

# Die stochastische Verteilung der intra-individuellen Varianz in 2x2x2 Cross-Over Bioverfügbarkeitsstudien

Tobias Bluhmki  
Fakultät Mathematik &  
Wirtschaftswissenschaften,  
Universität Ulm  
tobias.bluhmki@gmx.de

Julia Habeck  
Clinical Biostatistics,  
Boehringer Ingelheim GmbH & Co KG  
Birkendorfer Straße 65  
88397 Biberach/Riß  
julia.habeck@boehringer-  
ingelheim.com

Dr. Frank Fleischer  
Clinical Biostatistics,  
Boehringer Ingelheim GmbH & Co KG  
Birkendorfer Straße 65  
88397 Biberach/Riß  
frank.fleischer@boehringer-  
ingelheim.com

## Zusammenfassung

Bei der Planung und Durchführung von klinischen Bioverfügbarkeits- und Bioäquivalenzstudien mithilfe des Cross-Over Design spielt die intra-individuelle Varianz der zu untersuchenden Substanz eine zentrale Rolle. Insbesondere bei der Fallzahlplanung ist eine a-priori Schätzung der Varianz von Nöten. Hierbei stellt sich die Frage, welcher stochastischen Verteilung dieser Schätzer folgt und wie solch eine Verteilung mithilfe von simulierten Daten verifiziert werden kann. Darauf basierend ist es möglich Eigenschaften abzuleiten und eine statistische Testentscheidung zu konstruieren, die einen a-posteriori Vergleich der ursprünglichen Annahme zur Fallzahlplanung mit der aus den Daten geschätzten intra-individuellen Varianz zulässt.

**Schlüsselwörter:** klinische Studie, intra-individuelle Varianz, Bioverfügbarkeit, Bioäquivalenz, Cross-Over Design, GLM-Prozedur, MIXED-Prozedur, UNIVARIATE-Prozedur

## 1 Klinisch Pharmakologische (Phase I) Studien

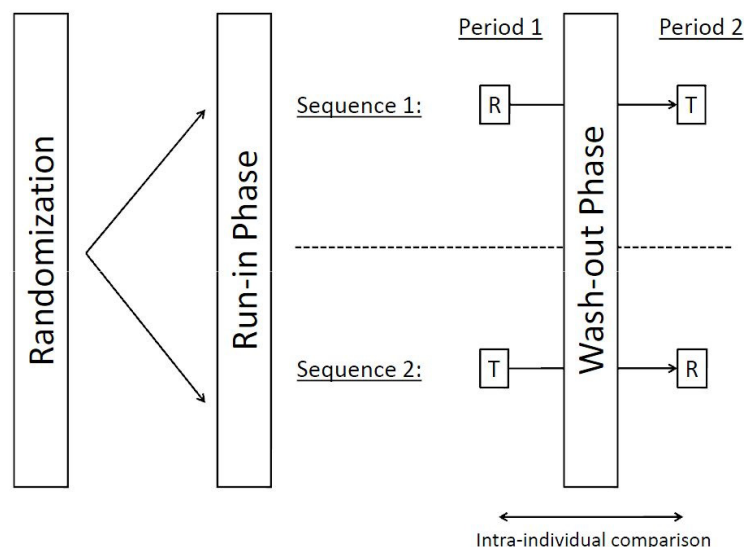
Die *Pharmakologie* beschreibt die biologische Aktivität (Zellantwort) von chemischen Substanzen auf den menschlichen Organismus, genauer, die Wechselwirkung(en) zwischen dem eingesetzten Wirkstoff und dem Organismus ([1], [5]). Die Ziele der frühen pharmakologischen Studien sind diese Assoziationen in Form von Verträglichkeit, Sicherheit und Toxizität zu erkennen und zu quantifizieren.

Ein wichtiger Baustein der Pharmakologie ist die Untersuchung der sog. *Pharmakokinetik* (PK) eines Wirkstoffes. Sie beschreibt den Einfluss des menschlichen Organismus auf das verabreichte Medikament, dessen Prozesse in die Aufnahme, Verteilung, Meta-

bolisierung und Ausscheidung (auch *ADME* genannt) des Medikamentes/Wirkstoffes aufgeteilt werden. Dafür wird für jeden Patienten ein Zeit-Wirkstoffkonzentrations-Profil erstellt und daraus aussagekräftige Parameter für die PK-Aktivität abgeleitet, um die o.g. Prozesse zu visualisieren und zu quantifizieren:

- *AUC* (engl. *Area under the curve*; Fläche unter der Konzentrationskurve): beschreibt die totale Freisetzung des Wirkstoffes ins Blut/Serum,
- *C<sub>max</sub>* (maximale Wirkstoffkonzentration im Blut/Serum): beschreibt die Geschwindigkeit/Rate der Absorption,
- v.a., siehe z.B. [1], [5].

Diese Parameter sind Repräsentanten dafür, in welchem Umfang und in welcher Geschwindigkeit ein nicht intravenös verabreichter Wirkstoff unverändert die Blutbahn erreicht. Diese Eigenschaft ist auch unter dem Begriff *Bioverfügbarkeit* (engl. *Bioavailability, BA*) bekannt und wird in sog. *Bioverfügbarkeitsstudien* untersucht (speziell in der Phase I am gesunden Probanden). Heutzutage ist der Test auf *Bioäquivalenz* (engl. *Bioequivalence, BE*) zweier Behandlungen von enormer Bedeutung. Im Zuge derer wird die Bioverfügbarkeit einer Test-Behandlung *T* mit der einer Referenz-Behandlung *R* verglichen, z.B. der Vergleich eines Generikums mit einer Standardtherapie. Das von den internationalen Behörden empfohlene Design für BA/BE Studien ist das sog. *Cross-Over Design*, insbesondere das 2x2x2 Cross-Over. Dabei wird jeder Proband zufällig einer Sequenz von zwei Behandlungen (*RT* oder *TR*) zugeordnet, wobei die Sequenz die Reihenfolge der zu verabreichenden Behandlungen festlegt. Da jeder Proband sowohl die Test- als auch die Referenzbehandlung erhält und folglich als seine eigene Kontrolle betrachtet werden kann, wird ein möglicher Behandlungsunterschied anhand von intra-individuellen Vergleichen quantifiziert (s. Abbildung 1). Im Gegensatz dazu basiert ein Parallelgruppen-Design auf inter-individuellen Vergleichen, d.h. auf den Unterschieden zwischen den Gruppen.



**Abbildung 1:** 2x2x2 Cross-Over Design: 2 Behandlungen, 2 Sequenzen, 2 Perioden

## 2 Statistische Auswertung einer 2x2x2 BA/BE Studie

### 2.1 Mathematische Modellierung (ohne Carry-Over Effekt)

Nach dem aktuellen Stand der Wissenschaft (vgl. z.B. [2], [3], [5]) wird davon ausgegangen, dass die PK-Parameter  $AUC$  und  $C_{max}$  einer logarithmischen Normalverteilung folgen. Deshalb ist es nach der Transformation der Daten von der *Original-Skala* (Roh-Daten) auf die sog. *Log-Skala* (logarithmierte Daten) möglich eine Varianzanalyse durchzuführen.

Seien im Folgenden  $Y_{ijk}$  die gemessenen (Roh-)Werte der zu betrachtenden pharmakokinetischen Parameter ( $AUC$  oder  $C_{max}$ ) von Proband  $i$  in Periode  $j$  und Sequenz  $k$  ( $i = 1, \dots, n_k$  und  $j, k = 1, 2$ ), wobei  $n_k$  die Anzahl der Probanden in der Sequenz  $k$  darstellt und die totale Fallzahl gegeben ist durch  $N = n_1 + n_2$ .

Seien außerdem  $Z_{ijk} = \log(Y_{ijk})$  die korrespondierenden Werte auf der Log-Skala, wobei  $\log$  den natürlichen Logarithmus darstellt.

Nach [3], [4] ist das zugrundeliegende mathematische Modell für das Standard 2x2x2 Cross-Over BA/BE Design (ohne Carry-Over Effekt) wie folgt definiert (hier: andere Parametrisierung):

$$Z_{ijk} = \mu_h + s_{ik} + \pi_j + \xi_k + e_{ijk}, \quad (1)$$

wobei

- $\mu_h$  der Mittelwert von Behandlung  $h \in \{T, R\}$ ,
- $s_{ik}$  der Probandeneffekt von Proband  $i$  in Sequenz  $k$ ,
- $\pi_j$  der feste Effekt von Periode  $j$ ,
- $\xi_k$  der feste Effekt von Sequenz  $k$ ,
- $e_{ijk}$  der unerklärte Fehler (von Proband  $i$  in Periode  $j$  und Sequenz  $k$ ) mit  $e_{ijk} \sim N(0, \sigma_W^2)$ , unabhängig und identisch verteilt.

Immer wenn ein Proband mehr als eine Behandlung über die Zeit hinweg erhält, wird die Variabilität, die durch die möglicherweise verschiedenen Reaktionen entsteht, im Modellfehler festgehalten und als *intra-individuelle Varianz*  $\sigma_W^2$  geführt.

Weiter gibt es zwei Möglichkeiten den Probandeneffekt  $s_{ik}$  zu spezifizieren:

Erstens ist es möglich den Probandeneffekt als einen *festen* Modelleffekt anzunehmen (führt zu einem sog. *Fixed-Effect Model*). Von den (inter-)nationalen Behörden wird jedoch empfohlen in diesem Studenumfeld den Probandeneffekt als *zufällig* anzunehmen, was zu einem sog. *Mixed Effect Model* führt. Zu allen obigen Modellannahmen muss dann zusätzlich gelten:

- $s_{ik} \sim N(0, \sigma_B^2)$  unabhängig und identisch verteilt, wobei  $\sigma_B^2$  die sog. *inter-individuelle Varianz* darstellt,
- $s_{ik}$  und  $e_{ijk}$  sind stochastisch unabhängige Zufallsvariablen.

Das Ziel einer Bioäquivalenzstudie ist nachzuweisen, dass sich die Behandlungseffekte  $\mu_h$  nicht signifikant voneinander unterscheiden, bzw. dass sich deren Differenz in einem vorher definierten Rahmen um Null bewegt. Von den Behörden wird gefordert, dass das Konfidenzintervall für den Punktschätzer der Differenz komplett innerhalb des Intervalls  $[\log(0.8), \log(1.25)]$  liegt (für mehr Informationen siehe Guidelines der *EMA* [10] und *FDA* [11]).

## 2.2 Statistische Auswertung

In den zwei folgenden Unterabschnitten wird kurz dargelegt, wie das *Fixed-* bzw. das *Mixed-Effect Model* mithilfe von SAS realisiert werden kann.

### 2.2.1 Modell mit festem Probandeneffekt und SAS PROC GLM

Ist der Probandeneffekt  $s_{ik}$  fest, ist eine klassische Varianzanalyse möglich, die basierend auf der Streuungszerlegung die Daten mithilfe von Quadratsummen statistisch auswertet. Dafür wird üblicherweise folgende Syntax verwendet:

```
PROC GLM DATA=data;
CLASS subject treatment sequence period;
MODEL logpk = treatment period sequence subject(sequence);
ESTIMATE 'Diff T-R' treatment -1 1;
TEST H=subject E=subject(sequence);
LSMEANS treatment / ADJUST=t PDIFF=CONTROL ("R") CL ALPHA=0.10;
RUN;
```

Beachte, dass hier der Probandeneffekt *subject(sequence)* als klassischer fester Modell-effekt behandelt wird. Der Schätzer für die intra-individuellen Varianz ist dabei durch den Wert *MeanSquare(Error)* (s. beispielhaft Tabelle 1) gegeben.

**Tabelle 1:** PROC GLM Output: Sum of Squares - Example

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	3.42425521	0.13697021	4.44	0.0004
Error	22	0.67894427	0.03086110		
Corrected Total	47	4.10319948			

Die Prozedur GLM ist jedoch nur effizient, wenn ein *kompletter* Datensatz zugrunde liegt, d.h. keine fehlenden Werte (engl. *missing values*) vorhanden sind. Sobald in einer Periode ein fehlender Wert auftritt, wird der komplette Proband aus der Analyse ausgeschlossen, was in der Realität schnell zu einer Verzerrung der Ergebnisse führen kann. Mehr Informationen und Erläuterungen bzgl. der GLM Prozedur sind in [7] aufgeführt.

## 2.2.2 Modell mit zufälligem Probandeneffekt und SAS PROC MIXED

Wird der Probandeneffekt  $s_{ik}$  als zufällig angenommen ist eine klassische Varianzanalyse wie in Abschnitt 2.2.1 nicht mehr möglich, da nun feste *und* zufällige Effekte im Modell auftauchen und folglich eine Berechnung von klassischen Quadratsummen und deren Auswertung im klassischen Sinne nicht mehr realisierbar sind.

Die allgemeinere Methode zur Schätzung der Parameter auf Basis des gemischten Modells beruht auf der sog. *Restricted/Residual Maximum Likelihood (REML)*-Methode (Berechnung von *General Least Square Estimates*, für mathematische Erläuterungen siehe z.B. [8]). Im Gegensatz zur GLM Prozedur ist die MIXED Prozedur in der Lage fehlende Werte angemessen zu schätzen, sodass auch Datensätze mit fehlenden Werten effizient statistisch ausgewertet werden können. Folgende Syntax wird in diesem Zusammenhang verwendet:

```
PROC MIXED DATA=data ORDER=INTERNAL;
CLASS subject treatment sequence period;
MODEL logpk = treatment period sequence / HTYPE = 2 SOLUTION
OUTPRED = outpred
DDFM = KR;
RANDOM subject(sequence)/S;
LSMEANS treatment period / CL;
ESTIMATE 'Diff T-R' treatment -1 1 /CL ALPHA = 0.1;
RUN;
```

Im Gegensatz zum Fixed-Effect Model ist hier der Probandeneffekt *subject(sequence)* als *zufällig* gesetzt. Ein Beispiel-Output für die Schätzer der Varianzen  $\sigma_B^2$  und  $\sigma_W^2$  ist in Tabelle 2 gegeben.

**Tabelle 2:** PROC MIXED Output: Covariance Parameter Estimates

Covariance Parameter Estimates	
Cov Parm	Estimate
ptno(tpatt)	0.04966
Residual	0.03086

Besitzt ein zu analysierender Datensatz keine fehlenden Werte, so sind die Schätzer für die intra-individuelle Varianz identisch mit denen aus der klassischen Varianzanalyse (vgl. z.B. [9]). Mehr Informationen und Erklärungen bzgl. der MIXED Prozedur, deren Optionen und der REML-Methode können in [6] nachgeschlagen werden.

## 3 Die Verteilung des Schätzers der intra-individuellen Varianz

Ziel dieses Kapitels wird sein, anhand des oben beschriebenen mathematischen Modells (1) eine stochastische Verteilung für den Schätzer der intra-individuellen Varianz herzuleiten.

In diesem Kapitel wird der o.g. Probandeneffekt  $s_{ik}$  vorläufig als fest und  $n_1 = n_2 = n$  gesetzt.

### 3.1 Mathematische Herleitung

Das grundlegende Prinzip der klassischen Varianzanalyse beruht auf der Zerlegung der totalen Variabilität in den Daten. Bekanntermaßen wird dabei die sog. totale Quadratsumme (engl. *Total Sum of Squares* ( $SS_{total}$ )) berechnet bzw. in die erklärenden Quadratsummen zerlegt. Da das Hauptaugenmerk dieses Artikels auf der intra-individuellen Varianz liegt, wird im Folgenden nur noch die  $SS_{intra}$  betrachtet (explizite Formeln für alle anderen Quadratsummen finden sich z.B. in [3], S. 66-69), welche sich wie folgt darstellen lässt:

$$SS_{intra} = \sum_{k=1}^2 \sum_{j=1}^2 \sum_{i=1}^n Z_{ijk}^2 - \sum_{k=1}^2 \sum_{i=1}^n \frac{Z_{i\cdot k}^2}{2} - \sum_{k=1}^2 \sum_{j=1}^2 \frac{Z_{jk}^2}{n} + \sum_{k=1}^2 \frac{Z_{\cdot k}^2}{N}, \quad (2)$$

wobei  $Z_{i\cdot k}$ ,  $Z_{\cdot jk}$  und  $Z_{\cdot k}$  die Summen der  $Z_{ijk}$  über die korrespondierenden Indices repräsentieren. Zusätzlich ist es möglich über die quadratische Form  $SS_{intra} = \vec{Z}^T A \vec{Z}$  nachzurechnen, dass die Transformationsmatrix  $A$  den Rang  $N - 2$  besitzt. Daraus ergibt sich direkt die mittlere quadratische Abweichung (engl. *Mean Square Error* ( $MSE$ )) für die intra-individuelle Varianz:

$$MSE_{intra} = \frac{1}{N-2} \cdot SS_{intra}.$$

Außerdem kann nachgerechnet werden, dass

$$\mathbb{E}(MSE_{intra}) = \widehat{\sigma_W^2} \quad (3)$$

gilt.

Durch den  $MSE_{intra}$  ist folglich ein unverzerrter Schätzer für die intra-individuelle Varianz des zugrundeliegenden mathematischen Modells (1) gegeben, d.h. es kann  $\widehat{\sigma_W^2} = MSE_{intra}$  gesetzt werden. Weiter definiere

$$Q \stackrel{\text{def}}{=} \frac{N-2}{\sigma^2} \cdot \widehat{\sigma_W^2} = \frac{1}{\sigma^2} \cdot SS_{intra},$$

wobei  $\sigma_W^2$  die wahre zugrundeliegende intra-individuelle Varianz von Modell (1) darstellt. Es kann nun gezeigt werden, dass  $Q$   $\chi^2$ -verteilt ist mit  $N - 2$  Freiheitsgraden. Weiter ist aus der elementaren Wahrscheinlichkeitsrechnung bekannt, dass die  $\chi^2$ -Verteilung ein Spezialfall der  $\Gamma$ -Verteilung ist, d.h. insbesondere gilt:

$$Q \stackrel{\text{def}}{=} \frac{N-2}{\sigma^2} \cdot \widehat{\sigma_W^2} \sim \chi_{N-2}^2 \Rightarrow Q \sim \Gamma\left(\frac{N-2}{2}, \frac{1}{2}\right).$$

Schließlich folgt aus den Eigenschaften der  $\Gamma$ -Verteilung für  $c := \frac{\sigma^2}{N-2}$ :

$$\widehat{\sigma}_W^2 = c \cdot Q = MSE_{intra} \sim \Gamma\left(\frac{N-2}{2}, \frac{N-2}{2\sigma_W^2}\right), \quad (4)$$

sodass der Schätzer der intra-individuellen Varianz einer  $\Gamma$ -Verteilung mit den Parametern  $p = \frac{N-2}{2}$  und  $b = \frac{N-2}{\sigma^2}$  folgt. Insbesondere hängt diese Verteilung nicht von den festen Modell-Effekten ab, sondern nur von der zugrunde liegenden Fallzahl und der wahren intra-individuellen Varianz.

Beachte, dass durch (4) weiterhin die Unverzerrtheit der Schätzung gegeben ist. Da die wahre zugrundeliegende Variabilität in fast allen Fällen unbekannt (und v.a. substanzabhängig) ist, ist es zwingend notwendig eine geeignete Variabilität zu wählen um die Verteilung des Schätzers explizit anzugeben. Diese kann z.B. gleich der Variabilität gesetzt werden, die für die Fallzahlplanung angenommen wurde. Dadurch wird es möglich die aus den Daten geschätzte Variabilität mit der angenommenen zu vergleichen (vgl. Kapitel 4).

## 3.2 Verifizierung durch Simulation

In diesem Abschnitt werden die mathematischen Erkenntnisse des vorhergegangenen Abschnittes anhand einer Simulation von PK-Daten (auf der Log-Skala) visualisiert und verifiziert.

### 3.2.1 Simulation von PK-Daten

Zuallererst ist es notwendig die pharmakokinetischen Daten zu simulieren, die den Anforderungen von Modell (1) genügen. Zur Vereinfachung werden der Sequenz- bzw. der Periodeneffekt als 0 angenommen. Außerdem ist es möglich, die Behandlungseffekte  $\mu_T$  und  $\mu_R$  nicht einzeln festzusetzen, sondern deren Differenz  $diff = \widetilde{\mu}_T \stackrel{\text{def}}{=} \mu_T - \mu_R$  als Input-Parameter zu wählen. Dies führt automatisch zu  $\widetilde{\mu}_R = 0$ .

Um die Simulation durchzuführen, benötigt man zusätzlich noch die vorher festgelegten Parameter der inter- und intra-individuellen Varianz, die zugrundeliegende Fallzahl und die Anzahl der Iterationen, was gesamt zu folgender Variablenliste führt:

- $\sigma_B^2 = varb$ ,
- $\sigma_W^2 = varw$ ,
- $\mu_T - \mu_R = diff$ ,
- $noit = \text{Anzahl d. Iterationen}$ ,
- $n = \text{Anzahl der Probanden pro Sequenz (folglich: } n_1 = n_2 = \frac{N}{2}\text{)}$ .

Betrachte nun die Simulation von PK-Daten für die erste Sequenz  $RT$ , dann gilt nach Definition (nach vorheriger Bestimmung der Kovarianzstruktur):

$$(\vec{Z})_i \stackrel{\text{def}}{=} \begin{pmatrix} Z_{i21} \\ Z_{i11} \end{pmatrix} \sim N \left( \begin{pmatrix} diff \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_B^2 + \sigma_W^2 & \sigma_B^2 \\ \sigma_B^2 & \sigma_B^2 + \sigma_W^2 \end{pmatrix} \right).$$

Folglich erhalten wir in der ersten Periode (wegen  $Z_{i11} \sim N(0, \sigma_B^2 + \sigma_W^2)$ ) für deren Realisierungen  $z_{i11} = \sqrt{\sigma_B^2 + \sigma_W^2} \cdot \omega_{i1}$ , wobei  $\omega_{i1}$  eine generierte Zufallszahl ist, die einer Standardnormalverteilung folgt.

Für die zweite Periode muss zusätzlich noch die Korrelation zwischen der ersten und der zweiten Periode berücksichtigt werden:

$$(Z_{i21} | Z_{i11} = z_{i11}) \sim N(diff + \rho \cdot z_{i11}, (\sigma_B^2 + \sigma_W^2) \cdot (1 - \rho^2)),$$

wobei  $\rho$  den Korrelationskoeffizienten zwischen den Beobachtungen von Proband  $i$  in Periode 1 und Periode 2 darstellt:

$$\rho = \text{Corr}(Z_{i21}, Z_{i11}) \stackrel{\text{def}}{=} \frac{\text{Cov}(Z_{i21}, Z_{i11})}{\sqrt{\text{Var}(Z_{i21}) \cdot \text{Var}(Z_{i11})}} = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

Deshalb kann der PK-Wert für die zweite Periode wie folgt simuliert werden:

$$z_{i21} = diff + \rho \cdot z_{i11} + \sqrt{\sigma_B^2 + \sigma_W^2} \cdot \epsilon_{i1} \cdot \sqrt{1 - \rho^2},$$

wobei  $\epsilon_{i1}$  wieder eine generierte Zufallszahl der Standardnormalverteilung darstellt. Ähnliche Argumentationen führen zu den folgenden Realisierungen der zweiten Sequenz:

$$z_{i12} = diff + \sqrt{\sigma_B^2 + \sigma_W^2} \cdot \omega_{i2},$$

$$z_{i21} = \rho \cdot (z_{i12} - diff) + \sqrt{\sigma_B^2 + \sigma_W^2} \cdot \epsilon_{i2} \cdot \sqrt{1 - \rho^2}.$$

Zusätzliche Data-Steps führen zu einem vollständigen logarithmierten PK-Datensatz (Tabelle 3), der alle Kriterien von Gleichung (2) erfüllt:

- *ptno*: Proband  $i$  in Sequenz  $k$  und Periode  $j$ ,
- *tpatt*: Sequenz- „Nummer“  $k$ ,
- *pker*: Periode  $j$ ,
- *pktrt*: verabreichte Behandlung in Periode  $j$  und Sequenz  $k$ ,
- *lnpk*: simulierter PK-Parameter  $Z_{ijk}$ .

Beachte, dass das Simulations-Makro keine fehlenden Werte produziert.



**Tabelle 3:** Beispiel – Simulierter PK-Datensatz mit  $n=4$

VIEWTABLE: Work.Data					
	ptno	tpatt	pkper	pktrt	lnpk
1	1	RT		1 R	0.0092886707
2	2	RT		1 R	-0.167155592
3	3	RT		1 R	-0.097319877
4	4	RT		1 R	0.6241108765
5	5	TR		1 T	0.239517135
6	6	TR		1 T	0.404121962
7	7	TR		1 T	-0.158745768
8	8	TR		1 T	-0.007670583
9	1	RT		2 T	-0.023681589
10	2	RT		2 T	-0.384325852
11	3	RT		2 T	-0.268332433
12	4	RT		2 T	0.5352977188
13	5	TR		2 R	0.5379186991
14	6	TR		2 R	-0.063270144
15	7	TR		2 R	0.2305229798
16	8	TR		2 R	-0.056140642

### 3.2.2 Extraktion des Schätzers für die intra-individuelle Varianz

Basierend auf einem simulierten Datensatz aus dem vorigen Abschnitt ist es nun möglich mittels der Prozeduren GLM und MIXED die Daten für den festen bzw. zufälligen Probandeneffekt  $s_{ik}$  auszuwerten und die jeweiligen Schätzer der intra-individuellen Varianzen für die vorher festgelegte Anzahl an Iterationen zu extrahieren. Schließlich erhält man den finalen Datensatz für die weiteren Analysen:

- *MeanSquare*:  $\widehat{\sigma}_W^2 = c \cdot Q$  aus der GLM Prozedur,
- *chi\_stat*: repräsentiert die Größe  $Q$  (vgl. Gleichung (4)), die aus der Variable *MeanSquare* und den Eingabeparametern gewonnen wird,
- *Estimate\_REML*:  $\widehat{\sigma}_W^2$  aus der MIXED Prozedur.

### 3.2.3 Statistische und grafische Analyse der geschätzten Varianzen

Die folgenden Ergebnisse sowie deren Interpretationen und grafischen Visualisierungen basieren auf folgenden (realitätsnahen) vordefinierten Makro-Parametern:

- Anzahl der Iterationen: 10000,
- Anzahl der Probanden pro Sequenz: 12 ,
- Behandlungsdifferenz: 0.05 auf der logarithmierten Skala,
- $\sigma_B^2 = \sigma_W^2 = 0.03922$ .

**Tabelle 4:** Ausgabe der Prozedur MEANS: *Estimate\_ANOVA* vs. *MeanSquare*

Analysis Variable : difference				
N	Mean	Std Dev	Minimum	Maximum
10000	0.000026393	0.000434354	0	0.0152504

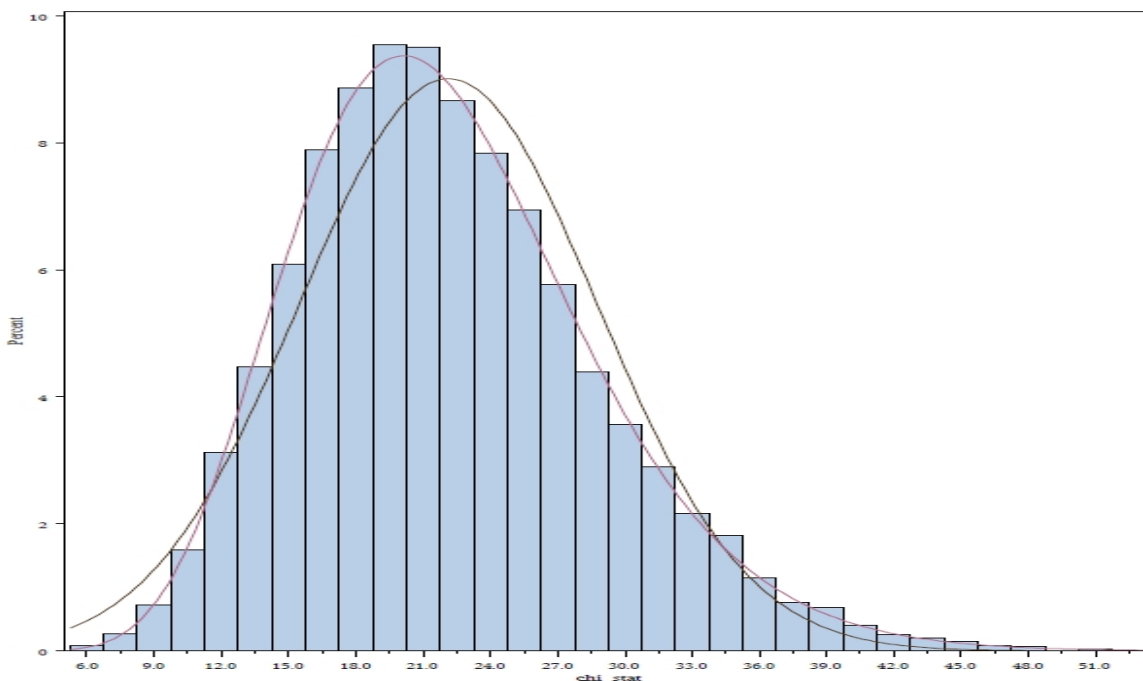
Zuerst werden die zwei verschiedenen Schätzer für die intra-individuelle Varianz durch die Prozedur MEANS miteinander verglichen. Dabei stellt sich heraus, dass die zwei Schätzer nahezu identisch sind (vgl. Tabelle 4 und Argumentationen aus Kapitel 2.2.2). Somit genügt es die hergeleiteten mathematischen Ergebnisse aus Kapitel 3.1 nur für die Variable *MeanSquare* zu verifizieren.

Um die zugrundeliegende  $\chi^2$ -Verteilung der Größe  $Q$  zu visualisieren, wird die UNIVARIATE Prozedur auf die Variable *chi\_stat* mit folgendem Code angewandt:

```
PROC UNIVARIATE data = finalset ;  
VAR chi_stat ;  
Histogram / normal gamma ;  
QQPLOT / normal ;  
QQPLOT / gamma (shape = %sysevalf(&n -1) scale =2 theta =0);  
RUN;
```

Dabei wird ein Histogramm erstellt (Abbildung 2), gleichzeitig eine Normal- bzw. (die zu erwartende) Gamma-Verteilung angenähert und deren Güte durch die durchgeführten *Goodness-of-Fit*-Statistiken statistisch ausgewertet.

In Abbildung 2 ist deutlich erkennbar, dass wohl eher die  $\Gamma$ -Verteilung (bessere Beschreibung d. Histogramms) als die Normalverteilung der Variable *chi\_stat* zugrunde liegt. Besonders deutlich wird dies, wenn die korrespondierenden QQ-Plots (s. Vortragsfolien T. Bluhmki in [12]) bzw. die Goodness-of-Fit Teststatistiken betrachtet werden. Letztere implizieren die Nicht-Ablehnung der Nullhypothese der zugrundeliegenden  $\chi^2$ -Verteilung mit  $N - 2$  Freiheitsgraden (s. Tabelle 7), d.h. es kann davon ausgegangen werden, dass die Variable *chi\_stat* einer  $\Gamma$ -Verteilung mit den Parametern  $p = 10.99 \approx (24-2)/2$  bzw.  $b = 1/1.99 \approx 1/2$  folgt (vgl. Tabelle 5).



**Abbildung 2:** Histogramm über die Variable *chi\_stat*

**Tabelle 5:** Parameterschätzer für angepasste Gammaverteilung für Variable *chistat*

Parameters for Gamma Distribution		
Parameter	Symbol	Estimate
Threshold	Theta	0
Scale	Sigma	1.993993
Shape	Alpha	11.09055
Mean		22.11449
Std Dev		6.640492

Die gleichen Argumentationen (vgl. Tabelle 6) führen nun auch für die Variable *MeanSquare* zur Verifizierung der hergeleiteten Gammaverteilung für vorher festgelegte Makro-Parameter, also

$$MeanSquare = Estimate\_REML \sim \Gamma\left(12 - 1, \frac{12-1}{0.00356}\right),$$

(vgl. Tabelle 6 mit Gleichung (3)).

**Tabelle 6:** Schätzer der angepassten  $\Gamma$ -Verteilung für *MeanSquare*

Parameters for Gamma Distribution		
Parameter	Symbol	Estimate
Threshold	Theta	0
Scale	Sigma	0.00356
Shape	Alpha	10.99036
Mean		0.039128
Std Dev		0.011803

Da das obige Makro zur Simulation und Auswertung der PK-Daten jedoch von vorher definierten Eingabeparametern wie  $n$ ,  $\sigma_B^2$ ,  $\sigma_W^2$  und der Behandlungsdifferenz abhängt, sind weitere Makros nötig, die die (Un-)Abhängigkeit der  $\Gamma$ -Verteilung in Gleichung (3) verifizieren. Die resultierenden Grafiken können in den Vortragsfolien [12] nachgeschlagen werden. Dabei wird deutlich, dass sich die geschätzten Parameter  $p$  und  $b$  genau so verhalten, wie es Gleichung (3) fordert.

**Tabelle 7:** Goodness-of-Fit-Statistiken bzgl. angepasster  $\Gamma$ - /Normalverteilung

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.0454231	Pr > D	<0.010
Cramer-von Mises	W-Sq	6.9851317	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	43.5341806	Pr > A-Sq	<0.005

Goodness-of-Fit Tests for Gamma Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.00587254	Pr > D	>0.500
Cramer-von Mises	W-Sq	0.04938108	Pr > W-Sq	>0.500
Anderson-Darling	A-Sq	0.30289676	Pr > A-Sq	>0.500

## 4 Anwendung

Wie am Ende von Abschnitt 3.1 schon angedeutet, ist es nun möglich den a-priori für die Fallzahlplanung angenommen Parameter der intra-individuellen Varianz (nach Abschluss der Studie) auf seine Güte hin zu überprüfen. Seien dafür  $\sigma_W^2$  die angenommene intra-individuelle Varianz, die für die Fallzahlplanung verwendet wurde,  $N = 2n$  die Studiengröße und  $\widehat{\sigma}_W^2$  die aus den Daten geschätzte Variabilität (hier: eine *Realisierung* der Zufallsvariable des Schätzers). Weiter seien folgende Testhypothesen gegeben:

$$H_0 : E(\widehat{\sigma}_W^2) = \sigma_W^2 \quad (\text{Nullhypothese})$$

$$H_1 : E(\widehat{\sigma}_W^2) \neq \sigma_W^2 \quad (\text{Alternativhypothese})$$

Nun setze  $T \stackrel{\text{def}}{=} \widehat{\sigma}_W^2$  als die zugrundeliegende Teststatistik. Aus den mathematischen Argumentationen ist nun bekannt, dass  $T \sim \Gamma\left(n - 1, \frac{n-1}{\sigma_W^2}\right)$  unter  $H_0$  gilt.

Abbildung 4.1 visualisiert beispielhaft solch eine Verteilung der Zufallsvariablen des Schätzers für die Parameter  $\sigma_W^2=0.03922$ , die (u.a.) daraus resultierende Fallzahl  $n=12$  und der Realisierung  $\widehat{\sigma}_W^2= 0.065$ .

Basierend auf dem Signifikanzniveau von 95% gilt nun offensichtlich

$P(Q1 \leq \widehat{\sigma}_W^2 \leq Q3) = 0.95$ , wobei  $Q1$  ( $Q3$ ) das 2.5%- (97.5%-) Quantil der zugrunde liegenden  $\Gamma$ -Verteilung definiert.

Das führt dazu, dass die Nullhypothese dieses (zweiseitigen) statistischen Tests abgelehnt wird, wenn entweder  $T < Q1$  oder  $T > Q3$  gilt (vgl. Abbildung 3).

Was bedeutet nun aber die Testentscheidung?

Intuitiv gibt dieser konstruierte Test eine Entscheidungsregel an, ob die Annahmen während der Planungsphase (insbesondere bei der Fallzahlplanung) zu den Ergebnissen „passen“. Kann die Nullhypothese nicht abgelehnt werden, so ist ein signifikanter Unterschied zwischen der angenommenen Variabilität und die der erhobenen Daten nicht nachweisbar, d.h. es kann davon ausgegangen werden, dass die Studie angemessen geplant (und durchgeführt) wurde. Wird  $H_0$  jedoch abgelehnt, passen Annahmen und Beobachtungen in nicht zueinander.

Dafür kann es diverse Gründe geben:

- die Variabilität für die Fallzahlplanung war nicht angemessen gewählt,
- die (u.a.) daraus berechnete Fallzahl war zu groß/zu klein,
- Ergebnis ist ein Zufallsereignis,
- die Studie wurde nicht angemessen durchgeführt,
- nicht bekannte/andere Gründe.

Die genaue Ursache für die Ablehnung der Nullhypothese kann dieser Test jedoch nicht weiter spezifizieren.

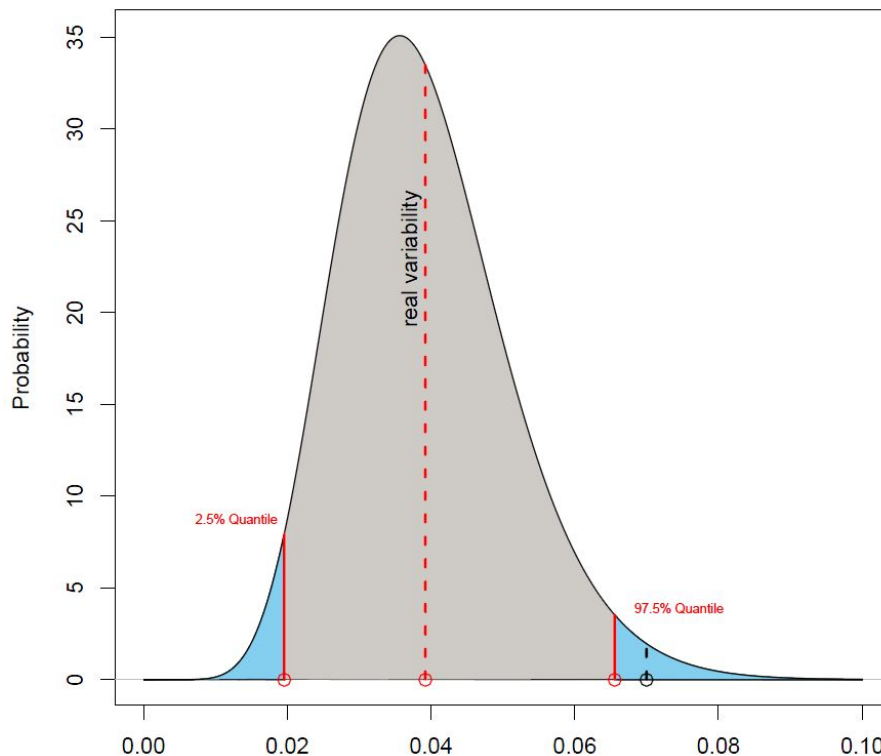


Abbildung 3: Beispiel einer Dichte für die Zufallsvariable  $\sigma_W^2$

## 5 Diskussion und Ausblick

Wie aus den Annahmen aus Kapitel 3 hervorgeht, basiert die Herleitung der stochastischen Verteilung auf einer balancierten Gruppengröße, d.h. es wurde  $N = 2n$  angenommen. Es ist natürlich möglich die mathematischen Argumentationen auf eine unbalancierte Gruppengröße zu erweitern. Dafür kann einfach  $N = n_1 + n_2$  gesetzt werden und die Ergebnisse aus Kapitel 3 bleiben gültig. Außerdem wurde die Theorie nur für einen festen Probandeneffekt hergeleitet, was unter der Annahme eines kompletten Datensatzes dazu führt, dass die Schätzer für die intra-individuelle Varianz für das Fixed- als auch für das Mixed Model identisch sind. Leider kommt es in der Realität oft vor, dass fehlende Werte in den Daten vorhanden sind, sodass ein direkter Transfer der hergeleiteten Ergebnisse nicht möglich ist. Folglich müssen weitere (mathematische) Untersuchungen für Datensätze mit fehlenden Werten durchgeführt werden.

Deshalb ist auch die Testentscheidung aus Kapitel 4 in der Realität mit Vorsicht zu behandeln und nur anwendbar unter der Annahme, dass die Verteilung des Schätzers auch

im Falle von fehlenden Werte (approximativ) der hergeleiteten  $\Gamma$ -Verteilung folgt. Weiter kann die Testentscheidung die Gründe für die Ablehnung der Nullhypothese nicht näher spezifizieren, sondern erfordern jeweils Einzelfalluntersuchungen.

Alle Ergebnisse, deren Herleitungen und Anwendungen basieren auf dem Standard  $2 \times 2 \times 2$  Cross-Over-Design. Es mag deshalb ausblickend als interessant erscheinen, wie die Ergebnisse auf höher dimensionale (allgemeinere) Cross-Over Designs übertragbar sind, insbesondere ob der Schätzer der intra-individuellen Varianz auch einer  $\Gamma$ -Verteilung (mit anderen Parametern) folgt. Es wäre auch möglich Eigenschaften über den Schätzer mithilfe von nicht-parametrischen Ansätzen oder bayesianisch herzuleiten oder ähnliche Untersuchungen für den Schätzer der inter-individuellen Varianz durchzuführen.

## Literatur

- [1] W. Cawello: Parameter zur modellunabhängigen Pharmakokinetik. 1st Edition. Shaker Verlag Aachen, 1998.
- [2] L.F. Lacey et al: Common non-compartmental Pharmacokinetic Variables – Are they Normal or Log-Normal Distributed? *Journal of Biopharmaceutical Statistics* 1997, 7(1): 171-178.
- [3] S.C. Chow, L.P. Liu: Design and Analysis of Bioavailability and Bioequivalence Studies. 3<sup>rd</sup> Edition. Chapman & Hall/CRC Press, 2009.
- [4] S. Patterson, B. Jones: Bioequivalence and Statistics in Clinical Pharmacology. Chapman & Hall/CRC Press, 2006.
- [5] S. Senn: Cross-Over Trials in Clinical Research. 2nd Edition. Wiley, 1993.
- [6] SAS Institute: SAS/Stat 9.2 User's Guide - The Mixed Procedure (Book Excerpt), 2008.
- [7] SAS Institute: SAS/Stat 9.2 User's Guide - The GLM Procedure (Book Excerpt), 2008.
- [8] S. Welham: REML Estimation and Linear Mixed Models. Slides, 2008. Verfügbar auf [http://www.ltcc.ac.uk/courses/REML/REML\\_and\\_LMM\\_ltcc1.pdf](http://www.ltcc.ac.uk/courses/REML/REML_and_LMM_ltcc1.pdf)
- [9] S.S. Wulff: The Equality of REML and ANOVA Estimators of Variance Components in unbalanced Normal Classification Models. *Statistics & Probability Letters*, Volume 78, Issue 4, pp. 405-411, 2008.
- [10] U.S. Department of Health and Human Services - Food and Drug Administration: Guidance for Industry: Statistical Approaches to Establishing Bioequivalence, 2001.
- [11] European Medicines Agency - Committee for Medical Products for Human Use: Guideline on the Investigation of Bioavailability and Bioequivalence, 2010.
- [12] KSFE 2013 Wiki: [http://de.saswiki.org/wiki/KSFE\\_2013](http://de.saswiki.org/wiki/KSFE_2013)