



届出対策セミナー
統計学で事後チェック対策

株式会社オルトメディコ

人 本稿の目的（1）

■第1 機能性表示食品の科学的根拠に関する事項, 2 科学的根拠として明らかに適切とは考えられない具体例, (2) 最終製品を用いた臨床試験（ヒト試験）

ア 試験の実施計画又は実施方法に不備がある場合
(以下、省略)

イ 試験結果の評価に不備がある場合
(以下、省略)

機能性表示食品に対する食品表示等関係法令に基づく事後的規制（事後チェック）の透明性の確保等に関する指針より引用

■別紙4 システマティックレビュー（systematic review; SR）の実施手順に係る考え方, ⑦個々の論文の質評価, ア バイアスリスクの評価, (エ) 例数減少バイアスのリスク

比較する群の間に、例数の減少等に起因した系統的な差がないかどうかを評価する。例えば、アウトカムデータの欠損数及び理由が介入群・対照群で同様と判断される場合、例数減少バイアスのリスクは「低」、アウトカムデータの欠損数及び理由が介入群・対照群で不均衡な場合や、割付けられた介入からかなりの離脱が生じた試験で、元の割付けではなく、**実際に行われた介入を基に解析を行った場合（Per Protocol Set 解析）は「高」、論文において関連の記載が不十分だった場合は「不明」と評価する。**

機能性表示食品の届出等に関するガイドライン（2024年4月1日付け消食表第217号）より引用

機能性表示食品においてガイドラインをはじめ事後チェック指針においても、統計学的な指摘はあるものの、統計学的な留意事項は詳細に定められておらず、考慮すべき点や注意すべき点などが明示されていない。本稿では、ヒト臨床試験の実施において大きな影響を及ぼす可能性のあるデータの欠測についてまとめた。

人 本稿の目的（2）

なぜ欠測値に着目するのか？

ランダム化された全症例

- 主要な登録基準を満たしていない症例
- 介入ランダム化後に一度も受けていない症例
- ランダム化後のデータが全くない症例

最大の解析対象集団
Full Analysis Set (FAS)

- 最低限のプロトコル規定を満たしていない症例
- 主要変数の測定値が利用不可能な症例
- 重大な治験実施計画書違反症例
- 登録基準違反、禁止薬の併用、服薬の不遵守

etc

プロトコルに適合した対象集団
Per Protocol Set (PPS)

FASで解析するためには、
欠測値はつきものであり
適切な処理ができないと
良い結果を検出できない。

<よくある査読者からの指摘>

- 中止、脱落者も含めた基準での分析も必要ではないか？
- FASの解析もを見せてください。

↓ このような事例も…

経済

シークワサー由来成分の「認知機能改善効果」→「効果可能性」に修正 研究者ら発表

ノビレチン シークワサー

いいね! 21 シェア B! 0 ツイート 共有する

シークワサー由来の成分ノビレチンと認知機能の関連性を研究している、沖縄リサーチセンター（うるま市）と静岡県立大学は18日、昨年6月に発表した研究結果の内容の一部が「内容の一部に齟齬（そご）があることが判明した」として、修正を発表した。昨年6月は、認知機能の「改善効果を確認した」としていたが、「改善効果の『可能性』を確認した」に修正した。

共同研究した静岡県立大の山田静雄特任教授によると、115人の被験者で試験を開始し、中止、脱落者が7人いた。4カ月間の摂取を完了した108人の分析では、ノビレチンを摂取した人と偽薬を摂取した人の間で効果に有意差があったが、中止、脱落者も含めた基準での分析も必要だが、研究の倫理審査委員会から指摘された。

琉球新報より引用



ノビレチンが含まれるシークワサー（資料写真）

- 1 欠測データが解析結果に及ぼす影響
- 2 欠測データに対する統計学的手法の紹介
- 3 欠測データの発生を最小限に抑える方策
- 4 本日のまとめ



1

欠測データが解析結果に及ぼす影響

2

欠測データに対する統計学的手法の紹介

3

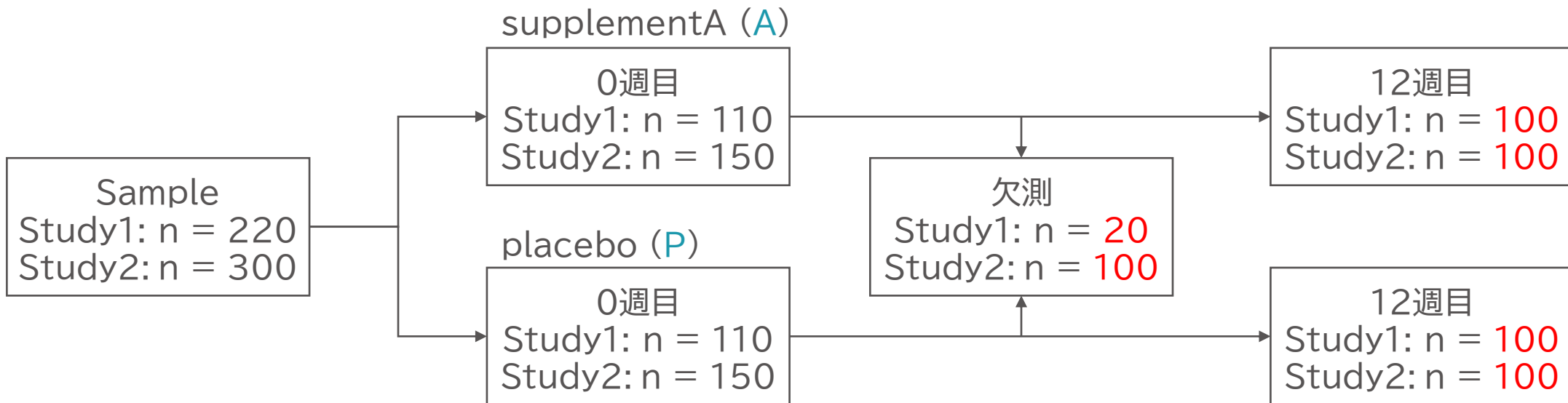
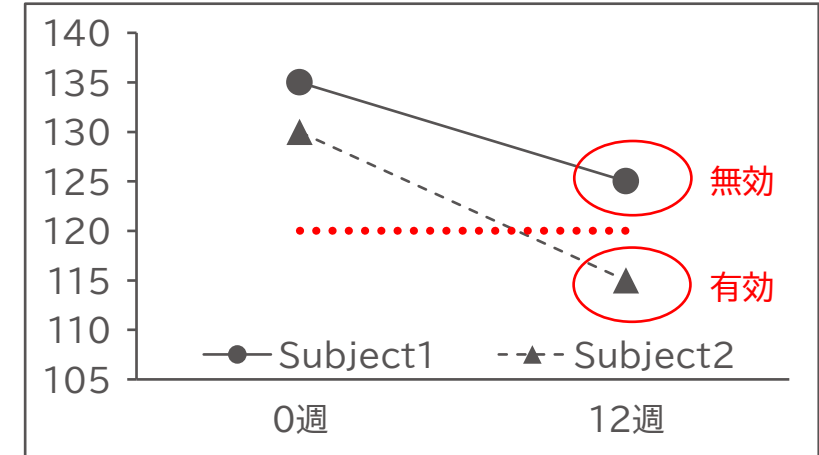
欠測データの発生を最小限に抑える方策

4

本日のまとめ

ヒト臨床試験（ヒト試験）の仮想事例

対象者	LDL-Cの測定値が120～139 mg/dLの成人日本人男女
介入	supplementA (A)
対照	placebo (P)
アウトカム	12週時のLDL-Cの測定値が境界域下限未満の割合
デザイン	二重盲検・ランダム化・プラセボ対照・並行群間デザイン
期間	12週間



人 仮想事例の解析結果と欠測

Study1の結果

	A	P
有効	80	20
無効	20	80
欠測	10	10



仮定	差 (%)	95%CI-	95%CI+	P
①観測データ	60.0	48.9	71.1	<0.05
②欠測が観測と同様の効果	60.0	49.4	70.6	<0.05
③すべての欠測が無効	54.5	43.5	65.6	<0.05
④Aの欠測は無効、Bは有効	45.5	33.7	57.2	<0.05

欠測データの取り扱い方で介入効果に差があるが、欠測データの割合が少ない場合は、どのような仮定においても結果は安定しており、有効性の証拠として受け入れやすい。

Study2の結果

	A	P
有効	80	20
無効	20	80
欠測	50	50



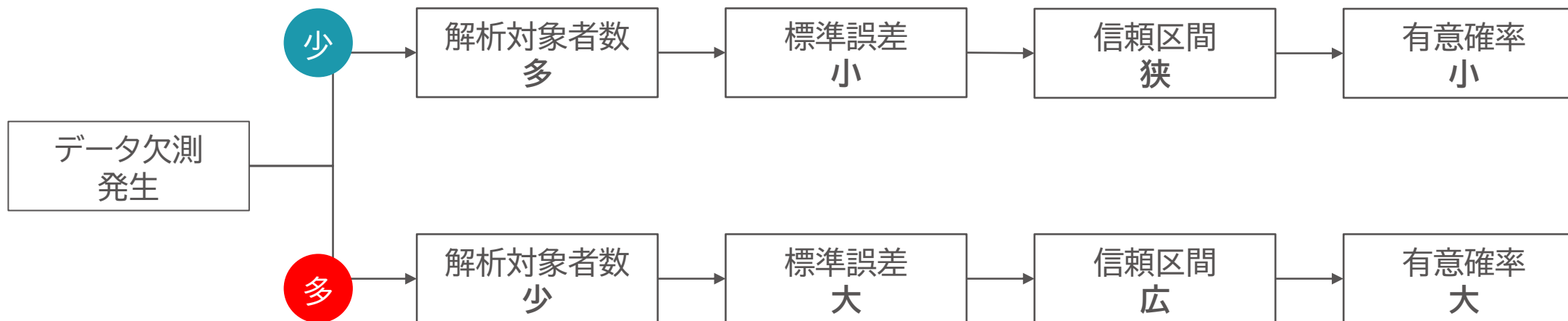
仮定	差 (%)	95%CI-	95%CI+	P
①観測データ	60.0	48.9	71.1	<0.05
②欠測が観測と同様の効果	60.0	50.9	69.1	<0.05
③すべての欠測が無効	40.0	30.3	49.7	<0.05
④Aの欠測は無効、Bは有効	6.7	-4.6	18.0	>0.05

欠測が大きいほど介入効果へ与える影響は大きい。
中止理由が多様に存在する場合には、結果を一意に表現することが困難になる。

人 欠測データがヒト試験の推論に与える影響（１）

精度の低下（検出力の低下）

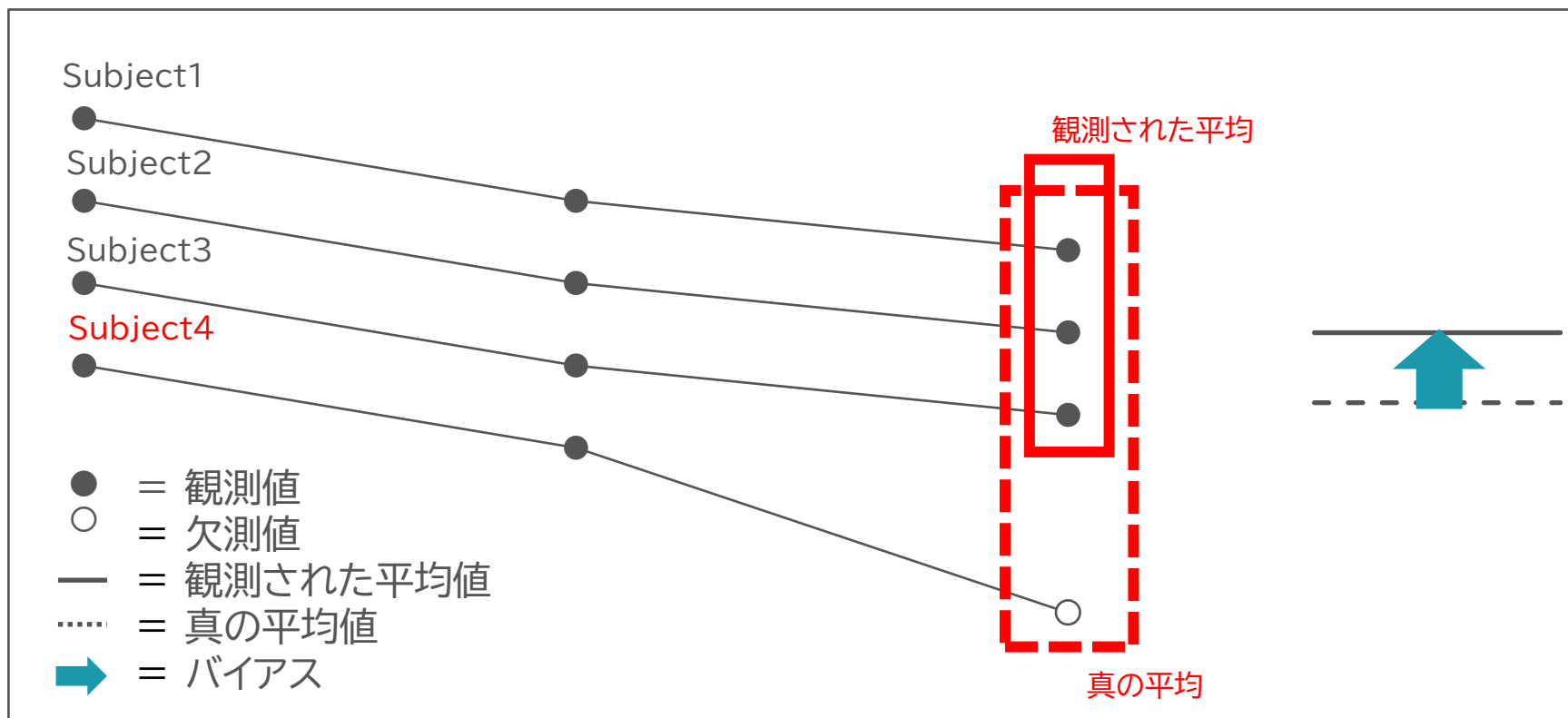
変化率や有効率などの推定値は、データ数が多くなれば標準誤差が小さくなり精度が高くなる。よって、欠測データが発生することによって、解析に利用できるデータ数が少なくなると、推定値の標準誤差が大きくなり、精度が悪くなる。これらのことにより検出力の低下に繋がり、信頼区間が広くなり、 P 値が大きくなる。場合によっては計画どおりの試験参加者数を収集し、想定通りの介入効果が観察されたとしても有意差が得られないということが起こり得る。



人 欠測データがヒト試験の推論に与える影響（2）

バイアスの発生

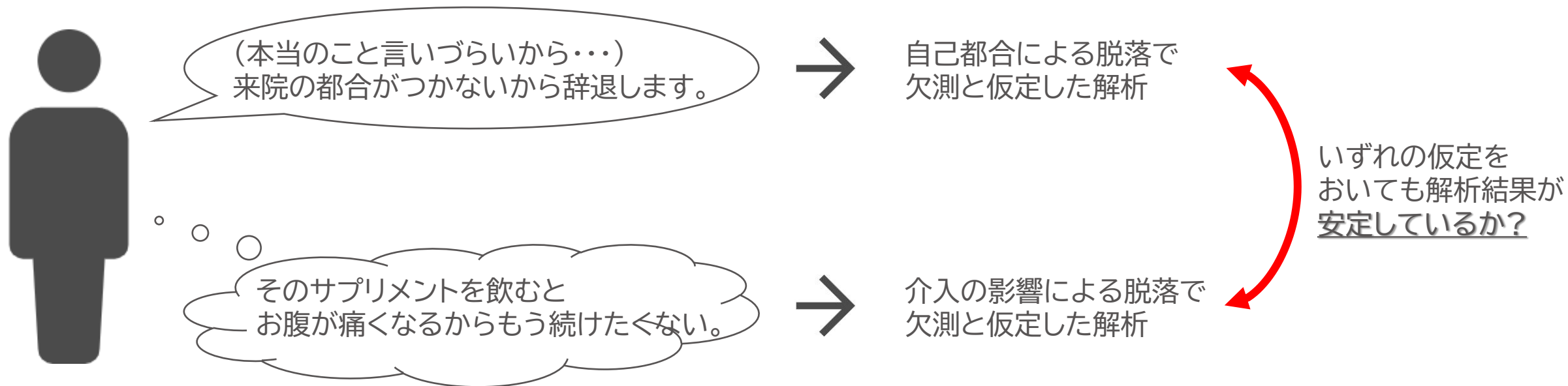
例えば、状態が悪くなると試験を中止する傾向が存在する場合に、試験に残された試験参加者のみで推測を行えば、その推定値は真実よりも良い結果となる。このように結果が真実よりも系統的に偏っていることをバイアスが発生しているといい、無視できないバイアスの危険がある場合には論文の査読などで指摘される可能性がある。



人 欠測データがヒト試験の推論に与える影響（3）

試験結果の安定性が損なわれる

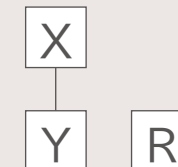
全ての欠測データの本当の発生理由はつきとめられない場合が多いし、欠測でなかったとしたらどのような値をとっていたかもわからない。だからこそ、さまざまな仮定を置いて「もしこの仮定が正しいならば、このような結果が得られていたはずである」という推論を行うのである。どのような仮定を置いても結果が大きく異ならないならば、結果の安定性は示せるだろう。しかし、仮定の置き方によって結果が大きく異なるのであれば、結果の安定性は損なわれ、結果の解釈において強い主張はしづらくなる。なお、安定性の評価では、仮定の置き方によってどの程度結果が異なるのか、置いた仮定のあり得なさ等を考慮すべきであり、安定性が保たれるか否かは程度問題である。



- 1 欠測データが解析結果に及ぼす影響
- 2 欠測データに対する統計学的手法の紹介
- 3 欠測データの発生を最小限に抑える方策
- 4 本日のまとめ

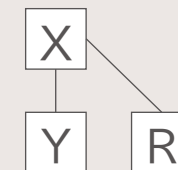
Missing Completely At Random (MCAR)

欠測値が**完全にランダムで生じた場合**である。つまり、欠測値の有無が他の分析に含まれる変数やその変数自体の値とは無関係である。欠測したときには 0、観測されたときには 1 の値をとる確率変数 R が、 X とも Y とも無関係であることである。仕事の都合で来院できなかったために欠測値が生じるような場合や測定装置がランダムな動作不具合を起こし欠測値が生じるような場合があり、これらはランダムに生じるためバイアスを生じない。



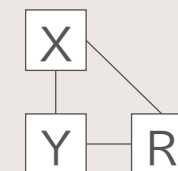
Missing At Random (MAR)

データが**測定されている値に依存して欠測（欠損データとは無関係）した場合**である。つまり観測データに依存した欠測である。よって、欠測値の生じるメカニズムが、観測されている変数で全て完全に説明することができる。 Y における欠損値の有無は、他の変数 X と関係しているが Y 自体の値とは無関係であるケースを示している。例えば、生存分析による被験者の死亡などでありバイアスを生じる。

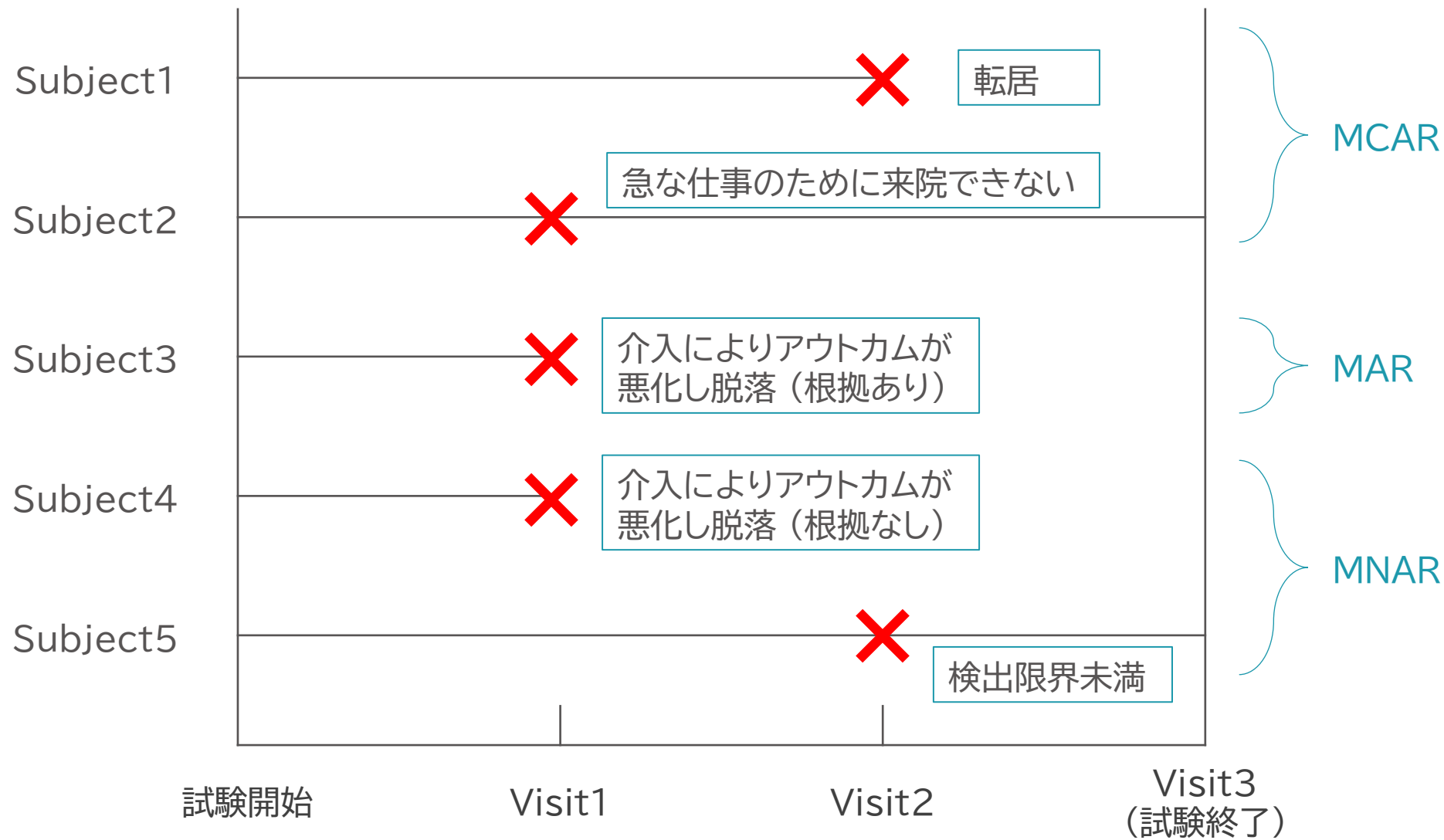


Missing Not At Random (MNAR)

データが**欠測データに依存して欠測する場合**である。つまり、観測データから説明が不可能である。分析に含まれる他の変数を統制した後でも、欠測値の有無が欠測値を持つ変数自身と関係を持つケースを示している。 X を完全に統制した後でも、 Y と欠測値の有無である確率変数 R の間に関係が残っている。例えば、喫煙を止めさせる研究において、来院しなくなったら喫煙を疑うことになる。このような場合、バイアスを生じさせることになる。



欠測メカニズムの具体例



人 症例ごとの欠測メカニズムを試験全体にあてはめる

- 脱落や中止による欠測のメカニズムがMCAR、MAR、MNARのどれに該当するかは疾患の特性・プロトコルの状況・組み入れられた試験参加者の状態等で異なるため、症例一人一人に対して検討が必要である。
- 欠測メカニズムによって、得られた推定量の妥当性が変わることがありうるため、脱落・中止理由の情報を収集することは当該試験の解析だけではなく似たようなデザインや介入のヒト臨床試験を計画する際にも重要な情報となる。

症例毎の欠測メカニズムと試験全体の欠測メカニズムを考える以下のようなになる。

- ① 全脱落・中止例のうち、全例MCARならば、試験全体で**MCAR**
- ② 全脱落・中止例のうち、1例以上MARが存在し、残りの全症例がMCAR ならば、試験全体では**MAR**
- ③ 全脱落・中止例のうち、1例以上MNARが存在すれば試験全体では**MNAR**

Complete case analysis

完了例でない試験参加者のデータを削除して、全評価時点の測定値が観測された被験者（完了例）に対して解析計画に基づいた解析方法を適用する方法であり、欠測値の取り扱いとして最も単純なアプローチである。解析の実行が非常に容易である点がメリットである。**Complete case analysisが根拠のある妥当な推定結果を与える状況は、欠測メカニズムがMCARのもとでは幾つか考えられるが、一般的には不適切であることが多い（※1）**。完了例ではない試験参加者のデータを用いないまま解析を行うため、バイアスを引き起こす可能性があること及び精度の低下の2点で問題があると考えられる（※2）。

1. 欠測割合が非常に少ない試験においては、Complete case analysisの結果は試験参加者全体の結果を反映しており、Complete case analysisの適用の可能性があると考えられる。
2. アクティブ群で無効例や中止例が多い場合に、完了例のデータのみを利用した解析では、有効性を過大評価する事になり、有効性評価において重大なバイアスを引き起こす可能性がある。また、ランダム化された被験者の中で完了例のみのデータを利用することで、例数が減少して、精度の低下に繋がると考えられる。

Group	Subject No.	Visit 1	Visit 2	Visit 3
A	1	0.6	0.1	0.8
A	2	0.4	1.0	0.3
A	3	0.9	0.4	
B	4	0.7	0.1	0.9
B	5	0.8	0.8	0.6
B	6	0.3	0.5	0.5

解析対象者から除外

人 Last Observation Carried Forward (LOCF)

単一補完の方法であり単調な欠測と非単調な欠測の両方に対して適用可能である。また、プロトコルで規定された時点で評価を行う試験において試験参加者の応答変数が脱落后に変化しないという強い仮定が成り立つ場合（※1）に導かれる応答変数の推定量にバイアスが入らないとされる補完法である。欠損値が存在しないデータを作成し、解析を実施するため比較的理解しやすい点がメリットである。**LOCFを妥当とする欠測メカニズムの仮定はMNARである。**

1. 単一代入法は、その背景にある理論的仮定が科学的に正当化されない場合には、主要な解析方法として用いられるべきではない。

Group	Subject No.	Visit 1	Visit 2	Visit 3
A	1	0.6	0.1	0.8
A	2	0.4	1.0	0.3
A	3	0.9	0.4	0.4
B	4	0.7	0.1	0.9
B	5	0.8	0.8	0.6
B	6	0.3	0.5	0.5

データ欠測者の
最終時点の測定値を代入する



Baseline Observation Carried Forward (BOCF)

LOCFと同様の単一補完の手法である。完了例ではない試験参加者の欠測値をその試験参加者のベースライン測定値で補完する手法である。つまり、ベースラインからの変化量をエンドポイントとする場合には欠測値は全て0で置き換えられる。BOCFは、脱落后には介入効果が消え、介入前の状態に戻ると想定することが妥当と考えられる場合に選択が可能である。一方で**LOCFに対する論説はBOCFに対しても同様にあてはまる。**LOCFやBOCF といった単一補完の方法は、データの不確実性を単一の値で補完することで無視することに繋がり、真のデータの複雑さとは異なるものになってしまうと考えられる（※1）。

1. 試験途中で有害事象などを示して脱落となった場合、症状が悪化したという事実を完全に無視してしまう。

Group	Subject No.	Visit 1	Visit 2	Visit 3
A	1	0.6	0.1	0.8
A	2	0.4	1.0	0.3
A	3	0.9	0.4	0.9
B	4	0.7	0.1	0.9
B	5	0.8	0.8	0.6
B	6	0.3	0.5	0.5

データ欠測者の
開始時点の測定値を代入する



Mixed-effect Models for Repeated Measures (MMRM)

MMRMは線形混合モデルを用いた解析手法である。これは経時測定データの場合に、欠損値を補完することなく、応答に対して特定の確率分布と回帰モデルを仮定することによって解析を行う手法である（※1）。そして、**妥当な結果を得るためには、MARの仮定が必要となる**。また、MARの場合においてバイアスのない解析を行うこと、および検定・信頼区間を通常の線形混合モデルと同様に取り扱うことが可能である。一方、モデルが非常に複雑であること、MNARの場合にバイアスが入る可能性があるという欠点を有する。また、この解析手法は脱落割合、時点間の相関性の影響を受けるため、症例設計を適切に行わなければならない。そして、不十分な症例数であった場合、適切な推定が行えず、バイアスを含む結果となることが考えられる。

1. 時点間の相関係数行列の構造をご特定すると固定効果パラメータの一致性も失われるため、注意が必要である。

Group	Subject No.	Visit 1	Visit 2	Visit 3
A	1	0.6	0.1	0.8
A	2	0.4	1.0	0.3
A	3	0.9	0.4	
B	4	0.7	0.1	0.9
B	5	0.8	0.8	0.6
B	6	0.3	0.5	0.5

欠測値を補完せずに
解析を行う

- 1 欠測データが解析結果に及ぼす影響
- 2 欠測データに対する統計学的手法の紹介
- 3 欠測データの発生を最小限に抑える方策
- 4 本日のまとめ

人 欠測を想定したサンプルサイズの設計

欠測データを考慮せずにサンプルサイズを一旦計算し、そこで得られた試験参加者数から想定した欠測割合を考慮し増加させる

$$\text{必要な試験参加者数} = \frac{\text{試験参加者数}}{(1 - \text{想定する中止割合})}$$

被験食品とプラセボの変化量の差を8（標準偏差10）と仮定した場合、両側 $\alpha=5\%$ 、検出力=80%の条件下で、有意差を検出するためには1群あたり26例が必要となる。脱落などによる欠測を20%と想定すると本試験の目標症例数は1群あたり33例となった。



慣例的に10%～20%が使用されることが多いが・・・

- アウトカムの種類
 - 過去の試験データ
 - 文献調査
- etc

様々なシナリオを考慮して設計することが推奨されている

人 欠測の発生を抑制する試験実施上の方策

データ収集段階で、試験参加者を煩わせない（不便をかける要因の制限）

- 来院や評価の回数を最小限にする
- 来院時は必要な情報のみを収集する
- 症例報告書は、記入する人だけでなくデータベースに入力する人にも使いやすいものにする
- 各評価時点の許容範囲を厳しくしすぎない。

試験責任医師は、ヒト臨床試験（ヒト試験）の実績が多くデータ収集の成績が良い医師にお願いする

- 臨床試験のトレーニングを受けている
- 試験参加者に試験の重要性を説明できる
- 看護師や臨床検査技師などの研究スタッフに適切なトレーニングができる

試験中のデータ収集のモニタリングと報告を確実に実行できる実施体制の構築

- 試験期間中に重要なアウトカム（主要アウトカムなど）に影響を及ぼす欠測や来院していない試験参加者の情報はリアルタイムで評価される必要がある
- データの収集状況を試験責任医師に伝え、完全なデータを収集できる環境にする
- 不備が多い施設は追加のトレーニングを行う

試験参加者に対する教育

- 同意取得や中止の同意の際に、完了データの重要性を強調して伝える
- 科学的な知見にとって、データがどれほど大切かを伝える
- 日誌への記入や入力を懇願する
- 来院日が近づいたらリマインドメールを送る
- ブログやSNSを通じて感謝の意を伝える

機能性表示食品とは・・・



機能性表示食品の安全性や機能性はどのように確保されているの？

- 事業者が、国の定めた一定のルールに基づき安全性や機能性に関する評価を行うとともに、生産・製造、品質の管理の体制、健康被害の情報収集体制を整え、商品の販売日の60日前までに消費者庁長官に届け出ることとなっています。
- 届け出られた内容は、消費者庁のウェブサイトで公開されます。
- 消費者の皆さんは、商品の安全性や機能性がどのように確保されているのかなどについて、商品の情報を販売前に確認できます。
- ◆消費者庁が中心となり、販売後の監視を行います。

1. 安全性の評価は？

以下のいずれかによって、評価されます。

- 今まで広く食べられていたかどうかの食経験
- 安全性に関する既存情報の調査
- 動物や人を用いた安全性試験の実施

医薬品との相互作用などについても評価されます。

2. 機能性の評価は？

以下のいずれかによって、評価されます。

- 最終製品を用いた臨床試験
- 最終製品又は機能性関与成分に関する文献調査（研究レビュー）※

「どのような科学的根拠に基づいて」「どのような人が」「どのように摂取すると」「どのような機能性があるのか」が明らかにされます。

- 「最終製品を用いた臨床試験」により科学的根拠が示されている場合、商品パッケージに「○○の機能があります。」のように表示されます。
- 「研究レビュー」により科学的根拠が示されている場合、「○○の機能があると報告されています。」のような表示が基本とされています。

【YouTube動画による帳票類の解説】



- 1 欠測データが解析結果に及ぼす影響
- 2 欠測データに対する統計学的手法の紹介
- 3 欠測データの発生を最小限に抑える方策
- 4 本日のまとめ



欠測データが解析結果に及ぼす影響

- ① 精度の低下（検出力の低下）
- ② バイアスの発生
- ③ 試験結果の安定性が損なわれる

ヒト臨床試験（ヒト試験）におけるデータの欠測は、100%防げるものではないため、欠測がデータに及ぼす影響を考えておく必要がある。



欠測データに対する統計学的手法の紹介

- ① 欠測メカニズムを判断するには適切な情報収集が必要
- ② 欠測メカニズムにあわせた解析方法を選択することでバイアスを抑制
- ③ とくにComplete case analysisに注意する

食品を介入としたヒト臨床試験（ヒト試験）ではComplete case analysisが使われがちである。しかし、それぞれの解析には問題をかかえていることを理解して使用することが望ましい。



欠測データの発生を最小限に抑える方策

- ① 欠測を想定したサンプルサイズの設計
- ② 欠測の発生を抑制する試験実施上の方策
- ③ 試験責任医師や施設のスタッフが実施できる方策

試験期間中のリアルタイムなモニタリングだけでなく、試験参加者自身も試験に協力したくなるような仕組みづくりがデータ欠測の予防につながる。

- 消費者庁. 機能性表示食品に対する食品表示等関係法令に基づく事後的規制（事後チェック）の透明性の確保等に関する指針（2024年6月19日アクセス可能: https://www.caa.go.jp/policies/policy/food_labeling/foods_with_function_claims/pdf/about_foods_with_function_claims_200324_0003.pdf)
- 消費者庁. 機能性表示食品の届出等に関するガイドライン（2024年4月1日付け消食表第217号）（2024年6月19日アクセス可能: https://www.caa.go.jp/policies/policy/food_labeling/foods_with_function_claims/assets/notice_240401_01.pdf)
- 琉球新報. 2021年2月19日. （2024年6月19日アクセス可能: <https://ryukyushimpo.jp/news/entry-1274504.html>)
- 厚生労働省. ICH E9 臨床試験のための統計的原則（平成10年11月30日付医薬審第1047号）（2024年6月19日アクセス可能: <https://www.pmda.go.jp/files/000156112.pdf>)
- Rubin DB. Inference and missing data. Biometrika, 1976; 63: 581–92. (DOI: <https://doi.org/10.1093/biomet/63.3.581>)

ヒト臨床試験・統計解析に関するお問い合わせは・・・

03-3818-0610
info@orthomedico.jp