

## DNFV-Positionspapier

## Methoden und Indikatorensets für die Evaluation regionaler sektorverbindender Versorgungsmodelle

In definierten Regionen angesiedelte Versorgungsmodelle sind oftmals dadurch gekennzeichnet, dass gleichzeitig mehr als eine Intervention neu in eine komplexe Umwelt eingeführt wird [1]. Die Komplexität der Umwelt ergibt sich unter anderem durch unterschiedliche Akteure, die von Versorgungsmodellen betroffen sein können (z. B. Patient:innen und ihre Angehörigen, niedergelassene Ärzt:innen, Pflegekräfte sowie weitere Gesundheitsberufe, Krankenhäuser, Kostenträger, Kommunen), Unterschiede bei den Zielgruppen der Interventionen, den organisatorischen Rahmenbedingungen oder sozioökonomischen Verhältnissen in den Regionen. Bei eingeschränkter Betrachtung scheitern viele solcher Versorgungsmodelle an vorab zu eng definierten Effektivitätsmaßen. Um den Wert regionaler Versorgungsmodelle umfassend beurteilen zu können und damit die Weichen für die Implementierung und Verstetigung eines regionalen Versorgungskonzepts zu stellen, müssen Evaluationskonzepte entwickelt werden, die Komplexität von Anfang an berücksichtigen. Der Medical Research Council hat im September 2021 eine Überarbeitung seines Konzepts zur Evaluation komplexer Interventionen vorgelegt, das als Richtschnur für

die Entwicklung von Evaluationskonzepten für regionale komplexe Versorgungsmodelle dienen kann [2]. Ebenfalls kürzlich hat auf europäischer Ebene das EU-finanzierte ‚Healthcare Performance Intelligence Network‘ Empfehlungen zur Evaluation von integrierten Versorgungsmodellen herausgegeben [3]. Zu einigen der in diesen Publikationen angesprochenen Aspekte wird im folgenden Positionspapier aus Sicht des Deutschen Netzwerks Versorgungsforschung Stellung genommen.

## Rahmenkonzept/Wirkmodell

Ausgangspunkt einer Evaluation für regionale Versorgungsmodelle muss ein Rahmenkonzept sein, das das Wirkmodell der geplanten Intervention/en mit allen wesentlichen Bestandteilen, Implementierungsstrategien und ihre intendierten Mechanismen sowie den aktuellen Kontext, in dem interveniert wird, exakt beschreibt. Wichtig ist dabei zu bedenken, dass Wirkungen Zeit benötigen („roll-in Phase“) und somit für den Evaluationszeitraum ein dementsprechend angemessenes Beobachtungsfenster gewählt werden sollte. Als wesentlicher Baustein der Evaluation ist immer auch eine formative Evaluation bzw. Prozessevaluation der tatsächlichen Wirkungen zu berücksichtigen. Relevante Fragen in diesem Zusammenhang sind z. B.: War die Intervention machbar und wurde sie tatsächlich wie geplant umgesetzt und in welchem Umfang? Wie stand es um die Akzeptanz bei den Betroffenen, aber auch denjenigen, die für die Umsetzung verantwortlich waren [4]?

Für die Akzeptanz, aber auch einen späteren Implementierungserfolg in der Routine ist es wichtig, die regionalen Stakeholder wie auch die Nutzerperspektive, vor allem also auch Patient:innen von Anfang an bei der Konzeption des Wirkmodells einzubinden. Dabei müssen die Verantwortlichkeiten für die Umsetzung und das Monitoring der Umsetzung klar benannt werden. Kenntnisse der regionalen Vernetzungsstrukturen und der in der Region vorhandenen Steuerungskompetenzen sind notwendig, um die für den Erfolg eines regionalen Versorgungsmodells wesentlichen Akteure zu identifizieren.

## Fragestellung/Ziel

Je klarer die Fragestellung und das Ziel eines regionalen Versorgungsmodells formuliert wurden, desto einfacher ist es, die Evaluation zu konzipieren. An einem differenzierten Wirkmodell für die Intervention/en orientiert, lassen sich konkrete Ziele ableiten, deren Erreichung dann im Rahmen der Evaluation überprüft werden kann. Zu hoch gesteckte Ziele oder von vornherein überzogene Erwartungen an die potenziellen Wirkungen eines regionalen Versorgungsmodells enden oft in unzureichenden Evaluationsergebnissen. Auf der anderen Seite sind zwar leicht umzusetzende, jedoch für Bürger:innen und Patient:innen letztlich nicht relevante Ergebnisse zu vermeiden, da sie als Argument für eine spätere Implementierung des regionalen Versorgungsmodells in der Routine nicht überzeugen.

Der Medical Research Council rät dazu, im Evaluationsprozess offen für eine Anpassung der Ziele komplexer Interventionen zu sein. Dabei stellt sich stets die Frage, ob alle Parameter zur Evaluation eines neuen Zieles überhaupt von Anfang an objektiv, valide und reliabel erfasst wurden und wie eine Abweichung von einem vorab festgelegten Studienprotokoll wissenschaftlich zu bewerten ist.

## Design

Letztlich bestimmt die mit einem regionalen Versorgungsmodell verfolgte Fragestellung das bei einer Evaluation anzulegende Studiendesign. Dem wissenschaftlichen Paradigma folgend wäre für den Nachweis der Wirksamkeit einer Intervention, sei es als Äquivalenz- oder Überlegenheitsnachweis, ein randomisiert-kontrolliertes Studiendesign notwendig, mit dem Kausalbeziehungen zu belegen sind. Aber auch andere Fragestellungen können von wesentlichem Erkenntnisgewinn sein, wie beispielsweise die Machbarkeit und Akzeptanz eines regionalen Versorgungsmodells, die Versorgungsgerechtigkeit oder Effizienz eines Modells, wozu typischerweise andere Studiendesigns, oftmals Kombinationen von qualitativen und quantitativen Ansätzen in Form von Mixed-Methods-Analysen zu nutzen sind. Darüber hinaus ist zu diskutieren, welches Studiendesign geeignet ist, die Fragestellungen möglichst nahe an der Versorgungsroutine bzw. im Versorgungsalltag zu

beantworten, um so die Implementierung eines Modells in der Routine zu befördern.

Ebenfalls zu beachten ist die oben genannte Forderung des Medical Research Council nach der Offenheit für Anpassungen der Ziele bzw. Fragestellungen im Evaluationsprozess, für die eventuell zusätzliche Parameter erhoben oder andere Designs notwendig sind. Deshalb reicht das Repertoire potenzieller Studiendesigns von experimentellen Ansätzen wie RCTs oder Cluster-RCTs, zum Teil kombiniert mit qualitativen Studien, über beobachtende, kontrollierte und nicht-kontrollierte, prospektive und retrospektive Designs, über Surveys und Fokusgruppen bis hin zu teilnehmenden Beobachtungen. Bei Beobachtungsstudien kommen verschiedene, zum Teil aus der Ökonometrie bekannte Methoden zur Kontrollgruppenbildung zum Einsatz (z. B. Difference-in-differences-Methoden, Propensity-Score-Matching oder Entropy-Balancing), die bei der Verwendung von Routinedaten aufgrund der nur begrenzt verfügbaren Variablen trotzdem an ihre Grenzen kommen und durch weitergehende Überlegungen zur Vergleichbarkeit von Regionen ergänzt werden müssen, wie dies beispielhaft bei der 10-Jahres-Evaluation des ‚Gesunden Kinzigtals‘ erprobt wurde [5].

## Messung

Wie für die Wahl des Studiendesigns gilt auch für die Erfolgsmessung im Rahmen der Evaluation, dass die erforderlichen Evaluationsparameter vom Wirkmodell abgeleitet werden müssen. Bei der Auswahl von Indikatoren, die bei der Evaluation regionaler Versorgungsmodelle eingesetzt werden sollen, ist eine Vielzahl von Aspekten zu beachten, die hier nur angedeutet werden können. Zunächst stellt sich die Frage, ob eine begrenzte Evaluation mit allein auf Routinedaten beruhenden Indikatoren oder eine umfassendere Evaluation durchgeführt werden soll, für die auch Primärdaten generiert werden müssen. In den meisten Versorgungsbereichen liegen nur begrenzt patientenrelevante Ergebnis- oder Erfahrungsdaten bzw. -indikatoren vor (Tod, Rezidiv, AU, inzidente Komorbiditäten) und patientenberichtete Outcomes (PROM, PREM) fehlen, so dass diesbezügliche Fragestellungen mit Routinedaten zumeist nicht beantwortbar sind.

Eine in den letzten Jahren positive Entwicklung aus dem Bereich der Routinedaten ist für die Evaluation von Langzeitverläufen zu konstatieren, da Krankenversicherungen zunehmend in der Lage sind, die Leistungsdaten ihrer Versicherten über Jahre hinweg zu analysieren. Damit werden auch Indikatoren zur Verzögerung eines Erkrankungsbeginns oder Vermeidung von Erkrankungen und Langzeitpflegebedürftigkeit in einer Region messbar. Sollen Routinedaten genutzt werden, dann bereitet die Zusammenführung verschiedener Datenquellen (z. B. von verschiedenen GKV, PKV, KVen/ZI, IQTIG, Krebs- und klinischen Registern) vielfach Probleme, da die Datenkörper mit unterschiedlich definierten Datenfeldern nicht für eine Kopplung konzipiert sind.

Zudem stellt sich bei der Evaluation regionaler Versorgungsmodelle die Frage der Aggregationsebene, also ob nur ein (kleinräumiger) regionaler Bezug oder aber ein Bezug auf einzelne Leistungserbringer möglich sein soll, womit Datenschutzaspekte berührt werden.

Um die Handlungsrelevanz von Indikatoren zu erhöhen, müssen nicht nur Indikatoren für Expert:innen, sondern auch solche für Praktiker entwickelt werden, die einfach zu interpretieren und praktikabel sind sowie zeitnah erhoben und rückgespiegelt werden können [6]. Dabei müssen Prozess- von Ergebnisindikatoren unterschieden werden. Klar ist, dass für die formative Evaluation Prozessindikatoren notwendig sind, die zumeist einfacher bzw. direkter Optionen für Verbesserungen aufzeigen und ein Monitoring auf der Handlungsebene ermöglichen. Dagegen sind für die summative Evaluation immer Ergebnisindikatoren zu erheben. Bei diesen stellt sich am deutlichsten die Frage der methodischen Güte von Indikatoren, zu denen unter anderem die Änderungssensitivität, das mögliche Problem kleiner Fallzahlen, der Risikoadjustierung und Standardisierung der Indikatoren sowie der Zuordnung von Verantwortlichkeit für eine Ergebnisausprägung zählen.

## Synthese und Darstellung der Ergebnisse

Um eine fundierte Interpretation der Ergebnisse einer Evaluation eines regionalen Ver-

sorgungsmodells sowie darauf aufbauend eine Verstetigung der erprobten komplexen Intervention zu erlauben, sollten die Ergebnisse der Evaluation für die verschiedenen Akteure adressatengerecht dargestellt werden. Wichtig ist auch, den jeweiligen Kontext genau zu beschreiben und eventuelle Veränderungen des Kontexts, die durch das regionale Versorgungsmodell hervorgerufen wurden, ebenfalls darzulegen. Dabei sollte bedacht werden, dass Modellverträge nicht nur zu Adaptationsphänomenen im System führen, sondern zuweilen auch Routinedaten beeinflussen, indem diese zum Teil nur noch nicht vom Modell erfasste Populationen erfassen.

Insgesamt sollte darauf geachtet werden, die vom Modell erreichte Population, die oftmals nur eine Teilpopulation der regionalen Gesamtpopulation umfasst, genau zu definieren und die Ergebnisse auch auf diese Population bezogen darzustellen. Um ggfls. auftretende Risikoselektionseffekte abschätzen und in ihrer Ergebnisauswirkung beurteilen zu können, sollte parallel auch eine Auswertung der Effekte auf die jeweilige Gesamtgruppe von Versicherten bzw. Populationen erfolgen. Ob eine Ergebnisdarstellung in Anlehnung an eine „Balanced Score Card“, also der gleichzeitigen Darstellung unterschiedlicher Evaluationskriterien der betrachteten Region erfolgt oder aber andere Ergebnisdarstellungen geeignet sind, sollte im Verlauf der Erprobung mit den späteren Adressaten besprochen werden.

## Interpretation, Handlungsoptionen

Die Interpretation der Evaluationsergebnisse und die sich daraus ergebenden Handlungsoptionen müssen vor dem Hintergrund des jeweils speziellen Kontextes erfolgen. Dabei sind die formativen und die summativen Ergebnisse zu gewichten und unter Umständen auch gegeneinander abzuwägen. Vorschnelle positive wie negative Bewertungen regionaler Versorgungsmodelle lassen sich durch die Berücksichtigung der hier genannten Punkte vermeiden. Dies gilt zum einen im Hinblick auf die Übertragbarkeit auf andere Regionen mit anderen Kontextbedingungen, zum anderen auch im Hinblick auf Kausalbeziehun-

gen, deren Nachweis auf der Basis der genutzten Studiendesigns häufig nicht möglich ist.

## Implementierung

Die Implementierung eines positiv evaluierten regionalen Versorgungsmodells in die Routineversorgung sollte Ziel der Erprobung sein. Dafür ist es notwendig, die Evaluationsergebnisse so aufzubereiten, dass die Erfahrungen aus der Erprobung für die Entwicklung anderer regionaler Versorgungsmodelle genutzt werden können. Der Medical Research Council rät dazu, in allen Phasen der Erprobung komplexer Interventionen, der Entwicklungsphase, Machbarkeitstestung und Evaluationsphase, bereits deren spätere Implementierung mitzudenken und Anpassungen frühzeitig vorzunehmen, wenn sich Hindernisse für eine spätere Implementierung in der Routine zeigen [2]. Da sich der jeweilige Kontext in unterschiedlichen Regionen nach der Implementierung vom Kontext der Region unterscheiden wird, in dem das Versorgungsmodell evaluiert wurde, ist auch nach der Implementierung ein Monitoring auf Basis geeigneter Prozess- und Ergebnisindikatoren anzustreben.

### FAZIT

Eine dem Stand der Wissenschaft entsprechende Evaluation regionaler sektorverbindender Versorgungsmodelle setzt voraus, dass vielfältige methodische Aspekte beachtet werden, die nur interdisziplinärkooperativ und mit ausreichenden Ressourcen zu bewältigen sind. Um in Zukunft die Erprobung und Evaluation solcher Modelle in Deutschland zu vereinfachen, plädiert das DNVF dafür, ein möglichst standardisiertes Indikatorenset zu etablieren, das Aspekte der Gesundheits- und Qualitätsberichterstattung vereint und regionale sektorverbindende Vergleiche und Transparenz über die Gesundheitsversorgung ermöglicht. Dazu wird es unter anderem nötig sein, einen krankenkassenübergreifenden Routinedatensatz zu etablieren, gleichzeitig darüber hinausgehende

Datenquellen einzubinden und auch standardisierte Befragungen zu den von Patient:innen berichteten Erfahrungen und Ergebnissen (PRE und PRO) einzuführen.

### Interessenskonflikt

Die Autor:innen erklären, dass kein Interessenkonflikt besteht.

### Autoren

Max Geraedts, Thomas Bierbaum, Oliver Gröne, Martin Härter, Helmut Hildebrandt, Wolfgang Hoffmann, Monika Klinkhammer-Schalke, Juliane Köberlein-Neu, Jochen Schmitt

### Korrespondenzadresse

Prof. Dr. med. Max Geraedts, M.San.  
Institut für Versorgungsforschung und Klinische Epidemiologie, Fachbereich Medizin, Philipps-Universität Marburg  
Karl-von-Frisch-Straße 4  
35043 Marburg  
Deutschland  
E-Mail: geraedts@uni-marburg.de

### Literatur

- [1] Plsek PE, Greenhalgh T. Complexity science: The challenge of complexity in health care. *BMJ* 2001; 323: 625–628. doi:10.1136/bmj.323.7313.625
- [2] Skivington K, Matthews L, Simpson SA et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ* 2021; 374: n2061. doi:10.1136/bmj.n2061
- [3] Larrain N, Groene O. Policy guidance on advancing the performance assessment of integrated healthcare systems. Im Internet [https://www.healthpros-h2020.eu/images/Recommendations/HPIS\\_14\\_Policy\\_guidance\\_on\\_advancing\\_the\\_performance\\_assessment\\_of\\_integrated\\_healthcare\\_systems\\_30032022.pdf](https://www.healthpros-h2020.eu/images/Recommendations/HPIS_14_Policy_guidance_on_advancing_the_performance_assessment_of_integrated_healthcare_systems_30032022.pdf) Stand: 01.04.2022
- [4] Elwy AR, Wasan AD, Gillman AG et al. Using formative evaluation methods to improve clinical implementation efforts: Description and an example. *Psychiatry Res* 2020; 283: 112532. doi:10.1016/j.psychres.2019.112532
- [5] Stelzer D, Graf E, Köster I et al. Assessing the effect of a regional integrated care

model over ten years using quality indicators based on claims data – the basic statistical methodology of the INTEGRAL project. *BMC Health Serv Res* 2022; 22: 1–20. doi:10.1186/s12913-022-07573-7

- [6] Barbazza E, Kara P, Larrain N et al. A practical guide towards actionable healthcare performance indicators: selecting healthcare performance indicators that are fit for purpose and use for various stakeholders. In Internet [https://www.healthpros-h2020.eu/images/Recommendations/HPIS\\_13\\_Fit\\_for\\_purpose\\_use\\_indicators\\_30032022.pdf](https://www.healthpros-h2020.eu/images/Recommendations/HPIS_13_Fit_for_purpose_use_indicators_30032022.pdf)  
Stand: 01.04.2022