

## **Untersuchung zum Balanceverhalten der stratifizierten Blockrandomisierung - Eine Lösung mit SAS Makros -**

Laura Hupperz Universität Ulm Schwabstr. 13 89075 Ulm laura.hupperz@gmail.com	Friederike Rohlmann Universität Ulm Schwabstr. 13 89075 Ulm friederike.rohlmann@uni-ulm.de
Beate Einsiedler Universität Ulm Schwabstr. 13 89075 Ulm beate.einsiedler@uni-ulm.de	Rainer Mucbe Universität Ulm Schwabstr. 13 89075 Ulm rainer.mucbe@uni-ulm.de

### **Zusammenfassung**

Die randomisierte kontrollierte klinische Studie wird in der Fachliteratur als „Goldstandard“ der klinischen Forschung bezeichnet. Um bei Erreichen der geplanten Fallzahl möglichst gleiche Studienteilnehmerzahlen in allen Behandlungsgruppen zu erzielen, wird häufig die sogenannte Blockrandomisierung eingesetzt. Um die Unvorhersehbarkeit der Therapiezuweisungen (Concealment) zum Ende der Blockauffüllung möglichst groß zu halten, können verschiedene Blocklängen zufällig gewählt und zu einer Randomisierungsliste aneinandergefügt werden (permutierte Blockrandomisierung). Bei der stratifizierten Blockrandomisierung werden solche Randomisationslisten pro Stratum (= Schicht) erstellt, z.B. für Frauen und Männer und/oder bei multizentrischen Studien pro Zentrum.

Die wichtigsten zu wählenden Parameter bei der Planung einer solchen stratifizierten, permutierten Blockrandomisierung sind demnach die Blocklängen je Stratum. Dabei ist zu beachten, dass sich mit steigender Anzahl Strata die Wahrscheinlichkeit der Imbalance durch viele angebrochene Blöcke am Ende der Rekrutierung erhöht. Es gilt also, einen Kompromiss zu finden zwischen Gesamtbalance und Concealment.

Zur Entscheidungsunterstützung wurden SAS Makros entwickelt, mit denen es möglich ist, die Rekrutierung anhand der Studienvorgaben (geplante Fallzahl, Strata, erwartete Patientenzahlen je Stratum) und ausgewählten Blocklängen zu simulieren [1].

**Schlüsselwörter:** Stratifizierte permutierte Blockrandomisierung, klinische Studie, Balance der Studienbehandlungen, Concealment, Simulation, Makro

## 1 Problembeschreibung bei der stratifizierten Blockrandomisierung

Die randomisierte kontrollierte klinische Studie wird in der Fachliteratur als „Goldstandard“ der klinischen Forschung bezeichnet. Randomisierung bedeutet zufällige Zuteilung von Patienten zu den verschiedenen Studienbehandlungen mit dem Ziel, die Störgrößen gleichmäßig auf die Behandlungsgruppen zu verteilen. Eine uneingeschränkte Randomisierung bedeutet vollkommen zufällige Zuteilung der Patienten zu den Behandlungen. Es gilt als das beste Verfahren hinsichtlich Unvorhersehbarkeit und Vermeidung von systematischen Fehlern. Das Problem bei diesem Vorgehen ist, dass insbesondere bei kleineren Studien zu Ende der Rekrutierung eine gleiche Anzahl der Patienten in den Behandlungsgruppen (Balanziertheit) nicht garantiert ist. Deshalb wird häufig die Blockrandomisierung eingesetzt. Der Zufall wird eingeschränkt, da pro Block vorgegeben ist, dass (in der Regel) gleich viele Patienten den Gruppen zugeteilt werden. Ist z.B. bei zwei möglichen Studienbehandlungen die Anzahl in einer der Gruppen erreicht, so ergeben sich die letzten Zuteilungen zur anderen Gruppe in einem Block deterministisch. Der Balance-Vorteil wird also durch die Reduzierung der Unvorhersagbarkeit der nächsten Behandlung(en) (Concealment) eingeschränkt.

Um das Concealment bei der Blockrandomisierung zu verbessern, können abhängig von der Fallzahl unterschiedliche Blockgrößen gewählt und zufällig aneinandergefügt werden (permutierte Blockrandomisierung). In klinischen Studien erfolgt die Randomisierung häufig zusätzlich stratifiziert, d.h. es gibt separate geblockte Randomisierungslisten für jede Ausprägung eines prognostischen Faktors (z.B. männlich/weiblich) bzw. Ausprägungskombination bei mehreren Faktoren (z.B. Zentrum1-männlich/Zentrum1-weiblich/ Zentrum2-männlich ...). Das Verfahren wird dann stratifizierte Blockrandomisierung genannt. In multizentrischen Studien wird in der Regel mindestens nach Zentrum stratifiziert.

Die wichtigsten zu wählenden Parameter bei der Planung einer solchen stratifizierten, ggf. permutierten Blockrandomisierung sind demnach die Blocklängen je Stratum. Dabei ist zu beachten, dass sich mit steigender Anzahl Strata die Wahrscheinlichkeit der Imbalance durch viele angebrochene Blöcke am Ende der Rekrutierung erhöht. Es gilt also, bei der Festlegung der Blocklängen einen Kompromiss zu finden zwischen Gesamtbalance und Concealment. Zur Entscheidungsunterstützung wurden Makros entwickelt [1], mit denen es möglich ist, die Rekrutierung anhand der Studienvorgaben (geplante Fallzahl, Strata, erwartete Patientenzahlen je Stratum) und ausgewählten Blocklängen zu simulieren.

## 2 Realisierung der Randomisierungssimulation

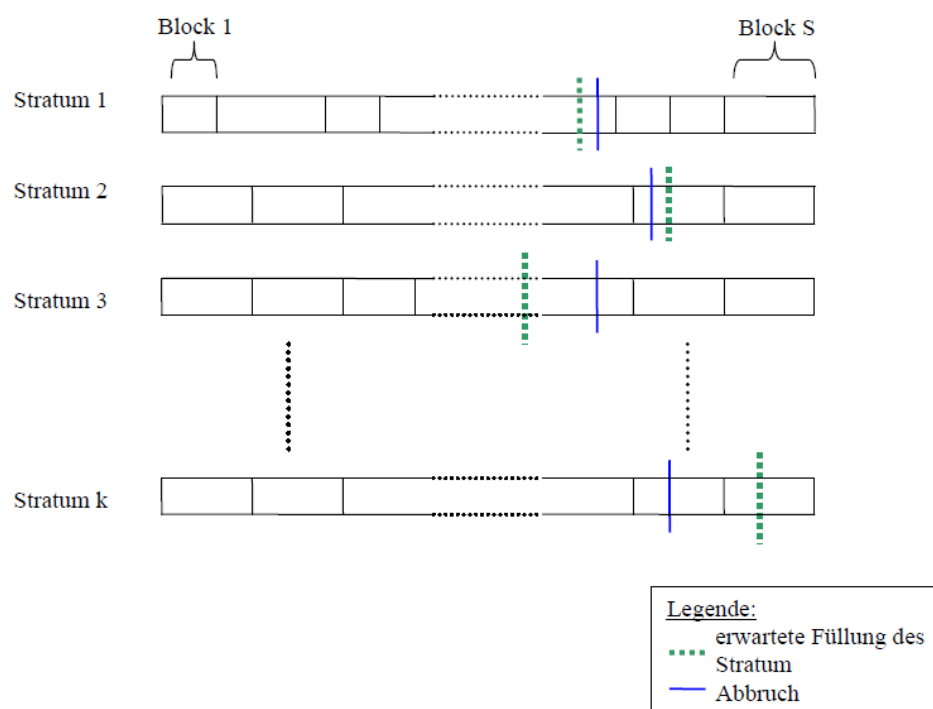
Die SAS-Makros sind für das gängigste Studiendesign mit zwei Studienbehandlungen und gleicher Wahrscheinlichkeit für die randomisierte Behandlungszuteilung (1:1) konzipiert und programmiert worden. Es können aus beliebig vielen Stratifizierungsvariab-

len (z.B. Geschlecht, Zentrum) entsprechend ihrer Ausprägungen Strata gebildet und an die Makros übergeben werden (z.B. Stratum1 = Zentrum1-weiblich; Stratum2 = Zentrum1-männlich ...).

Für jedes Stratum können ein oder mehrere Blocklängen gewählt werden. Dabei orientiert sich die Wahl der Blocklängen pro Stratum an der jeweiligen erwarteten Rekrutierungsrate. Diese errechnet sich bei zwei oder mehr Stratifizierungsvariablen (unter Annahme der Unabhängigkeit) als Produkt aus den Wahrscheinlichkeiten ihrer jeweiligen Ausprägungen (Bsp.: Erwartungswert, Frauen zu rekrutieren:  $E_{(F)} = 0.6$ ; Erwartungswert, Teilnehmer für Zentrum 1 zu rekrutieren:  $E_{(Z1)} = 0.3 \rightarrow$  Erwartungswert, Frauen in Zentrum 1 zu rekrutieren:  $E_{(F \cap Z1)} = 0.6 * 0.3 = 0.18$ ).

Unter Einbindung des SAS-Makros „Ransch“ von M.G. Kundu und A. Roy [2], das die SAS-Prozedur PROC PLAN nutzt, werden entsprechend der definierten Blocklängen je Stratum permutierte geblockte Randomisierungslisten erzeugt, die deutlich länger sind, als Rekrutierungen erwartet werden.

Die Studiensituation wird simuliert, indem diese Listen zufällig um den Erwartungswert abgebrochen werden unter der Bedingung, dass die geplante Gesamtfallzahl eingehalten wird (siehe Abbildung 1). Der zufällige Abbruch wird mit einer normalverteilten Zufallszahl realisiert (Erwartungswert: siehe oben; die Standardabweichung kann gewählt werden). Für den Fall, dass – insbesondere bei Strata mit kleinen Erwartungswerten – negative Zufallszahlen gezogen werden, wird der Betrag für den Abbruch verwendet. Nach ganzzahliger Rundung können die Listen also mindestens einen Teilnehmer haben oder leer bleiben.



**Abbildung 1:** Schema zum erwarteten und simulierten Abbruch der Randomisierungslisten bei der stratifizierten, permutierten Blockrandomisierung

Bevor der sukzessive Listenabbruch beginnt, werden die Strata zufällig angeordnet. Die Abbruchschleife über die Strata wird beendet, sobald die geplante Gesamtfallzahl erreicht oder überschritten wird. Bei Überschreitung wird die letzte Ziehung verworfen und die Differenz zur Gesamtfallzahl zugewiesen. Gegebenenfalls nachfolgende Strata bleiben leer; es wird also simuliert, dass hier keine Studienteilnehmer rekrutiert werden. Ist die Gesamtfallzahl vor Abbruch des letzten Stratums noch nicht erreicht, so wird automatisch die Differenz zur Gesamtfallzahl zugewiesen. Durch die zufällige Anordnung der Strata sind alle Strata gleichermaßen von dem beschriebenen Vorgehen beim Abbruch der Listen betroffen.

Bei jedem Simulationslauf werden neue Randomisierungslisten erzeugt und mit neuen Zufallszahlen abgebrochen.

### 3 Makro-Beschreibung

Die Realisierung der Simulationen zur permutierten stratifizierten Blockrandomisierung wurde mit SAS-Makros programmiert. Dabei wurde zur Umsetzung der Randomisierung das Makro „Ransch“ von M.G. Kundu und A. Roy [2] verwendet. Die anderen Makros wurden im Rahmen einer BSc.-Arbeit programmiert [1] und ergänzt.

#### Übersicht der Makros

- Balanceverhalten: Ablaufsteuerung der Simulationen
- Ransch: Permutierte Blockrandomisierung je Stratum
- Abbruch: Zufälliger Abbruch aller Randomisierungslisten je Simulation
- Vorbereitung: Datenvorbereitung für die Auswertung
- Auswertung: Häufigkeitsanalysen und Graphiken für die Gesamtauswertung und je Stratum

Das Makro „Balanceverhalten“ ist das übergeordnete Programm, das bei jeder Simulation die Makros Ransch, Abbruch und Vorbereitung nacheinander aufruft. Die Pflichtparameter für den Programmablauf werden hier als Makrovariablen eingelesen und auf Plausibilität geprüft.

**Tabelle 1:** Angaben zum Studienbeispiel

Variablenname	Bedeutung
behandlungen	Namen der Behandlungen
verhaeltnis	Zuordnungsverhältnis der Behandlungen
blocklaengen	Angabe der Blocklänge pro Stratum, mehrere variierende Blocklängen sind möglich
stratanamen	Namen der Strata
anzahlPat	Anzahl Pat., die pro Stratum randomisiert werden (ausreichende Zahl für die Realisierung des zufälligen Abbruchs wählen)
studentitel	Titel der Studie

teilnehmer	Anzahl Teilnehmer, die insgesamt an der Studie teilnehmen
erwstr	erwartete Anteile in Prozent für die Rekrutierung pro Stratum
sd	Standardabweichung, um diesen Parameter schwanken die erwarteten Längen der Randomisierungslisten pro Stratum
simulationen	Anzahl der Simulationen

Je Stratum wird mit dem Makro „Ransch“ entsprechend der angegebenen Blocklängen die Randomisierung mit PROC PLAN durchgeführt. Die Länge der Randomisierungslisten ist für alle Strata mit der Variable „anzahlPat“ vorgegeben und muss ausreichend groß sein, um den zufälligen Listenabbruch realisieren zu können. Je Stratum wird eine Randomisierungsliste, d.h. ein Dataset erstellt.

Diese Randomisierungslisten werden nun im Makro „Abbruch“ zufällig gekürzt. Im ersten Schritt werden die Strata mit dem Verfahren ‚Ziehen ohne Zurücklegen‘ zufällig angeordnet. Die SAS-Funktion RANUNI, die gleichverteilte Zufallszahlen erzeugt, wird dabei verwendet. Die neue Anordnung der Strata wird in Makrovariablen mit laufender Nummer gespeichert und in dieser Reihenfolge bearbeitet. Im zweiten Schritt wird mit der SAS-Funktion RANNOR für jedes Stratum unter Normalverteilungsannahme ein zufälliger Parameter erzeugt, der mit dem Erwartungswert des jeweiligen Stratums und der angegebenen Standardabweichung die zufällige Länge der Randomisierungsliste festlegt.

Folgender Programmauszug aus dem Makro „Abbruch“ zeigt das Vorgehen. Die Makrovariable &b beschreibt die Nummer des bearbeiteten Stratums. Welches Stratum aktuell bearbeitet wird, ist in der Makrovariablen &&folge&b festgelegt. Die weiteren Makrovariablen sind in Tabelle 1 beschrieben und werden gegebenenfalls in Bezug zum aktuell bearbeiteten Stratum in der Berechnung berücksichtigt.

```

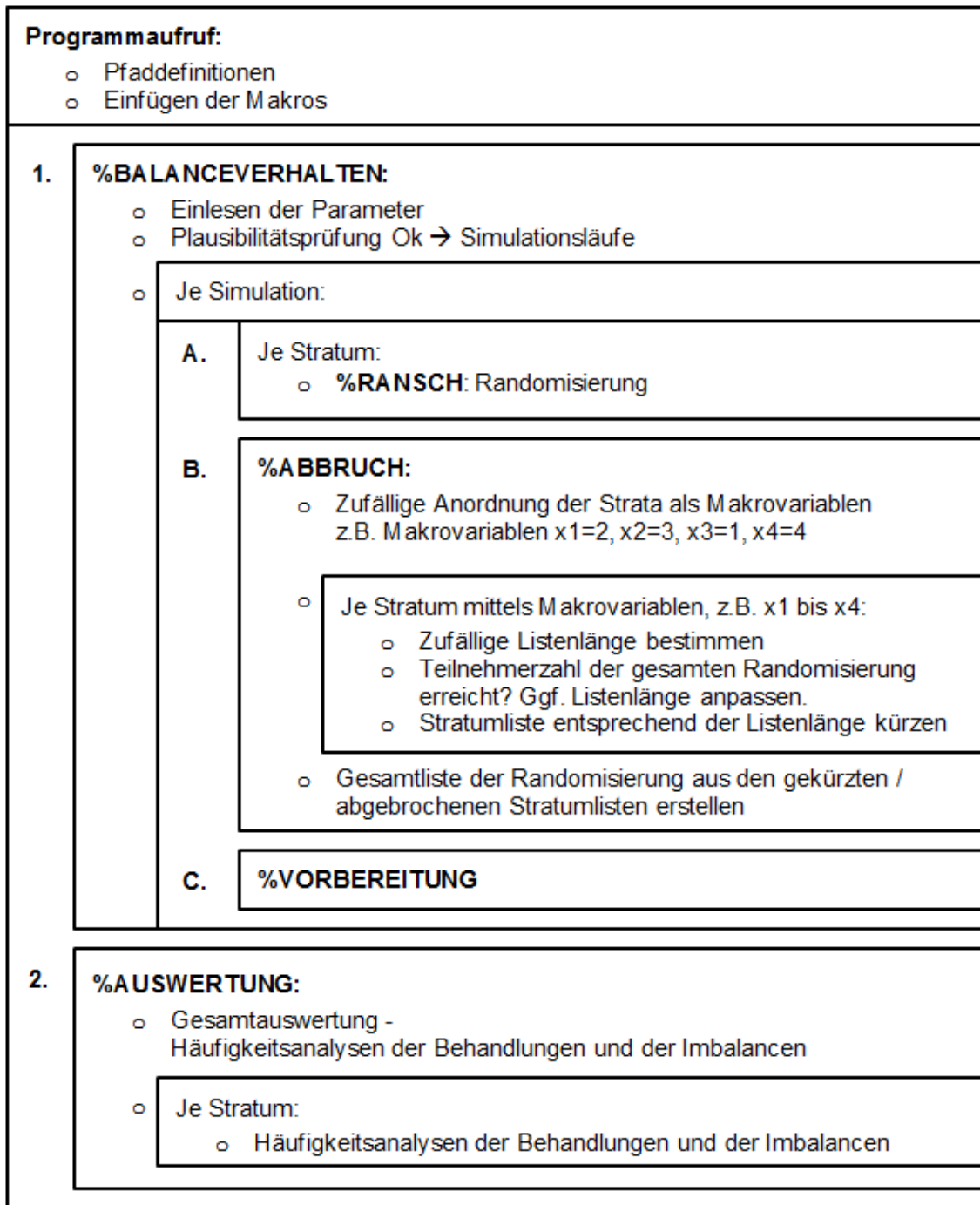
%* Zufälliger Abbruch um den Erwartungswert:
Anzahl Teilnehmer für das Stratum bestimmen;
%let y = %sysevalf(%sysfunc(RANNOR(&seed.)) *
%sysevalf(&sd.)+
%sysevalf((&&&erwstr&&folge&b./100)*&teilnehmer.));
%let abbruch = %sysfunc(abs(%sysfunc(round(&y.))));

```

Die Randomisierungsliste (SAS Dataset) des Stratums, die mit „Ransch“ erstellt wurde, wird dann im dritten Schritt mit der zufälligen Länge (Makrovariable „abbruch“) gekürzt. Bei jedem Stratum wird geprüft, ob die Gesamtteilnehmerzahl für die Randomisierung erreicht ist. Ist dies der Fall, so wird das entsprechende Stratum nicht mit der zufälligen Listenlänge gekürzt, sondern entsprechend der zur Gesamtteilnehmerzahl fehlenden Anzahl. Beim Abbruch der Randomisierungslisten der Strata können sowohl leere Listen als auch Listen mit wenigen Teilnehmern auftreten, die evtl. auch nur eine Behandlung im Stratum enthalten.

Für jede Simulation wird zum Abschluss des Programmblocks „Balanceverhalten“ die Datenvorbereitung zur Auswertung im Makro „Vorbereitung“ bearbeitet.

In einem weiteren Programmblock werden für die „Auswertung“ im gleichnamigen Makro Häufigkeitsanalysen und Graphiken zur Gesamtbewertung und die Bewertung je Stratum erstellt. Dabei werden Tabellen und Graphiken für die Behandlungen und Imbalancen mit ODS RTF und variablem Pfad als File ausgegeben. Den beschriebenen Programmablauf zeigt Abbildung 2.



**Abbildung 2:** Programmstruktur und –ablauf zum Balanceverhalten bei der stratifizierten Blockrandomisierung

## 4 Studienbeispiel und Ergebnisse

### 4.1 Angaben zum Simulationsbeispiel

Zu Demonstrationszwecken wird ein einfaches Studienbeispiel gewählt mit einer geplanten Fallzahl von 150 Teilnehmern und zwei Stratifizierungsmerkmalen Zentrum und Geschlecht mit je zwei Ausprägungen, so dass sich also insgesamt „nur“ vier Strata ergeben. (Anmerkung: Im Arbeitsalltag ist das Makropaket insbesondere für Studien mit vielen Strata gedacht.)

Die Studienbehandlungen A und B sollen im Verhältnis 1:1 zufällig zugeteilt werden. Die Simulierung der Studienrekrutierung wird 1000mal wiederholt. Der erwartete Anteil Studienteilnehmer pro Zentrum und Geschlecht ist in Tabelle 2 dargestellt.

**Tabelle 2:** Angaben zum Studienbeispiel

Stratifizierungsvariablen	Ausprägungen	Erwarteter Anteil Studienteilnehmer (%)
Zentren	Zentrum 1	70%
	Zentrum 2	30%
Geschlecht	männlich	40%
	weiblich	60%

Aus den erwarteten Anteilen pro Stratifizierungsmerkmal werden die erwarteten Anteile pro Stratum (unter Annahme der Unabhängigkeit) als Produkt aus den Wahrscheinlichkeiten ihrer jeweiligen Ausprägungen errechnet [Bsp. Stratum 1:  $0.7 \cdot 0.4 = 0.28$ ], siehe Tabelle 3. Auf Grundlage der resultierenden absoluten Fallzahlen werden pro Schicht geeignet scheinende Blocklängen festgelegt. Der zufällige Listenabbruch wird mit einer gewählten Standardabweichung von 5 für alle Strata durchgeführt.

**Tabelle 3:** Erwartete Fallzahlen in den Strata und gewählte Blocklängen

Stratum Nr.	Stratum-Bezeichnung	Erwarteter Anteil in %	Erwartete, absolute Fallzahlen	Gewählte (permutierte) Blocklängen
1	Zentrum 1 / männlich	28%	42	6 / 8
2	Zentrum 1 / weiblich	42%	63	6 / 8
3	Zentrum 2 / männlich	12%	18	4 / 6
4	Zentrum 2 / weiblich	18%	27	4 / 6
Gesamt		100%	150	

## 4.2 Aufruf der Makros

```

* Simulationen zur Randomisierung;
*-----;
%balanceverhalten(behandlungen=A: Test/B: Placebo,
                  verhaeltnis=1 1,
                  blocklaengen=6 8/6 8/4 6/4 6,
                  stratanamen=Zentrum1 maennlich/Zentrum1 weiblich/
                              Zentrum2 maennlich/Zentrum2 weiblich,
                  anzahlPat=200, studientitel=Beispiel-Studie,
                  teilnehmer=150, erwstr=28/42/12/18,
                  sd=5,          simulationen=1000);

* Auswertung der Simulationen;
*-----;
* Ergebnisdateien: Gesamt.rtf und;
*           je Stratum StrataNr.rtf;
* pfad: Ausgabeordner;
%auswertung(pfad=T:\...\Randomisierung\Simulationen,
             behandlungen=A: Test/B: Placebo,
             stratanamen= Zentrum1 maennlich/Zentrum1 weiblich/
                          Zentrum2 maennlich/Zentrum2 weiblich);
    
```

**Abbildung 3:** Aufruf der SAS-Makros „Balanceverhalten“ und „Auswertung“ (Parameterangaben, siehe Kapitel 3)

## 4.3 Ergebnisse

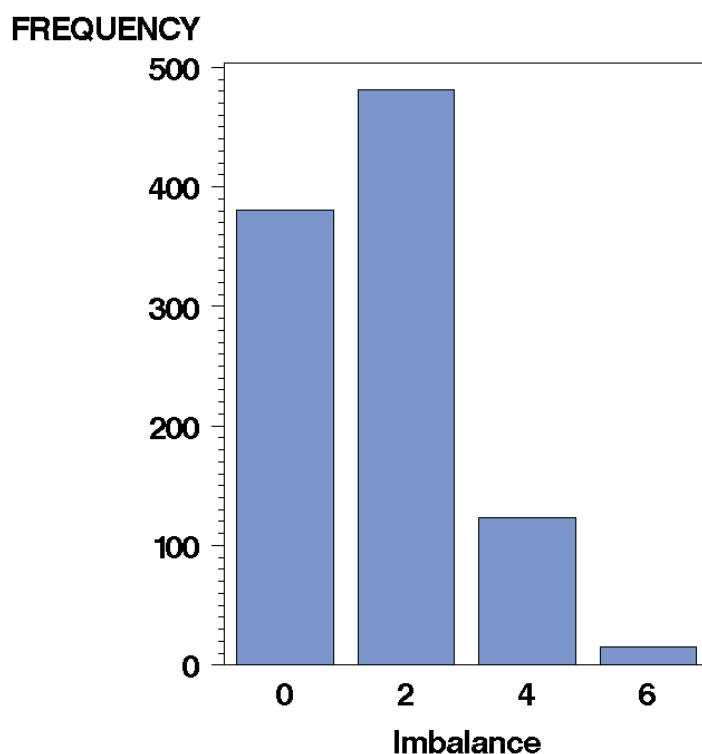
Nach 1000 Simulationen ergibt sich die in Tabelle 4 dargestellte Verteilung für Behandlung A. Wie erwartet, liegt der Schwerpunkt bei N=75, also der Hälfte der geplanten Studienteilnehmerzahl.

**Tabelle 4:** Häufigkeiten der Behandlung A im Gesamtkollektiv von 150 Studienteilnehmern und 1000 Simulationen

Behandlung A	Frequency	Percent	Cumulative Frequency	Cumulative Percent
72	9	0.90	9	0.90
73	63	6.30	72	7.20
74	243	24.30	315	31.50
75	381	38.10	696	69.60
76	238	23.80	934	93.40
77	60	6.00	994	99.40
78	6	0.60	1000	100.00



Daraus ergeben sich folgende Wahrscheinlichkeiten für (absolute) Imbalancen bezogen auf das Gesamtkollektiv (Abbildung 4 und Tabelle 5):



**Abbildung 4:** Verteilung der Imbalance im Gesamtkollektiv bei 1000 Simulationen

**Tabelle 5:** Imbalance im Gesamtkollektiv von 150 Studienteilnehmern und 1000 Simulationen

Imbalance	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	381	38.10	381	38.10
2	481	48.10	862	86.20
4	123	12.30	985	98.50
6	15	1.50	1000	100.00

Unter der Annahme, dass die tatsächlichen Fallzahlen in den Schichten zufällig mit einer Standardabweichung von 5 um die erwarteten Häufigkeiten streuen, ist also bei Wahl der oben angegebenen Blockgrößen die Wahrscheinlichkeit für eine Imbalance  $\geq 6$  etwa 1.5%.

Es zeigt sich, dass die maximal mögliche Imbalance von 14 bei 1000 Simulationen nicht vorkommt. Die maximale Imbalance würde sich ergeben, wenn bei Rekrutierungsende nach 150 Studienteilnehmern in jedem Stratum der letzte Block immer der größte und je zur Hälfte mit immer der gleichen Therapie angebrochen wäre. Für das Beispiel ergäbe sich ein Maximum von  $4+4+3+3=14$ ; das bedeutet, dass 82 Teilnehmer zur einen und 68 Teilnehmer zur anderen Studienbehandlung zugewiesen worden wären.

## 5 Diskussion und Ausblick

Mit Hilfe des entwickelten Makro-Pakets steht ein Tool zur Verfügung, das mit den Funktionen RANNOR und RANUNI sowie der Prozedur PLAN die stratifizierte permutierte Blockrandomisierung für unterschiedliche Studiensituationen simulieren kann. Wie im institutseigenen Randomisierungsprogramm ROM [3] können für jedes Stratum unterschiedliche Blocklängen angegeben werden. Für den zufälligen Listenabbruch lassen sich stratumspezifische Erwartungswerte definieren. Auch die Möglichkeit, leere Listen zu ziehen, ist gegeben.

Die resultierenden Wahrscheinlichkeiten für Imbalancen unterstützen die Entscheidung für die endgültige Festlegung der Blocklängen. Mit den simulierten Imbalance Ergebnissen kann die Fallzahlplanung, die im Studienprotokoll beschrieben ist, rekapituliert und die verbleibende Power abgeschätzt werden.

Die Anwendung ist insbesondere für Studien mit verhältnismäßig vielen Strata geplant. Nach ICH E9 [4] wird empfohlen, bei multizentrischen Studien nach Zentrum zu schichten. Zusätzlich kann gerade in kleinen Studien die Stratifizierung nach mindestens einem weiteren wichtigen prognostischen Faktor nötig sein.

Zurzeit ist nur die Simulierung von Studien mit zwei Behandlungsgruppen möglich. Eine Erweiterung auf drei und mehr Gruppen ist geplant. Außerdem bestehen Überlegungen, die Normalverteilung für den zufälligen Listenabbruch durch eine „gestutzte Normalverteilung“ [5] zu ersetzen, so dass Zufallsziehungen kleiner Null per se nicht mehr auftreten können.

Die Größe der Standardabweichung kann – abhängig von der Zuverlässigkeit der erwarteten Fallzahlen – gewählt werden. Hier ist geplant, nicht nur global einen einheitlichen, sondern für jedes Stratum einen separaten Wert angeben zu können.

Die Laufzeit des Programms liegt im vorliegenden Beispiel bei einer Simulation mit 1000 Wiederholungen bei ca. 1,5 Stunden. Sie ist sowohl von der Anzahl der zu bearbeitenden Strata abhängig, als auch von der Programmstruktur, die viele SAS Datasets benutzt. Hier sind durch Optimierung der Dateistrukturen sicherlich noch Laufzeitverbesserungen zu erreichen.

Weitere Auswertungen, z. B. die Untersuchung des Balance-Verhaltens in den Ausprägungsstufen der Stratifizierungsvariablen (z.B. Balance bei den Frauen / bei den Männern / je Zentrum) sind mit Hilfe der erzeugten SAS Datasets möglich. Es ist geplant, diese in die Standardauswertung zu integrieren.

## Literatur

- [1] L. Hupperz: Untersuchung zum Balanceverhalten der stratifizierten Blockrandomisierung. BSc.-Arbeit, Fachhochschule Koblenz/Remagen; 2012
- [2] M.G. Kundu, A. Roy: A unified SAS macro for generating randomisation schedule. In.: PharmaSUG 2007 Proceedings; 2007  
(<http://www.lexjansen.com/pharmasug/2007/ad/ad07.pdf>)
- [3] F. Rohlmann, R. Muche, L. Goldschmidt: Randomisierung in klinischen Studien. Praktische Umsetzung mit dem Randomisationsprogramm ROM. In: B. Schweizer, C. Großmann, M. Meule, W. Gaus (Hrsg.): Dokumentation – der Schritt ins 3. Jahrtausend – 8. DVMD-Tagung April 2004 in Ulm, Universitätsverlag Ulm, Ulm (2004), 168-171
- [4] EMEA European Medicines Agency: ICH E9 – Note for Guidance on Statistical Principles in Clinical Trials. London 1998  
([http://www.emea.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500002928.pdf](http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002928.pdf))
- [5] J. Hartung: Statistik – Lehr- und Handbuch der angewandten Statistik. 15. Auflage. München, Wien und Oldenbourg 2009