

# Adjustierung des sozialmedizinischen Verlaufs nach medizinischer Rehabilitation: methodische Weiterentwicklung der Reha-Qualitätssicherung der Deutschen Rentenversicherung

## Risk-Adjustment of Return to Work After Medical Rehabilitation: Methodical Advancements in Quality Assurance of the German Pension Insurance

### Autoren

Sarah Leinberger<sup>1</sup>, Christian Hetzel<sup>2</sup>, Rainer Kaluscha<sup>1</sup>

### Institute

- 1 Universität Ulm, Institut für Rehabilitationsmedizinische Forschung, Bad Buchau
- 2 Deutsche Sporthochschule Köln, Institut für Qualitätssicherung in Prävention und Rehabilitation GmbH, Köln

### Schlüsselwörter

Rehabilitation, Qualitätsindikator, sozialmedizinischer Verlauf, Return to Work, Risikoadjustierung

### Key words

rehabilitation, quality indicator, return to work, risk adjustment

online publiziert 16.02.2023

### Bibliografie

Rehabilitation 2023; 62: 225–231

DOI 10.1055/a-1998-6574

ISSN 0034-3536

© 2023. Thieme. All rights reserved.

Georg Thieme Verlag, Rüdigerstraße 14, 70469 Stuttgart, Germany

### Korrespondenzadresse

Sarah Leinberger, M. Sc.

Institut für Rehabilitationsmedizinische Forschung an der Universität Ulm

Am Kurpark 1

88422 Bad Buchau

Deutschland

sarah.leinberger@ifr-ulm.de



**Zusätzliches Material** finden Sie unter <https://doi.org/10.1055/a-1998-6574>

### ZUSAMMENFASSUNG

**Ziel der Studie** Die berufliche (Wieder-)Eingliederung der Rehabilitanden stellt neben der Lebensqualität ein wichtiges Behandlungsergebnis einer medizinischen Rehabilitationsmaßnahme in Trägerschaft der Deutschen Rentenversicherung dar. Um den sozialmedizinischen Verlauf in Form der beruflichen (Wieder-)Eingliederung als Qualitätsindikator nutzen zu können, musste zunächst eine geeignete Operationalisierung gefunden werden, um anschließend eine Risikoadjustierung hinsichtlich vorbestehender ergebnisrelevanter Rehabilitanden-, Fachabteilungs- und Arbeitsmarktmerkmale zu erarbeiten, welche die Fachabteilungen nicht unmittelbar beeinflussen können.

**Methodik** Dazu wurde mittels multipler Regressionsanalysen und Kreuzvalidierung eine Adjustierungsstrategie entwickelt, welche die Einflüsse ergebnisrelevanter Confounder (Einflussgrößen) mathematisch kompensiert und so sachgerechte Fachabteilungsvergleiche bezüglich der sozialmedizinischen Verläufe der Rehabilitanden nach ihrer medizinischen Rehabilitation zulässt. Als geeignete Operationalisierung des sozialmedizinischen Verlaufs in Form der beruflichen (Wieder-)Eingliederung wurden unter Einbezug von Experten die sozialversicherungspflichtigen Beschäftigungstage im ersten und zweiten Jahr nach der Rehabilitationsmaßnahme ausgewählt. Methodische Herausforderungen bei der Entwicklung der Adjustierungsstrategie stellten die Auswahl einer geeigneten Regressionsmethode für die Verteilungsbesonderheiten der Zielgröße, die adäquate Modellierung der Mehr-Ebenenstruktur der Daten, und die Selektion der relevanten Einflussgrößen auf den sozialmedizinischen Verlauf dar. Auf Basis der adjustierten Ergebnisse wurde ein anwenderfreundliches Rückmeldekonzept entwickelt.

**Ergebnisse** Als angemessene Regressionsmethode zur Modellierung der U-förmig verteilten Anzahl der Beschäftigungstage wurde die Fractional Logit Regression gewählt. Die Mehr-Ebenenstruktur der Daten (cross-classified Arbeitsmarktregionen und Fachabteilungen) ist aufgrund von durchweg niedrigen Intraklassenkorrelationen statistisch zu vernachlässigen. Potenzielle Einflussgrößen wurden theoretisch vorselektiert (unter Einbezug medizinischer Experten bei medizinischen Parametern) und indikationsspezifisch über Rückwärtsselekti-

on auf ihre prognostische Relevanz geprüft. Die Strategie erwies sich nach Kreuzvalidierung als stabil. Die Ergebnisse der Adjustierung wurden unter Einbezug der Nutzerperspektive (Interviews und Fokusgruppen) anwenderorientiert in einem Rückmeldekonzept dargestellt.

**Schlussfolgerung** Das entwickelte Adjustierungsverfahren ermöglicht durch die Kompensation vorbestehender Rehabilitandenmerkmale einen sachgerechten Fachabteilungsvergleich zur Beurteilung der Behandlungsergebnisse im Rahmen der Reha-Qualitätssicherung. Die methodischen Herausforderungen, Entscheidungen und Limitationen werden in diesem Beitrag ausführlicher dargestellt.

## ABSTRACT

**Purpose** Besides the quality of life, patients' return to work is one of the most important treatment results of medical rehabilitation paid by the German Pension Insurance. In order to be able to use the return to work as a quality indicator for medical rehabilitation, a risk adjustment strategy for pre-existing characteristics of patients, rehabilitation departments and labour markets had to be developed.

**Methods** Multiple regression analyses and cross validation were used to develop a risk adjustment strategy, which mathematically compensates the influence of confounders and thus allows for appropriate comparisons between rehabilitation departments regarding patients' return to work after medical

rehabilitation. Under the inclusion of experts, the number of employment days in the first and second year after medical rehabilitation were chosen as an appropriate operationalization of return to work. Methodological challenges in the development of the risk adjustment strategy were the identification of a suitable regression method for the distribution of the dependent variable, modelling the multilevel structure of the data appropriately and selecting relevant confounders for return to work. A user-friendly way of communicating the results was developed.

**Results** The fractional logit regression was chosen as an appropriate regression method to model the U-shaped distribution of the employment days. Low intraclass correlations indicate that the multilevel structure of the data (cross-classified labour market regions and rehabilitation departments) is statistically negligible. Potential confounding factors were theoretically preselected (medical experts were involved for medical parameters) and tested for their prognostic relevance in each indication area using backwards selection. Cross validations proved the risk adjustment strategy to be stable. Adjustment results were displayed in a user-friendly report, including the users' perspective (focus groups and interviews).

**Conclusions** The developed risk adjustment strategy allows for adequate comparisons between rehabilitation departments and thus enables a quality assessment of treatment results. Methodological challenges, decisions and limitations are discussed in details throughout this paper.

## Einleitung

### Reha-Qualitätssicherung der Deutschen Rentenversicherung

Bei der Deutschen Rentenversicherung (DRV) werden Rehabilitationsleistungen mit dem Ziel durchgeführt, eine bereits geminderte Erwerbsfähigkeit zu verbessern oder zumindest eine Verschlechterung zu vermeiden [1]. Zur fortlaufenden Verbesserung der Leistungsqualität in der medizinischen Rehabilitation werden von der Deutschen Rentenversicherung standardisierte Instrumente der Reha-Qualitätssicherung (QS) für die von ihr belegten Fachabteilungen eingesetzt. In einer Rehabilitationsklinik kann es eine oder mehrere Fachabteilungen geben, da diese die medizinischen Spezialisierungen abbilden (z. B. Orthopädie und/oder Psychosomatik). Aktuell werden unter anderem die Behandlungszufriedenheit und der subjektive Behandlungserfolg als Qualitätsindikatoren erfasst [2]. Dabei erfolgt eine Adjustierung, um unterschiedlichen Verteilungen vorbestehender ergebnisrelevanter Rehabilitandenmerkmale in den einzelnen Fachabteilungen Rechnung zu tragen und so einen sachgerechten Vergleich der Ergebnisse aus den Fachabteilungen einer Indikationsgruppe zu ermöglichen [2]. Als Einflussgrößen (Confounder) werden soziodemographische sowie gesundheitsbezogene Merkmale und Eigenschaften des Rehabilitationszugangs und der Fachabteilung in der Adjustierung berücksichtigt [3]. Bisher ist der sozialmedizinische Verlauf, d. h. der Verbleib im bzw. das Ausscheiden aus dem Erwerbsleben, in Form der

(Wieder-)Aufnahme einer sozialversicherungspflichtigen Beschäftigung nach medizinischer Rehabilitation durch die DRV nur begrenzt Gegenstand der Reha-QS. Gründe hierfür sind, dass für ihn noch keine geeignete Operationalisierung sowie Strategie zur Risikoadjustierung ausgearbeitet wurde, welche eine Bewertung der Qualität zugelassen hätte. Bei der Risikoadjustierung sollte zusätzlich zu den bekannten Confoundern anderer Qualitätsindikatoren auch der Arbeitsmarkt Berücksichtigung finden, weil sich Arbeitsmarkteinflüsse als relevant für den Eingliederungserfolg von Leistungen zur Teilhabe am Arbeitsleben erwiesen haben [4].

### Ziel und methodische Herausforderungen

Für die QS der DRV sollen Fachabteilungen anhand des sozialmedizinischen Verlaufs in Form der beruflichen (Wieder-)Eingliederung ihrer Rehabilitanden auf Basis administrativer Daten verglichen werden. Dazu muss für Merkmale adjustiert werden, welche die Fachabteilungen der Rehabilitationseinrichtungen nicht beeinflussen können, z. B. vor Beginn der Rehabilitation bereits feststehende Merkmale der Rehabilitanden, für die ein Einfluss auf den sozialmedizinischen Verlauf nachgewiesen werden kann.

Damit die Fachabteilungen von den Rückmeldungen zum sozialmedizinischen Verlauf profitieren und daraus Rückschlüsse für die Weiterentwicklung ihrer Behandlungsstrategien ziehen können, müssen die Adjustierungsergebnisse verständlich und anwenderorientiert aufbereitet werden.

Zur Ermittlung der individuell erwarteten beruflichen (Wieder-)Eingliederung wurde ein indikationsspezifisches Regressionsmodell unter Berücksichtigung der relevanten Confounder entwickelt. Diese erwarteten Werte wurden nach Fachabteilungen aggregiert. Für jede Fachabteilung wurde dann der beobachtete dem erwarteten Wert gegenübergestellt und aus der Differenz ein Qualitätsindikator für den sachgerechten Fachabteilungsvergleich errechnet.

Bei der Entwicklung der Adjustierungsstrategie waren folgende methodische Herausforderungen zu bewältigen.

### Zielgröße

In bisherigen Auswertungen des sozialmedizinischen Verlaufs wurde dieser mithilfe von Lücken in Beitragsverläufen oder sozialversicherungspflichtiger Beschäftigung operationalisiert [5]. Allerdings birgt diese Operationalisierung des sozialmedizinischen Verlaufs Optimierungspotentiale, da auch aus Sozialleistungen wie Arbeitslosengeld I oder Krankengeld Pflichtbeiträge gezahlt werden. Somit kann auf dieser Basis keine klare Aussage zum Erwerbsstatus getätigt werden. Es sollen verschiedene Alternativen zur Operationalisierung des sozialmedizinischen Verlaufs in Form der beruflichen (Wieder-)Eingliederung auf der Datengrundlage (vgl. Abschnitt „Diskussion der Methoden und Ergebnisse“) diskutiert werden.

### Regressionsmethode

Eine weit verbreitete Regressionsmethode für Vorhersagemodelle mit numerischer Zielgröße ist die lineare Regression. Lineare Regressionsmodelle sind jedoch nicht für alle Daten und Wirkzusammenhänge optimal geeignet. So können vorhergesagte Werte einer Zielvariablen mit festem beobachtetem Wertebereich beispielsweise über dessen Ränder hinausgehen (z. B. negative Beschäftigungstage). Bei der Modellierung von schief oder multimodal verteilten Zielgrößen über lineare Regressionsmodelle ist außerdem mit Verletzung der Regressionsannahmen wie Normalverteilung oder Heteroskedastizität der Residuen in den Randbereichen zu rechnen. Insofern stellt sich die Frage, welche Regressionsmethode der Verteilung der Zielgröße sowie der großen Fallzahl (z. B. Orthopädie: N = 305.980 Reha-Maßnahmen in 2016, vgl. **Online-Anhang 1**) am besten gerecht wird.

### Mehr-Ebenenstruktur der Daten

Rehabilitanden sind in Fachabteilungen sowie in Arbeitsmarktregionen gruppiert. In der Regel sind sich Personen innerhalb von Fachabteilungen ähnlicher als Personen zwischen Fachabteilungen (analog für Arbeitsmarktregionen). Statistisch bedeutet das, dass bei solchen hierarchischen Datenstrukturen die beobachteten Werte abhängig voneinander sind [6]. Dies kann dazu führen, dass Standardfehler konventioneller statistischer Tests überschätzt werden und zentrale Annahmen zur Verteilung der Residuen (Linearität und Normalität) verletzt sind und so zu verzerrten Schätzern führen. Mit Mehrebenenmodellen sind diese Limitationen zu bewältigen [7]. Im vorliegenden Fall liegt keine perfekte hierarchische Datenstruktur vor, weil Personen zwar aus unterschiedlichen Arbeitsmarktregionen stammen können, aber in derselben Fachabteilung behandelt werden – oder umgekehrt (cross-classified [8]).

### Selektion der Einflussgrößen

Für eine Adjustierung ist die Berücksichtigung der Confounder sinnvoll, die einen Einfluss auf die Zielgröße haben könnten, valide vorliegen und die eine Fachabteilung nicht unmittelbar beeinflussen kann (vgl. auch [9]). Nach dieser theoretischen Vorselektion ist zu entscheiden, ob und wie mit statistischen Methoden weiter selektiert wird. Ausgelassene Einflussgrößen führen womöglich zu Verzerrungen. Diese Gefahr besteht, wenn beispielsweise aus Gründen der Praktikabilität „sparsame“ Modelle gefordert werden [10].

### Rückmeldekonzep

Auf Basis der Adjustierungsergebnisse sollte ein Rückmeldekonzep für die Nutzer der Qualitätsberichte der DRV erarbeitet werden. Aus Platzgründen können die Entwicklung und die Inhalte des Rückmeldekonzeps hier nicht näher beschrieben werden. Eine separate Veröffentlichung hierzu ist geplant.

### Datensatz

Die Datenanalysen wurden auf Grundlage der Reha-Statistik-Datenbasis (RSD) der Deutschen Rentenversicherung durchgeführt. Diese basiert auf dem internen SK-94-Datensatz, der die Grundlage für die gemeinsame Berichterstattung der 16 Rentenversicherungsträger bildet. Die Nutzung eines Datenauszeuges für wissenschaftliche Zwecke ist über das Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) möglich; dort findet sich auch eine Beschreibung [11]. Die im Analysezeitraum vorliegenden Daten erlauben eine Auswertung der sozialmedizinischen Verläufe für Rehabilitanden mit Abschluss einer regulären medizinischen Rehabilitation im Jahr 2016.

Die RSD umfasst Routinedaten von jedem Berechtigten, der eine Leistung zur medizinischen Rehabilitation gemäß §§ 15, 31 SGB VI oder eine Leistung zur Teilhabe am Arbeitsleben gemäß § 16 SGB VI erhalten hat, die vor, während und bei Abschluss der Leistung erhoben werden, sowie einige weitere statistische Merkmale, die der Versicherungsträger aus dem Versicherungskonto bereitstellt. Die erbrachten Reha-Leistungen sind einer Fachabteilung zugeordnet, die wiederum zu indikationsbezogenen Vergleichsgruppen zusammengefasst werden.

Ergänzend werden regionale Arbeitsmarktmerkmale aus der IN-KAR-Datenbank [12] für die Raumeinheit Arbeitsmarktregion verwendet und anhand des Wohnorts der Rehabilitanden auf Kreisebene mit der RSD verknüpft.

Rehabilitationsmaßnahmen werden in die Auswertungen zur DRV-Reha-QS einbezogen, sofern der Rehabilitand im Jahr vor der Reha-Antragsstellung mindestens einen Pflichtbeitrag in die Rentenversicherung eingezahlt hat und zum Zeitpunkt der Antragstellung erwerbsfähig war.

Verstirbt ein Rehabilitand im Beobachtungszeitraum, werden seine Daten von den Analysen ausgeschlossen.

Um Verzerrungen aufgrund von Zufallsschwankungen zu vermeiden, müssen pro Fachabteilung mindestens 25 einschließbare Rehabilitationsmaßnahmen vorliegen.

Die QS-Vergleichsgruppen, d. h. Indikationsgruppen nach Definition der DRV, sind in **Online-Anhang 1** deskriptiv beschrieben.

## Methoden

Die Entwicklung der Adjustierung und des Rückmeldekonzeptes wurde durch einen Expertenbeirat begleitet, in dem Fachleute aus Kliniken, Verbänden, DRV und Wissenschaft vertreten waren.

### Zielgröße

Zusammen mit dem wissenschaftlichen Projektbeirat wurden potenzielle Operationalisierungen des sozialmedizinischen Verlaufs der Rehabilitanden in Form der beruflichen (Wieder-)Eingliederung diskutiert. Der Expertenbeirat empfahl als Zielgröße die Anzahl der Beschäftigungstage im ersten und zweiten Kalenderjahr nach Ende der Reha-Maßnahme (► **Abb. 1**).

### Regressionsmethode

Für Anteilswerte in einem begrenzten Wertebereich, wie sie bei der gewählten Zielgröße „sozialversicherungspflichtige Beschäftigungstage nach Rehabilitation“ vorliegen, ist die Fractional Logit Regression (FLR) eine angemessene Regressionsmethode. FLR wurde von Wedderburn [13] 1974 erstmals beschrieben und von Papke und Wooldrige [14] 1996 quasi wiederentdeckt. Sie zählt zur Familie der verallgemeinerten linearen Modelle (GLM), basiert auf Quasi-Maximum Likelihood Schätzern und ist der binären logistischen Regression sehr ähnlich. Basis der FLR ist eine Binomialverteilung, mit einem zusätzlichen Parameter zur Schätzung der abweichenden Fehlervarianz in den Daten. Dadurch ist die Methode sehr flexibel und erfordert keine besonderen Verteilungsannahmen. Im Kern steht wie bei der binären logistischen Regression die Logit-Link-Funktion.

Die Schätzer für die einzelnen Einflussfaktoren sind Logits, deren Interpretation allerdings wenig intuitiv ist [15]. Daher werden zusätzlich die durchschnittlichen marginalen Effekte (average marginal effects, AME) berechnet. Die AME stellen den Durchschnitts-

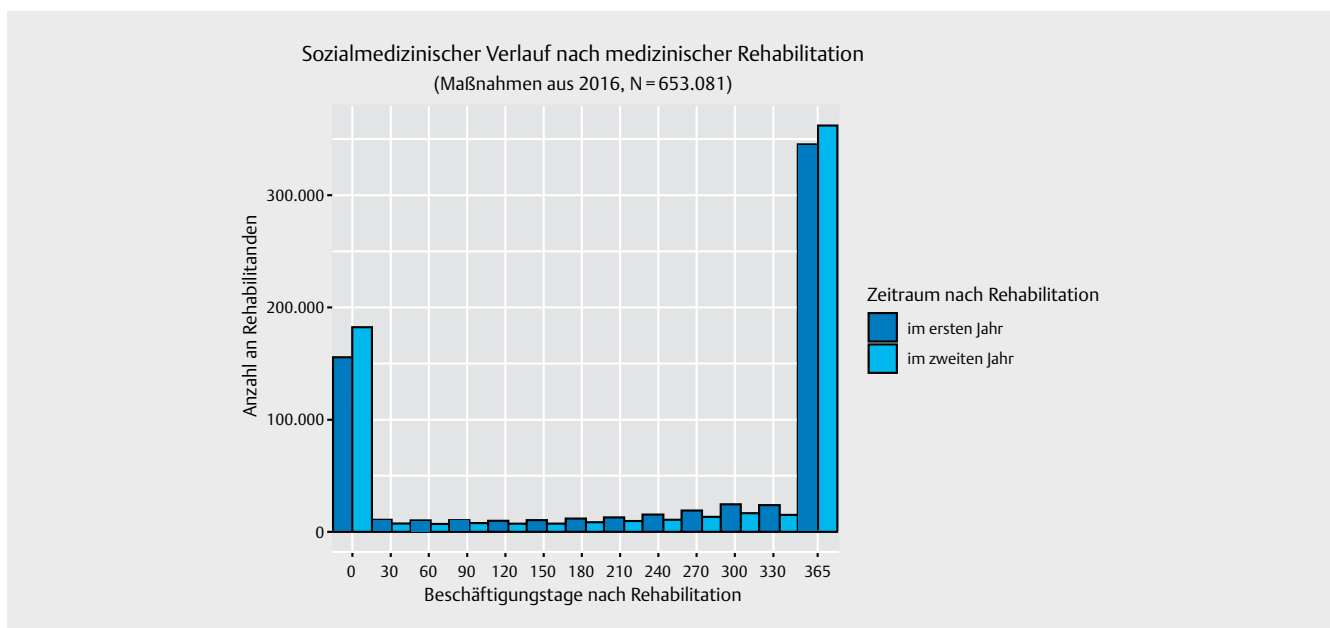
effekt einer Einflussgröße als Mittelwert der marginalen Effekte über alle Beobachtungen dar [16]. Sie sind ähnlich wie die Koeffizienten einer linearen Regression zu interpretieren und werden in der Einheit der Zielgröße (hier: Tage) berichtet. Sämtliche Berechnungen wurden mit der Programmiersprache R durchgeführt (für die verwendeten R-Pakete und Syntax vgl. **Online-Anhang 2**).

Für die Beschreibung der Modellgüte wurde das Pseudo-R<sup>2</sup> nach McKelvey und Zavoina [17] verwendet, welches der erklärten Varianz R<sup>2</sup> aus linearen Modellen nahe kommt [18]. Die finalen Modelle erzielten je nach QS-Vergleichsgruppe ein Pseudo-R<sup>2</sup> von 0,2940 bis 0,4650 für den einjährigen und von 0,2560 bis 0,4457 für den zweijährigen Beobachtungszeitraum.

In die finalen Adjustierungsmodelle flossen je nach Indikationsbereich im Mittel 24 Einflussgrößen ein, z. B. Beschäftigungstage zwei Jahre vor der Rehabilitation, Art der geleisteten Rentenversicherungsbeiträge drei Monate vor der Rehabilitation sowie medizinisch begründete Untergruppen wie Rückenschmerz- und Total-Endoprothese-Patienten im Indikationsbereich Orthopädie und psychische Multimorbidität in der Psychosomatik. Das Geschlecht hatte in einigen Indikationsbereichen wie der inneren Medizin und der Dermatologie keine statistisch relevante Vorhersagekraft auf die sozialversicherungspflichtigen Beschäftigungstage nach der Rehabilitation. Ebenso wurden indikationsspezifisch einige medizinische Merkmale wie Komplikationen bei einer Total-Endoprothese (Orthopädie) geprüft, welche allerdings statistisch nicht relevant waren und daher nicht weiterverfolgt wurden.

### Mehr-Ebenenstruktur der Daten

Die Berücksichtigung der cross-classified Datenstruktur (Personen in Fachabteilungen bzw. Arbeitsmarktregionen) ist mindestens dann nötig, wenn die Gruppierungen in diesen Ebenen einen substantziellen Varianzanteil der Zielgröße erklären. Die Intraklassenkorrelation (Intra Class Correlation, ICC) ist ein Maß für die interne



► **Abb. 1** Sozialmedizinischer Verlauf der Rehabilitanden in Form der sozialversicherungspflichtigen Beschäftigungstage im ersten und zweiten Jahr nach medizinischer Rehabilitation 2016, Quelle: Reha-Statistik-Datenbasis (RSD) 2018.

Homogenität der Gruppierungen auf den höheren Datenebenen, hier der Fachabteilungen bzw. der Arbeitsmarktregionen, in Bezug auf die Zielgröße. Die ICC ist das Verhältnis der Varianz zwischen den Gruppen einer Datenebene zur Gesamtvarianz [7] und somit der Anteil der Gesamtvarianz, der durch die Ebenenstruktur erklärt wird. Ist dieser Anteil „nahe“ Null, kann auf die explizite Modellierung der Datenstruktur (Mehrebenenmodell) verzichtet werden und ein Einebenenmodell führt zu unverzerrten Ergebnissen. In den einzelnen QS-Vergleichsgruppen liegen die ICC für die Ebene der Fachabteilungen und für die Ebene der Arbeitsmarktregionen unter fünf Prozent, so dass auf Mehrebenenmodelle verzichtet wurde (vgl. **Online-Anhang 3**).

## Selektion der Einflussgrößen

Als potenzielle Einflussgrößen auf den sozialmedizinischen Verlauf nach Rehabilitation wurden neben den in der RSD zur Verfügung stehenden soziodemographischen und gesundheitsbezogenen Merkmalen, Merkmalen des Reha-Zugangs und fachabteilungsspezifischen Merkmalen zusätzlich Arbeitsmarktindikatoren geprüft. Zudem wurden unter Einbezug medizinischer Experten medizinische Parameter geprüft, um z. B. die Bedeutung von Begleiterkrankungen aufzudecken und bei der Modellierung zwischen Untergruppen zu unterscheiden (z. B. in der Orthopädie „Gelenkersatz“ vs. „Rückenschmerz“).

Für die theoretisch vorselektierten Einflussgrößen (siehe **Online-Anhang 4**) erfolgt eine statistische Rückwärtsselektion der Einflussgrößen anhand des Signifikanzwerts  $p$  (zu Vorteilen gegenüber der Vorwärtsselektion vgl. [19]). Das Vorgehen ist dabei wie folgt: (1) Aufstellen eines Ausgangsmodells mit allen *potenziell* relevanten Einflussgrößen, (2) schrittweise Eliminierung statistisch zu vernachlässigender Variablen ( $p > 0,05$ ).

Um ein potenzielles Overfitting (d. h. eine Überanpassung des Modells aufgrund zu vieler berücksichtigter Einflussgrößen) sowie die Stabilität der Schätzer in den Adjustierungsmodellen zu überprüfen, wurde eine Validierungsstrategie entwickelt, die eine k-fache-Kreuzvalidierung sowie Bootstrapping beinhaltet.

## Diskussion der Methoden und Ergebnisse

### Zielgröße

Beim sozialmedizinischen Verlauf steht die berufliche (Wieder-)Eingliederung der Rehabilitanden im Mittelpunkt. Dieses Kriterium lässt sich anhand der RSD-Daten unterschiedlich operationalisieren.

Alternative Operationalisierungen des sozialmedizinischen Verlaufs in Form der beruflichen (Wieder-)Eingliederung auf Basis der RSD bergen inhaltliche Möglichkeiten und Grenzen sowie statistische Herausforderungen. Die wichtigsten Alternativen sind:

- Anzahl der Tage mit sozialversicherungspflichtiger Beschäftigung (Arbeitsverhältnis und andere nichtselbstständige Beschäftigung, vgl. SGB IV)
- Höhe des Entgelts für sozialversicherungspflichtige Beschäftigungszeiten
- Rentenbeitragsart (z. B. Pflichtbeitragszeit aufgrund versicherungspflichtiger Beschäftigung oder Sozialleistungsbezug)

- Beitragsindex (gewichteter Anteil an Monaten mit sozialversicherungspflichtiger Beschäftigung, vgl. [9]).

Die Beitragsart und der daraus berechnete Beitragsindex liegen in der RSD mit Monatsbezug vor, sodass ein beliebiger Zeitpunkt/-raum und damit auch ein definierter Abstand zum Ende der Reha-Leistung ausgewertet werden kann. Die monatliche Kodierung des Beitragsstatus ist jedoch recht grob. Beispielsweise wird aufgrund von Verdrängungsregelungen Beschäftigung kodiert, sobald an einem Tag des Monats eine versicherungspflichtige Beschäftigung vorlag.

Die sozialversicherungspflichtigen Beschäftigungstage und das Entgelt liegen in der RSD mit Kalenderjahresbezug vor, weshalb kein einheitlicher Abstand zum Ende der Reha-Leistung bestimmbar ist. Die Angaben unterscheiden zudem nicht zwischen Teil- und Vollzeitarbeit. Die Verteilung der Anzahl der Beschäftigungstage weist spezielle Eigenschaften auf (U-förmig in geschlossenem Intervall, für die Gesamtgruppe siehe ► **Abb. 1**), die regressionsanalytisch herausfordernd sind. Das sozialversicherungspflichtige Beschäftigungsentgelt hat hingegen einen Gipfel bei 0 Euro und flacht nach rechts ab.

Der Expertenbeirat sprach sich für die sozialversicherungspflichtigen Beschäftigungstage als Operationalisierung der beruflichen (Wieder-)Eingliederung aus. Er folgte in seiner Empfehlung der Argumentation, dass eine punktbezogene Betrachtung des Erwerbsstatus zu kurz greife und die Beschäftigungstage dem Handlungsauftrag der Rehabilitation besser entsprächen als das Beschäftigungsentgelt. Zudem sind „Beschäftigungstage“ leichter zu interpretieren, da dort im Gegensatz zum Entgelt weder branchenspezifische oder regionale Gehaltsunterschiede noch die wöchentliche Arbeitszeit berücksichtigt werden müssen.

Hohe Interkorrelationen zwischen den unterschiedlichen Operationalisierungen (im ersten Jahr nach Reha: Tage mit Entgelt  $r_{\text{Spearman}} = 0,9068$ , mit Beitragsindex  $r_{\text{Spearman}} = 0,9113$ ) deuten darauf hin, dass aus statistischer Sicht mit alternativen Operationalisierungen der Zielgröße sehr ähnliche Adjustierungsergebnisse erwartbar sind.

Da sowohl für den einjährigen als auch den zweijährigen Nachbetrachtungszeitraum starke Argumente vorliegen, werden für beide Zeiträume Modelle berechnet und berichtet. Für die einjährige Zielgröße sprechen die schnellere Rückmeldung und die etwas höheren Modellgüten. Der Fokus auf Nachhaltigkeit der (Wieder-)Eingliederung spricht hingegen für die zweijährige Nachbeobachtung.

### Regressionsmethode

Für die Wahl einer angemessenen Regressionsmethode zur Modellierung der Zielgröße wurden unterschiedliche statistische Methoden jeweils in der Ein- und in der Mehrebenenstruktur exploriert. Eine Alternative zu der letztlich favorisierten Fractional Logit Regression (FLR) wären Modelle für Beta-Verteilungen mit zero/one-Inflation [20]. Bei dieser Regressionsmethode werden die Verteilung mit dem geschlossenen Intervall angemessen berücksichtigt und die zwei Gipfel an den Rändern durch die sog. zero/one-Inflation abgebildet. Dazu werden drei Regressionsgleichungen mit Link-Funktionen verwendet – jeweils für die beiden Ränder und für den Bereich dazwischen.

Beim Vergleich der FLR mit den zero/one-inflated-Beta-Modellen überzeugte erstere aber in den Bereichen Rechenzeiten, Interpretation, Praktikabilität und Vorhersagequalität: (1) Die Rechenzeiten insbesondere in der Mehrebenenstruktur waren deutlich kürzer. Die Berechnung der zero/one-inflated-Beta-Modelle benötigten bis zu 150 GB Hauptspeicher (RAM) und dauerten z.T. mehrere Tage, was einen Routineeinsatz bei der DRV Bund fraglich erscheinen lässt. (2) Die Operationalisierung und Interpretation einer Regressionsgleichung ist praktikabler als von drei Gleichungen. (3) Die vorhergesagten Werte der FLR waren empirisch dichter an den beobachteten Werten (operationalisiert über die Residuenquadratsumme).

Auch die Vorhersagen mit der FLR sind nicht perfekt. Da aber keine Anhaltspunkte auf systematische Verzerrungen bei einzelnen Fachabteilungen erkennbar waren, wurde letztlich die FLR für die weiteren Modellierungen verwendet.

## Ebenenstruktur der Daten

Aus den geringen Intraklassenkorrelationen (ICC) der Fachabteilungen und Arbeitsmarktregionen kann abgeleitet werden, dass der Einfluss der Personenmerkmale kaum von der Zuordnung der Rehabilitanden zu Fachabteilungen und Arbeitsmarktregionen abhängt. Das Adjustierungsergebnis ist demnach inhaltlich praktisch unabhängig von einer expliziten Modellierung der Ebenenstruktur. Trotzdem könnten für einzelne Personenmerkmale Gruppierungseffekte vorliegen, beispielsweise für arbeitslose Personen in Bezug auf den regionalen Kontext [4]. In den zwei größten QS-Vergleichsgruppen (Orthopädie und Psychosomatik) wurden daher für die Einflussgröße „Beschäftigungsstatus 3 Monate vor Reha-Beginn“ random slopes auf der Ebene der Arbeitsmarktregionen getestet, die allerdings statistisch nicht signifikant waren. Snijders und Bosker [6] empfehlen unabhängig davon, bei starken theoretischen Argumenten Interaktionen zwischen Variablen aus verschiedenen Strukturebenen (cross-level) zu testen. Die cross-level-Interaktion des Beschäftigungsstatus mit der regionalen Arbeitslosenquote war praktisch unbedeutend. Voraussetzung für beide Befunde war, dass weitere erwerbsbiografische Einflussgrößen im Modell enthalten waren. Dies lässt vermuten, dass der Einfluss des regionalen Kontextes in der individuellen Erwerbsbiografie bereits „eingepreist“ und damit implizit berücksichtigt ist. Diese Analysen sprechen für die Belastbarkeit der Koeffizientenschätzer im Einebenenmodell im vorliegenden Kontext im Vergleich zum Mehrebenenmodell.

## Selektion der Einflussgrößen

Aus statistischer Sicht gibt es angesichts der hohen Fallzahl in den QS-Vergleichsgruppen wenige Gründe, die Anzahl der Einflussgrößen zu beschränken. Ein solcher Grund wäre, dass der Einflussfaktor statistisch nicht signifikant ist und so zu Overfitting führen könnte. Ein weiterer Grund wäre, dass die Fallzahl in den Merkmalsausprägungen einzelner Einflussgrößen zu klein ist und so die Koeffizienten nur mit großer Unsicherheit geschätzt werden können. Daher wurde bei der Kategorisierung von Einflussgrößen auf ausreichend große Zellbesetzungen geachtet (wie z. B. das Zusammenfassen der drei Geschlechterkategorien zu „weiblich“ und „nicht weiblich“).

## Prädiktorenselktion

Aufgrund der hohen Fallzahlen in den QS-Vergleichsgruppen werden auch kleine Effektstärken statistisch signifikant. Daher wurden

mit anderen Selektionskriterien (Effektstärken anhand der AME) sparsamere Modelle entwickelt. Diese zeigten eine ähnlich hohe Varianzaufklärung (Pseudo-R<sup>2</sup>) wie die weniger sparsamen finalen Modelle, so dass eine Beschränkung auf wenige Einflussgrößen statistisch vertretbar wäre. Für einzelne Fachabteilungen, die diesbezüglich abweichende Verteilungen gegenüber der Vergleichsgruppe aufweisen, könnte das Weglassen solcher Einflussfaktoren aber zu Verzerrungen im Adjustierungsergebnis führen. Dies gilt insbesondere dann, wenn die Qualitätspunkte einzelner Fachabteilungen nahe beieinander oder in der Nähe der Schwellenwerte zum strukturierten Qualitätsdialog (SQD) liegen. Die Prädiktorenauswahl erfolgte daher anhand des *p*-Wertes und ohne weitere Selektionskriterien.

Im Rahmen der Prädiktorenselktion stellte sich die Frage, inwieweit Interaktionen von Einflussgrößen die Adjustierung verbessern. Dass die theoretisch relevante Interaktion von individueller mit regionaler Arbeitslosigkeit statistisch unauffällig ist, ist partiell methodisch begründet, weil logistische Modelle infolge der Nichtlinearität weniger sensitiv für Interaktionen sind [15]. Interaktionen sind damit implizit zu einem gewissen Grad abgedeckt, so dass auf die explizite Testung weiterer Interaktionen verzichtet wurde.

## Modellvalidierung

Die ausgewählten Modelle sind in ihrer Modellgüte auf den vorhandenen Datensatz optimiert, weshalb die Leistungsfähigkeit dieser Modelle bei neuen Daten überprüft werden sollte. Dazu gibt es externe und interne Validierungsverfahren. Bei der externen Validierung wird das entwickelte Modell auf neue unabhängige Daten übertragen. Interne Validierungsverfahren beruhen auf dem Originaldatensatz, der auch für die Modellentwicklung genutzt wurde (für einen Überblick siehe z. B. Muche et al. [19]). Eine externe Validierung wurde aufgrund eingeschränkter Übertragbarkeit nicht durchgeführt, da relevante Einflussgrößen in vorherigen Jahrgängen teilweise noch nicht vorhanden waren.

Die Befunde der internen Modellvalidierung legen nahe, dass die Modelle in hohem Maße valide sind. Von einer Korrektur der einzelnen Regressionsschätzer wurde abgesehen, da sich die gemittelten Bootstrapping-Schätzer (d. h. Koeffizientenschätzer nach mehrfachem Stichprobenziehen) nur sehr geringfügig von den Schätzern in den finalen Adjustierungsmodellen unterscheiden (z. B. Schätzer „Beitragsart 3 Monate vor Reha“:  $\text{Sonstiges}_{\text{Bootstrap}} = -1,7313$ ,  $\text{Sonstiges}_{\text{Final}} = -1,7326$ ;  $\text{Sozialleistung}_{\text{Bootstrap}} = -0,8188$ ,  $\text{Sozialleistung}_{\text{Final}} = -0,8297$ ; Referenz: Beschäftigung), gleichzeitig aber auf kleineren Stichproben beruhen. Die Ergebnisse der *k*-fachen-Kreuzvalidierung deuten nicht auf Verzerrungen aufgrund von zu vielen oder zu wenigen berücksichtigten Einflussgrößen hin, daher wird die Prädiktorenauswahl mittels Rückwärtsselektion als angemessen angesehen.

Eine regelmäßige Anpassung der Modelle scheint trotzdem sinnvoll, da sich der Einfluss einzelner Merkmale auf die berufliche (Wieder-)Eingliederung aufgrund wirtschaftlicher und (welt-)politischer Entwicklungen verschieben könnte und weitere bisher vernachlässigte Faktoren an Relevanz gewinnen könnten.

Trotz hoher Modellgüten verbleibt eine Restvarianz, von der unklar ist, welcher Anteil auf Behandlungsqualität der Fachabteilung – dieser Anteil sollte möglichst hoch sein – und welcher auf ausgelassene, aber relevante Einflussgrößen zurückzuführen ist. Bei-

spielsweise ist unbekannt, welche Suchanstrengungen eine arbeitslose Person unternimmt, um Arbeit zu finden. Da sich solche unbekannt Einflussgrößen über alle Fachabteilungen ähnlich verteilen dürften, bilden sie Zufallsstörgrößen mit ähnlichen Effekten auf die Ergebnisse der Fachabteilungen. Somit bleibt die vergleichende Qualitätsbewertung der Fachabteilungen unproblematisch, da keine systematisch bevorzugt oder benachteiligt wird.

Trotz sorgfältiger Auswahl der Adjustierungsparameter können aber nicht alle möglichen „Praxisbesonderheiten“ abgebildet werden – dies ist bei der Interpretation von adjustierten Werten zu berücksichtigen.

Dennoch bilden die adjustierten Werte eine deutliche bessere Grundlage für einen sachgerechten Vergleich der Fachabteilungen als die Rohwerte.

### KERNBOTSCHAFT

Das entwickelte Adjustierungsverfahren für den sozialmedizinischen Verlauf nach medizinischer Rehabilitation ermöglicht einen sachgerechten Fachabteilungsvergleich zur Qualitätsbeurteilung der Behandlungsergebnisse im Rahmen der Reha-QS. Vorbestehende Unterschiede bei Rehabilitandenmerkmalen zwischen den Fachabteilungen mit Einfluss auf den sozialmedizinischen Verlauf der Rehabilitanden nach der Maßnahme wurden mithilfe der entwickelten Adjustierungsstrategie so gut wie möglich kompensiert.

Für die methodischen Herausforderungen (Verteilungsform der Zielgröße, Mehrebenenstruktur, Selektion der Prädiktoren) wurden angemessene Lösungen gefunden.

## Danksagung

Wir danken der Projektgruppe des Dezernats 0430 „Reha-Qualitätssicherung, Epidemiologie und Statistik“ der Deutschen Rentenversicherung Bund: Frau Mitschele, Frau Kranzmann, Frau Schmidt und Herrn Dr. Thiede sowie den Mitgliedern des Expertenbeirats des Projekts für wertvolle inhaltliche Beiträge. Ferner danken wir allen Teilnehmern der Fokusgruppen für die Anregungen bezüglich der Gestaltung des Rückmeldekonzpts.

## Interessenkonflikt

Die Autor\*innen geben an, dass kein Interessenkonflikt besteht.

## Literatur

- [1] Deutsche Rentenversicherung Bund. Rahmenkonzept zur medizinischen Rehabilitation in der gesetzlichen Rentenversicherung. Im Internet: [https://www.deutsche-rentenversicherung.de/SharedDocs/Downloads/DE/Experten/infos\\_reha\\_einrichtungen/konzepte\\_systemfragen/konzepte/rahmenkonzept\\_medizinische\\_reha.html](https://www.deutsche-rentenversicherung.de/SharedDocs/Downloads/DE/Experten/infos_reha_einrichtungen/konzepte_systemfragen/konzepte/rahmenkonzept_medizinische_reha.html); Stand: 17.11.2020
- [2] Becker A, Stausberg J, Fischer B et al. Risikoadjustierung von Qualitätsindikatoren. *Das Krankenhaus* 2016; 11: 954–963
- [3] Gregorkiewicz A, Sägebarth M. Überarbeitung der Adjustierung in der Rehabilitandenbefragung der DRV. *DRV-Schriften* 2020; 29: 41–43
- [4] Hetzel C, Streibelt M. Hängt die berufliche Wiedereingliederung nach beruflichen Bildungsleistungen vom Arbeitsmarkt ab? *Rehabilitation (Stuttg)* 2016; 55: 290–298. doi:10.1055/s-0042-113932
- [5] Streibelt M, Egner U. Eine Meta-Analyse zum Einfluss von Stichprobe, Messmethode und Messzeitpunkt auf die berufliche Wiedereingliederung nach beruflichen Bildungsleistungen. *Rehabilitation (Stuttg)* 2012; 51: 398–404. doi:10.1055/s-0031-1291283
- [6] Snijders TAB, Bosker RJ. *Multilevel analysis. An introduction to basic and advanced multilevel modeling.* Sage 1999
- [7] Hox JJ. *Multilevel analysis: Techniques and applications.* 2nd ed. New York, NY: Routledge; 2010
- [8] Rasbash J, Goldstein H. Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model. *Journal of Educational and Behavioral Statistics* 1994; 19: 337. doi:10.2307/1165397
- [9] Nübling R, Kaluscha R, Krischak G et al. Return to Work nach stationärer Rehabilitation – Varianten der Berechnung auf der Basis von Patientenangaben und Validierung durch Sozialversicherungsbeitragszahlungen. *Phys Med Rehab Kuror* 2016; 26: 293–302. doi:10.1055/s-0042-117282
- [10] Babyak MA. What You See May Not Be What You Get. A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. *Psychosomatic Medicine* 2004; 66: 411–421
- [11] Deutsche Rentenversicherung. Datensatz zur Reha-Statistik-Datenbasis (RSD) ab Berichtsjahr 2015 (11.12.2015). Im Internet: [http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/getResource.do?key=sk\\_94\\_ab\\_2015.pdf](http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/getResource.do?key=sk_94_ab_2015.pdf); Stand: 31.11.2020
- [12] Bundesinstitut für Bau-, Stadt- und Raumforschung. INKAR. Indikatoren und Karten zur Raum- und Stadtentwicklung (05.2020). Im Internet: <https://www.inkar.de/>
- [13] Wedderburn RWM. Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika* 1974; 61: 439. doi:10.2307/2334725
- [14] Papke L, Wooldridge J. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics* 1996; 6: 619–632
- [15] Best H, Wolf C. Logistische Regression. In: Wolf C, Best H, Hrsg. *Handbuch der sozialwissenschaftlichen Datenanalyse.* Wiesbaden: VS Verlag für Sozialwissenschaften/Springer Fachmedien Wiesbaden GmbH Wiesbaden; 2010: 827–854. doi:10.1007/978-3-531-92038-2\_31
- [16] Williams R. Marginal Effects for Continuous Variables. Im Internet: <https://www3.nd.edu/~rwilliam/stats3/Margins02.pdf>
- [17] McKelvey RD, Zavoina W. A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology* 1975; 4: 103–120. doi:10.1080/0022250X.1975.9989847
- [18] Veall MR, Zimmermann KF. Evaluating Pseudo-R<sup>2</sup>'s for binary probit models. *Qual Quant* 1994; 28: 151–164. doi:10.1007/BF01102759
- [19] Muche R, Ring C, Ziegler C. Entwicklung und Validierung von Prognosemodellen auf Basis der logistischen Regression. [Zugl.: Ulm, Univ., Habil.-Schr., 2004]. Aachen: Shaker; 2005
- [20] Abdel-Karim AH. Extended zero-one inflated beta and adjusted three-part regression models for proportional data analysis. *Communications in Statistics – Simulation and Computation* 2016; 46: 6155–6172. doi:10.1080/03610918.2016.1197248