群逐次デザインの臨床試験において Multiple Imputationを用いることによる 検定サイズへの影響の検討

石井 敦, 丸尾 和司, 菅波 秀規
興和株式会社 臨床解析部

Evaluation for test size of multiple imputation in group sequential clinical trials

Atsushi Ishii, MSc*, Kazushi Maruo, PhD*, Hideki Suganami, PhD*
*Clinical Data Science Department, Kowa Company, LTD. Tokyo, Japan

要旨:

群逐次デザインの臨床試験において，中間解析と最終解析のそれぞれにMIを適用すると検定統計量間の相関が理論値より低くなり，検定サイズの増大が懸念される。本稿ではMIプロシジャに基づきこれらの影響を検討する。

キーワード: Multiple Imputation，中間解析，群逐次デザイン，
MIプロシジャ

はじめに

- 従来，LOCFのような単一補完法が臨床試験に広く使用されてきた(Tanaka *et al.*，2014).

⇓

- 近年では，状況に応じた手法が利用されてきている(NRCレポート，2010).
  - Multiple Imputation(MI)は広く用いられている(Carpenter & Kenward，2013).

<Mutiple Imputation(MI)>

- 欠測値に対して，統計モデルに基づきランダムに複数回の補完を行う方法
  - 統計モデル：MCMC法，予測モデル法，傾向スコア法
  - 補完に起因する不確実性を考慮に入れることが可能

有効中止を目標とした群逐次デザインの臨床試験

- 複数の治療の有効性を比較
  - 治療効果の差の検討
- 有効性の評価指標
  - 連続量
  - 経時的に複数時点で測定
- 主要な評価時点
  - 最終測定時点
- 解析時点
  - 中間解析，最終解析
- 群逐次デザインの仮説検定の棄却限界値
- 脱落等により，最終測定時点が欠測の症例が存在
  - 単調な欠測を仮定
- MIに基づく補完

補完方法：従来法

- 中間解析と最終解析それぞれでMIを行う
  - 中間解析と最終解析の同一症例の補完値が異なる場合がある

中間解析

id	時点1	時点2	・・・	時点k
1	○	○	・・・	. → x
2	○	○	・・・	○
3	○	○	・・・	○
4	○	○	・・・	. → y
5	○	○	・・・	○
:	:	:	:	:
n	○	.	・・・	. → z

補完値

n例の時点kに
おける欠測を補完

最終解析

id	時点1	時点2	・・・	時点k
1	○	○	・・・	. → a
2	○	○	・・・	○
3	○	○	・・・	○
4	○	○	・・・	→ b
5	○	○	・・・	○
:	:	:	:	:
n	○	.	・・・	. → c
n+1	○	○	・・・	○
n+2	○	○	・・・	○
n+3	○	○	・・・	○
n+4	○	○	・・・	○
n+5	○	○	・・・	.
:	:	:	:	:
$2n$	○	.	・・・	.

補完値

$2n$例の時点kに
おける欠測を補完

従来法

- 中間解析と最終解析の同一症例の補完値の違いにより検定サイズの増大が懸念
  - 群逐次デザインの仮説検定の棄却限界値
    - 分散既知のもと，中間解析と最終解析の検定統計量間の相関を中間解析時点での症例の集積割合の平方根と仮定して算出する

欠測が存在し，MIに基づく補完を行う場合

    - MIに基づく補完後に得られる，中間解析と最終解析の検定統計量間の相関(真の相関と呼ぶ)が仮定された相関より小さくなる
      - ✓ 仮定された相関の上方バイアスが生じる

- 中間解析の集積割合50%の場合

仮定された相関: $\sqrt{\frac{1}{2}} = 0.707$ $\geq$ 真の相関(例): 0.65

本研究の目的

- 群逐次デザインの臨床試験における欠測の対処法としてMI法を用いる
  - 従来法における真の相関の低下を抑えるための提案
  - 従来法と提案法の中間解析と最終解析の検定統計量間の相関の挙動，及び試験全体の検定サイズの評価

- 発表の概要
  - 予測モデル法によるMIの手順及び，群逐次デザイン
  - 従来法と提案法の詳細
  - シミュレーションに基づく従来法と提案法の性能比較

予測モデル法を用いた補完手順(1)

- Notation

治療群: $g = 1,2$
測定時点: $t = 1,2,\ldots,T$
各治療群の症例数: n(各群の症例番号: $i = 1,2,\ldots,n$)
第j時点($j \geq 2$)の欠測データ: $y_{gij}^{(m)}$
※単調欠測とする
第$j-1$時点までの応答ベクトル: $\boldsymbol{y}_{gi} = \left(y_{gi1},\ldots,y_{gij-1}\right)$
第j 時点の観測データ: y_{gij}
重回帰モデルパラメータ: $\boldsymbol{\beta_g} = (\beta_{g0},\ldots\beta_{gj-1})^T$
重回帰モデルの誤差変動項: $\varepsilon_{gij} \sim N(0,\sigma_{\cdot\cdot j}^2)$
自由度($n_{gj}-j-1$)のχ^2分布に従う確率変数: γ
j 個の独立な標準正規確率変数ベクトル: $\boldsymbol{Z}$

予測モデル法を用いた補完手順(2)

step1：$\boldsymbol{y}_{gi}$を説明変数とし，y_{gij}を応答変数として，群ごとに重回帰モデルをあてはめ，
パラメータ$\boldsymbol{\beta}_g$の推定値$\widehat{\boldsymbol{\beta}}_g = (\hat{\beta}_{g0}, \ldots, \hat{\beta}_{gj-1})^T$，共分散行列$\hat{\sigma}_{..j}{}^2\boldsymbol{V}_{gj}$を推定

$$y_{gij} = \beta_{g0} + \beta_{g1}y_{gi1} + \cdots + \beta_{gj-1}y_{gij-1} + \varepsilon_{gij}$$

$V_{gj} = \left(X_g^T X_g\right)^{-1}$，$\boldsymbol{X}_g$: $\boldsymbol{y}_{gi}$により構成される計画行列

step2：推定された$\widehat{\boldsymbol{\beta}}_g$，$\hat{\sigma}_{..j}{}^2$及び$\boldsymbol{V}_{gj}$に基づき，パラメータの事後予測分布から新たな
パラメータ$\boldsymbol{\beta}_{*g}$，$\sigma^2_{*..j}$をサンプリング

$$\sigma^2_{*..j} = \hat{\sigma}_{..j}{}^2(n_{gj} - j - 1)/\gamma$$

$$\boldsymbol{\beta}_{*g} = \widehat{\boldsymbol{\beta}}_g + \sigma_{*..j}\boldsymbol{V}_{\boldsymbol{ghj}}\boldsymbol{Z}$$

$\boldsymbol{V}_{\boldsymbol{ghj}}$はコレスキー分解$\boldsymbol{V}_{\boldsymbol{gj}} = \boldsymbol{V}_{\boldsymbol{ghj}}^{\boldsymbol{T}}\boldsymbol{V}_{\boldsymbol{ghj}}$の上三角行列

step3：$\boldsymbol{\beta}_{*g}$，$\sigma^2_{*..j}$及び標準正規乱数z_{gi}に基づき欠測データ$y_{gij}^{(m)}$を補完

$$\beta_{*g0} + \beta_{*g1}y_{gi1} + \cdots + \beta_{*gj-1}y_{gij-1} + z_{gi}\sigma^2_{*..j}$$

予測モデル法を用いた補完手順(3)

step4：step1～3を欠測データそれぞれにM回繰り返し，M個の補完された完全データを生成

step5：M個の完全データに関して，治療効果の差の推定値$\hat{\theta}^{(1)},\ldots,\hat{\theta}^{(M)}$及び分散推定値$\hat{V}^{(1)},\ldots,\hat{V}^{(M)}$を算出

step6：step5の推定値を用いて推定値$\hat{\theta}$及び分散推定値 $\hat{V}$を算出(Rubin，1987).

$$\hat{\theta} = \frac{1}{M}\sum_{\ell=1}^{M}\hat{\theta}^{(\ell)}\ ,\ \hat{V} = \widehat{W} + \left(1+\frac{1}{M}\right)\hat{B}$$

補完内分散：$\widehat{W} = \frac{1}{M}\sum_{\ell=1}^{M}\hat{V}^{(\ell)}$

補完間分散：$\hat{B} = \frac{1}{M-1}\sum_{\ell=1}^{M}\left(\hat{\theta}^{(\ell)}-\hat{\theta}\right)\left(\hat{\theta}^{(\ell)}-\hat{\theta}\right)^{T}$

step7：検定統計量tを算出．近似的に自由度νのt分布を仮定

$$t = \frac{\hat{\theta}}{\sqrt{\hat{V}}},\quad \nu = (M-1)\left\{1+\frac{\widehat{W}}{\left(1+\frac{1}{M}\right)\hat{B}}\right\}$$

群逐次デザイン:α消費関数を用いた有意水準

(Jennison & Turnbull, 2000)

- O'Brien-Fleming型(OF)

$$f_{OF}(u_k) = \left(4 - 4\Phi\left(\frac{Z_{1-\alpha/4}}{\sqrt{u_k}}\right)\right) \quad ,(0 < u_k < 1)$$

- Pocock型(PC)

$$f_{PC}(u_k) = \alpha\log(1 + (e-1)u_k) \quad ,(0 < u_k < 1)$$

- 検定統計量間の相関

$$\sqrt{u_k} = \sqrt{\frac{n_k}{n_K}} \quad ,(k = 1, \ldots, K)$$

α :有意水準
u_k:目標情報量に対する第k解析時点の情報量の割合
n_k:第k番目の中間解析時点における症例数
n_K:最終解析時点における症例数

補完方法：従来法(再掲)

- 中間解析と最終解析それぞれでMIを行う
  - 中間解析と最終解析の同一症例の補完値が異なる場合がある

中間解析

id	時点1	時点2	・・・	時点k
1	○	○	・・・	. → x
2	○	○	・・・	○
3	○	○	・・・	○
4	○	○	・・・	. → y
5	○	○	・・・	○
:	:	:	:	:
n	○	.	・・・	. → z

補完値

n例の時点kに
おける欠測を補完

最終解析

id	時点1	時点2	・・・	時点k
1	○	○	・・・	. → a
2	○	○	・・・	○
3	○	○	・・・	○
4	○	○	・・・	→ b
5	○	○	・・・	○
:	:	:	:	:
n	○	.	・・・	. → c
n+1	○	○	・・・	○
n+2	○	○	・・・	○
n+3	○	○	・・・	○
n+4	○	○	・・・	○
n+5	○	○	・・・	.
:	:	:	:	:
$2n$	○	.	・・・	.

補完値

$2n$例の時点kに
おける欠測を補完

- 仮定された相関の上方バイアス

  仮定された相関 ≥ 真の相関

補完方法：提案法

- 中間解析における補完データ＋中間解析以降に得られる症例
  - 中間解析と最終解析の同一症例の補完値が同一となる

中間解析

id	時点1	時点2	…	時点k	
1	○	○	…	.	→ x
2	○	○	…	○	
3	○	○	…	○	
4	○	○	…	.	→ y
5	○	○	…	○	
:	:	:	:	:	
n	○	.	…	.	→ z

補完値

n例の時点kに
おける欠測を補完

最終解析

id	時点1	時点2	…	時点k	
1	○	○	…	.	→ x
2	○	○	…	○	
3	○	○	…	○	
4	○	○	…	.	→ y
5	○	○	…	○	
:	:	:	:	:	
n	○	.	…	.	→ z

補完値

＋

n+1	○	○	…	○
n+2	○	○	…	○
n+3	○	○	…	○
n+4	○	○	…	○
n+5	○	○	…	.
:	:	:	:	:
2n	○	.	…	.

中間解析以降のn例の
時点kにおける欠測を補完

- 仮定された相関の上方バイアスを減少

群逐次デザイン：検定サイズの計算

- 検定サイズの計算

$$SIZE = 2\{1 - \Phi_2(c_1, c_2; \rho_k)\}$$

c_1：中間解析の棄却限界値
c_2：最終解析の棄却限界値
ρ_k：検定統計量間の真の相関

- 棄却限界値c_1，c_2は仮定された相関$\sqrt{u_k}$を基に算出

- 真の相関が仮定された相関と等しい場合，検定サイズは名義の水準を維持

$$\rho_k = \sqrt{u_k}$$

- 真の相関が仮定された相関より小さくなると検定サイズは増大

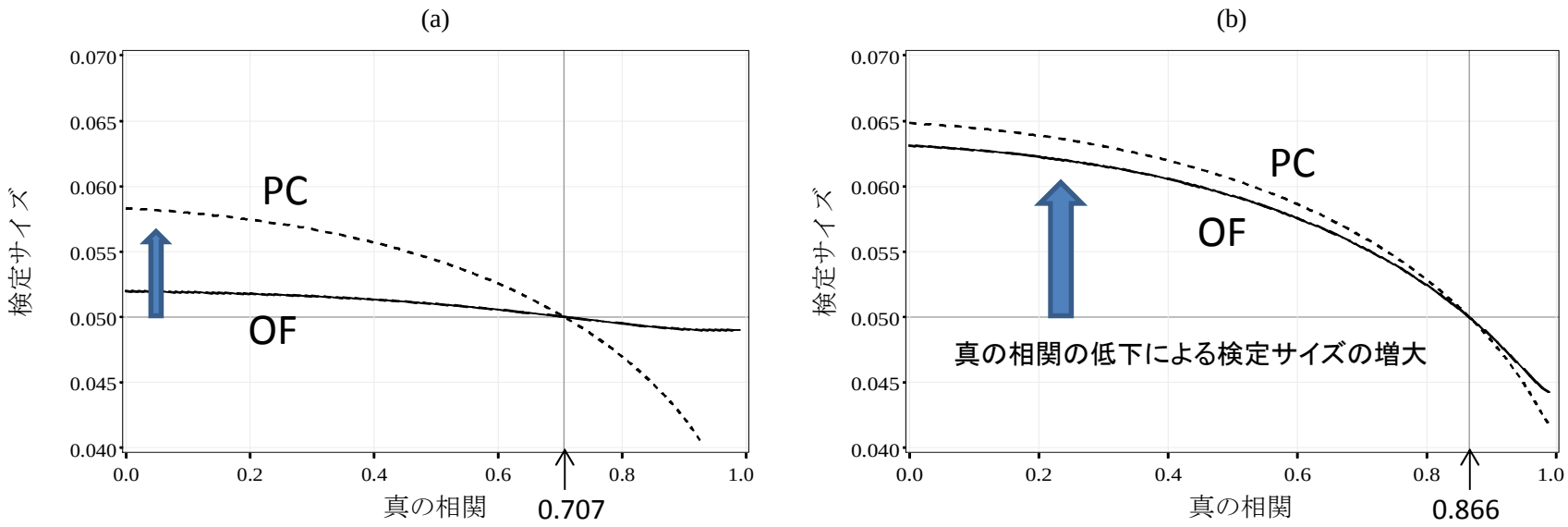
$$\rho_k \leq \sqrt{u_k}$$

群逐次デザイン：相関と検定サイズの関係

中間解析1回の検定サイズの挙動

(a)：中間解析時点の症例数の集積割合50%⇒仮定される相関$\sqrt{1/2} = 0.707$

(b)：中間解析時点の症例数の集積割合75%⇒仮定される相関$\sqrt{3/4} = 0.866$



※検定サイズの低下は理論的な関係を示しているだけであり，実際には生じない

シミュレーション：目的

- 群逐次デザイン臨床試験においてMIを実装する場合の，従来法と提案法の検定サイズへの影響の評価を行う

- 状況設定
  - 正規応答を主要変数とし，治療の有効性を中間解析を含むランダム化並行群間試験に基づき比較
    - 治療群：$g = 2$
    - 測定時点：$t = 3$

シミュレーション：データ生成

- 多変量正規分布に従う乱数を発生
  - 測定時点：$t = 1,2,3$
  - 各時点の平均$(\mu_1, \mu_2, \mu_3) = (0, 0.5, 1)$
    分散$(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1.5, 2)$
  - 相関構造：一次自己回帰構造(AR(1))
    複合対称性構造(CS)
  - 時点間の相関：0.4，0.8

シミュレーション：欠測生成方法

- ロジスティックモデルに基づく欠測確率

$$logit\left(\Pr\left(r_{i,t}=1\right)\right)=Int+y_{i,t-1} \quad ,t=1,2,3$$

- $r_{i,t}$：第i被験者のt時点目の欠測の有無(1：欠測あり，0：欠測なし)
- 1時点目は欠測なし
- 単調な欠測を仮定：欠測が初めて起きた時点以降はすべて欠測
- $y_{i,t-1}$：第i被験者の$t-1$時点目の測定値
  - 1時点前のデータが大きいと欠測しやすい
- Int：最終解析時点の欠測確率が20%，40%になるように設定

シミュレーション：その他のシミュレーション条件

- 最終解析時点の各群の症例数：50，100
- 中間解析の回数：1回
  - 集積割合50%(25/50，50/100)，75%(75/100)
- 補完回数：10，100
- 有意水準:OF，PC
- シミュレーション回数：100000
- MIプロシジャ，MIANALYZEプロシジャによる補完及び推定
- 最終解析の補完方法
  - 従来法，提案法

- 評価
  - 最終解析の補完方法ごとの検定サイズの比較
  - 中間解析と最終解析の検定統計量間の真の相関と仮定された相関の比較

シミュレーション：シミュレーション条件要約

シミュレーションの制御変数

パラメータ	値
相関構造：$type$	AR(1), CS
中間解析時点の症例数の集積割合：π	50%, 75%
最終測定時点の欠測確率：p_3	20%, 40%
補完回数：M	10, 100
最終解析時点の各群の症例数：n	50, 100
時点間の相関：ρ	0.4, 0.8

中間解析時点の集積割合毎の有意水準

中間解析集積割合	α消費関数	中間解析有意水準	最終解析有意水準
50%	OF	0.0031	0.0490
50%	PC	0.0310	0.0277
75%	OF	0.0193	0.0442
75%	PC	0.0414	0.0239

シミュレーション条件とIntの値

ρ	p_3	Int
0.8	40%	-1.69
0.8	20%	-1.74
0.4	40%	-2.80
0.4	20%	-2.83

検定サイズ（AR(1)，中間解析：50%）

最小の相関

5%以上

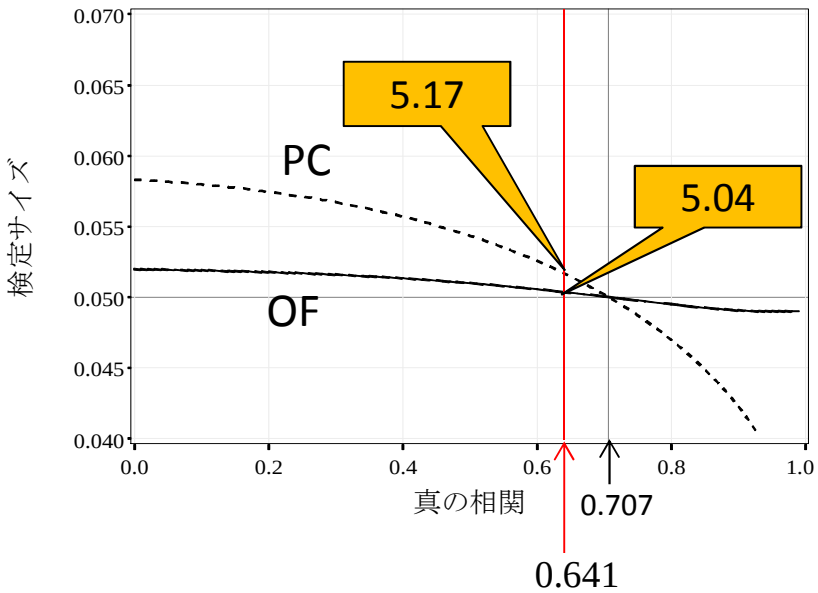
シミュレートされた真の相関の推定値

				検定統計量間の相関(仮定：0.707)		OFの検定サイズ		PCの検定サイズ	
p_3	M	n	ρ	従来	提案	従来	提案	従来	提案
40	10	50	0.8	0.661	0.701	4.72	3.67	4.57	3.86
			0.4	0.641	0.697	4.60	3.29	4.30	3.51
		100	0.8	0.674	0.703	4.95	4.46	4.93	4.56
			0.4	0.655	0.700	4.86	4.20	4.82	4.33
	100	50	0.8	0.688	0.704	4.62	3.48	4.24	3.58
			0.4	0.679	0.701	4.31	2.96	3.89	3.08
		100	0.8	0.697	0.705	4.87	4.42	4.71	4.49
			0.4	0.692	0.703	4.72	4.08	4.5	4.09
20	10	50	0.8	0.693	0.701	5.16	4.81	5.17	4.95
			0.4	0.682	0.702	4.96	4.37	4.94	4.57
		100	0.8	0.694	0.705	5.01	4.87	5.19	5.04
			0.4	0.683	0.703	5.09	4.78	5.04	4.85
	100	50	0.8	0.703	0.702	5.15	4.80	5.10	4.91
			0.4	0.700	0.703	4.99	4.33	4.78	4.40
		100	0.8	0.703	0.705	5.05	4.86	5.12	5.01
			0.4	0.699	0.703	5.02	4.69	4.94	4.74

仮定された相関の上方バイアスに伴う検定サイズへの影響
AR(1), 中間解析:50%

p_3	M	n	ρ	検定統計量間の相関(仮定:0.707)	OFの検定サイズ	PCの検定サイズ
40	10	50	0.4	0.641	4.60	4.30

中間解析の集積割合50%



シミュレートされた真の相関の低下に伴い，検定サイズが増大するはずが，減少する場合が存在

他の因子による影響の可能性

GLMSELECTプロシジャに基づくモデル選択

- 検定サイズへの影響因子の探索
- データ
  - 従来法のシミュレーション結果(OF, PCそれぞれで検討)
- 応答変数
  - シミュレートされた検定サイズ
- 説明変数
  - シミュレーションの制御変数
    - 欠測確率: p_3, 補完回数: M, 最終解析時点の症例数: n, 時点間の相関: ρ, 相関構造: $type$, 中間解析時点の症例数の集積割合: π
  - シミュレーション結果から得られた検定サイズと相関関係が存在し得る変数
    - 相関比
      - {検定統計量間の仮定された相関}/{シミュレートされた真の相関}
        - 相関比が増大するほど, 真の相関が減少
    - 自由度:中間, 自由度:最終
      - 中間解析, 最終解析の推定された自由度の中央値の対数
    - mean(t):中間, mean(t):最終, SD(t):中間, SD(t):最終
      - 生成された1万個の中間解析, 最終解析の検定統計量の平均とSD
- ステップワイズ法(組み入れ, 維持基準共にp=0.15)に基づくモデル選択

GLMSELECTプロシジャに基づくモデル選択

- 使用データ(一部抜粋)

相関構造	集積割合	欠測確率	補完回数	症例数	時点間の相関	OF	PC	…
AR	50	40	10	50	0.8	4.72	4.57	…
AR	50	40	10	50	0.4	4.60	4.30	…
AR	50	40	10	100	0.8	4.95	4.93	…
AR	50	40	10	100	0.4	4.86	4.82	…
AR	50	40	100	50	0.8	4.62	4.24	…
AR	50	40	100	50	0.4	4.31	3.89	…
AR	50	40	100	100	0.8	4.87	4.71	…
AR	50	40	100	100	0.4	4.72	4.5	…
AR	50	20	10	50	0.8	5.16	5.17	…
AR	50	20	10	50	0.4	4.96	4.94	…
AR	50	20	10	100	0.8	5.01	5.19	…
AR	50	20	10	100	0.4	5.09	5.04	…
:	:	:	:	:	:	:	:	⋱

OF, PCの検定サイズを応答変数としてモデル選択

検定サイズへの影響因子の探索

OFを応答変数としたモデル選択結果

パラメータ	t値
p_3: 20 ↓	2.59
n:50 ↓	4.21
相関比 ↓	-1.76
SD(t):中間 ↑	16.13

PCを応答変数としたモデル選択結果

パラメータ	t値
p_3: 20 ↓	6.92
n:50 ↓	4.55
ρ: 0.4 ↑	-2.75
SD(t):中間 ↑	30.93

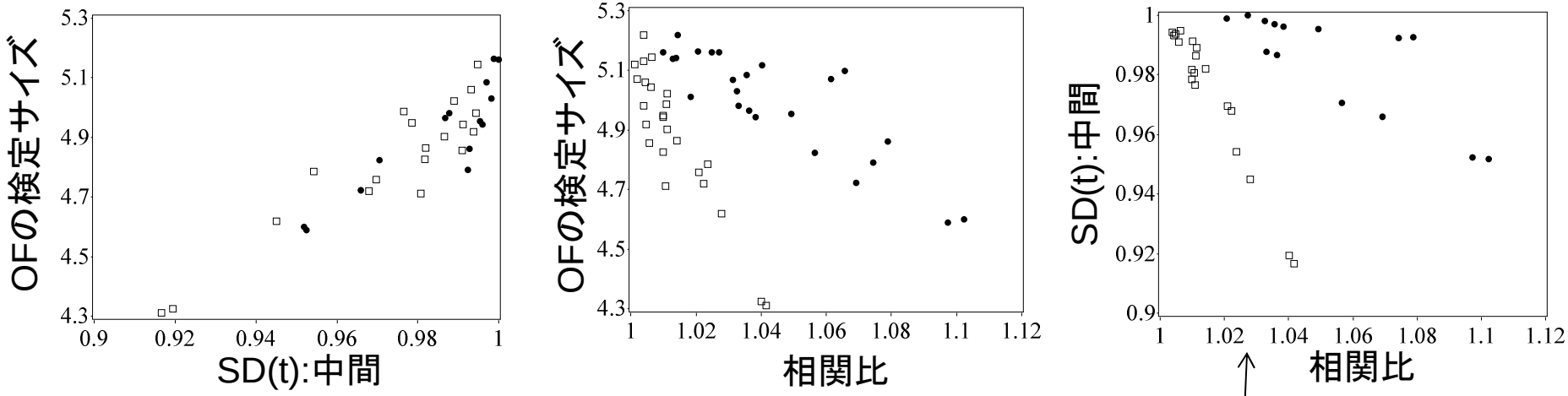
※↑, ↓:パラメータの矢印の方向への増減に伴い,
検定サイズの増加を意味する

- 検定サイズに影響する因子
  - OF, PC共にp_3, n, SD(t):中間の3因子が関係
  - 特にSD(t):中間の影響が大きい

  ⇒SEの過大推定に起因

- OFの相関比のt値が負⇒本来ならば相関比の増大に伴い検定サイズの増大
  - SD(t):中間と相関比の変動要因が共通であるため

OFの検定サイズ，SD(t):中間，相関比の関係



(●: $M = 10$, □: $M = 100$)

相関比とSD(t):中間に逆相関が存在

結論

- 中間解析を含む群逐次デザインのランダム化並行群間試験における従来法，提案法の検定サイズへの影響を評価
- 検定統計量間の仮定された相関の上方バイアスによる検定サイズへの影響はそれほど大きくなかった
  - 従来法：名義の水準より若干増加する場合あり
  - 提案法：名義の水準より小さくなる
    - ただし，推定対称パラメータが増えることで過度に保守的になる恐れ
- 検定サイズへ影響する他の因子
  - 欠測確率，症例数，中間解析の検定統計量のSD
    - SEの過大推定に起因
- MIANALYZEプロシジャにより算出される自由度の過大推定
  - Barnard & Rubin(1999)に基づく自由度の補正
- 今後の課題
  - 状況設定のさらなる一般化
    - 中間解析回数，測定時点数

参考文献

- Barnard, J. and Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, 86, 948-955.
- Charpenter, J. and Kenward, M. (2013). *Multiple Imputation and its Application*. Wiley.
- Jennison, C. and Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials* [森川敏彦・山中竹春 訳(2012). 臨床試験における群逐次法：理論と応用, CAC].
- National Research Council. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education(2010). *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press, Washington, DC.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.
- Tanaka, S. Fukinbara, S. Tsuchiya, S. Suganami, H. and Ito, Y. M. (2014). Current Practice in Japan for the Prevention and Treatment of Missing Data in Confirmatory Clinical Trials:A Survey of Japanese and Foreign Pharmaceutical Manufacturers. *Therapeutic Innovation & Regulatory Science*, 48(6), 717-723.

ご清聴ありがとうございました