

【日本製薬工業協会シンポジウム】
ランダム化比較試験における統計的効率改善のための共変量調整

パネルディスカッション
(議論のまとめ含む)



2026年3月16日
医薬品評価委員会 データサイエンス部会
2025年度タスクフォース3-1

- 本パネルディスカッションにおける発言内容は、ディスカッション時点での各登壇者の個人的な見解に基づくものであり、登壇者が所属する組織または規制当局の公式な見解や方針を示すものではありません
- また、本ディスカッションにおける発言は、特定の個人または組織の見解として帰属されるものではありません

参加者

役割	所属	氏名
司会	TF3-1	土居 正明
パネリスト	PMDA 新薬審査第5部 主任専門員	原 綾子 先生
	東京医科大 主任教授	田栗 正隆 先生
	TF3-1	飯田 涼介
		大野 浩太
		澤本 涼

パネルディスカッションの概要

➤ 目的

1. 論点の整理
2. 今後の方向性に関する議論

➤ 論点

- 論点1: 条件なし治療効果と条件付き治療効果の比較
- 論点2: 条件なし治療効果の推定方法の検討
- 論点3: Super covariates の使用に関して

スライドの分類:

論点
整理

議論

議論の
結果

[論点1]条件なし治療効果と条件付き治療効果の比較

議論の前提条件: 想定する臨床試験と担当者の状況

➤ 想定する臨床試験

- 検証的試験 (RCT)
- 強い予測因子は, 現在のところ(試験計画段階では)知られていない

➤ 担当者の状況

- 試験計画時に, estimand を
「条件なし治療効果」にすべきか, 「条件付き治療効果」にすべきか,
を決めるための参考情報が欲しい

[論点1]条件なし治療効果と条件付き治療効果の比較

(1) ESTIMAND の整理と分類

「条件なし治療効果」と「条件付き治療効果」の現状

- 多くの RCT で, 試験の対象集団全体に興味がある, と考えるのは自然
 - 「条件なし治療効果」を主要な estimand とすることが「自然」な状況は多いはず
- 一方, 主要な解析で共変量調整を行う場合,
「条件付き治療効果」を主要な estimand とすることも, 比較のみられる
 - 2値) 共通リスク差を Mantel-Haenszel 法により推定
 - 生存時間) 条件付きハザード比を, 層別 Cox モデルを利用して推定

A) 臨床的観点と B) 統計的観点

➤ ランダム化比較試験での条件付き治療効果と条件なし治療効果には一般的に下記の特徴がある (Wei, et.al. 2024)

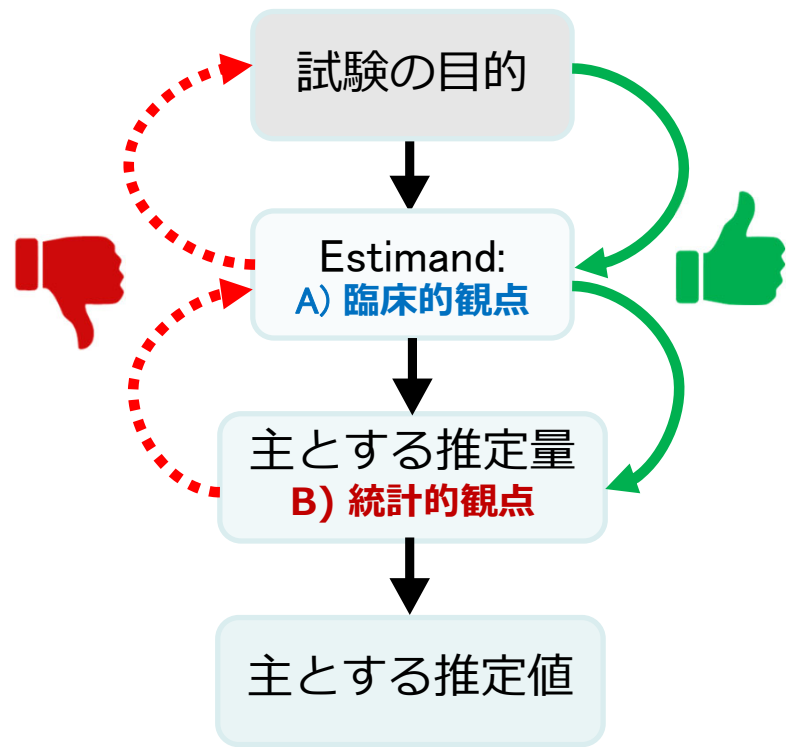
	条件付き治療効果	条件なし治療効果
A) 臨床的観点	<ul style="list-style-type: none"> 特定の共変量特性を有する患者集団での治療効果 (より個別化された治療効果) を把握できる 	<ul style="list-style-type: none"> 全体の患者集団で治療効果が平均的にどれだけあるかを把握できる
B) 統計的観点	<ul style="list-style-type: none"> (共変量の組や交互作用の有無を含め)モデル誤特定の影響が大きい 	<ul style="list-style-type: none"> 一般的に必要なとする仮定は条件付き治療効果よりも弱い 一般的にモデル誤特定に頑健

「条件なし」の方が「よさそう」なイメージ ← 「B) 統計的観点」のイメージが大きい?

A) 臨床的観点と B) 統計的観点

- ICH E9(R1) の thinking process (右図) に従うと,
B) 統計的観点より先に
A) 臨床的観点から検討すべき?

PR3-1, LP1 参照



ICH E9(R1) training material スライド76より
一部抜粋・改変

A) 臨床的観点と B) 統計的観点

LP4 スライド11
を少し加工

論点
整理



➤ ランダム化比較試験での条件付き治療効果と条件なし治療効果には一般的に下記の特徴がある (Wei, et.al. 2024)

	条件付き治療効果	条件なし治療効果
A) 臨床的観点	<ul style="list-style-type: none">特定の共変量特性を有する患者集団での治療効果 (より個別化された治療効果) を把握できる	<ul style="list-style-type: none">全体の患者集団で治療効果が平均的にどれだけあるかを把握できる
B) 統計的観点	<ul style="list-style-type: none">(共変量の組や交互作用の有無を含め)モデル誤特定の影響が大きい	<ul style="list-style-type: none">一般的に必要なとする仮定は条件付き治療効果よりも弱い一般的にモデル誤特定に頑健

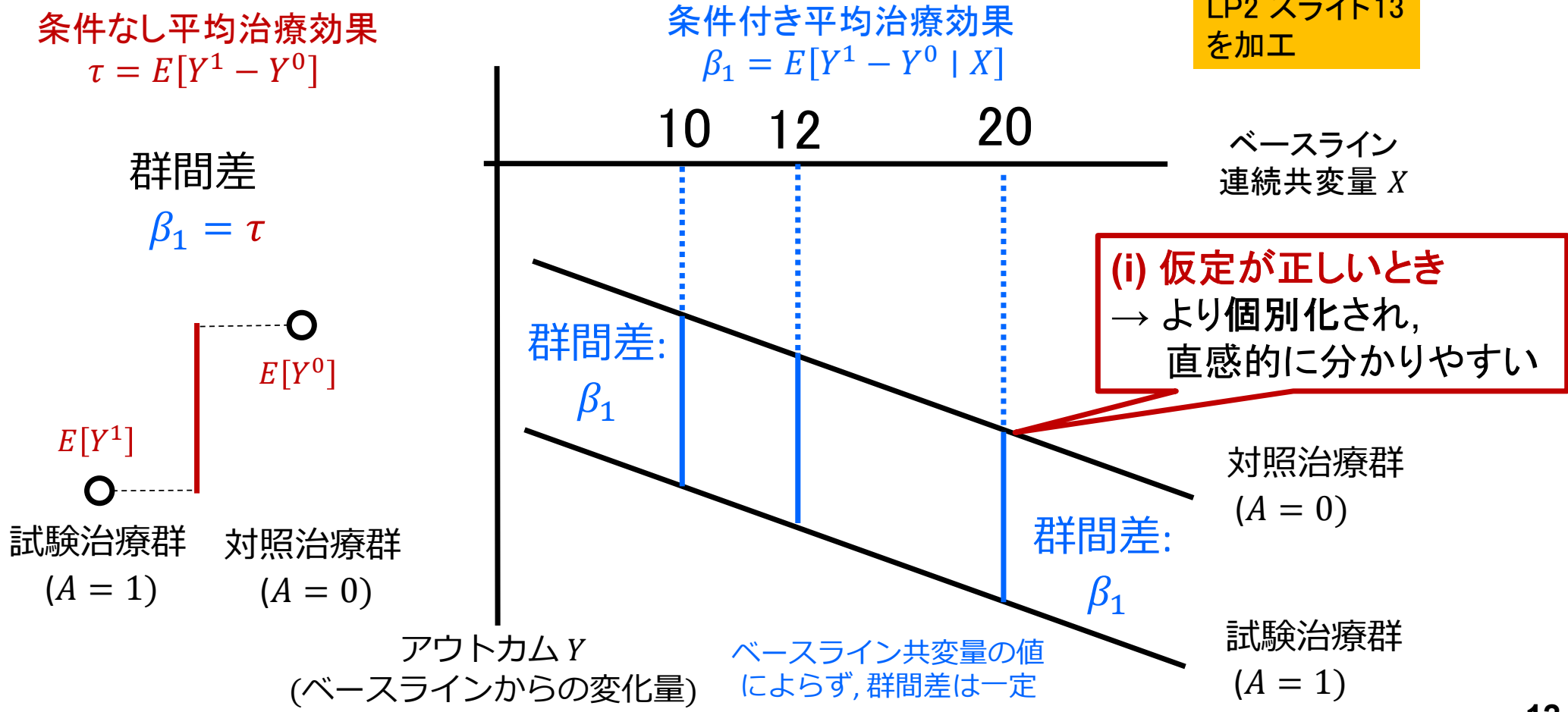
「A) 臨床的観点」で比較したい

→ 「より個別化された治療効果」とは何か, をもう少し検討

より個別化された治療効果(1)

「部分集団間で、治療効果がほぼ一定」の仮定が正しいとき

LP2 スライド13
を加工



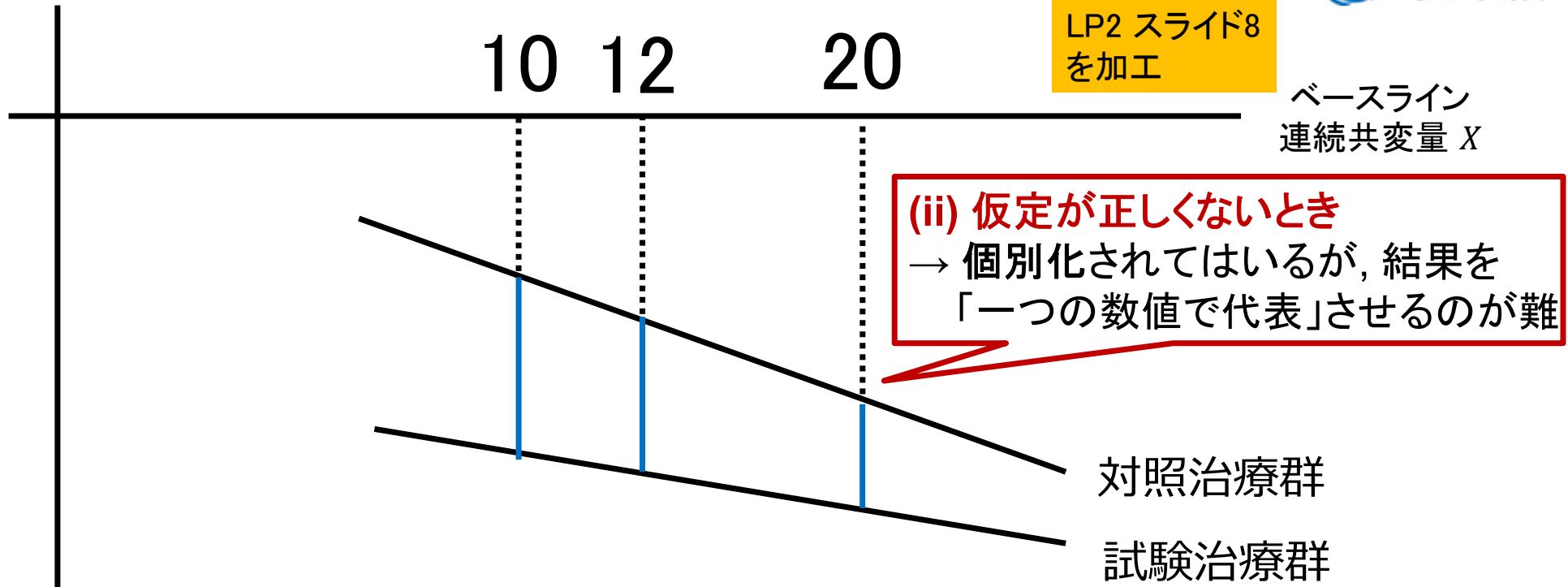
【当日 2】 ANCOVA による ATE 推定

より個別化された治療効果(2)

「部分集団間で、治療効果がほぼ一定」の仮定が正しくないとき

LP2 スライド8
を加工

ベースライン
連続共変量 X



アウトカム Y
(ベースラインからの変化量)

群間で大きく傾きが異なると考えられるようなデータでは、
共分散分析では、結果の解釈が困難となる可群性がある

「条件付き治療効果」の分類

➤ 「条件なし治療効果」 vs 「条件付き治療効果」

分類

➤ 「(1)条件なし治療効果」 vs 「(2)条件付き治療効果(部分集団間で一定)」
vs 「(3)条件付き治療効果(部分集団間で異なる)」

まずは、「仮定が正しい」想定。
「部分集団間で一定」の仮定が
正しかったか、ここでは問わない

PR3-1, PR3-2,
LP1, LP4 参照

➤ 論点1-1:

- 「試験全体の有効性の**結果の解釈しやすさ**」を比較した場合、
「試験全体を1つの数値で要約できる」方が、解釈しやすいのでは?

有効性を一つの数字で代表可能(集団全体 or 部分集団(共通))

- 「(1)条件なし治療効果」, 「(2)条件付き治療効果(部分集団間で一定)」

vs 「(3) 条件付き治療効果(部分集団間で異なる)」

部分集団ごとに異なる有効性
→ 「一つの数値で代表」させるのが難

論点1-1: 議論の結果

- これまでの審査や治験相談の内容を踏まえたコメント:
 - 「条件なし治療効果」と「条件付き治療効果」の区別に特に注目することなく議論することが多い。個別に「その薬がどの集団に有効かつ安全と判断できるか」を議論。対象となる集団を反映する治療効果を考えることが重要
 - 基本方針：同じ有効性が期待できる集団を一つの集団とみなし、一つの臨床試験の中で有効性を評価
 - 集団内に有効性の異なる集団が含まれる場合、全体の平均治療効果だけではリスク・ベネフィット評価が不十分となり、「全体を一つの値で代表させる」ことが適切でない場合もある
 - 集団間で有効性が異ならないかについて部分集団解析などを通じて確認。有効性が小さいと考えられる集団は、効能効果の対象患者に含めるかどうか、や注意喚起の必要性を個別に検討
 - 以上より、試験全体を一つの数値で要約できる状況となるように試験を計画することが望ましい。あらかじめ治療効果が異なる集団が想定される場合、それらを一つの集団として平均的に評価してよいか、慎重に検討すべき
- パネリストのコメントの抜粋(パネリストの総意を示すものではありません):
 - 大きな異質性はない、という想定が正しい場合、条件付き治療効果を一つの数値で要約する方法では意思決定や結果の解釈に大きな問題は生じないかもしれない
 - ただし、部分集団間で治療効果が一定、という仮定は強いため、注意が必要
 - 治療効果の大きな異質性につながる既知のバイオマーカーがある場合、バイオマーカーで規定される部分集団の治療効果を評価する方が妥当かもしれない

治療効果の比較

まずは、「仮定が正しい」想定.
「部分集団間で一定」の仮定が
正しかったか、ここでは問わない

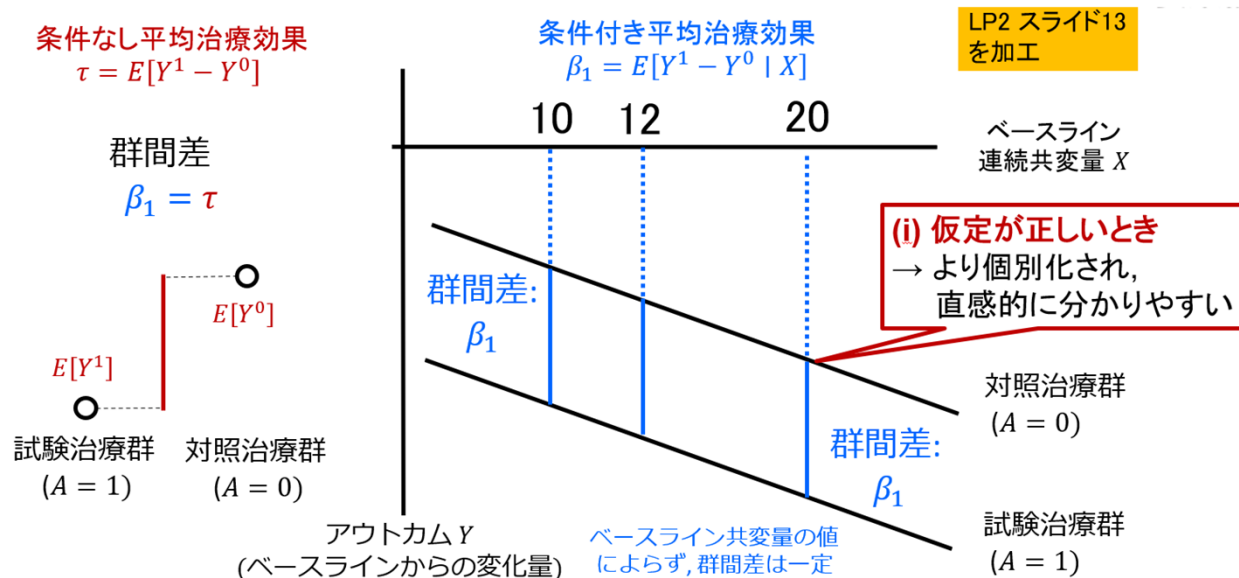
議論



➤ 論点1-2:

■ (特に, 併合可能な指標を用いて, 適切に解析されている場合)

「(1)条件なし」vs「(2)条件付き(部分集団間で一定)」では,
「(2)条件付き(部分集団間で一定)」の方が
「個別化されており, 情報が多い」のでは?



論点1-2: 議論の結果

- パネリストのコメントの抜粋(パネリストの総意を示すものではありません):
 - 「(2) 条件付き治療効果(部分集団間で一定)」は、その仮定が完全に正しいならば、どの部分集団でもベネフィットが平均的に見込めるため、より個別化した解釈が可能となり、メリットはある
 - 一方、やはり現実的な使用を考えると、どうしても仮定の強さが無視できない。
この仮定が満たされている、というのは現実的ではないと思われるため、
主要な解析の estimand としては、「(1)条件なし治療効果」の方が望ましいかもしれない
 - 部分集団に関する議論を行う状況で、条件付き治療効果に関心があるかもしれない。
 - これは特に、「(3) 条件付き治療効果(部分集団間で異なる)」の方が顕著かもしれない。
たとえば、安全性が部分集団ごとに異なる場合に、リスクベネフィット評価を部分集団ごとに実施する必要があるならば、条件付き治療効果の比較はありうるかもしれない

[論点1]条件なし治療効果と条件付き治療効果の比較
(2)「条件なし治療効果」と「条件付き治療効果」

論点の整理:

I. A) 臨床的観点と B) 統計的観点

LP4 スライド11
を少し加工

論点
整理



➤ ランダム化比較試験での条件付き治療効果と条件なし治療効果には一般的に下記の特徴がある (Wei, et.al. 2024)

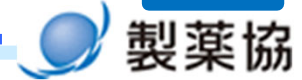
	条件付き治療効果	条件なし治療効果
A) 臨床的観点	<ul style="list-style-type: none">特定の共変量特性を有する患者集団での治療効果 (より個別化された治療効果) を把握できる	<ul style="list-style-type: none">全体の患者集団で治療効果が平均的にどれだけあるかを把握できる
B) 統計的観点	<ul style="list-style-type: none">(共変量の組や交互作用の有無を含め)モデル誤特定の影響が大きい	<ul style="list-style-type: none">一般的に必要なとする仮定は条件付き治療効果よりも弱い一般的にモデル誤特定に頑健

「条件なし治療効果」の推定性能はよさそう
→ B) 統計的観点 の評価

統計的観点からの「条件なし」「条件付き」治療効果の比較

参考: FDA ガイダンス

論点
整理



条件付き

- (C6) 非線形回帰により条件付き治療効果を推定する際、一般にモデルの仮定が完全に正しいことはなく、モデルが誤特定されていて治療効果が部分集団間で大きく異なる場合、結果の解釈が困難となる可能性がある。

条件なし

- (C7) 治験依頼者はランダム化比較試験の主要な解析で条件なし治療効果(例えば, Table 1のオッズ比4.8)に共変量を調整した推定・推測を用いてもよい。ランダム化比較試験の未調整推定で必要とされる最小限の統計的仮定とほぼ同様の仮定の下で妥当な推測を与える方法を用いるべきである。

◎一見すると、

・「条件付き」⇒ 部分集団間で治療効果が大きく異なるとよくない

・「条件なし」⇒ 部分集団間で治療効果が大きく異なってもよい

かのように見えるかもしれない。しかし、実際は以下のはず。

・「条件なし」⇒ 部分集団間で、治療効果が大きく異なっても、

「条件なし治療効果の推定が統計的に適切 (B. 統計的観点)」なだけであり

「結果の解釈の際に、治療効果の均質性は別途確認が必要」 ⇒ 「部分集団解析」

治療効果の選択

➤ 論点1-3(確認):

- 「(1)条件なし治療効果」は、「集団全体を対象とする」ということは意味するが、「結果として有効性に均質性がなくても、集団全体に対して有効」とはならないはず
 - 通常の部分集団解析で均質性の確認を実施することは、(これまでもこれからも)重要では?

論点1-3: 議論の結果

- これまでの審査や治験相談の内容を踏まえたコメント:
 - 部分集団解析による均質性の評価は重要
 - とくに, 治療効果が他の集団と比較してかなり小さい部分集団があった場合, その部分集団を含めて薬の投与の対象集団とできるか, の検討が必要になることもある
 - 「同様の有効性を期待できる集団」を1つの集団として評価しようとしているか, は論点になる
 - 条件あり, 条件なしに関わらず重要

➤ 論点1-4:

- A) Estimand の選択で、「集団全体に関心がある」+
「主要な解析の統計的仮定が少ない」という理由で、
「(1) 条件なし治療効果」を選択することに対して、(一般的な) 懸念点は?

- B) Estimand の選択で、「より個別化された治療に関心がある」
「現在のところ(計画段階では)、強い予測因子が知られていない」ことから、
「(2)条件つき治療効果(部分集団間で一定)」を選択することに対して、
(一般的な) 懸念点は?

論点1-4: 議論の結果

- パネリストのコメントの抜粋(パネリストの総意を示すものではありません):
 - A) 「(1)条件なし治療効果」の使用は多くの試験で基本的には問題なさそう
 - 「条件なし治療効果」の推定値が、「各部分集団での有効性を示すわけではない」という点の認識、および「条件なし治療効果」であっても均質性の検討のための部分集団解析などは重要
 - B) 「(2)条件付き治療効果(部分集団間で一定)」の使用には、以下の点などに注意が必要
 - 「(2) 条件付き治療効果(部分集団間で一定)」の仮定は、有効性がある場合の推定においては、厳密には誤っている試験が多いはず
 - 「部分集団間で一定」という仮定が誤っているときに、「各部分集団で一定の有効性を示している」と誤って解釈することがないように注意が必要
 - 以上を踏まえつつ、「仮定が誤っている場合に、統計量がどのような性能を示すか」の検討は重要。たとえば、TF の発表 LP4 の資料中で「1点目」として議論されている内容は以下の通り:
Mantel-Haenszel リスク差は、当初は条件付きリスク差の推定方法として提案された。一方、Qiu et al. (2025)では、緩い仮定と若干の調整のもとで、条件なしリスク差の推定量としても解釈可能、という記載がある

Qiu, X., et.al. (2025). Clarifying the role of the Mantel-Haenszel risk difference estimator in randomized clinical trials. *Biometrics*, 81(4): ujaf142.

<https://academic.oup.com/biometrics/article-abstract/81/4/ujaf142/8307320>

評価指標の選択（併合可能性含む）

田栗先生ご講演資料:
スライド44-52参照

PR3-3 参照



➤ 論点1-5:

- 併合不能な評価指標の使用の適切性などを含め、評価指標の選択や使用における注意点は？

論点1-5: 議論の結果

➤ これまでの審査や治験相談の内容を踏まえたコメント:

■ 臨床的解釈の重要性

□ オッズ比・ハザード比

- ✓ 併合不能な指標の典型例としてよく挙げられるオッズ比とハザード比は、臨床的な解釈が難しい
- ✓ 審査で治療効果の臨床的意義、リスクベネフィットを議論する場合、主要な解析でオッズ比やハザード比が用いられている場合であっても、別途リスク差やイベント発現までの時間の中央値の群間差、などの指標も算出してもらうこともあった
- ✓ 審査報告書では、臨床的意義の説明などにはハザード比以外の指標も使って議論することが多い印象

□ 比の指標

- ✓ プラセボ対照試験で、プラセボ群の改善例が非常に少ない状況で、リスク比などの比の指標を用いた試験があった。分母をプラセボ群とすると、分母の改善割合が小さいことから比の値が大きくなってしまい、数値自体は正しいものの、「治療効果をミスリードしていないか?」という指摘が出たことがある。その際は、割合の差も提示いただくことなどで対応した

論点1-5: 議論の結果

➤ これまでの審査や治験相談の内容を踏まえたコメント(つづき):

■ 条件つき治療効果か, 条件なし治療効果か

- 前スライドの「オッズ比・ハザード比」の記載内容も踏まえ, 現状としては「条件付き治療効果を主要な estimand として, 層別Cox回帰で主たる推定・検定は行い, 合わせて中央値の比較や時点生存率の比較も行った上で, 総合的に判断する」という方向で審査を行っている

✓ 検定に関しては, 帰無仮説下では比例ハザード性などの仮定が満たされやすい

- 「条件なし治療効果」の方が臨床的観点からは望ましい印象がある.

一方で, 現在「条件付き治療効果」を用いている領域・試験で, 今後は「条件なし治療効果」を使用するように変更する方がよいか, に関しては, 「条件なし治療効果」を推定する統計手法の性能に依存する面もある. そのため, 特に新しい手法を使用しようとする際は, 統計的性能の議論が重要になる

■ 新しい統計手法の使用

- 新しい統計手法を使用する際, 「従来から用いられている手法を用いるとどのようになるか」を伺うことはあるが, これはあくまで比較対照として聞いていることが多い.
「既存手法を用いるべき」という意図ではない点は, ご理解いただけるとありがたい

論点1-5: 議論の結果

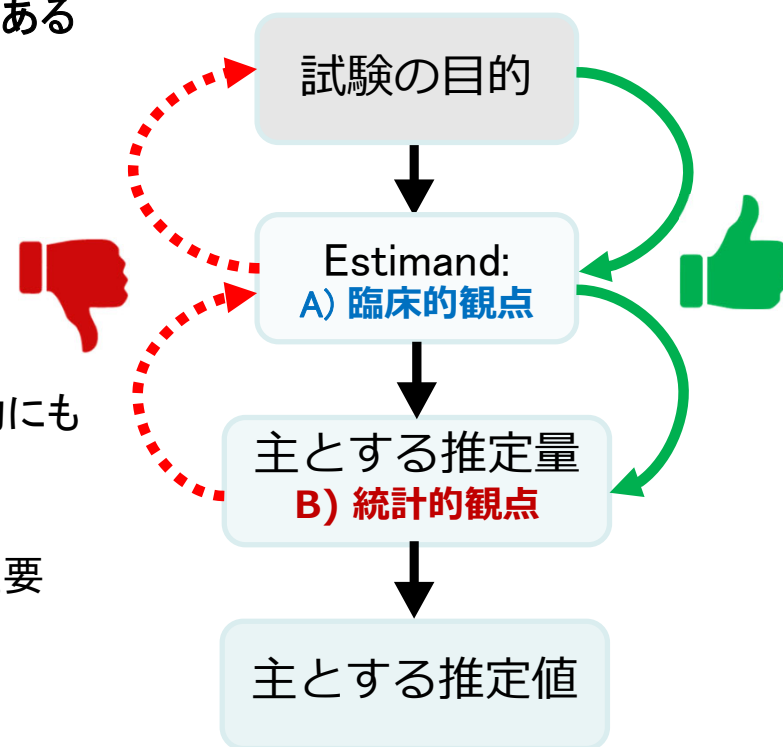
田栗先生ご講演資料:
スライド50-52参照

議論の
結果



➤ パネリストのコメントの抜粋(パネリストの総意を示すものではありません):

- ハザード比は臨床的解釈が難しいが、一部臨床医等を含め、「臨床的なインパクトの大きさの感覚」を持っているような印象もある
- 現在、条件付き治療効果が推定されている試験でも、
今後は第一選択は条件なし治療効果で推定・検定を行うのが、
望ましいかもしれない
 - 一方、条件なし治療効果を共変量調整して推定する方法は、
新しい方法が多く、経験的にも統計的性能が
それほど知られていない。
業界全体で経験を積み、推定量の特徴を理論的にも経験的にも
評価していくことが重要
 - ✓ 最終的な estimand の選択には、
A) 臨床的観点 と B) 統計的観点 の両面での検討が重要



ICH E9(R1) training material スライド76より
一部抜粋・改変

[論点2]条件なし治療効果の推定方法の検討

(再掲) パネルディスカッションの概要

➤ 目的

1. 論点の整理
2. 今後の方向性に関する議論

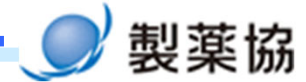
➤ 論点

- 論点1: 条件なし治療効果と条件付き治療効果の比較
- 論点2: 条件なし治療効果の推定方法の検討
- 論点3: Super covariates の使用に関して

[議論] 条件なし治療効果の推定方法の検討

田栗先生ご講演資料:
スライド5-32参照

議論



➤ 論点2

PR3-4
LP3 参照

- ANCOVA/ANHECOVA/標準化(g-computation)/IPTW/AIPW の使い分け
 - 傾向スコアを使用する場合, 真値/推定値のどちらを用いるのがよいか

論点2: 議論の結果

- パネリストのコメントの抜粋(パネリストの総意を示すものではありません):
 - 提案されている複数の手法の理論的評価の基本は, 漸近分散を用いるものが多い. 漸近分散が同じ時に, 現実的な状況設定でどちらがよいか, は今後検討が必要
 - 一つの案としては, 特別な理由がない限り, (手法の特徴がよく理解できている) ANCOVA や ANHECOVA を使用する, という方針で, 大きな問題は生じないかもしれない
 - AIPW や, アウトカムの条件付き期待値のモデリングによるメリットが得られそうな状況の例:
 - 二値結果変数に対して線形モデルでモデル化すると, 近似精度がよくないかもしれない. そのような場合, 非線形なロジスティック回帰モデルを用いたAIPWや標準化を利用することで, 検出力の増加が見込めるかもしれない
 - IPTWで, 使用する傾向スコアは (1)割付確率の設定値(真値), (2)推定値, のどちらにするか, の選択肢がある
 - (1)や, (2)で $\hat{e}_i = \frac{n_i}{n}$ (n : 総サンプルサイズ, n_i : 第 i 群のサンプルサイズ) を代入した場合, アウトカムの標本平均の差とほぼ一致するため, メリットはあまりないかもしれない
 - (2)で, 傾向スコアをロジスティック回帰モデルでモデル化し, その推定値を用いる場合, 推定精度の改善が見込める(漸近的にANHECOVAと同様)

[論点3] SUPER COVARIATES の使用に関して

(再掲) パネルディスカッションの概要

➤ 目的

1. 論点の整理
2. 今後の方向性に関する議論

➤ 論点

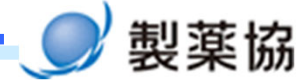
- 論点1: 条件なし治療効果と条件付き治療効果の比較
- 論点2: 条件なし治療効果の推定方法の検討
- 論点3: Super covariates の使用に関して

その他の論点:

[議論内容] Super covariatesの利用

田栗先生ご講演資料:
スライド33-42参照

議論



➤ 論点3:

■ 「積極的に使用してよい」

vs 「通常の方法での解析が難しい場合にのみ使用するのが望ましい」

□ 複数の因子を縮約させた共変量(super covariates)の使用自体の適切性

✓ 臨床的解釈の難しい共変量の使用

✓ 「条件なし治療効果」の推定では, estimand は調整因子で条件づけられない.

+ モデル誤特定に対して一定の頑健性をもつ手法の適用が考えられる.

□ Super covariatesの構成に, (ニューラルネットワークを含む)機械学習モデルを用いることの適切性

✓ 共変量作成のプロセスに「ブラックボックス」が存在

✓ Super covariatesの「再現性」の担保をどうするか?

論点3: 議論の結果

- パネリストのコメントの抜粋(パネリストの総意を示すものではありません):
 - どのような状況で複数の因子を縮約させた super covariate を使うのが望ましいか、を考えることが重要かもしれない
 - 簡単に思いつく状況: 比較的少数例の RCT を計画しており、比較的多くの共変量で調整した解析を行いたい、サンプルサイズを考慮すると難しい、というような状況
 - ✓ そのような状況で、アウトカムと相関する、よい super covariate が作れば、閾値で区分して層別ランダム化に利用するなど、既存の枠組みに載りやすいかもしれない
 - 過去データと現在データで super covariate とアウトカムの関連が異なる場合などにおいては、通常の ANCOVA と比較して、推定効率が落ちることがあり得る。そのため「常に使用する方がよい」とはいえないのではないか。いずれにせよ、使用経験の少ない統計手法であるため、十分な動作特性の確認が必要