Amplifying the patient's perspective

A blueprint to reduce low-value care by leveraging the potential of patient-reported outcomes and patients' health literacy

D I S S E R T A T I O N
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Management

submitted by
Anja Yvonne Bischof
from
Eggersriet (St.Gallen)

Approved on the application of

Prof. Dr. Alexander Geissler

and

Prof. Dr. Tobias Kowatsch

Dissertation no. 5422

Digitaldruckhaus GmbH, Konstanz / 2024

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science, hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, May 27, 2024

The President:

Prof. Dr. Manuel Ammann

Danksagung

Diese Dissertation ist das Ergebnis meiner dreieinhalbjährigen Reise als Doktorandin, in der ich viel über mich selbst, die Wissenschaft und den unschätzbaren Wert nahezu grenzenloser Unterstützung, Zuspruch und Liebe gelernt habe, die ich in dieser Zeit erfahren durfte.

Ich möchte meinem Doktorvater, Professor Dr. Alexander Geissler, und meinem Ko-Referenten, Professor Dr. Tobias Kowatsch, für ihre wertvollen Ratschläge, die wegweisenden Gespräche und ihr konstruktives Feedback während meines gesamten Doktorates meinen tiefsten Dank aussprechen. Ihr Fachwissen und ihre Ermutigung meinen eigenen akademischen Weg zu finden, spielten eine entscheidende Rolle in der Ausarbeitung meiner Ideen und der Verfeinerung meiner Forschung. Weiter danke ich Prof. Dr. Janna Hastings dafür, dass Sie Teil des Konsortiums war und meine Forschungsarbeit durch ihre kritische Reflexion und ihr Feedback bereichert hat.

Ebenfalls gebührt besonderer Dank dem Team am Lehrstuhl für Health Economics, Policy and Management für die bereichernde Zusammenarbeit und das grosse Engagement in unseren gemeinsamen Forschungsprojekten. Die anregenden Diskussionen und das kollegiale Miteinander haben wesentlich zum Gelingen dieser Arbeit beigetragen. Ebenfalls möchte ich Viktoria Steinbeck für die tolle Zusammenarbeit bei den beiden «Gender-Manuskripten» danken. Der Austausch war immer sehr bereichernd und aus einer anfänglichen kollegialen Zusammenarbeit ist Freundschaft entstanden.

Mein herzlichster Dank gilt auch meinen Freunden, die mich während der gesamten Zeit meiner Doktoratsreise unterstützt haben. Ihre Geduld, ihr Verständnis und ihre aufmunternden Worte haben mir in schwierigen Momenten Kraft gegeben. Besonders danke ich Henrietta Dorfmüller, Giuliana Breu, Marie Wosny und Matteo Bernasconi für die vielen Gespräche und ihre stetige Ermutigung. Sie haben mir geholfen, den Blick für das Wesentliche nie zu verlieren, stets motiviert zu bleiben und eine positive Einstellung zu behalten.

Mein unendlicher Dank gilt meinen Eltern, Yvonne und Erich, deren Glaube an mein Potential Antrieb für meine Motivation das Doktorat durchzustehen und abzuschliessen war. Ihre Geduld und stetige Unterstützung haben es mir erlaubt, ein Fundament zu schaffen, auf dem ich meinen akademischen Weg zielstrebig bestreiten konnte. Auch auf die Unterstützung meines Bruders Hendrik konnte ich immer zählen. Sein enormer Enthusiasmus hat mich über die Jahre immer beeindruckt und mich in meinem Handeln geprägt. Darüber hinaus danke ich ihm für seinen grossartigen Sinn für Humor, der mich oft zum Lachen gebracht hat und mir in schwierigen Zeiten Leichtigkeit und Glück geschenkt hat.

Meinen Grosseltern, Anny und Werner, bin ich zutiefst dankbar für ihren wohlwollenden Zuspruch und ihr Glaube an mich. Sie sind für mich noch heute grosse Vorbilder. Ihre Liebe und ihre Fürsorge haben mir auch in schwierigen Momenten Halt gegeben und mich dazu ermutigt, weiterzumachen.

Schliesslich möchte ich meinem Partner Dennis von ganzem Herzen für seine unerschütterliche Unterstützung und bedingungslose Liebe während dieser anspruchsvollen Zeit danken. Seine Geduld, die vielen ermutigenden Gespräche und sein Verständnis waren der Grundstein für meine Widerstandsfähigkeit. Ich bin zutiefst dankbar, dass ich dich an meiner Seite habe.

Diese Dissertation wäre ohne die grossartige Unterstützung und den Zuspruch dieser bemerkenswerten Menschen nicht möglich gewesen. Danke, dass ihr gemeinsam mit mir diesen Weg beschritten habt!

St. Gallen, Juni 2024 Anja Yvonne Bischof

Content Overview

List of dissertation papers XII

List of abbreviations XIV

List of figures XVI

List of tables XVIII

Summary XX

Zusammenfassung XXI

1 Introduction 1

2 Low-value care in routine and claims data 41

3 Enhancing the usefulness of PROs 91

4 Adoption of PROMs 228

5 Factors influencing health literacy 278

6 Results and outlook 325

Curriculum Vitae 358

Table of contents

List of dissertation papers XII

List of abbreviations XIV

List of figures XVI

List of tables XVIII

Summary XX

Zusammenfassung XXI

1 Introduction 1

1.1 Motivation and objectives 1

1.1.1 Low-value care indicates shortcomings in the current healthcare system 2

1.1.2 Patient-reported outcomes as a countermeasure to low-value care 5

1.1.3 Health literacy and its potential impact on counteracting low-value care 8

1.2 Goals and methods 9

1.2.1 Studies on low-value care 13

1.2.2 Studies on patient-reported outcome measures 17

1.2.3 Study on health literacy 24

References 28

2 Low-value care in routine and claims data 41

2.1 Making the cut on caesarean section: a logistic regression analysis on factors favouring caesarean sections without medical indication in comparison to spontaneous vaginal birth 41

2.1.1 Abstract 42

2.1.2 Background 44

2.1.3 Data and Methods 45
2.1.4 Results 48
2.1.5 Discussion 55
2.1.6 Conclusion 58
References 59
Appendix 63
2.2 Medication adherence halves COPD patients' hospitalization risk—Evidence from Swiss health insurance data 64
2.2.1 Abstract 65
2.2.2 Background 66
2.2.3 Methods 67
2.2.4 Results 72
2.2.5 Discussion 78
2.2.6 Conclusion 81
References 82
Appendix 86
3 Enhancing the usefulness of PROs 91
3.1 What is the association between gender and self-perceived health status when controlling for disease-specific conditions? A retrospective data analysis of pre- and post-operative EQ-5D-5L differences in total hip and knee arthroplasty 91
3.1.1 Abstract 92
3.1.2 Background 93
3.1.3 Methods 94
3.1.4 Results 98
3.1.5 Discussion 106
3.1.6 Conclusion 109
References 111

Appendix 118
3.2 Gender health gap pre- and post-joint arthroplasty: Identifying affected patient-reported health domains 135
3.2.1 Abstract 136
3.2.2 Background 138
3.2.3 Methods 139
3.2.4 Results 143
3.2.5 Discussion 153
3.2.6 Conclusion 156
References 157
Appendix 163
3.3 What matters when interpreting patient-reported outcomes?—Adjusted minimal important differences for hip and knee arthroplasty 168
3.3.1 Abstract 169
3.3.2 Background 170
3.3.3 Methods 171
3.3.4 Results 175
3.3.5 Discussion 194
3.3.6 Conclusion 197
References 198
Appendix 204
4 Adoption of PROMs 228
4.1 A collection of components to design clinical dashboards incorporating patient-reported outcome measures: Building on semi-structured interviews 228
4.1.1 Abstract 229
4.1.2 Background 230
4.1.3 Methods 231

4.1.4 Results 233
4.1.5 Discussion 243
4.1.6 Conclusion 245
References 247
Appendix 252
5 Factors influencing health literacy 278
5.1 The power of knowledge: A survey on COPD patients' health literacy in Switzerland 278
5.1.1 Abstract 279
5.1.2 Background 280
5.1.3 Methods 282
5.1.4 Results 286
5.1.5 Discussion 298
5.1.6 Conclusion 300
References 302
Appendix 309
6 Results and outlook 325
6.1 Discussion of results 325
6.2 Implications 336
6.3 Limitations 341
6.4 Outlook and recommendations for future research 344
6.4.1 Recommendations derived from Chapter 2 ("Low-value care") 346
6.4.2 Recommendations derived from Chapters 3 and 4 ("Patient-reported outcome measures") 347
6.4.3 Recommendations derived from Chapter 5 ("Health literacy") 348
6.5 Closing remarks 349

References 351

Curriculum Vitae 358

List of dissertation papers

Study	Title	Recommended Citation	Status
1	Making the cut on caesarean section: a logistic regression analysis on factors favouring caesarean sections without medical indication in comparison to spontaneous vaginal birth	**Bischof** AY, Geissler A (2023) Making the cut on caesarean section: a logistic regression analysis on factors favouring caesarean sections without medical indication in comparison to spontaneous vaginal birth. BMC Pregnancy Childbirth 23:.	Published
2	Medication adherence halves COPD patients' hospitalization risk – Evidence from Swiss health insurance data	**Bischof** AY, Cordier J, Vogel J, Geissler A (under review) Medication adherence halves COPD patients' hospitalization risk – Evidence from Swiss health insurance data. npj Primary Care Respiratory Medicine.	Published
3	What is the association between gender and self-perceived health status when controlling for disease-specific conditions? A retrospective data analysis of pre- and post-operative EQ-5D-5L differences in total hip and knee arthroplasty	**Bischof** AY, Steinbeck V, Kuklinski D, et al (2023) What is the association between gender and self-perceived health status when controlling for disease-specific conditions? A retrospective data analysis of pre- and post-operative EQ-5D-5L differences in total hip and knee arthroplasty. BMC Musculoskelet Disord 24:.	Published
4	Gender health gap pre- and post-joint arthroplasty: Identifying affected patient-reported health domains	Steinbeck, V, **Bischof** A, Schöner L, et al (under review) Gender health gap pre- and post-joint arthroplasty: Identifying affected patient-reported health domains. International Journal for Equity in Health.	Published
5	What matters when interpreting patient-reported outcomes? – Adjusted minimal important differences for hip and knee arthroplasty	Salvi I, Kuklinski D, Vogel J, **Bischof** A, Geissler A (under review) What matters when interpreting patient-reported outcomes? – Adjusted minimal important differences for hip and knee arthroplasty. Quality of Life Research.	Under review
6	A collection of components to design clinical dashboards incorporating patient-reported	**Bischof** AY, Kuklinski D, Salvi I, et al (under review) A collection of components to design clinical dashboards incorpoarting patient-reported outcome measures: Building	Under review

	outcome measures: Building on semi-structured interviews	on semi-structured interviews. Journal of Medical Internet Research.	
7	The power of knowledge: A survey on COPD patients' health literacy in Switzerland	**Bischof** A (2023) The power of knowledge: A survey on COPD patients' health literacy in Switzerland. Schriftenreihe in Health Economics, Management and Policy, No. 2023-04, Universität St. Gallen, School of Medicine, Lehrstuhl für Management im Gesundheitswesen, St. Gallen.	Published

The papers were reformatted for a consistent presentation. Figures and tables were numbered in a unified sequence, however the format remained according to the journals' requirements – which also applies for the labelling of the appendices. The text, however, was not changed.

List of abbreviations

ANOVA	Analysis of variance
ASA	American Society of Anesthesiologists
AUC	Area under the curve
BMI	Body mass index
CHOP	Swiss operation classification
COPD	Chronic Obstructive Pulmonary Disease
CS	Caesarean section
DHI	Digital health intervention
DRG	Diagnosis-related groups
e.g.	exempli gratia (for example)
EQ-5D-3L	European Quality of Life 5-Dimensions 3-Levels
EQ-5D-5L	European Quality of Life 5-Dimensions 5-Levels
EQ-VAS	European Quality of Life Visual Analogue Scale
f	Female
i.e.	id est (that is)
GDP	Gross domestic product
GP	General practitioner
HA	Hip arthroplasty
HL	Health literacy
HLQ	Health Literacy Questionnaire
HLS-EU-Q	European Health Literacy Survey Questinnaire
HLS-EU-Q16	16 items European Health Literacy Survey Questionnaire
HOOS-PS	Hip Disability and Osteoarthritis Outcome Score Physical Function Shortform
HRQoL	Health-related quality of life
KA	Knee arthroplasty
KOOS-PS	Knee Injury and Osteoarthritis Outcome Score Physical Function Shortform

LVC	Low-value care
m	Male
MID	Minimal important difference
NHS	National Health Service
NQRs	National Quality Registers
OECD	Organization for Economic Co-Operation and Development
OHS	Oxford Hip Score
OKS	Oxford Knee Score
p.	Page
PDC	Proportion of days covered
pp.	Pages
PRO	Patient-reported outcome
PROM	Patient-reported outcome measure
PROMIS-D-SF	Patient-Reported Outcomes Measurement Information System Depression-Short Form
PROMIS-F-SF	Patient-Reported Outcomes Measurement Information System Fatigue-Short Form
ROC	Receiver operating curve
RQ	Research question
S(1-7)	Study (referring to the corresponding study)
SF-36	Short Form 36
SILS	Single Item Literacy Screener
THA	Total hip arthroplasty
TKA	Total knee arthroplasty
WHO	World Health Organization
WOMAC	Western Ontario and McMaster Universities Arthritis Index

List of figures

Figure 1.1: Framework of this dissertation 12
Figure 2.1: Selection process of non-medically indicated CS based on a defined set of evaluation criteria 47
Figure 2.2: Schematic representation of the study design 70
Figure 2.3: Grouping of COPD patients according to the proportion of days covered (PDC) 72
Figure 2.4: Hospitalization likelihood due to exacerbations in relation with higher medication adherence 78
Figure 3.1: Patient flow diagram by follow-up time, procedure, gender and EQ-5D-5L response 96
Figure 3.2: Mean EQ-5D-5L total score and corresponding standard deviations for the patients in the total hip and total knee arthroplasty groups by gender over time 101
Figure 3.3: Association of gender with EQ-5D-5L dimensions of total hip arthroplasty over time 105
Figure 3.4: Association of gender with EQ-5D-5L dimensions of total knee arthroplasty over time 106
Figure 3.5: Participant flow 140
Figure 3.6: Hip arthroplasty: Explanatory power of the patients' gender on the z-score standardized PROs 149
Figure 3.7: Knee arthroplasty: Explanatory power of the patients' gender on the z-score standardized PROs 151
Figure 3.8: MIDs for hip replacement patients, unstratified and stratified by pre-operative EQ-5D index score—impact on the patient sample 188
Figure 3.9: MIDs for knee replacement patients, unstratified and stratified by pre-operative EQ-5D index score—impact on the patient sample 189

Figure 4.1: Framework summarizing the relationship between the individual 25 design principles and main concepts of clinical dashboards incorporating PROMs 242
Figure 5.1: Survey response rates 285
Figure 6.1: Key results 327
Figure 6.2: Key implications 337
Figure 6.3: Recommendations for future research 345

List of tables

Table 1.1: Definitions of low-value care (LVC) 3
Table 2.1: Number of total births, caesarean sections and spontaneous vaginal births in Switzerland per year 50
Table 2.2: Comparison of ordinal factors between non-medically indicated caesarean section and spontaneous vaginal birth 51
Table 2.3: Independent samples t-test for metric variables between women with non-medically indicated caesarean sections and spontaneous vaginal births 53
Table 2.4: Log odds ratio for factors associated with a non-medically indicated caesarean section 54
Table 2.5: Overview of data types and variables included in dataset 68
Table 2.6: Number of COPD patients, number and proportion of exacerbations, and mean of the model variables per category 75
Table 2.7: Regression output on the association of higher medication adherence with hospitalized exacerbations 76
Table 3.1: Descriptive statistics 99
Table 3.2: Comparison of the EQ-5D-5L dimensions by gender 103
Table 3.3: Descriptive statistics of the study population (n=6,803) 145
Table 3.4: Descriptive statistics for hip and knee replacement samples 177
Table 3.5: Regression results on 6-months EQ-5D-5L index score improvement for hip replacement sample 182
Table 3.6: Unstratified and stratified MIDs for hip and knee replacement according to “pre-operative EQ-5D index score” 186
Table 3.7: Unstratified and stratified MIDs for hip and knee replacement according to “pre-operative EQ-5D index score”, “depression”, and “gender” 192
Table 4.1: Overview of software producer and user responses on how data collection should be implemented in the clinical dashboard 235
Table 5.1: Descriptives of the survey participants 287
Table 5.2: Differences in variables’ characteristics regarding health literacy 290

Table 5.3: Differences in mean values regarding health literacy..... 293
Table 5.4: Regression table for identifying associations with health literacy..... 295
Table 5.5: Difficulty in dealing with COPD, wish for support in daily life, and support with symptoms per HL group..... 297

Summary

Rising costs in the Swiss healthcare system are a pervasive topic today. However, the quality of care remains on a relatively stable level in Switzerland. The question, therefore, arises whether all services provided in the healthcare system are to the patient's benefit or whether treatments are provided that do not benefit the patient, cause harm, incur additional costs, and waste scarce healthcare resources—a phenomenon known as low-value care (LVC).

Against this backdrop, this dissertation proposes a framework based on three pillars to identify and tackle LVC. The ultimate research objective is to amplify the patient's perspective to impede LVC.

The *first pillar* considers the identification of LVC in routine and claims data. Here, a novel approach for the identification of LVC is developed. The *second pillar* focuses on the enhanced usefulness of patient-reported outcomes (PROs) and how their adoption in clinical practice can be further enhanced. Here, the application of PROs is examined and it is shown what is needed for their accurate interpretation and adoption in clinical practice for utilizing these measures to counteract LVC. The *third pillar* completes the framework by investigating the patient's health literacy (HL) and associated factors. HL is one primary lever for increased self-management capabilities. Improved HL empowers patients to adhere to treatment regimens, averting adverse outcomes and mitigating instances of LVC. Within this pillar it is shown which factors are associated with the patient`s HL and thereby enables for possible management of HL.

Thus, this dissertation provides a comprehensive framework for addressing the identification of LVC and how the patient's perspective, instantiated in PROs and HL, can be leveraged to effectively resolve the challenges of LVC. As this framework represents a proposition, it is recommended that future research endeavors delve into the nuanced mechanisms linking LVC with PROs and HL. A thorough exploration of these relationships will facilitate the development of a sustainable, patient-centered approach to counteract LVC effectively.

Zusammenfassung

Steigende Kosten im Schweizer Gesundheitssystem sind heutzutage ein allgegenwärtiges Thema. Die Versorgungsqualität in der Schweiz verbleibt jedoch auf einem relativ stabilen Niveau. Folglich stellt sich die Frage, ob alle im Gesundheitssystem erbrachten Leistungen dem Patienten nützen oder ob sie ihm sogar schaden, zusätzliche Kosten verursachen, oder knappe Gesundheitsressourcen verschwenden—ein Phänomen, das als Low-Value Care (LVC) bezeichnet wird.

Vor diesem Hintergrund wird in dieser Dissertation ein Modell zur Identifikation und Verhinderung von LVC vorgeschlagen, das aus drei Säulen besteht. Das Forschungsziel ist, die Patientenperspektive stärker zu nutzen, um LVC zu vermeiden.

Die *erste Säule* befasst sich mit dem Identifizieren von LVC in Routine- und Leistungsdaten. Es wird ein neuer Ansatz zur Identifikation von LVC entwickelt. Die *zweite Säule* konzentriert sich auf den erhöhten Nutzen von patientenrapportieren Ergebnissen (PROs) und der Verbesserung ihrer Einsatzfähigkeit in der klinischen Praxis. Ihre Anwendung wird untersucht und aufgezeigt, was für die Interpretation und Übernahme in die klinische Praxis zur Verhinderung von LVC erforderlich ist. Die *dritte Säule* vervollständigt das Modell und untersucht die Gesundheitskompetenz (HL) der Patienten und die damit verbundenen Faktoren. HL ist ein primärer Hebel für die Verbesserung der Selbstmanagementfähigkeiten. Eine verbesserte HL befähigt die Patienten, sich an die Behandlungsschemata zu halten, unerwünschte Ergebnisse zu vermeiden und Fälle von LVC zu vermindern. Es wird aufgezeigt, welche Faktoren mit der HL des Patienten zusammenhängen und dadurch deren Management ermöglichen.

Diese Dissertation bietet somit ein Modell, welches sich mit der Identifikation von LVC befasst und zeigt, wie die Patientenperspektive, in Form von PROs und HL, genutzt werden kann, um das Problem von LVC anzugehen. Es handelt sich bei dem Modell um einen ersten Ansatz. Zukünftige Forschungsbemühungen sollten sich nuancierter mit den verbindenden Mechanismen zwischen LVC, PROs und HL befassen, um die Entwicklung eines nachhaltigen, patientenzentrierten Ansatzes zur wirksamen Bekämpfung von LVC zu erleichtern.

1 Introduction

1.1 Motivation and objectives

The Swiss healthcare system is severely pressured by soaring costs, as reflected directly in rising health insurance premiums [1–3]. In autumn 2023, the Federal Council announced a premium increase of 8.7% compared to 2022 [4]. Explanations for the increasing costs are manifold: the aging population, more outpatient treatments, more cost-intensive medications, inaccurate incentive mechanisms in the healthcare systems based on prevailing remuneration models, or increased treatment inefficiencies [4, 5]. According to the Organization for Economic Co-Operation and Development (OECD), the Swiss healthcare system is the sixth most expensive system in terms of GDP [6]. In 2022, the United States of America topped the list (16.6%), followed by Germany (12.7%), France (11.9%), Japan (11.5%), and Austria (11.4%). In the same year, Switzerland spent 11.3% of its GDP on healthcare.

As Swiss healthcare costs are relatively high, the healthcare system's aim must be to provide good quality of care to sustain the population's health-related quality of life (HRQoL). Swiss people's self-perceived health status is among the highest compared to other OECD countries. Switzerland ranks fifth (for females) and fourth (for males) by OECD comparison. Of female inhabitants over the age of 15, 80.5% perceive their health as good or very good, while 83.3% of their male counterparts report the same perception [6]. HRQoL and quality of care are interlinked, as the quality of care influences how the patient recovers and afterward perceives the personal health status—measured by the HRQoL.

As stated in the revised Health Insurance Act of 2007, the Swiss healthcare system aims to increase or maintain the quality of care while slowing the steadily increasing costs [7]. The World Health Organization defines quality of care as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes" [8]. Quality of care can be measured according to three main categories: structure, process, and outcome quality [9]. Structural quality indicators

provide information about the systems, processes, and capacities that healthcare providers put in place to provide high quality of care (e.g., accreditation and certification, staff qualifications, facility infrastructure). Process quality indicators deal with the actions taken by healthcare providers to sustain or enhance health, whether for healthy individuals or for those diagnosed with a medical condition (e.g., timeliness of care, adherence to medical guidelines, medication safety)..[1] Outcome quality indicators assess the impact of a certain treatment or intervention on an individual's health status (e.g., complication rates, mortality rates, readmission rates, patient satisfaction). Patients' perception of good quality of care is mostly influenced by process quality, followed by structural quality [10].

If structural or process quality is insufficient, low-value care (LVC) may occur because patients do not experience treatment as beneficial, while additional, unnecessary costs are incurred and limited resources are used.

1.1.1 Low-value care indicates shortcomings in the current healthcare system

LVC was first discussed in the late 1990s when a roundtable of US medical experts concluded that care structures are not optimal for providing high-quality care. Back then, this phenomenom was not called LVC but "overuse." The medical experts found that overuse "occurs when a healthcare service is provided under circumstances in which its potential harm exceeds the possible benefit" [11]. At the time, the US faced large-scale problems with the inappropriate use of treatments, which led to direct negative consequences for patients. To quantify this problem, the above-mentioned medical experts proposed to measure the value of healthcare services by the achieved health benefit per dollar spent. Accordingly, they established three main principles:

[1] In this dissertation, indication and process quality are subsumed under *process quality*. A physician's adherence to medical guidelines starts when examining and diagnosing a patient. Both, examining and diagnosing, belong to the indication process and will be covered by process quality in this dissertation.

1) Providing effective care to potential beneficiaries; 2) avoiding inappropriate care; and 3) preventing foreseeable complications. Based on these principles, Chassin et al. [11] claimed that assessing quality of care requires considering all treatment processes and the outcome of care. However, in this first definition of LVC, no instruments to measure quality of care or the patient's health benefit were provided, but described the phenomenon in superficial terms. Further definitions have since been developed (see Table 1.1).

Table 1.1: Definitions of low-value care (LVC)

Definition	Source
"Low-value care involves the use of care that is unlikely to benefit the patient given the cost, available alternatives, and preferences of the patient."	[12]
"An intervention in which evidence suggests it confers no or very little benefit for patients, or risk of harm exceeds probable benefit, or, more broadly, the added costs of the intervention do not provide proportional added benefits."	[13]
"[…] raises costs, causes iatrogenic patient harm, and often interferes with the delivery of high-value care."	[14]
"Care that is unlikely to benefit the patient given the harms, cost, available alternatives, or preferences of the patient."	[15]

These definitions highlight the limited benefit of treatment and the potentially resulting iatrogenic harm (i.e., harm unintentionally induced by a physician or medical treatment). Other unfavorable consequences include excessive costs and disproportionate resource consumption. However, treatments identified in one context as LVC can provide adequate value in another setting: that is, the amount of value provided varies between patients and across patient groups [16].

LVC covers a broad range of treatments, which—for various reasons— do not create value for the patient. It therefore needs to be distinguished by type. Three main types of LVC are evident: 1) *inefficient care* (i.e., treatment proven to be inefficient in certain situations); 2) *ineffective care* (i.e., care proven to be (cost)ineffective in a specific case); and 3) *unwanted care* (i.e., the provided care is effective but does not fit the patient's preferences) [15]. Especially inefficient care may be seen as one major driver of the overuse of healthcare resources as services are delivered to patients who

do not belong to the population benefitting from those services. This phenomenon is also known as *indication creep* [13].

To reduce the number of LVC practices, various de-implementation approaches have been proposed [15, 17–20]. *De-implementation* describes the process of eliminating ineffective and harmful medical care to improve patient outcomes and alleviate the uncurbed rise of healthcare costs [20, 21]. However, LVC de-implementation should not lead to any compromises in patient outcomes due to lower quality of care [22]. Chan et al. [23] criticized that early research lacks specific definitions of how to measure overuse of healthcare services; however, the lack of a clear definition is a common problem of performance measurement in healthcare systems [24]. One approach that has been promoted involves differentiating *"what intervention works for whom under which circumstances"* [25]; thus a treatment's effectiveness is not doubted but that treatment must be applied in the right context. Building on the above understanding of LVC, this dissertation defines LVC services as follows: *"LVC are services that, in an individual context, (a) provide little or no benefit to patients, (b) potentially cause harm, (c) incur unnecessary costs, and/or (d) waste scarce healthcare resources compared to the use of alternative services or the application in an alternative patient group."*

To classify healthcare services as LVC, professional guidelines need to be developed. Such guidelines require to include the assessment of certain context where services may be classified as LVC, to enable for the formulation of effective de-implementation recommendations. One approach for formulating guidelines to de-implementing LVC is the *Choosing Wisely* movement [26]. It was founded in the US in 2012. Medical societies developed lists with treatments that have empirically been proven to create inappropriate value for patients in specific situations. Switzerland adopted this concept in 2017 and called it *Smarter Medicine* [27]. Although they adopted some *Choosing Wisely* recommendations, most Swiss medical societies also developed their own recommendations [28]. In contrast to *Choosing Wisely*, which does not limit the number of recommendations, *Smarter Medicine* provides "Top-5" lists of the different medical specialties. In the last decade, more countries (e.g., Italy,

Canada, and Japan) have adopted the *Choosing Wisely* approach in an attempt to de-implement LVC services [29].

This one-sided approach to reducing LVC treatments (i.e., only considering the professional perspective) neglects the potential to achieve sustainable results by ignoring patients and their subjective evaluation of treatment success. A 2019 review of studies on LVC de-implementation revealed that most studies target providers by applying education mechanisms (69%), while only a small number of studies focus on patients (13%). Additionally, two out of five interventions measured the treatment outcome based on hospitalization, length of stay, intensive care unit admission, or mortality. Measuring patient-reported outcomes (PROs) such as quality of life or patient satisfaction was only considered in 8% of all studies [30].

As research has shown, raising awareness on the importance of de-implementing LVC, and educating medical experts is not sufficient to achieve sustainable results [12]. To use a second lever for reducing LVC, the patient should be considered more actively. For that, patients and their experiences need to be actively engaged in the shared decision-making with their physician [31] and research teams should assess their study outcomes according to clinically important and patient-centered measures [30] to reduce LVC. This goal might also be achieved by including patient-reported outcome measures (PROMs). PROs may support the physician's understanding of the patient's needs and thereby avoid preference misdiagnosis, which leads to LVC (i.e., physicians providing a service unsuitable to the individual patient's needs) [16].

1.1.2 Patient-reported outcomes as a countermeasure to low-value care

Using PROMs in healthcare is key to ensuring the transition to patient-centered care [32] and to prioritizing health outcomes that matter to patients [33]. PROMs assess the patient's self-reported health status, functional impairments, and disabilities. Additionally, the nature and severity of symptoms can be reported by means of standardized survey instruments [34, 35]. Incorporating the patient perspective into healthcare systems by capturing PROs offers considerable potential to enhance patient outcomes and quality of life [36–39] and to use PROs as an additional lever to reduce

LVC. Collecting PROs complements traditional clinical data by representing a patient's assessment of symptoms, physical function, and health-related quality of life (HRQoL), independently of a physician's interpretation or that of other healthcare professionals [40, 41].

Patient-generated health-related data have a threefold potential: (1) to enable patients to actively engage in their healthcare; (2) to provide healthcare providers with more insight into their patients; and (3) to ensure more meaningful and structured patient-provider communication [34, 35, 39, 42, 43]. Growing evidence suggests that collecting patients' HRQoL increases healthcare providers' awareness of patients' concerns [38, 44] and enhances patient-physician communication without increasing visit length [45–48]. PROs not only reflect the patient's health status but also visualize treatment effectiveness [49, 50]. Further, they enable patients to evaluate their health status in a valid, reliable, and systematic manner [51–53]. PROs might even encourage patients to address health issues that are usually not covered in consultations (e.g., mental health [44]). However, standardized PRO questionnaires will integrate such health issues into patient-physician communication [54, 55].

Using PROs can guide individualized patient care [56] as they enable setting goals for improving health status [42, 57]. Consulting PROs will also allow medical experts to discuss patients' treatment expectations and to identify incompatibilities early on [58].

Three main principles need to be followed to increase the likelihood of PROs being used in clinical practice: (1) short and simple [59–61]; (2) easy to interpret [62], and (3) available in multiple languages [63]. PROMs can be categorized in several ways. Al Sayah et al. [64] proposed a taxonomy of PROMs based on (1) focus, (2) scope, and (3) type of measurement:

(1) The focus of the measurement indicates how health outcomes are measured. Patient-reported health outcomes may include symptom burden, functional status, or health-related quality of life [65].

(2) The scope of measurement consists of generic and disease-specific. Generic PROMs (e.g., EQ-5D-5L [66, 67], SF-36 [68, 69], or PROMIS [70]) are

designed to assess a person's health independent of a particular disease. Especially beneficial is their capability to compare people's health status across a broad range of diseases. However, changes in specific areas of health are easily missed by applying generic PROMs. Generic PROMs are complemented by disease-specific PROMs (e.g., WOMAC [53] or HOOS/KOOS [71–73] for joint arthroplasties or CAT [74, 75] for COPD). Disease-specific PROMs include disease-specific health issues. Therefore, Al Sayah et al. [64] recommended combining generic and disease-specific PROMs to enlarge the scope of assessment as broadly as possible.

(3) The type of measurement accounts for how PRO scores are calculated. Scores are either profile- or preference-based. The profile approach (applied in SF-36, or PROMIS) enables assessing one or more health domains with one or more questions [76]. Profile measures are beneficial when assessing health or a related aspect on the quality of life to evaluate treatments from a patient's perspective. Preference-based instruments (represented in the EQ-5D-5L) foreground health states or conditions and are usually calculated in a single multi-domain index score. Besides assessing the patient's perspective on the health-related aspects of a specific treatment, preference-based instruments are suitable to generate utility values for calculating quality-adjusted life years for health economic evaluations [64].

The PROM instruments need to be implemented in clinical practice to strengthen their real impact based on analyzing and interpreting PROs. Currently, PROMs are primarily used in research approaches (e.g., Kuklinski et al. [77], Kuklinski et al. [78], Hertler et al. [79], Lipscomb et al. [80], Berry et al. [81], Berry et al. [82]), but the literature advocates an increased use of PROs in clinical practice [83–85].

Today, manifold challenges inhibit applying PROMs: IT systems must be adapted [86, 87]; PRO visualization and interpretation need to be simplified [85]; physicians remain skeptical about the added value [88] and fear a lack of organizational support [86, 87]. Despite these barriers, the literature shows that using PROMs positively

impacts process outcomes (e.g, more appropriate patient-physician communication, fewer outpatient visits) [81, 89–93].

1.1.3 Health literacy and its potential impact on counteracting low-value care

Rigorous patient-physician communication is key to providing effective and efficient treatment. Therefore, patients must be able to assess and evaluate healthcare information and act on treatment prescriptions. This patient capability is known as health literacy (HL). HL denotes the ability to find, understand, and effectively use health information and health services while promoting health and preventing disease [94, 95]. HL includes various skills (e.g., reading, numeracy, critical thinking, and effectively navigating the healthcare system) [96]. Sørensen et al. [97] identified three main application areas of health literacy: disease management, disease prevention, and health promotion. Each area consists of four components concerning health information: 1) access/obtain, 2) understand, 3) process/appraise, and 4) apply/use.

HL is assessable by various internationally validated instruments: the Health Literacy Questionnaire (HLQ) [98], the European Health Literacy Survey Questionnaire (HLS-EU-Q) [97], the 16-item European Health Literacy Survey Questionnaire (HLS-EU-Q16) [99], or the Single-Item Literacy Screener (SILS) [100]. These instruments differ by measurement type (i.e., self-reporting vs testing of capabilities and knowledge), scope (i.e., generic, disease-specific, or e-health literacy) or by target group (i.e., age- or culture-specific adjustments) [101].

Today, patient-physician communication is potentially marred by miscommunication. Physicians and nurses often overestimate patients' HL skills by assuming that patients can process complex instructions [102–104]. This problem worsens because patients also tend to overestimate their HL [104, 105]. Overestimating one's abilities is also known as the *Dunning-Kruger effect* [106]. Besides verbal communication, written disease-related educational material remains inaccessible to patients with low HL by demanding advanced reading skills or by being poorly written [107, 108].

Potential patient-physician miscommunication makes collecting information about a patient's HL crucial for two reasons: First, it can improve patient-physician communication and allow physicians to tailor information to ensure patients understand complex medical information [109]. Second, higher HL enables patients to participate in shared decision-making and is associated with better treatment adherence and enhanced patient outcomes [110–115].

In contrast, lower HL fosters adverse outcomes such as a higher rate of revisits to emergency departments (e.g., Shahid et al. [116]) or more complications (e.g., Yee et al. [117]). Therefore, low HL might lead to LVC as avoidable complications occur, which do not benefit the patient, cause additional harm, require using resources to counteract complications, and generate additional treatment costs. Hence, low HL covers the four pillars of LVC.[2] Hence, improving low HL to adequate HL might help avoid (unintentionally) patient-inferred detrimental effects and, consequently, LVC.

1.2 Goals and methods

This dissertation proposes a set of levers for reducing LVC by explicitly including PROs and for basing physicians' treatment decisions on patients' HRQoL. Reducing LVC and thereby improving quality of care requires considering two levels: (1) the healthcare system (represented by individual physicians) and (2) patients. On the one hand, physicians need to suggest and implement the most appropriate treatments; on the other, patients need to act on treatment plans to leverage the treatment's full potential.

Achieving the research goal of amplifying the patient's perspective to reduce LVC rests on three building blocks. First, LVC must be detectable in routine and claims

[2] The four pillars of LVC cover services that (a) provide no or little benefit to the patient, (b) potentially cause harm, (c) incur unnecessary costs, and (d) waste limited healthcare resources (see Chapter 1.1.1).

data for future research efforts to determine whether LVC can be decreased by applying the proposed measures. Second, the patient's perspective must be collected via PROMs. However, concrete guidelines on appropriately interpreting PROs are currently lacking but are needed to leverage the full potential of PROs in clinical practice. Additionally, using PROs in clinical practice needs to be further promoted. One potential approach is to enhance the display of PROs with digital tools such as clinical dashboards. PROs strengthen the patient's representation of their health status, bolstering treatment changes, which are derived directly from the patient's HRQoL. Third, patients benefit from improved HL, leading to better self-management capabilities and increased treatment and medication adherence, ultimately reducing LVC.

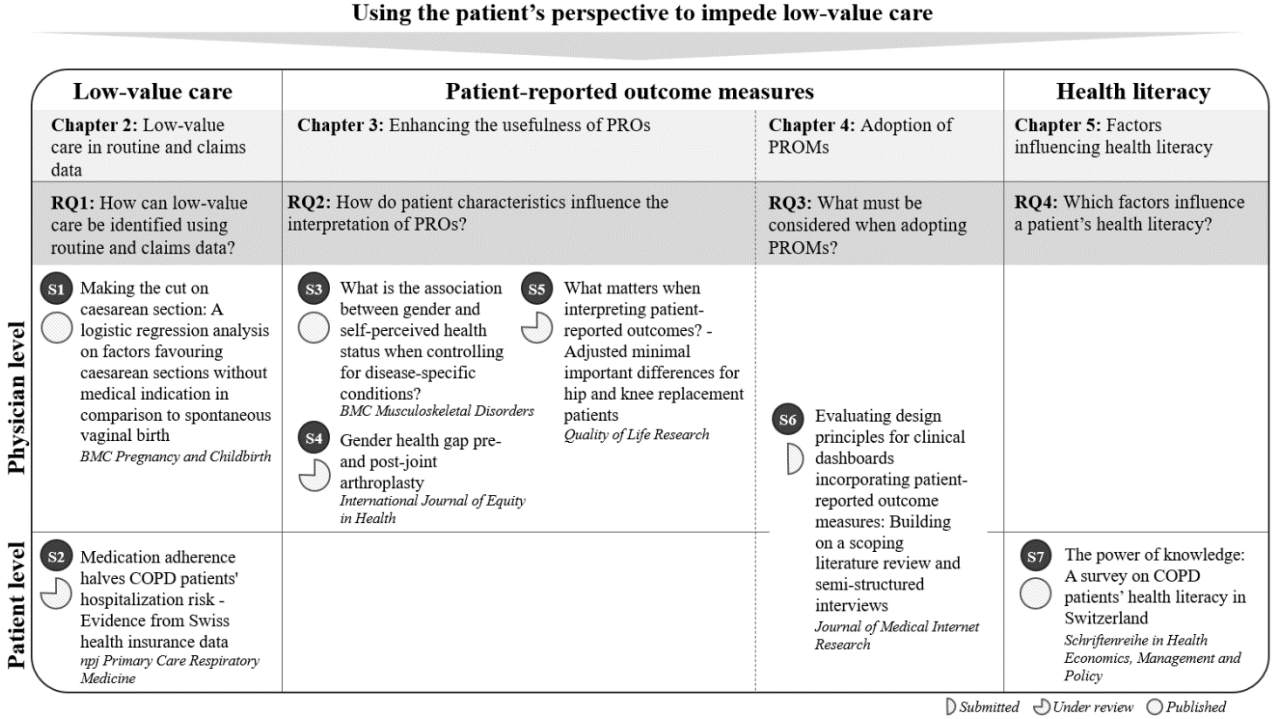
To address the three building blocks (see Figure 1.1), this dissertation proceeds as follows: Chapter 2 identifies LVC based on analyzing routine and claims data (1st building block). Only if LVC is detectable in routine or claims data de-implementation efforts can become quantifiable and hence trackable. The goal is to analyze whether LVC can be identified in routine and claims data and which identification approaches apply. Chapter 2 asks (RQ1): *"How can low-value care be identified using routine and claims data?"* To address RQ1, Chapter 2 presents two studies, one focusing on the physician level, the other on the patient level.

Chapters 3 and 4 address the use of PROs (2nd building block). Chapter 3 considers how to make PROs even more useful. As PRO interpretation relies on patient characteristics, this chapter explores what guides the interpretation of PRO results. It therefore asks (RQ2): *"How do patient characteristics influence the interpretation of PROs?"*. Chapter 4 analyzes what promotes increased PRO adoption in clinical practice and therefore asks (RQ3): *"What must be considered when adopting PROMs?"*. Answering RQ2 (Chapter 3) and RQ3 (Chapter 4) provides a robust basis for better managing expectations about treatment success and for analyzing patients' health status.

Chapter 5 shifts from the physician level to the patient's HL. HL is integral to self-management capabilities and thus enables patients to cope with disease-specific

symptoms. The better patients can manage their disease, the less likely adverse outcomes become, thus lowering the risk of LVC. Chapter 5, then, investigates whether HL enables identifying support needs and which factors influence HL (3rd building block). It therefore asks (RQ4): *"Which factors influence a patient's health literacy?"*. Raising awareness of factors influencing patients' HL may indicate which future research is needed to enhance patients' self-management capabilities, resulting in less patient-induced LVC.

Figure 1.1: Framework of this dissertation



Legend: S = Study, RQ = Research Question, PRO = Patient-reported outcome, PROM = Patient-reported outcome measure, COPD = Chronic Obstructive Pulmonary Disease

What follows outlines the seven individual studies and their contribution to the overall framework (see Figure 1.1).

1.2.1 Studies on low-value care

Study 1 (S1) ("Making the cut on caesarean section: A logistic regression analysis of factors favouring caesarean sections without medical indication in comparison to spontaneous vaginal birth") has been published in *BMC Pregnancy and Childbirth*..[3] It presents an identification approach to non-medically indicated caesarean sections (CS) from routine data and compares the characteristics of women giving birth by non-medically indicated CS with women giving birth spontaneously. Non-medically indicated CS may be classified as LVC as the risk of morbidity and mortality increases for the mother and the neonate if spontaneous vaginal birth (SVB) was feasible. It also causes higher costs compared to SVB and involves more resources (e.g., staff and operation rooms). This raises the question whether non-medically indicated CS are identifiable in routine data and how high their annual share compared to the overall birth rate is. Such an identification would help tackle LVC as information is missing on how many CS are on a non-medically indicated basis nowadays. The current challenge in identifying non-medically indicated CS from routine data arises from the unavailability of a clear classification code based on diagnoses-related groups (DRG), diagnoses, or procedures. This chapter therefore asks:

- How can cases of non-medically indicated CS be identified in routine data?
- Which factors make giving birth by non-medially indicated CS more likely?

To investigate these research questions, we retrieved the national dataset "Medical Statistics of Hospitals" from the Federal Statistical Office. This dataset includes 98.3% of all women who gave birth in Swiss hospitals between 2014 and 2018. We

[3] The co-authors of the corresponding study are denoted at the beginning of each study in Chapters 2 to 5.

enriched the dataset with information on net income per capita, nursing staff per hospital, and midwives per 1,000 births, all annually and at the cantonal level, to control for their impact on the decision of the mode of birth. We also added the urbanity level of the living area (i.e., urban, suburban, or rural) to the dataset. We developed a three-step approach based on diagnosis-related groups (SwissDRG), diagnoses, and procedure classification (specified by the Swiss operation procedure (CHOP) catalog) to identify non-medically indicated CS. By applying this approach, we identified 11,112 non-medically indicated CS between 2014 and 2018. To build the comparison group, we used all cases (identified through SwissDRG) where women gave birth spontaneously without any complicated or complex diagnoses being noted in the dataset. Thus, 176,908 births were spontaneous within the investigated timeframe. Using the two identified groups, women were described by age, living areas, insurance class, and urbanity level.

To test whether the mean values of the two investigated groups are significantly different, we applied an independent-samples t-test to the variables of midwives per 1,000 births, nursing staff per hospital bed, and income per capita. By further applying a binary logistic regression analysis, we determined the impact of the above factors (i.e., age, living area, insurance class, urbanity level, number of midwives per 1,000 births, number of nursing staff per hospital bed, and net income per capita) on non-medically indicated CS.

S1 shows that our three-step approach enables identifying non-medically indicated CS from routinely collected data. Almost 8% of all conducted CS are non-medically indicated in Switzerland and thus can be classified as LVC. Women giving birth by non-medically indicated CS tend to be older, have supplemental insurance, and live in urban areas. Additionally, significant differences ($p<0.001$) between the mean values of women in either of the two birth modes are observable. Women giving birth by non-medically indicated CS live in cantons with fewer midwives per 1,000 births, a higher number of nursing staff per hospital bed, and a higher net income per capita. The logistic regression analysis reveals that higher age, supplemental insurance, higher income, and more available nursing staff in the hospital increase the likelihood

of giving birth by non-medically indicated CS. In contrast, women living in areas with a higher density of practicing midwives are significantly less likely to give birth by non-medically indicated CS.

In conclusion, S1 develops an approach for identifying LVC (in non-medically indicated CS) in routine data by including contextual factors. Applying this approach enables measuring—for the first time—LVC based on routine CS data. The results show that 8% of CS in Switzerland are non-medically indicated and hence can be categorized as LVC. By developing a novel identification approach for LVC and applying it to Swiss data, S1 provides a basis for discussing one lever (i.e., the identification of non-medically indicated CS) to systematically tackle the problem of LVC in Switzerland.

Study 2 (S2) is titled "Medication adherence halves COPD patients' hospitalization risk: Evidence from Swiss health insurance data" (under review, *npj Primary Care Respiratory Medicine*). It analyzes medication adherence among patients suffering from Chronic Obstructive Pulmonary Disease (COPD) in Switzerland and discusses how patients' low medication adherence impacts the likelihood of hospitalized exacerbations. LVC might be caused not only by physicians but also by patients not adhering to treatment and medication plans, leading to avoidable, cost-intensive additional treatments. However, until now, no study has proven that medication non-adherence leads to hospitalized exacerbation. S2 therefore asks:

- How is medication adherence distributed among COPD patients?
- How does inadequate medication adherence impact the likelihood of a hospitalized exacerbation?

The study is based on 2015–2020 data provided by a Swiss health insurance. The dataset includes each insured person's basic claims data, contractual information, and billing data. Billing data encompasses information on treatments and prescriptions during the observation period. We included 13,557 insured persons in the dataset. As Swiss health insurances are not allowed to store diagnosis data, we identified insured persons suffering from COPD via three inclusion criteria: 1) treatment with long-

acting COPD-specific medications; 2) over 40 years of age; and 3) claims data available for at least two consecutive years prior to the hospitalized exacerbation.[4]

To approximate COPD patients' medication adherence, we calculated their proportion of days covered (PDC). PDC represents the days covered by COPD patients' home inventory of long-acting medication taken during the observation period. Therefore, PDC reveals the distribution of medication adherence among Swiss COPD patients, which we classified into five categories. Category 1 includes COPD patients with a PDC of 0–20%, category 2 covers COPD patients with a PDC >20–40%, category 3 encompasses COPD patients with a PDC >40–60%, category 4 represents COPD patients with a PDC >60–80%, and category 5 includes the remaining COPD patients.

We applied the causal Rubin effect to analyze the effect of medication adherence (represented by the PDC) on the likelihood of a hospitalized exacerbation. Therefore, we quantified the average treatment effect of the PDC on hospitalized exacerbations. This procedure allowed us to estimate the probability of being hospitalized with an exacerbation per COPD patient based on the corresponding PDC category.

S2 shows that about half of the included COPD patients have a PDC of less than 40%. Thus, less than 40% of the days in the observation period are covered by a high enough personal home inventory of long-acting medications. If patients do not take their medication as prescribed, the medication cannot be fully effective, increasing the risk of adverse outcomes (e.g., hospitalized exacerbations classifiable as LVC).

Only 22.7% of all included patients qualify for category 5, representing a PDC of 80–100%. Concerning the rate of hospitalized exacerbations, the lowest rate occurs in the patient category with the highest medication adherence (PDC >80–100%), while most exacerbations are recorded in the patient category with the lowest PDC (0–20%). Furthermore, the regression analysis reveals that COPD patients with high medication adherence (PDC >80-100%) are almost half as likely to suffer from hospitalized

[4] The reasons for these three inclusion criteria are further elaborated in Chapter 2 (Study 2).

exacerbation compared to COPD patients with low medication adherence (PDC = 0–20%). Thus, S2 indicates that low medication adherence in patients may lead to LVC in the form of an increased risk of hospitalized exacerbation.

1.2.2 Studies on patient-reported outcome measures

While Chapter 2 focuses on LVC and whether it is identifiable in routine and claims data, Chapters 3 and 4 explore which factors enhance the usefulness of PROs and how clinical dashboards potentially improve patient-physician communication through PROMs. Study 3 (S3, "What is the association between gender and self-perceived health status when controlling for disease-specific conditions? A retrospective data analysis of pre- and post-operative EQ-5D-5L differences in total hip and knee arthroplasty") has been published in *BMC Musculoskeletal Disorders*. It examines the validity of using the EQ-5D-5L index score as a tool to interpret gender-related differences in outcomes among THA and TKA patients. These gender disparities highlight the need to discuss whether treatment adaptions are possible based on changes in PROM scores over time and what needs to be respected. S3 therefore asks:

- What are the differences in the association between gender and the EQ-5D-5L dimensions in patients undergoing total hip and knee arthroplasty pre- and post-surgery?

The dataset consisted of data collected in two German hospitals (*Schoen Clinic Hamburg Eilbek* and *Schoen Clinic Neustadt*) and included all patients who underwent elective primary total hip or knee arthroplasty between January 2016 and December 2018. In total, 2,368 THA patients (m=978; f=1,390) and 1,629 TKA patients (m=715; f=914) were included. Data from several sources were merged: PROM data were collected at patient admission to the hospital, as well as three and 12 months post-surgery via postal service. Both PROM types (generic and disease-specific) were collected. The generic PROM is represented by the EQ-5D-5L, which consists of mobility, self-care, usual activity, pain/discomfort, and anxiety/depression and measures the patient's HRQoL. As a disease-specific PROM, we applied

WOMAC (Western Ontario and McMaster Universities Arthritis Index), which encompasses disease-related aspects (i.e., stiffness, pain, and functionality). Additionally, routine data on patient characteristics (e.g., sociodemographic data, further medical conditions, and surgical details) enriched the dataset.

We described the differences between gender and surgical procedures by independent-samples t-test or Mann-Whitney U-test depending on the variable's characteristics and distribution. Additionally, we applied a cumulative odds ordinal logistic regression with proportional odds. We analyzed every dimension of the EQ-5D-5L at each of the three measuring points (i.e., at admission, three months post-surgery, and 12 months post-surgery). The control variables included age, the Elixhauser Comorbidity Index, walking distance, walking aid pre-surgery, and surgery duration. Furthermore, we included the WOMAC index score to adjust for disease-specific conditions that might influence the generic health state.

S3 shows that gender is significantly associated with the individual dimensions of the EQ-5D-5L. Female patients have significantly higher self-care scores (THA and TKA), regardless of the measuring point, and better post-operative mobility scores (TKA). Contrarily, male patients score significantly higher for reduced pain/discomfort (TKA) and decreased anxiety/depression (THA) prior to surgery and three months after surgery. These results suggest that considering the EQ-5D-5L index score does not provide enough information on where gender differences originate. The individual dimensions provide more comprehensive and functional insights, which physicians may use to adapt treatment to the patient's needs. Therefore, this study functions as a baseline that gender differences in PROMs must be considered when using PROMs in daily practice and hence adds to the enhanced interpretability of PROs.

Complementing S3, Study 4 (S4, "Gender health gap pre- and post-joint arthroplasty: Identifying affected patient-reported health domains"; under review at the *International Journal of Equity in Health*) evaluates whether a gender health gap exists in hip arthroplasty (HA) and knee arthroplasty (KA) pre- and post-surgery and

whether differences concerning over-time improvement are observable. S4 therefore asks:

- Does a gender health gap exist pre-surgery, at discharge, and 12 months post-HA and KA?
- Does a gender health gap exist in the PRO improvement between admission and 12 months post-HA and -KA?

This retrospective cohort study analyzed data from the *PROMoting Quality* study [118], which included 3,693 HA (1,627 male, 2,066 female) and 3,110 KA (1,430 male, 1,680 female) patients from nine German hospitals in a randomized controlled trial between 2020 and 2021. The study followed patients until March 2022, with patients completing PROM questionnaires at admission, discharge, and 12 months post-surgery. The study assessed the impact of the PROMoting Quality intervention, which began one month post-surgery, alongside other variables, on PROs and gender-related health disparities. For further information on the study design, please see Kuklinski et al. [118] and Steinbeck et al. [119].

To assess general health status, we used the European Quality of Life Visual Analogue Scale (EQ-VAS) and the EQ-5D-5L. To measure disease-specific outcomes, we applied the Hip disability/Knee injury and Osteoarthritis Outcome Score Physical Function Shortform (HOOS-PS and KOOS-PS, respectively). We used the Patient-Reported Outcomes Measurement Information System Fatigue Short-Form (PROMIS-F-SF) to measure fatigue. PROMIS-D-SF enables assessing depression symptoms. Lastly, pain in the operated joint was evaluated on a scale of 1 to 10, with higher scores indicating greater pain.

First, we used summary statistics to compare patient and treatment characteristics between genders, employing t-tests and Chi-square tests for continuous and categorical variables, respectively. PROs at admission, discharge, and 12 months post-surgery were analyzed by gender using the non-parametric Whitney-Mann U-test due to non-normal distribution. Next, linear multiple regression models were employed for each PRO, controlling for various variables depending on the

timepoint.[5] Additionally, regression analyses set the admission PRO as a baseline to assess the gender health gap present at hospital admission.

We found significant differences between genders in almost all PROs for both joint replacement types at each timepoint, with men experiencing significantly better health across PROs. The pain dimension 12 months post-HA revealed no significant gender difference. The mean PRO differences between male and female HA patients exhibited the greatest disparities in PROs pre-surgery, followed by those at discharge and were lowest at month 12. In comparison, the largest differences for KA patients were also evident at admission, while disparities were lower at discharge and at month 12 than at admission. However, disparities did not diminish in all PROs from discharge to month 12 among KA patients.

After adjusting for confounders, significantly better health in men is observable in all health dimensions pre-HA, and at discharge. However, 12 months post-surgery the most significant gender differences had disappeared expect for fatigue (HA), depression (KA) and knee functionality (KA). In terms of PRO improvement between admission and 12-months post-surgery, women showed significantly higher improvement in all PROs.

When also controlling for pre-operative scores, we found that men exhibited significantly better health at hospital discharge after HA and KA in terms of fatigue and depression. Additionally, male HA patients displayed improved hip functionality while male KA patients suffered less pain; however, other health dimensions exhibited no significant gender differences. At 12 months post-surgery, the only significant difference is that male HA patients experienced worse fatigue than women. Regarding PRO-improvement, the sole significant difference was decreased fatigue in male HA patients.

[5] See Chapter 3 (Study 4) for further information on the selected control variables.

Thus, S4 illustrates that the pre-operative score in HRQoL predicts the 12-month post-surgery score. When evaluating improvements in health status over time, accounting for pre-operative scores may mediate the influence of gender on post-surgical PROs. This study therefore questions whether male patients might receive HA and KA unnecessarily early (with too high (good) pre-operative PRO scores). In contrast, women might undergo surgery unnecessarily late (with much worse pre-operative PRO scores). However, neither the optimal time nor the patient's influence are known. Raising awareness of the importance of pre-operative gender differences in PROs might further enhance the application of timely adjusted interventions where male and female patients receive the appropriate treatment when needed. As otherwise, due to the inappropriate timing of the treatment, LVC may be caused.

Study 5 (S5, "What matters when interpreting patient-reported outcomes? – Adjusted minimal important differences for hip and knee replacement patients"; under review at *Quality of Life Research*) explores whether a change in PRO scores is clinically relevant by suggesting minimal important differences (MIDs). The MIDs proposed here are adjusted to account for patient characteristics as this enables more realistically evaluating treatment success. S5 therefore asks:

- Which patient characteristics influence post-surgery PRO score improvements and thereby enable a more precise MID estimation?

S5 employs a retrospective observational analysis using the National Health Service (NHS) PROM dataset covering HA and KA in England from 2013 to 2020. The dataset included patient demographics, pre-operative health history, comorbidities, as well as pre- and 6-month post-operative scores from the EQ-5D-3L, Oxford Hip Score (OHS), and Oxford Knee Score (OKS).

Data cleaning resulted in two datasets for regression and MID models. For HA, the regression model included 216,639 cases, and the MID model encompassed 252,331 cases. Similarly, the KA datasets included 245,711 cases for the regression model and 279,668 for the MID model.

Sample characteristics were outlined using descriptive statistics. A three-step approach was applied to estimate the MIDs for HA and KA patients: First, regression models served to identify pre-operative factors that best predict EQ-5D-3L improvement and standardized coefficients were calculated for these variables. Second, we clustered patients using the k-means algorithm based on the identified predicting factors. Additionally, differences in MIDs between male and female patients were examined. Third, we estimated MIDs for unstratified and stratified samples to determine the smallest meaningful improvement required for HA and KA to be considered successful.

The MIDs were estimated using an anchor-based approach, employing a binary "Success" variable as the anchor, based on patient responses to compare the current HRQoL status with the pre-surgery score. Sensitivity analyses were performed to test alternative anchors such as "Satisfaction" and transformed versions of "Success" to validate the stability of the results. MIDs were calculated by minimizing the sum of squares of 1-sensitivity and 1-specificity.

We also analyzed the factors influencing post-operative EQ-5D-3L index score improvement in hip and knee replacement patients. The most significant predicting factors of the post-surgery EQ-5D-3L index score are a patient's pre-operative EQ-5D-3L index score and a prior diagnosis of depression, both negatively associated with post-surgery improvement. Additionally, we created sub-groups based on pre-operative EQ-5D-3L scores to estimate MIDs. Stratifying the sample led to more precise MID calculations. However, MIDs varied among pre-operative EQ-5D-3L score subgroups, with patients starting with worse scores needing larger improvements to achieve their MIDs. Stratifying by gender revealed no significant MID differences between male and female patients, with consistent MIDs across pre-operative EQ-5D-3L index score subgroups.

S5 highlights that controlling for the pre-operative EQ-5D-3L index score predicts more accurate MIDs. In contrast to studies 3 and 4, we found no gender differences in MIDs. Considering the pre-operative EQ-5D index score is crucial when analyzing intervention success. Therefore, to increase the usefulness of PROs, users, (i.e.,

especially physicians) must be aware of the different interpretive approaches to obtain meaningful PRO results.

After investigating essential considerations when interpreting PROs, we incorporated PROMs in dashboards to enhance adoption (see Study 6 (S6): "Evaluating design principles for clinical dashboards incorporating patient-reported outcome measures: Building on a scoping literature review and semi-structured interviews"; submitted to the *Journal of Medical Internet Research*). We therefore asked:

- What are design principles of clinical dashboards incorporating generic and disease-specific PROMs?
- Do the design principles for clinical dashboards incorporating PROMs differ for acute conditions with a one-time intervention versus chronic conditions?

To answer these research questions, we applied a two-step approach. First, we conducted a scoping literature review to gain insights into the current state of evidence on clinical dashboard designs. Four medical databases were consulted: PubMed, PsychInfo, Cochrane, and Science Direct. We included studies published between January 2015 and May 2022. Furthermore, we employed a twin-track search strategy. Search A collected information on which generic and disease-specific PROMs are applicable in clinical dashboards and whether a combined use is frequently reflected in the literature. Search B focused on application areas of clinical dashboards and how such dashboards are designed. All included manuscripts had to represent original work and be written in English. Further studies were added by conducting backward research. In total, we identified 67 suitable manuscripts.

The scoping literature review revealed that various design principles are employed to enhance the usability of clinical dashboards incorporating PROMs such as additional patient information (e.g., demographic information or recent health updates) or clinical data (e.g., lab results and medication data). The literature also suggests including past assessment scores, peer-group comparisons, or alerts. The findings of the scoping literature review served as a basis for the subsequent interviews, where the identified design principles were discussed and evaluated with interviewees.

We conducted 16 interviews with six software producers and ten users (i.e., physicians). Interviews lasted between 20 to 45 minutes. They were audio-recorded and transcribed verbatim. For the purpose of analysis, we followed a deductive approach based on the identified design principles from the scoping literature review and enhanced our findings by applying an inductive approach. We evaluated 17 design principles from the scoping literature review and identified eight additional design principles through the inductive approach.

The interviews revealed that design principles need not be distinguished in terms of episodes of care—except for the disease-specific PROM, which needs to be adjusted accordingly. Interviewees rated design principles such as patient information and clinical data as necessary add-ons in the clinical dashboards. This additional information provides valuable insights into how PROMs must be interpreted according to a patient's current health. Interviewees highly recommended including the past assessment PROM score and peer-group comparison into clinical dashboards incorporating PROMs, as these principles further facilitate using PROMs in clinical practice.

S6 provides guidance on design principles that must be considered when planning, developing, and implementing clinical dashboards incorporating PROMs. We propose a model encompassing all 25 design principles and group these into four main aggregated concepts: *general setting* (subdivided into technical and non-technical settings), *PROM-related principles*, *PROM-independent principles*, and *requirements for adoption*. Connecting the findings to the framework of this dissertation, this study illustrates that PROMs are appropriate to be included in a clinical dashboard. However, additional design principles need to be regarded to enhance the informative value of such dashboards.

1.2.3 Study on health literacy

While Chapters 3 and 4 focus on interpreting PROMs and how their adoption in clinical practice reduces LVC, Chapter 5 explores the potential of patient's health literacy (HL) to increase treatment adherence and thus avoid LVC. Study 7 (S7, "The

power of knowledge: A survey on COPD patients' health literacy in Switzerland"; published in Schriftenreihe in Health Economics, Management and Policy) evaluates whether COPD patients' daily support requirements differ according to their HL states and which factors influence HL. HL is a major component of self-management capabilities, as reflected in a patient's ability to follow the treatment path and take medication as prescribed. However, if HL is not distinct enough, patients may (unintentionally) cause LVC. S7 therefore asks:

- Which factors are associated with higher health literacy?
- What types of support do patients with varying health literacy states desire in digital health interventions to better deal with COPD?

Data were collected via an online survey conducted between July 2022 and July 2023. All COPD patients living in Switzerland at the time of the survey could participate. The questionnaire contained five sections (general patient information, disease-specific questions, support needs, quality of life, and coping with COPD). It also included internationally validated instruments for measuring the patient's HRQoL (with the EQ-5D-5L), disease-specific condition (with the COPD Assessment Test), and HL (with the HLS-EU-Q16). In total, the survey included 66 questions and was available in German, Italian, and French. We announced it on Facebook and distributed flyers to various care providers and COPD-related organizations. Data of 203 survey participants were analyzed.

First, respondents were clustered into one out of three HL states, which categorize HL as "sufficient," "inadequate," or "problematic." Second, descriptives of variables were provided, and a chi-square test of homogeneity and an analysis of variance (ANOVA) were used to identify significant differences in the values allocated to HL states. Third, a cumulative odds ordinal regression with proportional odds served to determine which factors are significantly associated with higher HL.

We found that around half (53.8%) of the participants have adequate HL. The remaining participants expressed inadequate (27.6%) or problematic (19.2%) HL. Significant differences were evident in patients' support needs, length of diagnosis,

and confidence in medication intake. Participants with problematic and inadequate HL required more daily assistance than those with sufficient HL. People diagnosed with COPD over five years were more likely to have sufficient HL. Additionally, those with sufficient HL felt more confident about their medication intake than those with inadequate or problematic HL.

Furthermore, individuals with sufficient HL were older compared to those with inadequate and problematic HL. Moreover, participants with sufficient HL scored significantly higher (better) in EQ-5D-5L compared to the other two HL states. In terms of EQ-VAS, individuals with problematic HL reported significantly lower (worse) values compared to those with inadequate and sufficient HL. Additionally, participants with sufficient HL indicated significantly lower (better) CAT scores compared to those with inadequate and problematic HL.

The ordinal logistic regression used the following variables as confounders: daily life support, length of diagnosis, feeling confident about medication intake, age, EQ-5D-5L, and EQ-VAS. These variables all exhibited significant differences in the distribution of characteristics or mean values. Higher age and higher (better) EQ-5D-5L scores increase the likelihood of enhanced HL.

Although differences in patient characteristics are evident between HL states, patients do not differ in their perceived difficulties in dealing with COPD, their wish for daily life support, and symptom-specific support. Study participants indicated that they struggled most with their restricted movement possibilities and lacking chances of cure. Their desiderata included regular advice and information for dealing with COPD, regular exercises to alleviate COPD symptoms, and the possibility to compare their health status with the average COPD patient.

S7 reveals that significant differences between people with varying HL exist. Age and the EQ-5D-5L—representing the HRQoL—are significantly associated with better HL. However, study participants do not differ in their support requirements. In the framework of this dissertation (see Figure 1.1), the link between HRQoL and HL is of particular interest. If higher HRQoL increases the likelihood of improved HL, this might positively impact self-management, as HL is an essential lever of self-

management capabilities. If patients are enabled to develop a more intuitive sense of their HRQoL and thus increase HL, this might reduce LVC as patients will more likely adhere to treatment plans and thereby avoid complications.

References

1. Hehli S (2023) Prämien schlagen drastisch auf - Durchschnittsfamilie mit zwei Kindern zahlt 1000 Franken mehr. Neue Zürcher Zeitung

2. MedInside (2023) Steigende Gesundheitskosten: Beschleunigung setzt sich fort. In: MedInside. https://www.medinside.ch/steigende-gesundheitskosten-beschleunigung-setzt-sich-fort-20230811. Accessed 9 Oct 2023

3. SDA/anf (2023) Santésuisse rechnet mit acht bis neun Prozent höheren Krankenkassenprämien. Tagesanzeiger

4. BAG (2023) Stark steigende Kosten führen zu deutlich höheren Prämien im Jahr 2024. https://www.bag.admin.ch/bag/de/home/das-bag/aktuell/medienmitteilungen.msg-id-97889.html. Accessed 6 Oct 2023

5. Bundesärztekammer (2022) Finanzielle Fehlanreize im Gesundheitssystem gefährden Patientenwohl. https://www.bundesaerztekammer.de/presse/informationsdienste/informationsdienst-baekground/detail/finanzielle-fehlanreize-im-gesundheitssystem-gefaehrden-patientenwohl. Accessed 9 Oct 2023

6. OECD.Stat (2023) Health expenditure and financing. https://stats.oecd.org/Index.aspx?DataSetCode=SHA. Accessed 6 Oct 2023

7. BAG (2019) Evaluation der KVG-Revision im Bereich der Spitalfinanzierung - Schlussbericht des BAG an den Bundesrat

8. World Health Organization (2020) Quality health services. https://www.who.int/news-room/fact-sheets/detail/quality-health-services. Accessed 6 Oct 2023

9. Donabedian A (1980) Explorations in quality assessment and monitoring: The definition of quality and approaches to its assessment. Ache Management

10. Rademakers J, Delnoij D, de Boer D (2011) Structure, process or outcome: which contributes most to patients' overall assessment of healthcare quality? BMJ Qual Saf 20:326–331.

11. Chassin MR, Galvin RW, Donaldson MS (1998) The urgent need to improve health care quality: Institute of Medicine national roundtable on health care quality. J Am Med Assoc 280:1000–1005.

12. Colla CH, Mainor AJ, Hargreaves C, et al (2017) Interventions aimed at reducing use of low-value health services: A systematic review

13. Elshaug AG, Rosenthal MB, Lavis JN, et al (2017) Levers for addressing medical underuse and overuse: Achieving high-value health care. The Lancet 390:191–202.

14. Mafi JN, Parchman M (2018) Low-value care: An intractable global problem with no quick fix. BMJ Qual Saf 27:333–336.

15. Verkerk EW, Tanke MAC, Kool RB, et al (2018) Limit, lean or listen? A typology of low-value care that gives direction in de-implementation. International Journal for Quality in Health Care 30:736–739.

16. Brownlee S, Chalkidou K, Doust J, et al (2017) Evidence for overuse of medical services around the world. The Lancet 390:156–168.

17. Sypes EE, De Grood C, Whalen-Browne L, et al (2020) Engaging patients in de-implementation interventions to reduce low-value clinical care: A systematic review and meta-analysis. BMC Med 18:1–15.

18. Sypes EE, De Grood C, Clement FM, et al (2020) Understanding the public's role in reducing low-value care: A scoping review. Implementation Science 15:1–13.

19. Rietbergen T, Spoon D, Brunsveld-Reinders AH, et al (2020) Effects of de-implementation strategies aimed at reducing low-value nursing procedures: A systematic review and meta-analysis. Implementation Science 15:1–18.

20. Prasad V, Ioannidis JP (2014) Evidence-based de-implementation for contradicted, unproven, and aspiring healthcare practices. Implementation Science 9:.

21. Niven DJ, Mrklas KJ, Holodinsky JK, et al (2015) Towards understanding the de-adoption of low-value clinical practices: A scoping review. BMC Med 13:.

22. De Vries EF, Struijs JN, Heijink R, et al (2016) Are low-value care measures up to the task? A systematic review of the literature. BMC Health Serv Res 16:.

23. Chan KS, Chang E, Nassery N, et al (2013) The state of overuse measurement: A critical review. Medical Care Research and Review 70:473–496.

24. Pronovost PJ, Lilford R (2011) A road map for improving the performance of performance measures. Health Aff 30:569–573.

25. Grimshaw JM, Patey AM, Kirkham KR, et al (2020) De-implementing wisely: Developing the evidence base to reduce low-value care. BMJ Qual Saf 29:409–417.

26. Mathias JS, Baker DW (2013) Developing quality measures to address overuse. JAMA 309:1897–1898.

27. Choosing Wisely (2022) Choosing Wisely. https://www.choosingwisely.org/

28. Smarter Medicine (2022) Smarter Medicine. https://www.smartermedicine.ch/de/home.html

29. Born K, Kool T, Levinson W (2019) Reducing overuse in healthcare: advancing Choosing Wisely. BMJ l6317.

30. Maratt JK, Kerr EA, Klamerus ML, et al (2019) Measures used to assess the impact of interventions to reduce low-value care: A systematic review. J Gen Intern Med 34:1857–1864.

31. Miller G, Rhyan C, Beaudin-Seiler B, Hughes-Cromwick P (2018) A Framework for measuring low-value care. Value in Health 21:375–379.

32. Greenhalgh J, Gooding K, Gibbons E, et al (2018) How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. J Patient Rep Outcomes 2:42.

33. Field J, Holmes MM, Newell D (2019) PROMs data: Can it be used to make decisions for individual patients? A narrative review. Patient Relat Outcome Meas Volume 10:233–241.

34. Desantis D, Baverstock RJ, Civitarese A, et al (2016) A clinical perspective on electronically collecting patient-reported outcomes at the point-of-care for overactive bladder. Canadian Urological Association Journal 10:359.

35. Fautrel B, Alten R, Kirkham B, et al (2018) Call for action: how to improve use of patient-reported outcomes to guide clinical decision making in rheumatoid arthritis. Rheumatol Int 38:935–947.

36. Chen J, Ou L, Hollis SJ (2013) A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res 13:211.

37. Devlin NJ, Appleby J (2010) Getting the most out of PROMs. Putting health outcomes at the heart of NHS decision making. King's Fund

38. Greenhalgh J, Meadows K (1999) The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. J Eval Clin Pract 5:401–416.

39. Hassett MJ, Cronin C, Tsou TC, et al (2022) eSyM: An electronic health record–integrated patient-reported outcomes-based cancer symptom management program used by six diverse health systems. JCO Clin Cancer Inform 1–10.

40. Cella D, Yount S, Rothrock N, et al (2007) The Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 45:S3–S11.

41. Lipscomb J, Reeve BB, Clauser SB, et al (2007) Patient-Reported Outcomes Assessment in Cancer Trials: Taking Stock, Moving Forward. Journal of Clinical Oncology 25:5133–5140.

42. Cronin RM, Conway D, Condon D, et al (2018) Patient and healthcare provider views on a patient-reported outcomes portal. Journal of the American Medical Informatics Association 25:1470–1480.

43. Detmar SB (2003) Use of HRQOL questionnaires to facilitate patient–physician communication. Expert Rev Pharmacoecon Outcomes Res 3:215–217.

44. Marshall S, Haywood K, Fitzpatrick R (2006) Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract 12:559–568.

45. Berry DL, Hong F, Halpenny B, et al (2014) Electronic Self-Report Assessment for Cancer and Self-Care Support: Results of a Multicenter Randomized Trial. Journal of Clinical Oncology 32:199–205.

46. Berry DL, Blumenstein BA, Halpenny B, et al (2011) Enhancing Patient-Provider Communication With the Electronic Self-Report Assessment for Cancer: A Randomized Trial. Journal of Clinical Oncology 29:1029–1035.

47. Detmar SB, Muller MJ, Schornagel JH, et al (2002) Health-related quality-of-life assessments and patient-physician communication. JAMA 288:3027.

48. Velikova G, Booth L, Smith AB, et al (2004) Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. Journal of Clinical Oncology 22:714–724.

49. Abernethy AP, Ahmad A, Zafar SY, et al (2010) Electronic patient-reported data capture as a foundation of rapid learning cancer care. Med Care 48:.

50. Kwon S, Florence M, Grigas P, et al (2012) Creating a learning healthcare system in surgery: Washington State's Surgical Care and Outcomes Assessment Program (SCOAP) at 5 years. Surgery 151:146–152.

51. Devlin NJ, Parkin D, Janssen B (2020) Methods for analysing and reporting EQ-5D data. Springer International Publishing

52. Dieppe PA (1995) Recommended methodology for assessing the progression of osteoarthritis of the hip and knee joints. Osteoarthritis Cartilage 3:73–77.

53. Bellamy N, Buchanan WW, Goldsmith CH, et al (1988) Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 15:1833–40.

54. Bergh I, Kvalem IL, Aass N, Hjermstad MJ (2011) What does the answer mean? A qualitative study of how palliative cancer patients interpret and respond to the Edmonton Symptom Assessment System. Palliat Med 25:716–724.

55. Greenhalgh J, Gooding K, Gibbons E, et al (2018) How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. J Patient Rep Outcomes 2:42.

56. Hewlett S, Kirwan J, Pollock J, et al (2005) Patient initiated outpatient follow up in rheumatoid arthritis: six year randomised controlled trial. BMJ 330:171.

57. Ragouzeos D, Gandrup J, Berrean B, et al (2019) “Am I OK?” using human centered design to empower rheumatoid arthritis patients through patient reported outcomes. Patient Educ Couns 102:503–510.

58. Field J, Holmes MM, Newell D (2019) PROMs data: can it be used to make decisions for individual patients? A narrative review. Patient Relat Outcome Meas Volume 10:233–241.

59. Kroenke K, Monahan PO, Kean J (2015) Pragmatic characteristics of patient-reported outcome measures are important for use in clinical practice. J Clin Epidemiol 68:1085–1092.

60. Kroll T, Wyke S, Jahagirdar D, Ritchie K (2014) If patient-reported outcome measures are considered key health-care quality indicators, who is excluded from participation? Health Expectations 17:605–607.

61. Brundage MD, Smith KC, Little EA, et al (2015) Communicating patient-reported outcome scores using graphic formats: results from a mixed-methods evaluation. Quality of Life Research 24:2457–2472.

62. Bantug ET, Coles T, Smith KC, et al (2016) Graphical displays of patient-reported outcomes (PRO) for use in clinical practice: What makes a pro picture worth a thousand words? Patient Educ Couns 99:483–490.

63. Kirwan JR, Bartlett SJ, Beaton DE, et al (2014) Updating the OMERACT Filter: Implications for Patient-reported Outcomes. J Rheumatol 41:1011–1015.

64. Al Sayah F, Jin X, Johnson JA (2021) Selection of patient-reported outcome measures (PROMs) for use in health systems. J Patient Rep Outcomes 5:99.

65. Weldring T, Smith SMS (2013) Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). Health Serv Insights 6:61–68.

66. Ludwig K, von der Schulenburg J-MG, Greiner W (2017) Valuation of the EQ-5D-5L with composite time trade-off for the German population – an exploratory study. Health Qual Life Outcomes 15:39.

67. Ludwig K, Graf von der Schulenburg J-M, Greiner W (2018) German Value Set for the EQ-5D-5L. Pharmacoeconomics 36:663–674.

68. Ware JE, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30:473–83

69. Bullinger M (1995) German translation and psychometric testing of the SF-36 Health Survey: Preliminary results from the IQOLA project. Soc Sci Med 41:1359–1366.

70. Rothrock NE, Amtmann D, Cook KF (2020) Development and validation of an interpretive guide for PROMIS scores. J Patient Rep Outcomes 4:16.

71. Mehta SP, Sankar A, Venkataramanan V, et al (2016) Cross-cultural validation of the ICOAP and physical function short forms of the HOOS and KOOS in a multi-country study of patients with hip and knee osteoarthritis. Osteoarthritis Cartilage 24:2077–2081.

72. Roos EM, Toksvig-Larsen S (2003) Knee injury and Osteoarthritis Outcome Score (KOOS) – validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 1:17.

73. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM (2003) Hip disability and osteoarthritis outcome score (HOOS) – validity and responsiveness in total hip replacement. BMC Musculoskelet Disord 4:10.

74. Jones PW, Harding G, Berry P, et al (2009) Development and first validation of the COPD assessment test. European Respiratory Journal 34:648–654.

75. Jones PW (2013) COPD assessment test – Rationale, development, validation and performance. COPD: Journal of Chronic Obstructive Pulmonary Disease 10:269–271.

76. Krabbe PFM (2017) The measurement of health and health status. Elsevier

77. Kuklinski D, Oschmann L, Pross C, et al (2020) The use of digitally collected patient-reported outcome measures for newly operated patients with total knee and hip replacements to improve post-treatment recovery: study protocol for a randomized controlled trial. Trials 21:322.

78. Kuklinski D, Marques CJ, Bohlen K, et al (2022) Thresholds for meaningful improvement in WOMAC scores need to be adjusted to patient characteristics after hip and knee replacement. J Orthop 29:50–59.

79. Hertler C, Seiler A, Gramatzki D, et al (2020) Sex-specific and gender-specific aspects in patient-reported outcomes. ESMO Open 5:e000837.

80. Lipscomb J, Reeve BB, Clauser SB, et al (2007) Patient-reported outcomes assessment in cancer trials: Taking stock, moving forward. Journal of Clinical Oncology 25:5133–5140.

81. Berry DL, Blumenstein BA, Halpenny B, et al (2011) Enhancing patient-provider communication with the electronic self-report assessment for cancer: A randomized trial. Journal of Clinical Oncology 29:1029–1035.

82. Berry DL, Hong F, Halpenny B, et al (2014) Electronic self-report assessment for cancer and self-care support: Results of a multicenter randomized trial. Journal of Clinical Oncology 32:199–205.

83. Nelson JA, Chu JJ, Dabic S, et al (2023) Moving towards patient-reported outcomes in routine clinical practice: implementation lessons from the BREAST-Q. Quality of Life Research 32:115–125.

84. Fleischmann M, Vaughan B (2018) The challenges and opportunities of using patient reported outcome measures (PROMs) in clinical practice. International Journal of Osteopathic Medicine 28:56–61.

85. Feenstra MM, Kidholm K, Rudnicki M (2023) Use of patient reported outcome measures in daily clinical practice – A way to optimize treatment of patients? Acta Obstet Gynecol Scand 102:404–405.

86. Amini M, Oemrawsingh A, Verweij LM, et al (2021) Facilitators and barriers for implementing patient-reported outcome measures in clinical care: An academic center's initial experience. Health Policy (New York) 125:1247–1255.

87. Foster A, Croot L, Brazier J, et al (2018) The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. J Patient Rep Outcomes 2:46.

88. Nguyen H, Butow P, Dhillon H, Sundaresan P (2021) A review of the barriers to using Patient-Reported Outcomes (PROs) and Patient-Reported

Outcome Measures (PROMs) in routine cancer care. J Med Radiat Sci 68:186–195.

89. Etkind SN, Daveson BA, Kwok W, et al (2015) Capture, transfer, and feedback of patient-centered outcomes data in palliative care populations: Does it make a difference? A systematic review. J Pain Symptom Manage 49:611–624.

90. Basch E, Deal AM, Kris MG, et al (2016) Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. Journal of Clinical Oncology 34:557–565.

91. Ishaque S, Karnon J, Chen G, et al (2019) A systematic review of randomised controlled trials evaluating the use of patient-reported outcome measures (PROMs). Quality of Life Research 28:567–592.

92. Gibbons C, Porter I, Gonçalves-Bradley DC, et al (2021) Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Cochrane Database of Systematic Reviews 2021:.

93. Graupner C, Kimman ML, Mul S, et al (2021) Patient outcomes, patient experiences and process indicators associated with the routine use of patient-reported outcome measures (PROMs) in cancer care: a systematic review. Supportive Care in Cancer 29:573–593.

94. Berkman ND, Davis TC, McCormack L (2010) Health literacy: What is it? J Health Commun 15:9–19.

95. Nutbeam D (1998) Health Promotion Glossary. Health Promot Int 13:349–364.

96. Speros C (2005) Health literacy: concept analysis. J Adv Nurs 50:633–640.

97. Sørensen K, Van den Broucke S, Pelikan JM, et al (2013) Measuring health literacy in populations: illuminating the design and development process of

the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health 13:948.

98. Osborne RH, Batterham RW, Elsworth GR, et al (2013) The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 13:658.

99. Jordan S, Hoebel J (2015) Gesundheitskompetenz von Erwachsenen in Deutschland: Ergebnisse der Studie „Gesundheit in Deutschland aktuell“ (GEDA). Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 58:942–950.

100. Morris NS, MacLean CD, Chew LD, Littenberg B (2006) The Single Item Literacy Screener: Evaluation of a brief instrument to identify limited reading ability. BMC Fam Pract 7:21.

101. Tavousi M, Mohammadi S, Sadighi J, et al (2022) Measuring health literacy: A systematic review and bibliometric analysis of instruments from 1993 to 2021. PLoS One 17:e0271524.

102. Voigt-Barbarowicz M, Dietz G, Renken N, et al (2022) Patients’ health literacy in rehabilitation: Comparison between the estimation of patients and health care professionals. Int J Environ Res Public Health 19:3522.

103. Dickens C, Lambert BL, Cromwell T, Piano MR (2013) Nurse overestimation of patients’ health literacy. J Health Commun 18:62–69.

104. Storms H, Aertgeerts B, Vandenabeele F, Claes N (2019) General practitioners’ predictions of their own patients’ health literacy: a cross-sectional study in Belgium. BMJ Open 9:e029357.

105. Stock S, Altin S, Nawabi F, et al (2021) A cross-sectional analysis of health literacy: patient- versus family doctor-reported and associations with self-efficacy and chronic disease. BMC Fam Pract 22:187.

106. Canady BE, Larzo M (2023) Overconfidence in managing health concerns: The Dunning–Kruger effect and health literacy. J Clin Psychol Med Settings 30:460–468.

107. Sunjaya AP, Bao L, Martin A, et al (2022) Systematic review of effectiveness and quality assessment of patient education materials and decision aids for breathlessness. BMC Pulm Med 22:237.

108. Szabó P, Bíró É, Kósa K (2021) Readability and comprehension of printed patient education materials. Front Public Health 9:725840.

109. Heijmans M, Waverijn G, Rademakers J, et al (2015) Functional, communicative and critical health literacy of chronic disease patients and their importance for self-management. Patient Educ Couns 98:41–48.

110. Puente-Maestu L, Calle M, Rodríguez-Hermosa JL, et al (2016) Health literacy and health outcomes in chronic obstructive pulmonary disease. Respir Med 115:78–82.

111. Azkan Ture D, Bhattacharya S, Demirci H, Yildiz T (2022) Health literacy and health outcomes in chronic obstructive pulmonary disease patients: An explorative study. Front Public Health 10:.

112. Omachi TA, Sarkar U, Yelin EH, et al (2013) Lower health literacy is associated with poorer health status and outcomes in chronic obstructive pulmonary disease. J Gen Intern Med 28:74–81.

113. Muellers KA, Chen L, O'Conor R, et al (2019) Health literacy and medication adherence in COPD patients: When caregiver presence is not sufficient. COPD: Journal of Chronic Obstructive Pulmonary Disease 16:362–367.

114. Poureslami I, FitzGerald JM, Tregobov N, et al (2022) Health literacy in asthma and chronic obstructive pulmonary disease (COPD) care: a narrative review and future directions. Respir Res 23:361.

115. O'Conor R, Muellers K, Arvanitis M, et al (2019) Effects of health literacy and cognitive abilities on COPD self-management behaviors: A prospective cohort study. Respir Med 160:105630.

116. Shahid R, Shoker M, Chu LM, et al (2022) Impact of low health literacy on patients' health outcomes: a multicenter cohort study. BMC Health Serv Res 22:1148.

117. Yee LM, Silver R, Haas DM, et al (2021) Association of health literacy among nulliparous individuals and maternal and neonatal outcomes. JAMA Netw Open 4:e2122576.

118. Kuklinski D, Oschmann L, Pross C, et al (2020) The use of digitally collected patient-reported outcome measures for newly operated patients with total knee and hip replacements to improve post-treatment recovery: study protocol for a randomized controlled trial. Trials 21:322.

119. Steinbeck V, Langenberger B, Schöner L, et al (2023) Electronic patient-reported outcome monitoring to improve quality of life after joint replacement. JAMA Netw Open 6:e2331301.

2 Low-value care in routine and claims data

2.1 Making the cut on caesarean section: a logistic regression analysis on factors favouring caesarean sections without medical indication in comparison to spontaneous vaginal birth

Anja Y. Bischof

Alexander Geissler

Published in *BMC Pregnancy and Childbirth*

2.1.1 Abstract

Background: In the absence of medical necessity, opting for caesarean sections exposes mothers and neonates to increased risks of enduring long-term health problems and mortality. This ultimately results in greater economic burden when compared to the outcomes of spontaneous vaginal births. In Switzerland around 33% of all births are by caesarean section. However, the rate of caesarean sections without medical indication is still unknown. Therefore, we devise an identification strategy to differentiate caesarean sections without medical indication using routine data. In addition, we aim to categorize the influencing factors for women who undergo spontaneous vaginal births as opposed to those with caesarean sections without medical indication.

Method: We use Swiss Federal Statistics data including 98.3% of all women giving birth from 2014 to 2018. To determine non-medically indicated caesarean sections in our dataset, we base our identification strategy on diagnosis-related groups, diagnosis codes, and procedure classifications. Subsequently, we compare characteristics of women who give birth by non-medically CS and external factors such as the density of practicing midwives to women with spontaneous vaginal birth. Logistic regression analysis measures the effect of factors, such as age, insurance class, income, or density of practicing midwives on non-medically indicated caesarean sections.

Results: Around 8% of all Swiss caesarean sections have no medical indication. The regression analysis shows that higher age, supplemental insurance, higher income, and living in urban areas are associated with non-medically indicated caesarean sections, whereas a higher density of midwives decreases the likelihood of caesarean sections without medical indication.

Conclusions: By identifying non-medically indicated caesarean sections using routine data, it becomes feasible to gain insights into the characteristics of impacted mothers as well as the external factors involved. Illustrating these results, our recommendation is to revise the incentive policies directed towards healthcare professionals. Among others, future research may investigate the potential of midwife-assisted pregnancy programs on strengthening spontaneous vaginal births in absence of medical complications.

Keywords: Birth setting, Natural birth, Caesarean section, Switzerland

2.1.2 Background

The mode of birth relies on the condition of the mother and the neonate. The child is either born spontaneously, by instrumental birth, or via a caesarean section (CS). Desirable results linked to CS, such as reduced maternal or neonatal mortality or morbidity, are attainable only up to a specific national ratio [1]. No improved outcomes are evident for the remaining CS when compared to spontaneous vaginal birth (SVB). The WHO recommends a national CS rate between 10-15% [2]. CS rates vary considerably among OECD (Organization for Economic Co-cooperation and Development) countries, which are mainly high-income and developed countries from various regions of the world [3]. However, if unified guidelines for the indication of CS are applied or a second opinion is mandatory, the variance across and within countries is expected to be lower [4, 5]. One suggested approach is the Robson classification, where every pregnant woman is assigned into one of ten clusters based on pregnancy-related characteristics, aiming to uncover reasons and outcomes for different birth modes [6, 7].

In this paper, we categorize CS into two primary groups: planned and emergency. Planned CS represents a planned intervention, which may be conducted upon or without medical indication. Women might prefer a planned, non-medically indicated CS (referred to as non-medically indicated CS from now on) due to social norms, emotional or personal experiences [8, 9]. During the 1990s, medical indication was among the primary factors guiding the decision to opt for CS [10]. However, in the present day, psychosocial elements like birth-related fear or maternal preference without concurrent medical reasons have taken precedence at the top of the list [9, 11]. Thus, the rationale for planned CS as the eligible mode of birth is no longer only due to medical indication [8, 9, 12, 13]. In contrast, emergency CS are performed as an immediate or urgent intervention once labour has commenced.

Media reports presenting an inappropriate image of CS reduce women's perception of the associated risks [14]. Nevertheless, different studies draw attention to the risks of non-medically indicated CS [12, 15–17]. The mother faces higher risks such as higher blood loss [18], increased probability of a stillbirth [19], or negative implications for future pregnancies [16, 19], whereas the neonate is confronted by

increased respiratory morbidity [19], and higher risks of short- and long-term morbidity in general [19–21] compared to SVB. Conducting a non-medically indicated CS not only increases the potential iatrogenic harm for mothers and neonates [11, 19], it also leads to increased financial burdens on the healthcare system [18, 22].

In this paper, we present a novel approach on identifying cases of non-medically indicated CS in routine data by using Swiss hospital claims data from 2014 to 2018. We examine factors that are associated with a higher likelihood of giving birth via CS in absence of any medical indication.

2.1.3 Data and Methods

Data and selection of cases

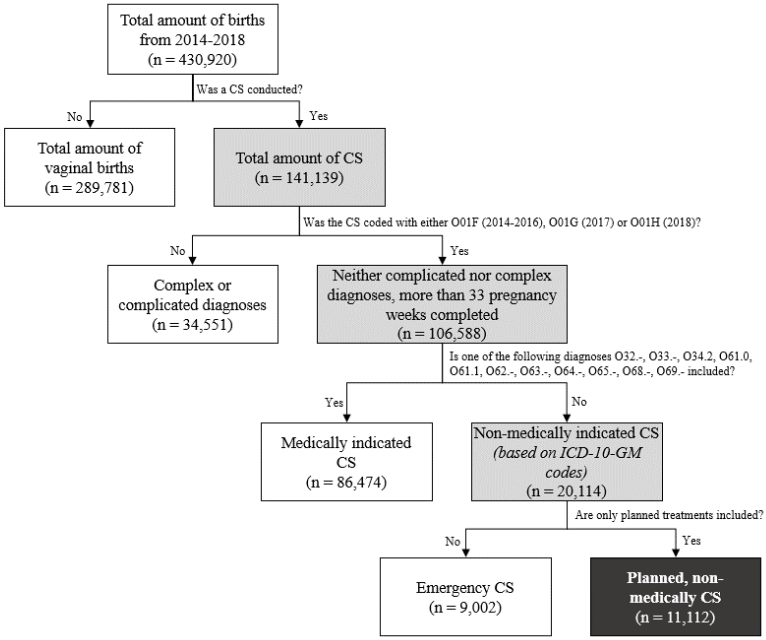
The study is based on the national dataset "Medical Statistics of Hospitals" which includes all women having given birth in Swiss hospitals between 2014 to 2018 (98.3% of all births, n = 430,920). The dataset is provided by the Federal Statical Office containing claims data of Swiss hospitals. As neither a corresponding diagnosis-related group (DRG), procedure classification (represented by the Swiss operation classification (CHOP)), nor a diagnosis code (ICD-10-GM) is available for non-medically indicated CS, we estimate the number of non-medically indicated CS based on various criteria. Nevertheless, the dataset does not provide information about whether a woman expresses a preference for giving birth via CS or not.

We obtain cases of CS without medical indication by data selection as represented in Figure 1. First, only cases with diagnosis-related groups (SwissDRG) O01F (2014-2016), O01G (2017), and O01H (2018) are considered as they refer to *"sectio caesarea, pregnancy duration more than 33 completed weeks (without complicating diagnosis, without a complex diagnosis)"*. We assume that having completed 33 weeks of pregnancy allows for spontaneous birth without fearing negative effects neither for the mother nor the neonate – every birth before 33 completed weeks classifies as extremely (less than 28 weeks) or very (28 to 32 weeks) preterm birth [23]. Furthermore, the dataset contains information about reimbursement rates, but it lacks details about actual incurred costs.

Second, diagnoses that do not allow for vaginal birth such as malpresentation of the foetus, maternal hip anomalies, or women with indicated labour who later change to an emergency CS are excluded from the dataset (see Additional file 1). We base our filtering process on Henry et al. [24] who applied the same procedure to ICD-9 codes. Subsequent studies further refine and augment this approach [25–27]. We translate the identified procedures to ICD-10-GM, which is the standard in Switzerland.

Third, we refer to the procedure classification to eliminate all CS characterized by emergency treatment features (specified by the CHOP catalogue, see Additional file 2). Moreover, the analysis begins from the year 2014 because the procedure classification did not differentiate between planned and emergency CS in prior years. This lack of distinction resulted in an inaccurate count of non-medically indicated CS.

Figure 2.1: Selection process of non-medically indicated CS based on a defined set of evaluation criteria



Legend: CS = Caesarean section; ICD-10-GM = International Statistical Classification of Diseases and Related Health Problems, 10. Revision, German Modification

Annotation: The boxes represent the individual stages of how we identify non-medically indicated CS. The figure starts at the top while getting more granular from going top left to bottom right. At each box, two arrows are allocated which either affirm or deny the corresponding question. The second question about whether a CS was classified as O01F (2014-2016), O01G (2017), or O01G (2018) uses the classification of Swiss DRG (diagnose-related groups) of the corresponding year. The third question considers the diagnoses of the ICD-10-GM. The last question is about planned and emergency CS. The CHOP (Swiss operation classification) catalogue distinguishes between these two types of CS. Coloured boxes represent the path on how to get to the total number of CS without medical indication.

We identify SVBs without complicated or complex diagnoses or accompanying factors via the DRG code O60D (2014-2018). The less complex DRG codes for CS are O01F (2014-2016), O01G (2017), and O01H (2018). We hypothesize that women with a CS in absence of any medical indication could have given birth spontaneously instead. This makes the group of SVB comparable to non-medically indicated CS.

Furthermore, we integrate supplementary data sources to enrich the explanatory power of the present study. Information is collected for per capita net income and nursing staff per hospital bed, both at the cantonal level and on an annual basis. For the identification of differences in the level of urbanity, we use information from the statistical atlas of Switzerland. Additionally, the Swiss society of midwives provides information on the number of practicing midwives per canton.

Empirical strategy

We provide descriptive statistics to illustrate the characteristics of the two birth mode groups. Secondly, to analyse significant differences between the mean values of mothers giving birth by non-medically indicated CS versus SVB, we use independent-samples t-test. Thirdly, logistic regression analysis with a 5% significance level identifies the impact of factors associated with a higher likelihood of giving birth by CS without medical indication. The regression includes the following independent variables: age, insurance status, urbanity level (i.e., rural, suburban, and urban), number of midwives per 1000 births, number of nursing staff per hospital, and income per capita.

Logistic regression analysis especially suits this research as only two outcomes are assumed: non-medically indicated CS and SVB. Furthermore, the logistic regression calculates the impact of the chosen independent variables on the chance of giving birth by non-medically indicated CS. A higher log odds ratio indicates a higher chance of giving birth by non-medically indicated CS. Similar approaches were already taken by previous studies for assessing the impact of various non-medical factors on non-medically indicated CS [12, 28]. We conduct all analyses with the Statistical Package for Social Sciences (SPSS), Version 27 [29].

2.1.4 Results

Descriptive statistics

The average number of annual births in Swiss hospitals from 2014 to 2018 was approximately 86,200 (see Table 2.1). Among them, 33% were born by CS and on average approximately 2,200 CS without medical indication have been detected.

These amounts to almost 8% of all CS and to more than 2% compared to the total number of births in Swiss hospitals. These numbers are stable over time. In contrast, SVB without complicated or complex diagnoses faced some fluctuations. The highest rates were reached in 2014 and 2015 with almost 40,000 births. Compared to this, around 10,000 births less were given birth vaginally in 2018. The total number of detected non-medically indicated CS added up to 11,112 and 176,908 women had a SVB between 2014 and 2018. On average, 41.1% of all births were given by SVB according to the DRG code O60D.

Table 2.1: Number of total births, caesarean sections and spontaneous vaginal births in Switzerland per year

Year	Total births (# births)	Caesarean section (# CS)	Non-medically indicated CS (# nmi CS)	#CS / # births	# nmi CS / # CS	# nmi CS / # births	SVB (# SVB)	# SVB / # births
2014	84,438	28,324	2,206	33.54%	7.79%	2.61%	39,654	46.96%
2015	85,847	28,473	2,340	33.17%	8.22%	2.73%	39,941	46.53%
2016	87,255	28,768	2,331	32.97%	8.10%	2.67%	33,662	38.58%
2017	86,501	27,801	2,145	32.14%	7.72%	2.48%	32,701	37.80%
2018	86,879	27,773	2,090	31.97%	7.53%	2.41%	30,950	35.62%
Average	**86,184**	**28,227**	**2,222**	**32.76%**	**7.87%**	**2.58%**	**35,381**	**41.10%**
Total	**430,920**	**141,139**	**11,112**	**-**	**-**	**-**	**176,908**	**-**

Legend: SVB = spontaneous vaginal birth

Annotation: Total births indicate all births in Switzerland per year – independent of birth mode. The number of CS represents all CS in Switzerland per year. In non-medically indicated CS, only the defined DRG-groups O01F (2014-2016), O01G (2017) and O01H (2018) are included. Spontaneous vaginal birth displays the number of births classified by the DRG-group O60D (2014-2018).

In terms of demographics (see Table 2), the age category 30-34 years is the largest group for both types of birth modes. This is followed by the category 35-39 years for non-medically indicated CS and 25-29 years for SVB. Thus, the average age tends to be higher at non-medically indicated CS compared to SVB. Furthermore, also differences in the living area of women were observable. The distribution of women's living areas was somewhat focused on some regions for non-medically indicated CS (Zurich followed by Espace Mittelland and North-Western Switzerland), while the rates of SVB were more equally distributed across the country.

Additionally, the urbanity level indicates the kind of living region (i.e., rural, suburban, or urban). Most women with a non-medically indicated CS live in urban areas (67.9%) followed by suburban (17.1%) and rural areas (15.1%). Women with SVB live mostly in urban areas (63.3%) followed by rural (19.2%) and suburban areas (17.2%).

In Switzerland, individuals have the choice to opt for supplementary insurance coverage for inpatient medical services in addition to mandatory insurance which covers basic services. Comparing this with supplemental insurance types, i.e., semi-private or private, the insured pay higher fees for additional services such as staying in a single or double-person room, and timely-preferred treatment by a specialist. The percentage of privately insured women with a non-medically indicated CS was double of those with SVB (8.8% vs 4.1%). Further, women with non-medically indicated CS were less often solely mandatorily insured than those with SVB (75.9% vs 85.5%).

Table 2.2: Comparison of ordinal factors between non-medically indicated caesarean section and spontaneous vaginal birth

	Non-medically indicated caesarean section		Spontaneous vaginal birth	
	Number	**%**	**Number**	**%**
Total population	11,112	100	176,908	100
Age category				
15-19	89	0.8	1,355	0.8

20-24	654	5.9	14,849	8.4
25-29	2,236	20.1	47,374	26.8
30-34	3,754	33.8	69,238	39.1
35-39	3,176	28.6	37,453	21.2
40-44	1,060	9.6	6,422	3.6
45-49	128	1.2	199	0.1
Living region				
Eastern Switzerland	1,115	10.0	22,296	12.6
Central Switzerland	1,009	9.1	16,519	9.3
North-Western Switzerland	1,805	16.2	21,544	12.2
Ticino	324	2.9	6,626	3.7
Zurich	2,871	25.8	29,409	16.6
Western Switzerland	1,616	14.5	38,837	22.0
Espace Mittelland	2,058	18.5	38,014	21.5
Other Countries	314	2.8	3,663	2.1
Insurance class				
Mandatory	8,431	75.9	151,295	85.5
Semi-private	1,702	15.3	18,429	10.4
Private	979	8.8	7,184	4.1
Urbanity level				
Urban	6,896	67.9	102,431	63.6
Suburban	1,733	17.1	27,706	17.2
Rural	1,529	15.1	30,963	19.2

DRG cost weights determine the payment of each inpatient case. Non-medically indicated CS were covered on average with a cost weight of 0.864 in 2014 remaining at the same level in 2016 and then decreasing to 0.791 in 2018. SVB showed a lower

cost weight, which was 0.581 in 2014 decreasing to 0.545 in 2016 and followed by a slight increase to 0.564 in 2018. Therefore, SVB payment was generally lower compared to non-medically indicated CS across all years.

Some differences in the characteristics of the mothers belonging to either group of birth modes were already observable. Table 2.3 summarizes the main differences in metric variables for non-medically indicated CS and SVB. Women with non-medically indicated CS live in cantons with a lower relative rate of midwives. On the contrary, the amount of nursing staff at the hospital calculated per bed is slightly higher for women with non-medically indicated CS. Finally, income per capita is on average more than CHF 1000 higher per year for women with non-medically indicated CS in contrast to women who gave birth spontaneously. Income per capita and the insurance status show a positive correlation ($r_S = 0.155$, $p < 0.01$), i.e., woman with a higher income tend to have supplemental insurance. All variables differ in their mean values with statistical significance ($p < 0.001$) between the two birth modes.

Table 2.3: Independent samples t-test for metric variables between women with non-medically indicated caesarean sections and spontaneous vaginal births

	Non-medically indicated CS	Spontaneous vaginal birth	p-values
Number of midwives per 1,000 births	37.28	37.77	<0.001
Number of nursing staff per hospital bed	3.70	3.64	<0.001
Income per capita	CHF 38,273.26	CHF 37,152.96	<0.001

Regression results

For the logistic regression (see Table 2.4), age category, insurance status, urbanity level, number of midwives per 1,000 births, number of nursing staff per hospital bed and income per capita were included as independent variables. All variables show the highest significance level ($p<0.001$). A lower age category decreases the risk of giving

birth by CS without medical indication, whereas the category 45-49 years served as a reference category. Additionally, for woman with mandatory or semi-private insurance the odds ratio of giving birth by non-medically indicated CS is significantly lower compared to woman with private insurance (reference category). For woman in urban areas, the likelihood of a CS without medical indication increases compared to woman in rural areas (reference category). Furthermore, a higher number of practicing midwives per 1,000 births reduce the log odds ratio by 0.019. The number of nursing staff per bed is also associated to the birth mode. The more nursing staff available per bed, the higher the chance of undergoing a non-medically indicated CS. Lastly, income per capita impacts the decision on the birth mode. A CHF 10,000 increase in annual income per capita raises the log odds ratio by 0.168.

Table 2.4: Log odds ratio for factors associated with a non-medically indicated caesarean section

Variable	Odds Ratio	Confidence Interval
Intercept	0.800 ***	(0.461-1.139)
Age category (years)		
15-19	-2.172 ***	(-2.503 – -1.841)
20-24	-2.533 ***	(-2.784 – -2.281)
25-29	-2.500 ***	(-2.742 – -2.259)
30-34	-2.413 ***	(-2.653 – -2.173)
35-39	-2.020 ***	(-2.260 – -1.779)
40-44	-1.373 ***	(-1.620 – -1.125)
Insurance status		
Mandatory	-0.713 ***	(-0.791 – -0.636)
Semi-private	-0.360 ***	(-0.450 – -0.270)
Urbanity level		
Urban	0.152 ***	(0.093 – 0.210)
Suburban	0.176 ***	(0.105 – 0.248)
Number of midwives per 1,000 births	-0.019 ***	(-0.023 – -0.016)

Number of nursing staff per hospital	0.243 ***	(0.195 – 0.291)
Income per capita (per CHF 10'000)	0.168 ***	(0.133 – 0.203)
AUC	0.625	

Significance levels: *** p<0.001; AUC=Area under the curve

Legend: A higher log odds ratio indicates a higher chance of giving birth by non-medically indicated CS. The following categories served as reference categories for categorical variables: 45-49 years (age), private (insurance status), rural (urbanity level).

The predictive power of the logistic regression is calculated by a ROC curve as it displays the trade-off between the detection of true positives and avoiding false positives based on the women's characteristics. ROC curves are also suitable for imbalanced data sets, as the group sizes between non-medically indicated CS and SVB (11,112 vs 176,908) is highly imbalanced. The performed logistic regression has an area under the curve (AUC) of 0.625, which indicates an acceptable quality of the model regarding the aim of this research.

2.1.5 Discussion

This paper aims to identify the percentage of non-medically indicated CS in Switzerland and second, to uncover levers for a change in the birth practice. Both questions are analysed based on a large Swiss dataset. The share of non-medically indicated CS amounts to almost 8% of all CS, which are more than 2,200 births annually in absolute terms. Research from the neighbouring country Germany reports one out of ten CS as not medically indicated. Additionally, the average non-medically indicated CS ratio in Germany is around 1% of the total number of births [30]. Other comparable studies only controlled for one diagnosis for identifying non-medically indicated CS [12, 24]. However, the presented approach allows for a more precise approximation of the number of CS without medical indication due to the holistic range of evaluation criteria.

The DRG cost weight for non-medically indicated CS generates higher payments compared to SVB. Reasons for this imbalance are the costs of expensive infrastructure (such as operating rooms and hospital bed charges) and higher personnel requirements

(such as interdisciplinary operation teams). On the other hand, a higher level of uncertainty exists for SVB. On the contrary, there is a greater degree of uncertainty associated with SVB as it is inherently less predictable and demands higher time and personnel resources. Furthermore, resources must be kept available such as additional staff for an emergency CS or operating room capacities although they might not be used. Hence, for future infrastructure (hospital) planning, also the non-usage of available resources needs to be considered when setting payment rates.

The average age of the childbearing mother does not significantly differ between the two investigated modes of birth. The largest age category is at both birth modes 30-34 years. However, for non-medically indicated CS the group of 35-39 years follows whereas for SVB the category of 25-29 years ranks second. Hence, women with non-medically indicated CS are in general slightly older than women having given birth by SVB. This is also supported by previous research [28, 31] which found that women with non-medically indicated CS tends to be older than women with SVB. This can be explained as age is a factor associated with a higher risk for complications at birth [18, 32].

Further, a difference in insurance status of the included women exists. A higher ratio of women with semi-private or private insurance is observable in the non-medically indicated CS group compared to SVB. Similar findings were also reported by Womack et al. [31] for the United States. Additionally, a positive correlation between the income per capita and the insurance status is visible. The logistic regression further strengthened the assumption that higher income and supplemental insurance increases the likelihood of non-medically indicated CS over SVB. Faisal-Cury & Menezes [13] argued that income only impacts the decision on the mode of birth in low-income countries. However, this holds also true for developed countries such as Switzerland as shown in this paper.

The number of midwives per 1,000 births per canton might have an impact on the likelihood of a woman being accompanied by a midwife during pregnancy. Research showed that women who benefit from a midwife-led care program are more likely to give birth spontaneously [33], face better outcomes and have a positive maternal

experience [11]. Contrastingly, another study claimed that midwife-led care models rather lead to an CS increase in high-resource countries [34]. Nevertheless, the present results support the first findings as woman living in a canton with more midwives have a higher chance of SVB. Furthermore, logistic regression indicates that the probability of a non-medically indicated CS is lower in cantons with a higher number of midwives, as every one-unit increase in the number of midwives reduces the log odds ratio by -0.018. This leads to the assumption that the presence of midwives has a positive impact on the mother's preferred mode of birth.

The paper identifies decisive factors for the decision on the mode of birth. To the authors' knowledge, this is the first paper that chooses the presented approach to identify non-medically indicated CS in Switzerland, as corresponding procedures, diagnoses or DRG codes do not exist. Age, number of nursing staff per bed, and income per capita are significantly higher for women with non-medically indicated CS and these factors also increase the log odds ratio of a CS, although no medical indication is present. On the other hand, women also show significant differences in the insurance class, urbanity level and the number of midwives per canton. Having mandatory insurance, living in rural areas or in areas with a high number of practicing midwives per 1,000 births decreases the chance of CS without medical indication.

As every empirical study, also this analysis faces some limitations. First, there is no classification code for non-medically indicated CS, whereby only an approximation of the number of non-medically indicated CS cases is possible. Second, despite the large amount of cases of a national dataset, the analysed data does not contain personal (such as personal beliefs, fear of birth or maternal request) nor interpersonal (such as care during pregnancy or obstetrician's tendency) factors. Additionally, clinical information is not available and it is unknown for example whether the woman is nulliparous, primiparous, or multiparous. Therefore, it is not possible to distinguish whether a non-medically indicated CS follows a previous CS. Having such variables available would improve the predictive power of the ROC curves for the logistic regression model. Furthermore, a mandatory nationwide collection of the information specifying the classification according to Robson's proposition would further

strengthen the explanatory power of research on uncovering reasons and outcomes for different birth modes in Switzerland.

2.1.6 Conclusion

We develop a strategy for identifying non-medically indicated CS in routine data in absence of any unique identifying factor. CS without medical indication can be determined based on a composition of DRGs, procedure classifications and diagnoses. Thereby, this study reveals that at least 8% of CS were on a non-medically indicated basis between 2014 and 2018. Factors significantly associated to non-medically indicated CS are higher age, supplemental insurance, higher income, more available staff at the hospital and living in urban areas. In contrast, living in an area with a higher density of practicing midwives significantly decreases the likelihood of giving birth by non-medically indicated CS, and thereby increases the likelihood of giving birth spontaneously. Health policy implications derived from this research are fourfold: first, obstetricians should be sensitized to special characteristics of women undergoing CS in absence of any medical indication and be encouraged to communicate the risks for both the mother and the neonatal more explicitly; second, the density of practicing midwives should be increased in case they can reduce the number of non-medically indicated CS and further promote the benefits of spontaneous birth; third, the DRG cost weights for non-medically indicated CS and SVB should be aligned to avoid inadequate incentive mechanisms to prefer non-medically indicated CS over SVB due to higher reimbursement rates. Giving birth to a child is different from typical treatments provided in a hospital, as personal beliefs and preferences must be taken into consideration.

References

1. Ye J, Zhang J, Mikolajczyk R, et al (2015) Association between rates of caesarean section and maternal and neonatal mortality in the 21st century: A worldwide population-based ecological study with longitudinal data. BJOG 123:745–753.

2. WHO (2015) WHO Statement on Caesarean Section Rates. Geneva

3. OECD (2019) Healt at a glance 2019 - OECD indicators. OECD Publishing, Paris

4. Chen I, Opiyo N, Tavender E, et al (2018) Non-clinical interventions for reducing unnecessary caesarean section. Cochrane Database of Systematic Reviews.

5. Ye J, Betrán AP, Guerrero Vela M, et al (2014) Searching for the optimal rate of medically necessary cesarean delivery. Birth 41:237–244.

6. WHO (2017) Robson classification: Implementation manual. Geneva

7. Robson M, Murphy M, Byrne F (2015) Quality assurance: The 10-Group Classification System (Robson classification), induction of labor, and cesarean delivery. International Journal of Gynecology & Obstetrics 131:S23–S27.

8. O'Donovan C, O'Donovan J (2018) Why do women request an elective cesarean delivery for non-medical reasons? A systematic review of the qualitative literature. Birth 45:109–119.

9. Eide KT, Morken N-H, Bærøe K (2019) Maternal reasons for requesting planned cesarean section in Norway: A qualitative study. BMC Pregnancy Childbirth 19:102.

10. Stjernholm YV, Petersson K, Eneroth E (2010) Changed indications for cesarean sections. Acta Obstetricia et Gynecologica 89:49–53.

11. Betrán AP, Temmerman M, Kingdon C, et al (2018) Interventions to reduce unnecessary caesarean sections in healthy women and babies. The Lancet 392:1358–1368.

12. Karlström A, Rådestad I, Eriksson C, et al (2010) Cesarean section without medical reason, 1997 to 2006: A Swedish register study. Birth 37:11–20.

13. Faisal-Cury A, Menezes PR (2006) Factors associated with preference for cesarean delivery. Rev Saude Publica 40:226–232.

14. Fioretti B, Reiter M, Betrán A, Torloni M (2015) Googling caesarean section: a survey on the quality of the information available on the Internet. BJOG 122:731–739.

15. Boerma T, Ronsmans C, Melesse DY, et al (2018) The global epidemiology of Caesarean Sections : major increases and wide disparities T. Lancet 392:1341–1348.

16. Betrán AP, Torloni MR, Zhang JJ, Gülmezoglu AM (2016) WHO statement on caesarean section rates. BJOG 123:667–670.

17. Dahlquist K, Stuart A, Källén K (2022) Planned cesarean section vs planned vaginal delivery among women without formal medical indication for planned cesarean section: A retrospective cohort study of maternal short-term complications. Acta Obstet Gynecol Scand 101:1026–1032.

18. Herstad L, Klungsøyr K, Skjærven R, et al (2016) Elective cesarean section or not? Maternal age and risk of adverse outcomes at term: A population-based registry study of low-risk primiparous women. BMC Pregnancy Childbirth 16:1–11.

19. Sandall J, Tribe RM, Avery L, et al (2018) Short-term and long-term effects of caesarean section on the health of women and children. The Lancet 392:1349–1357.

20. MacDorman MF, Declercq E, Menacker F, Malloy MH (2008) Neonatal mortality for primary cesarean and vaginal births to low-risk women: application of an “Intention-to-Treat” model. Birth 35:3–8.

21. Souza JP, Gülmezoglu AM, Lumbiganon P, et al (2010) Caesarean section without medical indications is associated with an increased risk of adverse short-term maternal outcomes: The 2004-2008 WHO global survey on maternal and perinatal health. BMC Med 8:1–10.

22. Berg J, Källén K, Andolf E, et al (2023) Economic evaluation of elective cesarean section on maternal request compared with planned vaginal birth — Application to Swedish setting using national registry data. Value in Health 26:639–648.

23. WHO (2022) Preterm birth. https://www.who.int/news-room/fact-sheets/detail/preterm-birth. Accessed 9 Jan 2023

24. Henry OA, Gregory KD, Hobel CJ, Platt LD (1995) Using ICD-9 codes to identify indications for primary and repeat cesarean sections: Agreement with clinical records. Am J Public Health 85:1143–1146.

25. Korst LM, Gregory KD, Gornbein JA (2004) Elective primary caesarean delivery: Accuracy of administrative data. Paediatr Perinat Epidemiol 18:112–119.

26. Gregory KD, Korst LM, Gornbein JA, Platt LD (2002) Using administrative data to identify indications for elective primary cesarean delivery. Health Serv Res 37:1387–1401.

27. Gregory KD, Korst LM, Platt LD (2001) Variation in elective primary cesarean delivery by patient and hospital factors. Am J Obstet Gynecol 184:1521–1534.

28. Lin HC, Xirasagar S (2005) Maternal age and the likelihood of a maternal request for cesarean delivery: A 5-year population-based study. Am J Obstet Gynecol 192:848–855.

29. SPSS Inc. (2020) Statistical Package for Social Science for Windows, Version 27.0

30. Wacker J (2010) Geburtshilfe. Kaiserschnitt vs. Natürliche Geburt. Geburtshilfe Frauenheilkd 70:840–843.

31. Womack LS, Sappenfield WM, Clark CL, et al (2014) Maternal and hospital characteristics of non-medically indicated deliveries prior to 39 weeks. Matern Child Health J 18:1893–1904.

32. Räisänen S, Gissler M, Kramer MR, Heinonen S (2014) Influence of delivery characteristics and socioeconomic status on giving birth by caesarean section - a cross sectional study during 2000-2010 in Finland. BMC Pregnancy Childbirth 14:1–9.

33. Sandall J, Soltani H, Gates S, et al (2013) Midwife-led continuity models versus other models of care for childbearing women. Cochrane Database Systematic Reviews.

34. Lawton BA, Koch A, Stanley J, Geller SE (2013) The effect of midwifery care on rates of cesarean delivery. International Journal of Gynecology and Obstetrics 123:213–216.

Appendix

Additional file 1: Diagnoses used to exclude cases from the not medically indicated CS dataset

ICD-10-GM codes	Definition
O32·-	Care of the mother in the event of a detected or suspected positional and attitude anomaly of the fetus
O33·-	Care of the mother in case of identified or suspected mismatch between fetus and pelvis
O34·2	Care for the mother with uterine scarring due to previous surgery
O61·0	Failed medical induction of labor
O61·1	Failed instrumental induction of labor
O62·-	Abnormal labor
O63·-	Protracted birth
O64·-	Obstruction of birth due to positional, postural and attitude anomalies of the fetus
O65·-	Obstruction of birth due to anomaly of the maternal pelvis
O68·-	Complications of labor and delivery due to fetal distress [fetal danger]
O69·-	Complications of labor and delivery due to umbilical cord complications

Additional file 2: Treatments used to exclude cases from the not medically indicated CS dataset

CHOP codes	Definition
74·0X·20	Classic cesarean section, emergency
74·1X·20	Deep cervical cesarean section, emergency
74·2X·20	Extraperitoneal cesarean section, emergency
74·4X·20	Cesarean section of a different type, emergency
74·99·20	Other unspecified types of cesarean sections, emergency

CHOP = Swiss operation procedures

2.2 Medication adherence halves COPD patients' hospitalization risk—Evidence from Swiss health insurance data

Anja Y. Bischof

Johannes Cordier

Justus Vogel

Alexander Geissler

Published in *npj Primary Care Respiratory Medicine*

2.2.1 Abstract

Background: Medication adherence is vital for patients suffering from Chronic Obstructive Pulmonary Disease (COPD) to mitigate long-term consequences. The impact of poor medication adherence on inferior outcomes like exacerbations leading to hospital admissions is yet to be studied using real-world data.

Method: Using Swiss claims data from 2015-2020, we group patients into five categories according to their medication possession ratio. By employing a logistic regression, we quantify each category's average treatment effect of the medication possession ratio on hospitalized exacerbations.

Results: 13,557 COPD patients are included in the analysis. Patients with high medication adherence (daily medication reserve of 80% to 100%) are 51% less likely to incur exacerbation following a hospital stay than patients with the lowest medication adherence (daily medication reserve of 0% to 20%).

Conclusions: The study shows that medication adherence varies strongly among Swiss COPD patients. Furthermore, high medication adherence immensely de-creases the risk of hospitalized exacerbations.

2.2.2 Background

In 2019, Chronic Obstructive Pulmonary Disease (COPD) was the third leading cause of death worldwide [1]. Typical COPD symptoms are shortness of breath, sputum, and cough [2]. In Switzerland, about 400,000 people suffer from COPD [3]. Annual direct healthcare expenditures are CHF 603-847 million, while the indirect costs of COPD amount up to CHF 932 million per year [4]. The prevalence of COPD is particularly high in individuals with low socioeconomic status [5].

To treat COPD sustainably and slowing down the progress of the disease, COPD patients are prescribed long- and short-acting medication. Long-acting medication should be taken regularly to reach its full effect. Thus, it differs from short-acting medication, which is taken according to acute need such as during shortness of breath [2, 6]. Taking long-acting medication as prescribed alleviates disease-specific symptoms, slows disease progression and prevents hospitalization due to acute exacerbation [7]. Acute exacerbations are especially harmful as they result in additional therapy, imply a setback in COPD patients' health-related quality of life, raise the risk of iatrogenic harm, and diminish lung function long-term [8, 9].

High medication adherence can help to better control disease progression, reduce hospitalizations due to acute exacerbations and ultimately reduce cost [7]. Accord-ing to the WHO, medication adherence refers to "the extent to which the person's behavior corresponds with the agreed recommendations from a healthcare provid-er" [10]. A globally conducted study concluded that in clinical trials medication adherence reaches up to 80% [11], while in real world settings, based on US data, medication adherence is solely around 30% [12]. Similarly, a Korean study showed that adherence is low by analyzing a nationwide health insurance dataset. The per-centage of patients with high adherence was only around 34.7% and declined to 22.3% over four years [13].

Implementing high medication adherence in COPD patients depends on, among others, patients' beliefs about medication and concerns about potential side effects, experiences and satisfaction with medication effectiveness, personal circumstances and health status [14]. Moreover, the patient-physician relationship and the physi-

cian's availability are also relevant factors for a high medication adherence [15–18]. Further, COPD patients report personal challenges related to medication ad-herence, e.g., understanding appropriate intake or adhering to the correct timing of intake [18].

Some studies have already investigated the relation of medication adherence and acute exacerbations [13, 19, 20]. However, only few studies focused on the rela-tionship between medication adherence and hospitalization caused by acute exac-erbations (see exemplarily Suh et al. [21] , Chen et al. [22], or Weir et al. [23]). Nevertheless, the medication adherence of COPD patients in Switzerland is un-known.

We use Swiss health insurance data to investigate medication adherence of Swiss COPD patients and to analyse effects of low medication adherence on the risk of hospitalization. Hence, our research questions are:

- How is medication adherence distributed among COPD patients?
- How does inadequate medication adherence impact the likelihood for a hospitalized exacerbation?

We aim to use these results to contribute to the ongoing discussion regarding strat-egies for improving medication adherence among individuals with this chronic condition. Given the substantial prevalence of this patient group, even minor ad-justments could potentially yield significant benefits. Similar to previous studies [19–21], we estimate medication adherence from the proportion of days covered (PDC), assuming that COPD patients had enough medication at home. We investi-gate whether low adherence to prescribed long-acting medications made exacerba-tions more likely using a logistic regression. Thus, we categorize COPD patients into categories according to their PDC.

2.2.3 Methods

Data

We used claims data from Groupe Mutuel, a Swiss health insurance with a 10% market share [24], from 2015 to 2020.

The dataset included basic claims data, contract information, and billing data for each insured person (see Table 2.5). Billing data covered the treatments and pre-scriptions invoiced during the observation period. For each invoice, the date of treatment/prescription, the amount of the treatment/prescription in CHF, the share covered by the insurance, the tariff at which the treatment/prescription was billed, and the tariff code were available. In case of inpatient treatments billed under the SwissDRG tariff, the DRG code, the length of stay, and the reason for the stay (illness, accident or pregnancy) were included.

Table 2.5: Overview of data types and variables included in dataset

Data type	Variable
Basic claims data	Year of birth
	Nationality (Swiss vs non-Swiss)
Contract information data	Type of insurance
	Start (and if available) end date of contract
	Deductible level
	Premium reduction (only if received)
Billing data (per treatment/ prescription)	Date
	Amount in CHF
	Portion covered by insurance
	Applicable tariff (TARMED, TARPSY, SwissDRG)
	Tariff code
	Only for inpatient stays billed at SwissDRG: Length of hospital stay Reason for stay (illness, accident, or pregnancy)
	For medication: Pharma code (including package size and daily recommended dosis)

Sample

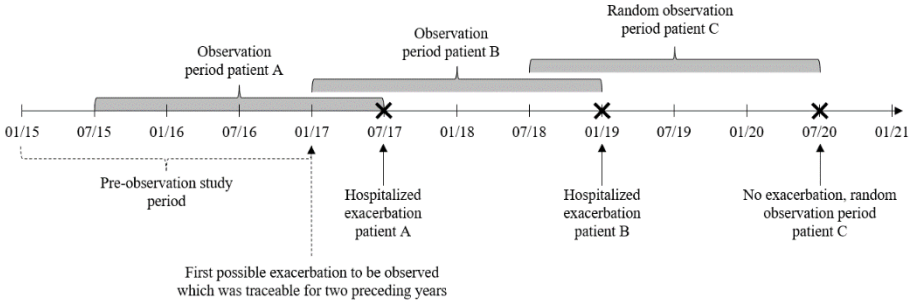
We included 13'557 insurees with prescriped COPD medication in our sample. In Switzerland, health insurances are not allowed to save data on diagnoses. Thus, to identify COPD patients in our dataset, we defined the following three inclusion criteria:

- Therapy with long-acting medication specific for chronic pulmonary diseases: Long-acting medication should have be taken on a regular basis, but the medication or rather the active component itself might have changed over time depending on the COPD patient's health status. Therefore, all medications defined as long-acting medications in the GOLD standard [2] were grouped together to identify COPD patients: Long-acting beta2 agonists (LABA), long-acting antimuscarinic antagonists (LAMA), the combination of LABA and LAMA, the combination of LABA and corticosteroid (ICS), the combination of LAMA and ICS, and the triple combination of LABA, LAMA, and ICS. All COPD patients who received any of these medications at least once during the observation period were included.
- Age (> 40 years) to distinguish between COPD and asthma patients: Long-acting medication can be prescribed to asthma as well as to COPD pa-tients. The literature suggests that patients older than 40 years are more likely to suffer from COPD than from asthma if they take at least one of the above-mentioned medications [2]. Thus, to exclude asthma patients, we only included insured persons older than 40 years in our sample.
- Availability of claims data for at least two consecutive years prior to the exacerbation: A hospitalized exacerbation was identified with the base DRG E65 (COPD). The date of the hospitalized exacerbation was the cut-off date and the observation of a COPD patient started two years before (see Figure 2.2). Thereby, 2015 and 2016 served as a pre-observation peri-od, as exacerbations between 1st January 2015 and 31st December 2016 were not traceable for two years within our dataset. Additionally, for COPD patients with multiple hospitalized exacerbations, only the first hospitalized

exacerbation, which allows an ex-ante observation period of two years (i.e., from 01/2017 onwards), was used to ensure independent observations. For COPD patients without hospitalized exacerbations, by definition, there is no exacerbation day that could have been used as a cut-off date. Therefore, we chose a random period of two years. This ensured independent observations even for non-hospitalized COPD patients.

- exacerbation, which allows an ex-ante observation period of two years, is used to ensure independent observations. For COPD patients without hospitalized exacerbations, by definition, there is no exacerbation day that could be used as a cut-off date. Therefore, we chose a random period of two years. This ensures independent observations even for non-hospitalized COPD patients.

Figure 2.2: Schematic representation of the study design



Legend: The period between 01/15 and 12/16 functioned as a pre-observation period. In case COPD patients underwent a hospitalization due to an exacerbation during this period, these hospitalizations were controlled for in the analysis of the impact of medication adherence on hospitalizations. However, hospitalizations due to exacerbations in the pre-observation period were not treated as observations but only as a control variable. The first possible hospitalization due to an exacerbation to be observed in the sample stated 01/17, where patients' medication adherence could be tracked for the two preceding years in the pre-observation period. The point in time of the hospitalized exacerbation marked the cut-off date and determined the two years of observation prior to the hospitalized exacerbation (e.g., observation period patients A and B). For COPD patients without a hospitalized exacerbation, a random date marked the cut-off date and the two preceding years were used for the observation (e.g., random observation period patient C).

Statistical Model

Using real world data, we approximated COPD patient's medication adherence by calculating their PDC. The PDC represents the proportion of days covered by COPD patients' personal home-inventory of long-acting medication taken with inhalers during the period of investigation [21]. To estimate a patient's medication adherence and its impact on the likelihood of hospitalization due to exacerbations, we applied the following calculations:

Patient Medication Reserve: We calculated a patient's remaining supply of a long-acting medication on a specific day. The reserve depended on the previous day's reserve, the period when the medication was bought last respecting the package size (in number of doses), and the recommended daily dose. If a patient switched between different medication categories (e.g., from LABA to LAMA), we assumed both medications were taken until one was no longer purchased.

Proportion of days covered (PDC): We determined the PDC for each COPD patient. We calculated the proportion of days by dividing the number of days with enough medication supply by the total observation period. The observation period in this study was set to 720 years, i.e., two years.

To assess the causal effect of medication adherence on the probability of hospital-ized exacerbations, we relied on the causal Rubin effect [25]. We measured the PDC's Average Treatment Effect (ATE) on hospitalized exacerbations [26]. COPD patients were divided into five categories (C1 = 0 to <20%, C2 = 20% to < 40%, C3 = 40% to < 60%, C4 = 60% to < 80%, and C5 = 80% to < 100%) based on their PDC values. ATE is the difference between the expected outcomes (hospitaliza-tion) for patients in different adherence categories.

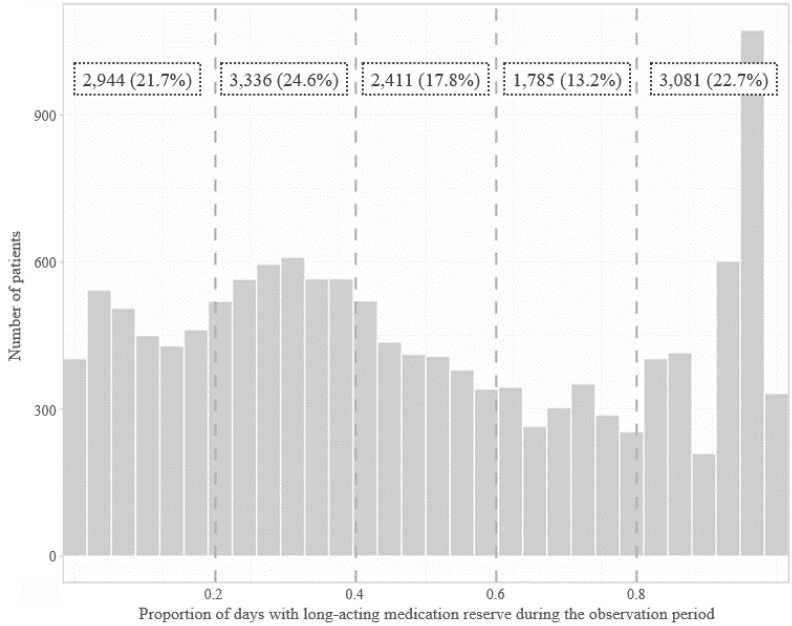
To calculate the effect of medication adherence on hospitalized exacerbations, a logistic regression model was used due to the binary nature of the dependent varia-ble (hospitalization or no hospitalization). In the model, we controlled for socioec-onomic factors (such as age, franchise level, chosen franchise increase or decrease, and premium reduction), and medical factors (such as hospitalized exacerbations in the

first two years of the observation period, purchase of short-acting medication, purchase of methylxanthine, purchase of phosphodiesterase 4 inhibitors, and purchase of mucolytic). The influence of medication adherence was estimated as a linear combination of the category values (C2 to C5) (for further information see Supplementary Information I).

2.2.4 Results

Figure 2.3 shows the distribution diagram of the number of COPD patients by PDC. Category 1 on the left, for instance, includes all COPD patients who had a positive medication reserve on 0 to 20% of the days of the observation period

Figure 2.3: Grouping of COPD patients according to the proportion of days covered (PDC)



Legend: The X-axis shows the medication possession ratio (PDC) over all days in the observation period and the Y-axis plots the number of COPD patients. Categories are indicated with dashed lines. The number of COPD patients per category and their share of the total sample are indicated in the rectangles at the top.

46.3% of the COPD patients are grouped in the first two categories, i.e., about half of the patients has a PDC of less than 40% of the days during the observation peri-od. At the same time, the number of COPD patients in categories 3 and 4 is rela-tively low and second highest in category 5, i.e., the proportion of COPD patients who have a rather high PDC.

Of the 13,557 COPD patients, 1,721 (12.6%) were hospitalized due to an exacerba-tion. The lowest rate of exacerbations is present in category 5 with 225 out of 3,081 (7.2%) COPD patients, whereas the first category includes the highest rate of exac-erbations with 619 out of 2,944 (21.0%) COPD patients (see Table 2.6). According-ly the prescription of short-acting medications – provided during an acute worsen-ing of the lung function – is highest in category 1. Regarding age, COPD patients in category 5 are on average about 3.5 years younger than in category 1. Furthermore, the rate of premium reduction is around 40% for all categories, whereas COPD patients in category 1 most often receive premium reduction with about 42.8%, meaning they have the lowest incomes.

Regression results: Impact on the exacerbation likelihood

Regarding the results of the main model, the regression analysis showed a signifi-cant effect between higher medication adherence and hospitalized exacerbations (see Table 2.7). Control variables with significant associations to hospitalized exacerbations were premium reduction, age, prescription of short-acting medica-tions, and prescription of phosphodiesterase-4 inhibitors.

Figure 2.4 shows a large effect of higher medication adherence on the probability of exacerbation, using category 1 as the reference value for each of the categories 2 to 5. If the odds ratio of a category is less than 1, this means that COPD patients in this category are less likely to be hospitalized due to an exacerbation during the observation period compared to category 1. For example, Figure 2.4 displays that the probability of an exacerbation is already about 28% lower if the PDC is >20-40% (category 2). COPD patients with a PDC of more than 80% (category 5) have a 51% lower probability of being hospitalized due to an exacerbation. We observe a

decreasing marginal utility of continuous medication reserve (i.e., belonging to a higher category). Further, each category's whisker stands for the confidence inter-val. The confidence intervals decrease for higher medication adherence – indicat-ing that a higher medication adherence allows for more accurate estimates.

Table 2.6: Number of COPD patients, number and proportion of exacerbations, and mean of the model variables per category

	Category 1	Category 2	Category 3	Category 4	Category 5	Total
Number of observations (share of total)	2'944 (21.7%)	3'336 (24.6%)	2'411 (17.8%)	1'785 (13.2%)	3'081 (22.7%)	13'557 (100%)
Number (share of category) exacerbations°	619 (21.0%)	446 (13.4%)	268 (11.1%)	163 (9.1%)	225 (7.2%)	1'721 (12.6%)
Average age in years (standard deviation)	68.45 (10.64)	66.51 (11.12)	66.05 (11.27)	66.08 (11.36)	64.86 (11.30)	66.42 (11.18)
Premium reduction	42.8%	41.6%	40.0%	40.1%	38.6%	40.7%
Increase of the deductible	2.0%	2.6%	2.6%	2.6%	2.4%	2.4%
Reduction of the deductible	5.5%	7.6%	6.7%	4.4%	5.3%	6.0%
Hospitalized exacerbation during the observation period+	11.2%	5.5%	4.7%	3.3%	1.4%	5.4%
Purchase of short-acting medication	28.3%	20.8%	19.0%	17.1%	12.5%	19.7%
Purchase of methylxanthine	0.3%	0.2%	0.3%	0.1%	0.1%	0.2%
Purchase of phosphodiesterase 4 inhibitors	2.3%	1.4%	0.7%	0.5%	0.1%	1.1%
Purchase of of mucolytic	9.5%	9.0%	8.9%	8.6%	8.6%	9.0%

Legend: ° This variable indicates the total number and the corresponding share of a category of all exacerbations between 2015 and 2020.

+ This variable relates to the exacerbations which have taken place during the two-year observation period of individuals belonging to the different categories.

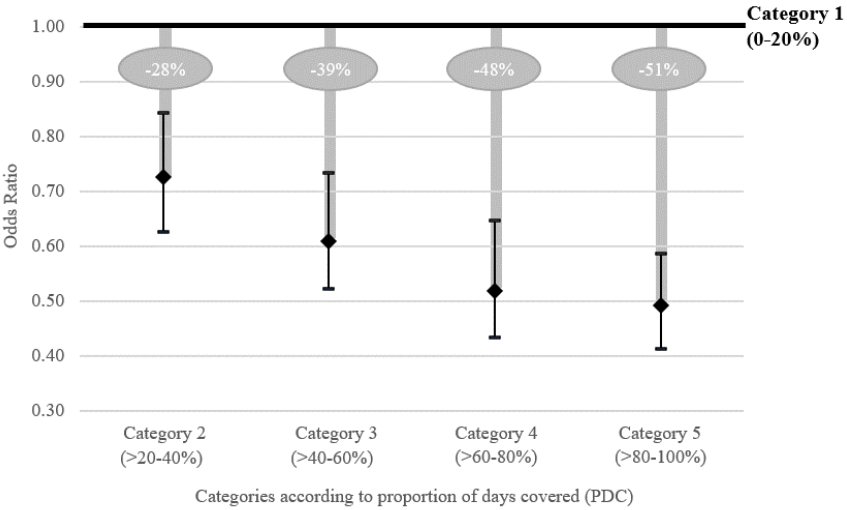
Table 2.7: Regression output on the association of higher medication adherence with hospitalized exacerbations

Variable	Abbreviation	Estimate coefficient	Std. error	OR	95% CI	Significance marker
Category 2 of the proportion of days with medication reserve	Q2	-0.319	0.074	0.73	(0.63 - 0.84)	***
Category 3 of the proportion of days with medication reserve	Q3	-0.479	0.085	0.62	(0.52 - 0.73)	***
Category 4 of the proportion of days with medication reserve	Q4	-0.637	0.100	0.53	(0.43 - 0.65)	***
Category 5 of the proportion of days with medication reserve	Q5	-0.709	0.089	0.49	(0.41 - 0.59)	***
Current deductible = 500	F_{500}	-0.018	0.065	0.98	(0.86 - 1.12)	
Current deductible = 1'000	F_{1000}	-0.115	0.186	0.89	(0.61 - 1.29)	
Current deductible = 1'500	F_{1500}	0.238	0.140	1.27	(0.96 - 1.68)	
Current deductible = 2'000	F_{2000}	0.024	0.371	1.02	(0.49 - 2.15)	
Current deductible = 2'500	F_{2500}	-0.043	0.236	0.96	(0.60 - 1.54)	
Premium reduction	R	0.261	0.057	1.30	(1.16 - 1.45)	***
Age	A	0.017	0.002	1.02	(1.01 - 1.02)	***
Increase of the franchise	F+	-0.138	0.237	0.87	(0.54 - 1.40)	
Reduction of the franchise	F-	-0.181	0.138	0.83	(0.63 - 1.10)	

Variable	Abbreviation	Estimate coefficient	Std. error	OR	95% CI	Significance marker
Hospitalized exacerbation during the observation period	E_{t-2}	1.580	0.087	4.85	(4.08 - 5.78)	***
Prescription of short-acting medication	B	1.280	0.058	3.60	(3.20 - 4.04)	***
Methylxanthine prescription	M	0.510	0.476	1.67	(0.64 - 4.31)	
Prescription of phosphodiesterase 4 inhibitors.	P	1.410	0.194	4.10	(2.78 - 6.04)	***
Prescription of mucolytic	K	-0.186	0.100	0.83	(0.68 - 1.01)	
Intercept	β_0	31.600	5.020		(0.63 - 0.84)	***

Legend: The significance levels were set at p<0.1*, p<0.05**, and p<0.01***; OR = Odds ratio; CI = Confidence Interval

Figure 2.4: Hospitalization likelihood due to exacerbations in relation with higher medication adherence



Legend: The figure illustrates the decrease in the likelihood of hospitalized exacerbations per category. Category 1 functions as the reference for the comparison with categories 2 to 5. The whiskers display the 95% confidence intervals.

2.2.5 Discussion

Furthermore, the share of premium reduction recipients among the COPD patients is remarkably high. Premium reductions are granted to people living in modest economic circumstances, where the family and financial situation is decisive for receiving premium reductions [27]. In our sample, around 39% to almost 43% of all COPD patients – depending on the corresponding category – receive premium reduction. Compared to the average Swiss population, approximately 28% of all insured persons received a premium reduction in 2020 [28]. The higher share of premium reduction recipients is a sign for lower socioeconomic status among COPD patients – as already highlighted by Eisner et al. [5].

In addition, potential influencing factors on hospitalized exacerbations were analysed. First, higher medication adherence lowers the risk for acute exacerbations, which is in line with the literature [21–23, 29]. Our analysis shows that the risk for

hospitalized exacerbations can be decreased around 51% in case of high adherence (category 5) compared to non- respectively low adherence (category 1). However, also COPD patients in categories 2 to 4 benefit from lower risks for exacerbations (between 28%-48% less likelihood) compared to COPD patients belonging to category 1. Similar results were also retrieved in RCTs where, for example, Vestbo et al. [11] showed that adherence to inhaled medication is significantly associated with reduced risk of death and admission to hospital due to exacerbations in COPD. Furthermore, Torres-Robles et al. [30] found that medication adherence improves clinical outcomes in COPD patients and that interventions on increasing medica-tion adherence show positive results. These insights shows that COPD patients would already benefit if general practitioners, outpatient specialists, and other medical professionals could manage supporting patients in moderately increasing their medication adherence. Potential approaches for increasing COPD patients' medication adherence are individualized educational interventions [31, 32], reminders, motivational strategies, shared decision making or direct feedback on medication use [32]. These approaches could also be incorporated into digital health interventions (for example Spielmanns et al. [33]).

Medication adherence is categorized into three main stages: initiation, implementation and discontinuation [34]. The focus of this study is on the intersection of the two latter stages – implementation and discontinuation – by calculating the PDC. Although most COPD patients do not take their medications as prescribed (i.e., no full implementation), they do also not completely quit taking their medications (i.e., no full discontinutation). This inconsistent behaviour of COPD patients leads to the assumption that they face certain challenges in their adherence habit such as personal beliefs about the effectiveness of the prescribed medication, patient-physician relationship or insecurities in the right usage of the medication [14–17]. Rising awareness of these challenges in medication adherence plus knowing the factors (such as higher age, receiving premium reduction, or prior exacerbations) being significantly associated to a higher risk of exacerbations, health policy initia-tives

should sustainably strengthen COPD patients' medication adherence, e.g. when implementing chronic care programs and/or digital health assistants.

The limitations of this study are twofold: Firstly, due to the missing patient diag-nosis in the dataset, the number of COPD patients included in the dataset are ap-proximated with two inclusion criteria: taking prescribed long-acting medication and age over 40 years. This means that, on the one hand, we might underestimate the number of COPD patients as patients might not take any medication yet or they stopped taking medication. On the other hand, we might include asthma patients in our sample because there might be asthma patients older than 40 years taking long-acting prescriptions. Both issues should be marginal, however, and not strongly influence our results. Moreover, the advantage of claims data enabling us to ob-serve patients intersectorally is more essential than the limitation of potentially underestimating the number of COPD patients or including single patients with a different diagnosis.

Secondly, due to the empirical setting of this study, the number of observed varia-bles is limited. Two areas with potentially unobserved control variables include the COPD patient's personal motivation and financial situation. Personal motivation is difficult to quantitatively replicate based on the used dataset but may have an im-pact on both medication reserve and the likelihood of exacerbation. Furthermore, high personal motivation might positively influence COPD patients' health literacy and their body awareness. These two factors potentially reduce the exacerbation risk. However, in our study, we only focus on the association of the PDC with ex-acerbation likelihood. Therefore, we might overestimate this effect as we neglect the association of motivation with exacerbation likelihood. Additionally, the finan-cial situation may influence the utilization of healthcare services, as there is some cost sharing in the Swiss insurance system between insured persons and insurances. We aimed to consider this by including premium reduction as control variable. However, it is only a proxy, since the variable premium reduction only adjusts for lower income groups. If there is no premium reduction, the exact income level is still unknown.

For future research, a better understanding of COPD patients' needs on potential supporting mechanisms to promote higher medication adherence has to be estab-

lished. Through a deepened understanding of current needs or challenges faced by COPD patients, researchers may be able to develop supportive tools to empower COPD patients in their daily lives to deal with the challenges of COPD and to increase their medication adherence. On the one hand, the risk of hospitalized exacerbations can be decreased, whereas on the other hand, the COPD patient's health-related quality of life should be conserved through a potential slowdown of the progression of the disease.

2.2.6 Conclusion

To conclude, medication adherence according to prescription halves the risk for hospitalized exacerbations. Generally, medication adherence is low meaning that there is vast potential to improve patient-relevant outcomes without "inventing" a new treatment but by supporting COPD patients. There are two main levers to do this: (1) Strengthening health literacy so that patients understand the importance of regular medication intake and (2) structured, physician-led support programs such as disease management programs or chronic care programs. The first lever is pa-tient-focused whereas the second physician-focused, yet both can be supported digitally. Additionally, COPD patients' support needs should be further analysed and focused on in future research, especially in the context of digital support tools. Future policy changes in Switzerland and elswhere should strengthen the incentive structures for physicians and COPD patients to use disease management programs and to finance digital health interventions.

References

1. World Health Organization (2020) Global health estimates 2020: Deaths by cause, age, sex, by country and by region, 2000-2019

2. GOLD (2021) Global Initiative for Chronic Obstructive Lung Disease

3. BAG (2020) Chronische Atemwegserkrankungen. https://www.bag.admin.ch/bag/de/home/krankheiten/krankheiten-im-ueberblick/chronische-atemwegserkrankungen.html

4. Wieser S, Tomonaga Y, Riguzzi M, et al (2014) Die Kosten der nichtübertragbaren Krankheiten in der Schweiz - Schlussbericht. Winterthur

5. Eisner MD, Blanc PD, Omachi TA, et al (2011) Socioeconomic status, race and COPD health outcomes. J Epidemiol Community Health (1978) 65:26–34.

6. Demeyer H, Louvaris Z, Frei A, et al (2017) Physical activity is increased by a 12-week semiautomated telecoaching programme in patients with COPD: a multicentre randomised controlled trial. Thorax 72:415–423.

7. Davis JR, Wu B, Kern DM, et al (2017) Impact of nonadherence to inhaled corticosteroid/LABA therapy on COPD exacerbation rates and healthcare costs in a commercially insured US population. Am Health Drug Benefits 10:92–102.

8. Donaldson GC, Seemungal TAR, Bhowmik A, Wedzicha JA (2002) Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax 57:847–852.

9. Miravitlles M, Ferrer M, Pont A, et al (2004) Effect of exacerbations on quality of life in patients with chronic obstructive pulmonary disease: a 2 year follow up study. Thorax 59:387–395.

10. World Health Organization (2003) Adherence to long-term therapies: Evidence for action. Geneva

11. Vestbo J, Anderson JA, Calverley PMA, et al (2009) Adherence to inhaled therapy, mortality and hospital admission in COPD. Thorax 64:939–943.

12. Rolnick SJ, Pawloski PA, Hedblom BD, et al (2013) Patient characteristics associated with medication adherence. Clin Med Res 11:54–65.

13. Kim J-A, Lim MK, Kim K, et al (2018) Adherence to inhaled medications and its effect on healthcare utilization and costs among high-grade chronic obstructive pulmonary disease patients. Clin Drug Investig 38:333–340.

14. Bhattarai B, Walpola R, Mey A, et al (2020) Barriers and strategies for improving medication adherence among people living with COPD: A systematic review. Respir Care 65:1738–1750.

15. George J, Kong DCM, Thoman R, Stewart K (2005) Factors associated with medication nonadherence in patients with COPD. Chest 128:3198–3204.

16. Cecere LM, Slatore CG, Uman JE, et al (2012) Adherence to long-acting inhaled therapies among patients with chronic obstructive pulmonary disease (COPD). COPD: Journal of Chronic Obstructive Pulmonary Disease 9:251–258.

17. Laforest L, Denis F, van Gansea E, et al (2010) Correlates of adherence to respiratory drugs in COPD patients. Primary Care Respiratory Journal 19:148–154.

18. O'Toole J, Krishnan M, Riekert K, Eakin MN (2022) Understanding barriers to and strategies for medication adherence in COPD: a qualitative study. BMC Pulm Med 22:98.

19. Simoni-Wastila L, Wei Y-J, Qian J, et al (2012) Association of chronic obstructive pulmonary disease maintenance medication adherence with all-cause hospitalization and spending in a Medicare population. Am J Geriatr Pharmacother 10:201–210.

20. Toy EL, Beaulieu NU, McHale JM, et al (2011) Treatment of COPD: Relationships between daily dosing frequency, adherence, resource use, and costs. Respir Med 105:435–441.

21. Suh H-S, Chang M-S, Yu I, et al (2022) Adherence to long-acting inhaler use for initial treatment and acute exacerbation of chronic obstructive pulmonary disease: A retrospective cohort study. J Pers Med 12:2073.

22. Chen R, Gao Y, Wang H, et al (2020) Association between adherence to maintenance medication in patients with COPD and acute exacerbation occurrence and cost in China: A retrospective cohort database study. Int J Chron Obstruct Pulmon Dis Volume 15:963–971.

23. Weir DL, Bai YQ, Thavorn K, et al (2023) Non-Adherence to COPD medications and its association with adverse events: A longitudinal population based cohort study of older adults. Ann Epidemiol 21:S1047-2797(23)00228-4.

24. Groupe Mutuel (2023) Key figures. https://www.groupemutuel.ch/en/groupe-mutuel/our-company/key-figures.html. Accessed 9 Aug 2023

25. Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 66:688–701.

26. Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55

27. BAG (2020) Krankenversicherung: Prämienverbilligung. https://www.bag.admin.ch/bag/de/home/versicherungen/krankenversicherung/krankenversicherung-versicherte-mit-wohnsitz-in-der-schweiz/praemienverbilligung.html. Accessed 6 Mar 2023

28. Ecoplan (2022) Wirksamkeit der Prämienverbilligung - Monitoring 2020. Bern

29. Vestbo J (2014) COPD: Definition and phenotypes. Clin Chest Med 35:1–6.

30. Torres-Robles A, Benrimoj SI, Gastelurrutia MA, et al (2022) Effective-ness of a medication adherence management intervention in a community pharmacy setting: a cluster randomised controlled trial. BMJ Qual Saf 31:105–115.

31. Sánchez-Nieto JM, Bernabeu-Mora R, Fernández-Muñoz I, et al (2022) Effectiveness of individualized inhaler technique training on low adher-ence (LowAd) in ambulatory patients with COPD and asthma. NPJ Prim Care Respir Med 32:1.

32. van de Hei SJ, Dierick BJH, Aarts JEP, et al (2021) Personalized medica-tion adherence management in asthma and chronic obstructive pulmonary disease: A review of effective interventions and development of a practi-cal adherence toolkit. J Allergy Clin Immunol Pract 9:3979–3994.

33. Spielmanns M, Boeselt T, Huber S, et al (2020) Impact of a smartphone application (KAIA COPD app) in combination with Activity Monitoring as a maintenance prOgram following PUlmonary Rehabilitation in COPD: the protocol for the AMOPUR Study, an international, multicenter, paral-lel group, randomized, controlled study. Trials 21:636.

34. Vrijens B, de Geest S, Hughes DA, et al (2012) A new taxonomy for de-scribing and defining adherence to medications. Br J Clin Pharmacol 73:691–705.

Appendix

Supplementary Information I: Formulas on the calculation of the proportion of days covered (PDC)

To approximate the PDC, the following formula were used:

$$R_{t,LA} = \max\left(R_{t-1,LA} + \mathbb{1}_{\{Medication\ bought\ in\ Period\ t\}} S_{LA} - D_{LA}, 0\right) \quad (1)$$

$$MPR = \frac{\sum_{t=1}^{T} \mathbb{1}_{\{R_{t,LA}>0\}}}{T} \quad (2)$$

Formula (1) represents the patient's reserve *R* of a long-acting medication *LA* on a particular day *t* in recursive form. The reserve $R_{t,LA}$ depends on the reserve $R_{t-1,LA}$ of the previous day *(t-1)* and the period when medication was bought last respecting the package size (in the number of doses) S_{LA} and the recommended daily dose D_{LA}. If a COPD patient switched from one medication category to another, such as from LABA to LAMA, we assume both medications are taken simultaneously until one is no longer purchased. With formula (2), we calculate the PDC per COPD patient by dividing the sum of all days with enough medication reserve by the total days of the observation period, i.e., by 720 days.

To estimate the causal effect of medication adherence on the probability of a hospitalized exacerbation, we rely on the causal Rubin effect. Accordingly, we quantify the average treatment effect (ATE) of the PDC on hospitalized exacerbations [22]. As the treatment variable for calculating the ATE must be finite, we divided COPD patients into five quintiles Q_q with $q \in \{1, 2, 3, 4, 5\}$ based on COPD patients' respective PDCs. Each COPD patient has a potential outcome *y* for the individual adherence status *q* denoted by the value of the dummy variable Q_q (i.e., $Q_q = 1$ means "COPD patient belongs to the group with adherence status *q*" and $Q_q = 0$ means "COPD patient does not belong to the group with adherence status *q*"). Formally, we define the ATE as

$$\beta_{ate,q} = E\left(y_q - y_1\right) \quad (3)$$

with E representing whether a COPD patient was hospitalized for exacerbation. The control group is the first quintile Q_1, which includes all COPD patients with a PDC of 0 to < 20%. Quintiles for q_2 through q_5 have a PDC of 20% to < 40%, 40% to < 60%, 60% to < 80%, and 80% to < 100%. Thus, the respective ATE is estimated on a total of four levels. The effect measure $\beta_{ate,q}$ indicates the average reduction in the probability of being hospitalized due to an exacerbation (for information on the satisfaction of the "conditional independence assumption" see Further Specification I).

Logistic regression is used due to the binary nature of the dependent variable (hospitalized exacerbation $E = 1$, no hospitalized exacerbation $E = 0$):

$$P\left(E = 1 \mid \beta_0 + \beta_q\, Q_q' + \gamma_x\, X' + \mu_w\, W'\right) = S\left(\beta_0 + \beta_q\, Q_q' + \gamma_x\, X' + \mu_w\, W'\right) \quad (4)$$

with

$$S\left(\beta_0 + \beta_q\, Q_q' + \gamma_x\, X' + \mu_w\, W'\right) = \frac{1}{1 + e^{-\left(\beta_0 + \beta_q\, Q_q' + \gamma_x\, X' + \mu_w\, W'\right)}} \quad (5)$$

where the influence of medication adherence is estimated by

$$\beta_q\, Q_q' = \beta_2\, Q_2 + \beta_3\, Q_3 + \beta_4\, Q_4 + \beta_5\, Q_5 \quad (6)$$

and the influence of socioeconomic factors is taken into account with

$$\gamma_x\, X' = \gamma_{1f}\, F_f + \gamma_2\, F^+ + \gamma_3\, F^- + \gamma_4\, R + \gamma_5\, A \quad (7)$$

and finally medical factors are considered with

$$\mu_w\, W' = \mu_1\, E_{t-2} + \mu_2 B + \mu_3\, M + \mu_4\, D + \mu_5\, K \quad (8).$$

Simplified, formula (4) expresses the probability P per COPD patient of being hospitalized with an exacerbation (E = 1), depending on the quintile of medication reserve Qq (formula (6)), socioeconomic factors X (formula (7)), and medical factors W (formula (8)). Due to the logistic nature of the regression analysis, formula (5) shows the exponential form of the regression equation. An overview of the notation of the variables is provided in Further Specification II.

To satisfy identifying assumption (1), i.e., the CIA, variables that have an effect on both the quintile assignment and the dependent variable are used as control variables that we grouped into two categories: Socioeconomic factors (age A, franchise level F_f, chosen franchise increase F^+ or decrease F^-, and premium reduction R) and medical factors (hospitalized exazerbations in the first two years of the observation period E_{t-2}, purchase of short-acting medication B, purchase of methylaxthine M, purchase of phosphodiesterase 4 inhibitors D, and purchase of mucolytic K). The deductible F_f serves as a proxy of risk preference relative to expected healthcare expenditures. Since it is possible to change the deductible rate at the beginning of each calendar year, we use the deductible F_f an insured person has at the very end of the observation period. To control for changing deductible rates between two subsequent years, the increase (F^+) and the decrease (F^-) of the deductible is included as a control variable. The deductible F_f and the changes in deductibles (F^+ and F^-) approximate the severity of COPD, as individuals who expect high health care costs to occur in the upcoming year may accordingly adjust their deductible. In addition, the premium reduction R is included to control for the socioeconomic status approximated by income. Only insured with low income are eligible to obtain a premium reduction.

Consequently, whether a COPD patient has ever been hospitalized for an exacerbation during the observation period is taken into account (E_{t-2}). This control variable approaches the COPD patients ability on dealing with the disease. Frequent hospitalized exacerbations may indicate poor disease management. Furthermore, short-acting medications (B) are used as a proxy for COPD severity. The GOLD standard defines two short-acting medication groups, namely beta$_2$ agonists (SABA) and antimuscarinic antagonists (SAMA), and their combination [2]. These medications may be taken in addition to the long-acting medications and indicate that this COPD patient has higher disease severity. Other agents used in the treatment of COPD include methylxanthines (M), phosphodiesterase-4 inhibitors (D), and mucolytics (K). These are used during acute worsening of lung function and hint at a non-hospitalized exacerbation.

Further Specification I: Conditional independence assumption

The most important difference between the identification of the ATE in an experimental setting, e.g., in a randomized controlled trial, versus the "Selection on Observables" applied in this study, is a weaker form of independence *assumption*, i.e., a *conditional independence assumption* (CIA). The CIA is violated when unobserved or empirically untested or untestable influencing factors are present.5F In such cases, it is assumed that most likely unexplained confounding factors affect the ATE.

According to these considerations, an empirical strategy suitable for the "Selection on Observables" method is used. Therefore, five identifying assumptions for each patient in this sample apply: 1) The assignment of a case to the treatment and control group according to CIA depends on the characteristics of the control variables. Hence, the treatment must actually depend on the values of the observed control variables measured for the case. This means that all necessary control variables are observed. 2) Each patient (i.e., each observation) could receive or not receive the treatment. This implies that there should be no patient subgroups (e.g., patients above 80 years) that occur in only one of the quintiles. By using the descriptive results, violations of this assumption can be excluded. 3) The treatment does not affect the control variables in a way that is related to the outcome. This means that e.g., that age potentially affects medication reserve, but medication reserve does not affect patient age. 4) A patient's treatment status only affects the individual outcome. This signalizes that a patient's own medication reserve does not affect another patient's likelihood of a hospitalized exacerbation. 5) There are only as many different potential outcomes as there are different treatments and one is observed for each subject. This indicates that the variation in medication reserve among multiple patients within a quintile should be small.

Further Specification II: List of model variables with corresponding abbreviations

Variable/ Category	Abbreviation	Description	Variable type
Probability function	P	The likelihood that a patient will be hospitalized due to an exacerbation	Percentage value between 0 and 100
Exponential function	S	Designation of the exponential function of a logistic regression	-
Dependent variable	E	Hospitalized exacerbation	Binary, 0 or 1
Independent variable of interest	Q_q	Quintile q of the proportion of days with medication reserve in all days in the observation period.	Binary, 0 or 1 (per quintile)
Socioeconomic control variables	A	Age	Steady
	F_f	Most recent deductible (with f equal to 350, 500, 1,000, 1,500, 2,000, or 2,500), i.e., at the end of the observation period	Binary, 0 or 1 (per franchise)
	F^+	Increase in deductible during the observation period	Binary, 0 or 1
	F^-	Reduction of the deductible during the observation period	Binary, 0 or 1
	R	Premium reduction	Binary, 0 or 1
Medical control variables	E_{t-2}	Hospitalized exacerbations in the two years of the observation period	Binary, 0 or 1
	B	Purchase of short-acting medication	Binary, 0 or 1
	M	Purchase of methylxanthine	Binary, 0 or 1
	D	Purchase of phosphodiesterase 4 inhibitors	Binary, 0 or 1
	K	Purchase of mucolytic	Binary, 0 or 1

3 Enhancing the usefulness of PROs

3.1 What is the association between gender and self-perceived health status when controlling for disease-specific conditions? A retrospective data analysis of pre- and post-operative EQ-5D-5L differences in total hip and knee arthroplasty

Anja Y. Bischof

Viktoria Steinbeck

David Kuklinski

Carlos J. Marques

Karina Bohlen

Karl C. Westphal

Frank Lampe

Alexander Geissler

Published in *BMC Musculoskeletal Disorders*

3.1.1 Abstract

Background: The self-perceived health status of patients undergoing total hip and knee arthroplasty (THA and TKA) might differ post-operatively due to gender, age, or comorbidities. Patient-reported outcome measures (PROMs) such as the EQ-5D-5L measure the self-perceived health status. This study investigates whether the index score of the EQ-5D-5L is a valid tool for interpreting gender differences in outcomes for patients undergoing THA and TKA.

Methods: Routine and PROM-data of elective primary THA or TKA patients in two German hospitals between 2016 and 2018 were analyzed. Univariate analysis with Pearson's chi-square was conducted to identify control variables for gender. To quantify the association between gender and the EQ-5D-5L dimensions, a cumulative odds ordinal logistic regression with proportional odds was conducted.

Results: 2368 THA patients (m=978; f=1390) and 1629 TKA patients (m=715; f=914) were considered. The regression analysis of the individual EQ-5D-5L dimensions showed that female gender was significantly associated with better self-care (THA and TKA) and better post-operative mobility (THA). In contrast, male gender was significantly associated with less pain/discomfort (TKA) and less anxiety/depression (THA) pre-surgery and 3-months post-surgery.

Conclusion: Our results confirmed that the self-perceived health status improved after surgery. However, due to the different associations of gender to the individual dimensions of the EQ-5D-5L, the weighted index score clouds the comparability between patients with different gender undergoing THA or TKA. Therefore, we argue to use the individual five dimensions for health status analysis, to reveal relevant additional information.

Keywords: Total hip arthroplasty, Total knee arthroplasty, Patient-reported outcome measures (PROMs), Gender differences, EQ-5D-5L

3.1.2 Background

Patient-reported outcome measures (PROMs) quantify the self-perceived health status of patients. Over the past years, they gained higher utilization in clinical care [1–3] and digital health interventions [4–6]. Two main PROM groups exist: 1) generic such as the EuroQol five-dimension five-level questionnaire (EQ-5D-5L), and 2) disease-specific such as the Western Ontario and McMaster Universities Osteoarthritis questionnaire (WOMAC), focusing on hip and knee joints [7]. Disease-specific instruments complement generic instruments, as they highlight disease-specific problems regarding a particular disease that a generic instrument cannot cover. However, both instruments allow for objectively evaluating patients' health status in a valid, reliable, and systematic manner [8–10].

Total hip and knee arthroplasty (THA and TKA) are standardized treatments. Female patients are up to three times more likely to undergo TKA at a more advanced stage [11–13]. Multiple studies reported that the preoperative [13–15] and sometimes also post-operative self-perceived health status [16] differ by gender. The reasons for these differences are assumed to be diverse. Literature suggests various reasons for female patients experiencing worse health status such as being referred to a surgeon with a higher degree of disability, higher risk of complications, and unwillingness to accept surgery [12, 14, 17]. The unwillingness to accept surgery might be developed due to a general increase in the perceived risk for morbidity and mortality after surgery or previous negative experiences in health systems due to gender biases [18–20]. Additionally, female patients tend to be older when undergoing arthroplasty, especially in case of THA [17, 21], and age is negatively associated with health status [22–25]. Regarding EQ-5D-5L scores, female gender, older age, and low sociodemographic status are related to worse scores in most dimensions in the general population [22, 26, 27]. Furthermore, differences in (total) scores by gender are also observed in the WOMAC, as female patients indicated more problems especially with pain. Therefore, the literature emphasizes evaluating and elaborating more gender-specific treatment strategies [28–30].

To enhance gender-specific treatment strategies, previous research highlighted the consideration of the individual dimensions of the EQ-5D-5L instead of the index score as the deconstruction provides insights into the contribution of the individual dimensions to the index score [26]. Additionally, the index score of the EQ-5D-5L is established based on national preferential weights, where cross-society preferences are reflected. The aim of selecting respondents for the national weighing is to achieve a quota-based gender ratio according to national statistics. The German value set was established on a participant dataset consisting of 53.4% female respondents [31]. Nevertheless, cross-society preference weights do not have to reflect specific preferences of individual sub-groups, such as patients undergoing THA or TKA. Therefore, analyzing dimensional scores becomes critical in stratifying and understanding differences in patients' health status composition.

While there is research on health status differences by gender [14, 32, 33], it has not been analyzed whether systematic gender-specific differences are observable in the dimensions of the EQ-5D-5L for THA and TKA patients pre- and post-surgery. A granular analysis of the differences per dimension helps to understand patients' gender-specific health status pre- and post-surgery, and their recovery pathway. Consequently, these findings can be used to improve outcomes from THA and TKA, by e.g., more accurately monitoring recovery pathways and adapting treatment protocols according to health status development. To the best of our knowledge, no study has yet assessed THA and TKA patients' values of EQ-5D-5L dimensions by gender over time. Hence, the research question is: "What are the differences in the association between gender and the EQ-5D-5L dimensions in patients undergoing total hip and knee arthroplasty pre- and post-surgery?".

3.1.3 Methods

Study design

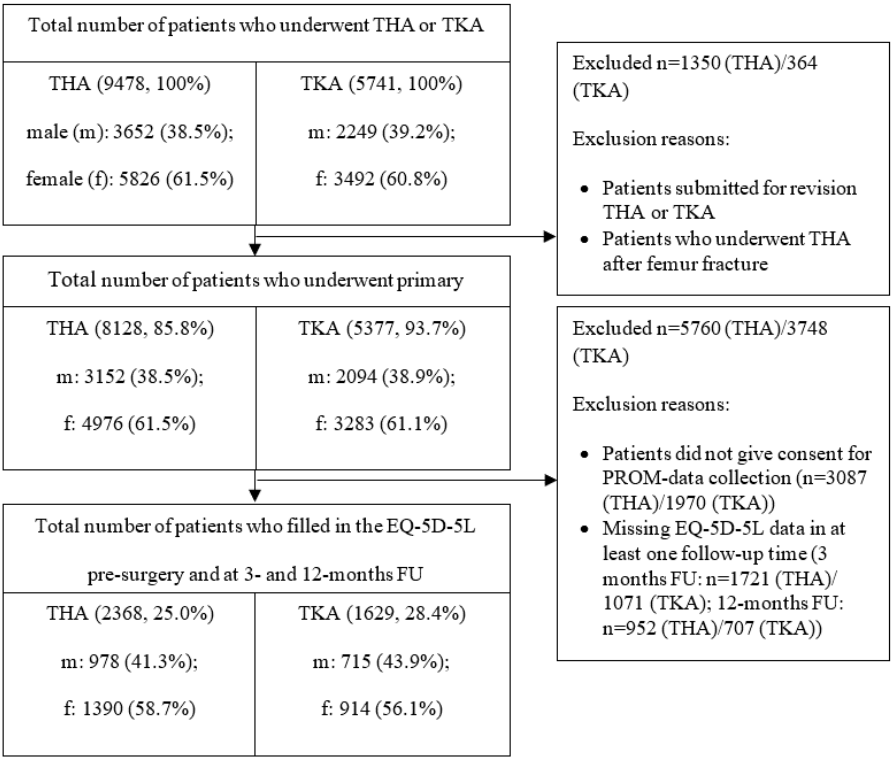
This study is a retrospective data analysis based on a prospectively collected dataset. The STROBE Statement guidelines for reporting observational studies were followed [34]. The Ethic-Commission of the Federal State of Hamburg, Germany, was

informed and approved the study (2021–300010-WF). Data was saved, processed, and analyzed anonymously.

Patients

PROM- and routine data of 2368 patients (male (m) = 978; female (f) = 1390) undergoing elective primary THA and 1629 patients (m = 715; f = 914) undergoing elective primary TKA in two private German hospitals belonging to the same hospital chain between January 2016 and December 2018 were analyzed. PROM-data is not routinely collected in German hospitals. Therefore, the patients were asked to give their written consent. Patients filled in the pre-surgery (at the day of admission), 3- and 12-months post-surgery EQ-5D-5L, and WOMAC questionnaires. The data of patients who underwent revisional THA or TKA, or THA after femur fracture and those who did not consent to PROM-data collection were excluded from the analysis (Fig. 3.1).

Figure 3.1: Patient flow diagram by follow-up time, procedure, gender and EQ-5D-5L response



Legend: FU: Follow-up; THA: Total Hip Arthroplasty; TKA: Total Knee Arthroplasty. The percentages after the total number of THA and TKA patients indicate the retention rate. The percentages after male and female gender indicate the distribution within the remaining sample.

Patient-reported outcome instruments

The EQ-5D-5L encloses the following five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension provides five answer levels: no, mild, moderate, severe, and extreme or unable to [35]. The EQ-5D-5L health index scores (also called values) ranges from 0 (death) to 1 (perfect health). It might even drop into a negative range, where certain health status is considered worse than death. EQ-5D-5L values are calculated based on national preference weights defined per dimension – called value sets [8, 36]. The answers on each

dimension are used to calculate the index score according to national preference weights. In the German value set, pain/discomfort and anxiety/depression are allocated with higher utility weights than the functional dimensions: mobility, self-care, and usual activities. Additionally, the dimension pain/discomfort has the highest impact on the health status [31]. The latest EQ-5D version with five levels was chosen due to its improved informational value, higher sensitivity, and accuracy in health status measurement [37, 38].

The WOMAC – a disease-specific instrument for the evaluation of hip and knee osteoarthritis [9, 10] – includes 24 items covering three dimensions – pain (5 items), stiffness (2 items), and physical function (17 items). Each dimension of the WOMAC consists of five levels: none, mild, moderate, severe, or extreme. Within a dimension, the scores of all items are added up. This results in pain taking values from 0 to 20, stiffness oscillates between 0 to 8, and physical function ranges from 0 to 68. The total WOMAC score is generated by summing up all values of the individual dimensions. The lower the total or individual-dimension WOMAC score, the less the patient suffers from pain, stiffness, or physical limitations [39, 40].

Data analysis

We used the PROMs (EQ-5D-5L, WOMAC) and routine hospital data on patient characteristics (such as sociodemographic data, medical conditions, and surgical details) from THA and TKA patients operated at two German hospitals. PROMs were measured pre-surgery, 3-months, and 12-months post-surgery. The dimensions of the EQ-5D-5L serve as main area of investigation.

Differences regarding the surgical procedures and gender were described by independent-samples t-test or Mann–Whitney U depending on the variable's characteristics or distribution. The health state indices of the EQ-5D-5L were calculated based on the German value set [31].

To quantify the association between gender and the EQ-5D-5L dimensions, a cumulative odds ordinal logistic regression with proportional odds was conducted [25] (for the formula and additional information see Online Resource 1). Age was

included as control variable in all calculations. The relationship between age (in years) and health status has been extensively discussed [22, 26, 41, 42]. Other control variables such as the Elixhauser Comorbidity Index, walking distance, and walking aid pre-surgery, or surgery duration were included in the calculations based on the correlation significance of Pearson's chi-square between gender and the control variables (for results see Online Resources 2 and 3).

At each measuring point (pre-, 3-, and 12-months post-surgery) the corresponding WOMAC total score was included as a control variable for each patient. Literature provides evidence that the WOMAC subscale scores – especially pain and functional disability – are associated with changes in the EQ-5D-5L values, i.e., disease-specific PROMs are interrelated with generic health states [43]. However, since the WOMAC total score is built of the sum of its subscale scores – and no weighing of individual subscales applies – the total score was considered in the ordinal logistic regression.

All analyses and calculations were performed separately for patients who underwent THA and TKA to quantify the association between gender and health status more precisely. All analyses were conducted with IBM-SPSS version 28 [44].

3.1.4 Results

Descriptive statistics

The average age of male THA patients was 68.72 years (f = 68.9 years), with a mean length of hospital stay of 7.34 days (f = 7.46 days) and a mean surgery time of 46.47 min (f = 45.76 min). The specific and general in-hospital complication ratios were around 2% or lower for both gender. Regarding the WOMAC total scores, male patients started with lower values, i.e., better scores. The scores of the male patients remained lower (better) 12-months post-surgery compared to female patients. However, both genders experienced an improvement in total WOMAC scores over time, while significant score differences between the gender existed preoperatively ($p < 0.001$) and at 3-months post-surgery ($p = 0.019$) (Table 3.1).

Table 3.1: Descriptive statistics

Variable	THA		TKA	
	Male patients (n=978)	Female patients (n=1390)	Male patients (n=715)	Female patients (n=915)
Age (years)	68.72 (9.05)	68.86 (9.18)	71.00** (7.90)	70.00** (8.70)
LOS (days)	7.34 (4.16)	7.43 (4.01)	7.44 (3.55)	7.37 (3.42)
Surgery duration (min.)	46.47 (18.06)	45.76 (17.05)	57.83*** (17.67)	55.26*** (18.72)
In-hospital complication ratio (general)	0.03 (0.16)	0.02 (0.14)	0.03 (0.16)	0.02 (0.12)
In-hospital complication ratio (specific)	0.01 (0.10)	0.01 (0.11)	0.02** (0.12)	0.04** (0.07)
EQ-5D-5L (pre-surgery)	0.576*** (0.275)	0.512*** (0.291)	0.625*** (0.253)	0.534*** (0.284)
EQ-5D-5L (3-month FU)	0.883** (0.156)	0.867** (0.172)	0.844*** (0.163)	0.817*** (0.182)
EQ-5D-5L (12-month FU)	0.910 (0.149)	0.903 (0.160)	0.885*** (0.154)	0.862*** (0.169)
WOMAC (pre-surgery)	45.66*** (15.52)	50.15*** (14.61)	41.77*** (15.28)	46.44*** (14.39)
WOMAC (3-month FU)	13.05 (12.19)	14.36 (12.59)	17.87*** (13.45)	20.00*** (14.28)
WOMAC (12-month FU)	11.02 (12.98)	11.98 (13.89)	14.43*** (13.64)	17.54*** (15.13)
Elixhauser Comorbidity Score	3.80** (4.96)	3.31** (4.55)	3.46 (4.77)	3.26 (4.34)

Statistically significant differences between gender are indicated by $p < 0.05$** and $p < 0.01$***

LOS Length of stay, SD Standard Deviation

The following in-hospital complications are categorized as specific: Misplacement of the implant; secondary dislocation of the implant; closed or open reposition of the implant; lesion of vessels; lesion of nerve roots; periprosthetic fractures and wound infections. General in-hospital complications are defined as: Pneumonia; urinary tract infection; sepsis; pulmonary embolism; venous thromboembolism; cardiac complications; etc. For reasons of simplicity and clarity, the most informative variables are shown. A full table of all variables used in the further analysis is attached to Online Resource 4.

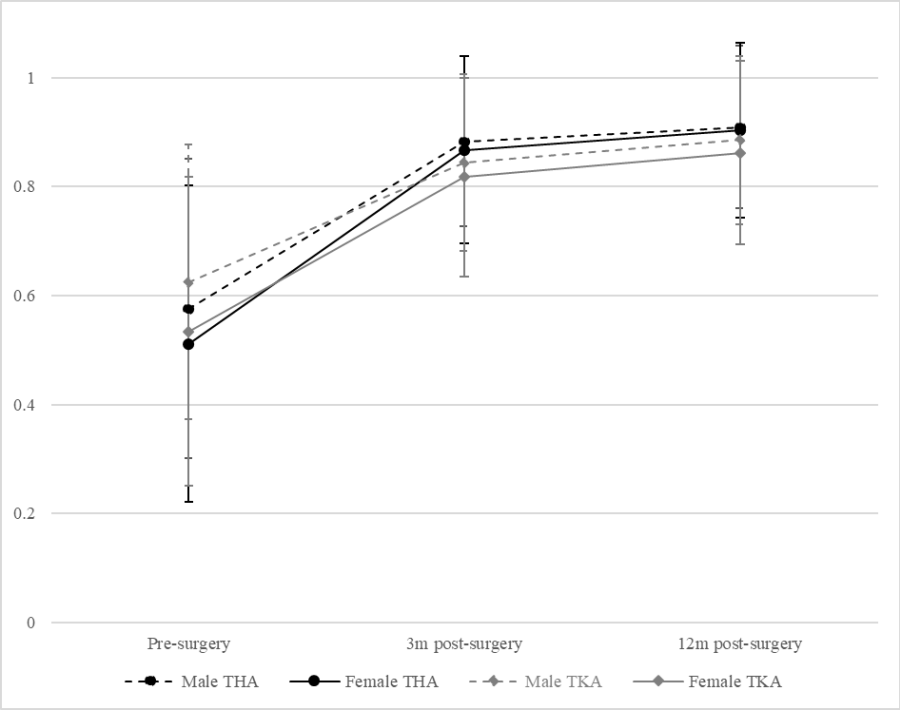
On average, male TKA patients were 71.0 years old, whereas female patients were one year younger at surgery ($p = 0.015$). After TKA, male patients stayed on average 7.44 days (f = 7.37 days) in the hospital, and the mean surgery time was 57.83 min (f = 55.26 min, $p = 0.005$). When referring to the in-hospital complication ratio, a significant difference between gender was observable only for specific complications. The WOMAC scores for male patients were lower pre- and post-surgery compared to female patients. However, both gender showed WOMAC score improvements over time, while the scores between the gender remained significantly different over all measuring points (Table 3.1).

Gender differences in generic health index score (EQ-5D-5L) over time

Male patients in the THA group perceived a mean health index score improvement of 0.334 between pre-operative and 12-months post-surgery, while female patients experienced an improvement of 0.391. Hence, female patients perceived a larger health status improvement within the first 12 months after surgery ($p < 0.001$).

In the TKA group, male patients had on average a preoperative health status index of 0.625 which improved to 0.885, while female patients showed an improvement from 0.534 to 0.862. Thus, female patients started with lower values, and despite higher improvement rates, they did not experience the same post-operative health status enhancement as male patients. However, the improvement rates between gender tend to be significantly different ($p < 0.001$) (Fig. 3.2).

Figure 3.2: Mean EQ-5D-5L total score and corresponding standard deviations for the patients in the total hip and total knee arthroplasty groups by gender over time



Legend: THA: Total Hip Arthroplasty; TKA: Total Knee Arthroplasty

When comparing the two types of arthroplasties, TKA patients started with higher EQ-5D-5L total scores (m = 0.625, f = 0.534) compared to THA patients (m = 0.576, f = 0.511). However, THA patients showed higher mean health status values 3-months post-surgery (THA: m = 0.883, f = 0.867; TKA: m = 0.844, f = 0.817). At 12-months post-surgery, all patient groups recorded a health status improvement over time. Nevertheless, THA patients showed a larger overall health status improvement (THA: m = + 0.334, f = + 0.391; TKA: m = + 0.260, f = + 0.328).

Comparison of EQ-5D-5L dimension scores by gender

In the THA group, significant differences regarding gender were observed in the dimensions of self-care and anxiety/depression at all measuring points. Additionally,

significant differences were detected in usual activity and pain/discomfort pre-surgery. However, this effect was significant only for pain/discomfort 3-months post-surgery. In the TKA group, the dimensions of pain/discomfort and anxiety/depression differed significantly per gender pre- and post-surgery. Furthermore, mobility and usual activity differed significantly concerning gender, but this effect was no longer present post-surgery (Table 3.2). For both THA and TKA, the dimension with the lowest scores was pain/discomfort pre-and post-surgery. Furthermore, both patient groups indicated the least problems in the dimension self-care pre- and post-surgery, while in THA usual activity reached the same level as self-care 12-months post-surgery.

Table 3.2: Comparison of the EQ-5D-5L dimensions by gender

	THA		TKA	
	Male Mean (SD)	**Female Mean (SD)**	**Male Mean (SD)**	**Female Mean (SD)**
Mobility				
pre-surgery	1.81 (0.73)	1.86 (0.72)	1.75*** (0.72)	1.82*** (0.67)
FU3	1.32 (0.56)	1.30 (0.57)	1.34 (0.57)	1.34 (0.57)
FU12	1.37 (0.71)	1.31 (0.64)	1.37 (0.63)	1.42 (0.72)
Self-care				
pre-surgery	1.27*** (0.55)	1.21*** (0.50)	1.15 (0.42)	1.13 (0.42)
FU3	1.08** (0.31)	1.06** (0.28)	1.08** (0.28)	1.06** (0.30)
FU12	1.11*** (0.39)	1.07*** (0.33)	1.13 (0.44)	1.11 (0.41)
Usual activity				
pre-surgery	1.66*** (0.73)	1.73*** (0.73)	1.55*** (0.69)	1.66*** (0.68)
FU3	1.25 (0.51)	1.27 (0.51)	1.32 (0.55)	1.34 (0.56)
FU12	1.11*** (0.39)	1.07*** (0.33)	1.32*** (0.60)	1.42*** (0.69)
Pain/discomfort				
pre-surgery	2.34*** (0.65)	2.47*** (0.66)	2.24*** (0.64)	2.41*** (0.66)
FU3	1.44*** (0.57)	1.51*** (0.61)	1.63*** (0.59)	1.74*** (0.60)
FU12	1.43 (0.65)	1.46 (0.67)	1.55*** (0.67)	1.73*** (0.77)
Anxiety depression				
pre-surgery	1.28*** (0.55)	1.47*** (0.65)	1.18*** (0.42)	1.46*** (0.67)
FU3	1.08*** (0.30)	1.14*** (0.39)	1.11*** (0.37)	1.17*** (0.42)
FU12	1.14*** (0.45)	1.18*** (0.50)	1.13*** (0.43)	1.19*** (0.50)

Statistically significant difference between gender at a **95% and ***99% significance level

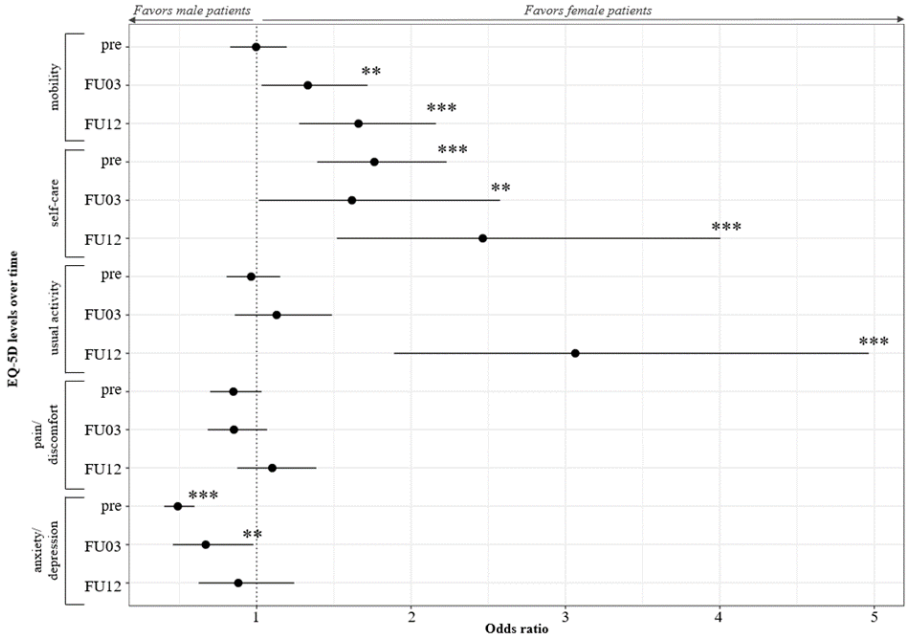
SD Standard Deviation, FU Follow-up

Association of gender with the EQ-5D-5L dimensions

When controlling for confounders such as age, pre-surgical conditions, and disease-specific health status in the ordinal logistic regression, the association of gender with the individual EQ-5D-5L dimensions was investigated. In the THA group, female patients showed significantly higher odds for better (lower) mobility levels post-surgery, although pre-surgery the odds ratio was slightly, but non-significantly, in favor of male gender (OR = 0.996). In the dimension of self-care, female gender was significantly associated with better self-care levels at baseline and 3- and 12-month follow-ups. In the dimension of usual activities, female patients achieved better scores post-surgery. However, the association is significant only 12-months post-surgery. In contrast, female gender was significantly associated with worse values (decreased odds) in the dimension anxiety/depression pre- and 3-month post-surgery. Nevertheless, this association was no longer significant 12-months post-surgery. Furthermore, in pain/discomfort, there were no significant associations with either gender, while the odds ratio favored male gender in the first two measuring points, 12-months post-surgery female gender was associated with better scores (Fig. 3.3).

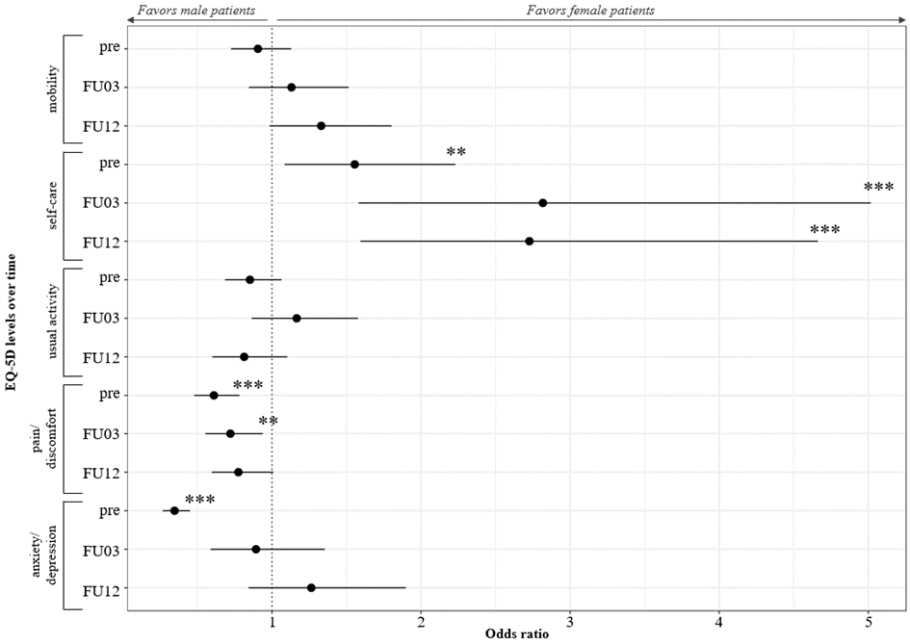
The pre-surgical mobility odds ratio is slightly better for male gender (OR = 0.906), whereas female patients tend to display better scores post-surgery. However, none of the associations for mobility are significant. Female gender is significantly associated with better scores in self-care pre- and post-surgery. No clear association to one of the gender is observable in usual activity. In contrast, being a male patient is significantly associated with better scores in pain/discomfort pre- and 3-month post-surgery. Additionally, being of female gender was significantly associated with worse scores in the dimension anxiety/depression preoperatively. However, this effect is no longer observable post-surgery (Fig. 3.4).

Figure 3.3: Association of gender with EQ-5D-5L dimensions of total hip arthroplasty over time



Legend: Values are odds ratio and their 95% CI for female gender; significant Odds ratio at a **95% and ***99% significance level, FU = Follow-up; Notes: Coefficients of the control variables related to age, Elixhauser Comorbidity Index, WOMAC, pain, the modified Kellgren-Lawrence Classification, severity of joint destruction in rheumatic diseases, pre-operative conditions, walking distance and aid at admission were included in the pre-surgical dimensions, while early mobilization, clinic type, length of stay, and surgery duration were added to the post-operative regressions, too. Additionally, pre-surgical EQ-5D-5L dimensions were included in the post-operative regressions to avoid omitting the effect of pre-surgical conditions on post-surgical outcomes. The coefficients of the control variables are attached to Online Resource 5.

Figure 3.4: Association of gender with EQ-5D-5L dimensions of total knee arthroplasty over time



Legend: Values are odds ratio and their 95% CI for female gender; significant Odds ratio at a **95% and ***99% significance level, FU = Follow-up; Notes: Confidence intervals are set to 95%. For TKA, coefficients of the control variables related to age, Elixhauser Comorbidity Index, WOMAC, pain, severity of joint destruction in rheumatic diseases, pre-operative findings, patient clinical complexity level (PCCL, a measure for the degree comorbidities), walking distance and aid at admission were included in the pre-surgical dimensions, while post-operative condition, surgery duration, walking distance and aid at discharge were added to the post-operative regressions, too. Additionally, pre-surgical EQ-5D-5L dimensions were included in the post-operative regressions to avoid omitting the effect of pre-surgical conditions on post-surgical outcomes. The coefficients of the control variables are attached to Online Resource 6.

3.1.5 Discussion

This paper comprehensively investigated the relationship between gender and the dimensions of the EQ-5D-5L. Our results confirmed previous studies on the potential of THA and TKA in improving quality of life since self-perceived health status improved significantly after surgery [17, 21, 42, 45]. From a gender perspective, female patients experienced higher health status improvements in total scores than

male patients, however, remained at a lower level in the total scores post-surgery. This was measured with both PROM instruments (i.e., the EQ-5D-5L and WOMAC). The results are in line with previous studies highlighting this imbalance, especially pre-surgery [13, 15, 26]. The pre-surgical health condition is the strongest predictor of health gains after surgery [42, 46].

Diving into the dimensions of the EQ-5D-5L, male patients rated most of the dimensions with lower, i.e., more favorable, levels. Male gender was significantly related to better (lower) anxiety/depression (THA) and pain/discomfort (TKA) levels pre- and 3-month post-surgery. This effect diminished 12-month post-surgery. Interestingly, Casetta et al. [47] found that the mental health state of female patients remained better while facing chronic conditions such as multiple sclerosis, although physical functioning decreased. Contrarily, our conclusions lead to the opposite assumption of male patients showing better mental health (anxiety/depression) pre- and post- arthroplasty. Potential explanations could be the underreporting of mental health concerns among male patients or female patients' increased pain sensitivity in combination with a certain degree of severity of the chronic disease leading to more mental health constraints [15, 48]. This argument is also supported by the significant association of lower pain/discomfort scores with male gender in our study.

The different associations of gender with the individual dimensions of the EQ-5D-5L indicate that the outcomes of THA and TKA need to be considered and interpreted separately. Self-care seems to be the dimension that is significantly positively associated with female gender at all measuring points, independent of either undergoing THA or TKA. However, male patients show significant positive associations to dimensions with higher weights in the EQ-5D-5L – anxiety/depression (THA) and pain/discomfort (TKA).

In many national EQ-5D-5L value sets, like in the German value set, more weight is set on the dimension anxiety/depression and pain/discomfort compared to self-care [31]. The weighted preferences determine how EQ-5D-5L value sets are calculated and used in decision-making processes [8]. Therefore, this aspect leads to the discussion of whether the different perceptions of well-being of female THA or TKA

patients are sufficiently weighted when using the EQ-5D-5L as index score or whether an analysis per dimension creates a more accurate picture of gender-specific health status.

In the investigated patient population, female TKA patients were younger than their male counterparts. Former studies differed in their statements of whether male or female TKA patients were older at the point of surgery, but at times referred to the age as explanatory variable for gender difference [14, 18, 49]. This relationship could not be confirmed in our analysis. Contrarily, no significant difference in frequency of in-hospital complications between male and female patients was measurable, which confirms previous findings [13, 50].

Our results show the need to put more emphasis on the individual dimensions of PROMs when interpreting the outcomes concerning gender differences, especially regarding the rising demand in clinical practice and digital health interventions. The relationship of gender with mobility (THA), self-care (THA and TKA), anxiety/depression (THA), and pain/discomfort (TKA) were significant when controlling for other relevant variables. Although female patients tended to show better scores in self-care over all measuring points, they only achieved lower index scores due to the weighing of the individual dimensions of the EQ-5D-5L in the German value set. This might falsify the meaning of the index score that female patients were worse off pre- and post-surgery, as they tend to indicate better scores in dimensions that are associated with lower utility weights.

Our findings suggest that using a single index score for evaluating patients' health status without gender consideration is questionable. An alternative strategy to enhance the comparability of results would be the implementation of gender-specific index scores or a greater reliance on the distinct dimensions of the EQ-5D-5L, some of which exhibit gender-related associations. This approach may additionally facilitate the operational utility of PROMs, such as in monitoring treatment regimens or delineating recovery trajectories, thus enabling informed decision-making about treatment adaptations. To promote the broader integration of PROMs within clinical

practice, it is essential to establish comprehensive guidelines for accurately assessing PROMs and the subsequent derivations that can be drawn from them.

Limitations

As stated above, this is a retrospective data analysis. We used routine- and PROM-data that was collected for quality control purposes. We identified three limiting factors of this research: First, missing profile items in the EQ-5D-5L led to the exclusion of the corresponding patients, as substituting an average derived from other non-missing items for missing profile items is considered bad practice, and the individual dimensions serve as dependent variables in the conducted regressions. Second, we included all patients in our sample independent of comorbidities due to the dataset's lack of information. The number and degree of comorbidities might influence the patient's perception of the health status and, thus, change the EQ-5D-5L index score. Third, there are also limitations to the generalization of results from this dataset since the data originates from two health care providers belonging to the same provider network. Data collection on information such as previous treatment approaches or current living situation are not part of the routine data collection in Germany. These additional variables would have allowed for more precise interpretation of the relationship between gender and post-surgery outcomes.

Future research should apply the presented methodological approach to comparable datasets from other countries. If similar findings result, it should be investigated whether gender-specific adjustments of the EQ-5D-5L are necessary to ensure comparability of results between gender in people undergoing THA or TKA. In addition, including a longer than 12-months post-surgery follow-up period might further improve the understanding of the relationship between gender and the outcome parameters in the long-term.

3.1.6 Conclusion

Gender was significantly related to individual dimensions of the EQ-5D-5L in the THA and TKA groups, pre- and post-surgery when controlling for age, comorbidity, disease-specific health status, and pre-surgical conditions. Male THA and TKA

patients perceived significantly better levels than female patients for most dimensions of the EQ-5D-5L. Male patients in the THA group showed significant associations with lower anxiety/depression levels pre- and 3-months post-surgery, while the same holds for pain/discomfort in TKA. Female gender was significantly associated with better mobility post-surgery for THA and better self-care independent of the measuring point for both treatments. Therefore, the awareness of gender-specific differences in health status should inform healthcare professionals and policymakers in integrating these insights into healthcare to ensure appropriate treatment measures based on patients' needs. When evaluating patients' health status, considering the individual dimensions reveals more detailed, differentiated, and actionable insights than the index score.

References

1. Porter I, Gonçalves-Bradley D, Ricci-Cabello I, et al (2016) Framework and guidance for implementing patient-reported outcomes in clinical practice: evidence, challenges and opportunities. J Comp Eff Res 5:507–519.

2. Porter I, Davey A, Gangannagaripalli J, et al (2021) Integrating Patient Reported Outcome Measures (PROMs) into routine nurse-led primary care for patients with multimorbidity: a feasibility and acceptability study. Health Qual Life Outcomes 19:133.

3. Gibbons C, Porter I, Gonçalves-Bradley DC, et al (2021) Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Cochrane Database of Systematic Reviews 2021:.

4. Hamilton DF, Giesinger JM, Giesinger K (2020) Technological developments enable measuring and using patient-reported outcomes data in orthopaedic clinical practice. World J Orthop 11:584–594.

5. Knapp A, Harst L, Hager S, et al (2021) Use of Patient-Reported Outcome Measures and Patient-Reported Experience Measures Within Evaluation Studies of Telemedicine Applications: Systematic Review. J Med Internet Res 23:e30042.

6. Bartlett VL, Ross JS, Shah ND, et al (2021) Physical activity, patient-reported symptoms, and clinical events: Insights into postprocedural recovery from personal digital devices. Cardiovasc Digit Health J 2:212–221.

7. Räsänen P, Roine E, Sintonen H, et al (2006) Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. Int J Technol Assess Health Care 22:235–241.

8. Devlin NJ, Parkin D, Janssen B (2020) Methods for analysing and reporting EQ-5D data. Springer International Publishing

9. Dieppe PA (1995) Recommended methodology for assessing the progression of osteoarthritis of the hip and knee joints. Osteoarthritis Cartilage 3:73–77.

10. Bellamy N, Buchanan WW, Goldsmith CH, et al (1988) Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 15:1833–40.

11. Srikanth VK, Fryer JL, Zhai G, et al (2005) A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthritis Cartilage 13:769–781.

12. Hawker GA, Wright JG, Coyte PC, et al (2000) Differences between men and women in the rate of use of hip and knee arthroplasty. N Engl J Med 342:1016–1022

13. Katz JN, Wright EA, Guadagnoli E, et al (1994) Differences between men and women undergoing major orthopedic surgery for degenerative arthritis. Arthritis Rheum 37:687–694.

14. Lim JBT, Chi CH, Lo LE, et al (2015) Gender difference in outcome after total knee replacement. Journal of Orthopaedic Surgery 23:194–197.

15. Impellizzeri FM, Mannion AF, Naal FD, Leunig M (2015) Validation of the core outcome measures index in patients with femoroacetabular impingement. Arthroscopy - Journal of Arthroscopic and Related Surgery 31:1238–1246.

16. Turnbull GS, Scott CEH, MacDonald DJ, Breusch SJ (2019) Gender and preoperative function predict physical activity levels after revision total knee arthroplasty. J Arthroplasty 34:939–946.

17. Holtzman J, Saleh K, Kane R (2002) Effect of baseline functional status and pain on outcomes of total hip arthroplasty. The Journal of Bone and Joint Surgery-American 84:1942–1948.

18. Patel AP, Gronbeck C, Chambers M, et al (2020) Gender and total joint arthroplasty: Variable outcomes by procedure type. Arthroplast Today 6:517–520.

19. Samulowitz A, Gremyr I, Eriksson E, Hensing G (2018) “Brave Men” and “Emotional Women”: A theory-guided literature review on gender bias in health care and gendered norms towards patients with chronic pain. Pain Res Manag 2018:1–14.

20. Dixon LK, Dimagli A, Di Tommaso E, et al (2022) Females have an increased risk of short-term mortality after cardiac surgery compared to males: Insights from a national database. J Card Surg 37:3507–3519.

21. Kostamo T, Bourne RB, Whittaker JP, et al (2009) No difference in gender-specific hip replacement outcomes. Clin Orthop Relat Res 467:135–140.

22. König HH, Bernert S, Angermeyer MC, et al (2009) Comparison of population health status in six european countries results of a representative survey using the EQ-5D questionnaire. Med Care 47:255–261.

23. Hofstede SN, Gademan MGJ, Stijnen T, et al (2018) The influence of preoperative determinants on quality of life, functioning and pain after total knee and hip replacement: A pooled analysis of Dutch cohorts. BMC Musculoskelet Disord 19:68.

24. Fang M, Noiseux N, Linson E, Cram P (2015) The effect of advancing age on total joint replacement outcomes. Geriatr Orthop Surg Rehabil 6:173–179.

25. Götz JS, Benditz A, Reinhard J, et al (2021) Influence of anxiety/depression, age, gender and ASA on 1-year follow-up outcomes following total hip and knee arthroplasty in 5447 patients. J Clin Med 10:3095.

26. Szende A, Janssen B (2014) Socio-demographic indicators based on EQ-5D. In: Szende A, Janssen B, Cabases J (eds) Self-reported population health: An international perspective based on EQ-5D. Springer Netherlands, Dordrecht, pp 37–46

27. König H-H, Brettschneider C, Lühmann D, et al (2020) EQ-5D-3L health status and health state utilities of the oldest-old (85 +) in Germany: results from the AgeCoDe-AgeQualiDe study. Quality of Life Research 29:3223–3232.

28. Cho HJ, Chang CB, Yoo JH, et al (2010) Gender differences in the correlation between symptom and radiographic severity in patients with knee osteoarthritis. Clin Orthop Relat Res 468:1749–1758.

29. Debi R, Mor A, Segal O, et al (2009) Differences in gait patterns, pain, function and quality of life between males and females with knee osteoarthritis: A clinical trial. BMC Musculoskelet Disord 10:.

30. Elboim-Gabyzon M, Rozen N, Laufer Y (2012) Gender differences in pain perception and functional ability in subjects with knee osteoarthritis. ISRN Orthop 2012:1–4.

31. Ludwig K, Graf von der Schulenburg J-M, Greiner W (2018) German Value Set for the EQ-5D-5L. Pharmacoeconomics 36:663–674.

32. Nehme Z, Andrew E, Bernard S, Smith K (2019) Sex differences in the quality-of-life and functional outcome of cardiac arrest survivors. Resuscitation 137:21–28.

33. Jansson KÅ, Granath F (2011) Health-related quality of life (EQ-5D) before and after orthopedic surgery. Acta Orthop 82:82–89.

34. von Elm E, Altman DG, Egger M, et al (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet 370:1453–1457.

35. Herdman M, Gudex C, Lloyd A, et al (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of Life Research 20:1727–1736.

36. Parkin D, Rice N, Devlin N (2010) Statistical analysis of EQ-5D profiles: Does the use of value sets bias inference? Medical Decision Making 30:556–565.

37. Thompson AJ, Turner AJ (2020) A Comparison of the EQ-5D-3L and EQ-5D-5L. Pharmacoeconomics 38:575–591.

38. Janssen MF, Bonsel GJ, Luo N (2018) Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive Systems and Value Sets from Seven Countries. Pharmacoeconomics 36:675–697.

39. Impellizzeri FM, Mannion AF, Leunig M, et al (2011) Comparison of the reliability, responsiveness, and construct validity of 4 different questionnaires for evaluating outcomes after total knee arthroplasty. J Arthroplasty 26:861–869.

40. Terwee CB, Roorda LD, Knol DL, et al (2009) Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol 62:1062–1067.

41. O'Connor MI (2006) Osteoarthritis of the hip and knee: Sex and gender differences. Orthopedic Clinics of North America 37:559–568.

42. Konopka JF, Lee Y, Su EP, McLawhorn AS (2018) Quality-adjusted life years after hip and knee arthroplasty. JBJS Open Access 3:e0007.

43. Wailoo A, Hernandez Alava M, Escobar Martinez A (2014) Modelling the relationship between the WOMAC osteoarthritis index and EQ-5D. Health Qual Life Outcomes 12:37.

44. SPSS Inc. (2022) Statistical Package for Social Science for Windows, Version 28.0

45. Pichard-Encina CP, Hungerford MW, Khanuja HS, et al (2014) Total hip arthroplasty: Differences in outcome measures between men and women. J Long Term Eff Med Implants 24:213–218.

46. Kuklinski D, Marques CJ, Bohlen K, et al (2022) Thresholds for meaningful improvement in WOMAC scores need to be adjusted to patient characteristics after hip and knee replacement. J Orthop 29:50–59.

47. Casetta I, Riise T, Wamme Nortvedt M, et al (2009) Gender differences in health-related quality of life in multiple sclerosis. Multiple Sclerosis 15:1339–1346.

48. Parsley BS, Bertolusso R, Harrington M, et al (2010) Influence of gender on age of treatment with TKA and functional outcome. Clin Orthop Relat Res 468:1759–64.

49. Dalury DF, Mason JB, Murphy JA, Adams MJ (2009) Analysis of the outcome in male and female patients using a unisex total knee replacement system. Journal of Bone and Joint Surgery 91:357–360.

48. Bartley EJ, Fillingim RB (2013) Sex differences in pain: a brief review of clinical and experimental findings. Br J Anaesth 111:52–8.

49. Parsley BS, Bertolusso R, Harrington M, et al (2010) Influence of gender on age of treatment with TKA and functional outcome. Clin Orthop Relat Res 468:1759–64.

50. Dalury DF, Mason JB, Murphy JA, Adams MJ (2009) Analysis of the outcome in male and female patients using a unisex total knee replacement system. Journal of Bone and Joint Surgery 91:357–360.

Appendix

Online Resource 1: Formula on applied cumulative odds ordinal logistic regression with proportional odds

The following formula was used:

$$logit\ [\pi(EQdimension_{time} \leq j \mid X)] = \ln\left(\frac{\pi\ (EQdimension_{time} \leq j \mid X)}{\pi\ (EQdimension_{time} \geq j \mid X)}\right) = \alpha_j + (-\beta X),$$

where α_j's are the thresholds, β are logit coefficients, and X represents the independent variables. When interpreting the binary variable sex (m/f) and its relation to the ordinal explanatory variables, female sex functions as the reference category. The interpretational logic holds that an odds ratio lower than 1 indicates that being female is associated with worse values in that dimension. However, if the expected β is larger than 1, being of female sex is associated with increased odds for better values in that dimension [1,2]. The closer the value gets to 1, the smaller the difference between the sex-specific dimensional scores.

EQdimension stands for the individual dimensions of the EQ-5D-5L instrument (*mobility, self-care, usual activity, pain/discomfort, and anxiety/depression*), which are time-dependent (*pre-surgery*, *3-months post-surgery* and *12-months post-surgery)*. Furthermore, j = 1,2,3,4,5 represent the response levels of the EQ-5D-5L – where 1 indicates "no problems" and 5 represents "extreme problems" in the corresponding dimensions. We used *SPSS PLUM (Polytomous Universal Model)*, an extension of the generalized linear model for ordinal response data [1].

References

1. Liu X (2009) Ordinal regression analysis: Fitting the proportional odds model using Stata, SAS and SPSS. Journal of Modern Applied Statistical Methods 8:632–42.

2. Janssen B, Szende A, Ramos-Goñi JM (2014) Data and methods. In: Szende A, Janssen B, Cabasés J, editors. Self-Reported Population Health:

An International Perspective based on EQ-5D. Dordrecht: Springer Netherlands; p. 7–17.

Online Resource 2: Correlations of potential control variables with sex for THA

THA		
Variable	**Pearson's chi-square**	**Cramer-V**
Pain (no, stress pain, or rest pain)	24.847***	0.102
Modified Kellgren-Lawrence Classification		
Osteophytes	2.078	
Joint space	10.774**	0.067
Sclerosis	8.529**	0.060
Deformation	7.541**	0.056
Severity of joint destruction in rheumatic diseases		
Rheumatic disease with manifestation at the affected joint	11.093***	0.068
Erosive joint destruction (according to Larsen-Dale-Eek)	6.008	
Pre-operative findings		
ASA classification	24.401***	0.102
Post-operative specific complications	0.485	
Post-operative generic complications	0.946	
Walking distance at admission	9.492**	0.063
Walking aid at admission	13.372***	0.075
Walking distance at discharge	6.156	
Walking aid at discharge	5.199	
Patient Clincial Complexity Level	4.605	
Early mobilization	4.807**	0.047
Clinic type (general vs. specialized)	4.608**	0.044

Variable	Chi-Square	Eta
WOMAC (pre-surgery)	422.109	
WOMAC (3-month FU)	290.700	
WOMAC (12-month FU)	232.479	
Elixhauser Comorbidity Score	67.111***	0.051
Length of stay	76.660***	0.018
Surgery duration	124.728***	0.020
Age	49.334	

Statistically significant difference between sex at a **95% and ***99% significance level; FU=Follow-up.
Notes: Pearson's chi-square indicates whether there is a significant correlation to a variable with sex, and Cramer-V or Eta shows the effect size. If there is no significant correlation for a variable, the effect size is not indicated.

Online Resource 3: Correlations of potential control variables with sex for TKA

TKA		
Variable	**Pearson's chi-square**	**Cramer-V**
Pain (no, stress pain, or rest pain)	28.808***	0.133
Modified Kellgren-Lawrence Classification		
Osteophytes	0.268	
Joint space	2.426	
Sclerosis	1.586	
Deformation	1.155	
Severity of joint destruction in rheumatic diseases		
Rheumatic disease with manifestation at the affected joint	7.637***	0.068
Erosive joint destruction (according to Larsen-Dale-Eek)	1.732	
Pre-operative findings		
Gonarthrosis	6.761**	0.064
Malposition of the knee	96.257***	0.243

ASA classification	32.674***	0.142
Post-operative specific complications	5.329**	0.057
Post-operative generic complications	2.024	
Walking distance at admission	8.724	
Walking aid at admission	11.183**	0.083
Walking distance at discharge	8.537**	0.072
Walking aid at discharge	10.043***	0.079
Patient Clincial Complexity Level	14.029***	0.093
Early mobilization	0.271	
Clinic type (general vs. specialized)	0.633	
Variable	**Chi-Square**	**Eta**
WOMAC (pre-surgery)	343.458	
WOMAC (3-month FU)	239.62	
WOMAC (12-month FU)	216.115	
Elixhauser Comorbidity Score	69.557***	0.022
Length of stay	36.295	
Surgery duration	139.742***	0.07
Age	65.554	

Statistically significant difference between sex at a **95% and ***99% significance level; FU=Follow-up.
Notes: Pearson's chi-square indicates whether there is a significant correlation to a variable with sex, and Cramer-V or Eta shows the effect size. If there is no significant correlation for a variable, the effect size is not indicated.

Online Resource 4: Descriptive statistics (full table)

Independent-samples t-test					
		THA		**TKA**	
Variable	**Sex**	**n**	**Mean (SD)**	**n**	**Mean (SD)**
Age (years)	Male	978	68.719 (9.045)	715	71.004[a] (7.902)
	Female	1390	68.862 (9.179)	915	70.002[a] (8.704)
LOS (days)	Male	978	7.338 (4.157)	715	7.438 (3.548)
	Female	1390	7.463 (4.014)	915	7.371 (3.418)
Surgery duration (min.)	Male	978	46.474 (18.060)	715	57.834[b] (17.669)
	Female	1390	45.756 (17.048)	915	55.256[b] (18.715)
In-hospital complication ratio (general)	Male	978	0.025 (0.155)	715	0.025 (0.157)
	Female	1390	0.019 (0.136)	915	0.015 (0.123)
In-hospital complication ratio (specific)	Male	978	0.009 (0.096)	715	0.015[a] (0.123)
	Female	1390	0.012 (0.110)	915	0.004[a] (0.066)
EQ-5D-5L (pre-surgery)	Male	978	0.576*** (0.275)	715	0.625*** (0.253)
	Female	1390	0.512*** (0.291)	915	0.534*** (0.284)
EQ-5D-5L (3-month FU)	Male	978	0.883** (0.156)	715	0.844*** (0.163)
	Female	1390	0.867** (0.172)	915	0.817*** (0.182)
EQ-5D-5L (12-month FU)	Male	978	0.910 (0.149)	715	0.885*** (0.154)
	Female	1390	0.903 (0.160)	915	0.862*** (0.169)
WOMAC (pre-surgery)	Male	895	45.658*** (15.516)	631	41.765*** (15.279)
	Female	1265	50.154*** (14.609)	837	46.435*** (14.392)
	Male	868	13.048 (12.189)	634	17.869*** (13.450)

WOMAC (3-month FU)	Female	1147	14.360 (12.594)	776	19.999*** (14.277)
WOMAC (12-month FU)	Male	852	11.016 (12.977)	609	14.430*** (13.640)
	Female	1144	11.978 (13.886)	704	17.543*** (15.127)
Elixhauser Comorbidity Score	Male	959	3.803** (4.957)	703	3.459 (4.773)
	Female	1373	3.314** (4.548)	907	3.261 (4.337)
Mann-Whitney U test					
	THA			**TKA**	
Variable	**Mean (SD)**	**Median**	**Min/ Max**	**Mean (SD)**	**Median**	**Min/ Max**
Pain	1.760*** (0.430)	2	0/2	1.692*** (0.465)	2	0/2
Modified Kellgren-Lawrence Classification						
Osteophytes	1.303 (0.481)	1	0/2	0.990 (0.099)	1	0/1
Joint space	2.340*** (0.540)	2	0/3	1.817 (0.393)	2	0/2
Sclerosis	2.329*** (0.644)	2	0/3	2.053 (0.401)	2	1/3
Deformation	1.159 (0.468)	1	0/2	0.998 (0.423)	1	0/2
Severity of joint destruction in rheumatic diseases						

Rheumatic disease with manifestation at the affected joint	0.032*** (0.175)	0	0/1	0.026*** (0.160)	0	0/1
Erosive joint destruction (according to Larsen-Dale-Eek)	0.986** (1.451)	0	0/4	2.860 (0.861)	3	1/5
Pre-operative findings						
Gonarthrosis				1.080** (0.278)	1	0/2
Malposition of the knee				0.679*** (0.867)	0	0/2
ASA classification	2.059*** (0.511)	2	1/4	2.140*** (0.474)	2	1/4
Post-operative specific complications	0.011 (0.104)	0	0/1	0.009** (0.096)	0	0/1
Post-operative generic complications	0.021 (0.144)	0	0/1	0.020 (0.139)	0	0/1
Walking distance at admission	2.101** (0.570)	2	1/5	2.101** (0.558)	2	1/5
Walking aid at admission	0.239*** (0.497)	0	0/3	0.277** (0.514)	0	0/3
Walking distance at discharge	2.321 (0.484)	0	1/4	2.355 (0.493)	2	1/4
Walking aid at discharge	1.004 (0.154)	2	0/2	1.012*** (0.162)	1	0/2
Patient Clincial Complexity Level	0.741 (1.103)	0	0/5	0.754 (1.094)	0	0/4
Early mobilization	0.770** (0.421)	1	0/1	0.706 (0.456)	1	0/1

Clinic type (general vs. specialized)	0.153** (0.360)	0	0/1	0.194 (0.396)	0	0/1

Statistically significant difference between sex at a **95% and ***99% significance level; SD= Standard Deviation; LOS = Length of stay; ASA= American Society of Anesthesiologists score.

Online Resource 5: Full regression table for total hip arthroplasty (THA)

THA	**mobility**			**self-care**			**usual activity**			**pain/discomfort**			**anxiety/depression**		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
N=2368 (m=978; f=1390)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)
Age	1.016 (1.005-1.027)	1.026 (1.011 - 1.042)	1.018 (1.002 - 1.034)	1.015 (1.001 - 1.029)	.994 (.967-1.021)	1.033 (1.004 - 1.063)	0.988 (0.978 - 0.998)	0.991 (0.976 - 1.007)	1.037 (1.007 - 1.067)	1.009 (0.998 - 1.020)	0.999 (0.987 - 1.012)	0.989 (0.976 - 1.003)	0.982 (0.972-0.993)	0.989 (0.969 - 1.009)	0.996 (0.977 - 1.015)
WOMAC score (at corresponding time)	1.052 (1.045-1.059)	1.129 (1.117 - 1.142)	1.132 (1.120 - 1.145)	1.062 (1.052 - 1.072)	1.108 (1.091 - 1.125)	1.113 (1.097 - 1.130)	1.057 (1.050 - 1.064)	1.135 (1.121 - 1.148)	1.113 (1.097 - 1.129)	1.088 (1.079 - 1.097)	1.155 (1.141 - 1.169)	1.147 (1.134 - 1.160)	1.028 (1.021-1.035)	1.079 (1.066 - 1.092)	1.077 (1.066 - 1.088)
Elixhauser Comorbidity Index	1.020 (1.000-1.040)	1.019 (.991-1.048)	1.049 (1.020 - 1.079)	1.010 (0.986 - 1.035)	1.015 (.965-1.067)	0.997 (0.951 - 1.046)	1.018 (0.999 - 1.038)	1.003 (0.974 - 1.033)	0.995 (0.949 - 1.043)	1.035 (1.014 - 1.056)	1.014 (0.989 - 1.039)	1.031 (1.005 - 1.058)	1.014 (0.994-1.035)	1.014 (0.976 - 1.054)	1.048 (1.013 - 1.085)

THA	mobility			self-care			usual activity			pain/discomfort			anxiety/depression		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
Length of stay		1.017 (.976-1.060)	1.013 (.977-1.051)		1.030 (.970-1.093)	1.042 (0.991 - 1.096)		1.059 (1.017 - 1.104)	1.033 (0.983 - 1.085)		1.002 (0.964 - 1.041)	0.970 (0.935 - 1.006)		1.051 (1.000 - 1.105)	1.027 (0.985 - 1.071)
Surgery duration		1.018 (1.008 - 1.028)	1.009 (.999-1.019)		1.005 (.987-1.023)	1.002 (0.984 - 1.021)		1.009 (0.999 - 1.020)	1.002 (0.985 - 1.020)		1.001 (0.992 - 1.010)	1.001 (0.992 - 1.010)		1.003 (0.989 - 1.018)	0.989 (0.975 - 1.002)
Gender (male vs female)	0.996 (0.829-1.196)	1.331 (1.032 - 1.718)	1.660 (1.275 - 2.162)	1.762 (1.392 - 2.231)	1.618 (1.015 - 2.577)	2.465 (1.519 - 4.003)	0.964 (0.806 - 1.153)	1.130 (0.858 - 1.488)	3.063 (1.890 - 4.964)	0.849 (0.698 - 1.032)	0.853 (0.682 - 1.067)	1.102 (0.875 - 1.387)	0.489 (0.401-0.598)	0.670 (0.458 - 0.979)	0.881 (0.625 - 1.243)
ASA															
1 vs 3	0.701 (0.476-1.031)	1.276 (.727-2.240)	.974 (.523-1.812)	0.331 (0.184 - 0.593)	1.887 (.637-5.590)	2.407 (0.869 - 6.664)	0.606 (0.416 - 0.882)	0.853 (0.473 - 1.538)	1.842 (0.687 - 4.938)	0.872 (0.579 - 1.313)	0.911 (0.561 - 1.479)	1.155 (0.700 - 1.906)	0.387 (0.254-0.591)	0.642 (0.279 - 1.480)	0.944 (0.433 - 2.060)
2 vs 3	0.627 (0.483-0.813)	1.152 (.811-1.636)	.974 (.683-1.390)	0.597 (0.449 - 0.795)	1.337 (.739-2.417)	0.681 (0.388 - 1.194)	0.540 (0.421 - 0.692)	0.726 (0.503 - 1.047)	0.659 (0.379 - 1.145)	0.746 (0.576 - 0.968)	0.959 (0.692 - 1.328)	0.912 (0.653 - 1.274)	0.587 (0.453-0.760)	0.831 (0.517 - 1.336)	0.956 (0.619 - 1.474)
Clinic type (general vs specialized)		1.575 (0.982 - 2.527)	1.308 (0.812 - 2.109)		0.550 (0.251 - 1.204)	1.180 (0.521 - 2.673)		1.421 (0.863 - 2.340)	1.362 (0.610 - 3.038)		1.120 (0.747 - 1.681)	1.207 (0.794 - 1.835)		1.005 (0.529 - 1.909)	0.581 (0.327 - 1.031)

THA	mobility			self-care			usual activity			pain/discomfort			anxiety/depression		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
Pain (none or stress pain vs rest pain)	1.048 (0.846-1.299)	1.177 (0.873-1.585)	1.055 (0.768-1.449)	0.964 (0.717-1.297)	1.499 (0.844-2.661)	0.701 (0.365-1.348)	0.933 (0.754-1.154)	0.742 (0.528-1.042)	0.670 (0.349-1.286)	0.960 (0.760-1.213)	0.996 (0.768-1.292)	0.879 (0.667-1.157)	0.967 (0.766-1.221)	1,074 (0.695-1.659)	0.954 (0.626-1.452)
Joint space															
none or clearly narrowed vs lifted	1,188 (0.703-2.007)	0.858 (0.431-1.708)	1.183 (0.586-2.388)	1.027 (0.529-1.991)	0.253 (0.030-2.156)	1.437 (0.360-5.742)	1.480 (0.896-2.446)	1.340 (0.671-2.677)	1.306 (0.341-4.995)	1.841 (1.085-3.125)	1.250 (0.685-2.283)	1.274 (0.681-2.381)	0.982 (0.557-1.730)	1.432 (0.587-3.495)	0.880 (0.337-2.298)
advanced narrowed vs lifted	0.826 (0.669-1.020)	1.136 (0.837-1.542)	1.073 (0.784-1.469)	0.978 (0.741-1.290)	1.174 (0.665-2.072)	1.211 (0.686-2.136)	1.020 (0.829-1.254)	1.123 (0.808-1.560)	1.104 (0.633-1.928)	1.099 (0.877-1.376)	1.009 (0.776-1.312)	1.151 (0.876-1.512)	1.272 (1.016-1.592)	1.206 (0.784-1.854)	0.947 (0.637-1.408)
Sclerosis															
none or light sclerosis vs sclerosis with cysts	1.414 (1.020-1.959)	1.091 (0.689-1.726)	1.195 (0.750-1.906)	1.272 (0.843-1.919)	1.234 (0.573-2.657)	1.140 (0.513-2.536)	1.436 (1.048-1.969)	1.062 (0.660-1.708)	1.110 (0.506-2.433)	1.686 (1.207-2.356)	1.004 (0.661-1.524)	1.386 (0.921-2.085)	0.804 (0.573-1.129)	1.024 (0.557-1.881)	1.811 (1.040-3.154)
light sclerosis with cyst formation vs sclerosis with cysts	1.180 (0.957-1.454)	0.861 (0.639-1.161)	1.077 (0.789-1.469)	1.074 (0.814-1.417)	0.645 (0.368-1.127)	0.620 (0.357-1.075)	1.196 (0.974-1.468)	0.681 (0.494-0.939)	0.710 (0.413-1.219)	1.103 (0.883-1.378)	1.161 (0.898-1.501)	1.304 (0.996-1.708)	0.772 (0.619-0.963)	0.801 (0.530-1.210)	1.343 (0.909-1.985)

THA	mobility			self-care			usual activity			pain/discomfort			anxiety/depression		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
Deformation															
none vs clear deformation	0.927 (0.579-1.484)	1.224 (0.617 - 2.427)	1.644 (0.867 - 3.118)	0.544 (0.271 - 1.089)	0.410 (0.098 - 1.712)	0.572 (0.172 - 1.906)	0.882 (0.554 - 1.404)	1.086 (0.511 - 2.307)	0.646 (0.206 - 2.026)	0.570 (0.340 - 0.958)	1.525 (0.836 - 2.782)	1.876 (1.041 - 3.382)	1.088 (0.670-1.769)	1.778 (0.817 - 3.872)	0.683 (0.295 - 1.582)
light deformation vs clear deformation	1.052 (0.830-1.332)	1.072 (0.755 - 1.522)	0.808 (0.569 - 1.147)	0.840 (0.628 - 1.123)	1.175 (0.636 - 2.167)	0.802 (0.440 - 1.459)	0.882 (0.701 - 1.109)	1.360 (0.934 - 1.980)	0.764 (0.424 - 1.374)	0.986 (0.771 - 1.262)	1.332 (0.980 - 1.810)	1.194 (0.869 - 1.640)	0.994 (0.777-1.272)	0.865 (0.537 - 1.394)	0.762 (0.496 - 1.172)
Rheutmatic disease with manifestation at the affected joint (no vs yes)	0.928 (0.568-1.517)	0.758 (0.382 - 1.501)	1.274 (0.548 - 2.960)	1.186 (0.640 - 2.199)	0.528 (0.180 - 1.552)	1.472 (0.347 - 6.244)	1.287 (0.793 - 2.089)	0.498 (0.253 - 0.981)	1.654 (0.400 - 6.847)	0.969 (0.589 - 1.594)	1.103 (0.600 - 2.026)	1.033 (0.523 - 2.043)	0.934 (0.575-1.519)	1.049 (0.422 - 2.608)	0.990 (0.403 - 2.429)
Walking distance at admission															
infinite (>500m) vs mobile on station level (50) or less	0.657 (0.453-0.953)	1.204 (0.687 - 2.111)	0.543 (0.294 - 1.005)	0.963 (0.568 - 1.634)	0.702 (0.241 - 2.041)	0.606 (0.186 - 1.972)	0.708 (0.489 - 1.024)	0.800 (0.427 - 1.499)	0.642 (0.199 - 2.076)	0.836 (0.557 - 1.254)	0.690 (0.433 - 1.098)	0.466 (0.281 - 0.774)	1.387 (0.934-2.058)	1.486 (0.729 - 3.030)	0.993 (0.495 - 1.992)

THA	mobility			self-care			usual activity			pain/discomfort			anxiety/depression		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
walking at a stretch up to 500m possible vs mobile on station level (50) or less	1.068 (0.833-1.370)	1.432 (1.013 - 2.023)	0.959 (0.681 - 1.349)	1.138 (0.846 - 1.530)	0.858 (0.488 - 1.510)	0.916 (0.520 - 1.612)	0.997 (0.786 - 1.266)	1.302 (0.906 - 1.869)	0.971 (0.555 - 1.698)	1.014 (0.788 - 1.304)	0.872 (0.648 - 1.173)	0.821 (0.605 - 1.115)	1.188 (0.920-1.535)	1.059 (0.663 - 1.691)	0.953 (0.626 - 1.453)
Walking aid at admission (no vs yes)	0.634 (0.500-0.804)	1.089 (0.782 - 1.515)	0.966 (0.693 - 1.347)	0.696 (0.528 - 0.917)	0.824 (0.464 - 1.464)	0.896 (0.507 - 1.581)	0.637 (0.508 - 0.799)	1.085 (0.763 - 1.542)	0.847 (0.490 - 1.467)	0.839 (0.660 - 1.067)	1.580 (1.171 - 2.131)	1.200 (0.885 - 1.626)	0.751 (0.592-0.952)	1.302 (0.821 - 2.068)	0.949 (0.634 - 1.419)
Early mobilization (no vs yes)		0.966 (0.720 - 1.297)	0.958 (0.707 - 1.300)		1.070 (0.634 - 1.807)	0.847 (0.494 - 1.450)		1.062 (0.777 - 1.450)	0.978 (0.582 - 1.644)		1.079 (0.835 - 1.395)	1.045 (0.803 - 1.360)		0.906 (0.607 - 1.352)	1.066 (0.736 - 1.543)
pre-operative EQ-5D corresponding dimension															
1 vs ≥3		0.214 (0.139 - 0.329)	0.549 (0.353 - 0.854)		0.118 (0.057 - 0.244)	0.096 (0.045 - 0.202)		0.209 (0.136 - 0.319)	0.242 (0.117 - 0.503)		0.288 (0.103 - 0.810)	0.772 (0.332 - 1.795)		0.097 (0.055 - 0.169)	.099 (.059-.168)
2 vs ≥3		0.353 (0.245	0.554 (0.373		0.458 (0.218	0.295 (0.135		0.379 (0.259	0.464 (0.256		0.691 (0.547	0.843 (0.663		0.312 (0.183	0.443 (0.268

THA	mobility			self-care			usual activity			pain/discomfort			anxiety/depression		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
		- 0.510)	- 0.824)		- 0.962)	- 0.642)		- 0.556)	- 0.840)		- 0.873)	- 1.072)		- 0.532)	- 0.734)
R-squared (McFadden)	0.096	0.303	0.316	0.123	0.352	0.391	0.108	0.334	0.365	0.174	0.298	0.311	0.056	0.247	0.253

Statistically significant difference between sex at a 95% (light grey) and 99% (dark grey) significance level; OR = Odds ratio; CI = 95% confidence interval; ASA= American Society of Anesthesiologists score.

Online resource 6: Full regression table for total knee arthroplasty (TKA)

TKA	mobility			self-care			usual activity			pain/discomfort			anxiety/depression		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
N=1629 (m=715; f=914)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)	OR (CI)
Age	1.014 (1.001 - 1.028)	0.989 (0.972 - 1.005)	1.003 (0.986 -1.02)	1.034 (1.011 - 1.057)	1.015 (0.983 - 1.048)	0.999 (0.971- 1.027)	0.979 (0.966 - 0.992)	0.971 (0.954 - 0.987)	0.997 (0.98- 1.014)	0.981 (0.967 - 0.995)	0.984 (0.968 - 0.999)	0.986 (0.971 - 1.001)	0.992 (0.978 - 1.006)	0.993 (0.972 - 1.015)	0.995 (0.973 - 1.016)
WOMAC score (at corresponding time)	1.053 (1.045 - 1.061)	1.121 (1.108 - 1.134)	1.131 (1.118 - 1.144)	1.062 (1.047 - 1.077)	1.125 (1.102 - 1.148)	1.109 (1.091- 1.127)	1.059 (1.05- 1.068)	1.124 (1.111 - 1.138)	1.124 (1.112 - 1.137)	1.082 (1.072 - 1.093)	1.149 (1.133 - 1.164)	1.153 (1.139 - 1.168)	1.029 (1.02- 1.037)	1.08 (1.066 - 1.093)	1.068 (1.056 -1.08)

TKA	**mobility**			**self-care**			**usual activity**			**pain/discomfort**			**anxiety/depression**		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
Elixhauser Comorbidity Index	0.999 (0.973 - 1.025)	0.99 (0.957 - 1.023)	1.01 (0.976 - 1.045)	1.005 (0.966 - 1.045)	1.039 (0.976 - 1.107)	1.03 (0.973-1.089)	1.01 (0.984 - 1.037)	1.024 (0.99-1.06)	1.041 (1.006 - 1.077)	1.008 (0.98-1.037)	1.02 (0.989 - 1.052)	1.031 (1.001 - 1.063)	1.015 (0.986 - 1.044)	0.971 (0.929 - 1.015)	0.98 (0.938 - 1.025)
Surgery duration		1.003 (0.995 - 1.011)	1 (0.992 - 1.007)		1 (0.985 - 1.016)	0.985 (0.972-0.999)		0.998 (0.99-1.007)	0.999 (0.992 - 1.007)		0.997 (0.99-1.004)	0.994 (0.988 - 1.001)		0.991 (0.981 - 1.002)	1 (0.99-1.009)
Gender (male vs female)	0.906 (0.726 -1.13)	1.132 (0.845 - 1.515)	1.331 (0.983 - 1.803)	1.556 (1.086 -2.23)	2.817 (1.582 - 5.016)	2.727 (1.595-4.662)	0.853 (0.684 - 1.065)	1.167 (0.864 - 1.578)	0.814 (0.6-1.104)	0.612 (0.479 - 0.783)	0.721 (0.554 - 0.939)	0.776 (0.597 - 1.009)	0.347 (0.268 -0.45)	0.893 (0.589 - 1.354)	1.265 (0.843 - 1.899)
ASA															
1 vs 3	0.426 (0.241 - 0.754)	0.921 (0.429 -1.98)	1.058 (0.479 - 2.337)	0.117 (0.016 - 0.869)	3.019 (0.568 - 16.044)	0 (0-)	0.65 (0.365 - 1.158)	1.446 (0.69-3.029)	1.093 (0.494 - 2.417)	0.334 (0.173 - 0.645)	1.106 (0.572 - 2.138)	1.126 (0.578 - 2.192)	0.526 (0.257 - 1.077)	0.811 (0.275 - 2.395)	0.711 (0.196 - 2.585)
2 vs 3	0.616 (0.459 - 0.827)	0.811 (0.57-1.155)	0.664 (0.466 - 0.946)	0.523 (0.353 - 0.775)	1.098 (0.571 - 2.112)	0.745 (0.424-1.311)	0.710 (0.533 - 0.944)	0.925 (0.643 - 1.331)	0.663 (0.466 - 0.943)	0.584 (0.43-0.793)	0.875 (0.623 - 1.228)	0.870 (0.626 - 1.208)	0.999 (0.723 -1.38)	0.587 (0.371 -0.93)	0.737 (0.467 - 1.165)
Rheutmatic disease with manifestation	0.814 (0.428	1.738 (0.699 -4.32)	2.180 (0.838	0.711 (0.289	0.439 (0.106	0.560 (0.155-2.017)	0.801 (0.425	0.871 (0.377	0.344 (0.167	0.901 (0.456	2.394 (1.051	0.567 (0.277	1.158 (0.582	0.674 (0.244 -1.86)	0.975 (0.342

TKA	**mobility**			**self-care**			**usual activity**			**pain/discomfort**			**anxiety/depression**		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
at the affected joint (no vs yes)	- 1.549)		- 5.671)	- 1.754)	- 1.815)		- 1.508)	- 2.015)	- 0.711)	- 1.781)	- 5.455)	- 1.158)	- 2.303)		- 2.774)
Gonarthrosis (primary vs secondary)	0.881 (0.592 - 1.311)	0.728 (0.439 - 1.207)	1.127 (0.659 - 1.927)	0.602 (0.335 - 1.082)	0.353 (0.152 - 0.816)	0.362 (0.175- 0.751)	0.841 (0.567 - 1.247)	0.530 (0.322 - 0.873)	0.925 (0.55- 1.558)	1.879 (1.203 - 2.937)	1.084 (0.682 - 1.724)	1.818 (1.134 - 2.915)	1.173 (0.741 - 1.857)	0.905 (0.442 - 1.852)	0.926 (0.476 - 1.804)
Walking aid at admission (no vs yes)	0.826 (0.641 - 1.064)	0.912 (0.664 - 1.252)	1.152 (0.825 - 1.609)	0.692 (0.476 - 1.007)	0.929 (0.511 - 1.688)	0.706 (0.419- 1.191)	0.692 (0.54- 0.888)	0.933 (0.675 -1.29)	1.174 (0.843 - 1.635)	0.812 (0.621 - 1.062)	1.174 (0.877 - 1.572)	1.207 (0.899 -1.62)	1.066 (0.81- 1.403)	0.94 (0.619 - 1.426)	1.389 (0.892 - 2.162)
Patient Clinical Complexity Level (PCCL) (0vs1)	0.977 (0.769 - 1.241)	0.991 (0.729 - 1.346)	1.32 (0.958 - 1.821)	0.637 (0.435 - 0.933)	0.828 (0.458 - 1.495)	1.023 (0.595- 1.758)	0.87 (0.686 - 1.103)	0.816 (0.597 - 1.115)	1.134 (0.825 -1.56)	0.99 (0.765 -1.28)	1.186 (0.900 - 1.562)	1.755 (1.327 - 2.321)	1.091 (0.84- 1.417)	0.885 (0.592 - 1.323)	0.734 (0.489 - 1.102)
Malposition of the knee															
none vs severe varus knee		0.994 (0.713 - 1.385)	1.281 (0.907 - 1.809)		1.231 (0.631 - 2.402)	1.121 (0.623- 2.017)		1.26 (0.893 - 1.778)	1.04 (0.741 - 1.461)		0.830 (0.62- 1.111)	1.217 (0.91- 1.627)		1.009 (0.644 -1.58)	1.157 (0.735 - 1.821)

TKA	**mobility**			**self-care**			**usual activity**			**pain/discomfort**			**anxiety/depression**		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
severe valgus knee vs severe varus knee		1.492 (0.949 - 2.347)	2.169 (1.367 - 3.439)		1.685 (0.702 - 4.045)	1.441 (0.656-3.166)		1.29 (0.798 - 2.085)	0.988 (0.617 - 1.581)		0.861 (0.569 - 1.301)	1.615 (1.074 - 2.427)		0.767 (0.414 - 1.422)	1.621 (0.898 - 2.923)
Post-operative specific complications (no vs yes)		0.206 (0.070 - 0.606)	0.326 (0.107 - 0.992)		1.105 (0.198 - 6.164)	1.822 (0.31-10.701)		0.912 (0.266 - 3.127)	0.395 (0.127 - 1.221)		0.314 (0.099 - 0.992)	0.823 (0.254 - 2.669)		1.341 (0.279 - 6.451)	0.335 (0.099 - 1.134)
Walking distance at discharge															
infinite (>500m) vs mobile on station level (50) or less		2.219 (0.458 - 10.746)	1.385 (0.242 - 7.923)		0 (0-0)	11.845 (1.063-131.953)		2.28 (0.427 - 12.163)	2.168 (0.39-12.058)		4.421 (0.802 - 24.381)	1.88 (0.443 -7.97)		2.165 (0.23-20.384)	7.025 (1.278 - 38.63)
walking at a stretch up to 500m possible vs mobile on station level (50) or less		0.838 (0.628 - 1.117)	1.073 (0.794 -1.45)		1.318 (0.742 - 2.342)	1.864 (1.08-3.217)		1.356 (1.004 - 1.831)	1.16 (0.858 - 1.569)		1.253 (0.97-1.62)	0.974 (0.752 -1.26)		0.888 (0.605 - 1.302)	1.093 (0.739 - 1.617)

TKA	**mobility**			**self-care**			**usual activity**			**pain/discomfort**			**anxiety/depression**		
	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12	pre	FU03	FU12
Walking aid at discharge (no vs yes)		0.845 (0.22-3.251)	0 (0-0)		0 (0-)	0 (0-)		1.672 (0.456 - 6.126)	0.476 (0.071 - 3.195)		2.035 (0.504 - 8.224)	1.087 (0.252 - 4.684)		1.703 (0.309 -9.38)	0 (0-0)
pre-operative EQ-5D corresponding dimension															
1 vs ≥3		0.148 (0.09-0.243)	0.296 (0.18-0.485)		0.140 (0.053 - 0.368)	0.157 (0.067-0.370)		0.147 (0.090 - 0.239)	0.246 (0.153 - 0.396)		0.524 (0.252 - 1.088)	0.389 (0.169 - 0.892)		0.042 (0.022 -0.08)	0.094 (0.049 -0.18)
2 vs ≥3		0.272 (0.175 - 0.423)	0.447 (0.287 - 0.696)		1.032 (0.375 - 2.842)	1.147 (0.461-2.852)		0.321 (0.202 - 0.509)	0.426 (0.27-0.67)		0.615 (0.467 -0.81)	0.664 (0.509 - 0.868)		0.242 (0.133 -0.44)	0.457[b] (0.242 - 0.862)
R-squared (McFadden)	0.085	0.299	0.321	0.137	0.408	0.526	0.104	0.325	0.315	0.164	0.309	0.331	0.065	0.285	0.218

Statistically significant difference between sex at a 95% (light grey) and 99% (dark grey) significance level; OR = Odds ratio; CI = 95 % confidence interval; ASA= American Society of Anesthesiologists score.

3.2 Gender health gap pre- and post-joint arthroplasty: Identifying affected patient-reported health domains

Viktoria Steinbeck

Anja Yvonne Bischof

Lukas Schöner

Benedikt Langenberger

David Kuklinski

Alexander Geissler

Christoph Pross

Reinhard Busse

Published in the *International Journal for Equity in Health*

3.2.1 Abstract

Background: As patient-reported outcomes (PROs) gain prominence in hip and knee arthroplasty (HA and KA), studies indicate PRO variations between genders. Research on the specific health domains particularly impacted is lacking. Hence, we aim to quantify the gender health gap in PROs for HA/KA patients, differentiating between general health, health-related quality of life (HrQoL), physical functioning, pain, fatigue, and depression.

Methods: The study included 3,693 HA patients (1,627 men, 2,066 women) and 3,110 KA patients (1,430 men, 1,680 women) receiving surgery between 2020 to 2021 in nine German hospitals, followed up until March 2022. Questionnaires used were: EQ-VAS, EQ-5D-5L, HOOS-PS, KOOS-PS, PROMIS-F-SF, PROMIS-D-SF, and a joint-specific numeric pain scale. PROs at admission, discharge, 12-months post-surgery, and the change from admission to 12-months (PRO-improvement) were compared by gender, tested for differences, and assessed using multivariate linear regressions. To enable comparability, PROs were transformed into z-scores (standard deviations from the mean).

Results: Observed differences between genders were small in all health domains and differences reduced over time. Men reported significantly better health versus women pre-HA (KA), with a difference of 0.252 (0.224) standard deviations from the mean for pain, 0.353 (0.243) for fatigue (PROMIS-F-SF), 0.327 (0.310) for depression (PROMIS-D-SF), 0.336 (0.273) for functionality (H/KOOS-PS), 0.177 (0.186) for general health (EQ-VAS) and 0.266 (0.196) for HrQoL (EQ-5D-5L). At discharge, the gender health gap reduced and even disappeared for some health dimensions since women improved in health to a greater extent than men. No gender health gap was observed in most PRO-improvements and at month 12.

Conclusions: Men experiencing slightly better health than women in all health dimensions before surgery while experiencing similar health benefits 12-months post-surgery, might be an indicator of men receiving surgery inappropriately early, women unnecessarily late or both. As studies often investigate the PRO-improvement, they miss pre-surgery gender differences, which could be an important target for

improvement initiatives in patient-centric care. Moreover, future research on cutoffs for meaningful between-group PRO differences per measurement time would aid the interpretation of gender health disparities.

Trial registration: German Register for Clinical Trials, DRKS00019916, 26 November 2019

Keywords: gender health gap, patient-reported outcomes, health disparities

3.2.2 Background

In many medical fields, health differences can be observed based on sex and/or gender, the so-called sex/gender health gap. Prominent examples are differences in life expectancy [1] and chronic pain prevalence [2]. Moreover, research in different medical domains has revealed biases in the way patients are diagnosed, treated, and cared for due to their sex and gender, e.g., medical guidelines based on predomi-nantly male symptoms that lead to under- and late diagnoses of women [3, 4] and under or non-representation of women in clinical trials [5]. Gender and sex biases sometimes get replicated, for instance, by feeding biased datasets or guidelines into decision tools, leading to outdated or misleading diagnoses and recommenda-tions [6]. While women appear to be disproportionally affected by revealed sex and gender biases in medicine, also male patients suffer from these biases, e.g., in the under-diagnosis of mental illnesses [7] or osteoporosis [8]. Non-binary individuals are rarely considered in analyses at all [9].

One option to measure gender health disparities aka a potential gender health gap is the utilization of Patient-Reported Outcome Measures (PROMs). PROMs are validated questionnaires that assess health from the patient's perspective [10]. However, outcome measurement in health systems currently often focuses on clinician and administratively reported data, hence PROMs-results are not widely available and consequently not always used when speak about the gender health gap. However, over the last decade PROMs have been used more and more in some medical fields including orthopedics [11].

Hip and knee arthroplasties (HA and KA) are the most common orthopedic surger-ies worldwide with OECD averages in 2019 being 174 per 100,000 population for HA, and 137 per 100,000 for KA [12]. Germany ranked on top of the list that year for HA, with 315 surgeries per 100,000 population, and fourth highest for KA with 227 per 100,000. One of the main reasons for HA and KA is the treatment of end-stage osteoarthritis [13]. As the prevalence of osteoarthritis is higher in women [14], they receive HA and KA more frequently [14].

In orthopedic studies, PROMs are often either not reported separately by gender or are presented as secondary findings [15]. Consequently, these studies fail to adequately account for confounding factors when analyzing gender-related health differences [6]. Moreover, evidence concerning which health dimensions are af-fected by gender-related disparities and whether these differences can be consid-ered meaningful is rare in recently published literature [16-20]. In addition, anal-yses concerning the gender health gap in HA and KA in the context of the German healthcare system, to our knowledge, do not yet exist. Since systematic gender differences in health can play a role in the dynamic of under- and oversupply of healthcare services (i.e. low-value care), an evaluation per country can give specif-ic insights into how to adjust care in the national setting.

Hence, this paper aims to address this research gap by evaluating whether a gender health gap exists pre-surgery, at discharge, 12-months post-HA and KA, and over time from admission to 12-months post-HA and KA (PRO-improvement). Data from a large German multicenter randomized-controlled trial was used to investi-gate the research question for multiple generic and disease-specific PROMs cover-ing general health, health-related quality of life (HrQoL), depression, fatigue, pain and physical functioning.

3.2.3 Methods

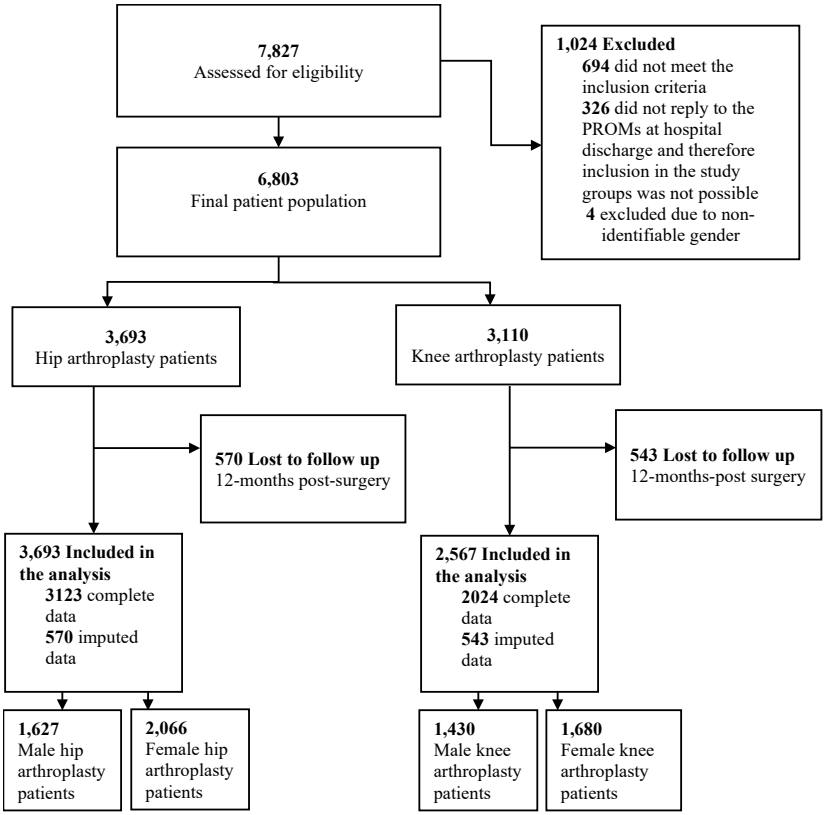
Dataset

In this retrospective cohort study, we make secondary use of a dataset originating from the PROMoting Quality study [21], which covered primary HA and KA pa-tients across nine hospitals in Germany in a randomized controlled trial. Patients underwent surgery between 2020 and 2021 and were followed up until March 2022. Patients received generic and treatment-specific PROMs, at admission to the hos-pital, at discharge, and 12-months post-surgery. The intervention of the random-ized controlled trial started at month 1 post-surgery. Further information on the PROMoting Quality trial can be found elsewhere [21, 22]. We control for the PROMoting Quality intervention when looking at the PRO-improvement and 12-

month PRO values together with other control variables to distill the gender health gap.

The selection criteria for this study are detailed in Fig. 1 and were based on the inclusion criteria of the trial [21] and the availability of the gender variable. Due to the non-representation of diverse individuals, this study focuses on the gender differences between individuals identifying as men and women only. The gender variable was retrieved from the information patients had to self-report. Other miss-ing data besides the gender variable was imputed using the MissForest Package in R Version 4.1.3.

Figure 3.5: Participant flow



Patient-reported outcome measures

General health was measured via the generic PROMs European Quality of Life Visual Analogue Scale (EQ-VAS) and HrQoL was measured using the 5-level EQ-5D version (EQ-5D-5L) [23]. The EQ-VAS ranges from 0–100 and the EQ-5D-5L ranges from -0.661 to 1, with higher scores indicating better health. We utilized the treatment specific PROMs Hip disability/Knee injury and Osteoarthritis Outcome Score Physical Function Shortform (HOOS-PS / KOOS-PS) respectively for HA and KA patients. Both cumulate in a score on a scale from 0–100 from “no difficul-ty” to “extreme difficulty”, meaning that lower scores indicate better physical health. The Patient Reported Outcomes Measurement Information System Fatigue Shortform (PROMIS-F-SF) measuring fatigue and the PROMIS Depression Short-form (PROMIS-D-SF) measuring depression symptoms both cover four questions and are summarized to a score from 33.7–75.8 and 41–79.4 respectively, from no fatigue/ depression to extreme fatigue/ depression [24]. Pain in the operated joint was measured on a numeric rating scale from 1 to 10, from lowest to highest possi-ble pain. All patients were asked about the pain in the right hip joint, the left hip joint, the right knee joint and the left knee joint. We utilized the answer for the operated joint instead of a cumulative score of the answers for all pain questions. This was done because utilizing the pain scores for all four joints would have downplayed the severity of pain experienced, especially given that the pain in the operated joint is often most relevant for the decision to perform surgery and also where the most pronounced impact after surgery is expected. In case the surgery was performed on both sides (e.g. the right and left hip joint) we used the average pain of both sides (this applied to 123 HA patients and 104 KA patients). All PROMs were used in the validated German versions [24-29].

Statistical analyses

Statistical gender differences – raw data: First, summary statistics of patient and treatment characteristics are displayed and compared between genders, using t-test and Chi-square test for continuous and categorical variables respectively. Similarly, the PROs at admission, at discharge and 12-months post-surgery were analyzed per

gender. Due to non-normal distribution of PROs, the Mann–Whitney U test was utilized to test for independence between genders. This enabled the analysis of statistically significant differences in PROs between genders.

Meaningful gender differences – raw data: In addition, we analyzed whether the differences in PRO-improvement can be considered meaningful based on previously published minimally clinical important difference (MCID) thresholds. Where possible we utilized the MCIDs calculated based on the same German HA (KA) dataset which were 0.17 (0.20) for EQ-5D-5L, 7.81 (5.86) for EQ-VAS, -10.01 (-5.06) for HOOS-PS (KOOS-PS) [30] and -0.9 (-0.7) for pain [31]. For the PROMIS measures we used a 2-point cutoff [32].

Statistical gender differences – adjusted for confounders: For the following main analyses, PROs per measurement time and for the PRO-improvement were transformed into z-scores to make them comparable on one scale. Z-scores express how many standard deviations (SD) an individual's outcome is above or below the average of the population and is calculated as:

$$z\,\frac{x-\mu}{\sigma}$$

where x is the observed score per measurement time (or improvement score) of an individuum, μ is the population's mean and σ its SD. In addition, some PROs were adjusted in their directionality so that for all scores a positive coefficient indicates better health. As main analyses, linear multiple regression models were run per PRO controlling for a different set of variables depending on the timepoint to distill the contribution of gender towards the different health dimensions.

For the admission PROs as dependent variable, the control variables were age in years, body mass index (BMI) group ("underweight", "normal", "overweight", "obese"), education ("no school degree", "primary school degree", "high/middle school degree", "university degree"), living situation ("I live with a partner/family/friends", "I live alone", "I live in a care facility", "Other") and having had one of the following comorbidities (yes/no answer options per comorbidity: heart-, circulation-, blood-, or

lung-related diseases, stroke, diabetes, neurological diseases, cancer, depression, back pain and arthritis). For the discharge score as dependent variable, the previously listed information, the mobilization after surgery, the experience of the main surgeon in numbers of surgeries and the admission PRO were controlled for. For the 12-months score, in addition to the previously mentioned variables, the PRO-monitoring group [21] and the rehabilitation were controlled for. For the PRO-improvement (change from admission to month 12), we controlled for the same variables as in the 12-month regressions.

Meaningful gender differences – adjusted for confounders: As suggested by Norman et al. [33] and Bloom et al. [34], above 0.5 standard deviations can be considered as MCID across different PROMs. Hence, this threshold was used to interpret whether the differences in PRO-improvement can be considered meaningful.

All analyses were conducted in R, version 4.1.3. Due to different procedures and recovery times, all analyses were run separately for HA and KA patients.

3.2.4 Results

Patient characteristics

The study population consists of 3,693 HA patients (1,672 men and 2,066 women) and 3,110 KA patients (1,430 men and 1,680 women) as shown in Fig. 1. The descriptive statistics are presented in Table 1. HA patients identifying as women in this study are, on average 66 years old, one year older compared to patients identi-fying as men, show significant differences in their BMI, e.g., more frequently have a “normal” BMI when being admitted for surgery (38% versus 23%) and less often have a university degree (24% versus 35%). In addition, more HA patients who identify as women live alone (30% versus 14%), and were in employment prior to surgery (36% versus 30%). KA patients were on average 66 years old, not showing significant differences between genders. 50% of KA patients who identify as wom-en were obese and 42% of KA patients who identify as men. Most KA patients have a high school degree, while significant gender differences were observable (87% of men and 73% of women). The PROs at admission generally show worse health levels in women e.g.,

with the HOOS-PS being 50.43 in women and 44.39 in men. Significant differences were observable between genders in most of the rec-orded commodities but not in the PRO-monitoring group (see Appendix).

Table 3.3: Descriptive statistics of the study population (n=6,803)

	Hip replacement patients		p[b]	Knee replacement patients		p[b]
	male (N=1,627)	female (N=2,066)		male (N=1,430)	female (N=1,680)	
Age						
mean (SD)	65.18 (10.54)	66.24 (10.59)	0.003 ***	66.06 (9.19)	65.99 (9.26)	0.843
BMI[a]						
underweight	0 (0%)	20 (0.97%)	< 0.001 ***	2 (0.14%)	4 (0.24%)	< 0.001***
normal	379 (23.29%)	786 (38.04%)		191 (13.36%)	286 (17.02%)	
overweight	742 (45.61%)	674 (32.62%)		637 (44.55%)	547 (32.56%)	
obese	506 (31.10%)	586 (28.36%)		600 (41.96%)	843 (50.18%)	
Education						
no school degree	6 (0.37%)	9 (0.44%)	< 0.001 ***	6 (0.42%)	12 (0.71%)	< 0.001 ***
primary school degree	222 (13.64%)	270 (13.07%)		266 (18.60%)	268 (15.95%)	
high/middle school degree	831 (51.08%)	1,287 (62.29%)		766 (53.57%)	1,093 (65.06%)	
university degree	568 (34.91%)	500 (24.20%)		392 (27.41%)	307 (18.27%)	
Living situation						
alone	220 (13.52%)	624 (30.20%)	< 0.001 ***	172 (12.03%)	427 (25.42%)	< 0.001 ***
care facility	3 (0.18%)	8 (0.39%)		7 (0.49%)	12 (0.71%)	
with a partner/family/friends	1,392 (85.56%)	1,423 (68.88%)		1,247 (87.20%)	1,225 (72.92%)	

other	12 (0.74%)	11 (0.53%)		4 (0.28%)	16 (0.95%)	
Job						
working	579 (35.59%)	621 (30.06%)	< 0.001 ***	475 (33.22%)	447 (26.61%)	< 0.001 ***
voluntarily not working including retirement	866 (53.23%)	1230 (59.54%)		772 (53.99%)	1,018 (60.60%)	
looking for work	18 (1.11%)	17 (0.82%)		15 (1.05%)	19 (1.13%)	
unable to work	164 (10.08%)	198 (9.58%)		168 (11.75%)	196 (11.67%)	
Admission PRO[c] mean (SD)						
EQ-5D-5L	0.644 (0.239)	0.562 (0.269)	< 0.001	0.665 (0.230)	0.589 (0.262)	< 0.001
EQ-VAS	59.14 (19.90)	54.96 (19.67)	< 0.001	61.09 (19.29)	55.91 (18.92)	< 0.001
HOOS-PS/ KOOS-PS	44.29 (15.50)	50.43 (16.00)	< 0.001	40.62 (12.67)	45.13 (12.35)	< 0.001
PROMIS-D-SF	47.96 (7.78)	51.20 (8.40)	< 0.001	47.42 (7.69)	51.04 (8.39)	< 0.001
PROMIS-F-SF	46.92 (9.20)	50.97 (10.00)	< 0.001	46.28 (9.51)	49.98 (9.80)	< 0.001
Pain in joint	6.09 (2.19)	6.72 (2.05)	< 0.001	6.49 (2.04)	7.04 (1.93)	< 0.001

[a]BMI stands for body-mass index

[b]For continuous variables a two-sided t-test was performed, for and categorical variables a Chi-square test was performed and for the PRO-scores at admission, a Mann–Whitney U test

[c]PRO Patient-reported Outcome. The score ranges are: -0.661 to 1.0 for the EQ-5D-5L, 0–100 for the EQ-VAS, HOOS-PS and KOOS-PS, 33.7–75.8 for the PROMIS-F-SF, 41–79.4 for the PROMIS-D-SF and 0–10 for pain. The EQ-5D-5L is reported with three digits after the decimal point due to the small score range

Raw gender health gap

Table A2 and A3 in the appendix show the descriptive statistics of the PROs per observation time and as improvement, the mean difference between genders in points as well as the results of the significance tests. The tables also indicate whether the raw gender difference in PRO-improvement constitute MCIDs.

The results reflect significant differences between genders in almost all PROs for both joint replacement types at each timepoint, with men experiencing significant-ly better health across PROs. Solely in the pain dimension 12-months post HA, no significant gender difference is observable. Mean PRO differences between HA patients identifying as men and women show the widest disparities in PROs pre-surgery (HrQoL -0.081, general health -4.18, physical functioning 6.13, depression 3.25, fatigue 4.05, pain 0.63), followed by those at discharge (HrQoL -0.025, gen-eral health -2.18, physical functioning 5.88, depression 2.50, fatigue 3.15, pain 0.27) and the lowest at month 12 (HrQoL -0.021, general health -1.27, physical functioning 1.89, depression 1.70, fatigue 1.26, pain 0.04). For KA patients, the widest differences are also observable at admission (HrQoL -0.076, general health -5.19, physical functioning 4.51, depression 3.61, fatigue 3.61, pain 0.55), while the disparities are lower at discharge (HrQoL -0.030, general health -2.80, physical functioning 2.21, depression 3.17, fatigue 3.39, pain 0.46) and month 12 (HrQoL -0.030, general health -2.83, physical functioning 2.96, depression 2.51, fatigue 2.00, pain 0.18) than at admission. However, disparities do not reduce in all PROs from discharge to month 12 among KA patients. Comparing the two joint replace-ment types, the gender health gap is larger among HA patients at admission, but larger at discharge among KA patients in all PROs besides the physical functioning score. At month 12 the gap is larger among KA patients compared to HA patients in all PROs.

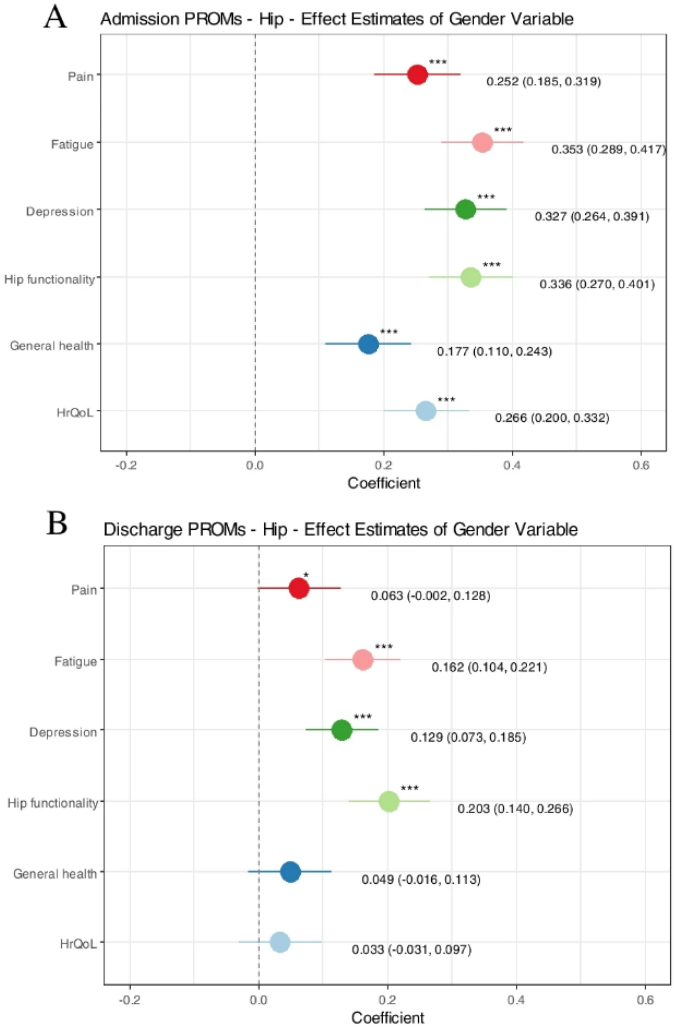
The PRO improvement from admission to month 12 shows men experience signifi-cantly less improvement in all PROs in HA and KA, with point differences of e.g. 0.061 (0.046) in HrQoL for HA (KA), -4.25 (-1.55) in physical functioning and -2.79 (-1.61) in fatigue. Based on previously mentioned MCID thresholds, only the

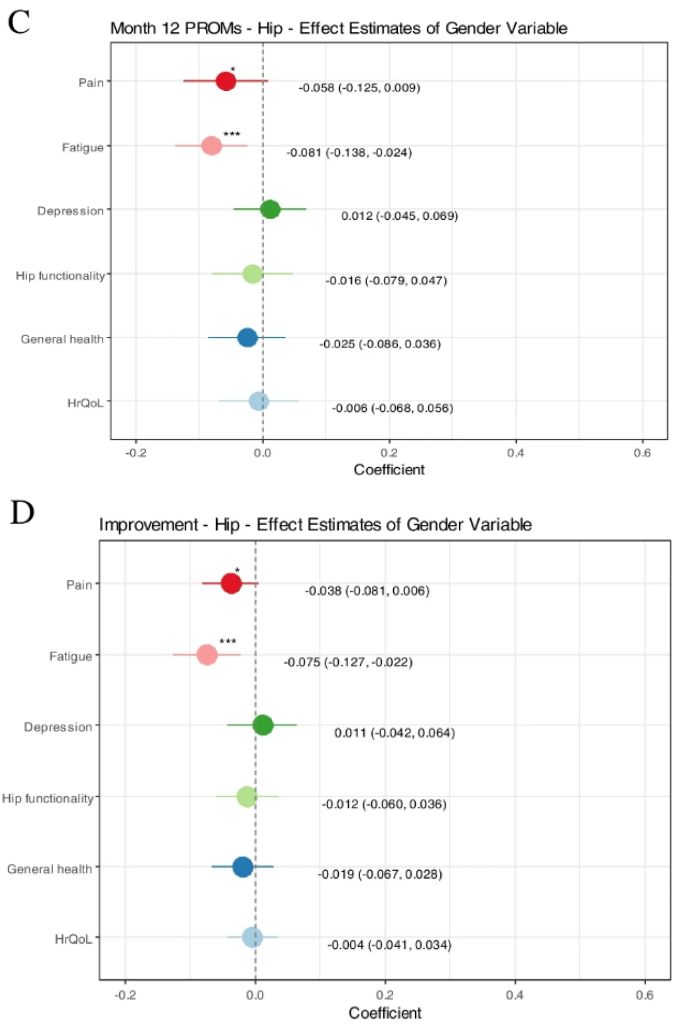
difference in PRO-improvement in fatigue among HA patients can be considered clinically meaningful whereas the other differences in improvement are not.

Gender health gap after adjusting for confounders

The gender health gap, after adjusting for confounders through the multivariate linear regression models is illustrated in Figs. 3.6 and 3.7. The results are visual-ized using z-scores, depicting standard deviations from the mean, to make the PROs visually comparable on one scale. Positive effect estimates indicate better health in men. Significantly better health in men can be observed in all health di-mensions pre-HA (KA) with standard deviations from the mean of 0.252 (0.224) for pain, 0.353 (0.243) for fatigue, 0.327 (0.310) for depression, 0.336 (0.273) for functionality, 0.177 (0.186) for general health and 0.266 (0.196) for HrQoL (Figs. 3.6A and 3.7A). Better health in men was observable at hospital discharge after HA (KA) for fatigue 0.162 (0.184) and depression 0.129 (0.170) and after HA by 0.203 in hip functionality and in pain by 0.123 after KA (Figs. 3.6B and 3.7B). The other health dimensions do not show significant differences at the $p < 0.05$ value level. 12-months post-surgery the only significant difference can be observed in men undergoing HA experiencing slightly worse fatigue than women by -0.081 (Figs. 3.6C and 3.7C). Similarly, the only difference in the PRO-improvement that is significant but not clinically meaningful is the difference in fatigue (-0.075) (Figs. 3.6D and 3.7D).

Figure 3.6: Hip arthroplasty: Explanatory power of the patients' gender on the z-score standardized PROs

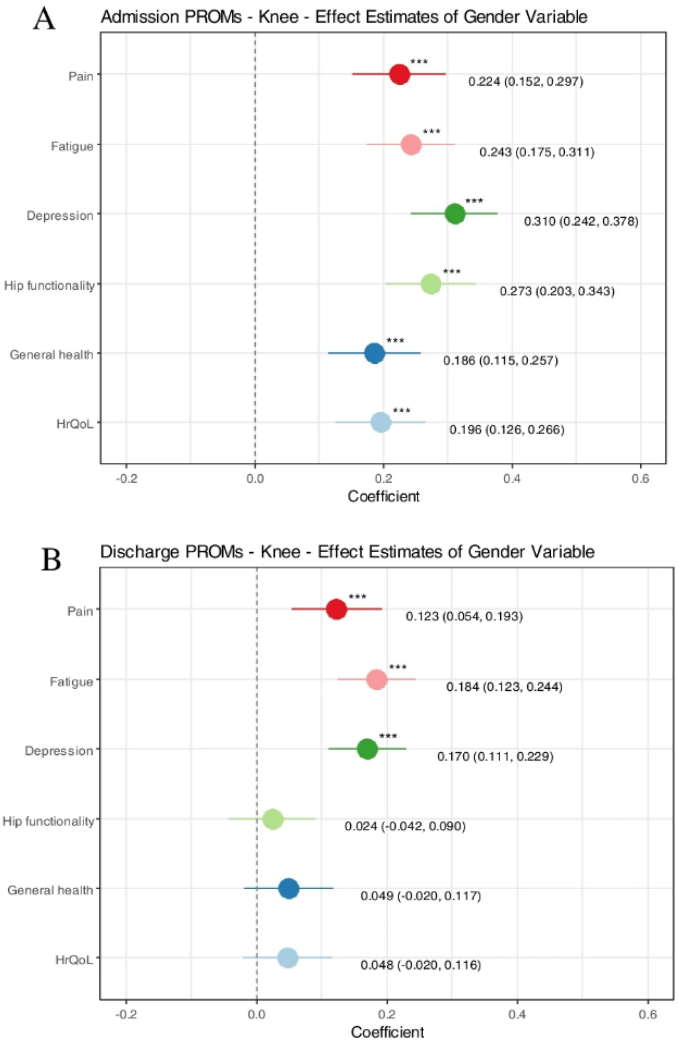


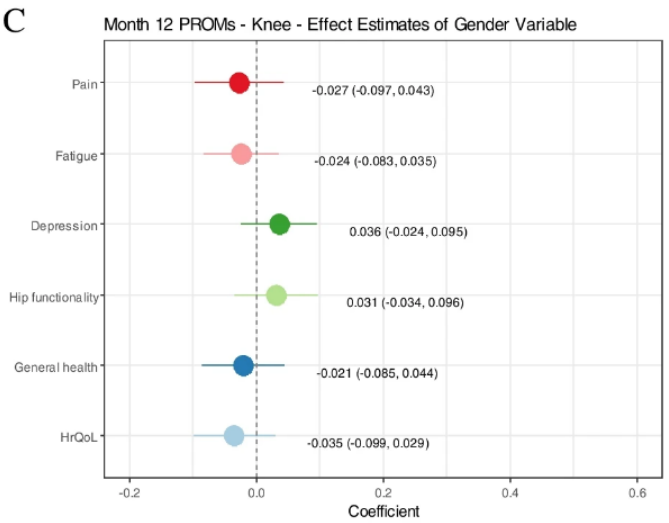


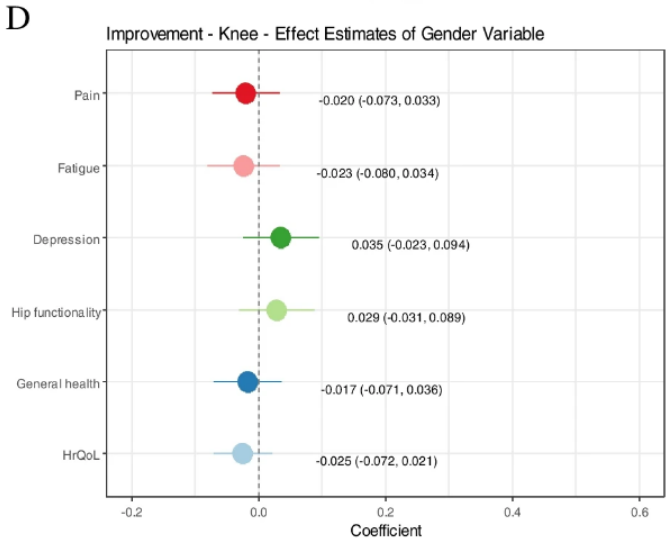
Hip arthroplasty: Explanatory power of the patients' gender on the z-score standardized Patient-reported outcomes (PROs) per measurement time. All graphs: the outcome variables (PROs) are transformed into z-scores and adjusted in their directionality so that all scores indicate better health in men if the coefficient is positive; HrQoL = Health-related quality of life. **Graph A** - admission: results of multivariate linear regression models controlling for age, education, BMI, living situation and nine comorbidities; **Graph B** - discharge: results of multivariate linear regression models controlling for age, Body Mass Index (BMI), the respective PRO at admission, the hospital, the mobilization after surgery, the experience of the main surgeon and nine comorbidities; **Graph C** - month 12: results of multivariate linear regression models controlling for age, BMI, the respective PRO at admission, the mobilization after surgery and the experience of the main surgeon, the monitoring group, the rehabilitation form and nine comorbidities; **Graph D** - PRO-

improvement: results of multivariate linear regression models controlling for age, BMI, the respective PRO at admission, the mobilization after surgery and the experience of the main surgeon, the monitoring group, the rehabilitation form and nine comorbidities.

Figure 3.7: Knee arthroplasty: Explanatory power of the patients' gender on the z-score standardized PROs







Knee arthroplasty: Explanatory power of the patients' gender on the z-score standardized Patient-reported Outcomes (PROs) per measurement time. All graphs: the outcome variables (PROs) are transformed into z-scores and adjusted in their directionality so that all scores indicate better health in men if the coefficient is positive; HrQoL = Health-related quality of life. **Graph A** - admission: results of multivariate linear regression models controlling for age, education, BMI, living situation and nine comorbidities; **Graph B** - discharge: results of multivariate linear regression models controlling for age, Body Mass Index (BMI), the respective PRO at admission, the hospital, the mobilization after surgery, the experience of the main

surgeon and nine comorbidities; **Graph C** - month 12: results of multivariate linear regression models controlling for age, BMI, the respective PRO at admission, the mobilization after surgery and the experience of the main surgeon, the monitoring group, the rehabilitation form and nine comorbidities; **Graph D** - PRO-improvement: results of multivariate linear regression models controlling for age, BMI, the respective PRO at admission, the mobilization after surgery and the experience of the main surgeon, the monitoring group, the rehabilitation form and nine comorbidities.

3.2.5 Discussion

In this retrospective cohort study of HA and KA patients, we found that observed differences between genders were present but rather small in all health domains and differences reduced over time. Women reported significantly worse PROs at hospital admission in all examined health domains in both HA and KA patients when adjusting for patient and treatment characteristics. Whereas the raw unad-justed health gap shows that the point difference between genders reduces over time, while remaining significantly different at all observed times (besides pain 12-months post HA), the adjusted analyses showed that a significant gender health gap disappeared for some health domains at discharge (general health and HrQoL in both, HA and KA, physical functioning in KA and pain in HA) and all health domains 12-months post-surgery besides fatigue in HA patients (where the effect direction reverses). The gender differences in PRO-improvements can be consid-ered not meaningful based on specified MCIDs. Unfortunately, no cutoff points exist that specify which between-group differences per measurement time (admis-sion, discharge, month 12) can be considered meaningful, demanding future re-search.

The main analyses (multivariate regressions) of health differences at month 12, showcases the gender difference in individual health gain, meaning that a woman undergoing HA/KA has the same individual health gain (in all PROs except fatigue in HA patients) as a man if baseline PROs were the same (other patient and treat-ment variables controlled for). Putting this into perspective of the raw analyses reflecting women's worse average admission and similar 12-month health status relative to men means on average women undergoing HA/KA in this study popula-tion have higher health gains than men. Women, e.g., experience around double the fatigue symptom improvement than men. Hence, the relationship between admis-sion PROs and month

12 PROs in the context of HA/KA is similar for both genders but the observed distribution of health states per gender per timepoint is not.

The fact that men experience better overall, mental, and physical health prior to surgery while reaching similar self-reported health status 12-months post-surgery might be an indicator of men receiving surgery earlier than necessary or women receiving surgery later than necessary. It could also indicate that women and men are systematically different in reaching the right time for surgery at different health impairment levels. While there are studies showing gender health differ-ences in the general population (e.g.Bloom et al. [35]), we assume that the health gap observed in the HA/KA population goes beyond this difference as the gender health gap in the HA/KA population reduces over time after surgery. Whether the observed differences are based on individual patient decisions, guidance from care providers or other factors is unclear and requires further research and care deci-sion-making sensitive to gender differences. Hypotheses are presented in some papers and include women being referred to a surgeon only with a higher degree of disability [3, 36], women undergoing surgery with a higher age [37], and the un-willingness to accept surgery from the patient's side [38,39,40]. As summarized by Novicoff and Saleh, women express more concerns regarding the risks of treatment and the disruption of their family-role [41]. In the context of the German healthcare system, evaluations from the national registry showed that there are higher risks of endoprosthesis failure due to a higher infection risk in men, which could explain these findings as well e.g. if women at risk for infection avoid sur-gery or men improve less on average due to higher infection risks [42]. Moreover, research on gender stereotypes also pointed to reporting and interpretation differ-ences in pain perception, with women's pain being judged as less severe by physi-cians than men's pain when the descriptions of pain were the same [43]. In addi-tion, Samulowitz et al. [44] and Mogil et al. [45] pointed out that women and men are brought up to express pain differently which can change their biological re-sponse to pain and their willingness to report it. However, as Moretti et al. [46] noted, gender is just one aspect of the multifactorial influences on outcomes and hence needs to be viewed from an intersectional lens when addressing

the gender health gap. Potential reasons for gender differences stated by Tannenbaum et al. [47] that can be excluded based on the design of our study are firstly the experimenter-participant interaction as PROs are self-reported and secondly product-participant interaction as all questionnaires were validated in men and women. As the EQ-5D-5L is the only preference-based PROM in our study, the preference weights from the German value set might have however led to overemphasizing health dimensions that men are doing better in for the HrQoL assessment as sug-gested by Bischof et al. [48].

Based on the findings of this study, several implications for clinical practice can be drawn. Patients identifying as women might need more specific attention in the pre- and intraoperative phases to improve health outcomes as argued by Solarino et al. [49]. Both genders might also need a closer evaluation at which health and po-tentially arthritis status a HA or KA is beneficial to reduce low-value care. The implementation and evaluation of PROs before surgery could be one way to assess the optimal moment to perform surgery as shown by Tew et al. [50]. In light of described biases in physician–patient interaction, pre-surgery PROs have the bene-fit of objectifying the health assessment. Studies often include PROs in the form of the change in PRO from before to after therapy. As shown in this paper, looking only at the health change over time, and not the baseline health status, can lead to clouded interpretations. In the case of HA and KA, the results would indicate that women benefit more from surgery, however this is driven mainly by pre-surgery PRO levels. Looking at PROs before an intervention can point out improvement potential in the treatment of subgroups, e.g., more suggestions for alternative treatment options for men if their PROs are above a certain health threshold.

This study has several strengths and limitations. It specifies the gender health gap per observation time across different health domains for many PROs while control-ling for known confounders. It moreover covers a large study population across different hospitals in Germany. Since there is no way to randomize for gender, controlling for confounders is the only way to digest the health difference due to gender. However, this study cannot control for some confounders or identify other reasons, e.g., for the

pre-surgery health gap, like treatments pre-hospital admission, expectations from patients, or previous recommendation for surgery or arthritis progression in the affected joint. Similarly, there was no information available on the follow-up appointments besides the rehabilitation and PRO-monitoring, which could also explain the disappearing health gap at 12-months (e.g. if men are im-proving less because they miss follow-up care). We cannot say for certain whether men or women are closer to the optimum of receiving surgery when indicated. In addition, the gender variable was only assessed via three different answer options in the digital platform "man", "woman", "other", whereas more and clearer gender options should have been represented in the study.

3.2.6 Conclusion

This study shows that there are small differences across health domains between men and women undergoing HA/KA that reduce over time. We observed a signifi-cant gender health gap at admission to the hospital pre-HA and KA with men re-porting better health status for pain, fatigue, depression, functionality, general health and HrQoL when controlling for various patient characteristics and comor-bidities. At discharge from the hospital, the gender health gap reduced, and showed significant differences only for depression and fatigue in HA and KA, pain in KA and functionality in HA patients. 12-months post-surgery, the gap disappeared and even reversed in fatigue in HA patients, where men showed significantly worse scores. The difference in PRO-improvement were small and not clinically mean-ingful. To understand whether the differences, per measurement time, are meaning-ful to patients, further research is needed to identify cut-off points that go beyond existing ones only applicable to PRO-improvements. Men experiencing better health at hospital admission while reaching similar health levels 12-months post-surgery, when accounting for the pre-surgery health differences, might be an indi-cator of men receiving surgery earlier than necessary or women receiving surgery later than necessary, or both. Many studies only present the improvement in PROs over time, thereby missing pre-intervention gender differences, which could be an important target for improvement initiatives in patient-centric care.

References

1. OECD (2021) Health at a Glance 2021: OECD Indicators. OECD Publishing, Paris.

2. Mills SEE, Nicolson KP, Smith BH (2019) Chronic pain: a review of its epidemiology and associated factors in population-based studies. British journal of anaesthesia 123:e273-e283.

3. Mandl LA (2013) Determining who should be referred for total hip and knee replacements. Nat Rev Rheumatol 9:351–357.

4. Chapman EN, Kaatz A, Carnes M (2013) Physicians and implicit bias: how doctors may unwittingly perpetuate health care disparities. J Gen In-tern Med 28:1504–1510.

5. Merkatz RB (1998) Inclusion of Women in Clinical Trials: A Historical Overview of Scientific Ethical and Legal Issues. Journal of Obstetric, Gynecologic & Neonatal Nursing 27:78–84.

6. Hertler C, Seiler A, Gramatzki D et al. (2020) Sex-specific and gender-specific aspects in patient-reported outcomes. ESMO Open 5:e000837.

7. Shi P, Yang A, Zhao Q et al. (2021) A Hypothesis of Gender Differences in Self-Reporting Symptom of Depression: Implications to Solve Under-Diagnosis and Under-Treatment of Depression in Males. Front Psychiatry 12:589687.

8. Szulc P, Kaufman JM, Orwoll ES (2012) Osteoporosis in men. Journal of osteoporosis 2012.

9. Nielsen MW, Stefanick ML, Peragine D et al. (2021) Gender-related variables for health research. Biology of Sex Differences 12:1–16.

10. Wilson I, Bohm E, Lübbeke A et al. (2019) Orthopaedic registries with patient-reported outcome measures. EFORT Open Rev 4:357–367.

11. Bohm ER, Kirby S, Trepman E et al. (2021) Collection and Reporting of Patient-reported Outcome Measures in Arthroplasty Registries: Multinational Survey and Recommendations. Clinical Orthopaedics and Related Research 479:2151–2166.

12. OECD (ed) (2021) Health at a Glance 2021: OECD Indicators: Hip and knee replacement. OECD Publishing, Paris

13. Hunter DJ, Bierma-Zeinstra S (2019) Osteoarthritis. Lancet 393:1745–1759.

14. Maradit Kremers H, Larson DR, Crowson CS et al. (2015) Prevalence of Total Hip and Knee Replacement in the United States. The Journal of bone and joint surgery. American volume 97:1386–1397.

15. Ramaskandhan J, Rashid A, Kometa S et al. (2020) Comparison of 5-Year Patient-Reported Outcomes (PROMs) of Total Ankle Replacement (TAR) to Total Knee Replacement (TKR) and Total Hip Replacement (THR). Foot Ankle Int 41:767–774.

16. Lim JBT, Chi CH, Lo LE et al. (2015) Gender difference in outcome after total knee replacement. J Orthop Surg (Hong Kong) 23:194–197.

17. Golinelli D, Grassi A, Tedesco D et al. (2022) Patient reported outcomes measures (PROMs) trajectories after elective hip arthroplasty: a latent class and growth mixture analysis. Journal of Patient-Reported Outcomes 6:95.

18. Turnbull GS, Scott CEH, MacDonald DJ et al. (2019) Gender and preoperative function predict physical activity levels after revision total knee arthroplasty. The Journal of arthroplasty 34:939–946.

19. Cherian JJ, Jinnah AH, Robinson K et al. (2016) Prospective, Longitudinal Evaluation of Gender Differences After Total Hip Arthroplasty. Orthopedics 39:e391-6.

20. Rolfson O, Kärrholm J, Dahlberg LE et al. (2011) Patient-reported out-comes in the Swedish Hip Arthroplasty Register: Results of a nationwide

prospective observational study. Journal of Bone and Joint Surgery - Se-ries B 93 B.

21. Kuklinski D, Oschmann L, Pross C et al. (2020) The use of digitally collected patient-reported outcome measures for newly operated patients with total knee and hip replacements to improve post-treatment recovery: Study protocol for a randomized controlled trial. Trials 21.

22. Steinbeck V, Langenberger B, Schöner L et al. (2023) Electronic Patient-Reported Outcome Monitoring to Improve Quality of Life After Joint Replacement: Secondary Analysis of a Randomized Clinical Trial. JAMA Netw Open 6:e2331301.

23. Ludwig K, Graf von der Schulenburg JM, Greiner W (2018) German Va-lue Set for the EQ-5D-5L. PharmacoEconomics 36:663–674.

24. Wahl I, Löwe B, Rose M (2011) Das Patient-Reported Outcomes Measurement Information System (PROMIS®): Übersetzung der Item-Banken für Depressivität und Angst ins Deutsche. Klin Diagnostik Eval 3:236–261.

25. Davis AM, Perruccio AV, Canizares M et al. (2009) Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. Osteoarthritis and Cartilage 17:843–847.

26. Bilbao A, García-Pérez L, Arenaza JC et al. (2018) Psychometric proper-ties of the EQ-5D-5L in patients with hip or knee osteoarthritis: reliability, validity and responsiveness. Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation 27:2897–2908.

27. Blasimann A, Dauphinee SW, Staal JB. (2014) Translation, cross-cultural adaptation, and psychometric properties of the German version of the hip disability and osteoarthritis outcome score. Journal of Orthopaedic & Sports Physical Therapy 44:989–97.

28. Kessler S, Lang S, Puhl W, Stöve J. (2003) Der Knee Injury and Osteoarthritis Outcome Score–ein Funktionsfragebogen zur Outcome-Messung in der Knieendoprothetik. Z Orthop Ihre Grenzgeb 141:277–82.

29. Greiner W, Claes C, Busschbach JJV, von der Schulenburg J-MG. (2005) Validating the EQ-5D with time trade off for the German population. Eur J Health Econ 6:124–30.

30. Langenberger B, Schrednitzki D, Halder A et al. (2023) Predicting whether patients will achieve minimal clinically important differences following hip or knee arthroplasty. Bone and Joint Research 12:512–21.

31. Langenberger B, Steinbeck V, Busse R. (2024) Who Benefits From Hip Arthroplasty or Knee Arthroplasty? Preoperative Patient-reported Out-come Thresholds Predict Meaningful Improvement. Clinical Orthopaedics and Related Research 482:867-881.

32. Terwee CB, Peipert JD, Chapman R, Lai J-S et al. (2021) Minimal im-portant change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res 30:2729–54.

33. Norman GR, Sloan JA, Wyrwich KW. (2003) Interpretation of changes in health-related quality of life: the remarkable universality of half a stand-ard deviation. Med Care 41:582–92.

34. Bloom DA, Kaplan DJ, Mojica E et al. (2023) The minimal clinically important difference: A review of clinical significance. Am J Sports Med 51:520–4.

35. König H-H, Heider D, Lehnert T et al. (2010) Health status of the ad-vanced elderly in six European countries: results from a representative survey using EQ-5D and SF-12. Health Qual Life Outcomes 8:143.

36. Borkhoff CM, Hawker GA, Kreder HJ et al. (2008) The effect of patients' sex on physicians' recommendations for total knee arthroplasty. CMAJ 178:681–687.

37. Srikanth VK, Fryer JL, Zhai G et al. (2005) A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthritis and Cartilage 13:769–781.

38. Torrente-Jimenez RS, Feijoo-Cid M, Rivero-Santana AJ et al. (2022) Gender differences in the decision-making process for undergoing total knee replacement. Patient Educ Couns 105:3459–3465.

39. Frankel L, Sanmartin C, Conner-Spady B et al. (2012) Osteoarthritis patients' perceptions of "appropriateness" for total joint replacement surgery. Osteoarthritis and Cartilage 20:967–973.

40. Postler A, Goronzy J, Günther K-P et al. (2020) Which disease-related factors influence patients' and physicians' willingness to consider joint replacement in hip and knee OA? Results of a questionnaire survey linked to claims data. BMC Musculoskeletal Disorders 21:352.

41. Novicoff WM, Saleh KJ. (2011) Examining sex and gender disparities in total joint arthroplasty. Clin Orthop Relat Res 469:1824–8.

42. Grimberg A, Lützner J, Melsheimer O et al. (2022) Jahresbericht 2022: Mit Sicherheit mehr Qualität, 2022. Auflage. EPRD Deutsche Endoprothesenregister, Berlin

43. Schwarz KA, Sprenger C, Hidalgo P et al. (2019) How Stereotypes Affect Pain. Sci Rep 9:8626.

44. Samulowitz A, Gremyr I, Eriksson E et al. (2018) Brave Men" and "Emotional Women": A Theory-Guided Literature Review on Gender Bias in Health Care and Gendered Norms towards Patients with Chronic Pain. Pain Research and Management 2018.

45. Mogil JS (2020) Qualitative sex differences in pain processing: emerging evidence of a biased literature. Nature Reviews Neuroscience 2020 21:7 21:353–365.

46. Moretti B, Spinarelli A, Varrassi G et al. (2022) Influence of sex and gen-der on the management of late-stage knee osteoarthritis. Musculoskelet Surg. 106:457-467.

47. Tannenbaum C, Greaves L, Graham ID (2016) Why sex and gender matter in implementation research Economic, social, and ethical factors affecting the implementation of research. BMC Medical Research Methodology 16:1–9.

48. Bischof AY, Steinbeck V, Kuklinski D et al. (2023) What is the associa-tion between gender and self-perceived health status when controlling for disease-specific conditions? A retrospective data analysis of pre- and post-operative EQ-5D-5L differences in total hip and knee arthroplasty. BMC Musculoskelet Disord 24:914.

49. Solarino G, Bizzoca D, Moretti AM et al. (2022) Sex and Gender-Related Differences in the Outcome of Total Hip Arthroplasty: A Current Con-cepts Review. Medicina (Kaunas) 58.

50. Tew M, Dalziel K, Clarke P et al. (2020) Patient-reported outcome measures (PROMs): can they be used to guide patient-centered care and optimize outcomes in total knee replacement? Quality of Life Research 29:3273–3283.

Appendix

Table A1: Descriptive statistics of the study population for additional variables (n = 6,803) igure A1: Hip arthroplasty: Explanatory power of the patients' gender on the z-score standardized PROs (Controlling for the PRO at admission)

	Hip replacement patients		p[b]	Knee replacement patients		p[b]
	men (N=1,627)	women (N=2,066)		men (N=1,430)	women (N=1,680)	
Comorbidities						
Heart-related (yes)	332 (20.41%)	233 (11.28%)	< 0.001 ***	252 (17.62%)	204 (12.14%)	< 0.001 ***
Blood-related (yes)	916 (56.30%)	1043 (50.48%)	< 0.001 ***	935 (65.38%)	1092 (65.00%)	0.823
Stroke (yes)	0 (0.00%)	0 (0.00%)	/	0 (0.00%)	0 (0.00%)	/
Circulation-related (yes)	0 (0.00%)	0 (0.00%)	/	0 (0.00%)	0 (0.00%)	/
Lung-related (yes)	145 (8.91%)	237 (11.47%)	0.011 **	144 (10.07%)	249 (14.82%)	< 0.001***
Diabetes (yes)	192 (11.80%)	160 (7.74%)	< 0.001 ***	160 (11.19%)	158 (9.40%)	0.102
Neurological (yes)	41 (2.52%)	68 (3.29%)	0.1691	42 (2.94%)	74 (4.40%)	0.031**
Cancer (yes)	157 (9.65%)	137 (6.63%)	< 0.001 ***	88 (6.15%)	98 (5.83%)	0.707
Depression (yes)	99 (6.08%)	224 (10.84%)	< 0.001 ***	71 (4.97%)	226 (13.45%)	< 0.001***
Back-related (yes)	448 (27.54%)	821 (39.74%)	< 0.001 ***	393 (27.48%)	687 (40.89%)	< 0.001***
Arthritis (yes)	130 (7.99%)	260 (12.58%)	< 0.001 ***	146 (10.21%)	251 (14.94%)	< 0.001***
PRO[a]-remote monitoring group						
monitoring	821(50.46%)	1,029 (49.81%)	0.692	720 (50.35%)	844 (50.24%)	0.950

no monitoring	806 (49.54%)	1,037 (50.19%)		710 (49.65%)	836 (49.76%)	
Mobilization after surgery						
within 6 hours	775 (47.63%)	931 (45.06%)	0.076	659 (46.08%)	766 (45.60%)	0.755
within 12 hours	453 (27.84%)	600 (29.04%)		404 (28.25%)	481 (28.63%)	
within 24 hours	367 (22.56%)	480 (23.23%)		320 (22.38%)	367 (21.85%)	
within 48 hours	27 (1.66%)	34 (1.65%)		39 (2.73%)	50 (2.98%)	
after 48 hours	5 (0.31%)	21 (1.02%)		8 (0.56%)	16 (0.95%)	

[a]PRO = patient-reported outcome

[b]significance levels were set at p<0.1*, p<0.05** and p<0.01***

Table A2: Summary statistics for PROs pre-surgery, at discharge and 12 motnhs post-surgery and as change for hip arthroplasty pateints per gender

Gender	**Male (n = 1,627)**			**Female (n =2,066)**			**Female vs. male difference of means (gender health gap)**		
Variable[a]	Mean	SD	Median	Mean	SD	Median	Points	%[b]	p-value[c]
Admission									
EQ-5D-5L	0.644	0.239	0.744	0.562	0.269	0.634	-0.081	-12.72	< 0.001***
EQ-VAS	59.14	19.90	60.00	54.96	19.67	54.00	-4.18	-7.07	< 0.001***
HOOS-PS	44.29	15.50	41.70	50.43	16	50.80	6.13	13.85	< 0.001***
PROMIS-D-SF	47.96	7.78	49.00	51.20	8.4	51.80	3.25	6.77	< 0.001***
PROMIS-F-SF	46.92	9.20	49.00	50.97	10	51.00	4.05	8.62	< 0.001***
Pain in joint	6.09	2.19	7.00	6.72	2.05	7.00	0.63	10.34	< 0.001***
Discharge									
EQ-5D-5L	0.780	0.170	0.816	0.755	0.189	0.800	-0.025	-3.18	< 0.001***

EQ-VAS	65.87	18.08	70.00	63.69	18.10	66.00	-2.18	-3.31	< 0.001***
HOOS-PS	43.02	16.49	41.70	48.90	17.77	50.80	5.88	13.66	< 0.001***
PROMIS-D-SF	47.13	7.15	41.00	49.63	7.79	49.00	2.50	5.30	< 0.001***
PROMIS-F-SF	46.92	8.53	48.60	50.07	9.36	48.60	3.15	6.72	< 0.001***
Pain in joint	3.35	2.08	3.00	3.62	2.18	3.50	0.27	7.91	< 0.001***
Month 12									
EQ-5D-5L	0.897	0.148	0.943	0.876	0.171	0.918	-0.021	-2.33	< 0.001***
EQ-VAS	76.11	17.40	80.00	74.84	17.80	80.00	-1.27	-1.67	0.009***
HOOS-PS	13.10	12.93	8. 80	14.99	14.05	12.57	1.89	14.43	< 0.001***
PROMIS-D-SF	46.10	6.93	41.00	47.79	7.60	45.44	1.70	3.68	< 0.001***
PROMIS-F-SF	44.25	8.13	46.00	45.51	8.84	46.00	1.26	2.85	< 0.001***
Pain in joint	0.91	1.60	0.00	0.95	1.69	0.00	0.04	4.13	0.585
Improvement (change from admission to month 12)									
EQ-5D-5L	0.252	0.246	0.187	0.313	0.275	0.250	0.061	42.20	< 0.001***
EQ-VAS	16.97	22.54	15.00	19.88	22.31	20.00	2.91	17.14	< 0.001***
HOOS-PS	-31.19	17.09	-30.40	-35.44	18.22	-35.90	-4.24	13.60	< 0.001***
PROMIS-D-SF	-1.86	7.44	0.00	-3.41	8.19	-1.70	-1.55	83.24	< 0.001***
PROMIS-F-SF	-2.68	8.60	-2.40	-5.46	9.63	-5.16	-2.79	104.06	< 0.001***
Pain in joint	-5.19	2.57	-5.00	-5.78	2.52	-6.00	-0.59	11.42	< 0.001***

[a] The score ranges for the PROMs are: -0.661 to 1.0 for the EQ-5D-5L, 0-100 for the EQ-VAS, HOOS-PS and KOOS-PS, 33.7-75.8 for the PROMIS-F-SF, 41-79.4 for the PROMIS-D-SF and 0-10 for pain. The EQ-5D-5L is reported with three digits after the decimal point due to the small score range

[b] Exploratory: The percentage shows the relation of the difference between men's and women's mean PROM-score compared to men's mean PROM-score per timepoint. For instance, for the EQ-5D-5L at admission (0.562-0.644)/0.644.

[c] Calculated based on the Mann-Whitney U test for the admission, discharge and month 12 values and with a two-sided t-test for the change from admission to month 12. Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01

Table A3: Summary statistics for PROs pre-surgery, at discharge, and 12 months post-surgery and as a change for knee arthroplasty patients per gender

Gender	Male (n = 1,430)			Female (n =1,680)			Female vs. male difference of means (gender health gap)		
Variable[a]	Mean	SD	Median	Mean	SD	Median	Points	%[b]	p-value[c]
Admission									
EQ-5D-5L	0.665	0.230	0.752	0.589	0.262	0.666	-0.076	-11.44	< 0.001***
EQ-VAS	61.09	19.29	62.00	55.91	18.92	53.50	-5.19	-8.49	< 0.001***
HOOS-PS	40.62	12.67	40.30	45.13	12.35	44.00	4.51	11.11	< 0.001***
PROMIS-D-SF	47.42	7.69	41.00	51.04	8.39	51.80	3.61	7.62	< 0.001***
PROMIS-F-SF	46.28	9.51	46.00	49.98	9.80	48.60	3.61	7.80	< 0.001***
Pain in joint	6.49	2.04	7.00	7.04	1.93	7.00	0.55	8.41	< 0.001***
Discharge									
EQ-5D-5L	0.754	0.192	0.805	0.724	0.214	0.788	-0.030	-4.04	< 0.001***
EQ-VAS	63.34	17.97	66.00	60.54	17.56	60.00	-2.80	-4.43	< 0.001***
KOOS-PS	43.46	11.52	42.00	45.67	12.07	44.00	2.21	5.09	< 0.001***
PROMIS-D-SF	46.81	7.20	41.00	49.98	8.12	51.80	3.17	6.77	< 0.001***
PROMIS-F-SF	47.11	8.84	48.60	50.50	8.989	48.60	3.39	7.19	< 0.001***
Pain in joint	4.34	2.18	4.00	4.80	2.26	5.00	0.46	10.61	< 0.001***
Month 12									
EQ-5D-5L	0.869	0.172	0.917	0.838	0.197	0.901	-0.030	-3.47	< 0.001***
EQ-VAS	73.30	17.29	79.00	70.47	18.31	75.00	-2.83	-3.86	< 0.001***
KOOS-PS	24.13	13.62	22.05	27.09	12.49	27.50	2.96	12.26	< 0.001***
PROMIS-D-SF	46.03	7.05	41.00	48.54	7.99	47.81	2.51	5.44	< 0.001***
PROMIS-F-SF	44.62	8.59	46.00	46.63	8.86	46.20	2.00	4.49	< 0.001***

Pain in joint	1.65	1.89	1.00	1.83	2.01	1.00	0.18	11.03	0.023**
Improvement (change from admission to month 12)									
EQ-5D-5L	0.204	0.240	0.146	0.250	0.266	0.187	0.046	22.58	< 0.001***
EQ-VAS	12.21	20.86	10.00	14.56	22.16	14.00	2.36	19.30	0.002***
KOOS-PS	-16.49	14.22	-16.40	-18.04	13.99	-17.10	-1.55	9.41	0.002***
PROMIS-D-SF	-1.39	7.39	0.00	-2.497	8.012	0.00	-1.11	79.52	< 0.001***
PROMIS-F-SF	-1.66	8.86	0.00	-3.266	9.21	-2.60	-1.61	96.72	< 0.001***
Pain in joint	-4.84	2.50	-5.00	-5.21	2.63	-5.00	-0.36	7.52	< 0.001***

[a] The score ranges for the PROMs are: -0.661 to 1.0 for the EQ-5D-5L, 0-100 for the EQ-VAS, HOOS-PS and KOOS-PS, 33.7-75.8 for the PROMIS-F-SF, 41-79.4 for the PROMIS-D-SF and 0-10 for pain. The EQ-5D-5L is reported with three digits after the decimal point due to the small score range

[b] Exploratory: The percentage shows the relation of the difference between men's and women's mean PROM-score compared to men's mean PROM-score per timepoint. For instance, for the EQ-5D-5L at admission (0.589-0.665)/0.665.

[c] Calculated based on the Mann-Whitney U test and with a two-sided t-test for the change from admission to month 12. Statistical significance markers: * $p<0.1$; ** $p<0.05$; *** $p<0.01$

3.3 What matters when interpreting patient-reported outcomes?—Adjusted minimal important differences for hip and knee arthroplasty

Irene Salvi

David Kuklinski

Justus Vogel

Anja Bischof

Alexander Geissler

Under review in the *Quality of Life Research*

3.3.1 Abstract

Background: As one of the main goals of hip and knee replacements is to improve patients' health-related quality of life (HRQoL), a meaningful evaluation of treatment success can be achieved by calculating minimal important differences (MIDs) for improvements in patient-reported outcome measures (PROMs). This study aims at providing MIDs adjusted for patient characteristics for EQ-5D-3L index score improvements after hip and knee replacements. It adds to existing literature by relying on a large national sample and precise clustering algorithms.

Methods: A retrospective observational study was conducted using the publicly available National Health Service (NHS) PROMs dataset for primary hip and knee replacements. We use information on 252,331 hip replacements and 279,668 knee replacements from all NHS-funded providers in England between 2013 and 2020. To identify the best pre-operative predictor for EQ-5D-3L improvements, multivariable linear regression was run. Based on regression results, clusters of patients were created. Finally, unstratified and stratified MIDs for EQ-5D-3L improvements were estimated using anchor-based receiver operator curves.

Results: The best pre-operative predictors for EQ-5D-3L index score were pre-operative EQ-5D-3L index scores and depression status. Stratifying patients led to increased precision in the MID calculations and showed that MIDs were larger for patients with worse pre-operative scores and for patients with depression, and were equally large for both genders. MIDs calculated for knee replacements were lower than those for hip replacements.

Conclusions: Our findings show the importance of adjusting MIDs for patients' characteristics and should be considered for quality-related choices and policy initiatives.

Keywords: Minimal important differences; Adjusted minimal important differences; Patient-reported outcomes; EQ-5D; Adjustment for patient characteristics; Hip replacement; Knee replacement

3.3.2 Background

Hip and knee replacements are two of the most frequently performed and effective orthopedic surgeries worldwide. They are usually performed once conservative treatment approaches have been tried and failed for end-stage osteoarthritis [1]. In 2019, roughly 101,384 hip replacements [2] and 108,506 knee replacements [3] were performed in the UK alone, and these numbers are set to increase over the next years [4].

As one of the main aims of hip and knee replacements is to improve patients' health-related quality of life (HRQoL) [5, 6], it is important to measure HRQoL improvement after surgery with suitable instruments. Being patients the best judges of their own HRQoL, patient-reported outcome measures (PROMs) have been proposed as suitable instruments to understand whether surgeries such as hip or knee replacements were successful in terms of HRQoL [7, 8].

Over the years, patients have reported some degree of dissatisfaction with hip and knee replacement results, ranging from 20% to 7% for hip replacements [9, 10] and 29% to 13% for knee replacements [10, 11]. One possible approach to offset negative surgery outcomes reported by patients and to set the right patients' expectations is to investigate PROM results. To measure outcomes from the perspective of the patient and to encourage measurable improvements, in April 2009 the English National Health Service (NHS) began to collect PROMs on a yearly basis from all NHS-funded providers [12].

However, statistical significance of a PROM score change does not necessarily mean that such change is also clinically relevant [13-15]. Clinicians and health policy makers need to understand how to identify and interpret clinically meaningful results to correctly use them for clinical decision making and health policy development [16]. One way to do so is to calculate minimal important differences (MIDs) of the score of interest. In this article, we adopt the definition provided by Guyatt et al. [17] of MID as "smallest difference in score in the domain of interest that patients perceive as important […], and that would lead the clinician to consider a change in the patient's management".

However, MIDs calculated for PROM scores of the average patient may lead to biased treatment evaluations [18]. By estimating MIDs adjusted for patient characteristics, it is possible to show a more realistic estimation of the percentage of successful hip and knee replacements [19].

Previous studies have estimated MIDs for the EQ-5D after hip or knee replacement [20, 21]. However, these studies are limited by sample size, lack of adjustment for patient characteristics, not fully suitable anchors, and/or imprecise clustering algorithms. Furthermore, to the best of our knowledge, no study has so far explored the impact of mental health and gender-related effects in MIDs. With this study, we aim at determining patient characteristics that influence post-surgery PROM score improvements and at estimating more precise meaningful MIDs based on these characteristics. According to the relevance indicated in the literature on correct interpretation of PROMs, we additionally explore potential gender-differences in the MID calculation [22]. We employ a larger sample size, a more reliable anchor set and more precise clustering algorithms with respect to existing studies.

3.3.3 Methods

Data source

We conducted a retrospective observational study using the publicly available National Health Service (NHS) PROMs dataset on hip and knee replacements, with data from all NHS-funded providers in England between 2013 and 2020 and a response rate of more than 60% [23]. We excluded datasets from the years 2009 to 2012 due to the absence of distinction between primary replacements and revisions and the presence of fewer variables. See Table A1 for a detailed description of the data access procedure.

The STROBE Statement guidelines for reporting observational studies was followed [24].

Variables

The dataset includes information on patient demographics, pre-operative health history, comorbidities, pre- and post-operative EQ-5D-3L (EQ-5D), Oxford Hip Score (OHS) and Oxford Knee Score (OKS), and other post-operative information. The post-operative follow-up time was determined by the NHS to be 6 months [25]. The EQ-5D is an internationally well-established generic PROM which takes into account five overall HRQoL dimensions, namely mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has three response levels, indicating no problems, some problems or severe problems. Its scores are aggregated into a single index ranging from -0.594 to 1 (1 being full health and 0 being the equivalent of death) for the UK population [26].

The OHS and OKS are joint-specific PROMs with values ranging from 0 (most severe symptoms) to 48 (least symptoms) designed to assess disability in patients undergoing hip replacement (knee replacement) [27, 28].

Data cleaning and final data set

We built two datasets, one for the regression model and one for the MID model, to maximize the informational content of the dataset. The first steps of data cleaning are the same in both models, while the second step differs according to the different variable needs. For an illustration, see Figure A1 in the Appendix.

Between April 01, 2013, and March 31, 2020, 297,806 patients received hip replacements, of which we excluded 16,560 observations with the "Revision Flag", as we are only interested in primary joint replacements. Of the remaining 281,246 primary hip replacements, 255,132 completed both the pre-operative and 6-month post-surgery follow-up EQ-5D questionnaire. We excluded 617 cases for which the pre-operative EQ-5D was already recorded as "full health" (corresponding to a score of 1.0), since no further score improvement was possible. At this point the data cleaning steps for the two models differ: for the regression model, from the remaining 254,515 hip replacements, 37,876 cases for which information on any of the regression variables was not available were eliminated, resulting in 216,639 cases.

For the MID model, 2,184 cases for which the anchor variable "Success" for the baseline analysis was not available were eliminated, leaving 252,331 cases. We performed data cleaning steps for the knee replacement sample analogue to the hip replacement sample, reaching a final data set of 245,711 observations for the regression model, and 279,668 for the MID model.

Before implementing the analyses, some variables in the dataset were transformed for a better representation in the model. See Table A2 in the Appendix for an overview of such changes.

Empirical methods and approach

To characterize the sample, descriptive statistics were performed. To estimate MIDs stratified by the factor best predicting the post-operative EQ-5D improvement, a three-step approach was followed for both hip and knee replacements. The first step was to run multivariable ordinary least squares regression models with the EQ-5D index score improvement as dependent variable. The aim of this exploratory analysis was to identify the pre-operative factor best predicting EQ-5D improvement which would then be used to stratify patients into more homogeneous groups for MID-estimation. The selection of the model variables was adapted with respect to the NHS methodology for the risk-adjustment of PROMs [29, 30], since some variables are not available in the NHS PROMs datasets available online. A forward-selection approach maximizing the adjusted R squared was implemented for variable selection. To understand which variable had the strongest impact on the EQ-5D score improvement, we calculated standardized coefficients for the model.

The second step of our approach was to cluster patients through the k-means algorithm according to the previously identified predicting factors. Additionally, a split according to the gender variable was implemented to test differences in MIDs between male and female patients.

As third step, we estimated the MIDs for the unstratified and stratified samples. MIDs have a long history dating back to their first mention in 1987 by Guyatt et. al. [31]. In 1989, Jaeschke et al. [32] further developed the concept and defined the MCID as "the

smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate […] a change in the patient's management". In this article, we adopt the definition by Guyatt et al. [17] of "smallest difference in score in the domain of interest that patients perceive as important […], and that would lead the clinician to consider a change in the patient's management". We focus here on the MID that needs to be achieved in order for the hip or knee replacement surgery to be considered a success.

The MIDs in this study were estimated with the anchor method exploiting a binary variable indicating the success or failure of the surgery as "anchor" to assess the meaningful improvement achieved in the EQ-5D score [33]. This is done by exploiting receiver operator curves (ROC) [13, 33], which are often used in medical research as they provide a good picture of the responsiveness of the outcome measure for clinical trial evaluations or for the identification of meaningful thresholds to indicate acceptable outcomes [34]. The advantage of the anchor-based approach is that change in the outcome measure score is linked to a meaningful external anchor that accounts for the patient's perspective [35]. In this way, the EQ-5D index score changes can be plotted on an ROC curve, which can then be used to determine thresholds that maximize sensitivity and specificity [36]. As the aim of the MID identification is to understand which EQ-5D improvement corresponds to a successful surgery, the anchor selected for the baseline analysis was one question collected during the recording of PROMs for hip and knee replacements, namely "Success" ("Overall, how are your problems now, compared to before your operation?"). Appendix, Table A3 shows that the majority of patients answer "Much better" to the question of the variable "Success", followed by "A little better". Since the Pearson correlation between the EQ-5D change and the variable "Success" is -0.28 for hip replacements and -0.32 for knee replacements, we conclude that it is suitable to be used as an anchor [37], once transformed into a binary variable (see Appendix, Table A2). As we are looking at elective hip and knee replacement, we want to identify the smallest change in measurement that signifies an important improvement for the patient [38]. For this reason, we consider a success those surgeries for which patients

reported their problems to be “Much better” or “A little better” with respect to before the surgery.

The question “Satisfaction” (“How would you describe the results of your operation?”), as well as two alternative binary transformations of the variable “Success” were employed as anchors in the sensitivity analyses to test the stability of the results to a change in the anchor. In this study, MIDs were identified using the minimization of the sum of squares of 1-sensitivity and 1-specificity [33]:

$$threshold = \min\{(1 - sensitivity)^2 + (1 - specificity)^2\}$$

The anchor “Success” was also used to estimate the MIDs for sub-groups created according to depression status and gender.

Finally, we compared the size and accuracy of the unstratified and stratified MIDs, as well as the share of patients achieving those thresholds. As a sensitivity analysis, we also calculate the MIDs using a distribution-based method consisting in the calculation of half of a standard deviation (half SD) for the unstratified sample and the subgroups stratified by the pre-operative EQ-5D index score. All statistical analyses were performed using the statistical software RStudio (2022.02.1 version) and STATA (18.0 version).

3.3.4 Results

Descriptive statistics

Table 3.6 displays the descriptive statistics for the MID model (see Table A4 and Table A5 in the Appendix for the descriptive statistics of the regression model and the gender-adjusted model).

Patients undergoing hip (knee) replacement were on average 64.9 (65.5) years old and mostly female (60.7% and 56.9%). The majority lived with their family (74.8% and 77.3%) and had symptoms for one to five years before the operation (69.7% and 52.3%). The most common comorbidity was arthritis (72.5% and 77.8%), followed by high blood pressure (38.1% and 44.7%). The mean pre-operative EQ-5D index

score for hip replacement patients was 0.350, i.e., lower than for knee replacement patients (0.417), but their mean post-operative EQ-5D index score was 0.801, i.e., higher compared to knee replacement patients (0.746). Accordingly, hip replacement patients had a higher mean score improvement (0.451) compared to knee replacement patients (0.329). The mean pre-operative OHS was 17.8, and the OKS was 19.1. As for the post-operative questionnaires, 95.8% (90.3%) of hip (knee) replacement patients answered the question “Overall, how are your problems now, compared to before your operation?” (variable “Success”) with “Much better” or “A little better”, 78.3% (63.5%) were satisfied with the results of the operation (variable “Satisfaction”), and 6% (7.8%) were readmitted (variable “Readmitted”).

Table 3.4: Descriptive statistics for hip and knee replacement samples

Descriptive Statistics Hip and Knee Replacement						
	Hip			Knee		
	Observations	**Mean/%**	**SD**	**Observations**	**Mean/%**	**SD**
Demographics						
Age	232,824			261,115		
20 to 29		0.0%	-		0.0%	-
30 to 39		0.1%	-		0.0%	-
40 to 49		1.7%	-		0.2%	-
50 to 59		12.5%	-		10.2%	-
60 to 69		33.7%	-		36.1%	-
70 to 79		39.1%	-		41.8%	-
80 to 89		12.8%	-		11.7%	-
90 to 120		0.0%	-		0.0%	-
Gender	232,824			261,115		
Female		60.7%	-		56.9%	-
Pre-operative living arrangements	248,941			275,433		
Living with partner/spouse/family/friends		74.8%	-		77.3%	-
Living alone		24.7%	-		22.2%	-

Living in a nursing home, hospital or other long term care home		0.1%	-		0.1%	-
Other		0.4%	-		0.3%	-
Pre-operative patient health history						
Pre-operative Assisted	250,411			277,443		
Yes		16.0%	-		15.9%	-
Pre-operative Symptom Period	250,162			277,265		
Less than 1 year		12.4%	-		4.6%	-
1 to 5 years		69.7%	-		52.3%	-
6 to 10 years		11.7%	-		22.1%	-
More than 10 years		6.2%	-		21.0%	-
Pre-operative previous surgery	250,514			277,728		
Yes		3.6%	-		3.9%	-
Pre-operative disability	238,548			268,054		
Yes		53.5%	-		51.3%	-
Comorbidities	252,331			279,668		
Arthritis		72.5%	-		77.8%	-
Cancer		5.4%	-		5.3%	-
Circulation		4.6%	-		5.8%	-
Depression		8.0%	-		8.9%	-
Diabetes		9.2%	-		12.6%	-

Heart disease		8.6%	-		9.4%	-
High blood pressure		38.1%	-		44.7%	-
Kidney disease		1.9%	-		2.0%	-
Liver disease		0.6%	-		0.6%	-
Lung disease		8.2%	-		9.1%	-
Nervous system		0.8%	-		1.0%	-
Stroke		1.3%	-		1.6%	-
EQ-5D index score						
Pre-op EQ-5D index score	252,331	0.35	0.32	279,668	0.417	0.309
[Median; Inter-quartile range]		[0.587; 0.101-0.691]			[0.516; 0.055-0.656]	
Post-op EQ-5D index score	252,331	0.801	0.24	279,668	0.746	0.249
[Median; Inter-quartile range]		[0.760; 0.689-1.000]			[0.848; 0.691-1.000]	
Post-op EQ-5D index score change	252,331	0.451	0.339	279,668	0.329	0.327
[Median; Inter-quartile range]		[0.309; 0.071-0.601]			[0.413; 0.204-0.736]	
Oxford Hip (Knee) Score						
Hip (Knee) replacement pre-op score	249,827	17.829	8.06	276,654	19.124	7.663
Hip (Knee) replacement post-op score	250,420	39.987	8.493	277,661	36.001	9.367

Hip (Knee) replacement post-op score change	247,955	22.141	9.866	274695	16.865	9.748
Post-operative questionnaire						
Post-op “Success”	252,331			279,668		
Yes		95.8%	-		90.3%	-
Post-op “Satisfied”	251,449			278,391		
Yes		78.3%	-		63.5%	-
Post-op Readmitted	251,551			278,636		
Yes		6.0%	-		7.8%	-

Except for a slight drop between 2017/18 and 2018/19 for hip replacement patients, the pre-operative EQ-5D index score distribution remained rather constant over the years for both hip and knee replacement patients and doesn't present any trend (see Figure A2 and A3 in the Appendix). The development of the post-operative EQ-5D score for hip and knee replacement patients (see Figure A4 and A5 in the Appendix) similarly does not present any trend or relevant change over the years. Likewise, the sample size for both patient samples does not present unwarranted fluctuations. This shows that differences over the years do not explain variance in the scores, allowing to safely aggregate the observations over the years into a unitary dataset for the analyses.

Regression results

Table 3.7 (Table A6) shows that the pre-operative factors explaining the largest share of hip (knee) replacement patients' post-operative EQ-5D index score improvement are the patient's pre-operative EQ-5D index score and previously diagnosed depression. Both variables are negatively associated with the subsequent EQ-5D index score improvement. In both groups of patients, the pre-operative EQ-5D index score has the highest prediction weight, with a standardized coefficient of -0.859 (-0.849) (compared with -0.101 (-0.100) from the next highest value).

Table 3.5: Regression results on 6-months EQ-5D-5L index score improvement for hip replacement sample

Hip Replacement Regression Results		
	Dependent variable:	
	EQ5D Index Change	
	(1)	*(2)*
	Multivariate OLS	*Standardized multivariate OLS*
Demographics		
Age 30 to 39	-0.090 (0.062)	-0.009 (0.062)
Age 40 to 49	-0.068 (0.060)	-0.026 (0.060)
Age 50 to 59	-0.053 (0.060)	-0.051 (0.060)
Age 60 to 69	-0.030 (0.060)	-0.042 (0.060)
Age 70 to 79	-0.028 (0.060)	-0.040 (0.060)
Age 80 to 89	-0.038 (0.060)	-0.037 (0.060)
Age 90 to 120	-0.078 (0.064)	-0.005 (0.064)
Female	-0.005*** (0.001)	-0.007*** (0.001)
Pre-operative Living Arrangements	-0.018*** (0.001)	-0.024*** (0.001)
Pre-operative patient health history		
Pre-operative Assisted	-0.015*** (0.001)	-0.031*** (0.001)
Pre-operative Symptom Period	0.020*** (0.003)	0.011*** (0.003)

Pre-operative Previous Surgery	0.061*** (0.001)	0.090*** (0.001)
Pre-operative Disability	0.004*** (0.001)	0.008*** (0.001)
Comorbidities		
Arthritis	-0.026*** (0.001)	-0.033*** (0.001)
Cancer	-0.010*** (0.002)	-0.007*** (0.002)
Circulation	-0.076*** (0.002)	-0.047*** (0.002)
Depression	-0.126*** (0.002)	-0.101*** (0.002)
Diabetes	-0.026*** (0.002)	-0.023*** (0.002)
Heart disease	-0.026*** (0.002)	-0.022*** (0.002)
High blood pressure	-0.007*** (0.001)	-0.010*** (0.001)
Kidney disease	-0.011*** (0.003)	-0.004 (0.003)
Liver disease	0.004 (0.006)	0.001 (0.006)
Lung disease	-0.033*** (0.002)	-0.027*** (0.002)
Nervous system	-0.059*** (0.005)	-0.016*** (0.005)
Stroke	-0.024*** (0.004)	-0.008** (0.004)
EQ-5D-3L Index		
Pre-operative EQ-5D Index	-0.914*** (0.002)	-0.860*** (0.002)
Oxford Hip Score		
Hip Replacement Pre-operative Pain	-0.012*** (0.001)	-0.023*** (0.001)
Hip Replacement Pre-operative Sudden Pain	0.001** (0.0005)	0.004*** (0.0005)
Hip Replacement Pre-operative Night Pain	0.005*** (0.001)	0.016*** (0.001)
Hip Replacement Pre-operative Washing	0.010*** (0.001)	0.029*** (0.001)

Hip Replacement Pre-operative Transport	0.0003 (0.001)	0.001 (0.001)
Hip Replacement Pre-operative Dressing	-0.002*** (0.001)	-0.005*** (0.001)
Hip Replacement Pre-operative Shopping	0.016*** (0.001)	0.059*** (0.001)
Hip Replacement Pre-operative Walking	0.007*** (0.001)	0.023*** (0.001)
Hip Replacement Pre-operative Limping	-0.029*** (0.001)	-0.077*** (0.001)
Hip Replacement Pre-operative Stairs	0.017*** (0.001)	0.046*** (0.001)
Hip Replacement Pre-operative Standing	0.001 (0.001)	0.002** (0.001)
Hip Replacement Pre-operative Work	0.008*** (0.001)	0.020*** (0.001)
Constant	0.685*** (0.060)	(0.06)
Observations	216,639	216,639
R^2	0.593	0.593
Adjusted R^2	0.593	0.593

Legend: *p<0.1; **p<0.05; ***p<0.01; standard deviation in brackets

MIDs

Since the pre-operative EQ-5D index score is the strongest predictor of the subsequent EQ-5D index score improvement for both hip and knee replacement patients, sub-groups of patients were created according to their pre-operative EQ-5D index scores. Table 3.8 reports the results of the MID analysis using the baseline anchor "Success". Sensitivity analyses exploring the alternative anchor "Satisfaction" (Table A7), as well as two alternative binary transformations of the anchor "Success" (Table A8) are shown in the Appendix. It is clear from the area under the curve (AUC) that stratifying the sample of patients based on the pre-operative EQ-5D index score leads to increased precision in the MID calculations. The estimated MID value for hip (knee) replacement patients for the unstratified sample was 0.185 (0.114) for both hip and knee replacement patients, and, regardless of the anchor used, the MIDs varied among the different pre-operative EQ-5D index score subgroups. The MID values ranged from 0.661 (0.567) for the patients in the subgroups with the lowest pre-operative EQ-5D index score (-0.594 to -0.055 for hip replacement patients and -0.594 to 0.03 for knee replacement patients) to 0.045 (0.012) for patients with the highest pre-operative EQ-5D index score (0.656 to 0.883 for both hip and knee replacement patients). Generally, patients starting with worse pre-operative scores need larger improvements to reach their MID.

Table 3.6: Unstratified and stratified MIDs for hip and knee replacement according to "pre-operative EQ-5D index score"

MIDs - Hip replacement				**MIDs - Knee replacement**			
Anchor: Success	**Observations**	**AUC**	**MIDs**	**Anchor: Success**	**Observations**	**AUC**	**MIDs**
Unstratified	252,331	0.784	0.185	**Unstratified**	290,763	0.751	0.114
Pre-operative EQ-5D				**Pre-operative EQ-5D**			
-0.594 to -0.055	25,015	0.849	0.661	-0.594 to 0.03	40,109	0.821	0.567
-0.05 to 0.109	67,845	0.88	0.61	0.05 to 0.213	69,321	0.862	0.542
0.118 to 0.383	29,002	0.875	0.525	0.219 to 0.433	8,732	0.839	0.405
0.414 to 0.639	65,876	0.886	0.114	0.452 to 0.639	53,888	0.834	0.081
0.656 to 0.883	64,593	0.9009	0.045	0.656 to 0.883	107,618	0.876	0.012

Legend: AUC = area under the curve; MID = minimal important difference

The same trend is true when using the anchor “Satisfaction”, but the precision of the estimates is reduced compared to using the anchor “Success”. Compared to the MIDs calculated with the anchor “Success”, the MIDs calculated using the alternative anchor “Satisfaction” show consistently higher values (see Table A7). These are similar to the MIDs calculated using the anchor “Success” but defined in a more conservative way, i.e., only considering the reply “Much better” to the question “Overall, how are your problems now, compared to before your operation?” as indicative of a successful surgery (see Table A8). On the other hand, MIDs calculated only on the basis of the answer “A little better” to the “Success” variable, while reaching much lower accuracy, are consistently lower.

The MIDs calculated for knee replacement patients were consistently lower than those calculated for hip replacement patients for both unstratified and stratified samples.

Figure 3.8 and Figure 3.9 show that without adjusting for pre-operative EQ-5D index score the probability that a hip or knee replacement is considered a success is higher for the patients group starting with worse EQ-5D-3L scores. The unstratified MID of 0.185 (0.114) would only be achieved by 61% (48%) of hip (knee) replacement patients belonging to the subgroup with highest pre-operative EQ-5D-3L index score (0.656 to 0.883), while 92% (88%) of the patients in the lowest pre-operative EQ-5D-3L index score group would achieve it. When adjusting for patients’ characteristics by using subgroup specific MIDs, instead, a more constant share of patients across all subgroups achieves their stratified MID thresholds.

Figure 3.8: MIDs for hip replacement patients, unstratified and stratified by pre-operative EQ-5D index score—impact on the patient sample

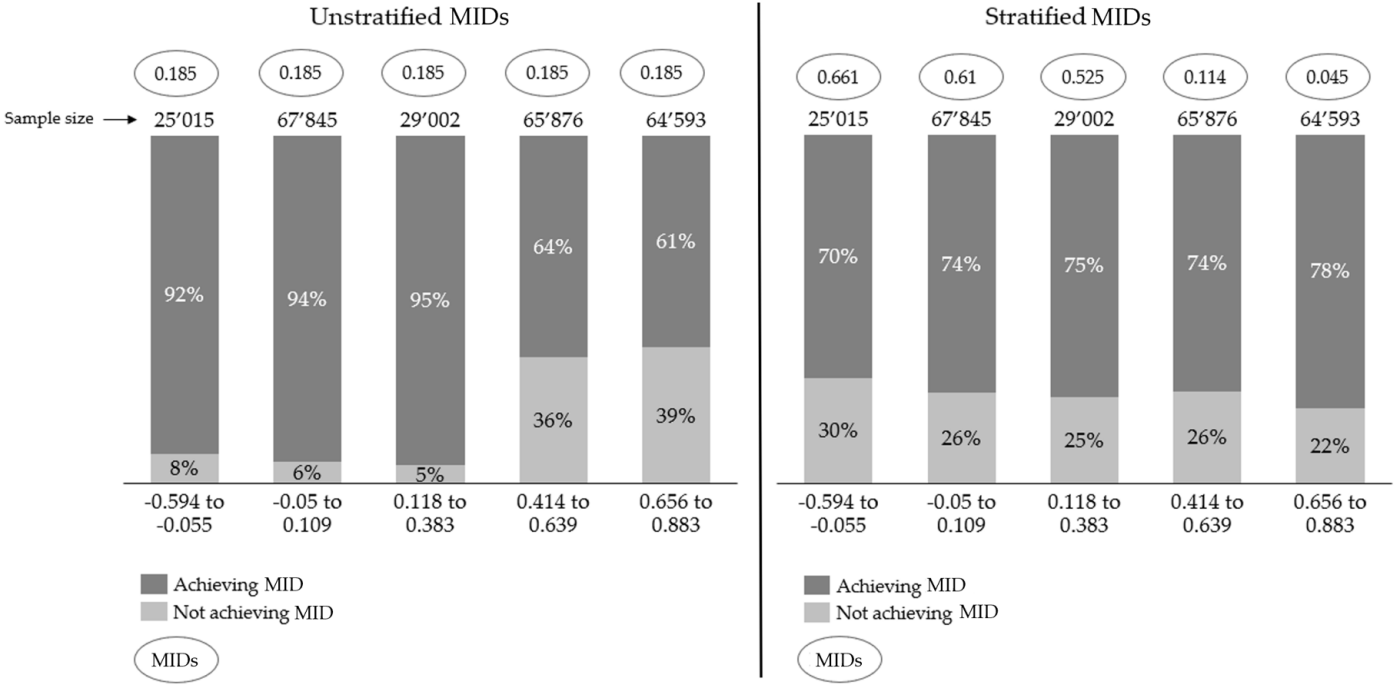
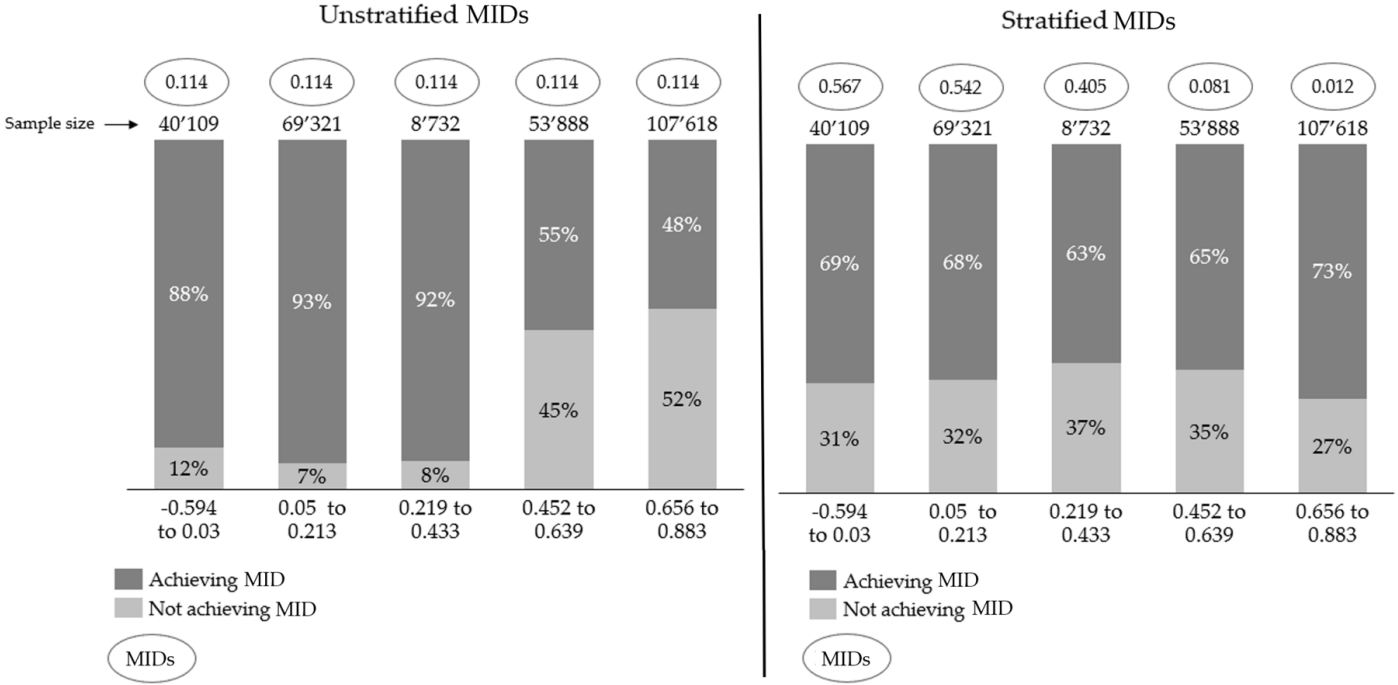


Figure 3.9: MIDs for knee replacement patients, unstratified and stratified by pre-operative EQ-5D index score—impact on the patient sample



By looking at the results of the sensitivity analysis using the half SD method (Table A11), we note that the unstratified MID calculated with the half SD method is close to the one calculated through the anchor-based method, however it is lower for hip replacements and higher for knee replacements. On the other hand, estimates for the stratified MIDs are consistently lower for the lower pre-operative EQ-5D index score subgroups, and higher for the higher pre-operative EQ-5D index score subgroups.

Additional MID stratifications

Table 3.9 shows additional stratifications of the MIDs according to different patients' characteristics.

As the second most important predicting factor of EQ-5D index score improvement, we first explore the stratification of patients according to the variable "Depression" (i.e., having been told by a doctor to have depression), and according to "Depression" and "Pre-operative EQ-5D" clustered together. The sub-group of patients with depression shows higher MIDs for hip replacements (0.215 compared to 0.185) but not for knee replacements (0.114). The accuracy of the estimation decreases in the sub-group of patients with depression in both hip and knee replacement samples, which make up smaller sample sizes than the non-depression groups. On the other hand, the accuracy of the MID estimates reached in the non-depressed sample is increased with respect to the baseline model. Once additionally stratifying for the pre-operative EQ-5D score, the MIDs calculated for patients with depression are lower than the ones calculated for patients without depression in almost all pre-operative score sub-groups. However, one can notice how there is a relatively higher proportion of depressed patients among the lower pre-operative EQ-5D score sub-groups than for non-depressed patients.

We also present the results of the MID calculations stratified by gender and by gender and pre-operative EQ-5D index score, to understand whether the MIDs differ between male and female patients. We observe that the MID threshold for perceiving the intervention as a success is the same for both genders before the combined stratification with the pre-operative EQ-5D index score, meaning that the minimal

post-operative EQ-5D improvement representing a successful intervention was on average the same for female and male patients. The same result is found when stratifying by both gender and pre-operative EQ-5D index score, with minor differences between the MIDs of female and male patients across the pre-operative EQ-5D index score groups.

Table 3.7: Unstratified and stratified MIDs for hip and knee replacement according to "pre-operative EQ-5D index score", "depression", and "gender"

MIDs - Hip replacement															
Subgroups	**-**			**Depression**			**No depression**			**Female**			**Male**		
Anchor: Success	**Obs.**	**AUC**	**MIDs**	**Obs.**	**AUC**	**MIDs**	**Obs.**	**AUC**	**MIDs**	**Obs.**	**AUC**	**MIDs**	**Obs.**	**AUC**	**MIDs**
Unstratified	252,331	0.784	0.185	20,285	0.765	0.215	232,046	0.787	0.185	141,132	0.782	0.185	91,676	0.787	0.185
Pre-operative EQ-5D															
-0.594 to -0.055	25,015	0.849	0.661	5,868	0.809	0.438	19,147	0.858	0.71	15,550	0.847	0.652	7,429	0.853	0.661
-0.05 to 0.109	67,845	0.88	0.61	6,181	0.841	0.542	61,664	0.882	0.617	41,443	0.875	0.617	21,290	0.891	0.61
0.118 to 0.383	29,002	0.875	0.525	1,437	0.806	0.375	27,565	0.878	0.542	16,461	0.87	0.544	10,352	0.884	0.471
0.414 to 0.639	65,876	0.886	0.114	5,445	0.845	0.043	60,431	0.889	0.114	34,681	0.879	0.114	26,102	0.897	0.114
0.656 to 0.883	64,593	0.9009	0.045	1,354	0.887	0.01	63,239	0.901	0.045	32,997	0.902	0.045	26,503	0.908	0.046
MIDs - Knee replacement															
Subgroups	**-**			**Depression**			**No depression**			**Female**			**Male**		
Anchor: Success	**Obs.**	**AUC**	**MIDs**	**Obs.**	**AUC**	**MIDs**	**Obs.**	**AUC**	**MIDs**	**Obs.**	**AUC**	**MIDs**	**Obs.**	**AUC**	**MIDs**
Unstratified	290,763	0.751	0.114	25,001	0.739	0.114	254,667	0.753	0.114	148,443	0.747	0.114	112,650	0.757	0.114
Pre-operative EQ-5D															
-0.594 to 0.03	40,109	0.821	0.567	9,688	0.795	0.39	30,421	0.826	0.603	24,571	0.812	0.558	12,538	0.833	0.603
0.05 to 0.213	69,321	0.862	0.542	5,135	0.835	0.444	64,186	0.865	0.542	40,317	0.858	0.542	24,310	0.87	0.542
0.219 to 0.433	8,732	0.839	0.405	482	0.791	0.337	8,250	0.842	0.405	4,012	0.809	0.402	4,060	0.863	0.37

0.452 to 0.639	53,888	0.834	0.081	7,264	0.815	0.01	46,624	0.837	0.081	29,211	0.829	0.081	21,300	0.843	0.081
0.656 to 0.883	107,618	0.876	0.012	2,432	0.84	0.043	105,186	0.877	0.012	50,332	0.874	0.012	50,442	0.881	0.043

Legend: AUC = area under the curve; MID = minimal important difference; Obs. = Observations

3.3.5 Discussion

With this study, we aimed at filling the research gap existing in the literature on MIDs by estimating thresholds for the EQ-5D index score reflecting meaningful improvements in the HRQoL of patients following hip or knee replacement, with some crucial innovations with respect to previous studies. Firstly, we leveraged the large sample size from the nationally representative NHS PROMs dataset. Secondly, the availability of multiple external variables allowed us to not only calculate accurate MIDs but also to check their robustness with respect to possible alternative anchors. In addition, the patient stratification was conducted through a machine-learning clustering algorithm considering the density distribution of the pre-operative EQ-5D index score, allowing to identify meaningful patient clusters.

Our findings show that the improvements in the post-operative EQ-5D index score are influenced by various patients' pre-operative characteristics, such as pre-operative PROM score values and depression status, and are best predicted by the pre-operative EQ-5D index score. Moreover, we show that stratifying the sample based on patients' pre-operative EQ-5D index score leads to increased accuracy in the MID calculations, and that patients starting with worse pre-operative scores need a larger improvement in the post-operative EQ-5D index score for the surgery to be considered successful. This is in line with literature showing that low pre-operative PROM scores are associated with high pre-operative expectations [39]. Furthermore, while MIDs based on some PROMs have been shown in the literature to present no correlation with baseline severity, MIDs based on the EQ-5D show correlation in the case of hip and knee replacements [40]. In line with findings from the literature [18], we also show how, without adjustment for patients' characteristics, the probability that a surgery is considered a success is disproportionately higher for the patients group starting with worse EQ-5D scores. Instead, when adjusting for patients' characteristics by using subgroup specific MIDs, a more constant share of patients across all subgroups achieves their stratified MID thresholds, reflecting a more realistic evaluation of surgery successes.

These findings show that the use of unstratified thresholds to evaluate treatment outcomes or detect critical recovery paths are not advisable and would lead to an inaccurate assessment of hip and knee replacement successes and adverse events. MID values calculated based on patients' EQ-5D index score at admission, instead, provide thresholds that are better tailored to individual patients.

In addition, we show that stratifying patients by both pre-operative EQ-5D index score and depression status leads to increased accuracy in the estimation of the sample of non-depressed patients; that depressed patients exhibit lower MIDs than non-depressed patients; and that depressed patients make up a relatively larger share of cases with lower pre-operative EQ-5D index scores. We also find that the threshold for perceiving the improvement as a success is the same for both genders. This leads to the conclusion that, on average, male and female patients expect a similar absolute improvement of their HRQoL status post-surgery. However, since there is a higher proportion of female patients starting with a lower baseline score our results suggest that female patients tend to rate the intervention as a success with a lower post-operative HRQoL compared to male patients. This implies that, while gender differences need to be considered when assessing patients' current health status, MID thresholds can reasonably be calculated independently of gender.

Findings from other studies

While several studies attempted to estimate (unstratified) MIDs based on PROMs for hip and knee replacement patients, and despite several studies emphasizing that the share of patients achieving an unstratified MID depends on their admission score [41-43], literature on thresholds adjusted for patient characteristics is relatively limited [44]. To the best of our knowledge, no study has so far estimated MIDs for hip and knee replacement patients clustered according to both depression status and pre-operative PROMs, and according to gender and pre-operative PROMs. Davis et al. [45] showed that unstratified thresholds to assess treatment outcome after orthopedic surgery led to an unfair assessment of successful outcomes, depending on pre-operative PROM scores. Stratford et al. [46] concluded that calculating PROM thresholds based on pre-operative PROM scores significantly improved the accuracy

of estimated thresholds. However, the studies by Paulsen et al. [21] and Kuklinski et al. [19] are the only ones estimating MCID values based on subgroups clustered by pre-operative PROM scores. Through a distribution-based method, Paulsen et al. [21] showed that MCID values vary significantly among subgroups of patients with different pre-operative PROM scores (lowest tertile: 0.67; middle tertile: 0.34; highest tertile: 0.23). Through an anchor-based method, Kuklinski et al. [19] showed that thresholds for meaningful improvement in PROM scores need to be adjusted to patient characteristics and that MCID thresholds for hip replacement at 3-months (12-months) follow-up are 0.637 (0.889) for the lowest pre-operative EQ-5D-5L group, and 0.001 (0.113) for the highest pre-operative score group.

Comparison with the distribution-based method sensitivity analysis results show that the anchor-based MID achieves a better discrimination performance for stratified MIDs. Our results for unstratified and stratified MID thresholds are in line and consistent with the results from the literature.

Limitations

Since PROMs information comes from survey data, our dataset may suffer from a problem of responder bias, as patients with extremely poor recoveries may not be able to answer the questionnaire, even if assisted. However, the English 'best practice tariff' pay-for-performance scheme creates an incentive for providers to meet the minimum standard for data collection [47, 48].

Another limitation lies in the fact that MID estimations varied depending on the anchor used. Anchor-based methods have also been criticized for the effect of recall bias on long-term responsiveness and for their inability to include the measurement precision of the global instrument [13]. However, studies have shown that the EQ-5D appears to be largely free of recall bias [49, 50].

A final source of bias may also derive from the correlation between fulfillment of expectations and patient satisfaction [39]. A systematic difference in expectations between the clusters might bias our MIDs when calculated using the anchor "Satisfaction".

3.3.6 Conclusion

Our research shows that when evaluating PROM results by using MIDs, they should be adjusted for patients' characteristics, and especially for pre-operative PROM scores, as patients starting with worse pre-operative scores need larger improvements for surgeries to be considered successful. By doing so, a more accurate evaluation of surgery successes can be achieved.

We envision our findings to be used as inputs for the refinement of clinical decision support systems, thus benefitting several groups of stakeholders in the healthcare system. Physicians could take our findings into consideration when monitoring patients' recovery paths to enable a reliable feedback cycle and treatment path adaption. When using PROMs for quality monitoring and policy design, health policy makers might consider our findings to assess the quality of hospital treatments independently from patient selection bias. Our findings can also guide health insurers in the development of pay-for-performance contracts based on surgery success evaluated through PROMs. Likewise, patients could benefit from exposure to these findings for better hospital choices based on expected surgery success.

Future research should aim at providing MID thresholds that are adjusted for patients' characteristics and as precise as possible. Furthermore, future research would benefit from the exploration of MIDs for disease-specific instruments in addition to MIDs for generic instruments.

References

1. Deakin AH, Smith MA, Wallace DT, et al (2019) Fulfilment of preoperative expectations and postoperative patient satisfaction after total knee replacement. A prospective analysis of 200 patients. Knee 26:1403–1412.

2. National joint registry. Hips – all procedures - activity. https://reports.njrcentre.org.uk/hips-all-procedures-activity. Accessed 29 Jul 2022

3. National Joint registry. Knees - all procedures - activity. https://reports.njrcentre.org.uk/knees-all-procedures-activity. Accessed 29 Jul 2022

4. OECD (2021) Health at a Glance 2021: OECD Indicators. OECD Publishing, Paris.

5. Ray G, Ekelund P, Nemes S, et al (2019) Changes in health-related quality of life are associated with patient satisfaction following total hip replacement: an analysis of 69,083 patients in the Swedish Hip. Taylor & Francis 91:48–52.

6. Norman-Taylor FH, Palmer CR, Villar RN (1996) Quality-of-life improvement compared after hip and knee replacement. Journal of Bone and Joint Surgery - Series B 78:74–77.

7. Ney JP, Taylor LP (2019) Patients are the best judges. Neurol Clin Pract 9:7–8.

8. Padilla G v, Rhiner M, Bogen C (1992) Health Quality of Life and Colorectal Cancer.

9. Shao Z, Bi S (2022) Patient satisfaction after total hip arthroplasty: Influencing factors. Front Surg 9:.

10. Sabah SA, Knight R, Alvand A, et al (2022) Early patient-reported outcomes from primary hip and knee arthroplasty have improved over the past seven years : an analysis of the NHS PROMs dataset. Bone Joint J 104-B:687–695.

11. Choi Y-J, Ra HJ (2016) Patient Satisfaction after Total Knee Arthroplasty. Knee Surg Relat Res 28:1–15.

12. Statistics - Patient Reported Outcome Measures (PROMs). https://www.england.nhs.uk/statistics/statistical-work-areas/proms/. Accessed 29 Jul 2022

13. Wright A, Hannon J, Hegedus EJ, Kavchak AE (2012) Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). J Man Manip Ther 20:160–166.

14. Page P (2014) Beyond statistical significance: Clinical interpretation of rehabilitation research literature. Int J Sports Phys Ther 9:726-736.

15. Batterham AM, Hopkins WG (2006) Making meaningful inferences about magnitudes. Int J Sports Physiol Perform 1:50–57.

16. Kristensen N, Nymann C, Konradsen H (2016) Implementing research results in clinical practice- the experiences of healthcare professionals. BMC Health Serv Res 16:.

17. Guyatt GH, Osoba D, Wu AW, et al (2002) Methods to Explain the Clinical Significance of Health Status Measures. Mayo Clin Proc 77:371–383.

18. Farrow L, Redmore J, Talukdar P, et al (2022) Prioritisation of patients awaiting hip and knee arthroplasty: Lower pre-operative EQ-5D is associated with greater improvement in quality of life and joint function. Musculoskeletal Care.

19. Kuklinski D, Marques CJ, Bohlen K, et al (2022) Thresholds for meaningful improvement in WOMAC scores need to be adjusted to patient characteristics after hip and knee replacement. J Orthop 29:50–59.

20. Giesinger JM, Hamilton DF, Jost B, et al (2015) WOMAC, EQ-5D and Knee Society Score Thresholds for Treatment Success After Total Knee Arthroplasty. J Arthroplasty 30:2154–2158.

21. Paulsen A, Roos EM, Pedersen AB, Overgaard S (2014) Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. Acta Orthop 85:39–48.

22. Hertler C, Seiler A, Gramatzki D, et al (2020) Sex-specific and gender-specific aspects in patient-reported outcomes. ESMO Open 5(Suppl 4): e000837.

23. Finalised Patient Reported Outcome Measures (PROMs) in England for Hip and Knee Replacement Procedures (April 2019 to March 2020) - NHS Digital.https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/finalised-hip-and-knee-replacement-april-2019---march-2020. Accessed 3 Nov 2023

24. Von Elm E, Altman DG, Egger M, et al (2009) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. J Clin Epidemiol 61:344-9.

25. National Health Service (NHS) Digital (2022) Background information about PROMs. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms/background-information-about-proms. Accessed 27 Jul 2022

26. Dolan P, Gudex C, Kind P, Williams A (1995) A social tariff for EuroQol: results from a UK general population survey

27. Dawson J, Fitzpatrick R, Carr A, Murray D (1996) Questionnaire on the perceptions of patients about total hip replacement. Journal of Bone and Joint Surgery - Series B 78:185–190.

28. Dawson J, Fitzpatrick R, Murray D, Carr A (1998) Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg Br 80:63–69.

29. Coles J Risk Adjustment models for Patient Reported Outcome Measures

30. DH (2012) Patient Reported Outcome Measures (PROMs) in England: The case-mix adjustment methodology

31. Guyatt G, Walter S, Norman G (1987) Measuring change over time: Assessing the usefulness of evaluative instruments. J Chronic Dis 40:171–178.

32. Jaeschke R, Singer J, Guyatt GH (1989) Measurement of health status: Ascertaining the minimal clinically important difference. Control Clin Trials 10:407–415.

33. Froud R, Abel G (2014) Using ROC Curves to Choose Minimally Important Change Thresholds when Sensitivity and Specificity Are Valued Equally: The Forgotten Lesson of Pythagoras. Theoretical Considerations and an Example Application of Change in Health Status. PLoS One 9:.

34. de Vet HCW, Bouter LM, Dick Bezemer P, Beurskens AJHM (2001) Reproducibility and responsiveness of evaluative outcome measures. International Journal of Technology Assessment in Health Care 17:479–487.

35. Crosby RD, Kolotkin RL, Williams GR (2003) Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 56:395–407.

36. Deyo RA, Centor RM (1986) Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. J Chronic Dis 39:897–906.

37. King MT (2011) A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 11:171-84.

38. Kvien TK, Heiberg T, Hagen KB (2007) Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? Ann Rheum Dis 66:iii40–iii41.

39. Scott CEH, Bugler KE, Clement ND, et al (2012) Patient expectations of arthroplasty of the hip and knee. Journal of Bone and Joint Surgery - Series B 94 B:974–981.

40. Browne JP, van der Meulen JH, Lewsey JD, et al (2010) Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol 63:865–874.

41. Berliner JL, Brodke DJ, Chan V, et al (2017) Can Preoperative Patient-reported Outcome Measures Be Used to Predict Meaningful Improvement in Function After TKA? Clin Orthop Relat Res 475:149–157.

42. Berliner JL, Brodke DJ, Chan V, et al (2016) John Charnley Award: Preoperative Patient-reported Outcome Measures Predict Clinically Meaningful Improvement in Function After THA. Clin Orthop Relat Res 474:321–329.

43. Most J, Hoelen TCA, Spekenbrink-Spooren A, et al (2022) Defining Clinically Meaningful Thresholds for Patient-Reported Outcomes in Knee Arthroplasty. J Arthroplasty 37:837-844.e3.

44. Rouquette A, Blanchin M, Sébille V, et al (2014) The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline score. J Clin Epidemiol 67:433–440.

45. Davis AM, Perruccio AV, Lohmander LS (2012) Minimally clinically important improvement: all non-responders are not really non-responders an illustration from total knee replacement. Osteoarthritis Cartilage 20:364–367.

46. Stratford PW, Binkley JM, Riddle DL, Guyatt GH (1998) Sensitivity to Change of the Roland-Morris Back Pain Questionnaire: Part 1. Phys Ther 78:1186–1196.

47. Gomes M, Gutacker N, Bojke C, Street A (2016) Addressing Missing Data in Patient-Reported Outcome Measures (PROMS): Implications for the Use of PROMS for Comparing Provider Performance. Health Econ 25:515–528.

48. NHS (2021) 2021/22 National tariff payment system. NHS England, London.

49. Olerud P, Tidermark J, Ponzer S, et al (2011) Responsiveness of the EQ-5D in patients with proximal humeral fractures. J Shoulder Elbow Surg 20:1200–1206.

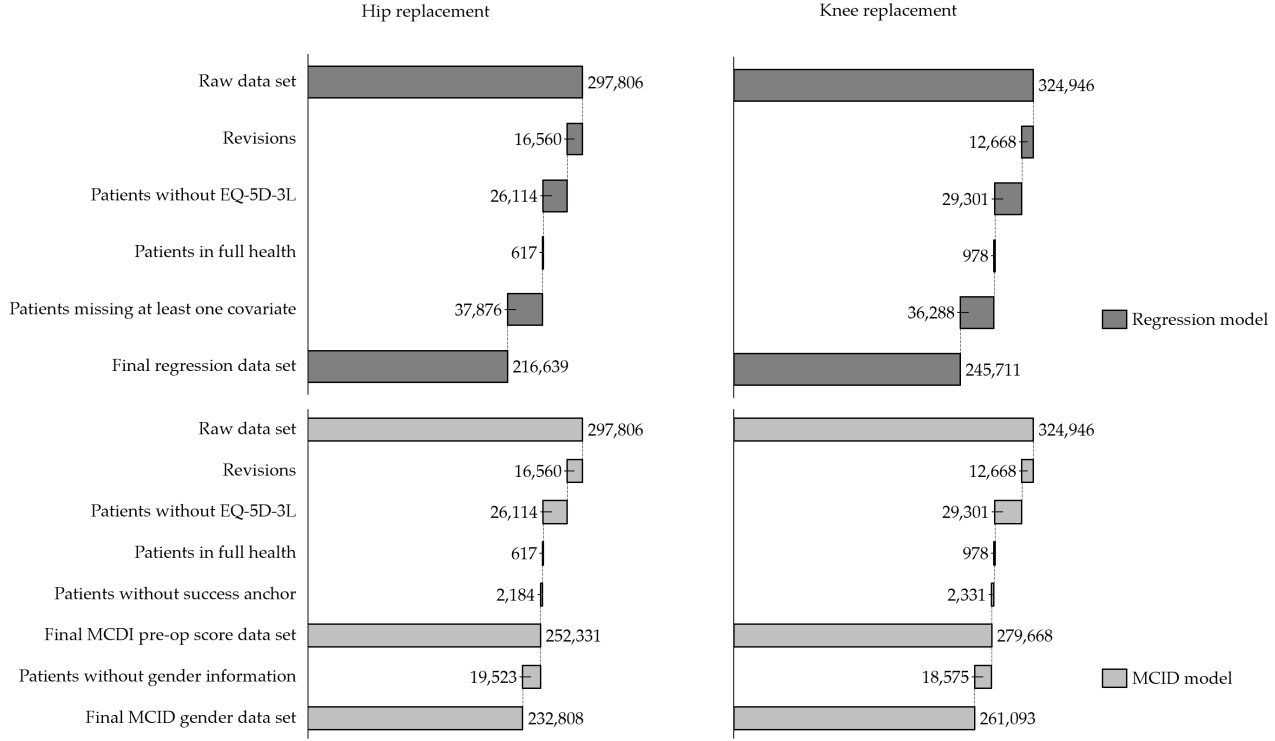
50. McCann AC, Phillips KM, Trope M, et al (2020) Characterising the potential for recall bias in anchor-based MCID calculation of patient-reported outcome measures for chronic rhinosinusitis. Clinical Otolaryngology 45:768–774.

Appendix

Table A1: Description of data access procedure

Step	Procedure
1	Visit the main website: https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms
2	From the main website, select "Finalised Patient Reported Outcome Measures (PROMs) in England for Hip & Knee Replacements, April 20XX – March 20XX" links for each year (in our case from 2013 to 2020). For example, for 2019 to 2020 the link is: https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/finalised-hip-and-knee-replacement-april-2019---march-2020
3	Under “Resources”, download the "CSV Data Pack" files for each year
4	Within the downloaded folder, select "Record Level Hip Replacement XXXX" and "Record Level Knee Replacement XXXX" (2013 to 2015), or "Hip Replacement XXXX" and "Knee Replacement XXXX" (2015-2016), or "Hip Replacement Provider XXXX" and "Knee Replacement Provider XXXX" (2016-2020)
5	Merge the selected files in one dataset from 2013 to 2020.

Figure A1: Data cleaning process



Note: "Patients without EQ-5D-3L" indicates observations missing either pre-operative or post-operative EQ-5D-3L questionnaire. "Patients in full health" indicates observations for which the EQ-5D-3L index score = 1.

Table A2: Variable transformation

Variable name	NHS dataset variables	Recoded variables	Recoded variable "Success" for sensitivity analysis	
Gender	1= Male	0 = Male		
	2 = Female	1 = Female		
Comorbidities	1= Yes	1 = Yes		
	9 = Missing	0 = No		
Success	1 = Much better	Yes	Yes	
	2 = A little better		No	Yes
	3 = About the same	No		No
	4 = A little worse			
	5 = Much worse			
Satisfaction	1 = Excellent	Yes		
	2 = Very Good			
	3 = Good	No		
	4 = Fair			
	5 = Poor			

Table A3: Percentage of answers to the variable "Success"

Descriptive Statistics Hip (Knee) Replacement						
	Hip			Knee		
	Observations	**Mean**	**SD**	**Observations**	**Mean**	**SD**
Variable "Success"	252331			279,668		
Much better		0.878	-		0.749	-
A little better		0.08	-		0.154	-
About the same		0.021	-		0.043	-
A little worse		0.012	-		0.033	-
Much worse		0.009	-		0.021	-

Table A4: Descriptive statistics for hip and knee replacement samples – Gender-stratified MIDs samples

Descriptive Statistics Hip and Knee Replacement						
	Hip			Knee		
	Observations	**Mean/%**	**SD**	**Observations**	**Mean/%**	**SD**
Demographics						
Age	232,808			261,093		
20 to 29		0.0%	-		0.0%	-
30 to 39		0.1%	-		0.0%	-
40 to 49		1.7%	-		0.2%	-
50 to 59		12.5%	-		10.2%	-
60 to 69		33.7%	-		36.1%	-
70 to 79		39.1%	-		41.8%	-
80 to 89		12.8%	-		11.7%	-
90 to 120		0.0%	-		0.0%	-
Gender	232,808			261,093		
Female		60.6%	-		56.9%	-
Pre-operative living arrangements	229,580			257,038		

Living with partner/spouse/family/friends		74.5%	-		77.3%	-
Living alone		25.0%	-		22.3%	-
Living in a nursing home, hospital or other long term care home		0.1%	-		0.1%	-
Other		0.4%	-		0.3%	-
Pre-operative patient health history						
Pre-operative Assisted	231,023			259,014		
Yes		15.8%	-		15.8%	-
Pre-operative Symptom Period	230,727		-	258,778		
Less than 1 year		12.5%	-		4.6%	-
1 to 5 years		70.1%	-		52.6%	-
6 to 10 years		11.6%	-		22.2%	-
More than 10 years		5.8%	-		20.6%	-
Pre-operative previous surgery	231,085			259,220		
Yes		3.4%	-		3.9%	-
Pre-operative disability	219,936			250,127		
Yes		53.9%	-		51.4%	-

Comorbidities	232,808			261,093		
Arthritis		72.8%	-		78.0%	-
Cancer		5.5%	-		5.4%	-
Circulation		4.7%	-		5.9%	-
Depression		7.9%	-		8.8%	-
Diabetes		9.4%	-		12.8%	-
Heart disease		8.8%	-		9.6%	-
High blood pressure		39.0%	-		45.4%	-
Kidney disease		1.9%	-		2.0%	-
Liver disease		0.6%	-		0.6%	-
Lung disease		8.4%	-		9.3%	-
Nervous system		0.8%	-		1.0%	-
Stroke		1.3%	-		1.6%	-
EQ-5D index score						
Pre-op E5-D index score	232,808	0.35	0.319	261,093	0.418	0.308
Post-op E5-D index score	232,808	0.801	0.239	261,093	0.748	0.247
Post-op E5-D index score change	232,808	0.452	0.339	261,093	0.33	0.326

Oxford Hip (Knee) Score						
Hip (Knee) replacement pre-op score	230,477	17.815	8.052	258,282	19.149	7.65
Hip (Knee) replacement post-op score	231,037	39.978	8.474	259,224	36.042	9.328
Post-operative questionnaire						
Post-op “Success”	232,808			261,093		
Yes		95.8%	-		90.4%	-
Post-op “Satisfied”	231,969			259,892		
Yes		78.1%	-		63.5%	-
Post-op Readmitted	232,089			260,143		
Yes		6.0%	-		7.7%	-

Table A5: Descriptive statistics for hip and knee replacement sample – Regression sample

Descriptive Statistics Hip and Knee Replacement						
	Hip			Knee		
	Observations	**Mean/%**	**SD**	**Observations**	**Mean/%**	**SD**
Demographics						
Age	225,604			256,258		
20 to 29		0.0%	-		0.0%	-
30 to 39		0.1%	-		0.0%	-
40 to 49		1.7%	-		0.2%	-
50 to 59		12.4%	-		10.2%	-
60 to 69		33.6%	-		36.2%	-
70 to 79		39.1%	-		41.6%	-
80 to 89		13.0%	-		11.8%	-
90 to 120		0.1%	-		0.0%	-
Gender	225,604			256,258		
Female		60.8%	-		57.0%	-
Pre-operative living arrangements	225,604			256,258		

Living with partner/spouse/family/friends		74.3%	-		77.1%	-
Living alone		25.2%	-		22.5%	-
Living in a nursing home, hospital or other long term care home		0.1%	-		0.1%	-
Other		0.4%	-		0.3%	-
Pre-operative patient health history						
Pre-operative Assisted	224,326			254,784		
Yes		15.9%	-		15.8%	-
Pre-operative Symptom Period	225,604			256,258		
Less than 1 year		12.5%	-		4.5%	-
1 to 5 years		70.1%	-		52.5%	-
6 to 10 years		11.7%	-		22.3%	-
More than 10 years		5.8%	-		20.6%	-
Pre-operative previous surgery	225,604			256,258		
Yes		3.4%	-		7.4%	-
Pre-operative disability	225,604			256,258		
Yes		54.2%	-		51.7%	-

Comorbidities	225,604			256,258		
Arthritis		74.4%	-		79.6%	-
Cancer		5.6%	-		5.5%	-
Circulation		4.9%	-		5.9%	-
Depression		8.3%	-		9.1%	-
Diabetes		9.6%	-		13.0%	-
Heart disease		9.1%	-		9.8%	-
High blood pressure		39.7%	-		46.2%	-
Kidney disease		2.0%	-		2.0%	-
Liver disease		0.6%	-		0.6%	-
Lung disease		8.7%	-		9.5%	-
Nervous system		0.8%	-		1.1%	-
Stroke		1.4%	-		1.7%	-
EQ-5D index score						
Pre-op E5-D index score	225,604	0.346	0.32	256,258	0.418	0.309
Post-op E5-D index score	216,639	0.8	0.239	245,711	0.748	0.247
Post-op E5-D index score change	216,639	0.452	0.339	245,711	0.329	0.326

Oxford Hip (Knee) Score						
Hip (Knee) replacement pre-op score	225,604	17.703	8.042	256,258	19.128	7.666
Hip (Knee) replacement post-op score	223,158	39.871	8.525	252,444	35.99	9.348
Post-operative questionnaire						
Post-op “Success”	223,096			253,792		
Yes		95.8%	-		90.3%	-
Post-op “Satisfied”	222,780			253,337		
Yes		77.8%	-		63.2%	-
Post-op Readmitted	223,071			253,741		
Yes		6.1%	-		7.7%	-

Figure A2: Evolution of pre-operative EQ-5D index score – hip replacement sample

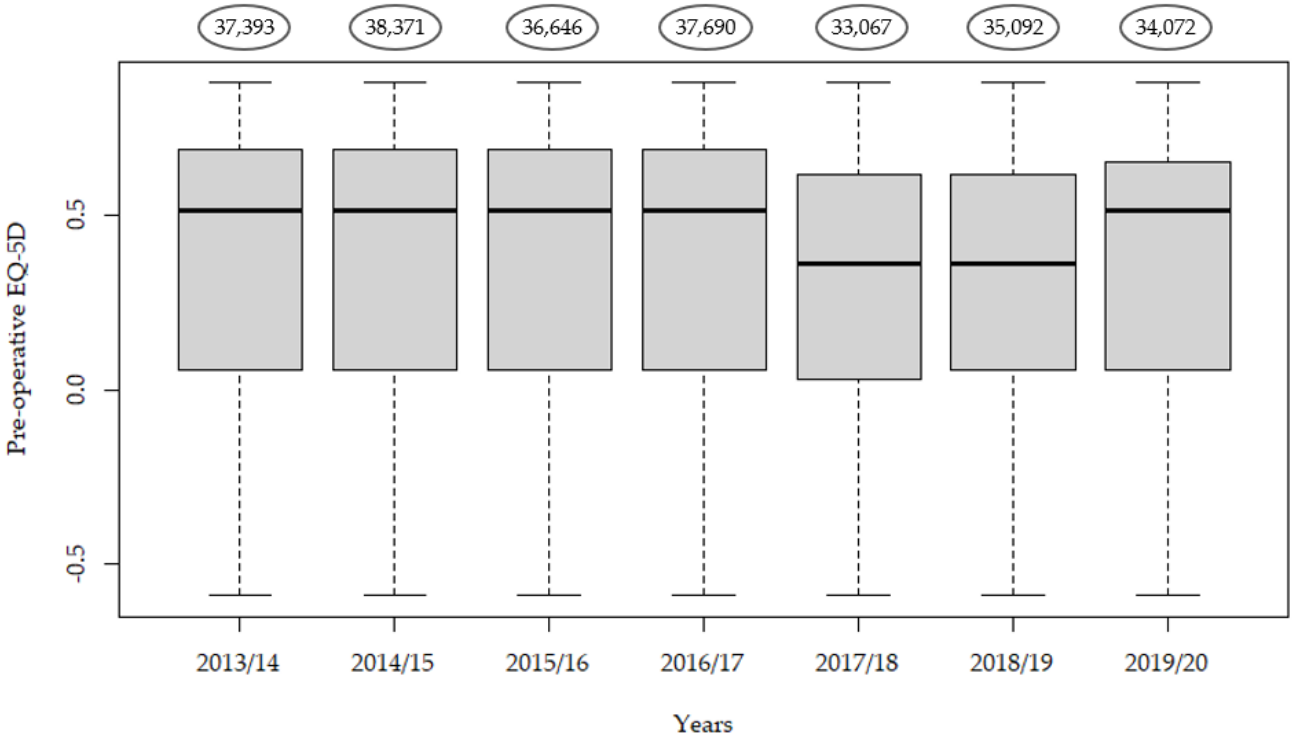


Figure A3: Evolution of pre-operative EQ-5D index score – knee replacement sample

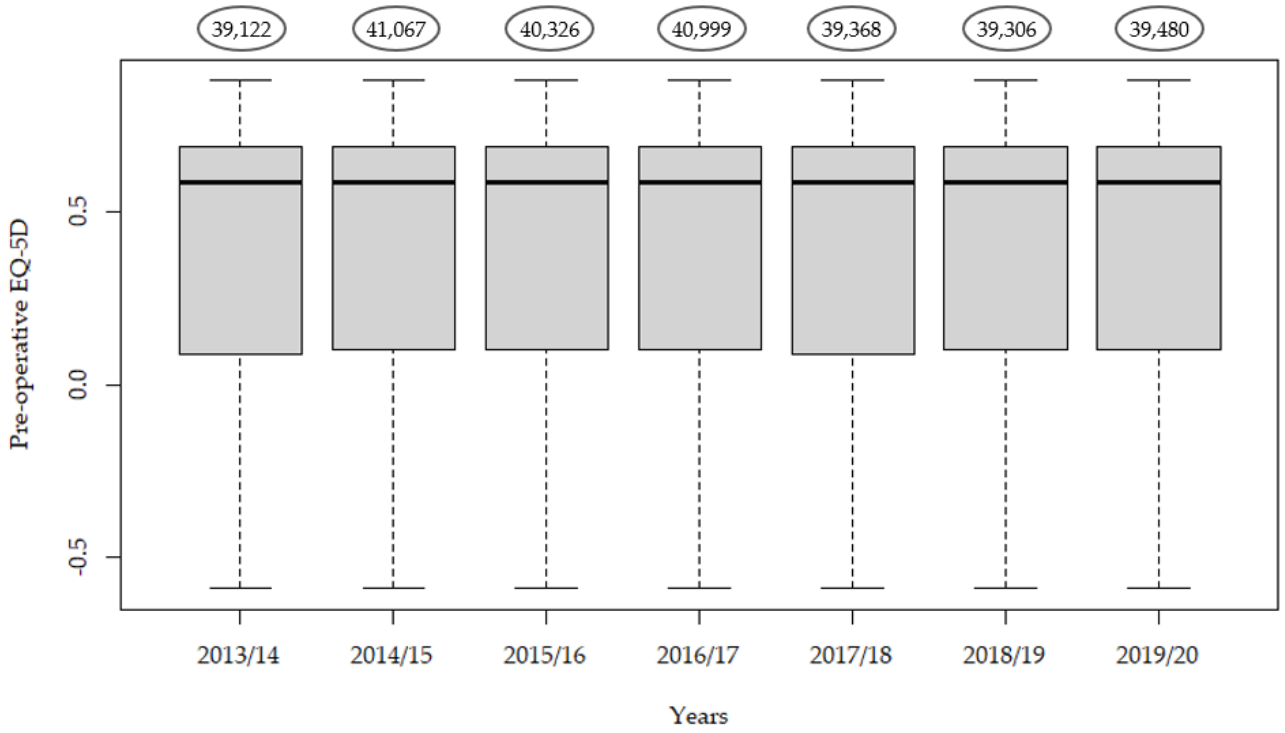


Figure A4: Evolution of post-operative EQ-5D index score – hip replacement sample

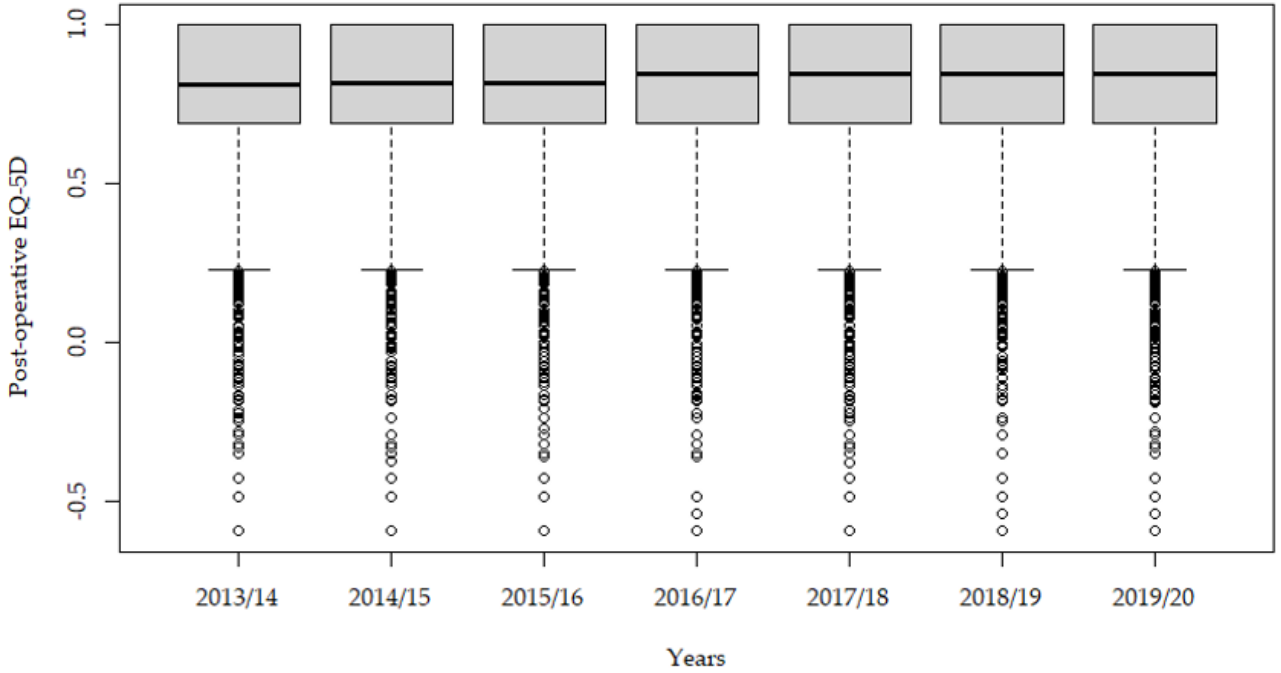


Figure A5: Evolution of post-operative EQ-5D index score – knee replacement sample

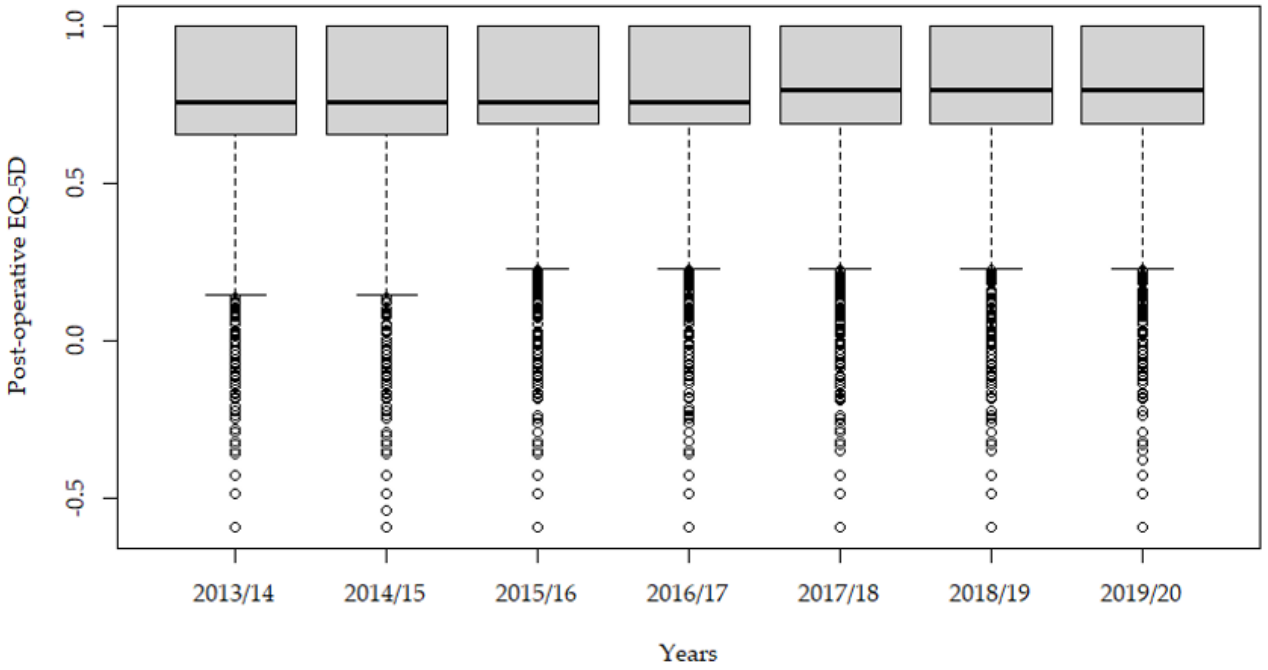


Table A6: Regression results on 6-months EQ-5D-3L index score improvement for knee replacement sample

Knee Replacement Regression Results		
	Dependent variable:	
	EQ5D Index Change	
	(1)	*(2)*
	Multivariate OLS	*Standardized multivariate OLS*
Demographics		
Age 30 to 39	-	-
Age 40 to 49	-	-
Age 50 to 59	0.010 (0.011)	0.009 (0.011)
Age 60 to 69	0.054*** (0.011)	0.081*** (0.011)
Age 70 to 79	0.079*** (0.011)	0.119*** (0.011)
Age 80 to 89	0.082*** (0.011)	0.081*** (0.011)
Age 90 to 120	0.086* (0.046)	0.003 (0.046)
Gender	0.003*** (0.001)	0.005*** (0.001)
Pre-operative Living Arrangements	-0.016*** (0.001)	-0.022*** (0.001)

Pre-operative patient health history		
Pre-operative Assisted	0.004*** (0.001)	0.007*** (0.001)
Pre-operative Symptom Period	0.001 (0.001)	0.001*** (0.001)
Pre-operative Previous Surgery	0.019*** (0.002)	0.011*** (0.002)
Pre-operative Disability	0.063*** (0.001)	0.097*** (0.001)
Comorbidities		
Arthritis	-0.015*** (0.001)	-0.018*** (0.001)
Cancer	0.0004 (0.002)	0.0003 (0.002)
Circulation	-0.061*** (0.002)	-0.044*** (0.002)
Depression	-0.114*** (0.002)	-0.100*** (0.002)
Diabetes	-0.028*** (0.001)	-0.028*** (0.001)
Heart disease	-0.020*** (0.002)	-0.018*** (0.002)
High blood pressure	-0.001 (0.001)	-0.002** (0.001)
Kidney disease	-0.012*** (0.003)	-0.005 (0.003)
Liver disease	-0.004 (0.006)	-0.001 (0.006)
Lung disease	-0.019*** (0.002)	-0.017*** (0.002)
Nervous system	-0.049*** (0.004)	-0.015*** (0.004)

Stroke	-0.027*** (0.003)	-0.011*** (0.003)
EQ-5D-3L Index		
Pre-operative EQ5D Index	-0.897*** (0.002)	-0.849*** (0.002)
Oxford Knee Score		
Knee Replacement Pre-operative Pain	-0.018*** (0.001)	-0.035*** (0.001)
Knee Replacement Pre-operative Kneeling	0.003*** (0.001)	0.007*** (0.001)
Knee Replacement Pre-operative Night Pain	0.009*** (0.0005)	0.032*** (0.0005)
Knee Replacement Pre-operative Washing	0.026*** (0.001)	0.078*** (0.001)
Knee Replacement Pre-operative Transport	0.004*** (0.001)	0.010*** (0.001)
Knee Replacement Pre-operative Confidence	0.005*** (0.0005)	0.018*** (0.0005)
Knee Replacement Pre-operative Shopping	0.015*** (0.001)	0.056*** (0.001)
Knee Replacement Pre-operative Walking	0.007*** (0.0005)	0.025*** (0.0005)
Knee Replacement Pre-operative Limping	-0.020*** (0.001)	-0.060*** (0.001)
Knee Replacement Pre-operative Stairs	0.006*** (0.001)	0.018*** (0.001)
Knee Replacement Pre-operative Standing	-0.001 (0.001)	-0.003*** (0.001)
Knee Replacement Pre-operative Work	0.010*** (0.001)	0.028*** (0.001)
Constant	0.410*** (0.012)	(0.012)

Observations	245,711	245,711
R^2	0.545	0.545
Adjusted R^2	0.545	0.545

Note: *p<0.1; **p<0.05; ***p<0.01

Table A7: Unstratified and stratified MIDs for hip and knee replacement with "Satisfaction" variable as anchor

MIDs - Hip replacement				MIDs - Knee replacement			
Anchor: Satisfaction	**Observations**	**AUC**	**MIDs**	**Anchor: Satisfaction**	**Observations**	**AUC**	**MIDs**
Unstratified	251,992	0.655	0.282	**Unstratified**	279,144	0.657	0.193
Pre-operative EQ-5D				**Pre-operative EQ-5D**			
-0.594 to -0.055	24,989	0.744	0.778	-0.594 to 0.03	40,034	0.733	0.71
-0.05 to 0.109	67,726	0.795	0.717	0.05 to 0.213	69,175	0.787	0.646
0.118 to 0.383	28,954	0.802	0.611	0.219 to 0.433	8,711	0.784	0.441
0.414 to 0.639	65,801	0.813	0.197	0.452 to 0.639	53,794	0.772	0.117
0.656 to 0.883	64,522	0.829	0.133	0.656 to 0.883	107,430	0.804	0.095

Table A8: Unstratified and stratified MIDs for hip and knee replacement with alternative anchor "Success". "Much better": 1 = "much better"; 0 = "a little better", "about the same", "a little worse", "much worse". "A little better": 1 = "a little better"; 0 = "about the same", "a little worse", "much worse"

MIDs - Hip replacement				MIDs - Knee replacement			
Anchor: Success ("Much better")	**Observations**	**AUC**	**MIDs**	**Anchor: Success ("Much better")**	**Observations**	**AUC**	**MIDs**
Unstratified	252,331	0.716	0.235	**Unstratified**	279,668	0.693	0.185
Pre-operative EQ-5D				**Pre-operative EQ-5D**			
-0.594 to -0.055	25,015	0.795	0.767	-0.594 to 0.03	40,109	0.777	0.646
-0.05 to 0.109	67,845	0.849	0.646	0.05 to 0.213	69,321	0.83	0.6
0.118 to 0.383	29,002	0.851	0.542	0.219 to 0.433	8,732	0.82	0.441
0.414 to 0.639	65,876	0.864	0.164	0.452 to 0.639	53,888	0.81	0.114
0.656 to 0.883	64,593	0.877	0.095	0.656 to 0.883	107,618	0.848	0.081
Anchor: Success ("A little better")	**Observations**	**AUC**	**MIDs**	**Anchor: Success ("A little better")**	**Observations**	**AUC**	**MIDs**
Unstratified	30,841	0.659	0.078	**Unstratified**	70,297	0.656	0.078
Pre-operative EQ-5D				**Pre-operative EQ-5D**			
-0.594 to -0.126	1,952	0.718	0.38	-0.594 to 0.112	2,987	0.711	0.292

-0.113 to 0.101	10,778	0.701	0.542	-0.095 to 0.101	19,687	0.719	0.534
0.118 to 0.383	3,181	0.678	0.367	0.106 to 0.383	9,874	0.697	0.438
0.414 to 0.639	8,253	0.68	0.043	0.416 to 0.639	15,205	0.679	0.043
0.656 to 0.883	6,677	0.684	-0.024	0.656 to 0.883	22,544	0.701	0.008

Table A9: MIDs for hip replacement patients – impact on the patient sample

Subgroups	EQ-5D mean change	Unstratified MID	% achieving unstratified MID	Stratified MID	% achieving stratified MIDs
EQ5D 1 (-0.594 to -0.055)	0.760	0.185	92%	0.661	70%
EQ5D 2 (-0.05 to 0.109)	0.727	0.185	94%	0.610	74%
EQ5D 3 (0.118 to 0.383)	0.618	0.185	95%	0.525	75%
EQ5D 4 (0.414 to 0.639)	0.244	0.185	64%	0.114	74%
EQ5D 5 (0.656 to 0.883)	0.179	0.185	61%	0.045	78%

Table A10: MIDs for knee replacement patients – impact on the patient sample

Subgroups	EQ-5D mean change	Unstratified MID	% achieving unstratified MID	Stratified MID	% achieving stratified MIDs
EQ5D 1 (-0.594 to 0.03)	0.614	0.114	88%	0.567	69%
EQ5D 2 (0.05 to 0.213)	0.612	0.114	93%	0.542	68%
EQ5D 3 (0.219 to 0.433)	0.451	0.114	92%	0.405	63%
EQ5D 4 (0.452 to 0.639)	0.143	0.114	55%	0.081	65%
EQ5D 5 (0.656 to 0.883)	0.125	0.114	48%	0.012	73%

Table A11: Sensitivity analysis with half SD distribution-based method

	MIDs - Hip replacement		MIDs - Knee replacement	
	SD	SD*0.5	SD	SD*0.5
Unstratified	0.339	0.170	0.327	0.163
Pre-operative EQ-5D				
-0.594 to -0.055	0.323	0.162	0.317	0.158
-0.05 to 0.109	0.255	0.128	0.245	0.123

0.118 to 0.383	0.229	0.115	0.231	0.116
0.414 to 0.639	0.207	0.103	0.219	0.109
0.656 to 0.883	0.162	0.081	0.177	0.088

4 Adoption of PROMs

4.1 A collection of components to design clinical dashboards incorporating patient-reported outcome measures: Building on semi-structured interviews

Anja Y. Bischof

David Kuklinski

Irene Salvi

Carla Walker

Justus Vogel

Alexander Geissler

Under review in the *Journal of Medical Internet Research*

4.1.1 Abstract

Background: A clinical dashboard is a data-driven clinical decision support tool visualizing multiple key performance indicators in a single report while minimizing time and effort for data gathering. Studies have shown that including patient-reported outcome measures (PROMs) in clinical dashboards supports the clinician's understanding of how treatments impact patients' health status, helps identifying health-related quality of life changes at an early stage and strengthens patient-physician communication.

Objective: We aim to determine design components for clinical dashboards incorporating PROMs to inform software producers and users (i.e., physicians).

Methods: We conducted interviews with software producers and users to test pre-selected design components. Further, the interviews allowed to derive additional components that are not outlined in existing literature. Finally, we used inductive and deductive coding to derive a guide on which design components need to be considered when building a clinical dashboard incorporating PROMs.

Results: 25 design components were identified, of which 16 were already surfaced during literature search. Nine additional components were derived inductively during our interviews. The design components are clustered in generic dashboard, PROM-related, adjacent information, and requirements for adoption components. Both software producers and users agreed on the primary purpose of a clinical dashboard incorporating PROMs to enhance patient communication in outpatient settings. Dashboard benefits include enhanced data visualization and improve workflow efficiency, while interoperability and data collection were named as adoption challenges. Consistency in dashboard design components is preferred across different episodes of care, with adaptations only for disease-specific PROMs.

Conclusion: Clinical dashboards have the potential to facilitate informed treatment decisions if certain design components are followed. This study establishes a comprehensive framework of design components to guide the development of effective clinical dashboards incorporating PROMs in healthcare practice.

Keywords: clinical dashboards, design components, patient-reported outcome measures (PROMs), patient-reported outcomes (PROs)

4.1.2 Background

Technological advances have created a multitude of innovative prevention and treatment options in medicine. Despite providing benefits to patients, this evolution created settings of increasingly complex medical decision-making [1] and an increased distancing of physicians from their patients [2]. The use of clinical dashboards can support physicians to deal with various relevant sources of information such as clinical parameters or patient-reported outcomes (PROs). A clinical dashboard can serve as a data-driven clinical decision support tool if capable of querying multiple data sources and visually representing key performance indicators in a single frame. The added value of a dashboard comes from its ability to provide a concise overview of key information along the treatment process [3]. Theoretically, visualizing data on clinical dashboards should decrease time spent on data gathering, cognitive overload, and time to task completion, and should improve situation awareness and compliance with evidence-based safety guidelines [4]. However, frustration of physician increases as systems are not running smoothly and sometimes even hinder physicians in their daily work [5].

Incorporating the patient perspective into clinical dashboards via patient-reported outcome measures (PROMs) offers considerable potential to enhance clinical outcomes and patient's health related quality of life (HRQoL) [6–9]. Furthermore, it can result in a more patient-centered treatment approach and improved care processes [10, 11]. PROMs assess, for instance, patients' HRQoL, functional impairments, and disabilities. Additionally, PROMs can report the nature and severity of symptoms in a standardized way [11, 12], independent from physicians' or caregivers' interpretation [13, 14]. The use of PROMs in clinical dashboards has, thus, been shown to support physicians' understanding of how a treatment impacts symptom scores [11] and to help identify HRQoL deterioration at an early stage [12].

Research on the development of dashboards including PROMs has been conducted for outpatient treatment of prostate cancer, overactive bladder, multiple sclerosis, and various surgical procedures [12, 15–18]. Moreover, most of the existing literature has focused on disease-specific PROMs [9, 11, 16, 19, 20]. Accordingly, the potential of

incorporating a generic PROM, such as the EQ-5D, into a dashboard has not been investigated yet. Incorporating a generic PROM enables comparability across diseases and an overall assessment of the patient's HRQoL. However, research shows that physicians will only use clinical dashboards if they perceive an additional value [11].

Current literature focuses considerably more often on dashboards for chronic diseases [11, 12, 16, 17, 20–22] compared to dashboards on one-time acute care interventions [18, 23]. Furthermore, the literature reveals the relevance of complementary data as important design components of clinical dashboards incorporating PROMs. Components such as patient information (such as demographic information, or most recent health updates) [20, 24, 25] or clinical data (such as lab results and medication data) [20, 24, 26, 27] are highlighted. Further, the inclusion of additional design components such as past assessment scores (i.e., the visualization of PRO scores over time) [9, 11, 27, 28], peer-group comparisons (i.e., analyze a patient's PRO scores with peer group) [17, 27, 29], PRO-related and overall health-related goals (i.e., goals concerning PROs or overall health developed during patient-physician interaction) [16, 27], alerts (i.e., in case of critical values) [9, 12], free-write in features (i.e., additional space for physician's notetaking) [9, 28], and dashboard customizability (i.e., degree of how far individual wishes of users may be respected) [16, 17, 24] are promoted (for more information, please refer to Multimedia Appendix 1).

However, due to the literature's inconsistency on design components for clinical dashboards, we aim at developing a guide on which design components to consider when building clinical dashboards incorporating PROMs. Therefore, we investigate the following two research questions: (1) What are design components of clinical dashboards incorporating generic and disease-specific PROMs? (2) Do the design components for clinical dashboards incorporating PROMs differ for acute conditions with a one-time intervention versus chronic conditions?

4.1.3 Methods

We conducted interviews with software producers and users, i.e., physicians, to identify design components for user-friendly clinical dashboards to enhance quality

of care [18]. By interviewing two stakeholder groups, we collected insights on two key perspectives (software producers vs users).

The recruitment of the interviewees followed a consecutive process: First we recruited in our own network, and then, through a snowballing technique, we were able to contact software producers and users outside of our network. The conduction of interviews ended once thematic saturation was reached, i.e., when no additional information was gained from further interviews [30]. The interviews were semi-structured, and findings from a previous literature search (see Multimedia Appendix 1) guided the development of the interview questionnaire. Initially, AYB developed the interview guide, while IS, DK, and AG collaborated to refine it. To finalize the questionnaire, a pilot interview was conducted with a key opinion leader in this subject area. After the first five interviews, we iteratively tailored the interview structure to align with the evolving interview dynamics, facilitating focused exploration of our research priorities. This adaptive approach involved selecting pertinent questions towards the gravitation points that seemed to emerge from our pre-established question catalogue, thus augmenting the inductive explanatory insights.

All interviews were conducted by two authors (AYB and IS) to ensure that the same content was covered. Both interviewers were female, had received training in qualitative methods as part of their PhD studies, and had no prior relationship to any of the interviewees. All interviews took place between 11/02/2022 and 05/15/2023. Interviewees decided on the interview date and time.

Prior to each interview, three introductory slides were presented to all interviewees to ensure a common understanding of clinical dashboards, outline the underlying research questions, and clarify content-related questions. All interviews covered the following main areas: 1) general questions on dashboards and PROM usage, 2) questions about the market (penetration) (software producers only), 3) usage of the dashboard, 4) dashboard development and data collection, 5) feature assessment, 6) role of the patient (Multimedia Appendix 2).

We conducted the one-to-one interviews online (via Zoom or Teams) in English or German. During the interviews, both interviewers took notes. All interviews were audio-recorded and transcribed verbatim to perform detailed data analysis. We sent the transcripts back to the interviewees to receive approval and then anonymized all participant details. For the analysis of the interviews, we applied thematic coding to uncover potential patterns and themes within the data [31]. To identify relevant information for the design components, first, deductive coding was used to cross-validate the findings from the literature search. Second, inductive coding allowed us to identify and classify information not covered by the literature search. AYB and CW coded all interviews and conducted the analysis with Atlas.ti Windows (Version 22) [32]. The study is reported in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) 32-item checklist to ensure transparency and reliability (Multimedia Appendix 3) [33].

This study was performed in line with the principles of the Declaration of Helsinki. The Ethics Committee of the University of St. Gallen was informed about this research and issued a letter of exemption.

4.1.4 Results

Step 1. Scoping Literature Review

In total, we conducted 16 interviews with six software producers (I1 to I6) and ten users (I7 to I16) until thematic saturation was reached (male = 10, female = 6). All software producer interviewees were from different companies and had different positions. Eight out of ten user interviewees were department heads with different specialties such as pneumology, oncology, pediatrics, or orthopedics. This allowed us to analyze distinctive needs for design components in clinical dashboards. One of the remaining user interviewees was a chief hospital innovation officer, and the other interviewee was a member of hospital quality management. These specific interviewees lacked a formal medical background. However, based on their employment within a hospital setting, they routinely interacted closely with diverse medical practitioners and thus, can be assumed to have a profound understanding of

physicians' needs for clinical dashboards. The software producer interviews lasted between 29 and 45 minutes, whereas the user interviews had a duration of 20 to 36 minutes (Multimedia Appendix 4).

The interviews with software producers as well as users revealed that they perceived a differentiation in design depending on disease (one-time acute vs. chronic condition) as not essential. Therefore, we did not distinguish between design requirements of different disease types but instead focused on the design principles for clinical dashboards incorporating PROMs in general.

General information

To evaluate the importance of the key design components, we asked both interview groups to outline the type of disease, the setting, the type of PROM used, and the targeted key user for their clinical dashboard. Software producers mostly indicated they prefer dashboards which are applicable to chronic and acute conditions similarly (4/6) (Multimedia Appendix 5 contains the full set of codes). For users, we observed that they preferred a dashboard for either a chronic or acute disease – which often correlated with their professional focus. Both interview groups rated the outpatient setting as the preferable area of application (8/16). Both interview groups mentioned using a disease-specific PROM only (6/16) or combining generic and disease-specific PROMs (9/16). However, none of the interviewees preferred to use a generic PROM only – whereas I11 acknowledged that his team only uses generic PROMs for research purposes. Furthermore, both interview groups favored physicians (13/16), followed by other healthcare professionals (such as nurses or physiotherapists) (7/16) as key users. The patient was also recognized as a potential user of clinical dashboards (8/16), however, there were some discrepancies on how much and which content should be displayed to patients. I13 also highlighted that clinical dashboards may be valuable tools for relatives supporting patients in their daily activities to grasp the patient's health status better.

Data collection

The most frequently mentioned use case for data collection was for micro perspectives (i.e., patient-physician communication and intra-patient comparison) (12/16), especially by users (8/10), followed by meso perspectives (i.e., comparison of patient groups within departments or institutions) for both interview groups (5/16) (Table 1). Furthermore, the most often mentioned purposes of reporting were "communication" (14/16) and "better basis for decision-making (for physician)" (9/16) (Table 1).

For the point-in-time of data collection, there was a clear tendency (11/16) of interviewees for patients to answer the questionnaires independently at home. However, I2 mentioned that it is perceivable that the first-time data collection is conducted in the waiting room because *"[patients] get an explanation by a physician or a nurse, but after that, they are on their own."* Also, interviewee I14 had a similar perception for dealing with the data collection process. For the type of data collection, a clear tendency towards online tools such as platforms or phone apps was indicated (10/16).

Table 4.1: Overview of software producer and user responses on how data collection should be implemented in the clinical dashboard

Data collection	
Level of reporting	
Software producer 1. Micro (4/6) 2. Macro (2/6) 3. Meso (1/6)	**User** 1. Micro (8/10) 2. Meso (4/10) 3. Macro (1/10)
Purpose of reporting	
Software producer 1. Improved communication (6/6) 2. Better basis for physician's decision (4/6) 3. Real-time tracking (2/6) | Shared decision-making (2/6)	**User** 1. Improved communication (8/10) 2. Better basis for physician's decision (5/10) 3. Shared decision-making (2/10)
Data collection	
Software producer 1. Digital (5/6)	**User** 1. Digital (5/10) 2. Paper-based (4/10)

Time of data collection	
Software producer	**User**
1. Independent at home (4/6) 2. Before the appointment (in the waiting room) (2/6)	1. Independent at home (7/10) 2. Before the appointment (in the waiting room) (6/10)

Legend: The table shows the ranking of design principles' features according to the number of mentions by software producers and users. Behind the features, the share of interviewees mentioning this design principle's feature is indicated. As the interviewees could provide more than one answer or no answer for each design principle, shares can add up to more or less than 100%.

Dashboard components

Concerning the dashboard components, representatives of both interview groups mentioned that some patient information (such as patient photograph, demographic information or contact details of other care team members) need to be presented on the dashboard (9/16). Interviewee I16 additionally emphasized that she expects the patient information to include *"key events. So, surgeries need to be shown. For example, in cancer, the start of chemotherapy, completion of chemotherapy, started radiation, that you can understand what is going on in the background of those patients."*

Except for I1 and I5, all other software producers rated clinical data (e.g., lab results or medication data) as meaningful information that must be included in the clinical dashboard. In contrast, users tended rather not to overfill the clinical dashboard with clinical data as this information might be available from another source, as I8 mentioned: *"[…] and then if they [the physicians] want to have the clinical information on the patient, they just open up the EMR [electronic medical record] and check it, which is another tab in the Chrome app."*

The free write-in design component did not resonate well in either of the two groups. Six interviewees replied that a free write-in space is not useful, whereas only two interviewees perceived this design component as very beneficial. Nevertheless, when having such a free write-in box included, users (3/10) wanted it as an additional source of information for some specific variables.

Software producers and users rated the past assessment PRO score as one of the most crucial design component in a clinical dashboard incorporating PROMs. From the

software producer perspective, I6 mentioned: *"This [past assessment PRO score] is very well received, simply the score progression up and down visually, so to speak."* I9 had a similar perception: "*This [past assessment PRO score] is absolutely relevant because it is above all changes in these questionnaires that are significant. In any case, it is very important to look at the progression, not just the individual value."*

Another well-perceived design component was the peer-group comparison. All software producer interviewees (6/6) agreed on the inclusion of this feature. The users (7/10) appreciated the inclusion, too. However, three users did not perceive an added value in the peer-group comparison due to uncertainties in interpreting the applied PRO (I9), the missing guidance on the relevant factors to compare different patient groups (I12), or a clear reasoning on why peer group comparison is useful (I15).

Although only one user emphasized applying PRO-related goals, overall health-related goals were discussed more controversially within both groups. Software producer I4 reported that their dashboard includes a component where the physician can develop overall health goals together with the patient and check goal achievement, whereas I6 stated that this is not part of their dashboard. From the user perspective, this design component could add value to the patient-physician communication, as the goal statement makes the aim or expectations of the patients explicit (I9, I16).

Concerning alerts, software producers and users have different perceptions on whether alerts should be included in the dashboard. Software producers (5/6) mentioned that an alert function needs to be included. Generally, producers were indifferent regarding whether the alert is real-time (4/6) or only appears during the appointment (4/6). I4 mentioned, however, that liability for medical risks and errors is unclear when real-time alerts are used: *"if [a score is] three days red or is really critical and no push notification is sent, the error is on our side. However, if a push notification is sent and the doctor does not react, the error lies with the doctor. And that is a bit of a grey area, where we still have to figure out how it's actually done."* In contrast, four users preferred not to include alerts. I8 and I11 mentioned that this feature previously existed in their dashboard, but the acceptance was not high enough in their teams, which made them stop using it. Further, I12 raised the issue that additional

interpretations for physicians are required to ensure that they completely understand what the deterioration or improvement in a score means. I10 favored the alerts during the appointment to highlight critical factors and to facilitate comparison over time. I13 and I16 preferred real-time alerts allowing the treating physician to react directly to the patient's issues.

Concerning the dashboard's customizability, all interviewees agreed that some degree of customizability is required to meet different user needs. Although most users preferred customizability from a standard set (7/10) – software producers still seemed undecided whether a clinical dashboard should be adapted to individual needs (2/6) or "off the shelf" (3/6). Nevertheless, software producers (5/6) agreed that scalability is only achievable in case clinical dashboards equipped with features defined in a standard set are provided.

Findings from inductive coding

Whereas above-presented results were derived from deductive procedure, the subsequent inductive procedure allowed us to identify additional aspects that were not explicitly discussed in current literature. These aspects included on the one hand the added value of clinical dashboards incorporating PROMs and visualization possibilities of PROMs and on the other hand various barriers, reducing the uptake of the clinical dashboard. Furthermore, the interviewees also discussed the patient's access to their data controversially.

Although the visualization of the PRO scores was not extensively discussed in the literature, we asked our interviewees about their preferences. Only two out of six software producers indicated that the index and dimensional scores are always provided in their dashboards. The remaining software producers elaborated on the visualization of their PRO scores. Contrarily, almost all users (9/10) mentioned their preferences for the visualization: 4/10 users preferred to have index and dimensional scores available, 3/10 users prioritized dimensional scores, whereas 2/10 users voted for the visualization of index scores. For instance, I7 voted for the visualization of dimensional scores by stating, *"we must of course know the dimensionality and*

different aspects." In contrast, I11 argued: *"I like [index] scores better, as I said, but because we have these individual questions like there are ten questions, and you can have a summary score of it and the system plots every question on a trend. I think that is rather messy because then you have like ten different color graphs just projected over each other, and you can click them on, or off. So it's easy, but for me it's less informative."*

Concerning added value provided by the clinical dashboard, software producers emphasized the visualization of results (5/6), enhanced workflow efficiency (4/6), improved data comparability (3/6), and higher patient satisfaction (3/6). Users perceived the biggest advantages of a clinical dashboard in the visualization of results (8/10), enhanced workflow efficiency (6/10), and improved data comparability (5/10). Higher patient satisfaction (2/10) was not among the most often named advantages.

Potential barriers to implementing the full potential of clinical dashboards are interoperability between various systems for both interview groups (12/16) and the burdensome collection of data for users (6/10). Reasons for the burdensome collection were various. For instance, I9 highlighted organizational issues by stating: "*There is one burden that falls more on the admin staff and the secretaries, which is to make sure that the right patient is put on the right pathway with the right PROMs. So, you need to pre-program all this. Then you must either have an automatic integration saying this patient is going to see this because it needs to be personalized to the patient pathway. So how do you make sure that the right patients get the right thing without having an army of extra people?"* Further, I11 raised technical challenges in collecting PROs by expressing: *"We had a problem for a year now that we had these computers in the waiting room and for some reason they were broken or disconnected. It took us a very long time to get them because there were all these people taking ownership of these silly computers."* Further, software producers mentioned legal consequences in case of displaying inadequate information (4/6), and the licensing of the PROM questionnaires (3/6) as potential barriers. Users rather perceived the lack of good data (4/10), legal consequences (3/10), and the non-intuitive use (3/10) as additional barriers.

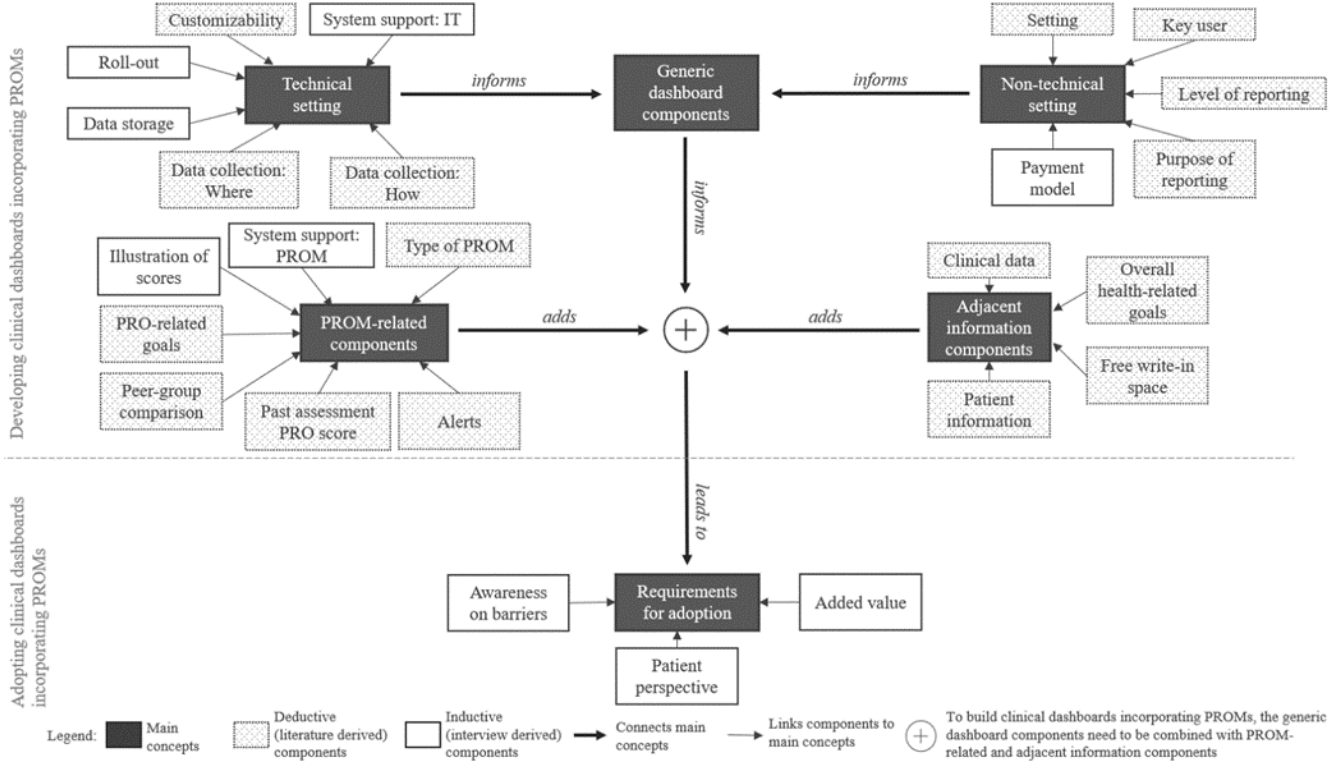
Except for two users (I7 and I14), all interviewees agreed that patients benefit from the access to the information stored on the clinical dashboard. Nine of the interviewees expect the patients to develop a better understanding and interpretation of their own health status. Software producer I2 summarized their efforts as *"what we are doing now is that we have patient apps, and the patients can also see their own results in their own app. They can also say to the caregiver: 'hey, I see this or that', and so we involve the patients into it. I think it is important to make it patient-friendly also."* Users highlighted similar scenarios as, for instance, mentioned by I8: *"When reading this data with PROMs, it's very clear. It's data that they have entered. If they see that they're functioning or the ability to walk a block or go on a bus is going down, they can relate to it and understand 'Oh, actually I thought I was doing better than that' or 'I'm happy, actually I forgot that I couldn't and now I can."* This quote also represents the statement of I14, whereas I15 additionally emphasized the positive correlation between an enhanced understanding of one's own health status and increased treatment adherence. Especially the score visualization supports the patient's understanding of their own health status (6/16). Still, users (7/10) acknowledged that more explanation on the meaning of the data is required to avoid patient's misinterpretation of the data. Furthermore, two users claimed that physicians and patients need different level of aggregation of available data concerning depth of analysis and interpretation complexity to avoid overwhelming patients. Finally, I6 and I8 stated they would refuse to share clinical data with patients as clinical outcomes might frighten patients or lead to misinterpretations – such as outlined by I8: *"If you look at in oncology or radio oncology to PSA levels, which are an indicator but not an actual sign of cancer, you wouldn't want the patients to learn by themselves. 'Okay, my PSA level are up again, and I don't know what it means and what is going on.' With this type of outcomes, you need to be very careful on how you get them across […] and this needs professional support."*

Findings from deductive and inductive coding

The derivation of deductive and inductive codes led to insights on how different design components are related to each other (Figure 1). To build clinical dashboards

incorporating PROMs, not only the selected PROMs on the dashboard are crucial, but rather the complete construct of the clinical dashboard. Additionally, generic dashboard components need to be respected as well. Furthermore, aspects such as awareness of barriers, patients' perspective and the potential of added value need to be considered when reflecting on the adoption of clinical dashboards (Multimedia Appendix 6 contains a full description of all topics).

Figure 4.1: Framework summarizing the relationship between the individual 25 design principles and main concepts of clinical dashboards incorporating PROMs



4.1.5 Discussion

Principle results

The qualitative research approach allowed for a more profound understanding of the design components for clinical dashboards incorporating generic and disease-specific PROMs. Our developed framework (Figure 1) displays that we found 16 design components based on a literature search and identified nine additional design components through interviews with software producers and users. The interviews revealed that both interview groups had a similar perception of the use case of a clinical dashboard incorporating PROMs.

The use of a dashboard primarily strengthens the patient-physician communication in outpatient settings. It should include a combination of disease-specific and generic PROMs. Physicians and other healthcare professionals were the most frequently mentioned key users. Additionally, preferences for visualizing PROs, whether as index or dimensional scores, are subjective and dependent on individual inclinations. According to software producers and users, dashboards provide benefits especially through the visualization of data and the enhancement of workflow efficiency. Main barriers for dashboard implementation were seen in the interoperability between various systems and the collection of data.

Additionally, the interviews revealed that dashboard content should not vary across diseases, except by an adaptation of the disease-specific PROM. This project showed that not only design components directly displayed by the clinical dashboard incorporating PROMs, but also other features, such as providing system support or the definition of key users, are relevant topics for a successful implementation.

Comparison with prior work

Concerning the use of PROMs, studies show that it is common to use generic PROMs, particularly the EQ-5D, to track the health status of patients. Notably, researchers see the standardized format and content as a major benefit that facilitates its usability across different diseases and patient groups [34]. Further, the multi-lingual questionnaire allows large-scale analyses [35, 36]. Not only were disease-specific

PROMs complementing generic PROMs, but their combined value seen as more valuable than their sum as in the case of the EQ-5D and the COPD Assessment Test (CAT) combination [37]. Interestingly, similar findings emerged from studies developing PROM dashboards for various disease areas. Baeksted et al. [28] and Hassett et al. [9] found it relevant to include the cancer-care-specific PRO-CTCAE in their dashboards, and Nicolas-Boluda et al. [22] incorporated endometriosis-specific indicators into their dashboard. We found that disease-specific PROMs are preferred over generic PROMs by the interviewees. Nevertheless, a combined use of both PROMs was seen as beneficial, too.

An enhanced visualization of the patient's PROs over time can facilitate the moderation role of the physician [12, 19, 38] – in line with the growing demand for shared decision-making. We found a similar outcome as the interviewees perceived the purpose of a clinical dashboard incorporating PROMs in the improved communication between patient and physician. Furthermore, Desantis et al. [12] found that using a clinical dashboard improves the workflow and communication of changes in the HRQoL between the patient and physician. Following this argument, interviewees emphasized that a clinical dashboard incorporating PROMs may also serve as a basis for better informed decisions by physicians as various key information points are displayed in one report. Therefore, additional information besides the PRO data must be displayed to display a holistic picture of the patient's general health conditions. Patient information (such as sociodemographic information or most recent health updates) and clinical data (such as lab results and medication lists) are required, which was previously highlighted in the literature [20, 24, 25] and also rated as essential by the interviewees in our study.

Furthermore, the interviewees promoted the design component "past assessment PRO score" and ranked it as one of the most important ones. Changes in the patient's health status become apparent and traceable, and thus, enable the initiation of countermeasures immediately. Similar reasons are also provided in the literature [9, 28]. Another well-perceived design component for analyzing PROs was the peer-group comparison. Interviewees claimed that this design component facilitates

understanding whether a patient's health status is above or below the norm. Hartzler et al. [17] highlighted that peer-group comparison should match patients in age and treatment. Additionally, Ragouzeos et al. [27] advise indicating a "normal" range.

Design components where interviewees did not indicate a clear tendency were PRO-related and overall health goals and the function of alerts. Cronin et al. [16] found that patients want to set and evaluate goals over time, which could be supported by a clinical dashboard. However, Liu et al. [26] warned that setting goals might further pressure the patient. Furthermore, the challenge of including alerts is to decide on the appropriate alert level [39] and avoiding "alert fatigue" of physicians [20].

Limitations

Our study has three limitations: First, the interviewees were mainly from Western countries – especially Switzerland and Germany. A generalization of our results to other countries should therefore be carefully assessed. Second, the user interviewees were most often specialists, and we did not control for their years of experience. This research can be seen as a starting point in building guidance on what needs to be considered when developing clinical dashboards incorporating PROMs. Therefore, we wanted to include a broad range of opinions and needs on critical design components. Future research should consider various dashboard stakeholders from different countries with varying years of professional experience to further evaluate our proposed design components. Third, the design components were not built and evaluated in a real-world setting but only discussed during the interviews. We aimed to develop a theoretical framework of design components based on software producers' capabilities and users' needs. Therefore, future research approaches can build up on our findings to refine clinical dashboards and test them in clinical practice.

4.1.6 Conclusion

In our study, we identified 25 design components and consolidated them into a framework of four main concepts. This framework strives to work as a guiding approach where researchers and practitioners can orient themselves on components to be included when building clinical dashboards incorporating PROMs. Design

components such as peer-group comparison or alert thresholds require further investigation in future research, as knowledge on determining appropriate peer groups or setting alerts based on PRO thresholds needs to be improved. Additionally, future research efforts may test the real-world applicability of the proposed design components by either discussing them in focus groups or implementing them in a dashboard in clinical practice.

References

1. Mulley AG, Trimble C, Elwyn G (2012) Stop the silent misdiagnosis: patients' preferences matter. BMJ 345:.

2. Barry MJ, Edgman-Levitan S (2012) Shared Decision Making — The Pinnacle of Patient-Centered Care. New England Journal of Medicine 366:780–781.

3. Wilbanks BA, Langford PA (2014) A review of dashboards for data analytics in nursing. CIN: Computers, Informatics, Nursing 32:545–549.

4. Khairat SS, Dukkipati A, Lauria HA, et al (2018) The Impact of Visualization Dashboards on Quality of Care and Clinician Satisfaction: Integrative Literature Review. JMIR Hum Factors 5:e22.

5. Tawfik DS, Sinha A, Bayati M, et al (2021) Frustration with technology and its relation to emotional exhaustion among health care workers: Cross-sectional observational study. J Med Internet Res 23:e26817.

6. Chen J, Ou L, Hollis SJ (2013) A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res 13:211.

7. Devlin NJ, Appleby J (2010) Getting the most out of PROMs: Putting health outcomes at the heart of NHS decision making. King's Fund

8. Greenhalgh J, Meadows K (1999) The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: A literature review. J Eval Clin Pract 5:401–416.

9. Hassett MJ, Cronin C, Tsou TC, et al (2022) eSyM: An electronic health record–integrated patient-reported outcomes-based cancer symptom management program used by six diverse health systems. JCO Clin Cancer Inform 1–10.

10. Dowding D, Randell R, Gardner P, et al (2015) Dashboards for improving patient care: Review of the literature. Int J Med Inform 84:87–100.

11. Fautrel B, Alten R, Kirkham B, et al (2018) Call for action: How to improve use of patient-reported outcomes to guide clinical decision making in rheumatoid arthritis. Rheumatol Int 38:935–947.

12. Desantis D, Baverstock RJ, Civitarese A, et al (2016) A clinical perspective on electronically collecting patient-reported outcomes at the point-of-care for overactive bladder. Canadian Urological Association Journal 10:359.

13. Cella D, Yount S, Rothrock N, et al (2007) The patient-reported outcomes measurement information system (PROMIS). Med Care 45:S3–S11.

14. Lipscomb J, Reeve BB, Clauser SB, et al (2007) Patient-reported outcomes assessment in cancer trials: Taking stock, moving forward. Journal of Clinical Oncology 25:5133–5140.

15. Jongen P, Sinnige L, van Geel B, et al (2015) The interactive web-based program MSmonitor for self-management and multidisciplinary care in multiple sclerosis: Concept, content, and pilot results. Patient Prefer Adherence 1741.

16. Cronin RM, Conway D, Condon D, et al (2018) Patient and healthcare provider views on a patient-reported outcomes portal. Journal of the American Medical Informatics Association 25:1470–1480.

17. Hartzler AL, Izard JP, Dalkin BL, et al (2016) Design and feasibility of integrating personalized PRO dashboards into prostate cancer care. Journal of the American Medical Informatics Association 23:38–47.

18. Hartzler AL, Chaudhuri S, Fey BC, et al (2015) Integrating patient-reported outcomes into spine surgical care through visual dashboards: Lessons learned from human-centered design. eGEMs (Generating Evidence & Methods to improve patient outcomes) 3:2.

19. LeRouge C, Hasselquist MB, Kellogg L, et al (2017) Using heuristic evaluation to enhance the visual display of a provider dashboard for patient-reported outcomes. eGEMs (Generating Evidence & Methods to improve patient outcomes) 5:6.

20. Rudin RS, Perez S, Rodriguez JA, et al (2021) User-centered design of a scalable, electronic health record-integrated remote symptom monitoring intervention for patients with asthma and providers in primary care. Journal of the American Medical Informatics Association 28:2433–2444.

21. Elm JJ, Daeschler M, Bataille L, et al (2019) Feasibility and utility of a clinician dashboard from wearable and mobile application Parkinson's disease data. NPJ Digit Med 2:95.

22. Nicolas-Boluda A, Oppenheimer A, Bouaziz J, Fauconnier A (2021) Patient-reported outcome measures in endometriosis. J Clin Med 10:5106.

23. Heath EL, Ackerman I, Lorimer M, et al (2022) National implementation of an electronic patient-reported outcome measures program for joint replacement surgery: Pilot study. JMIR Form Res 6:e30245.

24. Tsangaris E, Edelen M, Means J, et al (2022) User-centered design and agile development of a novel mobile health application and clinician dashboard to support the collection and reporting of patient-reported outcomes for breast cancer care. BMJ Surg Interv Health Technol 4:119.

25. Taxter A, Johnson L, Tabussi D, et al (2021) Co-Design of an Electronic Dashboard to Support Coproduction of Care in Pediatric Rheumatic Disease: Human-Centered Design and Usability Testing (Preprint). J Particip Med.

26. Liu LH, Garrett SB, Li J, et al (2020) Patient and clinician perspectives on a patient-facing dashboard that visualizes patient reported outcomes in rheumatoid arthritis. Health Expectations 23:846–859.

27. Ragouzeos D, Gandrup J, Berrean B, et al (2019) “Am I OK?” using human centered design to empower rheumatoid arthritis patients through patient reported outcomes. Patient Educ Couns 102:503–510.

28. Baeksted C, Pappot H, Nissen A, et al (2017) Feasibility and acceptability of electronic symptom surveillance with clinician feedback using the Patient-Reported Outcomes version of Common Terminology Criteria for Adverse Events (PRO-CTCAE) in Danish prostate cancer patients. J Patient Rep Outcomes 1:1.

29. Strachna O, Cohen MA, Allison MM, et al (2021) Case study of the integration of electronic patient-reported outcomes as standard of care in a head and neck oncology practice: Obstacles and opportunities. Cancer 127:359–371.

30. Sandelowski M (2000) Whatever happened to qualitative description? Res Nurs Health 23:334–340.

31. Braun V, Clarke V (2006) Using thematic analysis in psychology. Qual Res Psychol 3:77–101.

32. ATLAS.ti Scientific Software Development GmbH (2022) Atlas.ti 22 Windows

33. Tong A, Sainsbury P, Craig J (2007) Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. International Journal for Quality in Health Care 19:349–357.

34. Zhou T, Guan H, Wang L, et al (2021) Health-related quality of life in patients with different diseases measured with the EQ-5D-5L: A systematic review. Front Public Health 9:.

35. Devlin NJ, Brooks R (2017) EQ-5D and the EuroQol Group: Past, present and future. Appl Health Econ Health Policy 15:127–137.

36. Sen RK, Shetti V, Mukhopadhyay R, et al (2022) Satisfaction and health-related quality of life following hip and knee arthroplasty surgeries in Indian patients: A cross-sectional Study. Indian J Orthop 56:918–926.

37. Szentes BL, Schwarzkopf L, Kirsch F, et al (2020) Measuring quality of life in COPD patients: Comparing disease-specific supplements to the EQ-5D-5L. Expert Rev Pharmacoecon Outcomes Res 20:523–529.

38. Watson L, Delure A, Qi S, et al (2021) Utilizing Patient Reported Outcome Measures (PROMs) in ambulatory oncology in Alberta: Digital reporting at the micro, meso and macro level. J Patient Rep Outcomes 5:97.

39. Mooney K, Whisenant MS, Beck SL (2019) Symptom care at home: A comprehensive and pragmatic PRO system approach to improve cancer symptom. Med Care 57:S66–S72.

Appendix

Additional file 1: Overview of design components and categories of components which laid the foundation for user and software producer interviews

	Components	Categories of components	Sources
General information			
	Type of disease	Chronic; Acute	1–9
	Setting	Inpatient; Outpatient; Combination	1,7,10–12
	Type of PROM	Disease-specific; Generic, Combination	8,13–27
	Key user	Physician; Other health care professionals (e.g., physiotherapist, nurse, etc.); Patient; Relatives	1,9,28
Data collection			
	Level of reporting	Micro (patient-physician communication and intra-patient comparison); Meso (comparison of patient groups within departments or institutions); Macro (comparison of patient groups across departments or institutions)	29
	Purpose of reporting	Shared decision-making (for patient and physician); Better basis for decision (for physician); Interpretation support of data (for physician)	30–32
	Data collection	Digital; Paper-based	6,7,29

	Components	Categories of components	Sources
	Time of data collection	Directly during appointment; Before the appointment (in waiting room); Independent at home	6,7
Dashboard			
	Patient information		1,33,34
	Clinical data		1,33,35,36
	Free write-in space		10,37
	Past assessment PRO score		5,10,36,37
	Peer-group comparison		2,36,38
	PRO-related goals		3,36
	Overall health-related goals		
	Alerts	Alerts should be presented immediately on the dashboard during the appointment if there is a critical value; Alerts should be presented during the appointment; No alerts should be included	6,37
	Customizability	The dashboard components should be tailored to individual needs; The dashboard components can be chosen from a standardized set; The components of the clinical dashboard are not customizable	3,9,33

Legend: Based on a literature search, we developed a list of design components and potential categories of components to be considered when building a clinical dashboard incorporating PROMs covering three grouping areas:(1) general information, (2) data collection, and (3) dashboard components. Each grouping area includes components and categories of components. For example, in the grouping area of "general

information", the component of "setting" means whether the dashboard is used in an in- or outpatient setting or in a combination of both. Additionally, for the components of "patient information", "clinical data", "free write-in space", "past assessment PRO score", "peer-group comparison", "PRO-related goals" and "overall health-related goals", the considered literature did not further specify on categories of components but only highlighted the importance of these components.

References:

1. Rudin RS, Perez S, Rodriguez JA, et al. User-centered design of a scalable, electronic health record-integrated remote symptom monitoring intervention for patients with asthma and providers in primary care. *Journal of the American Medical Informatics Association*. 2021;28(11):2433-2444. doi:10.1093/jamia/ocab157
2. Hartzler AL, Izard JP, Dalkin BL, Mikles SP, Gore JL. Design and feasibility of integrating personalized PRO dashboards into prostate cancer care. *Journal of the American Medical Informatics Association*. 2016;23(1):38-47. doi:10.1093/jamia/ocv101
3. Cronin RM, Conway D, Condon D, Jerome RN, Byrne DW, Harris PA. Patient and healthcare provider views on a patient-reported outcomes portal. *Journal of the American Medical Informatics Association*. 2018;25(11):1470-1480. doi:10.1093/jamia/ocy111
4. Elm JJ, Daeschler M, Bataille L, et al. Feasibility and utility of a clinician dashboard from wearable and mobile application Parkinson's disease data. *NPJ Digit Med*. 2019;2(1):95. doi:10.1038/s41746-019-0169-y
5. Fautrel B, Alten R, Kirkham B, et al. Call for action: How to improve use of patient-reported outcomes to guide clinical decision making in rheumatoid arthritis. *Rheumatol Int*. 2018;38(6):935-947. doi:10.1007/s00296-018-4005-5
6. Desantis D, Baverstock RJ, Civitarese A, Crump RT, Carlson K V. A clinical perspective on electronically collecting patient-reported outcomes at the point-of-care for overactive bladder. *Canadian Urological Association Journal*. 2016;10(11-12):359. doi:10.5489/cuaj.3757
7. Nicolas-Boluda A, Oppenheimer A, Bouaziz J, Fauconnier A. Patient-reported outcome measures in endometriosis. *J Clin Med*. 2021;10(21):5106. doi:10.3390/jcm10215106
8. Heath EL, Ackerman I, Lorimer M, et al. National implementation of an electronic patient-reported outcome measures program for joint replacement surgery: Pilot study. *JMIR Form Res*. 2022;6(4):e30245. doi:10.2196/30245

9. Hartzler AL, Chaudhuri S, Fey BC, Flum DR, Lavallee D. Integrating patient-reported outcomes into spine surgical care through visual dashboards: Lessons learned from human-centered design. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2015;3(2):2. doi:10.13063/2327-9214.1133

10. Baeksted C, Pappot H, Nissen A, et al. Feasibility and acceptability of electronic symptom surveillance with clinician feedback using the Patient-Reported Outcomes version of Common Terminology Criteria for Adverse Events (PRO-CTCAE) in Danish prostate cancer patients. *J Patient Rep Outcomes*. 2017;1(1):1. doi:10.1186/s41687-017-0005-6

11. Tan A, Durbin M, Chung FR, et al. Design and implementation of a clinical decision support tool for primary palliative Care for Emergency Medicine (PRIM-ER). *BMC Med Inform Decis Mak*. 2020;20(1):13. doi:10.1186/s12911-020-1021-7

12. DeMellow J, Kim TY. Technology-enabled performance monitoring in intensive care: An integrative literature review. *Intensive Crit Care Nurs*. 2018;48:42-51. doi:10.1016/j.iccn.2018.07.003

13. Nolan CM, Longworth L, Lord J, et al. The EQ-5D-5L health status questionnaire in COPD: Validity, responsiveness and minimum important difference. *Thorax*. 2016;71(6):493-500. doi:10.1136/thoraxjnl-2015-207782

14. Szentes BL, Schwarzkopf L, Kirsch F, Schramm A, Leidl R. Measuring quality of life in COPD patients: Comparing disease-specific supplements to the EQ-5D-5L. *Expert Rev Pharmacoecon Outcomes Res*. 2020;20(5):523-529. doi:10.1080/14737167.2019.1662302

15. Huber MB, Kurz C, Kirsch F, Schwarzkopf L, Schramm A, Leidl R. The relationship between body mass index and health-related quality of life in COPD: real-world evidence based on claims and survey data. *Respir Res*. 2020;21(1):291. doi:10.1186/s12931-020-01556-0

16. Merino M, Villoro R, Hidalgo-Vega Á, Carmona C. Health-related quality of life of patients diagnosed with COPD in Extremadura, Spain: results from an observational study. *Health Qual Life Outcomes*. 2019;17(1):189. doi:10.1186/s12955-019-1244-4

17. Smith SMS, Jan S, Descallar J, Marks GB. An investigation of methods to improve recall for the patient-reported outcome measurement in COPD patients: a pilot randomised control trial and feasibility study protocol. *Pilot Feasibility Stud*. 2019;5(1):92. doi:10.1186/s40814-019-0475-9

18. Ayala A, Forjaz MJ, Ramallo-Fariña Y, Martín-Fernández J, García-Pérez L, Bilbao A. Response mapping methods to estimate the EQ-5D-5L from the Western Ontario McMaster Universities Osteoarthritis in patients with hip or knee osteoarthritis. *Value in Health*. 2021;24(6):874-883. doi:10.1016/j.jval.2021.01.003

19. Baghbani-Naghadehi F, Armijo-Olivo S, Prado CM, Gramlich L, Woodhouse LJ. Does obesity affect patient-reported outcomes following total knee arthroplasty? *BMC Musculoskelet Disord*. 2022;23(1):55. doi:10.1186/s12891-022-04997-4

20. Bansback N, Trenaman L, MacDonald K V., et al. An individualized patient-reported outcome measure (PROM) based patient decision aid and surgeon report for patients considering total knee arthroplasty: protocol for a pragmatic randomized controlled trial. *BMC Musculoskelet Disord*. 2019;20(1):89. doi:10.1186/s12891-019-2434-2

21. Conner-Spady BL, Marshall DA, Bohm E, Dunbar MJ, Noseworthy TW. Comparing the validity and responsiveness of the EQ-5D-5L to the Oxford hip and knee scores and SF-12 in osteoarthritis patients 1 year following total joint replacement. *Quality of Life Research*. 2018;27(5):1311-1322. doi:10.1007/s11136-018-1808-5

22. Conner-Spady BL, Marshall DA, Bohm E, et al. Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. *Quality of Life Research*. 2015;24(7):1775-1784. doi:10.1007/s11136-014-0910-6

23. Eneqvist T, Nemes S, Kärrholm J, Burström K, Rolfson O. How do EQ-5D-3L and EQ-5D-5L compare in a Swedish total hip replacement population? *Acta Orthop*. 2020;91(3):272-278. doi:10.1080/17453674.2020.1746124

24. Haragus H, Prejbeanu R, Poenaru D V., Deleanu B, Timar B, Vermesan D. Cross-cultural adaptation and validation of a patient-reported hip outcome score. *Int Orthop*. 2018;42(5):1001-1006. doi:10.1007/s00264-017-3742-5

25. Rolfson O, Bohm E, Franklin P, et al. Patient-reported outcome measures in arthroplasty registries. *Acta Orthop*. 2016;87(sup1):9-23. doi:10.1080/17453674.2016.1181816

26. Sen RK, Shetti V, Mukhopadhyay R, et al. Satisfaction and health-related quality of life following hip and knee arthroplasty surgeries in Indian patients: A cross-sectional Study. *Indian J Orthop*. 2022;56(5):918-926. doi:10.1007/s43465-021-00589-x

27. Marshall DA, Jin X, Pittman LB, Smith CJ. The use of patient-reported outcome measures in hip and knee arthroplasty in Alberta. *J Patient Rep Outcomes*. 2021;5(S2):87. doi:10.1186/s41687-021-00362-6

28. Howarth M, Bhatt M, Benterud E, et al. Development and initial implementation of electronic clinical decision supports for recognition and management of hospital-acquired acute kidney injury. *BMC Med Inform Decis Mak*. 2020;20(1):287. doi:10.1186/s12911-020-01303-x

29. Watson L, Delure A, Qi S, et al. Utilizing Patient Reported Outcome Measures (PROMs) in ambulatory oncology in Alberta: Digital reporting at the micro, meso and macro level. *J Patient Rep Outcomes*. 2021;5(S2):97. doi:10.1186/s41687-021-00373-3

30. Gibbons C, Porter I, Gonçalves-Bradley DC, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. *Cochrane Database of Systematic Reviews*. 2021;2021(10). doi:10.1002/14651858.CD011589.pub2

31. Berry DL, Blumenstein BA, Halpenny B, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: A randomized trial. *Journal of Clinical Oncology*. 2011;29(8):1029-1035. doi:10.1200/JCO.2010.30.3909

32. Graupner C, Kimman ML, Mul S, et al. Patient outcomes, patient experiences and process indicators associated with the routine use of patient-reported outcome measures (PROMs) in cancer care: a systematic review. *Supportive Care in Cancer*. 2021;29(2):573-593. doi:10.1007/s00520-020-05695-4

33. Tsangaris E, Edelen M, Means J, et al. User-centered design and agile development of a novel mobile health application and clinician dashboard to support the collection and reporting of patient-reported outcomes for breast cancer care. *BMJ Surg Interv Health Technol*. 2022;4(1):119. doi:10.1136/BMJSIT-2021-000119

34. Taxter A, Johnson L, Tabussi D, et al. Co-Design of an Electronic Dashboard to Support Coproduction of Care in Pediatric Rheumatic Disease: Human-Centered Design and Usability Testing (Preprint). *J Particip Med*. Published online November 5, 2021. doi:10.2196/34735

35. Liu LH, Garrett SB, Li J, et al. Patient and clinician perspectives on a patient-facing dashboard that visualizes patient reported outcomes in rheumatoid arthritis. *Health Expectations*. 2020;23(4):846-859. doi:10.1111/hex.13057

36. Ragouzeos D, Gandrup J, Berrean B, et al. “Am I OK?” using human centered design to empower rheumatoid arthritis patients through patient reported

outcomes. *Patient Educ Couns*. 2019;102(3):503-510. doi:10.1016/j.pec.2018.10.016

37. Hassett MJ, Cronin C, Tsou TC, et al. eSyM: An electronic health record–integrated patient-reported outcomes-based cancer symptom management program used by six diverse health systems. *JCO Clin Cancer Inform*. 2022;(6):1-10. doi:10.1200/cci.21.00137

38. Strachna O, Cohen MA, Allison MM, et al. Case study of the integration of electronic patient-reported outcomes as standard of care in a head and neck oncology practice: Obstacles and opportunities. *Cancer*. 2021;127(3):359-371. doi:10.1002/cncr.33272

Additional file 2. Semi-structured interview guide for software producer and user interviews

1) General questions on dashboards and PROM usage	
Software Producer	**User**
1. Please describe the dashboard you are producing/designing. 2. In which countries or regions is your dashboard available? 3. What is the major goal of your clinical dashboard? a. For what type of work (e.g., analyzing outcomes over time, communicating with the patient, comparing to other groups, getting a better overview of all data collected) is the clinical dashboard designed for? b. What do you think is beneficial about using clinical dashboards? c. Where do you see major barriers in using clinical dashboards? 4. Do you incorporate PROMs in your clinical dashboard? If yes, which ones	1. What is your motivation for using clinical dashboards? a. For what type of work (e.g., analyzing outcomes over time, communicating with the patient, comparing to other groups, getting a better overview of all data collected) do you use the clinical dashboard? b. What do you like about the clinical dashboard you currently use? c. What don't you like about the clinical dashboard you currently use? 2. Do you use PROMs? If yes, which ones (standardized sets vs. own creation)? If no, why not? a. Do you use generic and disease-specific PROMs, or just one of

(standardized sets vs. own creation)? If no, why not? a. Does the user have the possibility to choose from a set of available PROMs or is it pre-defined by you? b. Do you use generic and disease-specific PROMs, or just one of each? Why? c. Why do you think it is beneficial to incorporate PROMs into clinical dashboards? d. What do you perceive as challenging when incorporating PROMs into clinical dashboards?	each? Why? b. Are PROMs already integrated into the clinical dashboard? c. What is your motivation for using PROMs? d. For what type of work (e.g., analyzing outcomes over time, communicating to patients, comparing to other groups) do you use PROMs? e. What do you like about using PROMs? f. What don't you like about using PROMs?
2) Questions about the market (penetration)	
Software Producer	**User**
1. How many clients do already use the clinical dashboards incorporating PROMs to communicate with patients? 2. How do you feel about the demand for these dashboards? 3. What designs are requested? 4. What is the general feedback from customers on your solution? What do they like - what don't they like? 5. Is there a scientific basis for the design of your dashboards? Are physicians involved in the development process?	*Not asked to users.*
3) Usage of the dashboard	
Software Producer	**User**
1. In how far does your product facilitate the workflow of practitioners?	1. To what extent does using the clinical dashboard make your day-to-day

2. Who uses and has access to the clinical dashboard? a. Do you have a special area of expertise, or can the dashboard be used at any discipline? b. Can it be used in an outpatient and inpatient setting? c. Can it only be used in one department/hospital or is it conceivable that it could also be used, for example, by outpatient care providers such as primary care physicians in parallel? 3. What type of data is available in the clinical dashboard? a. Do you receive regular feedback from your users on how valuable the included data is? b. Do you provide support in using the clinical dashboards? c. Do your customers own or rent the software of the clinical dashboard? 4. Do you think different episodes of care (e.g., orthopedics and COPD, one-time intervention vs chronic disease) require different dashboard capabilities? Why?	work easier? 2. Who uses and has access to the clinical dashboard? a. Is it only for use in your department/hospital or is it conceivable that it could also be used, for example, by outpatient care providers such as primary care physicians? 3. Is all the data you need in your clinical practice included in the clinical dashboard? a. What type of data is available in the clinical dashboard? b. In your opinion, is there a principle missing that you would find particularly valuable? c. What addition did you notice when using the clinical dashboard for the first time? d. Did you feel that using a clinical dashboard was complicated and could be made easier in the future? e. Do you find that using the clinical dashboard for the first time was intuitive? i. Have you received any support, or have you familiarized yourself with its use? 4. Do you think different episodes of care (e.g., orthopedics and COPD) require different dashboard capabilities? Why?
4) Dashboard development and data collection	

Software Producer	**User**
1. How do you develop a dashboard for a client? Do you have a basic product that is customizable to different needs, or do you develop it every time from scratch according to the needs of your client? a. How would you describe your working mode? Do you work agile or in a waterfall structure? b. How many iteration cycles do you go through until the delivery of the final product? 2. How does the roll-out of a new clinical dashboard work? 3. How does the data collection work? (Where, when, how, who, ...) a. Where do you see difficulties in data collection? 4. Where is the data stored? a. Do you have access to the data collected by your customers? 5. Is the collected data also used for other purposes than for the improvement in the patient-physician communication? a. E.g., aggregation of data and comparison between hospitals, other research purposes, etc.	1. Who provides the dashboard you use? a. How was it developed? b. Is it customizable to individual needs? 2. How was the clinical dashboard implemented in your organization? 3. How does the data collection work? (Where, when, how, who, ...) a. Where are difficulties in the data collection?
5) Dashboard Component Assessment	
Same questions for software producers and users From the literature, we have extracted some design principles that could be included in clinical dashboards - please provide your opinion on displaying these design principles and whether you already incorporate these design principles in your clinical dashboard: 1. Past PROM assessments (including evolution over time). a. What kind of scores should the dashboard display?	

Values for
individual dimensions or index scores? Why?
2. Future PROM-related goals
3. Overall health-related goals
4. Benchmarking | peer-group comparison
5. Alerts on symptom/symptom change warnings
   a. If yes, at what time should the alarm/warning appear?
6. Patient information (patient photo, demographic information, recent health updates, and contact information for other care team members)
7. Clinical data (laboratory results and drug data)
   a. What medical data do you use?
8. Free write-in area
9. In your opinion, is a design principle missing?

6) Role of the patient

Software Producer	**User**
1. What is the role of the patient? a. Does the patient also have access to the information on the dashboard? b. Do you assist the patient in interpreting the data? c. Do you think the patient's needs concerning the dashboard differ whether he/she underwent a one-time intervention or whether he/she suffers from a chronic disease? d. To what extent is the information different from that displayed to the health professional/physician?	1. What is the role of the patient? a. Does the patient also have access to the information on the dashboard? b. Should the patient be assisted in interpreting the information? c. To what extent is the information different from that displayed to the health professional/physician?

End

Is there anything else you would like to add to conclude the interview?

Additional file 3. Consolidated Criteria for Reporting Qualitative Research (COREQ)

No. Item	Guide questions/description	Reported on Page #
Domain 1: Research team and reflexivity		
Personal Characteristics		
1. Interviewer/facilitator	Which author/s conducted the interview or focus group?	Page 4
2. Credentials	What were the researcher's credentials? E.g. PhD, MD	Page 4
3. Occupation	What was their occupation at the time of the study?	Page 4
4. Gender	Was the researcher male or female?	Page 4
5. Experience and training	What experience or training did the researcher have?	Page 4
Relationship with participants		
6. Relationship established	Was a relationship established prior to study commencement?	Page 4
7. Participant knowledge of the interviewer	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	Page 4
8. Interviewer characteristics	What characteristics were reported about the interviewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	Page 4
Domain 2: study design		
Theoretical framework		
9. Methodological orientation and Theory	What methodological orientation was stated to underpin the study? e.g.	Page 4

	grounded theory, discourse analysis, ethnography, phenomenology, content analysis	
Participant selection		
10. Sampling	How were participants selected? e.g. purposive, convenience, consecutive, snowball	Page 4
11. Method of approach	How were participants approached? e.g. face-to-face, telephone, mail, email	Page 4
12. Sample size	How many participants were in the study?	Page 5
13. Non-participation	How many people refused to participate or dropped out? Reasons?	n/a
Setting		
14. Setting of data collection	Where was the data collected? e.g. home, clinic, workplace	Page 4
15. Presence of non-participants	Was anyone else present besides the participants and researchers?	No
16. Description of sample	What are the important characteristics of the sample? e.g. demographic data, date	Pages 4&5, Multimedia Appendix 4
Data collection		
17. Interview guide	Were questions, prompts, guides provided by the authors? Was it pilot tested?	Page 4
18. Repeat interviews	Were repeat interviews carried out? If yes, how many?	No
19. Audio/visual recording	Did the research use audio or visual recording to collect the data?	Page 4
20. Field notes	Were field notes made during and/or after the interview or focus group?	Page 4
21. Duration	What was the duration of the inter views or focus group?	Page 5, Multimedia Appendix 4

22. Data saturation	Was data saturation discussed?	Page 4
23. Transcripts returned	Were transcripts returned to participants for comment and/or correction?	Page 4
Domain 3: analysis and findings		
Data analysis		
24. Number of data coders	How many data coders coded the data?	Page 5
25. Description of the coding tree	Did authors provide a description of the coding tree?	Page 10 (Figure 1), Multimedia Appendix 5
26. Derivation of themes	Were themes identified in advance or derived from the data?	Page 5
27. Software	What software, if applicable, was used to manage the data?	Page 5
28. Participant checking	Did participants provide feedback on the findings?	No
Reporting		
29. Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	Pages 7-10
30. Data and findings consistent	Was there consistency between the data presented and the findings?	Yes
31. Clarity of major themes	Were major themes clearly presented in the findings?	Pages 7-10, Multimedia Appendix 5
32. Clarity of minor themes	Is there a description of diverse cases or discussion of minor themes?	Pages 7-10, Multimedia Appendix 5

Additional file 4. Overview of software producer and user interviewee's characteristics

Interview ID	Software producer/ User (gender)	Function	Country	Interview duration
I1	Software producer (male)	Product Owner	Netherlands	45 mins
I2	Software producer (female)	Product Delivery Manager	Netherlands	43 mins
I3	Software producer (male)	Digital strategy lead	Switzerland	29 mins
I4	Software producer (male)	CEO & founder	Germany	36 mins
I5	Software producer (male)	Director Smart Health Academy & Patient Reported Outcomes	Germany	44 mins
I6	Software producer (female)	Chief Medical Officer	Germany	33 mins
I7	User (male)	Chief physician at department of orthopedics	Switzerland	29 mins
I8	User (male)	Chief hospital innovation officer	Switzerland	28 mins
I9	User (male)	Chief physician at department of pneumology	Switzerland	30 mins
I10	User (female)	Member of hospital quality management	Switzerland	34 mins
I11	User (male)	Chief physician at department of pediatrics and rheumatology	Netherlands	36 mins
I12	User (male)	Chief physician at department of pneumology	Switzerland	35 mins
I13	User (female)	Chief physician at department of orthopedic surgery	Switzerland	34 mins

Interview ID	Software producer/ User (gender)	Function	Country	Interview duration
		and traumatology		
I14	User (female)	Chief physician at department of orthopedics	Germany	21 mins
I15	User (male)	Chief physician at department of pneumology	Germany	20 mins
I16	User (female)	Chief physician at department of cancer care	USA	24 mins

Additional file 5. Output of thematic coding

	Software Producer (SP)						User (U)										Interview guide	
	I1	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**I8**	**I9**	**I10**	**I11**	**I12**	**I13**	**I14**	**I15**	**I16**	**SP**	**U**
Added value	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**	**X**	**1.3a & 1.4c**	**1.1a & 1.2d**
- comparability			X	X		X		X			X		X	X		X		
- efficiency	X	X		X		X	X		X	X	X				X	X		
- improvements						X	X							X		X		
- patient satisfaction		X			X	X		X								X		
- visualisation		X	X	X	X	X	X	X		X	X		X	X	X	X		
Alert		**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**5.5**	**5.5**
- during appointment		X	X		X	X				X				X	X			
- no								X	X		X	X						
- real time		X	X	X	X								X			X		
Awareness on barriers	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**1.1c & 1.3c**	**1.1c & 1.2f**
- burdensome collection of PROs		X					X	X			X	X		X		X		
- interoperability		X	X	X		X	X	X	X	X	X	X		X		X		
- lack of good data							X			X			X	X				
- lack of motivation of users	X	X														X		
- legal consequences	X	X	X	X			X			X	X							
- licensing of questionnaires		X			X	X												
- not intuitive to use								X							X	X		
- various interests							X		X									
Clinical data		**X**	**X**	**X**		**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**5.7**	**5.7**

	Software Producer (SP)						User (U)										Interview guide	
	I1	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**I8**	**I9**	**I10**	**I11**	**I12**	**I13**	**I14**	**I15**	**I16**	**SP**	**U**
- clinical data very useful		X	X	X		X			X				X	X	X			
- not included								X		X	X	X				X		
Customizability		**X**		**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**4.1**	**4.1b**
- from a standard set		X			X	X	X	X		X	X		X	X		X		
- to individual needs				X	X							X			X			
Iterative dashboard development	**X**	**X**	**X**			**X**		**X**		**X**	**X**		**X**			**X**	**4.1b**	**4.1a**
Data storage			**X**	**X**	**X**						**X**						**4.4**	**4.3**
- At producer			X		X													
- Cloud			X	X														
- in clinic				X							X							
Different needs btw diseases			**X**	**X**	**X**	**X**		**X**		**X**				**X**			**3.4**	**3.4**
- no				X	X	X		X		X				X				
- yes			X															
- different needs in various countries				X														
Data collection: How	**X**	**X**		**X**	**X**	**X**	**X**		**X**	**X**	**X**	**X**	**X**	**X**		**X**	**4.3**	**4.3**
- digital	X	X		X	X	X	X			X	X		X			X		
- paper-based									X			X	X	X				
Data collection: Where	**X**	**X**		**X**	**X**	**X**	**X**		**X**	**X**	**X**	**X**	**X**	**X**		**X**	**4.3**	**4.3**
- in waiting room		X				X	X			X	X	X	X	X		X		
- independent at home	X	X		X	X				X	X	X	X	X			X		

	Software Producer (SP)						User (U)										Interview guide	
	I1	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**I8**	**I9**	**I10**	**I11**	**I12**	**I13**	**I14**	**I15**	**I16**	**SP**	**U**
Free write-in area		**X**		**X**	**X**	**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**5.8**	**5.8**
- not useful					X	X		X		X			X			X		
- neutral														X				
- rather useful																		
- very useful		X													X			
- yes, but specific question				X		X			X		X	X						
Key user	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**	**X**		**X**	**3.2**	**3.2**
- physician	X	X	X	X	X	X	X		X		X	X	X	X		X		
- other health care professionals	X	X			X	X	X		X							X		
- patient			X	X	X	X		X				X	X			X		
- relatives													X					
Level of reporting	**X**	**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**		**X**	**1.3 & 4.5a**	**1.1a**
- macro	X	X												X				
- meso	X						X	X			X			X				
- micro		X		X	X	X	X	X	X	X	X	X		X		X		
Overall health related goals	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**		**X**	**X**	**X**	**X**		**X**	**5.3**	**5.3**
- less useful														X				
- neutral									X							X		
- not included		X			X	X		X				X	X					
- rather useful																		
- very useful	X		X	X							X							
Past assessment score	**X**	**X**		**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**5.1**	**5.1**

	Software Producer (SP)						User (U)										Interview guide	
	I1	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**I8**	**I9**	**I10**	**I11**	**I12**	**I13**	**I14**	**I15**	**I16**	**SP**	**U**
- very useful	X	X		X		X	X	X	X	X	X	X	X	X		X		
- not useful															X			
Patient information		**X**	**X**	**X**	**X**	**X**		**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**5.6**	**5.6**
- no					X			X		X			X					
- yes		X	X	X		X					X	X		X	X	X		
Patient perspective		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**6.1a-d**	**6.2a-c**
- all PRO results shown		X		X				X										
- better interpretation of own health status		X	X		X			X	X	X			X	X		X		
- clinical data not to be shown						X		X										
- different processing of data for physician and patient															X	X		
- more explanation on the meaning of the data		X		X	X			X		X		X				X		
- no							X							X				
- no additional help for PROM fill-out		X						X		X								
- visualisation of scores		X			X				X	X	X					X		
Payment model			**X**	**X**	**X**	**X**	**X**										**3.3c**	**(not explicitly covered in the interview guide)**
- add-on to product (for free)			X				X											
- license				X	X	X	X											
Peer-group comparison	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**5.4**	**5.4**

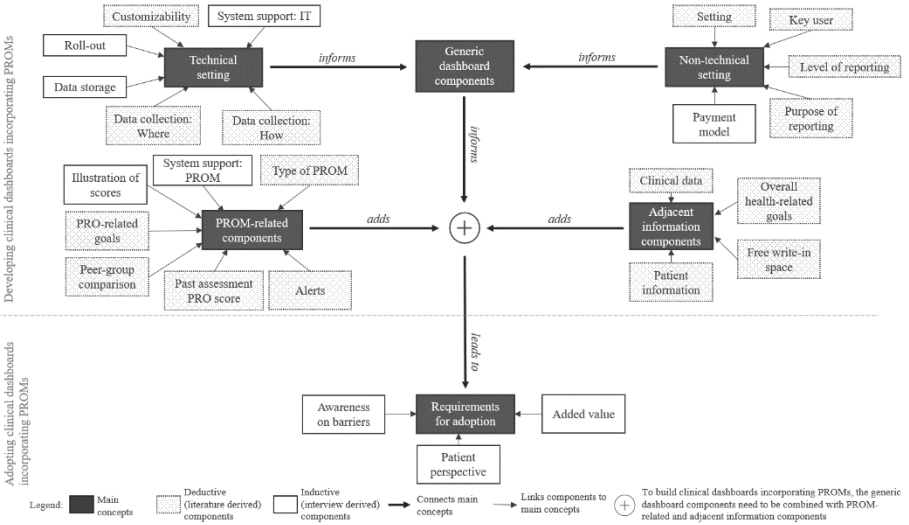
	Software Producer (SP)						User (U)										Interview guide	
	I1	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**I8**	**I9**	**I10**	**I11**	**I12**	**I13**	**I14**	**I15**	**I16**	**SP**	**U**
- no									X			X			X			
- yes	X	X	X	X	X	X	X	X					X	X		X		
- yes, but currently not possible										X	X							
PRO-related goals		**X**		**X**	**X**	**X**		**X**			**X**	**X**	**X**	**X**	**X**	**X**	**5.2**	**5.2**
- neutral																X		
- not included		X		X	X			X			X	X	X		X			
- rather useful						X												
- very useful														X				
Purpose of reporting	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**	**X**		**X**	**1.3**	**1.1**
- better basis for physician's decision		X	X		X	X		X	X		X	X	X					
- improved communication	X	X	X	X	X	X	X	X	X		X	X	X	X		X		
- real-time tracking			X	X														
- shared decision-making				X		X	X	X										
Roll-out				**X**		**X**	**X**		**X**	**X**				**X**			**4.2**	**4.2**
- step by step (in different clinics)				X		X	X		X	X								
- top down										X				X				
Scores				**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**5.1a**	**5.1a**
- both				X		X		X			X			X		X		
- dimensional							X			X			X					
- index									X			X						
Setting		**X**	**X**	**X**	**X**		**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**3.2b**	**3.1&2a**
- combination					X					X					X	X		

	Software Producer (SP)						User (U)										Interview guide	
	I1	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**I8**	**I9**	**I10**	**I11**	**I12**	**I13**	**I14**	**I15**	**I16**	**SP**	**U**
- inpatient							X											
- outpatient		X	X	X					X		X	X	X	X				
System support: IT		**X**	**X**	**X**	**X**				**X**	**X**			**X**			**X**	**3.3b**	**3.3d & 3.3e**
- in-house at corresponding institution									X									
- introduction		X	X						X	X						X		
- no					X								X					
System support: PROM		**X**	**X**	**X**						**X**				**X**			**3.3b**	**3.3d & 3.3e**
- interpretation of PRO scores		X	X	X						X				X				
Type of disease	**X**	**X**	**X**	**X**	**X**		**X**		**X**		**X**	**X**	**X**	**X**	**X**	**X**	**1.1**	**1.1 (Derived from specialization)**
- chronic		X							X		X				X	X		
- both	X		X	X	X							X	X					
- one-time intervention							X							X				
Type of PRO	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**1.4b**	**1.2a**
- combination		X		X	X	X	X			X		X	X	X				
- disease-specific	X		X					X	X		X					X		
- generic - only for research											X							

Legend: SP = Software producer, U = User, btw = between

The last column referring to the interview guide (for more information, please see Supplementary Appendix 2) indicates which question corresponds to the thematic coding output. Exemplarily, for "type of PRO" questions 1.4b in the software producer guide and 1.2a in the user guide are applicable. This means that the question is allocated to "1) General questions on dashboards and PROM usage" and refers to the specific question 2 respectively 4 in the corresponding software producer and user guide. Section 5 of the interview guide assessed the different dashboard components. Therefore, in this coding output, we display the assessment categorically of how valuable the interviewees rate the individual components. This categorization applies to the following components: past assessment score, PROM-related goals, overall health-related goals, peer-group comparison, patient information, clinical data, and free write-in area.

Additional file 6. Explanation of topics



The list of 25 design principles indicates what needs to be considered when developing a clinical dashboard incorporating PROMs. This appendix provides a guiding question and a non-conclusive list of potential features for each design principle.

Technical setting:

- **System support: IT*:** How does the software producer support the system use?
    - E.g., in-house support by the corresponding institution, introduction to the system use, workshop and webinars, or no support
- **Customizability:** What degree should customizability be possible/allowed?
    - E.g., customizability to individual needs, customizability from a standard set, or no customizability
- **Roll-out*:** How should the roll-out of the clinical dashboard be conducted?
    - E.g., step-by-step or top-down
- **Data storage*:** Where should the data be stored?
    - E.g., at software producer, in a cloud solution, in the corresponding institution (i.e., clinic)
- **Data collection:** Where: When/Where should the collection take place?
    - E.g., directly during the appointment, before the appointment in the waiting room, or independent at home

- **Data collection: How:** How should the data collection take place?
    - E.g., digital or paper-based

Non-technical setting:

- **Setting:** In which setting is the clinical dashboard used?
    - E.g., inpatient, outpatient, or combination
- **Key user:** Who is the key user of the clinical dashboard?
    - E.g., all kind of physicians, other health care professionals (such as nurses or physiotherapists), patients, relatives, or third parties (such as hospital quality manager)
- **Level of reporting:** What is the preferred/chosen level of reporting?
    - E.g., micro (patient-physician communication and intra-patient comparison), meso (comparison of patient groups within departments or institutions), or macro (comparison of patient groups across departments or institutions)
- **Purpose of reporting:** For which activity is the data collected? What does the clinical dashboard serve for?
    - E.g., shared decision-making, better basis for decision, improved communication, or real-time tracking
- **Payment model*:** How does the payment model of the clinical dashboard look like?
    - E.g., add-on to another product (i.e., for free) or license

PROM-related components:

- **Type of PROM:** What kind of PROMs are used in the clinical dashboard?
    - E.g., disease-specific (such as CAT, WOMAC, HOOS, KOOS, etc.), generic (such as EQ-5D, PROMIS, WHO-5, etc.), or combination
- **System support: PROM*:** Does the software producer provide any support in analyzing or interpreting the PROM data?
- **Illustration of scores*:** How are the PRO scores visualized?
    - E.g., index, dimensions, or both
- **PRO-related goals:** Are PRO-related goals included in the dashboard?
- **Peer-group comparison:** Is a peer-group comparison possible?
- **Past assessment PRO score:** Can the evolution of the PRO scores be tracked over time?
- **Alerts:** Should alerts be incorporated in/displayed by the clinical dashboard?

  - E.g., immediately when critical values appears, during appointment, or no alerts

Adjacent information components:

- **Clinical data:** Is clinical data such as lab results or medication information included in the dashboard?
- **Overall health-related goals:** Are overall health related goals included in the dashboard?
- **Free write-in space:** Is it possible to note additional information?
  - E.g., yes, only to specific questions/items, or not possible
- **Patient information:** Is patient information such patient photograph, demographic information, or recent health updates included in the dashboard?

Requirements for adoption:

- **Awareness on barriers*:** What potential barriers might hinder the adoption of clinical dashboards?
  - E.g., burdensome collection of PROs, interoperability, lack of good data, lack of motivation of users, legal consequences, licensing of questionnaires, dashboard not intuitive to use, various interest of different stakeholders
- **Patient perspective*:** What data should be provided to the patient?
  - E.g., all PRO results shown, better interpretation of own health status, clinical data not to be shown, different processing of data for physician and patient, more explanation on the meaning of the data, no additional help for PROM fill-out, visualization of scores, no data to be provided
- **Added value*:** What value can the use of a clinical dashboard incorporating PROMs add?
  - E.g., comparability between PRO scores, efficiency in workflow, improvements for physicians performance (personal learning), enhanced patient satisfaction, visualization of data

*all marked topics are derived by inductive coding, i.e., are interview-based.

5 Factors influencing health literacy

5.1 The power of knowledge: A survey on COPD patients' health literacy in Switzerland

Anja Bischof

Published as a Working Paper in *Schriftenreihe in Health Economics, Management and Policy*

5.1.1 Abstract

Background: Health literacy (HL) is one of the major levers to increase the self-management capabilities of patients suffering from Chronic Obstructive Pulmonary Disease (COPD). However, HL is often insufficient in individuals suffering from COPD. Therefore, this study investigates which factors are associated with HL and whether patients with varying HL states desire different support in digital health interventions to better manage COPD-specific symptoms.

Methods: In total, 203 participants completed a web-based survey that collected information on patient's health-related quality of life (HRQoL), possible impairments, and areas of improvement in their daily lives. The HLS-EU-Q16 was applied to assess HL. A chi-square test and an analysis of variances were per-formed to compare variable characteristics across COPD patients with different HL. An ordinal logistic regression was used to identify variables significantly associated with HL. Based on the survey's responses, the patients' desire for daily support were identified.

Results: Of the 203 participants, 53.2% demonstrated sufficient HL, 27.6% indicated problematic HL, and 19.2% displayed inadequate HL. Significant differences in the requirement of support in daily life (χ^2=0.032), length of diagnosis (χ^2=0.015), and confidence in medication intake (χ^2<0.001) are observable. Furthermore, significant differences in age (p<0.001), EQ-5D-5L (p<0.001), EQ-VAS (p<0.001), and CAT (p=0.004) were observable. The ordinal logistic regression revealed age and the EQ-5D-5L as significantly associated with HL. COPD patients expressed consistent preferences for support in digital health interventions, irrespective of their HL status.

Conclusion: The study reveals that age and HRQoL, represented by the EQ-5D-5L, are significantly associated with HL. Approaches to improving HRQoL might be considered to strengthen HL. Additionally, the study identifies desired support mechanisms for COPD patients when developing digital health interventions for this patient group. Future research could explore the interplay between HRQoL, HL, and self-management and the potential of digital health intervention in this interplay.

Keywords: Chronic Obstructive Pulmonary Disease (COPD), Switzerland, Health-related quality of life (HRQoL), Health literacy

5.1.2 Background

Chronic Obstructive Pulmonary Disease (COPD) is the third leading cause of death worldwide in 2019 [1]. In Switzerland, around 400,000 patients suffer from the chronic lung disease [2], which is approximately 5% of the total population. In former times, male patients were more prone to suffer from COPD. However, the prevalence in women has increased and equaled the level of men in 2008 [3]. Typical symptoms are shortness of breath, cough, and sputum [4]. COPD causes an enormous economic burden of CHF 603-847 million in direct costs and CHF 932 million in indirect costs annually in Switzerland [5]. Individuals with a low socioeconomic status are especially prone to the dis-ease [6]. Due to the chronic nature of the disease, patients suffering from COPD must know how to deal with the disease during their daily lives. This ability is known as self-management, which relates to "an individual's ability to detect and manage symptoms, treatment, physical and psychosocial consequences, and lifestyle changes inherent in living with a chronic condition" [7]. Furthermore, patient-physician communication is important to foster a basic understanding of the disease, however physicians alone cannot mitigate the problem of non-compliance to treatment due to low self-management capabilities [8].

In the case of COPD, self-management comprises regular physical activity, medication adherence, and smoking cessation [4]. The main aim of self-management is improving or maintaining physical health, decreasing symptom impact and impairments, increasing well-being from an emotional, social, and psychological perspective, and creating effective ties with family, practitioners, and community [9]. One major driver of improved self-management capabilities is health literacy (HL) [10].

HL encompasses the extent of one's capability to find, understand, and effectively use health information and health services while promoting health and preventing disease [11]. HL is assessed by various internationally validated self-assessment instruments such as the Health Literacy Questionnaire (HLQ) [12], the European Health Literacy Survey Questionnaire (HLS-EU-Q) [13], or the Single Item Literacy Screener (SILS) [14]. HL is especially low among COPD patients [15]. Low HL is associated with

worse health outcomes in COPD [16], low medication adherence [8], and limited self-management capabilities [15].

However, supporting patients to navigate through the health care system and misunderstanding information are structural problems in the patient-physician communication. Physicians and nurses often overestimate patients' HL skills by assuming that the patients can process complex instructions [17–19]. This problem worsens as patients also tend to overestimate their HL [20]. In addition to verbal communication, written disease-related education material remains inaccessible to patients with low HL due to excessively high reading prerequisites or insufficient quality [21, 22]. Furthermore, the literature highlights that patients usually have limited knowledge about their disease [23, 24], directly related to insufficient disease management. To tackle the challenge of patients' limited health knowledge, self-management resources such as virtual chatrooms are established to strengthen conversations between COPD patients and various stakeholders [25], or information material in facilitated language is provided, such as in Germany [26].

To support COPD patients better in their daily dealing with the disease and the corresponding symptoms, digital health interventions (DHI) are a potential approach. DHIs cover mHealth [27, 28], telemonitoring [29], and telemedicine. This allows providers to remotely assess the patient's health status and adapt treatments to individual needs [30, 31]. Furthermore, the patients can access health information themselves and independently monitor their personal condition [27]. Puente-Maestu et al. [32] argue that HL needs to be considered when planning the care of COPD patients – especially when taking a patient-centered approach. Nevertheless, most interventions do not measure HL as an outcome [10]. The principal aim of the DHIs must be to avoid COPD exacerbations (i.e., acute worsening of COPD) leading to hospitalizations and increased mortality [33].

Facilitators for DHI adoption are improved disease understanding and management [34]. Additionally, scalability of interventions is needed to allow DHIs to be cost-effective [33]. Therefore, it becomes crucial to understand better the specific needs of support raised by patients suffering from COPD. As HL highly influences self-

management capabilities [35], the intervention's requirements must be distinguished between different HL states. Being aware of different needs between HL states allows for developing more suitable interventions and enhances scalability across patients with different HL states.

Thus, I investigate the following questions in this paper:

- Which factors are associated with higher health literacy?
- What types of support do patients with varying health literacy states desire in digital health interventions to better deal with COPD?

This paper analyzes where differences in HL exist and investigates which factors influence HL. Furthermore, it provides insights on whether patients suffering from COPD desire different types of support to manage their daily lives actively. Understanding this will be a crucial part of building DHIs in the future. This is the first study investigating the need to support COPD patients by directly collecting responses from a tailored survey distributed in Switzerland.

5.1.3 Methods

Survey design

We designed an online survey using Unipark [36], an online survey software for research conducted at universities and other educational institutions. The survey was developed based on current literature and in collaboration with a chief pulmonologist and a digital health expert. The questionnaire contained different sections dealing with patient's health-related quality of life (HRQoL), possible impairments, and areas of improvement in their daily lives and was structured into five sections: general information, disease-specific questions, need for support, HRQoL, and dealing with COPD. In the questionnaire, we used internationally validated instruments for measuring patients' HRQoL (EQ-5D-5L and EQ-VAS questionnaires) [37–39], patients' disease-specific condition (COPD Assessment Test questionnaire) [39–41], and patients' HL (HLS-EU-Q16) [32, 42] (the survey questions are provided in Supplementary Material 1). The chosen HL questionnaire is the short-form version of

the HLS-EU-Q. The instrument collects information on four main HL dimensions: access, understandability, evaluation, and use of health information in the areas of disease prevention and health promotion [13].

The survey was available in German, French, and Italian. Native speakers translated the German version into French and Italian. Furthermore, eight participants checked the survey for understandability and practicability before the collection period. These eight participants were not included in the final sample.

The data was stored on password-protected computers. The order of questions was the same for all respondents. The survey, furthermore, included adaptive items, for example, for smoking and further comorbidities, which were only presented if the participant answered previous questions accordingly. At the bottom of every page, the participant could click a “back” button to adjust responses on previous pages if needed. The survey was designed and reported following the Checklist for Reporting Results of Internet E-Surveys [43].

Ethical approval and informed consent

This survey was submitted to the Ethics Committee Eastern Switzerland and determined to be exempt according to Art 2 (2c) on the Swiss Human Research Act on March 24, 2022. All participants completed the survey voluntarily. The front page stated that the participant’s data would be used anonymously, explained the research purpose, and listed host organization details. Before starting the survey, the participants had to agree on having understood the information on the front page, which served as electronic consent. We only collected and stored data anonymously.

Participants

All patients suffering from COPD and living in Switzerland were eligible for participation in the survey. One question dealt with the participant’s disease, and the answers were excluded if the participant did not suffer from COPD.

Recruitment

We applied an open survey design and therefore allowed for a convenience sample. The survey was advertised on Facebook in German, French, and Italian (see Supplementary Material 2). A particular challenge of the recruitment was the anonymous collection of the responses. To further promote the recruitment, the following care providers and organizations were contacted and asked to distribute flyers (see Supplementary Material 3): general practitioners, pulmonologists, pulmonology departments in hospitals, pharmacies, COPD support groups, rehabilitation centers, cantonal lung leagues, and physician networks. However, potential participants were never contacted directly.

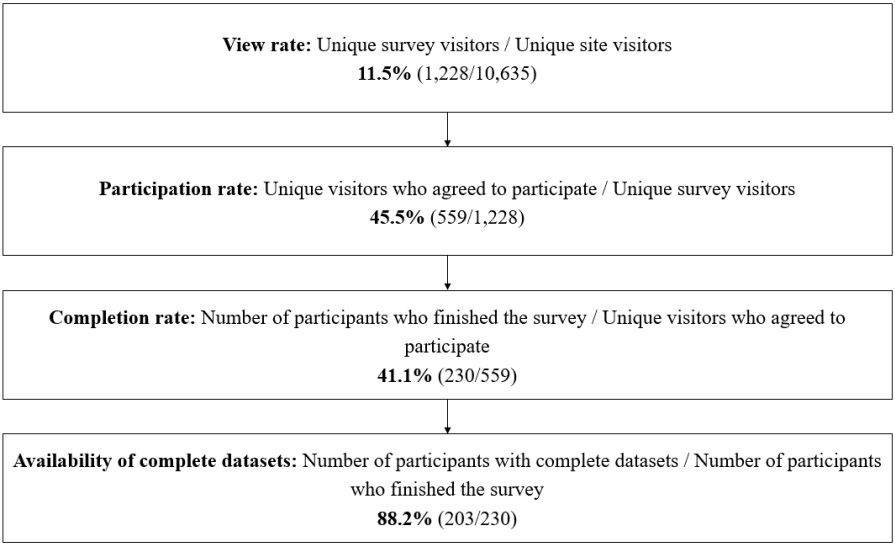
Administration

A link to the survey was provided in the Facebook ad, and a QR code on the flyers directed the participants directly to the survey's landing page. The survey was not administered via email. Participants responded voluntarily, and a lottery with three Swiss holiday (Reka) vouchers for CHF 100 was offered. In addition, participants could reach out to the survey hosts via email to be informed about the study results. The response collection period was from July 1, 2022, to July 31, 2023. The questionnaire consisted of 66 questions distributed over six survey pages. From the server side, all questions were marked as mandatory, and the participant could not reach the next survey page if not all questions were answered on the previous page. IP addresses were not visible for the survey hosts. However, the survey platform Unipark stored IP addresses to deny access to participants if they had already previously filled out the survey.

Response rates

Unipark, the chosen survey collection platform, measured the number of visitors and participants directly. The survey view rate was 11.5% (1,228/10,635), the participation rate was 45.5% (559/1,228), and the completion rate for the overall survey was 41.1% (230/559). In total, 203 survey participants were included in the sample (see Figure 5.1).

Figure 5.1: Survey response rates



Statistical analysis

To calculate HL for the individual respondents, the thresholds were chosen according to Röthlin et al. [44]. Thereby, three states of HL are possible: "sufficient" (scores 13-16), "inadequate" (scores 9-12), and "problematic" (scores 1-8). First, to characterize the study population, mean values and standard deviations (95% confidence intervals) for continuous variables, and frequencies and percentages for categorical variables are provided. Second, to analyze differences in the patients' characteristics of participants between HL states, a chi-square test of homogeneity was used for categorical variables, whereas an analysis of variance (ANOVA) was performed to compare the central tendencies of continuous variables. Third, a cumulative odds ordinal logistic regression with proportional odds was applied. As potential confounders, only variables with a $p<0.05$ in the chi-square test and the ANOVA were included in the regression. The regression was performed to identify significantly associated variables with HL. All analyses were conducted with IBM-SPSS version 28 [45].

To analyze whether differences in the desire for daily support for dealing with COPD-specific challenges depend on HL states, three areas were inquired: 1) difficulties with COPD, 2) desire type of support, and 3) support with COPD-specific symptoms (see Supplementary Material 1, Questions 27, 28, and 35). The participants had the possibility to choose more than one answer for each question. Subsequently the frequency of the responses was counted to establish a ranking.

5.1.4 Results

Descriptives

Of the 203 participants, 108 showed sufficient HL, 56 indicated problematic HL, and 39 displayed inadequate HL (see Table 5.1). Following this, only 53.2% of all participants have a satisfactory HL state. Concerning gender distribution, 128 female, 73 male, and two diverse participants were included. Most of the participants (52.2%) indicated that a professional training is their highest degree of education, followed by compulsory school (17.7%) and higher education (16.7%). 55.7% of the participants indicated a COPD status B, while 32.5% indicated a D-status. Around 28% of the participants claimed they depend on daily support.

38.4% of the participants expressed that they are still smoking despite being diagnosed with COPD, whereas more than half quit smoking. 47.3% of the participants have suffered from COPD for more than five years, and 35% have already been hospitalized due to a COPD exacerbation. While 84.2% of the participants must take medication to cope with COPD-specific symptoms, 22.6% feel insecure about correctly taking the medication. For treatment and further information on COPD, the primary contact person is the specialist or pulmonologist in 52.7% of all cases. 10.3% of participants indicated that they regularly exchange with other COPD patients, and 11.8% participated in a program designed to learn more about COPD and how to deal with it in daily life (called “Living better with COPD”). More than one fourth have already been to outpatient pulmonary rehabilitation, and 57.1% are yearly vaccinated against the flu.

Furthermore, the average participant was 64.9 years old and had a BMI of 26.7 – thus being slightly overweight. 31 participants expressed that they suffer from comorbidities. On average, these participants suffered from 2.68 comorbidities. Regarding self-perceived health status, the EQ-index score showed an average of 0.724. In contrast, the EQ-VAS displayed an average of 61.26. For the disease-specific PRO – measured by CAT – the mean was at 18.88, indicating a medium impact level of the disease on the quality of life. The mean time participants spent filling out the survey was 21 minutes 4 seconds, while the median was 19 minutes 38 seconds. All responses with a time stamp below 12 minutes were excluded.

Table 5.1: Descriptives of the survey participants

Variable	Variable characteristics	Number	(%)
Health Literacy	Sufficient	108	53.2
	Inadequate	56	27.6
	Problematic	39	19.2
Gender	Male	73	36.0
	Female	128	63.1
	Diverse	2	1.0
Education	No school	5	2.5
	Compulsory school	36	17.7
	Professional training	106	52.2
	Matura	8	3.9
	Higher education	34	16.7
	University degree	14	6.9
COPD status	A	19	9.4
	B	113	55.7
	C	5	2.5
	D	66	32.5
Support in daily life	Yes	57	28.1
	No	146	71.9
Smoking	Yes, daily	61	30.0
	Yes, sometimes	17	8.4
	No longer	104	51.2
	Never	21	10.3
Length of diagnosis	Less than 1 year	33	16.3
	1 to less than 3 years	33	16.3
	3 to less than 5 years	41	20.2
	5 to less than 7 years	45	22.2
	7 to less than 10 years	23	11.3
	Over 10 years	28	13.8
Hospitalized exacerbation	Yes	71	35.0
	No	132	65.0

Taking COPD-specific medication	Yes	171	84.2
	No	32	15.8
Feeling secure in medication intake	Very insecure	24	11.8
	Insecure	22	10.8
	Neutral	37	18.2
	Secure	43	21.2
	Very secure	76	37.4
	I do not take medication	1	0.5
Primary contact person	GP	77	37.9
	Specialist/Pulmonologist	107	52.7
	Hospital	11	5.4
	Other	8	3.9
Exchange with other patients	Yes	21	10.3
	No	182	89.7
Participation at "Living better with COPD"	Yes	24	11.8
	No	179	88.2
Participation at outpatient pulmonary rehabilitation	Yes	55	27.1
	No	148	72.9
Yearly vaccination against the flu	Yes	116	57.1
	No	87	42.9

	Mean	**SD**
Age	64.93	10.02
BMI	26.71	6.63
Number of comorbidities*	2.68	1.54
EQ-5D-5L	0.724	0.233
EQ-VAS	61.26	22.38
CAT	18.88	7.67

Legend: SD = Standard deviation; BMI = Body Mass Index; CAT = COPD Assessment Test

*only 31 participants indicated that they were suffering from comorbidities – therefore, mean and standard deviation were calculated for these 31 participants

Differences in distribution of variables' characteristics regarding health literacy

I grouped the participants according to their HL into sufficient, inadequate, and problematic and com-pared the distribution of variables' characteristics across the groups and whether these distributions were significantly different. Significant differences in the requirement of support in daily life (χ^2=0.032), length of diagnosis (χ^2=0.015), and confidence in medication intake (χ^2<0.001) were observable (see Table 5.2). In support in daily life, participants with problematic (38.5%) and inadequate HL (35.7%) indicated that they receive significantly more often help compared to participants with sufficient HL (20.4%). Additionally, participants who

are longer than five years diagnosed with COPD had significantly more often sufficient HL (54.6%) compared to people who were diagnosed for fewer years (inadequate: 39.3%; problematic: 37.7%). Lastly, a significant difference in the distribution was observable for feeling secure in medication intake. Participants with sufficient HL (66.6%) indicated feeling confident in medication intake compared to participants with inadequate (48.2%) and problematic HL (51.3%). No significant difference in the distribution of variables' characteristics was observable for any other categorical variable.

Table 5.2: Differences in variables' characteristics regarding health literacy

		Sufficient (n=108)		Inadequate (n=56)		Problematic (n=39)		Chi-square
		Number	(%)	Number	(%)	Number	(%)	
Gender	Male	36	33.3	21	37.5	16	41.0	0.665
	Female	70	64.8	35	62.5	23	59.0	
	Diverse	2	1.9	0	0.0	0	0.0	
Education	No school	1	0.9	1	1.8	3	7.7	0.278
	Compulsory school	19	17.6	8	14.3	9	23.1	
	Professional training	54	50.0	33	58.9	19	48.7	
	Matura	5	4.6	2	3.6	1	2.6	
	Higher education	18	16.7	9	16.1	7	17.9	
	University degree	11	10.2	3	5.4	0	0.0	
COPD status	A	13	12.0	4	7.1	2	5.1	0.316
	B	56	51.9	34	60.7	23	59.0	
	C	5	4.6	0	0.0	0	0.0	
	D	34	31.5	18	32.1	14	35.9	
Support in daily life	Yes	22	20.4	20	35.7	15	38.5	0.032
	No	86	79.6	36	64.3	24	61.5	
Smoking	Yes, daily	28	25.9	14	25.0	19	48.7	0.059
	Yes, sometimes	12	11.1	2	3.6	3	7.7	
	No longer	56	51.9	35	62.5	13	33.3	
	Never	12	11.1	5	8.9	4	10.3	
Length of diagnosis	Less than 1 year	17	15.7	10	17.9	6	15.4	0.015
	1 to less than 3 years	16	14.8	5	8.9	12	30.8	
	3 to less than 5 years	16	14.8	19	33.9	6	15.4	
	5 to less than 7 years	23	21.3	12	21.4	10	25.6	
	7 to less than 10 years	17	15.7	3	5.4	3	7.7	
	Over 10 years	19	17.6	7	12.5	2	5.1	

Hospitalized exacerbation	Yes	39	36.1	18	32.1	14	35.9	0.872
	No	69	63.9	38	67.9	25	64.1	
Taking medication	Yes	92	85.2	49	87.5	30	76.9	0.351
	No	16	14.8	7	12.5	9	23.1	
Feeling secure in medication intake	Very insecure	14	13.0	3	5.4	7	17.9	<0.001
	Insecure	9	8.3	11	19.6	2	5.1	
	Neutral	13	12.0	14	25.0	10	25.6	
	Secure	20	18.5	9	16.1	14	35.9	
	Very secure	52	48.1	18	32.1	6	15.4	
	I do not take medication	0	0.0	1	1.8	0	0.0	
Primary contact person	GP	39	36.1	18	32.1	20	51.3	0.611
	Specialist/Pulmonologist	59	54.6	32	57.1	16	41.0	
	Hospital	5	4.6	4	7.1	2	5.1	
	Other	5	4.6	2	3.6	1	2.6	
Exchange with other patients	Yes	14	13.0	5	8.9	2	5.1	0.356
	No	94	87.0	51	91.1	37	94.9	
Participation at "Living better with COPD"	Yes	13	12.0	7	12.5	4	10.3	0.941
	No	95	88.0	49	87.5	35	89.7	
Participation at outpatient pulmonary rehabilitation	Yes	32	29.6	15	26.8	8	20.5	0.546
	No	76	70.4	41	73.2	31	79.5	
Yearly vaccination against the flu	Yes	64	59.3	34	60.7	18	46.2	0.299
	No	44	40.7	22	39.3	21	53.8	

An ANOVA was applied for continuous variables to test for differences in the mean values between the three HL states. I identified significant differences in age ($p<0.001$), EQ-5D-5L ($p<0.001$), EQ-VAS ($p<0.001$), and CAT ($p=0.004$) (see Table 5.3). According to the post hoc test, participants with sufficient HL were significantly older than participants with inadequate ($p=0.005$) and problematic ($p=0.002$) HL. Participants with sufficient HL expressed significantly better scores in EQ-5D-5L than participants with inadequate ($p=0.003$) and problematic ($p<0.001$) HL. Concerning EQ-VAS, participants with problematic HL stated significantly worse values compared to inadequate ($p=0.012$) and sufficient ($p<0.001$) HL. For the CAT score, participants with sufficient HL showed significantly low-er (i.e., better) values compared to inadequate ($p=0.041$) and problematic ($p=0.010$) HL. No significant differences were observable between the groups for mean BMI and the average number of comorbidities.

Table 5.3: Differences in mean values regarding health literacy

	Sufficient (n=108) Mean (SD)	**Inadequate (n=56)** Mean (SD)	**Problematic (n=39)** Mean (SD)	**F**	**p-value**
Age	67.54 (8.85)	62.45 (10.68)	61.28 (10.33)	8.55	<0.001
BMI	26.67 (6.76)	27.33 (6.37)	25.96 (6.72)	0.49	0.614
Number of comorbidities	2.89 (1.23)	2.22 (1.30)	2.75 (3.10)	0.553	0.581
EQ-5D-5L	0.793 (0.200)	0.6714 (0.215)	0.610 (0.280)	10.81	<0.001
EQ-VAS	65.27 (21.19)	62.11 (21.05)	48.95 (23.53)	8.226	<0.001
CAT	17.23 (7.53)	20.30 (7.70)	21.38 (7.07)	5.801	0.004

Factors influencing health literacy

For the logistic regression, the control variables were support in daily life, length of diagnosis, feeling secure in medication intake, age, EQ-5D-5L, EQ-VAS, and CAT. These variables were chosen due to their significant values ($p<0.05$) in the previously conducted chi-square test and ANOVA. The logistic regression revealed that age and the EQ-5D-5L are significantly associated with HL (see Table 5.4). A lower odds ratio signifies a higher (better) HL. Thus, the odds ratio for age (OR= -0.042; p=0.011) and the EQ-5D-5L (OR= -2.166; p=0.017) express a higher likelihood of sufficient HL to be present. No significant associations were observable for any other included variable. The model achieved a pseudo-R-squared of 0.260.

Table 5.4: Regression table for identifying associations with health literacy

	Odds ratio	Confidence intervals	Significance (p-value)
Support in daily life	0.625	-0.098 – 1.349	0.090
Length of diagnosis			
Less than 1 year	0.479	-0.655 – 1.614	0.408
1 to less than 3 years	0.915	-0.200 – 2.029	0.108
3 to less than 5 years	0.453	-0.630 – 1.536	0.412
5 to less than 7 years	0.823	-0.217 – 1.863	0.121
7 to less than 10 years	-0.417	-1.716 – 0.881	0.529
Feeling secure in medication intake			
Very insecure	-0.791	-4.630 – 3.048	0.686
insecure	-1.070	-4.870 – 2.729	0.581
Neutral	-0.561	-4.346 – 3.224	0.771
Secure	-0.631	-4.394 – 3.131	0.742
Very secure	-1.406	-5.191 – 2.378	0.466
Age	-0.042	-0.074 – -0.010	0.011
EQ-5D-5L	-2.166	-3.950 – -0.382	0.017
EQ-VAS	-0.007	-0.023 – 0.008	0.364
CAT	-0.018	-0.069 – 0.033	0.494

Reference categories: support in daily life: no; diagnosis in years: over 10 years; confidence in medication intake: very confident

Differences in desire for support

The “top three” answers of the patients’ desire for daily support are listed in Table 5.5 according to the HL states. Concerning difficulty in dealing with COPD, participants of all HL states indicated facing most difficulties in the limited movement possibilities, followed by a lack of chance to cure and fears about the future due to disease and changing living circumstances for sufficient HL. Compared to participants with inadequate and problematic HL, they indicated most difficulties in dealing with fears about the future due to the disease and, the lack of chance of cure (inadequate) and insufficient sleep quality (problematic). Thus, independent of HL, participants perceive the same difficulties in dealing with COPD.

Regarding the wish for support, the picture is even more similar. Independent of the HL state, the participants wished for support through regular advice and information dealing with COPD, followed by inputs for regular exercises to decrease the COPD symptom level. Third ranks in all HL states the desire to compare their health status with the average COPD patient. Also, for the support in dealing with COPD-specific symptoms, the highest symptom burdens are breathlessness, anxiety/depression, and tiredness across all HL states. Concludingly, no differences in needs for support were observable between HL states – the needs were almost identical.

Table 5.5: Difficulty in dealing with COPD, wish for support in daily life, and support with symptoms per HL group

	Sufficient (n=108)	Inadequate (n=56)	Problematic (n=39)
Difficulty in dealing with COPD	1. Limited movement possibilities (48.1%) 2. Lack of chance to cure (38.9%) 3. Fears about future due to disease (31.5%) / Changing living circumstances (31.5%)	1. Limited movement possibilities (62.5%) 2. Fears about future due to disease (58.9%) 3. Lack of chance to cure (42.9%)	1. Limited movement possibilities (69.2%) 2. Fears about future due to disease (59.0%) 3. Insufficient sleep quality (41.0%)
Wish for support in daily life	1. Regular advice and information dealing with COPD (36.1%) 2. Regular exercises to decrease COPD symptom level (28.7%) 3. Comparing own health status with average COPD patient (18.5%)	1. Regular advice and information dealing with COPD (57.1%) 2. Regular exercises to decrease COPD symptom level (44.6%) 3. Comparing own health status with average COPD patient (33.9%)	1. Regular advice and information dealing with COPD (64.1%) 2. Regular exercises to decrease COPD symptom level (48.7%) 3. Comparing own health status with average COPD patient (30.9%)
Support with symptoms	1. Breathlessness (50.0%) 2. Anxiety/depression (33.3%) 3. Tiredness (29.6%)	1. Breathlessness (51.8%) 2. Tiredness (39.3%) 3. Anxiety/depression (37.5%)	1. Breathlessness (69.2%) 2. Tiredness (53.8%) 3. Anxiety/depression (51.3%)

5.1.5 Discussion

The sample included 203 participants, of which 108 showed sufficient HL, 56 indicated inadequate HL and 39 expressed problematic HL. Evidence exists that people suffering from COPD are often affected by low HL [16] – in this case, around 46% of the participants. Furthermore, the sample consisted of 63% female participants. Based on current research, men are no longer exclusively affected by COPD, but more and more women suffer from the disease. However, women often remain under-diagnosed [46] due to various problems, such as women being less frequently referred to pulmonologists or receiving spirometry testing [47, 48]. Additionally, women tend to experience symptoms such as fatigue or depression, which are indicators of other diseases as well [49].

Additionally, the participants in the collected sample display more often B and D COPD status – indicating that they suffer from a worse CAT score. Lower HL is associated with higher (i.e., worse) CAT scores [50]. Although Omachi et al. [51] claim that lower HL is associated with more severe COPD status, no significant differences between HL states are observable in my sample. Additionally, the sample encompassed a broad range of educational levels, with most participants having finished a professional training (52.2%), followed by compulsory school (17.7%) and higher education (16.7%). This represents a balanced sample, as in Switzerland, 41.2% conduct a professional training, followed by 15% receiving a higher education degree and 13.9% finishing compulsory school [52].

Significant differences were observable in the sample for the length of diagnosis and feeling secure in the medication intake between the HL states. Participants with sufficient HL were longer diagnosed with COPD compared to the other two HL states. Gysels and Higginson [53] claimed that patients develop better HL over time as they develop an understanding of their symptoms and become able to understand changes in the environment or the body and take proper counteracting measures. Also, Lu et al. [54] highlighted in their study that a longer disease duration is positively associated with better HL and thus leading to higher self-efficacy of managing the disease.

Furthermore, 66% of participants with sufficient HL indicated feeling confident in medication intake, whereas only 48.2% of participants with inadequate and 51.3% with problematic HL expressed the same feeling. Interestingly, proportionally more participants with problematic HL felt secure in medication intake compared to participants with adequate HL. This finding might be explained with the Dunning-Kruger effect, where people with lower HL overestimate their capabilities of understanding and taking the medication correctly [19, 20, 55]. Nevertheless, the proportion of participants feeling secure in medication intake was rather low across all HL states, while COPD patients with higher HL are more adherent to medication intake [56].

Concerning the posed research questions, the focus of the first research question was whether significantly associated factors with patients' HL are identifiable. The ordinal logistic regression reveals that age and the EQ-5D-5L are significantly associated with HL. Compared to other studies, no clear tendency of the association of age with HL in COPD patients is derived in the literature. While some studies indicated that there is no significant association of age with HL [16, 57], others highlighted a significant negative association between these two variables [58, 59]. Regarding HRQoL (represented by EQ-5D-5L and EQ-VAS) and the disease-specific condition (measured by CAT), significant differences between HL states were observable. These values were all most favorable for participants with sufficient HL, i.e., higher HRQoL and lower disease-specific burden than the other two HL states. Results in line with this study are found in literature [50, 58, 60]. Better HRQoL is positively associated with better HL. Furthermore, Fischer et al. [50] highlighted that higher HL leads to a lower CAT. The EQ-5D-5L is significantly associated with better HL as also shown in Stellefson et al. [58] – thus, strengthening the relation between HRQoL and HL even more.

The second research question dealt with differences in the desire for support by respecting the COPD patients' HL. Although participants differed heavily in their HL, their difficulties in dealing with COPD, requirements for support in daily life, and disease-specific symptoms were almost similar, and no differences across groups were

observable. According to this study, study participants perceived most difficulties in the limited movement possibilities and the fears about the future due to the disease. They wished for support in the form of regular advice and information on how to deal with COPD-specific symptoms, followed by regular exercise to decrease the symptom burden and the possibility of comparing their health status with other COPD patients. Also, the participants' perception of managing the disease-specific burden were similar independent of the HL states. Due to the similar requirements across HL states, scalability and cost-effectiveness of DHIs potentially increase [33]. Based on the survey's direct derivation of support options, patients' acceptability of the DHI may be enhanced since it reflects directly on the patients' support requirements.

This study faces two main limitations: 1) a potential bias in the sample as the survey was conducted online and participants were mainly recruited through Facebook advertisements, and 2) the assessment of HL depended on patients' reported information. The first limitation may be established due to the online survey form and the main recruitment strategy through Facebook. Thus, only people with access to the Internet and a Facebook account were mainly targeted. However, since Internet coverage is almost entirely provided in Switzerland and Facebook is still one of the major social media platforms for baby-boomer and earlier generations, I believe this limitation has only minor influence on the sample's recruitment. Furthermore, future health interventions to strengthen HL and self-management capabilities will more extensively use digital approaches, which precondition patients to interact with digital tools. The second limitation refers to the patient's self-report of HL, and their overestimation of their HL capabilities. However, using an internationally validated instrument, the HLS-EU-Q16, any bias of patients' overestimating their HL capabilities should be reduced to a minimum.

5.1.6 Conclusion

This study highlights significant differences in the amount of support in daily life, length of diagnosis, feeling secure in medication intake, age, and HRQoL across COPD patients with different HL states. Participants with sufficient HL receive less support daily, feel more secure in medication intake, are older, have a higher HRQoL,

and indicate less disease-specific impact on health. Furthermore, the ordinal logistic regression reveals that higher age and better HRQoL significantly increase the likelihood of enhanced HL.

Additionally, this study derives desired supportive mechanisms for daily life, such as regular advice on COPD treatment and exercises on coping with COPD-specific symptoms. It further reveals that COPD patients independent of their HL are struggling with similar symptoms such as breathlessness, anxiety/depression, and tiredness. Thus, these insights guide the development of mechanisms being implemented in DHIs. Furthermore, future research may build on the findings of this study, aiming to influence patients' self-perceived quality of life to prove whether an improved HRQoL has an impact on HL and develop supporting mechanisms to enhance patients' self-management capabilities. This study contributes to existing evidence by highlighting the association of age and HRQoL with HL. As patients' HL is an important lever for increased self-management capabilities, it also strengthens the ability to cope with COPD daily. Therefore, it is essential to assess patients' HL regularly and aim to improve it in case patients display inadequate or problematic HL.

References

1. World Health Organization (2020) Global health estimates 2020: Deaths by cause, age, sex, by country and by region, 2000-2019

2. BAG (2020) Chronische Atemwegserkrankungen. https://www.bag.admin.ch/bag/de/home/krankheiten/krankheiten-im-ueberblick/chronische-atemwegserkrankungen.html

3. Zysman M, Raherison-Semjen C (2022) Women's COPD. Front Med (Lausanne) 8:.

4. GOLD (2021) Global Initiative for Chronic Obstructive Lung Disease

5. Wieser S, Tomonaga Y, Riguzzi M, et al (2014) Die Kosten der nichtübertragbaren Krankheiten in der Schweiz - Schlussbericht. Winterthur

6. Eisner MD, Blanc PD, Omachi TA, et al (2011) Socioeconomic status, race and COPD health outcomes. J Epidemiol Community Health (1978) 65:26–34.

7. Barlow J, Wright C, Sheasby J, et al (2002) Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 48:177–187.

8. Muellers KA, Chen L, O'Conor R, et al (2019) Health literacy and medication adherence in COPD patients: When caregiver presence is not sufficient. COPD: Journal of Chronic Obstructive Pulmonary Disease 16:362–367.

9. Effing TW, Vercoulen JH, Bourbeau J, et al (2016) Definition of a COPD self-management intervention: International expert group consensus. European Respiratory Journal 48:46–54.

10. Yadav UN, Hosseinzadeh H, Baral KP (2018) Self-management and patient activation in COPD patients: An evidence summary of randomized controlled trials. Clin Epidemiol Glob Health 6:148–154.

11. Nutbeam D (1998) Health Promotion Glossary. Health Promot Int 13:349–364.

12. Osborne RH, Batterham RW, Elsworth GR, et al (2013) The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 13:658.

13. Sørensen K, Van den Broucke S, Pelikan JM, et al (2013) Measuring health literacy in popula-tions: illuminating the design and development process of the European Health Literacy Sur-vey Questionnaire (HLS-EU-Q). BMC Public Health 13:948.

14. Morris NS, MacLean CD, Chew LD, Littenberg B (2006) The Single Item Literacy Screener: Evaluation of a brief instrument to identify limited reading ability. BMC Fam Pract 7:21.

15. Poureslami I, FitzGerald JM, Tregobov N, et al (2022) Health literacy in asthma and chronic obstructive pulmonary disease (COPD) care: a narrative review and future directions. Respir Res 23:361.

16. Azkan Ture D, Bhattacharya S, Demirci H, Yildiz T (2022) Health literacy and health outcomes in chronic obstructive pulmonary disease patients: An explorative study. Front Public Health 10: 846768.

17. Voigt-Barbarowicz M, Dietz G, Renken N, et al (2022) Patients' health literacy in rehabilitation: Comparison between the estimation of patients and health care professionals. Int J Environ Res Public Health 19:3522.

18. Dickens C, Lambert BL, Cromwell T, Piano MR (2013) Nurse overestimation of patients' health literacy. J Health Commun 18:62–69.

19. Storms H, Aertgeerts B, Vandenabeele F, Claes N (2019) General practitioners' predictions of their own patients' health literacy: a cross-sectional study in Belgium. BMJ Open 9:e029357.

20. Stock S, Altin S, Nawabi F, et al (2021) A cross-sectional analysis of health literacy: patient- versus family doctor-reported and associations with self-efficacy and chronic disease. BMC Fam Pract 22:187.

21. Sunjaya AP, Bao L, Martin A, et al (2022) Systematic review of effectiveness and quality as-sessment of patient education materials and decision aids for breathlessness. BMC Pulm Med 22:237.

22. Szabó P, Bíró É, Kósa K (2021) Readability and comprehension of printed patient education materials. Front Public Health 9:725840.

23. Russell S, Ogunbayo OJ, Newham JJ, et al (2018) Qualitative systematic review of barriers and facilitators to self-management of chronic obstructive pulmonary disease: views of patients and healthcare professionals. NPJ Prim Care Respir Med 28:2.

24. Wong SS, Abdullah N, Abdullah A, et al (2014) Unmet needs of patients with chronic obstruc-tive pulmonary disease (COPD): a qualitative study on patients and doctors. BMC Fam Pract 15:67.

25. Kjellsdotter A, Andersson S, Berglund M (2021) Together for the future – Development of a digital website to support Chronic Obstructive Pulmonary Disease self-management: A quali-tative study. J Multidiscip Healthc 14:757–766.

26. gesund.bund.de (2023) Krankheiten: Chronisch obstruktive Lungenerkrankung (COPD). https://gesund.bund.de/chronisch-obstruktive-lungenerkrankung-copd. Accessed 18 Nov 2023

27. Ding H, Karunanithi M, Kanagasingam Y, et al (2014) A pilot study of a mobile-phone-based home monitoring system to assist in remote interventions in cases of acute exacerbation of COPD. J Telemed Telecare 20:128–134.

28. Spielmanns M, Boeselt T, Huber S, et al (2020) Impact of a smartphone application (KAIA COPD app) in combination with Activity Monitoring as a maintenance prOgram following PUlmonary Rehabilitation in COPD: the protocol for the AMOPUR Study, an international, multicenter, parallel group, randomized, controlled study. Trials 21:636.

29. Miłkowska-Dymanowska J, Białas AJ, Obrębski W, et al (2018) A pilot study of daily telemonitoring to predict acute exacerbation in chronic obstructive pulmonary disease. Int J Med In-form 116:46–51.

30. Kowatsch T, Otto L, Harperink S, et al (2019) A design and evaluation framework for digital health interventions. it - Information Technology 61:253–263.

31. Fleisch E, Franz C, Hermann A (2021) The digital pill: What everyone should know about the future or our healthcare system. Emerald Publishing Limited, Bingley (UK)

32. Puente-Maestu L, Calle M, Rodríguez-Hermosa JL, et al (2016) Health literacy and health out-comes in chronic obstructive pulmonary disease. Respir Med 115:78–82.

33. Watson A, Wilkinson TMA (2022) Digital healthcare in COPD management: a narrative re-view on the advantages, pitfalls, and need for further research. Ther Adv Respir Dis 16:175346662210754.

34. Ramachandran HJ, Oh JL, Cheong YK, et al (2023) Barriers and facilitators to the adoption of digital health interventions for COPD management: A scoping review. Heart & Lung 59:117–127.

35. Bos-Touwen I, Schuurmans M, Monninkhof EM, et al (2015) Patient and disease characteristics associated with activation for self-management in patients with Diabetes, Chronic Obstructive Pulmonary Disease, Chronic

Heart Failure and Chronic Renal Disease: A cross-sectional survey study. PLoS One 10:e0126400.

36. Unipark (2023) Unipark. https://www.unipark.com/en/. Accessed 17 Nov 2023

37. EuroQol (2021) EQ-5D-5L | About. https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/

38. Nolan CM, Longworth L, Lord J, et al (2016) The EQ-5D-5L health status questionnaire in COPD: Validity, responsiveness and minimum important difference. Thorax 71:493–500.

39. Szentes BL, Schwarzkopf L, Kirsch F, et al (2020) Measuring quality of life in COPD patients: Comparing disease-specific supplements to the EQ-5D-5L. Expert Rev Pharmacoecon Out-comes Res 20:523–529.

40. Jones PW, Harding G, Berry P, et al (2009) Development and first validation of the COPD assessment test. European Respiratory Journal 34:648–654.

41. Jones PW (2013) COPD assessment test – Rationale, development, validation and performance. COPD: Journal of Chronic Obstructive Pulmonary Disease 10:269–271.

42. Jordan S, Hoebel J (2015) Gesundheitskompetenz von Erwachsenen in Deutschland: Ergebnisse der Studie „Gesundheit in Deutschland aktuell“ (GEDA). Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 58:942–950.

43. Eysenbach G (2004) Improving the quality of web surveys: The checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 6:e34.

44. Röthlin F, Pelikan JM, Ganahl K (2013) Die Gesundheitskompetenz von 15-jährigen Jugendlichen in Österreich. Abschlussbericht der österreichischen

Gesundheitskompetenz Jugendstudie im Auftrag des Hauptverbands der österreichischen Sozialversicherungsträger (HVSV)

45. SPSS Inc. (2022) Statistical Package for Social Science for Windows, Version 28.0

46. Gut-Gobert C, Cavaillès A, Dixmier A, et al (2019) Women and COPD: do we need more evidence? European Respiratory Review 28:180055.

47. Watson L, Vestbo J, Postma DS, et al (2004) Gender differences in the management and experience of Chronic Obstructive Pulmonary Disease. Respir Med 98:1207–1213.

48. Chapman KR, Tashkin DP, Pye DJ (2001) Gender bias in the diagnosis of COPD. Chest 119:1691–1695.

49. Martinez CH, Raparla S, Plauschinat CA, et al (2012) Gender differences in symptoms and care delivery for Chronic Obstructive Pulmonary Disease. J Womens Health 21:1267–1274.

50. Fischer C, Jörres RA, Alter P, et al (2022) Basic determinants of disease knowledge in COPD patients: Results from COSYCONET. Patient Prefer Adherence Volume 16:1759–1770.

51. Omachi TA, Sarkar U, Yelin EH, et al (2013) Lower health literacy is associated with poorer health status and outcomes in chronic obstructive pulmonary disease. J Gen Intern Med 28:74–81.

52. Federal Statistical Office (2023) Bildungsstand der Bevölkerung - Daten des Indikators. https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/bildungsindikatoren/themen/wirkung/bildungsstand.assetdetail.24485179.html. Accessed 13 Oct 2023

53. Gysels M, Higginson I (2009) Self-management for breathlessness in COPD: the role of pulmonary rehabilitation. Chron Respir Dis 6:133–140.

54. Lu J, Sun S, Gu Y, et al (2023) Health literacy and health outcomes among older patients suffering from chronic diseases: A moderated mediation model. Front Public Health 10:1069174.

55. Canady BE, Larzo M (2023) Overconfidence in managing health concerns: The Dunning–Kruger effect and health literacy. J Clin Psychol Med Settings 30:460–468.

56. O'Conor R, Muellers K, Arvanitis M, et al (2019) Effects of health literacy and cognitive abilities on COPD self-management behaviors: A prospective cohort study. Respir Med 160:105630.

57. Jayasinghe UW, Harris MF, Parker SM, et al (2016) The impact of health literacy and life style risk factors on health-related quality of life of Australian patients. Health Qual Life Outcomes 14:68.

58. Stellefson M, Paige SR, Alber JM, et al (2019) Association between health literacy, electronic health literacy, disease-specific knowledge, and health-related quality of life among adults with Chronic Obstructive Pulmonary Disease: Cross-sectional study. J Med Internet Res 21:e12165.

59. Yadav UN, Lloyd J, Hosseinzadeh H, et al (2020) Levels and determinants of health literacy and patient activation among multi-morbid COPD people in rural Nepal: Findings from a cross-sectional study. PLoS One 15:e0233488.

60. Wahl AK, Osborne RH, Larsen MH, et al (2021) Exploring health literacy needs in chronic obstructive pulmonary disease (COPD): Associations between demographic, clinical variables, psychological well-being and health literacy. Heart & Lung 50:417–424.

Appendix
Supplementary Material 1: Survey questionnaire in original language

Identifikation des Unterstützungsbedarfs von Patient:innen mit COPD zur Entwicklung eines Digital Health Assistants

Geschätzte Teilnehmerinnen und Teilnehmer

Vielen Dank, dass Sie sich die Zeit für unsere Umfrage nehmen. Ziel der Umfrage ist es ein vertieftes Verständnis Ihrer alltäglichen Herausforderungen mit COPD zu erhalten. Ihre Rückmeldungen helfen uns neue Gesundheitsdienstleistungen zu entwickeln, die Sie zukünftig besser im Alltag mit COPD unterstützen. Die Umfrage dauert ca. 10-20 Minuten.

Ihre Daten werden vertraulich und anonym behandelt. Ihre Antworten dienen ausschliesslich obengenanntem Zweck und werden nicht an Dritte weitergereicht. Nach Abschluss der Forschungsarbeit werden sämtliche Daten vernichtet.

Unter allen teilnehmenden Personen verlosen wir am Ende der Umfrage 3 Reka-Gutscheine im Wert von je CHF 100.

Bei Fragen oder technischen Schwierigkeiten wenden Sie sich gerne an Anja Bischof (E-Mail: XXX, Tel: XXX).

0. Haben Sie die Voraussetzungen zur Teilnahme an der Umfrage zum Unterstützungsbedarf im Umgang mit COPD verstanden und stimmen Sie den Bedingungen zu?
   … Ja, ich stimme zu.
   … Nein, ich stimme nicht zu.

Allgemeine Angaben
In diesem ersten Abschnitt bitten wir Sie um einige persönliche Angaben.

1. Bitte geben Sie Ihr Geschlecht an.
   … Männlich
   … Weiblich
   … Divers

2. Bitte geben Sie Ihr Geburtsjahr an: ……

3. Bitte geben Sie Ihre Grösse in cm an: …… cm

4. Bitte geben Sie Ihr Gewicht in kg an: …… kg

5. Bitte geben Sie Ihren höchsten Abschluss an.
   … Kein Schulabschluss
   … Obligatorische Schule
   … Lehrabschluss
   … Matura/Abitur
   … Höhere Berufsausbildung (HF)
   … Hochschulabschluss (Bachelor/Master/Doktorat)

6. Rauchen Sie?
   … Ja, täglich
   … Ja, gelegentlich
   … Nein, nicht mehr
   … Nein, noch nie
   a. Falls ja, wie viele Zigaretten (bzw. Zigarren/Pfeifen) pro Tag rauchen Sie?
      …… *(Anzahl eingeben)*
   b. Falls ja, seit wie vielen Jahren rauchen Sie?
      …… *(Jahre eingeben)*
   c. Falls gelegentlich, wie viele Zigaretten (bzw. Zigarren/Pfeifen) pro Woche?
      …… *(Anzahl eingeben)*
   d. Falls gelegentlich, seit wie vielen Jahren rauchen Sie?
      …… *(Jahre eingeben)*
   e. Falls nicht mehr, wie viele Zigaretten (bzw. Zigarren/Pfeifen) haben Sie pro Woche geraucht?
      …… *(Anzahl eingeben)*
   f. Falls nicht mehr, wie viele Jahre haben Sie geraucht?
      …… *(Jahre eingeben)*

7. Haben Sie bereits einmal an einem Rauchstopp-Programm teilgenommen?
   … Ja
   … Nein
   a. Falls Sie bereits an einem Rauchstopp-Programm teilgenommen haben, von welcher Art war dieses?
      (Mehrfachantwort möglich)
      … Professionelle Rauchstopp-Beratung
      … Rauchstopp-App
      … Hypnose oder Akupunktur
      … Entspannungstechniken wie autogenes Training oder Yoga
      … Andere: ……
      … Ich habe noch nie an einem Rauchstopp-Programm teilgenommen.
   b. An wie vielen Rauchstopp-Programmen haben Sie bereits teilgenommen?
      Bitte geben Sie die Anzahl in ganzen Zahlen an.
      …… (Anzahl eingeben)

8. Bitte geben Sie die Grösse Ihres Wohnorts an:
   … < 5'000 Einwohner
   … > 5'000 – 15'000 Einwohner

… > 15'000 – 50'000 Einwohner
… > 50'000 Einwohner

9. Wie sieht Ihre Wohnsituation zurzeit aus?
… Zu Hause mit Lebens- bzw. Ehepartner/in
… Zu Hause mit anderen Angehörigen
… Zu Hause allein
… Betreutes Wohnen
… Andere: ……

10. *Wenn zu Hause mit Lebens- bzw. Ehepartner/in oder Angehörigen angegeben wurde*: Raucht eine Person, welche mit Ihnen im gleichen Haushalt lebt?
… Ja
… Nein

11. Sind Sie zurzeit auf Hilfe zur Bewältigung Ihres Alltags angewiesen?
… Ja
… Nein

12. Von wem erhalten Sie Hilfe?
(Mehrfachantwort möglich)
… Angehörige
… Freunde/Nachbarn
… Haushaltshilfe
… SPITEX oder ähnliche Dienstleister
… Andere: ……
… Ich erhalte keine Hilfe.

13. Haben Sie Haustiere?
… Ja (Bitte geben Sie die Art Ihres Haustiers an – z.B. Hund, Katze, Vogel, etc.)
Welche? ……
… Nein

Fragen zur Krankheit

In diesem Abschnitt werden Ihnen einige Fragen zu Ihrer Krankheit gestellt.

14. An welcher chronischen Lungenkrankheit leiden Sie?
(Mehrfachantwort möglich)
… COPD
… Asthma
… Weiss ich nicht
… Andere Erkrankung: ……

15. Kennen Sie Ihr Stadium der COPD?
    … Ja
    … Nein
        a. Falls ja, in welchem Stadium befinden Sie sich?
        Hinweis: Durch Einteilung in COPD-Stadien lassen sich der Schweregrad und mögliche Veränderungen im Verlauf der Krankheit gut dokumentieren. Kombination aus Buchstabe und Zahl möglich.
        … A
        … B
        … C
        … D
        … 1
        … 2
        … 3
        … 4

16. Wie geht es Ihnen mit Ihrer COPD? *(Bitte geben Sie für jede der folgenden Aussagen an, was derzeit am besten auf Sie zutrifft. Kreuzen Sie (X) in jeder Zeile nur eine Möglichkeit an.)*

Ich huste nie	0 1 2 3 4 5	Ich huste ständig
Ich bin überhaupt nicht verschleimt	0 1 2 3 4 5	Ich bin völlig verschleimt
Ich spüre keinerlei Engegefühl in der Brust	0 1 2 3 4 5	Ich spüre ein sehr starkes Engegefühl in der Brust
Wenn ich bergauf oder eine Treppe hinaufgehe, komme ich nicht außer Atem	0 1 2 3 4 5	Wenn ich bergauf oder eine Treppe hinaufgehe, komme ich sehr außer Atem
Ich bin bei meinen häuslichen Aktivitäten nicht eingeschränkt	0 1 2 3 4 5	Ich bin bei meinen häuslichen Aktivitäten sehr stark eingeschränkt
Ich habe keine Bedenken, trotz meiner Lungenerkrankung das Haus zu verlassen	0 1 2 3 4 5	Ich habe wegen meiner Lungenerkrankung große Bedenken, das Haus zu verlassen
Ich schlafe tief und fest	0 1 2 3 4 5	Wegen meiner Lungenerkrankung schlafe ich nicht tief und fest
Ich bin voller Energie	0 1 2 3 4 5	Ich habe überhaupt keine Energie

17. Vor wie vielen Jahren wurde COPD bei Ihnen diagnostiziert?
    … Unter 1 Jahr
    … 1 bis unter 3 Jahre
    … 3 bis unter 5 Jahre
    … 5 bis unter 7 Jahre

… 7 bis unter 10 Jahre
… Über 10 Jahre

18. Gab es schon einmal einen stationären Aufenthalt im Krankenhaus aufgrund von COPD?
… Ja
… Nein

19. Falls Sie bereits einen stationären Aufenthalt im Krankenhaus aufgrund Ihrer COPD hatte, wie lange ist der letzte her?
… Unter 6 Monate
… 6 Monate bis unter 1 Jahr
… 1 bis unter 2 Jahre
… Über 2 Jahre
… Ich hatte noch nie einen stationären Aufenthalt aufgrund meiner COPD.

20. Nehmen Sie verschriebene Medikamente zur Eindämmung der Symptome Ihrer COPD?
… Ja
… Nein

21. Wie sicher fühlen Sie sich bei der Medikamenteneinnahme in Bezug auf den Einnahmezeitpunkt, die richtige Dosierung, etc.? *(1 = sehr unsicher; 5 = sehr sicher)*
…… *(Zahl eingeben)*

22. Wer ist Ihre primäre Ansprechperson, wenn es um Ihre COPD geht?
… Hausarzt
… Facharzt/Pneumologe (Lungenarzt)
… Spital
… Andere: ……

Fragen zum Unterstützungsbedarf
Es folgen nun einige Fragen zu den Herausforderungen mit Ihrer Krankheit. Ebenfalls können Sie Bereiche angeben, in denen Sie sich vermehrt Unterstützung wünschen.

23. Sind Sie in regelmässigem Austausch mit anderen COPD-Betroffenen?
… Ja
… Nein

24. Nehmen Sie am Selbstmanagement-Coaching Programm «Besser Leben mit COPD» der Lungenliga teil oder haben Sie bereits daran teilgenommen?
… Ja
… Nein

25. Nehmen Sie an einer ambulanten pulmonalen Rehabilitation (APR) teil bzw. haben Sie bereits daran teilgenommen?
    … Ja
    … Nein

26. Lassen Sie sich jährlich gegen die Grippe impfen?
    … Ja
    … Nein

27. Wo bestehen zurzeit Schwierigkeiten im Umgang mit COPD? *(Mehrfachantwort möglich)*
    … Häufung von ärztlichen Terminen
    … Regelmässige Medikamenteneinnahme
    … Mangelhafte Schlafqualität
    … Eingeschränkte Bewegungsmöglichkeiten
    … Fehlende Informationsgrundlage
    … Identifikation der zuständigen Fachstellen
    … Fehlende Heilungschancen
    … Unverständnis aus dem persönlichen Umfeld
    … Soziale Isolation
    … Bedarf an Hilfe/Unterstützung für alltägliche Tätigkeiten
    … Umgang mit Begleiterkrankungen
    … Nicht-Akzeptanz der Krankheit («Ich hadere mit mir selbst.»)
    … Veränderung der Lebensumstände
    … Zukunftsängsten aufgrund der Erkrankung
    … Andere Herausforderungen:
    ……
    ……

28. Im Umgang mit welchen Symptomen Ihrer COPD wünschen Sie sich Unterstützung? *(Mehrfachantwort möglich)*
    … Atemnot
    … Husten
    … Auswurf
    … Abgeschlagenheit/Angstzustände/Depression
    … Bewegungsmangel
    … Müdigkeit
    … Kein Unterstützungsbedarf
    … Sonstiges*:*
    ……
    ……

29. Leiden Sie an weiteren chronischen Erkrankungen?
    Hinweis: Eine chronische Erkrankung wird als ein langfristiger, nicht- oder nur schwer-heilbarer Krankheitszustand bezeichnet.
    … Ja
    … Nein

a. Falls ja, kreuzen Sie die entsprechenden bitte an. *(Mehrfachantwort möglich)*
… Koronare Herzerkrankung
… Herzinsuffizienz (Herzschwäche)
… Bronchialkarzinom
… Bluthochdruck
… Angststörung
… Depression
… Muskelschwund (Atrophie)
… Knochenschwund (Osteoporose)
… Metabolisches Syndrom
… Diabetes Mellitus Typ 1
… Diabetes Mellitus Typ 2
… Gewichtsverlust
… Anämie
… Rückenschmerzen
… Andere: ……

Hinweis: Falls Frage 29 mit «ja» beantwortet wurde, dann beantworten Sie bitte die Fragen 30-34 ebenfalls. Ansonsten gehen Sie bitte direkt zu Frage 35 weiter.

30. Wie gut fühlen Sie sich über die zuvor angegebenen chronischen Erkrankungen informiert? *(1 = ich besitze keine Information; 5 = ich besitze alle benötigten Informationen)*
…… *(Zahl eingeben)*

31. Sind die Informationsmaterialen, die Sie kennen und verwenden, verständlich in Bezug auf Ihre zuvor angegebenen chronischen Erkrankungen? *(1 = überhaupt nicht verständlich; 5 = komplett verständlich)*
…… *(Zahl eingeben)*

32. Wie kommen Sie zurzeit mit den zuvor angegebenen chronischen Erkrankungen zurecht?
(1 = überhaupt nicht; 5 = sehr gut)
…… *(Zahl eingeben)*

33. Wie gut fühlen Sie sich von Ihrem Hausarzt/Facharzt bezüglich der zuvor angegebenen chronischen Erkrankungen betreut? *(1 = überhaupt nicht; 5 = sehr gut)*
…… *(Zahl eingeben)*

34. Wünschen Sie sich im Umgang mit den zuvor angegebenen chronischen Erkrankungen weitere Unterstützung?
… Ja
… Eher ja
… Unentschlossen
… Eher nein

… Nein

35. Welche Arten der Unterstützung wünschen Sie sich in Bezug auf Ihre COPD sowie auf die zuvor angegebenen chronischen Erkrankungen *(nur falls zuvor angegeben)*? *(Mehrfachantwort möglich)*
… Unterstützung bei der Vereinbarung von Arztterminen
… Identifikation der zuständigen Fachstelle je nach Anliegen
… Erinnerung an Arzttermine
… Erinnerung an die Medikamenteneinnahme
… Erinnerung an weitere unterstützende Massnahmen zum Umgang mit COPD wie bspw. jährliche Grippeimpfung
… Regelmässige Tipps und Informationen zum Umgang mit COPD
… Regelmässige Tipps und Informationen zum Umgang mit weiteren chronischen Erkrankungen
… Regelmässige Übungen, um Symptome von COPD zu lindern (z.B. Erinnerung an gesunde Ernährung, regelmässige Bewegung, etc.)
… Regelmässige Übungen, um Symptome von den weiteren chronischen Krankheiten zu lindern
… Unterstützung beim Rauchstopp
… Austauschmöglichkeiten mit weiteren Betroffenen
… Vergleichsmöglichkeiten des eigenen Gesundheitszustandes zum durchschnittlichen COPD-Patienten
… Hilfestellungen für Angehörige oder Pflegende
… Unterstützung bei der Kommunikation mit der Krankenkasse
… Unterstützung beim Bezahlen von Rechnungen
… Kochen bzw. Sicherstellen einer gesunden Ernährung
… Sonstiges:
……
……

36. Denken Sie, dass sich eine Verhaltensänderung Ihrerseits positiv auf Ihre COPD auswirken kann?
Hinweis: Im Zusammenhang mit dieser Umfrage wird unter Verhaltensänderung jene Veränderung von Gewohnheiten betrachtet, die sich positiv auf die Gesundheit auswirken, wie bspw. mehr Bewegung, gesunde Ernährung, genügend Schlaf, striktes Befolgen von Therapieplänen, Wissenserwerb zur Krankheit etc.
… Ja
… Eher ja
… Unentschlossen
… Eher nein
… Nein

37. Sind Sie zur Verbesserung Ihrer COPD zu einer Verhaltensänderung bereit?
… Ja

… Eher ja
… Unentschlossen
… Eher nein
… Nein

38. *Falls weitere chronische Erkrankungen angegeben:* Denken Sie, dass sich die Verhaltensänderungen auch positiv auf Ihre anderen chronischen Erkrankungen auswirken kann?
… Ja
… Eher ja
… Unentschlossen
… Eher nein
… Nein
    a. Falls ja, sind Sie zur Verbesserung Ihrer weiteren chronischen Erkrankungen zu einer Verhaltensänderung bereit?
    … Ja
    … Eher ja
    … Unentschlossen
    … Eher nein
    … Nein

39. Nutzen Sie Ihr Smartphone regelmässig?
… Ja
… Nein

40. Falls ja: Wie oft verwenden Sie Ihr Smartphone?
… Mehrmals täglich
… Einmal am Tag
… Mehrmals in der Woche
… Einmal in der Woche
… Seltener

41. Besitzen Sie eine Smartwatch?
Hinweis: Eine Smartwatch ist eine Armbanduhr, die über zusätzliche Sensoren verfügt, welche beispielsweise Ihren Puls oder Ihre Schrittzahl messen.
… Ja
… Nein

42. Nutzen Sie digitale Lösungen zur Unterstützung im Umgang mit Ihrer COPD?
Hinweis: Mit digitalen Lösungen sind in diesem Kontext Handy-Apps oder Wearables/Smartwatches/andere Tracking Geräte (kleine tragbare Computer) gemeint. Dabei erhalten Sie Informationen zur Krankheit oder Sie verfolgen damit die Entwicklung Ihres Gesundheitszustandes über einen gewissen Zeitraum.
… Ja
Welche? ……
… Nein

43. Wären Sie bereit, eine digitale Lösung zur Unterstützung im Umgang mit COPD auszuprobieren?
    … Ja
    … Eher ja
    … Unentschlossen
    … Eher nein
    … Nein

Hinweis: Falls Frage 29 mit «ja» beantwortet wurde, dann beantworten Sie bitte die Fragen 44 & 45 ebenfalls. Ansonsten gehen Sie bitte direkt weiter zu Frage 46.

44. Nutzen Sie bereits digitale Lösungen zur Unterstützung im Umgang mit Ihren weiteren chronischen Erkrankungen?
    … Ja
    … Nein
    a. Falls ja, welche? *(freie Texteingabe)*
       ……

45. Wünschen Sie sich ebenfalls Unterstützung im Umgang mit den zuvor angegebenen chronischen Krankheiten?
    … Ja
    … Eher ja
    … Unentschlossen
    … Eher nein
    … Nein

Fragen zur Lebensqualität
In dieser kurzen Sequenz bitten wir Sie Ihre Lebensqualität einzuschätzen.

46. Wie hoch schätzen Sie Ihre Lebensqualität ein? (0 = schlecht möglichst, 100 = bestmöglich)
    …… *(Zahl eingeben)*

47. Bitte kreuzen Sie unter jeder Überschrift DAS Kästchen an, das Ihre Gesundheit HEUTE am besten beschreibt.
    a. **Beweglichkeit/Mobilität**
       … Ich habe keine Probleme herumzugehen
       … Ich habe leichte Probleme herumzugehen
       … Ich habe mässige Probleme herumzugehen
       … Ich habe starke Probleme herumzugehen
       … Ich bin nicht in der Lage herumzugehen
    b. **Für sich selbst sorgen**
       … Ich habe keine Probleme, mich selbst zu waschen oder anzuziehen

… Ich habe leichte Probleme, mich selbst zu waschen oder anzuziehen
… Ich habe mässige Probleme, mich selbst zu waschen oder anzuziehen
… Ich habe grosse Probleme, mich selbst zu waschen oder anzuziehen
… Ich bin nicht in der Lage, mich selbst zu waschen oder anzuziehen

c. **Alltägliche Tätigkeiten**
(z.B. Arbeit, Studium, Hausarbeit, Familien- und Freizeitaktivitäten)
… Ich habe keine Probleme, meinen alltäglichen Tätigkeiten nachzugehen
… Ich habe leichte Probleme, meinen alltäglichen Tätigkeiten nachzugehen
… Ich habe mässige Probleme, meinen alltäglichen Tätigkeiten nachzugehen
… Ich habe grosse Probleme, meinen alltäglichen Tätigkeiten nachzugehen
… Ich bin nicht in der Lage, meinen alltäglichen Tätigkeiten nachzugehen

d. **Schmerz/Körperliche Beschwerden**
… Ich habe keine Schmerzen oder Beschwerden
… Ich habe leichte Schmerzen oder Beschwerden
… Ich habe mässige Schmerzen oder Beschwerden
… Ich habe starke Schmerzen oder Beschwerden
… Ich habe extreme Schmerzen oder Beschwerden

e. **Angst/Niedergeschlagenheit**
… Ich bin nicht ängstlich oder deprimiert
… Ich bin ein wenig ängstlich und deprimiert
… Ich bin mässig ängstlich und deprimiert
… Ich bin sehr ängstlich und deprimiert
… Ich bin extrem ängstlich und deprimiert

Fragen zum Umgang mit COPD und Ihrer allgemeinen Gesundheit
Der letzte Abschnitt beinhaltet Aussagen zum Umgang mit COPD. Wir bitten Sie, diese zu bewerten. Dabei geht es primär darum, herauszufinden, wie gut Sie sich über die chronische Lungenerkrankung COPD informiert fühlen.

Auf einer Skala von sehr einfach bis sehr schwierig, wie einfach ist es Ihrer Meinung nach …

48. … Informationen über Therapien für Krankheiten, die Sie betreffen, zu finden?
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

49. … herauszufinden, wo Sie professionelle Hilfe erhalten, wenn Sie krank sind? (Arzt, Apotheker, Psychologe)
… Sehr einfach
… Ziemlich einfach

… Ziemlich schwierig
… Sehr schwierig

50. … zu verstehen, was Ihr Arzt Ihnen sagt?
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

51. … die Anweisungen Ihres Arztes oder Apothekers zur Einnahme der verschriebenen Medikamente zu verstehen?
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

52. … zu beurteilen, wann Sie eine zweite Meinung von einem anderen Arzt einholen sollten?
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

53. … mit Hilfe der Informationen, die Ihnen der Arzt gibt, Entscheidungen bezüglich Ihrer Krankheit zu treffen?
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

54. … den Anweisungen Ihres Arztes oder Apothekers zu folgen?
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

55. … Informationen über Unterstützungsmöglichkeiten bei psychischen Problemen, wie Stress oder Depression, zu finden?
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

56. … Gesundheitswarnungen vor Verhaltensweisen wie Rauchen, wenig Bewegung oder übermässiges Trinken zu verstehen?
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig

… Sehr schwierig

57. … zu verstehen, warum Sie Vorsorgeuntersuchungen brauchen? (Krebsfrüherkennung, Blutzuckertest, Blutdruck)
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

58. … zu beurteilen, ob Informationen über Gesundheitsrisiken in den Medien vertrauenswürdig sind? (Fernsehen, Internet oder andere Medien)
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

59. … aufgrund von Informationen aus den Medien zu entscheiden, wie Sie sich vor Krankheiten schützen können? (Zeitungen, Broschüre, Internet oder andere Medien)
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

60. … Informationen über Verhaltensweisen zu finden, die gut für Ihr psychisches Wohlbefinden sind? (Meditation, körperliche Bewegung, Spazieren gehen, Pilates, etc.)
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

61. … Gesundheitsratschläge von Familienmitgliedern oder Freunden verstehen?
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

62. … Informationen in den Medien darüber, wie Sie Ihren Gesundheitszustand verbessern können, zu verstehen? (Internet, Zeitungen, Zeitschriften)
… Sehr einfach
… Ziemlich einfach
… Ziemlich schwierig
… Sehr schwierig

63. … zu beurteilen, welche Alltagsgewohnheiten mit Ihrer Gesundheit zusammenhängen? (Trink- und Essgewohnheiten, Bewegung, etc.)
… Sehr einfach
… Ziemlich einfach

… Ziemlich schwierig
… Sehr schwierig

64. Welche Empfehlung würden Sie jemandem geben, der erst kürzlich die Diagnose COPD erhalten hat?
……
……

65. Zurückblickend, was würden Sie im Umgang mit Ihrer COPD aus heutiger Sicht ändern?
……
……

66. Angenommen Sie hätten drei Wünsche im Umgang mit Ihrer COPD frei. Was würden Sie sich wünschen?
……
……

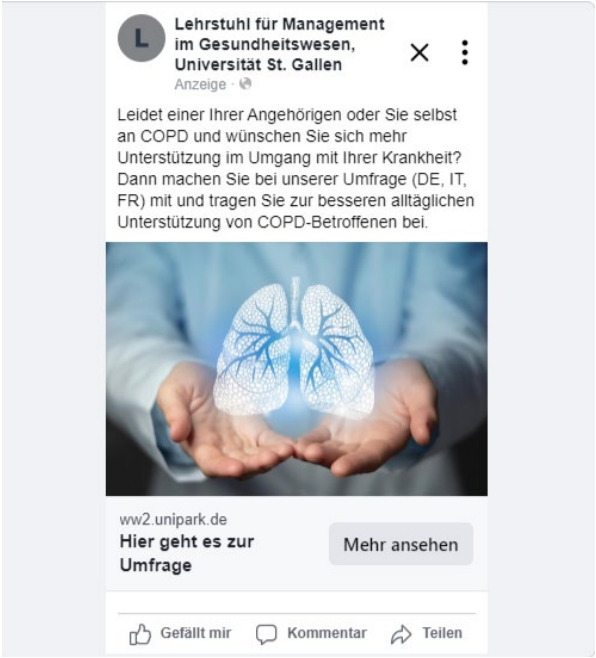
Weitere Anmerkungen
……
……

Vielen Dank für Ihre Teilnahme!

Mit Ihrer Teilnahme helfen Sie nicht nur der Forschung, sondern tragen auch dazu bei Unterstützungsmöglichkeiten im Umgang mit COPD zu verbessern. Wir danken Ihnen daher sehr, dass Sie sich die Zeit genommen haben, an unserer Umfrage teilzunehmen.

Bei Fragen, Interesse an der Auswertung oder falls Sie an der Verlosung für die 3 Reka-Gutscheine im Wert von je CHF 100 teilnehmen möchten, schreiben Sie gerne eine E-Mail an Anja Bischof (E-Mail: XXX, Tel: XXX).

Supplementary Material 2: Facebook advertisement



Supplementary Material 3: Flyer for recruitment

Universität St.Gallen
School of Medicine

LUNGENLIGA ST.GALLEN–APPENZELL
Mehr Luft fürs Leben

Kantonsspital St.Gallen

Studienteilnehmende gesucht

Identifikation des Unterstützungsbedarfs von Patient:innen mit COPD

(Chronisch-obstruktive Lungenerkrankung)

Teilnehmendenprofil
Wir suchen COPD-betroffene Personen ab 18 Jahren mit deutscher Muttersprache oder guten Deutschkenntnissen.

Ziel der Studie
Wir möchten prüfen, ob und welchen Unterstützungsbedarf COPD-Betroffene haben und ob dieser durch bestehende Programme/Angebote genügend abgedeckt wird. Ausgehend von diesen Erkenntnissen soll ein digitaler Gesundheitsassistent entwickelt werden, der die Bedürfnisse der Patient:innen in den Mittelpunkt stellt.

Ablauf und Dauer der Studie
Die Studie besteht aus einem Fragebogen, für dessen Beantwortung Sie ungefähr 20 Minuten benötigen (siehe QR-Code und Link).
Der Fragebogen beinhaltet Fragen zu Ihrer Lebensqualität, dem Umgang mit Ihrer Erkrankung und einem möglichen Unterstützungsbedarf im Alltag aufgrund Ihrer COPD.

Information
Anja Bischof
+41 (0)71 224 32 20
anja.bischof@unisg.ch

From insight to impact.

6 Results and outlook

This dissertation has explored how the patient's perspective may be leveraged to impede LVC and thus better control rising healthcare costs. It makes three contributions to theory and practice: first, it outlines approaches to identifying LVC from routine and claims data; second, it highlights essential considerations for the interpretation and adoption of PROMs in clinical practice; and third, it emphasizes the importance of patients' HL in enhancing their self-management capabilities to more sustainably manage their diseases. The second and third contributions indicate that effective countermeasures are needed to lastingly reduce LVC.

This dissertation has shown that identifying LVC in routine and claims data is necessary for validating whether LVC is present in the system and which treatments are especially prone to LVC. Based on the definition introduced at the beginning of this dissertation (see Chapter 1.1.1), LVC are *"services which (a) provide little or no benefit to patients, (b) potentially cause harm, (c) incur unnecessary costs, and/or (d) waste scarce resources compared to the use of alternative services or the application in an alternative patient group."* As patient-perceived outcomes and medical outcomes might diverge, it is essential to supplement objective medical information with subjective, patient-perceived outcomes to clearly identify LVC. Ultimately, the Swiss healthcare system must provide good quality of care while impeding LVC and slowing the increase in healthcare costs. Achieving this goal requires respecting patients' perceptions—reflected by PROMs and patients' HL—as well as physicians observing medical guidelines and applying the most suitable treatment strategies.

6.1 Discussion of results

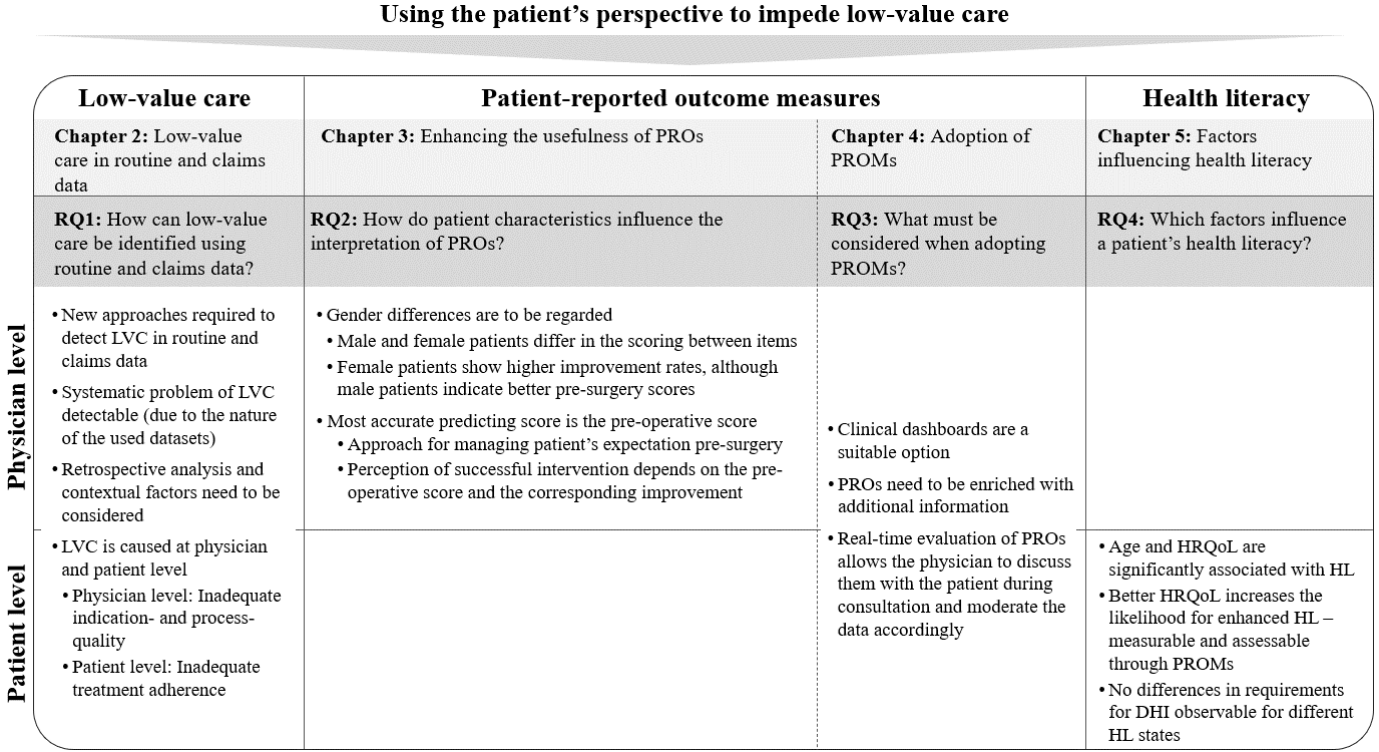
LVC impedes providing good quality of care because patients do not benefit from the provided treatments and might experience iatrogenic harm. Furthermore, additional costs are incurred. As healthcare costs continue to soar in Switzerland, as reflected directly by rising health insurance premiums [1–3], workable solutions are urgently needed. As LVC depends on context, it is crucial to consider patients' circumstances

and HRQoL to avoid *preference misdiagnosis* [4]. This means that a physician implicitly assumes a desired patient treatment goal, which, however, does not reflect the patient's actual goal and thus leads to LVC. As shown, PROMs are well-suited to establishing a patient's preferences and perceptions of HRQoL. However, using PROMs in clinical practice on a larger scale requires demonstrating their enhanced usefulness through appropriate interpretation mechanisms and adoption tools.

Treatment success is not only the physician's responsibility. Patients must also contribute to success, among others, by adhering to treatment plans, which depends strongly on the patient's self-management capabilities [5]. Therefore, the patient becomes responsible for avoiding LVC. This dissertation thus highlights that LVC may be caused by physicians—being representatives of the healthcare system—and by patients. As shown, however, applying PROMs strengthens patient-physician communication and thereby takes a patient-centered approach that allows identifying treatment that benefits patients and does not result in unnecessary costs or harm.

Figure 6.1 highlights the key results by chapter. The results are subsequently discussed in more detail. As already outlined in the introductory chapter of this dissertation, Chapter 2 considers whether LVC is detectable at the physician and patient level, while Chapter 3 focuses on how to interpret PROs on the physician level. Chapter 4 brings the physician and patient level together by investigating PROM application. Chapter 5 discusses patients' HL and which factors are associated with HL. While Chapter 2 discusses how to identify LVC in routine and claims datasets, the remaining chapters explore ways of reducing LVC by either applying PROs or increasing patients' HL.

Figure 6.1: Key results



Legend: LVC = Low-value care, PRO = Patient-reported outcome, PROM = Patient-reported outcome measure, HRQoL = Health-related quality of life, HL = Health literacy, DHI = Digital health intervention

Chapter 2: Low-value care in routine and claims data

Chapter 2 asked *"How can low-value care be identified using routine and claims data?"* (RQ1). Its main finding is that LVC can be identified in routine and claims data. However, as such identification methods are a relatively new field of investigation, strategies for identifying LVC in such data must first be developed. The studies reported in Chapter 2 adopted two different approaches: In Study 1 (S1), only the outcome (i.e., non-medically indicated CS) was analyzed based on diagnoses, procedure classification, and billing data—amounting to a snapshot of a patient's history. Study 2 (S2) observed COPD patients' behavior over two years prior to a critical event (i.e., hospitalized exacerbation). LVC was identified in both studies, albeit from a different perspective: Study 1 focused on the system level, Study 2 on the patient level. Thus, it is essential to determine the type of LVC: Does the treatment itself (see non-medically indicated CS) or rather inaccurate treatment behavior have negative consequences (see COPD patients' exacerbations)?

Basing both studies (S1 & S2) on routine and claims datasets enables holistically assessing LVC in the Swiss healthcare system. As shown, LVC is a systemic problem in Switzerland.

Both proposed approaches outlined in S1 and S2 enable retrospective analysis of the provision of LVC because the outcomes (i.e., non-medically indicated CS or COPD exacerbations based on low treatment adherence) need to be known. This might initiate a discussion about which outcomes are desirable in which context or under what circumstances—as proposed by the definition of LVC in the Introduction. For medication or treatment to be approved in clinical practice, their efficacy must be demonstrated in medical trials. However, whether a specific treatment needs to be classified as LVC depends on context: Certain treatments are effective under certain circumstances for a specific group of patients [6]. Exemplarily, CS are life-saving treatments in case of complications during pregnancy or labor, but it may be classified as LVC in case SVB would also have been an option for giving birth. Therefore, Chapter 2 highlights that the context needs to be considered to classify a treatment as LVC. As in the example of CS, which constitute life-saving treatments under certain

conditions, they increase the risk of morbidity and mortality for mother and child where SVD would also have been a potential birth mode. By contrast, treating hospitalized exacerbation does not amount to LVC per se. However, as shown in the study, hospitalized exacerbations are avoidable if medication adherence is high. Therefore, S1 and S2 both illustrate how routine data and claims data can be used to identify LVC by considering contextual factors.

The different approaches of S1 and S2 show that LVC may be caused by the physician (i.e., who represents the healthcare system) or by the patient. A physician's decision or action may lead to LVC due to inadequate indication and/or process quality. For example, LVC may occur if the physician misinterprets a patient's preferences [4] or disrespects medical guidelines because the patient demands a particular treatment [7, 8]. Conversely, inappropriate patient behavior (e.g., not adhering to treatment or medication plans) might also be responsible for LVC. Showing low medication adherence or other risky behavior (e.g., smoking in the case of COPD patients) [9] might result in complications that more adequate patient behavior could have avoided [10]. Patients do not adhere to treatment for manifold reasons. For instance, Bhattarai et al. [11] and O'Toole et al. [12] discussed barriers to medication adherence, such as a limited understanding of and anxiety about the prescribed medication, limited patient-physician interaction, or concerns about medication side effects. Additionally, Agarwal et al. [13] claimed that discussing medication concerns with COPD patients and stressing the importance of daily medication intake may enhance adherence in individuals with limited HL.

Chapter 2 provides insight into how LVC is detectable in routine and claims data and emphasizes the requirements of different identification approaches. However, once those approaches are determined, they facilitate the retrospective analysis of whether LVC appears under certain circumstances. Furthermore, if certain de-implementation recommendations are introduced as, for example, suggested by *Choosing Wisely* [14] or by *Smarter Medicine*, its Swiss adaption [15], then the proposed approaches enable identifying whether these recommendations are implemented as suggested while respecting the treatment context.

Chapter 3: Enhancing the usefulness of PROs

Chapter 3 asked *"How do patient characteristics influence the interpretation of PROs?"* (RQ2) and addressed three main aspects of PROs: interpreting the current health status (Study 3, S3), changing health status over time (Study 4, S4), and improving health status after intervention to merit calling treatment successful (Study 5, S5). To enhance PRO usefulness, certain aspects, such as gender or pre-intervention PRO scores, need to be considered. Tew et al. [16] and Solarino et al. [17] proposed using PROs as a starting point for improving patient-centered care by respecting the patient's gender and pre-surgical HRQoL.

Chapter 3.[6] explores gender differences in the individual dimensions of a generic PROM (EQ-5D) and in gender-related health improvement rates after interventions. It also considers MIDs as a means of determining when an intervention can be considered successful according to personal preconditions.

S3 reveals that the individual dimensions of generic PROs (i.e., EQ-5D) are differently associated with gender. Female patients exhibit significantly fewer problems with self-care pre- and post-surgery (THA and TKA) and better post-operative mobility scores (THA). By contrast, male patients report significantly less pain/discomfort (TKA) and less anxiety/depression (THA) pre- and post-surgery. Especially when using a weighted index score (as for EQ-5D), HRQoL might be biased as women tend to indicate fewer problems in dimensions with lower weights [18]. Hence, the index score clouds male-female comparability. Analyzing individual dimensions provides more specific information to assess the current health status.

Hertler et al.'s [19] review of HRQoL outcome reports reached similar conclusions: Cancer treatment shows significant association with gender in dimensions such as nausea, pain, or psychological symptoms. Although certain gender differences in PRO

[6] All studies included in Chapter 3 deal with patients undergoing THA or TKA. Therefore, results might differ for other episodes of care.

evaluation are known, Hertler et al. [19] also emphasized the need to further stress the importance of gender differences and called for further research in this area.

When analyzing the health improvement after intervention, S4 highlights that female patients benefit from higher health improvement than male patients. Although male patients show significantly better scores pre-surgery, this effect diminishes within one year as most gender differences in PROs were no longer observable. This might indicate that male patients receive surgical treatment earlier than required or that female patients receive surgery later than required.

Adding to S4, S5 reveals that pre-surgery PRO scores and depression status are most accurate in predicting post-surgery scores. This finding is also in line with Berliner et al. [20]. The results of S5 demonstrate that physicians can use MIDs to manage patients' expectations better by clarifying the realistic amount of improvement in HRQoL when undergoing treatment. Patients can be stratified by pre-surgery scores. This results in patients with worse pre-operative scores requiring larger MIDs. Thus, a patient facing lower pre-surgery HRQoL can expect more improvement in HRQoL compared to a patient dealing with higher pre-surgery HRQoL. In contrast, this finding also implies that patients with lower pre-surgery HRQoL must experience a larger HRQoL improvement post-surgery to perceive the intervention as a success. We found no other gender differences in MIDs.

In conclusion, Chapter 3 shows which factors are relevant to enhancing PRO interpretation. Individual PRO dimensions are more informative for analyzing health status than index scores, especially when dealing with weighted dimensional index scores (see S3). By contrast, assessing the index score is a valid approach when examining changes in a patient's health status over time (see S4). Gender differences play a significant analytical role in S3 and S4. Conversely, gender reveal no significant differences in MIDs (see S5). For MIDs, the pre-operative PRO score is most predictive of HRQoL post-intervention improvement and indicates whether that improvement is high enough to claim successful intervention.

PROs can be used to evaluate whether treatment improves or worsens the patient's situation. In the latter case, the patient would face LVC. Thus, PROs can be used to assess health status and its improvement over time. PROs also serve to evaluate whether a treatment or a change in treatment leads to the desired outcome and its effect on HRQoL. However, gender differences and pre-intervention PRO scores in particular need to be considered when evaluating health status (S3), change over time (S4), and MIDs (S5).

Chapter 4: Adoption of PROMs

Chapter 4 asked "*What must be considered when adopting PROMs?"* (RQ3). Increasing the adoption of PROs in clinical practice requires intervening in patient-physician interaction. One potential approach is to introduce clinical dashboards. To enable broad adoption, such dashboards need to incorporate the requirements presented in Study 6 (S6). Including PROs in patient-physician interaction allows for evaluating treatment approaches from both an objective and a subjective perspective, as the patient's perception is included. Through the more targeted patient-physician discussion, treatment approaches having less desirable outcomes, such as LVC, may be de-prioritized.

By visualizing data, clinical dashboards theoretically reduce the time required for data collection, alleviate cognitive burden, expedite task completion, enhance situational awareness, and encourage adherence to evidence-based safety guidelines [21]. Clinical dashboards offer valuable support to physicians by facilitating the management of highly diverse data sources, including clinical parameters and PROs [22]. A vital advantage of utilizing a dashboard is its ability to provide a succinct data overview throughout the treatment process.

Nevertheless, clinical dashboards should not only include PROs, but also need to be enriched with additional data (e.g., patients' personal data or lab results). Our study identified 25 design principles for developing clinical dashboards incorporating PROMs. These principles can be divided into four concepts: *generic dashboard principles* (subdivided into technical and non-technical setting), *PROM-related*

principles, *PROM-independent principles*, and *adoption requirements*. The principles indicate that additional data needs to be included in clinical dashboards to create added value for physicians consulting the dashboard.

PROM-related principles encompass seven design principles: 1) type of PROM, 2) PROM-system support, 3) illustration of scores, 4) PRO-related goals, 5) peer-group comparison, 6) past assessment PRO score, and 7) alerts. All these principles are add-ons to the PRO visualization itself. Thus, applying these principles in the clinical dashboard increases the informative value for physicians. The interviews conducted as part of S6 revealed that physicians consider the past assessment PRO score, which enables observing changes in PROs over time, and peer-group comparison particularly valuable. PRO interpretation can also be facilitated by setting thresholds, which issue alerts if values are exceeded or not achieved.

Real-time evaluation, enabled by the digital data collection and its direct visualization, allows physicians to discuss PRO results with patients during consultations (i.e., joint reflection on the treatment and its impacts on the patient's HRQoL). By covering various dimensions of generic health—in the case of generic PROMs—or disease-specific symptoms—by disease-specific PROMs—, the physician can discuss health issues in a structured manner [23, 24]. Also, the fact that such innovative dashboards allow physicians to visualize PRO graphs and discuss these directly with patients enhances the basis for effective communication [25]. As PROs become part of consultations, their visualizations of past assessment scores or peer-group comparisons substantiate patient-physician exchange. A possible scenario might be that the patient remembers the development of the HRQoL differently from what was indicated some months ago. The physician can intervene and explain that the patient's perception of the HRQoL development might be biased and that the actual development was better or worse. The same applies to peer-group comparison, where PROs enable more comprehensively estimating whether a patient is performing similarly to the peer group or exhibits outlier values. Dashboard-based PRO visualization allows the physician to moderate the patient's health status more effectively. Patient-physician interaction can also provide a basis for readjusting

treatment choices based on changes in the patient's HRQoL, and thereby counteract LVC.

In summary, Chapter 4 indicates how PROs can be adopted to enhance their use in clinical practice. PROs provide the physician with a more profound understanding of the patient's perceived health status. Based on these insights, decisions on treatment adaptions can be made to increase quality of care and avoid LVC. The tool proposed in S6 is a clinical dashboard that includes PROs and other data to strengthen the informative value of PROs as it allows physicians to contextualize them. Clinical dashboards and their PRO visualizations can also accentuate patient-physician interaction by collecting HRQoL data in a structured way and by facilitating joint discussion. As such, dashboards enable physicians to relate HRQoL changes to treatment changes and to assess with patients whether they are satisfied with their personal health and development over time. Lastly, clinical dashboards can integrate PROs into running systems and, through facilitated use, make physicians more efficient by providing a source of subjective patient data without additional collection effort. As outlined in Chapter 3, regularly collecting PROs enables tracking HRQoL changes over time, and thereby directly analyzing whether HRQoL improves or deteriorates. Thus, if treatment improves medical values, but the patient feels no improvement (e.g., due to side effects), this may indicate LVC. PROs encourage physicians to discuss a patient's HRQoL regularly during consultations and to adjust treatment if needed.

Chapter 5: Factors influencing health literacy

Chapter 5 asked *"Which factors influence a patient's health literacy?"* (RQ4) because patients are also responsible for preventing LVC in the healthcare system. As LVC might be caused by patients' inadequate treatment or medication adherence (see S2), this chapter focuses on patient's HL and associated factors. The ordinal logistic regression analysis in Study 7 (S7) found that higher age and HRQoL (represented by EQ-5D) significantly increase the likelihood of enhanced HL. This means that HRQoL, measurable through PROMs, and HL are interrelated. Also, previous

literature has found that a higher self-perceived quality of life is closely linked with improved patient HL [26–28].

According to S7, HRQoL is the only variable that is significantly associated with HL, which external factors might influence, with age being beyond external influence. HRQoL is quantifiable through PROs, which provide a starting point for enhancing HL. Given that higher HRQoL increases the likelihood for improved HL, interventions should aim to maintain high HRQoL. This, in turn, strengthens patients' HL which enables them to use health information and services more effectively while promoting health and preventing disease [29, 30], and also averting patient-caused LVC.

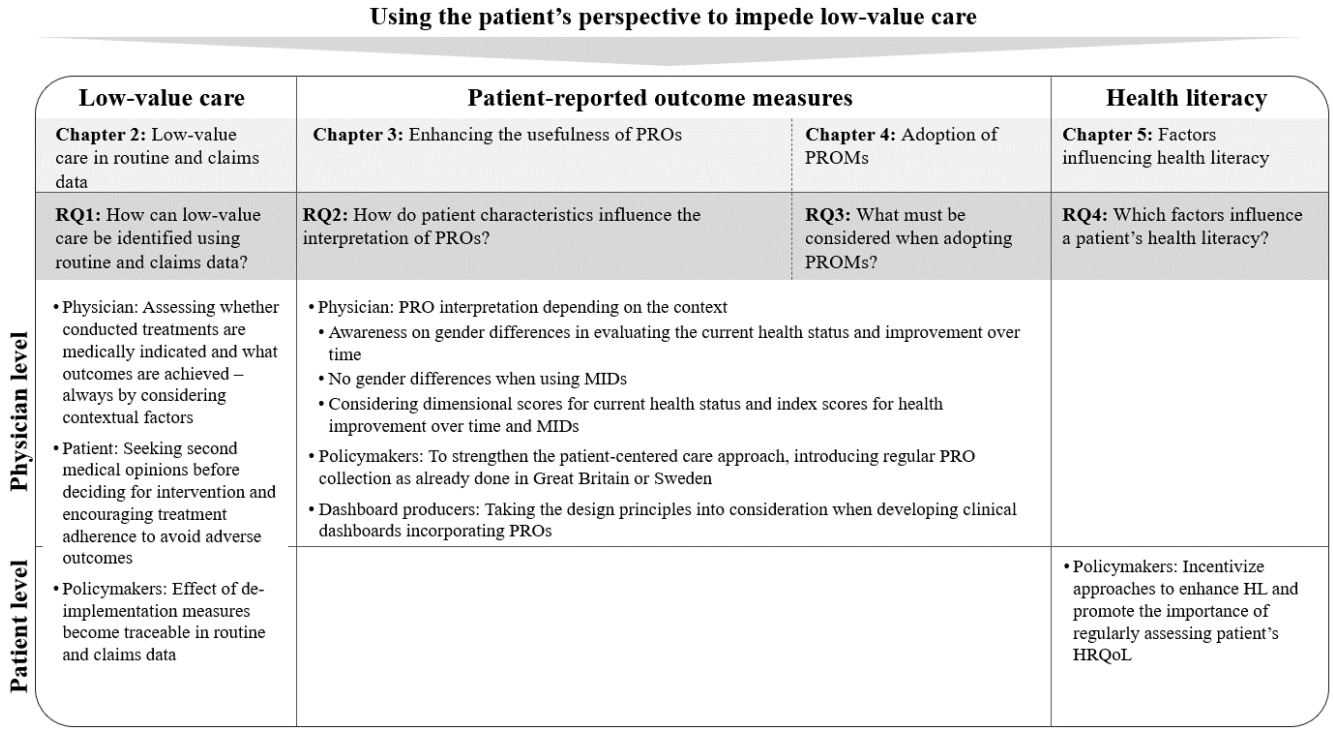
Digital health interventions (DHIs) enable continuously assessing HRQoL (measured through PROs) and providing educational material to strengthen HL. On the one hand, implementing DHIs allows physicians to remotely assess patients' health status and adapt treatments [32, 33]. On the other hand, patients can track their HRQoL independently and access further health information [34]. Implementing DHIs requires determining the requirements of different groups to tailor interventions to user needs. S7 found that COPD patients with different HL neither report varying difficulties in dealing with COPD nor request different symptom support. Rather, they are almost aligned over the different HL states. These results show that a DHI (in the case of COPD patients) need not be tailored to different HL to effectively support users in dealing with their disease in daily life. Accounting for the similar desires of COPD patients for daily support sets the basis for scalability across COPD patient groups with different HL states and hence increases the chance of the DHI becoming cost-effective.

In sum, collecting HRQoL allows physicians to better assess how patients perceive their health status, a key indicator of HL. As discussed, strengthening patients' HL might be further supported by implementing DHIs. The aim should be to develop sufficient HL to ensure patients are more likely to adhere to treatment plans and thus contribute to preventing LVC.

6.2 Implications

Adopting a patient-centered perspective, as proposed by the framework presented here, has implications for stakeholders in the Swiss healthcare system (Figure 6.2). The recommended actions aim to reduce LVC by enabling physicians to consider the patient's perspective in a structured manner by collecting information through PROMs. As shown, establishing PROs' usefulness and adoption further enhances their use in clinical practice. Improving patients' knowledge of disease management (assessed via HL) also helps reduce LVC. The next sections outline the implications of the results for theory and practice.

Figure 6.2: Key implications



Legend: RQ = Research question, PRO = Patient-reported outcome, PROM = Patient-reported outcome measure, MID = Minimal importance difference, HRQoL = Health-related quality of life

Chapter 2 proves that retrospective routine and claims data analysis enable identifying LVC in the healthcare system under consideration of various contextual factors. This finding has implications for physicians, patients, and policymakers. It means that physicans will be able to assess whether treatment is medically indicated and what outcomes are achieved. This more thorough evaluation may incentivize physicians to follow guidelines on whether treatments are indicated or not, and by respecting the individual context. This finding has twofold implications for patients: 1) Second medical opinions can provide information on whether treatment is medically indicated, and might even reduce costs while averting over- and under-treatment [35]; and 2) low treatment adherence is directly interrelated with negative medical consequences. Therefore, patients should be encouraged to adhere to treatment plans. Potential approaches include patient education [36], building a caring and supportive environment [37], or providing supportive mechanisms in the form of mobile apps [38]. Further, a third stakeholder group benefitting from these insights are policymakers, who can orient themselves toward initiatives such as *Choosing Wisely* [14] or *Smarter Medicine* [15] that recommend de-implementing certain treatments under specific circumstances. To evaluate whether such de-implementation recommendations reduce LVC, routine and claims data provide insights into the impact of specific measures.

Chapters 3 and 4 have interrelated implications about enhanced usefulness of PROs and adoption of PROMs. As gender disparities are observable in PROs, gender differences, and their far-reaching implications must be considered when interpreting the outcomes. Especially when PROs are based on weighted index scores (e.g., EQ-5D), the potential for misleading implications becomes prominent. Failure to account for gender disparities in these weighted scores can skew treatment outcomes and effectiveness assessment. Awareness of existing gender differences in EQ-5D dimensions leads physicians to more appropriately assess patients' HRQoL. Assessing individual dimensions is better suited to evaluating health status because physicians can evaluate patients' responses in terms of those dimensions while being aware of gender differences.

In contrast, physicians and other healthcare professionals need to consider index scores to compare health improvement over time. For THA and TKA, especially male patients exhibit better scores pre-surgery. One year post-surgery, however, no significant gender differences are observable anymore. This implies that female patients experience a larger improvement in HRQoL post-surgery than male patients, raising the question of whether male patients undergo THA or TKA too early or female patients too late. By considering PROs and collecting information on health status improvement, future guidelines can be adapted based on the findings to determine the ideal point to undergo surgery for patients by respecting gender.

The findings on MID stratification based on pre-surgery PRO scores have three further implications: 1) Pre-surgery scores can be used to discuss expectations with patients and illustrate which HRQoL can realistically be achieved post-surgery; 2) patients with lower pre-surgery PRO scores must experience higher HRQoL improvement to rate an intervention as a success; and 3) patients with higher pre-surgery PRO scores can only benefit from lower HRQoL improvement due to the intervention than patients with lower pre-surgery PRO scores. These insights benchmark whether an intervention can be rated as successful and which post-surgery scores are to be expected for individual patients.

The clinical dashboard proposed in Chapter 4 guides development process of incorporating PROMs into clinical dashboards to enable the structured evaluation of patients' health status. Visualizing PROs provides a more robust basis for patient-physician communication. The proposed design principles show that merely considering PROs is insufficient and that further information is needed to contextualize PROs. Chapter 4 has also highlighted that dashboard design principles must not differ between episodes of care. This implies that dashboard developers should consider these design principles when developing or building clinical dashboards.

Chapters 3 and 4 also have implications for policymakers. To strengthen patient-centered care in healthcare systems, the patient's perspective must be evaluated regularly. Regular and mandatory PRO collection by the NHS in Great Britain [39] or

by the National Quality Registers (NRQs) in Sweden [40] could serve as role models for adapting similar approaches in Switzerland. By considering patients' insights and experiences, healthcare policies may become more effective, efficient, and equitable. Among others, considering PROs allows policymakers to assess the quality of healthcare services. Tracking PROs enables identifying LVC (i.e., where healthcare services are falling short) and taking data-driven decisions to adjust quality of care. The aim should be to provide high quality of care where patients suffer no iatrogenic harm, complications are prevented, and unnecessary costs can be kept to a minimal level.

If an intervention was successful, the patient might still be dissatisfied with the treatment outcome because too little HRQoL is gained. This HRQoL outcome may then serve as a basis for physicians to discuss possible solutions with patients and try to understand why objective and subjective treatment evaluations are misaligned [41]. Hartzler et al. [42] found that joint PRO evaluation enables meaningful patient-physician discussions on HRQoL.

Chapter 5 implies that not only treatment should be considered in the healthcare context. Greater emphasis should also be placed on preserving and enhancing health. As HRQoL and HL are significantly associated [26–28], policymakers should incentivize approaches to enhance HL and promote the importance of regularly assessing patients' HRQoL. The literature shows that sufficient HL is essential for superior self-management capabilities [43]. Self-management capabilities allow patients to cope with disease-specific symptoms and challenges [44], which will positively impact HRQoL. Thus, enhanced self-management capabilities increase the likelihood of treatment adherence and, in turn, reduce the likelihood of avoidable adverse outcomes such as LVC.

In conclusion, the presented framework proposes identifying LVC in the healthcare system and counteracting shortcomings by applying PROs on the physician and on the patient level. The framework means that PROs provide a data-led basis for structured communication about patients' self-perceived health status and changes through treatment. As LVC is also detectable from routine and claims data, the large-

scale application of PROs might visibly reduce LVC on the system level because the patient's perspective is better accounted for. Not only physicians must avoid LVC but patients must follow and adhere to treatment guidelines enabled by sufficient HL. Else, without sufficient HL, patients can cause unintentionally LVC by not adhering to treatment guidelines. Future research needs to prove whether PROs and regularly assessing HL live up to the expectations of being gamechangers in making a patient-centered approach reality and in sustainably reducing LVC.

6.3 Limitations

Apart from the study presented in Chapter 4, all studies in this dissertation are based on quantitative analysis. A major limiting factor of these studies is data availability. Routine and claims data were used to identify LVC. These data are routinely collected by the Swiss government or by health insurances. Given the nature of the Swiss healthcare and insurance market, no dataset comprises procedures, diagnoses, and medical history. However, to determine LVC, including as many contextual factors as possible leads to more precise results about the provision of LVC. S1 included 98.3% of all women who gave birth in Switzerland between 2014 and 2018. The dataset encompasses information on diagnoses, procedures, and DRGs. However, no information on these women's medical history or personal preferences influencing their chosen mode of birth was available. S2, by contrast, included patients' medical history, prescribed medication, and claims data. Nevertheless, applicable law prevents Swiss health insurances from storing diagnosis- or procedure-related information. Therefore, medication adherence had to be approximated.

Still, the findings of S1 can be considered representative because it included almost all women who gave birth in Switzerland during the observation period. In contrast, S2 relied on claims data of a Swiss health insurance with a market share of 10% [45].

In 2023, 48 health insurances were registered in Switzerland.[7] [46]. The market share of the considered health insurance makes it rank among the largest in Switzerland. So while the dataset is representative of Swiss health insurances, more accumulated data from different health insurances would allow for more generalizable results.

Additionally, as discussed in Chapter 2, new approaches to identifying LVC are needed. S1 and S2 explored specific settings, meaning that the proposed identification approaches cannot be generalized. While this dissertation has identified LVC based on routine and claims data, it has not presented a generalizable approach for identifying various LVC practices.

Chapter 3 has presented approaches to enhancing PROs' usefulness by considering suitable interpretation mechanisms. However, the included studies deal with THA and TKA, neglecting other episodes of care. Similar to previous studies [47–49], the included studies show that gender differences are evident in PROs for THA and TKA. However, further research is needed on whether these differences also exist in other episodes of care. A first approach in this direction has been taken by Hertler et al. [19], who highlighted gender differences in cancer treatment for PRO scores concerning the dimensions of nausea, pain, and anxiety. Nevertheless, further studies are needed to establish whether gender differences are systematically identifiable independent of episodes of care or in how far responses differ for acute and chronic conditions.

S6, the only qualitative research project in this dissertation, investigated PRO adoption by incorporating PROs into clinical dashboards. This primary limitation of S6 is only physician specialists were included. GPs were not included regardless of their high impact on outpatient treatment. Therefore, for further validation, the perspective and expectations of GPs should be considered further to challenge the findings provided by physician specialists and software producers.

[7] As reported by the Swiss Health Insurance Register, valid from October 1, 2023.

Nevertheless, the identified design principles are derived from the two-step approach —based on a scoping literature review and semi-structured interviews with software producers and physicians. The chosen methodology aimed to imitate a real-world scenario by interacting with two key stakeholder groups, who stated their requirements and expectations for clinical dashboards incorporating PROMs to be successfully adapted in the real world.

Chapter 5 investigated which factors influence patients' HL and focused on COPD patients. The findings indicate that higher age and better HRQoL significantly increase the likelihood for enhanced HL. However, the found relation might only be valid for COPD patients. The association between HRQoL and HL has already been addressed in the literature [26–28]. In contrast, the literature has found no clear association between age and HL. Some studies report no significant association [50, 51], while others highlight a negative association between the two variables [26, 52]. However, whether these associations also hold for healthy people and episodes of care other than COPD is questionable. Studies investigating this association in different kinds of cancer such as head and neck cancer [53, 54], breast cancer [55], or prostate cancer [56] reveal similar findings. Still, further research is required to confirm whether HRQoL is generally significantly associated with HL or whether this is only true for some specific episodes of care.

Besides the limitations discussed above, the most significant limitation of this dissertation is the challenge to prove that PROMs might be an appropriate instrument at both the physician and the patient level to sustainably reduce LVC. Although PROM application might serve as a de-implementation intervention [23], this effect needs to be validated by a large-scale project.

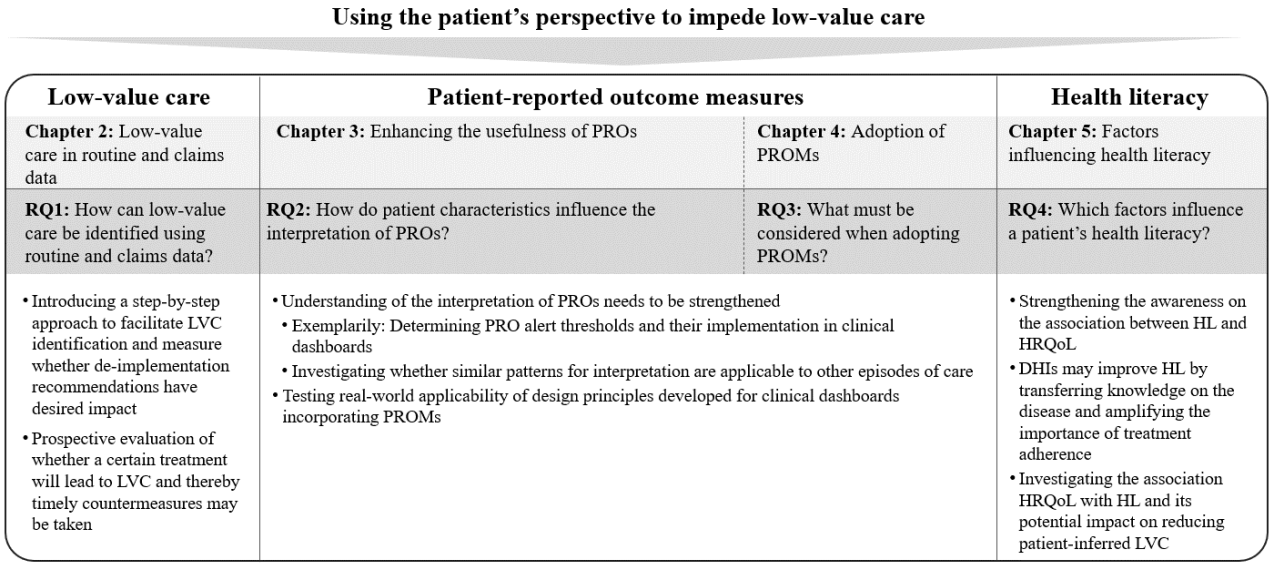
This dissertation has posited that more effectively considering the patient's perspective by applying PROMs will contribute to counteracting LVC. It has also argued that scientific rigor is required to prove the direct impact of PROs on counteracting LVC.

Furthermore, studies are required to validate the impact of enhanced HL on reducing LVC. Patients' enhanced HL might foster a better understanding of the importance of treatment adherence and act accordingly.

6.4 Outlook and recommendations for future research

This dissertation has developed a framework for reducing LVC by considering the patient's perspective. Nevertheless, it has only planted a first seed and further research is necessary to grow this approach. On the one hand, the limitations of the individual studies will help further enrich datasets and cross-validate research approaches to refine the findings. On the other hand, the derived results and implications of the included studies lead to novel and inspiring research proposals (see Figure 6.3).

Figure 6.3: Recommendations for future research



Legend: RQ = Research question, PRO = Patient-reported outcome, MID = Minimal importance difference, HRQoL = Health-related quality of life

6.4.1 Recommendations derived from Chapter 2 (“Low-value care”)

Chapter 2 has demonstrated that LVC is detectable in routine and claims data. However, as outlined in the Limitations, the introduced approaches to identify non-medically CS and low medication adherence are context-specific. The goal should be to provide guidelines on identifying LVC from routine and claims datasets. A step-by-step approach could be introduced to facilitate LVC identification and to better determine whether de-implementation interventions have the desired impact.

Another approach to facilitate LVC identification would be to enrich routine and claims data with PROs and to evaluate whether the patient perceived a high enough gain in HRQoL post-treatment. One role model of PRO collection could be the NHS in Great Britain, which gathers PROs for selected treatments such as hip and knee replacement on a routine basis [39]. Routine PRO collection aims to gather patients’ perceptions of NHS care effectiveness and complements existing information on quality of care [57]. A similar approach is also taken by the Swedish NQRs [40], where additional data types (routine data and PROM) enable objective and subjective treatment evaluation. Although the NHS setting is specific to elective hip and knee replacement, it is conceivable to introduce routine PROM collection for chronic conditions such as cancer care. Objective medical outcomes and subjective perceptions of treatment success might diverge. For example, in the case of prostate cancer, tumor irradiation decreases tumor size; however, as a side effect, patients must deal with urinary leakage, which negatively impacts HRQoL [58]. This raises the question about whether objective value overcompensates HRQoL loss, and thus whether treatment causes LVC. Therefore, supplementing routine data with PRO data might enable identifying LVC even more explicitly by considering objective and subjective evaluation criteria.

Currently, only retrospective analysis enables determining LVC. However, the aim should be to build enough knowledge about which treatments or patient behavior lead to LVC and thus enable pre-identifying potential LVC. Such an approach would lead to better decision-making on whether treatment is needed and whether it will reach its full potential. If treatment does not seem appropriate under certain circumstances or

if the patient is unable to adhere to the treatment plan, timely countermeasures can be taken to avoid LVC. Thus, shifting from retrospective analysis to prospective forecasting would enable learning from past outcomes to predict future ones. Exemplarily, such an approach would inform physicians whether a woman is prone to give birth by non-medically indicated CS. Early countermeasures could then be taken by having the woman participate in a midwife-led care program during pregnancy, which would increase the likelihood of giving birth by SVD. Furthermore, for patients suffering from a chronic disease such as COPD, early awareness of patient medication adherence may lead physicians to take additional measures (e.g., providing educational material) to increase patients' HL and self-management capabilities. This, in turn, would raise awareness of the importance of good medication adherence and its potential for avoiding hospitalized exacerbation.

6.4.2 Recommendations derived from Chapters 3 and 4 ("Patient-reported outcome measures")

Chapter 3 has provided insights into interpreting PRO results depending on the aim of studying PRO data. Chapter 4 explored the integration of PROMs into clinical dashboards. Future research efforts should focus on combining the insights of these chapters: Chapter 3 provides a robust basis for PROM integration into clinical dashboards (as proposed in Chapter 4). Nevertheless, the following requirements must be strengthened before implementing PROMs in clinical dashboards.

One example of the intersection of Chapters 3 and 4 is the development of PRO alert thresholds. On the one hand, appropriate threshold levels need to be determined [59]. This requires establishing guidelines for PRO alerts, such as whether alerts differ between patient groups or according to specific preconditions. On the other hand, implementing PRO alerts in clinical dashboards and the legal consequences must be thoroughly considered. Regulations must clearly stipulate who becomes liable if an alert is not sent out correctly or if the physician does not react to an alert and the patient is harmed as a result. Furthermore, alert fatigue in healthcare professionals due to too many irrelevant alerts should be avoided [60]. Thus, this dissertation opens up avenues for further investigating alert threshold levels (as conducted for other aspects

of PRO interpretation in Chapter 3). More research is required to include PRO alerts into clinical dashboards and to evaluate their informative value once implemented in real-world practice.

The investigated episodes of care in S3 to S5 concern THA and TKA. However, other episodes of care were not analyzed. Future research therefore needs to examine whether similar patterns concerning gender differences or interpretation, in general, are observable across various episodes of care. One question meriting investigation is whether acute (e.g., THA and TKA) and chronic (e.g., COPD) episodes of care are interpreted differently. Especially for observing the progression of chronic diseases, PROs might be an appropriate measure to regularly assess a patient's HRQoL [61, 62]. However, the challenge here is to interpret the values correctly to enhance their informative value and avoid interpretation bias.

Besides the enhanced usefulness of PROs by applying supportive interpretation mechanisms, the real-world adoption of PROs in clinical dashboards needs to be tested. S6 has only outlined a theoretical framework of which design principles need to be considered when developing clinical dashboards incorporating PROMs. Given that S6 was based merely on a scoping literature review and on semi-structured interviews, a discrepancy might exist between what physicians wish for their daily practice and what they need. Therefore, the proposed design principles need to be tested to verify whether additional design principles are required or whether certain design principles could also be de-prioritized.

6.4.3 Recommendations derived from Chapter 5 ("Health literacy")

S7 revealed that age and HRQoL are significantly associated with HL. This finding is limited to COPD patients and thus needs to be further verified for other episodes of care. Nevertheless, the interconnection of HRQoL and HL offers excellent potential for improving patient care. As these concepts are closely interrelated—one measures patient well-being, the other the patient's ability to effectively access and navigate healthcare services—, they may be used to enhance health outcomes and optimize healthcare services. Improving health literacy can empower patients to make informed

decisions, access care, effectively communicate their needs, and adhere to treatment plans, ultimately leading to better quality of life.

A potential approach to improving HL involves applying DHIs where educational material and advice on reducing symptom burdens can easily be shared. Interestingly, S7 outlines that patients—independent of their HL state—wish to have the same support mechanisms (e.g., COPD coping advice, exercises to alleviate COPD symptoms, and comparing their health status with average COPD patients). Thus, future research should strive to establish a DHI incorporating the proposed support mechanisms. It also needs to investigate the state of HL, and whether the developed DHI can sustainbably increase HL. Furthermore, the link of increased HL on HRQoL needs to be explored further, as well as whether such a DHI can increase treatment adherence—and thereby reduce LVC long-term.

6.5 Closing remarks

As its title suggests, this dissertation proposes a blueprint for reducing LVC by leveraging the potential of PROs and patients' HL. For physicians, PROs are a communication tool that enables assessing patients' HRQoL. Consequently, treatment decisions will no longer need to be based solely on medical grounds but also on the patient's specific health status. LVC thus becomes potentially avoidable as additional factors, based on the patient's perception, will also be considered. Nevertheless, using PROs to reduce LVC needs to be further researched. One fruitful approach might involve tracing whether considering PROs leads to different treatment decisions compared to non-consideration and whether PROs have the potential to sustainably reduce LVC.

Additionally, enhanced HL should enable patients to better access healthcare information. Patients' self-management capabilities can be improved if they better understand their disease. This, in turn, will make patients more effectively adhere to treatment plans and medication intake. Future research should investigate how HL relates to averting LVC. If such an association is measurable, further measures can be

developed to increase patients' HL, thereby leading to fewer adverse outcomes, and fewer avoidable costs.

In conclusion: This dissertation has emphasized that both physicians (as representatives of the healthcare system) and patients are responsible for reducing LVC. It has shown how utilizing PROMs and assessing patients' HL can contribute to achieving this important goal.

References

1. Hehli S (2023) Prämien schlagen drastisch auf - Durchschnittsfamilie mit zwei Kindern zahlt 1000 Franken mehr. Neue Zürcher Zeitung

2. MedInside (2023) Steigende Gesundheitskosten: Beschleunigung setzt sich fort. In: MedInside. https://www.medinside.ch/steigende-gesundheitskosten-beschleunigung-setzt-sich-fort-20230811. Accessed 9 Oct 2023

3. SDA/anf (2023) Santésuisse rechnet mit acht bis neun Prozent höheren Krankenkassenprämien. Tagesanzeiger

4. Brownlee S, Chalkidou K, Doust J, et al (2017) Evidence for overuse of medical services around the world. The Lancet 390:156–168.

5. Muellers KA, Chen L, O'Conor R, et al (2019) Health literacy and medication adherence in COPD patients: When caregiver presence is not sufficient. COPD: Journal of Chronic Obstructive Pulmonary Disease 16:362–367.

6. Grimshaw JM, Patey AM, Kirkham KR, et al (2020) De-implementing wisely: Developing the evidence base to reduce low-value care. BMJ Qual Saf 29:409–417.

7. Kool RB, Verkerk EW, Winnemuller LJ, et al (2020) Identifying and de-implementing low-value care in primary care: The GP's perspective-a cross-sectional survey. BMJ Open 10:e037019.

8. Selby K, Cornuz J, Cohidon C, et al (2018) How do Swiss general practitioners agree with and report adhering to a top-five list of unnecessary tests and treatments? Results of a cross-sectional survey. European Journal of General Practice 24:32–38.

9. GOLD (2021) Global Initiative for Chronic Obstructive Lung Disease

10. Vestbo J, Anderson JA, Calverley PMA, et al (2009) Adherence to inhaled therapy, mortality and hospital admission in COPD. Thorax 64:939–943.

11. Bhattarai B, Walpola R, Mey A, et al (2020) Barriers and strategies for improving medication adherence among people living with COPD: A systematic review. Respir Care 65:1738–1750.

12. O'Toole J, Krishnan M, Riekert K, Eakin MN (2022) Understanding barriers to and strategies for medication adherence in COPD: a qualitative study. BMC Pulm Med 22:98.

13. Agarwal P, Lin J, Muellers K, et al (2021) A structural equation model of relationships of health literacy, illness and medication beliefs with medication adherence among patients with chronic obstructive pulmonary disease. Patient Educ Couns 104:1445–1450.

14. Choosing Wisely (2022) Choosing Wisely. https://www.choosingwisely.org/

15. Smarter Medicine (2022) Smarter Medicine. https://www.smartermedicine.ch/de/home.html

16. Tew M, Dalziel K, Clarke P, et al (2020) Patient-reported outcome measures (PROMs): Can they be used to guide patient-centered care and optimize outcomes in total knee replacement? Quality of Life Research 29:3273–3283.

17. Solarino G, Bizzoca D, Moretti AM, et al (2022) Sex and gender-related differences in the outcome of total hip arthroplasty: A current concepts review. Medicina (B Aires) 58:1702.

18. Ludwig K, Graf von der Schulenburg J-M, Greiner W (2018) German Value Set for the EQ-5D-5L. Pharmacoeconomics 36:663–674.

19. Hertler C, Seiler A, Gramatzki D, et al (2020) Sex-specific and gender-specific aspects in patient-reported outcomes. ESMO Open 5:e000837.

20. Berliner JL, Brodke DJ, Chan V, et al (2017) Can preoperative patient-reported outcome measures be used to predict meaningful improvement in function after TKA? Clin Orthop Relat Res 475:149–157.

21. Khairat SS, Dukkipati A, Lauria HA, et al (2018) The impact of visualization dashboards on quality of care and clinician satisfaction: Integrative literature review. JMIR Hum Factors 5:e22.

22. Wilbanks BA, Langford PA (2014) A review of dashboards for data analytics in nursing. CIN: Computers, Informatics, Nursing 32:545–549.

23. Field J, Holmes MM, Newell D (2019) PROMs data: Can it be used to make decisions for individual patients? A narrative review. Patient Relat Outcome Meas Volume 10:233–241.

24. Greenhalgh J, Gooding K, Gibbons E, et al (2018) How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. J Patient Rep Outcomes 2:42.

25. Bantug ET, Coles T, Smith KC, et al (2016) Graphical displays of patient-reported outcomes (PRO) for use in clinical practice: What makes a pro picture worth a thousand words? Patient Educ Couns 99:483–490.

26. Stellefson M, Paige SR, Alber JM, et al (2019) Association between health literacy, electronic health literacy, disease-specific knowledge, and health-related quality of life among adults with Chronic Obstructive Pulmonary Disease: Cross-sectional study. J Med Internet Res 21:e12165.

27. Wahl AK, Osborne RH, Larsen MH, et al (2021) Exploring health literacy needs in chronic obstructive pulmonary disease (COPD): Associations between demographic, clinical variables, psychological well-being and health literacy. Heart & Lung 50:417–424.

28. Fischer C, Jörres RA, Alter P, et al (2022) Basic determinants of disease knowledge in COPD patients: Results from COSYCONET. Patient Prefer Adherence Volume 16:1759–1770.

29. Nutbeam D (1998) Health Promotion Glossary. Health Promot Int 13:349–364.

30. Berkman ND, Davis TC, McCormack L (2010) Health literacy: What is it? J Health Commun 15:9–19.

31. Miller TA (2016) Health literacy and adherence to medical treatment in chronic and acute illness: A meta-analysis. Patient Educ Couns 99:1079–1086.

32. Fleisch E, Franz C, Hermann A (2021) The digital pill: What everyone should know about the future or our healthcare system. Emerald Publishing Limited, Bingley (UK)

33. Kowatsch T, Otto L, Harperink S, et al (2019) A design and evaluation framework for digital health interventions. it - Information Technology 61:253–263.

34. Ding H, Karunanithi M, Kanagasingam Y, et al (2014) A pilot study of a mobile-phone-based home monitoring system to assist in remote interventions in cases of acute exacerbation of COPD. J Telemed Telecare 20:128–134.

35. Tosteson ANA, Yang Q, Nelson HD, et al (2018) Second opinion strategies in breast pathology: a decision analysis addressing over-treatment, under-treatment, and care costs. Breast Cancer Res Treat 167:195–203.

36. López-Campos JL, Quintana Gallego E, Carrasco Hernández L (2019) Status of and strategies for improving adherence to COPD treatment. Int J Chron Obstruct Pulmon Dis 14:1503–1515.

37. Feldman SR, Vrijens B, Gieler U, et al (2017) Treatment adherence intervention studies in dermatology and guidance on how to support adherence. Am J Clin Dermatol 18:253–271.

38. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ (2019) Mobile apps for increasing treatment adherence: Systematic review. J Med Internet Res 21:e12505.

39. NHS Digital (2022) Patient reported outcomes measures (PROMs). In: NHS Official Statistics. https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms

40. Nilsson E, Orwelius L, Kristenson M (2016) Patient-reported outcomes in the Swedish National Quality Registers. J Intern Med 279:141–153.

41. Lavallee DC, Chenok KE, Love RM, et al (2016) Incorporating patient-reported outcomes into health care to engage patients and enhance care. Health Aff 35:575–582.

42. Hartzler AL, Izard JP, Dalkin BL, et al (2016) Design and feasibility of integrating personalized PRO dashboards into prostate cancer care. Journal of the American Medical Informatics Association 23:38–47.

43. Yadav UN, Hosseinzadeh H, Baral KP (2018) Self-management and patient activation in COPD patients: An evidence summary of randomized controlled trials. Clin Epidemiol Glob Health 6:148–154.

44. Barlow J, Wright C, Sheasby J, et al (2002) Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 48:177–187.

45. Groupe Mutuel (2023) Key figures. https://www.groupemutuel.ch/en/groupe-mutuel/our-company/key-figures.html. Accessed 9 Aug 2023

46. BAG (2023) Verzeichnisse der zugelassenen Kranken- und Rückversicherer. https://www.bag.admin.ch/bag/de/home/versicherungen/krankenversicherung/krankenversicherung-versicherer-aufsicht/verzeichnisse-krankenundrueckversicherer.html. Accessed 25 Oct 2023

47. Jansson KÅ, Granath F (2011) Health-related quality of life (EQ-5D) before and after orthopedic surgery. Acta Orthop 82:82–89.

48. Turnbull GS, Scott CEH, MacDonald DJ, Breusch SJ (2019) Gender and preoperative function predict physical activity levels after revision total knee arthroplasty. J Arthroplasty 34:939–946.

49. Lim JBT, Chi CH, Lo LE, et al (2015) Gender difference in outcome after total knee replacement. Journal of Orthopaedic Surgery 23:194–197.

50. Azkan Ture D, Bhattacharya S, Demirci H, Yildiz T (2022) Health literacy and health outcomes in chronic obstructive pulmonary disease patients: An explorative study. Front Public Health 10:.

51. Jayasinghe UW, Harris MF, Parker SM, et al (2016) The impact of health literacy and life style risk factors on health-related quality of life of Australian patients. Health Qual Life Outcomes 14:68.

52. Yadav UN, Lloyd J, Hosseinzadeh H, et al (2020) Levels and determinants of health literacy and patient activation among multi-morbid COPD people in rural Nepal: Findings from a cross-sectional study. PLoS One 15:e0233488.

53. Nilsen ML, Moskovitz J, Lyu L, et al (2020) Health literacy: Impact on quality of life in head and neck cancer survivors. Laryngoscope 130:2354–2359.

54. Clarke N, Dunne S, Coffey L, et al (2021) Health literacy impacts self-management, quality of life and fear of recurrence in head and neck cancer survivors. Journal of Cancer Survivorship 15:855–865.

55. Wei C, Wu M, Tung H (2021) Relationships between health literacy and quality of life among survivors with breast cancer. Int J Nurs Pract 27:.

56. Goodwin BC, March S, Zajdlewicz L, et al (2018) Health literacy and the health status of men with prostate cancer. Psychooncology 27:2374–2381.

57. NHS Digital (2016) Patient Reported Outcome Measures in England - Data Dictionary (Version 3.4.)

58. Lehto U-S, Tenhola H, Taari K, Aromaa A (2017) Patients' perceptions of the negative effects following different prostate cancer treatments and the impact on psychological well-being: a nationwide survey. Br J Cancer 116:864–873.

59. Mooney K, Whisenant MS, Beck SL (2019) Symptom care at home: A comprehensive and pragmatic PRO system approach to improve cancer symptom . Med Care 57:S66–S72.

60. Rudin RS, Perez S, Rodriguez JA, et al (2021) User-centered design of a scalable, electronic health record-integrated remote symptom monitoring intervention for patients with asthma and providers in primary care. Journal of the American Medical Informatics Association 28:2433–2444.

61. Aiyegbusi OL, Nair D, Peipert JD, et al (2021) A narrative review of current evidence supporting the implementation of electronic patient-reported outcome measures in the management of chronic diseases. Ther Adv Chronic Dis 12:.

62. Santana M-J, Feeny D (2014) Framework to assess the effects of using patient-reported outcome measures in chronic care management. Quality of Life Research 23:1505–1513.

Curriculum Vitae

Anja Yvonne Bischof

Born March 21, 1997, in St. Gallen (Switzerland)

Education

02/2021 – 02/2024	University of St. Gallen: PhD Programme in Management
09/2018 – 02/2020	University of St. Gallen: Master of Arts in Business Innovation
09/2015 – 05/2018	University of St. Gallen: Bachelor of Arts in Business Administration

Work experience

07/2020 – 12/2023	School of Medicine University of St. Gallen: Research associate at the Chair of Health Economics, Policy and Mangement
08/2017 – 05/2020	Executive School of Management, Technology and Law University of St. Gallen: Student assistant

Presentation, conference contributions

09/2023	40th EuroQol Plenary 2023, Rome, Italy
05/2023	Life Science Impact Conference, Zurich, Switzerland
05/2023	St. Galler Gesundheitsforum, St. Gallen, Switzerland
03/2024	EuroQol PhD Network, Virtual
03/2023	DGGÖ (Germna Association for Health Economics) Conference, Hannover, Germany
08/2022	EuHEA PhD Conference, Galway, Ireland