

【日本製薬工業協会シンポジウム】

臨床試験の欠測データの取り扱いに関する最近の展開と今後の課題について
—統計手法・estimandと架空の事例に対する流れの整理—

(5) Estimand の解説

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
欠測のあるデータに対する解析方法論・SASプログラム検討チーム
東レ株式会社 土居正明

本シンポジウムの概要(PM:セッション2) 臨床試験の計画から解析までの流れの整理



13:10～13:50 (5) Estimandの解説

13:50～14:15 (6) 架空の事例の検討1

estimandの特定・試験デザイン・主要評価項目の設定

質疑応答(14:15～14:25)

休憩(14:25～14:40)

← アンケート回収

14:40～15:20 (7) 架空の事例の検討2

主解析の選択・例数設計・データの発生方法

15:20～16:10 (8) 架空の事例の検討3

主解析・感度分析

休憩(16:10～16:30)

16:30～17:10 パネルディスカッション

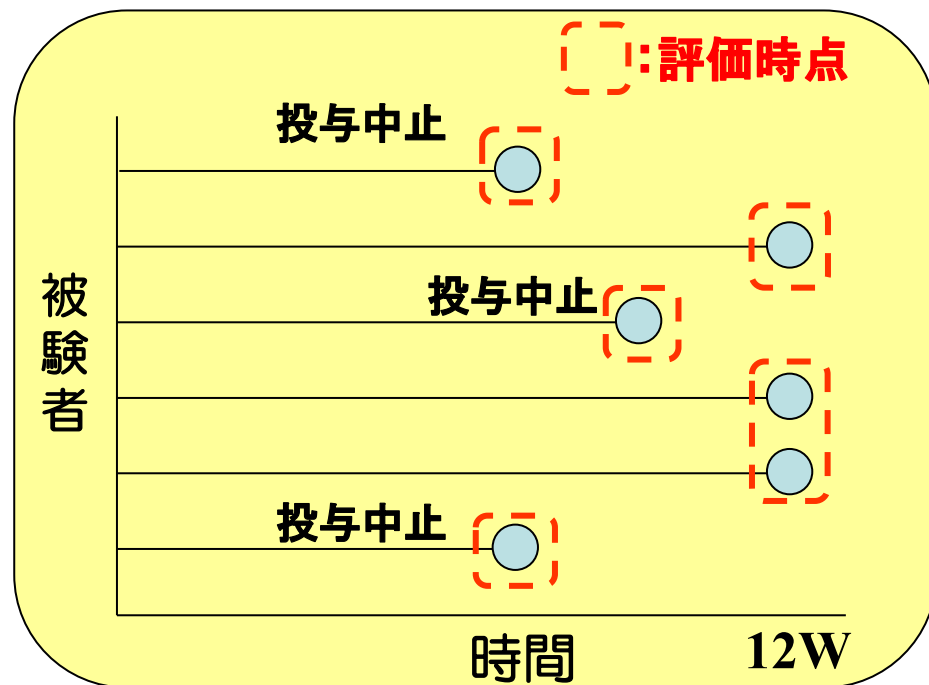
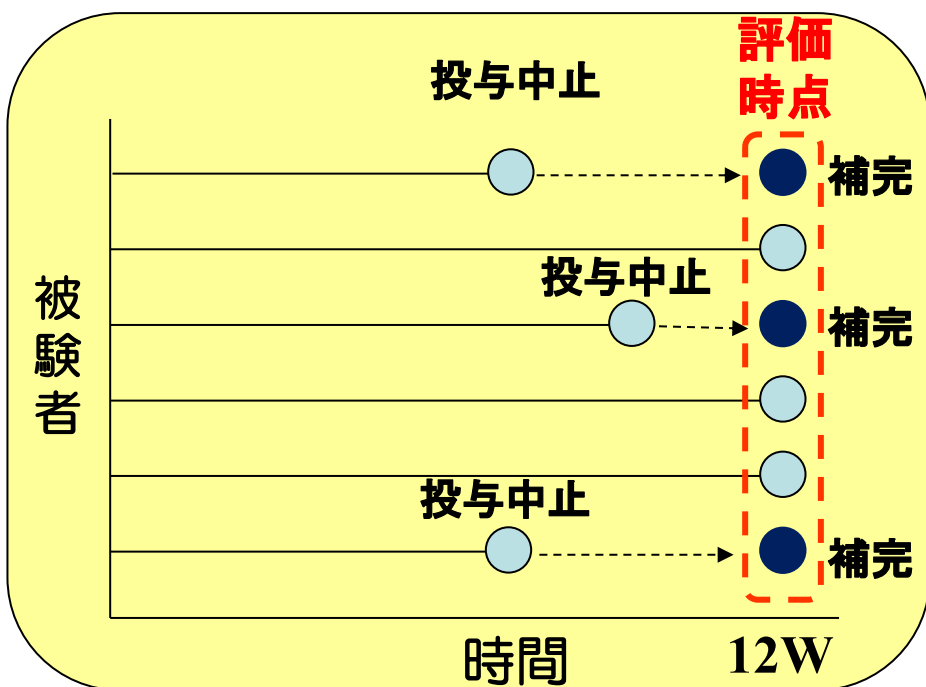
17:10～17:20 まとめ

1. Estimandとは
2. NRC (2010)とMallinckrodt (2013)の記載
3. 具体例からの検討
4. 注意点
5. 現段階で不明な点
6. 今後の展望

1. Estimandとは : introduction

- 試験1) ・12週投与. 12週で評価
- ・欠測はLOCFで補完
 - ・主解析: ANCOVA

- 試験2) ・12週投与の試験
- ・投与された最終時点で評価
 - ・主解析: ANCOVA



◎ 解析結果は同じ

◎ この2試験の違いは？

1. Estimandとは : introduction

試験1

試験2

評価時点

12週

12週までの最終時点

解析の妥当性

LOCFによって12週のデータを補完
→ **LOCFの妥当性の検討**が必要

全症例欲しいデータが取れている
→ 解析は妥当
※ ANCOVAの2直線の平行性等は仮定

結果の解釈

・12週時点の有効性が示される

- ・「12週時点の有効性」とみなすには **中止症例の影響の評価**が必要
- ・「最終時点が重要」と考えるなら **最終時点で評価することの妥当性の説明**が必要(計画段階)

解析結果(数値)は同じでも、
① 正当化すべき**解析の仮定**
② **結果から主張**できること
が異なる



「Estimandが異なる」
と呼んで、よく考えて
いきましょう

2. NRC (2010)とMallinckrodt (2013)の記載 Estimandの定義と説明



**Estimand = “what is being estimated”
(Mallinckrodt , 2013)**

◎試験で知りたいものを明確化する

NRC (2010)の説明

- ・Primary estimandをそのばらつきと共に適切に推定することが臨床試験の目的
- ・試験開始前(プロトコール完成前)に設定しておくこと

※ 試験の計画段階で

- 治験の目的に合致する estimand を設定
- estimand を妥当に推定するデザインを検討することが重要

2. NRC (2010)とMallinckrodt (2013)の記載 Estimandの分類



Efficacy: 計画通りに投与された薬剤の効果

→ per-protocol estimand (de jure)

Effectiveness: 実際に投与された薬剤(もしくは治療方針)の効果

→ ITT estimand (de facto)

NRC (2010) → 5種類

Mallinckrodt et al. (2012), Mallinckrodt (2013)

→ NRC + 1種類 = 6種類

Mallinckrodt et al. (2014)

→ Mallinckrodt et al. (2012)から3種類 pick up

2. NRC (2010)とMallinckrodt (2013)の記載 Estimandの構成要素



経時データの1時点を主要評価項目とする試験では, estimandに

- ・パラメータ(例. 平均の群間差)
- ・**時点または曝露期間(例. 投与期間8週目)**
- ・アウトカム(例. 拡張期血圧)
- ・**対象となる集団(例. 高血圧と診断された患者)**
- ・**投与中止後に治療(rescue medication)が行われた場合, その後に得られたデータは解析に含めるかどうか**

等が含まれる(Mallinckrodt, 2013)

2. NRC (2010)とMallinckrodt (2013)の記載

- 本発表では以下の状況を考える
 - 2群比較
 - プラセボ対照ランダム化比較試験
 - 優越性試験

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand (Mallinckrodt, 2013)

Estimand	仮説	推測の対象	被験者	評価時点	Rescue Medication後のデータ
1	Effectiveness	割り付け群 (治療方針)	全被験者	計画された 時点	主解析に 含める
2	Efficacy	最初に 割り付けられた 治療	最初の治療に 耐えられた被験 者のみ	計画された 時点	主解析に 含めない
3	Efficacy	最初に 割り付けられた 治療	全被験者	計画された 時点	主解析に 含めない
4	Effectiveness	最初に 割り付けられた 治療	全被験者	未定義	主解析に 含めない
5	Effectiveness	最初に 割り付けられた 治療	全被験者	未定義	主解析に 含めない
6	Effectiveness	最初に 割り付けられた 治療	全被験者	計画された 時点	補完することが 望ましい

※慢性疾患の第III相試験を想定

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: Estimand 1



• 全てのランダム化された症例の応答変数の改善の群間差

◎仮説: Effectiveness

◎推測の対象: 割付群 (治療方針)

◎対象集団: ランダム化された全症例

◎評価時点: 計画された時点

★ 投与中止後 or Rescue medication後のデータ

→ 必ずデータを取るデザインが必要

★ follow-upデータの欠測

→ Rescue medicationの使用有無別に”reference group”を作り、controlled imputationにより補完することも考えられる (Mallinckrodt, 2013)

(利点)

- ・実際の臨床に近いプロトコルが組めれば、**臨床現場に近い結果**が得られる

(欠点)

- ・評価対象が薬剤ではなく、治療方針
 - 治験で本当に知りたいこと?
- ・群間差が小さくなることも多いため、コストがかかる

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: Estimand 2



- 初期の治療に耐えられた被験者の応答変数の改善の群間差
 - ◎仮説: Efficacy ◎推測の対象: 最初に割り付けられた治療
 - ◎対象集団: **最初の治療に耐えられた被験者**のみ ◎評価時点: 計画された時点

★ **open label の active run in phase** で治療に耐えられる被験者を選択
その後ランダム化して、一部の症例をプラセボ群に変更
→ 中止確率減少

★ 投与中止後のデータは不使用

(利点)

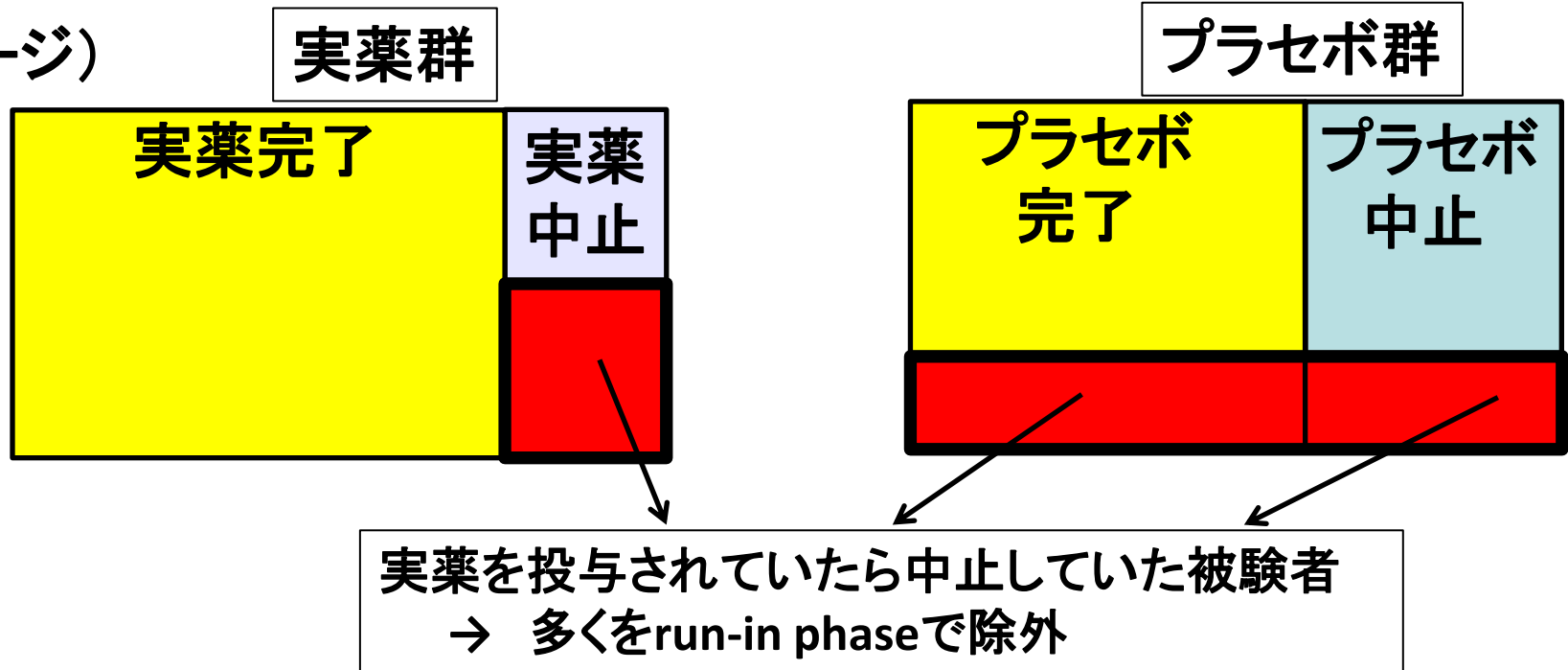
- ・欠測が減り、実薬群とプラセボ群の対等な比較ができる
- ・長期投与の評価に適している

(欠点)

- ・全ての被験者に対する評価ができない(一般化可能性が低い)
→ 臨床現場では誰が耐えられるか事前に分からない
- ・同一試験内で別のEstimandを検討することが難しい

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: Estimand 2

(イメージ)



(estimand 2における有効性)

= (実薬の投与に耐えられる症例の実薬投与後の有効性)

− (実薬の投与に耐えられる症例のプラセボ投与後の有効性)

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: Estimand 3

- 「全ての被験者が治療を完了した」と仮定した場合の応答変数の改善の群間差

◎仮説: Efficacy ◎推測の対象: 最初に割り付けられた治療

◎対象集団: ランダム化された全症例 ◎評価時点: 計画された時点

★全ての症例が完了したと想定

★投与中止後のデータは不使用

(利点)

- ・ **全ての症例が完了したと想定した場合の被験者の評価ができる**
→ 薬剤の有効性(efficacy)の評価ができる

(欠点)

- ・ 投与中止症例の投与中止後のデータの影響が入らない

※「投与中止後ベースラインに戻る」ような疾患・薬剤の場合、
投与中止後の影響を評価した解析との違いが大きくなる

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: Estimand 3



- NRC (2010)では「ほとんど使われない」と記載
 - 「全症例が治療を完了した」という仮定は現実的ではない。
 - 治療を継続させるような補助療法・支持療法が使用されることが前提
- Mallinckrodt (2013)では具体例の章(14章)で primary estimandとして使用
 - 「指示された通りに服薬した場合の薬剤の有効性」を薬効と考え, efficacy の評価が重要, と主張
 - secondary で effectiveness の評価も実施

◎ 考え方によって妥当性が変わる

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: Estimand 4

- 投与が続けられた範囲でのAUCの群間差

◎仮説: effectiveness ◎推測の対象: 最初に割り付けられた治療

◎対象集団: ランダム化された全症例 ◎評価時点: 未定義

(投与された最終時点)

(利点)

- ・データの欠測はない
- ・使用した期間と有効性を同時に評価できる

(欠点)

- ・「よく効いたが早期に中止」した症例と「効き目は少なかったが長く使い続けられた」症例が同等の扱い
- ・臨床的評価をしづらい

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: Estimand 5

- 投与された最終時点での応答変数の群間差

◎仮説: effectiveness ◎推測の対象: 最初に割り付けられた治療

◎対象集団: ランダム化された全症例 ◎評価時点: 未定義

(投与された最終時点)

(利点)

・データの欠測が発生し得ない

(欠点)

・「投与された最終時点」で評価することが妥当な疾患・薬剤にしか
使用できない

→ 適用可能な状況は多くない?

※ 慢性疾患等で、投与中止後に有効性が減弱する場合、
「計画された時点」の評価に(誤って)用いると、有効性が過大評価される

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: Estimand 6

- 計画された時点での、最初に割り付けられた治療による応答変数の変化の群間差

◎仮説: Effectiveness ◎推測の対象: 最初に割り付けられた治療

◎対象集団: ランダム化された全症例 ◎評価時点: 計画された時点

★投与中止後のデータ

→ rescue medicationなしの時点は使用
rescue medication使用後は不使用. 補完

◎「投与中止後は無治療」を想定＝「最初に割り付けられた治療」の評価

(利点)

- ・「中止後ベースライン値に戻る」ような薬剤に対して、中止症例の影響を合わせて評価

(欠点)

- ・補完方法によって結果が変わる
→補完モデルの妥当性の検討が必要

※ 完了例と中止例で経時推移が大きく異なる場合、
「全体の平均」の解釈には注意が必要

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: Estimand 6



- 補完の方法
 - BOCF
 - 中止後, 応答変数がベースライン近くに戻る薬剤など
 - pMI (placebo Multiple Imputation)
 - 「rescue medication使用前のデータ」+「rescue medication使用後はプラセボ群のデータを使用して補完」

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: estimand間の関係

- 評価したいものが
 - 割付群(治療方針): Estimand 1
 - 最初に割り付けられた薬剤: Estimand 6
 - どちらも(補完以外の)解析方法は同じ
- 有効性を評価したい対象集団が
 - 初期の治療に耐えられた被験者: Estimand 2
 - 全症例(治療を完了したと想定): Estimand 3
 - どちらも投与中止後のデータは使用しない
 - 解析手法は同じ
 - MARが成り立つと解析は容易
- 「プロトコールで規定された時点」以外で評価するとき
 - 時間と有効性を合わせて評価したい: Estimand 4
 - 投与された最終時点の有効性で評価したい: Estimand 5

2. NRC (2010)とMallinckrodt (2013)の記載 6種類のEstimand: その他のestimand



- **responder analysis**

- 「治療を完了」+「計画された時点で応答変数の値が閾値を超える」の複合エンドポイント
 - 上記条件を満たす”responder”の割合を評価
 - 投与中止は「治療失敗」とみて, ”non-responder”に
 - 欠測はない
- 投与中止しやすさは, 治験と臨床現場では異なることも多い
 - 投与中止を endpoint に含めるのは妥当か? (Mallinckrodt, 2013)

※NRC (2010)では estimand 5 に含まれている

※Mallinckrodt (2013) では estimand1~6 以外

3. 具体例からの検討

- 疼痛をベースに, Estimand 1~6を検討する
 - (Case 1)慢性疼痛
 - プロトコールで規定された時点の影響を評価したいと想定
 - (Case 2)がん疼痛
 - 短期でも(プロトコールで規定された期間投与が続けられなくても)痛みの緩和に意味があると想定

◎どちらも投与は12Wと想定

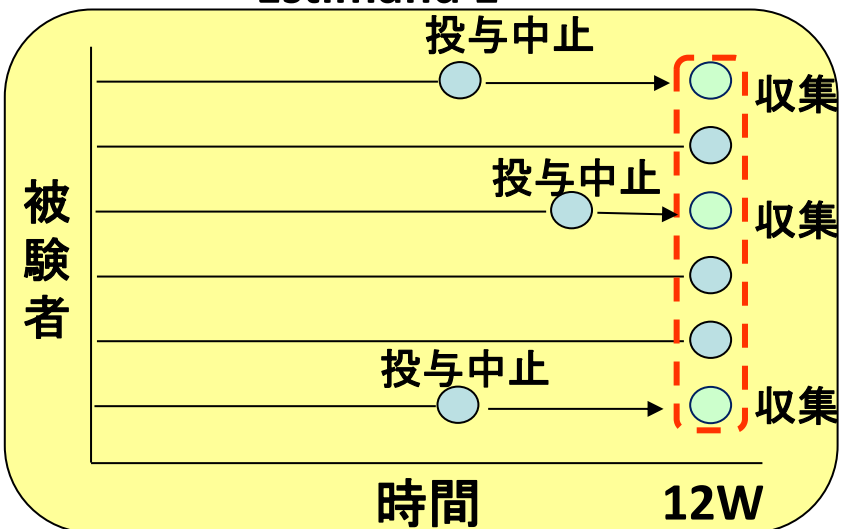
3. 具体例からの検討: 慢性疼痛

● (Case1) 慢性疼痛の治療薬の治験

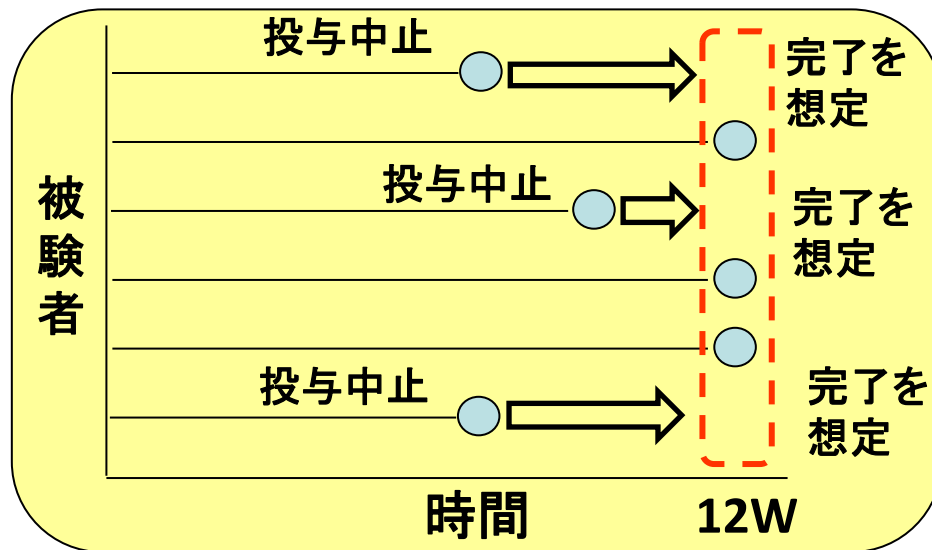
- 薬剤を正しく使った際の、薬剤の有効性が知りたい
 - Estimand 3
- (臨床現場に近い) 治療方針の有用性が知りたい
 - Estimand 1
- 最初に投与された薬剤の有用性が知りたい
 - 中止後にRescue Medicationがない状態(無治療)での計画された評価時点での有効性
 - Estimand 6
- 実薬の投与に耐えられる症例に対する薬剤の有効性が知りたい
 - Estimand 2
- 投与中止は「慢性疾患の治療」としては失敗と考え、有効な症例の割合を評価したい
 - Responder rate

3. 具体例からの検討: 慢性疼痛

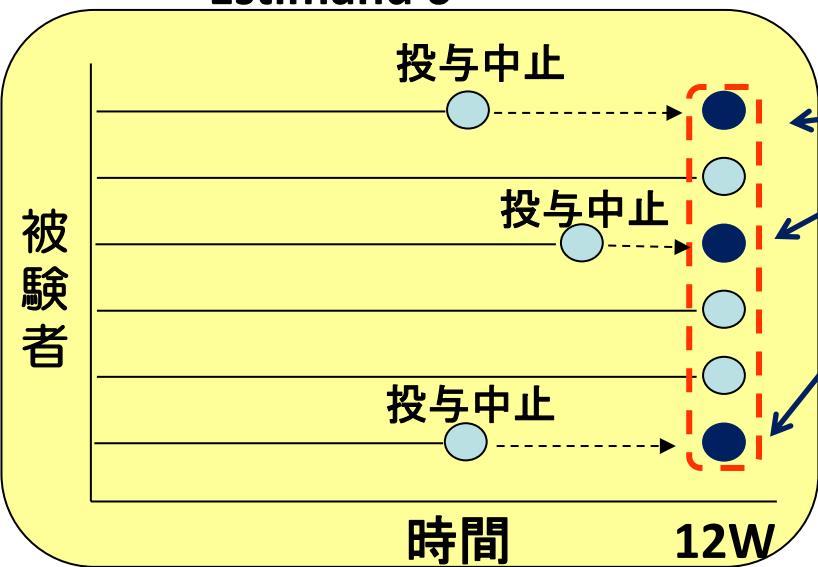
Estimand 1



Estimand 3



Estimand 6



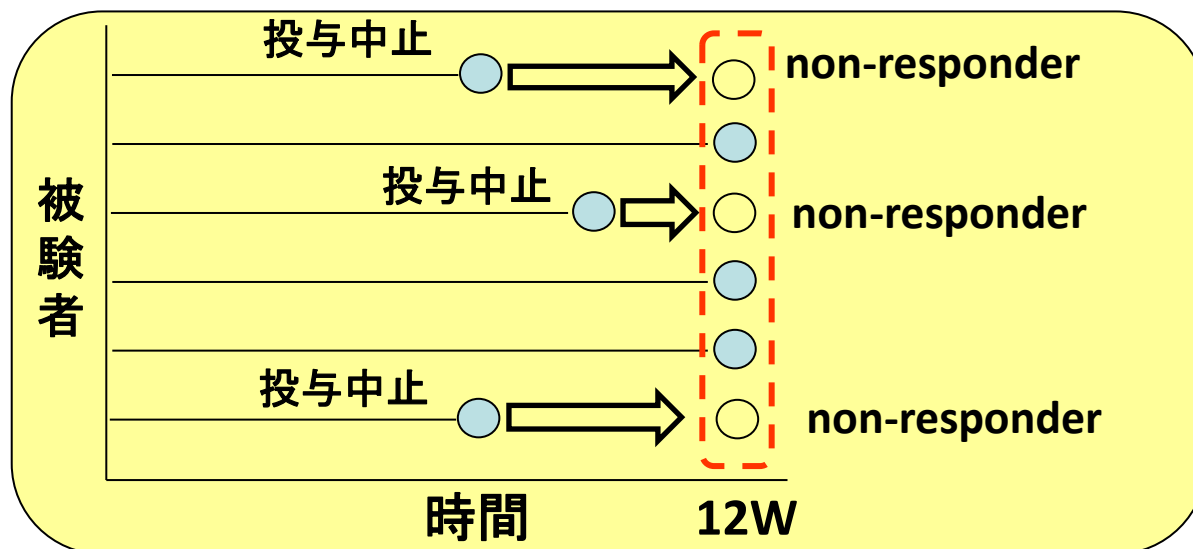
Rescue medication or 欠測が

- ・ある場合: 補完
- ・ない場合: データ使用

◎1試験で同時に検討可能

3. 具体例からの検討: 慢性疼痛

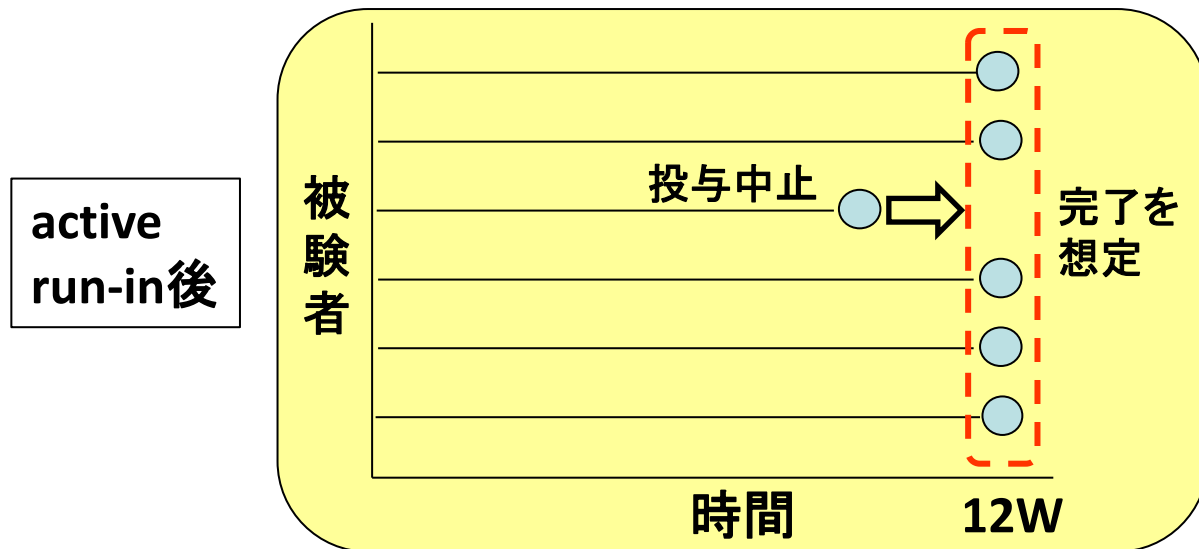
Responder analysis



◎1試験でestimand 1, 3, 6と同時に検討可能

3. 具体例からの検討: 慢性疼痛

Estimand 2



◎run-in phaseで対象を制限

⇒ 1試験内で他のestimandと同様に評価するのは難しい

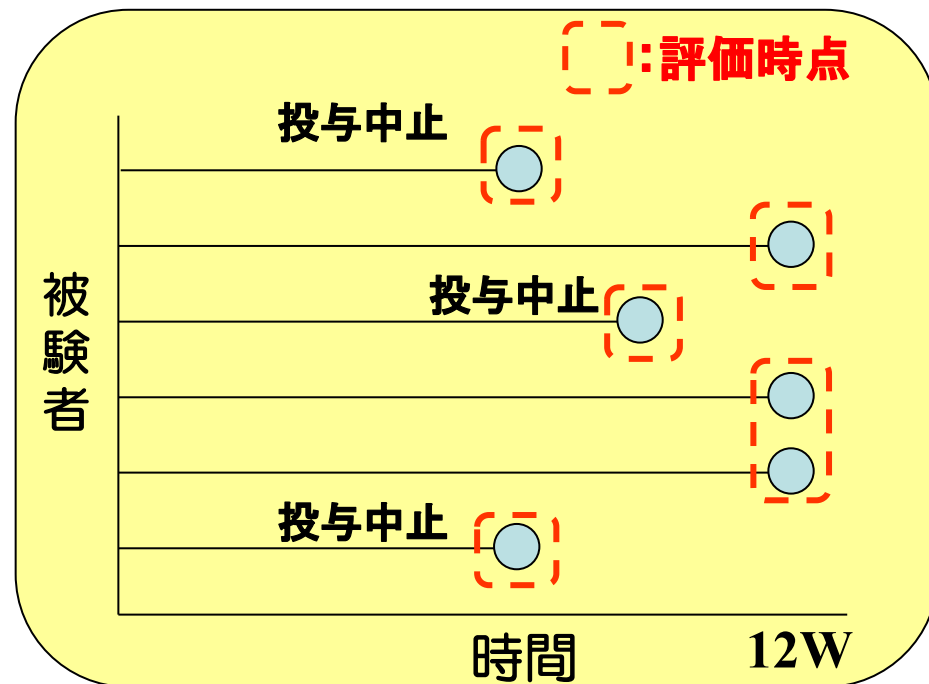
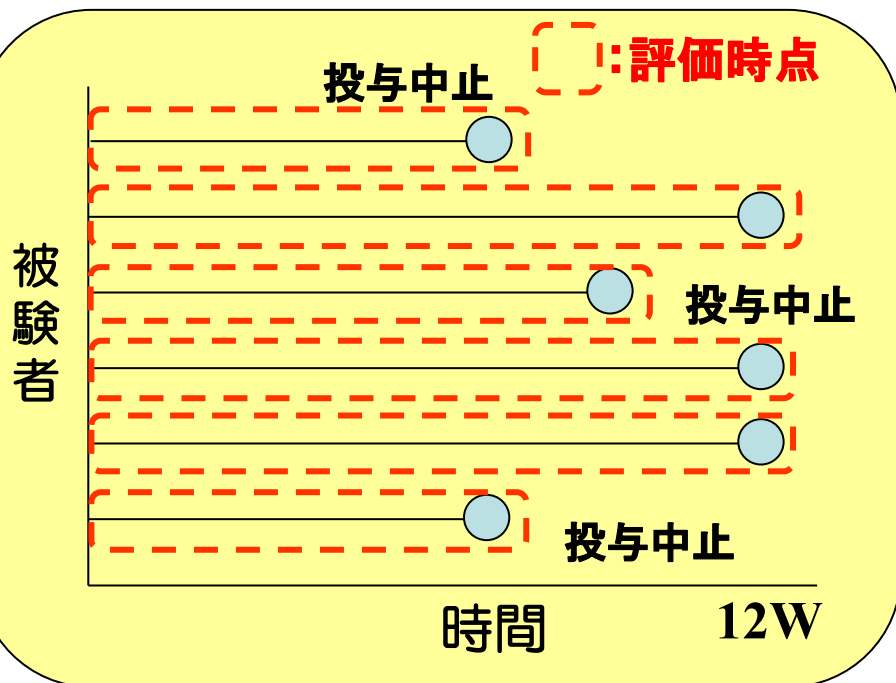
3. 具体例からの検討:がん疼痛

- (Case2)がん疼痛の治験薬の治験
 - 使用している期間と有効性を合わせて評価したい(AUC)
 - Estimand 4
 - どれだけの期間使用できるかは問題にせず, 投与された最終時点での評価がしたい
 - Estimand 5

3. 具体例からの検討:がん疼痛

Estimand 4

Estimand 5



◎1試験で同時に検討可能

3. 具体例からの検討:がん疼痛

- まとめ
 - Estimand 1～6, Responder rate は, 疾患等の状況次第で全てprimary estimand になりうる

4. 注意点

- 重要なことは「Estimand 1～6の中から選ぶ」のではなく、**試験ごとに適切な Estimand を考えること**
 - Estimand 1～6はあくまで例 .他のEstimandも定義可能.

5. 検討課題

- 薬剤の一生の中で、estimandをどう位置付けるか？
 - － 例.
 - 承認前
 - － 承認までに必要な情報は何か？そのためにどのような治験を計画すべきか？
 - 承認後
 - － 承認後の継続した薬剤のベネフィット・リスクの検討のために、どのような試験を計画すべきか？

◎それぞれの段階で、「適切なestimand」を
どのように決めていくか？

5. 検討課題

- Estimandは全て検定すべきか？
 - 「試験の目的」と考えるならDecision Makingベースの議論が必要？その場合は検定の多重性は？
 - Mallinckrodt (2013)ではprimaryのみ検定. secondaryは推定のみ
 - 感度分析全般も含め, 「複数の推定値を込みにして判断」するのであれば, 推定値ベースでのDecision Makingの定量的基準は不要か？
- 検証試験を複数行う場合Estimandは共通にするべきか？
 - 固定したEstimandに対する再現性をみるべきか？
 - 色々なEstimandを検討し, 広い知見を得るべきか
 - Estimand 2は単独では評価しにくい

5. 検討課題

- 複数のestimandの検討は「感度分析」か？
 - Mallinckrodt (2013)では感度分析に含めているが、妥当か？
- どこまで治験で行うべきか？
 - Estimand 1が示されれば望ましいが、コスト大. 治験の段階で必要か？
- 治験と臨床現場での治療の関係
 - どの程度一致を目指すべきか(一般化可能性)？
 - 治験の役割は？
- NRC (2010)では, (causal) estimand という表現がある
 - 「因果効果」である必要はどの程度あるか？

5. 検討課題

- 臨床現場で利用しやすい estimand と利用しにくい estimand
 - 例). AUC (estimand 4)は
 - 申請時: 投与期間と有効性を同時に評価でき, 便利
 - 臨床現場: 「自分がその薬剤を使えばよいか」にどう役立てる?

6. 今後の展望

- ICH E9 (R1)
 - **Choosing Appropriate Estimands** and Defining Sensitivity Analyses in Clinical Trials.

※) 本発表内容は、現段階のものであり、
今後大きく変更される可能性もあります

◎ 検討課題が多く、さらなる活発な議論が必要

1. Mallinckrodt, C. H. (2013). *Preventing and Treating Missing Data in Longitudinal Clinical Trials*. Cambridge Press.
2. Mallinckrodt, C. H., Chuang-Stein, C., Molenberghs, G., O’Kelly, M., Ratitch, B., Janssens, M., and Bunouf, P. (2014). Recent development in the prevention and treatment of missing data, *Therapeutic Innovation & regulatory Science* ,**48**, 68-80.
3. Mallinckrodt, C. H., Lane, P. W., Schnell, D., Peng, Y., Mancuso, J. P. (2008). Recommendations for the Primary Analysis of continuous Endpoints in Longitudinal Clinical Trial, *Drug Information Journal*, **42**, 303-319.
4. Mallinckrodt, C. H., Lin, Q., Lipkovich, I., and Molenberghs, G. (2012). A structured approach to choosing estimands and estimators in longitudinal clinical trials, *Pharmaceutical Statistics*, **11**, 456-461.
5. National Research Council. (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press.