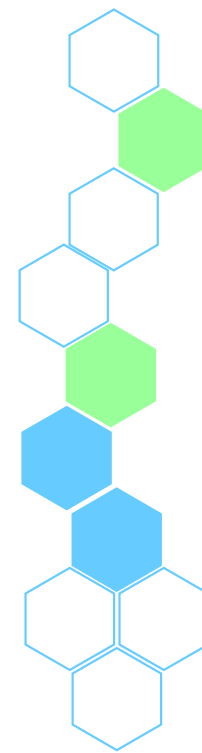
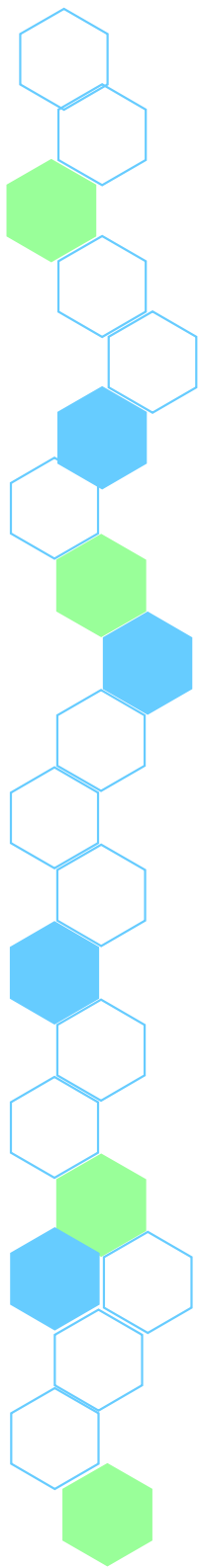




BE 試験の解析と例数設計

舟尾 暢男



- **BE 試験の概要と解析**
- BE 試験のデザイン
- BE 試験の例数設計

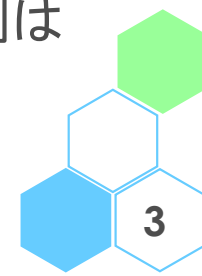
後発医薬品の生物学的同等性試験ガイドライン

薬生薬審発 0319 第 1 号(令和 2 年 3 月 19 日改正)



- 試験製剤 T と標準製剤 R について
- 原則としてクロスオーバー法(通常は 2 剤 2 期)、無作為割付(通常は非盲検)、被験者は原則として健康成人志願者、有意水準は片側 5%
- 例えば、経口製剤の単回投与試験では、 AUC_t 及び C_{max} を、多回投与試験では AUC_{τ} 及び C_{max} を生物学的同等性評価パラメータとする
- 原則として、 t_{max} を除くパラメータでは対数正規分布することが多いので、対数変換をして解析、90% 信頼区間(非対称, 最短区間)で生物学的同等性を評価する
 - これの代わりに、有意水準 5% の 2 つの片側検定(two one-sided tests)で評価してもよい、合理的な理由があれば他の適当なものを用いてもよい
- 試験製剤と標準製剤の生物学的同等性評価パラメータの対数値の平均値の差の 90% 信頼区間が、 $\log(0.80) \sim \log(1.25)$ の範囲にあるとき、試験製剤と標準製剤は生物学的に同等と判定する
 - 対数スケールのまま解析を行い、最後に元のスケールに戻してもよい

	Period 1	Period 2
Group 1	試験製剤 T	標準製剤 R
Group 2	標準製剤 R	試験製剤 T



BE 試験の解析

- 試験製剤 T と標準製剤 R との比較、以下では Cmax に関する解析を展開
- $X_{ijk} = \mu_t + \pi_k + s_{ij} + e_{ijk}$
 - X_{ijk} : 対数スケール、 μ_t : 治療効果 (T or R)、 π_k : 時期 (1 or 2)
 - s_{ij} : 群 i 、被験者 j の変量効果 $\sim N(0, \sigma_B^2)$ 、 e_{ijk} : 誤差 $\sim N(0, \sigma_W^2)$
 - n : 試験全体の例数 (\neq 各群の例数)、被験者内の変動係数 $CV = \sqrt{\exp(\sigma_W^2) - 1}$
- 「試験製剤と標準製剤は生物学的に同等」は、以下の計算により行う
 - $t_{1-\alpha, n-2}$: 自由度 $n - 2$ の t 分布の $(1 - \alpha)$ パーセンタイル点、 $\hat{t} = \bar{X}_T - \bar{X}_R$ 、 $\hat{\sigma}_W = \{\text{Group 1 の平方和} + \text{Group 2 の平方和}\} \div 2(n - 2)$ とし、信頼区間:
$$\left[\hat{t} - t_{1-\alpha, n-2} \sqrt{\frac{2\hat{\sigma}_W}{n}}, \hat{t} + t_{1-\alpha, n-2} \sqrt{\frac{2\hat{\sigma}_W}{n}} \right]$$
にて同等性の判断を行う ($\alpha = 0.05$)
- 実際の解析では、上記モデル式に関する変量効果モデルによる解析 (ざっくり言うと分散分析) にて信頼区間等を算出することがほとんど

BE 試験の解析: ダミーデータの作成

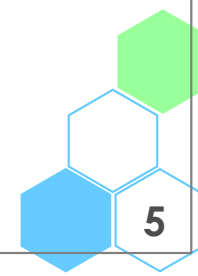


```
> makedata <- function (n=24, CV, RootMSE=0.3, theta0=0.95, set.seed=7777) {  
+   set.seed(set.seed)  
+   if (!missing(CV)) rmse <- sqrt(log(1 + CV^2))  
+   if (missing(CV) && !missing(RootMSE)) rmse <- RootMSE  
+   mlog <- log(theta0)  
+   result <- data.frame(subject=rep(1:n,2),  
+                         group =c(rep(1,n/2),rep(2,n/2),rep(1,n/2),rep(2,n/2)),  
+                         period =c(rep(1,n/2),rep(2,n/2),rep(2,n/2),rep(1,n/2)),  
+                         treat  =c(rep(1,n), rep(0,n)),  
+                         x      =c(rnorm(n=n, mean=mlog, sd=rmse),  
+                                   rnorm(n=n, mean=0, sd=rmse)) )  
+   result <- result[order(result$group, result$subject, result$period),]  
+   rownames(result) <- 1:nrow(result)  
+   return(result)  
+ }
```

```
> data <- makedata(CV=0.3)
```

```
> head(data)
```

	subject	group	period	treat	x
1	1	1	1	1	-0.60258648
2	1	1	2	0	-0.04710220
3	2	1	1	1	0.04454161
4	2	1	2	0	0.22852274
5	3	1	1	1	-0.13990655
6	3	1	2	0	0.20091221



BE 試験の解析: クロスオーバー分散分析



- 最後に結果を元のスケールに戻し、信頼区間が [0.8, 1.25] に含まれるか確認

```
> library(lme4)
> result <- lmer(x ~ group + period + treat + (1 | subject), data=data)
boundary (singular) fit: see ?isSingular
> s      <- summary(result)
> anova(result)          # ANOVA table
Analysis of Variance Table

      npar    Sum Sq   Mean Sq F value
group      1 0.0006368 0.0006368  0.0062
period      1 0.0024554 0.0024554  0.0239
treat       1 0.0121576 0.0121576  0.1184
> s$coefficients[4,2]      # se
[1] 0.09248986
> s$coefficients[4,3]      # t-value
[1] 0.3441432
> df <- s$ngrps - 2        # df
> r  <- s$coefficients[4,1] # ratio (log-scale)
> l  <- s$coefficients[4,1] - qt(1-0.05, df) * s$coefficients[4,2] # Lower CI
> u  <- s$coefficients[4,1] + qt(1-0.05, df) * s$coefficients[4,2] # Upper CI
> exp(c(r, l, u))          # ratio with 90% CI (original-scale)
[1] 1.0323417 0.8807436 1.2100337
```



- BE 試験の概要と解析
- **BE 試験のデザイン**
- BE 試験の例数設計

後発医薬品の生物学的同等性試験ガイドライン

薬生薬審発 0319 第 1 号(令和 2 年 3 月 19 日改正)



Q-10 本試験を行う前に、予試験を行う必要があるか

(A) 既知の情報から、必要例数及び体液採取間隔を含む適切な試験法を計画できない場合には、予試験を行う必要がある。(中略)

なお、予試験の結果を生物学的同等性の評価に用いることはできない。

Q-11 例数追加試験を実施し、本試験のデータと合わせて解析することはできるか

(A) 原則として、本試験とは別に例数追加試験を実施し、本試験のデータと合わせて解析することは認めない。(中略)

ただし、予試験の結果を含めた事前情報から予期した以上に個体内分散が大きくなり、生物学的同等性を示すことができないリスクを考慮し、中間解析結果に基づく追加データ取得を含むプロトコルを組むことは可能である。追加データの取得は、第一種の過誤確率を制御した上で、中間解析結果に基づき十分な検出力が担保できる例数で実施することを、1 回に限り認める。(以下略)



1. Fixed Design

- 今までの BE 試験(予試験又は本試験)のデザインと同様
- 各国のガイドライン等から 12 例以上での実施が望ましいが、日本のガイドラインの下記記載により 20 例以上とするのが得策
 - なお, 同等性の判定基準に適合しない場合でも, 試験製剤と標準製剤の生物学的同等性評価パラメータの対数値の平均値の差が $\log(0.90) \sim \log(1.11)$ であり, 且つ溶出試験で溶出挙動が類似していると判定された場合には, 生物学的に同等と判定する. ただし, この規定が適用されるのは, 総被験者数 20 名(1 群 10 名)以上が用いられた場合に限られる.

2. Group Sequential Design

- 中間解析(stage 1)にて BE の判定、BE 不成立の場合は最終解析にて BE の判定(stage 1 と stage 2 のデータを併合)
- 例えば stage 1 と stage 2 の例数を等しくし、Pocock の方法による α の調整を行う($\alpha_i = 0.030367$ 、 $c_i = 1.875423$)

3. Adaptive Design

- 例えば inverse normal method を使用
- Pocock の方法による α の調整 + stage 2 の例数を再算定



参考：中間解析 & 最終解析の棄却点

- 詳細は「R で統計解析入門 (16) 群逐次法」にて
http://nfunao.web.fc2.com/files/R-intro/R-stat-intro_16.pdf

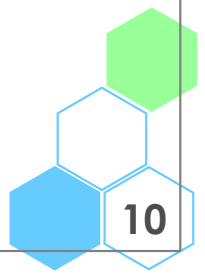
```
> # install.packages("rpact", dep=T)
> library(rpact)

> # Pocock Boundary
> x <- getDesignGroupSequential(kMax=2, alpha=0.05, sided=1, typeOfDesign="P")
> ( a <- x$stageLevels )
[1] 0.03036726 0.03036726
> qnorm(a, lower=F)
[1] 1.875423 1.875423

> pnorm(1.87542, lower=F)
[1] 0.03036748
> 1-pnorm(1.87542, lower=F)*2
[1] 0.939265

> # Alpha-spending function (Pocock type)
> x <- getDesignGroupSequential(informationRates=c(0.5, 1), alpha=0.05, sided=1,
+                               typeOfDesign="asP")
> ( a <- x$stageLevels )
[1] 0.03100573 0.02972334
> qnorm(a, lower=F)
[1] 1.866214 1.884875

> pocock <- function(t,alpha) alpha*log(1+(exp(1)-1)*t)
> pocock(0.5,0.05)
[1] 0.03100573
```



- BE 試験の概要と解析
- BE 試験のデザイン
- **BE 試験の例数設計**

1. Fixed Design

- 試験製剤 T と標準製剤 R との比較
- 以下、Cmax に関する解析を展開(AUC についても実施するが、例数設計においては Cmax についてのみ実施することで概ね良い)
- $X_{ijk} = \mu_t + \pi_k + s_{ij} + e_{ijk}$
 - X_{ijk} : 対数スケール、 μ_t : 治療効果(T or R)、 π_k : 時期(1 or 2)
 - s_{ij} : 群 i 、被験者 j の変量効果 $\sim N(0, \sigma_B^2)$ 、 e_{ijk} : 誤差 $\sim N(0, \sigma_W^2)$
 - 被験者内の変動係数 $CV = \sqrt{\exp(\sigma_W^2) - 1}$
- BE の判定は、以下の仮説に関する検定を行うことに相当する
 - $H_0 : H_0^U \cup H_0^L$ vs. $H_1 : H_1^U \cap H_1^L$
 - $H_0^U : \mu_T - \mu_R \geq \log 1.25$ vs. $H_1^U : \mu_T - \mu_R < \log 1.25$
 - $H_0^L : \mu_T - \mu_R \leq \log 0.80$ vs. $H_1^L : \mu_T - \mu_R > \log 0.80$

1. Fixed Design

- $H_0 : H_0^U \cup H_0^L$ vs. $H_1 : H_1^U \cap H_1^L$ に関する検定は以下となる
 - $T^U = \sqrt{\frac{n}{2}} \frac{\bar{X}_T - \bar{X}_R - \log 1.25}{\hat{\sigma}_W}$ 、 $T^L = \sqrt{\frac{n}{2}} \frac{\bar{X}_T - \bar{X}_R - \log 0.80}{\hat{\sigma}_W}$
 - \bar{X}_T 、 \bar{X}_R : 各製剤の対数変換後の平均、 $\hat{\sigma}_W^2 : \sigma_W^2$ の推定値、 n in total
 - $T^U \leq -t_{1-\alpha, n-2}$ & $T^L \geq t_{1-\alpha, n-2}$ のとき、 H_0 を棄却 (BE 成立)
 - $t_{1-\alpha, n-2}$: 自由度 $n - 2$ の t 分布の $(1 - \alpha)$ パーセンタイル点
- 有意水準 α 、検出力 $1 - \beta$ 、 $GMR = \exp(\mu_T) / \exp(\mu_R)$ 、 $F_t(x, n - 2)$ を自由度 $n - 2$ の t 分布の累積分布関数、例数は以下の基準にて算出
 - $F_t\left(\frac{\log\left(\frac{1.25}{GMR}\right)}{\sigma_W \sqrt{2/n}} - t_{1-\alpha, n-2}, n - 2\right) - F_t\left(\frac{-\log(1.25 \cdot GMR)}{\sigma_W \sqrt{2/n}} + t_{1-\alpha, n-2}, n - 2\right) \geq 1 - \beta$



1. Fixed Design: 信頼区間、 t 値、 p 値で BE の判断

- 信頼区間、 t 統計量、 p 値のいずれで判断しても結果は同じであるか確認
 - n : 試験全体の例数、CV or RootMSE を指定、 θ_0 : 製剤間の比

```
> TOST_sim <- function (alpha=0.05, n=22, CV=0.2, RootMSE, theta0=0.95, nsims=100000, set.seed=777) {
+   if (!missing(CV))               mse <- log(1 + CV^2)
+   if (missing(CV) && !missing(RootMSE)) mse <- RootMSE^2
+   ltheta1 <- log(0.80)
+   ltheta2 <- log(1.25)
+   mlog    <- log(theta0)
+   BE1     <- rep.int(TRUE, times = nsims)
+   BE2     <- rep.int(TRUE, times = nsims)
+   BE3     <- rep.int(TRUE, times = nsims)
+
+   set.seed(set.seed)
+   for (i in 1:nsims) {
+     df      <- n - 2
+     tval    <- qt(1 - alpha, df)
+     x       <- rnorm(n=n, mean=mlog, sd=sqrt(2*mse))
+     ratio   <- mean(x)
+     s2_h    <- (n-1)*var(x)/df/2 # not sigma_D^2!!
+     hw      <- tval * sqrt(2*s2_h/n)
+     lower   <- ratio - hw
+     upper   <- ratio + hw
+     BE1[i]  <- lower >= ltheta1 & upper <= ltheta2
+     Pu      <- pt( (ratio-ltheta2) * sqrt(n/(2*s2_h)), df)
+     Pl      <- pt( (ratio-ltheta1) * sqrt(n/(2*s2_h)), df, lower=F)
+     BE2[i]  <- Pu <= alpha & Pl <= alpha
+     P       <- max(Pu, Pl)
+     BE3[i]  <- qnorm(1-P) >= qnorm(1-alpha)
+   }
+   result <- c(mean(BE1)*100, mean(BE2)*100, mean(BE3)*100)
+   names(result) <- c("CI based", "t-value based", "p-value based")
+   return(result)
+ }
```

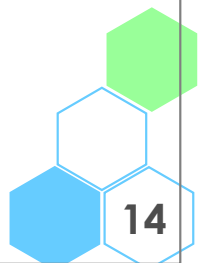
```
> TOST_sim(n=24, RootMSE=0.20, theta0=0.96) # n in total (not per seq.)
```

CI based t-value based p-value based

91.169

91.169

91.169



1. Fixed Design: 例数設計

- theta0: 製剤間の比、CV or RootMSE を指定、alpha: 片側0.05、beta: 1-検出力
 - 試験全体の例数 (n in total) が出力される

```
> TOST_N2 <- function(theta0=1, CV=0.2, RootMSE, alpha=0.05, beta=0.1, Nmax=1000)
{
+   if (!missing(CV))                sigma <- sqrt(log(1 + CV^2))
+   if (missing(CV) && !missing(RootMSE)) sigma <- RootMSE
+   N      <- 2
+   Power  <- 0
+   while ( Power < (1-beta) && N < Nmax) {
+     N      <- N+1
+     ta     <- qt(1-alpha, N-2)
+     Power  <- pt( log(1.25/theta0)/(sigma*sqrt(2/N)) - ta, N-2) -
+               pt(-log(1.25*theta0)/(sigma*sqrt(2/N)) + ta, N-2)
+   }
+   return( N+N%%2 )
+ }

> TOST_N2(RootMSE=0.20, theta0=0.96) # n in total (not per seq.)
[1] 24
```

2. Group Sequential Design: 例数設計

- 中間解析(stage 1)にて BE の判定、BE 不成立の場合は最終解析にて BE の判定(stage 1 と stage 2 のデータを併合)
- Stage 1 と stage 2 は等例数、Pocock の方法による調整($\alpha_i = 0.030367$)
 - alpha: 各 stage の有意水準を指定(片側)、n: 各 stage の 2 群合計の例数、RootMSE を指定、theta0: 製剤間の比、nsims: シミュレーション回数

```
> TOST_GS <- function (alpha=c(0.030367,0.030367), n=c(22,22), RootMSE=0.2, theta0=0.95,
                        nsims=100000, set.seed=777) {
+   ltheta1 <- log(0.80)
+   ltheta2 <- log(1.25)
+   mlog    <- log(theta0)
+   mse     <- RootMSE^2
+   BE      <- rep.int(TRUE, times = nsims)
+   stage2  <- rep.int(TRUE, times = nsims)
+
+   set.seed(set.seed)
+   for (i in 1:nsims) {
+     df      <- n[1] - 2
+     tval    <- qt(1 - alpha[1], df)
+     x       <- rnorm(n=n[1], mean=mlog, sd=sqrt(2*mse))
+     ratio   <- mean(x)
+     s2_h    <- (n[1]-1)*var(x)/df/2 # not sigma_D^2!!
+     hw      <- tval * sqrt(2*s2_h/n[1])
+     lower   <- ratio - hw
+     upper   <- ratio + hw
+     BE[i]   <- lower >= ltheta1 & upper <= ltheta2
+     stage2[i] <- !BE[i]
```


2. Group Sequential Design: 例数設計

- Power: 検出力、Stage 2: Stage 2 への移行割合、ASN: 平均症例数

```
+ if (!BE[i]) {  
+   df      <- n[1] + n[2] - 2  
+   tval    <- qt(1 - alpha[2], df)  
+   x       <- c(x, rnorm(n=n[2], mean=mlog, sd=sqrt(2*mse)))  
+   ratio   <- mean(x)  
+   s2_h    <- (n[1] + n[2] - 1)*var(x)/df/2 # not sigma_D^2!!  
+   hw      <- tval * sqrt(2*s2_h/(n[1]+n[2]))  
+   lower   <- ratio - hw  
+   upper   <- ratio + hw  
+   BE[i]   <- lower >= ltheta1 & upper <= ltheta2  
+ }  
+ }  
+ result <- c(mean(BE)*100, mean(stage2)*100, sum( n[1]+n[2]*mean(stage2) ) )  
+ names(result) <- c("Power", "Stage 2", "ASN")  
+ return(result)  
+ }  
  
> TOST_GS(n=c(12,12), RootMSE=0.20, theta0=0.96) # each ni in total (not per seq.)  
      Power  Stage 2      ASN  
87.08100 62.17700 19.46124
```

3. Adaptive Design (Inverse Normal Method)



- 各 stage の計画例数を n_i 、すなわち $n = n_1 + n_2$ in total
 - Stage 1 にて 12 ~ 13 頁の T^U 、 T^L から計算された p 値のうち、大きい方を p_1 とし、同様の計算で Stage 2 の p 値を p_2 とする
 - $\Phi^{-1}(1 - p_1), \Phi^{-1}(1 - p_2) \sim N(0,1)$ 、各 p 値は独立なので、中間解析と最終解析ではそれぞれ T_1^* 、 T_2^* にて検定を行えばよい (w_i : 各 stage の重み、例えば $w_i = \sqrt{n_i}$)
 - $T_1^* := \Phi^{-1}(1 - p_1), T_2^* := \frac{w_1 \Phi^{-1}(1 - p_1) + w_2 \Phi^{-1}(1 - p_2)}{\sqrt{w_1^2 + w_2^2}} \sim N(0,1)$
 - …と Kieser (2015) にはあるが、各 p_i は「2 つの p 値の最大値」であるため、帰無仮説の下でも一様分布に従わず、よって $\Phi^{-1}(1 - p_i)$ も標準正規分布に従わないため、上記は正しくないかもしれない。ただし、以下の関係式から T_1^* 、 T_2^* が標準正規分布に従うと仮定して検定を行った場合、保守的な検定となる。
 - ✓ $T_1^* := \Phi^{-1}(1 - p_1) \leq_{st} N(0,1)$
 - ✓ $T_2^* := \frac{w_1 \Phi^{-1}(1 - p_1) + w_2 \Phi^{-1}(1 - p_2)}{\sqrt{w_1^2 + w_2^2}} \leq_{st} N(0,1)$
- $A \leq_{st} B$: 確率変数 A は確率変数 B より、stochastic order の意味で小さい

3. Adaptive Design (Inverse Normal Method)



- 前頁の問題があるため、inverse normal method を適用する際、「各 stage の T_i^U から計算された p 値について上記の計算」「各 stage の T_i^L から計算された p 値について上記の計算」を行い、別々に BE の判定を行う方が良い (α は溢れない)
 - 各 stage の最大 p 値か、各 stage の T_i^L か、どちらが検出力が高いかは状況によるが、前頁の方法は保守的な検定であるため検出力の観点からは劣る
- いずれの方法でも、Stage 1 で BE 不成立の場合の例数再算定方法は何でも可 (α は溢れない、ただし検出力には影響する)
 - Kieser (2015) では Potvin (2008) の論文の方法を使用、ただこの論文には批判が...
 - Stage 1 の比と Root MSE より単純に例数設計 (他に stage 1 の Root MSE のみ使用)
 - 条件付き検出力に基づく方法

3. Adaptive Design (Inverse Normal Method): 例数設計

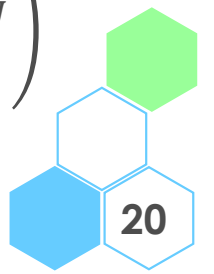


- Kieser (2015) での条件付き検出力の導出
- 13 頁と同様の議論により、stage i の 2 つの統計量を T_i^U と T_i^L 、各 CHW 統計量を T_{1+2}^U と T_{1+2}^L とすると、 $T_{1+2}^L = T_{1+2}^U + \tau_n$ ($\tau_n = \sqrt{n/2} \{\log 1.25 - \log 0.80\} / \sigma_W$) となり、stage 1 の統計量で条件付けた検出力は以下となる (stage 2 の棄却点: c_2)

$$\begin{aligned} & \checkmark P(T_{1+2}^U \leq -c_2 \wedge T_{1+2}^L \geq c_2 \mid T_1^U, T_1^L) = P(c_2 - \tau_n \leq T_{1+2}^U \leq -c_2 \mid T_1^U) \\ & = \Phi\left(\frac{-c_2\sqrt{n} - T_1^U\sqrt{n_1}}{\sqrt{n_2}} - \frac{\sqrt{n_2}}{\sqrt{n_1}}T_1^U\right) - \Phi\left(\frac{(c_2 - \tau_n)\sqrt{n} - T_1^U\sqrt{n_1}}{\sqrt{n_2}} - \frac{\sqrt{n_2}}{\sqrt{n_1}}T_1^U\right) \end{aligned}$$

- 例数再算定は τ_n を $\tau_n^* = (n_1 + \sqrt{n_2 n_2^*}) / \sqrt{2n} \{\log 1.25 - \log 0.80\} / \sigma_W$ に置換え、($T_{1+2}^L = T_{1+2}^U + \tau_n^*$)、以下の条件付検出力が $1 - \beta$ を超える n_2^* を探索

$$\begin{aligned} & \checkmark P(c_2 - \tau_n^* \leq T_{1+2}^U \leq -c_2 \mid T_1^U) \\ & = \Phi\left(\frac{-c_2\sqrt{n} - T_1^U\sqrt{n_1}}{\sqrt{n_2}} - \frac{\sqrt{n_2^*}}{\sqrt{n_1}}T_1^U\right) - \Phi\left(\frac{(c_2 - \tau_n^*)\sqrt{n} - T_1^U\sqrt{n_1}}{\sqrt{n_2}} - \frac{\sqrt{n_2^*}}{\sqrt{n_1}}T_1^U\right) \end{aligned}$$



3. Adaptive Design (Inverse Normal Method): 例数設計



- 前頁の計算は、統計量を分散既知の正規分布の式を t 統計量に置換えているため、厳密性に欠ける
- Inverse Normal により条件付検出力を再計算すると以下となり、2 行目の不等号の右辺 (α^* とおく) を stage 2 の有意水準として例数再算定すれば良い → この α^* で信頼区間を構成し [0.8, 1.25] に含まれるか判断、も可

$$\begin{aligned} \bullet \quad \Pr(T_2^* > c_2 \mid T_1^*) &= \Pr\left\{\frac{w_1\Phi^{-1}(1-p_1)+w_2\Phi^{-1}(1-p_2)}{\sqrt{w_1^2+w_2^2}} > c_2 \mid p_1\right\} \\ &= \Pr\left\{p_2 < 1 - \Phi\left(\sqrt{\frac{n_1+n_2}{n_2}}c_2 - \sqrt{\frac{n_1}{n_2}}\Phi^{-1}(1-p_1)\right) \mid p_1\right\} \end{aligned}$$

- よって、再計算後の条件付検出力は以下(ただし、t 分布での近似計算のため、 α^* が小さい場合や比が 1 から離れている場合は負になる場合あり)

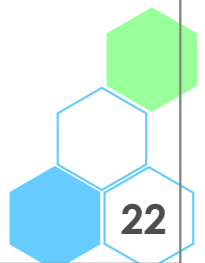
$$F_t\left(\frac{\log\left(\frac{1.25}{GMR}\right)}{\sigma_W\sqrt{2/n_2^*}} - t_{1-\alpha^*, n_2^*-2}, n_2^* - 2\right) - F_t\left(\frac{-\log(1.25 \cdot GMR)}{\sigma_W\sqrt{2/n_2^*}} + t_{1-\alpha^*, n_2^*-2}, n_2^* - 2\right)$$

3. Adaptive Design (Inverse Normal Method): 例数設計



- alpha: 各 stage の有意水準、n: 各 stage の例数、RootMSE を指定、theta0: 製剤間の比、min_n2: stage 2 の最小例数、max_n: 試験全体の最大例数、target_power: 目標検出力

```
> # ssr=1: based on Kieser (2015), 2: based on inverse normal, cpmin: not used
> TOST_IN <- function (alpha=c(0.030367,0.030367), n=c(22,22), min_n2=6, max_n=4*n[1], RootMSE=0.2, theta0=0.95, cpmin=0,
+                       target_power=0.8, ssr=1, nsims=200000, set.seed=777) {
+   ltheta1 <- log(0.80)
+   ltheta2 <- log(1.25)
+   mlog    <- log(theta0)
+   mse     <- RootMSE^2
+   BE      <- rep.int(TRUE, times=nsims)
+   BE2     <- rep.int(TRUE, times=nsims)
+   stage2  <- rep.int(TRUE, times=nsims)
+   n2      <- rep.int(0, times=nsims)
+
+   set.seed(set.seed)
+   for (i in 1:nsims) {
+     df      <- n[1] - 2
+     tval    <- qt(1 - alpha[1], df)
+     x       <- rnorm(n=n[1], mean=mlog, sd=sqrt(2*mse))
+     ratio   <- mean(x)
+     s2_h    <- (n[1]-1)*var(x)/df/2 # not sigma_D^2!!
+     hw      <- tval * sqrt(2*s2_h/n[1])
+     lower   <- ratio - hw
+     upper   <- ratio + hw
+     BE[i]   <- lower >= ltheta1 & upper <= ltheta2
+     stage2[i] <- !BE[i]
+
+     if (!BE[i]) {
+       # p-value for Stage 1
+       Plu    <- pt( (ratio-ltheta2) * sqrt(n[1]/(2*s2_h)), df)
+       Pll    <- pt( (ratio-ltheta1) * sqrt(n[1]/(2*s2_h)), df, lower=F)
+       P1     <- max(Plu, Pll)
+
+       # SSR
+       if (ssr==1){
+         zlu   <- (ratio-ltheta2) * sqrt(n[1]/(2*s2_h))
+         zll   <- (ratio-ltheta1) * sqrt(n[1]/(2*s2_h))
+         c2    <- qnorm(alpha[2], lower=F)
+         n2[i] <- min_n2
+         tau   <- (ltheta2-ltheta1) / sqrt(2*sum(n)) / sqrt(s2_h) * ( n[1] + sqrt(n[2]*n2[i]) )
```



3. Adaptive Design (Inverse Normal Method): 例数設計



```
+ cp      <- pnorm( (-c2      *sqrt(sum(n))-z1u*sqrt(n[1]))/sqrt(n[2]) - sqrt(n2[i]/n[1])*z1u ) -
+          pnorm( ((c2-tau)*sqrt(sum(n))-z1u*sqrt(n[1]))/sqrt(n[2]) - sqrt(n2[i]/n[1])*z1u )
+ while ( cp < target_power && sum(n[1]+n2[i]) < max_n) {
+   n2[i] <- n2[i] + 2
+   tau   <- (ltheta2-ltheta1) / sqrt(2*sum(n)) / sqrt(s2_h) * ( n[1] + sqrt(n[2]*n2[i]) )
+   cp    <- pnorm( (-c2      *sqrt(sum(n))-z1u*sqrt(n[1]))/sqrt(n[2]) - sqrt(n2[i]/n[1])*z1u ) -
+             pnorm( ((c2-tau)*sqrt(sum(n))-z1u*sqrt(n[1]))/sqrt(n[2]) - sqrt(n2[i]/n[1])*z1u )
+ }
+ }
+ if (ssr==2) {
+   alpha2 <- 1-pnorm( sqrt(sum(n)/n[2])*qnorm(alpha[2], lower=F)-sqrt(n[1]/n[2])*qnorm(1-P1) )
+   # cp    <- pt( log(1.25/exp(ratio))/(sqrt(s2_h)*sqrt(2/n[2])) - qt(1-alpha2, n[2]-2), n[2]-2) -
+   #       pt(-log(1.25*exp(ratio))/(sqrt(s2_h)*sqrt(2/n[2])) + qt(1-alpha2, n[2]-2), n[2]-2)
+   # cp    <- power.TOST(alpha=alpha2, theta0=exp(ratio), CV=sqrt(exp(s2_h)-1), n=n[2], method="exact")
+   n_tmp  <- TOST_N2(theta0=exp(ratio), RootMSE=sqrt(s2_h), alpha=alpha2, beta=(1-target_power), Nmax=(max_n - n[1]))
+   n2[i]  <- max(n_tmp, min_n2)
+ }
+
+ # Stage 2
+ df      <- n2[i] - 2
+ x       <- rnorm(n=n2[i], mean=mlog, sd=sqrt(2*mse))
+ ratio   <- mean(x)
+ s2_h    <- (n2[i] -1)*var(x)/df/2 # not sigma_D^2!!
+ P2u     <- pt( (ratio-ltheta2) * sqrt(n2[i]/(2*s2_h)), df)
+ P2l     <- pt( (ratio-ltheta1) * sqrt(n2[i]/(2*s2_h)), df, lower=F)
+ P2      <- max(P2u, P2l)
+ w1      <- sqrt(n[1])
+ w2      <- sqrt(n[2])
+ T2      <- (w1*qnorm(1-P1)+w2*qnorm(1-P2)) / (sqrt(w1^2+w2^2))
+ BE[i]   <- T2 >= qnorm(1-alpha[2])
+ T2u     <- (w1*qnorm(1-P1u)+w2*qnorm(1-P2u)) / (sqrt(w1^2+w2^2))
+ T2l     <- (w1*qnorm(1-P1l)+w2*qnorm(1-P2l)) / (sqrt(w1^2+w2^2))
+ BE2[i]  <- T2u >= qnorm(1-alpha[2]) && T2l >= qnorm(1-alpha[2])
+ }
+ }
+ result <- c(mean(BE)*100, mean(BE2)*100, mean(stage2)*100, sum( n[1]+mean(n2) ) )
+ names(result) <- c("Power(max p)", "Power(piu & pil)", "Stage 2", "ASN")
+ return(result)
+ }
```

3. Adaptive Design (Inverse Normal Method): 例数設計



- 中間解析 (stage 1) にて BE 不成立の場合、例数再算定を行うが、その方法は自由である
 - 今回は最終解析にて (条件付き、又は無条件の) 検出力が 90% となる例数を算出、ただし stage 2 の最小例数 (min_n2) は 12 例、試験全体 (stage 1 + stage 2) の最大例数 (max_n) を 48 例とした
 - 他の基準としては「条件付き検出力が、ある閾値未満 (例えば 30% 未満) の場合は計画例数のままとする」「例数再算定の結果、計画例数から減らす」等がある

```
> TOST_IN(n=c(12,12), min_n2=12, max_n=48, RootMSE=0.20, theta0=0.96, target_power=0.9)
```

Power(max p)	Power(piu & pil)	Stage 2	ASN
96.97200	97.63850	62.17050	29.04655

```
> TOST_IN(n=c(12,12), min_n2=12, max_n=48, RootMSE=0.20, theta0=0.96, target_power=0.9, ssr=2)
```

Power(max p)	Power(piu & pil)	Stage 2	ASN
97.9680	98.3615	62.2565	30.4594

参考: Stage 1 + 2 の併合データに関する推定



- Wassmer (2016) で紹介されている stagewise ordering による推定方法を、Power2Stage() の内部関数 adaptive_ci_limits() の実装を改変して実施する (stage 2 に移行した場合のみ計算する状況を考える)
- 各 stage の点推定値を $\hat{\theta}_k$ 、その標準誤差を se_k 、stage 1 の有意水準を α_1 、 $\alpha_{1,\delta} = 1 - \Phi\{\Phi^{-1}(1 - \alpha_1) - \delta/se_1\}$ 、 $p_{k,\delta} = 1 - \Phi\{(\hat{\theta}_k - \delta)/se_1\}$ とおく
- inverse normal 法に関する結合関数を $C(p_1, p_2) = 1 - \Phi\{w_1\Phi^{-1}(1 - p_1) + w_2\Phi^{-1}(1 - p_2)\}$ とすると、 $100 \times a\%$ 信頼下限は、以下の関数 $Q_\delta(p_{1,\delta}, p_{2,\delta})$ の値がちょうど a となる δ 、と計算される
 - $Q_\delta(p_{1,\delta}, p_{2,\delta}) = \alpha_{1,\delta} + \int_{\alpha_{1,\delta}}^1 \int_0^1 1\{c(x, y) \leq c(p_{1,\delta}, p_{2,\delta})\} dy dx$
 - $1\{c(x, y) \leq c(p_{1,\delta}, p_{2,\delta})\}$ は定義関数
 - $a = 0.05, 0.95$ の場合は両側 90% CI、 $a = 0.5$ の場合は 点推定値となる

参考: Stage 1 + 2 の併合データに関する推定



- Stagewise Ordering による点推定値 (Median Unbiased Point Estimate) と 90% CI
 - diff_i 、 semi_i 、 n_i にそれぞれ各 stage の点推定値、標準誤差、例数を指定
 - a_0 に stage 1 の有意水準、 a に信頼係数を指定 (0.5 or 0.05 or 0.95)

```
> so <- function(diff1, diff2, sem1, sem2, n1, n2, a1, a, lower_bnd=TRUE, maxiter=10000) {
+   # lower_bnd = TRUE creates lower bound l from one-sided interval (l, Inf)
+   # lower_bnd = FALSE creates upper bound u from one-sided interval (-Inf, u)
+
+   comb <- function(x, y, weight) 1-pnorm(weight[1]*qnorm(1-x) + weight[2]*qnorm(1-y))
+   pd <- function(d, diff, sem, df, lt=FALSE) {
+     pt((diff-d)/sem, df=df, lower.tail=lt)
+   }
+   indv <- function(x, d, weight, diff1, diff2, sem1, sem2, df1, df2, lt=FALSE) {
+     matrix(apply(x, 2, function(z) comb(z[1], z[2], weight)) <=
+       comb(pd(d, diff1, sem1, df1, lt), pd(d, diff2, sem2, df2, lt), weight)), ncol=ncol(x))
+   }
+   library(cubature)
+   Qd_a <- function(d, a1, a0, a, weight, diff1, diff2, sem1, sem2, df1, df2, lt=FALSE) {
+     # direction of H0 (>= 0 vs. <= 0)
+     sgn <- ifelse(lt, 1, -1)
+     ald <- ifelse(a1==0, 0, 1-pnorm(qnorm(1-a1) + sgn*d/sem1))
+     ald + cubature::hcubature(f = indv, lowerLimit=c(ald,0), upperLimit=c(1,1),
+       d=d, weight=weight, diff1=diff1, diff2=diff2, sem1=sem1, sem2=sem2,
+       df1=df1, df2=df2, lt=lt, tol=1e-10, maxEval=maxiter, vectorInterface=TRUE)$integral - a
+   }
+   df1 <- n1-2; df2 <- n2-2; weight=c(sqrt(n1),sqrt(n2))/sqrt(n1+n2)
+   diff_tmp <- (min(df1,1e+06)*diff1 + min(df2,1e+06)*diff2)/(min(df1,1e+06) + min(df2,1e+06))
+   search_int <- diff_tmp + c(-6,6) * max(sem1,sem2)
+   lt <- ifelse(lower_bnd, FALSE, TRUE)
+   uniroot(f=Qd_a, interval=search_int, a1=a1, a0=a0, a=a, weight=weight, diff1=diff1, diff2=diff2,
+     sem1=sem1, sem2=sem2, df1=df1, df2=df2, lt=lt, tol=1e-10, maxiter=maxiter)$root
+ }
```

参考: Stage 1 + 2 の併合データに関する推定

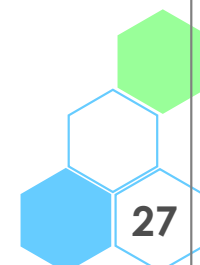


- Stagewise Ordering による点推定値 (Median Unbiased Point Estimate) と 90% CI
 - diff_i 、 semi_i 、 n_i にそれぞれ各 stage の点推定値、標準誤差、例数を指定
 - a_0 に stage 1 の有意水準、 a に信頼係数を指定 (0.5 or 0.05 or 0.95)

```
> # median Unbiased Point Estimate: exp(-0.0958092)=0.9086374
> so(diff1=log(0.85), diff2=log(0.95), sem1=(0.3/4), sem2=(0.2/4),
>     n1=16, n2=16, a1=0.030367, a=0.50)
[1] -0.0958092

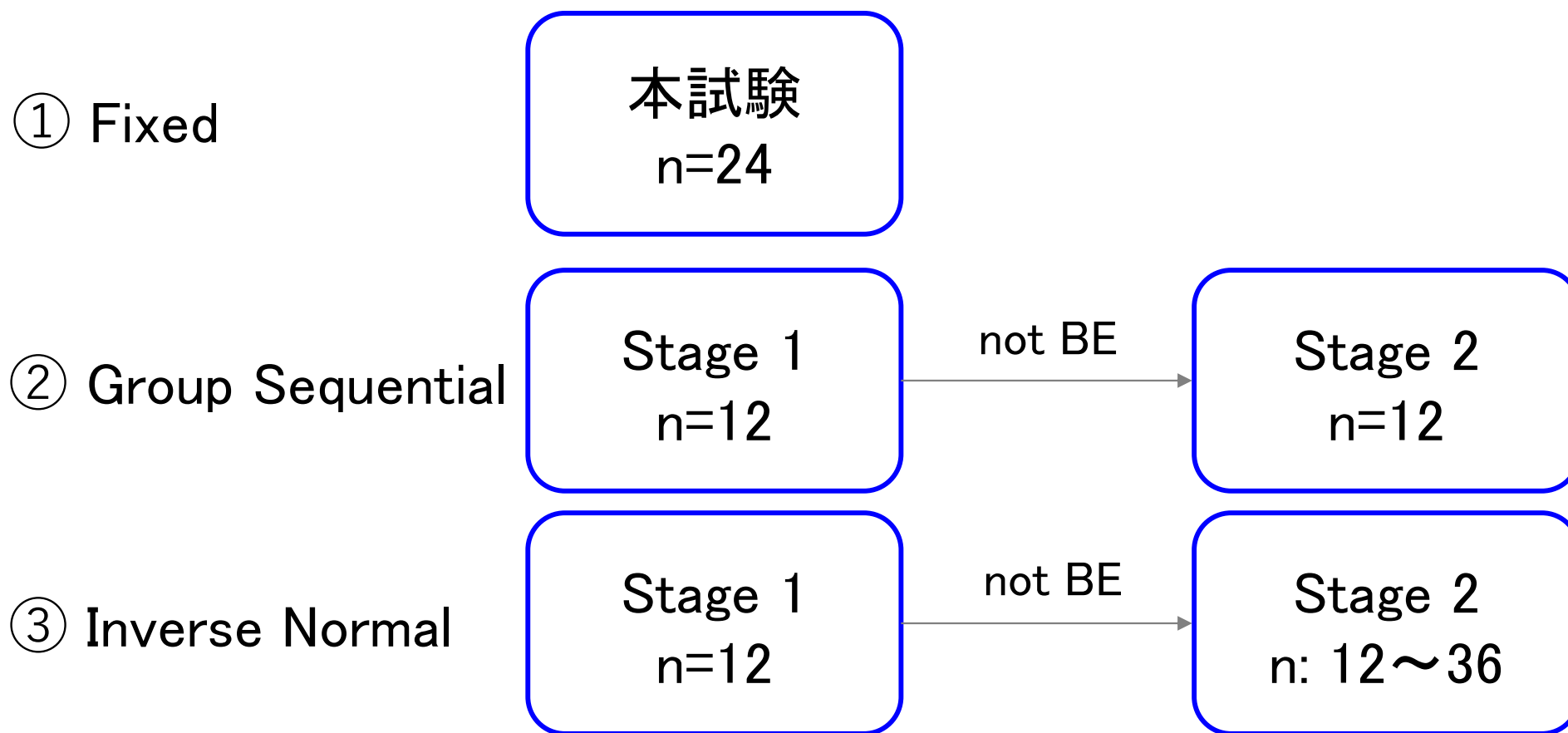
> # 90% Confidence Interval based on Stagewise Ordering:
> # [exp(-0.1718212), exp(-0.02278674)] = [0.8421297, 0.9774709]
> so(diff1=log(0.85), diff2=log(0.95), sem1=(0.3/4), sem2=(0.2/4),
>     n1=16, n2=16, a1=0.030367, a=0.05)
[1] -0.1718212

> so(diff1=log(0.85), diff2=log(0.95), sem1=(0.3/4), sem2=(0.2/4),
>     n1=16, n2=16, a1=0.030367, a=0.95)
[1] -0.02278674
```



例

- 過去の PK 試験より、Cmax の比が 0.96~1.04、Root MSE が 0.2 以下と想定される、AUC は比も Root MSE も Cmax に比べて楽観的
 - Cmax に基づいてのみ例数設計を行う
 - Fixed design で n=24 in total でも良さそうだが、中間解析にて例数を減らせる？



例



- 過去の PK 試験より、Cmax の比が 0.96~1.04、Root MSE が 0.2 以下と想定される、AUC は比も Root MSE も Cmax に比べて楽観的
 - Fixed design で 91%、被験者は必ず $n=24$ in total
 - Group Sequential なら平均症例数は $n \div 19.5$ だが検出力が約 4% 減る
 - Adaptive Design なら検出力は上がるが平均症例数も 5 ~ 6 例増える

```
> TOST_sim(n=24, RootMSE=0.20, theta0=0.96) # n in total (not per seq.)
      CI based t-value based p-value based
      91.169      91.169      91.169

> TOST_GS(n=c(12,12), RootMSE=0.20, theta0=0.96) # each ni in total (not per seq.)
      Power Stage 2      ASN
87.08100 62.17700 19.46124

> TOST_IN(n=c(12,12), min_n2=12, max_n=48, RootMSE=0.20, theta0=0.96, target_power=0.9)
      Power(max p) Power(piu & pil)      Stage 2      ASN
      96.97200      97.63850      62.17050      29.04655

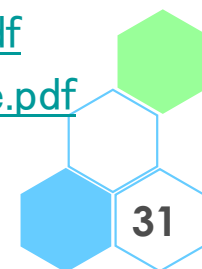
> TOST_IN(n=c(12,12), min_n2=12, max_n=48, RootMSE=0.20, theta0=0.96, target_power=0.9, ssr=2)
      Power(max p) Power(piu & pil)      Stage 2      ASN
      97.9680      98.3615      62.2565      30.4594
```

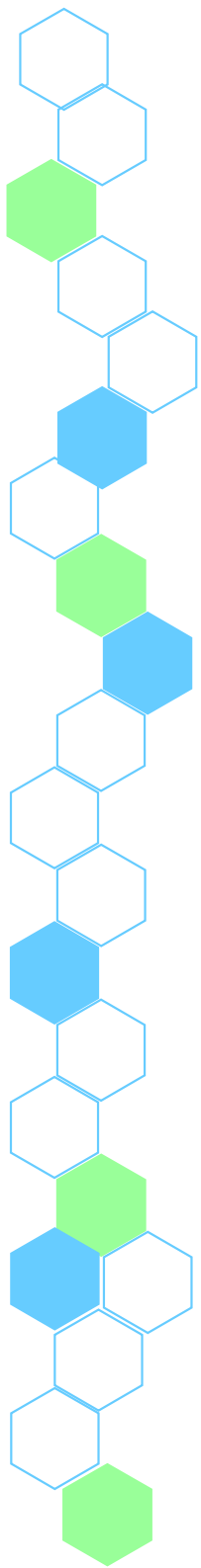
- BE 試験の概要と解析
- BE 試験のデザイン
- BE 試験の例数設計

参考文献



- 後発医薬品の生物学的同等性試験ガイドラインと Q&A(令和 2 年 3 月 19 日改正版)
<https://www.pmda.go.jp/review-services/drug-reviews/about-reviews/p-drugs/0008.html>
- 橋本 隆正 他(2021) “中間解析を伴う生物学的同等性試験のデザイン検討”, SAS ユーザー総会 2021 : https://www.sas.com/ja_jp/events/21/users-group-2021/sessions.html
- EMA (2010) “Guideline on the Investigation of Bioequivalence”
https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1_en.pdf
- Byron Jones and Michael G. Kenward (2014) “Design and Analysis of Cross-Over Trials (3rd Edition)”, Chapman & Hall/CRC
- Meinhard Kieser and Geraldine Rauch (2015) “Two-stage designs for cross-over bioequivalence trials”, Statistics in Medicine
- Mehta CR and Pocock SJ (2011) “Adaptive increase in sample size when interim results are promising: a practical guide with examples”, Statistics in Medicine
- Gernot Wassmer and Werner Brannath (2016) “Group Sequential and Confirmatory Adaptive Designs in Clinical Trials”, Springer
- Diane Potvin, et. al. (2007) “Sequential design approaches for bioequivalence studies with crossover designs”, Pharmaceutical Statistics
- R Packages:
 - PowerTOST: <https://cran.r-project.org/web/packages/PowerTOST/PowerTOST.pdf>
 - Power2Stage: <https://cran.r-project.org/web/packages/Power2Stage/Power2Stage.pdf>
 - rpact: <https://cran.r-project.org/web/packages/rpact/rpact.pdf>
 - cubature: <https://cran.r-project.org/web/packages/cubature/cubature.pdf>





– End of File –

