

9. 生物統計・ データマネジメント

国立研究開発法人 日本医療研究開発機構
臨床研究・治験推進研究事業
先端医療開発を担う人材養成のための
標準化教育プログラムの策定と実践研究班

内 容

①生物統計



- i. 生物統計の概要
- ii. 臨床試験準備
- iii. 臨床試験実施と終了

②データマネジメント



- i. データマネジメントの概要
- ii. 臨床試験準備
- iii. 臨床試験実施と終了

① 生物統計

i. 生物統計の概要



臨床試験と生物統計の関わり

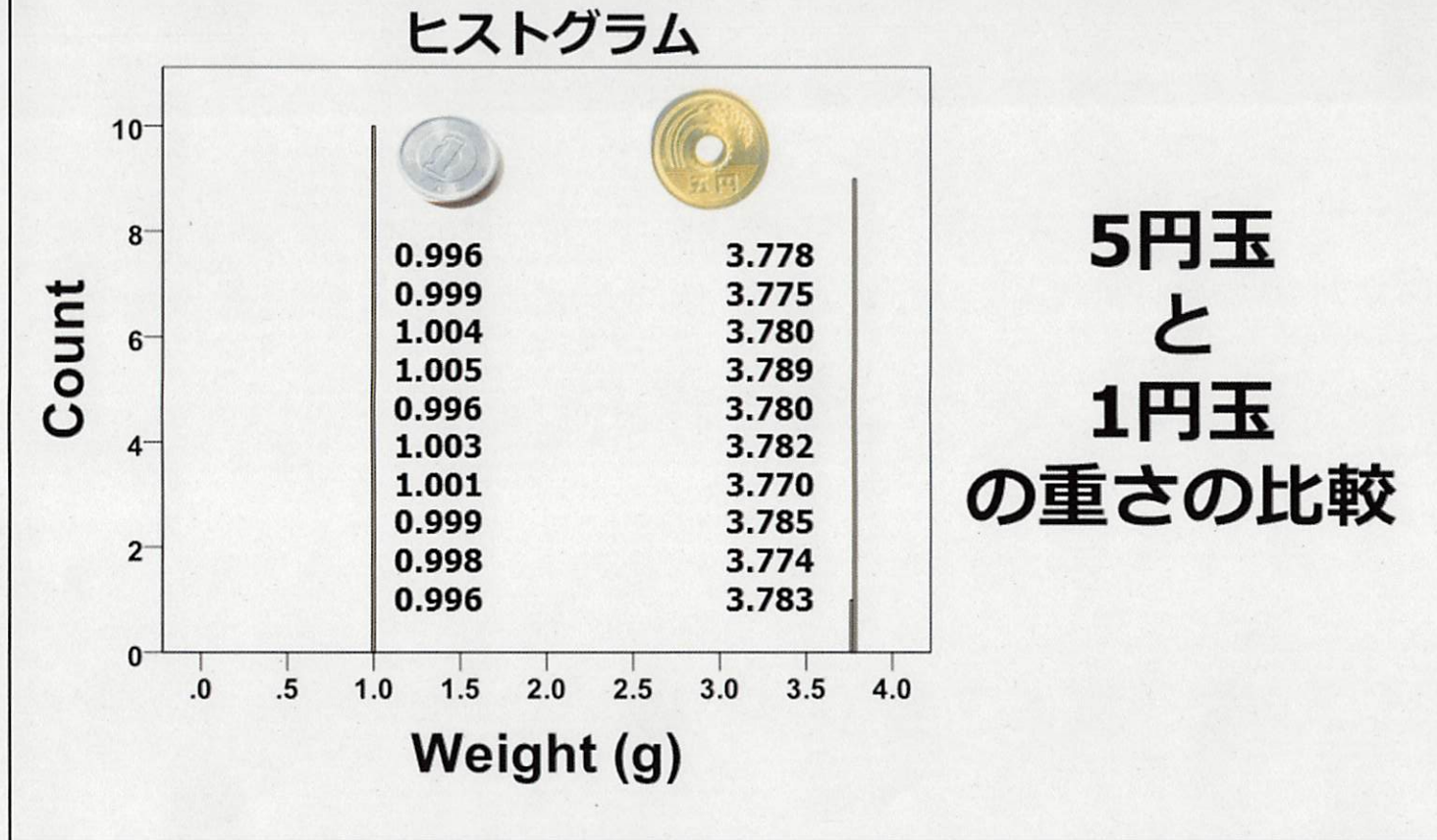
生物統計とは：

生物統計学 (Biostatistics) とは、**基礎・臨床・疫学**といった医学研究において、**どうデータをとるか** (研究計画・実験計画)、**どう解析するか** (統計解析) の方法論を提供する応用統計学です。

新薬あるいはより優れた治療法を開発するために行われる臨床試験においては、生物統計家の参加が必須ですが、わが国は極端に少ないのが現況です。

東京大学大学院 医学系研究科 公共健康医学専攻 生物統計学分野 HPより
<http://www.epistat.m.u-tokyo.ac.jp/about/>

なぜ統計が必要か？



はじめに、なぜ統計、正しくは「生物統計」が必要か、というのを考えるために、まずは逆に、あまり「必要でない」例から考えてみたいと思います。

例えばここに1円玉と5円玉があり、10枚ずつ重さを測定し比較するとします。背景にあるのはヒストグラムですが、バラつきが小さいためほとんど一本の線のようにです。平均はそれぞれ1.0, 3.78, 標準偏差はいずれも0.005前後と非常に小さいです。

なぜ統計が必要か？

ヒストグラム



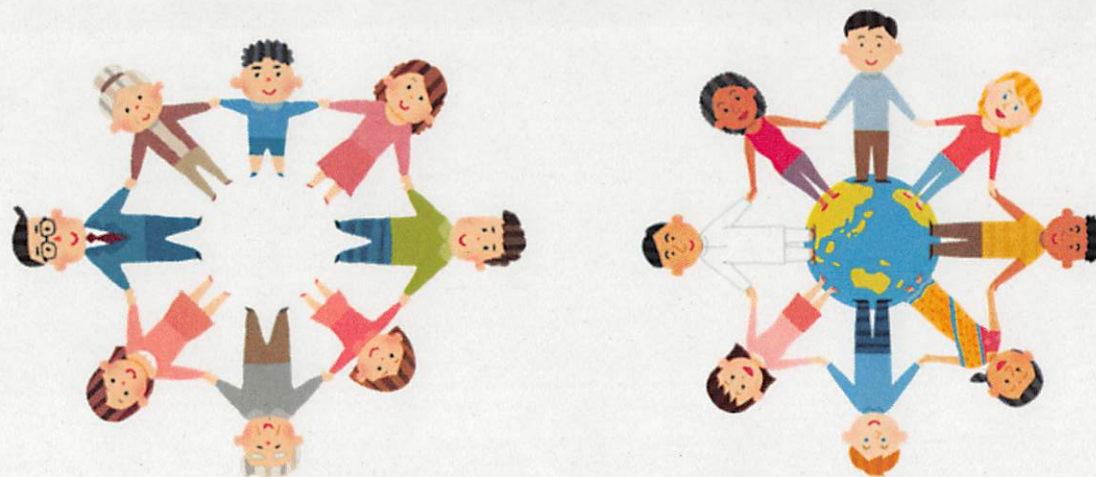
**5円玉
と
1円玉
の重さの比較**

**バラつきが非常に小さい
わざわざ10個ずつ調べる必要もない**

このように、規格が統一されている工業製品のような場合には、ほとんどのサンプルの値が平均値近くに集まり、バラつきが非常に小さいためわざわざ10個ずつ調べる必要もないと言えます。こういった場合には、統計というのは意識されにくいかと思います。

なぜ統計が必要か？

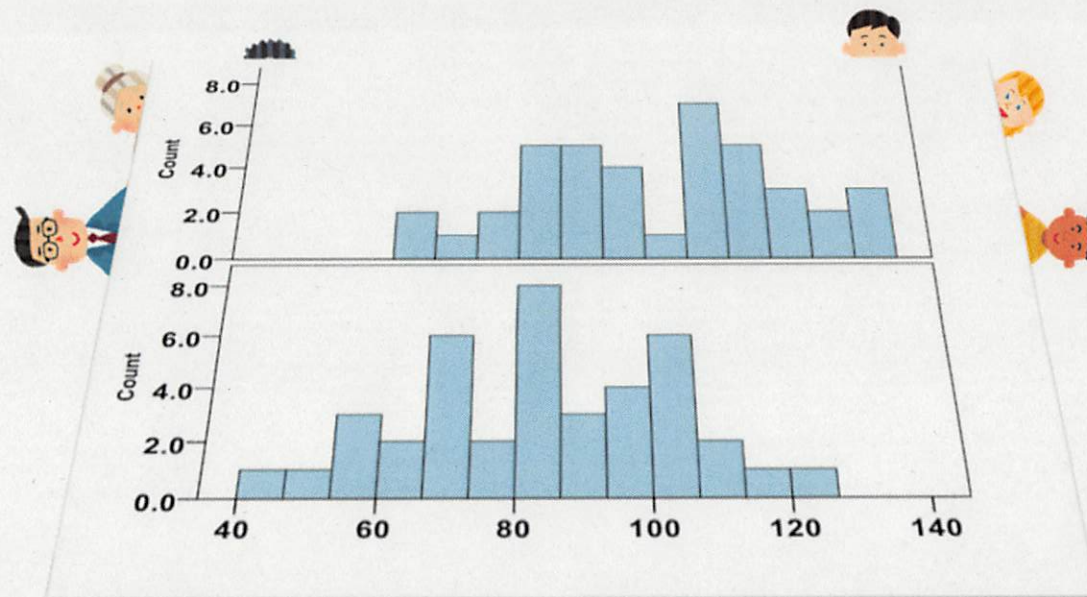
ヒトを対象した研究



比較群間が均一でない
測定誤差・個体差など
による**バラつきが大きい**

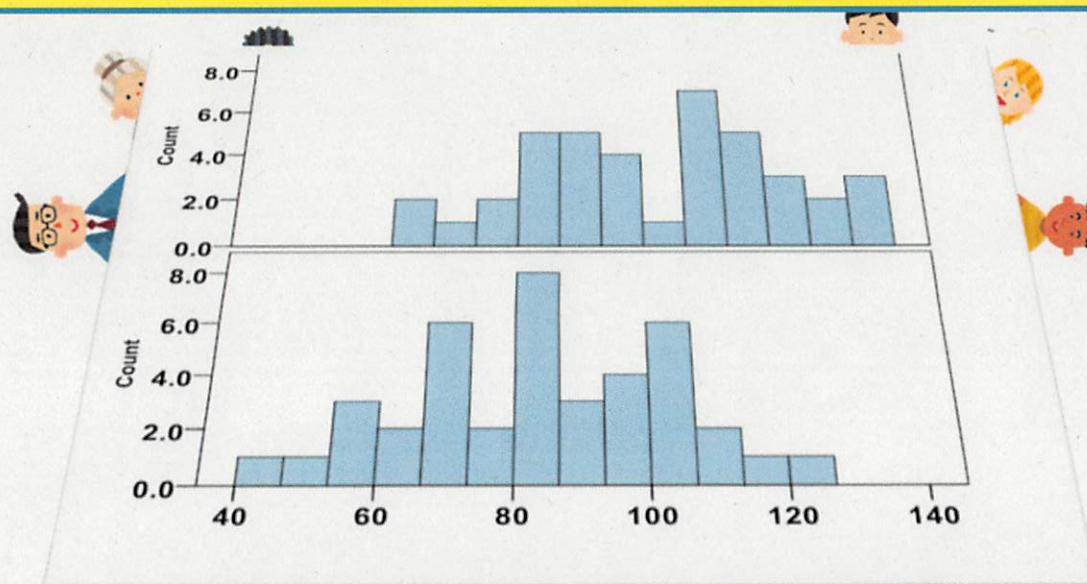
ところが、対象がヒトとなるとそうはいかず、ランダム化しない限り比較群間にバイアスが存在し、測定誤差や個体差などのバラつきも大きい状況となります。

なぜ統計が必要か？



ヒストグラムで示しますとこういったイメージとなり、1円玉と5円玉の場合とは大きく異なります。

**1円玉と5円玉の重さの分布が
このような状況だったら・・・？
どっちが重いかはすぐには言えない**



**バラつきの大きいデータを
扱うのも生物統計の特徴**

ヒストグラムで示しますとこういったイメージとなり、1円玉と5円玉の場合とは大きく異なります。

なぜ統計が必要か？

さらに・・・



得られた結果が介入によるものなのか、
「交絡」によるものなのかわからない場合
その新薬は効いたのか？選択バイアスじゃないのか？

効く群と効かない群＝「交互作用」の存在は？

アウトカムが暴露による硬貨なのか交絡によるものなのかが判別できず、因果関係の特定も難しい、あるいは交互作用が存在する場合があります。

なぜ統計が必要か？

真実は見たままじゃない
だから統計

ごつに・・・

介入群が際立って健康
だったりしないか？

特定の遺伝子変異にのみ
効果のある薬剤

得られた結果が介入に
「交絡」によるものなのか
その新薬は効いたのか？選択

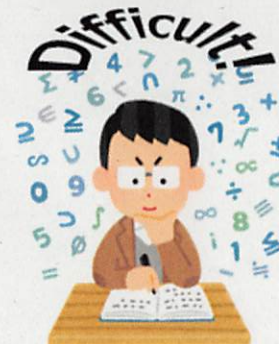
のなのか、
わからない場合
アスじゃないのか？

効く群と効かない群＝「交互作用」の存在は？

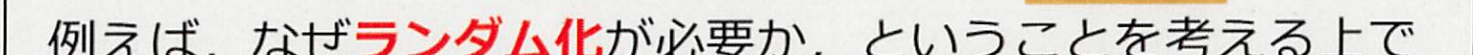
アウトカムが暴露による硬貨なのか交絡によるものなのかが判別できず、因果関係の特定も難しい、あるいは交互作用が存在する場合があります。

臨床試験と生物統計の関わり

なぜ生物統計が重要か



ある介入と結果における因果関係を検討する上で、**適切な試験デザインの下、適切なデータ解析**およびその**解釈**が必要となります。しかし、この一連の業務は、交絡や交互作用、推定精度といった**統計学的概念を十分に理解していなければ不可能**なのです。

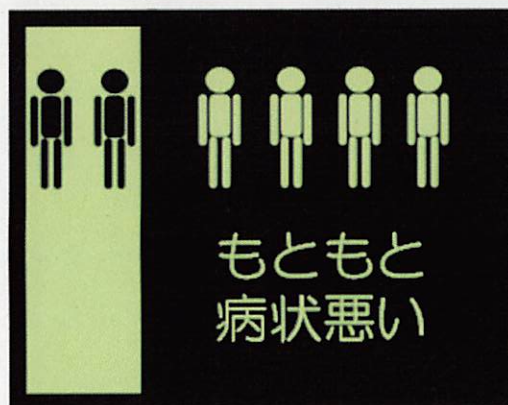


ランダム化と交絡

好きなように治療を選ぶと・・・

参加者の試験前の特徴

新しい治療



(従来の) 標準治療



「病状が悪いので新しい治療を試したい」
「病状がそう悪くないので従来の治療で」
不公平な比較になってしまう
(どちらに偏るかは状況による)

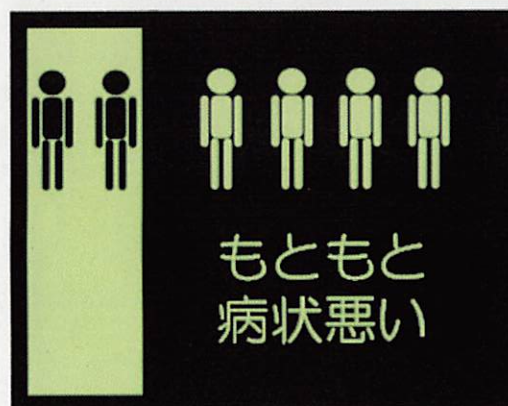
ニ、ガムルレ六效

新しい治療に状態の悪い症例が集まり本来の効果
が見えなくなる恐れ

好きなように治療を選ばと・・・

参加者の試験前の特徴

新しい治療

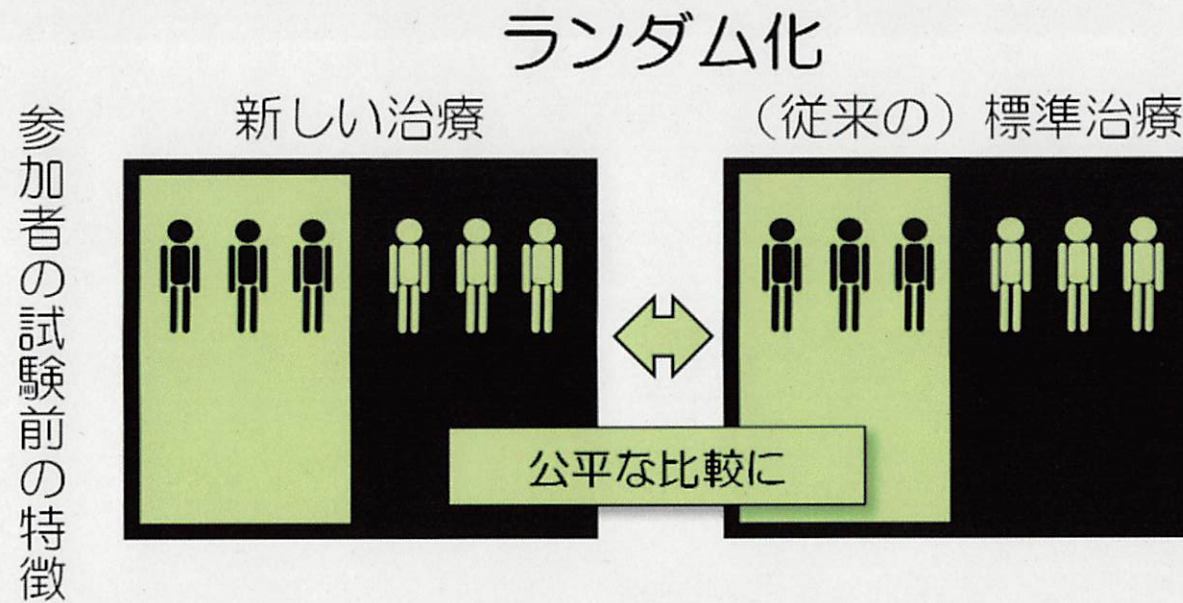


(従来の) 標準治療



「病状が悪いので新しい治療を試したい」
「病状がそう悪くないので従来の治療で」
不公平な比較になってしまう
(どちらに偏るかは状況による)

ランダム化と交絡



比較可能性が高まる

生物統計の誤用

検定や多変量解析などの統計学的手法では、**結果変数の型やデータ構造によって用いられる手法が異なります。**

誤用の例

- カテゴリカル変数である二値データ（生存・死亡）などを結果変数とした検討に対しt検定を用いる
- 打ち切りを考慮せずに生存時間の比較を行う
- 統計学的多重性を考慮せず、繰り返し同じ有意水準による検定を実施する
- ランダム化試験ではないのに交絡を考慮せず因果関係として結果解釈する

十分な知識がないのに統計学的検討を実施することは、誤った結論を導いてしまう原因になります。



生物統計の誤用

検定や多変量解析などの統計学的手法では、**結果変数の型やデータ構造によって用いられる手法が異なります。**

誤)

**医学研究における統計の誤りは、
生命の危険に直結する場合があります
重要なデータについては慎重な吟味が必要です**

■ 検定を実施する

- ランダム化試験ではないのに交絡を考慮せず因果関係として結果解釈する

十分な知識がないのに統計学的検討を実施することは、誤った結論を導いてしまう原因になります。



どんな解析が必要なの？

目 的		変数の型		
		二値	連続値	生存時間
2群比較	対応なし	χ^2 検定 Fisherの正確確率検定	t検定 Mann-Whitney検定	Log-rank検定
	対応あり	McNemar検定	対応のあるt検定 Wilcoxonの 符号付き順位和検定	
多群比較		χ^2 検定 Fisherの正確確率検定	一元配置分散分析 Kruskal-Wallis検定	Log-rank検定
層別解析		Mantel-Haenszel 検定	二元・多元配置 分散分析	層別Log-rank
多変量解析		ロジスティック回帰	一般線形モデル 重回帰分析	Cox回帰
相関解析		カッパ係数など	Pearsonの相関 Spearmanの相関	

生物統計の悪用

統計ソフトを利用すると、比較的簡単に「それらしい結果」が得られてしまいます。そこに、都合の良い解釈を加えて発表すると問題となるわけです。

俗に「チャンピオンデータ」と呼ばれる、自らの学説に一致したデータのみを採用し、結果をまとめることはしばしば行われていましたが（よく言及されるのはメンデル）、かつては研究者側だけでなく読み手の統計学への理解も乏しかったでしょうし、説得力を持たせる上でやむを得なかったこともあったかもしれません。

しかし現在は違います。



生物統計の悪用

COIがある場合など，研究者側に不正を行う動機づけが存在する場合には特に注意してデータを扱っていることを証明しなくてははいけません．

なお，研究不正には意図したもの以外も含まれます！



論文を読むために

- サンプルサイズが推定精度・および検定結果の解釈に及ぼす影響の理解
- 想定される母集団と論文において一般化を行っている対象集団の一致度
- 交絡・交互作用の検討は十分か
- アウトカムの設定は適切か
- 考えられるバイアスは何か
- 収集された情報は、それが意図している内容を適切にあらわしているか



論文を読むために

- サンプルサイズが推定精度・および検定結果の解釈に及ぼす影響の理解
- 想定される母集団と論文において一般化を行っている対象集団の一致度
- 交絡・交互作用の検討は十分か
- アウトカムの設定は適切か
- 考えられるバイアスは何か
- 収集された情報は、それが意図している内容を適切にあらわしているか

CONSORT (臨床試験)

STROBE (観察研究)

といったガイドラインに沿って読み込むのも良いでしょう



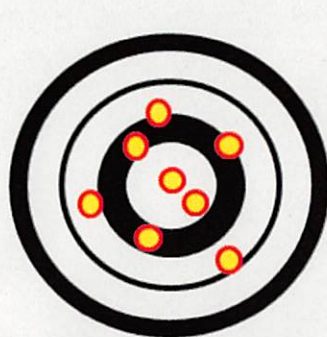
医学研究における 2つのエラー

医学研究における2つのエラー／誤差
偶然誤差-Precision と 系統誤差-Accuracy
このうち、系統誤差の方が
「バイアス」である。



医学研究における 2つのエラー

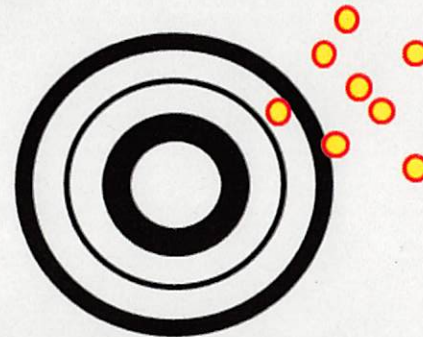
医学研究における2つのエラー／誤差
偶然誤差-Precision と 系統誤差-Accuracy
このうち、系統誤差の方が
「バイアス」である。



いい感じ



偶然誤差
(信頼性・精度)



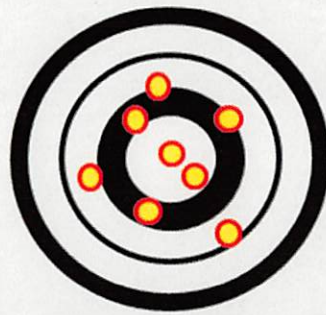
系統誤差
(妥当性)

医学研究における 2つのエラー

● 1つは1つの「研究」
だと考える

偶然誤差 (Precision) と 系統誤差-Accuracy

このうち、系統誤差の方が
「バイアス」である。



いい感じ



偶然誤差
(信頼性・精度)



系統誤差
(妥当性)

バイアスを考える上で重要な概念

選択バイアス

交 絡

情報バイアス

選択バイアス Selection bias

対象者を選択する 過程から発生するバイアス

危険な職場で働く労働者と一般集団を比較したいが...

- 危険な職場の労働者は、若く体力十分なことが多い
(多少の曝露もへっちゃら?)
- 一般集団には病気で働けない人もいる

→ かって危険な職場の労働者が良い結果に

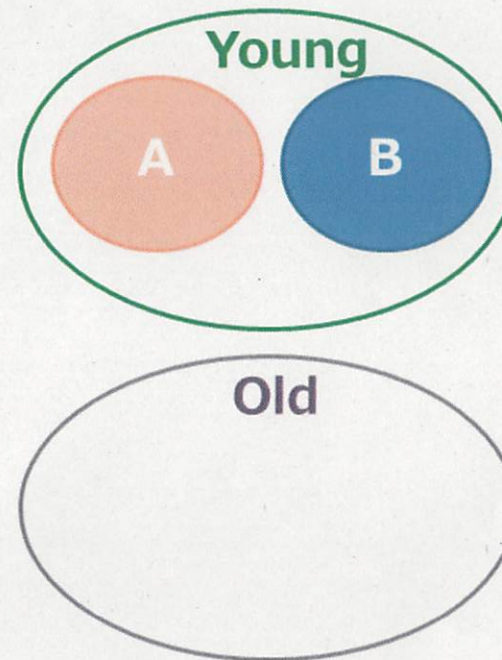
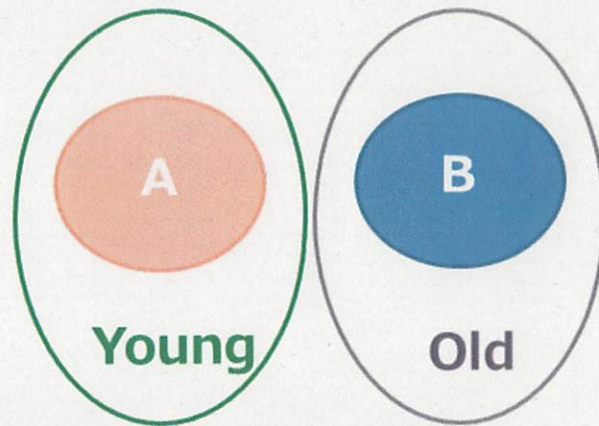
イメージ



健康労働者効果
Healthy worker effect

選択バイアス Selection bias

2つの視点



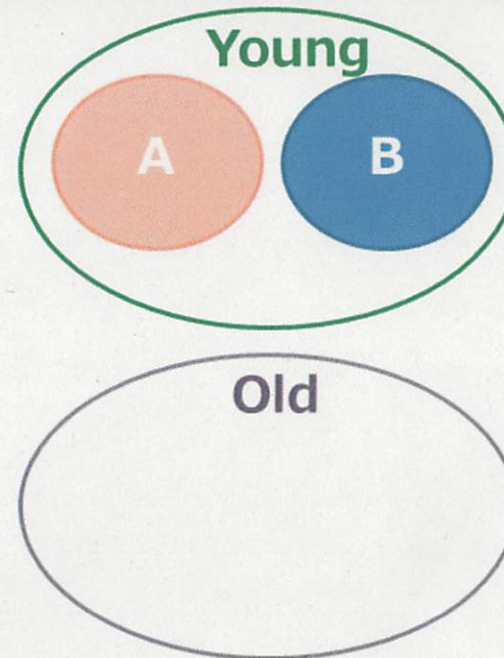
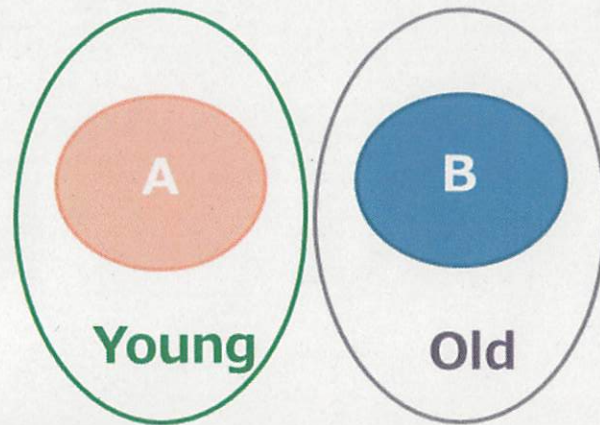
群間で母集団が異なっている
交絡を生じる

選択バイアス

Selection

研究対象者全体にバイアスが見られる。若年者のデータが高齢者に適用できるかどうか

2. **交互作用**がある場合には問題



群間で母集団が異なっている
交絡を生じる

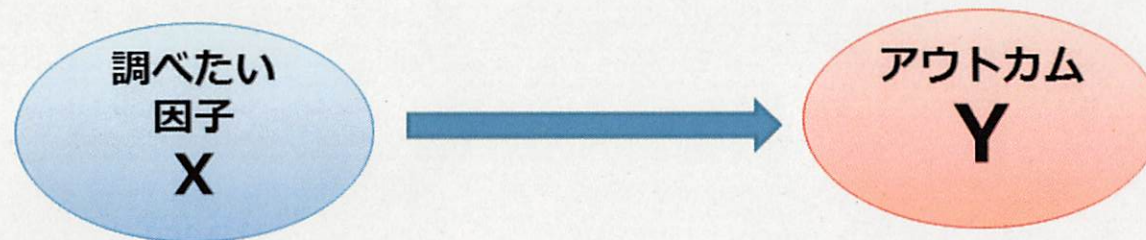
交 絡 Confounding

興味のある因子が他の危険因子と
関連があるために生じる

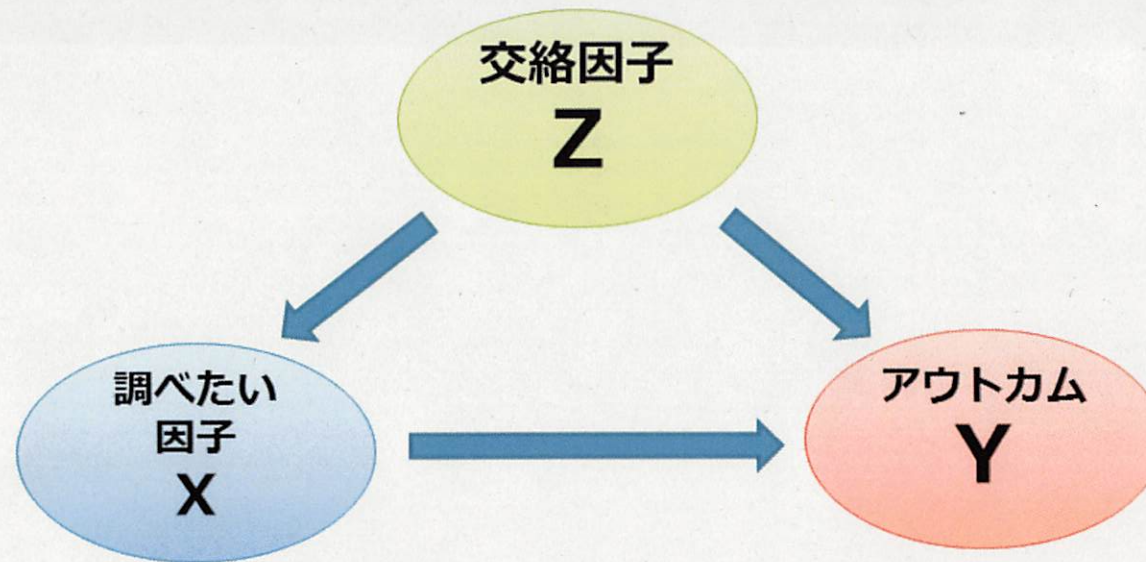
本当は関連がないのに
関連があるように見えてしまう！

例) とある国ではコーヒー好きには肺癌が多い
→ コーヒー好きはタバコも好きだった！

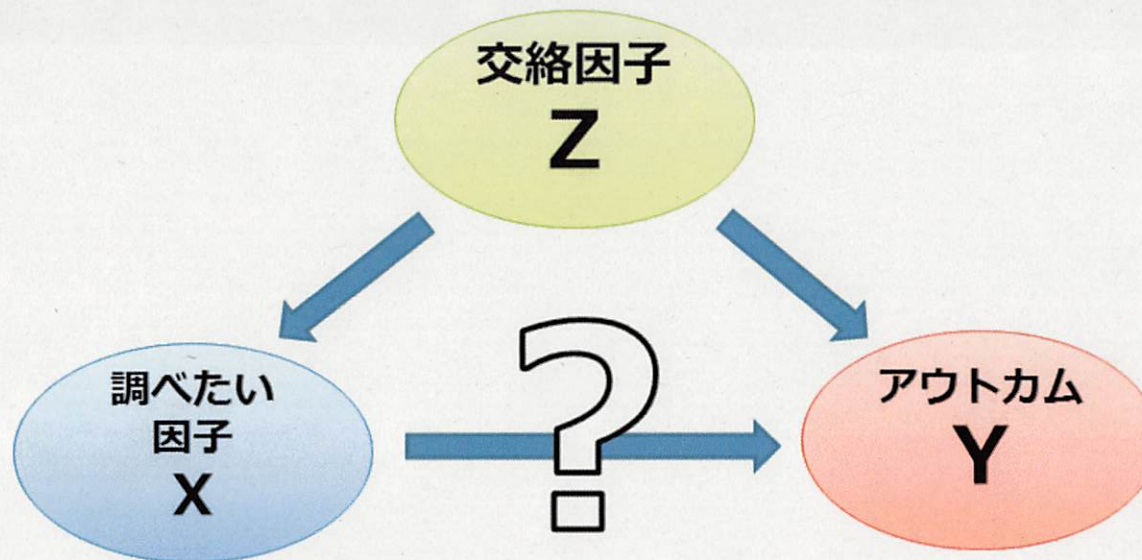
交 絡 Confounding



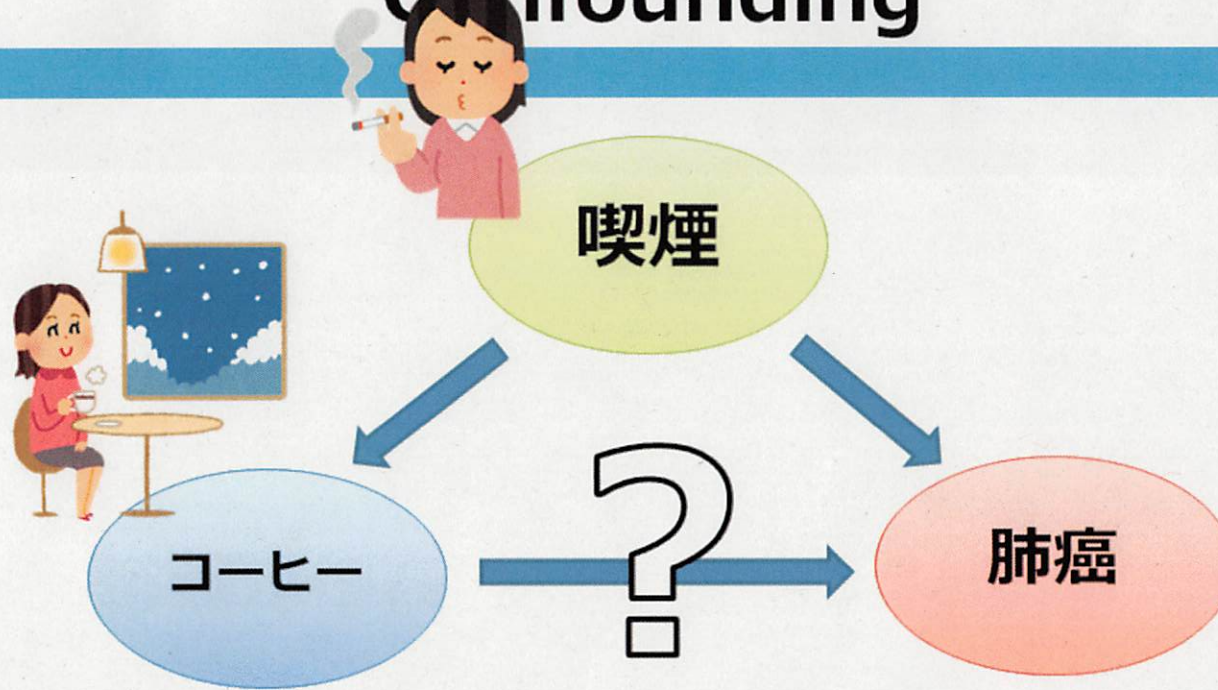
交絡 Confounding



交絡 Confounding



交絡 Confounding



交絡 Confounding



交絡の制御

Controlling for confounders

デザインの上では大きく 3 つ

■ ランダム化 (Randomization)

→ 「臨床試験」で有名. 介入研究にしか使用できず

■ 限定化 (restriction)

→ 観察研究でも介入研究でも使用できる

■ マッチング (matching)

→ 性・年齢をマッチングしてこれらの交絡を制御

交絡の制御

Controlling for confounders

解析の上では大きく 2 つ

■ 層化 (stratification)

→ 層ごとの効果を確認する。先ほどのダウン症のデータのような感じ。

■ 回帰モデル (regression)

→ いわゆる多変量解析 (おなじみのロジスティック回帰や Cox 回帰)

情報バイアス Information bias

研究対象から得られた
情報が**誤っている**場合に発生

■ 差別的誤分類

- 比較する群間で程度が異なっている
「思い出しバイアス」「追跡の偏り」など
医師や患者の先入観, COI等の問題も.

盲検化が必要な理由となる

■ 無差別的誤分類

- 比較する群間で程度が同じ

情報バイアス Information bias

無差別的誤分類

両方の群で同じようにバイアスがある場合

例 1) 肺気腫の例において、曝露群でも非曝露群でも診断がされにくいような場合（逆もあり）

→ 診断基準がきつい、画像検査の質が悪い

情報バイアス Information bias

例2) 心疾患と高脂肪性食餌についての仮想的な症例対照研究

正確な分類をした場合

無差別的誤分類をした場合

	高脂肪性食餌		高脂肪性食餌		高脂肪性食餌	
	なし	あり	なし	あり	なし	あり
心疾患群	450	250	360	340	410	290
対照群	900	100	720	280	740	260
オッズ比	5		2.4		2.0	

本当はもっと関係があるのに効果が薄まってしまう

ii. 臨床試験準備

エンドポイント

評価される項目を指す

さまざまな「アウトカム」の中からいくつかの「エンドポイント」を設定します

例えば、癌の第III相臨床試験においては、全生存期間、無増悪生存期間、無再発生存期間などのほか、無〇〇生存率、無〇〇期間など、さまざまなアウトカムがあります。この中から、当該試験において第一目的となるアウトカム、また検証の対象とするアウトカムを「主要評価項目」あるいは「プライマリーエンドポイント」といいます。通常ただ1つ設定します。

セカンダリーエンドポイントは「副次評価項目」ともいい、通常複数設定されます。探索的な項目と考えられ、

エンドポイント

評価さ
さまざ
ト」を

多重性の調整等により第一種の過誤を，例数設定により第二種の過誤を十分にコントロールしている．多重性の調整により通常検出力が低下するため，できるだけ1つか少ない数にとどめるのが望ましい．

例えば
生存期間，

など，さまざまなアウトカムがあります．

において第一目的となるアウトカム，また検証の対象とするアウトカムを「主要評価項目」あるいは「プライマリーエンドポイント」といいます．通常ただ1つ設定します．

セカンダリーエンドポイントは「副次評価項目」ともいい，通常複数設定されます．探索的な項目と考えられ，検証よりも新規的知見の発見を目指します．

悪生

期間

エンドポイント

True endpoint:

がん臨床試験など致死的な疾患には死亡のみをイベントとした全生存期間など

「**本当に改善すべき**」指標

Surrogate endpoint:

全生存期間ではイベント数が少なくなり評価が困難となる場合には再発、増悪などもイベントとしたアウトカムが利用される
さらに、腫瘍マーカーなどのバイオマーカーが利用される場合もある

本当に改善すべき指標と**強い相関**を持つ指標
連続値のマーカーなどは差を検出しやすい



さまざまなアウトカム

安全性：

有害事象発現割合

- 単一の項目を主要評価項目として取り上げることもある

重篤な有害事象発現割合

有効性：

奏効割合 - 抗がん剤の評価など

全生存期間 - 死亡のみをイベントとし生存期間を比較

無増悪生存期間 - 死亡と増悪をイベントとする

Event-free survival - 死亡と予め定めたイベントを用いる

サロゲートとなるバイオマーカーの平均値
(打ち切りのあるデータに注意)

などが代表的

試験デザインに関連する概念

■ 観察研究（コホート研究）

割付を実施せず，日常臨床をベースとした研究．臨床試験には該当しない

■ 介入研究

割付を実施する場合，また，単群であっても新薬等，日常臨床の範疇にない診療が行われている場合に介入研究と呼ばれる．臨床試験はほぼ全て「介入研究」である．

試験デザインに関連する概念

■ 平行群間比較試験

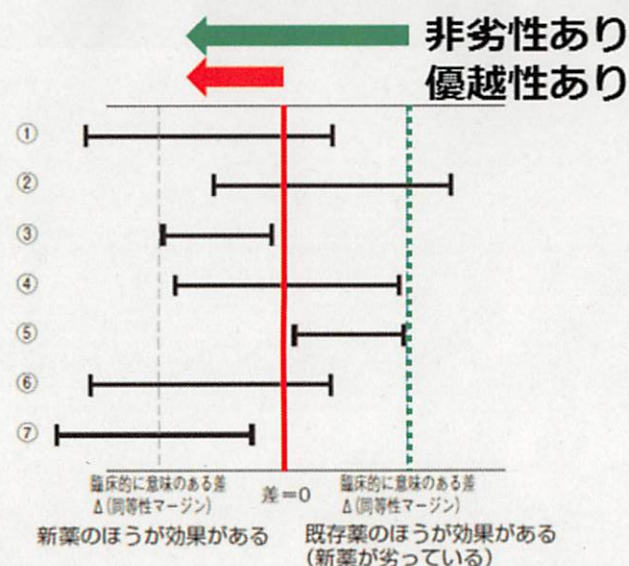
臨床試験においてはランダム化は必須（ランダム化されていない場合にはバイアスのため解釈困難）。異なる介入を実施した2つ以上の集団間のエンドポイントを差や比の指標を用いて比較する。

■ クロスオーバー研究

同一の集団（コホート）を対象とし、時期をずらして異なる介入を実施する。被験者間の背景要因の違いによる交絡はなくなるが、持ち越し効果など一般に考えられる以上に注意すべき点が多い

試験デザインに関連する概念

■ 優越性試験・非劣性試験



優越性試験とは差の指標が0でない、比の指標が1でないことを帰無仮説として実施される試験である。非劣性試験の場合には、帰無仮説に非劣性マージンと呼ばれる、臨床的に同等とみなすことが可能な許容範囲を含める。

非劣性マージンの設定は
リーズナブル??

http://www.igaku-shoin.co.jp/paperDetail.do?id=PA02971_04

サンプルサイズ設定

サンプルサイズが大きくなると、偶然誤差が縮小するため、「有意」になりやすくなる。

「**検出力80%**」などと良く言われるが、これは、アウトカムとして想定される状況（例えば、薬剤Aの奏効割合が70%、Bが40%など）の下で有意差が得られる確率を示したものである。サンプルサイズが大きいと検出力が上がる。

サンプルサイズの設定には、**用いられる指標および各群で想定される結果、検定方法などの情報**が必要となる。

臨床試験においては、サンプルサイズは**小さすぎても大きすぎても非倫理的**である

臨床試験の相とデザイン

■ 第Ⅰ相試験

- 安全性・投与用量の決定 (N=10~)

■ 第Ⅱ相試験

- 安全性・効果の推定 (N≒50)

■ 第Ⅲ相試験

- 効果の検証 (N=100~)

臨床試験の相とデザイン

新薬であれば順番に進む

- 
- 第Ⅰ相試験
 - 安全性・投与量の決定 (N=10~)
 - 第Ⅱ相試験
 - 安全性・効果の推定 (N=50)
 - 第Ⅲ相試験
 - 効果の検証 (N=100~)

臨床試験の相とデザイン

新薬であれば順番に進む

■ 第Ⅰ相試験

- 安全性

既存治療の比較であればⅡ・Ⅲ相から

■ 第Ⅱ相試験

- 安全性・効果の推定 (N≒50)

■ 第Ⅲ相試験

- 効果の検証 (N = 100~)

第Ⅰ相試験

■ 第Ⅰ相試験

- 安全性・投与用量の決定

3+3コホート

用
量

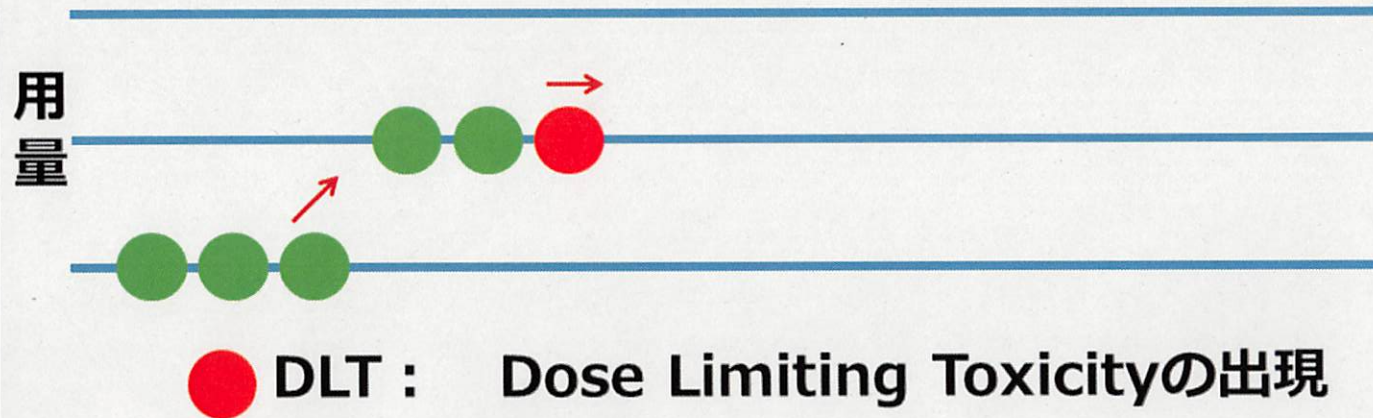


■ 当初用量はマウスの実験結果やこれまでの類似薬の結果から

第 I 相試験

■ 第 I 相試験

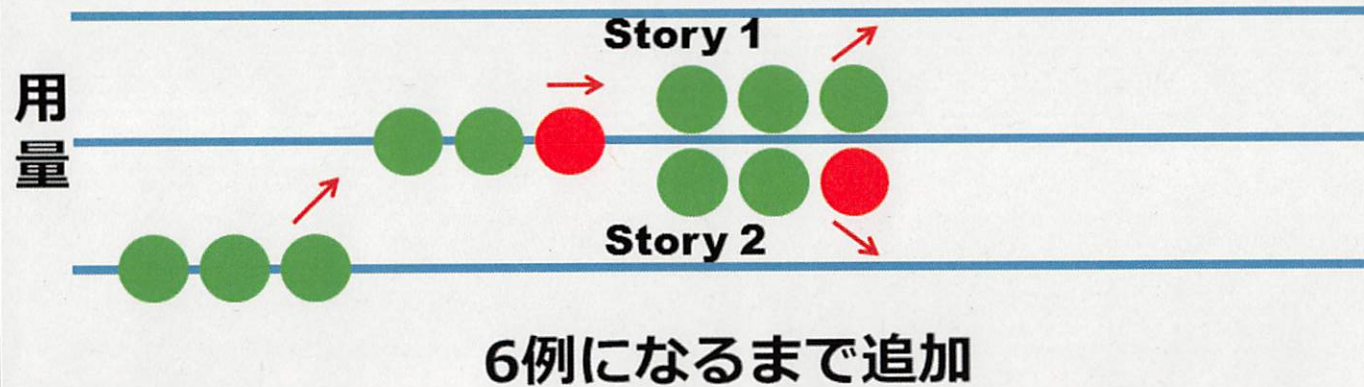
- 安全性・投与用量の決定



第Ⅰ相試験

■ 第Ⅰ相試験

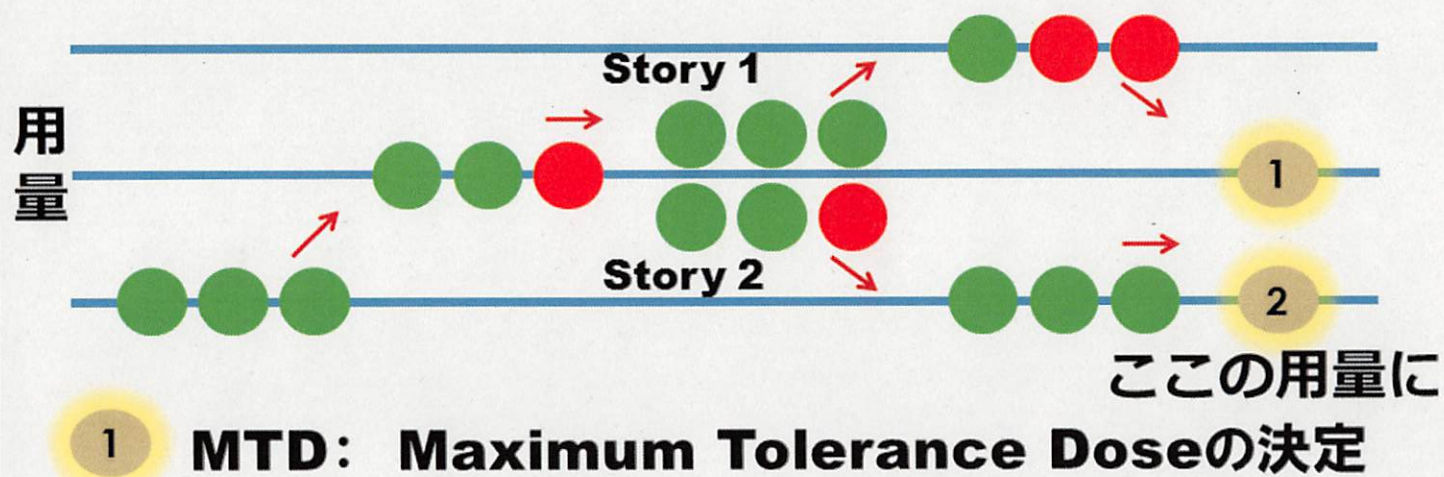
- 安全性・投与用量の決定



第Ⅰ相試験

■ 第Ⅰ相試験

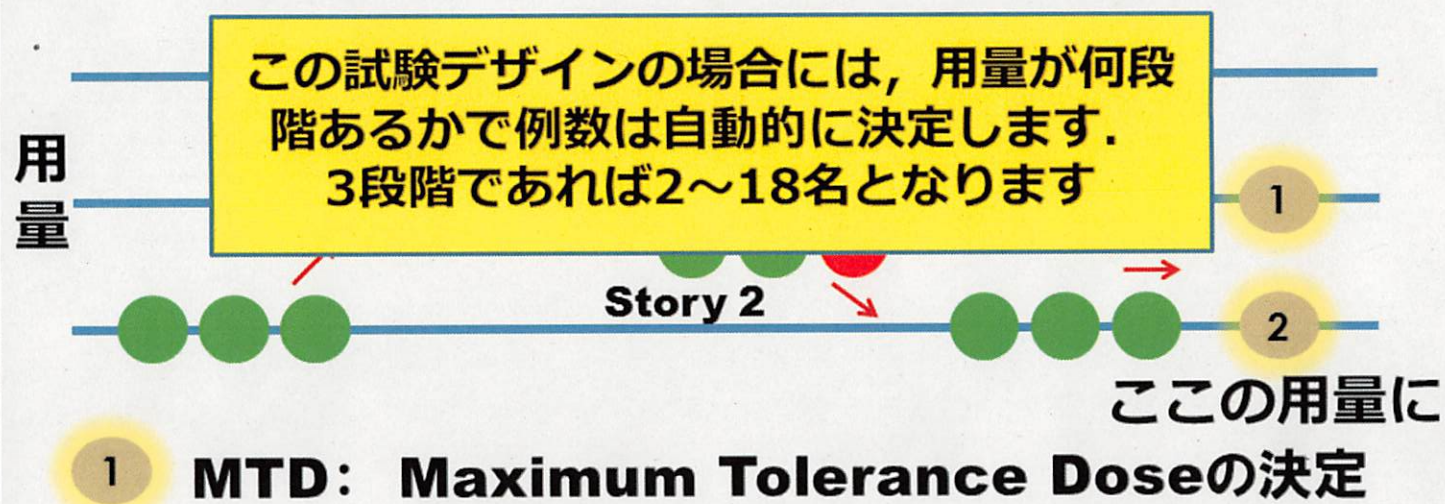
- 安全性・投与用量の決定



第 I 相試験

■ 第 I 相試験

- 安全性・投与用量の決定



第Ⅱ相試験

■ 第Ⅱ相試験

- 安全性・効果の推定

- 安全性が（ある程度）認められた投与用量が決定し、50例前後の症例に対し投与 -> 単群・複数群
奏効割合・有害事象の出現割合などを信頼区間などとともに推定

【奏効割合】 = CR + PR + SDの割合

CR : 完全寛解

PR : 部分寛解

SD : 安定 - Stable disease

第Ⅱ相試験

■ 第Ⅱ相試験

- 安全性・効果の推定

- 安全性が（ある程度）認められた投与用量が決定し、50例前後の症例に対し投与 -> 単群・複数群
奏効割合・有害事象の出現割合などを信頼区間などとともに推定

大規模で費用もかさむ第III相で
「いけるかどうか」を決める

【奏効

CR：完全寛解

PR：部分寛解

SD：安定 - Stable disease

第Ⅲ相試験

■ 第Ⅲ相試験 - 効果の検証

- 「本当に重要な指標に対して効果があるか」を
「ランダム化比較試験」で検証

重要な指標 -> 生存率, 無再発生存率
といった指標
時に盲検化, 二重盲検化など

ランダム化

ランダム化にはさまざまな手法がある

要相談

■ 単純ランダム化

完全に乱数で被験者登録ごとに割付。例数が少ないと偏る可能性が高い

■ ブロックランダム化

ある例数ごと（4例など）に被験者を区切り、そのブロックごとに割付数が一定となるように割付

■ 層別ランダム化

割付を男女、60歳以上/未満など、層ごとに行い、層別因子の分布が群間で異ならないようにする

ランダム化

ランダム化にはさまざまな手法がある

要相談

■ 適応的ランダム化

それまでに割付されている被験者の情報を基に、割付確率を変化させる方法（バイアスコイン法，最小化法など）

■ クラスターランダム化

施設や地域など，個人よりも大きな単位で割付を実施

盲検化

情報バイアスを防ぐため、**単盲検**、**二重盲検**（被験者のみ、又は被験者と医師）といった盲検化の手続きが用いられる。

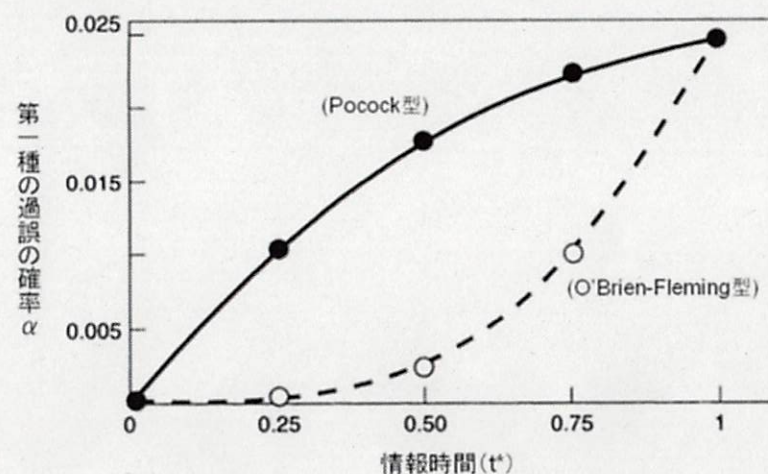
被験者においては、知らされた割付情報が自覚症状に影響を与えたり（**プラセボ効果**等）、医師においては、同様に有害事象や診断（再発・増悪等）に影響を与えたりすると**情報バイアス**が生まれ、結果にバイアスが生じる。

盲検化は広く知られているが、盲検を遵守するにはさまざまな実施上の工夫が必要となる。製剤の形状における工夫や、手順の工夫、情報の遮断などである。盲検が解除されている研究協力者も必要となるが、秘密の保持が重要となる。

中間解析計画

統計学的多重性を考慮し、極めて高度な手法によりデザインされる。現在、Lan & DeMetz の α 消費関数を利用する方法が一般的（近年の治験などはほとんどがこの方法で行われている）。

生物統計家への相談なしには実施不可能であるため、要相談となる。



中間解析計画

統計学的多重性を考慮し、極めて高度な手法によりデザインされる。現在、Lan & DeMetz の α 消費関数を利用する方法が一般的（近年の治験などはほとんどがこの方法で行われている）。

生 統計学的多重性：

な

通常検定では、5%の確率で偽陽性（本来差がないのに有意になる）がでるように設定されているため、検定を繰り返すと1度は有意になる確率はどんどん上昇してしまう。このため、検定を繰り返すような状況では各検定の有意水準を5%よりも低くする必要がある。

→この計算が難しい

情報時間 (t^*)

統計解析計画書

とくに第III相などの検証的試験では、**あらかじめ定められた解析計画に沿って解析**しなければ、統計学的多重性の問題により第一種の過誤が増大するため、「検証」とみなされなくなる場合がある（「チャンピオンデータ」の話に関連）。

そこで、プライマリーエンドポイントについては試験開始前に形跡方法を入念に計画ことが重要となる。完全に確定するのはデータ固定前となる。盲検の場合にはキーオープン前に盲検下レビューを実施し、最終的な解析計画を決定する。2つ以上のプライマリーエンドポイントが設定される場合には、Bonferroniの補正などで有意水準を調整することが多い。

統計解析計画書

セカンダリーエンドポイントについてはその限りではなく、探索的であることを理由に多重性の調整等は免除され、事後の追加解析計画も容認されるが、バイアスへの配慮や解析自体の妥当性は適切に検討されなければならない。

予定される統計解析計画は、全て計画書に記載する必要がある。探索的で事後的に実施された検討であれば、そういったことが予想される点や、そのことによるメリット・デメリットなどを記す。

iii. 臨床試験実施と終了

解 析

中間解析は、先述の通り高度な手法を利用するため、必ず統計家の関与が必要である。場合によっては、試験デザインや最終解析を担当する統計解析責任者とは別に担当者を置く場合もある。

臨床試験における統計解析は、統計解析計画書に沿ったものでなくてはならない。利用されるソフトウェアは、SASが従来からスタンダードとして利用されているが、JMP, SPSS, STATA, Rなども統計ソフトとして広く利用されており、問題なく利用可能である。

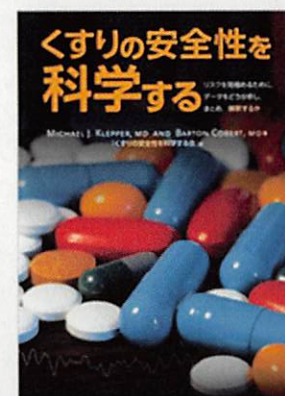
最終的な解析は、データ固定、キーオープンなどの手続きの後に実施される。

有害事象の報告

報告の仕方により集計方法が異なるため意外とややこしい。
以下のような集計が実施される。

- 1度でも有害事象を発現した症例の割合
 - 1度でも重篤な有害事象を発現した症例の割合
 - 1度でもGrade 3以上の有害事象を発現した症例の割合
 - 各症例の最大Gradeの集計
-
- 各有害事象ごとの発現割合
 - 各有害事象ごとの最大Gradeの集計

また、安全性報告として、各臨床検査データ、
バイタルサインの推移等の算出も行う



これらの本がすごく良いです

解析対象集団

以下の2つが代表的である

- 最大の解析対象集団

Full Analysis Set (FAS)

～ Intention-to-treat 解析の原則に最も近く、
主要評価項目の評価に利用される

- 実施計画書に適合した対象集団

Per Protocol Set (PPS)

最大の解析対象集団（FAS）

「最大の解析対象集団」という用語を用いて、ランダム化が行われた全被験者を含める **intention-to-treat 解析** という理想に可能な限り近づけた解析対象集団を表す。

ランダム化が行われた被験者を最大の解析対象集団から除外できる状況は限られている。

- 主要な登録基準を満たしていない場合（適格基準違反）
- 試験治療を一回も受けていない場合
- ランダム化後のデータがない場合

最大の解析対象集団 (FAS)

「最大の解析対象集団」という用語を用いて、ランダム化が行われた全被験者を含める **intention-to-treat 解析** という理想に可能な限り近づけた解析対象集団を表す

ランダム化
できる状況

- 主要な
- 試験
- ラン

Intention-to-treatの原則

割付後にいかなる事態となっても、
割り付けられたグループのまま（プロトコール違反
や脱落によらず）、全対象者を解析対象とする原則。
「ランダム化の保持」が最大の目的である。
保守的となるため、得られた結果が望ましいもので
あった場合には信頼性が高い

最大の解析対象集団（FAS）

最大の解析対象集団を用いる場合、ランダム化後に起こる治験実施計画書違反は、**特にそれらの発生が試験治療の割付と関連しているときに、データ及び結論に影響するおそれがある。**

ほとんどの場合、ランダム化後に違反があった被験者のデータを解析に含めることが適切であり、intention-to-treat の原則とも一致する。試験治療を一度以上受けた後で試験治療を中止し、その後のデータがない被験者及び中止ではないが追跡不能となった被験者の扱いに関連して特別の問題が起きる。

脱落（投与中止や来院不能）となった場合にも、可能な限り主要評価項目等を収集すべきである。そうすることで、脱落した対象者が含まれる場合においても統計モデル等を適用し、より真実に近い推定を実施することができるともいえるからである。

実施計画書に適合した対象集団 (PPS)

「治験実施計画書に適合した」対象集団は、最大の解析対象集団の被験者のうち治験実施計画書をより遵守している一部であり、以下のような基準から特徴づけられる。

- 事前に定められた最低限の試験治療規定を完了していること
- 主要変数の測定値が利用可能であること
- 登録基準違反などの重大な治験実施計画書違反がないこと

治験実施計画書に適合した対象集団を用いることは、解析において新しい試験治療が有効性を示す可能性を最も大きくするといっている（⇔保守的）。しかし、対応する仮説検定と試験治療の効果の推定値は試験により保守的となる場合もある。治験実施計画書の遵守状況が試験治療及び臨床結果と関連している場合には、生じるバイアスは重大なものとなるおそれがある。

対象集団の選択による影響

検証的試験では、「**最大の解析対象集団の解析**」と「**治験実施計画書に適合した対象集団**」の解析との相違を明示的な議論と解釈の対象にできるよう、通常両方の解析を計画することが適切である。

優越性試験では、最大の解析対象集団から得られる有効性の推定値は、治験実施計画書に適合した解析ほど過度に楽観的ではないため、最大の解析対象集団が（例外的な状況を除いて）主要な解析に用いられる。最大の解析対象集団に含まれる遵守状況の悪い被験者は、推定される試験治療の効果を一般に小さくするからである。

しかし非劣性試験では、最大の解析対象集団を使用することは一般に保守的ではないため（差が出にくくなる＝研究仮説に近づいてしまう）、その役割は十分慎重に考慮すべきである。

欠測・外れ値

欠測値は、臨床試験においてバイアスを起こし得る代表的な原因である。

欠測値を扱う方法が適切で、特にそれらの方法が治験実施計画書の解析計画中に事前に定められている場合には、試験は妥当であるとみなすことができる。

欠測値に対処する方法で、普遍的に適用可能と薦められる方法はない。欠測値に対処する方法により解析結果がどの程度変わりやすいかを検討すべきである（**感度解析**）。脱落がイベントや割付と関連がある場合には、重大なバイアスの要因となるため、事前にどのようにするか定めておくことが重要である。

同様の手法は、外れ値（その統計的定義はある程度恣意的だが）の影響を探索するためにも用いるべきである。特定の値が外れ値であると明確に判断するのに最も説得力があるのは、統計的に正当であるだけでなく医学的にも正当な場合であり、医学的な背景によりしばしば適切な対応が決定される例がみられる。

統計解析報告書

統計解析の結果とその解釈を，統計解析責任者および担当者によって報告書にまとめる．治験の際の総括報告書は，統計解析報告書の内容が大部分を占めており，文書として極めて重要なものである．

統計解析担当者は，報告書をまとめる上で，選択した手法と解釈，そしてその根拠について十分に記載しなくてはならない．

ICH-E9（「臨床試験のための統計的原則」）に，重要な記載がまとめられているためそちらも参照してください．

②データマネジメント

i. データマネジメントの概要

データマネジメント：DMとよく略される



データマネジメントの目的

統計解析のための基礎となる
クリーンなデータベースを
提供することである

R.Hattermer-Apostel 他, 2008

データマネジメントの目的

統計解析のための基礎となる
クリーンなデータベースを
提供することである

クリーン：
正確で、データに不正がない

意図しない誤り
意図した誤り

データマネジメントの目的

統計解析のための基礎となる
クリーンなデータベースを
提供することである

データの品質管理であり、
データの品質保証でもある

一般的には、品質管理業務の1つとみなされる

品質マネジメントにおける位置付け

ひっくるめて品質マネジメント



品質管理

品質保証

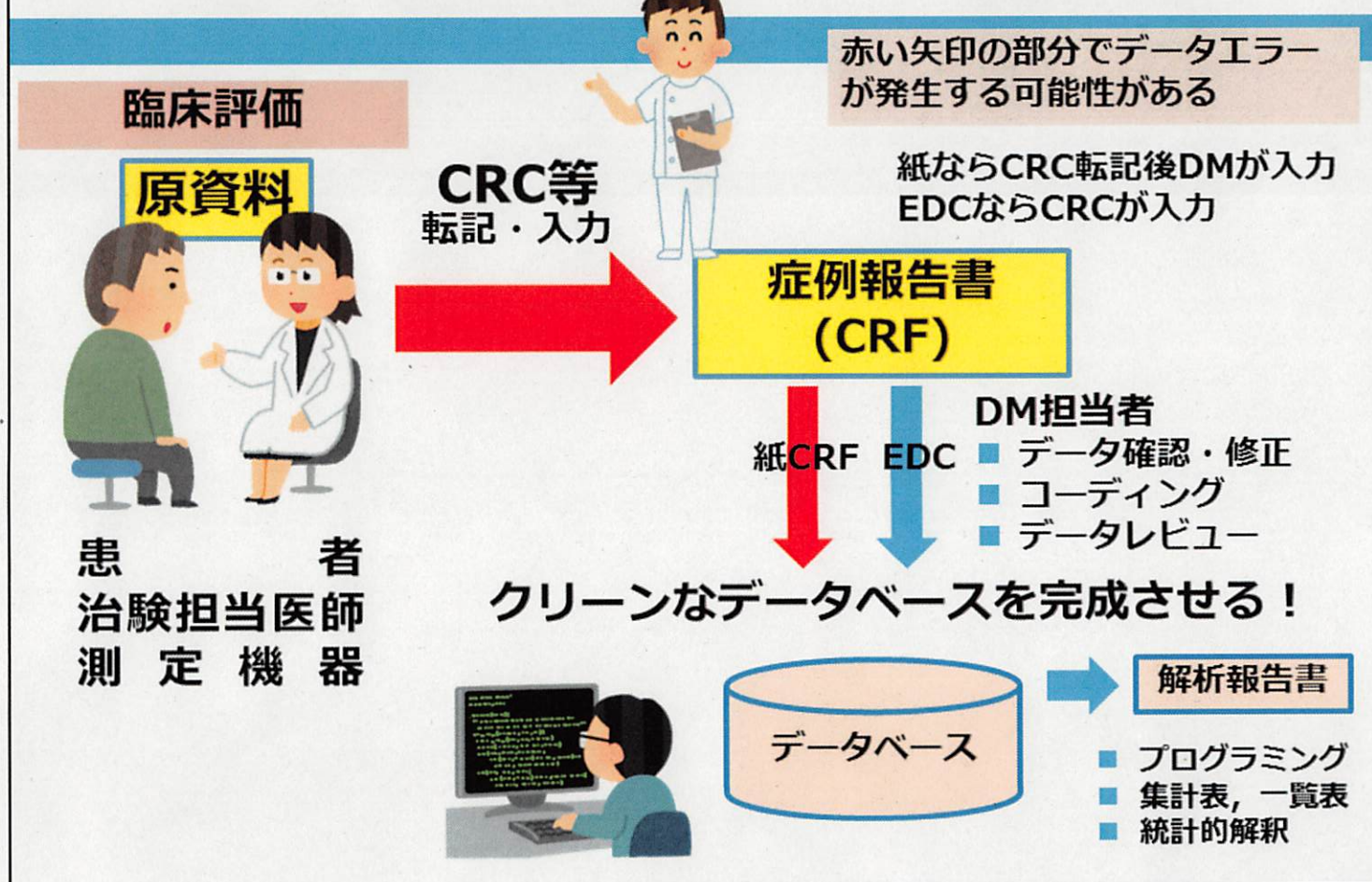
計画書の作成

データマネジメント

モニタリング

監査

データマネジメント概念図



データマネジメントの必要性

クリーンなデータベースを作成することは
思っている以上に
容易でない！！！！

- 記載ミス・入力ミス
 - 記載項目の理解不足
 - 症例報告書の作成／必要事項の検討
 - 監査証跡を残さなくてはならない
 - リレーショナルデータベースの作成
 - コーディング
 - 標準化
 - 研究不正の予防・発見
 - 膨大な被験者数・多施設の場合
 - 法令の遵守
 - ちょうどいいところを見つけられないといけない
- etc…



データマネジメント業務

登録：誰が臨床試験に登録されたのかを明確にしないと対象集団が曖昧に

CRFの作成支援：類型化されている部分もあり，必要事項を過不足なく収集するノウハウがある．研究者の方はたいてい多くの項目を収集しすぎて質が低下してしまう．

EDCのセットアップ：ソフトウェアの利用に習熟する必要がある

法令の理解：GCPを初め，個人情報や電磁的記録の取り扱いについて，監査証跡とは何か知っていますか？

CDISC等の標準化への対応

データマネジメント業務

データクリーニング：汚い（誤りの多い）データは解析できない。
クエリ（問い合わせ）発行や確認作業

中央モニタリング：収集されたデータから統計的な手法により試験の品質を管理する

データ管理（データ固定等）：研究者グループによるデータ管理はCOIの観点から問題となることがある。第三者によるデータ管理は重要である。

②データマネジメント

ii. 臨床試験準備



原資料に求められる基本要素

原資料：カルテや検査報告書，PRO等の原データ

- 正確であること (accurate)
- オリジナルであること (original)
→ 最初に記録されたものであること
- 読むことができること (legible)
- データを誰が記録し，どこから得られたかが明確であること (attributable)
- データは発生から時間を置かずに記録されていること (contemporaneous)

FDAガイダンスにおいて言及されたもの

原資料に求められる基本要素

原資料：カルテや検査報告書，PRO等の原データ

- 正確であること (accurate)
- オリジナルであること (original)
→ 最初に記録されたものであること
- 読むことができること (legible)
- データを誰が記録し，どこから得られたかが明確であること (attributable)

症例報告書は原資料と矛盾しないものでなくてはならない。

FDAガイダンスにおいて言及されたもの

症例報告書（CRF）

CRF（紙 or EDC）はなぜ必要か

- **監査証跡**を残すため
- 効率的なデータ収集のため

EDC: Electric Data Captureの略

電子的なデータ収集システムを指す。

紙CRFを介さず、システム上にeCRFを構築し臨床現場から直接入力を実施する

GCPでは必須です！

監査証跡

紙CRFの場合は記名を伴う二重線による修正等により
監査証跡を残すのが容易

EDCの場合にはeCRFへのアクセス履歴を全て残すよ
うなシステムが必要となるため、Excel直接入力など
ではダメ！（そういったログは残りません）

品質管理上、
いくら頑張っても、
記録に残さなければやっていないのと同じ
とみなされます

紙CRFとEDC

紙CRFは手軽で法令等の縛りも少ないため運用しやすいが、例数・データ量が多かったり、試験の性質上修正が多くなる場合などはレイアウトや運用が困難なものとなりやすい。

また、転記の回数が増えるため誤りが生じる可能性が増す。原本が1つしかないため、コピー等による運用を行うと複数バージョンが存在することに拠る混乱が生じやすく、またモニタリング時などに原本の受け渡しが必要となり非効率的である。データベース作成時にはいずれにしてもEDCに準じたシステムの利用が必要となる。

紙CRFとEDC

EDCはセットアップに手間がかかり、法令の理解もより重要となるため一定のハードルがあるが、専門家が運用する上では効率が良い。特に、紙面の制約に因われずレイアウトが可能であったり、転記が減る、データベース化が容易、CRF原本のやり取りが不要で修正時にも混乱がない、等のメリットがある。

上記により、EDCによる運用が可能であれば、例数によらずEDCを利用する方が良いと考えられる。

症例報告書の項目を定めた CDASH標準

目次

共通の識別変数 COMMON IDENTIFIER VARIABLES.....	2
共通の時期変数 COMMON TIMING VARIABLES.....	5
有害事象 ADVERSE EVENT.....	7
前治療, 併用治療 PRIOR AND CONCOMITANT MEDICATIONS	20
薬剤使用記録 DRUG ACCOUNTABILITY	31
患者背景 DEMOGRAPHICS.....	35
試験対象被験者の内訳 (転帰) DISPOSITION	43
試験実施計画書からの逸脱 PROTOCOL DEVIATIONS	49
心電図検査結果 ECG TEST RESULTS	53

症例報告書の項目を定めた CDASH標準

薬物暴露 EXPOSURE	72
不適合選択基準 / 除外基準 INCLUSION / EXCLUSION CRITERIA NOT MET	84
臨床検査結果 LABORATORY TEST RESULTS	86
<SCENARIO 1>	87
<SCENARIO 2>	90
<SCENARIO 3>	97
既往症・合併症 MEDICAL HISTORY	101
身体所見 PHYSICAL EXAMINATION	107
<BEST PRACTICE APPROACH>	108
<TRADITIONAL APPROACH>	110
被験者特性 SUBJECT CHARACTERISTICS	114
嗜好品等 SUBSTANCE USE	115
バイタルサイン VITAL SIGNS	120

症例報告書 (CRF)

- 間違いにくいレイアウト
- わかりやすい表現
- 「その他」等の自由記載は**最小限**
 - Display question (質問者が知っている答えを要求)
 - Referential question (質問者が知らない答えを要求)
- 重複を最小限
- コード化

データを集めれば集めるほど症例報告書は厄介なものになり、
混乱したり誤ったりしがちになる Williams, G.W, 2006

データベース定義書

■ データベース（構造）定義書・仕様書

データベースを作成する上の基本文書。以下の情報を含む

- 名前とラベル
- 各変数が含まれるドメイン（大分類）の指定
- 型式（例： 数値, 文字, 整数, 小数, 日付）
- コードリストに含まれる全てのコード値の定義（例：1は「はい」, 2は「いいえ」など）
- 入力後に導出または算出する変数のアルゴリズム

とくに変数名やドメイン名などはCDISCによる標準化の核となる

データマネジメント計画書

- 登録手順を定める
 - 研究スケジュールの明確化
 - 実施体制／担当者
 - 使用するシステム（紙・EDC）
 - データ収集に関する規定（締め切り等）
 - データの管理
 - ～ データチェック, 問い合わせ, 修正
 - データ固定の手順を定める
 - 記録の保管について
- などなど

データマネジメント業務

■ 症例報告書のチェック・レビュー

被験者情報

登録番号	CRF001
性別	1. ☒ 男 2. ☒ 女
年齢	<u>990</u> 歳

データマネジメント業務

■ 症例報告書のチェック・レビュー

現病歴・既往歴はすべてこちらに記載して下さい
番号

1	既往症・併存症名	糖尿病のケがあると言われた
	継続していますか？	1. ☐ はい / 2. ☒ いいえ
	コントロールされていますか？	1. ☒ はい / 2. ☐ いいえ
	調査日	20 990 / 13 / 32

データマネジメント業務

■ 症例報告書のチェック・レビュー
おかしい場合には「クエリ」（問い合わせ）を発行し、
現場の医師・CRCに問い合わせる

番号

1	既往症・併存症名	糖尿のケがあると言われた
	継続していますか？	1. ☐ はい / 2. ☒ いいえ
	コントロールされていますか？	1. ☒ はい / 2. ☐ いいえ
	調査日	20 990 / 13 / 32

データマネジメント業務

■ データの電子化

随時データを抽出し、
統計学的手法等を利用した品質管理を実施する

USUBJID	Domain	VISITNUM	VSSTAT	VSDAT	VSTIM	VSTPT	VSSPID	VSTEST	VSORRES	VSORRESU	VSCLSIG	VSLOG	VSPOS	unused
1	VS	1	0	0	0	0	1	PS (ECOG)	0		2	1		FALSE
1	VS	1	0	0	0	0	2	血圧	0			0	0	FALSE
1	VS	1	0	0	0	0	3	体温	0			0	0	FALSE
1	VS	1	0	0	0	0	4	身長	0			0	0	FALSE
1	VS	1	0	0	0	0	5	体重	0			0	0	FALSE
1	VS	1	0	0	0	0	6							TRUE
1	VS	1	0	0	0	0	7							TRUE
1	VS	2	0	0	0	0	1	PS (ECOG)	0		2	1		FALSE
1	VS	2	0	0	0	0	2	血圧	0			0	0	FALSE
1	VS	2	0	0	0	0	3	体温	0			0	0	FALSE
1	VS	2	0	0	0	0	4							TRUE
1	VS	2	0	0	0	0	5							TRUE
1	VS	2	0	0	0	0	6							TRUE
1	VS	2	0	0	0	0	7							TRUE
1	VS	3	0	0	0	0	1	PS (ECOG)	0		2	1		FALSE
1	VS	3	0	0	0	0	2	血圧	0			0	0	FALSE
1	VS	3	0	0	0	0	3	体温	0			0	0	FALSE

CDISC標準におけるドメイン「VS」（バイタルサイン）の例
変数名もCDISC標準に則っている

コーディング



報告語の世界

腹痛

はらいた

腹部の圧痛

左下腹部圧痛

用語集の世界

AE	件数
腹痛	2
腹部圧痛	2

読み替えの
ルール
Coding
convention

コーディング

有害事象

臨床問題となる所見はすべてこちらに記載して下さい

番号

ここにどう書くか！

報告語の世

腹痛

はらいた

腹部

左下腹部圧痛

1	有害事象名	
	発現日	20__/__/__
	消失日	20__/__/__
	継続中	1. ☐ はい / 2. ☐ いいえ
	重症度	CTCAE GRADE __
	重篤	1. ☐ はい / 2. ☐ いいえ
	被験治療との関連性	1. ☐ あり / 2. ☐ 可能性あり / 3. ☐ なし
	被験治療の変更等	1. ☐ あり(内容:) / 2. ☐ なし
	関連性の理由・その他の説明	
	転帰	1. ☐ 軽快 / 2. ☐ 回復 / 3. ☐ 未回復 / 4. ☐ 死亡

3	有害事象名	
	発現日	20__/__/__
	消失日	20__/__/__
	継続中	1. ☐ はい / 2. ☐ いいえ
	重症度	CTCAE GRADE __
	重篤	1. ☐ はい / 2. ☐ いいえ
	被験治療との関連性	1. ☐ あり / 2. ☐ 可能性あり / 3. ☐ なし
	被験治療の変更等	1. ☐ あり(内容:) / 2. ☐ なし
	関連性の理由・その他の説明	
	転帰	1. ☐ 軽快 / 2. ☐ 回復 / 3. ☐ 未回復 / 4. ☐ 死亡

の世界

件数

2

痛

2

読み替えの
ルール

Codina

MeDDRAの下層語 (LLT) やCTCAEの用語を用いるのが一般的

コーディング : なぜ必要？

有害事象の提示にあたっては、治験責任医師が用いたとおりの用語で提示することも、関連する事象（つまり、おそらく同一の現象を表す事象）をグループ化することも、ともに重要である。

グループ化することによって真の発現率があいまいでなくなる。

このための1つの方法は、標準的な副作用・有害事象用語辞書を用いることである。

治験総括報告書の構成と内容に関するガイドライン

コーディング辞書（用語集）

MedDRA	標準的な医薬品用語集 高位グループ用語，基本語，下層語などに分類されている
CTCAE	有害事象の性質や重篤度のコーディングのために米国国立癌研究所が開発（MedDRAを利用している）
ICD-9	診断や処置のためのコードからなる．1977年WHOが公開
ICD-9-CM	ICD-9の更新版．1988年からメディケアやメディケイドでは必須
ICD-10	1992年完成，世界中で実装されたが，米国では採用されず

GCDMP2009年度版より

②データマネジメント

iii. 臨床試験実施と終了



データの修正

■ データチェック/ 整合性確認

- マニュアルチェック
- 電子チェック

■ データ変更

- データの更新
- クエリーの更新
- 実施医療機関から申し入れがあった変更

監査証跡！！！！

データの整合性の確認

たとえば

■ 有害事象の場合

「有害事象」のフォームと「臨床検査値」のフォームで整合性が取れているか？

日付の整合性はフォーム間で問題がないか？

厳格な一致を求める場合とそうでない場合を規定しておく（照合すべき項目、一致レベルを設定）

有害事象

臨床問題となる所見はすべてこちらに記載して下さい

番号

1	有害事象名	
	発現日	20__/__/__
	消失日	20__/__/__
	継続中	1. ☐ はい / 2. ☐ いいえ
	重症度	CTCAE GRADE __
	重篤	1. ☐ はい / 2. ☐ いいえ
	被験治療との関連性	1. ☐ あり / 2. ☐ 可能性あり / 3. ☐ なし
	被験治療の変更等	1. ☐ あり(内容:) / 2. ☐ なし
	関連性の理由・その他の所見	
	転帰	1. ☐ 軽快 / 2. ☐ 回復 / 3. ☐ 未回復 / 4. ☐ 死亡

3	有害事象名	
	発現日	20__/__/__
	消失日	20__/__/__
	継続中	1. ☐ はい / 2. ☐ いいえ
	重症度	CTCAE GRADE __
	重篤	1. ☐ はい / 2. ☐ いいえ
	被験治療との関連性	1. ☐ あり / 2. ☐ 可能性あり / 3. ☐ なし
	被験治療の変更等	1. ☐ あり(内容:) / 2. ☐ なし
	関連性の理由・その他の所見	
	転帰	1. ☐ 軽快 / 2. ☐ 回復 / 3. ☐ 未回復 / 4. ☐ 死亡

確認

血液検査では
Hb低値だけど、
有害事象にない、
など

血液検査

臨床検査を実施しましたか？	1. ☐ はい / 2. ☐ いいえ
検体採取日	20__/__/__

以下に記載するか、検査結果を印刷した用紙を添付してください。

No.	検査項目名	検査値	単位	検体の状態	異常値フラグ
1	白血球数(WBC)				1. ☐ はい 2. ☐ いいえ
2	白血球分画				1. ☐ はい 2. ☐ いいえ
3	赤血球数				1. ☐ はい 2. ☐ いいえ
4	ヘモグロビン				1. ☐ はい 2. ☐ いいえ
5	ヘマトクリット				1. ☐ はい 2. ☐ いいえ
6	血小板数				1. ☐ はい 2. ☐ いいえ
7	PT				1. ☐ はい 2. ☐ いいえ
8	APTT				1. ☐ はい 2. ☐ いいえ

電子的なデータチェック

エディットチェック

→ データベース, EDC等によって作成される,
自動的に発せられる警告又は注意

GCDMP2009年度版より

電子的なデータチェック

エディットチェック

→ データベース, EDC等によって作成される,
自動的に発せられる警告又は注意

Input	Control	表示1	表示2	Check	文字数
1		嗜好品等 1			
1 喫煙・たばこ	11 11	摂取項目	喫煙・たばこ 1. ☐ 摂取したことがない 2. ☐ 現在も摂取している 3. ☐ 過去に摂取していた	0	10
2	12	摂取状況		0	1
10	16 10 10	消費量 単位 頻度	___本 / 日	0 0 0	4 4 4
1990	17 10	開始年 終了年	___年 ___年	0 0	4 4
120	18 10	継続期間 継続期間単位	___年	1 0	2 4

データチェックの種類と実施時期

チェックの種類

欠落している値

欠落しているCRFページ

値の範囲のチェック

重複項目に対するチェック

1つのCRF内の論理的矛盾

CRFの頁/モジュール間の矛盾

外部データのチェック

治験実施計画書違反

GCDMP2009年度版より

CRC



CRF記載時にその場でチェックできること



DM



データが集まったからこそわかること
(ページ間の矛盾など)

データの品質

参考入力方法とエラー率（文献）

- シングルエントリーのエラー率は二重入力の4-5倍
- シングルエントリーのエラー率：22/10000項目
- ダブルエントリーのエラー率：15/10000項目
- ★入力者によっては154又は194/10000項目のエラー
→入力者のスキルの影響
- 直接入力した場合のエラー率：2.0%
- 紙の様式に転記した後に入力した時のエラー率：1.1%

**手間と効果はトレードオフ
バランスを考えたい**

データの品質

- 主要評価項目
- 安全性や有効性の**核**となるようなデータ
(有害事象, 懸念されている検査値の変動, サロゲートとして有用な指標等)

については特に厳しい品質を求め, そうでないデータについてはほどほどに・・・

というか, そうでないデータは集めるべきではない.

エラー率の測定における留意点

エラー率：

検出されたエラー数を検査された総フィールド数で割ることで算出される。10000データフィールドあたりのエラー数として表現されることもある。

計算上の留意点

- 導出されたデータ項目を含むエラー
- 導出された日付と年のフィールドにエラーがあった場合のエラー数は1か2か？
- 一度入力されると、試験の全ページに電子的に入力されるヘッダーのエラーをどのようにカウントするか？

エラー率の許容範囲

- 米国臨床データマネジメント学会の調査(2004)

許容されるエラー率として採用されている数字

0.1% (回答者の34%)

0.5% (回答者の27%)

★エラーの数え方により最大7倍の差

- Duke Clinical Research Instituteの調査

1997年から2006年の試験の調査

0.025%~0.375% (平均 : 0.25%)

エラー率の許容範囲としてVery high quality data (0.01% ↓) が重要と考えられる人もいればHigh quality data (0.1 – 0.5%) で十分と考える人もいる

データの固定

- すべてのクエリが解消され、全CRFについて責任医師による承認が得られた後に「データ固定」がなされる
- 固定されたデータは変更が不可能となるため、修正が必要な際には理由を記した上固定解除の手続きを行う。各手続きにおいて書面等の証明書を発行することが重要。
- データ固定後、盲検化試験においては盲検下レビューを実施した後、単群あるいは非盲検の場合にはそのまま統計解析のステップに移行する。
- **100%完璧なデータセットを作成するのは不可能である**
- そのため、安全性・有効性の判断に問題を来さない水準の品質が確保されたと考えられた時点でデータ固定を行う

中央モニタリング・統計的モニタリング

モニタリングとは

臨床試験の進行状況を調査し、臨床試験が研究計画書、標準業務手順書（SOP）、GCP及び適用される規制要件に従って実施、記録されていることを確認する行為

⇒ 試験開始前から終了後に至るまで、試験が研究計画書や規制要件に従って実施されており、試験データ正しく記録・保管されていることを確認することを中心とした品質管理（Quality Control）に関する活動

Source document verification (SDV) といって、施設訪問の上原資料の確認を目視で行うなど非常に手間のかかる手法が主体

中央モニタリング・統計的モニタリング

中央モニタリング 統計的モニタリング

収集されたデータに基づき統計手法を利用して、主としてデータマネジメント部門で行うモニタリング

SDVなどの実地のモニタリングに比して、一度に多くの情報を扱うことが可能である。また、平均値の推移のような集団の傾向を巨視的に捉える点では個別モニタリングよりも優れている。しかし、個々のデータの誤りを正確に検出する作業についてはSDVに及ばない。

通常のモニタリングと組み合わせる、一部を置き換えるなどモニタリング手法の特性やコスト・手間などを考えて手法を選択する

中央モニタリング・統計的モニタリング

2. 有害事象

2.1. 概要

有害事象発生者数

コホート	最高の Grade	N	%
1	1	9	45.0%
	2	8	40.0%
	3	3	15.0%
	Total	20	100.0%
2	1	9	45.0%
	2	5	25.0%
	3	3	15.0%
	Total	17	85.0%

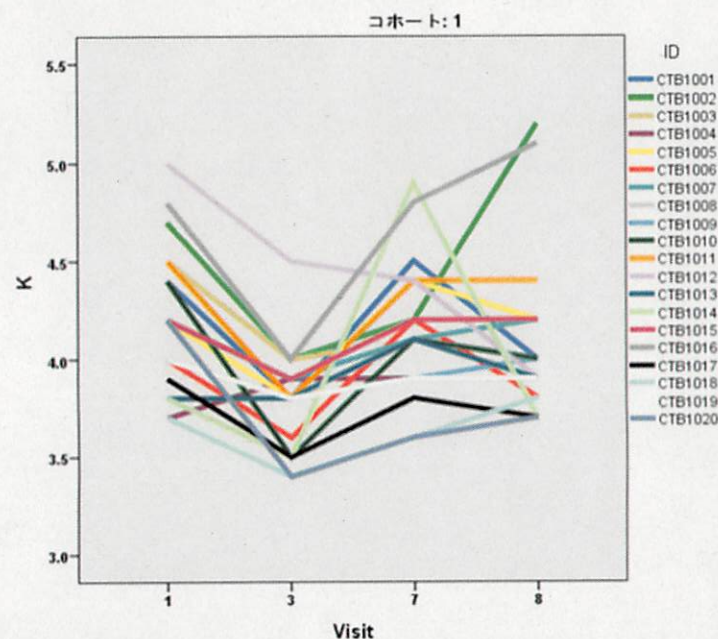
重篤な有害事象の発生はない

治験薬との関連の有無

コホート		N	%
1	関連なし	42	60.9%
	関連あり	0	0
	不明	27	39.1%
	全有害事象数	69	100
2	関連なし	19	43.2%
	関連あり	0	0
	不明	25	56.8%
	全有害事象数	44	100

有害事象発生者数-関連なしを除く

コホート	最高の Grade	N	%
1	1	5	25.0%
	2	5	25.0%
	3	2	10.0%
	Total	12	60.0%
2	1	7	35.0%
	2	4	20.0%
	3	3	15.0%
	Total	14	70.0%



有害事象の取りまとめ
臨床検査値の推移

中央モニタリング・統計的モニタリング

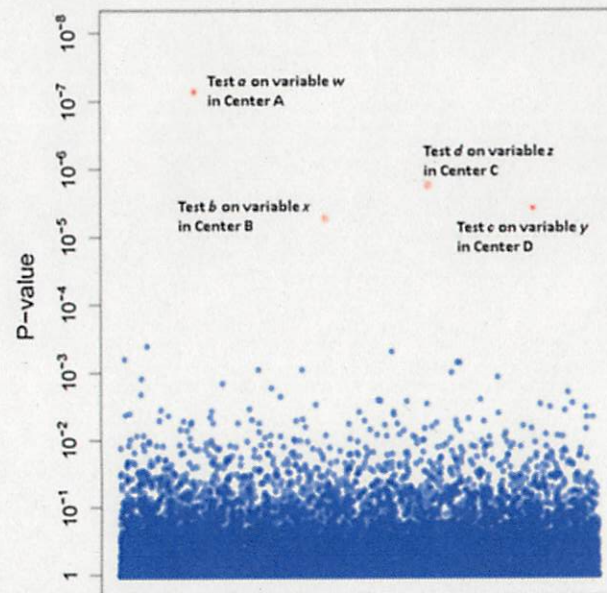
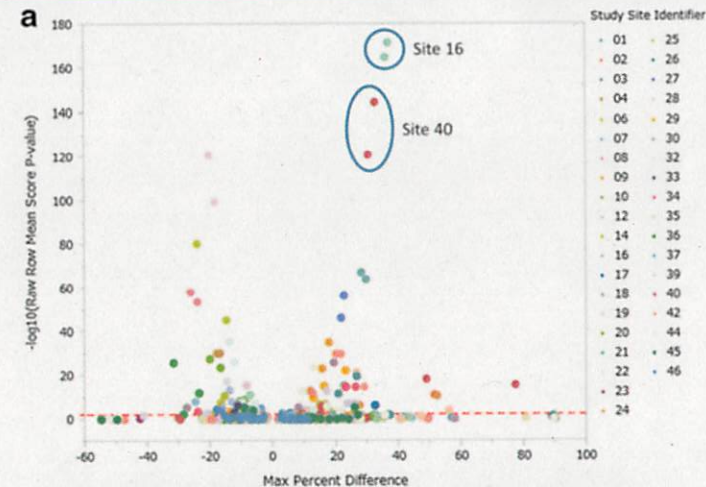


Fig. 4 Typical output of central statistical monitoring (CSM): the P values of all statistical tests performed are reported. The smaller the P value, the more atypical the distribution of the variable in the center, for the specific aspect (mean, variability, etc.) investigated by the test. Extremely low P values are identified by the related center, variable and statistical test. For example, in Center A an extremely low P value ($<10^{-7}$) was identified for the variable w using the statistical test a



各参加施設における検査値
および統合スコアの分布を
利用した異常検知

Int J Clin Oncol (2019) 23:38–45
DOI 10.1007/s10147-018-0877-5

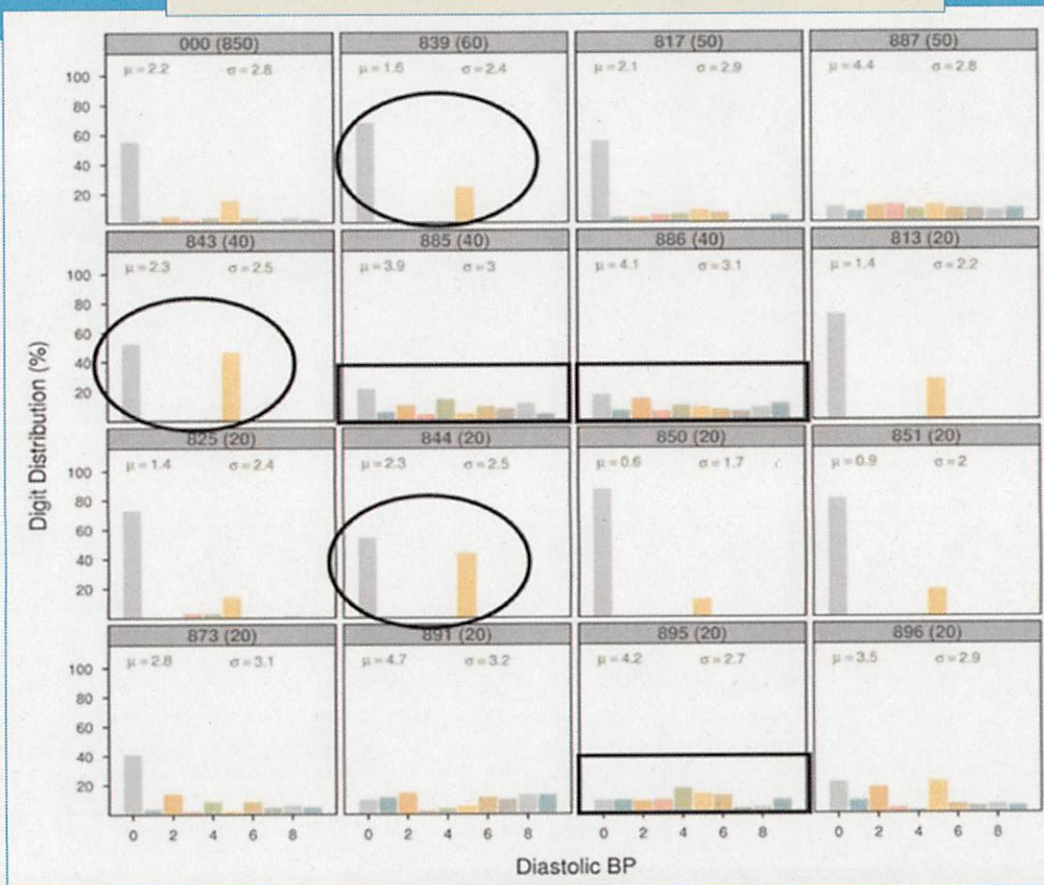
REVIEW ARTICLE

Data-driven risk identification in phase III clinical trials using central statistical monitoring

Catherine Timmermans¹ · David Venet¹ · Tomasz Burzykowski^{1,4}



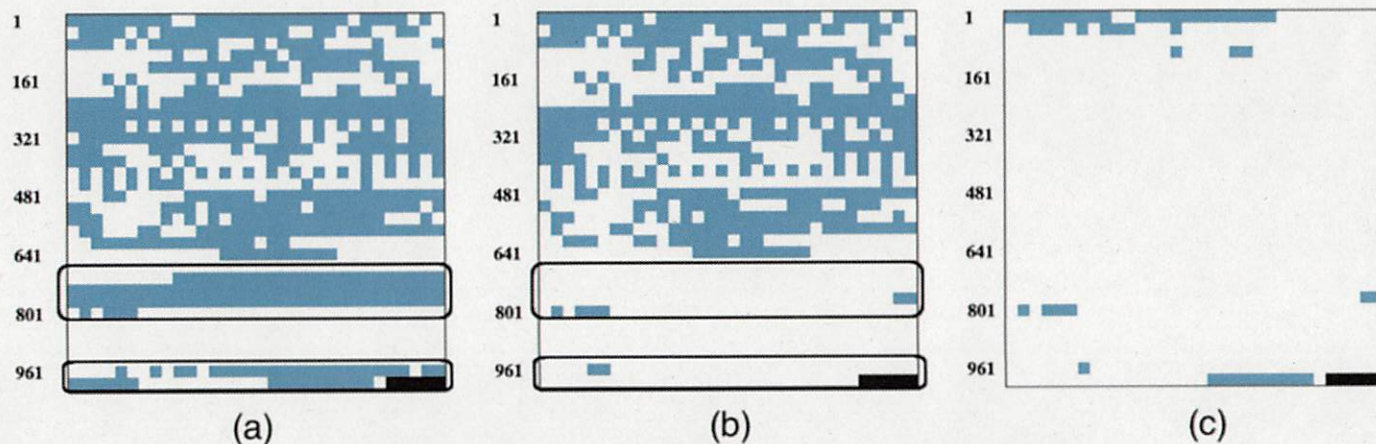
中央モニタリング 本来一様な分布を示す数値にバイアスが見られる



Lindblada et.al, Clinical Trials 2013; 0: 1-13

中央モニタリング・統計的モニタリング

統計データの視覚化によるアプローチ



欠測データパターンの可視化

Kelley R., J Biomed Informatics 2015;54:337-44

終わりに

生物統計・データマネジメントは、ともに臨床試験の中で極めて重要な位置を占め、専門家の手による計画・実施が不可欠です。

これら臨床試験の専門家が、早い段階から研究に関わることが、高品質で信頼性の高い研究を行うための基本となります。