

Evaluationsbericht (gemäß Nr. 14.1 ANBest-IF)

Konsortialführung:	Charité - Universitätsmedizin Berlin
Förderkennzeichen:	01NVF18024
Akronym:	PRÄP-GO
Projekttitel:	Prähabilitation von älteren Patienten mit Gebrechlichkeitssyndrom vor elektiven Operationen
Autorinnen und Autoren:	Prof. Dr. Ulrich Mansmann, Stefan Buchka, Dr. Verena Loidl, Prof. Dr. Reinhard Busse, Dr. Tanja Rombey, Lukas Schöner, Helene Eckhardt, Wilm Quentin
Förderzeitraum:	1. September 2019 - 31. August 2024
Ansprechpartner:	<p>Prof. Dr. Ulrich Mansmann mansmann@ibe.med.uni-muenchen.de Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie (IBE) LMU Marchioninistrasse 15 81377 München</p> <p>Prof. Dr. Reinhard Busse mig@tu-berlin.de Fachgebiet Management im Gesundheitswesen Technische Universität Berlin Straße des 17. Juni 135 10623 Berlin</p>

Das dieser Veröffentlichung zugrundeliegende Projekt PRÄP-GO wurde mit Mitteln des Innovationsausschusses beim Gemeinsamen Bundesausschuss unter dem Förderkennzeichen 01NVF18024 gefördert. Die Darstellungen im Evaluationsbericht sind das Ergebnis der unabhängigen Evaluation zur neuen Versorgungsform.

Zusammenfassung

Hintergrund: Immer mehr komplexe chirurgische Eingriffe bei älteren Patient:innen haben ein erhöhtes Risiko für postoperative Komplikationen durch das Vorliegen eines Gebrechlichkeitssyndroms. Durch eine Prähabilitation könnte bei gebrechlichen Patient:innen die Auswirkungen einzelner Komponenten des Gebrechlichkeitssyndroms reduziert werden. So könnten funktionale Fähigkeiten wiedererlangt werden, die die Häufigkeit postoperativer Komplikationen verringern und den Krankenhausaufenthalt verkürzen. Die meisten Studien sind auf bestimmte Patientengruppen oder Operationsarten beschränkt, während es an Belegen für eine generelle chirurgische Population mangelt.

Methodik: Die PRÄP-GO-Studie ist eine nationale multizentrische, assessor-verblindete, randomisierte, pragmatische, zweiarmige, kontrollierte klinische Studie. Eingeschlossen wurden Patient:innen im Alter über 70 Jahre mit geplanter elektiver Operation, die im Screening pre-frail oder frail waren. Primärer Endpunkt war die Veränderung des Pflegegrades nach dem neuen Begutachtungsassessment (NBA) zwischen Baseline und 12 Monaten nach Operation. Die Interventionsgruppe erhielt ein strukturiertes Prähabilitationsprogramm. Die Kontrollgruppe erhielt die Standardversorgung. Die primäre ITT-Analyse verwendet einen stratifizierten Mann-Whitney-Test. Per Protokoll Analysen wurden an relevanten Subgruppen durchgeführt (Adhärenz der Prähabilitation, OP erfolgt, Principal stratum). Weiterhin wurde ein Spektrum an sekundären Endpunkten untersucht und eine gesundheitsökonomische Evaluation aus drei Perspektiven durchgeführt.

Ergebnisse: Die primäre ITT-Analyse des primären Endpunktes zeigte keinen signifikanten Unterschied zwischen den Studienarmen (stratifizierter Wilcoxon-Test: $p = 0.3316$). Im Rahmen von explorativen Sensitivitäts- und Per-Protokoll-Analysen wurden signifikante Signale für eine positive Wirkung der Intervention gesehen, z.B. bei der Veränderung im numerischen NBA-Score (Differenz der Differenzen -1.716, 95% KI [-3.137; -0.295], $p=0.018$). Für keine sekundären Endpunkte zeigten sich in der ITT-Analyse signifikante Unterschiede ($p<0.05$). Die Sicherheit der Intervention wurde belegt. Aus der gesellschaftlichen Perspektive betrug die inkrementelle Kosten-Effektivitäts-Relation (IKER) 45.547 € pro Qualitätsadjustiertes Lebensjahr und die Kosten-Effektivitäts-Akzeptanz 52 % bei einer Zahlungsbereitschaft von 50.000 €. Aus der GKV-Perspektive betrug die IKER 111.115 € pro vermiedener Pflegegradverschlechterung und die Akzeptanz ca. 35 % bei 50.000 €. Subgruppenanalysen favorisierten u.a. jüngere Patient:innen und solche, die eine ambulante Prähabilitation erhalten hatten.

Diskussion: Die Studie konnte im primären Endpunkt keinen signifikanten Unterschied zeigen. Ein möglicher Grund hierfür ist eine wenig pflegebedürftige Studienpopulation trotz vorliegender Gebrechlichkeit. Dennoch zeigen Per-Protokoll-Analysen für den primären Endpunkt Signale für eine Wirksamkeit der Intervention. In den gesundheitsökonomischen Analysen zeigte sich die Situation, dass die Intervention in der Tendenz besser, aber teurer ist, so dass die Kosten-Effektivitäts-Akzeptanz maßgeblich von der Zahlungsbereitschaft abhängt.

Schlagnote: Prähabilitation, Frailty, ältere Patient:innen, randomisierte kontrollierte Studie, Veränderung des Pflegegrads, Gebrechlichkeitssyndrom

Inhaltsverzeichnis

I	Abkürzungsverzeichnis	5
II	Abbildungsverzeichnis	7
III	Tabellenverzeichnis	7
1	Ziele der Evaluation	9
2	Darstellung des Evaluationsdesigns	10
2.1	Population	10
2.2	Kontrollgruppe:	10
2.3	Interventionsgruppe	10
2.4	Endpunkte	11
2.5	Fallzahlberechnung	12
2.6	Randomisierung	13
2.7	Verblindung	13
2.8	Statistische Analysen	13
2.9	Gesundheitsökonomische Evaluation	14
2.9.1	Perspektiven	15
2.9.2	Population	16
2.9.3	Zeithorizont und Diskontierungsrate	16
2.9.4	Klinischen Endpunkte	16
2.9.5	Kosten	18
2.9.6	Analyse	21
3	Ergebnisse der Evaluation	25
3.1	Deskription	25
3.2	Analysen des primären Endpunktes mit Sensitivitätsanalysen	36
3.3	Per Protokoll Analysen: Primärer Endpunkt und Sensitivitätsanalysen	38
3.4	Analysen der sekundären Endpunkte	49
3.5	Interaktion zwischen Intervention und Patienteneigenschaften	49
3.6	Nebenwirkungen und Komplikationen, Mortalität	49
3.7	Ergebnisse der gesundheitsökonomischen Evaluation	50
3.7.1	Ergebnisse der gesundheitsökonomischen Evaluation aus gesellschaftlicher Perspektive	50
3.7.2	Ergebnisse der gesundheitsökonomischen Evaluation aus GKV- Perspektive. 65	
3.7.3	Ergebnisse der gesundheitsökonomischen Evaluation aus Leistungserbringer-Perspektive.....	81
4	Diskussion der Projektergebnisse	81

5	Schlussfolgerungen und Empfehlungen des Evaluators	86
IV	Literaturverzeichnis	87
V	Anlagen	89



I Abkürzungsverzeichnis

ABC6	verkürzte Version der Activities-specific Balance Confidence (ABC)-Skala, die 6 Aktivitäten umfasst
aG-DRG	ausgegliederte German Diagnosis Related Groups
AMG	Arzneimittelgesetz
ASA	American Society of Anesthesiologists
AUC	Area under the curve
BAS	Bundesamt für Soziale Sicherung
BGB	Bürgerliches Gesetzbuch
BMI	Body Mass Index
CC	Complete Cases
CEA	Cost-effectiveness analysis (Kosteneffektivitätsanalyse)
CHARMI	Charité Mobilitäts-Index
CONSORT	CONsolidated Standards Of Reporting Trials
COVID-19	Coronavirus disease 2019
CUA	Cost-utility analysis (Kostennutzwertanalyse)
DiD	Differenz in Differenz: Unterschied zwischen Veränderungen
DRG	Diagnosis related group
EQ-5D	EuroQol 5 Dimension (EQ-5D)
EQ-5D-5L	5-level EQ-5D version
FIMA	Fragebogen zur Inanspruchnahme medizinischer und nicht-medizinischer Versorgungsleistungen im Alter
GAD7	Generalized Anxiety Disorder Scale-7
G-BA	Gemeinsamer Bundesausschuss
GEE	Generalized Estimation Equation
GKV	Gesetzliche Krankenversicherung
GLM	Generalized Linear Model
IA	Innovationsausschuss
IADL	Instrumental Activities of Daily Living (Instrumentelle Aktivitäten des täglichen Lebens)
IG	Interventionsgruppe
IKER	Inkrementelle Kosten-Effektivitäts-Relation
IQR	Interquartile range (Interquartilsabstand)
ISPOR	The Professional Society for Health Economics and Outcomes Research
ITPW	Inverse Therapy Probability Weighting
ITS	Intensivstation
ITT	„Intention to treat“
ITT CC	„Intention to treat“ Analyse basierend auf Complete Cases
ITT SA	Sensitivitätsanalyse basierend auf „Intention to treat“ Analyse
KG	Kontrollgruppe
KH	Krankenhaus
KHEntgG	Krankenhausentgeltgesetz
KI	Konfidenzintervall
LtFu	Lost-to-Follow-up
MAR	Missing at random
MAX, Max	Maximum

MCAR	Missing completely at random
MICE	Multiple Imputation by Chained Equations
MIN, Min	Minimum
MNAR	Missing not at random
MNASF	Mini Nutritional Assessment - Short Form
MOCA	Montreal Cognitive Assessment
MPG	Medizinproduktegesetz
MW	Mittelwert
MWD	Mittelwertdifferenz
NA	Not available
NBA	Neues Begutachtungsassessment
NYHA	Klassifikation der New York Heart Association
OP, Op	Operation, chirurgischer Eingriff
OR	Odds Ratio
PG	Pflegegrad
PGV	Pflegegradverschlechterung
PHQ8	Patient Health Questionnaire – 8
PIM	Probabilistic Index Model
PP15	Per Protocol Analyse mit mindestens 15 Prähabilitations-Einheiten
PP15OP	Per Protocol Analyse mit mindestens 15 Prähabilitations-Einheiten und Operation
PP#	Per Protocol Analyse Version #
preha.end	Endzeitpunkt der Prähabilitation
preha.start	Startzeitpunkt der Prähabilitation
PRÄP-GO	Prähabilitation von älteren Patienten mit Gebrechlichkeitssyndrom vor elektiven Operationen
QALY	Qualitätsadjustiertes Lebensjahr
REDCap	Research Electronic Data Capture Software
ROC	Receiver Operator Characteristic
SA	Sensitivitätsanalyse ohne winsorisierte Kostendaten
SD	Standardabweichung
SGB	Sozialgesetzbuch
SDM	Shared Decision Making
TE	Trainingseinheit
TMTAB	Trail Making Test Score aus Teil A und B
TUG	Timed "Up and Go"-Test
V#	Studienvisite Nummer #
WHO-CHOICE	World Health Organization – Choosing Interventions that are Cost-Effective
WHODAS	World Health Organization Disability Assessment Schedule
%-Diff	Prozentuale Differenz

II Abbildungsverzeichnis

Abbildung 1: CONSORT Flow-Chart.....	25
Abbildung 2: Individuelle Zeitpunkte der verschiedenen Untersuchungstermine, aufgeteilt nach Studienarmen und stratifiziert nach dem Pflegegrad der Baselineerhebung	35
Abbildung 3: Boxplots zur Veränderung des NBA-Scores zwischen Baseline (v2) und einem Jahr nach OP (v18). Der Score reicht von 0 (keine Pflegebedürftigkeit) bis 100 (maximale Pflegebedürftigkeit). Fehlende Zeitangaben wurden durch den Mittelwert innerhalb der Strata (wie beispielsweise v02) im Plot ersetzt	36
Abbildung 4: Streudiagramm der ITT-Analyse des QALY aus gesellschaftlicher Perspektive..	54
Abbildung 5: Streudiagramm aller Analysen des QALY aus gesellschaftlicher Perspektive	55
Abbildung 6: Streudiagramm der ITT-Analyse der Pflegegradversänderung aus der gesellschaftlichen Perspektive	57
Abbildung 7: Streudiagramm aller Analysen der Pflegegradverschlechterung aus gesellschaftlicher Perspektive.....	58
Abbildung 8: Kosten-Effektivität-Akzeptanz-Diagramm für QALY und Pflegegradverschlechterung aus gesellschaftlicher Perspektive	59
Abbildung 9: Kosten-Effektivität-Akzeptanz-Diagramm für QALY aus gesellschaftlicher Perspektive – Subgruppenanalyse	63
Abbildung 10: Kosten-Effektivität-Akzeptanz-Diagramm für Pflegegradverschlechterung aus gesellschaftlicher Perspektive – Subgruppenanalyse	63
Abbildung 11: Streudiagramm der ITT-Analyse der Pflegegradverschlechterung aus GKV-Perspektive.....	71
Abbildung 12: Streudiagramm aller Analysen der Pflegegradverschlechterung aus GKV-Perspektive.....	71
Abbildung 13: Streudiagramm der ITT-Analyse des QALYs aus GKV-Perspektive	74
Abbildung 14: Streudiagramm aller Analysen des QALYs aus GKV-Perspektive	75
Abbildung 15: Kosten-Effektivität-Akzeptanz-Diagramm für QALY und Pflegegradverschlechterung aus GKV- Perspektive.....	76
Abbildung 16: Kosten-Effektivität-Akzeptanz-Diagramm für QALY aus GKV-Perspektive	80
Abbildung 17: Kosten-Effektivität-Akzeptanz-Diagramm für Pflegegradverschlechterung aus GKV-Perspektive.....	80

III Tabellenverzeichnis

Tabelle 1: Plan zu den Studienvisiten.....	11
Tabelle 2: Übersicht der gesundheitsökonomischen Analysen	23
Tabelle 3: Baseline-Eigenschaften nach Randomisierungsgruppe der ITT Population.....	26
Tabelle 4: Durchführung der Intervention, Beschreibung von Interventionseigenschaften ...	29
Tabelle 5: Eigenschaften der durchgeführten OP	31
Tabelle 6: Deskription des NBA-Scores und seiner Veränderung.....	33
Tabelle 7: Veränderungen des Pflegegrades in der Interventionsgruppe	33
Tabelle 8: Veränderungen des Pflegegrades in der Kontrollgruppe.....	34
Tabelle 9: Ergebnisse des primären Endpunktes	37

Tabelle 10: Dichotomisierte Veränderung des Pflegegrades: 0 – Verschlechterung, 1 – Gleichbleibend oder verbessert	37
Tabelle 11: Zusammensetzung der verschiedenen Per Protokoll Populationen	38
Tabelle 12: Therapieeffekte zum primären Endpunkt.	40
Tabelle 13: Therapieeffekte bei ausgesuchten sekundären Endpunkten	42
Tabelle 14: Interventionskosten gesellschaftliche Perspektive (in €).....	50
Tabelle 15: Gesamtkosten gesellschaftliche Perspektive nach Winsorisierung und Imputation (in €).....	52
Tabelle 16: Ergebnisse der Kosten-Effektivitäts- und -Nutzwert-Analysen aus gesellschaftlicher Perspektive (QALY und WHODAS 2.0).....	53
Tabelle 17: Ergebnisse der Kosten-Effektivitäts- und -Nutzwert-Analysen aus gesellschaftlicher Perspektive (Pflegegradverschlechterung)	56
Tabelle 18: Subgruppenanalyse ITT Qualitätsadjustiertes Lebensjahr aus gesellschaftlicher Perspektive	60
Tabelle 19: Subgruppenanalyse ITT Pflegegradverschlechterung aus gesellschaftlicher Perspektive	61
Tabelle 20: Zusammenhang von Patient:innen-Charakteristika und Kosteneffektivität innerhalb der Subgruppe „Ambulantes Setting“	64
Tabelle 21: Patient:innencharakteristika der GKV-Population	65
Tabelle 22: Interventionskosten GKV-Perspektive (in €)	66
Tabelle 23: Gesamtkosten GKV-Perspektive nach Winsorisierung und Imputation (in €)	67
Tabelle 24: Veränderung des Pflegegrads in der GKV-Population.....	68
Tabelle 25: Ergebnisse der Kosten-Effektivitäts-Analysen aus GKV-Perspektive (Pflegegradverschlechterung)	70
Tabelle 26: Ergebnisse der gesundheitsökonomischen Evaluation aus GKV-Perspektive (QALY)	73
Tabelle 27: Subgruppenanalyse ITT Qualitätsadjustiertes Lebensjahr aus GKV-Perspektive .	78
Tabelle 28: Subgruppenanalyse ITT Pflegegradverschlechterung aus GKV-Perspektive.....	79
Tabelle 29: Gesundheitsökonomische Evaluation aus Leistungserbringer-Perspektive	81

1 Ziele der Evaluation

Aufgrund der steigenden Lebenserwartung der Bevölkerung und aufgrund des medizinischen Fortschritts erfolgen mehr komplexe chirurgische Eingriffe bei älteren Patient:innen. Auch wenn oftmals der direkte postoperative Verlauf aus medizinischer Sicht komplikationsfrei verläuft, zeigen viele Patient:innen mittel- und langfristige Einschränkungen in Hinblick auf ihre Autonomie, auf kognitive und funktionelle Fertigkeiten und weitere Domänen der Lebensqualität. Verschiedene Studien zeigen, dass ein Gebrechlichkeitssyndrom (englisch: Frailty-Syndrome) einen Risikofaktor für ein schlechtes postoperatives Outcome darstellt.

Dabei stellt ein Gebrechlichkeitssyndrom jedoch keine Diagnose eines festen Zustands dar, sondern eine Beschreibung eines fluiden, sich verändernden Zustands. Aus diesem Grund lässt sich dieser Zustand auch mittels gezielter Interventionen positiv beeinflussen. In der Vorbereitung auf eine anstehende Operation geschieht dies durch eine so genannte Prähabilitation. Dabei handelt es sich um eine zielgerichtete multimodale Therapie zur Behandlung bestehender Symptome des Frailty-Syndroms vor der Operationsdurchführung. Die Prähabilitation reduziert perioperative Komplikationen, unterstützt die postoperative funktionelle Wiederherstellung und verringert die Gefahr von postoperativen kognitiven Störungen wie zum Beispiel dem postoperativen Delir. Verschiedene Studien konnten zudem eine kürzere postoperative Verweildauer im Krankenhaus nachweisen.

Aus diesen Gründen werden im Projekt PRÄP-GO Patient:innen mit einem Gebrechlichkeitssyndrom im präoperativen Umfeld identifiziert und vor der Durchführung der geplanten Operation einer Prähabilitation zugeführt. In einem multidisziplinären Setting werden vorhandene Risikofaktoren für perioperative Komplikationen verringert oder beseitigt. Zu den in diesem Setting stattfindenden Maßnahmen gehören unter anderem trainingstherapeutische Maßnahmen, Alltagstraining oder eine Ernährungsberatung, aber auch die Begutachtung des aktuellen Medikamentenplans mit dem Ziel der Vermeidung einer Polypharmazie. Zur Stärkung der Partizipation wird die Auswahl von Behandlungszielen und Maßnahmen der Prähabilitation in einer gemeinsamen Entscheidungsfindungskonferenz (Shared Decision Making Konferenz) mit den behandelnden Ärzten, Pflegekräften/Physiotherapeuten und dem Patient:innen stattfinden. Die neue Versorgungsform der Prähabilitation soll innerhalb des Projekts PRÄP-GO evaluiert werden. Dabei stehen die Beurteilung der Effektivität der Intervention, d.h. der Verringerung der postoperativen Pflegeabhängigkeit 1 Jahr nach OP sowie die gesundheitsökonomische Betrachtung im Mittelpunkt.

Primäres Ziel war der Nachweis, dass eine dreiwöchige Prähabilitation nach gemeinsamer Entscheidungsfindungskonferenz bei Patient:innen mit Gebrechlichkeitssyndrom ≥ 70 Jahre, bei denen eine elektive Operation geplant war, die präoperative Selbständigkeit zu erhalten oder zu verbessern sowie postoperative Pflegebedürftigkeit ein Jahr nach der Operation zu vermeiden.

Rechtsgrundlage der Studie sind (1) Selektivvertrag gem. §140a SGB V oder alternativ, wenn kein Selektivvertrag vorhanden ist, (2) ein Behandlungsvertrag gem. §630a BGB nach Freigabe durch die Projektleitung und Projektträger.

2 Darstellung des Evaluationsdesigns

Hier folgen wir dem veröffentlichten Protokoll [1].

Die PRÄP-GO-Studie ist eine nationale multizentrische, assessor-verblindete, randomisierte, pragmatische, zweiarmige, kontrollierte klinische Studie.

2.1 Population

Einschlusskriterien: (1) Alter ≥ 70 Jahre; (2) Einwilligungsfähiger Patient:in oder vorhandene:r gesetzliche:r Betreuer:in bei nicht-einwilligungsfähigem Patient:innen, (3) Elektive Operation geplant, (4) Erwartete Anästhesiedauer ≥ 60 min, (5) Versicherung in der gesetzlichen Krankenversicherung, (6) Frailty-Syndrom (≥ 1 positiver von 5 standardisierten Parametern entsprechend dem Physical Frailty Phänotyp nach Fried et. al. [2])

Ausschlusskriterien: (1) Schwere kardiologische oder pulmonale Erkrankung (NYHA IV, Gold IV), (2) Intrakranielle Eingriffe, (3) Moribunde Patient:innen (palliative Situation), (4) Keine ausreichenden Sprachkenntnisse, (5) Teilnahme an einer anderen interventionellen Rehabilitationsstudie oder einer AMG- bzw. MPG-Studie, die nicht von der Projektleitung freigegeben wurden (Ausnahme: Parallele Teilnahme an adjuvanter Therapiestudie), (6) Keine rechtsgültige Einwilligungserklärung.

2.2 Kontrollgruppe:

Teilnehmer der Kontrollgruppe führen die geplante Operation gemäß dem Versorgungsstandard durch.

2.3 Interventionsgruppe

Die Intervention im Rahmen des PRÄP-GO-Projekts besteht aus (1) einer Konferenz zur gemeinsamen Entscheidungsfindung (SDM), um gemeinsam mit dem Patienten und allen beteiligten Disziplinen über die Art und die Ziele der prärehabilitativen Intervention zu entscheiden und (2) der Durchführung der prärehabilitativen Intervention.

In der SDM-Konferenz werden basierend auf den ermittelten Risikofaktoren ein primäres und zwei sekundäre Therapieziele formuliert. Relevante Bereiche sollen durch die Verbesserung der bio-psycho-sozialen Ressourcen gestärkt werden um perioperative Komplikationen und eine Erhöhung des Pflegebedarfs zu vermeiden.

Die formulierten Therapieziele dienen als Grundlage für die therapeutische Interventionsplanung des Prähabilitationsprogramms. Es ist wichtig die Therapieziele auf die verschiedenen Interventionsformen in der Prähabilitation abzustimmen und die Ziele der Prähabilitation für den Patienten / die Patientin verständlich und überzeugend zu formulieren. Die formulierten Ziele sollen die Alltagsziele des Patienten / der Patientin widerspiegeln.

Weitere Ziele können auf der Grundlage der bei der Baseline-erhebung ermittelten Risikobereiche festgelegt werden und sollten im Rahmen der Prähabilitation reduziert oder beseitigt werden. Diese klinischen Ziele können entweder unabhängig (z. B. eine separate Diagnose) von der (Trainingstherapie) prähabilitativen Intervention erfolgen oder mit ihr einhergehen (z. B. ein funktionelles Defizit, das nicht durch die geplante Operation behoben wird).

Das Ziel der Trainingsplanung besteht im Erstellen individueller Trainingspläne mit folgenden Eigenschaften: (1) zielgerichtet, (2) auf den in der gemeinsamen Entscheidungsfindungskonferenz (SDM) formulierten Rehabilitationszielen basierend und (3) auf der Grundlage evidenzbasierten Wissens konzipiert und kontrolliert.

Jeder Patient erhält innerhalb der 3 Wochen vor der Rehabilitation 30 Einheiten einer von einem Spezialisten überwachten Therapie. Jede Sitzung dauert 30 Minuten. Zusätzlich wird jeder Patient gebeten, mindestens eine weitere TE pro Werktag selbstständig durchzuführen als Vorbereitung auf die Phase nach der OP: Atemtraining, Verwendung von Hilfsmitteln, etc. Eine weitere, selbstständig durchgeführte TE pro Woche ist für das Atemtraining zur Vorbereitung auf die postoperative Phase vorgesehen. So erhält jeder Patient 48 Einheiten gezielter Therapie gemäß den in der SDM-Konferenz formulierten Zielen.

Das Visiten-schema zur Studie ist in Tabelle 1 dargestellt.

Tabelle 1: Plan zu den Studienvisiten

Visit	Time	Description
Phase I: Screening und Einschluss		
V0	Tag 1	Screening auf Gebrechlichkeit, geriatrische Mini-Untersuchung
V1	Tag 1	Aufnahme in die Studie, Stammdatenerfassung
V2	Tag 1	Baseline Untersuchung
Phase II: Shared Decision Making (SDM)		
V3	Tag 1-3	Organisation der SDM-Konferenz
Phase III: Prehabilitation		
V4	V3 Zeitplan ¹	Beginn der Prehabilitation
V5	V4 + 21 Tage	Ende der Prehabilitation
Phase IV: Follow-up		
V6	3 - 14 Tage post-Op	Entlassung aus dem Krankenhaus
V7, 8	Erster und zweiter Monat post-Op	Kurze Telefoninterviews
V9	Dritter Monat post-Op	Ausführliches Telefoninterview/Hausbesuch
V10 - V17	Vierter bis elften Monat post-Op	Kurze Telefoninterviews
V18	12 Monate post-Op	Nachuntersuchung als Hausbesuch

2.4 Endpunkte

Der **primäre Endpunkt** ist die Veränderung des Pflegeabhängigkeitsgrades (gemäß der neuen Pflegestufeneinschätzung (NBA)) [3] ein Jahr nach der Operation (V18) zur Baseline Erhebung (V2) des Pflegegrades. Die positive Differenz $\text{Pflegegrad}_{V18} - \text{Pflegegrad}_{V2}$ deutet auf eine Verschlechterung hin. In Sensitivitätsanalysen betrachten wir ebenfalls eine dichotomisierte Version des primären Endpunkts: Verschlechterung eingetreten ja/nein und als stetige Größe den NBA Score.

Der **NBA Score** wird aus fünf gewichteten Werten aus sechs Modulen berechnet, wobei für die Module 2 und 3 nur der höhere Punktwert in die Berechnung einfließt. Der Gesamtwert reicht von 0 (keine Pflegebedürftigkeit) bis 100 (maximale Pflegebedürftigkeit). Der NBA-Score wird auf einer metrischen Skala gemessen, und das Pflegelevel wird basierend auf dem Gesamtwert zugeordnet.

Die **Klassifikation in die Pflegegradstufen (Pflegegrad)** basiert auf dem NBA-Score. Der Pflegegrad wird auf einer ordinalen Skala gemessen und in fünf Stufen unterteilt: Pflegegrad 0 - Keine Beeinträchtigung der Unabhängigkeit (NBA Score unter 12,5); Pflegegrad 1 - Geringfügige Beeinträchtigung der Unabhängigkeit (NBA-Score mehr als 12,5 und weniger als 27 Punkte), Pflegegrad 2 - Deutliche Beeinträchtigung der Unabhängigkeit (NBA-Score 27 bis weniger als 47,5 Punkte), Pflegegrad 3 - Schwere Beeinträchtigung der Unabhängigkeit (NBA-Score 47,5 bis weniger als 70 Punkte), Pflegegrad 4 - Schwerste Beeinträchtigung der Unabhängigkeit (NBA Score 70 bis weniger als 90 Punkte) und Pflegegrad 5 - Schwerste Beeinträchtigung der Unabhängigkeit mit besonderen Anforderungen an die Pflege oder Tod (NBA-Score 90 bis 100 Punkte)

Die Studie untersucht ebenfalls eine Reihe von **sekundären Endpunkten**: Bartel Index, CHARMI Score, Charles Comorbidity Score, EQ5D Score, GAD7 Score, Gebrechlichkeit, Gehgeschwindigkeit, geringe Kognitive Beeinträchtigung, Handkraft, IADL Score, kategorieller ABC6 Score, kategorieller TUG Score, kognitive Beeinträchtigung, MNASF Score, Mini Cog Score, NBA Unterskalen (Bewältigung, Eigenständigkeit, Kognitions- und Konversationsfähigkeiten, Mobilität, Organisation und soziale Kontakte), numerischer TUG Score, PHQ8 Score, schwere kognitive Beeinträchtigung, TMTAB Score, Treppengehgeschwindigkeit, Two Minute Step Test, WHODAS score und Mortalität.

2.5 Fallzahlberechnung

Auf der Grundlage des aktuellen Wissenstandes [4] wurden folgende Veränderungen des Pflegebedarfs in der Interventions- und Kontrollgruppe angenommen: (1) Eine Verbesserung des Pflegegrades um mindestens einen Punkt hat in der Kontrollgruppe die Wahrscheinlichkeit 2.5% und in der Interventionsgruppe 5%; (2) Keine Veränderung des Pflegegrades hat in der Kontrollgruppe die Wahrscheinlichkeit 42.5% und in der Interventionsgruppe 47.5%; (3) Eine Verschlechterung des Pflegegrades um genau einen Punkt hat in der Kontrollgruppe die Wahrscheinlichkeit 50.0% und in der Interventionsgruppe 45.0%; (4) Eine Verschlechterung des Pflegegrades um mehr als zwei Punkte hat in der Kontrollgruppe die Wahrscheinlichkeit 5.0% und in der Interventionsgruppe 2.5%.

Formal ist die Wahrscheinlichkeit von Veränderungen des Pflegebedarfs ein geordnetes kategoriales Ergebnis, sodass die Stichprobengröße nach Kolassa [5] unter Verwendung des nQuery-Advisor V7 [6] berechnet wurde.

Eine Stichprobengröße von 470 Patient:innen in jeder Gruppe hat eine 80-prozentige Power, um die oben aufgeführten erwarteten Effekte unter Verwendung eines Wilcoxon-Mann-Whitney-Rangsummentests mit einem 5%igen zweiseitigen Signifikanzniveau zu erkennen.

Unter Berücksichtigung von 2,5 % falscher Behandlungszuweisungen und 30 % Verlust bei der Nachsorge nach 12 Monaten betrug die angestrebte Stichprobengröße der Studie

n = 1378. Der ursprüngliche Rekrutierungsplan sah vor, 1378 Teilnehmer innerhalb eines Zeitraums von 12 Monaten zu erreichen. Aufgrund der COVID-19-Krise wurde die Rekrutierungsphase um weitere 12 Monate auf insgesamt 24 Monate verlängert.

2.6 Randomisierung

Nach Abschluss der Untersuchung V2 wurden alle Teilnehmer:innen gemäß einem vorab festgelegten Protokoll randomisiert. Zu diesem Zweck wurde ein unabhängiges, web-basiertes Verfahren für die zufällige Gruppenzuordnung unter Verwendung eines Randomisierungstools der Research Electronic Data Capture Software (REDCap) implementiert. Nach der Randomisierung durch das für die Studie zuständige Personal wurden der Patient:innen, das Studienzentrum und das Studienleitungsteam über die jeweilige Zuordnung informiert.

2.7 Verblindung

Aufgrund der Art der Intervention konnten alle Teilnehmer:innen und Mitarbeiter:innen in den Prähabilitationszentren und in der Kontrollgruppe nicht verblindet werden. Die Verblindung der Gruppenzuordnung wurde auf die Ergebnisprüfer:innen beschränkt (assessor-blinded). Projektkoordination, Datenmanagement und Interventionskoordination hatten Zugriff auf die Gruppenzuordnung, waren jedoch nicht an der Bewertung der Teilnehmer:innen in der Nachbeobachtungsphase beteiligt. Die Bewertungen der Baseline der relevanten Variablen (NBA) fand vor der Randomisierung statt.

2.8 Statistische Analysen

Die primäre Analyse des primären Endpunktes ist ein stratifizierter Wilcoxon-Mann-Whitney-Test.

Sensitivitätsanalysen für den primären Endpunkt bestehen (1) aus einem dichotomisierten Endpunkt der aus der Veränderung des Pflegegrades hergeleitet und mittels logistischer Regression getestet wird, (2) aus der Analyse der numerischen Veränderungen des NBA Scores zwischen den Studienarmen mittels GEEs, (3) aus einem dichotomisierten Version der NBA-Score Veränderung die mittels logistischer Regression analysiert wird, (4) aus einer Analyse der Diskriminationsfähigkeit der Veränderungen im Pflegegrad und NBA-Score zwischen beiden Studienarmen mittels ROC Analyse und AUC Berechnung, (5) eine stratifizierte Analyse der AUC mittels probabilistischen Index Modellen (PIM, Test der Nullhypothese, die AUC ist in allen Strata gleich 0.5).

Weiterhin verwenden wir die Win-Ratio Methode um mit den auftretenden Todesfällen formal sauber umgehen zu können [7]. Für jedes Paar von Studienteilnehmer:innen, bestehend aus einem der Interventionsgruppe (IG) und einem aus der Kontrollgruppe (KG), wird ein Gewinner:in und Verlierer:in nach folgenden Regeln bestimmt: (a) Der / die IG-Teilnehmer:in verliert, wenn er / sie vor dem KG-Teilnehmer:in stirbt; (b) Der / die KG-Teilnehmerin verliert, wenn er / sie vor dem IG-Teilnehmer:in stirbt; (c) Der / die überlebende IG-Teilnehmerin gewinnt gegenüber dem / der überlebenden KG-Teilnehmer:in falls er/sie eine bessere Score Veränderung hat; (d) Der / die überlebende KG-Teilnehmer:in gewinnt gegenüber einem / einer überlebenden IG Teilnehmer:in falls er/sie eine bessere Score Veränderung hat. Die Win-Ratio wird dann als das Verhältnis der Gewinner aus der IG zu den Gewinnern aus der KG berechnet.

Die Analyse des NBA Scores lässt sich auch auf die Per-Protokoll Populationen anwenden, wobei bei kein GEE Modell, sondern eine lineare Regression mit robust geschätzten Standardfehler (Sandwich Schätzer) verwendet wurde. Hier werden die genannten Verfahren mit Gewichtungen kombiniert (ITPW – Inverse Therapy Probability Weighting).

Es werden drei Per-Protokoll-Populationen betrachtet: (PP1) Alle Kontroll- und alle InterventionsPatient:innen mit mindestens 15 Prähabilitationseinheiten (15E), (PP2) alle Patient:innen aus PP1, die eine Operation erhielten (15E+OP), (PP3) alle Patient:innen aus PP2, die zum Zeitpunkt V18 noch am Leben waren (Principal Stratum). Die Analysen werden in den entsprechenden Populationen durchgeführt, dabei werden Selektionseffekte, die die Randomisierung zwischen Kontroll- und Interventionsarm aufbrechen, durch die Verwendung von ITPW (inverse treatment probability weighting) ausgeglichen und eine Analyse wird in einer Pseudopopulation mit strukturgleichen Armen durchgeführt.

In diesen Pseudopopulationen befinden sich diejenigen Patienten, die bereit sind mindestens unter der Bedingung 15 Prähabilitationseinheiten (und dazu die geplante OP) zu erhalten sich in zwei Arme randomisieren zu lassen. Das ist ein hypothetischer Ansatz der den Therapieeffekt unter idealen Bedingungen zeigen soll. Das gleiche gilt für die Population im Principal Stratum (Personen, die 15 Einheiten und eine OP erhalten und bis zum Abschlussbesuch überleben).

Die Ansätze für die primäre und Sensitivitäts-Analysen werden soweit passend auf die relevanten sekundären Endpunkte übertragen.

Weiterhin mittels eines zweistufigen Verfahrens eine explorative Analyse durchgeführt um die Möglichkeit von Interaktion zwischen individuellen Patienteneigenschaften und der Intervention zu detektieren. Für jeden Endpunkt wird in der Studienpopulation ein Prognosemodell entwickelt und ein Prognosescore berechnet. Mittels logistischer Regression wird die individuelle Wahrscheinlichkeit einer Verschlechterung im betrachteten Endpunkt untersucht. In einem zweiten Schritt wird eine Interaktion zwischen Prognose-score und Therapie (Standard versus Prähabilitation). Signifikante Interaktionseffekte werden als Signale für eine potentielle Patienten-Therapie Interaktion gesehen: Individuelle Patienten reagieren verschieden auf die Prähabilitation.

Die Analysestrategien zur Auswertung der Daten sind in einem statistischen Analyseplan definiert, Dieser ist unter (Password: PraepGO2025!; <https://cloud.ibe.med.uni-muenchen.de/index.php/s/FT78oidTBHYm2CE>) zusammen mit dem Protokoll der Studie einsehbar.

2.9 Gesundheitsökonomische Evaluation

Die gesundheitsökonomische Evaluation folgte einem vorab registrierten und publizierten Protokoll [8]. Die verwendete Methodik orientierte sich an den aktuellen ISPOR Guidelines [9], Standardwerken der gesundheitsökonomischen Evaluation (u.a. [10] und [11]) sowie einem von den Evaluatoren durchgeführten systematischen Review zur Kosten-Effektivität der Prähabilitation vor elektiven Operationen [12].

Die übergeordnete Fragestellung lautete: Ist die neue Versorgungsform, bestehend aus Frailty-Screening, Shared-Decision-Making (SDM)-Konferenz und 3-wöchiger multimodaler

Prähabilitation, im Vergleich zur standardmäßigen präoperativen Versorgung bei älteren, (vor-)gebrechlichen Patient:innen kosteneffektiv?

Es wurden insgesamt zwei Analysearten aus drei Perspektiven durchgeführt:

1. Kosten-Nutzwert-Analyse (Qualitätsadjustiertes Lebensjahr (QALY)) aus der gesellschaftlichen Perspektive
2. Kosten-Effektivitäts-Analyse (Pflegegradverschlechterung, funktionaler Gesundheitszustand) aus der gesellschaftlichen Perspektive
3. Kosten-Effektivitäts-Analyse (Pflegegradverschlechterung) aus der Perspektive der Gesetzlichen Krankenversicherung (GKV) als Kostenträger
4. Kosten-Nutzwert-Analyse ((QALY) aus der GKV-Perspektive
5. Kosten-Effektivitäts-Analyse (Krankenhausverweildauer, Verweildauer auf der Intensivstation (ITS), Anzahl an Intensivstation-Patient:innen, perioperative Komplikationen) aus der Perspektive der Krankenhäuser als Leistungserbringer

2.9.1 Perspektiven

Die gesundheitsökonomische Analyse aus drei Perspektiven soll eine optimale gesellschaftliche Entscheidungsfindung ermöglichen und relevanten Akteuren sowie Entscheidungsträgern Empfehlungen für die potenzielle Implementierung der Intervention in die Routineversorgung geben.

1. Gesellschaftliche Perspektive: Die gesellschaftliche Perspektive wurde gewählt, um die Tragweite der neuen Versorgungsform möglichst vollständig und unter Einbeziehung aller relevanten Kosten, z.B. auch für informelle Pflege, abzubilden. Zusätzlich zu den direkten medizinischen Kosten während der 12-monatigen Nachbeobachtung wurden auch nicht-medizinische und indirekte Kosten betrachtet, wie Opportunitätskosten für informelle Pflege oder die Beteiligung von Angehörigen an der partizipativen Entscheidungsfindung (SDM-Konferenz) sowie Eigenbeteiligungen der Patient:innen an Gesundheitsleistungen durch Zuzahlungen.
2. GKV-Perspektive: Die Bewertung aus der Sicht der gesetzlichen Krankenversicherung berücksichtigte die erstattungsfähigen Kosten der Intervention, die Kosten des initialen Krankenhausaufenthalts für die Operation sowie die Nutzung erstattungsfähiger Leistungen während des 12-monatigen Nachbeobachtungszeitraums. Die GKV-Perspektive wurde gewählt, um die Entscheidung des Innovationsfonds hinsichtlich einer Empfehlung für eine Überführung der neuen Versorgungsform in die Regelversorgung optimal unterstützen zu können.
3. Leistungserbringerperspektive: Aus Sicht der Gesundheitsdienstleister wurden die Endpunkte Krankenhaus- und ITS-Verweildauer, Anzahl an ITS-Patient:innen und perioperative Komplikationen als Proxy für Kosten berücksichtigt und den erhaltenen Entgelten gegenüber gestellt. Die Leistungserbringer-Perspektive wurde gewählt, um zu eruieren, ob sich für die Krankenhäuser, welche die Index-Operation erbringen, ein Kostenvorteil durch die neue Versorgungsform bietet. Ein solcher finanzieller Anreiz

stellt einen wichtigen Faktor bei einer möglichen Überführung der Intervention in die Regelversorgung dar.

2.9.2 Population

Die Population entsprach, wenn nicht anders angegeben, der Studienpopulation. Ursprünglich sollte diese nur aus BARMER-Versicherten bestehen, da die BARMER Krankenkasse Konsortialpartner des Projektes ist. Es sind ca. 12 % der deutschen Gesamtpopulation aller gesetzlich Versicherten bei der BARMER Krankenkasse versichert (Stand 2024) [13]. Im Verlauf des Projektes kam es zu einer Kassenöffnung um die geplanten Rekrutierungszahlen trotz der COVID-19 Pandemie und den damit verbundenen Einschränkungen erreichen zu können, sodass letztendlich nur ein Subset der Studienteilnehmer:innen auch BARMER-Versicherte waren. Für die Analyse aus der GKV-Perspektive konnten jedoch nur BARMER-Versicherte betrachtet werden, da nur für sie ein Antrag beim Bundesamt für Soziale Sicherung (BAS) von der BARMER Krankenkasse gestellt werden konnte, welches die Bedingung für die Routinedatenlieferung seitens der Krankenkassen ist. Für die Zusammenführung der primär erhobenen Daten mit den Routinedaten der BARMER-Versicherten wurde das geteilte Pseudonym der Studienteilnehmer:innen verwendet. Für die Analyse aus der Leistungserbringer-Perspektive wurden als Population nur Patient:innen betrachtet, die eine Operation erhalten haben.

2.9.3 Zeithorizont und Diskontierungsrate

Der gewählte Zeithorizont variierte je nach Perspektive:

- Gesellschaftliche und GKV-Perspektive: Der Betrachtungszeitraum umfasste 12 Monate postoperativ nach der Indexoperation.
- Leistungserbringerperspektive: Der Zeithorizont reichte bis zur Entlassung nach der Indexoperation.

Eine Extrapolation von Kosten und Effekten über mehr als ein Jahr wurde nicht durchgeführt. Diese wäre aufgrund der heterogenen Patient:innenpopulation (unterschiedliche Operationstypen sowie gebrechliche und vorgebrechliche ältere Menschen) mit erheblicher Unsicherheit verbunden gewesen. Da der Follow-Up Zeitraum nur ca. 12 Monate umfasste, wurde keine Diskontierung von Kosten und Effekten vorgenommen.

2.9.4 Klinischen Endpunkte

Je nach gewählter Perspektive wurden unterschiedliche klinische Endpunkte erhoben. Folgende Endpunkte wurden ausgewählt und gemessen, um die klinische Effektivität der Intervention zu untersuchen:

1. Pflegegradverschlechterung bis 12 Monate postoperativ - gesellschaftliche und GKV-Perspektive:

Analog zum primären Endpunkt der Studie wurde die Pflegegradveränderung als Endpunkt für die Kosten-Effektivitäts-Analysen gewählt. Das Vorhandensein und die Höhe eines Pflegegrades spiegeln Einschränkungen in den Bereichen Mobilität, kognitive und kommunikative Fähigkeiten, Verhaltensweisen und psychische Problemlagen, Selbstversorgung, Umgang mit krankheits- und therapiebedingten Anforderungen und

Belastungen, sowie der Gestaltung des Alltagslebens und sozialer Kontakte von Patient:innen wider und ist somit ein patient:innenrelevanter Endpunkt. Des Weiteren ist ein Pflegegrad mit Leistungsansprüchen aus der Pflegekasse verbunden und stellt somit einen wichtigen Endpunkt aus Sicht der Kranken- bzw. Pflegekassen und der Solidargemeinschaft dar. Die Erhebung und Berechnung des Pflegegrades mittels des Neuen Begutachtungsinstruments wird bereits ausführlich im Kapitel „Endpunkte“ [3] beschrieben. Für die gesundheitsökonomische Evaluation wurde die dichotome Ausprägung der Pflegegradveränderung verwendet (Pflegegradverschlechterung (PGV) eingetreten ja/nein) und als prozentualer Anteil ausgedrückt. Das Ergebnis der gesundheitsökonomischen Analysen bezieht sich auf die Vermeidung von Patient:innen mit Pflegegradverschlechterungen in Prozent.

2. Qualitätsadjustiertes Lebensjahr (QALY) - gesellschaftliche und GKV-Perspektive:

Das QALY ist ein zusammengesetzter Endpunkt aus Lebensqualität und Lebenszeit, welche beide patient:innenrelevant sind. Das QALY wurde als Endpunkt für die Kosten-Nutzwert-Analysen gewählt, da es interventionsübergreifend gültig und international bekannt ist, sodass es Vergleiche zwischen Interventionen und Gesundheitssystemen ermöglicht. In der vorliegenden gesundheitsökonomischen Evaluation wurde das QALY mittels des EQ-5D-5L-Fragebogens [14] und den Präferenzwerten der deutschen Allgemeinbevölkerung [15] bestimmt. Der EQ-5D-5L wurde 3, 6, 9 und 12 Monate nach der Operation erhoben. Im Todesfall wurde das QALY ab dem Todesdatum (erhoben aus der Studiendokumentation bzw. mittels der Melderegisterabfragen) auf 0 gesetzt. Ein QALY von 1 entspricht einem Lebensjahr bei voller Lebensqualität. Das Ergebnis wurde dann als inkrementeller Effekt, also gewonnene QALYs in der Interventionsgruppe, wiedergegeben. Laut Studienprotokoll war ursprünglich keine Analyse des QALYs aus GKV-Perspektive vorgesehen, sondern diese wurde nachträglich im Sinne der Vergleichbarkeit mit der gesellschaftlichen Perspektive mitaufgenommen.

3. Funktionaler Gesundheitszustand zu 12 Monate postoperativ - gesellschaftliche Perspektive:

Der funktionale Gesundheitszustand, gemessen mit dem World Health Organization Disability Assessment Schedule (WHODAS) 2.0 [16], wurde als sekundärer Endpunkt für die Kosten-Effektivitäts-Analyse aus gesellschaftlicher Perspektive gewählt, da er ähnliche Domänen wie das Neue Begutachtungsinstrument für den Pflegegrad erfasst (Verständnis und Kommunikation, Mobilität, Selbstversorgung, Umgang mit anderen Menschen, Tätigkeiten des alltäglichen Lebens und Teilnahme am gesellschaftlichen Leben), aber international angewendet wird und so Vergleiche mit anderen Gesundheitssystemen erlaubt. Es wurde die 12-Item Version des Fragebogens und die „Simple scoring“-Auswertung verwendet, bei der Werte zwischen 12 (besten Wert) und 60 (schlechtester Wert) erreicht werden können.

4. Krankenhausverweildauer und Intensivstationsverweildauer - Leistungserbringerperspektive:

Entsprechend der aktuellen Vergütungsstrukturen erhalten Krankenhäuser pauschale Entgelte für erbrachte Leistungen, die überwiegend von den Diagnosen der Patient:innen abhängen (aG-DRG-Fallpauschalen). Eine Möglichkeit für Krankenhäuser, ihre Kosten-Effektivität zu optimieren, liegt daher in der Minimierung von Opportunitätskosten, z.B. durch kürzere Verweildauern von Patient:innen (innerhalb der Grenzverweildauer) oder die Vermeidung von ungeplanten ITS-Aufenthalten, welche es erlauben, mehr Fälle einer Art zu behandeln und abzurechnen. Daher wurden die Krankenhausverweildauer (in Tagen), die Verweildauer auf der Intensivstation (in Tagen) und der Anteil an ITS-Patient:innen als Endpunkte für die Kosten-Effektivitäts-Analyse aus Leistungserbringer-Perspektive gewählt.

5. Perioperative Komplikationen - Leistungserbringerperspektive:

Das Auftreten von perioperativen Komplikationen wurde als Endpunkt für die Kosten-Effektivitäts-Analyse aus Leistungserbringer-Perspektive gewählt, da diese mit einem erhöhten Versorgungsaufwand einhergehen, welcher nur unter bestimmten Umständen (z.B. bei der Notwendigkeit zur Beatmung) durch Fallpauschalen abgedeckt wird. Für die gesundheitsökonomische Evaluation wurde die Einteilung nach Clavien-Dindo [17] von Grad I (nicht behandlungsbedürftige Abweichung vom normalen postoperativen Verlauf) bis V (Tod) verwendet, wobei pro Patient:in jeweils nur die Komplikation mit der schlechtesten Ausprägung gezählt wurde.

2.9.5 Kosten

Auch die Auswahl, Messung und Bewertung der Ressourcenverbräuche orientierte sich an den jeweiligen Perspektiven. Die daraus resultierende Berechnung bzw. Schätzung der Kosten lässt sich in drei Hauptkostenblöcken zusammenfassen: Die Interventionskosten, Kosten des Index-Krankenhausaufenthaltes und Kosten im 12-Monats Follow-Up.

Sofern nicht anders angegeben, werden die aufgeführten Kosten in der Währung Euro (€), Preisjahr 2020, dargestellt. Für alle Berücksichtigungen von Patient:innen-Zuzahlungen wurde die Annahme getroffen, dass 50% der Patient:innen zuzahlungspflichtig sind.

1. Interventionskosten:

Die Interventionskosten setzten sich aus den Kosten für die drei Interventionsbestandteile zusammen (Frailty-Screening, SDM-Konferenz und Prähabilitation).

Die Kosten für das Frailty-Screening wurden basierend auf der in der Studie veranschlagten Pauschale i.H.v. 19,00 € pro Patient:in (Frailty-Erhebung Teil 1) berechnet, da ein solches Screening noch nicht Teil der Regelversorgung ist. Die Kosten aller gescreenter Patient:innen wurden durch die Anzahl aller positiv gescreenten Patient:innen geteilt und dann pro Person in der Interventionsgruppe veranschlagt. Aus studienorganisatorischen Gründen war es notwendig, dass auch die Patient:innen aus der Kontrollgruppe das Frailty-Screening durchlaufen. Da dieses jedoch nicht Teil der Regelversorgung ist, wurde den Teilnehmer:innen der Kontrollgruppe hierfür keine Kosten angesetzt.

Die Kosten für die SDM-Konferenz setzten sich aus einem Vorgespräch zwischen einer Ärzt:in und der/dem Patient:in (SDM-1), sowie der eigentlichen multiprofessionellen SDM-Konferenz zusammen. Für jede SDM-Konferenz wurde zunächst ermittelt, welche Berufsgruppen (Anästhesie, Chirurgie, Geriatrie, Allgemeinmedizin, Therapie und Pflege) involviert waren. Die Kosten für die SDM-Konferenz wurden ebenfalls basierend auf den in der Studie veranschlagten Pauschalen berechnet, da ein solches Screening noch nicht Teil der Regelversorgung ist. Für eine Teilnahme von Personen der ärztlichen Berufsgruppen wurden dabei 99,90 € pro Person berechnet (für max. eine Person pro Gruppe), für eine Teilnahme von Personen der Therapieberufe oder der Pflege 54,00 € (für max. eine Person für beide Gruppen) bei einer geschätzten Dauer der Konferenz von 90 Minuten. Die Kosten für die SDM-1 wurden im Rahmen der Studie nicht gesondert vergütet, sodass hierfür nachträglich die durchschnittliche Dauer ermittelt wurde (30 Minuten) und anteilig die Pauschale für ärztliches Personal berechnet wurde (33,30 €). Zusätzlich wurde aus der gesellschaftlichen Perspektive ermittelt, ob Angehörige der Patient:innen teilgenommen haben, und dafür 45,39 € (für max. eine Person) veranschlagt. Dieser Kostensatz basiert auf den von [18] veröffentlichten standardisierten Bewertungssätzen aus gesellschaftlicher Perspektive für gesundheitsökonomische Evaluationen (hier Substitutionskostenansatz für informelle Pflege, hochgerechnet auf 90 min). Für Patient:innen, die keine SDM-Konferenz erhalten haben, wurden die Kosten mit 0 € berechnet.

Die Kosten für die Prähabilitation wurden je nach Setting berechnet. Für die Durchführung von Heimübungen wurden aus keiner Perspektive Kosten berechnet, da alle Patient:innen über 70 Jahre alt waren und angenommen wurde, dass sie nicht mehr erwerbstätig sind (keine Berechnung von Opportunitätskosten). Für das ambulante Setting (hier: ambulant, ambulant mit Fahrtkosten und mobil) wurde für jede Prähabilitationseinheit die Berufsgruppe ermittelt und aus der gesellschaftlichen Perspektive mit den entsprechenden Kostensätzen von [18] bewertet: Für Physio- und Sporttherapeut:innen entsprach dies 21,45 €, Ergotherapeut:innen 54,21 €, Logopäd:innen 60,70 €, Ärzt:innen 73,13 € und Psychotherapeut:innen 102,92 €. Die Kostensätze für die Ernährungsberatung wurden aus der studieninternen Abrechnung übernommen. Für die mobile Prähabilitation wurde eine Pauschale für Hausbesuch inkl. Wegegeld i.H.v. 16,22 € berechnet (gemäß der Höchstpreise nach § 125b SGB V zum 01.07.2019 Physiotherapie). Die Transportkosten wurden aus der studieninternen Abrechnung übernommen. Für Patient:innen, die keine Prähabilitation erhalten haben, wurden die Kosten hierfür mit 0 € berechnet.

Aus GKV-Perspektive wurden die Kostensätze für die Therapieberufe abweichend basierend auf den GKV-Heilmittel-Schnellinformation für Deutschland nach § 84 Abs. 5 i.V.m. Abs. 7 SGB V von 2020 berechnet: Für Physio- und Sporttherapeut:innen entsprach dies 23,96 €, Ergotherapeut:innen 50,96 € und Logopäd:innen 56,87 €. Für die Ernährungsberatung, welche keine Kassenleistung darstellt, jedoch von vielen Krankenkassen bezuschusst wird, wurden die Zuschüsse der BARMER Krankenkasse berücksichtigt: Erstberatung bis 40 €, Folgeberatung (max. 4) jeweils bis zu 30 € pro Termin. Die Pauschale für Hausbesuch inkl. Wegegeld wurde um die Patient:innen-Zuzahlung i.H.v. 2,55 € verringert. Die Transportkosten wurden nur für Patient:innen berechnet, die in der GKV aktuell Anspruch darauf hätten (tatsächlicher Pflegegrad 4 oder 5 bzw. 3 bei permanent eingeschränkter Mobilität), und um Patient:innen-Zuzahlungen verringert (10% des Preises, mind. 5 €, max. 10 €).

Für das teilstationäre Setting wurde die Gesamtanzahl der Tage berechnet, an denen die Prähabilitation stattgefunden hat, und mit dem Kostensatz von [18] für ambulante Rehabilitation bewertet (74,95 € pro Tag). Aus GKV-Perspektive wurde dieser um die Patient:innen-Zuzahlung i.H.v. 10 € pro Tag für max. 28 Tage verringert.

Für das stationäre Setting wurde die Gesamtanzahl der Tage berechnet und mit einer fiktiven DRG bewertet. Diese wurde aus den Durchschnittswerten der DRGs 2020 für die geriatrische frührehabilitative Komplexbehandlung gebildet (ohne Diagnose eines akuten Schlaganfalls oder mit einer Prozedur): B44C, E42Z, F48Z, G52Z, H44Z, I41Z, J44Z, K44Z, L44Z, T44Z, U40Z. Die Bewertungsrelation von 1,456 wurde dann mit dem Bundesbasisfallwert von 2020 i.H.v. 3.679,62 € unter Berücksichtigung der unteren und oberen Grenzverweildauer multipliziert und der Pflegeentgeltwert mit 159,97 € pro Tag angesetzt (gemittelter Wert für 2020 gemäß § 15 Krankenhausentgeltgesetz (KHEntgG)). Aus gesellschaftlicher Perspektive wurde zusätzlich eine Pauschale für Investitionskosten i.H.v. 331,68 € je Patient:in berücksichtigt (entspricht den vom Institut für das Entgeltsystem im Krankenhaus ermittelten mittleren Investitionskosten je Fall 2020). Aus GKV-Perspektive wurden die Kosten um die Patient:innen-Zuzahlung i.H.v. 10 € pro Tag für max. 28 Tage verringert.

2. Kosten des Index-Krankenhausaufenthaltes:

Die Kosten des Index-Krankenhausaufenthalts wurden basierend auf den Entgelten berechnet, welche die jeweiligen Krankenhäuser für den entsprechenden Behandlungsfall der Studien-Patient:innen erhalten haben (Preisjahr: Jahr des Aufenthalts). Alle teilnehmenden Krankenhäuser übermittelten diese Daten (aufbereitet gemäß § 21 - KHEntgG) vertragsgemäß an die Konsortialführung, welche die Daten in das Studiendokumentationssystem REDCap integrierte. Aus der gesellschaftlichen Perspektive wurden zusätzlich die Patient:innen-Zuzahlung i.H.v. 10 € pro Tag für max. 28 Tage und eine Pauschale für Investitionskosten berechnet (s.o.). Es wurden keine Kosten für Patient:innen berechnet, die keine Operation erhalten haben.

3. Kosten des 12-Monats Follow-Up:

Gesellschaftliche Perspektive: Aus der gesellschaftlichen Perspektive wurden die Kosten des 12-Monats Follow-Up mittels des „Fragebogens zur Inanspruchnahme medizinischer und nicht-medizinischer Versorgungsleistungen im Alter“ (FIMA) erhoben [19]. Dieser wurde 3, 6, 9 und 12 Monate nach der Operation erhoben. Patient:innen wurden jeweils zur Inanspruchnahme in den letzten 3 Monaten befragt zu folgenden Ressourcen: Krankenhausaufenthalte vollstationär, Krankenhausaufenthalte teilstationär (Tagesklinik, ambulante Operationen), stationäre Aufenthalte in einer Psychiatrie, Arztbesuche, therapeutische Leistungen (Heilmittel), Leistungen aus der gesetzlichen Pflegeversicherung (Pflegedienst, Kurzzeitpflege, Tagespflege, stationäre Dauerpflege), Haushaltshilfe, Hilfe von anderen (Familienmitgliedern, Freunden, Bekannten oder Nachbarn), Hilfsmittel, Rehabilitation (ambulant, stationär). Diese wurden mit den entsprechenden Kostensätzen von [18] bewertet. Patient:innen wurden außerdem zu den von ihnen innerhalb der letzten 7 Tage eingenommenen Medikamenten befragt. Da für Medikamente keine standardisierten Kostensätze vorliegen und im Rahmen der Studie aus organisatorischen Gründen nur in Einzelfällen die Pharmazentralnummer der eingenommenen Medikamente erfasst werden

konnte, war es letztendlich jedoch nicht möglich, die Kosten für die Inanspruchnahme von Medikamenten zu schätzen.

GKV-Perspektive: Aus GKV-Perspektive wurden für die Follow-Up-Kosten alle tatsächlich erstatteten Kosten herangezogen, die innerhalb eines Jahres nach der Operation angefallen sind. Diese umfassten den Bereich der ambulant-fachärztlichen Versorgung (inklusive verordneter Arznei-, Heil- und Hilfsmittel), sowie von Behandlungen im Krankenhaus (ambulant und stationär), ambulante und stationäre Rehabilitationsmaßnahmen, Transport, Haushaltshilfe und Leistungen aus der Pflegekasse (Preisjahr: Jahr der Leistung).

2.9.6 Analyse

Inkrementelle Kosten-Effektivitäts-Relation

Die verwendeten Effektmaße richten sich nach der Perspektive. Aus der gesellschaftlichen und GKV-Perspektive wurde bei den Kosten-Effektivitäts-Analysen jeweils die Mittelwertdifferenz (MWD) der Kosten und die MWD oder die prozentuale Differenz (%-Diff) der Effektivitätspunkte zwischen Interventions- und Kontrollgruppe berechnet, was den inkrementellen Kosten bzw. inkrementellen Effekten der Intervention entspricht. Da bei der Wahrscheinlichkeit einer Pflegegradverschlechterung sowie beim WHODAS 2.0 Score niedrigere Werte wünschenswerter sind als höhere, wurden diese vor der Berechnung invertiert. Bei der Kosten-Nutzwert-Analyse, die eine Unterform der Kosten-Effektivitäts-Analyse darstellt, war der Effektivitätspunkt stets das QALY. Aus den inkrementellen Kosten und Effekten wurde anschließend die inkrementelle Kosten-Effektivitäts-Relation (IKER) gebildet, welche die zusätzlich entstandenen Kosten für eine zusätzliche Nutzeinheit widerspiegelt. Den Zähler der IKER bildet dabei die Differenz der Mittelwerte zwischen den Gesamtkosten der Interventions- und Kontrollgruppe, während die Differenz der verschiedenen klinischen Endpunkte den Nenner bildet.

Die gesundheitsökonomische Evaluation aus der Leistungserbringer-Perspektive erfolgte rein deskriptiv, ohne Berechnung einer IKER. In dieser Analyse können die erhaltenen Entgelte nicht als Kosten interpretiert werden, da diese auf Fallpauschalen basieren und nicht den tatsächlich angefallenen Kosten. Daher wurde stattdessen die oben beschriebenen Endpunkte sowie das Entgelt pro Patient:in und Endpunkteinheit betrachtet und deskriptiv zwischen der Interventions- und Kontrollgruppe verglichen.

Datenvorbereitung und -bearbeitung

Aufgrund der breiten Einschlusskriterien der Studie wurde antizipiert, dass es trotz Randomisierung zu relevanten Unterschieden in den Patient:innencharakteristika zwischen der Interventions- und Kontrollgruppe kommen könnte. Daher wurde die Standardisierte Mittelwertdifferenz (SMD) für folgende, im Protokoll spezifizierte, Charakteristika berechnet, um die Notwendigkeit einer potentiellen Adjustierung der Endpunkt- und Kostendaten zu eruieren. Diese wurde ab einer SMD-Abweichung von $\geq 10\%$ als gegeben angesehen. Charakteristika: Alter, Geschlecht, Charlson Comorbidity Index, Operationsart, Operationsrisiko, und einschließendes Zentrum.

Fehlende Werte wurden bei verstorbenen Patient:innen wie folgt ersetzt: Bei den Kosten mit 0 € ab dem Todeszeitpunkt, bei den Effekten mit dem jeweils schlechtesten Wert (Pflegegrad von 5, WHODAS 2.0 Score von 60) bzw. beim EQ-5D-5L mit 0. Da aus Leistungserbringer-

Perspektive nur Patient:innen berücksichtigt wurden, die eine Operation erhalten haben, gab es bei den Endpunkten Krankenhausverweildauer, ITS-Verweildauer und Komplikationen keine fehlenden Werte auf Grund verstorbener Patient:innen. Für die Complete Case-Analyse aus der gesellschaftlichen Perspektive wurden fehlende Kosten auf 0 € gesetzt („Complete Case“ bezieht sich hier nur auf die Effektivitäts-Endpunkte). Für die übrigen Analysen wurden fehlende Kosten- und Effektivitäts-Werte imputiert. Dafür wurde zunächst geprüft, ob die Werte missing-completely-at-random (MCAR), missing-at-random (MAR) oder missing-not-at-random (MNAR) sind. Da es Unterschiede in der Häufigkeit von einigen fehlenden Werten zwischen den Gruppen gab, diese jedoch durch studienorganisatorische Gründe erklärbar waren, wurde MAR angenommen. Fehlende Werte kamen in der GKV-Analyse vornehmlich in den 3-, 6- und 9-Monatsbefragungen des EQ-5D-5L vor und wurden mit geeigneten random forest basierten, multiplen Imputationsverfahren ersetzt. Für die Evaluation aus gesellschaftlicher Perspektive wurde der MICE (Multiple Imputation by Chained Equations) Ansatz gewählt, um die fehlenden Daten zu imputieren. Alle relevanten Baseline-, Effekt- und Kostendaten wurden hierbei berücksichtigt. Lediglich Kostendaten, bei denen über 90 % fehlten oder 0 € betrugen wurden hierbei ausgenommen, da sie eine geringe Vorhersagekraft aufwiesen. Fehlende Werte wurden in diesem Fall auf 0 € gesetzt.

Um den Einfluss von Extremkosten auf die Berechnungen der Kosteneffektivität zu begrenzen, wurden die Kosten für alle Analysen pro Kostenblock ab dem 97,5 Perzentil winsorisiert. Dieser Cut-Off wurde gewählt, da dadurch einerseits nur die extremen Ausreißer angepasst wurden, welche die Ergebnisse erheblich beeinflusst hätten, und dabei gleichzeitig nur ein sehr geringer Anteil der Daten manipuliert werden musste. Die Interventionskosten wurden nicht winsorisiert, da hier keine Ausreißer erwartet oder beobachtet wurden. Mögliche Verzerrung durch die Winsorisierung wurde in einer Sensitivitätsanalyse untersucht (s.u.).

Durchgeführte Analysen

Die Hauptanalyse basierte bei allen Perspektiven auf der Intention-to-treat (ITT) Population. Als sekundäre Analysen wurde aus der GKV- und gesellschaftlichen Perspektive jeweils eine Complete Case-Analyse anhand der ITT-Population sowie Per-Protocol-Analysen, eine Sensitivitätsanalyse, bei der die Kostendaten nicht winsorisiert wurden, und Subgruppenanalysen durchgeführt. Alle Analysen außer der Complete Case-Analysen wurden anhand der imputierten Werte durchgeführt. Tabelle 2 zeigt eine Darstellung aller durchgeführten Analysen.

Tabelle 2: Übersicht der gesundheitsökonomischen Analysen

Perspektive	Effekte	Population	Analyseart	Kosten
Gesellschaftliche Perspektive	QALY	Gesamte Studien-population	ITT ITT CC PP15 PP15OP ITT SA	Interventionskosten, Kosten des Index-Krankenhausaufenthalts, Inanspruchnahme von medizinischen Leistungen im Follow-Up (FIMA)
	Pflegegrad	Gesamte Studien-population	ITT ITT CC PP15 PP15OP ITT SA	Interventionskosten, Kosten des Index-Krankenhausaufenthalts, FIMA
	WHODAS 2.0	Gesamte Studien-population	ITT	Interventionskosten, Kosten des Index-Krankenhausaufenthalts, FIMA
GKV-Perspektive	QALY	BARMER-Versicherte	ITT ITT CC PP15 PP15OP ITT SA	Interventionskosten, Routinedaten
	Pflegegrad	BARMER-Versicherte	ITT ITT CC PP15 PP15OP ITT SA	Interventionskosten, Routinedaten
Leistungserbringer-Perspektive	Verweildauer ITS-Verweildauer ITS-Patient:innen Perioperative Komplikationen	Nur Patient:innen mit Operation	ITT	Krankenhaus Abrechnungsdaten gemäß §21 KHEntgG
Abkürzungen: QALY – quality adjusted life year; ITT – intention-to-treat, CC – complete cases; PP15 – Per Protocol Analyse mit mindestens 15 Prähabilitations-Einheiten; PP15OP – Per Protocol Analyse mit mindestens 15 Prähabilitations-Einheiten und Operation; SA – Sensitivitätsanalyse ohne winsorisierte Kostendaten.				

Die Per Protocol-Population wurde zweifach eingeteilt: Bei der PP15-Population mussten die Teilnehmenden der Interventionsgruppe mindestens 15 Einheiten der Prähabilitation erhalten haben und Teilnehmende aus der Kontrollgruppe durften keine Prähabilitation erhalten haben. Bei der PP15OP-Population mussten die Teilnehmenden der Interventions- und Kontrollgruppe zusätzlich eine Operation erhalten haben. Um die Vergleichbarkeit zwischen der Interventions- und Kontrollgruppe trotz ungleicher Kriterien aufrechtzuerhalten, wurde Propensity-Score-Matching (nearest neighbour) mit relevanten Baseline-Charakteristika durchgeführt: Alter, Geschlecht, BMI, Charlson-Comorbidity-Index, Gebrechlichkeit, EQ-5D-5L bei Baseline, Kognition (MOCA), funktionaler Gesundheitszustand,

Fachabteilung, Pflegegrad bei Baseline (nicht bei Analysen der Pflegegradverschlechterung), Kosten vor Studienbeginn (nicht bei Analyse aus gesellschaftlicher Perspektive).

Es wurden fünf vorab spezifizierte Subgruppenanalysen durchgeführt nach: Alter (dichotomisiert in $<$ vs. \geq Median), Geschlecht (männlich vs. weiblich), Gebrechlichkeit (vorgebrechlich vs. gebrechlich), Operationsart (orthopädische Operation bzw. neurochirurgische Operation an der Wirbelsäule vs. Herz-Operation vs. Tumor-Operation vs. sonstige Operationen) und Setting der Prähabilitation (ambulant (inkl. teilstationär) vs. stationär). Um für die Subgruppenanalyse nach Setting die Vergleichbarkeit zwischen der Interventions- und Kontrollgruppe trotz ungleicher Kriterien aufrechtzuerhalten, wurde auch hier das oben beschriebene Propensity-Score-Matching angewandt. Zusätzlich wurden aus der gesellschaftlichen Perspektive zwei post-hoc Subgruppenanalysen nach Pflegegrad bei Baseline (0 vs. ≥ 1) und Kognition (auffällig (MOCA < 26) vs. unauffällig (MOCA ≥ 26)) durchgeführt.

Umgang mit Unsicherheit

Um Unsicherheit in den Ergebnissen der durchgeführten Analysen darzustellen und mögliche Abweichungen von Verteilungsannahmen zu berücksichtigen, wurden mittels nicht-parametrischer Bootstrapping-Verfahren (10.000 Iterationen) 95%-Konfidenzintervalle für die Mittelwertdifferenzen der Kosten und Effekte berechnet. Zur klaren Einordnung und Interpretation der Hauptergebnisse wurden die Punktschätzer zusammen mit einer Konfidenzellipse in Kosten-Effektivitäts-Diagramme (Punktewolken) mit vier Quadranten dargestellt, um zwischen dominanten (rechter, unterer Quadrant: besser und billiger), dominierten (linker, oberer Quadrant: schlechter und teurer) oder fraglichen (rechter, oberer Quadrant: besser und teurer, oder linker, unterer Quadrant: schlechter und billiger) Situationen zu unterscheiden.

Kosten-Effektivitäts-Akzeptanz

Zur übersichtlichen Einordnung wurden anschließend alle Ergebnisse in Kosten-Effektivitäts-Akzeptanz-Diagramme übertragen, welche die Wahrscheinlichkeit der Kosten-Effektivität für Zahlungsbereitschafts-Schwellenwerte von 0 € bis 100.000 € darstellen. Dazu wurde der prozentuale Anteil an gebootstrappten Punktschätzern bestimmt, die bei einer gegebenen Zahlungsbereitschaft als kosteneffektiv gelten (d.h. bei denen die IKER unter dem jeweiligen Schwellenwert lag), und in Form von Kosten-Effektivitäts-Akzeptanz-Kurven dargestellt.

3 Ergebnisse der Evaluation

3.1 Deskription

Abbildung 1 zeigt das CONSORT Flow-chart für die PRÄP-GO Studie.

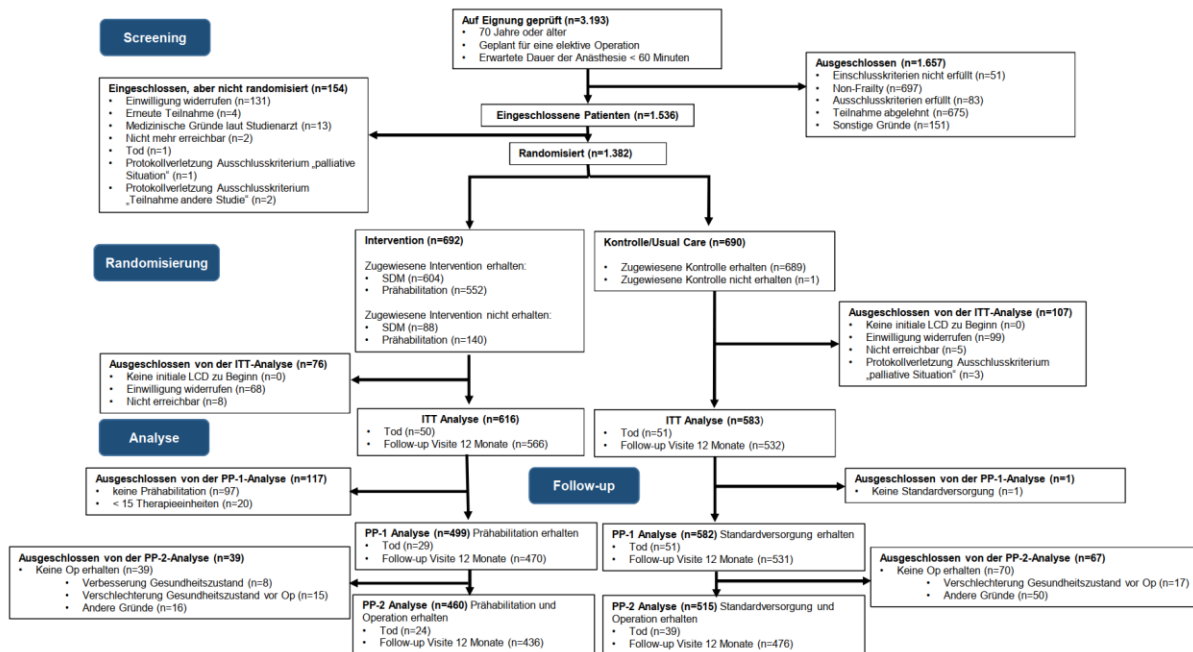


Abbildung 1: CONSORT Flow-Chart

In die Studie wurden 1382 Patienten in der Zeit vom 30.06.2020 bis 13.07.2023 (Datum des Endes des Patienteneinschluss) randomisiert eingeschlossen, davon 692 in die Interventionsgruppe und 690 in die Kontrollgruppe. Die Rekrutierung fand an 17 Zentren statt, die im Median 26 Patienten rekrutierten (Min 1, Max 346, IQR [3, 90]). Von den randomisierten Patient:innen wurden 1199 in die ITT (Intention to treat) Analyse eingeschlossen: 616 in die Interventions- und 583 in die Kontrollgruppe (siehe Abbildung 1).

Das mediane Alter in der ITT-Population beträgt 78 Jahre (IQR [74; 82]). Sie besteht aus 546 Männern (45.5%) und 653 Frauen (54.5%). Die darin vertretenden Baseline Pflegegrade haben die folgende Verteilung: Grad 0 bei 891 Studienteilnehmern (74.3%), Grad 1 bei 184 (15.3%), einen Pflegegrad über 1 haben 124 Teilnehmer (10.3%). Die Mortalität in der PRÄP-OP Studie ist gering: 6 Todesfälle vor der OP-Durchführung und insgesamt 51 Todesfälle in der Kontrollgruppe (51/583, 8.7%) und 50 Todesfälle in der Interventionsgruppe (50/616, 8.1%) (log-rank Test, $p=0.84$). Hinsichtlich der OP-Verläufe zwischen beiden Gruppen haben wichtige Parameter auch keine Evidenz für Unterschiede geliefert: Dauer des Krankenhausaufenthaltes ($p=0.371$), Dauer des Aufenthaltes auf der Intensivstation ($p=0.564$), Dauer der Anästhesie ($p=0.372$), Chirurgisches Risiko ($p=0.774$) und kardiales Risiko ($p=0.371$). Details finden sich in Tabelle 5.

Tabelle 3: Baseline-Eigenschaften nach Randomisierungsgruppe der ITT Population

Eigenschaften	Ausprägungen	Gesamt	Interventionsgruppe	Kontrollgruppe	P-Wert
n		1199	616	583	
Alter (median [IQR])		78.00 [74.00, 82.00]	78.00 [74.00, 82.00]	78.00 [73.00, 82.00]	0.740
BMI (median [IQR])		26.93 [23.99, 30.42]	27.04 [24.03, 30.48]	26.78 [23.82, 30.39]	0.651
Pflegescore (median [IQR])		3.75 [0.00, 12.50]	3.75 [0.00, 11.25]	3,75 [0.00, 13.75]	0.247
EQ5D Score (median [IQR])		0.73 [0.41, 0.87]	0.72 [0.42, 0.87]	0.74 [0.41, 0.87]	0.776
EQ5D VAS (median [IQR])		55.00 [50.00, 70.00]	50.00 [50.00, 70.00]	55.00 [50.00, 70.00]	0.399
IADL-Score (median [IQR])		8.00 [6.00, 8.00]	8.00 [6.00, 8.00]	8.00 [6.00, 8.00]	0.652
Barthel Index (median [IQR])		100.00 [95.00, 100.00]	100.00 [95.00, 100.00]	100.00 [95.00, 100.00]	0.914
MMQ-Score (median [IQR])		57.00 [45.00, 65.00]	57.00 [46.00, 64.00]	57.00 [44.00, 65.00]	0.668
MOCA-Score (median [IQR])		24.00 [21.00, 26.00]	24.00 [21.00, 26.00]	24.00 [21.00, 26.00]	0.837
NBA-Score (median [IQR])		3.75 [0.00, 12.50]	3.75 [0.00, 11.25]	3.75 [0.00, 13.75]	0.247
Charlson Comorbidity Index (median [IQR])		5.00 [4.00, 6.00]	5.00 [4.00, 6.00]	5.00 [4.00, 6.00]	0.785
Geschlecht (%)	männlich	546 (45.5)	288 (46.8)	258 (44.3)	0.418
	weiblich	653 (54.5)	328 (53.2)	325 (55.7)	
Pflegegrad (%)	0	891 (74.3)	469 (76.1)	422 (72.4)	0.180
	1	184 (15.3)	84 (13.6)	100 (17.2)	
	2	81 (6.8)	44 (7.1)	37 (6.3)	
	3	39 (3.3)	19 (3.1)	20 (3.4)	

Eigenschaften	Ausprägungen	Gesamt	Interventionsgruppe	Kontrollgruppe	P-Wert
	4	3 (0.3)	0 (0.0)	3 (0.5)	
	5	1 (0.1)	0 (0.0)	1 (0.2)	
Kategorischer MOCA Score (%)	< 26	757 (65.7)	387 (65.3)	370 (66.1)	0.820
	≥ 26	396 (34.3)	206 (34.7)	190 (33.9)	
Gewichtsverlust (%)	negativ	853 (71.1)	439 (71.3)	414 (71.0)	0.973
	positiv	346 (28.9)	177 (28.7)	169 (29.0)	
Muskelkraft (%)	negativ	588 (49.5)	303 (49.6)	285 (49.4)	0.992
	positiv	600 (50.5)	308 (50.4)	292 (50.6)	
Abgeschlagenheit (%)	negativ	327 (27.3)	162 (26.3)	165 (28.4)	0.475
	positiv	870 (72.7)	453 (73.7)	417 (71.6)	
Gehgeschwindigkeit (%)	negativ	468 (42.9)	237 (42.4)	231 (43.4)	0.779
	positiv	623 (57.1)	322 (57.6)	301 (56.6)	
Aktivitaet (%)	negativ	990 (82.6)	507 (82.3)	483 (82.8)	0.864
	positiv	209 (17.4)	109 (17.7)	100 (17.2)	
Gebrechlichkeit (%)	Vorgebrechlich	757 (63.1)	388 (63.0)	369 (63.3)	0.960
	Gebrechlich	442 (36.9)	228 (37.0)	214 (36.7)	
ASA (%)	normaler, gesunder Patient	15 (1.3)	5 (0.8)	10 (1.7)	0.266
	Patient mit leichter Allgemeinerkrankung	463 (38.6)	248 (40.3)	215 (36.9)	
	Patient mit schwerer Allgemeinerkrankung	667 (55.6)	339 (55.0)	328 (56.3)	
	Patient mit schwerer	54 (4.5)	24 (3.9)	30 (5.1)	

Eigenschaften	Ausprägungen	Gesamt	Interventionsgruppe	Kontrollgruppe	P-Wert
Demenz (%)	Allgemeinerkrankung, ständige Lebensbedrohung				
	nein	1176 (98.4)	604 (98.2)	572 (98.6)	0.738
	ja	19 (1.6)	11 (1.8)	8 (1.4)	
Index KH-Abteilung (%)	Allgemeinchir	71 (5.9)	44 (7.1)	27 (4.6)	0.134
	Augenheilkunde	100 (8.3)	44 (7.1)	56 (9.6)	
	Derma	1 (0.1)	0 (0.0)	1 (0.2)	
	Gefaesschir	32 (2.7)	16 (2.6)	16 (2.7)	
	Gyn	20 (1.7)	15 (2.4)	5 (0.9)	
	HNO	50 (4.2)	23 (3.7)	27 (4.6)	
	Innere Med	4 (0.3)	2 (0.3)	2 (0.3)	
	Kardio	98 (8.2)	41 (6.7)	57 (9.8)	
	Kardiochir	13 (1.1)	5 (0.8)	8 (1.4)	
	MKG	16 (1.3)	10 (1.6)	6 (1.0)	
	Neurochir	91 (7.6)	45 (7.3)	46 (7.9)	
	Ortho-/Unfallchirurgie	617 (51.5)	326 (52.9)	291 (49.9)	
	Urologie	86 (7.2)	45 (7.3)	41 (7.0)	
Leichte kognitive Einschränkungen (%)	negative	185 (16.2)	109 (18.4)	76 (13.9)	0.050
	positive	956 (83.8)	485 (81.6)	471 (86.1)	
Schwere kognitive Einschränkungen (%)	negative	601 (52.0)	312 (52.3)	289 (51.6)	0.847
	positive	555 (48.0)	284 (47.7)	271 (48.4)	
Raucherstatus (%)	rauchend	91 (7.6)	60 (9.8)	31 (5.3)	0.012
	früher	557 (46.6)	286 (46.5)	271 (46.6)	

Eigenschaften	Ausprägungen	Gesamt	Interventionsgruppe	Kontrollgruppe	P-Wert
	nie	548 (45.8)	269 (43.7)	279 (48.0)	
Alkoholkonsum (%)	negativ	931 (77.9)	484 (78.7)	447 (77.1)	0.542
	positiv	264 (22.1)	131 (21.3)	133 (22.9)	

Tabelle 3 zeigt die Verteilung relevanter Eigenschaften zwischen den Studienarmen. Bei zwei Eigenschaften (Raucherstatus, leichte kognitive Einschränkungen) zeigen sich mögliche Heterogenität zwischen den beiden Studienarmen (p-Wert des Vergleichstests ist kleiner 0.05).

Von den 616 Teilnehmer:innen, die in den Interventionsarm randomisiert wurden, liegen Angaben zur geplanten und durchgeführten Intervention vor. In dieser Gruppe gibt es 97 Patienten ohne Prähabilitation (97/616, 15.8%). Damit wurden 519 Patienten aktiv in Prähabilitationsmaßnahmen eingebunden. Bei 556 (90.3%) gibt es Angaben zur geplanten Prähabilitationseinrichtung. Bei 519 (84.3%) liegen dazu Angaben vor. Die Prähabilitation war bei 80 von 519 Teilnehmern (15.4%) stationär, bei 23 (4.4%) teilstationär und bei 416 (80.2%) ambulant oder mobil. Die Art der Prähabilitation (hier sind Doppelnennungen möglich) unter den Teilnehmern bestand aus den folgenden Komponenten: 511 Teilnehmer erhielten Training (511/519, 98.5%), 203 erhielten eine Ernährungsberatung (203/519, 39.1%), 28 eine kognitive/psychologische Unterstützung (28/519, 5.4%) und 253 erhielten weitere Maßnahmen (253/519, 48.7%). Im Mittel erhielten die Teilnehmer in der Interventionsgruppe 27 von 48 geplanten Prähabilitationseinheiten. Es gab 20 Teilnehmer (20/519, 3.8%) mit weniger als 15 durchgeführten Prähabilitationseinheiten. Details dazu finden sich in Tabelle 4.

Tabelle 4: Durchführung der Intervention, Beschreibung von Interventionseigenschaften

	Anzahl (Statistik)	Fehlende Werte (Prozent)
ITT Population:	N: 616	
Geplante Prähabilitationseinrichtungen	N: 556 (90.26%)	NA: 60 (9.74%)
Bis dahin verstorben	N: 0 (0%)	
Ambulant/Mobil	437 (78.6%)	
Teilstationär	35 (6.29%)	
Stationär	84 (15.11%)	
Geplante Prähabilitationskomponenten	N: 616 (100%)	NA: 0 (0%)
Bis dahin verstorben	N: 0 (0%)	
Training	550 (89.29%)	
Ernährungsberatung	112 (18.18%)	
Kognitive/Psychologische Unterstützung	28 (4.55%)	
Medikamentenanpassung	13 (2.11%)	
Andere/Verschiedene Unterstützung	78 (12.66%)	

	Anzahl (Statistik)	Fehlende Werte (Prozent)
Tatsächliche Prähabilitationseinrichtungen	N: 519 (84.25%)	NA: 97 (15.75%)
Bis dahin verstorben	N: 0 (0%)	
Stationär	80 (15.41%)	
Teilstationär	23 (4.43%)	
Ambulant/Mobil	416 (80.15%)	
Durchgeführte Prähabilitationskomponenten	N: 519 (84.25%)	NA: 97 (15.75%)
Bis dahin verstorben	N: 0 (0%)	
Training	511 (98.46%)	
Ernährungsberatung	203 (39.11%)	
Kognitive/Psychologische Unterstützung	28 (5.39%)	
Andere/Verschiedene Unterstützung	253 (48.75%)	
Anzahl an Prähabilitationskomponenten:	Median: 29.5 (IQR: 26;30)	Median: 29.5 (IQR: 26;30)
Anzahl an Prähabilitationskomponenten pro Teilnehmenden:	Mittelwert: 27.06 (SD: 5.27)	
Ohne Teilnehmer mit < 15 Prähabilitationseinheiten	N < 15 Prähabilitationskomponenten: 20 (3.25%)	
Training	492 (98.6%)	
Ernährungsberatung	199 (39.88%)	
Kognitive/Psychologische Unterstützung	27 (5.41%)	
Andere/Verschiedene Unterstützung	246 (49.3%)	
Anzahl an Prähabilitationskomponenten unter Teilnehmenden mit ausreichend Einheiten (>=15):	Median: 30 (IQR: 26.5;30)	Median: 30 (IQR: 26.5;30)
Gründe fehlende Prähabilitation:		
Medizinische Gründe	25 (25.77%)	
Organisatorische Gründe	23 (23.71%)	
Patientenwunsch	44 (45.36%)	
Sonstige Gründe	5 (5.15%)	

Insgesamt erhielten 132 der 1199 ITT-Teilnehmer:innen (11.0%) keine Operation. Das waren 64 der 616 Teilnehmer:innen der Interventionsgruppe (10.4%) und 68 der 583 Teilnehmer:innen der Kontrollgruppe (11.7%). Es verstarben 6 Teilnehmer vor der OP (Zwei in der Kontrollgruppe und 4 in der Interventionsgruppe). Wurde eine OP durchgeführt so lagen in der Interventionsgruppe im Mittel 56 Tage zwischen OP und Aufnahme in die Studie, bei der Kontrollgruppe 50 Tage. Dieser Unterschied wird als signifikant gesehen ($p=0.027$). Auch gab es Unterschiede in der Art der OP (offen, laparoskopisch, arthroskopisch, interventionell)

($p=0.018$). Andere Parameter in der OP-Phase wurden nicht als unterschiedlich bewertet. Detaillierte Informationen hierzu gibt Tabelle 5.

Tabelle 5: Eigenschaften der durchgeführten OP

Eigenschaften	Ausprägungen	Gesamt	Interventions- gruppe	Kontrollgruppe	P-Wert
N Keine Operation		132	64	68	
Information über Operation	Verbesserter Gesundheitszustand	15	15	0	
	Verschlechterter Gesundheitszustand	40	23	17	
	Andere Gründe	77	26	51	
N Operation erhalten		1067	552	515	
Tage zwischen Aufnahme und Operation (median [IQR])		54.00 [36.00, 92.50]	56.00 [39.00, 92.00]	50.00 [35.00, 93.50]	0.027
Dauer des Krankenhausaufenthalts (Tage) (median [IQR])		6.00 [4.00, 8.00]	6.00 [4.00, 8.00]	6.00 [4.00, 8.00]	0.371
Dauer des ITS-Aufenthalts (Stunden) (median [IQR])		0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.564
Dauer der Anästhesie (Stunden) (median [IQR])		2.35 [1.77, 3.17]	2.37 [1.80, 3.22]	2.33 [1.74, 3.05]	0.372
Schnitt-Naht-Zeit (Stunden) (median [IQR])		1.30 [0.92, 1.95]	1.32 [0.93, 1.98]	1.30 [0.91, 1.92]	0.382
Chirurgische Risikoskala (median [IQR])		8.00 [7.00, 9.00]	8.00 [7.00, 9.00]	8.00 [7.00, 9.00]	0.774
Art der Operation (%)	offen	854 (80.2)	456 (82.8)	398 (77.4)	0.018
	laparoskopisch	42 (3.9)	25 (4.5)	17 (3.3)	
	arthroskopisch	5 (0.5)	1 (0.2)	4 (0.8)	
	interventionell	164 (15.4)	69 (12.5)	95 (18.5)	
Tumoroperation (%)	ja	78 (7.3)	48 (8.7)	30 (5.8)	0.093

Eigenschaften	Ausprägungen	Gesamt	Interventions- gruppe	Kontrollgruppe	P-Wert
	nein	989 (92.7)	504 (91.3)	485 (94.2)	
Komplexität der OP/BUPA (%)	Geringfügig	13 (1.2)	3 (0.5)	10 (1.9)	0.164
	Mittelgradig	108 (10.1)	54 (9.8)	54 (10.5)	
	Größer	115 (10.8)	65 (11.8)	50 (9.7)	
	Größer Plus	436 (40.9)	233 (42.2)	203 (39.4)	
	Komplex Größer	395 (37.0)	197 (35.7)	198 (38.4)	
Anästhesieverfahren (%)	Allgemeinanästhesi e	666 (63.2)	347 (63.8)	319 (62.5)	0.453
	Allgemeinanästhesi e, Lokal- /Regionalanästhesi e	83 (7.9)	41 (7.5)	42 (8.2)	
	Regionale Anästhesie	113 (10.7)	60 (11.0)	53 (10.4)	
	Lokalanästhesie	45 (4.3)	19 (3.5)	26 (5.1)	
	Regionale Anästhesie, Lokalanästhesie	86 (8.2)	50 (9.2)	36 (7.1)	
	Andere	61 (5.8)	27 (5.0)	34 (6.7)	
	Andere	61 (5.8)	27 (5.0)	34 (6.7)	
Kardiales Risiko (%)	hohes Risiko	72 (6.7)	38 (6.9)	34 (6.6)	0.371
	moderates Risiko	790 (74.0)	417 (75.5)	373 (72.4)	
	niedriges Risiko	205 (19.2)	97 (17.6)	108 (21.0)	
Clavien-Dindo (%)	Keine Komplikation	689 (65.4)	361 (66.5)	328 (64.2)	0.896
	1	115 (10.9)	52 (9.6)	63 (12.3)	
	2	112 (10.6)	56 (10.3)	56 (11.0)	
	3	21 (2.0)	12 (2.2)	9 (1.8)	
	4	15 (1.4)	8 (1.5)	7 (1.4)	
	5	88 (8.3)	47 (8.7)	41 (8.0)	

Eigenschaften	Ausprägungen	Gesamt	Interventions- gruppe	Kontrollgruppe	P-Wert
	6	5 (0.5)	3 (0.6)	2 (0.4)	
	7	9 (0.9)	4 (0.7)	5 (1.0)	

Abbildung 2 zeigt individuelle Zeitpunkte, wann sich relevante Meilensteine der Studie ereignet haben. Die Zeitpunkte sind aufgeteilt nach den beiden Studienarmen sowie stratifiziert nach dem zu Baseline erhobenen Pflegegrad (0, 1, ≥ 2). Fehlende Zeitangaben wurden durch den Mittelwert der beobachteten Zeitpunkte innerhalb der Strata (wie beispielsweise v02) im Plot ersetzt und durch einen roten Punkt symbolisiert.

Die Deskription des NBA-Scores und des Pflegegrades sowie deren Veränderung zwischen Baseline und V18 sind in Tabelle 6 und Abbildung 3 gegeben. Tabelle 7 und Tabelle 8 beschreiben die Veränderungen des Pflegegrades in Kontroll- und Interventionsgruppe. In der Kontrollgruppe (Tabelle 8) haben sich 130 der 583 Studienteilnehmer verschlechtert (22,3%), 372 blieben im Pflegegrad gleich (63,8%) und 81 konnten sich verbessern (13,9%). In der Interventionsgruppe (Tabelle 7) haben sich 113 der 616 Studienteilnehmer verschlechtert (18,3%), 427 blieben im Pflegegrad gleich (69,3%) und 76 konnten sich verbessern (12,3%).

Tabelle 6: Deskription des NBA-Scores und seiner Veränderung

Studien-arm	Visite	Median	MW	SD	Mittlere Veränderung seit Baseline	Verstorben
Interventions- gruppe	v02	3.75	9.278	13.360	---	0 (0%)
Interventions- gruppe	v18	3.75	8.123	13.260	-0.65	50 (8.12%)
Kontrollgruppe	v02	3.75	9.908	13.806	---	0 (0%)
Kontrollgruppe	v18	3.75	10.301	17.412	0.92	51 (8.75%)

Tabelle 7: Veränderungen des Pflegegrades in der Interventionsgruppe

	v18: 0	v18: 1	v18: 2	v18: 3	v18: 4	v18: 5
v02: 0	391	31	12	2	0	33
v02: 1	47	18	10	1	0	8
v02: 2	11	11	13	5	0	4
v02: 3	0	4	3	5	1	6

Tabelle 8: Veränderungen des Pflegegrades in der Kontrollgruppe

	v18: 0	v18: 1	v18: 2	v18: 3	v18: 4	v18: 5
v02: 0	339	34	13	5	1	30
v02: 1	56	15	11	2	4	12
v02: 2	6	7	11	5	1	7
v02: 3	4	2	5	6	1	2
v02: 4	0	0	0	1	0	2
v02: 5	0	0	0	0	0	1



Abbildung 2: Individuelle Zeitpunkte der verschiedenen Untersuchungstermine, aufgeteilt nach Studienarmen und stratifiziert nach dem Pflegegrad der Baselineerhebung

Death = Tod des Studienteilnehmers; LtFu = Lost-to-Follow-up; med = Studienteilnehmer aus medizinischen Gründen aus Studie ausgeschieden; preha.start = Startzeitpunkt der Prähabilitation; preha.end = Endzeitpunkt der Prähabilitation; vxx = Indikator für die Visiten; wish = Ausscheiden eines Studienteilnehmers auf seinen Wunsch hin.

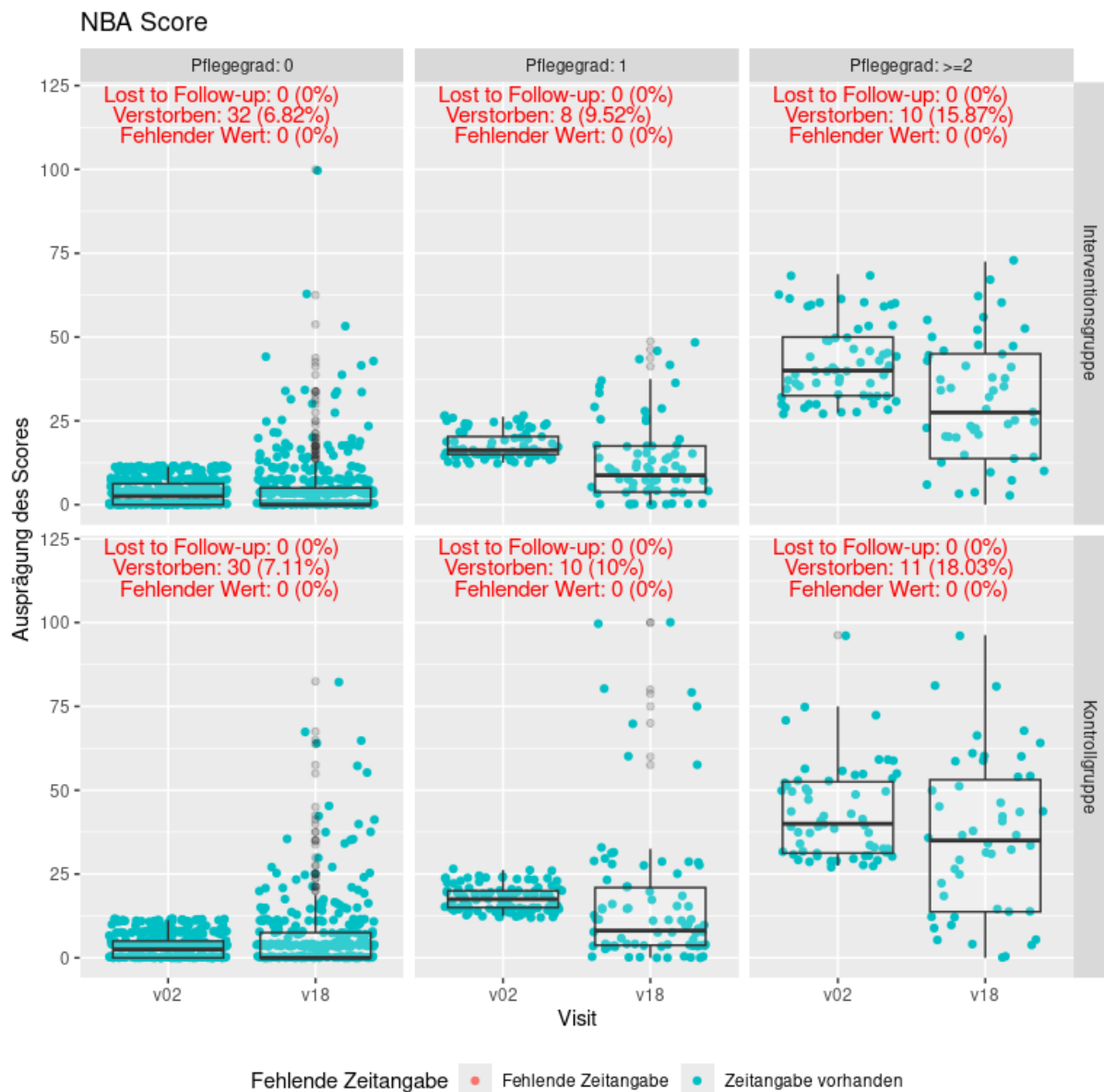


Abbildung 3: Boxplots zur Veränderung des NBA-Scores zwischen Baseline (v2) und einem Jahr nach OP (v18). Der Score reicht von 0 (keine Pflegebedürftigkeit) bis 100 (maximale Pflegebedürftigkeit). Fehlende Zeitangaben wurden durch den Mittelwert innerhalb der Strata (wie beispielsweise v02) im Plot ersetzt

3.2 Analysen des primären Endpunktes mit Sensitivitätsanalysen

Die primäre Analyse des primären Endpunktes in der ITT-Population (Veränderung im Pflegegrad) erfolgte mittels eines stratifizierten Mann-Whitney-Wilcoxon-Testes. Verstorbene Patienten werden dabei auf den schlechtesten Outcome gesetzt. Das Ergebnis war nicht signifikant ($p=0.3316$). Die Daten dazu zeigt Tabelle 9.

Tabelle 9: Ergebnisse des primären Endpunktes

Veränderung des Pflegegrades	≥ 2 Punkte	1 Punkt	keine	- 1 Punkt	≤ -2 Punkte
	Verschlechterung			Verbesserung	
Kontrolle N=583	77 (13.21%)	53 (9.09%)	372 (63.8%)	69 (11.8%)	12 (2.06%)
Intervention N=616	66 (10.71%)	47 (7.63%)	427 (69.3%)	61 (9.9%)	15 (2.44%)

Sensitivitätsanalysen zum primären Endpunkt zeigten keine Signale eines Interventionseffektes (Tabelle 12). Die Dichotomisierung der Veränderung des Pflegegrades hinsichtlich „verschlechtert“ versus „gleichgeblieben oder verbessert“ und die Analyse der Daten mittels logistischer Regression ergeben für die Intervention ein 1.257 (95% KI [0.946;1.67], $p=0.114$). Damit verändert die Intervention die Odds der Nicht-Verschlechterung um 1.257. Tabelle 10 zeigt die verwendeten Daten. Die AUC zur ROC-Kurve quantifiziert die Diskriminations-fähigkeit der Pflegegradveränderung zwischen den Studienarmen. Diese ist nicht signifikant von 0.5 verschieden (AUC = 0.489, 95% KI [(0.461;0.518)]. Die Verwendung von probabilistischen Index Modellen (PIMs) für eine stratifizierte Betrachtung der ROC-Kurven konnte keine signifikanten Unterschiede zwischen den Studienarmen nachweisen ($p=0.176$).

Tabelle 10: Dichotomisierte Veränderung des Pflegegrades: 0 – Verschlechterung, 1 – Gleichbleibend oder verbessert

Studienarm	Pflegegrad zur Baseline	Gleichgeblieben oder verbessert	Verschlechtert	Verstorben
Kontrollgruppe	Pflegegrad: 0	339	53	30
Kontrollgruppe	Pflegegrad: 1	71	19	10
Kontrollgruppe	Pflegegrad: ≥ 2	43	7	11
Interventionsgruppe	Pflegegrad: 0	391	46	32
Interventionsgruppe	Pflegegrad: 1	65	11	8
Interventionsgruppe	Pflegegrad: ≥ 2	47	6	10

Bei Sensitivitätsanalysen unter Verwendung des NBA-Scores zeigt sich ein signifikanter adjustierter mittlerer Unterschied zwischen den Veränderungen des NBA-Scores (DiD = -1.716, 95% KI [-3.137;-0.295], $p=0.018$). Das bedeutet, dass die Veränderung unter der Intervention im Mittel um -1.716 tiefer liegt als die Kontrolle. Da sich die Veränderung als Unterschied V18 – V02 berechnet und höhere Werte eine schlechtere Situation des Patienten reflektieren, stellt dies eine signifikante Verbesserung unter der Intervention dar.

Eine alternative Berücksichtigung von Todesfällen im Laufe des Follow-Up ermöglicht die Win-Ratio Methode. Das Ergebnis für den primären Endpunkt ist nicht signifikant. Für das Verhältnis zwischen Intervention und Kontrolle ergibt sich 1.078 (95% KI [0.933; 1.246], $p=0.305$).

3.3 Per Protokoll Analysen: Primärer Endpunkt und Sensitivitätsanalysen

Die Daten aus Tabelle 11 verdeutlichen die Fallzahlen in den drei Populationen.

Tabelle 11: Zusammensetzung der verschiedenen Per Protokoll Populationen

Operation	Prähabilitations-einheiten	Status	Populati-on	Häufigkeit
CG	Keine OP	Kontrolle (Keine Prähabilitation)	Lebend	56
CG	Keine OP	Kontrolle (Keine Prähabilitation)	Tot	12
CG	OP	Kontrolle (Keine Prähabilitation)	Lebend	476
CG	OP	Kontrolle (Keine Prähabilitation)	Tot	39
IG	Keine OP	zu wenig (weniger als 15)	Lebend	14
IG	Keine OP	zu wenig (weniger als 15)	Tot	11
IG	Keine OP	ausreichend (mehr gleich 15)	Lebend	34
IG	Keine OP	ausreichend (mehr gleich 15)	Tot	5
IG	OP	zu wenig (weniger als 15)	Lebend	82
IG	OP	zu wenig (weniger als 15)	Tot	10
IG	OP	ausreichend (mehr gleich 15)	Lebend	436
IG	OP	ausreichend (mehr gleich 15)	Tot	24

15E: Studienteilnehmer, die mindestens 15 Prähabilitationseinheiten erhielten

15E + OP: Studienteilnehmer, die mindestens 15 Prähabilitationseinheiten und eine OP erhielten

Principal Stratum: Studienteilnehmer, die mindestens 15 Prähabilitationseinheiten, eine OP erhielten und bei Studienvisite 18 noch am Leben waren.

Die drei Population sind hierarchisch gestaffelt: E15 enthält E15+OP. Beide enthalten das Principal Stratum.

Für die Auswertung sind drei Per Protokoll (PP) Populationen von Interesse:

- 1.) Patient:innen, die mindestens 15 Interventionseinheiten erhielten (15E, N=991=582(K)+499(I), eine Patientin der Kontrollgruppe hat an der Intervention teilgenommen),
- 2.) Patient:innen die mindestens 15 Interventionseinheiten und eine OP erhielten (15E+OP, N=975= 515(K)+460(I)),
- 3.) Patient:innen die mindestens 15 Interventionseinheiten, eine OP erhielten und während der Studie nicht gestorben sind (Principal Stratum, N=912=476(K)+436(I)).

Die Analyseergebnisse zum primären Endpunkt und damit zusammenhängende Sensitivitätsanalysen sind in Tabelle 12 für die drei Populationen dargestellt.

In den PP-Populationen kann die Hauptanalyse des primären Endpunktes nicht durchgeführt werden, da es keine gewichteten Ansätze für den Wilcoxon-Mann-Whitney Test gibt. Deshalb

wurden die Werte der Veränderung im Pflegegrad numerisch interpretiert und mittels linearen Regressionsmodellen mit robust geschätzten Standardfehler (Sandwich Schätzer) analysiert. Es zeigen sich marginale Vorteile hinsichtlich der Veränderung der Score-Differenzen (DiD: Differenz in Differenz) für die Intervention in den Populationen 15E und 15E+OP.

Untersucht man eine dichotomisierte Version der Veränderung (Verschlechterung (0) versus keine Verschlechterung (1)), so zeigen sich in den drei PP-Populationen signifikante Effekte durch ein Odds-Ratio, das Vorteile für die Intervention zeigt.

Ein vergleichbares Bild ergibt sich, wenn man die Veränderungen des NBA Scores mittels linearer Regression mit Sandwich Schätzer des Standardfehlers analysiert. Signifikante Veränderungen zeigen sich in den Populationen 15E und 15E+OP. In der Population des Principal Stratum ist der Effekt nicht signifikant ($p=0.115$). Dichotomisiert man die Veränderungen im NBA-Score (Verschlechterung (0) versus keine Verschlechterung (1)), so zeigen sich keine signifikanten Effekte.

Tabelle 12: Therapieeffekte zum primären Endpunkt.

Form Endpunkt	ITT	Interaktion	AUC	Principal Stratum	Per Protocol (15E)	Per Protocol (OP)	Per Protocol (15E + OP)
Veränderung Pflegegrad	Wilcoxon: p: 0.332	Nicht erhoben	Nicht erhoben	DiD: -0.046 (-0.15; 0.059) p:0.394	DiD: -0.069 (-0.169;0.031) p: 0.179	DiD: -0.075 (-0.179;0.029) p: 0.154	DiD: -0.075 (-0.181;0.031) p: 0.164
	DiD (GEE): -0.075 (-0.149;-0.001) p: 0.048						
Dichotom (basierend auf Verschlechterung des Pflegegrads; ja/nein)	OR: 1.257 (0.946; 1.67) p: 0.114	1.027 (0.727;1.45) p: 0.88	Gesamt: 0.489 (0.461-0.518) PIM Schätzer: 0.928 (0.832;1.036) p: 0.176	OR: 1.308 (0.926; 1.847) p:0.127	OR: 1.504 (1.038;2.178) p: 0.031	OR: 1.468 (1.02;2.114) p: 0.04	OR: 1.511 (1.025;2.227) p: 0.038
Veränderung NBA Score	Play the Winner: Win Ratio: 1.078 Win Proportion: 0.519 p: 0.305	Nicht erhoben	Nicht erhoben	DiD: -1.505 (-3.321; 0.312) p:0.105	DiD: -1.984 (-3.744;-0.224) p: 0.027	DiD: -1.82 (-3.637;-0.003) p: 0.05	DiD: -2.001 (-3.859;-0.143) p: 0.035
	DiD (GEE): -1.716 (-3.137;-0.295) p: 0.018						

Form Endpunkt	ITT	Interaktion	AUC	Principal Stratum	Per Protocol (15E)	Per Protocol (OP)	Per Protocol (15E + OP)
Dichotom (basierend auf Verschlechterung des NBA Scores; ja/nein)	OR: 1.026 (0.794; 1.326) p: 0.847	1.095 (0.736;1.63) p: 0.655	Gesamt: 0.483 (0.45-0.514) PIM Schätzer: 0.908 (0.796;1.035) p: 0.146	OR: 1.059 (0.82; 1.368) p:0.66	OR: 1.168 (0.891;1.531) p: 0.262	OR: 1.019 (0.779;1.333) p: 0.89	OR: 1.138 (0.856;1.512) p: 0.372

DiD (GEE): Unterschied der Veränderung zwischen beiden Studienarmen analysiert mit Generalized Estimation Equations (GEE);

AUC: Area under the ROC curve; OR: Odds Ratio, p: p-value; (): 95% confidence interval

Tabelle 13: Therapieeffekte bei ausgesuchten sekundären Endpunkten

Endpunkt	ITT	Principal Stratum	Per Protokoll (15E)	Per Protokoll (15E + OP)
Barthel Index	Wilcoxon: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: -1.557 (-3.123;0.009) p: 0.051	DiD: -1.642 (-3.296;0.012) p: 0.052
	OR : 1.274 (0.978; 1.66) p: 0.073			
	DiD (GEE): -1.33 (-2.559; -0.101) p: 0.034			
	Play the Winner: p-Wert > 0.1	OR: 1.286 (0.957; 1.728) p: 0.096	OR: 1.5 (1.088;2.069) p: 0.013	OR: 1.596 (1.13;2.253) p: 0.008
	AUC: p-Wert > 0.1			
CHARMI Score	Wilcoxon: p: 0.056	DiD: 0.162 (-0.031; 0.355) p: 0.099	DiD: 0.203 (0.023;0.383) p: 0.028	DiD: p-Wert > 0.1
	OR : p-Wert > 0.1			
	DiD (GEE): 0.147 (-0.01; 0.304) p: 0.066			
	Play the Winner: p-Wert > 0.1	OR: 0.767 (0.586; 1.005) p: 0.055	OR: 0.782 (0.59;1.037) p: 0.089	OR: p-Wert > 0.1
	AUC: p-Wert > 0.1			
Charles Comorbidity Index	Wilcoxon: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: 0.197 (0.052;0.342) p: 0.007
	OR : p-Wert > 0.1			
	DiD (GEE): p-Wert > 0.1			
	Play the Winner: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1	OR: 0.734 (0.539;1) p: 0.05
	AUC: p-Wert > 0.1			
	AUC: p-Wert > 0.1			
GAD7 Score	Wilcoxon: p-Wert > 0.1			

Endpunkt	ITT	Principal Stratum	Per Protokoll (15E)	Per Protokoll (15E + OP)
	OR : p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: -0.417 (-0.872;0.038) p: 0.073	DiD: -0.428 (-0.894;0.038) p: 0.072
	DiD (GEE): -0.446 (-0.824; -0.068) p: 0.021			
	Play the Winner: p-Wert > 0.1			
	AUC: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1
Gebrechlichkeit	Wilcoxon: p-Wert > 0.1	DiD: -0.074 (-0.162; 0.014) p: 0.099	DiD: -0.092 (-0.182;-0.002) p: 0.047	DiD: -0.106 (-0.202;-0.01) p: 0.029
	OR : p-Wert > 0.1			
	DiD (GEE): -0.084 (-0.16; -0.008) p: 0.032	OR: p-Wert > 0.1	OR: 1.4 (0.946;2.072) p: 0.093	OR: p-Wert > 0.1
	Play the Winner: Win Ratio: 1.187 Win Proportion: 0.543 p: 0.065			
	AUC: Gesamt: 0.469 (0.441-0.499) PIM Schätzer: 0.884 (0.784;0.996) p: 0.045			
Gehgeschwindigkeit	Wilcoxon: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: -0.078 (-0.166;0.01) p: 0.085	DiD: p-Wert > 0.1
	OR : p-Wert > 0.1			
	DiD (GEE): -0.077 (-0.155; 0.001) p: 0.056			
	Play the Winner: p-Wert > 0.1	OR: p-Wert > 0.1	OR: 1.539 (1.01;2.346) p: 0.045	OR: p-Wert > 0.1
	AUC: Gesamt: 0.473 (0.442-0.503) PIM Schätzer: 0.895 (0.786;1.019) p: 0.094			

Endpunkt	ITT	Principal Stratum	Per Protokoll (15E)	Per Protokoll (15E + OP)
Handkraft	Wilcoxon: p: 0.053	DiD: -0.849 (-1.564; -0.133) p: 0.02	DiD: -0.755 (-1.48;-0.03) p: 0.042	DiD: -0.701 (-1.456;0.054) p: 0.069
	OR : p-Wert > 0.1			
	DiD (GEE): p-Wert > 0.1			
	Play the Winner: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1
	AUC: p-Wert > 0.1			
IADL Score	Wilcoxon: p-Wert > 0.1	DiD: -0.198 (-0.401; 0.006) p: 0.057	DiD: -0.239 (-0.445;-0.033) p: 0.023	DiD: -0.191 (-0.407;0.025) p: 0.082
	OR : p-Wert > 0.1			
	DiD (GEE): -0.174 (-0.356; 0.008) p: 0.059	OR: 1.297 (0.994; 1.692) p: 0.055	OR: 1.267 (0.957;1.677) p: 0.099	OR: p-Wert > 0.1
	Play the Winner: Win Ratio: 1.166 Win Proportion: 0.538 p: 0.079			
	AUC: Gesamt: 0.471 (0.439-0.503) PIM Schätzer: 0.889 (0.784;1.008) p: 0.067			
Kategorieller ABC6 Score	Wilcoxon: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: -0.09 (-0.18;0) p: 0.048
	OR : p-Wert > 0.1			
	DiD (GEE): p-Wert > 0.1			
	Play the Winner: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1
	AUC: p-Wert > 0.1			
MNASF Score	Wilcoxon: p-Wert > 0.1			

Endpunkt	ITT	Principal Stratum	Per Protokoll (15E)	Per Protokoll (15E + OP)
	OR : p-Wert > 0.1			
	DiD (GEE): p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1
	Play the Winner: p-Wert > 0.1	OR: 0.777 (0.598; 1.01)		
	AUC: p-Wert > 0.1	p: 0.06	OR: p-Wert > 0.1	OR: p-Wert > 0.1
NBA: Bewältigung Krankheit und Therapie	Wilcoxon: p: 0.024			
	OR : p-Wert > 0.1			
	DiD (GEE): -0.165 (-0.298; -0.032) p: 0.016	DiD: -0.164 (-0.35; 0.022) p: 0.084	DiD: -0.173 (-0.34;-0.006) p: 0.042	DiD: -0.172 (-0.346;0.002) p: 0.052
	Play the Winner: p-Wert > 0.1			
	AUC: p-Wert > 0.1	OR: 1.417 (0.998; 2.012) p: 0.052	OR: 1.476 (1.011;2.155) p: 0.044	OR: p-Wert > 0.1
NBA: Eigenständigkeit	Wilcoxon: p-Wert > 0.1			
	OR : p-Wert > 0.1			
	DiD (GEE): - 0.532 (-1.03; -0.034) p: 0.036	DiD: p-Wert > 0.1	DiD: -0.601 (-1.263;0.061) p: 0.075	DiD: p-Wert > 0.1
	Play the Winner: p-Wert > 0.1			
	AUC: p-Wert > 0.1	OR: p-Wert > 0.1	OR: 1.484 (1.08;2.039) p: 0.015	OR: 1.487 (1.057;2.091) p: 0.023
NBA: Kognitions- und Konversationsfähigkeiten	Wilcoxon: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: -0.601 (-1.263;0.061) p: 0.075	DiD: p-Wert > 0.1
	OR : p-Wert > 0.1			

Endpunkt	ITT	Principal Stratum	Per Protokoll (15E)	Per Protokoll (15E + OP)
	DiD (GEE): -0.532 (-1.03; -0.034) p: 0.036			
	Play the Winner: p-Wert > 0.1			
	AUC: p-Wert > 0.1	OR: p-Wert > 0.1	OR: 1.484 (1.08;2.039) p: 0.015	OR: 1.487 (1.057;2.091) p: 0.023
NBA: Mobilität	Wilcoxon: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1
	OR : p-Wert > 0.1			
	DiD (GEE): p-Wert > 0.1			
	Play the Winner: p-Wert > 0.1			
	AUC: p-Wert > 0.1	OR: 1.497 (1.056; 2.123) p: 0.024	OR: p-Wert > 0.1	OR: 1.452 (0.972;2.17) p: 0.069
NBA: Organisation und soziale Kontakte	Wilcoxon: p-Wert > 0.1	DiD: -0.33 (-0.564; - 0.096) p: 0.006	DiD: -0.323 (-0.55;-0.096) p: 0.005	DiD: -0.329 (-0.576;-0.082) p: 0.009
	OR : p-Wert > 0.1			
	DiD (GEE): -0.27 (-0.47; -0.07) p: 0.008			
	Play the Winner: p-Wert > 0.1	OR: 1.418 (0.948; 2.121) p: 0.09	OR: 1.679 (1.063;2.651) p: 0.026	OR: 1.586 (0.966;2.604) p: 0.069
	AUC: p-Wert > 0.1			
Numerischer TUG Score	Wilcoxon: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: -0.851 (-1.796;0.094) p: 0.078	DiD: -0.881 (-1.883;0.121) p: 0.085
	OR : 1.29 (0.975; 1.707) p: 0.075			
	DiD (GEE): -0.932 (-1.83; -0.034) p: 0.042			
	Play the Winner: p-Wert > 0.1			

Endpunkt	ITT	Principal Stratum	Per Protokoll (15E)	Per Protokoll (15E + OP)
	AUC: Gesamt: 0.46 (0.42-0.5) PIM Schätzer: 0.847 (0.723;0.993) p: 0.041	OR: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1
PHQ8 Score	Wilcoxon: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1
	OR : p-Wert > 0.1			
	DiD (GEE): -0.375 (-0.808; 0.058) p: 0.089	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1
	Play the Winner: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1
	AUC: p-Wert > 0.1			
Schwere Kognitive Beeinträchtigung	Wilcoxon: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: -0.058 (-0.125;0.009) p: 0.091	DiD: -0.063 (-0.132;0.006) p: 0.075
	OR : p-Wert > 0.1			
	DiD (GEE): p-Wert > 0.1			
	Play the Winner: p-Wert > 0.1			
	AUC: p-Wert > 0.1	OR: p-Wert > 0.1	OR: 1.497 (0.973;2.304) p: 0.067	OR: 1.528 (0.966;2.417) p: 0.07
TMTAB Score	Wilcoxon: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: -0.061 (-0.132;0.01) p: 0.084	DiD: -0.07 (-0.143;0.003) p: 0.058
	OR : p-Wert > 0.1			
	DiD (GEE): p-Wert > 0.1			
	Play the Winner: p-Wert > 0.1			
	AUC: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1
Two Minute Step Test	Wilcoxon: p-Wert > 0.1			

Endpunkt	ITT	Principal Stratum	Per Protokoll (15E)	Per Protokoll (15E + OP)
	OR : p-Wert > 0.1			
	DiD (GEE): -4.848 (-9.097; -0.599) p: 0.025			
	Play the Winner: Win Ratio: 1.14 Win Proportion: 0.533 p: 0.099			
	AUC: Gesamt: 0.461 (0.427-0.496) PIM Schätzer: 0.858 (0.742;0.992) p: 0.037	OR: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1
WHODAS score	Wilcoxon: p: 0.07			
	OR : p-Wert > 0.1			
	DiD (GEE): -0.8 (-1.709; 0.109) p: 0.085	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1	DiD: p-Wert > 0.1
	Play the Winner: p-Wert > 0.1			
	AUC: p-Wert > 0.1			
		OR: p-Wert > 0.1	OR: p-Wert > 0.1	OR: p-Wert > 0.1

DiD (GEE): Unterschied der Veränderung zwischen beiden Studienarmen analysiert mit Generalized Estimation Equations (GEE); AUC: Area under the ROC curve; OR: Odds Ratio, p: p-Wert; (): 95% Konfidenzintervall, PIM: Probabilistische Index ModellEs wurden nur diejenigen sekundären Endpunkte dargestellt, bei denen mindestens eine der gezeigten Analysen einen p-Wert von ≤ 0.1 aufwies.

3.4 Analysen der sekundären Endpunkte

Die Ergebnisse in diesem Abschnitt sind als explorativ zu bewerten.

Die Ergebnisse (ITT Population und Per Protokoll Populationen) für ausgesuchte sekundäre Endpunkte sind in Tabelle 13 aufgeführt. In den Spalten finden sich die Ergebnisse zu den vier Analysepopulationen: ITT, Principal Stratum, 15E und 15E+OP.

Pro Endpunkte zeigt die erste Zeile das Ergebnis des stratifizierten Wilcoxon-Mann-Whitney-Testes für die Veränderung des Endpunktes (numerisch oder ordinal). In der zweiten Zeile werden die Veränderungen des Endpunktes dichotomisiert (Verschlechterung gegen keine Verschlechterung) und mittels logistischer Regression stratifiziert in den vier interessierenden Populationen berechnet. In der dritten Zeile werden die Veränderungen des Endpunktes numerisch betrachtet und mittels GEEs in den vier interessierenden Populationen berechnet. In der fünften Zeile wird für die Diskriminierung der Populationen zwischen beiden Armen eine AUC berechnet. Gegebenenfalls wird mittels eines Probabilistic Impact Modells (PIM) für die stratifizierte Situation die Nullhypothese analysiert, dass die AUC in allen Strata gleich 0.5 ist. Zu diesem Hypothesentest wird ein zweiseitiger p-Wert berechnet.

Signifikante Unterschiede ($p < 0.05$) zwischen den Therapiearmen zeigten bei mindestens einer der ITT-Analysen (zweite Spalte) für die sekundären Endpunkte: Barthel Index, die Gebrechlichkeit, den GAD7 Score, den numerischen TUG-Score und dem Two-Minute-Step Test, sowie für die NBA Unterkategorien Bewältigung Krankheit und Therapie, Eigenständigkeit, Kognitions- und Konversationsfähigkeiten und Organisation und soziale Kontakte. Signale für einen Unterschied ($p < 0.1$) in der ITT-Analyse zeigten sich bei Handkraft, CHARMI, der Gehgeschwindigkeit, IADL, PHQ8 und WHODAS. In den Sensitivitätsanalysen zeigten sich mehr Signale und signifikante Effekte sowohl in der ITT Population und verschiedenen PP-Populationen.

3.5 Interaktion zwischen Intervention und Patienteneigenschaften

Die Ergebnisse in diesem Abschnitt sind als explorativ zu bewerten. Signale ($p \leq 0.1$) und Signifikanzen ($p < 0.05$) für die Interaktion zwischen Prognosescore und Intervention wurden bei den folgenden Endpunkten gefunden: CHARMI ($p = 0.056$), EQ5D ($p = 0.063$), MiniCog ($p = 0.012$) und numerischer TUG-Score ($p = 0.089$). Bei den verschiedenen dichotomen Versionen zum primären Endpunkt wurden keine Interaktionen gefunden. (Daten im Detail nicht gezeigt).

3.6 Nebenwirkungen und Komplikationen, Mortalität

Die Mortalität in der Präp-GO Studie ist gering: 6 Todesfälle vor der OP-Durchführung und insgesamt 51 Todesfälle in der Kontroll- (51/583, 8.7%) und 50 Todesfälle in der Interventionsgruppe (50/616, 8.1%). Hinsichtlich der OP-Verläufe zwischen beiden Gruppen haben wichtige Parameter auch keine Evidenz für Unterschiede geliefert: Dauer des Krankenhausaufenthaltes ($p = 0.371$), Dauer des Aufenthaltes auf der Intensivstation ($p = 0.564$), Dauer der Anästhesie ($p = 0.372$), Chirurgisches Risiko ($p = 0.774$) und kardiales Risiko ($p = 0.371$). Details finden sich in [Tabelle 5](#).

3.7 Ergebnisse der gesundheitsökonomischen Evaluation

3.7.1 Ergebnisse der gesundheitsökonomischen Evaluation aus gesellschaftlicher Perspektive

Kosten

Die Zusammensetzung der Interventionskosten aus der gesellschaftlichen Perspektive ist in Tabelle 14 dargestellt. Die Kosten aller gescreeenten Patient:innen ($n = 3.199 \cdot 19,00 \text{ €}$) wurden durch die Anzahl aller positiv gescreeenten Patient:innen ($n = 2.362$) geteilt und dann pro Person in der Interventionsgruppe veranschlagt (25,73 €, gerundet 26 €).

Insgesamt wurde für 556 Patient:innen eine SDM-Konferenz durchgeführt. Die Teilnahme der Berufsgruppen war wie folgt verteilt: Personen aus der Berufsgruppe Anästhesie nahmen an 551 Konferenzen teil, Geriatrie an 536, Therapieberufe an 525, Chirurgie an 267, Pflege an 52, Allgemeinmedizin an 17. Zusätzlich nahmen die Patient:innen selbst an 402 Konferenzen teil und deren Angehörige an 124 Konferenzen.

Nach einer erfolgten SDM-Konferenz haben 519 Patient:innen mind. eine Einheit an Prähabilitation erhalten. Die Kosten hierfür waren im ambulanten Setting am niedrigsten, gefolgt von mobiler Prähabilitation, ambulante Prähabilitation mit Fahrtkosten, teilstationäre Prähabilitation und - mit deutlichem Abstand - der stationären Prähabilitation. Die mittleren Gesamt-Interventionskosten lagen bei 2.414 €.

Tabelle 14: Interventionskosten gesellschaftliche Perspektive (in €)

Interventionsbestandteil	Interventionsgruppe (n = 616)				
	MW	SD	Median	1. Quartil	3. Quartil
Frailty-Screening	26	0	26	26	26
SDM-Konferenz*	309	116	332	287	387
Prähabilitation*	2.079	2.808	1.195	644	1.749
mobile Prähabilitation ** (n = 4)	831	155	798	747	882
ambulante Prähabilitation** (n = 69)	608	222	644	579	713
ambulante Prähabilitation mit Fahrtkosten** (n = 343)	1.360	577	1.277	1.006	1.580
teilstationäre Prähabilitation** (n = 23)	1.900	399	2.029	1.662	2.254
stationäre Prähabilitation** (n = 80)	9.069	924	9.056	8.883	9.217
Gesamtkosten	2.414	2.837	1.546	1.025	2.122

* Enthält auch Werte i.H.v. 0 € für Teilnehmende ohne SDM-Konferenz bzw. Prähabilitation.

** Enthält nur die Kosten der tatsächlich beanspruchten Leistung. Wenn nicht beansprucht für die durchschnittlichen Gesamtkosten mit 0 € bewertet.

MW, Mittelwert; SD, Standardabweichung.

Eine Übersicht der Gesamtkosten nach Imputation fehlender Werte und Winsorisieren ist in Tabelle 15 dargestellt. Entsprechende Übersichten ohne Imputation bzw. Winsorisierung finden sich im Anhang (Tabelle A1 und A2). Die Interventionskosten der Kontrollgruppe betrugen gemäß Studiendesign 0 €, was zu einem hohen Unterschied in den Kosten dieses Kostenblocks zwischen den Gruppen führte (+ 2.414 €). Die mittleren Gesamtkosten für den Index-Krankenhausaufenthalt betrugen zwischen 10.000 und 11.000 €. Sie fielen in der Interventionsgruppe etwas niedriger aus (-312 €). Dieser Unterschied zeigte sich allerdings nicht, wenn man die medianen Kosten betrachtet (7.996 € vs. 7.982 € (+14€)).

Die mittleren Kosten durch die Leistungsanspruchnahme im Follow-Up betrugen in der Interventionsgruppe knapp 13.000 € und in der Kontrollgruppe knapp 14.000 € (-855 €). Die größten Kostenblöcke machten dabei vollstationäre Krankenhausaufenthalte, informelle Pflege und Rehabilitation aus. Hier gab es außerdem die größten Unterschiede zwischen den Gruppen. Die mittleren Gesamtkosten betrugen 26.621 € in der Interventionsgruppe und 25.418 € in der Kontrollgruppe (+ 1.203 €).

Die mittleren winsorisierten Gesamtkosten ohne Imputation lagen mit 23.296 € in der Interventionsgruppe und 21.398 € in der Kontrollgruppe etwa 3.000 € bis 3.800 € unter den mittleren winsorisierten Gesamtkosten mit Imputation, da hier die fehlenden Werte mit 0 € ersetzt wurden. Die mittleren Gesamtkosten ohne Winsorisierung mit Imputation lagen mit 28.996 € in der Interventionsgruppe und 29.008 € in der Kontrollgruppe etwa 2.400 € und 3.600 € über den mittleren winsorisierten Gesamtkosten mit Imputation. Sie waren in der Interventionsgruppe und Kontrollgruppe in etwa gleich hoch (-12 €). Ausreißer wurden vor allem bei den Kosten für vollstationäre Krankenhausaufenthalte und die Inanspruchnahme eines Pflegedienstes im Follow-Up beobachtet.

Kosten-Effektivitäts-Analyse des QALY und WHODAS 2.0

Die inkrementellen Kosten und Effekte und daraus resultierenden IKER für die ITT-Analyse mit imputierten Werten (ITT), die ITT-Analyse mit Complete cases (ITT CC), die Per-Protokollanalysen PP15 und PP15+OP, und die Sensitivitätsanalyse ohne Winsorisierung der Kosten (ITT SA) sind in Tabelle 16 dargestellt.

Die Hauptanalyse aus gesellschaftlicher Perspektive ist dabei die ITT für das QALY. Hier betrugen die inkrementellen Kosten 1.203 € (95% KI: -1.620; 3.997) und die inkrementellen Effekte 0,03 QALYs (95% KI: -0,01; 0,06). Daraus ergab sich eine IKER von 45.547 € pro gewonnenem QALY. Für die ITT-Analyse mit Complete Case des QALYs konnten nur 296 Patient:innen analysiert werden, da bei den restlichen Personen mindestens ein EQ-5D-5L-Wert im Follow-Up fehlte. Hier zeigten sich geringere inkrementelle Kosten bei höheren QALY-Werten und somit insgesamt eine niedrigere IKER von 8.054 € pro gewonnenem QALY. Bei den Per-Protocol-Analysen waren die inkrementellen Kosten etwas niedriger als bei der ITT-Analyse und es wurden höhere QALY-Werte als in der ITT-Analyse beobachtet, die außerdem signifikant höher waren als in der Kontrollgruppe. Daraus resultierten kleinere ICER i.H.v. 16.752 € und 19.520 €. Bei der Sensitivitätsanalyse waren die inkrementellen Kosten negativ und die inkrementellen Effekte positiv, was zu einer Situation führt, wo die Intervention die Kontrolle dominiert, also besser und billiger ist.

Tabelle 15: Gesamtkosten gesellschaftliche Perspektive nach Winsorisierung und Imputation (in €)

Kostenblock	Interventionsgruppe (n = 616)					Kontrollgruppe (n = 538)				
	MW	SD	Median	Q1	Q3	MW	SD	Median	Q1	Q3
Interventionskosten	2.414	2.837	1.546	1.025	2.122	0	0	0	0	0
KH-Indexaufenthalt*	10.540	8.065	7.996	5.621	11.334	10.852	8.621	7.982	5.105	11.958
Follow-Up Kosten Total	12.994	18.452	5.914	2.868	14.235	13.849	18.461	6.742	3.461	14.973
Krankenhaus (stationär)	5.580	10.961	607	0	5.742	5.875	11.499	1.012	0	6.121
Krankenhaus (ambulant)	147	423	0	0	0	138	437	0	0	0
Psychiatrie	0	0	0	0	0	0	0	0	0	0
Arztbesuche	679	430	616	385	876	689	453	626	355	907
Therapieberufe	708	672	508	191	1.030	715	718	489	190	1.023
Pflegedienst	677	1.859	0	0	418	886	2.200	0	0	475
Informelle Pflege	2.722	11.138	0	0	0	2.879	12.032	0	0	0
Haushaltshilfe	106	301	0	0	0	103	289	0	0	0
Hilfe von Nachbarn etc.	314	1.398	0	0	0	435	1.615	0	0	0
Tagespflege	0	0	0	0	0	0	0	0	0	0
Kurzzeitpflege	0	0	0	0	0	0	0	0	0	0
Pflegeheim	228	1.213	0	0	0	137	949	0	0	0
Rehabilitation	1.609	2.022	684	0	3.248	1.754	2.040	1.072	0	3.040
Hilfsmittel	225	423	61	0	184	238	466	61	0	154
Gesamtkosten	26.621	25.088	18.360	12.312	32.846	25.418	24.617	16.724	11.316	30.169

*enthält auch Werte i.H.v. 0 € für Teilnehmende ohne Operation.

IG, Interventionsgruppe; KG, Kontrollgruppe; KH, Krankenhaus; KI, Konfidenzintervall; MW, Mittelwert; SD, Standardabweichung.

Beim WHODAS 2.0 wurde eine nicht statistisch-signifikante Verbesserung um 0,97 Punkte (95% KI: 0,62; 2,54) beobachtet, was einer IKER von 1.241 € pro gewonnenem Punkt entspricht.

Tabelle 16: Ergebnisse der Kosten-Effektivitäts- und -Nutzwert-Analysen aus gesellschaftlicher Perspektive (QALY und WHODAS 2.0)

Analyse	N		Kosten IG		Kosten KG		Inkrementelle Kosten		Effekte IG		Effekte KG		Inkrementelle Effekte		IKER
QALY	IG	KG	MW	SD	MW	SD	MWD	95% KI	MW	SD	MW	SD	MWD	95% KI	€/QALY
ITT	616	583	26.621	25.088	25.418	24.617	1.203	- 1.620; 3.997	0,72	0,28	0,69	0,29	0,03	-0,01; 0,06	45.547
CC	167	129	25.325	24.506	24.901	26.401	424	- 5.623; 6.279	0,69	0,35	0,64	0,36	0,05	-0,03; 0,13	8.054
PP15	473	473	25.879	23.177	24.939	24.714	940	- 2.116; 3.976	0,76	0,25	0,70	0,29	0,06	0,02; 0,09	16.752
PP15OP	391	391	26.596	23.343	25.642	25.120	954	- 2.406; 4.315	0,77	0,24	0,72	0,28	0,05	0,01; 0,09	19.520
ITT SA	616	583	28.996	31.326	29.008	35.463	-12	- 3.793; 3.813	0,72	0,28	0,69	0,29	0,03	-0,01; 0,06	dominant
WHODAS 2.0	IG	KG	MW	SD	MW	SD	MWD	95% KI	MW	SD	MW	SD	MWD	95% KI	€/Punkt
ITT	616	583	26.621	25.088	25.418	24.617	1.203	- 1.617; 3.976	26,06	13,48	27,03	14,48	0,97	-0,62; 2,54	1.241

IG, Interventionsgruppe; IKER, Inkrementelle Kosten-Effektivitäts-Relation; KG, Kontrollgruppe; KI, Konfidenzintervall; MW, Mittelwert; MWD, Mittelwertdifferenz; SD, Standardabweichung; QALY, Qualitätsadjustiertes Lebensjahr, WHODAS 2.0, World Health Organization Disability Assessment Schedule 2.0.

Die Unsicherheit der ITT-Analyse des QALY ist als Streudiagramm in Abbildung 4 dargestellt. Die Mehrzahl der gebootstrappten Ergebnisse befindet sich im rechten, oberen Quadranten, gefolgt vom rechten, unteren Quadranten. Ein kleinerer Teil der Punkte befindet sich außerdem im linken, oberen Quadranten und ein sehr kleiner Teil im linken unteren Quadranten.

Die Ergebnisse aller weiteren Analysen finden sich in der vereinfachten Darstellung in Abbildung 5. Hier befindet sich die Mehrzahl der gebootstrappten Ergebnisse im rechten, oberen Quadranten und im rechten, unteren Quadranten.

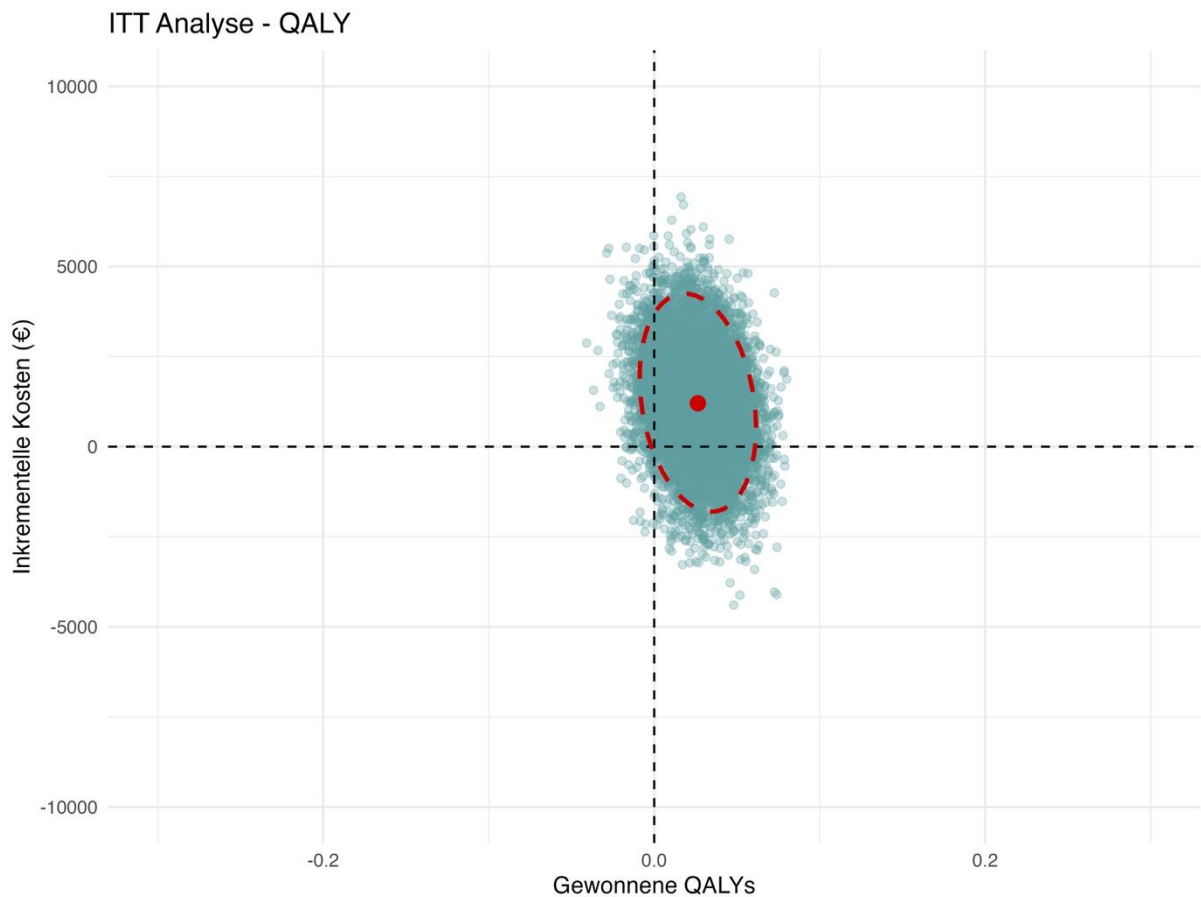


Abbildung 4: Streudiagramm der ITT-Analyse des QALY aus gesellschaftlicher Perspektive

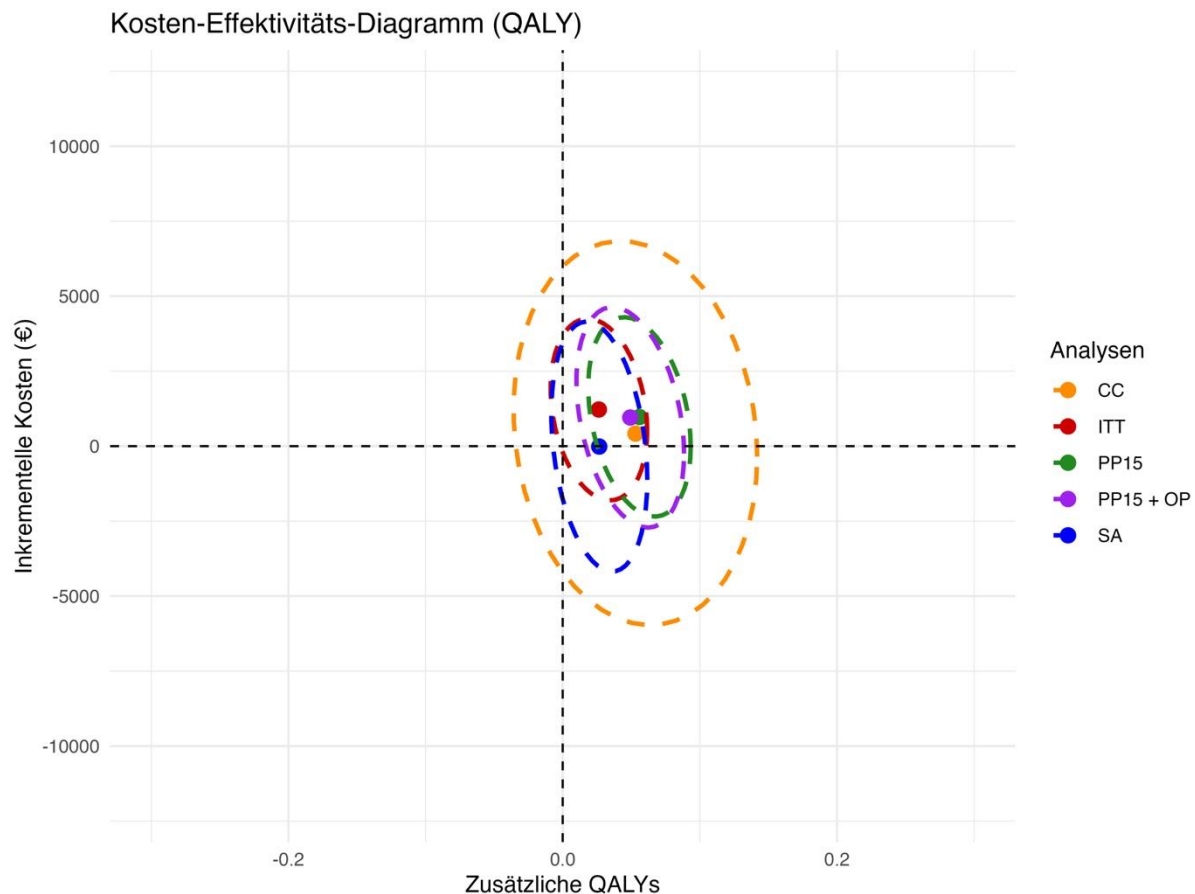


Abbildung 5: Streudiagramm aller Analysen des QALY aus gesellschaftlicher Perspektive

Kosten-Effektivitäts-Analyse der Pflegegradverschlechterung

Die Ergebnisse der gesundheitsökonomischen Evaluation der Wahrscheinlichkeit der Verschlechterung von Baseline zu 12-Monaten postoperativ (PGV) sind in Tabelle 17 dargestellt. Bei der ITT-Analyse für die PGV betrugen die inkrementellen Kosten 1.203 € (95% KI: -1.596; 4.051) und die inkrementellen Effekte 4,42 % vermiedene PGV (95% KI: -0,05; 8,94). Daraus ergab sich eine IKER von 27.197€ pro vermiedener PGV. Die inkrementellen Kosten der Complete Case-Analyse und der Per-Protocol-Analysen waren hier höher (PP15OP) bzw. deutlich höher (PP15) als bei der ITT-Analyse, was bei vergleichbarer (PP15OP) bzw. höherer (PP15) Effektivität zu größeren IKER führte. Die inkrementellen Kosten waren bei der Sensitivitätsanalyse negativ, was ebenfalls wieder zu einer Situation führte, wo die Intervention die Kontrolle dominierte.

Tabelle 17: Ergebnisse der Kosten-Effektivitäts- und -Nutzwert-Analysen aus gesellschaftlicher Perspektive (Pflegegradverschlechterung)

Analyse	N		Kosten IG		Kosten KG		Inkrementelle Kosten		PGV IG		PGV KG		Vermiedene PGV		IKER €/verm. PGV
	IG	KG	MW	SD	MW	SD	MWD	95% KI	N	%	N	%	MWD	95% KI	
ITT	616	583	26.621	25.088	25.418	24.617	1.203	- 1.596; 4.051	108	17,53	128	21,96	4,42	-0,05; 8,94	27.197
CC	609	581	23.265	23.184	21.382	23.340	1.883	- 769; 4.579	106	17,41	128	22,03	4,63	0,05; 9,14	40.712
PP15	470	470	26.896	24.268	24.245	22.190	2.651	- 282; 5.676	67	14,26	99	21,06	6,81	1,83; 11,58	38.938
PP15OP	382	382	26.513	23.089	24.985	22.604	1.528	- 1.726; 4.795	54	14,14	72	18,85	4,71	-0,56; 9,96	32.427
ITT SA	616	583	28.996	31.326	29.008	35.463	-12	- 3.784; 3.854	108	17,53	128	21,96	4,42	0,02; 9,03	dominant

IG, Interventionsgruppe; IKER, Inkrementelle Kosten-Effektivitäts-Relation; KG, Kontrollgruppe; KI, Konfidenzintervall; MW, Mittelwert; MWD, Mittelwertdifferenz; PGV, Pflegegradverschlechterung; SD, Standardabweichung.

Die Unsicherheit der ITT-Analysen ist als Streudiagramm für die Pflegegradverschlechterung in Abbildung 6 dargestellt. Auch hier liegt die Mehrzahl der gebootstrappten Ergebnisse im rechten, oberen Quadranten, gefolgt vom rechten, unteren Quadranten. Ein kleinerer Teil der Punkte befindet sich außerdem im linken, oberen Quadranten und ein sehr kleiner Teil im linken unteren Quadranten.

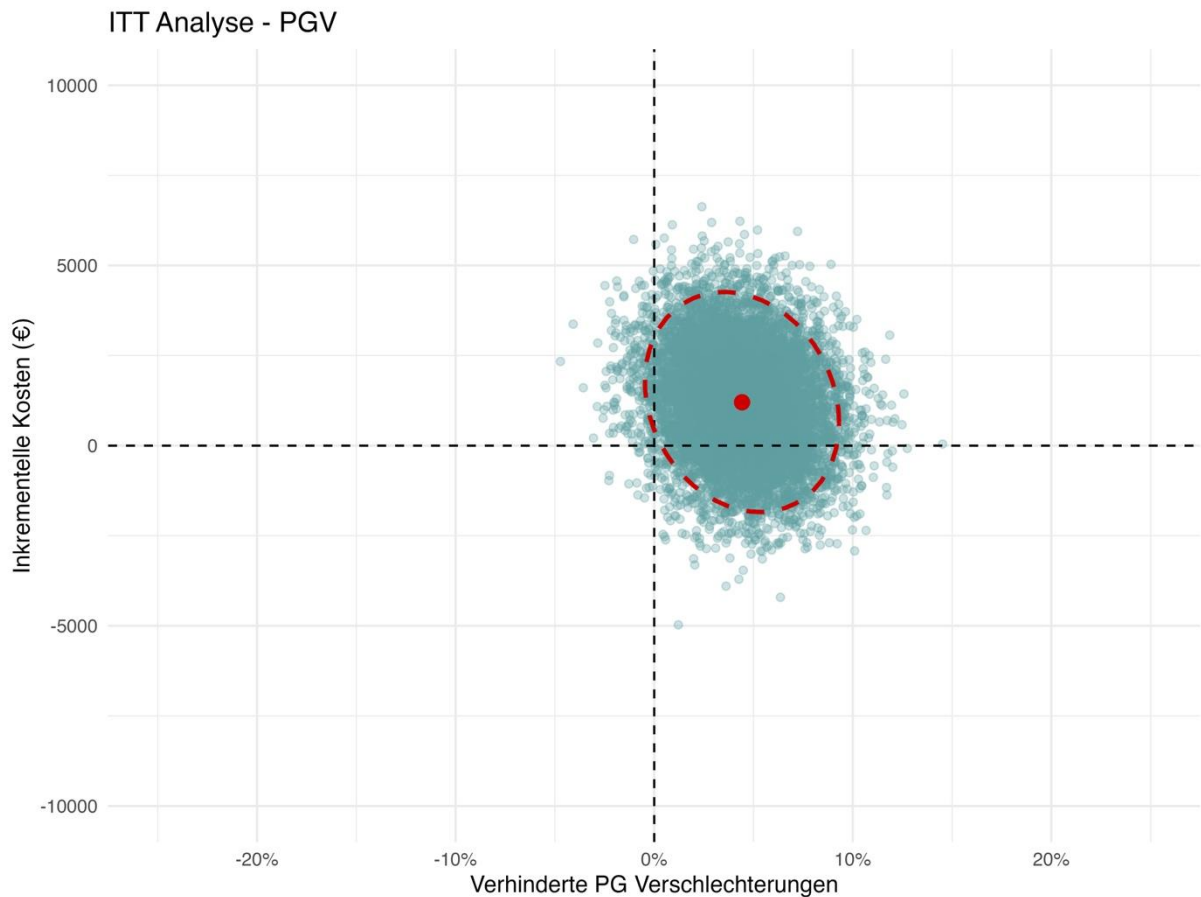


Abbildung 6: Streudiagramm der ITT-Analyse der Pflegegradverschlechterung aus der gesellschaftlichen Perspektive

Die Ergebnisse aller weiteren Analysen finden sich in der vereinfachten Darstellung in Abbildung 7. Hier befinden sich die Mehrzahl der gebootstrappten Ergebnisse im rechten, oberen Quadranten, gefolgt vom rechten, unteren Quadranten.

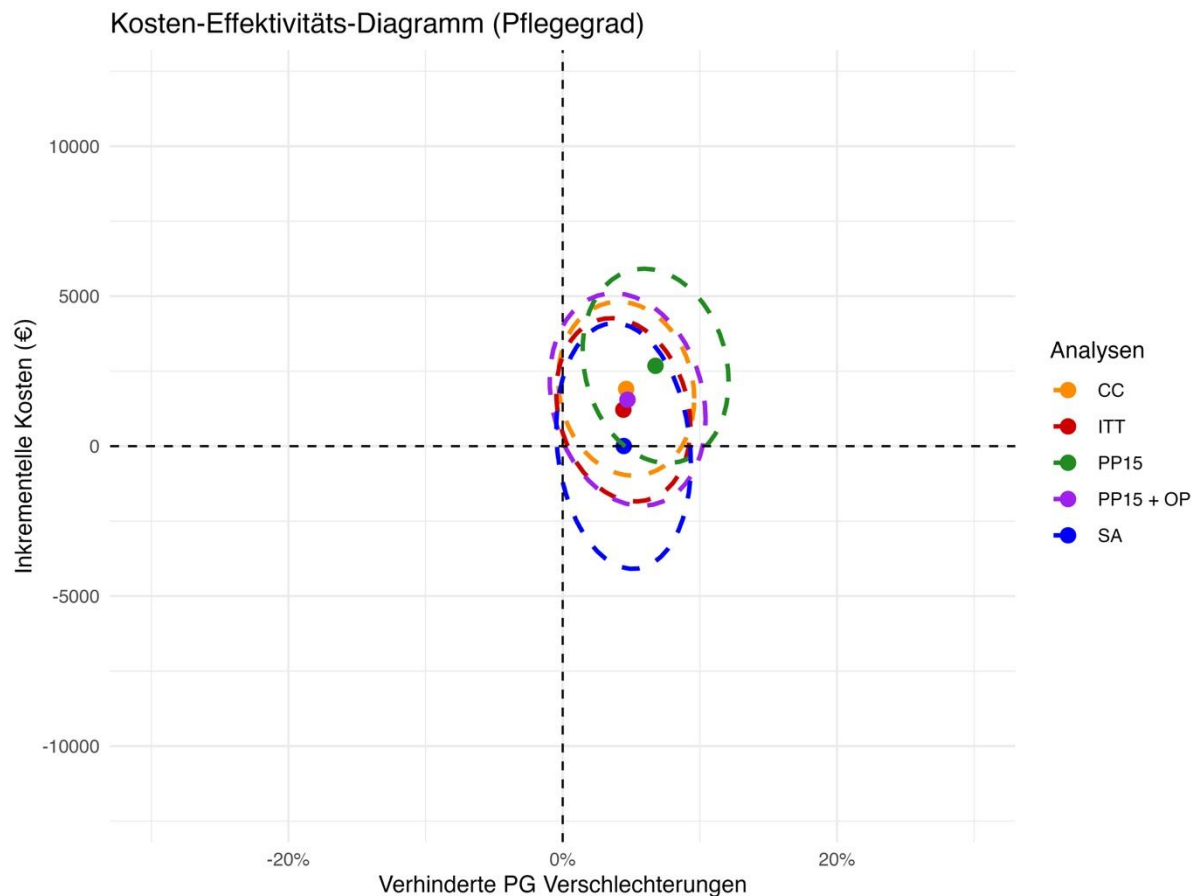


Abbildung 7: Streudiagramm aller Analysen der Pflegegradverschlechterung aus gesellschaftlicher Perspektive

Wahrscheinlichkeit der Kosteneffektivität

Um die Wahrscheinlichkeit der Kosteneffektivität der Intervention beurteilen zu können, wurden oben gezeigte Ergebnisse in ein Kosteneffektivitäts-Akzeptanz-Diagramm eingetragen, welches die Wahrscheinlichkeit der einzelnen Analysen unter Annahme von verschiedenen hohen, hypothetischen Zahlungsbereitschaften darstellt. Die Ergebnisse sind in Abbildung 8 dargestellt.

Bei einer Zahlungsbereitschaft von 0 € für ein zusätzliches QALY, liegt die Wahrscheinlichkeit, dass die Intervention als kosteneffektiv akzeptiert wird bei der ITT bei knapp 20 %. Diese erhöht sich bei einer Zahlungsbereitschaft von 50.000 € auf ca. 52 % und bei einer Zahlungsbereitschaft von 100.000 € auf ca. 72 %. Die anderen Analysen zeigen höhere Kosten-Effektivitäts-Akzeptanzwerte, beginnend bei 27 % (Per-Protocol-Analyse PP15) bis 50 % (Sensitivitätsanalyse). Wahrscheinlichkeit für Kosten-Effektivitäts-Akzeptanz bei einer Zahlungsbereitschaft von 0 €. Bei einer Zahlungsbereitschaft von 100.000 € erreichen die Complete Case-Analyse und Sensitivitätsanalyse eine Wahrscheinlichkeit von ca. 82 % und die beiden Per-Protocol-Analysen von 91 % (PP15OP) und 96 % (PP15).

Bei einer Zahlungsbereitschaft von 0 € pro verhinderter PGV liegt die Wahrscheinlichkeit, dass die Intervention als kosteneffektiv akzeptiert wird, bei der ITT ebenfalls bei knapp 20 %. Diese erhöht sich bei einer Zahlungsbereitschaft von 50.000 € auf ca. 69 % und bei einer Zahlungsbereitschaft von 100.000 € auf ca. 86 %. Die Complete Case-Analyse und Per-Protocol-Analyse PP15OP zeigen durchweg niedrigere Werte, die übrigen Analysen höhere Werte als die ITT (die PP15-Analyse jedoch erst ab 71.000 €). Bei einer Zahlungsbereitschaft

von 0 € zeigt hier die Sensitivitätsanalyse mit 50 % die höchste Wahrscheinlichkeit. Die höchste Wahrscheinlichkeit insgesamt erreicht ebenfalls die Sensitivitätsanalyse mit 92 %, gefolgt von der PP15-Analyse mit 90 % bei einer Zahlungsbereitschaft von 100.000 €.

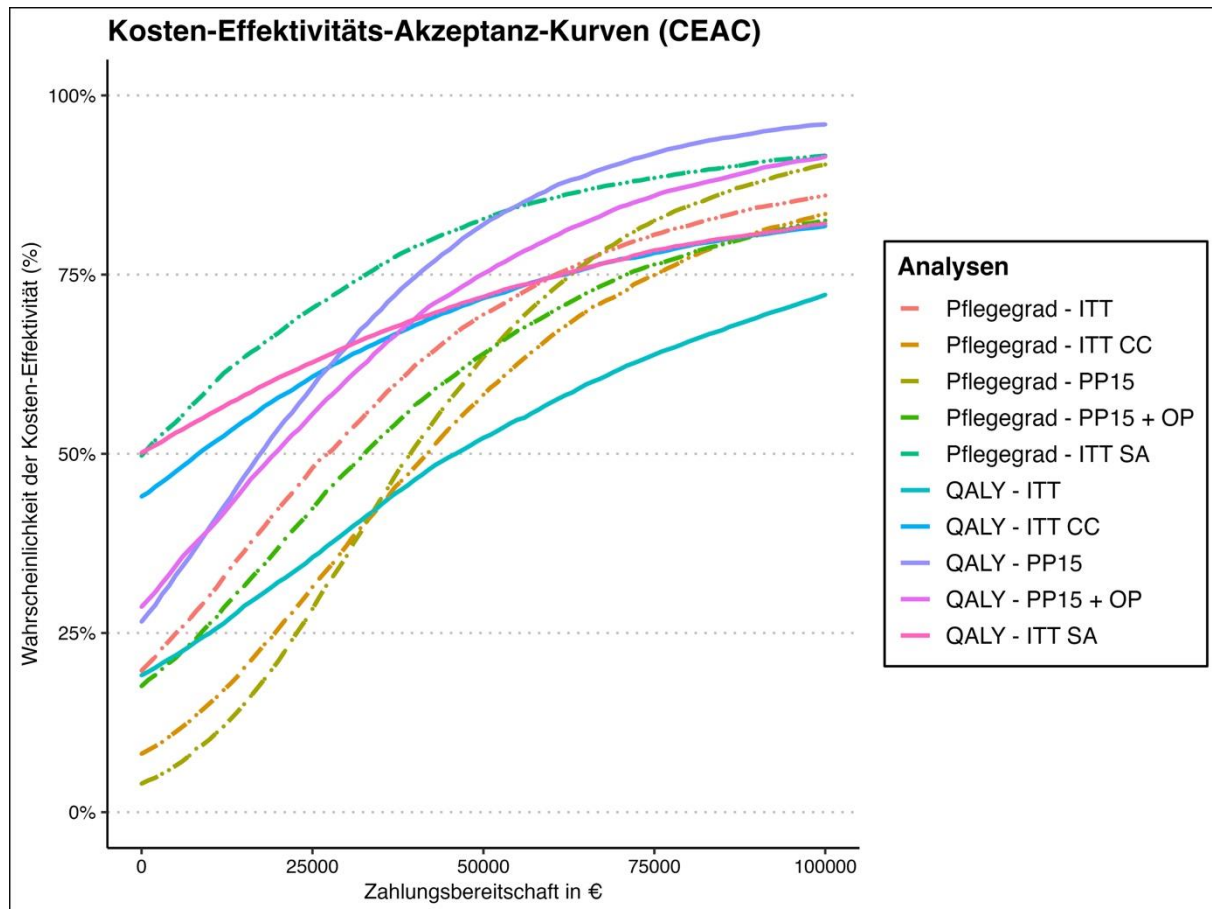


Abbildung 8: Kosten-Effektivität-Akzeptanz-Diagramm für QALY und Pflegegradverschlechterung aus gesellschaftlicher Perspektive

Subgruppenanalysen

Eine Übersicht der inkrementellen Kosten und Effekte und daraus resultierenden IKER für die Subgruppenanalysen (alle basierend auf der ITT-Analyse mit imputierten Daten) finden sich in Tabelle 18 und Tabelle 19. Bei den Subgruppenanalysen zeigte sich beim Vergleich nach Alter bessere Ergebnisse für die Patient:innen, deren Alter unter dem Median lag. Die inkrementellen Kosten waren hier klein (150 €, 95% KI: -4.074; 4.243) und die inkrementellen Effekte waren statistisch signifikant höher als in der Kontrollgruppe (0,05 QALYs, 95% KI: 0,00; 0,10, Vermiedene PGV bei 8,73 %, 95% KI: 2,40; 15,00). Bei Patient:innen, deren Alter auf oder über dem Median lag, war der beobachtete inkrementelle Effekt sehr klein (0,003 QALYs, 95% KI: -0,04; 0,05, Vermiedene PGV bei 0,65 %, 95% KI: -5,70; 7,15), was bei erhöhten inkrementellen Kosten zu einer sehr hohen IKER von 759.219 € bzw. 337.605 € führte. Beim Vergleich nach Geschlecht wurden für männliche Patienten etwas niedrigere inkrementelle Kosten und ca. 50 % höhere inkrementelle Effekte als bei weiblichen Patientinnen beobachtet. Vorgebrechliche Patient:innen hatten im Vergleich zu gebrechlichen Patient:innen nur ein Fünftel der inkrementellen Kosten, jedoch auch niedrigere (Pflegegradveränderung) bzw. deutlich niedrigere (QALY) inkrementelle Effekte.

Tabelle 18: Subgruppenanalyse ITT Qualitätsadjustiertes Lebensjahr aus gesellschaftlicher Perspektive

QALY	N	Kosten IG		Kosten KG		Inkrementelle Kosten			QALY IG		QALY KG		Gewonnene QALYs			IKER
	IG	KG	MW	SD	MW	SD	MWD	95% KI	MW	SD	MW	SD	MWD	95% KI		€/QALY
Alter < Median	285	257	24.635	23.467	24.485	26.367	150	-4.074 4.243	0,76	0,26	0,70	0,31	0,05	0,00 0,10		2.828
Alter ≥ Median	331	326	28.330	26.320	26.153	23.160	2.178	-1.594 6.125	0,69	0,29	0,69	0,28	0,00	-0,04 0,05		759.219
Weiblich	328	325	25.191	22.151	24.028	20.113	1.163	-2.025 4.493	0,72	0,26	0,70	0,28	0,02	-0,02 0,06		51.495
Männlich	288	258	28.249	28.012	27.168	29.272	1.080	-3.786 5.870	0,71	0,30	0,68	0,31	0,03	-0,02 0,08		34.003
Gebrechlich	228	214	31.076	31.429	28.738	23.339	2.338	-2.735 7.552	0,60	0,31	0,55	0,32	0,05	-0,01 0,11		45.466
Vor- gebrechlich	388	369	24.002	20.056	23.492	25.160	510	-2.757 3.632	0,79	0,23	0,78	0,24	0,01	-0,02 0,05		39.768
Orthop. OP*	370	335	25.705	22.847	24.199	21.162	1.506	-1.743 4.810	0,76	0,25	0,70	0,27	0,06	0,02 0,10		25.789
Tumor-OP	53	35	31.139	40.627	19.227	20.885	11.912	-356 25.458	0,63	0,34	0,80	0,23	-0,17	-0,29 -0,05		wird dominiert
Herz-OP	46	65	45.845	30.634	50.996	37.270	-5.150	- 17.752 7.498	0,55	0,37	0,59	0,35	-0,04	-0,18 0,09		120.178
Andere OP	148	150	21.217	17.285	18.281	17.940	2.936	-1.212 6.914	0,71	0,27	0,70	0,30	0,01	-0,05 0,08		231.291
Ambulantes Setting**	422	422	23.738	21.876	25.207	25.718	-1.469	-4.736 1.757	0,78	0,23	0,71	0,29	0,07	0,04 0,11		dominant
Stationäres Setting**	70	70	42.040	33.481	23.564	19.204	18.476	9.788 28.069	0,57	0,32	0,69	0,28	-0,12	-0,21 -0,02		wird dominiert
PG = 0	469	422	22.722	19.829	22.850	22.428	-128	-3.019 2.656	0,77	0,25	0,76	0,25	0,01	-0,03 0,04		dominant
PG ≥ 1	147	161	39.058	34.444	32.147	28.599	6.911	-143 14.139	0,57	0,30	0,51	0,31	0,05	-0,02 0,12		129.224
Kogn. auffällig	409	387	28.956	27.518	26.673	23.509	2.283	-1.245 5.894	0,67	0,30	0,66	0,31	0,01	-0,03 0,05		245.493
Kogn. unauffällig	207	196	22.005	18.629	22.939	26.559	-934	-5.648 3.417	0,82	0,19	0,76	0,24	0,06	0,02 0,10		dominant

*Beinhaltet neurochirurgische Operationen an der Wirbelsäule **Propensity Score gematched

IG, Interventionsgruppe; IKER, Inkrementelle Kosten-Effektivitäts-Relation; KG, Kontrollgruppe; KI, Konfidenzintervall; MW, Mittelwert; MWD, Mittelwertdifferenz; SD, Standardabweichung; QALY, Qualitätsadjustiertes Lebensjahr.

Tabelle 19: Subgruppenanalyse ITT Pflegegradverschlechterung aus gesellschaftlicher Perspektive

Pflegegrad	N		Kosten IG		Kosten KG		Inkrementelle Kosten		PGV IG		PGV KG		Vermiedene PGV		IKER	
	IG	KG	MW	SD	MW	SD	MWD	95% KI	N	%	N	%	%-Diff	95% KI	€/verm. PGV	
Alter < Median	285	257	24.635	23.467	24.485	26.367	150	-4.074 4.243	35	12,28	54	21,01	8,73	2,40 15,00	1.714	
Alter ≥ Median	331	326	28.330	26.320	26.153	23.160	2.178	-1.594 6.125	73	22,05	74	22,70	0,65	-5,70 7,15	337.605	
Weiblich	328	325	25.191	22.151	24.028	20.113	1.163	-2.025 4.493	53	16,16	64	19,69	3,53	-2,40 9,31	32.911	
Männlich	288	258	28.249	28.012	27.168	29.272	1.080	-3.786 5.870	55	19,10	64	24,81	5,71	-1,32 12,74	18.926	
Gebrechlich	228	214	31.076	31.429	28.738	23.339	2.338	-2.735 7.552	61	26,75	69	32,24	5,49	-2,97 14,01	42.597	
Vor-gebrechlich	388	369	24.002	20.056	23.492	25.160	510	-2.757 3.632	47	12,11	59	15,99	3,88	-1,15 8,87	13.168	
Orthop. OP*	370	335	25.705	22.847	24.199	21.162	1.506	-1.743 4.810	48	12,97	70	20,90	7,92	2,40 13,45	19.013	
Tumor-OP	53	35	31.139	40.627	19.227	20.885	11.912	-356 25.458	15	28,30	6	17,14	-11,16	-28,44 6,86	wird dominiert	
Herz-OP	46	65	45.845	30.634	50.996	37.270	-5.150	-17.752 7.498	17	36,96	21	32,31	-4,65	-22,94 13,33	110.785	
Andere OP	148	150	21.217	17.285	18.281	17.940	2.936	-1.212 6.914	28	18,92	32	21,33	2,41	-6,64 11,69	121.610	
Ambulantes Setting**	414	414	23.914	21.451	24.493	22.283	-579	-3.563 2.454	52	12,56	87	21,01	8,45	3,36 13,54	dominant	
Stationäres Setting**	76	76	46.033	36.478	24.179	22.387	21.854	12.558 31.758	18	23,68	17	22,37	-1,32	-14,48 11,92	wird dominiert	
PG = 0	469	422	22.722	19.829	22.850	22.428	-128	-3.019 2.656	76	16,20	82	19,43	3,23	-1,85 8,19	dominant	
PG ≥ 1	147	161	39.058	34.444	32.147	28.599	6.911	-143 14.139	32	21,77	46	28,57	6,80	-3,21 16,64	101.592	
Kogn. auffällig	409	387	28.956	27.518	26.673	23.509	2.283	-1.245 5.894	94	22,98	95	24,55	1,56	-4,38 7,45	145.914	
Kogn. unauffällig	207	196	22.005	18.629	22.939	26.559	-934	-5.648 3.417	14	6,76	33	16,84	10,07	3,70 16,28	dominant	

*Beinhaltet neurochirurgische Operationen an der Wirbelsäule **Propensity Score gematched

IG, Interventionsgruppe; IKER, Inkrementelle Kosten-Effektivitäts-Relation; KG, Kontrollgruppe; KI, Konfidenzintervall; MW, Mittelwert; MWD, Mittelwertdifferenz; PGV, Pflegegradverschlechterung; SD, Standardabweichung.

Beim Vergleich nach Operationstyp zeigten sich die besten Ergebnisse für Patient:innen mit einer orthopädischen Operation bzw. einer neurochirurgischen Operation an der Wirbelsäule. Die inkrementellen Kosten lagen hier bei 1.506 € (95 % KI: -1.743; 4.810) und die inkrementellen Effekte waren statistisch signifikant höher als in der Kontrollgruppe (0,06 QALYs, 95% KI: 0,02; 0,10, Vermiedene PGV bei 7,92 %, 95% KI: 2,40; 13,45). Bei Patient:innen mit einer Tumor-Operation wurde die Interventionsgruppe durch die Kontrollgruppe dominiert, da die inkrementellen Kosten hier signifikant höher (11.912 €, 95% KI: -356; 25.458) und die inkrementellen Effekte negativ waren (-0,17 QALYs, 95% KI: -0,29; -0,05, Vermiedene PGV bei -11,16 %, 95% KI: -28,44; 6,86); beim QALY statistisch signifikant geringer. Auch bei Patient:innen mit Herz-Operationen waren die inkrementellen Effekte niedriger, jedoch auch die inkrementellen Kosten. Beim Vergleich nach Setting war das ambulante Setting dem stationären Setting eindeutig überlegen. Sowohl beim QALY als auch beim Pflegegrad dominierte die Intervention und die inkrementellen Effekte waren statistisch signifikant höher als in der Kontrollgruppe (0,07 QALYs, 95% KI: 0,04; 0,11, Vermiedene PGV bei 8,45 %, 95% KI: 3,36; 13,54). Bei Patient:innen mit Prähabilitation im stationären Setting wurde die Intervention von der Kontrolle dominiert. Die inkrementellen Kosten waren hier jeweils statistisch signifikant höher (QALY: 18.476 €; 95%-KI: 9.788; 28.069, PGV: 21.854 €, 95%-KI: 12.558; 31.758) und die inkrementellen Effekte negativ (-0,12 QALYs, 95%-KI: -0,21; -0,02, Vermiedene PGV bei -1,32%, 95%-KI: -14,48; 11,92); beim QALY statistisch signifikant geringer. Es ist anzumerken, dass n = 3 Patient:innen sowohl Teil der Subgruppe orthopädische Operation als auch der Subgruppe Tumor-Operation waren.

Bei den post-hoc Subgruppenanalysen nach Pflegegrad bei Baseline und Kognition dominierte die Intervention die Kontrolle bei Patient:innen mit Pflegegrad 0 und unauffälliger Kognition. Bei Letzteren waren die inkrementellen Effekte statistisch signifikant erhöht (0,06 QALYs, 95% KI: 0,02; 0,10, Vermiedene PGV bei 10,07 %, 95% KI: 3,70; 16,28), während bei Patient:innen mit Pflegegrad 0 die inkrementellen Effekte geringer waren als bei der Gesamtpopulation (0,01 QALYs, 95%-KI: -0,03; 0,04, Vermiedene PGV bei 3,23 %, 95%-KI: -1,85; 8,19), aber die inkrementellen Kosten negativ (-128 €, 95%-KI: -3.019; 2.656).

Die Wahrscheinlichkeiten für Kosten-Effektivitäts-Akzeptanz der Analysen und Subgruppenanalysen sind in Abbildung 9 und Abbildung 10 (je für QALY und vermiedene PGV) dargestellt. Die Subgruppen von Patient:innen mit Tumor-Operation und Prähabilitation im stationären Setting zeigen hier bei allen Zahlungsbereitschaften Wahrscheinlichkeiten von unter 5 %, als kosteneffektiv akzeptiert zu werden. Die Subgruppe von Patient:innen mit Prähabilitation im ambulanten Setting zeigt durchweg die höchste Wahrscheinlichkeit, als kosteneffektiv akzeptiert zu werden. Diese liegt bei einer Zahlungsbereitschaft von 0 € beim QALY bei ca. 80 % und bei der PGV bei ca. 63 %. Diese Subgruppe wird gefolgt von kognitiv unauffälligen Patient:innen (MOCA \geq 26) und Patient:innen, deren Alter unter dem Median liegt. Die Subgruppe von Patient:innen mit einer orthopädischen Operation kommt auf ähnlich hohe Wahrscheinlichkeiten wie die drei genannten Gruppen, jedoch erst ab einer Zahlungsbereitschaft von etwa 75.000 € und mehr. Die Subgruppe von Patient:innen mit einer Herz-Operation hat bei einer Zahlungsbereitschaft von 0 € eine Wahrscheinlichkeit von über 75 % kosteneffektiv zu sein, jedoch sinkt diese auf knapp über 50 % bei einer Zahlungsbereitschaft von 100.000 €. Dies hängt damit zusammen, dass sowohl die inkrementellen Kosten als auch die inkrementellen Effekte hier in der Interventionsgruppe geringer waren, sodass eine erhöhte Zahlungsbereitschaft nicht zu einer höheren Kosteneffektivitätsakzeptanz führt.

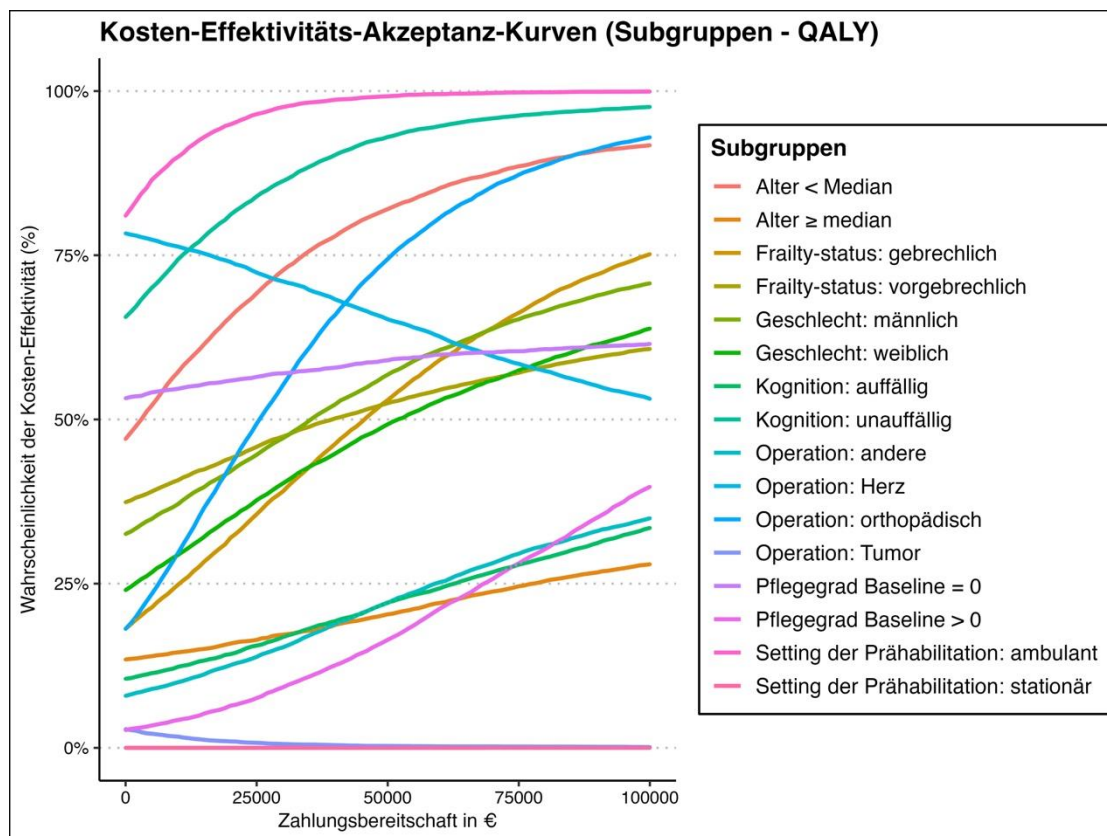


Abbildung 9: Kosten-Effektivität-Akzeptanz-Diagramm für QALY aus gesellschaftlicher Perspektive – Subgruppenanalyse

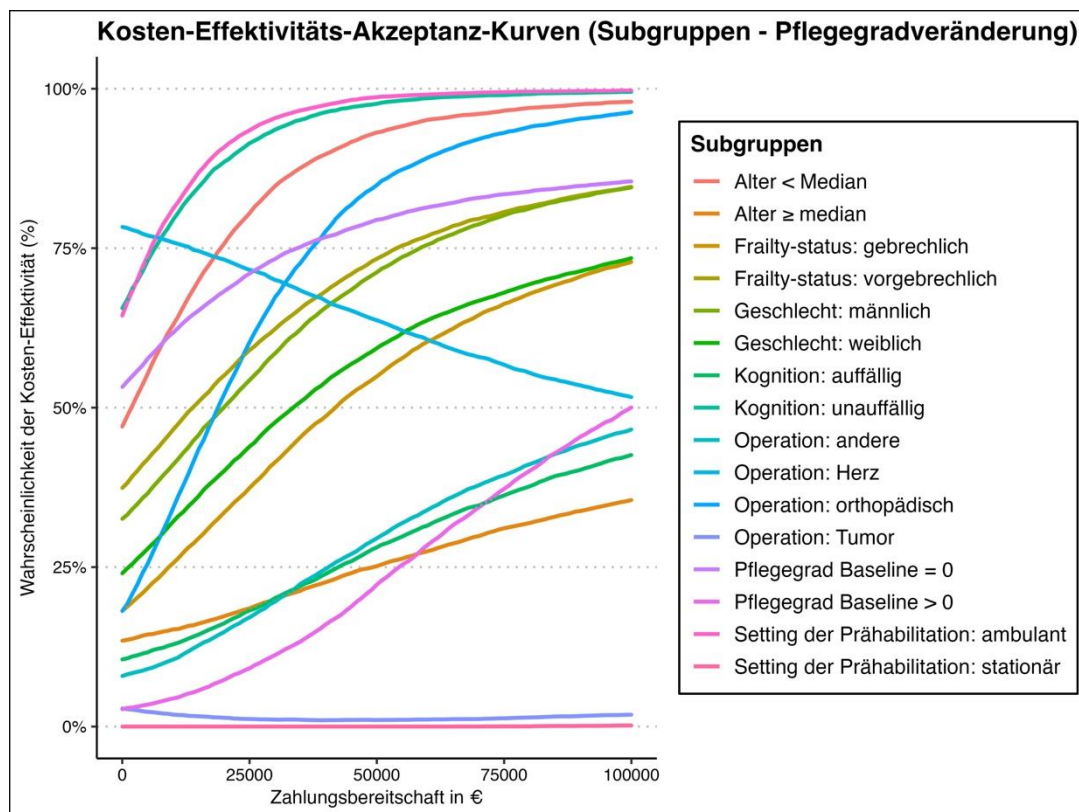


Abbildung 10: Kosten-Effektivität-Akzeptanz-Diagramm für Pflegegradverschlechterung aus gesellschaftlicher Perspektive – Subgruppenanalyse

Die Subgruppenanalysen zeigten eine klare Tendenz zu einer kosteneffektiven Intervention, sofern diese im ambulanten Setting stattfand. Um zu untersuchen, welche Patient:innengruppen innerhalb dieser Subgruppe (n = 439) besonders von der Intervention profitieren, bei gleichzeitig geringeren Kosten, wurde nachträglich explorativ in einer deskriptiven Analyse ausgewertet, welche Patient:innen-Charakteristika mit besseren Kosten-Effektivitäts-Ergebnissen assoziiert waren.

Direkte Kosten-Effektivität konnte hierfür nicht betrachtet werden, da diese sich immer aus dem Vergleich mit der Kontrollgruppe ergibt. Daher wurden die Patient:innen aus dieser Subgruppe stattdessen unterteilt in Patient:innen mit verhältnismäßig besseren (1) bzw. schlechteren (0) Ergebnissen. Um die Bedingungen für 1 (bessere Ergebnisse) zu erfüllen, durfte keine Pflegegradverschlechterung vorliegen und musste das QALY über dem Mittelwert der Subgruppe liegen und mussten die Kosten unter dem Mittelwert der Subgruppe liegen. Sobald eine der Bedingungen nicht erfüllt war, z.B. keine Pflegegradverschlechterung, QALY über dem Mittelwert, aber Kosten ebenfalls über dem Mittelwert, galt das Ergebnis dieser Person als verhältnismäßig schlechter (0).

Insgesamt hatten demnach n = 230 Patient:innen innerhalb der Subgruppe ambulantes Setting verhältnismäßig bessere Ergebnisse und n = 209 Patient:innen verhältnismäßig schlechtere Ergebnisse. Der Zusammenhang zwischen besseren Ergebnissen und Charakteristika wurde mit einem verallgemeinerten linearen Modell (Generalized Linear Model, GLM) gemessen und in Odds Ratios (ORs) mit 95%-Konfidenzintervallen angegeben. Die Ergebnisse finden sich in Tabelle 20.

Tabelle 20: Zusammenhang von Patient:innen-Charakteristika und Kosteneffektivität innerhalb der Subgruppe „Ambulantes Setting“

Charakteristikum	Odds Ratio	95% KI für OR	
Alter < Median	1,57	1,04	2,38
Frailty-Status: vorgebrechlich	1,95	1,25	3,05
Geschlecht: männlich	0,88	0,58	1,35
Kognition: unauffällig	2,03	1,32	3,14
Operation: orthopädisch	1,45	0,90	2,34
Operation: Herz	0,04	0,00	0,21
Operation: Tumor	0,86	0,37	1,97
Pflegegrad Baseline = 0	2,05	1,18	3,61

*Beinhaltet neurochirurgische Operationen an der Wirbelsäule.

Für vier der untersuchten Charakteristika zeigte sich eine signifikante OR für bessere Ergebnisse: „Alter < Median“, „Frailty-Status: vorgebrechlich“, „Kognition: unauffällig“ und „Pflegegrad Baseline = 0“. Die Odds waren in den letzten beiden Patient:innengruppen jeweils mehr als doppelt so hoch (OR > 2). Gleichzeitig war die OR für bessere Ergebnisse bei Patient:innen mit Herz-Operationen signifikant erniedrigt (OR 0,04, 95%-KI: 0,002; 0,21).

3.7.2 Ergebnisse der gesundheitsökonomischen Evaluation aus GKV- Perspektive

Population

In der Gesamtstudienpopulation waren etwa 25 % (n = 273) bei der BARMER Krankenkasse versichert und hatten der Datennutzung zugestimmt. Diese Patient:innen konnten aus der GKV-Perspektive berücksichtigt werden, davon n = 143 in der Interventionsgruppe und n = 130 in der Kontrollgruppe. Ihre Charakteristika werden in Tabelle 21 beschrieben.

Die GKV-Population ist bezüglich der Charakteristika größtenteils vergleichbar mit der Gesamtstudienpopulation (siehe Tabelle 3) und zwischen den Studienarmen ausgewogen. Größere Unterschiede zur Gesamtstudienpopulation bestehen in der Verteilung der Geschlechter, wobei in der GKV-Population ein höherer Anteil von weiblichen Patientinnen zu beobachten war (60 % gegenüber 55 % in der Gesamtstudienpopulation). Gleichzeitig war ein höherer Anteil der BARMER-Versicherten kognitiv auffällig gemäß des kategoriellen MOCA-Scores (62 % gegenüber 54 % in der Gesamtstudienpopulation). Während in der Gesamtstudienpopulation der Gebrechlichkeitsstatus zwischen den Studienarmen ausgewogen war, zeigte sich hier bei der GKV-Population ein Unterschied, mit 29 % gebrechlichen Patient:innen in der Interventionsgruppe und einem höheren Anteil von 42 % in der Kontrollgruppe (p=0.048). Die Überprüfung der SMD der Baseline-Charakteristika zwischen den Gruppen ergab keine Notwendigkeit für eine Adjustierung.

Tabelle 21: Patient:innencharakteristika der GKV-Population

Eigenschaften	Ausprägungen	Gesamt	Interventionsgruppe	Kontrollgruppe	P-Wert
n		273	143	130	
Alter (median[IQR])		78,00 [74,00; 82,00]	77,00 [74,00; 81,00]	79,00 [73,25; 83,00]	0,152
BMI (median[IQR])		26,50 [23,30; 30,00]	26,30 [23,05; 30,10]	26,55 [23,40; 29,87]	0,848
EQ5D.Score (median[IQR])		0,77 [0,43; 0,90]	0,78 [0,48; 0,90]	0,76 [0,37; 0,86]	0,246
CCI (median[IQR])		5,00 [4,00; 7,00]	5,00 [4,00; 7,00]	5,00 [4,00; 7,00]	0,261
WHODAS (median[IQR])		23,00 [17,00; 29,75]	22,50 [17,00; 29,00]	23,50 [18,00; 31,00]	0,340
Geschlecht (%)	männlich	110 (40,3)	53 (37,1)	57 (43,8)	0,309
	weiblich	163 (59,7)	90 (62,9)	73 (56,2)	
Pflegegrad (%)	0	216 (79,1)	116 (81,1)	100 (76,9)	0,728
	1	35 (12,8)	17 (11,9)	18 (13,8)	
	2	12 (4,4)	5 (3,5)	7 (5,4)	
	3	9 (3,3)	5 (3,5)	4 (3,1)	
	4	0 (0,0)	0 (0,0)	0 (0,0)	
	5	1 (0,4)	0 (0,0)	1 (0,8)	
MOCA (%)	n. auffällig	98 (35,9)	54 (37,8)	44 (33,8)	0,752
	auffällig	168 (61,5)	85 (59,4)	83 (63,8)	
	NA	7 (2,6)	4 (2,8)	3 (2,3)	
Fachabteilung (%)	Orthopädie	133 (48,7)	72 (50,3)	61 (46,9)	0,598
	Allgemeinchirurgie	17 (6,2)	10 (7,0)	7 (5,4)	
	Kardiologie	24 (8,8)	8 (5,6)	16 (12,3)	
	Neurochirurgie	14 (5,1)	6 (4,2)	8 (6,2)	

Eigenschaften	Ausprägungen	Gesamt	Interventionsgruppe	Kontrollgruppe	P-Wert
	Augenheilkunde	26 (9,5)	16 (11,2)	10 (7,7)	
	HNO	13 (4,8)	6 (4,2)	7 (5,4)	
	MKG	6 (2,2)	4 (2,8)	2 (1,5)	
	Urologie	20 (7,3)	10 (7,0)	10 (7,7)	
	Gynäkologie	9 (3,3)	6 (4,2)	3 (2,3)	
	Gefäßchirurgie	10 (3,7)	4 (2,8)	6 (4,6)	
	Kardiochirurgie	1 (0,4)	1 (0,7)	0 (0,0)	
Gebrechlichkeit (%)	Gebrechlich	96 (35,2)	42 (29,4)	54 (41,5)	0,048
	Vorgebrechlich	177 (64,8)	101 (70,6)	76 (58,5)	

IQR, interquartil range; BMI, body mass index; EQ5D, EuroQol 5 Dimensions 5 Level; CCI, Charlson Comorbidity Index.

Kosten

Die Zusammensetzung der Interventionskosten aus der GKV-Perspektive ist in Tabelle 22 dargestellt. Insgesamt wurde für 131 Patient:innen eine SDM-Konferenz durchgeführt. Die Teilnahme der Berufsgruppen war wie folgt verteilt: Personen aus der Berufsgruppe Anästhesie nahmen an 130 Konferenzen teil, Geriatrie an 126, Therapieberufe an 127, Chirurgie an 58, Pflege an 7, Allgemeinmedizin an 5. Zusätzlich nahmen die Patient:innen selbst an 97 Konferenzen teil und deren Angehörige an 26 Konferenzen. Nach einer erfolgten SDM-Konferenz haben 127 Patient:innen mind. eine Einheit an Prähabilitation erhalten. Die Kosten hierfür waren im ambulanten Setting am niedrigsten, gefolgt von ambulanter Prähabilitation mit Fahrtkosten, mobiler Prähabilitation, teilstationärer Prähabilitation und - mit deutlichem Abstand - der stationären Prähabilitation. Die mittleren Gesamt-Interventionskosten lagen bei 1.958 € pro Patient:in.

Tabelle 22: Interventionskosten GKV-Perspektive (in €)

Interventionsbestandteil	Interventionsgruppe (n = 143)				
	MW	SD	Median	1. Quartil	3. Quartil
Frailty-Screening	26	0	26	26	26
SDM-Konferenz*	302	107	287	287	387
Prähabilitation*	2.079	2.808	1.195	644	1.749
mobile Prähabilitation** (n = 1)	1.020	n.a.	1.020	1.020	1.020
ambulante Prähabilitation** (n = 16)	671	158	719	686	721
ambulante Prähabilitation mit Fahrtkosten** (n = 83)	666	140	699	623	719
teilstationäre Prähabilitation** (n = 4)	1.224	452	1.399	1.137	1.486
stationäre Prähabilitation** (n = 19)	8.382	463	8.611	8.146	8.611
Gesamtkosten	1.958	2.693	1.032	936	1.143

* Enthält auch Werte i.H.v. 0 € für Teilnehmende ohne SDM-Konferenz bzw. Prähabilitation.

** Enthält nur die Kosten der tatsächlich beanspruchten Leistung. Wenn nicht beansprucht für die durchschnittlichen Gesamtkosten mit 0 € bewertet.

MW, Mittelwert; SD, Standardabweichung.

Die winsorisierten Nachverfolgungskosten inklusive der oben gezeigten Interventionskosten und den daraus resultierenden Gesamtkosten sind in Tabelle 23 dargestellt. Eine entsprechende Übersicht ohne Winsorisierung findet sich im Anhang in Tabelle A3. Die Kosten für den Krankenhaus-Indexaufenthalt lagen bei 10.637 € (SD 9.664). Im 12-Monats Follow-Up Zeitraum entstanden in der Interventionsgruppe durchschnittliche Kosten in Höhe von 16.998 € (SD 14.886) mit Median 11.922 € (IQR 18.286). Die größten Treiber dieser Kosten lagen in den stationären Behandlungen mit 5.904 € (SD 9.314), gefolgt von verordneten Arzneimitteln mit 3.039 € (SD 4.542) und der Pflege mit 2.600 € (SD 4.392). Da, wie für Kostendaten üblich, eine stark rechtsschiefe Verteilung vorliegt, zeigen sich bei Betrachtung der Median-Werte, dass vor allem Kosten im ambulanten Sektor und in verordneten Arzneimitteln entstehen, wohingegen die hohen Durchschnittskosten von stationären und Pflegebehandlungen in erster Linie durch einzelne, sehr hohe Extremwerte zustande kamen. Addiert man auf die Kosten für den Index-Krankenhausaufenthalt und die Nachverfolgungskosten die Interventionskosten i.H.v. 1.958 € (SD 2.693), entstanden einem/-r Patient:in der Interventionsgruppe im Studienzeitraum durchschnittliche Kosten von 29.593 € (SD 20.781).

Tabelle 23: Gesamtkosten GKV-Perspektive nach Winsorisierung und Imputation (in €)

Kostenblock	Interventionsgruppe (n = 143)				Kontrollgruppe (n = 130)			
	MW	SD	Median	IQR	MW	SD	Median	IQR
Interventionskosten	1.958	2.693	1.032	207	0	0	0	0
KH-Indexaufenthalt*	10.637	9.664	7.798	6.918	10.062	9.583	7.624	5.563
Follow-up Kosten Total	16.998	14.886	11.922	18.286	17.123	14.772	12.170	18.514
Arztbesuche	1.518	1.120	1.253	1.176	1.356	1.009	1.092	920
Heilmittel	869	1.231	404	1.047	912	1.476	276	1.106
Hilfsmittel	611	921	260	597	670	1.054	246	673
Häusliche Krankenpflege	124	515	0	0	161	577	0	0
Stationäre KH-Aufenthalte	5.904	9.314	0	8.934	6.494	9.331	2.232	9.534
Ambulante KH-Aufenthalte	245	316	151	323	164	309	0	190
Krankentransporte	641	1.143	111	705	732	1.281	96	749
Arzneimittel	3.039	4.542	1.624	2.657	2.514	3.739	1.415	2.003
Pflege	2.600	4.392	0	4.137	2.621	4.404	0	4.245
Ambulante Reha	2	11	0	0	4	17	0	0
Stationäre Reha	1.446	1.727	0	2.763	1.494	1.699	798	2.787
Gesamtkosten	29.593	20.781	23.383	28.282	27.184	17.594	22.967	23.614

*enthält auch Werte i.H.v. 0 € für Teilnehmende ohne Operation.

IG, Interventionsgruppe; IQR, Interquartilabstand; KG, Kontrollgruppe; KH, Krankenhaus; KI, Konfidenzintervall; MW, Mittelwert; SD, Standardabweichung.

Eine ähnliche Tendenz lässt sich auch in der Kontrollgruppe beobachten. Die gesamten Nachverfolgungskosten betrugen hier 17.123 € (SD 14.772), bzw. im Median 12.170 € (IQR 18.514), welche, neben den Kosten für den KH-Indexaufenthalt, vornehmlich durch Kosten in den Bereichen ambulante und stationäre Versorgung, sowie in verordneten Arzneimitteln und der Pflege entstanden. Da die Kontrollgruppe keine Prähabilitationsmaßnahmen erhielt, fielen hier auch keine weiteren Interventionskosten an. Somit entstanden der Interventionsgruppe im Mittel rund 2.000 € höhere Kosten als der Kontrollgruppe, was in etwa den zusätzlichen Kosten der Intervention entspricht.

Kosten-Effektivitäts-Analyse der Pflegegradverschlechterung

Tabelle 24 zeigt die Veränderung des Pflegegrads für die GKV-Population. In der Kontrollgruppe wurde bei 86 Patient:innen (66 %) keine Veränderung im Pflegegrad beobachtet, 13 Patient:innen (10 %) erfuhren eine Verbesserung, wohingegen 31 Patient:innen (24 %) eine Pflegegradverschlechterung (PGV) hatten. In der Interventionsgruppe blieben hingegen 96 Patient:innen (67 %) unverändert, 16 (11 %) verbesserten sich, und 31 Patient:innen (22 %) verschlechterten sich.

Tabelle 24: Veränderung des Pflegegrads in der GKV-Population

Veränderung des Pflegegrades	≥ 2 Punkte	1 Punkt	keine	- 1 Punkt	≤ -2 Punkte
	Verschlechterung		unverändert	Verbesserung	
Kontrolle (n = 130)	19 (14,6 %)	12 (9,2 %)	86 (66,2 %)	12 (9,2 %)	1 (0,8 %)
Intervention (n = 143)	14 (9,8 %)	17 (11,9 %)	96 (67,1 %)	13 (9,1 %)	3 (2,1 %)

Die Ergebnisse aller gesundheitsökonomischen Analysen mit vermiedenen PGV als Effektmaß sind in Tabelle 25 dargestellt. Die ITT-Analyse zeigt inkrementelle Kosten von 2.409 € (95 % KI: 2.092; 7.067) in der Interventionsgruppe, bei durchschnittlich 2,0 %-Punkte geringerer PGV (95 % KI: -0,08; 0,12). Während in der Interventionsgruppe 22 % (n = 31) der Patient:innen eine Verschlechterung im Pflegegrad erfuhren, waren es in der Kontrollgruppe 24 % (n = 31). Die daraus resultierende IKER beträgt 111.115 € pro vermiedene Pflegegradverschlechterung. Da im Gegensatz zur QALY-Analyse im Pflegegrad alle Daten verfügbar waren, wurde hier keine weitere Complete-Case Analyse durchgeführt. Für das Kosten-Effektivitäts-Diagramm der Ergebnisse der gebootstrappten PG ITT Analyse siehe Abbildung 11. Die Ergebnisse aller weiteren Analysen finden sich in der vereinfachten Darstellung in Abbildung 12.

Eine höhere Effektivität bzgl. der Pflegegradverschlechterung, bei geringeren inkrementellen Kosten, wurde bei Patient:innen beobachtet die mindestens 15 Prähabilitations-Einheiten erhielten. In der PP15-Analyse werden inkrementelle Kosten von 1.244 € (95 % KI: -4.067; 6.524) und ein höherer Anteil von vermiedenen Pflegegradverschlechterungen von 4 %-Punkten (95 % KI: -0,07; 0,16) beobachtet, was zu einer IKER von 28.620 € pro vermiedener PGV führt. In der Interventionsgruppe kamen PGV in 18 % der Fälle vor, während es in der Kontrollgruppe 23 % waren.

Das beste Ergebnis bzgl. der vermiedenen PGV zeigte sich bei Patient:innen, die sowohl mindestens 15 Prähabilitations-Einheiten erhielten, als auch tatsächlich operiert wurden. Die PP15OP-Analyse verzeichnet die innerhalb der PG Analysen geringsten inkrementellen Kosten von 672 € (95 % KI: -4.945; 6.586) bei dem verhältnismäßig höchsten inkrementellen Effekt

von 7 %-Punkten (95 % KI: -0,06; 0,2) weniger Verschlechterungen im Pflegegrad (entspricht etwa 10 Patient:innen), woraus sich eine IKER von 9.943 € pro vermiedener PGV berechnet. In dieser Population erfuhren 16 % der Patient:innen in der Interventionsgruppe und 23 % der Patient:innen in der Kontrollgruppe eine Verschlechterung im Pflegegrad.

Abschließend zeigt die Sensitivitätsanalyse (ITT SA), in der die Follow-Up Kosten nicht winsorisiert wurden, hinsichtlich Kosten ein sehr ähnliches Ergebnis zur ITT-Analyse, wobei die Streuung extremer ausfällt. Die inkrementellen Kosten sind hierbei marginal geringer mit 2.185 € (95% KI: -6.195; 10.775).

Tabelle 25: Ergebnisse der Kosten-Effektivitäts-Analysen aus GKV-Perspektive (Pflegegradverschlechterung)

Analyse	N		Kosten IG		Kosten KG		Inkrementelle Kosten		PGV IG		PGV KG		Vermiedene PGV		IKER €/verm. PGV
	IG	KG	MW	SD	MW	SD	MWD	95% KI	N	%	N	%	MWD	95% KI	
ITT	143	130	29.593	20.781	27.184	17.594	2.409	-2.092; 7.067	31	0,22	31	0,24	0,02	-0,08; 0,12	111.115
PP15	92	92	27.773	19.408	26.529	17.271	1.244	-4.067; 6.524	17	0,18	21	0,23	0,04	-0,07; 0,16	28.620
PP15OP	74	74	28.777	19.113	28.106	17.106	672	-4.945; 6.586	12	0,16	17	0,23	0,07	-0,06; 0,2	9.943
ITT SA	143	130	34.052	37.892	31.868	33.719	2.185	-6.195; 10.775	31	0,22	31	0,24	0,02	-0,08; 0,12	100.773

IG, Interventionsgruppe; IKER, Inkrementelle Kosten-Effektivitäts-Relation; KG, Kontrollgruppe; KI, Konfidenzintervall; MW, Mittelwert; MWD, Mittelwertdifferenz; PGV, Pflegegradverschlechterung; SD, Standardabweichung.



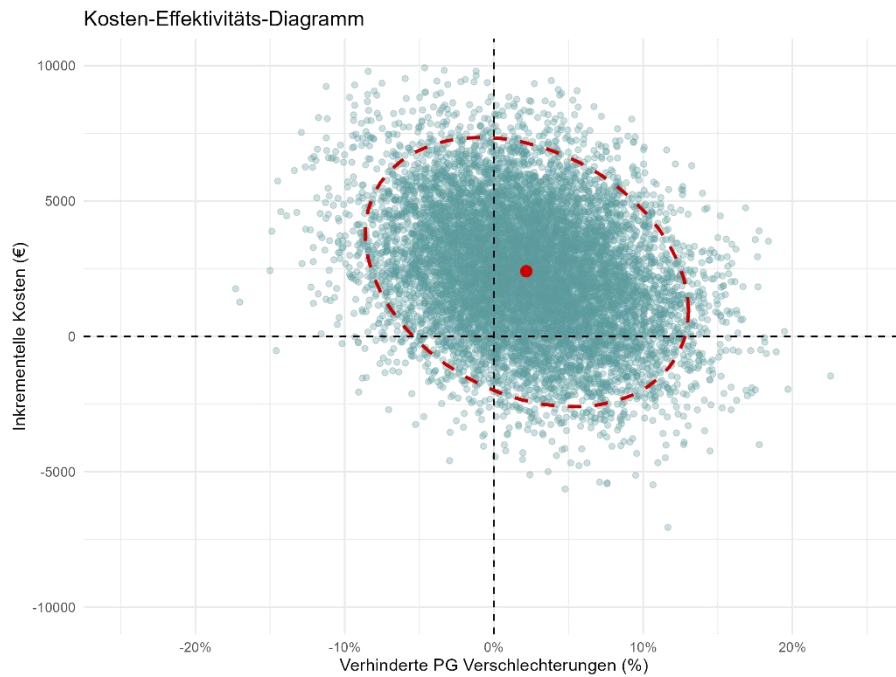


Abbildung 11: Streudiagramm der ITT-Analyse der Pflegegradverschlechterung aus GKV-Perspektive

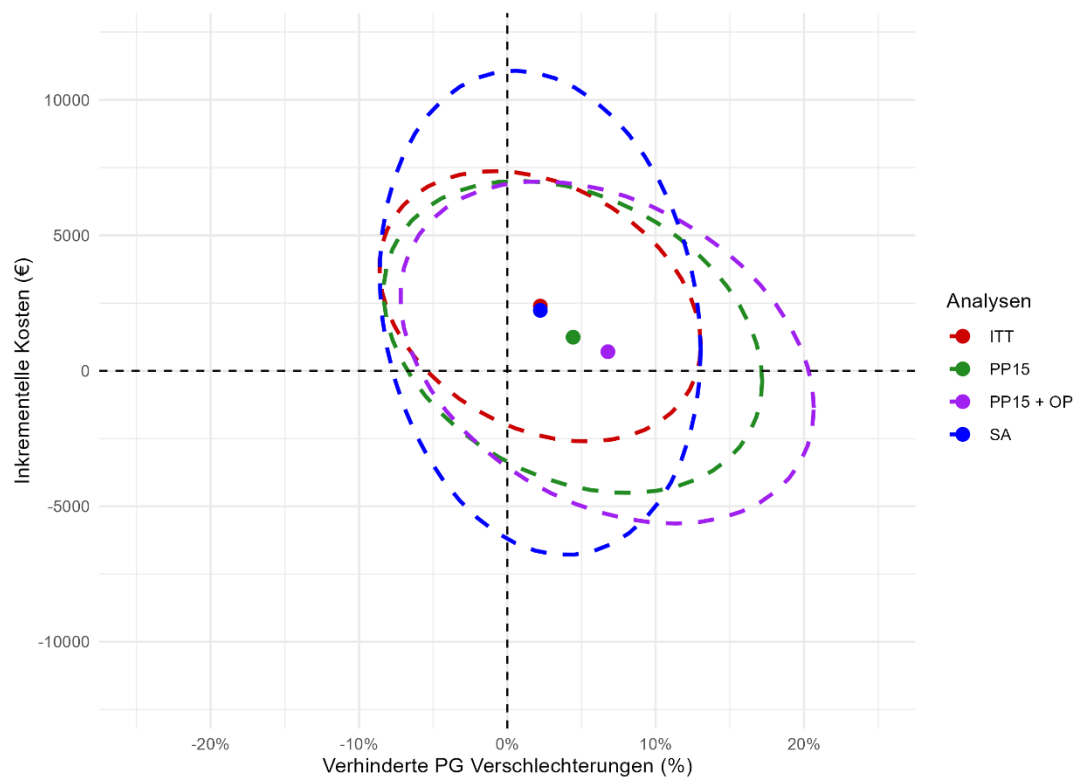


Abbildung 12: Streudiagramm aller Analysen der Pflegegradverschlechterung aus GKV-Perspektive

Kosten-Effektivitäts-Analyse des QALY

Tabelle 26 zeigt die Ergebnisse der gesundheitsökonomischen Analysen mit QALY als Effektmaß. Generell zeigt sich bzgl. der CEAs mit QALY als Effektmaß eine eher schwache Tendenz zur Kosteneffektivität, welche sich vorwiegend durch zusätzliche Kosten der Intervention bei gleichzeitig sehr geringen oder keinen zusätzlichen Zugewinnen an QALYs begründet. Auch wenn einzelne IKERs auf eine kosteneffektive Intervention (also eine Tendenz zu positiven inkrementellen Effekten bei nur geringen oder sogar negativen inkrementellen Kosten) schließen lassen, zeigen die Diagramme mit den Bootstrapping-basierten Punktschätzer durchaus große Unsicherheit in der Bewertung der Ergebnisse auf. Dabei weisen, wie folgt beschrieben, die einzelnen Analysearten leichte Unterschiede auf.

Die ITT QALY (siehe Abbildung 13) Analyse zeigt, dass Patient:innen der Interventionsgruppe im Schnitt während des post-operativen Nachverfolgungszeitraums 0.74 (SD 0.25) QALYs bei entstandenen Kosten i.H.v. 29.593 € (SD 20.781) erreichten. Verglichen mit der Kontrollgruppe, mit 0.73 (SD 0.27) QALYs und Kosten i.H.v. 27.184€ (SD 17.594), bedeutet das einen inkrementellen Zuwachs von lediglich 0.01 QALYs (95% KI: -0.05; 0.07) bei zusätzlichen Kosten von 2.457 € (95% KI: -2.092; 7.067). Daraus ergibt sich ein inkrementelles Kosten-Effektivitäts-Verhältnis von 232.588 € pro gewonnenem QALY.

Tabelle 26: Ergebnisse der gesundheitsökonomischen Evaluation aus GKV-Perspektive (QALY)

Analyse	N		Kosten IG		Kosten KG		Inkrementelle Kosten		QALY IG		QALY KG		Gewonnene QALYs		IKER
	IG	KG	MW	SD	MW	SD	MWD	95% KI	MW	SD	MW	SD	MWD	95% KI	€/QALY
ITT	143	130	29.593	20.781	27.184	17.594	2.409	-2.092; 7.067	0,74	0,25	0,73	0,27	0,01	-0,05; 0,07	232.588
CC	43	40	28.057	17.610	28.858	18.605	-801	-8.645; 6.773	0,69	0,36	0,62	0,39	0,07	-0,09; 0,23	(dominant) -11.426
PP15	92	92	27.619	19.214	26.524	16.901	1.095	-4.003; 6.426	0,76	0,25	0,72	0,30	0,04	-0,04; 0,12	25.252
PP15OP	67	67	27.206	17.234	27.124	16.740	82	-5.670; 5.822	0,80	0,21	0,76	0,27	0,05	-0,03; 0,13	1.794
ITT SA	143	130	34.052	37.892	31.868	33.719	2.185	-6.195; 10.775	0,74	0,25	0,73	0,27	0,01	-0,05; 0,07	210.940
IG, Interventionsgruppe; IKER, Inkrementelle Kosten-Effektivitäts-Relation; KG, Kontrollgruppe; KI, Konfidenzintervall; MW, Mittelwert; MWD, Mittelwertdifferenz; QALY, Qualitätsadjustiertes Lebensjahr; SD, Standardabweichung.															



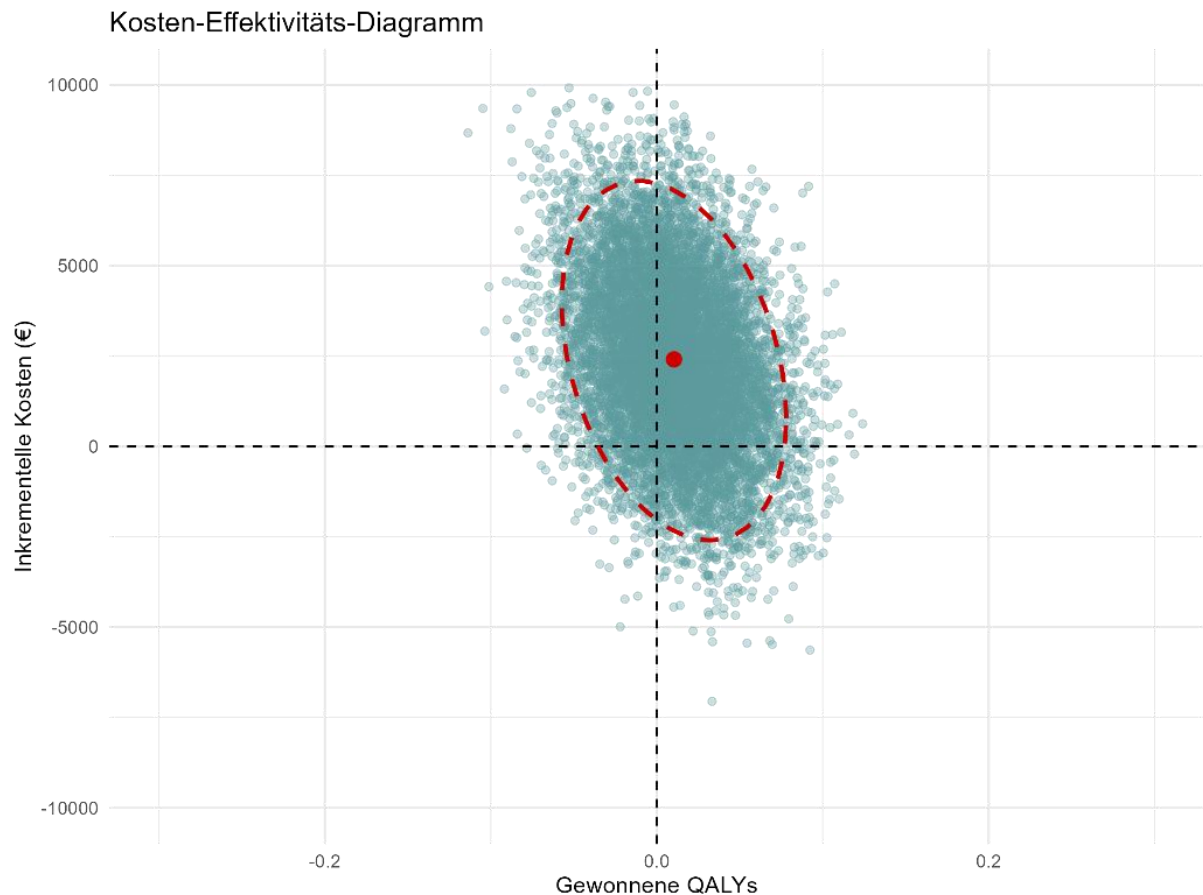


Abbildung 13: Streudiagramm der ITT-Analyse des QALYs aus GKV-Perspektive

Die Ergebnisse weiterer Analysen sind in Abbildung 14 dargestellt. Ein positiveres Bild zeigt sich in der CC Analyse: Betrachtet man nur jene Patient:innen, die alle EQ-5D-5L Befragungen durchlaufen haben, zeigen die Ergebnisse, trotz der zusätzlichen Interventionskosten, negative inkrementelle, also geringere, Kosten in der Interventionsgruppe von -801 € (95 % KI: -8.645; 6.773) auf, bei einem, verglichen mit der ITT QALY Analyse, höheren Zuwachs an QALYs von 0,07 (95 % KI: -0,09; 0,23) in der Interventionsgruppe, was zu einer negativen IKER im dominanten Quadranten von -11.426 € führt („besser bei geringeren Kosten“).

In der QALY PP15-Analyse, bei der nur Patient:innen berücksichtigt wurden, die mindestens 15 Prähabilitationseinheiten erhielten, liegen die inkrementellen Kosten bei 1.095 € (95 % KI: -4.003; 6.426) und der QALY-Gewinn beträgt 0,04 (95 % KI: -0,04; 0,12), was eine IKER von 25.252 € ergibt.

Patient:innen, die mindestens 15 Prähabilitationseinheiten erhielten und auch operiert wurden (PP15OP-Analyse) wiesen mit lediglich 82 € die niedrigsten inkrementellen Kosten auf (95 % KI: -5.670; 5.822), während es zu gewonnenen QALYs von 0,05 kam (95 % KI: -0,03; 0,13)), wodurch sich eine IKER von 1.794 € pro gewonnenen QALY ergibt.

Die Sensitivitätsanalyse ohne winsorisierte Kostendaten, zeigt bzgl. der Effektivität sehr ähnliche Ergebnisse zur ITT Analyse auf, mit einer höheren Streuung der Kosten (95% KI: - 6.195; 10.775. Die IKER beträgt 210.940 € pro zusätzlichen QALY.

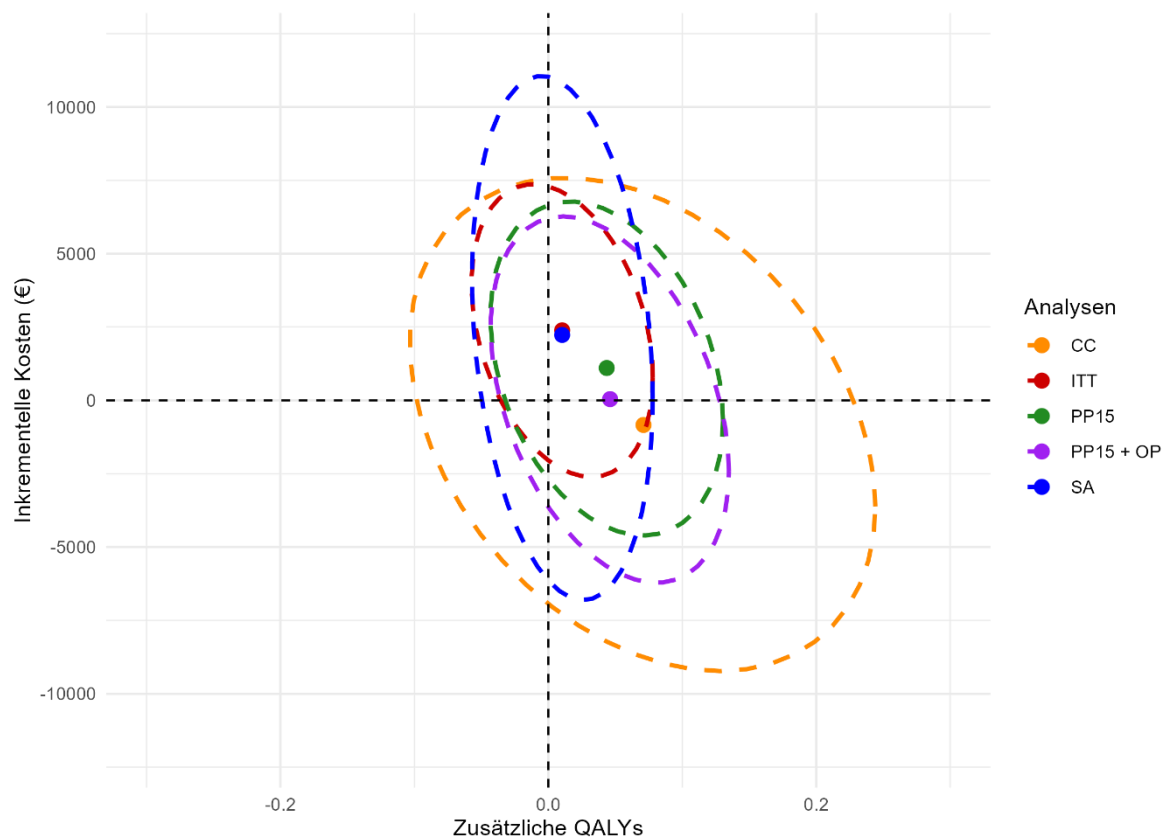


Abbildung 14: Streudiagramm aller Analysen des QALYs aus GKV-Perspektive

Wahrscheinlichkeit der Kosteneffektivität

Um die Wahrscheinlichkeit der Kosteneffektivität der Intervention beurteilen zu können, wurden die oben gezeigten Ergebnisse in ein Kosteneffektivitäts-Akzeptanz-Diagramm eingetragen, welches die Wahrscheinlichkeit der einzelnen Analysen unter Annahme von verschiedenen hohen, hypothetischen Zahlungsbereitschaften darstellt. Die Ergebnisse sind in Abbildung 15 dargestellt.

Unter den QALY-Analysen zeigte die Complete Case-Analyse die höchste Wahrscheinlichkeit von Kosteneffektivität, welche auch bei einer Zahlungsbereitschaft von 0 € pro zusätzlichem QALY bei etwa 60 % liegt und ab einer Zahlungsbereitschaft von 50.000 € bei 75 % liegt. Das erklärt sich durch den verhältnismäßig hohen Anteil von Bootstrapping-Punktschätzern im dominanten Bereich (besser und billiger), welcher immer als kosteneffektiv zu bewerten ist, unabhängig von der Zahlungsbereitschaft (vgl. Abbildung 13: Streudiagramm der ITT-Analyse des QALYs aus GKV-Perspektive) Dabei ist allerdings zu beachten, dass diese Analyse nur eine deutlich kleinere Stichprobengröße berücksichtigt als die restlichen Analysen. Darauf folgt die PP15OP-Analyse, bei der sich eine Wahrscheinlichkeit von 50% selbst bei einer Zahlungsbereitschaft von 0 € ergibt und die Wahrscheinlichkeit bei einer Zahlungsbereitschaft von 75.000 € über 75% ansteigt. In der PP15-Analyse übersteigt die Wahrscheinlichkeit für Kosteneffektivität die 50%-Marke etwa bei einer Zahlungsbereitschaft von 25.000 €. Deutlich geringer fällt die Wahrscheinlichkeit für Kosteneffektivität bei den verbleibenden beiden QALY-Analysen aus. Hier wird selbst bei einer Zahlungsbereitschaft von 100.000 € die 50%-Marke nicht erreicht.

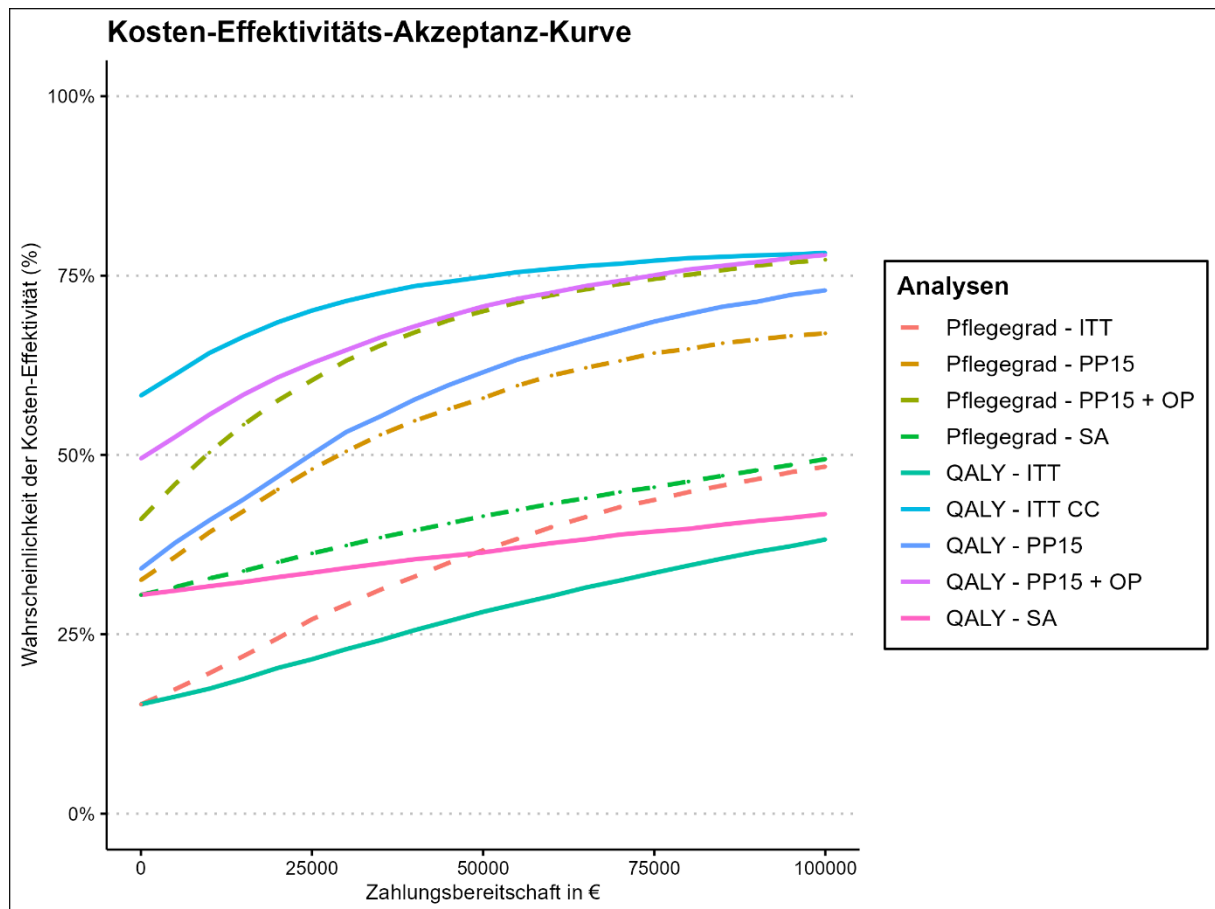


Abbildung 15: Kosten-Effektivitäts-Akzeptanz-Diagramm für QALY und Pflegegradverschlechterung aus GKV-Perspektive

Betrachtet man die Ergebnisse der Pflegegrad-Analysen zeigt sich ein ähnliches Bild. Hier ist die Wahrscheinlichkeit der Kosteneffektivität höher, wenn Patient:innen protokollgetreu mindestens 15 Prähabilitations-Einheiten erhielten und auch operiert wurden. Hier zeigt sich der steilste Anstieg der Wahrscheinlichkeit bei höheren Zahlungsbereitschaften. Bei einer Zahlungsbereitschaft von 50.000 € liegt die Wahrscheinlichkeit der Kosteneffektivität in der PP15-Gruppe bei etwa 60 % und in der PP15OP-Gruppe bei 70 %. Im Gegensatz dazu liegt die Wahrscheinlichkeit in der ITT-Analyse selbst bei einer Zahlungsbereitschaft von 100.000 € pro vermiedener PGV unter 50 %.

Sowohl bei den QALY als auch bei der Pflegegradverschlechterung zeigen die Sensitivitätsanalysen ein mit den ITT-Analysen vergleichbares Bild, wenn auch mit einer geringfügig erhöhten Wahrscheinlichkeit der Kosteneffektivität zu rechnen ist.

Subgruppenanalysen

Die Ergebnisse der Subgruppenanalysen sind in Tabelle 27 für das QALY und Tabelle 28 für die Pflegegradverschlechterung dargestellt. Dabei wurde eine deutlich höhere Tendenz zu einer dominanten, kosteneffektiven Intervention unter Patient:innen im ambulanten Prähabilitations-Setting beobachtet. Hier zeigte sich die Intervention als deutlich effektiver verglichen mit der Kontrollgruppe, bei gleichzeitig geringeren Kosten. Patient:innen in der Interventionsgruppe konnten dabei im Schnitt 0.06 QALYs (95 % KI: -0.01; 0.13) mehr gewinnen und 5 %-Punkte (95 % KI: -0.06; 0.16) geringere Verschlechterungen im Pflegegrad

erzielen als die Kontrollgruppe, wobei um -3.690 € geringere Kosten entstanden (95 % KI: -8.531; 1.157). Das führt bei beiden Analysen zu einer negativen ICER im dominanten Bereich. Weiter zeigt sich die Intervention insbesondere unter jüngeren (Alter<median), männlichen, orthopädisch und anders operierten Patient:innen als kosteneffektiv, während sie nach Herz und Tumor Operationen, sowie bei vor-gebrechlichen, älteren und weiblichen Patient:innen als nicht-kosteneffektiv zu bewerten ist.

Diese Ergebnisse spiegeln sich auch in den Kosten-Effektivitäts-Akzeptanz-Kurven aus Abbildung 16 und Abbildung 17 wider. Erfolgt die Prähabilitation im ambulanten Sektor, ist mit einer sehr hohen Wahrscheinlichkeit der Kosteneffektivität zu rechnen, auch schon bei einer Zahlungsbereitschaft von 0 €. Im Gegensatz dazu liegt die Wahrscheinlichkeit der Kosteneffektivität bei einem stationären Setting unter allen möglichen Formen der Zahlungsbereitschaft nur knapp über 0 %. Weiter zeigen insbesondere jüngere Patient:innen und männliche Patienten bei höheren Zahlungsbereitschaften einen schnellen Anstieg in der Wahrscheinlichkeit der Kosteneffektivität.

Tabelle 27: Subgruppenanalyse ITT Qualitätsadjustiertes Lebensjahr aus GKV-Perspektive

QALY	N		Kosten IG		Kosten KG		Inkrementelle Kosten			QALY IG		QALY KG		Gewonnene QALYs			IKER
	IG	KG	MW	SD	MW	SD	MWD	95% KI		MW	SD	MW	SD	MWD	95% KI		
Alter < Median	73	55	27.827	20.584	25.529	18.201	2.298	-4.531	8.866	0,78	0,25	0,72	0,33	0,06	-0,04	0,16	40.149
Alter ≥ Median	70	75	31.435	20.973	28.398	17.156	3.037	-3.046	9.382	0,71	0,24	0,74	0,22	-0,03	-0,11	0,04	wird dominiert
Weiblich	90	73	27.746	18.057	25.791	16.233	1.955	-3.368	7.224	0,73	0,25	0,76	0,23	-0,03	-0,10	0,04	wird dominiert
Männlich	53	57	32.730	24.602	28.969	19.196	3.762	-4.434	12.008	0,76	0,26	0,69	0,31	0,07	-0,04	0,18	54.456
Gebrechlich	42	54	36.223	25.917	30.892	17.378	5.330	-3.457	14.571	0,57	0,30	0,58	0,33	0,00	-0,13	0,13	wird dominiert
Vor-gebrechlich	101	76	26.836	17.654	24.550	17.380	2.286	-2.927	7.396	0,81	0,19	0,84	0,14	-0,03	-0,08	0,02	wird dominiert
Orthop. OP*	72	61	27.113	19.316	23.977	16.041	3.136	-2.976	9.228	0,79	0,22	0,76	0,24	0,03	-0,05	0,11	99.407
Tumor-OP	18	6	39.458	28.506	18.633	12.917	20.825	4.370	36.692	0,61	0,32	0,91	0,06	-0,31	-0,46	-0,16	wird dominiert
Herz-OP	9	16	49.203	22.926	49.332	17.888	-129	-17.640	16.346	0,72	0,29	0,77	0,23	-0,05	-0,29	0,15	2.471
Andere OP	44	47	25.605	15.594	24.900	14.521	705	-5.380	6.802	0,72	0,24	0,66	0,32	0,06	-0,05	0,18	11.321
Ambulantes Setting**	92	92	23.218	15.781	26.908	17.514	-3.690	-8.531	1.157	0,81	0,21	0,75	0,28	0,06	-0,01	0,13	dominant
Stationäres Setting**	15	15	47.400	21.939	27.815	20.271	19.585	4.755	34.115	0,58	0,31	0,62	0,35	-0,03	-0,26	0,21	wird dominiert

*Beinhaltet neurochirurgische Operationen an der Wirbelsäule **Propensity Score gematched

IG, Interventionsgruppe; IKER, Inkrementelle Kosten-Effektivitäts-Relation; KG, Kontrollgruppe; KI, Konfidenzintervall; MW, Mittelwert; MWD, Mittelwertdifferenz; SD, Standardabweichung; QALY, Qualitätsadjustiertes Lebensjahr.

Tabelle 28: Subgruppenanalyse ITT Pflegegradverschlechterung aus GKV-Perspektive

Pflegegrad- veränderung	N		Kosten IG		Kosten KG		Inkrementelle Kosten			PGV IG	PGV KG	Vermiedene PGV			IKER
	IG	KG	MW	SD	MW	SD	MWD	95% KI		%	%	%-Diff	95% KI		€/verm. PGV
Alter < Median	73	55	27.827	20.584	25.529	18.201	2.298	-4.531	8.866	0,16	0,25	0,09	-0,05	0,23	25.487
Alter ≥ Median	70	75	31.435	20.973	28.398	17.156	3.037	-3.046	9.382	0,27	0,23	-0,04	-0,18	0,10	Wird dominiert
Weiblich	90	73	27.746	18.057	25.791	16.233	1.955	-3.368	7.224	0,22	0,21	-0,02	-0,15	0,11	Wird dominiert
Männlich	53	57	32.730	24.602	28.969	19.196	3.762	-4.434	12.008	0,21	0,28	0,07	-0,08	0,23	51.421
Gebrechlich	42	54	36.223	25.917	30.892	17.378	5.330	-3.457	14.571	0,40	0,35	-0,05	-0,25	0,15	Wird dominiert
Vor- gebrechlich	10 1	76	26.836	17.654	24.550	17.380	2.286	-2.927	7.396	0,14	0,16	0,02	-0,08	0,13	118.587
Orthop. OP*	72	61	27.113	19.316	23.977	16.041	3.136	-2.976	9.228	0,21	0,23	0,02	-0,12	0,16	148.119
Tumor-OP	18	6	39.458	28.506	18.633	12.917	20.825	4.370	36.692	0,22	0,00	-0,22	-0,43	-0,05	Wird dominiert
Herz-OP	9	16	49.203	22.926	49.332	17.888	-129	-17.640	16.346	0,33	0,25	-0,08	-0,49	0,30	1.546
Andere OP	44	47	25.605	15.594	24.900	14.521	705	-5.380	6.802	0,21	0,28	0,07	-0,11	0,25	9.789
Ambulantes Setting**	92	92	23.218	15.781	26.908	17.514	-3.690	-8.531	1.157	0,15	0,21	0,05	-0,06	0,16	dominant
Stationäres Setting**	15	15	47.400	21.939	27.815	20.271	19.585	4.755	34.115	0,20	0,33	0,13	-0,18	0,45	146.888

*Beinhaltet neurochirurgische Operationen an der Wirbelsäule **Propensity Score gematched

IG, Interventionsgruppe; IKER, Inkrementelle Kosten-Effektivitäts-Relation; KG, Kontrollgruppe; KI, Konfidenzintervall; MW, Mittelwert; MWD, Mittelwertdifferenz; PGV, Pflegegradverschlechterung SD, Standardabweichung.

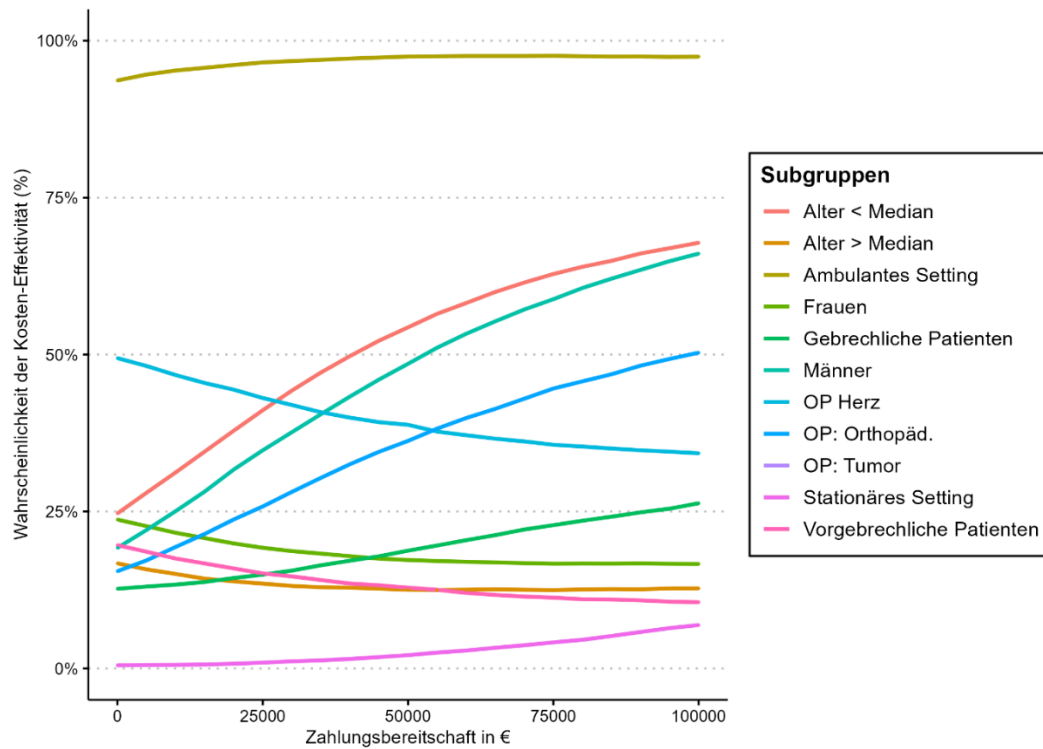


Abbildung 16: Kosten-Effektivität-Akzeptanz-Diagramm für QALY aus GKV-Perspektive

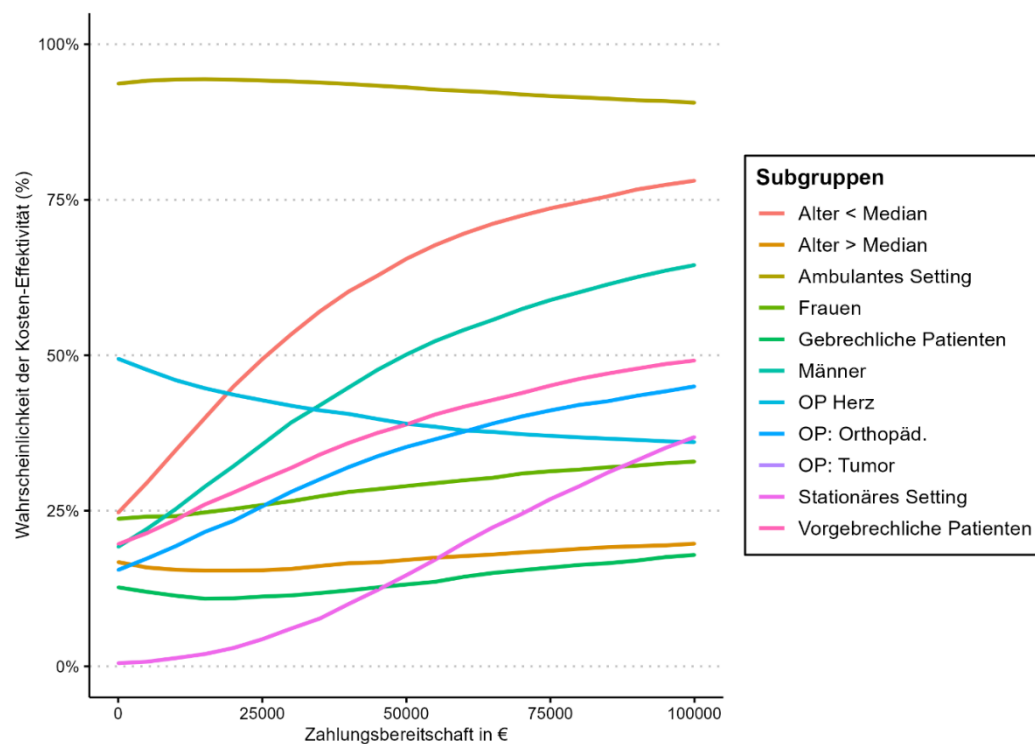


Abbildung 17: Kosten-Effektivität-Akzeptanz-Diagramm für Pflegegradverschlechterung aus GKV-Perspektive

3.7.3 Ergebnisse der gesundheitsökonomischen Evaluation aus Leistungserbringer-Perspektive

Die Ergebnisse der gesundheitsökonomischen Evaluation sind in Tabelle 29 dargestellt. Es wurden insgesamt 1.067 Patient:innen operiert, davon 552 in der Interventionsgruppe und 515 in der Kontrollgruppe. Die mittlere Differenz der Entgelte betrug -540 €. Die Verweildauer war in der Interventionsgruppe 0,75 Tage länger. Es wurde kein relevanter Unterschied in der mittleren ITS-Verweildauer oder dem Anteil an ITS-Patient:innen beobachtet. In der Interventionsgruppe war die Rate an Komplikationen niedriger als in der Kontrollgruppe (-2,86 %).

Tabelle 29: Gesundheitsökonomische Evaluation aus Leistungserbringer-Perspektive

Endpunkt	IG (n = 552)		KG (n = 515)		Differenz
	MW	SD	MW	SD	MWD
Entgelt (€)	11.159	8.173	11.699	8.776	-540
Verweildauer	7,73	7,75	6,99	5,99	0,75
ITS-Verweildauer	0,56	2,56	0,60	2,34	-0,03
Endpunkt	n	%	n	%	%-Diff
ITS-Patient:innen	102	18,48	99	19,22	-0,75
Komplikationen	160	28,99	164	31,84	-2,86
Clavien-Dindo I	44	7,97	51	9,90	-1,93
Clavien-Dindo II	48	8,70	53	10,29	-1,60
Clavien-Dindo IIIa	13	2,36	12	2,33	0,02
Clavien-Dindo IIIb	8	1,45	7	1,36	0,09
Clavien-Dindo IVa	41	7,43	34	6,60	0,83
Clavien-Dindo IVa	2	0,36	2	0,39	-0,03
Clavien-Dindo V	4	0,72	5	0,97	-0,25

IG, Interventionsgruppe; ITS, Intensivstation; KG, Kontrollgruppe; MW, Mittelwert; MWD, Mittelwertdifferenz; SD, Standardabweichung.

4 Diskussion der Projektergebnisse

Die hier vorgestellte nationale multizentrische, assessor-verblindete, randomisierte, pragmatische, zweiarmige, kontrollierte klinische Studie hat für den Primären Endpunkt (Veränderung des Pflegegrades zwischen V18 und V2) kein signifikantes Ergebnis gezeigt, obwohl die Power für den angenommenen Unterschied hätte ausreichen sollen.

Die Analyse des Primären Endpunktes erfolgte stratifiziert nach dem Pflegegrad der Baseline Untersuchung. Während sich in den Patienten mit Baselinegrad 0 oder 1 keine sehr unterschiedlichen Verschlechterungsraten zeigen, so ist das in der Gruppe mit einem höheren Pflegegrad zu Beginn der Studie von mindestens 2 extremer. In der Kontrollgruppe verschlechtern sich 29.5% (18/61) in der Interventionsgruppe 25.4% der Patienten (16/63) (Siehe Tabelle 10). Somit könnte ein Grund für das negative Ergebnis darin liegen, dass die

Ausgangssituation in der Pflege für die Studienpopulation zu gut war: Grad 0 bei 891 Studienteilnehmern (74.3%), Grad 1 bei 184 (15.3%), einen Pflegegrad über 1 haben 124 Teilnehmer (10.5%).

Sensitivitätsanalysen zum primären Outcome (Betrachtung der Veränderung des NBA Scores) zeigten Signale eines möglichen Therapieeffektes. Die gefundenen signifikanten Veränderungen im NBA sind jedoch klinisch nicht interpretierbar. Es wurde eine mittlere Veränderung um -1.7 Punkte berechnet (95% KI [-3.14; -0.30]). Auch zeigten die Per Protokoll Analysen (siehe Tabelle 12) für Patienten mit einer Dosierung der Intervention von mindestens 15 Einheiten signifikante Effekte, sowie für Patienten mit mindestens 15 Einheiten und der ursprünglich geplanten Operation.

Die Intervention wurde gut angenommen. Bei 97 der 616 Patienten in der Interventionsgruppe nahmen an der Prähabilitation nicht teil, davon 25 aus medizinischen und 23 aus organisatorischen Gründen, 44 auf persönlichen Wunsch und 5 aus anderen Gründen. 20 der verbleibenden 519 Patienten (3.9%) hatten weniger als 15 der vorgesehenen 48 Prähabilitationseinheiten. Im Median wurden 30 der 48 vorgesehenen Einheiten durchgeführt. Leider liegen dem Evaluator keine Daten zur Prozessanalyse der Interventionsdurchführung vor. Insofern kann nicht bewertet werden, wie gut die Compliance mit der Intervention (Dosierung der Intervention) war und wie gut die 18 selbstverantworteten Trainingseinheiten überhaupt wahrgenommen wurden (Reporting Bias?).

Effekte bei Sensitivitätsanalysen zum primären Endpunkt und zu den sekundären Endpunkten müssen mit großer Vorsicht bewertet werden. Es liegt ein Multiplizitätsproblem vor.

Die Intervention ist sicher. Blick auf Mortalität oder spezifische Parameter der OP-Durchführen liefern keine Evidenz für Nachteile der einen gegenüber der anderen Gruppe. Die mediane Zeit zwischen OP und Aufnahme in die Studie waren in der Kontrollgruppe 50 Tage (IQR [25, 93.5]) und in der Interventionsgruppe 56 Tage (IQR [39, 92]). Die Intervention verlängert wohl signifikant die Zeit bis zu OP. Der detektierte Unterschied ist jedoch als klinisch nicht relevant zu erachten.

Mögliche Variabilität in der NBA Bewertung wurden bei der Analyse nicht berücksichtigt. Es wurde auch keine Analyse der Intra-Rater und Inter-Rater Variabilität vorgenommen. Weiterhin liegen keine Angaben vor, wieviel Bewerter des NBA die Abschlussuntersuchung durchgeführt haben und ob damit Abhängigkeiten zwischen den Bewertungen durch die Bewerter vorliegen. In der Regel würde die Ignoranz der durch verschiedenen Bewerter verursachte Clusterung zu kleiner p-Werten führen. Aufgrund des nicht-signifikanten Ergebnisses der Studie ist aus dieser Richtung keine Verzerrung des Ergebnisses zu erwarten.

Die Studie hat ein pragmatisches und multizentrisches Design, wodurch die Verallgemeinerbarkeit der Ergebnisse auf größere Populationen möglich werden.

Die Ergebnisse der gesundheitsökonomischen Analysen aus gesellschaftlicher und GKV-Perspektive erzielten überwiegend höhere inkrementelle Kosten und höhere inkrementelle Effekte, was bedeutet, dass die Intervention in der Tendenz besser, aber teurer ist. Die mittleren Interventionskosten waren aus der gesellschaftlichen Perspektive generell etwas höher als aus der GKV-Perspektive, da hier auch Kosten für Patient:innen (z.B. für Transport) und Opportunitätskosten für Angehörige berücksichtigt wurden. Aus gesellschaftlicher

Perspektive werden für 45.547 € ein zusätzliches QALY erreicht bzw. für 27.197 € die Verhinderung der Pflegegradverschlechterung bei einer Person. Da diese Ergebnisse mit substantieller Unsicherheit verbunden sind, zeigte sich aber nur bei einer Zahlungsbereitschaft von 100.000 € eine hohe Wahrscheinlichkeit (72 % beim QALY und 86 % für die Verhinderung der Pflegegradverschlechterung), dass die Intervention als kosteneffektiv akzeptiert wird. Dabei sollte allerdings berücksichtigt werden, dass eine Pflegegradverschlechterung auch längerfristig mit höheren Kosten für das Gesundheitssystem durch höhere Pflegekosten einhergeht, die in den ermittelten IKER nur teilweise berücksichtigt wurden. Bezüglich des QALYs, zeigten die Subgruppen der Patient:innen, die mindestens 15 Einheiten der Prähabilitation erhielten (PP15), und die zusätzlich die geplante Operation erhalten haben (PP15OP), hohe Wahrscheinlichkeiten (> 75 %) für Kosteneffektivität ab einer Zahlungsbereitschaft von ca. 40.000 € (PP15) bzw. 50.000 € (PP15OP), was darauf hindeutet, dass die Wahrscheinlichkeit für Kosteneffektivität hier maßgeblich mit der Intensität der Prähabilitationsmaßnahmen zusammenhängt. Allerdings wurde dies bei der Verhinderung der Pflegegradverschlechterung nicht beobachtet. Hier zeigten die PP15 und PP15OP-Analysen erst ab 63.000 € bzw. 71.000 € Wahrscheinlichkeiten für Kosteneffektivität über 75%, während das bei der ITT-Analyse bereits ab 61.000 € der Fall war.

Es ist anzumerken, dass es in Deutschland keine festen Schwellenwerte für die Zahlungsbereitschaft gibt, anders als z.B. in Großbritannien, wo das National Institute for Health and Care Excellence vorgibt, dass Interventionen, die zwischen 20.000 und 30.000 £ pro zusätzlichem QALY kosten, ein gutes Preis-Leistungs-Verhältnis für den National Health Service darstellen [20]. Um die hier gezeigten Ergebnisse trotzdem an internationalen Guidelines zu orientieren und besser einordnen zu können, werden die Empfehlungen der World Health Organization – Choosing Interventions that are Cost-Effective (WHO-CHOICE) herangezogen, welche Zahlungsbereitschaften für kosten-effektive Interventionen beschreiben. WHO-CHOICE beurteilt eine Intervention als kosten-effektiv, sofern die IKER unter dem jährlichen Bruttoinlandsprodukt pro Kopf liegt [21]. Dieses lag in Deutschland im Jahr 2020 bei etwa 41.500 €. Allerdings sehen sich auch die WHO-CHOICE Schwellenwerte mit Kritik konfrontiert und sollten zu reinen Vergleichszwecken und Unterstützung eines mehrstufigen Health Technology Assessment-Entscheidungsprozesses dienen [22]. Zieht man diesen Schwellenwert von 41.500 € pro zusätzliches QALY heran liegt die Wahrscheinlichkeit für Kosteneffektivität in der QALY ITT-Analyse bei knapp 50%.

In dem vorab durch die Evaluator:innen durchgeführten systematischen Review zur Kosten-Effektivität der Prähabilitation war die Richtung der inkrementellen Effekte und Kosten bei der Mehrheit der studienbasierten ökonomischen Evaluationen (14 von 22) so, dass die Prähabilitation die Kontrolle (Regelversorgung) dominierte. Die eingeschlossenen Kosten-Nutzwert-Analysen fanden jeweils etwas höhere inkrementelle QALY-Werte zwischen 0,04 und 0,1 und tendenziell geringere Kosten in der Interventionsgruppe, sodass die Intervention hier die Kontrollgruppe dominierte [23 – 25].

Die ITT-Analysen aus GKV-Perspektive zeigen zwar eine positive Tendenz hinsichtlich der Effekte der Intervention, jedoch nur in sehr geringem Ausmaß. So betrug der inkrementelle Zugewinn an qualitätsadjustierten Lebensjahren (QALYs) lediglich 0,01 (95 % KI: -0,05; 0,07), während die Vermeidung von Pflegegradverschlechterungen um lediglich 2 %-Punkte (95 % KI: -0,08; 0,12) verbessert wurde. Gleichzeitig entstanden zusätzliche Kosten in Höhe von

2.409 € (95 % KI: -2.092; 7.067), was insgesamt zu einer niedrigen Wahrscheinlichkeit für Kosteneffektivität führt – selbst bei einer hohen Zahlungsbereitschaft pro zusätzlicher Nutzeneinheit. Deutlich positiver fällt das Ergebnis jedoch in der Subgruppe der Patient:innen aus, die mindestens 15 Einheiten der Prähabilitation erhielten (PP15-Analyse). Hier zeigte sich insbesondere in der Vermeidung von Pflegegradverschlechterungen ein stärkerer Effekt mit einer Reduktion von 4 %-Punkten (95 % KI: -0,07; 0,16) und gleichzeitig geringeren inkrementellen Kosten von 1.244 € (95 % KI: -4.067; 6.524). Dies deutet darauf hin, dass auch hier die Wahrscheinlichkeit für Kosteneffektivität maßgeblich mit der Intensität der Prähabilitationsmaßnahmen zusammenhängt. Insgesamt bleibt die Intervention jedoch nur in sehr begrenztem Maße kosteneffektiv.

Die Subgruppenanalysen verdeutlichten, dass die Ergebnisse stark variieren können. So zeigt sich bei beiden genannten Perspektiven, dass eine Durchführung der Prähabilitation im ambulanten statt im stationären Setting die Kosteneffektivität erheblich steigern kann. Dies ist auf die deutlich niedrigeren Kosten und gleichzeitig bessere Outcomes beim ambulanten Setting zurückzuführen. Ähnliche Trends wurden aus gesellschaftlicher Perspektive für die Subgruppen von Patient:innen mit einem Pflegegrad bei Baseline von 0 und Patient:innen ohne kognitive Auffälligkeiten beobachtet.

Aus beiden Perspektiven waren die Ergebnisse für jüngere, vor-gebrechliche Patient:innen, Patient:innen mit einer orthopädischen Operation bzw. einer neurochirurgischen Operation an der Wirbelsäule beobachtet und männlichen Patienten besser als über die gesamten Kohorten hinweg. Negative Ergebnisse zeigten sich jeweils für Patient:innen mit Prähabilitation im stationären Setting und einer Tumor-Operation, aus der GKV-Perspektive außerdem für ältere Patient:innen, weibliche Patient:innen und teilweise für gebrechliche Patient:innen. Diese Merkmale sind zum Teil eng korreliert, z.B. da jüngere, vor-gebrechliche Patient:innen ohne kognitive Auffälligkeiten bessere Voraussetzungen haben, um an der ambulanten Prähabilitation teilzunehmen, und Krebspatient:innen ggf. zu schwer erkrankt und daher eher eine stationäre Prähabilitation erhalten müssen. Bei den unterschiedlichen Ergebnissen je nach Operationsart wäre es überdies denkbar, dass die Prähabilitation einen kausalen Einfluss auf orthopädische Krankheitsbilder hat, wie z.B. fortgeschrittene Gonarthrose, und daher zu tendenziell besseren Outcomes führt. Bei Tumoren ist eine kausale Beeinflussung nicht möglich, sondern umgekehrt könnte möglicherweise eine Verzögerung der Therapie hier nachteilig gewesen sein. In einer Übersichtsarbeit von systematischen Reviews zur Prähabilitation wurden im Gegenzug dazu allerdings bessere Ergebnisse für Patient:innen mit Tumor-Operationen als für Patient:innen mit orthopädischen Operationen gefunden [26].

Da sich insbesondere im ambulanten Prähabilitationssetting deutliche Anzeichen für eine kosten-effektive Intervention zeigten, wurden hierfür explorative tiefergehende Analysen der Patient:innencharakteristika durchgeführt. Dabei stellte sich heraus, dass im ambulanten Setting insbesondere „gesündere“ Patient:innen von der Intervention profitieren. Die OR für kosten-effektivere Ergebnisse wurde signifikant erhöht, sofern Patient:innen jünger als der Median und vor-gebrechlich statt gebrechlich waren, sich im MOCA kognitiv unauffällig zeigten und bei der Baseline-Erhebung einen Pflegegrad von 0 gemäß NBA hatten. Weniger kosten-effektiv zeigte sich die Intervention im ambulanten Setting bei geplanten Herz-Operationen. Diese Ergebnisse unterstreichen die Bedeutung einer gezielten Patient:innenauswahl und einer differenzierten Betrachtung des Versorgungssettings.

Zukünftige Analysen sollten daher genauer untersuchen, welche Patient:innengruppen tatsächlich von der Intervention profitieren, um die Kosteneffektivität gezielt zu optimieren und ineffektive Anwendungen zu vermeiden.

Vergleicht man die Ergebnisse aus der Perspektive der GKV mit der gesellschaftlichen Perspektive, lassen sich insbesondere bei der Effektivität, aber auch bei den Kosten Unterschiede feststellen. So zeigen sich aus der gesellschaftlichen Perspektive insgesamt positivere Effekte als aus der GKV-Perspektive, insbesondere bei den inkrementellen QALYs und der inkrementellen Vermeidung von Pflegegradverschlechterungen (z.B. ITT-Analyse: 4,58 % vs. 2,18 %; 0,03 QALYs vs. 0,01 QALYs). Diese führen in Kombination mit etwas höheren inkrementellen Kosten zu wesentlich größeren IKER. Eine mögliche Erklärung für diese Unterschiede könnte in der Zusammensetzung der untersuchten Populationen liegen. Die Subgruppenanalysen haben gezeigt, dass die Intervention insbesondere bei männlichen und weniger stark erkrankten Patient:innen wirksamer ist. Ein Vergleich der Patient:innencharakteristika zwischen der Gesamtstudienpopulation und der GKV-Population zeigt jedoch, dass in der GKV-Population ein höherer Anteil an Frauen vertreten war und mehr Patient:innen gemäß MOCA kognitiv auffällig waren. Zudem war der Anteil der Patient:innen mit orthopädischen Operationen in der GKV-Population geringer, was ebenfalls zu der geringeren Kosten-Effektivität beitragen haben könnte. Im Gegensatz dazu waren bei der GKV-Population in der Interventionsgruppe weniger gebrechliche Patient:innen als in der Kontrollgruppe ($p=0.048$), was die Kosten-Effektivität der Intervention hätte steigern können.

Aus der Leistungserbringer-Perspektive gab es keine Hinweise, dass die Intervention kosteneffektiv ist. Die erhaltenen Entgelte waren in der Interventionsgruppe etwas geringer, die Verweildauer jedoch höher ohne Unterschied bei der (mit hohen Kosten verbundenen) Verweildauer auf der Intensivstation. Die Komplikationsrate war in der Interventionsgruppe etwas geringer, was potenziell Kosten sparen könnte. Dabei stellt sich bei schweren Komplikationen, die eine maschinelle Beatmung nach sich ziehen, allerdings die paradoxe Situation ein, dass hierfür ggf. höhere Entgelte erzielt werden können, da je nach Beatmungstunden höhere DRGs abgerechnet werden können.

Auch die Ergebnisse der gesundheitsökonomischen Evaluation müssen vorsichtig interpretiert werden. Bei der Schätzung der Interventionskosten mussten verschiedene Annahmen getroffen werden, z.B. wurde Sporttherapie wie Physiotherapie bewertet, da keine standardisierten Kostensätze für Sporttherapie vorlagen. Außerdem wurde die in der Studie angewandte Sessiondauer von 30 Minuten zugrunde gelegt, obwohl therapeutische Behandlungen z.T. kürzere (u.a. viele Physiotherapiezeffern) oder längere Behandlungszeiten vorsehen (vgl. Heilmittelkatalog [27]). Durch die COVID-19-Pandemie kam es zu Änderungen der ursprünglich geplanten Settings. So wurde ein neues Setting geschaffen (ambulant mit Fahrtkosten), in welchem dann viele Patient:innen ihre Prähabilitation erhalten haben anstelle des teilstationären Settings, welches durch Schließungen vorübergehend nicht oder nur sehr eingeschränkt zur Verfügung stand. So konnten Kosten reduziert werden. Andererseits konnten weniger Patient:innen als geplant in Sammeltransporten befördert werden und wurden stattdessen in Taxen transportiert, was ggf. zu höheren Transportkosten geführt hat.

Die Kostenschätzung mittels des FIMA-Fragebogens bei der Evaluation aus gesellschaftlicher Perspektive basierte auf den Erinnerungen von Patient:innen und/oder deren Angehörigen für die vergangenen drei Monate. Da viele Patient:innen kognitive Auffälligkeiten zeigten, ist

die Verlässlichkeit ihrer Angaben nicht gewährleistet. Jedoch war der Anteil an kognitiv auffälligen Patient:innen zwischen den Gruppen gleich verteilt (s. Tabelle 10), sodass es hier zu keiner systematischen Verzerrung gekommen sein sollte. Schließlich gab es für einige Inanspruchnahmen keinen entsprechenden standardisierten Bewertungssatz, z.B. für die Inanspruchnahme der Facharztgruppe Radiologie oder die Inanspruchnahme von Osteopathie. Dies betraf auch die Medikamente, für welche aus methodischen Gründen (s. Kapitel „Kosten“) aus der gesellschaftlichen Perspektive keine Kosten geschätzt werden konnten.

Die 3-Monats-Visite war formal nicht Teil der Studie, sondern diente der patientenzentrierten Kommunikation. Aus studienorganisatorischen Gründen wurden daher gewisse Fragebögen, darunter der EQ-5D-5L, bei der 3-Monats-Visite zudem nur abgefragt, wenn die Visite als Hausbesuch stattgefunden hat. Daher mussten bei der gesundheitsökonomischen Evaluation für die 3-Monats-Visite besonders viele Werte imputiert werden. Darüber hinaus spiegelt die Verwendung von Präferenzwerten der deutschen Allgemeinbevölkerung zur Berechnung des QALYs ggf. nicht die Präferenzen älterer Menschen wider, da kleine Gesundheitsgewinne möglicherweise unzureichend erfasst werden [28].

5 Schlussfolgerungen und Empfehlungen des Evaluators

Ziel der Studie war der Nachweis des Erhalts und der Verhinderung des Pflegegrads. Die Studie liefert hierzu keine schlüssigen Ergebnisse zum Effekt der untersuchten Prähabilitation auf dieses Ziel.

Gründe dafür sieht der Evaluator in der vorgefundenen Studienpopulation. Über 70% der eingeschlossenen Patient:innen sind mit einem Pflegegrad von 0 in die Studie gegangen. Es war für die Evaluation auch nicht möglich, die Qualität der Intervention außer durch sehr grobe Parameter zu bewerten.

Die in großer Zahl durchgeführten Analysen zu unterschiedlichen Aspekten des primären Endpunktes und einer umfassenden Anzahl sekundärer Endpunkte macht die Interpretation der gefundenen Signifikanzen ($p < 0.05$) und Signale ($p < 0.1$) für Therapieeffekte schwierig. Ob die in den verschiedenen Betrachtungen gefundenen Effekte systematisch sind ist nicht zu bewerten.

Explorativ liefert die Studie neue Richtungen für weiterführende Forschung (Spezifizierung der Population, Optimierung der Intervention, Interaktion des Interventionseffektes mit der Pflegesituation einer behandelten Person).

Der biometrische Evaluator erachtet die Intervention als sicher. Interessant sind die Ergebnisse in den betrachteten Per Protokoll Populationen. Hier zeigen sich überzeugender positive Effekte der Intervention.

Bei der gesundheitsökonomischen Evaluation lieferten die Hauptanalysen keine eindeutigen Hinweise darauf, dass die Intervention kosteneffektiv ist, sondern es zeigte sich, wie bei vielen neuen Interventionen, dass die Intervention in der Tendenz besser, aber teurer ist, sodass die Kosten-Effektivitäts-Akzeptanz maßgeblich von der Zahlungsbereitschaft abhängt.

Allerdings gab es deutliche Variationen in den Ergebnissen der Subgruppenanalysen, die zeigen, dass unter bestimmten Bedingungen – insbesondere bei Patient:innen, die unter 78

Jahren waren und kognitiv nicht eingeschränkt und vor-gebrechlich, bei denen die Prähabilitation ambulant erfolgen konnte – eine kosteneffektive Umsetzung durchaus möglich ist. Die ambulante Prähabilitation sollte daher weiter in Studien untersucht werden. Eine stationäre Prähabilitation kann dagegen nicht empfohlen werden, sondern es sollte geprüft werden, ob man die Indikation einer geriatrischen Komplexbehandlung auch für Patient:innen öffnen könnte, denen eine elektive Operation bevorsteht.

IV Literaturverzeichnis

1. Schaller SJ, Kiselev J, Loidl V, Quentin W, Schmidt K, Mörgeli R, Rombey T, Busse R, Mansmann U, Spies C; PRAEP-GO consortium,; PRAEP-GO investigators. Prehabilitation of elderly frail or pre-frail patients prior to elective surgery (PRAEP-GO): study protocol for a randomized, controlled, outcome assessor-blinded trial. *Trials*. 2022 Jun 6;23(1):468. doi: 10.1186/s13063-022-06401-x. Erratum in: *Trials*. 2023 Feb 14;24(1):111. doi: 10.1186/s13063-023-07106-5.
2. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Seeman T, Tracy R, Kop WJ, Burke G, McBurnie MA; Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001 Mar;56(3):M146-56. doi: 10.1093/gerona/56.3.m146.
3. Wingenfeld, K., Büscher, A., & Gansweid, B. (2008). Das neue Begutachtungsassessment zur Feststellung von Pflegebedürftigkeit. Abschlussbericht zur Hauptphase, 1, 1-128.
4. Müller-Mai, C. M., US, S. R., Kostuj, T., Dahlhoff, G., Günster, C., & Smektala, R. (2015). One-year outcomes for proximal femoral fractures: posthospital analysis of mortality and care levels based on health insurance data. *Der Unfallchirurg*, 118(9), 780-794.
5. Kolassa, J. E. (1995). A comparison of size and power calculations for the Wilcoxon statistic for ordered categorical data. *Statistics in Medicine*, 14(14), 1577-1581.
6. Statistical Solutions Ltd. (2021). *nQuery* (Version 7) [Computer software]. Retrieved November 26, 2025, from <https://nquery-advisor-nterim.software.informer.com/>
7. Pocock, S. J., Ariti, C. A., Collier, T. J., & Wang, D. (2012). The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European heart journal*, 33(2), 176-182.
8. Eckhardt H, Quentin W, Silzle J, Busse R, Rombey T. Cost-effectiveness of prehabilitation of elderly frail or pre-frail patients prior to elective surgery (PRAEP-GO) versus usual care - Protocol for a health economic evaluation alongside a randomized controlled trial. *BMC Geriatr*. 2024 Mar 6;24(1):231. doi: 10.1186/s12877-024-04833-5.
9. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, Briggs A, Sullivan SD. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. *Value Health*. 2015 Mar;18(2):161-72. doi: 10.1016/j.jval.2015.02.001.
10. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. 4th ed. 2015, Oxford: Oxford University Press.

11. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials. 2nd ed. 2014, Oxford University Press: Oxford.
12. Rombey T, Eckhardt H, Kiselev J, Silzle J, Mathes T, Quentin W. Cost-effectiveness of prehabilitation prior to elective surgery: a systematic review of economic evaluations. BMC Med. 2023 Jul 19;21(1):265. doi: 10.1186/s12916-023-02977-6.
13. Statista. Größte gesetzliche Krankenkassen in Deutschland nach der Versichertenzahl im Jahr 2025. Hamburg, 2025. <https://de.statista.com/statistik/daten/studie/856392/umfrage/groesste-gesetzliche-krankenkassen-in-deutschland-nach-der-versichertenzahl/>
14. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, Swinburn P, Busschbach J. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013 Sep;22(7):1717-27. doi: 10.1007/s11136-012-0322-4.
15. Ludwig K, Graf von der Schulenburg JM, Greiner W. German Value Set for the EQ-5D-5L. Pharmacoeconomics. 2018 Jun;36(6):663-674. doi: 10.1007/s40273-018-0615-8.
16. World Health Organization. Measuring health and disability: Manual for WHO disability assessment schedule ; WHODAS 2.0. Genf; 2010. https://iris.who.int/bitstream/handle/10665/43974/9789241547598_eng.pdf?sequence=1
17. Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD, de Santibañes E, Pekolj J, Slankamenac K, Bassi C, Graf R, Vonlanthen R, Padbury R, Cameron JL, Makuuchi M. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg. 2009 Aug;250(2):187-96. doi: 10.1097/SLA.0b013e3181b13ca2.
18. Muntendorf LK, Brettschneider C, Konnopka A, König H. Aktualisierung der standardisierten Bewertungssätze aus gesellschaftlicher Perspektive für gesundheitsökonomische Evaluationen. Gesundheitswesen, 2024. 86(5): 389-393. doi: 10.1055/a-2169-1480
19. Seidl H, Bowles D, Bock JO, Brettschneider C, Greiner W, König HH, Holle R. FIMA--Fragebogen zur Erhebung von Gesundheitsleistungen im Alter: Entwicklung und Pilotstudie Gesundheitswesen. 2015 Jan;77(1):46-52. German. doi: 10.1055/s-0034-1372618.
20. [Office for Health Improvement and Disparities](https://www.gov.uk/guidance/cost-utility-analysis-health-economic-studies). London, UK. 2025. <https://www.gov.uk/guidance/cost-utility-analysis-health-economic-studies>
21. Kazibwe J, Gheorghe A, Wilson D, Ruiz F, Chalkidou K, Chi YL. The Use of Cost-Effectiveness Thresholds for Evaluating Health Interventions in Low- and Middle-Income Countries From 2015 to 2020: A Review. Value Health. 2022 Mar;25(3):385-389. doi: 10.1016/j.jval.2021.08.014.
22. Bertram MY, Lauer JA, Stenberg K, Edejer TTT. Methods for the Economic Evaluation of Health Care Interventions for Priority Setting in the Health System: An Update From WHO CHOICE. Int J Health Policy Manag. 2021 Nov 1;10(11):673-677. doi: 10.34172/ijhpm.2020.244.
23. Fernandes L, Roos EM, Overgaard S, Villadsen A, Sjøgaard R. Supervised neuromuscular exercise prior to hip and knee replacement: 12-month clinical effect and cost-utility analysis

alongside a randomised controlled trial. BMC Musculoskelet Disord. 2017 Jan 6;18(1):5. doi: 10.1186/s12891-016-1369-0.

24. Nguyen C, Boutron I, Roren A, Anract P, Beaudreuil J, Biau D, Boisgard S, Daste C, Durand-Zaleski I, Eschalièr B, Gil C, Lefèvre-Colau MM, Nizard R, Perrodeau É, Rabetrano H, Richette P, Sanchez K, Zalc J, Coudeyre E, Rannou F. Effect of Prehabilitation Before Total Knee Replacement for Knee Osteoarthritis on Functional Outcomes: A Randomized Clinical Trial. JAMA Netw Open. 2022 Mar 1;5(3):e221462. doi: 10.1001/jamanetworkopen.2022.1462.

25. Tveter, A., et al., Cost-utility analysis of multimodal occupational therapy in patients with thumb base osteoarthritis. Osteoarthritis and Cartilage, 2020. 28: S439. doi: [10.1016/j.joca.2020.02.683](https://doi.org/10.1016/j.joca.2020.02.683)

26. McIsaac DI, Gill M, Boland L, Hutton B, Branje K, Shaw J, Grudzinski AL, Barone N, Gillis C; Prehabilitation Knowledge Network. Prehabilitation in adult patients undergoing surgery: an umbrella review of systematic reviews. Br J Anaesth. 2022 Feb;128(2):244-257. doi: 10.1016/j.bja.2021.11.014.

27. Gemeinsamer Bundesausschuss. Richtlinie des Gemeinsamen Bundesausschusses über die Verordnung von Heilmitteln in der vertragsärztlichen Versorgung (Heilmittel-Richtlinie/Heilm-RL). Berlin, 2025. https://www.g-ba.de/downloads/62-492-3676/Heilm-RL_2024-10-29_iK-2025-01-01.pdf

28. Marten O, Greiner W. EQ-5D-5L reference values for the German general elderly population. Health Qual Life Outcomes. 2021 Mar 6;19(1):76. doi: 10.1186/s12955-021-01719-7.

V Anlagen

Anlage 1: Anhang der Gesundheitsökonomischen Evaluation

Anlage 2: Statistischer Analyseplan

Anlage 1: Anhang der Gesundheitsökonomischen Evaluation

Tabelle A1: Gesamtkosten gesellschaftliche Perspektive nach Winsorisierung ohne Imputation (in €)

Kostenblock	Interventionsgruppe (n = 616)					Kontrollgruppe (n = 538)				
	MW	SD	Median	Q1	Q3	MW	SD	Median	Q1	Q3
Interventionskosten	2.447	2.857	1.555	1.042	2.131	0	0	0	0	0
KH-Indexaufenthalt*	9.529	8.564	7.701	4.612	10.860	9.596	9.196	7.557	3.960	11.294
Follow-Up Kosten Total	10.772	17.269	4.321	1.162	11.832	11.083	17.397	4.524	877	11.985
Krankenhaus (stationär)	4.241	9.538	0	0	4.047	4.264	9.598	0	0	4.047
Krankenhaus (ambulant)	147	423	0	0	0	138	437	0	0	0
Psychiatrie	0	0	0	0	0	0	0	0	0	0
Arztbesuche	529	428	440	208	758	527	464	394	201	760
Therapieberufe	505	580	308	0	759	471	589	257	0	679
Pflegedienst	536	1.658	0	0	0	682	1.919	0	0	0
Informelle Pflege	2.722	11.138	0	0	0	2.879	12.032	0	0	0
Haushaltshilfe	103	299	0	0	0	101	289	0	0	0
Hilfe von Nachbarn etc.	314	1.398	0	0	0	435	1.615	0	0	0
Tagespflege	0	0	0	0	0	0	0	0	0	0
Kurzzeitpflege	0	0	0	0	0	0	0	0	0	0
Pflegeheim	228	1.213	0	0	0	137	949	0	0	0
Rehabilitation	1.236	2.032	0	0	2.772	1.224	2.085	0	0	2.587
Hilfsmittel	213	422	0	0	131	224	464	0	0	121
Gesamtkosten	23.296	23.109	15.896	9.938	29.774	21.398	23.352	13.575	8.336	27.088

*enthält auch Werte i.H.v. 0 € für Teilnehmende ohne Operation.

IG, Interventionsgruppe; KG, Kontrollgruppe; KH, Krankenhaus; KI, Konfidenzintervall; MW, Mittelwert; SD, Standardabweichung.

Tabelle A2: Gesamtkosten gesellschaftliche Perspektive ohne Winsorisierung nach Imputation (in €)

Kostenblock	Interventionsgruppe (n = 616)					Kontrollgruppe (n = 538)				
	MW	SD	Median	Q1	Q3	MW	SD	Median	Q1	Q3
Interventionskosten	2.414	2.837	1.546	1.025	2.122	0	0	0	0	0
KH-Indexaufenthalt*	11212	12127	7996	5621	11334	11569	12819	7982	5105	11958
Follow-Up Kosten Total	15370	24664	6071	2868	15015	17439	29304	6992	3473	17056
Krankenhaus (stationär)	6171	14581	607	0	5742	7391	20273	1012	0	6121
Krankenhaus (ambulant)	172	590	0	0	0	181	791	0	0	0
Psychiatrie	87	1586	0	0	0	31	748	0	0	0
Arztbesuche	719	674	616	385	876	745	766	626	355	907
Therapieberufe	733	797	508	191	1030	741	831	489	190	1023
Pflegedienst	1509	8515	0	0	418	1713	8541	0	0	475
Informelle Pflege	2889	12355	0	0	0	3282	15267	0	0	0
Haushaltshilfe	202	1417	0	0	0	158	919	0	0	0
Hilfe von Nachbarn etc.	507	2676	0	0	0	706	3433	0	0	0
Tagespflege	16	156	0	0	0	45	440	0	0	0
Kurzzeitpflege	51	520	0	0	0	62	745	0	0	0
Pflegeheim	429	2560	0	0	0	233	1764	0	0	0
Rehabilitation	1647	2201	684	0	3248	1893	2658	1072	0	3040
Hilfsmittel	238	487	61	0	184	258	561	61	0	154
Gesamtkosten	28996	31326	18620	12319	33928	29008	35463	16854	11416	32062

*enthält auch Werte i.H.v. 0 € für Teilnehmende ohne Operation.

IG, Interventionsgruppe; KG, Kontrollgruppe; KH, Krankenhaus; KI, Konfidenzintervall; MW, Mittelwert; SD, Standardabweichung.

Tabelle A3: Gesamtkosten GKV-Perspektive ohne Winsorisierung (in €)

Kostenblock	Interventionsgruppe (n = 143)				Kontrollgruppe (n = 130)			
	Mean	SD	Median	IQR	Mean	SD	Median	IQR
Interventionskosten	1.958	2.693	1.032	207	0	0	0	0
KH-Indexaufenthalt*	12.070	17.233	7.798	6.918	11.198	17.832	7.624	5.563
Follow-up Kosten Total	20.024	26.126	11.988	19.001	20.669	24.268	12.375	18.903
Arztbesuche	1.674	2.220	1.253	1.176	1.669	2.930	1.092	920
Heilmittel	907	1.435	404	1.047	1.092	2.375	276	1.106
Hilfsmittel	632	1.019	260	597	779	1.672	246	673
Häusliche Krankenpflege	208	1.116	0	0	183	691	0	0
Stationäre KH-Aufenthalte	6.714	14.209	0	8.934	6.932	11.065	2.232	9.534
Ambulante KH-Aufenthalte	262	412	151	323	193	460	0	190
Krankentransporte	693	1.391	111	705	1.161	3.586	96	749
Arzneimittel	4.734	18.818	1.624	2.657	3.708	13.716	1.415	2.003
Pflege	2.617	4.454	0	4.137	3.073	6.832	0	4.245
Ambulante Reha	3	24	0	0	46	312	0	0
Stationäre Reha	1.580	2.348	0	2.763	1.834	3.488	798	2.787
Gesamtkosten	34.052	37.892	23.383	29.090	31.868	33.719	22.967	25.395

*enthält auch Werte i.H.v. 0 € für Teilnehmende ohne Operation.

IG, Interventionsgruppe; IQR, Interquartilabstand; KG, Kontrollgruppe; KH, Krankenhaus; KI, Konfidenzintervall; MW, Mittelwert; SD, Standardabweichung.

Statistical Analysis Plan (SAP)



TRIAL FULL TITLE (Acronym)	Prehabilitation of elderly frail or pre-frail patients prior to elective surgery – a randomized, controlled, outcome-assessor blinded study (PRAEP-GO)
Protocol identification	UTN U1111-1253-4820
Clinical Trials	NCT04418271
Grant Number	01NVF18024
Financial support	Innovation Committee of the Joint Federal Committee
Protocol version; Date	Version 1.1; February 14, 2023 Doi: 10.1186/s13063-023-07106-5
SAP Version; Date	Draft Version 1.0; November 2024
SAP Author	Prof. Dr. Ulrich Mansmann Verena Loidl, MPH, Stefan Buchka, MSc, Dr. Maria Stark MSc, Ahmed Elhelaly
SAP Reviewer	PD Dr. Ulrike Grittner

PRINCIPAL INVESTIGATOR:

Prof. Dr. Claudia Spies
Charité – Universitätsmedizin Berlin
Klinik für Anästhesiologie und
Intensivmedizin CCM/CVK

Augustenburger Platz 1
D-13353 Berlin

TRIAL STATISTICIAN:

Prof. Dr. Ulrich Mansmann
IBE - Institut für Medizinische
Informatik, Biometrie und
Epidemiologie,
Lehrstuhl für Biometrie und
Bioinformatik
Ludwig-Maximilians-Universität
München
Marchioninistr. 15
D-81377 München

Prehabilitation of elderly frail or pre-frail patients prior to elective surgery – a randomized, controlled, outcome-assessor blinded study (PRAEP-GO)

Statistical analysis plan for the PRAEP-GO trial (SAP)

Version 1.0 of [03.12.2024]

Investigational Intervention: Prehabilitation
Comparator: Care as usual
Indication: Frailty of aged persons waiting for a surgery

Approved by

Prof. Dr. Claudia Spies,
Principal Investigator

Berlin 17.12.24

Place and date

Signature

Prof. Dr. Ullrich Mansmann,
Statistician

Munich, 19.12.2024

Place and date

Signature

Reviewed by

PD Dr. Ulrike Grittner,
External Reviewer

Berlin, 17.12.24

Place and date

Signature

Authors

Prof. Dr. Ulrich Mansmann¹

Verena Loidl, MPH¹, Stefan Buchka, MSc.¹, Dr. Maria Stark, MSc.¹, Ahmed Elhelaly, MSc.¹

¹ Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Ludwig-Maximilians University, Marchioninstr. 15, 81377 Munich, Germany

Table of Contents

1.	Background	8
1.1	Trial objectives	8
1.2	Trial design.....	9
1.3	Hypotheses	9
1.3.1	Primary hypothesis.....	9
1.3.2	Secondary hypotheses	9
1.4	Participants	15
2.	Analysis sets.....	16
2.1	Definitions	16
2.1.1	Full Analysis Set (FAS) / Intention to treat (ITT).....	16
2.1.2	Per Protocol Population (PP)	16
2.1.3	Safety Population (SAF)	17
2.2	Application.....	17
2.3	Trial centres.....	19
3.	Analysis variables	20
3.1	Study Variables and study schema	20
3.2	Explanatory variables / confounders (at baseline).....	21
3.3	Explanatory variables / confounders between baseline (v2) and surgery.....	24
3.4	Explanatory variables / confounder at surgery	25
3.5	Explanatory variables / confounders after surgery	26
3.6	Primary endpoint.....	28
3.6.1	Composite endpoint: Incorporate death	29
3.7	Secondary endpoints.....	29
3.7.1	Change in NBA-total-Score	29
3.7.2	Change in NBA domains	30
3.7.3	Frailty	30
3.7.4	Neurocognitive Disorders (NCD)	30
3.7.5	Health-related quality of life	31
3.7.6	Functioning and disability	31
3.7.7	Mortality	31
3.7.8	Drop-outs / Censoring	31
3.7.9	Change in functional status.....	31
3.7.10	Change in activities of daily living	32
3.7.11	Probability of moving to a nursing home	32
3.7.12	Duration of rehabilitation or follow-up treatment.....	32
3.8	Safety/Tolerability.....	32
4.	Handling of missing values and outliers.....	34
4.1	Missing values	34
5.	Statistical analyses / methods	34
5.1	Subject disposition	34
5.2	Demography and baseline characteristics.....	34
5.3	Prior or concomitant medication and diseases.....	35
5.4	Exposition to treatment/Compliance.....	35
5.5	Primary analysis - ITT	35
5.5.1	The ordinal aspect: Wilcoxon-Mann-Whitney rank sum test by Brunner et al. (67,68)	35
5.5.2	The binary aspect.....	36
5.5.3	Stratified area under the curve (ROC-GLM)	36
5.6	Sensitivity analysis	36
5.6.1	Hierarchical endpoint (play the winner).....	36

5.6.2	On the score level: LMM: Score difference ~ baseline score + treatment.....	37
5.7	Secondary analyses -ITT	38
5.7.1	Efficacy	38
5.7.2	Safety/Tolerability	39
5.8	Per Protocol Analysis	40
5.8.1	Inverse Probability Weighting (IPW)	41
5.8.2	Directed Acyclic Graphs (DAG)	41
5.8.3	IPW for intervention adherence	42
5.8.4	IPW for death/censoring	42
5.8.5	Outcome model.....	43
5.8.6	Parametric g-formula	44
5.9	Propensity Score Weighting (IPW) for probability receiving the surgery ..	44
5.10	Patient - Treatment Interaction.....	44
6.	Other Variables.....	45
7.	Planned subgroup analyses	45
8.	Interim analyses	45
9.	Deviations from the protocol	45
10.	Interpretation of results	45
11.	Data problems	45
12.	Software	45
13.	References	46

List of Figures

Figure 1: Study visits. The baseline is conducted at V2, and the 12-months post-surgery follow-up is conducted at V18.	9
Figure 2: Patient inclusion for intention to treat (ITT) and per protocol (PP) analysis after randomization. SDM: Shared decision making.	18
Figure 3: Probability curves to get a surgery (time until surgery).....	25
Figure 4: DAG considering adherence to treatment.....	42
Figure 5: DAG considering competing risks (death vs. trial outcomes).....	43
Figure 6: DAG considering censoring events	43

List of Tables

Table 1: Overview of all trial centres, the number of patients randomized, and the level of care, categorized by maximum care/university hospitals and other hospitals according to bed count.	19
Table 2: Schedule of screening, enrolment, intervention (Shared Decision-Making Conference (SDM) and prehabilitation), and assessments.....	20
Table 3: Baseline characteristics per randomisation group.	23
Table 4: Describes the measures of care for the shared decision-making conference and prehabilitation for the intervention group.	24
Table 5: Describes the situation peri-surgery.	26
Table 6: Explanatory variables measured after surgery for intervention (prehabilitation) and control (usual care) groups.....	27
Table 7: Determination of the Level of Care Dependency (LCD) and corresponding points of the NBA-score (Neues Begutachtungsassessment).....	28
Table 8: Change in the Level of Care Dependency (LCD) for the intervention and control group.	28
Table 9: Modules of the New care needs assessment (NBA) and its weighting to the total sum score.	29
Table 10: Description of safety parameters in the PRAEP-GO study.	33

Abbreviations

2-MST	2-Minute Step Test
ADL	Activity of daily living
AE	Adverse event
ASA	American Society of Anesthesiologists classification
BMI	Body-Mass-Index
CCI	Charlson Comorbidity Index
CDS	Clavien-Dindo-Score
CHARMI	Charité Mobility Index
CI	Confidence interval
EQ-5D-5L	EuroQol quality of life 5-level EQ-5D version
EQ-VAS	EuroQol visual analogue scale
FAS	Full Analysis Set
GLM	Generalized linear model
IADL	Instrumental activity of daily living
ICU	Intensive Care Unit
IPW	Inverse Probability Weighting
ITT	Intention to treat
LCD	Level of care dependency
LMM	Linear mixed model
MNA-SF	Mini Nutritional Assessment
MoCA	Montreal Cognitive Assessment
NBA	Neue Begutachtungsassessment, New Nursing Care Grading Assessment, Score that can be transformed into the level of care dependencies
NCD	Neurocognitive disorder
No	Number
PP	Per Protocol
PRAEP-GO	Prehabilitation of elderly patients with frailty syndrome before an elective surgery
RCT	Randomised Controlled Trial

ROC	Receiver Operating Characteristics
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SDM	Shared Decision Making
SOS	Social Situation questionnaire
TU	Therapy units
TUG	Time Up and Go Assessment
TSG	Stair climbing speed (Treppensteigegeschwindigkeit)
V	Visit
WHODAS	WHO Disability Assessment Schedule
ZUF	Patient Satisfaction Questionnaire

This Statistical Analysis Plan (SAP) is based on the published study protocol of Schaller et al., 2022 (1,2). It also takes the amended version from July 2024 into account. The study protocol already provides descriptions of certain aspects of the statistical methods and study design in a condensed format. This SAP aims to provide a clear, pre-specified plan for handling and interpreting data in the PRAEP-GO study. This SAP provides an extensive analysis plan for the official report of the PRAEP-GO study to the "*Innovationsausschuss*" of the Federal Joint Committee (G-BA) in Germany. Only the main analyses as described before will be presented in the first / main publication.

1. Background

Due to the increasing life expectancy within the German population and significant medical progress, more and more complex surgical interventions are performed in elderly patients. Even though severe postoperative complications, such as mortality, have been reduced in the last decades (3), elderly patients are at risk of losing autonomy and to develop medium and long-term disabilities in the cognitive and functional domains (4–6). Additionally, elderly patients have higher rates of hospital readmissions associated with a delay in functional recovery (7). The risk of postoperative complications is not only determined by pre-existing conditions, such as diabetes mellitus or cardiovascular disease (8–10) but also by the presence of a frailty syndrome (11–14). Frailty is associated with restrictions in functional reserves, including mobility, muscle strength, and vital capacity. It is an independent risk factor for the development of postoperative complications, such as postoperative delirium (15), overall mortality, and long-term care (16), as well as long-term cognitive disorders (17). The prevalence of frailty is estimated to be between 4.0 and 27.3% within the population ≥ 65 years of age (18). However, the prevalence is much higher in the perioperative setting, with up to 42%, depending on the surgical discipline and screening tool used (19). Prehabilitation is a term used to describe preventive programs targeting specific health-related issues, and frail patients may benefit from such an approach aiming to reduce the impact of individual frailty syndrome components. Prehabilitation interventions are safe to perform (20,21) and have been demonstrated to accelerate the recovery of functional skills (22,23); reduce rates of postoperative complications, such as delirium and cognitive disorders (24–27); and shorten hospital stays (28). However, most of these findings are restricted to specific patient groups or types of surgery, while evidence in a general surgical population is lacking (29).

1.1 Trial objectives

The study aims to evaluate the clinical effectiveness of a prehabilitation program for frail or pre-frail elderly prior to elective surgery, compared to usual care. Specifically, we want to investigate the change in the level of care dependency (LCD) as well as other factors related to clinical outcomes before and one year after surgery.

Further relevant secondary and exploratory endpoints are of interest. This SAP does not specify for each endpoint its analysis. It provides in a more generic manner the analysis strategies which will be applied.

Alongside this evaluation, a health economic evaluation is conducted to investigate whether the prehabilitation for frail or pre-frail elderly prior to elective surgery, compared to usual care, is cost-effective from different perspectives. The responsibility for the health economic evaluation, including cost-effectiveness analyses, rests with the Technische Universität Berlin and has been detailed elsewhere (30).

1.2 Trial design

The PRAEP-GO study is a national multicenter, assessor-blinded, randomized, pragmatic, controlled clinical trial conducted in Germany. Randomisation is planned as a two-armed, parallel group design with a 1:1 allocation ratio for each participating hospital. The follow-up period is ongoing until twelve months after surgery. The visits are shown in Figure 1.

Visit	Day	Description
Phase I: Screening and inclusion in the study (intervention and control groups)		
V0	Day 1	Frailty-screening
V1	Day 1/2	Inclusion into study
V2	Day 1/2	Baseline assessment, followed by randomization
Phase II: Shared decision-making (intervention group only)		
V3	V1/V1 + 1–5 days	SDM conference
Phase III: Prehabilitation (intervention group only)		
V4	22–25 days before surgery	Start prehabilitation
V5	V4 + 21 days	End prehabilitation
Phase IV: Follow-up (intervention and control groups)		
V6	7–14 days post-surgery	Discharge from hospital
V7, V8	1/2 months post-surgery	Telemedical/telephone interview
V9	3 months post-surgery	Follow-up at 3 months (telemedical/telephone interview or home visit)
V10–V17	4–11 post-surgery	Telemedical/telephone interview
V18	12 months post-surgery	Follow-up at 12 months (home visit)

Figure 1: Study visits. The baseline is conducted at V2, and the 12-months post-surgery follow-up is conducted at V18.

1.3 Hypotheses

1.3.1 Primary hypothesis

The shared decision-making conference and the 3-week prehabilitation program want to maintain, or, if possible, reduce the level of care dependency (according to the new nursing care grading assessment (NBA)) 1 year after surgery.

The statistical hypotheses are formulated two-sided to assess any significant change in the level of care dependency, whether it improves/remains stable, or worsens.

$$H_0: \Delta_{NBA_E} = \Delta_{NBA_C} \text{ vs. } H_1: \Delta_{NBA_E} \neq \Delta_{NBA_C}$$

1.3.2 Secondary hypotheses

1. Key secondary endpoints

1.1. Level of care dependency one year after surgery reported as Barthel Index (Mahoney and Barthel 1965)

1.2. Change in NBA score (Wingenfeld et al., 2008 (31))

1.3. Change in individual NBA domains (Wingenfeld et al., 2008 (31))

2. Degree of Frailty assessed by Fried's Phenotype (Fried, Tangen et al. 2001 ((32))

3. Cognitive Endpoints

3.1. Screening tool for cognitive vulnerability by MiniCog (Borson, Scanlan et al. 2000 (33))

- 3.2. Dementia risk assessed by MOCA (Nasreddine, Phillips et al. 2005 (34)) deterioration (change) and its absolute value if < 26 .
- 3.3. Frequency of neurocognitive vulnerability (deterioration) at 3 or 12 months as surrogate for NCD by 1) Frequency of cognitive impairment/vulnerability is calculated from a change in composite of MiniCog (Borson, Scanlan et al. 2000 (33)), Animal Naming Test (ANT) (Sager, Hermann et al. 2006 (35)), TMT-B (Specka, Weimar et al. 2022 (36)) and MOCA (Nasreddine, Phillips et al. 2005 (34)) 2) subjective concern (MMQ) (37)) and 3) IADL/ADL (Mahoney and Barthel 1965, Lawton and Brody 1969 (38,39))
4. Perioperative complications
 - 4.1. Incidence of postoperative delirium using billing data (Measurement duration: up to hospital discharge)
 - 4.2. Perioperative complications assessed by Clavien-Dindo Dindo, Demartines et al. 2004 (40)) (Measurement duration: up to hospital discharge)
5. Postoperative complications (Measurement duration: up to hospital discharge)
 - 5.1. ICU admission (unplanned)
 - 5.2. ICU length of stay
 - 5.3. Hospital length of stay
 - 5.4. Postoperative functional mobility, i.e. mobilisation speed assessed by CHARMI (Liebl, Elmer et al. 2016)
6. Physical Function
 - 6.1. Functional endurance measured with 2-min step test (2-MST) (Bohannon and Crouch 2019 (41))
 - 6.2. Function of the respiratory system measured with peak expiratory flow (PEF) (Knudson et al. 1976 (42), Nunn et al. 1989 (43))
 - 6.3. Muscle strength measured with handgrip strength (Roberts, Denison et al. 2011 (45))
 - 6.4. Functional mobility measured with gait speed over 15ft (Bohannon and Wang 2019 (46))
 - 6.5. Stair-climbing speed (Hinman, O'Connell et al. 2014 (47))
 - 6.6. Timed Up & Go Test (TUG) (Podsiadlo and Richardson 1991 (48))
7. Nutritional endpoints
 - 7.1. Nutritional status by MNA-SF (Rubenstein, Harker et al. 2001 (49))
 - 7.2. Change in arm circumference
 - 7.3. Change in calf circumference
 - 7.4. Change in body weight

8. Depression and anxiety assessed by PHQ-8 (Kroenke, Spitzer et al. 2010 (50)) and GAD-7 (Lowe, Decker et al. 2008 (51)), respectively
9. Level of everyday mobility using Life-Space Assessment (LSA) (Taylor, Buchan et al. 2019 (52))
10. Falls
 - 10.1. Occurrence of falls
 - 10.2. Fear of falling assessed by Activities-Specific Balance Confidence 6 Scale (ABC-6) (Schepens, Goldberg et al. 2010 (53))
11. Health status
 - 11.1. Quality of life assessed by EQ-5D-5L (Janssen, Pickard et al. 2013 (54))
 - 11.2. Disability assessed by WHODAS 2.0 (Garin, Ayuso-Mateos et al. 2010 (55))
 - 11.3. Instrumental Activities of daily living (IADL) using the questionnaire by Lawton and Brody (Lawton and Brody 1969 (39))
12. Mortality

The following measures regarding SDM and prehabilitation will be reported together with the primary outcome analysis.

1. Involvement SDM with PEF-9 (= SDM-Q-9) and SDM-Q-Doc (Kriston, Harter et al. 2012, Scholl, Kriston et al. 2012 (56,57))
2. Satisfaction with the intervention (SDM & prehabilitation) using the (ZUF-8) (Schmidt and Nübling 2002 (58))
3. Autonomy preference assessed by Autonomy-Preference-Index, modified German version (API-Dm) (Nease and Brooks 1995 (59))
4. Prehabilitation setting (out-patient, Day-clinic, in-patient)
5. Planned prehabilitation components (e.g., (training, nutrition, cognitive and psychological support, and other or mixed support)
6. Exercise adherence assessed by the number of supervised prehabilitation sessions completed

Depending on the data quality (which may not allow a more detailed modelling) specific secondary and exploratory outcomes are only reported by simple descriptive statistics per treatment groups.

Hospital stay evaluation (Peri- and postoperative)

Operation outcome based on postoperative complication rates according to Clavien-Dindo; measurement duration: During the hospital stay.

Although "Count of new discharge diagnoses; measurement duration: at hospital discharge" was a "pre-specified other outcome" it cannot be calculated from the PRÄP-GO database. Due to the opening of health insurance providers during the study period, complete insurance data are not available for all patients. However, by combining the available data with health insurance records, this outcome may be calculable in the future.

The following postoperative outcomes until hospital discharge are analysed:

- Admission to the intensive care unit (ICU), either as a planned precautionary measure or due to unexpected complications.
- The duration of stay in the ICU, measured in hours (counting all ICU hours of the index hospital stay).
- The total length of hospital stays from admission to discharge, measured in days.

Cognitive evaluation

This includes:

1. Screening for cognitive vulnerability (by MiniCog), with <3 points indicating a suspicion of cognitive impairment.
2. Dementia risk by MOCA
 - a. Vulnerability: Change of MOCA (Baseline (V2) to 12 month after surgery (V18))
 - b. Impairment: MOCA < 26 points
3. Frequency of neurocognitive vulnerability at Baseline (V2) and 12 months after surgery (V18) as a surrogate for (mild/severe) NCD. Patient scores positive if:
 - a. One of the following
 - i. MiniCog < 3 points
 - ii. ANT < 15 animals/minute
 - iii. TMT-B time > 97 s (70-74 y) / > 115 s (75-79 y) / > 133 s (80-84 y) / > 138 s (85+ y) / abortion / > 5 min
 - iv. MOCA < 26 points
 - b. AND MMQ positive (<72 points)
 - c. for severe NCD only: AND IADL/ADL restriction (Barthel Index < 100 points/ IADL Score < 8 points)

Functional evaluation

Changes in the physical functional outcomes 6.1 to 6.6 are reported by descriptive statistics per treatment groups. For each of these outcome measures, values are collected preoperative at baseline and 1 year postoperative.

Fall evaluation

Risk of falling (Occurrence of falls within last 12 months)

Fear of falling (ABC-6)

The occurrence of falls measures the frequency or incidence of falls experienced by the patient from baseline up to 1 year postoperatively.

Fear of falling - assessed by the ABC-6 Scale - measures the level of concern or fear a patient has about falling during various activities.

Psychosocial evaluation

Frequency of depressive and anxiety symptoms

The severity of depressive symptoms is measured with the Patient Health Questionnaire-8 (PHQ-8) up to one year postoperatively. The Generalized Anxiety Disorder-7 (GAD-7) measures the severity of anxiety symptoms also up to one year postoperatively.

SDM and prehabilitation evaluation

Extent of involvement in the shared decision-making process; measurement duration: during shared decision-making process

Extent of patients, relatives and health professionals' involvement in shared decision-making process; 9-Item Shared Decision-Making Questionnaire (SDM-Q-9/SDM-Q-Doc); measurement duration: following shared decision-making process

Extent of patient participation preference: questionnaire for participatory decision making (PEF-FB-9))

The following outcomes assess aspects of shared decision-making (SDM) and prehabilitation:

1. Patient involvement in SDM: Measured using PEF-9 (SDM-Q-9) and SDM-Q-Doc, which evaluate the level of patient participation in the decision-making process.
2. Patient satisfaction with SDM and prehabilitation: Assessed using the ZUF-8 questionnaire to gauge satisfaction with both the SDM and the prehabilitation.
3. Preferences for autonomy in decision-making: Evaluated using the modified German version of the Autonomy-Preference-Index (API-Dm).
4. Prehabilitation Setting decided in the SDM: Records the type of setting where prehabilitation is planned to be provided, categorized as (out-patient, Day-clinic, in-patient) Planned Prehabilitation Components: Describes the specific content of the prehabilitation, such as endurance training, strength training, ADL training, nutritional therapy, etc., as planned during the SDM.
5. Adherence to prehabilitation: Measured by the total number of supervised prehabilitation sessions (including nutritional therapy sessions) completed, with each session lasting 30 minutes.
6. Setting of prehabilitation (out-patient, Day-clinic, in-patient)
7. Duration of prehabilitation (without and with interruption)

8. Performed prehabilitation components (training, nutrition, cognitive and psychological support, and other or mixed support)

1.4 Participants

Inclusion criteria:

- Age \geq 70 years
- Patient able to give consent or existing legal guardians in case of patient unable to give consent
- Planned elective surgery/interventional procedure
- Expected anaesthesia duration \geq 60 minutes
- Pre-frail (at least one item indicated as positive) or frail (three or more items indicated as positive) based on screening according to Fried's frailty phenotype (32)
- Insurance in the statutory health insurance (legal basis is a selective contract according to §140a SGB V with the insurance company; alternative legal basis after approval by the project management and project funding agency is a treatment contract according to §630a BGB if no selective contract exists)

Exclusion criteria:

- Severe cardiological disease (NYHA IV)
- Pulmonary disease (Gold IV)
- Intra-cranial procedures
- Moribund patients (palliative situation)
- Language barrier
- Participation in another interventional rehabilitation study or an AMG or MPG study that has not been approved by the project management (exception: parallel participation in adjuvant therapy studies).
- No legally valid declaration of consent

2. Analysis sets

2.1 Definitions

There are two analysis sets defined for this study:

2.1.1 Full Analysis Set (FAS) / Intention to treat (ITT)

It is as complete as possible and as close as possible to the Intention-To-Treat principle, which includes all randomized patients as belonging to their randomisation arm, regardless of whether they refused therapy or whether other protocol violations are known. For the ITT population, at least data from the initial screening must be available. Patients, who do not fulfil the inclusion criteria, cannot be randomized. Patients who withdrew consent, have no initial LCD at baseline or with a protocol violation concerning eligibility are excluded from ITT analysis.

2.1.2 Per Protocol Population (PP)

The per-protocol population is a subset of the FAS and includes only patients who have no major protocol violation. Here, we follow the recommendations given in the ICH E9 R1 guideline (60).

Major protocol violations encompass any unauthorised alterations to the research study design and/or procedures that fall within the investigator's purview and deviate from the protocol approved by the IEC/IRB. These deviations have the potential to impact participant rights, safety, or well-being, as well as the completeness, accuracy, and reliability of the study data.

Furthermore, it includes ineligible participants who were included in the trial by mistake and those for whom the intervention or other procedure differed from that outlined in the protocol as described in the following paragraph or failure of the consent process.

The intervention group of the PP population is divided into two distinct subgroups based on adherence to the intervention and the receipt of surgery:

1. First PP population (prehabilitation): Patients who completed at least 50% of the prehabilitation intervention (≥ 15 therapy units, TU). This definition of at least 15 TU will be used to quantify the effect of the intervention on the outcomes.

The prehabilitation consists of 30 multimodal sessions (e.g., nutritional consultation, physical exercises). In addition to this primary definition, we will perform an exploratory analysis of adherence patterns by dividing the intervention into three chronologically ordered ten-session groups. This approach allows us to investigate the impact of the first and second ten sessions on the subsequent sessions (note that this definition is subject to change once information about the intervention group becomes available):

- Adherence during the first 10 sessions (at least 5 of 10 sessions)
 - Adherence during the sessions 11 to 20 (at least 5 of 10 sessions)
 - Adherence during the last 10 sessions (at least 5 of 10 sessions)
2. Second PP population (prehabilitation AND surgery): Patients who completed at least 50% of the prehabilitation and underwent surgery as outlined in the protocol.

The control group of the PP population includes only patients who received usual care. Patients – randomized to the control group but receiving the prehabilitation – are excluded from the PP analysis.

Similarly, the control group is also divided into two distinct PP subgroups:

1. Control group (usual care): Patients who received the usual care.
2. Control group (usual care AND surgery): Patients who received usual care and underwent surgery. Patients randomized to the control group but who received prehabilitation are excluded from the PP analysis.

To examine differences in efficacy and safety between groups with varying adherence in more detail, Inverse Probability Weighting (IPW) approaches will be used according to Robins and Finkelstein (61). The IPW approach generates a “pseudo population”, wherein outcome, study discontinuation, and death influencing factors (baseline confounders and post-randomisation factors) are adjusted for.

2.1.3 Safety Population (SAF)

For the analysis of safety (predefined serious adverse events), all randomized patients are included, regardless of whether they actually received a prehabilitation session (intervention group) or care as usual (control group). Patients are analysed based on the treatment arm they were assigned at randomisation.

2.2 Application

The primary analysis of primary and secondary hypotheses is based on FAS (ITT) including all subjects randomized irrespective whether they received the intervention or not.

Sensitivity analyses will be conducted to verify the ITT analysis results' robustness.

In a per protocol analysis, primary and secondary hypotheses are evaluated based on appropriate IPW approaches.

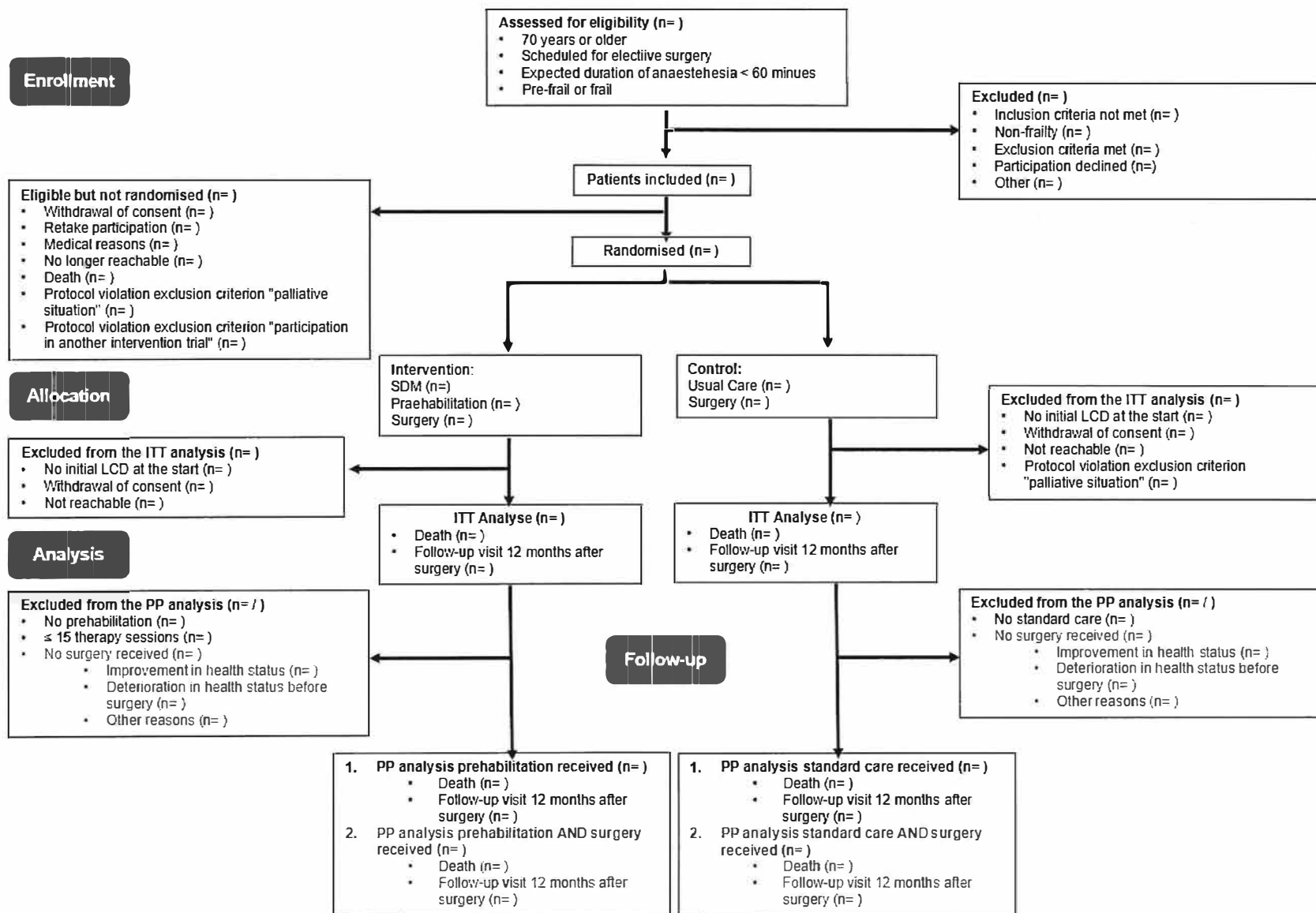


Figure 2: Patient inclusion for intention to treat (ITT) and per protocol (PP) analysis after randomization. SDM: Shared decision making.

2.3 Trial centres

Table 1: Overview of all trial centres, the number of patients randomized, and the level of care, categorized by maximum care/university hospitals and other hospitals according to bed count.

Hospital	Number of patients randomized	Level of care ¹
Charité Universitätsklinikum Berlin – Campus Virchow	398	Maximum care hospital/University hospital
Charité Universitätsklinikum Berlin – Campus Mitte	316	Maximum care hospital/University hospital
Sana Klinik Sommerfeld	186	Other (287 Hospital beds ²)
UKSH Lübeck	132	Maximum care hospital/University hospital
Charité Universitätsklinikum Berlin – Campus Benjamin Franklin	103	Maximum care hospital/University hospital
Klinikum der LMU München	73	Maximum care hospital/University hospital
Vivantes Auguste-Viktoria-Klinikum, Berlin	57	Other (598 Hospital beds)
Technische Universität München	47	Maximum care hospital/University hospital
Sankt Joseph Krankenhaus Berlin Tempelhof	34	Other (506 Hospital beds)
Caritas-Klinik Maria-Heimsuchung Berlin-Pankow	11	Other (245 Hospital beds)
Unfallkrankenhaus Berlin	9	Other (608 Hospital beds)
Dominikus Krankenhaus, Berlin Reinickendorf	4	Other (240 Hospital beds)
Uniklinik Hamburg Eppendorf	4	Maximum care hospital/University hospital
Krankenhaus Bad Oeynhausen	3	Other (240 Hospital beds)
Evangelisches Krankenhaus Hubertus	2	Other (202 Hospital beds)
Bundeswehrkrankenhaus	2	Other (367 Hospital beds)
Universtiätsmedizin Greifswald	1	Maximum care hospital/University hospital
Klinikum Frankfurt (Oder)	0	Maximum care hospital (775 Hospital beds)
Martin-Luther-Krankenhaus Berlin	0	Other (270 Hospital beds)
Vivantes Klinikum am Friedrichshain, Berlin	0	Maximum care hospital (1.012 Hospital beds)
Vivantes Wenckebach-Klinikum, Berlin	0	Other (443 Hospital beds)
total	1382	

The analysis of the multi-centre PRAEP-GO study will account for centre effects to address potential variations between centres, which may influence the outcomes. In this regard, the analysis considers the centre of recruitment as the reference point, which may differ from the centre where prehabilitation or surgery was performed, due to practical care considerations.

¹Hospital of maximum care (700 to over 1000 hospital beds). Other hospitals (<700 hospital beds)

² Source of bed count: <https://dkgev.deutsches-krankenhaus-verzeichnis.de/app/suche>

Furthermore, levels of care will be examined as a covariate in the sensitivity analyses.

Since different levels of care can involve different patient groups and treatment outcomes, centers are categorised according to whether they provide maximum care or not, with bed count serving as the proxy for this categorisation (Table 1).

The levels of care are considered as categorical variables in the models as follows:

- Hospital of maximum care: 700 to over 1000 hospital beds
- Other hospitals: <700 hospital beds (62)

The consideration of the interaction between the centre and treatment will be omitted.

3. Analysis variables

3.1 Study Variables and study schema

Details concerning the screening, inclusion, treatment and follow-up (post treatment) period together with the frequency and timing of relevant variables or assessments are displayed in Table 2.

Table 2: Schedule of screening, enrolment, intervention (Shared Decision-Making Conference (SDM) and prehabilitation), and assessments.

V7, V8, V10, V11, V12, V13, V14, V15, V16: telephone visits. V2, V6, V9, V18: office or home visits. V2: Baseline visit prior randomisation. V18: Last follow-up visit 1 year after surgery.

Assessments	V0	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18
Frailty status	X									X									X
Inclusion / exclusion criteria	X																		
Informed consent		X																	
Sociodemographic data																			
Age, sex	X																		
Level of education			X																
Weight, Height	X					X				X									X
Medical data																			
Diagnoses		X																	
Medication			X							X		X				X			X
Laboratory values			X																
Care dependency			X							X									X
Functional Assessments*			X		X	X				X									X
Falls			X			X	X	X	X	X	X	X	X	X	X	X	X	X	X
Cognitive function**	X	X								X									X
Depression			X							X									X
Psychosocial assessment			X							X									X
Autonomy, SDM-process				X															
Adherence to prehabilitation						X													
Medical record data			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

V0: Screening (Day 1)

V1: Study inclusion (Day 1/2)

V2: Baseline visit (Day 1/2)
 V3: Shared Decision-Making Conference (V1 +1-5 days)
 V4: Begin prehabilitation for intervention group ((22-25 days before surgery)
 V5: End of prehabilitation for intervention group (V4 + 21 days)
 V6: Hospital discharge after surgery (7-14 days post-surgery); exclusive medical data analysis
 V7, V8: 1-2 months post-surgery
 V9: 3 months post-surgery; personal or telephone visit possible
 V10-V17: 4-11 months post-surgery
 V18: 12 months post-surgery; personal or telephone visit possible

*2-Minute-Step-Test (2-MST), Peak expiratory flow (PEF), Timed Up and Go-Test (TUG), stair-climbing speed, handgrip strength, gait speed over 15 feet

** MiniCog™, Montreal Cognitive Assessment (MOCA), Trailmaking (TMT)-test part A and B

3.2 Explanatory variables / confounders (at baseline)

Among others, baseline characteristics are those criteria, which are used to determine initial prehabilitation goals at baseline:

1) Sociodemographic data

- Sex (male/female)
- Age
- BMI (height, weight)

1) Medical history

- Comorbidities (CCI)
- ASA status

2) Baseline cognitive vulnerability

- One of
 - i. Mini-Cog < 3 points OR (screening)
 - ii. ANT < 15 animals/minute
 - iii. TMT-B time > 97 s (70-74 y) / > 115 s (75-79 y) / > 133 s (80-84 y) / > 138 s (85+ y) / abortion / > 5 min
 - iv. MOCA < 26 points

AND

- Subjective Concern (MMQ) positive (for mild neurocognitive vulnerability)

AND

For severe neurocognitive vulnerability only: IADL/ADL restriction (Barthel Index < 100 points/ IADL Score < 8 points)

3) Frailty status (pre-frail / frail)

4) LCD at baseline

5) Index department

- Orthopaedics/ Trauma surgery
- General surgery
- Vascular surgery
- Cardiology

- Cardiac surgery
- Neurosurgery
- Ophthalmology
- Ear Nose Throat (ENT) Surgery
- Oral or maxillofacial surgery
- Gynaecology
- Urology
- Internal medicine
- Plastic surgery
- Quality of life measured by EQ-5D-5L and valuated for German population (EQ-5D-5L score) (63)

The following baseline characteristics will be reported per randomisation group and presented in a baseline table (Table 3).

Table 3: Baseline characteristics per randomisation group.

Characteristic	Prehabilitation Group (N = X)	Control Group (N = X)
Age – years		
Female sex		
Bodymass index [§]		
Fried-Items:		
weight loss (unintentional)		
Weakness		
Poor endurance; exhaustion		
Slowness		
Low activity		
Frailty Status by Fried:		
Pre-frail		
Frail		
Charlson Comorbidity Index		
Smoker (current)		
Alcohol use (AUCIT-C)		
ASA status [†]		
NBA score [#]		
Level of care dependency ^{&}		
0		
1		
2		
3		
4		
5		
Barthel-Index		
IADL [%]		
Quality of life ⁺		
EQ-5D-5L score		
EQ-5D-5L VAS		
Cognitive impairment or vulnerability		
Dementia risk		
Suspected mild or severe neurocognitive disorder		
Index department		
General surgery		
Gynecology		
Urology		
ENT		
Ophthalmology		
Orthopedic / Trauma surgery		
Cardiology		
Cardiosurgery		
Others (dermatology, etc.)		

* Values are presented as mean±SD, median [IQR] or n (%). The intention-to-treat population included all the patients who had undergone randomization.

§ The body-mass index is the weight in kilograms divided by the square of the height in meters

† American Society of Anesthesiologists Status

Score ranging from 0 to 100 to assess the level of care dependency in the German healthcare system (31)

& Level of care dependency is calculated from NBA (31) assessed by the study team to no (0), minor (1), considerable (2), severe (3), severest (4) impairment of independence or abilities, and severest impairment with special requirements for nursing care (5).

+ Quality of life is measured by the European Quality of Life 5 Dimensions 5 Level Version and valuated for German population (63,64)

% Measured by Lawton & Browdy Instrumental Activities of Daily Living Scale (39)

3.3 Explanatory variables / confounders between baseline (v2) and surgery

Table 4 describes the characteristics of the intervention, both as decided during the SDM conference and as actually conducted during the prehabilitation.

Table 4: Describes the measures of care for the shared decision-making conference and prehabilitation for the intervention group.

Planned Prehabilitation setting – no. (%)	
Out-patient	n (x%)
Day-clinic	n (x%)
In-patient	n (x%)
Planned prehabilitation components – no. (%)	
Training therapy	n (x%)
Nutrition therapy	n (x%)
Cognitive and psychological support	n (x%)
Other or mixed support	n (x%)
Performed Prehabilitation setting – no. (%)	
Out-patient	n (x%)
Day-clinic	n (x%)
In-patient	n (x%)
Performed prehabilitation components – no. (%)	
Training therapy	n (x%)
Nutrition therapy	n (x%)
Cognitive and psychological support	n (x%)
Other or mixed support	n (x%)
Prehabilitation compliance – median [IQR]	
Total number of prehabilitation sessions per patient	XX [XX – XX]
Total number of prehabilitation sessions (supervised and unsupervised)	XX [XX – XX]
Supervised prehabilitation sessions	XX [XX – XX]
Duration of prehabilitation total (d)	XX [XX – XX]
Interruption (d)	XX [XX – XX]

ADL, activities of daily living; IQR, interquartile range;

3.4 Explanatory variables / confounder at surgery

Figure 3: Probability curves to get a surgery (time until surgery).(time to surgery (KM, median, quartiles)).

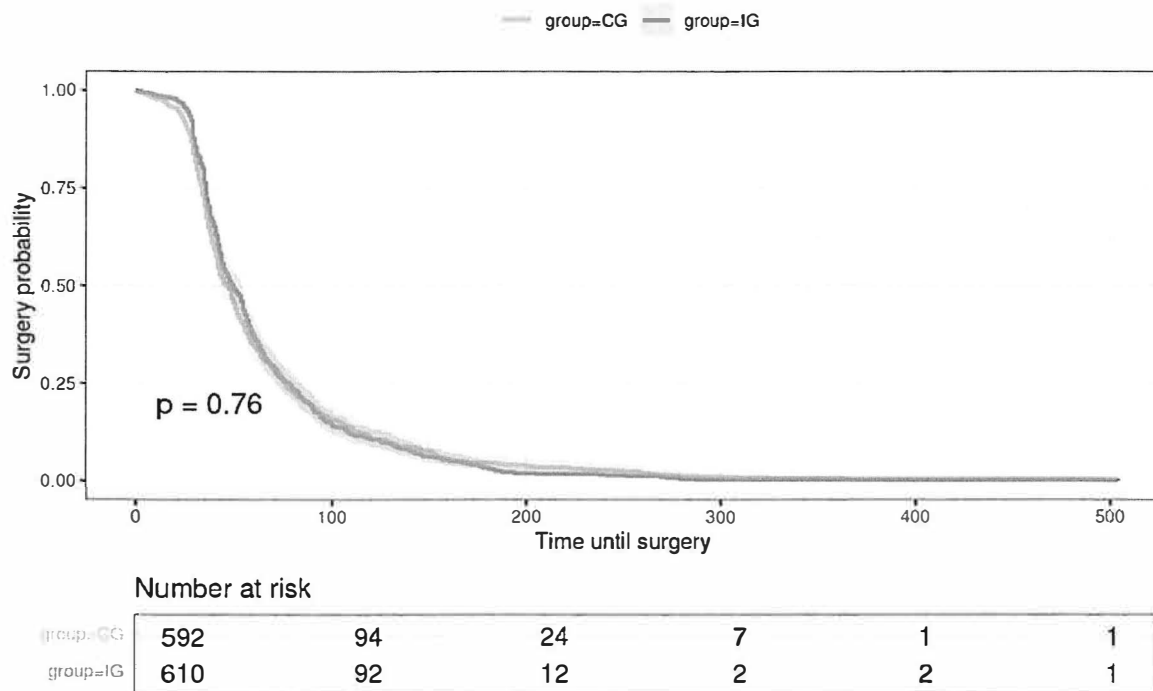


Figure 3: Probability curves to get a surgery (time until surgery).

The following data are used to describe the surgery and measurements taken during surgery:

- Type of operation (open / laparoscopic / arthroscopic / interventional)
- Duration of anesthesia [minutes]
- Incision-to-suture time [minutes]
- Cardiovascular risk associated with surgery: KRI
- Surgical risk scale (Sutton R, Bann S, et al., 2002 (65))

Table 5: Describes the situation peri-surgery.

	Intervention Group (Prehabilitation) (n=)	Control Group (Usual care) (n=)	p-value
Time between inclusion and surgery (d)			
Type of anaesthesia			
General anaesthesia			
Spinal anaesthesia			
Regional anaesthesia			
Local anaesthesia			
Other			
Type of surgery			
open			
laparoscopic			
arthroscopic			
interventional			
Duration of anaesthesia (Min.)			
Duration of surgery (min)			
Tumor surgery (y/n)			
Complexity of surgery (British United Provident Association Score)			
Surgical risk scale			

3.5 Explanatory variables / confounders after surgery

Possible confounders for the 12-month follow-up are:

- ICU LOS
- Postop. Complications by Clavien-Dindo
- Charmi Index at Hospital Discharge
- Hospital LOS
- Rehabilitation (yes/no)
- Duration of rehabilitation (days)
- Type of rehabilitation (inpatient, day-clinic, outpatient)
- Occurrence of falls
- New aids (Hilfsmittel) necessary
- NBA at 3 months
- Hospital admission after index admission (yes/no)

- Hospital admission after index admission (days)
- New moving to a nursing home

Table 6: Explanatory variables measured after surgery for intervention (prehabilitation) and control (usual care) groups.

	Intervention Group (Prehabilitation) (n=)	Control Group (Usual care) (n=)	p-value
ICU Admission (yes/no)			
ICU Admission (in hours)			
Rehabilitation (yes/no)			
Duration of Rehabilitation (in days)			
Type of Rehabilitation			
Inpatient			
Partial Inpatient			
Outpatient			
Moving to a nursing home			
Charmi Index at hospital discharge			

3.6 Primary endpoint

The primary endpoint is change in Level of Care Dependency (LCD) from baseline (visit v2) to one year postoperative (visit v18).

The classification into LCD levels is based on the NBA (Neues Begutachtungsassessment) score (see secondary endpoint) using a points system. The number of points reflects the extent to which independence is impaired or abilities are lost. The care levels are measured on an ordinal scale and are divided into five levels (Table 7):

- Care Level 1 - Minor impairment of independence (more than 12.5 and less than 27 points)
- Care Level 2 - Significant impairment of independence (27 to less than 47.5 points)
- Care Level 3 - Severe impairment of independence (47.5 to less than 70 points)
- Care Level 4 - Most severe impairment of independence (70 to less than 90 points)
- Care Level 5 - Most severe impairment of independence with special requirements for nursing care or death (90 to 100 points)

Table 7: Determination of the Level of Care Dependency (LCD) and corresponding points of the NBA-score (Neues Begutachtungsassessment).

Level of Care Dependency (LCD)	No	1	2	3	4	5
NBA points	< 12.5	12.5 - <27	27 - <47.5	47.5 - <70	70 - <90	90 - 100

The LCD is calculated from the weighted added point values of 5 modules of the NBA (1 to 6, whereby for module 2 and 3 only the higher point value is included in the calculation).

The primary endpoint is calculated as follows:

$$\Delta_{LCD} = LCD_{V18} - LCD_{V2}$$

The LCD is determined by the total score from the NBA, which forms the basis for assigning the appropriate care level. Changes in the NBA score are used as a secondary endpoint and are further described in section 3.7.1:

$$\Delta_{NBA} = NBA_{V18} - NBA_{V2}$$

No change and an improvement from baseline to one year postoperative in the LCD are considered a success, while a deterioration in the LCD is interpreted as a failure.

Table 8 describes the changes in the LCD, which can range from -2 to 2. Negative values indicate a reduction in the level of care (improvement), 0 means no change, and positive numbers indicate an increase in the level of care (deterioration). We further assume that a theoretical improvement by two care levels will not be observed in practice.

Table 8: Change in the Level of Care Dependency (LCD) for the intervention and control group.

		LCD Follow-Up 12 months after surgery					
		0	1	2	3	4	5
LCD Baseline	0	0	1	2	2	2	2
	1	-1	0	1	2	2	2
	2	-2	-1	0	1	2	2
	3	-2	-2	-1	0	1	2
	4	-2	-2	-2	-1	0	1
	5	-2	-2	-2	-2	-1	0

3.6.1 Composite endpoint: Incorporate death

The composite endpoint is used to consider both mortality and care dependency when assessing the intervention. In case of death, the worst level of care dependency (5) is assumed. This means that the care level representing the highest need for nursing care and support is used for analysis when the patient passes away from the time of enrolment until the last follow-up visit one year after the surgery. This procedure will be part of the ITT analyses to cope with trial participants, who died.

3.7 Secondary endpoints

3.7.1 Change in NBA-total-Score

The total NBA sum score is determined by means of the weighted values of each module of the new care needs assessment (Neues Begutachtungsassessment, NBA) (31). It is a standardized and reliable assessment instrument to determine the need for care according to SGB XI in Germany. It is measured at the metric scale level – ranging from 0 (no need for care) to 100 (maximum need for care).

The modules seven and eight are not included into the calculation of the NBA-total-score.

Each item differentiates five levels: (0) independent, (1) minor impairments, (2) significant impairments, (3) severe impairments, or (4) complete / extensive loss (of independence, etc.).

The total sum score is calculated from the weighted added point values of 5 modules (1 to 6, whereby for module 2 and 3 only the higher point value is included in the calculation. The degree of care is derived from the continuous scale, which is divided into five care levels (see primary endpoint LCD). A sixth care level – no care level – is added if there are less than 12.5 points reached (Table 7).

$$\Delta_{NBA} = NBA_{V18} - NBA_{V2}$$

The NBA consists of eight modules (Table 9):

Table 9: Modules of the New care needs assessment (NBA) and its weighting to the total sum score.

Module LCD	Weighting for total score
1 Mobility	10%
2 Cognitive and communicative capacities	15% (together with 3)*
3 Behaviour and mental problems	15% (together with 2)*
4 Self-care (feeding, dressing, washing, eliminations)	40%
5 Dealing with the requirements of an illness and treatment	20%
6 Structuring everyday life and social contacts	15%
7 Activities outside the home	-
8 Housekeeping	-

*Only the module (2 or 3) with the higher point value is included in the calculation of the total sum score with 15%.

3.7.2 Change in NBA domains

Each of the aforementioned NBA domain (1 to 6) is used individually as a dependent variable.

3.7.3 Frailty

Fried frailty phenotype is defined as the presence of three or more of the following five criteria:

1) weight loss, 2) exhaustion, 3) low physical activity, 4) slow gait speed, and 5) weakness.

- 1) Weight loss is defined as having more than 4.5 kg unintentional weight loss over the preceding 12 months.
- 2) Exhaustion is defined as having a positive response to at least one of the following questions regarding endurance and energy: a) "Everything I did was exhausting." b) "I couldn't get started".
- 3) Low physical activity is defined as regular physical activity based on the metabolic equivalent table of less than 3.
- 4) Speed criterion is met if the 4.57 meter (15 ft) walking at an individual's pace is slower than standard cut-points based on sex and height (male: height ≤ 173 cm: < 7 sec.; height > 173 cm: < 6 sec.; female: height ≤ 159 cm: < 7 sec.; height > 159 cm: < 6 sec.
- 5) Grip strength is assessed using a handheld dynamometer - the best of three attempts will be counted.

One point is assigned for the presence of each feature of the phenotype (1-5), resulting in a score from 0-5: Not Frail (0), Pre-Frail (1-2), and Frail (>2).

3.7.4 Neurocognitive Disorders (NCD)

1. Trail Making Test (TMT) as a Surrogate Cognitive Endpoint

The Trail Making Test (TMT) is a neuropsychological assessment used to measure cognitive functions. We use TMT-B – requiring alternating between numbers and letters (1, A, 2, B, etc.) – to assess cognitive flexibility and executive function.

2. MiniCog

The MiniCog is a brief screening tool for cognitive impairment that combines a short-term memory recall test and a simple clock-drawing test.

3. Animal Naming Test (ANT)

The ANT is a semantic fluency test that consists of listing as many names of animals as possible in 1 minute (Agarwal A, Taneja S, et al. 2020 (66)).

4. Montreal Cognitive Assessment (MOCA)

The MOCA measures the dementia risk in participants post-surgery. It provides a detailed cognitive profile that helps in identifying individuals at risk of developing dementia by assessing multiple cognitive domains. A lower MOCA score may indicate a higher risk of dementia.

5. Frequency of new mild or major neurocognitive disorder

This outcome measures the incidence of new cases of mild or major neurocognitive disorders (NCDs) in participants after surgery.

3.7.5 Health-related quality of life

Health-related quality of life (HRQoL) is measured using the EuroQol Five-Dimensional Five-Level Questionnaire (EQ-5D-5L), including the Visual Analogue Scale (EQ-VAS) assessing self-rated health. It comprises five dimensions (with one five-point Likert-scaled item each): 1) mobility, 2) self-care, 3) usual activities, 4) pain/discomfort and 5) anxiety/depression ranging from no problems (1) to extreme problems (5). The EQ-5D-5L utility score ranges between death (−0.661) and perfect health (1) and is calculated based on the value set devised by Ludwig et al., 2018. In the EQ-VAS, the self-rated health is visualized on a continuous VAS with endpoints labelled the worst imaginable health (0) and the best imaginable health (100).

3.7.6 Functioning and disability

Functioning and disability is measured using the World health organization disability assessment schedule 2.0 (WHODAS 2.0) (5). It contains 12 items ranging from no difficulty (1) to extreme difficulty or cannot do it (5), with higher scores suggesting worse function.

3.7.7 Mortality

Mortality is defined as the number of deaths in the group of all randomized patients within the time between enrolment and end of follow-up period one year after surgery.

3.7.8 Drop-outs / Censoring

The drop-out rate includes the number of drop-outs and censoring in a group of all randomized patients within the period from enrolment to the end of the follow-up period, one year after surgery. This encompasses all patients who leave the study before it is completed for any reason except death (e.g., lost to follow-up, personal choice, medical reasons, etc.).

3.7.9 Change in functional status

Degree of independence in activities of daily living (ADLs) is assessed with the Barthel Index (38). It comprises ten domains measuring a person's ability to complete ADLs: 1) Feeding, 2) Bathing, 3) Grooming, 4) Dressing, 5) Bowel control, 6) Bladder control, 7) Toilet use capability, 8) Transfers bed-to-chair-and-back, 9) Mobility on level surfaces, and 10) Stair negotiation. The maximum achievable score is 100 points - the higher the score, the more independent the patient is in completing the measured ADLs. Higher scores also indicate the patient is more likely to return home, with varying degrees of assistance, following hospital discharge. Barthel's index score can be divided in five categories: independent (80–100 points), minimally dependent (60–79 points), partially dependent (40–59 points), very dependent (20–39 points), and totally dependent (<20 points).

3.7.10 Change in activities of daily living

For instrumental activities of daily living (IADL) the Lawton and Brody's index is used (39). It comprises eight domains of functions and independent living skills: 1) Using the telephone, 2) Shopping, 3) Meal preparation, 4) Housekeeping, 5) Laundry, 6) Transportation (driving or using public transportation), 7) Responsibility for medications, and 8) Responsibility for finances (e.g., managing money, paying bills). Women are scored on all eight areas of function; historically, for men, the areas of 3) meal preparation, 4) housekeeping, 5) laundry are excluded. The sum score ranges from low function/dependent (1) to high function/ independent (8) for women, and 0 through 5 for men.

3.7.11 Probability of moving to a nursing home

To calculate the probability of moving into a nursing home within a year after surgery, we only consider those patients who were not in a nursing home before surgery. The following data are used:

- a. Total number of patients before surgery (N)
- b. Number of patients who were already in a nursing home before surgery (B)
- c. Number of patients who moved into a nursing home within a year after surgery (M)
1. We calculate the number of patients who were not in a nursing home before surgery: $N' = B - M$.
2. We calculate the probability for patients who moved into a nursing home after surgery, excluding those who were already in a nursing home before surgery: $P = (M/N')$

3.7.12 Duration of rehabilitation or follow-up treatment

The total duration of rehabilitation/follow-up treatment is defined as the time period over which rehabilitation was provided from surgery until the last follow-up visit one year later (v18), or rehabilitation duration, measured in days.

If rehabilitation is conducted ambulatory, the number of prescribed therapy sessions for the entire study period can be specified.

3.8 Safety/Tolerability

Exercise adherence and composition during prehabilitation. For this, the responsible therapists will document all exercise and therapy sessions. Patients will fill out a training diary to document all performed unsupervised sessions. Predefined SAEs were documented as described in Table 10.

Table 10: Description of safety parameters in the PRAEP-GO study.

	Intervention Group (Prehabilitation) (n=)	Control Group (Usual care) (n=)
Serious adverse events (SAEs), Falls to the ground during basic assessment or home visit, n (%)		
Patient with a least one SAE, n (%)		
Severity of SAE, n (%)		
Grade 1: Mild		
Grade 2: Moderate		
Grade 3: Severe		
Grade 4: Life-threatening		
Grade 5: Death		
Results, n (%)		
Death		
Inpatient hospitalization or prolongation of existing hospitalization		
Medical/surgical procedure		
Admission to ICU		
Falls		
Life-threatening disease		
Persistent or significant disability/incapacity		
First aid		
Unknown/not assessable		
Other		
Causal relationship of SAE/AE with study intervention , n (%)		
Confirmed		
Probable		
Possible		
Unlikely		
Not assessable		

4. Handling of missing values and outliers

4.1 Missing values

Missing values in the response variables are generally not imputed, except when the assessment manual provides specific instructions for imputation (e.g. in the case of the WHODAS and MMQ). Inverse Probability weighting may be helpful in this situation and will be used as the main strategy to cope with missing data. Missing response values can be imputed by creating a pseudo population, in which no missing values occur. This can be achieved by calculating IPWs for missing values and include these in the outcome model as weights (for details see section 5.8). Missing values in explanatory variables may be imputed by multiple imputation (MICE) if applicable according to the evaluation manual of the respective assessment. If possible, inverse probability weighting may be applied.

5. Statistical analyses / methods

5.1 Subject disposition

Eligible participants are defined as any participants who signed the written informed consent.

Randomized participants are defined as all participants who had signed the informed consent and who were randomized either in the intervention or in the control arm.

For participant study status, the total number of participants will be presented in the study report using a CONSORT flowchart diagram (Figure 2):

- Participants assessed for eligibility (i.e., fulfilled the inclusion criteria)
- Screen failure participants and reasons for screen failure (i.e., fulfilled an exclusion criterion)
- Randomized participants
- Participants who did not complete the study until v18 and reasons for study discontinuation
- Participants included in primary outcome analysis

5.2 Demography and baseline characteristics

Continuous data are summarised using the mean, standard deviation (SD), median, minimum, and maximum – depending on the data distribution – for each intervention arm.

Categorical and ordinal data are summarized using the number and percentage of participants in each intervention arm.

Parameters are summarized based on the randomised population within each intervention group they were assigned to. Parameters are summarised by intervention arm and overall using descriptive statistics.

No specific description of the efficacy parameters, safety parameters, and quality-of-life parameters are provided at baseline. If relevant, baseline values are described along with each efficacy, safety, or quality-of-life analysis, respectively.

5.3 Prior or concomitant medication and diseases

Data on polypharmacy is collected and analysed for the randomised population. Polypharmacy is defined as the use of five or more different medications by a single participant. Reports detail how many participants exhibited polypharmacy before randomisation and during the study. The results are reported by the intervention group and presented in tables showing the frequency of polypharmacy. These tables include the number of participants who are taking five or more medications. Additionally, changes between baseline and visit 18 are analysed to identify potential effects of the interventions on polypharmacy.

5.4 Exposition to treatment/Compliance

The adherence to prehabilitation is divided into three periods, based on 30 prehabilitation sessions. Each period consists of 10 sessions, and compliance for each period is defined as follows:

- First 10 sessions: Patients must have attended at least 5 out of 10 sessions to be considered compliant.
- Sessions 11 to 20: Similarly, patients must attend at least 5 out of 10 sessions to maintain compliance.
- Last 10 sessions: The same criterion applies: at least 5 out of 10 sessions.

It is assessed whether patients have attended the minimum number of 5 sessions in each of the three defined periods. Adherence for each period is calculated as a percentage. For example, if a patient attends 6 out of 10 sessions, this corresponds to an adherence of 60% for that period. By comparing the adherence values of the first 10 sessions with those of the subsequent periods (sessions 11 to 20 and 21 to 30), the impact of early sessions on later compliance can be assessed.

The aim is to measure and evaluate adherence to prehabilitation in three periods and to determine whether patients meet the established criteria for each period. This allows for a detailed analysis of compliance throughout the entire duration of prehabilitation.

5.5 Primary analysis - ITT

The primary analysis is the ITT analysis aiming to analyse the data according to the participants' initial group allocation. Differences between groups will be compared using the Wilcoxon-Mann-Whitney rank sum test with stratification. To determine the discriminatory ability of intervention versus control group, the area under the curve stratified according to the baseline LCD will be conducted.

5.5.1 The ordinal aspect: Wilcoxon-Mann-Whitney rank sum test by Brunner et al. (67,68)

The possibilities for changes in LCD between baseline and follow-up (twelve months after surgery or proxy) are shown in Table 8. Changes in LCD can range from -2 to 2. Zero indicates no change, negative values indicate a reduction in the LCD (improvement), and positive numbers indicate an increase in the LCD (deterioration). We assume that an improvement of more than two care levels is not observed in practice. This type of Wilcoxon test will consider baseline LCD by stratification.

5.5.2 The binary aspect

Additionally, the change in LCD is evaluated binarily: zero or negative values indicate an improvement. A stable condition is considered a positive event and thus an improvement, while positive values of the change indicate a deterioration.

5.5.3 Stratified area under the curve (ROC-GLM)

The primary outcome of this study is the change in LCD from baseline to study completion. The ability to differentiate LCD based on group assignment (intervention or control group) will be assessed using a Receiver Operating Characteristic (ROC) curve. Additionally, the area under the ROC curve (AUC) will be calculated. Since a subject's baseline LCD may affect the change in LCD over the course of the study, a stratified analysis, as described by Pepe and colleagues, will be applied (69). She describes a method to estimate the ROC and AUC by using a logistic regression model. Independent variables, like baseline LCD, can then be used as stratifying variables.

5.6 Sensitivity analysis

In the sections (5.6.1 und 5.6.2) the sensitivity analyses regarding the ITT analysis will be outlined.

5.6.1 Hierarchical endpoint (play the winner)

Composite endpoint analyses are necessary, if an endpoint of interest cannot be observed, because another event occurred before. In this trial the primary endpoint of change of the LCD cannot be observed, if the study subject dies before study completion. Pocock and colleagues introduced a hierarchical endpoint approach - the win ratio- to deal with such composite endpoints in clinical trials (70). The trial subjects of the control group and intervention group are matched pairwise according to their risks. The risks are determined by propensity scores.

The subject of the intervention group is declared as "winner", if the worsen of the LCD ($\Delta_{LCD} = LCD_{V18} - LCD_{V2}$) is smaller than that of the comparator (number of "winners" of the intervention = N_A). N_B is the number of "winners" within the control group. In case of equal LCD change, the pair is tied. If LCD_{V18} could not be assessed due to death before V18 (study completion visit), the subject, who (first) died is marked as "loser". Accordingly, N_C is defined as the number of pairs in which the subject of the control group dies (first) and the intervention subject is the "winner". Correspondingly, N_D is defined as the number of pairs, if the subject of intervention group dies first. Note, that on a hierarchically point of view, if one subject of a pair dies before study completion, the other is always the winner since the LCD change cannot be assessed any more.

The "win ratio" R_W is defined as:

$$R_W = \frac{N_W}{N_L} = \frac{p_w}{1 - p_w},$$

where:

- $p_w = \frac{N_W}{N_W + N_L}$

- $N_W = N_A + N_C$
- $N_L = N_B + N_D$

95% confidence intervals can be calculated as follows:

$$CI_{95\%} = \frac{p_L}{1-p_L}, \frac{p_U}{1-p_U},$$

with:

$$p_L, p_U = p_w \pm 1.96 \left[\frac{p_w(1 - p_w)}{N_W + N_L} \right]^{\frac{1}{2}}$$

A test statistic will be calculated as follows:

$$z = \frac{p_w - 0.5}{\left[\frac{p_w(1 - p_w)}{N_W + N_L} \right]^{\frac{1}{2}}}$$

The p-value will be derived by comparing the value of z with the standard normal distribution.

5.6.2 On the score level: LMM: Score difference ~ baseline score + treatment

Additionally, to the analysis of the change of the LCD the change in NBA score (Δ_{NBA}) as defined in section 28 will be examined.

Linear mixed model (LMM):

To determine the effect of the intervention on the linear Δ_{NBA} score (defined in section 0), we will apply a linear mixed model (LMM) and include the baseline NBA score and intervention assignment as covariates. Additionally, we include the time until V18h to consider subject specific time intervals between baseline and V18h. To adjust for the multi-center structure of the collected data, a random intercept and slope for the centers will be included in the LLM. Centers with less than 30 subjects will be summarized to one cluster. For the main analysis, the linear predictor is defined as:

$$\eta_i = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) + \beta_2^{Intervention} + \beta_2^{NBA_{baseline}} + \beta_3^{time}$$

Mean structure:

The two groups (control and intervention) will be dummy-coded, whereas the control group will be used as reference category. For example, the coefficient β_2 expresses the difference between the slopes of the logarithm of Δ_{NBA} for a patient randomized to the intervention group as compared to the control group.

Random effects:

$b_i = (b_{0i}, b_{1i})'$ are random intercepts and random slopes to account for the variation among subjects both within centres as well as in the slopes over time (i.e., random variation in slopes through time across groups specified by centres). b_i is assumed to follow a bivariate normal distribution with mean zero and an unknown precision matrix.

Underlying key assumptions:

MAR. E.g., drop-out not due to death assuming MAR means that in some sense is predictable from the observed data

linear time trend

normally and identically distributed residuals

covariance structure: An unstructured covariance pattern will be used to model the within-centre errors

variance components: Random effects precision not depending on intervention group; correlated random effects

Covariance structure

Random intercept for centre, random slope for time.

Random effects (b_{0i}, b_{1i}) are assumed to be normally distributed and correlated.

Strategies to improve convergence or to avoid convergence limitations/ Convergence issues:

The main model will be analysed using a restricted maximum likelihood (REML)-based on multiple centre approach. In case the LMM defined above fails to converge, alternative more parsimonious covariance/correlation structures will be used in place of an unstructured one. Generally, the *first-to-converge approach* (71) (in the sense that the first structure in the ever-more parsimonious set to yield convergence) will be applied to avoid model building and hypothesis testing from the same data.

If the LMM fails to converge or in case of computational difficulties (e.g., numerical instabilities), we proceed as follows: We fit the model first without the random slope and then without the random intercept. In case of a skewed distribution of the residuals, a robust standard error (Huber sandwich estimator (72)) will be used.

It is also of interest to apply GEE modelling if appropriate.

5.7 Secondary analyses -ITT

Secondary endpoints will be analysed as described for the primary endpoints (section 5.5) and additionally as follows.

5.7.1 Efficacy

The efficacy of the prehabilitation will be determined by the IT and PP analyses strategies of the primary and secondary outcomes.

5.7.2 Safety/Tolerability

Adverse Events

Adverse events (AEs) are defined as any untoward medical occurrence in a patient or clinical investigation subject administered an investigational intervention. AEs do not necessarily have a causal relationship with the intervention (adapted from the Note for Guidance on Clinical Safety Data Management: Definitions and Standards for Expedited Reporting (CPMP/ICH/377/95 July 2000) (73). The underlying mortality and morbidity rate of patients participating in the trial is high. Common aberrations in laboratory values, signs and symptoms due to the underlying illness and the impact of standard therapies will not necessarily constitute an adverse event unless they are considered to be of concern or related to the study or the intervention in the investigator's clinical judgement. In this case an Adverse Event should be documented.

Adverse events will be collected from enrollment in the trial to 48h post cessation of PRÄP-GO protocol only. The patient should be followed until the event is resolved or explained (max. until day 365 after elective surgery). Frequency of follow-up is left to the discretion of the investigator.

Adverse Events will be recorded in the eCRF Serious adverse event

Serious adverse events

If an SAE is unexpected (i.e., not described in the protocol or patient information sheet or patient study brochure) and thought to be related to study treatment and is "fatal" or "life threatening" the site Principal Investigator must report the event in the eCRF within 72 hours of knowledge of the occurrence of the event. Serious adverse events will be reported up to 48 hours after the last study visit including the following:

If they occur during prehabilitation:

- Falling to the floor
- Cardiac Arrest
- Unscheduled hospitalization
- Unscheduled intervention (e.g. heart catheter)
- Unscheduled surgery

If they occur during study visits in the hospital or at home (from start of visit up to 48h after the visit):

- Falling to the floor

The project manager shall report the SAEs to the DSMC in tabulated form every 3 months with the scheduled report. Any recommendations from the DSMC will be reviewed by the MC and, if necessary, disseminated by the project manager to other participating sites. The study protocol mandates that events that are considered as study endpoints and are part of the natural history of the primary disease process (i.e. are not related to study intervention or procedure) will be systematically collected for all patients (deaths and causes of deaths). The DSMC will review these in the pre-planned safety analysis. However, deaths that the site investigator considers to be related to study treatment or study procedures will be reported as a SAE.

5.8 Per Protocol Analysis

It has been shown, that baseline and/or post-randomisation factors are able to influence the effect of an intervention on the trials' objectives, even if the randomisation is performed properly (60,74). Post-randomisation events (also called intercurrent events, ICEs) like censoring and/or death (as competing risk) are affected, as well. This means that the exchangeability between the intervention and control groups cannot be assumed after the results of the randomisation process are known. In this trial we identified the following post-randomisation events, which are assumed to be influenced by post-randomisation factors:

1. Intervention adherence: subjects may behave differently, after they know the assignment to intervention or control group. A subject may be willing to quit the trial, because they were assigned to the intervention group, which is linked to time consuming and exhausting physical exercises (Details in section 5.8.3 and Figure 4).
2. Probability of death: baseline characteristics, the intervention, and severity of the surgery could have influence on the probability to die (Details in section 5.8.4 and Figure 5). Note, death is a competing risk, because the trials' objective cannot be observed after a subject died.
3. Probability of study discontinuation (censoring): baseline characteristics, the intervention, and severity of the surgery could have influence on the probability to discontinuation of the trial (Details in section 5.8.4 and Figure 6).

To adjust for baseline and post-randomisation factors, IPW (described in section 5.8.1) according to Robins and Finkelstein (61) (inverse probability weighting), and Young (75) (competing risk) will applied. These approaches are a two-step procedure, in which first, the IPW weights are estimated ("*IPW models*" with the influencing baseline and, when available, post-randomisation factors as dependent variables). In the second step, an "outcome model" will be estimated, with the trial's primary and/or secondary outcomes as the dependent variable and the randomisation result as the independent variable. The estimated IPW values will be used as weights in this model. The resulting estimate of the effect of the intervention then be unbiased and robust against post-randomisation events/factors. The idea behind this approach is to generate a pseudo population, in which the influencing post-randomisation event would not occur (for example: What is the effect of the intervention, if nobody would have died during the duration of the trial). This approach has some assumptions, which have to be met: exchangeability (between the control and the intervention group), positivity (probability of being exposed to the treatment/intervention within the strata of the covariates is unequal zero) and consistency (an individual A) of the control group would experience the same intervention effect of an individual B) out of the intervention group, if individual A) had received the intervention). These assumptions are essential to (i) avoid extreme IPW values within a stratum with low amount of data (positivity), (ii) be sure that the effect is observed due to the intervention and not due to confounding factors (exchangeability), and (iii) to construct a pseudo population, in which study protocol violations (e.g., intervention adherence) or dropped-out subjects are replaced by subjects with similar characteristics (consistency). We presume, that these assumptions are fulfilled in this clinical trial.

To examine the effect of elapsed time between baseline and surgery, effects of the intervention are estimated separately for patients whose surgery occurred before the median time between baseline and surgery, and for those whose surgery occurred afterwards.

5.8.1 Inverse Probability Weighting (IPW)

Inverse probability weighting (IPW) will be conducted to assess the likelihood of undergoing intervention adherence, censoring, and death. To compute the IPWs and to consider the longitudinal data structure at hand, three pooled logistic regression models with intervention adherence, death, or censoring as outcome variable will be estimated. Time in trial and outcome specific baseline characteristics, post-randomisation factors, and time varying confounders will be included as covariates. The probability for death, censoring, or intervention adherence can be derived by applying the inverse logit function to the score resulted of the exponentiated linear predictor of the pooled logistic regression models. Non-stabilized IPWs can now be derived by one divided by the calculated probabilities:

$$\frac{1}{P(O|X_0, X_T, X_P, T)}$$

where:

$P(O|X_0, X_T, X_P, T)$ is the probability of regression outcome O (intervention adherence, death, and censoring) given baseline confounders X_0 , post-randomisation factors X_P , time varying confounders X_T , and time T. In this PP analysis we will derive stabilized IPWs to avoid outliers within the individual IPW values. For that, the numerator of the formula above will be replaced by the probability to observe the outcome given time:

$$\frac{P(O|T)}{P(O|X_0, X_T, X_P, T)}$$

To estimate $P(O|T)$, pooled logistic regression models will be used. This method aims to elucidate whether the post-randomisation confounders influence the outcome by affecting the probability of death, censoring, and/or intervention adherence.

Note, considering lost to follow-up or intervention adherence, the trials' outcome (defined in sections 0 and 3.7) could potentially be observed, but we have no observed data. In such cases, the IPW approach can be used. Contrary, considering death as trial discontinuation event, the trial outcomes cannot be observed anymore (competing risks). Young and colleagues proposed an estimand, the direct effect, with which this competing risk situation can be handled and showed, that in this case, the IPW approach is appropriate, as well. In sections (5.8.3 and 5.8.4) details about the implementation of the IPW approach is outlined (75).

5.8.2 Directed Acyclic Graphs (DAG)

Directed Acyclic Graphs (DAGs) are graphical models widely used in statistics and causal inference to visually represent causal relationships between variables. In a DAG, nodes represent variables of interest, and directed edges between nodes indicate the direction of causation, without any cycles. This acyclic property ensures that causal relationships are clearly defined, helping researchers to model and understand complex systems where variables interact in causal pathways. DAGs play a crucial role in guiding causal inference by providing a visual framework to hypothesize, test, and refine causal theories based on observational or experimental data. They are instrumental in identifying potential confounders, determining variable selection for analysis, and validating causal assumptions (76). We are interested in the causal effect of the intervention $A \rightarrow Y$. Confounding is indicated by a so-called backdoor path. A backdoor path in a DAG is a path between an intervention (A) variable and an outcome variable (Y) that includes an arrow (edge) entering into the treatment variable

(or any of its descendants). This path allows for potential confounding because it creates an association between the treatment and the outcome that is not due to the direct causal effect of the treatment on the outcome. Such path should be identified and blocked. One way to block backdoor paths is to use the IPW approach.

Our DAGs (Figure 4, Figure 5, Figure 6) functions as a structural tool to enhance the visualization of relationships among the variables central to the causal relationship under investigation: baseline and time varying confounders (L_k), intervention (A_k), and outcome (Y). Furthermore, our DAG incorporates a vector of unmeasured covariates (U) that impact both the outcome and treatment. Additionally, Death (D), Censoring (C), and Surgery (S) are displayed in the DAGs. To avoid clutter, only important arrows are shown. We depend on the strong and untestable assumption that the effects of all unmeasured confounders (U) are appropriately captured by the measured covariates (L_k). Furthermore, we assume, that only baseline variables impact the intervention and deaths or censoring before the surgery (Figure 4, Figure 5, Figure 6).

5.8.3 IPW for intervention adherence

The DAG of Figure 4 shows the effect of the intervention to the outcome. To adjust for confounders, which are identified by backdoor paths, the red arrows must be blocked. To block the backdoor paths, the baseline variables (L_0) must be included in the IPW model. We listed the baseline variables in Table 3. Since we have a longitudinal data structure at hand (the intervention is investigated in three ten-session groups.) and the intervention (the outcome for the IPW model) is a dichotomous variable we will use a pooled logistic regression model with time as describing variable. Read section 2.1.2. for details about the definition of the intervention. To avoid model misspecification, quadratic terms and splines for baseline and time variables will be used, if necessary. The IPWs values will be derived as described in section 5.8.1.

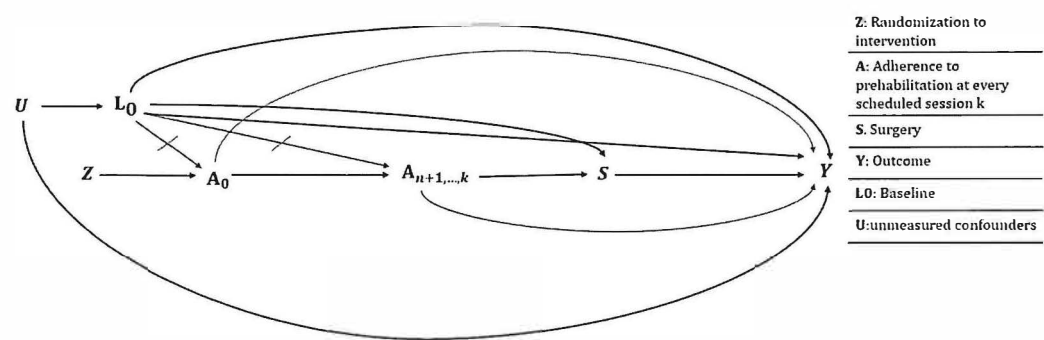


Figure 4: DAG considering adherence to treatment.

5.8.4 IPW for death/censoring

Death and censoring of subjects can occur during the whole duration of this clinical trial. We assume, that only baseline variables (defined in

Table 3) are influential for deaths or censoring before surgery (D1/D2 and C1/C2 in Figure 5 and Figure 6). As before, the backdoor paths must be blocked, indicated by crossed-out arrows in Figure 5 and Figure 6. For the interval after surgery, we assume, that surgery related variables are influential, as well (defined in Table 5). To include them in a pooled logistic regression model, non-time varying surgery related variables (variables, which are already known at baseline) will be treated as baseline variables. Variables, that can be measured after surgery only (as blood loss during surgery), will be included as constant before surgery and as the actual - now known - value after the surgery. Surgery per se will be treated as time varying variable for these two IPW models. As in section 5.8.3, the IPW models (pooled logistic regression model) for death and censoring will be estimated and the IPW values will be derived.

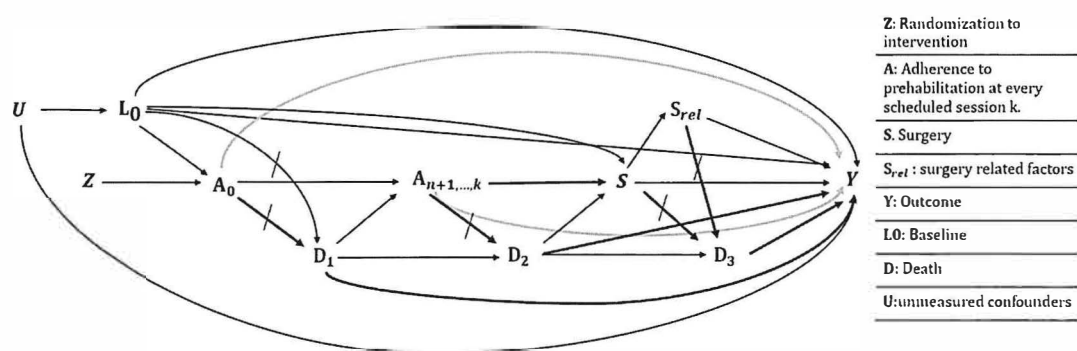


Figure 5: DAG considering competing risks (death vs. trial outcomes)

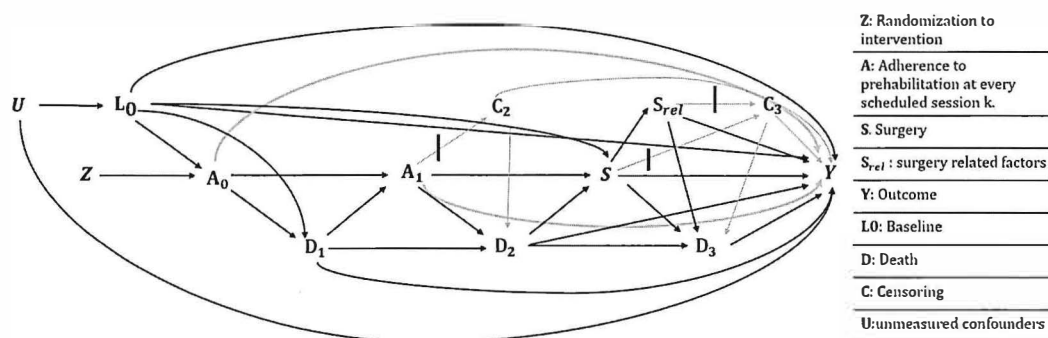


Figure 6: DAG considering censoring events

5.8.5 Outcome model

As shown in section 5.8.1, the outcome model estimates the effect of the intervention of the trials' outcome. In this case, the outcomes are the primary and secondary outcomes of the trial defined in sections 1.3.1 and 1.3.2. Since we also considered the intervention as post-randomisation event and in turn as outcome of the IPW models, we include randomisation (Z) in

Figure 4, Figure 5, and Figure 6) as independent variable to derive the effect of the intervention (prehabilitation). The IPW values, derived from the probabilities of intervention adherence, death, and censoring, will be multiplied and included as weights in the outcome model. The type of the regression model used as outcome model depends on the distribution of the primary/secondary outcomes. The estimate for the randomisation Z will represent an unbiased estimate for the effect of the prehabilitation.

5.8.6 Parametric g-formula

The IPW approach is, as mentioned above, not prone against model misspecification. The parametric g-formula is another approach to estimate the effect of sustained treatment strategies from observational/trial data and should yield similar estimates of the intervention when compared to the effects derived by the IPW approach (77–79). Therefore, we will estimate the intervention effect with both approaches to compare the estimates afterwards. When the estimates are similar, model misspecification can be ruled out.

Similar to the IPW approach, the parametric g-formula relies on the assumptions of no unmeasured confounding, positivity, consistency, and correct model specification. Initially introduced by Robins in 1986 (77), the parametric g-formula extends the concept of standardization to account for time-varying treatments. This approach involves fitting parametric models of the outcome and the joint conditional density of the confounders at each time point given prior treatment and covariate history. The counterfactual risk of the treatment strategy at each time point is estimated by performing Monte Carlo simulation based on the fitted models, which can be thought of as simulating the distribution of covariates, treatment, and outcome in a population of patients undergoing the given treatment strategy.

5.9 Propensity Score Weighting (IPW) for probability receiving the surgery

Propensity score analysis will be conducted to assess the likelihood of undergoing surgery based on baseline characteristics and prehabilitation covariates (refer to Table 3 and Table 4). This method aims to elucidate whether the covariates considered influence the decision for surgical intervention. We employed logistic regression modelling to predict the binary variable "Surgery" using the selected covariates as predictors. This statistical approach helps to quantify the probability of undergoing surgery given the specific set of baseline characteristics and prehabilitation factors included in our analysis.

However, not all patients proceeded to surgery. In some cases, the health status of patients deteriorated to the point where surgery was no longer a viable option. Conversely, patients in the intervention group who underwent prehabilitation improved to such an extent that surgery became unnecessary.

By applying the propensity score technique, we aimed to enhance our understanding of how these covariates impact the likelihood of surgery, thereby controlling for potential confounders in our subsequent analyses.

5.10 Patient - Treatment Interaction

Here, we employ the two-step approach as proposed by Chalkou et al., 2021 (80). This approach formulates a general prediction model on the population at hand for the outcome of interest. In a second step, it fits a model where the individual prediction scores and their interaction with the treatment are considered. This analysis is explorative and provides signals for potential patient treatment interactions. Detailed models for patient-treatment interaction for a series of outcomes will not be explored.

In a first step, we build the risk model using internal cross-validation to enhance generalizability (based on penalized regression approaches, elastic nets, R-package glmnet). In the second step, a simple regression model is performed on the risk score and its interaction with the intervention. The regression used will depend on the type of outcome (linear, logistic, Poisson, or Cox regression).

To derive clinically relevant insights for patient-treatment interaction, a more careful and refined approach in deriving the respective predictive models is needed.

6. Other Variables

Laboratory parameters

There are no clinical laboratory data available.

Quality of life

We compared the HRQoL measured at baseline and follow-up 1 year after elective surgery (v18) using the EuroQOL-5 dimensions 5 levels (EQ-5D-5L) and the visual analogue scale (VAS). For detailed description see 3.7.5

7. Planned subgroup analyses

The consistency of the results from the primary analysis conducted in the randomized participants is evaluated by the defined subgroups in patients with available results:

- Sex (male, female)
- Age (70-79 vs 80+)
- Frailty status (prefrail, frail)
- CCI (Charlson Comorbidity Index)
- MOCA (≤ 25 vs. > 25)

Subgroup analyses will involve stratified analyses and sensitivity tests to assess treatment effects within each subgroup. Interaction tests will also be conducted to examine potential differences in treatment effects between subgroups.

8. Interim analyses

There are no interim analyses planned.

9. Deviations from the protocol

Not applicable

10. Interpretation of results

Not applicable

11. Data problems

Not applicable

12. Software

In general, the software for analysis will be the open-source programming language R and RStudio (latest version) (81).

13. References

1. Schaller SJ, Kiselev J, Loidl V, Quentin W, Schmidt K, Mörgeli R, u. a. Prehabilitation of elderly frail or pre-frail patients prior to elective surgery (PRAEP-GO): study protocol for a randomized, controlled, outcome assessor-blinded trial. *Trials*. 6. Juni 2022;23(1):468.
2. Schaller SJ, Kiselev J, Loidl V, Quentin W, Schmidt K, Mörgeli R, u. a. Correction: Prehabilitation of elderly frail or pre-frail patients prior to elective surgery (PRAEP-GO): study protocol for a randomized, controlled, outcome assessor-blinded trial. *Trials*. 14. Februar 2023;24(1):111.
3. Semel ME, Lipsitz SR, Funk LM, Bader AM, Weiser TG, Gawande AA. Rates and patterns of death after surgery in the United States, 1996 and 2006. *Surgery*. Februar 2012;151(2):171–82.
4. Berian JR, Mohanty S, Ko CY, Rosenthal RA, Robinson TN. Association of Loss of Independence With Readmission and Death After Discharge in Older Patients After Surgical Procedures. *JAMA Surg*. 21. September 2016;151(9):e161689.
5. Saczynski JS, Marcantonio ER, Quach L, Fong TG, Gross A, Inouye SK, u. a. Cognitive Trajectories after Postoperative Delirium. *N Engl J Med*. 5. Juli 2012;367(1):30–9.
6. Govers AC, Buurman BM, Jue P, de Mol BAJM, Dongelmans DA, de Rooij SE. Functional decline of older patients 1 year after cardiothoracic surgery followed by intensive care admission: a prospective longitudinal cohort study. *Age and Ageing*. 1. Juli 2014;43(4):575–80.
7. Pisani MA, Albuquerque A, Marcantonio ER, Jones RN, Gou RY, Fong TG, u. a. Association Between Hospital Readmission and Acute and Sustained Delays in Functional Recovery During 18 Months After Elective Surgery: The Successful Aging after Elective Surgery Study. *J Am Geriatr Soc*. Januar 2017;65(1):51–8.
8. Martin ET, Kaye KS, Knott C, Nguyen H, Santarossa M, Evans R, u. a. Diabetes and Risk of Surgical Site Infection: A Systematic Review and Meta-analysis. *Infect Control Hosp Epidemiol*. Januar 2016;37(1):88–99.
9. Sibia US, Weltz AS, MacDonald JH, King PJ. Insulin-Dependent Diabetes Is an Independent Risk Factor for Complications and Readmissions After Total Joint Replacements. *J Surg Orthop Adv*. Winter 2018;27(4):294–8.
10. Smilowitz NR, Gupta N, Guo Y, Beckman JA, Bangalore S, Berger JS. Trends in cardiovascular risk factor and disease prevalence in patients undergoing non-cardiac surgery. *Heart*. Juli 2018;104(14):1180–6.
11. Watt J, Tricco AC, Talbot-Hamon C, Pham B, Rios P, Grudniewicz A, u. a. Identifying older adults at risk of harm following elective surgery: a systematic review and meta-analysis. *BMC Med*. Dezember 2018;16(1):2.
12. Birkelbach O, Mörgeli R, Spies C, Olbert M, Weiss B, Brauner M, u. a. Routine frailty assessment predicts postoperative complications in elderly patients across surgical disciplines - a retrospective observational study. *BMC Anesthesiol*. 7. November 2019;19(1):204.
13. Sepehri A, Beggs T, Hassan A, Rigatto C, Shaw-Daigle C, Tangri N, u. a. The impact of frailty on outcomes after cardiac surgery: A systematic review. *The Journal of Thoracic and Cardiovascular Surgery*. Dezember 2014;148(6):3110–7.

14. Lin HS, Watts JN, Peel NM, Hubbard RE. Frailty and post-operative outcomes in older surgical patients: a systematic review. *BMC Geriatr.* Dezember 2016;16(1):157.
15. Leung JM, Tsai TL, Sands LP. Brief report: preoperative frailty in older surgical patients is associated with early postoperative delirium. *Anesth Analg.* Mai 2011;112(5):1199–201.
16. Kim S wook, Han HS, Jung H won, Kim K il, Hwang DW, Kang SB, u. a. Multidimensional Frailty Score for the Prediction of Postoperative Mortality Risk. *JAMA Surg.* 1. Juli 2014;149(7):633.
17. Rana MV, Bonasera LK, Bordelon GJ. Pharmacologic Considerations of Anesthetic Agents in Geriatric Patients. *Anesthesiology Clinics.* Juni 2017;35(2):259–71.
18. Santos-Eggimann B, Cuenoud P, Spagnoli J, Junod J. Prevalence of Frailty in Middle-Aged and Older Community-Dwelling Europeans Living in 10 Countries. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences.* 1. Juni 2009;64A(6):675–81.
19. McIsaac DI, Taljaard M, Bryson GL, Beaulé PE, Gagné S, Hamilton G, u. a. Frailty as a Predictor of Death or New Disability After Surgery: A Prospective Cohort Study. *Annals of Surgery.* Februar 2020;271(2):283–9.
20. Santa Mina D, Clarke H, Ritvo P, Leung YW, Matthew AG, Katz J, u. a. Effect of total-body prehabilitation on postoperative outcomes: a systematic review and meta-analysis. *Physiotherapy.* September 2014;100(3):196–207.
21. Sanchez-Lorente D, Navarro-Ripoll R, Guzman R, Moises J, Gimeno E, Boada M, u. a. Prehabilitation in thoracic surgery. *J Thorac Dis.* August 2018;10(S22):S2593–600.
22. Gillis C, Buhler K, Bresee L, Carli F, Gramlich L, Culos-Reed N, u. a. Effects of Nutritional Prehabilitation, With and Without Exercise, on Outcomes of Patients Who Undergo Colorectal Surgery: A Systematic Review and Meta-analysis. *Gastroenterology.* August 2018;155(2):391-410.e4.
23. Jahic D, Omerovic D, Tanovic A, Dzankovic F, Campara M. The Effect of Prehabilitation on Postoperative Outcome in Patients Following Primary Total Knee Arthroplasty. *Med Arch.* 2018;72(6):439.
24. Moyce Z, Rodseth RN, Biccard BM. The efficacy of peri-operative interventions to decrease postoperative delirium in non-cardiac surgery: a systematic review and meta-analysis. *Anaesthesia.* März 2014;69(3):259–69.
25. Deschodt M, Braes T, Flamaing J, Detroyer E, Broos P, Haentjens P, u. a. Preventing Delirium in Older Adults with Recent Hip Fracture Through Multidisciplinary Geriatric Consultation. *J Am Geriatr Soc.* April 2012;60(4):733–9.
26. Gustafson Y, Brännström B, Berggren D, Ragnarsson JI, Sigaard J, Bucht G, u. a. A Geriatric-Anesthesiologic Program To Reduce Acute Confusional States in Elderly Patients Treated for Femoral Neck Fractures. *Journal of the American Geriatrics Society.* Juli 1991;39(7):655–62.
27. Friedman SM. Impact of a Comanaged Geriatric Fracture Center on Short-term Hip Fracture Outcomes. *Arch Intern Med.* 12. Oktober 2009;169(18):1712.

28. Chen H, Li S, Ruan T, Liu L, Fang L. Is it necessary to perform prehabilitation exercise for patients undergoing total knee arthroplasty: meta-analysis of randomized controlled trials. *The Physician and Sportsmedicine*. 2. Januar 2018;46(1):36–43.
29. McIsaac DI, Gill M, Boland L, Hutton B, Branje K, Shaw J, u. a. Prehabilitation in adult patients undergoing surgery: an umbrella review of systematic reviews. *British Journal of Anaesthesia*. Februar 2022;128(2):244–57.
30. Eckhardt H, Quentin W, Silzle J, Busse R, Rombey T. Cost-effectiveness of prehabilitation of elderly frail or pre-frail patients prior to elective surgery (PRAEP-GO) versus usual care – Protocol for a health economic evaluation alongside a randomized controlled trial. *BMC Geriatr*. 6. März 2024;24(1):231.
31. Wingenfeld K, Büscher, Gansweid. Das neue Begutachtungsassessment zur Feststellung von Pflegebedürftigkeit [Internet]. Bielefeld /Münster: Institut für Pflegewissenschaft an der Universität Bielefeld (IPW) / Medizinischer Dienst der Krankenversicherung Westfalen-Lippe (MDK WL); [zitiert 9. März 2021]. Verfügbar unter: https://www.uni-bielefeld.de/gesundhw/ag6/downloads/Abschlussbericht_IPW_MDKWL_25.03.08.pdf#page=1&zoom=auto,-270,842
32. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, u. a. Frailty in Older Adults: Evidence for a Phenotype. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. 1. März 2001;56(3):M146–57.
33. Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A. The Mini-Cog: a cognitive ?vital signs? measure for dementia screening in multi-lingual elderly. *Int J Geriatr Psychiatry*. November 2000;15(11):1021–7.
34. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, u. a. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *J American Geriatrics Society*. April 2005;53(4):695–9.
35. Sager MA, Hermann BP, La Rue A, Woodard JL. Screening for dementia in community-based memory clinics. *WMJ*. Oktober 2006;105(7):25–9.
36. Specka M, Weimar C, Stang A, Jöckel KH, Scherbaum N, Hoffmann SS, u. a. Trail Making Test Normative Data for the German Older Population. *Archives of Clinical Neuropsychology*. 17. Januar 2022;37(1):186–98.
37. Troyer AK, Rich JB. Psychometric Properties of a New Metamemory Questionnaire for Older Adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 1. Januar 2002;57(1):P19–27.
38. Mahoney FI, Barthel DW. FUNCTIONAL EVALUATION: THE BARTHEL INDEX. *Md State Med J*. Februar 1965;14:61–5.
39. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9(3):179–86.
40. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. August 2004;240(2):205–13.

41. Bohannon RW, Crouch RH. Two-Minute Step Test of Exercise Capacity: Systematic Review of Procedures, Performance, and Clinimetric Properties. *Journal of Geriatric Physical Therapy*. April 2019;42(2):105–12.
42. Knudson RJ, Slatin RC, Lebowitz MD, Burrows B. The maximal expiratory flow-volume curve. Normal standards, variability, and effects of age. *Am Rev Respir Dis*. Mai 1976;113(5):587–600.
43. Nunn AJ, Gregg I. New regression equations for predicting peak expiratory flow in adults. *BMJ*. 22. April 1989;298(6680):1068–70.
44. Roberts HC, Syddall HE, Cooper C, Aihie Sayer A. Is grip strength associated with length of stay in hospitalised older patients admitted for rehabilitation? Findings from the Southampton grip strength study. *Age and Ageing*. 1. September 2012;41(5):641–6.
45. Roberts HC, Denison HJ, Martin HJ, Patel HP, Syddall H, Cooper C, u. a. A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. *Age and Ageing*. Juli 2011;40(4):423–9.
46. Bohannon RW, Wang YC. Four-Meter Gait Speed: Normative Values and Reliability Determined for Adults Participating in the NIH Toolbox Study. *Archives of Physical Medicine and Rehabilitation*. März 2019;100(3):509–13.
47. Hinman MR, O'Connell JK, Dorr M, Hardin R, Tumlinson AB, Varner B. Functional Predictors of Stair-Climbing Speed in Older Adults. *Journal of Geriatric Physical Therapy*. Januar 2014;37(1):1–6.
48. Podsiadlo D, Richardson S. The Timed "Up & Go": A Test of Basic Functional Mobility for Frail Elderly Persons. *Journal of the American Geriatrics Society*. Februar 1991;39(2):142–8.
49. Rubenstein LZ, Harker JO, Salva A, Guigoz Y, Vellas B. Screening for Undernutrition in Geriatric Practice: Developing the Short-Form Mini-Nutritional Assessment (MNA-SF). *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. 1. Juni 2001;56(6):M366–72.
50. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *General Hospital Psychiatry*. Juli 2010;32(4):345–59.
51. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, u. a. Validation and Standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the General Population. *Medical Care*. März 2008;46(3):266–74.
52. Taylor JK, Buchan IE, van der Veer SN. Assessing life-space mobility for a more holistic view on wellbeing in geriatric research and clinical practice. *Aging Clin Exp Res*. April 2019;31(4):439–45.
53. Schepens S, Goldberg A, Wallace M. The short version of the Activities-specific Balance Confidence (ABC) scale: Its validity, reliability, and relationship to balance impairment and falls in older adults. *Archives of Gerontology and Geriatrics*. Juli 2010;51(1):9–12.
54. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, u. a. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res*. September 2013;22(7):1717–27.

55. Garin O, Ayuso-Mateos JL, Almansa J, Nieto M, Chatterji S, Vilagut G, u. a. Validation of the „World Health Organization Disability Assessment Schedule, WHODAS-2“ in patients with chronic diseases. *Health Qual Life Outcomes*. Dezember 2010;8(1):51.
56. Kriston L, Härter M, Scholl I. A latent variable framework for modeling dyadic measures in research on shared decision-making. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 2012;106(4):253–63.
57. Scholl I, Kriston L, Dirmaier J, Buchholz A, Härter M. Development and psychometric properties of the Shared Decision Making Questionnaire – physician version (SDM-Q-Doc). *Patient Education and Counseling*. August 2012;88(2):284–90.
58. Schmidt J, Nübling R. ZUF-8. Fragebogen zur Messung der Patientenzufriedenheit. Hogrefe; (Diagnostische Verfahren in der Psychotherapie).
59. Nease RF, Brooks WB. Patient desire for information and decision making in health care decisions: The autonomy preference index and the health opinion survey. *J Gen Intern Med*. November 1995;10(11):593–600.
60. European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials [Internet]. EMA/CHMP/ICH/436221/2017; 2020 S. 1–19. Verfügbar unter: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-and-sensitivity-analysis-clinical-trials-guideline-statistical-principles-clinical-trials-step-5_en.pdf
61. Robins JM, Finkelstein DM. Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests. *Biometrics*. September 2000;56(3):779–88.
62. Amelung V, Angelkorte M, Augurzky B. Ergebnisse der ersten nationalen Reifegradmessung deutscher Krankenhäuser [Internet]. 2022 [zitiert 6. August 2024] S. 148. Report No.: DigitalRadar. Verfügbar unter: https://www.bundesgesundheitsministerium.de/fileadmin/user_upload/2022-09-06_Zwischenbericht_barrierefrei_lo.pdf
63. Ludwig K, Graf Von Der Schulenburg JM, Greiner W. German Value Set for the EQ-5D-5L. *PharmacoEconomics*. Juni 2018;36(6):663–74.
64. Ludwig K, von der Schulenburg JMG, Greiner W. Valuation of the EQ-5D-5L with composite time trade-off for the German population - an exploratory study. *Health Qual Life Outcomes*. 20. Februar 2017;15(1):39.
65. Sutton R, Bann S, Brooks M, Sarin S. The surgical risk scale as an improved tool for risk-adjusted analysis in comparative surgical audit. *British Journal of Surgery*. 5. November 2002;89(6):763–8.
66. Agarwal A, Taneja S, Chopra M, Duseja A, Dhiman RK. Animal Naming Test - a simple and accurate test for diagnosis of minimal hepatic encephalopathy and prediction of overt hepatic encephalopathy. *Clin Exp Hepatol*. Juni 2020;6(2):116–24.
67. Brunner E, Bathke AC, Konietzschke F. Rank and Pseudo-Rank Procedures for Independent Observations in Factorial Designs: Using R and SAS [Internet]. Springer International Publishing; 2019. (Springer Series in Statistics). Verfügbar unter: <https://books.google.de/books?id=t9KiDwAAQBAJ>

68. Brunner E, Munzel U. Nichtparametrische Datenanalyse: Unverbundene Stichproben [Internet]. Springer Berlin Heidelberg; 2013. (Statistik und ihre Anwendungen). Verfügbar unter: <https://books.google.de/books?id=XgSdBgAAQBAJ>
69. Pepe MS. The statistical evaluation of medical tests for classification and prediction. 1. publ. in paperback. Oxford New York: Oxford University Press; 2004. 302 S. (Oxford statistical science series).
70. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal*. 2. Januar 2012;33(2):176–82.
71. Mallinckrodt CH. Preventing and treating missing data in longitudinal clinical trials: a practical guide. Cambridge: Cambridge University Press; 2013. 165 S. (Practical guides to biostatistics and epidemiology).
72. Kauermann G, Carroll RJ. A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *Journal of the American Statistical Association*. 2001;96(456):1387–96.
73. The Therapeutic Goods Administration. Note for Guidance on Clinical Safety Data Management: Definitions and Standards for Expedited Reporting (CPMP/ICH/377/95) [Internet]. 2000 Juli [zitiert 21. Juni 2023]. Verfügbar unter: <https://www.tga.gov.au/sites/default/files/ich37795.pdf>
74. Hernán MA, Hernández-Díaz S, Robins JM. Randomized Trials Analyzed as Observational Studies. *Ann Intern Med* [Internet]. 10. September 2013 [zitiert 15. Juli 2024]; Verfügbar unter: <http://annals.org/article.aspx?doi=10.7326/0003-4819-159-8-201310150-00709>
75. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in Medicine*. 15. April 2020;39(8):1199–236.
76. Pearl J. Causal Diagrams for Empirical Research. Cognitive Systems Laboratory Computer Science Department University of California, Los Angeles. 1994;
77. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1. Januar 1986;7(9):1393–512.
78. Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. *BMJ*. 16. Oktober 2017;4587.
79. Chiu YH, Wen L, McGrath S, Logan R, Dahabreh IJ, Hernán MA. Evaluating Model Specification When Using the Parametric G-Formula in the Presence of Censoring. *American Journal of Epidemiology*. 3. November 2023;192(11):1887–95.
80. Chalkou K, Steyerberg E, Egger M, Manca A, Pellegrini F, Salanti G. A two-stage prediction model for heterogeneous effects of treatments. *Statistics in Medicine*. 10. September 2021;40(20):4362–75.
81. RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA [Internet]. 2019 [zitiert 21. November 2023]. Verfügbar unter: <http://www.rstudio.com/>

