

医薬品臨床試験におけるサンプルサイズの推 定に関するガイドライン(試行)

2024年12月

目次

一、概要	3
二、サンプルサイズ推定のための主な考慮事項	3
(一) 試験デザイン	4
(二) 有意水準及び検出力	6
(三) 統計解析手法	7
1. データ分布の仮定	8
2. 統計解析モデルの仮定	8
(四) 関心のある治療効果とばらつき	9
三、サンプルサイズの調整	10
(一) サンプルサイズ調整の原則	10
(二) サンプルサイズ調整のシナリオ	13
1. 外部データに基づく調整	13
2. 内部データに基づく調整	13
四、その他	16
(一) 試験実施中の考慮事項	16
(二) ベイジアン法に基づくサンプルサイズの推定	16
(三) その他の目的のためのサンプルサイズ推定	17
(四) サンプルサイズの感度計算	17
(五) サンプルサイズ再推定とその他の適応的デザインの組み合わせ	17
(六) 規制当局との意思疎通	18
五、参考文献	19
附属書1 用語説明	21
附属書2 日本語と英語の用語一覧	23

医薬品臨床試験におけるサンプルサイズの推定に関するガイドライン(試行)

一、概要

サンプルサイズの推定は、サンプルサイズの決定としても知られ、医薬品臨床試験デザインの重要な部分であり、研究の妥当性、正確性、信頼性、完全性、科学性を確保するための重要な手段である。一般的に、臨床試験のサンプルは十分に代表的であり、含まれるサンプルサイズは、研究仮説が意図する対象集団の臨床的疑問に確実に答えるのに十分な大きさでなければならない。

医薬品臨床試験において、サンプルサイズの推定に関連するパラメータを無根拠または不十分な根拠で設定した場合、サンプルサイズの推定に不確実性をもたらし、試験の失敗リスクを高めることになる。さらに、試験における不適切なサンプルサイズの調整又は不適切な操作は、ファミリーワイズエラー率(FWER)のインフレ、試験の完全性の低下、バイアスの導入といった問題を引き起こす可能性がある。

本ガイドラインでは、治験依頼者が科学的に健全なサンプルサイズ推定を行うための指針として、統計的仮定に基づくサンプルサイズ推定におけるパラメータ設定やサンプルサイズの調整といった基本的な検討事項に焦点を当てる。本ガイドラインは、登録目的で実施される検証的臨床試験に適用される。

二、サンプルサイズ推定のための主な考慮事項

臨床試験で答えるべき科学的な疑問に対して、統計学は主に推定対象に対応する研究仮説に基づいて策定された統計的仮説に基づき、仮説検定を通じて推論を行う。サンプルサイズの推定は、一般的に主要な推定対象に基づいて行われ、原則的には、十分な検出力を有しながら、全体的な検定でFWERを確実にコントロールする必要がある。場合によっては、主要な二次推定対象のサン

プルサイズの推定が必要となることがあり、その場合にもFWERのコントロールを考慮する必要がある。仮説検定の方法によって、サンプルサイズの推定方法は異なる。正しいサンプルサイズの推定は、推定対象に合致した試験デザインと、適切かつ合理的な統計解析手法に基づくべきである。サンプルサイズの算出方法は、算出に使用した関連パラメータとその根拠と同様に、プロトコールに記述されるべきである。

一般に、サンプルサイズの推定には、①試験デザイン、②有意水準と検出力、③統計解析手法、④関心のある治療効果とばらつきなど、考慮すべき多くの要素がある。

(一) 試験デザイン

試験デザインは、サンプルサイズの推定において考慮すべき重要な因子であり、通常、比較の種類、デザインの種類、変数の種類、中間事象とその治療ストラテジー、多重度調整ストラテジー、群間の割り付けの割合、無作為化方法などが含まれるが、これらに限定されない。

優越性デザインの場合、試験群の効果推定値は対照群よりも優れている必要があり、優越性を達成するための判断基準は臨床ベネフィットに基づいて決定されるべきであり、統計的仮定はそれに応じて明確にされるべきである。非劣性試験及び同等性試験において、陽性対照薬の有効性とそれに対応する非劣性カットオフ値及び同等性カットオフ値は、サンプルサイズ推定における重要なパラメータであり、非劣性カットオフ値及び同等性カットオフ値を決定する際には、使用した過去の研究と現在の研究の推定対象の違いを考慮する必要がある、非劣性カットオフ値の決定方法については、「医薬品臨床試験の非劣性設計に関するガイドライン」を参照することが推奨される。

単群デザインで目標値コントロールを使用する場合、目標値は専門分野で高いコンセンサスまたは認知度を持つ基準に基づいて設定する必要がある。ヒストリカルコントロールを使用する場合、または目標値が過去のデータに基づいて決定される場合、信頼性の高い過去のデータを選択する必要がある。

時間-イベント型の変数の場合、検定力を決定する最終的な要因は、追跡調査中に観察されたアウトカムイベントの数であるため、アウトカムイベントの発生率、登録のスピード、登録期間、追跡調査期間、脱落率などの要因も最終的なサンプルサイズにさらに影響する。

サンプルサイズの推定は、推定対象における中間事象とその治療戦略が治療効果とばらつきパラメータに与える影響を考慮する必要がある。治療（処理）の一部として中間事象を含む治療戦略がサンプルサイズ推定に与える影響は、主に治療（処理）における中間事象が治療効果とばらつきに与える影響に反映される。治療下戦略と複合変数戦略は、アウトカム変数の定義、したがって治療効果とばらつきに直接影響する。仮想戦略は、通常、ある仮定を伴って、中間事象が発生しないシナリオを想定したものであり、対応する仮定は治療効果とばらつきに直接影響を与える可能性がある。主要層戦略は、推定対象集団の属性の定義に反映され、この戦略の下での主要層の同定方法、主要層の割合、対応する集団の関心のある治療効果とばらつきは、最終的なサンプルサイズの推定に影響する。

多重性の問題がある場合、有意水準の調整、検出力の調整、統計解析手法の調整が必要となる場合があり、それらはサンプルサイズの計算に考慮されるべきである。例えば、中間解析が設定された場合、サンプルサイズの推定においてFWERのコントロールを考慮する必要がある。検証的亜群のサンプルサイズの推定は、多重度戦略と合わせて考慮すべきである。

群間のサンプルサイズ分布の割合は、サンプルサイズの推定において考慮すべきパラメータであり、医薬品の臨床試験では、各群が同じサンプルサイズを持つバランスデザインがしばしば使用される。アンバランスデザインは、倫理的配慮またはその他の正当な理由により、ある群または一部の群に割り当てられるサンプルサイズの割合を減らす必要がある場合に使用される。各群に割り当てるサンプルサイズの割合は、最終的なサンプルサイズの推定値に直接影響するため、研究プロトコールに明記しなければならない。層別無作為化を使用する場合、各層の割合と対象集団との整合性、層別化要因に起因する一つまたはいくつかの層の被験者の症例数がまばらであることなどの問題を考慮する必要があるかもしれない。

(二) 有意水準及び検出力

有意水準と検出力は、サンプルサイズの推定において考慮される基本的なパラメータであり、プロトコールにおいて指定されなければならない。

適切な有意水準を設定することで、FWERを制御するという目的を達成することができる。検証的試験では通常、FWERを片側で0.025、両側で0.05以内に抑えることが求められる。多重性の問題が関与する場合、名目水準の設定は、「医薬品臨床試験における多重性問題に関するガイドライン(試行)」を参照することができる。

検出力については、通常80%以上に設定され、多重性の問題が関与する場合は、検出力への影響を考慮する必要がある。要因計画では、研究の目的に交互作用の検定が含まれている場合、検定の主効果に基づいてサンプルサイズを計算すると、交互作用の検出力が不十分な場合がある。多施設共同試験の場合、サンプルサイズと検出力の計算は、通常、各施設における群間の治療差が同じであり、偏りのない推定値であるという仮定に基づいて行われるため、共通の研究プロトコールを作成し、実施することが重要であり、試験の実施プロセスも可能

な限り標準化すべきである。

(三) 統計解析手法

適切かつ合理的な統計解析手法は、科学的なサンプルサイズ推定の基礎であり前提条件であり、サンプルサイズ推定の前に、選択した統計解析手法が研究デザインに合致していることを確認しなければならない。例えば、統計解析手法は、研究デザインのタイプ、比較のタイプ、無作為化の方法、変数のタイプなどに適したものでなければならない。サンプルサイズの推定の根拠となる統計解析手法は、プロトコールに明確に記載され、原則として主要な解析手法と一致していなければならない。そうでない場合は、依拠した手法が主要な推定対象に必要なサンプルサイズを満たし、サンプルサイズの過小評価につながらないと信じる合理的な根拠が必要である。サンプルサイズを推定するための具体的な計算方法とツールは、プロトコールに明記する必要がある。統計シミュレーションを使用してサンプルサイズを推定する場合は、シミュレーションパラメータの設定、シミュレーション方法、シード値、シミュレーションコードを関連文書に詳細に記載し、規制当局との意思疎通時に提出する必要がある。

それぞれの統計解析手法には、対応する仮定があることに注意する必要がある。統計解析手法を選択する際には、選択した手法がその仮定に違反するリスクと、仮定に違反した場合のサンプルサイズへの推定に与える影響を十分に評価する必要がある。統計解析手法における一般的な仮定には、データの分布に関する仮定と統計解析モデルに関する仮定がある。

1. データ分布の仮定

統計解析では、多くの場合、データの分布に関する仮定が必要になる。例えば、連続変数のデータは正規分布に従うという仮定や、時間-イベント変数のデータは指数分布に従うという仮定などである。データ分布の仮定からの乖離の可能性を評価する必要がある。一方では、乖離のリスクが高い場合には、分布の仮定に対してより頑健な解析手法や分布の仮定に依存しない解析手法を用いるべきであり、他方では、分布の仮定からの乖離による治療効果の過大評価やばらつきの過小評価の可能性をパラメータ設定において考慮すべきである。ノンパラメトリック手法やセミパラメトリック手法を解析に使用する場合、サンプルサイズの推定は、計算の容易さの考慮に基づいて特定のパラメトリック仮定に依存することがあり、例えば、生存解析のログランク検定では、サンプルサイズの推定は、生存データが指数分布に従うと仮定することがあり、それを適用する際には、仮定から逸脱することに関連するリスクを考慮する必要がある。

データに相関があり、治療効果やばらつきの推定に影響を与える場合、相関を無視することは、サンプルサイズの推定や解析に影響を与えるだけでなく、FWERのインフレのリスクをもたらす可能性がある。したがって、サンプルサイズを推定する際には、データ間の相関関係を評価し、そのサイズと根拠をプロトコールに明確に記述する必要がある、統計解析においても相関関係を考慮する必要がある。

2. 統計解析モデルの仮定

統計解析モデルは通常、一連の仮定に基づいており、モデルの適用可能性に注意を払う必要がある。モデルの仮定が成立するかどうか、事前判断と事後検証を行う必要がある。統計解析モデルの仮定が成立しないリスクがある場合は、サンプルサイズを算出する際にそのリスクを考慮することが推奨される。さらに、共変量は治療効果やばらつきの推定に影響を与えるため、サンプルサイズの推定

では共変量を調整するかどうかを考慮する必要がある。共変量調整については、「医薬品臨床試験における共変量補正に関するガイドライン」を参照することができる。さらに、層別無作為化では、層別化因子の補正や層別化解析の使用は、例えば、ある層でのサンプルサイズが小さいなどの事情がある場合、試験の有効性を低下させるリスクがあり、これはデザインにおいて考慮されるべきである。

(四) 関心のある治療効果とばらつき

与えられた統計解析手法に基づいてサンプルサイズを推定する場合、帰無仮説と対立仮説がそれぞれ成立するという条件の下で、各パラメータを合理的に設定する必要がある、主に関心のある治療効果とばらつきに影響する関連パラメータの設定が含まれる。

パラメータ設定は十分な根拠が必要であり、一般的には臨床的意義と妥当性に留意した過去のデータ(過去の研究結果や公表データなど)に基づくべきである。サンプルサイズを小さくする目的で、過激なパラメータを設定することは推奨されない。パラメータ設定に根拠がない場合(過去のデータがない)、あるいはパラメータ設定に根拠が十分でない場合(過去のデータが少なすぎる)には、まず必要なパラメータを得るための探索的試験を行うことが推奨される。

参照した過去のデータについては、今回提案されている研究の推定対象と、過去のデータに関連する研究の推定対象との間の差異を十分に評価すべきである。提供されたパラメータは、過去のデータに関する関連研究が、今回の研究と同じか類似の推定対象を持っている場合、より信頼性が高くなる。差異が許容範囲内である場合は、合理的な仮定に基づいてパラメータを設定することが推奨され、差異が大きい場合は、さらなる探索的試験を実施することが推奨される。母集団、治療(処理)、変数(エンドポイント)の定義は、中間事象の治療戦略に関係するため、中間事象の治療戦略が関心のある治療効果とばらつきの推定に与える影響を考慮する必要がある。パラメータを設定する際に、想定

されるパラメータまたは参照される過去のデータが中間事象の影響を反映していない場合は、サンプルサイズを過小評価するリスクを最小限に抑えるために、中間事象の発生率とその治療ストラテジー、および起こり得る有効性の損失を評価することが推奨される。さらに、今回提案された研究の統計解析手法と、過去のデータに関連する研究の統計解析手法に違いがある場合、それらがパラメータ推定に与える影響も評価する必要がある。

サンプルサイズの推定は、欠損データの影響も考慮しなければならない。欠損データは、治療効果とばらつきのパラメトリック仮定で考慮することができ、また、例えば、算出されたサンプルサイズを一定の割合で増加させることによって対処することもできる。

三、サンプルサイズの調整

サンプルサイズの調整とは、臨床試験中に初期デザインのサンプルサイズに対して行われる調整である。サンプルサイズの調整は、一般的にサンプルサイズを小さくすることは推奨されない。サンプルサイズの調整の必要性和妥当性、盲検化と試験の完全性を維持するために取られた措置は、プロトコールに記述されるべきである。

(一) サンプルサイズ調整の原則

合理的なサンプルサイズの調整のみが、試験の効率を向上させるという目標を達成できる。サンプルサイズを調整する必要性、妥当性、実現可能性は十分に評価されるべきであり、試験の完全性は確保されるべきであり、許容できる最大サンプルサイズは通常、実現可能性と臨床における最小重要差または最大の許容できる分散に基づいて設定される。

(1) 必要性

サンプルサイズを正当な理由なく恣意的に調整すべきではなく、サンプルサイズ調整の必要性を十分に評価する必要がある。過去のデータが少なすぎてパラメータ設定を正当化できない場合、探索的試験を実施することによって必要なデータを得ることができる。過去のデータが十分に信頼できる場合、サンプルサイズ再推定の設計と比較して、固定サンプルサイズ設計は検出力を確保しながら、中間解析のコストを節約でき、中間解析によってもたらされる可能性のある試験の完全性を損なうリスクを回避でき、より高い効率性という利点を有する。非劣性試験および同等性試験については、陽性対照薬のパラメータは十分に定義されており、調整の必要性がなく、サンプルサイズの調整は第一種過誤率のインフレの問題につながる可能性があるため、サンプルサイズの調整は慎重に検討することが推奨される。

(2) 妥当性

サンプルサイズの調整は、FWERを制御し、試験の完全性を確保することを前提として、調整の妥当性を確保し、調整方法に基づいて正しい統計解析手法を作成しなければならない。サンプルサイズ算出の原則に則り、調整後のサンプルサイズは、統計的有意性はあるが臨床的有意性がない状況を避けるべきである。試験自体から蓄積されたデータに基づいてサンプルサイズを再推定する場合、調整のための妥当的な時点を選択することが推奨され、サンプルサイズの調整が早すぎたり遅すぎたりすることは推奨されない。早すぎる調整は、データが少ないため、調整の根拠となるパラメータの信頼性が不十分である可能性がある。遅すぎる調整は、登録終了など試験実施上の問題に直面する可能性がある。

(3) 実現可能性

サンプルサイズの調整では実現可能性を考慮する必要がある。複数回の調整を行うと、設計と実施の複雑さが増し、操作上のバイアスが生じる可能性があるため、通常、サンプルサイズの調整を複数回行うことは推奨されない。登録終了後のサンプルサイズの調整は、臨床試験の実施に困難をもたらす可能性があるため、サンプルサイズの調整は、試験登録の進捗状況、データ収集とデータクリーニングの進捗状況とスピードの観点から評価する必要がある。さらに、試験エンドポイントの観察に長い追跡調査期間が必要な場合、サンプルサイズの調整が試験期間全体に与える影響を評価する必要がある。

(4) 完全性

サンプルサイズの調整は、操作上のバイアスの発生を避け、試験の完全性を維持しなければならない。試験の累積データに基づいてサンプルサイズ再推定を行う場合は、プロトコール、統計解析計画書、及び/又はサンプルサイズの調整を組み込んだ中間解析計画書に予め定められた規則に厳密に従うべきである。

盲検下でのサンプルサイズ再推定は、試験の群分け情報を使用せず、二重盲検試験で実施される場合、一般的に操作上のバイアスは発生せず、治験依頼者またはデータモニタリング委員会(DMC)とその独立した統計チームが行うことができる。試験の完全性を維持するために、他の状況での盲検下サンプルサイズ再推定は、DMCとその独立した統計チームによって実行されることが推奨される。

非盲検下でのサンプルサイズ再推定は、DMCとその独立した統計チームによって行われる必要がある。中間解析の実施は、非盲検のデータと結果が含まれるため、完全に機密扱いとすべきである。試験の実施に関与するすべての人員(治験責任医師、治験依頼者のプロジェクトチームメンバー、治験依頼者に雇用

された人員等を含む)及び被験者は、これらの解析結果について盲検化されるべきである。盲検化されなければ、採用された患者の特性の変化、患者のコンプライアンスの低下、被験者登録の迅速性の低下、治療群間の比較におけるバイアスの発生といった問題につながる可能性があるからである。治験責任医師は、治験の継続または中止の決定、あるいはプロトコルの改訂の決定についてのみ通知されるべきである。治験依頼者に対するDMCの助言も、試験の完全性を確保するために、有効性を逆推定する可能性のある内容の提供を避けるべきである。

一般的に、単群試験ではサンプルサイズの調整は推奨されない。単盲検試験やオープン試験のサンプルサイズ再推定は、盲検化の維持だけでなく、試験の完全性にも注意を払いながら、DMCとその独立した統計チームによって行われることが推奨される。

(二) サンプルサイズ調整のシナリオ

一般的なサンプルサイズ調整のシナリオには、外部データに基づくものと内部データに基づくものがある。

1. 外部データに基づく調整

外部試験(この試験以外)からのデータに基づいて、初期デザインのサンプルサイズ推定に関連するパラメータ(例えば、治療効果やばらつきに影響するパラメータ)を修正し、それに応じてサンプルサイズを調整することをいう。このような調整は、外部データの信頼性を十分に考慮する必要がある。

2. 内部データに基づく調整

サンプルサイズ再推定とも呼ばれ、最終的な統計検定が事前に決定または修正された目的を満たし、同時にFWERをコントロールできることを確実にするた

めに、この試験から蓄積されたデータを用いて、プロトコールに事前設定された中間解析計画書に基づいてサンプルサイズを再計算することを指す。

サンプルサイズ再推定の時点、サンプルサイズ再推定の決定ルール、許容できる最大サンプルサイズ、サンプルサイズ再推定の方法、FWERコントロールの方法、各群の調整済み点推定値および区間推定値の算出方法、群間差の点推定値および区間推定値の算出方法、群間比較時の統計量およびP値の算出方法については、プロトコールの中で明確にされるべきである。

サンプルサイズ再推定の具体的な決定ルールや中間解析報告書に記載する内容を中間解析計画書で明確にし、中間解析報告書へのアクセスや中間解析データの送信方法を関連文書で明確にすることが推奨される。中間解析計画書に、サンプルサイズ再推定における有効性の逆推定の詳細が含まれている場合、必要な人員以外が関連する内容を知らないように、中間解析計画書へのアクセス権限を設定することが推奨される。例えば、サンプルサイズ再推定計画書を作成し承認する統計家、DMC、及びその独立した統計チームのみにアクセスを制限することができる。申請者にサンプルサイズの再推定を提案する際、DMCが従うべき手順をDMC章程で明確にすることが推奨される。

盲検下でのサンプルサイズ再推定と比較すると、非盲検下でのサンプルサイズ再推定は、バイアス、FWERのインフレ、試験の完全性の低下のリスクが高くなる可能性があり、デザインと実施においてより注意が必要である。非盲検下でのサンプルサイズ再推定を使用するかどうかは、様々な要因を考慮する必要がある。例えば、より信頼性の高い事前データが入手可能な場合、非盲検下でのサンプルサイズ再推定が必要かどうか。非盲検下でのサンプルサイズ再推定を使用するコスト(例:有意水準の調整)が、初期デザインでサンプルサイズをわずかに拡大することと比較して価値があるかどうか。中間解析が迅速に完了できるかどうか、あるいは登録の迅速化により試験を調整する時間が不足する可能性がある

かどうか。中間解析の時点と推測方法が妥当かどうか。既存データが計画された中間解析の実施をサポートできるかどうかなど。したがって、様々な要因を慎重に検討した上で、試験自体の特性に基づいて適切な判断を下すべきである。

(1) 盲検下でのサンプルサイズ再推定

盲検下でのサンプルサイズ再推定については、サンプルサイズ再推定のための中間解析は、データの盲検化を解除せず、実際の試験の群分けに関する情報を使用しておらず、群間の有効性に関わる比較を行っていない。サンプルサイズは通常、最初に設定した群間差が変わらないと仮定して、他のパラメータ(例えば、イベント発生率、ばらつきなど)を再推定することによって調整される。一般に、中間解析は、群間の有効性の比較を伴わないため、FWERのインフレをもたらさない。

(2) 非盲検下でのサンプルサイズ再推定

非盲検下でのサンプルサイズ再推定については、サンプルサイズ再推定のための中間解析では試験の群分け情報が使用され、解析内容には群間の有効性の比較が含まれる。サンプルサイズの重要なパラメータ(例えば、関心のある治療効果)は、通常、試験からの累積データと群分け情報に基づいて計算され、その後、サンプルサイズが再推定されるが、中間解析では有効性の群間比較が行われるため、通常、FWERのインフレにつながるため、第一種の過誤率をコントロールする必要がある。

サンプルサイズ調整の一般的な決定ルールは次のとおりである。①条件付き検出力または成功確率に基づき、条件付き検出力または成功確率が特定の閾値内に収まった場合にサンプルサイズを再推定し、そうでない場合はサンプルサイズを調整しない。②中間解析における群間の治療効果の差に基づき、その差が特定の閾値内に収まった場合にサンプルサイズを再推定し、そうでない場合は

サンプルサイズを調整しない。

一般的に使用される調整決定ルールに対応する調整後サンプルサイズの算出方法は、通常、①調整後サンプルサイズに基づいて算出された条件付き検出力または成功確率が予め設定された要件を満たすように、条件付き検出力または成功確率に基づいて算出する方法、②プロトコルの初期設計の治療効果に対する中間解析から得られた治療効果の比率に基づいてサンプルサイズを調整する方法がある。

四、その他

(一) 試験実施中の考慮事項

治験依頼者は、計画されたサンプルサイズに従って臨床試験を完了させるべきであり、倫理的理由による早期終了や検出力が許容されなくなったことによる放棄を除き、原則として恣意的に臨床試験を終了させてはならない。サンプルサイズは試験中に恣意的に増減すべきではなく、サンプルサイズの調整はプロトコルで明確にし、プロトコルに従って実施する必要がある。

治験依頼者は、サンプルサイズの再推定において、有効性の逆推定及び試験の完全性を損なうリスクがあるかどうかを評価し、ある場合には、それを防止するために必要な対策と方法を講じるべきである。

(二) ベイジアン法に基づくサンプルサイズの推定

ベイジアン法に基づくサンプルサイズの推定は、事前分布の設定を除いて、従来のサンプルサイズの推定で考慮されたものと同じ考慮事項がある。その設計は、包括的で厳密なシミュレーション研究に基づくべきであり、事前情報の妥当性を十分に評価し、帰無仮説を棄却するための妥当的な決定ルールを確定することで、その方法がFWERをコントロールするための要件を満たすようにする。感

度計算は、他の妥当的な事前情報を使って行うこともできる。パラメータ設定、FWERコントロール、モデリングについて、規制当局と意思疎通を行うことが推奨される。

(三) その他の目的のためのサンプルサイズ推定

一般的に、サンプルサイズは主要な推定対象に基づいて推定され、他の要因（例えば、安全性評価や重要な副次的推定対象に基づくサンプルサイズ）によって決定される場合は、プロトコールにおいて正当化されるべきであり、推定方法を詳細に記述すべきである。最終的に決定されるサンプルサイズは、主要な推定対象に基づいて必要とされるサンプルサイズより小さくすることはできない。

(四) サンプルサイズの感度計算

サンプルサイズの推定には考慮すべき要素が多く、過去のデータは通常比較的限られているため、治験依頼者は一般的に、サンプルサイズの推定における不確実性のある程度低減するために、仮定からの様々な逸脱、すなわち感度計算を実施し、パラメータが一定の合理的な範囲内の値を取るようにしたり、パラメータ値が一定の分布から来るようにしたりして、仮定から逸脱したサンプルサイズの範囲を提供する必要がある。感度計算結果を保守的な原則と組み合わせて使用し、妥当なサンプルサイズの決定の指針となる方法が、慎重かつ比較的安全である。

(五) サンプルサイズ再推定とその他の適応的デザインの組み合わせ

サンプルサイズ再推定を他の適応的デザインと組み合わせる場合、または他の多重度問題と組み合わせる場合、デザインの複雑さ、特に試験の完全性を損なうリスクが高くなるなどの問題を考慮する必要があるため、厳密な理論とシミュレーションが必要とされることが多く、調整の時点と決定ルールおよび他の運用特性を決定し、それらがサンプルサイズの調整によって満たされる必要がある要

件、例えば、FWERのコントロール、検出力の達成、サンプルサイズの調整が早すぎるまたは遅すぎる(例えば、登録終了後)ことの回避などを満たすことを実証しなければならない。複数の適応の必要性を慎重に検討し、必要性を確認した上で、適応的デザインの妥当性、完全性、実現可能性の原則を厳守し、規制当局と十分かつ慎重に意思疎通を行いながら、設計・実施する必要がある。

(六) 規制当局との意思疎通

治験依頼者は、試験開始前に、重要な臨床試験におけるサンプルサイズについて規制当局と意思疎通を行うことが推奨され、プロトコルにはサンプルサイズの計算に使用されるすべてのパラメータとそれに対応する根拠が明記されるべきである。治験依頼者は、試験デザイン、有意水準及び達成すべき検出力、統計解析手法、関心のある治療効果とばらつきのパラメータ設定、サンプルサイズの算出方法などについて、意思疎通の中で詳細に説明することが推奨される。サンプルサイズの推定にシミュレーションを使用する場合は、シミュレーションコードとシミュレーション方法(シミュレーションパラメータの仮定と根拠、シミュレーションのシード値、シミュレーション回数など)の詳細な説明を提出することが推奨される。

サンプルサイズの調整を伴うプロトコルは、当該プロトコルが最終決定された後、できるだけ早く規制当局と連絡を取り、調整の根拠を提供すべきである。サンプルサイズの計算またはシミュレーションのコードと結果、FWER管理戦略、群の調整後点推定値と区間推定値の具体的な計算プロセス、群間差の点推定値と区間推定値の具体的な計算プロセス、群間比較の仮説検定統計量とP値の具体的な計算プロセスを含むがこれらに限定されない、(個別の)サンプルサイズ調整計画を、プロトコルの意思疎通を行う際に一緒に提出することが推奨される。

五、参考文献

- [1] 国家医薬品监督管理局医薬品審査センター.医薬品臨床試験における多重性問題に関するガイドライン(試行).2020
- [2] 国家医薬品监督管理局医薬品審査センター.医薬品臨床試験の非劣性設計に関するガイドライン.2020
- [3] 国家医薬品监督管理局医薬品審査センター.医薬品臨床試験における盲検化に関するガイドライン(試行).2022
- [4] 国家医薬品监督管理局医薬品審査センター.医薬品臨床試験における無作為化に関するガイドライン(試行).2022
- [5] 国家医薬品监督管理局医薬品審査センター.医薬品臨床試験の適応的デザインに関するガイドライン(試行).2021
- [6] 国家医薬品监督管理局医薬品審査センター.医薬品臨床試験データ管理と統計分析計画に関するガイドライン.2021
- [7] 国家医薬品监督管理局医薬品審査センター.医薬品臨床試験における共変量補正に関するガイドライン(試行).2020
- [8] 国家医薬品监督管理局医薬品審査センター.医薬品のリアルワールド研究のデザイン及びプロトコールの枠組みに関するガイドライン(試行).2023

- [9] ICH E1a: The Extent of Population Exposure to Assess Clinical Safety: For Drugs Intended for Long-term Treatment of Non-Life-Threatening Conditions. 1995.
- [10] ICH E8(R1): General Considerations for Clinical Trials. 2022.
- [11] ICH E9: Statistical Principles for Clinical Trials. 1998.
- [12] ICH E9(R1): Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials. 2019.
- [13] U.S. Food and Drug Administration. Adaptive Designs for Clinical Trials of Drugs and Biologics. 2019.

附属書1 用語説明

第一種の過誤 (Type I Error) : 帰無仮説 (H_0) は正しいが、検定結果が帰無仮説を棄却する過誤のことで、実際には効果のない薬を統計的推論により、有効な結論に導く過誤に相当する。その確率はある水準で管理される必要があり、これは検定水準または有意水準と呼ばれ、 α で示される。

検出力 (Power) : 帰無仮説が真でない場合に、検定結果が帰無仮説を棄却できる確率を指す。仮説検定の確実性のレベルは、 $1 - \beta$ で示される。 β は第二種の過誤を犯す確率である。

名目水準 (Nominal Level) : 多重検定における仮説検定の検定水準は、名目水準 (局所検定水準ともいう) と呼ばれ、 α_i で示される。

ファミリーワイズエラー率 (Familywise Error Rate: FWER) : 同じ臨床試験の複数の仮説検定で少なくとも1つの真の帰無仮説が棄却される確率を指す。適切なレベルで制御する必要がある。

多重性の問題 (Multiplicity Issues) : 完全な臨床試験において、研究の結論を決定するために複数の統計的推論 (多重検定) が必要であることに関する問題を指す。

推定対象 (Estimand) : 臨床試験の目的で尋ねられた臨床上の疑問を反映した治療効果の正確な説明である。同じ患者の異なる治療条件下でのアウトカムを集団レベルで集計し、比較する。

中間事象 (Intercurrent Event) : 治療の開始後に発現し、関心のある臨床的疑問に関連する測定値の解釈や測定値の有無に影響を及ぼす事象である。推定すべき治療効果を正確に定義するためには、関心のある臨床的疑問を説明する際に、中間事象に対応する必要がある。

関心のある治療効果/検出すべき治療差 (Treatment Effects of Interest, Treatment Difference to be Detected) : この研究は、分析指標における群間の差異の大きさを検出するために、一定の検出力を持つことを目指している。

サンプルサイズ再推定 (Sample Size Re-estimation) : これは、最終的な統計検定が事前に決定または修正された目的を満たし、同時にファミリーワイズエラー率をコントロールすることを確実にするために、蓄積された試験データを用いて、事前に決定された中間解析計画書に基づいてサンプルサイズを再計算することを指す。

附属書2 日本語と英語の用語一覧

日本語	英語
正確性	Accuracy
条件付き検出力	Conditional Power
データモニタリング委員会	Data Monitoring Committee, DMC
推定対象	Estimand
イベント発生率	Event Rates
ファミリーワイズエラー率	Familywise Error Rate, FWER
完全性	Integrity
中間事象	Intercurrent Event
中間解析	Interim Analysis
名目水準	Nominal Level
臨床における最小重要差	Minimum Clinically Important Difference
多重性	Multiplicity
成功確率	Probability of Success
検出力	Power
信頼性	Reliability
有意水準	Significance Level
サンプルサイズの調整	Sample Size Adjustment
サンプルサイズ再推定	Sample Size Re-estimation
感度計算	Sensitivity calculations
統計解析計画書	Statistical Analysis Plan, SAP
関心のある治療効果/検出すべき治療効果の差	Treatment Effects of Interest, Treatment Difference to be Detected

日本語	英語
第一種の過誤	Type I Error
第二種の過誤	Type II Error
妥当性	Validity