

【日本製薬工業協会シンポジウム】

臨床試験のestimandを理解するための因果推論

-因果推論の基礎とcausal estimand-

5. LOCFとMMRM



2018年1月23日

医薬品評価委員会 データサイエンス部会 タスクフォース4

Estimand検討チーム

ノバルティスファーマ株式会社 片岡尚子

1. はじめに
2. 計画された最終時点前の欠測データの扱い
3. 具体例
4. まとめ

1. はじめに

- JPMALレポート(2017)はPermutt(2016)の全訳を目的としたものではない
- 基本的にはPermutt氏の意見
 - 本発表者の意見または補足の場合には適宜 **TF4** を付している
 - 特にPermutt氏の意見であることを強調したい部分には”Permutt氏の意見”等をつけて強調した
 - 式や定義の番号はJPMALレポート(2017)にしたがう
- 本シンポジウム(特にJPMA担当分)では, Permutt(2016)に記載されている用語を用いており, ICH E9(R1) step 2 ドキュメントで用いられているものと同じ用語が, 別の意味で用いられている部分もあるため, 注意されたい
 - Estimand
 - 「欠測」とは「データがないこと」を示す

➤ JPMALレポート (2017)

第3章 Permutt (2016) の解説

3.1 Introduction (Permutt (2016) の1 節)

3.2 因果推論の基本(Permutt (2016) の2 節+補足)

3.2.1 Causal estimand

3.2.2 因果推論に関する補足

3.2.2.1 治療効果

3.2.2.2 部分集団における因果効果

3.2.2.3 因果効果が推定可能となる条件

3.2.2.4 平均交換可能性についての補足

3.3 Taxonomy of estimand (Permutt (2016) の3 節)

3.3.1 De facto estimand/de jure estimand (Permutt (2016) の3.1.1 節)

3.3.2 Utility (Permutt (2016) の3.1.2 節)

3.3.2.1 [補足1]: 治療と死亡とアウトカムの関係についての整理

3.3.2.2 [補足2]: Utility estimand の計算のための準備(実薬群のみの場合)

3.3.2.3 [補足3]: Utility estimand の具体例(投与群を考慮する場合)

3.3.2.4 [補足4]: 観測された値のみによる群間差

3.3.3 Better-half (Permutt (2016) の3.1.2 節)

3.3.3.1 [補足1]: 本稿執筆者らによるbetter-half estimand の計算

3.3.4 LOCF とMMRM (Permutt (2016) の3.1.3 節)

3.3.5 Means in subsets (Permutt (2016) の3.2 節)

3.4 Some notes on estimation and testing in the regulatory setting (Permutt (2016) の4 節)

3.5 Discussion (Permutt (2016) の5 節)

2. 計画された最終時点前の欠測データ 一々の扱い

- (被験者の)最終測定値, 平均値, 合計値, 最良値
 - いずれも各被験者の全ての観測値を1つの値で表す
 - 観測値が1つでも存在すれば欠測にならない
 - 全てのランダム化された被験者が含まれているとの意味でITTと表現されることもある
- 群間差の検定: (分布を仮定した)t検定, (分布を仮定しない)順位和検定など
 - 「どの被験者のどの時点でも治療は結果に影響を与えない」という強い帰無仮説の下で α エラーを適切にコントロール
 - 最小限または全く仮定がない状況で統計的には正当化される

最終測定値に基づく検定(LAO)

➤ Last Available Observation(LAO)

- 最終測定値に基づく検定は, Last Observation Carried Forward (LOCF) と表現されてきた

- LOCFは計画された最終時点の値を観測された最終測定値で置き換えて(Carried)得られた(擬似)完全データを解析する方法を連想するため, LAOが正確な表現

➤ LAOは以下の仮定の下での推定値としては非現実的

- 試験治療中止したが観測を続けた場合の推定 (*de facto*)

- 試験治療を最後まで継続した場合の値の推定 (*de jure*)

➤ LAOは全く治療効果がないという仮説の下では 妥当な検定*

* TF4 注: α エラーがコントロールされるという意味

Mixed Model for Repeated Measures (MMRM)

- 利用できる観測値の平均に基づく検定
- 一般には時点を通しての平均的な治療効果を推定
 - 推定値が被験者内の重みつき平均の被験者全体の平均値に相当
 - より強い仮定の帰無仮説の下で妥当な検定
- (計画された)最終時点の *de facto* または *de jure* 効果を推定するのに用いられることもある
 - 推定値が妥当である状況は制限される
 - α エラーは強い仮定の帰無仮説の下でコントロールされる

- 現在(少なくとも国内では), MMRMは主に計画された最終時点での de jure 効果の推定に用いることが多いようである
 - 試験治療中止後のデータが欠測, または解析に用いないことを前提
- MMRMを de facto estimand の推定に用いることも可能
 - 試験治療中止後のデータを収集するデザインとした上で欠測に対してMARを仮定してMMRMを用いる

- LAOよりもMMRMに推定上の利点があると言われることが多いが、どちらも利用できるデータから推定
 - 統計手法の系統樹を作った場合、両者は近い位置
 - 全被験者の計画された最終時点での平均アウトカムの推定値としてはいずれも拡大解釈*(overinterpreted)
- * TF4 注: 常に推定値が大きくなったり, バイアスが入ったりするわけではない
- 検定の問題は統計的な妥当性ではない
 - 帰無仮説を棄却できた場合に, その薬剤が主張通りの有効性を持つと証明したことになるかどうか
 - 何が既知で, 何が未知であるのか, 試験で何を実証することを意図しているかという臨床的な設定に依存

3. 具体例

具体例1: 慢性疼痛(オピオイド鎮痛薬)

- 以前から鎮痛効果が知られている
- 有効性のある用量で毒性もある
 - 効果と毒性のバランスが取れた用量設定が必要
- 試験デザイン
 - 12週のプラセボ対照並行群間比較試験
 - 試験期間中, 複数時点で痛みのスコアの評価
 - 各群約半数の被験者が試験治療中止
 - 実薬群: 忍容性不足
 - プラセボ群: 効果不十分
- 解析対象: 計画された最終時点の値
 - 欠測の場合はLOCFによる補完

群間で中止理由が異なる

具体例1：慢性疼痛(オピオイド鎮痛薬)

- **結果**：試験治療中止割合，完了した被験者の最終時点での痛みのスコアの平均が等しいにもかかわらず，治療効果に有意差あり
- **原因**：群間で補完する状況(値)が異なる
 - 実薬群：毒性による中止が生じる前の**良い値**を補完
 - プラセボ群：効果不十分で中止する前の，痛みが改善されなかった**悪い値**を補完
- **試験は薬剤の鎮痛効果を実証しているが，長期投与による薬剤のベネフィットは示せていない**
 - 2群で中止理由が異なることは薬剤効果を示している
 - 鎮痛の薬理作用は明白なため，試験の主要な目的は長期間の投与計画における処方の有用性を示すこと

具体例1: 慢性疼痛(オピオイド鎮痛薬)

➤ 長期投与による薬剤のベネフィットを示すことが主要目的の場合, どう対処すべきか

- Permutt(2016)によると, LOCFやMMRMを主要な解析とすべきではない*

* **TF4** 注(試験治療中止後のデータを収集しない想定で)中止前のデータは長期投与のベネフィットを示すものではない

- Utility estimandも利用可能か

□ **TF4** 例: 毒性による中止を痛みのスコアと共通のスケールに変換

- **TF4** 長期投与時のベネフィットを評価できる試験デザイン, エンドポイントを用いることも一案

□ 試験治療中止後のデータの収集

□ 毒性による中止を含めた複合エンドポイントの利用

➤ 試験デザイン

■ 12週のプラセボ対照並行群間比較試験

□ 12週は長期使用のサロゲート

■ 反復測定デザイン

□ 試験後半の観測値が特に重要

□ 投与期間中に良い効果が現れたとしても、試験治療中止した場合は効果が高かったとはみなされない

➤ LAOやMMRMを用いた場合

■ 「薬剤が抗うつ作用を持つか」への回答は可能 → 主解析に

■ 「投与計画が長期間効果的に使われたか」は不明

➤ 部分集団解析により知るべき問い

■ 「薬剤がどのくらいの人に忍容性があるか」

■ 「忍容性のある人の中でどのくらい効果があるのか」

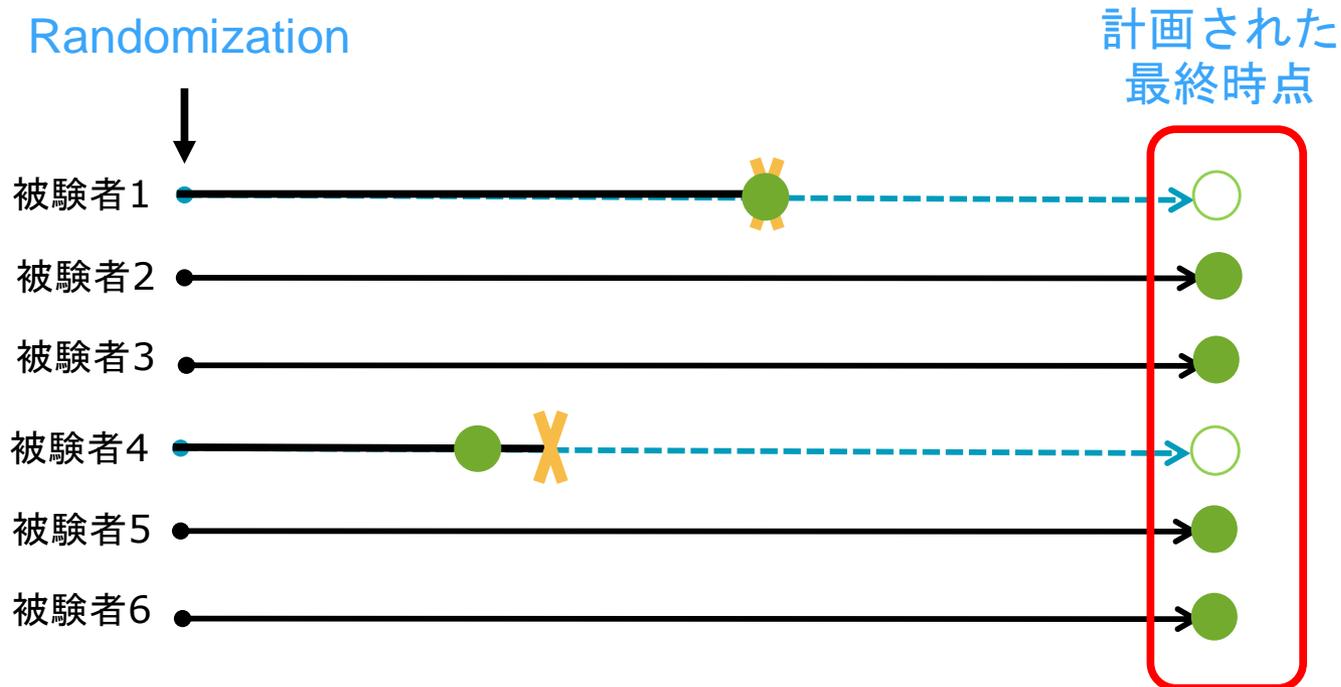
4. まとめ

- LAO(LOCF)/MMRMなど利用可能なデータのみを用いる解析の妥当性は, 何を推定したいかに依存
- 2つの具体例の違いは, 数学的な構造の問題ではなく臨床上の疑問の違い
 - 薬剤の薬理作用を検出することが目的の場合
 - 利用できる観測値の解析も主要な解析として受け入れられるかもしれない
 - 薬剤の処方と投与計画が役に立つかどうかを評価することが目的の場合
 - 試験治療を中止したという事実が, 中止前の観測値よりも意味がある
 - Utility estimandが必要になる可能性(3.3.2節参照)

- Permutt, T. (2016). A taxonomy of estimands for regulatory clinical trials with discontinuations. *Statistics in medicine*, 35(17), 2865–2875.
- 製薬協(2017). 臨床試験の estimand を理解するための因果推論. Available at: <http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/estimand.html>
[Accessed 2017/6/22]

補足資料

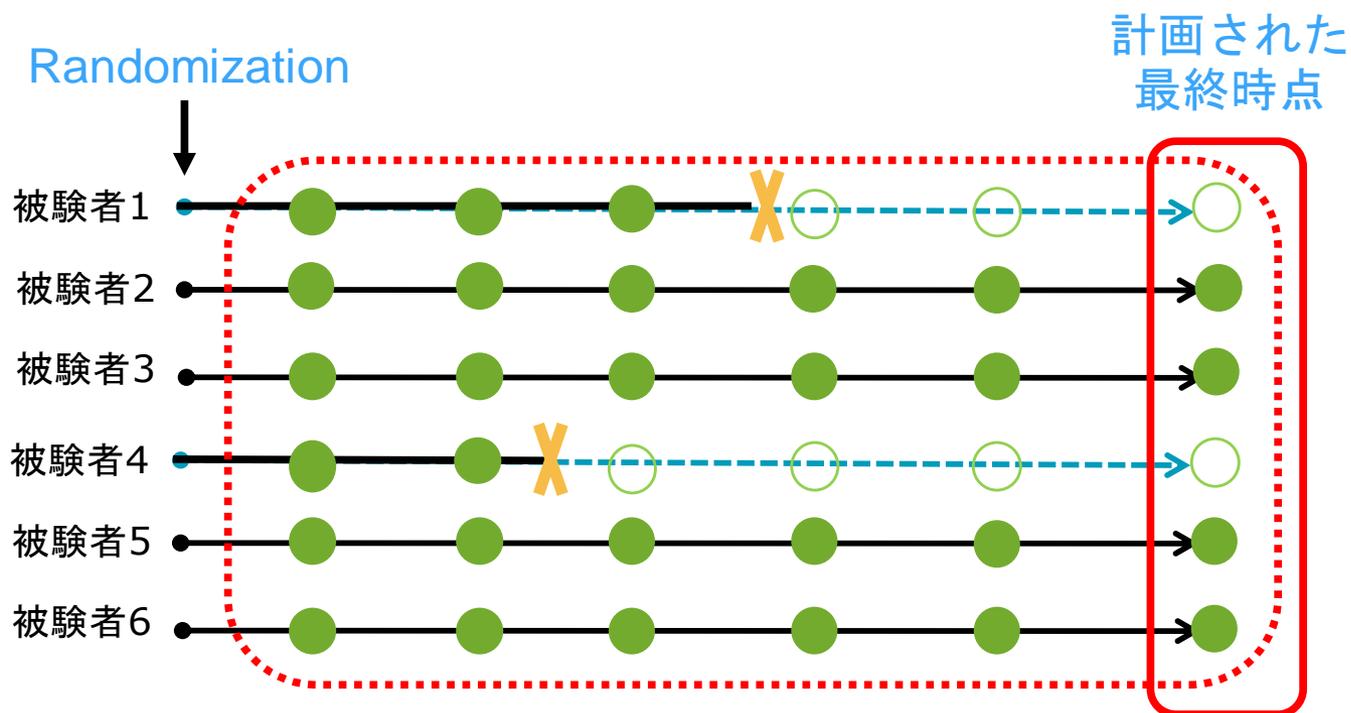
- 「試験薬投与中の最終時点の値を用いる」という解析方法は同じだが、推定したい対象が異なる



● の解析: 試験薬投与中の最終時点を推定 (LAO Estimand)

□ の解析: 計画された最終時点の推定 (LOCF Estimand)

➤ 複数時点のデータを用いて時点を通しての平均的な治療効果をモデル化



● : 推定に用いるデータ

時点の平均効果 \square または、最終時点の効果 \square の推定に利用可能
欠測も含め、時点間の共分散構造に対する仮定が必要

「長期投与のベネフィット」とは何かを定義する必要

- 投与を継続できた場合の効果
 - 投与を継続できる症例での薬剤の効果を比較
 - 主要層ストラテジー
- 投与を継続できるかどうか（中止した場合はベネフィットがない）
 - 中止を含めた複合エンドポイントの利用
 - 複合ストラテジー
- （実際の投与の継続の有無にかかわらず）一定期間後の効果の有無
 - 試験治療中止後のデータを収集
 - 治療方針ストラテジー