

# SASで始めよう被験者数再計算を伴う生物学的同等性試験

～ 後発医薬品のBEガイドライン改正に適応する ～

森田 祐介, 浜田 泉

(ノーベルファーマ株式会社 データサイエンス室)

Let's Try Two-stage Bioequivalence Trials with Sample Size Re-estimation using SAS

Yusuke Morita, Izumi Hamada

Data Science Office, Nobelpharma Co., Ltd.

## 要旨

令和2年3月19日付けで「後発医薬品の生物学的同等性試験ガイドライン」が一部改正され、より厳密な第一種の過誤確率の制御が求められるようになった。すなわち、従前許容されていた「本試験」に加えて「例数追加試験」を実施することが原則として認められなくなり、代わりに、「本試験」で、第一種の過誤確率を制御したうえで、中間解析結果に基づく被験者数の追加が1回に限り認められることとなった。本稿では、この改正に対応しうる中間解析を含む2段階デザインの解析手法のひとつとして、Maurerらが提案したMaximum Combination Testおよび被験者数再計算アルゴリズムを関連手法とともに紹介し、SASによる実装方法を提案する。なお、本ガイドラインの改正は、後発医薬品のみならず、先発医薬品のライフサイクルマネジメントにも影響しうるため、生物統計担当者へのインパクトは小さくないと思われる。

キーワード：生物学的同等性試験，2段階デザイン，アダプティブデザイン，逆正規法，被験者数再計算，Maximum Combination Test，Potvin，中間解析，後発医薬品の生物学的同等性試験ガイドライン

## 1. はじめに

生物学的同等性試験(BE試験)は、有効成分が同じ医薬品(例えば先発医薬品 vs 後発医薬品、あるいは錠剤 vs 顆粒剤)について、臨床上的有効性および安全性が同等であることを評価するために実施される試験である。通常、AUCやC<sub>max</sub>を同等性評価パラメータとして試験製剤と対照製剤のバイオアベイラビリティ(有効成分の未変化体または活性代謝物が体循環血中に入る速度および量)を、クロスオーバーデザインを用いて比較する。そして、試験製剤と対照製剤の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(または有意水準5%の2つの片側検定がいずれも棄却されたとき)、試験製剤と対照製剤は生物学的に同等(BE)と判定する。また、総被験者数20名以上の場合、試験製剤と対照製剤の同等性評価パラメータの対数値の平均値の差が $\log(0.90) \sim \log(1.11)$ の範囲にあり、かつ、「後発医薬品の生物学的同等性試験ガイドライン」(BEガイドライン)で規定された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、BEと判定する[1]。

令和2年3月19日付けでBEガイドラインが一部改正され、「本試験」は後発医薬品の開発において検証試験であることが明確にされ、より厳密な第一種の過誤確率の制御が求められるようになった。(BE試験での第一種の過誤とは、真実はBEでない製剤同士を、BEであると判定することを指す。)すなわち、「本試験」とは別に「例数追加試験」を実施し、「本試験」のデータと併合して解析することは原則として認められなくなった。加えて、「予試験」の結果をBE評価に用いることも認められなくなった。代わりに、「本試験」で、第一種の過誤確率を制御したうえで、中間解析結果に基づく被験者数の追加が1回に限り認められることとなった[2,3]。

このBEガイドライン改正の影響は、後発医薬品だけにとどまらない。「含量が異なる経口固形製剤の生物学的同等性試験ガイドライン」「経口固形製剤の処方変更の生物学的同等性試験ガイドライン」「剤形が異なる製剤の追加のための生物学的同等性試験ガイドライン」は[4,5,6]、既承認の薬剤について、それぞれ所定の条件に該当する場合は、BEガイドラインに準じたBE試験の実施を求めている。したがって、後発医薬品のBEガイドラインの改正は、先発医薬品のライフサイクルマネジメントにも影響を与える改正であり、生物統計担当者へのインパクトは小さくないと思われる。

本稿では、BEガイドラインの改正に対応しうる解析手法を、先行文献に基づいて紹介し、SASによる実装方法を提案する。2章では、本稿で想定する試験デザインを示し、3章では、BEガイドラインの改正に対応しうる解析手法および被験者数再計算アルゴリズムを紹介する。そして4章でSASによる実装例を示し、5章では、試験計画時や報告時に考慮すべき事項を考察する。

## 2. 想定する試験デザイン

本稿では、AUCおよびC<sub>max</sub>を同等性評価パラメータとして、試験製剤Tと対照製剤RのBE評価を目的とする、中間解析を含む、2段階デザインの2剤2期のクロスオーバー試験を想定する。中間解析までをステージ1、中間解析後から最終解析までをステージ2と呼ぶこととする。(図1)

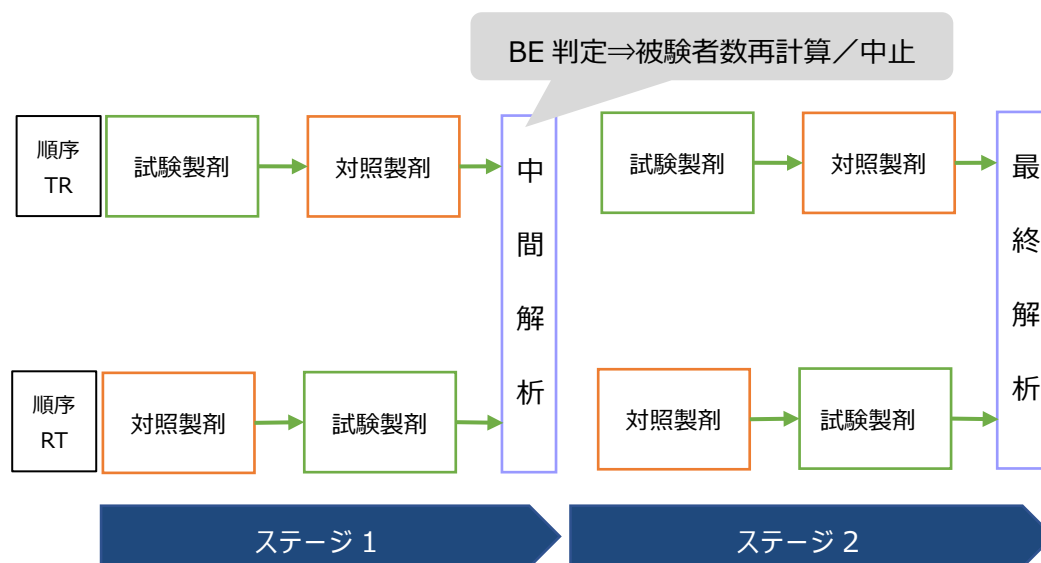


図1. 中間解析を含む、2段階デザインの2剤2期のクロスオーバー試験

本試験デザインは、中間解析結果に基づいて BE 評価が可能であり、BE でないと判定される場合には、適切な被験者数を追加可能であること、中間解析で BE と判定された場合または被験者数を追加して試験を継続しても BE と判定される見込みが低い場合は、ステージ 1 で早期中止することが可能である点で魅力的である。

一方、被験者数再計算を伴う 2 段階デザインでは、以下の 2 つの要因から第一種の過誤確率が上昇する。

(1) 中間解析と最終解析と繰り返し BE 評価を行うことによるもの

(2) 中間解析に基づく被験者数追加によるもの

したがって、3 章で紹介する第一種の過誤確率を制御可能な解析手法の適用が必要となってくる。

なお、予め最終解析の目標被験者数を固定したうえで、中間解析で早期中止を判断する群逐次デザインは、被験者数再計算を伴うものではないため、本稿では取り上げない。しかし、2 段階デザインの BE 試験を検討する場合、群逐次デザインも第一種の過誤確率を制御可能であり、選択肢になりうる。

### 3. 改正後の BE ガイドラインに対応しうる解析手法

BE ガイドラインでは、中間解析結果に基づく被験者数追加に対応するための具体的な解析手法への言及はないが、BE ガイドラインの改正について医薬品医療機器総合機構(PMDA)の職員らが概説した文献[7]では、(PMDA の公式見解を示すものではないが)、いくつかの解析手法について「maybe helpful」として関連する文献が挙げられている[8,9,10]。本章では、それらの手法について概説する。なお、3.3 節 Standard Combination Test, 3.4 節 Maximum Combination Test および、3.5 節 被験者数再計算のアルゴリズムについては、棚橋らの総説[11]に数値例と共に詳説されており、理論的な詳細はそちらを参照されたい。

#### 3.1 数理と表記

解析手法の説明に先立って本稿での帰無仮説、対立仮説および検定統計量を示す。同等性評価パラメータは、AUC または Cmax とする。対数変換した同等性評価パラメータの試験製剤の平均を  $\mu_T$ 、対照製剤の平均を  $\mu_R$  とし、 $\delta = \mu_T - \mu_R$  とする。BE 試験における帰無仮説と対立仮説は以下となる。

$$(\text{複合})\text{帰無仮説} : H_{01} : \delta \leq \log(0.80) \cup H_{02} : \delta \geq \log(1.25)$$

$$(\text{複合})\text{対立仮説} : H_{11} : \delta > \log(0.80) \cap H_{12} : \delta < \log(1.25)$$

また、説明の便宜上、ポイント帰無仮説： $H_{01}^* : \delta = \log(0.80)$ 、 $H_{02}^* : \delta = \log(1.25)$  を定義する。

(複合)帰無仮説は、 $H_{01}$  vs  $H_{11}$  および  $H_{02}$  vs  $H_{12}$  に対応する 2 つの片側検定(TOST : Two One-Sided Tests)によって検定される。 $H_{01}$  と  $H_{02}$  に対するステージ  $i$  の検定統計量は以下である。

$$T_{i1} = \frac{\log(1.25) + \hat{\delta}_i}{\sqrt{2\hat{\sigma}_i^2/n_i}}, \quad T_{i2} = \frac{\log(1.25) - \hat{\delta}_i}{\sqrt{2\hat{\sigma}_i^2/n_i}}$$

$\hat{\delta}_i$  はステージ  $i$  における  $\delta$  の推定値、 $\hat{\sigma}_i^2$  は対数変換した同等性評価パラメータの被験者内分散  $\sigma^2$  の推定値、 $n_i$  は被験者数を表す。 $T_{i1}$  と  $T_{i2}$  はそれぞれ、自由度  $n_i - 2$  の t 分布に従い、対応する p 値を  $p_{i1}$  と  $p_{i2}$  とする。

### 3.2 Potvin の方法[8,9]

Potvin の方法は、有意水準  $\alpha$  を調整した中間解析を実施し、BE であると判定される場合は試験を終了し、そうでない場合は、中間解析の結果に基づいて被験者数を追加し、試験を継続したうえで、最終解析を行うものである。BE 判定手順や中間解析と最終解析の有意水準  $\alpha$  の割り振りに基づき、Method B、Method C、Method D が提案されている(図 2、図 3)。また、Method B と Method C にステージ 1 の 90%信頼区間が一定の範囲内に含まれない場合は無効中止とするルールを取り入れ、被験者内変動係数が 10~30%、30~55% のそれぞれの想定において、ステージ 1 の被験者数や各ステージの有意水準など試験デザインに関するパラメータを最適化した Method E、Method F も提案されている[12]。

Potvin の方法は簡便であり、欧州医薬品庁(EMA)の BE ガイドライン[13]では Method B に関する言及があり、Lee らは、2006 年から 2019 年のアメリカ食品医薬品局(FDA)の後発医薬品の承認審査で 9 つの 2 段階デザインの事例があり、そのうち Method B が 1 つ、Method C が 7 つであったことを報告している[14]。

Potvin の方法はシミュレーションによって第一種の過誤確率が片側 5%程度以下に抑えられることを確認したものであり、第一種の過誤確率の制御を理論的に保証した方法ではない点に注意が必要である。Potvin らは、シミュレーション結果の評価では、5.2%までの第一種の過誤確率を許容している[8]。実際に Potvin ら、Montague ら、および Maurer らのシミュレーション[8,9,10]でも、Potvin の方法では第一種の過誤確率が 5%を超える結果が散見される。

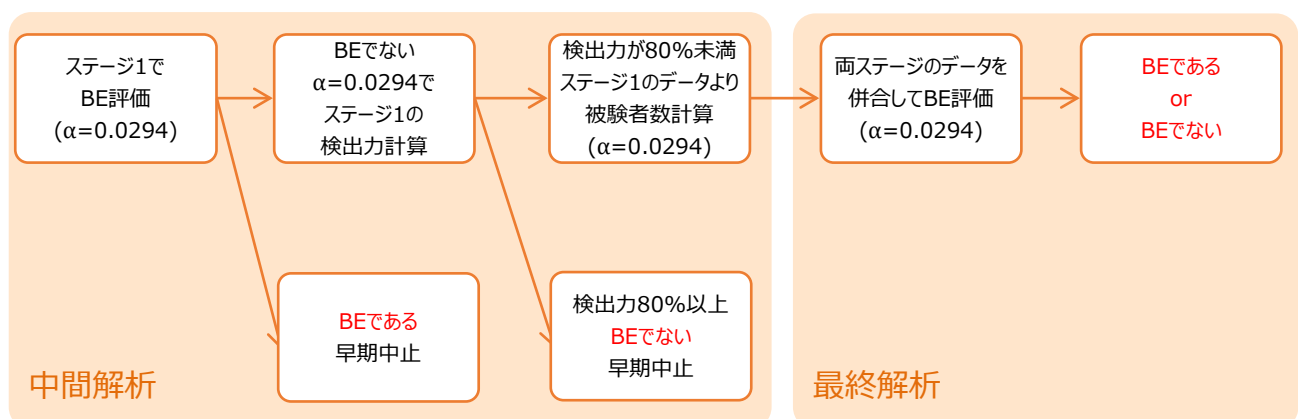


図 2. Potvin Method B のフロー[8]

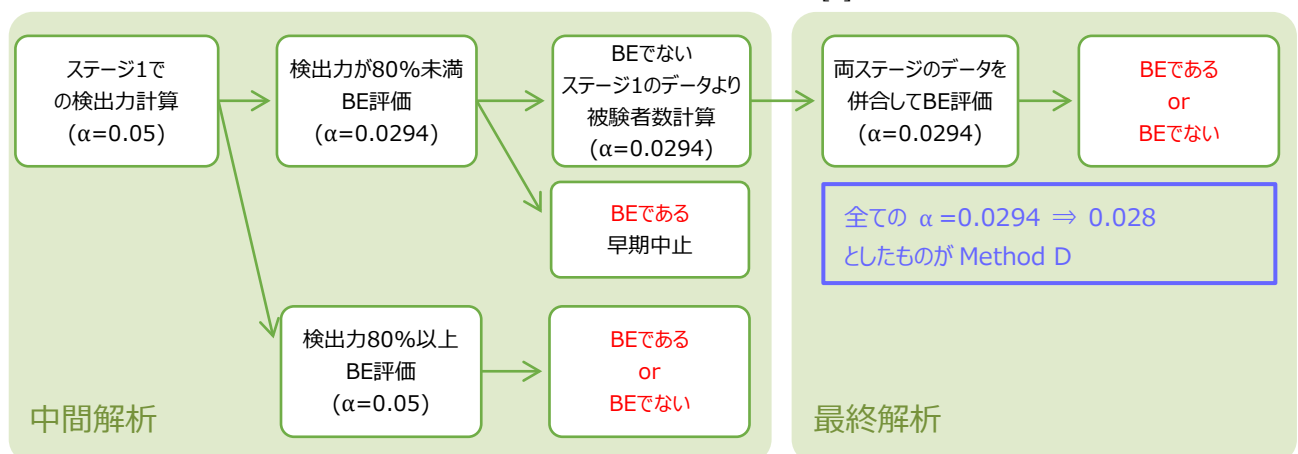


図 3. Potvin Method C および D のフロー[8]

### 3.3 Standard Combination Test [15]

Kieser らは、群逐次法および逆正規法を適用することで 2 段階デザインの BE 試験で、第一種の過誤確率を制御できることを示した。  $Z_{ij} = \Phi^{-1}(1 - p_{ij})$  として ( $\Phi^{-1}(\cdot)$  は標準正規分布の累積分布関数  $\Phi(\cdot)$  の逆関数)、ステージ 1 とステージ 2 の統計量を、重み  $w$  ( $0 < w < 1$ ) を用いて逆正規法により統合したものをステージ 2 終了後の最終解析の検定統計量  $Z_{0j}$  とする。(ここで  $Z_{1j}$  と  $Z_{2j}$  は独立である)

$$Z_{0j} = \sqrt{w}Z_{1j} + \sqrt{1-w}Z_{2j}$$

最終解析の検定統計量  $Z_{0j}$  および中間解析の検定統計量  $Z_{1j}$  は、ポイント帰無仮説  $H_{0j}^*$  のもとで、正の相関  $\sqrt{w}$  をもつ、二変量標準正規分布に従う。

$$\begin{pmatrix} Z_{1j} \\ Z_{0j} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{w} \\ \sqrt{w} & 1 \end{pmatrix} \right)$$

ステージ  $i$  終了時の検定に対する棄却限界値を  $z_{\alpha_i}$  として、試験全体での第一種の過誤確率が  $\alpha$  となるように  $z_{\alpha_i}$  を定める。

$$P(Z_{1j} < z_{\alpha_1} \cap Z_{0j} < z_{\alpha_2}) = 1 - \alpha$$

ステージ 1 終了時の BE 評価(TOST)では、 $H_{01}$  および  $H_{02}$  が棄却されるとき、つまり、 $p_{11} < \alpha_1$  かつ  $p_{12} < \alpha_1$  のとき BE と判定する。BE と判定されない場合、ステージ 1 の結果からステージ 2 の被験者数を計算のうえ、試験を継続する。最終解析では、逆正規法により統合した統計量  $Z_{0j}$  に基づいて、BE 評価を行う。つまり、 $Z_{0j}$  の観測値  $z_{0j}$  が  $z_{01} \geq z_{\alpha_2}$  かつ  $z_{02} \geq z_{\alpha_2}$  のとき BE と判定する。(Maurer らは、この手法を Standard Combination Test と呼んでいる。)

### 3.4 Maximum Combination Test [10]

Maurer らは Standard Combination Test の改良法として Maximum Combination Test を提案した。Standard Combination Test では、重み  $w = n_1/(n_1 + n_2)$  であるとき、検出力が最大となる。しかし、重み  $w$  は試験開始前に設定する必要があるが、ステージ 2 の被験者数  $n_2$  はステージ 1 の中間解析結果に基づいて決定されるため、重み  $w$  の最適な設定値を試験開始前に知ることはできない。そこで、Maurer らは、事前に 2 つの異なる重み  $w$  と  $w^*$  ( $0 < w^* < w < 1$ ) を設定し、それぞれの重みに対する統計量  $Z_{0j}$  と  $Z_{0j}^*$  を考え、最終解析では、 $Z_{\max j} = \max(Z_{0j}, Z_{0j}^*)$  に基づいて、BE 評価を行う方法を提案し、これを Maximum Combination Test と名付けた。

$$\begin{aligned} Z_{0j} &= \sqrt{w}Z_{1j} + \sqrt{1-w}Z_{2j} \\ Z_{0j}^* &= \sqrt{w^*}Z_{1j} + \sqrt{1-w^*}Z_{2j} \end{aligned}$$

ここで、最終解析の検定統計量  $Z_{0j}$ 、 $Z_{0j}^*$  および中間解析の検定統計量  $Z_{1j}$  は、ポイント帰無仮説  $H_{0j}^*$  のもとで、以下の相関構造をもつ三変量標準正規分布に従う。

$$\begin{pmatrix} Z_{1j} \\ Z_{0j} \\ Z_{0j}^* \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{w} \\ \sqrt{w} & 1 \\ \sqrt{w^*} & (\sqrt{ww^*} + \sqrt{(1-w)(1-w^*)}) \end{pmatrix} \begin{pmatrix} \sqrt{ww^*} + \sqrt{(1-w)(1-w^*)} \\ 1 \\ 1 \end{pmatrix} \right)$$

Standard Combination Test と同様にステージ 1 とステージ 2 終了時の検定に対する棄却限界値をそれぞれ  $z_{\alpha_{1,max}}$  と  $z_{max}$  として、試験全体での第一種の過誤確率が  $\alpha$  となるように  $z_{\alpha_{1,max}}$  と  $z_{max}$  を定める。

$$P(Z_{1j} < z_{\alpha_{1,max}} \cap Z_{0j} < z_{max} \cap Z_{0j}^* < z_{max}) = 1 - \alpha$$

ステージ 1 終了時の BE 評価(TOST)では、 $H_{01}$  および  $H_{02}$  が棄却される時、つまり、 $p_{11} < \alpha_{1,max}$  かつ  $p_{12} < \alpha_{1,max}$  のとき BE と判定する。BE と判定されない場合、ステージ 1 の結果からステージ 2 の被験者数を計算のうえ、試験を継続する。最終解析では、統計量  $Z_{max_j}$  に基づいて、BE 評価を行う。つまり、 $Z_{max_j}$  の観測値  $z_{max_j}$  が  $z_{max_1} \geq z_{max}$  かつ  $z_{max_2} \geq z_{max}$  のとき BE と判定する。

### 3.5 被験者数再計算のアルゴリズム[10]

Potvin らや Kieser らは、中間解析の被験者数再計算に、ステージ 1 から得られた被験者内変動の推定値  $\hat{\sigma}_1^2$  だけを用いており、ステージ 1 の製剤間の差の推定値  $\hat{\delta}_1 (= \hat{\mu}_{T1} - \hat{\mu}_{R1})$  は用いていない。しかし、Maurer らは、ステージ 1 の p 値に関する情報(つまり  $\hat{\delta}_1$ )を無視することは、ステージ 2 の被験者数を必要以上に増大させるとして、新しい被験者数再計算のアルゴリズムを提案した。

このアルゴリズムでは、ポイント帰無仮説  $H_{0j}^*$  のもとで、ステージ 1 のデータを所与としたときに、ステージ 2 で帰無仮説  $H_{0j}$  を棄却する確率 conditional error rate を求める。そして、conditional error rate を有意水準として、ステージ 2 の TOST の検出力が目標検出力を越えるときの被験者数を求める。このとき、目標検出力には、ステージ 1 で BE と判定されなかったもとで、ステージ 2 で BE と判定する条件付き確率 estimated conditional target power を用いる。

ステージ 1 の統計量  $Z_{1j}$  の観測値  $z_{1j}$  とすると、Standard Combination Test の conditional error rate は以下となる。

$$\alpha_{comb_j}^c = 1 - \Phi((z_{\alpha_2} - \sqrt{w} z_{1j})/\sqrt{1-w})$$

同様に、Maximum Combination Test の conditional error rate は以下となる。

$$\alpha_{max_j}^c = 1 - \Phi(\min_j)$$

$$\min_j = \min[(z_{max} - \sqrt{w} z_{1j})/\sqrt{1-w}, (z_{max} - \sqrt{w^*} z_{1j})/\sqrt{1-w^*}]$$

一方、目標検出力である estimated conditional target power は、 $1 - \beta^c = (\hat{\beta}_1 - \beta)/\hat{\beta}_1$  となる。その導出は、 $R_i$  をステージ  $i$  で BE と判定するイベントとするといずれかのステージで BE と判定するイベント  $R = R_1 \cup R_2$  となる。 $R_i$  の補集合を  $\bar{R}_i$  とすると試験全体の検出力は以下となる。

$$1 - \beta = P(R) = P(R_1 \cup R_2) = P(R_1) + P(\bar{R}_1 \cap R_2) = P(R_1) + P(R_2|\bar{R}_1)P(\bar{R}_1) \text{ より}$$

$$P(R_2|\bar{R}_1) = (P(R) - P(R_1))/P(\bar{R}_1)$$

ここで、 $\beta_1 = P(\bar{R}_1)$ ,  $1 - \beta = P(R)$ ,  $1 - \beta^c = P(R_2|\bar{R}_1)$ より、 $1 - \beta^c = (\beta_1 - \beta)/\beta_1$ が導かれる．実際には  $\beta_1$  はステージ 1 のデータから推定した  $\hat{\beta}_1$  を代入する．

最後に、ステージ 2 の検出力が、estimated conditional target power を達成する被験者数を求める．しかし、conditional error rate が 2 つの片側検定で異なる場合、検出力曲線を  $\delta$  に対して描いた場合、 $\delta = 0$  を最大値とした左右対称の形状とならなくなる．このため、 $\delta_p \neq 0$  であるとき、 $\delta' \in [-|\delta_p|, +|\delta_p|]$  である全ての  $\delta'$  に対して、検出力が estimated conditional target power 以上となることを保証する必要があるが生じる．そこで、Maurer らは、上側と下側の conditional error rate に基づいて検出力計算に用いる  $\delta_{ap}$  を設定することで、全ての  $\delta'$  に対して、検出力が estimated conditional target power 以上となることを保証する方法(Adaptive Planning)を提案した．( $\delta_p = \log(\text{計画時の幾何平均比})$ とする)

$$\delta_{ap} = \begin{cases} +|\delta_p| & \text{if } \alpha_{comb_1}^c > \alpha_{comb_2}^c \text{ or } \alpha_{max_1}^c > \alpha_{max_2}^c \\ -|\delta_p| & \text{if } \alpha_{comb_1}^c \leq \alpha_{comb_2}^c \text{ or } \alpha_{max_1}^c \leq \alpha_{max_2}^c \end{cases}$$

つまり、 $\delta = \delta_{ap}$  のもとで、ステージ 2 の検出力  $P(T_{21} \geq t_{n_2-2, \alpha_{max_1}^c} \cap T_{22} \geq t_{n_2-2, \alpha_{max_2}^c}) \geq 1 - \beta^c$  を

満たす最小の  $n_2$  を求める(自由度  $v$  である  $t$  分布の上側累積確率  $\alpha$  点を  $t_{v, \alpha}$  とする)．なお、 $T_{21}$  と  $T_{22}$  は、相関係数が  $-1$ 、自由度  $n_2 - 2$ 、非心度がそれぞれ  $(\log(1.25) + \delta_{ap})/\sqrt{2\hat{\sigma}_1^2/n_2}$ 、 $(\log(1.25) - \delta_{ap})/\sqrt{2\hat{\sigma}_1^2/n_2}$  である 2 変量非心  $t$  分布に従うものとして計算する．

### 3.6 各解析手法の性能比較[10]

Maurer らは、Potvin の Method B、Standard Combination Test および Maximum Combination Test についてシミュレーションによる性能比較を行っており、その結果から、Maximum Combination Test の利用を推奨している．なお、Standard Combination Test と Maximum Combination Test には、3.5 節の被験者数再計算のアルゴリズムおよび 4.2 節で示す判定フロー(図 4)をそれぞれ適用してシミュレーションを実施している．

第一種の過誤確率は、Potvin の Method B でいくつかの条件で最大 5.68% まで上昇する結果が観察されているが、Standard Combination Test と Maximum Combination Test では、すべての条件で第一種の過誤確率が 5% 以下に制御されている．また、Maximum Combination Test の方が Standard Combination Test より第一種の過誤確率が小さい傾向が観察されている．

目標として 80% を設定した検出力は、Potvin の Method B は、ステージ 1 の被験者数が少ない場合や被験者内変動係数が 50% を越える条件において、80% を下回る結果が観察されているが、Standard Combination Test および Maximum Combination Test は、ステージ 1 の被験者数が 12 名を超える場合は、概ね 80% 以上の検出力が観察されている．

期待被験者数では、想定した幾何平均比が 0.95 に対して真の幾何平均比が 0.9 の場合、Standard Combination Test よりも Maximum Combination Test の方が少なくなる傾向が観察されている．また、ステージ 2 の被験者数に上限を設けた場合、期待被験者数は同程度ながら、Maximum Combination Test の方が、検出力が Standard Combination Test と同程度かやや優れているという結果が観察されている．



## 4. SAS による実装例

本章では、Maximum Combination Test の SAS による実装でキーとなるポイントについて、アルゴリズムや特記事項を説明する。なお、1)本章で紹介する SAS プログラム、2)Maurer らの Supporting Information[16]にある R による実装例を SAS に移植したプログラム、3)Standard Combination Test の実装例、4)Potvin ら[8]の Example 1 および Example 2 を SAS で再現したプログラム(Method B/C)を Appendix として添付した。

### 4.1 想定する解析データセット構造

本章で想定する解析データセット構造を表 1 に示す。なお、データは Maurer[10]の Table 6(Example) と同一の結果が得られるように乱数を用いて作成した。

表 1. 本章で想定する解析データセット

STAGE	USUBJID	TRTSEQA	TRTA1	TRTA2	lnCmax1	lnCmax2
1	xx-01	RT	R	T	x.xxxx	x.xxxx
1	xx-02	RT	R	T	x.xxxx	x.xxxx
	:	:	:	:	:	:
1	xx-19	TR	T	R	x.xxxx	x.xxxx
1	xx-20	TR	T	R	x.xxxx	x.xxxx
2	xx-21	RT	R	T	x.xxxx	x.xxxx
	:	:	:	:	:	:
2	xx-55	TR	T	R	x.xxxx	x.xxxx
2	xx-56	TR	T	R	x.xxxx	x.xxxx

STAGE：ステージ 1／ステージ 2，USUBJID：被験者識別コード，TRTSEQA：順序群，TRTA1：第 I 期の製剤，TRTA2：第 II 期の製剤，lnCmax1：第 I 期の Cmax の自然対数変換値，lnCmax2：第 II 期の Cmax の自然対数変換値

### 4.2 想定する判定フロー

Maurer らは、Maximum Combination Test を用いたフローを提案している(図 4)。ステージ 1 で同等性評価パラメータの幾何平均比の両側 90%信頼区間が完全に 0.95 - 1.05 の範囲外の場合、無効中止とするルール(futility ルール)を設けている。この futility ルールを設定することで、検出力を損なうことなく、試験を繰り返した場合の期待被験者数を減らすことが可能であると Maurer らは述べている[10]。なお、futility ルールに抵触した場合、中止せずにステージ 2 を継続しても第一種の過誤確率は制御される(non-binding)。

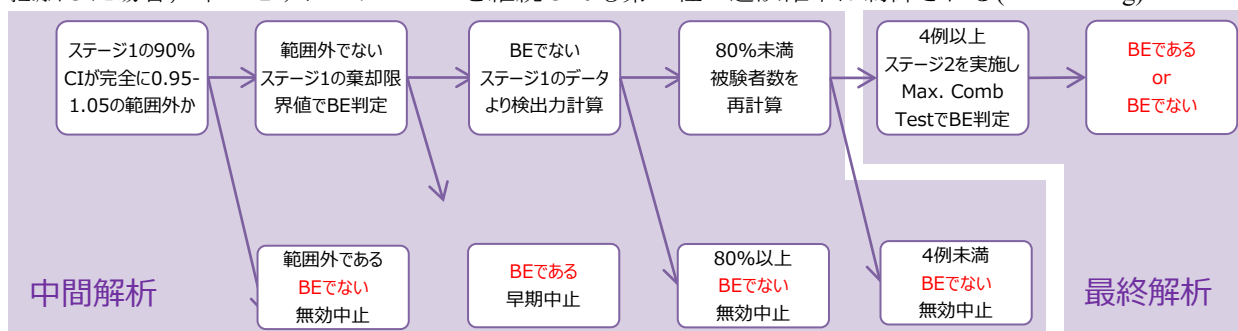


図 4. Maurer らの提案する Maximum Combination Test を用いたフロー[10]



### 4.3 重みに基づく棄却限界値および有意水準の決定

現時点では三変量以上の標準正規分布の累積分布に対応した SAS 関数がないため、Maximum Combination Test の重み  $w$  と  $w^*$  によって決定される相関構造をもつ三変量標準正規分布に従う正規乱数を大量に発生させ、3 つの検定統計量 ( $Z_{1j}$ ,  $Z_{0j}$ ,  $Z_{0j}^*$ ) の棄却限界値が Pocock の方法のように等しいという制約下で、試験全体の第一種の過誤確率が 5% になるような棄却限界値および有意水準を探索するというアルゴリズムとした。計算精度を担保するために可能な限り多くの乱数を発生させることが重要である点に留意されたい。

```
/*#####  
## Calculation of critical values for maximum combination test  
#####*/  
proc iml;  
  call randseed(12345678);  
  randN = 100000000; /* number of random variables, recommend >= 100,000,000 */  
  
  * choice of weights;  
  w1 = 0.5;  
  w2 = 1 - w1;  
  w1_star = 0.25;  
  w2_star = 1 - w1_star;  
  root_w1 = sqrt(w1);  
  root_w1_star = sqrt(w1_star);  
  rho=sqrt(w1 * w1_star) + sqrt(w2 * w2_star);  
  
  /* correlation matrix */  
  Sigma = (1 || root_w1 || root_w1_star) //  
           (root_w1 || 1 || rho) //  
           (root_w1_star || rho || 1);  
  mean = {0 0 0};  
  Z = randnormal(randN, mean, Sigma); /* sample from MVN(0, Sigma) */  
  create MVNorm3 from Z[c={Z1 Z0a Z0b}]; append from Z; close;  
  
  /*#####  
  ## function used in froot to solve for critical value  
  ## of maximum combination test  
  #####*/  
  start TVNCDF(x);  
    use MVNorm3;  
    read all into TRIVN;  
    ind = (TRIVN[,1] < x) & (TRIVN[,2] < x) & (TRIVN[,3] < x);  
    prob = ind[,];  
    f = (prob) - (1 - 0.05);  
    return (f);  
  finish TVNCDF;  
  
  * get critical value of robust combination test for z_max1 and z_max2;  
  z_crit_max = froot("TVNCDF", {0, 5});  
  
  * convert to nominal significance level;  
  alpha_max = 1 - probnorm(z_crit_max);  
  
  * store to macro variable;  
  call symput("z_crit_max", char(z_crit_max));  
  call symput("alpha_max", char(alpha_max));  
  print z_crit_max alpha_max;  
quit;
```

なお、Standard Combination Test の場合は、中間解析と最終解析の検定統計量  $Z_{1j}$  と  $Z_{0j}$  は、ポイント帰無仮説  $H_{0j}^*$  のもとで二変量標準正規分布に従うため、PROBBNRM 関数が利用できる。

```

/*#####
## significance levels for standard combination test for weight=0.25
#####*/
proc iml;
  * function to solve for critical value (equal alphas in stages 1 and 2);
  start BVNCDF(x);
    w1_star = 0.25;
    f = PROBBNRM(x, x, sqrt(w1_star)) - (1 - 0.05);
    return (f);
  finish BVNCDF;

  * get critical value of standard combination test;
  z_crit_stage1_star = froot("BVNCDF", {0, 5});

  * convert this to a significance level;
  alpha_stage1_star = 1 - probnorm(z_crit_stage1_star);

  print z_crit_stage1_star alpha_stage1_star;
quit;

```

#### 4.4 中間解析

TTEST プロシジャの TOST オプションを指定し、2つの片側検定の p 値を出力し、BE 判定を行う。また、futility ルールに該当するか判定するための両側 90%信頼区間や被験者数再設計を行うために各種統計量を ODS ステートメントで出力するように指定する。

```

*****;
* Interim Analysis;
*****;
ods output Statistics=stat1 Conflimits=cl1 EquivLimits=ev1 EquivTests=et1;
proc ttest data=stage1plus2 tost(%sysfunc(log(0.80)), %sysfunc(log(1.25))) alpha=0.05
order=internal;
  where STAGE = 1;
  var lnCmax1 lnCmax2 / crossover=(TRTA1 TRTA2) ;
run;

```

#### 4.5 被験者数再設計

power の計算にあたっては、2 変量非心 t 分布の確率計算を行う必要があるが、本稿の執筆時点では対応する SAS 関数がないため、Genz and Bretz のプログラム(probmvmt.sas)を一部改変(probmvmt.sas 中の IML サブルーチンおよび変数等を外部から参照できるように、最終行の QUIT ステートメントの直前に STORE ステートメントを追記)して利用した[17]。被験者数を 2 例ずつ増やしながら、MVN\_DIST サブルーチンを用いて、estimated conditional target power を達成する最小の被験者数を求める。

```

*-----;
* power function for unequal significance levels;
*-----;
%include "/home/u#####/probmvmt.sas"; *specify the folder;

proc iml;
  load ; *load modules in probmvmt.sas;

  lnGmrPlanned = log(0.95);
  lnlp25 = log(1.25);
  variancel = &var_hat_stage1.; *estimated intra-subject variance on log-scale;

  COVAR = {1 -1,
            -1 1};
  acheived = 0;

```

```

do N = %eval(&n_stage1.+2) to 200 by 2 until(acheived);

    DIM = 2;
    DF = N - 2;

    nc1 = (lnlp25 + lnGmrPlanned) /sqrt(variance1 * 2 / N);
    nc2 = (lnlp25 - lnGmrPlanned) /sqrt(variance1 * 2 / N);
    DELTA = j(1, 2);
    DELTA[1] = nc1;
    DELTA[2] = nc2;

    INFIN = J(1, DIM, 1);
    tc1 = tinv(1 - &alpha_c1_max., DF); *estimated conditional error rate for H01;
    tc2 = tinv(1 - &alpha_c2_max., DF); *estimated conditional error rate for H02;
    LOWER = j(1, 2);
    LOWER[1] = tc1;
    LOWER[2] = tc2;
    UPPER = j(1, 2, 999);*infinity:999 is ignored;

    MAXPTS = 2000*DIM*DIM*DIM;
    ABSEPS = .000001;
    RELEPS = 0;

    RUN MVN_DIST( DIM, DF, DELTA, LOWER, UPPER, INFIN, COVAR, MAXPTS, ABSEPS, RELEPS,
ERROR, condpower, NEVALS, INFORM );

    if (condpower >= &estimated_conditional_target_power.) then acheived = 1;
    print N condpower;
    if acheived = 1 then do;
        create SSR from N[c={N}];
        append from N;
        close;
    end;
end;
print N;
quit;

```

#### 4.6 最終解析(Maximum Combination Test)

最終解析では、ステージ2のデータから4.4節と同様に各種統計量を算出し、同じくステージ1のデータから算出した統計量と逆正規法による統合を行い、Maximum Combination TestによるBE判定を行う。

```

*-----;
* Maximum Combination Test;
*-----;
data maxcomb;
    merge tost1st tost2nd;
    w = 0.5;
    w_star = 0.25;
    z_crit_max = 1.9374;

    * z-statistics;
    z1_w = sqrt(w) * z1_stage1 + sqrt(1 - w) * z1_stage2;
    z2_w = sqrt(w) * z2_stage1 + sqrt(1 - w) * z2_stage2;

    z1_w_star = sqrt(w_star) * z1_stage1 + sqrt(1 - w_star) * z1_stage2;
    z2_w_star = sqrt(w_star) * z2_stage1 + sqrt(1 - w_star) * z2_stage2;

    z1_max = max(z1_w, z1_w_star);
    z2_max = max(z2_w, z2_w_star);

    * reject null hypothesis at left side?;
    nrejlw = (z1_max >= z_crit_max);

```

```

* reject null hypothesis at right side?;
nrejupp = (z2_max >= z_crit_max);

* reject null hypothesis at both sides?;;
* i.e., is ABE accepted?;
z_passed = (nrejlw & nrejupp); *1:ABE accepted, 0:ABE rejected;

keep w w_star z_crit_max z1_w z1_w_star z1_max z2_w z2_w_star z2_max nrejlw nrejupp
z_passed;
run;

```

## 4.7 シミュレーション

被験者内変動係数の設定値に基づいて乱数を発生させる方法など詳細なプログラムは Appendix を参照されたい。シミュレーションの出力として、試験性能の比較・評価のため、ステージ1で中止する確率(BE, BEでないそれぞれ)、ステージ2に移行する確率、ステージ1またはステージ2でBEと判定する確率(検出力または第一種の過誤確率に相当する)、総被験者数の要約統計量(平均、分位点)などを算出することが有用である(表2)。

表2. シミュレーション結果の提示例

設定	幾何平均比	被験者内変動係数 (%)	ステージ1でBEと判定する確率	ステージ1でBEでない判定する確率	ステージ2へ移行する確率	ステージ2でBEと判定する確率	ステージ1または2でBEと判定する確率	期待総被験者数
n <sub>1</sub> =24 w=0.5 w*=0.25	0.8	20	0.####	0.####	0.####	0.####	0.####	##.#
		30	0.####	0.####	0.####	0.####	0.####	##.#
		40	0.####	0.####	0.####	0.####	0.####	##.#
	1.0	20	0.####	0.####	0.####	0.####	0.####	##.#
		30	0.####	0.####	0.####	0.####	0.####	##.#
		40	0.####	0.####	0.####	0.####	0.####	##.#

## 5. その他、考慮すべき事項

本章では、試験計画時や報告時に考慮すべき事項について述べる。

まず、中間解析を含む2段階デザインを採用するか、それとも中間解析を伴わない通常の固定デザインを採用するか検討が必要となる。Kazaらは、2段階デザインの採用を判定するためのスキームを提案している[18]。このスキームは日本の規制を意識したものではないが、判断ポイントとして、1)溶出挙動の類似性(f2関数)、2)信頼できる被験者内変動係数の推定値の有無、3)被験者内変動係数の大きさが30%を越えるかが挙げられており、指針を与えてくれる。

2段階デザインを採用する場合、試験計画段階で検討・決定し、試験実施計画書に明記すべきパラメータが複数ある。BEガイドラインは、中間解析手順、第一種の過誤確率の制御方法、追加する例数の算出方法、追加データの併合手順等を含めた実施条件を事前に明記することを求めている[2]。加えて、Maximum Combination Testの2つの重み、ステージ1の被験者数、無効中止基準等も明記すべきパラメータとして考えられる。これら種々のパラメータ検討のためにシミュレーションによる試験の動作性能の評価が必要であり、通常の固定デザインよりも計画により多くのリソースや時間が必要になりうる点に留意が必要である。また、被験者数再計算を伴わない群逐次デザインも選択肢として検討する必要があるだろう。

中間解析結果に基づいて試験デザインに変更を行うことを計画するアダプティブデザインでは、一般に、インテグリティの維持への配慮が重要となってくる[19]。バイオアベイラビリティを比較する BE 試験では、試験計画および解析計画が詳細に事前宣言されている限りは、中間解析の解析結果がスポンサーに秘匿されなくても、さほど影響がない可能性もある。しかし、割付情報の管理、中間解析実施とその報告等について事前に手順を定めておくことが必要になる。

試験のオペレーション面での検討事項もある。中間解析で追加被験者数を決定あるいは早期中止となる可能性があるため、予定被験者数は幅をもった記載となる。これを実施医療機関、CRO(開発業務受託機関)、薬物濃度測定機関等の関連機関にも予め説明し、合意を得たうえで、業務委託契約書や必要に応じて業務手順書等に反映する必要がある。加えて、現実には追加可能な被験者数には上限があり、被験者数の上限が、コストやスケジュール、実施医療機関等での受託可否などに影響を及ぼすことを考慮する必要がある。

また、データマネジメントに関しても、中間解析のために速やかにデータクリーニングを行い、タイムリーに質の高いデータを提供できることを保証するための業務手順を予め策定することが求められる。

治験総括報告書など試験結果の報告でも、試験計画、中間解析および最終解析の結果について詳細に報告することが必要である。報告すべき事項については、FDA のアダプティブデザインのガイダンス[19]およびランダム化比較試験を報告する際の CONSORT チェックリストをアダプティブデザイン試験に拡張した ACE チェックリスト[20]が指針を与えてくれる。

以上、被験者数再計算を伴う 2 段階デザインの恩恵を得るためには、生物統計担当者だけではなく関連部門・機関で連携し、綿密な試験計画の立案とその着実な遂行、適切な報告もまた重要である。

## 6. まとめ

令和 2 年 3 月の BE ガイドラインの一部改正により認められることとなった被験者数再計算を伴う 2 段階デザインに対応しうる解析手法として Maximum Combination Test および被験者数再計算アルゴリズムを関連手法とともに紹介し、SAS による実装方法を提案した。本稿の執筆時点では、本邦で中間解析を伴う 2 段階デザインを採用した BE 試験の報告は見つけられなかった。今後、事例の報告が増え、経験や課題が蓄積・共有されていくこと、および更なる改善法が提案されていくことで、改正後の BE ガイドラインの実装が推進されていくことを期待する。本稿がその一助になれば幸いである。

連絡先：morita.yusuke@nobelpharma.co.jp

### Appendix：本稿に添付する SAS プログラム

プログラム名	内容
max_comb.sas	Maximum Combination Test の実装例
max_comb_simulation.sas	Maximum Combination Test の性能評価のシミュレーションの実装例
maurer_example.sas	Maurer らの Appendix の R プログラム[16]を SAS に移植したもの
std_comb.sas	Standard Combination Test の実装例
potvin_example1.sas	Potvin[8]の Example 1(Method B および C)を SAS で再現したもの
potvin_example2.sas	Potvin[8]の Example 2(Method B および C)を SAS で再現したもの

なお、プログラムの動作確認は SAS® OnDemand for Academics (2021 年 9 月時点で SAS9.4TS1M6 相当)で実施した。

## 引用文献

- [1] 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, 別紙 1)
- [2] 後発医薬品の生物学的同等性試験ガイドライン Q&A(令和 2 年 3 月 19 日事務連絡, 別紙 1)
- [3] 後発医薬品 品質情報 No.14 令和 2 年 8 月(厚生労働省 医薬・生活衛生局医薬品審査管理課)
- [4] 含量が異なる経口固形製剤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, 別紙 2)
- [5] 経口固形製剤の処方変更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, 別紙 3)
- [6] 剤形が異なる製剤の追加のため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, 別紙 4)
- [7] Kuribayashi R, Yamaguchi T, Takagi K. Modernization and Strengthening of Bioequivalence Guidelines in Japan. Clin Pharmacokinet. 2021 Feb;60(2):145-151.
- [8] Potvin D, DiLiberti CE, Hauck WW, Parr AF, Schuirmann DJ, Smith RA. Sequential design approaches for bioequivalence studies with crossover designs. Pharm Stat. 2008;7(4):245-62.
- [9] Montague TH, Potvin D, Diliberti CE, Hauck WW, Parr AF, Schuirmann DJ. Additional results for Sequential design approaches for bioequivalence studies with crossover designs. Pharm Stat. 2012;11(1):8-13.
- [10] Maurer W, Jones B, Chen Y. Controlling the type I error rate in two-stage sequential adaptive designs when testing for average bioequivalence. Stat Med. 2018;37(10):1587-607.
- [11] 棚橋 昌也, 菅波 秀規. 中間解析, サンプルサイズ再計算を伴う生物学的同等性試験における統計的課題. ジェネリック研究 2021 ;15: 020-030.
- [12] Xu J, Audet C, Diliberti CE, Hauck WW, Montague TH, Parr AF, et al. Optimal adaptive sequential designs for crossover bioequivalence studies. Pharm Stat. 2016;15(1):15-27.
- [13] European Medicine Agency. Committee for Medicinal Products for Human Use (CHMP) Guideline on the investigation of bioequivalence. CPMP/EWP/QWP/1401/98 Rev 1/ Corr \*\*. 2010 January 20 .
- [14] Lee J, Feng K, Xu M, Gong X, Sun W, Kim J, Zhang Z, Wang M, Fang L, Zhao L. Applications of Adaptive Designs in Generic Drug Development. Clin Pharmacol Ther. 2021 Jul;110(1):32-35.
- [15] Kieser M, Rauch G. Two-stage designs for cross-over bioequivalence trials. Stat Med. 2015;34(16):2403-2416.
- [16] Maurer W, Jones B, Chen Y. Supplementary Information for "Controlling the Type I error rate in two-stage sequential adaptive designs when testing for Average Bioequivalence"  
<https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fsim.7614&file=sim7614-sup-0001-supplementary.pdf>
- [17] Genz, A. and Bretz, F. Methods for the computation of multivariate t-probabilities. Journal of Computational and Graphical Statistics, 2002; 11, 950-971.
- [18] Kaza M, Sokolovsky A, Rudzki PJ. 10th Anniversary of a Two-Stage Design in Bioequivalence. Why Has it Still Not Been Implemented? Pharm Res. 2020 Jul 13;37(7):140.
- [19] Food and Drug Administration. Adaptive designs for clinical trials of drugs and biologics: guidance for industry. 2019. (邦訳 : [https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000n5a8-att/adaptive\\_design.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000n5a8-att/adaptive_design.pdf))
- [20] Dimairo M, Pallmann P, Wason J, Todd S, Jaki T, Julious SA, Mander AP, Weir CJ, Koenig F, Walton MK, Nicholl JP, Coates E, Biggs K, Hamasaki T, Proschan MA, Scott JA, Ando Y, Hind D, Altman DG; ACE Consensus Group. The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. BMJ. 2020 Jun 17;369:m115.