

Zusammenfassung

Eine Reihe von Studienanlagen in den Rehabilitationswissenschaften liefern korrelierte Messwerte, z.B. wenn Patienten im Zeitverlauf an verschiedenen Zeitpunkten beobachtet werden, wenn pro Patient mehrere Zielgrößen gemessen werden oder wenn ein Patient unter verschiedenen experimentellen Bedingungen beobachtet wird. Die meisten grundlegenden statistischen Verfahren verlangen jedoch unabhängige Beobachtungen, und ein kritikloses Anwenden dieser Methoden bei korrelierten Messwerten kann unter Umständen zu falschen Schlüssen führen. Anhand eines einfachen Beispiels wird gezeigt, dass es sich (im Sinne eines Powergewinns) durchaus lohnen kann, korrelierte Daten tatsächlich auch mit Methoden für korrelierte Daten auszuwerten. Im Folgenden werden verschiedene Auswertungsmethoden (Reduktion auf eine einzige Beobachtung, ANOVA mit Messwiederholungen, MANOVA und gemischte Modelle) für korrelierte Daten dargestellt. Unter diesen sind die gemischten Modelle die Methode der Wahl, da diese eine sehr flexible Modellierung der Korrelationsstruktur der Beobachtungen erlauben und inzwischen auch mit Standardsoftware schätzbar sind.

Abstract

Many study designs in rehabilitation science give rise to correlated data. For example, patients are followed over time, different responses are measured for each patient, or patients are observed in logical units. Standard statistical methods, however, are only valid for independent responses, and careless application of these methods for actually correlated observations might give erroneous results. By means of a simple example, we show how using methods for correlated data can indeed give a gain in statistical power. In the following, different approaches (Summary measures, Repeated Measurement ANOVA, MANOVA, and Mixed Models) to deal with correlated data are presented. We conclude that among these, the Mixed Models approach is the method of choice because it allows flexible modelling of correlation structure and is, meanwhile, also available in standard statistical software packages.

Koordinatoren der Reihe „Methoden in der Rehabilitationsforschung“:

Prof. Dr. Dr. Hermann Faller, Würzburg; Prof. Dr. Thomas Kohlmann, Greifswald;
Dr. Christian Zwingmann, Berlin
Interessenten, die einen Beitrag zur Reihe beisteuern möchten, werden gebeten,
vorab Kontakt aufzunehmen, E-mail: christian.zwingmann@web.de

Institutsangaben

¹ Institut für Medizinische Epidemiologie, Biometrie und Informatik
der Martin-Luther-Universität Halle-Wittenberg
² Klinik und Poliklinik für Psychiatrie und Psychotherapie
der Martin-Luther-Universität Halle-Wittenberg

Korrespondenzadresse

Oliver Kuß · Institut für Medizinische Epidemiologie, Biometrie und Informatik der Martin-Luther-Universität
Halle-Wittenberg · Magdeburger Straße 27 · 06097 Halle (Saale), · E-mail: Oliver.Kuss@medizin.uni-halle.de

Bibliografie

Rehabilitation 2005; 44: 367–372 © Georg Thieme Verlag KG Stuttgart · New York
DOI 10.1055/s-2005-867062
ISSN 0034-3536

Schlüsselwörter

Korrelierte Messungen · Messwiederholungen · gemischte Modelle · MANOVA · ANOVA

Key words

Correlated data · repeated measurements · mixed model · MANOVA · ANOVA

Wie entstehen korrelierte Daten?

Eine Reihe von Studienanlagen in den Rehabilitationswissenschaften führen zu korrelierten oder abhängigen Messwerten:

- Patienten werden im Zeitverlauf beobachtet und Messungen werden an mehreren Zeitpunkten gemacht (z. B. Erhebung des SF-36 – Fragebogen zum Gesundheitszustand – vor Beginn, während und nach Abschluss der Rehabilitation), Untersuchungen dieser Art werden auch als Messwiederholungsdesigns bezeichnet.
- Patienten werden wiederholt unter verschiedenen experimentellen Bedingungen behandelt (z. B. Messung der Beweglichkeit nach Kältebehandlung und Messung der Beweglichkeit nach biomechanischer Stimulation bei demselben Patienten).
- Es werden mehrere Zielgrößen pro Patient beobachtet (z. B. Messung des Reha-Erfolges mithilfe eines Funktionalitätsindex, mit einem Lebensqualitätsfragebogen und mit einem physiologischen Indikator).
- Patienten werden in logischen Einheiten (Clustern) beobachtet, z. B. als Angehörige von Familien oder als Patienten verschiedener Kliniken.
- Es werden zwei oder mehr Personengruppen untersucht, die anhand wichtiger Merkmale (z. B. Alter, Geschlecht, Diagnose) auf Stichprobenebene äquivalent sind. In einem solchen Fall wird der Begriff „parallelierte Stichproben“ verwendet. Wird jedem Probanden der einen Stichprobe ein „Paarling“ der anderen Stichprobe anhand wichtiger Merkmale (z. B. Zwillingspaare, Ehepartner) zugeordnet, wird von „matched samples“ gesprochen.

Meistens findet man sogar mehrere dieser Phänomene gleichzeitig und sich überlagernd in einer Studie, also z. B. Erhebung von mehreren Zielgrößen pro Patient an mehreren Punkten im Zeitverlauf.

Die Standardannahme in den meisten grundlegenden statistischen Auswertungsverfahren (z. B. Chi²-Test, lineare Regression) ist jedoch die Unabhängigkeit der Beobachtungen, und eine Analyse von korrelierten Messungen mit den Standardmethoden für unkorrelierte Beobachtungen führt unter Umständen zu falschen Schlüssen.

Es lohnt sich, korrelierte Daten als solche auszuwerten!

Das folgende kleine Beispiel soll zeigen, dass es sich durchaus lohnt, beim Vorliegen von korrelierten Messungen dieses auch in der statistischen Auswertung zu berücksichtigen. Wir nehmen an, mit einer Gruppe von 50 adipösen Kindern wird ein 4-wöchiges Sportprogramm zur Gewichtsreduktion durchgeführt, das Gewicht der Kinder wird vor Beginn und zum Ende des Sportprogrammes gemessen. Das Ausgangsgewicht der Kinder soll normal verteilt sein mit Erwartungswert 80 kg und Varianz 25 (= 5²),

sodass gemäß der 2σ(Sigma)-Regel ca. 95% der Kinder ein Ausgangsgewicht zwischen 70 und 90 kg haben. Als Ergebnis der Studie zeigt sich, dass die Kinder im Mittel 2 kg abgenommen haben – genauer: die Gewichtsabnahme soll normal verteilt mit Erwartungswert 2 kg und Varianz 1 sein; die Gewichtsabnahme wird dabei als unabhängig vom Ausgangsgewicht angenommen. Es ist klar, dass die methodisch saubere Evaluierung eines Sportprogramms eine Kontrollgruppe mitführen müsste, vorzugsweise zusätzlich eine randomisierte Zuteilung der Kinder zu den Gruppen – dieser Einwand sei für den Moment jedoch vernachlässigt.

Führen wir eine Analyse durch, die *nicht* berücksichtigt, dass die Messwerte der beiden Gruppen („Gewicht in kg vor dem Sportprogramm“ und „Gewicht in kg nach dem Sportprogramm“) korreliert sind, da jedes Kind Gewichtswerte in beiden Gruppen beisteuert, dann ist angesichts der Intervallskaliertheit der Zielgröße ein unverbundener t-Test (oder: t-Test für unabhängige Stichproben) die nahe liegende Auswertungsmethode. Die Formel zur Berechnung der Teststatistik des t-Tests t_{uv} sieht dann, etwas vereinfacht, so aus:

$$t_{uv} = \frac{\text{Differenz der Mittelwerte}}{\text{mittlere Standardabweichung}} \times f_{uv}(N)$$

d. h., die Differenz der Mittelwerte in beiden Gruppen, als nahe liegender Schätzer für den Effekt des Sportprogramms, wird an einer mittleren Standardabweichung der Gewichtswerte in beiden Gruppen standardisiert, zusätzlich ist der Wert der Teststatistik noch eine Funktion – $f_{uv}(N)$ – der Fallzahl. In unserem konkreten Beispiel ergibt sich als Wert der Teststatistik $t_{uv} = 2,00$, was bei einer Anzahl von 98 Freiheitsgraden einem p-Wert von $p = 0,048$ entspricht.

Eine korrekte Analyse der vorliegenden Daten, die berücksichtigt, dass eigentlich wiederholte Messungen vorliegen, würde naheliegenderweise einen t-Test für verbundene (oder: abhängige) Stichproben heranziehen. Die Teststatistik des t-Tests für verbundene Stichproben t_v sieht dann, ebenfalls etwas vereinfacht, so aus:

$$t_v = \frac{\text{Mittelwert der Differenzen}}{\text{Standardabweichung der Differenzen}} \times f_v(N)$$

Der Effekt des Sportprogramms wird nicht mehr als Differenz der Mittelwerte in der „Vorher“- und „Nachher“-Gruppe gemessen, sondern jetzt wird, für jedes Kind einzeln, zuerst eine Differenz zwischen „Vorher“- und „Nachher“-Wert berechnet. Der Mittelwert über diese Differenzen stellt dann einen Schätzer für den Effekt des Sportprogramms dar. Noch entscheidender in unserem Beispiel aber ist der Nenner der Teststatistik. Hier wird jetzt an der Standardabweichung der gemessenen Differenzen standardisiert, und diese ist in unserem Beispiel viel kleiner als im unverbundenen Fall, weil die Streuung der Ausgangsgewichte

nicht mit berücksichtigt werden muss. In unserem Beispiel ergibt sich ein Wert der Teststatistik von $t_v = 14,24$, was bei einer Anzahl von 49 Freiheitsgraden zu einem höchstsignifikanten p-Wert führt (der, um ganz genau zu sein, gleich $4,7 \times 10^{-19}$ ist).

Zugegebenermaßen handelt es sich hier um ein etwas konstruiertes Beispiel, aber das Prinzip der statistischen Verfahren für korrelierte Messungen wird klar: Jeder Patient ist seine eigene Kontrolle, und die zusätzliche Variabilität, die unabhängige Kontrollen mit sich bringen würden, muss nicht mehr berücksichtigt werden.

Es soll nicht verschwiegen werden, dass die Anwendung von Auswertungsmethoden für korrelierte Daten jedoch nicht immer notwendigerweise zu einem Powergewinn führen muss. Abhängig vom Design der Studie kann es durchaus auch zu einem Rückgang der Power kommen. Zu einem Powergewinn kommt es immer dann, wenn Effekte gemessen werden sollen, die sich innerhalb einer Beobachtung ändern, so wie in unserem Beispiel die Gewichtsveränderung zwischen zwei Zeitpunkten *innerhalb* der Kinder. Würden wir die Fragestellung etwas erweitern und uns für einen Faktor interessieren, der sich *zwischen* den Kinder unterscheidet, z. B. das Geschlecht, so hätten wir zur Prüfung dieses Einflusses weniger Power zur Verfügung.

Dieses Problem ist v. a. in so genannten „clusterrandomisierten“ Interventionsstudien relevant, in denen die Intervention auf Gruppen (Cluster) von Probanden randomisiert wird und davon auszugehen ist, dass die Beobachtungen der Probanden innerhalb eines Clusters korreliert sind. Die Beurteilung des Interventionseffekts bezieht sich hier also auf einen Faktor, der sich *zwischen* den Clustern (im obigen Beispiel wären das die Kinder!) ändert, sodass wir hier weniger Power haben.

Wie berücksichtigt man korrelierte Daten?

Reduktion der abhängigen Beobachtungen auf eine einzelne Beobachtung

Der einfachste Zugang zur Modellierung von korrelierten Daten ist der, diese mithilfe eines zusammenfassenden Maßes auf einen einzelnen Wert pro Patient zu reduzieren [1]. So wird Unabhängigkeit erreicht, und die Analyse kann mit diesem zusammengefassten Wert fortgeführt werden. Beispiele dafür sind für longitudinale Daten die Anpassung einer Regressionsgeraden durch mehrere Punkte im Zeitverlauf und ein Weiterrechnen mit der Steigung dieser Regressionsgeraden oder auch die Berechnung einer AUC („area under the curve“, [1]). Für multivariate Zielgrößen ist die Reduktion auf einen Score denkbar; z. B. könnte dieser dadurch berechnet werden, dass gezählt wird, wie viele Zielgrößen einen vorher pro Zielgröße spezifizierten kritischen Wert überschreiten.

Die Vorteile dieser Methode sind klar. Es findet eine Komplexitätsreduktion statt, sowohl bezüglich der Interpretation der Daten als auch der weiteren statistischen Analyse. Des Weiteren sind diese Verfahren bis zu einem gewissen Grade robust gegen fehlende Werte, z. B. wird es bei der Berechnung einer Regressionsgeraden durch mehrere Messwerte im Zeitverlauf wenig

ausmachen, wenn der Patient an einem oder zwei Zeitpunkten nicht beobachtet werden konnte.

Diesen Vorteilen stehen jedoch auch eine Reihe von Nachteilen gegenüber, die letztendlich zu der Empfehlung führen, solche zusammenfassende Methoden nur dann zu verwenden, wenn sie inhaltlich begründet und vor allem auch vorher spezifiziert worden sind. Der empfindlichste Nachteil ist sicherlich der, dass eine Komplexitätsreduktion in der Regel auch zu einem Informationsverlust führt und unter Umständen wichtige Einzelheiten verloren gehen. Ein weiterer Nachteil der Reduktion auf zusammenfassende Maße ist der, dass die Unsicherheit bei der Schätzung des zusammenfassenden Parameters im weiteren Verlauf der Analyse im Allgemeinen nicht berücksichtigt wird. Stattdessen wird angenommen, dass der reduzierte Parameter tatsächlich gemessen (und eben nicht geschätzt) worden ist. Dies führt dazu, dass weiterführende statistische Analysen mit dem reduzierten Parameter insgesamt zu liberal und Effekte überschätzt werden.

Eine letzte Gefahr besteht schließlich darin, dass die Fülle der möglichen zusammenfassenden Maße dazu verleiten kann, so lange alle denkbaren Parameter „auszuprobieren“, bis schließlich einer gefunden ist, der im weiteren Verlauf der Analyse die gewünschten signifikanten Effekte zeigt. Dies ist aus biometrischer Sicht natürlich aufs Schärfste zu verurteilen; wenn tatsächlich zusammenfassende Maße berechnet werden sollen, so sollten diese explizit *vor Beginn der Studie* im Protokoll dargelegt worden sein.

Varianzanalyse mit Messwiederholungen

Diese Methode ist die Standardmethode im Umgang mit wiederholten und korrelierten Messungen (siehe z. B. [2]). Sie beruht auf dem allgemeinen Prinzip der Varianzanalyse, nämlich einer Zerlegung der Variabilität der Messwerte auf die einzelnen Faktoren, bezüglich derer sich die Messwerte kategorisieren lassen. Im einfachsten Fall, in Messwiederholungsdesigns mit zwei Gruppen, also bei der Erhebung von mehreren Messwerten beim gleichen Patienten an verschiedenen Zeitpunkten, wird das zum einen der Zeitfaktor (Within-Subject-Faktor), zum anderen der Behandlungsfaktor (Between-Subject-Faktor) sein. Der Behandlungsfaktor wird in der Regel derjenige sein, für den man sich interessiert, z. B. der Unterschied zwischen Kontroll- und Interventionsgruppe. Eine Zerlegung der Varianz auf die beiden Faktoren bedeutet dann, dass der Behandlungseffekt gemessen werden kann, während gleichzeitig hinsichtlich des Zeiteffekts adjustiert wird. Das bedeutet, dass ein Einfluss der Korrelation der Messungen pro Patient über die Zeit ausgeschaltet wird.

Der große Nachteil dieser Methode besteht in den sehr restriktiven Anforderungen an die Korrelationsmatrix der wiederholten Messungen. Diese Matrizen müssen nämlich die so genannte Sphäritäts- oder Zirkularitätsannahme erfüllen, die im Falle von wiederholten Messungen besagt, dass die Messwerte an jedem beliebigen Paar von Zeitpunkten identisch korreliert sein müssen. Von Messungen im Zeitverlauf wird jedoch erwartet, dass auf der Zeitachse weiter auseinander liegende Messwerte schwächer korreliert sind als näher zusammen liegende, was der Annahme der Sphärität entgegensteht. Teilweise Abhilfe

können Korrekturfaktoren (Huynh-Feldt, Greenhouse-Geisser) schaffen, mit denen die entscheidenden F-Tests standardisiert werden. Bei konkreten inhaltlichen Vorstellungen über die zugrunde liegende Korrelation wird man jedoch auch daran interessiert sein, gerade diese zu modellieren und eventuell sogar Parameter zu schätzen, also z. B. die Stärke der Abnahme der Korrelation bei ansteigender Entfernung in der Zeit.

Von statistischen Tests, die das Vorliegen der Sphäritätsannahme prüfen, ist in der Regel wenig zu erwarten. Diese haben bei kleinen Fallzahlen eine geringe Power, Nichtsphärität zu entdecken. Bei großen Abweichungen dagegen werden auch vernachlässigbare Abweichungen als statistisch signifikant deklariert [3].

Die Varianzanalyse mit Messwiederholungen scheint immer noch das beliebteste Verfahren zur Analyse von korrelierten Beobachtungen zu sein, v. a. bei der Analyse von Zwei-Gruppen-Vergleichen zu zwei Zeitpunkten, einem Studiendesign, das in der Rehabilitationsforschung häufig zu finden ist. Dann wird in der Regel ein Modell mit sowohl den Haupteffekten (Zeitpunkt, Behandlung) als auch deren Interaktion verwendet. Dadurch zielt man auf eine Adjustierung des Effektes nach den Baseline-Werten ab, sodass auch bei etwaigen Imbalancen der Baseline-Werte der Behandlungseffekt (hier als die Interaktion zwischen Behandlung und Zeitpunkt) valide beurteilt werden kann. Es sei aber darauf hingewiesen, dass diese Art der Herangehensweise nicht die ideale Analysemethode für besagte Designs darstellt. Hier sollte man eher zur Verwendung von Kovarianzanalysemodellen (ANCOVA) übergehen, da diese ebenfalls für Baseline-Werte adjustieren, aber daneben eine größere statistische Power haben [4].

Multivariate Varianzanalyse (MANOVA)

In der multivariaten Varianzanalyse (siehe z. B. [5]) werden die verschiedenen Beobachtungen bezüglich (innerhalb) eines Patienten als verschiedene *abhängige* Variablen angesehen und dementsprechend modelliert. Dies hat den Vorteil, dass keine Anforderungen mehr an die Korrelation der verschiedenen Zielgrößen gestellt wird (also insbesondere auch nicht die Sphärität), andererseits kann die Korrelation auch hier nicht inhaltlich begründet modelliert werden. Des Weiteren sind MANOVA-Modelle eher instabile Modelle, das bedeutet, die Anzahl der Patienten sollte „um einiges“ größer sein als die Anzahl der abhängigen Variablen. Dies ist auf die große Anzahl an Parametern zurückzuführen, die in der unstrukturierten Kovarianzmatrix geschätzt werden müssen. Konkret bedeutet dies, dass in einer Studie mit zehn Patienten (z. B. fünf in der Therapie- und fünf in der Kontrollgruppe) und zehn Messungen einer Zielgröße im Zeitverlauf, die sich als MANOVA-Modell mit zehn Zielgrößen und zehn Beobachtungen auffassen lässt, 56 Parameter (einer für den Therapieeffekt, zehn für die Varianzen bezüglich der Zeitpunkte und 45 für die Kovarianzen bezüglich der Zeitpunkte) geschätzt werden müssen und daher keine sehr stabile und valide Schätzung des Therapieeffektes erwartet werden kann.

Gemischte Modelle

Die gemischten Modelle (auch hierarchische Modelle oder Modelle mit zufälligen Effekten, siehe z. B. [6]) stellen eine Erweiterung des allgemeinen linearen Modells dahingehend dar, dass in

die Modellgleichung auch zufällige Effekte aufgenommen werden können, wobei „zufällig“ bedeutet, dass der Parameter (also das β -Gewicht) des jeweiligen zufälligen Effektes keinen festen Wert hat, sondern einer Verteilung folgt. Im einfachsten Fall, der in der Praxis jedoch häufig bereits ausreichend ist, wird nur der konstante Faktor in der Modellgleichung als zufälliger Effekt angenommen. Dadurch wird zum einen die sehr plausible Annahme ins Modell aufgenommen, dass alle Patienten eine zufällige Auswahl aus einer größeren Population von Patienten darstellen. Zum anderen, und das macht diese Modellklasse für korrelierte Messwerte interessant, wird dabei die Korrelation der Messwerte innerhalb der Patienten explizit mitmodelliert.

Wir wollen zur Erläuterung der Klasse der gemischten Modelle auf unser obiges Beispiel zurückgreifen. Die Wirksamkeit des Diätprogramms für adipöse Kinder soll in einer randomisierten Studie gemessen werden. Dazu wird den Kindern neben den allgemeinen Rehabilitationsmaßnahmen in der Klinik ein umfangreiches Programm an ernährungswissenschaftlichen Lehrveranstaltungen und Kochkursen angeboten, eine Kontrollgruppe erhält nur die rehabilitativen Standardmaßnahmen. Der Erfolg des Programms soll anhand der Bodymass-Index(BMI)-Veränderung sechs Monate nach Ende der Rehabilitation gemessen werden. Nun ist es nicht zweckmäßig, in einer einzelnen Klinik Intervention und Kontrolle gleichzeitig anzuwenden, denn es ist anzunehmen, dass sich die Kinder aus der Interventionsgruppe mit den Kontrollkindern über das Gelernte austauschen. Dadurch würde eine „Verschmutzung“ der Kontrollgruppe mit dem Interventionseffekt stattfinden, und der eigentliche Effekt des Programms könnte nicht mehr valide gemessen werden. Des Weiteren werden die Mitarbeiter der Klinik, die das neue Programm durchführen, vorher geschult. Nun ist aber anzunehmen, dass diese geschulten Mitarbeiter das neu erworbene Wissen bei der Durchführung des Programms möglicherweise auch auf die Kinder aus der Kontrollgruppe anwenden, sodass auf der Ebene der Mitarbeiter ebenfalls keine klare Trennung zwischen Interventions- und Kontrollgruppe gewährleistet ist.

Die Lösung dieses Problems ist, pro Klinik entweder nur Intervention oder nur Kontrolle durchzuführen. Studien mit diesem Design werden als „clusterrandomisiert“ (s. o.) bezeichnet, weil die Intervention bzw. Kontrolle auf ganze Cluster (hier: die Kliniken) randomisiert wird, die Zielgröße aber an Objekten innerhalb der Cluster (hier: die Kinder in den Kliniken) gemessen wird. Nun ist davon auszugehen, dass die Ergebnisse der Kinder innerhalb einer Klinik korreliert sein werden, weil sich möglicherweise die allgemeinen Rehabilitationsprogramme innerhalb der Kliniken unterscheiden oder die Kliniken unterschiedliche Einzugsgebiete und damit unterschiedliches Patientengut haben. Diese Korrelation muss bei der Auswertung berücksichtigt werden, und die gemischten Modelle bieten hier ein ideales Werkzeug. Die Modellgleichung des zugehörigen gemischten Modells lautet:

$$Y_{ij} = \alpha_i + \beta x + \varepsilon_{ij},$$

wobei Y_{ij} die Zielgröße (also die Veränderung des BMI nach sechs Monaten) des j -ten Kindes in der i -ten Klinik und x den Interventionseffekt bezeichnet, dessen Einfluss mit Hilfe von β gemessen wird. ε_{ij} bezeichnet den residualen Fehler bezüglich des j -ten

Kindes in der i -ten Klinik und wird als normal verteilt angenommen – $\varepsilon_{ij} \sim N(0, \sigma^2)$. Der Interventionseffekt x wird, wie im herkömmlichen allgemeinen linearen Modell, als fester Effekt betrachtet. Der konstante Faktor α_i (auch als „Intercept“ bezeichnet) dagegen wird als zufälliger Faktor angesehen. Dabei kennzeichnet der Index i des zufälligen Faktors α_i die Zugehörigkeit des Parameters zu der jeweiligen Untersuchungseinheit, im Beispiel also der jeweiligen Klinik. Dieser Faktor kann nun genutzt werden, um die unterschiedlichen Ausgangslagen der einbezogenen Kliniken zu modellieren. Das heißt, jeder einzelnen Klinik wird ein eigener, klinikspezifischer konstanter Faktor zugeordnet, der die klinikspezifische Veränderung des BMI unabhängig vom Interventionsprogramm misst. Dieser klinikspezifische konstante Faktor wird im Modell jedoch nicht für jede einzelne Klinik geschätzt, sondern als normal verteilt um einen unbekannt konstanten Faktor α angenommen – $\alpha_i \sim N(\alpha, v^2)$ –, so dass also im Modell nicht i (für jede Klinik einen), sondern nur zwei (α und v^2) zusätzliche Parameter geschätzt werden. Mithilfe dieses Zugangs wird also die Heterogenität zwischen den Kliniken modelliert.

Korrelative Zusammenhänge innerhalb der Kliniken werden durch die Anwendung gemischter Modelle praktischerweise mitberücksichtigt, denn es gilt:

$$\text{corr}(Y_{ij}, Y_{ij'}) = v^2 / (v^2 + \sigma^2).$$

Das heißt, die Korrelation zwischen zwei Kindern j und j' in derselben Klinik i ist genau dann groß, wenn die Varianz des konstanten Faktors α (v^2) im Vergleich zur Varianz des residualen Fehlers (σ^2) groß ist. Wenn also die Heterogenität zwischen den Kliniken (und damit v^2) groß ist, dann gibt es Kliniken, die eine große BMI-Veränderung erreichen, und andere, die nur eine kleine BMI-Veränderung erreichen. Zwei Kinder, die in einer „guten“ Klinik behandelt worden sind, werden also beide eher eine große BMI-Veränderung aufweisen, während zwei Kinder in einer „schlechten“ Klinik eher eine kleine BMI-Veränderung haben werden. Folglich werden die BMI-Veränderungen in beiden Paaren korreliert sein. Eine große Heterogenität zwischen den Kliniken (großes v^2) kennzeichnet somit gleichzeitig auch eine hohe Korrelation der Messwerte innerhalb der Kliniken.

Die beschriebene Anwendung zeigt an einem einfachen Beispiel, wie durch gemischte Modelle komplizierte Studiendesigns mit korrelierten Beobachtungen adäquat abgebildet werden können. Es sei darauf hingewiesen, dass das gemischte Modell in unserem Beispiel (mit genau einem zufälligen konstanten Faktor) äquivalent zur Varianzanalyse mit Messwiederholungen ist [7]. Die Varianzanalyse mit Messwiederholungen stellt folglich nur einen simplen Spezialfall der gemischten Modelle dar. Auch die im vorherigen Kapitel beschriebene multivariate Varianzanalyse kann als Spezialfall eines (anderen) gemischten Modells gelten.

Umgekehrt zu der Tatsache, dass die Standardmethoden Varianzanalyse mit Messwiederholungen und multivariate Varianzanalyse einfache Sonderfälle der gemischten Modelle darstellen, können letztere jedoch vielfältig erweitert werden. Dies wird v. a. dadurch erreicht, dass die gemischten Modelle zusätzlich zur Aufnahme von zufälligen Effekten in die Regressionsglei-

chung auch noch eine flexible Modellierung der Kovarianzstruktur erlauben.

• **Kompliziertere Abhängigkeitsstrukturen**

Im obigen Beispiel wird angenommen, dass die Beobachtungen innerhalb eines Clusters alle gleich korreliert sind. In Messwiederholungsdesigns, in denen der einzelne Patient als das Cluster anzusehen ist und die Zielgröße an mehreren Zeitpunkten erhoben wird (die Beobachtungen innerhalb des Clusters sind die Messungen an den verschiedenen Zeitpunkten), ist diese Annahme nicht angemessen. Man wird eher erwarten, dass Beobachtungen, die auf der Zeitachse näher zusammenliegen, stärker korreliert sein werden als solche, die weiter auseinander liegen.

• **Mehrere Ebenen von Abhängigkeiten**

Die gemischten Modelle werden oft auch als hierarchische Modelle bezeichnet, weil die Beobachtungen auf unterschiedlichen Hierarchieebenen anfallen. Im Beispiel ist das zum einen die Ebene der Kliniken, zum anderen die Ebene der Kinder. In manchen Fällen werden noch weitere hierarchische Ebenen anfallen; z. B. wäre es denkbar, dass sich die Kliniken aus der Beispielstudie noch in unterschiedlichen Regionen befinden und dadurch auch Abhängigkeiten zwischen Kliniken in derselben Region entstehen können. Auch diese zusätzlichen Ebenen können in gemischte Modelle integriert werden.

• **Andere Zielgrößen**

Die gemischten Modelle sind auch auf nichtstetige, d. h. binäre, nominale oder ordinale Zielgrößen erweiterbar, jedoch steigt damit auch deren Komplexität. Die aus Sicht der Autoren am leichtesten verständliche Einführung (auch wenn sie in englischer Sprache verfasst ist) in diese Erweiterungen der gemischten Modelle geben Brown u. Prescott [6], eine konkrete Anwendung im rehabilitationswissenschaftlichen Kontext findet man bei Hasenbein et al. [8].

Es soll schlussendlich noch auf einen weiteren Vorteil der gemischten Modelle hingewiesen werden, nämlich deren Umgang mit fehlenden Werten, wobei man hier einen softwaretechnischen und einen statistischen Aspekt unterscheiden muss. Von der Seite der Software her müssen die Datensätze sowohl für die Varianzanalyse mit Messwiederholungen als auch für die multivariate Varianzanalyse so aufgebaut sein, dass alle Werte für eine Beobachtung der höchsten Ebene in *einer* Zeile stehen müssen. Im Beispiel aus dem einführenden Kapitel würden also in jeder Zeile die Messwerte eines Kindes stehen, in einer Spalte sein Gewicht *vor* dem Trainingsprogramm, in einer zweiten sein Gewicht *nach* dem Trainingsprogramm. Demgegenüber müssen die Datensätze für die Analyse der gemischten Modelle so aufgebaut sein, dass in jeder Zeile nur eine einzelne individuelle Beobachtung steht. Im Beispiel aus dem einführenden Kapitel würde also jedes Kind zwei Zeilen im Datensatz belegen, eine für das Gewicht vor und eine für das Gewicht nach der Intervention, wobei die Gewichtswerte in ein und derselben Spalte stehen und eine Indikatorvariable anzeigt, an welchem Zeitpunkt (vor oder nach dem Trainingsprogramm) der jeweilige Wert gemessen wurde.

Wenn nun fehlende Werte vorliegen, wird in den gängigen Softwarepaketen (SPSS und SAS) bei der Analyse die ganze Zeile gelöscht. Das bedeutet im ersten Fall (also bei der Varianzanalyse mit Messwiederholungen und bei der MANOVA), dass ein Kind vollkommen aus der Analyse ausgeschlossen wird, während im gemischten Modell nur die Information zum Zeitpunkt gelöscht wird, bei dem der fehlende Wert vorliegt. Dies stellt im vorliegenden Beispiel kein so großes Problem dar; wenn wir uns aber vorstellen, dass wir im allgemeineren Fall in der Regel gleichzeitig noch andere Kovariablen mitmodellieren wollen, so würde ein einzelner fehlender Wert pro Kind (egal ob Zielgröße oder Kovariable) zum Verlust sämtlicher Information des Kindes führen. Für die Analyse des gemischten Modells hätte ein fehlender Wert weniger drastische Auswirkungen, da nicht die komplette Information pro Kind gelöscht wird, sondern nur die zu dem Zeitpunkt, an dem der fehlende Wert vorliegt.

Bezüglich des statistischen Aspektes sind die Schätzungen sowohl für die Varianzanalyse mit Messwiederholungen als auch für die multivariate Varianzanalyse beim Vorliegen von fehlenden Werten nur dann gültig, wenn diese fehlenden Werte der Annahme MCAR (missing completely at random) folgen. Das bedeutet, dass die Werte nur zufällig fehlen dürfen und nicht von der Ausprägung des eigentlich zu messenden Wertes abhängen. Die gemischten Modelle sind hier robuster. Fehlende Daten müssen lediglich der Annahme MAR (missing at random) genügen. Eine umfassende Diskussion des Problems der fehlenden Werte findet man bei Wirtz [9].

Software

Zur Berechnung von Modellen mit korrelierten Beobachtungen ist im Allgemeinen Statistiksoftware notwendig; Berechnungen innerhalb von EXCEL oder anderen Tabellenkalkulationsprogrammen, die zur reinen Datenhaltung noch ausreichen, sind prinzipiell möglich, aber mit unverhältnismäßig hohem Aufwand verbunden.

Nach den Erfahrungen der Autoren scheint SPSS das am weitesten verbreitete Softwarepaket im Bereich der Rehabilitationswissenschaften zu sein. Varianzanalysen mit Messwiederholungen finden sich in SPSS unter „Analysieren → Allgemeines lineares Modell → Messwiederholung“ (MANOVA-Prozedur), multivariate Varianzanalysen unter „Analysieren → Allgemeines lineares Modell → Multivariat“ (MANOVA-Prozedur). Seit Version 11 können mit SPSS auch gemischte Modelle (allerdings nur für stetige Zielgrößen) berechnet werden, diese findet man unter „Analysieren → Gemischte Modelle → Linear“ (MIXED-Prozedur). Steht auch die gegenwärtig leistungsfähigste Software SAS

zur Verfügung, so stehen die Prozeduren ANOVA, GLM, MIXED, NLMIXED und diverse validierte Makros zur Verfügung [10].

Fazit

In rehabilitationswissenschaftlichen Arbeiten, wie in jeglicher empirischer Forschung, treten in vielen Untersuchungsdesigns korrelierte bzw. abhängige Daten auf. Dies ist z.B. bei Messwiederholungen, multiplen Zielgrößen und parallelisierten Stichproben der Fall. Es existieren eine Reihe einschlägiger statistischer Verfahren zur spezifischen Auswertung korrelierter Daten, sodass im Prinzip kein Grund besteht, Information durch Reduktion der Messwerte auf einen einzelnen Parameter zu verschwenden. Unter den statistischen Methoden, die die Daten auf Originalniveau auswerten, stellen die gemischten Modelle die Methode der Wahl dar, da diese eine sehr flexible Modellierung der Korrelationsstruktur der Beobachtungen erlauben und inzwischen auch mit Standardsoftware schätzbar sind.

Auch wenn sich die Darstellung hier nur auf stetige Zielgrößen bezieht, existieren für jedes beliebige Skalenniveau der Zielgröße – binär, ordinal, nominal – statistische Verfahren für korrelierte Daten. Jedoch sind diese mathematisch komplexer und gehen über den Rahmen dieses Artikels hinaus.

Literatur

- ¹ Matthews JNS, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *British Medical Journal* 1990; 300: 230–235
- ² Lindman HR. *Analysis of Variance in Experimental Design*. New York: Springer, 1992
- ³ Davis CS. *Statistical Methods for the Analysis of Repeated Measurements*. New York: Springer, 2002
- ⁴ Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *British Medical Journal* 2001; 323: 1123–1124
- ⁵ Bray JH, Maxwell SE. *Multivariate Analysis of Variance*. Newbury Park: Sage Publications, 1994
- ⁶ Brown H, Prescott R. *Applied Mixed Models in Medicine*. Chichester: Wiley, 1999
- ⁷ Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Heidelberg: Springer, 1999
- ⁸ Hasenbein U, Kuss O, Bäumer M, Schert C, Schneider H, Wallesch CW. Physicians' preferences and expectations in stroke rehabilitation – results of a case-based questionnaire survey. *Disability and Rehabilitation* 2002; 24: 954–960
- ⁹ Wirtz M. Über das Problem fehlender Werte: Wie der Einfluss fehlender Informationen auf Analyseergebnisse entdeckt und reduziert werden kann. *Rehabilitation* 2004; 43: 109–115
- ¹⁰ Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models*. Cary, NC, USA: SAS Institute Inc, 1996