

# 臨床試験

第8回中間解析  
臨床統計学/臨床統計教育育成コース

2017年5月17日  
田中司朗

Clinical  
Biostatistics  
Course



## 試験経過を把握する目的

- 品質管理
    - 登録, 治療, 評価などの手順
    - 計画通りデータ収集がなされているか?
    - 登録数, イベント発生状況は?
  - 安全性モニタリング
    - 被験者保護
    - 安全性評価
  - 中間解析
    - 試験途中の全ての群間比較解析  
(ICH E9臨床試験のための統計的原則)
- ・モニタリング  
 ・データマネジメント  
 ・データモニタリング委員会

2

## データモニタリング委員会



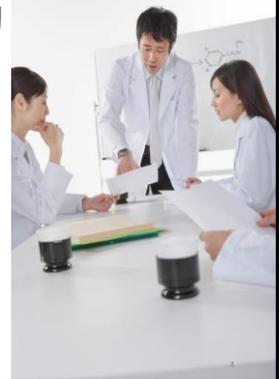
- 効果安全性評価委員会とも
- 専門性を有する委員（特に医師と生物統計家）から構成され、試験実施中の中間データについて中立的な評価を行う組織
- 被験者保護・倫理的・科学的妥当性の確保のために適切な助言・勧告を行う
  - 安全性モニタリング
  - 中間解析に基づく評価
  - 試験実施状況のモニタリング
  - 外部情報の利用

PMDA. データモニタリング委員会に関するガイドライン.  
[http://www.pmda.go.jp/kijunsakusei/file/guideline/new\\_drug/data-monitoring.pdf](http://www.pmda.go.jp/kijunsakusei/file/guideline/new_drug/data-monitoring.pdf)

3

## 中間解析の目的

- 判断
  - 有効性が示された (有効性, efficacy)
  - 有効性を証明できる見込みがないことが判明した (無益性, futility)
  - 副作用が許容できないことが明らかになった (安全性, safety)
- アクション
  - プロトコール改訂を勧告
  - 早期中止を勧告
  - 早期公表を勧告



4

## 誰のための 早期中止?



- 治療中の被験者
  - 有効な治療への切り替えや副作用の対応などが必要かもしれない
- 治療が終了した被験者
  - 多くの場合, 早期中止によるメリットはない
- これから登録される被験者
  - 結論の出た試験に登録することは非倫理的
- 将来の患者
  - 結果を公表することで新治療を早く届けられる
  - データ不足のため結論が出ないリスク

5

## 中間解析のロジック

- 中間解析の結果は不確実であることが前提
- しかしながら, 患者登録中に有効性または無益性の強い証拠が得られれば, 個の倫理を優先し, 登録を中止すべき
- 治療中の患者がいる場合, 治療切り替えを推奨するか, 許容するか, 許容しないか, 難しい三択を迫られる
- 全患者で治療が終了しているとき, 有効性の証拠が極めて強い場合を除いて, 早期中止のメリットはないことが多い

6

## 中間解析の統計手法: α消費関数

- 検定の多重性
  - 検定を複数回行うことによりαエラーが増大する
  - 回避するには有意水準を厳しくすればよい
- α消費関数により有意水準を決定

回数 (症例数の%)	等間隔型の有意水準 ( Pocock型)	非等間隔型の有意水準 ( O'Brien and Fleming型)
第1回 (20%)	0.015	0.000001
第2回 (40%)	0.015	0.0008
第3回 (60%)	0.016	0.007
第4回 (80%)	0.017	0.022
最終解析 (100%)	0.017	0.042

Lan and Demets, Biometrika 1983

## 例1. 早期胃癌臨床試験

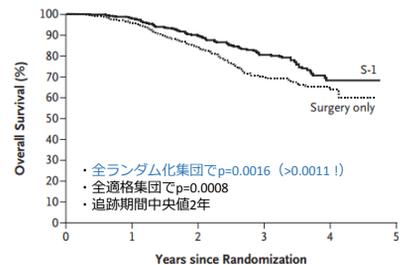
- 対象
  - ステージIIまたはIIIでD2廓清手術を受ける胃癌患者
- 治療
  - S-1による術後補助化学療法 v.s. 手術単独
  - 治療期間1年
- 主要エンドポイント: 全生存期間
- 有効性の主たる解析対象集団
  - 全ランダム化集団 (適格性に関わらず)

Sakuramoto, et al. New Engl J Med 2007

## 中間解析の方法

- Efficacy was to be evaluated in two interim analyses performed by an independent data and safety monitoring committee 1 year and 3 years after the completion of enrollment. Significance was evaluated with the use of the method of Lan and DeMets and the O'Brien-Fleming boundary.
  - 2回の中間解析 (登録終了後1年と3年)
  - Lan and Demetsの方法 (α消費関数) とO'Brien-Flemingの境界により, 有意性を評価
  - 第一回中間解析の有意水準は0.0011

## 第一回中間解析の結果



No. at Risk	529	515	370	196	46
S-1	530	504	352	163	40

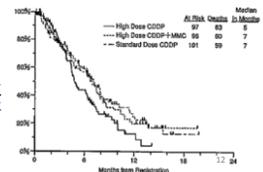
Sakuramoto, et al. NEJM 2007

## データモニタリング委員会の判断

- 本試験においては, 評価項目である overall survival および relapse-free survival に関して, 多重性を考慮した統計的評価において胃がん術後補助化学療法の有効性が認められた
- 効果・安全性委員会は, S-1が既に市販されていること, および胃がん補助療法の臨床試験が他にも計画中であることを鑑み, 今後治療対象となる患者の利益を考慮してしかるべき学会に早期に公表するよう勧告した

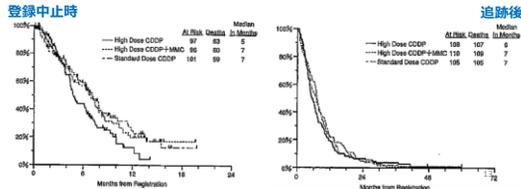
## 例2. 非小細胞肺癌臨床試験

- シスプラチン v.s. 高用量シス v.s. 高用量シス+MMC
- 主要エンドポイント: 無病生存期間
- 中間解析の計画
  - 期待イベント数の1/3, 2/3が観察された時点で, それぞれ有意水準0.005の片側ログランク検定
- 第一回中間解析
  - 「高用量シスが勝る」という対立仮説 (ハザード比1.25) が有意に棄却されてしまった



## データモニタリング委員会の判断

- Futilityによる登録中止を勧告
- 追跡後の生存曲線は登録中止時点とは大きく異なっていた
- もし登録を中止しなかったら？



## 例3. 抗不整脈薬臨床試験

- 背景
  - 抗不整脈薬は、抗不整脈効果を根拠としてFDAにより承認
  - 毎年20万人以上が使用
- 国立心臓血液研究所による試験CAST
  - エンカイニド、フレカイニド、モリスジンのランダム化プラセボ対照試験
  - 対象: 心筋梗塞後の心室性不整脈を有する患者
  - 主要エンドポイント: 心突然死
  - 予定症例数: 4400人
  - 予想された心突然死リスク: 3年で11%
- 中止基準
  - 主たる解析は片側検定
  - $\alpha$ 消費関数による中止基準は、有効性・安全性のため両側に

CAST. New Engl J Med 1989

14

## 例3. 抗不整脈薬臨床試験

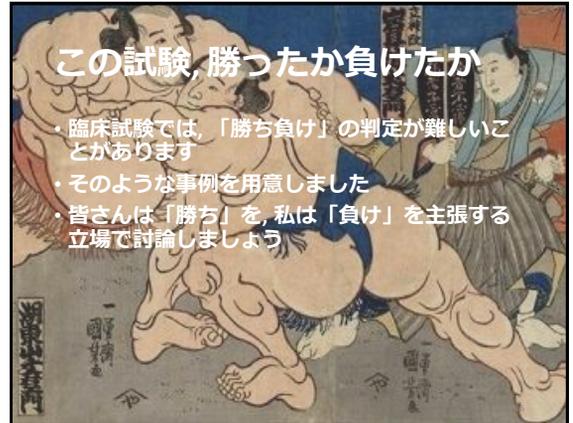
- 1987年3月 第1回DMC会議（プロトコルレビュー）
- 1987年6月 患者登録開始
- 1988年1月 第2回DMC会議（中間データはブラインドに）
- 1988年9月 第3回DMC会議でのデータレビュー
  - X群の突然死: 3/576
  - Y群の突然死: 19/571
- 1989年2月 国立心臓血液研究所がブラインド解除を決定
- 1989年4月 第4回DMC会議（一部登録中止・結果公表）
  - エンカイニド・フレカイニド群突然死33人
  - プラセボ群の突然死9人

CAST. New Engl J Med 1989

15

## この試験、勝ったか負けたか

- 臨床試験では、「勝ち負け」の判定が難しいことがあります
- そのような事例を用意しました
- 皆さんは「勝ち」を、私は「負け」を主張する立場で討論しましょう



## CHARM試験

- カンデサルタンをプラセボと比較するランダム化二重盲検試験
- 対象
  - 18歳以上NYHAクラスII~IVの症候性心不全患者
- 三つのサブ試験
  - CHARM-Added (2,028人)
  - CHARM-Alternative (2,548人)
  - CHARM-Preserved (3,025人)

17

## 三つのサブ試験

- CHARM-Added (2,028人)
  - ACE阻害薬あり,かつ左室駆出率 (LVEF) <40%
- CHARM-Alternative (2,548人)
  - ACE阻害薬の忍容がなく (72%が咳),かつLVEF<40%
- CHARM-Preserved (3,025人)
  - LVEF>40%
- 試験全体 (CHARM-Overall)

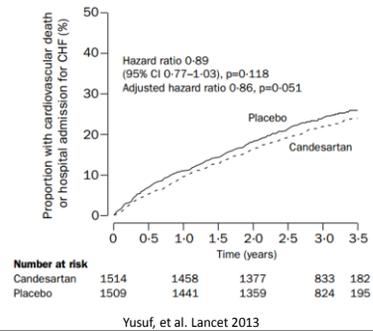
18

## エンドポイント

- サブ試験の主要エンドポイント
  - 心血管死または心不全関連入院
- CHARM-Overallの主要エンドポイント
  - 全死亡
- サンプルサイズ計算はサブ試験ごとに
- 結果としてCHARM-Overallでも検出力は確保

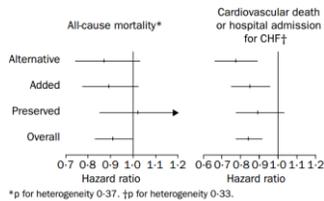
19

## CHARM-Preserved



20

## CHARM-Overall



	Events/patients		P-value (Log rank)	Hazard ratio (95% CI)
	Candesartan	Placebo		
Primary endpoint	483/1276	538/1272	0.011	0.85 (0.75, 0.96)
CHARM-Added	334/1013	406/1015	<0.001	0.77 (0.66, 0.89)
CHARM-Alternative	333/1514	366/1509	0.118	0.89 (0.77, 1.03)
CHARM-Preserved	1150/3803	1310/3796	<0.001	0.84 (0.77, 0.90)
All trials combined				

Pfeffer, et al. Lancet 2003

21

## CHARM-Preservedの結論

- カンデサルタンは、LVEF>40%の心不全患者において、CHF悪化による入院の中程度の予防効果がある

**Interpretation** Candesartan has a moderate impact in preventing admissions for CHF among patients who have heart failure and LVEF higher than 40%.

Yusuf, et al. Lancet 2013

22

## CHARM-Preservedの結論の根拠

- CHARM-Overallの主効果が有意で、交互作用の検定で有意差がないこと
  - The absence of heterogeneity in results underscores that this benefit was achieved across a broad spectrum of patients. Subgroup analyses must be interpreted cautiously since the most rigorous test of the study hypothesis is derived from the entire population, in which consistency of this benefit was seen
- CHARM-Preservedの主効果の検定では判断していない

Pfeffer, et al. Lancet 2003

23

## 結論（皆さんの立場）

- カンデサルタンは、LVEF>40%の心不全患者において、CHF悪化による入院の中程度の予防効果がある

**Interpretation** Candesartan has a moderate impact in preventing admissions for CHF among patients who have heart failure and LVEF higher than 40%.

Yusuf, et al. Lancet 2013

24

## 批判1

- CHARM-Preservedの主たる解析の結果
  - 調整しなければ $p=0.118$
  - 調整しても $p=0.051$
- 調整したp値では、以下の因子をCox回帰で調整
  - 年齢, 性別, 民族, NYHA, LVEF, 心拍数, 血圧, BMI, 心不全の原因, 既往歴, CHFの症状, 治療
- ランダム化しているのだから調整すべきではないのでは?

25

## 批判2

- CHARM-Added, -Alternative, -Preservedで、それぞれ検証的な臨床試験を計画し、サンプルサイズを計算している
- それで有意ではなかったのだから、CHARM-Overallの結果は見るべきではないのでは?

26

## 中間解析のルール

- 有効性の早期中止基準
  - 試験全体の死亡率の差が $p<0.001$ 
    - 層別ログランク検定の両側p値
    - ただし試験開始18ヶ月以内は $p<0.0001$ と厳しくする
  - サブ試験の早期中止基準も同様だが、交互作用がないことが条件
- 主要エンドポイントに基づく基準は用いない
- データモニタリング委員会, 運営委員会, スポンサーの間で事前に合意

27

## 中間解析の結果

Analysis Date	CHARM-Alternative		CHARM-Added		CHARM-Preserved		Overall Program		p-value
	C	P	C	P	C	P	C	P	
9 Aug 1999	3	0	5	4	0	0	8	4	0.3
27 March 2000	20	38	45	69	11	16	133	198	0.0002
27 July 2000	39	60	76	115	18	25	260	339	0.0006
Final Report**							886	945	0.055

\* Each line gives the number of deaths by treatment group for each constituent trial and overall, plus the overall logrank p-value, stratified by trial (C = candesartan, P = placebo).  
 \*\* Final report on September 6, 2005, based on follow-up to March 31, 2003.

Demets, et al. Data Monitoring in Clinical Trials: A Case Studies Approach 2005 28

## 中間解析の結果

Analysis Date	CHARM-Alternative		CHARM-Added		CHARM-Preserved		Overall Program		p-value
	C	P	C	P	C	P	C	P	
9 Aug 1999	3	0	5	4	0	0	8	4	0.3
27 March 2000	20	38	45	69	11	16	133	198	0.0002
27 July 2000	39	60	76	115	18	25	260	339	0.0006
1 March 2001	66	100	140	168	54	71	387	474	0.0010
9 Aug 2001	117	148	186						
22 Feb 2002	166	198	258						
1 Aug 2002	210	236	298						
Final Report**									

\* Each line gives the number of deaths by plus the overall logrank p-value, stratified

\*\* Final report on September 6, 2003, based on follow-up to March 31, 2003.

Demets, et al. Data Monitoring in Clinical Trials: A Case Studies Approach 2005 29

## 批判3

- 2001年に試験全体の早期中止基準に到達
- 死亡率の差を根拠に、有効性の早期中止をすべきだったのではないかと?

30

## 論点

- きわどく有意ではなかったときに結果をどのように解釈するか
- サブ試験やサブグループがあるときの判断基準
  - サブ試験ごとの主効果のp値
  - 全体の主効果のp値
  - 交互作用のp値
- サブ試験やサブグループがあるとき中間解析の意思決定は複雑になる

31

## Key words & further reading

- 中間解析 (interim analysis)
  - データモニタリング委員会 (data monitoring committee)
  - 有効性による中止 (efficacy)
  - 無益性による中止 (futility)
  - 安全性による中止 (safety)
  - $\alpha$ 消費関数 ( $\alpha$  spending function)
- Pocock SJ. Current controversies in data monitoring for clinical trials. Clin Trials. 2006;3(6):513-21.

32