

## Power- und Fallzahlab-schätzungen für hierarchische und longitudinale Studien

Jens Dreyhaupt  
Sabrina Sufeida  
Rainer Muche

Institut für Epidemiologie  
und Medizinische Biometrie,  
Universität Ulm  
Schwabstraße 13  
89075 Ulm  
jens.dreyhaupt@uni-ulm.de

Jürgen-Michael Steinacker  
für die Arbeitsgruppe  
„Komm mit in das gesunde  
Boot - Grundschule“  
Sektion Sport- und  
Rehabilitationsmedizin,  
Universitätsklinikum Ulm  
Leimgrubenweg 12-14  
89075 Ulm

### Zusammenfassung

Klinische und epidemiologische Studien liefern häufig Studiendesigns mit hierarchischen und/oder longitudinalen Daten. Diese Datenstrukturen weisen häufig statistische Abhängigkeiten auf, die bei Studienplanung und -auswertung berücksichtigt werden müssen. Statistische Methoden für einfache Situationen von Cluster-Daten werden in [1] beschrieben. Bei komplexeren Situationen besteht für die Power-Abschätzung metrischer Zielgrößen die Möglichkeit der Anwendung eines Verfahrens mit Erzeugung künstlicher Datensätze und der Anwendung linearer gemischter Modelle ([2], [3], [4]). Diese Methode wurde zur a-priori Power-Abschätzung für die Baden-Württemberg Studie verwendet, einer prospektiven, longitudinalen, zweiarmigen, cluster-randomisierten Studie zur Untersuchung der Wirksamkeit der Intervention des Präventionsprogramms „Komm mit in das gesunde Boot - Grundschule“ (Finanzierung: Baden-Württemberg Stiftung) [5]. Das für die Power-Abschätzung angewendete Verfahren basiert auf Daten, welche vom Vorgängerprojekt UR-MEL-ICE zur Verfügung gestellt wurden. Im Artikel werden Datenstruktur und Annahmen der Power-Abschätzung vorgestellt. Die fünf Schritte des Verfahrens werden detailliert beschrieben. Es wurde über Datenschnitte und Aufrufe der SAS Prozeduren MIXED und PRINT realisiert. Weiterhin werden die SAS<sup>®</sup> Funktionen FINV und PROBF angewendet. Der SAS Quelltext für die Realisierung des Verfahrens wird gezeigt. Es wird die Power verschiedener Szenarien vorgestellt sowie die Power für die entsprechende unverbundene Datensituation (basierend auf einem (nicht korrekten) Zweistichproben-t-Test). Mit der letztgenannten Methode erfolgt eine starke Überschätzung der Power, womit die Wichtigkeit einer Berücksichtigung der statistischen Abhängigkeit bei hierarchischen und/oder longitudinalen Daten gezeigt wird. Für die Bewertung einer Vielzahl von Szenarien kann die Entwicklung eines SAS Makros erfolgen.

**Schlüsselwörter:** Poweranalyse, hierarchische Daten, longitudinale Daten, Cluster-Randomisierung, gemischtes Modell, MIXED-Prozedur, FINV-, PROBF-Funktion

## **1 Einleitung und Fragestellung**

In klinischen und epidemiologischen Studien werden zunehmend Studiendesigns mit hierarchischen und/oder longitudinalen Daten angewendet, die beispielsweise durch Cluster-Randomisierung oder Beobachtung von Untersuchungseinheiten im Zeitverlauf entstehen. Neben der Durchführung und statistischen Auswertung stellen derartige Studiendesigns spezifische Anforderungen an die Planung. Eine adäquate Fallzahlplanung oder Power-Abschätzung erfordert die Berücksichtigung von Abhängigkeiten der Messwerte innerhalb der Untersuchungseinheiten (longitudinale Daten) oder zwischen den Untersuchungseinheiten (Cluster-Daten). Das Ignorieren der Abhängigkeiten kann zur Überschätzung der Power und damit zu nicht adäquaten Fallzahlen führen. Gängige Methoden der Studienplanung sind für einfache Situationen in [1] beschrieben. Diese sind jedoch bei komplexeren Situationen (z.B. Multilevel-Datenstrukturen, extrem unbalancierten Clustern, Berücksichtigung von Confoundern) nicht oder nur unzureichend anwendbar. Eine mögliche Lösung besteht hierfür bei metrischen Zielgrößen in der Anwendung eines Verfahrens mit Erzeugung künstlicher Datensätze und der Anpassung linearer gemischter Regressionsmodelle ([2], [3], [4]).

Im Folgenden wird dieses Verfahren exemplarisch zur a-priori Power-Abschätzung für eine cluster-randomisierte Studie angewendet, wobei die Realisierung mit SAS erfolgte. Einleitend werden Hintergründe der Studie, Datenstruktur und Annahmen der Power-Abschätzung dargestellt. Im Fokus steht die detaillierte Beschreibung des Verfahrens einschließlich seiner Umsetzung in SAS. Als Ergebnis wird die Power für einige mögliche Szenarien angegeben und mit der Power für die entsprechende unverbundene Datensituation (basierend auf einem (nicht korrekten) Zweistichproben-t-Test) verglichen. Abschließend werden Konsequenzen aus den Ergebnissen diskutiert und ein Ausblick auf weitere Entwicklungen gegeben.

## **2 Material und Methodik**

In diesem Kapitel werden Informationen über den Hintergrund der cluster-randomisierten Baden-Württemberg Studie gegeben. Die Datenstruktur der Studie und wesentliche Annahmen für ihre Power-Abschätzung werden beschrieben.

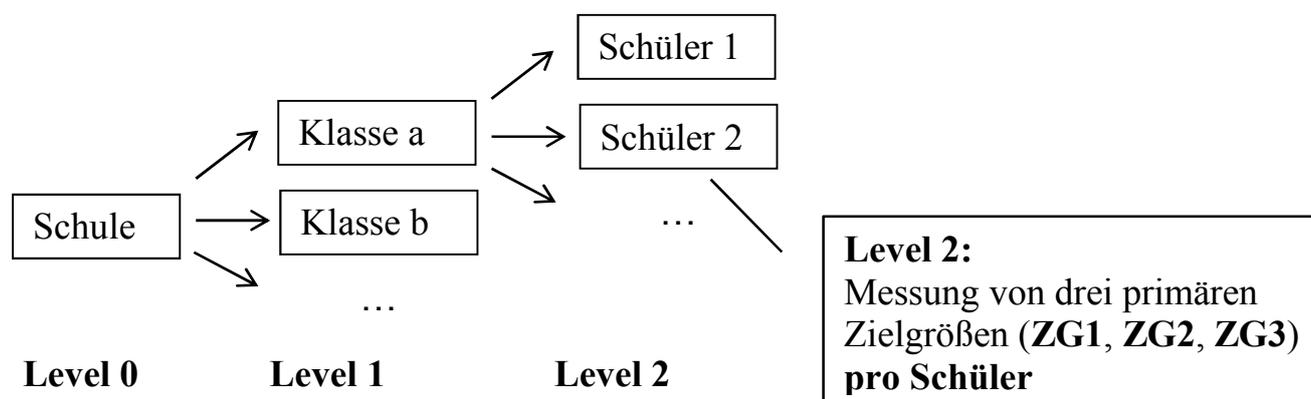
### **2.1 Hintergrund der Baden-Württemberg Studie**

Das Präventionsprogramm „Komm mit in das gesunde Boot - Grundschule“ (Finanzierung: Baden-Württemberg Stiftung) ist ein umfassendes Programm zur Gesundheitsförderung und soll Grundschulkindern in ganz Baden-Württemberg die Grundlagen für einen gesünderen Lebensstil im Hinblick auf Ernährung, Bewegung und Freizeitverhalten vermitteln. Wesentliche Elemente des Programms basieren auf der Vorgängerstudie URMEL-ICE, einer randomisierten kontrollierten Studie zur Untersuchung der Wirksamkeit eines Programms zur Prävention von Übergewicht im Kindesalter. Die einjährige schulbasierte Intervention des „gesunden Bootes“ ist Teil des Unterrichts bzw. der Pausen, zusätzliche Sportstunden sind nicht vorgesehen. Die Durchführung erfolgt durch den Klassenlehrer anhand speziell ausgearbeiteter Materialien, unter Einbe-

ziehung von Freunden und Familie. Die Verbreitung des Programms in Baden-Württemberg ist über ein landesweites Multiplikatorensystem realisiert. Multiplikatoren sind Grundschullehrkräfte, die in zentralen Schulungen ausführlich mit Programm und Intervention vertraut gemacht wurden und ihr Wissen dann vor Ort an weitere interessierte Grundschullehrkräfte weitergeben. Die Untersuchung der Wirksamkeit der Intervention erfolgte mit der Baden-Württemberg Studie, einer prospektiven, longitudinalen, zwei-armigen, cluster-randomisierten Studie [5]. Die Randomisierung in Kontrollgruppe und Interventionsgruppe wurde auf Schulebene (Cluster) realisiert, mit einer Stratifizierung nach sechs verschiedenen Schultypen. Die Untersuchungseinheit war das einzelne Grundschulkind, welches zu zwei Zeitpunkten untersucht wurde, zu Beginn der Schuljahre 2010/2011 (Baseline) und 2011/2012 (Follow-up). Für die Untersuchung der Wirksamkeit wurden im Studienprotokoll drei primäre metrische Zielgrößen festgelegt:

- Differenz in der subscapularen Hautfaltendicke (**ZG1**)
- Differenz im Bauchumfang (**ZG2**)
- Differenz im 6-Minuten-Lauf (**ZG3**)

Die Differenzen sind die Unterschiede der genannten Zielgrößen zwischen Follow-up Messung und Baseline-Messung. Abbildung 1 zeigt die hierarchische Datenstruktur der Baden-Württemberg Studie. Eine Schule gehört dabei entweder zur Interventionsgruppe oder zur Kontrollgruppe.



**Abbildung 1:** Datenstruktur der Baden-Württemberg Studie

## 2.2 Power-Abschätzung

Zur Bewertung der Möglichkeiten der Studie wurde in der Planungsphase eine a-priori Power-Abschätzung durchgeführt. Da während dieser Phase noch keine genauen Anzahlen von teilnehmenden Schulen bekannt waren, konnten hierfür nur Annahmen getroffen werden, auf Basis der Schulanzahlen der für das Programm angemeldeten Lehrkräfte. Grundlage der Power-Abschätzung ist ein Verfahren, bei welchem eine Erzeugung künstlicher Datensätze und eine Anpassung linearer gemischter Regressionsmodelle erfolgt ([2], [3], [4]). Für die Anwendung dieses Verfahrens auf die Baden-Württemberg Studie wurden vom Projektteam folgende Annahmen getroffen:

- In Interventions- und Kontrollgruppe wird die gleiche Anzahl Schulen randomisiert.
- Feste Schüleranzahl pro Klasse und keine Unterscheidung zwischen verschiedenen Klassen innerhalb einer Schule. Hierdurch ergibt sich eine Vereinfachung der Datenstruktur in Abbildung 1.
- Im Regressionsmodell wird nur ein fester Effekt (Gruppe) modelliert.
- Die praktisch relevante Differenz zwischen Interventions- und Kontrollgruppe in den drei primären Zielgrößen entspricht dem in URMEL-ICE beobachteten Effekt.
- Adjustierung des Fehlers erster Art nach Bonferroni:  $\alpha=0.05/3=0.0167$  (zweiseitig), da drei primäre Zielgrößen untersucht werden sollen.

Aus den Daten des Vorgängerprojekts URMEL-ICE konnte eine mittlere Klassengröße erhalten werden. Durch Anwendung eines linearen gemischten Regressionsmodells auf die URMEL-ICE Daten wurden die Effekte für **ZG1**, **ZG2** und **ZG3** in Interventions- und Kontrollgruppe sowie die Schätzer für die Varianzkomponenten bestimmt. Mit diesen Informationen und den oben gemachten Annahmen kann das Verfahren der Power-Abschätzung durchgeführt werden. Details des Verfahrens sind im folgenden Abschnitt am Beispiel von **ZG1** beschrieben.

### **2.3 Verfahren der Poweranalyse für hierarchische Datenstrukturen**

Grundlage des Verfahrens ist der Test einer allgemeinen linearen Hypothese über die festen Effekte im linearen gemischten Regressionsmodell mittels F-Test. Das Verfahren besteht aus den folgenden fünf Schritten.

- 1 Anpassung eines linearen gemischten Regressionsmodells an existierende Daten (z.B. Pilotprojekt, Vorgängerprojekt) und Verwendung der geschätzten Modellparameter (Regressionskoeffizienten der festen Effekte und Schätzer der Varianzkomponenten) für die nachfolgenden Schritte 2 bis 5.
- 2 Erzeugung eines künstlichen Datensatzes entsprechend der zu planenden Studie und Anpassung des linearen gemischten Regressionsmodells aus Schritt 1 unter Verwendung der in Schritt 1 erhaltenen Schätzer der Varianzkomponenten.
- 3 Speicherung von Teststatistik sowie der Freiheitsgrade von Zähler und Nenner des F-Tests für den Regressionskoeffizienten des zu untersuchenden festen Effekts.
- 4 Berechnung des Nichtzentralitätsparameters sowie des kritischen Wertes unter Vorgabe eines Maximalwertes des Fehlers 1. Art.
- 5 Berechnung der korrespondierenden Power für Nichtzentralitätsparameter und kritischen Wert.

Für Schritt 1 ist das Vorhandensein geeigneter Daten notwendig, vor allem um realitätsnahe Schätzwerte für die Varianzkomponenten zu erhalten. Eine Verwendung von Annahmen kann starken Einfluss auf die Power haben. Weitere Details und Anwendungsbeispiele sind [2], [3] und [4] zu entnehmen.

Im folgenden Abschnitt ist die Umsetzung des Verfahrens mit SAS beschrieben.

## 2.4 Umsetzung in SAS

### 2.4.1 Schritt 1

Zur Umsetzung von Schritt 1 sind zunächst die Daten eines geeigneten Pilot- oder Vorgängerprojekts einzulesen. Dies war für die vorliegende Arbeit das Projekt URMEL-ICE. Die Struktur der gegebenen Daten ist Tabelle 1 zu entnehmen.

**Tabelle 1:** Struktur der gegebenen Daten

id	schule	gruppe	ZG1	ZG2	ZG3	...
1001	20	0	7.66	5.15	-90	...
1004	20	0	-2.33	8.60	251	...
...	...	...	...	...	...	...
1103	16	1	-1.66	4.95	113	...
1105	16	1	-1.33	0.25	105	...
...	...	...	...	...	...	...

Die Variable **id** ist eine eindeutige Schülernummer, **schule** ist die eindeutige Schulnummer ( $schule = 1, \dots, 32$ ), **gruppe** bezeichnet die Interventionsgruppe (0=Kontrollgruppe, 1=Interventionsgruppe). Die Variablen **ZG1**, **ZG2**, und **ZG3** bezeichnen die drei primären metrischen Zielgrößen.

An diesen gegebenen Datensatz wurde ein für die Fragestellung geeignetes lineares gemischtes Regressionsmodell angepasst, hier am Beispiel der Zielgröße **ZG1**:

$$\mathbf{ZG1}_{ij} = \alpha + \beta \cdot \mathbf{gruppe}_i + \mathbf{b}_i + \varepsilon_{ij}$$

wobei

- $\mathbf{ZG1}_{ij}$  ist der Messwert von **ZG1** beim j-ten Schüler der i-ten Schule.
- $\mathbf{gruppe}_i$  ist die Gruppenzugehörigkeit der i-ten Schule, d.h. ein fester Effekt ( $\mathbf{gruppe}_i = 0$ : Kontrollgruppe,  $\mathbf{gruppe}_i = 1$ : Interventionsgruppe)
- $\mathbf{b}_i$  ist ein zufälliger Effekt (Effekt der Schule)
- $\varepsilon_{ij}$  ist der zufällige Restfehler
- Index i: Schulnummer ( $i=1, \dots, 32$ )
- Index j: Schülernummer innerhalb einer Schule ( $j=1, \dots, n_i$ , mit  $n_i$  ist Anzahl der Schüler in der i-ten Schule)

Die Regressionskoeffizienten  $\alpha$  bzw.  $\beta$  sind Achsenabschnitt bzw. Effekt der Interventionsgruppe. Für die zufälligen Effekte gelten die Verteilungsannahmen  $\mathbf{b}_i \sim N(0, \sigma_1^2)$  und  $\varepsilon_{ij} \sim N(0, \sigma^2)$ .

Für den Test des festen Effekts **gruppe** wird eine Linearkombination der Form  $\mathbf{L}^T \cdot \boldsymbol{\beta}$  betrachtet, wobei  $\boldsymbol{\beta}$  der Vektor der Regressionskoeffizienten der festen Effekte und  $\mathbf{L}$  eine (als bekannt vorausgesetzte) Koeffizientenmatrix ist. Getestet wird das Hypothesenpaar  $H_0: \mathbf{L}^T \cdot \boldsymbol{\beta} = \mathbf{0}$  versus  $H_1: \mathbf{L}^T \cdot \boldsymbol{\beta} \neq \mathbf{0}$ . Die Teststatistik

$$F = \frac{\hat{\beta}^T L^T \left[ L \left( \sum_{i=1}^n X_i^T \hat{V}_i^{-1} X_i \right)^{-1} L^T \right]^{-1} L \hat{\beta}}{\text{Rang}(L)}$$

ist unter  $H_0$  F-verteilt mit  $\text{Rang}(L)$  Freiheitsgraden im Zähler. Die Anzahl der Freiheitsgrade des Nenners muss aus den Daten geschätzt werden. Unter  $H_1$  ist die Teststatistik F-verteilt mit denselben Freiheitsgraden für Zähler und Nenner aber mit dem Nicht-zentralitätsparameter

$$\varphi = \hat{\beta}^T L^T \left[ L \left( \sum_{i=1}^n X_i^T \hat{V}_i^{-1} X_i \right)^{-1} L^T \right]^{-1} L \hat{\beta}$$

Die praktische Umsetzung erfolgt in den Schritten 4 und 5, siehe Abschnitt 2.4.3. Weiterführende Informationen sind [2], [3] und [4] zu entnehmen.

Für die Realisierung der Modellanpassung mit der SAS Prozedur MIXED wurde für **ZG1** folgender Quelltext verwendet:

```
PROC MIXED data= ksfe.urmel_long;
  CLASS gruppe schule;
  MODEL ZG1 = gruppe /solution cl;
  RANDOM intercept / subject=schule;
RUN;
```

Mit der Option SOLUTION werden die Schätzer der festen Effekte ( $\alpha$ ,  $\beta$ ) ermittelt, die Option CL liefert darüber hinaus noch die Schätzer der 95% Konfidenzintervalle. Die im Ergebnis der Modellanpassung erhaltenen Schätzer für  $\alpha$  (=1.3568),  $\beta$  (=0.6306),  $\sigma_1^2$  (=0.3959) und  $\sigma^2$  (=7.8167) werden im Schritt 2 für die Erzeugung des künstlichen Datensatzes und für die Anpassung des linearen gemischten Regressionsmodells benötigt.

## 2.4.2 Schritte 2 und 3

In Schritt 2 wird ein künstlicher Datensatz erzeugt, welcher der Struktur des gegebenen Datensatzes entspricht. Mit dem künstlichen Datensatz wird ein mögliches Szenario für die geplante Studie simuliert. Neben den in Schritt 1 erhaltenen Schätzern aus der Anpassung des gemischten Modells (vergleiche Abschnitt 2.4.1) werden noch Annahmen über die Gesamtzahl der Schulen und die Schüleranzahlen pro Schule benötigt. In Interventions- und Kontrollgruppe gibt es dieselbe Anzahl Schulen (siehe Abschnitt 2.2). Der künstliche Datensatz mit 40 Schulen pro Gruppe und 15 Schülern pro Schule wird für **ZG1** mit folgendem SAS Datenschnitt realisiert:

```

DATA ZG1_szenario_1;
  DO gruppe = 0 to 1;          /*Interventions- und Kontrollgruppe */
    DO schule = 1 to 40;      /*40 Schulen in jeder Gruppe */
      DO id = 1 to 15;       /*15 Schüler pro Schule */
        ZG1 = 1.3568 - 0.6306*Gruppe;
        OUTPUT;
      END;
    END;
  END;
RUN;

```

Im Ergebnis des Datenschritts entsteht eine zu den Daten des Vorgängerprojekts ähnliche Struktur, siehe Tabelle 2.

**Tabelle 2:** Struktur der künstlichen Daten für **ZG1** aus Schritt 2

<b>Id</b>	<b>schule</b>	<b>gruppe</b>	<b>ZG1</b>
1	1	0	1.3568
...	...	...	...
15	1	0	1.3568
...	...	...	...
1	40	0	1.3568
...	...	...	...
15	40	0	1.3568
1	1	1	0.7262
...	...	...	...
15	1	1	0.7262
...	...	...	...
1	40	1	0.7262
...	...	...	...
15	40	1	0.7262

Im Gegensatz zu den Daten des Vorgängerprojekts URMEL-ICE (siehe Tabelle 1) sind die Variablen **id** und **schule** hier nicht eindeutig sondern verschachtelt (**id** innerhalb von **schule**, **schule** innerhalb von **gruppe**), was bei der Struktur des linearen gemischten Modells in Schritt 3 beachtet werden muss. Die Anpassung des linearen gemischten Regressionsmodells an die künstlichen Daten ist mit der Prozedur MIXED wie folgt realisiert.

```
PROC MIXED data = ZG1_szenario_1 noprofile;
  CLASS gruppe schule;
  MODEL ZG1 = gruppe;
  RANDOM intercept / subject = schule (gruppe);
  PARMs (0.3959) (7.8167) / noiter;
  CONTRASZ 'equal effect' gruppe 1 -1;
  ODS OUTPUT contrasts=c1;
RUN;
```

Der zufällige Effekt der Schule ist bei dieser Syntax verschachtelt innerhalb der Interventionsgruppe, was eine Folge der nicht eindeutigen Variable **schule** ist. Dies hat für die in der Studienplanung betrachteten Szenarien jedoch nur wenig Einfluss auf die Power. Mit dem PARMs Statement werden die Startwerte für die Varianzkomponenten spezifiziert. Die Optionen NOITER und NOPROFILE bewirken (zusammen verwendet), dass für die Inferenz die im PARMs Statement spezifizierten Parameterwerte verwendet werden. Mit dem CONTRAST Statement erfolgt der Test der Gleichheitshypothese beider Gruppen, wobei ohne Angaben weiterer Optionen der F-Test ausgeführt wird. Über ein ODS Statement wird für das Szenario das Ergebnis des F-Tests (Größe der Teststatistik) sowie die Freiheitsgrade von Zähler und Nenner im Datensatz c1 gespeichert (Schritt 3 des Verfahrens). Diese Informationen werden in Schritt 4 und 5 für die Berechnung der Power benötigt.

### 2.4.3 Schritte 4 und 5

In den Schritten 4 und 5 wird die in Schritt 3 erhaltene Tabelle c1 für die Berechnung der Power verwendet. Weiterhin ist noch die Angabe eines Maximalwerts für den Fehler erster Art ( $\alpha$ ) notwendig. Es wird zunächst der Nichtzentralitätsparameter aus den Daten der Tabelle c1 berechnet sowie das  $(1 - \alpha)$  - Quantil der F-Verteilung zu den gegebenen Freiheitsgraden bestimmt. Abschließend wird die Power für das Szenario berechnet. Die Realisierung erfolgt über den folgenden Datenschnitt:

```
DATA power;
  SET c1;
  alpha=0.05/3;          /*Bonferroni-Adjustierung*/
  ncparm=numdf*fvalue;  /*Berechnung Nichtzentralitätsparameter*/
  fc=FINV(1-alpha,numdf,dendf,0);
  power=1-PROBF(fc,numdf,dendf,ncparm);
RUN;
PROC PRINT data=power;
RUN;
```

Da für die Baden-Württemberg Studie drei primäre Zielgrößen spezifiziert wurden (vergleiche Abschnitt 2.1), ist für den Fehler erster Art eine Adjustierung nach Bonferroni realisiert worden:  $\alpha=0.05/3$ . Mit der Funktion FINV wird das  $1-0.05/3$  Quantil der zentralen F-Verteilung mit den Freiheitsgraden von Nenner und Zähler des F-Tests für den

festen Effekt (Variable **gruppe**, siehe Schritte 2 und 3 in Abschnitt 2.4.2) berechnet. In der Variable **power** wird die Power des Szenarios gespeichert, berechnet über die Funktion PROBF. Diese wird angewendet auf den aus dem Ergebnis der Modellanpassung erhaltenen Nichtzentralitätsparameter. Abschließend wird das Ergebnis der Power-Abschätzung über die Prozedur PRINT ausgegeben.

Im folgenden Abschnitt werden Ergebnisse der Powerberechnung für ausgewählte Szenarien dargestellt.

### 3 Ergebnisse

In diesem Abschnitt werden Ergebnisse ausgewählter Szenarien der Power-Abschätzung für **ZG1** dargestellt und mit der Power des (statistisch nicht adäquaten) Zweistichproben-t-Tests verglichen. Die Annahmen der Power-Abschätzung sind in Abschnitt 2.2 genannt. Für diese Untersuchungen wurden

- in Interventions- und Kontrollgruppe 40 oder 50 Schulen angenommen.
- pro Klasse eine Anzahl von 15 oder 20 Schülern angenommen, von denen Daten für die Auswertung der drei primären Zielgrößen zur Verfügung stehen.

Hieraus ergeben sich vier Szenarien. Für deren Power-Abschätzung wurden entsprechende künstliche Datensätze erzeugt, an denen die Schritte 2 bis 5 ausgeführt worden sind. Für die vier Szenarien gibt Tabelle 3 eine Zusammenfassung der Ergebnisse einschließlich des Vergleichs mit der Power eines korrespondierenden Zweistichproben-t-Tests, welche mit der Prozedur POWER in SAS ermittelt wurde.

**Tabelle 3:** Ergebnisse der Power-Abschätzung für ZG1 und Vergleich mit dem korrespondierenden Zweistichproben-t-Test

Szenario	nSchulen	nSchüler/Klasse	n <sub>gesamt</sub>	Power (%) lin. gem. Modell	Power (%) Zweistichpr.-t-Test
1	2 x 40	15	1200	69	92
2	2 x 40	20	1600	77	98
3	2 x 50	15	1500	80	97
4	2 x 50	20	2000	87	99

Tabelle 3 zeigt, dass die mit dem Verfahren berechneten Werte der Power stark vom Szenario abhängen. Bei unsicheren Informationen in der Studienplanungsphase können somit nur grobe Informationen über die Größenordnung der Power erhalten werden. Weiterhin ist sichtbar, dass die Power mit dem nicht adäquaten Zweistichproben-t-Test teilweise stark überschätzt wird. Dieser berücksichtigt nicht, dass es sich um hierarchische Daten handelt.

## **4 Diskussion und Ausblick**

Ein zentraler Punkt der Studienplanung ist die Berechnung der notwendigen Fallzahl bzw. der erreichbaren Power in einem gegebenen Setting. Bei Studiendesigns mit hierarchischen und/oder longitudinalen Daten entstehen statistische Abhängigkeiten, welche Einfluss auf Fallzahl bzw. Power haben und durch adäquate Methoden Berücksichtigung finden müssen. Übliche Methoden für einfache Situationen sind in [1] beschrieben, allerdings bei komplexeren Situationen nicht oder nur unzureichend anwendbar.

Das hier vorgestellte Verfahren ermöglicht eine Abschätzung der Power bei metrischen Zielgrößen auch für komplexere Situationen. Auf Grundlage künstlicher Datensätze und der Anpassung geeigneter linearer gemischter Regressionsmodelle mit entsprechenden Varianzkomponenten kann eine Berücksichtigung der statistischen Abhängigkeiten erfolgen. Eine wesentliche Voraussetzung für das Verfahren ist das Vorhandensein geeigneter Daten aus einem Pilot- oder Vorgängerprojekt, um realitätsnahe Schätzer der Varianzkomponenten und des festen Effektes zu erhalten.

Die Anwendung der Prozedur MIXED erlaubt die Modellierung flexibler Kovarianzstrukturen, sowie die Berücksichtigung von mehr als einer Hierarchieebene. Weiterhin ist es möglich, zusätzliche feste Effekte, wie Baseline-Werte der Zielgröße, zu berücksichtigen. Eine Erweiterung des Verfahrens kann erfolgen, in dem für Interventions- und Kontrollgruppe auch ungleiche Klassengrößen und Schulanzahlen berücksichtigt werden. Durch Einbindung des Verfahrens in ein SAS Makro kann eine komfortable und schnelle Anwendung für verschiedene Szenarien erfolgen.

## **Literatur**

- [1] Donner A, Klar N.: Design and Analysis of Cluster Randomization Trials in Health Research. Arnold London 2000.
- [2] Verbeke G, Molenberghs G: Longitudinal and incomplete data Analysis. A one day course by Geert Verbeke and Geert Molenberghs at the International Biometric Conference in Cairns, Australia (July 10, 2004).  
[<https://perswww.kuleuven.be/~u0018341/documents/longitudinal.pdf>]
- [3] Verbeke G, Molenberghs G: Linear Mixed Models for Longitudinal Data. Springer New York 2009.
- [4] Littell RC, Milliken GA, Walter WW, Wolfinger RD, Schabenberber O. SAS<sup>®</sup> for Mixed Models, Second Edition. SAS<sup>®</sup> Institute Inc., Cary, NC, USA 2006.
- [5] Dreyhaupt J, Koch B, Wirt T, Schreiber A, Brandstetter S, Kesztyues D, Wartha O, Kobel S, Kettner S, Prokopchuk D, Hundsdoerfer V, Klepsch M, Wiedom M, Sufeida S, Fischbach N, Muche R, Seufert T & Steinacker JM. Evaluation of a health promotion program in children: Study protocol and design of the cluster-randomized Baden-Wuerttemberg primary school study [DRKS-ID: DRKS00000494]. BMC Public Health, 12, 157 (2012).  
[<http://www.biomedcentral.com/1471-2458/12/157>]