

Validierung generischer Lebensqualitätsinstrumente bei alten und sehr
alten Menschen mit Hüftfraktur



Dissertation
zur Erlangung des Doktorgrades
der Humanwissenschaften
(Dr. sc. hum.)

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Magdalena Rohr
aus
Straubing

im Jahr
2022

FREISEITE

Validierung generischer Lebensqualitätsinstrumente bei alten und sehr
alten Menschen mit Hüftfraktur



Dissertation
zur Erlangung des Doktorgrades
der Humanwissenschaften
(Dr. sc. hum.)

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Magdalena Rohr
aus
Straubing

im Jahr
2022

Dekan:

Prof. Dr. Dirk Hellwig

Betreuer:

Prof. Dr. Christian Apfelbacher PhD

Inhalt

Abbildungsverzeichnis	6
Tabellenverzeichnis	7
Abkürzungsverzeichnis	9
1. Einleitung	11
1.1 Epidemiologie der Hüftfraktur	11
1.2 Lebensqualität als Konstrukt	12
1.3 Messung der Lebensqualität	14
1.4 Psychometrische Messeigenschaften	15
1.4.1 Validität	15
1.4.2 Reliabilität	15
1.4.3 Änderungssensitivität	16
1.4.4 Wichtigkeit der einzelnen Messeigenschaften	16
1.5 Instrumente zur Messung der generischen HRQOL bei Hüftfrakturpatienten .	17
1.6 Aufgabenstellung und Ziele	18
2. Material und Methoden	20
2.1 Instrumente	20
2.1.1 Generisch - präferenzbasiert	20
2.1.2 Generisch - nicht präferenzbasiert	21
2.2 Studiendesigns	23
2.3.1 Qualitative Studie	23
2.3.2 Quantitative Studie	29
3. Ergebnisse	42
3.1 Qualitative Studie	42
3.1.1 Beschreibung der Studienteilnehmenden	42
3.1.2 Vergleichende Analyse der Geschlechter und Altersgruppen	42

3.1.3 Aspekte der Machbarkeit und generelle Eindrücke	43
3.1.4 Aspekte der Inhaltsvalidität.....	46
3.2 Quantitative Studie.....	54
3.2.1 Beschreibung der Studienteilnehmenden	54
3.2.2 Deskriptive Ergebnisse der Lebensqualitätsfragebögen und Interpretierbarkeit	54
3.2.3 Analysen zur Beschreibung der Validität	56
3.2.4 Analysen zur Beschreibung der Reliabilität	64
3.2.5 Analysen zur Beschreibung der Änderungssensitivität.....	64
4. Diskussion	69
4.1 Qualitative Studie	69
4.1.1 Hauptergebnisse	69
4.1.2 Vergleich mit anderen Studien.....	70
4.1.3 Stärken und Schwächen.....	74
4.1.4 Schlussfolgerungen	76
4.2 Quantitative Studie.....	76
4.2.1 Zusammenfassung und Rückblick auf zentrale Ergebnisse	77
4.2.2 Einordnung in die bestehende Forschung	78
4.2.3 Limitationen	80
4.2.4 Anwendbarkeit und zukünftige Forschung.....	82
4.3. Ansatzpunkte zur Ableitung von Empfehlungen für den Einsatz in der klinischen Forschung	84
5. Zusammenfassung in deutscher und englischer Sprache	86
6. Literaturverzeichnis	92
7. Anhang.....	104
7.1 Qualitative Studie	104
7.1.1 Erhebung Patientencharakteristika.....	104

7.1.2 Aufklärungsdokument.....	105
7.1.3 Einwilligungsdokument.....	107
7.1.4 Ethik.....	110
7.1.5 Interviewleitfaden.....	114
7.1.6 Übersicht Fragebögen.....	115
7.1.8 WHOQOL-BREF.....	116
7.1.9 NHP.....	120
7.2 Quantitative Studie.....	122
7.2.1 Analyseplan.....	122
7.2.2 R-Code.....	136
Danksagung.....	154
Lebenslauf.....	155
Selbstständigkeitserklärung.....	161

Abbildungsverzeichnis

Abbildung 1: Übersicht über die Messeigenschaften und deren Beziehung basierend auf der COSMIN Taxonomie nach De Vet (36).	16
Abbildung 2: Arbeitspakete der Dissertation.....	19
Abbildung 3: Foto der verwendeten Fragebögen.....	28
Abbildung 4: Darstellung des ProFem Studiendesigns aus Andrich (2019) (14)	31

Tabellenverzeichnis

Tabelle 1: Vergleich der verwendeten Instrumente zur Messung der gesundheitsbezogenen Lebensqualität	24
Tabelle 2: Themen des Interviewleitfadens	28
Tabelle 3: Erhobene Variablen und jeweilige Erhebungszeitpunkte	32
Tabelle 4: Übersicht über die quantitativen Analysen und jeweiligen Grenzwerte zur Bewertung der jeweiligen Messeigenschaft.....	36
Tabelle 5: Hypothesen zur Ermittlung der Diskriminanz- und Konvergenzvalidität...	38
Tabelle 6: Ergebnisse zu Aspekten der Machbarkeit im Vergleich der untersuchten Fragebögen	45
Tabelle 7: Ergebnisse zu Aspekten der Inhaltsvalidität im Vergleich der untersuchten Fragebögen	52
Tabelle 8: Konfirmatorische Faktorenanalyse: Indizes der Modelanpassung	56
Tabelle 9: EQ-5D: Items und Itemkennwerte (standardisierte Faktorladungen und Standardfehler).....	57
Tabelle 10: Deskription der EQ-5D-5L Antworten zur Baseline	58
Tabelle 11: Deskription der SF-12 Antworten zur Baseline	59
Tabelle 12: SF-12: Items und Itemkennwerte (standardisierte Faktorladungen und Standardfehler).....	61
Tabelle 13: Konvergenz- und Diskriminanzvalidität: Korrelationen (Spearman´s rho) zwischen den Lebensqualitäts- und Referenzmaßen	62
Tabelle 14: Known-groups Validität: Vergleiche des SF-12, EQ-5D Index und EQ VAS hinsichtlich verschiedener Subgruppen	63
Tabelle 15: Scores der HRQOL Instrumente zu den Zeitpunkten 3 und 6 Monaten und entsprechender Change Score.....	65
Tabelle 16: Änderungssensitivität der Lebensqualitätsmessinstrumente anhand der selbstberichteten Veränderung (Ankerfrage)	66

Tabelle 17: Korrelationen der Change Scores der Lebensqualitäts- und Referenzinstrumente: Abgleich mit a priori-Hypothesen.....	67
---	----

Abkürzungsverzeichnis

ADL	Aktivitäten des täglichen Lebens [engl. activities of daily living]
ANOVA	Varianzanalyse [engl. analysis of variance]
AOK RH	Allgemeine Ortskrankenkasse Reihnland/Hamburg
CFA	konfirmatorische Faktoranalyse [engl. confirmatory factor analysis]
COSMIN	Consensus-based Standards for the Selection of Health Measurement Instruments
CTA	concurrent think-aloud
CTT	Klassische Testtheorie [engl. classical test theory]
EP	Ersatzproxy
ES	Effektgröße [engl. effect size]
GRCS	Global Rating of Change Scale
HRQOL	gesundheitsbezogene Lebensqualität [engl. health-related quality of life]
ICF	Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit
IMET	Index zur Messung von Einschränkungen der Teilhabe
IQR	Interquartilsabstand [engl. interquartile range]
ISOQOL	International Society for Quality of Life Research
M	Mittelwert
Md	Median
MID/MIC	kleinster bedeutsamer Unterschied/ kleinste bedeutsame Veränderung [engl. minimal important difference/minimal important change]
MCS-12	psychische Summenskala [engl. mental component score]
NHP	Nottingham Health Profile
OHS	Oxford Hip Score
OSS-3	Oslo Social Support Skala
PCS-12	körperliche Summenskala [engl. physical component score]
PFF	Proximale Femurfraktur
PROMs	patientenberichtete Messinstrumente [engl. patient-reported outcome measures)
PROs	patientenberichtete Outcomes [engl. patient-reported outcomes]
QOL	Lebensqualität [engl. quality of life]

QUALYs	Qualitätskorrigierte Lebensjahre [engl. quality-adjustes life years]
RCT	randomisierte kontrollierte Studie [engl. randomised controlled trial]
SD	Standardabweichung [eng. standard deviation]
SEM	Strukturgleichungsmodell [engl. structural equation modeling]
SF-12	Short-Form-12 Fragebogen
SRM	Standardisierte Mittelwertdifferenz [engl. standardised response mean)
SSS	Subjektiver sozialer Status
TN	Teilnehmende (selbst)
TN/ZP	Teilnehmende mit Zusatzproxy
VAS	visuelle Analogskala
WHO	Weltgesundheitsorganisation [engl. World Health Organization]
WHOQOL-BREF	WHO Lebensqualitätsfragebogen Kurzform

Anmerkung: Zur Verwendung gendergerechter Sprache orientiert sich diese Arbeit am aktuellen Leitfaden zur Verwendung gendergerechter Sprache der Universität Regensburg (Stand 02/2019).

1. Einleitung

1.1 Epidemiologie der Hüftfraktur

Proximale Femurfrakturen (PFF), umgangssprachlich Hüftfrakturen, gehören zu den häufigsten Frakturtypen älterer Menschen und sind oft ein prognostisch einschneidendes Ereignis (1, 2). Hüftfrakturen werden gemäß anatomischer Lokalisation und Frakturtyp klassifiziert und können im klinischen Alltag grob in Schenkelhalsfrakturen und Frakturen am metaphysären Femur (per- bis subtrochanter) eingeteilt werden (3). In Deutschland werden jährlich ca. 100.000 Patientinnen und Patienten aufgrund einer Hüftfraktur stationär aufgenommen, wobei die Inzidenz mit höherem Alter ansteigt (4, 5). Dies ist hauptsächlich durch eine weichere Beschaffenheit der Knochen im Alter sowie ein höheres Sturzrisiko begründet (6). Dementsprechend ist aufgrund der zunehmenden Alterung der Bevölkerung ein Anstieg an Hüftfrakturen zu erwarten (7). Im Durchschnitt sind Frauen, die eine Hüftfraktur erleiden, 77 Jahre alt und Männer 72 Jahre alt, wobei aufgrund eines höheren Anteiles an Osteoporose mehr Frauen als Männer betroffen sind (8, 9).

Hüftfrakturen erhöhen das Sterblichkeits- und Krankheitsrisiko bei älteren Patientinnen und Patienten erheblich (10, 11). Unter den Betroffenen sind Bewohner:innen von Pflegeheimen, Männer, Personen über 90 Jahre und Personen mit kognitiven Beeinträchtigungen besonders gefährdet (12, 13). Eine Metaanalyse prospektiver Studien zeigt, dass das relative Hazard für das Versterben in den ersten drei Monaten nach einer Hüftfraktur bei älteren Frauen 5,75 und bei älteren Männern 7,95 beträgt (14) und rund 20% innerhalb eines Jahres verstarben (15). Auch Zahlen einer großen Beobachtungsstudie mit 58.000 Patientinnen und Patienten aus Deutschland zeigen je nach Behandlungsmanagement eine hohe Mortalität zwischen ca. 10-13% innerhalb der ersten 30 Tage (16).

Neben der erhöhten Mortalität gibt es weitere bedeutsame Folgen. Dieser vergleichsweise hohe Morbiditätsgrad führt allerdings dazu, dass diese Personengruppe nur beschränkt in Studien eingeschlossen werden kann, wodurch die Kenntnisse über die Situation der Betroffenen nach einer PFF limitiert sind (17, 18). Die vorhandenen Studien weisen auf schlechte Outcomes nach PFF hin, so behielten z.B. 50% der Betroffenen Einschränkungen in der Funktionsfähigkeit (19) und ca. 15% wurden neu in ein Pflegeheim aufgenommen (20).

Das Outcome nach einer Hüftfraktur wird entscheidend von der Komplexität der Erkrankung, den zahlreichen Komorbiditäten und den Lebensumständen beeinflusst (21, 22). Seitdem zunehmend deutlich wird, dass Hüftfrakturpatient:innen aufgrund der komplexen Krankheitsbilder und Komorbiditäten eine schwierige Aufgabe für das Gesundheitssystem darstellen, hat sich der Fokus der Outcomemessung verschoben. Neben den traditionellen Maßen wie Mortalität und Operationserfolg ist es besonders wichtig, bei Hüftfrakturpatient:innen zusätzlich auch patientenberichtete Outcomes (=PROs) [engl. patient-reported outcome(s)] zu erfassen. PROs können verschiedene Konstrukte wie Aktivitäten des täglichen Lebens und Teilhabe erfassen, wobei die (gesundheitsbezogene) Lebensqualität (=QOL) [engl. quality of life] als wichtigstes Konstrukt angenommen wird. Dies zeigt auch ein Auszug aus der Patientenleitlinie zur Schenkelhalsfraktur des Erwachsenen:

„Ziel der Behandlung ist es, das Überleben zu sichern, die Lebensqualität zu erhalten, Komplikationen und funktionelle Einschränkungen nach Schenkelhalsfraktur zu minimieren.“(23)

Um dieses Ziel messen zu können, wird empfohlen, ein Maß für die gesundheitsbezogene Lebensqualität (=HRQOL) [engl. health-related QOL] in das core-outcome set für Hüftfrakturstudien aufzunehmen (24). Unter einem core-outcome set wird die minimale Anzahl zu messender Outcomes (Endpunkte) in Studien in einem bestimmten Bereich definiert (25). Im Bereich der Hüftfrakturforschung existieren eine ganze Reihe von Fragebögen, die bereits verwendet werden. Jedoch bleibt unklar, welche der etablierten Instrumente für die Messung der Lebensqualität bei alten Menschen mit Hüftfraktur geeignet sind (26).

Dies ist mitunter darin begründet, dass die Messung eines latenten Konstruktes wie Lebensqualität klare Definitionen voraussetzt. Im nächsten Abschnitt wird das Konstrukt der Lebensqualität daher näher erläutert.

1.2 Lebensqualität als Konstrukt

Der Begriff Lebensqualität wird heute in allen Lebensbereichen verwendet und hat in den letzten Jahrzehnten einen rasanten Aufstieg erlebt und sogar zur Gründung einer eigenen Fachgesellschaft (International Society for Quality of Life Research, ISOQOL; <https://www.isoqol.org>) geführt. Zudem wird die Messung der Lebensqualität als potentielle Möglichkeit gesehen, die Gesundheitsversorgung zu verbessern (27). Dies

zeigt, wie multidisziplinär das Thema Lebensqualität verstanden wird. Es wird in der Soziologie, Psychologie, der klinischen Forschung und der Ökonomie bearbeitet. Diese Multidisziplinarität erschwert es allerdings, eine einzelne allgemeingültige Definition zu finden. So wurde bereits vor 40 Jahren die folgende Aussage getroffen, die sich wohl unter Beteiligung verschiedenster Disziplinen noch weiter bestärkt:

„Insgesamt haben es die Sozialwissenschaftler nicht geschafft, eine konsistente und prägnante Definition der Lebensqualität zu erarbeiten. Die Aufgabe ist tatsächlich schwierig, denn Definitionen der Lebensqualität sind weitgehend eine Frage der persönlichen oder Gruppenpräferenzen; verschiedene Menschen schätzen unterschiedliche Dinge.“ (28)

QOL kann in umwelt-, strukturelle und gesundheitsbezogene Aspekte unterteilt werden. Speziell im Alter ist Lebensqualität als Outcome eine interaktive Kombination aus Lebenswegfaktoren und unmittelbarer situationsbezogener Faktoren. Laut Fisher (1992) basiert die Bewertung der QOL bei Älteren auf der Grundlage ihrer Erwartungen, die durch ihre Lebenserfahrungen begründet sind (29). Dies erschwert die Messung. Allerdings ist klar, dass im Vergleich zu Jüngeren die Gesundheit und darunter vor allem die Mobilität bei älteren Menschen eine weitaus größere Rolle spielt (30). Die Konzentration auf die Lebenszufriedenheit älterer Menschen verlagerte sich in den letzten Jahrzehnten auf die Lebensqualität in Bezug auf Gesundheit und Gesundheitsversorgung. Um zwischen Lebensqualität im weiteren Sinne und der Auffassung der Lebensqualität in der klinischen Medizin zu unterscheiden, wird der Begriff "gesundheitsbezogene Lebensqualität" geläufig verwendet (31). Gleichwohl damit die Aspekte von QOL verstanden werden, die mit Gesundheit und Krankheit zu tun haben, werden dafür unterschiedliche Begrifflichkeiten wie Lebensqualität [engl. quality of life], gesundheitsbezogene Lebensqualität [engl. HRQOL], Gesundheitszustand [engl. health state] und Wohlbefinden [engl. well-being] verwendet. Dies verstärkt die konzeptionelle Verwirrung und die Zweifel an der Messbarkeit dieses Konstrukts.

„Quality of life...is implicit and highly subjective and seems to be a concept that is hard to express as a quantity.“(32)

Im nächsten Abschnitt werden verschiedene bisherige Versuche und Ansätze zur Messung der HRQOL skizziert.

1.3 Messung der Lebensqualität

Es ist hinreichend bekannt, dass die Ansichten der Patientinnen und Patienten generell und in Bezug auf die Lebensqualität von denen der Forscherinnen und Forscher bzw. Ärztinnen und Ärzte abweichen können und deshalb eine Stärkung der Patient:innenperspektive unter dem Stichwort „Patientenzentrierte Versorgung“ vorangetrieben wird. Außerdem sind verschiedene Arten von Messinstrumenten für verschiedene Zwecke geeignet (33). Dies führt zu einer großen Anzahl an verschiedenen Instrumenten, die dementsprechend in ihrem Verwendungszweck, der Länge, dem Inhalt, der Erhebungsform und den Messeigenschaften variieren.

Grundsätzlich kann zwischen krankheitsspezifischen und generischen Messinstrumenten unterschieden und gewählt werden. Generische Instrumente sind für die Beurteilung von Personen mit und ohne gesundheitliche Beeinträchtigungen geeignet und erlauben dadurch einen Vergleich zwischen Gruppen mit und ohne eine bestimmte Erkrankung sowie mit einer Bevölkerungsnorm. Andere patientenberichtete Messinstrumente (=PROMs) [engl. patient-reported outcome measures) beziehen sich auf eine bestimmte Krankheit oder Behandlung. Diese Instrumente werden häufig dafür entwickelt, in einer klinischen Studie das Ansprechen auf eine Behandlung nachweisen zu können.

Ein Artikel von Bryant et al. (34) über die Selektion von Outcome Messinstrumenten bei Patientinnen und Patienten mit Hüftfraktur berichtet als Ergebnis, dass es für die Interpretierbarkeit der Ergebnisse über Studien mit ähnlichen und anderen Erkrankungen hinweg wichtig sei, nicht nur ein krankheitsspezifisches, sondern auch ein generisches Messinstrument zu inkludieren. In dieser Arbeit liegt das Augenmerk ausschließlich auf der Untersuchung generischer Instrumente. Diese Fokussierung beruht unter anderem darauf, dass Hüftfrakturpatient:innen oftmals eine Mehrzahl an Komorbiditäten mit sich bringen und eine Abgrenzung auf die Lebensqualitätskomponente, die sich auf die Hüftfraktur bezieht, schwer vorzunehmen wäre. Diese Beschränkung soll keine negative Bewertung krankheitsspezifischer Instrumente bedeuten, sondern der schrittweisen Klärung in Frage kommender Instrumente dienen. Welche Instrumente letztendlich verwendet werden sollen bzw. empfohlen werden können, hängt von den Messeigenschaften in der jeweiligen Zielgruppe ab.

Im Folgenden wird inhaltlich kurz auf die grundlegenden Messeigenschaften eingegangen, wobei die Messung der jeweiligen Eigenschaft im Methodenteil erläutert wird.

1.4 Psychometrische Messeigenschaften

Auch im Bereich der Messeigenschaften werden in der Literatur unterschiedliche Terminologien und Definitionen verwendet. Bereits die synonymen Bezeichnungen für die Beurteilung der Qualität eines Fragebogens, nämlich „Gütekriterien“, „psychometrische Eigenschaften“ [engl. psychometric properties] und „Messeigenschaften“ [engl. measurement properties], zeigen diese unterschiedlichen Konventionen. Diese Arbeit beruft sich auf die Taxonomie der COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) Initiative (35), die in einer Delphi-Studie entwickelt wurde. Diese umfasst alle relevanten Messeigenschaften für die Bewertung von gesundheitsbezogenen PROMs.

Insgesamt gilt zu beachten, dass es sich bei der Evaluation von Gütekriterien um einen kontinuierlichen Prozess handelt, da Gütekriterien nur für die untersuchte Population und Aufgabenstellung gültig sind und die Kriterien bei Verwendung in einer anderen Gruppe neu evaluiert werden müssen. So ist beispielsweise Reliabilität keine Eigenschaft eines Instruments, sondern die Eigenschaft von mit diesem Instrument vorgenommenen Messungen in einer bestimmten Population. Grundsätzlich lassen sich die drei Gütekriterien Validität, Reliabilität und Änderungssensitivität unterscheiden. Abbildung 1 zeigt eine Übersicht über die Messeigenschaften.

1.4.1 Validität

Validität bezieht sich auf den Grad, zu dem ein Fragebogen/Assessment/Instrument das Konstrukt misst, das es zu messen vorgibt. Es wird zwischen Inhaltsvalidität, Konstruktvalidität (einschließlich struktureller Validität, Hypothesentestung und Kulturübergreifender Validität) und Kriteriumsvalidität unterschieden (35, 36).

1.4.2 Reliabilität

Mokkink et al. (35) definiert Reliabilität als „den Grad, zu dem eine Messung frei von Messfehlern ist“. Häufig berichtet wird die interne Konsistenz. Diese „ist ein Maß für die Homogenität eines eindimensionalen Assessments, bei dem mehrere Items dasselbe Konstrukt erheben“ (37). Weisen die Items eine eindimensionale Struktur auf,

kann die Reliabilität für das Gesamtkonstrukt geprüft werden, ansonsten wird Reliabilität für jede Dimension getrennt erfasst.



Abbildung 1: Übersicht über die Messeigenschaften und deren Beziehung basierend auf der COSMIN Taxonomie nach De Vet (36).

1.4.3 Änderungssensitivität

Die Änderungssensitivität oder auch Responsivität bezieht sich auf die Fähigkeit eines Messinstrumentes, Veränderungen im Laufe der Zeit in dem zu messenden Konstrukt zu erkennen (35).

1.4.4 Wichtigkeit der einzelnen Messeigenschaften

Unter allen Messeigenschaften wird die Inhaltsvalidität als wichtigste Messeigenschaft angesehen, da zunächst alle Inhalte eines Messinstrumentes im Hinblick auf das Konstrukt von Interesse und die Zielpopulation relevant, umfassend und verständlich sein sollten. Nach der Inhaltsvalidität wird die interne Struktur eines Instruments, einschließlich der Messeigenschaften strukturelle Validität, interne Konsistenz und kulturübergreifende Validität als wichtig angesehen. Schließlich werden die übrigen Messeigenschaften als bedeutend eingestuft, d.h. Reliabilität, Messfehler, Kriteriumsvalidität, Hypothesentestung und Änderungssensitivität.

1.5 Instrumente zur Messung der generischen HRQOL bei Hüftfrakturpatienten

Der Fokus auf PROs hat zur Entwicklung und Verwendung verschiedenster Instrumente beigetragen, so dass z.B. Wissenschaftler:innen bei der Planung einer Studie oder Gesundheitsprofessionelle bei der Messung von PROs in der Routineversorgung häufig unsicher sind, welches Instrument sie verwenden sollen. Bisher gibt es keine Leitlinie oder klare Empfehlungen für die angemessene Messung der generischen Lebensqualität nach Hüftfraktur, in Frage kommende Instrumente können aber durch drei Übersichtsarbeiten eingegrenzt werden.

Das eigentliche Ziel der Literaturübersicht von Peeters et al. (21) war es, eine Übersicht über den Gesundheitszustand und die Lebensqualität von älteren Menschen nach Hüftfraktur zu geben. Die Autoren listeten aber auch die Fragebögen auf, die in den eingeschlossenen prospektiven oder randomisierten kontrollierten Studien (=RCT) [engl. randomised controlled trial] verwendet wurden. Demnach gibt es vier Fragebögen, die häufig eingesetzt werden (Short-Form-36 Fragebogen (SF-36), EQ-5D, Short-Form-12 Fragebogen (SF-12) und Nottingham Health Profile (NHP)) und einige Fragebögen, die nur vereinzelt verwendet werden (z.B. health utility index (HUI2)). Darüber hinaus empfehlen die Autoren den Einsatz zweier Fragebögen (Kurzform des WHO Lebensqualitätsfragebogens (WHOQOL-BREF), ICEpop CAPability Fragebogen für ältere Menschen (ICECAP-O)), die in keiner der eingeschlossenen Studien verwendet wurden (21).

Eine weitere Übersichtsarbeit nannte die Diversifizierung der Outcomemessung, neben den traditionellen Maßen wie Mortalität, bei älteren Patientinnen und Patienten nach Hüftfraktur als Grund, um die Häufigkeit und die Anwendung anerkannter Outcome Instrumente u.a. zur Messung der Lebensqualität zu ermitteln (38). Die Studie identifizierte 14 Instrumente, die für den Einsatz bei älteren Patient:innen validiert sind, darunter am häufigsten Instrumente zur Messung von Aktivitäten des täglichen Lebens (=ADL) [engl. activities of daily living] gefolgt von Lebensqualitätsinstrumenten und vieler weiterer nicht validierter Instrumente. Auch in dieser Studie konnte demnach nicht ein einziges Instrument identifiziert werden, sondern die Auswahl wurde auch hier auf vier häufig eingesetzte Instrumente (SF-36, EQ-5D, SF-12 und NHP) zur Messung der generischen Lebensqualität eingeschränkt.

Die dritte Arbeit stellt die erste Übersichtsarbeit dar, die nicht nur die Messeigenschaften der Fragebögen, sondern auch die Qualität der zugrundeliegenden Studien bewertete. Die Autorengruppe um Haywood et al. (26) identifizierte zum Zeitpunkt des Reviews 28 verschiedene PROMs, die bei älteren Hüftfrakturpatient:innen verwendet wurden, worunter zwölf generische Instrumente waren. Unter diesen generischen Instrumenten gab es für die Fragebögen SF-36 und EQ-5D (3L) die meiste Evidenz für akzeptable Messeigenschaften. Allerdings ist die Evidenz hinsichtlich der praktischen Eigenschaften bzw. der Relevanz innerhalb der untersuchten Gruppe limitiert. Für alle anderen generischen Instrumente, u.a. NHP, SF-12, WHOQOL-BREF gibt es nur wenig, widersprüchliche oder keine Evidenz für Reliabilität, Validität und Änderungssensitivität (26, 39). Die Autoren empfehlen, sich in zukünftigen Studien auch u.a. auf Akzeptanz, Relevanz, Machbarkeit und dem Einbezug von Patientinnen und Patienten zu fokussieren.

Die Reviews zeigen, dass für die Messung der HRQOL bei Hüftfrakturpatient:innen verschiedene Instrumente im Einsatz sind und bestimmte Messeigenschaften bereits überprüft wurden, wodurch sich die Auswahl an potenziellen Fragebögen für die Untersuchungen im Rahmen dieser Arbeit bereits auf vier bzw. zwei dieser Fragebögen beschränkt (21, 26, 38). Die noch vorhandenen Forschungslücken bei der Anwendung dieser Fragebögen werden in dieser Arbeit adressiert. Den Empfehlungen der Reviews folgend werden die psychometrischen Messeigenschaften dieser Fragebögen in der Gruppe der älteren Menschen nach Hüftfraktur weiter geprüft. Außerdem wird wie für die weitere Forschung empfohlen, die Zielgruppe bei der Überprüfung von Aspekten der Inhaltsvalidität und Machbarkeit direkt mit eingebunden. Im folgenden Abschnitt werden die Ziele aufgeführt und die entsprechenden Arbeitspakete werden in Abbildung 2 zusammenfassend dargestellt.

1.6 Aufgabenstellung und Ziele

Ziel dieser Dissertation ist es, die...

- A) Inhaltsvalidität und Machbarkeit von generischen PROMs zur Messung der HRQOL (SF-12, EQ-5D-5L Gesundheitsprofil, NHP, WHOQOL-BREF) bei älteren Menschen (≥ 60 Jahre) nach Hüftfraktur zu untersuchen und

- B) Validität, Reliabilität und Änderungssensitivität der zwei am häufigsten in dieser Zielgruppe verwendeten Messinstrumente (SF-12, EQ-5D-5L) zu testen.

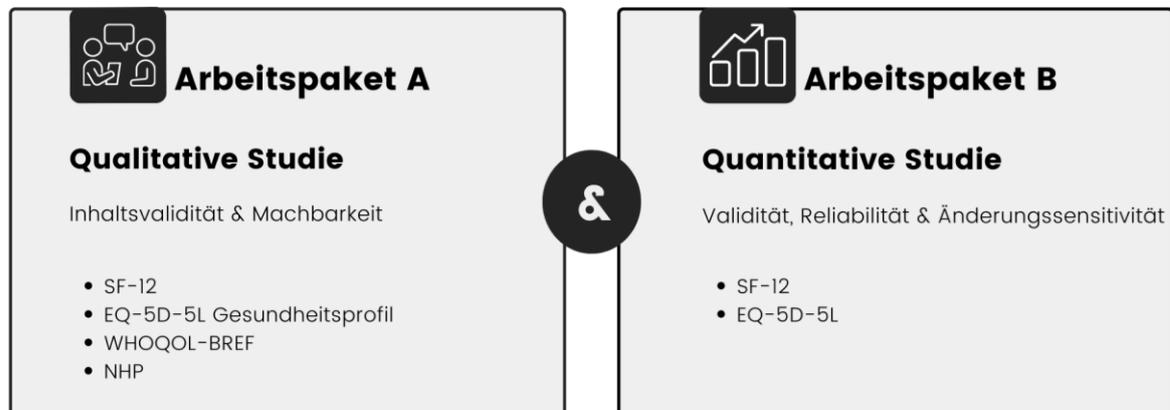


Abbildung 2: Arbeitspakete der Dissertation

2. Material und Methoden

2.1 Instrumente

Im Folgenden werden die wichtigsten Eigenschaften der in dieser Dissertation verwendeten Instrumente zur Messung der generischen Lebensqualität aufgeführt, welche in Tabelle 1 noch einmal vergleichend dargestellt werden.

2.1.1 Generisch - präferenzbasiert

EQ-5D-5L

Der EQ-5D ist ein präferenzbasiertes Maß für den Gesundheitszustand und wird weltweit in klinischen und bevölkerungsbezogenen Studien sowie im klinischen Alltag eingesetzt (40). Derzeit gibt es drei Versionen des EQ-5D: den EQ-5D-3L, den EQ-5D-5L und den EQ-5D-Y. Der EQ-5D-5L stellt eine Weiterentwicklung des EQ-5D-3L dar und soll durch fünf statt drei Antwortlevel die aufgetretenen Boden- bzw. Deckeneffekte vermeiden (41, 42). Der EQ-5D besteht unabhängig der Version aus zwei Teilen. Im deskriptiven Teil, dem EQ-5D Gesundheitsprofil, wird mit fünf Dimensionen die gesundheitsbezogene Lebensqualität erfragt. Im zweiten Teil, der EQ VAS, kann der momentane Gesundheitszustand auf einer visuellen Analogskala (VAS) angegeben werden (43). Es ist möglich, nur den deskriptiven Teil zu verwenden, die VAS liefert aber eine zusätzliche Perspektive der Befragten auf die subjektiv empfundene HRQOL (44).

Der deskriptive Teil bzw. das Gesundheitsprofil des EQ-5D bildet folgende fünf Dimensionen ab: Mobilität, Selbstversorgung, Allgemeine Tätigkeiten, Schmerz/Körperliche Beschwerden und Angst/Niedergeschlagenheit. Jede Dimension kann anhand fünf möglicher Antworten (keine Probleme, leichte Probleme, mäßige Probleme, starke Probleme, nicht möglich/extreme Probleme) beantwortet werden. Damit können 3125 (= 5^5) verschiedene Gesundheitsprofile in einer fünfstelligen Zahlenkombination erfasst werden. Außerdem können die Antworten mithilfe eines länderspezifischen Wertesets (45) zu einem präferenzbasiertem Indexwert aggregiert werden, der die gesellschaftliche Perspektive repräsentiert und über den außerdem die qualitätskorrigierten Lebensjahre (=QUALYs) [engl. quality-adjustes life years] berechnet werden können. Die Werte variieren dabei üblicherweise zwischen weniger als 0.0 (wobei 0 den Wert eines Gesundheitszustands bedeutet, der dem Tod

gleichkommt; negative Werte stehen für Werte, die schlechter als der Tod sind) und 1.0 (bester Gesundheitszustand) und spiegeln die Bewertung des Gesundheitszustandes basierend auf den Werten der Allgemeinbevölkerung wider. Für die quantitative Studie in dieser Arbeit wurden mithilfe des R-Pakets „eq5d“ (Version 0.7.2) der Indexwert berechnet (siehe R-Skript im Anhang S.140/141). Die EQ VAS erfasst den aktuellen allgemeinen Gesundheitszustand eines Befragten auf einer vertikalen visuellen Analogskala und bildet Werte von 0-100 ab, wobei 0 den schlechtmöglichen Gesundheitszustand widerspiegelt und 100 den bestmöglichen Gesundheitszustand.

Insgesamt besteht große Evidenz für die Erfüllung der Messeigenschaften des EQ-5D in verschiedenen Patientengruppen (39, 42, 46). Der EQ-5D zeichnet sich außerdem durch seine vielfältigen Assessmentoptionen aus, wie z.B. eigenständiges Ausfüllen, das Ausfüllen mithilfe eines Interviewers/einer Interviewerin oder sogar das Ausfüllen durch einen Proxy, z.B. einen Angehörigen oder eine Angehörige.

Für die Ermittlung der Inhaltsvalidität (Arbeitspaket A, qualitative Studie) wurde nur der deskriptive Teil verwendet; für die Ermittlung der weiteren Messeigenschaften (Arbeitspaket B, quantitative Studie) wurden beide Teile des Fragebogens herangezogen.

2.1.2 Generisch - nicht präferenzbasiert

SF-12

Der Short-Form-12 Fragebogen (SF-12 V2) wurde als Kurzversion des SF-36 entwickelt und erklärt über 90% der Varianz der SF-36 Summenskalen in der U.S.-Normstichprobe (47). International schreiben einige Arbeiten dem schnell auszufüllenden Fragebogen gute psychometrische Eigenschaften zu (48, 49).

Der Fragebogen umfasst zwölf Items, wobei sich die deutsche Version nur anhand eines Items zur U.S. Stichprobe unterscheidet (50). Die zwölf Items erfassen über acht verschiedene Dimensionen Aussagen über den Gesundheitszustand:

- Allgemeine Gesundheitswahrnehmung (1 Item)
- Physische Gesundheit (2 Items)
- Eingeschränkte physisch-bedingte Rollenfunktion (2 Items)
- Körperliche Schmerzen (1 Item)

- Vitalität (1 Item)
- Mentale Gesundheit (2 Items)
- Eingeschränkte emotional-bedingte Rollenfunktion (2 Items)
- Soziale Funktionsfähigkeit (1 Item)

Die Auswertung des Fragebogens resultiert in zwei Skalen, der körperlichen (=PCS-12) [engl. physical component score] und der psychischen Summenskala (=MCS-12) [engl. mental component score], auch Komponenten-Scores genannt. Sowohl Antwortkategorien, Polarität der Antworten als auch die Abfolge der Items aus den Subskalen sind unterschiedlich. Der Fragebogen kann sowohl als Interview geführt als auch als Self-Report ausgefüllt werden.

Der am Ende gebildete Score kann zwischen null und 100 variieren, wobei 100 die höchste subjektiv vorstellbare Lebensqualität abbildet. Die Berechnung der beiden Summenskalen erfolgt in vier Schritten und beinhaltet die Gewichtung von Dummy-Variablen anhand der amerikanischen oder deutschen Normstichprobe. In dieser Arbeit wird die U.S.-Populationsnorm zur Berechnung der Summenscores verwendet. Eine Substitution fehlender Werte ist nicht vorgesehen, so dass der Fragebogen nicht auswertbar ist, sobald eine Frage des SF-12 nicht beantwortet ist.

Nottingham Health Profile

Auch das Nottingham Health Profile (NHP) (51, 52) dient als Instrument zur Patient:innenbeurteilung der subjektiven Gesundheit und umfasst insgesamt 38 Items in sechs Dimensionen: Körperliche Mobilität, Schmerz, Schlafstörungen, soziale Isolation, emotionale Beeinträchtigungen und Energieverlust. Das NHP ist als vollstandardisiertes Verfahren ebenso zur Selbstbeurteilung wie zur Interviewdurchführung geeignet. Die Antwortoptionen beim NHP sind dichotom in „ja“ und „nein“ aufgeteilt. Die Auswertung bezieht sich auf die mit „ja“ beantworteten Items. Für den Gesamtscore werden den Items in jeder Domäne Gewichte zugeordnet und der Durchschnitt aus den Scores der sechs Domänen gebildet, wobei gegensätzlich zum SF-12 die null den besten subjektiven Gesundheitsstatus beschreibt und 100 den Schlechtesten. Dem englischen Original wurden in einer Reihe von Validierungsstudien die Erfüllung der Eigenschaften Reliabilität und Validität zugeschrieben (52, 53). Die deutsche Übersetzung zeigte gemischte Beurteilungen psychometrischer Messeigenschaften (51, 54).

WHOQOL-BREF

Die Kurzform des WHO [engl. World Health Organization] Lebensqualitätsfragebogens WHOQOL-100 (55) umfasst 26 Items, darunter ein allgemeines Lebensqualitäts-Item, ein allgemeines Gesundheits-Item und 24 spezifische Items - ein Item aus jeder der 24 LQ-Facetten des WHOQOL-100. Die 24 spezifischen Items bilden vier Domänen ab: Körperliche Gesundheit, psychische Gesundheit, soziale Beziehungen und Umwelt. Die Items werden auf einer Fünf-Punkte-Skala beantwortet, die die Intensität, Kapazität, Häufigkeit und Bewertung der HRQOL-Facetten in Bezug auf die letzten zwei Wochen bewertet. Negativ bewertete Items werden umgekehrt bewertet, so dass höhere Werte eine bessere Lebensqualität anzeigen. Der WHOQOL-BREF hat gute bis ausgezeichnete psychometrische Eigenschaften hinsichtlich Reliabilität und schneidet auch bei Validierungstests gut ab (55).

Der Durchschnittswert aus jeder Domäne wird verwendet, um den Gesamtscore zu berechnen. Die positiv gepolte Skala reicht von null bis 100, wobei höhere Werte eine bessere HRQOL abbilden. Der WHOQOL-BREF kann ebenso selbst ausgefüllt werden, Interviewer:innen gestützt oder durch einen Interviewer oder eine Interviewerin ausgefüllt werden.

2.2 Studiendesigns

Für die Bewertung unterschiedlicher Messeigenschaften sind verschiedene Studientypen notwendig. In dieser Arbeit wird eine qualitative Studie zur Erfassung der Inhaltsvalidität und Machbarkeit aus Patientenperspektive vorgestellt sowie die Analyse von Daten aus einer quantitativen Studie zur Bewertung weiterer Messeigenschaften.

2.3.1 Qualitative Studie

Design

Think-aloud Interviews (56, 57) gefolgt von einem semi-standardisierten Interview wurden verwendet, um die Inhaltsvalidität der Fragebögen zu bewerten. Think-aloud Interviews sind eine Methode aus dem Bereich der kognitiven Erhebungsverfahren und haben das Ziel, herauszufinden, wie die Zuhörer das ihnen präsentierte Material verstehen, mental erfassen und darauf reagieren (57). Bei think-aloud Interviews werden die Teilnehmenden darum gebeten, ihre Gedanken beim Ausführen einer Aufgabe laut zu formulieren.

Tabelle 1: Vergleich der verwendeten Instrumente zur Messung der gesundheitsbezogenen Lebensqualität

Instrument	Anzahl der Items	Scoring	Dimensionen	Antwortlevel	Verwendung Arbeitspaket	Entwicklungsjahr	Lizenzierung
Generisch präferenzbasiert							
EQ-5D-5L (Gesundheitsprofil/ deskriptives System) (43, 58)	5	<ul style="list-style-type: none"> • Gesundheitsprofil oder präferenzbasierter Indexwert zwischen 0.0 (schlechtmöglicher Gesundheitszustand) und 1.0 (beste Gesundheit) 	<ul style="list-style-type: none"> • Mobilität • Selbstversorgung • Aktivitäten des täglichen Lebens (ADL) • Schmerz/Unwohlsein • Angst/Depression 	5	A, B	2009	keine Kosten (universitär), Zustimmung der EuroQol-Gruppe
EQ VAS	1	<ul style="list-style-type: none"> • vertikale visuelle Analogskala mit Werten von 0-100 (bestmöglicher Gesundheitszustand) 		VAS	B		
Generisch- nicht präferenzbasiert							
Short-Form 12 Questionnaire (SF-12) (47, 59)	12	<ul style="list-style-type: none"> • Psychische und Körperliche Summenskala (MCS-12 und PCS-12) • Wert zwischen 0-100 (höhere Werte bedeuten bessere HRQOL) 	<ul style="list-style-type: none"> • Allgemeine Gesundheitswahrnehmung • Physische Gesundheit • Eingeschränkte physisch-bedingte Rollenfunktion • Körperliche Schmerzen • Vitalität • Mentale Gesundheit • Eingeschränkte emotional-bedingte Rollenfunktion – • Soziale Funktionsfähigkeit 	2-6	A, B	1998	kostenpflichtig (Lizenzgebühren), Hogrefe Verlag

WHOQOL-BREF (55, 60)	26	<ul style="list-style-type: none"> • 4 Domänenscores • 2 separate Items • Wert zwischen 0-100 (höhere Werte bedeuten bessere HRQOL) 	<ul style="list-style-type: none"> • allgemeine Gesundheit • allgemeine Lebensqualität • körperliche Gesundheit • psychische Gesundheit • soziale Beziehungen • Umwelt 	5	A	2000	keine Kosten, Zustimmung WHO
Nottingham Health Profile (NHP) (51, 61)	38	<ul style="list-style-type: none"> • ein Score pro Dimension • Wert zwischen 0-100 (<u>niedrigere</u> Werte bedeuten bessere HRQOL) 	<ul style="list-style-type: none"> • körperliche Mobilität • Schmerz • Schlafstörungen • soziale Isolation • emotionale Beeinträchtigungen • Energieverlust 	2	A	1981	für wissenschaftliche Studien frei erhältlich

WHOQOL-BREF= WHO [engl. World Health Organization] Lebensqualitätsfragebogen Kurzform

Die verbalen Informationen bieten einen Einblick in den Prozess der Aufgabenerledigung. Think-aloud oder kognitive Interviews wurden ursprünglich verwendet, um den Prozess der Aufgabenerfüllung zu erforschen, zunächst bei relativ komplexen, mehrstufigen Aufgaben, wie z.B. bei Schachspielen (57), aber zunehmend auch beim Verstehen des Ausfüllens von Fragebögen (62, 63). Die qualitative Studie in dieser Arbeit verwendete die "concurrent think-aloud (CTA)" Methode, bei der die Teilnehmerinnen und Teilnehmer ihre Gedanken während der Aufgabenerledigung verbalisieren müssen. Es zeigte sich, dass CTA Interviews mehr Informationen und Einblicke in Entscheidungsprozesse generieren als retrospektive Methoden (64).

Ethikvotum, Einverständniserklärung und Datenschutz

Die Promovendin legte die Studie im Vorfeld der Ethikkommission der Universität Regensburg zur Beratung vor (17-703-101). Der Hintergrund der Studie wurde allen Interviewkandidatinnen und Interviewkandidaten erklärt, und alle erhielten die Möglichkeit, Fragen zu stellen. Von allen Studienteilnehmenden wurde das schriftliche Einverständnis eingeholt. Die Teilnehmerinnen und Teilnehmer erhielten keine Incentives und nahmen freiwillig teil. Die Audiodateien wurden bis zur Fertigstellung der Transkripte passwortgeschützt auf einem gesicherten Laufwerk der Universität Regensburg gespeichert. Die Transkripte der Think-aloud Statements und der Interviews wurden de-identifiziert, mit einer nicht-sprechenden Identifikationsnummer (Pseudonym) versehen und ebenfalls passwortgeschützt auf einem Laufwerk der Universität Regensburg gespeichert. Zugriff auf die Transkripte hatte nur die Promovendin sowie -für den Zeitraum des Abgleichs- die wissenschaftliche Hilfskraft. Die Einwilligungserklärungen wurden stets getrennt davon in verschließbaren Schränken (für Dritte unzugänglich) aufbewahrt. Entsprechend der Empfehlung der Deutschen Forschungsgemeinschaft werden die Forschungsdaten (Transkripte) für 10 Jahre nach Beendigung der Studie aufbewahrt.

Teilnehmende/Rekrutierung

Die Daten wurden zwischen November 2017 und Januar 2019 erhoben. Eingeschlossen wurden ältere Hüftfrakturpatient:innen (proximale Femurfraktur und älter als 60 Jahre) ohne schwere kognitive Einschränkungen, d.h. Patientinnen und Patienten, die kognitiv in der Lage waren, an einem Interview teilzunehmen. Alle Interviewteilnehmenden waren in stationärer Behandlung in einem Krankenhaus in Regensburg (Caritas-Krankenhaus St. Josef Regensburg (Klinik für Unfallmedizin);

Speziell: Alterstraumatologie Station) und hatten im Schnitt sieben Tage zuvor (Spannbreite 3-23 Tage) eine Operation zur Behandlung der Hüftfraktur. Pflegekräfte vor Ort screenen die Station nach in Frage kommenden Patientinnen und Patienten. Der Einschluss in die Studie erfolgte im Krankenhaus durch die Promovendin selbst. Die Fälle wurden mit Hilfe einer gezielten Stichprobenziehung rekrutiert, um eine vielfältigere Stichprobe in Bezug auf Geschlecht und Alter zu erhalten. Ziel war es, Männer und Frauen unterschiedlichen Alters und aus unterschiedlichen Wohnsituationen (zu Hause allein, Pflegeheim, bei Familie) zu befragen. So wurden im Laufe der Stichprobenziehung gezielt Männer eingeschlossen, da in der Ausgangsstichprobe wesentlich mehr Frauen als Männer vertreten waren. Um gehaltvolle Aussagen zum Vergleich zwischen den Fragebögen machen zu können, wurde die Rekrutierung bis zur theoretischen Sättigung fortgeführt. Die Promovendin führte alle Interviews als Face-to-Face Interviews durch. Vor der Studie gab es keine Beziehung zwischen der Interviewerin und den Teilnehmenden. Nach der Studie erhielten die Teilnehmerinnen und Teilnehmer eine kleine Pralinschachtel mit dem Aufdruck „Kleines Dankeschön“. Den Teilnehmerinnen und Teilnehmern wurde mitgeteilt, dass die Studie im Rahmen des Promotionsvorhabens der Interviewerin durchgeführt wird.

Ablauf/Datenerhebung

Zunächst wurden die Teilnehmenden gebeten, vier verschiedene generische HRQOL-Fragebögen auszufüllen. Einige Teilnehmende konnten jedoch aufgrund ihres beeinträchtigten Gesundheitszustands nicht alle Fragebögen ausfüllen. Während des Ausfüllens wurden die Patientinnen und Patienten dazu ermutigt, laut zu denken (65) und es wurden verbal probes verwendet, um das Verständnis der Teilnehmenden für die Anweisungen, Items und Antwortskalen des Fragebogens zu ermitteln. Danach wurde die Perspektive der Hüftfrakturpatient:innen auf die Fragebögen in einem halbstrukturierten Interview untersucht. Der Themenleitfaden basierte auf einer Studie über die Ansichten von Asthmapatient:innen zu HRQOL-Instrumenten sowie auf Indikatoren zur Beschreibung der inhaltlichen Validität eines PROMs (66, 67).

Der Leitfaden deckte die Dimensionen Durchführbarkeit, inhaltliche Validität, Präferenzen und Verbesserungsvorschläge ab. Die Themen im Leitfaden sind in Tabelle 2 aufgeführt, der gesamte Leitfaden ist im Anhang zu finden.

Tabelle 2: Themen des Interviewleitfadens

Themenbereiche	
Genereller Eindruck	Redundanz
Länge	Belastung/Aufwand
Layout/Darstellung	Validität/Passung
Format der Fragen bzw. Antworten	Vorschläge für Verbesserungen
Verständlichkeit	Präferenz
Spezifikation der Fragestellung	Anmerkungen der Patienten

Nach der Pilotphase mit fünf Befragten wurden kleine Änderungen an der Formulierung einiger Fragen vorgenommen. Zusätzlich zur think-aloud Methode und zum Interview wurden Feldnotizen gemacht und Informationen zu allgemeinen Patientencharakteristika (z.B. Alter, Geschlecht, Lebenssituation) aus der Krankenakte gesammelt (siehe Anhang). Die Interviews wurden mitgeschnitten und wörtlich transkribiert. Je nach Vorliebe und Gesundheitszustand der Teilnehmerinnen und Teilnehmer wurden die Interviews in einem Gemeinschaftsraum auf der Station (im Rollstuhl oder normalen Stuhl sitzend) oder direkt im Bett liegend mit der Interviewerin neben sich sitzend durchgeführt. Die Schriftgröße der Fragebögen wurde so weit wie möglich vergrößert, ohne dabei das ursprüngliche Layout zu verändern. Jeder Fragebogen wurde auf verschiedenfarbigem Papier gedruckt (siehe Abbildung 3), so dass die Teilnehmenden den Fragebogen farblich zuordnen konnten.

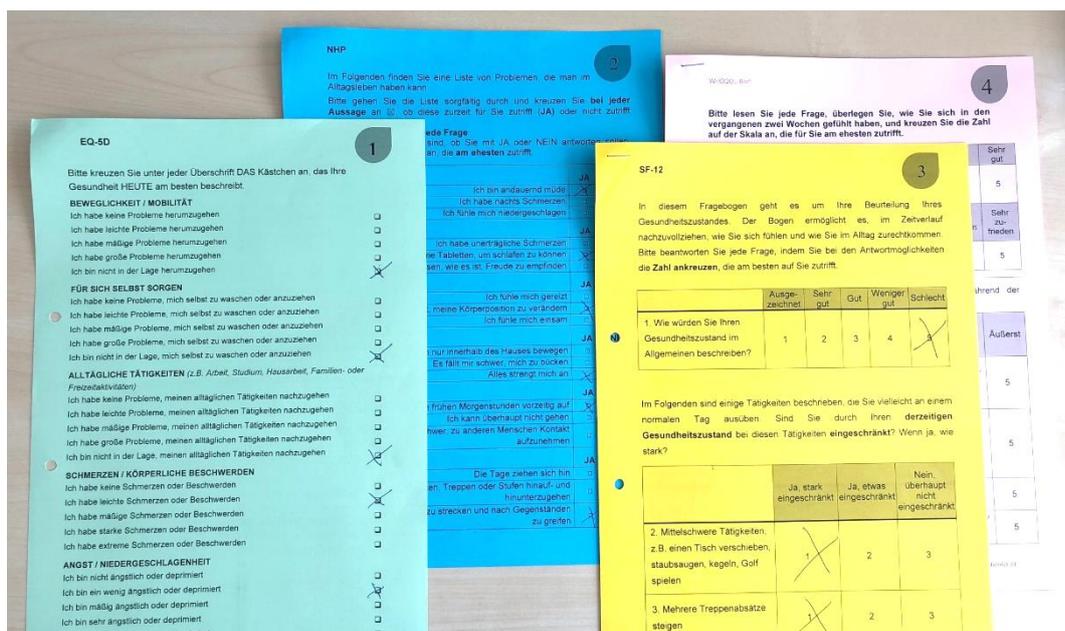


Abbildung 3: Foto der verwendeten Fragebögen

Die Fragebögen wurden in zufälliger Reihenfolge ausgegeben. Einige Teilnehmende baten jedoch darum, die Fragebögen zu sehen, bevor sie in die Studie einwilligten, und dann selbst zu entscheiden, welche zuerst oder welche überhaupt ausgefüllt werden sollten. Die Abschriften wurden den Teilnehmerinnen und Teilnehmern nicht zur Stellungnahme und/oder Korrektur zurückgegeben. Die Interviewerin führte zum ersten Mal selbstständig Interviews durch und ging, abgesehen von Informationen aus der Literatur, offen an den Forschungsprozess heran.

Datenanalyse

Die deskriptive Auswertung der demographischen Daten zur Charakterisierung der Studienstichprobe erfolgte mittels SPSS 22 (SPSS Inc., Chicago, USA). Die Tonaufnahmen wurden nach vereinfachten Regeln von Dresing und Pehl (68) hauptsächlich durch die Promovendin (22 von 30 Transkripten) und zum Teil durch einen professionellen Transkriptionsdienst transkribiert. In diesem Fall wurden die Interviews anschließend von der Promovendin kontrolliert und anonymisiert. Qualitative Daten (Transkripte) wurden mit der QDA-Software ATLAS.ti Version 8 analysiert. Die Daten wurden mit Hilfe der qualitativen Inhaltsanalyse nach Schreier (69) untersucht. Die Kategorien wurden hauptsächlich deduktiv aus dem Interviewleitfaden gebildet, weitere Aspekte wurden induktiv aus dem Datenmaterial abgeleitet. Um die Validität der Ergebnisse zu erhöhen, wurden die Transkripte von der Promovendin und einer wissenschaftlichen Hilfskraft unabhängig voneinander gelesen und kodiert. Unterschiede in der Kategorisierung wurden innerhalb der beiden Kodiererinnen und wenn nötig mit dem Betreuer der Promotionsarbeit diskutiert, bis ein Konsens erreicht wurde. Diskrepanzen in den Ergebnissen, widersprüchliche Aspekte und abweichende Aussagen wurden mit besonderer Aufmerksamkeit analysiert (70). Zudem wurde eine vergleichende Analyse zwischen den Altersgruppen (≥ 80 Jahre und < 80 Jahre) sowie zwischen Frauen und Männern durchgeführt.

2.3.2 Quantitative Studie

Datenbasis

Die Daten für die quantitative Studie stammen aus einer bevölkerungsbezogenen prospektiven Beobachtungsstudie namens ‚ProFem‘, bei der Personen mit einer proximalen Femurfraktur (PFF) zu Lebensqualität, Funktionsfähigkeit und sozialer Teilhabe befragt wurden (14). Die ProFem-Studie wurde mit Mitteln des

Innovationsfonds an der Heinrich-Heine-Universität Düsseldorf durchgeführt (Förderkennzeichen 01VSF16043), koordiniert vom Institut für Versorgungsforschung und Gesundheitsökonomie (Studienleitung: Prof. Dr. Dr. Andrea Icks, MBA). Studienteilnehmende waren Versicherte der AOK Rheinland/Hamburg (AOK RH) in einem Alter von ≥ 60 Jahren und mit Wohnsitz in Nordrhein-Westfalen, die mindestens 12 Monate vor der PFF durchgängig bei der AOK RH versichert waren und bei denen im Zeitraum Januar 2018 bis September 2019 eine PFF auftrat. Die Identifizierung der Versicherten erfolgte anhand der ICD-10 Codes S72.0, S72.1 und S72.2 sowie ausgewählter Operationen- und Prozedurenschlüssel (OPS).

Die Studie basierte auf GKV-Daten (Studienpopulation „Versicherte“) sowie auf bei einer Subgruppe von Versicherten erhobenen Primärdaten (Studienpopulation „Befragte“). Die Primärdatenerhebung umfasste einen Hausbesuch nach 3 (T0) und 12 Monaten (T2), eine postalische Befragung (T1) und bei Verhinderung zu T0 einen Ersatz-Hausbesuch (ET1) nach 6 Monaten (siehe auch Abbildung 4). Die Studienteilnahme umfasste zu jedem Zeitpunkt verschiedene Möglichkeiten, angepasst an die Verfassung der Studienteilnehmenden:

- Person mit PFF selbst (TN): normale Version oder Kurzversion
- Person mit PFF selbst und zusätzlich eine nahestehende Proxy-Person (TN/ZP): jeweils normale Version oder Kurzversion
- Nahestehende Proxy-Person stellvertretend für Person mit PFF (EP): normale Version

Die ProFem Studie verfolgte die nachfolgenden Ziele:

1. Leistungsgeschehen und
2. patientenberichtete Outcomes nach PFF zusammenhängend zu untersuchen
3. Subgruppen zu identifizieren, die durch besonders schlechte klinische und patientenberichtete Outcomes charakterisiert sind und potentiell eine intensivere Versorgung benötigen; Indikatoren zur Identifizierung dieser Subgruppen zu definieren; einen Algorithmus für die GKV im Sinne eines case findings zu entwickeln
4. Maßnahmen abzuleiten, um dadurch eine zielgruppengerechtere und damit auch wirtschaftlichere Gestaltung der Versorgung anzustreben

Je nach Zielsetzung bzw. Fragestellung wurden die GKV-Daten, Daten aus der Primärerhebung oder eine Verknüpfung der beiden Datenarten verwendet.

Für die psychometrischen Analysen in dieser Dissertation wurden nur Teile der erhobenen Primärdaten verwendet. Es wurden Studienteilnehmende berücksichtigt, die zu den beiden Zeitpunkten T0 (3 Monate nach Krankenhausentlassung) und T1 (6 Monate nach Krankenhausentlassung) eigenständig teilnahmen (TN und TN/ZP). Teilnehmende, für die ausschließlich stellvertretend durch eine nahestehende Proxy-Person Primärdaten vorlagen (EP), wurden dementsprechend für die hier durchgeführten Analysen nicht berücksichtigt.

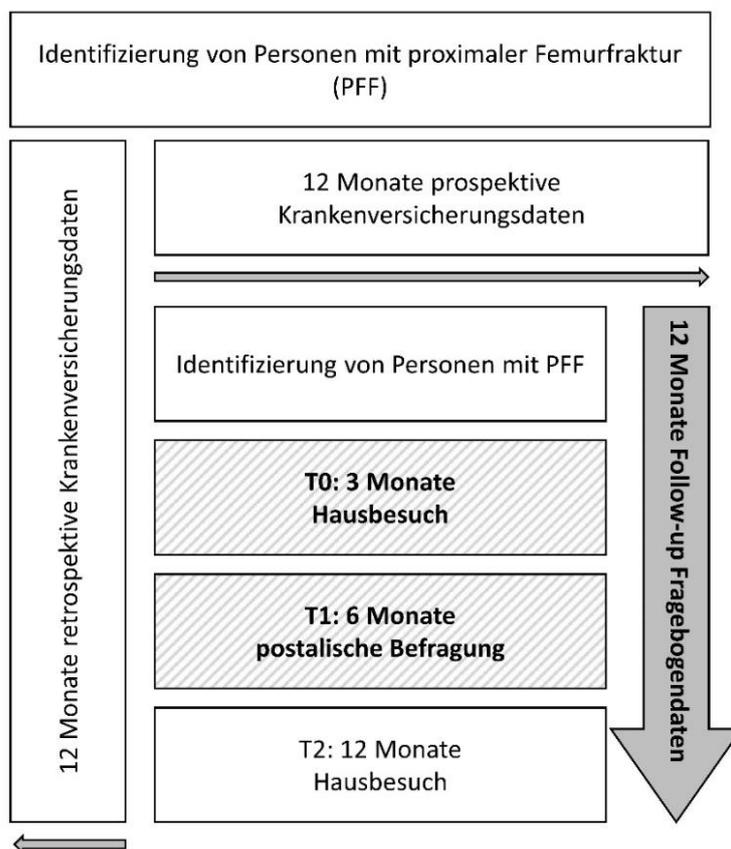


Abbildung 4: Darstellung des ProFem Studiendesigns aus Andrich (2019) (14)

(Daten für die psychometrische Analyse in dieser Studie sind in schräg schraffiert hervorgehoben).

Anhand der durch ProFem erhobenen Primärdaten wurden Messeigenschaften (Validität, Reliabilität, Änderungssensitivität) der beiden Fragebögen EQ-5D und SF-12 getestet. Dazu wurden analytische Prinzipien der klassischen Testtheorie (=CTT) [engl. classical test theory] angewendet. Verwendet wurden die in Tabelle 3 aufgeführten Variablen. Die Analysen wurden a priori in einem englischsprachigen

Analyseplan (siehe Anhang) festgelegt. In der Monografie werden die zu überprüfenden Messeigenschaften und die methodische Überprüfung dieser erläutert.

Im Folgenden werden die zur Überprüfung der Messeigenschaften des SF-12 und EQ-5D-5L Fragebogens verwendeten Referenzmaße kurz erläutert. Eine Beschreibung der zu untersuchenden Fragebögen (SF-12 und EQ-5D) findet sich in Abschnitt 2.1.1.

Tabelle 3: Erhobene Variablen und jeweilige Erhebungszeitpunkte

Konstrukt/ Kategorie	Instrument/Variable	3 Monate nach Krankenhausentlassung	6 Monate nach Krankenhausentlassung
Outcome			
Gesundheitsbezogene Lebensqualität (HRQOL)	EQ-5D-5L (41)	x	x
	Short-Form-12 Fragebogen (47)	x	x
Referenzmaße			
Aktivitäten des täglichen Lebens (ADL)	Barthel-Index (71)	x	x
Funktionsfähigkeit (hüftspezifisch)	Oxford Hip Score (72)	x	x
Soziale Teilhabe	IMET (Index for Measuring Participation Restrictions) (73)	x	x
Soziale Unterstützung	Oslo Social Support Scale (74)	x	
Veränderung selbst eingeschätzter Gesundheitszustand	Ankerfrage (Global-Rating of Change Scale)	-	x
Allgemeine Merkmale			
Soziodemographische Daten	Alter, Geschlecht	x	-
Sozioökonomischer Status	MacArthur Skala (75)	x	-
Klinische Daten			
Kognitive Einschränkung/ reduzierter Allgemeinzustand	Lang- oder Kurzfragebogen	x	x

Kurzbeschreibung der Referenzmaße

Der Barthel-Index (Version von Heuschmann et al. (71)) ist ein gebräuchliches Maß für Activities of Daily Living (ADL) und umfasst zehn Bereiche, die in Selbstversorgung und Mobilität unterteilt sind. Die Interviewversion (T0) umfasst 17 Fragen, die postalische Erhebung (T1) zehn Fragen (76). Die Gesamtpunktzahl reicht von 0

(niedrigste ADL) bis 100 (vollständige Unabhängigkeit). Die Verständlichkeit und Zuverlässigkeit beider Fragebögen wurde unter anderem bei Patientinnen und Patienten mit Schlaganfall getestet (71).

Der Oxford Hip Score (OHS) ist ein hüftspezifischer Fragebogen zur Beurteilung der Funktionsfähigkeit. Er enthält sechs Fragen zu Schmerzen und sechs Fragen zur Funktionsfähigkeit und bezieht sich auf die letzten vier Wochen. Jedes Item hat fünf Antwortstufen, die von null (schlechtestes Ergebnis) bis vier (bestes Ergebnis) reichen. Es kann ein Summenwert aller Werte angegeben werden, der von null (schlechtestes Ergebnis) bis 48 (bestes Ergebnis) reicht. Die psychometrischen Eigenschaften (Reliabilität und Validität) des OHS wurden in einer deutschen Population von Patient:innen mit Hüftarthrose untersucht (77).

Mit Hilfe des Index zur Messung von Einschränkungen der Teilhabe (IMET) kann die soziale Teilhabe von Menschen mit verschiedenen chronischen Krankheiten (73) auf der Grundlage der Internationalen Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF) beurteilt werden. Der Fragebogen enthält neun Items, die neun Aspekte der Alltagsaktivitäten und der Partizipation abdecken. Die Beeinträchtigungen werden mit Hilfe einer elfstufigen Skala bewertet, wobei 0 für "keine Beeinträchtigung" und 10 für "keine Aktivität möglich" steht. Es kann eine zusammenfassende Punktzahl berechnet werden, wobei Werte von 0-100 (niedrigere Werte = weniger Beeinträchtigung in der sozialen Teilhabe) vorkommen können. Die psychometrischen Eigenschaften wurden für verschiedene Patientengruppen getestet (73), und die Normwerte ermöglichen eine Klassifizierung der Teilnahmebeschränkungen für Menschen, die sich in Rehabilitation befinden oder an chronischen Krankheiten leiden (78).

Die Oslo Social Support Skala (OSS-3) wird verwendet, um die soziale Unterstützung (79) über drei Items zu erfassen. Die Items können von 1 bis 4 oder 5 bewertet werden. Somit kann eine Gesamtpunktzahl zwischen drei und 14 erreicht werden, wobei höhere Werte eine höhere soziale Unterstützung anzeigen.

Speziell für die Erfassung von psychometrischen Messeigenschaften war eine sogenannte globale Ankerfrage in den Follow-up Erhebungsbögen enthalten, mit der die Veränderung des subjektiven Gesundheitszustandes im Vergleich zur letzten Befragung erfragt wurde. Dabei konnten die Teilnehmenden auf einer Skala von -5 (stark verschlechtert) über 0 (unverändert) bis +5 (stark verbessert) antworten.

Die MacArthur-Skala bewertet den subjektiven sozialen Status (SSS), indem sie die Teilnehmenden bittet, sich entsprechend ihrer wahrgenommenen sozialen Stellung in der Gesellschaft auf einer Leiter einzuordnen (80). Chen und Kollegen (81) schlugen Grenzwerte vor, um Gruppen zu identifizieren, die im Vergleich zu anderen eindeutig schlechter oder besser gestellt sind. Die Antworten wurden entsprechend in folgende drei Gruppen unterteilt:

- niedriger SSS: Sprossen 1-3
- mittlerer SSS: Sprossen 4-7
- hoher SSS: Sprossen 8-10

Psychometrische Analyse

Für Gütekriterien der Zustandsmessung wurde der 3-Monats-Fragebogen verwendet. Zur Bestimmung von Gütekriterien der Veränderungsmessung wurden die Daten aus T0 und T1 bzw. deren Schnittmenge verwendet. Alle Analysen basierten auf den COSMIN Empfehlungen (36). Eine Übersicht der zu untersuchenden Messeigenschaften und die jeweils herangezogenen Grenzwerte zur Beurteilung finden sich in Tabelle 4. Um feststehende Begrifflichkeiten eindeutig zu verwenden, werden im Folgenden die englischen Bezeichnungen bei Erstnennung der Messeigenschaften mit angegeben.

Validität

Generell sind drei verschiedene Arten der Validität zu unterscheiden: Inhaltsvalidität [engl. content validity], Kriteriumsvalidität [engl. criterion validity] und Konstruktvalidität [engl. construct validity]. Die von den Patientinnen und Patienten berichteten Ergebnismaße (PROMs) können nicht mit einem Goldstandard zur Ermittlung der Kriteriumsvalidität verglichen werden. Ein Goldstandard wird als repräsentativ für den wahren Zustand angesehen und ist in der Medizin z.B. ein Krankheitsstatus. PROMs konzentrieren sich häufig auf subjektive Wahrnehmungen und Ansichten, so dass ein Goldstandard meist nicht verfügbar ist. In diesen Fällen wird üblicherweise die Konstruktvalidität untersucht. Um die Konstruktvalidität zu beurteilen, wurden die strukturelle Validität [engl. structural validity] (auch Dimensionalität) sowie die Konvergenzvalidität [engl. convergent validity], Diskriminanzvalidität [engl. discriminant validity] (auch diskriminante/diskriminative oder differenzielle Validität genannt) und known-groups-Validität [engl. known-groups or discriminative validity] untersucht.

Die strukturelle Validität der beiden interessierenden Maße wurde durch eine konfirmatorische Faktoranalyse (=CFA) [engl. confirmatory factor analysis] überprüft. Diese Analyse basiert darauf, dass für die empirische Erfassung eines hypothetischen bzw. latenten Konstruktes wie der HRQOL eine Operationalisierung über ein geeignetes Messmodell erfolgen muss. Die CFA ist eine Art der Strukturgleichungsmodellierung (SGM, engl. structural equation modeling, kurz SEM), die sich speziell mit Messmodellen befasst, d.h. mit den Zusammenhängen zwischen beobachteten Messwerten und latenten Variablen oder Faktoren (82). Für die beiden hier untersuchten Fragebögen wurden die bereits bei der Entwicklung der Fragebögen etablierten Faktorenstrukturen verifiziert: 2-Faktor Struktur des SF-12 (PCS-12 und MCS-12) und die eindimensionale Struktur (1-Faktor Struktur) des EQ-5D. Anhand der CFA wurde geprüft, ob die hier erhobenen Daten zur theoretisch erwarteten Struktur passen. Mittels verschiedener Kennwerte konnten die erwarteten Zusammenhänge bzw. die Modellanpassung beurteilt werden: Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR) und normierter Chi-Quadrat-Wert (χ^2/df). Entsprechend den Empfehlungen von Gädde und Kollegen (83) konnte bei Werten von $CFI \geq 0.95$, $RMSEA < 0.08$, $SRMR < 0.10$ und $\chi^2/df < 5$ (84) ein akzeptabler Modellfit angenommen werden.

Zur Bewertung der weiteren Aspekte der Konstruktvalidität wurden mehrere Hypothesen getestet. Dazu wurden Korrelationen der beiden interessierenden Maße (SF-12 und EQ-5D) mit Referenzmaßen (siehe Tabelle 3) untersucht und auf der Grundlage von a priori-Hypothesen beurteilt. Eine starke Korrelation mit einem Maß eines verwandten Konstrukts sprach für Konvergenzvalidität, während schwache Korrelationen mit nicht verwandten Konstrukten Evidenz für Diskriminanzvalidität lieferten. A priori-Hypothesen wurden auf Grundlage wissenschaftlicher Erfahrung und der inhaltlichen Konsistenz der Fragebögen gemeinsam von der Promovendin, dem Doktorvater und einer erfahrenen promovierten wissenschaftlichen Mitarbeiterin definiert. Mit Hilfe der Spearmans-Rang-Korrelationskoeffizienten (r_s) wurde die Stärke der Korrelation in drei Kategorien eingeteilt (85):

- schwach (+): $r_s < 0.3$
- moderat (++) : $r_s = 0.3-0.7$
- stark (+++) : $r_s > 0.7$

Tabelle 4: Übersicht über die quantitativen Analysen und jeweiligen Grenzwerte zur Bewertung der jeweiligen Messeigenschaft

	SF-12		EQ-5D-5L		
	PCS-12	MCS-12	EQ-5D deskriptives System	EQ-5D Index	EQ VAS
Validität					
Konvergenz-/ Diskriminanz- validität	Rangkorrelationen nach Spearman Korrelation <ul style="list-style-type: none"> • schwach: $r < 0.3$ • moderat: $r = 0.3-0.7$ • stark: $r > 0.7$ 	Rangkorrelationen nach Spearman Korrelation <ul style="list-style-type: none"> • schwach: $r < 0.3$ • moderat: $r = 0.3-0.7$ • stark: $r > 0.7$ 	Rangkorrelationen nach Spearman Korrelation <ul style="list-style-type: none"> • schwach: $r < 0.3$ • moderat: $r = 0.3-0.7$ • stark: $r > 0.7$ 	Rangkorrelationen nach Spearman Korrelation <ul style="list-style-type: none"> • schwach: $r < 0.3$ • moderat: $r = 0.3-0.7$ • stark: $r > 0.7$ 	Rangkorrelationen nach Spearman Korrelation <ul style="list-style-type: none"> • schwach: $r < 0.3$ • moderat: $r = 0.3-0.7$ • stark: $r > 0.7$
Known-groups Validität	Mann-Whitney-U-Test/ Kruskal-Wallis-Test $p < 0.05$	Mann-Whitney-U-Test/ Kruskal-Wallis-Test $p < 0.05$	/	Mann-Whitney-U-Test/ Kruskal-Wallis-Test $p < 0.05$	Mann-Whitney-U-Test/ Kruskal-Wallis-Test $p < 0.05$
Strukturelle Validität	CFA <ul style="list-style-type: none"> • CFI ≥ 0.95 • RMSEA < 0.08 • SRMR < 0.10 • $\chi^2/df < 5$ 	CFA <ul style="list-style-type: none"> • CFI ≥ 0.95 • RMSEA < 0.08 • SRMR < 0.10 • $\chi^2/df < 5$ 	CFA <ul style="list-style-type: none"> • CFI ≥ 0.95 • RMSEA < 0.08 • SRMR < 0.10 • $\chi^2/df < 5$ 	/	/

Reliabilität					
Interne Konsistenz	Cronbachs a	Cronbachs a	/	/	/
	≥ 0.7	≥ 0.7			
Änderungssensitivität					
Kriteriums-basiert	Wilcoxon-Vorzeichen-Rang-Test	Wilcoxon-Vorzeichen-Rang-Test	/	Wilcoxon-Vorzeichen-Rang-Test	Wilcoxon-Vorzeichen-Rang-Test
	p < 0.05	p < 0.05		p < 0.05	p < 0.05
Konstrukt-basiert	Rangkorrelationen nach Spearman	Rangkorrelationen nach Spearman	/	Rangkorrelationen nach Spearman	Rangkorrelationen nach Spearman
	Korrelation	Korrelation		Korrelation	Korrelation
	<ul style="list-style-type: none"> • schwach: r < 0.3 • moderat: r = 0.3-0.7 • stark: r > 0.7 	<ul style="list-style-type: none"> • schwach: r < 0.3 • moderat: r = 0.3-0.7 • stark: r > 0.7 		<ul style="list-style-type: none"> • schwach: r < 0.3 • moderat: r = 0.3-0.7 • stark: r > 0.7 	<ul style="list-style-type: none"> • schwach: r < 0.3 • moderat: r = 0.3-0.7 • stark: r > 0.7

CFA= Confirmatory factor analysis; CFI= Comparative Fit Index; GRCS= global rating of change scale; ICC= Intraklassenkorrelation; RMSEA= Root Mean Square Error of Approximation; ROC = Receiver operating characteristic; SRMR= Standardized Root Mean Square Residual; VAS= visuelle Analogskala

Korrelationen mit Instrumenten, die ähnliche Konstrukte messen, sollten stark (+++) sein. Korrelationen mit Instrumenten, die verwandte, aber unähnliche Konstrukte messen, sollten niedriger sein (++) , und Korrelationen mit Instrumenten, die nicht verwandte Konstrukte messen, sollten schwach sein (+). Tabelle 5 zeigt die vordefinierten a priori-Hypothesen zu den Korrelationen zwischen den jeweiligen Skalen/Items. Die Konstruktvalidität wurde als gut eingestuft, wenn mindestens 75% der vorher festgelegten Hypothesen bestätigt werden können (86).

Die Methode der known-groups basierte auf dem Prinzip, dass der Median der Messung innerhalb einer bestimmten Gruppe (z.B. Frauen) voraussichtlich von dem Median einer anderen Gruppe (z.B. Männer) abweicht.

Tabelle 5: Hypothesen zur Ermittlung der Diskriminanz- und Konvergenzvalidität

Dimensionen und Index	SF-12		OHS	Barthel- Index	OSS- 3	IMET
	MCS-12	PCS-12				
MCS-12	/	++	++	+	++	+
PCS-12	++	/	+++	+++	+	++
EQ-5D: Beweglichkeit/Mobilität	+	++	++	++	+	++
EQ-5D: Selbstversorgung	+	++	++	++	+	++
EQ-5D: Allgemeine Tätigkeiten	++	++	++	+++	+	++
EQ-5D: Schmerzen/körperliche Beschwerden	+	++	+++	++	+	++
EQ-5D: Angst/Niedergeschlagenheit	+++	+	+	+	++	+
EQ-5D-5L index	++	++	++	++	+	++
EQ VAS	++	++	++	++	+	++

IMET= Index zur Messung von Einschränkungen der Teilhabe; MCS-12= psychische Summenskala [engl. mental component score]; OHS= Oxford Hip Scale; OSS-3= Oslo Social Support Skala, PCS-12= körperliche Summenskala [engl. physical component score]; SF-12= Short-Form 12 Fragebogen; VAS= visuelle Analogskala

Ein valides Instrument sollte auf diese Unterschiede reagieren und in der Lage sein, sie in der vorhergesagten Richtung anzuzeigen. In diesem Fall wurden die Studienteilnehmenden nach Alter (<80 vs. ≥80), Geschlecht (männlich vs. weiblich),

selbstberichtetem sozioökonomischem Status und der Verwendung der Kurz- oder Langversion des ProFem-Fragebogens gruppiert bzw. dichotomisiert. In beiden Fragebögen zur Lebensqualität sollten erwartungsgemäß ältere Patient:innen, weibliche Patientinnen, Patient:innen mit einem niedrigeren selbstberichteten sozioökonomischen Status und Patient:innen mit einem höheren Einschränkungsgrad (Verwendung der Kurzversion des Fragebogens) eine geringere Lebensqualität angeben. Die Unterschiede zwischen den Gruppen wurden bei zwei Gruppen mit Hilfe des Mann-Whitney-U-Test (auch "Wilcoxon Rangsummen-Test" genannt) und bei mehr als 2 Gruppen mit dem Kruskal-Wallis-Test beurteilt.

Reliabilität

Die Beurteilung der Reliabilität wurde über das Maß der internen Konsistenz in Form von Cronbachs Alpha (α) berechnet. Ein Wert von > 0.7 wird in der Literatur häufig als Schwellenwert für die Annahme einer konsistenten Skala verwendet (36). Cronbachs α wurde für die beiden Subskalen des SF-12 berechnet. Für den EQ-5D wurde die Berechnung des Cronbachs α nicht als sinnvoll angesehen, da jede Dimension nur durch ein Item repräsentiert wird.

Interpretierbarkeit

Die Interpretierbarkeit ist streng genommen keine psychometrische Messeigenschaft, da sie sich nicht auf die Qualität eines Instruments bezieht, sondern darauf, was die Werte oder deren Veränderung bedeuten. Sie ist jedoch eine wichtige Eigenschaft und wurde daher in die COSMIN-Taxonomie und somit auch in diese Dissertation aufgenommen. Die Interpretierbarkeit wurde durch die Berechnung der Anzahl fehlender Items und deren Verteilung sowie durch die Bewertung von Boden- und Deckeneffekten beurteilt. Im Kontext dieser Studie wurde ein Bodeneffekt als die niedrigste mögliche Antwortkategorie verstanden, die ein Befragter angeben konnte. Deckeneffekte waren analog dazu die höchsten Antwortkategorien bei Likert-skalierten Items. Häufig wird die Interpretierbarkeit auch unter Begriffen wie Praktikabilität oder Machbarkeit untersucht. Zugrunde liegt immer die Hypothese, dass je größer der Anteil an fehlenden Werten in einem Fragebogen ist, desto fraglicher ist es, inwieweit das Item bzw. der Fragebogen verständlich war bzw. eine Antwortkategorie passend war. Dafür wurden für die beiden HRQOL Instrumente Verteilungsparameter (z.B. Mittelwert und Standardabweichung) berichtet.

Änderungssensitivität (Gütekriterium der Veränderungsmessung)

In der Literatur finden sich unterschiedliche Vorschläge und Vorgehensweisen zur Bestimmung der Änderungssensitivität und die Empfehlung, mehrere Kennwerte zu ermitteln, so dass es zu einer größeren Aussagekraft kommt. In dieser Arbeit wurde anhand eines ankerbasierten bzw. Kriteriums-basierten und eines Konstrukt-Ansatzes Evidenz für die Änderungssensitivität generiert.

Kriteriums-basierter Ansatz

Es wurden auf der Summenskalenebene einmal im Gesamtvergleich zwischen den beiden Zeitpunkten und einmal je nach Kategorie in der GRCS verschiedene Maße berechnet. Dazu wurden zunächst anhand der Antwortmöglichkeiten der GRCS fünf Kategorien gebildet:

- stark verbessert: 5, 4
- leicht verbessert: 3,2,1
- keine Veränderung: 0
- leicht verschlechtert: -1, -2, -3
- stark verschlechtert: -4, -5

Um Unterschiede zwischen den jeweiligen Scores der HRQOL Instrumente für die beiden Erhebungszeitpunkte zu ermitteln, wurde der Wilcoxon-Vorzeichen-Rang-Test verwendet. Wenn sich die Scores mit einem p-Wert von <0.05 in die erwartete Richtung unterschieden, wurde dies als Zeichen für die Empfindlichkeit des untersuchten Instruments gegenüber Veränderungen gewertet.

Konstrukt-Ansatz

Der Konstrukt-Ansatz orientierte sich an dem Vorschlag analog zur Bestimmung der Validität vorzugehen, da die Änderungssensitivität als Aspekt der Validität in einem längsschnittlichen Kontext angesehen werden kann (36, 87). Dementsprechend wurden innerhalb des Konstrukt-Ansatzes verschiedene a priori-Hypothesen überprüft (siehe Tabelle 17). Diese betrafen die Korrelation (Spearman Rangkorrelationskoeffizienten) zwischen der Veränderung in den Lebensqualitätsinstrumenten zwischen der ersten und zweiten Erhebung (im Folgenden ‚change score‘ genannt) und den Veränderungen in Referenzmaßen. Es wurde erwartet, dass auch die change scores ähnlicher Konstrukte in die gleiche Richtung eine Veränderung zeigen.

Die Änderungssensitivität wurde folgendermaßen bewertet (87, 88):

- hoch: weniger als 25% der Hypothesen müssen verworfen werden
- moderat: zwischen 25-50% der Hypothesen müssen verworfen werden
- gering: mehr als 50% der Hypothesen müssen verworfen werden.

Statistik

Die statistischen Analysen wurden mit R-Studio (R-Studio Team, Version 1.3.1073) durchgeführt. Kontinuierliche Daten wurden deskriptiv zusammengefasst, z.B. Mittelwert (M) und Standardabweichung (SD) [eng. standard deviation] oder Median (Md) und Interquartilsabstand (IQR) [engl. interquartile range]. Kategoriale Variablen wurden nach absoluten und relativen Häufigkeiten (Prozentsätze) der Patientinnen und Patienten in den jeweiligen Kategorien zusammengefasst. Fehlende Werte wurden nicht imputiert, sondern es wurde eine vollständige Fallanalyse durchgeführt.

Spearman-Rang-Korrelationskoeffizienten wurden verwendet, um die Korrelationen zwischen den durch die Instrumente gemessenen Messwerte zu ermitteln. Zur Ermittlung von Gruppenunterschieden (Unterschieden in der zentralen Tendenz) wurde bei zwei Gruppen der Mann-Whitney-U-Test (auch Wilcoxon-Mann-Whitney-Test oder Wilcoxon-Rangsummentest genannt) verwendet, bei mehr als zwei Gruppen der Kruskal-Wallis Test. Für alle Tests wurde eine statistische Signifikanz bei p-Werten $< 0,05$ angenommen. Eine Übersicht über die durchgeführten Analysen findet sich in Tabelle 4.

3. Ergebnisse

3.1 Qualitative Studie

Im Folgenden werden die Ergebnisse auf der Grundlage der think-aloud Phase und des semi-strukturierten Abschnitts der Interviews gemeinsam berichtet.

3.1.1 Beschreibung der Studienteilnehmenden

Insgesamt nahmen 32 Hüftfrakturpatient:innen (78% weiblich, Durchschnittsalter 81,3 Jahre) an dieser Studie teil, wobei die Zustimmungsrate 53% betrug. Von diesen 32 Interviews mussten zwei abgebrochen werden (wegen schlechter körperlicher oder emotionaler Verfassung) und konnten nicht in die qualitative Analyse einbezogen werden. Nicht alle Patientinnen und Patienten waren in der Lage, alle Fragebögen auszufüllen. Im Median betrug die Interviewdauer 35 Minuten, schwankte aber stark zwischen zwölf und 70 Minuten, unter anderem aufgrund der Anzahl an ausgefüllten Fragebögen. Im Median füllten Teilnehmende drei Fragebögen aus, wobei nur sechs Teilnehmende alle vier Fragebögen geschafft haben. Insgesamt wurde der SF-12 25-mal ausgefüllt, der EQ-5D 28-mal, das NHP 19-mal und der WHOQOL-BREF zwölf Mal. Die Mehrzahl der Teilnehmenden wohnte vor dem Frakturereignis zu Hause und es war kein Pflegegrad vermerkt. Die Gehfähigkeit vor der Fraktur gliederte sich in „selbstständig“ (48%), „Rollator-mobil“ (38%) und „mit Hilfe eines Gehstocks“ (14%).

3.1.2 Vergleichende Analyse der Geschlechter und Altersgruppen

Der vergleichenden Altersanalyse lag die Annahme zu Grunde, dass möglicherweise die Machbarkeit der Fragebögen bei hochaltrigen (≥ 80 Jahre) Teilnehmerinnen und Teilnehmern stärker eingeschränkt ist als bei den jüngeren Teilnehmerinnen und Teilnehmern.

Es zeigte sich jedoch kein konstanter Anstieg von Machbarkeitsproblemen mit dem Alter, so dass sich diese Hypothese nicht bestätigen lies. In beiden Altersgruppen gab es sowohl Teilnehmerinnen und Teilnehmer, die die Fragebögen selbstständig ausfüllen konnten als auch Teilnehmerinnen und Teilnehmer, die Hilfe benötigten. Auch bei den Aspekten zum Thema Inhaltsvalidität gab es keine Unterschiede, die direkt auf das Alter zurückzuführen waren. Gleiches galt für Geschlechterunterschiede. Weder in der Machbarkeit noch bei Aspekten der Inhaltsvalidität konnte ein Unterschied zwischen den interviewten Männern und Frauen festgestellt werden. So

waren die Unterschiede in der Machbarkeit und die Bewertung der Inhaltsvalidität vielmehr an einem Zusammenspiel aus Faktoren wie zum Beispiel dem allgemeinen Gesundheitszustand, dem kognitiven Level, den motorischen Fähigkeiten und dem Sehvermögen auszumachen.

Da diese vergleichenden Auswertungen keinen Unterschied zwischen den Geschlechtern und Altersgruppen zeigten, werden dementsprechend im Folgenden die Daten für die gesamte Stichprobe dargestellt.

3.1.3 Aspekte der Machbarkeit und generelle Eindrücke

Eine zusammenfassende bzw. vergleichende Darstellung der Machbarkeitsaspekte ist in Tabelle 6 zu finden. Insgesamt konnten neun Teilnehmende die Fragebögen selbstständig ausfüllen, 13 Teilnehmende benötigten Hilfe und die restlichen Teilnehmenden waren aufgrund funktioneller Probleme oder stark eingeschränktem Sehvermögen nicht in der Lage, Fragebögen auszufüllen. In diesem Fall wurden die Fragebögen durch die Interviewerin erhoben.

ID11 (w, 78): „Ich glaube Sie lesen es mir lieber vor, weil ich einfach nicht so gut sehen kann.“

Bei der Länge des Fragebogens ließen sich die Patientinnen und Patienten stark von der Seitenanzahl bzw. dem Layout beeinflussen, sodass nur zwölf Teilnehmende bereit waren, den WHOQOL-BREF auszufüllen. Dieser benötigte zwei Doppelseiten, umfasste aber nicht die meisten Items.

Bezüglich des Aspektes Anstrengung ergaben sich zwei unterschiedliche Empfindungen. Ein Teil der Patientinnen und Patienten sah das Ausfüllen eher als Zeitvertreib an und empfand es als nicht anstrengend.

I: „Und wie anstrengend war das für Sie den auszufüllen?“

ID09 (w, 84): „Nein (lachend), da muss es schon schlimmer kommen.“

Einige Teilnehmende empfanden es jedoch als anstrengend, zwei der Interviews mussten sogar abgebrochen werden.

ID16 (w, 74): „Jetzt werde ich müde, da surrt es da drin [im Kopf].“

Ein klares Layout trug zur Vollständigkeit und Richtigkeit der Fragebögen bei. So implizierte die Anordnung in Dreierblöcken bei dem NHP Fragebogen fälschlicherweise, dass nur eine Antwort pro Block benötigt werde. Außerdem wurden

die drei Items zum Teil als zusammenhängende Items verstanden, wodurch teilweise ein oder zwei Kreuze pro Block vergessen wurden. So wäre der Fragebogen letztendlich nicht auswertbar. Beim EQ-5D führte die Entfernung von Kästchen und Antwort zu einem höheren Risiko, in der Zeile zu verrutschen.

ID03 (w, 77): „Naja ich zähle hier [Fragen EQ-5D] und dann kann ich da [Antworten] auch runter zählen.“

Einige Teilnehmende waren zusätzlich verunsichert, wie sie beim SF-12 und WHOQOL-BREF ihre Antwort verdeutlichen sollen, etwa durch einen Kringel oder ein Kreuz.

Die Akzeptanz war insgesamt hoch (nur neun von 60 angesprochenen Personen lehnten die Teilnahme ab, z.B. aufgrund von datenschutzrechtlichen Bedenken oder keinem Interesse an der Studie bzw. an Studien insgesamt). Die restlichen, nicht interviewten Personen, waren teilweise zu krank (z.B. Sauerstoffgerät) und das Sprechen war zu anstrengend oder aber kognitiv zu stark eingeschränkt (Beurteilung durch kurzes einführendes Gespräch zu Beginn), so dass ein Interview nicht möglich gewesen ist. Die meisten Patientinnen und Patienten hatten eine neutrale Einstellung gegenüber Fragebögen.

I: „Was sagen Sie denn jetzt zu dem ersten Fragebogen?“

ID07 (w, 94): „Ja. Was Solls.“

Demnach beschrieben die Teilnehmenden die Fragebögen häufig als „normal“. Vereinzelt waren Patientinnen und Patienten v.a. zu Beginn skeptisch und hegten eine Ablehnung gegen „Deppenfragen“ (ID09, w, 84), worunter sie Fragen zur geistigen Fitness verstanden. Die anfängliche Skepsis war allerdings bei fast allen Teilnehmenden nach dem ersten Fragebogen überwunden.

Tabelle 6: Ergebnisse zu Aspekten der Machbarkeit im Vergleich der untersuchten Fragebögen

Instrument	Akzeptanz	Anstrengung	Länge	Layout
EQ-5D Gesundheits- profil	↑ • gute Akzeptanz, mit der Ausnahme von einem Teilnehmenden (aufgrund fehlender Relevanz)	↑ • nicht anstrengend	↑ • gut, da sehr kurz (ein paar empfanden es als zu kurz, aber aufgrund mangelnder inhaltlicher Vollständigkeit)	↗ • überwiegend gut, aber manchmal Probleme aufgrund fehlender Führungslinien
SF-12	↗ • neutrale Einstellung („normale Fragen“)	↗ • überwiegend nicht anstrengend	↗ • überwiegend nicht zu lang, bis auf zwei Teilnehmende	↗ • leicht verständlich, selten kleinere Probleme
NHP	⇒ • okay, aber viele Fragen ausgelassen	↗ • machbar	↗ • überwiegend okay, von wenigen als zu lang empfunden	⇒ • Anordnung in Dreierblöcken war oftmals verwirrend (fehlende Antworten)
WHOQOL- BREF	↘ • anfangs nicht (eher eingeschüchtert durch die Länge des Fragebogens), aber gut nach dem Ausfüllen	↘ • eher anstrengend (sehr viel zu lesen auch inkl. der Antwortkategorien), manchmal Pause dazwischen	⇒ • für einige "sehr lange" oder "zu lange", für andere in Ordnung	↗ • eigentlich klar, nur einzelne verstehen nicht wo oder wie man das Kreuz setzen muss

NHP= Nottingham Health Profile; SF-12= Short-Form-12 Fragebogen; WHOQOL-BREF= WHO [engl. World Health Organization] Lebensqualitätsfragebogen Kurzform

3.1.4 Aspekte der Inhaltsvalidität

Die Teilnehmerinnen und Teilnehmer sprachen viele verschiedene Aspekte der Inhaltsvalidität an, insbesondere während des Ausfüllens der Fragebögen in der think-aloud Phase. Im Folgenden werden wiederkehrende Themen dargestellt, die bei vielen Befragten auftraten oder Aspekte der Validität ernsthaft beeinträchtigen können. Eine vergleichende Darstellung der Ergebnisse ist in Tabelle 7 zu finden.

Verständlichkeit

Die größten Probleme bei allen Fragebögen lagen im Bereich der Verständlichkeit und umfassten z.B. schwierige oder ungewöhnliche Formulierungen, fehlende Definitionen, unzureichende Differenzierungen und/oder ein fehlendes Beispiel.

Die Teilnehmerinnen und Teilnehmer waren der Meinung, dass einige Sätze seltsame oder ungewöhnliche Formulierungen aufweisen oder schwierig zu verstehen sind. Dies führte zu Schwierigkeiten beim Verständnis der Items, was sich darin zeigte, dass die Teilnehmenden die Frage zweimal oder öfter wiederholen mussten. Darüber hinaus äußerten die Befragten sehr oft, dass Formulierungen mehrdeutig oder nicht präzise seien. Der think-aloud Ansatz zeigte, dass einige Items (z.B. Sicherheit) von verschiedenen Teilnehmenden unterschiedlich interpretiert wurden und daher die Semantik zu einer reduzierten Inter-User-Validität führte.

WHOQOL-BREF: "Wie sicher fühlen Sie sich in ihrem täglichen Leben?"

*ID16 (w, 74): Sicher fühle ich mich nicht, weil ich doch ein wenig Angst habe-
[...] Dass mir wieder etwas passiert, dass ich wieder hinfalle."*

ID25 (w, 78): „Was ist da gemeint? Diebstahl, Einbruch, oder-?"

ID32 (m, 82): „Welche Sicherheit? Finanzielle Sicherheit?"

SF-12: "Ich konnte nur bestimmte Dinge tun."

ID16 (w, 74): "Was sollen bestimmte Dinge sein? Das gibt es ja gar nicht."

Die Befragten empfanden die Beispiele manchmal als unangemessen, wobei es schwierig für sie war, eine Antwort zu geben. Einige Teilnehmende fragten die Interviewerin nach der Bedeutung oder Definition eines Begriffs oder einer Frage und hätten ein (geeignetes) Beispiel benötigt. Dies galt insbesondere für Fragen zur Gehfähigkeit. Die Fragebögen EQ-5D und NHP verwenden beispielsweise die Formulierungen "nicht gehen können" bzw. "Probleme beim Gehen". Die Definition von

"nicht gehen können" wurde von den Befragten angezweifelt, insbesondere ob dies die Verwendung von Gehhilfen einschließt. Ebenso waren die Teilnehmenden unsicher, ob sich die Fragen auf ihren Zustand unter regelmäßiger Medikation (z.B. Schmerzmittel) oder auf einen Zustand ohne Medikamente bezogen.

NHP: "Ich kann überhaupt nicht gehen" (Antwortoptionen: Ja/Nein)

ID10 (w,82): „[überlegt] Naja mit dem Gehwagen einigermaßen.“

Beim SF-12 bestanden insbesondere bei der letzten Frage aufgrund der langen, etwas holprigen Formulierung Probleme. Dies zeigte sich vor allem anhand des mehrmaligen Lesens der Frage.

Ein weiteres Problem auf der Interpretationsebene war, dass die Befragten die Frage anders interpretierten als von den Entwicklern beabsichtigt. Einige der Befragten interpretierten zum Beispiel den Punkt "Hatten Sie viel Energie?" im SF-12 als körperliche Energie und Körperkraft, die erforderlich sind, um körperlich aktiv zu sein. Tatsächlich ist das Item Teil der mentalen Komponente des SF-12 und soll die mentale Energie erfassen.

ID23 (m, 79): "Ich hab noch keine Energie. Ganz energielos praktisch. Ich kann nichts machen."

I: "Interpretieren Sie da Energie dann als körperliche Energie?"

ID23 (m, 79): "Als körperliche Energie, jaja."

Im Großen und Ganzen wurde die vorgegebene Auswahl der Antworten der vier Fragebögen gut verstanden. Manchmal hatten die Befragten Schwierigkeiten, eine Antwort zu finden, wenn sie der Meinung waren, dass die Art der Einstufung oder Differenzierung nicht ausreichend oder anwendbar ist.

EQ-5D Item „Für sich selbst sorgen“: Ich habe keine Probleme/leichte Probleme/mäßige Probleme/große Probleme/ ich bin nicht in der Lage mich selbst zu waschen oder anzuziehen.“

ID18 (w, 90): "Naja, ich würde [sagen], dass ich Hilfe noch in Anspruch nehme."

Infolgedessen antworteten Patientinnen und Patienten, die Schwierigkeiten hatten, eine Frage zu verstehen oder die Frage oder die Antwortmöglichkeiten als irrelevant betrachteten, manchmal mit der mittleren Antwortoption, wie das folgende Zitat veranschaulicht:

ID32 (m, 82): „Man kommt immer in der Mitte an, wenn man der Meinung ist, dass die Frage auf einen nicht so richtig zutrifft. (bezugnehmend auf das WHOQOL-BREF Item “Wie zufrieden sind Sie mit Ihrer Gesundheit?”)

Relevanz

Bezüglich der Relevanz bestanden Probleme wie irrelevante Items, Items mit nicht relevanten Beispielen oder unzureichende Antwortoptionen.

Bestimmte Indikatoren wurden von einigen Befragten als irrelevant in Bezug auf das Leben der Befragten angesehen, wie z.B. "Zufriedenheit mit dem Sexualleben" (WHOQOL-BREF) oder die Erfüllung von Aufgaben (SF-12 und NHP). Die Teilnehmenden erklärten die mangelnde Relevanz vor allem mit ihrem hohen Alter.

SF-12: “Ich habe weniger geschafft als ich wollte. (Antwortoptionen: Ja/Nein)

ID029 (w, 81): „Puh, da weiß ich nicht, was ich da antworten soll, weil das ein Schmarrn ist. Ich brauche ja nicht alle Tage putzen. Weil das eine blöde Frage ist, weil ich kann nicht sagen, ich habe es nicht geschafft. Weil ich bin der Einstellung, ich habe ein schönes Alter. Und wenn ich nicht mehr kann, dann setze ich mich hin und raste aus, weil ich muss ja nicht arbeiten.“

Darüber hinaus gab es einige Items, die als Thema von Bedeutung sein könnten, aber die dazu angeführten Beispiele hatten keine Bedeutung für das Leben einiger Teilnehmerinnen und Teilnehmer. Zum Beispiel listet der SF-12 Aktivitäten über die körperliche Funktionsfähigkeit auf, die über den Rahmen der üblichen Aktivitäten und Fähigkeiten der Teilnehmenden hinausgehen. Folglich wurde die Frage selbst als irrelevant betrachtet.

SF-12: “Sind Sie durch Ihren derzeitigen Gesundheitszustand bei mittelschweren Tätigkeiten (z.B. einen Tisch verschieben, staubsaugen, kegeln, Golf spielen) eingeschränkt? Wenn ja, wie stark?”

ID 24 (w, 72): „Ja. Ich würde da wahrscheinlich überall den gleichen Satz hinschreiben „Was soll das?“

Die Befragten empfanden die Antwortmöglichkeiten manchmal als unangemessen und stufte daher das jeweilige Item als nicht relevant ein. Die meisten Befragten hielten beispielsweise die dichotomen Antwortmöglichkeiten "ja" und "nein" für nicht ausreichend, da, ihrer Meinung nach, eine Option dazwischen fehlt. Dies beeinträchtigte die Validität, da die Antwortoptionen es nicht erlauben, die Situation so darzustellen, wie sie tatsächlich ist. Andererseits beschrieben einige Patientinnen und

Patienten die dichotome Antwortmöglichkeit als "ehrlich" (ID06, w, 62) und leicht verständlich.

ID32 (m, 82): "Der Fragebogen gefällt mir nicht, weil er nur immer ein Ja und ein Nein zulässt. Die Optionen sind schlecht, nicht ausreichend. Weil kaum eine Lebenssituation ist nur mit ja und mit nein definiert."

Ein weiterer Aspekt der Relevanz ist die Angemessenheit des Erinnerungszeitraums (engl. Recall Period). Die Studie wurde mit Patientinnen und Patienten durchgeführt, die sich im Durchschnitt sieben Tage zuvor einer Hüftfrakturoperation unterzogen. Die Erinnerungszeiträume "heute" oder "derzeit", wie sie beim deskriptiven System des EQ-5D bzw. beim NHP verwendet werden, wurden von einigen Teilnehmenden als nicht ganz angemessen empfunden, da sich ihr Zustand von Tag zu Tag ändert ("*weil es [Angst und Depression] jeden Tag anders ist*" ID17, m, 88).

EQ-5D, P03: „Das ist jetzt auch so eine Sache. Manchmal schon und manchmal nicht. Heute ist es schlecht. Heute habe ich große Probleme.“

Im SF-12 und WHOQOL-BREF wird eine Recall-Periode von einer bzw. zwei Wochen verwendet, so dass sich Teile dieses Zeitraums auf die Zeit vor der Fraktur und Teile auf die unmittelbare postoperative Phase bezogen. Ein Problem war die begrenzte Fähigkeit der Teilnehmenden, sich die durch den Fragebogen angeforderten Informationen in Erinnerung zu rufen. Einige Teilnehmende zählten die letzten Tage mit den Fingern, was darauf hindeutet, dass der Erinnerungszeitraum schwer zu erfassen sein kann. Da die Hüftfraktur für die Teilnehmenden ein relativ junges Ereignis ist, hatten einige Teilnehmende Schwierigkeiten zu sagen, ob sich die Fragen auf ihren "normalen" Zustand (vor der Fraktur) oder auf ihre aktuelle Situation beziehen, die sich aufgrund der Therapie ständig ändern kann. Das Ergebnis war, dass einige Patienten versucht waren, objektiv falsche Antworten anzukreuzen, d.h. Antworten, die nicht dem aktuell beobachtbaren Zustand entsprechen. Einige Patientinnen und Patienten schienen auch ihre aktuelle Situation zu verdrängen.

WHOQOL-BREF: „Haben Sie genug Energie für das tägliche Leben?“

ID27 (w, 77): "Völlig. Achso, Moment, was heute ist. Ich bin schon wieder weiter, ich bin [gedanklich] schon daheim nach der Reha."

Vollständigkeit

Die meisten Patientinnen und Patienten empfanden die Fragebögen als vollständig. Ein Grund dafür war jedoch auch, dass viele der Teilnehmenden gar keinen Anspruch auf Vollständigkeit haben. Einem Teil der Teilnehmenden zufolge ist es schwierig, ein Konzept wie HRQOL in einem einzigen Fragebogen umfassend abzudecken.

ID32 (m, 82): „Ein Fragebogen alleine für sich ist eh nicht ausreichend.“

Das laute Nachdenken und die Befragungen ergaben dennoch einige Themen, die für die HRQOL der Teilnehmenden als relevant erachtet wurden, die aber in keinem der vier Fragebögen enthalten waren. Genannte Aspekte waren u.a. Ernährung/Verdauung, Medikamente und Sehkraft.

ID021 (m, 76): " Zum Beispiel das bei mir mit den Augen, das haben wir ja nicht. Aber da kann ja der Fragebogen nichts dafür. "

Der Aspekt Medikamenteneinnahme umfasste zwei Gesichtspunkte: Erstens waren die Patientinnen und Patienten durch die Medikamenteneinnahme belastet.

ID17 (m, 88): Da muss ich mich einordnen bei "Ich bin sehr ängstlich oder deprimiert". Weil es ist ja jeden Tag anders. Das kommt in der Früh schon an bei der Tabletteneinnahme. "

Zweitens waren die Patientinnen und Patienten verunsichert, dass in keinem der Fragebögen etwas zur Medikamenteneinnahme angegeben werden musste. Dementsprechend wussten die Teilnehmenden nicht, ob bei Fragen zu Schmerz oder Schlaf beispielsweise die Einnahme von Tabletten schon berücksichtigt war.

Als häufiges Thema trat die Wichtigkeit sozialer Kontakte auf, sei es die Familie, Freunde oder auch Nachbarn.

P011 (w, 78): „Ja als mich die [Pflegerkraft] gestern gefragt hatte, was ist Ihnen am Wichtigsten [...] da war ich erstmal bei Gesundheit, weil ich eben in dieser Situation bin und dann bei Familie. Also Familie ist für mich sehr wichtig.“

Dies war oft gekoppelt an eine Hilfsbedürftigkeit gegenüber diesen Personen. In diesem Zusammenhang wurde der Aspekt der Unabhängigkeit bzw. Abhängigkeit von Personen in den Fragebögen aus Sicht der Teilnehmenden nicht (ausreichend) angesprochen. Es wurde berichtet, dass die Notwendigkeit der Hilfe anderer Personen bei einfachen täglichen Aufgaben die Lebensqualität beeinträchtigt.

ID31 (f, 85): „Das ist ganz wichtig. Dass man nicht auf Hilfe angewiesen ist.“

ID03 (f, 77): “Dass man wieder nach Hause kann, dass man sich bewegen kann, dass man seine Sachen wieder selber kann, dass man keine Hilfe braucht in dem Sinn.

Gerade beim Anziehen und beim Hausarbeit machen, das mache ich halt auch ganz gerne selbst.“

In Bezug auf einzelne Fragebögen gaben die Befragten meist nur allgemeine Eindrücke an. Einige der Befragten empfanden den EQ-5D als "zu oberflächlich". Nur wenige Patientinnen und Patienten konnten ihre Wahrnehmung weiter präzisieren und erwähnten das Fehlen einer sozialen Komponente.

ID25 (f, 78): “Also ich finde, das [EQ-5D] geht mehr peripher, während das [WHOQOL-BREF] ins Detail geht.

ID32 (m, 82): “Also Familie ist ganz wichtig. Familie ist doch das was einen hält. Ein Mensch, der in Schwierigkeit ist, muss doch ein Held sein, um selber raus zu kommen.”

Tabelle 7: Ergebnisse zu Aspekten der Inhaltsvalidität im Vergleich der untersuchten Fragebögen

	Verständlichkeit	Relevanz	Vollständigkeit
	Werden die Items und Antwortoptionen von der Zielpopulation wie intendiert verstanden? Sind die Items angemessen formuliert?	Sind die enthaltenen Items für die Zielpopulation von Interesse? Sind die Antwortoptionen angemessen?	Fehlen keine zentralen Konzepte?
EQ-5D Gesundheitsprofil	<ul style="list-style-type: none"> Item zu Mobilität: Definition von „Probleme herumzugehen“ zu unklar definiert <ul style="list-style-type: none"> ➤ Ein-/ Ausschlussprobleme bei der Beantwortung Item Angst oder Niedergeschlagenheit <ul style="list-style-type: none"> ➤ Probleme mit der „oder“ Verknüpfung 	<ul style="list-style-type: none"> Item zu alltäglichen Tätigkeiten: nicht alle Beispiele relevant Items zum Teil nicht relevant, wenn dafür Hilfe benötigt wird 	<ul style="list-style-type: none"> sehr unterschiedlich wahrgenommen <ul style="list-style-type: none"> ➤ von vollständig, bestimmtes Konzept fehlend bis oberflächlich
SF-12	<ul style="list-style-type: none"> Item zu Physische Gesundheit <ul style="list-style-type: none"> ➤ konzentrierten sich ausschließlich auf die Beispiele und antworteten direkt, ob sie genau diese Aktivitäten durchführen (können), anstatt diese als Beispiele für Aktivitäten in diesem Aktivitätslevel zu sehen der Unterschied zwischen Items zur eingeschränkten physisch-bedingten und emotional-bedingten Rollenfunktion wurde nicht klar <ul style="list-style-type: none"> ➤ hatten z.T. das Gefühl, die Fragen würden sich wiederholen 	<ul style="list-style-type: none"> Item Soziale Kontakte: eher speziell Familie statt Freunde, Freunde nicht mehr so relevant (viele schon verstorben) Zielerreichung/weniger geschafft als man wollte war ganz nach dem Tenor „müssen in diesem Alter nichts mehr schaffen“ siehe Verständlichkeit: da Beispiele nicht als Beispiele, sondern als „Abgefragte Tätigkeiten“ verstanden wurden, auch nicht als relevant angesehen 	<ul style="list-style-type: none"> vereinzelt fehlende Komponenten angemerkt (z.B. Schlaf, Hobbys, andere Emotionen, Medikamenteneinnahme)

	<ul style="list-style-type: none"> • Item zu Vitalität <ul style="list-style-type: none"> ➢ „Voller Energie“ wurde z.T. als körperliche Energie statt als – wie angedacht – mentale Energie verstanden • Item Soziale Kontakte: zu umständlich formuliert 		
NHP	<ul style="list-style-type: none"> • unklare Formulierungen bei den Items: “Körperposition verändern”, “die Kontrolle verlieren”, „alles strengt mich an“ • unklare Definition: alleine gehen oder mit Hilfe • Verständnisprobleme bei doppelt negativen Formulierungen • fehlende oder unpassende zeitliche Differenzierung (manchmal/selten) • insgesamt: fehlende Beispiele 	<ul style="list-style-type: none"> • einzelne Items (z.B. Item zu Energie) aufgrund mangelnder Relevanz von manchen als „blöde Frage“ bezeichnet <ul style="list-style-type: none"> ➢ altersentsprechend einfach weniger Energie • dichotome Ja/Nein Antwortoptionen reichten nicht aus <ul style="list-style-type: none"> ➢ keine relevante Antwort möglich 	<ul style="list-style-type: none"> • keine direkte Aussage, einmal eher das Gefühl von Wiederholungen
WHOQOL-BREF	<ul style="list-style-type: none"> • Items „sich sicher fühlen“, „Informationen für das tägliche Leben“, „Umweltbedingungen im Wohnumfeld“, „Gesundheitsdienste“ <ul style="list-style-type: none"> ➢ teilweise nicht ausreichend definitive Wortwahl oder fehlendes Beispiel 	<ul style="list-style-type: none"> • Item zu Sexualeben, Arbeitsfähigkeit, Zufriedenheit mit dem Aussehen häufig als nicht relevant eingestuft • z.T. überfordert mit der Fülle an Antwortoptionen <ul style="list-style-type: none"> ➢ zu viel Antworttext, Antworten mussten z.T. mehrfach gelesen werden 	<ul style="list-style-type: none"> • größtenteils keine direkte Aussage • Item zu sozialen Kontakten sollte spezifiziert (Freunde UND Familie) werden • Item zu psychischer Gesundheit sollte differenziert gefragt werden (Depression, Angst)

NHP= Nottingham Health Profile; SF-12= Short-Form-12 Fragebogen; WHOQOL-BREF= WHO [engl. World Health Organization] Lebensqualitätsfragebogen Kurzform

3.2 Quantitative Studie

3.2.1 Beschreibung der Studienteilnehmenden

Zur Baseline (T0/ET1) der ProFem Studie nahmen insgesamt 515 Personen teil, wobei davon 361 Personen eigenständig (64% TN + 6% TN/ZP, 70,4% Frauen, Alter $79,1 \pm 8,2$ Jahre) und 154 Personen mittels Fremdeinschätzung (71,4% Frauen, Alter $85,5 \pm 7,9$ Jahre) teilnahmen. Insgesamt handelte es sich bei fast allen Befragten (92.9%) laut Selbstangabe um deren erste Hüftfraktur (n=443). An der Follow-up Befragung (T1) zum Zeitpunkt 6 Monate nahmen insgesamt 375 Personen teil, wovon wiederum 269 Personen eigenständig teilnahmen. Bei 106 Teilnehmenden konnten die Daten über die Teilnahme eines Proxys gewonnen werden. Im Rahmen der ProFem Studie ist ein Abschlussbericht verfasst worden, der nach Veröffentlichung auf der Projektwebsite heruntergeladen werden kann (<https://www.uniklinik-duesseldorf.de/patienten-besucher/klinikeninstitutezentren/institut-fuer-versorgungs-forschung-und-gesundheitsoekonomie/forschung/versorgung-funktionsfaehigkeit-und-lebensqualitaet-nach-proximaler-femurfraktur-profem>).

Die nachfolgenden Analysen zur Beurteilung der Gütekriterien der Zustandsmessung basierten auf einer Subgruppe, die zum Zeitpunkt T0 (3 Monate) eigenständig teilnahmen (n= 306) und beinhaltete nicht die Teilnehmenden, die bei einem Ersatz-Hausbesuch aufgrund von Verhinderung bei T0 (ET1) nach 6 Monaten befragt wurden. Die Subgruppe war im Durchschnitt 79.1 Jahre alt (SD=8.2) und 69.0% (n=211) waren weiblich. Berechnungen zur Beurteilung der Änderungssensitivität (Gütekriterium der Veränderungsmessung) wurden für die selbstständig Teilnehmenden zu den beiden Zeitpunkten T0 (3 Monate) und T1 (6 Monate) bzw. deren Schnittmenge vorgenommen.

In den folgenden Abschnitten werden die Ergebnisse der psychometrischen Analysen des SF-12 und EQ-5D-5L beschrieben. Dabei werden zuerst die Ergebnisse zu den Gütekriterien der Zustandsmessung berichtet. Abschließend werden die Ergebnisse zur Überprüfung der Änderungssensitivität beschrieben.

3.2.2 Deskriptive Ergebnisse der Lebensqualitätsfragebögen und Interpretierbarkeit

EQ-5D

Die Ergebnisse zum EQ-5D Gesundheitsprofil, der EQ VAS und dem EQ-5D Index sind in Tabelle 10 aufgeführt. Das Deskriptive System des EQ-5D füllten insgesamt

305 Studienteilnehmer:innen aus. 35.3% gaben an, ‚mäßige Probleme‘ im Bereich *Beweglichkeit/Mobilität* zu haben, wohingegen im Bereich *Körperpflege/Hygiene* 32.7% ‚keine Probleme‘ angaben. Die Antworthäufigkeiten beim Item *Allgemeine Tätigkeiten* waren über alle fünf Antwortlevel relativ gleichmäßig verteilt. Eine deutliche Mehrheit (41.2%) gab allerdings im Bereich *Schmerzen/körperliche Beschwerden* ‚mäßige Probleme‘ an. Im Gegensatz zu den restlichen vier Fragen ergab sich bei der Verteilung der Antworthäufigkeiten zum Item *Angst/Niedergeschlagenheit* ein anderes Bild: die große Mehrheit (47.7%) gaben an ‚keine Probleme‘ zu haben und nur 2.9% gaben ‚extreme Probleme‘ an. Die Verteilung der Antworthäufigkeiten lies nur für die fünfte Frage Bodeneffekte vermuten. Insgesamt wurde der Fragebogen nur ein Mal (0.3%) nicht ausgefüllt, so dass von einer guten Machbarkeit ausgegangen werden kann.

Die EQ VAS wurde von 306 Studienteilnehmer:innen im Median mit 50.0 beantwortet, wobei sowohl das Minimum der Skala (0), als auch das Maximum der Skala (100) erreicht wurde. Die Verteilung lies keine Decken- oder Bodeneffekte erkennen. Die VAS wurde von allen für diese Analyse ausgewählten Studienteilnehmenden ausgefüllt, was ebenfalls für eine gute Machbarkeit spricht.

Der EQ-5D Index wurde mit dem deutschen Valueset (45) berechnet und lag im Median bei 0.67, wobei auch negative Werte vorkamen.

SF-12

Zur Berechnung von Skalenmittelwerten und Reliabilitätsmaßen wie Cronbachs α wurden invers kodierte Items rekodiert, so dass alle Antworten in die gleiche Richtung interpretiert werden können.

Beim Vergleich der beiden Summenscores des SF-12 zeigte sich eine stärker eingeschränkte körperliche Komponente der HRQOL im Vergleich zur psychischen Komponente (Median PCS-12= 28.26 vs. Median MCS-12= 52.56). Das Item zur allgemeinen Gesundheitswahrnehmung beantworteten die meisten Befragten (42.0%) mit ‚weniger gut‘. Zum Erhebungszeitpunkt waren noch 54.8% ‚stark eingeschränkt‘ beim Steigen von mehreren Treppenabsätzen. Die Antworten zum Item körperliche Schmerzen verteilten sich tendenziell zur Mitte, so gaben 33.8% mäßige körperliche Schmerzen an. In Tabelle 11 sind die bisher berichteten sowie alle weiteren Items aufgeführt. Insgesamt gab es nur einen komplett fehlenden Fragebogen und je ein

Missing bei den Items ROLEM3, VITAL2 und SOC2, so dass von einer guten Machbarkeit des Fragebogens ausgegangen werden kann. Die Verteilung der Antworthäufigkeiten wies bei keinem Item auf einen Decken- oder Bodeneffekt hin, weshalb davon auszugehen ist, dass das Instrument in der Lage ist eine vollständige Spanne an Gesundheitszuständen zu erfassen.

3.2.3 Analysen zur Beschreibung der Validität

Strukturelle Validität

Im Folgenden werden die Ergebnisse der konfirmatorischen Faktorenanalyse (CFA) für die bekannten Faktorenstrukturen des SF-12 (2-Faktor Modell: MCS-12 & PCS-12) und EQ-5D (1-Faktor Modell) aufgeführt. Die Ergebnisse des Tests nach Mardia (multivariater Normalitätstest) deuteten jeweils darauf hin, dass die Daten keiner multivariaten Normalverteilung entsprechen. Um dies zu berücksichtigen, wurde für beide Instrumente die MLR-Schätzung in der CFA verwendet. Einen Überblick über die Fit Indizes der Modellanpassung für beide Erhebungsinstrumente findet sich in Tabelle 8.

Beim EQ-5D wiesen die Fit-Indizes auf eine gute Übereinstimmung zwischen dem angenommenen theoretischen 1-Faktor Modell mit den empirischen Daten hin. Die standardisierten Faktorladungen lagen zwischen 0.46 und 0.84 (siehe Tabelle 9) und waren alle signifikant ($p < 0.0001$).

Tabelle 8: Konfirmatorische Faktorenanalyse: Indizes der Modellanpassung

Indizes der Modellanpassung	akzeptabler Modellfit	SF-12 (n= 303)	EQ-5D deskriptives System (n=306)
Comparative Fit Index (CFI)	≥ 0.95	0.741	0.986
Root Mean Square Error of Approximation (RMSEA)	< 0.08	0.144	0.072
Standardized Root Mean Square Residual (SRMR)	< 0.10	0.095	0.032
Normierter Chi-Quadrat-Wert (χ^2/df)	< 5	7.30	2.59

KI, Konfidenzintervall; MCS-12, psychische Summenskala [engl. mental component score]; PCS-12, körperliche Summenskala [engl. physical component score]; SF-12, Short-Form 12 Fragebogen

Beim SF-12 lagen die standardisierten Faktorladungen der CFA zwischen 0.44 und 0.76 (siehe Tabelle 12) und die Fit Indizes lagen außerhalb des akzeptablen Bereichs. Auch die χ^2 -Statistik war signifikant, was ebenfalls auf ein nicht passendes Modell hindeutet.

Tabelle 9: EQ-5D: Items und Itemkennwerte (standardisierte Faktorladungen und Standardfehler)

N= 307	Faktorstruktur latenter Faktor gesundheitsbezogene Lebensqualität Schätzer (SE)
Beweglichkeit/Mobilität	0.774
Körperpflege/Hygiene	0.843
Allgemeine Tätigkeiten	0.809
Schmerzen/körperliche Beschwerden	0.611
Angst/Niedergeschlagenheit	0.461

Alle Ladungen waren signifikant von 0 verschieden ($p < 0.0001$).

Tabelle 10: Deskription der EQ-5D-5L Antworten zur Baseline

EQ-5D Gesundheitsprofil (N= 305)						
	Antwortlevel n (%)	Level 1 (keine Probleme)	Level 2 (leichte Probleme)	Level 3 (mäßige Probleme)	Level 4 (starke Probleme)	Level 5 (extreme Probleme/ nicht in der Lage)
EQ-5D Dimension						
Beweglichkeit/Mobilität		22 (7.2%)	74 (24.2%)	108 (35.3%)	75 (24.5%)	27(8.8%)
Körperpflege/Hygiene		100 (32.7%)	77 (25.2%)	64 (20.9%)	37 (12.1%)	28 (9.2%)
Allgemeine Tätigkeiten		49 (16.0%)	67 (21.9%)	83 (27.1%)	57 (18.6%)	50 (16.3%)
Schmerzen/körperliche Beschwerden		33 (10.8%)	71 (23.2%)	126 (41.2%)	59 (19.3%)	17 (5.6%)
Angst/Niedergeschlagenheit		146 (47.7%)	66 (21.6%)	60 (19.6%)	25 (8.2%)	9 (2.9%)
EQ VAS (N= 306)						
Mittelwert (Standardabweichung)		55.44 (20.77)				
Median (95%-Konfidenzintervall)		50.0 (53.10-57.78)				
EQ-5D Index (N= 306)						
Mittelwert (Standardabweichung)		0.55 (0.35)				
Median (95%-Konfidenzintervall)		0.67 (0.52-0.59)				

VAS= visuelle Analogskala

Tabelle 11: Deskription der SF-12 Antworten zur Baseline

SF-12 Items	Kennung	M (SD)	Antworthäufigkeiten n (%)					N
1 – Allgemeine Gesundheitswahrnehmung (rekodiert)	GHP1	2.29 (0.80)	1 schlecht 50 (16.4%)	2 weniger gut 128 (42.0%)	3 gut 115 (37.7%)	4 sehr gut 10 (3.3%)	5 ausge- zeichnet 2 (0.7%)	305
2- Durch Gesundheitszustand eingeschränkt: mittelschwere Tätigkeiten	PFI02	1.49 (0.69)	1 ja, stark 189 (62.0%)	2 ja, etwas 82 (26.9%)	3 nein 34 (11.1%)		305	
3 - Durch Gesundheitszustand eingeschränkt: mehrere Treppenabsätze steigen	PFI04	1.57 (0.69)	1 ja, stark 167 (54.8%)	2 ja, etwas 102 (33.4%)	3 nein 36 (11.8%)		305	
4 – Durch körperliche Beeinträchtigung Schwierigkeiten bei Arbeit: weniger geschafft als ich wollte	ROLPH2	1.20 (0.40)	1 ja 244 (80.0%)	2 nein 61 (20.0%)			305	
5 – Durch körperliche Beeinträchtigung bei Arbeit: konnte nur bestimmte Dinge tun	ROLPH3	1.15 (0.36)	1 ja 258 (84.6%)	2 nein 47 (15.4%)			305	
6- Durch seelische Probleme Schwierigkeiten bei der Arbeit: weniger geschafft als ich wollte	ROLEM2	1.63 (0.48)	1 ja 112 (36.7%)	2 nein 193 (63.3%)			305	
7- Durch seelische Probleme Schwierigkeiten bei Arbeit: nicht so sorgfältig	ROLEM3	1.67 (0.47)	1 ja 99 (32.6%)	2 nein 205 (67.4%)			304	

8 – Körperliche Schmerzen (rekodiert)	PAIN2	2.88 (1.20)	1 sehr 48 (15.7%)	2 ziemlich 64 (21.0%)	3 mäßig 103 (33.8%)	4 ein bisschen 58 (19.0%)	5 über- haupt nicht 32 (10.5%)		305
9 – Ruhig und gelassen (rekodiert)	MHI3	4.16 (1.42)	1 nie 5 (1.6%)	2 selten 43 (14.1%)	3 manch- mal 63 (20.7%)	4 ziemlich oft 45 (14.8%)	5 meistens 85 (27.9%)	6 immer 64 (21.0%)	305
10 – Voller Energie (rekodiert)	VITAL2	3.23 (1.47)	1 nie 37 (12.2%)	2 selten 68 (22.4%)	3 manch- mal 85 (28.0%)	4 ziemlich oft 38 (12.5%)	5 meistens 53 (17.4%)	6 immer 23 (7.6%)	304
11 – Entmutigt und traurig	MHI4	4.22 (1.33)	1 immer 11 (3.6%)	2 meistens 17 (5.6%)	3 ziemlich oft 57 (18.7%)	4 manch- mal 100 (32.8%)	5 selten 50 (16.4%)	6 nie 70 (23.0%)	305
12 – Kontakte zu anderen Menschen beeinträchtigt	SOC2	3.67 (1.35)	1 immer 27 (8.9%)	2 meistens 39 (12.8%)	3 manch- mal 65 (21.4%)	4 selten 50 (16.4%)	5 nie 123 (40.5%)		304
Maße der zentralen Tendenz für	Mittelwert (Standardabweichung)		Median (95%-Konfidenzintervall)			Minimum; Maximum			
PCS-12 (n= 302)	30.06 (9.54)		28.26 (28.98-31.13)			13.73; 54.84			
MCS-12 (n= 302)	50.06 (11.86)		52.56 (48.72- 51.40)			17.10; 71.09			

GHP1= General Health Perception; M= Mittelwert; PF12/PF14= Physical Functioning; PAIN2= Pain; ROLPH2/ROLPH3= Physical Role Functioning; ROLEM2/ROLEM3= Emotional Role Functioning; MHI3/MHI4= Mental Health; VITAL2= Vitality; SOC2= Social Functioning; SD= Standardabweichung

Tabelle 12: SF-12: Items und Itemkennwerte (standardisierte Faktorladungen und Standardfehler)

N=304		Faktorstruktur	
		latenter Faktor	latenter Faktor
Kennung	SF-12 Domäne	körperliche	psychische
		Gesundheit	Gesundheit
		Schätzer (SE)	Schätzer (SE)
PFI2	Physische Gesundheit	0.761	
PFI4	Physische Gesundheit	0.719	
ROLPH2	Eingeschränkte physisch-bedingte Rollenfunktion	0.622	
ROLPH3	Eingeschränkte physisch-bedingte Rollenfunktion	0.664	
PAIN2	Körperliche Schmerzen	0.635	
GHP1	Allgemeine Gesundheitswahrnehmung	0.637	
VITAL2	Vitalität		0.633
SOC2	Soziale Funktionsfähigkeit		0.438
ROLEM2	Eingeschränkte emotional-bedingte Rollenfunktion		0.713
ROLEM3	Eingeschränkte emotional-bedingte Rollenfunktion		0.673
MHI3	Mentale Gesundheit		0.630
MHI4	Mentale Gesundheit		0.729

Alle Ladungen waren signifikant von 0 verschieden ($p < 0.0001$). GHP1= General Health Perception; PFI2/PFI4= Physical Functioning; PAIN2= Pain; ROLPH2/ROLPH3= Physical Role Functioning; ROLEM2/ROLEM3= Emotional Role Functioning; MHI3/MHI4= Mental Health; VITAL2= Vitality; SOC2= Social Functioning

Konvergenz- und Diskriminanzvalidität

Die konvergente und divergente Validität der beiden Instrumente wurde durch die Zusammenhänge zwischen den beiden Instrumenten sowie mit weiteren Referenzmaßen überprüft (siehe Tabelle 13).

Die Zusammenhänge zwischen den beiden Instrumenten zeigten stärkere Korrelationen für vergleichbare Summenscores und Dimensionen, z.B. PCS-12 mit Beweglichkeit/Mobilität ($r = -0.61$) und MCS-12 mit Angst/Niedergeschlagenheit ($r = -0.60$), was auf eine konvergente Validität hinweist. Andererseits korrelierten weniger vergleichbare Summenscores und Dimensionen nur schwach, z.B. PCS-12 und Angst/Niedergeschlagenheit ($r = -0.21$) und MCS-12 und Beweglichkeit/Mobilität ($r = -0.27$), was für die divergente Validität des SF-12 und EQ-5D-5L spricht. Im Gegensatz

dazu korrelierte die EQ VAS moderat mit beiden SF-12 Summenscores, nämlich $r=0.55$ mit PCS-12 und $r= 0.41$ mit MCS-12.

Tabelle 13: Konvergenz- und Diskriminanzvalidität: Korrelationen (Spearman's rho) zwischen den Lebensqualitäts- und Referenzmaßen

	SF- 12 (MCS)	SF-12 (PCS)	OHS	Barthel- Index	OSS-3	IMET
SF-12 (MCS)	1	0.05	0.37	0.23	0.23	-0.40
SF-12 (PCS)	0.05	1	0.67	0.59	0.18	-0.58
EQ-5D: Beweglichkeit/Mobilität	-0.27	-0.60	-0.71	-0.62	-0.21	0.62
EQ-5D: Selbstversorgung	-0.39	-0.55	-0.70	-0.66	-0.26	0.65
EQ-5D: Allgemeine Tätigkeiten	-0.35	-0.63	-0.69	-0.57	-0.21	0.67
EQ-5D: Schmerzen/körperliche Beschwerden	-0.27	-0.58	-0.61	-0.48	-0.12	0.45
EQ-5D: Angst/Niedergeschlagenheit	-0.63	-0.21	-0.41	-0.29	-0.18	0.42
EQ-5D Index	0.44	0.67	0.77	0.66	0.22	-0.71
EQ VAS	0.41	0.55	0.59	0.53	0.22	-0.59

Alle Korrelationen sind statistisch signifikant ($p < 0.05$). Fett hervorgehobene Feldinhalte zeigen der Hypothese entsprechende Korrelationen. Aufgrund von fehlenden Werten basieren die Korrelationen auf Stichprobengrößen zwischen 263 und 306. IMET= Index zur Messung von Einschränkungen der Teilhabe; MCS-12= psychische Summenskala [engl. mental component score]; OHS= Oxford Hip Scale; OSS-3= Oslo Social Support Skala, PCS-12= körperliche Summenskala [engl. physical component score]; SF-12= Short-Form 12 Fragebogen; VAS= visuelle Analogskala

Die Zusammenhänge der beiden untersuchten Instrumente mit verschiedenen Referenzmaßen stimmten in 17 von 24 Hypothesen beim SF-12 und in 32 von 42 Hypothesen beim EQ-5D mit den vorher formulierten Hypothesen überein. Das EQ-5D Item Angst/Niedergeschlagenheit korrelierte stärker als erwartet mit dem OHS ($r=-0.41$) sowie schwächer als erwartet mit dem OSS-3 ($r=-0.18$). Beim SF-12 gab es auch bei der psychischen Komponente die meisten Abweichungen von den ursprünglich formulierten Hypothesen. Insgesamt betrachtet, sprechen die Ergebnisse für eine gute Konvergenz- und Diskriminanzvalidität des EQ-5D (weniger als 25% der Hypothesen mussten verworfen werden) und eine moderate Konvergenz- und Diskriminanzvalidität des SF-12 (29% der Hypothesen mussten verworfen werden).

known-groups Validität

Der Hypothese entsprechend, hatten jüngere Teilnehmende und männliche Teilnehmer im Median einen besseren Gesundheitsstatus. Die Gruppenunterschiede waren allerdings nur bei einzelnen Instrumenten/Scores signifikant (siehe Tabelle 14). Dieses Bild zeigte sich auch für die Variable der Fragebogenversion, die als Surrogat für den Gesundheitszustand dienen sollte. Der Hypothese folgend sollte die Gruppe, die nur den Kurzfragebogen ausgefüllt hat, schlechtere Summenscores aufweisen.

Tabelle 14: *Known-groups Validität: Vergleiche des SF-12, EQ-5D Index und EQ VAS hinsichtlich verschiedener Subgruppen*

Gruppe	n	EQ VAS (Md, IQR)	n	EQ-5D Index (Md, IQR)	n	SF-12	
						MCS-12 (Md, IQR)	PCS-12 (Md, IQR)
<i>Geschlecht</i>							
Männlich	89	60.0 (31.2)	89	0.74 (0.40)	89	56.9 (20.4)	29.7 (14.7)
Weiblich	184	50.0 (30.0)	184	0.66 (0.53)	184	51.5 (17.0)	28.1 (11.8)
p		0.161		0.033		0.015	0.191
<i>Alter</i>							
≤ 80 Jahre	147	60.0 (30.0)	147	0.72 (0.48)	147	52.9 (20.8)	29.9 (13.5)
> 80 Jahre	125	50.0 (20.0)	127	0.66 (0.52)	126	53.4 (16.0)	27.1 (11.3)
p		0.020		0.065		0.313	0.080
<i>Fragebogenversion</i>							
Originalversion	274	50.0 (35.0)	275	0.70 (0.50)	274	53.2 (18.2)	28.4 (12.9)
Kurzversion	31	50.0 (23.5)	31	0.53 (0.44)	29	49.0 (12.3)	27.4 (11.5)
p		0.383		0.062		0.071	0.802
<i>Subjektiver sozioökonomischer Status (Mac Arthur Scale)</i>							
niedrig	35	50.0 (30.0)	35	0.50 (0.59)	35	44.2 (19.2)	25.7 (9.19)
mittel	241	50.0 (25.0)	241	0.70 (0.52)	239	52.9 (18.3)	28.5 (13.3)
hoch	19	75.0 (30.0)	19	0.67 (0.40)	19	54.4 (18.3)	29.9 (20.2)
p		0.004		0.014		0.045	0.009

Statistisch signifikante ($p < 0.05$) Korrelationen sind hervorgehoben (fett). IQR= Interquartilsabstand [engl. interquartile range]; MCS-12= psychische Summenskala [engl. mental component score]; Md= Median; PCS-12= körperliche Summenskala [engl. physical component score]; SF-12= Short-Form 12 Fragebogen; VAS= visuelle Analogskala

Die Unterschiede in den Scores fielen zwar ebenfalls in die erwartete Richtung aus, waren jedoch nicht statistisch signifikant. Im Gegensatz dazu waren beim Vergleich der Teilnehmenden anhand des subjektiven sozialen Status die Unterschiede bei allen

Summenscores signifikant. Die Ergebnisse konnten nur einzelne erwartete Zusammenhänge mit einer Fehlerwahrscheinlichkeit unter 5% ($p < 0.05$) bestätigen, jedoch zeigte sich ein konsistentes Bild bei allen Instrumenten/Scores in die erwartete Richtung. Die Ergebnisse lassen auf eine akzeptable Konstruktvalidität des SF-12 und EQ-5D-5L schließen.

3.2.4 Analysen zur Beschreibung der Reliabilität

Das standardisierte Cronbach's α betrug 0.8 für die MCS-12 und 0.83 für die PCS-12. Für beide Komponenten des SF-12 kann somit eine konsistente Skala angenommen werden.

3.2.5 Analysen zur Beschreibung der Änderungssensitivität

In diesem Abschnitt werden mittels eines Kriteriums-basierten Ansatzes und eines Konstrukt-Ansatzes Kennwerte zur Beurteilung der Änderungssensitivität der untersuchten HRQOL Instrumente berichtet.

Kriteriums-basierter Ansatz

Von 269 eigenständig Teilnehmenden liegen von 266 Teilnehmenden Angaben zur subjektiven Veränderung ihres Gesundheitszustandes (GRCS Item) vor. Es gaben 164 Teilnehmende eine subjektive Veränderung ihres Gesundheitszustandes zwischen der Baseline und 6 Monatserhebung an, wohingegen 102 Teilnehmende keine Veränderung angaben. Die Teilnehmenden mit einer subjektiven Veränderung empfanden meistens eine subjektive Verbesserung innerhalb des dreimonatigen Zeitraums, jedoch gaben auch 33 (12.4%) eine Verschlechterung ihres Gesundheitszustandes an.

Die Ergebnisse der einzelnen HRQOL Instrumente bzw. der jeweiligen Scores zeigten ein uneinheitliches Bild. Entgegen der Erwartung zeigten sich überwiegend keine eindeutigen Verbesserungen in den HRQOL Scores zum Zeitpunkt 6 Monate im Vergleich zur Baseline (3 Monate) (siehe Tabelle 15). Der EQ-5D Index sowie die EQ VAS wiesen keine signifikante Veränderung zwischen den beiden Messzeitpunkten auf. Beim SF-12 Fragebogen zeigten sich je nach betrachteter Komponente sogar zwei gegensätzliche Trends: so verbesserte sich die PCS-12 (körperliche Komponente) signifikant, wohingegen sich die MCS-12 (psychische Komponente) signifikant verschlechterte.

Tabelle 15: Scores der HRQOL Instrumente zu den Zeitpunkten 3 und 6 Monaten und entsprechender Change Score

Instrument	Baseline		6 Monate		Change Score		p-Wert
	n	Md (Q ₂₅ -Q ₇₅)	n	Md (Q ₂₅ -Q ₇₅)	n	Md (Q ₂₅ -Q ₇₅)	
EQ-5D Index	306	0.67 (0.32-0.83)	264	0.68 (0.34-0.85)	263	-0.02 (-0.13-0.10)	0.186
EQ VAS	305	50.0 (40.0-75.0)	266	50.0 (40.0-70.0)	264	0.0 (-10.0-10.0)	0.487
PCS-12	303	28.26 (22.66-35.52)	228	30.39 (25.07-37.35)	225	-1.92 (-6.08-2.95)	0.001
MCS-12	303	52.56 (40.50-59.48)	228	44.72 (35.17-56.94)	225	3.65 (-2.79-11.59)	<0.001

^a Wilcoxon matched-pair test; MCS-12= psychische Summenskala [engl. mental component score]; Md= Median; PCS-12= körperliche Summenskala [engl. physical component score]; Q₂₅-Q₇₅= 25.Quartil – 75. Quartil; VAS= visuelle Analogskala

Kategorisierte man die Teilnehmenden anhand der GRCS und betrachtete ebenfalls die einzelnen Scores, zeigte sich weiterhin ein inkonsistentes Bild (siehe Tabelle 16).

In der Gruppe, die anhand der GRCS eine starke Verbesserung berichtete, lies sich eine signifikante Verbesserung in der PCS-12 beobachten. Allerdings zeigten weder der EQ-5D Index, noch die EQ VAS eine signifikante Veränderung.

Bei Teilnehmenden, die bei der GRCS eine leichte Verbesserung angaben, konnte bei keinem der untersuchten Instrumente/Scores eine signifikante Verbesserung abgebildet werden. In der laut GRCS unveränderten Gruppe zeigte sich allerdings entgegen der Erwartung eine signifikante Verbesserung der PCS-12 sowie eine signifikante Verschlechterung der MCS-12 zwischen den beiden Zeitpunkten. Die signifikante Verschlechterung der MCS-12 im Vergleich zur Baseline bestand auch in der Gruppe der Teilnehmenden, die bei der GRCS eine Verschlechterung (leicht & stark) und sogar eine leichte Verbesserung angaben.

Verglich man die Scores zwischen dem 6-Monats Follow-up und der Baseline in der subjektiv leicht verschlechterten Gruppe, zeigte sich ebenfalls kein einheitliches Bild, wobei keine Veränderung signifikant war. Im Gegensatz zu den anderen Gruppen ging eine selbstberichtete starke Verschlechterung laut GRCS mit einer einheitlichen

Verschlechterung in den Medianen der Scores einher. Allerdings handelte es sich um eine sehr kleine Gruppe (n=13) und nur die Veränderung in der MCS-12 war signifikant.

Tabelle 16: Änderungssensitivität der Lebensqualitätsmessinstrumente anhand der selbstberichteten Veränderung (Ankerfrage)

Ankerfrage		Instrumente			
Antwort- level	Kategorie (n)	Score	Baseline Md (Q ₂₅ -Q ₇₅)	6-Monate Md (Q ₂₅ -Q ₇₅)	p-Wert ^a
4, 5	stark verbessert (n= 32)	EQ-5D Index	0.84 (0.63-0.93)	0.89 (0.71-0.96)	0.123
		EQ VAS	75.0 (47.5-87.5)	75.0 (60.0-90.0)	0.067
		PCS-12	35.06 (28.09-44.49)	39.72 (31.92-47.60)	0.037
		MCS-12	56.82 (42.91-61.97)	56.13 (38.26-60.70)	0.218
1, 2, 3	leicht verbessert (n=99)	EQ-5D Index	0.73 (0.41-0.83)	0.72 (0.45-0.84)	0.053
		EQ VAS	50.0 (45.0-70.0)	57.50 (50.0-70.0)	0.799
		PCS-12	29.31 (23.04-34.70)	30.88 (25.71-37.23)	0.092
		MCS-12	51.47 (41.46-59.01)	48.12 (36.48-58.16)	0.031
0	keine Veränderung (n=102)	EQ-5D Index	0.65 (0.28-0.82)	0.63 (0.28-0.83)	1.0
		EQ VAS	50 (46.5-70.0)	50.0 (40.0-65.0)	0.503
		PCS-12	27.83 (22.53-35.39)	29.24 (24.71-35.62)	0.012
		MCS-12	53.57 (42.14-59.60)	41.85 (34.62-54.26)	<0.001
-1, -2, -3	leicht verschlechtert (n=20)	EQ-5D Index	0.29 (0.16-0.62)	0.38 (0.17-0.74)	0.942
		EQ VAS	50.0 (32.50-70.0)	45.0 (28.0-52.5)	0.123
		PCS-12	24.45 (20.82-29.21)	25.60 (21.63-31.41)	0.325
		MCS-12	51.65 (38.27-60.13)	41.17 (37.46-57.52)	0.024
-4, -5	stark verschlechtert (n=13)	EQ-5D Index	0.49 (-0.12-0.75)	0.41 (-0.01-0.58)	0.787
		EQ VAS	40.0 (30.0-60.0)	30 (10.0-45.0)	0.117
		PCS-12	26.77 (20.56-29.80)	24.84 (23.39-26.87)	0.652
		MCS-12	37.90 (35.70-45.94)	31.26 (27.42-34.45)	0.004

^a Wilcoxon matched-pair test; IQR= Interquartilsabstand [engl. interquartile range]; MCS-12= psychische Summenskala [engl. mental component score]; Md= Median; PCS-12= körperliche Summenskala [engl. physical component score]; SD= Standardabweichung [engl. standard deviation]; VAS= visuelle Analogskala

Insgesamt konnten mit dem Kriteriums-basierten Ansatz nur fünf von 20 Hypothesen bestätigt werden, so dass 75% der Hypothesen verworfen werden mussten und die Änderungssensitivität damit als gering bewertet werden muss.

Konstrukt-Ansatz

Die Korrelationen der HRQOL change scores mit den change scores der Referenzmaße sind in Tabelle 17 zu finden.

Tabelle 17: Korrelationen der Change Scores der Lebensqualitäts- und Referenzinstrumente: Abgleich mit a priori-Hypothesen

Zusammenhang zwischen Instrumenten	Hypothese	n	Spearman Rangkorrelations- koeffizient (gerundet)	Bestätigung der Hypothese
MCS-12 CS – EQ-5D Index CS	0.3-0.7	224	0.3	Ja
MCS-12 CS – EQ VAS CS	0.3-0.7	222	0.2	Nein
PCS-12 CS – EQ-5D Index CS	0.3-0,7	224	0.2	Nein
PCS-12 CS – EQ VAS CS	0.3-0.7	222	0.3	Ja
PCS-12 CS – OHS CS	≥ 0.5	206	0.4	Nein
PCS-12 CS – Barthel Index CS	0.3-0.7	197	0.2	Nein
PCS-12 CS – Barthel Index CS vs. PCS-12 CS – IMET CS	0.1 höher	197 vs. 224	0.2 vs. 0.2	Nein
PCS-12 CS – Barthel-Index CS vs. MCS-12 CS – Barthel-Index CS	0.1 höher	197 vs. 197	0.2 vs. 0.1	Ja
EQ-5D Index CS – OHS CS	0.3-0.7	239	0.4	Ja
EQ VAS CS – OHS CS	0.3-0.7	239	0.3	Ja
EQ-5D Index CS – OHS CS vs. EQ-5D Index CS – Barthel Index CS	0.1 höher	239 vs. 229	0.4 vs. 0.2	Ja
EQ VAS CS – OHS CS vs. EQ VAS CS – Barthel Index CS	0.1 höher	239 vs. 228	0.3 vs. 0.2	Ja

CS= Change Score; MCS-12= psychische Summenskala [engl. mental component score]; OHS= Oxford Hip Scale, PCS-12= körperliche Summenskala [engl. physical component score]; SF-12= Short-Form 12 Fragebogen; VAS= visuelle Analogskala

Sechs der acht vorher für den EQ-5D-5L change score festgelegten Hypothesen konnten bestätigt werden, so dass die Änderungssensitivität des EQ-5D-5L anhand dieses Ansatzes als moderat bewertet werden kann.

Dem SF-12 wird anhand dieser Ergebnisse eine geringe Änderungssensitivität zugeschrieben, da mehr als 50% der a-priori Hypothesen nicht bestätigt werden konnten.

Die Ergebnisse der beiden verwendeten Ansätze sprechen insgesamt für eine geringe Änderungssensitivität der hier untersuchten HRQOL Instrumente EQ-5D-5L und SF-12 in der untersuchten Stichprobe.

4. Diskussion

4.1 Qualitative Studie

4.1.1 Hauptergebnisse

Bei älteren Hüftfrakturpatient:innen stellt die Messung der HRQOL mittels eines Selbstaussfüll-Fragebogens eine Herausforderung dar, sowohl bezüglich der Machbarkeit als auch teilweise aufgrund der abgefragten Inhalte. In der vorliegenden Interviewstudie waren weniger als ein Drittel der 32 Teilnehmerinnen und Teilnehmer in der Lage, mindestens eines der vier getesteten HRQOL-Instrumente selbst bzw. ohne Hilfe einer Interviewerin oder eines Interviewers auszufüllen. Häufig war dies begründet durch Einschränkungen der Seh- und/oder motorischen Funktionsfähigkeit.

Die meisten Items der jeweiligen Fragebögen waren gut verständlich. Manche Items, wie beispielsweise „sich im täglichen Leben sicher fühlen“ oder „genug Energie haben“ wurden jedoch als mehrdeutig beschrieben. Insgesamt waren manche Teilnehmende verunsichert, ob sie bei der Beantwortung der Fragebögen beispielsweise ihre üblichen Medikamente oder die Zuhilfenahme von Hilfsmitteln, wie z.B. einen Gehwagen, berücksichtigen sollen. Einzelne Items wurden von einigen Teilnehmenden als irrelevant betrachtet. Dies beinhaltet beispielsweise die Leistungsfähigkeit oder die Zufriedenheit mit dem Sexualleben. Dem gegenüberstehend brachten einige Teilnehmende weitere HRQOL-Determinanten zur Sprache, die sie in allen untersuchten Fragebögen vermissten: die Belastung durch die Medikamenteneinnahme, das Angewiesensein auf Hilfe bei Aktivitäten des täglichen Lebens und die Einschränkung durch einen möglichen Sehkraftverlust.

Während einige Teilnehmende die kürzeren Instrumente präferiert ausfüllten, bemerkten andere, dass kurze Fragebögen zu oberflächlich sein könnten, um ihre HRQOL angemessen darzustellen. Auf die Frage nach einem bevorzugten Fragebogen gaben die Teilnehmenden an, wenn überhaupt, den EQ-5D (deskriptives System) und den SF-12 zu wählen. Auch die Inhaltsanalyse wies bei diesen beiden Fragebögen die geringsten Einschränkungen hinsichtlich der inhaltlichen Validität und Machbarkeit auf.

Bisher gibt es aufgrund mangelnder, robuster Studien keine klaren Empfehlungen für die Wahl eines geeigneten Instrumentes zur Messung der HRQOL bei Hüftfrakturpatient:innen. Aus diesem Grund wurden insbesondere weitere

Forschungsarbeiten zur Akzeptanz, Relevanz und Durchführbarkeit verschiedener Instrumente empfohlen, bei denen die Perspektive der Patienten aktiv mit einbezogen werden soll (26). Die Ergebnisse der vorliegenden Studie liefern einen wichtigen Beitrag aus der Patientenperspektive und unterstützen damit die Wahl eines geeigneten HRQOL Messinstrumentes bei älteren Hüftfrakturpatient:innen.

Die psychometrischen Messeigenschaften eines Instrumentes und somit die Güte der HRQOL Messung kann bevölkerungsspezifisch sein und sollte deshalb für jede spezifische Bevölkerungsgruppe individuell bewertet werden (36). Bei der Gruppe der älteren Hüftfrakturpatient:innen ist davon auszugehen, dass die Akuität der Fraktur sowie der Fokus auf die Gehfähigkeit das Antwortverhalten und die Perspektive auf die Fragebögen beeinflussen.

Im folgenden Abschnitt können die Ergebnisse dieser Studie allerdings nur mit Studien zu Inhaltsvalidität und Machbarkeit bei älteren Menschen im Allgemeinen verglichen werden, da dies die erste Studie zur Perspektive von Hüftfrakturpatient:innen zu den gewählten Fragebögen ist.

4.1.2 Vergleich mit anderen Studien

Machbarkeit

Noch vor Überprüfung aller Messeigenschaften ist die Abklärung der Machbarkeit in der jeweiligen Zielgruppe vorzunehmen. Die Ergebnisse unserer Studie legen nahe, dass die Datenerfassung von PRO möglicherweise von Interviewer:innen unterstützt werden muss und als eine Ressource überall dort berücksichtigt werden muss, wo die PRO Datenerfassung als wichtig erachtet wird. In der hier durchgeführten Studie erwies sich das selbstständige Ausfüllen von PRO nur als eingeschränkt möglich. Dies gilt jedoch wahrscheinlich für ältere Menschen im Allgemeinen, so zeigt ein aktueller Review zur Machbarkeit des EQ-5D-3L/5L bei älteren Menschen ebendies (89). Auch berichtet eine Studie von van Leeuwen und Kollegen zur Inhaltsvalidität und Machbarkeit von drei HRQOL-Instrumenten (EQ-5D, ICECAP-O und ASCOT) bei Älteren in den Niederlanden über sehr ähnliche Machbarkeitsprobleme, wie z. B. die fehlende Kraft zum Ausfüllen der Instrumente (90). Ebenso fanden Bureau-Chalot et al. (91), dass beim Ausfüllen des NHP durch Demenzpatient:innen stets ein Interviewer oder eine Interviewerin anwesend sein muss, um auf Verständnis- oder Aufmerksamkeitsprobleme reagieren zu können (Bureau-Chalot et al. , 2002). Baró et

al. (92) bezeichnen die Verwendung des NHP bei Älteren mit mäßiger kognitiver Beeinträchtigung als geeignet, zuverlässig und valide, wenn der Interviewer oder die Interviewerin die Befragung durchführt. In unserer Studie wurden Patient:innen mit schweren kognitiven Beeinträchtigungen ausgeschlossen.

Ein weiterer Punkt, der die Machbarkeit beeinflusst, ist die Länge des jeweiligen Fragebogens. Es war besonders deutlich, dass die Hüftfrakturpatient:innen in dieser Studie umso mehr bereit waren, die Fragebögen auszufüllen, je kürzer sie in Bezug auf die Seitenzahl waren. So lehnten die Patientinnen und Patienten es am häufigsten ab, den WHOQOL-BREF auszufüllen, der mit zwei Doppelseiten der längste Fragebogen bezüglich der Seitenanzahl war. Ein Review von Rolstad und Kollegen untersuchte den Zusammenhang zwischen der Belastung der Befragten (gemessen an der Rücklaufquote) und der Länge des Fragebogens (93). Sie fanden, dass andere Faktoren (wie z.B. der Inhalt an sich) als die reine Länge des Fragebogens zwar mindestens genauso wichtig sein könnten, jedoch zeigte sich in der Meta-Analyse ein ähnliches Bild wie in der hier vorliegenden Studie: wenn den Patienten ein im Vergleich kürzerer Fragebogen vorgelegt wurde, gab es eine höhere Chance, dass dieser beantwortet wurde.

Darüber hinaus beeinflusst das gewählte Layout des Fragebogens die Machbarkeit und sogar die Validität des Fragebogens. Ein übersichtliches Layout steigerte die Wahrscheinlichkeit, dass die Hüftfrakturpatient:innen die von ihnen beabsichtigte Antwort ankreuzten und keinen Punkt übersahen. So geht aus der hier vorliegenden Studie die Empfehlung hervor, beim deskriptiven System des EQ-5D-5L zwischen den Items und den Ankreuzfeldern Hilfslinien zu ziehen, um ein Verrutschen in den Antworten zu vermeiden. Beim NHP konnte ein irreführendes Layout in Dreierblöcken festgestellt werden. Einige Patientinnen und Patienten dachten fälschlicherweise, dass nur eine der drei Fragen pro Block beantwortet werden müsse. Ohne die Anwesenheit einer Interviewerin oder eines Interviewers wäre also eine große Anzahl von Fragebögen aufgrund einer zu hohen Anzahl fehlender Items nicht auswertbar.

Inhaltsvalidität

Nach Testung der Machbarkeit gilt es die wichtigste Messeigenschaft, die Inhaltsvalidität, zu überprüfen. Auch hier zeigen sich beim Vergleich der Ergebnisse mit bereits vorhandenen Studien bezüglich der Inhaltsvalidität übereinstimmende

Aspekte, da bei der Verwendung von Fragebögen zur Messung der HRQOL bei älteren Menschen im Allgemeinen von Problemen berichtet wird (46, 94).

Darunter fallen Probleme in der Verständlichkeit. Besonders deutlich wurden diese bei der Bewertung der Mobilität oder der Gehfähigkeit. Insbesondere bei Hüftfrakturpatienten werden PROMs auch zur Beurteilung der Wirksamkeit von Interventionen eingesetzt. Allerdings ist der Einsatz von Instrumenten in diesem Zusammenhang stark eingeschränkt, wenn die Gehfähigkeit nicht klar definiert ist oder die Verwendung von Hilfsmitteln nicht berücksichtigt wird. Bei allen in dieser Studie verwendeten Fragebögen wurden deshalb teils von den Teilnehmenden selbst Bedenken hinsichtlich der inhaltlichen Validität geäußert. Auch die Ergebnisse der Inhaltsanalyse lassen bezüglich dieses Aspekts der Mobilität auf eine Einschränkung der Inhaltsvalidität schließen. In einer Studie in Großbritannien wurde untersucht, was für die Patientinnen und Patienten bei der Beurteilung der Genesung nach einer Hüftfraktur wichtig ist. Für die Teilnehmenden in dieser Studie war die Mobilität nach einer Hüftfraktur das wichtigste Thema (95). Die in der hier vorliegenden Studie untersuchten Fragebögen fragten zwar immerhin alle den Aspekt Mobilität ab, definierten ihn aber zu breit oder nicht genau genug. Dies bestätigt auch die Studie von van Leeuwen und Kollegen beim EQ-5D (90).

Ein weiteres Problem bezüglich Verständlichkeit ergibt sich durch die vorgegebene Itemstruktur des SF-12. Im SF-12-Fragebogen beispielsweise ist das Item "viel Energie haben" im mentalen Summenscore (MCS-12) enthalten. Einige Befragte in der vorliegenden Interviewstudie interpretierten das Item "Energie" als körperliche Energie anstelle von psychischem Wohlbefinden. Diese Beobachtung zeigte sich auch bei der Bestimmung der Konstruktvalidität. In einer Studie in einer allgemeinen älteren Population zur Bestimmung der Konstruktvalidität des SF-12 konnte eine zweidimensionale Itemstruktur nicht unterstützt werden, da von Kreuzladungen berichtet wurde (Jakobsson et al., 2012).

Auch bezüglich der Relevanz der Inhalte tauchten Probleme auf. Die Teilnehmenden der hier vorliegenden Studie hatten das Gefühl, dass viele der gegebenen Beispiele zu körperlichen Aktivitäten weit über ihre Möglichkeiten hinausgingen. Eine Studie testete den SF-36 (Langform des SF-12) Fragebogen bei älteren Patienten und zeigte Probleme mit der Skala zur körperlichen Funktionsfähigkeit auf (96). Auch hier wird ein Aktivitätsniveau abgefragt, das weit über die Fähigkeiten der meisten Teilnehmerinnen

und Teilnehmer deren Studie hinausging und daher von den Befragten als irrelevant empfunden wurde. Darüber hinaus empfanden die Befragten der hier vorliegenden Studie insgesamt den Aspekt "Leistungen" für ihren Lebenszustand als irrelevant an. Auch bei der Studie von Baró gaben ungefähr zwei Drittel der Befragten bei mindestens einem Item des NHP "nichtzutreffend" für deren Situation an. Diese als nichtzutreffend empfundenen Items betrafen nur die drei Items zur körperlichen Aktivität (92).

Die Probleme bezüglich der Relevanz betreffen allerdings nicht nur die Inhalte der Items selbst, sondern auch die Antwortkategorien können als unpassend empfunden werden. So wurden im NHP aufgrund der dichotomen Antwortkategorien oft alle Items als irrelevant wahrgenommen. Infolgedessen haben die Befragten die Frage ganz ausgelassen, ‚Ja‘ und ‚Nein‘ angekreuzt oder ein Kreuz in der Mitte gemacht. Dies führt (ohne Interviewer:in) zu einem hohen Anteil an fehlenden Antworten und so zu nicht auswertbaren Fragebögen. In der bereits berichteten Studie von Bureau-Chalot füllten nur 83.5% mindestens 30 der 38 Items aus (91). Diese hohe Anzahl an Missings könnte durch die Ergebnisse in der hier vorliegenden Studie erklärt werden.

Auch bezüglich der Vollständigkeit der verwendeten HRQOL gab es Probleme. Den meisten hier Befragten war klar: eine umfassende Beurteilung mit nur einem Fragebogen ist gar nicht möglich. Die Befragten hegten zwar keinen Anspruch auf Vollständigkeit, dennoch wurden beiläufig einige die HRQOL betreffende fehlende Aspekte, wie Medikamente, Sehbehinderungen oder Verdauungsprobleme berichtet. Andere Studien zeigten bereits, dass eine umfassende Bewertung aller Aspekte von HRQOL unter Verwendung generischer Instrumente nur begrenzt möglich ist. Daher muss ein gewisser Kompromiss zwischen generischen und (Hüftfraktur-) spezifischen Instrumenten eingegangen werden. Um dieser Herausforderung zu begegnen, wurde zur Outcome Messung bei Hüftfrakturpatient:innen eine Kombination verschiedener Instrumente (21, 38) oder die Verwendung von hüftfrakturspezifischen Ergebnismaßen mit einer generischen Komponente (97) empfohlen. Der EQ-5D beispielsweise soll in das core outcome set für Hüftfrakturstudien aufgenommen werden (24). Die Befragten der hier vorliegenden Studie nahmen den EQ-5D teilweise als zu oberflächlich wahr. Eine Kombination verschiedener Instrumente liegt also nahe, führt aber wiederum zu einer höheren Belastung.

Frühere Studien zu anderen psychometrischen Eigenschaften (z.B. Konstruktvalidität) bestätigen die Anwendbarkeit von SF-12 und EQ-5D in dieser Patientengruppe (98-100). Die COSMIN Gruppe gewichtet die Erfüllung der Inhaltsvalidität jedoch als wichtigste Messeigenschaft, so dass die Ergebnisse bisheriger Studien angesichts der hier vorliegenden Ergebnisse möglicherweise neu diskutiert werden müssten.

4.1.3 Stärken und Schwächen

Nach aktuellem Wissensstand ist dies die erste Studie, die die Perspektive von Hüftfrakturpatient:innen auf generische gesundheitsbezogene Lebensqualitätsinstrumente untersucht. Die Stichprobe war in Bezug auf Alter und Geschlecht heterogen, aber es konnten nur Personen mit geringer oder keiner kognitiven Beeinträchtigung befragt werden. Die Ergebnisse können daher nicht einfach auf die gesamte Hüftfrakturpopulation übertragen werden. Wenn jedoch bereits in dieser ausgewählten Stichprobe Probleme in der Machbarkeit und Einschränkungen in der inhaltlichen Validität auftreten, kann davon ausgegangen werden, dass diese Probleme bei Hüftfrakturpatient:innen mit schwereren kognitiven Beeinträchtigungen noch stärker ausgeprägt sind. Darüber hinaus sind einige der in dieser Studie aufgetretenen Fragen zur Inhaltsvalidität möglicherweise auf die deutsche Semantik beschränkt und nicht unbedingt auf alle sprachlichen Versionen der jeweiligen Fragebögen übertragbar. Daher liefert diese Studie nur erste Erkenntnisse darüber, inwiefern der EQ-5D (deskriptives System), SF-12, NHP und WHOQOL-BREF für Hüftfrakturpatient:innen in Bezug auf Machbarkeit und Aspekte der Inhaltsvalidität adäquat oder nicht adäquat sein können. Limitierend kommt hinzu, dass vergleichbare Studien dazu bereits einige Zeit alt sind. Möglicherweise gelten die untersuchten Fragebögen als bereits etabliert und werden nicht mehr in diesem Feld erforscht. Doch die hier durchgeführte Studie liefert auch einen Beitrag zur aktuell häufigen Abwägung zwischen der Verwendung von etablierten/bekanntem Instrumenten mit möglicherweise nicht erfüllten Messeigenschaften und einem hinsichtlich der Messeigenschaften passenderen Instrument, das aber unbekannter ist und einen Vergleich von Outcomes damit erschwert.

Ein weiterer Grund für die geringe Evidenz könnte die sehr schwer zu rekrutierende Gruppe sein. Aus diesem Grund wurde für die hier vorliegende Studie der Zugang über das Krankenhaus gewählt. Möglicherweise wäre aber ein Befragungszeitpunkt von sieben bis 14 Tage nach der Hüftoperation günstiger hinsichtlich der geeigneten

Recall-Periode. Ein möglicher Zugangsweg könnte der häufig anschließende Rehabilitationsaufenthalt sein. Allerdings sollten dann die Patientinnen und Patienten mehrerer Einrichtungen erfasst werden, da sich diese je nach Gesundheitszustand in unterschiedliche Weiterbehandlung (z.B. ambulante Reha) begeben.

Die Anwendung der think-aloud Methode war besonders wertvoll, da sich zeigte, dass viele Befragte ihre Meinung im Anschluss an die Befragung nicht mitteilten und daher wichtige Erkenntnisse nur durch lautes Denken und Beobachtung gewonnen werden konnten. Ein Grund dafür könnte sein: manche Befragte hatten die Sorge, als kognitiv beeinträchtigt eingestuft zu werden, wenn sie zugaben, ein Item nicht verstanden zu haben. Ein weiterer Teil der Patientinnen und Patienten war schlicht nicht in der Lage die Fragen im anschließenden Interview, die sich auf abstrakte Themen bezogen und metakognitive Fähigkeiten erforderten, zu verstehen und zu beantworten. Eine weitere Stärke dieser Studie ist die Konzentration auf den Prozess des Ausfüllens der Fragebögen, so dass z.B. auch fehlende Items zugelassen wurden und dies ein eigenes Ergebnis darstellte. Darüber hinaus ermöglichte das Ausfüllen von bis zu vier Fragebögen einen direkten Vergleich der Fragebögen untereinander. Jedoch war auch die think-aloud Methode für manche Befragte nicht einfach und es fiel ihnen schwer, laut mitzudenken. In diesen Fällen war es notwendig, häufig nachzufragen und/oder auf Beobachtungen zurückzugreifen.

Ein Nebenaspekt der think-aloud Methode ist die Entdeckung der Phänomene des *response shift* oder des *positive editing* bei manchen Befragten. Diese Phänomene stellen kein Problem für die inhaltliche Validität dar, wenn es eine tatsächliche Anpassung der Selbsteinschätzung der HRQOL widerspiegelt. Die Verschiebung der Antworten spiegelt sich in reduzierten Selbsterwartungen wider, die durch einen allmählichen Funktionsverlust und reduzierte soziale Erwartungen entstehen. Sie bezieht sich auf eine Veränderung der Bedeutung der eigenen Selbsteinschätzung eines Zielkonstrukts. (101). In den Interviews der hier durchgeführten Studie tauchte dieses Phänomen bei einigen Fragen auf, z.B. wenn Teilnehmende angaben, dass sie nicht beabsichtigen, etwas zu tun, sondern nur so viel wie möglich. Das Problem des *response shift* mag zwar bei Hüftfrakturpatient:innen verstärkt werden, aber das Problem ist im Allgemeinen bei älteren Menschen bekannt (90, 102). Es könnte durch die kontinuierliche Anpassung an den schlechteren Gesundheitszustand und Abwärtsvergleiche oder andere Bewältigungsmechanismen erklärt werden (103). Es

kann jedoch ein Problem für die inhaltliche Validität darstellen, wenn Befragte das Item nur deshalb positiver bewerten, weil sie nicht als alt oder klagend erscheinen wollen. Hier kann die Rolle des Interviewenden die Antworten möglicherweise beeinflussen.

Alle Interviews wurden von mir als junge, weibliche Forscherin durchgeführt, was die Kommentare und Antworten der Teilnehmenden beeinflusst haben könnte, z.B. in Bezug auf sensitive/heikle Themen (z.B. Zufriedenheit mit dem Sexualleben). Auf der anderen Seite dankten mir die meisten Patientinnen und Patienten ausdrücklich für die Zeit und machten den Eindruck, das Interview genossen zu haben. Für weitere Studien sollte der nicht unbeträchtliche Zeitaufwand mitbedacht werden. So dauerte ein Besuch immer weitaus länger als das aufgezeichnete Interview und es waren häufig mehrere Anläufe nötig (Patient:in schläft, hat Besuch, fühlt sich an diesem Tag nicht gut, usw.).

4.1.4 Schlussfolgerungen

Für die Messung der HQQOL bei Patientinnen und Patienten mit Hüftfrakturen gelten der EQ-5D und der SF-12 als die beste Wahl, da die Befragten diesen Fragebögen die beste Inhaltsvalidität und Machbarkeit zuschreiben. Dennoch müssen sich Gesundheitsfachkräfte und Forscher:innen, die bei Hüftfrakturpatient:innen diese häufig verwendeten generischen HRQOL Instrumente verwenden, der möglichen Einschränkungen bewusst sein. Bei der Verwendung dieser Fragebögen bei Hüftfrakturpatient:innen sollte darauf geachtet werden, dass angemessene Unterstützung angeboten wird (z.B. in Form einer Interviewerin/eines Interviewers) und zusätzlich Aspekte der Mobilität angesprochen werden, die in den generischen HRQOL-Instrumenten nicht ausreichend erfasst werden. Künftige Forschungsarbeiten könnten die Perspektive von Proxys untersuchen und die Perspektive der Patientinnen und Patienten in verschiedenen sprachlichen und kulturellen Kontexten weiter erforschen.

4.2 Quantitative Studie

Die nachfolgende Diskussion fasst als Erstes die Hauptergebnisse der vorliegenden Forschungsarbeit zusammen. Im Anschluss erfolgt eine Einordnung der hier analysierten psychometrischen Messeigenschaften in das Forschungsumfeld der Erhebung von HRQOL bei älteren Hüftfrakturpatient:innen. Insbesondere wird dabei auf Studien eingegangen, die ebenfalls den Einsatz des EQ-5D-5L oder SF-12 sowie damit zusammenhängende (EQ-5D-3L und SF-36) Instrumente auf deren psychometrische

Messeigenschaften in der gleichen oder einer ähnlichen Stichprobe (z.B. ältere Menschen) überprüfen. Nachfolgend werden (methodische) Limitationen der vorliegenden Forschungsarbeit diskutiert und Themen für zukünftige Forschung adressiert.

4.2.1 Zusammenfassung und zentrale Ergebnisse

Für die Messung der HRQOL bei Hüftfrakturpatient:innen sind verschiedene Instrumente im Einsatz, für die es allerdings teilweise nur wenig, widersprüchliche oder keine Evidenz zu deren Messeigenschaften und praktischen Eigenschaften gibt (26, 104). Bei den meisten Studien handelte es sich außerdem um Beobachtungs- oder Interventionsstudien ohne Ziel einer psychometrischen Validierung. Zudem konzentrierten sich die Studien überwiegend auf Aspekte der Validität bzw. insbesondere auf Aspekte der known-groups Validität. In dieser Studie untersuchten wir unter Anwendung der COSMIN Empfehlungen sowohl Validität, Reliabilität als auch Änderungssensitivität des EQ-5D-5L und SF-12 Fragebogens bei einer im Vergleich zu den bisherigen Studien großen Stichprobe.

Unsere Studie zeigt eine gute Machbarkeit beider Fragebögen sowie eine ausreichende Abdeckung des Spektrums der HRQOL der Befragten. Obwohl es Hinweise gibt, dass die Hüftfrakturpatient:innen kränker sind als die Allgemeinbevölkerung (z.B. SF-12 Summenscores), konnten keine Decken- oder Bodeneffekte beobachtet werden. Bezüglich der Validität, konnte anhand der Ergebnisse der CFA die strukturelle Validität nur für den EQ-5D-5L bestätigt werden. Darüber hinaus konnte ebenfalls nur dem EQ-5D-5L eine gute Konvergenz- und Diskriminanzvalidität zugeschrieben werden, wohingegen der SF-12 eine akzeptable Konvergenz- und Diskriminanzvalidität aufwies. Außerdem deutet diese Studie darauf hin, dass die known-groups Validität beider Instrumente nur im akzeptablen Bereich liegt. Die Reliabilität betreffend, konnte für beide Skalen des SF-12 interne Konsistenz gezeigt werden. Des Weiteren sprechen die Ergebnisse für eine eher geringe Änderungssensitivität der beiden Instrumente.

Diese Ergebnisse tragen dazu bei, mehr Evidenz zu den psychometrischen Messeigenschaften des EQ-5D-5L und des SF-12 bei älteren Hüftfrakturpatient:innen zu generieren und bestätigen größtenteils die bisherigen Ergebnisse (26, 104-106).

4.2.2 Einordnung in die bestehende Forschung

EQ-5D

Direkt vergleichbare Literatur ist spärlich, die Publikationen liegen bereits einige Zeit zurück: Eine Studie von Parsons et al. (100) untersuchte die Messeigenschaften von drei verschiedenen Instrumenten (u.a. OHS und EQ-5D-3L) in einer großen (n=225) Hüftfrakturkohorte und konnte eine gute Änderungssensitivität des EQ-5D zeigen, die fast mit der Änderungssensitivität des hüftspezifischen Instruments (OHS) übereinstimmte. Ebenfalls zeigte eine Studie von Tidermark und Kollegen (99) aus dem Jahr 2003 mit 110 Teilnehmenden eine hohe Änderungssensitivität für den EQ-5D-3L. Unsere Ergebnisse konnten diese Aussagen nicht unterstützen.

Laut einer aktuelleren systematischen Übersichtsarbeit, die erstmals international anerkannte Kriterien zur Beurteilung der Qualität und Akzeptanz von PROMs bei älteren Hüftfrakturpatient:innen anwendete, waren der EQ-5D-3L und der SF-36 die am häufigsten evaluierten Instrumente (26). Als Ergebnis wurde für den EQ-5D-3L moderate Evidenz für Validität (hauptsächlich known-groups Validität) und Änderungssensitivität berichtet. In der hier vorliegenden Studie konnte Evidenz für eine gute Konvergenz- und Diskriminanzvalidität sowie akzeptable known-groups Validität generiert werden.

Kürzlich berichtete eine systematische Übersichtsarbeit über die psychometrischen Messeigenschaften des EQ-5D-5L in unterschiedlichen Populationen, worunter muskuloskeletale/orthopädische Probleme und Krebs (jeweils acht Studien) am häufigsten untersucht wurden (39). Der Review schlussfolgert, dass der EQ-5D-5L hervorragende psychometrische Messeigenschaften für eine Vielzahl an Bevölkerungsgruppen aufweist, jedoch eine gründliche Erforschung der Änderungssensitivität notwendig ist. Unsere Studie kann die hervorragenden psychometrischen Messeigenschaften nicht bestätigen. Im direkten Vergleich zwischen den beiden hier untersuchten Instrumenten schneidet der EQ-5D-5L aber besser ab.

Eine ebenfalls aktuelle Studie berichtet anhand von Daten aus einer Population älterer Menschen (n=1690) über den ersten direkten empirischen Vergleich zwischen dem Barthel-Index und dem EQ-5D-3L (107). Die Autoren konnten diskriminative Validität für beide Instrumente zeigen und schlussfolgerten dadurch, dass beide Instrumente für den Einsatz in der älteren Bevölkerung geeignet sind. In unserer Studie wurde der

Barthel-Index als Referenzmaß zur Ermittlung der Konvergenzvalidität verwendet. Übereinstimmend mit der Hypothese zeigte sich eine Korrelation von $r_s=0.66$ zwischen dem EQ-5D-5L Index und dem Barthel Score, wodurch die Konvergenzvalidität des EQ-5D-5L unterstützt wird.

Vergleicht man die Ergebnisse bzgl. der Machbarkeit mit den Erfahrungen aus der ebenfalls im Rahmen dieser Dissertation durchgeführten qualitativen Studie, zeigen sich Parallelen. In beiden Studien mussten die Teilnehmenden beim Ausfüllen teilweise von einer Studienmitarbeiterin oder einem Studienmitarbeiter unterstützt werden. Die Übertragung der Ergebnisse bzgl. der Machbarkeit ist also nur eingeschränkt unter Berücksichtigung dieser notwendigen Hilfestellung möglich.

SF-12

Ähnlich dem EQ-5D, ist auch die Literatur zu den psychometrischen Messeigenschaften des SF-12 bereits älter. Zumeist wird der SF-36 erforscht. So berichtet die bereits beschriebene systematische Übersichtsarbeit von Haywood et al. über moderate bis starke Evidenz für Validität und Änderungssensitivität für den SF-36 und begrenzte Evidenz für den SF-12 bei älteren Hüftfrakturpatient:innen (26). Die hier vorliegende Arbeit trägt dazu bei, diese Forschungslücke zu schließen. Allerdings konnte mit dieser Studie keine weitere Evidenz für die strukturelle Validität des SF-12 generiert werden, denn die Ergebnisse der CFA lieferten keine eindeutigen Belege für die von den Entwicklern des Fragebogens vorgeschlagene 2-Faktor Struktur. Unsere Studie untersuchte allerdings nicht das Auftreten neuer Faktorstrukturen mit Hilfe einer explorativen Faktorenanalyse. Vielmehr wurde in dieser Studie bewertet, wie gut die Daten aus einer Hüftfrakturpopulation zur bestehenden Instrumentenstruktur passen. Eine Studie von Jakobsson et al. 2012, die die Konstruktvalidität des SF-12 bei älteren Menschen, Menschen mit Parkinson und Schlaganfallüberlebenden überprüfte, fanden ebenfalls eine fehlende Übereinstimmung zwischen den empirischen Daten und dem vorgeschlagenen Modell in allen drei Stichproben (108).

Betrachtet man die Konvergenz- und Diskriminanzvalidität, so zeigten sich Unterschiede innerhalb der beiden Summenskalen. Die PCS-12 zeigte gute Konvergenz mit anderen Instrumenten, die ähnliche Konstrukte messen und wies geringere Zusammenhänge mit unähnlichen Konstrukten auf, so dass die meisten a-priori Hypothesen bestätigt werden konnten. Die Eigenschaften der MCS-12 stimmten jedoch häufig nicht mit den aufgestellten Hypothesen überein, so dass die

Konvergenz- und Diskriminanzvalidität der MCS-12 nur als akzeptabel bewertet wurde. Insgesamt wurde so die Konvergenz- und Diskriminanzvalidität des SF-12 ebenfalls als akzeptabel bewertet. Es gibt jedoch auch Studien, die nur eine der beiden Summenskalen verwenden (109, 110). In diesem Fall wäre das Ergebnis der jeweiligen Skala relevant und das schlechtere Abschneiden der MCS-12 könnte beispielsweise berücksichtigt werden.

Die Ergebnisse der CFA und der Korrelationsanalysen könnten darauf hinweisen, dass nicht alle Fragen so wahrgenommen wurden, wie intendiert. Dies würde den Ergebnissen der ebenfalls im Rahmen der Dissertation durchgeführten qualitativen Studie entsprechen. Außerdem diskutierte auch eine Studie, die den SF-12 (V2) bei Diabetespatient:innen validierte, dass es möglicherweise Schwächen in der Abbildung der geistigen Gesundheit gibt und die entsprechenden Items zur Abfrage der geistigen Gesundheit nicht adäquat sind (111). Sie erklärten damit die ebenfalls unerwarteten Zusammenhänge mit anderen Konstrukten.

Unsere Studie zeigte bei der Überprüfung der known-groups Validität nur einzelne signifikante Unterschiede in die erwartete Richtung. Die systematische Übersichtsarbeit unterstützt diese Aussage, denn die meisten der insgesamt 22 berichteten Studien zum SF-36 bzw. SF-12 zeigten eine angemessene (n=13) oder geringe (n=5) known-groups Validität für den SF-36 bzw. SF-12, wobei zwei Studien auch über eine gute known-groups Validität berichteten (26).

Zur Eigenschaft Änderungssensitivität gibt es nur eine Studie aus dem Jahr 2003, die zur Änderungssensitivität des EQ-5D und des SF-36 bei älteren Hüftfrakturpatient:innen (99) forschte. Diese Studie berichtete eine gute Änderungssensitivität für den SF-36, die unsere Studie nicht bestätigen konnte.

4.2.3 Limitationen

Im Rahmen einer erweiterten Responseanalyse innerhalb der ProFem Studie wurden die befragten mit den übrigen Versicherten verglichen, die den Krankenhausaufenthalt überlebt hatten (insgesamt 5096 Versicherte). Die Analyse zeigte u.a., dass die an der Befragung eigenständig Teilnehmenden im Vergleich zu den übrigen Versicherten jünger waren und vor der PFF seltener einen hohen Pflegegrad hatte, wohingegen sich beispielsweise keine Unterschiede hinsichtlich des Geschlechts zeigten.

Allerdings konnten durch die Möglichkeit der Verwendung einer Kurzversion des Fragebogens sowie durch die stellvertretende Beantwortung durch einen Proxy auch Information zu Teilnehmenden gewonnen werden, die sonst nur schwer für Studien gewonnen werden können. Außerdem kann man davon ausgehen, dass die Ergebnisse zu den Messeigenschaften auf die nicht befragte Gruppe übertragen werden können, da keine Bodeneffekte vorhanden waren. Eine weitere Stärke ist die Stichprobengröße: für eine ausreichend große Stichprobe werden mindestens 200 Fälle als essentiell beschrieben (112). Außerdem ist die Auswahl der Studienteilnehmenden wertvoll, da sie sich nicht wie häufig in anderen Studien auf die Rekrutierung in einem einzigen Zentrum stützte (z.B. Studie von Parsons et al. 2014).

Bei der Erfassung der Gütekriterien der Zustandsmessung gibt es einige methodische Schwächen, die bei der Interpretation der Ergebnisse berücksichtigt werden sollten. So empfiehlt COSMIN eigentlich, psychometrische Messeigenschaften nicht in einer Studie zu messen, deren Ziel die Evaluation der Ergebnisqualität ist. Diese Empfehlung ist aber vor allem in der hier untersuchten Gruppe der älteren Hüftfrakturpatient:innen nur schwer umzusetzen und würde die doppelte Belastung der Befragten möglicherweise nicht rechtfertigen. Des Weiteren sollten bei der Ermittlung von Cronbachs α alle Fragen in die gleiche Richtung kodiert sein und die Items sollten die gleichen Antwortformate haben. Beim SF-12 variieren die Antworten aber von ja/nein bis zu einer 6-stufigen Likertskala.

Zudem muss diskutiert werden, inwieweit bei dieser Studienpopulation das Alter als known-group zur Validitätsbewertung verwendet werden kann. In unserer Stichprobe zeigt sich nicht der erwartete Unterschied, dass ältere Teilnehmende schlechtere HRQOL angeben. Auch bei einer aktuellen prospektiven Kohortenstudie, die verschiedene potentiell prädiktive Faktoren unter anderem auf die HRQOL untersuchte, hatte Alter keinen signifikanten Einfluss auf die HRQOL (gemessen mit dem EQ-5D-5L) (113).

Bisherige Studien zur Überprüfung der Änderungssensitivität sind unter anderem bezüglich des verwendeten Ankers heterogen. In unserer Studie wurden zwei häufig verwendete Ansätze gewählt, wobei für den Kriteriums-basierten Ansatz eine Ankerfrage in die Follow-up Erhebung integriert wurde. COSMIN empfiehlt zwei verschiedene GRCS Fragen, wenn es zwei Subskalen gibt. Möglicherweise ist die integrierte Ankerfrage somit nicht gemeinsam passend für die zwei Subskalen des SF-

12 und den EQ-5D. Außerdem sollte bei der Bewertung der Änderungssensitivität bedacht werden, dass möglicherweise die ursprüngliche Hypothese einer fortschreitenden Verbesserung der HRQOL bei einer Hüftfrakturpopulation nicht unbedingt gegeben ist. Eine Studie, die anhand von Bewegungsmessern den Rehabilitationsprozess von Hüftfrakturpatientinnen und Hüftfrakturpatienten explorierte, zeigt, dass es drei Cluster zu geben scheint: Patienten und Patientinnen mit geringem, mittlerem und starkem Prozess im Verlauf (114). Auch in der ProFem Studienpopulation zeigen sich nach ersten Einschätzungen ganz unterschiedliche Verläufe der HRQOL über die Befragungszeitpunkte hinweg. Außerdem fällt bei der Betrachtung der Änderungssensitivität speziell bei der MCS-12 auf, dass sich der Median in der stark verbesserten Gruppe zwischen den Zeitpunkten zwar nicht signifikant unterscheidet, der Summenscore im Median bei beiden Zeitpunkten aber verglichen mit allen anderen Gruppen der höchste ist. Es sollte also immer bedacht werden, inwieweit ein Instrument Veränderungen messen können soll, wenn es möglicherweise gar keine Veränderung zu messen gibt. Um dieser Problematik etwas entgegen zu wirken, wurden in dieser Studie zwei unterschiedliche Ansätze zur Überprüfung der Änderungssensitivität gewählt, die beide zum gleichen Ergebnis kamen.

Insgesamt konnten in unserer Studie erstmals vielfältige Messeigenschaften anhand einer Stichprobe überprüft werden. Außerdem gibt es nach aktuellem Stand keine Evidenz zu den psychometrischen Messeigenschaften der Version 2.0 des SF-12 und der 5L Version des EQ-5D in einer älteren Hüftfrakturpopulation.

4.2.4 Anwendbarkeit und zukünftige Forschung

Bereits zitierte Studien schlussfolgern, dass in zukünftigen Studien am besten eine Kombination aus z.B. generischen und spezifischen HRQOL Instrumenten verwendet werden soll (21, 24, 38). Die Ergebnisse dieser Studie können diese Aussage stützen, denn der SF-12 und EQ-5D-5L weisen zwar nicht durchgängig gute, aber akzeptable psychometrischen Messeigenschaften auf und sind machbar. Die generischen Fragebögen haben den großen Vorteil, dass sie bei verschiedenen Krankheitsbildern zur Abfrage der HRQOL eingesetzt werden können und auch bereits schon vielfach eingesetzt wurden. Dies macht einen Vergleich mit anderen Erkrankungen oder z.B. die Untersuchung einer allgemeinen älteren Population mit einem gemeinsamen Fragebogen möglich. Griffiths et al. (2015) schlussfolgert, dass ein generisches

Instrument auch deshalb sinnvoll ist, weil es kein spezifisches Instrument gibt, dass alle für die Patientinnen und Patienten relevante Bereiche abdeckt (95).

Insgesamt ist die Analyse psychometrischer Messeigenschaften ein fortlaufender Prozess, so dass mehr Evidenz größeres Vertrauen in ein HRQOL Instrument bedeutet (112). Dementsprechend benötigt es noch weitere Forschung, um größere Sicherheit bei der Verwendung des SF-12 und EQ-5D-5L zu erhalten:

Zukünftige Validierungsstudien sollten die Schätzung von MID-Werten (= kleinster bedeutsamer Unterschied) [engl. minimal important difference]) und MIC-Werten (= kleinste bedeutsame Veränderung) [engl. minimal important change]) beinhalten. Die MID wird als der geringste Unterschied definiert, der zu einer Änderung in der Behandlung führen würde (115), während der MIC als der geringste Unterschied über die Zeit definiert ist, der von der Patientin oder dem Patient als relevant angesehen wird (116). Diese Werte für jede Skala/jedes Instrument zu haben wäre also hilfreich bei der Interpretation von Veränderungen. Es gibt allerdings keine Goldstandard-Methode zur Schätzung von MID oder MIC, sondern es werden verschiedene Ansätze analog zur Messung der Änderungssensitivität verwendet.

Außerdem sollten weitere Studien die Test-Retest-Reliabilität der beiden Instrumente überprüfen. Neben der Änderungssensitivität könnte prinzipiell bei der Gruppe, die in der GRCS keine Veränderung angab, auch die Test-Retest-Reliabilität überprüft werden. Die Überprüfung der Test-Retest-Reliabilität wird als angemessen angesehen, wenn innerhalb des Messzeitraums keine signifikanten Veränderungen zu erwarten sind (36). In der hier vorliegenden Studie wurde vermutet, dass zwischen den Messzeitpunkten bzw. mit weiterem Abstand zum Hüftfrakturereignis eine Veränderung bzw. Verbesserung eintritt. Zu bedenken ist zudem, dass sich die Erhebungsarten zwischen den beiden Zeitpunkten unterschieden (persönlich vs. postalisch). Als weiterer möglicher problematischer Punkt kann der Zeitraum zwischen den Messzeitpunkten angesehen werden. Es gibt keine Standardregel für einen optimalen Zeitraum zwischen den beiden Messzeitpunkten, jedoch gibt COSMIN als grobe Richtlinie zwei Wochen an (36). Aus diesen Gründen wurde in der hier vorliegenden Arbeit von der Bestimmung der Test-Retest-Reliabilität abgesehen.

4.3. Ansatzpunkte zur Ableitung von Empfehlungen für den Einsatz in der klinischen Forschung

Eine Möglichkeit die Ergebnisse einzuordnen und in Richtung einer Empfehlung zu kommen, ist, die relative Bedeutung psychometrischer Messeigenschaften zu betrachten (117): Laut COSMIN gilt die Inhaltsvalidität als wichtigste Messeigenschaft. Es wird betont, dass als Erstes klar sein sollte, dass alle Inhalte eines Instrumentes bezogen auf das interessierende Konstrukt und die Zielpopulation relevant, umfassend und verständlich sind. Die qualitative Studie dieser Arbeit konnte relativ zu den anderen untersuchten Instrumenten für den SF-12 und den EQ-5D die beste Inhaltsvalidität zeigen. Jedoch wiesen auch die beiden Fragebögen noch Schwächen auf, die bei der Verwendung in einer klinischen Studie berücksichtigt werden sollten. Beispielsweise empfanden einige Teilnehmende der qualitativen Studie den EQ-5D als zu oberflächlich, so dass eine Studie mit dem primären Outcome Lebensqualität möglicherweise einen umfassenderen Fragebogen wählen sollte. Für die Messung der HRQOL als sekundäres Outcome wäre der EQ-5D-5L aber aufgrund seiner guten Machbarkeit hervorragend geeignet.

Als zweites führt COSMIN die interne Struktur eines Instrumentes als wichtig an, was zum Beispiel die Messeigenschaften strukturelle Validität und interne Konsistenz beinhaltet. Die interne Struktur gibt an, wie beispielsweise die verschiedenen Items in einem Instrument zusammenhängen, was wichtig ist, um zu entscheiden wie die Items z.B. in einer Subskala kombiniert werden können. In dieser Arbeit wurde keine interne Konsistenz für den EQ-5D berechnet, da die Items des EQ-5D-5L als formative Indikatoren konzipiert wurden und in diesem Fall die Anwendung von Cronbach's α die wahre Reliabilität unterschätzen würde (118). Durch die CFA konnte allerdings die strukturelle Validität des EQ-5D bewertet werden, wodurch auch diese „zweite Stufe“ innerhalb der relativen Wichtigkeiten der Messeigenschaften beurteilt werden konnte.

Zuletzt werden die übrigen Messeigenschaften, wie zum Beispiel Hypothesentestung zur Überprüfung der Konstruktvalidität und die Änderungssensitivität als wichtig erachtet.

Ein weiterer Punkt, der bei der Entwicklung einer Empfehlung unbedingt berücksichtigt werden sollte, ist der Erhebungskontext. Unabhängig von den psychometrischen Messeigenschaften eines Fragebogens war in den hier durchgeführten Studien die

Unterstützung durch eine Interviewerin/einen Interviewer für viele Teilnehmende relevant, um eine gute und vollständige Datenqualität zu erhalten. Ebenfalls wurden die Teilnehmenden entweder direkt in der Klinik aufgesucht oder innerhalb der ProFem Studie direkt zu Hause besucht. Für beide Studien mussten die Teilnehmenden keine Wege auf sich nehmen. Auch dieser Kontext sollte für zukünftige Erhebungen mit den Instrumenten bedacht werden und bei Empfehlungen mit einfließen.

Es wäre nun wichtig, die gesamte vorliegende Evidenz aus der Literatur und unseren Studien in einen Konsensfindungsprozess einzubringen. In dem Konsensprozess wird bewertet, wie die einzelnen Instrumente anhand der Ergebnisse beurteilt werden können. Dafür können verschiedene Methoden wie e-Delphis, Surveys oder Konsenskonferenzen zum Einsatz kommen (119, 120). Gegenüber weniger strukturierten Methoden wie z.B. Diskussionen am runden Tisch, hat dieser Ansatz einen Vorteil hinsichtlich der Validität der finalen Konsensentscheidung. Beispielsweise dadurch, dass die Teilnehmenden einer Delphi-Studie nicht direkt miteinander interagieren, so dass nicht ein Teilnehmer oder eine Teilnehmerin dominieren und die Ansicht einer ganzen Gruppe beeinflussen kann (120). Bedeutsam ist der Einbezug aller relevanten Stakeholder, d.h. Patient:innen, Angehörige/Proxies, Kliniker und methodisch arbeitende Wissenschaftler:innen (121, 122). Diese gemeinsame Konsensfindung kann zur Vermeidung von unangemessenen oder uneinheitlichen PROMs beitragen und helfen, Reporting-Bias zu verhindern (120).

5. Zusammenfassung in deutscher und englischer Sprache

Hintergrund und Ziele

Für die Messung der gesundheitsbezogenen Lebensqualität (HRQOL) bei Patientinnen und Patienten nach proximaler Femurfraktur (kolloquial Hüftfraktur) sind verschiedene Instrumente im Einsatz, für die es allerdings teilweise nur wenig, widersprüchliche oder keine Evidenz zu deren psychometrischen Messeigenschaften und praktischen Eigenschaften gibt. In dieser Studie untersuchten wir unter Anwendung der COSMIN Empfehlungen a) die Inhaltsvalidität und Machbarkeit von vier generischen Instrumenten zur Messung der HRQOL (SF-12, EQ-5D-5L deskriptives System, Nottingham Health Profile/NHP, WHOQOL-BREF) bei älteren Menschen (≥ 60 Jahre) nach einer Hüftfraktur und b) die Validität, Reliabilität und Änderungssensitivität der zwei am häufigsten in dieser Zielgruppe verwendeten Instrumente (SF-12, EQ-5D) zur Messung der generischen HRQOL.

Methoden

Für die qualitative Studie (a) rekrutierten wir 32 Patientinnen und Patienten ohne relevante kognitive Beeinträchtigung im Alter von über 60 Jahren während ihres stationären Aufenthaltes nach einer Hüftfraktur (weiblich= 25; Durchschnittsalter= 81 Jahre). Die Patientinnen und Patienten wurden gebeten, bis zu vier HRQOL-Fragebögen auszufüllen. Während die Patientinnen und Patienten die Fragebögen ausfüllten, wurde die think-aloud Technik angewandt, um Probleme bei der Beantwortung zu erkennen. Im Anschluss an die Beantwortung der Fragebögen, wurden die Teilnehmenden mit Hilfe eines semi-strukturierten Interviewleitfadens befragt. Die abgefragten Themen beinhalteten Aspekte der Inhaltsvalidität (Verständlichkeit, Relevanz und Vollständigkeit) und Machbarkeit. Die Interviews wurden aufgezeichnet, verbatim transkribiert und inhaltsanalytisch ausgewertet.

Für die quantitative Studie (b) wurden Daten aus einer bevölkerungsbezogenen prospektiven Beobachtungsstudie namens ProFem ausgewertet, die zwischen Januar 2018 und September 2019 konsekutiv Versicherte mit einer proximalen Femurfraktur (PFF) eingeschlossen hat. Für die psychometrische Analyse wurden Primärdaten von Studienteilnehmenden verwendet, die zu den Zeitpunkten T0 (3 Monate nach Krankenhausentlassung) und T1 (6 Monate nach Krankenhausentlassung)

eigenständig teilnahmen. Die Gütekriterien zur Zustandsmessung wurden folgendermaßen evaluiert: Machbarkeit und Interpretierbarkeit wurden anhand von fehlenden Werten sowie Decken- und Bodeneffekten beurteilt; strukturelle Validität anhand einer konfirmatorischen Faktorenanalyse (CFA); Konstruktvalidität anhand des known-groups Ansatzes; Diskriminanz- und Konvergenzvalidität anhand von Spearmans Rho (ρ) Korrelationen mit Referenzmaßen; Reliabilität unter Verwendung der internen Konsistenz anhand Cronbachs Alpha (α) (nur für den SF-12). Die Änderungssensitivität als Gütekriterium der Veränderungsmessung wurde mittels eines Konstrukt- und Kriteriums-basierten Ansatzes ermittelt.

Ergebnisse

Die Ergebnisse der qualitativen Studie (a) zeigten, dass die Teilnehmenden kurze Fragebögen mit einem klaren Layout favorisierten. Nur neun Teilnehmende waren in der Lage, die Fragebögen selbst auszufüllen und nur sechs füllten alle vier Fragebögen aus. Bezüglich der Inhaltsvalidität gab es die größten Probleme bei der Verständlichkeit der Items. Items, die zu vage formuliert waren (häufig beim NHP) und Items ohne Beispiel (beim WHOQOL-BREF) waren schwierig zu verstehen. Ein aus Sicht der Teilnehmenden irrelevantes Beispiel (SF-12 und EQ-5D-5L deskriptives System) oder nicht ausreichende dichotome Antwortkategorien (NHP) begründeten eine fehlende Relevanz desjenigen Items. Das deskriptive System des EQ-5D-5L wurde teilweise als zu oberflächlich und damit nicht als vollständig empfunden. Insgesamt umfassten die Instrumente keine der folgenden von den Teilnehmenden als wichtig für deren HRQOL erachtete Themen: Medikamenteneinnahme, Sehschwierigkeiten und Ernährung.

Die Studienteilnehmerinnen und -teilnehmer der quantitativen Studie (b) waren zur Baseline ($n= 306$) im Durchschnitt 79.1 Jahre alt (69% weiblich). Die anhand der Baseline untersuchten Gütekriterien der Zustandsmessung ergaben folgende Ergebnisse: Beide Instrumente bewiesen eine hohe Machbarkeit und zeigten weder Decken- noch Bodeneffekte. Die CFA zur Beurteilung der strukturellen Validität ergab nur beim EQ-5D-5L einen akzeptablen Modellfit. Bei der Überprüfung der known-groups Validität konnten nur die Unterschiede anhand des subjektiven sozialen Status anhand beider Instrumente signifikant unterschieden werden, was auf eine nur akzeptable known-groups Validität beider Instrumente schließen lässt.

Anhand der ermittelten ρ -Korrelationen ergab sich eine gute Konvergenz- und Diskriminanzvalidität für den EQ-5D-5L und eine akzeptable für den SF-12. Bezüglich der Reliabilität kann für beide Komponenten des SF-12 eine konsistente Skale angenommen werden. Die Ergebnisse sprachen bei beiden verwendeten Ansätzen und für beide Instrumente für eine geringe Änderungssensitivität.

Schlussfolgerung

Insgesamt können der EQ-5D-5L und der SF-12 zur Messung der generischen HRQOL bei älteren Hüftfrakturpatient:innen empfohlen werden. Im direkten Vergleich mit dem WHOQOL-BREF und NHP schnitten die beiden Fragebögen bei der Bewertung von Machbarkeit und Inhaltsvalidität am besten ab. Die beiden Instrumente zeigten allerdings Schwächen bei der Überprüfung der weiteren psychometrischen Messeigenschaften, die aber trotzdem als akzeptabel eingestuft wurden. Weitere Studien sollten bedenken, dass sich die Messung der HRQOL anhand eines Fragebogens zum Selbstausfüllen als herausfordernd darstellte, selbst wenn keine oder nur eine leichte kognitive Beeinträchtigung bei den Teilnehmenden vorhanden war. Die Ergebnisse dieser Arbeit können gemeinsam mit der vorliegenden Evidenz aus der Literatur den Konsensusprozess informieren, ein geeignetes Instrument für die Messung der generischen HRQOL bei älteren Hüftfrakturpatient:innen festzulegen. Zukünftige Validierungsstudien sollten die Test-Retest-Reliabilität des SF-12 und EQ-5D erforschen sowie die Schätzung von MID- und MIC-Werten (kleinster bedeutsamer Unterschied bzw. kleinste bedeutsame Veränderung) beinhalten.

Background and aims

Various instruments are used to measure health-related quality of life (HRQOL) in patients after proximal femur fracture (colloquial hip fracture), but some of them have little, contradictory or no evidence on their psychometric measurement properties and on feasibility aspects. Using the COSMIN recommendations, this study investigated a) the content validity and feasibility of four generic instruments for measuring HRQOL (SF-12, EQ-5D-5L descriptive system, NHP, WHOQOL-BREF) in older people (≥ 60 years) after hip fracture and b) the validity, reliability and change sensitivity of the two most commonly used instruments (SF-12, EQ-5D) for measuring generic HRQOL in this target group.

Methods

For the qualitative study (a), we recruited 32 patients without relevant cognitive impairment aged over 60 years during their inpatient stay after hip fracture (female=25; mean age= 81 years). Patients were asked to complete up to four HRQOL questionnaires. While patients were completing the questionnaires, the think-aloud technique was used to identify problems in answering. After completing the questionnaires, participants were interviewed using a semi-structured interview guide. The topics asked included aspects of content validity (comprehensibility, relevance and completeness) and feasibility. The interviews were recorded, verbatim transcribed and subjected to content analysis.

For the quantitative study (b), we analysed data from a population-based prospective observational study called ProFem, which consecutively included insured persons with a proximal femur fracture (PFF) between January 2018 and September 2019. Primary data from study participants who participated on their own at time points T0 (3 months after hospital discharge) and T1 (6 months after hospital discharge) were used for the psychometric analysis. The psychometric properties were evaluated as follows: Feasibility and interpretability were assessed using missing values and ceiling and floor effects; structural validity using confirmatory factor analysis (CFA); construct validity using the known-groups approach; discriminant and convergent validity using Spearman's Rho (ρ) correlations with reference measures; reliability using internal consistency by means of Cronbach's alpha (α) (only for the SF-12).

Responsiveness, as a psychometric property in a longitudinal context, was determined using a construct- and criterion-based approach.

Results

The results of the qualitative study (a) showed that participants favored short questionnaires with a clear layout. Only nine participants were able to complete the questionnaires themselves and only six completed all four questionnaires. In terms of content validity, the problems were most prevalent in the comprehensibility of the items. Items that were formulated too vaguely (often in the NHP) and items without an example (in the WHOQOL-BREF) were difficult to understand. An example that was irrelevant from the participants' point of view (SF-12 and EQ-5D-5L descriptive system) or insufficient dichotomous response categories (NHP) caused a lack of relevance of that item. The EQ-5D-5L descriptive system was partly perceived as too superficial and thus not complete. Overall, none of the following topics considered important by the participants for their HRQOL were included in the instruments: Medication use, visual difficulties and nutrition.

The study participants of the quantitative study (b) were on average 79.1 years old (69% female) at baseline (n= 306). Both instruments (SF-12 and EQ-5D-5L) demonstrated high feasibility and showed neither ceiling nor floor effects. The CFA to assess structural validity showed an acceptable model fit only for the EQ-5D-5L. When testing the known-groups validity, only the differences based on subjective social status could be significantly distinguished using both instruments, which indicated an only acceptable known-groups validity for both instruments. Based on the determined ρ -correlations, a good convergence and discriminant validity was found for the EQ-5D-5L and an acceptable one for the SF-12. Regarding reliability, a consistent scale can be assumed for both components of the SF-12. The results indicated poor responsiveness for both instruments using both approaches.

Conclusion

Overall, the EQ-5D-5L and the SF-12 can be recommended for measuring generic HRQOL in older hip fracture patients. In direct comparison with the WHOQOL-BREF and NHP, the two questionnaires performed better in terms of feasibility and content validity. However, the two instruments showed weaknesses in examining the other psychometric measurement properties, but these were still rated as acceptable. Future

studies should consider that measuring HRQOL using a self-completion questionnaire proved to be challenging even when there was no or only mild cognitive impairment among the participants. The results of this work, together with the available evidence from the literature, can inform the consensus process to determine an appropriate instrument for measuring generic HRQOL in older hip fracture patients. Future validation studies should explore test-retest reliability of the SF-12 and EQ-5D-5L and include estimation of MID and MIC (minimal important difference and minimal important change) scores, respectively.

6. Literaturverzeichnis

1. Ballane G, Cauley JA, Luckey MM, Fuleihan GEH. Secular trends in hip fractures worldwide: opposing trends East versus West. *Journal of bone and mineral research*. 2014;29(8):1745-55.
2. Icks A, Arend W, Becker C, Rapp K, Jungbluth P, Haastert B. Incidence of hip fractures in Germany, 1995-2010. *Archives of osteoporosis*. 2013;8:140.
3. Balk R, Hahn F, Tarcea B. Die proximale Femurfraktur. *OP-JOURNAL*. 2002;18(02):80-4.
4. Hoffmann F, Glaeske G. [Incidence of hip fracture in Germany--person-related analysis of health insurance population]. *Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany))*. 2006;68(3):161-4.
5. Dreinhofer K. Hip Fracture in the Elderly: Time to Act. *Deutsches Arzteblatt international*. 2020;117(4):51-2.
6. Wehren LE, Magaziner J. Hip fracture: Risk factors and outcomes. *Current Osteoporosis Reports*. 2003;1(2):78-85.
7. Konnopka A, Jerusel N, König HH. The health and economic consequences of osteopenia- and osteoporosis-attributable hip fractures in Germany: estimation for 2002 and projection until 2050. *Osteoporosis International*. 2009;20(7):1117-29.
8. Brunner LC, Eshilian-Oates L, Kuo TY. Hip fractures in adults. *Am Fam Physician*. 2003;67(3):537-42.
9. Melton LJ, 3rd. Who has osteoporosis? A conflict between clinical and public health perspectives. *J Bone Miner Res*. 2000;15(12):2309-14.
10. Wolinsky FD, Fitzgerald JF, Stump TE. The effect of hip fracture on mortality, hospitalization, and functional status: a prospective study. *Am J Public Health*. 1997;87(3):398-403.
11. Bentler SE, Liu L, Oribizan M, Cook EA, Wright KB, Geweke JF, et al. The aftermath of hip fracture: discharge placement, functional status change, and mortality. *Am J Epidemiol*. 2009;170(10):1290-9.

12. Mariconda M, Costa GG, Cerbasi S, Recano P, Aitanti E, Gambacorta M, et al. The determinants of mortality and morbidity during the year following fracture of the hip: a prospective study. *The bone & joint journal*. 2015;97-b(3):383-90.
13. Lunde A, Tell GS, Pedersen AB, Scheike TH, Apalset EM, Ehrenstein V, et al. The Role of Comorbidity in Mortality After Hip Fracture: A Nationwide Norwegian Study of 38,126 Women With Hip Fracture Matched to a General-Population Comparison Cohort. *Am J Epidemiol*. 2019;188(2):398-407.
14. Andrich S, Ritschel M, Meyer G, Hoffmann F, Stephan A, Baltes M, et al. Healthcare provision, functional ability and quality of life after proximal femoral fracture - 'ProFem': Study protocol of a population-based, prospective study based on individually linked survey and statutory health insurance data. *BMJ open*. 2019;9(6):e028144.
15. Smektala R, Ohmann C, Paech S, Neuhaus E, Rieger M, Schwabe W, et al. Zur Prognose der Schenkelhalsfraktur. *Der Unfallchirurg*. 2005;108(11):927-37.
16. Rapp K, Becker C, Todd C, Rothenbacher D, Schulz C, König HH, et al. The Association Between Orthogeriatric Co-Management and Mortality Following Hip Fracture. *Deutsches Arzteblatt international*. 2020;117(4):53-9.
17. Griffin XL, Parsons N, Achten J, Fernandez M, Costa ML. Recovery of health-related quality of life in a United Kingdom hip fracture population. The Warwick Hip Trauma Evaluation--a prospective cohort study. *The bone & joint journal*. 2015;97-B(3):372-82.
18. Martín-Martín LM, Arroyo-Morales M, Sánchez-Cruz JJ, Valenza-Demet G, Valenza MC, Jiménez-Moleón JJ. Factors influencing performance-oriented mobility after hip fracture. *Journal of aging and health*. 2015;27(5):827-42.
19. Stenvall M, Olofsson B, Nyberg L, Lundström M, Gustafson Y. Improved performance in activities of daily living and mobility after a multidisciplinary postoperative rehabilitation in older people with femoral neck fracture: a randomized controlled trial with 1-year follow-up. *Journal of rehabilitation medicine*. 2007;39(3):232-8.
20. Ohmann C, Smektala R, Pientka L, Paech S, Neuhaus E, Rieger M, et al. Ein neues Modell der sektorübergreifenden Datenzusammenführung und Evaluation am Beispiel der Schenkelhalsfraktur. *Z ärztl Fortbild Qual Gesundheitswes*. 2005;99:547-54.

21. Peeters CM, Visser E, Van de Ree CL, Gosens T, Den Oudsten BL, De Vries J. Quality of life after hip fracture in the elderly: A systematic literature review. *Injury*. 2016;47(7):1369-82.
22. Auais M, Al-Zoubi F, Matheson A, Brown K, Magaziner J, French SD. Understanding the role of social factors in recovery after hip fractures: A structured scoping review. *Health & Social Care in the Community*. 2019;27(6):1375-87.
23. Fachgesellschaften eV AdWM. "S2-Leitlinie" Schenkelhalsfraktur des Erwachsenen" der Deutschen Gesellschaft für Unfallchirurgie. 2008.
24. Haywood KL, Griffin XL, Achten J, Costa ML. Developing a core outcome set for hip fracture trials. *The bone & joint journal*. 2014;96-B(8):1016-23.
25. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials*. 2012;13(1):1-8.
26. Haywood KL, Brett J, Tutton E, Staniszezwska S. Patient-reported outcome measures in older people with hip fracture: a systematic review of quality and acceptability. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2017;26(4):799-812.
27. Brandl M, Bernardi C, Apfelbacher C. Einführung von patientenberichteten Endpunkten: erhöhen sie die Qualität von Behandlungen? *Public Health Forum*. 2019;27(3):211-3.
28. George LK, Lucille B. Bearon 1980. *Quality of Life in Older Person: Meaning and Measurement*. New York: Human Science Press.
29. Fisher BJ. Successful aging and life satisfaction: A pilot study for conceptual clarification. *Journal of Aging Studies*. 1992;6(2):191-202.
30. Schlag B. Persönliche Veränderungen der Mobilität und der Leistungsfähigkeit im Alter. *Mobilität und demografische Entwicklung*. 2013:119.
31. Mollenkopf H, Walker A. *Quality of life in old age*: Springer; 2007.
32. Lepage A, Hunt S. The problem of quality of life in medicine. *Jama*. 1997;278(1):47-50.

33. Cella DF, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N, et al. Patient-reported outcomes in performance measurement: RTI Press Research Triangle Park, NC; 2015.
34. Bryant DM, Sanders DW, Coles CP, Petrisor BA, Jeray KJ, Laflamme GY. Selection of outcome measures for patients with hip fracture. *Journal of orthopaedic trauma*. 2009;23(6):434-41.
35. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology*. 2010;63(7):737-45.
36. de Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: a practical guide*: Cambridge University Press; 2011.
37. Ritschl V, Weigl R, Stamm T. *Wissenschaftliches Arbeiten und Schreiben: Verstehen, Anwenden, Nutzen für die Praxis*: Springer-Verlag; 2016.
38. Hutchings L, Fox R, Chesser T. Proximal femoral fractures in the elderly: How are we measuring outcome? *Injury*. 2011;42(11):1205-13.
39. Feng Y-S, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. *Quality of Life Research*. 2021;30(3):647-73.
40. Van Reenen M, Janssen B. *EQ-5D-5L user guide*. Rotterdam: EuroQol Research Foundation; 2015. 2019.
41. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2011;20(10):1727-36.
42. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2013;22(7):1717-27.
43. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Annals of medicine*. 2001;33(5):337-43.

44. Greiner W. 4 Der EQ-5D der EuroQol-Gruppe. Gesundheitsökonomische Evaluationen: Springer; 2012. p. 411-22.
45. Ludwig K, Graf von der Schulenburg JM, Greiner W. German Value Set for the EQ-5D-5L. *Pharmacoeconomics*. 2018;36(6):663-74.
46. Haywood KL, Garratt AM, Fitzpatrick R. Quality of life in older people: A structured review of generic self-assessed health instruments. *Quality of Life Research*. 2005;14(7):1651-68.
47. WARE JEJ, KOSINSKI M, KELLER SD. A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. *Medical Care*. 1996;34(3):220-33.
48. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics*. 2000;17(1):13-35.
49. Gandhi SK, Salmon JW, Zhao SZ, Lambert BL, Gore PR, Conrad K. Psychometric evaluation of the 12-item short-form health survey (SF-12) in osteoarthritis and rheumatoid arthritis clinical trials. *Clinical therapeutics*. 2001;23(7):1080-98.
50. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. *Journal of clinical epidemiology*. 1998;51(11):1171-8.
51. Kohlmann T, Bullinger M, Kirchberger-Blumstein I. [German version of the Nottingham Health Profile (NHP): translation and psychometric validation]. *Sozial- und Präventivmedizin*. 1997;42(3):175-85.
52. Hunt SM, McKenna SP, McEwen J, Backett EM, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. *Journal of Epidemiology and Community Health*. 1980;34(4):281-6.
53. Hunt SM, McKenna SP, Williams J. Reliability of a population survey tool for measuring perceived health problems: a study of patients with osteoarthritis. *Journal of Epidemiology and Community Health*. 1981;35(4):297.

54. Hinz A, Klaiberg A, Schumacher J, Brahler E. [The psychometric quality of the Nottingham Health Profile (NHP) in the general population]. *Psychotherapie, Psychosomatik, medizinische Psychologie*. 2003;53(8):353-8.
55. Skevington SM, Lotfy M, O'Connell KA. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2004;13(2):299-310.
56. Ericsson KA, Simon HA. Verbal reports as data. *Psychological review*. 1980;87(3):215.
57. Willis GB. *Cognitive Interviewing*. 2005 2020/02/13. Thousand Oaks, California: SAGE Publications, Inc. Available from: <https://methods.sagepub.com/book/cognitive-interviewing>.
58. Graf von der Schulenburg J, Claes C, Greiner W, Über A. The German Version of the Euro-Qol Questionnaire. *J Publ Health/Zeitschrift für Gesundheitswissenschaften*. 1998;6(1):2-20.
59. Bullinger M, Kirchberger I. SF-36: Fragebogen zum Gesundheitszustand; Handanweisung: Hogrefe, Verlag für Psychologie; 1998.
60. Whoqol Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychological medicine*. 1998;28(3):551-8.
61. Hunt SM, McKenna S, McEwen J, Williams J, Papp E. The Nottingham Health Profile: subjective health status and medical consultations. *Social Science & Medicine Part A: Medical Psychology & Medical Sociology*. 1981;15(3):221-9.
62. Willis GB. *Cognitive interviewing: A tool for improving questionnaire design*: Sage Publications; 2004.
63. Collins D. Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*. 2003;12(3):229-38.
64. Kuusela H, Paul P. A comparison of concurrent and retrospective verbal protocol analysis. *The American journal of psychology*. 2000;113(3):387-404.
65. Van Someren M, Barnard Y, Sandberg J. *The think aloud method: a practical approach to modelling cognitive*1994.

66. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2018;27(5):1159-70.
67. Apfelbacher CJ, Jones CJ, Frew A, Smith H. Validity of three asthma-specific quality of life questionnaires: the patients' perspective. *BMJ open*. 2016;6(12):e011793.
68. Dresing T, Pehl T. *Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende*. 2013;5:813-29.
69. Schreier M. *Qualitative content analysis in practice*: Sage Publications; 2012.
70. Mays N, Pope C. Assessing quality in qualitative research. *Bmj*. 2000;320(7226):50-2.
71. Heuschmann P, Kolominsky-Rabas P, Nolte C, Hünermund G, Ruf H, Laumeier I, et al. The reliability of the german version of the barthel-index and the development of a postal and telephone version for the application on stroke patients. *Fortschritte der Neurologie-Psychiatrie*. 2005;73(2):74-82.
72. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *The Journal of bone and joint surgery British volume*. 1996;78(2):185-90.
73. Deck R, Mittag O, Hüppe A, Muche-Borowski C, Raspe H. Index zur Messung von Einschränkungen der Teilhabe (IMET)–Erste Ergebnisse eines ICF-orientierten Assessmentinstruments. *Praxis Klinische Verhaltensmedizin und Rehabilitation*. 2007;76(2):113-20.
74. Kocalevent R, Brähler E. *OSSS Oslo Social Support Scale. Psychologische und sozialwissenschaftliche Kurzskalen: standardisierte Erhebungsinstrumente für Wissenschaft und Praxis*, Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft. 2014:216-24.
75. Hoebel J, Müters S, Kuntz B, Lange C, Lampert T. Measuring subjective social status in health research with a German version of the MacArthur Scale. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*. 2015;58(7):749-57.

76. Mahoney FI, Barthel DW. FUNCTIONAL EVALUATION: THE BARTHEL INDEX. Maryland state medical journal. 1965;14:61-5.
77. Naal FD, Sieverding M, Impellizzeri FM, von Knoch F, Mannion AF, Leunig M. Reliability and Validity of the Cross-Culturally Adapted German Oxford Hip Score. Clinical Orthopaedics and Related Research. 2009;467(4):952-7.
78. Deck R, Walther A, Staupendahl A, Katalinic A. limitations of social participation in general population-normative data of the IMET based on a population-based survey in northern Germany. Die Rehabilitation. 2015;54(6):402-8.
79. Dalgard O. Community health profile as tool for psychiatric prevention. Promotion of mental health. 1996;5:681-95.
80. Miyakawa M, Magnusson Hanson LL, Theorell T, Westerlund H. Subjective social status: its determinants and association with health in the Swedish working population (the SLOSH study). The European Journal of Public Health. 2012;22(4):593-7.
81. Chen B, Covinsky KE, Cenzer IS, Adler N, Williams BA. Subjective social status and functional decline in older adults. Journal of general internal medicine. 2012;27(6):693-9.
82. Brown TA. Confirmatory factor analysis for applied research: Guilford publications; 2015.
83. Gäde JC, Schermelleh-Engel K, Brandt H. Konfirmatorische Faktorenanalyse (CFA). Testtheorie und Fragebogenkonstruktion: Springer; 2020. p. 615-59.
84. Schumacker RE, Lomax RG. A beginner's guide to structural equation modeling: psychology press; 2004.
85. McDowell I. Measuring health: a guide to rating scales and questionnaires: Oxford University Press, USA; 2006.
86. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of clinical epidemiology. 2007;60(1):34-42.
87. Mokkink L, Terwee C, de Vet H. Key concepts in clinical epidemiology: Responsiveness, the longitudinal aspect of validity. Journal of clinical epidemiology. 2021;140:159-62.

88. de Boer MR, Terwee CB, de Vet HC, Moll AC, Volker-Dieben HJ, van Rens GH. Evaluation of cross-sectional and longitudinal construct validity of two vision-related quality of life questionnaires: the LVQOL and VCM1. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2006;15(2):233-48.
89. Marten O, Brand L, Greiner W. Feasibility of the EQ-5D in the elderly population: a systematic review of the literature. *Quality of Life Research.* 2022;31(6):1621-37.
90. van Leeuwen KM, Jansen APD, Muntinga ME, Bosmans JE, Westerman MJ, van Tulder MW, et al. Exploration of the content validity and feasibility of the EQ-5D-3L, ICECAP-O and ASCOT in older adults. *BMC Health Services Research.* 2015;15(1):201.
91. Bureau-Chalot F, Novella JL, Jolly D, Ankri J, Guillemin F, Blanchard F. Feasibility, acceptability and internal consistency reliability of the nottingham health profile in dementia patients. *Gerontology.* 2002;48(4):220-5.
92. Baro E, Ferrer M, Vazquez O, Miralles R, Pont A, Esperanza A, et al. Using the Nottingham Health Profile (NHP) among older adult inpatients with varying cognitive function. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2006;15(4):575-85.
93. Rolstad S, Adler J, Ryden A. Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2011;14(8):1101-8.
94. Hickey A, Barker M, McGee H, O'Boyle C. Measuring health-related quality of life in older patient populations. *PharmacoEconomics.* 2005;23(10):971-93.
95. Griffiths F, Mason V, Boardman F, Dennick K, Haywood K, Achten J, et al. Evaluating recovery following hip fracture: a qualitative interview study of what is important to patients. *BMJ open.* 2015;5(1):e005406.
96. Mallinson S. Listening to respondents: a qualitative assessment of the Short-Form 36 Health Status Questionnaire. *Social science & medicine (1982).* 2002;54(1):11-21.

97. Ahmad MA, Xypnitos FN, Giannoudis PV. Measuring hip outcomes: common scales and checklists. *Injury*. 2011;42(3):259-64.
98. Tidermark J, Bergstrom G. Responsiveness of the EuroQol (EQ-5D) and the Nottingham Health Profile (NHP) in elderly patients with femoral neck fractures. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2007;16(2):321-30.
99. Tidermark J, Bergstrom G, Svensson O, Tornkvist H, Ponzer S. Responsiveness of the EuroQol (EQ 5-D) and the SF-36 in elderly patients with displaced femoral neck fractures. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2003;12(8):1069-79.
100. Parsons N, Griffin XL, Achten J, Costa ML. Outcome assessment after hip fracture: is EQ-5D the answer? *Bone & joint research*. 2014;3(3):69-75.
101. Daltroy LH, Larson MG, Eaton HM, Phillips CB, Liang MH. Discrepancies between self-reported and observed physical function in the elderly: the influence of response shift and other factors. *Social Science & Medicine*. 1999;48(11):1549-61.
102. Moser DK, Heo S, Lee KS, Hammash M, Riegel B, Lennie TA, et al. 'It could be worse... lot's worse!' Why health-related quality of life is better in older compared with younger individuals with heart failure. *Age and ageing*. 2013;42(5):626-32.
103. Heckhausen J, Brim OG. Perceived problems for self and others: Self-protection by social downgrading throughout adulthood. *Psychology and Aging*. 1997;12(4):610.
104. Van Balen R, Essink-Bot ML, Steyerberg E, Cools H, Habbema DF. Quality of life after hip fracture: a comparison of four health status measures in 208 patients. *Disability and rehabilitation*. 2003;25(10):507-19.
105. Shyu Y-IL, Lu J-fR, Liang J. Evaluation of Medical Outcomes Study Short Form-36 Taiwan version in assessing elderly patients with hip fracture. *Osteoporosis international*. 2004;15(7):575-82.
106. Frihagen F, Grotle M, Madsen JE, Wyller TB, Mowinckel P, Nordsletten L. Outcome after femoral neck fractures: a comparison of Harris Hip Score, Eq-5d and Barthel Index. *Injury*. 2008;39(10):1147-56.
107. Kaambwa B, Bulamu NB, Mpundu-Kaambwa C, Oppong R. Convergent and discriminant validity of the Barthel Index and the EQ-5D-3L when used on older people

in a rehabilitation setting. *International Journal of Environmental Research and Public Health*. 2021;18(19):10314.

108. Jakobsson U, Westergren A, Lindskov S, Hagell P. Construct validity of the SF-12 in three different samples. *Journal of evaluation in clinical practice*. 2012;18(3):560-6.

109. Mishra V, Thomas G, Sibly T. Results of displaced subcapital fractures treated by primary total hip replacement. *Injury*. 2004;35(2):157-60.

110. Latham NK, Mehta V, Nguyen AM, Jette AM, Olarsch S, Papanicolaou D, et al. Performance-Based or Self-Report Measures of Physical Function: Which Should Be Used in Clinical Trials of Hip Fracture Patients? *Archives of physical medicine and rehabilitation*. 2008;89(11):2146-55.

111. Kathe N, Hayes CJ, Bhandari NR, Payakachat N. Assessment of reliability and validity of SF-12v2 among a diabetic population. *Value in Health*. 2018;21(4):432-40.

112. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Group MFPROCM. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value in Health*. 2007;10:S94-S105.

113. Chiang M-H, Huang Y-Y, Kuo Y-J, Huang S-W, Jang Y-C, Chu F-L, et al. Prognostic Factors for Mortality, Activity of Daily Living, and Quality of Life in Taiwanese Older Patients within 1 Year Following Hip Fracture Surgery. *Journal of Personalized Medicine*. 2022;12(1):102.

114. Altenbuchner A, Haug S, Weber K. Exploratory Analysis of Motion Tracking Data in the Rehabilitation Process of Geriatric Trauma Patients. *Stud Health Technol Inform*. 2019;260:138-45.

115. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Controlled clinical trials*. 1989;10(4):407-15.

116. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *Journal of clinical oncology*. 1998;16(1):139-44.

117. Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). *User manual*. 2018;78(1).

118. Konerding U. What does Cronbach's alpha tell us about the EQ-5D? A methodological commentary to "Psychometric properties of the EuroQol Five-Dimensional Questionnaire (EQ-5D-3L) in caregivers of autistic children". *Quality of Life Research*. 2013;22(10):2939-40.
119. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *Journal of advanced nursing*. 2000;32(4):1008-15.
120. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS medicine*. 2011;8(1):e1000393.
121. Chevance A, Tran V-T, Ravaud P. Controversy and Debate Series on Core Outcome Sets. Paper 1: Improving the generalizability and credibility of core outcome sets (COS) by a large and international participation of diverse stakeholders. *Journal of clinical epidemiology*. 2020;125:206-12.e1.
122. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET handbook: version 1.0. *Trials*. 2017;18(3):1-50.

7. Anhang

7.1 Qualitative Studie

7.1.1 Erhebung Patientencharakteristika

**Erfassung der Lebensqualität nach Hüftfraktur bei Menschen 60+**

Klinikinterne Patienten-ID:	ID:	Datum Interview: __. __. 201__																
Aufnahme																		
Aufnahme in die Unfallchirurgie/Klinikum: Datum __. __. 201__	Geschlecht: <input type="radio"/> männlich <input type="radio"/> weiblich <input type="radio"/> unbekannt																	
Alter bei Aufnahme: Geburtsdatum __. __. ____ oder Alter in Jahren ____	Wohnsituation vor Frakturereignis: <input type="radio"/> unbekannt <input type="radio"/> zu Hause oder betreutes Wohnen <input type="radio"/> Heim <input type="radio"/> Krankenhaus <input type="radio"/> Sonstiges: _____																	
Pflegebedürftigkeit: Liegt eine Einteilung in eine Pflegestufe/ Pflegegrad vor? <table border="0"> <tr> <td>Pflegestufe</td> <td>Pflegegrad</td> </tr> <tr> <td><input type="radio"/> unbekannt</td> <td><input type="radio"/> unbekannt</td> </tr> <tr> <td><input type="radio"/> nein</td> <td><input type="radio"/> nein</td> </tr> <tr> <td><input type="radio"/> Stufe 1</td> <td><input type="radio"/> Grad 1</td> </tr> <tr> <td><input type="radio"/> Stufe 2</td> <td><input type="radio"/> Grad 2</td> </tr> <tr> <td><input type="radio"/> Stufe 3</td> <td><input type="radio"/> Grad 3</td> </tr> <tr> <td></td> <td><input type="radio"/> Grad 4</td> </tr> <tr> <td></td> <td><input type="radio"/> Grad 5</td> </tr> </table>	Pflegestufe	Pflegegrad	<input type="radio"/> unbekannt	<input type="radio"/> unbekannt	<input type="radio"/> nein	<input type="radio"/> nein	<input type="radio"/> Stufe 1	<input type="radio"/> Grad 1	<input type="radio"/> Stufe 2	<input type="radio"/> Grad 2	<input type="radio"/> Stufe 3	<input type="radio"/> Grad 3		<input type="radio"/> Grad 4		<input type="radio"/> Grad 5	Gehfähigkeit vor Frakturereignis: <input type="radio"/> unbekannt <input type="radio"/> selbstständige Gehfähigkeit ohne Hilfsmittel <input type="radio"/> Gehfähigkeit außer Haus mit einem Gehstock oder einer Gehstütze <input type="radio"/> Gehfähigkeit außer Haus mit zwei Gehstützen oder Rollator <input type="radio"/> Gewisse Gehfähigkeit in der Wohnung, aber außer Haus nur mit Hilfsperson <input type="radio"/> Keine funktionale Gehfähigkeit (d.h. keine Nutzung der unteren Extremitäten möglich)	
Pflegestufe	Pflegegrad																	
<input type="radio"/> unbekannt	<input type="radio"/> unbekannt																	
<input type="radio"/> nein	<input type="radio"/> nein																	
<input type="radio"/> Stufe 1	<input type="radio"/> Grad 1																	
<input type="radio"/> Stufe 2	<input type="radio"/> Grad 2																	
<input type="radio"/> Stufe 3	<input type="radio"/> Grad 3																	
	<input type="radio"/> Grad 4																	
	<input type="radio"/> Grad 5																	
OP																		
Datum der OP: __. __. 201__																		
Entlassung/ Verlegung																		
Entlassung aus dem Akutkrankenhaus: Datum __. __. 201__	Entlassen nach: <input type="radio"/> Unbekannt <input type="radio"/> Zu Hause oder betreutes Wohnen <input type="radio"/> Heim <input type="radio"/> Geriatrische Rehabilitationsklinik <input type="radio"/> Akut-Geriatrie <input type="radio"/> AHB-Klinik <input type="radio"/> Anderes Krankenhaus <input type="radio"/> Tod																	

* Auszüge aus dem AltersTraumaRegister DGU®

7.1.2 Aufklärungsdokument



Klinik für UNFALLMEDIZIN
Alterstraumatologie

Prof. Dr. Michael Nerlich
Telefon Sekretariat: 0941/782-3810

Dr. Rainer Kretschmer & Dr. Ildiko Grell

Gerlinde Weigert



Universität Regensburg

Fakultät für MEDIZIN
Institut für Epidemiologie und Präventivmedizin
Professur für Medizinische Soziologie

Prof. Dr. Christian Apfelbacher
Telefon Sekretariat: 0941/944-5229
E-Mail: christian.apfelbacher@ukr.de

Magdalena Brandl, MPH
Telefon: 0941/ 944-5238
E-Mail: Magdalena.Brandl@ukr.de

Information für Teilnehmerinnen und Teilnehmer

Sehr geehrte Damen, sehr geehrte Herren,

wir möchten Sie über die Interviewstudie „**Lebensqualität nach Hüftfraktur bei Menschen über 60 Jahre**“ der Medizinischen Soziologie der Universität Regensburg informieren.

Bitte lesen Sie sich die folgenden Anmerkungen zur Studie aufmerksam durch. Sie können jederzeit alle Fragen stellen, die Sie eventuell im Zusammenhang mit der Studie haben.

Hintergrund und Ziel der Studie

Um beispielsweise den Erfolg einer Behandlung zu beurteilen, ist es immer wichtiger, dass **Patienten selbst** über Ihren aktuellen Gesundheitszustand **berichten**. Dazu gehört auch die gesundheitsbezogene **Lebensqualität**. Um diese messen zu können, gibt es bereits einige **Fragebögen**. Wir möchten mit dieser Studie herausfinden, ob diese Fragebögen auch wirklich geeignet sind, Ihre gesundheitsbezogene Lebensqualität richtig abzubilden.

Wir möchten also mit Ihnen diese Fragebögen ausfüllen und anschließend in einem Interview herausfinden, welche Meinung Sie zu diesen Fragebögen haben. Dazu zählen praktische Dinge, wie z.B. die Länge des Fragebogens. Wichtig sind auch die Inhalte des Fragebogens, also beispielsweise wie angemessen und verständlich sind die Fragen für Sie.

Außerdem möchten wir einige Angaben aus Ihrer Krankenakte entnehmen, wie z.B. Diagnose, Behandlung, Dauer des Krankenhausaufenthaltes.

Beschreibung bzw. Ablauf der Studie

1. Alle von Ihnen gemachten Aussagen dienen wissenschaftlichen Zwecken.
2. In einem Interview möchten wir mit Ihnen über verschiedene Fragebögen zur Lebensqualität sprechen. Wichtig ist **Ihre persönliche Sicht**.
3. Das Gespräch wird ca. **zwischen 30 und 45 Minuten** dauern und mit einem Tonbandgerät aufgezeichnet. Rückfragen sind jederzeit möglich.
4. Die gesammelten Informationen (Daten) aus dem Gespräch werden zur Auswertung in einer Datenbank zusammengefasst. Angaben, mit denen ein Rückschluss auf Ihre Person möglich ist, werden **getrennt** von den erhobenen Daten gespeichert. Eine Weitergabe zu anderen Zwecken als denen der Studie erfolgt nicht.

Leitung der Studie

Herr Prof. Dr. Apfelbacher (Professur für Medizinische Soziologie, Universität Regensburg) ist Leiter der Studie und wird von Herrn Dr. Kretschmer (Klinik für Unfallmedizin, Caritas-Krankenhaus St. Josef) unterstützt.

Freiwilligkeit der Studie

Ihre Teilnahme ist **freiwillig und kann jederzeit (auch während des Gesprächs) abgebrochen** werden. Die **Nichtteilnahme** an der Studie hat **keinerlei negative Folgen** für Sie. Ebenso können Sie Ihre Einwilligung zur Teilnahme an der Studie jederzeit ohne Angabe von Gründen widerrufen, ohne dass Ihnen daraus irgendwelche Nachteile entstehen.

7.1.3 Einwilligungsdokument



Caritas -
Krankenhaus
St. Josef

Klinik für UNFALLMEDIZIN
Alterstraumatologie

Prof. Dr. Michael Nerlich
Telefon Sekretariat: 0941/782-3810

Dr. Rainer Kretschmer & Dr. Ildiko Grell

Gerlinde Weigert



Universität Regensburg

Fakultät für MEDIZIN
Institut für Epidemiologie und Präventivmedizin
Professur für Medizinische Soziologie

Prof. Dr. Christian Apfelbacher
Telefon Sekretariat: 0941/944-5229
E-Mail: christian.apfelbacher@ukr.de

Magdalena Brandl, MPH
Telefon: 0941/ 944-5238
E-Mail: Magdalena.Brandl@ukr.de

Einwilligungserklärung

Die Interviewstudie „Gesundheitsbezogene Lebensqualität nach Hüftfraktur bei Menschen über 60 Jahre“ der Medizinischen Soziologie der Universität Regensburg beschäftigt sich mit der Validierung von Fragebögen zur Messung der gesundheitsbezogenen Lebensqualität nach einem Oberschenkelhalsbruch bzw. einer Hüftfraktur bei älteren Menschen.

Uns interessieren Ihre Einschätzung und Meinung zu den unterschiedlichen Fragebögen zum Thema Lebensqualität. Mit Ihrer Teilnahme tragen Sie dazu bei, Fragebögen zu entwickeln, die geeignet sind innerhalb Ihrer Zielgruppe die gesundheitsbezogene Lebensqualität richtig abzubilden.

Hiermit willige ich,

Vorname, Name des Teilnehmers

in die Teilnahme an einem Interview für die o.g. Studie ein.

Ich bestätige, dass meine Teilnahme am Interview freiwillig erfolgt. Ich bestätige weiterhin die ergänzende Information gemäß Europäischer Datenschutz-Grundverordnung erhalten und gelesen zu haben.

Ich erkläre mich damit einverstanden, dass im Rahmen dieses Interviews mich betreffende personenbezogene Daten/Angaben erhoben, verschlüsselt (pseudonymisiert) auf elektronische Datenträgern aufgezeichnet und verarbeitet werden dürfen. Ich stimme zu, dass für die Studie interessierende Daten wie z.B. Diagnose und Dauer des Klinikaufenthaltes aus der Krankenakte entnommen werden dürfen. Ich bin auch damit einverstanden, dass die Studienergebnisse in nicht rück-entschlüsselbarer (anonymer) Form, veröffentlicht werden. Mir ist bekannt, dass ich jederzeit die Löschung meiner Daten verlangen kann.

Ort, Datum, Unterschrift des Teilnehmers

Recht auf Löschung

Sie haben das Recht auf Löschung Sie betreffender personenbezogener Daten, z. B. wenn diese Daten für den Zweck, für den sie erhoben wurden, nicht mehr notwendig sind und der Löschung keine gesetzlichen Aufbewahrungsfristen entgegenstehen.

Recht auf Einschränkung der Verarbeitung

Unter bestimmten Voraussetzungen haben Sie das Recht auf Einschränkung der Verarbeitung zu verlangen, d. h. die Daten dürfen nur gespeichert, nicht verarbeitet werden. Dies müssen Sie beantragen. Wenden Sie sich hierzu bitte an die Studienleitung.

Recht auf Datenübertragbarkeit

Sie haben das Recht, die Sie betreffenden personenbezogenen Daten, die Sie dem Verantwortlichen für die Studie bereitgestellt haben, zu erhalten. Damit können Sie beantragen, dass diese Daten entweder Ihnen oder, soweit technisch möglich, einer anderen von Ihnen benannten Stelle übermittelt werden.

Widerspruchsrecht

Sie haben das Recht, jederzeit gegen konkrete Entscheidungen oder Maßnahmen zur Verarbeitung der Sie betreffenden personenbezogenen Daten Widerspruch einzulegen. Eine Verarbeitung findet anschließend grundsätzlich nicht mehr statt, es sei denn, die Verarbeitung ist gesetzlich weiterhin gefordert (wie beispielsweise in § 40 Abs. 2a S. 2 Nr. 3 AMG).

Möchten Sie eines dieser Rechte in Anspruch nehmen, wenden Sie sich bitte an Ihren Studienleiter. Außerdem haben Sie das **Recht, Beschwerde bei der Datenschutzaufsichtsbehörde einzulegen**, wenn Sie der Ansicht sind, dass die Verarbeitung der Sie betreffenden personenbezogenen Daten gegen die DSGVO verstößt.

Datenschutz: Kontaktdaten der Studienleitung

Prof. Dr. Christian Apfelbacher PhD

Institut für Medizinische Soziologie

Dr. Gessler-Straße 17

93051 Regensburg

Tel: 0941 944 5230

Magdalena Brandl, Doktorandin

E-Mail: magdalena.brandl@klinik.uni-regensburg.de

Datenschutz: Datenschutz-Aufsichtsbehörde

Der Bayerische Landesbeauftragte für den Datenschutz

Prof. Dr. Thomas Petri

E-Mail: poststelle@datenschutz-bayern.de

**Ergänzende Information für Studienteilnehmer gemäß
Europäischer Datenschutz-Grundverordnung¹
für bereits laufende medizinische Forschungsvorhaben
(Start vor 25.05.2018)**

Studientitel: Gesundheitsbezogene Lebensqualität nach Hüftfraktur bei Menschen über 65 Jahre

Sehr geehrte/r Studienteilnehmer/in,

aufgrund des Wirksamwerdens der Europäischen Datenschutz-Grundverordnung (**DSGVO**) zum 25. Mai 2018 ändern sich die Datenschutzvorschriften in Europa. Auch für bereits laufende medizinische Forschungsvorhaben ergeben sich dadurch neue Anforderungen an die Verarbeitung personenbezogener Daten.

Zusätzlich werden Sie hiermit über die in der DSGVO festgelegten Rechte informiert:

Rechtsgrundlage

Die Rechtsgrundlagen zur Verarbeitung der Sie betreffenden personenbezogenen Daten bildet bei dieser Studie Ihre freiwillige schriftliche Einwilligung gemäß DSGVO sowie die Deklaration von Helsinki (Erklärung des Weltärztebundes zu den ethischen Grundsätzen für die medizinische Forschung am Menschen) und die Leitlinie für Gute Klinische Praxis.

Bezüglich Ihrer Daten haben Sie folgende Rechte:

Einwilligung zur Verarbeitung personenbezogener Daten und Recht auf Widerruf der Einwilligung

Die Verarbeitung Ihrer personenbezogenen Daten ist nur mit Ihrer Einwilligung rechtmäßig.

Sie haben das Recht, Ihre Einwilligung zur Verarbeitung personenbezogener Daten jederzeit zu widerrufen. Es dürfen jedoch die bis zu diesem Zeitpunkt erhobenen Daten durch die in der Patienteninformation/Einwilligungserklärung zu der jeweiligen Studie genannten Stellen verarbeitet werden.

Recht auf Auskunft

Sie haben das Recht auf Auskunft über die Sie betreffenden personenbezogenen Daten, die im Rahmen der Studie erhoben, verarbeitet oder ggf. an Dritte übermittelt werden.

Recht auf Berichtigung

Sie haben das Recht, Sie betreffende unrichtige personenbezogene Daten berichtigen zu lassen.

¹ Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)

7.1.4 Ethik



Universität Regensburg

Ethikkommission · Universität Regensburg · 93040 Regensburg

Universität Regensburg
 Institut für Epidemiologie und Präventivmedizin
 Medizinische Soziologie
 Frau Magdalena Brandl
 Dr.-Gessler-Str. 17
 93051 Regensburg

20.09.2017

Beratung nach § 15 Abs. 1 Berufsordnung für die Ärzte Bayerns

für das Forschungsvorhaben:

Studientitel: Inhaltsvalidität und Machbarkeit häufig verwendeter Fragebögen zur Messung der HRQOL bei Patienten über 65 Jahre mit Hüftfraktur aus Patientensicht: qualitative Studie

Antragssteller: Frau Magdalena Brandl, Universität Regensburg

Die Ethikkommission der Universität Regensburg hat in Ihrer Sitzung am 20.09.2017 über das o.g. Forschungsvorhaben auf Grundlage der im Anhang aufgeführten Unterlagen beraten. Es ergeben sich daraus keine berufsethischen oder rechtlichen Bedenken gegen das vorgelegte Forschungsvorhaben.

Es wird auf folgendes grundsätzlich hingewiesen:

1. Unabhängig vom Beratungsergebnis verbleibt die ärztliche und juristische Verantwortung beim Forscher und seinen Mitarbeitern. Eine Nichtbeachtung des Beratungsergebnisses kann berufs- und haftungsrechtliche Folgen nach sich ziehen.
2. Die Auflagen der Deklaration von Helsinki des Weltärztebundes in ihrer aktuellen Fassung hinsichtlich ethischen und rechtlichen Aspekten biomedizinischer Forschung am Menschen sind strikt zu beachten.
3. Die Ethikkommission erwartet bei Interventionsstudien, dass ihr alle schwerwiegenden oder unerwarteten unerwünschten Ereignisse (u.a. Todesfälle), die während der Studie auftreten und die Sicherheit der Studienteilnehmer oder die Durchführung der Studie beeinträchtigen können, unverzüglich schriftlich mitgeteilt werden. Dieses sollte in Verbindung mit einer Stellungnahme des Antragsstellers geschehen, ob aus seiner Sicht die Nutzen-Risiko-Relation des Vorhabens verändert ist.
4. Die Ethikkommission bittet darum, dass ihr der Abbruch oder Abschluss einer Studie mitgeteilt werden.

**Ethikkommission
 an der Universität Regensburg**

Prof. Dr. med. Christian Stroszczyński, Vorsitzender

Ass. jur. Jan von Hassel, Geschäftsführer

Geschäftsstelle:

Telefon +49 941 943-5370

Telefax +49 941 943-5369

Postanschrift:

Universität Regensburg

ETHIKKOMMISSION

D-93040 Regensburg

ethikkommission@klinik.uni-regensburg.de

<http://ethikkommission.uni-regensburg.de>

Unser Zeichen: **17-703-101**

2

5. Dieses Schreiben ist mit den Studienunterlagen jederzeit sorgfältig aufzubewahren. Duplikate oder Abschriften dieses Schreibens können im Nachhinein nicht erstellt werden.
Auf die Rechtspflichten zum Umgang mit dienstlichem Schriftgut bzw. Urkunden wird verwiesen.
6. Auf Grundlage dieser rein berufsrechtlichen Beratung können Sie nachträgliche Änderungen am Protokoll dieses Forschungsvorhabens vornehmen, ohne dafür eine erneute Beratung (umgangssprachlich 'Amendmentvotum') durch die Ethikkommission beantragen zu müssen. Zur Begrenzung rechtlicher Risiken wird eine solche Beratung aber gleichwohl dringend empfohlen.

Sobald Sie jedoch ein neues Forschungsvorhaben durchführen wollen, müssen Sie dieses einer eigenständigen Beratung durch die Ethikkommission zuführen. Hierfür gilt gemäß Grundsatzbeschluss unserer Ethikkommission vom 02.08.2016:

In der Regel handelt es sich noch um ein und dasselbe Forschungsvorhaben ohne eine erneute Beratungspflicht, wenn sich lediglich ergänzende Fragestellungen im Rahmen der selben Hypothese, methodische Erweiterungen oder Beschränkungen oder Erweiterungen oder Beschränkungen in der Studienpopulation nachträglich ergeben. Um ein neues Forschungsvorhaben handelt es sich aber in der Regel, wenn die Formulierung einer neuen Hypothese, wesentliche Änderungen am Studiengegenstand bzw. der Entität sowie wesentliche Änderungen an der wissenschaftlichen oder technischen Vorgehensweise vorgenommen werden sollen, was dann eine Pflicht zur neuerlichen Beratung durch die Ethikkommission begründet. Gesetzliche Vorschriften bleiben unberührt.

7. Die Ethikkommission bestätigt die Bearbeitung gemäß der GCP/ICH-Richtlinien.
8. Die Ethikkommission empfiehlt im Einklang mit der Deklaration von Helsinki nachdrücklich die Registrierung der Studie vor Studienbeginn in einem öffentlich zugänglichen Register, das die von der WHO geforderten Voraussetzungen erfüllt.
9. Falls kein gesetzlicher Kostenbefreiungstatbestand greift, wird ein gesonderter Kostenbescheid für die Gebühren und Auslagen der Ethikkommission ergehen.
10. Die Übermittlung personenbezogener Daten einschließlich DNA-tragender Biomaterialien in datenschutzrechtlich unsichere Drittstaaten, wie etwa die USA, bedarf einer gesonderten datenschutzrechtlichen Beurteilung und Risikoaufklärung.



Prof. Dr. Christian Stroszczyński
Vorsitzender der Ethikkommission

Anlagen

3

Liste der eingereichten Unterlagen:

Prüfplan, 01_Protokoll_MagdalenaBrandl_MedSoz.pdf
Gegenstand des Forschungsvorhabens, 02_Gegenstand des
Forschungsvorhabens_MagdalenaBrandl_MedSoz.pdf
Nutzen-Risiko-Abwägung, 03_Angaben zum Nutzen_MagdalenaBrandl_MedSoz.pdf
Patienteninformation, 04_Aufklärungsdokument_MagdalenaBrandl_MedSoz.pdf
Patienteneinwilligung, 05_Einwilligungserklärung_MagdalenaBrandl_MedSoz.pdf
Literaturverzeichnis, 06_Literatur_MagdalenaBrandl_MedSoz.pdf
Datenschutzerklärung, 07_Angaben zur Verarbeitung von
Personendaten_MagdalenaBrandl_MedSoz.pdf
Sonstiges, 08_Angaben zur ethischen Problematik_MagdalenaBrandl_MedSoz.pdf
Keine, Gebührenreduzierung.pdf
Keine, Kurzbeschreibung.pdf
Sonstiges, Leitfaden_MagdalenaBrandl_MedSoz.pdf

4



An dieser Entscheidung der Ethikkommission in Ihrer Sitzung vom 20.09.2017 haben mitgewirkt:

Prof. Dr. C. Stroszczyński

PD Dr. K. P. Ittner

Prof. Dr. H.E. Müller

Prof. Dr. U. Reischl

Dr. B. Knoppke

PD Dr. B. Dietl

Herr W. Stelzl

7.1.5 Interviewleitfaden

Leitfaden „Lebensqualität nach Hüftfraktur“

Themenbereiche	Leitfragen	
Einstieg	Vielen Dank, dass Sie sich die Zeit für das Interview nehmen. Mit dem Interview unterstützen Sie ein Projekt der Universität Regensburg, in dem wir untersuchen wollen, wie gut vorhandene Fragebögen Ihre aktuelle Lebensqualität beschreiben können. Es geht darum welche Dinge für Ihre Lebensqualität wichtig sind und inwiefern die vorhandenen Fragebögen in der Lage sind diese Aspekte aufzugreifen.	Prompts: - Vorhin haben Sie gesagt, dass... - Habe ich richtig verstanden, dass... - Können Sie mir das erklären? - Können Sie mehr darüber sagen?
	Das Interview wird auf Tonband aufgezeichnet, damit wir es auswerten können. Selbstverständlich werten wir das Interview so aus, dass man nicht mehr auf ihre Person rückschließen kann (anonymisiert). (EINVERSTÄNDNISERKLÄRUNG & AUFNAHMEGERÄT ANSCHALTEN).	Minimal probes (mhm) – requests – summarizing & Reflecting
	Bevor wir anfangen ist mir noch wichtig zu erwähnen, dass es keine richtigen oder falschen Antworten gibt, da es in diesem Gespräch um Ihre Sichtweisen und Einschätzungen geht und nicht darum, Wissen abzufragen. Die Gesprächsleitung und Themenstellung geschieht durch mich. Allerdings sind auch Aspekte wünschenswert, die Ihnen zusätzlich einfallen.	Können Sie mir ein Beispiel nennen?
Genereller Eindruck zu den Fragebögen	Wie war Ihr Eindruck beim Ausfüllen der Fragebögen? - Was haben Sie sich beim Ausfüllen der Fragebögen gedacht? - Was waren Ihre Gefühle beim Ausfüllen der Fragebögen?	- unsicher/froh, Dinge los zu werden/überfordert,
Länge der Fragebögen	Wie fanden Sie die Länge der Fragebögen? - Was denken Sie über die Länge der Fragebögen? - Wie empfanden Sie die Anzahl an Fragen?	
Layout/ Darstellung	Wie fanden Sie das Layout/ die Darstellung der Fragebögen? - Aufbau aus Fragen und Antworten/ Anordnung der Fragen	z.B. leichter wenn Fragen und darunter Antwortmöglichkeiten oder wenn links Fragen und rechts Antwortmöglichkeiten
Format der Fragen/ Antworten	Was halten Sie von den Antwortmöglichkeiten? - Wie finden Sie die Abstufung der Antworten?	Menge der Antworten bzw. Abstufung/Formulierung der Antworten passend bzw. klar
Verständlichkeit	Wie verständlich waren die Fragen für Sie?	

Leitfaden „Lebensqualität nach Hüftfraktur“

Spezifikation der Fragestellung	Gab es für Sie Fragen, die man genauer hätte formulieren müssen? Wussten Sie bei allen Fragen was gemeint war/ auf was die Frage hinaus will?	
Redundanz	Gab es Fragen, die Sie als doppelt oder als sehr ähnlich empfunden haben?	
Belastung/Aufwand	Wie belastend war es für Sie die Fragen zu beantworten?	
Validität (Angemessenheit zur Situation)	Nun geht es mehr um den thematischen Inhalt der Fragen: Wie fanden Sie haben die Fragen die Probleme Ihrer aktuellen Situation nach einer Hüftfraktur abgedeckt? (How much did you feel the questions in the questionnaires covered the issues you are concerned with because of your asthma Gibt es wichtige Aspekte die im Fragebogen fehlen?	
Vorschläge für Verbesserungen	Gibt es Ihrerseits Vorschläge für Verbesserungen der Fragebögen?	
Präferenz	Fanden Sie einen Fragebogen besser als den/die anderen? Könnten Sie einen präferieren/bevorzugen? Sollten alle Fragebögen nicht geeignet erscheinen: welcher ist dennoch Ihrer Meinung nach der passendste? Würden Sie einen der Fragebögen empfehlen?	
Abschluss	Gibt es noch etwas, dass Sie dem Gespräch hinzufügen möchten? Vielleicht ein wichtiger Aspekt, der bisher gar nicht erwähnt wurde, aber wichtig ist für den Einfluss, den die Hüftfraktur auf Ihr Leben hat?	Sie haben bereits ... gesagt, gibt es darüber hinaus ...? (→ summarizing)
	Damit sind wir am Ende des Interviews angekommen. Ich danke Ihnen sehr herzlich, dass Sie sich die Zeit genommen haben mit mir zu sprechen.	

7.1.6 Übersicht Fragebögen

SF-12: gelber Fragebogen	EQ-5D: grüner Fragebogen	NHP: blauer Fragebogen	WHOQOL: rosa Fragebogen
Körperliche Gesundheit	Mobilität	Körperliche Mobilität	Körperliche Gesundheit
Geistige Gesundheit	Angst/Depression	Emotionen	Geistiges Wohlbefinden
Körperliche Schmerzen	Schmerz/Unwohlsein	Schmerz	
Soziale Beziehungen		Soziale Beziehungen	Soziale Beziehungen
„Körperliche“ Einschränkungen bei täglichen Aktivitäten	Selbstversorgung		
Vitalität	Gewöhnliche Aktivitäten	Energie	
„Geistige“ Einschränkungen bei täglichen Aktivitäten			
			Umwelt (Wohnumfeld, Versorgung, Transportmöglichkeiten)
Allgemeine Gesundheitswahrnehmung		Schlaf	
8 Kategorien	5 Kategorien	6 Kategorien	4 Kategorien

SF-12: gelber Fragebogen	EQ-5D: grüner Fragebogen	NHP: blauer Fragebogen	WHOQOL: rosa Fragebogen
12 Fragen	5 Fragen	38 Fragen	26 Fragen
Unterschiedliche Antwortmöglichkeiten	Beantwortung in 5 Stufen:	Beantwortung nur mit Ja/Nein	Beantwortung in 5 Stufen
	Keine Probleme – leichte Probleme – mäßige Probleme – große Probleme – nicht in der Lage		Überhaupt nicht – eher nicht – halbwegs – überwiegend – völlig

7.1.8 WHOQOL-BREF

WHOQOL-Bref

3

Bitte lesen Sie jede Frage, überlegen Sie, wie Sie sich in den vergangenen zwei Wochen gefühlt haben, und kreuzen Sie die Zahl auf der Skala an, die für Sie am ehesten zutrifft.

		Sehr schlecht	Schlecht	Mittelmäßig	Gut	Sehr gut
1	Wie würden Sie Ihre Lebensqualität beurteilen?	1	2	3	4	5
		Sehr unzufrieden	Unzufrieden	Weder zufrieden noch unzufrieden	Zufrieden	Sehr zufrieden
2	Wie zufrieden sind Sie mit Ihrer Gesundheit?	1	2	3	4	5

In den folgenden Fragen geht es darum, **wie stark** Sie während der vergangenen zwei Wochen bestimmte Dinge erlebt haben.

		Überhaupt nicht	Ein wenig	Mittelmäßig	Ziemlich	Äußerst
3	Wie stark werden Sie durch Schmerzen daran gehindert, notwendige Dinge zu tun?	1	2	3	4	5
4	Wie sehr sind Sie auf medizinische Behandlung angewiesen, um das tägliche Leben zu meistern?	1	2	3	4	5
5	Wie gut können Sie Ihr Leben genießen?	1	2	3	4	5
6	Betrachten Sie Ihr Leben als sinnvoll?	1	2	3	4	5

WHOQOL-Bref

		Überhaupt nicht	Ein wenig	Mittelmäßig	Ziemlich	Äußerst
7	Wie gut können Sie sich konzentrieren?	1	2	3	4	5
8	Wie sicher fühlen Sie sich in Ihrem täglichen Leben?	1	2	3	4	5
9	Wie gesund sind die Umweltbedingungen in Ihrem Wohngebiet?	1	2	3	4	5

In den folgenden Fragen geht es darum, **in welchem Umfang** Sie während der vergangenen zwei Wochen bestimmte Dinge erlebt haben oder in der Lage waren, bestimmte Dinge zu tun.

		Überhaupt nicht	Eher nicht	Halbwegs	Überwiegend	Völlig
10	Haben Sie genug Energie für das tägliche Leben?	1	2	3	4	5
11	Können Sie Ihr Aussehen akzeptieren?	1	2	3	4	5
12	Haben Sie genug Geld, um Ihre Bedürfnisse erfüllen zu können?	1	2	3	4	5
13	Haben Sie Zugang zu den Informationen, die Sie für das tägliche Leben brauchen?	1	2	3	4	5
14	Haben Sie ausreichend Möglichkeiten zu Freizeitaktivitäten?	1	2	3	4	5

WHOQOL-Bref

		Sehr schlecht	Schlecht	Mittelmäßig	Gut	Sehr gut
15	Wie gut können Sie sich fortbewegen?	1	2	3	4	5

In den folgenden Fragen geht es darum, wie **zufrieden, glücklich oder gut** Sie sich während der vergangenen zwei Wochen hinsichtlich verschiedener Aspekte Ihres Lebens gefühlt haben.

		Sehr unzufrieden	Unzufrieden	Weder zufrieden noch unzufrieden	Zufrieden	Sehr zufrieden
16	Wie zufrieden sind Sie mit Ihrem Schlaf?	1	2	3	4	5
17	Wie zufrieden sind Sie mit Ihrer Fähigkeit, alltägliche Dinge erledigen zu können?	1	2	3	4	5
18	Wie zufrieden sind Sie mit Ihrer Arbeitsfähigkeit?	1	2	3	4	5

		Sehr unzufrieden	Unzufrieden	Weder zufrieden noch unzufrieden	Zufrieden	Sehr zufrieden
19	Wie zufrieden sind Sie mit sich selbst?	1	2	3	4	5
20	Wie zufrieden sind Sie mit Ihren persönlichen Beziehungen?	1	2	3	4	5
21	Wie zufrieden sind Sie mit Ihrem Sexualleben?	1	2	3	4	5

WHOQOL-Bref

		Sehr unzufrieden	Unzufrieden	Weder zufrieden noch unzufrieden	Zufrieden	Sehr zufrieden
22	Wie zufrieden sind Sie mit der Unterstützung durch Ihre Freunde?	1	2	3	4	5
23	Wie zufrieden sind Sie mit Ihren Wohnbedingungen?	1	2	3	4	5
24	Wie zufrieden sind Sie mit Ihren Möglichkeiten, Gesundheitsdienste in Anspruch nehmen zu können?	1	2	3	4	5
25	Wie zufrieden sind Sie mit den Beförderungsmitteln, die Ihnen zur Verfügung stehen?	1	2	3	4	5

In den folgenden Fragen geht es darum, **wie oft** sich während der vergangenen zwei Wochen bei Ihnen negative Gefühle eingestellt haben, wie zum Beispiel Angst oder Traurigkeit.

		Niemals	Nicht oft	Zeitweilig	Oftmals	Immer
26	Wie häufig haben Sie negative Gefühle wie Traurigkeit, Verzweiflung, Angst oder Depression?	1	2	3	4	5

Hat Ihnen jemand beim Ausfüllen dieses Fragebogens geholfen?

Ja Nein

Wie lange hat es gedauert, den Fragebogen auszufüllen?

_____ Minuten

7.1.9 NHP

NHP

2

Im Folgenden finden Sie eine Liste von Problemen, die man im Alltagsleben haben kann.

Bitte gehen Sie die Liste sorgfältig durch und kreuzen Sie **bei jeder Aussage** an , ob diese zurzeit für Sie zutrifft (**JA**) oder nicht zutrifft (**NEIN**).

Bitte beantworten Sie jede Frage.

Wenn Sie nicht sicher sind, ob Sie mit JA oder NEIN antworten sollen, kreuzen Sie die Antwort an, die **am ehesten** zutrifft.

	JA	NEIN
Ich bin andauernd müde	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe nachts Schmerzen	<input type="checkbox"/>	<input type="checkbox"/>
Ich fühle mich niedergeschlagen	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Ich habe unerträgliche Schmerzen	<input type="checkbox"/>	<input type="checkbox"/>
Ich nehme Tabletten, um schlafen zu können	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe vergessen, wie es ist, Freude zu empfinden	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Ich fühle mich gereizt	<input type="checkbox"/>	<input type="checkbox"/>
Ich finde es schmerzhaft, meine Körperposition zu verändern	<input type="checkbox"/>	<input type="checkbox"/>
Ich fühle mich einsam	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Ich kann mich nur innerhalb des Hauses bewegen	<input type="checkbox"/>	<input type="checkbox"/>
Es fällt mir schwer, mich zu bücken	<input type="checkbox"/>	<input type="checkbox"/>
Alles strengt mich an	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Ich wache in den frühen Morgenstunden vorzeitig auf	<input type="checkbox"/>	<input type="checkbox"/>
Ich kann überhaupt nicht gehen	<input type="checkbox"/>	<input type="checkbox"/>
Es fällt mir schwer, zu anderen Menschen Kontakt aufzunehmen	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Die Tage ziehen sich hin	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe Schwierigkeiten, Treppen oder Stufen hinauf- und hinunterzugehen	<input type="checkbox"/>	<input type="checkbox"/>
Es fällt mir schwer, mich zu strecken und nach Gegenständen zu greifen	<input type="checkbox"/>	<input type="checkbox"/>

NHP

Bitte denken Sie daran: Wenn Sie nicht sicher sind, ob Sie mit JA oder NEIN antworten sollen, kreuzen Sie die Antwort an, die am ehesten zutrifft.		
	JA	NEIN
Ich habe Schmerzen beim Gehen	<input type="checkbox"/>	<input type="checkbox"/>
Mir reißt in letzter Zeit oft der Geduldsfaden	<input type="checkbox"/>	<input type="checkbox"/>
Ich fühle, dass ich niemandem nahestehe	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Ich liege nachts die meiste Zeit wach	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe das Gefühl, die Kontrolle zu verlieren	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe Schmerzen, wenn ich stehe	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Es fällt schwer, mich selbst anzuziehen	<input type="checkbox"/>	<input type="checkbox"/>
Meine Energie lässt schnell nach	<input type="checkbox"/>	<input type="checkbox"/>
Es fällt mir schwer, lange zu stehen (z.B. am Spülbecken, an der Bushaltestelle)	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Ich habe ständig Schmerzen	<input type="checkbox"/>	<input type="checkbox"/>
Ich brauche lange zum Einschlafen	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe das Gefühl, für andere Menschen eine Last zu sein	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Sorgen halten mich nachts wach	<input type="checkbox"/>	<input type="checkbox"/>
Ich fühle, dass das Leben nicht lebenswert ist	<input type="checkbox"/>	<input type="checkbox"/>
Ich schlafe nachts schlecht	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Es fällt mir schwer, mit anderen Menschen auszukommen	<input type="checkbox"/>	<input type="checkbox"/>
Ich brauche Hilfe, wenn ich mich außer Haus bewegen will (z.B. einen Stock, oder jemanden, der mich stützt)	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe Schmerzen, wenn ich Treppen oder Stufen hinauf- und hinabgehe	<input type="checkbox"/>	<input type="checkbox"/>
	JA	NEIN
Ich wache deprimiert auf	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe Schmerzen, wenn ich sitze	<input type="checkbox"/>	<input type="checkbox"/>

7.2 Quantitative Studie

7.2.1 Analyseplan



Statistical Analysis Plan (SAP) – Psychometric Analysis

Study title:

Validation of two generic measures of quality of life (SF-12 and EQ-5D) in a population of old people with hip fracture: a psychometric study

Institutions:

- University of Regensburg, Institute for Epidemiology and Preventive Medicine, Medical Sociology
- Institute for Social Medicine and Health Economics, Faculty of Medicine, Otto-von-Guericke University Magdeburg, Magdeburg, Germany
- Institute for Health Services Research and Health Economics, Centre for Health and Society, Faculty of Medicine, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

Scientists:

- Prof. Dr. Christian Apfelbacher
- Prof. Dr. Andrea Icks
- Magdalena Brandl
- Susanne Brandstetter
- Michaela Ritschel
- Dr. Silke Andrich
- Dr. Veronika Gontscharuk



Table of Contents

General information on the analysis plan	3
Study design	3
Methods	5
Generic HRQOL measures	5
Reference measures	6
Definition of analysis sets	6
Psychometric analysis.....	7
Validity.....	7
Reliability	9
Responsiveness	10
Interpretability	11
Statistics	13
References	13



General information on the analysis plan

This analysis plan was agreed between all scientists involved at the University of Regensburg, the University of Magdeburg and the University of Düsseldorf and reviewed by a statistician. This SAP prospectively describes all planned psychometric analyses of the Short Form 12 Questionnaire (SF-12) and EQ-5D using data from the ProFem study. The aim of the study was to extend the validation of the SF-12 and EQ-5D to the older hip fracture patient population.

Study design

Data for this psychometric analysis will be obtained from questionnaire data of a population-based prospective study called ProFem in people aged 60 years or older, who have sustained a proximal femoral fracture (PFF, colloquially hip fracture). The ProFem study is conducted at the Heinrich Heine University Düsseldorf, coordinated by the Institute for Health Services Research and Health Economics (Principal Investigator: Prof. Dr. Dr. Andrea Icks). Figure 1 illustrates the design of the psychometric validation study. We will apply analytical principles of classical test theory (CTT) to guide the analysis of the data set generated from the prospective study.

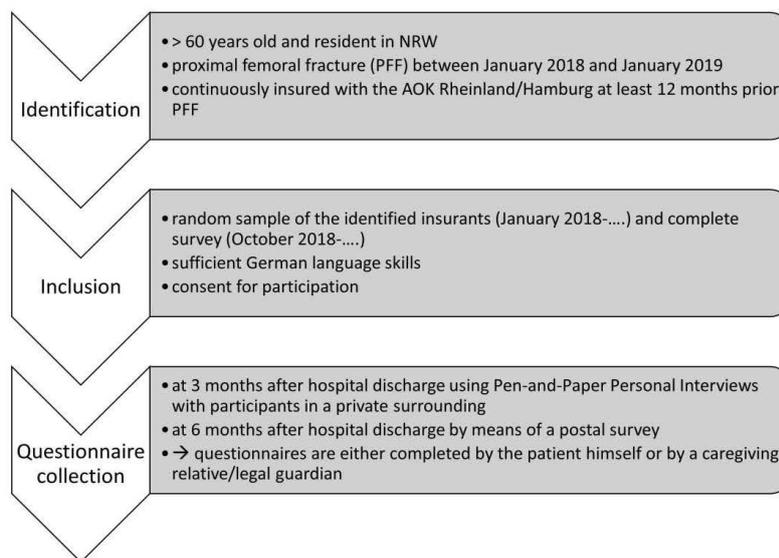


Figure 1: Study design overview



To assess the psychometric properties (validity, reliability, sensitivity to change) of the SF-12 and EQ-5D questionnaire we will use all variables listed in table 1. In the following, the planned analyses are briefly outlined and then listed in detail in further sections.

Table 1: Collected variables and timing

Construct/ category	Instrument/variables	3 months after hospital discharge	6 months after hospital discharge
Outcome			
HRQOL	EQ-5D-5L (1)	x	x
	SF-12 (2)	x	x
Reference measures			
Self-perceived health status	Global-Rating of Change Scale		x
Functioning	Barthel-Index (3)	x	x
	Oxford Hip Score (4)	x	x
Participation	IMET (Index for Measuring Participation Restrictions) (5)	x	x
Social Support	Oslo Social Support Scale (6)	x	
General characteristics			
Sociodemographic data	Age, sex	x	
Socioeconomic status	MacArthur Scale (7)	x	
Clinical data			
Cognitive impairment/reduced general condition	Long or short questionnaire version	x	x

For psychometric measurement characteristics based on cross-sectional data, the 3-months questionnaire will be used for primary analysis. For longitudinal psychometric measurement characteristics, the intersection of the questionnaires collected at both time points will be used.

In order to assess **construct validity**, several hypotheses will be tested. Therefore, correlations of the two measures of interest (SF-12 and EQ-5D) with reference measures (see table 1) will be investigated and judged based on a priori **hypotheses** to assess convergent and discriminant validity. In addition, the known-groups method will be used for discriminative validation. **Structural validity** (dimensionality) of both measures of interest will be verified by a confirmatory factor analysis (CFA).



Interpretability will be assessed by calculating the number of missing items and their distribution and by assessing floor and ceiling effects. To define minimal clinically important difference (**MCID**) we will use an anchor-based and a distribution-based approach.

Cronbach's alpha (SF-12 only) as well as Inter-Item and Item-Total-Correlations will be used to **assess internal consistency**.

In stable patients (patients who experienced no change between the first and second assessment), **test-retest reliability** can be assessed by calculating intraclass correlation coefficients (ICC). For patients who changed in their clinical status over time **responsiveness** will be computed.

All analyses will be done separately for self-reported and eventually (depending on the sample size) for proxy-reported questionnaires and will be carried out with R.

Methods

Generic HRQOL measures

SF-12

The German version of the SF-12 will be used (2, 8). The questionnaire covers the previous four weeks of a person's life. It comprises 12 items and produces two scores, a physical component summary score (PCS-12) and a mental component summary score (MCS-12). The items will be scored and transformed according to the standard procedure in the manual (8). Briefly, each item score is weighted and then transformed to a mean of 50 and a standard deviation of 10 in the general German population. Better HRQOL is denoted with higher scores (maximum 100). The SF-12 is one of the most commonly used instruments to assess HRQOL and its psychometric properties has been tested in various patient groups, for example in the elderly (9).

EQ-5D-5L

The EQ-5D-5L (1, 10) consists of two parts: the EQ-5D descriptive system and the EQ visual analogue scale (EQ-VAS). The descriptive system includes five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression). A health profile in form of a 5-digit code can be generated (from best to worst state where '11111' is the best and '55555' the worst state). Summary statistics will be derived, including numbers of patients and proportions of categorical responses for the five EQ-5D dimensions. A health state index score will be calculated from individual health profiles using the German value set (11). Mean and standard deviation (SD) scores will be provided for the study population for every measurement point. Patients' self-rated health can be recorded on the EQ-VAS,



which can range between “best imaginable health” (100) and “worst imaginable health” (0). The EQ-VAS score will be summarized using mean, SD, minimum, median and maximum scores. The EQ-5D is amongst the most commonly used instruments to assess HRQOL with evidence for good psychometric properties in different populations including the hip fracture population (12-14).

Reference measures

The Oxford Hip Score (OHS) is a hip-specific questionnaire assessing activities of daily living (ADL). The 12-item questionnaire refers to the last four weeks and contains six questions relating to pain and six questions relating to functional ability. Each item has five answer levels, ranging from 0 (worst outcome) to 4 (best outcome). A sum score of all values can be provided and ranges from 0 (worst outcome) to 48 (best outcome). Psychometric properties (reliability and validity) of the OHS have been assessed in a German population with osteoarthritis of the hip (15).

The Barthel Index (Version of Heuschmann *et al* (3)) is a common measure of ADL and comprises ten domains, divided in self-care and mobility. The interview version (T0) comprises 17 questions, the postal survey (T1) includes 10 questions, one for each domain (16). The overall score ranges from 0 (lowest ADL) to 20 (complete independence). Comprehensibility and reliability of both questionnaires were tested in patients with stroke (3).

The Index for Measuring Participation Restrictions (IMET) measures the restriction of participation of people with different chronic diseases (5) based on the International Classification of Functioning, Disability and Health. The questionnaire contains nine items which cover nine aspects of everyday activities and participation. The impairments are evaluated by means of a of an 11-level scale, where 0 indicates “no impairment” and 10 indicates “no activity possible”. A summary score can be calculated. The psychometric properties have been tested for different patient groups (5) and norm values enable classification of limitations of participation for people undergoing rehabilitation or patients suffering from chronic diseases (17).

The Oslo Social Support Scale will be used to capture social support (18) in three items. The items can be scored from 1 to 4 or 5. A total score between 3-14 can be obtained, where higher values indicate higher social support.

Definition of analysis sets

All primary analyses are performed with the main analysis set (self-report at 3 months). Further analyses are possible if data are available and sample sizes are sufficient (see table 2).



Table 2: Analysis sets

Assessment	Self-Report	<i>Self-Report Short Version</i>	<i>Proxy-Report</i>
3 months	Main analysis set	<i>Sub-analyses</i>	<i>Sub-analyses</i>
6 months	2 nd analysis set	<i>Sub-analyses</i>	<i>Sub-analyses</i>

Psychometric analysis

A lack of clarity in the literature about terminology and definitions of measurement properties led to the founding of the COSMIN (Consensus-based Standards for the selection of health Measurement Instruments) initiative. COSMIN is an initiative of an international multidisciplinary team with the aim to improve the quality of studies on measurement properties. Therefore, COSMIN develops methodology and practical tools for assessing measurement properties. The analyses in this study are based on the recommendations of the COSMIN group (19, 20).

Validity

In general, there are three different types of validity to be distinguished: content validity, criterion validity and construct validity. Since there is no 'gold standard' against which patient reported outcome measures (PROMs) can be compared to (criterion validity), the commonly used way is to examine construct validity by assessing structural validity and testing various hypotheses. Three different aspects are often considered: (1) convergent, (2) discriminant and (3) known-groups validity.

Construct validity

Evidence for construct validity can be determined by examining the strength of correlations that should exist with measures of related concepts by testing a priori hypotheses (discriminant and convergent validity). Evidence of high correlation with a measure of a related construct provides convergent validity, whereas evidence of low correlation with an unrelated construct provides evidence for discriminant validity. Furthermore a priori hypotheses for a logical relationship that should exist between diverse patient groups should be conformed to examine the degree to which the PRO instrument can distinguish among groups (known groups validity).

Construct validity: discriminant and convergent validity

A priori hypotheses were defined on the basis of scientific experience and the consistency of the contents of the questionnaires. These hypotheses describe to what extent a correlation of the respective instrument with reference measures is expected. Scales of instruments that pretend to



measure the same construct should correlate higher (convergent validity) than instruments with different underlying constructs (discriminant validity). Using the Pearson's correlation coefficient the strength of the correlation is divided into three categories (21):

- weak (+): $r < 0.3$
- moderate (++) : $r = 0.3-0.7$
- strong (+++) : $r > 0.7$

Correlations with instruments measuring similar constructs should be strong (+++). Correlations with instruments measuring related, but dissimilar constructs should be lower (++) and correlations with instruments measuring unrelated constructs should be weak (+). Table 3 shows the pre-defined a priori hypotheses on the correlations between the respective scales/items.

Table 3: Hypotheses for convergent and discriminant construct validity

Dimensions and index	SF-12		OHS	Barthel	OSS-3	IMET
	PCS-12	MCS-12				
EQ-5D: Mobility	++	+	++	++	+	++
EQ-5D: Self-care	++	+	++	++	+	++
EQ-5D: Usual activities	++	++	++	+++	+	++
EQ-5D: Pain/discomfort	++	+	+++	++	+	++
EQ-5D: Anxiety/depression	+	+++	+	+	++	+
EQ-5D-5L index	++	++	++	++	+	++
EQ-VAS	++	++	++	++	+	++
MCS-12	++	/	++	+	++	+
PCS-12	/	++	+++	+++	+	++

Construct validity: known-groups validity (discriminative validity)



The known-groups method is based on the principle that the average score of a pre-defined group should be expected to differ from another group. A valid instrument should be sensitive to these differences and be able to indicate them in the predicted direction. In this case, study participants will be grouped by age (<80 vs. ≥80), gender (male vs. female), socioeconomic status and support. In both quality of life questionnaires, as expected, older patients (++), female patients (+), patients with a lower socioeconomic status (+), and patients with need for support (++) should report significantly lower quality of life.

Depending on the score distributions, differences between groups will be assessed by using the Kruskal-Wallis-Test for variables with more than two categories and the Wilcoxon-Mann-Whitney test for variables with two categories (gender). Known-groups validity will be evaluated mainly in terms of the magnitude of the mean or median differences and the statistical significance. The level of significance is 5 % (one-sided).

Structural validity (dimensionality)

Structural validity will be assessed by a CFA verifying the a priori scales of the SF-12 and the unidimensional structure of the EQ-5D. The following goodness-of-fit indices of the CFA model will be used (based on COSMIN guideline): comparative fit index (CFI) ≥ 0.95 , root mean square error approximation (RMSEA) < 0.08 standardized root mean square residual (SRMR) < 0.10 and Chi square/df ration (χ^2/df) < 5 .

Reliability

Test-retest reliability

Reliability in the sense of "repeatability", describes the differences between repeated measurements over time, based on the analysis of correlations between different measurement points. Therefore, test-retest reliability will be assessed to examine stability of scores over time when no change is expected/observed (anchor question) in general quality of life. Therefore, in stable patients, i.e. patients whose status has not changed between the survey periods of 3 and 6 months, the test-retest reliability of the SF-12 and EQ-5D should be determined between these two survey periods.

For this purpose, the intraclass correlation coefficient (ICC) will be calculated. For continuous data, the ICC measures the strength of agreement between the repeated measurements. If the ICC is large (i.e. close to 1), the random variance is small, and a large proportion of the variance is due to variation between patients. A reasonable test-retest-reliability is assumed from an ICC > 0.80 (22), the instruments are then called reliable. The ICC can be determined by analysis of variance (ANOVA).



Internal consistency

To examine the extent to which items comprising a scale measure the same concept (internal consistency), Cronbach's α is calculated. A value of > 0.7 is often used in the literature as a threshold for the assumption of a consistent scale (COSMIN Manual). In addition, Item-Total Correlations will be calculated. Cronbach's α will be calculated for the PCS-12 and MCS-12 subscales of the SF-12. The calculation of the Cronbach α does not make sense in the EQ-5D, since each dimension is represented by only one item.

Responsiveness

Responsiveness, i.e. "the ability of an instrument to detect change over time in the construct to be measured" (19), can be determined by two different approaches: (1) the criterion approach and (2) the construct approach. In order to generate evidence that the EQ-5D or SF-12 instrument can identify differences in scores over time in individuals who have changed with respect to quality of life ES statistics and standardized response mean (SRM) were calculated historically. However, as responsiveness is a form of longitudinal validity, the COSMIN panel recommends assessment of responsiveness in a similar way in which validity testing is performed. They describe the difference between validity and responsiveness as follows: validity refers to the validity of a single score, whereas responsiveness refers to the validity of a change score. They also emphasize that one can never conclude that an instrument IS responsive, but can only accumulate evidence.

(1) Criterion approach

As gold standards for PROs are very rare, we will use a global rating of change scale (GRCS) as a gold standard for measuring change. As there is some concern about the reliability and validity of such a retrospective measure of change, some authors consider assessing responsiveness using a GRCS to be a construct approach rather than a criterion approach. However, COSMIN considers a GRCS to be a suitable criterion if the GRCS measures the same construct as the instrument under study. Therefore, in this psychometric study we will use the GRCS scale as criterion approach. Depending on the level of measurement, several statistical methods can be used for the criterion approach.

We will first calculate correlations between changes on the HRQOL measures and the GRCS (Spearman's ρ correlation). Secondly, we will look at mean changes in the HRQOL scores for categories indicated on the GRCS. In doing so, we will create five categories:

- Much improved: 5, 4
- Slightly improved: 3, 2, 1
- No change: 0



- Slightly deteriorated: -1, -2, -3
- Much deteriorated: -4, -5

Finally, we will do a receiver operating characteristic (ROC) analysis, considering patients who show any improvement at all on the GRS as “improved”, those reporting no change as “stable”, those who reported deterioration will be excluded for this analysis. An area under the curve (AUC) of at least 0.70 will be considered as appropriate.

(2) Construct approach

In the construct approach the assessment of responsiveness relies on testing hypotheses. The hypotheses can concern expected correlations between changes in scores on the instrument and changes in scores on other instruments known to have adequate responsiveness. We will test a priori formulated hypotheses. We will confirm our hypotheses based on the direction and magnitude of the observed correlation, rather than relying on p values of the correlations. The instrument will be considered responsive if less than 25% of the hypotheses will be rejected.

Limitations: One should be cautious when evaluating the responsiveness of a measurement instrument in the same study in which the instrument is used as an outcome measure.

Interpretability

Interpretability is not a measurement property, because it does not refer to the quality of an instrument but refers to what the scores or change in the scores mean. However, it is an important attribute and is therefore included in the COSMIN taxonomy.

Single scores

As the distribution of the scores affects the values of several measurement properties, we will provide tables with a simple presentation of means and SDs of the scores or median values and interquartile ranges (IQR) for continuous variables, and the numbers (%) in the various classes for ordinal or nominal variables. In addition, we will provide a histogram or some type of other visual presentation to get a clear insight into the distribution. We will also check for floor and ceiling effects, as this would affect the responsiveness of the respective instrument. Another option will be to compare the results with the values of the norm population.

Change scores

When interpreting change scores the smallest detectable change (SDC) and the minimal important change (MIC) are of special interest.



Smallest detectable change is based on measurement error and is defined as a change that goes beyond measurement error (e.g. falls outside the limits of agreement of the Bland and Altman method: $\pm 1.96 \times SD_{\text{difference}}$). So, a change in scores within the limits of agreement or smaller than the SDC can be attributed to measurement error, and only outside the limits of agreement we can be confident these are statistically significant changes. For interpreting the change scores the assessment of measurement error based on a test-retest parameter is required. For this reason, we can only calculate the SDC if the N is large enough to calculate test-retest-reliability.

The COSMIN panel defines the minimal important change as “the smallest change in score in the construct to be measured which patients perceive as important.” Currently, there is no consensus on the best method to determine MIC. In addition, the term takes on many meanings, there are 86 unique terms referring to the concept. We will use both, the distribution-based and anchor-based approach, respectively the integrated approach.

Anchor based

Using the mean change method the MIC is defined as the mean change in a score on the measurement instrument in the subcategory of patients who are minimally importantly changed, according to the anchor. Another method would be the receiver operating characteristic (ROC) method. In this method, the health status measurement instrument at issue is considered as the diagnostic test, and the anchor functions as the gold standard and distinguishes patients with important improvement or deterioration from patients with no important change. However, cut off values would be needed for the SF-12 and EQ-5D in order to assess if the instruments correctly identify improved or deteriorated patients. Therefore, we will only apply the mean change method. Using the anchor-based approach, the mean score change in the slightly improving group will provide the MIC estimate.

Distribution based approach

This approach is based on distributional characteristics of the sample and expresses the observed change in the measurement instrument under study to some form of variation to obtain a standardized metric. As parameter, we will use the effect size, a parameter that relates the observed change to the sample variability. Using the distribution-based approach, the half standard deviation of the SF-12/EQ-5D at baseline will be used as a MIC estimate.

The COSMIN group designed a method that integrates both approaches, which is called the *visual anchor-based MIC distribution* (De Vet et al. 2007). We will also apply this approach, following the three steps to assess the MICs for improvement and for deterioration separately.



Statistics

Continuous data will be summarized descriptively, for example mean and standard deviation (SD). Categorical variables will be summarized by absolute and relative frequencies (percentages) of the patients by categories. Missing values will not be imputed, a complete case analysis will be conducted.

References

1. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2011;20(10):1727-36.
2. WARE JEJ, KOSINSKI M, KELLER SD. A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. *Medical Care*. 1996;34(3):220-33.
3. Heuschmann P, Kolominsky-Rabas P, Nolte C, Hünermund G, Ruf H, Laumeier I, et al. The reliability of the german version of the barthel-index and the development of a postal and telephone version for the application on stroke patients. *Fortschritte der Neurologie-Psychiatrie*. 2005;73(2):74-82.
4. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *The Journal of bone and joint surgery British volume*. 1996;78(2):185-90.
5. Deck R, Mittag O, Hüppe A, Muche-Borowski C, Raspe H. Index zur Messung von Einschränkungen der Teilhabe (IMET)—Erste Ergebnisse eines ICF-orientierten Assessmentinstruments. *Praxis Klinische Verhaltensmedizin und Rehabilitation*. 2007;76(2):113-20.
6. Kocalevent R, Brähler E. OSSS Oslo Social Support Scale. *Psychologische und sozialwissenschaftliche Kurzskaalen: standardisierte Erhebungsinstrumente für Wissenschaft und Praxis*, Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft. 2014:216-24.
7. Hoebel J, Müters S, Kuntz B, Lange C, Lampert T. Measuring subjective social status in health research with a German version of the MacArthur Scale. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*. 2015;58(7):749-57.
8. Bullinger M, Kirchberger I. SF-36: Fragebogen zum Gesundheitszustand; Handanweisung: Hogrefe, Verlag für Psychologie; 1998.
9. Jakobsson U, Westergren A, Lindskov S, Hagell P. Construct validity of the SF-12 in three different samples. *Journal of evaluation in clinical practice*. 2012;18(3):560-6.
10. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Annals of medicine*. 2001;33(5):337-43.
11. Ludwig K, Graf von der Schulenburg JM, Greiner W. German Value Set for the EQ-5D-5L. *Pharmacoeconomics*. 2018;36(6):663-74.
12. Tidermark J, Bergstrom G. Responsiveness of the EuroQol (EQ-5D) and the Nottingham Health Profile (NHP) in elderly patients with femoral neck fractures. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2007;16(2):321-30.
13. Tidermark J, Bergstrom G, Svensson O, Tornkvist H, Ponzer S. Responsiveness of the EuroQol (EQ 5-D) and the SF-36 in elderly patients with displaced femoral neck fractures. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2003;12(8):1069-79.
14. Hinz A, Kohlmann T, Stobel-Richter Y, Zenger M, Brahler E. The quality of life questionnaire EQ-5D-5L: psychometric properties and normative values for the general German population. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2014;23(2):443-7.



15. Naal FD, Sieverding M, Impellizzeri FM, von Knoch F, Mannion AF, Leunig M. Reliability and Validity of the Cross-Culturally Adapted German Oxford Hip Score. *Clinical Orthopaedics and Related Research*. 2009;467(4):952-7.
16. Mahoney FI, Barthel DW. FUNCTIONAL EVALUATION: THE BARTHEL INDEX. *Maryland state medical journal*. 1965;14:61-5.
17. Deck R, Walther A, Staupendahl A, Katalinic A. limitations of social participation in general population-normative data of the IMET based on a population-based survey in northern Germany. *Die Rehabilitation*. 2015;54(6):402-8.
18. Dalgard O. Community health profile as tool for psychiatric prevention. *Promotion of mental health*. 1996;5:681-95.
19. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology*. 2010;63(7):737-45.
20. de Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: a practical guide*: Cambridge University Press; 2011.
21. McDowell I. *Measuring health: a guide to rating scales and questionnaires*: Oxford University Press, USA; 2006.
22. DeVellis RF. *Scale development: Theory and applications*: Sage publications; 2016.

7.2.2 R-Code

R Script Psychometrische Analyse: EQ-5D and SF-12 in elderly hip fracture patients

```
#1. (install) & load packages
#2. load & save data
#3. clean & prepare data (Variablen ableiten)
#4. Filling Table "Distribution of EQ-5D responses at baseline"
#5. Filling Table "Distribution of SF-12 responses at baseline"
#6. Filling Tables "Confirmatory Factor Analysis"
#7. Filling Table "Convergent and Discriminant Validity"
#8. Filling Table "Known-groups construct validity EQ-5D and SF-12"
```

```
#####
#####
#1 (install) & load packages
#####
#####
```

```
install.packages ("Rtools")
install.packages("sjmisc") # 3. for row_count
install.packages ("eq5d") # 5. for eq5d
install.packages ("Rmisc")
install.packages ("Hmisc") # 15.
install.packages ("psych") # 6. for alpha
install.packages ("moments") # 7. skewness
install.packages ("questionr")
install.packages ("sem")
install.packages ("lavaan") # 9. cfa # 11. cfa
install.packages ("semTools") # 14.
install.packages ("semPlot")
install.packages ("ggplot2") # 12.-13-
install.packages ("rstatix") # 16.
install.packages ("ggpubr")
install.packages("haven") # R open SAS files
install.packages("dplyr") # 2. for na_if # 10. select #12.
install.packages ("MVN")
install.packages ("eq5d")
```

```
library(plyr)
library(dplyr)
library(tidyr)
library(survival)
library(tidyverse)
library(eq5d)
library(sjmisc)
```

```
library(Rmisc)
library(Hmisc)
library(psych)
library(moments)
library(questionr)
library(sem)
library(lavaan)
library(semTools)
library(semPlot)
library(ggraph)
library(ggplot2)
library(rstatix)
library(ggpubr)
library(MVN)
library(haven)
library(eq5d)

#####
#####
# 2. load & save data
#####
#####

# open csv
#probe_ds <- read.csv2 (file.choose()) #Probedatensatz ink. SF-12 Summenscore

#open SAS files
#probe_ds <- read_sas (file.choose())

# pruefen, ob Datensatz richtig eingelesen?
summary(probe_ds)
head (probe_ds)
dim(probe_ds)
str(probe_ds)
View(probe_ds)

#####
#####
#3. clean & prepare data (Variablen ableiten)
#####
#####

# Variablen (ausser grcs) Ausprägungen -1 auf NA setzen
probe_ds_na <- probe_ds %>%
```

```

select("-t0_grcs")
probe_ds_na <- na_if (probe_ds_na, -1)

ds <- left_join (probe_ds_na, probe_ds) # merge na-Datensatz und unveränderte grcs-
Variable

#####
# create OHS sum score (Zeitpunkt t0)

# neue variablen erstellen, Anzahl missings pro row, Mittelwert und Summe
ds <- row_count (ds, t0_ohs_01:t0_ohs_12, count= NA, var ="ohs_na", append= TRUE)

ds <- row_means(ds, t0_ohs_01:t0_ohs_12, n=1, var ="ohs_mean", append = TRUE)

ds <- row_sums(ds, t0_ohs_01:t0_ohs_12, n=1, var ="ohs_sum", append = TRUE)

# Score berechnen
ds <- ds %>%
  mutate (ohs_score_t0 = ohs_sum + (ohs_na * ohs_mean))

#####
# calculate OSS-3 sum scores

ds <- ds %>%
  rowwise() %>%
  mutate (oss_score_t0 = sum(t0_oss_01, t0_oss_02, t0_oss_03, na.rm = FALSE))

#####
# calculate IMET sum scores
ds <- ds %>%
  rowwise() %>%
  mutate (imet_score_t0 = sum(t0_imet_01, t0_imet_02, t0_imet_03, t0_imet_04, t0_imet_05,
t0_imet_06, t0_imet_07, t0_imet_08, t0_imet_09, na.rm = TRUE))

#####
# create eq5d scores
# rename variables
ds <- dplyr::rename(ds, MO = t0_eq5d_01) #For Mobility
ds <- dplyr::rename(ds, SC = t0_eq5d_02) #For Self Care
ds <- dplyr::rename(ds, UA = t0_eq5d_03) #For Usual Activities
ds <- dplyr::rename(ds, PD = t0_eq5d_04) #For Pain/discomfort
ds <- dplyr::rename(ds, AD = t0_eq5d_05) #For Anxiety/Depression

# Calculate EQ-5D index scores

```

```

valuesets(country="Germany") #EQ-5D-5L Type VT

ds$eq5d_index_t0 <- eq5d(ds, type="VT", version="5L", country="Germany",
ignore.incomplete = TRUE)

#####
#####
#4. Filling Table "Distribution of EQ-5D responses at baseline"
#####
#####

#EQ-5D index
summary(ds$eq5d_index_t0, na.rm=TRUE)
sd(ds$eq5d_index_t0, na.rm=TRUE)
t.test(ds$eq5d_index_t0) # for CI

# EQ-5D item response level
ds$MO_f <- factor(ds$MO) # dafuer Antworten als factor abspeichern
ds$SC_f <- factor(ds$SC)
ds$UA_f <- factor(ds$UA)
ds$PD_f <- factor(ds$PD)
ds$AD_f <- factor(ds$AD)

questionr::freq(ds$MO_f)
questionr::freq(ds$SC_f)
questionr::freq(ds$UA_f)
questionr::freq(ds$PD_f)
questionr::freq(ds$AD_f)

# EQ-VAS
summary(ds$t0_eq5d_vas_06)
sd(ds$t0_eq5d_vas_06, na.rm=TRUE)
t.test(ds$t0_eq5d_vas_06) # for CI

#####
#####
#5. Filling Table "Distribution of SF-12 responses at baseline"
#####
#####

# PCS12 => KSK12 (physical bzw. körperliche) und MCS12 = PSK (mental bzw. psychische)

# vorher: invers kodierte Items rekodieren

```

```
#install.packages("car")
library(car)

ds$t0_ghp1_rekodiert <- recode(ds$t0_sf12_01_ghp1, "1=5; 2=4; 3=3; 4=2; 5=1")
ds$t0_pain2_rekodiert <- recode (ds$t0_sf12_08_pain2, "1=5; 2=4; 3=3; 4=2; 5=1")
ds$t0_mhi3_rekodiert <- recode (ds$t0_sf12_09_mhi3, "1=6; 2=5; 3=4; 4=3; 5=2; 6=1")
ds$t0_vital2_rekodiert <- recode (ds$t0_sf12_10_vital2,"1=6; 2=5; 3=4; 4=3; 5=2; 6=1")

# für körperliche Summenskala
summary (ds$t0_KSK12)
sd(ds$t0_KSK12, na.rm= TRUE)
t.test (ds$t0_KSK12) # for CI
alpha (ds [c("t0_ghp1_rekodiert", "t0_sf12_02_pfi02", "t0_sf12_03_pfi04",
"t0_sf12_04_rolph2", "t0_sf12_05_rolph3", "t0_pain2_rekodiert")], check.keys=TRUE,
warnings=FALSE)
skewness(ds$t0_KSK12, na.rm=TRUE)

# für psychische Summenskala
summary (ds$t0_PSK12)
sd(ds$t0_PSK12, na.rm=TRUE)
t.test (ds$t0_PSK12) # for CI
alpha (ds [c("t0_sf12_06_rolem2", "t0_sf12_07_rolem3", "t0_mhi3_rekodiert",
"t0_vital2_rekodiert", "t0_sf12_11_mhi4", "t0_sf12_12_soc2")], check.keys=TRUE,
warnings=FALSE)
skewness(ds$t0_PSK12, na.rm=TRUE)

# pro Item
# Item 1: ghp1
questionr::freq(ds$t0_ghp1_rekodiert)
mean(ds$t0_ghp1_rekodiert, na.rm=TRUE)
sd(ds$t0_ghp1_rekodiert, na.rm=TRUE)
t.test(ds$t0_ghp1_rekodiert)

# Item 2: pfi02
questionr::freq(ds$t0_sf12_02_pfi02)
mean(ds$t0_sf12_02_pfi02, na.rm=TRUE)
sd(ds$t0_sf12_02_pfi02, na.rm=TRUE)
t.test(ds$t0_sf12_02_pfi02)

# Item 3:
questionr::freq(ds$t0_sf12_03_pfi04)
mean(ds$t0_sf12_03_pfi04, na.rm=TRUE)
sd(ds$t0_sf12_03_pfi04, na.rm=TRUE)
t.test(ds$t0_sf12_03_pfi04)
```

Item 4:

```
questionr::freq(ds$t0_sf12_04_rolph2)
mean(ds$t0_sf12_04_rolph2, na.rm=TRUE)
sd(ds$t0_sf12_04_rolph2, na.rm=TRUE)
t.test(ds$t0_sf12_04_rolph2)
```

Item 5:

```
questionr::freq(ds$t0_sf12_05_rolph3)
mean(ds$t0_sf12_05_rolph3, na.rm=TRUE)
sd(ds$t0_sf12_05_rolph3, na.rm=TRUE)
t.test(ds$t0_sf12_05_rolph3)
```

Item 6:

```
questionr::freq(ds$t0_sf12_06_rolem2)
mean(ds$t0_sf12_06_rolem2, na.rm=TRUE)
sd(ds$t0_sf12_06_rolem2, na.rm=TRUE)
t.test(ds$t0_sf12_06_rolem2)
```

Item 7:

```
questionr::freq(ds$t0_sf12_07_rolem3)
mean(ds$t0_sf12_07_rolem3, na.rm=TRUE)
sd(ds$t0_sf12_07_rolem3, na.rm=TRUE)
t.test(ds$t0_sf12_07_rolem3)
```

Item 8:

```
questionr::freq(ds$t0_pain2_rekodiert)
mean(ds$t0_pain2_rekodiert, na.rm=TRUE)
sd(ds$t0_pain2_rekodiert, na.rm=TRUE)
t.test(ds$t0_pain2_rekodiert)
```

Item 9:

```
questionr::freq(ds$t0_mhi3_rekodiert)
mean(ds$t0_mhi3_rekodiert, na.rm=TRUE)
sd(ds$t0_mhi3_rekodiert, na.rm=TRUE)
t.test(ds$t0_mhi3_rekodiert)
```

Item 10:

```
questionr::freq(ds$t0_vital2_rekodiert)
mean(ds$t0_vital2_rekodiert, na.rm=TRUE)
sd(ds$t0_vital2_rekodiert, na.rm=TRUE)
t.test(ds$t0_vital2_rekodiert)
```

Item 11:

```
questionr::freq(ds$t0_sf12_11_mhi4)
mean(ds$t0_sf12_11_mhi4, na.rm=TRUE)
sd(ds$t0_sf12_11_mhi4, na.rm=TRUE)
t.test(ds$t0_sf12_11_mhi4)
```

```
# Item 12:
```

```
questionr::freq(ds$t0_sf12_12_soc2)
mean(ds$t0_sf12_12_soc2, na.rm=TRUE)
sd(ds$t0_sf12_12_soc2, na.rm=TRUE)
t.test(ds$t0_sf12_12_soc2)
```

```
#####
```

```
# Histogramm KSK12
```

```
hist(ds$t0_KSK12, main="Verteilung Körperliche Summenskala", xlab="PCS-12 Score",
ylab="Anzahl", col="deepskyblue", breaks=seq(0,60,length=60))
box()
```

```
# Histogram mit Normalverteilungskurve
```

```
x <- ds$t0_KSK12
hist(x, main="Verteilung Körperliche Summenskala", xlab="PCS-12 Score", ylab="Anzahl",
col="deepskyblue", breaks=seq(0,60,length=60), freq=FALSE)
m <- mean(x, na.rm=TRUE)
s <- sd(x, na.rm=TRUE)
curve(dnorm(x,m,s), add=TRUE, lwd=2, col="black")
box()
```

```
# Histogramm PSK12
```

```
hist(ds$t0_PSK12, main="Verteilung Psychische Summenskala", xlab="MCS-12 Score",
ylab="Anzahl", col="tomato2", breaks=seq(0,80,length=60))
box()
```

```
# Histogram mit Normalverteilungskurve
```

```
x <- ds$t0_PSK12
hist(x, main="Verteilung Psychische Summenskala", xlab="MCS-12 Score", ylab="Anzahl",
col="tomato2", breaks=seq(0,80,length=60), freq=FALSE)
m <- mean(x, na.rm=TRUE)
s <- sd(x, na.rm=TRUE)
curve(dnorm(x,m,s), add=TRUE, lwd=2, col="black")
box()
```

```
#####
```

```
#####
```

```
#6. Filling Tables "Confirmatory Factor Analysis"
```

```
#(1 Tabelle Indizes der Modellanpassung; 1 Tabelle Items und Itemkennwerte EQ-5D, 1
Tabelle Items und Itemkennwerte SF-12)
```

```
#####
#####
```

```
#EQ-5D
```

```
#Voraussetzung CFA
```

```
cfa_vs_eq5d <- select (ds, MO, SC, UA, PD, AD) # lineare Zusammenhänge
```

```
pairs(cfa_vs_eq5d, panel=panel.smooth)
```

```
MardiaTest_eq <- mvn(data = cfa_vs_eq5d, mvnTest = "mardia", univariateTest = "SW")
```

```
MardiaTest_eq$multivariateNormality
```

```
# for multivariate normality, both p-values of skewness and kurtosis statistics should be greater
than 0.05.
```

```
# no multivariate normal distribution --> therefore using MLR estimation in the CFA to take
into account the non-normality of data
```

```
# all effect indicator model
```

```
model_eq5d = "
```

```
  eq5d_fit =~ MO + SC + UA + PD + AD
```

```
"
```

```
eq5d_complete <- ds %>%
```

```
  dplyr::filter (!is.na(eq5d_index_t0))
```

```
eq5d_fit_ml <- cfa (model_eq5d, data= eq5d_complete)
```

```
summary(eq5d_fit_ml, fit.measures=TRUE, standardized = TRUE)
```

```
fitmeasures(eq5d_fit_ml, c('cfi', 'rmsea', 'rmsea.ci.upper', 'bic'))
```

```
# alternativer Estimator
```

```
eq5d_fit_wlsmv <- cfa (model_eq5d, data= eq5d_complete, estimator = "wlsmv")
```

```
summary (eq5d_fit_wlsmv, standardized = TRUE)
```

```
#
```

```
semPlot::semPaths(eq5d_fit, what="paths", whatLabels = "par",
```

```
  nCharNodes = 0, # keine Abkürzung im Kreis
```

```
  style= "lisrel", # gerader Pfeil statt Kreis
```

```
  residScale = 8, # Pfeilgröße
```

```
  rotation=2) # Querformat
```

```
#SF-12
```

```

#Voraussetzung CFA
# lineare Zusammenhänge

ds$t0_ghp1_rekodiert <- as.numeric(ds$t0_ghp1_rekodiert)
cfa_vs_sf12 <- select (ds, t0_ghp1_rekodiert, t0_sf12_02_pfi02, t0_sf12_03_pfi04,
t0_sf12_04_rolph2, t0_sf12_05_rolph3, t0_pain2_rekodiert,
t0_sf12_06_rolem2, t0_sf12_07_rolem3, t0_mhi3_rekodiert,
t0_vital2_rekodiert, t0_sf12_11_mhi4, t0_sf12_12_soc2)

pairs(cfa_vs_sf12, panel=panel.smooth)
MardiaTest <- mvn(data = cfa_vs_sf12, mvnTest = "mardia", univariateTest = "SW")
MardiaTest$multivariateNormality

# lavaan model specification
model_sf12 = "
PCS =~ t0_ghp1_rekodiert + t0_sf12_02_pfi02 + t0_sf12_03_pfi04 + t0_sf12_04_rolph2 +
t0_sf12_05_rolph3 + t0_pain2_rekodiert
MCS =~ t0_sf12_06_rolem2 + t0_sf12_07_rolem3 + t0_mhi3_rekodiert + t0_vital2_rekodiert
+ t0_sf12_11_mhi4 + t0_sf12_12_soc2
"

# fit cfa model
cfa.model_sf12 = cfa(model_sf12, cfa_vs_sf12, estimator = "MLR")
summary(cfa.model_sf12, fit.measures = T, standardized = T)

reliability(cfa.model_sf12)

#Graphik
semPlot::semPaths(cfa.model_sf12, what="paths", whatLabels = "par",
nCharNodes = 0, # keine Abkürzung im Kreis
style= "lisrel", # gerader Pfeil statt Kreis
residScale = 8, # Pfeilgröße
rotation=2) # Querformat

#####
#####
#7. Filling Table "Convergent and Discriminant Validity"
#####
#####

#plot (ds$KSK12, ds$PSK12)

# Pearsons r: metrisches Messniveau + Normalverteilung, ansonsten Spearman's rho

```

```

shapiro.test(ds$t0_KSK12)
shapiro.test(ds$t0_PSK12)

# wenn p>0.05, dann Annahme auf Normalverteilung

#library(dplyr)
cor_variables <- ds %>%
  select (c(t0_PSK12, t0_KSK12, MO, SC, UA, PD, AD, eq5d_index_t0, t0_eq5d_vas_06,
  ohs_score_t0, oss_score_t0, imet_score_t0, t0_bind_Score))

cor_pearson <- cor (cor_variables, use="pairwise.complete.obs", method = "pearson")
cor_spearman <- cor (cor_variables, use="pairwise.complete.obs", method = "spearman")

cor_pearson <- round (cor_pearson, 2)
cor_spearman <- round (cor_spearman, 2)

cor_pearson #print cor matrix
cor_spearman #print cor matrix

cor_observations <- rcorr(as.matrix(cor_variables))
cor_observations

#library(Hmisc)
cor_pvalue <- rcorr (as.matrix (cor_variables)) # list with r: the correlation matrix/ n: the matrix
of the number of observations used in analyzing each pair of variables/ P: the p-values
corresponding to the significance levels of correlations
cor_pvalue$r
cor_pvalue$P

flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
    row = rownames(cormat)[row(cormat)[ut]],
    column = rownames(cormat)[col(cormat)[ut]],
    cor =(cormat)[ut],
    p = pmat[ut]
  )
}

flattenCorrMatrix(cor_pvalue$r, cor_pvalue$P)

#####
#####
#8. Filling Table "Known-groups construct validity EQ-5D and SF-12"

```

```
# groups: gender, age, questionnaire version and socioeconomic status (self-reported)
#####
#####
# Geschlecht und EQ-5D index

ds$t0_hpax_01_geschlecht <- as.factor(ds$t0_hpax_01_geschlecht)

ds %>%
  dplyr::filter(!is.na(t0_hpax_01_geschlecht)) %>%
  group_by(t0_hpax_01_geschlecht) %>%
  summarise (
    count =n(),
    mean = mean (eq5d_index_t0, na.rm=TRUE),
    median = median (eq5d_index_t0, na.rm=TRUE),
    IQR = IQR (eq5d_index_t0, na.rm= TRUE),
    SD = sd (eq5d_index_t0, na.rm= TRUE)
  )

eq5d_bysex <- ds %>%
  dplyr::filter (!is.na(t0_hpax_01_geschlecht))

eq5d_by_sex_test <- wilcox.test (eq5d_index_t0 ~ t0_hpax_01_geschlecht, data = eq5d_bysex,
na.rm=TRUE, conf.int =TRUE)
print (eq5d_by_sex_test)

#####
# Geschlecht und EQ-5D-VAS
ds %>%
  dplyr::filter (!is.na(t0_hpax_01_geschlecht)) %>%
  group_by (t0_hpax_01_geschlecht) %>%
  summarise (
    count =n(),
    mean = mean (t0_eq5d_vas_06, na.rm=TRUE),
    median = median (t0_eq5d_vas_06, na.rm=TRUE),
    IQR = IQR (t0_eq5d_vas_06, na.rm= TRUE),
    SD = sd (t0_eq5d_vas_06, na.rm= TRUE)
  )

eq5d_bysex <- ds %>%
  dplyr::filter (!is.na(t0_hpax_01_geschlecht))
```

```
eq5dvas_by_sex_test <- wilcox.test (t0_eq5d_vas_06 ~ t0_hpax_01_geschlecht, data =
eq5d_bysex, na.rm=TRUE, conf.int =TRUE)
print (eq5dvas_by_sex_test)
```

```
#####
```

```
# Geschlecht und KSK-12
```

```
ds %>%
  dplyr::filter (!is.na(t0_hpax_01_geschlecht)) %>%
  group_by (t0_hpax_01_geschlecht) %>%
  summarise (
    count =n(),
    mean = mean (t0_KSK12, na.rm=TRUE),
    median = median (t0_KSK12, na.rm=TRUE),
    IQR = IQR (t0_KSK12, na.rm= TRUE),
    SD = sd (t0_KSK12, na.rm= TRUE)
  )
```

```
KSK12_bysex <- ds %>%
  dplyr::filter (!is.na(t0_hpax_01_geschlecht))
```

```
KSK12_by_sex_test <- wilcox.test (t0_KSK12 ~ t0_hpax_01_geschlecht, data =
KSK12_bysex, na.rm=TRUE, conf.int =TRUE)
print (KSK12_by_sex_test)
```

```
#####
```

```
# Geschlecht und PSK-12
```

```
ds %>%
  dplyr::filter (!is.na(t0_hpax_01_geschlecht)) %>%
  group_by (t0_hpax_01_geschlecht) %>%
  summarise (
    count =n(),
    mean = mean (t0_PSK12, na.rm=TRUE),
    median = median (t0_PSK12, na.rm=TRUE),
    IQR = IQR (t0_PSK12, na.rm= TRUE),
    SD = sd (t0_PSK12, na.rm= TRUE)
  )
```

```
PSK12_bysex <- ds %>%
  dplyr::filter (!is.na(t0_hpax_01_geschlecht))
```

```
PSK12_by_sex_test <- wilcox.test (t0_PSK12 ~ t0_hpax_01_geschlecht, data = PSK12_bysex,
na.rm=TRUE, conf.int =TRUE)
```

```
print (PSK12_by_sex_test)

#####
# Age --> 2 Gruppen: 60-80, 81-104 (Median=81)

ds <- ds %>%
  mutate(age_group=cut(t0_hpax_01_alter, breaks=c(0, 80, 105), labels=c("old","oldest old")))

#library(rstatix)
#age and EQ-5D index
ds %>%
  group_by(age_group) %>%
  get_summary_stats(eq5d_index_t0, type = "common")

eq5dage_wilcox <- ds %>%
  dplyr::filter ((!is.na(eq5d_index_t0)& !is.na(age_group))) %>%
  kruskal_test(eq5d_index_t0 ~ age_group)
eq5dage_wilcox

#####
#age and EQ-5D-VAS
ds %>%
  group_by(age_group) %>%
  get_summary_stats(t0_eq5d_vas_06, type = "common")

eq5dvasage_wilcox <- ds %>%
  dplyr::filter ((!is.na(t0_eq5d_vas_06)& !is.na(age_group))) %>%
  kruskal_test(t0_eq5d_vas_06 ~ age_group)
eq5dvasage_wilcox

#####
#age and KSK-12
ds %>%
  group_by(age_group) %>%
  get_summary_stats(t0_KSK12, type = "common")

KSK12age_wilcox <- ds %>%
  dplyr::filter ((!is.na(t0_KSK12)& !is.na(age_group))) %>%
  kruskal_test (t0_KSK12 ~ age_group)
KSK12age_wilcox

#####
```

```
#age and PSK-12
ds %>%
  group_by(age_group) %>%
  get_summary_stats(t0_PSK12, type = "common")

PSK12age_wilcox <- ds %>%
  dplyr::filter ((!is.na(t0_PSK12)& !is.na(age_group))) %>%
  kruskal_test (t0_PSK12 ~ age_group)
PSK12age_wilcox

#####
# Fragebogenversion lang/kurz als Annäherung an Allgemeinzustand des/der Befragten
#####
#FB Version and EQ-5D-index
ds %>%
  group_by(befragung_fbversion) %>%
  get_summary_stats(eq5d_index_t0, type = "common")

eq5dFB_wilcox <- ds %>%
  dplyr::filter ((!is.na(eq5d_index_t0)& !is.na(befragung_fbversion))) %>%
  wilcox.test(eq5d_index_t0 ~ befragung_fbversion)
eq5dFB_wilcox

#####
#FB Version and EQ-5D-VAS
ds %>%
  group_by(befragung_fbversion) %>%
  get_summary_stats(t0_eq5d_vas_06, type = "common")

eq5dvasFB_wilcox <- ds %>%
  dplyr::filter ((!is.na(t0_eq5d_vas_06)& !is.na(befragung_fbversion))) %>%
  wilcox.test(t0_eq5d_vas_06 ~ befragung_fbversion)
eq5dvasFB_wilcox

#####
#FB Version and KSK-12
ds %>%
  group_by(befragung_fbversion) %>%
  get_summary_stats(t0_KSK12, type = "common")

PCS12FB_wilcox <- ds %>%
  dplyr::filter ((!is.na(t0_KSK12)& !is.na(befragung_fbversion))) %>%
  wilcox.test (t0_KSK12 ~ befragung_fbversion)
```

```
PCS12FB_wilcox
```

```
#####
```

```
#FB Version and PSK-12
```

```
ds %>%
```

```
  group_by(befragung_fbversion) %>%
```

```
  get_summary_stats(t0_PSK12, type = "common")
```

```
MCS12FB_wilcox <- ds %>%
```

```
  dplyr::filter ((!is.na(t0_PSK12)& !is.na(befragung_fbversion))) %>%
```

```
  wilcox.test (t0_PSK12 ~ befragung_fbversion)
```

```
MCS12FB_wilcox
```

```
#####
```

```
# Socioeconomic status --> anhand SSS (MacArthurScale)
```

```
# kategorisiert in 3 Gruppen wie in Chen, 2012
```

```
# neue Variable Subjektiver Sozialer Status (SSS) mit 3 Kategorien (low/middle/high)
```

```
ds <- ds %>%
```

```
  mutate(SSS=cut(t0_leiter, breaks=c(0, 3, 7, 10), labels=c("low","middle","high")))
```

```
#####
```

```
#SSS and EQ-5D Index
```

```
ds %>%
```

```
  group_by(SSS) %>%
```

```
  get_summary_stats(eq5d_index_t0, type = "common")
```

```
eq5dSSS_kruskal <- ds %>%
```

```
  dplyr::filter ((!is.na(eq5d_index_t0)& !is.na(SSS))) %>%
```

```
  kruskal_test(eq5d_index_t0 ~ SSS)
```

```
eq5dSSS_kruskal
```

```
pairwise.wilcox.test(ds$eq5d_index_t0, ds$SSS,
```

```
  na.omit = TRUE, p.adjust.method = "BH")
```

```
#####
```

```
#SSS and EQ-5D-VAS
```

```
ds %>%
```

```
  group_by(SSS) %>%
```

```
  get_summary_stats(t0_eq5d_vas_06, type = "common")
```

```
eq5dvasSSS_kruskal <- ds %>%
```

```
dplyr::filter ((!is.na(t0_eq5d_vas_06)& !is.na(SSS))) %>%
  kruskal_test(t0_eq5d_vas_06 ~ SSS)
eq5dvasSSS_kruskal
```

```
pairwise.wilcox.test(ds$t0_eq5d_vas_06, ds$SSS,
  na.omit = TRUE, p.adjust.method = "BH")
```

```
#####
#SSS and KSK-12
```

```
ds %>%
  group_by(SSS) %>%
  get_summary_stats(t0_KSK12, type = "common")
```

```
KSK12SSS_kruskal <- ds %>%
  dplyr::filter ((!is.na(t0_KSK12)& !is.na(SSS))) %>%
  kruskal_test(t0_KSK12 ~ SSS)
KSK12SSS_kruskal
```

```
pairwise.wilcox.test(ds$t0_KSK12, ds$SSS,
  na.omit = TRUE, p.adjust.method = "BH")
```

```
#####
#SSS and PSK-12
```

```
ds %>%
  group_by(SSS) %>%
  get_summary_stats(t0_PSK12, type = "common")
```

```
PSK12SSS_kruskal <- ds %>%
  dplyr::filter ((!is.na(t0_PSK12)& !is.na(SSS))) %>%
  kruskal_test(t0_PSK12 ~ SSS)
PSK12SSS_kruskal
```

```
pairwise.wilcox.test(ds$t0_PSK12, ds$SSS,
  na.omit = TRUE, p.adjust.method = "BH")
```

```
# You can use the sink() function in R. It will divert all R console output between the first
# and second call of sink() to a file that you specify.
```

```
# ENDE:
```

```
sink(file = NULL)
```

Longitudinal psychometric analysis EQ-5D and SF-12: Sensitivity to change

#1. (install) & load packages

#2. load and save data

#3. Änderungssensitivität

#####

#####

#1. (install) & load packages

install.packages ("psych")

install.packages ("irr")

install.packages("pROC")

install.packages("randomForest")

install.packages ("coin")

library (dplyr)

library (tidyr)

library(psych)

library (irr)

library (questionr)

library(pROC)

library(randomForest)

library (rstatix)

library (coin)

#####

#####

#2. load and save data

probe_ds <- read.csv2 (file.choose()) #Datensatz auswählen

#####

#####

#3. Änderungssensitivität

für alle Berechnungen Change-Scores (cs) als neue Variable benötigt

#####

#zunächst Beschreibung der Change-Scores für jede GRCS-Kategorie

probe_ds %>%

group_by(grcs_2) %>%

get_summary_stats(eq_vas_cs, type = "common")

probe_ds %>%

```

group_by(grcs_2) %>%
  get_summary_stats(eq_index_cs, type = "common")

probe_ds %>%
  group_by(grcs_2) %>%
  get_summary_stats(PCS12_cs, type = "common")

probe_ds %>%
  group_by(grcs_2) %>%
  get_summary_stats(MCS12_cs, type = "common")

#####
# Kriteriumsbasierter Ansatz: Beschreibung der Scores anhand der GRCS Gruppierung

# Kategorien aus den Antwortmöglichkeiten der GRCS bilden
probe$ds <- probe$ds %>%
  mutate(grcs_cat=cut(t0_grcs, breaks=c(-6, -4, -1, 0, 3, 5), labels=c("stark
verschlechtert","leicht verschlechtert", "keine Veränderung", "leicht verbessert", "stark
verbessert")))

#####
# Wilcoxon signed rank test on paired samples --> Vergleich der Scores zwischen den beiden
Zeitpunkten
# Transform into long data: gather t0 und t1 values in the same column
ds.long <- ds %>%
  gather (key= "timing", value = "KSK_12_Score", t0_KSK12, t1_KSK12)

#compute summary statistics
ds.long %>%
  group_by(timing) %>%
  get_summary_stats(KSK_12_Score, type = "common")

stat.test <- ds.long %>%
  wilcox_test(KSK_12_Score ~ timing, paired = TRUE) %>%
  add_significance()
stat.test

ds.long %>%
  wilcox_effsize(KSK_12_Score ~ timing, paired = TRUE)

```

Danksagung

*„Nicht die Glücklichen sind dankbar. Es sind die Dankbaren, die glücklich sind.“
(Francis Bacon)*

Besonders in der heutigen Zeit wird mir stets aufs Neue bewusst, wie dankbar ich über die Möglichkeit zur Promotion bin und über meine Lebensumstände, die es mir erlaub(t)en diese Möglichkeit wahrzunehmen. Darüber hinaus haben mich viele Personen auf unterschiedliche Weise in diesem Vorhaben unterstützt:

Ein besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr. Dr. Christian Apfelbacher, für die Betreuung dieser Arbeit, aber vor allem auch für seine Begleitung über die letzten Jahre. Durch ihn lernte ich über diese Arbeit hinaus noch so viel mehr, wie zum Beispiel das Verstehen von Zusammenhängen und Mechanismen im „System Wissenschaft“, das Halten von wissenschaftlichen Vorträgen und das Publizieren. In diesem Zuge möchte ich mich auch bei meinen aktuellen und ehemaligen Arbeitskolleg:innen bedanken, die mich begleitet und immer unterstützt haben.

Ein großer Dank gilt auch meinen Mentorinnen Frau Prof. Dr. Haug und Frau Prof. Dr. Dr. Icks. Frau Prof. Dr. Haug übernimmt zudem die Zweitprüfung dieser Arbeit und Frau Prof. Dr. Dr. Icks ermöglichte durch die Bereitstellung von Daten aus der ProFem Studie die Dissertation in diesem Umfang. Auch dem Team von Frau Prof. Dr. Dr. Icks, welches mich so selbstverständlich unterstützte, möchte ich hiermit danken.

Des Weiteren möchte ich mich bei dem ehemaligen Team der Alterstraumatologie am St. Josef Krankenhaus (insbesondere Herrn Dr. Rainer Kretschmer und Frau Gerlinde Weigert) für die Unterstützung bei der Rekrutierung und den wertvollen Einblick in die Herausforderungen in der Versorgung der untersuchten Patient:innengruppe danken.

Zuletzt gilt ein besonderer Dank meiner Familie und meinen Freunden, für deren Verständnis, Unterstützung und Motivation.

Lebenslauf

Magdalena Rohr (geb. Brandl), MPH

[REDACTED]	[REDACTED]

Schulische und akademische Ausbildung

[REDACTED]	[REDACTED]

Berufstätigkeit

[REDACTED]	[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Lehrtätigkeiten

- | | |
|--------------------------|---|
| 10/2017 – 09/2020 | Universität Regensburg, Studiengang Humanmedizin <ul style="list-style-type: none">• Kurs der Medizinischen Soziologie• Seminar der Medizinischen Soziologie |
| Sommersemester 2019-2021 | Technische Hochschule Deggendorf, Studiengang Pflege Dual <ul style="list-style-type: none">• Blockseminar im Modul Pflegeversorgungsforschung |

Veröffentlichungen in Zeitschriften mit peer review-Verfahren

(Erstveröffentlichungen + Auswahl an Co-Autorenschaften)

Höroid, M., Drewitz, K. P., Piel, J., Hruday, I., Rohr, M., Brunthaler, V., ... & Apfelbacher, C. (2022). Intensive Care Units Healthcare Professionals' Experiences and Negotiations at the Beginning of the COVID-19 Pandemic in Germany: A Grounded Theory Study. INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 59, 00469580221081059.

Rohr, M., Brandstetter, S., Plomer, A. S., Loss, J., Kretschmer, R., & Apfelbacher, C. (2021). A qualitative study exploring content validity and feasibility of frequently used

generic health-related quality of life measures in older people with hip fracture: The patients' perspective. *Injury*, 52(2), 134-141.

Rohr, M., Brandstetter, S., Bernardi, C., Fisser, C., Drewitz, K. P., Brunthaler, V., ... & Apfelbacher, C. J. (2021). Piloting an ICU follow-up clinic to improve health-related quality of life in ICU survivors after a prolonged intensive care stay (PINA): study protocol for a pilot randomised controlled trial. *Pilot and feasibility studies*, 7(1), 1-13.

Brandl, M., Apfelbacher, C., Weiß, A., Brandstetter, S., & Baumeister, S. E. (2020). Incidence Estimation in Post-ICU Populations: Challenges and Possible Solutions When Using Claims Data. *Das Gesundheitswesen*.

Apfelbacher, C., Brandstetter, S., Blecha, S., Dodoo-Schittko, F., Brandl, M., Karagiannidis, C., ... & Bein, T. (2020). Influence of quality of intensive care on quality of life/return to work in survivors of the acute respiratory distress syndrome: prospective observational patient cohort study (DACAPO). *BMC Public Health*, 20(1), 1-9.

Blecha, S., Brandl, M., Zeman, F., Dodoo-Schittko, F., Brandstetter, S., Karagiannidis, C., ... & Apfelbacher, C. (2020). Tracheostomy in patients with acute respiratory distress syndrome is not related to quality of life, symptoms of psychiatric disorders or return-to-work: the prospective DACAPO cohort study. *Annals of Intensive Care*.

Brandl, M., Bernardi, C., & Apfelbacher, C. (2019). Einführung von patientenberichteten Endpunkten: erhöhen sie die Qualität von Behandlungen?. In *Public Health Forum*.

Brandstetter, S., Dodoo-Schittko, F., Brandl, M., Blecha, S., Bein, T., Apfelbacher, C., & DACAPO study group. (2019). Ambulatory and stationary healthcare use in survivors of ARDS during the first year after discharge from ICU: findings from the DACAPO cohort. *Annals of intensive care*.

Brandl, M., Böhmer, M. M., Brandstetter, S., Finger, T., Fischer, W., Pfeifer, M., & Apfelbacher, C. (2018). Factors associated with generic health-related quality of life (HRQOL) in patients with chronic obstructive pulmonary disease (COPD): a cross-sectional study. *Journal of thoracic disease*.

Dodoo-Schittko, F., Brandstetter, S., Brandl, M., Blecha, S., Quintel, M., Weber-Carstens, S., et al. (2017) Characteristics and provision of care of patients with the

acute respiratory distress syndrome: descriptive findings from the DACAPO cohort baseline and comparison with international findings. *Journal of thoracic disease*.

Brandstetter, S., Finger, T., Fischer, W., Brandl, M., Bohmer, MM., Pfeifer, M., Apfelbacher, C., et al. (2017). Differences in medication adherence are associated with beliefs about medicines in asthma and COPD. *Journal of Clinical and Translational Allergy*.

Wissenschaftliche Originalvorträge

Brandl, M., Brandstetter, S., Dodoo-Schittko, F., Blecha, S., Bein, T. & Apfelbacher, C. (2019). Soziale Unterstützung ist bei Überlebenden eines akuten Lungenversagens signifikant mit der psychischen Lebensqualität assoziiert: Ergebnisse der prospektiven Patientenkohortenstudie DACAPO. *Das Gesundheitswesen*, 81(08/09), 8E-1. Deutscher Kongress für Sozialmedizin und Prävention 2019, Düsseldorf. (Vortrag)

Brandl, M., Weiß, A., Bernardi, C., Brandstetter, S. & Apfelbacher, C. (2019) PINA: Perspektiven von Versorgungsakteuren zur Entwicklung einer Intensiv-Nachsorgeambulanz. German Medical Science GMS Publishing House. 18.; Deutscher Kongress für Versorgungsforschung (DKVF) 2019, Berlin (Vortrag)

Brandl, M., Brandstetter, S., Kretschmer, R., & Apfelbacher, C. (2019). Content validity and feasibility of frequently used health-related quality of life measures in older people with hip fracture: a qualitative study. *QUALITY OF LIFE RESEARCH*, 28(Suppl1), S154. (Postervortrag)

Brandl, M., Brandstetter, S., Kretschmer, R., Weigert, G., & Apfelbacher, C. (2018). „Also ich bin nicht recht überzeugt von der Richtigkeit dieser Fragebögen“: Qualitative Studie zu Inhaltsvalidität und Machbarkeit häufig verwendeter Fragebögen zur Messung der Lebensqualität bei Patienten über 65 Jahre mit Hüftfraktur. *Das Gesundheitswesen*, 80(08/09), V-65. Deutscher Kongress für Sozialmedizin und Prävention 2018, Dresden. (Vortrag)

Brandl, M., Brandstetter, S., Finger, T., Fischer, W., Pfeifer, M., & Apfelbacher, C. (2018). Measuring health-related quality of life in patients with chronic respiratory disease: reliability and construct validity of the SF-12. *QUALITY OF LIFE RESEARCH*, 27(Suppl1), S175. (Postervortrag)

Brandl, M., Dodoo-Schittko, F., Brandstetter, S., Blecha, S., Bein, T. & Apfelbacher, C. (2017). "Psychische Gesundheit, Lebensqualität und Rückkehr in das Arbeitsleben bei Überlebenden eines ARDS – 3 Monate nach dem Intensivaufenthalt." Gesundheitswesen 79(08/09): V-12. Deutscher Kongress für Sozialmedizin und Prävention 2017, Lübeck. (Vortrag)

Brandl, M., Brandstetter, S., Dodoo-Schittko, F., Blecha, S., Bein, T. & Apfelbacher, C. (2017). "Versorgungsqualität auf deutschen Intensivstationen – Ergebnisse der DACAPO-Studie." Gesundheitswesen 79(08/09): V-69. Deutscher Kongress für Sozialmedizin und Prävention 2017, Lübeck. (Vortrag)

Brandl, M., Böhmer, M., Brandstetter, S., Finger, T., Fischer, W., Pfeifer, M. & Apfelbacher, C. (2016). "Factors associated with health-related quality of life (HRQOL) in adults with chronic obstructive pulmonary disease (COPD)." Gesundheitswesen 78(08/09): A125. Deutscher Kongress für Sozialmedizin und Prävention 2016, Essen. (Vortrag)

Weiterbildung

- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]

[REDACTED]	[REDACTED]

Mitgliedschaften

seit 01/2018	Mitgliedschaft im BayWISS Verbundkolleg Gesundheit
seit 04/2018	Mitgliedschaft in der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP)

Gutachtertätigkeit

2019 & 2020	Reviews für Vorträge/Poster Abstracts ISOQOL Jahrestagung
2020	Review für Zeitschrift: "International Journal of Chronic Obstructive Pulmonary Disease"
2020	Review für Zeitschrift "COPD: Journal of Chronic Obstructive Pulmonary Disease"

Selbstständigkeitserklärung

Ich, Magdalena Rohr, geboren am 16.08.1991 in Straubing erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Insbesondere habe ich nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ort, Datum

Magdalena Rohr