

Estimand

【第5回データサイエンスラウンドテーブル会議】

2018年2月27日

浅野 淳一 (医薬品医療機器総合機構)
平川 晃弘 (東京大学)
横山 雄一 (持田製薬株式会社)
渡邊 大丞 (サノフィ株式会社)

本日のスケジュール

時間	内容
10:00~10:20(20分)	開会式、会議室の部屋割りの説明
10:20~10:30(10分)	会議室の仕切りの設置＋各部屋の準備
10:30~12:30(120分)	午前のディスカッション
10:30~10:40(10分)	ディスカッションの方法・テーマ・論点の紹介・説明
10:40~12:00(80分)	各テーブルで自己紹介 テーマ1~3の議論(休憩は適宜) 発表準備時間も含む
12:00~12:30(30分)	各テーブルの発表(質疑応答を含む)
12:30~13:30(60分)	昼休憩
13:30~16:00(150分)	午後のディスカッション
13:30~13:40(10分)	ディスカッションの方法・テーマ・論点の紹介・説明
13:40~16:00(140分)	テーマ4の議論(休憩は適宜)
16:00~16:15(15分)	会議室の仕切りを除去＋発表準備＋小休憩
16:15~17:30(75分)	ポスター形式による発表
17:30~17:40(10分)	閉会式

ディスカッションテーマ

- 午前（テーマ1は必須）
 1. 中間事象に対応するためのストラテジーの選択
 2. 非劣性を示すことを目的とした試験におけるestimandの選択
 3. 国際共同試験におけるestimandに影響を与えるような地域間差についての解釈
- 午後
 4. Estimand、感度分析及び補足的解析を意識した架空の臨床試験の試験計画

ICH E9(R1) step 2ドキュメント

【お断り】

本資料は、2017年9月13日に発出された「ICH E9(R1) 臨床試験のための統計的原則 補遺 臨床試験におけるestimandと感度分析(案)」をもとに作成した。
今後、パブリックコメントを経てICH E9(R1)が最終化され、用語又は内容の変更がされることもある。

ICH E9(R1) step 2ドキュメント 目次

A.1. 目的と適用範囲

A.2. 計画、デザイン、実施、解析及び解釈を整合させるフレームワーク

A.3. Estimand

A.3.1. 解説

A.3.2. 中間事象に対応するためのストラテジー

A.3.3. Estimandの構成

A.3.3.1. 一般的留意事項

A.3.3.2. 治療背景及び試験の実験的側面に関する留意事項

A.4. 試験デザイン及び実施への影響

A.5. 試験の解析への影響

A.5.1. 主とする推定

A.5.2. 感度分析

A.5.2.1. 感度分析の役割

A.5.2.2. 感度分析の選択

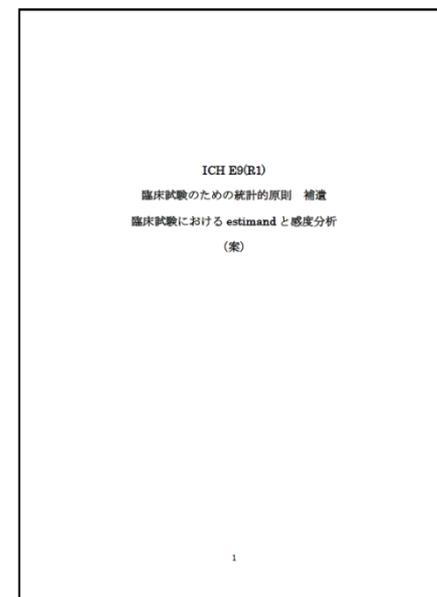
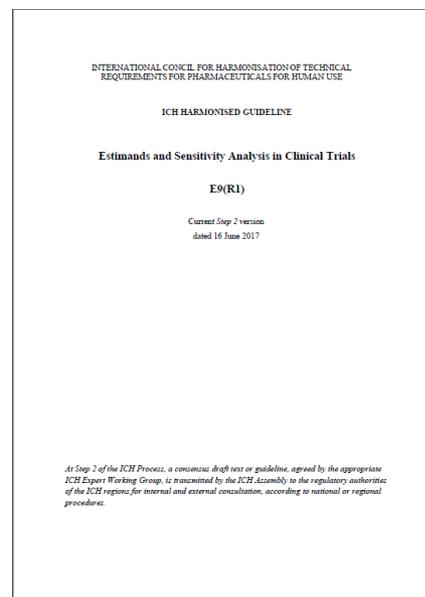
A.5.3. 補足的解析

A.6. Estimandと感度分析の記載

A.7. 一般的な例

A.7.1. 中間事象が一つの例

A.7.2. 中間事象が二つの例



目的と適用範囲(A.1)

多くのランダム化比較試験として実施される検証試験は、**治療効果を定量化**し、規制上の意思決定の裏付けとなる有効性及び安全性の証拠を提示するために実施される。

A.3.1. 解説

医薬品開発と承認審査において中心となる課題は**治療効果の定量化**である。すなわち、**試験治療による結果と、異なる治療条件下**(例えば、試験治療を受けなかった場合や、異なる治療を受けた場合)で、**同じ被験者に起きたかもしれない結果との比較**である。

ランダム化比較試験、ベースラインにおける交絡は回避できるはずだが、**臨床現場と同様に、試験においても治療効果の説明と解釈を複雑にする特定の事象**が発現することがある。これらの事象を**中間事象**(用語集参照)と呼ぶ。

中間事象には特に、**代替治療**(レスキュー薬、試験実施計画書における併用禁止薬、又は後続ラインの治療)**の使用、治療の中止、治療の切り替え**、状況によっては死亡といった**終末事象**が含まれる。

先に中間事象の発現に対応することなく、有効性及び安全性の変数、データ収集の基準、及び統計解析の手法を選択し定義することは、**推定される治療効果の曖昧さや、推定される治療効果と試験の目的との潜在的な不整合**に繋がる。

計画、デザイン、実施、解析及び解釈を統合させるフレームワーク(A.2)

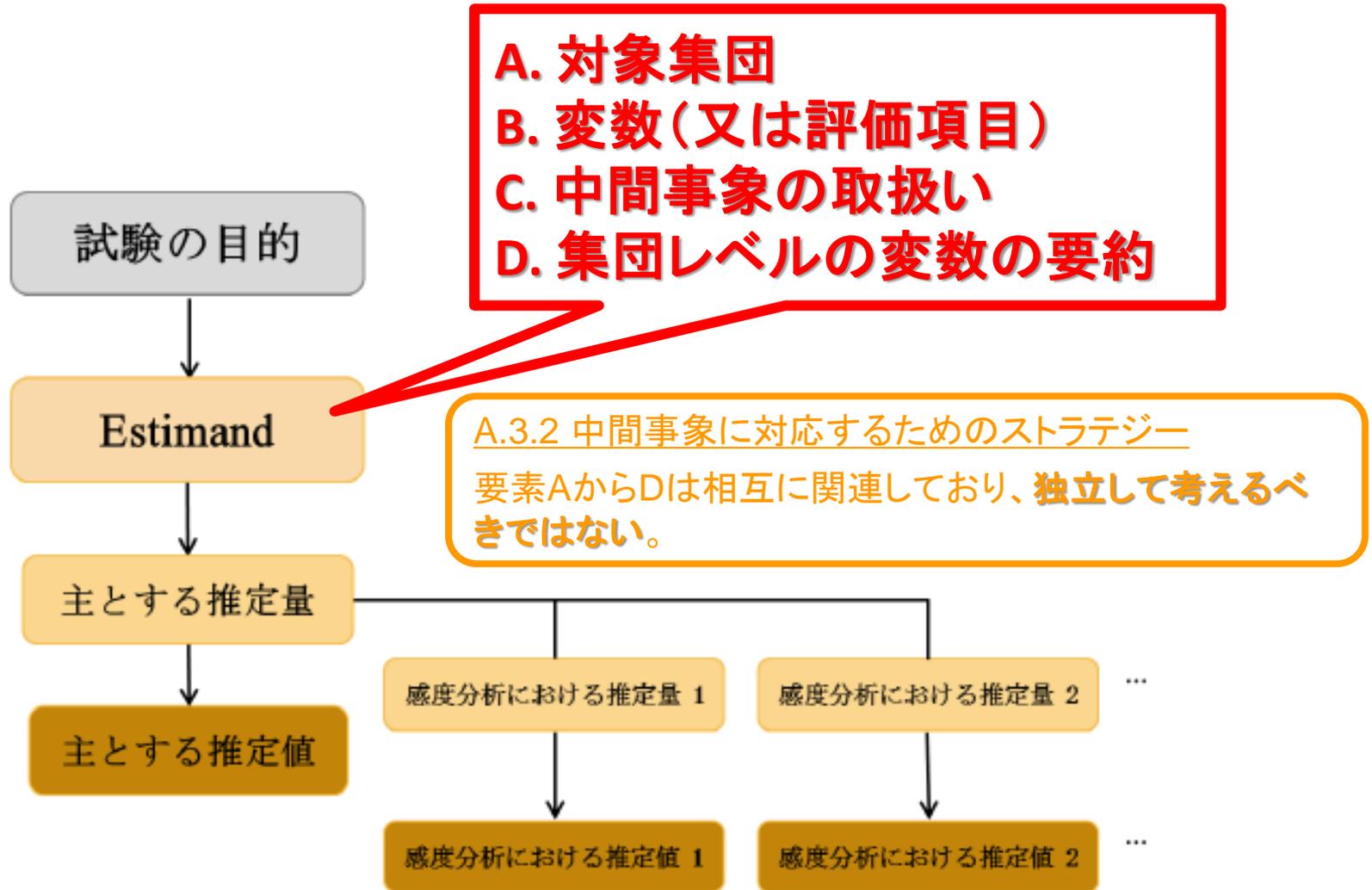


図1: 定められた試験の目的に対して、推定の対象、推定の方法及び感度分析を統合させる

Estimandの構成 (A.3.3)

Estimandの構成には、手順の繰り返しが必要となるかもしれない。Estimandの構成は、**試験の対象疾患や治療目標を含む個々の治療状況**、及び**個々の試験設定における臨床的な妥当性**を考慮して、正当化されるべきである(A.3.3.2節)。

A.3.3.2. 治療背景及び試験の実験的側面に関する留意事項

疾患の状況と治療の目的の特徴は**estimandの構成**に影響を与える。治療背景については、代替治療の選択肢があるかどうか、治療に対する個々の反応を観察できるかどうか、そして治療の目的が症状のコントロールか、疾患経過の修飾か、又は疾患の予防なのかがそれぞれ含まれる。

さらに推測のために信頼できる推定値が得られることを保証するために、**試験デザインと統計手法の妥当性**について検討する必要がある。特に、臨床試験におけるランダム化の重要な利点を認識し、可能な限り利用すべきである。

Estimandの構成 (A.3.3)

いくつかのestimand、特に**観測されたデータを使って推定するestimand**は、少ない仮定で**安定した推定**を行うことができるが、**他のestimand**は、正当化することがより困難で、起こりそうな仮定の変化に対してより影響を受ける可能性がある特定の仮定を必要とする (A.5.1 節)。

A.5.1 主とする推定

安定した推定値が得られるかどうかに関する重要な判断は、**必要となる仮定の範囲**でのものである。主とする推定量及び感度分析における推定量について、必要となる仮定を明確に説明すべきである。その仮定は正当化できるべきであり、ありそうにない仮定は避けるべきである。**仮定に対する結果の安定性は、同じestimandに対する感度分析**によって評価されるべきである (A.5.2 節)

適切な試験デザインの策定や、特定のestimandに対する信頼できる推定値の導出に重大な問題が存在する場合には、別のestimand、試験デザイン、及び解析手法について検討する必要があるだろう。

感度分析(A.5.2)

特定のestimandに基づく推測は、**データの限界と、主とする推定量に対する統計モデルで用いる仮定からのずれに対して安定しているべきである**。この安定性は**感度分析**によって評価される。

主とする推定量を裏付ける**統計的な仮定**は明記すべきである。そして、主とする推定量によって算出された推定値が**その仮定からのずれに対して安定しているかを確認する目的で、同じestimandを対象とした一つ以上の解析を事前に規定すべきである**。

感度分析とは異なり、**試験のデータをより十分に評価し、理解するために計画、提示又は要求されるその他の各解析は補足的解析**(用語集参照)と呼ぶ。それぞれの補足的解析は異なるestimand、又は同じestimandに対する異なる推定量に関連させることができるだろう。

ICH E9: 1.2. 適用範囲と方向性

安定性(ロバストネス)とは、データ、仮定及び解析方針についての様々な制限に対して全体の結論がどの程度変わり易いかに関連した概念である。安定性は、**異なる仮定又は異なる解析方針に基づいて解析を行った場合でも、試験治療の効果と試験の主要な結論は大きく影響されないこと**を意味している。試験治療の効果と試験治療の比較における不確実さに関する統計的指標の解釈には、p値、信頼区間又は推測に偏りが与える影響を考慮に入れるべきである

テーマ1

中間事象に対応するための戦略の 選択

中間事象に対応するための戦略

治療方針戦略

中間事象の発現は問題としないものとする、つまり、**中間事象の発現の有無に関わらず、関心のある変数の値を用いる。**

複合戦略

中間事象の発現を変数の構成要素とする、つまり**中間事象を一つ以上の他の臨床結果の指標と統合して関心のある変数とする。**

仮想戦略

中間事象が発現しなかった状況を想定する。関心のある科学的疑問を反映する値は、**定義された仮想的な状況において得られたであろう変数の値**であるとする。

中間事象に対応するためのストラテジー

主要層ストラテジー

対象集団を、**中間事象が起こらないであろう主要層**（用語集参照）とする。例えば、関心のある対象集団は、治療計画の不遵守が起こらないであろう患者層であるとされるかもしれない。言い換えると、主要層はより広い集団のうち中間事象の発現がないであろう部分集団となる。関心のある科学的疑問は、その層のみでの治療効果に関連する。

治療下ストラテジー

中間事象の発現前までの治療に対する反応に関心の対象とする。変数が繰り返し測定される場合、中間事象の発現までの変数の値は、全ての被験者に対して同様に設定された時点における値よりも、中間事象の発現を考慮するものであると考えられるかもしれない。

A.4. 試験デザイン及び実施への影響

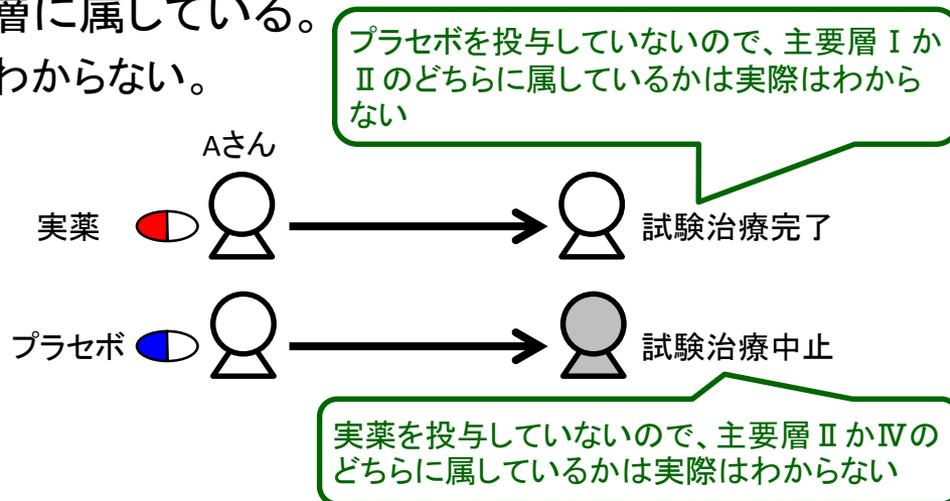
中間事象を取り扱うストラテジーの記述による関心のある治療効果の詳細な説明は、被験者数の計算に対して情報を与えるべきである。**全ての被験者が解析に対する情報に寄与し、中間事象とその取扱いの影響が対象とする効果の大きさと予期される分散に反映されるのであれば、通常は算出された被験者数に想定される中止例の割合に応じた上乘せをする必要はない。**

主要層ストラテジー(例)

- 治療群: 実薬、プラセボ

- 割付け通りに治療を受けるとする(全被験者がcomplierである)。
- 各被験者は以下のいずれかの主要層に属している。
 - ただしどの主要層に属しているかはわからない。

主要層	実薬	プラセボ
I	試験治療完了	試験治療完了
II	試験治療完了	試験治療中止
III	試験治療中止	試験治療完了
IV	試験治療中止	試験治療中止



- 対象集団

- いずれの治療でも試験治療を完了するであろう被験者⇒ **主要層 I**
- プラセボを投与した場合の試験治療を完了するであろう被験者⇒ **主要層 I + III**
- 実薬を投与した場合の試験治療を完了するであろう被験者⇒ **主要層 I + II**

主要層ストラテジー

主要層に該当する被験者かどうかは、不完全となる場合が多いものの、共変量によって推論しなければならない。一方で、割付けられた治療を受けている状況での中間事象の発現に基づいて主要層に該当する被験者を特定すると、その集団に基づく解析による治療効果の推定には、交絡が生じる恐れがある。なぜなら、異なる被験者は、異なる治療で異なる中間事象を経験することになるからである。

ディスカッションポイント

- 適用範囲(疾患領域、薬剤等)

- 被験者数の決定

- 主解析・感度分析・補足的解析の選択

- 試験デザインにおける留意点

- 終末事象(例えば、死亡)の取り扱い

- 試験計画時に終末事象(例えば、死亡)を想定すべきか。

- 中間事象に対応するためのストラテジーに関する論点

- 治療方針ストラテジー

- ✓ 例) 試験中止(欠測データ)が発生した際の対応

- 複合ストラテジー

- ✓ 例) 「中間事象の発現は極端に好ましくない値」の設定根拠

- ✓ 例) 連続量を二値化するための閾値を変えた解析は感度分析になるのか？

- 仮想ストラテジー

- ✓ 例) 「仮想的な条件」をどのように正当化するのか？

- ✓ 過去に経験した試験を、ICH E9 (R1) step 2 ドキュメントに従い計画すると、何が変化する/しない。
- ✓ 過去に経験した試験のestimandは何か。

テーマ 2

非劣性を示すことを目的とした試験におけるestimandの選択

ICH E9: 5.2.3 二つの異なる解析対象集団の役割

最大の解析対象集団 (*Full Analysis Set; FAS*) と **治験実施計画書に適合した対象集団** (*Per-Protocol Set; PPS*) は、(被験薬が優っていることを示そうとする) 優越性試験と(被験薬が匹敵しうることを示そうとする、3.3.2節参照) 同等性試験又は非劣性試験とで異なる役割を果たす。

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。なぜならば、**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、**同等性試験又は非劣性試験では、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、その役割は十分慎重に考慮すべきである。

ICH E9 (R1) step 2ドキュメントの記載(A.3.3.2)

治療方針ストラテジーを用いて一つ以上の中間事象を考慮するように構成されたestimandでは、非劣性及び同等性の試験に対してFASに関する同様の問題が生じる。割付けられた治療の中止やその他の治療の使用の後には、最初に割付けられた治療の類似性とは関係のない理由で、**両治療群間の反応がより類似して見えることがある。**

Estimandは**治療群間の差の過小評価に繋がる可能性がある中間事象**(例えば、レスキュー薬の使用や対象集団からの違反)に直接的に対応できるように構成することもあり得る。この状況では、estimandは、**治療間で差があるのであればそれを高い感度で検出できるような治療効果の指標**を対象とすることになるだろう。

ICH E10 1.5 分析感度

「分析感度」(assay sensitivity)とは、**有効な治療と有効性の低いあるいは無効な治療とを区別する力**として定義される臨床試験の特性である。(中略)一方、実薬対照に対する被験治療の非劣性を証明することにより有効性を証明しようとする試験に分析感度がない場合には、無効な治療が「劣っていない」と判断され、誤って有効であるとの結論に至ることがありうる。

ディスカッションポイント

- 治療群間の差の過小評価に繋がる可能性がある中間事象としてどのような事象があるか？
 - レスキュー薬の使用
 - 対象集団からの違反
 - ...
- 治療群間の差の過小評価に繋がる可能性がある中間事象に対応する戦略として何が有効であるのか？
 - 治療方針戦略
 - 複合戦略
 - 仮想戦略
 - 主要層戦略
 - 治療下戦略

テーマ 3

国際共同試験におけるestimandに影響を与えるような地域間差についての解釈

ICH E17: Overall Sample Size (拔粹)

The **primary objective** of an MRCT generally corresponds to an evaluation (estimation and testing) of the treatment effect averaged across all subjects in all regions of the MRCT. **The overall sample-size is determined to ensure that this objective can be met.**

The same general principles provided in ICH E9 for determining sample sizes of clinical trials apply to MRCTs.

Two additional factors are particularly important in the MRCT setting; **(i) the size of the treatment effect that is considered clinically relevant to all regions in the trial**, and **(ii) the expected variability of the primary outcome variables based on combining data across regions**. These factors may result in a sample size increase for an MRCT compared to a single trial in one region.

ICH E17: Sample Size Allocation to Regions (拔粹)

There is no uniformly acceptable or optimal approach to sample size allocation in an MRCT. Some approaches currently in use include:

- 1. Proportional Allocation:** Allocation of subjects to regions in proportion to size of region and disease prevalence.
- 2. Equal Allocation:** Allocation of equal numbers of subjects to each region.
- 3. Preservation of Effect:** Allocation of subjects to one or more regions based on preserving some specified proportion of the overall treatment effect.
- 4. Local Significance:** Allocation of a sufficient number of subjects to be able to achieve significant results within each region.
- 5. Fixed Minimum Number:** Allocation of a fixed minimum number of subjects to a region.

Note that the five approaches discussed above are not exhaustive. New approaches for sample size allocation in MRCTs may be developed in the future, and innovation in this area is encouraged.

ICH E17: Primary endpoint (抜粋)

Agreement on the primary endpoint ensures that the overall sample size and power can be determined for **a single (primary) endpoint based on the overall population** and also agreed upon by the regulatory authorities.

If agreement cannot be reached due to well-justified scientific or regulatory reasons, a single protocol should be developed with endpoint-related sub-sections tailored to meet the respective requirements of the regulatory authorities.

In this case, because regulatory approvals are based on **different primary endpoints by different authorities**, **no multiplicity adjustment is needed for regulatory decision-making**.



Regulatory region間で主要評価項目〔estimandの構成要素である”変数(variable)”〕が異なる可能性がある。

ICH E17: Collecting and Handling of Efficacy and Safety Information (拔粹)

Coordinated site initiation is particularly important in MRCTs to ensure **proper conduct, completion** and **reporting of results** without any delays among regions.

Centralised and risk-based monitoring may be particularly useful for MRCTs to identify variability across regions and sites in protocol compliance (e.g., differences in follow-up, compliance with study medications, adverse event reporting and/or **extent of missing data**). Mitigation approaches should take regional variations into consideration.

Careful attention to quality during trial planning, investigator training and trial monitoring will help achieve **the consistently high trial quality** required for a successful MRCT.

ディスカッションポイント

- 試験全体の主要評価項目と異なる主要評価項目を要求された場合、該当する地域の被験者数の決定とestimandの定義
- 国際共同試験の地域間で中間事象の発現に乖離があった場合の試験結果の解釈
 - 発現割合(例えば、有害事象による試験治療中止)が地域間で異なる。
 - 地域間で中間事象の発現に乖離が想定される場合に、どのようにestimandを定義すべきか？

テーマ4

架空の臨床試験の計画

試験デザイン

- 対象疾患:2型糖尿病
- 主目的
 - 被験薬ヨクナールを24週間単剤投与したときの有効性及び安全性の検討
- 試験デザイン
 - ランダム化 二重盲検 プラセボ対照 比較試験
 - 観察期間24週

ヨクナールの特徴

- 1日1回の経口剤(1用量)
- 新規作用機序の薬剤で、強い血糖降下作用を持つ。
- 薬剤の位置づけとしては、インスリンと同等
 - インスリン皮下投与が必要な人が、本薬剤への切り替えを狙っている。
 - ヨクナールだけで、正常な人のインスリンの体内分泌の再現することが期待される。
- 強い血糖降下作用のため、低血糖が副作用として生じやすい。
 - 低血糖は重篤な中枢神経系の症状を起こしうる危険な事象

試験治療の中止、レスキュー薬の使用

- 試験治療の中止割合

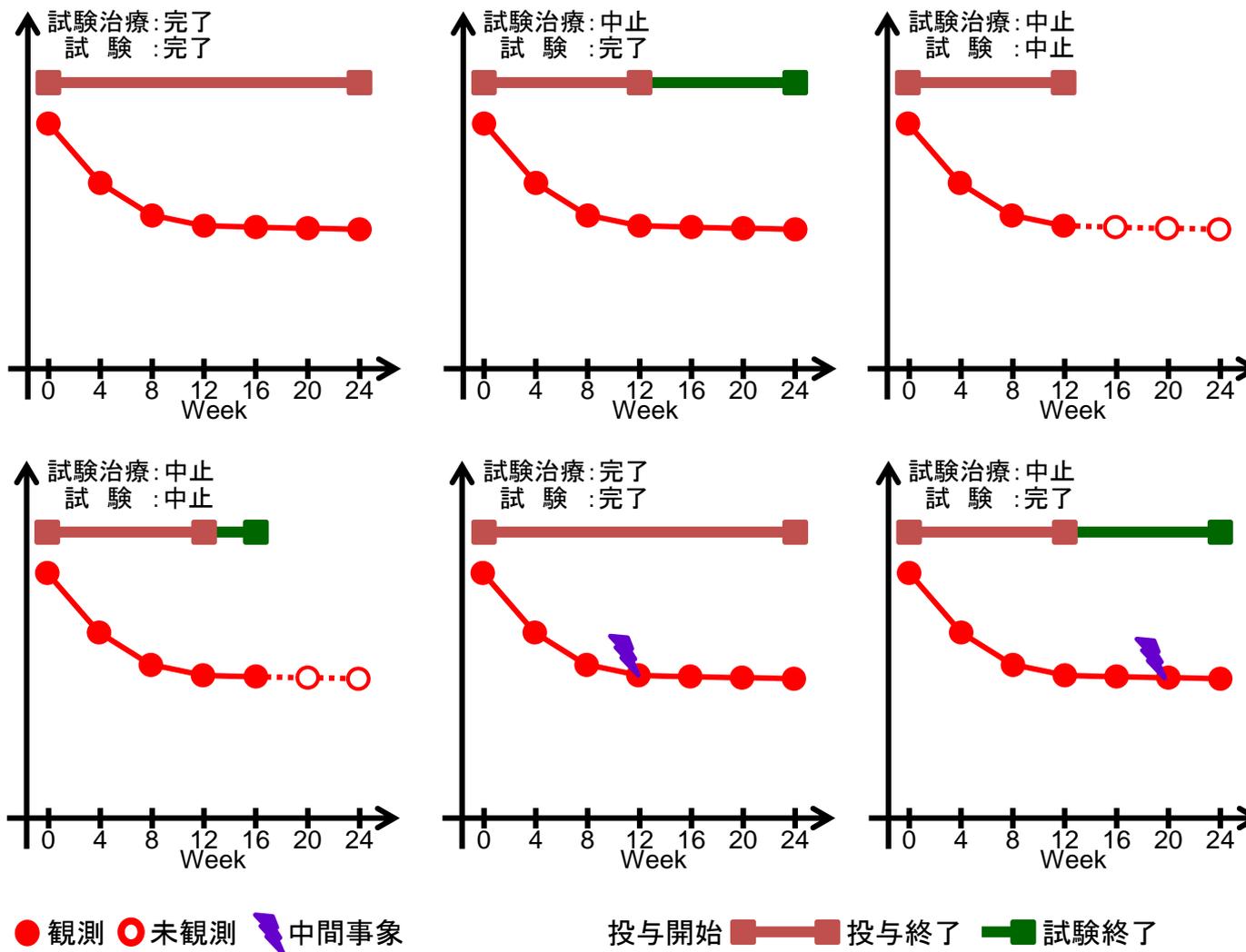
		ヨクナール群	プラセボ群
全体		41%	23%
有効性の欠如		0%	5%
有害事象	低血糖	20%	0%
	低血糖以外	10%	5%
追跡不能		5%	5%
心血管イベント又は死亡		1%	3%
その他		5%	5%

- 24週までのレスキュー薬の使用割合
 - ヨクナール群5%、プラセボ群20%
- 追跡不能, 心血管イベント又は死亡以外の試験治療中止後は薬物治療を行う

議論のポイント

- Estimandの4つの構成要素の設定
- データの収集範囲
- 主解析の計画、欠測データの取り扱い
- 感度分析・補足的解析の計画
- 被験者数の設定

想定されるHbA1cの推移



血糖降下薬の臨床評価方法に関するガイドライン(改訂案)

3-3-1-1 無作為化二重盲検並行群間比較試験(抜粋)

目的

第III相試験は、第II相試験により明確にされた適応、用法・用量等に基づいて、治験薬の有用性をより客観的に検証することを目的とする。このため、適切な対照薬を選び二重盲検法による並行群間比較試験を行う。

対象

原則として第II相試験と同様に他の血糖降下薬を投与されていない2型糖尿病患者で、状態が安定した成人を対象とするが、承認後に医療現場で投与対象と想定される患者を選択する。

評価項目

原則として主要評価項目にはHbA1cを用いる。

試験期間

投与期間は治験薬の有効性、安全性を評価するに足る十分な期間が必要である。HbA1cを主要評価項目とする場合は、少なくとも12週は必要であり、原則として24週が望ましい。また、適切な観察期間も設定する。

試験計画

対照薬

試験計画(実施)時点において、わが国で臨床的評価が確立していると考えられる既存の治療薬の中から適切な薬剤を対照薬とする。ただし、治験薬の特性等を考慮して比較対照となる適切な治療薬がないと考えられる場合、プラセボを用いることも可能である。

試験症例数

プラセボに対する優越性試験、既存の治療薬に対する非劣性又は優越性試験のいずれの場合であっても、統計学的な観点から、仮説を検証するために適切と考えられる症例数を設定することが必要である。また、治験薬の安全性についても評価しうる症例数が必要である。