

平成4年3月4日

臨床試験の統計解析に関するガイドライン

目 次

第1章 . 序	1
第2章 . 治験実施計画書に記載すべき統計解析関係事項	
1 . 治験の目的	3
2 . 対 象	3
3 . 試験群の構成と割付	4
4 . 処 置	6
5 . 観察・測定及び臨床検査	9
6 . 評価方法及びその基準	11
7 . 中止基準	12
8 . データの質の管理	13
9 . 解析方法	13
第3章 . データのまとめ方	
1 . 個別被験者データのまとめ方	16
2 . 問題症例、不完全例の取り扱い	17
3 . 多施設共同試験	17
第4章 . 解析方法と結果のまとめ方	
1 . 処置群の比較可能性	18
2 . 統計解析の記述	18
3 . 不完全例及び欠測データ	20
4 . 多施設共同試験	20
5 . 中間解析結果	20
6 . 治験の不完全終了	20
7 . 結果の要約	20
(注)	21

第1章．序

薬効評価に当っては、その評価のために必要な観察を行う臨床の場やデータの収集・解析の過程において、各種の偏りの混入を防ぐ努力をすることが最も大切である。この立場から、適切な対照群を置いた、無作為割付による二重盲検試験法を用いることと、データの解析に当っては統計手法を用いることが基本となる。現在、薬効評価に当ってこれらの手続がとられているが、さらに改善すべき諸問題も残されている。

その一つに、統計手法の誤用による判断（評価）の誤りがある。本指針は統計手法の誤用を指摘し、誤用を防いで、科学的に正しく薬効を評価するための、現時点で最も妥当と思われる指針を示したものである。しかし、いかに正しく統計手法を適用しても、その計算のもとになるデータの質や信頼性に問題があるようでは、正しい薬効評価は望むべくもない。データの質や信頼性を高めるために遵守すべき事項については、他のガイドラインなど（一般指針（案）、GCP）に示されている。また、比較臨床試験は、地域や施設間の変動を超えて、普遍性を確保するため、複数の施設（群）で行なうことが原則である。

なお、本指針では検証を目的とした臨床試験における並行群間比較法を中心に記述している。例えば、クロスオーバー法は理論的には優れた面を有し、症例数が少ない疾患について臨床試験を行う場合などには実際上も優れた方法と思われるが、薬効評価のための臨床試験の多くの場合において、持越し効果の影響やベースラインの変化など、実施及び評価に当っては困難な問題を内包している。

統計的推論には大別して、検定と推定がある。これまで検定が多く用いられてきたこともあり、ここでも検定を例に説明していることが多い。しかし、治験薬は対照薬より有用かどうかにかかわらず、どれくらい有用か推定するほうが、臨床の場で役に立つより多くの情報を与えることが普通である。したがって臨床試験の主要な成績については、できる限り点推定と区間推定の結果を検定結果に併記することが望ましい。

本来、新医薬品の承認審査に当たっては、原則として治験薬は当該効能を有する既承認の薬剤よりも優れているか、または同等と評価された場合にのみ承認されるべきである。比較試験において治験薬が対照薬と有意差がなければ直ちに同等であるとするのは統計学的に問題がある。本指針では臨床的な実用性を加味した観点から、許容できる範囲を越えては劣っていないことを積極的に証明する方法を使用するように提案している。本方法の活用と、より良い方法を考案していただきたい。

本指針に沿って、質の高い、信頼性のある薬効評価のための資料を作成されることを望むものである。

第2章．治験実施計画書に記載すべき統計解析関係事項

1．治験の目的

治験の目的は明確かつ具体的に表現する必要がある。主要目的の他に副次的目的があれば、それも具体的に記載する。

あまりにも多くの目的を一つの治験計画に設定すると、推論の多重性等、統計学上困難な問題が生じ、個々の目的についてはかえって情報量が減る場合があるので注意を要する（注1）。

2．対象

(1) 選択基準

対象の記述に当ってはその臨床試験がいかなる被験者集団を想定しているかを具体的に表現する。すなわち、その試験で得られた結論がどのような範囲に適用されるのかを表現すべきである。試験対象として想定する臨床状態の定義は被験者選択の基準と直結するものであるから、試験への組入れ基準（例えば、疾患の重症度、特定の臨床検査成績の結果など）の形に整理して明確に記載する。

通常、次の諸因子に関して規定する。

- ア 健常人と患者の別
- イ 年齢及び性別
- ウ 入院・外来の別
- エ 対象疾患、症状及びその重症度（病期）
- オ 病歴、合併症及び基礎疾患の有無
- カ 前治療の内容

上記条件のチェックの為に試験前観察期間中に行う必要がある検査についても記述する。（注2）

(2)除外基準

信頼性のあるデータが得られず有効性又は安全性の評価に偏りをもたらすと考えられる被験者をあらかじめ除外すること、治験を安全に遂行することを考慮して、必要な除外基準を設定する。その際、一般的規定なのか、当該治験薬に固有の規定なのかを区別して設定するとよい。治験薬に固有な安全性の配慮は、対照薬があればその使用上の注意も参考にする。

(3)除外対象として考慮すべき被験者

- ア 効果の発見がほとんど期待できない被験者
 - イ 主要な項目が観察不能又は信頼性に欠けると予測される被験者
 - ウ 目的とする疾患とまぎらわしい症状を示す被験者
 - エ 計画書の処置、観察の規定の遵守が難しいと考えられる被験者
 - オ 新生児・幼小児、高齢者の被験者
 - カ 肝・腎・心・肺等に（治療目的以外の）高度の障害のある被験者
 - キ 妊娠又は妊娠の可能性のある女性、授乳中の被験者
 - ク 過去に薬物アレルギーを示した被験者
 - ケ その他、特定の疾患の既往などを有する被験者
- （注3～5）

3．試験群の構成と割付

試験計画の要約を記述する。治験の目的を考慮して、偏りの除去、実施可能性などの諸要素から見て適切な試験計画となっていることが必要である。

(1)対照群

同時対照としての無処置、プラセボ（placebo）投与、実対照薬投与又は異なる用量投与のいずれによるかを記す。

用量検討試験では、用量の増減規定を明確にして漸増法又は漸減法を採用することがあるが、症状経過に自然変動があり同一用量でも症状が大きく変動する場合や、増減規定違反の可能性が大きい場合には漸増法、漸減法は好ましくない。

既存対照(historical control)は特殊な場合に限って用いられる。

(2)群の構成

群間比較、被験者内での左右比較、クロスオーバー比較などがあるが、どの方法によるか及びその根拠を記す。

(3)無作為割付

試験群の割付の際に生じうる系統的な偏りを防ぎ、比較可能な群形成を確率的に保証するために無作為化(無作為割付)を行う。これは統計解析の根拠を与え、無作為化法としては完全無作為化の他、重要な背景因子(予後因子;共変量)に関する層別無作為化、時間経過に関し短期的にも均一性を保証する置換ブロック法等がある(注6)。

なお、一般に比較試験では無作為化二重盲検法が原則であるが、二重盲検法を採用できない場合には、無作為割付の手段として封筒法や、電話割付法を採用することがある。その際、封筒法では割付(被験者の組入れ)、処置内容及び判定に偏りが生じやすく、偏りを防止するには十分な注意を要する。

(4)盲検の種類

採用された盲検の種類(単盲検、二重盲検、評価者のみ盲検など)と、それを実行するための特定の手順(容器のラベルの貼り方、ダブルダミー法など)を記すことにより、被験者、観察者及び解析者に由来するどのような偏りを防止しようとしたかを示す。また盲検の解除(コード開封)の条件や、手順及びこのような解除が個々の被験者についてなされるのか、試験の一部についてなされるのか、全体についてなされるのかを記すこともある。

盲検化が望ましいが実施不能であったり、明らかな副作用のため不完全であると考えられる場合など、その理由と影響について記述しておく。

(5)症例数

計画された症例数を示し、その根拠となった統計的考察あるいは実施上の制約などについて記述する。統計的考察では、症例数あるいは検出力を、それを算出した方法とともに示し、算出の際に用いた薬剤間の有効性の差等とその根拠を説明する。

実対照薬との臨床的同等性を示すことを目的とする臨床試験においては、症例数の設定にあたって、治験薬が実対照薬と比べて劣っている場合に臨床的に許容できる差()を延べ、当該症例数がそれ以上の差を検出する検出力を示す(注22)。

(6)試験終了規則

試験全体の終了規則(逐次解析、例数、期間等)について記載する。

(7)組入れ期間

基準に合致する被験者を選択する期間について記載する。

(8)被験者の登録

治験に組入れる被験者を登録し、治験の進行状況及び計画書の遵守状況把握に役立て、また、対象選択基準との合致を確認する方法を記述する。割付センターが設定されている無作為化比較試験では、登録の申請を受けた被験者が選択基準に合致していれば治験に組入れる。

4. 処置

計画全体としていつから処置を開始するのか、いつ処置を終了又は中止するのかを明記しておく。被験者ごとの試験スケジュールの決定に従って、各人についての投薬・観察など試験のスケジュールを記す。このスケジュールに対応して、各期間ごとに処置に関する事項、すなわち、処置内容、用法・用量、用量変更の指標(たとえば段階的に投与量を増減させるときの手順)、前処置、併用薬・併用療法(理学療法等)の内容、許容範囲などについて記す。これらの規定は治験計画立案段階でその目的、実施可能性を十分考慮して記述する。

(1)ウォッシュアウト (wash-out) 期間

前処置のウォッシュアウトがもたらす病態の変化、ウォッシュアウトしない場合の効果判定などの観点からウォッシュアウトの必要性について記す。また、ウォッシュアウト用のプラセボを使用するか、使用する場合はその形状・色を治験薬と変えるか等について記す。ウォッシュアウト期間を設けている試験においては、その期間での処置内容、その理由について記載しておく。

(2)前観察期間

ここでいう前観察期間とは、治験に組み入れることを予測した時点以降、治験薬の投与開始までの期間をいう。組入れの適格性の判断、処置法の統一及び前観察期間中のベースラインの確保等、この期間をどんな目的で設け、その期間にどのような観察・測定を行うかを記す。

前観察期間が設けられている試験で処置法を統一する必要がある場合には、その内容、理由を記載しておく。一般的には、前観察期間の後に無作為割付すること望ましい。

(3)投与期間

効果の発現、病態の自然変動などを考慮して記す。

短期投与 / 長期投与、単回投与 / 連続投与 / 繰り返し投与など治験の目的によって異なるが、十分な臨床効果が得られる期間を設定することが重要である。その際、屯用効果、ベースラインの変動、耐薬性の発現その他により、治験成績がどのような影響を受けるかに関しても考慮すべきである。記載は具体的に行い、“1週以上、できれば2週以上”のような幅のある表現は用いない。

(4)投与量、投与方法

1回投与量、1日投与回数、投薬時期、投薬方法、被験者への指示事項などについて、項目別に、具体的に記載する。用量を増減する方法を用いる場合、変更の基準を明記する。

(5)服薬の確認

どのような方法・対策を講ずるかを記す。服薬確認の方法としては、問診、被験者日誌の利用、残薬の回収、同時投与のマーカ―や薬剤成分の血中・尿中濃度の測定などが考えられる。服薬忘れの対策としては、被験者に処置スケジュールの説明を詳しく行なうことが有効である。

(6)追跡調査期間

追跡調査期間を設ける場合、その意義を記す。

追跡調査期間が設けられている試験で処置法を統一する必要がある場合には、その内容、理由を記載しておく。

(7)併用薬剤・併用療法

試験目的、対象疾患の種類に応じ、必要ならば前観察期間、投与期間、追跡調査期間の各期間ごとにその条件、許容範囲を例えば次のように規定しておく。

(ア) 絶対禁止（構造物、作用的に類似している薬剤、併用禁忌など）

(イ) 一定量・一定期間なら可

(ウ) 新規投与は禁止

(エ) 試験期間に入ってから中止

(オ) 併用自由

イ：また、必要に応じ、次の事項について検討しておく。

(ア) 他の開発中の薬剤の併用に関する禁止規定

(イ) 基礎治療や急性症状に対する屯用療法に関する規定

(ウ) 副作用、合併症治療に対する規定

(エ)

場合によっては、併用薬剤の使用の有無又は量などを効果判定の指標として利用する試験計画も考えられる。例えば、狭心症に対するニトログリセリンの使用頻度などは、この場合に該当する。

ウ：相互作用

併用薬剤の許容範囲決定に際しては、次の例に示すような治験薬との相互作用の有無についても検討し、可能性がある場合には前治療薬、併用薬に関する注意事項として、副作用とともに記載しておく。

(7) 調剤学的相互作用：配合禁忌の注射液における混濁等

薬理的相互作用：アスピリンと経口抗凝血薬との併用による血液凝固

(8) 阻害作用の増強等

薬物動態学的相互作用：テトラサイクリンと制酸薬の併用による吸収阻害等

(I) 酸素誘導阻害作用：ジクマロールによるトルブタミドの作用増強等

5．観察・測定及び臨床検査

記録すべき背景因子及び有効性・安全性の観察・測定項目、時期及び方法についてその妥当性・信頼性を考慮した上で記述する。観察時期について明確な記述をする。それは試験開始時からの経過時間で定めてもよいが、適切な場合には薬剤投与時期との関係で定めてもよい。

複数の有効性、安全性の観察・測定項目をもつ試験の場合には、これらの扱いについて検討しておく。

(1) 有効性、安全性の観察・測定項目

治験の目的に直結し、かつ、再現性のある結果を与える観察・測定項目を過不足ないように選択する。

なお、有効性、安全性を主観的なカテゴリー、数値スコアなどによって評価する場合には、評価の妥当性を検討し、観察者内の再現性、観察者間の一致性が得られるように努力する必要がある。

(2) 臨床検査、理学検査

これらの項目は合理的な必要最小限にとどめる。施設によって測定方法や表現方法が異なり、それが評価に影響を与えると予想される場合には、統一して解析するための工夫、すなわち、施設ごとの正常値、測定機器の種類、観察・操作方法、試薬の種類、測定単位、採取量、記載方法などを必要に応じ明記する。また、特定施設において統一的に測定する場合で、検体の保存方法、包装形態、運搬方法及び検査結果の送付方法などを定めておかないと試験の質を保証し得ない場合には、その内容を具体的に記す。

(3) 観察・測定時期

多くの場合は時系列的に観察・測定されるが、次の事項に配慮して観察・測定時期及び1回ごとの許容しうる時間幅を決定し明記する。

- ア 症候、測定値の変動パターン：傾向変動／循環変動／不規則変動
- イ 試験計画：時点固定型／反応固定型
- ウ 投薬時刻、薬物動態との関連性
- エ 繰り返し観察における観察・測定の時間的間隔
- オ その他（休日・祝祭日、学会、年中行事など）

(4) 定義の明確化、基準化

観察・測定項目の変動に対する段階付けや表現方法の基準を規定する。計測値をそのまま記録する場合には、有効数字を取り決めておく。

(5) 医療従事者用記録用紙

服薬状況、副作用、自覚症状の調査に利用できるが、作成する場合には実施可能な内容にしておく。

(6) 手順のまとめ

観察・測定項目の重要度順、実施時期などを示した一覧表を作成しておく。臨床検査、理学検査については、必要な場合には、その操作手順、観察・測定方法などを別にまとめておく。

6．評価方法及びその基準

個々の評価項目、それらの総合評価について評価時点、評価方法（用語、基準）を明記する。

(1)個々の評価項目

観察・測定項目あるいは臨床検査・理学検査に基づいてさらに評価を行う場合には、各評価項目毎に評価基準を規定する。経時的な評価項目では、評価時点を明確にしておく。評価方法としては治験担当医師又は被験者が主観的に行う場合と、ある基準のもとに機械的に行う場合がある。また場合によっては、問題とする観察・測定項目あるいは臨床検査・理学検査において、症候・所見が少ない被験者の各評価項目ごとの扱いを事前に検討しておく必要がある。

(2)総合評価

一般には複数項目を要約し、個々の被験者について次のような観点から総合評価を行う。

重症度の変化、被験者の苦痛の軽減度の経過などの全般的な改善度

イ 安全性： から判断

副作用、臨床検査の異常値又は異常変動及びその薬剤使用との関連性を考慮する（注17）

ウ 有用性：有効性、安全性、QOLなどを考慮した治験担当医の判定

これらの評価を意味あるものとするために、評価の妥当性、再現性、一致性について統計的に検討を要する場合がある。試験目的によっては時系列的に観察した項目を分析・評価することもある（注13）。

(3)判定委員会

判定委員会は次のような目的のために必要に応じて設けられる。

ア 総合評価の基準の統一化、精度の向上

イ 副作用など判定に特別な知識を必要とする場合

ウ 判定基準が複雑で再チェックを必要とする場合

エ 写真判定や記録用紙の解読など、一括して評価した方がよい場合（注7）

オ 臨床試験の中止・続行の決定を必要とする場合

(4)不完全例の評価

不完全例はその発生時期と内容により分類されるが、各被験者ごとに内容を検討し、治験の目的によっては部分的にでも評価できるものは取り上げるようにする（注16）。

7．中止基準

処置・観察を中止する場合の条件、中止後の処置、記録の作成、追跡調査に関する事項について規定する。

(1)中止の条件

治験担当医師の判断によって治験を完了せずに中止する場合、次の事項に整理し、有効性、安全性、有用性の評価に加えるべき被験者を見落とさないことが大切である。また、被験者来院せずの場合、その理由によって中止と扱う場合もある。予め治験実施計画書に記載出来る中止条件として、次の事項が考えられる。

ア 症状の悪化

イ 試験処置の有効性が認められない場合

ウ 重篤な副作用、臨床検査値異常変動

エ 合併症の増悪または偶発症の発生（不慮の事故を含む）

オ 処置方針の変更を必要とする場合

カ 上記以外の理由で試験計画の遵守が不可能になった場合

キ 試験開始早期に、対象外であることが判明した場合

また、上記以外に試験を完了できない場合として

ク 被験者又は家族による同意の撤回

ケ 担当医の異動により、正確な引き継ぎが行われず処置を中止した場合

コ その他医師の判断により中止が必要とされる場合

(2)記録の作成

途中で中止した対象の中止理由及び時期について記す。副作用や臨床検査値異常の発生による中止例では、中止後の追跡調査の内容及びその結果を記録として残すよう明記する。

8．データの質の管理（注8）

治験の信頼性確保のためにデータの質の管理は重要である。治験の質の評価に関与するものとして欠測値、外れ値（outlier）、評価時期のずれ、服薬状況などがあげられる。これらのデータの入力に際しては、入力方法を具体的に規定しておくことが重要である。比較試験では無作為割付が正しく行なわれていることが前提である。また、評価項目間の整合性を確かめておく等の配慮が必要で、結果の再現性、信頼性を確保するよう、努めなければならない。

9．解析方法

臨床試験では不完全例が生じる可能性があり、常に完全なデータが得られるとは限らない。これらの不完全例はデータ解析に際して偏りの原因となるので予め発生を予想し、理由別の分類と解析に際しての取扱い方針を治験の目的に応じて規定しておく必要がある。

背景因子に関して処理群間の比較可能性を損うような不均衡が生じた場合の対応を事前に検討しておく（注20）。

薬効を検証するための主要な統計解析を、副次的な統計解析と区別して明記する。一般的には、有効性、安全性、有用性ごとに主要な統計解析を行う。主要な統計解析については、解析項目と共に、統計的仮説とそれを検証する適切な解析手法を予め選択し記載する。この際、有効率、安全率、有用率の算出に当たっては、その定義(分子・分母の内容)を明確にしておく。また部分集団について別途解析するのか、どの時期を問題とするのか、3処理群以上の場合にはどのような比較を重視するのかなどに関しても考慮しておくべきである。主要な統計解析として複数の統計的推論を行う場合には、有意水準、信頼係数に注意しなければならない(注9～14)。実対照薬との臨床的同等性を検証することを目的とする臨床試験の場合には、実対照薬と比べて、有効性などが臨床的に許容できる最大限の差（ ）を越えては劣っていないことを証明するための解析方法を明記する(注22）。

(1)不完全例の取扱い

不完全例の取扱いにより試験結果に大きな偏りを生じたり、データの信憑性にまで影響を及ぼしたりする事が考えられる。したがって、予めこれらの発生を考慮し、理由別に分類区分するとともに、試験の性格、つまり、研究的か、実践的か、あるいは、検索的か、検証的か、に応じて、解析に際しての取扱い方針を明確にしておかねばならない（注15、16）。

(2)背景因子の扱い

記録すべき背景因子とは、臨床的に治療予後と関連することが知られており、その処置群間での不均衡が比較試験結果に影響する可能性のある要因である。薬効に関する統計解析を行う前に、記録された背景因子に関する処置群間での比較可能性を点検する必要がある、その手順を規定しておく。

背景因子の統計解析では、5%、1%の有意水準あるいは95%、99%の信頼係数といった基準による評価ではなく、背景因子不均衡が有効性・安全性・有用性評価にどの程度影響するかという観点から検討されることが重要である。多くの場合、有意確率が15%前後といった背景因子の不均衡についても注意深い検討が必要である。また、複数の種類の統計手法を用いて、異なる種類の不均衡を点検することも勧められる。

背景因子の不均衡が生じた場合の対応についても検討しておく必要がある。統計的な調整を実施する可能性がある場合には、その解析手法について検討しておく。

(3)解析項目

一般に、主要な統計解析の解析項目は有効性、安全性、有用性それぞれの総合評価である。場合によっては治験の目的にあった解析項目を観察・測定項目、臨床検査・理学検査および評価項目の中から予め選択することもできるが、これらの扱いについては十分な検討が必要である（注12）。

副次的な統計解析として、主たる解析項目以外の観察・測定項目、臨床検査・理学検査および評価項目についても十分な解析がなされるべきである。副次的な統計解析は、主要な統計解析の裏付けとなるとともに、治験薬に関する新しい臨床的仮説を提供する可能性がある。過不足のない副次的解析を計画し、治験実施計画書に記載するなど事前に検討しておくことが望ましい。

(4)有意水準・信頼係数

解析時の有意水準、信頼係数、及び両側か片側検定のどちらを用いるかを明記する。通常、5%、1%の有意水準及び95%、99%の信頼係数が用いられている。検定では有意であるか否かだけでなく、有意確率（ p 値）を記載しておくことよい。

(5)解析手法

解析手法は試験の仮説、計画および評価尺度等に対応して適切に選択されなければならない。例えば、順序分類データの2群比較において無定見に何通りもの検定を繰り返し適用すると第一種の過誤の確率が大幅に増大する(注11)。多群比較においては、群間の一様性検定よりも、目的に応じた多重比較の方が合理的である(注9)。用量群比較においては水準の順序を考慮した多重比較法、用量反応関係の解析方法も用いることができる。

(6)統計解析担当責任者

統計解析を担当した者の氏名、所属及び連絡先を明記する。

第3章．データのまとめ方

収集データは、解析に採用された被験者ばかりでなく、解析から除外された被験者も含めて整理する必要がある。

1．個別被験者データのまとめ方

無作為化された被験者は、たとえ治験薬又は対照薬の投与された試験期間に入っていないとしても、データ表から除外してはならない。

試験への組入れ基準に合致せず、無作為化されなかった被験者はデータ表に入れる必要はない。すべてのデータを示すに当たって各被験者に識別番号を用いる。各被験者に関し以下の様なデータを示す必要がある。

(1)人口統計的項目；年齢、性別、身長、体重等

(2)重要な背景因子；重症度、検査値、合併症、前治療の効果等

治験薬の投与量と投与日数（投与量を効果などに応じて増減した場合は、各投与量と最終投与量及びそれらの投与日数）

(4)治験の開始及び終了年月日

(5)服薬状況

(6)有効性、安全性、有用性に関する主な評価項目の値と評価時期

(7)その被験者が有効性、安全性及び有用性評価に採用されているか否か。

(8)随伴症状（臨床検査の異常値又は異常変動を含む）

ア 症状の種類

イ 発症した時期

治験薬との関連性。（例えば、明らかに関連あり、関連性が高い、関連性があり得る、又は無関係）

エ 重症度（例えば、重症、中等症、軽症）

オ 転帰：死亡、治験薬の投与中止、投与量の変更、そのまま継続投与など

カ 発症時の投与量

キ 特に処置中止を要するほど重篤な症状の場合、発症時までの総投与量

一人の被験者に複数の重篤な副作用が発症した場合は、それらの副作用の種類と発症時期を示すために、その被験者の概要を別記する（第4章（5）参照）。

2．問題症例、不完全例の取扱い

すべての問題症例、不完全例（不適格例、中止例、脱落例、追跡不能例など）について一覧表を作り、下記の(1)～(6)に関して記述する。

- (1) 問題の内容
- (2) 問題発生時期
- (3) 問題発生時の投与量
- (4) 問題発生時における併用薬
- (5) 問題発生理由
- (6) 不完全例の取扱い（注18、19）

3．多施設共同試験

有効性、安全性及び有用性に関するデータは、治験担当医師ごとに結果を把握できるようにまとめておく。特に多数の被験者を対象に治験を実施した治験担当医師については、有効性、安全性及び有用性の評価を治験担当医師ごとに別々に示す。

第4章．解析方法と結果のまとめ方

解析結果には、処置群の比較可能性に関する考察、問題症例及び不完全例の取扱いを含めた統計解析法を記述する。多施設共同試験の場合には、評価の整合性に関する吟味についての記述を含める。

(1) 処置群の比較可能性

背景因子、解析採用例数、治験遵守例数などについて、処置群間の比較可能性を表、グラフその他適切な統計的方法を用いて示す。取り上げるべき背景因子は対象とした疾患の特性によって異なる。通常それらには、年齢、性別などの人口統計的項目と、罹病期間、疾患の重症度、ベースライン、治験前の治療などの項目が含まれる。

(2)

ある背景因子が処置群間で異なった場合には、その事実がその後のデータ解析および結果の解釈におよぼす影響を記載する。重要な背景因子について統計上の調整を行うならば、それらの因子を選択した理由を説明する（注20）。

(3)

既存対照を使う場合には、予め、その対照群をどの様にして選び、他にどのような対照群の候補について文献を調べたか、及びそれらの候補での結論が実際に用いられた対照群と比べて同様か否かを説明する。

2．統計解析の記述

(1) 主要な統計解析の結果

有効性、安全性及び有用性の解析結果を提出する場合には、以下の事項を示す。

ア．解析項目

イ．解析手法

ウ．統計的仮設とその検定結果

エ．主要な特性値（平均値、出現率など）の推定結果

特に、イ、ウ及びエについては統計用語により記述し、検出力、推定精度などの立場から考察する。

(2) 副次的な統計解析の結果

検証的な統計解析結果を裏付ける副次的な統計解析結果を、過不足なく提示する。

さらに、探索的な統計解析により示唆された興味ある事実をまとめる。これらは、新たな治験により検証すべき臨床上の仮説となることもある。

(3) 解析手順

ア．計算経過

解析結果を迅速かつ容易に確認できるような要約資料を必要に応じ提示する。

イ．検定

検定統計量の値、有意水準及び有意確率（ p 値）をまとめる。有意水準及び p 値は、片側又は両側のどちらであることを明記する（注21）。

ウ．推定

点推定値、信頼区間及びその信頼係数をまとめる。

エ．図表

解析結果の指示に当たっては、適切な図又は表を用いてまとめるとわかりやすい。図表には必要に応じ標準偏差や標準誤差などを、そのいずれであることを明示して、記入する。

(4) 臨床的同等性

プラセボ以外の対照薬(実対照薬)を用いた試験で、両薬剤で観察された類似性を根拠として治験薬の有効性、有用性が主張されることがある。この場合には、治験薬の有効性、有用性が対照薬と同等又はそれ以上であるという積極的な証拠が必要である（注22）。場合によっては、当該治験での実対照薬との比較の他に、過去の類似試験におけるその実対照薬の効果水準とその一貫性についての記述、あるいは治験薬の有効性が同様の試験条件下でプラセボを用いた時に要求される最低限の有効性を上回っていると判断した根拠についての記述が必要である。

(5)副作用

各処置群における副作用について以下の諸事項を示す。

副作用の種類並びにそれぞれの副作用について、出現例数、評価対照例数
イ 及び出現率

処置群間の副作用発生率の統計解析による比較。必要に応じて、体重、年齢、性別、併用薬、治験薬の投与量、投与期間、臨床検査の異常値又は異常変動などの被験者特性との関連を見る。

ウ 臨床検査の異常値又は異常変動についてもこれに準じる。

3．不完全例及び欠測データ

一般に、不完全例及び観測値の欠測は臨床試験における偏りの潜在的要因になる。不完全例及び欠測データの取り扱い方の試験結果に対する影響を考察する。通常、試験の目的に応じて、試験に組入れられたすべての被験者又は有効性に関するデータが調べられているすべての被験者を対象とした統計解析を加えておく。

4．多施設共同試験

多施設共同試験の結果を一つにまとめるときは、その妥当性を示す。すなわち治験担当医師は異なっても結果には十分な一貫性があるかを記載する。評価の施設間変動が大きいとき、又は極端に異なる評価を行った治験担当医師があったときには、それらを明らかにする。

5．中間解析結果

治験の進行途中で統計解析を行った場合にはその結果を提出して、それらが試験の進行及び最終的結論に及ぼした影響を記述する。計画書に記載されていないにもかかわらず、中間解析を行った場合には、その理由を説明する。

6．治験の不完全終了

目的とした症例数と著しく食い違って試験を終了した場合には、その理由を明記し、結果に及ぼした影響について考察する。

7．結果の要約

得られた統計解析結果を用いて、その臨床上の特性について要約する。

(注)

1. 推論の多重性

臨床試験からは薬剤の作用に関する多量のデータが生まれ、これに対して如何なる解析方法を用いるかには多くの選択の余地がある。現在のところ臨床試験から得られた多量のデータに対して多数の統計的推論が行われている。しかしながら、多数の統計的推論を行えば見かけ上有意な結果が現れる確率が高まり、第1種の過誤の確率が大きくなるという問題がある。この問題は臨床試験における多重性 (multiplicity) として議論されている。

現在わが国の比較試験で議論されている多重性に関連した主な問題には、以下のものがある。

a. 多群 (剤) 比較

3群 (剤) 以上の多群間の比較を、例えば、2群 (剤) ごとの対を作り比較する場合 (注9参照)。

b. 順序分類データの切直し

分割表を種々のところで区切って、繰返し検定を行う場合 (注10参照)。

c. 多種検定

1組のデータに対して複数の異なる種類の検定を行う場合 (注11参照)。

d. 多項目比較

複数の観察・測定項目、臨床検査・理学検査、評価項目などのひとつひとつの項目ごとに比較を行う場合 (注12参照)。

e. 経時的比較

効果に経時的な相違があるか否かをみるために、各評価時期ごとに比較を行う場合 (注13参照)。

f. サブグループ解析

性、年齢、病型、重症度などの多数の層について比較を行う場合。(注14参照)

g. 中間解析

試験の途中時点で何回かの解析を行う場合。

2. 被験者の適格性

被験者組入れの際には、適格性チェックシートを利用して、被験者が基準を満足しているかどうかを確認しておくに役立つ。

3．合併症

肝・腎・心・肺疾患等の合併については、病名がつくか、治療が必要かなどが目安となるが、主要検査項目の値で規定することもできる。

4．妊婦・授乳婦

妊婦、授乳婦に対する投与の是非は、生殖試験、胎児への移行、乳汁中への移行などに関する試験成績に基づいて判断するが、特定の薬剤では避妊について取り決める必要がある場合もある。

5．被験者の協力

環境要因（地理的・社会的）及び被験者（あるいは保護者）の協力について記載することが必要な場合もある（長期追跡調査における遠方の被験者等、治験開始後に来院が難しくなる場合等）

6．層別無作為割付

予後への影響の大きいことが予想される因子については層別無作為割付を行うことによって層ごとのバランスをはかることができるが、過度の層別は推論効率の低下を招くので層別の必要性、合理性を明らかにしておく。多施設臨床試験においては施設を層別因子とし、施設ごとに置換ブロックを用意するのが一般的である。また、複数の重要な背景因子についてバランスをはかるために最小化法などが提案されているが、解析にあたっては特別な注意が必要である。

7．委員会評価

委員会評価は全データを短期間に統一的に行うので、再現性、一致性の向上には役立つが、治験担当医師でなければわからない被験者個々の問題が無視されることがある。

8 . データの質の管理

治験の質はデータの質によって左右される。治験が計画に沿って実施され、調査・観察・測定・評価が精確に行なわれ、正しく解析され、解釈されれば、妥当性・信頼性の高い結論が得られる。狭義のデータの質の管理とは、集積されたデータの誤りを発見して正すことであるが、いくつかの段階に分けて扱える。

治験の規模や期間にもよるが、症例記録、付随する資料などは特定の被験者についての治験が終了したならば速やかに回収することが原則で、場合によっては登録記録、診断検査記録、処置記録、追跡記録などに分割して回収することもある。これによってデータの質の管理は容易になる。

症例記録は多くの場合に、コード化され、コンピュータに入力され、編集されて、統計解析によって縮約される。まず、症例記録の段階で、データの欠損や調査・観察・測定・評価時期のズレをチェックし、これらには必要に応じて特定のコードを割り当てる。ついで入力したデータについて読み合せを行なう。独立に入力したデータの引き算が0になることをチェックすることもできるが、手書きの数字を同時に誤って入力することもありうるので症例記録あるいはコード化の段階でのクリーニングが重要になる。

入力したデータについては、データの性質に応じて慎重に吟味する必要がある。範囲チェックでは、例えば被験者の年齢が20歳以上、65歳未満であれば、これから外れた例を拾いだす。臨床検査値なども同様にチェックし、外れ値を拾いだす。論理チェックでは、例えば最大血圧 - 最小血圧が負になる例、副作用なしであるのに副作用のため中止となっている例などを拾いだす。いくつかの観察・測定項目を改善度に総括してある場合には、この概括評価の妥当性をチェックすることもある。グラフィカルな方法も編集チェックには役立ち、経時的なプロットの視察によって異常を発見できることも少なくない。

こうしたチェックで拾われたものが全て誤りであるとは限らず、多くの場合には症例記録に戻り、必要ならば治験担当医師まで戻って、正誤を調査することになるが、治験が終了して時間が経過するほど、この作業は困難になる。こうしたチェックで拾われなかったものが全て正しいという保証もないため、とくに中心的なデータについてはデータ採取の時期での注意が重要になる。

9 . 多群比較

従来3群以上の臨床試験の解析では、2群ごとの対のすべての組合せを作り、各対ごとに2群の比較試験と同じ方法による比較を繰り返すことが多かった。この場合、統計的検定の有意水準は個々の対比較ではとなっているが、複数の対比較からなる全体の有意水準は設定したよりかなり大きな値となる。検証的データ解析で統計学的に例えば5%有意という保証を与えるためには、全体の有意水準が5%となるような正しい有意点を設定しなくてはならない。

この多群比較への対応として、F検定やKruskal-Wallis検定のような一様性検定の適用、あるいは多重比較法の適用などが考えられる。しかし、一様性検定は多群の効果がいずれも等しいか否かをみるための方法であって、臨床試験で関心のある特定の2群間の比較に注目しているわけではない。このために一様性検定によって「効果が等しい」という帰無仮説が否定されたとしても、多群間のどこに差があるかは直ちに明確にはならない。さらに問題なのは、特に注目したい2群間の差が、一様性検定で有意にならないばかりに見逃される可能性があることである。臨床試験の仮説構造に従って、適切な多重比較法を実施することが望ましい。

それら多重比較法としては、基準の群に対して他のすべての群を比較するのに適したDunnnett法、群の全ての対比較に興味のある場合に用いられるTukey法、より一般的なScheffe法などがある。また、正確な有意点が得られない場合には、Bonferroniの不等式を利用した保守的な近似法が利用できる。これらの手法は、計量値のみならず、カテゴリカルデータの場合にも用いることができる。なお、多重比較に対応した区間推定は同時推定と呼ばれる。

多群比較は、たとえ適切な多重比較法を用いたとしても、結果の解析や検出力の低下など統計的に複雑な問題を引起す可能性がある。したがって、検証を目的とした第Ⅱ相臨床試験では、可能な限り実対照薬又はプラセボとの2群比較を行うことが望ましい。十分な理由から3群以上の比較を行う場合には、症例数設定の根拠と解析計画を治験実施計画書に記載する比較がある。

10．順序分類データの切直し

例えば最終全般改善度について著明改善以上、改善以上、やや改善以上、悪化以下といった区切りによって多数の4分割表をつくり通常の χ^2 検定を繰り返すことがある。臨床的には各区分の比率についての情報は理解しやすい重要な情報であるが、このように検定を繰り返すことは第1種の過誤の確率を増大させる。このような場合には、例えば、最大 χ^2 法により正確に有意確率を計算することができる。

11．多種検定

最終全般改善度や有用度の分割表に対して Mann-Whitney U検定や χ^2 検定など複数の種類の検定方法を適用することがあるが、この場合にも当然、第1種の過誤の確率の増大が起こる。順序分類データに対しては2標本 Wilcoxon 検定 (Mann-Whitney U検定)、累積 χ^2 法、最大 χ^2 法などの適用が提案され、それぞれの特徴が示されている。例えば、薬効差を検討する場合、U検定は平均的な薬効差に対する検出力が高い。最大 χ^2 法は、悪化あるいは著明改善といった臨床上重要な尺度段階での薬効差に関心がある場合に適する方法である。累積 χ^2 法は平均的な薬効差の他に効果のばらつきや分布の形の差にも適度な検出力をもっている。

比較臨床試験のデータにこれら異なる検定方法を用いた場合、統計的推論の結果が異なることが少なからずあることが知られている。

12．多項目比較

複数の観察・測定項目、臨床検査・理学検査、評価項目などについて非常に多数の比較を行う場合、個々の検定での有意水準は保たれていても、全体としての第1種の過誤の確率は増大する。この増大の程度は比較する項目間の相関構造によって異なる。例えば、20項目について全く同等な2薬剤を比較しても、どの項目間にも相関が全くないときには、いずれにか有意差が現れるのは64%ほどであるのに対して、どの項目間の相関係数も0.8のときには22%程度となる。項目間の相関が非常に強い場合には第1種の過誤の増大は軽微であるが、相関の弱い多数の項目で薬効差を検証しようとするすると第1種の過誤の増大は重大な問題となる。

項目間の相関構造を事前に推定しがたいことが、多項目比較における第1種の過誤の制御を困難なものにしている。

この問題に対する対処のひとつとして、個別の解析における検定の有意点ないし有意確率を修正して第1種の過誤を押さえようとする考え方がある。そのうちで無相関の状況を仮定した Bonferroni 流の修正は現実には過度に保守的な方法となる可能性がある。総合解析としては、事前に複数の測定をどのようにまとめて扱うかを決めておいて統計的調整の困難を回避しようという行き方もあるが、この場合にはまとめ方が医学的に妥当であるか否かの検討が重要である。また、複数の検定統計量の重み付き和により少数の統計量に縮約して検定する方法も提案されている。

治験薬の効果の検証を意図する第 相臨床試験では、事前に医学的に重要な少数項目にしぼり、それを治験実施計画書に明記する必要がある。相関が多角ないことが予想される複数の項目についての比較が必要とされる場合には、複数の結果をどのように扱うかを事前に決め、適切な統計的調査を図る必要がある。

一方、薬効差の検証ではなく、今後検証すべき効果の性質を絞るための比較試験の事後的な探索的データ解析では、有意差の出現した項目数や有意確率のレベルなどを考慮して妥当な仮説を作ることが重要である。

1 3 . 経時的比較

比較試験では各被験者について経時的に種々の評価や測定がなされ、経時的な薬効差の検討が行われる。この際の検定に当っては、薬剤群ごとに処置開始時点とそれ以降の評価時点とに対応のある場合の検定を繰返したり、各時点とに2群間での対応のない場合の検定を繰しがちであるが、それでは第1種の過誤の確率の増大する。このようなデータの場合には時点毎の比較ではなく、トレンドやプロフィールを比較するための特別な手法が必要である。

14 . サブグループ解析

ある背景因子の特定の層（サブグループ）について、治験薬の効果を検証したい場合もある。しかし、サブグループごとに統計的推論を繰返すことに伴う多重性の問題は、最も解決の困難なもののひとつである。複数の背景因子のそれぞれについてサブグループ解析を繰返した場合には、多項目比較と同様に、背景因子間の相関構造の扱いに絡む未解決の問題があり、第1種の過誤の正確な制御が困難になる。複数の背景因子の組み合わせによるサブグループ解析では、各サブグループでの被験者数の減少に伴う検出力の低下という第2種の過誤の確率の増大も問題になる。検証的な目的でのサブグループ解析では、サブグループごとの十分な被験者数の確保を計画段階から図る必要がある。

性や年齢階層などのサブグループによって治験薬と対照薬の効果の様式が異なるという交互作用は、治験薬の効果を評価する上で極めて重要な情報である。しかし、治験薬と同一系統の作用機序をもつ実対照薬との比較試験においては、治験薬の非臨床試験やそれ以前の臨床試験の成績の十分な検討なくしては、交互作用の予測は困難なことが多い。検証的データ解析では、それ以前の治験薬の効果に関する情報を総合し、薬剤と背景因子間の交互作用に注目して、事前に層別する背景因子を絞り込むことが不可欠である。

15 . 不完全例の分類

例えば、不完全例を試験開始前の問題にかかわる不適格と、試験開始後の試験処置にかかわる中止、脱落、逸脱（違反）に大別して整理する。

a. 不適格

被験者選択基準に合致しない被験者、除外基準に抵触する被験者は不適格とする。対象外の疾患、治験開始後に対象外であることが判明した場合、診断根拠のデータ不備など、対象疾患に関連したものと、年齢、病歴、合併症、観察・測定初期値など被験者の背景特性、重要な背景因子に関連したものがある。

b. 中止

治験担当医師の医学的診断によって、ある被験者の処置を意識的に打切る場合である。必ずしも計画違反とは限らず、完全な1例となる場合もある。実際には、副作用、悪化、治癒、偶発症による治療変更ないし終了、試験処置開始後の不適格の判明などの場合がある。用法、用量の変更、併用療法の実施などを経て中止に至る場合もある。副作用、治癒などの理由で中止となった場合、これらは試験処置の重要な結末であるから、中止の時点で内容を正確に記録しておけば、重要な情報になる。

c.脱落、追跡不能

被験者の転居、多忙、試験への非協力など試験処置と直接に関係のない理由により、観察・測定、検査の継続、評価の実施が不可能となりデータに打切りを生じる場合である。一方、処置と関係した理由、例えば副作用の懸念、対象とした疾患や症候の軽快などにより、服薬違反などを経たりして、結果的にデータの打切りを生じることもある。この場合、追跡調査によって被験者の判断が医学的に妥当であることが確認されたならば、中止、つまり試験処置の結末をみた扱いに変えてもよい場合がある。脱落と追跡不能とはほぼ同義に使われることもあるが、脱落のうち、追跡調査してもその後の情報が全く得られないものを追跡不能とするのが妥当と思われる。中止と脱落をあわせて、未完了例と呼ぶこともあり、データの打切りがあることが特徴である。

d.処置違反

被験者が選択されると試験に組入れられ、比較試験の場合には割付に従った試験処置が開始される。その後、併用薬、併用療法などに関する治験実施計画との不一致や、観察・測定、検査、評価の方法や時期などに関する治験実施計画との不一致が生ずることがあるが、ここではこのような例を、処置違反とする。これらは治療側、治験担当医師側の意識的、無意識的な治験薬の用法、用量の変更、禁止されている併用療法の追加などによるほか、不十分な治験管理や休日、学会と診療予定日との重なりというようなことでも生じうる。

e.処置不遵守

被験者が治験担当医師の指示通りに服薬しなかったり、試験参加を撤回した場合、処置不遵守となる。来院日のずれから、投薬が受けられず、服薬違反と観察・測定時期違反が同時に発生する場合が多い。

単純な服薬忘れ、被験者自身による副作用の懸念、無効の判断あるいは完治の判断などによって不遵守が発生するが、観察・測定期、試行期を設けて、予め非協力的な対象者を除外しておくことも考えられる。しかし、このような方法では計画において想定した被験者集団の内容が変わってしまう可能性がある。

被験者自身による副作用の懸念の場合、追跡調査によって、被験者の単純な思いすぎであることもあるし、思いすぎではなくて臨床的にみても明らかな副作用であるということもある。

治験担当医師の処置違反と被験者側の処置不遵守は、打切りにつながるとは限らず、一部のデータが欠落又は不適切ということが多いが、ある時点から違反が続けば打切りと同義になる。

16 . 不完全例の解析段階での取扱い

比較試験における不完全例にかかわる問題は、その発生の時点が試験の初期であるか後期であるか、理由に関しては処置に直結しているか否か、ないしは治験担当医師の理由によるのか被験者側の理由によるのか、内容的には全般的か部分的かなど様々な観点から検討される。また、被験者の取扱いも、試験の目的、期間、規模、二重盲検法であるか否かなどによって、変わる可能性がある。不完全例のデータの扱いにあたって、もっとも注意すべきことは、試験結果に偏りが介入することである。偏りの程度によっては比較の公平性をねらった無作為割付の意味が壊滅的になる。ついで、計画段階で想定した被験者集団の内容が変化してしまうことが問題にある。単に、不適格、中止、脱落などの例数を数え、それぞれの率を治験薬群と対象薬群とで比べるだけではなしに、より詳細なコメントを用意したり、個々の不完全例が解析でどのように取り扱われたかを詳細に説明する資料を用意する必要がある。

治験実施計画をどの程度厳格に規定するかは、当然その試験の目的、性格に依存しており、解析段階において不完全例のデータを採用するか否かの判断も、この点に密接に関連している。試験の目的ないし性格は、例えば、探索的、検証的と分類されたり、研究的と実践的などと対比されるが、それぞれの立場によってデータの妥当な取扱いは変わってくる。

本来、解析にあたっての方針は、試験が終了してから不完全例の様子をみて定めるといった事後的な性質のものでなくて、事前に定めておくべきものである。不完全例の多寡は、少なくとも定性的にはその試験の質的レベルを示唆する指標となり、多い場合には試験の信頼性が損なわれることになる。

a. 研究的試験と実践的試験：

研究的な試験の場合、臨床試験の場を単純に診察現場に求めるのではなく、できるだけ純粋な等質性を土俵とすべきだとの認識のもとに、例えば改善率を考えたときに、適格な被験者でしかも計画通りに試験処置を完了した被験者だけを厳密に選んで分子、分母とする取扱いが設定される(protocol-compatible)。

一方、実践的な試験の場合、診療現場への適用という観点から不完全例を単純に違反とせず、試験処置を開始した被験者は、処置の長短には無関係に解析対象とし、改善例としての分子になる可能性はなくても分母に加えるべきだという主張もある（intent-to-treat）。

この両極端の立場においては、想定している対象集団の定義、目的が明らかに異なっている。研究的な試験では限定した条件での処置法の比較、薬理的な特徴の探索、説明に力点があり、実践的な試験では比較にとりあげた試験処置のうち、現実の治療にどちらを選ぶかの決定に力点がある。通常の比較臨床試験はおそらくこれらの中間的な性格をもつものも多いであろう。つまり不完全となった時点、理由を吟味し、データの種類に応じて、例えば、有効性については不採用（または“不明”として採用）であるが、安全性、有用性については採用とするといった取扱いがより現実的であろう。しかし、例えば、副作用以外の理由で処置期間が不十分なために有効性が評価できない被験者について安全性のみを評価すると、実際よりも安全性がより高く評価される危険性があり、適切ではない。

b.不適格の内容：

不適格例を特定するに当たり、試験が二重盲検下に実施されているかどうかは大きな影響を与える。非盲検で例えば封筒法を利用した場合など、試験処置が指示通りに実施されないことがある。このような例は、定義上は処置違反ということになるが、事後的にその違反理由を知り、比較に当て偏りが生じないように扱うことは一般に困難である。勿論、年齢制限のわずかなずれと診断の誤りとを一律に取扱うことは不合理であろうが、たとえ診断の誤りによる不適格例であっても有効性、有用性については不採用であるが、安全性については採用できるという場合も少なくない。

腎や肝障害のある被験者には副作用の虞れがあり、あらかじめ除外規定に明記されているにもかかわらず、この種の被験者が試験に組込まれた場合には、実際の副作用の発現の有無とも関連して特別な扱いが必要になるろう。

c.中止、脱落の理由：

副作用の発現、治癒など、中止にいたった理由が明瞭であれば、少なくとも一部のデータは解析に採用される。試験初期の副作用中止では、研究的試験の立場からは有効性評価は無理としても、安全性は勿論、有用性の評価も可能であろう。欧米では有用性の概念が希薄であるためか、実践的試験の立場から、副作用中止に無効の評価を与え、有効性の解析に採用すべきだとする意見もある。また、副作用を生じた罰の意味で、改善率の計算の分母に加えて率を落とすべきだとの意見もある。

処置期間を4週と定め、処置終了後1年の追跡といった場合に、副作用のため2週で中止の例では有効性の評価は確かに無理であろう。しかし、単純に不適格ないし不完全の扱いにすると、この種の例は表面に現れないことになるので、有用性の評価に加えるのが適切である。

被験者が試験処置中に転倒して骨折し、通院不能となり、脱落扱いとなった場合、使用薬剤が中枢抑制作用をもっている、あるいは骨代謝に影響をもっているなどの理由が考えられ、単純な脱落ではないと判断される場合がある。交通事故による打切りなどでも同様の考察が必要となる。

d.逸脱ないし欠測

研究的試験では、一般に処置違反や服薬違反に関してデータはできるだけ採用しないし、観察・測定、検査、評価の時点のずれも常識的な範囲を越えれば問題にする。これに対して実践的試験ではデータの採用範囲はかなり広がる。いずれの場合も、一定の取決めのもとにデータの採用、不採用を決めるのであるが、この際に試験結果に偏りが介入する虞れが常に存在すること、しかもこの偏りは単純に評価できないことに注意しなければならない。

一般的に言えば、不適格、処置違反、処置不遵守などの被験者を除くことで、試験結果のノイズを減少し、感度のよい試験が期待できるが、このことは例数の減少による逆方向の影響を生む場合もある。不採用例が多ければ、出発での無作為化の意義が失われるという強い議論もある。また、対象の均質化により、結論をあてはめるべき母集団の性格はより明確にされるが、一方、実践的にみれば適用範囲がせばめられることにもなる。

17．過度の薬理作用

例えば、降圧薬での低血圧など、治験薬の過度の薬理作用から予測される症候・所見が認められた場合の取扱いを決めておく。

18．フローチャート

フローチャートを使用して、試験の各段階の内容を示すことが望ましい。その際、各段階での不完全例の発生状況及び統計解析の対象とした症例数を記載する。

19．計画違反例

計画の重要事項に違反した被験者又は治験担当医師があれば、そのどちらであってもその事例を記載する。そして可能な範囲でその違反の及ぼす影響も併記する。

20．背景因子の偏りの調整

治療予後と強く関連する重要な背景因子（予後因子）の分布が処置群間で異なっている場合に、予後因子の影響を考慮せずに統計解析を進めると解析結果は歪んだものになる。臨床試験においては、既知及び未知の予後因子の分布が処置群間で均等になるように、無作為割付が可能な限り実施されている。無作為割付は処置群間の比較可能性を確保するための最良の手段であり、多くの場合、予後因子の影響を考慮せずに解析を行っても妥当な比較結果が得られる。しかし、特に症例数が少ない場合には、無作為割付を行ったとしても処置群間に予後因子の偏りが偶然に生じ、薬効差に関する単純な比較の妥当性が懸念される場合もある。予後との関連が強いことが明らかな予後因子の分布が処置群間で異なる場合には、やむを得ないときの手段として、適切な統計的調整を実施することも必要となる。

検証的な臨床試験で予後因子の偏りに関する統計的調整を行う場合、処置群間での主要な予後因子の偏りの点検は薬効差の解析に先立って実施する必要がある。この際、予後因子自体の検定は参考的なものであり、予後因子の偏りが薬効差の解析結果にどのように影響するかという観点から注意深く点検することが重要である。多くの場合、有意確率が15%前後といった背景因子の不均衡についても注意深い検討が必要である。薬効差の統計解析が一旦なされてから、後知恵によって予後因子を選択して統計的調整を行った場合には、結果の公平性が保証できず、治験薬の薬効の検証とはならない。

解析段階での予後因子の偏りの統計的調整としては種々の方法があるが、層別化による方法が最も基本的な方法といえる。層別化による方法は、偏りのあった予後因子について層別集計し、各層に共通の薬効差があると仮定してこの共通薬効差に関する統計的推論を行うものである。各層に共通の薬効差が存在するという統計モデルの妥当性については各層別集計において確認することが可能であり、臨床医学的な理解も直接的に得られる。統計的調整の具体的方法としては2値データについては Mantel-Haenszel 法などがあり、3分類以上の順序分類データについては拡張 Mantel 法などが提案されている。共通薬効差という前提の点検（交互作用の点検）に、2値データについては、例えば、Breslow-Day 検定などを用いることができる。計量データについても、種々の方法による統計的調整が可能であるが、交互作用の点検と共に、パラメトリックな手法を用いた場合にはその統計的方法の基礎にある諸仮定が満たされているかどうかの検討が必要である。

2 1 . 検定結果の記載

例えば、2標本Wilcoxon検定（Mann-WhitneyのU検定）を用いた場合には、z値、p値及び各群の被験者数を記載する。2標本t検定を用いた場合には、t値、自由度、p値及び各群の被験者数、平均値、分散、合成分散の推定値を記載する。

2 2 . 同等性の検証

薬効評価に当っては、プラセボよりも有意に優れており、市販の実対照薬と同等又はそれ以上であることが要求されている。ここで、“有意差のない”ことをもって直ちに“同等”であるとするのは明らかに間違いであり、有意差のないことは統計的に“同等”を保証するものではない。

臨床的同等性を立証するには、例えば、有効率の場合には次のような統計的方法が考えられる。ただし、以下の は、治験薬と実対照薬の臨床的に許容できると考えられる有効率の差を予め設定しておいたものであり、疾患、薬効群により具体的な数値は異なるものの、例えば10%が一つの目安となる。

治験薬と実対照薬の有効率の差の信頼区間（90%信頼区間）が - より小さい領域を含まない場合には、臨床的に同等と評価する。

b .

「治験薬の有効率が実対照薬の有効率に比して 以上劣る」という帰無仮説に対する片側仮説検定（有意水準 5 %）を用い、これが棄却されたときに臨床的に同等と判断する。

これらはいずれも 以上に劣ることはないことの証明をもって同等性を保証しようとするものである。