



Evaluationsbericht (gemäß Nr. 14.1 ANBest-IF)

Konsortialführung:	Prof. Dr. med. Maria M. Karsten Klinik für Gynäkologie mit Brustzentrum Charité – Universitätsmedizin Berlin
Förderkennzeichen:	01NVF19013
Akronym:	PRO B
Projekttitel:	Patient-Reported Outcomes bei Brusterkrankungen – Neue Wege in der Versorgung bei metastasiertem Brustkrebs
Autorinnen und Autoren:	PD Dr. phil. Ulrike Grittner, Dr. rer. med. Pimrapat Gebert
Förderzeitraum:	01.10.2020 - 30.09.2024
Ansprechpartner:	Maria Margarete Karsten maria-margarete.karsten@charite.de Charité - Universitätsmedizin Berlin Klinik für Gynäkologie mit Brustzentrum der Charité CCM Charitéplatz 1 10117 Berlin

Das dieser Veröffentlichung zugrundeliegende Projekt PRO B wurde mit Mitteln des Innovationsausschusses beim Gemeinsamen Bundesausschuss unter dem Förderkennzeichen 01NVF19013 gefördert. Die Darstellungen im Evaluationsbericht sind das Ergebnis der unabhängigen Evaluation zur neuen Versorgungsform.

Zusammenfassung

Hintergrund: Metastasierter Brustkrebs ist trotz therapeutischer Fortschritte weiterhin mit hoher Mortalität verbunden. In der metastasierten Situation stehen die Verlängerung des Überlebens und der Erhalt einer bestmöglichen Lebensqualität im Vordergrund, wobei Prognose und Therapieentscheidungen maßgeblich von der Tumorbiologie abhängen. Da Patientinnen häufig langfristig und individuell angepasst systemisch behandelt werden, sind zusätzliche patientenberichtete Informationen zur Symptomlast und gesundheitsbezogenen Lebensqualität essenziell. Patient-Reported Outcomes (PROs) ermöglichen die direkte Erfassung dieser Aspekte aus Patientinnensicht und erlauben eine frühzeitige Erkennung von Krankheitsprogress, therapieassoziierter Toxizität und Verschlechterungen der Lebensqualität, insbesondere bei klinisch oft unterschätzten Symptomen wie Fatigue.

Ziel der neuen Versorgungsform war die frühzeitige Erkennung krankheits- oder therapieassoziierter Symptome mittels sequentieller PRO-Erhebungen, um eine Verschlechterung des Gesundheitszustands zu verhindern oder zu verzögern. Die primäre Fragestellung war, ob diese Versorgungsform die Fatigue-Belastung nach sechs Monaten im Vergleich zur Regelversorgung reduziert.

Methodik: PRO B ist eine deutschlandweite, randomisiert-kontrollierte Studie, die zwischen 2021 und 2024 an 52 Brustkrebszentren durchgeführt wurde und 924 Patientinnen mit metastasiertem Brustkrebs unter systemischer Tumortherapie einschloss. Patientinnen im Interventionsarm beantworteten wöchentliche digitale PRO-Befragungen (EORTC-CAT). Relevante Verschlechterungen lösten automatisierte Alarmer aus, die dem Behandlungsteam über ein Webportal in Echtzeit zur Verfügung gestellt wurden und eine zeitnahe Kontaktaufnahme mit der Patientin ermöglichten. Die Art der Intervention blieb den Behandelnden überlassen. Patientinnen der Kontrollgruppe erhielten die Regelversorgung mit vierteljährlichen PRO-Erhebungen. Die Analysen erfolgten nach einem modifizierten Intention-to-Treat-Prinzip und wurden für relevante klinische Kovariaten adjustiert; Zentrenunterschiede wurden mittels gemischter linearer Modelle berücksichtigt.

Ergebnisse: Insgesamt wurden 909 Patientinnen in die Hauptanalyse eingeschlossen (456 Intervention, 453 Kontrolle). Nach sechs Monaten zeigte sich eine signifikante und klinisch relevante Reduktion der Fatigue in der Interventionsgruppe (54,5 [95 %-KI: 53,7–55,4] vs. 59,9 [95 %-KI: 59,0–60,8]; mittlere Differenz –5,4 [95 %-KI: –6,6 bis –4,1]; $p < 0,001$), die auch nach zwölf Monaten anhielt. Das Gesamtüberleben war in der Interventionsgruppe ebenfalls verbessert (12-Monats-OS 88 % vs. 85 %; adjustierte HR 0,71 [95 %-KI: 0,51–0,99]).

In der gesundheitsökonomischen Analyse ($n = 235$; 118 Intervention und 117 Kontrolle) zeigte die Intervention eine signifikant höhere Lebensqualität (EQ-5D-5L) und mehr qualitätsadjustierte Lebensjahre (QALYs) bei gleichzeitig geringeren Kosten, was auf eine ökonomische Überlegenheit gegenüber der Regelversorgung hinweist. Die Akzeptanz des PRO-Monitorings war sowohl bei Patientinnen als auch beim Behandlungspersonal hoch.

Diskussion: Ein alarmbasiertes PRO-Monitoring verbessert die Symptomlast, Lebensqualität und das Überleben von Patientinnen mit metastasiertem Brustkrebs und ist zugleich kosteneffektiv. Die Ergebnisse sprechen für eine Implementierung dieses Versorgungsmodells in die klinische Routine. Das Potenzial des Ansatzes sollte auch für andere onkologische Erkrankungen geprüft werden.

Schlagnote: digitales Symptommonitoring, alarmbasiertes PRO-Monitoring-System, metastasierter Brustkrebs, Patient-Reported Outcomes (PROs), Lebensqualität, Fatigue, neue Versorgungsform

Inhaltsverzeichnis

I	Abkürzungsverzeichnis	5
II	Abbildungsverzeichnis.....	6
III	Tabellenverzeichnis	7
1	Ziele der Evaluation	9
1.1	Hintergrund	9
1.2	Ziele und Beschreibung der neuen Versorgungsform	10
1.3	Fragestellung	10
1.4	Hypothesen	11
1.4.1	Primäre Hypothese.....	11
1.4.2	Sekundäre Hypothesen	11
2	Darstellung des Evaluationsdesigns	11
2.1	Studiendesign	11
2.2	Einschluss- und Ausschlusskriterien.....	11
2.3	Randomisierung	12
2.4	Verblindung	12
2.5	Studienablauf / Zusammensetzung der Intervention	12
2.6	Begründung der Fallzahl/Stichprobengrößen (Power-Berechnungen)	14
2.7	Drittvariablen und Confounder: Komorbiditäten, soziodemographische und sozioökonomische Charakteristika.....	15
2.8	Subgruppen und Bedeutung der Subgruppenanalysen	15
2.9	Darstellung der Datenerfassung und des Archivierungskonzeptes	16
2.10	Erhebungsinstrumente.....	18
2.10.1	EORTC-Instrumente.....	19
2.11	Prozessevaluation.....	19
2.11.1	Prozessevaluation aus Sicht von Behandlungspersonal und Patientinnen	19
2.11.2	Grafische Visualisierungen von PRO-Daten	20
2.12	Auswertungsstrategie	20
2.12.1	Statistischen Analyse der Haupthypothese	20
2.12.2	Analysen der Sekundären Hypothesen	21
2.12.3	Umgang mit Dropouts und fehlenden Werten	21
2.12.4	Sensitivitätsanalysen	22
2.12.5	Statistische Analysen der Subgruppen.....	22
2.12.6	Statistische Analyse der Sekundärdaten (Krankenkassendaten).....	22
2.12.7	Prozessevaluation.....	24
2.12.8	Evaluation der graphischen Visualisierung von PRO-Werten	26
3	Ergebnisse der Evaluation	26
3.1	Studienpopulation.....	26

3.1.1	CONSORT	26
3.1.2	Drop-out	29
3.1.3	Fehlende Werte	30
3.1.4	Baseline-Charakteristika der Patientinnen vor und nach der multiplen Imputation	33
3.1.5	Rücklaufquote	37
3.1.6	Beschreibung der Studienpopulation	38
3.2	Primärer Endpunkt: Fatigue-Score	46
3.2.1	Fatigue nach sechs Monaten	46
3.2.2	Sensitivitätsanalyse	47
3.2.3	Subgruppenanalysen	48
3.3	Sekundäre Endpunkte	52
3.3.1	Fatigue-Score über 12 Monate	52
3.3.2	Körperliche Funktionsfähigkeit über 12 Monate	53
3.3.3	Gesundheitsbezogene Lebensqualität (HRQoL) über 12 Monate	54
3.3.4	Überleben	55
3.3.5	Zeit bis zum ersten Wechsel der systemischen Therapie	60
3.4	Interventionsgruppe	61
3.4.1	Wöchentliche Antworten der PRO-Befragungen	61
3.4.2	Alarmauslösende Domänen	63
3.4.3	Häufigkeit von Alarmen und Alarmbenachrichtigungen an das Behandlungsteam (Alarm-E-Mail)	64
3.4.4	Kontakt nach Alert	64
3.4.5	Alarmauslöser	65
3.5	Analyse der Sekundärdaten (Krankenkassendaten)	65
3.5.1	Beschreibung der Studienpopulation der Krankenkassendaten	65
3.5.2	Gesundheitsökonomische Evaluation	68
3.5.3	Anzahl der Krankenhaus- und Rettungststellenaufenthalte	74
3.5.4	Dauer des Krankenhausaufenthalts	77
3.6	Prozessevaluation	77
3.6.1	Neue Versorgungsform aus Sicht der Versorgenden	77
3.6.2	Neue Versorgungsform aus Sicht der Patientinnen	78
3.6.3	Evaluation der graphischen Visualisierung von PRO-Werten	83
4	Diskussion der Projektergebnisse	87
4.1	Effektevaluation	87
4.1.1	Fatigue-score	87
4.1.2	Körperliche Funktion und Gesundheitsbezogene Lebensqualität (HRQoL)	87
4.1.3	Überleben und Zeitpunkt des ersten Systemtherapie-Wechsels	88
4.2	Gesundheitsökonomische Evaluation	89
4.3	Limitationen der Studie	90
5	Schlussfolgerungen und Empfehlungen des Evaluators	91
IV	Literaturverzeichnis	92
V	Anlagen	95

I Abkürzungsverzeichnis

ANCOVA	Analysis of Covariance
BMI	Body mass index
CAT	Computer Adaptive Testing
CHA-THS	Datentreuhandstelle der Charité
CTO	Clinical Trials Office
DKG	Deutsche Krebsgesellschaft
DLR	Deutsches Zentrum für Luft- und Raumfahrt
COPD	Chronic Obstructive Pulmonary Disease
DFI	Disease free interval
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case record form
ER	Estrogen receptor (Östrogenrezeptor)
e.V.	Eingetragener Verein
EORTC	European Organisation of Research and Treatment of Cancer
GmbH	Gesellschaft mit beschränkter Haftung
HDP	Health Data Platform, Medizinisches Datenintegrationszentrum
GCP	Good Clinical Practice
HER	Human epidermal growth factor receptor
HR+ / HR-	Hormonrezeptor positiv / Hormonrezeptor negativ
HR	Hazard Ratio
HRQoL	Health-related Quality of Life (gesundheitsbezogene Lebensqualität)
HRU	Ressourcen des Gesundheitswesens
iBiKE	Institut für Biometrie und Klinische Epidemiologie
ICER	Incremental cost-effectiveness ratio (Inkrementeller Kosten-Effektivitäts-Quotient)
IQR	Interquartile range (Interquartilsabstand)
IRR	Incidence Rate Ratio (Inzidenzratenverhältnis)
ISCED	International Standard Classification of Education
KKS	Koordinationszentrum für klinische Studien
KI	Konfidenzintervall
MAR	Missing at random
MCAR	Missing completely at random
mITT	Modifiziertes Intention-to-Treat-Prinzip

NCI	National Cancer Institute
OS	Overall Survival (Gesamtüberleben)
PROs	Patient-Reported Outcomes
QALYs	Quality adjusted life years (Qualitäts-adjustierte Lebensjahre)
SAP	Statistischer Analyseplan
SD	Standard deviation (Standardabweichung)
TNBC	triple-negativer Brustkrebs
TPA	Tipping point analysis

II Abbildungsverzeichnis

Abbildung 1: Studiendesign.....	13
Abbildung 2: Individueller Studienablauf.....	14
Abbildung 3: Datenfluss	17
Abbildung 4: Übermittlung der Krankenkassendaten.....	18
Abbildung 5: CONSORT-Diagramm.	28
Abbildung 6: Rücklaufquote	37
Abbildung 7: Box-Plot und Forest Plot für den primären Endpunkt Fatigue T-Score nach sechs Monaten zwischen Interventions- und Kontrollgruppe.....	47
Abbildung 8: Längsschnittliche Daten des Fatigue-Scores über 12 Monate in der PRO B Studie	52
Abbildung 9: Longitudinale Daten der körperlichen Funktion über 12 Monate in der PRO B Studie.....	53
Abbildung 10: Longitudinale Daten der Gesundheitsbezogenen Lebensqualität über 12 Monate in der PRO B Studie.....	54
Abbildung 11: Überleben (Gesamtkohorte).....	56
Abbildung 12: Überleben bei viszeraler Metastasierung.....	57
Abbildung 13: Überleben bei triple-negativem Brustkrebs	58
Abbildung 14: Zeit bis zum ersten Wechsel der systemischen Therapie im Vergleich zwischen Kontroll- und Interventionsgruppe	61
Abbildung 15: Wöchentliche Antworten (in Prozent) in der Interventionsgruppe	62
Abbildung 16: CONSORT-Diagramm für die gesundheitsökonomische Evaluation.....	67
Abbildung 17: Kosten-Effektivitäts-Ebene mit 95%-Konfidenzniveau Ellipse.....	73
Abbildung 18: Kosten-Effektivitäts-Akzeptanzkurve.....	73
Abbildung 19: Effekte der Teilnahme an der neuen Versorgungsform aus Patientinnenperspektive	81
Abbildung 20: Inhalte der Alarmanrufe aus Patientinnenperspektive	82
Abbildung 21: Wunsch auf Übernahme in die Routineversorgung aus Patientinnenperspektive	82

Abbildung 22: Verständlichkeit und Orientierung der Abbildungen (absolute Werte und Prozentangaben).....	85
Abbildung 23: Nützlichkeit für Reflektion und Kommunikation (absolute Werte und Prozentangaben).....	85
Abbildung 24: Bedarf an zusätzlichen Erläuterungen in den Abbildungen (absolute Werte und Prozentangaben).....	86
Abbildung 25: Bedarf an zusätzlichen klinischen Informationen und Referenzdaten (absolute Werte und Prozentangaben).....	86

III Tabellenverzeichnis

Tabelle 1: Übersicht primärer Outcome und sekundäre Outcomes.....	18
Tabelle 2: Anzahl der Dropouts.....	29
Tabelle 3: Exploration der fehlenden PRO-Scores nach sechs Monaten nach Patientenmerkmalen.....	31
Tabelle 4: Baseline-Charakteristika der Patientinnen zwischen beobachteten Daten und nach multipler Imputation (imputierter Datensatz, m = 30).....	34
Tabelle 5: Stratifizierte Randomisierung.....	38
Tabelle 6: Soziodemografische Daten.....	39
Tabelle 7: Anamnese.....	42
Tabelle 8: Klinische Charakteristika.....	44
Tabelle 9: Sensitivitätsanalysen.....	48
Tabelle 10: Geplante Subgruppenanalyse des Fatigue-T-Scores 6 Monate nach der Randomisierung.....	50
Tabelle 11: Überlebensanalysen.....	59
Tabelle 12: Erster Wechsel der systemischen Therapie.....	60
Tabelle 13: Übersicht über die Anzahl der verschickten PRO-Befragungen und die Rücklaufquote in der Interventionsgruppe.....	62
Tabelle 14: Domänen, die Alarme ausgelöst haben.....	63
Tabelle 15: Anzahl der Alarme und der nach den Alarmen versandten E-Mails an das Behandlungsteam.....	64
Tabelle 16: Kontakt nach Alarm und Zeit bis zum Kontakt.....	64
Tabelle 17: Alarmauslöser.....	65
Tabelle 18: Grundlegende Merkmale der Stichprobe für die gesundheitsökonomische Evaluation.....	67
Tabelle 19: EQ-5D-5L-Indexwerte nach Zeitpunkten und qualitätsadjustierten Lebensjahre (QALY).....	70
Tabelle 20: Kosten für das Gesundheitswesen (€) – Gesamtkosten pro Jahr.....	71
Tabelle 21: Ergebnisse der Kosteneffizienzanalyse – Gesamtkosten pro Jahr (€).....	72
Tabelle 22: Ergebnisse der Sensitivitätsanalysen zur Kosteneffizienz.....	75
Tabelle 23: Inzidenzrate der Krankenhausaufenthalte pro Personenjahr.....	76

Tabelle 24: Anzahl und Dauer der Krankenhausaufenthalte pro Patientin..... 77
Tabelle 25: Anzahl der vorhandenen und fehlenden Antworten der Evaluationsfragebögen pro Item. 80

1 Ziele der Evaluation

1.1 Hintergrund

Brustkrebs ist weltweit die häufigste Krebserkrankung und krebsbedingte Todesursache bei Frauen. Trotz aller Fortschritte in Forschung und klinischer Praxis entwickeln auch heute noch bis zu 30% der Betroffenen Metastasen [1]. Laut dem Zentrum für Krebsregisterdaten des Robert-Koch-Instituts sterben damit allein in Deutschland jährlich mehr als 18.000 Patientinnen an den Folgen eines metastasierten Mammakarzinoms [2]. Mit dem Auftreten von Metastasen ändern sich nicht nur die Lebensumstände der Patientinnen, sondern auch das therapeutische Vorgehen grundlegend. Im Vordergrund der Behandlung steht nun nicht mehr die Heilung als vielmehr die Kontrolle des Tumors und die damit verbundene Verlängerung des Überlebens unter gleichzeitiger Kontrolle der Therapienebenwirkungen. Ziel ist es, die bestmögliche Lebensqualität über einen möglichst langen Zeitraum aufrecht zu erhalten.

Das Gesamtüberleben der Patientinnen mit metastasiertem Brustkrebs hat sich durch modernere Therapieoptionen über die letzten Jahrzehnte deutlich verbessert und lag mit 38 Monaten in 2010 bereits 17 Monate über dem medianen Gesamtüberleben von 1990 [3]. Die Lebenserwartung im Falle einer metastasierten Brustkrebserkrankung hängt jedoch wesentlich von den biologischen Eigenschaften des zugrundeliegenden Tumors ab. Caswell-Jin et al. zeigten in einer Metaanalyse ein medianes Überleben von 57 Monaten für Patientinnen mit Hormonrezeptor-positivem Tumor verglichen mit einer Überlebenszeit von lediglich 33 Monaten bei triple-negativen Mammakarzinomen [3]. Mit dem Ziel, den Zeitraum des Überlebens zu verlängern, unterziehen sich die meisten Patientinnen während dieser Zeit einer systemischen Tumorthherapie. Anders als in der adjuvanten Situation, in der sich Therapieentscheidungen relativ einfach an den Standards der Fachgesellschaften orientieren können, müssen sie in der metastasierten Situation jedoch auf viel individuellerer Basis getroffen werden. Unter Berücksichtigung der medizinischen Vorgeschichte, der aktuellen Beschwerden und vor allem des Patientinnenwunsches ist der Entscheidungsprozess dahingehend zeitaufwendig und muss aufgrund des fortschreitenden Krankheitsverlaufs immer wieder neu hinterfragt werden. An dieser Stelle sind Informationen, die über sogenannte Patient-Reported Outcomes (PROs) erfasst werden, hilfreich, um die bestmöglichen Therapieentscheidungen treffen zu können. PROs ermöglichen die Bewertung gesundheitsbezogener Aspekte (wie Krankheitssymptome, psychosoziale Belastungen und Lebensqualität) aus Patientinnensicht ohne zwischengeschaltete Interpretation durch eine andere Person (z. B. Arzt oder Ärztin) [4]. Sie tragen damit zur Verbesserung

- der frühzeitigen Erkennung eines möglichen Krankheitsfortschritts,
- der Einschätzung therapieassoziierter Toxizität und
- der Detektion einer Verschlechterung der sogenannten Health-related Quality of Life (HRQoL) bei.

Die besondere Herausforderung der Bewertung des Schweregrades bestimmter Krankheitssymptome oder Medikamentennebenwirkungen wird von dem betreuenden ärztlichen Personal häufig unterschätzt [5-7]. Dies betrifft insbesondere unspezifische Symptome wie Konstipation, Appetitlosigkeit und Fatigue, obwohl letztere oftmals erster klinischer Indikator für einen Krankheitsprogress sein kann [8]. In diesem Zusammenhang sind PRO-Messungen hilfreich, da sie den Behandelnden ermöglichen, sich binnen kürzester Zeit ein sehr differenziertes Bild über den Gesundheitszustand ihrer Patientinnen zu verschaffen. Besonders in der metastasierten Situation, in der die Patientinnen so wenig Zeit wie möglich in der Klinik verbringen sollen, und die Therapien daher in der Regel ambulant durchgeführt werden, ermöglichen regelmäßige PRO-Messungen ein Echtzeit-Monitoring von Symptomen und Nebenwirkungen sowie zeitnahe Interventionen der Behandelnden. Müller et al. berichten, dass schon eine geringe Verschlechterung der HRQoL bei Patientinnen mit

metastasiertem Mammakarzinom mit einem Progress der Erkrankung assoziiert ist [9]. Die Möglichkeit, mittels PRO-Messungen eine solche Verschlechterung und damit ggf. einen Krankheitsprogress frühzeitig festzustellen, kann den Betroffenen weitere Zyklen einer nicht mehr effektiven Tumorthherapie ersparen.

1.2 Ziele und Beschreibung der neuen Versorgungsform

Ziel der Implementierung der neuen Versorgungsform war es, krankheits- oder therapieassoziierte Symptome frühzeitig zu erkennen und durch eine entsprechende Intervention eine Verschlechterung des Gesundheitszustandes der Patientin zu verhindern oder zumindest zu verzögern. Sequentielle PRO-Messungen sind in der Lage, differenziert und in Echtzeit widerzuspiegeln, wie es der erkrankten Patientin geht, und berichten (abhängig vom gewählten Messinstrument) die häufigsten Symptome, die im Rahmen einer Krebserkrankung und deren Behandlung auftreten.

Die Patientinnen in der neuen Versorgungsform erhielten einmal wöchentlich (immer am gleichen Wochentag) eine Aufforderung entweder via App oder per E-Mail, die PRO-Befragung (studienpezifischer Fragebogen aus der CAT-Itembank der European Organisation of Research and Treatment of Cancer, EORTC) durchzuführen. Bei Verschlechterung der PRO-Scores im individuellen longitudinalen Verlauf und/oder im Vergleich zu zuvor definierten Grenzwerten wurde ein Alarm generiert. Dieser Alarm musste durch die kooperierenden Studienzentren innerhalb von 48 Stunden quittiert und dokumentiert werden. Auf den Alarm folgte eine Kontaktaufnahme des jeweiligen Zentrums mit der Patientin, um die Verschlechterung ihrer Werte zu spezifizieren und ggf. zu intervenieren. Die auf die Kontaktaufnahme folgende Intervention bzw. Maßnahme durch das Studienzentrum wurde in PRO B nicht vorgegeben. Die Art der Maßnahmen blieb somit dem Behandelnden überlassen.

Insofern testete die Studie nicht eine spezifische Intervention bei Brustkrebspatientinnen, sondern eher eine schnelle, individualisierte Kontaktaufnahme und mögliche Intervention verschiedener Ausprägung bei Verschlechterung subjektiver Lebensqualitätsparameter. Der Effekt der Intervention ergab sich vor allem durch die schnelle, individualisierte Reaktion der Behandelnden auf alarmierende oder sich alarmierend verschlechternde patientenorientierte, lebensqualitätsbezogene Parameter.

Diese neue Versorgungsform galt für alle Patientinnen im Interventions-Behandlungsarm für die gesamte Dauer der Studie (24 Monate), mindestens jedoch für 12 Monate. Am Ende des Behandlungszeitraums wurde eine Auswertung der erhobenen Endpunkte und ein Vergleich zwischen der Interventions- und einer Kontrollgruppe nach den im folgenden beschriebenen Kriterien und Methoden durchgeführt.

1.3 Fragestellung

Mit der PRO B Studie sollte folgende Frage beantwortet werden:

Führt die neue Versorgungsform zu einer geringeren Belastung durch Fatigue (gemessen mit den studienpezifischen Items¹ aus der CAT-Itembank der European Organisation of Research

¹ In der PRO B-Studie wurden ausschließlich validierte Items aus der EORTC Item-Bank verwendet. Die studienpezifischen Anpassungen bestanden in der Auswahl geeigneter Items für das wöchentliche Monitoring im Interventionsarm, jedoch ohne inhaltliche Veränderungen der Items selbst. Grundlage der PRO-Erhebung war das EORTC QLQ-C30, zu dem zusätzliche Items ergänzt wurden, um den Messbereich und die Messgenauigkeit für das wöchentliche Monitoring zu erhöhen und die Antwortbelastung zu reduzieren, indem nicht in jeder Woche exakt dieselben Fragen gestellt wurden. Die Domänenbewertung erfolgte ohne die zusätzlich eingesetzten Einzelitems (Details hierzu finden sich in Abschnitt 2.10.1).

and Treatment of Cancer (EORTC) in der Interventionsgruppe im Vergleich zur Kontrollgruppe 6 Monate nach Behandlungsbeginn?

1.4 Hypothesen

1.4.1 Primäre Hypothese

Der Studie lag folgende primäre Hypothese zugrunde:

Patientinnen mit Intervention haben 6 Monate nach Behandlungsbeginn einen geringeren Fatigue-Score als Patientinnen in der Kontrollgruppe.

1.4.2 Sekundäre Hypothesen

Der Studie lagen folgende sekundäre Hypothese zugrunde:

- (1) Patientinnen in der Interventionsgruppe haben im Vergleich zur Kontrollgruppe*
 - nach 6 und 12 Monaten eine geringere Anzahl an Krankenhaus- und Notaufnahmehaufenthalten*
 - nach 6 und 12 Monaten einen höheren Score im Bereich körperliche Funktion*
 - nach 12 Monaten einen niedrigeren Score im Bereich Fatigue*
 - nach 12 Monaten eine erhöhte Überlebensrate bei viszeraler Metastasierung*
 - nach 12 Monaten eine erhöhte Überlebensrate bei triple negativer Mammakarzinomerkkrankung*
 - eine verkürzte Zeit bis zum ersten Wechsel der systemischen Therapie, da eine Therapieunverträglichkeit oder ein Fortschreiten der Erkrankung frühzeitig erkannt wird*
 - nach 6 und 12 Monaten einen höheren Summen-Score der gesundheitsbezogenen Lebensqualität (HRQoL).*
- (2) Nach 12 Monaten ist die Intervention im Hinblick auf die Gesundheitskosten und den Interventionsnutzen als kosteneffektiv einzustufen.*
- (3) Der Anteil an beantworteten wöchentlichen PRO-Erhebungen bleibt in der Interventionsgruppe über den Zeitraum von 52 Wochen hinweg konstant.*
- (4) Die Prozessevaluation aus Sicht von Behandlungspersonal und Patientinnen liefert Erkenntnisse hinsichtlich Optimierungsmöglichkeiten der Interventionsprozesse, Nutzung der Intervention in der klinischen Routine und Wirkmechanismen der Intervention.*
- (5) Die Rückmeldungen von Patientinnen und Behandlungspersonal zur Darstellung der PRO-Werte liefern Erkenntnisse, die zur Optimierung von Nutzerfreundlichkeit und Verständlichkeit der PRO-Darstellungen und somit zur Verbesserung der patientenzentrierten Kommunikation im klinischen Alltag beitragen.*

2 Darstellung des Evaluationsdesigns

2.1 Studiendesign

Bei PRO B handelte es sich um eine prospektive, multizentrische, randomisierte zweiarmlige Interventionsstudie.

2.2 Einschluss- und Ausschlusskriterien

Es bestanden folgende **Einschlusskriterien**:

- weibliches Geschlecht,
- metastasierter Brustkrebs,

PRO B (01NVF19013)

- Behandlung erfolgt über ein an der Studie teilnehmendes Brustkrebszentrum,
- Einwilligungsfähigkeit,
- 18 Jahre oder älter,
- medikamentöse Behandlung aufgrund des metastasierten Mammakarzinoms,
- geschätzte Lebenserwartung bei Studieneinschluss von mehr als 3 Monaten²,
- Eastern Cooperative Oncology Group (ECOG) 0-2,
- Bereitschaft, an einer wöchentlichen, online-basierten PRO-Befragung teilzunehmen,
- Zugang zum Internet (Smartphone).

Ausgeschlossen werden Patientinnen, die keine aktive Tumorthherapie erhalten (Comfort Care) oder nicht alle der oben genannten Einschlusskriterien erfüllen.

2.3 Randomisierung

Die Randomisierung erfolgte nach folgenden Strata:

- Zentren (n=52)
- Art der Fernmetastasierung (3 Gruppen: ossäre Metastasen & Hautmetastasen & Lymphknotenmetastasen; viszerale Metastasen (ein Organ); Gehirnmastasen & multiple Metastasen)
- histologischer Befund (2 Gruppen: [HR+/HER-, HR+/HER+]; [HR-/HER+, triple negative])

Durch die hohe Anzahl der Strata konnte es zu Unausgeglichenheiten bei einer einfachen Randomisierung hinsichtlich der Stratifizierungsvariablen zwischen der Interventions- und der Kontrollgruppe kommen. Um diese möglichst gering zu halten, wurde unter Verwendung der Minimierungsmethode von Pocock und Simon adaptiv randomisiert [10]. Die Randomisierung erfolgte nach Aufklärung und Einwilligung, Download der Applikation und nach der Baseline-Befragung.

2.4 Verblindung

Eine Verblindung im PRO B-Studiendesign war nicht umsetzbar, da es sich um ein offenes („open-label“) randomisiertes Studiendesign handelte. Gleichzeitig war eine Verblindung der Evaluierenden oder Datenanalytistinnen nicht praktikabel: Die Datenanalyse erfolgte auf Basis der PRO-Scores und alarmbezogenen Ereignisse, sodass Evaluatorinnen Zugang zu Gruppenzugehörigkeiten erhielten. Zudem ist eine Verblindung der Analyse trotz potenzieller Verzerrungsrisiken kaum praktikabel, wenn Intervention und Kontrolle klar differenzierbar sind.

2.5 Studienablauf / Zusammensetzung der Intervention

Die für PRO B in Frage kommenden Patientinnen luden in den jeweiligen Studienzentren die Smartphone-Applikation PatientConcept für die PRO Befragungen herunter. Damit einhergehend wurde jeder Patientin direkt über die App eine individuelle PatientConcept ID zugewiesen. Diese wiederum übermittelte das Studienzentrum gemeinsam mit der Randomisierungsanfrage an die Studienzentrale, die anschließend mittels Randomisierungstool die Randomisierung vornahm. Das Studienpersonal in der PRO B

² Es war von einer deutlichen Reduktion der in Frage kommenden Patientinnen auszugehen, wenn nur jene mit einer Lebenserwartung von über 12 Monaten eingeschlossen worden wären. Dies hätte zudem zu einer Verzerrung führen können, da vornehmlich stabile Patientinnen eingeschlossen worden wären. Die Studie war explizit auch für schwer kranke Patientinnen konzipiert, deren Lebensqualität verbessert werden sollte. Die Lebenserwartung hängt darüber hinaus auch mit der Tumorbiologie und der Metastasenlokalisation zusammen. Um also die Versorgungsrealität in der Studie bestmöglich abzubilden, wurden auch Patientinnen mit geringerer zu erwartender Lebenserwartung in die Studie aufgenommen.

Zentrale führte daraufhin über eine Web-Applikation die Zuweisung in den jeweiligen Behandlungsarm aus.

Die Patientinnen in der Interventionsgruppe erhielten wöchentlich eine PRO-Befragung (studienpezifisches Messinstrument aus der EORTC CAT-Itembank) über die Smartphone-Applikation, welche bei Verschlechterung der PRO-Werte einen Alarm generiert. Dieser Alarm wurde durch die kooperierenden Studienzentren innerhalb von 48 Stunden quittiert und dokumentiert. Auf den Alarm folgte eine Kontaktaufnahme des jeweiligen Zentrums mit der Patientin, um die Verschlechterung ihrer Werte zu spezifizieren und ggf. zu intervenieren. Die auf die Kontaktaufnahme folgende Intervention durch das Studienzentrum wurde in PRO B nicht vorgegeben.

Die Studienzentrale kontrollierte hierbei die sachgemäße Quittierung und Dokumentation der Alarme und Kontaktaufnahmen und kontaktierte das Studienzentrum, sollten Alarme nicht binnen 48 Stunden quittiert und dokumentiert worden sein. Zudem wurde durch die Studienzentrale die Beantwortung der Befragungen im Webportal des Befragungssystems monitoriert. Patientinnen erhielten über die App automatisierte Benachrichtigungen zur Beantwortung der Befragungen. Blieben Befragungen jedoch über einen längeren Zeitraum aus, bereitete die Studienzentrale ein Schreiben vor, welches über das entsprechende Studienzentrum personalisiert an die Patientin weitergeleitet wurde. Der genaue Ablauf des Studiendesigns ist in Abbildung 1 dargestellt und Abbildung 2 gibt einen Überblick über den individuellen Studienablauf der Patientinnen.

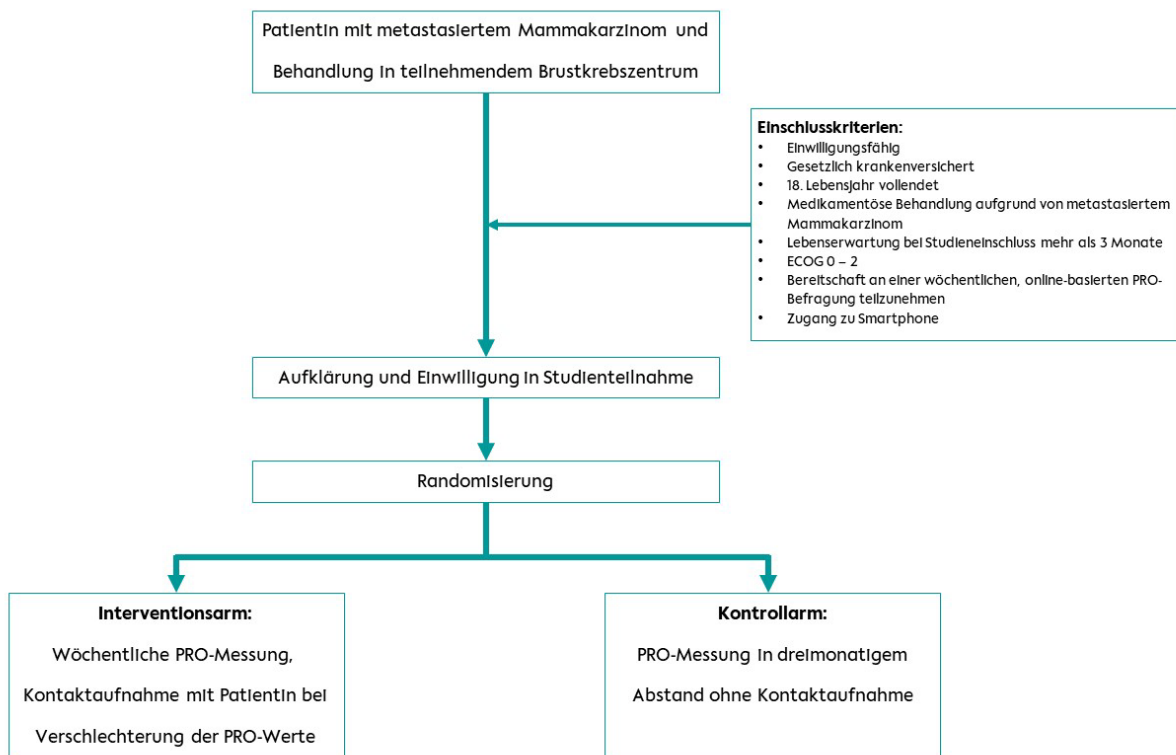


Abbildung 1: Studiendesign

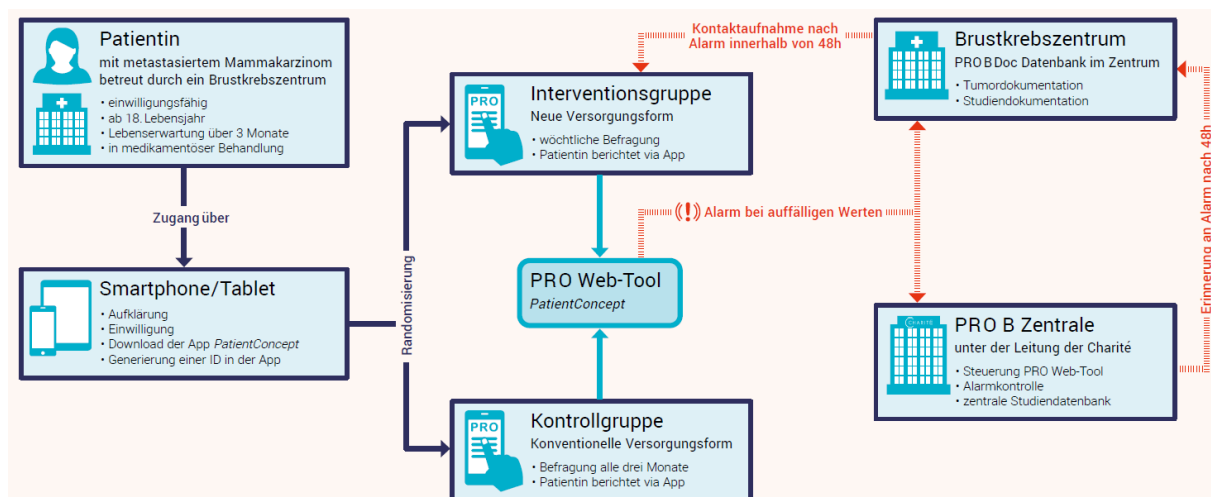


Abbildung 2: Individueller Studienablauf

2.6 Begründung der Fallzahl/Stichprobengrößen (Power-Berechnungen)

In einer Meta-Analyse zu Studien zu Veränderungen des EORTC QLQ-C30 analysierten Cocks et al. 35 Studien, die eine Verbesserung der Werte der Fatigue-Skala berichteten [11]. Hier wurden mittlere Verbesserungen der Werte um 5 Punkte genannt. Cocks et al. erstellten darauf basierend Richtlinien für die Interpretation längsschnittlicher Veränderungen auf der Fatigue-Skala, wobei verbesserte Werte ab durchschnittlich 4 Punkten als kleine Verbesserung und Werte ab 9 Punkten als mittlere Verbesserung einzuschätzen sind.

Die Effektschätzungen der PRO B Studie richteten sich nach den Angaben von Cocks et al.. Allerdings wurden aufgrund des hier durchgeführten Zwei-Gruppen-Vergleichs zwischen Interventions- und Kontrollgruppe sehr viel größere Streuungsmaße erwartet, sodass konservativ eine zu erreichende standardisierte Effektstärke von etwa 0,2 (mittlere Differenz zum Beispiel 5 Punkte. Gemeinsame Standardabweichung 25,2 Punkte) geschätzt wurde. Eine solche Streuung ist vergleichbar mit den Angaben von Cortes et al. 2015 [12].

Da außerdem die Analyse von bestimmten Subgruppen, zum Beispiel Patientinnen mit viszeralen Metastasen, als besonders relevant erachtet wurde, wurde eine Fallzahl angestrebt, die auch für Subgruppenanalysen ausreichend groß war. Hier wurde die Rekrutierung von circa 500 Patientinnen pro Gruppe als realistisch angesehen. Bei einer erwarteten Drop-Out-Rate von 20% resultierte dies in etwa 400 vollständigen Datensätzen pro Behandlungsarm. Bei Vorliegen eines minimalen klinisch relevanten Effekts von 5 Punkten Mittelwertunterschied auf der Fatigue-Skala des EORTC QLQ-C30, wie von Cocks et al. berichtet (bei einer gemeinsamen Standardabweichung von 25), bedeutete dies eine Power von 80,65% bei einem zweiseitigen t-Test mit einem Signifikanzniveau von 5%.

Die Fallzahl war somit auch für die Detektion kleiner Effektstärken (0,2 oder größer) ausreichend und trug der Anzahl relevanter Subgruppen in der Patientinnenpopulation Rechnung. Diese Berechnung wurde mit nQuery (Version 8.3.1.0) durchgeführt.

Da die geplante Stichprobengröße nicht im ursprünglich intendierten Rekrutierungszeitraum erreicht werden konnte, wurde der Rekrutierungszeitraum um vier Monate verlängert. Die insgesamt erwartete Stichprobengröße betrug aufgrund der Rekrutierungsschwierigkeiten 400 Patientinnen pro Gruppe. Bei einer erwarteten Drop-out-Rate von 20% ergab dies etwa 320 vollständige Datensätze pro Behandlungsgruppe. Diese Fallzahl war somit für den Nachweis von Effektgrößen von 0,2 mittels zweiseitigem t-Test für unabhängige Stichproben mit einem Signifikanzniveau von 5% und einer Power von 71,4% ausreichend. Bei einer Effektgröße von $\geq 0,225$ konnte eine Power von mindestens 81,1% erreicht werden.

2.7 Drittvariablen und Confounder: Komorbiditäten, soziodemographische und sozioökonomische Charakteristika

Als mögliche Confounder wurden einmalig im Rahmen der Erstbefragung Daten zu folgenden sozioökonomischen Parametern (vorläufiger Fragebogen s. u.) erhoben:

- Familienstand
- Pflegebedürftige Angehörige/ minderjährige Kinder (≤ 14 Jahren) im Haushalt
- Bildungsstatus (nach ISCED [13])
- Erwerbsstatus
- Einkommen
- Migrationshintergrund

Neben der Brustkrebs-spezifischen Anamnese (z. B. familiäre Krebsbelastung) ermöglichte die Befragung über die Smartphone-Applikation auch die Erhebung eventueller Vorerkrankungen (vorläufiger Fragebogen s. u.):

- kardiale und kardiovaskuläre Vorerkrankungen (arterielle Hypertonie, koronare Herzerkrankung, Angina pectoris, Myokardinfarkt, Herzinsuffizienz, Vorhofflimmern, Klappenvitien)
- vaskuläre Vorerkrankungen (tiefe Venenthrombose, periphere arterielle Verschlusskrankheit, Lungenarterienembolie)
- cerebrovaskuläre Erkrankungen (TIA, Apoplex)
- hämatologische Vorerkrankungen (Anämie)
- pneumologische Vorerkrankungen (Asthma bronchiale, COPD, Schlafapnoe)
- endokrine/ metabolische Vorerkrankungen (Hyperlipidämie, Diabetes mellitus (therapiebedürftig ohne und mit Endorganschäden), Hypothyreose, Adipositas)
- rheumatologische Vorerkrankungen (Polymyalgia rheumatica, Lupus erythematodes, rheumatoide Arthritis, Polymyositis)
- psychische Vorerkrankungen (Depression, Angststörung)
- neurologische Vorerkrankungen (Migräne, Krampfanfälle, Gleichgewichtsstörungen, Neuropathie, Hemiplegie, Demenz)
- gastrointestinale Vorerkrankungen (gastrointestinale Ulcera, chronisch entzündliche Darmerkrankung, Divertikulitis, Reflux, Reizdarmsyndrom, leichte Lebererkrankung (Leberzirrhose ohne portale Hypertonie), schwere Lebererkrankung (Leberzirrhose mit portaler Hypertonie, Ösophagusvarizenblutung))
- gynäkologische Vorerkrankungen (Infertilität, Abort, Dys- bzw. Hypermenorrhoe, Endometriose, Ovarialzysten)
- urogenitale Vorerkrankungen (Nephro- bzw. Urolithiasis, Niereninsuffizienz, Dialysepflicht, Miktionsbeschwerden, Harnwegsinfekte, Inkontinenz)
- Infektionserkrankungen (HIV, AIDS)
- onkologische Vorerkrankungen (andere Tumorerkrankungen (mit und ohne Metastasen), Leukämie, Lymphom)
- Voroperationen
- Suchtmittelanamnese (Alkohol, Nikotin)

Da ein nicht unerheblichen Einfluss dieser Komorbiditäten auf den primären Endpunkt der Studie (Fatigue) anzunehmen war, wurden diese Parameter sowohl bei der Interpretation der Messwerte im laufenden Projekt als auch in der abschließenden Analyse (z. B. als Komorbiditätsindex des National Cancer Institute (NCI) [14], unterstrichene Erkrankungen) berücksichtigt.

2.8 Subgruppen und Bedeutung der Subgruppenanalysen

Es wurden Subgruppen anhand folgender Kriterien gebildet und betrachtet:

- Metastasierung (sekundär/therapiert, primär/Therapieinaktivität)

PRO B (01NVF19013)

- Tumorsubtyp (HR+/HER2+, HR+/HER2-, HR-/HER2+, TNBC)
- Ort der Metastasierung (cerebral, viszeral, ossär, multipel)
- Alter bei Randomisierung (<50, ≥50)
- Therapie bei Studieneinschluss (Chemotherapie, endokrine Therapie, Anti-Her2-Therapie, Immuntherapie, zielgerichtete Therapie und andere)
- Beziehungsstatus (Partnerschaft/Ehe, ledig)
- Einkommen
- Depression (ja/ehemals, nein)
- BMI (<30, ≥30 kg/m²)
- Veränderung der Alarmkriterien (vor und nach dem 01.09.2022)

Patientinnen mit primär metastasiertem Brustkrebs haben eine bessere Prognose als Patientinnen mit rezidivierter metastasierter Erkrankung. Dieser Unterschied wird darauf zurückgeführt, dass etwaige adjuvante Systemtherapien der letzteren zu Resistenzen führen können, die bei therapienaiven Patientinnen, also primär Metastasierten nicht vorliegen [15, 16]. Innerhalb der Gruppe der sekundär metastasierten Tumoren gilt ein längeres disease free interval (DFI) als prognostisch günstig [15]. Einen weiteren Prognose-entscheidenden Faktor stellt das Alter bei Erstdiagnose der Metastasen dar. So ist ein jüngeres Alter bei initialem Auftreten von Fernmetastasen mit einer besseren Prognose assoziiert [15]. Grund für die schlechtere Überlebensrate der Patientinnen über 50 könnte in diesem Zusammenhang eine geringere Ansprechrates auf Chemotherapien mit ggf. mehr Nebenwirkungen und häufigerer Dosisreduktion sein [16]. Weiterhin gilt die Tumorbiologie als wesentlicher prognostischer Faktor, da von ihr unter anderem das Streuungsmuster der neoplastischen Zellen abhängt. Die beste Prognose wird für Hormonrezeptor (HR)-positive (Luminal A), gefolgt von HER2neu (HER) exprimierende und die schlechteste für Luminal B und triple-negative Tumoren (TNBC) beschrieben [17]. Diese Unterschiede werden unter anderem auf unterschiedliche Metastastropismen zurückgeführt. Bei den Luminaltumoren (HR+/HER-) liegen vordergründig ossäre Metastasen vor, TNBC dagegen weisen weitaus häufiger viszerale und cerebrale Metastasen auf [18], wobei mit letzteren mithin das schlechteste Overall Survival (OS) assoziiert ist [17]. Unabhängig von der Hormonrezeptorexpression liegt bei HER2neu positiven Tumoren am häufigsten eine viszerale Metastasierung vor, die ebenfalls mit einem kürzeren Gesamtüberleben als Knochenmetastasen einhergeht [17]. HER2neu exprimierende Tumoren stellen zudem einen Risikofaktor für cerebrale Metastasen dar, da die zur Verfügung stehende HER2neu gerichtete Therapie (z. B. Trastuzumab) nicht liquorgängig ist [19].

Einen weiteren mitunter auch prognostischen Faktor stellt die für die PRO B Studie als primärer Endpunkt definierte Fatigue dar [8]. Diese wird durch unterschiedliche Parameter beeinflusst. So zeigen sich z. B. therapieabhängige Unterschiede des Auftretens und Verlaufs Fatigue-spezifischer Symptome zwischen Patientinnen nach Chemotherapie und endokriner Therapie [20]. Auch soziodemographische Charakteristika wie das Einkommen und der Beziehungsstatus der Betroffenen sind mit dem Auftreten von Tumor-bedingter Fatigue assoziiert [21]. In mehreren Studien wurde außerdem festgestellt, dass ein Zusammenhang von Fatigue mit Depression und mangelnder körperlicher Aktivität/erhöhtem Body-Mass-Index besteht [22, 23].

2.9 Darstellung der Datenerfassung und des Archivierungskonzeptes

Die klinische Tumordokumentation erfolgt in den von der Deutschen Krebsgesellschaft (DKG) zertifizierten Brustkrebszentren bereits seit Jahren standardisiert und ist flächendeckend etabliert. Basis hierfür ist eine verifizierte XML-Schnittstelle, die aus dem klinischen Tumordokumentationssystem einen identischen bzw. einen vergleichbaren XML-Datensatz generiert. Dieser XML-Standarddatensatz wird über die OncoBox der OnkoZert GmbH qualitätsgesichert und steht dann für verschiedene Zwecke (z.B. Abbildung Qualitätsindikatoren, Versorgungsforschung) zur Verfügung. Innerhalb des Projektes wurde diese flächendeckende IT-Infrastruktur genutzt und erweitert. Hierzu wurde eine neue

PRO B (01NVF19013)

Anwendung erstellt, das studienspezifische eCRF PRO B Doc der OnkoZert GmbH, in dem zusätzliche, für das Projekt erforderliche Datenfelder erfasst werden konnten.

Da sich eine geringe Schnittmenge zwischen den Zertifizierungsdaten (OncoBox-XML-Daten) aus den Tumordokumentationssystemen der Zentren und den Studiendaten herausstellte, wurden zahlreiche Funktionen einer OncoBox direkt in Pro B Doc integriert. So kam der OncoBox entgegen der initialen Planung keine zentrale Rolle in der PRO B Studie zu, und der XML-Standarddatensatz der OncoBox wurde nicht verwendet. Durch einen XML-Import der Zertifizierungsdaten in PRO B Doc war technisch dennoch weiterhin eine Zusammenführung der Zertifizierungsdaten und der Studiendaten möglich. Die importierten Zertifizierungsdaten waren in PRO B Doc nicht editierbar. Aufgrund der geringen Schnittmenge der Zertifizierungs- und Studiendaten fiel jedoch auch keine Doppeldokumentation durch die Zentren an.

Die Online-Befragung der Patientinnen wurde über ein etabliertes Befragungstool der Firma NeuroSys (*PatientConcept*) abgedeckt, das für dieses Projekt individuell angepasst wurde. Die Summe aller Projektdaten (klinische Tumordaten, zusätzliche Studiendaten und Befragungsdaten) wurden in einer zentralen Datenbankapplikation, der OncoBox-Research, zusammengeführt und standen anschließend für Auswertungen zur Verfügung. Für die IT-Infrastruktur der Datenzusammenführung war die OnkoZert GmbH als Teil des PRO B Konsortiums zuständig. Die Auswertung der Daten erfolgte durch das Institut für Biometrie und Klinische Epidemiologie (iBikE) der Charité – Universitätsmedizin Berlin. Zusätzlich wurden durch die entsprechenden Krankenkassen der Patientinnen, welche auf Grundlage von Selektivverträgen nach § 140a SGB V in die Studie eingeschlossen wurden (BARMER, DAK Gesundheit, mkk – meine krankenkasse (ehemals BKK VBU)), Routinedaten zur Verfügung gestellt und über eine Datenintegrationsstelle in der zentralen Studiendatenbank mit den anderen Daten verknüpft. Die Abbildungen 3 und 4 fassen den Datenfluss bei PRO B zusammen. Die zentrale Studiendatenbank wurde regelmäßig auf Vollständigkeit, Konsistenz und Plausibilität der Dateneinträgen überprüft. Wenn Daten während der Erhebung nicht verfügbar oder nicht zutreffend waren (not applicable), wurde dies entsprechend vermerkt. Medikamente und Nebenwirkungen wurden entsprechend der Terminologie der internationalen Referenzlisten (WHO dictionary, MedDRA) dokumentiert. Alle Aktivitäten der Datenerfassung erfolgten im Rahmen der geltenden Datenschutzrichtlinien sowie der Good Clinical Practice (GCP).



Abbildung 3: Datenfluss

* Die Hauptfunktionen der OncoBox wurden in PRO B Doc integriert, der XML-Standarddatensatz der OncoBox wurde nicht verwendet.

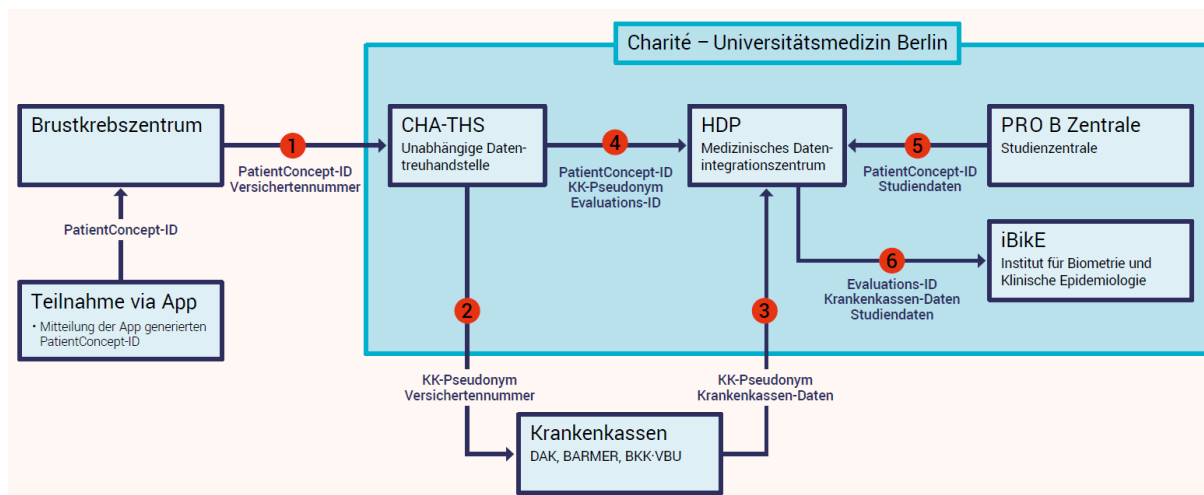


Abbildung 4: Übermittlung der Krankenkassendaten

2.10 Erhebungsinstrumente

In Tabelle 1 sind der primäre Endpunkt sowie die sekundären Endpunkte der Studie aufgeführt. Weitere Variablen, die im Rahmen der Randomisierung oder als mögliche Confounder relevant waren, wurden über die Baseline-Befragung und das Studiendokumentationssystem (eCRF in PRO B Doc) erhoben. Die Erhebung der Daten in PRO B Doc erfolgte alle drei Monate während der regulären klinischen Vorstellung im behandelnden Brustkrebszentrum.

Tabelle 1: Übersicht primärer Outcome und sekundäre Outcomes

Variabel	Datenquelle	Erhebungszeitpunkt				
		Baseline	Wöchentlich in Interventionsgruppe; alle 3 Monate in Kontrollgruppe	6 Monate	12 Monate	24 Monate
Primärer Outcome						
Fatigue-Score (EORTC)	Studienspezifischer Fragebogen aus EORTC Itembank	X	X	X	X	X
Sekundäre Outcomes						
Score für körperliche Funktion (EORTC)	Studienspezifischer Fragebogen aus EORTC Itembank	X	X	X	X	X
Gesundheitsbezogene Lebensqualität Summenscore (EORTC)	Studienspezifischer Fragebogen aus EORTC Itembank	X	X	X	X	X
Anzahl der Krankenhaus- und Rettungstufenaufenthalte	Krankenkassendaten				X	
Gesamtüberleben bei viszeraler Metastasierung	PRO B Doc, Krankenkassendaten				X	X*
Gesamtüberleben bei triple negativem Mammakarzinom	PRO B Doc, Krankenkassendaten				X	X*
Zeit bis zum ersten Wechsel der	PRO B Doc				X	X*

systemischen Therapie						
EQ-5D-5L		X		X	X	X
Kosten	Krankenkassen- daten				X	

*aus PRO B Doc

2.10.1 EORTC-Instrumente

Im Rahmen der PRO B-Studie wurden PROs mithilfe von Items aus der EORTC Item Library erhoben. Die Auswahl erfolgte auf Basis standardisierter, validierter Items der EORTC Item-Bank und wurde um thematisch relevante Einzelitems ergänzt, die speziell für das wöchentliche Monitoring im Interventionsarm ausgewählt wurden. Grundlage der PRO-Erhebung war das EORTC QLQ-C30, zu dem zusätzliche Items hinzugefügt wurden, um den Messbereich und die Messgenauigkeit für das wöchentliche Monitoring zu erhöhen und die Antwortbelastung zu reduzieren, indem nicht in jeder Woche exakt dieselben Fragen gestellt wurden. Die Domänenbewertung erfolgte ohne die zusätzlich eingesetzten Einzelitems. Diese Anpassung diente dazu, eine höhere Sensitivität für kurzfristige Veränderungen im Gesundheitszustand von Patientinnen mit metastasiertem Brustkrebs zu erreichen.

Die vorgenommenen Änderungen bzw. Ergänzungen beziehen sich nicht auf eine inhaltliche Modifikation der Items, sondern auf die Auswahl und Kombination bestehender, validierter EORTC-Items, mit dem Ziel, spezifische Symptome und Lebensqualitätsaspekte engmaschig zu erfassen. Es wurden keine neuen oder selbst entwickelten Fragen hinzugefügt.

Zur Auswertung:

- Die Auswertung der PRO-Domänen erfolgte ohne die studienspezifisch zusätzlich aufgenommenen Einzelitems, d. h. es wurden ausschließlich die Kern-Items gemäß den Vorgaben der jeweiligen Skalen (z. B. EORTC QLQ-C30) für die Berechnung von Scores berücksichtigt. Die resultierenden EORTC CAT Core Short Forms wurden dabei auf Basis standardisierter T-Werte ausgewertet, wie vom EORTC CAT-Team bereitgestellt; dabei entspricht ein Mittelwert von 50 und eine Standardabweichung von 10 der Referenz in der allgemeinen europäischen Bevölkerung[24].
- Für zusätzliche Items, die nicht Teil einer Skala mit etablierter Summenbildung waren, wurde kein Gesamtscore berechnet, sondern diese Items wurden deskriptiv ausgewertet, insbesondere im Hinblick auf die PRO-basierten Alarme im Interventionsarm.

Die Validität der eingesetzten Instrumente bleibt somit gewährleistet, da ausschließlich validierte EORTC-Items verwendet wurden und die Skalenberechnungen den methodischen Vorgaben der EORTC folgten. Die Item-Auswahl wurde zudem in enger Abstimmung mit dem erfahrenen PRO B-Team (einschließlich Psychologinnen, Biostatistikerinnen, Ärztinnen und Studienpflegekräften) methodisch begleitet.

2.11 Prozessevaluation

2.11.1 Prozessevaluation aus Sicht von Behandlungspersonal und Patientinnen

Im Verlauf und am Ende der PRO B Studie wurde eine explorative Prozessevaluation durchgeführt, um die neue Versorgungsform aus Sicht von Behandlungspersonal und Patientinnen zu evaluieren und zugleich zu untersuchen, wie die getestete neue Versorgungsform bei positiven Versorgungseffekten sinnvoll in die Routine implementiert werden könnte. Fragestellungen waren hier:

- Welche Aspekte der Studie und Intervention werden gut evaluiert und welche Optimierungsmöglichkeiten bestehen?

- Welche Aspekte müssen bei einer Nutzung der Intervention in der klinischen Routine berücksichtigt werden?
- Welche Ausprägungen hatte die Intervention? Können daraus mögliche Wirkmechanismen der Intervention abgeleitet werden?

Um die Strukturen und Abläufe der Versorgungsform aus Sicht des Behandlungsteams zu evaluieren und für eine mögliche Verstetigung zu optimieren sowie mögliche Wirkmechanismen der Versorgungsform zu identifizieren, wurden in einem explorativen inhaltsanalytischen Ansatz halbstrukturierte Interviews mit Mitgliedern des Behandlungsteams (Study Nurses, Ärztinnen/Ärzte) aus verschiedenen Studienzentren durchgeführt. Die Interviewteilnehmerinnen wurden per Schneeballmethode rekrutiert: zunächst wurden Mitglieder des Behandlungsteams aus 19 der 52 PRO B Studienzentren mit besonders hohen und niedrigen Rekrutierungszahlen per E-Mail zur Interviewteilnahme eingeladen. Einschlusskriterium war die aktive Beteiligung an der PRO B Studie als Teil des Behandlungsteams. Zur Evaluation der neuen Versorgungsform aus Patientinnenperspektive wurden sowohl qualitative (halbstrukturierte Interviews während der Studie) als auch quantitative Daten (Studienevaluationsumfrage am Ende der Studie) erhoben. Für die Interviews wurden die Interviewteilnehmerinnen mittels Schneeballmethode in den teilnehmenden Studienzentren rekrutiert. Hierzu wurden alle Studienzentren aufgerufen, ihre teilnehmenden Patientinnen die Teilnahme an den Interviews anzubieten. Einschlusskriterium war die Gruppenzuordnung in die Interventionsgruppe. Alle interessierten Teilnehmerinnen wurden interviewt. Wie initial geplant erhielten die Studienpatientinnen nach Abschluss der Datenerhebung ihre gemessenen PRO-Werte postalisch zur Verfügung gestellt (vgl. 2.11.2). In diesem Zusammenhang erfolgte auch eine Evaluationsbefragung der Teilnehmerinnen in der Interventionsgruppe hinsichtlich ihrer Einschätzung der Intervention. Hier wurden folgende Ausschlusskriterien angewandt: Tod, die zurückgezogene Einwilligung im Studienverlauf oder vollständig fehlende PRO-Fragebögen in den letzten sechs Monaten (vgl. 2.12.7).

2.11.2 Grafische Visualisierungen von PRO-Daten

Damit PROs ihr volles Potenzial entfalten können und in der Routine regelmäßig genutzt werden, müssen diese verständlich für Fachpersonal und Laien aufbereitet und damit implementierbar gemacht werden. Daher kommt insbesondere der graphischen Abbildung der PRO-Werte einer enormen Bedeutung zu, da die abstrakten Werte hierdurch verständlich und nutzbar gemacht werden können. Anstatt reiner Zahlen oder Tabellen bieten grafische Darstellung eine visuelle Repräsentation, die auf einen Blick verstanden werden kann.

Darauf basierend wurde die graphische Aufbereitung der PRO-Werte im Rahmen der PRO B Studie überarbeitet und wissenschaftlich fundiert validiert. Hierzu wurden neue PRO-Darstellungen erarbeitet, in qualitative Interviews mit Patientinnen und Behandlungsteams partizipativ weiterentwickelt und abschließend am Studienende gemeinsam mit dem Studienevaluationsfragebogen quantitativ evaluiert (vgl. 2.12.8).

2.12 Auswertungsstrategie

2.12.1 Statistischen Analyse der Haupthypothese

Die primäre Fragestellung wurde im Rahmen eines modifizierten Intention-to-Treat-Prinzips (mITT) analysiert. Nach diesem Ansatz umfasste die ITT-Population alle randomisierten Teilnehmerinnen, die auf mindestens eine PRO-Erhebung geantwortet haben. Patientinnen, die während des Studienzeitraums keine PRO-Erhebungen beantworteten, wurden von der Analyse ausgeschlossen. Alle Stratifizierungsvariablen der Randomisierung wurden als Kovariaten in allen statistischen Modellen berücksichtigt.

Während in der Analyse primär ANCOVA-Modelle verwendet wurden, wurde für die Berechnung der Fallzahl ein zweiseitiger t-Test genutzt. Basierend auf der Literatur ist dies eine konservative Vorgehensweise zur Schätzung der Fallzahl von ANCOVA-Modellen, da ein

ANCOVA-Modell mit $(1 - 2\rho) \cdot n$ Beobachtungen mindestens die gleiche Power wie ein t-Test mit n Beobachtungen hat [25], wobei ρ der Varianzinflationsfaktor (Korrelation zwischen der *baseline*-Messung und der *follow-up*-Messung) ist. Im schlechtesten Fall von $\rho = 0$ führt dies zu einer benötigten Fallzahl gleich dem t-Test. Die Analyse im ANCOVA-Modell erfolgte analog zur Stratifizierung der Randomisierung mit folgenden Kovariaten: Baseline-Messung, Art der Metastasierung und Histologie. Außerdem wurde die Analyse für die Parameter Alter, Art der systemischen Therapie, ECOG-Status bei Studieneinschluss sowie gegebenenfalls für weitere wichtige *Confounder* (vgl. 2.6) adjustiert. Die Heterogenität zwischen den Studienzentren wurde durch Zufallseffekte (*random effects*) berücksichtigt (gemischtes lineares Regressions-Modell, *random intercept model*).

2.12.2 Analysen der Sekundären Hypothesen

Alle sekundären Hypothesen wurden analog zur primären Hypothese und entsprechend des jeweiligen Outcomes mittels geeigneter statistischer Methoden (z.B. Zeit-Ereignis-Modelle, ANCOVA, logistische Regression) analysiert. Vergleichbar mit der Analyse des primären Endpunkts wurde auch bei Analyse der sekundären Hypothesen für relevante *Confounder* adjustiert und die Heterogenität zwischen den Studienzentren statistisch angemessen berücksichtigt. Durch die Verkürzung der Nachbeobachtung bei einem Teil der Stichprobe in Bezug auf sekundäre Endpunkte aufgrund der verlängerten Rekrutierungszeit, konnten einige Patientinnen nicht 12 Monate lang vollständig beobachtet werden. Dadurch ergaben sich fehlende Werte in den Outcomes. Für die sekundären Endpunkte wurden für diese spät rekrutierten Patientinnen jedoch keine imputierten Daten geschätzt, da diese späteren Daten dieser Patientinnen vollständig fehlten. Daher beruhten die Analysen der sekundären Endpunkte nach 12 Monaten auf einem geringeren Stichprobenumfang als die Analysen für den primären Endpunkt. Im Gegensatz dazu stellte die Einbeziehung einiger spät rekrutierter Patientinnen mit kürzeren Beobachtungszeiten in die Ereignis-Zeit-Analysen kein Problem dar, da in diesen Ereignis-Zeit-Analysen die individuellen Beobachtungszeiträume berücksichtigt werden konnten. Es wurden alle verfügbaren Informationen verwendet, und Patientinnen wurden als zensiert definiert, wenn das interessierende Ereignis (Rezidiv oder Tod) bis zum Ende der individuellen Beobachtungszeit nicht eintrat.

Die Anzahl und der Anteil der beantworteten PRO-Befragungen pro Woche wurden berechnet und grafisch dargestellt. Berücksichtigt wurden alle Teilnehmerinnen der Interventionsgruppe, die im Studienverlauf mindestens eine PRO-Befragung beantwortet haben. Die Auswertung erfolgte deskriptiv, ohne statistische Testung, mit dem Ziel, einen Überblick über das Antwortverhalten im Verlauf der 52 Wochen zu geben.

2.12.3 Umgang mit Dropouts und fehlenden Werten

Wir haben die Zahlen zu Verlusten beim Follow-up, Studienabbrüchen und Todesfällen zwischen den Studiengruppen und zu jedem Follow-up-Zeitpunkt im Ablaufschema der Studie dargestellt. Patientinnen galten als Lost-to-Follow-up, wenn sie bis zum Studienende nicht mehr auf die PRO-Erhebungen reagierten. Das Datum der letzten beantworteten PRO-Erhebung wurde als Zeitpunkt des letzten Follow-ups gewertet. Wenn Patientinnen zu einem bestimmten Zeitpunkt nicht auf die PRO-Erhebung antworteten, aber bei der darauffolgenden Erhebung wieder teilnahmen, wurden die fehlenden PRO-Daten imputiert und als „missing at random“ (MAR) behandelt. Studienabbrüche mit Angabe der jeweiligen Gründe wurden dokumentiert und prozentual dargestellt.

Für fehlende PRO-Daten infolge von schwerer Erkrankung oder bei Auftreten eines interkurrierenden Ereignisses wie einem vorzeitigen Tod wurde ein Worst-Case-Szenario angewendet [26]. Der Umgang mit fehlenden Daten erfolgte gemäß dem Estimand Framework, wie es in der ICH-Guideline (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use) empfohlen wird [27].

Aufgrund der längeren Rekrutierungszeit von vier Monaten wurden die spät rekrutierten Patientinnen nur sechs Monate lang vollständig beobachtet. Für diese Patientinnen wurden keine sekundären Endpunkte nach 12 Monaten mittels Imputationsmethoden geschätzt (vgl. 2.12.2). Die Einzelheiten der multiplen Imputation sind im Statistischen Analyseplan (SAP) festgelegt.

2.12.4 Sensitivitätsanalysen

Wenn ein signifikantes Ergebnis für den primären Endpunkt festgestellt wurde, wurde eine Sensitivitätsanalyse durchgeführt, um die Robustheit des Ergebnisses der primären Analyse unter der Annahme eines nicht zufällig fehlenden Ergebnisses (MNAR) zu bewerten, indem eine referenzbasierte Sensitivitätsanalyse und ein Delta-anpassender Muster-Mischungs-Ansatz für die Tipping-Point-Analyse (TPA) verwendet wurde[28, 29]. Bei der referenzbasierten Sensitivitätsanalyse handelte es sich um einen Imputationsansatz für Patientinnen, die die Intervention abbrachen, und die fehlenden Werte wurden wie in der Kontrollgruppe imputiert[28]. Die TPA geht davon aus, dass ein systematischer Unterschied zwischen den bedingten Verteilungen der fehlenden und der beobachteten Daten besteht. Um einen solchen systematischen Unterschied widerzuspiegeln, wurde ein Shift-Parameter δ auf das Imputationsmodell angewendet, wenn die fehlenden Datenpunkte imputiert wurden. Für jeden Wert von δ wurden mehrfach imputierte Datensätze erzeugt, jeder Datensatz wurde nach derselben Methode für die Primäranalyse analysiert, und die Analyseergebnisse wurden dann nach der Rubins Regel kombiniert. Durch Variation der Werte von δ wurden die Auswirkungen fehlender Daten auf die Analyseergebnisse untersucht, um Kippunkte zu ermitteln (d. h. die Werte, bei denen sich die Schlussfolgerungen zwischen den Studiengruppen von signifikant zu nicht signifikant änderten).

2.12.5 Statistische Analysen der Subgruppen

Alle Subgruppenanalysen erfolgten explorativ. Für die Subgruppenanalysen wurde im ANCOVA-Modell (primäres oder sekundäres Outcome, Kovariablen: jeweiliger Baseline-Wert, Therapie bei Randomisierung, Stratifizierungsvariablen der Randomisierung, zufälliger Achsenabschnitt (random intercept) nach Studienzentrum) zusätzlich ein Interaktionsterm zwischen Studiengruppe und Subgruppe ergänzt, um erste Hinweise für differenzielle Interventionseffekte in den Subgruppen zu analysieren. Zusätzlich wurden marginale Effektschätzer für jede Subgruppe sowie 95%-Konfidenzintervalle (KI) aus den Modellen ausgegeben und betrachtet.

2.12.6 Statistische Analyse der Sekundärdaten (Krankenkassendaten)

Gesundheitsökonomische Auswertung

Über den Interventionszeitraum hinweg wurden die Gesamtkosten der Behandlung über die Krankenkassendaten für alle Studienpatientinnen gesammelt und ausgewertet. Es wurden sämtliche stationäre und ambulante Abrechnungsdaten berücksichtigt, die im Erhebungszeitraum anfielen. Dies umfasste Krankenhausleistungen (Dauer, Diagnosen, Prozeduren, Kosten), Rehabilitationsleistungen (Art der Leistung, Diagnose, Verordnungsdatum, Dauer, Kosten), Leistungen im Sinne der Arzneimittel-, Heilmittel-, Hilfsmittel-, Fahrtkosten-, Häusliche Krankenpflege-, Haushaltshilfeverordnungen, aber ebenso vertragsärztliche Leistungen (Datum, Fachgruppe des Arztes oder der Ärztin, Diagnose, Einzelleistungen, Kosten), Leistungen der Hospiz- und Palliativversorgung sowie Leistungen der sozialen Pflegeversicherung gem. SGB XI. Da insbesondere bei den ambulanten Kosten mit einem erheblichen Zeitverzug der Kostenaufstellung durch die Krankenkassen nach Inanspruchnahme zu rechnen war, wurden die ambulanten Kosten nur bis Monat 24 der Projektlaufzeit direkt berücksichtigt. Anschließend wurden die ambulanten und die stationären Gesamtkosten ins Verhältnis zur jeweiligen Beobachtungszeit gesetzt (mittlere Behandlungskosten pro Jahr pro Patientin). Die so ermittelten individuellen jährlichen Durchschnittswerte für ambulante und stationäre Gesundheitskosten wurden dann für jede

Patientin zu einem mittleren Gesamtposten pro Jahr aufsummiert. Die individuellen jährlichen Gesamtkosten (Mittelwert) bildeten die Grundlage der weiteren Berechnungen. Somit wurden für jede Patientin für die Monate nach Studieneintritt bis Studienmonat 30 alle klinischen Kosten und bis Monat 24 alle ambulanten Kosten erfasst. Für jede Patientin wurden mindestens für fünf Monate ambulante Kosten und mindestens für 12 Monate klinische Kosten berücksichtigt. Der jährliche Mittelwert für die Gesamtkosten pro Patientin errechnete sich über die Summe der jährlichen ambulanten Kosten und der jährlichen klinischen Kosten. Durch dieses Vorgehen konnte sichergestellt werden, dass die Daten rechtzeitig zur Analyse vorlagen.

Aufgrund der typischen rechtsschiefen Verteilung von Kostendaten wurden für die Analyse log-transformierte Kosten (ln-Kosten) verwendet. Dabei entspricht der Exponent der mittleren beobachteten ln-Kosten dem geometrischen Mittel (GM) der Kosten. Es wurde ein lineares gemischtes Modell mit Patientinnen und Zentren als zufälligem Achsenabschnitt (random intercept) durchgeführt, um den Unterschied von ln-Kosten zwischen den Gruppen zu vergleichen. Weitere statistische Methoden der gesundheitsökonomischen Analyse sind im statistischen Analyseplan ausgeführt. Kosteneffektivitätsanalysen werden dann durchgeführt, wenn sich die Intervention als überlegen in Bezug auf „quality adjusted life years (QALYs)“ erwies. Ein QALY berücksichtigt Veränderungen sowohl der Lebenszeit als auch der Lebensqualität infolge von Interventionen. QALYs wurden hier auf der Grundlage von Lebensqualitätsfragebögen (EQ-5D) berechnet, die zu Baseline und zu verschiedenen folgenden Zeitpunkten während der Projektlaufzeit ausgefüllt wurden. Zwischen den Messzeitpunkten wurde eine lineare zeitliche Veränderung der Lebensqualität angenommen. QALYs ergaben sich aus der Berechnung der Fläche unter der Kurve. Falls die Intervention zu besseren Lebensqualitätswerten bei gleichzeitig geringeren Kosten führte, galt sie als kosteneffektiv. Falls es zu zusätzlichen Kosten in der Interventionsgruppe kam, wurde der „incremental cost-effectiveness ratio (ICER)“ berechnet, indem die mittleren zusätzlichen Kosten für ein zusätzliches QALY geschätzt wurden. Ein Wert von 50.000 € pro extra QALY wird in Studien oft als Grenzwert für die Kosteneffektivität einer Intervention verwendet: Lag der ICER unterhalb von 50.000 €, wurde die Intervention als kosteneffektiv eingestuft.

Fallzahlschätzung für die gesundheitsökonomische Analyse

Die Fallzahlschätzung für die gesundheitsökonomische Analyse beruhte auf den Angaben von Basch et al. (2016) [30] zum Interventionseffekt eines PRO-Monitorings gegenüber der Standardversorgung auf die Lebensqualität gemessen mit dem EQ-5D bei Patientinnen und Patienten mit einer Krebserkrankung. Basch et al. (2016) [30] berichten mittlere EQ-5D-Werte von 84,8 (95%-KI: 83,2-86,4) in der Interventionsgruppe und 79,5 (95%-KI: 76,7-82,2) in der Kontrollgruppe sechs Monate nach Interventionsbeginn (Baseline-Werte: Intervention: 86,2; Kontrolle: 86,6). Nach Skalierung der Werte auf eine Skala von 0-1 (statt 0-100) ergab sich nach der Formel von Matthews et al. (1990) [31] zur Berechnung der QALYs als Fläche unter der Kurve ein Wert von 0,428 für die Interventionsgruppe $((0,848+0,862)*0,5*0,5 \text{ Jahre})$ und von 0,415 für die Kontrollgruppe $((0,795+0,866)*0,5*0,5 \text{ Jahre})$. Dies resultierte in einen Interventionseffekt von 0,012 QALYs bei Basch et al. (2016). Bezogen auf 12 Monate wäre das ein geschätzter Effekt von 0,0245 QALYs. Für die PRO B Studie wurde ein solcher oder ähnlicher Effekt angenommen.

Weiterhin wurde davon ausgegangen, dass für die neue Versorgungsform etwa 1000€ zusätzliche Kosten aufzubringen sind. Zur Berechnung des mittleren inkrementellen Kosten-Effektivitäts-Quotienten (ICER) wurde die mittlere Kostendifferenz durch die geschätzte QALY-Differenz geteilt: $ICER = 1000€ / 0,0245 \text{ QALYs}$. Dies ergab eine mittlere geschätzte Summe von 40.816€, die für ein zusätzliches QALY zu erbringen ist. Die obere Grenze für diese Aufwendungen wird oft bei 50.000€ festgesetzt. Aus der Studie von Basch et al. (2016) resultierte zunächst nur eine Schätzung des ICER, da sie auf mittleren Werten der Kostendifferenz und der QALY-Differenz zwischen den untersuchten Gruppen beruhte. Somit war für den ICER keine direkte Fallzahlschätzung möglich. Das 95%-KI für den ICER wurde über

ein Bootstrapping-Verfahren bestimmt, indem das ICER separat für $n=5000$ Bootstrap-Stichproben aus den Originaldaten berechnet wurde.

Basierend auf den genannten Daten von Basch et al. (2016) [30] ließ sich eine Effektstärke für den Interventionseffekt von 0,331 (Cohen's d) berechnen. Eine Stichprobengröße von 200 Personen pro Gruppe hatte eine Power von mehr als 90%, um einen Mittelwertunterschied von 5,3 (die Differenz zwischen dem Mittelwert der Interventionsgruppe von 84,8 und dem Mittelwert der Kontrollgruppe von 79,5) festzustellen. Dies basierte auf der Annahme einer gemeinsamen Standardabweichung von 16,01 (berechnet auf Grundlage der Angaben zu den 95%-KI und der Stichprobengrößen bei Basch et al. 2016) bei Verwendung eines t-Tests für zwei Gruppen mit einem zweiseitigen Signifikanzniveau von 5%.

Aufgrund der zu erwartenden Drop-Out Rate von 20% konnten bei angestrebten 400 einzuschließenden Patientinnen der beteiligten Krankenkassen nur 320 vollständige Datensätze gewonnen werden. Unvollständige Datensätze wurden jedoch, sofern von fehlenden Werten im Rahmen der Mechanismen „completely missing at random“ (MCAR) oder „missing at random“ (MAR) ausgegangen werden konnte, mittels multipler Imputationsmethoden [32] unter Einbeziehung der vorhandenen Daten geschätzt. Somit wurde angenommen, dass für die gesundheitsökonomische Fragestellung inklusive geschätzter Daten 400 Datensätze zur Verfügung gestellt werden konnten. Demzufolge wurde die Rekrutierung von 400 Patientinnen der beteiligten Krankenkassen für die gesundheitsökonomische Fragestellung als ausreichend angesehen.

Da die ursprünglich geplante Stichprobengröße nicht erreicht werden konnte, wurde für die gesundheitsökonomische Analyse final eine Stichprobengröße von etwa 150 Patientinnen pro Gruppe erwartet. Bei einer erwarteten Drop-out-Rate von 20% ergab dies etwa 120 vollständige Datensätze pro Gruppe. Diese Fallzahl war für den Nachweis von Effektgrößen von mindestens 0,331 mit einer Power von 72,4% ausreichend. Bei einer tatsächlichen Effektgröße von 0,37 oder größer betrug die Power der Studie mindestens 81,4%.

Es wurde davon ausgegangen, dass diese Powerüberlegung für den EQ-5D, der als Grundlage für die Berechnung der QALYs verwendet wurde, eine robuste Schätzung des ICER bildete, für den keine direkte Fallzahlschätzung möglich war.

Statistischen Analyse der sekundären Endpunkte (Krankenkassendaten)

Die Inanspruchnahme von Ressourcen des Gesundheitswesens (HRU) wurde anhand der Anzahl ambulanter Besuche, ungeplanter Notaufnahmen, ungeplanter stationärer Besuche und der Aufenthaltsdauer (Tage) im Krankenhaus bewertet. Der Beobachtungszeitraum wurde berechnet als Zeit zwischen Zeitpunkt der Aufnahme in die Studie bis zum Ende des Untersuchungszeitraums (z. B. Todestag, Ende der Studie oder Abbruchdatum, bzw. letztes Datum, für das diese Daten von der Krankenkasse zur Verfügung standen). Für ambulante und stationäre Besuche wurden die kumulative Inzidenzrate und die Rate der Inzidenzdichte angegeben. Ein Poisson-Regressionsmodell, bzw. ein negatives Binomialmodell wurde verwendet, um die Inzidenzraten der ambulanten und stationären Besuche zwischen den Studiengruppen zu vergleichen. Die Effektgrößen wurden als Inzidenzratenverhältnisse (IRRs) mit einem 95 %-KI (Konfidenzintervall) dargestellt. Wurde eine Patientin mehrfach ins Krankenhaus eingeliefert, wurde die Aufenthaltsdauer für diese Patientin als Durchschnitt angegeben. Da die HRU um die Beobachtungszeit jeder Patientin bereinigt wurde, stellten fehlende Beobachtungszeiträume kein Problem dar, sodass hier keine Schätzung fehlender Daten vorgenommen wurde und alle vorhandenen Daten in die Analyse einfließen konnten.

2.12.7 Prozessevaluation

Prozessevaluation: Neue Versorgungsform aus Sicht der Versorgenden (Interviews)

Für eine Prozessevaluation aus Sicht des Behandlungsteams wurden von Mai bis Juli 2022 in einem explorativen inhaltsanalytischen Ansatz halbstrukturierte Interviews mit 15 Mitgliedern des Behandlungsteams (Study Nurses, Ärztinnen/Ärzte) aus verschiedenen Studienzentren

durchgeführt. Die Interviews erfolgten auf Grundlage eines Interviewleitfadens mit offenen Stimuli, welcher durch weitere narrative Fragen ergänzt wurde. In diesen Interviews wurden unter anderem Einstellungen zu PROs, Studienabläufen und -anwendungen sowie mögliche Hürden thematisiert. Einschlusskriterium war die aktive Beteiligung an der PRO B Studie als Teil des Behandlungsteams. Für die Durchführung der Interviews lag eine Genehmigung der Ethikkommission der Charité – Universitätsmedizin Berlin vor (EA2/070/23). Die Interviews wurden telefonisch geführt, aufgezeichnet und transkribiert. Es wurden keine Folgeinterviews durchgeführt, und Transkripte oder Ergebnisse wurden den Teilnehmern nicht zur Kommentierung oder Korrektur vorgelegt. Die Analyse der Interviews stützte sich auf das Consolidated Framework for Implementation Research (CFIR)[33]. Der CFIR ist ein häufig verwendetes Instrument zur Datenextraktion für die Charakterisierung der Determinanten einer effektiven Umsetzung von Innovationen im Gesundheitswesen. Die durchgeführten Interviews wurden von zwei Codierern mit Hilfe eines inhaltsanalytischen Ansatzes nach Kuckartz strukturiert.[34] Zunächst wurden einige allgemeine Kategorien deduktiv aus den Hauptbereichen des CFIR gebildet. Anschließend wurden zusätzliche Kategorien und Untereinheiten induktiv formuliert. Beide Codierer codierten jedes Interview unabhängig voneinander. Anschließend wurden Unstimmigkeiten gemeinsam besprochen und ein Konsens gefunden. Zur Strukturierung der Daten wurde das Softwareprogramm *f4analyse* verwendet.

Prozessevaluation: Neue Versorgungsform aus Sicht der Patientinnen (Interviews)

Zur Evaluation der neuen Versorgungsform aus Patientinnenperspektive wurde ein gemischter Methodenansatz aus halbstrukturierten Interviews während der Studie und einer Studienevaluationsumfrage am Ende der Studie genutzt. Es lagen hierzu Ethikvoten der Ethikkommission der Charité – Universitätsmedizin Berlin vor (EA1/318/20 und EA2/070/23). Die Patientinneninterviews wurden von September bis November 2023 geführt. Für diese Interviews wurde ein vergleichbarer Interviewleitfaden genutzt wie in den Interviews mit dem Behandlungspersonal (offene Stimuli ergänzt durch narrative Fragen). Thematisiert wurde, wie die Patientinnen mit den Studienabläufen zurechtkamen, welchen Mehrwert sie aus der Studienteilnahme zogen und welche Verbesserungsvorschläge oder Änderungswünsche für die Abläufe der Versorgungsform bestanden. Die Interviews wurden überwiegend telefonisch geführt (einmalig wurde das Interview auf Wunsch der Teilnehmerin per Videokonferenz durchgeführt). Alle Interviews wurden aufgezeichnet und transkribiert. Es wurden keine Folgeinterviews durchgeführt, und Transkripte oder Ergebnisse wurden den Teilnehmern nicht zur Kommentierung oder Korrektur vorgelegt. Die Analyse und Codierung der Interviews erfolgte nach vergleichbarer Methodik wie die Interviews des Behandlungspersonals gemäß CFIR[33] und Kuckartz[34].

Prozessevaluation: Quantitative Analyse der neuen Versorgungsform durch Befragung am Studienende

Für eine quantitative Analyse der Patientinnenperspektive zur neuen Versorgungsform wurde am Studienende ein papierbasierter studienspezifischer Evaluationsfragebogen an die Patientinnen im Interventionsarm verschickt. Studienteilnehmerinnen, die verstorben waren (n=84), ihre Einwilligung während der Studie zurückgezogen (n=23) oder in den letzten 6 Monaten keinen PRO-Fragebogen ausgefüllt (n=87) hatten, wurden von der Auswertung ausgeschlossen. Der Evaluierungsfragebogen umfasste 25 Fragen, die sich auf drei Schlüsselbereiche konzentrierten: (1) Bewertung der wöchentlichen PRO B-Fragebögen der PRO B Studie, (2) Bewertung der PRO B Studie und (3) Perspektiven für die langfristige Nutzung des PRO-Monitorings mit Alarmen. Sie umfasste geschlossene Fragen mit zwei oder mehr Antwortmöglichkeiten und Aussagen, die auf einer 4-Punkte-Likert-Skala beantwortet wurden (1=trifft überhaupt nicht zu bis 4=trifft voll und ganz zu). Zusätzlich konnten die Patientinnen bei offenen Fragen Freitextantworten geben. Die papierbasierten Fragebögen wurden digitalisiert und mittels IBM SPSS Statistics for Windows, Version 27.0, deskriptiv analysiert. Die Freitextantworten wurden von zwei Codierern systematisch codiert und in thematische Kategorien eingeteilt. Die klinischen Daten und die Daten der Baselineerhebung aus der PRO

B Studie wurden verwendet, um die deskriptive Analyse und die Interpretation der Evaluationsbefragung zu ergänzen.

2.12.8 Evaluation der graphischen Visualisierung von PRO-Werten

Zur Evaluation der Verständlichkeit und Nutzbarkeit der graphischen Aufbereitung von PRO-Werten wurden gemeinsam mit Experten aus dem Bereich Health Care Service Design zunächst Mock-Ups neuer grafischer Aufbereitungen von PRO-Werten erstellt. Diese Designentwürfe basierten auf einer vorangegangenen Literaturrecherche und Erfahrungen des Studienteams aus dem Verlauf von PRO B und anderen Projekten mit Nutzung von PROs. Anschließend wurde ein darauf ausgerichteter Fragenkatalog für semi-strukturierte Interviews mit Patientinnen und klinischem Personal der PRO B Studie erstellt. Im Fokus standen hier Fragen zur Patientenzentrierung und Verständlichkeit für Fachpersonal und Laien sowie zu Präferenzen und Empfehlungen für einen Einsatz in der klinischen Routine. Durchgeführt wurden die Interviews anschließend via Videokonferenz, in denen den Teilnehmerinnen eine PowerPoint-Präsentation mit den Grafiken gezeigt wurde und sie entsprechend des Interviewleitfadens befragt wurden. Es wurden Audioaufnahmen der Interviews angefertigt, anhand derer anschließend Transkripte der Interviews angefertigt wurden. Die Transkripte wurden von zwei Codierern mit Hilfe eines inhaltsanalytischen Ansatzes nach Kuckartz strukturiert [34] und qualitativ ausgewertet. Die wichtigsten thematischen Kategorien wurden deduktiv abgeleitet, basierend auf den wichtigsten Designaspekten bei der Gestaltung von Graphen, welche mittels der Literaturrecherche identifiziert wurden (z. B. Beschriftungen, Farben, Richtungsabhängigkeit) sowie auf bekannten notwendigen Dimensionen für das Symptommonitoring (z. B. Alarmdarstellung). Innerhalb dieser Kategorien wurden induktiv Unterkategorien festgelegt.

Eine biometrische Fallzahlplanung wurde nicht vorgenommen, sondern basierend aus den Erfahrungen aus vorherigen qualitativen Studien mit einer maximalen Anzahl von 15 Interviews mit Patientinnen und 10 Interviews mit Mitgliedernd des Behandlungsteams geplant. Ausschlaggebend für die finale Stichprobengröße an Interviews war schließlich eine sich abzeichnende inhaltliche Sättigung. Auf Grundlage der in den Interviews gewonnen Erkenntnisse wurden anschließend die finalen Versionen der PRO-Graphiken erstellt. Diese wurden mit den individuellen Befragungsdaten aller PRO B Patientinnen in einem vollständigen Report inkl. einzelner Erläuterungstexte an die Studienpatientinnen versandt. Gemeinsam mit diesem Bericht wurde ein anonymisierter Fragebogen versandt, welcher mittels geschlossener Fragen quantitatives Feedback zur PRO Visualisierung und möglichen Konsequenzen für die PRO Berichterstattung einholte. Nach Digitalisierung der papierbasierten Rückläufer wurden die quantitativen Daten deskriptiv mit SPSS Statistics für Mac (Version 30.0.0.0, IBM Corp.) analysiert. Für nicht normalverteilte Daten zu Gruppenvergleichen wurden zusätzliche Analysen mit dem Mann-Whitney-U-Test durchgeführt. Eine formale Fallzahlkalkulation war nicht erforderlich, da die Umfrage nach Studienende der gesamten PRO B-Studienpopulation angeboten wurde. Da die Daten gemeinsam mit der Befragung zur Prozessevaluation erhoben wurden, galten dieselben Ausschlusskriterien (Tod, Widerruf der Einwilligung während der PRO B Studie und kein ausgefüllter PRO-Fragebogen in den letzten 6 Monaten; vgl. 2.12.7).

3 Ergebnisse der Evaluation

3.1 Studienpopulation

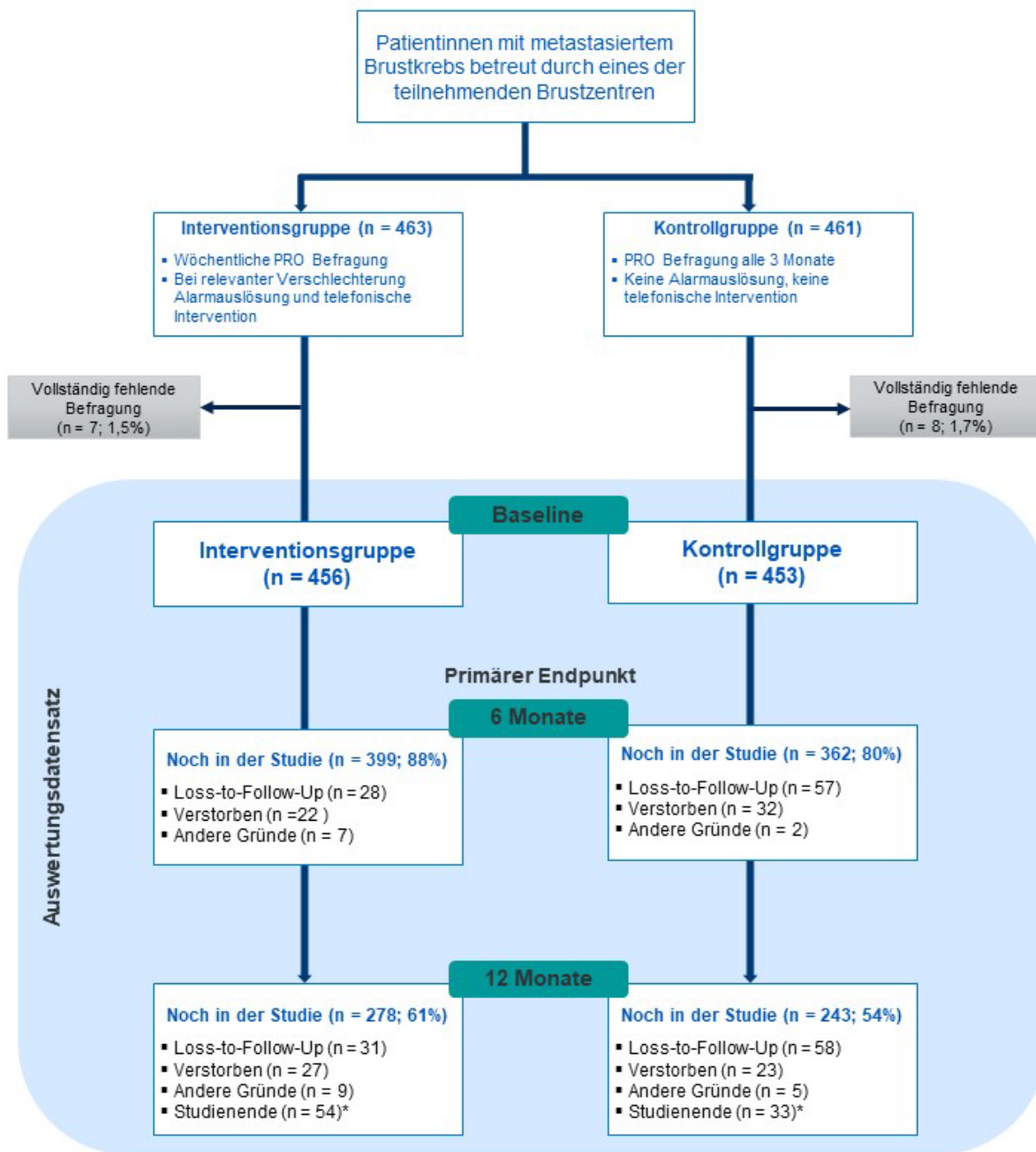
3.1.1 CONSORT

Die Studie begann mit der Rekrutierung im Mai 2021, mit dem ursprünglichen Ziel, bis Februar 2023 insgesamt 1.000 Patientinnen zu gewinnen. Aufgrund von Herausforderungen bei der Erreichung dieses Ziels wurde der Rekrutierungszeitraum jedoch um vier Monate verlängert.

PRO B (01NVF19013)

Die Rekrutierung wurde schließlich vor Erreichen des Ziels aufgrund der zeitlich beschränkten Förderperiode beendet. Durch zahlreiche Bemühungen und Maßnahmen zur Rekrutierung wurden letztendlich 924 Patientinnen in die PRO B Studie eingeschlossen (463 in der Interventionsgruppe und 461 in der Kontrollgruppe). Davon wurden 909 Patientinnen in die Hauptanalyse einbezogen (456 in der Interventionsgruppe und 453 in der Kontrollgruppe). 15 Fälle (7 in der Interventionsgruppe und 8 in der Kontrollgruppe) wurden von der Analyse ausgeschlossen, da sie nach Randomisierung keine Studienfragebögen während des gesamten Interventionszeitraumes beantwortete.

Insgesamt schlossen 761 Patientinnen (83,7 %) die Nachbeobachtung nach sechs Monaten und 521 Patientinnen (57,3 %) die Nachbeobachtung nach 12 Monaten ab. Die letzte Befragung fand am 15. Februar 2024 statt (Abbildung 5).



Loss-to-Follow-Up = Patientinnen, die die PRO Befragung nicht mehr beantwortet haben, ohne das Studienzentrum darüber zu informieren.

*Studienende = Patientinnen, die zu einem späten Zeitpunkt in die Studie eingeschlossen wurden.

Abbildung 5: CONSORT-Diagramm.

3.1.2 Drop-out

Tabelle 2 stellt die Abbruchraten sowie verschiedene Gründe für das Ausscheiden von Teilnehmerinnen aus der Studie dar. Die Daten sind zu vier Zeitpunkten (3, 6, 9 und 12 Monate) für die Gesamtkohorte (n = 909) sowie getrennt für die Interventionsgruppe (n = 456) und die Kontrollgruppe (n = 453) ausgewertet worden.

Nach drei Monaten betrug die Gesamtabbruchrate 7,9%. Die Interventionsgruppe wies mit 4,4% eine niedrigere Abbruchrate auf als die Kontrollgruppe mit 11,5%. Die Abbruchrate erhöhte sich auf 16,3 % gesamt nach sechs Monaten, wobei die Interventionsgruppe eine Rate von 12,5% und die Kontrollgruppe eine Rate von 20,1% aufwies. Hauptgründe für den Abbruch waren: Fehlende Nachbeobachtung (Loss-to-Follow-Up) (9,4% gesamt) und Versterben (5,9% gesamt). Nach einem Jahr lag die Gesamtabbruchrate bei 42,7%, wobei 39,0% der Teilnehmerinnen in der Interventionsgruppe und 46,4% in der Kontrollgruppe die Studie abbrachen. Die häufigsten Abbruchgründe nach 12 Monaten waren: Fehlende Nachbeobachtung (19,1% gesamt), Versterben (11,4% gesamt) und das reguläre Studieneende (9,6% gesamt).

Die Analyse zeigt, dass die Kontrollgruppe durchweg höhere Abbruchraten als die Interventionsgruppe aufwies. Die Hauptgründe für das Ausscheiden aus der Studie waren fehlende Nachbeobachtungen und Versterben. Insgesamt zeigt sich, dass die Abbruchrate über die Zeit zunahm, mit der höchsten Rate nach 12 Monaten.

Tabelle 2: Anzahl der Dropouts

Follow-Up Zeit*	Gesamt (n = 909)	Interventionsgruppe (n = 456)	Kontrollgruppe (n = 453)
3 Monate			
Verstorben	20 (2,2%)	8 (1,8%)	12 (2,6%)
Rücknahme der Zustimmung	2 (0,2%)	1 (0,2%)	1 (0,2%)
Kein Interesse mehr	1 (0,1%)	1 (0,2%)	0 (0,0%)
Fehlende Nachbeobachtung ¹	48 (5,3%)	9 (2,0%)	39 (8,6%)
Wechsel der Klinik	1 (0,1%)	1 (0,2%)	0 (0,0%)
Gesamtabbruch	72 (7,9%)	20 (4,4%)	52 (11,5%)
6 Monate			
Verstorben	54 (5,9%)	22 (4,8%)	32 (7,1%)
Rücknahme der Zustimmung	3 (0,3%)	2 (0,4%)	1 (0,2%)
Überforderung	1 (0,1%)	1 (0,2%)	0 (0,0%)
Kein Interesse mehr	3 (0,3%)	3 (0,7%)	0 (0,0%)
Fehlende Nachbeobachtung ¹	85 (9,4%)	28 (6,1%)	57 (12,6%)
Wechsel der Klinik	2 (0,2%)	1 (0,2%)	1 (0,2%)
Gesamtabbruch	148 (16,3%)	57 (12,5%)	91 (20,1%)
9 Monate			
Verstorben	85 (9,4%)	41 (9,0%)	44 (9,7%)
Rücknahme der Zustimmung	5 (0,6%)	4 (0,9%)	1 (0,2%)
Gesundheitliche Gründe	1 (0,1%)	1 (0,2%)	0 (0,0%)
Überforderung	3 (0,3%)	3 (0,7%)	0 (0,0%)
Kein Interesse mehr	5 (0,6%)	3 (0,7%)	2 (0,4%)
Fehlende Nachbeobachtung ¹	135 (14,9%)	43 (9,4%)	92 (20,3%)
Wechsel der Klinik	3 (0,3%)	2 (0,4%)	1 (0,2%)
Studienende ²	36 (4,0%)	22 (4,8%)	14 (3,1%)
Gesamtabbruch	273 (30,0%)	119 (26,1%)	154 (34,0%)

12 Monaten			
Verstorben	104 (11,4%)	49 (10,7%)	55 (12,1%)
Rücknahme der Zustimmung	6 (0,7%)	5 (1,1%)	1 (0,2%)
Gesundheitliche Gründe	3 (0,3%)	1 (0,2%)	2 (0,4%)
Überforderung	3 (0,3%)	3 (0,7%)	0 (0,0%)
Kein Interesse mehr	8 (0,9%)	5 (1,1%)	3 (0,7%)
Fehlende Nachbeobachtung ¹	174 (19,1%)	59 (12,9%)	115 (25,4%)
Wechsel der Klinik	3 (0,3%)	2 (0,4%)	1 (0,2%)
Studienende ²	87 (9,6%)	54 (11,8%)	33 (7,3%)
Gesamtabbruch	388 (42,7%)	178 (39,0%)	210 (46,4%)

¹ Loss-to-Follow-Up: Patientinnen, die die PRO-Befragung ab diesem Zeitpunkt nicht mehr beantwortet haben und das Studienzentrum nicht über ihre Absicht informiert haben, aus der Studie auszusteigen.

² Patientinnen, die zu einem späten Zeitpunkt rekrutiert wurden.

*Alle Gründe und die Gesamtzahl der Abbrüche werden als kumulative Zahl und als Prozentsatz angegeben.

3.1.3 Fehlende Werte

Tabelle 3 beschreibt eine Untersuchung der fehlenden PRO-Befragungsdaten. Betrachtet wurden vier Gruppen: monoton fehlende Daten (Tod), nicht monoton fehlende Daten (Dropout), beobachtete Fälle sowie die Gesamtpopulation. Die Gruppe mit monoton fehlenden Daten (Tod) umfasste 54 Patientinnen, während die Gruppe mit nicht monoton fehlenden Daten 111 Patientinnen umfasste. Die Gruppe mit beobachteten Daten bestand aus 650 Patientinnen, während die Dropout-Gruppe 94 Patientinnen umfasste.

Das Durchschnittsalter lag zwischen 48 und 52 Jahren, wobei die Dropout-Gruppe das jüngste Durchschnittsalter (48 Jahre) auswies.

Der Anteil an HR-/HER+ oder HR-/HER-Tumoren war in der Todesfall-Gruppe am höchsten (33,3%) und in der Gruppe der beobachteten Fälle am niedrigsten (18,0%).

Gehirn- oder multiple Metastasen waren in der Todesfall-Gruppe am häufigsten (68,5%), verglichen mit 54,3% in der Dropout-Gruppe. Ossäre, Lymphknoten- oder kutane Metastasen waren in der Dropout-Gruppe und in der Gruppe der beobachteten Fälle am häufigsten (~30%), und in der Todesfall-Gruppe am seltensten (18,5%). Viszerale Metastasen waren in allen Gruppen mit 13-17% etwa gleich verteilt.

Der Anteil an Patientinnen mit niedriger Bildung war in allen Gruppen relativ gering (~9-13%). Die Erwerbstätigkeit war in der Gruppe der beobachteten Fälle am höchsten (55,0%) und in der Todesfall-Gruppe am niedrigsten (44,9%). Der Anteil an Patientinnen mit einem Haushaltseinkommen unter 2500€ war in der Todesfall-Gruppe am höchsten (45,5%) und in der Gruppe der beobachteten Fälle am niedrigsten (33,6%).

Endokrine Therapie war in der Dropout-Gruppe am häufigsten (63,8%) vertreten, während in der Todesfall-Gruppe nur 25,9% endokrine Therapie erhielten. Chemotherapie war in der Todesfall-Gruppe am häufigsten (64,8%) und in der Gruppe der beobachteten Fälle am seltensten (47,4%) vertreten. Anti-HER2-Therapie wurde in der Dropout-Gruppe am häufigsten durchgeführt (31,9%), jedoch nur bei 9,3% in der Todesfall-Gruppe. Immuntherapie wurde nur selten angewendet (2,9% in der Gruppe der beobachteten Fälle, 3,7% in der Todesfall-Gruppe). Zielgerichtete Therapien waren in der Gruppe mit beobachteten Daten am häufigsten (37,5%).

Die Todesfall-Gruppe wies die niedrigsten Werte für körperliche (35,9), emotionale (38,6) und soziale Funktionsfähigkeit (37,6) auf. Fatigue war in der Todesfall-Gruppe am höchsten (63,7), während die Gruppe der beobachteten Fälle den niedrigsten Wert aufwies (57,8). Übelkeit, Schmerzen und Dyspnoe waren in der Todesfall-Gruppe am ausgeprägtesten.

Die Analyse zeigt, dass Patientinnen mit monoton fehlenden Daten (Tod) tendenziell schwerere Erkrankungen aufwiesen, was sich in höheren Raten an Gehirnmetastasen und

einer stärkeren Einschränkung der Selbstversorgung widerspiegelt. Unterschiede in sozioökonomischen Faktoren und Behandlungsformen sind ebenfalls erkennbar. Die Todesfall-Gruppe zeigte eine höhere Prävalenz von aggressiven Tumoren (HR-/HER+) und schwereren Metastasen (z. B. Gehirnetastasen). Die Gruppe der beobachteten Fälle war sozioökonomisch bessergestellt (höheres Einkommen, höhere Erwerbstätigkeit) und hatte allgemein bessere Lebensqualitätswerte. Unterschiede in der Therapie zeigen, dass in der Todesfall-Gruppe häufiger aggressive Behandlungen wie Chemotherapie zum Einsatz kamen, während in der Gruppe der beobachteten Fälle endokrine Therapien bevorzugt wurden.

Tabelle 3: Exploration der fehlenden PRO-Scores nach sechs Monaten nach Patientenmerkmalen

Patientenmerkmale	Monoton fehlende Daten ³		Nicht monoton fehlende Daten ⁴	Beobachtete Fälle
	Tod	Dropout		
	n=54	n=94	n=111	n=650
Alter (Jahr) – Mean (SD)	52 (10)	48 (11)	50 (11)	51 (11)
Education				
Niedrig	5 (10,2%)	11 (12,6%)	9 (8,4%)	60 (9,3%)
Mittel	26 (53,1%)	45 (51,7%)	56 (52,3%)	359 (55,5%)
Hoch	18 (36,7%)	31 (35,6%)	42 (39,3%)	228 (35,2%)
Histologischer Befund				
HR+/HER- oder HR+/HER+	36 (66,7%)	73 (77,7%)	84 (75,7%)	533 (82,0%)
HR-/HER+ oder HR-/HER-	18 (33,3%)	21 (22,3%)	27 (24,3%)	117 (18,0%)
Arten der Fernmetastasierung				
Gehirnetastasen oder multiple Metastasen	37 (68,5%)	51 (54,3%)	62 (55,9%)	353 (54,3%)
Ossäre Metastasen oder Lymphknotenmetastasen oder kutane Metastasen	10 (18,5%)	29 (30,9%)	30 (27,0%)	203 (31,2%)
Viszerale Metastasen (nur ein Organ betreffend)	7 (13,0%)	14 (14,9%)	19 (17,1%)	94 (14,5%)
Erwerbstätig	22 (44,9%)	45 (51,7%)	56 (52,3%)	356 (55,0%)
Deutsche Staatsangehörigkeit	45 (91,8%)	68 (78,2%)	95 (88,8%)	608 (94,0%)
Familienstand				
Ledig	7 (14,3%)	7 (8,0%)	15 (14,0%)	89 (13,8%)
Verheiratet/ in eingetragener Lebenspartnerschaft zusammenlebend	32 (65,3%)	58 (66,7%)	70 (65,4%)	421 (65,1%)
Getrennt lebend/geschieden/verwitwet	10 (20,4%)	22 (25,3%)	22 (20,6%)	137 (21,2%)
Kinder 14-jährige im Haushalt	7 (14,3%)	21 (24,1%)	15 (14,0%)	105 (16,2%)

³ „Monoton fehlende Daten“ (monotone missing data) beziehen sich auf ein bestimmtes Muster fehlender Werte in Datensätzen. Die Daten fehlen auf eine systematische Weise: Wenn eine Variable fehlt, fehlen auch alle nachfolgenden Variablen; Z.B.: Wenn ein Patient den PRO-Score in Woche 4 nicht ausfüllt, fehlen auch die Werte für Woche 8 und Woche 12.

⁴ "Nicht monoton fehlende Daten" (non-monotone missing data) treten auf, wenn fehlende Werte in einem Datensatz kein eindeutiges, sequentielles Muster haben. Das bedeutet, dass fehlende Daten unregelmäßig verteilt sind und nicht notwendigerweise von einer vorherigen fehlenden Variablen abhängen; Z.B.: Ein Patient füllt den PRO-Score in Woche 4 aus, überspringt aber Woche 8, gibt jedoch wieder eine Antwort in Woche 12.

Patientenmerkmale	Monoton fehlende Daten ³		Nicht monoton fehlende Daten ⁴	Beobachtete Fälle
	Tod	Dropout		
	n=54	n=94	n=111	n=650
Einkommen Haushalt (Monatlich)				
<2500	20 (45,5%)	26 (38,2%)	34 (39,1%)	185 (33,6%)
2500-4999	15 (34,1%)	34 (50,0%)	38 (43,7%)	256 (46,5%)
≥5000	9 (20,5%)	8 (11,8%)	15 (17,2%)	109 (19,8%)
Depression	0 (0,0%)	11 (12,6%)	14 (13,2%)	69 (10,6%)
NCI Index	0,1 (0,3)	0,1 (0,3)	0,1 (0,2)	0,1 (0,2)
Therapie bei Randomisierung				
Endokrine Therapie	14 (25,9%)	60 (63,8%)	54 (48,6%)	408 (62,8%)
Chemotherapie	35 (64,8%)	45 (47,9%)	54 (48,6%)	308 (47,4%)
Anti-Her2-Therapie	5 (9,3%)	30 (31,9%)	28 (25,2%)	146 (22,5%)
Immuntherapie	2 (3,7%)	5 (5,3%)	6 (5,4%)	19 (2,9%)
Gezielte Therapie	13 (24,1%)	19 (20,2%)	41 (36,9%)	244 (37,5%)
Andere Therapien	9 (16,7%)	8 (8,5%)	28 (25,2%)	78 (12,0%)
Randomisierung vor oder nach Änderung des Alarmkriteriums				
vor dem 01.09.2022	38 (70,4%)	45 (47,9%)	73 (65,8%)	441 (67,8%)
nach dem 01.09.2022	16 (29,6%)	49 (52,1%)	38 (34,2%)	209 (32,2%)
Gesundheitszustand				
Normale, uneingeschränkte Aktivität	2 (4,1%)	22 (25,3%)	13 (12,3%)	141 (21,8%)
Einschränkung bei körperlicher Anstrengung, gehfähig, leichte körperliche Arbeit möglich	25 (51,0%)	44 (50,6%)	58 (54,7%)	369 (56,9%)
Gehfähig, Selbstversorgung möglich, aber nicht arbeitsfähig, kann mehr als 50% der Wachzeit aufstehen	16 (32,7%)	14 (16,1%)	24 (22,6%)	115 (17,7%)
Nur begrenzte Selbstversorgung möglich, 50% der Wachzeit oder mehr an Bett oder Stuhl gebunden	6 (12,2%)	6 (6,9%)	11 (10,4%)	22 (3,4%)
Völlig pflegebedürftig, keinerlei Selbstversorgung möglich, völlig an Bett oder Stuhl gebunden	0 (0,0%)	1 (1,1%)	0 (0,0%)	1 (0,2%)
EORTC T-Score bei Studienbeginn				
Globaler Gesundheitszustand/QoL	34,1 (3,3)	36,1 (4,8)	36,1 (4,5)	36,0 (3,7)
Körperliche Funktionsfähigkeit	35,9 (9,3)	40,9 (10,4)	40,4 (9,0)	41,8 (9,2)
Funktionsfähigkeit der Rolle	34,9 (9,5)	39,5 (11,1)	38,4 (10,2)	41,1 (10,0)
Emotionale Funktionsfähigkeit	38,6 (8,4)	39,5 (8,1)	39,7 (8,8)	42,1 (8,8)
Kognitive Funktionsfähigkeit	42,7 (10,0)	42,1 (10,5)	41,6 (11,0)	44,3 (9,8)
Soziale Funktionsfähigkeit	37,6 (7,8)	41,3 (9,0)	39,0 (8,4)	42,0 (8,9)
Fatigue	63,7 (8,3)	60,4 (9,8)	61,3 (8,8)	57,8 (9,3)
Übelkeit und Erbrechen	63,5 (11,2)	61,5 (11,5)	60,6 (11,2)	58,5 (10,3)
Schmerzen	56,9 (10,1)	55,6 (10,1)	56,8 (10,4)	54,2 (9,8)

Patientenmerkmale	Monoton fehlende Daten ³		Nicht monoton fehlende Daten ⁴	Beobachtete Fälle
	Tod	Dropout		
	n=54	n=94	n=111	n=650
Dyspnoe	64,1 (7,6)	59,3 (10,7)	61,0 (10,0)	57,4 (10,7)
Schlaflosigkeit	57,6 (7,5)	58,8 (8,7)	59,6 (9,1)	56,9 (8,5)
Appetitlosigkeit	62,2 (13,3)	57,7 (11,2)	56,9 (11,2)	53,7 (11,3)
Verstopfung	57,7 (11,1)	54,2 (11,5)	54,7 (11,4)	52,2 (10,6)
Durchfall	55,7 (11,3)	55,6 (11,3)	56,7 (12,7)	55,1 (11,7)
Finanzielle Schwierigkeiten	55,1 (11,7)	55,9 (12,2)	57,5 (11,1)	53,0 (10,0)

3.1.4 Baseline-Charakteristika der Patientinnen vor und nach der multiplen Imputation

Die multiple Imputation hatte keinen wesentlichen Einfluss auf die Verteilung der soziodemografischen und klinischen Merkmale der Studienteilnehmerinnen. Die Ergebnisse zeigen eine weitgehend stabile Verteilung zwischen beobachteten und imputierten Daten. Die erhobenen Daten geben einen detaillierten Einblick in die Merkmale der Patientenpopulation und ermöglichen eine zuverlässige Analyse der Einflussfaktoren auf die Lebensqualität und den Krankheitsverlauf (Tabelle 4).

Die Anzahl der fehlenden Werte betrug insgesamt 19 (2,09%), mit einer ähnlichen Verteilung in der Kontrollgruppe (2,00%) und der Interventionsgruppe (2,19%). Mittels der multiplen Imputation wurden diese fehlenden Werte ausgeglichen, sodass alle Daten vollständig vorlagen. Die Mehrheit der Patientinnen hatte eine mittlere Bildung (54,6%), gefolgt von 35,8% mit hoher Bildung und 9,6% mit niedriger Bildung. Nach der Imputation blieben diese Verhältnisse weitgehend konstant. Etwa 53,8% der Patientinnen waren zum Zeitpunkt der Studie erwerbstätig, während 46,2% nicht erwerbstätig waren. Auch nach der Imputation zeigten sich keine wesentlichen Veränderungen in diesen Kategorien. Ein Großteil der Studienteilnehmerinnen hatte keinen Migrationshintergrund (91,7%), während 8,3% einen Migrationshintergrund aufwiesen. Hinsichtlich des Familienstands waren 65,3% der Patientinnen verheiratet oder lebten in einer eingetragenen Partnerschaft, 13,3% waren ledig und 21,5% getrennt, geschieden oder verwitwet. Die multiplen Imputationen führten zu minimalen Änderungen in den Prozentwerten. Beim Familieneinkommen lag der größte Anteil der Patientinnen in der Einkommensgruppe von 2500–4999€ (45,8%), gefolgt von <2500€ (35,4%) und ≥5000€ (18,8%). Nach der Imputation gab es leichte Verschiebungen, insbesondere bei den niedrigeren Einkommensgruppen. 89,4% der Teilnehmerinnen berichteten, keine Depression zu haben, während 10,6% angaben, an einer Depression zu leiden. Diese Werte blieben nach der Imputation nahezu unverändert. In Bezug auf den Gesundheitszustand hatten 20,0% der Patientinnen eine normale, uneingeschränkte Aktivität, während 55,7% eine Einschränkung bei körperlicher Anstrengung aufwiesen. Nach der Imputation veränderten sich diese Werte nur minimal.

Die subjektive Einschätzung der Lebensqualität in der letzten Woche zeigte, dass 6,5% eine deutliche Verbesserung berichteten, 21,9% eine leichte Verbesserung und 60,5% keine Veränderung wahrnahmen. 10,5% der Patientinnen gaben eine Verschlechterung an.

Bei der Bewertung der körperlichen und emotionalen Funktionsfähigkeit lagen die Werte für den globalen Gesundheitszustand im Durchschnitt bei 35,9 Punkten. Fatigue (58,8%), Schmerzen (54,8%) und Schlaflosigkeit (57,4 %) waren häufige Symptome.

Tabelle 4: Baseline-Charakteristika der Patientinnen zwischen beobachteten Daten und nach multipler Imputation (imputierter Datensatz, m = 30)

Baseline-Charakteristika der Patientinnen	Beobachtete Daten			Multiple Imputation (m = 30)		
	Gesamt (n=909)	Kontrolle (n=453)	Intervention (n=456)	Gesamt (n=909)	Kontrolle (n=453)	Intervention (n=456)
Anzahl fehlender Daten (%)	n=19 (2,09%)	n=9 (2,00%)	n=10 (2,19%)			
Education						
Niedrig	85 (9,6%)	42 (9,5%)	43 (9,6%)	9,8%	9,7%	9,8%
Mittel	486 (54,6%)	257 (57,9%)	229 (51,3%)	54,5%	57,5%	51,4%
Hoch	319 (35,8%)	145 (32,7%)	174 (39,0%)	35,7%	32,7%	38,7%
Erwerbstätig						
Nein	411 (46,2%)	207 (46,6%)	204 (45,7%)	46,3%	46,7%	46,0%
Ja	479 (53,8%)	237 (53,4%)	242 (54,3%)	53,7%	53,3%	54,0%
Migration						
Kein Migrationshintergrund	816 (91,7%)	402 (90,5%)	414 (92,8%)	91,3%	90,2%	92,5%
Migrationshintergrund	74 (8,3%)	42 (9,5%)	32 (7,2%)	8,7%	9,8%	7,5%
Familienstand						
Ledig	118 (13,3%)	49 (11,0%)	69 (15,5%)	13,4%	11,1%	15,7%
Verheiratet/in eingetragener Lebenspartnerschaft zusammenlebend	581 (65,3%)	295 (66,4%)	286 (64,1%)	65,1%	66,2%	63,9%
Getrennt lebend/geschieden/verwitwet	191 (21,5%)	100 (22,5%)	91 (20,4%)	21,5%	22,6%	20,4%
Kinder 14-jährige im Haushalt						
Nein	742 (83,4%)	369 (83,1%)	373 (83,6%)	83,3%	82,9%	83,8%
Ja	148 (16,6%)	75 (16,9%)	73 (16,4%)	16,7%	17,1%	16,2%
Einkommen Haushalt (Monatlich)						
< 2500	265 (35,4%)	129 (34,6%)	136 (36,2%)	37,3%	35,9%	38,8%
2500-4999	343 (45,8%)	179 (48,0%)	164 (43,6%)	44,2%	46,6%	41,8%

Baseline-Charakteristika der Patientinnen	Beobachtete Daten			Multiple Imputation (m = 30)		
	Gesamt (n=909)	Kontrolle (n=453)	Intervention (n=456)	Gesamt (n=909)	Kontrolle (n=453)	Intervention (n=456)
≥ 5000	141 (18,8%)	65 (17,4%)	76 (20,2%)	18,4%	17,5%	19,4%
Depression						
Nein	796 (89,4%)	404 (91,2%)	392 (87,7%)	89,3%	90,9%	87,7%
Ja	94 (10,6%)	39 (8,8%)	55 (12,3%)	10,7%	9,1%	12,3%
Gesundheitszustand						
Normale, uneingeschränkte Aktivität	178 (20,0%)	82 (18,5%)	96 (21,5%)	19,8%	18,3%	21,3%
Einschränkung bei körperlicher Anstrengung, leichte körperliche Arbeit möglich	496 (55,7%)	256 (57,8%)	240 (53,7%)	55,6%	57,8%	53,5%
Gehfähig, Selbstversorgung möglich, aber nicht arbeitsfähig	169 (19,0%)	83 (18,7%)	86 (19,2%)	19,1%	18,9%	19,4%
Nur begrenzte Selbstversorgung möglich	45 (5,1%)	20 (4,5%)	25 (5,6%)	5,2%	4,6%	5,7%
Völlig pflegebedürftig	2 (0,2%)	2 (0,5%)	0 (0,0%)	0,2%	0,4%	
Anzahl fehlender Daten (%)	n=54 (5,94%)	n=28 (6,18%)	n=26 (5,70%)			
Lebensqualität im Vergleich der letzten Woche						
Viel besser	56 (6,5%)	28 (6,6%)	28 (6,5%)	6,6%	6,8%	6,4%
Etwas besser	187 (21,9%)	87 (20,5%)	100 (23,3%)	21,8%	20,6%	23,0%
Etwa wie vor einer Woche	517 (60,5%)	261 (61,4%)	256 (59,5%)	60,5%	61,2%	59,7%
Etwas schlechter	90 (10,5%)	45 (10,6%)	45 (10,5%)	10,5%	10,5%	10,6%
Viel schlechter	5 (0,6%)	4 (0,9%)	1 (0,2%)	0,6%	0,9%	0,2%
EORTC T-Score bei Studienbeginn						
Globaler Gesundheitszustand/QoL	35,9 (3,9)	36,0 (4,3)	35,9 (3,6)	35,9 (3,9)	36,0 (4,2)	35,9 (3,5)
Körperliche Funktionsfähigkeit	41,2 (9,4)	41,0 (9,1)	41,4 (9,7)	41,1 (9,4)	41,0 (9,0)	41,2 (9,7)
Funktionsfähigkeit der Rolle	40,3 (10,2)	39,7 (9,9)	40,8 (10,5)	40,2 (10,1)	39,7 (9,8)	40,6 (10,4)
Emotionale Funktionsfähigkeit	41,3 (8,8)	40,4 (8,4)	42,3 (9,1)	41,2 (8,8)	40,2 (8,4)	42,2 (9,1)

Baseline-Charakteristika der Patientinnen	Beobachtete Daten			Multiple Imputation (<i>m</i> = 30)		
	Gesamt (n=909)	Kontrolle (n=453)	Intervention (n=456)	Gesamt (n=909)	Kontrolle (n=453)	Intervention (n=456)
Kognitive Funktionsfähigkeit	43,7 (10,0)	42,5 (9,9)	44,8 (10,0)	43,5 (10,0)	42,4 (9,9)	44,6 (10,0)
Soziale Funktionsfähigkeit	41,3 (8,9)	40,2 (8,0)	42,4 (9,5)	41,2 (8,8)	40,2 (8,0)	42,3 (9,5)
Fatigue	58,8 (9,4)	60,2 (9,0)	57,5 (9,5)	58,9 (9,3)	60,2 (9,0)	57,6 (9,5)
Übelkeit und Erbrechen	59,3 (10,6)	61,3 (11,8)	57,4 (9,0)	59,3 (10,7)	61,3 (11,8)	57,4 (9,1)
Schmerzen	54,8 (10,0)	55,7 (10,2)	53,9 (9,6)	54,9 (10,0)	55,8 (10,2)	54,0 (9,6)
Dyspnoe	58,4 (10,6)	60,0 (10,1)	56,7 (10,8)	58,4 (10,6)	59,9 (10,1)	56,9 (10,8)
Schlaflosigkeit	57,4 (8,6)	58,7 (8,6)	56,2 (8,5)	57,5 (8,6)	58,8 (8,6)	56,2 (8,5)
Appetitlosigkeit	55,0 (11,6)	56,9 (12,1)	53,1 (10,9)	55,0 (11,6)	56,9 (12,1)	53,1 (10,9)
Verstopfung	53,0 (10,9)	54,3 (11,3)	51,7 (10,4)	53,0 (10,9)	54,2 (11,2)	51,8 (10,4)
Durchfall	55,3 (11,7)	57,3 (12,2)	53,4 (10,9)	55,4 (11,8)	57,3 (12,2)	53,4 (11,0)
Finanzielle Schwierigkeiten	54,0 (10,5)	55,9 (10,9)	52,1 (9,8)	54,1 (10,6)	56,0 (10,9)	52,1 (9,9)

3.1.5 Rücklaufquote

Die mediane Nachbeobachtungszeit betrug 64 Wochen (Interquartilsabstand (interquartile range, IQR) 37 - 100) in der Interventionsgruppe und 52 Wochen (IQR 26 - 91) in der Kontrollgruppe. Diese Differenz ist auf das Studiendesign zurückzuführen: In der Interventionsgruppe erfolgte eine wöchentliche PRO-Erhebung, während die Kontrollgruppe im Rahmen der Regelversorgung in größeren Abständen beobachtet wurde. Die längere Nachbeobachtung spiegelt somit die Studienstruktur wider und sollte bei der Ergebnisinterpretation berücksichtigt werden, da dadurch mehr PRO-Daten erfasst werden konnten.

Insgesamt wurden 39.817 Fragebögen verschickt, 36.845 an die Interventionsgruppe und 2.972 an die Kontrollgruppe. Die Gesamtrücklaufquote betrug 66,7% (24.571/36.845) in der Interventionsgruppe und 62,9% (1.869/2.972) in der Kontrollgruppe. Abbildung 6 zeigt die Rückläuferquote ("%PRO Response") über verschiedene Zeitpunkte für die Kontrollgruppe (orange) und die Interventionsgruppe (grün). Zu Beginn (Baseline) sind die Prozentsätze in beiden Gruppen identisch (94%). Über die Zeit sinkt die Anzahl der Fälle in beiden Gruppen, wobei die Interventionsgruppe stets höhere Prozentwerte aufweist als die Kontrollgruppe. Nach drei Monaten beträgt die Antwortrate 92% (401/436) in der Interventionsgruppe und 78% (307/394) in der Kontrollgruppe. Nach sechs Monaten sind es 89% bzw. 74% (262/352), nach 12 Monaten 90% (250/279) bzw. 73% (174/237). Dieser Trend setzt sich bis zum Zeitpunkt 24-Monate fort, zu dem die Interventionsgruppe 95% (98/103) und die Kontrollgruppe 92% (72/78) erreicht. Insgesamt zeigt die Grafik, dass die Interventionsgruppe durchgehend höhere PRO-Antwortraten aufweist als die Kontrollgruppe.

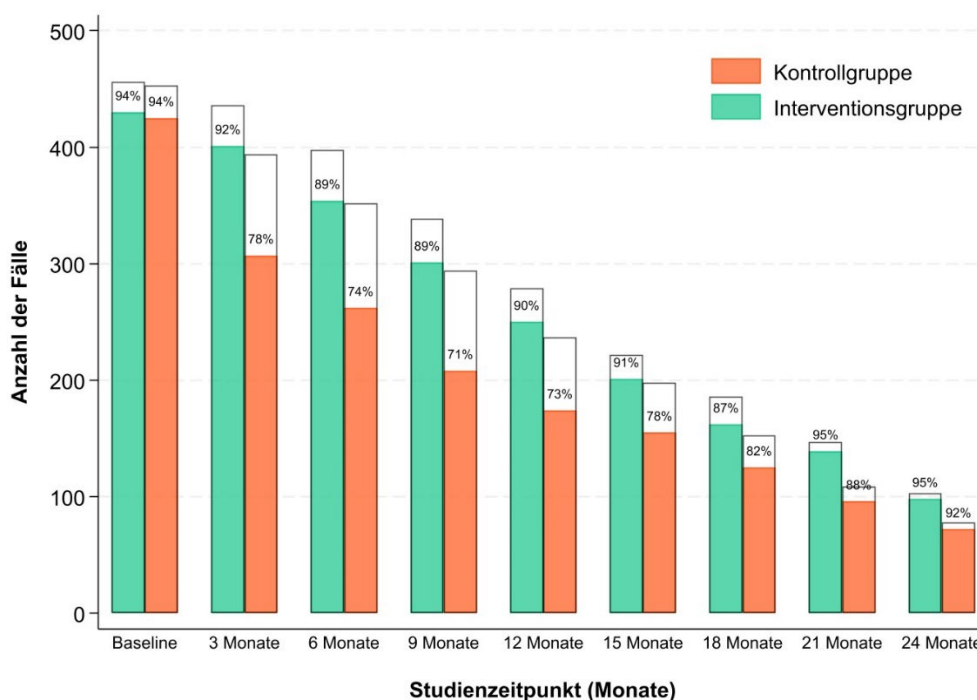


Abbildung 6: Rücklaufquote

3.1.6 Beschreibung der Studienpopulation

Verteilung der stratifizierten Faktoren für die Randomisierung

Die Studie umfasste insgesamt 909 Teilnehmerinnen, die in eine Kontrollgruppe (n=453) und eine Interventionsgruppe (n=456) aufgeteilt waren. Die Verteilung der histologischen Befunde zeigt, dass der Großteil der Teilnehmerinnen (79,9%) HR+/HER- oder HR+/HER+ Tumoren aufwies, während 20,1% HR-/HER+ oder HR-/HER- Tumoren hatten. Bezüglich der Art der Fernmetastasierung hatten 55,3% der Teilnehmerinnen Gehirnmastasen oder multiple Metastasen. Etwa 29,9% litten an ossären, Lymphknoten- oder kutanen Metastasen, während 14,7% viszerale Metastasen mit Befall nur eines Organs aufwiesen. Die Verteilung der Krankenkassenzugehörigkeit zeigt, dass 27,7% der Teilnehmerinnen bei BARMER, DAK Gesundheit oder mkk – meine krankenkasse versichert waren, während 72,3% anderen Krankenkassen angehörten. Die Kontroll- und Interventionsgruppe wiesen hierbei sehr ähnliche Verteilungen auf. Der Prozess der stratifizierten Randomisierung gewährleistete eine ausgewogene Verteilung wichtiger Faktoren zwischen der Kontroll- und der Interventionsgruppe. Die demografischen und klinischen Merkmale, einschließlich histologischer Befunde, Metastasierungsarten und Krankenkassenzugehörigkeiten, waren zwischen den Gruppen vergleichbar (Tabelle 5).

Tabelle 5: Stratifizierte Randomisierung

	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
Histologischer Befund			
HR+/HER- oder HR+/HER+	726 (79,9%)	360 (79,5%)	366 (80,3%)
HR-/HER+ oder HR-/HER-	183 (20,1%)	93 (20,5%)	90 (19,7%)
Arten der Fernmetastasierung			
Gehirnmetastasen oder multiple Metastasen	503 (55,3%)	250 (55,2%)	253 (55,5%)
Ossäre Metastasen oder Lymphknotenmetastasen oder kutane Metastasen	272 (29,9%)	135 (29,8%)	137 (30,0%)
Viszerale Metastasen (nur ein Organ betreffend)	134 (14,7%)	68 (15,0%)	66 (14,5%)
Krankenkassen			
BARMER, DAK Gesundheit, mkk – meine krankenkasse	252 (27,7%)	125 (27,6%)	127 (27,9%)
Andere Krankenkassen	657 (72,3%)	328 (72,4%)	329 (72,1%)

Soziodemografische Daten

Die soziodemografische Analyse zeigt, dass die Mehrheit der Teilnehmerinnen (97,9%) den Baseline-Fragebogen ausgefüllt hatte, wobei nur ein geringer Anteil (2,1%) fehlende Daten aufwies. Die deutsche Staatsangehörigkeit hatten 95,5% der Teilnehmerinnen, und 91,7% wurden in Deutschland geboren. 65,3% der Teilnehmerinnen waren verheiratet oder lebten in einer eingetragenen Lebenspartnerschaft, während 13,3% ledig waren und 12,0% geschieden oder getrennt lebten. Der Bildungsstand zeigt eine breite Verteilung: 32,2% der Teilnehmerinnen hatten das Abitur oder eine gleichwertige Hochschulreife, während 31,2% über einen Realschulabschluss verfügten. Lediglich 0,9% hatten keinen Schulabschluss. Im Bereich der beruflichen Qualifikation hatten 33,9% eine beruflich-betriebliche Ausbildung abgeschlossen, 19,8% verfügten über einen Hochschulabschluss auf Diplom-, Master- oder Magister-Niveau, und 2,0% waren promoviert. Ein kleiner Anteil (6,6%) besaß keinen beruflichen Abschluss. In Bezug auf die Erwerbssituation waren 21,1% der Teilnehmerinnen

vollzeitbeschäftigt, während 24,7% in Teilzeit arbeiten. Ein erheblicher Anteil von 44,3% war nicht erwerbstätig (einschließlich Rentnerinnen, Studentinnen und Arbeitslose). Die Mehrheit der Teilnehmerinnen (80,4%) lebte in einem Mehrpersonenhaushalt, mit einer medianen Haushaltsgröße von zwei Personen (IQR 2-3). Zudem hatten 16,6% Kinder unter 14 Jahren im Haushalt. Hinsichtlich des Einkommens lag der größte Anteil (12,8%) in der Kategorie von 3.000 bis unter 4.000 Euro, während 27,6% der Angaben nicht relevant oder nicht vorhanden waren. Die finanzielle Belastung wurde von 35,6% als nicht vorhanden und von 35,8% als gering empfunden, während 7,8% sie als sehr belastend einschätzten. Bei der Mobilität zeigte sich, dass 77,4% der Teilnehmerinnen das Auto als Hauptverkehrsmittel nutzten, während 13,9% öffentliche Verkehrsmittel bevorzugten. 44,9% der Befragten fuhren selbst Auto, während 32,5% gefahren wurden. Die soziodemografische Analyse zeigt, dass die Mehrheit der Teilnehmerinnen einen mittleren bis höheren Bildungs- und Berufsabschluss besaß, wobei eine ausgewogene Verteilung hinsichtlich der Erwerbssituation und des Einkommens vorlag (Tabelle 6).

Tabelle 6: Soziodemografische Daten

Soziodemografische Daten	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
Antwort auf den Baseline-Fragebogen			
Beantwortet	890 (97,9%)	444 (98,0%)	446 (97,8%)
Fehlender Baseline-Fragebogen	19 (2,1%)	9 (2,0%)	10 (2,2%)
Deutsche Staatsangehörigkeit			
Geburtsort Deutschland	850 (95,5%)	422 (95,0%)	428 (96,0%)
Geburtsort Deutschland			
Familienstand			
Verheiratet/ in eingetragener Lebenspartnerschaft zusammenlebend	581 (65,3%)	295 (66,4%)	286 (64,1%)
Verheiratet/ in eingetragener Lebenspartnerschaft getrennt lebend	25 (2,8%)	8 (1,8%)	17 (3,8%)
Ledig	118 (13,3%)	49 (11,0%)	69 (15,5%)
Geschieden/ eingetragene Lebenspartnerschaft aufgehoben	107 (12,0%)	56 (12,6%)	51 (11,4%)
verwitwet/ eingetragener Lebenspartner bzw. Lebenspartnerin verstorben	59 (6,6%)	36 (8,1%)	23 (5,2%)
Schulabschluss			
von der Schule abgegangen ohne Schulabschluss	8 (0,9%)	5 (1,1%)	3 (0,7%)
Hauptschulabschluss, Volksschulabschluss oder gleichwertiger Abschluss	150 (16,9%)	85 (19,1%)	65 (14,6%)
Polytechnische Oberschule (POS) der DDR mit Abschluss der 8. oder 9. Klasse	3 (0,3%)	2 (0,5%)	1 (0,2%)
Realschule (mittlere Reife) oder gleichwertiger Abschluss	278 (31,2%)	133 (30,0%)	145 (32,5%)
POS mit Abschluss der 10. Klasse	50 (5,6%)	23 (5,2%)	27 (6,1%)
Fachhochschulreife	88 (9,9%)	37 (8,3%)	51 (11,4%)
Abitur/Allgemeine oder fachgebundene Hochschulreife (Gymnasium bzw. Erweiterte Oberschule (EOS), auch EOS mit Lehre)	287 (32,2%)	144 (32,4%)	143 (32,1%)
Einen anderen Schulabschluss	26 (2,9%)	15 (3,4%)	11 (2,5%)

Soziodemografische Daten	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
Beruf Abschluss			
Noch in beruflicher Ausbildung (Berufsvorbereitungsjahr, Auszubildende, Praktikantin, Studentin)	1 (0,1%)	0 (0,0%)	1 (0,2%)
Schülerin einer berufsorientierten Aufbau-, Fachschule oder Ähnliches	1 (0,1%)	1 (0,2%)	0 (0,0%)
Keinen beruflichen Abschluss und nicht in beruflicher Ausbildung	59 (6,6%)	30 (6,8%)	29 (6,5%)
Beruflich-betriebliche Berufsausbildung (Lehre) abgeschlossen	302 (33,9%)	164 (36,9%)	138 (30,9%)
Berufsqualifizierender Abschluss einer beruflich-schulischen Ausbildung (Berufsfachschule, Kollegschule)	62 (7,0%)	26 (5,9%)	36 (8,1%)
Vorbereitungsdienst für den mittleren Dienst in der öffentlichen Verwaltung	5 (0,6%)	4 (0,9%)	1 (0,2%)
Abschluss einer einjährigen Ausbildung an einer Schule des Gesundheitswesens	6 (0,7%)	5 (1,1%)	1 (0,2%)
Abschluss einer zwei- bis dreijährigen Ausbildung an einer Schule des Gesundheitswesens	81 (9,1%)	37 (8,3%)	44 (9,9%)
Abschluss einer Ausbildung zur Erzieherin	29 (3,3%)	12 (2,7%)	17 (3,8%)
Abschluss einer Fachschule der DDR	19 (2,1%)	7 (1,6%)	12 (2,7%)
Abschluss einer Fach-, Meister-, Technikerschule, Verwaltungs- und Wirtschaftsakademie oder Fachakademie	51 (5,7%)	25 (5,6%)	26 (5,8%)
Bachelor	26 (2,9%)	12 (2,7%)	14 (3,1%)
Diplom, Master, Magister, Staatsexamen	176 (19,8%)	78 (17,6%)	98 (22,0%)
Promotion	18 (2,0%)	11 (2,5%)	7 (1,6%)
Einen anderen beruflichen Abschluss	54 (6,1%)	32 (7,2%)	22 (4,9%)
Erwerbssituation			
Vollzeiterwerbstätig	188 (21,1%)	94 (21,2%)	94 (21,1%)
Teilzeiterwerbstätig	220 (24,7%)	102 (23,0%)	118 (26,5%)
Altersteilzeit	11 (1,2%)	7 (1,6%)	4 (0,9%)
Geringfügig beschäftigt, 450-Euro-Job, Minijob, Ein-Euro-Job (bei Bezug von ALG II)	48 (5,4%)	29 (6,5%)	19 (4,3%)
Gelegentlich oder unregelmäßig beschäftigt	12 (1,3%)	5 (1,1%)	7 (1,6%)
in einer beruflichen Ausbildung/Lehre	2 (0,2%)	1 (0,2%)	1 (0,2%)
in Umschulung	1 (0,1%)	1 (0,2%)	0 (0,0%)
Mutterschafts-, Erziehungsurlaub, Elternzeit oder sonstige Beurlaubung	14 (1,6%)	6 (1,4%)	8 (1,8%)
nicht erwerbstätig ⁵	394 (44,3%)	199 (44,8%)	195 (43,7%)
Personen im Haushalt			

⁵ einschließlich: Schülerinnen, Studierende, die nicht gegen Geld arbeiten, Arbeitslose, Vorruheständlerinnen, Rentnerinnen ohne Nebenverdienst

Soziodemografische Daten	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
Eine Person (nur ich)	174 (19,6%)	84 (18,9%)	90 (20,2%)
Mehr als eine Person	716 (80,4%)	360 (81,1%)	356 (79,8%)
Anzahl Personen – Median (IQR)	2 (2-3)	2 (2-3)	2 (2-3)
Kinder <14 im Haushalt	148 (16,6%)	75 (16,9%)	73 (16,4%)
Anzahl Kinder <14 - Median (IQR)	1 (1-2)	1 (1-2)	1 (1-2)
Alter des jüngsten Kindes – Mean (SD)	7,8 (3,9)	8,1 (3,9)	7,4 (4,0)
Versorgung der Kinder	45 (30,4%)	23 (30,7%)	22 (30,1%)
Finanzierung Kinderversorgung	13/45 (28,9%)	9/23 (39,1%)	4/22 (18,2%)
Kosten Finanzierung Kinderversorgung im Monat (€) - Median (IQR)	300 (200-450)	280 (200-450)	375 (250-450)
Haushaltshilfe	121 (13,6%)	66 (14,9%)	55 (12,3%)
Pflegebedürftige	120 (13,5%)	64 (14,4%)	56 (12,6%)
Pflege der Angehörigen			
Ja, ich pflege selbst	54 (45,0%)	36 (56,2%)	18 (32,1%)
Ja, finanziell	10 (8,3%)	4 (6,2%)	6 (10,7%)
Nein	56 (46,7%)	24 (37,5%)	32 (57,1%)
Einkommen Haushalt			
leer=nicht relevant	246 (27,6%)	123 (27,7%)	123 (27,6%)
500 bis unter 1000	9 (1,0%)	5 (1,1%)	4 (0,9%)
1000 bis unter 1500	18 (2,0%)	10 (2,3%)	8 (1,8%)
1500 bis unter 2000	41 (4,6%)	25 (5,6%)	16 (3,6%)
2000 bis unter 2500	74 (8,3%)	32 (7,2%)	42 (9,4%)
2500 bis unter 3000	90 (10,1%)	46 (10,4%)	44 (9,9%)
3000 bis unter 4000	114 (12,8%)	59 (13,3%)	55 (12,3%)
4000 bis unter 5000	104 (11,7%)	54 (12,2%)	50 (11,2%)
5000 bis unter 6000	64 (7,2%)	27 (6,1%)	37 (8,3%)
6000 bis unter 7000	34 (3,8%)	17 (3,8%)	17 (3,8%)
7000 bis unter 10000	31 (3,5%)	13 (2,9%)	18 (4,0%)
10000 Euro und mehr	8 (0,9%)	6 (1,4%)	2 (0,4%)
keine Angabe	57 (6,4%)	27 (6,1%)	30 (6,7%)
Finanzielle Belastung			
gar nicht	317 (35,6%)	159 (35,8%)	158 (35,4%)
wenig	319 (35,8%)	159 (35,8%)	160 (35,9%)
mäßig	185 (20,8%)	94 (21,2%)	91 (20,4%)
sehr	69 (7,8%)	32 (7,2%)	37 (8,3%)
Mobilität			
Zu Fuß	31 (3,5%)	14 (3,2%)	17 (3,8%)
öffentliche Verkehrsmittel	124 (13,9%)	58 (13,1%)	66 (14,8%)
mit dem Auto	689 (77,4%)	356 (80,2%)	333 (74,7%)
mit dem Fahrrad	46 (5,2%)	16 (3,6%)	30 (6,7%)

Soziodemografische Daten	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
Auto			
leer=nicht relevant	201 (22,6%)	88 (19,8%)	113 (25,3%)
Ich fahre selbst	400 (44,9%)	200 (45,0%)	200 (44,8%)
Ich werde gefahren	289 (32,5%)	156 (35,1%)	133 (29,8%)

Anamnese-Baselinedaten

Tabelle 7 zeigt die Anamnese bei Studienbeginn in der Interventionsgruppe und der Kontrollgruppe. Die Mehrheit der Teilnehmerinnen beantwortete den Baseline-Fragebogen (97,9% insgesamt), wobei die Antwortrate in der Kontrollgruppe (98,0%) geringfügig höher lag als in der Interventionsgruppe (97,8%). Der durchschnittliche NCI-Komorbiditätsindex war in beiden Gruppen gleich (0,1). Bezüglich einzelner Erkrankungen wiesen Teilnehmerinnen der Interventionsgruppe eine leicht höhere Prävalenz für Herz-Kreislauf-Erkrankungen (34,9% vs. 32,3%), psychische Erkrankungen (17,9% vs. 13,3%) und gynäkologische Erkrankungen (21,3% vs. 17,6%) auf. Ebenso waren Erkrankungen der Blutgefäße des Gehirns in der Kontrollgruppe weniger häufig (1,6% vs. 2,9%). Ein Großteil der Teilnehmerinnen hatte keine Regelblutung mehr (94,6%), wobei dies in der Kontrollgruppe häufiger der Fall war (95,9% vs. 93,3%). Das durchschnittliche Alter zur Menopause lag insgesamt bei 45,1 Jahren, mit geringfügigen Unterschieden zwischen den Gruppen (44,9 Jahre in der Kontrollgruppe vs. 45,3 Jahre in der Interventionsgruppe). In der Interventionsgruppe berichteten etwas mehr Teilnehmerinnen über eine kürzlich durchgeführte Operation (26,0% vs. 23,7% in der Kontrollgruppe). Die meisten Teilnehmerinnen tranken selten Alkohol (44,0%), wobei es keinen wesentlichen Unterschied zwischen den Gruppen gab. Beim Rauchen gaben 50,8% der Interventionsgruppe an, Nichtraucher zu sein, verglichen mit 44,5% in der Kontrollgruppe. Aktuelles Rauchen war in der Interventionsgruppe seltener (8,5% vs. 12,9%). Insgesamt hatten 74,5% der Teilnehmerinnen eine familiäre Krebsvorgeschichte, mit einer etwas höheren Prävalenz in der Interventionsgruppe (76,3% vs. 72,7% in der Kontrollgruppe). Fast alle Teilnehmerinnen füllten den Fragebogen selbst aus (97,6%), und es gab keine relevanten Unterschiede zwischen den Gruppen. Die Belastung durch den Fragebogen wurde überwiegend als „gar nicht“ oder „wenig“ empfunden, wobei sich die Bewertungen zwischen den Gruppen kaum unterschieden. Die meisten Teilnehmerinnen fühlten sich sicher im Umgang mit einem Smartphone oder Tablet (91,9%). Hier gab es nahezu keine Differenzen zwischen den Gruppen (92,1% in der Kontrollgruppe vs. 91,7% in der Interventionsgruppe).

Tabelle 7: Anamnese

Anamnese-Baselinedaten	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
Antwort auf den Baseline-Fragebogen			
Beantwortet	890 (97,9%)	444 (98,0%)	446 (97,8%)
Fehlender Baseline-Fragebogen	19 (2,1%)	9 (2,0%)	10 (2,2%)
National Cancer Institute (NCI) Komorbiditätsindex - Mittelwert (SD)	0,1 (0,3)	0,1 (0,3)	0,1 (0,2)
Regelblutung			
Nein	842 (94,6%)	425 (95,9%)	417 (93,3%)
Ja	48 (5,4%)	18 (4,1%)	30 (6,7%)
Alter zur Menopause - Mittelwert (SD)	45,1 (7,8)	44,9 (7,4)	45,3 (8,1)
Komorbidität			

Anamnese-Baselinedaten	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
Herz-Kreislauf-Erkrankung	299 (33,6%)	143 (32,3%)	156 (34,9%)
Blutgefäßsystem	94 (10,6%)	48 (10,8%)	46 (10,3%)
Blutgefäße des Gehirns	20 (2,2%)	7 (1,6%)	13 (2,9%)
Atemwege oder Atmung	125 (14,0%)	60 (13,5%)	65 (14,5%)
Stoffwechsel oder Hormonhaushalt	289 (32,5%)	141 (31,8%)	148 (33,1%)
Psychisch	139 (15,6%)	59 (13,3%)	80 (17,9%)
Nervensystem	89 (10,0%)	42 (9,5%)	47 (10,5%)
Verdauungstrakt	45 (5,1%)	21 (4,7%)	24 (5,4%)
Gynäkologisch	173 (19,4%)	78 (17,6%)	95 (21,3%)
Rheumatologisch	38 (4,3%)	18 (4,1%)	20 (4,5%)
Urogenitaltrakt	49 (5,5%)	25 (5,6%)	24 (5,4%)
HIV-Infektion	1 (0,1%)	0 (0,0%)	1 (0,2%)
Andere onkologische Vorerkrankungen	240 (27,0%)	121 (27,3%)	119 (26,6%)
Operation innerhalb der letzten 6 Monate	221 (24,8%)	105 (23,7%)	116 (26,0%)
Alkoholkonsum			
nie	255 (28,7%)	134 (30,2%)	121 (27,1%)
selten	392 (44,0%)	192 (43,3%)	200 (44,7%)
gelegentlich	177 (19,9%)	83 (18,7%)	94 (21,0%)
wöchentlich	50 (5,6%)	27 (6,1%)	23 (5,1%)
täglich	16 (1,8%)	7 (1,6%)	9 (2,0%)
Rauchen			
nein	424 (47,6%)	197 (44,5%)	227 (50,8%)
ja, ehemals	371 (41,7%)	189 (42,7%)	182 (40,7%)
ja, aktuell	95 (10,7%)	57 (12,9%)	38 (8,5%)
Krebsvorerkrankungen Familie	663 (74,5%)	322 (72,7%)	341 (76,3%)
Ausfüllen des Fragebogens			
Ich habe den Fragebogen selbst ausgefüllt	869 (97,6%)	432 (97,5%)	437 (97,8%)
Ich habe den Fragebogen mit fremder Hilfe selbst ausgefüllt	19 (2,1%)	10 (2,3%)	9 (2,0%)
Jemand anderes hat den Fragebogen für mich ausgefüllt	2 (0,2%)	1 (0,2%)	1 (0,2%)
Sie ziehen es vor, den Fragebogen auf Papier auszufüllen			
Nein	869 (97,6%)	433 (97,7%)	436 (97,5%)
Ja	21 (2,4%)	10 (2,3%)	11 (2,5%)
Belastung Fragebogen			
gar nicht	647 (72,7%)	320 (72,2%)	327 (73,2%)
wenig	202 (22,7%)	99 (22,3%)	103 (23,0%)
mäßig	39 (4,4%)	24 (5,4%)	15 (3,4%)
sehr	2 (0,2%)	0 (0,0%)	2 (0,4%)

Anamnese-Baselinedaten	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
Sich sicher im Umgang mit einem Smartphone/Tablet fühlen			
Nein	72 (8,1%)	35 (7,9%)	37 (8,3%)
Ja	818 (91,9%)	408 (92,1%)	410 (91,7%)

Klinische Charakteristika

Das Durchschnittsalter der Patientinnen lag bei 50,8 Jahren (SD = 11,0), mit einer Spannweite von 19 bis 83 Jahren. Der durchschnittliche BMI betrug 33,9 kg/m², wobei Unterschiede zwischen den Gruppen festzustellen waren (Kontrollgruppe: 35,2 kg/m²; Interventionsgruppe: 32,6 kg/m²). Ein Großteil der Patientinnen (55,7%) konnte sich ohne Einschränkungen bewegen, war jedoch bei schwerer körperlicher Arbeit eingeschränkt. 19,0% waren gehfähig und konnten sich selbst versorgen, waren jedoch arbeitsunfähig. 5, % waren stark eingeschränkt, während 0,2% völlig pflegebedürftig waren. Beim primären Metastasenstatus (M-Status) wurden 39,8% der Patientinnen als M0 klassifiziert, 34,4% als M1, während 13,9% keine Angaben hierzu machten. Der primäre Tumorstatus (pT) zeigte, dass 36,2% der Patientinnen ein T2-Stadium aufwiesen, gefolgt von 19,6 % mit T1-Tumoren. Hinsichtlich des Nodalstatus (pN) waren 27,7% N0, während 31,6% als N1 klassifiziert wurden. Die häufigste Tumorart war das invasive Karzinom ohne speziellen Typ (73,4%), gefolgt vom invasiv lobulären Karzinom (13,6%). Die Mehrheit der Tumoren wurde als mäßig differenziert eingestuft (59,8%), während 33,3% schlecht differenziert waren. In Bezug auf die Therapie erhielten 59,0% der Patientinnen eine endokrine Therapie, 48,6% eine Chemotherapie und 23,0% eine Anti-HER2-Therapie. Immuntherapie wurde bei 3,5% der Patientinnen angewendet, während 34,9% eine zielgerichtete Therapie erhielten (Tabelle 8).

Tabelle 8: Klinische Charakteristika

Klinische Charakteristika	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
Alter bei Randomisierung (Jahre)			
Mittelwert (SD)	50,8 (11,0)	51,0 (11,2)	50,6 (10,7)
Min - Max	19 - 83	25 - 83	19 - 81
BMI (kg/m ²) – Mittelwert (SD)	33,9 (115,7)	35,2 (133,4)	32,6 (96,1)
Gesundheitszustand			
Normale, uneingeschränkte Aktivität	178 (20,0%)	82 (18,5%)	96 (21,5%)
Einschränkung bei körperlicher Anstrengung, gehfähig, leichte körperliche Arbeit möglich	496 (55,7%)	256 (57,8%)	240 (53,7%)
Gehfähig, Selbstversorgung möglich, aber nicht arbeitsfähig, kann mehr als 50% der Wachzeit aufstehen	169 (19,0%)	83 (18,7%)	86 (19,2%)
Nur begrenzte Selbstversorgung möglich, 50% der Wachzeit oder mehr an Bett oder Stuhl gebunden	45 (5,1%)	20 (4,5%)	25 (5,6%)
Völlig pflegebedürftig, keinerlei Selbstversorgung möglich, völlig an Bett oder Stuhl gebunden	2 (0,2%)	2 (0,5%)	0 (0,0%)
Primärer M Status			
Unbekannt	126 (13,9%)	62 (13,7%)	64 (14,0%)

Klinische Charakteristika	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
Nicht relevant	60 (6,6%)	30 (6,6%)	30 (6,6%)
M0	362 (39,8%)	191 (42,2%)	171 (37,5%)
M1	313 (34,4%)	148 (32,7%)	165 (36,2%)
MX	48 (5,3%)	22 (4,9%)	26 (5,7%)
Primärer pT Status			
Unbekannt	121 (13,3%)	64 (14,1%)	57 (12,5%)
Nicht relevant	61 (6,7%)	31 (6,8%)	30 (6,6%)
TX	25 (2,8%)	14 (3,1%)	11 (2,4%)
Tis	9 (1,0%)	7 (1,5%)	2 (0,4%)
T0	2 (0,2%)	1 (0,2%)	1 (0,2%)
T1	178 (19,6%)	85 (18,8%)	93 (20,4%)
T2	329 (36,2%)	161 (35,5%)	168 (36,8%)
T3	79 (8,7%)	40 (8,8%)	39 (8,6%)
T4	105 (11,6%)	50 (11,0%)	55 (12,1%)
Primärer pN Status			
Unbekannt	126 (13,9%)	67 (14,8%)	59 (12,9%)
Nicht relevant	63 (6,9%)	32 (7,1%)	31 (6,8%)
NX	63 (6,9%)	33 (7,3%)	30 (6,6%)
N0	252 (27,7%)	123 (27,2%)	129 (28,3%)
N1mi	10 (1,1%)	4 (0,9%)	6 (1,3%)
N1	287 (31,6%)	140 (30,9%)	147 (32,2%)
N2	59 (6,5%)	29 (6,4%)	30 (6,6%)
N3	49 (5,4%)	25 (5,5%)	24 (5,3%)
Art des Tumors			
Invasives Karzinom ohne ohne speziellen Typ	616 (73,4%)	314 (74,8%)	302 (72,1%)
Invasiv lobulär	114 (13,6%)	52 (12,4%)	62 (14,8%)
Invasiv tubulär	5 (0,6%)	3 (0,7%)	2 (0,5%)
Invasiv muzinös	3 (0,4%)	1 (0,2%)	2 (0,5%)
Invasiv apokrin	1 (0,1%)	0 (0,0%)	1 (0,2%)
Invasiv metaplastisch	1 (0,1%)	1 (0,2%)	0 (0,0%)
Invasiv papillär	3 (0,4%)	1 (0,2%)	2 (0,5%)
DCIS	34 (4,1%)	18 (4,3%)	16 (3,8%)
Sonstige	62 (7,4%)	30 (7,1%)	32 (7,6%)
Grading			
Gut differenziert	52 (6,4%)	24 (5,9%)	28 (6,9%)
Mäßig differenziert	486 (59,8%)	235 (58,0%)	251 (61,5%)
Schlecht differenziert	271 (33,3%)	145 (35,8%)	126 (30,9%)
Undifferenziert	3 (0,4%)	1 (0,2%)	2 (0,5%)
Borderline	1 (0,1%)	0 (0,0%)	1 (0,2%)

Klinische Charakteristika	Gesamt (n=909)	Kontrollgruppe (n=453)	Interventionsgruppe (n=456)
HER2-Status			
Negativ, keine Ausprägung	381 (42,8%)	198 (44,7%)	183 (40,9%)
Negativ, geringe Ausprägung	210 (23,6%)	96 (21,7%)	114 (25,5%)
Unklar, mäßige Ausprägung	105 (11,8%)	52 (11,7%)	53 (11,9%)
Positiv, hohe Ausprägung	194 (21,8%)	97 (21,9%)	97 (21,7%)
Östrogenrezeptor (ER - %) – Median (IQR)	80 (20-98)	80 (15-95)	80 (40-100)
Progesteronrezeptor (PR - %) – Median (IQR)	40 (1-80)	30 (1-80)	50 (2-80)
Ki-67 (%) – Median (IQR)	25 (15-40)	25 (15-40)	25 (15-40)
Therapie bei Randomisierung			
Endokrine Therapie	536 (59,0%)	262 (57,8%)	274 (60,1%)
Chemotherapie	442 (48,6%)	219 (48,3%)	223 (48,9%)
Anti-HER2-Therapie	209 (23,0%)	108 (23,8%)	101 (22,1%)
Immuntherapie	32 (3,5%)	16 (3,5%)	16 (3,5%)
Zielgerichtete Therapie	317 (34,9%)	150 (33,1%)	167 (36,6%)
Andere Therapien	123 (13,5%)	63 (13,9%)	60 (13,2%)
Klinischer Subtyp	n = 840	n = 416	n = 424
HR+/HER2+	141 (16,8%)	73 (17,6%)	68 (16,0%)
HR+/HER2-	556 (66,2%)	269 (64,7%)	287 (67,7%)
HR-/HER2+	56 (6,7%)	25 (6,0%)	31 (7,3%)
Triple-negativ	87 (10,4%)	49 (11,8%)	38 (9,0%)

3.2 Primärer Endpunkt: Fatigue-Score

3.2.1 Fatigue nach sechs Monaten

Die Abbildung 7 zeigt die Verteilung der Fatigue-T-Scores in der Kontroll- und Interventionsgruppe sowie die Differenz der Gruppenmittelwerte. Der Boxplot (links) illustriert, dass die Interventionsgruppe nach sechs Monaten im Median einen niedrigeren Fatigue-T-Score (53,5; IQR 46,4 – 59,7; Mittelwert 53,7; SD 10,7) aufwies als die Kontrollgruppe (59,3; IQR 53,7 – 67,7; Mittelwert 60,7; SD 10,2). Dies deutet auf eine geringere Fatigue-Belastung in der Interventionsgruppe hin. Der Forest-Plot (rechts) bestätigt diesen Unterschied quantitativ. Die mittlere Differenz der Fatigue-T-Scores zwischen der Interventions- und der Kontrollgruppe betrug -5,4 Punkte (95%-KI: -6,6 bis -4,1), was eine statistisch signifikante Verbesserung zugunsten der Interventionsgruppe darstellt. Da die Differenz unter dem MCID-Wert von -3,3 liegt, kann davon ausgegangen werden, dass diese Verbesserung auch klinisch relevant ist. Diese Ergebnisse legen nahe, dass die Intervention zu einer signifikanten Reduktion der Fatigue führte und einen bedeutsamen Einfluss auf das Wohlbefinden der Teilnehmerinnen hatte.

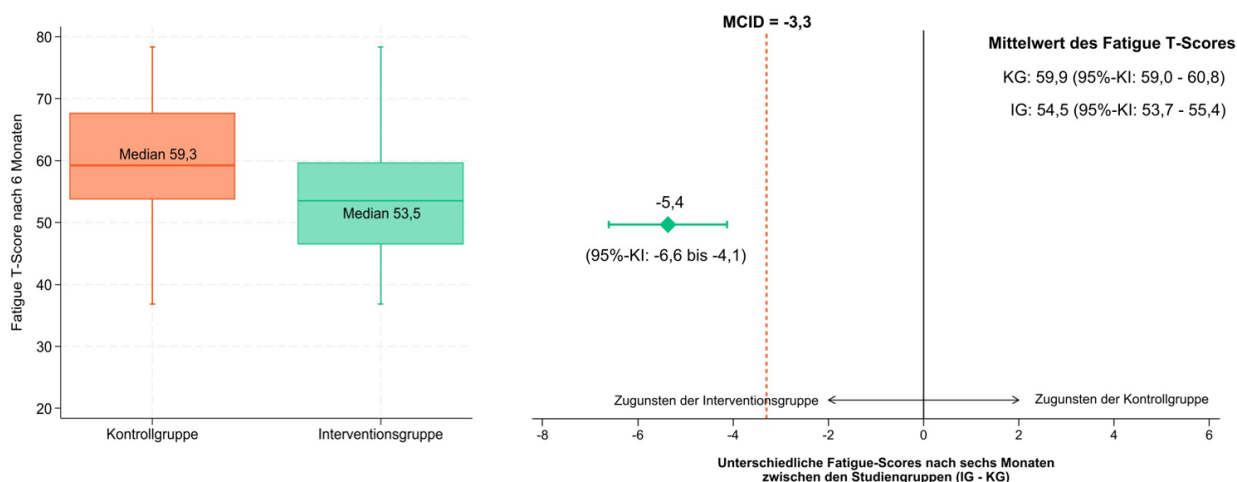


Abbildung 7: Box-Plot und Forest Plot für den primären Endpunkt Fatigue T-Score nach sechs Monaten zwischen Interventions- und Kontrollgruppe.

Fatigue-T-Scores nach sechs Monaten in der Interventionsgruppe (IG, n = 456) im Vergleich zur Kontrollgruppe (KG, n = 453). Es wurde ein lineares gemischtes Modell mit zufälligem Achsenabschnitt (random intercept) nach Studienzentrum durchgeführt, um den Mittelwert und den 95%-Koeffizienten der Fatigue-T-Scores zu schätzen. Die Modelle wurden für die Fatigue-T-Scores bei Studienbeginn, Alter, Bildung, ECOG, Beschäftigungsstatus, Familieneinkommen, Familienstand, Kinder ≤ 14 Jahre, Migration, Depression, NCI-Index, histologische Befunde, Art der Metastasierung, Behandlung bei Studienbeginn und Änderung der Alarmkriterien nach sechs Monaten adjustiert. Fehlende Daten wurden mittels multipler Imputation unter Verwendung einer verketteten Gleichung (MICE) berücksichtigt, wobei davon ausgegangen wurde, dass die Daten nach dem Zufallsprinzip fehlten, mit separater Imputation nach randomisiertem Arm (Randomisierter Arm-MI) (m = 30), und das Worst-Case-Szenario wurde nur auf verstorbene Patientinnen mit dem höchsten Fatigue-Score innerhalb jeder Studiengruppe angewandt. Die Quadrate stellen den geschätzten mittleren Unterschied in den Fatigue-T-Scores zwischen den beiden Gruppen dar, wobei die Fehlerbalken die 95%-KI illustrieren. Die gestrichelte Linie stellt den minimal klinisch wichtigen Unterschied (MCID) von 3,3 Punkten dar. Der MCID ist erreicht, wenn die KI die gestrichelte Linie nicht kreuzen.

3.2.2 Sensitivitätsanalyse

Tabelle 9 präsentiert die Ergebnisse der Sensitivitätsanalysen, die durchgeführt wurden, um die Robustheit der Studienergebnisse in Bezug auf den primären Endpunkt Fatigue zu überprüfen. Diese Analysen umfassen verschiedene Methoden für den Umgang mit fehlenden Daten und verschiedene Szenarien zur Bewertung der Auswirkungen fehlender Werte auf die Ergebnisse. Die Sensitivitätsanalysen zeigen durchweg, dass die Interventionsgruppe im Vergleich zur Kontrollgruppe deutlich niedrigere Fatigue-Werte aufwies. Die Ergebnisse bleiben auch bei verschiedenen Methoden zur Behandlung fehlender Daten und unter verschiedenen Annahmen in der Kippunktanalyse stabil. Dies unterstreicht die Robustheit der Studienergebnisse und die Wirksamkeit der Intervention.

In der Analyse vollständiger Fälle (Complete Case Analyse) betrug der geschätzte mittlere Fatigue-Scores in der Kontrollgruppe (n = 236) 59,2 (95%-KI: 58,1 – 60,3), während er in der Interventionsgruppe (n = 305) bei 54,5 (95%-KI: 53,6 – 55,5) lag. Die mittlere Differenz zwischen den Gruppen betrug -4,7 (95%-KI: -6,1 bis -3,2) mit einem signifikanten p-Wert von < 0,001. Dies deutet darauf hin, dass die Interventionsgruppe signifikant niedrigere Fatigue-Werte aufwies als die Kontrollgruppe.

Die Randomisierter Arm-MI-Analyse (MAR-basiert) und die referenzbasierte MI-Analyse, bei der angenommen wurde, dass das Verhalten der Patientinnen in der Interventionsgruppe nach Verlassen der Studie dem der Kontrollgruppe entspricht, zeigten ähnliche Ergebnisse. Der mittlere Unterschied zwischen den Gruppen betrug -5,2 (95% CI: -6,4 bis -4,0) in der Randomisierter Arm-MI-Analyse und der mittlere Unterschied zwischen den Gruppen betrug

-3,0 (95% CI: -4,2 bis -1,8) in der referenzbasierten MI-Analyse, was die Robustheit der Ergebnisse unter dieser Annahme weiter untermauert.

Eine Tipping-Point-Analyse (TPA) wurde durchgeführt, um die Robustheit der Ergebnisse gegenüber verschiedenen Annahmen über fehlende Daten zu testen. Delta-Werte von 4 bis 20 wurden verwendet, um die unterstellten Werte für verstorbene Patientinnen in der Interventionsgruppe anzupassen. Bei einem Delta-Wert von vier betrug die mittlere Differenz -4,5 (95% CI: -5,7 bis -3,2). Bei einem Delta-Wert von 20 betrug der mittlere Unterschied schließlich -1,3 (95% CI: -2,8 bis 0,10) mit einem p-Wert von 0,067. Diese Ergebnisse zeigen, dass selbst unter den konservativsten Annahmen der mittlere Unterschied zwischen den Gruppen signifikant bleibt, wenn auch mit einem höheren p-Wert bei den höheren Delta-Werten.

Tabelle 9: Sensitivitätsanalysen

Sensitivitätsanalysen	Geschätzter Mittelwert des T-Scores (95%-KI)		Mittlerer Unterschied zwischen den Gruppen (IG - KG)	p-Wert
	Kontrollgruppe (n = 453)	Interventionsgruppe (n = 456)		
Vollständiger Fall (Complete case) (n _{obs} = 541)	59,2 (58,1 – 60,3)	54,5 (53,6 – 55,5)	-4,7 (-6,1 to -3,2)	<0,001
Randomisierter Arm-MI (MAR-basiert) (n _{obs} = 909)	58,8 (57,8 – 59,8)	53,6 (52,7 – 54,4)	-5,2 (-6,4 to -4,0)	<0,001
Referenzbasierte MI [§] (n _{obs} = 909)	58,4 (57,5 – 59,3)	55,4 (54,5 – 56,3)	-3,0 (-4,2 to -1,8)	<0,001
Tipping-Point-Analyse (TPA)** (n _{obs} = 909)				
Delta-Anpassung: $\delta = 4$	58,8 (57,8 – 59,8)	54,3 (53,4 – 55,2)	-4,5 (-5,7 to -3,2)	<0,001
Delta-Anpassung: $\delta = 8$	58,8 (57,8 – 59,8)	55,1 (54,2 – 56,0)	-3,7 (-4,9 to -2,4)	<0,001
Delta-Anpassung: $\delta = 12$	58,8 (57,8 – 59,8)	55,8 (54,9 – 56,7)	-2,9 (-4,2 to -1,6)	<0,001
Delta-Anpassung: $\delta = 18$	58,8 (57,8 – 59,8)	57,0 (56,0 – 57,9)	-1,7 (-3,1 to -0,3)	0,015
Delta-Anpassung: $\delta = 20$	58,8 (57,8 – 59,8)	57,4 (56,4 – 58,3)	-1,3 (-2,8 to 0,10)	0,067

§Referenzbasierte MI: Mehrfache Imputation, bei der fehlende Werte unter der Annahme, dass das Verhalten der Interventionspatientinnen auf die Kontrollgruppe überspringt, imputiert werden.

*Das Worst-Case-Szenario wurde nur auf verstorbene Patientinnen mit dem höchsten Fatigue-Score innerhalb jeder Studiengruppe angewendet.

**TPA von 4 bis 20, ein Delta-Wert, der die imputierten Werte für Beobachtungen in der Interventionsgruppe anpasst, die vor Erreichen der sechs Monaten verstorben sind.

Ein lineares gemischtes Modell mit zufälligem Achsenabschnitt (random intercept) nach Studienzentrum wurde zur Schätzung des Mittelwerts und des 95 %-Koeffizienten des Fatigue-Scores durchgeführt. Das Modell wurde adjustiert für den Fatigue-Score bei Studienbeginn, das Alter, die Bildung, den ECOG-Wert, den Beschäftigungsstatus, das Familieneinkommen, den Familienstand, den Besitz von Kindern ≤ 14 Jahren, die Migration, die Depression, den NCI-Index, den histologischen Befund, die Art der Metastasierung, die Behandlung bei Studienbeginn und die Änderung der Alarmkriterien

3.2.3 Subgruppenanalysen

Tabelle 10 präsentiert die Ergebnisse der geplanten Untergruppenanalyse des Fatigue-T-Scores sechs Monate nach der Randomisierung. Die Analyse untersucht die Unterschiede in den Fatigue-Werten zwischen der Interventionsgruppe und der Kontrollgruppe in verschiedenen Subgruppen, basierend auf klinischen und soziodemografischen Merkmalen.

Die Ergebnisse wurden mittels eines linearen gemischten Modells mit zufälligem Achsenabschnitt (random intercept) nach Studienzentrum geschätzt und für verschiedene Baseline-Variablen adjustiert. Die Subgruppenanalyse zeigt, dass die Interventionsgruppe in allen Subgruppen signifikant niedrigere Fatigue-Werte im Vergleich zur Kontrollgruppe aufweist. Die mittleren Differenzen im Fatigue-T-Score variieren je nach Subgruppe, liegen jedoch in den meisten Fällen zwischen -4,7 und -7,5. Die p-Werte für die Interaktionstests deuten darauf hin, dass die Wirksamkeit der Intervention in den verschiedenen Subgruppen konsistent ist, da keine signifikanten Unterschiede zwischen den Subgruppen festgestellt wurden. Dies unterstreicht die Robustheit des Interventionseffekts, eine Reduktion von Fatigue, unabhängig von klinischen oder soziodemografischen Merkmalen der Patientinnen.

Tabelle 10: Geplante Subgruppenanalyse des Fatigue-T-Scores 6 Monate nach der Randomisierung

Subgruppenanalysen	Kontrollgruppe (KG) (n = 453)		Interventionsgruppe (IG) (n = 456)		Mittlerer Unterschied zwischen den Gruppen (IG - KG)	p-Wert für Interaktionstest
	Anzahl	Geschätzter marginaler Mittelwert (95%KI)	Anzahl	Geschätzter marginaler Mittelwert (95%KI)		
Arten der Fernmetastasierung						0,39
Gehirnmetastasen oder multiple Metastasen	250	59,7 (58,4 – 60,9)	253	54,5 (53,4 – 55,7)	-5,1 (-6,8; -3,5)	
Ossäre Metastasen oder Lymphknotenmetastasen oder kutane Metastasen	135	59,4 (57,7 – 61,1)	137	54,6 (53,1 – 56,2)	-4,8 (-7,1; -2,5)	
Viszerale Metastasen (nur ein Organ betreffend)	68	61,8 (59,2 – 64,4)	66	54,4 (52,2 – 56,6)	-7,5 (-10,8; -4,1)	
Histologischer Befund						0,17
HR+/HER- oder HR+/HER+	360	59,5 (58,5 – 60,5)	366	54,6 (53,6 – 55,5)	-4,9 (-6,3; -3,6)	
HR-/HER+ oder HR-/HER-	93	61,5 (59,2 – 63,7)	90	54,4 (52,3 – 56,4)	-7,1 (-9,9; -4,3)	
Familienstand*						0,19
Ledig	49	59,6 (56,9 – 62,2)	69	54,0 (51,6 – 56,4)	-5,6 (-9,2; -2,0)	
Verheiratet/ in eingetragener Lebenspartnerschaft zusammenlebend	295	59,6 (58,4 – 60,7)	286	54,9 (53,9 – 56,0)	-4,7 (-6,2; -3,1)	
Getrennt lebend/geschieden/verwitwet	100	61,2 (59,3 – 63,0)	91	53,7 (51,7 – 55,6)	-7,5 (-10,1; -4,9)	
Einkommen Haushalt (Monatlich) (in Euro)*						0,78
<2500	129	60,9 (59,2 – 62,5)	136	54,9 (53,5 – 56,3)	-5,9 (-8,0; -3,8)	
2500-4999	179	59,3 (57,9 – 60,6)	164	54,4 (53,0 – 55,8)	-4,9 (-6,8; -3,0)	
≥5000	65	59,6 (57,1 – 62,1)	76	54,1 (52,2 – 56,1)	-5,4 (-8,4; -2,3)	
Depression						0,86
Nein	404	60,0 (59,1 – 61,0)	392	54,6 (53,7 – 55,5)	-5,4 (-6,7; -4,1)	
Ja	39	58,8 (55,6 – 62,0)	55	53,8 (51,4 – 56,2)	-5,0 (-9,0; -1,0)	
Alarmkriterien nach sechs Monaten						0,38
Initiale Alarmkriterien	189	59,6 (58,2 – 61,1)	199	54,9 (53,6 – 56,2)	-4,7 (-6,6; -2,8)	
Aktualisierte Alarmkriterien	264	60,1 (58,9 – 61,3)	257	54,3 (53,1 – 55,4)	-5,8 (-7,5; -4,2)	
Alter bei Randomisierung (Jahre)						0,39
<50	197	60,4 (58,8 – 61,9)	211	54,4 (53,1 – 55,6)	-6,0 (-7,9; -4,1)	
≥50	256	59,6 (58,3 – 60,8)	245	54,7 (53,5 – 55,9)	-4,9 (-6,5; -3,2)	
BMI (kg/m²)						0,47
<30	238	60,3 (59,1 – 61,5)	259	54,5 (53,4 – 55,7)	-5,8 (-7,4; -4,2)	

Subgruppenanalysen	Kontrollgruppe (KG) (n = 453)		Interventionsgruppe (IG) (n = 456)		Mittlerer Unterschied zwischen den Gruppen (IG - KG)	p-Wert für Interaktionstest
	Anzahl	Geschätzter marginaler Mittelwert (95%KI)	Anzahl	Geschätzter marginaler Mittelwert (95%KI)		
≥30	215	59,5 (58,1 – 60,8)	197	54,5 (53,3 – 55,8)	-4,9 (-6,7; -3,1)	
Endokrine Therapie						0,65
Nein	191	61,7 (60,2 – 63,2)	182	55,9 (54,5 – 57,4)	-5,7 (-7,6; -3,8)	
Ja	262	58,7 (57,4 – 60,0)	274	53,5 (52,4 – 54,7)	-5,1 (-6,8; -3,5)	
Chemotherapie						0,71
Nein	234	59,8 (58,5 – 61,2)	233	54,7 (53,5 – 55,9)	-5,2 (-7,0; -3,3)	
Ja	219	60,0 (58,7 – 61,3)	223	54,4 (53,1 – 55,6)	-5,6 (-7,4; -3,9)	
Anti-Her2-Therapie						0,67
Nein	345	60,6 (59,5 – 61,7)	355	55,1 (54,1 – 56,0)	-5,5 (-6,9; -4,1)	
Ja	108	57,6 (55,7 – 59,5)	101	52,7 (50,8 – 54,6)	-4,9 (-7,5; -2,3)	
Andere Immuntherapie						0,21
Nein	437	60,0 (59,0 – 60,9)	440	54,7 (53,9 – 55,6)	-5,2 (-6,5; -4,0)	
Ja	16	58,7 (53,4 – 64,0)	16	49,0 (43,9 – 54,1)	-9,7 (-16,5; -2,8)	
Gezielte Therapie						0,77
Nein	303	60,2 (59,0 – 61,3)	289	54,9 (53,9 – 56,0)	-5,2 (-6,8; -3,7)	
Ja	150	59,4 (57,7 – 61,1)	167	53,8 (52,4 – 55,2)	-5,6 (-7,8; -3,5)	

*Die Zahlen wurden vor Anwendung der multiplen Imputation dargestellt.

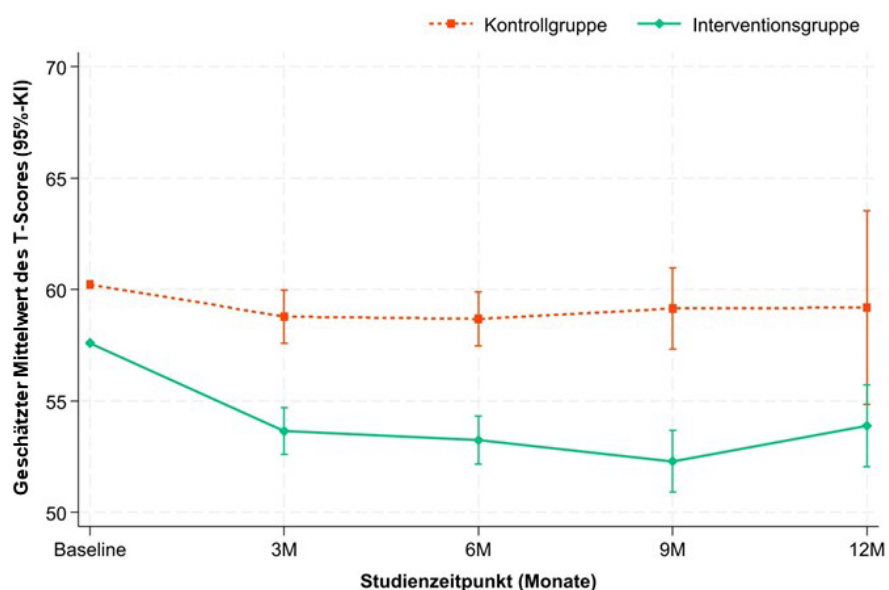
Es wurde ein lineares gemischtes Modell mit zufälligem Achsenabschnitt (random intercept) nach Studienzentrum durchgeführt, um den Mittelwert und den 95 %-Koeffizienten des T-Scores für Fatigue zu schätzen, und es wurden die Interaktionsterme zwischen der Studiengruppe und jedem Faktor untersucht. Die Modelle wurden für den Fatigue-T-Score bei Studienbeginn, Alter, Bildung, ECOG, Beschäftigungsstatus, Familieneinkommen, Familienstand, Kinder ≤ 14 Jahre, Migration, Depression, NCI-Index, histologische Befunde, Art der Metastasierung, Behandlung bei Studienbeginn und die Änderung der Alarmkriterien nach 6 Monaten angepasst.

3.3 Sekundäre Endpunkte

Die vorliegenden Ergebnisse umfassen die längsschnittliche Analyse von Fatigue, körperlicher Funktion, Gesundheitsbezogene Lebensqualität (HRQoL) sowie des Überlebens in der PRO B Studie über einen Zeitraum von 12 Monaten. Die Analysen wurden mittels linearer gemischter Modelle durchgeführt, die für verschiedene Baseline-Variablen adjustiert wurden.

3.3.1 Fatigue-Score über 12 Monate

Die Interventionsgruppe wies zu Studienbeginn einen mittleren Fatigue-T-Score von 57,6 (SD: 9,5) auf, während die Kontrollgruppe einen mittleren Score von 60,2 (SD: 9,0) hatte. Nach drei Monaten betrug der geschätzte Mittelwert des Fatigue-T-Scores in der Interventionsgruppe 53,6 (95%-KI: 52,6–54,7), während er in der Kontrollgruppe bei 58,8 (95%-KI: 57,6–60,0) lag. Der mittlere Unterschied zwischen den Gruppen betrug zu diesem Zeitpunkt -5,2 (95%-KI: -6,6 bis -3,8) mit einem p-Wert von <0,001. Dieser signifikante Unterschied setzte sich über die folgenden Monate fort: Nach sechs Monaten betrug der mittlere Unterschied -5,4 (95%-KI: -6,9 bis -3,9, $p < 0,001$), nach neun Monaten -6,8 (95%-KI: -9,0 bis -4,7, $p < 0,001$) und nach 12 Monaten -5,2 (95%-KI: -9,0 bis -1,4, $p = 0,004$). Diese Ergebnisse zeigen, dass die Interventionsgruppe im Vergleich zur Kontrollgruppe signifikant niedrigere Fatigue-Werte über den gesamten Zeitraum von 12 Monaten aufwies, was auf eine erfolgreiche Reduktion der Fatigue durch die Intervention hindeutet.



Studiengruppe	Baseline	Geschätzter Mittelwert des T-Scores (95%-KI)			
		3 Monate	6 Monate	9 Monate	12 Monate
Intervention	57,6 (SD 9,5)	53,6 (52,6 - 54,7)	53,2 (52,2 - 54,3)	52,3 (50,9 - 53,7)	53,9 (52,0 - 55,7)
Kontrolle	60,2 (SD 9,0)	58,8 (57,6 - 60,0)	58,7 (57,5 - 60,9)	59,2 (57,3 - 61,0)	59,2 (54,8 - 63,5)
Mittlerer Unterschied zwischen den Gruppen (IG - KG)	-	-5,2 (-6,6; -3,8)	-5,4 (-6,9; -3,9)	-6,8 (-9,0; -4,7)	-5,2 (-9,0; -1,4)
p-Wert	-	<0,001	<0,001	<0,001	0,004

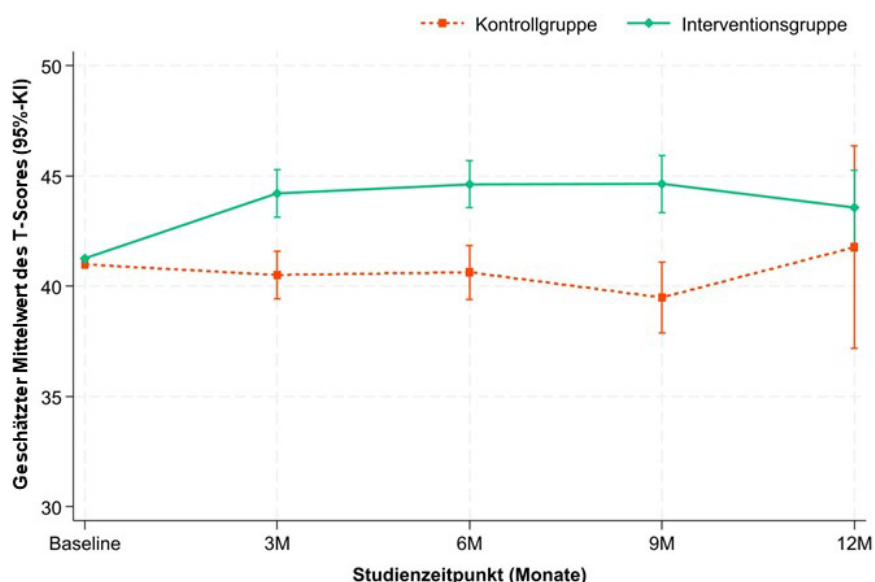
Abbildung 8: Längsschnittliche Daten des Fatigue-Scores über 12 Monate in der PRO B Studie

Randomisierter Arm-MI (m = 30), es wurde keine multiple Imputation auf Patientinnen angewandt, die nach diesem Zeitpunkt starben oder schwer erkrankten. Zur Schätzung des Mittelwerts und des 95%-KI wurde ein lineares gemischtes Modell mit zufälligem Achsenabschnitt (random intercept) nach Studienzentrum durchgeführt. Das Modell wurde adjustiert für den Fatigue-Score bei Studienbeginn, das Alter, die Bildung, den ECOG-Wert, den Beschäftigungsstatus, das Familieneinkommen, den Familienstand, das Vorhandensein von

Kindern ≤ 14 Jahren, Migration, die Depression, den NCI-Index, den histologischen Befund, die Art der Metastasierung, die Behandlung bei Studienbeginn und die Änderung der Alarmkriterien. IG: Interventionsgruppe, KG: Kontrollgruppe, KI: Konfidenzintervall. Die Grafik zeigt die Mittelwerte mit 95%-Konfidenzintervallen, bereinigt um Mehrfachvergleiche nach der Sidak-Methode.

3.3.2 Körperliche Funktionsfähigkeit über 12 Monate

Die körperliche Funktionsfähigkeit wurde ebenfalls über 12 Monate hinweg analysiert. Zu Studienbeginn betrug der mittlere T-Score für körperliche Funktion in der Interventionsgruppe 41,2 (SD: 9,7) und in der Kontrollgruppe 41,0 (SD: 9,0). Nach drei Monaten betrug der geschätzte Mittelwert in der Interventionsgruppe 44,2 (95%-KI: 43,1–45,3), während er in der Kontrollgruppe bei 40,5 (95%-KI: 39,4–41,6) lag. Der mittlere Unterschied zwischen den Gruppen betrug zu diesem Zeitpunkt 3,7 (95%-KI: 2,4–5,0) mit einem p-Wert von <0,001. Dieser signifikante Unterschied setzte sich über die folgenden Monate fort: Nach sechs Monaten betrug der mittlere Unterschied 4,0 (95%-KI: 2,6–5,4, p < 0,001), nach neun Monaten 5,1 (95%-KI: 3,2–7,1, p < 0,001). Nach 12 Monaten betrug der mittlere Unterschied 1,8 (95%-KI: -2,5–6,1, p = 0,730), was auf einen nicht mehr signifikanten Unterschied hinweist. Diese Ergebnisse zeigen, dass die Interventionsgruppe im Vergleich zur Kontrollgruppe eine signifikant bessere körperliche Funktionsfähigkeit über die ersten neun Monate aufwies, wobei der Unterschied nach 12 Monaten nicht mehr signifikant war.



Studiengruppe	Baseline	Geschätzter Mittelwert des T-Scores (95%-KI)			
		3 Monate	6 Monate	9 Monate	12 Monate
Intervention	41,2 (SD 9,7)	44,2 (43,1 - 45,3)	44,6 (43,6 - 45,7)	44,6 (43,3 - 45,9)	43,6 (41,9 - 45,3)
Kontrolle	41,0 (SD 9,0)	40,5 (39,4 - 41,6)	40,6 (39,4 - 41,9)	39,5 (37,9 - 41,1)	41,8 (37,2 - 46,3)
Mittlerer Unterschied zwischen den Gruppen (IG - KG)	-	3,7 (2,4; 5,0)	4,0 (2,6; 5,4)	5,1 (3,2; 7,1)	1,8 (-2,5; 6,1)
p-Wert	-	<0,001	<0,001	<0,001	0,730

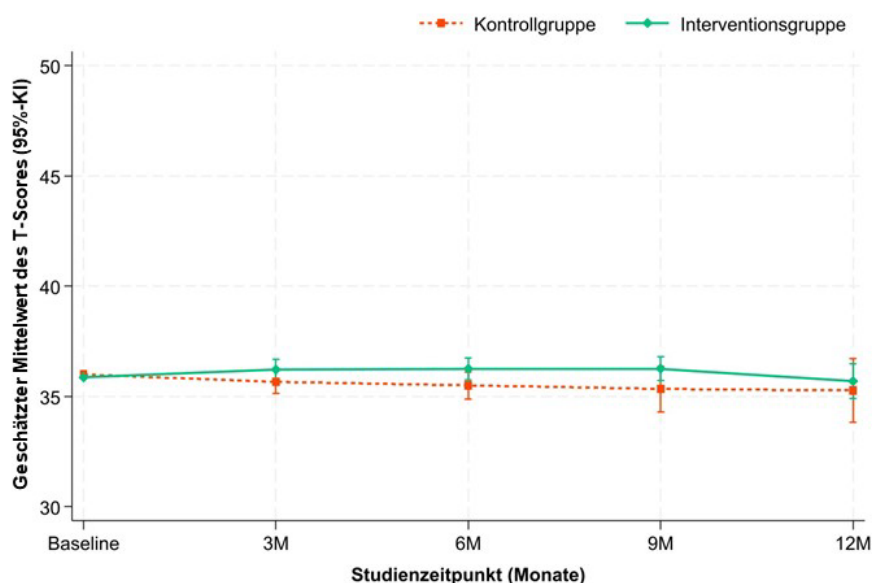
Abbildung 9: Longitudinale Daten der körperlichen Funktion über 12 Monate in der PRO B Studie.

Randomisierter Arm-MI (m = 30), es wurde keine multiple Imputation auf Patientinnen angewandt, die nach diesem Zeitpunkt starben oder schwer erkrankten. Zur Schätzung des Mittelwerts und des 95%-KI wurde ein lineares gemischtes Modell mit zufälligem Achsenabschnitt (random intercept) nach Studienzentrum durchgeführt. Das Modell wurde adjustiert für den Score für die körperliche Funktionsfähigkeit bei Studienbeginn, Alter, Bildung, ECOG, Beschäftigungsstatus, Familieneinkommen, Familienstand, Kinder ≤ 14 Jahre, Migration, Depression, NCI-Index, histologische Befunde, Art der Metastasierung, Behandlung bei Studienbeginn und Änderung der Alarmkriterien. IG: Interventionsgruppe, KG: Kontrollgruppe, KI:

Konfidenzintervall. Die Grafik zeigt die Mittelwerte mit 95%-Konfidenzintervallen, bereinigt um Mehrfachvergleiche nach der Sidak-Methode.

3.3.3 Gesundheitsbezogene Lebensqualität (HRQoL) über 12 Monate

Die Gesundheitsbezogene Lebensqualität (HRQoL) wurde ebenfalls über 12 Monate hinweg untersucht. Zu Studienbeginn betrug der mittlere T-Score für HRQoL in der Interventionsgruppe 35,9 (SD: 3,5) und in der Kontrollgruppe 36,0 (SD: 4,2). Nach drei Monaten betrug der geschätzte Mittelwert in der Interventionsgruppe 36,2 (95%-KI: 35,8–36,7), während er in der Kontrollgruppe bei 35,7 (95%-KI: 35,1–36,2) lag. Der mittlere Unterschied zwischen den Gruppen betrug zu diesem Zeitpunkt 0,6 (95%-KI: -0,03–1,1) mit einem p-Wert von 0,066. Nach 6 Monaten betrug der mittlere Unterschied 0,8 (95%-KI: 0,02–1,5, $p = 0,039$), was auf einen signifikanten Unterschied hinweist. Nach neun Monaten betrug der mittlere Unterschied 0,9 (95%-KI: -0,1–1,9, $p = 0,085$), und nach 12 Monaten 0,4 (95%-KI: -1,0–1,8, $p = 0,898$). Diese Ergebnisse zeigen, dass die Interventionsgruppe im Vergleich zur Kontrollgruppe eine leicht höhere HRQoL aufwies, wobei der Unterschied nur nach sechs Monaten signifikant ist. Nach 12 Monaten ist der Unterschied nicht mehr signifikant.



Studiengruppe	Baseline	Geschätzter Mittelwert des T-Scores (95%-KI)			
		3 Monate	6 Monate	9 Monate	12 Monate
Intervention	35,9 (SD 3,5)	36,2 (35,8 - 36,7)	36,3 (35,8 - 36,8)	36,3 (35,7 - 36,8)	35,7 (34,9 - 36,5)
Kontrolle	36,0 (SD 4,2)	35,7 (35,1 - 36,2)	35,5 (34,9 - 36,1)	35,3 (34,3 - 36,4)	35,3 (33,8 - 36,7)
Mittlerer Unterschied zwischen den Gruppen (IG - KG)	-	0,6 (-0,03; 1,1)	0,8 (0,02; 1,5)	0,9 (-0,1; 1,9)	0,4 (-1,0; 1,8)
p-Wert	-	0,066	0,039	0,085	0,898

Abbildung 10: Longitudinale Daten der Gesundheitsbezogenen Lebensqualität über 12 Monate in der PRO B Studie.

Randomisierter Arm-MI ($m = 30$), es wurde keine multiple Imputation auf Patientinnen angewandt, die nach diesem Zeitpunkt starben oder schwer erkrankten. Zur Schätzung des Mittelwerts und des 95%-KI wurde ein lineares gemischtes Modell mit zufälligem Achsenabschnitt (random intercept) nach Studienzentrum durchgeführt. Das Modell wurde adjustiert für den Score der Gesundheitsbezogenen Lebensqualität bei Studienbeginn, Alter, Bildung, ECOG, Beschäftigungsstatus, Familieneinkommen, Familienstand, Kinder ≤ 14 Jahre, Migration, Depression, NCI-Index, histologische Befunde, Art der Metastasierung, Behandlung bei Studienbeginn und Änderung der Alarmkriterien. IG: Interventionsgruppe, KG: Kontrollgruppe, KI: Konfidenzintervall. Die Grafik zeigt die Mittelwerte mit 95%-Konfidenzintervallen, bereinigt um Mehrfachvergleiche nach der Sidak-Methode.

3.3.4 Überleben

Tabelle 11 präsentiert die Analysen zum Überleben in der Gesamtkohorte sowie Subgruppenanalysen zum Überleben für Patientinnen mit viszeralem Metastasen und triple-negativem Brustkrebs. In der Gesamtkohorte betrug das 6-Monats-Überleben in der Interventionsgruppe 94,7%, während es in der Kontrollgruppe 92,0% betrug. Das 12-Monats-Überleben lag in der Interventionsgruppe bei 87,6% und in der Kontrollgruppe bei 84,7%. Die multivariate Analyse ergab eine adjustierte Hazard Ratio (HR) von 0,71 (95%-KI: 0,51–0,99) mit einem p-Wert von 0,04, was auf ein signifikant besseres Überleben in der Interventionsgruppe hinweist (

Abbildung 11).

In der Subgruppe der Patientinnen mit viszeraler Metastasierung betrug das 6-Monats-Überleben in der Interventionsgruppe 96,9%, während es in der Kontrollgruppe 91,9% betrug. Das 12-Monats-Überleben lag in der Interventionsgruppe bei 93,3% und in der Kontrollgruppe bei 87,7%. Die multivariate Analyse ergab eine adjustierte HR von 0,43 (95%-KI: 0,15–1,29) mit einem p-Wert von 0,33, was auf keinen signifikanten Unterschied in der Überlebensrate zwischen den Gruppen hinweist (Abbildung 12).

In der Subgruppe der Patientinnen mit triple-negativem Mammakarzinom betrug das 6-Monats-Überleben in der Interventionsgruppe 91,7%, während es in der Kontrollgruppe 84,4% betrug. Das 12-Monats-Überleben lag in der Interventionsgruppe bei 78,5% und in der Kontrollgruppe bei 67,2%. Die multivariate Analyse ergab eine adjustierte HR von 0,74 (95%-KI: 0,31–1,78) mit einem p-Wert von 0,51, was ebenfalls auf keinen signifikanten Unterschied in der Überlebensrate zwischen den Gruppen hinweist (Abbildung 13).

In der Subgruppe der Patientinnen mit HR+/HER- betrug das 6-Monats-Überleben in der Interventionsgruppe 95,3%, während es in der Kontrollgruppe 90,4% betrug. Das 12-Monats-Überleben lag in der Interventionsgruppe bei 88,5% und in der Kontrollgruppe bei 82,7%. Die multivariate Analyse ergab eine adjustierte HR von 0,70 (95%-KI: 0,48–1,01) mit einem p-Wert von 0,06 (Tabelle 11Abbildung 13).

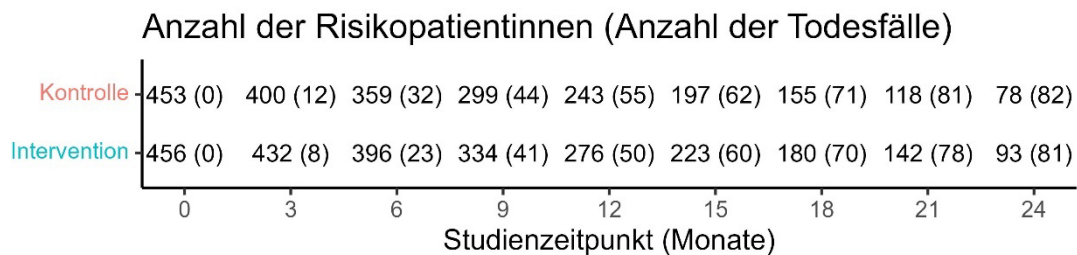
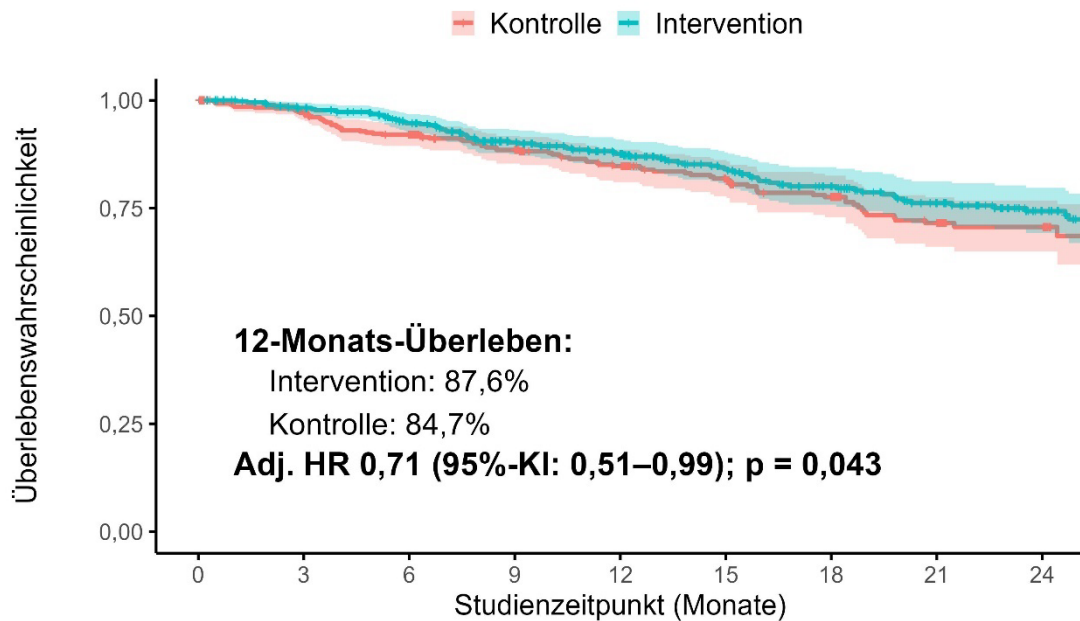


Abbildung 11: Überleben (Gesamtkohorte)

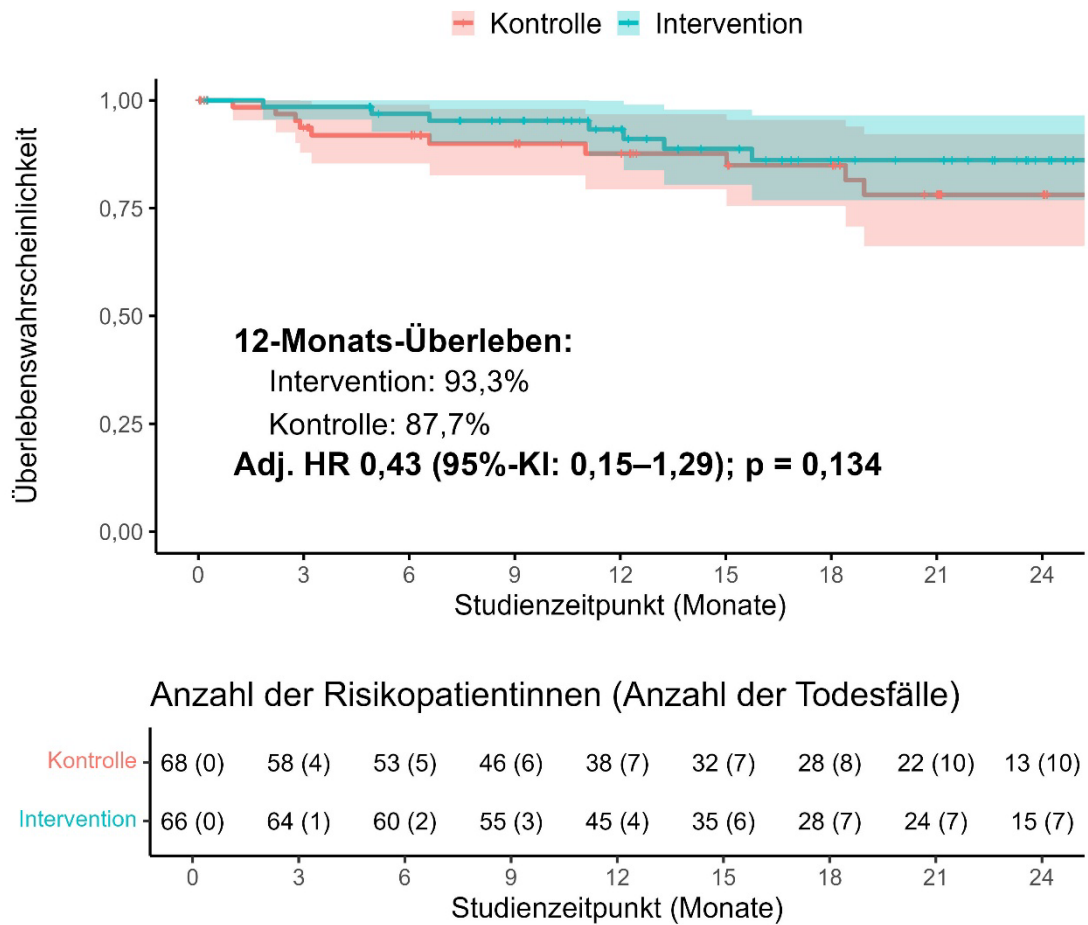


Abbildung 12: Überleben bei viszeraler Metastasierung

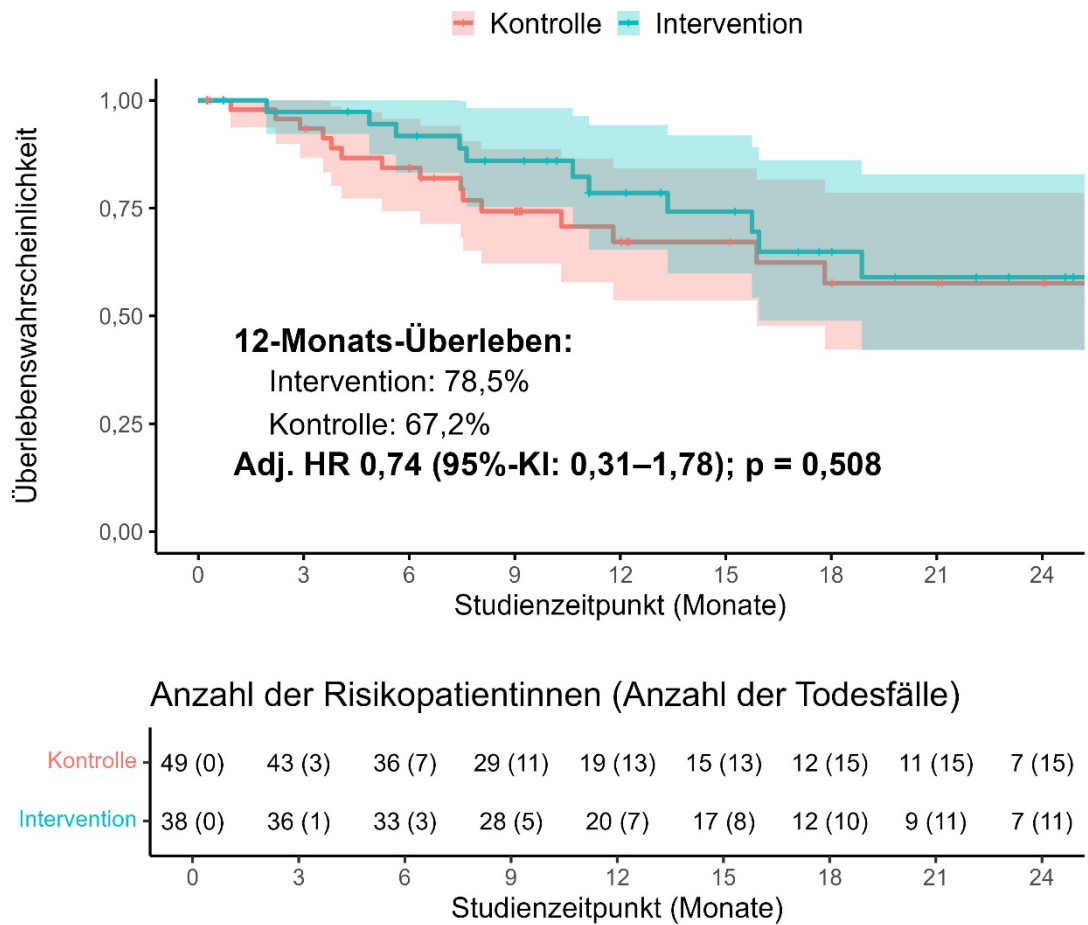


Abbildung 13: Überleben bei triple-negativem Brustkrebs

Tabelle 11: Überlebensanalysen

	Overall Survival		Bivariate Analyse			Multivariate Analyse		
	6 Monate	12 Monate	Unadj. HR	95%KI	p-Wert	Adj.HR	95%-KI	p-Wert
Gesamte Kohorte			n = 909			n = 838		
Interventionsgruppe	94,7%	87,6%	0,82	(0,61; 1,12)	0,21	0,71 ^a	(0,51; 0,99)	0,04
Kontrollgruppe	92,0%	84,7%						
Subgruppenanalysen								
Viszeral (nur ein Organ)			n = 134			n = 133		
Interventionsgruppe	96,9%	93,3%	0,57	(0,21; 1,56)	0,27	0,51 ^b	(0,15; 1,29)	0,13
Kontrollgruppe	91,9%	87,7%						
Triple-negativ			n = 87			n = 87		
Interventionsgruppe	91,7%	78,5%	0,75	(0,35; 1,64)	0,48	0,61 ^c	(0,31; 1,78)	0,51
Kontrollgruppe	84,4%	67,2%						
HR+/HER2-			n = 556			n = 555		
Interventionsgruppe	95,3%	88,5%	0,68	(0,47; 0,99)	0,04	0,70 ^c	(0,48; 1,01)	0,06
Kontrollgruppe	90,4%	82,7%						

Adj. = adjustiert, ECOG = HR = Hazard Ratio, NA = nicht anwendbar, da kein Ereignis in der Kontrollgruppe, Unadj. = unadjustiert

^aDas Modell wurde für Alter, Art der Metastasen, ECOG bei Studienbeginn und Subtyp adjustiert.

^bDas Modell wurde für Alter, ECOG bei Studienbeginn und Subtyp adjustiert. In der viszeralen Untergruppe wurde der Subtyp aufgrund der geringen Stichprobengröße ausgeschlossen.

^cDas Modell wurde für Alter, Art der Metastasen und ECOG bei Studienbeginn adjustiert.

Die Hazard Ratios (HR) wurden anhand eines Cox Proportional Hazards Modells mit Studienzentrum als zufällig geteiltem Frailty Term geschätzt.

Es wurden die Wechselwirkungen zwischen Studiengruppe und Ort der Metastasen sowie Studiengruppe und Subtyp untersucht (p-Wert: Gruppe*Metastasen = 0,70 und Gruppe*Subtyp = 0,14).

3.3.5 Zeit bis zum ersten Wechsel der systemischen Therapie

Im Folgenden werden die Ergebnisse der Gruppenvergleiche (Interventions- vs. Kontrollgruppe) in Bezug auf die Zeit bis zum ersten Wechsel der systemischen Therapie in der PRO B Studie präsentiert. Die präsentierten Daten umfassen die Häufigkeit von Therapiewechseln, die Gründe für diese Wechsel sowie die Anzahl der durchgeführten Therapieänderungen (Tabelle 12).

In der Kontrollgruppe (n = 453) wurde bei 194 Patientinnen (42,8%) ein erster Wechsel der systemischen Therapie durchgeführt, während in der Interventionsgruppe (n = 456) 203 Patientinnen (44,5%) einen solchen Wechsel erfuhren. In der Kontrollgruppe wurde bei drei Patientinnen (0,7%) die Dosis reduziert, während in der Interventionsgruppe ebenfalls drei Patientinnen (0,7%) eine Dosisreduktion erfuhren. In der Kontrollgruppe wurde bei 14 Patientinnen (3,1%) die Substanz gewechselt, während in der Interventionsgruppe ebenfalls 14 Patientinnen (3,1%) einen Substanzwechsel erfuhren. In der Kontrollgruppe wurde die Therapie bei 129 Patientinnen (28,5%) abgebrochen, während in der Interventionsgruppe 128 Patientinnen (28,1%) die Therapie abbrachen – diese Abbrüche erfolgten aufgrund von Progression, Nebenwirkungen oder auf Wunsch der Patientinnen. In der Kontrollgruppe wurden insgesamt 344 Therapiewechsel durchgeführt, während in der Interventionsgruppe 312 Therapiewechsel stattfanden. Der Median der Therapiewechsel lag in beiden Gruppen bei 1 (Spannweite: 1–6).

Abbildung 14 zeigt den Vergleich der Zeit bis zur ersten systemischen Therapieänderung zwischen der Kontroll- und der Interventionsgruppe. Die Grafik verdeutlicht, dass es keine signifikanten Unterschiede in der Zeit bis zum ersten Therapiewechsel zwischen den beiden Gruppen gab. Die multivariate Analyse ergab eine adjustierte HR von 0,89 (95%-KI: 0,74–1,08) mit einem p-Wert von 0,241, was ebenfalls auf keinen signifikanten Unterschied in der Zeit bis zum ersten Therapiewechsel zwischen den Gruppen hinweist.

Tabelle 12: Erster Wechsel der systemischen Therapie

Wechsel der systemischen Therapie	Kontrollgruppe (n = 453)	Interventionsgruppe (n = 456)
Erster Wechsel der systemischen Therapie	194 (42,8%)	203 (44,5%)
Gründe für den ersten Wechsels der systemischen Therapie		
Reduzierung der Dosis	3 (0,7%)	3 (0,7%)
Wechsel der Substanz	14 (3,1%)	14 (3,1%)
Abbruch wegen Progression der Krankheit, unerwünschter Wirkungen, Wunsch der Patientin	129 (28,5%)	128 (28,1%)
Anzahl der Änderungen der systemischen Therapie	344	312
Mittelwert (SD)	1,77 (1,08)	1,54 (0,90)
Median (Minimum – Maximum)	1 (1 – 6)	1 (1 – 6)

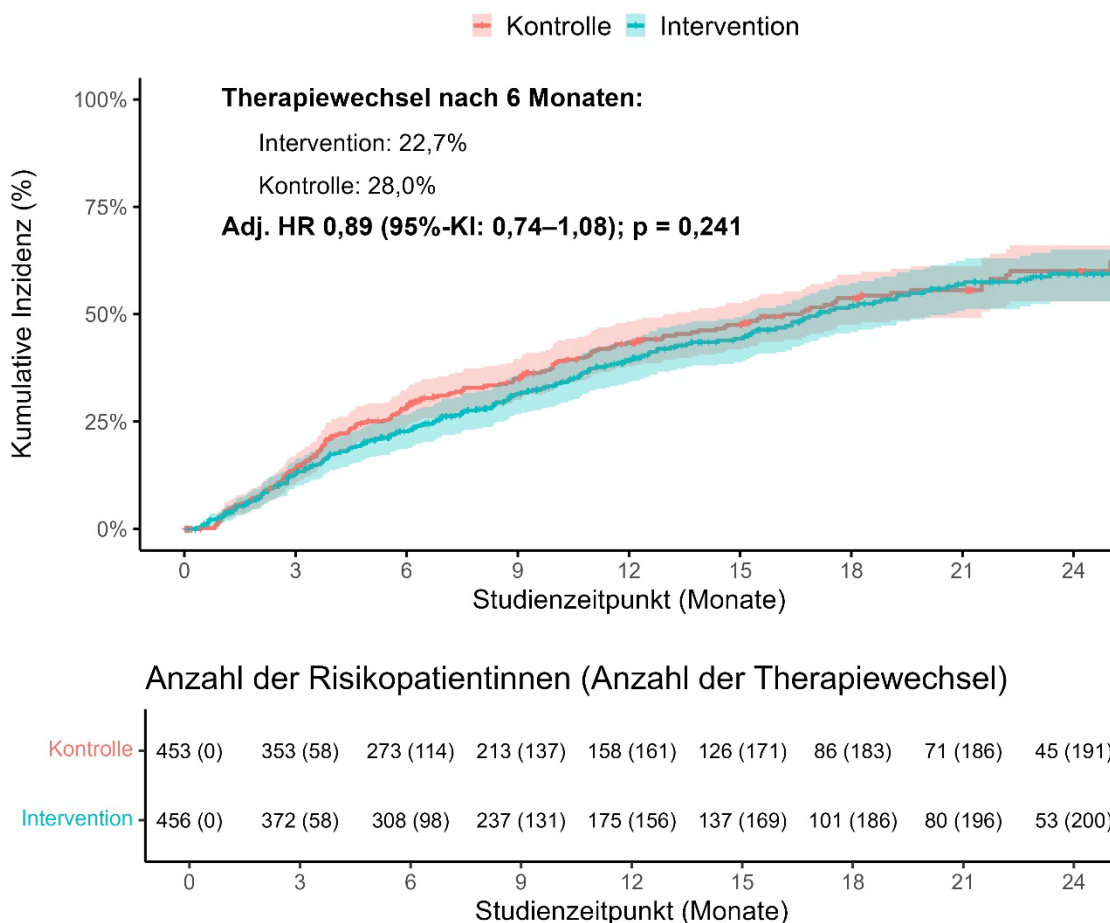


Abbildung 14: Zeit bis zum ersten Wechsel der systemischen Therapie im Vergleich zwischen Kontroll- und Interventionsgruppe

3.4 Interventionsgruppe

3.4.1 Wöchentliche Antworten der PRO-Befragungen

Abbildung 15 zeigt die wöchentlichen Antworten der PRO-Befragungen der Teilnehmerinnen der Interventionsgruppe über einen Zeitraum von 52 Wochen. Die Abbildung visualisiert die Entwicklung der Antworten im Laufe der Zeit, um zu veranschaulichen, wie sich die Beteiligung der Patientinnen an den PRO-Befragungen im Laufe der Wochen entwickelte. In der ersten Woche lag die Antwortrate bei 85,8%, was bedeutet, dass von 456 versendeten Befragungen 391 beantwortet wurden. Nach sechs Monaten (Woche 24) sank die Antwortrate auf 73,4%. In dieser Woche wurden 329 von 448 Befragungen beantwortet, während auf 119 Befragungen (26,6%) keine Rückmeldung folgte. Nach 12 Monaten (Woche 52) war die Antwortrate weiter auf 61,4% gesunken. Von 347 versendeten Befragungen wurden 213 beantwortet, während 134 Umfragen (38,6%) unbeantwortet blieben.

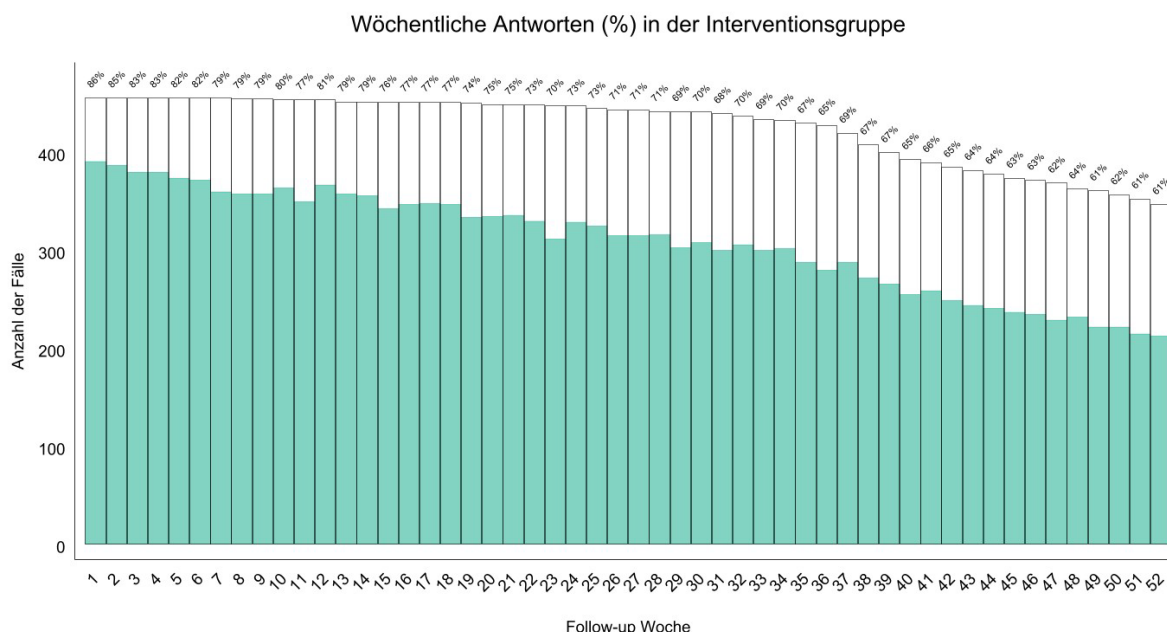


Abbildung 15: Wöchentliche Antworten (in Prozent) in der Interventionsgruppe

In der Interventionsgruppe wurden insgesamt 36.845 PRO-Erhebungen durchgeführt. Davon wurden 9.394 nach den initialen Alarmkriterien und 27.451 nach den aktualisierten Alarmkriterien ausgewertet. Die Rücklaufquote der PRO-Befragungen betrug 66,7%, d.h. 24.571 der versendeten Befragungen wurden beantwortet. Die Erstbefragung wurde von 391 Teilnehmerinnen (1,6%) beantwortet, während die Antwortrate der Follow-Up-Befragungen bei 98,4% lag (Tabelle 13).

Tabelle 13: Übersicht über die Anzahl der verschickten PRO-Befragungen und die Rücklaufquote in der Interventionsgruppe

	Gesamt	Änderung der Alarmkriterien	
		Initiale Alarmkriterien ¹	Aktualisierte Alarmkriterien ²
Anzahl der versendeten PRO-Erhebungen	36.845	9.394	27.451
Erstbefragung	456 (1,2%)	299 (3,2%)	157 (0,6%)
Follow-Up-Befragungen	36.389 (98,8%)	9.095 (96,8%)	27.294 (99,4%)
Rückläufer PRO-Befragungen			
keine Antwort	12.274 (33,3%)	2.001 (21,3%)	10.273 (37,4%)
Antwort	24.571 (66,7%)	7.393 (78,7%)	17.178 (62,6%)
Antwort Erstbefragung	391 (1,6%)	263 (3,6%)	128 (0,7%)
Antwort Follow-Up-Befragungen	24.180 (98,4%)	7.130 (96,4%)	17.050 (99,3%)

¹ Die initiale Alarmkriterien basierten auf einer Kombination des Health-Transition-Items (Gesundheitszustand im Vergleich zur letzten Woche) und einer Verschlechterung ausgewählter EORTC-Items.

² Die angepassten Alarmkriterien ermöglichten das Auslösen eines Alarms allein auf Grundlage des Health-Transition-Items, wenn der Patient angab, sich im Vergleich zur Vorwoche etwas oder deutlich schlechter zu fühlen – auch ohne Veränderungen in den EORTC-Items. Wenn keine Verschlechterung im Health-Transition-Item vorlag, wurden Alarme ausschließlich aufgrund einer Verschlechterung der EORTC-Items ausgelöst.

3.4.2 Alarmauslösende Domänen

Tabelle 14 zeigt die Domänen, die im Rahmen der PRO-Erhebungen Alarme ausgelöst haben, sowie deren Verteilung in Abhängigkeit davon, ob die Patientinnen einen Wechsel der systemischen Therapie erfahren haben oder nicht. Die Alarmdaten wurden in zwei Zeiträumen betrachtet: Alarme aufgrund von Veränderungen im Vergleich zur Vorwoche und Alarme aufgrund von Veränderungen im Vergleich zum Vormonat.

Alarme aufgrund einer Veränderung der PRO-Werte im Vergleich zur Vorwoche wurden in 10,5% der Fälle in der Domäne *emotionale Funktion* ausgelöst. Die soziale Funktion war für 9,3% der Alarme der Auslöser, die Rollenfunktion für 9,1% der Alarme. Die Körperliche Funktion war Auslöser für 6,0% der Alarme. Gastrointestinale Symptome (Übelkeit, Erbrechen, Appetitlosigkeit, Verstopfung, Durchfall) lösten 4,4% der Alarme aus. Lebensqualität bzw. Gesundheitsstatus war Auslöser für 0,8% der Alarme.

Für Alarme aufgrund einer Veränderung der PRO-Werte im Vergleich zum Vormonat war die emotionale Funktion in 10,7%, die soziale Funktion in 9,6%, die Rollenfunktion in 9,3%, die kognitive Funktion in 8,0% und die körperliche Funktion in 6,5% der Fälle der Auslöser. Bei Betrachtung der Symptomskalen lösten gastrointestinale Symptome (4,9%) am häufigsten einen Alarm aus, gefolgt von Dyspnoe (3,8%), Schmerzen (3,3%) und Schlaflosigkeit (2,9%). Fatigue war bei 2,5% der Alarme der Auslöser. Die Lebensqualität oder der Gesundheitszustand lösten 1,4% der Alarme aus.

Zusammenfassend zeigen die Ergebnisse, dass bestimmte Domänen wie emotionale Funktion, soziale Funktion, körperliche Funktion, gastrointestinale Symptome, Schmerz und Fatigue häufiger Alarme auslösten als andere Domänen, insbesondere bei Patientinnen, die einen Wechsel der systemischen Therapie erfahren hatten.

Tabelle 14: Domänen, die Alarme ausgelöst haben

Domänen	Gesamt (n = 456)
	nAlarme = 24,180
Die PRO-Werte haben sich im Vergleich zu den PRO-Werten der Vorwoche verändert	
Emotionale Funktion	2538 (10,5%)
Soziale Funktion	2248 (9,3%)
Rollenfunktion	2200 (9,1%)
Kognitive Funktion	1913 (7,9%)
Körperliche Funktion	1461 (6,0%)
Gastrointestinale Symptome*	1068 (4,4%)
Dyspnoe	911 (3,8%)
Schlaflosigkeit	785 (3,2%)
Schmerz	705 (2,9%)
Fatigue	573 (2,4%)
Lebensqualität/ Gesundheitsstatus	193 (0,8%)
Finanzielle Schwierigkeiten	15 (0,1%)
Die PRO-Werte haben sich im Vergleich zu den PRO-Werten des Vormonats verändert	
Emotionale Funktion	2594 (10,7%)
Soziale Funktion	2311 (9,6%)
Rollenfunktion	2257 (9,3%)
Kognitive Funktion	1926 (8,0%)
Körperliche Funktion	1572 (6,5%)
Gastrointestinale Symptome*	1176 (4,9%)

Dyspnoe	908 (3,8%)
Schmerz	789 (3,3%)
Schlaflosigkeit	701 (2,9%)
Fatigue	615 (2,5%)
Lebensqualität / Gesundheitsstatus	346 (1,4%)
Finanzielle Schwierigkeiten	18 (0,1%)

*Gastrointestinale Symptome = Übelkeit/Erbrechen, Appetitlosigkeit, Verstopfung, Durchfall

3.4.3 Häufigkeit von Alarmen und Alarmbenachrichtigungen an das Behandlungsteam (Alarm-E-Mail)

Insgesamt wurden 24.180 Alarme ausgelöst, davon 7.130 nach den alten Kriterien und 17.050 nach den neuen Kriterien. In 87,3% der Follow-Up-Befragungen wurde kein Alarm ausgelöst, während in 12,5% der Befragungen ein Alarm erfolgte. In 91,1% der Fälle, in denen ein Alarm ausgelöst wurde, wurde auch eine Alarm-E-Mail an das betreuende Behandlungsteam versendet (Tabelle 15).

Tabelle 15: Anzahl der Alarme und der nach den Alarmen versandten E-Mails an das Behandlungsteam

	Gesamt	Änderung des Alarmkriteriums	
		Initiale Alarmkriterien	Aktualisierte Alarmkriterien
Anzahl der Antworten Follow-Up-Befragungen	n =24.180	n =7.130	n =17.050
Kein Alarm	21.101 (87,3%)	6.465 (90,7%)	14.636 (85,8%)
Alarm	3.019 (12,5%)	665 (9,3%)	2.354 (13,8%)
Fehlende PRO-Werte	60 (0,2%)	0 (0,0%)	60 (0,4%)
Alarm-E-Mail versandt	n =3.019	n =665	n =2.354
Keine E-Mail versandt	269 (8,9%)	47 (7,1%)	222 (9,4%)
E-Mail versandt	2.750 (91,1%)	618 (92,9%)	2.132 (90,6%)

3.4.4 Kontakt nach Alert

Nach einem Alarm wurde in 84,8% der Fälle Kontakt mit den Patientinnen aufgenommen. In 86,7% dieser Kontakte erfolgte die Kontaktaufnahme innerhalb von 48 Stunden. Die durchschnittliche Zeit bis zur Kontaktaufnahme betrug 17 Stunden (Spannweite: 3–35 Stunden) (Tabelle 16).

Tabelle 16: Kontakt nach Alarm und Zeit bis zum Kontakt

	Gesamt	Änderung des Alarmkriteriums	
		Initiale Alarmkriterien	Aktualisierte Alarmkriterien
Kontakt nach Alarm	n =3019	n =665	n =2,354
Kein Kontakt	460 (15,2%)	56 (8,4%)	404 (17,2%)
Kontakt	2559 (84,8%)	609 (91,6%)	1950 (82,8%)
Kontakt innerhalb von 48 Stunden	n =2184	n =235	n =1,949
Verzögert kontaktiert	290 (13,3%)	38 (16,2%)	252 (12,9%)
Kontakt innerhalb von 48 Stunden	1894 (86,7%)	197 (83,8%)	1697 (87,1%)
Zeit von der Alarm-E-Mail bis zum Kontakt (Stunden) – median (IQR)	17 (3–35)	18 (2–40)	17 (3–32)

3.4.5 Alarmauslöser

Die häufigsten Alarmauslöser waren Lebensqualität/Gesundheitsstatus (20,0%), körperliche Funktion (15,8%) und Schmerz (13,5%). Die telefonische Beratung war die häufigste Interventionsform (44,2%), gefolgt von keiner Intervention (48,5%). In 66,6% der Fälle wurde der Alarm von einer Study Nurse bestätigt, während in 28,9% der Fälle ärztliches Personal involviert war (Tabelle 17).

Tabelle 17: Alarmauslöser

	Gesamt	Änderung des Alarmkriteriums	
		Initiale Alarmkriterien	Aktualisierte Alarmkriterien
Position der kontaktierten Person	n=2577	n =616	n =1961
Arzt/Ärztin	746 (28,9%)	216 (35,1%)	530 (27,0%)
Study Nurse	1715 (66,6%)	384 (62,3%)	1331 (67,9%)
Andere	116 (4,5%)	16 (2,6%)	100 (5,1%)
Grund für die Verzögerung der Alarmquittierung / Kontaktaufnahme	n =2577	n =616	n =1961
Patientin einbestellt	78 (3,0%)	22 (3,6%)	56 (2,9%)
telefonische Beratung	1140 (44,2%)	277 (45,0%)	863 (44,0%)
keine Intervention / kein Handlungs-, oder Beratungsbedarf	1249 (48,5%)	282 (45,8%)	967 (49,3%)
nicht relevant, da Patientin schon stationär ist	110 (4,3%)	35 (5,7%)	75 (3,8%)
Symptome oder funktionellen Beschwerden, die im Gespräch mit der Patientin am häufigsten genannt wurden	n =1218	n =299	n =919
Lebensqualität / Gesundheitsstatus	244 (20,0%)	74 (24,7%)	170 (18,5%)
Körperliche Funktion	193 (15,8%)	23 (7,7%)	170 (18,5%)
Rollenfunktion	17 (1,4%)	2 (0,7%)	15 (1,6%)
Emotionale Funktion	165 (13,5%)	50 (16,7%)	115 (12,5%)
Kognitive Funktion	9 (0,7%)	1 (0,3%)	8 (0,9%)
Soziale Funktion	12 (1,0%)	2 (0,7%)	10 (1,1%)
Fatigue	76 (6,2%)	25 (8,4%)	51 (5,5%)
Übelkeit/Erbrechen	99 (8,1%)	21 (7,0%)	78 (8,5%)
Schmerz	164 (13,5%)	35 (11,7%)	129 (14,0%)
Dyspnoe	27 (2,2%)	8 (2,7%)	19 (2,1%)
Schlaflosigkeit	14 (1,1%)	6 (2,0%)	8 (0,9%)
Appetitverlust	17 (1,4%)	8 (2,7%)	9 (1,0%)
Verstopfung	16 (1,3%)	3 (1,0%)	13 (1,4%)
Durchfall	47 (3,9%)	16 (5,4%)	31 (3,4%)
Andere	118 (9,7%)	25 (8,4%)	93 (10,1%)

3.5 Analyse der Sekundärdaten (Krankenkassendaten)

3.5.1 Beschreibung der Studienpopulation der Krankenkassendaten

Das CONSORT-Diagramm (Abbildung 16) zeigt die Stichprobe der PRO B Studienpatientinnen, die der gesundheitsökonomischen Evaluation zugrunde liegt. Das Diagramm beschreibt die Stichprobengröße von Interventions- und Kontrollgruppe über einen Zeitraum von 12 Monaten.

Von den 463 Patientinnen in der Interventionsgruppe wurden 334 (72,1%) von der gesundheitsökonomischen Analyse ausgeschlossen, da sie nicht bei den teilnehmenden

Krankenkassen versichert waren. Zusätzlich gab es zwei Fälle (0,14%) mit vollständig fehlenden Befragungen und neun Fälle (1,9%) ohne Datenübermittlung. In der Kontrollgruppe (n = 461) wurden ebenfalls 334 Patientinnen (72,6%) ausgeschlossen, da sie nicht bei den teilnehmenden Krankenkassen versichert waren. In zwei Fällen (0,14%) fehlten die Befragungen vollständig, in acht Fällen (1,7%) lag keine Datenübermittlung vor.

Folglich wurden für die gesundheitsökonomische Evaluation 118 Patientinnen (25,5%; 118/463) aus der Interventionsgruppe und 117 Patientinnen (25,4%; 117/461) aus der Kontrollgruppe eingeschlossen. Von den 118 Patientinnen in der Interventionsgruppe fehlten drei EQ-5D-Datensätze (2,6%), während 115 (97,5%) in die Analyse einfließen. In der Kontrollgruppe fehlten bei vier Patientinnen (3,4%) die EQ-5D-Daten, 113 (96,6%) EQ-5D-Datensätze wurden verwendet. Nach 12 Monaten nahmen 82 Patientinnen (69,6%) in der Interventionsgruppe und 65 Patientinnen (56,6%) in der Kontrollgruppe weiterhin an der Studie teil. Tabelle 18 zeigt die grundlegenden Merkmale der Stichprobe für die gesundheitsökonomische Evaluation, unterteilt in die Interventionsgruppe (n = 118) und die Kontrollgruppe (n = 117). Die Daten umfassen demografische, sozioökonomische und klinische Merkmale der Patientinnen mit metastasiertem Brustkrebs.

Die Ergebnisse zeigen, dass die Interventions- und Kontrollgruppe in Bezug auf demografische, sozioökonomische und klinische Merkmale weitgehend vergleichbar waren. Die hohe Übereinstimmung zwischen den Gruppen hinsichtlich Alter, Familienstand, Bildungsniveau, Erwerbstätigkeit, Einkommen und klinischer Merkmale deutet darauf hin, dass die Randomisierung erfolgreich war und die Gruppen gut vergleichbar waren.

Das Durchschnittsalter der Patientinnen bei Randomisierung betrug in beiden Gruppen 53 Jahre (SD 11). Hinsichtlich des Familienstands waren die meisten Patientinnen verheiratet oder lebten in einer eingetragenen Lebenspartnerschaft (64,9% insgesamt, 64,3% in der Interventionsgruppe und 65,5% in der Kontrollgruppe). Es lag überwiegend ein mittleres Bildungsniveau vor (67,5% insgesamt, 66,1% in der Interventionsgruppe und 69,0% in der Kontrollgruppe), während etwa 25% ein hohes und 8% ein niedriges Bildungsniveau aufwiesen. Etwa 47% der Patientinnen waren erwerbstätig, während 53% nicht erwerbstätig waren. Das monatliche Haushaltseinkommen lag bei 44% der Teilnehmerinnen zwischen 2.500 und 4.999 Euro.

Die klinischen Brustkrebssubtypen waren in beiden Gruppen ähnlich verteilt: HR+/HER2- war der häufigste Subtyp (67,9% insgesamt, 69,4% in der Interventionsgruppe und 66,4% in der Kontrollgruppe), gefolgt von HR+/HER2+ (13,8%), HR-/HER2+ (9,6%), und triple-negativem Brustkrebs (TNBC) (8,7%). Hinsichtlich der Art der Fernmetastasierung zeigte sich folgendes Bild: 58% der Patientinnen wiesen Gehirnmetastasen oder multiple Metastasen auf, bei 24% lagen ossäre Metastasen, Lymphknotenmetastasen oder kutane Metastasen vor und bei 17% viszerale Metastasen, die nur ein Organ betrafen. Die Follow-Up-Zeit betrug im Median 19 Monate (IQR: 0,4–34 Monate). Die Interventionsgruppe wies eine etwas längere mediane Follow-Up-Zeit von 20 Monaten auf, verglichen mit 18 Monaten in der Kontrollgruppe.

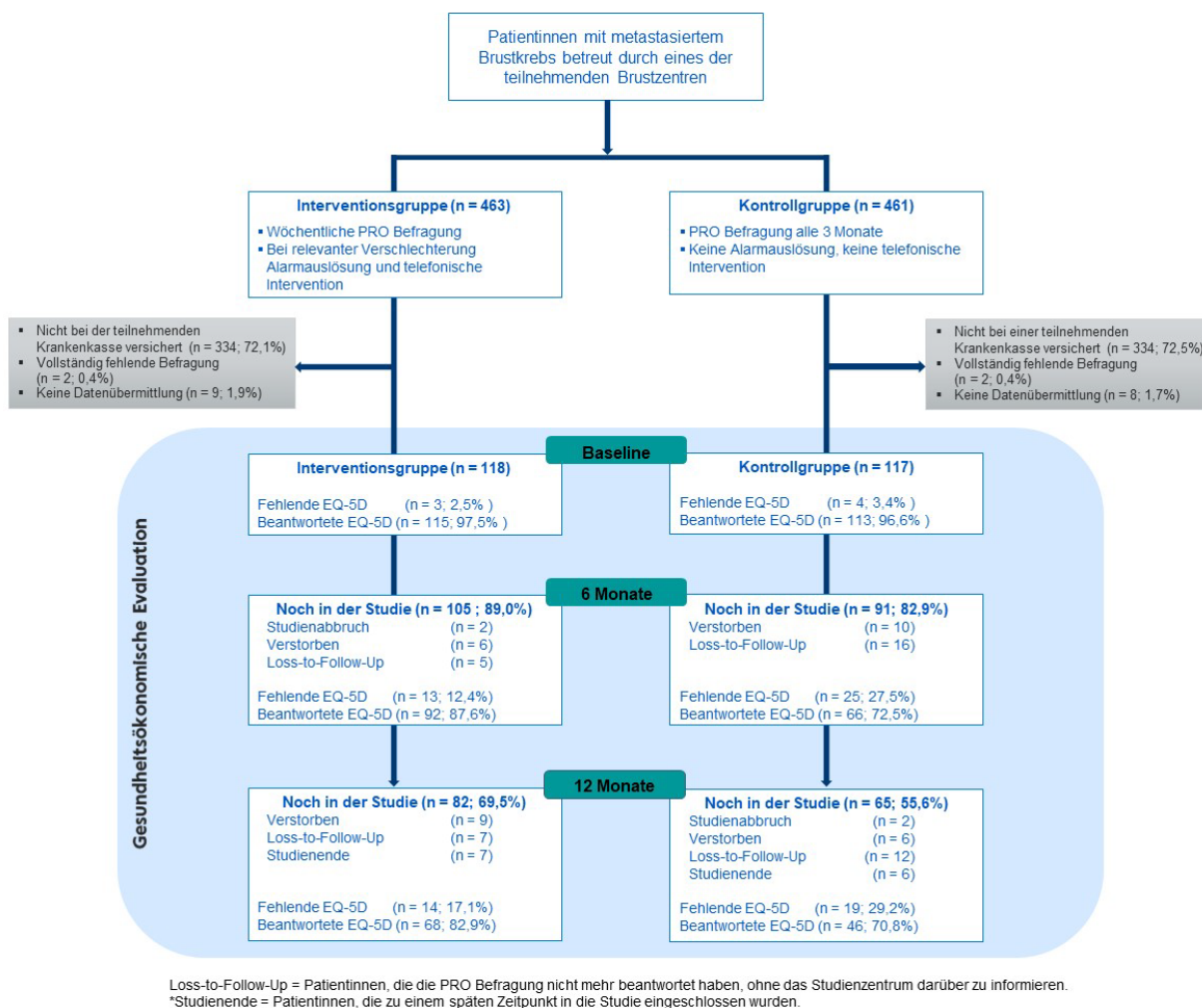


Abbildung 16: CONSORT-Diagramm für die gesundheitsökonomische Evaluation

Tabelle 18: Grundlegende Merkmale der Stichprobe für die gesundheitsökonomische Evaluation

Grundlegende Merkmale der Stichprobe	Gesamt (n = 235)	Studiengruppe	
		Kontrollgruppe (n = 117)	Interventionsgruppe (n = 118)
Alter bei Randomisierung (Jahre) – Mittelwert (SD)	53 (11)	53 (10)	53 (11)
Familienstand			
Ledig	27 (11,8%)	12 (10,6%)	15 (13,0%)
Verheiratet/ in eingetragener Lebenspartnerschaft zusammenlebend	148 (64,9%)	74 (65,5%)	74 (64,3%)
Getrennt lebend/geschieden/verwitwet	53 (23,2%)	27 (23,9%)	26 (22,6%)
Bildungsstatus			
Niedrig	18 (7,9%)	9 (8,0%)	9 (7,8%)
Mittel	154 (67,5%)	78 (69,0%)	76 (66,1%)
Hoch	56 (24,6%)	26 (23,0%)	30 (26,1%)

Grundlegende Merkmale der Stichprobe	Gesamt (n = 235)	Studiengruppe	
		Kontrollgruppe (n = 117)	Interventionsgruppe (n = 118)
Erwerbstätig			
Nein	120 (52,6%)	59 (52,2%)	61 (53,0%)
Ja	108 (47,4%)	54 (47,8%)	54 (47,0%)
Haushaltseinkommen (Monatlich) (in Euro)			
<2500	84 (43,1%)	43 (43,9%)	41 (42,3%)
2500–4999	86 (44,1%)	44 (44,9%)	42 (43,3%)
≥5000	25 (12,8%)	11 (11,2%)	14 (14,4%)
Kinder ≤ 14 im Haushalt			
Nein	204 (89,5%)	103 (91,2%)	101 (87,8%)
Ja	24 (10,5%)	10 (8,8%)	14 (12,2%)
Histologischer Befund			
HR+/HER- oder HR+/HER+	190 (80,9%)	95 (81,2%)	95 (80,5%)
HR-/HER+ oder HR-/HER-	45 (19,1%)	22 (18,8%)	23 (19,5%)
Art der Fernmetastasierung			
Gehirnmetastasen oder multiple Metastasen	137 (58,3%)	65 (55,6%)	72 (61,0%)
Ossäre Metastasen oder Lymphknotenmetastasen oder kutane Metastasen	57 (24,3%)	28 (23,9%)	29 (24,6%)
Viszerale Metastasen (nur ein Organ betreffend)	41 (17,4%)	24 (20,5%)	17 (14,4%)
Klinische Subtyp			
	n = 218	n = 107	n = 111
HR+/HER2+	30 (13,8%)	18 (16,8%)	12 (10,8%)
HR+/HER2-	148 (67,9%)	71 (66,4%)	77 (69,4%)
HR-/HER2+	21 (9,6%)	10 (9,3%)	11 (9,9%)
TNBC	19 (8,7%)	8 (7,5%)	11 (9,9%)
Follow-Up-Zeit (Monate) – Median (Min–Max)	19 (0,4–34)	18 (0,4–33)	20 (2–34)

3.5.2 Gesundheitsökonomische Evaluation

EQ-5D-5L-Indexwerte und Qualität-adjustierte Lebensjahre (QALY)

Die vorliegende Analyse vergleicht die EQ-5D-5L-Indexwerte zu verschiedenen Zeitpunkten sowie die Qualität-adjustierten Lebensjahre (quality-adjusted life-years, QALY) zwischen der Interventions- und der Kontrollgruppe. Die Daten wurden sowohl ohne als auch mit multipler Imputation ausgewertet (Tabelle 19).

Zu Beginn der Studie waren die EQ-5D-5L-Indexwerte in beiden Gruppen vergleichbar, mit durchschnittlichen Werten von 79,4 in der Interventionsgruppe und 80,5 in der Kontrollgruppe. Nach sechs Monaten zeigte sich jedoch ein deutlicher Unterschied: Die Interventionsgruppe wies einen mittleren EQ-5D-5L-Wert von 77,5 auf, während die Kontrollgruppe nur 64,9 erreichte. Dieser Unterschied war statistisch signifikant mit einer unbereinigten mittleren Differenz von 12,6 (95%-KI: 3,8–21,4; $p = 0,005$) und einer adjustierten Differenz von 9,0 (95%-KI: 0,4–17,6; $p = 0,041$). Nach 12 Monaten setzte sich

dieser Trend fort: Die Interventionsgruppe erreichte einen Mittelwert von 71,0, während die Kontrollgruppe bei 57,4 lag. Auch hier ergab sich ein signifikanter Unterschied mit einer unbereinigten Differenz von 13,6 (95%-KI: 3,1–24,1; $p = 0,011$) und einer adjustierten Differenz von 11,7 (95%-KI: 0,7–22,7; $p = 0,037$).

Hinsichtlich der qualitätsadjustierten Lebensjahre (QALY) wurde über einen Zeitraum von 12 Monaten ein Zeit-gewichteter AUC-Wert (area under the curve) von 0,69 in der Interventionsgruppe und 0,58 in der Kontrollgruppe berechnet. Auch hier zeigte sich ein signifikanter Unterschied mit einer unbereinigten Differenz von 0,11 (95%-KI: 0,03–0,2; $p = 0,005$) und einer adjustierten Differenz von 0,12 (95%-KI: 0,05–0,2; $p = 0,001$). Die Ergebnisse wurden zusätzlich mit multipler Imputation überprüft. Die Trends blieben konsistent, mit signifikanten Differenzen zwischen den Gruppen. Nach sechs Monaten betrug die mittlere Differenz 9,6 (95%-KI: 1,2–18,0; $p = 0,025$, unbereinigt) bzw. 10,4 (95%-KI: 2,3–18,5; $p = 0,012$, adjustiert), während sie nach 12 Monaten auf 14,2 (95%-KI: 3,6–24,8; $p = 0,009$, unbereinigt) bzw. 14,8 (95%-KI: 4,5–25,0; $p = 0,005$, adjustiert) anstieg. Auch die QALY-Berechnung nach multipler Imputation zeigte einen signifikanten Unterschied zwischen beiden Gruppen.

Insgesamt legen die Ergebnisse nahe, dass die Intervention einen positiven Einfluss auf die Lebensqualität der Patientinnen hatte, sowohl in Bezug auf den EQ-5D-5L-Indexwert als auch auf die qualitätsadjustierten Lebensjahre.

Gesundheitskosten

Die vorliegende Analyse untersuchte die jährlichen Gesundheitskosten und die Kosten-Nutzen-Effizienz der durchgeführten Intervention im Vergleich zur Kontrollgruppe. Die Ergebnisse zeigen signifikante Unterschiede in verschiedenen Bereichen der Gesundheitsausgaben, einschließlich stationärer Versorgung, ambulanter Behandlung und Medikamentenverschreibungen.

Die durchschnittlichen Gesundheitskosten vor Beginn der Studie betragen in der Kontrollgruppe 5.582€ und in der Interventionsgruppe 6.935€, berechnet pro Patientin als die Summe der durchschnittlichen Kosten pro in Anspruch genommener Gesundheitsleistung. Die gesamten jährlichen Gesundheitskosten beliefen sich auf 70.477€ in der Kontrollgruppe und 71.843€ in der Interventionsgruppe. Besonders auffällig war, dass die Kosten für Medikamente in der Interventionsgruppe mit 52.873€ höher lagen als in der Kontrollgruppe mit 48.780€. Demgegenüber waren die Kosten für stationäre Krankenhausaufenthalte in der Interventionsgruppe mit 15.472€ niedriger als in der Kontrollgruppe mit 17.298€ (Tabelle 20).

Tabelle 19: EQ-5D-5L-Indexwerte nach Zeitpunkten und qualitätsadjustierten Lebensjahre (QALY)

	Interventionsgruppe		Kontrollgruppe		Unadjustiert		Adjustiert	
	n	Mittelwert (95%-KI)	n	Mittelwert (95%-KI)	Mittlere Differenz (95%-KI)	p-Wert	Mittlere Differenz (95%-KI)	p-Wert
Keine multiple Imputation								
EQ-5D-5L-Indexwerte								
Baseline	115	79,4 (75,3 – 83,4)	113	80,5 (77,3 – 83,8)				
6 Monate	96	77,5 (71,6 – 83,3)	77	64,9 (58,3 – 71,4)	12,6 (3,8 – 21,4)	0,005	9,0 (0,4 – 17,6)	0,041
12 Monate	84	71,0 (64,0 – 77,9)	61	57,4 (49,5 – 65,3)	13,6 (3,1 – 24,1)	0,011	11,7 (0,7 – 22,7)	0,037
QALY								
Zeit-gewichteter AUC-Wert	107	0,69 (0,64 – 0,73)	88	0,58 (0,53 – 0,63)	0,11 (0,03 – 0,2)	0,005	0,12 (0,05 – 0,2)	0,001
Multiple imputation (m = 30)								
EQ-5D-5L-Indexwerte								
Baseline	118	79,0 (75,4 – 82,6)	117	80,0 (76,3 – 83,6)				
6 Monate	118	76,6 (70,6 – 82,5)	117	67,0 (60,4 – 73,6)	9,6 (1,2 – 18,0)	0,025	10,4 (2,3 – 18,5)	0,012
12 Monate	118	65,3 (57,9 – 72,7)	117	51,2 (43,4 – 59,1)	14,2 (3,6 – 24,8)	0,009	14,8 (4,5 – 25,0)	0,005
QALY								
Zeit-gewichteter AUC-Wert	118	0,74 (0,69 – 0,79)	117	0,66 (0,61 – 0,72)	0,08 (0,008 – 0,15)	0,029	0,08 (0,02 – 0,14)	0,008

PRO B (01NVF19013)

Tabelle 20: Kosten für das Gesundheitswesen (€) – Gesamtkosten pro Jahr

Gesundheitsleistung	Studiengruppe	Anzahl	Gesamtkosten pro Jahr ^b (€)				
			Mittelwert	SD	Median	P25	P75
Kosten vor PRO B ^a	Kontrolle	117	5.582	7.116	3.066	1.407	6.335
	Intervention	118	6.935	9.162	3.791	1.554	6.268
Gesamtkosten	Kontrolle	117	70.477	38.037	61.824	39.280	92.664
	Intervention	118	71.843	40.417	64.072	41.213	87.975
PRO B-Intervention	Kontrolle	117	–	–	–	–	–
	Intervention	118	1.040	–	–	–	–
Notaufnahme im Krankenhaus	Kontrolle	94	2.270	4.540	1.428	769	2.433
	Intervention	102	1.796	1413	1.347	772	2.423
Stationäre Krankenhausbehandlung	Kontrolle	90	17.298	22.744	7.927	2.652	21.078
	Intervention	78	15.472	22.029	8.071	3.140	18.653
Verschreibung von Medikamenten	Kontrolle	117	48.780	32.828	37.909	28.015	62.844
	Intervention	118	52.873	34.059	43.521	29.910	69.945
Ambulante Behandlung	Kontrolle	114	2.681	3.311	1.829	851	3.271
	Intervention	116	2.495	2.579	1.353	817	3.346
Verordnungen für Heilmittel	Kontrolle	73	1.592	1.444	1.198	396	2.370
	Intervention	71	2.321	3.917	1.053	288	2.722
Verordnungen für medizinische Hilfsmittel	Kontrolle	87	1.339	2.251	532	208	1571
	Intervention	88	932	1.346	377	196	968
Sonstige (AU, Transport, Haushaltshilfe)	Kontrolle	73	3.150	4.844	1.698	950	3.391
	Intervention	73	2.597	4.360	906	268	2.279

^a Durchschnittliche Kosten vor Beginn der Studie pro Patientin (berechnet als die Summe der durchschnittlichen Kosten pro in Anspruch genommener Gesundheitsleistung pro Patientin).

^b Zur Berücksichtigung individueller Unterschiede in der Studiendauer wurden die Kosten mit dem Beobachtungszeitraum gewichtet, um die Gesamtkosten pro Jahr abzuleiten.
AU = Arbeitsunfähigkeit

Kosteneffektivitätsanalysen

Im Rahmen der Kosteneffektivitätsanalysen wurden die qualitätsadjustierten Lebensjahre (QALY) berechnet. Die Interventionsgruppe erreichte einen durchschnittlichen QALY-Wert von 0,745, während die Kontrollgruppe bei 0,662 lag. Die inkrementellen Kosten der Intervention betrugen -846€, was darauf hindeutet, dass die Intervention kosteneffektiv war. Die inkrementellen QALYs lagen bei 0,081, und die Kosten-Nutzen-Analyse ergab, dass die Intervention ökonomisch vorteilhaft war (Tabelle 21).

Abbildung 17 zeigt die auf Bootstrapping basierenden Schätzungen der inkrementellen Kosten und inkrementellen Effekte, dargestellt auf der Kosteneffektivitäts-Ebene für die PRO B Studie. Die Mehrheit der Punkte liegt rechts der y-Achse, was darauf hindeutet, dass die Intervention der PRO B Studie wahrscheinlich zu einem Anstieg der gewonnenen QALYs führt. Die Verteilung der Punkte sowohl oberhalb als auch unterhalb der x-Achse weist auf die Unsicherheit hinsichtlich des Einflusses der PRO B-Intervention auf die Gesundheitskosten hin. Die PRO B-Intervention begann mit einer Wahrscheinlichkeit von 64%, kosteneffektiv zu sein, und stieg bei einem Schwellenwert von 20.000€ pro QALY auf 84% und bei einem Schwellenwert von 50.000€ pro QALY auf 97% wahrscheinlicher Kosteneffektivität (Abbildung 18).

Tabelle 21: Ergebnisse der Kosteneffizienzanalyse – Gesamtkosten pro Jahr (€)

	Interventionsgruppe (n = 118)	Kontrollgruppe (n = 117)
Gesamtkosten (€)	62.182 ^a	62.568 ^a
Gesamt QALY	0,745 ^b	0,662 ^b
Inkrementelle Kosten^c (€)	-846 (95%-KI: -5.419; 3.902)	
Inkrementelle QALY^c	0,081(95%-KI: 0,056; 0,106)	
ICER^c (€/QALY)	Zugunsten von PRO B	

^aKosten (€) ist ein geometrisches Mittel, das der rücktransformierte Mittelwert aus transformierten Daten ist (log10-Kosten). Der geschätzte Mittelwert der log10-Kosten wird auf eine exponentielle Skala rücktransformiert, die rücktransformierten Kosten werden anschließend bereinigt um den *Duan's smearing factor*. Die geschätzten Kosten stammen aus einem linearen gemischten Modell mit Zufallseffekt (random effect) für Studienzentrum und kontrolliert für die Variablen Alter bei der Randomisierung, den histologischen Befund, die Art der Metastasierung und die Kosten vor der Studie (log10-Transformation).

^bQALYs wurden geschätzt anhand eines gemischten Modells mit zufälligem Achsenabschnitt (random intercept) nach Studienzentrum, kontrolliert für die Variablen EQ-5D-5L-Indexwerte (Mehrfach-Imputation QALY, $m = 30$).

^cDie ICER wurden mittels nichtparametrischem Bootstrapping geschätzt, um 5.000 Schätzungen der inkrementellen Kosten und Effekte zu erhalten. Diese Schätzungen wurden auf der Kosten-Effektivitäts-Ebene aufgetragen, um die Unsicherheit im Zusammenhang mit dem Kosten-Effektivitäts-Punktschätzer zu veranschaulichen.

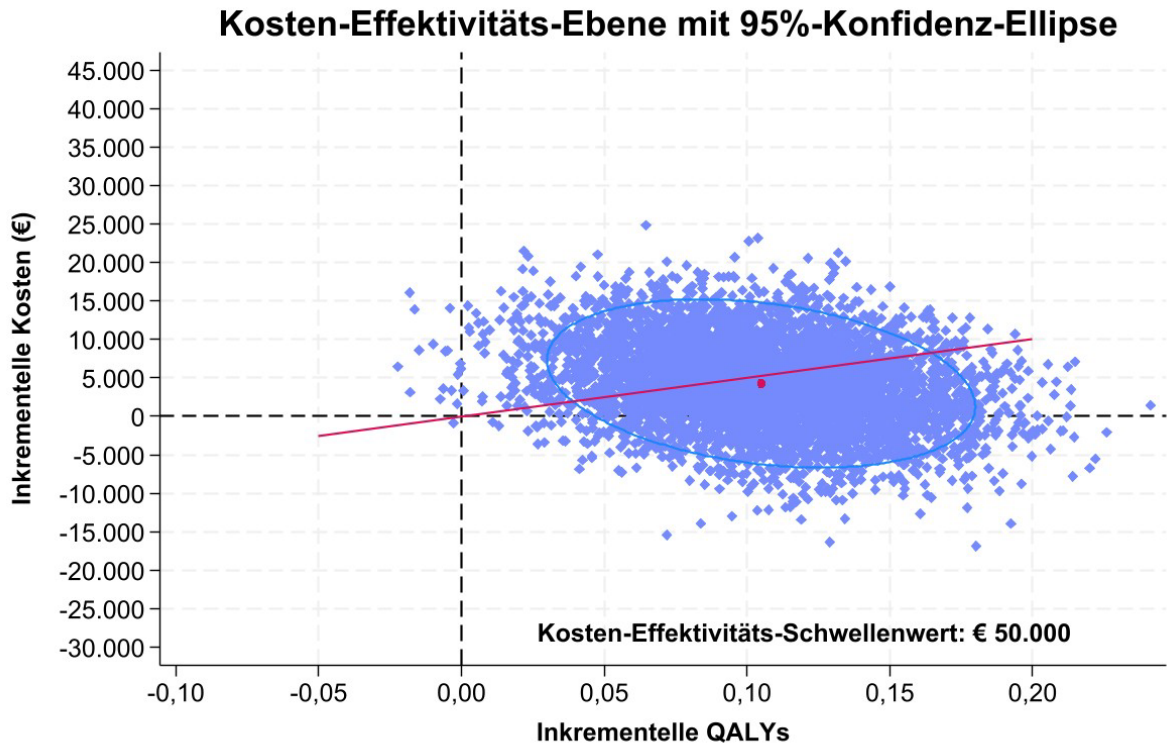


Abbildung 17: Kosten-Effektivitäts-Ebene mit 95%-Konfidenzniveau Ellipse

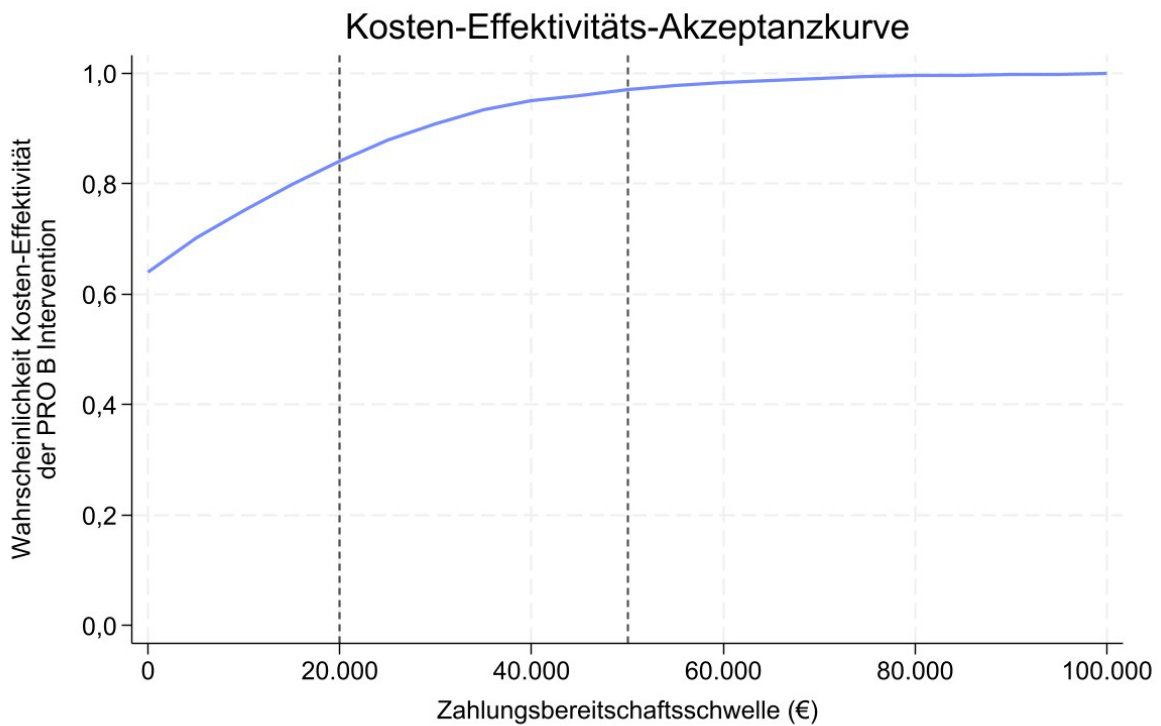


Abbildung 18: Kosten-Effektivitäts-Akzeptanzkurve

Sensitivitätsanalysen zur Kosteneffektivität

Die Sensitivitätsanalysen bestätigten diese Ergebnisse weitgehend (Tabelle 22). Die Berechnungen ohne multiple Imputation zeigen, dass die Kosten in der Interventionsgruppe bei 64.055€ und in der Kontrollgruppe bei 58.339€ lagen, während die QALY-Werte in der Interventionsgruppe mit 0,789 deutlich höher waren als in der Kontrollgruppe mit 0,581. Die Kosten-Nutzen-Analyse auf Basis der Log10-Transformation ergab ebenfalls, dass die Intervention vorteilhaft war, mit einem inkrementellen Kosten-Effektivitäts-Verhältnis (ICER) von 18.800€ pro QALY.

3.5.3 Anzahl der Krankenhaus- und Rettungststellenaufenthalte

Es wurde ebenfalls die Anzahl der Krankenhausaufenthalte untersucht (Tabelle 23). Die Anzahl der stationären Aufnahmen pro Patientenjahr war in der Interventionsgruppe mit 1,4 niedriger als in der Kontrollgruppe mit 1,6. Die Rate der ambulanten Krankenhausbesuche lag in der Interventionsgruppe bei 3,3 pro Patientenjahr und war damit ebenfalls niedriger als in der Kontrollgruppe mit 3,8. Allerdings erreichten weder die unbereinigte noch die bereinigte IRR statistische Signifikanz, so dass wir auf der Grundlage dieser Daten nicht abschließend feststellen können, dass die Intervention die Zahl der Krankenhauseinweisungen verringerte, obwohl das bereinigte Modell einen Trend zur Signifikanz nahelegt.

Tabelle 22: Ergebnisse der Sensitivitätsanalysen zur Kosteneffizienz

Sensitivitätsanalysen		Gesamtkosten (€)	Gesamt QALY	Inkrementelle Kosten ^c (€) (95%-bCI)	Inkrementelle QALY ^c (95%-bCI)	ICER ^c (€/QALY) (95%-bCI)
Keine multiple Imputation für QALY						
n = 195						
Median-basiert						
Interventionsgruppe	107	64.055	0,789	4.976	0,202	40.475
Kontrollgruppe	88	58.339	0,581	(-9.475; 18.601)	(0,019; 0,338)	(-69.763; 329.873)
Log10-Transformation						
Unadjustierte Kosten						
Interventionsgruppe	107	63.958	0,692 ^b	4.676	0,114	56.198
Kontrollgruppe	88	58.849	0,581 ^b	(-5944; 15.091)	(0,045; 0,181)	(-50.557; 201.439)
Adjustierte Kosten						
Interventionsgruppe	107	61.842 ^a	0,692 ^b	1.585	0,113	20.955
Kontrollgruppe	88	59.763 ^a	0,581 ^b	(-7.989; 10.876)	(0,046; 0,177)	(-69.773; 152.769)
Multiple Imputation QALY, m = 30						
n = 235						
Median-basiert						
Interventionsgruppe	118	64.072	0,823	1.568	0,059	31.436
Kontrollgruppe	117	61.824	0,767	(-7.091; 9.088)	(0,028; 0,094)	(-140.372; 201.013)
Log10-Transformation						
Unadjustierte Kosten						
Interventionsgruppe	118	63.381	0,745 ^b	1.389	0,081	18.800
Kontrollgruppe	117	61.840	0,662 ^b	(-3.707; 6.562)	(0,055; 0,107)	(-46.089; 94.247)

^aKosten (€) ist ein geometrisches Mittel, das der rücktransformierte Mittelwert aus transformierten Daten ist (log10-Kosten). Der geschätzte Mittelwert der log10-Kosten wird auf eine exponentielle Skala rücktransformiert, die rücktransformierten Kosten werden anschließend bereinigt um den *Duan's smearing factor*. Die geschätzten Kosten stammen aus einem linearen gemischten Modell mit Zufallseffekt (random effect) für Studienzentrum und kontrolliert für die Variablen Alter bei der Randomisierung, den histologischen Befund, die Art der Metastasierung und die Kosten vor der Studie (log10-Transformation).

^bQALYs wurden geschätzt anhand eines gemischten Modells mit zufälligem Achsenabschnitt (random intercept) nach Studienzentrum, kontrolliert für die Variablen EQ-5D-5L-Indexwerte (Mehrfach-Imputation QALY, m = 30).

^cDie ICER wurden mittels nichtparametrischem Bootstrapping geschätzt, um 5.000 Schätzungen der inkrementellen Kosten und Effekte mit einem 95%-Perzentil-Bootstrap-Konfidenzintervall (bCI) zu generieren – dem Intervall zwischen dem 2,5. und 97,5-Perzentil der 5.000 Bootstrap-geschätzten Parameter.

Tabelle 23: Inzidenzrate der Krankenhausaufenthalte pro Personenjahr

	Anzahl der Ereignisse	Personenjahr	Inzidenzrate (pro Personenjahr) (95%-KI)	Unadjustiert		Adjustiert	
				IRR (95%-KI)	p-Wert	IRR (95%-KI)	p-Wert
Aufnahmen im Krankenhaus (stationäre + teilstationäre)							
Kontrollgruppe	274	174,4	1,6 (1,4 – 1,8)	Referenz		Referenz	
Interventionsgruppe	252	182,4	1,4 (1,2 – 1,6)	0,79 (0,57 – 1,10)	0,160	0,74 (0,54 – 1,03)	0,073
Ambulante Krankenhausaufenthalte							
Kontrollgruppe	670	174,4	3,8 (3,6 – 4,1)	Referenz		Referenz	
Interventionsgruppe	606	182,4	3,3 (3,1 – 3,6)	0,91 (0,78 – 1,07)	0,250	0,89 (0,76 – 1,04)	0,151

Die Inzidenzraten stellen die durchschnittliche Anzahl der Krankenhausbesuche pro Personenjahr für jede Gruppe dar. Die Inzidenzraten wurden anhand eines Mehrebenenmodells mit gemischten Effekten und negativer Binomialregression mit Zufallseffekt (random effect) für das Studienzentrum und kontrolliert für die Variablen Alter bei Randomisierung, histologischer Befund, Art der Metastasierung und ECOG-Wert bei Studienbeginn geschätzt.

3.5.4 Dauer des Krankenhausaufenthalts

Die durchschnittliche Verweildauer im Krankenhaus war in der Interventionsgruppe mit vier Tagen (Median) kürzer als in der Kontrollgruppe mit sechs Tagen. Es zeigte sich kein Unterschied in der durchschnittlichen Verweildauer pro Patientin zwischen der Interventions- und der Kontrollgruppe. Allerdings war die durchschnittliche Verweildauer in der Interventionsgruppe 11% kürzer als in der Kontrollgruppe (Tabelle 24).

Tabelle 24: Anzahl und Dauer der Krankenhausaufenthalte pro Patientin

	Interventionsgruppe	Kontrollgruppe	Exp.(β) ^a	95%-KI	p-Wert
Anzahl der Krankenhausaufenthalte	252	274			
Anzahl der Krankenhausaufenthalte pro Patientin – Median (IQR)	1 (0–3)	1 (1–4)			
LOS (Tage) – Median (IQR)	4 (2–11)	6 (2–13)	0,77	(0,65 – 0,92)	0,003
Durchschnittliche LOS (Tage) pro Patientin – Median (IQR)	5 (3–9)	5 (3–8)	0,89	(0,70 – 1,14)	0,371

^a Die Aufenthaltsdauer (LOS) in Tagen pro Patientin wurde log-transformiert und zwischen den Gruppen mit Hilfe eines linearen gemischten Modells verglichen, das einen Zufallseffekt (random effect) für das Studienzentrum enthielt. Die Schätzung basierte auf der Potenzierung des Regressionskoeffizienten (β) für die log-transformierte LOS, wobei $\exp(\beta)$ den exponentiellen Regressionskoeffizienten der log-transformierten LOS darstellt.

3.6 Prozessevaluation

3.6.1 Neue Versorgungsform aus Sicht der Versorgenden

Die endgültige Stichprobe zur Prozessevaluation aus Perspektive des Behandlungspersonals umfasste 15 Interviewpartner aus 12 zertifizierten Brustkrebszentren in ganz Deutschland. Es wurden ein männlicher und 14 weibliche Mitglieder des Behandlungsteams mit einem Durchschnittsalter von 42 Jahren interviewt. Die Stichprobe der Versorgenden umfasste Ärzte, Studienschwestern und Dokumentationsassistenten. Von den 12 Krebszentren befanden sich fünf in Universitätskliniken. Die Interviews dauerten zwischen neun und 31 Minuten. Im Folgenden werden die Ergebnisse der Interviews basierend auf den CFIR-Domänen dargestellt (vgl. 2.12.7). Innerhalb der fünf Domänen des CFIR wird insbesondere die Domäne **Innovation** hervorgehoben, welche sich mit den innovativen Aspekten der zu implementierenden neuen Intervention beschäftigt.

Die Versorgenden empfanden die PRO B Studie in ihrer Routinearbeit als Innovation. Der innovative Ansatz der PRO B Studie unterschied sich dabei deutlich von anderen PRO-Studien, an denen sie in der Vergangenheit teilgenommen hatten. Dies lag insbesondere daran, dass sie in der neuen Versorgungsform in PRO B mit den PRO-Ergebnissen der Patientinnen interagieren mussten, wenn ein Alarm sie über eine Verschlechterung der Werte benachrichtigte. Diese Interaktion stand im Gegensatz zu anderen Studien, bei denen PROs lediglich erhoben wurden. Die Versorgenden empfanden das alarmbasierte PRO-Monitoring, das durch die PRO B Studie eingeführt wurde, im Allgemeinen als sinnvoll. Sie sahen das Potenzial dieses innovativen Ansatzes zur Unterstützung von Patientinnen und die generelle Relevanz der Forschung zu PROs. Sie äußerten jedoch auch Einschränkungen, z.B. den Eindruck, dass PRO-Fragebögen nur den Zustand der Patientin zu einem bestimmten Zeitpunkt widerspiegeln, während ein Gespräch mit einem Arzt ein ganzheitlicheres Bild ergeben kann. Eine ärztliche Interviewteilnehmerin erwähnte außerdem, dass sie durch die Messung der Lebensqualität keinen tatsächlichen praktischen Vorteil sehe: „Deswegen also, hmm, also ich bin so ein bisschen gespalten, also gibt es ja Studien, die durchaus zeigen, dass Lebensqualität super wichtig ist und wo man auch mit diesem ganzen Management eine

deutliche Verbesserung bringen kann, aber ich sehe es jetzt für uns noch nicht so gravierend.“ (Ärztin). Hinsichtlich der wahrgenommenen Vorteile der innovativen Versorgungsform beschrieben die Versorgenden ähnliche Erfahrungen wie die Patientinnen in der PRO B Studie – sie beobachteten, dass die Patientinnen die Anrufe schätzten und sich während der Intervention sicher und gut aufgehoben fühlten: *„Das wird gut aufgenommen. Also ich denke, die sind da auch zugetan, wenn auch quasi nochmal, nochmal zusätzlich jemand nach ihnen fragt wie es ihnen geht und die führen das ja auch immer in der Regel alle aus. Also habe ich bis jetzt nur Positives gehört.“* (Ärztin).

Zudem berichtete das Behandlungspersonal, dass sich die Kommunikation zwischen Behandlungsteam und Patientinnen durch die Telefonkontakte nach einem PRO-basierten Alarm verbesserte. Studienkrankenschwestern wurden als zusätzliche Ansprechpartnerinnen wahrgenommen. Patientinnen riefen sogar proaktiv an, da sie aufgrund der Studie einen Hauptansprechpartner im behandelnden Zentrum hatten: *„Und man hat halt auch noch wieder ein zusätzliches Gesicht, also da ich ja jetzt auch immer auf die Chemo übergegangen bin, war man halt noch einen zusätzlichen Ansprechpartner auch noch, weil die Ärzte ja immer recht knappes Zeitkontingent haben, dass da die Studienassistentin doch ein bisschen flexibler ist.“* (Study Nurse). Die Versorgenden waren der Meinung, dass die Telefongespräche mit den Patientinnen in den meisten Fällen Beratungscharakter hatten und sich auf emotionale und psychologische Unterstützung, die Bereitstellung medizinischer Informationen oder den Umgang mit Nebenwirkungen konzentrierten. *„Ähm eher Beratung, also in meinem Fall waren es noch nie medikamentöse Therapie, die eingeleitet werden musste, ähm wir hatten es mal, dass jemand Bedürfnis hatte, da ging es um glaube ich um Lymphödeme und sowas, wo dann einfach Beratung, welche Möglichkeiten es noch gibt zur Schmerzlinderung, das hat dann tatsächlich ein Arzt übernommen, die Patientin war dann tatsächlich am gleichen Tag später noch in der Klinik zum Routinetermin, aber da hatte ich dann vorab die Ärztin schon informiert, was da eben ansteht als und dann wurde da noch beraten, aber tatsächlich noch nie medikamentöse Intervention.“* (Study Nurse).

Es wurde rückgemeldet, dass klinische Maßnahmen in Form von Therapieanpassungen aufgrund von Alarmen selten erfolgten. Allerdings seien die Alarme hilfreich dafür gewesen, sich auf bevorstehende Konsultationen mit einer Patientin vorzubereiten, z.B. für administrative Prozesse im Gesundheitswesen wie die Überweisung zu Rehabilitationsleistungen oder die Kontaktaufnahme mit dem interdisziplinären Team in der Pflege.

„Genau, genau und [die Ärztin] konnte sich vorab auch schonmal informieren, da gings dann um Rehaanträge und sowas, wie sie da weiterhelfen konnte und das hatten wir dann am gleichen Tag lösen können, also das war Zufall, dass sie dann einen Termin hatte, genau dann wusste die Ärztin schon Bescheid und dann gings.“ (Study Nurse)

„Dann informiert die Ärztin, wenns ganz gravierend ist, sofort die Psychologin bzw. die Psychoonkologin, wenn ich die [Patientin] sehe, rufe ich dann an und sag ich hab hier Frau Soundso, ja aber da die bei uns halt auch in Anbindung sind, sagen die ganz oft ja ich hab schon einen Termin bei Frau PunktPunktPunkt [Psychoonkologin] dann und dann.“ (Study Nurse)

Zusammenfassend wurde die neue Intervention als positiv und hilfreich auch von Seiten des Behandlungsteams bewertet.

3.6.2 Neue Versorgungsform aus Sicht der Patientinnen

Semi-strukturierte Interviews

An den Interviews zur Prozessevaluation aus Patientinnensicht nahmen acht Patientinnen mit einem Durchschnittsalter von 59 Jahren (Bereich 42–74) teil, deren Erstdiagnose Brustkrebs im Durchschnitt 12 Jahre zurücklag. Die Interviews dauerten zwischen 10 und 34 Minuten. Die

befragten Patientinnen hatten heterogene Lebenssituationen: Sie lebten allein, lebten mit einem Partner, lebten zusammen mit Partner und Kindern oder lebten mit einem Kind, aber ohne Partner. Auch der berufliche Hintergrund (Ingenieurin, Geschäftsfrau, Podologin, Anästhesieschwester, Sachbearbeiterin und Büroassistentin) und der aktuelle Arbeitsstatus (Berufstätigkeit, Krankenstand, berufliche Wiedereingliederung, Erwerbsminderungsrente, Altersrente) variierten zwischen den Teilnehmerinnen.

Die Patientinnen nahmen überwiegend mehrere Vorteile und Nutzen der neuen Versorgungsform wahr. Zunächst berichteten sie, dass sie sich aufgrund ihrer Teilnahme an der PRO B Studie sicher und in guten Händen fühlten. Sie schätzten es, dass sich im Hintergrund jemand um sie kümmerte und auf Verschlechterungen reagieren konnte, was sie als beruhigend empfanden. Es wurde jedoch auch erwähnt, dass das wöchentliche Ausfüllen von Fragebögen das Potenzial hatte, Ängste auszulösen, indem es die Aufmerksamkeit auf Probleme lenkte, die man möglicherweise (noch) nicht hatte. Darüber hinaus berichteten die Patientinnen in den Interviews, dass das PRO-Monitoring ihre Kommunikation mit dem Behandlungsteam erleichterte, indem es durch die alarmbasierten Anrufe zusätzliche Kontaktpunkte bot. Diese ermöglichten es den Patientinnen, Fragen zu stellen oder Bedenken außerhalb der regelmäßig geplanten Termine in ihrem Brustzentrum zu besprechen. Einige Patientinnen beschrieben sogar, dass sie im Laufe der Zeit ihr Antwortverhalten angepasst hatten, um eine telefonische Kontaktaufnahme ihres Behandlungsteams auszulösen. Schließlich empfanden die Patientinnen die Aufmerksamkeit des Behandlungsteams auf ihre Symptome als Vorteil. Während Maßnahmen wie die Therapieanpassung als Ergebnis der telefonischen Interventionen in den Interviews kaum berichtet wurden, empfanden die Patientinnen das Gespräch mit einem Behandelnden selbst als sehr positiv und hilfreich, da es emotionale Unterstützung bot, Informationen vermittelte, Unsicherheit abbauete und Ratschläge zum Umgang mit spezifischen Symptomen gab. Es wurde jedoch auch angesprochen, dass zentrale Ansprechpartner für die Patientinnen in den Zentren unerlässlich seien, um dieses Gefühl der Sicherheit und des Aufgehobenseins zu erzeugen. Wenn dies nicht gegeben sei, werde die telefonische Intervention eher als anonym und infolgedessen lästig empfunden. Insgesamt empfanden die Patientinnen die Nutzung des alarmbasierten PRO-Monitorings (einschließlich des Ausfüllens wöchentlicher Fragebögen und des Empfangens von Telefonanrufen) überwiegend als wenig aufwändig und nicht lästig. Nur eine Patientin merkte an, dass die Intervention in einigen Momenten lästig sein könne. Bezüglich des Umgangs mit der neuen Versorgungsform sprachen die Patientinnen in den Interviews zwei Themen an: technische Aspekte sowie die Verständlichkeit und den Inhalt der Befragungen. In Bezug auf technische Aspekte empfanden die Patientinnen die Handhabung der App auf ihren Handys als einfach und intuitiv. Einige Patientinnen erwähnten kleinere technische Probleme, beispielsweise bei der Aktualisierung oder Synchronisierung der App. Eine Patientin erwähnte, dass die verwendete App weitere Funktionen enthielt, die jedoch in der PRO B Studie nicht verwendet wurden und daher irritierend waren (beispielsweise eine Erinnerungsfunktion für die Medikamenteneinnahme). In Bezug auf den wöchentlichen Fragebogen wurden mehrere Unsicherheiten geäußert: Einige Patientinnen hatten Probleme, die im Fragebogen verwendeten Antwortoptionen des deutschen EORTC QLQ-C30 zu interpretieren (deutsche Übersetzung von „ziemlich viel“ als „mäßig“). Darüber hinaus kritisierten sie eine Redundanz der Fragen, Schwierigkeiten bei der Beantwortung der Fragen in der Retrospektive („während der letzten Woche“), da die Symptome innerhalb einer Woche von Tag zu Tag stark variieren können, und eine Unsicherheit, ob sie Fragen allgemein oder nur in Bezug auf ihre Krebserkrankung beantworten sollten.

Evaluationsbefragung zum Studienteilnehmer

Wie initial geplant erhielten die Patientinnen beider Gruppen nach Abschluss der Datenerhebung ihre gemessenen PRO-Werte postalisch zur Verfügung gestellt. In diesem Zusammenhang erfolgte auch eine Evaluationsbefragung der Teilnehmerinnen in der Interventionsgruppe hinsichtlich ihrer Einschätzung der Intervention.

Von den 269 versendeten Evaluationsfragebögen erhielten wir Antworten von 152 Patientinnen (56,5 %) aus 28 verschiedenen Zentren. Dies entspricht einer Rücklaufquote von 56,5%. Die Vollständigkeit der Fragebögen ist in Tabelle 25 dargestellt.

Tabelle 25: Anzahl der vorhandenen und fehlenden Antworten der Evaluationsfragebögen pro Item.

Item	N beantwortet	N fehlend
Die Länge der Fragebögen war für mich angemessen.	152	
Wie lange haben Sie in der Regel etwa gebraucht, um einen Wochen-Fragebogen in der App auszufüllen	152	
Die Fragen und Antwortmöglichkeiten waren verständlich für mich.	152	
Durch das Ausfüllen der Fragebögen habe ich besser erkannt, wie es mir wirklich geht.		
Die Fragen in den Fragebögen waren insgesamt relevant für mich.	151	1
Fehlten aus Ihrer Sicht inhaltliche Aspekte im Fragebogen, die für Sie ebenfalls relevant gewesen wären?	138	14
Haben Sie mindestens einmal einen Anruf im Rahmen der PRO B Studie erhalten?	148	4
Mich haben die Anrufe zur PRO B Studie gestört.	136	6 (nicht zutreffend: 10)
Die Anrufe zur PRO B Studie waren zu häufig.	137	5 (nicht zutreffend: 10)
Bei den Anrufen zur PRO B Studie fühlte ich mich ernst genommen.	138	4 (nicht zutreffend: 10)
Bei den Anrufen zur PRO B Studie hatte ich ausreichend Gelegenheit, um über mein Befinden zu sprechen.	137	4 (nicht zutreffend: 10)
Ich habe mich durch die Anrufe zur PRO B Studie zusätzlich unterstützt gefühlt.	138	4 (nicht zutreffend: 10)
Durch meine Teilnahme an der PRO B Studie ...		
... war ich häufiger in Kontakt mit meinem Behandlungsteam als vor der Studie.	150	2
... war ich schneller in Kontakt mit meinem Behandlungsteam als vor der Studie.	150	2
... konnte ich dem Behandlungsteam meinen Gesundheitszustand besser mitteilen als vor der Studie.	149	3
... hatte ich ein zusätzliches Sicherheitsgefühl.	150	2
Die Teilnahme an der PRO B Studie hatte insgesamt einen Mehrwert für mich.	150	2
Würden Sie sich wünschen, dass die digitale Betreuung aus der PRO B Studie deutschlandweit als Teil der regulären Versorgung für alle Patientinnen mit fortgeschrittenem Brustkrebs angeboten werden würde?	150	2
Können Sie sich vorstellen, weiter langfristig an einer Befragung mit Anrufen bei Verschlechterung (wie in der PRO B Studie) teilzunehmen?	147	5
Bevorzugte Befragungsabstände	136	3 (nicht zutreffend: 13)

Hier berichteten 72,9% der Patientinnen, dass sie durch die Nutzung der neuen Versorgungsform besser in der Lage waren, ihren Gesundheitszustand gegenüber dem

PRO B (01NVF19013)

Behandlungsteam zu kommunizieren. 59,9% der Patientinnen haben im Rahmen der Intervention weitere Empfehlungen zum Umgang mit Nebenwirkungen erhalten. In Bezug auf die persönliche Einschätzung berichteten 83,8% der Patientinnen, dass ihnen durch die Intervention ein zusätzliches Sicherheitsgefühl vermittelt wurde, und 91,1% der Patientinnen befürworteten die Einführung der neuen Intervention als Teil der Regelversorgung bei Patientinnen mit einer fortgeschrittenen Brustkrebserkrankung.

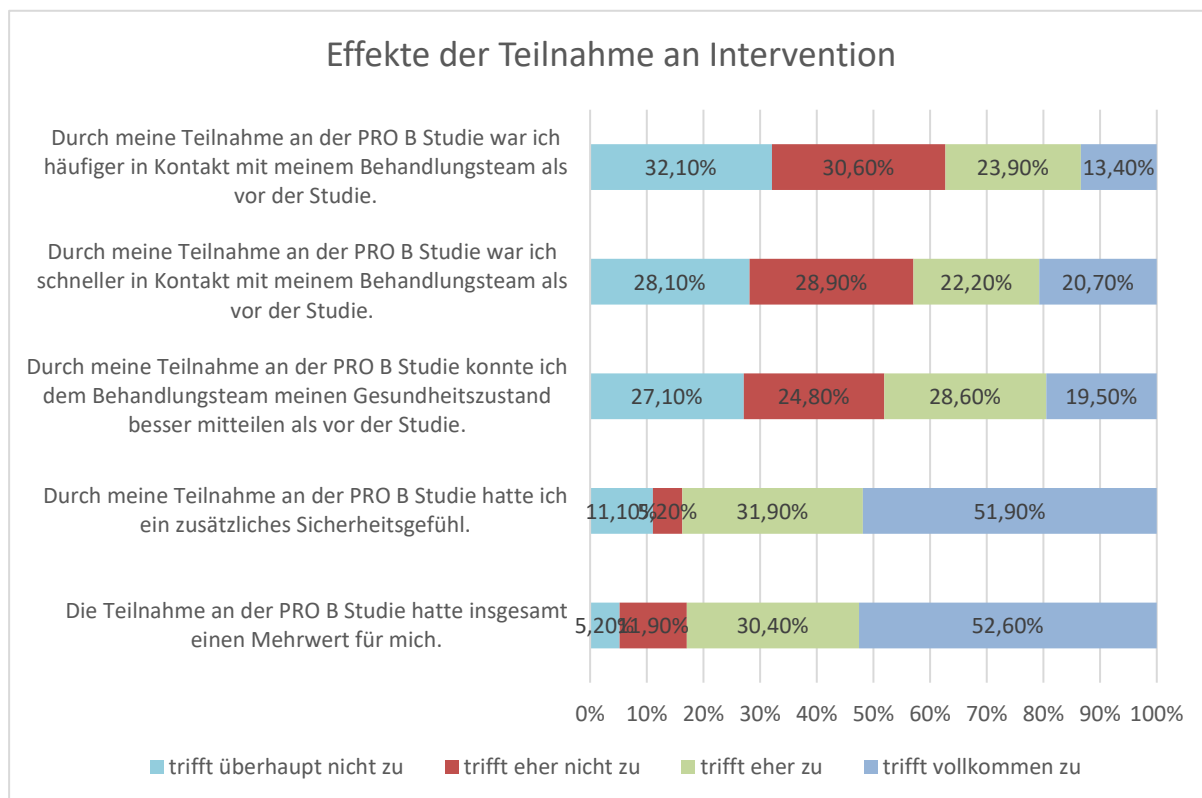


Abbildung 19: Effekte der Teilnahme an der neuen Versorgungsform aus Patientinnenperspektive

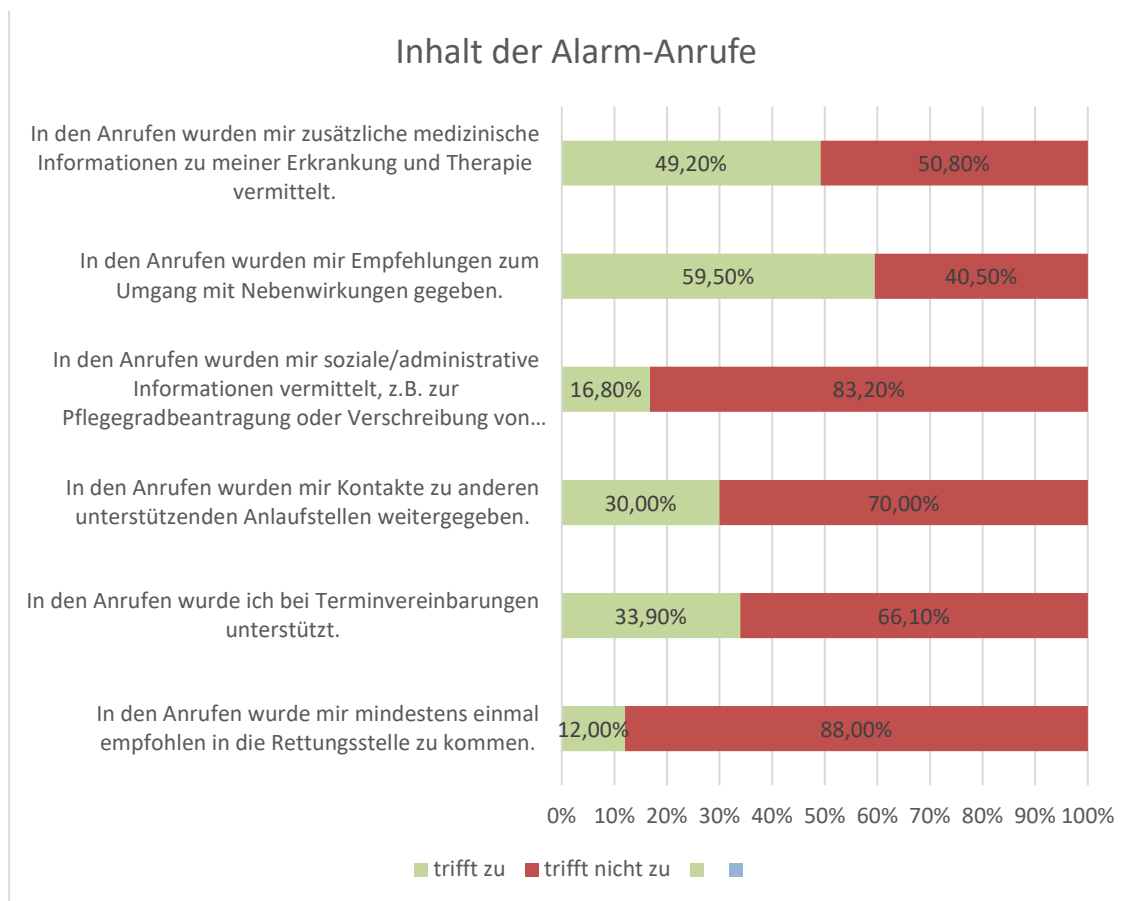


Abbildung 20: Inhalte der Alarmanrufe aus Patientinnenperspektive

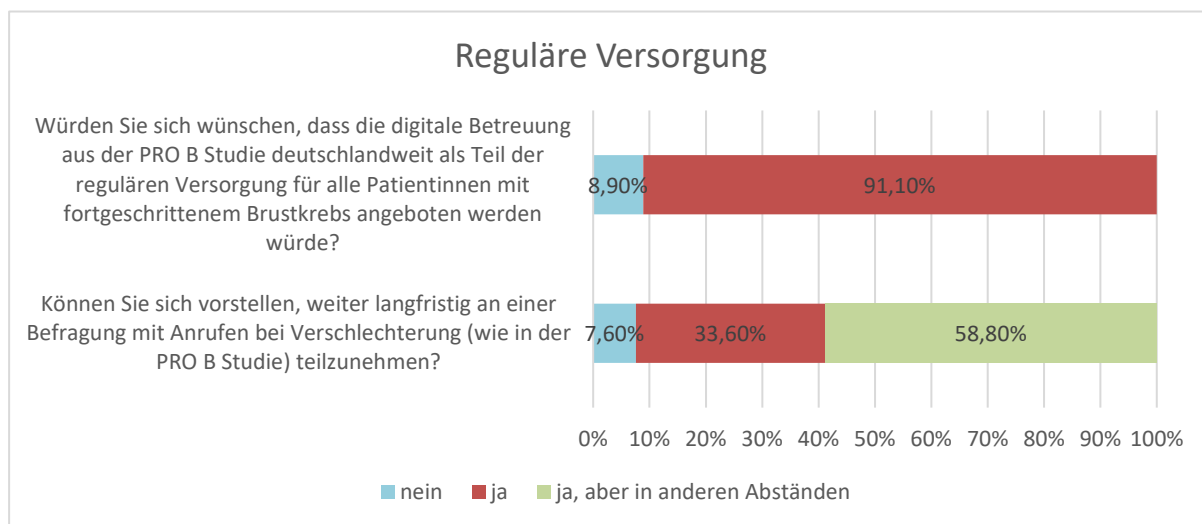


Abbildung 21: Wunsch auf Übernahme in die Routineversorgung aus Patientinnenperspektive

3.6.3 Evaluation der graphischen Visualisierung von PRO-Werten

Qualitative Ergebnisse: Semi-strukturierte Interviews

Insgesamt wurden acht Patienten und fünf Gesundheitsdienstleister aus vier verschiedenen PRO B-Studienzentren befragt. Das Durchschnittsalter der Teilnehmerinnen betrug 48,3 Jahre (SD = 7,8). Alle durchgeführten Interviews wurden aufgezeichnet, transkribiert und in die qualitative Inhaltsanalyse einbezogen. Aus den Interviews konnten acht verschiedene thematische Kategorien abgeleitet werden (Achsenbeschriftungen, Farbschema, Hintergrundlinien, Alarmdarstellung, Referenzdaten, fehlende Daten, Format und zusätzliche Informationen sowie Verwendung der Visualisierungen), und die Ergebnisse wurden nach Unterthemen innerhalb jeder Kategorie gruppiert. Die Design-Entscheidungen für die finalen Grafiken wurden in einem Bericht festgehalten, die als Anlage 1 beigefügt ist.

i. Achsenbeschriftungen

Teilnehmerinnen betonten die Bedeutung von Klarheit vor Ästhetik. Obwohl der Farbverlauf zur Darstellung der Ausprägung der Werte meist korrekt interpretiert wurde, wurden klare und konsistente Beschriftungen bevorzugt, um Mehrdeutigkeiten zu vermeiden. Beschriftungen erleichterten das Verständnis, ob hohe Werte „gut“ oder „schlecht“ sind, besonders bei Patientinnen mit kognitiven Einschränkungen durch die Krebstherapien. Zusätzliche Richtungspfeile wurden als unnötig und störend empfunden. Auch Fachpersonen empfanden Beschriftungen als hilfreich für eine schnelle Interpretation. Eine Patientin äußerte jedoch, dass die Beschriftungen Frustration auslösen könnten, wenn „bessere“ Werte als unerreichbar erscheinen.

ii. Farbschema

Die Teilnehmerinnen waren gespalten zwischen einem Ampel-Farbschema (Rot-Gelb-Grün) und einem blauen Farbverlauf. Das Ampel-Farbschema war zwar einfach verständlich durch kulturelle Vertrautheit damit und wurde teils als kontrastreicher und damit besser sichtbar wahrgenommen, die Farbe Rot löste jedoch starke negative Assoziationen aus (Blut, Schmerz). Zudem wurden Bedenken hinsichtlich Farbenblindheit geäußert. Das blaue Farbschema wirkte hingegen neutraler und beruhigender. Die Präferenz hing teils vom Datenkontext ab (z. B. emotionale Funktionsfähigkeit = Präferenz für wärmere Farben).

iii. Hintergrundlinien

Die Hintergrundlinien wurden von den meisten Befragten zunächst übersehen, später aber als hilfreich empfunden, um sich im Diagramm zu orientieren und Datenpunkte zu vergleichen. Fachpersonen wünschten sich, dass die Linien dezent aber erkennbar sind, um die Interpretation zu unterstützen, ohne vom Wesentlichen abzulenken.

iv. Alarmdarstellung

Alarme sollten auffällig und sofort erkennbar dargestellt werden. Als klarste Lösung ergab sich die direkte Markierung des Datenpunkts, ohne dessen Form zu verändern. Symbolische Darstellungen (z. B. Tonsymbole) führten zu gemischten Reaktionen: manche Teilnehmerinnen sahen darin nützliche Assoziationen (z. B. stechender Schmerz), die jedoch symptomabhängig und nicht generalisierbar waren, andere befürchteten Fehlinterpretationen. Bei mehreren Alarmen in einer Grafik wurde eine ausgewogene und beruhigende Darstellung bevorzugt, z. B. durch Kreise um die betroffenen Punktwerte.

v. Referenzdaten

Die als hellblaue Fläche dargestellten Referenzdaten wurden als Toleranzbereich verstanden. Eine zusätzlich angezeigte Linie für den Durchschnittswert sorgte für Missverständnisse.

Insgesamt lösten Referenzdaten sowohl positive (z. B. Motivation) als auch negative (z. B. Neid, Druck) Wirkung aus – abhängig vom emotionalen Zustand.

Folgende konkrete Wünsche äußerten die Teilnehmerinnen bzgl. der Darstellung von Referenzwerten:

- Klare Erklärung der Referenzgruppe
- Möglichkeit zur optionalen Anzeige, da emotionale Reaktionen individuell variieren.

vi. Fehlende Daten

Fehlende Werte wurden zu unauffällig dargestellt, was zu Missverständnissen führen konnte (z. B. Interpretation als Druckfehler). Patientinnen bevorzugten Konsistenz in der Darstellung (z. B. in Form von Punkten wie bei vorhandenen Daten). Fachpersonen waren kritischer, da dadurch unklar werde, ob der Wert fehlt oder außerhalb des Messbereichs liegt.

vii. Format und Zusatzinformationen

Patientinnen waren sich uneinig, ob eine digitale (nützlich für interaktive und personalisierte Ansicht, z. B. Zoomen) oder papierbasierte (nützlich für Mitnahme zu Arztbesuchen) Darstellung besser sei. Fachpersonen wünschten sich ein digitales Tool mit automatischer Auswertung und Darstellung von Gesundheitsverläufen. Patientinnen wünschten sich als zusätzliche Informationen eine kurze medizinische Einschätzung zu den Daten, die Abbildung weiterer medizinischer Informationen wie Vitalwerte und das Hervorheben von positiven Entwicklungen.

viii. Nutzung der Visualisierungen

Patientinnen äußerten, dass visualisierte PRO-Daten

- die Behandlung verbessern;
- die Kommunikation erleichtern;
- emotionale Sicherheit geben („Sicherheitsnetz“);
- besonders introvertierten Patientinnen helfen, Anliegen zu äußern;
- die aktive Beteiligung an Therapieentscheidungen fördern;
- eine Langzeitübersicht bieten, z. B. in Bezug auf Therapien oder Ereignisse.

Fachpersonen betrachteten Visualisierungen als wertvolle Unterstützung im Arzt-Patienten-Gespräch, besonders in der Langzeitbetreuung chronisch kranker Menschen.

Eine Patientin wollte jedoch in einer emotional belastenden Krankheitsphase bewusst auf Visualisierungen verzichten – sie betonte den Wunsch, selbst entscheiden zu können, wann und ob sie auf die Informationen zugreift.

Quantitative Ergebnisse: Befragung zum Studienende

Insgesamt beantworteten 276 Teilnehmer den Fragebogen, darunter 152 aus der Interventionsgruppe und 124 aus der Kontrollgruppe. Die Ergebnisse von Interventions- und Kontrollgruppe wichen lediglich in einem Befragungssitem (Wunsch nach zusätzlichen klinischen Informationen) signifikant voneinander, weshalb die Antworten beider Gruppen in den folgenden Abbildungen aggregiert dargestellt sind.

Die Teilnehmerinnen bewerteten die Visualisierungen hinsichtlich ihrer Verständlichkeit und der Orientierung der dargestellten Werte (starke Ausprägung der Werte stets am oberen Ende der Y-Achse)⁶ sehr positiv. Negative Bewertungen waren bei allen erfragten Aspekten in Bezug auf die Verständlichkeit gering (<10% Zustimmung). Diese Ergebnisse zeigen, dass die PRO-

⁶ Dies ist relevant, da im EORTC QLQ-C30 [18] hohe Werte bei Funktionsskalen gute Funktionalität bedeuten, hohe Werte bei Symptomskalen jedoch eine hohe Symptomlast darstellen.

Visualisierungen aus Sicht der Teilnehmerinnen ein hohes Maß an Klarheit und Benutzerfreundlichkeit aufweisen.

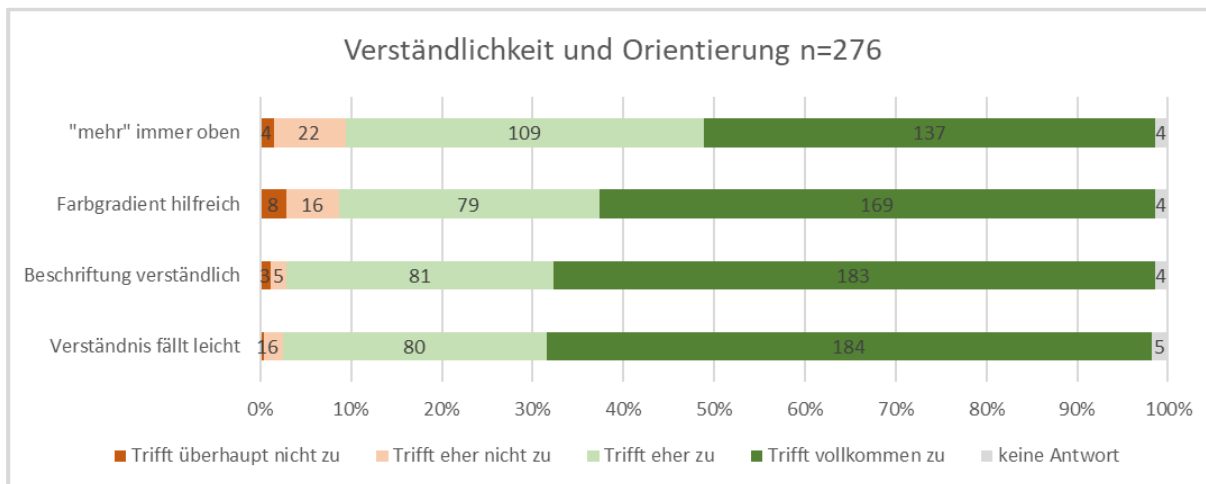


Abbildung 22: Verständlichkeit und Orientierung der Abbildungen (absolute Werte und Prozentangaben)

Die Mehrheit der Patientinnen empfand die PRO-Visualisierung als nützlich in Bezug auf verschiedene Aspekte der Reflexion des Gesundheitszustands und der Kommunikation mit dem Behandlungsteam. Die PRO-Berichte wurden somit weitgehend als nützliche Instrumente wahrgenommen im Hinblick auf die Verbesserung der Selbstwahrnehmung und der Fähigkeit, den eigenen Gesundheitszustand mitzuteilen.

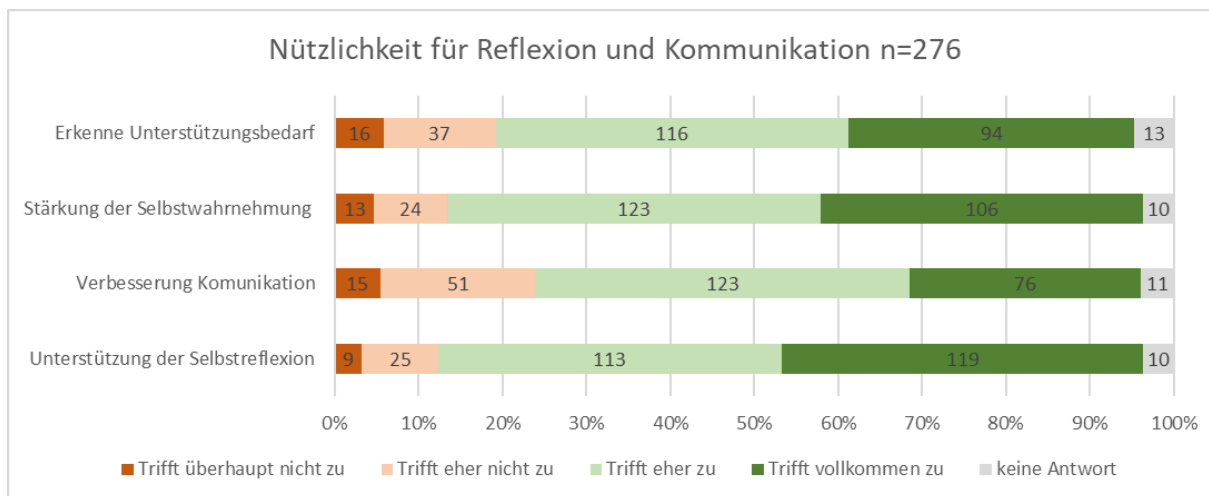


Abbildung 23: Nützlichkeit für Reflektion und Kommunikation (absolute Werte und Prozentangaben).

Die Ergebnisse zeigen ein gemischtes Bild beim Bedarf an zusätzlichen Erläuterungen. Erläuterungen zu Funktionsskalen, welche abstraktere Konzepte abbilden als Symptomskalen, wurden von den Patientinnen unterschiedlich bewertet: Rund 56% sahen keine Notwendigkeit dafür, während sich ca. 40% weitere Informationen wünschten. Mehr Anleitung insgesamt innerhalb der Abbildungen wurde mehrheitlich abgelehnt (~12% Zustimmung). Insgesamt zeigt sich ein begrenzter Bedarf an weiterführender Unterstützung, mit etwas mehr Offenheit für gezielte Erläuterungen bei komplexeren Inhalten wie Funktionsskalen.

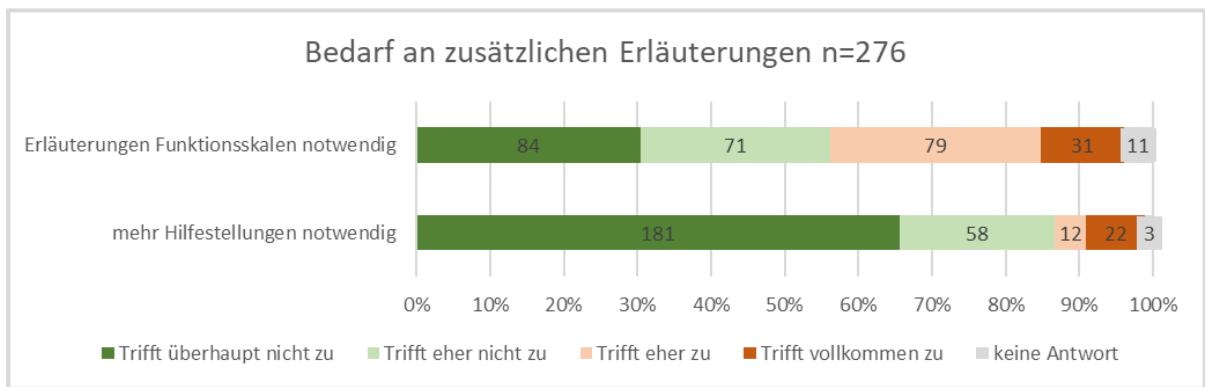


Abbildung 24: Bedarf an zusätzlichen Erläuterungen in den Abbildungen (absolute Werte und Prozentangaben).

Die Ergebnisse lassen keine einheitliche Haltung der teilnehmenden Patientinnen gegenüber der Darstellung eines klinischen und vergleichenden Kontexts, d.h. der Bereitstellung von Referenzdaten, in PRO-Visualisierungen erkennen: Im Hinblick auf Referenzdaten zu ähnlichen Patientinnen waren Zustimmung (51%) und Ablehnung (48%) nahezu gleich verteilt. Referenzdaten zur Allgemeinbevölkerung lehnten 61% der Teilnehmerinnen ab, während 38% zustimmten. Zusätzliche klinische Informationen wie das Operationsdatum oder den Therapiebeginn wünschten sich 50% aller Patientinnen, 49% erachteten diese als nicht nötig.

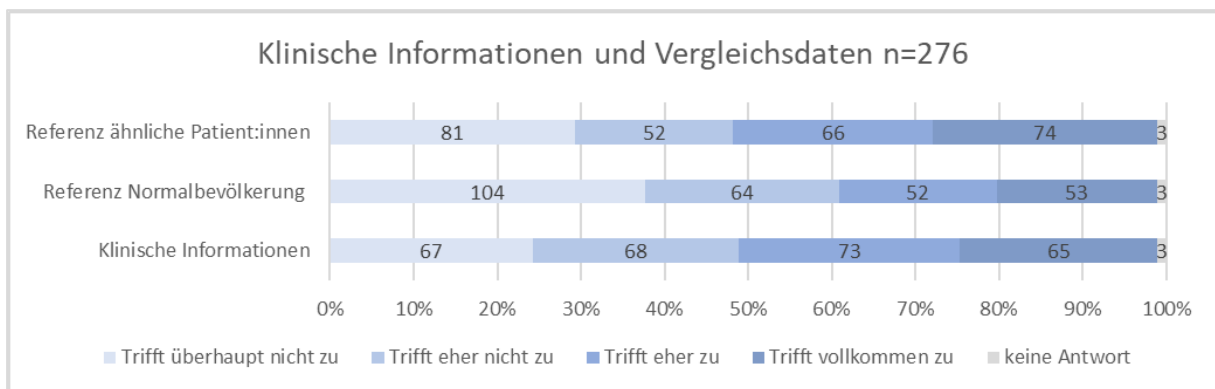


Abbildung 25: Bedarf an zusätzlichen klinischen Informationen und Referenzdaten (absolute Werte und Prozentangaben).

Der Bedarf an zusätzlichem klinischen Kontext innerhalb der PRO-Visualisierungen variierte jedoch signifikant zwischen der Intervention- und der Kontrollgruppe ($p < 0,001$): Teilnehmerinnen der Kontrollgruppe äußerten einen größeren Wunsch danach (vollständige Zustimmung ca. 33%), während Patientinnen der Interventionsgruppe keine starke Präferenz für diese ergänzenden Informationen zeigte (vollständige Zustimmung ca. 16%). Möglicherweise waren Patientinnen der Interventionsgruppe, welche sich über den PRO B Studienzeitraum durch die wöchentlichen Befragungen regelmäßig mit ihrem Gesundheitszustand auseinandersetzten, aufgrund der PRO B Intervention geschulter in der Reflexion ihres Krankheitsverlaufs und dazugehöriger „Meilensteine“.

Die Daten zeigen insgesamt ein geteiltes Meinungsbild zur Bereitstellung zusätzlicher klinischer Informationen, das auf unterschiedliche Informationsbedürfnisse aus Patientinnenperspektive hinweist, welche potenziell auch abhängig von der individuellen Erfahrung mit PRO-Monitoring sein können.

4 Diskussion der Projektergebnisse

4.1 Effektevaluation

4.1.1 Fatigue-score

Unsere Ergebnisse zeigen eine signifikante und anhaltende Reduktion der Fatigue-Werte in der Interventionsgruppe über den gesamten 12-monatigen Studienzeitraum. Die mittlere Differenz im Fatigue-T-Score zwischen den Gruppen lag zu allen Messzeitpunkten zwischen -5,2 und -6,8 zugunsten der Interventionsgruppe. Dies unterstreicht die Wirksamkeit des wöchentlichen ePRO-Monitorings mit Alarmfunktion als Strategie zur Reduktion von Fatigue bei Patientinnen mit metastasiertem Brustkrebs.

Diese Ergebnisse stehen im Einklang mit früheren Studien und Metaanalysen, die die Effektivität ePRO-basierter Symptommanagement-Interventionen in der Onkologie belegen [35-37]. Insbesondere konnten solche Interventionen eine signifikante Reduktion der Symptomlast – einschließlich Fatigue – nachweisen [38, 39]. Die kontinuierliche Überwachung in Kombination mit einer Alarmfunktion ermöglicht eine frühzeitige Erkennung und Intervention bei sich verschlechternden Symptomen und fördert so eine proaktive und personalisierte Versorgung. Dies ist besonders relevant im Hinblick auf Fatigue – ein häufiges und belastendes Symptom bei Krebspatient:innen, das oft nicht ausreichend erkannt und behandelt wird [40]. Die Möglichkeit, rasch auf erhöhte Fatigue-Werte zu reagieren, kann die Belastung der Patient:innen erheblich verringern und ihre Lebensqualität verbessern.

4.1.2 Körperliche Funktion und Gesundheitsbezogene Lebensqualität (HRQoL)

Die Studie zeigte zudem, dass die körperliche Funktionsfähigkeit in der Interventionsgruppe in den ersten neun Monaten signifikant besser war als in der Kontrollgruppe. Diese frühen Verbesserungen lassen sich vermutlich auf eine bessere Symptomkontrolle – insbesondere eine Reduktion der Fatigue – zurückführen, die es den Patientinnen ermöglichte, aktiver zu bleiben.

Frühere Studien haben wiederholt gezeigt, dass der Einsatz von elektronischer Symptomerfassung auf Basis von PROs nicht nur eine frühzeitige Erkennung und zeitnahe Behandlung krebsbedingter Symptome ermöglicht, sondern auch wesentlich zur Aufrechterhaltung oder Verbesserung der körperlichen Funktionsfähigkeit von onkologischen Patient:innen beiträgt [30, 39, 41-43]. Dies gilt insbesondere für Patientinnen mit fortgeschrittenen Erkrankungen wie metastasiertem Brustkrebs, bei denen ein proaktives Symptom-Monitoring funktionellen Einschränkungen entgegenwirken und langfristig bessere klinische Ergebnisse fördern kann.

Der Verlust der Signifikanz nach 12 Monaten könnte verschiedene Ursachen haben. Patientinnen beider Gruppen könnten ein Plateau in ihrer körperlichen Funktion erreicht haben, oder krankheits- bzw. therapiebedingte Faktoren könnten im weiteren Verlauf die langfristigen Effekte der Intervention überlagert haben. Ebenso ist denkbar, dass die Intensität der Intervention oder die Adhärenz im Zeitverlauf abnahm, wodurch die Aufrechterhaltung der anfänglichen Effekte erschwert wurde. Zukünftige Studien sollten untersuchen, ob eine intensivere oder adaptierte Intervention über einen längeren Zeitraum die positiven Effekte auf die körperliche Funktion erhalten kann.

Bezüglich der HRQoL zeigte die Interventionsgruppe eine leicht bessere Lebensqualität, wobei der Unterschied nur zum sechsmonatigen Messzeitpunkt signifikant war (mittlere Differenz: 0,8). Während einige Metaanalysen einen kleinen positiven Effekt von ePRO-Interventionen auf die HRQoL bei Krebspatient:innen nachweisen [38, 39, 43], sind die Ergebnisse in der Literatur insgesamt uneinheitlich [44]. Die HRQoL ist ein komplexes Konstrukt, das von vielen

Faktoren beeinflusst wird – darunter physische Symptome, psychisches Wohlbefinden, soziale Unterstützung und krankheitsbezogene Belastungen.

Es ist möglich, dass das wöchentliche ePRO-Monitoring zwar gezielt einzelne Symptome wie Fatigue lindern kann, jedoch nicht ausreicht, um eine umfassende und nachhaltige Verbesserung der HRQoL zu erreichen – insbesondere bei Patient:innen mit fortgeschrittener Erkrankung wie metastasiertem Brustkrebs.

Die geringe Effektstärke und das Ausbleiben langfristiger signifikanter Unterschiede deuten darauf hin, dass für eine robuste Verbesserung der HRQoL multimodale Ansätze erforderlich sind, die über das reine Symptommonitoring hinausgehen. Dazu könnten psychologische Unterstützung, Ernährungsberatung oder andere individualisierte Programme gehören [36, 40].

4.1.3 Überleben und Zeitpunkt des ersten Systemtherapie-Wechsels

Die Analyse des Gesamtüberlebens zeigte in unserer Studie einen signifikanten Vorteil für die Interventionsgruppe in der Gesamtkohorte. Nach 12 Monaten lag die Überlebensrate in der Interventionsgruppe bei 87,6%, verglichen mit 84,7% in der Kontrollgruppe. Die multivariate Analyse ergab ein adjustiertes HR von 0,71 (95 %-KI: 0,51–0,99), was auf eine 29% Reduktion des Sterberisikos in der Interventionsgruppe hinweist.

Diese Ergebnisse sind klinisch hochrelevant, da sie nahelegen, dass ein engmaschiges Monitoring mittels wöchentlicher ePRO-Erhebungen mit Alarmfunktion nicht nur die Symptomlast und Lebensqualität verbessert, sondern auch das Überleben von Patientinnen mit metastasiertem Brustkrebs verlängern kann. Diese Beobachtungen stehen im Einklang mit früheren Studien und Metaanalysen, die eine Assoziation zwischen ePRO-basiertem Symptommanagement und verbessertem Gesamtüberleben bei Krebspatient:innen nachweisen konnten [30, 35, 37, 42, 45]. Die zugrunde liegenden Mechanismen dieses Überlebensvorteils dürften vielfältig sein: Eine bessere Symptomkontrolle kann die Therapietreue verbessern, ungeplante Notaufnahmen und Krankenhausaufenthalte reduzieren und dazu beitragen, dass Patient:innen länger in einem besseren Allgemeinzustand bleiben – was wiederum eine effektivere und kontinuierlichere Tumorthherapie ermöglicht. Die Alarmfunktion der ePRO-Systeme erlaubt eine frühzeitige Erkennung und Behandlung von Nebenwirkungen oder Komplikationen, wodurch kritische Verschlechterungen verhindert werden können, die ansonsten zu Therapieabbrüchen oder Dosisreduktionen führen würden.

In Subgruppenanalysen ergaben sich jedoch keine signifikanten Unterschiede im Überleben bei Patientinnen mit viszeralen Metastasen (HR 0,43, $p = 0,13$), triple-negativem Brustkrebs (HR 0,74, $p = 0,51$) und HR+/HER2--Subtyp (HR 0,70, $p = 0,06$). Diese Ergebnisse könnten auf die kleinere Stichprobengröße in den jeweiligen Subgruppen zurückzuführen sein, wodurch die statistische Aussagekraft eingeschränkt ist. Es ist auch denkbar, dass bei aggressiveren Subtypen oder komplexeren biologischen Konstellationen der Einfluss der ePRO-Intervention durch die Krankheitsdynamik oder intensive Therapiestrategien überlagert wird. Dennoch zeigt sich in allen Subgruppen ein Trend zugunsten der Intervention, was weiterführende Untersuchungen mit größeren, differenzierten Kohorten rechtfertigt.

Im Gegensatz dazu zeigte die Analyse der Zeit bis zur ersten Änderung der systemischen Therapie in unserer Studie keinen signifikanten Unterschied zwischen der Interventions- und der Kontrollgruppe. Dies deutet darauf hin, dass das wöchentliche ePRO-Monitoring mit Alarmfunktion zwar die Symptomkontrolle verbessert und möglicherweise das Überleben verlängert, jedoch nicht zu einer Verzögerung systemischer Therapieanpassungen führt. Diese Interpretation wird durch die Ergebnisse der CANKADO PRO-React-Studie[46] gestützt, in der gezeigt wurde, dass die Behandlungscharakteristika – einschließlich relativer Dosisintensität, Verschiebungen, Unterbrechungen und Dosisreduktionen – zwischen den beiden

Studienarmen vergleichbar waren. Dies legt nahe, dass das ePRO-Monitoring keinen wesentlichen Einfluss auf Intensität oder Kontinuität der systemischen Therapie hatte.

Dies könnte widerspiegeln, dass Therapieanpassungen bei metastasiertem Brustkrebs primär durch objektive Krankheitsprogression oder durch das Auftreten schwerwiegender therapielimitierender Nebenwirkungen bestimmt werden – Faktoren, die durch ein verbessertes Symptommanagement allein nicht grundsätzlich verändert werden. Darüber hinaus ist es denkbar, dass die Intervention die Lebensqualität während der laufenden Therapie so weit verbessert, dass Patient:innen und behandelnde Ärzt:innen an einer bestehenden Therapie länger festhalten, selbst wenn kleinere Anpassungen erforderlich sind, die nicht als vollständiger Wechsel des Therapieschemas gewertet werden.

Die in der Studie dokumentierten Gründe für Therapieänderungen – Progression, Nebenwirkungen oder Patientenwunsch – waren in beiden Gruppen ähnlich verteilt. Dies stützt die Annahme, dass das ePRO-Monitoring zwar die Behandlungsverträglichkeit verbessert, aber keine ungerechtfertigten Verzögerungen bei notwendigen Therapieänderungen verursacht. Vielmehr scheint es die Therapiedurchführung zu stabilisieren und damit möglicherweise indirekt zum beobachteten Überlebensvorteil beizutragen – ohne jedoch die Zeit bis zur nächsten Therapie künstlich zu

Die Untersuchung des Einflusses einer Änderung der systemischen Therapie innerhalb der ersten 12 Monate auf das Gesamtüberleben legt nahe, dass eine Therapieänderung nicht zwangsläufig mit einem schlechteren Gesamtüberleben einhergeht, was zunächst kontraintuitiv erscheinen mag, da Therapieanpassungen häufig durch Krankheitsprogression oder Toxizitäten ausgelöst werden. Eine plausible Erklärung hierfür ist jedoch, dass rechtzeitige und individualisierte Therapieanpassungen – insbesondere im Rahmen eines strukturierten Symptommonitorings – tatsächlich zu besseren Behandlungsergebnissen beitragen können.

Das leicht bessere Überleben bei Patient:innen mit Therapieänderung in der Interventionsgruppe könnte den Vorteil einer frühzeitigen Erkennung von Therapieversagen oder Nebenwirkungen durch ePRO-basiertes Monitoring widerspiegeln. Dies ermöglicht es den behandelnden Ärzt:innen, vor einer kritischen Verschlechterung auf effektivere oder besser verträgliche Therapieoptionen umzusteigen. Diese Beobachtung steht im Einklang mit Studien wie Basch et al. (2017), die gezeigt haben, dass ein ePRO-basiertes Symptommonitoring während der Krebstherapie das Überleben durch frühere Interventionen verbessern kann [42, 45].

Wichtig ist zudem, dass das Fehlen eines negativen prognostischen Signals bezüglich Therapieänderungen in beiden Gruppen unterstreicht, dass die Sequenzierung der Therapien beim metastasierten Brustkrebs vergleichbare Ergebnisse erzielen kann, vorausgesetzt, das Timing und die Auswahl der nachfolgenden Therapien sind angemessen [47, 48].

Darüber hinaus deuten die Ergebnisse darauf hin, dass die alleinige Betrachtung der Therapiekontinuität für das Verständnis der Prognose nicht ausreicht. Vielmehr spielen der klinische Kontext und der individuelle Verlauf der Patient:innen – einschließlich eines proaktiven Symptommanagements – eine entscheidende Rolle. Dies unterstützt einen personalisierten und reaktionsschnellen Behandlungsansatz, der durch ePRO-Systeme erleichtert wird.

4.2 Gesundheitsökonomische Evaluation

Während des 12-monatigen Studienzeitraums wurden in der PRO B-Interventionsgruppe höhere QALYs und gleichzeitig geringere Gesamtkosten im Gesundheitswesen im Vergleich zur Kontrollgruppe beobachtet. Diese Ergebnisse stehen im Einklang mit der zunehmenden Evidenz, die auf eine gute Kostenwirksamkeit digitaler Gesundheitslösungen in der Onkologie

hinweist [49, 50]. Die Fähigkeit solcher Systeme, sowohl die Lebensqualität der Patientinnen zu verbessern als auch die Gesundheitskosten zu senken oder zumindest einzudämmen, macht ePRO-basierte Monitoringsysteme zu einer attraktiven Option für eine moderne, wertbasierte Krebsversorgung.

Ein wesentlicher Faktor für die beobachteten Kosteneinsparungen war in unserer Studie die Reduktion von Krankenhausaufenthalten in der Interventionsgruppe. Dies deckt sich mit früheren Studien, die gezeigt haben, dass ePRO-Monitoring eine frühere Erkennung und ein proaktives Management von Symptomen ermöglicht – und dadurch Komplikationen vermeidet, die andernfalls zu teuren Notaufnahmen oder stationären Behandlungen führen würden [42, 43, 45]. So konnten Basch et al. neben einem verbesserten Überleben auch eine signifikante Reduktion von Notaufnahmen und Hospitalisierungen durch ePRO-Überwachung während der Chemotherapie nachweisen [45]. Ebenso zeigte die randomisierte Studie von Denis et al., dass digitales Monitoring bei Lungenkrebspatient:innen die Zahl der Krankenhausaufenthalte reduzierte und gleichzeitig kosteneffektiv war [42].

Bemerkenswert ist, dass der wirtschaftliche Vorteil der PRO B-Intervention nicht zulasten der klinischen Wirksamkeit erzielt wurde – im Gegenteil: Die Intervention führte zu einer verbesserten Symptombelastung (z. B. bei Fatigue), besserer körperlicher Funktion und sogar zu einem Überlebensvorteil. Diese Kombination aus klinischem und ökonomischem Nutzen unterstreicht den Mehrwert der Integration von ePRO-Instrumenten in die reguläre Versorgung – insbesondere bei Patientinnen mit metastasiertem Brustkrebs, bei denen eine langfristige, ressourcenschonende Betreuung erforderlich ist.

4.3 Limitationen der Studie

Ein wichtiger Aspekt der vorliegenden Studie besteht darin, dass nur Patientinnen eingeschlossen werden konnten, die über ein Smartphone verfügten. Diese Einschränkung war notwendig, da das ePRO-System ausschließlich digital über eine App zugänglich war. Auch wenn dies potenziell zu einer Selektionsverzerrung geführt haben könnte, ist zu erwarten, dass dieser Punkt in Zukunft zunehmend an Relevanz verliert. Der Smartphone-Besitz steigt stetig, auch unter älteren Patientinnen, die traditionell weniger technikaffin sind. In aktuellen Erhebungen zeigt sich, dass digitale Endgeräte mittlerweile auch in dieser Bevölkerungsgruppe weit verbreitet sind, sodass langfristig eine breitere Einsetzbarkeit solcher Interventionen gewährleistet sein dürfte.

Eine weitere Limitation bestand darin, dass ausschließlich deutschsprachige Patientinnen in die Studie aufgenommen werden konnten. Diese sprachliche Einschränkung war primär durch finanzielle und logistische Rahmenbedingungen bedingt, insbesondere im Hinblick auf Übersetzungs- und Anpassungskosten der App und der Studiendokumente. Zwar ist anzunehmen, dass die Sprache selbst keinen Einfluss auf die Wirksamkeit der Intervention hat, dennoch schränkt dies die Generalisierbarkeit der Ergebnisse auf nicht-deutschsprachige Bevölkerungsgruppen ein. In zukünftigen Studien sollte daher eine sprachliche und kulturelle Adaptation des ePRO-Systems erfolgen, um eine größere Diversität der Studienteilnehmenden zu ermöglichen.

Ein bedeutsamer methodischer Aspekt betrifft die eingeschränkte Aussagekraft hinsichtlich der Gründe für Therapieänderungen im Studienverlauf. Im Rahmen der PRO B-Studie wurden keine spezifischen Daten darüber erfasst, ob ein Therapiewechsel primär durch eine Krankheitsprogression, eine Verschlechterung der Symptomatik oder aufgrund eines durch das ePRO-System ausgelösten Alarms erfolgte. Diese fehlende Differenzierung limitiert die Interpretation der Kausalzusammenhänge zwischen der Intervention und Änderungen im Behandlungsverlauf. Künftige Studien sollten deshalb strukturierte Dokumentationen der Auslöser für Therapieanpassungen integrieren, um die Wirkung digitaler Interventionen noch besser verstehen zu können.

Darüber hinaus ist zu beachten, dass die Zahl der gescreenten Patientinnen wahrscheinlich unterschätzt wurde. In mehreren teilnehmenden Studienzentren erfolgte keine vollständige Erfassung in den Screening-Logs, was die Analyse der Rekrutierungsbasis erschwerte. Dennoch lieferten die vorhandenen Daten wertvolle Erkenntnisse darüber, warum Patientinnen die Studienteilnahme ablehnten – unter anderem wegen technischer Bedenken oder einer bereits bestehenden Belastung durch ihre Erkrankung. Eine detailliertere Erfassung solcher Ablehnungsgründe wäre jedoch für zukünftige Implementierungsstrategien hilfreich.

Auch die Subgruppenanalysen, insbesondere bei Patientinnen mit viszeralen Metastasen oder dem triple-negativen Subtyp, sind nur eingeschränkt aussagekräftig. Die statistische Power war in diesen Gruppen limitiert, was das Risiko für falsch-negative Ergebnisse erhöht. Zudem wurden die Analysen explorativ und ohne Korrektur für multiples Testen durchgeführt, was zu einem erhöhten Risiko für falsch-positive Befunde führen kann. Entsprechende Ergebnisse sollten daher mit Vorsicht interpretiert und in größeren, confirmatorischen Studien überprüft werden.

Ein weiterer methodischer Schwachpunkt der Studie besteht darin, dass keine Verblindung der Teilnehmenden oder Behandelnden möglich war. Da alle erhobenen Endpunkte PROs waren, kann nicht ausgeschlossen werden, dass die Kenntnis der Gruppenzugehörigkeit das Antwortverhalten beeinflusst hat (z. B. durch Erwartungseffekte). Solche Bias-Einflüsse sind insbesondere bei subjektiven Endpunkten relevant und sollten bei der Interpretation berücksichtigt werden.

Abschließend ist anzumerken, dass die Studiendesign-bedingte Differenz in der Nachbeobachtungszeit zwischen der Interventions- und Kontrollgruppe eine potenzielle Verzerrung in Bezug auf das Drop-out-Risiko darstellen könnte. In der Kontrollgruppe zeigte sich eine kürzere Nachbeobachtungszeit bei Patientinnen, die aus der Studie ausschieden oder verloren gingen. Diese systematische Differenz kann Auswirkungen auf die Interpretation langfristiger Endpunkte haben.

5 Schlussfolgerungen und Empfehlungen des Evaluators

Die PRO B Studie untersuchte, ob eine neue Versorgungsform zu einer geringeren Belastung durch Fatigue bei Patientinnen mit metastasiertem Brustkrebs führt. Die Ergebnisse der Studie zeigen, dass die neue Versorgungsform im Vergleich zur Regelversorgung zu einem signifikant geringeren Level der Fatigue führte. Nach sechs Monaten wies die Interventionsgruppe im Median einen niedrigeren Fatigue-T-Score auf als die Kontrollgruppe (Median Interventionsgruppe: 53,5; IQR 46,4–59,7; Mittelwert 53,7; SD 10,7; Median Kontrollgruppe: 59,3; IQR 53,7–67,7; Mittelwert 60,7; SD 10,2). Die mittlere Differenz der Fatigue-T-Scores zwischen den Gruppen betrug -5,4 Punkte (95%-KI: -6,6 bis -4,1, $p < 0,001$), was einen statistisch signifikanten Unterschied zugunsten der Interventionsgruppe darstellt. Da die Differenz unter dem MCID-Wert von -3,3 liegt, ist dieser Unterschied auch klinisch relevant.

Die neue Versorgungsform besteht aus mehreren Teilen. Die Patientinnen in der Interventionsgruppe erhielten wöchentlich eine PRO-Befragung über eine Smartphone-Applikation. Bei Verschlechterung der PRO-Werte wurde ein Alarm generiert, der durch die Studienzentren innerhalb von 48 Stunden quittiert und dokumentiert werden musste. Auf den Alarm folgte eine Kontaktaufnahme des jeweiligen Zentrums mit der Patientin, um die Verschlechterung ihrer Werte zu spezifizieren und ggf. zu intervenieren. Die Art der Maßnahmen blieb dem behandelnden Arzt bzw. der behandelnden Ärztin überlassen.

Die Ergebnisse der Studie zeigen, dass die Interventionsgruppe im Vergleich zur Kontrollgruppe eine deutlich bessere körperliche Funktionsfähigkeit über die ersten neun Monate aufwies (mittlere Unterschiede zwischen 3,7 und 5,1 Punkten). Nach 12 Monaten war dieser Unterschied zwischen den Gruppen jedoch geringer (1,8 Punkte, 95%-KI: -2,5 bis 6,1).

Die gesundheitsbezogene Lebensqualität war geringfügig höher in der Interventionsgruppe im Vergleich zur Kontrollgruppe im gesamten Studienzeitraum nach Beginn der Intervention (<1 Punkt mittlerer Unterschied). Das 12-Monats-Gesamtüberleben war in der Interventionsgruppe mit 87,6% etwas höher als in der Kontrollgruppe mit 84,7% (adjustierte Hazard Ratio: 0,71, 95%-KI: 0,51–0,99). Überlebensanalysen sind allerdings über solche kurzen Zeiträume nur bedingt aussagekräftig. Die häufigsten Alarmauslöser waren Lebensqualität/Gesundheitsstatus (20,0%), körperliche Funktion (15,8%) und Schmerz (13,5%). Bei den Gesprächen mit den behandelnden Ärztinnen und Ärzten nach einem Alarm standen zusätzliche Informationen zur Erkrankung, zur Therapie, und zu Nebenwirkungen der Therapien im Vordergrund. Es ging aber auch um Terminvereinbarungen und Kontakte zu zusätzlichen Anlaufstellen.

Die Auswertung der Interviews mit dem Behandlungsteam ergab, dass die PRO B Studie als Innovation in der Routinearbeit empfunden wurde. Die Anbieter sahen das Potenzial der Innovation bei der Unterstützung von Patientinnen und die Relevanz der Forschung zu PROs. Allerdings äußerten sie auch Einschränkungen, z. B. dass PRO-Fragebögen nur den Zustand der Patientinnen zu einem bestimmten Zeitpunkt widerspiegeln und dass ein Gespräch mit einem Arzt oder einer Ärztin ein ganzheitlicheres Bild ergeben kann. Die Patientinnen der Interventionsgruppe berichteten, dass sie durch die Nutzung der neuen Versorgungsform besser in der Lage waren, ihren Gesundheitszustand gegenüber dem Behandlungsteam zu kommunizieren. Die Mehrheit der Patientinnen befürwortete die Einführung der neuen Intervention als Teil der Regelversorgung bei Patientinnen mit einer fortgeschrittenen Brustkrebserkrankung.

Im Rahmen der gesundheitsökonomischen Evaluation wurden die Gesamtkosten der Behandlung über die Krankenkassendaten für alle Studienpatientinnen gesammelt und ausgewertet. Die Analyse vergleicht die EQ-5D-5L-Indexwerte zu verschiedenen Zeitpunkten sowie die Qualität-adjustierten Lebensjahre (QALY) zwischen der Interventions- und der Kontrollgruppe. Zu Beginn der Studie waren die EQ-5D-5L-Indexwerte in beiden Gruppen vergleichbar. Nach sechs und 12 Monaten zeigte sich jedoch ein deutlicher Unterschied zugunsten der Interventionsgruppe. Die Interventionsgruppe erreichte einen durchschnittlichen QALY-Wert von 0,745, während die Kontrollgruppe bei 0,662 lag. Die inkrementellen jährlichen Kosten der Intervention pro Patientin betragen -846€ (modellbasiert), die Intervention war also etwas günstiger als die Standardbehandlung und demzufolge kosteneffektiv. Die Kosten-Nutzen-Analyse ergab, dass die Intervention ökonomisch vorteilhaft war, da die Intervention kostengünstiger als die Standardbehandlung war und zu einem Lebensqualitätsvorteil in der Interventionsgruppe geführt hat.

IV Literaturverzeichnis

1. Mariotto, A.B., et al., *Estimation of the Number of Women Living with Metastatic Breast Cancer in the United States*. *Cancer Epidemiol Biomarkers Prev*, 2017. **26**(6): p. 809-815.
2. *Krebs in Deutschland für 2015/2016*, Ausgabe, Editor. 2019, Robert Koch-Institut (Hrsg) und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (Hrsg). Berlin.
3. Caswell-Jin, J.L., et al., *Change in Survival in Metastatic Breast Cancer with Treatment Advances: Meta-Analysis and Systematic Review*. *JNCI Cancer Spectr*, 2018. **2**(4): p. pky062.
4. Health, U.S.D.o., et al., *Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance*. *Health Qual Life Outcomes*, 2006. **4**: p. 79.

5. Basch, E., et al., *Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study*. *Lancet Oncol*, 2006. **7**(11): p. 903-9.
6. Di Maio, M., et al., *Patient-reported outcomes in the evaluation of toxicity of anticancer treatments*. *Nat Rev Clin Oncol*, 2016. **13**(5): p. 319-25.
7. Di Maio, M., et al., *Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials*. *J Clin Oncol*, 2015. **33**(8): p. 910-5.
8. Groenvold, M., et al., *Psychological distress and fatigue predicted recurrence and survival in primary breast cancer patients*. *Breast Cancer Res Treat*, 2007. **105**(2): p. 209-19.
9. Muller, V., et al., *Impact of disease progression on health-related quality of life in patients with metastatic breast cancer in the PRAEGNANT breast cancer registry*. *Breast*, 2018. **37**: p. 154-160.
10. Pocock, S.J. and R. Simon, *Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial*. *Biometrics*, 1975. **31**(1): p. 103-115.
11. Cocks, K., et al., *Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30*. *Eur J Cancer*, 2012. **48**(11): p. 1713-21.
12. Cortes, J., et al., *Health-related quality of life in patients with locally advanced or metastatic breast cancer treated with eribulin mesylate or capecitabine in an open-label randomized phase 3 trial*. *Breast Cancer Res Treat*, 2015. **154**(3): p. 509-20.
13. UNESCO, *International Standard Classification of Education. ISCED 1997*. 2006: Paris
14. Stedman, M., et al., *Comorbidity technical report: the impact of different SEER-Medicare claims-based comorbidity indexes on predicting non-cancer mortality for cancer patients*. 2017, Bethesda, MD, National Cancer Institute.
15. Dawood, S., et al., *Survival differences among women with de novo stage IV and relapsed breast cancer*. *Ann Oncol*, 2010. **21**(11): p. 2169-2174.
16. Largillier, R., et al., *Prognostic factors in 1,038 women with metastatic breast cancer*. *Ann Oncol*, 2008. **19**(12): p. 2012-9.
17. Kast, K., et al., *Impact of breast cancer subtypes and patterns of metastasis on outcome*. *Breast Cancer Res Treat*, 2015. **150**(3): p. 621-9.
18. Soni, A., et al., *Breast cancer subtypes predispose the site of distant metastases*. *Am J Clin Pathol*, 2015. **143**(4): p. 471-8.
19. Lin, N.U., J.R. Bellon, and E.P. Winer, *CNS metastases in breast cancer*. *J Clin Oncol*, 2004. **22**(17): p. 3608-17.
20. Bower, J.E., et al., *Fatigue after breast cancer treatment: Biobehavioral predictors of fatigue trajectories*. *Health Psychol*, 2018. **37**(11): p. 1025-1034.
21. Bower, J.E., et al., *Fatigue in breast cancer survivors: occurrence, correlates, and impact on quality of life*. *J Clin Oncol*, 2000. **18**(4): p. 743-53.
22. Andrykowski, M.A., et al., *Prevalence, predictors, and characteristics of off-treatment fatigue in breast cancer survivors*. *Cancer*, 2010. **116**(24): p. 5740-8.
23. Bower, J.E., *Cancer-related fatigue--mechanisms, risk factors, and treatments*. *Nat Rev Clin Oncol*, 2014. **11**(10): p. 597-609.
24. Liegl, G., et al., *Establishing the European Norm for the health-related quality of life domains of the computer-adaptive test EORTC CAT Core*. *Eur J Cancer*, 2019. **107**: p. 133-141.

25. Borm, G.F., J. Fransen, and W.A. Lemmens, *A simple sample size formula for analysis of covariance in randomized clinical trials*. J Clin Epidemiol, 2007. **60**(12): p. 1234-8.
26. Grobler, A.C., et al., *Handling Missing Data and Drop Out in Hospice/Palliative Care Trials Through the Estimand Framework*. J Pain Symptom Manage, 2022. **63**(4): p. e431-e439.
27. ICH. *Integrated Addendum to ICH E6(R1); 2016. Guideline for Good Clinical Practice E6(R2)*. Available from: https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf. Accessed 29 November 2023.
28. Bell, M.L., et al., *Analytical approaches and estimands to take account of missing patient-reported data in longitudinal studies*. Patient Relat Outcome Meas, 2019. **10**: p. 129-140.
29. Ratitch, B., M. O'Kelly, and R. Tosiello, *Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models*. Pharm Stat, 2013. **12**(6): p. 337-47.
30. Basch, E., et al., *Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial*. J Clin Oncol, 2016. **34**(6): p. 557-65.
31. Matthews, J.N., et al., *Analysis of serial measurements in medical research*. BMJ, 1990. **300**(6719): p. 230-5.
32. Rubin, D.B., *Multiple Imputation for Nonresponse in Surveys*. 2004: Wiley.
33. Damschroder, L.J., et al., *The updated Consolidated Framework for Implementation Research based on user feedback*. Implement Sci, 2022. **17**(1): p. 75.
34. Kuckartz, U., *Qualitative Text Analysis: A Guide to Methods, Practice & Using Software*. 2014: SAGE.
35. Xia, Y., et al., *Effectiveness of symptom monitoring on electronic patient-reported outcomes (ePROs) among patients with lung cancer: a systematic review and meta-analysis*. NPJ Digit Med, 2025. **8**(1): p. 399.
36. Zhong, C., et al., *Digital Health Interventions to Improve Mental Health in Patients With Cancer: Umbrella Review*. J Med Internet Res, 2025. **27**: p. e69621.
37. Li, Y., J. Li, and X. Hu, *The effectiveness of symptom management interventions based on electronic patient-reported outcomes (ePROs) for symptom burden, quality of life, and overall survival among patients with cancer: A meta-analysis of randomized controlled trials*. Int J Nurs Stud, 2023. **147**: p. 104588.
38. Kikawa, Y., et al., *A randomized study comparing electronic patient-reported outcome (ePRO) monitoring with routine follow-up during trastuzumab deruxtecan treatment in patients with metastatic breast cancer (PRO-DUCE study)*. Journal of Clinical Oncology, 2024. **42**(16_suppl): p. 1504-1504.
39. Basch, E., et al., *Effect of Electronic Symptom Monitoring on Patient-Reported Outcomes Among Patients With Metastatic Cancer: A Randomized Clinical Trial*. JAMA, 2022. **327**(24): p. 2413-2422.
40. Lee, K., et al., *Digital Health Interventions for Adult Patients With Cancer Evaluated in Randomized Controlled Trials: Scoping Review*. J Med Internet Res, 2023. **25**: p. e38333.
41. Maguire, R., et al., *Real time remote symptom monitoring during chemotherapy for cancer: European multicentre randomised controlled trial (eSMART)*. BMJ, 2021. **374**: p. n1647.

42. Denis, F., et al., *Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer*. JAMA, 2019. **321**(3): p. 306-307.
43. Absolom, K., et al., *Phase III Randomized Controlled Trial of eRAPID: eHealth Intervention During Chemotherapy*. J Clin Oncol, 2021. **39**(7): p. 734-747.
44. Ciruelos, E.M., et al., *Patient-Reported Outcomes in Patients With PIK3CA-Mutated Hormone Receptor-Positive, Human Epidermal Growth Factor Receptor 2-Negative Advanced Breast Cancer From SOLAR-1*. J Clin Oncol, 2021. **39**(18): p. 2005-2015.
45. Basch, E., et al., *Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment*. JAMA, 2017. **318**(2): p. 197-198.
46. Harbeck, N., et al., *Significantly longer time to deterioration of quality of life due to CANKADO PRO-React eHealth support in HR+ HER2- metastatic breast cancer patients receiving palbociclib and endocrine therapy: primary outcome analysis of the multicenter randomized AGO-B WSG PreCycle trial*. Ann Oncol, 2023. **34**(8): p. 660-669.
47. Burstein, H.J., et al., *Endocrine Treatment and Targeted Therapy for Hormone Receptor-Positive, Human Epidermal Growth Factor Receptor 2-Negative Metastatic Breast Cancer: ASCO Guideline Update*. J Clin Oncol, 2021. **39**(35): p. 3959-3977.
48. Cardoso, F., et al., *5th ESO-ESMO international consensus guidelines for advanced breast cancer (ABC 5)*. Ann Oncol, 2020. **31**(12): p. 1623-1649.
49. Dawkins, B., et al., *Cost-Effectiveness of eRAPID eHealth Intervention for Symptom Management During Chemotherapy*. JCO Oncol Pract, 2024. **20**(4): p. 581-590.
50. Gentili, A., et al., *The cost-effectiveness of digital health interventions: A systematic review of the literature*. Front Public Health, 2022. **10**: p. 787135.

V Anlagen

Anlage 1: PRO Visualisierung Design Decisions v1.0

Design decisions

Summary of design decisions for the visualization of PRO data
Charité – Universitätsmedizin Berlin
Revision 1.0

17 October 2024
Rosanne Andriessen
Panton bv

Contents

1	Graph type	5
2	Colour.....	5
3	Grid.....	7
4	Readability	7
5	X-axis	7
6	Y-axis	10
7	Directionality.....	10
8	Reference data.....	12
9	Missing data points	13
10	Alarming data points.....	13

This document describes the design decisions made for the visualisation of PRO data. The design process is part of the PRO B research project.

Figure 1 shows the final design of both the functioning graphs (Emotionale Funktion) and the symptom graphs (Schmerz).

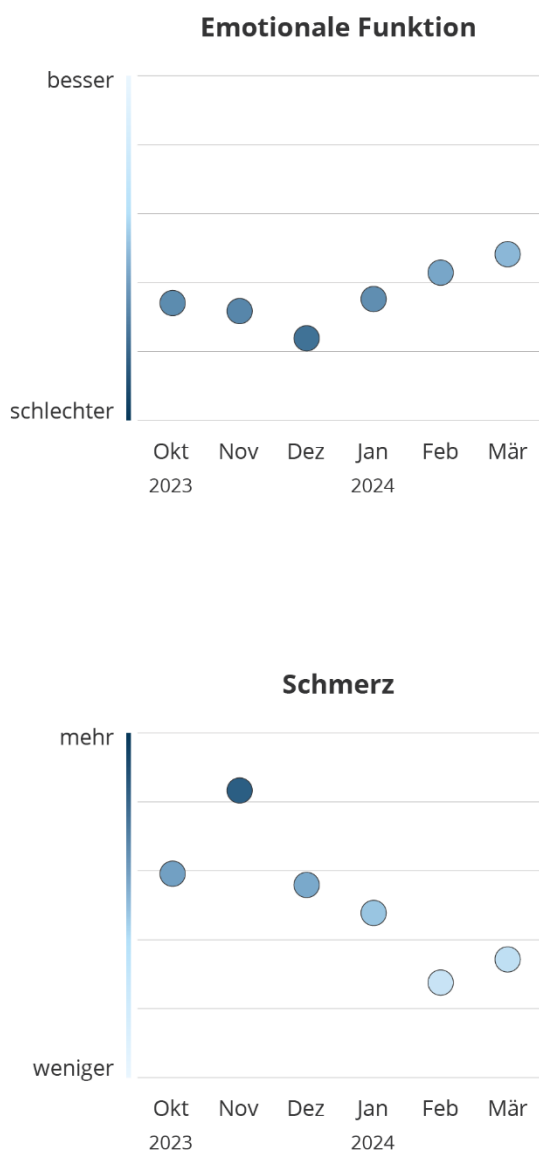


Figure 1 Final design

In this design research study, we aimed to create graphs that are easy to use for both patients and health care professionals, give a correct representation of the data and result in a right interpretation with a minimal risk of misinterpretation. In this report we show and substantiate with arguments all design decisions that together resulted in the designs shown in figure 1. The final design can be characterized as a clear and intuitive representation of the data by providing a balanced amount of information to optimize the chances for correct interpretation.

You can read the sentences as if they were preceded by “We decided to...”. Example: We decided to not use / use / position (...).

1 Graph type

Use a dot chart as it is the best representation of individual data points and it requires the least amount of visual information.

A line graph would give a suggestion of the values in between while this might not be correct. In addition, a line gives unnecessary and distracting long lines when data points are far apart. A column or bar chart would suggest that values are added, while the data just represents a value at a certain moment.

Plot time along the x-axis, as it fits with what people are used to. Plotting time along the x-axis is common practice. It also fits with the (European) way of reading from left to right.

2 Colour

Use colour to communicate a data point's position on the y-axis in addition to the vertical position of the data points. It makes the graphs easier to read. See figure 2.

Don't use a red to green gradient as it adds a value judgement to the data; green is associated with good and red with bad, which might result in less correct representation of the data. A qualitative study (n=8 patients, n=5 health care professionals) showed that some patients felt (strong) aversion towards using the red to green gradient. In addition, there is little difference between red and green when printed in black and white and some people are not able to see the difference between red and green due to colour blindness. A graph with a red to green gradient is shown in the bottom graph in figure 2.

Use a single colour gradient as it gives a better representation of the data than a multi-colour gradient and still adds to the interpretation of the data. In addition, it remains informative when printed in black and white and for people with colour blindness. See figure 2.

Relate dark to negative and light to positive values as it fits the mental model of *dark* being denser, heavier, more and *light* being light-weighted, less heavy, good.

Use blue as it is a neutral colour, being associated with calmness.

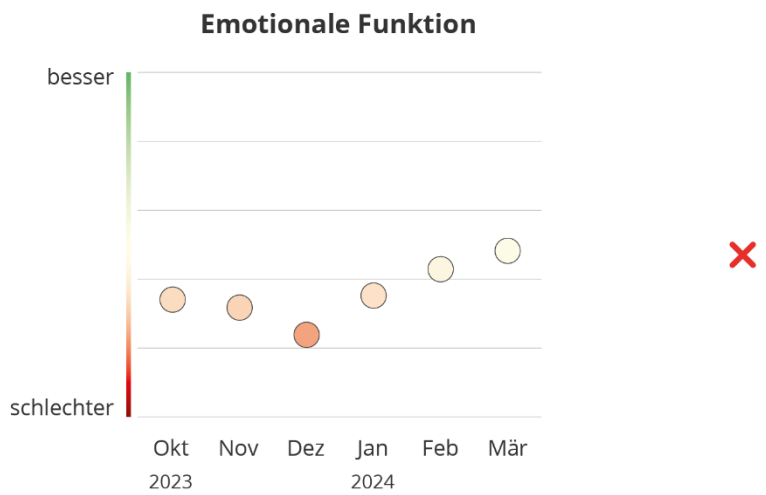
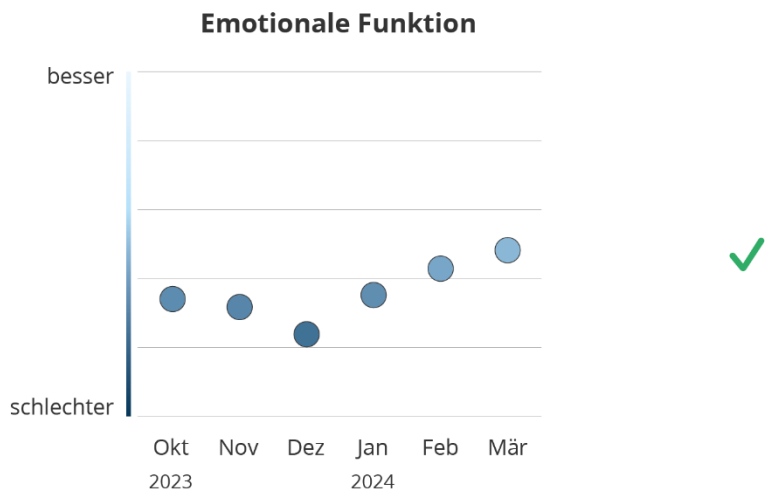


Figure 2 Use of colour: blue and red to green gradient

3 Grid

Show horizontal lines as it supports interpretation of the data points. Horizontal lines make it easier to assess the (vertical) position and the difference between the data points (figure 3).

Divide the graph area into five parts using six horizontal lines

Five areas provide a good balance between readability and the amount of lines used.

Don't show vertical lines as adding the lines results in a less clear graph. It diminishes the support of the horizontal lines in interpretation of the vertical difference between the data points (figure 3). In addition, they add little to the interpretation/readability of the data points as it goes without saying that a point represents the next point in time.

4 Readability

Position all labels horizontal to make them easy to read.

Use a font size of 8 to 11 points for the text labels in the graph. Always test the readability with the user group.

5 X-axis

Use abbreviations of the months as they are easier to interpret than a numerical notation. The abbreviations are commonly understood and enable horizontal presentation, making it easier to read (figure 3).

Position the year labels close to the month labels as the information is related (figure 3).

Only add a year label to the first month of a year to minimize the amount of information. Note that the first month in a graph is not always January. It can be any month as a graph might not start with January (figure 3). Also, if the interval of a graph is, for instance, every three months, the first month of a year is not always January.

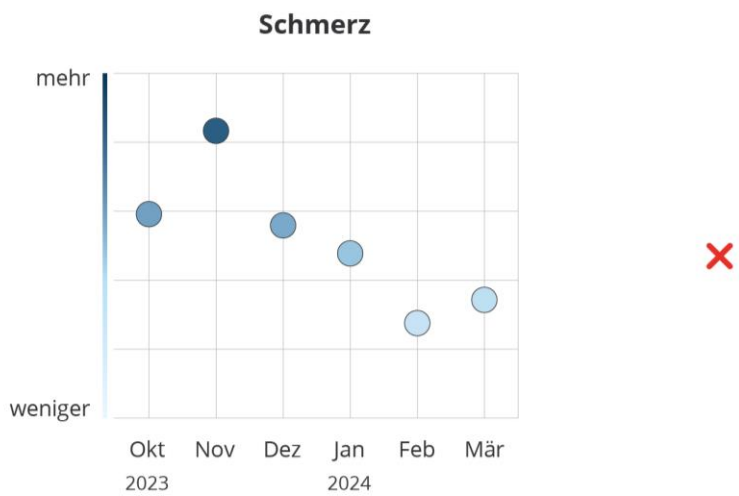
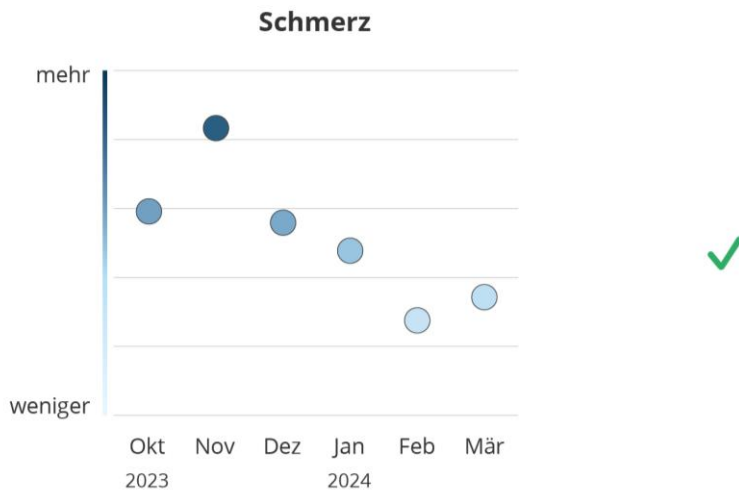
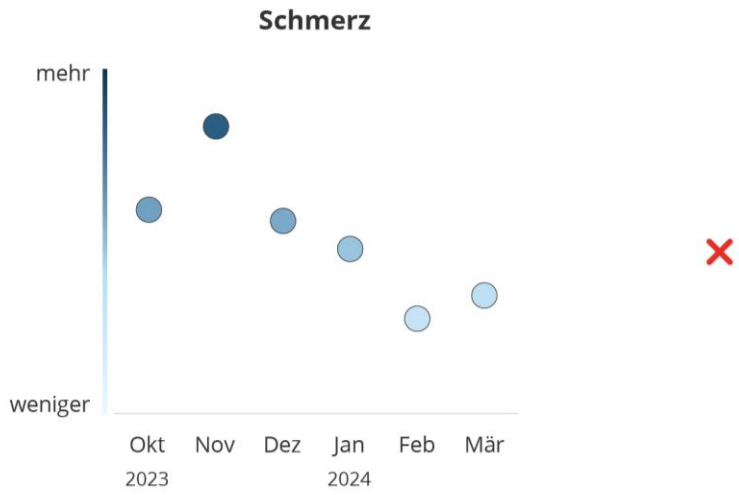


Figure 3 Grid

Don't show a strong x-axis line as it undesirably emphasizes the bottom of the graph and distracts from reading the data points and the trend they visualise (figure 4).

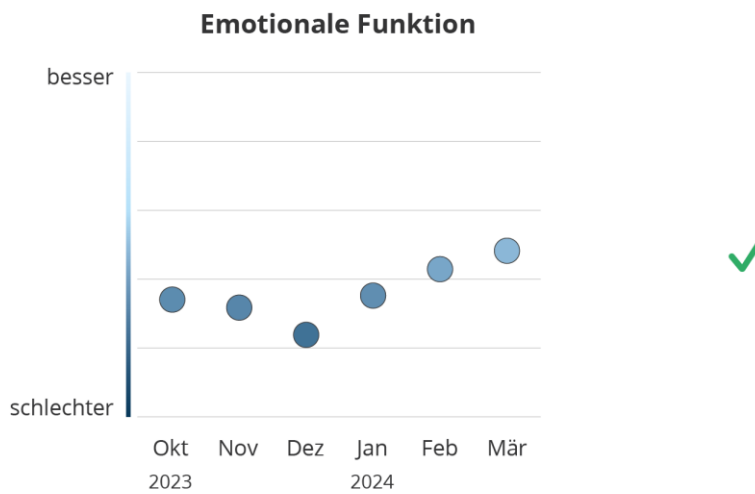
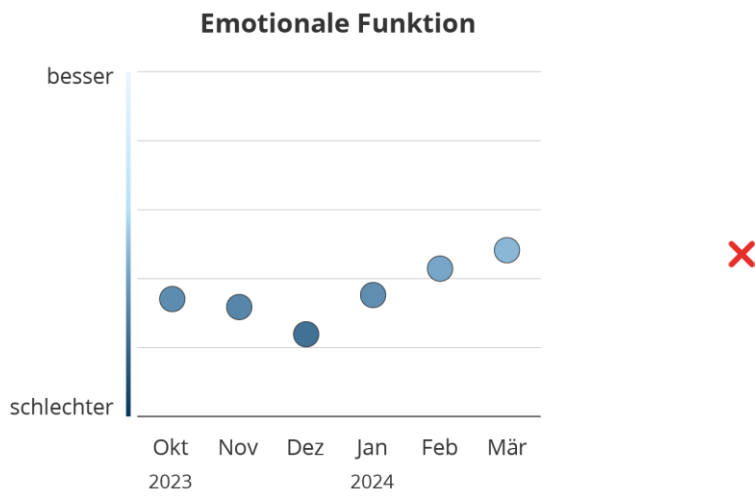


Figure 4 X-axis

6 Y-axis

Use **besser** and **schlechter** to define the extremes of the y-axis of the quality of life and functioning graphs as the labels are short, clear, unambiguous and go well with both the quality of life and all different functioning graphs. Using *besser* (better) and *schlechter* (worse) is preferred over using *gut* (good) and *schlecht* (poor) as functioning and quality of life are relative concepts.

Use **mehr** and **weniger** to define the extremes of the y-axis of the symptom graphs as they are short, clear, unambiguous and make a correct sentence with the symptoms which makes the data easier to interpret; more pain, less pain, more fatigue, less fatigue etc. Using *mehr* and *weniger* is preferred over using *starke* (strong) and *keine* (none) to be consistent with the labels of the functioning graphs, *besser* and *schlechter*. See bottom graph in figure 5.

Don't use 😊 (happy face) and ☹️ (sad face) to define the extremes of the y-axis as there is too little visual difference between the two icons, making interpretation of the data less easy. See top graph in figure 5.

Don't use + (plus) and – (min) to define the extremes of the y-axis as it might result in misinterpretation of the data. Example: + can be interpreted as *more pain* but also as *a better situation*. See middle graph in figure 5.

Don't add arrows to the labels as they don't add to the understanding of the data and result in a less clear and less balanced y-axis. This decision was based on findings from / confirmed by a qualitative usability study with 8 patients and 5 health care professionals.

Add a colour bar to the y-axis to define the extremes and colour gradient in between.

7 Directionality

Position the positive extreme (**besser**) at the top in the quality of life and functioning graphs as it corresponds to the mental model of a positive line going upwards. Both quality of life as well as the different aspects of a patient's functioning (physical functioning, emotional functioning, etc.) are positively framed; you want them to be high. This makes it logical to position better results higher than worse results.

Position the negative extreme (**mehr**) at the top in the symptom graphs as it makes interpretation of the data easier, more intuitive, because symptoms are framed negatively. This makes it logical to position worse results (more pain) higher than better results (less pain). A qualitative study (n=13) showed that changing the direction did not result in misinterpretation of the data.

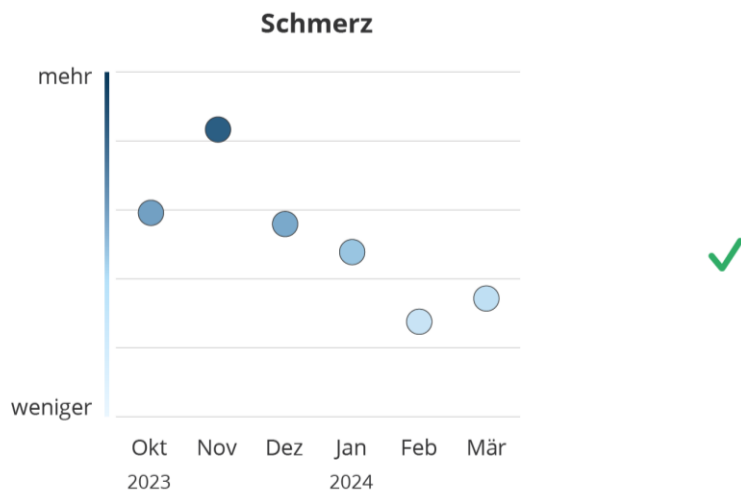
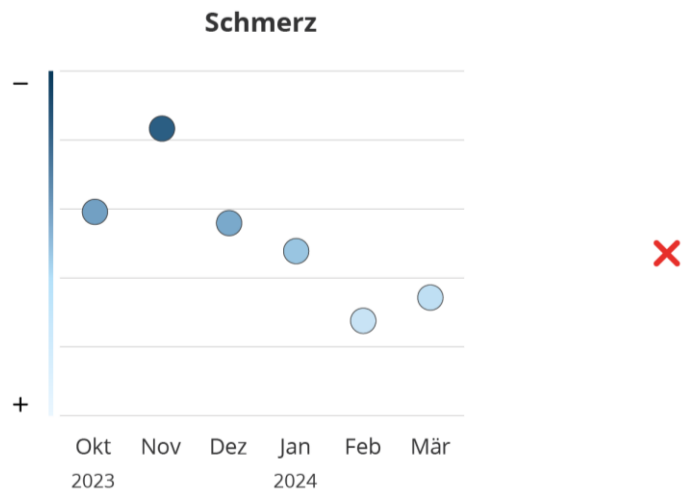
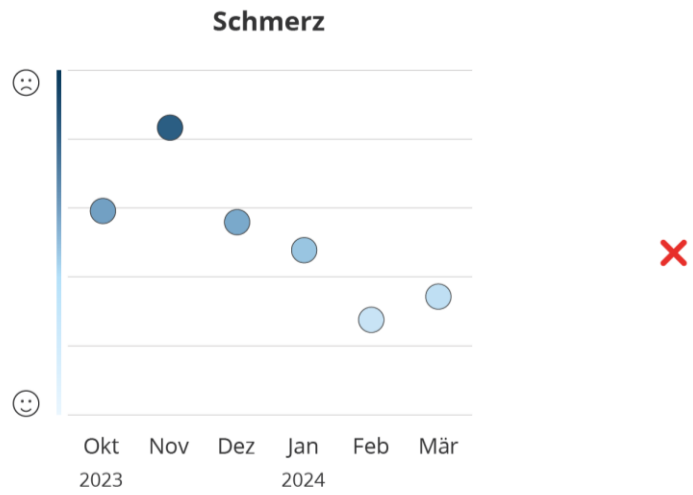


Figure 5 Labelling

8 Reference data

Use a light-coloured background area to highlight the area where most people in the reference group are located, as it is subtle but clear enough (figure 6).

Provide explanation on the reference area to avoid misinterpretation (figure 6).

Don't show a mean line as it might result in misinterpretation and does not add to interpretation of the data (figure 6).

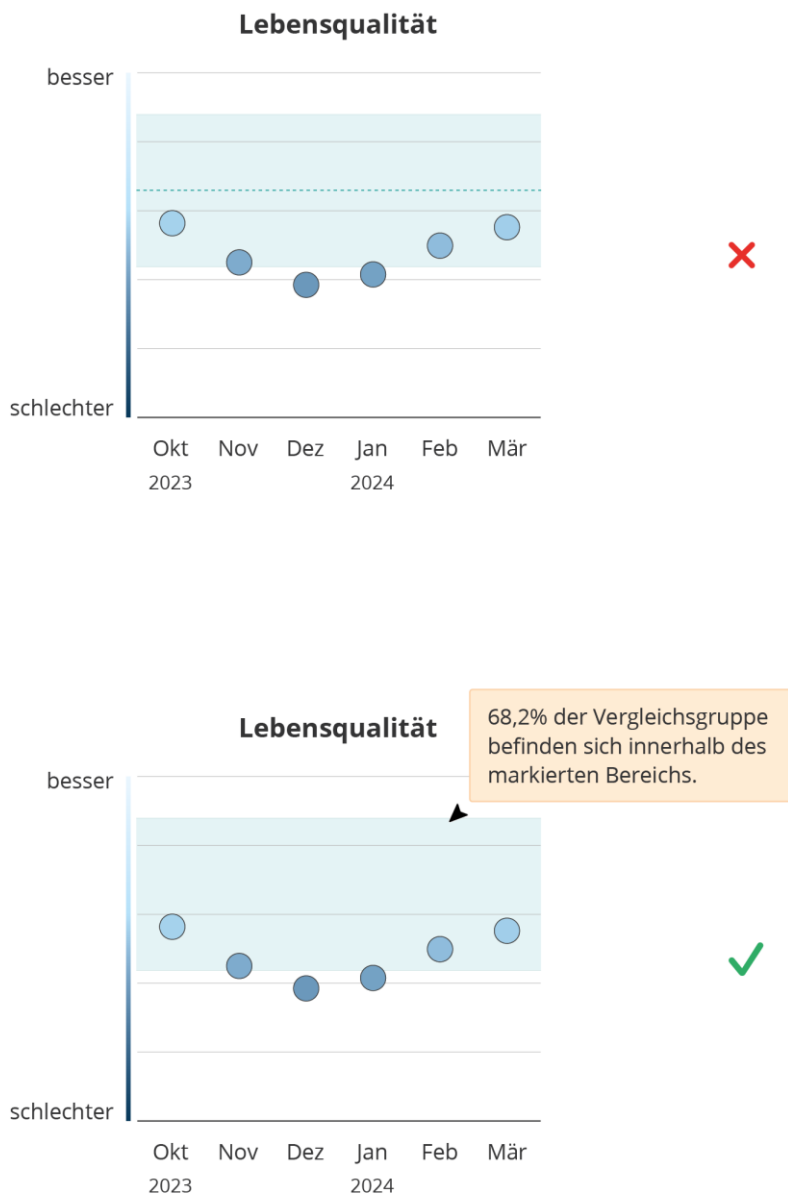


Figure 6 Reference data

9 Missing data points

Subtle mark the area of the missing data points to avoid misinterpretation without drawing too much attention (figure 7).

10 Alarming data points

Mark the data point itself, not the area as this is best understood (figure 8).

Use an alarming colour like red or orange to mark alarming data points as it contributes to the interpretation (figure 8).

Don't change the shape of the data points to mark alarming data points as it results in a less clear, less balanced graph (figure 8).

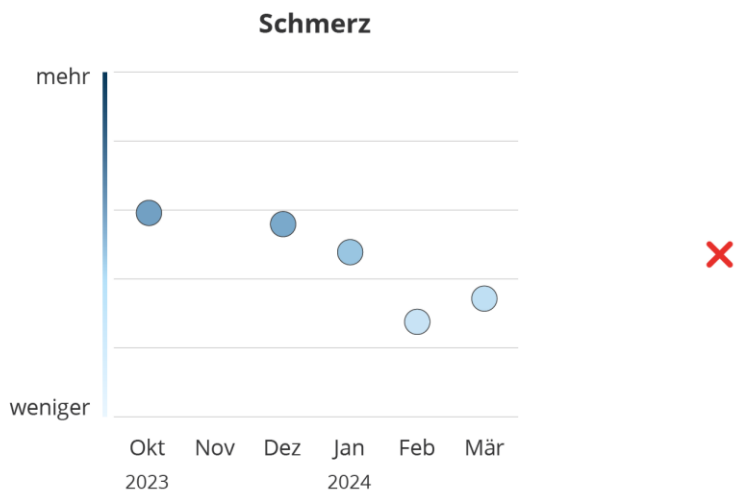
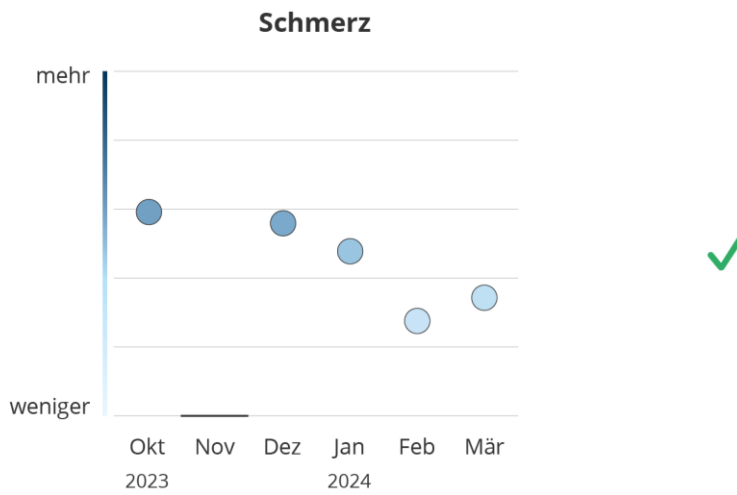
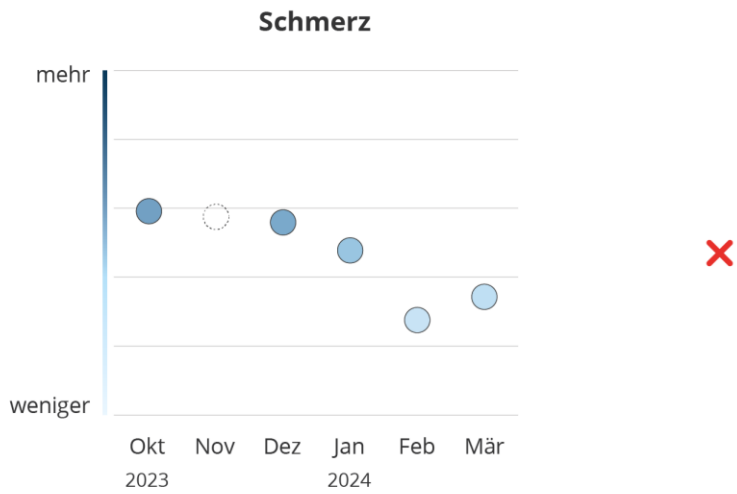


Figure 7 Missing data points

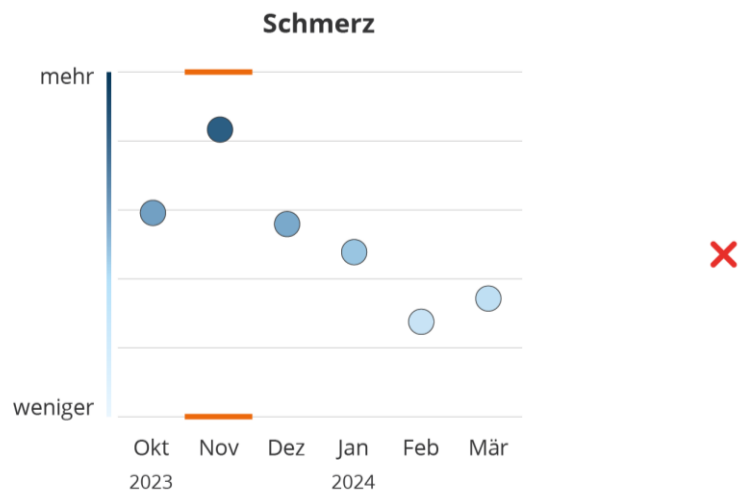
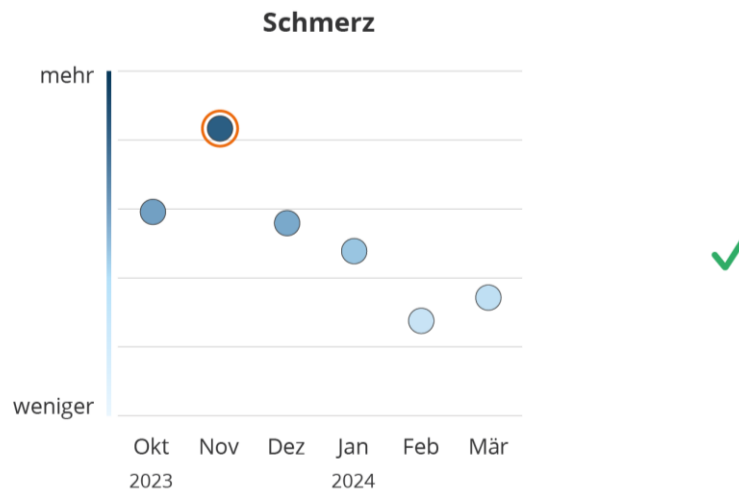
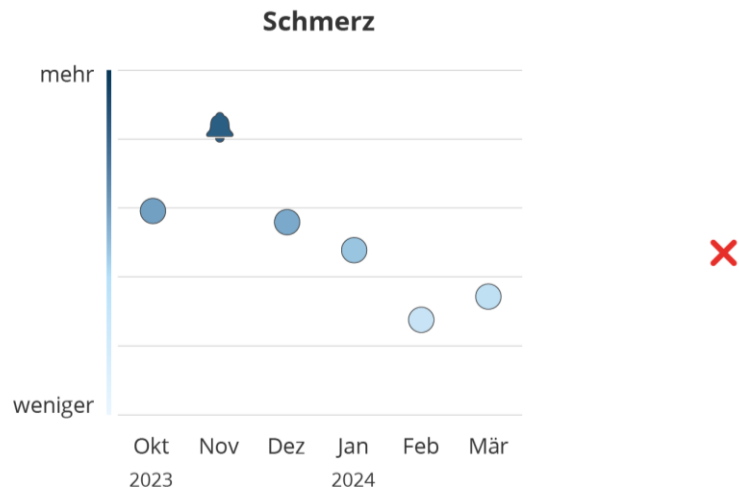


Figure 8 Alarming data points

