



**Ergebnismessung in der neurologischen
Rehabilitation des Schlaganfalls mit PROMs**

Abschlussbericht der Pilotstudie

30.05.2024

Impressum

Autor:innen

Wissenschaftlicher Bericht:

Dipl.-Psych. Volker Beierlein, Universitätsklinikum Hamburg-Eppendorf

Prof. Dr. Holger Schulz, Universitätsklinikum Hamburg-Eppendorf

Vorwort und Ko-Autorinnenschaft Kapitel „Hintergrund & Ziele“:

Annabelle Neudam, M.Sc. Health Care Management, 4QD – Qualitätskliniken.de GmbH

Zusammenfassung:

Gina-Sophie Labahn, M.Sc. Public Health and Administration, 4QD – Qualitätskliniken.de GmbH

Redaktion

Annabelle Neudam, M.Sc. Health Care Management, 4QD – Qualitätskliniken.de GmbH

Gina-Sophie Labahn, M.Sc. Public Health and Administration, 4QD – Qualitätskliniken.de GmbH

Mitwirkende

Dr. Ralf Bürgy, MEDICLIN AG

Dr. Torsten Kirsch, VAMED Gesundheit Holding Deutschland GmbH

Johannes Kneißl, M.Sc. Soziologie, MP Medical Park Holding SE

Prof. Dr. Matthias Köhler, VAMED Gesundheit Holding Deutschland GmbH

Alexander Mühlhause, M.A. Business-Health Management, 4QD-Qualitätskliniken.de GmbH

Friederike Preuß, B.A. Management im Gesundheitstourismus, 4QD-Qualitätskliniken.de GmbH

Robert Simon, M.A. Soziologie, Deutsche Rentenversicherung Bund

Besonderer Dank für Beiträge zur Diskussion der Ergebnisse

Dr. med. Michael Annas, MEDICLIN Hedon Klinik

Dr. med. Martin Falkenberg, ZAR Bielefeld

Dr. med. Tobias Knieß, Rhön-Klinikum Bad Neustadt

Prof. Dr. med. habil. Marcus Pohl, VAMED Klinik Schloss Pulsnitz

Dr. med. Christoph Schäfer, VAMED Klinik Hagen-Ambrock

Prof. Dr. med. Tobias Wächter, Passauer Wolf Reha-Zentrum Bad Gögging

Herausgeber

4QD – Qualitätskliniken.de GmbH

Friedrichstraße 60

10117 Berlin

E-Mail: info@qualitaetskliniken.de

Internet: www.qualitaetskliniken.de

Urheber- und Nutzungsrecht

Eine kommerzielle Nutzung des Dokuments ist ausgeschlossen. Alle Inhalte sind urheberrechtlich geschützt. Text und Gestaltung dürfen nicht verändert werden.

Zitation

Beierlein V., Neudam A., Labahn GS., Bürgy R., Kirsch T., Kneißl J., Köhler M., Mühlhause A., Preuß F. & Schulz H. (2024). Ergebnismessung in der neurologischen Rehabilitation des Schlaganfalls mit PROMs: Abschlussbericht der Pilotstudie. Berlin: Qualitätskliniken.de

Zusammenfassung

Informationen für Patient:innen und Behandler:innen

Im Rahmen dieser Studie wurden Schlaganfallpatient:innen in der Rehabilitation untersucht. Ziel war es aufzuzeigen, wie eine Verbesserung des Gesundheitszustandes gemessen werden kann. Die teilnehmenden Rehabilitationseinrichtungen wurden miteinander verglichen, um festzustellen, welche Einrichtungen überdurchschnittliche Ergebnisse erzielen und welche eher unterdurchschnittlich abschneiden. Zur Messung und zum Vergleich der Veränderungen des allgemeinen Gesundheitszustandes wurden zwei generische, ähnliche Messinstrumente, PROMIS-10 und VR-12, eingesetzt. Die Verwendung ähnlicher Instrumente ermöglicht es, diese und ihre Eignung für das Krankheitsbild zu vergleichen. Zusätzlich wurden soziodemographische und medizinische Daten erhoben, um einen fairen Vergleich der Rehabilitationseinrichtungen zu gewährleisten.

Die Messungen wurden zu Beginn der Rehabilitation und drei Monate nach der Aufnahme im häuslichen Umfeld durchgeführt. In Rücksprache mit Expert:innen aus der Neurologie wurde der Zeitpunkt nach der Rehabilitation auf drei Monate festgelegt, da nach diesem Zeitraum positive Veränderungen mit hoher Wahrscheinlichkeit adäquat abgebildet werden können.

Die Ergebnisse zeigen, dass die Beeinträchtigungen der behandelten Patient:innen zu Beginn der Rehabilitation stark variierten. Ein Drittel der Patient:innen wies zu Beginn der Rehabilitation eine stärkere Beeinträchtigung auf, ein Drittel befand sich in einem durchschnittlichen Gesundheitszustand ohne nennenswerte Einschränkungen und ein Drittel in einem überdurchschnittlich guten Gesundheitszustand. Nach drei Monaten zeigte sich in der Gruppe mit stärkeren Beeinträchtigungen eine deutliche Verbesserung des Gesundheitszustandes, während sich der Gesundheitszustand in der Gruppe mit geringen Einschränkungen weitgehend unverändert zeigte. In der Gruppe mit überdurchschnittlich gutem Gesundheitszustand ist insgesamt eine Verschlechterung zu beobachten, die darauf zurückzuführen sein könnte, dass sich die Wahrnehmung der eigenen Gesundheit wieder dem Durchschnitt annähert.

In den Ergebnissen zeigen sich sogenannte Deckeneffekte, deren Ursachen vielfältig sind. Sie können in Unterschieden in der Patientenselektion, im Befragungssetting, in der Zuweisungsselektion und in der Selbstselektion begründet sein. Diese Faktoren können zu systematischen Verzerrungen führen. Die Ergebnisse legen nahe, dass PROMs wertvolle Instrumente sind, jedoch müssen methodische Herausforderungen berücksichtigt werden, um valide und vergleichbare Ergebnisse zu erzielen.

Wir danken allen Beteiligten, insbesondere den Patient:innen und den teilnehmenden Rehabilitationseinrichtungen, für ihre wertvolle Mitarbeit und Unterstützung. Gemeinsam leisten wir einen entscheidenden Beitrag zur Weiterentwicklung und Verbesserung der neurologischen Rehabilitation.

Inhalt

Impressum	2
Zusammenfassung	3
Informationen für Patient:innen und Behandler:innen	3
Vorwort	6
Hintergrund und Zielsetzung des Pilotprojekts zur Ergebnismessung in der neurologischen Rehabilitation des Schlaganfalls	7
Methodik	10
Umgang mit Interessenkonflikten	10
Instrumente	11
Datenerhebung	12
Auswertung des erreichten beobachtbaren Outcomes	12
Risikoadjustierung.....	13
Analyse von Deckeneffekten (a posteriori)	15
Erweiterung der Risikoadjustierung durch stratifizierte multiple Regressionsmodelle (a posteriori)	16
Auswertung der Outcomes nach der Risikoadjustierung	16
Vergleich von Rehabilitationseinrichtungen mit dem Patient-Reported Outcome Quality Index (ProQI)	17
Vergleich der Änderungssensitivität von PROMIS-10 und VR-12.....	18
Ergebnisse	19
Rücklauf.....	19
Patient:innenmerkmale	20
Ergebnisqualität im PROMIS-10 (beobachtete Outcomes)	21
Ergebnisqualität im VR-12 (beobachtete Outcomes).....	24
Deckeneffekte im PROMIS-10 und VR-12.....	26
PROMIS-10: Vergleich der Einrichtungen nach stratifizierter Risikoadjustierung	33
VR-12: Vergleich der Einrichtungen nach stratifizierter Risikoadjustierung	35
Vergleich von Rehabilitationseinrichtungen mit dem Patient-Reported Outcome Quality Index (ProQI).....	36

Vergleich der Änderungssensitivität von PROMIS-10 und VR-12.....	39
Diskussion	43
Anhang.....	55
Risikoadjustierungsmodelle für PROMIS-10 PH	55
Risikoadjustierungsmodelle für PROMIS-10 MH	57
Risikoadjustierungsmodelle für VR-12 PCS.....	59
Risikoadjustierungsmodelle für VR-12 MCS	61
Analyse der Linearität des Zusammenhangs von Aufnahmewert zu Katamnesewert im PROMIS-10 und der Veränderungen der Rangreihen im Einrichtungsvergleich in Abhängigkeit des Risikoadjustierungsmodells.....	63
Vergleichsdaten für den PROMIS-10 aus fünf Einrichtungen der Medical Park Gruppe mit zusätzlicher Messung zur Entlassung aus der Rehabilitation	69
Literatur	70

Vorwort

Die Messung von Behandlungsergebnissen aus der Sicht von Patientinnen und Patienten, sogenannte Patient-Reported Outcome Measures (PROMs), gewinnt auch in der neurologischen Rehabilitation weltweit an Bedeutung. PROMs bieten eine wertvolle Ergänzung zu klinischen Parametern, indem sie direkte Einblicke in die subjektive Wahrnehmung von Symptomen, Funktionsfähigkeit und gesundheitsbezogener Lebensqualität ermöglichen. Diese patientenzentrierten Daten leisten einen wichtigen Beitrag zur Qualitätsbewertung und -verbesserung in der Gesundheitsversorgung. Genau hier setzt die vorliegende Pilotstudie zur Ergebnismessung in der neurologischen Rehabilitation von Schlaganfallpatient:innen an.

Im Rahmen einer Kooperation zwischen dem Universitätsklinikum Hamburg-Eppendorf und dem REHAPORTAL wurde untersucht, inwiefern sich die Gesundheit von Patient:innen vor und nach einer Rehabilitation mittels standardisierter PROMs bewerten lässt. Diese Studie stellt einen wichtigen Schritt zur Etablierung international vergleichbarer Standards in der Ergebnismessung dar, die einen fairen und objektiven Vergleich von Rehabilitationseinrichtungen ermöglichen.

Ein zentrales Element der Studie ist die öffentliche Berichterstattung auf der etablierten Plattform DAS REHAPORTAL. Das unabhängige Portal macht die Qualität von Rehakliniken öffentlich und unterstützt so Patient:innen und Angehörige bei der Wahl der passenden Einrichtung. Anhand Informationen zur Qualität, zur Ausstattung und zum Leistungsangebot können die Nutzer:innen die teilnehmenden Kliniken miteinander vergleichen. Die Bewertungen basieren auf Daten zur Behandlungsqualität, Patientensicherheit, Patientenzufriedenheit und Organisationsqualität. Durch die Integration der PROMs in das REHAPORTAL wird die Ergebnisqualität noch detaillierter und patienten-zentrierter abgebildet.

Unser Ziel ist es, gemeinsam mit Gesundheitseinrichtungen und Patient:innen einen international vergleichbaren Standard zu entwickeln und dadurch sowohl Behandler:innen als auch Patient:innen differenzierte Einschätzungen zu ermöglichen. Für Rehakliniken ergeben sich dadurch Möglichkeiten eines risikoadjustierten Benchmarkings und der Identifikation von ergebnisrelevanten Einflussfaktoren.

Herzliche Grüße



Annabelle Neudam

Geschäftsführerin
DAS REHAPORTAL

Hintergrund und Zielsetzung des Pilotprojekts zur Ergebnismessung in der neurologischen Rehabilitation des Schlaganfalls

Die Erfassung der Ergebnisqualität medizinischer Leistungen durch die Befragung von Patient:innen zu ihrem subjektiv wahrgenommenen Gesundheitszustand erfährt weltweit steigende Bedeutung. Von Patient:innen selbst berichtete Informationen bzw. Ergebnisse zu ihrem Gesundheitszustand (Englisch: Patient-Reported Outcomes; **PROs**) beinhalten beispielsweise wahrgenommene Symptome, Einschränkungen der eigenen Funktionsfähigkeit oder die Wahrnehmung der eigenen gesundheitsbezogenen Lebensqualität. PROs komplementieren somit apparativ messbare klinische Parameter (wie z. B. Blutdruck, Motilität von Gelenken, kognitive Funktionsfähigkeit usw.) bei der Beurteilung der Gesundheit eines Menschen. Die zunehmende Fokussierung auf PROs bei der Evaluation von Leistungen des Gesundheitswesens soll sicherstellen, dass die für Patient:innen tatsächlich wahrnehmbaren Auswirkungen einer Behandlung relevante Endpunkte (= Outcomes) im Qualitätsmanagement und bei der Weiterentwicklung der Gesundheitsversorgung darstellen. Auch im Bereich der Rehabilitation steht die Erfassung der Ergebnisqualität mittels PROs laut aktueller gemeinsamer Empfehlung der Bundesarbeitsgemeinschaft für Rehabilitation (BAR) im Zentrum der Qualitätssicherung. Als Goldstandard dient dabei die Mehrpunkterhebung des Gesundheitszustands, mit Zeitpunkten zur Aufnahme und zur Entlassung, sowie fakultativ einem katamnestischen Zeitpunkt, um Veränderungen der Gesundheit über den Behandlungsverlauf bzw. Ergebnisse zur erreichten Stabilität des Behandlungserfolgs abbilden zu können.

Ein wichtiges Instrument zur Entwicklung von Qualität und Transparenz im Gesundheitswesen ist das Public Reporting der Leistungsdaten von Einrichtungen der Gesundheitsversorgung. Durch die Veröffentlichung von Daten zu Struktur-, Prozess- und Ergebnisqualität können Versorgungseinrichtungen wichtige Hinweise und Impulse zur Optimierung ihrer Leistungen erhalten. Daneben wird dadurch Patient:innen und zuweisenden Institutionen ermöglicht, Versorgungseinrichtungen hinsichtlich verschiedener Qualitätsmerkmale miteinander zu vergleichen und so eine informierte Entscheidung darüber zu treffen, wo eine auf die individuellen Präferenzen und Anforderungen optimierte Gesundheitsversorgung erwartet werden kann.

Für das Public Reporting der bzw. zur Einholung von Informationen zu Leistungsdaten von Rehabilitationseinrichtungen sind verschiedene Plattformen verfügbar. Eine seit 2013 etablierte Internetplattform ist [DAS REHAPORTAL](#) der 4QD – Qualitätskliniken.de GmbH. Zur transparenten Darstellung der Qualität des Dienstleistungsangebots seiner aktuell mehr als 300 Mitgliedseinrichtungen bildet das Internetportal der Qualitätskliniken.de Qualitätsindikatoren aus vier unterschiedlichen Dimensionen ab. Neben der Präsentation von Daten zu Patientensicherheit, Patientenzufriedenheit und Organisationsqualität ist ein Schwerpunkt die Berichterstattung zur Dimension der Behandlungsqualität, die die Ergebnisse der QS-Verfahren der Kostenträger beinhaltet.

In einem Kooperationsprojekt der Professur für Klinische Versorgungsforschung am Institut für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf (Dipl.-Psych. Volker Beierlein, Prof. Dr. Holger Schulz) und der 4QD – Qualitätskliniken.de GmbH wird seit 2018 die ergänzende Implementierung von standardisierten Instrumenten zur Erfassung von PROs (Englisch: Patient-Reported Outcome Measures; **PROMs**) als Indikatoren der Ergebnisqualität der medizinischen Rehabilitation wissenschaftlich in mehreren Pilotstudien innerhalb unterschiedlicher Reha-Fachabteilungen evaluiert. Die Auswahl geeigneter PROMs orientiert sich hier an den Empfehlungen der internationalen Initiative **ICHOM** (International Consortium for Health Outcomes Measurement), die zur Vereinheitlichung der Ergebnismessung bei den am häufigsten auftretenden Erkrankungen sogenannte Standard Sets von PROMs vorgeschlagen hat. Der Einsatz dieser PROMs ermöglicht somit einen Vergleich mit einer Vielzahl von Behandlungsangeboten weltweit, die mit denselben Instrumenten evaluiert wurden und kann dazu beitragen, einen international vergleichbaren Standard zur Evaluation der Ergebnisqualität von Rehabilitationsangeboten zu entwickeln. Zentrale Ziele der bereits durchgeführten Pilotstudien waren dabei

- Veränderungen in der Gesundheit von Patient:innen im Rahmen der Rehabilitationsbehandlung anhand der eingesetzten PROMs durch **Effektstärkemaße** zu beschreiben,
- die Ergebnisqualität von Rehabilitationseinrichtungen durch Berücksichtigung einrichtungsspezifischer Patient:innenmerkmale (**Case-Mix**), die mit dem Behandlungsergebnis im Zusammenhang stehen, fair miteinander zu vergleichen (**risikoadjustiertes Benchmarking**) und
- einen Indexwert zu entwickeln, mit dem ein fairer Vergleich der Ergebnisqualität zwischen den Rehabilitationseinrichtungen anhand eines einzigen Kennwertes möglich ist (Patient-Reported Outcome Quality Index; **ProQI**).

Die bisher untersuchten Reha-Fachabteilungen umfassten dabei die [Orthopädie](#) mit den Indikationen Hüft-TEP, Knie-TEP und chronischer Rückenschmerz (2018-2020), [Psychosomatik](#) mit den Indikationen Depressive Erkrankungen und Angsterkrankungen (2019-2021) sowie [Onkologie](#) mit den Indikationen Mamma-Karzinom, Prostata-Karzinom und Kolon-Rektum-Karzinom (2020-2022). Die in diesen Reha-Fachabteilungen bzw. Indikationen eingesetzten PROMs erwiesen sich dabei als ausreichend änderungssensitiv, um die erzielten Veränderungen des Gesundheitszustands im Sinne der Ergebnisqualität vergleichend zwischen den beteiligten Rehabilitationseinrichtungen darzustellen.

Bei der hier vorliegenden Studie werden im Rahmen des Kooperationsprojekts erstmals PROMs in der Reha-Fachabteilung Neurologie der Mitgliedskliniken des REHAPORTALS eingebunden. Es soll analysiert werden, welche Ausprägungen von Effektstärken der Veränderungen der gesundheitsbezogenen Lebensqualität sich bei der Rehabilitation des Schlaganfalls anhand der ausgewählten PROMs beobachten lassen und in welchem Maße teilnehmende Einrichtungen

die aufgrund ihres Case-Mix zu erwartenden Veränderungen erreichen bzw. nicht erreichen. Zudem soll für die Einrichtungen der Indexwert zur vergleichenden Darstellung der erreichten Ergebnisqualität (ProQI) bei der Rehabilitation des Schlaganfalls berechnet werden. Für die Erhebung wurden zwei generische PROMs zur Erfassung der allgemeinen gesundheitsbezogenen Lebensqualität ausgewählt, wobei das erste PROM der ICHOM-Empfehlung für das Monitoring der Behandlung des Schlaganfalls entspricht und das zweite, von den teilnehmenden Einrichtungen optional eingesetzte PROM als generisches Instrument bereits in den Pilotstudien in der Orthopädie, Psychosomatik und Onkologie eingesetzt wurde. Für den Teil der Einrichtungen, die beide Instrumente eingesetzt hatten, soll die Änderungssensitivität beider PROMs miteinander verglichen werden. Aus den Ergebnissen der Analysen sollen Empfehlungen für nachfolgende Erhebungen bei der Rehabilitation des Schlaganfalls für das Public Reporting abgeleitet werden.

Methodik

Das Studiendesign für diese Pilotstudie wurde im Konsens der Kooperationspartner, der 4QD – Qualitätskliniken.de GmbH mit den Repräsentanten ihrer Mitgliedseinrichtungen (insbesondere Experten der AG Ergebnismessung) einerseits und der Professur für Klinische Versorgungsforschung am Institut für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf (UKE) andererseits, entwickelt.

In regelmäßigen gemeinsamen Sitzungen wurden zunächst die Erhebungsinstrumente, Einschlusskriterien, Mindestfallzahlen und die Messzeitpunkte festgelegt. Die Festlegung auf die Indikation Schlaganfall innerhalb des Reha-Fachbereichs Neurologie wurde davon geleitet, durch die Auswahl einer Indikation mit hohen Behandlungszahlen den Datenerhebungszeitraum möglichst kurz zu halten und dadurch den Erhebungsaufwand zu verringern, um so möglichst vielen Mitgliedseinrichtungen des REHAPORTALS eine Teilnahme an der Pilotstudie zu ermöglichen. Das Konzept zur Datenerhebung und die Auswertungsstrategien aus den vorangegangenen Pilotstudien in der Orthopädie, Psychosomatik und Onkologie wurden weiterentwickelt und an die Befragung von Schlaganfallpatient:innen in der neurologischen Rehabilitation angepasst. Für die teilnehmenden Rehabilitationseinrichtungen wurde ein Manual zur Datenerhebung erstellt, um die Implementierung der Datenerhebung in den Einrichtungen zu erleichtern und eine optimale Datenqualität sicherzustellen.

Während das UKE das Projekt im gesamten Prozess wissenschaftlich beratend begleitete, nahm es zudem die Position der Auswertungsstelle der Daten ein, erstellte zur Sicherstellung der Datenqualität Berichte zur Fehlerfreiheit der erhobenen Daten zur Weitergabe an die teilnehmenden Rehabilitationseinrichtungen, führte alle statistischen Analysen durch, erstellte zu den Ergebnissen Einzelberichte für die teilnehmenden Mitgliedseinrichtungen sowie den hier vorliegenden zusammenfassenden Gesamtbericht, und übermittelte Daten zum Einrichtungsvergleich zur Nutzung im REHAPORTAL an die 4QD – Qualitätskliniken.de GmbH.

Die 4QD – Qualitätskliniken.de GmbH stand als Ansprechpartnerin bei allen erhebungsbezogenen Fragen den teilnehmenden Rehabilitationseinrichtungen zur Verfügung,

Ein Repräsentant der Deutschen Rentenversicherung Bund nahm an den Projekt-Meetings beobachtend teil.

Umgang mit Interessenkonflikten

Das UKE führt seine Projektaufgaben im Sinne einer selbst-initiierten Studie (Investigator Initiated Study; **IIS**) innerhalb der Versorgungsforschung als Beitrag zur Weiterentwicklung der Qualitätssicherung und des Public Reportings in der Rehabilitation ohne kommerzielles Interesse durch. Für die Durchführung seiner Projektaufgaben erhielt das UKE Mittel zur Finanzierung des an der Studie beteiligten wissenschaftlichen Personals von der 4QD –

Qualitätskliniken.de GmbH. Die Finanzierung erfolgte dabei immer losgelöst von den ergebnisoffen angelegten Projektzielen, so dass der wissenschaftliche Prozess frei und unabhängig stattfinden konnte sowie die vom UKE produzierten Ergebnisse von finanziellen Interessen unbeeinflusst geblieben sind.

Die an der Konzipierung und Durchführung der Studie beteiligten Experten der AG Ergebnismessung sind Repräsentanten verschiedener z. T. deutschlandweit agierender Rehabilitationseinrichtungen, die auch Mitgliedseinrichtungen im REHAPORTAL der 4QD – Qualitätskliniken.de sind. Durch die Studienbeteiligung unterschiedlicher Wettbewerber im Rehabilitationsbereich einerseits und durch den alleinigen Fokus auf die wissenschaftlichen Aspekte der Studie andererseits war ausgeschlossen, dass durch die Studienergebnisse einzelne Mitgliedseinrichtungen unrechtmäßig bevorteilt werden konnten.

Instrumente

Für die Erfassung der gesundheitsbezogenen Lebensqualität der Patient:innen wurde als primäres Outcome die 10-Item PROMIS Scale v1.2 – Global Health (**PROMIS-10**; Hays et al., 2009) ausgewählt, mit dem die physische und die psychische Dimension des von den Patient:innen subjektiv wahrgenommenen Gesundheitszustands auf einer bevölkerungsnormierten T-Wertskala (M = 50; SD = 10) gemessen wird.

Optional konnte von den Einrichtungen als weiteres generisches Instrument zur Messung der gesundheitsbezogenen Lebensqualität der kostenfrei nutzbare Veterans RAND 12-Item Health Survey (**VR-12**; Buchholz & Kohlmann, 2015; vergleichbar mit dem bekannten SF-12) mit den beiden Summenskalen körperliche (PCS) und psychische Lebensqualität (MCS) eingesetzt werden, dessen Ergebnisse ebenfalls auf einer bevölkerungsnormierten T-Wert-Skala abgebildet werden. Auch wenn sich PROMIS-10 und VR-12 hinsichtlich ihres Konzeptes, ihrer übergeordneten Dimensionen körperlicher und psychischer gesundheitsbezogener Lebensqualität und in ihren Formulierungen der operationalisierten Indikatoren subjektiv wahrgenommener Gesundheit überlappen, können sich dennoch die Outcomes dieser PROMs komplementär ergänzen und darüber hinaus wertvolle Hinweise für die zukünftige Anwendung geeigneter Instrumente zur Erfassung der Ergebnisqualität einer Rehabilitation von Schlaganfallpatient:innen liefern. Von den 17 teilnehmenden Einrichtungen haben fünf Einrichtungen den VR-12 zusätzlich erhoben.

Des Weiteren wurden **soziodemografische und (sozial-)medizinische Angaben**, die bekanntermaßen bzw. potentiell einen Einfluss auf die Höhe des Outcomes haben bzw. haben können, für eine statistische Risikoadjustierung erhoben. Durch die Risikoadjustierung kann ein fairer Vergleich der Ergebnisqualität zwischen den Einrichtungen gewährleistet werden. Die Verwendung dieser Angaben in den Analysen wird im Abschnitt zur Risikoadjustierung detailliert beschrieben.

Datenerhebung

Die Datenerhebung für die Indikation Schlaganfall erfolgte zwischen Anfang Juli 2021 und Ende September 2022 in insgesamt 17 Einrichtungen der neurologischen Rehabilitation. Bei allen teilnehmenden Einrichtungen handelt es sich um Mitgliedseinrichtungen im REHAPORTAL. Patient:innen, die eine Rehabilitationsbehandlung als Anschlussheilbehandlung nach einem Akutereignis in der Rehabilitationsphase D^a durchführten und denen eine Einschlussdiagnose aus dem Bereich (ICD-10) I60.- bis I64.- zugeordnet werden konnte, wurden zu ihrer gesundheitsbezogenen Lebensqualität befragt. Um Veränderungen der Gesundheit im Laufe der Rehabilitationsbehandlung abzubilden, wurden die Patient:innen zu zwei Zeitpunkten befragt (indirekte Veränderungsmessung): zur Aufnahme in die Rehabilitationsbehandlung (**Aufnahmezeitpunkt**) und katamnestisch im häuslichen Umfeld 90 Tage nach Beginn der Rehabilitation (**Katamnesezeitpunkt**). Auf eine Befragung zum Zeitpunkt der Entlassung aus der Rehabilitation wurde nach ausführlicher Beratung mit Expert:innen teilnehmender Einrichtungen verzichtet um die Belastung der Patient:innen durch die Befragung zu minimieren und da die Annahme besteht, dass die Erfolge durch die Rehabilitation zum gewählten Katamnesezeitpunkt drei Monate nach Behandlungsbeginn aussagekräftiger als die Ergebnisse zum Entlassungszeitpunkt seien. Die für die Ergebnismessung benötigten Daten wurden aus Patient:innenangaben (Eingaben am Bildschirm oder Paper-Pencil-Fragebögen) und Behandler:innenangaben (Eingaben am Bildschirm, im KIS oder Paper-Pencil-Fragebögen) zusammengeführt. Für die Analysen konnten nur Datensätze berücksichtigt werden, die für die Berechnung mindestens (einer Dimension) eines Outcomes gültige Werte in den PROMs zum Beginn der Rehabilitationsbehandlung und zum Katamnesezeitpunkt beinhalten bzw. in denen für eine vorzunehmende Risikoadjustierung notwendige Patient:innenmerkmale vollständig erfasst sind. Als Zulassungskriterium für die Datenauswertungen wurde zunächst eine Mindestfallzahl von n = 125 vollständig erhobenen Fällen pro Einrichtung festgelegt um mit ausreichend statistischer Power möglichst aussagekräftige und trennscharfe Ergebnisse generieren zu können. Aufgrund der schwierigen Situation inmitten der SARS-Cov-2-Pandemie wurde das Kriterium der Mindestfallzahl auch im Hinblick auf den explorativen Charakter der Pilotstudie im Erhebungsverlauf aufgegeben.

Auswertung des erreichten beobachtbaren Outcomes

Als wichtige Zielgröße der Ergebnisqualität der neurologischen Rehabilitation des Schlaganfalls ist die Veränderung der gesundheitsbezogenen Lebensqualität, die von den Patient:innen in punktuellen Befragungen zum Aufnahmezeitpunkt und zum Katamnesezeitpunkt berichtet wird, bestimmt. Die Größe dieser Veränderungen werden für jedes Instrument bzw. jeweils für die körperliche und psychische Dimension der selbstberichteten Gesundheit als **Effektstärken**

^a Im späteren Erhebungsverlauf wurde eine Datenerhebung auch für Patient:innen in der Rehabilitationsphase C zugelassen, um Einrichtungen mit wenigen Phase-D-Patient:innen zu unterstützen, die anvisierten Fallzahlen zum Erhebungsende zu erreichen. Nur aus n = 3 Einrichtungen flossen insgesamt n = 7 Phase-C-Patient:innen in die Analysen ein. Der Anteil an Phase-C-Patient:innen an den einzelnen Analysestichproben der betreffenden Einrichtungen beträgt zwischen 2% und 4%.

Cohen's d_z für verbundene Stichproben für die Gesamtstichprobe aller teilnehmenden Einrichtungen und im Vergleich der einzelnen Einrichtungen mit den jeweiligen 95%-igen Konfidenzintervallen berechnet. Die Effektstärken der einzelnen Einrichtungen entsprechen dabei der mittleren Differenz der Outcomes zum Aufnahme- und Katamnesezeitpunkt innerhalb der einzelnen Einrichtung standardisiert an der Standardabweichung aller Differenzen der jeweiligen Einrichtung. Unterschiede in den Patient:innenmerkmalen zwischen den Einrichtungen sind bei diesen (unadjustierten) Effektstärken noch nicht berücksichtigt. Je größer der Unterschied zwischen der Aufnahme- und der Katamnesebefragung aller Patient:innen einer Einrichtung ist und je homogener die Veränderungen in der Stichprobe sind (wenn sich viele Patient:innen innerhalb einer Stichprobe/Einrichtung gleichermaßen stark verändern), desto größer ist der Betrag von Cohen's d_z .

Risikoadjustierung

Regressionsanalytisch kann ermittelt werden, welches Outcome unter Berücksichtigung einrichtungsspezifischer Patient:innenmerkmale (**Case-Mix**), wie der Eingangsbelastung, soziodemografischer und medizinischer Merkmale, am Ende der Rehabilitation im Durchschnitt aller Rehabilitationseinrichtungen erwartet werden kann (**Erwartungswert**). Die durchschnittliche Differenz der tatsächlich erreichten Werte der Patient:innen von den für sie berechneten Erwartungswerten innerhalb einer Klinik ist demnach ein Indikator, wie die jeweilige Klinik im Vergleich zu anderen Kliniken bezüglich der durchschnittlichen Ergebnisqualität für die Gesamtheit der befragten Patient:innen abschneidet. Die Erwartungswerte für die einzelnen Patient:innen werden durch eine lineare Gleichung berechnet, in die alle mit dem Outcome zusammenhängenden Patient:innenmerkmale entsprechend ihres in der linearen Regressionsanalyse ermittelten Gewichts eingehen. Werden die Regressionsgewichte zu β -Gewichten standardisiert, kann die jeweilige Stärke des Zusammenhangs eines Merkmals mit dem Outcome zwischen den Merkmalen unmittelbar verglichen werden. Für eine erwartungstreue Schätzung der Gewichtungen der Patient:innenmerkmale kommt es auf eine möglichst vollständige Erhebung der relevanten Einflussfaktoren an. Für die Risikoadjustierung konnten jene Fälle eingeschlossen werden, die hinsichtlich der Vollständigkeit und Qualität der von den Kliniken übermittelten Daten zu Patient:innenmerkmalen ausreichend waren. Folgende Variablen waren in der regressionsanalytischen Modellierung vorgesehen:

- Testergebnisse in der körperlichen und psychischen Dimension des PROMIS-10 bzw. des VR-12 zum Aufnahmezeitpunkt (T-Werte),
- Alter (Jahre),
- Geschlecht (weiblich, männlich, divers),
- Muttersprache (deutsch, andere),
- Wohnort (Bundesland Ost/West),
- Partnerschaft (feste:r Partner:in, ohne feste:n Partner:in),
- Familienstand (ledig, verheiratet/eingetragene Lebenspartnerschaft,

- getrennt/geschieden, verwitwet, wiederverheiratet),
- Schulabschluss (kein/niedrig/unbestimmt, mittel, hoch),
- Berufsausbildung (noch in Ausbildung/ohne Berufsabschluss/unbestimmt, Lehre/Fachschule, Meister:in, Fachhochschule/Universität/Berufsakademie),
- berufliche Situation (berufstätig, Bezug von Erwerbs-/Berufsunfähigkeitsrente, Bezug von Altersrente/Pension, nicht berufstätig, anderweitig beschäftigt),
- Krankschreibung vor Beginn der Reha (ja, nein)
- Dauer der Arbeitsunfähigkeit/krankheitsbedingter Einschränkungen im Alltag im gesamten letzten Jahr vor Beginn der Reha (weniger als 6 Monate, mehr als 6 Monate),
- Geplanter oder laufender Antrag auf Erwerbsminderungsrente (ja, nein),
- Aufforderung zur Rehabilitation durch Krankenkasse oder Arbeitsagentur (ja, nein),
- Anzahl dokumentierter psychischer Diagnosen,
- Anzahl dokumentierter somatischer Diagnosen,
- Weitere, für die Rehabilitation relevante und über die dokumentierten hinausgehende, Diagnosen (ja, nein),
- Art der Durchführung der Maßnahme (ganztägig ambulant, stationär),
- Reha-Phase zum Zeitpunkt der Aufnahmebefragung (C, D)
- Hilfe beim Ausfüllen der Fragebögen (ja, nein).

Unter Beteiligung von Repräsentant:innen der DRV Bund wurden im Jahr 2018 potenzielle Patient:innenmerkmale von der AG Ergebnismessung und dem UKE diskutiert und gemeinsam festgelegt. Die oben aufgelistete Auswahl orientiert sich dabei an den Variablen, die zu diesem Zeitpunkt im Reha-QS-Verfahren der DRV zur Risikoadjustierung genutzt wurden.

Alle Merkmalskategorien von Variablen auf einem Kategorial- oder Ordinalskalen-Niveau wurden zu sogenannten Dummy-Variablen binär rekodiert, d. h., jede Kategorie einer Variable wurde als einzelne Variable mit den Werten „Merkmal trifft zu“=1 und „Merkmal trifft nicht zu“=0 abgebildet. Ein Merkmal wurde als Referenzkategorie festgelegt, wobei in der Regel die besonders häufig besetzten Kategorien dazu ausgewählt wurden. Die Referenzkategorien der jeweiligen Variablen sind in der obigen Aufzählung durch Unterstreichung gekennzeichnet. Bis auf die Referenzkategorie wurden alle Dummy-Variablen einer kategorialen Variable in den Regressionsanalysen als Prädiktoren berücksichtigt. Die Referenzkategorie wird aus mathematischen Gründen in das Regressionsmodell nicht aufgenommen, da die Kombinationen aller anderen binär kodierten Kategorien eines Merkmals die Varianz des gesamten Merkmals einschließlich der Referenzkategorie vollständig erklären. Variablen auf mindestens Intervallskalenniveau (Outcomes im PROMIS-10 bzw. VR-12 zum Aufnahmezeitpunkt, Alter, Anzahl der Diagnosen) wurden als numerische Variablen im Regressionsmodell berücksichtigt.

Die Durchführung der Risikoadjustierung wurde zunächst getrennt nach Instrumenten bzw. ihren zugrundeliegenden Dimensionen (körperliche bzw. psychische gesundheitsbezogene

Lebensqualität) geplant. Für jedes Instrument bzw. jede Dimension sollte im ersten Schritt ein Regressionsmodell erstellt werden, in dem das Outcome des Instruments (das Testergebnis der jeweiligen Skala zum Katamnesezeitpunkt) die abhängige Variable darstellt. Als feste Prädiktoren sollten für jede Dimension Alter und Geschlecht, sowie das Ergebnis der Befragung zum Aufnahmezeitpunkt für das jeweilige Instrument in die Regressionsmodelle aufgenommen werden. In einem zweiten Schritt sollten zunächst alle weiteren (oben genannten) Variablen zur Risikoadjustierung aufgenommen und im Rahmen der Modellierung schrittweise wieder entfernt werden, wenn sie nicht zumindest einen tendenziell signifikanten Beitrag ($p < .10$) zur Aufklärung der Varianz des Outcomes beitragen. Im letzten Schritt sollte das resultierende Modell aus Schritt 2, nun mit den Variablen mit einem statistisch signifikanten Zusammenhang mit dem Outcome als festen Prädiktoren, erneut auf die Daten angewendet werden. Hintergrund des abschließenden Schritts ist es, dass durch die verringerte Anzahl an Prädiktoren im Vergleich zum Ursprungsmodell auch Fälle eingeschlossen werden können, die zwar fehlende Werte in den ausgeschlossenen Variablen aufweisen und daher in der Modellierung im Schritt 2 nicht berücksichtigt werden konnten, aber vollständige Daten in den finalen Prädiktoren aufweisen. Durch die hierdurch erreichte höhere Fallzahl steigen sowohl die Power nachfolgender statistischer Tests mit den risikoadjustierten Daten als auch die Repräsentativität der Daten für die einzelne Einrichtung. Allerdings können sich durch die veränderte Datenbasis die statistischen Parameter der Regressionsmodelle (insbesondere β -Gewichte der Prädiktoren und Irrtumswahrscheinlichkeiten für das Verwerfen der Annahme eines fehlenden Zusammenhangs mit dem Outcome) ändern. In der Praxis bedeutet dies, dass teilweise Prädiktoren, die nach der Regressionsanalyse im zweiten Schritt zumindest tendenziell einen statistisch bedeutsamen Zusammenhang mit dem Outcome zeigen, nun ein geringeres β -Gewicht aufweisen und der Zusammenhang nicht mehr statistisch abgesichert ist. Da die benannten Vorteile einer höheren erreichbaren Fallzahl den Nachteil der potentiellen Aufnahme statistisch nicht signifikanter Prädiktoren in das finale Regressionsmodell überwiegen, und davon ausgegangen werden kann, dass aufgrund der niedrigen β -Gewichte der betroffenen Prädiktoren den einzelnen Einrichtungen insgesamt keine relevanten Nachteile hinsichtlich der berechneten Erwartungswerte entstehen, wurde darauf verzichtet, im Sinne eines iterativen Verfahrens so lange die Modelle durch Wiederholungen der Analyseschritte 2 und 3 anzupassen, bis eine volle Übereinstimmung der Modelle aus den Schritten 2 und 3 resultiert. Demzufolge werden die Ergebnisse der Regressionsanalysen mit allen aufgenommenen Prädiktoren aus Analyseschritt 2 mit ihren β -Gewichten und Irrtumswahrscheinlichkeiten aus Schritt 3 berichtet, und die anhand der finalen Regressionsmodelle berechneten Erwartungswerte zur risikoadjustierten einrichtungsvergleichenden Analyse der Outcomes und der Berechnung des ProQIs genutzt.

Analyse von Deckeneffekten (a posteriori)

Bei der Durchführung der Analysen des durchschnittlich erreichten Outcomes im PROMIS-10 waren in der Gesamtstichprobe über alle Einrichtungen einerseits sehr niedrige Effektstärken der Veränderung des Gesundheitszustands zwischen Aufnahme- und Katamnesezeitpunkt und andererseits relativ hohe Testergebnisse zum Aufnahmezeitpunkt zu beobachten. Es sollte

daher analysiert werden, ob Deckeneffekte für die niedrigen Effektstärken ursächlich sind, also ob bei einem großen Teil der Patient:innenstichprobe die gesundheitsbezogene Lebensqualität zum Aufnahmezeitpunkt bereits so hoch war, das kaum weitere Verbesserungen erwartet werden konnten. Zunächst wurden dazu die Differenzwerte zwischen Aufnahme- und Katamnesezeitpunkt aller Patient:innen in Abhängigkeit der Testergebnisse zum Aufnahmezeitpunkt analysiert. Danach wurde für weitere Analysen zu Deckeneffekten die Gesamtstichprobe nach drei Gruppen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt stratifiziert. Ein Ziel der Stratifizierung war, die Vergleichsgruppen in etwa gleich groß zu bilden, um eine möglichst hohe statistische Power für die Vergleiche dieser Gruppen zu generieren. Aus Häufigkeitsanalysen der Gesamtstichprobe ergab sich, dass geeignete Trennwerte von T-Werten in der körperlichen und der psychischen Dimension des PROMIS-10 zur Gruppeneinteilung für die Gruppe mit niedriger gesundheitsbezogener Lebensqualität bei $T \leq 40$, für die Gruppe mit mittlerer Lebensqualität bei $T > 40$ und $T \leq 45$, sowie für die Gruppe mit hoher Lebensqualität zum Aufnahmezeitpunkt bei $T > 45$ lagen. Die Abstände zwischen den Trennwerten entsprechen damit auch jeweils einer halben Standardabweichung auf der T-Werte-Skala. Exemplarisch für den PROMIS-10 wurden für diese drei Gruppen die jeweiligen Anteile der Gruppengrößen an der Fallzahl pro Einrichtung ermittelt, um eine mögliche Variabilität von Deckeneffekten zwischen den Einrichtungen zu aufdecken zu können. Sowohl für den PROMIS-10 als auch für den VR-12 (unter Verwendung der selben Trennwerte auf den Skalen des VR-12 zur Gruppenbildung) wurden zur Beurteilung von Deckeneffekten jeweils die Effektstärken der Veränderungen für die drei stratifizierten Gruppen berechnet.

Erweiterung der Risikoadjustierung durch stratifizierte multiple Regressionsmodelle (a posteriori)

Nachdem die vorherigen Analysen Deckeneffekte im PROMIS-10 offenlegten, die in den Einrichtungen unterschiedlich stark ausgeprägt waren, wurde die oben dargestellte Risikoadjustierung erweitert und stratifiziert mit jeweils drei Regressionsmodellen pro Dimension (jeweils ein Modell für Patient:innen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt) sowohl für den PROMIS-10 als auch für den VR-12 durchgeführt. Wichtige Kennwerte der Regressionsmodelle und Stichprobenparameter der in die Regression einbezogenen Patient:innenmerkmale der Gesamtstichprobe sind im Anhang dieses Berichtes ab Seite 55 tabelliert.

Auswertung der Outcomes nach der Risikoadjustierung

Für einen fairen Vergleich der Einrichtungen hinsichtlich der erreichten Veränderungen der gesundheitsbezogenen Lebensqualität zum Katamnesezeitpunkt in Abhängigkeit des einrichtungsspezifischen Case-Mixes werden pro Einrichtung die an ihrer einrichtungsspezifischen Varianz standardisierten Differenzen von tatsächlich erreichten Outcome-Werten und Erwartungswerten analysiert. Diese standardisierten Differenzen entsprechen **ebenfalls Effektstärken Cohen's d_z** für verbundene Stichproben. Statistisch signifikant höhere Effektstärken als 0 zeigen hier an, dass ein besseres Outcome erreicht wurde,

als es aufgrund der Patient:innenmerkmale in einer Klinik zu erwarten gewesen wäre. Effektstärken statistisch signifikant kleiner 0 deuten darauf hin, dass die von den Patientinnen und Patienten berichtete gesundheitsbezogene Lebensqualität zum Katamnesezeitpunkt im Vergleich zu den anderen am Pilotprojekt teilnehmenden Kliniken unterdurchschnittlich ist. Schließlich zeigen statistisch nicht signifikante Effektstärken um 0 an, dass die beobachteten Outcomes etwa den aufgrund des Case-Mix erwarteten Behandlungsergebnissen entsprechen. In den im Ergebnisteil dargestellten Diagrammen zum Vergleich der Einrichtungen repräsentiert die Nulllinie eine absolute Übereinstimmung der erreichten mit den erwarteten Outcomes, und Abweichungen von der Nulllinie stellen die Größe der Abweichungen der Effektstärken von den erwarteten Outcomes dar.

Vergleich von Rehabilitationseinrichtungen mit dem Patient-Reported Outcome Quality Index (ProQI)

Abschließend wird ein zusammenfassender Ergebnisqualitätsindex ProQI berechnet, der die über die körperliche und psychische Dimension integrierten Ergebnisse der Instrumente darstellt und so die erreichten Verbesserungen aus Patient:innensicht als aggregiertes Maß abbildet.

Der Patient-Reported Outcome Quality Index (ProQI) wird entwickelt und vorgeschlagen, um Unterschiede in der Ergebnisqualität der patient:innenberichteten Gesundheit zwischen den Mitgliedskliniken in einem einfachen Indexwert zusammenzuführen und darzustellen. Grundlage des ProQI sind die standardisierten Residuen als Ergebnis der Risikoadjustierung der Outcome-Werte: Die Residuen sind die Differenzen zwischen den beobachteten und den aufgrund der Patient:innenmerkmale erwarteten Outcomes zum Entlassungs- oder Katamnesezeitpunkt. Je höher die durchschnittlichen Differenzen in einem Patient:innenkollektiv im Vergleich zu anderen Patient:innenkollektiven in Richtung Verbesserung abweichen, desto besser stellt sich die entsprechende Klinik in einem Benchmark dar. An dem Gesamtkollektiv z-standardisierte Residuen haben einen Mittelwert von $M = 0$ und eine Standardabweichung von $SD = 1$. Vorteil der z-standardisierten Differenzen ist, dass sich diese zwischen verschiedenen PROMs/Skalen direkt miteinander vergleichen lassen, da sie alle die gleichen Mittelwerte und Streuungen aufweisen. Zur Konstruktion des ProQI werden die z-standardisierten Residuen aller Patient:innen auf eine Skala mit einem Mittelwert von $M = 90$ und einer Standardabweichung von $SD = 30$ transformiert. Zugrundeliegende Annahme ist dabei, dass sich die Ergebnisqualität in der Rehabilitation insgesamt bereits auf einem hohen Niveau befindet und sich das Gros der Einrichtungen nicht allzu stark voneinander unterscheidet. Vermutlich nur wenige Rehabilitationseinrichtungen liegen in einem Benchmark deutlich über oder unter dem Durchschnitt aller betrachteten Kliniken. Für die Berechnung der ProQIs für dieses Pilotprojekt wurde folgendermaßen vorgegangen:

1. Transformation der standardisierten Residuen aller PROMs auf Skalen mit $M = 90$ und $SD = 30$.

2. Zur Berechnung eines skalenspezifischen ProQIs (physische oder psychische Dimension) eines Instruments wird der Mittelwert der standardisierten und transformierten Residuen in dieser Skala über alle Patient:innen einer Klinik gebildet.
3. Zur Berechnung eines instrumentenspezifischen ProQIs werden die durchschnittlich gewichteten Summen der skalenspezifischen ProQIs für jedes Instrument gebildet:
$$\text{ProQI}_{\text{Instrument}} = \frac{1}{2} * \text{ProQI}_{\text{physischeDim.}} + \frac{1}{2} * \text{ProQI}_{\text{psychischeDim.}}$$
4. Zur Berechnung eines instrumentenübergreifenden ProQI-Gesamt wird der Mittelwert über die instrumentenspezifischen ProQIs gebildet: $\text{ProQI}_{\text{Gesamt}} = (\text{ProQI}_{\text{PROMIS-10}} + \text{ProQI}_{\text{VR-12}}) / 2$.

Abschließend werden ProQIs mit Werten > 100 gekappt, so dass sich ProQI-Skalen mit theoretischen Werten von 0 bis 100 ergeben. Dadurch können einerseits unterdurchschnittliche Werte in einem Bereich durch überdurchschnittliche Ergebnisse in einem anderen Bereich kompensiert werden, andererseits weisen die resultierenden Skalen eine eingängige und intuitive Skalierung von 0 bis 100 auf. Der $\text{ProQI}_{\text{PROMIS-10}}$ kann für 17 Einrichtungen berechnet werden und der $\text{ProQI}_{\text{VR-12}}$ für die fünf Einrichtungen, die für eine Befragung mit dem VR-12 optiert hatten. Demzufolge kann auch nur für fünf Einrichtungen ein $\text{ProQI}_{\text{Gesamt}}$ berechnet werden.

Vergleich der Änderungssensitivität von PROMIS-10 und VR-12

Der PROMIS-10 und der VR-12 messen konzeptuell vergleichbare Dimensionen der physischen und psychischen gesundheitsbezogenen Lebensqualität. Die Indikatoren der einzelnen Facetten des Gesundheitszustands sind für beide Instrumente sehr ähnlich operationalisiert bzw. formuliert und überlappen sich inhaltlich. Auch wenn für beide Instrumente als Test-Ergebnis der jeweiligen Dimension ein Wert auf einer bevölkerungsnormierten T-Wert-Skala resultiert, unterscheidet sich das Scoring der Instrumente grundlegend: Während das Scoring des PROMIS-10 Global Health für beide Dimensionen körperlicher und psychischer gesundheitsbezogener Lebensqualität mit jeweils vier Items pro Skala getrennt durchgeführt wird und durch Summation der jeweils vier Item-Scores den zwei resultierenden Skalen-Rohwerten ein auf früheren Validierungsstudien beruhender T-Wert direkt zugeordnet wird, gehen beim Scoring des VR-12 (angelehnt an den SF-12) alle zwölf Items mit jeweils unterschiedlicher Gewichtung für jede Dimension in eine lineare Gleichung ein, aus der der T-Wert als Test-Ergebnis direkt berechnet wird. Trotz der unterschiedlichen Berechnungsweisen der Testergebnisse wird aber aufgrund der inhaltlichen Überschneidungen angenommen, dass beide Instrumente in gleicher Weise änderungssensitiv sind, um Veränderungen der gesundheitsbezogenen Lebensqualität im Behandlungsverlauf abbilden zu können. Um dies zu überprüfen, werden für die Unterstichprobe, in der Patient:innen sowohl den PROMIS-10 als auch den VR-12 ausgefüllt haben, Effektstärken der Veränderungen (Cohen's d_z) und Re-Test-Reliabilitätskoeffizienten (Korrelationen der Testergebnisse zwischen Aufnahme- und Katamnesezeitpunkt; r_{tt}) sowohl für die Gesamtheit dieser Auswahl von Patient:innen als auch stratifiziert nach Gruppen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt berechnet.

Ergebnisse

In den folgenden Abschnitten werden die Ergebnisse der Analysen berichtet. Die Darstellung der Einzelergebnisse für die teilnehmenden Einrichtungen erfolgt in den Diagrammen anonymisiert, d. h. den einzelnen Einrichtungen wurden vor den Berechnungen randomisiert Buchstaben von A bis Q zugeordnet.

Rücklauf

Insgesamt wurden in allen teilnehmenden Einrichtungen Daten von N = 2081 Patientinnen und Patienten erhoben (Mittelwert über alle Einrichtungen (N_M) = 122,4; Minimum (N_{\min}) = 43; Maximum (N_{\max}) = 297). Für die Analysen der erreichten beobachteten Outcomes (ohne Berücksichtigung des einrichtungsspezifischen Case-Mixes) zur Ergebnisqualität konnten insgesamt Daten mit ausreichender Vollständigkeit von insgesamt N = 1166 Patient:innen (N_M = 68,6; N_{\min} = 24; N_{\max} = 155) einbezogen werden. Durch fehlende Angaben in den für die auf der Risikoadjustierung basierenden Analysen notwendigen Patient:innenmerkmalen weichen die Fallzahlen für den Einrichtungsvergleich ab: hier konnten insgesamt N = 1047 Fälle (N_M = 61,6; N_{\min} = 18; N_{\max} = 135) ausgewertet werden (Abbildung 1).

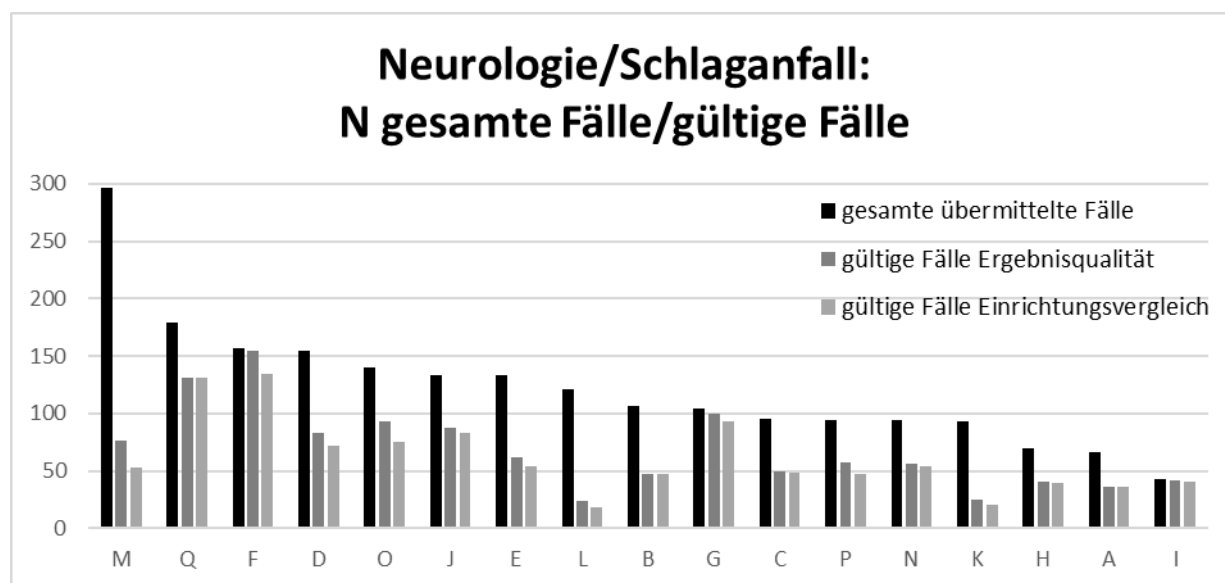


Abb.1: Anzahl der übermittelten Fälle und der für die Analysen zur Ergebnisqualität und zum Einrichtungsvergleich gültige Anzahl von Fällen aller teilnehmenden Einrichtungen in der Indikation Schlaganfall.

Die Rücklaufquoten analysierbarer Fälle betragen für die Analysen der Ergebnisqualität somit insgesamt 56% der zum Aufnahmezeitpunkt erfassten Patient:innen (M = 59%; \min = 19,8%; \max = 98,7%) und für den Einrichtungsvergleich insgesamt 50,3% (M = 54%; \min = 14,9%; \max = 95,3%; Abbildung 2).

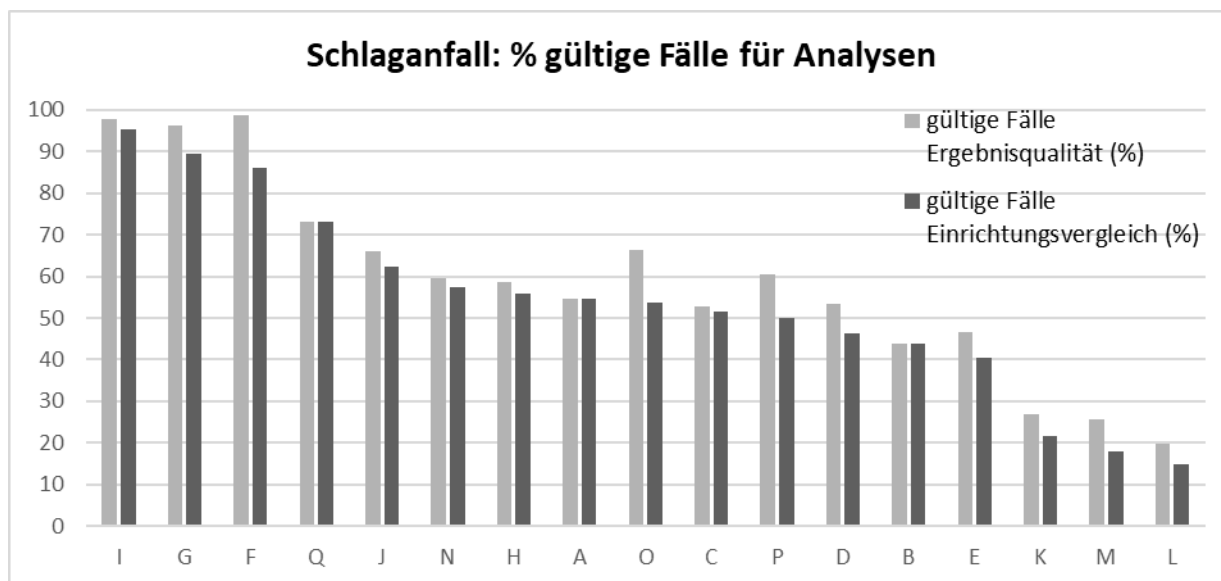


Abb.2: Prozentwerte gültiger Datensätze für Analysen zur Ergebnisqualität und zum Einrichtungsvergleich bezogen auf die Gesamtzahl der übermittelten Fälle aus den teilnehmenden Einrichtungen in der Indikation Schlaganfall.

Patient:innenmerkmale

In Tabelle 1 sind ausgewählte Stichprobencharakteristika von Alter, Geschlecht und weiteren Merkmalen der eingeschlossenen Patient:innen in der Gesamtstichprobe für die Analysen der Ergebnisqualität dargestellt.

Patient:innenmerkmale	Alle Kliniken (N = 1166)
Alter (M/SD)	65,0 (11,9)
weiblich (%)	41,8%
Muttersprache nicht Deutsch (%)	4,3%
Wohnort neue Bundesländer (%)	29,1%
Ohne feste Partnerschaft (%)	25,6%
Schulabschluss (%)	
Kein/niedrig	43,5%
Mittel	34,6%
Hoch	21,9%
Reguläre Erwerbstätigkeit (Voll- oder Teilzeit) (%)	41,7%
Erwerbsunfähigkeitsrente (auf Zeit oder auf Dauer) (%)	3,4%
Altersrente/Früherberentung/Pension (%)	39,6%
Antrag auf Erwerbsminderungsrente gestellt/geplant (%)	18,6%
Unmittelbar vor Reha krankgeschrieben (%)	36,7%

Patient:innenmerkmale	Alle Kliniken (N = 1166)
Im letzten Jahr weniger als 6 Monate arbeitsunfähig (%)	91,7%
Aufforderung zur Reha durch Krankenkasse/Agentur für Arbeit (%)	17,8%
Anteil Patient:innen in Phase C in Stichprobe (%)	0,6%
Durchführung Reha als ganztägig ambulante Maßnahme (%)	3,2%
Anzahl dokumentierter somatischer Diagnosen (M/SD) (%)	2,6 (1,6)
Anzahl dokumentierter psychischer Diagnosen (M/SD) (%)	0,1 (0,3)
Weitere relevante Diagnosen vorhanden (%)	21,0%
Benötigte Hilfe beim Ausfüllen eines Fragebogens (%)	23,3%

Tab.1: Alter, Geschlecht und weitere Patient:innenmerkmale für die Indikation Schlaganfall.

Ergebnisqualität im PROMIS-10 (beobachtete Outcomes)

Im Folgenden sind Stichprobenkennwerte und Effektstärken der Veränderung des Gesundheitszustands im PROMIS-10 zwischen Aufnahme- und Katamnesezeitpunkt für die Gesamtstichprobe über alle Einrichtungen und im Vergleich zwischen den Einrichtungen dargestellt.

In Tabelle 2 sind zunächst die durchschnittlichen T-Werte für die Dimension **Physical Health (PH) des PROMIS-10** zum Aufnahme- und zum Katamnesezeitpunkt, die Korrelation dieser Messwerte und die resultierende Effektstärke der Veränderung (Cohen's d_z) für die Gesamtstichprobe über alle Einrichtungen aufgeführt.

PROMIS-10 Physical Health (PH)	Alle Kliniken
Gesamtstichprobe (N)	1085
PROMIS-10 PH Aufnahme (M/SD)	42,3 (7,7)
PROMIS-10 PH Katamnese (M/SD)	43,1 (7,8)
Korrelation r_{tt} Promis-10 PH Aufnahme/Katamnese	0,56
Cohen's d_z	0,11

Tab.2: Deskriptive Statistiken zu Scores im PROMIS-10 Physical Health (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt. Der t-Test für verbundene Stichproben in der Gesamtstichprobe ergibt statistisch signifikante Unterschiede zwischen Aufnahme- und Katamnesezeitpunkt ($t = 3,543$; $df = 1084$; $p < 0,001$).

In Abbildung 3 wird ein Vergleich der Effektstärken der unadjustierten Outcomes (Cohen's d_z mit 95%-Konfidenzintervall) zu den Veränderungen zwischen Aufnahme- und Katamnesezeitpunkt für die Dimension Physical Health (PH) des PROMIS-10 zwischen den Einrichtungen dargestellt. Die nicht risikoadjustierten Unterschiede zwischen den Einrichtungen in diesen Effektstärken sind nur bedingt interpretierbar, da sich spezifische

Patient:innenmerkmale, die einen Einfluss auf die Höhe des Outcomes haben können, ebenfalls zwischen den Kliniken unterscheiden können. Dies trifft insbesondere dann zu, wenn potentielle Deckeneffekte (Patient:innen beginnen die Rehabilitation bereits mit einer hoch ausgeprägten gesundheitsbezogenen Lebensqualität) in verschiedenen Einrichtungen unterschiedlich stark ausgeprägt sind. Die durchschnittliche Effektstärke aller Einrichtungen im PROMIS-10 PH beträgt $d_z = 0,11$ und wird durch die hervorgehobene Linie im Diagramm dargestellt. Die erreichten Effektstärken der einzelnen Einrichtungen werden mit einer roten Raute abgebildet. Die Linien ober- und unterhalb der einzelnen Effektstärken (auch: „Whiskers“) markieren die jeweiligen Konfidenzintervalle der Effektstärken: die wahre, stichprobenunabhängige Effektstärke für die einzelne Einrichtung liegt mit 95%-iger Wahrscheinlichkeit in diesem Bereich. Der Umfang eines Konfidenzintervalls ist hier hauptsächlich abhängig von der Fallzahl, die für die jeweilige Einrichtung analysiert werden konnte. Der Range der in den Einrichtungen erreichten Effektstärken liegt zwischen $d_z = -0,28$ und $d_z = 0,43$. Während die Einrichtungen O, M und Q Effektstärken erreichen, die statistisch signifikant ($p < 0,05$) über dem Durchschnitt der Gesamtstichprobe liegen, liegt die Effektstärke der Einrichtung A statistisch signifikant unter der durchschnittlichen Effektstärke. Unter Außerachtlassung der inferenzstatistischen Absicherung bzw. Betrachtung des Konfidenzintervalls um die Effektstärken weisen in 11 Einrichtungen die Patient:innen nach ihren Angaben durchschnittlich zum Katamnesezeitpunkt eine höhere körperliche gesundheitsbezogene Lebensqualität auf als zum Rehabilitationsbeginn. Warum in den anderen sechs Einrichtungen Patient:innen verglichen mit dem Aufnahmezeitpunkt zum Katamnesezeitpunkt durchschnittlich eine geringere körperliche gesundheitsbezogene Lebensqualität beobachtbar ist, wird im Abschnitt „Deckeneffekte im PROMIS-10 und VR-12“ (S. 26) analysiert und im Diskussionsteil (S. 43 ff.) ausführlich begründet.

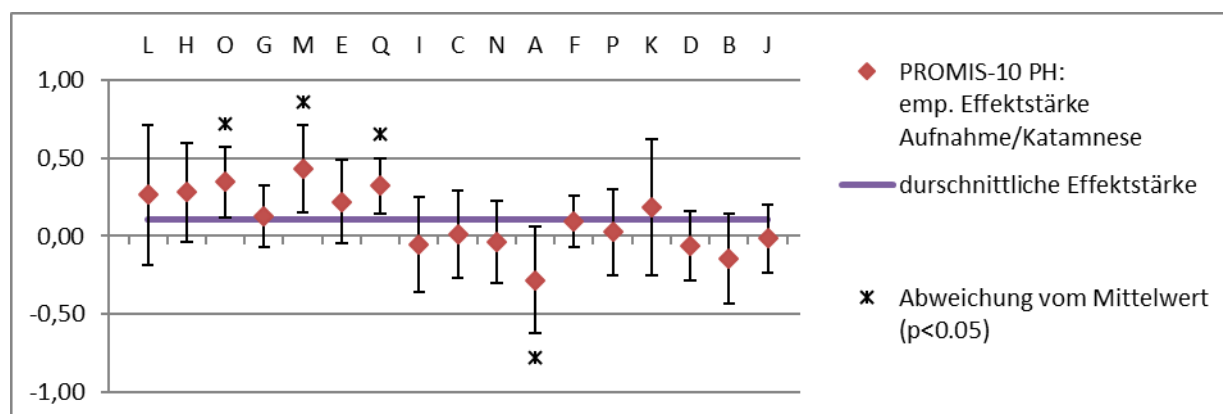


Abb.3: Effektstärken der unadjustierten Outcomes aller Einrichtungen im Vergleich Aufnahme/Katamnese für den PROMIS-10 PH (höhere Werte entsprechen einer größeren Verbesserung).

Für die Skala **Mental Health (MH) des PROMIS-10** sind in Tabelle 3 die durchschnittlichen T-Werte zum Aufnahme- und zum Katamnesezeitpunkt, die Korrelation dieser Messwerte und die resultierende Effektstärke der Veränderung (Cohen's d_z) für die Gesamtstichprobe über alle Einrichtungen aufgeführt.

PROMIS-10 Mental Health (MH)	Alle Kliniken
Gesamtstichprobe (N)	1119
PROMIS-10 MH Aufnahme (M/SD)	43,7 (7,8)
PROMIS-10 MH Katamnese (M/SD)	44,1 (7,6)
Korrelation r_{tt} Promis-10 MH Aufnahme/Katamnese	0,56
Cohen's d_z	0,05

Tab.3: Deskriptive Statistiken zu Scores im PROMIS-10 Mental Health (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt. Die Annahme von vorhandenen Unterschieden zwischen Aufnahme- und Katamnesezeitpunkt in der Gesamtstichprobe kann durch den t-Test für verbundene Stichproben **nicht** statistisch auf einem α -Niveau von 0,05 abgesichert werden ($t = 1,661$; $df = 1118$; $p = 0,097$).

In Abbildung 4 wird ein Vergleich der Effektstärken der unadjustierten Outcomes (Cohen's d_z mit 95%-Konfidenzintervall) zu den Veränderungen zwischen Aufnahme- und Katamnesezeitpunkt für die Dimension Mental Health (MH) des PROMIS-10 zwischen den Einrichtungen dargestellt. Die durchschnittliche Effektstärke aller Einrichtungen im PROMIS-10 MH beträgt $d_z = 0,05$. Der Range der in den Einrichtungen erreichten Effektstärken liegt zwischen $d_z = -0,25$ und $d_z = 0,38$. Während die Einrichtungen Q und E Effektstärken erreichen, die statistisch signifikant ($p < 0,05$) über dem Durchschnitt der Gesamtstichprobe liegen, liegt die Effektstärke der Einrichtung C statistisch signifikant unter der durchschnittlichen Effektstärke. Unter Auslassung der inferenzstatistischen Absicherung bzw. Betrachtung des Konfidenzintervalls um die Effektstärken weisen in 10 Einrichtungen die Patient:innen nach ihren Angaben durchschnittlich zum Katamnesezeitpunkt eine höhere psychische gesundheitsbezogene Lebensqualität auf als zum Rehabilitationsbeginn. Warum in den anderen sieben Einrichtungen Patient:innen verglichen mit dem Aufnahmezeitpunkt zum Katamnesezeitpunkt durchschnittlich eine geringere psychische gesundheitsbezogene Lebensqualität beobachtbar ist, wird im Abschnitt „Deckeneffekte im PROMIS-10 und VR-12“ (S. 26) analysiert und im Diskussionsteil (S. 43 ff.) ausführlich begründet. Die nicht risikoadjustierten Unterschiede zwischen den Einrichtungen in diesen Effektstärken sind aus den weiter oben genannten Gründen nur bedingt interpretierbar.

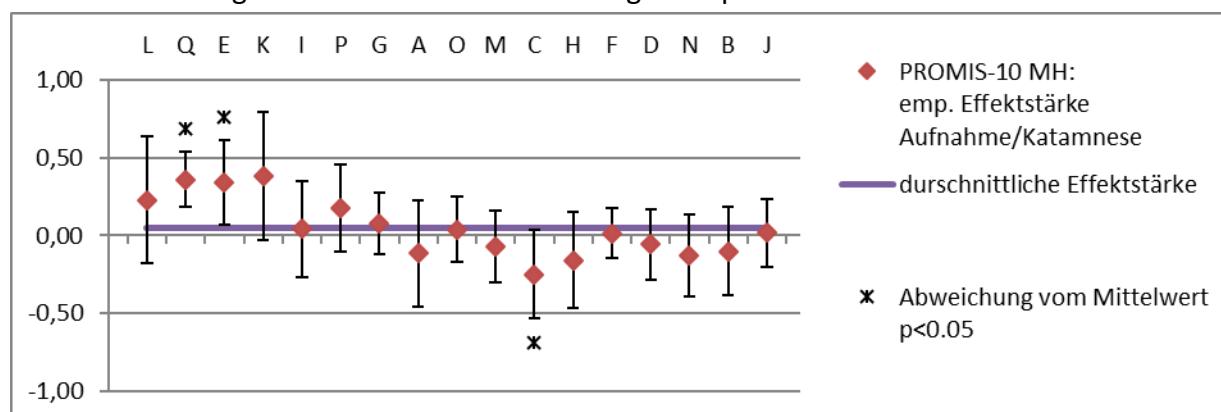


Abb.4: Effektstärken der unadjustierten Outcomes aller Einrichtungen im Vergleich Aufnahme/Katamnese für den PROMIS-10 MH (höhere Werte entsprechen einer größeren Verbesserung).

Ergebnisqualität im VR-12 (beobachtete Outcomes)

Fünf der 17 teilnehmenden Einrichtungen haben die Option genutzt, ihre Patient:innen zusätzlich zum PROMIS-10 mit dem VR-12 zu befragen. In Tabelle 4 sind die durchschnittlichen T-Werte, die Korrelation zwischen den T-Werten und die Effektstärken der Veränderung des Gesundheitszustands für den **Physical Component Score (PCS) des VR-12** für die Gesamtstichprobe dieser fünf Einrichtungen zwischen Aufnahme- und Katamnesezeitpunkt dargestellt.

VR-12 Physical Component Score (PCS)	5 Kliniken mit VR-12
Gesamtstichprobe (N)	347
VR-12 PCS Aufnahme (M/SD)	38,4 (9,8)
VR-12 PCS Katamnese (M/SD)	40,7 (9,7)
Korrelation r_{tt} VR-12 PCS Aufnahme/Katamnese	0,62
Cohen's d_z	0,26

Tab.4: Deskriptive Statistiken zum VR-12 Physical Component Score (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt. Der t-Test für verbundene Stichproben in der Gesamtstichprobe ergibt statistisch signifikante Unterschiede zwischen Aufnahme- und Katamnesezeitpunkt ($t = 4,933$; $df = 346$; $p < 0,001$).

In Abbildung 5 wird ein Vergleich der Effektstärken der unadjustierten Outcomes (Cohen's d_z mit 95%-Konfidenzintervall) zu den Veränderungen zwischen Aufnahme- und Katamnesezeitpunkt für die körperliche Dimension des VR-12 zwischen den Einrichtungen dargestellt. Auch hier sind die nicht risikoadjustierten Unterschiede zwischen den Einrichtungen in diesen Effektstärken nur bedingt interpretierbar, da sich spezifische Patient:innenmerkmale, die einen Einfluss auf die Höhe des Outcomes haben können (insbesondere potentielle Deckeneffekte), ebenfalls zwischen den Kliniken unterscheiden können. Die durchschnittliche Effektstärke der teilnehmenden Einrichtungen im VR12-PCS beträgt $d_z = 0,26$. Der Range der in den Einrichtungen erreichten Effektstärken liegt zwischen $d_z = 0,18$ und $d_z = 0,33$. Keine der Einrichtungen liegt statistisch signifikant ($p < 0,05$) über oder unter dem Durchschnitt der Gesamtstichprobe aus den fünf Einrichtungen. Nur für die Einrichtung Q lassen sich (positive) Veränderungen ($d_z > 0$; $p < 0,05$) zum Katamnesezeitpunkt inferenzstatistisch absichern. Unter Außerachtlassung der inferenzstatistischen Absicherung bzw. Betrachtung des Konfidenzintervalls um die Effektstärken weisen die Patient:innen aller fünf Einrichtungen durchschnittlich zum Katamnesezeitpunkt eine höhere körperliche gesundheitsbezogene Lebensqualität auf als zum Rehabilitationsbeginn.

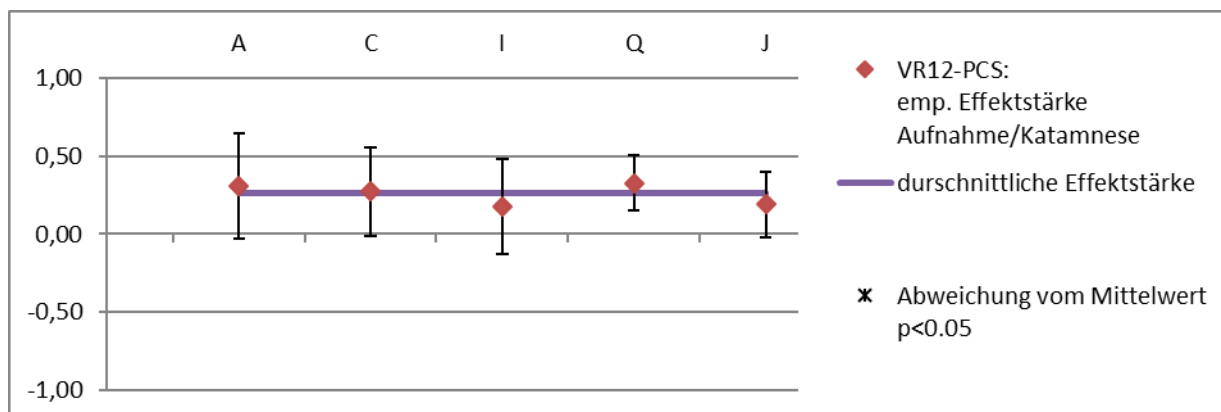


Abb.5: Effektstärken der unadjustierten Outcomes aller Einrichtungen im Vergleich Aufnahme/Katamnese für den VR-12 PCS (höhere Werte entsprechen einer größeren Verbesserung).

In Tabelle 5 sind die durchschnittlichen T-Werte zum Aufnahme- und zum Katamnesezeitpunkt, die Korrelation dieser Messwerte und die resultierende Effektstärke der Veränderung (Cohen's d_z) für die Gesamtstichprobe über alle Einrichtungen für den **Mental Component Score (MCS) des VR-12** aufgeführt.

VR-12 Mental Component Score (MCS)	5 Kliniken mit VR-12
Gesamtstichprobe (N)	346
VR-12 MCS Aufnahme (M/SD)	43,2 (12,2)
VR-12 MCS Katamnese (M/SD)	44,1 (12,0)
Korrelation r_{tt} VR-12 MCS Aufnahme/Katamnese	0,54
Cohen's d_z	0,08

Tab.5: Deskriptive Statistiken zum VR-12 Mental Component Score (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt. Die Annahme von vorhandenen Unterschieden zwischen Aufnahme- und Katamnesezeitpunkt in der Gesamtstichprobe kann durch den t-Test für verbundene Stichproben **nicht** statistisch auf einem α -Niveau von 0,05 abgesichert werden ($t = 1,449$; $df = 345$; $p = 0,148$).

In Abbildung 6 wird ein Vergleich der Effektstärken der unadjustierten Outcomes (Cohen's d_z mit 95%-Konfidenzintervall) zu den Veränderungen zwischen Aufnahme- und Katamnesezeitpunkt für die psychische Dimension des VR-12 zwischen den Einrichtungen dargestellt. Die durchschnittliche Effektstärke der fünf teilnehmenden Einrichtungen im VR12-MCS beträgt $d_z = 0,08$. Der Range der in den Einrichtungen erreichten Effektstärken liegt zwischen $d_z = -0,27$ und $d_z = 0,20$. Die Einrichtung C liegt statistisch signifikant ($p < 0.05$) unterhalb des Durchschnitts der Gesamtstichprobe. Patient:innen der Einrichtung Q berichten zum Katamnesezeitpunkt eine statistisch signifikante Verbesserung der psychischen gesundheitsbezogenen Lebensqualität, und unter Außerachtlassung der inferenzstatistischen Absicherung weisen auch Patient:innen der Einrichtungen I und J positive Veränderungen auf. Die nicht risikoadjustierten Unterschiede zwischen den Einrichtungen in diesen Effektstärken sind aus den bereits beschriebenen Gründen auch hier nur bedingt interpretierbar.

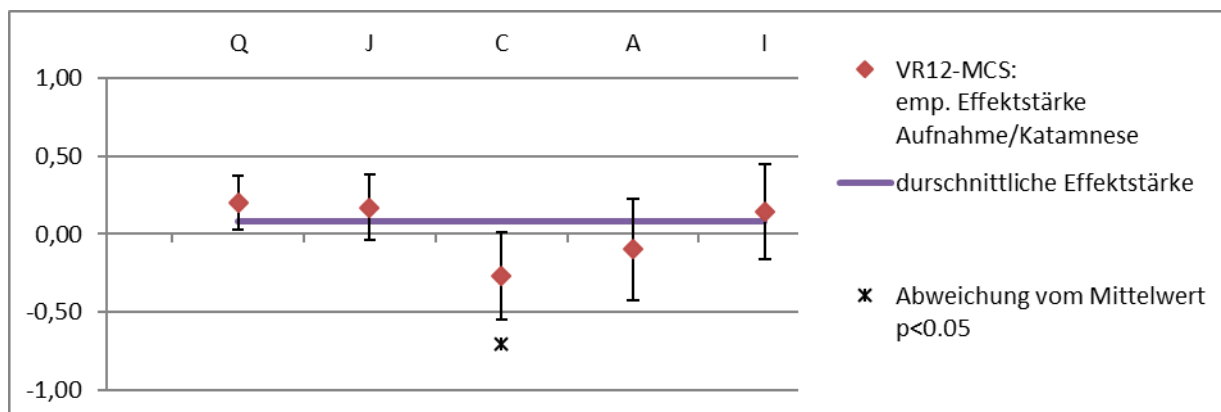


Abb.6: Effektstärken der unadjustierten Outcomes aller Einrichtungen im Vergleich Aufnahme/Katamnese für den VR-12 MCS (höhere Werte entsprechen einer größeren Verbesserung).

Deckeneffekte im PROMIS-10 und VR-12

Aufgrund der durchschnittlich niedrigen Effektstärken der Veränderungen der gesundheitsbezogenen Lebensqualität zwischen Aufnahme- und Katamnesezeitpunkt und den vor allem im PROMIS-10 relativ hohen T-Werten zum Aufnahmezeitpunkt wurde nach den ersten Auswertungen gezielt untersucht, ob Deckeneffekte den Ergebnissen zugrunde liegen können, ob also durch eine relativ hohe gesundheitsbezogene Lebensqualität zum Beginn der Rehabilitation kaum höhere Veränderungen zum Katamnesezeitpunkt erwartet werden können. Dazu wurden exemplarisch für den PROMIS-10 über alle Einrichtungen zunächst die Differenzen zwischen Aufnahme- und Katamnesezeitpunkt in Abhängigkeit des T-Werts zum Aufnahmezeitpunkt grafisch für die körperliche (Abbildung 7) und die psychische Dimension (Abbildung 8) analysiert. Jeder Punkt in den Grafiken stellt die Differenz des T-Werts eines Falls auf der Y-Achse mit dem jeweilige Ergebnis der Aufnahmebefragung auf der X-Achse dar, wobei Häufungen von Fällen an einer bestimmten Koordinate durch eine stärkere Umrandung gekennzeichnet sind. Punkte im grünen Bereich sind alle Fälle, bei denen Patient:innen eine positive Veränderung zum Katamnesezeitpunkt berichten, und im roten Bereich sind alle Fälle mit Verschlechterungen der berichteten gesundheitsbezogenen Lebensqualität abgetragen. Es zeigt sich sowohl für die Dimension PH als auch für MH, dass (höhere) positive Veränderungen vor allem mit T-Werten < 40 zum Aufnahmezeitpunkt einhergehen, dass sich bei T-Werten zwischen 40 und 45 Verbesserungen und Verschlechterungen etwa in gleicher Weise zeigen, und dass ab einem T-Wert von 45 bei den meisten Fällen Verschlechterungen berichtet werden.

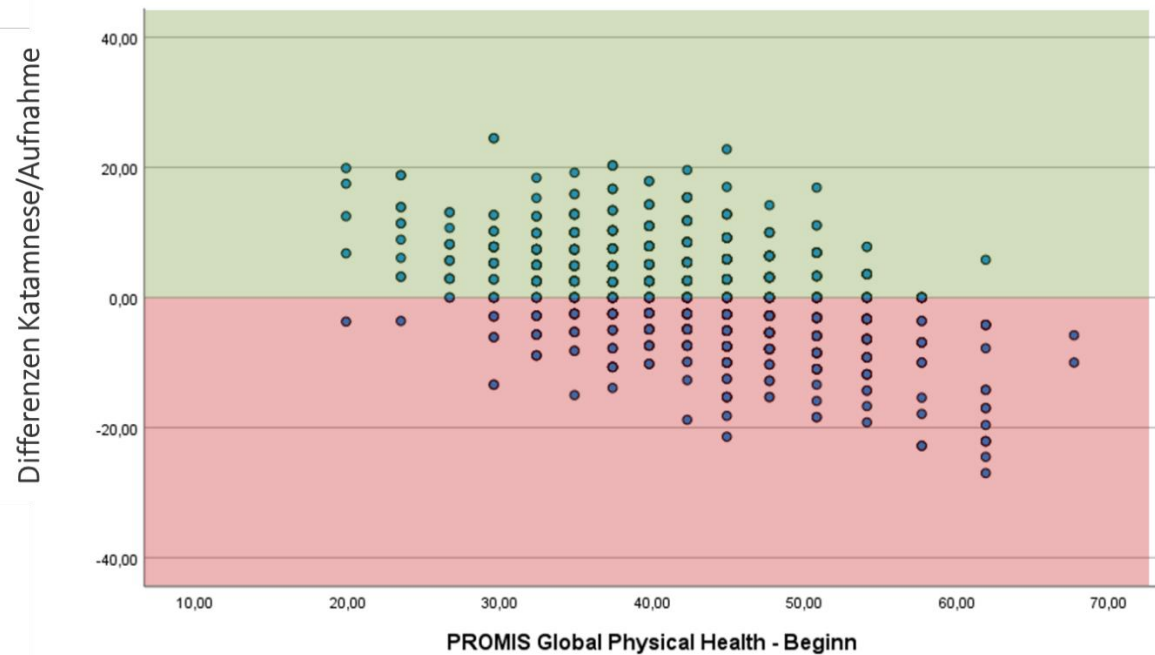


Abb.7: Differenzen zwischen T-Werten von Katamnese- zu Aufnahmezeitpunkt (Y-Achse) in Abhängigkeit des Ergebnisses zum Aufnahmezeitpunkt in der Skala PROMIS-10 PH. Im grünen Bereich liegen alle Fälle mit Verbesserungen der gesundheitsbezogenen Lebensqualität zum Katamnesezeitpunkt, im roten Bereich alle Fälle mit Verschlechterungen.

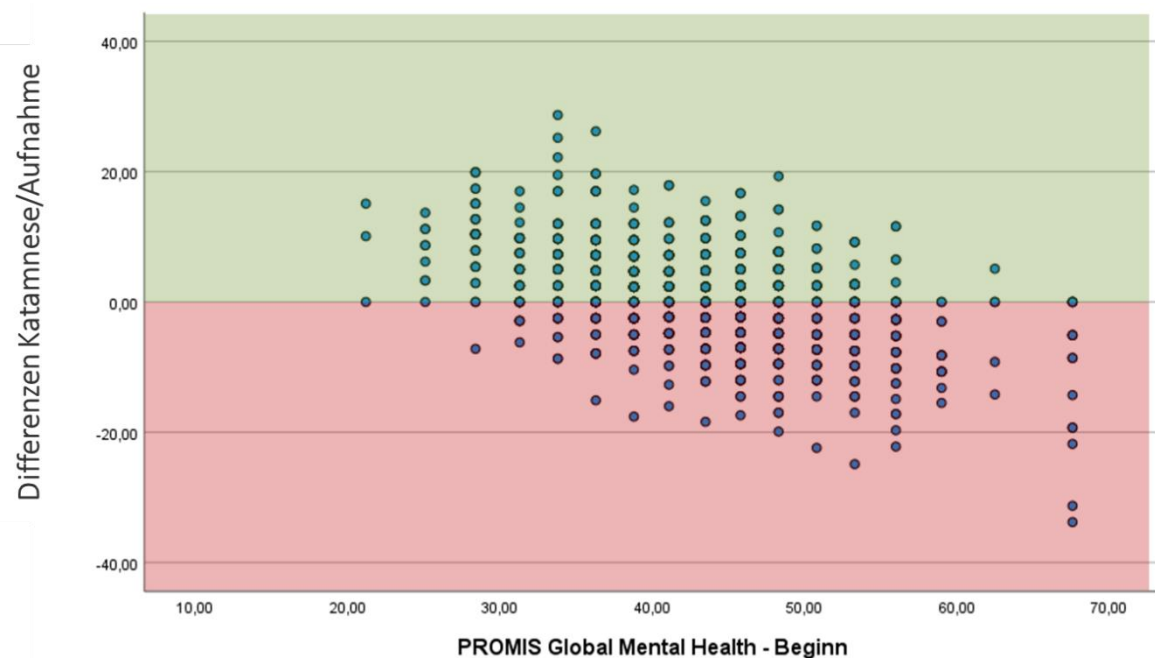


Abb.8: Differenzen zwischen T-Werten von Katamnese- zu Aufnahmezeitpunkt (Y-Achse) in Abhängigkeit des Ergebnisses zum Aufnahmezeitpunkt in der Skala PROMIS-10 MH. Im grünen Bereich liegen alle Fälle mit Verbesserungen der gesundheitsbezogenen Lebensqualität zum Katamnesezeitpunkt, im roten Bereich alle Fälle mit Verschlechterungen.

Da sich eine große Variabilität hinsichtlich der Höhe der Effektstärken zwischen den Einrichtungen in den Skalen des PROMIS-10 zeigte (siehe oben, Abbildungen 3 und 4), wurde zudem untersucht, in welcher Weise sich potentielle Deckeneffekte in den Einrichtungen

verteilen können. Auf der Grundlage der Befunde zu den Verteilungen der Differenzen zwischen Aufnahme- und Katamnesezeitpunkt in Abhängigkeit vom berichteten Gesundheitszustand zum Aufnahmezeitpunkt wurde die Gesamtstichprobe über alle Einrichtungen nach drei Gruppen mit unterschiedlich hoher gesundheitsbezogener Lebensqualität in der körperlichen bzw. psychischen Dimension zum Beginn der Rehabilitation stratifiziert: Der Gruppe mit niedriger gesundheitsbezogener Lebensqualität wurden Patient:innen mit einem T-Wert kleiner/gleich 40 zugeordnet (Dimension PH: 42,9% der Gesamtstichprobe; MH: 30,7%), der Gruppe mit mittlerer gesundheitsbezogener Lebensqualität Patient:innen mit einem T-Wert größer 40 und kleiner/gleich 45 (PH: 28,6%; MH: 24,7%) und mit vergleichsweise hoher gesundheitsbezogener Lebensqualität Patient:innen mit einem T-Wert größer 45 (PH: 28,6%; MH: 44,6%). Im Vergleich der Einrichtungen zeigen sich unterschiedlich hohe Anteile dieser drei Gruppen zwischen den Einrichtungen: Während in der Skala PROMIS-10 PH (Abbildung 9) in der Einrichtung E der höchste Anteil von Patient:innen mit niedriger körperlicher gesundheitsbezogener Lebensqualität behandelt wurde (55,4%), wurden in der Einrichtung A deutlich weniger Patient:innen dieser Gruppe behandelt (17,6%). Ein ähnliches Bild zeigte sich in der Skala PROMIS-10 MH (Abbildung 10): Hier wurden in der Einrichtung F im Vergleich die meisten Patient:innen mit niedriger psychischer gesundheitsbezogener Lebensqualität behandelt (41,3%) und in Einrichtung L der kleinste Anteil dieser Gruppe (8,3%).

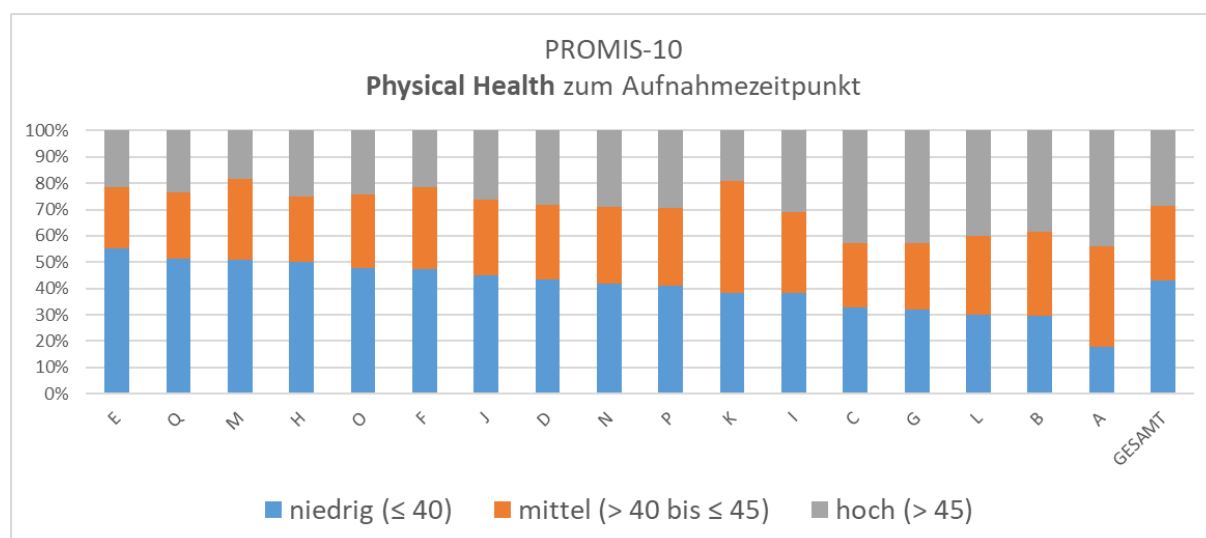


Abb.9: Anteil von Patient:innen mit berichteter niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität im PROMIS-10 Physical Health zum Aufnahmezeitpunkt im Einrichtungsvergleich.

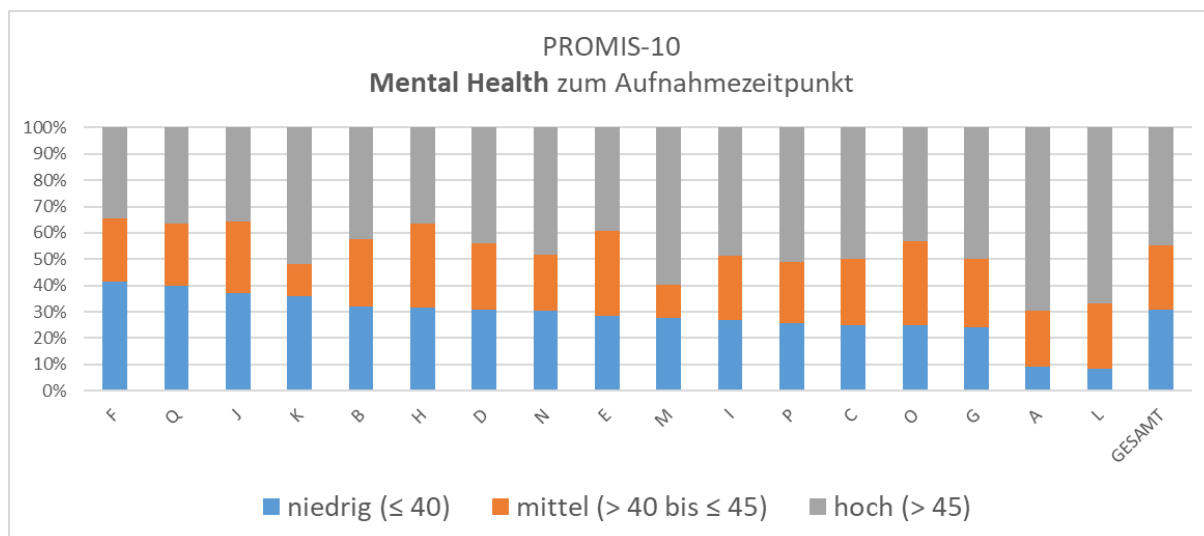


Abb.10: Anteil von Patient:innen mit berichteter niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität im PROMIS-10 Mental Health zum Aufnahmezeitpunkt im Einrichtungsvergleich.

Für die Gesamtstichprobe über alle Einrichtungen werden im folgenden Stichprobenkennwerte und die erreichten Effektstärken der Veränderungen für die drei Gruppen von Patient:innen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt im PROMIS-10 für die Dimension Physical Health (Tabelle 6) und für die Dimension Mental Health (Tabelle 7) analysiert.

PROMIS-10 Physical Health (PH)	Alle Kliniken
Niedrige HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 PH ≤ 40) (n)	465
Anteil an der Klinik-/Gesamtstichprobe	42,9%
PROMIS-10 PH Aufnahme (M/SD)	35,3 (4,3)
PROMIS-10 PH Katamnese (M/SD)	38,9 (6,9)
Korrelation r_{tt} PROMIS-10 PH Aufnahme/Katamnese	0,42
Cohen's d_z	0,56 (p < 0.001)
Mittlere HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 PH: > 40 bis ≤ 45) (n)	310
Anteil an der Klinik-/Gesamtstichprobe	28,6%
PROMIS-10 PH Aufnahme (M/SD)	43,6 (1,3)
PROMIS-10 PH Katamnese (M/SD)	44,3 (6,6)
Korrelation r_{tt} PROMIS-10 PH Aufnahme/Katamnese	0,06
Cohen's d_z	0,10 (p = 0.088)

Tab.6: Deskriptive Statistiken zu Scores im PROMIS-10 Physical Health (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt. Fortsetzung auf der nächsten Seite.

Fortsetzung Tabelle 6:

Hohe HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 PH > 45) (n)	310
Anteil an der Klinik-/Gesamtstichprobe	28,6%
PROMIS-10 PH Aufnahme (M/SD)	51,5 (4,1)
PROMIS-10 PH Katamnese (M/SD)	48,2 (6,6)
Korrelation r_{tt} PROMIS-10 PH Aufnahme/Katamnese	0,22
Cohen's d_z	-0,47 (p < 0.001)

Tab.6 (Fortsetzung): Deskriptive Statistiken zu Scores im PROMIS-10 Physical Health (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt.

Patient:innen mit niedrigen T-Werten im PROMIS-10 PH zum Aufnahmezeitpunkt erreichen zum Katamnesezeitpunkt statistisch signifikante Verbesserungen in der gesundheitsbezogenen Lebensqualität ($d_z = 0,56$; $p < 0.001$), während Patient:innen mit mittleren T-Werten zum Aufnahmezeitpunkt keine statistisch signifikanten Veränderungen berichten ($d_z = 0,10$; $p = 0.088$) und Patient:innen mit hohen Ausgangswerten statistisch signifikante Verschlechterungen zum Katamnesezeitpunkt berichten ($d_z = -0,47$; $p < 0.001$).

PROMIS-10 Mental Health (MH)	Alle Kliniken
Niedrige HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 MH ≤ 40) (n)	344
Anteil an der Klinik-/Gesamtstichprobe	30,7%
PROMIS-10 MH Aufnahme (M/SD)	34,9 (3,9)
PROMIS-10 MH Katamnese (M/SD)	39,3 (6,6)
Korrelation r_{tt} PROMIS-10 MH Aufnahme/Katamnese	0,30
Cohen's d_z	0,67 (p < 0.001)
Mittlere HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 MH: > 40 bis ≤ 45) (n)	276
Anteil an der Klinik-/Gesamtstichprobe	24,7%
PROMIS-10 MH Aufnahme (M/SD)	42,6 (1,2)
PROMIS-10 MH Katamnese (M/SD)	43,4 (5,7)
Korrelation r_{tt} PROMIS-10 MH Aufnahme/Katamnese	0,17
Cohen's d_z	0,15 (p = 0.014)

Tab.7: Deskriptive Statistiken zu Scores im PROMIS-10 Mental Health (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt. Fortsetzung auf der nächsten Seite.

Fortsetzung Tabelle 7:

Hohe HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 MH > 45) (n)	499
Anteil an der Klinik-/Gesamtstichprobe	44,6%
PROMIS-10 MH Aufnahme (M/SD)	50,4 (4,9)
PROMIS-10 MH Katamnese (M/SD)	47,7 (7,2)
Korrelation r_{tt} PROMIS-10 MH Aufnahme/Katamnese	0,38
Cohen's d_z	-0,39 (p < 0.001)

Tab.7 (Fortsetzung): Deskriptive Statistiken zu Scores im PROMIS-10 Mental Health (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt.

In der Skala Mental Health des PROMIS-10 weisen Patient:innen ebenfalls statistisch signifikante Verbesserungen in der gesundheitsbezogenen Lebensqualität bei niedrigen ($d_z = 0,67$; $p < 0.001$) bzw. mittleren ($d_z = 0,15$; $p = 0.014$) T-Werten zum Aufnahmezeitpunkt auf, während Patient:innen mit hohen T-Werten zum Aufnahmezeitpunkt statistisch signifikante Verschlechterungen berichten ($d_z = -0,39$; $p < 0.001$).

Für den VR-12 werden die drei Strata mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität hinsichtlich ihrer Stichprobenkennwerte und Effektstärken der Veränderungen für den Physical und Mental Component Score ebenfalls verglichen (Tabellen 8 und 9).

VR-12 Physical Component Score (PCS)	5 Kliniken mit VR-12
Niedrige HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 PCS \leq 40) (n)	187
Anteil an der Klinik-/Gesamtstichprobe	53,9%
VR-12 PCS Aufnahme (M/SD)	31,1 (6,7)
VR-12 PCS Katamnese (M/SD)	36,1 (9,4)
Korrelation r_{tt} VR-12 PCS Aufnahme/Katamnese	0,47
Cohen's d_z	0,57 (p < 0.001)

Tab.8: Deskriptive Statistiken zum VR-12 Physical Component Score (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt. Fortsetzung auf der nächsten Seite.

Fortsetzung Tabelle 8:

Mittlere HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 PCS: > 40 bis ≤ 45) (n)	63
Anteil an der Klinik-/Gesamtstichprobe	18,2%
VR-12 PCS Aufnahme (M/SD)	42,4 (1,4)
VR-12 PCS Katamnese (M/SD)	43,2 (6,9)
Korrelation r_{tt} VR-12 PCS Aufnahme/Katamnese	0,04
Cohen's d_z	0,11 (p = 0.366)
Hohe HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 PCS > 45) (n)	97
Anteil an der Klinik-/Gesamtstichprobe	28,0%
VR-12 PCS Aufnahme (M/SD)	49,9 (3,5)
VR-12 PCS Katamnese (M/SD)	47,9 (6,5)
Korrelation r_{tt} VR-12 PCS Aufnahme/Katamnese	0,09
Cohen's d_z	-0,28 (p = 0.007)

Tab.8 (Fortsetzung): Deskriptive Statistiken zum VR-12 Physical Component Score (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt.

Patient:innen mit niedrigen T-Werten im VR-12 PCS zum Aufnahmezeitpunkt erreichen zum Katamnesezeitpunkt statistisch signifikante Verbesserungen in der gesundheitsbezogenen Lebensqualität ($d_z = 0,57$; $p < 0.001$), während Patient:innen mit mittleren T-Werten zum Aufnahmezeitpunkt keine statistisch signifikanten Veränderungen berichten ($d_z = 0,11$; $p = 0.366$) und Patient:innen mit hohen Ausgangswerten statistisch signifikante Verschlechterungen zum Katamnesezeitpunkt berichten ($d_z = -0,28$; $p = 0.007$).

VR-12 Mental Component Score (MCS)	5 Kliniken mit VR-12
Niedrige HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 MCS ≤ 40) (n)	139
Anteil an der Klinik-/Gesamtstichprobe	40,2%
VR-12 MCS Aufnahme (M/SD)	31,0 (6,3)
VR-12 MCS Katamnese (M/SD)	37,1 (10,7)
Korrelation r_{tt} VR-12 MCS Aufnahme/Katamnese	0,27
Cohen's d_z	0,56 (p < 0.001)

Tab.9: Deskriptive Statistiken zum VR-12 Mental Component Score (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt. Fortsetzung auf der nächsten Seite.

Fortsetzung Tabelle 9:

Mittlere HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 MCS: > 40 bis ≤ 45) (n)	47
Anteil an der Klinik-/Gesamtstichprobe	13,6%
VR-12 MCS Aufnahme (M/SD)	42,2 (1,3)
VR-12 MCS Katamnese (M/SD)	44,9 (9,3)
Korrelation r_{tt} VR-12 MCS Aufnahme/Katamnese	0,34
Cohen's d_z	0,30 (p = 0.047)
Hohe HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 MCS > 45) (n)	160
Anteil an der Klinik-/Gesamtstichprobe	46,2%
VR-12 MCS Aufnahme (M/SD)	54,2 (6,2)
VR-12 MCS Katamnese (M/SD)	50,0 (10,5)
Korrelation r_{tt} VR-12 MCS Aufnahme/Katamnese	0,22
Cohen's d_z	-0,38 (p < 0.001)

Tab.9 (Fortsetzung): Deskriptive Statistiken zum VR-12 Mental Component Score (zur Aufnahme und zum Katamnesezeitpunkt) und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt.

Im Mental Component Score des VR-12 weisen Patient:innen ebenfalls statistisch signifikante Verbesserungen in der gesundheitsbezogenen Lebensqualität bei niedrigen ($d_z = 0,56$; $p < 0.001$) bzw. mittleren ($d_z = 0,30$; $p = 0.047$) T-Werten zum Aufnahmezeitpunkt auf, während Patient:innen mit hohen T-Werten zum Aufnahmezeitpunkt statistisch signifikante Verschlechterungen berichten ($d_z = -0,38$; $p < 0.001$).

PROMIS-10: Vergleich der Einrichtungen nach stratifizierter Risikoadjustierung

Ein Vergleich des Outcomes zwischen den Einrichtungen sollte nur unter Berücksichtigung einrichtungsspezifischer Unterschiede hinsichtlich der Merkmale behandelter Patient:innen vorgenommen werden. Auf den Beobachtungen von Deckeneffekten (siehe „Deckeneffekte im PROMIS-10 und VR-12“, S. 26) aufbauend und aufgrund einer verbesserten Schätzung von auf Patient:innenmerkmalen basierenden Erwartungswerten für das durchschnittlich erreichbare Outcome (siehe ausführlich im Anhang: „Analyse der Linearität des Zusammenhangs von Aufnahmewert zu Katamnesewert im PROMIS-10 und der Veränderungen der Rangreihen im Einrichtungsvergleich in Abhängigkeit des Risikoadjustierungsmodells“, S.63) wurde für den Einrichtungsvergleich eine nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt stratifizierte Risikoadjustierung durchgeführt. Parameter der stratifizierten Regressionsmodelle und der in die Risikoadjustierung einbezogenen Patient:innenmerkmale sind im Anhang (S. 25) tabelliert. Aus jeweils drei Regressionsanalysen pro Dimension resultiert für jede:n Patient:in ein vorhergesagtes

Outcome, das aufgrund des Zusammenhangs der Merkmale aller Patient:innen aus allen Einrichtungen für den PROMIS-10 zu erwarten wäre. Die Effektstärke Cohen's d_z der Differenzen des zum Katamnesezeitpunkt erreichten Outcomes und des zu erwarteten Outcomes beschreibt das Maß, in dem sich das in einer Einrichtung erreichte Outcome vom durchschnittlichen Outcome aller Einrichtungen unterscheidet.

In der Skala Physical Health des PROMIS-10 (Abbildung 11) wird für acht Einrichtungen ein höheres Outcome beobachtet, als es unter Berücksichtigung der spezifischen Patient:innenmerkmale durchschnittlich zu erwarten gewesen wäre (Range Cohen's d_z von $d_z = 0,01$ bis $d_z = 0,35$), allerdings schneidet bei diesen Einrichtungen das jeweilige Konfidenzintervall der Einrichtungen die Nulllinie, so dass für keine dieser Einrichtungen eine Aussage über ein überdurchschnittliches Outcome mit einem α -Niveau von 0.05 statistisch abgesichert wäre. Für neun Einrichtungen wird ein niedrigeres Outcome beobachtet, als es unter Berücksichtigung der spezifischen Patient:innenmerkmale durchschnittlich zu erwarten gewesen wäre (Range Cohen's d_z von $d_z = -0,24$ bis $d_z = 0,02$), aber nur für Einrichtung „J“ ist das unterdurchschnittliche Ergebnis statistisch signifikant abgesichert ($p < 0.05$).

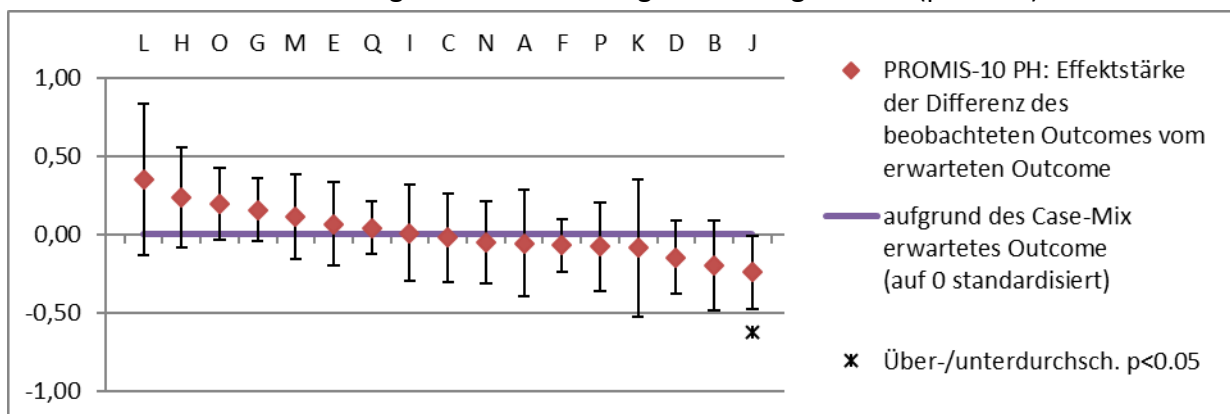


Abb.11: Adjustierte Ergebnisse für den PROMIS-10 PH: Vergleich erreichter und risikoadjustierter/erwarteter Outcomes aller Einrichtungen (höhere Werte entsprechen einer größeren Verbesserung als aufgrund des Case-Mix zu erwarten wäre). Sortiert nach 1. statistischer Absicherung eines überdurchschnittlichen/durchschnittlichen/unterdurchschnittlichen Ergebnisses und 2. nach Höhe der Effektstärken der Differenzen des beobachteten Outcomes vom erwarteten Outcome.

Für einen Vergleich der Outcomes der zwischen den Einrichtungen in der psychischen Dimension des PROMIS-10 wurden in gleicher Weise drei Regressionsmodelle stratifiziert nach dem PROMIS-10 MH Wert zum Aufnahmezeitpunkt erstellt. Parameter der Regressionsmodelle und der in die Risikoadjustierung einbezogenen Patient:innenmerkmale sind im Anhang (S. 27) tabelliert.

In der Skala Mental Health des PROMIS-10 (Abbildung 12) wird für acht Einrichtungen ein höheres Outcome beobachtet, als es unter Berücksichtigung der spezifischen Patient:innenmerkmale durchschnittlich zu erwarten gewesen wäre (Range Cohen's d_z von $d_z = 0,02$ bis $d_z = 0,52$), wobei nur für die Einrichtungen „L“ und „Q“ die Aussage eines überdurchschnittlichen Outcomes statistisch signifikant ($p < 0.05$) abgesichert wäre. Für neun Einrichtungen wird ein niedrigeres Outcome beobachtet, als es unter Berücksichtigung der spezifischen Patient:innenmerkmale durchschnittlich zu erwarten gewesen wäre (Range

Cohen's d_z von $d_z = -0,19$ bis $d_z = -0,02$), allerdings schneidet bei diesen Einrichtungen das jeweilige Konfidenzintervall der Einrichtungen die Nulllinie, so dass für keine dieser Einrichtungen eine Aussage über ein unterdurchschnittliches Outcome mit einem α -Niveau von 0.05 statistisch abgesichert wäre.

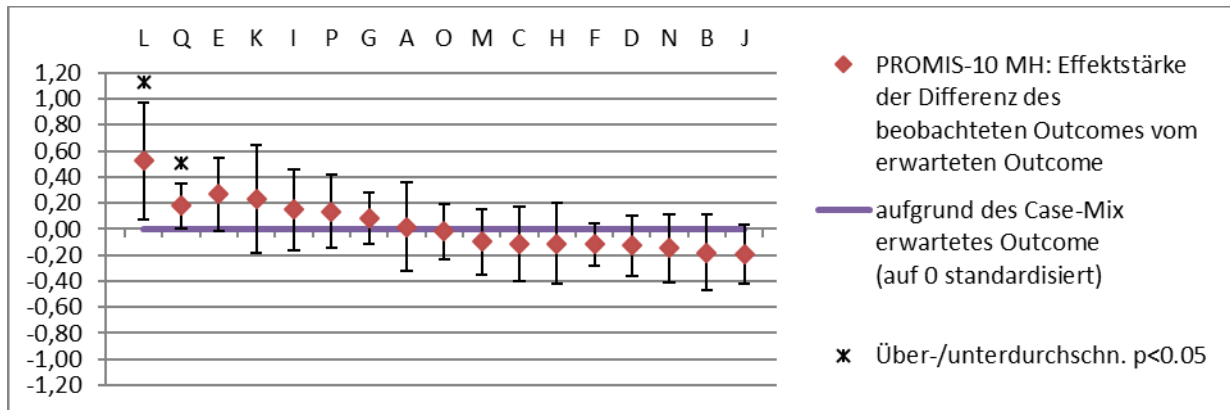


Abb.12: Adjustierte Ergebnisse für den PROMIS-10 MH: Vergleich erreichter und risikoadjustierter/erwarteter Outcome aller Kliniken (höhere Werte entsprechen einer größeren Verbesserung als aufgrund des Case-Mix zu erwarten wäre). Sortiert nach 1. statistischer Absicherung eines überdurchschnittlichen/durchschnittlichen/unterdurchschnittlichen Ergebnisses und 2. nach Höhe der Effektstärken der Differenzen des beobachteten Outcomes vom erwarteten Outcome.

VR-12: Vergleich der Einrichtungen nach stratifizierter Risikoadjustierung

Die Risikoadjustierung für den VR-12 wurde in Analogie zum PROMIS-10 ebenfalls stratifiziert nach Gruppen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt durchgeführt. Parameter der Regressionsmodelle und der in die Risikoadjustierung einbezogenen Patient:innenmerkmale sind im Anhang (S. 29) tabelliert.

Im Physical Component Score des VR-12 (Abbildung 13) wird für vier Einrichtungen ein höheres Outcome beobachtet, als es unter Berücksichtigung der spezifischen Patient:innenmerkmale durchschnittlich zu erwarten gewesen wäre, und für eine Einrichtung ein niedrigeres (Range Cohen's d_z von $d_z = -0,19$ bis $d_z = 0,16$). Allerdings schneidet bei allen Einrichtungen das jeweilige Konfidenzintervall der Einrichtungen die Nulllinie, so dass für keine der Einrichtungen eine statistisch abgesicherte Aussage ($p < 0.05$) über ein erreichtes über- oder unterdurchschnittliches Outcome getroffen werden kann.

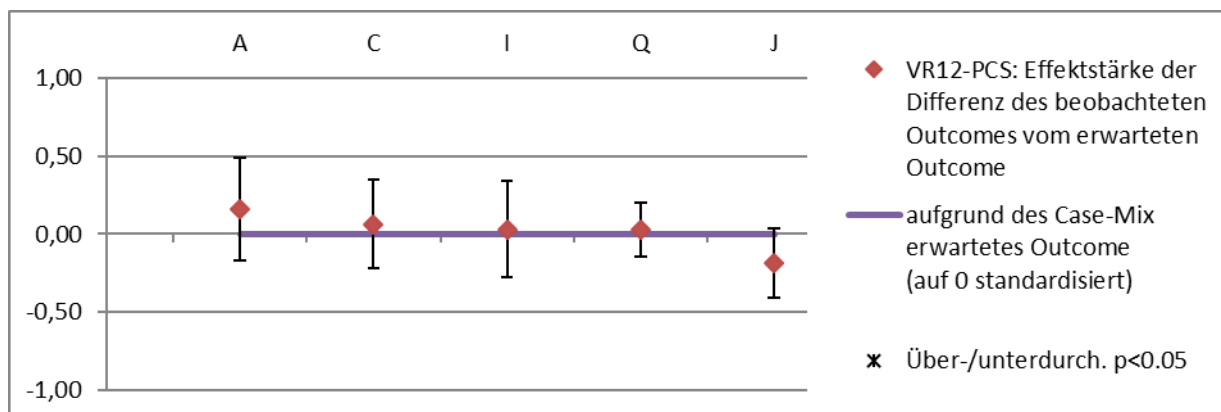


Abb.13: Adjustierte Ergebnisse für den VR-12 PCS: Vergleich erreichter und risikoadjustierter/erwarteter Outcome aller Kliniken (höhere Werte entsprechen einer größeren Verbesserung als aufgrund des Case-Mix zu erwarten wäre). Sortiert nach 1. statistischer Absicherung eines überdurchschnittlichen/durchschnittlichen/unterdurchschnittlichen Ergebnisses und 2. nach Höhe der Effektstärken der Differenzen des beobachteten Outcomes vom erwarteten Outcome.

Für den Mental Component Score des VR-12 sind die Parameter der Regressionsmodelle und der in die Risikoadjustierung einbezogenen Patient:innenmerkmale im Anhang (S. 31) tabelliert. Die Outcomes zweier Einrichtungen liegen über und die Outcomes dreier Einrichtungen unter den durchschnittlich erwarteten Outcomes (Range Cohen's d_z von $d_z = -0,14$ bis $d_z = 0,10$), allerdings sind diese Ergebnisse nicht mit einem α -Niveau von 0.05 statistisch abgesichert (Abbildung 14).

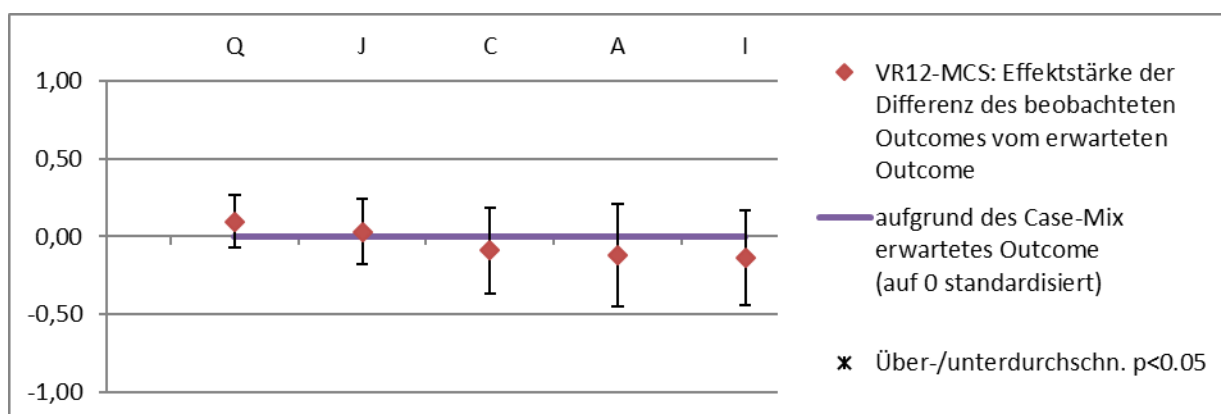


Abb.14: Adjustierte Ergebnisse für die psychische Skala des VR-12: Vergleich erreichter und risikoadjustierter/erwarteter Outcome aller Kliniken (höhere Werte entsprechen einer größeren Verbesserung als aufgrund des Case-Mix zu erwarten wäre). Sortiert nach 1. statistischer Absicherung eines überdurchschnittlichen/durchschnittlichen/unterdurchschnittlichen Ergebnisses und 2. nach Höhe der Effektstärken der Differenzen des beobachteten Outcomes vom erwarteten Outcome.

Vergleich von Rehabilitationseinrichtungen mit dem Patient-Reported Outcome Quality Index (ProQI)

Im Folgenden werden die Ergebnisse zu den berechneten instrumentenspezifischen und Gesamt-ProQIs berichtet. Die ProQIs bilden das erreichte Outcome einer Einrichtung unter

Berücksichtigung einrichtungsspezifischer Patient:innenmerkmale im Vergleich mit den anderen Einrichtungen in einem einzelnen Wert ab (für die methodischen Grundlagen siehe „Vergleich von Rehabilitationseinrichtungen mit dem Patient-Reported Outcome Quality Index (ProQI)“, S. 17). Der Durchschnittswert für den ProQI jedes Instruments beträgt 90 Punkte.

In Abbildung 15 sind die ProQIs für den PROMIS-10 dargestellt. Der $\text{ProQI}_{\text{PROMIS-10}}$ berechnet sich aus $\frac{1}{2} * \text{ProQI}_{\text{PROMIS-10 PH+}} + \frac{1}{2} * \text{ProQI}_{\text{PROMIS-10 MH}}$. Zehn Einrichtungen erreichen für das Outcome im PROMIS-10 einen ProQI > 90 Punkte, sieben Einrichtungen einen ProQI < 90 Punkte (ProQI-Range: 84,05 bis 100 Punkte).

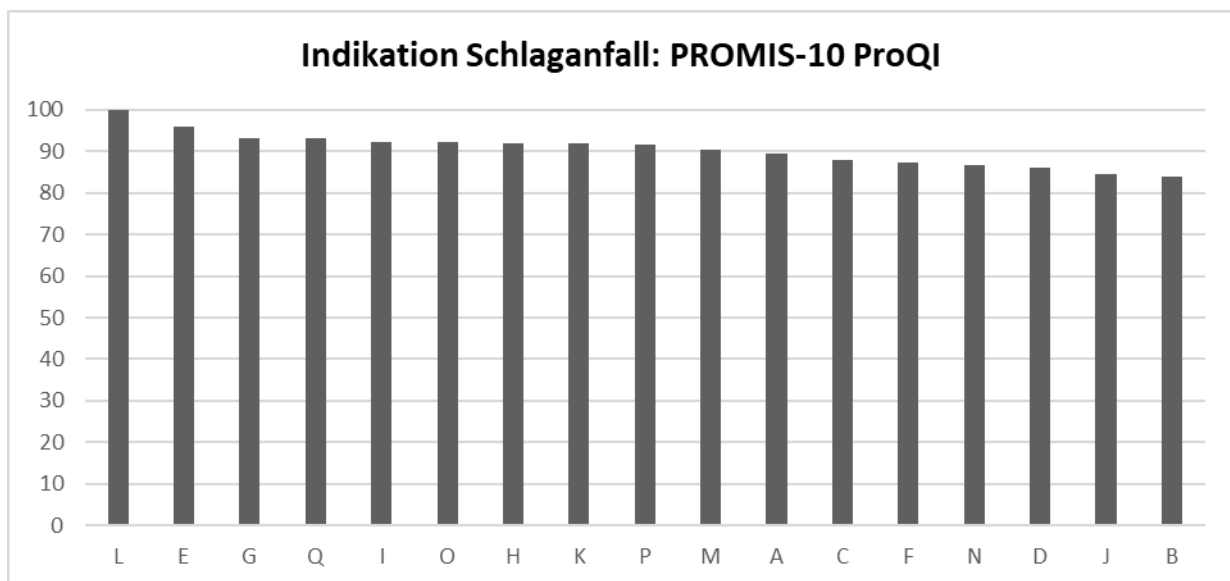


Abb.15: ProQIs für den PROMIS-10 über alle teilnehmenden Einrichtungen.

ProQIs für das Outcome im VR-12 (PCS & MCS) sind für die fünf Einrichtungen, die auch den VR-12 eingesetzt hatten, in Abbildung 16 dargestellt. Der $\text{ProQI}_{\text{VR-12}}$ berechnet sich aus $\frac{1}{2} * \text{ProQI}_{\text{VR12-PCS}} + \frac{1}{2} * \text{ProQI}_{\text{VR12-MCS}}$. Zwei Einrichtungen erreichen für das Outcome im VR-12 einen ProQI > 90 Punkte, eine Einrichtung einen ProQI von 90 Punkten und zwei Einrichtungen einen ProQI < 90 Punkte (ProQI-Range: 88 bis 92 Punkte).

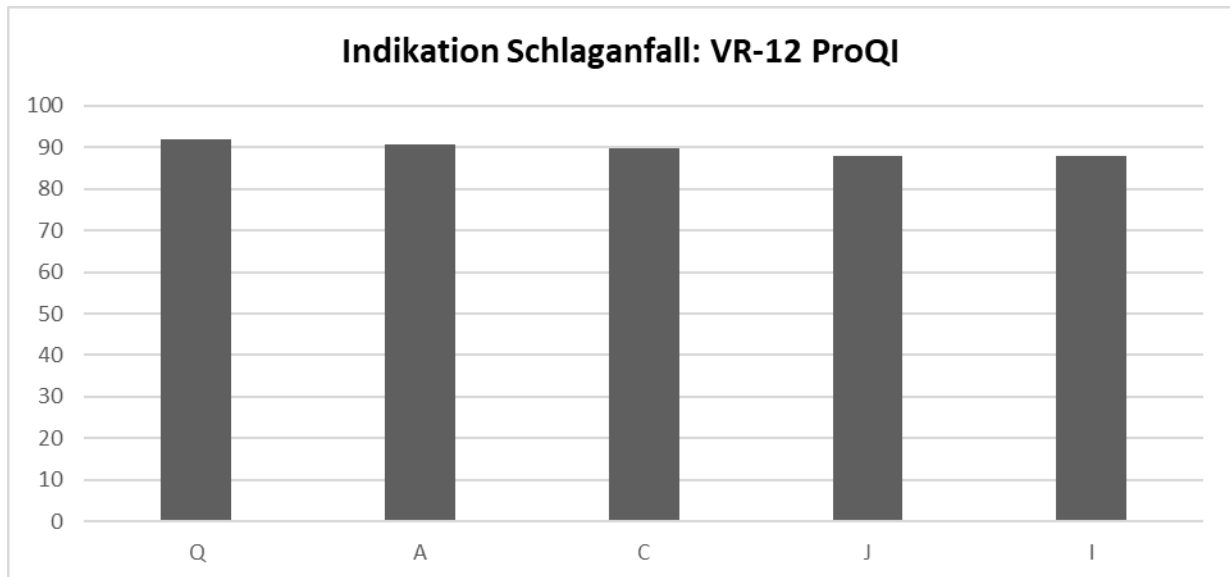


Abb.16: ProQIs für den VR-12 über alle teilnehmenden Einrichtungen.

Ein ProQI, der das Gesamtergebnis übergreifend für den PROMIS-10 und für den VR-12 abbildet, kann nur für die fünf Einrichtungen berechnet werden, die beide Instrumente eingesetzt hatten. Der $ProQI_{Gesamt}$ berechnet hier sich aus $(ProQI_{PROMIS-10} + ProQI_{VR-12}) / 2$. Die Ergebnisse sind in Abbildung 17 dargestellt.

Im Gesamtergebnis erreicht eine Einrichtung einen Gesamt-ProQI > 90 Punkte, zwei Einrichtungen einen ProQI von 90 Punkten und zwei Einrichtungen einen ProQI < 90 Punkte (ProQI-Range: 86 bis 92 Punkte).

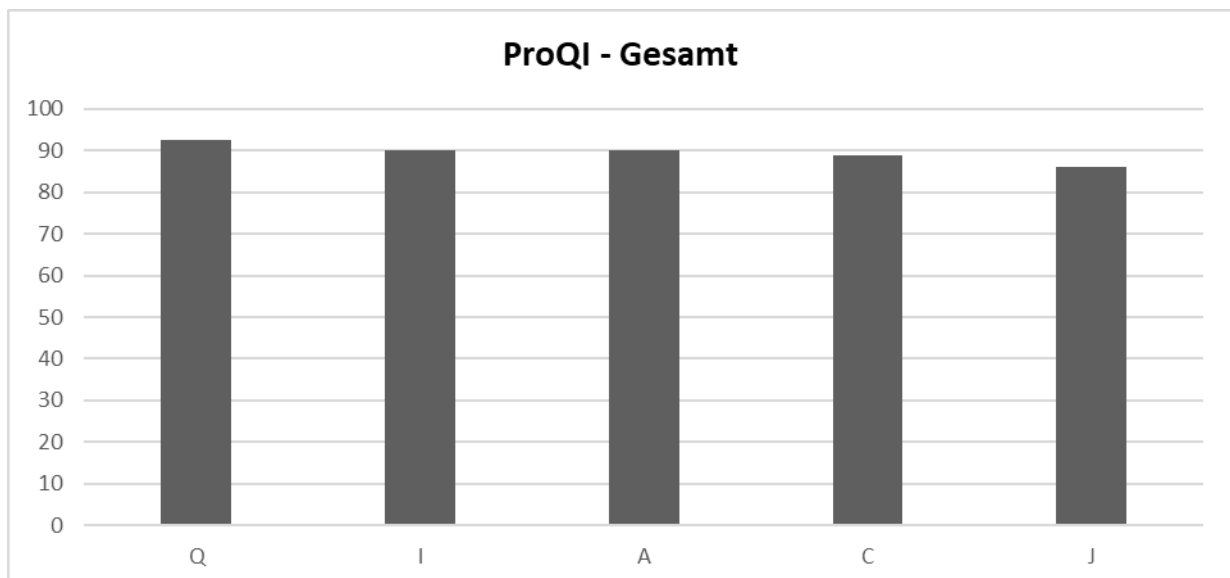


Abb.17: Gesamt-ProQIs aggregiert über $ProQI_{PROMIS-10}$ und $ProQI_{VR-12}$ für alle teilnehmenden Einrichtungen.

Vergleich der Änderungssensitivität von PROMIS-10 und VR-12

Zum Vergleich der Änderungssensitivität von PROMIS-10 und VR-12 werden für die Unterstichprobe, in der Patient:innen sowohl den PROMIS-10 als auch den VR-12 zu beiden Messzeitpunkten vollständig ausgefüllt haben, Korrelation zwischen den Instrumenten, Stichprobenkennwerte und Effektstärken der (unadjustierten) Veränderungen (Cohen's d_z) und Re-Test-Reliabilitätskoeffizienten (Korrelationen der Testergebnisse zwischen Aufnahme- und Katamnesezeitpunkt; r_{tt}) sowohl für die Gesamtheit dieser Auswahl von Patient:innen als auch stratifiziert nach Gruppen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität in den Skalen Physical Health bzw. Mental Health des PROMIS-10 zum Aufnahmezeitpunkt berechnet.

Für $n = 336$ Patient:innen liegen Ergebnisse in der **physischen Dimension** des PROMIS-10 und VR-12 zu beiden Messzeitpunkten vor (Tabelle 10).

Die Korrelation der körperlichen Dimension zwischen den Instrumenten beträgt $r_{PH/PCS} = 0,71$.

Die T-Werte zur Aufnahme (42,4) und Katamnese (42,9) in der Skala PROMIS-10 PH sind höher als für die selben Patient:innen im PCS des VR-12 (38,3 bzw. 40,6).

Die Effektstärken der Veränderung zwischen Aufnahme- und Katamnesezeitpunkt fallen in der Gesamtstichprobe für den VR-12 PCS höher aus ($d_z = 0,27$; $p < 0.001$) als für den PROMIS-10 PH ($d_z = 0,08$; $p = 0.138$).

In der Betrachtung der einzelnen Strata nach gesundheitsbezogener Lebensqualität im PROMIS-10 PH zum Aufnahmezeitpunkt zeigt sich im PROMIS-10 PH bei Patient:innen mit niedriger gesundheitsbezogener Lebensqualität eine höhere Effektstärke ($d_z = 0,45$; $p < 0.001$) als für den VR-12 PCS ($d_z = 0,31$; $p < 0.001$). In der Gruppe mit mittlerer gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt ist die Effektstärke im VR-12 PCS ($d_z = 0,27$; $p = 0.011$) höher als im PROMIS-10 ($d_z = 0,05$; $p = 0.618$). Während in der Gruppe mit hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt eine negative Effektstärke der Veränderung im PROMIS-10 PH resultiert ($d_z = -0,37$; $p < 0.001$), ergibt sich im VR-12 PCS eine positive Effektstärke ($d_z = 0,19$; $p = 0.058$).

Im VR-12 PCS sind in allen drei Gruppen die T-Werte zur Aufnahme und zur Katamnese niedriger als im PROMIS-10 PH.

Körperliche Dimension im Vergleich	PROMIS-10 PH	VR-12 PCS
Gesamtstichprobe – N = 336	$r_{PH/PCS} = 0,71$	
T-Wert Aufnahme (M/SD)	42,4 (8,2)	38,3 (9,9)
T-Wert Katamnese (M/SD)	42,9 (8,2)	40,6 (9,8)
Korrelation r_{tt} Aufnahme/Katamnese	0,66	0,62
Cohen's d_z	0,08 ($p = 0.138$)	0,27 ($p < 0.001$)
Niedrige HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 PH ≤ 40) – n = 141	$r_{PH/PCS} = 0,59$	
Anteil an der Gesamtstichprobe	42,0%	
T-Wert Aufnahme (M/SD)	34,8 (4,7)	31,5 (8,7)
T-Wert Katamnese (M/SD)	37,5 (6,7)	34,1 (8,1)
Korrelation r_{tt} Aufnahme/Katamnese	0,48	0,51
Cohen's d_z	0,45 ($p < 0.001$)	0,31 ($p < 0.001$)
Mittlere HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 PH: > 40 bis ≤ 45) – n = 94	$r_{PH/PCS} = 0,14$	
Anteil an der Gesamtstichprobe	28,0%	
T-Wert Aufnahme (M/SD)	43,8 (1,3)	40,8 (7,6)
T-Wert Katamnese (M/SD)	44,2 (7,1)	43,4 (8,6)
Korrelation r_{tt} Aufnahme/Katamnese	-0,02	0,30
Cohen's d_z	0,05 ($p = 0.618$)	0,27 ($p = 0.011$)
Hohe HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 PH > 45) – n = 101	$r_{PH/PCS} = 0,45$	
Anteil an der Gesamtstichprobe	30,1%	
T-Wert Aufnahme (M/SD)	51,8 (4,3)	45,6 (7,4)
T-Wert Katamnese (M/SD)	49,4 (5,7)	47,1 (6,2)
Korrelation r_{tt} Aufnahme/Katamnese	0,24	0,37
Cohen's d_z	-0,37 ($p < 0.001$)	0,19 ($p = 0.058$)

Tab.10: Zusammenhänge von PROMIS-10 PH und VR-12 PCS, deskriptive Statistiken und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt für die Gesamtstichprobe der Patient:innen, die beide Instrumente vollständig beantwortet haben, und stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt im PROMIS-10 PH.

Für die **psychische Dimension** des PROMIS-10 und des VR-12 liegen zu beiden Messzeitpunkten Ergebnisse von n = 333 Patient:innen vor (Tabelle 11).

Die Korrelation der psychischen Dimension zwischen den Instrumenten beträgt $r_{MH/MCS} = 0,73$.

Die T-Werte zur Aufnahme (43,4) und Katamnese (44,1) in der Skala PROMIS-10 MH sind etwa in gleicher Höhe wie für die selben Patient:innen im MCS des VR-12 (43,1 bzw. 44,2).

Die Effektstärken der Veränderung zwischen Aufnahme- und Katamnesezeitpunkt fallen in der Gesamtstichprobe für den VR-12 MCS gleich hoch aus ($d_z = 0,10$; $p = 0.068$) wie für den PROMIS-10 MH ($d_z = 0,10$; $p = 0.058$).

In der Betrachtung der einzelnen Strata nach gesundheitsbezogener Lebensqualität im PROMIS-10 MH zum Aufnahmezeitpunkt zeigt sich im PROMIS-10 MH bei Patient:innen mit niedriger gesundheitsbezogener Lebensqualität eine höhere Effektstärke ($d_z = 0,63$; $p < 0.001$) als für den VR-12 MCS ($d_z = 0,31$; $p = 0.002$). In der Gruppe mit mittlerer gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt sind die Effektstärken von PROMIS-10 MH ($d_z = 0,14$; $p = 0.224$) und VR-12 MCS ($d_z = 0,14$; $p = 0.210$) gleich hoch. In beiden Instrumenten zeigen sich in der Gruppe mit hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt negative Effektstärken der Veränderung, wobei diese im PROMIS-10 PH stärker ausgeprägt ist ($d_z = -0,27$; $p = 0.002$) als im VR-12 MCS ($d_z = -0,09$; $p = 0.310$).

Im Gegensatz zum VR-12 PCS sind im VR-12 MCS nur in der Gruppe mit niedriger gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt die durchschnittlichen T-Werte zur Aufnahme und Katamnese niedriger als im PROMIS-10 MH. In der Gruppen mit mittlerer gesundheitsbezogener Lebensqualität sind die T-Werte etwa gleich hoch, und in der Gruppe mit hoher gesundheitsbezogener Lebensqualität zur Aufnahme sind sie höher als im PROMIS-10 MH.

Psychische Dimension im Vergleich	PROMIS-10 MH	VR-12 MCS
Gesamtstichprobe – N = 333	$r_{MH/MCS} = 0,73$	
T-Wert Aufnahme (M/SD)	43,4 (7,9)	43,1 (12,2)
T-Wert Katamnese (M/SD)	44,1 (7,7)	44,2 (12,1)
Korrelation r_{tt} Aufnahme/Katamnese	0,66	0,55
Cohen's d_z	0,10 ($p = 0.058$)	0,10 ($p = 0.068$)

Tab.11: Zusammenhänge von PROMIS-10 MH und VR-12 MCS, deskriptive Statistiken und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt für die Gesamtstichprobe der Patient:innen, die beide Instrumente vollständig beantwortet haben, und stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt im PROMIS-10 MH. Fortsetzung auf der nächsten Seite.

Fortsetzung Tabelle 11:

Niedrige HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 MH ≤ 40) – n = 108		$r_{MH/MCS} = 0,55$	
Anteil an der Gesamtstichprobe	32,4%		
T-Wert Aufnahme (M/SD)	34,6 (4,0)	32,4 (9,6)	
T-Wert Katamnese (M/SD)	38,3 (5,8)	36,1 (10,6)	
Korrelation r_{tt} Aufnahme/Katamnese	0,30	0,26	
Cohen's d_z	0,63 (p < 0.001)	0,31 (p = 0.002)	
Mittlere HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 MH: > 40 bis ≤ 45) – n = 82		$r_{MH/MCS} = 0,01$	
Anteil an der Gesamtstichprobe	24,6%		
T-Wert Aufnahme (M/SD)	42,6 (1,2)	42,5 (8,9)	
T-Wert Katamnese (M/SD)	43,3 (5,7)	43,9 (9,4)	
Korrelation r_{tt} Aufnahme/Katamnese	0,24	0,39	
Cohen's d_z	0,14 (p = 0.224)	0,14 (p = 0.210)	
Hohe HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 MH > 45) – n = 143		$r_{MH/MCS} = 0,33$	
Anteil an der Gesamtstichprobe	42,9%		
T-Wert Aufnahme (M/SD)	50,6 (4,7)	51,5 (8,7)	
T-Wert Katamnese (M/SD)	48,9 (6,7)	50,5 (10,7)	
Korrelation r_{tt} Aufnahme/Katamnese	0,42	0,33	
Cohen's d_z	-0,27 (p = 0.002)	-0,09 (p = 0.310)	

Tab.11 (Fortsetzung): Zusammenhänge von PROMIS-10 MH und VR-12 MCS, deskriptive Statistiken und Effektstärken der unadjustierten Differenzen von Katamnesezeitpunkt zu Aufnahmezeitpunkt für die Gesamtstichprobe der Patient:innen, die beide Instrumente vollständig beantworten haben, und stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt im PROMIS-10 MH.

Diskussion

In der hier durchgeführten Pilotstudie sollte erstmals der Einsatz von Patient-Reported Outcome Measures (PROMs) in neurologischen Reha-Fachabteilungen zur einrichtungsvergleichenden Bewertung der Ergebnisqualität der Rehabilitationsbehandlung von Patient:innen mit Schlaganfall (ICD-10 I60 bis I64) in der Anschlussheilbehandlung (Phase D) erprobt werden. Die Studie wurde als Kooperationsprojekt zwischen dem Institut für Medizinische Psychologie am Universitätsklinikum Hamburg-Eppendorf (UKE) und dem REHAPORTAL der Qualitätskliniken.de GmbH mit ihren Mitgliedseinrichtungen und ihrem assoziierten Expertengremium (AG Ergebnismessung) durchgeführt. In der Konzeption der Studie wurde in gemeinsamen Diskussionen und den Empfehlungen des International Consortium for Health Outcomes Measurement (ICHOM) folgend der **PROMIS-10 Global Health** als generisches Instrument zur Messung der patient:innenberichteten körperlichen und psychischen gesundheitsbezogenen Lebensqualität als primäres PROM festgelegt. Zusätzlich wurde als optionales PROM der **Veterans RAND 12-Item Health Survey (VR-12)** in die Studie einbezogen, nachdem dieses Instrument in vorangegangenen Pilotstudien in den Reha-Fachabteilungen Orthopädie, Psychosomatik und Onkologie bereits erfolgreich eingesetzt werden konnte, um die Messeigenschaften hinsichtlich der Änderungssensitivität beider generischen Instrumente vergleichen zu können. Ziel der Analysen war es, Veränderungen in der Gesundheit von Patient:innen katamnestisch nach der Rehabilitationsbehandlung anhand der eingesetzten PROMs durch Effektstärkemaße zu beschreiben, die Ergebnisqualität von Rehabilitationseinrichtungen durch Berücksichtigung des einrichtungsspezifischen Case-Mix fair miteinander zu vergleichen und dazu auch einen Indexwert zu entwickeln, mit dem ein Vergleich der Ergebnisqualität zwischen den Einrichtungen anhand eines einzigen Kennwertes möglich ist.

Um die Veränderungen in der gesundheitsbezogenen Lebensqualität der Rehabilitierten beschreiben zu können, wurden für die indirekte Veränderungsmessung als Messzeitpunkte der Rehabilitationsbeginn und ein Katamnesezeitpunkt 90 Tage nach Rehabilitationsbeginn festgelegt. Auf eine Befragung zum Zeitpunkt der Entlassung aus der Rehabilitation wurde nach ausführlicher Diskussion in der AG Ergebnismessung verzichtet um die Belastung der Patient:innen durch die Befragung zu minimieren und da die Annahme besteht, dass die Erfolge durch die Rehabilitation zum gewählten Katamnesezeitpunkt drei Monate nach Behandlungsbeginn aussagekräftiger als die Ergebnisse zum Entlassungszeitpunkt seien.

Rücklauf und Datenqualität

Insgesamt nahmen 17 Einrichtungen mit neurologischen Reha-Fachabteilungen teil. Es wurden Daten von N = 2081 Patient:innen erfasst und für Analysen an das UKE übermittelt, allerdings konnten aufgrund unvollständiger Verlaufsdaten für die Analysen nur Daten von N = 1166 Patient:innen ausgewertet werden. Die sich ergebende Rücklaufquote über alle Einrichtungen von 56% entspricht etwa den Erfahrungen mit Rückläufen in Katamnesestudien anderer Forschungsbereiche. Beim Vergleich der Rücklaufquoten zwischen den Einrichtungen zeigte sich allerdings eine große Variabilität von ca. 20% bis ca. 99% an vollständigen Verlaufsdaten.

Die geringen Rücklaufquoten in einzelnen Einrichtungen können mit Schwierigkeiten zusammenhängen, die Organisation der Katamnesebefragung, für die in der Regel Fragebögen zu den Patient:innen nach Hause geschickt werden mussten, in die täglichen Abläufe der anspruchsvollen klinischen Routine zu implementieren. Bei den vereinzelt sehr hohen Rücklaufquoten von annähernd 100% ist zu vermuten, dass vor der Datenübermittlung möglicherweise bereits eine Vorselektion nach vollständigen Verlaufsdaten stattgefunden hatte. In Bezug auf den risikoadjustierten Einrichtungsvergleich verringert sich die Anzahl analysierbarer Fälle durch fehlende Angaben zu Patient:innenmerkmalen weiter auf $N = 1047$, was einer Ausschöpfungsquote der übermittelten Datensätze von 50,3% entspricht. Auch hier gibt es hinsichtlich der Vollständigkeit der Angaben zu Patient:innenmerkmalen eine beträchtliche Variabilität zwischen den Einrichtungen, mit Einrichtungen, in denen ein größerer Anteil von Angaben für die Risikoadjustierung fehlte, hin zu Einrichtungen mit (nahezu) keinen fehlenden Angaben. Zum Teil lässt sich die Variabilität einzelner fehlender Angaben zwischen den Einrichtungen damit erklären, dass unterschiedliche Erhebungsmethoden (elektronische Dateneingabe und Befragung per Paper-Pencil) zur Anwendung kamen. Allerdings zeigten weitere (hier nicht berichtete) Analysen, dass auch zwischen Einrichtungen, in denen mit Paper-Pencil erhoben wurde, teils deutliche Unterschiede im Anteil einzelner fehlender Angaben vorkommen. Ob diese Unterschiede durch ein spezifisches Patient:innenklientel oder durch einrichtungsspezifische Faktoren bedingt sind, bleibt dabei unklar.

Effektstärken der (unadjustierten) Veränderung der gesundheitsbezogenen Lebensqualität zwischen Aufnahme- und Katamnesezeitpunkt

Die Effektstärken der nicht risikoadjustierten Veränderung der mit dem PROMIS-10 gemessenen Veränderung der gesundheitsbezogenen Lebensqualität zwischen Aufnahme- und Katamnesezeitpunkt fallen in der Gesamtstichprobe über alle Einrichtungen sehr gering aus (Physical Health: $d_z = 0,11$; Mental Health: $d_z = 0,05$). Zwischen den Einrichtungen zeigt sich eine große Varianz der erreichten (unadjustierten) Effektstärken, wobei sogar bei etwa einem Drittel der teilnehmenden Einrichtungen Verschlechterungen der gesundheitsbezogenen Lebensqualität zwischen Rehabilitationsbeginn und Katamnesebefragung verzeichnet wurden. Durch den fehlenden Messzeitpunkt zur Entlassung der Patient:innen aus der Rehabilitation wird hier nicht klar, ob nach der Aufnahme in die Rehabilitation bei den Patient:innen durchschnittlich keine großen Veränderungen bis zum Katamnesezeitpunkt zu verzeichnen waren, oder ob sich potentiell positive, bis zur Entlassung aus der Behandlung eingetretene Veränderungen nach der Entlassung als nicht stabil erwiesen und bis zum Katamnesezeitpunkt wieder etwas oder sogar unter den Wert bei Rehabilitationsbeginn abfielen. Hinweise, dass zum Entlassungszeitpunkt größere Veränderungen gemessen werden können, die sich zum Katamnesezeitpunkt wieder verringern, erhielten wir in einer gemeinsamen Diskussion der Ergebnisse mit dem Expertengremium der AG Ergebnismessung: In fünf Einrichtungen der Medical Park Gruppe, die auch an dieser Pilotstudie teilgenommen hatten, zeigten sich bei $N = 102$ Patient:innen für den gleichen Erhebungszeitraum zum Entlassungszeitpunkt deutlich höhere T-Werte im PROMIS-10 als zum dort durchgeführten Follow-up (siehe Anhang

„Vergleichsdaten für den PROMIS-10 aus fünf Einrichtungen der Medical Park Gruppe mit zusätzlicher Messung zur Entlassung aus der Rehabilitation“, S. 69).

Für eine zusätzliche Befragung mit dem VR-12 optierten fünf teilnehmende Einrichtungen. Effektstärken der Veränderung der gesundheitsbezogenen Lebensqualität waren verglichen mit dem PROMIS-10 in der körperlichen Dimension etwas höher ($d_z = 0,26$) sowie im unadjustierten Einrichtungsvergleich ausschließlich positiv, und in der psychischen Dimension etwa gleich hoch ($d_z = 0,08$), wobei in zwei Einrichtungen die durchschnittlichen Ergebnisse zum Katamnesezeitpunkt unter den Werten zur Befragung zum Beginn der Rehabilitation lagen.

Deckeneffekte

Nicht nur die Wahl eines geeigneten Messzeitpunkts (zur Entlassung bzw. zur Katamnese) kann einen Einfluss auf die Höhe der messbaren Veränderungen des Gesundheitszustands haben, sondern auch Deckeneffekte könnten zum Tragen kommen, wenn die selbstberichtete Gesundheit der Patient:innen zum Beginn der Rehabilitation bereits so gut ist, dass sie sich in der Verlaufsmessung nur noch wenig verbessern kann. Um potentiell vorliegende Deckeneffekte aufzudecken, führten wir dazu zielgerichtete Analysen anhand des PROMIS-10 durch. Aus dem Scoring des PROMIS-10 ergeben sich bevölkerungsnormierte T-Werte, wobei T-Werte von 50 dem Durchschnitt der Konstruktionsstichprobe entsprechen, die idealerweise bevölkerungsrepräsentativ sein sollte. Die von den Schlaganfallpatient:innen (mit einem Altersdurchschnitt von 65 Jahren) in dieser Studie erreichten T-Werte betragen zu Beginn der Rehabilitation durchschnittlich 42,3 (Physical Health) bzw. 43,7 (Mental Health) und liegen damit bereits relativ dicht am (und innerhalb einer Standardabweichung um den) Durchschnitt der Konstruktionsstichprobe des Instruments. Aus einer grafischen Analyse der Differenzen von T-Werten zwischen Katamnese- zu Aufnahmezeitpunkt in Abhängigkeit des Ergebnisses zum Aufnahmezeitpunkt geht für beide Dimensionen des PROMIS-10 hervor, dass ein bedeutender Anteil von Patient:innen in der Gesamtstichprobe zum Rehabilitationsbeginn T-Werte > 45 aufweist, und dass in dieser Gruppe von Patient:innen höhere T-Werte zum Beginn der Rehabilitation mit einer Abnahme der gesundheitsbezogenen Lebensqualität zum Katamnesezeitpunkt einhergehen. Auf Grundlage der Punkteverteilung wurden die Patient:innen für weitere Analysen zu Deckeneffekten in drei Gruppen mit niedriger (T-Werte ≤ 40), mittlerer (T-Werte > 40 und ≤ 45) und hoher (T-Werte > 45) gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt eingeteilt. Insgesamt weisen im PROMIS-10 in der Dimension Physical Health 28,6% der Patient:innen und in der Dimension Mental Health 44,6% der Patient:innen eine hohe gesundheitsbezogene Lebensqualität auf. Im Vergleich zeigt sich, dass die Anteile dieser Gruppen zwischen den einzelnen Einrichtungen sehr variabel verteilt sind. Von vorliegenden Deckeneffekten wären die Einrichtungen also in einem unterschiedlichen Ausmaß betroffen. Vergleicht man die Effektstärken der (unadjustierten) Veränderungen zwischen Aufnahme- und Katamnesezeitpunkt zwischen diesen Gruppen, zeigen sich höhere positive Effektstärken in der Gruppe mit niedriger gesundheitsbezogener Lebensqualität (PH: $d_z = 0,56$; MH: $d_z = 0,67$), kaum Veränderungen in der Gruppe mit mittlerer (PH: $d_z = 0,10$; MH: $d_z = 0,15$) und höhere negative Effektstärken in der Gruppe mit hoher gesundheitsbezogener Lebensqualität (PH: $d_z = -0,47$; MH: $d_z = -0,39$). Durch höhere Anteile

von Patient:innen mit einer hohen gesundheitsbezogenen Lebensqualität in der Gesamtstichprobe bzw. in den einzelnen Einrichtungen können die positiven Veränderungen der Gruppe mit einem schlechteren Gesundheitszustand zum Beginn der Rehabilitation auf Gesamt- und Einrichtungsebene verschleiert werden. Sehr ähnliche Ergebnisse ergaben sich bei analogen Analysen für die fünf teilnehmenden Einrichtungen mit dem VR-12, wobei hier im Physical Component Score die negative Effektstärke in der Gruppe mit hoher gesundheitsbezogener Lebensqualität niedriger ausfiel ($d_z = -0,28$), so dass sich im Mittel in der Gesamtstichprobe höhere positive Veränderungen in diesem Instrument ergeben. Eine mögliche Erklärung für die negativen Effektstärken in der Gruppe von Patient:innen mit hoher gesundheitsbezogener Lebensqualität in beiden Dimensionen von PROMIS-10 und VR-12 kann das Vorhandensein einer Regression zur Mitte sein: Die Regression zur Mitte ist ein statistisches Phänomen, bei dem extreme Werte bei einer zweiten Messung tendenziell näher am Mittelwert der Population liegen, wenn diese Messung mit zufälligen Schwankungen behaftet ist. Die durchschnittliche Verschlechterung des berichteten Gesundheitszustands in dieser Gruppe kann also auch darin begründet sein, dass sich die Patient:innen zum Katamnesezeitpunkt wieder einem „Normalwert“ annähern.

In einer von den Kooperationspartnern durchgeführten Diskussionsrunde mit Klinikern aus den teilnehmenden Einrichtungen wurden die Befunde zu den Deckeneffekten vorgestellt und die Diskussionsteilnehmer danach befragt, welche Gründe sie dafür sehen, dass ein größerer Anteil von Patient:innen zum Rehabilitationsbeginn eine hohe gesundheitsbezogene Lebensqualität berichtet. Verschiedene Faktoren könnten demnach mit Deckeneffekten assoziiert sein:

So kämen Patient:innen bei stärkeren Beeinträchtigungen nach dem Akutereignis des Schlaganfalls häufiger mit einem längeren und schwierigeren Genesungsweg durch die Rehabilitationsphasen schließlich in der Phase D der Rehabilitation an. Bei der Betrachtung ihres Leidensweges durch die Rehabilitationsphasen könnten sie sich zum Beginn der Rehabilitationsphase D möglicherweise schon als „fast geheilt“ einschätzen um dann zum Zeitpunkt der Katamnesebefragung mit alltäglichen Problemen konfrontiert ihre Einschätzung realistischer zu treffen.

Das Befragungssetting kann sich ebenfalls auf das Wohlbefinden auswirken, wenn Patient:innen erst nach mehreren Tagen nach der Anreise die Befragung durchführen, wenn sie bereits etwas ausgeruhter sind, oder durch Incentives wie Kaffee-Gutscheine zur Befragungsteilnahme motiviert werden.

Auch eine Zuweisungselektion könne im Hinblick auf die Kostenträger DRV/GKV eine Rolle spielen: So könnten für Patient:innen, die von der DRV in die Rehabilitation zugewiesen werden, bei leichteren Beeinträchtigungen häufiger die sozialmedizinische Leistungsbeurteilung und andere Zielstellungen in Bezug auf Gesundheit und Rückkehr ins Berufsleben im Fokus stehen, als bei von der GKV zugewiesenen Patientinnen. Patient:innen der GKV werden dagegen häufiger mit einem größeren subjektiven Leidensdruck verbunden

mit einer hohen Motivation zur Rehabilitationsbehandlung wahrgenommen. Somit könnte das Verhältnis von GKV- zu DRV-Patient:innen auch Unterschiede in den Anteilen von Patient:innen mit Angaben zu einer höheren gesundheitsbezogenen Lebensqualität zum Aufnahmezeitpunkt zwischen den Einrichtungen bedingen.

Für die Höhe des Anteils von Patient:innen mit hoher gesundheitsbezogener Lebensqualität zu Rehabilitationsbeginn könnten auch Selektionseffekte eine Rolle spielen, wenn Patient:innen mit stärkeren Beeinträchtigungen aus Belastungsgründen von sich aus nicht an der Befragung teilnehmen oder zum Katamnesezeitpunkt keinen Fragebogen ausfüllen, oder wenn die Entscheidung zum Ausschluss aus der Befragung durch die Einrichtung erfolgte, um Patient:innen zu schonen. Im letzteren Falle könnten systematische Unterschiede in den Bewertungskriterien in den Einrichtungen, welche Patient:innen von der Befragung ausgenommen werden sollten, auch zu entsprechenden Unterschieden im Ausmaß der Deckeneffekte zwischen den Einrichtungen führen.

Auch kostenträgerunabhängig werden von den Klinikern deutliche Unterschiede zwischen den Patient:innen in der Belastung zum Rehabilitationsbeginn wahrgenommen: Reha-Motivation und Ziele können sich nach individuellen Hintergründen (beispielsweise Auswirkungen des Schlaganfalls auf Beruf oder Alltag) sehr unterscheiden. So kann sich eine identische physische Ausgangslage auf den Umgang mit Symptomen, das psychische Befinden und den Leidensdruck von Patient:innen sehr unterschiedlich auswirken. Bei sehr spezifischen Einschränkungen, die zu einem individuell hohen Leidensdruck führen können, ist es daher nicht ausgeschlossen, dass die Auswirkungen auf den Alltag oder Beruf, wie sie mit dem generisch angelegten PROMIS-10 erfragt werden, eher gering sind und sich entsprechend nicht im Ergebnis des Instruments als allgemeine Einschränkung in der gesundheitsbezogenen Lebensqualität abbilden lassen.

Zudem sei eine Rehabilitationsmaßnahme auch für Patient:innen mit insgesamt geringeren Einschränkungen sinnvoll, wenn sich durch die Maßnahme für bestimmte Patient:innengruppen wie z. B. alleinerziehende Mütter oder alleinstehende Senioren Verbesserungen des Gesundheitszustands erzielen lassen, die im häuslichen Umfeld auf sich allein gestellt so nicht möglich wären.

Schließlich wird von den Klinikern für eine Veränderungsmessung des Gesundheitszustands eine Messung zur Entlassung aus der Rehabilitation als aussagekräftiger als die Messung zum Katamnesezeitpunkt bewertet.

Stratifizierte Risikoadjustierung nach einrichtungsspezifischen Patient:innenmerkmalen für einen fairen Vergleich der Ergebnisqualität zwischen den Rehabilitationseinrichtungen

Das Verfahren der Risikoadjustierung ermöglicht faire Vergleiche in den Outcomes zur Ergebnisqualität zwischen Rehabilitationseinrichtungen, indem Patient:innenmerkmale, die mit den Outcomes im Zusammenhang stehen können, bei der Analyse der Unterschiede

zwischen den Einrichtungen berücksichtigt werden. So können beispielsweise auch bei vorliegenden Deckeneffekten, die sich heterogen über die zu vergleichenden Einrichtungen verteilen, Aussagen dazu statistisch abgesichert werden, ob einzelne Reha-Fachabteilungen, in denen Patient:innen mit durchschnittlich unterschiedlich hoher gesundheitsbezogener Lebensqualität zum Rehabilitationsbeginn behandelt werden, jeweils über- oder unterdurchschnittlich in der erreichten Ergebnisqualität abschneiden.

Wie in den zuvor durchgeführten Pilotstudien in den Reha-Fachabteilungen Orthopädie, Psychosomatik und Onkologie sollte dazu die Risikoadjustierung zunächst mittels eines an der Gesamtstichprobe abgeleiteten linearen Regressionsmodells durchgeführt werden. Zwischenergebnisse der hier vorliegenden Pilotstudie brachten allerdings hervor, dass die Annahme eines linearen Zusammenhangs der Ergebnisse der PROMs zum Aufnahmezeitpunkt als Prädiktoren mit dem Outcome der PROMs zum Katamnesezeitpunkt als abhängiger Variable bedingt durch die zuvor beschriebenen Deckeneffekte zumindest für die physische Dimension des PROMIS-10 nicht als gegeben angesehen werden kann. In der Folge könnten aufgrund der Deckeneffekte in Rehabilitationseinrichtungen mit einem höheren Anteil von Patient:innen mit hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt die zu erwartenden Outcomes überschätzt werden, wenn anhand der Gesamtstichprobe mittels linearer Regression adjustiert wird, was entsprechend für diese Einrichtungen zu einem Nachteil im Vergleich der Ergebnisqualität mit anderen Einrichtungen führen kann. Als pragmatischer Lösungsansatz wurde deshalb die Durchführung einer nach gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt stratifizierten Risikoadjustierung gewählt, bei der für jedes Stratum ein eigenes Regressionsmodell erstellt wurde.

So erfolgte die Einteilung der Strata nach niedriger, mittlerer und hoher zum Aufnahmezeitpunkt gemessener gesundheitsbezogener Lebensqualität. Für die Risikoadjustierung wurde dann jeweils für jede Dimension jedes eingesetzten Instruments und für jede der stratifizierten Gruppen ein **eigenes Regressionsmodell** erstellt, aus dem die mit Patient:innenmerkmalen assoziierten Erwartungswerte für das durchschnittlich erreichbare Outcome berechnet wurden. Ein bedeutender Nachteil dieser Methode ist, dass sich die Fallzahlen durch die Stratifizierung für die Analysen innerhalb der Strata entsprechend verringern, so dass durch eine geringere statistische Power die statistische Absicherung der jeweiligen Regressionsmodelle erschwert ist und auch Artefakte bezüglich des schrittweisen Ausschlusses relevanter Prädiktorvariablen mit größerer Wahrscheinlichkeit auftreten können. Die Berechnung der Abweichung der erreichten Outcomes vom risikoadjustierten erwarteten Outcome kann dann aber für jede Einrichtung über alle Strata gleichzeitig erfolgen. In einem deskriptiven Methodenvergleich der nicht-stratifizierten Risikoadjustierung vs. stratifizierten Risikoadjustierung mit Fokus auf Veränderungen der Rangfolgen der Einrichtungen zeigte sich, dass durch die Stratifizierung Einrichtungen mit einem großen Anteil von Patient:innen mit hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt insgesamt etwas besser abschnitten. Auch wenn das Ergebnis der stratifizierten Risikoadjustierung augenscheinlich valide ist, sollte ihre Durchführung aufgrund der geringen Fallzahlen in den Strata und eines

zum jetzigen Zeitpunkt noch fehlenden statistisch abgesicherten Vergleichs der beiden Methoden hier eher als vorläufiges Proof-Of-Concept gewürdigt werden.

Ergebnisqualität im risikoadjustierten Einrichtungvergleich

Zur Bewertung der erreichten Ergebnisqualität im Vergleich der Einrichtungen wurden Effektstärken der Abweichungen der erreichten Outcomes von den regressionsanalytisch bei der Risikoadjustierung berechneten durchschnittlichen Erwartungswerte analysiert. Für den PROMIS-10 variieren die Effektstärken der Abweichungen des erreichten vom erwarteten Outcome in der **Dimension Physical Health** von $d_z = -0,24$ bis $d_z = 0,35$, wobei eine statistische Absicherung einer (hier: unterdurchschnittlichen) Abweichung vom erwarteten Outcome nur für eine Einrichtung vorliegt. In der **Dimension Mental Health** des PROMIS-10 liegen die Effektstärken der Abweichungen zwischen $d_z = -0,19$ bis $d_z = 0,52$, und die (hier: überdurchschnittlichen) Abweichungen vom zu erwartenden Outcome zweier Einrichtungen können statistisch abgesichert werden. Für den VR-12 ergibt sich bezüglich der Höhe der Effektstärken der Abweichung des erreichten vom erwarteten Outcome in den fünf optional teilnehmenden Einrichtungen ein ähnliches Bild im **Physical Component Score** ($d_z = -0,19$ bis $d_z = 0,16$) bzw. im **Mental Component Score** ($d_z = -0,14$ bis $d_z = 0,10$), ohne dass eine Abweichung einer Einrichtung statistisch abgesichert werden kann. Die meisten der für die Einrichtungen berichteten Abweichungen vom erwarteten Outcome bewegen sich in einem Effektstärkenkorridor von $d_z = \pm 0,25$ – **sie variieren also nur wenig um die durchschnittlich erreichte Ergebnisqualität.**

Die nominell geringen Ausschläge nach oben oder unten und die geringe Varianz der Abweichungen zwischen den Einrichtungen sollten aber nicht in der Form interpretiert werden, dass die generischen Instrumente PROMIS-10 und VR-12 oder die berechneten Effektstärken nicht trennscharf genug sind, um zwischen über- und unterdurchschnittlich abschneidenden Einrichtungen zu diskriminieren. **Tatsächlich deuten diese Ergebnisse darauf hin, dass unter Berücksichtigung der spezifischen Merkmale von Patient:innen keine allzu großen Unterschiede im Behandlungsergebnis zwischen den teilnehmenden Rehabilitationseinrichtungen bestehen** – und das ist, eine generell hohe Behandlungsqualität vorausgesetzt, eine gute Nachricht! Denn auf dieser Grundlage können Patient:innen oder zuweisende Institutionen für Schlaganfallpatient:innen nach sekundären, der Behandlungsqualität nachgelagerten individuellen Anforderungen (z. B. Wohnortnähe, Ausstattung, etc.) aus den teilnehmenden Reha-Fachabteilungen eine geeignete Rehabilitationseinrichtung aussuchen ohne damit Abstriche an die zu erwartende Ergebnisqualität der Behandlung machen zu müssen. Etwas problematischer ist hinsichtlich dieser Ergebnisse allerdings, dass aufgrund der geringen Fallzahlen innerhalb der Einrichtungen und der damit verbundenen geringen statistischen Power eine statistische Absicherung eines über- oder unterdurchschnittlichen Ergebnisses nur bei relativ hohen Beträgen von Effektstärken der Abweichungen des erreichten vom erwarteten Behandlungsergebnis möglich sind. Für eine inferenzstatistische Absicherung einer Effektstärke von $d_z = \pm 0,25$ wären mindestens $n = 64$ Fälle pro Einrichtung nötig, was für kleinere Reha-Fachabteilungen mit

niedrigen Aufnahmezahlen eine größere Herausforderung darstellen kann und in der Regel dann relativ lange Erhebungsintervalle zur Erreichung der benötigten Fallzahl bedarf.

Patient-Reported Outcome Quality Index (ProQI)

Auf der Basis der in der Risikoadjustierung generierten standardisierten Residuen wurde der ProQI berechnet, der die in einer Einrichtung risikoadjustierten Outcomes in einem einzelnen Kennwert zusammenfasst und der direkt mit dem ProQI anderer Reha-Fachabteilungen in der Indikation Schlaganfall verglichen werden kann. Da nur fünf Einrichtungen den VR-12 einsetzen, kann ein instrumentenübergreifender Gesamt-ProQI nur für diese Einrichtungen berechnet werden. So bietet sich daher ein Vergleich des auf den PROMIS-10 beschränkten ProQIs an. Der ProQI des PROMIS-10 hat per konzeptueller Definition über alle teilnehmenden Einrichtungen einen Mittelwert von 90 und variiert um ihn in einem Range von ca. 84 bis 100 Punkten, wobei zehn der 17 Reha-Fachabteilungen einen Wert > 90 Punkte erreichen. Für die Interpretation des ProQIs gilt analog wie für die Interpretation der Effektstärken der Abweichungen der risikoadjustierten Erwartungswerte vom erreichten Outcome, dass **eine geringe Varianz zwischen den Einrichtungen bei geringen Abweichungen vom Mittelwert ein Zeichen für eine in gleicherweise hohe Ergebnisqualität sein kann**. Für einen einfacheren, deskriptiven Vergleich wird hier auf das Einbeziehen der statistischen Absicherung eines über- oder unterdurchschnittlichen Ergebnisses für die Bildung der Rangfolge der teilnehmenden Einrichtungen verzichtet.

Vergleich der Änderungssensitivität von PROMIS-10 und VR-12

Abschließend wurde in dieser Pilotstudie die Änderungssensitivität von PROMIS-10 und VR-12 vergleichend untersucht um Erkenntnisse zu gewinnen, ob bei inhaltlich vergleichbaren Konstrukten möglicherweise psychometrische Unterschiede bestehen, die den Vorzug eines der Instrumente in Folgestudien zur Rehabilitation des Schlaganfalls rechtfertigen würden. Es zeigte sich dabei, dass in der körperlichen Dimension in der Gesamtstichprobe der VR-12 eine deutlich höhere Änderungssensitivität als der PROMIS-10 aufweist. Bei einer Stratifizierung der Patient:innen nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt im PROMIS-10 wird deutlich, dass dies vor allem auf günstigere Messeigenschaften des VR-12 in Patient:innengruppen mit mittlerer und hoher gesundheitsbezogener Lebensqualität zurückzuführen ist. Der VR-12 scheint damit in der körperlichen Dimension weniger stark auf Deckeneffekte zu reagieren, die beim PROMIS-10 vorhandene positive Veränderungen der gesundheitsbezogenen Lebensqualität in der Verlaufsuntersuchung der Gesamtstichprobe verschleiern. Für die psychische Dimension von PROMIS-10 und VR-12 sind tendenziell ähnliche Unterschiede in der Änderungssensitivität beobachtbar, die aber so gering ausfallen, dass sie praktisch nicht ins Gewicht fallen. **Insgesamt kann aufgrund der Analysen aber empfohlen werden, in Folgestudien eher den VR-12 als den PROMIS-10 vorzuziehen, insbesondere, wenn mit Deckeneffekten gerechnet werden muss.**

Zusammenfassende Empfehlungen für Folgestudien und Ausblick

Für die Messung der Veränderungen der gesundheitsbezogenen Lebensqualität der Rehabilitierten wurden in dieser Pilotstudie ein Messzeitpunkt zum Rehabilitationsbeginn und ein Katamnesezeitpunkt 90 Tage nach Rehabilitationsbeginn festgelegt. Aus den Ergebnissen der Studie wird deutlich, dass ein **Messzeitpunkt zum Ende der Rehabilitation** möglicherweise geeigneter ist, um aussagekräftige Befunde zur Beschreibung der Veränderung des Gesundheitszustands von Schlaganfallpatient:innen im Rahmen der Rehabilitation der Phase D zu erhalten.

Einerseits könnte durch eine Messung zur Entlassung am Ende der Rehabilitation die **Ausschöpfungsquote vollständiger Verlaufsdaten** für die Datenanalysen erhöht werden, da durch die noch in der Einrichtung stattfindende Befragung Patient:innen unmittelbar und persönlich zur Teilnahme angesprochen werden können, was im Gegensatz zu einer postalisch stattfindenden katamnestischen Befragung im häuslichen Umfeld die Wahrscheinlichkeit für Drop-Outs reduzieren dürfte. Unvollständige Verlaufsdaten durch Drop-Outs zum Katamnesezeitpunkt müssten durch eine größere Zahl von Patient:innen, die zum Aufnahmezeitpunkt in die Befragung eingeschlossen werden, kompensiert werden, um auf die für die Analysen benötigte Fallzahl zu kommen. Wenn eine höhere Ausschöpfungsquote durch die Festlegung des zweiten Messzeitpunkts auf einen Zeitpunkt kurz vor der Entlassung erreicht werden kann, könnte der Erhebungsaufwand in den Rehabilitationseinrichtungen dadurch wahrscheinlich insgesamt reduziert werden.

Andererseits kann für die Fragestellung, wie groß die **Unterschiede in der gesundheitsbezogenen Lebensqualität von Schlaganfallpatient:innen vor und nach der Rehabilitation** sind, ein Messzeitpunkt am Ende der Rehabilitation ebenfalls von Bedeutung sein, wenn sich dadurch größere Effektstärken der Veränderungen zwischen den Messzeitpunkten messen lassen. Größere Effektstärken in Analysen von Veränderungen im berichteten Gesundheitszustand wären mit einer größeren statistischen Power verbunden, was in der Planung einer Studie, die lediglich auf die Analyse beobachtbarer Veränderungen zielt, zu einer Reduktion der benötigten Fallzahl beitragen kann.

Wenn sich die primäre Fragestellung allerdings darauf bezieht, in welcher Höhe Rehabilitationseinrichtungen ein über- oder unterdurchschnittliches Ergebnis im Vergleich mit anderen Einrichtungen erzielen, spielen auch andere Faktoren zur Ermittlung der benötigten Fallzahl für die Verlaufsuntersuchung eine Rolle: Die für die risikoadjustierten Einrichtungsvergleiche berechneten Effektstärken sind hier keine Veränderungsmaße, sondern Maße der Abweichung des erreichten Outcomes vom erwarteten Outcome. Im Idealfall, bei einer rehabilitativen Versorgung auf insgesamt hohem Niveau, die von allen Rehabilitationseinrichtungen mit gleicher Ergebnisqualität geleistet wird, kann damit gerechnet werden, dass die Effektstärken der Abweichung des erreichten Outcomes einer Einrichtung vom risikoadjustierten Erwartungswert des Outcomes über alle Einrichtungen hinweg eher sehr klein ausfallen, so dass es hier einer relativ großen Fallzahl pro Einrichtung

bedürfe, um für eine einzelne Einrichtung eine statistisch abgesicherte Aussage treffen zu können, ob die erreichte Ergebnisqualität im Vergleich anderer Einrichtungen über- oder unterdurchschnittlich ist. **Daher sollten bei der Planung von Folgestudien Überlegungen einbezogen werden, welche Höhe von Effektstärken der Abweichung des erwarteten vom erreichten Outcome als relevant betrachtet werden können**, um davon ausgehend die benötigte Fallzahl pro Einrichtung zu ermitteln, mit der ein der Effektstärke entsprechend über- oder unterdurchschnittliches Ergebnis einer Rehabilitationseinrichtung statistisch abgesichert werden kann.

Auf den Erfahrungswerten dieser und der vergangenen Pilotstudien beruhend wird daher empfohlen, grundsätzlich pro Einrichtung (bzw. pro behandelter Indikation und den dort eingesetzten PROMs) eine Fallzahl von $n > 50$ kumulativ eingeschlossener und auswertbarer Fälle anzustreben, auch um für jede Einrichtung ein möglichst repräsentatives Ergebnis ermitteln zu können. Einrichtungsvergleiche mit geringeren Fallzahlen pro Einrichtung können zwar noch aufschlussreiche Analysen erbringen, auch wenn die statistische Absicherung über- und unterdurchschnittlicher Outcomes weniger wahrscheinlich wird. Von risikoadjustierten Vergleichen von Einrichtungen mit einem $n < 25$ Fällen pro Einrichtung wird aufgrund mangelnder statistischer Power und fehlender Repräsentativität aber eher abgeraten.

Fehlende Angaben in für die Risikoadjustierung relevanten Merkmalen, die vor allem in der Paper-Pencil-gestützten Datenerhebung auftreten, können die Fallzahlen für den risikoadjustierten Einrichtungsvergleich deutlich verringern. Eine mögliche statistisch geleitete Imputation dieser fehlenden Angaben sollte kritisch hinterfragt werden, da nicht ausgeschlossen werden kann, dass Ausfälle in den Daten systematisch in einzelnen Einrichtungen auftreten oder zumindest für einzelne Einrichtungen spezifisch sind. In diesem Fall wäre angezeigt, mit geeigneten Verfahren die Imputation fehlender Daten auf Einrichtungsebene durchzuführen, was aber voraussetzen würde, dass für eine möglichst verzerrungsfreie Schätzung der imputierten Angaben eine ausreichende Fallzahl in der jeweiligen Einrichtung vorläge. Nimmt man an, dass – Faustregeln folgend – dazu mindestens 100 Fälle oder deutlich mehr pro Einrichtung notwendig wären, wird klar, dass sich der Erhebungsaufwand durch längere Erhebungszeiträume, um die nötige Fallzahl dafür zu erreichen, deutlich erhöhen würde. **Es ist daher sehr zu empfehlen, dass möglichst schon zum Erhebungszeitpunkt fehlende Angaben vermieden werden.** Für eine elektronische Erhebung ist dies durch technische Maßnahmen leichter zu erreichen (z. B. durch angezeigte erklärende Hinweise beim Auslassen einzelner Angaben am Bildschirm) als bei der Paper-Pencil-gestützten Datenerhebung. Müssen Patient:innen aus organisatorischen oder technischen Gründen mit gedruckten Fragebögen befragt werden, sollte entsprechend instruiertes Personal den Patient:innen Hilfestellungen geben können, falls die über den Fragen gedruckten Informationen bzw. Instruktionen zu Inhalt und Art der erfragten Information nicht ausreichen, um fehlende Angaben zu vermeiden.

Insbesondere sollten die **Patient:innen darauf hingewiesen werden, dass die Vollständigkeit der Daten notwendig ist**, damit ihre Angaben zur Bewertung der Ergebnisqualität ihrer Behandlung und somit zur langfristigen Sicherung und Weiterentwicklung der Qualität in der medizinischen Rehabilitation beitragen können.

Um Probleme mit der **Datenqualität** (z. B. durch fehlende Angaben) frühzeitig erkennen zu können, wird empfohlen, mit dem Beginn der Datenerhebung ein entsprechendes **Monitoring** zu implementieren. Anhand einer geringen Anzahl von Datensätzen, die aus jeder Einrichtung an die Auswertungsstelle übermittelt werden, können Erfassungsprobleme aufgedeckt werden, so dass auf der Basis der Rückmeldungen der Auswertungsstelle rechtzeitig Maßnahmen ergriffen werden können, um eine ausreichende Anzahl analysierbarer Datensätze für jede Einrichtung zum Ende des geplanten Erhebungszeitraums sicherzustellen.

Aus den bisherigen Ausführungen wird deutlich, welche Faktoren einen Einfluss auf die Höhe der am Ende eines festgelegten Erhebungszeitraums erreichten Fallzahl haben können, und dass die Fallzahl eine zentrale Größe darstellt, um im Einrichtungsvergleich ein über- oder unterdurchschnittliches Ergebnis statistisch absichern zu können. **Die Länge des Erhebungszeitraums sollte daher sorgfältig geplant und mit einem Puffer versehen werden**, damit auch Einrichtungen mit geringeren Belegungszahlen trotz nicht vermeidbarer Drop-Outs und Ausfällen in der Risikoadjustierung aufgrund fehlender Angaben eine ausreichende Fallzahl erreichen können, um das erzielte Ergebnis statistisch absichern zu können.

Werden Rehabilitationseinrichtungen hinsichtlich der erreichten Ergebnisqualität anhand von Effektstärken verglichen, können deutlich über- oder unterdurchschnittliche Ergebnisse zunächst nur aufzeigen, dass in den betreffenden Einrichtungen „etwas anders läuft“ als in anderen Einrichtungen. Um die hinter den Ergebnissen liegenden relevanten Faktoren zu identifizieren, die im Zusammenhang mit dem Erreichen einer überdurchschnittlichen Ergebnisqualität stehen, wäre möglicherweise ein **Austausch zwischen den Einrichtungen im Rahmen eines offenen Forums** sinnvoll. Besonders gut abschneidende Reha-Fachabteilungen könnten so für andere Einrichtungen wertvolle Impulse liefern, die eigene Ergebnisqualität weiter zu optimieren.

Wenn **Deckeneffekte** aufgrund einer relativ hohen berichteten gesundheitsbezogenen Lebensqualität zum Rehabilitationsbeginn bei einem Teil der Patient:innen erwartet werden können, wie sie sich hier in der Pilotstudie zur Ergebnisqualität in der Rehabilitation des Schlaganfalls zeigten, sollten bereits in der Konzeption einer Verlaufsstudie potentielle Zuweisungs- oder Selektionseffekte antizipiert und damit zusammenhängende Merkmale (wie z. B. Kostenträger der Maßnahme) möglichst vollständig identifiziert und dokumentiert werden, damit Einschlusskriterien für Patient:innen entsprechend angepasst werden können, um Deckeneffekte zu vermeiden, oder damit die mit den Deckeneffekten assoziierten Faktoren im Rahmen der Auswertung der Studie entsprechend (statistisch) kontrolliert werden können.

In der hier vorgestellten Pilotstudie nahm nur eine begrenzte Zahl an Einrichtungen teil. Zukünftige Untersuchungen zur einrichtungsvergleichenden Ergebnisqualität in der Rehabilitation des Schlaganfalls könnten darauf abzielen, mit einer umfassenderen Anzahl teilnehmender Reha-Fachabteilungen weitere Erkenntnisse zu den Veränderungen im Gesundheitszustand der Patient:innen im Rahmen der Rehabilitationsbehandlung zu gewinnen, einen tieferen Einblick in die Ausprägung und Variabilität von Deckeneffekten bei der Messung der Ergebnisqualität zu erhalten sowie weitere Befunde dazu zu generieren, in welcher Anzahl und mit welchen Ausprägungen über- bzw. unterdurchschnittliche Behandlungsergebnisse von den Einrichtungen erreicht werden. Für diese Folgeerhebungen sollte geprüft werden, ob aufgrund der günstigeren Messeigenschaften zur Erfassung der gesundheitsbezogenen Lebensqualität statt des PROMIS-10 der VR-12 eingesetzt werden kann, auch wenn in einigen Rehabilitationseinrichtungen der PROMIS-10 bereits ein Teil der Routinediagnostik ist. Zudem sollte erwogen werden, für die Verlaufsmessung nach der Messung zum Beginn der Behandlung die zweite Messung zum Entlassungszeitpunkt durchzuführen und eine katamnestiche Befragung ggf. fakultativ zusätzlich zu planen. Bei der Festlegung der notwendigen Länge des Erhebungszeitraums können die Resultate zu den in dieser Pilotstudie erreichten Fallzahlen in die Diskussion einbezogen werden.

Schließlich können die hier gewonnenen Erkenntnisse und Erfahrungen dazu genutzt werden, die Forschungsfragestellungen auf weitere Indikationen in der neurologischen Rehabilitation (wie z. B. Multiple Sklerose) zu erweitern und in diesen Bereichen ebenfalls Pilotstudien zur Messung der Ergebnisqualität mit PROMs durchzuführen.

Anhang

Im Folgenden werden die Ergebnisse der Regressionsanalysen zur stratifizierten Risikoadjustierung tabellarisch dargestellt. Für jede Dimension (körperlich/psychisch) eines PROMs wurde jeweils für jede Patient:innengruppe mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität in der jeweiligen Dimension ein eigenes Regressionsmodell erstellt.

In den Tabellen werden als wesentliche Ergebnisparameter der Regressionsanalysen die β -Gewichte der in die Regressionsanalysen aufgenommenen Patient:innenmerkmale sowie die durch das Regressionsmodell aufgeklärte Varianz am durchschnittlichen Outcome der jeweiligen analysierten Gruppe zum Katamnesezeitpunkt ausgewiesen.

Risikoadjustierungsmodelle für PROMIS-10 PH

PROMIS-10 Physical Health (PH) – Modelle der stratifizierten Risikoadjustierung		Alle Kliniken
Fälle gesamt zur Vorhersage des zu erwarteten Outcomes (N)		1032
Niedrige HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 PH \leq 40) (n)		438
Merkmal/Prädiktor	β -Gewicht	M (SD) / %
PROMIS-10 PH Aufnahme (M/SD) ^b	0,400*	35,3 (4,3)
Alter ^b	-0,022 ^a	66,2 (11,8)
weiblich ^b	-0,065 ^a	48,2%
vor Reha weniger als 6 Monate arbeitsunfähig	0,178*	84,0%
Phase C: ja	-0,106*	0,9%
Varianzaufklärung (korrigiertes R²)		0,219

Tab.12: Parameter der Regressionsmodelle zur Risikoadjustierung stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Rehabilitationsbeginn für die Dimension PROMIS-10 Physical Health; *: $p < 0.05$; †: $p < 0.10$; a: $p = n. s.$; b: Prädiktor wurde fix in allen Modellen für diese Dimension eingeschlossen; c: Prädiktor resultierend auf der Basis der Modellierungsstichprobe mit vollständigen Daten, aber kein statistisch signifikanter Beitrag in den erweiterten finalen Analysedaten; β -Gewicht: Ändert sich der Prädiktor um eine Standardabweichung, verändert sich das erwartete Outcome um den Betrag von β Standardabweichungen

Fortsetzung auf der nächsten Seite.

Fortsetzung Tabelle 12:

PROMIS-10 Physical Health (PH) – Modelle der stratifizierten Risikoadjustierung		Alle Kliniken
Mittlere HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 PH: > 40 bis ≤ 45) (n)		301
Merkmal/Prädiktor	β-Gewicht	M (SD) / %
PROMIS-10 PH Aufnahme (M/SD) ^b	0,063 ^a	43,6 (1,3)
Alter ^b	-0,162*	65,0 (11,2)
weiblich ^b	-0,012 ^a	42,2%
Hilfe beim Ausfüllen eines Fragebogens	-0,181*	18,6%
Krank/arbeitsunfähig vor Reha: ja	-0,254*	37,2%
Reha als ganztägig ambulante Maßnahme: ja	-0,101 [‡]	4,3%
Anzahl dokumentierte psychische F-Diagnosen	-0,060 ^c	0,1 (0,3)
mittlerer Schulabschluss	-0,069 ^c	35,9%
Varianzaufklärung (korrigiertes R²)		0,082
Hohe HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 PH > 45) (n)		293
Merkmal/Prädiktor	β-Gewicht	M (SD) / %
PROMIS-10 PH Aufnahme (M/SD) ^b	0,131*	51,6 (4,1)
Alter ^b	-0,111*	61,8 (12,3)
weiblich ^b	-0,122*	30,7%
PROMIS-10 Mental Health (MH) Aufnahme	-0,181*	50,4 (6,8)
Hilfe beim Ausfüllen eines Fragebogens	-0,129*	16,0%
Muttersprache nicht Deutsch	-0,106 [‡]	3,4%
Berufsbildung: Fachhochschule/Universität	0,117*	24,2%
EU-Rente/Berufsunfähigkeitsrente (Dauer u. Zeit)	-0,153*	2,0%
EMR-Antrag gestellt/geplant	0,086 ^c	9,9%
Varianzaufklärung (korrigiertes R²)		0,149

Tab.12 (Fortsetzung): Parameter der Regressionsmodelle zur Risikoadjustierung stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Rehabilitationsbeginn für die Dimension PROMIS-10 Physical Health; *: $p < 0.05$; ‡: $p < 0.10$; a: $p = n. s.$; b: Prädiktor wurde fix in allen Modellen für diese Dimension eingeschlossen; c: Prädiktor resultierend auf der Basis der Modellierungsstichprobe mit vollständigen Daten, aber kein statistisch signifikanter Beitrag in den erweiterten finalen Analysedaten; β-Gewicht: Ändert sich der Prädiktor um eine Standardabweichung, verändert sich das erwartete Outcome um den Betrag von β Standardabweichungen

Risikoadjustierungsmodelle für PROMIS-10 MH

PROMIS-10 Mental Health (MH) – Modelle der stratifizierten Risikoadjustierung		Alle Kliniken	
Fälle gesamt zur Vorhersage des zu erwarteten Outcomes (N)		1080	
Niedrige HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 MH \leq 40) (n)		327	
Merkmal/Prädiktor	β -Gewicht	M (SD) / %	
PROMIS-10 MH Aufnahme (M/SD) ^b	0,254*	34,8 (3,9)	
Alter ^b	-0,033^a	65,6 (11,9)	
weiblich ^b	-0,009 ^a	50,2%	
PROMIS-10 Physical Health Aufnahme	0,144*	36,9 (6,5)	
Schulbildung: kein oder niedriger Schulabschluss	0,081 ^c	50,2%	
Berufsbildung: Ausbildung/ohne Abschluss/sonstiges	0,132*	28,7%	
Berufsbildung: Meister	0,094 [‡]	6,7%	
Berufsbildung: Fachhochschule/Universität	0,219*	15,0%	
Erwerbstätigkeit: ohne berufliche Beschäftigung/unklar	0,082 ^c	19,3%	
Frührente/Altersrente/Pension	0,086 ^c	37,3%	
Varianzaufklärung (korrigiertes R²)		0,127	
Mittlere HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 MH: > 40 bis \leq 45) (n)		264	
Merkmal/Prädiktor	β -Gewicht	M (SD) / %	
PROMIS-10 MH Aufnahme (M/SD) ^b	0,159*	42,6 (1,2)	
Alter ^b	-0,028 ^a	66,6 (10,6)	
weiblich ^b	-0,087 ^a	39,0%	
Weitere relevante Diagnosen: ja	-0,069 ^c	24,6%	
Wohnort neue Bundesländer	0,171*	32,6%	
ledig	0,109 [‡]	6,1%	
mittlerer Schulabschluss	-0,168*	36,0%	
Varianzaufklärung (korrigiertes R²)		0,065	

Tab.13: Parameter der Regressionsmodelle zur Risikoadjustierung stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Rehabilitationsbeginn für die Dimension PROMIS-10 Mental Health; *: $p < 0.05$; ‡: $p < 0.10$; a: $p = n. s.$; b: Prädiktor wurde fix in allen Modellen für diese Dimension eingeschlossen; c: Prädiktor resultierend auf der Basis der Modellierungsstichprobe mit vollständigen Daten, aber kein statistisch signifikanter Beitrag in den erweiterten finalen Analysedaten; β -Gewicht: Ändert sich der Prädiktor um eine Standardabweichung, verändert sich das erwartete Outcome um den Betrag von β Standardabweichungen

Fortsetzung auf der nächsten Seite.

Fortsetzung Tabelle 13.

PROMIS-10 Mental Health (MH) – Modelle der stratifizierten Risikoadjustierung		Alle Kliniken	
Hohe HRQoL zum Aufnahmezeitpunkt (d. h. PROMIS-10 MH > 45) (n)		489	
Merkmal/Prädiktor	β -Gewicht	M (SD) / %	
PROMIS-10 MH Aufnahme (M/SD) ^b	0,350*	50,4 (5,0)	
Alter ^b	-0,069 ^a	63,6 (12,5)	
weiblich ^b	-0,101*	34,4%	
Hilfe beim Ausfüllen eines Fragebogens	-0,157*	19,0%	
Wohnort neue Bundesländer	-0,081 [‡]	29,2%	
Erwerbstätigkeit: unregelmäßige Beschäftigung	-0,023 ^a	1,2%	
Varianzaufklärung (korrigiertes R²)		0,180	

Tab.13 (Fortsetzung): Parameter der Regressionsmodelle zur Risikoadjustierung stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Rehabilitationsbeginn für die Dimension PROMIS-10 Mental Health; *: $p < 0.05$; ‡: $p < 0.10$; a: $p = n. s.$; b: Prädiktor wurde fix in allen Modellen für diese Dimension eingeschlossen; c: Prädiktor resultierend auf der Basis der Modellierungstichprobe mit vollständigen Daten, aber kein statistisch signifikanter Beitrag in den erweiterten finalen Analysedaten; β -Gewicht: Ändert sich der Prädiktor um eine Standardabweichung, verändert sich das erwartete Outcome um den Betrag von β Standardabweichungen

Risikoadjustierungsmodelle für VR-12 PCS

VR-12 Physical Component Score (PCS) – Modelle der stratifizierten Risikoadjustierung		Alle Kliniken
Fälle gesamt zur Vorhersage des zu erwarteten Outcomes (N)		336
Niedrige HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 PCS ≤ 40) (n)		177
Merkmal/Prädiktor	β-Gewicht	M (SD) / %
VR-12 PCS Aufnahme (M/SD) ^b	0,407*	31,2 (6,6)
Alter ^b	-0,220*	67,2 (11,7)
weiblich ^b	0,060 ^a	46,9%
vor Reha weniger als 6 Monate arbeitsunfähig	0,203*	87,0%
ledig	-0,172*	9,0%
kein oder niedriger Schulabschluss	0,130 [‡]	53,7%
Varianzaufklärung (korrigiertes R²)		0,295
Mittlere HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 PCS: > 40 bis ≤ 45) (n)		63
Merkmal/Prädiktor	β-Gewicht	M (SD) / %
VR-12 PCS Aufnahme (M/SD) ^b	-0,062 ^a	42,4 (1,4)
Alter ^b	-0,358*	64,1 (12,1)
weiblich ^b	0,090 ^a	44,4%
VR-12 Mental Component Score (MCS) Aufnahme	0,398 [‡]	41,7 (12,4)
Krank/arbeitsunfähig vor Reha: ja	-0,436*	42,9%
ohne feste Partnerschaft	-0,518*	30,2%
ledig	0,356*	12,7%
geschieden	0,334*	15,9%
verwitwet	0,283 [‡]	11,1%
Varianzaufklärung (korrigiertes R²)		0,251

Tab.14: Parameter der Regressionsmodelle zur Risikoadjustierung stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Rehabilitationsbeginn für die Dimension VR-12 Physical Component Score; *: p < 0.05; †: p < 0.10; a: p = n. s.; b: Prädiktor wurde fix in allen Modellen für diese Dimension eingeschlossen; c: Prädiktor resultierend auf der Basis der Modellierungstichprobe mit vollständigen Daten, aber kein statistisch signifikanter Beitrag in den erweiterten finalen Analysedaten; β-Gewicht: Ändert sich der Prädiktor um eine Standardabweichung, verändert sich das erwartete Outcome um den Betrag von β Standardabweichungen

Fortsetzung auf der nächsten Seite.

Fortsetzung Tabelle 14.

VR-12 Physical Component Score (PCS) – Modelle der stratifizierten Risikoadjustierung		Alle Kliniken	
Hohe HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 PCS > 45) (n)		96	
Merkmal/Prädiktor	β -Gewicht	M (SD) / %	
VR-12 PCS Aufnahme (M/SD) ^b	0,061 ^a	49,9 (3,5)	
Alter ^b	0,044 ^a	59,8 (12,4)	
weiblich ^b	-0,055 ^a	33,3%	
Weitere relevante Diagnosen: ja	0,218*	15,6%	
Muttersprache nicht deutsch	-0,222*	2,1%	
getrenntlebend	-0,188 [‡]	3,1%	
mittlerer Schulabschluss	0,242*	36,5%	
Fachhochschule/Universität	0,356*	19,8%	
Varianzaufklärung (korrigiertes R²)		0,150	

Tab.14: Parameter der Regressionsmodelle zur Risikoadjustierung stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Rehabilitationsbeginn für die Dimension VR-12 Physical Component Score; *: $p < 0.05$; †: $p < 0.10$; a: $p = n. s.$; b: Prädiktor wurde fix in allen Modellen für diese Dimension eingeschlossen; c: Prädiktor resultierend auf der Basis der Modellierungsstichprobe mit vollständigen Daten, aber kein statistisch signifikanter Beitrag in den erweiterten finalen Analysedaten; β -Gewicht: Ändert sich der Prädiktor um eine Standardabweichung, verändert sich das erwartete Outcome um den Betrag von β Standardabweichungen

Risikoadjustierungsmodelle für VR-12 MCS

VR-12 Mental Component Score (MCS) – Modelle der stratifizierten Risikoadjustierung		Alle Kliniken	
Fälle gesamt zur Vorhersage des zu erwarteten Outcomes (N)		345	
Niedrige HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 MCS ≤ 40) (n)		139	
Merkmal/Prädiktor	β-Gewicht	M (SD) / %	
VR-12 MCS Aufnahme (M/SD) ^b	0,164*	31,0 (6,3)	
Alter ^b	0,000 ^a	64,9 (12,1)	
weiblich ^b	-0,174*	53,2%	
VR-12 Physical Component Score (PCS) Aufnahme	0,236*	36,4 (36,4)	
Anzahl somatische Diagnosen	0,157*	2,5 (1,5)	
geschieden	-0,154*	11,5%	
Berufsbildung: Ausbildung/ohne Abschluss/sonstiges	0,152 [‡]	30,9%	
EMR-Antrag gestellt/geplant	-0,214*	18,0%	
Varianzaufklärung (korrigiertes R²)		0,228	
Mittlere HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 MCS: > 40 bis ≤ 45) (n)		46	
Merkmal/Prädiktor	β-Gewicht	M (SD) / %	
VR-12 MCS Aufnahme (M/SD) ^b	0,469*	42,2 (1,3)	
Alter ^b	-0,306 ^a	66,6 (12,7)	
weiblich ^b	-0,463*	45,7%	
VR-12 Physical Component Score (PCS) Aufnahme	0,457*	38,6 (9,2)	
Reha als ganztägig ambulante Maßnahme: ja	0,338*	4,3%	
Muttersprache nicht deutsch	0,440*	2,2%	
Wohnort neue Bundesländer	0,214 [‡]	23,9%	
ohne feste Partnerschaft	0,298*	23,9%	
ledig	-0,352*	13,0%	
ohne berufliche Beschäftigung/unklar	0,366*	6,5%	
Frührente/Altersrente/Pension	0,433*	52,2%	
EMR-Antrag gestellt/geplant	-0,503*	10,9%	
Varianzaufklärung (korrigiertes R²)		0,627	

Tab.15: Parameter der Regressionsmodelle zur Risikoadjustierung stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Rehabilitationsbeginn für die Dimension VR-12 Mental Component Score; *: $p < 0.05$; †: $p < 0.10$; a: $p = n. s.$; b: Prädiktor wurde fix in allen Modellen für diese Dimension eingeschlossen; c: Prädiktor resultierend auf der Basis der Modellierungsstichprobe mit vollständigen Daten, aber kein statistisch signifikanter Beitrag in den erweiterten finalen Analysedaten; β-Gewicht: Ändert sich der Prädiktor um eine Standardabweichung, verändert sich das erwartete Outcome um den Betrag von β Standardabweichungen.

Fortsetzung auf der nächsten Seite.

Fortsetzung Tabelle 15.

VR-12 Mental Component Score (MCS) – Modelle der stratifizierten Risikoadjustierung		Alle Kliniken	
Hohe HRQoL zum Aufnahmezeitpunkt (d. h. VR-12 MCS > 45) (n)		160	
Merkmal/Prädiktor	β-Gewicht	M (SD) / %	
VR-12 MCS Aufnahme (M/SD) ^b	0,211*	54,2 (6,2)	
Alter ^b	0,096 ^a	63,9 (12,5)	
weiblich ^b	0,028 ^a	34,4%	
Krank/arbeitsunfähig vor Reha: ja	0,135 ^a	37,5%	
verwitwet	-0,177*	8,1%	
kein oder niedriger Schulabschluss	-0,196 [‡]	40,6%	
mittlerer Schulabschluss	-0,223*	35,6%	
Varianzaufklärung (korrigiertes R²)		0,077	

Tab.15: Parameter der Regressionsmodelle zur Risikoadjustierung stratifiziert nach niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Rehabilitationsbeginn für die Dimension VR-12 Mental Component Score; *: $p < 0.05$; †: $p < 0.10$; a: $p = n. s.$; b: Prädiktor wurde fix in allen Modellen für diese Dimension eingeschlossen; c: Prädiktor resultierend auf der Basis der Modellierungsstichprobe mit vollständigen Daten, aber kein statistisch signifikanter Beitrag in den erweiterten finalen Analysedaten; β-Gewicht: Ändert sich der Prädiktor um eine Standardabweichung, verändert sich das erwartete Outcome um den Betrag von β Standardabweichungen

Analyse der Linearität des Zusammenhangs von Aufnahmewert zu Katamnesewert im PROMIS-10 und der Veränderungen der Rangreihen im Einrichtungsvergleich in Abhängigkeit des Risikoadjustierungsmodells

Im Abschnitt „Deckeneffekte im PROMIS-10 und VR-12“ (S. 26) wurden die beobachteten Deckeneffekte im PROMIS-10 (bzw. auch im VR-12) beschrieben. Um zu prüfen, ob die Deckeneffekte einen Einfluss auf die Ergebnisse der Risikoadjustierung für den Einrichtungsvergleich haben können, wurde dazu zum einen der (für das Regressionsmodell häufig wichtigste^b) bivariate lineare Zusammenhang zwischen den PROMIS-10 Werten zum Aufnahmezeitpunkt und zum Katamnesezeitpunkt (als abhängige Variable zur Berechnung des Erwartungswerts) grafisch durch die Beurteilung der Kongruenz der Regressionsgeraden für das wie im Abschnitt „Risikoadjustierung“ (S. 13) beschriebene Gesamtmodell und für nach den Gruppen niedrige, mittlere und hohe gesundheitsbezogene Lebensqualität zum Aufnahmezeitpunkt stratifizierte Regressionsmodelle analysiert. Zum anderen wurden die Ergebnisse des Einrichtungsvergleichs für die zwei verschiedenen Ansätze der Risikoadjustierung (Risikoadjustierung für das Gesamtmodell vs. stratifizierte Risikoadjustierung) miteinander verglichen.

In Abbildung 18 sind die Ergebnisse aller Patient:innen (N = 1085) über alle 17 teilnehmenden Einrichtungen im PROMIS-10 Physical Health zum Aufnahmezeitpunkt (X-Achse) und zum Katamnesezeitpunkt (Y-Achse) abgetragen. Zudem sind vier verschiedene Regressionsgeraden eingetragen, wobei jede der Geraden einer linearen Regressionsgleichung in der Form $y = b \cdot x + a$ des bivariaten Zusammenhangs r_{xy} entspricht. Die Regressionsgrade des bivariaten Zusammenhangs der Wertepaare Aufnahme/Katamnese für die Gesamtstichprobe ist in schwarzer Farbe eingetragen. Die rote Linie entspricht der Regressionsgeraden, die sich nur für den Zusammenhang der Wertepaare in der Punkteverteilung der Gruppe von Patient:innen mit niedriger körperlicher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt (T-Werte ≤ 40) ergibt. Die gelbe Linie entspricht gleichermaßen der Regressionsgerade nur für Patient:innen mit mittlerer gesundheitsbezogener Lebensqualität (T-Werte > 40 bis ≤ 45) und die grüne Linie der Regressionsgerade der Patient:innen mit hohen T-Werten zum Aufnahmezeitpunkt (> 45). Aus dem Diagramm kann abgelesen werden, dass die Regressionsgerade für Patient:innen mit niedriger gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt (rot) im T-Wertebereich von 29,6 bis 39,8 (X-Achse) etwa auf der Regressionsgeraden für den Zusammenhang der Werte in der Gesamtstichprobe (schwarz) liegt. Im T-Wertebereich von 19,9 bis 26,7 liegt die rote Gerade dabei leicht unterhalb der schwarzen, was bedeutet, dass in diesem Bereich das Regressionsmodell für die Gesamtstichprobe für den Katamnesezeitpunkt etwas höhere Outcomes (als Erwartungswert) schätzt, als das Regressionsmodell für Patient:innen mit einem T-Wert ≤ 40 zur Aufnahme. In der Patient:innengruppe mit mittleren T-Werten zur Aufnahme (T-Werte von > 40 bis ≤ 45) liegt die (gelbe) Regressionsgerade beim T-Wert von 42,3 zur Aufnahme etwas über der (schwarzen)

^b Zu der Relevanz einzelner Prädiktoren zur Vorhersage des Outcomes zum Katamnesezeitpunkt siehe die β -Gewichte der einzelnen Regressionsmodelle ab S. 44.

Regressionsgerade für die Gesamtstichprobe, d. h. das gruppenspezifische Regressionsmodell schätzt das Outcome zum Katamnesezeitpunkt etwas höher als das Regressionsmodell für die Gesamtstichprobe. Bei Patient:innen mit einem T-Wert von 44,9 berechnen beide Regressionsmodelle annähernd gleich hohe Erwartungswerte für den Katamnesezeitpunkt. In der Patient:innengruppe mit einer höheren gesundheitsbezogenen Lebensqualität zum Aufnahmezeitpunkt (T-Werte >45) liegt die gruppenspezifische (grüne) Regressionsgerade bis zum T-Wert von 54,1 eng beieinander. Ab einen T-Wert von 57,7 zum Beginn der Rehabilitation schätzt das Regressionsmodell für die Gesamtstichprobe höhere Erwartungswerte zum Katamnesezeitpunkt ein, als das gruppenspezifische Modell. Über den gesamten T-Wertebereich zum Aufnahmezeitpunkt betrachtet, liegen die Regressionsgerade für das Gesamtmodell und die gruppenspezifische Regressionsgerade für Patient:innen mit einer niedrigen gesundheitsbezogenen Lebensqualität zum Aufnahmezeitpunkt am dichtesten beieinander.

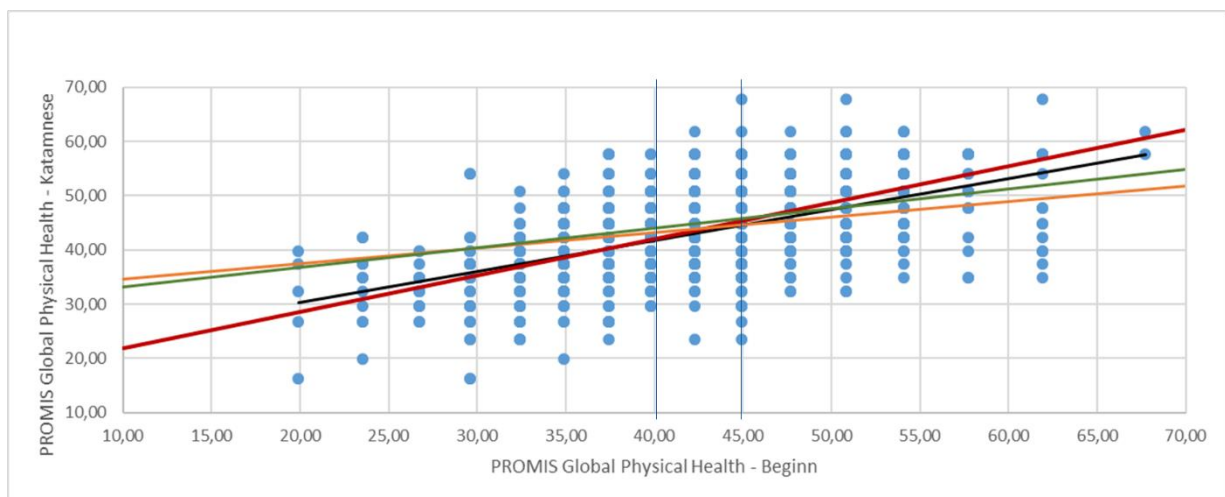


Abb.18: Punktverteilungen der Ergebnisse aller Patient:innen (N = 1085) über alle 17 teilnehmenden Einrichtungen im PROMIS-10 Physical Health zum Aufnahmezeitpunkt (X-Achse) und zum Katamnesezeitpunkt (Y-Achse) mit Regressionsgeraden der Zusammenhänge zwischen Aufnahme- und Katamnesewert. Schwarze Regressionsgerade: Gesamtstichprobe. Rote Regressionsgerade: Gruppe der Patient:innen mit niedriger körperlicher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt (T-Werte ≤ 40). Gelbe Regressionsgerade: Gruppe der Patient:innen mit mittlerer gesundheitsbezogener Lebensqualität (T-Werte > 40 bis ≤ 45). Grüne Regressionsgerade: Gruppe der Patient:innen mit hohen T-Werten zum Aufnahmezeitpunkt (> 45).

Für die Dimension Mental Health des PROMIS-10 zeigt sich beim Vergleich der gruppenspezifischen Regressionsgeraden und der Regressionsgeraden für die Gesamtstichprobe ein anderes Bild (Abbildung 19): Hier liegen über den gesamten T-Wertebereich zum Aufnahmezeitpunkt die Regressionsgerade der Gesamtstichprobe und der spezifischen Regressionsgerade für die Patient:innengruppen mit niedriger und hoher gesundheitsbezogener Lebensqualität eng beieinander, und für den T-Wertebereich von > 40 bis ≤ 45 gilt dies auch für die Regressionsgeraden der Gesamtstichprobe und der Gruppe mit einer mittleren mentalen gesundheitsbezogenen Lebensqualität.

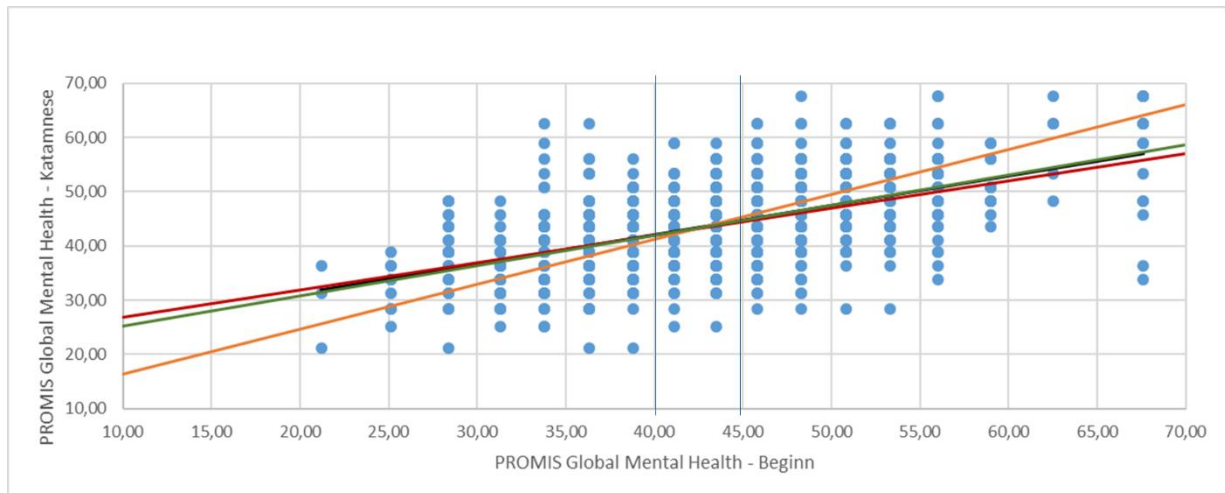


Abb.19: Punktverteilungen der Ergebnisse aller Patient:innen (N = 1119) über alle 17 teilnehmenden Einrichtungen im PROMIS-10 Mental Health zum Aufnahmezeitpunkt (X-Achse) und zum Katamnesezeitpunkt (Y-Achse) mit Regressionsgeraden der Zusammenhänge zwischen Aufnahme- und Katamnesewert. Schwarze Regressionsgerade: Gesamtstichprobe. Rote Regressionsgerade: Gruppe der Patient:innen mit niedriger körperlicher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt (T-Werte ≤ 40). Gelbe Regressionsgerade: Gruppe der Patient:innen mit mittlerer gesundheitsbezogener Lebensqualität (T-Werte > 40 bis ≤ 45). Grüne Regressionsgerade: Gruppe der Patient:innen mit hohen T-Werten zum Aufnahmezeitpunkt (> 45).

Zusammengefasst lässt sich beobachten, dass sich die Steigungskoeffizienten b für den PROMIS-10 PH in den gruppenspezifischen Regressionsgeraden gegenüber dem Koeffizienten der Geraden für die Gesamtstichprobe in der Form unterscheiden, dass ein Regressionsmodell für die Gesamtstichprobe zur Risikoadjustierung für den Einrichtungsvergleich die geschätzten Erwartungswerte des Outcomes zum Katamnesezeitpunkt für Patient:innen mit besonders niedrigen Werten zum Aufnahmezeitpunkt etwas höher und für Patient:innen mit besonders hohen Werten zum Aufnahmezeitpunkt deutlich höher berechnen würde, als es bei der Berechnung von gruppenspezifischen Erwartungswerten der Fall wäre. Es ist also über verschiedene Patient:innengruppen hinweg (unter Außerachtlassung weiterer, potentiell relevanter Prädiktoren mit einem Zusammenhang mit dem Outcome) kein vollständig linearer bivariater Zusammenhang zwischen dem Ergebnis zur Aufnahme in die Rehabilitation und dem Ergebnis zum Katamnesezeitpunkt gegeben. Aufgrund der Heterogenität der Proportionen von Patient:innengruppen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt zwischen den Einrichtungen wäre daher anzunehmen, dass bei einer Risikoadjustierung mit dem Outcome des PROMIS-10 PH als abhängige Variable an einer Gesamtstichprobe einige Einrichtungen benachteiligt sein können, wenn der Erwartungswert des Outcomes für Einrichtungen mit einem hohen Anteil von Patient:innen mit hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt systematisch höher liegt, als es bei der Berechnung von gruppenspezifischen Erwartungswerten der Fall wäre. Für die Mental Health Skala des PROMIS-10 wäre auf der Grundlage der Analyse der Regressionsgeraden dagegen ein linearer Zusammenhang zwischen den Ergebnissen von Aufnahme- und Katamnesezeitpunkt gegeben, so dass bei alleiniger Betrachtung des bivariaten

Zusammenhangs ohne Einbeziehung anderer Prädiktoren keine Unterschiede in den Erwartungswerten des Outcomes in Abhängigkeit zwischen einem Regressionsmodell basierend auf der Gesamtstichprobe und gruppenspezifischen Regressionsmodellen zu erwarten wären.

Zur weiteren Überprüfung der Unterschiede in den Ergebnissen nach der Durchführung der Risikoadjustierung in Abhängigkeit der verschiedenen Regressionsmodelle werden die Rangreihen der Einrichtungen je nach angewandter Risikoadjustierungsmethode unter Einbeziehung weiterer Prädiktoren verglichen. Dazu werden die Ergebnisse der Analysen zu den Effektstärken der Differenzen zwischen erreichten Outcomes und den Erwartungswerten herangezogen, und die Einrichtungen in Rangreihen in Abhängigkeit der Effektstärken und der statistischen Absicherung eines über- oder unterdurchschnittlichen Ergebnisses gebracht. Die Unterschiede in den Rangreihen nach a) Risikoadjustierung (RA) an der Gesamtstichprobe und nach b) stratifizierter Risikoadjustierung nach Gruppen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt werden dazu untersucht.

PROMIS-10 Physical Health

a) RA Gesamtstichprobe

b) RA stratifiziert nach Ergebnissen zur Aufnahme

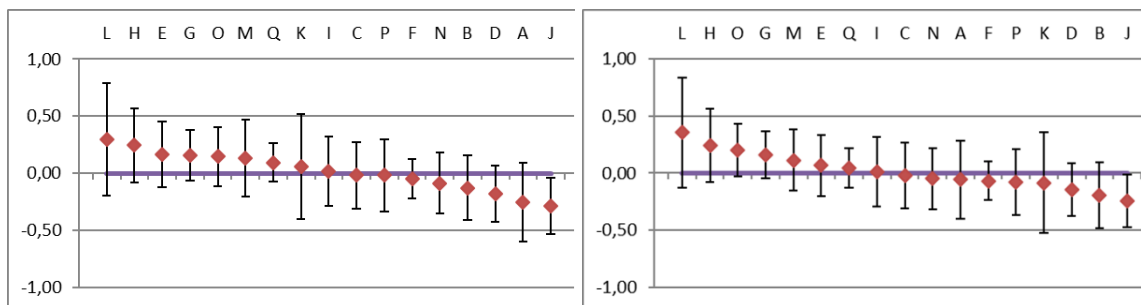
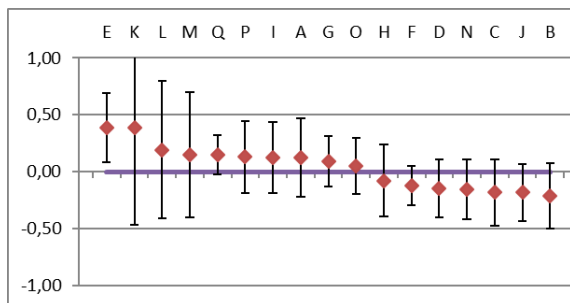


Abb.20: Effektstärken der Abweichungen der risikoadjustierten Erwartungswerte von den erreichten Outcomes im PROMIS-10 Physical Health und Rangreihen der Einrichtungen sortiert nach der Höhe der Effektstärken der Abweichungen für a) das Regressionsmodell der Gesamtstichprobe und b) die nach Gruppen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt stratifizierten Regressionsmodelle.

Im Vergleich zur Risikoadjustierung an der Gesamtstichprobe für den PROMIS-10 PH erreichen nach einer Risikoadjustierung mit Stratifizierung nach gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt sechs Einrichtungen einen um bis zu fünf Rangplätze besseren Rang, sieben Einrichtungen verändern ihren Rang nicht, und vier Einrichtungen erreichen einen um bis zu sechs Rangplätze schlechteren Rang (Abbildung 20).

PROMIS-10 Mental Health

a) RA Gesamtstichprobe



b) RA stratifiziert nach Ergebnissen zur Aufnahme

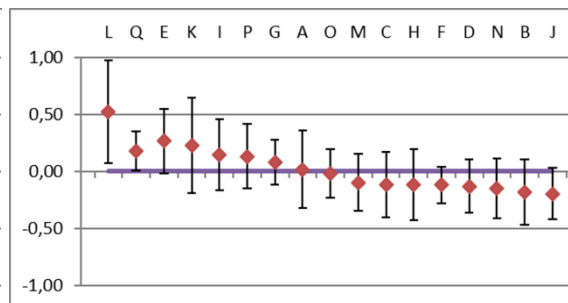


Abb.21: Effektstärken der Abweichungen der risikoadjustierten Erwartungswerte von den erreichten Outcomes im PROMIS-10 Physical Health und Rangreihen der Einrichtungen sortiert nach der Höhe der Effektstärken der Abweichungen für a) das Regressionsmodell der Gesamtstichprobe und b) die nach Gruppen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt stratifizierten Regressionsmodelle.

Im Vergleich zur Risikoadjustierung an der Gesamtstichprobe für den PROMIS-10 MH erreichen nach einer Risikoadjustierung mit Stratifizierung nach gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt sieben Einrichtungen einen um bis zu vier Rangplätze besseren Rang, zwei Einrichtungen verändern ihren Rang nicht, und acht Einrichtungen erreichen einen um bis zu sechs Rangplätze schlechteren Rang (Abbildung 21).

An den Änderungen der Rangreihen (auch für die Skala Mental Health) wird deutlich, dass nicht nur der Zusammenhang zwischen den Ergebnissen in den Skalen des PROMIS-10 zum Aufnahme- und Katamnesezeitpunkt je nach Wahl der Risikoadjustierungsmethode einen Einfluss auf die resultierenden Differenzen zwischen Outcomes und Erwartungswerten hat, sondern auch die weiteren in die Risikoadjustierung einbezogenen Prädiktoren, die sich sowohl zwischen den Methoden (Gesamtmodell vs. stratifizierte Modelle) als auch zwischen den stratifizierten Gruppen unterscheiden. Aufgrund der unterschiedlichen einbezogenen Prädiktoren verändert sich im Vergleich zur Risikoadjustierung an der Gesamtstichprobe auch die Anzahl einbezogener Fälle und somit der Umfang der Konfidenzintervalle um die Effektstärken der Differenzen. In der Zusammenschau der oben beschriebenen Ergebnisse zur Schätzung der risikoadjustierten Erwartungswerte durch stratifizierte Regressionsmodelle, den beobachteten Änderungen der Rangreihen im PROMIS-10 Physical Health bei Durchführung der stratifizierten Risikoadjustierung und unter Berücksichtigung der beobachteten unterschiedlichen Proportionen von Patient:innen mit niedriger, mittlerer und hoher gesundheitsbezogener Lebensqualität zwischen den Einrichtungen (siehe „Deckeneffekte im PROMIS-10 und VR-12“, S.26) lässt sich schließen, dass durch die stratifizierte Risikoadjustierung präzisere Schätzungen der Erwartungswerte für die unterschiedlichen Gruppen von Patient:innen möglich sind und dadurch Einrichtungen mit einem besonders hohen Anteil von Patient:innen mit hoher gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt (insbesondere Einrichtung „A“) in ihrem Outcome fairer mit allen anderen

Einrichtungen verglichen werden können. Aufgrund dieser Ergebnisse und für eine methodische Konsistenz der Analysen zum Einrichtungsvergleich haben wir uns entschlossen, dem Einrichtungsvergleich und der Berechnung des ProQIs die nach gesundheitsbezogener Lebensqualität zum Aufnahmezeitpunkt stratifizierten Regressionsmodelle der Risikoadjustierung beider Dimensionen des PROMIS-10 und des VR-12 zugrunde zu legen.

Vergleichsdaten für den PROMIS-10 aus fünf Einrichtungen der Medical Park Gruppe mit zusätzlicher Messung zur Entlassung aus der Rehabilitation

Die Ergebnisse aus fünf Einrichtungen der Medical Park Gruppe, die auch an dieser Pilotstudie teilgenommen hat, weisen darauf hin, dass zwischen der Aufnahme in die Rehabilitation und dem Entlassungszeitpunkt größere positive Veränderungen in der patient:innenberichteten gesundheitsbezogenen Lebensqualität messbar waren, die zum Katamnesezeitpunkt deutlich abfielen (Abbildung 22).

Datensatz zu drei Messzeitpunkten

Die Prozedur MEANS

Variable	N	Mittelwert	Std.abw.	Minimum	Maximum
Physical_Aufnahme	102	12.9803922	2.8178371	6.0000000	19.0000000
Physical_Entlassung	102	15.1862745	2.9374010	8.0000000	20.0000000
Physical_FU	102	13.3137255	3.0147505	6.0000000	19.0000000
Mental_Aufnahme	102	11.9411765	2.7604833	5.0000000	19.0000000
Mental_Entlassung	102	14.4607843	3.4142148	6.0000000	20.0000000
Mental_FU	102	12.3235294	3.2735416	5.0000000	20.0000000
Physical_aufnahmeT	102	42.6500000	7.6699846	23.5000000	61.9000000
Physical_EntlassungT	102	49.2931373	9.1029563	29.6000000	67.7000000
Physical_FUT	102	43.6068627	8.3523313	23.5000000	61.9000000
Mental_aufnahmeT	102	43.3735294	6.9168316	25.1000000	62.5000000
Mental_EntlassungT	102	50.1401961	9.3594899	28.4000000	67.6000000
Mental_FUT	102	44.3892157	8.2270732	25.1000000	67.6000000

Abb.22: Deskriptive Statistiken zu Scores im PROMIS-10 (Physical Health & Mental Health) zur Aufnahme, zur Entlassung und zum Katamnesezeitpunkt aus fünf Rehabilitationseinrichtungen der Medical Park Gruppe. Die T-Werte der Scores werden in den sechs Variablen Physical_aufnahmeT bis Mental_FUT beschrieben. Variablensuffix „FUT“: T-Werte zum Follow-Up (= Katamnesezeitpunkt).

Literatur

- Buchholz, I., Feng, Y.-S., Buchholz, M., Kazis, L. E. & Kohlmann, T. (2020).** The German Version of the Veterans Rand-36/12 Item Health Survey: Translation and Adaptation of a Generic Health Assessment Instrument in a Sample of Inpatient Rehabilitation Patients.
- Buchholz, I., Kohlmann, T. & Buchholz, M. (2017).** *Vergleichende Untersuchung der psychometrischen Eigenschaften des SF-36/SF-12 vs. VR-36/VR-12* (No. VFFR-205). Greifswald: Institut für Community Medicine der Universitätsmedizin Greifswald, Abteilung Methoden der Community Medicine.
- Katzan, I. L. & Lapin, B. (2018).** PROMIS GH (Patient-Reported Outcomes Measurement Information System Global Health) Scale in Stroke. *Stroke*, 49(1), 147-154. <https://www.ahajournals.org/doi/abs/10.1161/STROKEAHA.117.018766>
- Kazis, L. E., Miller, D. R., Skinner, K. M., Lee, A., Ren, X. S., Clark, J. A. et al. (2004).** Patient-reported measures of health: The Veterans Health Study. *J Ambul Care Manage*, 27(1), 70-83. <https://www.ncbi.nlm.nih.gov/pubmed/14717468>