

Teil II:

Effekte der neuen Versorgungsform OBERBERG_FAIRsorgt.

Ergebnisse der quantitativen Analysen der Evaluation

Christian Grebe

1	Einleitung	170
2	Kosteneffekte und Inanspruchnahme von Regelleistungen: Analyse von Krankenkassen-Routinedaten	171
2.1	Methodik	171
2.2	Ergebnisse	177
3	Veränderung von gesundheitsbezogenen Outcomes nach einem Jahr Versorgung – standardisierte Befragung der Interventionsgruppe	189
3.1	Methodik	189
3.2	Ergebnisse	192
4	Diskussion	207
5	Literatur	210

1 Einleitung

Der quantitative Teil der Evaluation von OBERBERG_FAIRsorgt gliederte sich in zwei separate methodische Ansätze. Zum einen wurden Fragestellungen untersucht, die Effekte der Versorgungsform auf die Wirtschaftlichkeit (Kosten für Leistungen der gesetzlichen Kranken- und Pflegeversicherung), auf die Ausschöpfung von Regelleistungen (u.a. Tages- und Kurzzeitpflege) sowie auf notfallbedingte Krankenhausaufenthalte, den Verbleib in der Häuslichkeit und die Pflegebedürftigkeit adressierten. Diese Analyse erfolgte auf der Grundlage von Routinedaten der AOK Rheinland/Hamburg, wobei die Teilnehmenden an der Versorgungsform mit zwei Vergleichsgruppen mit ähnlichen Fällen von Versicherten verglichen wurden, die mittels eines Matchingverfahrens ausgewählt wurden.

Zum anderen wurde untersucht, in wieweit sich verschiedene Aspekte des Gesundheitszustands der Teilnehmenden (u.a. körperliche Funktionsfähigkeit, Erschöpfung, Schmerzen, emotionale Belastung) sowie ihre Resilienz, Selbstwirksamkeit, gesundheitsbezogene Lebensqualität und subjektive Gesundheitskompetenz nach einem Jahr Versorgung verändern. Hierzu wurden die Teilnehmenden mittels einer standardisierten Pre-Post-Befragung mit Paper-and-Pencil-Fragebögen zweimal im Abstand von 12 Monaten befragt. In der zweiten Befragungswelle wurde zudem nach der Zufriedenheit mit der Neuen Versorgungsform und ihren Teilleistungen sowie deren Bedeutung gefragt und der Prozess der Leistungserbringung näher beleuchtet.

2 Kosteneffekte und Inanspruchnahme von Regelleistungen: Analyse von Krankenkassen-Routinedaten

2.1 Methodik

2.1.1 Evaluationsdesign

Für diesen Teil der Evaluation wurde ein 3-Gruppen Pre-Post-Design mit zwei Messzeitpunkten gewählt. Der Interventionsgruppe (den in die Neue Versorgungsform OBERBERG_FAIRsorgt eingeschriebenen Versicherten) wurden zwei Vergleichsgruppen gegenübergestellt. Bei der ersten Vergleichsgruppe (KG OBK) handelte es sich um Versicherte der AOK Rheinland/ Hamburg mit Wohnort im Oberbergischen Kreis, die die Einschlusskriterien für die Neue Versorgungsform erfüllten, sich aber nicht in die Neue Versorgungsform eingeschrieben hatten. Die zweite Vergleichsgruppe (KG RSK) wurde aus Versicherten der AOK Rheinland/ Hamburg gebildet, deren Wohnort sich im benachbarten Rhein-Sieg-Kreis befand und die ansonsten ebenfalls die Einschlusskriterien für OBERBERG_FAIRsorgt erfüllten. Die Fälle der Vergleichsgruppen wurden mittels Propensity Score Matching zu den Fällen der Interventionsgruppe gematched.

Bei der AOK Rheinland/Hamburg wurden auf der Basis einer Datensatzbeschreibung die entsprechenden Datensätze angefordert. Von den Teilnehmern der Interventionsgruppe wurde dazu eine Einwilligung auf Basis einer informierten Zustimmung eingeholt, für die Vergleichsgruppen erfolgte die Übermittlung der Routinedaten auf Grundlage einer Genehmigung durch das Ministerium für Arbeit, Gesundheit und Soziales des Landes Nordrhein-Westfalen (MAGS NRW) nach §75 SGB X.

Die Daten wurden für zwei Messzeitpunkte erhoben. Baseline (t0) bezieht sich auf den Zeitpunkt unmittelbar vor Interventionsstart. Da Routinedaten je Quartal vorliegen, ist somit das Quartal vor Interventionsstart bzw., insbesondere bei kostenbezogenen Outcomes, der Zeitraum der drei Quartale vor Interventionsstart gemeint (siehe Teil II Tabelle 1). Das Follow-Up (t1) lag 9 Monate, also drei Quartale nach Interventionsstart. Je nach Outcome wird auch hier entweder das Quartal, in dem der 9-monatige beobachtete Zeitraum der Intervention endete, betrachtet oder aber die drei Quartale vor t1 (entspricht somit dem gesamten 9-monatigen Interventionszeitraum). Dass in der Routinedatenanalyse nur ein Zeitraum von 9 Monaten beobachtet wurde, obwohl die Intervention einen Zeitraum von 12 Monaten umfasste, ist der Projektlogik und der Natur von Routineda-

ten geschuldet. Routinedaten aus dem ambulanten Sektor liegen bei den Krankenkassen üblicherweise erst nach 9 Monaten vor. Da aufgrund der COVID-19-Pandemie das Ausrollen der Neuen Versorgungsform zudem später erfolgte als ursprünglich geplant, wäre ein 12-Monate umfassender Beobachtungszeitraum, zumindest für die letzte Kohorte, nicht mehr während der Projektlaufzeit analysierbar gewesen.

Die Teilnehmenden starteten versetzt in drei Kohorten in die Versorgung, jeweils zum Quartalsbeginn (Januar, April und Juli 2022). Mit Genehmigung des Projektträgers starteten zudem zehn Teilnehmende, zu individuellen Zeitpunkten, als eine Art „Testlauf“ bereits im Herbst 2021 in die Versorgung. Diese Fälle sind nicht Teil der Stichprobe der Routinedatenanalyse, da sich für sie keine saubere Trennung in versorgte und unversorgte Quartale vornehmen lässt.

2.1.2 Fragestellungen und Outcomes

Mittels der Routinedatenanalyse wurden die Folgenden Fragestellungen untersucht:

1. Wirtschaftlichkeit (primäres Outcome): Unterscheiden sich die Teilnehmer an der Neuen Versorgungsform hinsichtlich der GKV-Kosten im Interventionszeitraum von Angehörigen der Vergleichsgruppen (nicht eingeschriebene Versicherte aus dem Oberbergischen Kreis, Versicherte aus der Vergleichsregion Rhein-Sieg-Kreis)?
2. Wirtschaftlichkeit: Unterscheiden sich die Teilnehmer an der Neuen Versorgungsform hinsichtlich der Krankenhauskosten im Interventionszeitraum von Angehörigen der Vergleichsgruppen?
3. Ausschöpfung von Regelleistungen: Unterscheiden sich die Teilnehmer an der Neuen Versorgungsform hinsichtlich der Inanspruchnahme von Kurzzeitpflege im Interventionszeitraum von Angehörigen der Vergleichsgruppen?
4. Ausschöpfung von Regelleistungen: Unterscheiden sich die Teilnehmer an der Neuen Versorgungsform hinsichtlich der Inanspruchnahme von Tagespflege im Interventionszeitraum von Angehörigen der Vergleichsgruppen?
5. Ausschöpfung von Regelleistungen: Unterscheiden sich die Teilnehmer an der Neuen Versorgungsform hinsichtlich der Inanspruchnahme von Pflege- und zum Verbrauch bestimmten Hilfsmitteln im Interventionszeitraum von Angehörigen der Vergleichsgruppen?

6. Ausschöpfung von Regelleistungen: Unterscheiden sich die Teilnehmer an der Neuen Versorgungsform hinsichtlich der Inanspruchnahme von Maßnahmen zur Verbesserung des Wohnumfelds im Interventionszeitraum von Angehörigen der Vergleichsgruppen?
7. Notfallbedingte Krankenhausaufnahmen: Unterscheiden sich die Teilnehmer an der Neuen Versorgungsform hinsichtlich der Anzahl ihrer notfallbedingten Krankenhausaufnahmen im Interventionszeitraum von Angehörigen der Vergleichsgruppen?
8. Verbleib in der Häuslichkeit: Unterscheiden sich die Teilnehmer an der Neuen Versorgungsform hinsichtlich der Inzidenz einer Inanspruchnahme vollstationärer Dauerpflege im Interventionszeitraum von Angehörigen der Vergleichsgruppen?
9. Verschlechterung der Pflegebedürftigkeit: Unterscheiden sich die Teilnehmer an der Neuen Versorgungsform hinsichtlich einer Verschlechterung (bzw. Neu-Zuerkennung) ihres SGB-XI- Pflegegrades im Interventionszeitraum von Angehörigen der Vergleichsgruppen?
10. Veränderung der Multimorbidität: Unterscheiden sich die Teilnehmer an der Neuen Versorgungsform hinsichtlich ihrer Multimorbidität am Ende des Interventionszeitraums von Angehörigen der Vergleichsgruppen?

Die Operationalisierung der entsprechenden zehn Outcomes zeigt Teil II Tabelle 1.

Teil II Tabelle 1: Operationalisierung der zehn Outcomes

Nr.	Endpunkt	Operationalisierung
1	GKV Gesamtkosten	Gesamtkosten für SGB V- und SGB XI-Leistungen in den 3 Quartalen vor Interventionsstart (t0) und in den 3 Quartalen des Interventionszeitraums (t1)
2	Krankenhauskosten	Kosten für akutstationäre Leistungen in den 3 Quartalen vor Interventionsstart (t0) und in den 3 Quartalen des Interventionszeitraums (t1)
3	Kurzzeitpflegekosten	Kosten für Leistungen nach §42 SGB XI den 3 Quartalen vor Interventionsstart (t0) und in den 3 Quartalen des Interventionszeitraums (t1)
4	Tagespflegekosten	Kosten für Leistungen nach §41 SGB XI den 3 Quartalen vor Interventionsstart (t0) und in den 3 Quartalen des Interventionszeitraums (t1)

5	Kosten für Pflege- und zum Verbrach bestimmte Hilfsmittel	Summierte Kosten für Leistungen nach §40 SGB XI und §33 SGBV (Produktgruppe 54) den 3 Quartalen vor Interventionsstart (t0) und in den 3 Quartalen des Interventionszeitraums (t1)
6	Maßnahmen zur Verbesserung des Wohnumfelds	Fälle mit Leistungen nach §40 Abs. 4 SGB XI im Interventionszeitraum (t1)
7	Notfallbedingte Krankenhausaufnahmen	Fälle mit als Notfall kodierter stationärer Krankenhausaufnahme im Versorgungszeitraum (t1)
8	Verbleib in der Häuslichkeit	Fälle mit Leistung stationäre Sachleistungen (SGB XI) im Quartal 3 Monate nach Interventionsstart (t1), bei denen zugleich im Quartal vor Interventionsstart (t0) keine solche Leistung vorhanden war
9	Pflegebedürftigkeit	Fälle mit höherem (bzw. neu zuerkannten) Pflegegrad im 3. Quartal der Versorgung (t1), verglichen mit dem Quartal vor Versorgungsbeginn (t0)
10	Multimorbidität	Exlihauser Komorbiditätsindex (Elixhauser et al. 1998) im Quartal vor Interventionsstart (t0) und im 3. Quartal der Versorgung (t1)

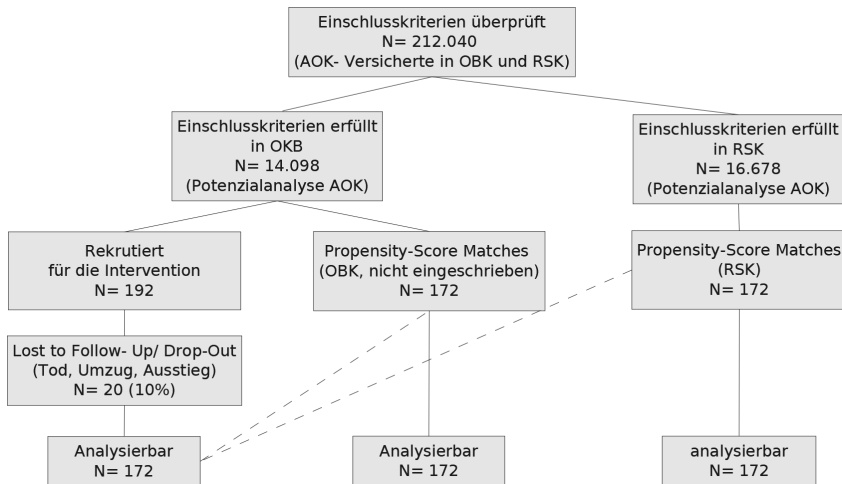
2.1.3 Fallzahlberechnung

A priori wurde mit der mit Software G*Power 3.1 (Faul et al. 2009) eine Fallzahlberechnung für die Analyse des primären Outcomes durchgeführt. Da als weiteres Outcome auch die Kosten für stationäre Krankenhausaufenthalte untersucht wurden, die eine Teilmenge der Kostenvariable des primären Outcomes darstellen, wurde der Alphafehler dieser beiden Tests nach Bonferroni korrigiert und so auf 0,025 gesetzt. Die Fallzahl wurde so kalkuliert, dass bei $df = 2$ Freiheitsgraden mit einer Power von 80% einer Effektstärke von $f = 0,15$ nachgewiesen werden kann.

Die erforderliche Fallzahl betrug demnach $N=515$, je Gruppe folglich $N=172$. Die Fallzahlsschätzung erfolgte unter der Annahme, dass von 10% der an t_0 in die Neue Versorgungsform eingeschriebenen Fälle am Follow-Up-Messzeitpunkt t_1 keine Daten vorliegen (Drop-Out bzw. Lost-to-Follow-Up), so dass sich eine Fallzahl von $N=192$ Fällen in der Interventionsgruppe errechnete, die in die Neue Versorgungsform eingeschrieben werden musste. Da das Matching erst erfolgte, nachdem die letzte Kohorte der Interventionsgruppe t_1 erreicht hatte, kam es in den Vergleichsgruppen zu keinem Lost-to-Follow-Up, da Verstorbene oder Verzoogene der für das Matching verfügbaren Population per se entzogen sind.

Das Potenzial der Versicherten mit den definierten Ein- und Ausschlusskriterien betrug nach einer Analyse der AOK Rheinland/ Hamburg 14.098 Personen im Oberbergischen Kreis (OBK) sowie 16.678 Personen in der Vergleichsregion Rhein-Sieg Kreis (RSK). Das Consort Schema zeigt Teil II Abbildung 1.

Teil II Abbildung 1: CONSORT Flow Chart (a-priori)



2.1.4 Matching

Das Sampling der Fälle der beiden Vergleichsgruppen erfolgte mittels Propensity Score Matching (*Caliendo et al. 2008*). Dazu wurden die in Teil II Tabelle 2 dargestellten Kovariate genutzt. Beim Matching wurde sichergestellt, dass das Kriterium stationäre Dauerpflege exakt übereinstimmt, stationäre Fälle wurden also stets zu stationären Fällen gematched.

Da nicht alle prinzipiell matchbaren Fälle der Vergleichspopulationen zu jeder der drei Kohorten gematched werden konnten (weil für das t0-Quartal einer oder mehrerer Kohorten die Einschlusskriterien nicht vollumfänglich erfüllt waren), wurden zunächst Fälle zur dritten Kohorte gematched (ohne zurücklegen), dann zur zweiten und abschließend zur ersten Kohorte. Für jedes Match der Vergleichsgruppen gingen die Outcome-Daten basierend auf dem gleichen Zeitraum in die Analyse ein wie beim jeweiligen Studienteilnehmer, zu dem der Fall gematched wurde (für die erste Kohorte

mit Interventionsstart zum 01.01.2022 für t0 also etwa die GKV-Gesamtkosten der Quartale 2-4/2021 und für t1 jene der Quartale 1-3/2022).

Das Matching wurde in R (R Core Team 2023) mittels des Packages MatchIt (Ho et al. 2011) durchgeführt.

Teil II Tabelle 2: Kovariate für das Propensity Score Matching

Kriterium	Operationalisierung
Alter	2022-Geburtsjahr
Geschlecht	In Q4/2021
Multimorbidität	Elixhauser Komorbiditäts-Index im Q4/2021
Pflegegrad	Im Q4/2021
SGB XI Leistungsart	im Quartal vor t ₀ : Pflegegeld, Sachleistung ambulant, Sachleistung stationär
GKV Kosten	Kosten im Q4/2021
Krankenhauskosten	Kosten im Q4/2021
SGB XI Kosten für Leistungen nach §42 (Kurzzeitpflege)	Kosten im Q4/2021
Leistungen nach §42 SGB XI (Tages- und Nachtpflege)	Kosten im Q4/2021
Leistung nach § 40 Abs. 1 SGB XI (Pflegehilfsmittel) und Leistungen nach § 33 SGB V; Produktgruppe 54 des Hilfsmittelverzeichnisses	Kosten im Q4/2021
Leistung nach § 40 Abs. 4 SGB XI (Maßnahmen zur Verbesserung des individuellen Wohnumfeldes)	im Q4/2021 (Kosten > 0 EUR ja/nein)
Leistung nach § 45 SGB XI (Pflegekurse)	im Q4/2021 in Anspruch genommen ja/nein
Leistung nach § 7a SGB XI (Pflegeberatung)	im Q4/2021 in Anspruch genommen ja/nein
Leistung nach § 45a SGB XI (Angebote zur Unterstützung im Alltag)	im Q4/2021 in Anspruch genommen ja/nein
Leistung nach § 33 SGB V; für die Produktgruppen 1-53; 98-99 des Hilfsmittelverzeichnisses	(separat für die Produktgruppen 3-, 8, 10-15, 17, 19-21, 23, 25, 29, 33): im Q4/2021 in Anspruch genommen ja/nein
wohnt in ambulant betreuter Wohngruppe (Leistung nach §38a SGB XI)	im Q4/2021 in Anspruch genommen ja/nein

Demenz-Diagnose	im Quartal 4/2021 Diagnose F00-F03; G30 ICD-10 dichotom ja (mindestens eine der o.g. Diagnosen) / nein
Bevölkerungsdichte am Wohnort	< Median BRD (232 Einwohner/km ²) (ja/nein) Quelle: NRW.IT;
Hausarztichte	Versorgungsgrad <100% am Wohnort; Quelle: Bedarfsplan KVNo vom 30.04.2019;

2.1.5 Statistische Analyse

Die Daten der kostenbezogenen Outcomes sowie des Outcomes Multimorbidität sind metrisch skaliert. Für diese Outcomes wurden ANCOVAs berechnet, bei denen die Baseline-Werte (t0) als Kovariate in die Analyse eingingen. Fälle mit standardisierten Residuen >3 wurden als Outlier angesehen und aus der Analyse des betreffenden Outcomes ausgeschlossen. Berichtet werden die p-Werte des Gruppeneffekts, die Estimated Marginal Means an t0 (identisch für die 3 Untersuchungsgruppen) und an t1 sowie deren 95%-Konfidenzintervalle. Die Analysen erfolgten in R (R Core Team 2023), zur Berechnung der Estimated Marginal Means wurde das Package emmeans (Lenth 2023) eingesetzt.

Die übrigen analysierten Outcomes sind dichotom. Für diese Outcomes (Maßnahmen zur Verbesserung des Wohnumfelds, notfallbedingte Krankenhausaufnahmen, Verbleib in der Häuslichkeit und Anstieg der Pflegebedürftigkeit) wurden chi2-Unabhängigkeitstests berechnet. Berichtet wird jeweils die relative Häufigkeit (Event Rate) der Fälle je Untersuchungsgruppe sowie der p-Wert des Hypothesentests. Alle Hypothesen wurden zweiseitig getestet.

2.2 Ergebnisse

2.2.1 Rekrutierte und Versorgte Teilnehmende und Inanspruchnahme der fakultativen Teilleistungen

Insgesamt wurden initial N=321 Versicherte der AOK Rheinland/ Hamburg für die Neue Versorgungsform rekrutiert. Vier Versicherte schieden bereits vor Interventionsstart wieder aus. N=307 Personen starteten in drei Kohorten jeweils zu Quartalsbeginn ab dem 01.01.2022 bis zum 01.07.2022 in die Intervention. Die übrigen zehn Teilnehmenden, die bereits im 2. Halbjahr

2021 rekrutiert worden waren, starteten mit Genehmigung des Projektträgers ab September 2021 bereits als „Testkohorte K0“, ohne Bindung an einen Quartalsbeginn, zu individuellen Terminen in die Versorgung.

Die nicht-obligatorischen Teilleistungen der Versorgungsform wurden mit den im Folgenden dargestellten Fallzahlen erbracht:

- Interdisziplinäre Fallkonsile wurden bei N=14 Fällen umgesetzt, wobei bei drei Teilnehmenden ein zweites Fallkonsil initiiert wurde. Bei 16 dieser 17 Fallkonsile war der Hausarzt beteiligt, in einem Fall war dieser kurzfristig verhindert. An jeweils 12 der Fallkonsile waren Angehörige und Pflegedienste beteiligt, in drei Fällen auch mehrere Pflegedienste. Die dominierende Zusammensetzung (in 12 Fallkonsilen) umfasste mindestens die Beteiligung aus dem zuständigen Fallmanager (RCCM) von OBERBERG_FAIRsorgt, dem Hausarzt, dem Pflegedienst und Angehörigen. An zwei Fallkonsilen war (jeweils neben einem ambulanten Pflegedienst) auch ein Pflegeheim beteiligt, Fachärzte wurden bei zwei der Fallkonsile hinzugezogen. In Einzelfällen waren Physiotherapie, die Krankenkasse, die Berufsgenossenschaft sowie die „Aktion Hilfe für alte Menschen“ beteiligt.
- Die erweiterte Erreichbarkeit wurde von 19 Teilnehmenden insgesamt 48x genutzt. Im Maximum kam ein Fall dabei auf sieben Kontaktaufnahmen, der Median unter den Teilnehmenden, die diese Teilleistung nutzten, liegt bei zwei Anrufen.
- Bei N=99 Fällen wurden telemedizinische Leistungenerbracht.
- Ausnahmslos alle rekrutierten Teilnehmenden stimmten der Nutzung der Kommunikationsplattform und der Speicherung ihrer Daten auf dieser zu.

Die Angaben beziehen sich jeweils auf den gesamten Zeitraum, über den die Neue Versorgungsform erbracht wurde (01.10.2021 bis 30.06.2023, 21 Monate) und schließen ggf. auch Fälle, die zu Drop-Outs wurden, sowie Fälle der „Testkohorte“ K0 mit ein.

2.2.2 Stichprobe für die Analyse der Routinedaten

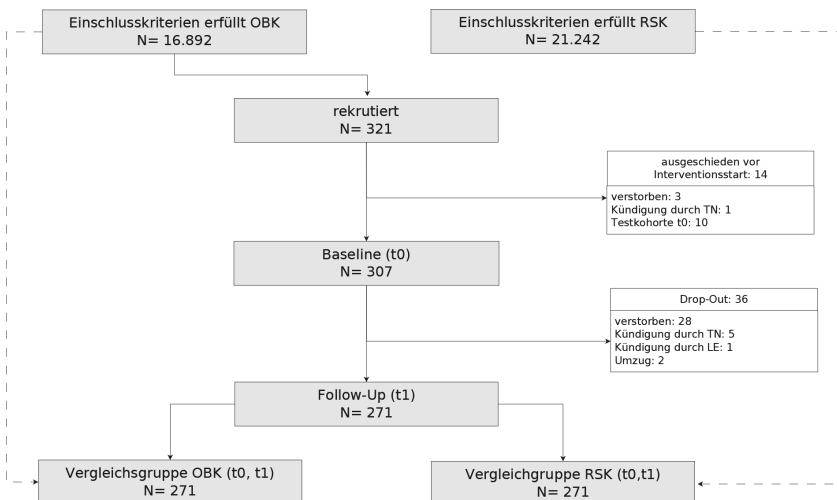
2.2.2.1 Soziodemografische Daten

Von der N=321 im Zuge der neuen Versorgungsform rekrutierten Versicherten schieden vier Personen bereits vor ihrem Interventionsstart aus. Eine dieser Personen zog ihre Einwilligung zurück und kündigte ihre Teilnahme

am Selektivvertrag, drei Personen verstarben zwischen der Rekrutierung und ihrem Interventionsstart. Weitere 10 Teilnehmende gehörten der so bezeichneten „Testkohorte 0“ an. Für diese Fälle lag eine Genehmigung des Projektträgers vor, mit der Versorgung bereits vor der finalen Genehmigung des Interventionsstarts zu beginnen. Die Versorgung dieser Fälle startete fließend im 2. Halbjahr 2021, der Interventionsstart war nicht an ein einheitliches Startdatum gebunden. Da für diese Kohorte keine saubere Unterscheidung in Quartale mit und ohne Versorgung möglich ist, wäre es nicht sinnvoll gewesen, diese Fälle in die Analyse der Routinedaten einzubeziehen. Dies war auch zu keinem Zeitpunkt vorgesehen.

Der Stichprobenumfang der in diesem Evaluationsbaustein analysierten Fälle lag an Baseline (Interventionsstart der jeweiligen Kohorte) bei N=307. Zwischen Baseline (t0) und dem Follow-Up nach drei Quartalen (t1) kam es zu 36 Drop-Outs aus der Versorgung. Der größte Teil dieser Drop-Outs, 28 Personen, verstarb bis zum Follow-Up, weitere fünf Fälle brachen die Versorgung von sich aus ab. Bei einer Person erfolgte ein Abbruch der Versorgung dieser Person durch den Hausarzt und zwei Personen zogen aus dem Landkreis weg.

Teil II Abbildung 2: CONSORT Flowchart für die Analyse der Routinedaten



Diese Fälle erreichten den Messzeitpunkt t1 folglich nicht, so dass der analysierbare Stichprobenumfang jene N=271 Fälle beträgt, die den regulär

versorgten drei Kohorten angehörten und für die im Follow-Up analysierbare Routinedaten vorlagen (Teil II Abbildung 2). Alle weiteren Daten beziehen sich auf diesen N=271 Fälle. Die Lost-to-Follow-Up-Rate für diesen Evaluationsbaustein liegt somit bei 15,6%.

Das Durchschnittsalter der Teilnehmenden lag bei $79,42 \pm 7,63$ Jahre¹, von denen 60,5% weiblichen und 29,5% männlichen Geschlechts waren.

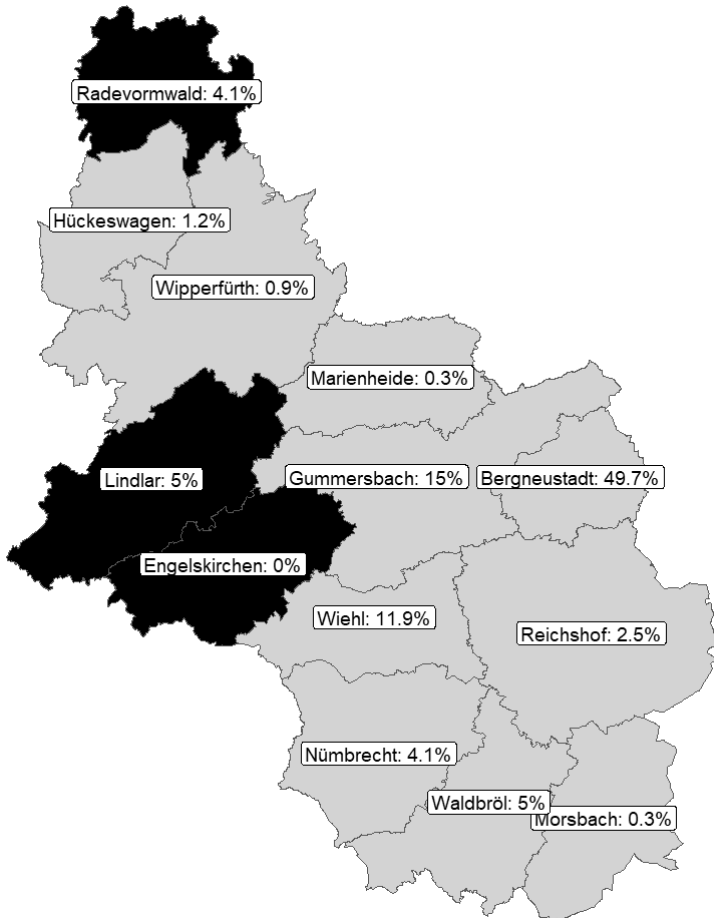
Die Verteilung der rekrutierten Teilnehmer über die Kommunen des Landkreises Teil II Abbildung 3. Deutlich wird zum einen eine stark ungleichmäßige Verteilung über die Kommunen, die auch keinen Bezug zur jeweiligen Einwohnerzahl aufweist. So wohnten 49,7% der rekrutierten Versicherten in Bergneustadt, weitere 15% in der einwohnerstärksten Kommune, der Kreisstadt Gummersbach und weitere 11,9% in Wiehl. Die übrigen Kommunen stellten maximal jeweils 5% der Teilnehmer. Aus Wipperfürth, Morsbach, Marienheide und Hückeswagen konnten nur vereinzelt Teilnehmer rekrutiert werden, aus Engelkirchen gar keiner.

4,1% der Teilnehmenden lebten in Wohnorten mit einer Bevölkerungsdichte unter dem Median der Kommunen der Bundesrepublik Deutschland (232 Einwohner/ km²). Teil II Abbildung 3 illustriert neben der Verteilung der versorgten Teilnehmer auch den hausärztlichen Versorgungsgrad der Kommunen (hellgrau: Versorgungsgrad <100%). 88,6% der Teilnehmenden lebten in Kommunen, die nach den Maßstäben der Kassenärztlichen Bundesvereinigung 2024 (Bedarfsplanung) einen Versorgungsgrad von <100% in der hausärztlichen Versorgung aufweisen.

Weitere soziodemografische Daten können aus den Routinedaten nicht bezogen werden. Weitere Angaben wurden sowohl in der Pre-Post-Befragung der Interventionsgruppe (siehe Abschnitt 3) als auch im qualitativen Baustein erhoben und sind dort, bezogen auf die jeweilige Stichprobe, dargestellt. Eine Verknüpfung mit dem Sample der Routinedatenanalyse ist nicht vollständig möglich, da die Stichproben nicht identisch sind.

1 durch $\pm$ getrennte Werte kennzeichnen im Folgenden das arithmetische Mittel und die zugehörige Standardabweichung.

Teil II Abbildung 3: Verteilung der versorgten Teilnehmer über die Kommunen



2.2.2.2 Diagnose-, Versorgungs- und Kostenbezogene Daten

Die folgenden Angaben beziehen sich auf das 4. Quartal 2021, das einheitlich und unabhängig vom tatsächlichen Interventionsstart (Kohortenzugehörigkeit) für das Matching mit den Vergleichsgruppen herangezogen wurde.

Demnach wiesen 8,5% eine in den Routinedaten dokumentierte Demenzdiagnose (mindestens eine Diagnose aus F00-F03 oder G30 ICD-10) auf.

Der Elixhauser-Komorbiditäts-Index ist ein Maß für Multimorbidität. Er setzt sich aus 30 Erkrankungen zusammen und kann Werte zwischen -19 und 89 annehmen. Höhere Werte zeigen dabei ein höheres Risiko an, während eines Krankenhausaufenthalts zu versterben. Der Teilnehmenden wiesen im Mittel einen Score von $6,33 \pm 7,38$ auf.

49,1% der Teilnehmenden hatte keinen Pflegegrad nach dem SGB XI, demzufolge ist eine knappe Mehrheit der Teilnehmenden nach den Kriterien des SGB XI als pflegebedürftig anzusehen. 9,6% wiesen den Pflegegrad 1 auf, 20,3% Pflegegrad 2, 14,4% Pflegegrad 3, 4,1% Pflegegrad 4 und 2,6% den höchsten Pflegegrad 5. Leistungen stationärer Dauerpflege bezogen 8,1% der Teilnehmenden.

Die Gesamtkosten für SGB V- und SGB XI-Leistungen im 4. Quartal des Jahres 2021 lagen im Mittel bei €3.804,58 $\pm$ €7.045,26. Sie weisen also eine sehr hohe Streuung auf. Gleiches gilt für die Kosten für Krankenhausbehandlungen im selben Zeitraum, die im Mittel bei €1.215,25 $\pm$ €5.057,59 lagen. Die Kosten für Pflege- und zum Verbrauch bestimmte Hilfsmittel (SGB V und SGB XI) lagen in diesem Zeitraum im Mittel bei €25,89 $\pm$ €70,96.

2.2.2.3 Repräsentativität der Interventionsgruppe

Vergleicht man die N=310 Fälle der Interventionsgruppe (exklusive der Fälle, die vor Interventionsstart ausgeschieden waren und der „Testkohorte“ K0) mit der übrigen Grundgesamtheit der bei der AOK Hamburg Rheinland/ Hamburg Versicherten, die die Einschlusskriterien für die neue Versorgungsform zum 01.01.2022 erfüllten (N= 16.582), so lassen sich folgende Befunde feststellen:

Die Interventionsgruppe ist mit $80,86 \pm 7,49$ Jahren im Mittel älter und auch homogener verteilt als die übrige Grundgesamtheit ($77,47 \pm 8,10$ Jahre). Der Unterschied ist statistisch signifikant ($p < 0,01$, t-Test für unabhängige Stichproben).

Die Interventionsgruppe ist multimorbider als die übrige Grundgesamtheit. Der Elixhauser Komorbiditätsscore lag in der Interventionsgruppe im 4. Quartal 2021 im Mittel bei $6,72 \pm 7,54$, in der übrigen Grundgesamtheit

hingegen bei 5.07 ± 6.69 . Auch dieser Unterschied ist signifikant ($p < 0,01$, t-Test für unabhängige Stichproben).

Die Interventionsgruppe unterscheidet sich von der Grundgesamtheit hinsichtlich ihrer sozialrechtlich zuerkannten Pflegebedürftigkeit nach dem SGB XI ($p < 0,001$, Mann-Whitney U-Test). Während in der Grundgesamtheit 29,1% eine Pflegestufe aufweisen, ist dies in der Interventionsgruppe bei 50,6% der Versicherten der Fall. Auch der prozentuale Anteil der Personen, die einen der Pflegegrade 1-5 aufweisen, liegt in der Interventionsgruppe jeweils höher. Statistisch signifikant, aber in der Differenz zur Grundgesamtheit weniger deutlich, fallen die Inanspruchnahmen der SGB XI-Leistungsarten aus. 34,2% der Personen der Interventionsgruppe bezogen Pflegegeld (übrige Grundgesamtheit: 21,6%, $p < 0,001$, Fishers exakter Test), 10,3% bezogen ambulante Sachleistungen (Grundgesamtheit: 4,7%, $p < 0,001$, Fishers exakter Test) und 8,7% erhielten Sachleistungen für stationäre Dauerpflege (Grundgesamtheit: 4,1%, $p < 0,001$, Fishers exakter Test).

Hinsichtlich der Gesamtkosten für SGB V- und SGB XI- Leistungen (4. Quartal 2021) liegt die Interventionsgruppe (wiederum $N=310$, inklusive späterer Dropouts) mit EUR 4.473,40 hingegen niedriger als die übrige Grundgesamtheit (EUR 6.614,44), auch dieser Unterschied ist statistisch signifikant ($p < 0,01$, t-Test für unabhängige Stichproben). Aufgrund der sehr hohen Streuung der Kosten empfiehlt sich für den Vergleich der Mittelwerte eher die Verwendung des Medians anstelle des arithmetischen Mittels. Im Median liegen die Gesamtkosten der Interventionsgruppe bei €1.980,00, in der übrigen Grundgesamtheit bei EUR €2.110,00, also lediglich um €130 auseinander.

2.2.3 Matching

Nach Durchführung des Propensity Score Matchings wurden die in das Matching einbezogenen Variablen hinsichtlich der Nullhypothese analysiert, dass sich die drei Untersuchungsgruppen nicht voneinander unterscheiden. Angelegt wurde ein 5% Signifikanzniveau. Für alle einbezogenen Merkmale bestätigte sich die Nullhypothese, die Untersuchungsgruppen unterscheiden sich nicht signifikant.

2.2.4 Outcomes

2.2.4.1 Wirtschaftlichkeit: Gesamtkosten für SGB V- und SGB XI-Leistungen

Die Estimated Marginal Means der Gesamtkosten für SGB V- und SGB XI-Kosten pro Fall lagen an t0 (Kosten der drei Quartale vor Interventionsstart) bei EUR 9.983,-. Während des betrachteten 9-monatigen Interventionszeitraums bis t1 kam es zu einem Anstieg in allen drei Untersuchungsgruppen. In der Interventionsgruppe lag der EMMEAN der Gesamtkosten mit EUR 10.429,- (95%-CI: EUR 9.471,- bis EUR 11.388,-) dabei am niedrigsten, die Gesamtkosten der beiden Kontrollgruppen (Oberbergischer Kreis: EUR 10.896,-, 95%-CI: EUR 9.937,- bis EUR 11.854,-; Rhein-Sieg-Kreis: EUR 10.714,-, 95%-CI: EUR 9.753,- bis EUR 11.674,-) fielen geringfügig höher aus. Der Unterschied ist nicht signifikant ($p=0,793$), ebenso wenig die paarweisen Vergleiche zwischen den Untersuchungsgruppen. Die Konfidenzintervalle der Untersuchungsgruppen überlappen sich deutlich.

Die ergänzend berechnete rangdatenbasierte Quade-ANCOVA zeigte ebenfalls keinen signifikanten Unterschied zwischen den drei Untersuchungsgruppen ($p=0,923$).

2.2.4.2 Wirtschaftlichkeit: Krankenhauskosten

Die Kosten für Akutversorgung im Krankenhaus lagen an t0 im Mittel bei EUR 2.960,- (EMMEANS). In allen drei Untersuchungsgruppen lagen die Kosten im Interventionszeitraum deutlich niedriger. Die Kosten der Interventionsgruppe beliefen sich auf EUR 1.422,- (95%-CI: EUR 1.028,- bis EUR 1.817,-). In der Kontrollgruppe des Oberbergischen Kreises lagen die Kosten mit EUR 1.339 (95%-CI: EUR 943,- bis EUR 1.734,-) noch niedriger, in der Kontrollgruppe aus dem Rhein-Sieg-Kreis mit EUR 1.650,- (95%-CI: EUR 1.255,- bis EUR 2.044,-) hingegen höher als in den beiden anderen Gruppen. Die Unterschiede an t1 sind nicht signifikant ($p=0,528$). Gleiches gilt für die paarweisen Vergleiche der Gruppen, zudem überlappen sich die Konfidenzintervalle deutlich.

Die ergänzend berechnete rangdatenbasierte Quade-ANCOVA zeigte ebenfalls keinen signifikanten Unterschied zwischen den drei Untersuchungsgruppen ($p=0,920$).

2.2.4.3 Ausschöpfung von Regelleistungen: Kosten für Kurzzeitpflege

Die Kosten für Kurzzeitpflege nach §42 SGB XI lagen an t0 im Mittel (EMMEANS) bei EUR 200,-. In allen drei Untersuchungsgruppen lagen die Kosten im Interventionszeitraum niedriger. Die Kosten der Interventionsgruppe beliefen sich auf EUR 124,- (95%-CI: EUR 78,- bis EUR 169,-). In der Kontrollgruppe des Oberbergischen Kreises lagen die Kosten mit EUR 130,- (95%-CI: EUR 85,- bis EUR 176,-) annähernd auf dem gleichen Niveau, in der Kontrollgruppe aus dem Rhein-Sieg-Kreis mit EUR 170,- (95%-CI: EUR 124,- bis EUR 216,-) hingegen höher als in den beiden anderen Gruppen. Die Unterschiede an t1 sind nicht signifikant ($p=0,319$). Gleiches gilt für die paarweisen Vergleiche der Gruppen, zudem überlappen sich die Konfidenzintervalle deutlich.

Die ergänzend berechnete rangdatenbasierte Quade-ANCOVA zeigte ebenfalls keinen signifikanten Unterschied zwischen den drei Untersuchungsgruppen ($p=0,913$).

2.2.4.4 Ausschöpfung von Regelleistungen: Kosten für Tagespflege

Die EMMEANS für die Kosten für Tagespflege nach §41 SGB XI lagen an t0 bei EUR 14,-. Die Kosten der Interventionsgruppe sanken an t1 geringfügig auf EUR 12,- (95%-CI: EUR 0,- bis EUR 24,-). In den beiden Kontrollgruppen lagen die Kosten an t1 etwas höher als an t0, im Oberbergischer Kreis bei EUR 19,- (95%-CI: EUR 7,- bis EUR 31,-) und im Rhein-Sieg-Kreis bei EUR 15,- (95%-CI: EUR 3,- bis EUR 26,-). Die Unterschiede an t1 sind nicht signifikant ($p=0,691$). Gleiches gilt für die paarweisen Vergleiche der Gruppen, zudem überlappen sich die Konfidenzintervalle deutlich.

Die ergänzend berechnete rangdatenbasierte Quade-ANCOVA zeigte ebenfalls keinen signifikanten Unterschied zwischen den drei Untersuchungsgruppen ($p=0,613$).

2.2.4.5 Ausschöpfung von Regelleistungen: Kosten für zum Verbrauch bestimmte Pflegehilfsmittel

Für zum Verbrauch bestimmte Pflegehilfsmittel nach §40 SGB XI sowie der Produktgruppe 54 für Hilfsmittel nach §33 SGB V lagen die EMMEANS an

t0 bei EUR 77,-. An t1 lagen diese Kosten in allen drei Untersuchungsgruppen etwas niedriger. Im Oberbergischen Kreis lagen die Kosten sowohl in dortigen Kontroll- wie auch in der Interventionsgruppe bei EUR 74,- (95%-CI in beiden Gruppen EUR 53,- bis EUR 75,-) und in der Kontrollgruppe aus dem Rhein-Sieg-Kreis bei EUR 72,- (95%-CI: EUR 62,- bis EUR 84,-). Die Unterschiede an t1 sind nicht signifikant ($p=0,471$). Gleiches gilt für die paarweisen Vergleiche der Gruppen, zudem überlappen sich die Konfidenzintervalle deutlich.

Die ergänzend berechnete rangdatenbasierte Quade-ANCOVA zeigte ebenfalls keinen signifikanten Unterschied zwischen den drei Untersuchungsgruppen ($p=0,944$).

2.2.4.6 Ausschöpfung von Regelleistungen: Inanspruchnahme von Maßnahmen zur Verbesserung des Wohnumfelds

In der Interventionsgruppe nahmen 2,24% der Teilnehmenden im Interventionszeitraum (t1) Leistungen zur Verbesserung des Wohnumfelds nach §40 Ab. 4 SGB XI in Anspruch, bei denen dies in den drei Quartalen vor Interventionsstart (t0) nicht der Fall war. Der Anteil lag in der Kontrollgruppe aus dem Oberbergischen Kreis exakt gleich hoch (2,24%) und in der Kontrollgruppe aus dem Rhein-Sieg-Kreis geringfügig höher (3,79%). Die Event-Rate dieses Endpunkts ist in allen drei Gruppen niedrig, die Gruppenunterschiede sind nicht signifikant ($p=0,452$).

2.2.4.7 notfallbedingte Krankenhausbehandlungen

Die Event-Rate notfallbedingter Krankenhausaufnahmen in Interventionszeitraum lag in allen drei Untersuchungsgruppen hoch, jede sechste Person wies binnen der betrachteten 9 Monate eine solche als Notfall kodierte Aufnahme in ein Akutkrankenhaus auf. Die Interventionsgruppe wies dabei die höchste Event-Rate auf (18,45%), in den beiden Kontrollgruppen (Oberbergischer Kreis: 17,34%; Rhein-Sieg-Kreis: 16,61%) lag sie etwas niedriger. Die Gruppenunterschiede sind nicht signifikant ($p=0,197$).

2.2.4.8 Verbleib in der Häuslichkeit

Die Event-Rate der Personen, die im Interventionszeitraum nicht in der eigenen Häuslichkeit verbleiben konnten (die im Vergleich zu vor dem Interventionsstart Leistungen stationärer Dauerpflege neu in Anspruch nahmen) war in allen drei Untersuchungsgruppen gering. In der Interventionsgruppe lag sie mit 1,20% am niedrigsten, in der Kontrollgruppe aus dem Oberbergischen Kreis lag sie bei 2,41%, und in der Kontrollgruppe aus dem Rhein-Sieg-Kreis bei 1,61%. Die Gruppenunterschiede sind nicht signifikant ($p=0,578$).

2.2.4.9 Veränderung der Pflegebedürftigkeit

Die Event-Rate der Personen, bei denen es im Interventionszeitraum zu einem neu zuerkannten SGBI XI-Pflegegrad oder zu einem Anstieg des Pflegegrads kam, war in allen drei Untersuchungsgruppen hoch und belief sich in der Interventionsgruppe auf 18,08%. In den Vergleichsgruppen lag die Event-Rate etwas niedriger (Oberbergischer Kreis: 17,71%, Rhein-Sieg-Kreis: 14,02%). Die Gruppenunterschiede sind nicht signifikant ($p=0,373$).

2.2.4.10 Veränderung der Multimorbidität

Das EMMEAN des Exilhauser-Komorbiditätsindex lag an t0 bei 5,93 Punkten und erhöhte sich an t1 in der Interventionsgruppe auf 6,22 (95%-CI: 5,85 bis 6,59). Dies entspricht exakt den Werten der Vergleichsgruppe aus dem Oberbergischen Kreis. In der Vergleichsgruppe aus dem Rhein-Sieg-Kreis nahm die Punktzahl jedoch geringfügig ab auf 5,75 Punkte (95%-CI: 5,38 bis 6,12). Die Konfidenzintervalle der drei Gruppen überlappen sich, der Gruppeneffekt ist nicht signifikant ($p=0,132$).

Die ergänzend berechnete rangdatenbasierte Quade-ANCOVA zeigte ebenfalls keinen signifikanten Unterschied zwischen den drei Untersuchungsgruppen ($p=0,16$).

2.2.5 Kosten der Intervention

Die Fallkosten für den betrachteten 9-Monats-Zeitraum wurden mit EUR 1.727,75,- berechnet². In diese Kosten gingen die Personalkosten für die RCCM, die gegenüber den Leistungserbringern (vor allem den Hausärzten) vergüteten Versorgungsleistungen (Eingangsassessments, Versorgungsplanung, Fallkonsile, telemedizinisches Monitoring), die Kosten für die Erweiterte Erreichbarkeit und das telemedizinische Equipment ein. Unberücksichtigt blieben als einmalige Anlaufkosten die Kosten für die Kommunikationsplattform sowie die gezahlten Einschreibepremien.

Addiert man die Fallkosten für die Teilnehmenden (Interventionsgruppe) den GKV-Gesamtkosten für die betrachteten 9 Monate (siehe Abschnitt 2.2.4.1) hinzu, so wächst deren EMMEAN auf EUR 12.157 an (95%-CI: 11.199,- bis 13.116,-). Die Kosten für SGB V- und SGB XI-Leistungen liegen somit unter Einbeziehung der Fallkosten der Neuen Versorgungsform um EUR 1.261,- (gegenüber der KG OBK) bzw. EUR 1.443,- (gegenüber der KG RSK) höher. Dies entspräche EUR 150,- bzw. EUR 160,- pro Monat. Die Konfidenzintervalle der drei Untersuchungsgruppen überlappen sich, der p-Wert der ANCOVA liegt bei $p=0,076$ und ist nicht signifikant.

2 Details zu dieser Berechnung können im Evaluationsbericht zu OBERBERG_FAIRsorgt nachvollzogen werden, den der GbA-Innovationsfonds im Laufe des Jahres 2025 auf seiner Homepage des verfügbar machen wird.

3 Veränderung von gesundheitsbezogenen Outcomes nach einem Jahr Versorgung – standardisierte Befragung der Interventionsgruppe

3.1 Methodik

3.1.1 Studiendesign

Die Teilnehmenden wurden mittels Pen-and-Paper-Fragebögen an zwei Messzeitpunkten im Abstand von 13 Monaten standardisiert befragt. Die Baseline-Befragung (t0) erfolgte in den 14 Tagen vor Interventionsstart, die Follow-Up-Befragung wurde binnen 14 Tagen nach Abschluss der 12-monatigen Versorgung versandt.

3.1.2 Fragestellungen

Die folgenden Fragestellungen zu gesundheitsbezogenen Outcomes wurden im längsschnittlichen Verlauf betrachtet und die Outcomes folglich an beiden Messzeitpunkten erhoben:

- Unterscheidet sich in der Interventionsgruppe die gesundheitsbezogene Lebensqualität am Ende des Interventionszeitraums von jener zu Beginn der Intervention?
- Unterscheidet sich in der Interventionsgruppe die allgemeine Selbstwirksamkeitserwartung am Ende des Interventionszeitraums von jener zu Beginn der Intervention?
- Unterscheidet sich in der Interventionsgruppe die Resilienz am Ende des Interventionszeitraums von jener zu Beginn der Intervention?
- Unterscheidet sich in der Interventionsgruppe die körperliche Funktionsfähigkeit am Ende des Interventionszeitraums von jener zu Beginn der Intervention?
- Unterscheidet sich in der Interventionsgruppe die Erschöpfung am Ende des Interventionszeitraums von jener zu Beginn der Intervention?
- Unterscheiden sich in der Interventionsgruppe die depressive Stimmungslage am Ende des Interventionszeitraums von jener zu Beginn der Intervention?
- Unterscheidet sich in der Interventionsgruppe die Schmerzintensität am Ende des Interventionszeitraums von jener zu Beginn der Intervention?

- Unterscheidet sich in der Interventionsgruppe die Teilhabe an sozialen Rollen und Aktivitäten am Ende des Interventionszeitraums von jener zu Beginn der Intervention?
- Unterscheidet sich in der Interventionsgruppe die Schlafbeeinträchtigung am Ende des Interventionszeitraums von jener zu Beginn der Intervention?
- Unterscheidet sich in der Interventionsgruppe die subjektive Gesundheitskompetenz am Ende des Interventionszeitraums von jener zu Beginn der Intervention?

Neben den längsschnittlichen Effekten wurden in der Befragung an tl zudem querschnittliche Fragestellungen bezogen auf Outcomes (Zufriedenheit), Outputs (Inanspruchnahme von Leistungen) und Prozesse adressiert. Bezogen auf Outcomes werden in diesem Evaluationsbaustein folgende Fragestellungen untersucht:

- Wie zufrieden ist der Befragte mit der Neuen Versorgungsform allgemein?
- Wie zufrieden ist der Befragte mit den einzelnen Elementen der Neuen Versorgungsform?

Bezogen auf die Outputs werden in diesem Evaluationsbaustein folgende Fragestellungen untersucht:

- War der Befragte in die Versorgungsplanung einbezogen?
- Waren Bezugspersonen des Befragte in die Versorgungsplanung einbezogen?
- Wie häufig hatte der Befragte Kontakt zu seinem RCCM?
- Wie regelmäßig hatte der Befragte Kontakt zu seinem RCCM?
- Nutzt der Befragte die technische Plattform?
- Hat der Befragte nach einem Notruf (112) die erweiterte Erreichbarkeit genutzt?
- Hat der Befragte am einer telemedizinischen Fallkonferenz teilgenommen?
- Hat der Befragte das telemedizinische Monitoring genutzt?

Bezogen auf die Prozesse werden in diesem Evaluationsbaustein folgende Fragestellungen untersucht:

- Wie charakterisiert der Befragte die Beziehung zu seiner Fallmanagerin?
- In welcher Weise wurde der Befragte in die Versorgungsplanung einbezogen?

- Wie konsistent wird er Befragte in den Versorgungsprozess einbezogen?
- In wie weit hat der Befragte das Gefühl, seinen Versorgungsprozess aktiv zu steuern?
- Wie intensiv nutzt der Befragte die elektronische Plattform passiv (lesend)?
- Wie intensiv nutzt der Befragte die elektronische Plattform aktiv (Uploads, Kommunikation)?

3.1.3 Messinstrumente

Die gesundheitsbezogene Lebensqualität wurde mit dem EQ5D-5L (*Herdman et al. 2011*) gemessen, wobei sowohl der fünf Items umfassende Index als auch die visuelle Analogskala verwendet wurden. Die allgemeine Selbstwirksamkeitserwartung wurde mit der 10 Items umfassenden Skala von Schwarzer und Jerusalem (*Schwarzer & Jerusalem 1999*) gemessen, Resilienz mit der Kurzskala der 11 Items umfassenden Kurzform (RS-11) der Resilienzskala RS-11 (*Schumacher et al. 2005*). Zur Messung der subjektiven Gesundheitskompetenz wurde die Kurzform HLS-EU-Q16 des European Health Literacy Survey (*Röthlin et al. 2013*) mit 16 Items eingesetzt. Die übrigen Outcomes zum Gesundheitszustand wurden mit den jeweiligen Short Forms des Patient Reported Outcomes Measurement Information System (PROMIS-29) (Cella et al., 2010, Fisher et al., 2018) gemessen. Zur Messung der Polypharmazie, der Kenntnis der eignen Medikation, der Unterstützung durch Dienste und informelle Hilfen sowie der Inanspruchnahme von Therapeuten wurden selbstentwickelte dichotome Items verwendet.

Bei den nur an t1 erhobenen Items, die diese Fragestellungen adressierten, handelt es sich ebenfalls um selbst entwickelte Einzelitems.

3.1.4 Datenerhebung

Die Fragebögen wurden den Teilnehmenden durch die Koordinierungsstelle der Konsortialführung, die am Gesundheitsamt des Oberbergischen Kreises angesiedelt war, zugestellt. Dort wurden die Fragebögen auch mit dem jeweiligen Projektpseudonym versehen, so dass den Evaluatoren zu keinem Zeitpunkt Klarnamen oder Kontaktdaten der Teilnehmenden bekannt waren. Auch der Rücklauf wurde über diese Koordinierungsstelle abgewickelt.

3.1.5 Statistische Analyse

Der überwiegende Teil der Outcome-Variablen ist metrisch skaliert, Unterschiede zwischen den beiden Messzeitpunkten wurden mittels t-Tests für verbundene Stichproben auf einem 5% Signifikanzniveau 2-seitig getestet. Für dichotome Outcomes (Polypharmazie und medikamentenbezogenes Wissen) wurden exakte Tests nach Fisher eingesetzt, für die Levels der Gesundheitskompetenz der Wilcoxon-Vorzeichenrangtest. Die Daten zum Prozessgeschehen, zur Bewertung und Zufriedenheit wurden als Einzelsysteme deskriptiv ausgewertet.

3.2 Ergebnisse

3.2.1.1 Drop-Out Analyse

Von den N=271 Fällen, die im quantitativen Evaluationsbaustein A mit ihren Routinedaten in die Analyse eingehen konnten, lagen im quantitativen Evaluationsbaustein B von 186 Teilnehmenden Fragebögen zu beiden Messzeitpunkten vor. Dies entspricht 71,2% der in den drei Kohorten über den gesamten Interventionszeitraum versorgten Fälle. Zusätzlich gingen in die Stichprobe sieben der N=10 Fälle ein, die keiner der drei regulären Kohorten zugehörten, sondern der „Testkohorte“ K0, die unabhängig von Quartalsgrenzen mit Genehmigung des Projektträgers bereits vor der ersten Kohorte mit der Intervention startete. Ausgehend von den N=321 rekrutierten Fällen beträgt die Lost-to-Follow-Up-Rate für diesen Evaluationsbaustein 39,9%.

Der überwiegende Teil der Befragten war in der Lage, den Fragebogen komplett selbstständig zu bearbeiten (54,4% an t0, 59,1% an t1). Unterstützung durch eine Bezugsperson erhielten 31,1% an t0 bzw. 27,5% an t1, und in 14,0% (t0) bzw. 11,4% (t1) der Fälle antwortete eine Bezugsperson als Proxy. Bei einem Fall an t0 und vier Fällen an t1 fehlte eine entsprechende Angabe.

Die folgenden Angaben zu Unterschieden zwischen den Stichproben der quantitativen Evaluationsbausteine A und B beziehen sich auf zum einen auf die N=186 Fälle aus der Routinedatenanalyse, für die auch Fragebögen zu beiden Messzeitpunkten vorlagen (im Weiteren als „QUANT-B-Fälle bezeichnet) und zum anderen auf die N= 85 Fälle ohne analysierbaren Rücklauf beider Fragebögen (im Weiteren als „QUANT-B-Dropouts“ be-

zeichnet). Über die N=7 Fälle der „Testkohorte“ K0, für die Fragebogendaten vorliegen, können keine weitergehenden Aussagen getroffen werden, da für diese Fälle keine Routinedaten im quantitativen Evaluationsbaustein A angefordert wurden.

Hinsichtlich des Geschlechts bestehen keine Unterschiede zwischen den beiden Stichproben ($p=0,231$, Fisher's exakter Test). Die QUANT-B-Fälle waren mit $78,16 \pm 7,09$ Jahren aber signifikant jünger ($p<0,001$, t-Test für unabhängige Stichproben) als die QUANT-B-Dropouts mit $82,18 \pm 7,23$ Jahren.

Von den QUANT-B-Dropouts weisen 22,4% an t0 eine Demenzdiagnose auf, unter den QUANT-B-Fällen nur 2,2%. Auch dieser Unterschied ist signifikant ($p<0,001$, Fisher's exakter Test). Die QUANT-B-Fälle wiesen an t0 auch signifikant ($p<0,001$, Mann-Whitney-U-Test) geringere SGB XI-Pflegegrade auf. So haben 58,6% der QUANT-B-Fälle keinen Pflegegrad (gegenüber 28,2% der QUANT-B-Dropouts), und nur 1,6% wiesen einen Pflegegrad von 4 oder 5 auf (QUANT-B-Dropouts: 17,6%). Auch die Kosten für Kurzzeitpflege an t0 (drei Quartale vor Interventionsstart) lagen unter den QUANT-B-Dropouts (EUR 320,00 $\pm$ EUR 791,91) signifikant höher ($p=0,041$) als unter den QUANT-B-Fällen (EUR 123,03 $\pm$ EUR 560,02). Der Anteil an Fällen mit stationärer Dauerpflege an t0 lag unter den QUANT-B-Dropouts ebenfalls signifikant höher ($p>0,001$, Fisher's exakter Test) mit 23,5% gegenüber 1,1% der QUANT-B-Fälle.

Hinsichtlich der t0-Werte des Elixhauser-Komorbiditäts-Scores ($p=0,342$), der GKV- Gesamtkosten ($p=0,133$), der Krankenhauskosten ($p=0,856$), der Kosten für zum Verbrauch bestimmten Hilfsmittel ($p=0,186$) und der Kosten für Tagespflege ($p=0,390$, jeweils t-Test für unabhängige Stichproben) unterscheiden sich die beiden Stichproben hingegen nicht. Unter den QUANT-B-Fällen hatten drei Fälle in den drei Quartalen vor Interventionsstart (t0) Leistungen zur Anpassung des Wohnumfelds erhalten, unter den QUANT-B-Dropouts kam dies gar nicht vor. Dieser Unterschied ist aber nicht signifikant ($p=0,554$, Fisher's exakter Test).

3.2.2 Outcomes

3.2.2.1 Gesundheitsbezogene Lebensqualität

Die berichtete gesundheitsbezogene Lebensqualität (Wertebereich jeweils zwischen 0 und 100 Punkten, wobei höhere Werte eine höhere gesundheitsbezogene Lebensqualität anzeigen) lag an Baseline sowohl gemessen als

EQ5D-Index als auch mittels visueller Analogskala im mittleren Bereich (57,89 bzw. 53,92 Punkte). Im Follow-Up zeigten sich geringfügig niedrigere Werte (56,12 bzw. 52,55 Punkte), die aber nicht statistisch signifikant sind ($p=0,438$ bzw. $p=0,330$) und somit im Bereich erwartbarer statistischer Schwankungen liegen.

3.2.2.2 Allgemeine Selbstwirksamkeitserwartung

Die verwendete Skala zur Messung der allgemeinen Selbstwirksamkeitserwartung hat einen Wertebereich von 10 bis 40 Punkten, höhere Scores zeigen eine höhere allgemeine Selbstwirksamkeitserwartung an. An Baseline lag der mittlere Score der Stichprobe bei 29,88 Punkten, im Follow-Up (29,52 Punkte) veränderte er sich nur geringfügig. Der Gruppenunterschied ist nicht signifikant ($p=0,598$). Der Score entspricht fast exakt dem arithmetischen Mittel der Bevölkerungsnorm (Hinz et al. 2006) für Männer zwischen 61 und 95 Jahren (29,8) und liegt über jenem für Frauen der gleichen Altersgruppe (27,6).

3.2.2.3 Resilienz

Der Wertebereich der Resilienzskala RS-11 erstreckt sich von 0 bis 100 Punkten, wobei höhere Werte eine höhere Resilienz anzeigen. Der an Baseline gemessene mittlere Score für die Resilienz der Teilnehmenden lag bei 51,58 Punkten. Im Follow-Up lag der Score geringfügig um ca. einen Punkt niedriger (50,50 Punkte), dieser Unterschied ist nicht signifikant ($p=0,411$). Einen Score von maximal 52 Punkten weisen in der Gesamtbevölkerung 30% der 14-95-Jährigen auf (Schumacher et al. 2005), das Mittel der Stichprobe liegt hinsichtlich ihrer Resilienz also auf dem Niveau des unteren Drittels der Bevölkerung.

3.2.2.4 Gesundheitskompetenz

Der Wertebereich des HLS-EU-Q16 liegt zwischen 0 und 16 Punkten, wobei höhere Werte eine höhere Gesundheitskompetenz anzeigen. An Baseline lag der mittlere Score bei 10,29, im Follow-Up etwas höher bei 10,55 Punkten. Der Unterschied ist statistisch nicht signifikant ($p=0,434$). Aus der

Index-Wert des Instruments lassen sich drei Levels der Gesundheitskompetenz ableiten (Röthlin et al. 2013). Bezogen auf die N=144 Befragten, von denen Scores für beide Messzeitpunkte vorlagen, wiesen an Baseline 36,1% eine ausreichende Gesundheitskompetenz auf, im Follow-Up waren es 39,6%. Eine problematische Gesundheitskompetenz zeigten Baseline 28,5%, im Follow-Up 25,7%. Auf dem niedrigsten Level, dem der inadäquaten Gesundheitskompetenz, befanden sich an Baseline 35,4% der Befragten und im Follow-Up 34,7%. Auch der Unterschied zwischen den Levels ist statistisch nicht signifikant ($p=0,474$). Verglichen mit der Bevölkerungsnorm (Jordan & Hoebel 2015) für die Levels auf Basis des HLS-EU-Q16 für die Altersgruppe 60 und älter (54,7% ausreichend, 32,8% problematisch, 12,5% inadäquat) ist die Gesundheitskompetenz der Stichprobe deutlich niedriger.

3.2.2.5 körperliche Funktionsfähigkeit

Die so genannten T-Scores der PROMIS-Skalen haben einen Wertebereich von 0 bis 100. Höhere Punkte zeigen dabei eine höhere Ausprägung des jeweils gemessenen Konstrukts an. Die Scores sind so normiert, dass der Durchschnitt der Referenzpopulation bei 50 Punkten liegt und 10 Punkte jeweils eine Standardabweichung bedeuten. Dies gilt auch für die in den folgenden Abschnitten berichteten Outcomes. Für die körperliche Funktionsfähigkeit, gemessen mit der Short Form PROMIS SF v2.0 – Physical Function 4a, lag an Baseline im Mittel bei 38,15 von 100 Punkten, was einer moderaten Beeinträchtigung (30-40 Punkte) entspricht. Im Follow-Up reduzierte sie sich um weniger als einen Punkt auf 37,15 Punkte. Dieser Unterschied ist statistisch nicht signifikant ($p=0,161$).

3.2.2.6 Erschöpfung

Die selbst wahrgenommene Erschöpfung bei alltäglichen Verrichtungen wurde mit der Short Form PROMIS SF v1.0 – Fatigue 4a gemessen. An Baseline lag der mittlere Punktwert bei 56,21, was einer milden Beeinträchtigung (55-60 Punkte) entspricht. Im Follow-Up bei 57,93. Dieser Unterschied ist statistisch signifikant ($p=0,032$). Wir bewerten diese Zunahme an Erschöpfung um 1,72 Punkte aber als nicht klinisch bedeutsam, zumal beide Scores im Bereich der „milden“ Beeinträchtigung liegen.

3.2.2.7 depressive Stimmungslage

Durch die Teilnehmer berichtete depressive Stimmungslagen wurden mit der Short Form PROMIS SF v1.0 – Depression 4a gemessen. Der Wertebereich liegt zwischen 0 und 100 Punkten, wobei höhere Werte eine depressivere Stimmungslage anzeigen. An Baseline lag der mittlere Punktwert bei 54,87, im Follow-Up bei 55,15. In der Logik der PROMIS-Skalen liegt der t0-Wert gerade noch im normalen Bereich (unter 55 Punkten), während der t1-Wert eine milde Beeinträchtigung (55-60 Punkte) darstellt. Dieser Unterschied ist statistisch aber nicht signifikant ($p=0,671$).

3.2.2.8 Beeinträchtigung durch Schmerzen und Schmerzintensität

Die subjektiv erlebte Schmerzintensität wurde mittels eines zum PROMIS-29 v2.1 Profil gehörenden Einzelitem („Global07“) gemessen. Es handelt sich um eine numerische Ratingskala mit einem Wertebereich zwischen 0 (keine Schmerzen) und 10 (schlimmste vorstellbare Schmerzen). An Baseline lag die im Mittel berichtete Schmerzintensität bei 4,96, im Follow-Up wurde im Mittel eine etwas geringere Schmerzintensität (4,84 Punkte) berichtet. Dieser Unterschied ist statistisch nicht signifikant ($p=0,480$). Die Beeinträchtigung durch Schmerzen lag an t0 bei 61,22 Punkten und an t1 bei 62,23 Punkten. Beide Werte liegen im Bereich einer moderaten Beeinträchtigung. Der Unterschied ist nicht signifikant ($p=0,068$).

3.2.2.9 Teilhabe an sozialen Rollen und Aktivitäten

Die Teilhabe an sozialen Rollen und Aktivitäten wurde mittels der Short Form PROMIS SF v1.0 – Ability to Participate in Social Roles and Activities 4a gemessen, der Wertebereich liegt zwischen 0 und 100 Punkten. An Baseline lag der mittlere Punktwert bei 41,51 Punkten, im Follow-Up bei 40,34 Punkten. Dieser Unterschied ist statistisch nicht signifikant ($p=0,087$). Beide Scores entsprechen einer milden Beeinträchtigung (40-45 Punkte).

3.2.2.10 Schlafbeeinträchtigung

Schlafbeeinträchtigung wurde mittels der Short Form PROMIS SF v1.0 – Sleep Disturbance 4a gemessen, die einen Wertebereich zwischen 0 und 100 Punkten umfasst. An Baseline lag der mittlere Score bei 53,55 Punkten, im Follow-Up fast unverändert bei 53,48 Punkten. Dieser Unterschied ist statistisch nicht signifikant ($p=0,917$). Beide Scores sind als normal („within normal limits“) zu interpretieren.

3.2.2.11 Polypharmazie

Bezogen auf die $N=184$ Teilnehmenden, von denen für beide Messzeitpunkte Daten vorlagen, lag der Anteil von Personen mit 5 oder mehr Medikamenten an Baseline bei 74,5% und im Follow-Up bei 75,5%. Dieser Unterschied ist statistisch nicht signifikant ($p=0,860$). 91,20% der Befragten gaben an Baseline an, für jedes ihrer Medikamente zu wissen, woran sie es einnehmen. Im Follow-Up lag der Anteil bei 90,7%, dieser Unterschied ist statistisch ebenfalls nicht signifikant ($p=1,000$).

3.2.2.12 Inanspruchnahme von Ärzten, anderen Leistungserbringern und informellen Hilfen

An Baseline gaben die Befragten im Mittel an, bei 3,38 Ärzten zeitgleich in Behandlung zu sein. Dieser Wert blieb im Follow-Up fast unverändert bei 3,41. Der Unterschied ist statistisch nicht signifikant ($p=0,751$).

Auch für andere professionelle Leistungserbringer und informelle Hilfen zeigten sich keine signifikanten Veränderungen von t_0 zu t_1 , mit einer Ausnahme: Einen Lieferdienst für Lebensmittel (z.B. durch einen Supermarkt) nahmen an t_1 neun Personen (5,3%) in Anspruch, an t_0 war dies nur bei einem Teilnehmenden (0,6%) der Fall. Der Unterschied ist signifikant ($p=0,008$). Auffällige Veränderungen zeigten sich auch bei der Inanspruchnahme von Lieferungen zubereiteter Lebensmittel, hier kamen zu einer Person an t_0 (0,6%) sechs weitere an t_1 hinzu. Der Anteil steigt auf 4,7%, der Unterschied war aber nicht signifikant ($p=0,070$). Ein deutlicher Zuwachs zeigte sich auch bei der informellen Unterstützung durch nicht im Haushalt lebende Personen. Der Anteil der Teilnehmenden, die auf diese Weise unterstützt wurden, erhöhte sich von 58 Personen (32,6%) auf 67 Personen

(37,6%), wobei 10 Personen diese Unterstützung an t1 nicht mehr erhielten, sie bei 19 Personen aber neu hinzukam. Die Nutzung von Physiotherapie lag bereits an t0 mit 22,2% in einem eher hohen Bereich und erhöhte sich zu t1 auf 27,5%, wobei 24 Personen diese Leistung neu in Anspruch nahmen und sie bei 15 Personen im Vergleich zu t0 wegfiel. Diese Veränderungen waren aber nicht signifikant (0,200). Auf niedrigem Niveau zeigte sich auf eine Veränderung bei der Nutzung von Diätassistenten und Ernährungsberatern. An t0 nutzten zwei Personen (1,2%) diese Leistung, bei beiden war dies an t1 nicht mehr der Fall. Im Gegenzug kamen sechs neue Teilnehmende (3,6%) hinzu, die diese Leistung nutzten. Der Unterschied war nicht signifikant ($p=0,289$). Leistungen eines Podologen nahmen an t0 17,4% der Teilnehmenden in Anspruch, an t1 sank der Anteil auf 13,8%. Während acht Personen diese Leistung neu in Anspruch nahmen, fiel sie bei 14 Personen weg. Auch dieser Unterschied war nicht signifikant ($p=0,286$). Bei allen übrigen Leistungserbringern und informellen Hilfen zeigten sich keine wesentlichen Veränderungen.

3.2.2.13 Zufriedenheit

Die globale, auf einer Schulnoten-Skala von 1-6 erfragte Zufriedenheit mit den Leistungen der Versorgungsform insgesamt lag im Mittel bei $1,9 \pm 1,1$, wobei nur 3,0% der Befragten die Noten 5 oder 6 vergaben und 44,6% ein „sehr gut“.

Ähnlich wurde die Zufriedenheit mit der durch die Fallmanager erhaltenen Unterstützung bewertet, sie lag im Mittel bei $1,83 \pm 1,1$.

Die Zufriedenheit mit der elektronischen Plattform wurde in Ermangelung eigener Erfahrungen von der Mehrzahl der Befragten nicht beantwortet ($N=49$), die Zufriedenheit lag hier lediglich bei einer Durchschnittsnote von $3,53 \pm 1,72$, wobei jeder Fünfte (24,5%) ein „ungenügend“ vergab und weitere 4,1% die Note 5. Umgekehrt vergaben 34,7% die Noten 1 („sehr gut“) oder 2, das Meinungsbild hinsichtlich der Zufriedenheit mit der Plattform kann als sehr heterogen bezeichnet werden.

Jene Befragten, die die Teilleistung des telemedizinischen Monitorings in Anspruch genommen hatten, bewerteten ihre Zufriedenheit mit der Nutzung der telemedizinischen Geräte mit Mittel mit der Note $1,81 \pm 1,31$, wobei 56,6% ihre Zufriedenheit mit der Note „sehr gut“ und weitere 25,9% die Note 2 ausdrückten. 7,1% vergaben eine 5 oder ein „ungenügend“.

N=39 Personen bewerteten, ebenfalls auf einer Schulnotenskala, ihre Zufriedenheit mit der Hilfe, die sie bei einem Anruf im Rahmen der Teilleistung „erweiterte Erreichbarkeit“ erhalten hatten. Die Durchschnittsnote lag bei $2,38 \pm 1,58$. Mehr als zwei Drittel der Befragten vergaben die Noten „sehr gut“ oder 2, andererseits zeigten sich auch 15,4%, die die Noten 5 oder „ungenügend“ vergaben, unzufrieden.

93,9% der Befragten stimmten der Aussage, dass das Angebot von OBERBERG_FAIRsorgt unbedingt weiterbestehen solle, tendenziell zu (82,7%: stimmt genau). Ein ähnlich hoher Anteil bejahte die Aussage, dass die Versorgungsform auch Versicherten anderer Krankenkassen offenstehen sollte (92,8%, davon 76,9% stimmt genau). Im Follow-Up gaben zwei Drittel der Befragten an, dass sich durch die Versorgungsform ihre Gesundheitsversorgung verbessert habe (jeweils ca. ein Drittel „stimmt eher“ und „stimmt genau“. Kategorisch verneint („stimmt nicht“) wurde diese Aussage von 12,4%. Ein Drittel der Befragten befand, dass sie ohne die Neue Versorgungsform im Zeitraum der Versorgung nicht so hätten weiter wohnen können wie zuvor (Teil II Tabelle 3).

Teil II Tabelle 3: globale Beurteilung der Versorgungsform

	N	stimmt nicht	stimmt kaum	stimmt eher	stimmt genau	Median	IQR
Das Angebot von OBERBERG_FAIRsorgt sollte unbedingt weiter bestehen.	196	3,1%	3,1%	11,2%	82,7%	3 (stimmt genau)	0
Ein Angebot wie OBERBERG_FAIRsorgt sollte auch Menschen offenstehen, die nicht bei der AOK krankenversichert sind.	195	3,1%	4,1%	15,9%	76,9%	3 (stimmt genau)	0
Durch OBERBERG_FAIRsorgt hat sich die Qualität meiner Gesundheitsversorgung verbessert.	185	12,4%	18,4%	33,0%	36,2%	2 (stimmt eher)	2
Ohne OBERBERG_FAIRsorgt hätte ich im letzten Jahr nicht weiter so wohnen können wie zuvor.	184	51,6%	14,1%	15,8%	18,5%	0 (stimmt nicht)	2

Die Wichtigkeit, die die Befragten den einzelnen Teilleistungen der Versorgungsform beimessen, wurde auf einer Skala von 0 (überflüssig) bis 4 (unverzichtbar) gemessen. Den höchsten Mittelwert (3,4 von 4) erreichte dabei der Aspekt, dass alle an der eigenen Gesundheitsversorgung Beteiligten zusammenarbeiten und einen gemeinsamen Versorgungsplan erstellen. Auch die erweiterte Erreichbarkeit, also die Möglichkeit, zu jeder Zeit einen Arzt telefonisch erreichen zu können, wurde mit einem Mittelwert größer als 3 bewertet. Diese beiden Teilleistungen bewerteten über 60% der Befragten als unverzichtbar. Gut die Hälfte der Befragten sah die Unterstützung durch einen Fallmanager als unverzichtbar an (Mittelwert: 2,98). Auch der regelmäßige Austausch aller an der eigenen Gesundheitsversorgung Beteiligter (interdisziplinäres Fallkonsil), die gemeinsame Nutzung gesundheits- bzw. versorgungsrelevanter Daten auf einer gemeinsamen elektronischen Plattform sowie ein telemedizinisches Monitoring in der eigenen Häuslichkeit wurde mit Mittelwerten zwischen 2,9 und 2,7 als wichtig erachtet (Teil II Tabelle 4).

Teil II Tabelle 4: Wichtigkeit der einzelnen Teilleistungen

„Wie wichtig ist es Ihnen, dass...“	N	überflüssig	1	2	3	unverzichtbar	MW	SD
Sie von einer Fallmanagerin/ einem Fallmanager unterstützt werden?	184	4,9%	12,4%	13,0%	18,9%	50,8%	2,98	1,26
alle an Ihrer Gesundheitsversorgung Beteiligten (Ärzte, Fallmanager, ggf. Pflegedienst, etc.) zusammenarbeiten und gemeinsam einen Versorgungsplan für Sie erstellen?	74	4,1%	2,7%	5,4%	25,7%	62,2%	3,4	1,00
sich alle an Ihrer Gesundheitsversorgung Beteiligten regelmäßig untereinander zu Ihrer Versorgung austauschen?	179	4,5%	11,7%	14,5%	25,1%	44,1%	2,9	1,21
Informationen zu Ihrer Gesundheit allen an Ihrer Gesundheitsversorgung Beteiligten auf	170	7,6%	10,6%	12,9%	31,2%	37,6%	2,8	1,26

einer gemeinsam genutzten elektronischen Plattform zur Verfügung stehen?								
die Möglichkeit besteht, über eine besondere Telefonnummer in Notfällen immer eine Ärztin oder einen Arzt erreichen zu können („erweiterte Erreichbarkeit“)?	184	5,4%	9,8%	6,0%	17,9%	60,9%	3,2	1,23
die Möglichkeit besteht, dass Sie zu Hause selbst z.B. Blutdruck, Puls, Blutzucker, Körpergewicht oder ein EKG messen können und diese Daten automatisch elektronisch zu Ihrem Arzt übertragen werden?	168	7,7%	17,3%	11,9%	25,0%	38,1%	2,7	1,34

3.2.3 Outputs und Prozesse

3.2.3.1 Rekrutierungswege

In der Baseline-Befragung wurden die Teilnehmenden gebeten, darüber Auskunft zu geben, auf welchen Wegen sie von dem Angebot der neuen Versorgungsform erfahren hatten. Dabei waren Mehrfachnennungen möglich. Als der mit weitem Abstand dominante Rekrutierungsweg stellte sich dabei die Information durch den Hausarzt heraus (82,1% der Fälle). Informationen durch Familie oder Bekannte (8,4%) spielten ebenfalls eine, allerdings untergeordnete, Rolle. Seltener erfuhrten die Befragten über ihren Pflegedienst (4,0%) oder über die Medien (2,6%) von der Versorgungsform, nur in einem Fall während eines Krankenhausaufenthalts. 11,0% der Befragten gaben sonstige Kanäle an. 5,9% gaben dabei an, angeschrieben worden zu sein, wobei der Großteil dieser Befragten explizit Anschreiben der Krankenkasse nennt. Vier Personen gaben an, direkt über das Team von OBERBERG_FAIRsorgt von dem Angebot erfahren zu haben, weitere Einzelnennungen waren die Senioren- und Pflegeberatung, die Stadtverwaltung, ein „Informationsstand auf dem Markt“, das eigene Pflegeheim sowie allgemein „Reklame“.

3.2.3.2 Zusammenarbeit mit dem RCCM

Der Aussage, durch den Fallmanager vor allem zu Beginn der Teilnahme unterstützt worden zu sein, stimmten 82,7% der Befragten tendenziell zu (53,9%: stimmt genau, Median: stimmt genau). Zugleich stimmten der Aussage, in den letzten Monaten der Versorgung mehr Unterstützung durch den Fallmanager benötigt zu haben, 73,2% der Befragten tendenziell nicht zu (54,2%: stimmt nicht, Median: stimmt nicht). Gefragt nach der Häufigkeit der Kontakte zum Fallmanager, gab für jedes der vier Quartale der einjährigen Versorgung eine deutliche Mehrheit (zwischen 57,5% und 69,5%) an, dass es 1-5 Kontakte gegeben habe. Für die jeweiligen Quartale 2-4 antwortete jeweils mehr als jeder Fünfte, dass im betreffenden Quartal gar kein Kontakt zum Fallmanager stattgefunden habe (Teil II Tabelle 5).

Teil II Tabelle 5: Häufigkeit der Kontakte zum Fallmanager nach Quartalen des eigenen Interventionszeitraums

	N	>10x	6-10x	1-5x	gar nicht	Median	IQR
in den ersten 3 Monaten	167	6,0%	15,6%	69,5%	9,0%	2 (1-5x)	0
in den Monaten 4-6	159	0,6%	13,2%	64,2%	22,0%	2 (1-5x)	0
in den Monaten 7-9	155	2,6%	12,9%	60,0%	24,5%	2 (1-5x)	0
in den Monaten 10-12	174	8,0%	11,5%	57,5%	23,0%	2 (1-5x)	0

Danach gefragt, ob es ihnen gelungen sei, zu ihrem Fallmanager ein Vertrauensverhältnis aufzubauen, antworteten 55,7% der Befragten mit „stimmt genau“ und 24,6% mit „stimmt eher“. 8,2% verneinten dies kategorisch („stimmt nicht“). 85,8% bejahten tendenziell die Aussage, dass sie sich stets mit allen Fragen an ihren Fallmanager wenden können (72,1% stimmt genau). 83,7% der Befragten stimmten tendenziell der Aussage zu, dass der Fallmanager im Nachhinein für sie das Wichtigste an der neuen Versorgungsform gewesen sei (70,4% stimmt genau).

15,9% der Befragten gab in der Follow-Up-Befragung an, dass es im Versorgungszeitraum zu einem Wechsel des zuständigen Fallmanagers gekommen sei. Zugleich bejahten tendenziell 85,8% der Befragten (77,5% stimmt genau, Median: stimmt genau), dass es für sie wichtig sei, als Fallmanager immer den gleichen Ansprechpartner zu haben.

3.2.3.3 Einbeziehung in Versorgungsplanung und Steuerung des Versorgungsprozesses

Der Aussage, dass ein Versorgungsplan erstellt und dieser mit ihnen besprochen worden sei, stimmten lediglich 59,5% der Befragten zu (43,8%: stimmt genau, Median: „stimmt eher“), 28,1% verneinten diese Aussage kategorisch („stimmt nicht“). Der Anteil der Befragten, die tendenziell der Aussage zustimmen, sich in alle Planungen die eigene Gesundheitsversorgung betreffend einbezogen gefühlt zu haben, lag mit 70,6% höher (48,9% „stimmt genau“). Diese Aussage wurde aber auch von 18,9% der Befragten strikt („stimmt nicht“) verneint. Knapp jeder Zweite gab an, dass ihm manchmal nicht ganz klar gewesen sei, was zwischen Arzt und Fallmanager besprochen worden sei (28,7% „stimmt genau“, 17,8% „stimmt eher“). Hingegen bejahten 88,2% die Aussage, letztendlich alle wichtigen, die Gesundheitsversorgung betreffenden, Entscheidungen selbst zu treffen (72,7% „stimmt genau“, Teil II Tabelle 6).

Teil II Tabelle 6: Einbeziehung in den Versorgungsprozess

	N	stimmt nicht	stimmt kaum	stimmt eher	stimmt genau	Median	IQR
Es wurde ein Versorgungsplan für mich erstellt, der mit mir besprochen wurde.	178	28,1%	12,4%	15,7%	43,8%	2 (stimmt eher)	3
Ich habe mich in alle Planungen, die meine Gesundheitsversorgung betreffen, einbezogen gefühlt.	180	18,9%	10,6%	21,7%	48,9%	2 (stimmt eher)	2
Am Ende treffe ich selbst alle wichtigen Entscheidungen, die meine Gesundheitsversorgung betreffen.	186	6,5%	5,4%	17,2%	71,0%	3 (stimmt genau)	1
Manchmal war mir nicht ganz klar, was meine Fallmanagerin/ mein Fallmanager und mein Arzt/ meine Ärztin zusammen besprochen haben.	174	32,2%	21,3%	17,8%	28,7%	1 (stimmt kaum)	3

3.2.3.4 Nutzung der elektronischen Plattform

Nur 10 der der Befragten gaben an, während des Versorgungszeitraums jemals selbst auf die elektronische Plattform zugegriffen haben. Zur Zufriedenheit mit der Plattform äußerten sich dennoch N=49 Teilnehmende (was keinen Widerspruch darstellt). Auf einer Schulnotenskala bewerteten diese ihre Zufriedenheit mit der Kommunikationsplattform mit der Note $3,53 \pm 1,72$. Die Fragen, wie oft der Teilnehmende selbst oder einer der Angehörigen lesend bzw. schreibend auf die Plattform zugegriffen haben, beantworteten 81% (lesend, N=100) bzw. 86,7% (schreibend, N=98) mit „nie“. Neun Personen gaben hier aber lesende Zugriffe mindestens 1x/Monat an, für schreibende Zugriffe war dies bei lediglich drei Personen der Fall.

3.2.3.5 Erweiterte Erreichbarkeit

14,5% der Befragten gaben an, während des Versorgungszeitraums die Teilleistung der erweiterten Erreichbarkeit genutzt, also die entsprechende besondere Rufnummer angerufen zu haben. Die Möglichkeit, diese Leistung zu nutzen, war einem großen Teil der Befragten bewusst, 82,4% stimmten dieser Aussage voll oder eher zu. Ein knappes Viertel der Befragten gab allerdings an, die entsprechende Telefonnummer nicht immer griffbereit gehabt zu haben. Zwei Drittel der Befragten stimmte der Aussage, dass die erweiterte Erreichbarkeit ihnen ein zusätzliches Sicherheitsgefühl gegeben habe, voll zu, weitere 10,1% stimmten dieser Aussage eher zu (Teil II Tabelle 7).

Teil II Tabelle 7: Prozesse und Outputs: erweiterte Erreichbarkeit

	N	stimmt nicht	stimmt kaum	stimmt eher	stimmt genau	Median	IQR
Mir war bewusst, dass ich die „erweiterte Erreichbarkeit“ nutzen konnte.	188	8,0%	9,6%	8,5%	73,9%	3 (stimmt genau)	1
Ich hatte die Telefonnummer immer griffbereit.	189	13,2%	11,6%	15,3%	59,8%	3 (stimmt genau)	1,5

Die Möglichkeit, auch außerhalb der Sprechzeiten einen Arzt zu erreichen, hat mir ein zusätzliches Gefühl von Sicherheit gegeben.	189	12,7%	9,5%	10,1%	67,7%	3 (stimmt genau)	1
---	-----	-------	------	-------	-------	---------------------	---

3.2.3.6 Telemedizinisches Monitoring

Von jenen Befragten, die die Teilleistung des telemedizinischen Monitorings in Anspruch genommen hatten, gab eine große Mehrheit an, mit der Nutzung der ihnen zu diesem Zweck zur Verfügung gestellten Geräte keine Probleme gehabt zu haben, nachdem sie in der Handhabung unterwiesen worden waren (81,6% „stimmt genau“, 7,1% „stimmt eher“). Drei Viertel dieser Befragten sahen es positiv, dass ihnen die Nutzung dieser Geräte Besuche bei ihrem Arzt erspart habe (58,5% „stimmt genau“, 18,1% „stimmt eher“). Zu einer Veränderung der medizinischen Behandlung kam es in der Folge der Nutzung des telemedizinischen Monitorings nach Angaben der Teilnehmenden in etwas mehr als jedem dritten Fall (31,9% „stimmt genau“, 4,3% „stimmt eher“. Allerdings gaben nur zwei Drittel der Befragten an, dass ihr Arzt mit ihnen über die gemessenen Werte gesprochen habe, das übrige Drittel verneinte dies tendenziell (25,5% „stimmt nicht“, 9,6% „stimmt kaum“, Teil II Tabelle 8).

Teil II Tabelle 8: Outputs und Prozesse: Telemedizinisches Monitoring

	N	stimmt nicht	stimmt kaum	stimmt eher	stimmt genau	Median	IQR
Ich finde es gut, dass die Nutzung der Geräte mir Besuche beim meinem Arzt/ meiner Ärztin erspart hat.	94	13,8%	9,6%	18,1%	58,5%	3 (stimmt genau)	0
Mein Arzt/ meine Ärztin hat mit mir über meine gemessenen Werte gesprochen.	94	25,5%	9,6%	12,8%	52,1%	3 (stimmt genau)	0
Nach dem Messen meiner Werte hat mein Arzt/ meine Ärztin etwas an mei-	94	58,5%	5,3%	4,3%	31,9%	0 (stimmt nicht)	3

ner Behandlung verändert (z.B. die Medikamente)							
Nachdem mir die Handhabung des Geräts/ der Geräte erklärt wurde, hatte ich keine Probleme mit der Nutzung.	98	6,1%	5,1%	7,1%	81,6%	3 (stimmt genau)	0

4 Diskussion

Der quantitative Teil der Evaluation weist einige offensichtliche Limitationen auf. So war es nicht möglich, ein randomisiert-kontrolliertes Studiendesign umzusetzen. Ein solches hätte den Zugang zur neuen Versorgungsform im Oberbergischen Kreis zu stark limitiert und wäre auch ethisch schwierig zu vertreten gewesen, weil unabhängig vom Aspekt der Wirksamkeit der Versorgungsform ein Mehr an Versorgungsleistungen (Nutzung eines Case-Management mit kontinuierlichem Ansprechpartner, Zugang zu erweiterter ärztlicher Erreichbarkeit sowie zu telemedizinischen Leistungen) angeboten wurde. Durch das gewählte Propensity Score Matching, durch welches zwei Kontrollgruppen generiert werden konnten, die sich zu Baseline in keiner einzigen der zahlreichen Matching-Kovariate von der Interventionsgruppe unterschieden, konnte diese Limitation bestmöglich kompensiert werden. Stärker ins Gewicht fällt die Limitation der fehlenden Kontrollgruppen bei der Messung der gesundheitsbezogenen Outcomes in der schriftlichen Befragung. Um hier ein kontrolliertes Design umzusetzen, wäre aber ein erheblicher Aufwand zur Rekrutierung erforderlich gewesen, der die Ressourcen der Evaluation bei weitem überschritten hätte.

Generell waren für eine Versorgungsform, die im Wesentlichen präventive Ziele verfolgt (Verbleib in der Häuslichkeit, Vermeidung von notfallbedingten Krankenhausaufenthalten u.a.), sowohl der Interventions- als auch der Beobachtungszeitraum für die Evaluation zu kurzfristig ausgestaltet. Dieses Grundproblem teilt OBERBERG_FAIRsorgt mit vielen anderen durch den Innovationsfonds geförderten Projekten. Insbesondere zum Nachweis von Effekten auf die Hospitalisierung wäre neben einem längeren Beobachtungszeitraum auch ein größerer Stichprobenumfang erforderlich gewesen, da aufgrund des kurzen Beobachtungszeitraums die Event-Rate erwartungskonform sehr niedrig war.

Die schließlich rekrutierte Stichprobe ist in ihrer Aussagekraft ebenfalls limitiert. Zum einen beteiligte sich nur eine recht geringe Anzahl an Hausärzten am Selektivvertrag, davon mehrere auch nur mit vereinzelt Fällen. Die rekrutierten Fälle weisen daher eine sehr unausgewogene regionale Verteilung auf, mit einem deutlichen Schwerpunkt auf der Kommune Bergneustadt. Hinzu kommt, dass einige wenige Hausärzte große Teile der rekrutierten Fälle versorgten. Die rekrutierte Stichprobe utzt die Einschlusskriterien bezogen auf wenig komplexe Fälle kaum aus, die Stichprobe ist älter, pflegebedürftiger, multimorbider und auch in höherem Maße von prekären Versorgungssituationen gekennzeichnet, als dies aufgrund der

Einschlusskriterien zu erwarten gewesen wäre. Die Hausarztpraxen legten bei der Rekrutierung, zumindest zu Beginn, einen starken Fokus auf Fälle, die aus ihrer Sicht aufgrund ihrer Bedarfslagen besonders von der neuen Versorgungsform profitieren könnten. Für den Workload der RCCM war dies eher ungünstig, Empfehlungen für einen konkreten Schlüssel für Fälle pro RCCM sind so nicht möglich. Die hohe Komplexität der rekrutierten Stichprobe wirkte sich auch auf die standardisierte Befragung im quantitativen Evaluationsbaustein B aus, die ein erheblicher Anteil der Teilnehmenden nicht ohne Unterstützung selbstständig bewältigen konnte.

Die Leistungserbringung der Neuen Versorgungsform erfolgte nicht in allen Punkten so, wie die Leistungsbeschreibung dies vorsah. So wurde ein Teil der Eingangsassessments nicht durch die Hausärzte, sondern durch die RCCM vorgenommen. Fallkonsile konnten nur vereinzelt durchgeführt werden. Beides hat seinen Grund darin, dass die Arztpraxen, auch beeinflusst durch die COVID-19-Pandemie, von deren Beteiligung die Umsetzung dieser Teilleitungen abhing, keine ausreichenden Ressourcen zur Verfügung stellen konnten oder wollten. Die verspätete Verfügbarkeit der telemedizinischen Teilleistung erschwerte es, für diese Teilleistung aussagekräftige Befunde zu generieren bzw. deren Anteil an den Fallkosten präziser abzubilden.

Die in der Routinedatenanalyse berechneten ANCOVAs erfüllten aufgrund starken Kostenvarianz zum Teil die Voraussetzungen für diese statistische Analyse nicht, weshalb die Punkt- und Bereichsschätzer verzerrt sein können und mit Vorsicht zu interpretieren sind. Die grundsätzliche Aussage zur jeweiligen Hypothese wurde aber mittels Quade-ANCOVAs abgesichert. Eingedenk dieser Limitationen lassen sich die folgenden Schlussfolgerungen für den quantitativen Teil der Evaluation von OBERBERG_FAIRsorgt ziehen:

Die Wirtschaftlichkeit der Neuen Versorgungsform OBERBERG_FAIRsorgt erscheint auf Basis der Evaluationsergebnisse gegeben zu sein. Die Gesamtkosten für Ausgaben der gesetzlichen Kranken- und Pflegeversicherung (SGB V- und SGB XI-Kosten) unterscheiden sich nicht im betrachteten 9-monatigen Interventionszeitraum nicht signifikant von jenen der beiden Kontrollgruppen. Dies gilt auch unter Einbeziehung der laufenden Fallkosten (Personalkosten der RCCM, Vergütung von Versorgungsleistungen, Abschreibung von Anschaffungskosten für telemedizinisches Equipment), wenngleich hier statistische Unsicherheit besteht. Berücksichtigt man das Mehr als Leistungen, die den Teilnehmenden zur Verfügung standen (Case-Management mit Lotsenfunktion, vielschichtigen

Beratungen, Kümmerer-Funktion und der Krisenintervention; erweiterte Erreichbarkeit eines Arztes, interdisziplinärer Versorgungsplanung sowie telemedizinischen Leistungen für ca. ein Drittel der Teilnehmenden), so lässt sich konstatieren, dass den Teilnehmenden eine im Grundsatz höherwertige, die Regelversorgung ergänzende, Versorgung zur Verfügung stand, ohne auf Seiten der GKV signifikant höhere Ausgaben zu produzieren.

Die Kosten für Krankenhausaufhalte sowie die Anzahl notfallbedingter Krankenseinweisungen und die Rate der Umzüge in stationäre Dauerpflege lagen auf dem Niveau der Kontrollgruppen. Hier wären Projekte mit längerem Beobachtungszeitraum sowie einer Stichprobe, die den Kreis der Anspruchsberechtigten hinsichtlich weniger komplexer Fällen besser ausschöpft, erforderlich und zu begrüßen.

Die Ausschöpfung von Regelleistungen (Pflegegrade, Kurzzeitpflege, Tagespflege, zum Verbrauch bestimmte Pflegehilfsmittel, Maßnahmen zur Verbesserung des Wohnumfelds) gelang OBERBERG_FAIRsorgt nicht besser als der Regelversorgung in den Kontrollgruppen.

OBERBERG_FAIRsorgt insgesamt, wie auch die einzelnen Teilleistungen, erzielt unter den Teilnehmenden hohe Zufriedenheitswerte. Aus Sicht der Befragten sind alle Teilleistungen wichtig, besondere Bedeutung wird aber der Unterstützung durch die RCCM und den interdisziplinären Fallkonsilen beigemessen. Die Teilnehmenden wünschen sich eine Weiterführung dieser Neuen Versorgungsform und eine Ausweitung auf weitere Krankenkassen.

In Anbetracht der Komplexität der versorgten Fälle sowie des Anstiegs der Pflegegrade im Interventionszeitraum ist es als ein Erfolg der Intervention zu werten, dass sich die auf den Gesundheitszustand abzielenden Outcomes aus der Pre-Post-Befragung, mit Ausnahme der Erschöpfung, über 12 Monate nicht signifikant verschlechtert haben. Dies spricht für eine Stabilisierung von Gesundheitsstatus und Versorgungssituation, die durch die qualitativen Befunde (Teil I dieses Buchs) eindrücklich bestätigt wurde.

5 Literatur

- Caliendo, M. & Kopening, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1): 31-72. R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Youn, T.S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., Devellis, R., DeWalt, D., Fries, J.F., Gershon, R., Hahn, E.A., Lai, J.S., Pilkonis, P., Revicki, D., Rose, M., Weinfurt, K. & Hays, R. (2010). Initial item banks and first wave testing of the patient-reported outcomes measurement information system (PROMIS) network: 2005–2008. *Journal of Clinical Epidemiology* 63(11): 1179–1194.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. & Petticrew, M. (2013). Developing and evaluating complex interventions: the new Medical Research Council guidance. *International journal of nursing studies* 50(5): 587-592.
- Elixhauser, A., Steiner, C., Harris, D. R. & Coffey, R. M. (1998). Comorbidity measures for use with administrative data. *Medical care*, 36(1), 8-27.
- Faul, F., Erdfelder, E., Buchner, A. & Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41: 1149-1160.
- Fisher, F., Gibbons, C., Coste, J., Valderas, J.M., Rose, M., & Leplège, A. (2018). Measurement invariance and general population reference values of the PROMIS profile 29 in the UK, France, and Germany. *Quality of Life Research* 27(4): 999-1014.
- Herdman, M., Gudex, C., Lloyd, A., Janssen, M.F., Kind, P., Parkin, D., Bonnel, G. & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research* 20(10): 1727-1736.
- Hinz, A., Schumacher, J., Albani, C., Schmid, G. & Brähler, E. (2006). Bevölkerungsrepräsentative normierung der skala zur allgemeinen selbstwirksamkeitserwartung. *Diagnostica*, 52(1), 26-32.
- Ho, D., Imai, K., King, G. & Stuart, E. (2011). "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software*, *42*(8), 1-28.
- Jordan, S. & Hoebel, J. (2015). Gesundheitskompetenz von Erwachsenen in Deutschland. *Bundesgesundheitsblatt* 58(9), 942-950.
- Lenth, R. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.8.9.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Röthlin, F., Pelikan, J. M., & Ganahl, K. (2013). Die Gesundheitskompetenz der 15-jährigen Jugendlichen in österreich. Abschlussbericht der österreichischen Gesundheitskompetenz Jugendstudie
- Schumacher, J., Leppert, K., Gunzelmann, T., Strauß, B. & Brähler, E. (2005). Die Resilienzskala Ein Fragebogen zur Erfassung der psychischen Widerstandsfähigkeit als Personmerkmal. *Z Klin Psychol Psychiatr Psychother*, 53(1), 16-39.

Schwarzer, R. & Jerusalem, M. (Hrsg.) (1999). Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin: Freie Universität Berlin.

