

統計解析に関連したガイドラインの紹介

馬場 亜沙美 (BABA Asami)^{1*}, 鈴木 直子 (SUZUKI Naoko)¹,
田中 瑞穂 (TANAKA Mizuho)¹, 山本 和雄 (YAMAMOTO Kazuo)¹

Key Words : ヒト臨床試験, 特定保健用食品, 機能性表示食品, ICH-E9, ICH E9 (R1), DELTA2, SAP

Introduction to Guidelines Related to Statistical Analysis

Authors: Asami Baba^{1*}, Naoko Suzuki¹, Mizuho Tanaka¹, Kazuo Yamamoto¹

***Correspondence author:** Asami Baba

Keywords: Human clinical trials, Foods for Specified Health Uses (FOSHU), Foods with Function Claims (FFC), ICH-E9, ICH E9 (R1), DELTA2, SAP

Affiliated institution:

¹ ORTHOMEDICO Inc.

2F Sumitomo Fudosan Korakuen Bldg., 1-4-1 Koishikawa, Bunkyo-ku, Tokyo, 112-0002, Japan.

はじめに

臨床試験の実施において、統計解析の適切な計画と実行は、試験結果の科学的妥当性および信頼性を確保する上で極めて重要である。近年、統計解析の透明性と再現性を向上させるため、国際的に様々なガイドラインが策定されている。

本稿では、臨床試験における統計解析に関する主要なガイドラインとして、統計解析計画書 (Statistical Analysis Plans; SAP) の作成指針¹⁾、ICH-E9 ガイドライン²⁾、ICH E9 (R1) 補遺³⁾、DELTA2 ガイドライン⁴⁾ について、その内容と実践的な適用方法を詳述する。これらを適切に理解し、活用することで、

臨床試験における統計解析の品質向上および国際的な標準への準拠が可能となる。

1. 統計解析計画書 (SAP) ガイドライン

1-1. SAP の定義と重要性

SAP は、プロトコルに記載された解析方針をより技術的かつ詳細に記述し、主要アウトカムおよび副次的アウトカムを含むすべてのデータに対する統計解析を実行するための具体的手順を示す文書である²⁾。SAP は解析の再現性と透明性を担保し、試験結果の解釈の一貫性を確保するために不可欠であり、規制当局への説明責任を果たす基盤となる。

* 責任著者：馬場 亜沙美 (Asami Baba)

所属機関：

¹ 株式会社オルトメディコ

〒112-0002 東京都文京区小石川 1-4-1 住友不動産後楽園ビル 2 階

SAPの重要性は、事前に定められた解析手法に従うことで、データ確認後の恣意的な解析選択を防ぎ、p-hackingやHARKing（仮説の事後的変更）などの不適切な研究実践を回避できる点にある。また、解析手順を詳細に記録することで結果の再現性が確保され、独立した研究者による検証が可能となる。

1-2. Gamble C, et al. (2017) ガイドライン

Gambleらによるガイドライン¹⁾は、臨床試験におけるSAP作成の最低要件を体系化した実務的指針であり、SAPが満たすべき記載項目のチェックリストを提示することで、作成者が見落としやすい事項を網羅的に整理している。これにより、SAPの品質向上および試験全体の妥当性確保が可能となる。

本ガイドラインでは、SAPに必須の記載事項として、試験デザインの詳細、解析対象集団の定義、主要・副次アウトカムの統計解析手法、欠測データの取扱い、中間解析および早期中止の計画、多重性の調整方法、感度分析の内容などが明示されている。特に、プロトコルとの整合性を保ちつつ、プロトコルに記載しきれない技術的詳細を補完することを重視している。

1-3. 実務における活用のポイント

SAPを効果的に運用するためには、以下の観点を実践することが重要である。

1-3-1. 作成時期の厳守とタイミング管理

SAPは、データ収集や解析開始前に作成し、最終修正は盲検解除前に確定する。これにより、事後的な解析選択やバイアスを防ぎ、結果の客観性を維持する。理想的には、症例登録開始前にSAPのドラフトを作成し、データベースロック前に最終版を確定することが推奨される。

1-3-2. 記述の詳細度と整合性

プロトコルに記載しきれない統計学的手順、前処理、モデル仕様、感度分析などを明確に記載し、プロトコルに示された主要手法との整合性を示す。プロトコルと異なる解析を行う場合は、その内容と理由を明瞭に説明する。略語は初出時に必ずフルネームを示し、文書全体で一貫した定義を用いる。また、統計ソフトウェアのバージョンや使用するパッケージ

も明記することが望ましい。

1-3-3. 変更管理とトレーサビリティ

すべての改訂履歴を時系列で記録し、変更者、変更日、変更理由および変更の影響を追跡可能とする。重要な修正は関係者の承認を得て記録し、変更が解析結果に与える影響を評価する手順を定める。バージョン管理システムの使用により、変更履歴の透明性を確保することが推奨される。

これらを実践することで、解析手順の再現性が高まり、規制当局の査察に耐え得る説明可能性の高い統計解析が実現される。最終的には、明確かつ一貫したSAPによりデータ解釈の信頼性を確保し、安全性および有効性に関する判断が適正に行われる。

2. ICH-E9 ガイドライン

2-1. ICH-E9 の概要と背景

ICH-E9「臨床試験のための統計的原則」は、1998年に策定された国際的なガイドラインであり、臨床試験における統計学的設計および解析の基本原則を定めている²⁾。本ガイドラインは、日本、米国および欧州の規制当局間で合意された統一基準として、現在でも臨床試験設計の基礎となっている。

2-2. ICH-E9 の主要内容

ICH-E9は、臨床試験における統計学的設計と解析の基本原則を示す国際的指針であり、試験の科学的妥当性および結果の信頼性を確保するための枠組みを提供する。本ガイドラインの中核的メッセージは、解析方針を試験開始前に明確に定めること、解析の再現性および透明性を担保すること、およびバイアス低減を目的とした試験デザインを徹底することである。

具体的には、主要アウトカムや検出力計算、中間解析といった統計的前提を事前に決定することで、事後的な解析選択や誤った解釈の発生を防止することが強調されている。また、ランダム化・盲検化・適切な対照群設定といった設計上の措置は、治療群間の比較可能性および評価の客観性を担保するために不可欠である。

解析においては、解析対象集団の定義（intention to treatやper protocol set等）を明示し、多重性の制御や欠測データの機序に応じた取扱い方針を定め、

感度分析を通じて結論の頑健性を検証することが求められる。これらの要素は単独で機能するのではなく、事前計画、適切な設計、透明な文書化および変更履歴の管理といった一連のプロセスを通じて相互に補強される。

2-3. 実務での活用指針

ICH-E9 を効果的に活用するためには、以下の観点に注意することが重要である。

2-3-1. 主要アウトカムと解析方針の確定

主要・副次アウトカムに対する統計手法、検定方向（片側・両側）、有意水準および共変量の取扱いを事前に明示する。特に主要アウトカムについては、測定方法、評価時点、複合エンドポイントの場合の構成要素などを詳細に定義する必要がある。

2-3-2. 症例数設計の精緻化

検出力（通常 80% 以上）、効果量の仮定（臨床的に意義のある差）および脱落率を含めたサンプルサイズ計算を記載する。効果量の設定根拠として、既存文献や予備試験のデータを明示することが推奨される。

2-3-3. 中間解析と早期中止の計画

中間解析の時期、解析基準、 α 分配法（例：O'Brien-Fleming 法, Pocock 法）を明確にする。また、優越性が立証された場合、有効性が見込みがない場合、または安全性上の問題が生じた場合の早期中止基準についても事前に定める必要がある。

2-3-4. 解析対象集団の定義

ITT (intention to treat), PPS (per protocol set), SAF (safety analysis population) の適用基準と除外ルールを明文化する。特に ITT 解析における欠測データの補完方法についても詳細に記載する。

これらの指針を臨床試験の運用に落とし込むことで、臨床試験の設計と解析における一貫した実務基準を提示し、事前の統計的検討と明確な文書化を通じて科学的妥当性と規制要求への適合性が高まる。これにより、解析結果の解釈が透明かつ再現可能になり、臨床的判断の信頼性が高まる。

3. ICH E9 (R1)

3-1. ICH E9 (R1) の背景と目的

ICH E9 (R1) 補遺「臨床試験における Estimand と感度分析」は日本では 2024 年 6 月に正式通知され³⁾、ICH-E9 の重要な補完文書として位置づけられている。この補遺は、臨床試験で推定すべき治療効果 (estimand) を明確に定義し、試験計画・実施・解析・解釈の一貫性を確保するための体系的枠組みを提供する。従来の ICH-E9 では、ITT 原則が強調されていたが、治療中断や併用療法の使用など、治療開始後に発生する事象 (intercurrent events) への対処方法が明確ではなかった。ICH E9 (R1) は、これらの事象を体系的に扱うための概念的枠組みとして Estimand を導入し、臨床的に意義のある治療効果の定義を支援する。

3-2. Estimand の 5 つの属性

Estimand は以下の 5 つの属性により構成される。

3-2-1. 対象集団 (Population)

臨床的疑問が対象とする患者集団を定義する。これは、試験全体の対象集団、ベースライン特性で定義されたサブグループ、あるいは特定の事象の発生・非発生により定義される主要層 (principal stratum) である場合がある。

3-2-2. 治療条件 (Treatment)

比較する治療条件を明確に定義する。単一の介入、併用療法、あるいは複雑な治療レジメンの場合もある。治療期間、用量、投与方法なども含めて詳細に記述する。

3-2-3. 変数 (Variable)

治療効果を測定するために使用される変数 (エンドポイント) を定義する。測定時点、測定方法、評価尺度なども明確にする。

3-2-4. Intercurrent Events (ICE) への対処方針

治療開始後に発生し、変数の解釈や測定可能性に影響を与える事象 (治療中断、併用療法の使用、死亡など) への対処戦略を定義する。主な戦略として、Treatment Policy (全ての事象を含めて評価)、Hypothetical (仮想的に事象が発生しなかった場合

を想定)、Composite (事象自体を結果の一部とする)、While on Treatment (事象発生前のデータのみ使用)、Principal Stratum (事象の発生・非発生で集団を層別化)がある。

3-2-5. 要約指標 (Population-level Summary)

治療効果を要約する集団レベルの指標を定義する。平均値の差、オッズ比、ハザード比、リスク差など、臨床的に解釈可能な指標を選択する。

3-3. 感度分析の重要性

主解析 (main estimator) は必然的に仮定に依存するため、感度分析により結論の頑健性を検証することが不可欠である。感度分析では、異なる仮定や解析手法を用いて、主解析の結果がどの程度変化するかを評価する。

例えば、欠測データの補完方法を変更したり (LOCF, MMRM, Multiple Imputation など)、異なる ICE 対処戦略を適用したりすることで、結果の安定性を確認する。主解析と感度分析の結果が一貫していれば、結論の信頼性が高まる。

3-4. 実務における活用方法

ICH E9 (R1) の活用には、多職種の協働が不可欠である。臨床医は臨床的に重要な疑問を提起し、統計家はそれを明確な Estimand に翻訳し、規制当局との対話を通じて合意形成を図る。Estimand はプロトコルと SAP の両方に明記し、試験実施中の変更履歴も記録する。

特に重要なのは、Estimand の定義を試験計画の早期段階で行うことである。これにより、必要なデータの収集計画、欠測データの最小化戦略、解析手法の選択が一貫して行われ、試験の科学的妥当性が確保される。

4. DELTA2 ガイドライン

DELTA2 (Difference ELicitation in TriAls) は、臨床的に意義のある (Minimally Important Difference; MID) の設定およびこれを踏まえたランダム化比較試験 (RCT) のサンプルサイズ計算と報告に関する国際的なガイドラインである⁴⁾。2018年に公表され、MID の妥当な決定手法とその試験デザインへの組み込み方を体系化することで、統計的有意性に加えて

臨床的意義を考慮した意思決定を促進することを目的としている。

MID の概念は、統計的に有意な差が必ずしも臨床的に重要とは限らないという認識から発展した。特に大規模試験では、臨床的には無視できるほど小さな差であっても統計的有意差が得られる可能性があり、逆に小規模試験では臨床的に重要な差があっても検出できない場合がある。DELTA2 は、こうした課題に対処するための体系的アプローチを提供する。

4-1. MID 設定のための7つの方法

DELTA2 では、MID の設定に利用可能な7つの方法を体系化している。単一の方法に盲目的に依存するのではなく、複数の方法を組み合わせて証拠を重ねることで、より堅牢な MID 推定が可能となる。これらの方法は、「重要な差」「現実的な差」あるいは「その両方」を推定するために用いられる。

4-1-1. 重要な差を推定する方法

(1) アンカー法 (Anchor based)

外部基準や臨床的評価尺度との関連を基に MID を導出する方法であり、患者報告アウトカム (PRO) やグローバル評価尺度との相関を用いることで、変化の臨床的意味を直接的に反映させやすい。例えば、患者が「少し改善した」と評価する変化量を MID として設定する。この方法の利点は、患者や臨床医の実際の評価と直接結びついている点にある。

(2) 分布法 (Distribution based)

変数の統計的分布特性から推定する方法であり、標準偏差の一定割合 (例えば 0.5 SD) や測定誤差を上回る変化量などが参照される。客観的で再現性が高い一方、臨床的意義との直接的な関連が弱いという限界がある。DELTA2 では、分布法を単独で MID 設定に用いることは推奨されておらず、他の方法との併用が求められる。

(3) 医療経済的手法 (Health economic approach)

費用対効果分析の原理を用いて、医療技術評価の観点から MID を設定する方法である。費用と健康アウトカムを比較し、意思決定者が支払ってもよいと考える健康効果 1 単位あたりの閾値を定義するこ

とで、ゼロの純増分便益を除外するようサンプルサイズを設定する。

(4) 標準化効果量 (Standardised effect size)

標準化された尺度での効果の大きさに基づいて MID を定義する方法である。連続アウトカムでは、Cohen's d (平均差を標準偏差で除したもの) が一般的に用いられ、0.2, 0.5, 0.8 がそれぞれ小, 中, 大の効果とされる。二値アウトカムや生存時間アウトカムでは、オッズ比やハザード比などの指標が用いられる。

4-1-2. 現実的な差を推定する方法

(1) パイロット研究 (Pilot study)

エビデンスや経験が乏しい場合、パイロット研究や予備研究を実施して MID の設定を支援できる。ただし、組入れ基準やアウトカムなど、方法論の違いに留意する必要がある。

4-1-3. 現実的な差を推定する方法

(1) 意見聴取法 (Opinion-seeking method)

医療専門家、患者、その他のステークホルダーから意見を収集して MID を設定する方法である。専門家パネルの形成、専門家団体や患者団体の会員への調査、個別インタビューなどのアプローチがある。この手法は試験の文脈を明示的に考慮できる利点がある。

(2) エビデンスのレビュー (Review of evidence base)

研究課題に関する既存のエビデンスから MID を導出する方法である。理想的には、ランダム化比較試験のシステマティックレビューやメタアナリシスからのエビデンスが用いられる。ランダム化エビデンスが存在しない場合は、観察研究からのエビデンスも同様に活用できる。

DELTA2 では、確実性の高い MID 設定のために、アンカー法と意見聴取法の併用が特に推奨されている。また、複数の方法を用いることで、各方法の限界を補完し、より信頼性の高い MID 推定が可能となる。

4-2. DELTA2 の主要な推奨事項

DELTA2 では、MID 設定とサンプルサイズ計算

に関する以下の推奨事項が示されている。

- 関連文献の検索から開始し、候補となる主要アウトカムや比較対象、対象集団における重要または現実的な差に関する情報を収集する
- 複数の候補アウトカムを順番に検討し、対応するサンプルサイズを探索する。主要アウトカムと目標差の選択は、最小サンプルサイズのみに基づくべきではなく、ステークホルダーの見解および試験実施の実行可能性の両方を考慮する
- 死亡や重篤な有害事象を除き、特定の大きさの差の重要性は自明とみなすべきではなく、ステークホルダーグループにとっての重要性を示す追加的な正当化が必要である
- 確定試験の目標差は、少なくとも1つの主要なステークホルダーグループ (患者, 医療専門家, 規制当局, 医療費支払者) にとって重要と考えられるべきである
- 目標差は、より大きな差が現実的である場合や診療変更に必要な場合、必ずしも最小重要差である必要はない
- 重要な差を知るための追加研究が必要な場合、アンカー法と意見聴取法が推奨される。分布法は単独で使用すべきではない。標準化効果量アプローチのみに基づく MID 設定は最後の手段とすべきであるが、二次的アプローチとして有用である
- 現実的な差を知るための追加研究が必要な場合、意見聴取法とエビデンスレビュー法が推奨される。パイロット試験は通常小規模すぎて現実的な差を示すことはできず、主に試験デザインおよび実施のほかの側面に対処する
- 既存研究を用いて、サンプルサイズ計算の一部である主要なパラメータ (連続アウトカムの標準偏差, 二値アウトカムの対照群割合など) の値を設定する
- 主要な入力値 (target difference や二値アウトカムの対照群割合など) に関する不確実性の影響を考慮した感度分析を実施すべきである
- サンプルサイズ計算の仕様は、主要な試験文書 (研究助成申請書, プロトコル, 結果論文) において推奨される報告項目に従って報告すべきである

これらの推奨事項に従うことで、試験のサンプル

サイズ計算の科学的妥当性と透明性が向上し、臨床的に意義のある研究結果の創出が促進される。

まとめ

本稿では、臨床試験における主要な統計解析ガイドラインについて概観し、それぞれの特徴と実務での活用方法を検討した。これらのガイドラインを適

切に理解し活用することで、科学的に妥当で規制要件を満たす高品質な臨床試験の実施が可能となり、最終的には患者への有効かつ安全な治療選択肢の迅速な提供に貢献できると考える。今後も、新たなガイドラインの開発や既存ガイドラインの改訂が続くと予想されるため、継続的な学習と実践への適用が重要である。

参考文献

1. Gamble C, Krishan A, Stocken D, Lewis S, Juszcak E, Doré C, Williamson PR, Altman DG, Montgomery A, Lim P, Berlin J, Senn S, Day S, Barbachano Y, Loder E: Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *JAMA*. 2017; **318** (23): 2337-43. (PMID: 29260229)
2. 厚生労働省: ICH E9 臨床試験のための統計的原則 (平成 10 年 11 月 30 日付医薬審第 1047 号) (2026 年 2 月 5 日アクセス可能: <https://www.pmda.go.jp/files/000156112.pdf>)
3. 厚生労働省: 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 (令和 6 年 6 月 20 日付医薬審第 0620 第 1 号) (2026 年 2 月 5 日アクセス可能: <https://www.pmda.go.jp/files/000269154.pdf>)
4. Cook JA, Julious SA, Sones W, Hampson LV, Hewitt C, Berlin JA, Ashby D, Emsley R, Fergusson DA, Walters SJ, Wilson ECF, MacLennan G, Stallard N, Rothwell JC, Bland M, Brown L, Ramsay CR, Cook A, Armstrong D, Altman D, Vale LD: DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *Trials*. 2018; **19** (1): 606. (PMID: 30400926)