

Cluster-randomisierte Studien: Methodische und praktische Aspekte

Cluster-Randomized Trials: Methodical and Practical Implications

Autoren

Jens Dreyhaupt¹, Benjamin Mayer¹, Rainer Kaluscha², Rainer Muehe¹

Institute

- 1 Institut für Epidemiologie und Medizinische Biometrie, Universität Ulm
- 2 Institut für Rehabilitationsmedizinische Forschung an der Universität Ulm

Schlüsselwörter

Cluster-Randomisierung, Strukturgleichheit, Rehabilitationsforschung, Fallzahlplanung, Statistische Auswertung

Key words

cluster randomization, structural equivalence, rehabilitation research, sample size calculation, statistical analysis

Bibliografie

DOI <https://doi.org/10.1055/a-0801-5697>

Online-Publikation: 23.1.2019

Rehabilitation 2020; 59: 54–61

© Georg Thieme Verlag KG Stuttgart · New York

ISSN 0034-3536

Korrespondenzadresse

Dr. Jens Dreyhaupt

Institut für Epidemiologie und Medizinische Biometrie

Universität Ulm

Schwabstraße 13

89075 Ulm

jens.dreyhaupt@uni-ulm.de

ZUSAMMENFASSUNG

Häufig wird kritisiert, dass in der Rehabilitationsforschung zu wenig randomisierte Studien durchgeführt werden, um methodisch hochwertige Evidenz zu erhalten. Allerdings wird auch anerkannt, dass sich das Design der doppel-blinden, Placebo-kontrollierten, randomisierten Studie häufig nicht einfach in die Rehabilitationsforschung übertragen lässt. Zum einen steht hier die Validität in Bezug auf den Versorgungsaspekt im Fokus, sodass Treatment-as-usual anstelle der Placebo-Behandlung als Referenz dient und eine Doppelverblindung teilweise nicht

definierbar ist. Weiterhin ist es aufgrund fehlender Kapazitäten oft schwierig, 2 ähnliche Therapieformen in einer Rehabilitationseinrichtung parallel durchzuführen. Ferner sind Kontaminationseffekte zu befürchten, wenn sich Rehabilitanden unterschiedlicher Studienarme über ihre Therapie austauschen. Hier bieten sich cluster-randomisierte Studien an, welche allerdings im Vergleich zu individuell randomisierten Studien meist höhere Fallzahlen, eine komplexere Methodik der Fallzahlplanung sowie umfassendere Methoden der statistischen Auswertung erfordern.

Im Beitrag werden Vor- und Nachteile sowie Besonderheiten der Cluster-Randomisierung dargestellt und Hinweise zu ihrer Umsetzung im Bereich der Rehabilitationsforschung gegeben.

ABSTRACT

Quite often critics demand more randomized studies in rehabilitation science to gather methodological evidence of high quality. However, it is also recognized that the design of double-blind, placebo-controlled, randomized studies often cannot simply be transferred into rehabilitation science. Validity concerning the health care is here in the focus. Thus, treatment as-usual is mostly used as placebo treatment and double-blinding is partly not definable. Additionally, it is often difficult to offer 2 similar forms of treatment in one rehabilitation hospital due to lack of capacity. Additionally, contamination effects are to be expected when patients of different study arms communicate. Here cluster-randomized studies may be helpful. However, in comparison to individual randomized studies they need often higher sample sizes, a more complex methodology of sample size calculation as well as extensive methods of statistical analysis.

Within this article advantages and disadvantages as well as the characteristics of cluster randomization are described and information is given how they can be implemented into the field of rehabilitation science.

Einleitung

Häufig wird in der Rehabilitationsforschung der Wunsch nach einer verstärkten Durchführung randomisierter Studien zur Gewinnung methodisch hochwertiger Evidenz geäußert. Allerdings wird auch anerkannt, dass sich das Design der doppel-blinden, Placebo-kontrollierten, randomisierten Studie oft nicht einfach in die Rehabilitationsforschung übertragen lässt [1]. Beispielsweise sind noch Fragen offen, wie ein „Reha-Placebo“ aussehen und wie die Verblindung von Behandlern und Rehabilitanden erfolgen könnte. Ein „Reha-Placebo“ müsste eine Behandlung sein, die zwar nicht wirkt, aber von Rehabilitanden und Therapeuten nicht als unwirksam erkannt werden kann. Dazu kommt noch, dass eine Verblindung in den meisten Fällen weder bei Behandlern noch Patienten möglich ist, da beide in der Regel aktiv an der Therapie mitarbeiten [2]. Daher dürfte die Konstruktion eines „Reha-Placebos“ in den meisten Fällen nicht gelingen. Die ethischen und juristischen Schwierigkeiten der Randomisierung bei etablierten und potentiell nützlichen Therapien kommen noch hinzu. Denn viele Therapieelemente der Rehabilitation wie z. B. Physiotherapie oder Schulungsprogramme haben sich bereits auch in anderen Bereichen bewährt, sodass zu meist die für eine Vorenthaltung einer Behandlung erforderliche Bedingung des Unwissens hinsichtlich des Behandlungsnutzens („Equipose“) verletzt sein dürfte. Weiterhin gibt es manchmal auch praktische Schwierigkeiten. So ist es oft nicht möglich, 2 ähnliche Therapieformen in einer Rehabilitationseinrichtung parallel laufen zu lassen, weil bspw. für weitere Gruppen keine geeigneten Therapeuten oder Therapieplätze zur Verfügung stehen.

Andererseits sind in der Rehabilitationsforschung Standards entwickelt worden, welche die Validität in Bezug auf die Versorgungssituation im Vordergrund stellen, sodass Treatment-as-usual anstelle der Placebo-Behandlung als Referenz dient und eine Doppelverblindung teilweise nicht definierbar ist [3]. Lediglich eine verblindete Datenauswertung wird als Standard gefordert. Die Fragen nach einem „Reha-Placebo“ und der doppel-blinden Studierendurchführung sind in der Rehabilitationsforschung somit eher meist theoretischer Natur.

In der klinischen Forschung ist für den Wirksamkeits- oder Überlegenheitsnachweis einer Therapie ein Kontrollarm notwendig, mit dem der Interventionsarm im statistischen Sinne vergleichbar ist (Strukturgleichheit). Eine Randomisierung bewirkt, dass bekannte und unbekannte Störgrößen mit hoher Wahrscheinlichkeit ähnlich auf Kontroll- und Interventionsarm verteilt werden und sich bei Behandlungsbeginn höchstens zufällige Unterschiede ergeben [1]. Bei Verwendung anderer Zuteilungsverfahren der Studienteilnehmer auf Kontroll- und Interventionsarm kann es zu systematischen Verzerrungen der Ergebnisse kommen, z. B. bei quasi-experimentellen Designs [2]. Daher ist auch für Wirkungs- und Wirksamkeitsstudien in der Reha-Forschung die Randomisierung die Methode der Wahl, wo immer dies möglich ist [1]. Bei Studien in der Reha-Forschung sind darüber hinaus einige Besonderheiten und Anforderungen beim Einsatz der Randomisierung zu beachten. So ist es kaum möglich, Studienteilnehmer individuell zu randomisieren, wenn die Therapie in (geschlossenen) Gruppen stattfindet, wie es z. B. in der psychosomatischen Rehabilitation häufig zum Therapiekonzept gehört oder auch bei Patientenschulungen. Hierfür müsste die Zuteilung der Rehabilitanden zu den Therapiearmen im Anschluss an eine individuelle Randomisierung erfolgen, was jedoch

meist erhebliche organisatorische Probleme bereitet und der Versorgungsrealität in der Regel nicht entspricht. Daher ist hier eine Randomisierung ganzer Gruppen (Cluster-Randomisierung) anstelle individueller Randomisierung sinnvoll.

Um zu einer verstärkten Nutzung und besseren Akzeptanz der Cluster-Randomisierung in der Rehabilitationsforschung beizutragen, werden in dieser Arbeit ihre Grundlagen beschrieben, praktische Aspekte erläutert und auf die besonderen Bedingungen bei Studien in der Reha-Forschung eingegangen. Es werden das Prinzip der Cluster-Randomisierung dargestellt sowie Fallzahlplanung und statistische Auswertung cluster-randomisierter Studien, die sich von denen individuell randomisierter Studien unterscheiden. Abschließend wird eine exemplarische Studie mit Cluster-Randomisierung in der Reha-Forschung als Anwendungsbeispiel vorgestellt.

Der Hintergrund der Randomisierung: Strukturgleichheit

Aufgrund der bekannten Schwierigkeiten, in der Rehabilitation randomisierte kontrollierte Studien durchzuführen, werden dort z. T. Studien mit Einarm-Prä-Post-Designs durchgeführt. Mangels Kontrollarm erlauben diese aber keine gesicherten Aussagen über den Nutzen der Therapie, da die Entwicklung bei unbehandelten oder anders behandelten Rehabilitanden nicht bekannt ist. Bei kontrollierten Studien wird aus Gründen der Praktikabilität oft auf leicht verfügbare Kontrollarme (z. B. Patienten einer anderen Rehabilitationseinrichtung oder historische Kontrollen) zurückgegriffen. Unterscheiden sich hier jedoch Kontroll- und Interventionsarm (d. h. Verletzung von Struktur-, Behandlungs- oder Beobachtungsgleichheit), treten Interpretationsprobleme auf, da ein beobachteter Effekt nicht mehr ausschließlich der neuen Behandlungsmethode zugeschrieben werden kann [4]. Behandlungsgleichheit bedeutet dabei, dass die Patienten im Kontroll- und Interventionsarm auf dieselbe Art und Weise behandelt werden (z. B. gleichartige Begleittherapien) mit Ausnahme der Therapie, welche in der Studie untersucht werden soll. Insbesondere zur Kontrolle von Konfundierungen und für eine Sicherstellung der internen Validität ist das Erreichen der Strukturgleichheit eine wesentliche Voraussetzung. Eine Randomisierung ist daher als Verfahren zur Erreichung der Strukturgleichheit ein wichtiges Instrument und sollte wo immer möglich auch bei der Wirksamkeitsuntersuchung neuer Therapien in der Rehabilitationsforschung Anwendung finden. Hier eröffnet die Cluster-Randomisierung gerade für die Rehabilitationsforschung neue Möglichkeiten.

Bedingungen und Anforderungen an Studien in der Rehabilitationsforschung

Klinische Studien in der rehabilitationsmedizinischen Forschung finden in einem besonderen Kontext statt. Oft ist in den Rehabilitationseinrichtungen eine natürliche Clusterstruktur vorgegeben, da Therapien häufig in verschiedenen Versorgungseinheiten (z. B. Stationen, Kliniken) oder in Behandlungsgruppen durchgeführt werden. Beispiele sind Trainingsgruppen in der orthopädischen Rehabilitation, Gesprächsgruppen in der psychosomatischen Rehabilitation oder

Patientenschulungen. Insbesondere in der psychosomatischen Rehabilitation gehört ein geschlossenes Gruppensetting mitunter auch zum therapeutischen Konzept, d. h. die Rehabilitanden beginnen und beenden ihren Aufenthalt gemeinsam und erhalten ihre Therapien immer in derselben geschlossenen Gruppe. Oft gibt es in den Therapiegruppen eine eigene Gruppendynamik bedingt durch die Gruppenkonstellation. Dadurch hat jede Gruppe spezifische Merkmale, welche die Gestaltung und die Wirksamkeit der Therapie beeinflussen können.

Die Gruppengrößen sind in den einzelnen Therapieformen unterschiedlich und reichen von 4–6 Rehabilitanden in Kleingruppen über 8–12 Rehabilitanden bei Gruppen bis zu 50 bei Schulungsvorträgen [5]. Einzeltherapien werden üblicherweise nur dort verordnet, wo sie gegenüber Gruppentherapien einen deutlichen therapeutischen Mehrwert haben, um eine effiziente Leistungserbringung zu gewährleisten.

Innerhalb einer Therapiegruppe oder sogar Klinik sind Studienteilnehmer denselben Bedingungen ausgesetzt, wie bspw. denselben Therapeuten, derselben Ausstattung, denselben Räumlichkeiten und Zeiten. Eine vom Studienleiter kontrollierte randomisierte Aufteilung in mehrere Gruppen, die separat unterschiedliche Therapieelemente in derselben Versorgungseinheit erhalten, ist oft aus logistischen und Kapazitätsgründen nicht möglich, da Ressourcen wie z. B. Therapeuten oder entsprechend ausgestattete Therapieräume nur begrenzt zur Verfügung stehen. Daher kann bei Studien die Intervention oft nur auf Cluster-Ebene (z. B. Therapiegruppe, Klinik) variiert werden. Somit ist das Ergebnis von Studienteilnehmern innerhalb eines Clusters meist ähnlicher als das Ergebnis von Studienteilnehmern verschiedener Cluster.

Ferner berichten Kliniker auch von Problemen, wenn Rehabilitanden einer Klinik studienbedingt unterschiedliche Therapien erhalten, da sie sich dann untereinander über ihre Therapien austauschen. Dies kann zu Unzufriedenheit und Kontaminationseffekten führen [2]. Kontaminationseffekte (Spill-Over) können sowohl bei Rehabilitanden als auch bei Therapeuten auftreten. Würde man z. B. in einer Rehabilitationseinrichtung zeitgleich 2 verschiedene Patientenschulungen durchführen und miteinander vergleichen wollen, können durch Gespräche zwischen Rehabilitanden Informationen in den jeweils anderen Therapiearm fließen. Es wäre auch möglich, dass die Rehabilitanden untereinander ausgehändigte Materialien austauschen. Damit wäre die Trennung von Interventionsarm und Kontrollarm durchbrochen, da dann in beiden Armen ähnliche Therapieinhalte wirken könnten. Ferner kann es auch für Therapeuten, die bereits beide Therapieschemata kennen, schwierig sein, stets nur Inhalte und Methoden aus einem Schema einzusetzen, wenn situativ die Inhalte aus dem anderen Therapiearm sinnvoller wären. Hier können sowohl Irrtum als auch der verständliche Wunsch, die bestmögliche Therapie anzubieten, zu Vermischungen der eigentlich getrennten Behandlungskonzepte führen. Solche Probleme lassen sich zwar mit einem randomisierten Wechsel der Therapieform über die Zeit lösen, wobei allerdings aufgrund der mehrwöchigen Aufenthalte der Rehabilitanden lange „Auswaschphasen“ nach einem Therapiewechsel erforderlich werden. Dies führt dann zu hohen Verlusten potentieller Studienteilnehmer und senkt damit die Rekrutierungsrate, was die praktische Anwendbarkeit in der Rehabilitationsforschung stark einschränkt.

Cluster-Randomisierung und ihre Anwendung in der Reha-Forschung

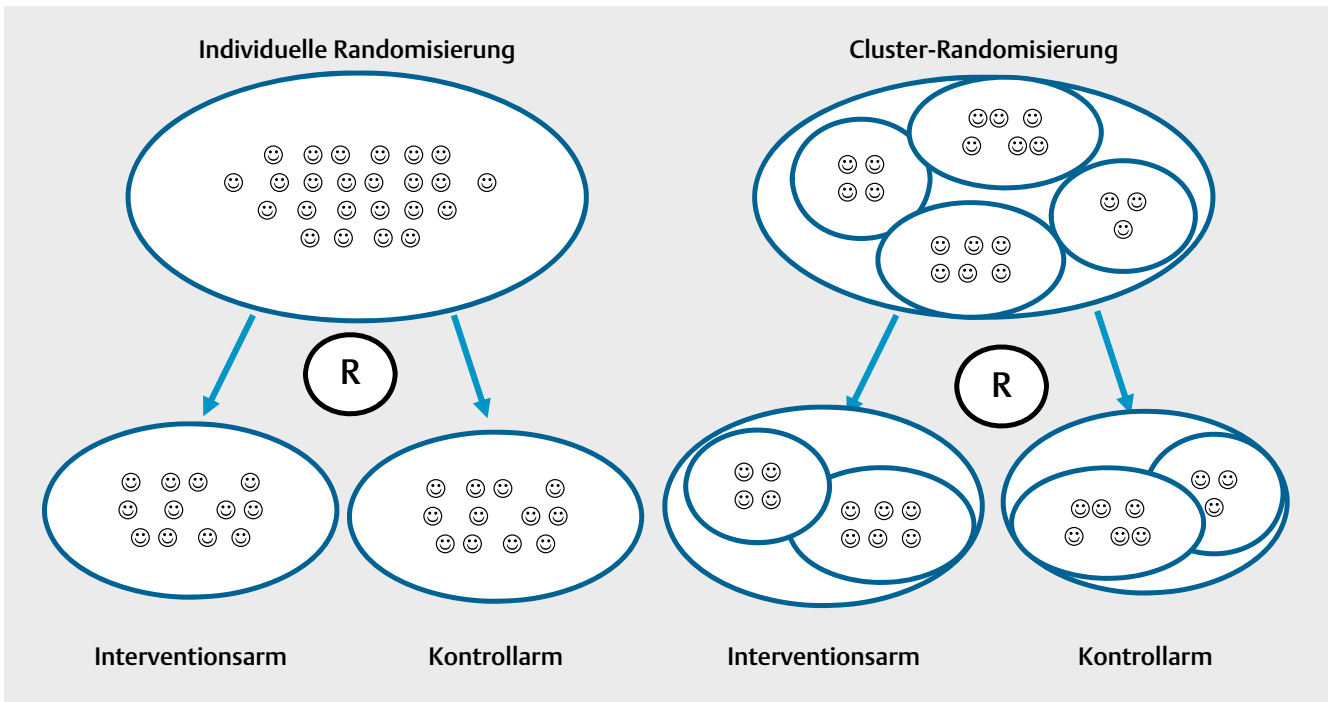
In den folgenden Abschnitten wird das Prinzip der Cluster-Randomisierung erläutert und Anwendungsmöglichkeiten im Bereich der Reha-Forschung skizziert. Weiterhin werden Informationen zu Studiendurchführung, Fallzahlplanung und Auswertung gegeben.

Definition und Motivation

In einer cluster-randomisierten Studie (engl. cluster randomised trial, Synonyme: „group randomised trial“, „community randomisation trial“, „community intervention trial“) werden komplette soziale Gruppen oder Cluster von Individuen anstelle von einzelnen Individuen randomisiert, siehe ► **Abb. 1**.

Durch die Cluster-Randomisierung allein ist allerdings noch keine Strukturgleichheit sichergestellt. Hierfür ist eine Mindestanzahl an Clustern notwendig (siehe Abschnitt Fallzahlplanung), wobei die Strukturgleichheit auf Individuen-Ebene nicht immer erreicht werden kann (siehe unten). Cluster-randomisierte Studien findet man in der klinischen und epidemiologischen Forschung, aber auch in der Rehabilitationsforschung, bspw. bei der Bewertung von Schulungsprogrammen, Präventionsprogrammen oder Maßnahmen zur Gesundheitsförderung [6]. Dabei werden Cluster häufig über Haushalte, Familien, Gemeinden, Schulklassen, Arbeitgeber, Krankenhäuser oder Praxen niedergelassener Ärzte definiert. Die Anzahl von Individuen pro Cluster kann stark schwanken. Aus den im letzten Abschnitt genannten Gründen ist die Verwendung cluster-randomisierter Studien auch im Bereich der Reha-Forschung empfehlenswert. Hauptmotivation für ihre Durchführung ist die Vermeidung oder Verringerung eines Kontaminationsbias. Ein solcher Bias (Verzerrung oder systematischer Fehler) kann bei individueller Randomisierung durch Interaktion zwischen den Individuen aus verschiedenen Studienarmen verursacht werden. Allerdings ist auch bei einer Cluster-Randomisierung diese Möglichkeit nicht ausgeschlossen, kann aber stark vermindert werden. Ein zweiter wesentlicher Grund für cluster-randomisierte Studien ist gegeben, wenn Interventionen auf höherer Ebene (bspw. Versorgungsstrukturen oder Teams) stattfinden. Hier stellt die Cluster-Randomisierung den Standard dar [7], da aufgrund des Wirkmodells die Wirksamkeit von Interventionen auf höherer Ebene meist aufgrund von Effekten der unteren Ebenen (oft Patientenebene) untersucht werden. Ein weiterer Grund für die Anwendung der Cluster-Randomisierung in der Reha-Forschung ist das Vorliegen natürlicher Cluster. ► **Tab. 1** zeigt wichtige Vor- und Nachteile der Cluster-Randomisierung im Kontext der Rehabilitationsforschung.

Weiterhin haben cluster-randomisierte Studien neben den in ► **Tab. 1** genannten Nachteilen im Vergleich zu individuell randomisierten Studien eine höhere Gefahr, dass die Strukturgleichheit auf Individuen-Ebene nicht erreicht wird. Dies kann die interne Validität gefährden, welche auch wegen der meist nicht durchführbaren Verblindung bei cluster-randomisierten Studien kritisch zu hinterfragen ist. Hier muss im Rahmen der statistischen Auswertung eine Adjustierung für ungleich verteilte Merkmale erfolgen, z. B. durch ein geeignetes Regressionsverfahren. Wegen der höheren Fallzahlen und der komplexeren Methodik sollte deshalb auch bei Studien in der Reha-Forschung in der Planungsphase überlegt werden, ob eine Cluster-Randomisierung gerechtfertigt und notwendig ist.



► **Abb. 1** Randomisierung von Individuen vs. Randomisierung von Clustern. R: Randomisierte Zuweisung von Individuen vs. Clustern zu 2 Therapiearmen.

► **Tab. 1** Vor- und Nachteile der Cluster-Randomisierung bei Studien in der Reha-Forschung.

Vorteile	Nachteile
<ul style="list-style-type: none"> – Berücksichtigung der Therapiedurchführung in natürlichen Clustern (Gruppen, Stationen, Kliniken) – Verminderung des Kontaminationsbias – Wirksamkeitsuntersuchung von Interventionen auf höherer Ebene (z. B. Reha-Einrichtungen) aufgrund von Effekten auf niedrigerer Ebene (z. B. Patienten) – Quantifizierung des Effekts der Clusterzugehörigkeit auf die Wirksamkeit von Therapien – Kontrolle und Quantifizierung des moderierenden Effekts von Clustereigenschaften auf die Wirksamkeit von Therapien 	<ul style="list-style-type: none"> – Im Vergleich zu individuell randomisierten Studien meist (deutlich) höhere Fallzahlen notwendig – Statistische Abhängigkeit der Patienten innerhalb von Gruppen: Komplexere statistische Methoden für Fallzahlplanung und Auswertung notwendig – Höhere Gefahr des Nichterreichens von Strukturgleichheit auf Individuen-Ebene

Studiendesigns

Bei cluster-randomisierten Studien kann zwischen vollständig randomisiertem Design, stratifiziert randomisiertem Design und gematchtem Design unterschieden werden. Weitere Aspekte sind die Ebene der Intervention (Ebene 1 oder Ebene 2) und eine mögliche Unterscheidung in Modelle mit zufälligen oder festen Effekten.

Beim vollständig randomisierten Design werden die Cluster zufällig den Gruppen zugeteilt und es wird dabei weder stratifiziert

noch gematcht. Dieses Verfahren ist insbesondere bei einer großen Clusteranzahl geeignet.

Beim stratifiziert randomisierten Design wird geschichtet nach (wenigen) bekannten Störgrößen randomisiert, sodass die Verteilung der potenziellen Störgrößen in Interventions- und Kontrollarm ähnlich ist. Stratifiziert wird nach Faktoren, die stark mit der Zielgröße assoziiert sind, wie bspw. Clustergröße oder Geschlecht. Die Randomisierung erfolgt dann innerhalb der durch die Faktoren gebildeten Subgruppen.

Im gematchten Design werden Paare von Clustern gebildet, die so ähnlich wie möglich in Bezug auf wichtige, die Zielgröße beeinflussende Faktoren sind. Ein Cluster des Paares wird jeweils in den Interventionsarm und das andere Cluster in den Kontrollarm randomisiert, wodurch bekannte Störgrößen (z. B. Ausstattung der Reha-Einrichtungen) zwischen beiden Armen balanciert werden, sodass die statistische Vergleichbarkeit der Arme erreicht wird. Das gematchte Design kann v. a. bei sehr wenigen oder stark unterschiedlichen Clustern sinnvoll sein.

Praktische Durchführung

Interventionsstudien in der Reha-Forschung bedürfen einer zustimmenden Bewertung der Ethikkommission. Details sowie Aspekte des Datenschutzes bei Studien in der Rehabilitationsforschung sind in [8] beschrieben.

Ein- und Ausschlusskriterien müssen sowohl auf Individuen-Ebene (Rehabilitand) als auch auf Cluster-Ebene (z. B. Behandler, Reha-Einrichtung) definiert werden. Geeignete Maßnahmen zur Erreichung der Behandlungs- und Beobachtungsgleichheit sollten angewendet werden, um die interne Validität der Studie bestmöglich zu gewährleisten. Beispiele sind eine starke Standardisierung des generellen Vorgehens und der Art der Beobachtung, ggf. eine ver-

blindete Beurteilung des Therapieerfolgs, z. B. durch einen dritten, nicht in die Studie involvierten Bewerter, der keine Kenntnis über die Zugehörigkeit des jeweiligen Studienteilnehmers zu Interventions- und Kontrollarm hat. Weiterhin trägt die Auswertung nach dem Intention-to-treat Prinzip (ITT Prinzip) zu einer Verminderung des Selektionsbias bei [9] und sollte auch bei cluster-randomisierten Studien in der Rehabilitationsforschung angewendet werden.

Zu überlegen ist auch, welche Cluster überhaupt sinnvoll gebildet werden können, z. B. Teilnehmer einer bestimmten Gruppentherapie oder Clusterung nach Anreisezeitraum innerhalb einer Klinik oder Zuweisung ganzer Kliniken zu einem Cluster. Jede dieser Varianten hat Vor- und Nachteile; ggf. müssen bei der Studienplanung auch mögliche Behandler- oder Klinikeffekte berücksichtigt werden, wie bspw. eine besondere Empathie einzelner Behandler oder die spezifische Ausstattung einer Klinik, welche nur dort verfügbar ist.

Fallzahlplanung

Warum eine eigene Fallzahlplanung?

Eine adäquate Fallzahlkalkulation ist auch bei rehabilitationswissenschaftlichen Studien ein wichtiger Aspekt zum Erzielen qualitativ hochwertiger Evidenz [10]. Die im Rahmen cluster-randomisierter Studien entstehenden Daten besitzen ein Mehrebenenendesign und haben zusätzlich oft eine komplexe Struktur, da beispielsweise ebenenspezifische Störgrößen auf Cluster- und Individuen-Ebene berücksichtigt werden müssen [6]. Für diese Situationen sind spezielle Fallzahlplanungsmethoden anzuwenden und Aspekte zu berücksichtigen, was aber über diesen Artikel hinausgeht. Weitere Informationen sind z. B. in [7] zu finden.

Durch die Cluster-Randomisierung wird eine spezielle Datenstruktur erzeugt: Beobachtungen innerhalb der Cluster sind meist ähnlicher als Beobachtungen aus verschiedenen Clustern (Vorliegen von statistischer Abhängigkeit, siehe z. B. [11]). Je nach Stärke dieser Ähnlichkeit (gemessen über den ICC, siehe unten) kann bei konfirmatorischen cluster-randomisierten Studien die zum Erreichen einer bestimmten Power notwendige Fallzahl deutlich über der Fallzahl einer entsprechenden individuell randomisierten Studie liegen sowie bei explorativen Studien bei gegebener maximaler Fallzahl die Power geringer bzw. der zu zeigende Mindesteffekt größer sein. Aufgrund des dadurch entstehenden Effizienz- und Powerverlustes ist die effektive Fallzahl in einer cluster-randomisierten Studie (d. h. die Anzahl der wirklich statistisch unabhängigen individuellen Beobachtungen) meist niedriger als die tatsächliche Fallzahl (d. h. die Anzahl rekrutierter Studienteilnehmer). Daher sind Standardverfahren, die von der statistischen Unabhängigkeit aller Beobachtungen ausgehen, für Fallzahlplanung und Auswertung solcher Studien ungeeignet. Eine Anwendung von Standardverfahren für die Fallzahlplanung würde bspw. zu Studien mit zu geringer Power führen, in denen die Chance einen tatsächlich vorhandenen Unterschied zwischen den Studienarmen nachzuweisen, (deutlich) geringer ist als in der Planung angenommen. Dies könnte dazu führen, dass die tatsächlich vorhandene Überlegenheit einer neuen Therapie mit der Studie fälschlicherweise nicht erkannt wird.

Intraclass-Korrelationskoeffizient (ICC) und Designeffekt (Deff)

Die Ähnlichkeit von Beobachtungen innerhalb der Cluster im Vergleich zu Beobachtungen aus verschiedenen Clustern wird mit dem

Intraclass-Korrelationskoeffizienten (synonym: Intraclass-Korrelationskoeffizient; abgekürzt ICC, ρ) quantifiziert, der auf verschiedene Art definiert werden kann [12]. Seine Größenordnung ist ein Maß für die Stärke der Ähnlichkeit der Beobachtungen innerhalb der Cluster im Vergleich zur Ähnlichkeit der Beobachtungen zwischen den Clustern. Ist bspw. der ICC = 0,10, so bedeutet dies, dass 10% der Datenvarianz mit der Clusterzugehörigkeit erklärt werden kann. Ein ICC von 0 bedeutet, dass die Clusterzugehörigkeit keinen Einfluss auf die Datenvarianz hat. In der Praxis liegt die Größenordnung der meisten ICC zwischen 0,00 und 0,20, wobei eine sehr große Spannweite besteht [13].

Für a-priori-Schätzungen des ICC können z. B. Daten aus einer Pilotstudie oder der Literatur dienen. Weiterhin können unterschiedliche Berechnungsverfahren Einfluss auf den Wert des ICC haben. Eine Übersicht über wichtige Berechnungsverfahren für metrische Zielgrößen gibt [12]. Für binäre Zielgrößen sind entsprechende Methoden z. B. in [14] verfügbar.

Um bei der Fallzahlberechnung in einer cluster-randomisierten Studie die gleiche Power wie in einer individuell randomisierten Studie zu erreichen, müssen in der Regel mehr Individuen rekrutiert werden. Die notwendige Fallzahl ergibt sich aus der Fallzahl für die individuell randomisierte Studie durch Multiplikation mit dem Designeffekt (Deff), der aus dem ICC ρ und der festen Clustergröße m berechnet wird: $Deff = 1 + \rho \cdot (m - 1)$.

Das Ergebnis ist eine Gesamtfallzahl und eine sich daraus ergebende Anzahl von Clustern (mit fester Clustergröße m) für eine vorgegebene Power. Beispielsweise erhält man für einen ICC von $\rho = 0,05$ und einer Anzahl von 15 Patienten pro Cluster ($m = 15$) einen Designeffekt von $Deff = 1 + 0,05 \cdot (15 - 1) = 1,7$, d. h. die erforderliche Fallzahl für eine Studie, bei der die Individuen randomisiert zugewiesen werden (RCT), muss mit dem Faktor 1,7 multipliziert werden.

Bei ungleicher Clustergröße kann m ersetzt werden durch das arithmetische Mittel, den Variationskoeffizienten oder durch die maximale Clustergröße [12]. Die Verwendung des arithmetischen Mittels ist sinnvoll, wenn nur wenig Variabilität in der Clustergröße besteht. Sind die Clustergrößen im Vorfeld unbekannt, liegen aber Informationen zur mittleren Clustergröße und ihrer Standardabweichung vor, kann der Variationskoeffizient anstelle der festen Clustergröße verwendet werden. Die Verwendung der maximalen Clustergröße ist empfehlenswert, wenn lost to follow-up zu erwarten ist oder die Mindestgröße bei einigen Clustern nicht erreicht werden kann. Dies stellt einen konservativen Ansatz dar. Wenn es nur wenige sehr große Cluster gibt, kann es zur Überschätzung der Fallzahl kommen.

Vorgehen bei der Studienplanung

Bei der Studienplanung können 2 Vorgehensweisen unterschieden werden. Zum einen kann im Rahmen eines explorativen Ansatzes (hypothesengenerierend) für eine gegebene maximale Fallzahl bei gegebener Power und Clustergröße ein Mindesteffekt oder bei gegebenem Mindesteffekt und Clustergröße eine Power berechnet werden [15]. Dies ist insbesondere dann sinnvoll, wenn nur eine limitierte Anzahl von Beobachtungen zur Verfügung steht. Zum anderen kann ein konfirmatorischer Ansatz (statistisch überprüfend) gewählt werden: Für eine vorgegebene Power und einen vorgegebenen Mindesteffekt wird eine Fallzahl (d. h. Anzahl Studienteilnehmer und eine sich daraus er-

gebende Clusteranzahl) berechnet. Der explorative Ansatz ist als eher pragmatischer Ansatz anzusehen (Machbarkeit und Durchführbarkeit der Studie), während der konfirmatorische Ansatz eine mehr theoriegeleitete Herangehensweise darstellt.

Sollen in die Planung noch zusätzlich Kovariaten einbezogen werden, ist eine Erweiterung der Definition des ICC nach [12] möglich. Speziell für komplexe Studiendesigns (bspw. Berücksichtigung mehrerer Kovariaten im longitudinalen Design, weitere Hierarchieebenen), gibt es auch die Möglichkeit der Simulation, z. B. [16, 17]. In eine cluster-randomisierte Studie sollten nicht zu wenige Cluster eingeschlossen werden. Eine einfache Faustregel für die minimale Clusteranzahl pro Studienarm ist $n \times ICC$, wobei n die Fallzahl pro Studienarm in der individuell randomisierten Studie und ICC den Intracluster-Korrelationskoeffizienten bezeichnet [18]. Eine weitere Empfehlung ist die 30-30-Regel: Es sollten mindestens 30 Ebene-2-Einheiten untersucht werden, die jeweils wieder durch mindestens 30 Einheiten auf Ebene 1 repräsentiert sind [6].

Auswertung

Die im Rahmen der Studienplanung angewendeten Methoden sollten auch zur Auswertung genutzt werden. Zur Berücksichtigung der statistischen Abhängigkeiten innerhalb der Cluster (z. B. Abhängigkeit der Ergebnisse von Reha-Teilnehmern innerhalb derselben Klinik) muss bei der Auswertung eine Cluster-Adjustierung (Bereinigung der Ergebnisse um die Clustereffekte) durchgeführt werden [19]. Eine Analyse ohne eine solche Adjustierung kann zur Schätzung von zu kleinen Konfidenzintervallen und verringerten p-Werten führen [20], in deren Konsequenz falsch signifikante Studienergebnisse entstehen.

Da die bei cluster-randomisierten Studien entstehenden Daten ein Mehrebenen-Design besitzen, ist der Standard für die Auswertung die Mehrebenenanalyse, ein Auswertungsverfahren für Zielgrößen auf der Individuen-Ebene [6, 21, 22]. Bei diesem Verfahren werden mehrere Hierarchieebenen unterschieden, die Mitglieder der untersten Hierarchieebenen (Ebene 1, z. B. Rehabilitanden) sind dabei den Mitgliedern der höheren Hierarchieebenen (z. B. Ebene 2: Stationen, Ebene 3: Reha-Einrichtungen) eindeutig untergeordnet. Die Mehrebenenanalyse basiert auf der Anpassung sukzessive komplexer werdender hierarchischer Regressionsmodelle. Merkmale aller Hierarchieebenen können simultan zur Vorhersage eines Ebene-1-Merkmals herangezogen werden, ebenso ist die Untersuchung von Wechselwirkungen möglich. Für jedes Modell werden die spezifischen Effekte geschätzt und aus den Modellergebnissen werden darüber hinaus Informationen erhalten, welcher Varianzanteil der Zielgröße auf den unterschiedlichen Hierarchieebenen noch ungeklärt bleibt. Nur wenn ein substanzieller Varianzanteil ungeklärt bleibt, sollte das nächstkomplexere Modell angepasst werden [6]. Bei Interventionsstudien ist der Prüfung des nächstkomplexeren Modells Vorrang einzuräumen, da die Intervention eine Kovariate auf Ebene 2 (Cluster) darstellt. Die sukzessive Modellanpassung muss in diesem Fall nicht notwendigerweise eingehalten werden.

Wenn die Zielgröße hingegen auf Cluster-Ebene liegt, sollte eine Analyse auf Cluster-Ebene erfolgen [12, 20]. Ein Beispiel dafür ist eine Studie in der stationären Rehabilitation, in welcher Reha-Kliniken randomisiert werden und der Anteil der Patienten, welche die Rehabilitationsmaßnahmen erfolgreich durchlaufen haben, pro Reha-Klinik als Zielgröße untersucht wird. Dies stellt allerdings

nicht den Standard dar, da in der Regel die Zielgrößen auf Ebene 1 (Ebene der Rehabilitanden) erhoben werden. Die Analyse auf Cluster-Ebene kann als 2-stufiger Prozess angesehen werden: Zunächst wird für jedes Cluster ein Summenmaß (Durchschnittswert je Cluster) berechnet (erste Stufe), welches dann mit einem geeigneten statistischen Test verglichen wird (zweite Stufe), siehe z. B. [20]. In Studien in der Reha-Forschung können als Summenmaß bspw. anstelle der individuellen Ergebnisse der Reha-Teilnehmer die Clusterdurchschnittswerte (z. B. die Durchschnittsergebnisse in jeder Klinik) in der Analyse (z. B. gewöhnlicher Zweistichproben-t-Test) verwendet werden. Eine vereinfachte Berücksichtigung von Kovariaten ist über Regressionen möglich [20]. Die Analyse auf Cluster-Ebene ist robust (d. h. die Methode liefert stabile Ergebnisse) insbesondere bei kleiner Clusteranzahl, hat jedoch den Nachteil, dass die Variabilität innerhalb der Cluster unberücksichtigt bleibt. Ein weiterer Nachteil ist, dass im Vergleich zur Mehrebenenanalyse eine Schätzung der Varianzanteile der Zielgröße auf den unterschiedlichen Hierarchieebenen nur sehr eingeschränkt möglich ist und eine Erweiterung auf longitudinale Studiendesigns (repeated measurements) sowie Designs mit mehr als 2 Hierarchieebenen kaum realisierbar ist. Darüber hinaus ist die Mehrebenenanalyse der Standard für die Auswertung cluster-randomisierter Studien (siehe oben), sodass eine Verwendung von Analyseverfahren auf Cluster-Ebene die Ausnahme bleiben sollte [7].

Berichterstattung

Für die Berichterstattung cluster-randomisierter klinischer Studien wurde das CONSORT Statement von Campbell et al. erweitert [22]. Dieses nimmt Bezug auf die Besonderheiten einer cluster-randomisierten Studie und die Publikation einer solchen Studie sollte sich daran orientieren. So wird u. a. gefordert die Gründe für die Cluster-Randomisierung zu beschreiben, die Einheit der Randomisierung und die der Intervention zu nennen, neben der Anzahl der Individuen auch die Anzahl der Cluster und ihre Größe anzugeben, die Strukturgleichheit nicht nur auf Individuen-Ebene, sondern auch auf Cluster-Ebene zu zeigen, den ICC zu berechnen und zu berichten, die Drop-outs auf Individuen- und auf Cluster-Ebene zu analysieren sowie ein Flowchart zur Anzahl der Studienteilnehmer und Cluster im Studienablauf zu zeichnen.

Anwendungsbeispiel

Als Beispiel für die Anwendung einer cluster-randomisierten Studie in der Reha-Forschung kann die Untersuchung von Wienert et al. aus dem Institut für Sozialmedizin und Epidemiologie der Universität Lübeck dienen [23]. Untersucht wird in der Studie, ob die Teilnehmer an einer medizinisch-beruflich orientierten Rehabilitation bei Rehabilitanden mit onkologischen Erkrankungen langfristig geringere Teilhabebeschränkungen haben als Patienten im Kontrollarm, die eine herkömmliche Rehabilitation bekommen. Die Studie ist als cluster-randomisierte Studie konzipiert. Die wichtigsten Aspekte in Bezug auf Planung, Fallzahl und Auswertungsstrategie werden im Folgenden skizziert.

Das Design der Studie ist eine cluster-randomisierte Studie, in der die Patienten bezüglich der Anreiseweche randomisiert werden in einen Arm mit Standardrehabilitation bzw. medizinisch-beruflich orientierter Rehabilitation. Die Studie wird in vier Kliniken

durchgeführt, so sodass die Randomisierung stratifiziert bzgl. der Klinik durchgeführt wurde.

Da die Randomisierung ein wesentlicher Aspekt des Studiendesigns ist, werden hier die wichtigsten Informationen detailliert dargestellt, da dies die wesentliche Begründung für die Wahl eines cluster-randomisierten Studiendesigns darstellt ([23], S. 6): „The study is designed as a cluster-randomized multicenter trial. Patients are recruited from 4 inpatient rehabilitation centers. Patients of a rehabilitation center who start their rehabilitation in the same week represent a cluster. These patients jointly receive either the intervention or the control treatment. Cluster-randomization was chosen to avoid spillover effects between therapy groups. Each of the 4 rehabilitation centers receives its own randomization schedule. The randomization lists are created by the last author using computer-generated random numbers and blocks of 4. The administrative employees who assign the patients to their date of rehabilitation are blinded to the randomization schedule“. Mit “blocks of 4” ist hier gemeint, dass in jedem Zentrum nach Randomisierung von jeweils vier Gruppen (Clustern) dieselbe Gruppenanzahl in Interventions- und Kontrollarm erreicht war (Blockrandomisierung zum besseren Erreichen der Balanciertheit).

Die Fallzahlabschätzung nutzt viele der Argumente aus dem Abschnitt Fallzahlplanung. Basis der Fallzahlplanung ist die Angabe der zu entdeckenden Effektgröße (standardized mean difference) von $d = 0,3$, einer eher kleinen bis mittleren Effektgröße nach [24]. Die Fallzahl für eine individuell randomisierte Studie kann hier wegen des Vorliegens normalverteilter Daten für den t-Test bei einem 2-seitigen Signifikanzniveau von 5 % und einer Power von 80 % mit insgesamt 352 Patienten (2×176) berechnet werden. Die notwendige Erhöhung der Fallzahl wird in dem Artikel über ein mögliches Drop-Out von etwa 30 % nach einem Jahr begründet. Ein ICC und/oder ein Designeffekt werden allerdings nicht angegeben. Der ICC müsste hier sehr klein sein ($< 0,01$), damit der Designeffekt die hier angegebene Fallzahlerhöhung von 352 auf 504 ($Deff = 1,432$) ergibt. Nach [13] ist in allgemeinmedizinischen (und vermutlich auch rehabilitationswissenschaftlichen) Studien der ICC relativ klein (ca. 0,01), sodass hier von einem solchen ICC ausgegangen wird. Bei vier Zentren und jeweils einer Clustergröße m von maximal 20 Patienten (Schätzung aus eigenen Erfahrungen) ist der Designeffekt $Deff = 1 + ICC(m-1) = 1 + 0,01(20-1) = 1,19$, sodass eine Gesamtfallzahl von $352 \times 1,19 = 419$ Patienten notwendig scheint. Bei einer Drop-Out-Rate von 30 % ergibt sich nach diesen Überlegungen hier mit 598 Patienten somit eine etwas höhere notwendige Gesamtfallzahl für die geplante Studie.

Die Beschreibung der Auswertungsstrategie für die Hauptfragestellung wird folgendermaßen angegeben, wobei der Schwerpunkt auf Regressionsanalysen mit Berücksichtigung des Cluster-Effekts über die Nutzung gemischter Regressionsmodelle liegt ([23], S. 8): “Treatment effects are tested using regression analyses. In the case of binary study outcomes, logistic regression will be used (e. g., effects on employment status or sick leave at follow-up). For analysis of time to return to work, we will use proportional hazard models. Baseline parameters will be considered as covariates. Due to the cluster-randomization, the induced correlation of error terms for patients with the same arrival week will be taken into account by a random effect. Rehabilitation centers will be considered using a fixed effect“. Gemischte Regressionsmodelle sind

eine Form der Mehrebenenanalyse und entsprechen damit der oben genannten Empfehlung für die Auswertung cluster-randomisierter Studien. Der Begriff ‚fester Effekt‘ (‘fixed effect’) wird für Einflussgrößen verwendet, deren Ausprägungen nicht als Zufallsauswahl aus einer großen Population angesehen werden, sondern wenn nur wenige, fest vorgegebene Ausprägungen untersucht werden. Für die einzelnen Ausprägungen werden im Regressionsmodell Effekte geschätzt, sodass sich die Interpretation nur auf die untersuchten Ausprägungen bezieht. Eine generalisierende Aussage für die weiteren Ausprägungen eines festen Effektes ist damit nicht möglich. Der Begriff ‚zufälliger Effekt‘ (‘random effect’) bezeichnet Einflussgrößen, bei welchen aus allen möglichen Ausprägungen nur einige wenige untersucht werden konnten. Im Gegensatz zu den festen Effekten wird für einen zufälligen Effekt im Regressionsmodell die Varianz geschätzt und keine Effekte für die einzelnen Ausprägungen. Somit ist es nicht möglich, eine Interpretation für einzelne Ausprägungen eines zufälligen Effektes vorzunehmen.

Die Ergebnisse der Studie zeigen, dass mittelfristig (d. h. 3 Monate nach Rehabilitationsende) kein Effekt im primären Zielkriterium bestätigt werden konnte [25]. Bei Vorliegen eines hohen Frühberentungsrisikos zeigten sich Effekte zugunsten der MBOR, auch für das primäre Zielkriterium. Die sekundären Endpunkte körperliche Funktionsfähigkeit, Schmerzsymptomatik und Fatigue-Symptomatik wurden durch die medizinisch-beruflich orientierte Rehabilitation positiv beeinflusst. Langfristige Ergebnisse (12 Monate nach Rehabilitationsende) stehen derzeit noch nicht zur Verfügung.

Insgesamt zeigt sich, dass die Planung und Durchführung einer cluster-randomisierten Studie kaum mehr Aufwand bedeutet als eine individuell randomisierte Studie, die hier bei gleichzeitiger Aufnahme der Patienten in den Kliniken in beide Arme sicher Kontaminationsprobleme zeigen würde. Weitere Beispiele mit praktischen Darstellungen von Planung und Auswertung cluster-randomisierter Studien sind in [26], [19] und [27] enthalten.

KERNBOTSCHAFT

In der Reha-Forschung gibt es einige besondere Bedingungen zu beachten (vgl. Abschnitt Bedingungen und Anforderungen an Studien in der Rehabilitationsforschung), welche die Planung, Durchführung und Auswertung beeinflussen. Bei prospektiven 2-armigen (Interventions-)Studien können auch hier cluster-randomisierte Designs genutzt werden, insbesondere wenn schon natürliche Cluster (Therapie erfolgt oft in Gruppen) vorliegen oder eine Verringerung des Kontaminationsbias angestrebt wird. Bei der Studienplanung ist bei Fallzahl-/Powerberechnungen die Clusterstruktur zu berücksichtigen, da Ergebnisse von Studienteilnehmern innerhalb der Cluster ähnlicher sind als Ergebnisse von Studienteilnehmern verschiedener Cluster. Aufgrund unserer Erfahrungen und der hier genannten Argumente empfehlen wir bei prospektiven vergleichenden Studien in der Reha-Forschung eine adäquate Randomisierung, um gute und überzeugende Ergebnisse zu erzielen. Insbesondere die Cluster-Randomisierung kann hierbei ein entscheidender Baustein sein, der bei Studien im Bereich der Reha-Forschung noch verstärkt genutzt werden kann.

Interessenkonflikt

Die Autoren geben an, dass kein Interessenkonflikt besteht.

Literatur

- [1] Muche R, Rohlmann F, Büchele G et al. Randomisierung in klinischen Studien in der Rehabilitationsforschung: Grundlagen und praktische Aspekte. *Rehabilitation* 2002; 41: 311–319. doi:10.1055/s-2002-34568
- [2] Wirtz MA. Konfundierungen und Störfaktoren in rehabilitationswissenschaftlichen Beobachtungs- und Interventionsstudien. *Rehabilitation* 2017, doi:10.1055/s-0043-120911
- [3] Pinnock H, Barwick M, Carpenter CR. Standards for Reporting Implementation Studies (StaRI) Statement. *BMJ* 2017; 356: i6795 doi:10.1136/bmj.i6795
- [4] Gaus W, Muche R. *Medizinische Statistik*. 2. Aufl. Stuttgart: Schattauer; 2017
- [5] Deutsche Rentenversicherung: Klassifikation therapeutischer Leistungen (KTL) (2015). Im Internet: ktl-driv.de; Stand: 26.03.2018
- [6] Wirtz MA. Die Mehrebenenanalyse als Verfahren zur Analyse rehabilitationswissenschaftlicher Forschungsfragen. *Rehabilitation* 2018, doi:10.1055/s-0043-124334
- [7] Hox JJ, Meerbeek M, van de Schoot R. *Multilevel analysis. Techniques and applications*. 3rd edition Oxford/UK: Routledge; 2018
- [8] Exner AK, Gerold H, Breckenkamp J et al. Datenschutz in der Rehabilitationsforschung. *Rehabilitation* 2014; 53: 258–267. doi:10.1055/s-0033-1357115
- [9] Faller H. Intention-to-treat. *Rehabilitation* 2004; 43: 52–55. doi:10.1055/s-2004-818553
- [10] Kutschmann M, Bender R, Grouven U et al. Aspekte der Fallzahlkalkulation und Powerberechnung anhand von Beispielen aus der rehabilitationswissenschaftlichen Forschung. *Rehabilitation* 2006; 45: 377–384. doi:10.1055/s-2006-940113
- [11] Kuß O, Watzke S. Korrekter Umgang mit korrelierten Daten in der Rehabilitationsforschung. *Rehabilitation* 2005; 44: 367–372. doi:10.1055/s-2005-86706
- [12] Eldridge SM, Kerry S. *A Practical Guide to Cluster Randomised Trials in Health Services Research*. John Wiley & Sons; 2012 doi:10.1002/9781119966241
- [13] Adams G, Gulliford MC, Ukoumunne OC et al. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004; 57: 785–794. doi:10.1016/j.jclinepi.2003.12.013
- [14] Ridout MS, Demétrio CG, Firth D. Estimating intraclass correlation for binary data. *Biometrics* 1999; 55: 137–148
- [15] Hemming K, Girling AJ, Sitch AJ et al. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. *BMC Med Res Methodol* 2011; 11: 102. doi:10.1186/1471-2288-11-102
- [16] Dreyhaupt J. Instrumente für Power- und Fallzahlberechnungen bei komplexen hierarchischen Studiendesigns in der Versorgungsforschung. *Monitor Versorgungsforschung* 2015; 6: 49–54
- [17] Dreyhaupt J. Generelle Fallzahl- und Powerabschätzung über Simulation bei Studien mit komplexen hierarchischen Daten als Unterstützung der Studienplanung in der Versorgungsforschung. Abschlussbericht des gleichnamigen Projekts bei der Nachwuchsakademie Versorgungsforschung Baden-Württemberg. 2015; Im Internet: http://vts.uni-ulm.de/query/longview.meta.asp?document_id=9509
- [18] Hemming K, Eldridge S, Forbes G et al. How to design efficient cluster randomised trials. *BMJ* 2017; 358: j3064. doi:10.1136/bmj.j3064
- [19] Chenot JF. Cluster-randomisierte Studien: eine wichtige Methode in der allgemeinmedizinischen Forschung. *Z. Evid.Fortbild.Qual.Gesundh.wesen* 2009; 103: 475–480. doi:10.1016/j.zefq.2009.07.004
- [20] Donner A, Klar N. *Design and Analysis of Cluster Randomization trials in Health Research*. John Wiley & Sons; 2010
- [21] Farin E. Die Anwendung Hierarchischer Linearer Modelle für Einrichtungsvergleiche in der Qualitätssicherung und Rehabilitationsforschung. *Rehabilitation* 2005; 44: 157–164
- [22] Campbell MK, Piaggio G, Elbourne DR et al. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012; 345: e5661. doi:10.1136/bmj.e5661
- [23] Wienert J, Schwarz B, Bethge M. Effectiveness of work-related medical rehabilitation in cancer patients: study protocol of a cluster-randomized multicenter trial. *BMC Cancer* 2016; 16: 544. doi:10.1186/s12885-016-2563-z
- [24] Cohen J. *Statistical power analysis for the behavioral sciences*. second edition Hillsdale, NJ: Lawrence Erlbaum Associates; 1988 doi:10.1016/C2013-0-10517-X
- [25] Fauser D, Wienert J, Bethge M. Medizinisch-beruflich orientierte Rehabilitation für onkologische Rehabilitanden – mittelfristige Ergebnisse einer clusterrandomisierten Multicenterstudie. *DRV-Schriften Bd. 113* 2018; 387–388
- [26] Kuß O, Jahn P, Renz P et al. Cluster-randomisierte Studien in der Pflegewissenschaft. *Hallesche Beiträge zu den Gesundheits- und Pflegewissenschaften* 2009; 8: 302–310
- [27] Bethge M, Herbold D, Trowitzsch L et al. Work status and health-related quality of life following multimodal work hardening: a cluster randomised trial. *J Back Musculoskelet Rehabil* 2011; 24/3: 161–172. doi:10.3233/BMR-2011-0290