

臨床試験

第7回検定の多重性 2017年5月17日 Clinical
臨床統計学/臨床統計家育成コース 田中司朗 Biostatistics
Course

検定の多重性

- 何回も検定を行うとどこかで有意差が出る可能性が高くなる

| 検定の数 | 差がないのにどこかで有意差が出る確率 |
|------|--------------------|
| 1 | 5% |
| 2 | 9.7% |
| 3 | 14.3% |
| 4 | 18.5% |
| 5 | 22.6% |
| 10 | 40.0% |
| 20 | 64.1% |



仮説検定の復習

- α エラー
 - 帰無仮説が正しいのに、それを棄却する誤り
- β エラー
 - 帰無仮説が間違いなのに、それを保留する誤り
 - 検出力 = $1 - \beta$
- 何回も検定を行うと α エラーが増大
 - 回避するには有意水準を0.05より小さくすればよい

| 試験の結果 | 真実 | |
|-----------------------|----------------|-------------------|
| | 差がない (帰無仮説) | 差がある (対立仮説) |
| $p \geq 0.05$ (有意差なし) | | β エラー |
| $p < 0.05$ (有意差あり) | α エラー | 検出力 = $1 - \beta$ |

多重性を調整した統計手法が必要になる場面

- 3群以上の比較試験
 - ケースバイケース
- 多重エンドポイント
 - 全てのエンドポイントで有意差が求められるときは調整する必要はない
 - どれか一つが有意でも有効と判断するときは多重性が生じる
- サブグループ解析
 - 副次的解析なので調整しない
- 中間解析
 - 必須

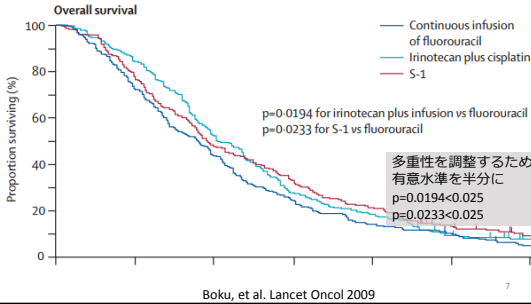
並行群間比較デザインと要因デザイン

- 3群以上の比較試験では、多重性の調整をどうかはケースバイケース
- 並行群間比較試験 (parallel trial)
 - 複数の試験治療をプラセボまたは標準治療と比較
 - 三つ以上の投与量を比較
- 要因計画試験 (factorial design trial)
 - 複数の介入の組み合わせを評価
 - 多重性は生じないため調整する必要はない

例1. JCOG9912試験

- 未治療の進行胃癌患者704人を対象とした三群比較ランダム化臨床試験
 - 試験治療群: イリノテカン+シスプラチン (IP療法)
 - 試験治療群: S-1
 - コントロール群: 5-FU持続静注
- 一つの試験で、二つの仮説を検証
 - 優越性仮説
 - IP療法の全生存期間は、5-FUに優るかどうかが
 - 非劣性仮説
 - S-1の全生存期間は、5-FUに劣らないかどうか
 - 劣らなければ、S-1は経口投与のため5-FUより簡便

例1. JCOG9912試験



Boku, et al. Lancet Oncol 2009

例2. Physicians' Health Study

- 40~84歳の健康人男性の医師に、以下の治療をランダムに割り付ける2×2要因計画試験
 - A群: 低用量アスピリン+βカロチン
 - B群: 低用量アスピリン+プラセボ
 - C群: プラセボ+βカロチン
 - D群: プラセボ+プラセボ
- 二つの仮説
 - 低用量アスピリンが心血管死を減少させるか
 - 「A群+B群」は「C群+D群」に優るかどうか
 - βカロチンが癌の発生を減少させるか
 - 「A群+C群」は「B群+D群」に優るかどうか
- 多重性は調整しない

Steering Committee of the Physicians' Health Study Research Group. N Engl J Med. 1989

多重エンドポイントの例

- 糖尿病
 - 大血管障害 (心筋梗塞, 脳卒中)
 - 細小血管障害 (糖尿病網膜症, 腎症, 神経障害)
- 眼科疾患
 - 右目と左目
- アルツハイマー病
 - 認知機能
 - 実行機能 (または全般的認知)
- FDAはこの両方を主要エンドポイントとして要求

Kozauer and Russell. New Eng J Med 2013

例3. アルツハイマー病臨床試験

- アルツハイマー病患者1537人を3群に割り付け
 - Semagacestat 140 mg/d
 - Semagacestat 100 mg/d
 - プラセボ
- Co-primary endpoints
 - ADAS-cog変化 (0~76週)
 - ADCS-ADL変化 (0~76週)
- Mixed model for repeated measuresで解析

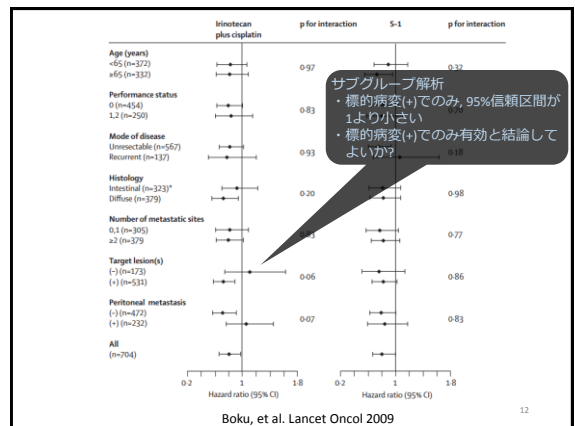
Doody, et al. New Eng J Med 2013

例3. アルツハイマー病臨床試験

| | プラセボ | Semagacestat 100mg | Semagacestat 140mg | p | p |
|------------------------------|----------------------|-----------------------|------------------------|---------------|---------------|
| | | | | 100mg vs プラセボ | 140mg vs プラセボ |
| ADAS-cog変化 (95%信頼区間) (<0で改善) | 6.4 (5.5 to 7.4) | 7.5 (6.4 to 8.5) | 7.8 (6.7 to 8.6) | 0.15 | 0.07 |
| ADCS-ADL変化 (95%信頼区間) (>0で改善) | -9.0 (-10.4 to -7.7) | -10.5 (-11.9 to -9.1) | -12.6 (-14.1 to -11.2) | 0.06 | <0.01 |

- 複数の仮説
- Co-primary endpoints (AFAS-cogとADCS-ADL)
- 多群比較 (100mg vs プラセボ, 140mg vs プラセボ)
- 仮説とその優先順位を事前に設定し、順番に検定 (ゲートキーピング法)

Doody, et al. New Eng J Med 2013



Boku, et al. Lancet Oncol 2009

サブグループ解析の解釈

- 全体で有意でない場合
 - サブグループで有意差がでて、何回も検定を行ったことによる偶然の結果かもしれない
- 全体で有意な場合
 - 全体で「ハザード比≠1」と分かっているのに、サブグループごとに「ハザード比=1」かどうかの検定を行うのは論理が一貫しない
 - 主効果ではなく、交互作用の検定を用いる

13

主効果と交互作用

- 二つのハザード比
 - 標的病変(+)の集団におけるハザード比
 - 標的病変(-)の集団におけるハザード比
- 主効果の検定
 - ハザード比=1かどうか
 - $y = \beta_0 + \beta_1 x_1$
- 交互作用の検定
 - 二つのハザード比が等しいかどうか
 - $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$

14

主効果と交互作用

| | 標的病変(+) | | | 標的病変(-) | | |
|-------------------|---------|------|------|---------|------|------|
| | 推定値 | 標準誤差 | p | 推定値 | 標準誤差 | p |
| ハザード比 | 1.1 | | | 0.7 | | |
| 対数ハザード比 | 0.09 | 0.19 | 0.64 | -0.36 | 0.14 | 0.01 |
| | 交互作用 | | | | | |
| (+)と(-)の差とそのSEを計算 | | | | 0.45 | 0.24 | 0.06 |

15

Key words

- 多重エンドポイント (multiple endpoints)
- 複合エンドポイント (composite endpoints)
- コプライマリーエンドポイント (co-primary endpoints)
- 並行群間比較試験 (parallel trial)
- 要因計画試験 (factorial design trial)
- サブグループ解析 (subgroup analysis)
 - 交互作用の検定 (interaction test)

16