



FDA ガイダンスを踏まえたランダム化比較試験 の共変量調整

日本製薬工業協会
医薬品評価委員会 データサイエンス部会
2024年度タスクフォース3-1

Ver 1.0
2025年6月

略語の定義・説明

略語	定義・説明	初出
ANCOVA	ANalysis of COVAriance（共分散分析）	3.9節
ANHECOVA	ANalysis of HEterogeneous COVAriance	3.9節
ANOVA	ANalysis Of VAriance（分散分析）	3.9節
EMA ガイドライン	“Guideline on adjustment for baseline covariates in clinical trials”	1.1節
FDA ガイダンス	“Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products”	1.1節
ICH E9	臨床試験のための統計的原則	1.1節
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	1.1節
IPTW	Inverse Probability of Treatment Weighting	3.9節

目次

略語の定義・説明.....	2
1. はじめに.....	5
1.1 背景.....	5
1.2 本報告書について.....	6
1.2.1 本報告書の目的と概要.....	6
1.2.2 本報告書の対象外の内容.....	7
2. ガイダンス・ガイドラインの共変量調整に関する部分の概説.....	8
2.1 ICH E9.....	8
2.2 ICH E9(R1).....	9
2.3 EMA ガイドライン.....	10
2.4 FDA ガイダンス.....	14
2.5 FDA ガイダンスとこれまでのガイドラインの比較.....	15
3. FDA ガイダンスの理解に役立つ用語の概説.....	22
3.1 統計的効率.....	22
3.2 Estimand に関連する用語.....	22
3.2.1 Estimand.....	22
3.2.2 潜在結果変数.....	23
3.2.3 条件付き治療効果と条件なし治療効果.....	24
3.2.4 Model-based estimand/ model-free estimand.....	25
3.3 Estimator に関連する用語.....	25
3.3.1 Adjusted estimator/unadjusted estimator.....	25
3.4 「条件付き治療効果と条件なし治療効果」と「Adjusted estimator/unadjusted estimator」 の違い.....	25
3.5 予後因子/ベースライン予後共変量と効果修飾因子/ベースライン予測共変量.....	26
3.6 交絡.....	27
3.7 併合可能性と併合不能性.....	28
3.8 モデル誤特定.....	29
3.9 Model-assisted な手法.....	30
3.10 線形モデル・非線形モデル.....	31
4. FDA ガイダンス本文の解説.....	32
4.1 ドラフトガイダンスと最終化までの経緯.....	32
4.2 ガイダンス本文.....	32
4.2.1 タイトルなど.....	33
4.2.2 I. Introduction.....	33
4.2.3 II. Background.....	35
4.2.4 III. Recommendations for Covariate Adjustment in Clinical Trials.....	37

4.2.4.1	A. General Considerations	37
4.2.4.2	B. Linear Models.....	41
4.2.4.3	C. Nonlinear Models	44
4.2.5	REFERENCES	51
4.3	論点の整理.....	52
4.3.1	FDA ガイダンスで審査部門との相談が推奨されている内容の列挙.....	52
4.3.2	治験実施計画書の estimand の記載	56
4.3.3	線形モデルの回帰係数の推定値と平均治療効果の推定値との関係.....	56
4.3.3.1	線形モデルに治療とベースライン共変量の交互作用を含めない場合	56
4.3.3.2	線形モデルに治療とベースライン共変量の交互作用を含める場合	57
4.3.4	線形モデル誤特定に関する整理	58
4.3.4.1	線形モデル誤特定に関する記述.....	58
4.3.4.2	モデル誤特定下での治療効果の推定量の妥当性について（推定量の漸近 正規性）	58
4.3.4.3	モデルが誤特定されている場合の標準誤差について	60
4.3.5	非線形モデルでの条件付き治療効果と条件なし治療効果の記載の比較.....	61
4.3.6	被験者数の減少を狙うか、検出力の増加を狙うか.....	64
5.	おわりに	66
6.	参考文献及びガイドラインの理解に役立つ資料	67
	執筆者・タスクフォースメンバー	71

[文書内のハイパーリンクに関する注意]

本報告書では、文書内にハイパーリンクを多用している。Adobe Acrobat 利用時は、文書内へのリンク（たとえば 4.2 節）をクリック後、「Alt + ←（左矢印）」で元の場所に戻ることができる。

1. はじめに

1.1 背景

ランダム化比較試験の主要な解析としてベースライン共変量を利用した調整解析が選択肢に上がった場合の方針・方法の検討は、古くから重要な課題である。2025年5月現在、よく参照されるガイドラインは ICH E9（臨床試験のための統計的原則）及び2015年に発出された EMA ガイドライン“Adjustment for Baseline Covariates in Clinical Trials”（以下、EMA ガイドライン）の二つである。

一方、2023年に FDA ガイダンス“Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products”（以下、FDA ガイダンス）が発出された。FDA ガイダンスは、治療効果の推定及び検定に対する統計的効率を向上させるためのベースライン予後共変量の取扱いを主な焦点として、最近の研究も踏まえた記載がなされている。統計的効率の向上は、検出力の増加を通して試験の成功確率の向上や被験者数の減少にも関連する重要な論点である。そのため、本 FDA ガイダンスを理解することは統計担当者のみならず、臨床試験の計画・実施・解析・報告のいずれかに関与する開発担当者にとって有益であろう。

ICH E9や EMA ガイドラインを踏まえ、現状は以下のような方針が用いられることが多い（ICH E9 5.7節や EMA ガイダンスの Executive Summary 等参照）。

- アウトカムと相関の強いことが計画段階で知られているベースライン共変量があれば、その変数で調整する
- ベースライン共変量とアウトカムの関係に対する仮定は、簡単なものとする
- 層別ランダム化を行った場合、層別因子で調整する
- 主要な解析の統計モデルには、治療とベースライン共変量の交互作用は含めない

一方、FDA ガイダンスでは、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）の estimand のフレームワークや、近年の統計理論の発展も踏まえ、新たな論点や可能性が提示されている。以下、特徴的な記載内容をいくつか挙げる。

- 有効性の主要な解析で、線形モデルを用いて平均治療効果（Average Treatment Effect）を推定する際、治療とベースライン予後共変量（3.5 節）の交互作用を含める方法が議論されている（FDA ガイダンス: 4.2.4.2 節, 参照: 3.2.3 節, 4.3.3.2 節, 4.3.4 節）
- 線形モデル・非線形モデルを問わず、モデル誤特定の可能性に対する言及が多い。特に「モデル誤特定下でも適切な推論が行える方法」についての議論も含まれている（FDA ガイダンス: 4.2.4.2 節, 4.2.4.3 節, 参照: 3.8 節, 3.9 節, 4.3.4 節）

- (非線形モデルの節に) estimand の要素として、条件付き治療効果を推定したいか、条件なし治療効果を推定したいか、を特定することが推奨 (should) されている (FDA ガイダンス: 4.2.4.3 節, 参照: 3.7 節, 3.2.3 節, 3.4 節)
- 二値アウトカムに対する条件なし治療効果を共変量調整して推定する方法は、これまでの治験ではほぼ用いられてこなかったが、FDA ガイダンス中に具体的な方法の一つが例示されている (FDA ガイダンス: 4.2.4.3 節, 参照: 3.4 節, 3.3 節)

また、後述するように (4.2.4.1 節), FDA ガイダンスでは未調整解析の実施, すなわち、今までランダム化比較試験で慣例的に行われてきた方法も許容されている。一方、適切な共変量で調整することにより統計的効率が向上し得ることを踏まえ、FDA は関心のあるアウトカムと最も強く関連すると予想されるベースライン共変量で調整することを推奨している。このような状況を踏まえ、今後計画されるランダム化比較試験でのベースライン共変量の調整の要否、及び調整する場合の具体的方法を検討することが、各製薬企業の課題となっている。

1.2 本報告書について

1.2.1 本報告書の目的と概要

本報告書の目的は、ICH E9やEMA ガイドラインの内容を踏まえた上で、FDA ガイダンスの理解を深める一助とすることである。概要は以下の通りである。

2章までは臨床試験の計画、実施、解析、報告のいずれかに関与する臨床開発担当者を想定読者とする。2章ではこれまでに発出された代表的なガイドライン・ガイダンスである、ICH E9, E9(R1), EMA ガイドライン, FDA ガイダンスの概要をまとめるとともに、FDA ガイダンスとの比較を行う。なお、FDA ガイダンスの日本語訳部分のみ (本報告書4.2節の中の日本語訳部分のみ) を抜粋したファイルを Appendix として、別ファイルを作成した。

3章以降は統計解析担当者を想定読者とする。3章では、FDA ガイダンスの理解を十分に行うため、各種用語の説明を記載した。まず、(1) 条件付き治療効果、調整済み推定量、併合不能性等、FDA ガイダンスに記載されている用語の説明を行った (3.1, 3.2.1, 3.2.3, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.10節)。次に、(2) FDA ガイダンスを十分に理解するための前提として、ICH E9(R1) の estimand の理解を深めるため、潜在結果変数を用いて estimand を記述するための用語の説明を行った (3.2.2, 3.2.3節)。さらに、(3) model-free estimand, model-based estimand や model-assisted な手法等、FDA ガイダンス上には明記されていないものの、FDA ガイダンスの内容を理解するために役立つと考えられる用語の説明を行った (3.2.4, 3.9節)。続く4章では、FDA ガイダンス本文の日本語訳を記載した。その際、原文をすぐに参照できるよう、原文も併記した。それに加え、2021年度に発出されたドラフトガイダンスや、それに対するパブリックコメント (<https://www.regulations.gov/docket/FDA-2019-D-0934/comments> [最終アクセス日: 2025/5/9]) の内容も踏まえ、適宜補足説明を記載した (4.2節)。また、FDA ガイダンスの中で特に検討が必要な内容をピックアップし、補足説明を記載した (4.3節)。5章では、本報告書のまとめ及び、今後のタスクフォース活動の予定について述べた。最後の6章では、参考文献及び FDA ガイダンスの理解に役立つ資料を記載した。なお、FDA ガイダンス内で引用されている文献は、FDA ガイダンスの日本語訳を記載している4.2.5節に記載している。

1.2.2 本報告書の対象外の内容

FDA ガイダンスを実際の臨床試験に適用しようとする際、以下の二つが大きな課題となり得る。

参考文献として引用されている論文の難易度が高く、実際に適用する際に理論的正当性や適用範囲を確認する際のハードルが高いこと

欠測、経時データの取扱い等、実際の臨床試験計画上で極めて重要な検討事項が FDA ガイダンスの対象外とされていること

これらは大きな課題となり得るものの、2024年度タスクフォース活動としては、まずはFDAガイダンス自体の理解を深めることが重要と考え、本報告書では対象外とし、今後のタスクフォース活動の課題とした。

2. ガイダンス・ガイドラインの共変量調整に関する部分の概説

医薬品開発におけるランダム化比較試験の統計解析における共変量の取扱いに関しては、諸種のガイドラインですでにいくつかの議論及び推奨事項が示されている。本章では、ICH E9, E9(R1), EMA ガイドラインの関連記載を確認した後、FDA ガイダンスの要約を与えるとともにこれらの記載内容との共通点及び相違点を示し、FDA ガイダンスをより深く理解するための準備を行う。以降の ICH E9, E9(R1) は日本語訳が公開されているため、そのままの記載を用いた。本報告書の他の箇所では、同じ単語に対しても訳語が異なることがある。

2.1 ICH E9

ICH E9 5.7節「部分集団、交互作用及び共変量」の記載を以下に示す（太字はFDA ガイダンスでの引用箇所）。

主要変数が試験治療によるものとは別の作用と、系統的に関連している例がよくみられる。例えば、年齢や性といった共変量と関連がある場合、又は多施設共同試験では異なる施設で試験治療を受けた被験者という部分集団間に差がみられる場合がある。共変量の影響又は部分集団の効果を調整することは予定した解析の重要部分となる場合があり、したがって試験実施計画書に記述すべきである。主要変数に重要な影響を及ぼすと予想される共変量と要因は、試験開始前に議論して確認しておくべきであり、精度を向上させ、試験治療グループ間のバランスの欠如を埋め合わせるため、それらを解析でどう取り扱うかを考慮すべきである。計画時に一つ以上の因子を層別因子として用いたならば、それらの因子を解析時に考慮することが適切である。調整することの潜在的な有益さが疑わしい場合にはしばしば、調整しない解析に主要な関心を払うと宣言し、調整した解析はそれを補うものとするのが薦められる。施設の効果及び基準となる時点での主要変数の測定値の役割には特別な注意を払うべきである。しかし、ランダム化後に測定された共変量を主要な解析で調整することは薦められない。なぜならばそのような共変量は試験治療の影響を受けている可能性があるからである。

試験治療の効果自身もまた部分集団や共変量により異なる可能性がある。例えば、効果は年齢とともに減少する場合があり、被験者の特別な診断分類では、より大きな効果があることもある。そのような交互作用は予想されたり、交互作用自体に重要な関心がある場合（例えば高齢者医療など）があり、したがって部分集団解析又は交互作用を含んだ統計モデルは予定した検証的解析に含まれる場合がある。しかし、多くの場合、部分集団別解析又は交互作用解析は探索的であるため、探索的であることを明確に確認しておくべきである。それらの解析では、全体でみられた試験治療の効果がすべて一様であるかどうかを探索すべきである。一般に、そのような解析は、まず問題となっている統計モデルに交互作用項を加えることから始めるべきであり、適切な被験者の部分集団内、又は共変量によって定められる層内を更に探索的に解析することによって補われる。探索的解析である場合、これらの解析結果は慎重に解釈すべきである。試験治療の有効性（若しくは有効性がないこと）、又は安全性に関する結論は、どのようなものであっても、探索的な部分集団解析のみに基づいては受け入れ難い。

2.2 ICH E9(R1)

FDA ガイダンスは ICH E9(R1) を踏まえた記載が複数個所にある。共変量調整された治療効果は, estimand の要素のうち, 集団レベルの要約に該当するため, estimand の考え方を十分に理解することは, FDA ガイダンスを理解し, 適切に運用する際に重要である。ICH E9(R1) では, estimand と estimator (推定量) を明確に区別した議論が行われていることを踏まえ, 本報告書の3章で estimand 及び estimator の用語について整理を行う (3.2～3.4節)。以下に, ICH E9(R1) の関連する記載を示す。

(A.2節「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」) 試験計画は順を追って進めるべきである (図1)。明確な試験の目的を, 適切なestimandを定義することにより, 関心のある重要な臨床的疑問に変換すべきである。Estimandは, 特定の試験の目的に対する推定の対象 (すなわち, 「推定されるべきもの」) を定義する (A.3節)。それにより, 適切な推定の方法 (すなわち, 主とする「推定量」と呼ばれる解析手法, 用語集参照) を選択することができる (A.5.1節)。主とする推定量は特定の仮定に裏付けられている。主とする推定量の推測について, その仮定のずれに対する安定性を調べるために, 同じestimandを対象とした一つ以上の解析として, 感度分析を実施すべきである (A.5.2節)。

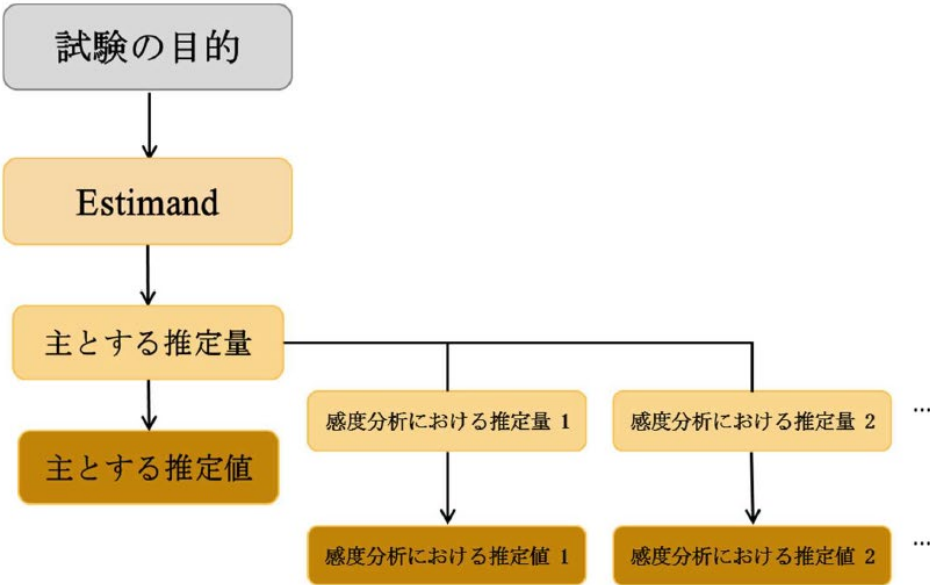


図 1：定められた試験の目的に対して、推定の対象、推定の方法及び感度分析を整合させる

このフレームワークに従うことにより, 推定の対象 (試験の目的, estimand), 推定の方法 (推定量), その数値的な結果 (「推定値」, 用語集参照) 及び感度分析を明確に区別した, 適切な試験の計画を行うことができる。これは, 治験依頼者の試験計画及び規制当局のレビューに役立ち, また, 試験デザインの適切性について議論する際の治験依頼者と規制当局との間の意思疎通や, 試験結果の解釈をより良いものとする。

適切なestimand (A.3節) の説明は, 通常, 試験デザイン, 実施 (A.4節) 及び解析 (A.5節) の特徴に関する主要な決定要因となる。

(A.3節「Estimand」、一部抜粋) 医薬品開発と薬事承認において中心となる課題は治療効果の存在を確認し、その大きさを推定することである。すなわち、試験治療による結果と、他の治療(すなわち、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)の下で同じ被験者に起きたであろう結果との比較である。Estimandは、定められた臨床試験の目的によって提起される臨床的疑問を反映する治療効果の詳細な説明である。それは、比較されている異なる治療状況下で同じ患者の結果がどのようなになるかを集団レベルで要約するものである。推定の対象は臨床試験の前に定義される。それが定義されれば、対象とする治療効果の信頼できる推定を可能にするための試験を計画することができる。

(A.3.3節「Estimandの要素」) 以下の要素によりestimandを構成し、関心のある治療効果を定義する。

関心のある治療の状況と、必要に応じて、比較を行う他の治療の状況(本補遺では以下「治療と呼ぶ」)。これは、個別の介入、同時に行われる介入の組み合わせ(例えば、標準治療への上乗せ)、又は複雑な一連の介入を含む総合的な治療計画で構成される場合がある(A.3.2節の治療方針及び仮想ストラテジーを参照)。

臨床的疑問に対応する対象集団。これは、試験の対象集団全体、ベースラインで測定される特性によって定義される部分集団、又は特定の間事象の発現(又は、状況によっては非発現)によって定義される主要層(A.3.2節の主要層ストラテジー参照)で説明される。

臨床的疑問に対応するために必要な、各患者について収集する変数(又は評価項目)。変数の詳細には、患者が中間事象を経験するかどうかが含まれる場合がある(A.3.2節の複合変数及び治療下ストラテジー参照)。

治療、対象集団及び変数の詳細な説明は、治験依頼者と規制当局が関心のある臨床的疑問について議論する際に考慮される多くの中間事象に対応しているだろう。その他の中間事象に関する関心のある臨床的疑問は、通常、治療方針、仮想、又は治療下ストラテジーを用いて反映される。

最後に、変数の集団レベルでの要約を規定し、治療状況間の比較の基盤を与える。

関心のある治療効果を定義する際には、その定義により、観察期間や患者の特性の違いといった潜在的な交絡因子による効果ではなく、治療による効果を特定することを保証することが重要である。

2.3 EMA ガイドライン

EMA ガイドライン Executive summary の原文、日本語訳及び必要に応じて補足説明を以下に示す。

Baseline covariates impact the outcome in many clinical trials. Although baseline adjustment is not always necessary, in case of a strong or moderate association between a baseline covariate(s) and the primary outcome measure, adjustment for such covariate(s) generally improves the efficiency of the analysis and avoids conditional bias from chance covariate imbalance.

【訳】 ベースライン共変量は、多くの臨床試験において結果に影響を及ぼす。ベースライン共

変量の調整は必ずしも必要ではないが、ベースライン共変量と主要なアウトカムとの間に強いもしくは中程度の関係が認められる場合は、そのような共変量の調整は一般に解析の効率を改善し、共変量分布の偶然の不均衡に起因する条件付きバイアスを回避することに寄与する。

【補足説明】ランダム化を行うことにより、群間の共変量の分布が「平均的に」均衡し、妥当な群間比較を行うことができる（内的妥当性が確保される）が、とくに被験者数が少ない場合に、ランダム化の結果として共変量の分布が偶然に不均衡となることがある。ランダム割付けの結果を観測した（条件とした）後の、偶然の共変量の不均衡により生じる交絡バイアスを「条件付きバイアス」と呼ぶ（Zhang et al. 2019）。

Baseline covariates may be accounted for at the design stage of a clinical trial and/or in the statistical analysis. When dealing with baseline covariates the following recommendations are made:

【訳】ベースライン共変量は、臨床試験のデザイン段階や統計解析で考慮される場合がある。ベースライン共変量を取り扱う際の推奨事項は以下の通りである：

- Stratification may be used to ensure balance of treatments across covariates; it may also be used for administrative reasons (e.g. block in the case of block randomisation). The factors that are the basis of stratification should normally be included as covariates or stratification variables in the primary outcome model, except where stratification was done purely for an administrative reason.

【訳】共変量間の治療のバランスをとるために層別化を行うことがある；層別化は、治験実施上の理由で用いられることもある（例えば、ブロックランダム化を行った場合のブロック）。通常、層別化を行った因子は、主要な解析モデルにおいて共変量又は層別因子として含めるべきである（ただし、純粋に治験実施上の理由のみで層別化を行った場合を除く）。

- Variables known *a priori* to be strongly, or at least moderately, associated with the primary outcome and/or variables for which there is a strong clinical rationale for such an association should also be considered as covariates in the primary analysis. The variables selected on this basis should be pre-specified in the protocol.

【訳】主要アウトカムと強い、若しくは少なくとも中程度の関連があることが事前に分かっている変数や主要アウトカムとの関連について強い臨床的根拠がある変数は、主要な解析において共変量として考慮するべきである。このように選択する変数は治験実施計画書に事前に規定するべきである。

- Baseline imbalance observed *post hoc* should not be considered an appropriate reason for including a variable as a covariate in the primary analysis. However, conducting exploratory analyses including such variables when large baseline imbalances are observed might be helpful to assess the robustness of the primary analysis.

【訳】ベースライン変数について試験実施後に観察された不均衡は、その変数を主要な解

析の共変量に含めるための適切な理由とはならないと考えるべきである。しかし、ベースライン変数に大きな不均衡が認められた場合に、その変数を含めた探索的解析を実施することは、主要な解析の結果の安定性を評価するのに役立つ可能性がある。

- Variables measured after randomisation and so potentially affected by the treatment should not be included as covariates in the primary analysis.

【訳】ランダム化後に測定され、治療の影響を受けている可能性がある変数は、主要な解析に共変量として含めるべきではない。

- If a baseline value of a continuous primary outcome measure is available, then this should usually be included as a covariate. This applies whether the primary outcome variable is defined as the ‘raw outcome’ or as the ‘change from baseline’.

【訳】主要アウトカムが連続量で、ベースライン値が利用可能な場合、通常、共変量として含めるべきである。これは、主要アウトカム変数を「測定値」と定義する場合と、「ベースラインからの変化量」と定義する場合のいずれにも該当する。

- Covariates to be included in the primary analysis must be pre-specified in the protocol.

【訳】主要な解析に含める共変量は、治験実施計画書に事前明記しなければならない。

- Only a few covariates should be included in a primary analysis. Although larger data sets may support more covariates than smaller ones, justification for including each of the covariates should be provided.

【訳】主要な解析に含める共変量の数は少数とするべきである。データセットが大きければより多くの共変量を含めることができる可能性があるが、共変量ごとに含める理由を示すべきである。

- In the absence of prior knowledge, a simple functional form (usually either linearity or categorising a continuous scale) should be assumed for the relationship between a continuous covariate and the outcome variable.

【訳】事前知識がない場合は、連続な共変量とアウトカム変数の関係について、単純な関数形式（通常、直線又は連続尺度のカテゴリ化）を仮定するべきである。

- The validity of model assumptions must be checked when assessing the results. This is particularly important for generalised linear or non-linear models where mis-specification could lead to incorrect estimates of the treatment effect. Even under ordinary linear models, some attention should be paid to the possible influence of extreme outlying values.

【訳】解析結果を評価する際には、モデルの仮定の妥当性を確認しなければならない。こ

のことは、モデル誤特定が治療効果の誤った推定値に繋がる恐れがある一般化線形モデル又は非線形モデルに対して特に重要である。通常の線形モデルにおいても、極端な外れ値の影響に注意を払う必要がある。

- Whenever adjusted analyses are presented, results of the treatment effect in subgroups formed by the covariates (appropriately categorised, if relevant) should be presented to enable an assessment of the model assumptions.

【訳】調整解析を行う場合は常に、（該当する場合、適切にカテゴリ化した）共変量で構成される部分集団における治療効果の推定結果を示し、モデル仮定の評価を行えるようにすべきである。

- Sensitivity analyses should be pre-planned and presented to investigate the robustness of the primary analysis. Discrepancies should be discussed and explained. In the presence of important differences that cannot be logically explained – for example, between the results of adjusted and unadjusted analyses – the interpretation of the trial could be seriously affected.

【訳】主要な解析の安定性を調査するために感度分析を事前に計画し、その解析結果を示すべきである。主要な解析と感度分析の結果の間に不整合がある場合は、それを議論し説明することが必要である。例えば、調整済みの解析結果と未調整の解析結果の間に論理的に説明できない重要な差がある場合は、治験結果の解釈に重大な影響を及ぼす可能性がある。

【補足説明】Estimandは共通で解析手法だけが異なるとすると、調整済みの解析結果と未調整の解析結果の間で、少なくとも点推定値の値自体が大きく異なることはあまり想定されないと考えられる（3.3節）。このような前提であっても、これらの解析結果の間に大きな差が生じる要因として、調整解析側の問題（例：調整する共変量の数が多すぎる、等）、未調整の解析側の問題（例：ランダム化に何らかの問題があり、共変量の分布が群間で同一でない、等）、あるいはこれらの複合問題の場合が考えられるが、論理的に説明ができない重要な差が認められる場合は、治験結果の解釈に重大な影響を及ぼす可能性がある。なお、EMAガイドラインの発出はICH E9(R1) によるestimandのフレームワークが導入されるよりも前であり、estimand（条件付き（conditional）／条件なし（unconditional））と、estimatorに対する調整解析／未調整解析の区別が明確でない記載が見受けられる（EMAガイドライン6.1節, 7.2節）。これらの区別の詳細については、3.4節を参照されたい。

- The primary model should not include treatment by covariate interactions. If substantial interactions are expected *a priori*, the trial should be designed to allow separate estimates of the treatment effects in specific subgroups.

【訳】主要な解析モデルには治療と共変量の交互作用を含めるべきではない。大きな交互

作用が *事前* に想定される場合は、各部分集団の治療効果を別々に推定することができるように試験を計画するべきである。

- Exploratory analyses may be carried out to improve the understanding of covariates not included in the primary analysis, and to help the sponsor with the ongoing development of the drug.

【訳】主要な解析に含まれない共変量について理解を深め、進行中の医薬品開発を支援するために、探索的解析を実施してもよい。

- In case of missing values in baseline covariates the principles for dealing with missing values as outlined e.g. in the Guideline on missing data in confirmatory clinical trials (EMA/CPMP/EWP/1776/99 Rev. 1) applies.

【訳】ベースライン共変量に欠測が認められた場合は、検証試験における欠測データに関するガイドライン ([EMA/CPMP/EWP/1776/99 Rev. 1](#) [最終アクセス日:2025/5/9]) などに記載されている欠測データの取扱いの原則が適用される。

- A primary analysis, unambiguously pre-specified in the protocol, correctly carried out and interpreted, should support the conclusions which are drawn from the trial. Since there may be a number of alternative valid analyses, results based on pre-specified analyses will carry most credibility.

【訳】治験実施計画書に明確に事前規定され、適切に実施・解釈された主要な解析は、試験から得られた結論の裏付けとなる。他にも多くの妥当な解析が存在する可能性があるが、事前に規定した解析に基づく結果が最も高い信頼性を有するであろう。

2.4 FDA ガイダンス

FDA ガイダンスでは、ランダム化比較試験におけるベースライン共変量の取扱いに対する推奨事項を提供し、共変量調整を伴う解析を適切に実施することにより治療効果の推定及び検定に対する統計的効率（ばらつきの観点から見た推定量の性能。ばらつきが小さい（精度が高い）推定量を、統計的効率が高い推定量と呼ぶ。詳細については3.1節を参照）を向上させることを目的としている。FDA ガイダンスは優越性試験及び非劣性試験のいずれにも適用されるが、非ランダム化試験や欠測データを考慮するためのモデルにおける共変量の利用、経時反復測定データの共変量調整、ベイズ流の手法や機械学習による共変量調整はFDA ガイダンスの範囲外である。

FDA ガイダンスは「一般的な検討事項」、「線形モデル」、「非線形モデル」の三つの観点で臨床試験における共変量調整に対する推奨事項を整理している（FDA ガイダンス本文の全訳及び補足説明については4.2節を参照）。

まず「一般的な検討事項」として、未調整解析の取扱い、共変量調整の一般的な特徴、事前規定の重要性、層別ランダム化の際の層別因子の調整に関する議論、共変量調整と検定及び被験

者数の関係、規制当局との協議を推奨する事項等、モデルを利用した共変量調整か否かに依らず検討しておくことが重要と考えられる点についてまとめている（4.2.4.1節）。

次に「線形モデル」の章では、線形回帰モデルに基づいて共変量を調整した平均治療効果の推定に関する検討事項をまとめている。モデル誤特定の可能性に対する言及が多く、線形回帰モデルに基づく共変量調整では、線形回帰モデルが誤特定されていても平均治療効果の妥当な推定値が与えられること、1:1ランダム化試験の2群比較の場合は、モデル誤特定下であっても平均治療効果の名目標準誤差（標準誤差のモデル推定値）は妥当であること、治療及び共変量の交互作用を含めたモデルでも有効性の主要な解析として平均治療効果の妥当な推定を行うことができること等、既存のガイドラインでは詳細に触れられてこなかった点について、最近の研究に基づいた具体的な言及がなされている（4.2.4.2節）。

最後に「非線形モデル」の章では、主に連続でないアウトカム（二値、順序、カウント、イベント発現までの時間等）の解析における、非線形回帰モデル（一般化線形モデルを含む）を用いた共変量調整の検討事項を整理している。線形回帰モデルと異なり、非線形回帰モデルに基づいて共変量調整を行う場合は、治療と共変量の交互作用がなく、その共変量によって規定される部分集団で治療効果が同一であったとしても、部分集団の条件付き治療効果と条件なし治療効果が一致しない場合があるので、特に条件付き／条件なしの特定が重要となる（詳細は3.2.3, 3.7節を参照）。そのため、estimandの一部として関心の対象とする治療効果を事前明記することを推奨している。非線形回帰モデルに基づいて共変量を調整した上で、治療効果をモデル・パラメータで表す場合の推定は条件付き治療効果の推定であり、これを主たる関心の対象とする場合には、モデルの仮定の妥当性が規制当局との重要な協議事項となるとされている。また、連続でないアウトカムに対して共変量を調整した条件なし治療効果の推定方法についても整理されている（4.2.4.3節）。

2.5 FDA ガイダンスとこれまでのガイドラインの比較

ここでは、共変量の取扱いに関する既存の各ガイドラインの記載と FDA ガイダンスの記載を比較することにより、FDA ガイダンスで新たに焦点を当てている論点を浮き彫りにする。

本章でこれまで述べた通り、FDA ガイダンスは基本的に既存の各ガイドラインを踏まえて作成されていると考えられるため、多くの共通した記載がみられる。具体的には、共変量調整の意義、調整を考慮すべき変数、未調整解析の受け入れ可能性、主要アウトカムのベースライン値の取扱い、事前規定の必要性、調整する共変量の数、層別ランダム化の因子の取扱い、ランダム化後に測定された共変量の取扱い、といったトピックについて、各ガイドラインにそれぞれ共通した記載を確認することができる。これらのトピック別に、各ガイドラインに共通する具体的な記載内容を表 2-1に示す。

一方で FDA ガイダンスは、2.4節にも記載の通り、特にモデルを利用した共変量調整に関して、近年の統計理論の新たな研究成果を踏まえて、既存のガイドラインから一步踏み込む内容となっている。そのため、共通のトピックでありながら既存のガイドラインとは記載内容・記載レベルが異なると考えられる点がある。具体的には、治療と共変量の交互作用、モデル誤特定と感度分析に関する記載である。表 2-2にこれらの相違点を要約する。

表 2-1 ICH E9, EMA ガイドライン, FDA ガイダンスで共通すると考えられる共変量調整に関連する記載

#	トピック	ICH E9	EMA ガイドライン	FDA ガイダンス
(i)	共変量調整の意義	主要変数に重要な影響を及ぼすと予想される共変量と要因は、試験開始前に議論して確認しておくべきであり、精度を向上させ、試験治療グループ間のバランスの欠如を埋め合わせるため、それらを解析でどう取り扱うかを考慮すべきである (5.7節)。	ベースライン共変量と主要なアウトカム変数との間に強い若しくは中程度の関係が認められる場合は、そのようなベースライン共変量の調整は一般に解析の効率を改善し、共変量分布の偶然の不均衡に起因する条件付きのバイアスを回避することに寄与する (Executive summary の冒頭)。	本ガイダンスは、治療効果の推定及び検定に対する統計的効率を向上させるためのベースライン予後共変量 ³ の取扱いを主な焦点としている (I. (4.2.2節))。 治験依頼者はランダム化比較試験の有効性評価項目の解析においてベースライン共変量を調整することができる。これにより、通常、治療効果の推定のばらつきが減少し、信頼区間が狭まり、仮説検定の検出力が高くなる (III.A.の箇条書き2個目 (4.2.4.1節))。
(ii)	調整を考慮すべき変数		主要アウトカム又は変数と強い、若しくは少なくとも中程度の関係があることが事前にわかっている変数についても、主要な解析において共変量として考慮すべきである (Executive summary の箇条書き2個目)。	FDA は治験依頼者に対し、興味のある結果と最も強く関連すると予想される共変量について調整することを推奨している (III.A.の箇条書き4個目 (4.2.4.1節))。
(iii)	未調整解析の受け入れ可能性	調整することの潜在的な有益さが疑わしい場合にはしばしば、調整しない解析に主要な関心を払うと宣言し、調整した解析はそれを補うものとするのが薦められる (5.7節)。	ベースライン共変量は、多くの試験においてアウトカムに影響を及ぼす。ベースライン共変量の調整は必ずしも常に必要ではないが、… (Executive summary の冒頭)	有効性評価項目の主要解析には未調整解析が許容される (III.A.の箇条書き1個目 (4.2.4.1節))。
(iv)	主要アウトカムのベースライン値の取扱い	-	連続変数の主要アウトカム測定のベースライン値が利用可能な場合は、通常、これを共変量として含めるべきである。これは、主要アウトカム変数を「アウトカム」と定義するか、「ベースラインからの変化量」と定義するかに該当する (Executive summary の箇条書き5個目)。	臨床試験では、定義された特性のベースライン値を記録し、その特性のベースライン後の測定値をアウトカムとして使用することが多くある。ベースラインからの変化量を主要評価項目として定義する代わりに (又は加えて)、ベースライン値の調整が一般的に許容される。治験依頼者がベースラインからの変化量ではなく、変化率としてアウトカムを定義することを提案する場合、該当の審査部門とアウトカムの定義及び共変量調整の使用につ

#	トピック	ICH E9	EMA ガイドライン	FDA ガイダンス
				いて議論すべきである。同様に、治験依頼者が治療群の平均値の差ではなく比に対して非劣性検定を実施することを提案する場合も、ベースラインからの変化に関するアウトカムの定義と共変量調整の使用について該当の審査部門と議論すべきである（III. A.の箇条書き10個目（4.2.4.1節））。
(v)	事前規定の必要性	臨床試験の計画立案の際、データの最終統計解析の主要な特徴は、治験実施計画書の統計の部に記述すべきである（5.1節）。 共変量の影響又は部分集団の効果を調整することは予定した解析の重要部分となることがあり、治験実施計画書に記述すべきである。主要変数に重要な影響を及ぼすと予想される共変量と要因は、試験開始前に議論しておくべきであり、精度を向上させ、試験治療グループ間のバランスの欠如を埋め合わせるため、それらを解析でどう取り扱うかを考慮すべきである（5.7節）。	主要な解析に含める共変量は、プロトコールに事前明記しなければならない（Executive summary の箇条書き6個目）。 プロトコールに明確に事前規定され、適切に実施・解釈された主要な解析は、治験から得られた結論の裏付けとなる。他にも妥当な多くの解析が存在する可能性があることから、事前に規定した解析に基づく結果が最も高い信頼性を有する（Executive summary の箇条書き15個目）。	治験依頼者は比較データの治療の割付けが明らかになる前に共変量を調整する解析の詳細な手順を事前に規定すべきである。FDA の審査は、事前に規定されていないモデルや共変量を使用した事後的に実施される解析ではなく、事前に規定された主要解析を重視する（III.A.の箇条書き3個目（4.2.4.1節））。
(vi)	調整する共変量の数	-	主要な解析に含める共変量の数は少数とすべきである。データセットが大きければより多くの共変量を含めることができる可能性があるが、共変量ごとに含める理由を示すべきである（Executive summary の箇条書き7個目）。	共変量調整の統計学的特性は、試験で調整に用いる共変量の数が被験者数に対して少ない場合に最もよく理解される（Tsatis et al. 2008）。従って、治験依頼者は、共変量の数が被験者数に対して多い場合、又は多くの水準を持つ共変量（例えば、多くの施設を有する試験の場合の試験施設）で調整を提案する場合、その提案を該当の審査部門と協議すべきである（III.A.の箇条書き12個目（4.2.4.1節））。
(vii)	層別ランダム化の因子の取り扱い	計画時に一つ以上の因子を層別因子として用いたならば、それらの因子を解析時に考慮することが適切である（5.7節）。	通常、層別化を行った因子は、主要な解析モデルにおいて共変量又は層別因子として含めるべきである（ただし、純粋に治験実施上の理由のみで層別化を行った場合	ランダム化はベースライン共変量によって層別化されることが多い。共変量調整モデルには一般的に層別変数を含むべきであり、層別ランダム化に使用されなか

#	トピック	ICH E9	EMA ガイドライン	FDA ガイダンス
			を除く) (Executive summary の簡条書き1 個目)	った共変量も含めることができる。場合 によっては、誤った層別化が起こり、実際 のベースライン層別変数とランダム化と してのベースライン層別変数が一致しな いことがある。共変量調整モデルには、事 前に指定されている限り、どちらの層別 変数の定義も使用できる (III.A.の簡条書 き7個目 (4.2.4.1節))。
(viii)	ランダム 化後に測 定された 共変量の 取扱い	ランダム化後に測定された共変量を主要 な解析で調整することは薦められない。 なぜならばそのような共変量は試験治療 の影響を受けている可能性があるからで ある (5.7節)。	ランダム化後に測定され、治療の影響を 受けている可能性がある変数は、主要な 解析に共変量として含めるべきではない (Executive summary の簡条書き4個目)。	また ICH E9ガイドラインは、「ランダム化 後に測定された共変量」は試験治療の影響 を受けている可能性があるため、その調 整に対して注意を促している (II. (4.2.3 節))。

表 2-2 ICH E9 / E9(R1), EMA ガイドライン, FDA ガイダンスで記載内容・記載レベルが異なると考えられる共変量調整に関連する記載

#	トピック	ICH E9 / E9(R1)	EMA ガイドライン	FDA ガイダンス
(i)	治療と共変量の交互作用	…交互作用は予想されたり、交互作用自体に重要な関心がある場合（例えば高齢者医療など）があり、したがって部分集団別解析又は交互作用を含んだ統計モデルは予定した検証的解析に含まれる場合がある。しかし、多くの場合、部分集団別解析又は交互作用解析は探索的であるため、探索的であることを明確に確認しておくべきである。それらの解析では、全体でみられた試験治療の効果がすべて一様であるかどうかを探索すべきである。一般に、そのような解析は、まず問題となっている統計モデルに交互作用項を加えることから始めるべきであり、適切な被験者の部分集団内、又は共変量によって定められる層内を更に探索的に解析することによって補われる。探索的解析である場合、これらの解析結果は慎重に解釈すべきである（ICH E9 5.7節）。	主要なモデルには治療と共変量の交互作用を含めるべきではない。大きな交互作用が事前 ¹ に予測される場合は、特定の部分集団における治療効果を別々に推定することができるように試験を計画すべきである（Executive summary の簡条書き 12 個目）。	線形モデルには治療×共変量の交互作用を含めることができる。交互作用を含めたモデルを用いる場合でも、モデルに基づく平均治療効果の推定値に基づいて主要な解析を行うことができる（Tsatis et al. 2008; Ye et al. 2022）。一方、ICH E9 ガイドラインで述べられているように、交互作用効果は補助解析又は探索的解析において評価することが重要である（III.B. の簡条書き 6 個目（4.2.4.2 節））。
(ii)	モデル誤特定／感度分析	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。感度分析は、規制当局の意思決定及び医薬品の情報の表示において重要となるすべての estimand の主とする推定量について計画すべきである。これは治験依頼者と規制当局とで議論し、合意する話題となり得る（ICH E9(R1) A5.2.1 節）。	解析結果を評価する際には、モデル仮定の妥当性を確認しなければならない。このことは、モデル誤特定が治療効果の誤った推定値に繋がる恐れがある一般化線形モデル又は非線形モデルに対して特に重要である。通常の線形モデルにおいても、極端な外れ値の影響に注意を払う必要がある（Executive summary の簡条書き 9 個目）。	主解析で条件付き治療効果の推定に非線形回帰を用いる場合、治験依頼者はそれを含む治験実施計画書又は統計解析計画書で具体的な提案を該当の審査部門と協議するべきである。非線形回帰により条件付き治療効果を推定する際、モデルの仮定は一般的に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かなり異なる場合、結果の解釈が困難となる可能性がある。解釈可能性はモデル特定の質と共に向上する。解析に用いるモデルの仮定に対して計画した評価について、治験依頼者は該当の審査部門と協議するべきである（III.C. の簡条書き 6 個目（4.2.4.3 節））。

最後に、既存のガイドラインでは触れられておらず、FDA ガイダンスで新たに記載された点を以下に要約する。各項目の詳細については、本稿の関連する記載を参照されたい。

【線形モデル (4.2.4.2 B. Linear Models)】

線形モデルによる共変量調整は、平均治療効果（条件なし治療効果）を推定する方法として許容される (4.3.2 節)。

主要な解析としてモデルを利用して平均治療効果を推定しようとする場合、交互作用を含めたモデルに基づく推定を行うことができる (4.3.2 節)。

線形回帰モデルが誤特定された場合であっても、線形モデルを用いた共変量調整により平均治療効果の妥当な推定値を与えることができる (Lin 2013a; Wang et al. 2019; 4.3.4 節)。

1:1 ランダム化比較試験における 2 群比較では、モデル誤特定下であっても、平均治療効果の標準誤差のモデル推定値は妥当である (Freedman 2008b; Wang et al. 2019; 4.3.4 節)。

1:1 ランダム化比較試験における 2 群比較以外の場合は、平均治療効果の標準誤差のモデル推定値はモデル誤特定下でバイアスを伴う恐れがあるため、サンドウィッチ型のロバスト分散 (Rosenblum and van der Laan 2009; Lin 2013a; Bartlett 2020) の利用を推奨する (4.3.4 節)。

【非線形モデル (4.2.4.3 C. Nonlinear Models)】

条件付き治療効果の推定が関心の対象である場合、治療及びベースライン共変量を説明変数とした非線形回帰モデルを用いることは、条件付き治療効果を推定しようとするものである (3.2.3 節, 3.4 節, 4.3.5 節)。

非線形モデルによる共変量調整では、オッズ比等の一部の要約指標において、併合不能性（すべての部分集団で治療効果が同一であっても、部分集団の条件付き治療効果が条件なし治療効果と異なる）の問題が生じることがある。このことを踏まえ、条件付き治療効果と条件なし治療効果のいずれに主要な関心があるか、事前に *estimand* の一部として明記すべきである (3.7 節, 4.3.5 節)。
二値、順序、カウント、イベント発現までの時間等の諸種のアウトカムに対して、回帰モデル誤特定に対して安定性を持つ条件なし治療効果の共変量調整済み推定量が提案されている。

「標準化」(standardized)、 「プラグイン」(plug-in) 又は「g-computation」推定量や逆確率治療重み付け (Inverse probability of treatment weighting [IPTW]) 法も、共変量を調整した条件なし治療効果の推定に利用可能である (3.9 節; JPMA 2020)。

3. FDA ガイダンスの理解に役立つ用語の概説

本章では、FDA ガイダンスの理解の助けとなる用語を概説する。FDA ガイダンスの理解に有益な用語は、FDA ガイダンスでは直接使用されていない概念も積極的に取り上げた。

本章の構成は以下の通りである。まず、統計的効率について定義を与え、次に、ICH-E9(R1) に基づき、estimand と estimator を取り上げた。FDA ガイダンスの理解において、各用語の議論の対象が estimand であるのか estimator であるのかを区別する意識が重要であるからである。本区別が重要な用語については、本報告書の各節タイトルにて estimand/estimator の別を付記した。また、FDA ガイダンスを十分に理解するための前提知識として必要な estimand の理解を深めるため、ICH E9(R1) や FDA ガイダンスの本文中では明記されていないものの、潜在結果変数及び潜在結果変数を用いた estimand の定式化に関する用語の説明を行った。引き続き、その他の重要な用語の概説を与えた。

なお、必要に応じて関連論文も引用しつつ補足を加えたが、本報告書では概説に留めるため、詳細に興味のある読者は各引用文献を参照されたい。

3.1 統計的効率

統計的効率 (statistical efficiency) は、推定量や推定方法を評価するための指標である。統計的効率が高い推定量は、ばらつき (分散や標準誤差) が小さく、精度 (precision) が高い推定量を指す。また、統計的効率が高い推定方法は、同じ被験者数でより高い検出力を得る方法、又は同じ検出力を達成するために必要な被験者数が少ない方法を指す (4.3.6 節)。この指標を用いることで、推定量やモデル、及び試験デザインの最適化が可能となる。

統計的効率は、推定量 A の分散に対する推定量 B の比 (推定量 B の分散/推定量 A の分散) を用いる相対効率 (relative efficiency) によって評価される。この比が1以上であれば、推定量 A は推定量 B よりも統計的効率が高いと判断できる。統計的効率に関連する概念として精度があり、これは推定量の分散の逆数として定義される (Steingrimsson et al. 2017; Rothman and Lash 2021; Chen et al. 2022)。

3.2 Estimand に関連する用語

3.2.1 Estimand

ICH E9(R1) では、estimand を「定められた臨床試験の目的によって提起される臨床的疑問を反映する治療効果の詳細な説明」と紹介されている。関心のある臨床的疑問に応じて何を推定すべきか (すなわち、「推定されるべきもの」) を明確にすることによって推定される治療効果と試験目的の潜在的な不整合を回避できると期待されることから、estimand を試験計画時に明示的に検討することが重要とされている (JPMA 2022)。ICH E9(R1) の図1によると、試験計画は、試験目的の設定、estimand の定義、推定方法の選択、といった順番で進められるべきである (2.2 節)。特に、ICH E9(R1) training material より抜粋した図 3-1 にもある通り、推定方法 (推定量) から estimand を遡り決定することは望ましくないとされている。

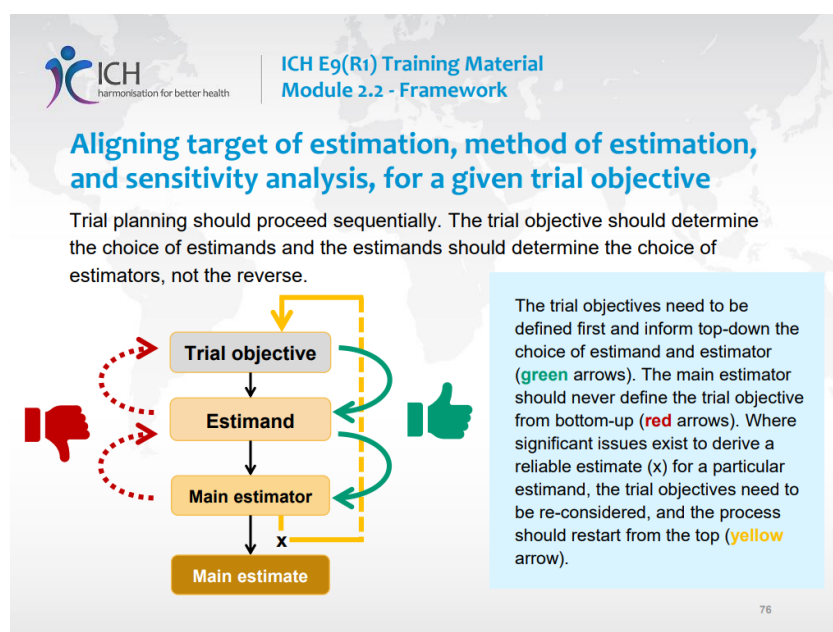


図 3-1 検討手順の流れについて (ICH E9(R1) training material より)

2.2節にある通り, estimand は, 五つの要素 (治療, 対象集団, 変数, その他の中間事象, 集団レベルでの要約) によって構成され, その詳細な規定が必要となる. FDA ガイダンス II. (4.2.3節) にもある通り, 共変量調整して推定される治療効果は変数の集団レベルでの要約にあたる. FDA ガイダンスの主眼となる共変量を調整して推定される治療効果に関する議論では, 統計的効率の観点から優れていると考えられる推定方法の選択だけでなく, 推定方法の選択に先立ち, estimand を定義する集団レベルでの要約が試験目的と整合しているかも重要な論点となる.

3.2.2 潜在結果変数

ICH E9(R1) や FDA ガイダンスの本文中では明記されていないものの, 潜在結果変数を用いた因果推論の枠組みを示唆する記載がなされている. 本節では, 以下で述べる estimand の議論の理解を助ける枠組みとして有益である潜在結果変数について, 概説する.

潜在結果変数 (もしくは反事実結果変数) とは, ある被験者が任意の治療を仮に受けた場合に観察されるであろう結果変数を指す (田中 2019; Imbens and Rubin 2015). 例えば, 結果変数を Y , 試験治療もしくは対照治療を受けたかを A ($A = 1$ なら試験治療, $A = 0$ なら対照治療) とする. その場合, 潜在結果変数としては, 「試験治療を受けた場合に観察されるであろう潜在結果変数 Y^1 」もしくは「対照治療を受けた場合に観察されるであろう潜在結果変数 Y^0 」が考えられる. しかし, 現実ではどちらか一つの状況下での潜在結果変数しか観察できないため, 実際に観察される事実とは反する潜在的な事実 (いわゆる, 反事実) を取り扱うために定式化されたモデルとして, 「潜在結果モデル (もしくは反事実モデル)」と呼ぶ. 潜在結果モデルは統計的因果推論で用いられるが, 潜在結果モデルの考え方に基づいて estimand としての治療効果を潜在結果変数を用いて表現することも可能であり, 次節にて説明する.

なお, 統計的因果推論のより網羅的かつ詳細な説明については, 総説 (田中 2019) や和書 (岩崎 2015; 宮川 2004), Hernán & Robins の “Causal inference: what if” (Hernán and Robins 2020)

や日本製薬工業協会医薬品評価委員会データサイエンス部会2020年度継続タスクフォース1による報告書「ICH E9(R1) の理解に役立つ因果推論」(JMPA, 2020)などを参照いただきたい。

3.2.3 条件付き治療効果と条件なし治療効果

本節では, estimand としての条件なし治療効果 (unconditional treatment effect) と条件付き治療効果 (conditional treatment effect) について補足する。

FDA ガイダンス II. (4.2.3節) にもある通り, FDA ガイダンスは ICH E9(R1) と整合する旨が述べられており, 共変量を調整して推定される治療効果は, estimand を定義する集団レベルでの要約に相当すると説明されている。そのため, FDA ガイダンスで取り上げられている条件なし治療効果及び条件付き治療効果はいずれも estimand を定義する集団レベルでの要約のことを指す。FDA ガイダンス III.C.の箇条書き3個目 (4.2.4.3節) にもある通り, 治験依頼者は解析で対象とする治療効果が条件付きと条件なし治療効果のいずれであるかを明記するべきである。

FDA ガイダンスにおいて, 条件なし治療効果とは, 対象集団を未治療から治療に移行させた際の集団レベルでの効果を定量化するものと説明されている。FDA ガイダンス以外では, 上記定義に基づく治療効果を周辺治療効果 (marginal treatment effect) と呼ぶ場合もある (Van Lancker et al. 2024; Wei et al. 2024)。FDA ガイダンス III.B.の箇条書き2個目 (4.2.4.2節) にもある通り, 条件なし治療効果の一例として, 平均治療効果が挙げられる。FDA ガイダンス中に記載はないが, Van Lancker et al. (2024) を参考に, 反事実モデルの枠組みを導入すると, estimand を定義する集団レベルでの要約としての平均治療効果は,

$$E(Y^1 - Y^0)$$

となる。ただし, E は対象集団に関する期待値を表す。同様に, 比 (正の値を取るアウトカムの場合) やオッズ比 (二値アウトカムの場合) の尺度の場合,

$$E(Y^1)/E(Y^0)$$

$$\frac{E(Y^1)/\{1 - E(Y^1)\}}{E(Y^0)/\{1 - E(Y^0)\}}$$

となる。

一方, FDA ガイダンスでは, 条件付き治療効果の定義が明確に記載されていないが, 部分集団における治療効果を指す際に用いられている。再び Van Lancker et al. (2024) にならうと, ベースライン共変量 X の特定の値 x の下での条件付き治療効果を表現すると,

$$E(Y^1 - Y^0 | X = x)$$

$$E(Y^1 | X = x) / E(Y^0 | X = x)$$

$$\frac{E(Y^1 | X = x) / \{1 - E(Y^1 | X = x)\}}{E(Y^0 | X = x) / \{1 - E(Y^0 | X = x)\}}$$

となる。ただし, $E(\cdot | X = x)$ は対象集団のうち $X = x$ の部分集団に関する条件付き期待値を表す。

3.2.4 Model-based estimand/ model-free estimand

3.2.3節では、潜在結果変数の期待値もしくは条件付き期待値を用いて異なる治療を受けた下での結果の比較として治療効果を表現した。これらはいずれも期待値のみで表現され、特定のモデルに依存しないことから、model-free estimand と呼ぶことがある (Van Lancker et al. 2024)。一方、実際の臨床試験における estimand を考える際に以下のように、期待値の構造に対して特定のモデル（例えば、一般化線形モデル）

$$g\{E(Y|A, X)\} = \beta_0 + \beta_1 A + \beta_2 X$$

を仮定し、治療効果がモデルのパラメータとして表現されている場面も少なくない。ここに、 $g(\cdot)$ はリンク関数を表す。上記モデルのように、estimand の集団レベルでの要約が統計モデルのパラメータとして表現される estimand を Van Lancker et al. (2024) では model-based estimand と呼ぶ。

なお、本報告書ではどちらの estimand がより適切かについて言及しないが、詳細については、Van Lancker et al. (2024) 及びその参考文献等を参照いただきたい。

3.3 Estimator に関連する用語

3.3.1 Adjusted estimator/unadjusted estimator

本節では、解析や推定・推定量といった推定に関連する用語を修飾するために FDA ガイダンス中で利用されている“調整済み (adjusted)”と“未調整 (unadjusted)”について補足する。なお、FDA ガイダンス II. (4.2.3節)にある通り、FDA ガイダンスにおける共変量調整とはベースライン共変量の利用を指しているため、ベースライン共変量の利用に焦点を当てる。

未調整推定量 (unadjusted estimator) とは、ベースライン共変量の情報を用いずに構成される推定量である。例えば、 Y^1 の期待値 $E[Y^1]$ の推定に関心がある場合、 n 例の試験データ全体から $\sum_{i=1}^n A_i Y_i / \sum_{i=1}^n A_i$ によってランダム化及び一致性の仮定 ($Y_i = A_i Y_i^1 + (1 - A_i) Y_i^0$) の下ではバイアス（偏り）なく推定できる場合があるが、共変量の情報は考慮されていない。

一方、調整済み推定量 (adjusted estimator) とは、ベースライン共変量の情報を用いて構成される推定量である。調整済み推定量は、第一種の過誤確率の増大やバイアスを抑えながら、推定精度の上昇につながる可能性がある。ベースライン共変量の情報を用いる代表的な方法として、層別化や回帰モデルによる共変量調整法 (ANalysis of COVAriance [ANCOVA], ロジスティック回帰分析など)以外にも、FDA ガイダンスの4.2.4.3節で紹介されている g-computation や IPTW 法などが挙げられる。詳細については、FDA ガイダンス III.B. (4.2.4.2節) 及び III.C. (4.2.4.3節)を参照いただきたい。

3.4 「条件付き治療効果と条件なし治療効果」と「Adjusted estimator/unadjusted estimator」の違い

3.2.3節では、estimand を定義する集団レベルでの要約として、条件なし治療効果と条件付き治療効果について紹介した。条件なし治療効果は、対象集団全体が異なる治療を受けた場合に観察されるであろう結果の比較に興味関心があるが、条件付き治療効果は対象集団の中の部分集

団における治療効果の要約に興味関心がある。さらに、それらを **model-free** もしくは **model-based estimand** として表現できることを3.2.4節で紹介した。一方、3.3.1節では、関心のある **estimand** を推定する方法が、共変量の情報を用いるかによって、未調整推定量と調整済み推定量に大別されることを説明した。FDA ガイダンスを読み進めていくにあたって、“条件付き/条件なし (**conditional/unconditional**) ”は **estimand** を定義する集団レベルでの要約を修飾するために用いられることに対し、“調整済み/未調整 (**adjusted/unadjusted**) ”は解析や推定方法の修飾のために用いられることに留意いただきたい。

最後に、“条件付き/条件なし (**conditional/unconditional**) ”な **estimand** と “調整済み/未調整 (**adjusted/unadjusted**) ”な推定方法の対応関係について言及する。FDA ガイダンス及び本報告書に触れるまでは、未調整推定量からは条件なし治療効果、調整済み推定量からは条件付き治療効果が推定されると連想された読者も少なくないだろう。実際、EMA ガイドラインの6.1節 (2.3節) においても、非線形モデルを考える場合に、“条件付き (**conditional**) ”であれば “調整済み (**adjusted**) ” (同じく、“条件なし (**unconditional**) ”であれば “未調整 (**unadjusted**) ”) のように同一視することを示唆するような記載がなされている。ランダム化比較試験の統計解析において、アウトカムの治療のみへの回帰や未調整なアウトカムの平均値の群間差などにより、暗黙のうちに推論されるのは条件なし治療効果である。また、共変量の影響を考慮するために層別化や回帰モデルによる共変量調整法が多く用いられているが、暗黙のうちに推論されるのは条件付き治療効果である。しかし、FDA ガイダンス III.C.の簡条書き9個目と10個目 (4.2.4.3節) でも紹介されている **g-computation** や **IPTW** 法などによって、条件なし治療効果の推定に共変量の情報を利用した調整済み推定量を用いることもできる。また、FDA ガイダンス III.B.の簡条書き1個目と2個目 (4.2.4.2節) で示されているように、線形モデルの場合、共変量を説明変数として含む回帰モデルから得られた治療割付け変数の回帰係数が条件なし治療効果として解釈できる場合もある。このように、“条件付き (**conditional**) ”と “調整済み (**adjusted**) ” (同じく、“条件なし (**unconditional**) ”と “未調整 (**unadjusted**) ”) をひとまとめでできないことを改めて留意すべきである (Daniel et al. 2021)。最後に、ICH E9(R1) の **estimand** に関する検討プロセス (**thinking process**) を踏まえ、**estimand** の要素として関心のある “条件付き/条件なし (**conditional/unconditional**) ”治療効果を先に決定した上で、治療効果の推定に “調整済み/未調整 (**adjusted/unadjusted**) ”推定量を用いるか決定すべきであることを忘れてはいけない。

3.5 予後因子/ベースライン予後共変量と効果修飾因子/ベースライン予測共変量

予後因子とは、アウトカムY (予後) と相関する因子である (田中 2019)。例えば、がんによる死亡においては、がんのステージやパフォーマンスステータス (がん患者の全身状態) が予後因子として知られている。一方、効果修飾因子とは、治療効果と相関する因子である (田中 2019)。例として、分子標的治療薬における特定の遺伝子変異が挙げられる (折笠 2022)。この例のように、臨床試験では効果修飾因子は臨床検査や画像などのバイオマーカーであることが多い。予測マーカーという用語を同じ意味で用いることも多い。

治療効果の指標を差としたときの予後因子と効果修飾因子の違いを表した図を図 3-2に示す (『医薬統計ポケット資料集 2023年版』を参考に作成)。

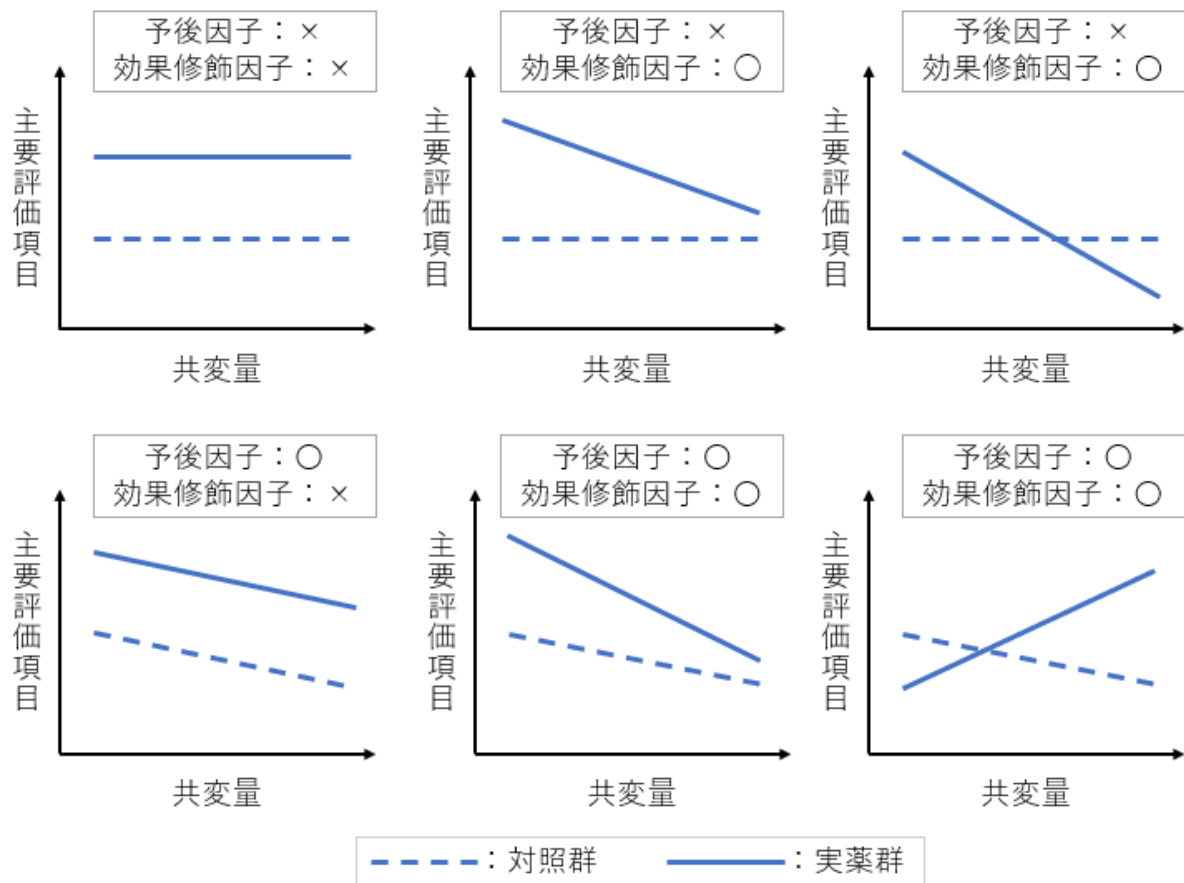


図 3-2 治療効果の指標を差としたときの予後因子と効果修飾因子の違い

図 3-2 の効果修飾因子がある（効果修飾因子: o）場合を考えたとき、効果修飾因子の値に依存して、治療効果の大きさである差が変化する。なお、効果修飾因子の治療効果の修飾の程度は、治療効果の指標によって異なることが知られている（田中 2019）。

FDA ガイダンス I の脚注3（4.2.2 節）では、『本ガイダンスで使用する「ベースライン予後共変量（prognostic baseline covariates）」という用語は、主要評価項目と関連する可能性が高いベースライン共変量を指す。治療が有益となる可能性が高い集団を特定するために用いられる「ベースライン予測共変量（predictive baseline covariates）」については、本ガイダンスの対象外である。』と記載されている。この記載から FDA ガイダンスでは、予後因子と思われるベースライン共変量を「ベースライン予後共変量」、効果修飾因子と思われるベースライン共変量を「ベースライン予測共変量」と呼んでいると考えられる。

3.6 交絡

3.2.3 節では、因果推論の枠組みに基づき、治療効果を対象集団における潜在結果変数の期待値の比較である因果効果（例えば、 $E(Y^1 - Y^0)$ ）として表現した。一方、観測データに基づき推定される治療群間の比較（例えば、 $E(Y|A = 1) - E(Y|A = 0)$ ）は、関連の指標である。関連の指標が因果効果と一致するためには、いくつかの仮定が必要となる。詳細な識別条件については、3.2.2 節などで紹介した参考文献を参照いただきたいが、一貫性（consistency）と交換可能性

(exchangeability) の仮定の下では,

$$\begin{aligned} E(Y^1 - Y^0) &= E(Y^1) - E(Y^0) \\ &= E(Y^1|A = 1) - E(Y^0|A = 0) \\ &= E(Y|A = 1) - E(Y|A = 0) \end{aligned}$$

が成立する. ただし, 一致性とは, 治療 $A = a$ を受けた場合に観察されるアウトカムが $Y = Y^a$ であること, つまり, 治療が二値の場合, $Y = AY^1 + (1 - A)Y^0$ の関係が成り立つことを表す. また, 交換可能性とは, 潜在結果変数 Y^a と治療 A が独立である ($Y^a \perp A, a \in \{0,1\}$) ことである. このとき, $E(Y^a|A = 1) = E(Y^a|A = 0) = E(Y^a)$ が成り立つ. [Greenland et al. \(1999\)](#) では, 交換可能性がないことを交絡と呼んでいる. 理想的な実施状況のランダム化比較試験においては, ランダム化によって潜在結果変数と治療の独立性が担保されるため, 一致性の仮定のもとで, 関連の指標が因果効果と一致する.

一方, 観察研究などにおいては, 治療群間の交換可能性は一般に担保されず, 因果効果の推定には, 観察されたある共変量が等しい層の中では治療群間の交換可能性が成り立つという仮定が必要となる. このように, 観察された共変量を条件付けた下で潜在結果変数と治療が独立である ($Y^a \perp A|X$) ことを条件付き交換可能性 (conditional exchangeability) や「未測定交絡がない」(no unmeasured confounding) と呼ぶことがある. また, この条件付き独立性が成り立つために必要な共変量を交絡因子と呼ぶ. 未測定交絡がないという仮定の下では, 以下のように各層において観察された条件付き治療効果を対象集団全体の共変量分布 $P(X)$ で標準化 (周辺化) することで, 対象集団全体における因果効果を推定できる.

$$E(Y^1 - Y^0) = \sum_x [E(Y^1 - Y^0|X)P(X)] = \sum_x [E(Y|A = 1, X = x) - E(Y|A = 0, X = x)]P(X = x)$$

FDA ガイダンス III.C. (4.2.4.3節) で紹介されている g-computation は上記の標準化の考え方に基づいている.

しばしばアウトカムと治療の両方と相関する因子のことを交絡因子と説明される場合があるが, こちらの定義には合流点 (collider) や中間因子 (intermediate) など含まれてしまい, これらの因子について調整してしまうとむしろバイアスが生じてしまうことが知られている. 交絡因子を特定するために, DAG (Directed acyclic graph) を用いた方法が一つの有用な手段として知られている. 詳細については, [Pearl \(2009\)](#) を参照いただきたい.

3.7 併合可能性と併合不能性

FDA ガイダンスでは, 併合不能性 (noncollapsibility) を「すべての部分集団の条件付きの治療効果が同一である場合にも, この部分集団特有の治療効果が条件なし治療効果と異なる場合」と説明しているが, 併合可能性 (collapsibility) の説明はない. 全体集団の効果指標の値 (周辺治療効果, もしくは, 条件なし治療効果) が各部分集団の効果指標の値 (条件付き治療効果) の重み付き平均として表現できることを併合可能といい, 特に, 部分集団の効果指標の値が部分集団間で共通あるときに, 全体集団の効果指標の値と同じであることを狭義の併合可能性 (strict collapsibility) という ([Greenland 1996](#); [Greenland et al. 1999](#); [Didelez and Stensrud 2022](#)). なお, 古い文献においては, 併合不能性が, 交絡によるバイアスと区別されていないため, 注意が必要で

ある (Greenland 1996; Greenland et al. 1999) . 併合不能性は交絡や効果修飾因子がない場合でも、オッズ比やハザード比といった推定量の特性により生じる場合がある (田中 2019) . 平均値やリスク差、リスク比は併合可能であるが、オッズ比やハザード比は併合不能であることが知られている (Wei et al. 2024) .

表 3-1は FDA ガイダンスに記載されているオッズ比における併合不能性の例である. この表では、オッズ比を指標として、試験開始前に得られたバイオマーカーの結果ごとに得られた治療群ごとの有効率 (success rate) と、層別のオッズ比を示している. バイオマーカー陽性及び陰性の層別オッズ比はともに8.0となるが、層を考慮せずに全体のデータを使用して治療効果の推定を行うとオッズ比は4.8となる. 層別のオッズ比は同一であるにもかかわらず、層別のオッズ比と全体のオッズ比が異なっており、併合不能といえる.

なお、バイオマーカー陽性 ($X = +$) 及び陰性 ($X = -$) の層別オッズ比は以下のように求められる.

$$\frac{E(Y^1|X = +)/\{1 - E(Y^1|X = +)\}}{E(Y^0|X = +)/\{1 - E(Y^0|X = +)\}} = \frac{0.8/(1 - 0.8)}{0.333/(1 - 0.333)} = 8.0$$
$$\frac{E(Y^1|X = -)/\{1 - E(Y^1|X = -)\}}{E(Y^0|X = -)/\{1 - E(Y^0|X = -)\}} = \frac{0.25/(1 - 0.25)}{0.04/(1 - 0.04)} = 8.0$$

一方で、層を考慮しない全体のオッズ比は以下のように求められる.

$$\frac{E(Y^1)/\{1 - E(Y^1)\}}{E(Y^0)/\{1 - E(Y^0)\}} = \frac{0.525/(1 - 0.525)}{0.187/(1 - 0.187)} = 4.8$$

表 3-1 Non-collapsibility of the Odds Ratio in a Hypothetical Target Population

	Percentage of target population	Success rate		Odds ratio
		New drug	Placebo	
Biomarker-positive	50%	80.0%	33.3%	8.0
Biomarker-negative	50%	25.0%	4.0%	8.0
Combined	100%	52.5%	18.7%	4.8

被験者数の計算等の目的で公表された文献の数値を参照する場合、併合不能な指標については、条件なし治療効果と条件付き治療効果の点推定値自体が異なるため、値自体を比べる場合には、計画している試験で推定したい治療効果と同様の効果を推定している文献を参照する必要がある点に留意されたい. Daniel et al. (2021) は、因果対数オッズ比や因果対数ハザード比について、回帰により調整された条件付き治療効果を用いて条件なし治療効果を推定する方法を紹介しており、このような目的に活用できると考えられる.

3.8 モデル誤特定

モデル誤特定とは、興味のある治療の推定のために仮定した統計モデルが、根底にある真のデータ生成プロセスを正確に捉えていないことを意味する. 例えば、線形性、交互作用の効果、

アウトカム変数の分布特性について誤った仮定を置く場合が考えられる (Van Lancker et al. 2024) .

FDA ガイダンス III.B.の箇条書き3個目、4個目及び6個目 (4.2.4.2節) とIII.C.の箇条書き6個目、8個目及び11個目 (4.2.4.3節) で、モデル誤特定について言及されている。また、FDA ガイダンスで述べられている線形モデルにおけるモデル誤特定の論点を4.3.4節に整理している。

3.9 Model-assisted な手法

FDA ガイダンスでは、“model-assisted”という用語は直接的には記載されていないが、この手法に関連する記述や議論が多く含まれていると考えられるため、本節で概説する (FDA ガイダンス III.B.の箇条書き3個目、4個目、5個目 (4.2.4.2節) と III.C.の箇条書き6個目、8個目、11個目 (4.2.4.3節)) .なお、がんの第1相試験で使用される“model-assisted”とは意味合いが異なる点に注意が必要である (Zhou et al. 2018; JPMA, 2023b) .

Model-assisted な手法 (model-robust と呼ばれることがある) は、モデル誤特定下でも漸近的に妥当な推論が可能であり (一致推定量を常に得られる) , さらにモデルが適切に適合すれば推定精度を向上させることができる手法である (Ye et al. 2022; Wang et al. 2023; Bannick et al. 2025) . 臨床試験におけるランダム化の下で広く用いられる ANalysis Of VAriance [ANOVA] や ANCOVA も、model-assisted な手法の一例である (Ye et al. 2022) .

ランダム化比較試験では、統計モデルを用いずとも、治療割付けに基づく単純な平均値の比較により平均治療効果を推定できる (Lin 2013b) . 治療割付けのランダム化のみに基づく推論は randomization-based 又は design-based な推論と呼ばれ、結果変数に対するパラメトリックな仮定を必要としない (Cox 2009; Aronow and Middleton 2013) . さらに臨床試験では、治療群間におけるベースライン共変量のランダムな差 (条件付きバイアス; 2.3節) を調整し、推定効率を改善させる目的で統計モデルを利用することがある。一般的にはサンプルが十分に大きく、共変量がアウトカムと関連している場合には、共変量調整を行うことで推定精度の向上が期待できる。ただし、モデルが誤特定されている場合には、かえって精度が悪化する可能性もある (Freedman 2008b; Lin, 2013a; Ye et al. 2022) . 一方、ランダム化を伴わずに model-based な手法 (統計モデルが真のデータ生成プロセスを正確に捉えているという仮定に基づく手法) のみに依存すると、モデル誤特定によって妥当な推論が得られなくなる可能性がある。これに対し、ランダム化された臨床試験において model-assisted な手法として統計モデルを利用する場合、モデルが誤特定されたとしても design-based な推論の観点から、比較的弱い仮定の下で漸的に妥当な推論が可能となる (Aronow and Middleton 2013) . このような model-assisted な手法における統計モデルは、一般に作業モデル (working model) と呼ばれる。

Lin (2013b) によれば、標本調査 (survey sampling) や計量経済学の分野でも同様の手法が用いられている。標本調査では、design-based な推論の枠組み内で推定効率を改善するためにモデルを利用する model-assisted な手法が広く利用されている (Särndal et al. 1992; Rao 2005; Särndal 2010) . また、計量経済学では、回帰分析を“agnostic”な視点 (すなわち、無限母集団からのランダムサンプリングを仮定し、モデル自体の正しさに依存しない視点) で用いることがある。

Model-assisted な手法としては、上記のランダム化の下で用いられる ANOVA や ANCOVA に加

え, ANalysis of HEterogeneous COVariance (FDA ガイダンス III.B.の箇条書き6個目 (4.2.4.2節) ; ANalysis of HEterogeneous COVariance [ANHECOVA]; 4.3.4.2 節) や標準化 (standardization; standardized, plug-in 又は g-computation 推定量を求めるための手法; FDA ガイダンス III.C.の箇条書き9個目) や, IPTW 法 (FDA ガイダンス III.C.の箇条書き10個目) などが挙げられる (Bannick et al. 2025) .

3.10 線形モデル・非線形モデル

線形モデルには、非線形リンク関数 (nonlinear link functions) を用いないモデル、つまり、恒等リンク関数を用いたモデルが含まれる。ただし、FDA ガイダンスの線形モデルに関する記載は、アウトカムに正規分布を仮定する線形モデルのみを対象としていると考えられる (FDA ガイダンスの2021年度ドラフト版への University College London のパブリックコメント ((PDF) https://downloads.regulations.gov/FDA-2019-D-0934-0026/attachment_1.pdf [最終アクセス日: 2025/5/9]) P6の1個目)

FDA ガイダンスの非線形モデルには、非線形リンク関数を用いた一般化線形モデルが含まれる (FDA ガイダンス II.の脚注5 (4.2.3節))。また、比例ハザード回帰 (proportional hazards regression) モデルは、臨床試験でよく用いられる非線形モデルの一つとして挙げられている (FDA ガイダンス III.C.の箇条書き5個目 (4.2.4.3節))。一方で、モデルが線形結合の形式で表現できない非線形モデルに関する具体的な言及はFDA ガイドラインにはない。

4. FDA ガイダンス本文の解説

本章では、まず、4.1節で FDA ガイダンス最終化までの経緯を説明し、4.2節でガイダンス本文の日本語訳を示し、適宜補足説明を与えた。最後に4.3節では、FDA ガイダンス本文の複数個所で言及されているトピックについて、各箇所での記載内容を集約し整理を試みた。

なお、FDA ガイダンス内では、FDA を指す用語として「FDA」及び「agency」が用いられているが、本報告書では、日本語訳として共に FDA を用いた。また、「should」の日本語訳として、「～すべき」を用いた。また、FDA ガイダンスで引用されている文献は4.2.5節にて示した。6章では本タスクフォースが本報告書で独自に引用している文献を掲載しているのので、文献の掲載箇所に留意いただきたい。

4.1 ドラフトガイダンスと最終化までの経緯

本成果物で解説する FDA ガイダンス「Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products」は、二編のドラフト版（2019年版、2021年版）公表とそれぞれに引き続く二度のパブリックコメント対応を経て、2023年に最終版となったものである。まず、2019年、連続量のみを対象とした「Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biologics with Continuous Outcomes DRAFT GUIDANCE」が公表された。その後、一度目のパブリックコメントを経て、2021年に、連続量以外に関する議論も追加された「Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products DRAFT GUIDANCE」が公表されている。その後、再度のパブリックコメント対応を経て、2023年に最終版として公表された。

各ドラフト版に対して寄せられたパブリックコメントは、FDA の HP にて閲覧が可能である。特に 2021 年版 から 2023 年 最終版 に向けたパブリックコメント（<https://www.regulations.gov/document/FDA-2019-D-0934-0020/comment> [最終アクセス日：2025/5/9]）は、最終版のガイダンスの記載に直接的につながり、ガイドラインの理解を深めるために有益なコメントがあったため、本成果物において適宜参照している。ただし、パブリックコメントの中には、真偽が不明瞭な記載もあるため、参照する際は注意が必要である。

なお、European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) regulatory statistics workshop 2023にて、共変量調整に関するセッションが開催され、FDA の統計家である Daniel Rubin 氏より、最終版にあたっての変更点やガイダンスの推奨事項等が紹介されている（セッション2内の第3演題）。当該 workshop のプレゼンテーションスライドは以下の github に公開されており（リンク [EFSPI regulatory statistics workshop](#) -内の Material, Session 2, Talk 3, Daniel Rubin (FDA), Adjusting for covariates in randomized clinical trials for drugs and biological products が該当）、動画も YouTube で公開されている（リンク [EFSPI regulatory statistics workshop](#) -内の Recordings, Number 2, Session: Synthetic and other baseline covariates: The promise of smaller and faster clinical trials through prognostic digital twins が該当）。英語ではあるものの字幕の表示や、文字起こしの機能で全文の確認も可能であるため、興味のある読者は参照いただきたい。

4.2 ガイダンス本文

以下、箇条書きがどの節の何個目の項目か、を確認しやすくするため、ガイドライン中の箇条

書きのうち、たとえば「III.A.の箇条書き1個目」の中黒（●）を「A1.」のように変更した。また、FDA ガイダンスで引用されている文献は4.2.5節に記載する。一方、本報告書の他の参考文献は、6章に記載する。なお、6章に記載しているもののみ、電子ファイル内でリンクを設定する。

4.2.1 タイトルなど

Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products
Guidance for Industry¹.

【訳】 薬剤及び生物製剤のランダム化比較試験の共変量調整
産業界のためのガイダンス¹

(脚注)

¹ This guidance has been prepared by the Office of Biostatistics in the Center for Drug Evaluation and Research in cooperation with the Center for Biologics Evaluation and Research at the Food and Drug Administration.

【訳】 1. 本ガイダンスは、Center for Drug Evaluation and ResearchのOffice of Biostatisticsが²、Food and Drug AdministrationのCenter for Biologics Evaluation and Researchの協力のもと作成した。

This guidance represents the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA office responsible for this guidance as listed on the title page.

【訳】 本ガイダンスは、本件に関する食品医薬品局（FDA又は政府機関）の現在の考え方を示すものである。これは、いかなる者に対してもいかなる権利も確立せず、FDAや一般市民を法的に拘束するものではない。該当する法令や規制の要件を満たしている場合は、別のアプローチを使用することができる。別のアプローチについて議論するには、タイトルページに記載されている本ガイダンスを担当するFDA部局に連絡されたい。

4.2.2 I. Introduction

This guidance describes FDA's current recommendations regarding adjusting for covariates in the statistical analysis of randomized clinical trials in drug² development programs. This guidance provides recommendations for the use of covariates in the analysis of randomized, parallel group clinical trials that are applicable to both superiority trials and noninferiority trials. The main focus of the guidance is on the use of prognostic baseline covariates³ to improve statistical efficiency for estimating and testing treatment effects. This guidance does not address use of covariates to control for confounding variables in non-randomized trials, the use of covariates in models to account for missing outcome data (National Research

Council 2010), the use of covariate adjustment for analyzing longitudinal repeated measures data, the use of Bayesian methods for covariate adjustment, or the use of machine learning methods for covariate adjustment.

【訳】本ガイダンスには、医薬品²開発におけるランダム化比較試験の統計解析で共変量を調整することに関するFDAの現時点での推奨事項を記載している。本ガイダンスでは、優越性試験及び非劣性試験のいずれにも適用可能な、ランダム化並行群間比較試験における共変量の取扱いに対する推奨事項を与える。本ガイダンスは、治療効果の推定及び検定に対する統計的効率を向上させるためのベースライン予後共変量³のを主な焦点としている。本ガイダンスでは、非ランダム化比較試験における交絡変数の制御のための共変量利用及びアウトカムの欠測データを考慮するためのモデルでの共変量利用（National Research Council 2010）、経時反復測定データの解析に対する共変量調整、共変量調整のためのベイズ流の手法や機械学習の利用については取り扱わない。

In general, FDA’s guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency’s current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word should in Agency guidances means that something is suggested or recommended, but not required.

【訳】一般に、FDAのガイダンス文書は法的強制力のある責任を定めるものではない。代わりに、ガイダンスは特定の規制又は法的要件が引用されていない限り、トピックに関するFDAの現時点での考え方を示しており、推奨事項としてのみ理解すべきである。FDAのガイダンスで“should”^{*}という用語を使用する場合は、何かが提案又は推奨されているが、必須ではないことを意味する。

【補足説明】^{*} 本報告書の日本語訳では、“should”と記載されている箇所は「～すべき」と訳した。

（脚注）

² The term *drug* used in this guidance refers to both human drugs and biological products.

【訳】本ガイダンスで使用する医薬品という用語は、特に指定のない限り、ヒト用医薬品及び生物学的製剤の両方を指す。

³ The term *prognostic baseline covariates* used in this guidance refers to baseline covariates that are likely to be associated with the primary endpoint. Use of *predictive baseline covariates* to identify groups more likely to benefit from treatment is beyond the scope of this guidance.

【訳】本ガイダンスで使用するベースライン予後共変量という用語は、主要評価項目と関連する可能性が高いベースライン共変量を指す。治療が有益となる可能性が高い集団を特定するために用いられるベースライン予測共変量については、本ガイダンスの対象外である

【参照】3.5節

4.2.3 II. Background

Baseline covariates in this guidance refer to demographic factors, disease characteristics, or other information collected from participants before the time of randomization. Covariate adjustment refers to the use of baseline covariate measurements for estimating and testing population-level treatment effects between randomized groups.

【訳】本ガイダンスにおけるベースライン共変量とは、ランダム化の時点より前に治験参加者から収集された人口統計学的要因、疾患特性又はその他の情報を指す。共変量調整とは、ランダム化された群間の集団レベルの治療効果の推定又は検定にベースライン共変量を利用することを指す。

In many randomized controlled trials, the primary analysis used to estimate treatment effects of a new drug might not adjust for baseline covariates (through what is termed an unadjusted analysis). However, incorporating prognostic baseline covariates in the design and analysis of clinical trial data can result in a more efficient use of data to demonstrate and quantify the effects of treatment. Moreover, this can be done with minimal impact on bias or the Type I error rate.

【訳】多くのランダム化比較試験では、新薬の治療効果を推定するための主要な解析において、ベースライン共変量の調整が行われない場合がある（以下、未調整解析）。しかし、ベースライン予後共変量を臨床試験デザインやデータ解析に組み込むことにより、データをより効率的に利用して治療効果の検証及び定量化を行うことができる。さらに、これをバイアス又は第一種の過誤確率に対して最小限の影響で行うことができる。

The ICH guidance for industry *E9 Statistical Principles for Clinical Trials* (September 1998)⁴ addresses these issues briefly. The ICH E9 guidance encourages the identification of “covariates and factors expected to have an important influence on the primary variable(s).” The ICH E9 guidance strongly advises prespecification of “the principal features of the eventual statistical analysis,” including “how to account for [covariates] in the analysis to improve precision and to compensate for any lack of balance between treatment groups.” The ICH E9 guidance also cautions against adjusting for “covariates measured after randomization because they could be affected by the treatments.”

【訳】ICH E9「臨床試験のための統計的原則」（1998年9月）⁴は、これらの問題について簡潔に取り上げている。ICH E9では、「主要変数に重要な影響を及ぼすと予想される共変量及び要因」の特定を推奨している。ICH E9は「最終解析の主要な特徴」を事前規定することを強く推奨しており、これには「精度を向上させ、試験治療グループ間のバランスの欠如を埋め合わせるため、それら（の共変量）を解析でどう取り扱うか」という考慮も含まれる。またICH E9ガイドラインは、「ランダム化後に測定された共変量」は試験治療の影響を受けている可能性があるため、その調整に対して注意を促している*）。

【補足説明】*）ICH E9原文では“not advisable”，日本語訳版では「薦められない」と記載されているが、FDAガイダンス原文では“caution”と記載されているため、ここでは「注意を促している」とした）。

This guidance is consistent with the ICH guidance for industry *E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials*. After specifying the treatment condition of interest, target population, and endpoint variable — the treatment effect estimated by covariate adjustment is a population summary measure defining an estimand.

【訳】本ガイダンスは、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」と整合している。関心のある治療の状況、対象集団、及び変数を特定した後、共変量を調整して推定される治療効果は、estimand を定義する集団レベルの要約となる。

【補足説明】FDA ガイダンスの内容で、estimand の population level summary への記載が求められている、と解釈するのが自然であるが、どの程度詳細に記載すべきは明示されていない。例えば非線形モデルに関する III.C. の箇条書き 3 個目（4.2.4.3 節）に“As part of the prespecification of the estimand of interest, sponsors should specify whether the treatment effect of interest in an analysis is a conditional or unconditional treatment effect.”という記載があるため、非線形モデルを使用する場合、条件付き治療効果と条件なし治療効果の区別は最低限必要であると考えられる。治験実施計画書の estimand の記載については、4.3.2 節を参照されたい。

This guidance provides general considerations and additional recommendations for covariate adjustment using linear and nonlinear models.⁵ In linear models, adjustment for prognostic baseline covariates often leads to improved precision by reducing residual variance. When adjusting for covariates based on fitting nonlinear regression models, such as logistic regression models in studies with binary outcomes, there are additional considerations that arise because inclusion of baseline covariates in a regression model can change the treatment effect that is being estimated. As explained below, after suitably addressing the treatment effect definition, covariate adjustment using linear or nonlinear models can be used to improve statistical efficiency.

【訳】本ガイダンスは、線形及び非線形モデル⁵を用いた共変量調整について、一般的な検討事項及び追加の推奨事項を提供する。線形モデルでは、ベースライン予後共変量を調整することで、残差分散の減少による精度の向上が認められることが多い。非線形回帰モデル（例えば、二値アウトカムの試験におけるロジスティック回帰モデル）に基づいて共変量を調整する場合は、回帰モデルにベースライン共変量を含めることにより推定される治療効果に変化することがあるため、追加の検討事項が生じる。以降で説明するように、治療効果を適切に定義した後、統計的効率の向上のために線形又は非線形モデルによる共変量調整を用いてもよい。

（脚注）

⁴ We update guidances periodically. To make sure you have the most recent version of a guidance, check the FDA guidance web page at <https://www.fda.gov/regulatory-information/search-fda-guidance-documents>.

【訳】FDA の作成するガイダンスは定期的に更新される。ガイダンスが最新版であることを確認するには、FDA のガイダンスの Web ページ (<https://www.fda.gov/regulatory-information/search-fda->

[guidance-documents](#) [最終アクセス日: 2025/5/9]) を参照されたい。

⁵ For this guidance, nonlinear models can include generalized linear models with nonlinear link functions.

【訳】FDAガイダンスの非線形モデルには、非線形なリンク関数を用いた一般化線形モデルが含まれる。

【参照】 3.10節

4.2.4 III. Recommendations for Covariate Adjustment in Clinical Trials

4.2.4.1 A. General Considerations

A1. An unadjusted analysis is acceptable for the primary analysis of an efficacy endpoint.

【訳】有効性評価項目の主要な解析には未調整解析が許容される。

【参照】 3.3節, 3.4節

【補足説明】本内容は、FDA ガイダンスの2021年度ドラフト版では別段落で記載されていたが、最終版では本節の最初の箇条書きに移動した。この変更は、未調整解析も受け入れ可能であることを強調する意図があると想像される。これは、共変量調整解析が未調整解析に比べて統計的効率を悪化させる場合があるためかもしれない (Freedman,2008b)。一方で、FDA は統計的効率の向上を目的として、関心のあるアウトカムと最も強く関連すると期待される共変量の調整を推奨している (FDA ガイダンス III.A.の箇条書き4個目)。要するに、統計的効率を適切に改善することができる方法がある場合は共変量調整解析が推奨されるが、適切な方法がない場合は未調整解析も選択肢になり得ることが示唆されていると考えられる。

A2. Sponsors can adjust for baseline covariates in the analyses of efficacy endpoints in randomized clinical trials. Doing so will generally reduce the variability of estimation of treatment effects and thus lead to narrower confidence intervals and more powerful hypothesis testing.

【訳】治験依頼者はランダム化比較試験の有効性評価項目の解析においてベースライン共変量を調整してもよい。これにより、通常、治療効果の推定のばらつきが減少し、結果として信頼区間が狭まり、仮説検定の検出力が高くなる。

【補足説明】通常、共変量調整を行うことで統計的効率が向上することが述べられている。ただし、モデルを誤特定した場合に、未調整解析よりも統計的効率が悪化する場合がある (Freedman 2008b)。そのため、モデル誤特定を踏まえた統計的性能評価や感度分析が重要と考えられる。

A3. Sponsors should prospectively specify the detailed procedures for executing covariate-adjusted analysis before any unblinding of comparative data. FDA review will emphasize the prespecified primary analysis rather than post-hoc analyses using different models or covariates.

【訳】治験依頼者は比較データの割付けが明らかになる前に共変量を調整する解析の詳細な手順を事前に規定すべきである。FDAの審査は、事前に規定されていないモデルや共

変量を使用した事後的に実施される解析よりも事前に規定された主要な解析を重視する。
【参照】事前規定の必要性は2.5節 表2-1 (ii), ICH E8

A4. Covariate adjustment leads to efficiency gains when the covariates are prognostic for the outcome of interest in the trial. Therefore, FDA recommends that sponsors adjust for covariates that are anticipated to be most strongly associated with the outcome of interest. In some circumstances these covariates may be known from the scientific literature. In other cases, it may be useful to use previous studies (e.g., a Phase 2 trial) to select prognostic covariates or form prognostic indices.

【訳】共変量調整は、その共変量が試験で関心のあるアウトカムの予後共変量である場合、効率向上につながる。従って、FDAは治験依頼者に対し、関心のあるアウトカムと最も強く関連すると期待される共変量について調整することを推奨する。場合によっては、予後共変量は科学的文献から分かっていることもある。他の場合には、予後共変量を選択したり予後の指標を形成したりするために、先行する試験（例えば、第2相試験）を利用することが有用なこともある。

A5. Covariate adjustment can still be performed with covariates that are not prognostic, but there may not be any gain in precision (or may be a loss in precision) compared with an unadjusted analysis.

【訳】予後共変量ではない共変量を用いた共変量調整は可能であるが、未調整解析と比較して精度が向上しない（又は精度が低下する）可能性がある。

【補足説明】予後共変量ではない共変量（アウトカムと相関のない共変量）による調整は、標準誤差を過大評価し、その結果、検出力を低下させる可能性がある。これは調整に用いられた共変量の数に応じてパラメータ数が増え、モデルの自由度が増加し、自由度の大きさに応じて推定に多くの情報量が必要となるため、実質的に治療効果を推定するための情報量が少なくなることでより生じる。特に、被験者数が少ない場合に問題となる。

予後共変量ではない共変量による調整を行った解析と未調整解析での検出力を比較した臨床試験の実データに基づくシミュレーション研究が行われており、一部のシナリオでは、予後共変量ではない共変量による調整を行った解析の方が未調整解析より検出力が低下していることが報告されている（[Kahan et al. 2014](#)）。

A6. Covariate adjustment is acceptable even if baseline covariates are strongly associated with each other (e.g., body weight and body mass index). However, adjusting for less correlated baseline covariates generally provides greater efficiency gains.

【訳】ベースライン共変量が互いに強く関連している場合（例えば、体重とbody mass index）でも、これらの共変量を用いた共変量調整は許容される。しかし、より相関の低いベースライン共変量で調整した方が、一般的に効率が向上する。

A7. Randomization is often stratified by baseline covariates. A covariate adjustment model should

generally include strata variables and can also include covariates not used for stratifying randomization. In some cases, incorrect stratification may occur and result in actual and as-randomized baseline strata variables. A covariate adjustment model can use either strata variable definition as long as this is prespecified.

【訳】ランダム化はベースライン共変量によって層別化されることが多い。共変量調整モデルには一般的に層別変数を含めるべきであり、また、層別ランダム化に使用しなかった共変量を含めてもよい。場合によっては、誤った層別化が起こり、本来割り当てられるべき層と実際に割り当てられた層が異なることがある。共変量調整モデルには、事前に規定されている限り、どちらの層別変数の定義も使用できる。

A8. Sponsors can conduct randomization/permutation tests with covariate adjustment (Rosenbaum 2002).

【訳】治験依頼者は共変量調整を伴うランダム化検定や並べ替え検定を実施してもよい (Rosenbaum 2002)。

【補足説明】Shao らのパブリックコメントを受けて追加された記載であると想像される (表 4-1)。なお、ランダム化検定や並べ替え検定の説明については、手良向 (2020) の総説等を参考にしていきたい。

表 4-1 Shao らのパブリックコメント

原文	日本語訳
... Although randomization tests are certainly quite restrictive in the hypotheses they can test, they are still of great value for RCTs. First of all, randomization tests are exact and thus can be applied for trials with a small sample size. Second, they can be easily applied to complex novel trial designs, by resampling treatment assignments according to the design. Moreover, covariates can also be used in randomization tests to increase their statistical power (Rosenbaum, 2002; Zhao and Ding, 2021). ...	... ランダム化検定は確かに検定できる仮説がかなり限定されているが、それでも RCT には大きな価値がある。第一に、ランダム化検定は厳密であるため、被験者数が少ない試験にも適用できる。第二に、デザインに従って治療割付けをリサンプリングすることで、複雑で新しい試験デザインにも簡単に適用できる。さらに、統計的検出力を高めるために、共変量もランダム化検定に用いることができる (Rosenbaum, 2002; Zhao and Ding, 2021) ...

A9. In a trial that uses covariate adjustment, the sample size and power calculations can be based on adjusted or unadjusted methods. The latter will often lead to a more conservative sample size.

【訳】共変量調整を使用する試験では、被験者数と検出力の計算を調整済みの方法又は未調整の方法のどちらで行ってもよい。未調整の方法に基づいて計算された被験者数は、たいていの場合より保守的になる。

【参照】被験者数を計算する際に共変量調整を考慮するか否かについては4.3.6節

A10. Clinical trials often record a baseline measurement of a defined characteristic and record a later measurement of the characteristic to be used as an outcome. Adjusting for the baseline value rather

than (or in addition to) defining the primary endpoint as a change from baseline is generally acceptable. Sponsors proposing to define the outcome as a percentage change rather than an absolute change from baseline should discuss the outcome definition and use of covariate adjustment with the relevant review division. Sponsors proposing to perform noninferiority testing on ratios of treatment group means rather than differences of treatment group means should also discuss change from baseline outcome definitions and use of covariate adjustment with the relevant review division.

【訳】臨床試験では、定義された特性のベースライン値を記録し、その特性のベースライン後の測定値をアウトカムとして使用することが多い。ベースラインからの変化量を主要評価項目として定義する代わりに（又は定義した上で）、ベースライン値で調整することが一般的に許容される。治験依頼者がベースラインからの変化量ではなく、変化率としてアウトカムを定義することを提案する場合、該当の審査部門とアウトカムの定義及び共変量調整の使用について協議すべきである。同様に、治療群の平均値の差ではなく比に対して非劣性検定を実施することを提案する場合も、ベースラインからの変化に関するアウトカムの定義と共変量調整の使用について該当の審査部門と協議すべきである。

【参照】変化率に関するパブリックコメントについてはEFSPi（（PDF）https://downloads.regulations.gov/FDA-2019-D-0934-0029/attachment_1.pdf [最終アクセス日: 2025/5/9]） P2, Pfizer Inc（（PDF）https://downloads.regulations.gov/FDA-2019-D-0934-0041/attachment_1.pdf [最終アクセス日: 2025/5/9]） P2の1段落目

変化率を用いる際の利点と課題についてはFDAのドラフト版ガイダンス“Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making”（<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-incorporating-clinical-outcome-assessments-endpoints-regulatory> [最終アクセス日: 2025/5/9]） P8, 9

非劣性試験において、治療効果を平均値の差と比で表現する際の違いについてはFDAのガイダンス“Non-Inferiority Clinical Trials”（<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/non-inferiority-clinical-trials> 同上.[最終アクセス日: 2025/5/9]） P23-P26

審査部門との協議については2.5節 表2-1 (vi), 2.5節 表2-2 (ii)

A11. Sponsors should discuss proposals for complex covariate-adaptive randomization, data-adaptive covariate selection, or use of covariate adjustment in an adaptive design with the relevant review division.

【訳】治験依頼者は、複雑な共変量アダプティブランダム化、データアダプティブ共変量選択、又はアダプティブデザインにおける共変量調整の使用に関して提案する場合、該当の審査部門と協議すべきである。

【補足説明】2019年11月にFDA から発出されたアダプティブデザインに関する企業向けガイド

ンス「Adaptive Design Clinical Trials for Drugs and Biologics」では、アダプティブデザインを適用する臨床試験の計画、実施、及び結果の報告に関する重要な原則や FDA と協議する際の考慮事項等がまとめられており、日本製薬工業協会から2021年8月に邦訳版も出ている（JPMA 2021）。このガイダンス内で、共変量に基づくアダプティブランダム化について言及されており、例として最小化法（Pocock and Simon 1975）が挙げられている。共変量に基づくアダプティブランダム化の方法を用いても、適切な方法（一般には、ランダム化に基づく検定や並べ替え検定）で解析すれば、第一種の過誤確率は増大しないと述べられている（詳細は V 章 ADAPTIVE DESIGNS BASED ON COMPARATIVE DATA の E. Adaptations to Patient Allocation 参照）。また、同ガイダンスにおいてアダプティブデザインを用いた試験での結果報告に関する留意事項が述べられており、治療効果の推定値はデザインを適切に考慮するか、あるいは未調整の標本平均値のようなナイーブな推定値を用いる場合にはバイアスの程度を評価し、その解釈に関する適切な注意点を示した上で、推定値を提示するべきと記述されている（詳細は VIII 章 REGULATORY CONSIDERATIONS の C. Evaluating and Reporting a Completed Trial 参照）。

データアダプティブ共変量選択とは、実施中の試験データに基づいて調整に用いる共変量を選択する手法と思われる。なお、先述のガイダンスの中でデータアダプティブ共変量選択については述べられていない。

データアダプティブ共変量選択に関して、EMA ガイドラインの6.1節 General considerations では、検証的試験において主要評価項目と最も強く関連する共変量を後ろ向きに選択することは避けるべきで、共変量の臨床的及び統計的な関連性はその試験のデータセット以外のソースから評価し、正当化されるべきと述べている。

Van Lancker et al. (2024) では、機械学習と共変量選択を組み合わせたデータアダプティブ共変量選択における推論は活発な研究分野となっており、将来的には臨床試験でも活用される手法となるかもしれないと示唆されている。

A12. The statistical properties of covariate adjustment are best understood when the number of covariates adjusted for in the study is small relative to the sample size (Tsiatis et al. 2008). Therefore, sponsors should discuss their proposal with the relevant review division if the number of covariates is large relative to the sample size or if proposing to adjust for a covariate with many levels (e.g., study site in a trial with many sites).

【訳】共変量調整の統計学的特性は、試験で調整に用いる共変量の数が被験者数に対して少ない場合に最もよく理解されている（Tsiatis et al. 2008）。従って、治験依頼者は、共変量の数が被験者数に対して多い場合、又は多くの水準を持つ共変量（例えば、多くの施設を有する試験の場合の試験施設）で調整を提案する場合、その提案を該当の審査部門と協議すべきである。

4.2.4.2 B. Linear Models

B1. Covariate adjustment through a linear model is an acceptable method for estimating the average

treatment effect, which is the difference in expected outcomes between subjects assigned to treatment and control groups. Generally, the outcome is regressed on an intercept, treatment assignment indicator, and baseline covariates, and the model is estimated using ordinary least squares. The resulting estimated regression coefficient for the treatment indicator is the estimate of the average treatment effect.

【訳】線形モデルによる共変量調整は、平均治療効果（すなわち、試験治療群もしくは対照治療群に割り付けられた被験者間での結果の期待値の差）を推定するために許容される方法である。一般的に、結果は切片、治療割付け変数及びベースライン共変量で回帰され、モデルは最小二乗法を用いて推定される。得られた治療割付け変数の回帰係数の推定値は平均治療効果の推定値となる。

【参照】4.3.3.1, 3.3節

B2. The average treatment effect is an example of an unconditional treatment effect, which quantifies the effect at the population level of moving a target population from untreated to treated.

【訳】平均治療効果は、条件なし治療効果の一例であり、対象集団を未治療から治療に移行させた際の集団レベルでの効果を定量化するものである。

B3. Even when the linear regression model is misspecified and does not accurately capture the relationships between the outcome, covariates, and treatment, covariate adjustment through a linear model is a valid method for estimating and performing inference for the average treatment effect (Lin, 2013). However, the power of hypothesis tests and precision of estimates generally improves if the model more closely approximates the true relationships among the outcome, covariates, and treatment.

【訳】線形回帰モデルが誤特定されており、結果変数、共変量、及び治療の関係を正確に捉えていない場合においても、線形モデルによる共変量調整は、平均治療効果を推定し推論を行うための妥当な方法である (Lin, 2013)。ただし、モデルが結果変数、共変量、治療間の真の関係をより正確に近似すると、検出力と推定値の精度は一般的に向上する。

B4. Nominal standard errors are often the default method in most statistical software packages. Even if the model is incorrectly specified, they are acceptable in two arm trials with 1:1 randomization. However, in other settings, these standard errors can be inaccurate when the model is misspecified. Therefore, the Agency recommends that sponsors consider use of a robust standard error method such as the Huber-White “sandwich” standard error when the model does not include treatment by covariate interactions (Rosenblum and van der Laan 2009; Lin 2013). Other robust standard error methods proposed in the literature can also cover cases with interactions (Ye et al. 2022). An appropriate nonparametric bootstrap procedure can also be used (Efron and Tibshirani 1993).

【訳】名目標準誤差は、多くの場合、ほとんどの統計ソフトウェアパッケージのデフォ

ルトの方法である。モデルが誤特定されている場合でも、1:1ランダム化を行う2群試験では許容される。しかしながら、他の状況においては、モデルが誤特定されている場合、これらの標準誤差は不正確になることがある。そのためFDAは、モデルに治療と共変量の交互作用を含めない場合、Huber-White「サンドウィッチ」標準誤差のようなロバスト標準誤差の方法（Rosenblum and van der Laan 2009; Lin 2013）の使用を検討することを治験依頼者に推奨する。文献等で提案されているその他のロバスト標準誤差の推定方法は、交互作用を含む場合も適用可能である（Ye et al. 2022）。適切なノンパラメトリックブートストラップ法を用いることも可能である（Efron and Tibshirani 1993）。

【補足説明】交互作用を考慮する解析手法では、共変量を調整する際に共変量の中心化を前提としている場合がある（Tsiatis et al. 2008; Ye et al. 2023）。その際、妥当な信頼区間を構成するために必要な標準誤差を推定するためには、共変量の中心化に必要な共変量の平均値の推定により生じるばらつきを考慮した標準誤差の推定量が必要となる。Ye et al. (2023) では、共変量の中心化に伴うばらつきを考慮した標準誤差の推定量を提案している。

一方、ノンパラメトリックブートストラップ法により、分散を推定する方法も存在するが、標準的な実装方法によるノンパラメトリックブートストラップ分散推定量は単純ランダム化の下でしか正当化されない点に注意が必要である（Van Lancker et al. 2024）。その他のランダム化の方法のための拡張については、Shao et al. (2010) を参照いただきたい。

B5. An analysis ignoring stratified randomization is likely to overestimate standard errors and can be unduly conservative when performing inference for the average treatment effect. The Agency recommends that the standard error computation account for stratified randomization. There are several methods for computing standard errors when combining stratification with covariate adjustment and possible model misspecification (Bugni et al. 2018; Ye et al. 2021). The statistical properties of such methods are best understood when the number of strata is small relative to the sample size. Sponsors can propose methods to account for stratified randomization in computing standard errors, confidence intervals, and hypothesis testing.

【訳】平均治療効果の推論を行う際に、層別ランダム化を無視した解析は標準誤差を過大評価する可能性が高く、過度に保守的になる可能性がある。FDAは、標準誤差の計算に層別ランダム化を考慮することを推奨する。層別化と共変量調整及びモデル誤特定の可能性を組み合わせたときの標準誤差の計算方法はいくつかある（Bugni et al. 2018; Ye et al. 2021）。このような方法の統計的特性は、被験者数に対して層の数が少ない場合に最もよく理解されている。治験依頼者は、標準誤差、信頼区間、仮説検定の計算において、層別ランダム化を考慮する方法を提案してもよい。

【参照】Bugni et al. (2018), Ye et al. (2022) で紹介されている層別化と共変量調整及びモデル誤特定の可能性を組み合わせたときの標準誤差の計算方法は、4.3.4.3節を参照のこと。

B6. The linear model may include treatment by covariate interaction terms. However, when using this

approach, the primary analysis can still be based on an estimate from the model of the average treatment effect (Tsiatis et al. 2008; Ye et al. 2021). As noted in the ICH E9 guidance, interaction effects may be important to assess in supportive analysis or exploratory analysis. This is because differences in treatment effects across subgroups defined by baseline covariates could be relevant to prescribers, patients, and other stakeholders and can imply that the average treatment effect gives an incomplete summary of efficacy.

【訳】線形モデルには治療と共変量の交互作用を含めてもよい。交互作用を含めたモデルを用いる場合でも、モデルに基づく平均治療効果の推定値に基づいて主要な解析を行うことができる (Tsiatis et al. 2008; Ye et al. 2021)。一方、ICH E9ガイドラインで述べられているように、交互作用効果は補助解析又は探索的解析において評価することが重要である。これは、ベースライン共変量で定義される部分集団間における治療効果の差異が、治療を行う医師や患者、その他のステークホルダーにとって意義のある場合があり、平均治療効果による有効性の要約だけでは不十分となり得るためである。

【補足説明】治療と共変量の交互作用を含めた線形モデルに基づいて平均治療効果の推定を行うためには、モデルを適合する前に、共変量の中心化を行う必要がある点に留意されたい (4.3.3.2節)。詳細は [Yang and Tsiatis \(2001\)](#), [Tsiatis et al. \(2008\)](#), [Ye et al. \(2022; 2023\)](#) を参照。

4.2.4.3 C. Nonlinear Models

C1. Covariate adjustment with nonlinear models is often used in the analysis of clinical trial data when the primary outcome of interest is not measured on a continuous scale or is right censored (e.g., binary outcome, ordinal outcome, count outcome, or time-to-event outcome). Adjustment using nonlinear models is a potentially acceptable method for analyzing these data from a clinical trial. However, there are additional issues described below that should be considered before using nonlinear models.

【訳】非線形モデルを用いた共変量調整は、関心のある主要評価項目が連続尺度で測定されない場合や右側打ち切りを受けている場合 (例えば、二値アウトカム、順序アウトカム、カウントアウトカム、イベント発現までの時間データ) の臨床試験データの解析でしばしば利用される。非線形モデルを利用した共変量調整は、臨床試験で得られるこのようなデータを解析するために許容される可能性がある。しかし、非線形モデルを使用する前に以下に示す追加で検討すべき事項がある。

C2. In general, treatment effects may differ across subgroups. However, with some parameters such as odds ratios, even when all subgroup treatment effects are identical, this subgroup-specific conditional treatment effect can differ from the unconditional treatment effect (i.e., the effect at the population level from moving the target population from untreated to treated) (Gail et al. 1984). This is termed non-collapsibility (Agresti 2002), which is distinct from confounding and can occur despite randomization and large sample sizes. An example of non-collapsibility of the odds ratio

for a hypothetical clinical trial is illustrated in Table 1 below. The unconditional odds ratio in the hypothetical target population is 4.8, which is lower than the conditional odds ratio of 8.0 in each of the biomarker-positive and biomarker-negative subgroups. In trials with time-to-event outcomes, the hazard ratio can also be non-collapsible. Unlike the odds ratio or hazard ratio, the risk difference and relative risk are collapsible.

Table 1: Non-collapsibility of the Odds Ratio in a Hypothetical Target Population

	Percentage of target population	Success rate		Odds ratio
		New drug	Placebo	
Biomarker-positive	50%	80.0%	33.3%	8.0
Biomarker-negative	50%	25.0%	4.0%	8.0
Combined	100%	52.5%	18.7%	4.8

【訳】 一般に、治療効果は部分集団間で異なる場合がある。ただし、オッズ比等のパラメータでは、すべての部分集団の条件付きの治療効果が同一である場合にも、この部分集団特有の条件付き治療効果と条件なし治療効果（すなわち、対象集団が未治療から治療に変化した場合の集団レベルの効果）は異なる場合がある（Gail et al. 1984）。この現象は併合不能性と呼ばれ（Agresti 2002）、交絡とは異なるものであり、ランダム化を実施した場合や被験者数が多い場合にも発生する可能性がある。Table 1は仮想的な臨床試験において、オッズ比の併合不能性の例を示している。仮想的な対象集団における条件なしオッズ比は4.8であり、これはバイオマーカー陽性及び陰性の各部分集団の条件付きオッズ比の8.0よりも小さい。イベント発現までの時間データを評価項目とする臨床試験では、ハザード比も併合不能である場合がある。オッズ比やハザード比とは異なり、リスク差とリスク比は併合可能である。

表1. 仮想的な対象集団におけるオッズ比の併合不能性

	対象集団の割合	成功確率		オッズ比
		新薬	プラセボ	
バイオマーカー陽性	50%	80.0%	33.3%	8.0
バイオマーカー陰性	50%	25.0%	4.0%	8.0
合計	100%	52.5%	18.7%	4.8

【参照】 併合可能性と併合不能性については3.7節

【補足説明】 “non-collapsibility”という用語の引用元とされている Categorical Data Analysis, 2nd edition（Agresti 2002）及び3rd edition（Agresti 2013）を調査したところ、“collapsibility”という用語は使用されている一方で“non-collapsibility”という用語は使用されていなかった。

C3. As part of the prespecification of the estimand of interest, sponsors should specify whether the treatment effect of interest in an analysis is a conditional or unconditional treatment effect.

【訳】 関心のあるestimandの事前規定の一環として、治験依頼者は解析で関心の対象とする治療効果が条件付き又は条件なしのいずれの治療効果であることを明記するべきである。

【補足説明】 特にオッズ比やハザード比等の併合不能な評価指標の場合、共変量で定義された部分集団間で条件付き治療効果が一定であったとしても、条件付き治療効果と条件なし治療効果が異なる場合がある。FDA ガイダンス II. で、共変量を調整して推定される治療効果は、estimandを定義する集団レベルでの要約となる、と述べられているとおり、治療効果として条件付き又は条件なしのいずれの治療効果を対象としているかを治験実施計画書の estimand の集団レベルでの要約の欄に明記することが求められていると思われる。

C4. Cochran-Mantel-Haenszel methods (Mantel and Haenszel 1959) are acceptable for the analysis of clinical trial data with binary endpoints if there is interest in estimating a conditional treatment effect, which is assumed to be constant across subgroups defined by a covariate taking a discrete number of levels (e.g., the value 8.0 in Table 1).

【訳】 条件付き治療効果の推定が関心の対象である場合、二値評価項目の臨床試験データの解析にCochran-Mantel-Haenszel法 (Mantel and Haenszel 1959) を用いることができる。なお、この場合、いくつかの水準を持つ共変量で定義された部分集団間で条件付き治療効果は一定（例えば、Table 1のオッズ比8.0）と仮定される。

【補足説明】 二値アウトカムを評価項目とする臨床試験で層別因子を設定している場合、層別因子を考慮した解析として Cochran-Mantel-Haenszel 法を主要な解析に用いることが多い。ここで、「Cochran-Mantel-Haenszel 法を用いるから、estimand として条件付き治療効果を考える」のではなく、「estimand として条件付き治療効果の推定に関心の対象とするから、解析では Cochran-Mantel-Haenszel 法を用いる」ことが ICH E9(R1) の理念からも極めて重要である点に留意が必要である。また、この方針で Cochran-Mantel-Haenszel 法を用いる場合、得られる効果指標の推定値は一つであるが、その推定値は条件なし治療効果（Table 1のオッズ比4.8）ではなく条件付き治療効果（Table 1のオッズ比8.0）であることにも留意する必要がある。

C5. Fitting a nonlinear regression of the outcome on treatment and baseline covariates similarly attempts to estimate a conditional treatment effect. Nonlinear models extend Cochran-Mantel-Haenszel methods by allowing adjustment for continuous covariates, such as age. In nonlinear regression models (without treatment by covariate interactions), the treatment effect is assumed to be approximately constant across subgroups defined by baseline covariates in the model and can provide more individualized information than the unconditional treatment effect if the assumption holds (and not otherwise). Nonlinear models such as logistic regression or proportional hazards regression (which can include stratification of the baseline hazard) are commonly used in many clinical settings.

【訳】 同様に（条件付き治療効果の推定が関心の対象である場合）、治療及びベースライン共変量を説明変数とした非線形回帰を用いることは、条件付き治療効果を推定しよう

とすることである。非線形モデルは年齢のような連続量の共変量での調整を可能とすることにより、Cochran-Mantel-Haenszel法を拡張している。（治療と共変量の交互作用のない）非線形回帰モデルでは、治療効果はモデルに含めたベースライン共変量で定義される部分集団間ではほぼ一定であると仮定され、仮定が正しければ条件なし治療効果よりも個別化された情報を得ることができる（仮定が正しくない場合はそうではない）。ロジスティック回帰や（ベースラインハザードの層を含むことが可能な）比例ハザード回帰のような非線形モデルは多くの臨床試験等（clinical settings）でよく用いられている。

【補足説明】 イベント発現までの時間アウトカムを評価項目とする臨床試験で層別因子を設定している場合、いくつかの水準を持つ共変量で定義された部分集団間で一定と仮定した条件付き治療効果の推定が関心の対象であれば、層別因子を考慮した解析として層別ログランク検定や層別 Cox モデル（部分集団間での比例ハザード性は不要であり、部分集団ごとにベースラインハザードは異なってもよいが、部分集団間で条件付き治療効果であるハザード比は一定）を用いることができる。

C6. Sponsors should discuss with the relevant review divisions specific proposals in a protocol or statistical analysis plan containing nonlinear regression to estimate conditional treatment effects for the primary analysis. When estimating a conditional treatment effect through nonlinear regression, the model assumptions will generally not be exactly correct, and results can be difficult to interpret if the model is misspecified and treatment effects substantially differ across subgroups. Interpretability increases with the quality of model specification. Sponsors should discuss any planned assessments of model assumptions and implications for analyses with the relevant review division.

【訳】 主要な解析で条件付き治療効果の推定に非線形回帰を用いる場合、治験依頼者は治験実施計画書又は統計解析計画書に当該解析の具体的内容を含め該当の審査部門と協議すべきである。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。解釈可能性はモデル特定の質が高まると、それに合わせて向上する。解析に対するモデル仮定及び影響の評価の計画について、治験依頼者は該当の審査部門と協議すべきである。

【補足説明】 通常用いられることが多い非線形回帰、例えばイベント発現までの時間アウトカムを評価項目とする臨床試験で層別因子を設定している場合に、層別因子を層として考慮した層別 Cox モデルを主要な解析に用いることを考える。この解析はこれまで慣例的に用いられてきたモデルと考えられ、層別因子で定義される部分集団内での比例ハザード性等のモデルの妥当性の議論は依然として発生する可能性はあるものの、多くの場合に FDA や他の規制当局との間で大きな議論を生まないと予想される。それに対し、それよりも強い仮定を必要とする非線形回帰、例えば層別因子を層ではなく説明変数として含めた Cox モデル（層別因子で定義される部分集団間でも比例ハザード性を仮定）を主要な解析に用いる場合、そのモデルを用いる根

拠や仮定の評価方法等を審査部門と協議する必要があると考えられる。

C7. Sponsors can perform covariate-adjusted estimation and inference for an unconditional treatment effect (e.g., the odds ratio of 4.8 in Table 1) in the primary analysis of data from a randomized trial. The method used should provide valid inference under approximately the same minimal statistical assumptions that would be needed for unadjusted estimation in a randomized trial. With nonlinear models using a covariate-adjusted estimator for an unconditional treatment effect, sponsors can use an appropriate bootstrap method or standard error formulas justified in the statistical literature for confidence interval construction. A variety of statistically reliable methods have been proposed in the literature for covariate adjustment with unconditional treatment effects (Colantuoni and Rosenblum 2015).

【訳】 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。非線形モデルを用いて共変量を調整した条件なし治療効果の推定を行う場合、治験依頼者は信頼区間の構成について適切なブートストラップ法又は統計的文献で正当化された標準誤差の計算式を用いてもよい。条件なし治療効果の共変量調整について、様々な統計的に妥当な手法が文献で提案されている (Colantuoni and Rosenblum 2015)。

【補足説明】 3.4節及び4.3.5節でも述べられているように、調整解析もしくは調整済み推定量からは条件付き治療効果が推定されると連想された読者も少なくないだろうが、ここでは共変量調整による条件なし治療効果に焦点を当てていることに留意する必要がある。また、共変量調整による条件なし治療効果の推定手法をランダム化比較試験の主要な解析に用いてよい旨が述べられている。以降でも紹介されている標準化やIPTWを含め、いくつかの手法の概要と性能評価比較は文献を参照のこと (Colantuoni and Rosenblum 2015)。

C8. Covariate-adjusted estimators of unconditional treatment effects that are robust to misspecification of regression models have been proposed for randomized clinical trials with binary outcomes (e.g., Steingrimsson et al. 2017), ordinal outcomes (e.g., Díaz et al. 2016), count outcomes (e.g., Rosenblum and van der Laan 2010), and time-to-event outcomes (e.g., Tangen and Koch 1999; Lu and Tsiatis 2008). If a novel method is proposed and statistical properties are unclear, the specific proposal should be discussed with the review division.

【訳】 二値アウトカム（例えば, Steingrimsson et al. 2017）、順序アウトカム（例えば, Díaz et al. 2016）、カウントアウトカム（例えば, Rosenblum and van der Laan 2010）、及びイベント発現までの時間アウトカム（例えば, Tangen and Koch 1999; Lu and Tsiatis 2008）を用いたランダム化比較試験において、回帰モデルの誤特定に対して安定性のある条件なし治療効果の共変量調整済み推定量が提案されている。新しい方法を提案し、その統計的

性質が不明瞭な場合には、その具体的な提案について審査部門と協議すべきである。

【補足説明】 上記の参考文献に加え、[Wei et al. \(2024\)](#) は、イベント発現までの時間アウトカムに関する条件なし治療効果（ハザード比や restricted mean survival time）を推定する方法を評価している。

C9. As an example, the following are steps for one reliable method for covariate adjustment for unconditional treatment effects with binary outcomes that produces a resulting estimator (Steingrimsdottir et al. 2017; Freedman 2008) termed the “standardized,” “plug-in,” or “g-computation” estimator:

【訳】例えば、以下は二値アウトカムに関する条件なし治療効果の共変量調整のための信頼できる方法の一つの手順であり、「標準化」、「プラグイン」、又は「g-computation」推定量と呼ばれる推定量を生成する（Steingrimsdottir et al. 2017; Freedman 2008）。

(1) Fit a logistic model with maximum likelihood that regresses the outcome on treatment assignments and prespecified baseline covariates. The model should include an intercept term.

【訳】アウトカムに対して、治療割付けと事前に規定したベースライン共変量を説明変数とするロジスティックモデルを最尤法で当てはめる。モデルには切片項を含めるべきである。

(2) For each subject, regardless of treatment group assignment, compute the model-based prediction of the probability of response under treatment using the subject’s specific baseline covariates.

【訳】各被験者について、割り付けられた治療に関係なく、その被験者のベースライン共変量の値を使用して試験治療下でのmodel-basedの予測反応確率を計算する。

(3) Estimate the average response under treatment by averaging (across all subjects in the trial) the probabilities estimated in Step 2.

【訳】ステップ2で推定された確率を試験全体の被験者について平均することで、試験治療下での平均反応確率を推定する。

(4) For each subject, regardless of treatment group assignment, compute the model-based prediction of the probability of response under control using the subject’s specific baseline covariates.

【訳】各被験者について、割り付けられた治療に関係なく、その被験者のベースライン共変量の値を使用して対照治療下でのmodel-basedの予測反応確率を計算する。

(5) Estimate the average response under control by averaging (across all subjects in the trial) the probabilities estimated in Step 4.

【訳】ステップ4で推定された確率を試験全体の被験者について平均することで、対照治療下での平均反応確率を推定する。

(6) The estimates of average responses rates in the two treatment groups from Steps 3 and 5 can be used to estimate an unconditional treatment effect, such as the risk difference, relative risk, or odds ratio.

【訳】ステップ3及び5の治療群ごとの平均反応確率の推定値を使用して、リスク差、相

対リスク，又はオッズ比等の条件なし治療効果を推定する。

【参照】潜在結果変数については3.2.2節

条件なし治療効果については3.2.3節

標準化の詳細な手法をガイダンスで示していることへの懸念に関するパブリックコメントについてはUniversity College London（PDF）https://downloads.regulations.gov/FDA-2019-D-0934-0026/attachment_1.pdf [最終アクセス日: 2025/5/9] P6の3個目

【補足説明】“As an example”と記載されているように、本ガイドラインに示された方法が「例示にすぎない」という点は重要である。治験依頼者は、ガイダンスに記載された方法を深く考慮せずにデフォルトとして採用するのではなく、他の代替手法を検討したうえで適切な解析手法を選択することが推奨される。例えば (1) では、治療割付けと事前に規定したベースライン共変量に基づいてアウトカムを回帰する一つの作業モデルを使用して標準化を行う方法が記載されている (Freedman 2008a)。一方で, Steingrimsson et al. (2017), Colantuoni and Rosenblum (2015), Van Lancker et al. (2024) では、治療割付けごとに作業モデルを適合させる方法を採用している。後者の方法はアウトカムとベースライン共変量の関係をより制限なくモデル化することができる。この点を鑑みて、治験依頼者は代替手法の検討を行ってもよいかもしれない。また、共変量やアウトカムに欠測データが存在する場合には、追加の考慮が必要となる (Van Lancker et al. 2024)。 (2) と (4) では、割り付けられた治療に関係なく、モデルに含まれる治療割付け変数の値を試験治療又は対照治療として、両群合わせた全被験者が、試験治療群又は対照治療群に割り付けられた場合の予測値を算出する。この予測値は、3.2.2節の「ある被験者が任意の治療を仮に受けた場合に観察されるであろう結果変数」と考えることができる。そして、(3) と (5) において、それぞれ (2) と (4) で得られた予測値の平均を求め、(6) において、3.2.3節の「対象集団を未治療から治療に移行させた際の集団レベルでの効果」を推定している。

例示では平均治療効果の点推定値を計算する方法のみが記載されているが、分散の推定については、ブートストラップ法又は標準誤差の計算式において割付け手法を考慮することが推奨される (FDA ガイダンス III.C.の箇条書き7個目と11個目)。

C10. Inverse probability of treatment weighting is another reliable method of covariate adjustment for unconditional treatment effects in randomized trials (Williamson et al. 2013).

【訳】治療を受ける確率の逆数で重みづけする方法 (IPTW法) も、ランダム化比較試験における条件なし治療効果に対する共変量調整の信頼できる手法の一つである (Williamson et al. 2013)。

【補足説明】Williamson et al. (2014) ではランダム化比較試験において、線形回帰モデルやロジスティック回帰モデルを用いて共変量調整を行った場合と、傾向スコアを用いた IPTW 法を用いて群間の共変量の分布を調整した場合の治療効果の推定精度が比較されている。回帰モデルを用いて共変量調整を行う際には被験者数が少ないなどデータの特性によってはモデルが収束しない場合があるが、IPTW 法を用いて共変量を調整した場合にはモデルが収束しない状況はないことも示されている。

C11. An analysis ignoring stratified randomization is likely to overestimate standard errors and can be unduly conservative when performing inferences for an unconditional treatment effect. The Agency recommends that the standard error computation account for stratified randomization. There are several methods for computing standard errors when combining stratification with covariate adjustment and possible model misspecification (e.g., Wang et al. 2021). The statistical properties of such methods are best understood when the number of strata is small relative to the sample size. Sponsors can propose methods to account for stratified randomization in computing standard errors, confidence intervals, and hypothesis testing.

【訳】層別ランダム化を考慮しない解析は、標準誤差を過大推定する可能性があり、条件なし治療効果の推定を行う際に不当に保守的になる可能性がある。FDAは、標準誤差の計算に、層別ランダム化を考慮することを推奨している。共変量調整を伴う層別化とモデル誤特定の可能性を組み合わせたときの標準誤差の計算方法はいくつかある（例えば、Wang et al. 2021）。そのような方法の統計的特性は、被験者数に対して層の数が相対的に少ないときによく知られている。治験依頼者は、標準誤差、信頼区間、仮説検定を計算するときの層別ランダム化を説明する方法を考慮できる。

4.2.5 REFERENCES

- Agresti, A, 2002, *Categorical Data Analysis*, Second Edition, New York (NY): John Wiley & Sons, Inc.
- Bugni, F, IA Canay, and AM Shaikh, 2018, Inference Under Covariate-Adaptive Randomization, *Journal of the American Statistical Association*, 113(524):1784-1796.
- Colantuoni, E and M Rosenblum, 2015, Leveraging Prognostic Baseline Variables to Gain Precision in Randomized Trials, *Statistics in Medicine*, 34(18): 2602–2617.
- Díaz, I, E Colantuoni, and M Rosenblum, 2016, Enhanced Precision in the Analysis of Randomized Trials with Ordinal Outcomes, *Biometrics*, 72(2):422-431.
- Efron, B and RJ Tibshirani, 1993, *An Introduction to the Bootstrap*, Boca Raton (FL): Chapman & Hall.
- Freedman DA, 2008, Randomization Does Not Justify Logistic Regression, *Statistical Science*, 23(2):237-249.
- Gail, MH, S Wieand, and S Piantadosi, 1984, Biased Estimates of Treatment Effect in Randomized Experiments with Nonlinear Regressions and Omitted Covariates, *Biometrika*, 71(3):431-444.
- Lin W, 2013, Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique, *Annals of Applied Statistics*, 7(1):295-318.
- Lu, X and AA Tsiatis, 2008, Improving the Efficiency of the Log-Rank Test Using Auxiliary Covariates, *Biometrika*, 95(3):679-694.
- Mantel, N and W Haenszel, 1959, Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease, *Journal of the National Cancer Institute*, 22(4):719-748.
- National Research Council, 2010, *The Prevention and Treatment of Missing Data in Clinical Trials*, Washington (DC): The National Academies Press.

- Rosenbaum PR, 2002, Covariance Adjustment in Randomized Experiments and Observational Studies. *Statistical Science*, 17(3):286-327.
- Rosenblum, M and MJ van der Laan, 2009, Using Regression Models to Analyze Randomized Trials: Asymptotically Valid Hypothesis Tests Despite Incorrectly Specified Models, *Biometrics*, 65(3):937-945.
- Rosenblum, M and MJ van der Laan, 2010, Simple, Efficient Estimators of Treatment Effects in Randomized Trials Using Generalized Linear Models to Leverage Baseline Variables, *International Journal of Biostatistics*, 6(1):13.
- Steingrimsson, JA, DF Hanley, and M Rosenblum, 2017, Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions, *Contemporary Clinical Trials*, 54:18-24.
- Tangen, CM and GG Koch, 1999, Nonparametric Analysis of Covariance for Hypothesis Testing with Logrank and Wilcoxon Scores and Survival-Rate Estimation in a Randomized Clinical Trial, *Journal of Biopharmaceutical Statistics*, 9(2):307-338.
- Tsiatis, AA, M Davidian, M Zhang, and X Lu, 2008, Covariate Adjustment for Two-Sample Treatment Comparisons in Randomized Trials: A Principled Yet Flexible Approach, *Statistics in Medicine*, 27(23):4658-4677.
- Wang, B, R Susukida, R Mojtabai, M Amin-Esmaeili, and M Rosemblum, 2021. Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Covariate Adjustment, *Journal of the American Statistical Association*, doi: 10.1080/01621459.2021.1981338.
- 【補足 1】 Published online は 2021 年だが、2023 年の Volume 118, Issue 542, Pages1152-1163. に掲載されている
- 【補足 2】 著者名の Rosemblum は FDA ガイダンスの誤記で、正しくは Rosenblum と思われる。
- Williamson, EJ, A Forbes, and IR White, 2013, Variance Reduction in Randomised Trials by Inverse Probability of Treatment Weighting Using the Propensity Score, *Statistics in Medicine*, 33(5):721-737.
- 【補足】 First published は 2013 年だが、Volume 33, Issue 5 は 2014 年に出版されている。
- Ye, T, Y Yi, and J Shao, 2021, Inference on The Average Treatment Effect Under Minimization and Other Covariate-Adaptive Randomization Methods, *Biometrika*, 109(1)33-47.
- 【補足】 First published は 2021 年だが、Volume 109, Issue 1 は 2022 年に出版されている。
- Ye, T, J Shao, Y Yi, and Q Zhao, 2022, Toward better practice Of Covariate Adjustment In Analyzing Randomized Clinical Trials, *Journal of the American Statistical Association*, doi: 10.1080/01621459.2022.2049278.
- 【補足】 Published online は 2022 年だが、2023 年の Volume 118, Issue 544, Pages 2370-2382. に掲載されている

4.3 論点の整理

4.3.1 FDA ガイダンスで審査部門との相談が推奨されている内容の列挙

FDA ガイダンスには、治験依頼者が審査部門と協議すべきと推奨している事項がいくつかあ

る。表 4-2 にそれを列挙する。協議すべき事項がこれに限定されているという文脈ではないことに留意が必要であるが、明記されている以下の事項が解析計画に含まれる場合は審査部門との合意には協議を要することが想定される。

表 4-2 FDA 審査部門との協議が推奨されている事項

本稿	原文	日本語訳	考察: 協議を推奨している理由
4.2.4.1	Sponsors proposing to define the outcome as a percentage change rather than an absolute change from baseline <u>should discuss</u> the outcome definition and use of covariate adjustment <u>with the relevant review division</u> .	治験依頼者がベースラインからの変化量ではなく、変化率としてアウトカムを定義することを提案する場合、該当の審査部門とアウトカムの定義及び共変量調整の使用について協議すべきである。	変化率は非対称性や非正規性により解釈やモデル化が困難になる場合があるためと推察される。変化率を用いる際の利点と課題については FDA のドラフト版ガイダンス “Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making”を参照されたい。
4.2.4.1	Sponsors proposing to perform noninferiority testing on ratios of treatment group means rather than differences of treatment group means <u>should also discuss</u> change from baseline outcome definitions and use of covariate adjustment <u>with the relevant review division</u> .	治療群の平均値の差ではなく比に対して非劣性検定を実施することを提案する場合も、ベースラインからの変化に関するアウトカムの定義と共変量調整の使用について該当の審査部門と協議すべきである。	非劣性試験では非劣性マージンの設定において他試験を参照することになる。対照群が異なる試験間では「平均値の比」の解釈が一貫しない場合があるためと推察される。非劣性試験における治療効果を平均値の差と比で表現する際の違いについては FDA のガイダンス “Non-Inferiority Clinical Trials”を参照されたい。
4.2.4.1	Sponsors <u>should discuss</u> proposals for complex covariate-adaptive randomization, data adaptive covariate selection, or use of covariate adjustment in an adaptive design <u>with the relevant review division</u> .	治験依頼者は、複雑な共変量アダプティブランダム化、データアダプティブ共変量選択、又はアダプティブデザインにおける共変量調整の使用に関して提案する場合、該当の審査部門と協議すべきである。	アダプティブ要素を試験デザインに組み込む場合には、第一種の過誤確率を適切に制御されているかに審査部門は関心があると考えられ、確認が必要な動作特性の協議を求めていると推察される。その他アダプティブデザインにおける原則や考慮事項については FDA ガイダンス “Adaptive Design Clinical Trials for Drugs and Biologics” を参照されたい。
4.2.4.1	sponsors <u>should discuss</u> their proposal with <u>the relevant review division</u> if the number of covariates is large relative to the sample size or if proposing to adjust for a covariate with many levels (e.g., study site in a trial with many sites).	治験依頼者は、共変量の数が被験者数に対して多い場合、又は多くの水準を持つ共変量（例えば、多くの施設を有する試験の場合の試験施設）で調整を提案する場合、その提案を該当の審査部門と協議すべきである。	小標本高次元に相当するような解析の場合は、慣習的に用いられる従来の方法では適切な推定ができない場合があるためと推察される。

本稿	原文	日本語訳	考察: 協議を推奨している理由
4.2.4.3	Sponsors <u>should discuss with the relevant review divisions</u> specific proposals in a protocol or statistical analysis plan containing nonlinear regression to estimate conditional treatment effects for the primary analysis. When estimating a conditional treatment effect through nonlinear regression, the model assumptions will generally not be exactly correct, and results can be difficult to interpret if the model is misspecified and treatment effects substantially differ across subgroups. Interpretability increases with the quality of model specification. Sponsors <u>should discuss</u> any planned assessments of model assumptions and implications for analyses <u>with the relevant review division</u> .	主要な解析で条件付き治療効果の推定に非線形回帰を用いる場合、治験依頼者は治験実施計画書又は統計解析計画書に当該解析の具体的内容を含め該当の審査部門と協議すべきである。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。解釈可能性はモデル特定の質が高まると、それに合わせて向上する。解析に対するモデル仮定及び影響の評価の計画について、治験依頼者は該当の審査部門と協議すべきである。	条件付ける共変量による交互作用がある場合は、非線形モデルの誤特定による結果の解釈が困難になるためと推察される。詳細は4.2.4.3を参照されたい。
4.2.4.3	Covariate-adjusted estimators of unconditional treatment effects that are robust to misspecification of regression models have been proposed for randomized clinical trials with If a novel method is proposed and statistical properties are unclear, the specific proposal <u>should be discussed with the review division</u> .	…を用いたランダム化比較試験において、回帰モデルの誤特定に対して安定性のある条件なし治療効果の共変量調整済み推定量が提案されている。新しい方法を提案し、その統計的性質が不明瞭な場合には、その具体的な提案について審査部門と協議すべきである。	統計的特性が不明確な手法を用いる場合、第一種の過誤確率など統計的な動作特性を慎重に評価する必要があるためと推察される。“novel”と“unclear”の程度についてはFDAガイダンス上では読み取れなかった。

4.3.2 治験実施計画書の estimand の記載

ICH E9(R1) では、主たる目的に対応する primary estimand とその感度分析、及び重要な副次的目的に対応する estimand について治験実施計画書への記載を求めている。Mallinckrodt et al. (2019) で述べられているように、どのような estimand を構成するかがデータ収集方針や試験デザインへの影響が大きいことを踏まえ、primary estimand の詳細及びその解析方法について、治験実施計画書に記載すべきである。2.2節にある通り、estimand は五つの要素（治療、対象集団、変数、その他の中間事象、集団レベルでの要約）で構成され、それぞれ詳細な規定が必要である。FDA ガイダンスの焦点である共変量を調整して推定される治療効果は変数の集団レベルでの要約にあたり（FDA ガイダンス II. (4.2.2節)）、estimand の要素の一つとして治験実施計画書に詳細を規定・記載する必要があると考えられるが、FDA ガイダンスには具体的な記載方法や記載の程度は明確に示されていない。

FDA ガイダンスの各所にあるように、集団レベルでの要約により推定される治療効果は、条件付き治療効果又は条件なし治療効果であり、まずこれらのいずれが関心のある estimand であるかを明記することが必要であると考えられる（3.2.3節, 3.4節）。特に、関心のある主要評価項目が連続変数でない場面等で非線形モデルを利用した共変量調整を検討する場合は、条件付き治療効果と条件なし治療効果が異なる場合があり（3.7節, 4.3.5節）、FDA ガイダンスでもその明記の重要性が明確に示されている（FDA ガイダンス III. C.の箇条書き3個目（4.2.4.3節））。

考慮する具体的な共変量を記載するかどうかについては、条件付き治療効果を estimand とする場合は、未調整解析／調整解析のいずれの場合においても、詳細の記載が必要であると考えられる。これは、条件付き治療効果の estimand が、当該共変量で定義される部分集団での治療効果を推定しようとするものであると考えられるためである（3.2.3節）。一方、条件なし治療効果を estimand とする場合は、治験実施計画書の estimand の項への共変量の記載は必ずしも必要ではないと考えられる。条件なし治療効果を estimand として共変量の調整解析を計画する場合は、治験実施計画書の estimand の項ではなく、解析方法の項等に共変量の具体的な詳細を記載することが適切であると考えられる。

治験実施計画書の estimand の記載の詳細については、製薬協レポート「Estimand の治験実施計画書への実装」も参照されたい（JPMA 2023a）。

4.3.3 線形モデルの回帰係数の推定値と平均治療効果の推定値との関係

本節では、線形モデルの回帰係数の推定値と平均治療効果の推定値の関係を、線形モデルに治療とベースライン共変量の交互作用を含めない場合と含める場合の二つの場合に分けて整理する。

4.3.3.1 線形モデルに治療とベースライン共変量の交互作用を含めない場合

FDA ガイダンス III.B.の箇条書き1個目（4.2.4.2節）で、「線形モデルによる共変量調整は、平均治療効果（すなわち、試験治療群もしくは対照治療群に割り付けられた被験者間での結果の期待値の差）を推定するために許容される方法である。一般的に、結果は切片、治療割付け変数及びベースライン共変量で回帰され、モデルは最小二乗法を用いて推定される。得られた治療

割付け変数の回帰係数の推定値は平均治療効果の推定値となる。」と述べられている。本節では、この記述に関連して、治療とベースライン共変量の交互作用を含めない線形モデルにおける治療割付け変数の回帰係数の推定値と平均治療効果の推定値の関係性を整理する。

3.2.2節で与えた記法と同様に、結果変数を Y 、治療割付け変数を A 、ベースライン共変量を X とする。ここでは簡単のためベースライン共変量は一つとするが、本節の内容は多変量に拡張可能である。このとき、FDA ガイダンス III.B.の箇条書き1個目（4.2.4.2節）で述べられた線形モデルは以下のように表現される。

$$E(Y | A, X) = \beta_0 + \beta_1 A + \beta_2 X$$

この線形モデルは共分散分析（ANCOVA）モデルである。最小二乗法を用いて推定された回帰係数を、それぞれ $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$ とする。

治療とベースライン共変量の交互作用がないという仮定が成り立つ場合、 β_1 は条件付き平均治療効果 $E(Y^1 - Y^0 | X = x)$ と平均治療効果 $E(Y^1 - Y^0)$ の両方の解釈を持つ。これは、上記モデルのような交互作用を含まない線形モデルの場合、これらの平均治療効果は一致するためである。従って、治療とベースライン共変量の交互作用がないという仮定が成り立つ下では、治療割付け変数の回帰係数の最小二乗推定値 $\widehat{\beta}_1$ は、 X の条件付き平均治療効果の推定値とも、平均治療効果の推定値とも解釈できる（Van Lancker et al. 2024）。

真のデータ生成プロセスに治療とベースライン共変量の交互作用がある場合、上記の線形モデルには交互作用を含めていないため、モデル誤特定となる。しかし、任意のモデル誤特定の下でも、上記の線形モデルに含まれる治療割付け変数の回帰係数の最小二乗推定値 $\widehat{\beta}_1$ は平均治療効果の一致推定量となることが証明されている（Yang and Tsiatis 2001）。

従って、真のデータ生成プロセスにおける治療とベースライン共変量の交互作用の有無に関わらず、上記の線形モデル（ANCOVA モデル）の治療割付け変数の回帰係数の推定値は平均治療効果の推定値となる。

4.3.3.2 線形モデルに治療とベースライン共変量の交互作用を含める場合

FDA ガイダンス III.B.の箇条書き6個目（4.2.4.2節）で、「線形モデルには治療と共変量の交互作用を含めてもよい。交互作用を含めたモデルを用いる場合でも、モデルに基づく平均治療効果の推定値に基づいて主要な解析を行うことができる（Tsiatis et al. 2008; Ye et al. 2021）。」と述べられている。本節では、この記述に関連して、治療とベースライン共変量の交互作用を含める線形モデルにおける治療割付け変数の回帰係数の推定値と平均治療効果の推定値の関係性を整理する。

治療とベースライン共変量の交互作用を含めた線形モデルに基づいて平均治療効果を推定するためには、ベースライン共変量で中心化を行う必要がある（Yang and Tsiatis 2001; Tsiatis et al. 2008; Ye et al. 2022; 2023）。中心化を行った線形モデルは以下のように表せる。

$$E\{Y - E(Y) | A, X\} = \beta_1\{A - E(A)\} + \beta_2\{X - E(X)\} + \beta_3\{A - E(A)\}\{X - E(X)\}$$

上記の線形モデルに基づき、最小二乗法を用いて推定された回帰係数を、それぞれ $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$ とする。

上記の線形モデルに含まれる治療割付け変数の回帰係数の最小二乗推定値 $\widehat{\beta}_1$ は、平均治療効

果の一致推定量となることが証明されている (Yang and Tsiatis 2001; Tsiatis et al. 2008) .

つまり、治療とベースライン共変量の交互作用を含む線形モデルにおいても、中心化を行うことで、治療割付け変数の回帰係数の推定値は平均治療効果の推定値となる。

なお、モデル誤特定下での平均治療効果の推定に関する説明は、4.3.4.2節を参照のこと。

4.3.4 線形モデル誤特定に関する整理

4.3.4.1 線形モデル誤特定に関する記述

本稿	原文	日本語訳
4.2.4.2	Even when the linear regression model is misspecified and does not accurately capture the relationships between the outcome, covariates, and treatment, covariate adjustment through a linear model is a valid method for estimating and performing inference for the average treatment effect (Lin, 2013).	線形回帰モデルが誤特定されており、結果変数、共変量、及び治療の関係を正確に捉えていない場合においても、線形モデルによる共変量調整は、平均治療効果を推定し推論を行うための妥当な方法である (Lin, 2013) .
4.2.4.2	Nominal standard errors are often the default method in most statistical software packages. Even if the model is incorrectly specified, they are acceptable in two arm trials with 1:1 randomization. However, in other settings, these standard errors can be inaccurate when the model is misspecified . Therefore, the Agency recommends that sponsors consider use of a robust standard error method such as the Huber-White “sandwich” standard error when the model does not include treatment by covariate interactions (Rosenblum and van der Laan 2009; Lin 2013). Other robust standard error methods proposed in the literature can also cover cases with interactions (Ye et al. 2022).	名目標準誤差は、多くの場合、ほとんどの統計ソフトウェアパッケージのデフォルトの方法である。モデルが誤特定されている場合でも、1:1ランダム化を行う2群試験では許容される。しかしながら、他の状況においては、モデルが誤特定されている場合、これらの標準誤差は不正確になることがある。そのため FDA は、モデルに治療と共変量の交互作用を含めない場合、Huber-White 「サンドウィッチ」 標準誤差のようなロバスト標準誤差の方法 (Rosenblum and van der Laan 2009; Lin 2013) の使用を検討することを治験依頼者に推奨する。文献等で提案されているその他のロバスト標準誤差の推定方法は、交互作用を含む場合も適用可能である (Ye et al. 2022) .
4.2.4.2	The Agency recommends that the standard error computation account for stratified randomization. There are several methods for computing standard errors when combining stratification with covariate adjustment and possible model misspecification (Bugni et al. 2018; Ye et al. 2021).	FDA は、標準誤差の計算に層別ランダム化を考慮することを推奨する。層別化と共変量調整及びモデル誤特定の可能性を組み合わせたときの標準誤差の計算方法はいくつかある (Bugni et al. 2018; Ye et al. 2021) .

4.3.4.2 モデル誤特定下での治療効果の推定量の妥当性について (推定量の漸近正規性)

FDA ガイダンスでは、モデルが誤特定されていても、線形回帰モデルによる共変量調整が平均治療効果を推定するための有効な方法として記載されている (FDA ガイダンス III.B.の箇条書き3個目) . Neyman のランダム化推論モデル (Splawa-Neyman 1990) では、治療効果は被験者ごとに異なることがあり得ると考え、線形性を仮定せず、ランダム化が平均治療効果の推定のばらつきの原因と考える。Lin (2013a) は、平均治療効果を推定するために、Neyman モデルを仮定して、以下の三種類の線形回帰モデルに基づいて平均治療効果の推定量の性質を検討した。

- (a) 治療のみを含む共変量の調整なし: 治療群のみを含めた線形回帰モデル

- (b) 治療と共変量を含む（治療と共変量の交互作用は含まない）：治療と共変量を説明変数とする線形回帰モデル
- (c) 治療、共変量、治療と共変量の交互作用を含む：治療と共変量の主効果、及び治療と共変量の交互作用を説明変数とする線形回帰モデル

上記の三種類の線形回帰モデルから導出される治療効果の推定量は、適当な正則条件の下では、モデルが誤特定されていたとしても、真の平均治療効果の一致推定量であり、漸近正規性を有することが示されている。したがって、これらの推定量に基づき妥当な信頼区間を構築することができる。(a) と (b) の推定精度について、[Freedman \(2008b\)](#) でも議論されているように、(a) と比べて、(b) のほうが推定精度が悪くなる場合がある。(c) は推定の精度を損なうことはなく、(a) と (b) と比べて推定精度が同程度か高くなる。割付比が1:1の場合、(b) と (c) は同程度の推定精度をもつ。

[Ye et al. \(2023\)](#) は、以下の三種類の作業モデルを使用し、単純ランダム化や共変量アダプティブランダム化（層別置換ブロック法、層別バイアスコイン法、最小化法）の下での、それぞれのモデルに基づく統計的効率の向上、一般的なランダム化手法に対する適用可能性、ならびにモデル誤特定下における安定した標準誤差に基づく妥当な推論の可能性を検討した。

- 1) 治療のみを含む共変量未調整の ANOVA モデル
- 2) 治療と共変量を含み、治療と共変量の交互作用は含まない ANCOVA モデル
- 3) 治療、共変量、治療と共変量の交互作用を含む ANHECOVA モデル

適切にデザインされ、実施されたランダム化比較試験では、いずれの作業モデルを使用しても妥当な平均治療効果の推定量が得られることが報告されている。誤特定された作業モデルの下であっても、治療効果の推定量の漸近正規性が保持されるため、妥当な推定を行い、信頼区間を構築することができる。

統計的効率に関して、ANHECOVA モデルは単純ランダム化及び一部の共変量アダプティブランダム化（層別置換ブロック法、層別バイアスコイン法）で、ANOVA モデルや ANCOVA モデルに比べて、効率の向上が保証され、ANHECOVA モデルを実用的な共変量調整方法として推奨している。

[Ye et al. \(2022\)](#) は、2群以上のランダム化比較試験での層別ランダム化の下で、ランダム化に用いた層（層別因子）に加えランダム化に利用しなかった共変量も含めた調整を行う model-assisted な方法（3.9節）を二種類提案している。

- 層別因子ごとに、治療群別に共変量を含む作業モデルを適用して、求めた最小二乗推定量を用いて共変量調整された平均治療効果を推定する方法。
- 層別因子ごとに、共変量を含む作業モデルを適用して、求めた最小二乗推定量を用いて共変量調整された平均治療効果を推定する方法。

これらの方法により推定した平均治療効果の推定量は漸近正規性を有し、model-assisted な方法であるためモデル誤特定下でも漸近的に妥当であることが示されている。また、単純ランダム化、バイアスコインデザイン、Wei の壺デザイン及び層別ブロックランダム化に加えて、これらの方法は最小化法にも適用可能であり、いずれのランダム化手法においても割付比に特段の制約条件はない。ただし、層ごとに作業モデルを推定するため、層内の被験者数が少ない場合に

は推定の安定性が損なわれる可能性がある点に注意が必要である。

4.3.4.3 モデルが誤特定されている場合の標準誤差について

FDA ガイダンスでは、割付比が1:1のランダム化比較試験では、モデルが誤特定されている場合でも名目標準誤差の使用が許容されている（FDA ガイダンス III.B.の箇条書き4個目（4.2.4.2節））。ここでの名目標準誤差とは、線形回帰モデルに対して通常の最小二乗法を適用して推定される回帰係数の標準誤差の推定量を指す。割付比が1:1の場合、モデルが誤特定されていたとしても、この名目標準誤差は真の漸近的標準誤差と同じあるいは保守的になることが知られている（Freedman 2008b）。

治療及び共変量の交互作用を含まない線形回帰モデルに対して、モデル誤特定下での名目標準誤差推定量は実際の漸近標準誤差より過大評価あるいは過小評価する可能性がある。この点に対して、Lin (2013a) は Huber のサンドイッチ推定量を用いたロバスト分散の推定方法を提案し、以下の結果を示している。

- 共変量調整あり（治療と共変量の交互作用なし）のモデル及び治療と共変量の交互作用を含むモデルから構成されるサンドイッチ分散推定量は一致推定量であるか、真の漸近分散よりも漸近的により大きくなる。
- 治療と共変量の交互作用を含むモデルから構成されるサンドイッチ分散推定量は、共変量調整あり（治療と共変量の交互作用なし）のモデルから構成されるサンドイッチ分散推定量と等しいか、より小さくなる。

FDA ガイダンス III.B.の箇条書き4個目（4.2.4.2節）では、治療とベースライン共変量の交互作用を含むモデルにおけるロバストな標準誤差の推定方法として、Ye et al. (2023) の研究が引用されている。Ye et al. (2023) は、model assisted 推定量（3.9節）に対して、モデルの誤特定や不等分散が存在する場合でも、ランダム化の方法に依存せず、正しい漸近分散を与えるロバストな標準誤差推定量を提案している。この推定量は、共変量を中心化したうえで、作業モデルとして ANHECOVA を用いることに基づいており、以下の2つの項から構成される。

1. 群別の残差分散: 各投与群(t)における残差($Y_i - \hat{\beta}_t X_i$)の標本分散 $S_t^2(\hat{\beta}_t)$ を用いて、群ごとの不等分散を反映する。
2. 共変量の標本共分散行列: 共変量の中心化により導入される追加の不確実性を補正するため、全体の共変量に基づく共分散行列 $\hat{\Sigma}_X$ を利用する。

このとき、提案されたロバストな分散推定量は以下のように表される。

$$\hat{V} = \text{diag}\{\pi_t^{-1} S_t^2(\hat{\beta}_t)\} + B^T \hat{\Sigma}_X B$$

ここで、 π_t は割付比、 $\hat{\beta}_t$ は群 t の回帰係数ベクトルの推定量、 B は各群の回帰係数ベクトルの推定量を列として並べた行列である。

一方、従来のロバスト標準誤差（たとえば Huber-White 型）は、共変量の中心化の処理を前提としておらず、推定対象が潜在応答平均である場合には漸近的に不正確となる可能性がある。また、標準的な線形モデルに基づく分散推定量は、不等分散やモデルの誤特定に対して不安定であることが指摘されている（Freedman 2008b; FDA 2021）。実際の応用上では群間差の推定に

興味を持つことは多い。群 s と群 t の群間差の標準誤差推定量は下記のものが提案された

$$SE_{ts} = \sqrt{\pi_t^{-1}S_t^2(\hat{\beta}_t) + \pi_s^{-1}S_s^2(\hat{\beta}_s) + (\hat{\beta}_t - \hat{\beta}_s)\hat{\Sigma}_X(\hat{\beta}_t - \hat{\beta}_s)}$$

FDA 本ガイダンス III.B.の箇条書き5個目(4.2.4.2節)では、層別化と共変量調整およびモデル誤特定の可能性を組み合わせたときの標準誤差の計算方法として、Bugni et al. (2018) の研究が引用されている。Bugni et al. (2018) は、共変量適応型ランダム化、特に層別ランダム化の下での推論において、(i) 二標本 t 検定、(ii) 層固定効果を用いた回帰に基づく t 検定 (t-Test With Strata Fixed Effects) を比較・検討した。

t-Test With Strata Fixed Effects は、各層ごとに固定効果を導入した回帰モデルに基づいており、具体的には、観測されたアウトカムを治療と各層の指標変数を説明変数とする線形回帰モデルを設定する。この手法により、層間の差異を調整し、層内の共変量の影響を排除することで、より正確な推定が可能となる。

Bugni et al. (2018) は、これらの手法がモデル誤特定の影響を受け、特に標準誤差を過大評価して保守的になることがあると理論的に示した。共変量効果と層構造の寄与を明確に分離した上で、漸近分散を層内変動・層間効果・割付不均衡と共変量効果の交互作用の三成分に分解した。また、それぞれの成分に対応する推定量を提示し、モデルが誤特定されていても有意水準を正しく制御できる調整済み標準誤差と検定法を提案した。保守性が生じるのは、層内の共変量効果が非ゼロかつ割付に不均衡がある場合に限られる。

なお、適用する割付け手法は、単純ランダム化、バイアスコインデザイン、Wei の壺デザイン及び層別ブロックランダム化が想定されており、最小化法は対象外である。また、バイアスコインデザイン及び Wei の壺デザインは割付比が1:1である必要がある。その他の詳細な条件については Bugni et al. (2018) を参照いただきたい。

4.3.5 非線形モデルでの条件付き治療効果と条件なし治療効果の記載の比較

FDA ガイダンス III.C. (4.2.4.3節) の非線形モデルでの条件付き治療効果と条件なし治療効果の記載の流れは、大まかに以下の通りとなっている。

- 1) 併合不能性の説明 (箇条書き2個目)
- 2) 関心のある治療効果が条件付き又は条件なしのいずれの治療効果であるかの明記の必要性 (箇条書き3個目)
- 3) 条件付き治療効果の推定で許容される方法、条件付き治療効果の推定に非線形モデルを用いる場合の意味及び留意点 (箇条書き4個目, 5個目, 6個目)
- 4) 条件なし治療効果に共変量を調整した手法を用いることができること、いくつかの手法の紹介及び手法を用いる際の留意点 (箇条書き7個目～)

併合不能な評価指標の場合、共変量で定義された部分集団間で条件付き治療効果が一定であったとしても、条件付き治療効果と条件なし治療効果が異なる場合があるため、estimand としての治療効果が条件付きと条件なしのいずれの治療効果であるかの明記が必要であるという1)→2) の流れは自然である。

次に、3) として条件付き治療効果の推定について記載されている。これは条件付き治療効果を条件なし治療効果より重要視している訳ではなく、共変量を考慮する場合に用いられている標準的な方法（例: Cochran-Mantel-Haenszel 法）は条件付き治療効果を推定していることを強調し、受け入れ可能であることを述べる意図があると考えられる。条件付き治療効果の推定に非線形モデルを用いる場合の意味及び留意点は、このガイダンスに限らない一般的な事項である。

最後に、4) として条件なし治療効果の共変量調整による推定について記載されており、そのための手法をランダム化比較試験の主要な解析で用いてよい旨が述べられている。なお、FDA ガイダンスでは、「新しい方法が提案され、その統計的性質が不明な場合は、その具体的な提案について審査部門と議論するべきである」との記載があるが、そうでない場合でも、主要な解析で条件なし治療効果の共変量調整による推定を行う場合（特に使用実績があまりない状況では）、審査部門との協議が有益となるかもしれない。

表 4-3 に非線形モデルでの条件付き治療効果と条件なし治療効果に関する記載を列挙する。

表 4-3 非線形モデルでの条件付き治療効果と条件なし治療効果の記載

本稿	原文	日本語訳
4.2.4.3	In general, treatment effects may differ across subgroups. However, with some parameters such as odds ratios, even when all subgroup treatment effects are identical, this subgroup-specific conditional treatment effect can differ from the unconditional treatment effect (i.e., the effect at the population level from moving the target population from untreated to treated) (Gail et al. 1984). This is termed non-collapsibility (Agresti 2002), which is distinct from confounding and can occur despite randomization and large sample sizes. An example of non-collapsibility of the odds ratio for a hypothetical clinical trial is illustrated in Table 1 below. The unconditional odds ratio in the hypothetical target population is 4.8, which is lower than the conditional odds ratio of 8.0 in each of the biomarker-positive and biomarker-negative subgroups. In trials with time-to-event outcomes, the hazard ratio can also be non-collapsible. Unlike the odds ratio or hazard ratio, the risk difference and relative risk are collapsible.	一般に、治療効果は部分集団間で異なる場合がある。ただし、オッズ比等のパラメータでは、すべての部分集団の条件付きの治療効果が同一である場合にも、この部分集団特有の条件付き治療効果と条件なし治療効果（すなわち、対象集団が未治療から治療に変化した場合の集団レベルの効果）は異なる場合がある（Gail et al. 1984）。この現象は併合不能性と呼ばれ（Agresti 2002）、交絡とは異なるものであり、ランダム化を実施した場合や被験者数が多い場合にも発生する可能性がある。Table 1 は仮想的な臨床試験において、オッズ比の併合不能性の例を示している。仮想的な対象集団における条件なしオッズ比は4.8であり、これはバイオマーカー陽性及び陰性の各部分集団の条件付きオッズ比の8.0よりも小さい。イベント発現までの時間データを評価項目とする臨床試験では、ハザード比も併合不能である場合がある。オッズ比やハザード比とは異なり、リスク差とリスク比は併合可能である。
4.2.4.3	As part of the prespecification of the estimand of interest, sponsors should specify whether the treatment effect of interest in an analysis is a conditional or unconditional treatment effect.	関心のある estimand の事前規定の一環として、治験依頼者は解析で関心の対象とする治療効果が条件付き又は条件なしのいずれの治療効果であるかを明記するべきである。

本稿	原文	日本語訳
4.2.4.3	Cochran-Mantel-Haenszel methods (Mantel and Haenszel 1959) are acceptable for the analysis of clinical trial data with binary endpoints if there is interest in estimating a <u>conditional</u> treatment effect, which is assumed to be constant across subgroups defined by a covariate taking a discrete number of levels (e.g., the value 8.0 in Table 1).	条件付き治療効果の推定が関心の対象である場合、二値評価項目の臨床試験データの解析に Cochran-Mantel-Haenszel 法 (Mantel and Haenszel 1959) を用いることができる。なお、この場合、いくつかの水準を持つ共変量で定義された部分集団間で条件付き治療効果は一定（例えば、Table 1 のオッズ比 8.0）と仮定される。
4.2.4.3	Fitting a nonlinear regression of the outcome on treatment and baseline covariates similarly attempts to estimate a <u>conditional</u> treatment effect. Nonlinear models extend Cochran-Mantel-Haenszel methods by allowing adjustment for continuous covariates, such as age. In nonlinear regression models (without treatment by covariate interactions), the treatment effect is assumed to be approximately constant across subgroups defined by baseline covariates in the model and can provide more individualized information than the <u>unconditional</u> treatment effect if the assumption holds (and not otherwise).	同様に（条件付き治療効果の推定が関心の対象である場合）、治療及びベースライン共変量を説明変数とした非線形回帰を用いることは、条件付き治療効果を推定しようとするものである。非線形モデルは年齢のような連続量の共変量での調整を可能とすることにより、Cochran-Mantel-Haenszel 法を拡張している。（治療と共変量の交互作用のない）非線形回帰モデルでは、治療効果はモデルに含めたベースライン共変量で定義される部分集団間でほぼ一定であると仮定され、仮定が正しければ条件なし治療効果よりも個別化された情報を得ることができる（仮定が正しくない場合はそうではない）。
4.2.4.3	Sponsors should discuss with the relevant review divisions specific proposals in a protocol or statistical analysis plan containing nonlinear regression to estimate <u>conditional</u> treatment effects for the primary analysis. When estimating a <u>conditional</u> treatment effect through nonlinear regression, the model assumptions will generally not be exactly correct, and results can be difficult to interpret if the model is misspecified and treatment effects substantially differ across subgroups.	主要な解析で条件付き治療効果の推定に非線形回帰を用いる場合、治験依頼者は治験実施計画書又は統計解析計画書に当該解析の具体的内容を含め該当の審査部門と協議すべきである。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。
4.2.4.3	Sponsors can perform covariate-adjusted estimation and inference for an <u>unconditional</u> treatment effect (e.g., the odds ratio of 4.8 in Table 1) in the primary analysis of data from a randomized trial. The method used should provide valid inference under approximately the same minimal statistical assumptions that would be needed for unadjusted estimation in a randomized trial. With nonlinear models using a covariate-adjusted estimator for an <u>unconditional</u> treatment effect, sponsors can use an appropriate bootstrap method or standard error formulas justified in the statistical literature for confidence interval construction. A variety of statistically reliable methods have been proposed in the literature for covariate adjustment with <u>unconditional</u> treatment effects (Colantuoni and Rosenblum 2015).	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。非線形モデルを用いて共変量を調整した条件なし治療効果の推定を行う場合、治験依頼者は信頼区間の構成について適切なブートストラップ法又は統計的文献で正当化された標準誤差の計算式を用いてもよい。条件なし治療効果の共変量調整について、様々な統計的に妥当な手法が文献で提案されている（Colantuoni and Rosenblum 2015）。

本稿	原文	日本語訳
4.2.4.3	Covariate-adjusted estimators of unconditional treatment effects that are robust to misspecification of regression models have been proposed for randomized clinical trials with binary outcomes (e.g., Steingrimsson et al. 2017), ordinal outcomes (e.g., Díaz et al. 2016), count outcomes (e.g., Rosenblum and van der Laan 2010), and time-to-event outcomes (e.g., Tangen and Koch 1999; Lu and Tsiatis 2008).	二値アウトカム（例えば, Steingrimsson et al. 2017）、順序アウトカム（例えば, Díaz et al. 2016）、カウントアウトカム（例えば, Rosenblum and van der Laan 2010）、及びイベント発現までの時間アウトカム（例えば, Tangen and Koch 1999; Lu and Tsiatis 2008）を用いたランダム化比較試験において、回帰モデルの誤特定に対して安定性のある条件なし治療効果の共変量調整済み推定量が提案されている。
4.2.4.3	As an example, the following are steps for one reliable method for covariate adjustment for unconditional treatment effects with binary outcomes that produces a resulting estimator (Steingrimsson et al. 2017; Freedman 2008) termed the “standardized,” “plug-in,” or “g-computation” estimator:	例えば、以下は二値アウトカムに関する条件なし治療効果の共変量調整のための信頼できる方法の一つの手順であり、「標準化」、「プラグイン」、又は「g-computation」推定量と呼ばれる推定量を生成する (Steingrimsson et al. 2017; Freedman 2008)。
4.2.4.3	Inverse probability of treatment weighting is another reliable method of covariate adjustment for unconditional treatment effects in randomized trials (Williamson et al. 2013).	治療を受ける確率の逆数で重みづけする方法 (IPTW 法) も、ランダム化比較試験における条件なし治療効果に対する共変量調整の信頼できる手法の一つである (Williamson et al. 2013)。
4.2.4.3	An analysis ignoring stratified randomization is likely to overestimate standard errors and can be unduly conservative when performing inferences for an unconditional treatment effect.	層別ランダム化を考慮しない解析は、標準誤差を過大推定する可能性があり、条件なし治療効果の推定を行う際に不当に保守的になる可能性がある。

4.3.6 被験者数の減少を狙うか、検出力の増加を狙うか

FDA ガイダンスには、被験者数や検出力に関する記載が複数存在し、その要点を表 4-4 に列挙している。要点の一つは、解析段階で共変量調整を行うことで検出力を増加させられる点である。なお、精度向上を目的にモデルを用いて共変量調整を行う場合、モデル誤特定に対して安定性のある手法を採用することが推奨される (FDA ガイダンス III.C. の箇条書き 8 個目; 4.3.4 節)。

もう一つの要点は、解析段階で共変量調整を行う試験において、被験者数を計算する際に共変量調整を考慮するか否かである。この点について、Van Lancker et al. (2024) は、臨床試験のデザイン段階での共変量調整による精度向上及びそれに伴う被験者数削減の程度には不確実性があることを議論している。そして、この不確実性を考慮した必要被験者数の計算方法として、以下の二つのアプローチを考察している。

1. 保守的なアプローチ

共変量調整が精度向上につながらないと保守的に仮定し、被験者数を計算する。この場合、実際に共変量調整により精度向上が得られれば、検出力が増加する。

2. 精度向上を考慮したアプローチ

過去の知見に基づいて共変量調整による精度向上の度合いを考慮して被験者数を計算する。この方法では、特に精度向上の度合いを過大に見積もった場合に、試験の実際の検出力に影響を与えるリスクがある。このリスクを軽減するために、情報モニタリングを伴う群逐次デザイン (“information-adaptive designs” と呼ばれる) を検討することもできる。

FDA はこれら両方の方法を許容しているが（FDA ガイダンス III.A.の箇条書き9個目）、欧米の規制当局の統計家から、保守的なアプローチを推奨する見解が出されている（表 4-5）。

表 4-4 被験者数と検出力に関する事項

本稿	原文	日本語訳
4.2.4.1 （箇条書き 2 個目）	Sponsors can adjust for baseline covariates in the analyses of efficacy endpoints in randomized clinical trials. Doing so will generally reduce the variability of estimation of treatment effects and thus lead to narrower confidence intervals and more powerful hypothesis testing.	治験依頼者はランダム化比較試験の有効性評価項目の解析においてベースライン共変量を調整してもよい。これにより、通常、治療効果の推定のばらつきが減少し、結果として信頼区間が狭まり、仮説検定の 検出力 が高くなる。
4.2.4.1 （箇条書き 9 個目）	In a trial that uses covariate adjustment, the sample size and power calculations can be based on adjusted or unadjusted methods. The latter will often lead to a more conservative sample size.	共変量調整を使用する試験では、 被験者数 と 検出力 の計算を調整済みの方法又は未調整の方法のどちらで行ってもよい。未調整の方法に基づいて計算された被験者数は、たいていの場合より保守的になる。
4.2.4.2 （箇条書き 3 個目）	Even when the linear regression model is misspecified and does not accurately capture the relationships between the outcome, covariates, and treatment, covariate adjustment through a linear model is a valid method for estimating and performing inference for the average treatment effect (Lin, 2013). However, the power of hypothesis tests and precision of estimates generally improves if the model more closely approximates the true relationships among the outcome, covariates, and treatment.	線形回帰モデルが誤特定されており、結果変数、共変量、及び治療の関係を正確に捉えていない場合においても、線形モデルによる共変量調整は、平均治療効果を推定し推論を行うための妥当な方法である（Lin, 2013）。ただし、モデルが結果変数、共変量、治療間の真の関係をより正確に近似すると、 検出力 と推定値の精度は一般的に向上する。

表 4-5 EFSPi regulatory statistics workshop 2023における統計家のコメント

統計家	コメントの概要
Andreas Brandt (BfArM, EMA)	- 共変量調整は、主に精度や統計的効率の向上を目的とするべきであり、被験者数の減少に重点を置くべきではない。 - 被験者数削減にはデメリットが伴う。例えば、被験者数が減少すると、安全性データの量も減少する。また、臨床試験においては主要な解析だけでなく、リスク・ベネフィットの評価も必要であり、サブグループ解析や副次的解析を実施するためにも、十分なデータが求められる。 - 基本的に、被験者数を極端に減らすことは推奨されない。
Daniel Rubin (FDA)	- FDA としては、被験者数の設定方法や設定根拠について、治験依頼者に一定の裁量を認めている。 - 一方で、検出力が不足している試験が多く存在するのが現状である。そのため、一般的には、被験者数の計算において一定の保守性を持たせることが望ましいと考えられる。 - 未調整解析に基づいて被験者数を決定し、その後適切な調整を行うことで、精度を向上させるというアプローチには利点がある。 - 共変量調整を前提とした被験者数を設定した場合、共変量とアウトカムの相関が想定より低かった場合に検出力が不足する可能性がある。この点については慎重に検討する必要がある。

5. おわりに

本報告書の目的は、今後計画されるランダム化比較試験の解析のベースライン共変量調整方法をより適切に検討するための一助として、ICH E9やEMA ガイドラインの内容を踏まえた上で、FDA ガイダンスの理解を深めることであった。

実装に向けて、大きく分けて二種類の課題がある。

(1) FDA ガイダンスに記載されている個別手法の具体的理解

FDA ガイダンス内では、具体的な統計手法に関しては、参考文献を引用し、ガイダンス内には記載のないものが多い。そのため、各手法を実装しようとする場合、参考文献の内容を理解することが必要となる。しかし、これらの文献は比較的難易度の高い論文が多い。そのため、タスクフォース活動の次の課題として、参考文献に記載されている統計手法を説明することが挙げられる。

(2) ランダム化比較試験の計画上重要だが、FDA ガイダンスの対象外の論点の検討

FDA ガイダンスの Introduction (4.2.2節) に記載されている「対象外」の内容は、おおまかに二種類に大別できる。一つ目はアウトカムデータの欠測や経時測定データの取扱いのように、考慮することが必須となるランダム化比較試験が多く想定されるものである。二つ目は、ベイズ流の手法、機械学習の手法を用いた共変量調整や中間解析時の影響のように、現在実施されている多くの試験では必須とは思われないものの、統計的性能の向上や適用範囲の拡大に対する期待が大きい内容である。二種類とも現在も盛んに研究されている内容であり、これらに対する検討は、さらに進んだ課題である。

6. 参考文献及びガイドラインの理解に役立つ資料

注意事項) 4.2 節の FDA ガイダンス内で引用されている文献は、4.2.5 節に記載した。

- Agresti, A. (2002). *Categorical Data Analysis, 2nd ed.* New York, NY: John Wiley & Sons, Inc.
- Agresti, A. (2013). *Categorical Data Analysis, 3rd ed.* New York, NY: John Wiley & Sons, Inc.
- Aronow, P. M., and Middleton, J. A. (2013). A Class of Unbiased Estimators of the Average Treatment Effect in Randomized Experiments. *Journal of Causal Inference*, 1(1), 135-154.
- Bannick, M. S., Shao, J., Liu, J., Du, Y., Yi, Y., and Ye, T. (2025). A General Form of Covariate Adjustment in Clinical Trials under Covariate-Adaptive Randomization. *Biometrika*, in press. (本論文の内容は 2024 年 5 月 10 日の Journal Club で議論されている : <https://carswg.github.io/journalclub/journalclub3.html>)
- Bartlett, J. W. (2020). Robustness of ANCOVA in randomized trials with unequal randomization. *Biometrics*, 76(3), 1036-1038.
- Bugni, F., Canay, I. A., and Shaikh, A. M. (2018). Inference Under Covariate-Adaptive Randomization. *Journal of the American Statistical Association*, 113(524), 1784-1796.
- Chen, G., Pine, D. S., Brotman, M. A., Smith, A. R., Cox, R. W., Taylor, P. A., and Haller, S. P. (2022). Hyperbolic trade-off: The importance of balancing trial and subject sample sizes in neuroimaging. *NeuroImage*, 247, 118786.
- Colantuoni, E., and Rosenblum, M. (2015). Leveraging prognostic baseline variables to gain precision in randomized trials. *Statistics in medicine*, 34(18), 2602-2617.
- Cox, D.R. (2006). *Principles of Statistical Inference*. Cambridge University Press.
- Daniel, R., Zhang, J., and Farewell, D. (2021). Making apples from oranges: comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical Journal*, 63(3), 528-557.
- Didelez, V., and Stensrud, M. J. (2022). On the logic of collapsibility for causal effect measures. *Biometrical Journal*, 64(2), 235-242.
- Freedman, D. A. (2008a). Randomization Does Not Justify Logistic Regression. *Statistical Science*, 23(2), 237-249.
- Freedman, D. A. (2008b). On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40(2), 180-193.
- FDA (2021). Adjusting for covariates in randomized clinical trials for drugs and biological products. Draft Guidance for Industry. Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research, Food and Drug Administration (FDA), U.S. Department of Health and Human Services. May 2021.
- Greenland, S. (1996). Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology*, 7(5), 498-501.
- Greenland, S., Pearl, J., and Robins J. M. (1999) Confounding and Collapsibility in Causal Inference.

- Statistical Science*, 14(1), 29-46.
- Hernán, M. A., and Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman and Hall/CRC Press.
- Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press. (日本語訳: 星野 崇宏・繁樹 算男(監訳)(2023).統計的因果推論 (上・下) . 朝倉書店.)
- Kahan, B. C., Jairath, V., Doré, C. J., and Morris, T. P. (2014). The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials*, 15, 1-7.
- Lin, W. (2013a). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Annals of Applied Statistics*, 7(1), 295-318.
- Lin, W. (2013b). Essays on Causal Inference in Randomized Experiments. *UC Berkeley*. (学位論文)
- Mallinckrodt, C., Molenberghs, G., Lipkovich, I., and Ratitch, B. (2019). *Estimands, estimators and sensitivity analysis in clinical trials*. Chapman and Hall/CRC.
- Pearl, J. (2009). *Causality*, 2nd ed. Cambridge University Press.
- Pocock, S. J., and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, 103-115.
- Rao, J. N. K. (2005). Interplay between sample survey theory and practice: An appraisal. *Survey Methodology* 31(2), 117-138.
- Rosenblum, M., and van der Laan, M. J. (2009). Using Regression Models to Analyze Randomized Trials: Asymptotically Valid Hypothesis Tests Despite Incorrectly Specified Models. *Biometrics*, 65(3), 937-945.
- Rothman, K. J., and Lash, T.L. (2021). Precision and Study Size. In: Lash, T. L., VanderWeele T. J., Haneuse, S., Rothman, K.J., editors. *Modern epidemiology*, 4th ed. Philadelphia, PA: Wolters Kluwer, 333-364.
- Särndal, C. E. (2010). Models in survey sampling. *Statistics in Transition new series*. 11(3), 112-127.
- Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling*. Springer-Verlag Publishing.
- Shao, J., Yu, X., and Zhong, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika*, 97(2), 347-360.
- Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. P. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, 465-472.
- Steingrimsson, J. A., Hanley, D. F., and Rosenblum, M. (2017). Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions. *Contemporary clinical trials*, 54, 18-24.
- Tsiatis, A. A., Davidian, M., Zhang, M., and X Lu. (2008). Covariate Adjustment for Two-Sample Treatment Comparisons in Randomized Trials: A Principled Yet Flexible Approach. *Statistics in Medicine*, 27(23), 4658-4677.
- Van Lancker, K., Bretz, F., and Dukes, O. (2024). Covariate adjustment in randomized controlled trials: General concepts and practical considerations. *Clinical Trials*, 21(4), 399-411.

- Wang, B., Ogburn, E. L., and Rosenblum, M. (2019). Analysis of covariance in randomized trials: More precision and valid confidence intervals, without model assumptions. *Biometrics*, 75(4), 1391-1400.
- Wang, B., Susukida, R., Mojtakai, R., Amin-Esmaeili, M., and Rosenblum, M. (2023). Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Covariate Adjustment. *Journal of the American Statistical Association*, 118(542), 1152-1163.
- Wei, J., Xu, J., Bornkamp, B., Lin, R., Tian, H., Xi, D., Zhang, X., Zhao, Z., and Roychoudhury, S. (2024). Conditional and Unconditional treatment effects in randomized clinical trials: Estimands, Estimation, and Interpretation. *Statistics in Biopharmaceutical Research*, 16(3), 371-381.
- Williamson, E. J., Forbes, A., and White, I. R. (2014). Variance reduction in randomised trials by inverse probability weighting using the propensity score. *Statistics in Medicine*, 33(5), 721-737.
- Yang, L., and Tsiatis, A. A. (2001). Efficiency study of estimators for a treatment effect in a pretest-posttest trial. *The American Statistician*, 55(4), 314-321.
- Ye, T., Yi, Y., and Shao, J. (2022). Inference on the average treatment effect under minimization and other covariate-adaptive randomization methods. *Biometrika*, 109(1), 33-47.
- Ye, T., Shao, J., Yi, Y., and Zhao, Q. (2023). Toward better practice of covariate adjustment in analyzing randomized clinical trials. *Journal of the American Statistical Association*, 118(544), 2370-2382.
- Zhang, Z., Tang, L., Liu, C., and Berger, V. W. (2019). Conditional estimation and inference to address observed covariate imbalance in randomized clinical trials. *Clinical Trials*, 16(2), 122-31.
- Zhou, H., Murray, T. A., Pan, H., and Yuan, Y. (2018). Comparative review of novel model-assisted designs for phase I clinical trials. *Statistics in Medicine*, 37(14), 2208-2222.
- 医薬統計ポケット資料集 2023 年版, ポケット資料集製作委員会.
- 岩崎学. (2015). 統計的因果推論. 朝倉書店.
- 折笠秀樹. (2022). 押さえておきたい臨床統計の勘所 入門から実践的アプローチまで. ライフサイエンス出版.
- 田中司朗. (2019). 医学のための因果推論の基礎概念. 計量生物学, 40(1), 35-62.
- 手良向聡. (2020). 臨床試験におけるランダム化の意義と限界. 計量生物学, 41(1), 37-54.
- 宮川雅巳. (2004). 統計的因果推論 一回帰分析の新しい枠組み—. 朝倉書店.

EFSPI regulatory statistics workshop 動画及びプレゼンテーション資料

- Rubin, D. (2023). Adjusting for covariates in randomized clinical trials for drugs and biological products. (YouTube 動画) https://youtu.be/WW5P_tMTt08?t=3015 [最終アクセス日: 2025/5/9], (発表スライド: PDF) https://efspieurope.github.io/workshop/data/2023/slides/23_Rubin.pdf [最終アクセス日: 2025/5/9]

本報告書と特に関連の深い, 日本製薬工業協会 医薬品評価委員会報告書

JPMA. (2020). ICH E9(R1) の理解に役立つ因果推論.

https://www.jpma.or.jp/information/evaluation/results/allotment/ds_202207_causal.html [最終アクセス日: 2025/5/9]

JPMA. (2021). アダプティブデザインに関する FDA ガイダンスの邦訳.

https://www.jpma.or.jp/information/evaluation/results/allotment/adaptive_design.html [最終アクセス日: 2025/5/9]

JPMA. (2022). ICH E9(R1) の理解に役立つ因果推論～時間依存性治療編～.

https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202209_causal-tv.html [最終アクセス日: 2025/5/9]

JPMA. (2023a). Estimand の治験実施計画書への実装.

https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202303_protocol.html [最終アクセス日: 2025/5/9]

JPMA. (2023b). 近年のがん第 I 相試験デザインとその選択.

https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202306_oncoP1DE.html [最終アクセス日: 2025/5/9]

執筆者・タスクフォースメンバー

日本製薬工業協会 医薬品評価委員会 データサイエンス部会

2024年度タスクフォース3-1 STAT 新時代

【チームメンバー（50音順）】4.2節（FDA ガイダンス日本語訳）は全員で分担

飯田 涼介	興和株式会社	担当: 3章 (3.1, 3.9, 3.10), 4章 (4.3.6)
大江 基貴	Pfizer R&D 合同会社	担当: 2章, 4章 (4.3.2)
大野 浩太	アッヴィ合同会社	担当: 4章 (4.3.5)
宋 小林	武田薬品工業株式会社	担当: 4章 (4.3.4)
澤本 涼	中外製薬株式会社	担当: 3章 (3.2, 3.3, 3.4, 3.6)
黒見 真央	大塚製薬株式会社	担当: 3章 (3.5, 3.8), 4章 (4.3.3)
平井 俊	第一三共株式会社	担当: 3章 (3.7)
山田 桃香	MSD 株式会社	担当: 3章 (3.7)

【タスクフォースリーダー、推進委員】

土川 克	ゼリア新薬工業株式会社	担当: 3章 (3.1より前), 4章 (4.1まで)
土居 正明	小野薬品工業株式会社	担当: 1章, 5章
尾崎 凌斗	中外製薬株式会社	担当: 4章 (4.3.1)

【担当副部長】

角野 修司	武田薬品工業株式会社	
-------	------------	--