

Einfluss fehlender Daten auf Analyse und Fallzahlplanung in Bioäquivalenzstudien

Henrike Häbel
Fakultät für Mathematik und
Wirtschaftswissenschaften,
Universität Ulm

Albert-Einstein-Allee 11
89081 Ulm
henrike.haebel@uni-ulm.de

Julia Habeck
Boehringer Ingelheim
Pharma GmbH & Co KG.
MDS / Phase I/IIa
Biostatistics

Birkendorfer Straße 65
88397 Biberach
julia.habeck@
boehringer-ingelheim.com

Michaela Mattheus
Boehringer Ingelheim Pharma
GmbH & Co KG.
MDS / Phase I/IIa Biostatistics
Binger Straße 173
55216 Ingelheim am Rhein
michaela.mattheus@
boehringer-ingelheim.com

Zusammenfassung

Wie in allen klinischen Studien, kann es auch bei Bioäquivalenzstudien vorkommen, dass Werte pharmakokinetischer Parameter nicht in die Auswertungen eingeschlossen werden können (z.B. im Falle schwerwiegender Protokollverletzungen) oder diese Werte möglicherweise nicht gemessen wurden (z.B. bei vorzeitigem Ausscheiden eines Probanden aus der Studie). Hierbei stellt sich die Frage, bei wie vielen fehlenden Werten weitere Probanden nachrekrutiert werden müssen (vorausgesetzt dies ist logistisch durchführbar). Im Rahmen dieser Fragestellung kann in Simulationsstudien der Einfluss fehlender Daten auf die Variabilität der Ergebnisse untersucht werden.

Die vorliegenden Analysen beruhen auf Simulationen von normalverteilten Daten in 2x2-Crossover Studien. Bei diesen Simulationen wurde der Probandeneffekt als zufällig und normalverteilt angenommen. Des Weiteren wurde vorausgesetzt, dass von jedem Probanden höchstens ein Wert fehlt. Es wird nur uninformatives Fehlen betrachtet, das heißt, das Fehlen ist nicht von dem potentiellen unbeobachteten Wert abhängig. Ein Teil der Untersuchungen konzentriert sich auf nahezu balanciertes Fehlen mit möglichst gleich vielen fehlenden Werten für jede Behandlung in beiden Perioden.

Zudem werden Datensätze betrachtet, in denen die Werte vorwiegend in Periode 2 oder ausschließlich für eine Behandlung fehlen. Zu erkennen ist, dass der gCV des geschätzten Behandlungseffekts, betrachtet über alle Simulationen, mit zunehmender Anzahl fehlender Werte zunächst langsam, dann rapide ansteigt.

Anschließend wurde eine Fallzahlbetrachtung mit SAS[®] durchgeführt. Die dabei erhaltenen Ergebnisse für die Power einer Studie können schließlich bei der Entscheidung, ob weitere Probanden nachrekrutiert werden müssen, herangezogen werden.

Schlüsselwörter: Bioäquivalenzstudien, unvollständige Datensätze, PROC SURVEYSELECT, PROC MIXED, PROC POWER

1 Bioäquivalenzstudien

Als Pharmakokinetik bezeichnet man alle Prozesse der Reaktion (Absorption, Verteilung und Elimination) des Körpers auf eine Substanz. Somit spielen Konzentrations-Zeit-Verläufe in klinischen Studien eine wichtige Rolle. Zu den primären Zielvariablen zählen die Fläche unter der Plasmakonzentrationskurve AUC (Area Under the Curve) und die maximale Konzentration C_{\max} . Dabei wird anhand der AUC das Ausmaß und mit C_{\max} die Höhe der Absorption des Wirkstoffes gemessen. Die Zielvariablen werden auch pharmakokinetische (PK-) Parameter genannt.

In Bioäquivalenzstudien wird im Rahmen der klinischen Entwicklung eines Wirkstoffes getestet, ob Test- und Referenzformulierung bioäquivalent sind, mit anderen Worten, hinsichtlich ihrer pharmakokinetischen Eigenschaften nur unwesentlich voneinander abweichen. Die PK-Parameter müssen dabei nach Applikation von Test- und Referenzformulierung nicht identisch, sondern innerhalb definierter Grenzen vergleichbar sein. Die klassische Bioäquivalenzstudie dient somit dazu, anhand der pharmakokinetischen Surrogatparameter die Ähnlichkeit von verschiedenen Formulierungen in Bezug auf Sicherheit und Wirksamkeit zu untersuchen. So wird beispielsweise vor der Vermarktung einer verbraucherfreundlicheren Darreichungsform eines Medikaments getestet, ob der Wirkstoff in Form einer neu zusammengesetzten Tablette die gleichen pharmakokinetischen Eigenschaften aufweist wie die bisher produzierte Formulierung.

Das typische Design einer Bioäquivalenzstudie mit einer Test- (T) und einer Referenzformulierung (R) ist ein 2x2-Crossover. Hierbei kommt es auf zwei Aspekte an: (a) Der Behandlungsunterschied wird intra-individuell geschätzt, wodurch die Genauigkeit in der Regel höher wird als bei einem inter-individuellen Vergleich (paralleles Design). Bei der Durchführung werden Test- und Referenzformulierung in zwei Perioden gegeben, wobei zwischen den beiden Perioden ein definiertes Zeitfenster liegt, die sogenannte Wash-out Periode. Sie dient der klaren Abgrenzung der zwei Behandlungen und soll sogenannte Carry-over Effekte vermeiden.

(b) Vorab werden N Probanden jeweils zur Hälfte zufällig einer Sequenz RT oder einer zweiten Sequenz TR zugeordnet. Die Behandlung hängt anschließend von der jeweiligen Sequenz und Periode ab, wie Abbildung 1 veranschaulicht:

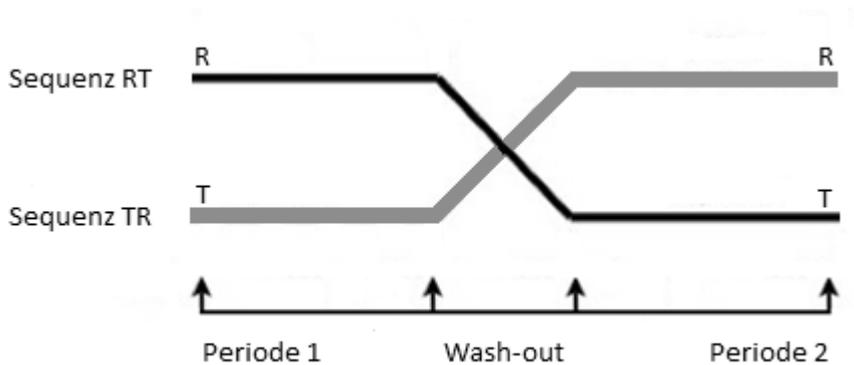


Abbildung 1: Darstellung eines 2x2-Crossovers

Dieses Studiendesign mit zwei Sequenzgruppen bewirkt, dass sich eventuell aufgetretene Periodeneffekte bei der Schätzung des Behandlungsunterschiedes im Mittel aufheben.

Für jeden Probanden werden die zu analysierenden PK-Parameter nach jeder Behandlung bestimmt. Falls der Wert eines PK-Parameters einer Behandlung aus der Analyse ausgeschlossen werden soll oder ganz fehlt, gibt es zwei mögliche Ansätze, wie fortzufahren ist: entweder (1) kann der jeweilige Proband ganz aus der Analyse ausgeschlossen werden, das heißt, die Werte aus allen beiden Behandlungen werden nicht berücksichtigt, oder (2) alle verfügbaren Daten des Probanden werden in die Analyse aufgenommen.

Hält man sich an die Alternative (1), so werden vorhandene Informationen ignoriert, was besonders dann kritisch ist, wenn dies auf mehrere Probanden $n < N$ zutrifft. In diesem Fall ist klar ersichtlich und kalkulierbar, dass die Power für den Nachweis von Bioäquivalenz derjenigen entspricht, die man bereits vor Beginn der Studie mit $N-n$ Probanden berechnet hätte. Mit Ansatz (2) ist dies jedoch offensichtlich nicht der Fall. Im Folgenden wird geprüft, ob und wie viele weitere Probanden bei Anwendung von Ansatz (2) nachrekrutiert werden müssen, um das Erreichen des Studienzieles zu gewährleisten.

In Simulationsstudien kann der Einfluss von fehlenden Daten auf die Analyse und Fallzahlplanung in Bioäquivalenzstudien untersucht werden. Diese Arbeit präsentiert die Durchführung, Auswertung und Ergebnisse der zu dieser Fragestellung durchgeführten Simulationsstudien.

2 Das Modell

An die ermittelten Werte der PK-Parameter wird nach Logarithmieren (natürlicher Logarithmus) ein lineares Modell angepasst. Dabei geht man von verschiedenen Einflussgrößen auf die jeweilige Zielvariable aus. So ist zu berücksichtigen, dass jeder Proband unterschiedlich auf die Behandlungen reagiert. Zudem können die Art der Behandlung, das heißt, die Verabreichung der Test- oder Referenzformulierung und der Zeitpunkt der Administration entscheidend sein.

Betrachtet man nun diese Faktoren, ergibt sich folgendes Modell für ein 2x2-Crossover Design nach Chow und Liu [1], mit einer auf der log-Skala normalverteilten Zielvariablen Y_{ijk} für $i = 1, \dots, N$ Probanden, $j = 1, 2$ Perioden und $k = 1, 2$ Sequenzen.

$$Y_{ijk} = S_{ik} + P_j + F_{jk} + C_k + e_{ijk}$$

mit

Y_{ijk} = logarithmierte Zielvariable des i -ten Probanden in der k -ten Sequenz in Periode j ; $i = 1, \dots, N$; $j = 1, 2$; $k = 1, 2$

S_{ik} = Effekt des i -ten Probanden in der k -ten Sequenz

P_j = Effekt in Periode j

F_{jk} = Behandlungseffekt in der k -ten Sequenz in Periode j

C_k = Effekt in der k -ten Sequenz

e_{ijk} = Residual-Fehler des i -ten Probanden in der k -ten Sequenz in Periode j

Die ursprünglich ermittelten Werte der PK-Parameter $\exp(Y_{ijk})$ sind logarithmisch normalverteilt.

2.1 Modellvoraussetzungen

Der Perioden-, Behandlungs- und Sequenzeffekt werden als fest vorausgesetzt. Der Effekt eines einzelnen Probanden sei eine Realisierung einer unabhängig und identisch $N(0, \tau^2)$ -verteilten Zufallsvariablen. Die Störgrößen beziehungsweise Residual-Fehler seien ebenfalls unabhängig und identisch verteilt mit $E(e_{ijk}) = 0$ und $\text{Var}(e_{ijk}) = \sigma^2$.

Der Probandeneffekt wird hier als zufällig angenommen. Unter diesen Voraussetzungen kann mittels der Restricted Maximum Likelihood (REML) Methode die Varianz τ^2 zwischen den Probanden und die Varianz σ^2 innerhalb eines Probanden auch dann geschätzt werden, wenn Werte fehlen sollten. Dies ermöglicht der bei der REML Methode implementierte Newton (-Raphson) Algorithmus, da mit diesem numerischen Verfahren aus der Klasse der Fixpunktiterationen nichtlineare Gleichungssysteme gelöst werden können.

Ein als fest angenommener Probandeneffekt würde dazu führen, dass ein Proband mit einem fehlenden Wert von der Analyse ausgeschlossen wird. Wie bereits geschildert,

sollen die simulierten Datensätze jedoch unter Berücksichtigung aller verfügbaren Werte analysiert werden.

2.2 Modellannahmen

Im Allgemeinen wurde $P_1 = F_{11} = F_{22} = C_1 = C_2 = 0$ gewählt. Somit ergibt sich für die logarithmierte Zielvariable des i -ten Probanden je nach Sequenz und Periode folgende einfache Zusammenstellung:

Tabelle 1: Zusammenstellung der logarithmierten Zielvariablen

Sequenz	Periode 1	Periode 2
RT ($k=1$)	$Y_{i11} = S_{i1} + e_{i11}$	$Y_{i21} = S_{i1} + P_2 + F_{21} + e_{i21}$
TR ($k=2$)	$Y_{i12} = S_{i2} + F_{12} + e_{i12}$	$Y_{i22} = S_{i2} + P_2 + e_{i22}$

Für die Größen der Effekte wurden zwei Gruppen gebildet und unterschiedliche Annahmen getroffen, die in den folgenden Unterkapiteln näher erläutert werden.

2.2.1 Modellannahmen zur Untersuchung wesentlicher Effekte

Im ersten Teil der Untersuchungen wurden Größen mit besonders deutlichen Effekten gewählt, um deren Einfluss klar erkennen zu können. In diesem Zusammenhang wurde $P_2 = 3$ und $F_{12} = F_{21} = F_2 = 5$ gewählt. Für den zufälligen Probandeneffekt wurden 3 verschiedene Varianzen τ^2 der Normalverteilung angenommen: $\tau^2 = 1, 9, 100$. Für die Residual-Fehler wurde eine Varianz von $\sigma^2 = 0.5$ gewählt.

2.2.2 Modellannahmen zur Untersuchung „realistischer Effekte“

Um Datensätze, die auch in der Realität vorkommen könnten, zu simulieren, wurden im zweiten Teil der Untersuchungen folgende Größen auf der Log-Skala angenommen: $P_2 = 0.2$, $F_2 = 0.1$ und $\tau^2 = 0.25$ sowie $\sigma^2 = 0.06$. Dies bedeutet auf der Original-Skala eine um 22% erhöhte Bioverfügbarkeit in der zweiten Periode und eine um 11% erhöhte Bioverfügbarkeit bei einer Behandlung mit der Testformulierung: der Wert der Zielvariablen erhöht sich durch Multiplikation mit einem Faktor von $\exp(0.2) = 1.22$ beziehungsweise $\exp(0.1) = 1.11$.

3 Simulation

Unter Berücksichtigung dieser Modellannahmen können nun mit SAS[®] vollständige und unvollständige Datensätze simuliert werden. Für die Simulationsstudien wurde zunächst ein vollständiger Datensatz für 24 Probanden, demnach mit 48 Werten, generiert. Dieser vollständige Datensatz wurde anschließend mit PROC MIXED analysiert, denn diese SAS[®] Prozedur verwendet zur Schätzung der gesuchten Varianzen τ^2 und σ^2 (oder anderer Größen von Interesse) die REML Methode.

Daraufhin wurden aus dem vollständigen Datensatz weitere 8 Datensätze mit jeweils 1, 2, 3, 4, 8, 12, 16 und 20 zufällig fehlenden Werten erstellt, um diese anschließend mit PROC MIXED zu analysieren.

PROC SURVEYSELECT bietet eine sehr gute Möglichkeit, zufällig Daten fehlen zu lassen, etwa

```
PROC SURVEYSELECT DATA = complete  
    METHOD = srs N = 4 OUT = incomplete;  
    STRATA stra;  
  
RUN;
```

PROC SURVEYSELECT zieht mit der Methode `srs` (simple random sampling) aus einem gegebenen Datensatz (hier: `complete`) zufällig `N` Werte ohne Zurücklegen und schreibt diese in einen neuen Datensatz (hier: `incomplete`). Mit der optionalen Funktion `STRATA` kann angegeben werden, aus welchen Teilmengen der gegebenen Daten Werte gezogen werden sollen.

Bei den Simulationen der unvollständigen Datensätze wurde berücksichtigt, dass von jedem Probanden mindestens ein Wert in die Analyse einfließt. Dabei wurden drei verschiedene Fälle für fehlende Werte betrachtet:

- Fehlen in beiden Behandlungen und in beiden Perioden (Fall I)
- Fehlen nur in Periode 2 (Fall II)
- Fehlen nur in Behandlung mit der Referenzformulierung (Fall III).

Abbildung 2 veranschaulicht den vollständigen Datensatz, der als Grundlage für die Simulation der drei gewählten Fälle verwendet wird. Dabei steht die Variable `PTNO` für die Nummer des Probanden, `TPATT` für die Sequenz, `PKPER` für die Periode, `PKTRT` für die Behandlung und `lnPK` für den logarithmierten Wert des PK-Parameters. Die Spalte `STRA` ist speziell für die Simulierung der unvollständigen Datensätze angelegt worden und gehört nicht zu einem typischen Datensatz einer Bioäquivalenzstudie.

Um die unvollständigen Datensätze für die bereits beschriebenen Fälle zu simulieren, wird der vollständige Datensatz zunächst durch die Variable `STRA` zerteilt, um das Ziehen von Werten auf bestimmte Bereiche (Strata) zu reduzieren. Anschließend werden die nicht vom Ziehen betroffenen Segmente und die nun nach Anwendung von PROC SURVEYSELECT unvollständigen Strata wieder zusammengeführt.

Im Fall I wurde für $m = 4, 8, 12, 16$ und 20 fehlende Werte darauf geachtet, dass gleich viele Werte pro Behandlung pro Periode vorhanden sind. Hierbei wurde, wie Abbildung 2 verdeutlicht, die Einteilung des vollständigen Datensatzes in die einzelnen durch Trennlinien gekennzeichneten Teile vorgenommen. Anschließend wurden aus den grauschraffierten Bereichen (weiterer Datensatz) dann $6-m/4$ Werte pro Stratum gezogen.

Hierfür wurde der Anweisung STRATA der Spaltenname `stra` übergeben. Der nach Anwendung von PROC SURVEYSELECT erhaltene Datensatz wurde anschließend mit dem Datensatz, dessen Inhalt hier unschraffiert dargestellt ist, verbunden. Für $m = 1$ wurden 47 von den insgesamt 48 Werten, für $m = 2$ wurden jeweils 23 von 24 pro Sequenz gezogen und für $m = 3$ wurden 5 Werte pro dunkel schraffiertem Stratum gezogen. Dieser Datensatz wurde analog zu dem bereits Beschriebenen mit dem Datensatz, dessen Inhalt hier unschraffiert dargestellt ist, verbunden.

Für die Fälle II und III mit fehlenden Werten ausschließlich in Periode 2 oder in der Referenzformulierung wird PROC SURVEYSELECT zur Weiterverarbeitung ein Datensatz übergeben, der nur Werte aus Periode 2 enthält (rechter Block in Abbildung 2) beziehungsweise der nur Werte für die Referenzbehandlung enthält (bestehend aus den beiden oberen Bereichen auf der linken Seite sowie den beiden unteren Bereichen auf der rechten Seite).

PTNO	TPATT	PKPER	PKTRT	InPK	STRA	PTNO	TPATT	PKPER	PKTRT	InPK	STRA
1	RT	1	R	X	1	1	RT	2	T	X	3
2	RT	1	R	X	1	2	RT	2	T	X	3
3	RT	1	R	X	1	3	RT	2	T	X	3
4	RT	1	R	X	1	4	RT	2	T	X	3
5	RT	1	R	X	1	5	RT	2	T	X	3
6	RT	1	R	X	1	6	RT	2	T	X	3
7	RT	1	R	X	1	7	RT	2	T	X	3
8	RT	1	R	X	1	8	RT	2	T	X	3
9	RT	1	R	X	1	9	RT	2	T	X	3
10	RT	1	R	X	1	10	RT	2	T	X	3
11	RT	1	R	X	1	11	RT	2	T	X	3
12	RT	1	R	X	1	12	RT	2	T	X	3
13	TR	1	T	X	2	13	TR	2	R	X	4
14	TR	1	T	X	2	14	TR	2	R	X	4
15	TR	1	T	X	2	15	TR	2	R	X	4
16	TR	1	T	X	2	16	TR	2	R	X	4
17	TR	1	T	X	2	17	TR	2	R	X	4
18	TR	1	T	X	2	18	TR	2	R	X	4
19	TR	1	T	X	2	19	TR	2	R	X	4
20	TR	1	T	X	2	20	TR	2	R	X	4
21	TR	1	T	X	2	21	TR	2	R	X	4
22	TR	1	T	X	2	22	TR	2	R	X	4
23	TR	1	T	X	2	23	TR	2	R	X	4
24	TR	1	T	X	2	24	TR	2	R	X	4

Abbildung 2: Vollständiger Datensatz

Die Simulation eines vollständigen Datensatzes und die darauf basierende Generierung der 8 unvollständigen Datensätze für jeden Fall von fehlenden Daten wurde 1000 Mal wiederholt.

Somit ergibt sich folgender Pseudo-Code:

```
DO 1 TO 1000;  
  1. Simuliere vollständigen Datensatz für 24 Probanden  
  2. Werte vollständigen Datensatz mit PROC MIXED aus  
  3. FOR m = 1, 2, 3, 4, 8, 12, 16, 20 DO  
      PROC SURVEYSELECT um Datensätze mit i fehlenden Werten auf  
      drei Varianten zu simulieren  
  END  
  4. Werte unvollständige Datensätze mit PROC MIXED aus  
END;
```

4 Analyse

Die simulierten Datensätze wurden wie reale Datensätze analysiert. Hierzu wurde in SAS[®], wie im vorherigen Abschnitt bereits erwähnt, PROC MIXED verwendet. Ein Rückgabewert von PROC MIXED ist der mittlere quadratische Fehler MSE (Mean Squared Error), welcher bei unserem Modell einen Schätzer für die Varianz σ^2 innerhalb eines Probanden darstellt. Der MSE ist in der Ausgabe von PROC MIXED in der Tabelle Covariance Parameter Estimates, als Wert von Estimate für Residual zu finden. Anschließend kann man mit dem MSE den geometrischen Variationskoeffizienten gCV (geometric Coefficient of Variation) berechnen, der wie folgt definiert ist

$$gCV := 100\sqrt{\exp(MSE) - 1}.$$

Diese Formel leitet sich gemäß Hauschke, Steinijans und Pigeot [4] aus der Definition des Variationskoeffizienten

$$CV := \frac{\sqrt{\text{Var}(X)}}{E(X)}$$

einer lognormalverteilten Zufallsvariablen X ab, für die gilt

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad \text{und} \quad \text{Var}(X) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1),$$

wobei der MSE als ein Schätzer für σ^2 eingesetzt wird. So ergibt sich für $\sigma^2 = 0.06$ ein gCV in Höhe von 25%. Der gCV wird von Naturwissenschaftlern für anschaulicher gehalten als der MSE, enthält aber die gleiche Information.

Der gCV ist eine entscheidende Größe bei der Fallzahlbestimmung. Intuitiv gilt, je höher der gCV ist, desto mehr Probanden müssen rekrutiert werden.

Die Berechnung der Fallzahl beziehungsweise der Power kann in SAS[®] mit PROC POWER durchgeführt werden.

4.1 Analyse mit PROC MIXED

Bei der Durchführung einer Bioäquivalenzstudie möchte man mittels des angepassten Modells zeigen, dass zwischen der Test- (T) und der Referenzformulierung (R) kein relevanter pharmakokinetischer Unterschied besteht.

Zur Untersuchung des Unterschieds zwischen den beiden Behandlungen μ_T und μ_R (in Kapitel 2.2 mit F_2 bezeichnet) wird eine Varianzanalyse mit der REML-Methode durchgeführt. Dabei werden die Variabilität zwischen den Probanden τ^2 und die Variabilität innerhalb eines Probanden σ^2 zwischen Test- und Referenzformulierung berücksichtigt. Bei einem klassischen Verfahren (Average Bioequivalence) wird überprüft, ob das 90%-Konfidenzintervall der Behandlungsdifferenz $\mu_T - \mu_R$ im Äquivalenzbereich von $[\ln(0.8), \ln(1.25)]$ liegt.

Anhand des Schätzers für $\mu_T - \mu_R$ und der Varianzen τ^2 und σ^2 kann ein Konfidenzintervall für die Behandlungsdifferenz ermittelt und folgende Hypothesen geprüft werden:

$$\begin{aligned} \text{Nullhypothese } H_0 : \quad & \mu_T - \mu_R \leq \ln(0.8) \quad \text{oder} \\ & \mu_T - \mu_R \geq \ln(1.25) \end{aligned}$$

versus

$$\text{Alternativhypothese } H_1 : \quad \ln(0.8) < \mu_T - \mu_R < \ln(1.25)$$

In SAS[®] kann hierfür die Prozedur PROC MIXED aufgerufen werden, etwa mit:

```
PROC MIXED DATA = data ORDER = INTERNAL
  CLASS ptno pktrt tpatt pkper;
  MODEL lnpk = pktrt pkper tpatt
    / HTYPE = 2 SOLUTION
    OUTPRED = outpred
    DDFM = KR;
  RANDOM ptno(tpatt) / S;
  LSMEANS pktrt pkper / CL;
```

```
ESTIMATE 'Diff T-R' pkrtrt -1 1 / CL ALPHA= 0.1;  
RUN;
```

Hierbei wird PROC MIXED zuerst ein zu analysierender Datensatz (data) übergeben, der mit der Methode REML analysiert werden soll, wobei ORDER = INTERNAL festlegt, dass die Werte für die Variablen in CLASS nach dem unformatierten Wert sortiert werden. Unter CLASS werden die qualitativen im Modell zu berücksichtigenden Variablen angegeben, diese sind in unserem Fall Probandennummer, Behandlung, Sequenz und Periode.

Die logarithmierte Zielvariable (lnkp) wird in Abhängigkeit der festen Effekte unter MODEL angegeben. Mit HTYPE = 2 wird angegeben, dass Typ-2 F-Tests für die fixen Effekte durchgeführt werden sollen. SOLUTION führt zur Ausgabe von Schätzern und Standardfehlern sowie der Berechnung einer t-Statistik für die festen Effekte. Mit OUTPRED wird nach dem Modellaufruf ein Datensatz ausgegeben, der unter anderem die geschätzten Werte enthält, und DDFM = KR legt fest, dass die Freiheitsgrade nach Kenward-Roger bestimmt werden sollen. Mit RANDOM wird der zufällige Probandeneffekt (innerhalb der Sequenz) übergeben, wobei S die Ausgabe von Schätzern und Standardfehlern sowie die Berechnung einer t-Statistik für die zufälligen Effekte hervorruft. Mit LSMEANS werden die Kleinste-Quadrate-Schätzer der festen Effekte sowie mit CL deren Standard-95%-Konfidenzgrenzen angefordert. Nach den Spezifikationen unter ESTIMATE wird die zu schätzende Behandlungsdifferenz inklusive des 90%-Konfidenzintervalls ausgegeben. Weitere Informationen können dem SAS/Stat 9.2 User's Guide [3] entnommen werden.

4.2 Ergebnisse

Die aus den 1000 Datensätzen nach anschließender Auswertung mit PROC MIXED erhaltenen MSEs und daraus berechneten gCVs wurden deskriptiv ausgewertet. Im Detail bedeutet dies, dass für Fall I, II und III die je 1000 geschätzten MSEs und berechneten gCVs gespeichert wurden (das heißt für die in Kapitel 3 beschriebenen Fälle I, II und III für fehlende Werte). Für jeden der drei Fälle wurde dann eine deskriptive Statistik erstellt. In diesem Kapitel erfolgt eine Fokussierung auf die wesentlichsten Ergebnisse und somit die Resultate aus Fall III (fehlende Werte ausschließlich in der Referenzbehandlung) mit Modellannahmen aus Kapitel 2.2.2. Tabelle 2 zeigt die deskriptive Statistik über die in diesem Szenario in 1000 Simulationen beobachteten gCVs.

Tabelle 2: Deskriptive Statistik über die beobachteten gCVs mit Annahmen aus 2.2.2 ($\sigma^2 = 0.06$), Fall III

Anzahl fehlender Werte	Anzahl Simulationen	Minimum	25% Perzentil	Mittelwert	75% Perzentil	Maximum
0	1000	13.765	22.001	24.569	26.828	37.540
1	1000	13.887	22.001	24.594	26.917	40.143
2	1000	13.887	21.820	24.536	27.045	39.052
3	1000	14.258	21.662	24.561	27.138	39.950
4	1000	5.301	19.552	24.317	28.520	48.217
8	1000	7.589	19.137	24.417	29.182	53.245
12	1000	4.040	16.892	24.018	29.659	68.274
16	1000	0.599	13.904	23.498	30.804	68.017
20	1000	0.0001	0.011	20.338	36.371	87.275

Um die geschätzten Werte anschaulich analysieren zu können, wurden Boxplot Diagramme erstellt, zum Beispiel für den eben beschriebenen Fall:

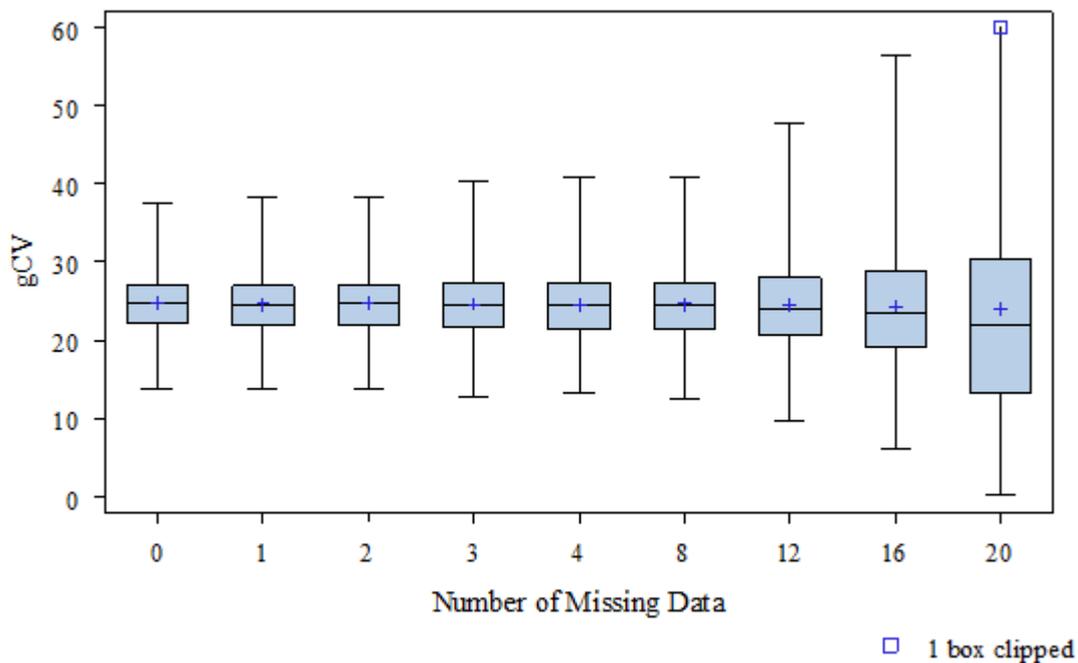


Abbildung 3: Median, Mittelwert, 25%-, 75%- Perzentile, Minimum und Maximum der beobachteten gCVs für Annahmen aus 2.2.2 ($\sigma^2 = 0.06$), Fall III

In Abbildung 3, wie auch für alle anderen durchgeführten Simulationen, ist zu erkennen, dass der gCV des geschätzten Behandlungseffekts, mit zunehmender Anzahl fehlender Werte zunächst langsam, dann rapide ansteigen kann.

5 Fallzahlplanung

Die Fallzahl in Bioäquivalenzstudien sollte ausreichend groß sein, sodass die im Abschnitt 4.1 beschriebene Nullhypothese, welche besagt, dass ein relevanter Unterschied zwischen den Behandlungen vorliegt, mit einer Wahrscheinlichkeit von mindestens 80% verworfen wird, vorausgesetzt, es liegt tatsächlich kein relevanter Unterschied vor. Mit anderen Worten bedeutet dies, dass das 90%-Konfidenzintervall des Quotienten der geometrischen Mittelwerte von Test- und Referenzformulierung μ_T/μ_R mit mindestens 80%-iger Wahrscheinlichkeit vollständig im Äquivalenzbereich von [0.80, 1.25] liegen sollte. Dies ist äquivalent zu der Aussage, dass das 90%-Konfidenzintervall der Behandlungsdifferenz $\mu_T - \mu_R$ auf der log-Skala vollständig im Äquivalenzbereich von $[\ln(0.8), \ln(1.25)]$ liegen sollte. Diese betrachtete Wahrscheinlichkeit wird auch Power genannt.

5.1 Fallzahlung mit Proc Power

Für die vorliegenden Auswertungen wurde in SAS[®] für eine Fallzahl von 24 Probanden zu den mit PROC MIXED geschätzten MSEs beziehungsweise gCVs die jeweilige Power wie folgt berechnet:

```
PROC POWER;
    PAIREDMEANS TEST = Equiv_Mult

        LOWER = 0.8

        UPPER = 1.25

        ALPHA = 0.05

        MEANRATIO = 1

        CV = 0.20      0.24      0.25

/*CV ergibt sich aus den Analysen und entspricht, hier von links
nach rechts, den Resultaten für 20; 16, 12, 8, 4 und 0-3 fehlende
Werte für  $\sigma^2 = 0.06$  im Fall III; vgl. Tabelle 2, Mittelwert*/

        CORR = 0.5

        POWER = .

        NPAIRS = 24;

        PLOT X = n MIN = 6 MAX = 30 VARY(linestyle,symbol)

        YOPTS = (ref = 0.8) XOPTS = (ref = 24)

        /DESCRIPTION = 'Plot of power by sample size';

RUN;
```

Die ersten Optionen legen die Art des Hypothesentests, den Äquivalenzbereich, das Testniveau und den mittleren erwarteten Behandlungsquotienten fest. Hierbei ist zu beachten, dass nicht mehr auf der log-Skala gerechnet wird. Deshalb werden die geschätzten gCVs als Variationskoeffizienten (CV), in diesem Fall die, die sich im Mittel für Annahmen aus 2.2.2 ($\sigma^2 = 0.06$) im Fall III mit fehlenden Werten ausschließlich in der Referenzbehandlung ergaben, angegeben.

Die Option CORR erfasst die Korrelation der Werte von Test- und Referenzformulierung bei einem beliebigen Probanden. Im Folgenden bezeichne wieder τ^2 die Varianz zwischen den Probanden und σ^2 die Varianz innerhalb eines Probanden. Hiermit folgt nach Patterson und Jones [2] unter Berücksichtigung der unabhängigen Verteilung von S_{ik} und e_{ijk} für alle $i = 1, \dots, N$ und $j, k = 1, 2$:

$$\text{Var}(Y_{ijk}) = \text{Var}(S_{ik} + e_{ijk}) = \tau^2 + \sigma^2$$

$$\text{Cov}(Y_{i1k}, Y_{i2k}) = \text{Cov}(S_{ik} + e_{i1k}, S_{ik} + e_{i2k}) = \text{Cov}(S_{ik}, S_{ik}) = \text{Var}(S_{ik}) = \tau^2$$

Somit ergibt sich die Korrelation der beiden gemessenen Werte eines Probanden nach Behandlung mit der Test- beziehungsweise Referenzformulierung unter Berücksichtigung der identischen Verteilung von Y_{i1k} und Y_{i2k} :

$$\rho = \text{cor}(Y_{i1k}, Y_{i2k}) = \frac{\text{Cov}(Y_{i1k}, Y_{i2k})}{\sqrt{\text{Var}(Y_{i1k})\text{Var}(Y_{i2k})}} = \frac{\text{Cov}(Y_{i1k}, Y_{i2k})}{\text{Var}(Y_{ijk})} = \frac{\tau^2}{\tau^2 + \sigma^2}$$

Für die angenommenen Größen mit $\tau^2 = 0.5$ und $\sigma^2 = 0.06$ würde sich eine Korrelation von $\rho \approx 0.806$ ergeben. Um eine verallgemeinerte, von den Modellannahmen unabhängige Aussage treffen zu können, wurde hier für die Berechnung der Power $\tau^2 = \sigma^2$ angenommen und somit $\rho = 0.5$ gewählt.

Des Weiteren kann PROC POWER entweder die Power zu einer gegebenen Fallzahl oder die Fallzahl zu einer gegebenen Power berechnen. Hier wird zu einer Fallzahl (NPAIRS) von 24 Probanden die Power (POWER) bestimmt.

Zusätzlich können die berechneten Werte mit der Option PLOT grafisch dargestellt werden. Abbildung 4 zeigt die so erstellte Grafik mit CV = 0.20, 0.24, 0.25 bei einem σ^2 von 0.06 für den Fall III mit fehlenden Werten ausschließlich in der Referenzbehandlung.

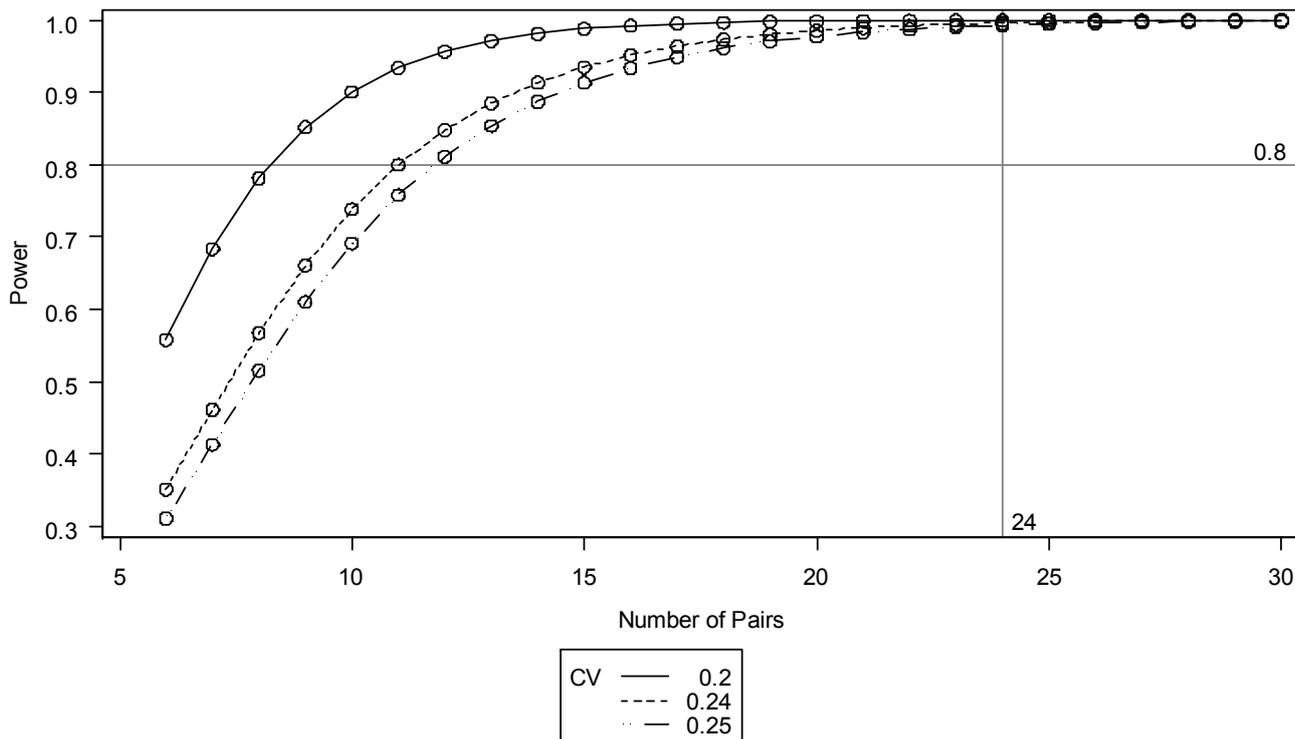


Abbildung 4: Darstellung der Power, basierend auf den mittleren gCVs aus 1000 Simulationen für die verschiedenen Szenarien der Anzahl fehlender Werte für Annahmen in 2.2.2 ($\sigma^2 = 0.06$) im Fall III

5.2 Ergebnisse

Mit SAS[®] wurde die Powerberechnung anhand der in den drei verschiedenen Fällen (von fehlenden Werten) beobachteten gCVs durchgeführt und grafisch ausgewertet.

Abbildung 5 zeigt die berechnete Power (gekennzeichnet durch Kreise) zu den gCVs im 75%-Perzentil der acht Szenarien von fehlenden Werten für die Referenzformulierung (Fall III) bei je 1000 Simulationen mit Annahmen aus 2.2.2 ($\sigma^2 = 0.06$) für eine Fallzahl von 24. Zusätzlich wurde die Power für vollständige Datensätze für eine Fallzahl von 6 bis 30 Probanden berechnet und auf einer Kurve zum Vergleich eingezeichnet. Anhand dieser Grafik kann zum Beispiel abgelesen werden, dass bei 24 Probanden und 20 fehlenden Werten nur eine effektive Fallzahl von 20 Probanden angenommen werden kann. Dementsprechend müssten 4 zusätzliche Probanden nachrekrutiert werden, um mindestens eine der Ausgangslage entsprechende Power zu erhalten, wobei in diesem Beispiel die Power bei 20 fehlenden Werten weiterhin mehr als 80% beträgt. Würden die 20 Probanden mit fehlenden Werten komplett von der Studie ausgeschlossen werden (entsprechend Variante (1) in Kapitel 1), so würden die übrigen 4 Probanden nicht genügen, und es müsste nachrekrutiert werden. Hier ist also deutlich ein Vorteil der in Kapitel 1 beschriebenen Variante (2) zu erkennen, bei der der Probandeneffekt als zufällig gewählt wird.

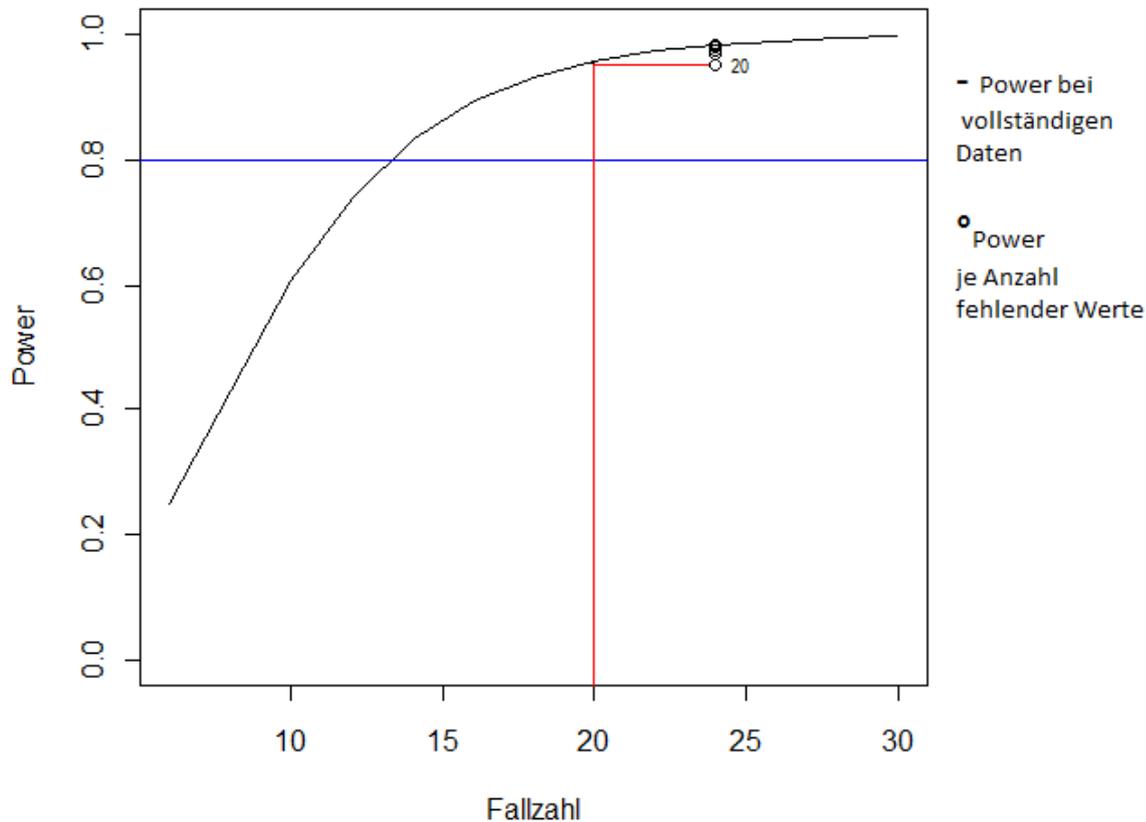


Abbildung 5: Darstellung der Power, basierend auf dem 75%-Perzentil des beobachteten gCVs aus 1000 Simulationen für $m=1, 2, 3, 4, 8, 12, 16, 20$ fehlende Werte unter Annahmen 2.2.2 ($\sigma^2 = 0.06$), Fall III; vgl. Tabelle 2, 75%-Perzentil

Diese Auswertungen wurden für beide Teile der Simulationsstudien, das heißt, zum einen anhand von Größen mit deutlichem Einfluss (Annahmen 2.2.1) und zum anderen anhand von in der Realität beobachtbaren Größen (Annahmen 2.2.2) durchgeführt, für alle drei Fälle der Verteilung von fehlenden Werten (Fall I, II und III), um eine allgemeine Antwort auf die Frage, bei wie vielen fehlenden Daten zusätzliche Probanden rekrutiert werden müssen, finden zu können.

6 Fazit

Die durchgeführten Simulationsstudien haben ergeben, dass bei 24 Probanden in einer Bioäquivalenzstudie mit einem 2x2-Crossover Design in den meisten Fällen keine weiteren Probanden nachrekrutiert werden müssen. Auf Basis der im 75%-Perzentil beobachteten gCVs sollte in manchen Fällen jedoch bei 10 bis 12 fehlenden Werten nachrekrutiert werden, besonders dann, wenn das Fehlen in beiden Behandlungen und Perioden auftritt.

Weiterführende Fragestellungen könnten sich mit Kohorten und anderen Kovariablen beschäftigen oder die Variabilität der geschätzten Restvarianz σ^2 in Betracht ziehen.

Literatur

- [1] S.C. Chow, J.P. Liu: Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel Dekker, Inc., New York, 1992
- [2] S. Patterson, B. Jones: Bioequivalence and Statistics in Clinical Pharmacology. Chapman & Hall / CRC, Boca Raton, 2006
- [3] SAS/Stat 9.2 User's Guide: The Mixed Procedure (Book Excerpt). SAS Publishing, 2008
- [4] D. Hauschke, V. Steinijans, I. Pigeot: Bioequivalence Studies in Drug Development. Wiley, West Sussex, England, 2007