

# 一般化Wilcoxon検定を用いた中間解析 の無効中止に関する研究

○大久保竣祐、浜田知久馬、安藤宗司  
(東京理科大学)

## 要旨:

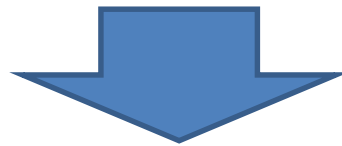
中間解析において一般化Wilcoxon検定を用いた場合の無益性評価の提案をする。評価の検討は、条件付き検出力及び予測検出力を用いてシミュレーションにより行う。

キーワード: 中間解析、条件付き検出力、予測検出力

ポスター展示: 次ページから12枚のスライドがポスター用です

## アダプティブデザインについて<sup>[8]</sup>

- 医薬品を市場に早く提供するには開発のスピードアップが必要  
→新薬開発を効率的に行うことでスピードアップを図る



- アダプティブデザインが注目されている  
→進行中の臨床試験の計画に変更を加える

アダプティブデザインでは中間解析に基づき  
試験の修正ができるため開発を効率的に行うことが可能

## 無効中止について<sup>[8]</sup>

- アダプティブデザインにおける変更例
  - ✓ 被験者の割り付け割合
  - ✓ サンプルサイズの見直し
  - ✓ 試験の有効・無効中止



- 無効中止を試験の早期に確認することで  
人的、財政的な無駄を無くすることができる

本研究ではアダプティブデザインにおける無効中止に着目

## 条件付き検出力<sup>[2]</sup>

- 無益性(無効中止)の評価

Lan-Simon-Halperinによる検定では条件付き検出力に基づき  
試験の無益性を判定する

$$\text{条件付き検出力関数: } P_n(\delta) = \Phi\left(\frac{\sqrt{n}Z_n + (N - n)\delta/\sqrt{2} - \sqrt{N}Z_{1-\alpha}}{\sqrt{N - n}}\right)$$

$$\text{無益性の判断基準: } P_n(\delta) < 1 - \gamma$$

(一般的に $\gamma$ は0.8~1.0)

$N$ : 最終解析時の被験者数(1群)     $n$ : 中間解析時の被験者数(1群)

$Z_n$ : 中間解析時の検定統計量     $Z_{1-\alpha}$ : 標準正規分布の100(1 -  $\alpha$ )パーセント点

$\delta$ : 応答変数の真の治療差

ログランク検定(生存時間データ)の場合

$\delta$ : 対数ハザード比

## 予測検出力 <sup>[2]</sup>

- 予測検出力の考え方

得られたデータから治療効果の事後分布に関して  
条件付き検出力関数を平均化する

→事後分布により試験結果の不確定性に対する能力が向上

$$\text{予測検出力: } P_n = \int P_n(\delta) f(\delta|Z_n) d\delta$$

$$\text{無益性の判断基準: } P_n < 1 - \gamma$$

(一般的に $\gamma$ は0.8~1.0)

$P_n(\delta)$ : 条件付き検出力関数

$f(\delta|Z_n)$ : 観測データからの治療差 $\delta$ の事後密度

$n$ : 中間解析時の被験者数(1群)

$Z_n$ : 中間解析時の検定統計量

$\delta$ : 応答変数の真の治療差

ログランク検定(生存時間データ)の場合  
 $\delta$ : 対数ハザード比

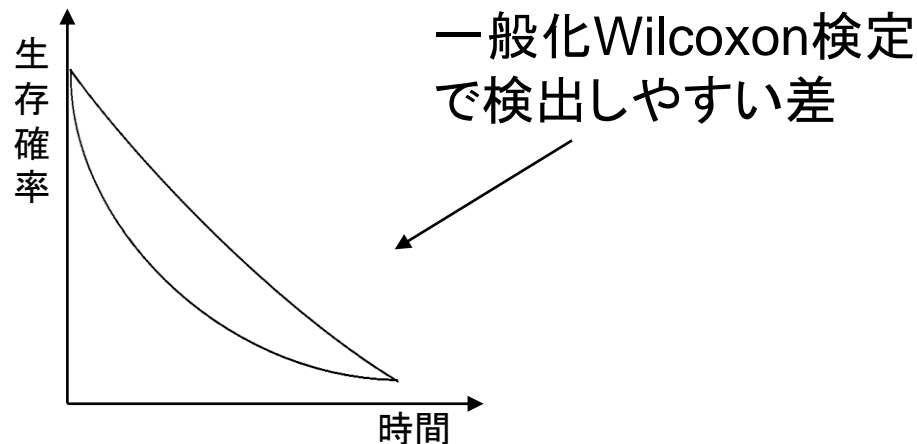
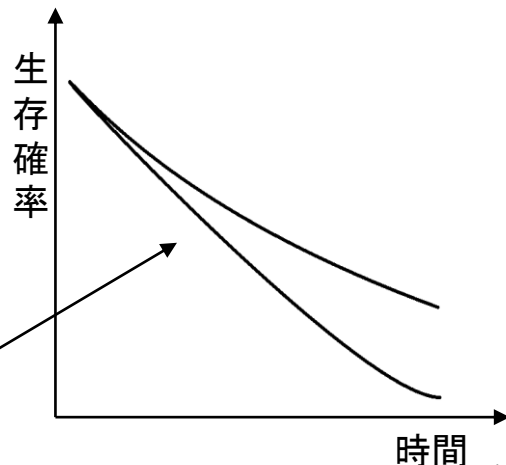
# 研究目的

- 無益性評価の課題

条件付き検出力と予測検出力の定式化に関して  
連続量、二値、ログランク検定における生存時間データ  
は既に定式化されている

一般化Wilcoxon検定を適応する場合の条件付き検出力と予測検出力の算出と性能評価を目的とする

ログランク検定で  
検出しやすい差



## 最尤推定量の下での検出力<sup>[5]</sup>

- 条件付き検出力の $\delta$ , 予測検出力 $f(\delta|Z_k)$ を設定することが困難な場合は最尤推定法を使用する  
(中間解析時までの状況がその後も続くと仮定)

条件付き検出力関数: 
$$P_n(\delta) = \Phi\left(\frac{\sqrt{n}Z_n + (N-n)\delta/\sqrt{2} - \sqrt{N}Z_{1-\alpha}}{\sqrt{N-n}}\right)$$

➡ 
$$P_n(\delta) = \Phi\left(\frac{\frac{N}{\sqrt{n}}Z_n - \sqrt{N}Z_{1-\alpha}}{\sqrt{N-n}}\right)$$

予測検出力: 
$$P_k = \int P_k(\delta) f(\delta|Z_k) d\delta$$

➡ 
$$P_n = \Phi\left(\frac{\sqrt{N}Z_n - \sqrt{n}Z_{1-\alpha}}{\sqrt{N-n}}\right)$$



# 条件付き検出力(一般化Wilcoxon検定)

- 問題点

ログランク検定の式では $\delta$ に対数ハザード比を使うが  
一般化Wilcoxon検定では適用できない

条件付き検出力関数: 
$$P_n(\delta) = \Phi \left( \frac{\sqrt{n}Z_n + (N - n)\delta/\sqrt{2} - \sqrt{N}Z_{1-\alpha}}{\sqrt{N - n}} \right)$$

➡  $\delta$ を症例数設計から算出する

$$Z = \delta\sqrt{N} = Z_{1-\alpha} + Z_\beta \rightarrow \delta = \frac{Z_{1-\alpha} + Z_\beta}{\sqrt{N}}$$

$N$ : 最終解析時の被験者数(1群)

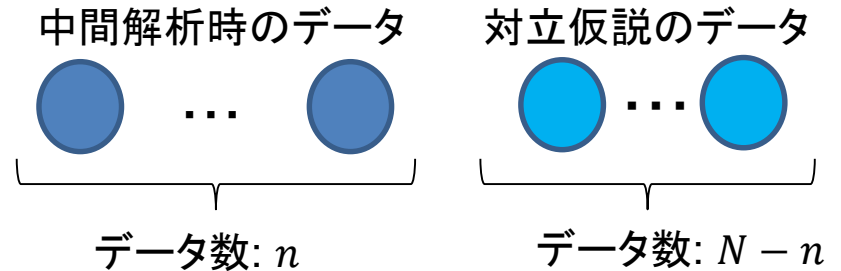
条件付き検出力関数:

$$P_n(\delta) = \Phi \left( \frac{\sqrt{n}Z_n - \sqrt{N}Z_{1-\alpha}}{\sqrt{N - n}} + \sqrt{\frac{N - n}{2N}} (Z_{1-\alpha} + Z_\beta) \right)$$

# セミパラメトリック、ノンパラメトリックな手法

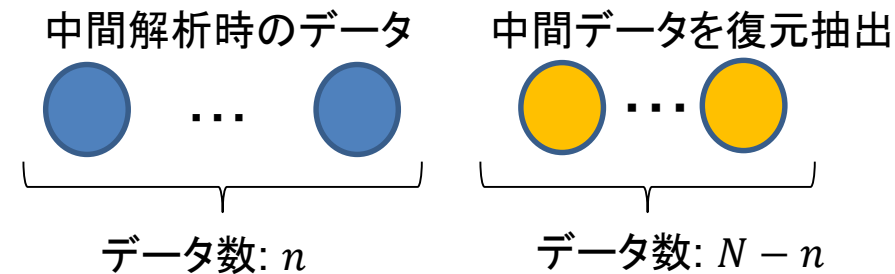
## ① セミパラメトリック条件付き検出力

中間解析時のデータに加えて対立仮説の下で発生させた乱数データをもとに算出



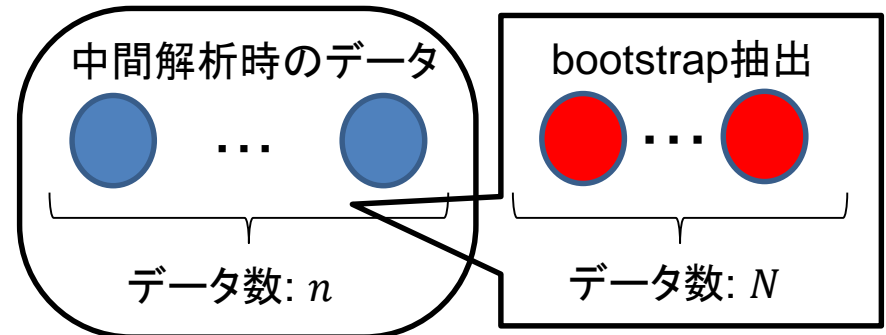
## ② ノンパラメトリック条件付き検出力

中間解析のデータを復元抽出し  
そのデータをもとに算出



## ③ ノンパラメトリック予測検出力

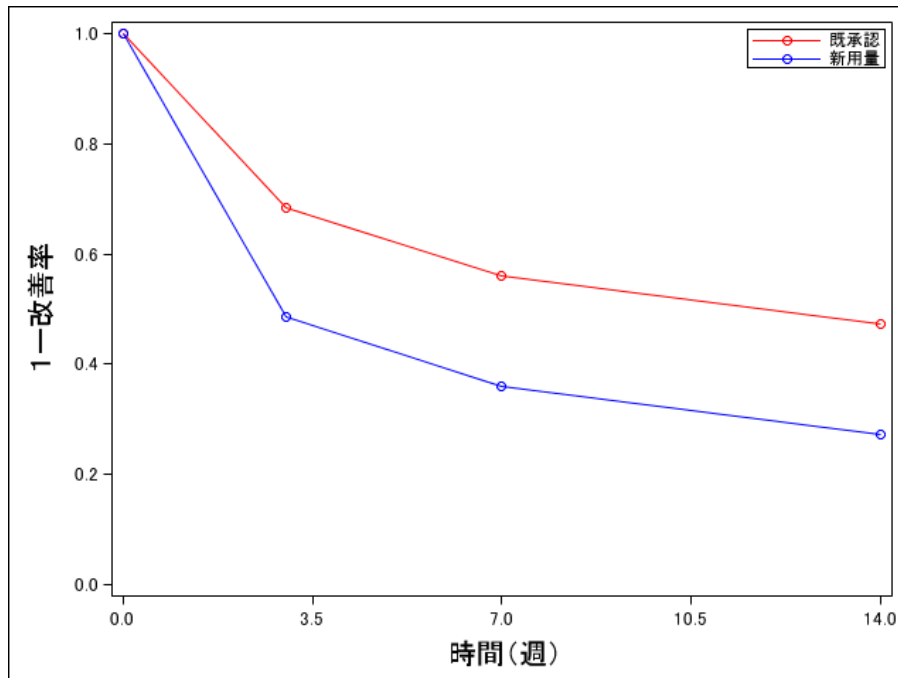
中間解析のデータをbootstrap抽出し  
そのデータをもとに算出



# 想定する臨床試験

## ● 既承認用量と新用量の改善率の縮小

	3日改善	7日改善	14日改善
既承認	31.5%	43.9%	52.6%
新用量	51.5%	63.9%	72.6%



Lakatosの方法<sup>[3]</sup>

生存時間分布: 区分直線分布を想定

帰無仮説:  $\lambda_{1i} = \lambda_{2i}$

必要症例数(1群):  $N = \frac{1}{2} \left( \frac{Z_{\alpha} + Z_{\beta}}{E^*} \right)^2$

$$E^* = \frac{\sum_{i=1}^M d_i \left[ \frac{\phi_i HR_i}{1 + \phi_i HR_i} - \frac{\phi_i}{1 + \phi_i} \right]}{\sqrt{\sum_{i=1}^M d_i \frac{\phi_i}{(1 + \phi_i)^2}}}$$

$\lambda_{ki}$ : 区間*i*の*k*群のハザード( $k = 1, 2$ )

$d_i$ : 区間*i*の2群を併せた期待イベント

数  $\phi_i$ : 区間*i*の必要例数比

$M$ : 分けられた区間数

$HR_i$ : 区間*i*におけるハザード比

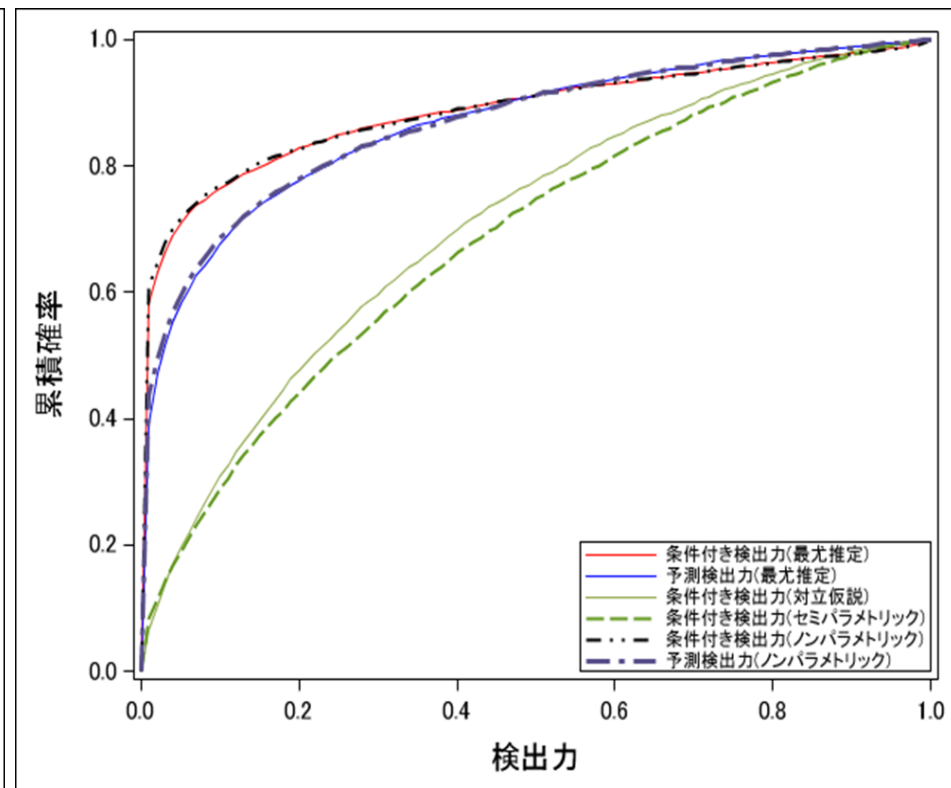
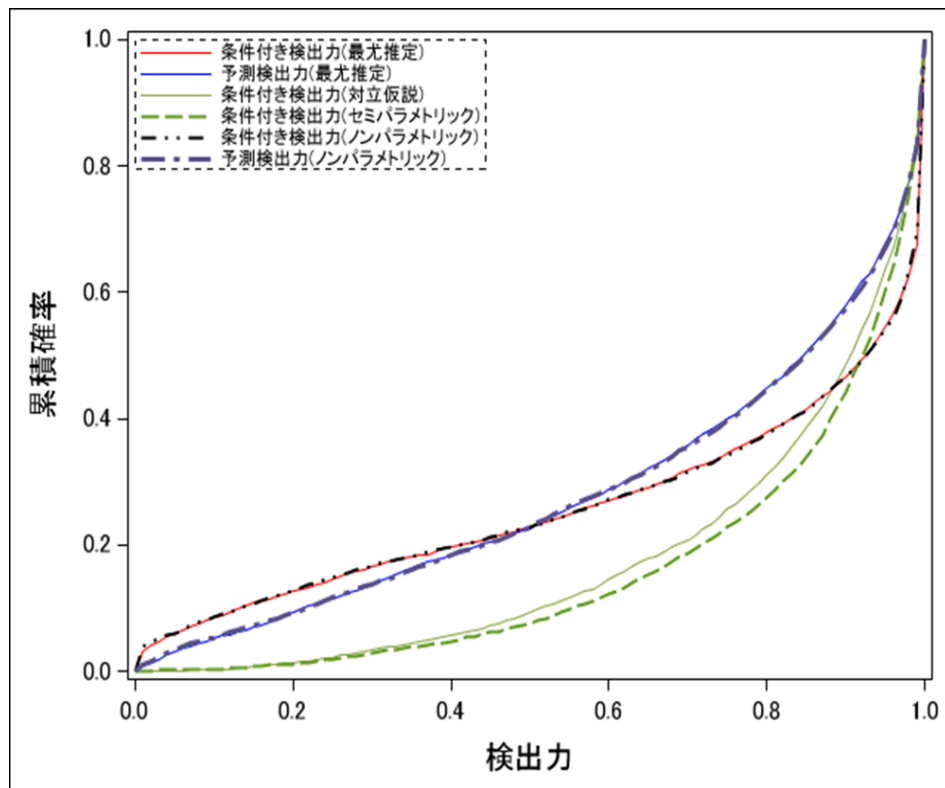
# 結果

シミュレーションを用いて条件付き検出力・予測検出力を定量的に評価する

□ ランダム化二群比較試験

□ 中間解析時の症例数86人

□ シミュレーション回数: 5000回



## まとめ

- 検出力0.2以下に着目(パラメトリック)
  - 帰無仮説の下では条件付き検出力(最尤推定)の累積確率が一番高い(無益な試験を中止しやすい)
  - 対立仮説の下では条件付き検出力(対立仮説)の累積確率が一番低い(有益な試験を中止しにくい)
- セミパラメトリック、ノンパラメトリックな手法と比較するとほぼ一致していることが確認できる

## 参考文献

- [1] Alex D, Geert M *et al.* Analysis of clinical trials using SAS® : A practical guide. SAS Institute. 2005
- [2] Christopher J, Bruce WT. Group sequential methods with applications to clinical trials. Chapman & Hall/CRC. 2000
- [3] Lakatos E. Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics*. Vol.44. No.1. pp229-241. 1988
- [4] Huson LW. The Bayesian Bootstrap in a Predictive Power Analysis. *Case Studies in Business, Industry and Government Statistics*. Vol.3. No.1. pp.18-22. 2009
- [5] Pei H, Tze LL *et al.* Futility stopping in clinical trials. *Statistics and Its Interface*. Vol.5. pp415-423. 2012
- [6] Ware JH, Muller JE *et al.* The futility index: An approach to the cost-effective termination of randomized clinical trials. *American Journal of Medicine*. Vol.78. No.4. pp.635-643. 1985
- [7] 独立行政法人医薬品医療機器総合機構. <http://www.pmda.go.jp/>. (Accessed 16 July 2017)
- [8] 日本製薬工業協会 医薬品評価委員会 統計・DM 部会. はじめての中間解析—実装のためのガイダンス—. 医薬出版センター. 2009