

2019/5/23 北大SPH・統計解析の応用①②

## 臨床試験デザイン



北海道大学 医学統計学  
横田 勲

## 今回の内容

2

- ▶ 臨床試験の質管理
- ▶ 試験の目的とデザインの種類
- ▶ 臨床試験登録
- ▶ CONSORTガイドライン

## 臨床試験の計画

3

- ▶ 対照 control をどうおくか
  - ▶ 試験のタイプをどうするか
    - ▶ 並行群間、クロスオーバー、用量漸増
    - ▶ ランダム化するか単群試験か
      - ▶ ランダム化するなら盲検化 blind をするか
  - ▶ 何をもちて評価するか
    - ▶ どの程度の効果があれば有意義か
- ↓
- ▶ 統計的仮説を設定しサンプルサイズ設計

## PICO

4

- ▶ P population, patient
- ▶ I intervention, indicator
- ▶ C comparison, control
- ▶ O outcome

## 臨床試験の科学的妥当性

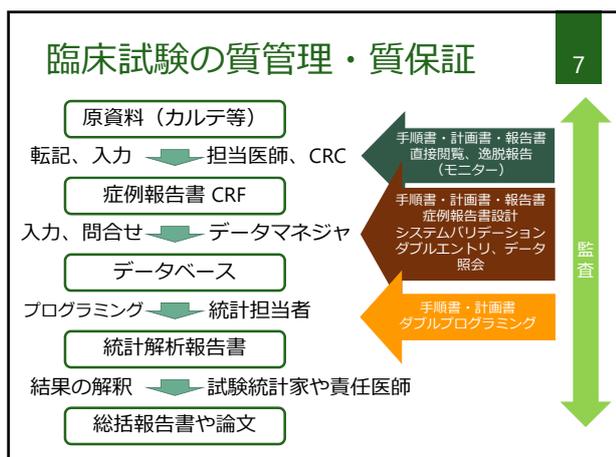
5

- ▶ 再現実験が困難
  - ▶ 同条件で繰り返すことは困難
- ▶ デザインとプロセスの正しさを保証
  - ▶ プロトコルを定め、倫理審査を受ける
  - ▶ 標準業務手順書(SOP)等を作成し、手順書通りに実施した記録を残す

## プロトコルの作成

6

- ▶ 総括報告書や論文化を念頭に置いたプロトコルの作成が要求される
  - ▶ 論文公表の際にSupplementとしてプロトコルと解析計画書の添付が要求される
    - ▶ しかも、1.0版とそこからの変更履歴も全て
- ▶ プロトコルのひな形
  - ▶ SPIRIT
  - ▶ JCOGのプロトコルテンプレート



### データマネジメント

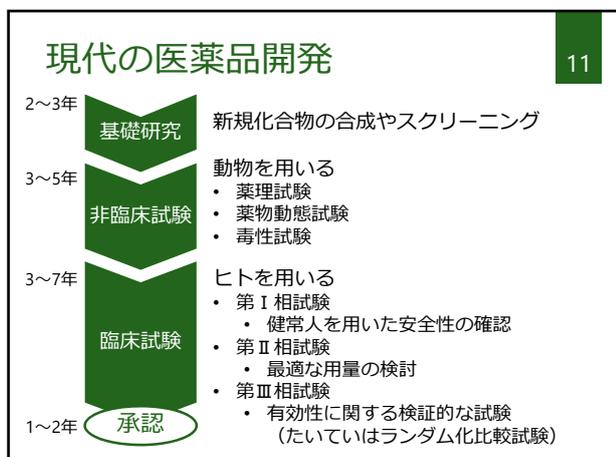
- ▶ 研究目的に関して、必要かつ十分な質の高いデータを効率よく収集する行為
- ▶ 試験開始前; データの最終的な質を決める
  - ▶ 収集方法、収集すべき内容を決める
  - ▶ 何でもかんでも収集することが質が高いとはいえない (むしろ質は下がる)
    - ▶ Risk-based monitoring
- ▶ 試験終了後; データ固定し、統計家へ提供

### モニタリング

- ▶ 試験の進行を監視し、試験が計画書、手順書、規制要件等に従って実施、記録、報告されていることを保証する行為
- ▶ 目的は次のことを確認すること
  - ▶ 被験者の人権と福祉が保護されていること
  - ▶ 報告された試験データが正確かつ完全で、原資料から検証できること
  - ▶ 試験が最新のプロトコル、GCP、および適用される規制要件を遵守して実施されていること

### 臨床試験に必要なもの

- ▶ 研究仮説
- ▶ 試験実施計画書 (プロトコル)
- ▶ 診断・評価の基準
  - ▶ 疾患の定義、有効性・安全性の評価基準
- ▶ 実施システム
  - ▶ スポンサー、施設 (IRB, 事務局)
  - ▶ サポート体制 (ARO, CRO, SMO)
- ▶ 統計家の参画と統計解析計画書
- ▶ 症例報告書とその標準化
- ▶ データの質管理・保証システム
  - ▶ データマネジメント、モニタリング、監査



### 探索的か検証的か

- ▶ 探索的臨床試験 exploratory
  - ▶ 仮説創出、ある程度仮説の確からしさを検討
  - ▶ Phase II 試験のような位置づけ
  - ▶ 効果推定が中心
  - ▶ ベイズ統計学を応用した柔軟なデザインも適
- ▶ 検証的臨床試験 confirmatory
  - ▶ 仮説の正しさを最終的に決定づける
  - ▶ Phase III 試験の位置づけ
  - ▶ 二分的議論に単純に落とす検定が採用

## 並行群間試験

13

- ▶ 対象者をいずれかの群に割付
- ▶ 介入によって対象者の状態が変化し、元に戻ることは想定されづらい疾患
  - ▶ 多くの臨床試験で採用されるデザイン

## クロスオーバー試験

14

- ▶ 対象者に両方の治療を実施し、治療順や部位をランダム化
- ▶ 同一対象者で比較するため、背景情報のバランスがとれる
  - ▶ 相関のあるデータとして解析
- ▶ 治療の順序効果がない疾患
  - ▶ しばらく放っておけば(wash-out)元に戻る
- ▶ 部位間で影響を及ぼさない疾患

## 前後比較

15

- ▶ 放っておけば絶対に予後が改善しないと医学的に認められる疾患
  - ▶ 外科手術を要する  
身体機能に障害が生じるような疾患
  - ▶ ランダム化によって経過観察群に割付られることが非倫理的とみなされる疾患
    - ▶ 経過観察による患者への不利益が限られる場合、並行群間試験を要求されることも

## 差がないことを示したい

16

- ▶ DPP4阻害薬のような類薬の開発
  - ▶ 先行薬と同等か  
それ以上の効果があればよい
- ▶ ジェネリック医薬品の開発
  - ▶ 先発薬より効いてしまっても困る
- ▶ 新規検査法の性能評価
  - ▶ 既存の検査法と同じ結果であるか

## 優越性 vs 非劣性、同等性

17

- ▶ 優越性 superiority
  - ▶ AがBより優れている
    - ▶ AとBが同じ、という帰無仮説を棄却
- ▶ 非劣性 non-inferiority
  - ▶ AがBより劣っているということはない
    - ▶ AがBより少し劣っている、という帰無仮説を棄却
- ▶ 同等性 equivalence
  - ▶ AとBが同じ
    - ▶ AがBより少し劣っている、  
AがBより少し優れている、という2つの帰無仮説を棄却

## 検定におけるP値

18

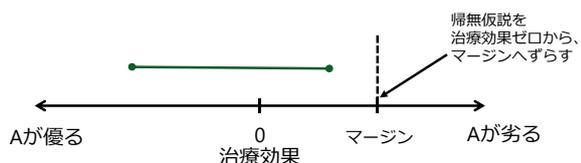
- ▶ 有意水準未満「統計的有意」
  - ▶ 帰無仮説が棄却され、差があったと結論
- ▶ 有意水準以上「統計的有意でなかった」
  - ▶ 帰無仮説が棄却されず、差があるかないか分からなかった
  - ▶ 「差がない」とは結論付けられない

### 非劣性を示したい

19

値が低い（減少する）ほど効果がある場合

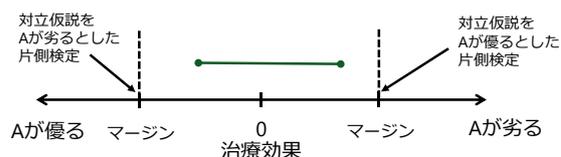
- ▶ 信頼区間の上限が非劣性マージンを超えない
- ▶ 効果の差が非劣性マージンに等しい、という帰無仮説を片側検定で棄却



### 同等性を示したい

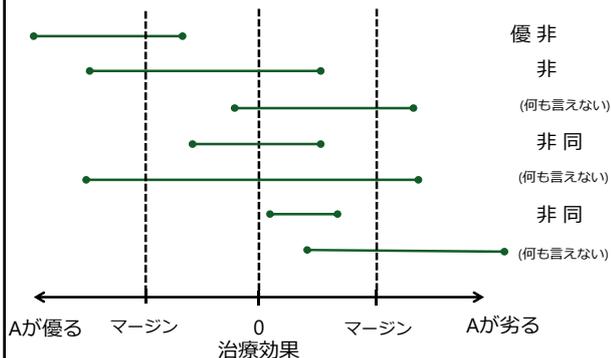
20

- ▶ 信頼区間が同等性マージンの間におさまっている
- ▶ 効果の差が各同等性マージンに等しい、という帰無仮説を片側検定で(2回)棄却
  - ▶ 決して両側検定を用いるわけではない



### 信頼区間を用いて表現

21



### 非劣性試験の注意点

22

- ▶ 優越性が証明できなかったから、非劣性を証明して、結論付けるのはナシ
  - ▶ 検定の多重性
- ▶ 盲検性があるか？
  - ▶ 差がないようにすれば、非劣性仮説は証明
    - ▶ 割付を無視して、全員同じ治療を行ったら
    - ▶ 対照群に割付られた人に試験治療を行ったら
  - ▶ せめて介入結果の評価を盲検化できないか

### ネガティブな論文は出づらい

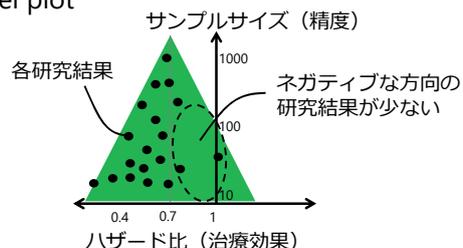
25

- ▶ 企業にとっては出たくない
  - ▶ 販売促進に使いづらい
  - ▶ メディカルライティング等費用もかかる
- ▶ 医師にとって
  - ▶ インパクトファクターの高い雑誌に出せない
  - ▶ だったら忙しいし、作成しない
- ▶ 世の中にとっては・・・？

### 出版バイアス

26

- ▶ ポジティブな結果ばかり出版されるため全体では過大評価してしまう
- ▶ Funnel plot



## 公正な出版のために

27

- ▶ 試験結果は必ず論文化
  - ▶ ヘルシンキ宣言2000,2002
  - ▶ 人を対象とする医学系研究に関する倫理研究
- ▶ そもそもどんな試験を実施中かを公開
  - ▶ 臨床試験登録(ICMJE; 2005)
- ▶ 誰がスポンサーで誰が執筆したか
  - ▶ COIの開示(ICMJE 統一投稿規定; 2008)
  - ▶ GPP3 (2015) 企業がスポンサー or 支援する場合
- ▶ 論文に必要な項目は何か
  - ▶ CONSORTガイドライン(1996-, 2010最新改訂)

ICMJE: 医学雑誌編集者国際委員会

## 臨床試験登録

28

- ▶ 科学的、倫理的、説明責任のため  
臨床試験の概要に関する項目を開示
- ▶ WHO・ICTRPの一次レジストリ、  
もしくは ClinicalTrials.govへの登録
  - ▶ 複数システムへの登録もOK
  - ▶ 日本はUMIN、JAPIC、JMACCT、jRCT
- ▶ 発行される登録番号なしでは、  
承認資料に使えない、論文化できない

## The EQUATOR network

<http://www.equator-network.org/>

29

## CONSORT statement

30

- ▶ ICMJEの活動成果
- ▶ 1996年初版
  - ▶ 2001年、2010年に改訂
- ▶ 群間並行ランダム化試験(RCT)を対象
  - ▶ クラスターランダム化試験や非劣性試験には  
CONSORT拡張版
- ▶ 400以上の雑誌が支持
- ▶ 25項目のチェックリストと  
フローダイアグラム

## CONSORTの目的・方針

31

- ▶ 完全、明快かつ透明性の高い報告を発展
  - ▶ 著者とpeer reviewerによる査読
  - ▶ 読者による文献の批判的評価 を助ける
- ▶ 厳格な形式を守らせるものではない
  - ▶ 論文の構造の標準化を意図するものではない
  - ▶ 雑誌のスタイル、投稿規定、研究分野の慣習、  
著者の好みなどに論文の形式はしたがうべき
- ▶ チェックリストは、試験の質スコアを  
構築するために使うべきではない

## タイトル・抄録・イントロ

34

- 1a タイトルにランダム化比較試験  
であることを記載
- 1b 構造化抄録
  - ▶ 背景、試験デザイン、方法、結果、結論
- 2a 科学的背景と論拠
- 2b 目的や仮説

## 試験デザイン

35

### 3a 試験デザイン

- ▶ 並行群間か、要因実験か
- ▶ ランダム化の割付比も書く
  - ▶ ランダム化の詳細は8a,8bにて

### 3b 試験開始後の方法の重要な変更

- ▶ 理由を明らかにする

## 参加者

36

### 4a 適格規準・条件

- ▶ 選択基準は、治療を適用したい標的集団を表した規準
  - ▶ 治験であれば適応に関わる
- ▶ 除外基準は、評価にバイアスを招く被験者を除くための規準
  - ▶ 安全性が保証できないような条件をもつ人は除く

### 4b データ収集のセッティングと場所

- ▶ 時期や研究を行った国、地域、施設

## 介入

37

### 5 各群の介入

- ▶ 他の研究者、臨床家が再現可能となるよう詳細に述べる
  - ▶ 研究のオリジナリティーがあることも
- ▶ 介入スケジュールや手順まで述べる
  - ▶ × 試験薬剤を投与した
  - ▶ ○ 試験薬剤 (xxmg/m<sup>2</sup>) を連日投与する4週と、その後2週間の休薬期間を1サイクルとし、合計8サイクル行った

## アウトカム①

38

### 6a 主要・副次評価項目

- ▶ 主要評価項目
  - ▶ その研究で最も明らかにしたい仮説を評価するための項目
  - ▶ 通常、αエラーを厳しく制御する
- ▶ 副次評価項目
  - ▶ 主要ほどではないが、確認しておきたい評価項目
  - ▶ 主要で代替エンドポイントを用いたならば、真のエンドポイントを副次に含めることも
  - ▶ 「探索的」の名のもと、αエラーの制御は行わないことが多い

## アウトカム②

39

### 6b 評価項目の変更

- ▶ 診断基準の変更
- ▶ 「イベント」の定義の変更
  - ▶ 長い試験期間の中で変更することがある
    - ▶ がん分野ではあまり見かけない
- ▶ 主要評価項目の変更の場合
  - ▶ 特に慎重にせねばならない
  - ▶ そもそもの仮説を正しく示すことができるのか？
  - ▶ スポンサー側に都合のいい変更でないか？

40

- ▶ 2005.10 登録開始
- ▶ 2007.7 選択基準変更
  - ▶ 463例登録済み
- ▶ 2008.10 主要評価項目をOSからPFSに
  - ▶ 1299例登録済み
- ▶ 2009.6 1873例登録完了
- ▶ OSをみるための増悪後の盲検維持は infeasible
- ▶ 後治療の影響を考慮し、PFSをプライマリに

NEJM 2011; 2473-83

## サンプルサイズ①

41

### 7a サンプルサイズの設定根拠

- ▶ 主要評価項目の検定に対応させてサンプルサイズ設計を行う

## サンプルサイズの決め方

42

- ▶ 「群間差がない」場合に、検定で有意差がつく割合を有意水準以下に
- ▶ 「群間差がある」場合に、検定で有意差がつく割合を一定水準以上に
  - ▶ この割合を検出力という
  - ▶ 「群間差」には臨床的に意味のある差を設定
- ▶ 以上を満たす必要サンプルサイズを計算

## 統計的仮説検定

43

- ▶ 帰無仮説：「比較群間で差がない」
  - ▶ OSに関し、対照群に対する試験群のハザード比が1
- ▶ 帰無仮説が正しい場合、観測された結果がどの程度極端なものであるかを評価
  - ▶ ありうるパターンを全通り考えたうち、何番目に極端なものかを計算
    - ▶ これがP値

## 割付を並べ替え

44

観測された結果

ID	OS time	割付
1	12	試験
2	7	対照
3	13	試験
4	3	対照
5	28*	試験



ハザード比：0.5

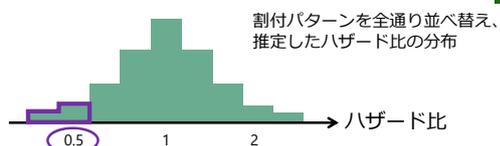
割付をランダムに並べ替えれば、試験と対照が同じ治療効果の場合に観測されるハザード比の分布が判明

ID	OS time	割付	試験	対照	ハザード比
1	12	試験	試験	対照	1.2
2	7	試験	対照	試験	1.5
3	13	試験	試験	対照	0.7
4	3	対照	対照	試験	0.5
5	28*	対照	試験	試験	0.5

\*打ち切り

## P値

45



- ▶ 並べ替え分布全体のうち、観測されたハザード比は何%目にあるか
  - ▶ パターン数が膨大なので、確率で表す
- ▶ 事前に定めた有意水準より小さければ、帰無仮説は間違いとする（棄却する）

## P値と有意水準

46

- ▶ 有意水準は片側2.5%がよく用いられる
- ▶ P値が有意水準より小さい場合
  - 群間差があった
- ▶ P値が有意水準より大きい場合
  - ✗ 群間差がなかった
  - 群間差があると証明できなかった
  - 群間差があるかないか、分からない

## αエラーとβエラー

47

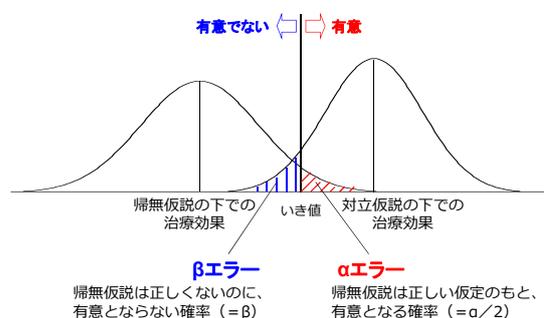
検定結果	母集団 (真)	
	差がない	差がある
有意差なし	正しい	第2種の過誤 βエラー
有意差あり	第1種の過誤 αエラー	正しい

有意水準で大きさを決める

サンプルサイズを大きくすることで軽減

## αエラーとβエラー

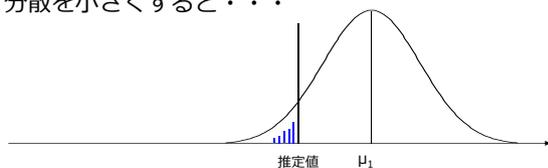
48



## αエラーとβエラー

49

もしサンプルサイズを大きくして分散を小さくすると・・・



βエラーは小さくなる

## 適切なサンプルサイズ

50

- ▶ サンプルサイズをただ大きくすればよいか？
  - ▶ 実際の研究では、多くの人数を集めることは時間的、金銭的に困難
  - ▶ わずかな差でも有意となってしまう、そのような差に臨床的意味があるのか？
- ▶ サンプルサイズが小さいとどうなるか？
  - ▶ 臨床的に意味のある差があるのに、有意であるという結果が得られにくくなる
  - ▶ 検出力不足

## サンプルサイズ設計の必要項目

51

- ▶ 検定手法
  - ▶ 評価項目と研究目的から選ぶ
- ▶ 有意水準      たいてい片側2.5%
- ▶ 検出力      たいてい80%
- ▶ 期待する治療効果
- ▶ 付加的情報

## 2群比較の優越性試験の場合

52

- ▶ 代表的な場合を表示

アウトカム	期待する治療効果	付加的情報
連続量	平均値の差	・ 標準偏差
二値	リスク差、リスク比、オッズ比	・ いずれかの群の発生割合
生存時間	ハザード比	・ いずれかの群のハザードや生存率 ・ 試験期間 ・ 患者登録期間

## サンプルサイズ設計の実際

53

- ▶ たいいていは専用ソフトを用いる
  - ▶ SAS、JMP、nQuery、PASS、、、、（有料）
  - ▶ SWOGのホームページ（無料）
- ▶ 複雑な仮説の場合はシミュレーション
  - ▶ 群間差がない下で、n人分のデータを発生し、検定を行い、有意な割合が有意水準以下
  - ▶ 期待する効果の下で、n人分のデータを発生、検定を行い、有意な割合が検出力以上
    - ▶ このnをいろいろ変えて検討

## サンプルサイズ②

54

### 7b 中間解析と中止基準

- ▶ 有効性の中間解析
  - ▶ 治療法が想定以上に有効であった場合、早期に試験を中止して、早く世の中に出したい
- ▶ 無益性 (futility) の中間解析
  - ▶ 治療法が想定よりはるかに効かなそうである場合、早期に試験を中止したい
- ▶ 安全性の中間解析（モニタリング）
  - ▶ 新薬開発ではほぼ必ず
  - ▶ この節でいう「中間解析」には通常含まれない

## 検定の多重性

55

- ▶ 中間解析で5%水準で検定
- ▶ 最終解析でも5%水準で検定
  - ▶ 治療が全く有効でない場合でも、有意差ありと結論してしまう機会が増える
  - ▶  $\alpha$  エラーの増加

## Bonferroni流の調整

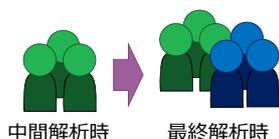
56

- ▶ 有意水準を検定の回数で割る
  - ▶ 2回検定をするなら、有意水準を2.5%に
  - ▶ 全体で $\alpha$  エラーを5%水準に担保できる
- ▶ かなり効率を失う
  - ▶ 本当は差がある場合に、きちんと有意差ありと判断する機会を失う
  - ▶ 2回の検定は独立でないため
    - ▶ 最終解析時データの一部は中間解析時データ

## グループ逐次法 group sequential method

57

- ▶ 中間解析からデータが追加されて、次回解析に用いられる特徴を考慮
- ▶ Lan-DeMets の  $\alpha$ 消費関数法
  - ▶ O'Brien-Fleming型
  - ▶ Pocock型
- ▶ Haybittle-Peto法



## 有効中止した試験結果

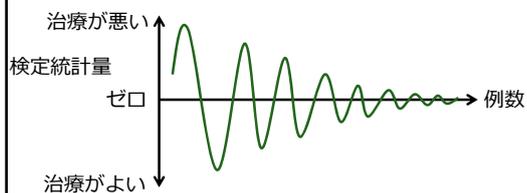
58

- ▶ 単純に求めた治療効果は過大評価しがち
  - ▶ グループ逐次法で制御したのは $\alpha$ エラーだけ
- ▶ Random highを刈り取るため
  - ▶ たまたま治療効果が大きくみられた場合、中間解析によって有効中止してしまう
  - ▶ たまたま治療効果が小さくみられても、試験継続により、最終解析結果での効果推定への影響は薄まる
- ▶ 中間解析を行う試験は、端から治療効果の推定を諦める

## 検定統計量の経時推移

59

- ▶ 帰無仮説が正しい場合

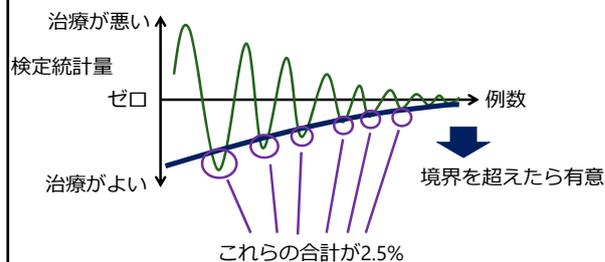


- ▶ 例数が増えるに連れ、検定統計量は真のゼロまわりに収束

## グループ逐次法による制御

60

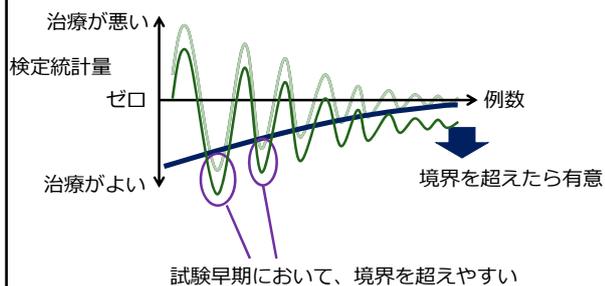
- ▶ 帰無仮説が正しい場合



## Random highの発生

61

- ▶ 対立仮説が正しい(治療効果がある)場合



## 誰が中間解析を行う？

63

- ▶ 独立データモニタリング委員会; IDMC, Independent Data Monitoring Committee
  - ▶ 試験のスポンサーとは独立な組織
    - ▶ 試験の中止・継続の意思決定を行うため
    - ▶ スポンサーは、二重盲検試験である場合、最終解析まで割付群を知ってはならない
- ▶ データモニタリング委員会; DMC
  - ▶ スポンサーも含まれるような場合
  - ▶ 単群試験や非盲検試験
  - ▶ 意思決定基準は試験開始前に定めておくべき

## 製薬企業向け実施体制モデル

64

Friedman et al. *Fundamentals of Clinical Trials*, 4<sup>th</sup> ed. Springer; 2015

## IDMCの基本的原則

65

- ▶ ① 責務
  - ▶ 被験者の利益を守る
  - ▶ 試験の完全性と信頼性を担保
  - ▶ 信頼できる最終結果を迅速かつ確実に利用できるようにする
- ▶ ② 構成委員
  - ▶ 医学専門家、生物統計家は必須
  - ▶ 生命倫理、疫学、基礎研究等の専門家も

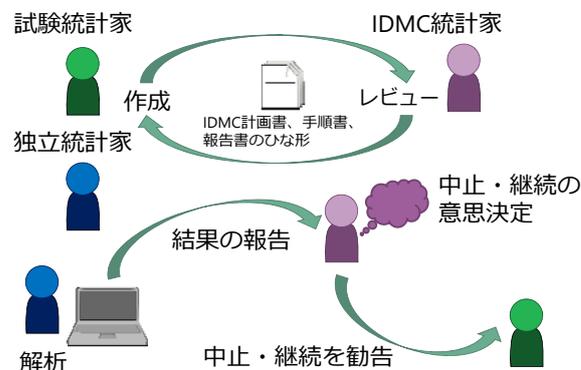
## IDMCの基本的原則（つづき）

66

- ▶ ③ 委員の利益相反
  - ▶ 金銭的、知識的、専門的、規制的に特筆すべき利益相反のない個人に限定すべき
- ▶ ④ 機密保持
  - ▶ 試験治療の有効性と安全性の比較結果は、IDMC委員のみが閲覧できる状態が理想

## 3人の生物統計家

67



## ランダム割付

68

- 8a 割付の方法
- 8b 割付のタイプ
  - ▶ 層別割付？層別因子は？
  - ▶ 置換ブロック法？ブロックサイズは？
  - ▶ 最小化法？
- 9 割付表の管理
- 10 割付の実施者

## 単純ランダム化 simple randomization

69

- ▶ 各患者に同じ確率でランダム化
- ▶ 人数のアンバランスが生じやすい
  - ▶ 例：100例を確率0.5ずつで割付
    - ▶ 50例/50例となる確率は約8%
    - ▶ 試験群が45～55例となる確率は約73%
    - ▶ 44/56やそれ以上に極端な場合が27%生じる

## ブロックランダム化 blocked ~

70

- ▶ 少数例のブロック内でランダムに割付
- ▶ 例：ブロックサイズが4
  - ▶ 4名での割付は以下の6パターン
 

○●●○	○●○●	●○○●
●●○○	●○●○	○●○○
  - ▶ 6パターンをさらに順に割付
 

●○○●	●○●○	●●○○	○○●●	...
------	------	------	------	-----
  - ▶ 人数のアンバランスは最大でも2名

## 予見可能性 predictability

71

- ▶ 次の患者の割付がわかってしまうこと
  - ▶ 非盲検・単盲検試験で懸念
- ▶ ブロックサイズが4で、最初2例が試験群
  - ▶ 次の割付は絶対に対照群
  - ▶ ブロックサイズを明かさない
    - ▶ 通常プロトコルに書かない
  - ▶ 可変ブロックサイズを用いる
    - ▶ ブロックサイズを2,4,6とすることもランダム化

## 予後因子（割付因子）の考慮

72

- ▶ 群間で予後因子のバランスがとれると、統計的効率がよい（検出力が高い）
- ▶ 層別ランダム化 stratified randomization
  - ▶ 層ごとにランダム化表を作成
- ▶ 最小化法 minimization
  - ▶ 全体的に因子のバランスを取るよう、バランスがとれるほうに高確率で割付
  - ▶ やや複雑なシステムを用意する必要

## 準ランダム化

73

- ▶ 基本的にすすめられない
- ▶ 生年月日、カルテID、受診日のような番号を利用
  - ▶ 末尾が同じ番号ならば同じ割付内容とわかる
- ▶ 交互にランダム化
  - ▶ 1つとばして同じ割付内容とわかる
  - ▶ 次の人は必ず違う割付内容であり、予見可能

## 割付表の管理、割付の実施

74

- ▶ 将来の割付が知られてしまえば、ランダム割付をする意義が失われる
  - ▶ 患者、施設の担当者から独立して管理
    - ▶ できれば解析責任者からも
- ▶ 中央割付（登録センター）の採用
  - ▶ 選択基準チェックの後、割付結果の返答
  - ▶ 電話、コンピュータシステム、Webシステムによる自動返答

## 盲検化

75

- 11a** 誰が（何が）盲検されたか
- ▶ 単盲検？二重盲検？
  - ▶ アウトカム評価を盲検？
    - ▶ PROBE(Prospective, Randomized, Open, Blinded-Endpoint)法
- 11b** 盲検した内容はバレなかったか
- ▶ プラセボは実薬とそっくりだったか

## 統計解析

76

- 12a** 主要/副次評価項目の解析方法
- ▶ 解析を再現できる記載
    - ▶ 統計解析計画書(SAP)は作った？
  - ▶ 信頼区間も表示すべき
- 12b** サブグループ解析や調整解析
- ▶ データを見る前に決めたか(pre-specified)、データを見た後に行ったか(post-hoc)
  - ▶ 割付因子は解析でも調整

## 対象者の流れ

77

- 13a** 対象者の流れ
- ▶ CONSORTダイアグラムをかく
- 13b** 対象者の追跡不能と除外の理由
- ▶ 群間のバランスが崩れると、解析結果にバイアスが生じうる

## 対象者の募集と追跡期間

78

### 14a 参加者の登録期間と追跡期間

### 14b 試験終了/中止理由

- ▶ 中間解析で中止した
- ▶ 資金面
- ▶ 他の研究グループからの報告により、試験遂行が非倫理的になった
- ▶ . . . etc.

## 対象者のベースライン特性

79

### 15 人口統計学的、臨床的な特性

- ▶ ランダム化した場合、群間比較のために検定することは無意味
  - ▶ 偶然によって起こること
- ▶ 予後因子のバランスが崩れた場合、その影響の強さとバランスの度合いを考慮

## 解析対象集団

80

### 16 解析された人数

- ▶ 各群での実際に解析された人数を明らかに
  - ▶ 分子分母ともに
  - ▶ 何らかの理由により、アウトカムが観察されなかった場合があるため

## FAS ? PPS ? SAS ? ①

81

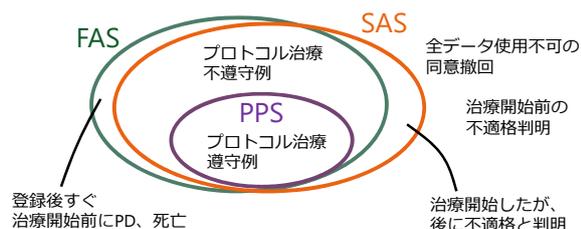
- ▶ Full Analysis Set (最大の解析対象集団)
  - ▶ 基本的に登録された対象者全員
  - ▶ 割付通りに解析する (ITTの原則)
  - ▶ 全データ使用不可の同意撤回や、追跡データなし、登録後不適格判明を除外する定義も
- ▶ Per Protocol Set (プロトコル遵守集団)
  - ▶ プロトコル治療を行った対象者
    - ▶ 試験によって定義はまちまち
  - ▶ 実際の投薬別に解析してもよい

## FAS ? PPS ? SAS ? ②

82

### ▶ Safety Analysis Set (安全性解析集団)

- ▶ プロトコル治療が行われた対象者全員
  - ▶ 基本的に登録後不適格判明も含める



## Intention To Treat (ITT) の原則

83

- ▶ 治療方針による効果は、実際に受けた治療ではなく、被験者を治療しようとした意図に基づくことにより最もよく評価できる

ICH-E9 (統計)ガイドライン

- ▶ 2群比較の優越性試験では、保守的 (群間差がつかない) 方向のみバイアスが入った結果が得られる
  - ▶ 規制当局からすれば望ましい方向のバイアス

## ITT原則の特徴

84

測定できない要因

ランダム割付 → 実際の治療 → 結果

- ▶ 割付後に、対象者の状態をみて、試験治療を行わなかった場合
  - ▶ 実際の治療内容で解析すると、治療効果を過大評価するおそれがある
  - ▶ 対照治療を受けても、割付通りの解析では、試験群として解析されるため、効果を過小評価

## ITT解析の解釈、問題点

85

- ▶ 治療遵守の程度を含めた、実践的医療技術の評価になるといわれる...
  - ▶ 臨床試験と実臨床は大きく異なる集団
  - ▶ Pragmatic 試験、市販後臨床試験実施が適切
- ▶ 非劣性試験では問題が生じることも
  - ▶ 全員が割付を無視して同じ治療を行えば、治療効果に差がない、ことを示せる
- ▶ ITT原則に従ったFAS集団とPPS集団の解析結果が大きく変わらないことを確認

## アウトカムの比較結果

86

### 17a 主要/副次評価項目の介入結果

- ▶ 検定だけでなく、推定(信頼区間)もあわせて

### 17b 二値アウトカムの場合、差と比の両方

- ▶ 相対指標：リスク比、オッズ比
- ▶ 絶対指標：リスク差

## サブグループ解析結果

87

### 18 サブグループ解析、調整解析

- ▶ Pre-specified か Post-hoc かを明らかに
- ▶ Forest Plot

## サブグループ解析の目的

88

- ▶ 効果の一様性を検討ならいい
  - ▶ どのサブグループでも同様の治療効果
- ▶ Post-hocにサブグループ解析
  - ▶ 都合のよい(有意な)結果を探すだけ?
    - ▶ 探索的の名の下、 $\alpha$ エラーは調整されない
  - ▶ それらしいが、無意味な

## 安全性

89

### 19 有害事象と意図しない効果

- ▶ 有害事象の報告が少ない
  - ▶ さらに試験治療による寄与も明らかでない
- ▶ CONSORT拡張版で有害事象報告におけるチェックリストを提供
  - ▶ Ioannidis JP, et al. *Ann Intern Med* 2004.
- ▶ 有効性と安全性両方の議論を行うべき

## 考察、その他

90

- 20 限界
- 21 一般化可能性
- 22 解釈
- 23 臨床試験登録番号と試験登録名
- 24 プロトコルの入手方法
  - ▶ 雑誌によってはプロトコルの提出は義務
- 25 資金提供者、薬剤提供者
  - ▶ 資金提供者の役割も
  - ▶ 利益相反(Conflice of Interest)とはまた別

## CONSORTに足りないもの

91

- ▶ IRBの承認
- ▶ 施設モニタリングがなされたか
- ▶ CONSORTはこれらを論文に盛り込むことを阻むものではない
  - ▶ 次回改訂で再度議論されるであろう

## 臨床試験論文を読む／書く上で

92

- ▶ CONSORTの説明論文を読む
  - ▶ BMJ 2010;340:c869
  - ▶ 文例まで載っており、論文を書く際に参考に
- ▶ 雑誌の特徴に注目
  - ▶ 商業誌か学会誌か、IF、論文の掲載時期
- ▶ それでも出版バイアスはなくなるらない
  - ▶ 結果の迅速な報告は義務となりつつある
  - ▶ 効きすぎた結果には疑いの目を

## まとめ①

93

- ▶ 臨床試験の質管理
  - ▶ データマネジメント、モニタリング
- ▶ 試験デザイン
  - ▶ 探索的か検証的か
  - ▶ 優越性、非劣性
- ▶ 臨床試験登録
  - ▶ 出版バイアスを減らす取り組み

## まとめ②

94

- ▶ 公正な出版のためのCONSORT
  - ▶ チェックリストとフローチャート
  - ▶ ランダム化のプロセスと盲検化
  - ▶ サンプルサイズ設計と検定
    - ▶ 中間解析の実施
  - ▶ 解析対象集団、ITTの原則

## 臨床試験を実施するなら①

95

- ▶ プロトコルの作成くらい自分たちで・・・
  - ▶ 統計家の参画はやや浸透してきた！？
  - ▶ モニタリング(+監査)、データマネジメント、プロジェクトマネジメントの視点は？
  - ▶ プロトコルとSOP類さえあれば、誰が試験を実施しても同様の結果が得られる状態？
- ▶ あとは対象者を集めて評価するだけ？
  - ▶ 事務局機能、モニタリングの実施
    - ▶ 試験終了後に重大な不備が見つかっても後の祭り

## 臨床試験を実施するなら②

96

- ▶ 臨床試験支援機関を利用しましょう
  - ▶ 北大なら臨床研究開発センター