

**【日本製薬工業協会シンポジウム】**  
**ランダム化比較試験における統計的効率改善のための共変量調整**

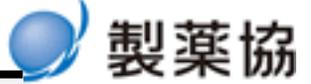
**[シンポジウム当日]**  
**LP3: 線形モデルの誤特定下での統計的性能**



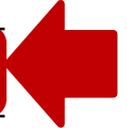
**2026年3月16日**  
**医薬品評価委員会 データサイエンス部会**  
**2025年度タスクフォース3-1**

**○武田薬品工業株式会社**      **宋小林**

# シンポジウム当日の製薬協TFの発表一覧と 本発表の位置づけ



	製薬協TFの発表	-
LP0	はじめに	関
LP1	FDAガイダンス/JPMA報告書の復習	澤本
LP2	共分散分析による平均治療効果の推測 (4.3.3節)	平井, 土居
LP3	線形モデルの誤特定下での統計的性能 (4.3.4節)	宋
LP4	非線形モデルにおける条件付き治療効果と条件なし治療効果 (4.3.5節)	大野



- **線形モデル誤特定下での条件なし治療効果の推定量の妥当性**
  - 単純ランダム化 ((1)Lin(2013))
  - 共変量アダプティブランダム化
    - ANHECOVA ((2)Ye et al. (2023))
    - 層別因子を考慮する手法 ((3)Ye et al. (2022))
- **線形モデルが誤特定されている場合の標準誤差**
  - サンドイッチ推定量 ((1)Lin(2013))
  - ANHECOVAに基づいた標準誤差の推定量 ((2)Ye et al. (2023))

- **線形回帰モデルの誤特定下での推定に関してFDAガイダンスでは以下の点を言及**
  - **線形回帰モデルが誤特定されていても、線形モデルによる共変量調整は平均治療効果の推定と推論に妥当な方法である (Lin, 2013) (B3)**
  - **モデル誤特定下でも妥当な、標準誤差の算出の推奨. Huber-White「サンドウィッチ」標準誤差のようなロバスト標準誤差の使用を推奨 (Rosenblum and van der Laan 2009; Lin 2013) .  
文献等で提案されているその他のロバスト標準誤差の推定方法は、交互作用を含む場合も適用可能である (Ye et al. 2023) (B4)**
  - **標準誤差の計算に層別ランダム化を考慮すること. 層別化と共変量調整及びモデル誤特定の可能性を組み合わせたときの標準誤差の計算方法はいくつかある (Bugni et al. 2018; Ye et al. 2022) (B5)**

## ➤ 引用文献の一部

- (1) Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics*. 7(1), 295-318
- (2) Ye, T., Shao, J., Yi, Y., & Zhao, Q. (2023). Toward better practice of covariate adjustment in analyzing randomized clinical trials. *Journal of the American Statistical Association*, 118(544), 2370-2382.
  - FDAガイダンスでは, Ye et al. (2022)
- (3) Ye, T., Yi, Y., & Shao, J. (2022). Inference on the average treatment effect under minimization and other covariate-adaptive randomization methods. *Biometrika*, 109(1), 33-47.
  - FDAガイダンスでは, Ye et al. (2021)

## ➤ 記号

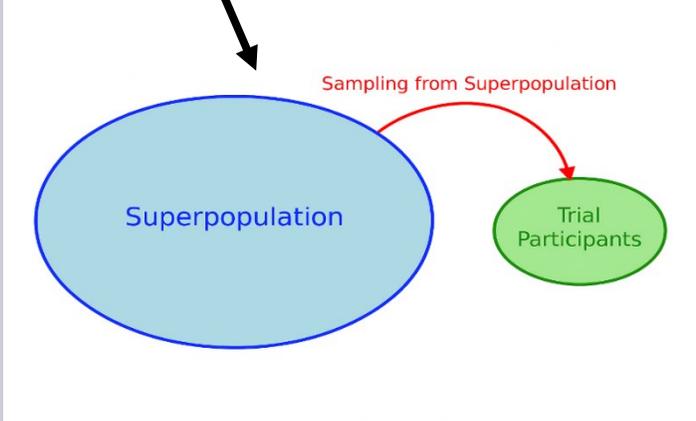
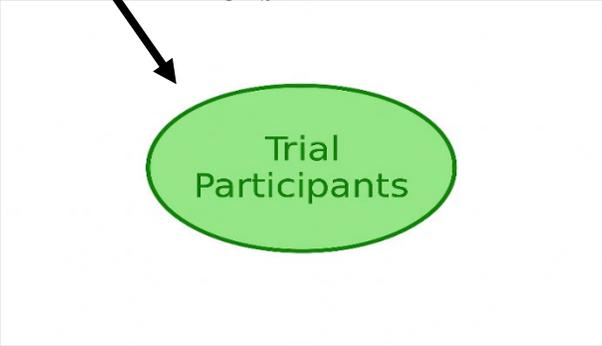
- $T$  : 治療変数(0: 対照治療群, 1: 試験治療群)
- $Y^{(t)}$  : 治療 $t = 0, 1$ に対する潜在結果変数
- $X$  : model-assistedな方法で調整に用いるすべての観察された共変量
- $Z$  : 共変量アダプティブランダム化に利用する層別因子

➤ 対象者 $i = 1, \dots, n$ のデータ

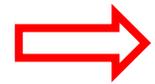
- $X_i$ を観察した後に治療 $T_i$ が割り付けられる
- $T_i = t$ が観察された場合に,  $Y_i = Y_i^{(t)}$ が観察される
- $\bar{X}, \bar{Y}$ はそれぞれ共変量の標本平均と結果変数の標本平均
- $\bar{Y}_t (t = 0, 1)$ は治療 $t$ に割り付けられた被験者の結果変数の標本平均
- $\bar{X}_t (t = 0, 1)$ は治療 $t$ に割り付けられた被験者の共変量の標本平均
- $\bar{Y}_t(z) (t = 0, 1)$ は層 $z$ に属して治療 $t$ に割り付けられた被験者の結果変数の標本平均
- $\bar{X}_t(z) (t = 0, 1)$ は層 $z$ に属して治療 $t$ に割り付けられた被験者の共変量の標本平均
- すべての治療割付が完了し, データを収集後, 観察データ $(Y_i, X_i, T_i)$ に基づき推論を行う

➤ 平均治療効果 :  $ATE = E[Y^{(1)} - Y^{(0)}]$

➤ **標本**平均治療効果 :  $ATE_n = \frac{1}{n} \sum_{i=1}^n (Y_i^{(1)} - Y_i^{(0)})$

	Ye et al. (2023)	Lin (2013)
Estimand: 条件なし治療効果	<p>平均治療効果</p> $ATE = E[Y_i^{(1)} - Y_i^{(0)}]$ 	<p><b>標本</b>平均治療効果</p> $ATE_n = \frac{1}{n} \sum_{i=1}^n (Y_i^{(1)} - Y_i^{(0)})$ 
対象集団	<i>superpopulation</i>	<i>finite population</i>

➤ **線形モデル誤特定下での条件なし治療効果の推定量の妥当性**



■ 単純ランダム化 ((1)Lin(2013))

■ 共変量アダプティブランダム化

□ ANHECOVA ((2)Ye et al. (2023))

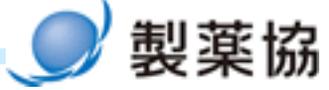
□ 層別因子を考慮する手法 ((3)Ye et al. (2022))

➤ **線形モデルが誤特定されている場合の標準誤差**

■ サンドイッチ推定量 ((1)Lin(2013))

■ ANHECOVAに基づいた標準誤差の推定量 ((2)Ye et al. (2023))

# 標本平均治療効果の推定( $ATE_n$ , (1)Lin(2013))



- 背景: **有限母集団**(finite population)におけるランダム化推論 (Neymanモデル) に基づき, **単純ランダム化**の下で, 条件なし標本平均治療効果を推定するための異なる推定量 (未調整, 回帰調整済み) の漸近的な性質を比較・評価
- 2群の場合, 下記のようなモデルを考える

	モデル	治療効果の推定量
治療のみ	$Y_i = \alpha + \beta T_i$	$\hat{\beta}_{unadj}$
治療+共変量 (交互作用なし)	$Y_i = \alpha + \beta T_i + \eta X_i$	$\hat{\beta}_{adj}$
治療+共変量 (交互作用あり)	$Y_i = \alpha + \beta T_i + \eta X_i + \psi T_i(X_i - \bar{X})$	$\hat{\beta}_{interact}$

- $\hat{\beta}_{unadj}, \hat{\beta}_{adj}, \hat{\beta}_{interact}$  は各モデルの最小二乗推定量(OLS)

【当日 3】誤特定下での線形モデルの性能

# 標本平均治療効果の推定 ( $ATE_n$ , (1)Lin(2013))



- 適当な正則条件の下では
  - 上記の三つの治療効果推定量は真の標本平均治療効果の一致推定量
  - 線形モデルが誤特定されていたとしても, 漸近正規性を有する
- 推定の精度に関して
  - $\hat{\beta}_{unadj}, \hat{\beta}_{adj}$  と比べて,  $\hat{\beta}_{interact}$  は推定精度を損なうことがない
  - 割付比が1:1の場合,  $\hat{\beta}_{interact}$  と  $\hat{\beta}_{adj}$  は同程度の推定精度
- 三群以上でも上記の結論が成り立つ

## ➤ 線形モデル誤特定下での条件なし治療効果の推定量の妥当性

- 単純ランダム化 ((1)Lin(2013))

- 共変量アダプティブランダム化

➔  ANHECOVA ((2)Ye et al. (2023))

- 層別因子を考慮する手法 ((3)Ye et al. (2022))

## ➤ 線形モデルが誤特定されている場合の標準誤差

- サンドイッチ推定量 ((1)Lin(2013))

- ANHECOVAに基づいた標準誤差の推定量 ((2)Ye et al. (2023))

# 平均治療効果の推定((2)Ye et al. (2023) )



- **超母集団 (superpopulation) を対象とする, 単純ランダム化 および共変量アダプティブランダム化 (例: 層別置換ブロック法, 最小化法など) の下で, 平均治療効果 (ATE) の異なる推定量 (未調整, 回帰調整済み) の漸近的な性質を比較・評価**
- **2群の場合, 下記のようなモデル(作業モデル)と推定量を提案**

	モデル	治療効果の推定量
ANOVA	$E(Y_i T_i) = \beta_1 I(T_i = 1) + \beta_0 I(T_i = 0)$	$\bar{Y}_1 - \bar{Y}_0$
ANCOVA	$E(Y_i X_i, T_i = 0) = \beta_0 + \theta^T(X_i - \bar{X})$ $E(Y_i X_i, T_i = 1) = \beta_1 + \theta^T(X_i - \bar{X})$	$\bar{Y}_1 - \hat{\theta}^T(\bar{X}_1 - \bar{X}) - [\bar{Y}_0 - \hat{\theta}^T(\bar{X}_0 - \bar{X})]$
ANHECOVA*	$E(Y_i X_i, T_i = 0) = \beta_0 + \theta_0^T(X_i - \bar{X})$ $E(Y_i X_i, T_i = 1) = \beta_1 + \theta_1^T(X_i - \bar{X})$	$\bar{Y}_1 - \hat{\theta}_1^T(\bar{X}_1 - \bar{X}) - [\bar{Y}_0 - \hat{\theta}_0^T(\bar{X}_0 - \bar{X})]$

\* ANHECOVA はANalysis of HEterogeneous COVarianceの略称

## 平均治療効果の推定((2)Ye et al. (2023) )



- 適当な正則条件の下では
  - 上記の三つの治療効果推定量は平均治療効果の一致推定量
- 推定の精度に関して
  - 効率性について, ANOVAモデルやANCOVAモデルに比べて, ANHECOVAモデルは効率の向上が保証される
  - 三群以上でも上記の結論が成り立つ
- 正則条件の下で, ANHECOVAは以下の性質を持つ
  - 不変性: 漸近分布はランダム化手法によらない
  - 最適性: 所定の推定量のクラスで, 漸近分散は最小

# (1)Lin (2013)と(2)Ye et al. (2023) の論点整理



## ➤ 共通点

- モデルが誤特定されていたとしても, 漸近正規性と一致性を有する
- 交互作用による推定効率の改善
- (1)Lin (2013)と(2)Ye et al. (2023) で提案されるATEの推定量の値は等しい

	治療効果の推定量( $ATE_n$ )		治療効果の推定量( $ATE$ )
治療のみ	$\hat{\beta}_{unadj}$	ANOVA	$\bar{Y}_1 - \bar{Y}_0$
治療+共変量 (交互作用なし)	$\hat{\beta}_{adj}$	ANCOVA	$\bar{Y}_1 - \hat{\theta}^T(\bar{X}_1 - \bar{X}) - [\bar{Y}_0 - \hat{\theta}^T(\bar{X}_0 - \bar{X})]$
治療+共変量 (交互作用あり)	$\hat{\beta}_{interact}$	ANHECOVA	$\bar{Y}_1 - \hat{\theta}_1^T(\bar{X}_1 - \bar{X}) - [\bar{Y}_0 - \hat{\theta}_0^T(\bar{X}_0 - \bar{X})]$

## ➤ 相違点

- Estimandの違い(具体的にはestimandの対象集団は異なる)
- 論文で議論したランダム化手法の違い

## ➤ 線形モデル誤特定下での条件なし治療効果の推定量の妥当性

- 単純ランダム化 ((1)Lin(2013))
- 共変量アダプティブランダム化
  - ANHECOVA ((2)Ye et al. (2023))

➡ □ 層別因子を考慮する手法 ((3)Ye et al. (2022))

## ➤ 線形モデルが誤特定されている場合の標準誤差

- サンドイッチ推定量 ((1)Lin(2013))
- ANHECOVAに基づいた標準誤差の推定量 ((2)Ye et al. (2023))

# 平均治療効果の推定((3)Ye et al. (2022))



- **超母集団 (superpopulation) を対象**とする, 解析時に被験者の背景や治療効果に影響を及ぼす可能性のある**特定の因子 (層別因子) を考慮し**, 共変量アダプティブランダム化 (例: 層別ブロック法, 最小化法 など) の下で, 提案する平均治療効果 (ATE) 推定量の漸近的な性質を比較・評価
- **Type\_A : 治療群別・層別**に共変量調整した作業モデルを作り, 最小二乗推定量で共変量調整ATEを推定. 層によって治療効果が異なる状況にも対応可
- **Type\_B : 層別**に共変量調整した作業モデルを作り, 最小二乗推定量で共変量調整ATEを推定. 全治療群で共変量の影響が同一と仮定できる場合に適する.

	作業モデル	治療効果の推定量
Type_B	$E(Y_i X_i, Z_i = z, T_i = 0) = \beta_0(z) + \theta(z)^T (X_i - \bar{X})$ $E(Y_i X_i, Z_i = z, T_i = 1) = \beta_1(z) + \theta(z)^T (X_i - \bar{X})$	$\sum_{z \in Z} \frac{n(z)}{n} \{ \bar{Y}_1(z) - \hat{\theta}(z)^T [\bar{X}_1(z) - \bar{X}(z)] - \{ \bar{Y}_0(z) - \hat{\theta}(z)^T [\bar{X}_0(z) - \bar{X}(z)] \} \}$
Type_A	$E(Y_i X_i, Z_i = z, T_i = 0) = \beta_0(z) + \theta_0(z)^T (X_i - \bar{X})$ $E(Y_i X_i, Z_i = z, T_i = 1) = \beta_1(z) + \theta_1(z)^T (X_i - \bar{X})$	$\sum_{z \in Z} \frac{n(z)}{n} \{ \bar{Y}_1(z) - \hat{\theta}_1(z)^T [\bar{X}_1(z) - \bar{X}(z)] - \{ \bar{Y}_0(z) - \theta_0(z)^T [\bar{X}_0(z) - \bar{X}(z)] \} \}$

## 平均治療効果の推定((3)Ye et al. (2022))



- **2群以上かつ共変量アダプティブランダム化の場合**  
(層別置換ブロック法, 層別バイアスコイン法, 層別壺デザイン, 最小化法)
- 平均治療効果の推定量(Type\_A, Type\_B)は漸近正規性を有し, model-assistedな方法であるためモデル誤特定下でも漸近的に妥当であることが示されている
  - **不変性**:平均治療効果の推定量(Type\_A, Type\_B)の漸近分布はランダム化の方法に依らず不変
- ほとんどの場合, Type\_AはType\_Bより効率がよい
- 単純ランダム化にも適用できる

## ➤ 線形モデル誤特定下での条件なし治療効果の推定量の妥当性

- 単純ランダム化 ((1)Lin(2013))
- 共変量アダプティブランダム化
  - ANHECOVA ((2)Ye et al. (2023))
  - 層別因子を考慮する手法 ((3)Ye et al. (2022))

## ⇒ ➤ 線形モデルが誤特定されている場合の標準誤差

- サンドイッチ推定量 ((1)Lin(2013))
- ANHECOVAに基づいた標準誤差の推定量 ((2)Ye et al. (2023))

# サンドイッチ推定量



## ➤ 名目標準誤差の問題点

- 通常の統計ソフトで出力される標準誤差（名目標準誤差）は、モデルが誤特定されていると**真の値を過大または過小評価するリスク**がある
  - 割付比が1:1の場合、モデルが誤特定されていたとしても、この名目標準誤差は真の漸近標準誤差と同じあるいは保守的になることが知られている

## ➤ 解決策：サンドイッチ推定量 ((1)Lin, 2013)

- Huberのサンドイッチ推定量を用いたロバスト分散の推定方法を提案

$$(X^T X)^{-1} X^T \text{diag}(\hat{\epsilon}_1^2, \dots, \hat{\epsilon}_n^2) X (X^T X)^{-1}$$

$X$ は説明変数の行列,  $\hat{\epsilon}_i$  は $i$ 番目の被験者の残差

- モデル誤特定に対してロバスト
- 治療と共変量の交互作用を含むモデルから構成されるサンドイッチ分散推定量は、共変量調整あり（治療と共変量の交互作用なし）のモデルから構成されるサンドイッチ分散推定量と等しいか、より小さくなる

- モデルが誤特定されていても信頼できる標準誤差の推定方法を提案(層別置換ブロック法, 層別バイアスコイン法)
  - 提案した推定量は対応する**真の漸近分散の一致推定量**
  - ロバスト性: モデルの誤特定や, 治療群ごとのデータのばらつき (不等分散) が存在しても, **ランダム化の方法に影響されずに正しい漸近分散を与える**
- この推定量の構成
  - 群ごとの残差分散: 各治療群( $t$ )における残差( $Y_i - \hat{\beta}_t X_i$ )の標本分散 $S_t^2(\hat{\beta}_t)$ を用いて, 群ごとの不等分散を反映する
  - 共変量の標本共分散行列: 共変量の中心化により導入される追加の不確実性を補正するため, 全体の共変量に基づく共分散行列 $\hat{\Sigma}_x$ を利用する

- このとき、提案されたロバストな分散推定量は以下のように表される

$$\hat{V} = \text{diag}\{\pi_t^{-1}S_t^2(\hat{\beta}_t)\} + B^T\hat{\Sigma}_X B$$

ここで、 $\pi_t$  は治療割付割合、 $\hat{\beta}_t$  は群 $t$ の回帰係数ベクトルの推定量、 $\hat{\Sigma}_X$  全被験者の $X_i$ の標本分散、 $B$ は各群の回帰係数ベクトルの推定量を列として並べた行列である

- **群 $s$ と群 $t$ の群間差の標準誤差推定量**は下記のもものが提案された

$$SE_{ts} = \sqrt{\pi_t^{-1}S_t^2(\hat{\beta}_t) + \pi_s^{-1}S_s^2(\hat{\beta}_s) + (\hat{\beta}_t - \hat{\beta}_s)^T \hat{\Sigma}_X (\hat{\beta}_t - \hat{\beta}_s)}$$

- **線形モデル誤特定下での条件なし治療効果の推定量の妥当性**
  - 単純ランダム化 ((1)Lin(2013))
  - 共変量アダプティブランダム化
    - ANHECOVA ((2)Ye et al. (2023))
    - 層別因子を考慮する手法 ((3)Ye et al. (2022))
- **線形モデルが誤特定されている場合の標準誤差**
  - サンドイッチ推定量 ((1)Lin(2013))
  - ANHECOVAに基づいた標準誤差の推定量 ((2)Ye et al. (2023))

## ➤ 本発表では主に**線形モデルの誤特定下**での平均治療効果の推定に関連する研究を紹介

- 線形回帰モデルが誤特定されていても平均治療効果の妥当な推定値が与えられること. かつ適当な正則条件の下では, 線形モデルによる共変量の調整が**効率改善**が期待できる.
- モデルが誤特定されている場合, 名目標準誤差では問題が生じ得るが, ロバスト標準誤差を用いることで対処できる. さらに, 標準誤差の算出にあたっては, モデル誤特定に加え, 層別ランダム化の影響も考慮する必要がある.

- Tsiatis, A. A., Davidian, M., Zhang, M., & Lu, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in medicine*, 27(23), 4658-4677.
- Yang, L., & Tsiatis, A. A. (2001). Efficiency study of estimators for a treatment effect in a pretest–posttest trial. *The American Statistician*, 55(4), 314-321.