

アダプティブデザインの 理論と性能に関する話題

興和株式会社
丸尾和司
k-maruo@kowa.co.jp

1

議題案

- 症例数再設計の有用性の再確認
 - Bindingの実装可能性について
 - 推定値のバイアスについて
-

2

症例数再設計の 有用性の再確認

想定および表記

- 2群, プラセボ対照, 並行群間比較試験
- エンドポイント: 連続量
 - ⇒ プラセボ群: $N(\mu_1, 1)$, 実薬群: $N(\mu_2, 1)$
- 中間解析: 1回
 - ⇒ 症例集積100r%のとき, 症例数再設計(SSR)を実行
- エフェクトサイズ(尺度調整済み)
 - ⇒ 真値: $\delta = \mu_2 - \mu_1 \geq 0$
 - ⇒ 計画時の想定: $\delta_a > 0$
 - ⇒ 中間解析時の推定値: $\hat{\delta}_{IA}$
- 計画段階における最終解析の各群の症例数: n
 - ⇒ δ_a に基づき設計
 - ⇒ 検出力: 80%, α : 5%
- SSR後の各群の症例数: \hat{n}
- 解析: 2標本t検定
 - ⇒ 中間解析, 中間解析~最終解析: T_1, T_2

統合検定 (Cui *et al.*, 1999)

□ 症例数再設計のルール

$$\Rightarrow \hat{n} = \left(\frac{\delta_\alpha}{\delta_{IA}} \right)^2 n$$

□ 一般に用いられる+αの指針

$$\Rightarrow \hat{n} < n \rightarrow \hat{n} = n$$

$$\Rightarrow \hat{n} > cn \rightarrow \hat{n} = cn$$

• e.g., c=2

□ 統合検定統計量

$$\Rightarrow t_c = \sqrt{r} t_1 + \sqrt{1-r} t_2$$

⇒ 実際に用いた \hat{n} でなく事前特定された n に基づく重みを用いる

• \hat{n} が t_1 に依存するので、帰無分布が標準正規とならないため

⇒ 十分性原理を満たさない (one patient one voteとならない)

• 効率低下

⇒ Inverse normalで等価な方法を構成可能 (e.g., Bretz *et al.*, 2009)

5

制約付きSSR (Chen *et al.*, 2004)

□ 条件付き検出力: CP

$$\Rightarrow CP(t_1, r) = 100\Phi \left(\frac{t_1}{\sqrt{r(1-r)}} - \frac{z_{1-\alpha/2}}{\sqrt{1-r}} \right)$$

z_p : 標準正規分布の100pパーセント点

□ 症例数再設計のルール

⇒ $P_{min} < CP(t_1, r) < 80$: promising zone

$$\rightarrow \hat{n} = rn \left\{ 1 + \frac{1}{t_1^2} \left(\frac{1}{\sqrt{1-r}} z_{1-\alpha/2} - \sqrt{\frac{r}{1-r}} t_1 + z_{1-\beta} \right)^2 \right\}$$

• $P_{min} = 50$: Chen *et al.* (2004)

• 状況によってももう少し下げられる: Gao *et al.* (2008)

– +αルール: if $\hat{n} > cn$ then $\hat{n} = cn$

⇒ other → $\hat{n} = n$

• $CP(t_1, r) < P_{min}$: unfavorable, $CP(t_1, r) > 80$: favorable,

□ 上記のルールに従えば、単純な検定を実施可能

⇒ ルールがIDMCの権限で破られた場合にはインフレの恐れ (Hung *et al.*, 2014)

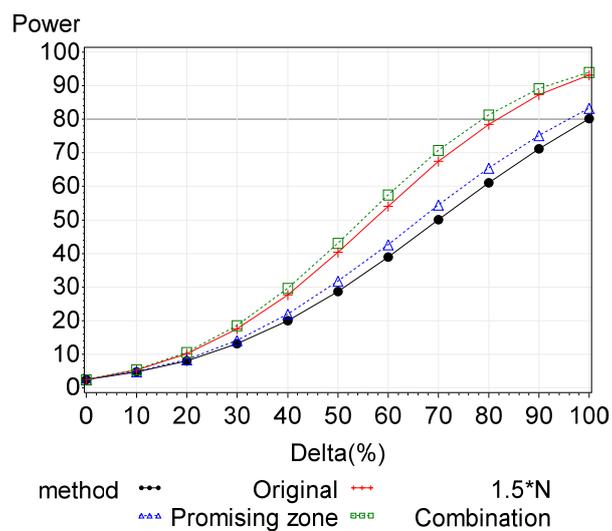
6

シミュレーション

- 目標: 統合検定と制約付きSSRの動作特性と、どのような場合にSSRが有用になるのかの確認
 - ⇒ Mehta and Pocock (2011); Emerson *et al.*(2011)への追加・改変
- 前述の表記+
 - ⇒ $n = 100$
 - このとき、検出力が80%となるような $\delta (= \delta_a) = 0.398$
 - ⇒ 中間解析(IA): 50%の症例集積
 - SSRの性能のみを検討するため、あえて無効、有効中止を入れない
 - 無効中止を考慮した場合は、Appendix参照
 - ⇒ デザイン
 - SSRなし(original), SSRなし($1.5n$), SSR(統合検定), 制約付きSSR
 - promising zone: 36.5%~80% (Gao *et al.*, 2008)
 - 両SSRの上限: $2n$
 - ⇒ $\delta: \delta_a$ の0% to 100% by 10% ←エフェクトサイズが想定以下
- シミュレーション回数100,000
 - ⇒ 最終解析で有意になった割合をシミュレートされた検出力として計算
 - SSRにおける中間解析での決定ごとにも検討
 - ⇒ 期待症例数

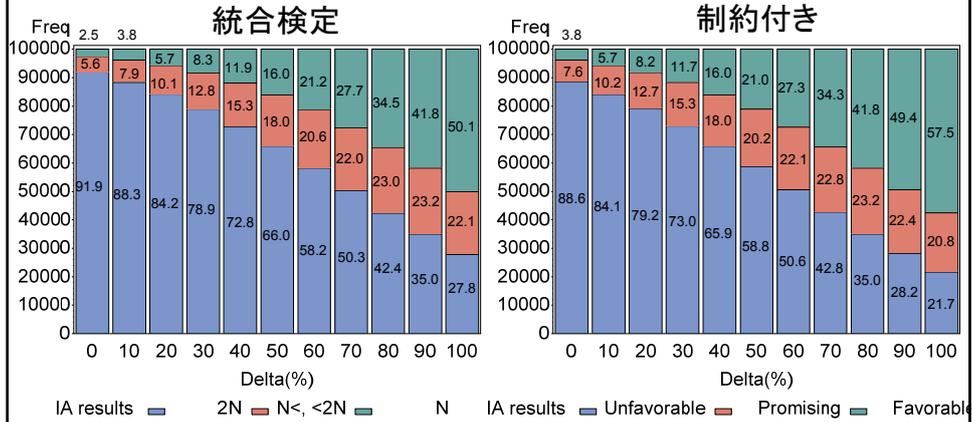
7

検出力

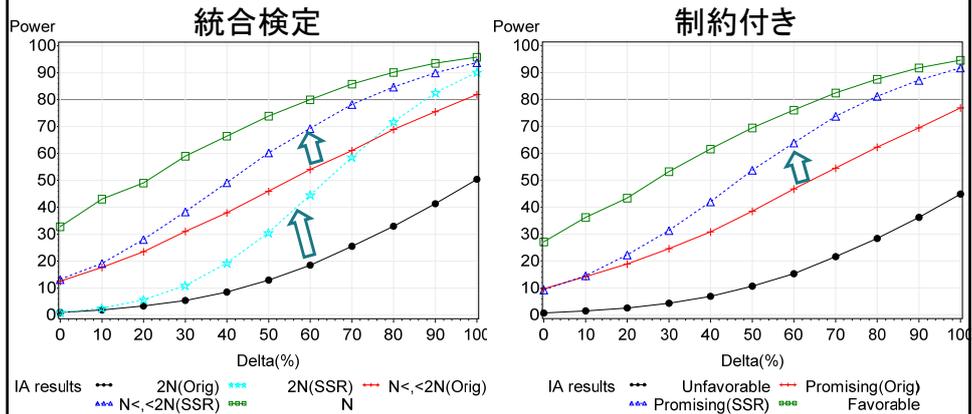


8

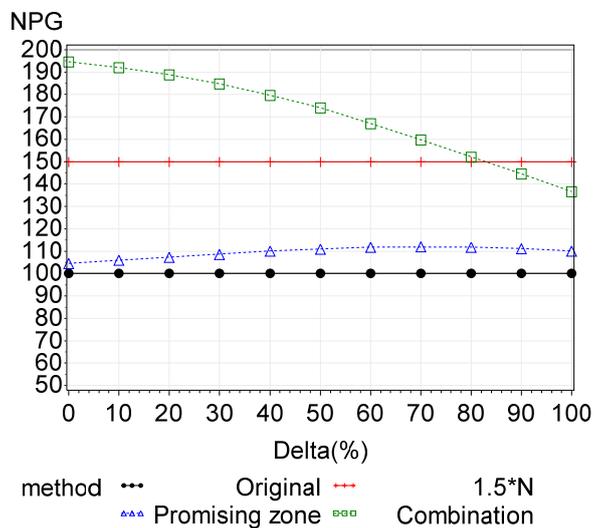
中間解析における決定の結果



中間解析における決定ごとの検出力



期待症例数



11

論点

- どのような場合にSSRが有用となるか
 - ⊙ 切り口によって、有用な方法が変わりそう
 - ⊙ 効かないときには増やさない or 増やす
 - futility等の追加によって変わりそう: 通常のデザインに追加しても同様?
- リスク(コスト)ベネフィットの観点から、どの程度までアグレッシブなアダプテーションを実装するか
 - ⊙ ベネフィット: 検出力, 統計部門の存在感など
 - ⊙ コスト: オペレーションや治験薬など
 - ⊙ リスク: SSRで2倍に増やしてN.S.
 - アダプテーションは万能でなく、「仕込み」が効く可能性がそれほど高いわけではない
 - 理論的には, in the long runで性能が高くと、1企業としては高々数回(特に初回+α)の成功の有無が重要
 - 特に統計家が能動的に動いている分
 - ⊙ 統計家としてのスタンス・社内啓蒙状況に異存

12

Bindingの実装可能性について

13

Binding futility rule

- 中間解析で無効中止を含めた場合に、このことに起因する検定サイズの減少を考慮したうえで、棄却限界値を計算すること
 - ⇒ たとえば、50%の症例集積で、 $CP < 5\%$ のとき無効中止と判定すると、最終解析の名義の水準を5%にすると、実質検定サイズは4.65%になる
 - ⇒ 無効中止で損した分を取り戻す
 - ⇒ ⇔ 考慮しない場合: non-binding
- 統計的に無効中止が宣言されても、IDMCの判断で試験継続になった場合は α エラーがインフレの恐れ
- たとえば、Bretz *et al.*(2009):
 - ⇒ Binding: 検出力が上がる
 - ⇒ Non-binding: 柔軟性をもつ
 - ⇒ 実装されるかはケースバイケースで決定すべき

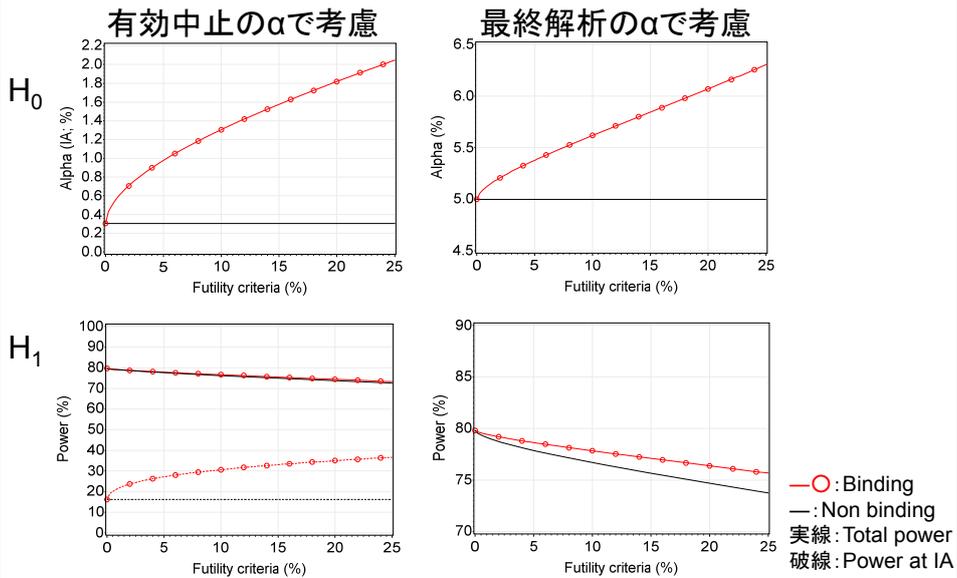
14

Bindingの効能検討

- 共通(定義や他の条件はSSRの検討における設定と同様)
 - ⇒ 症例集積50%で中間解析
 - ⇒ 無効中止基準: $CP < 0 \sim 25\%$
 - 0: 無効中止なし
 - ⇒ 無効中止がない場合, 最終解析の検出力: 80%
 - $n\delta^2 = 2(z_{1-\alpha/2} + z_{1-\beta})^2$
 - $\delta_\alpha = \delta$
- Bindingで「取り戻した」サイズを有効中止の α に配分
 - ⇒ Bindingがない場合の有効中止の判断: O'Brien-Fleming型
 - ⇒ 中間解析の $\alpha = 0.305\%$ **<= ここに配分**
 - ⇒ 最終解析の $\alpha = 4.900\%$
- Bindingで「取り戻した」サイズを最終解析の α に配分
 - ⇒ 有効中止無し
 - ⇒ 最終解析の $\alpha = 5\%$ **<= ここに配分**

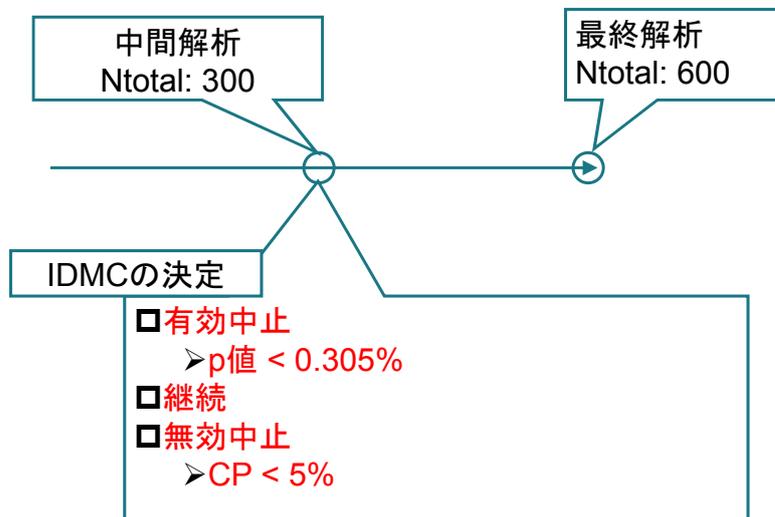
15

検討結果



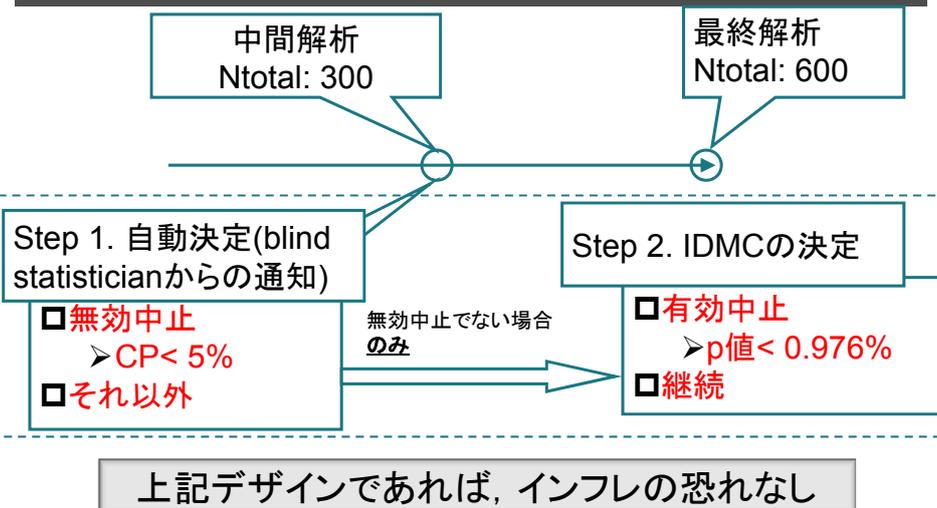
16

群逐次デザイン例(通常: Non binding)



17

群逐次デザイン例(Bindingを考慮した例)



18

論点

- Bindingを実装できる状況の検討
 - ⇒ 有効中止の α にオンすると効率が良い？
- 統計学的観点のみで無効中止を決める是非
 - ⇒ 適用できる疾患は限られる？
 - life threateningでないなど
 - ⇒ IDMC委員の同意が必要

19

推定値のバイアスについて

20

推定値のバイアス: 論点

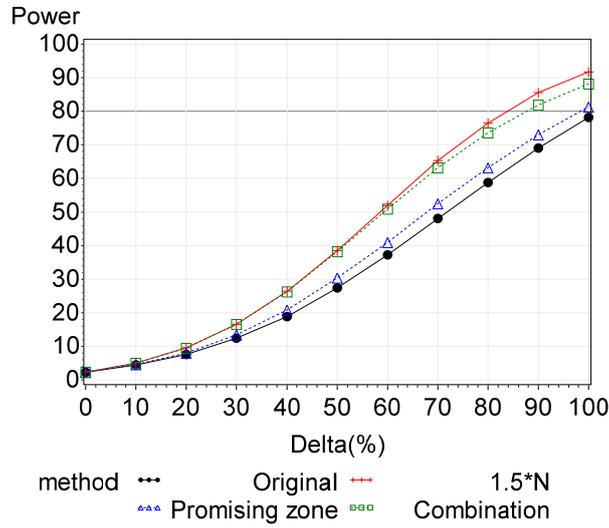
- SSR, シームレスP2/3, 有効中止などに起因するバイアス, 信頼区間の性能低下
 - ⇒ Jenison and Turnbull (2000), Bretz et al.(2009), Levin et al. (2014)など
- オペレーショナルバイアス
 - ⇒ period(IA前と後)*groupの交互作用
 - ⇒ 1試験内の経験蓄積によるトレンド?
- α エラーのインフレについては十分制御できているもとの, これらのことをどこまで考慮すべきか
 - ⇒ 上記の不都合は, 検定の文脈における対立仮説下での不都合であり, 帰無仮説下では起こりえないか無視可能
- 検証試験において「事前特定して」示すべきこととは
 - ⇒ 統計的有意性(必要条件; Neyman-Pearson流)
 - ・ α エラーのインフレ回避
 - ・ 帰無仮説の定義
 - ⇒ 臨床的有意性: 薬効推定値(十分条件)
 - ・ 主解析に付随する薬効推定値を絶対視する
 - ・ 事後的に(産官学ともに受容できる?) 妥当な推定値を模索する: 現状行われていない

21

付録: SSRの性能評価 無効中止(CP<5%)考慮

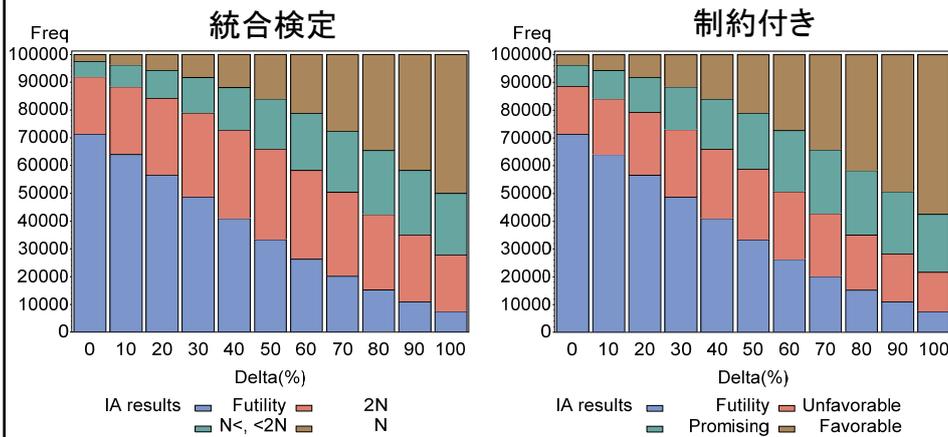
22

検出力



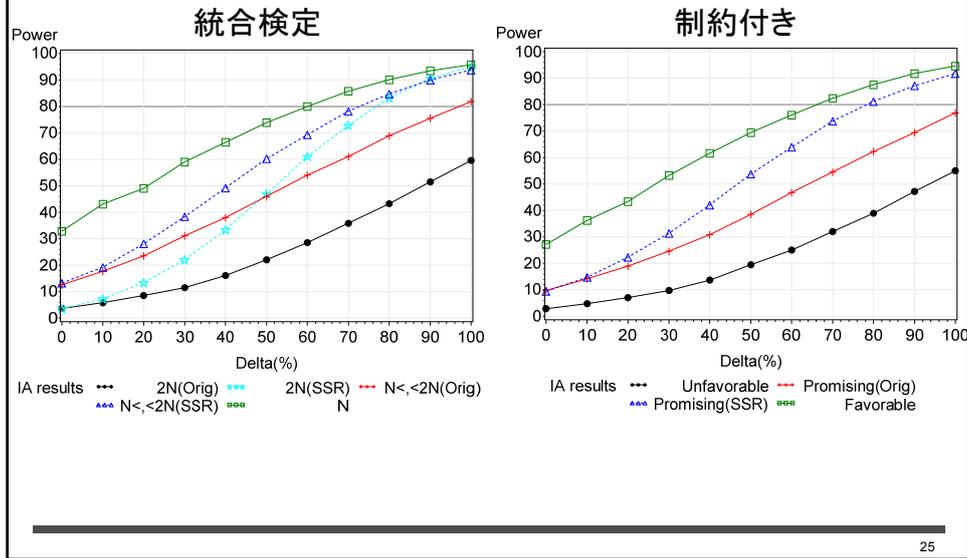
23

中間解析における決定の結果

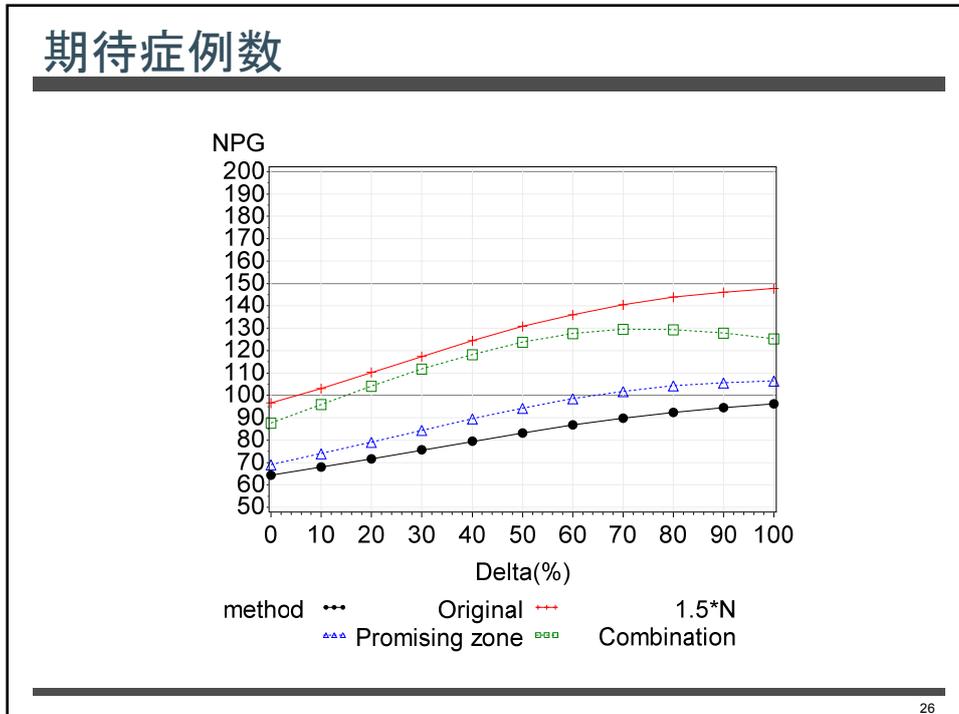


24

中間解析における決定ごとの検出力(無効中止以外)



期待症例数



参考文献

- Bretz et al. Adaptive designs for confirmatory clinical trials. *Statistics in Medicine* 2009; 28(8), 1181–1217.
- Chen, Y. H. J., DeMets, D. L., and Lan, K. K. G. Increasing the sample size when the unblinded interim result is promising. *Statistics in Medicine* 2004; 23, 1023-1038.
- Cui, L., Hung, H. M. J., and Wang, S. J. Modification of sample size in group sequential clinical trials. *Biometrics* 1999; 55, 853-857.
- Gao, P., Ware, J. H., Mehta, C. Sample size re-estimation for adaptive sequential design in clinical trials. *Journal of Biopharmaceutical Statistics* 2008; 18, 1184-1196.
- Emerson, S. S., Levin, G. P., and Emerson, S. C. Comments on 'Adaptive increase in sample size when interim results are promising: A practical guide with examples'. *Statistics in Medicine* 2011; 30, 3085-3301.
- Jennison, C. and Turnbull, B. W. *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC, 2000 [森川敏彦・山中竹春 訳. 臨床試験における群逐次法: 理論と応用. CAC, 2012].
- Levin, G. P., Emerson, S. C., and Emerson, S. S. An evaluation of inferential procedures for adaptive clinical trial designs with pre-specified rules for modifying the sample size. *Biometrics* 2014; 70, 556-567.
- Mehta, C. and Pocock, S. J. Adaptive increase in sample size when interim results are promising: A practical guide with examples. *Statistics in Medicine* 2011; 30, 3267-3284.
- 上村鋼平. 臨床試験における被験者数再設定: 方法論の概説と統計学的留意点. *計量生物学* 2012; 33, 77–99.