

医薬審 第1047号
平成10年11月30日

各都道府県衛生主管部(局)長 殿

厚生省医薬安全局審査管理課長

「臨床試験のための統計的原則」について

近年、優れた新医薬品の地球的規模での研究開発の促進と患者への迅速な提供を図るため、承認審査資料の国際的ハーモナイゼーション推進の必要性が指摘されている。このような要請に応えるため、日・米・EU三極医薬品規制調和国際会議(ICH)が組織され、品質、安全性及び有効性の3分野でハーモナイゼーションの促進を図るための活動が行われている。

別添の「臨床試験のための統計的原則」(以下「本ガイドライン」という。)は、ICHにおける合意に基づき、臨床試験における統計的原則について記載したものであり、臨床試験から得られる結果の偏りを最小にし、精度を最大にすることを目標としている。特に、計画段階から試験統計家が参加すること、治験実施計画書の作成に当たっては解析方法等について妥当性も含め事前明記すること等が強調されており、多施設共同試験における施設の捉え方及び施設当たりの症例数の設定に関する考え方、総合評価変数を用いる際の留意点等についても記載されている。また、検証的位置づけの試験を行う際の有意水準(第一種の過誤)については従来明確にされていなかったが、規制上の観点から、本ガイドラインの施行に伴い、原則として片側仮説を検証する場合は2.5%、両側仮説の場合は5%とすることとした。これらについては、ガイドラインの該当箇所及び関係する質疑応答を参照されたい。

本ガイドラインは、本通知の日以降施行し、これに伴い、「臨床試験の統計解析に関するガイドライン(平成4年3月4日薬新薬第20号)」(以下「旧ガイドライン」という。)は廃止する。ただし、治験実施計画書の作成にかかる事項については、既に治験実施計画書が作成され、実施されている臨床試験もあることから、このような場合に配慮し、臨床試験の実施に先立って治験実施計画書が確定される日が平成10年12月31日以前の場合は、被験者数の決定方法も含め旧ガイドラインを参考とした事項があっても差し支えないが、そのような場合であっても、治験実施計画書の改訂又は統計解析計画書の作成を含め、本ガイドラインの趣旨に添って適切と考えられる事項については可能な限り適用することとされたい。

以上の点を御了知の上、貴管下関係者に対し周知方ご配慮願いたい。

臨床試験のための統計的原則

臨床試験のための統計的原則

目次

I. はじめに	1
1.1 背景と目的	1
1.2 適用範囲と方向性	2
II. 臨床開発全体を通して考慮すべきこと	3
2.1 試験の性格	3
2.1.1 開発計画	3
2.1.2 検証的試験	4
2.1.3 探索的試験	4
2.2 試験で扱う範囲	5
2.2.1 対象集団	5
2.2.2 主要変数と副次変数	5
2.2.3 合成変数	6
2.2.4 総合評価変数	6
2.2.5 複数の主要変数	7
2.2.6 代替変数	8
2.2.7 カテゴリ化した変数	8
2.3 偏りを回避するための計画上の技法	8
2.3.1 盲検化	9
2.3.2 ランダム化(無作為化)	10
III. 試験計画上で考慮すべきこと	12
3.1 試験計画の構成	12
3.1.1 並行群間比較計画	12
3.1.2 クロスオーバー計画	12
3.1.3 要因計画	13
3.2 多施設共同試験	14
3.3 比較の型式	16
3.3.1 優越性を示すための試験	16
3.3.2 同等性又は非劣性を示すための試験	16
3.3.3 用量-反応関係を示すための試験	17
3.4 逐次群計画	18
3.5 必要な被験者数	18
3.6 データの獲得と処理	19
IV. 試験実施上で考慮すべきこと	20
4.1 試験モニタリングと中間解析	20
4.2 選択基準と除外基準の変更	20
4.3 集積率	21
4.4 必要な被験者数の調整	21
4.5 中間解析と早期中止	21

4.6 独立データモニタリング委員会の役割	23
V. データ解析上で考慮すべきこと.....	23
5.1 解析の事前明記.....	23
5.2 解析対象集団.....	24
5.2.1 最大の解析対象集団	24
5.2.2 治験実施計画書に適合した対象集団	26
5.2.3 二つの異なる解析対象集団の役割	26
5.3 欠測値と外れ値.....	27
5.4 データ変換	27
5.5 推定、信頼区間及び仮説検定.....	28
5.6 有意水準と信頼水準の調整.....	28
5.7 部分集団、交互作用及び共変量.....	29
5.8 データの完全性の維持とコンピュータソフトウェアの妥当性.....	29
VI. 安全性及び忍容性評価.....	30
6.1 評価の範囲	30
6.2 変数の選択とデータ収集	30
6.3 評価される被験者集団とデータの提示	30
6.4 統計的評価	31
6.5 統合した要約.....	32
VII. 報告.....	32
7.1 評価と報告	32
7.2 臨床データベースの要約	34
7.2.1 有効性データ	34
7.2.2 安全性データ	35
用語集.....	35

臨床試験のための統計的原則

I. はじめに

1.1 背景と目的

医薬品の有効性及び安全性は、1996年5月1日にICHが採択した「医薬品の臨床試験の実施に関する基準のガイドライン（ICH E6）」に基づき、日本における臨床試験の実施基準として定められた「医薬品の臨床試験の実施の基準に関する省令」（平成9年3月27日厚生省令第28号）に従って実施される臨床試験によって示されるべきである。上記のICHガイドライン及び日本における関連通知において、臨床試験の計画と解析における統計学の役割は欠くことのできないものと認められている。医薬品承認の過程及びヘルスケア一般において臨床研究が重要な役割を果たしていることにより、臨床試験領域での統計的研究が増大していることから、臨床試験に関連した統計的な問題に関する簡潔なガイドラインが必要とされている。本ガイドラインは、日米EUの3極間で、主として、承認申請のための臨床試験（治験）に適用する統計的方法論の原則の調和を進めるために書かれている。

本ガイドラインは出発点として、CPMP(欧州医薬品委員会)ガイダンス文書'Biostatistical Methodology in Clinical Trials in Applications for Marketing Authorisations for Medicinal Products' 「医薬品市販認可のための申請に用いる臨床試験における生物統計方法論（1994年12月）」を利用した。また、厚生省(日本)の「臨床試験の統計解析に関するガイドライン（1992年3月）」、U.S. Food and Drug Administration (米国食品医薬品庁)の'Guideline for the Format and Content of the Clinical and Statistical Sections of a New Drug Application' 「新医薬品申請書臨床の部及び統計の部の書式と内容に関するガイドライン（1988年7月）」も参考にした。統計的原則と方法論に関連したトピックは、他のICHガイドライン及びそれに基づいて作成された日本でのガイドライン等(作成中のものを含む)、特に以下のものにも記載されている。関連する内容を含むガイドラインを、本文の各節で引用する。

- E1A: 「致命的でない疾患に対し長期間の投与が想定される新医薬品の治験段階において安全性を評価するために必要な症例数と投与期間について(平成7年5月24日薬審第592号)」(以下引用に際しては、ICHのトピックに合わせ「ICH E1A」という。)
- E2A: 「治験中に得られる安全性情報の取り扱いについて(平成7年3月20日薬審第227号)」(以下「ICH E2A」という。)
- E2B: 「個別症例安全性報告の伝達のためのデータ項目(以下「ICH E2B」という。)」
- E2C: 「市販医薬品に関する定期的安全性最新報告(PSUR)(平成9年3月27日薬安第32号)」(以下「ICH E2C」という。)
- E3: 「治験の総括報告書の構成と内容に関するガイドライン(平成8年5月1日薬審第335号)」(以下「ICH E3」という。)
- E4: 「新医薬品の承認に必要な用量-反応関係の検討のための指針(平成6年7月25日薬審第494号)」(以下「ICH E4」という。)
- E5: 「外国臨床データを受け入れる際に考慮すべき民族的要因についての指針(平成10年8月11日医薬審第672号)」(以下「ICH E5」という。)

- E6: 「医薬品の臨床試験の実施の基準に関する省令(平成9年3月27日厚生省令第28号)」、及び「医薬品の臨床試験の実施の基準に関する省令の施行について(平成9年3月27日薬発第430号薬務局長通知)」(以下「ICH E6」という。)
- E7: 「高齢者に使用される医薬品の臨床評価法に関するガイドライン(平成5年12月2日薬新薬発第104号)」(以下「ICH E7」という。)
- E8: 「臨床試験の一般指針(平成10年4月21日医薬審第380号)」(以下「ICH E8」という。)
- E10: 「臨床試験における対照群の選択」(以下「ICH E10」という。)
- M1: 「規制目的のための医学用語の標準化」(以下「ICH M1」という。)
- M3: 「医薬品の臨床試験のための非臨床安全性試験の実施時期についてのガイドライン(平成10年11月13日医薬審第1019号)」(以下「ICH M3」という。)

本ガイドラインは、臨床開発のあらゆる場面で、治験依頼者が被験薬に関する臨床試験の計画、実施、解析及び評価を行う場合の方向づけを目的としている。また、本ガイドラインは、主として開発の後期の相の臨床試験について、承認申請書添付資料概要(以下「資料概要」という。)の作成を任せられた専門家、又は有効性及び安全性の証拠を評価することを任せられた専門家にも役立つであろう。

1.2 適用範囲と方向性

本ガイドラインは、統計的原則に焦点を合わせており、個々の統計的な手続き又は手法の使い方を扱うものではない。原則が正しく適用されていることを保証するために個々の手続きを積み重ねることは、治験依頼者の責任である。複数の臨床試験にわたるデータの統合も論じるが、本ガイドラインの主要な点ではない。データマネジメント又は臨床試験のモニタリング活動に関連した原則と手続きのうち、他のICHガイドラインで扱っているものはここでは触れない。

本ガイドラインは、科学の広い分野の人々から関心を持たれるべきものである。しかし、臨床試験に関連したすべての統計的業務に対する実際の責任は、適切な資格と経験のある統計家が果たすことが前提となっており、そのことは ICH E6にも示されている。試験統計家(用語集参照)の役割と責任は、医薬品開発を支える臨床試験に統計的原則が適切に適用されていることを、他の臨床試験専門家と共同して保証することである。したがって、試験統計家は本ガイドラインに明確に述べられた原則を実行するために十分な理論又は実地の教育及び経験を併せ持つべきである。

承認申請に含まれる個々の臨床試験の計画と実施に関するすべての重要事項についての詳細及び臨床試験において使用する統計解析の主要な特徴は、試験開始前に作成された治験実施計画書(プロトコル)に明記すべきである。治験実施計画書中の手続きの遵守状況及び主要解析の事前での計画状況がどの程度であったかが試験の最終結果と結論の信頼性に寄与することになる。治験実施計画書及びその作成後の改訂は、試験統計家を含む責任者全員から承認を受けるべきである。試験統計家は、治験実施計画書及びそのいかなる修正もが、すべての重要な統計的問題を、必要ならば専門用語を用いて、明確かつ正確に扱っていることを保証すべきである。

本ガイドラインにまとめられている原則は、主として、有効性の検証的試験として実施されることの多い開発の後期の相の臨床試験に当てはまるものである。検証的試験では、主要変数として有効性以外にも安全性に関する変数（例えば、有害事象、臨床検査変数又は心電図の読み取り結果）、薬力学変数又は薬物動態変数（検証的な生物学的同等性試験の場合）を採用してもよい。更に、検証的な知見の一部は複数の試験を統合したデータから導かれることもあり、このような場合でも本ガイドライン中の原則の一部は適用できる。最後に、医薬品開発の初期の相は主として探索的な性質の臨床試験からなるが、統計的原則はこれらの臨床試験にも当てはまる。したがって、本ガイドラインの趣旨は、可能な限り臨床開発のすべての相において適用されるべきである。

本ガイドラインに述べられている原則の多くは、偏り（用語集参照）を最小にし、精度を最大にすることを目的としている。本ガイドラインでは、「偏り（バイアス）」という用語を、「臨床試験の計画、実施、解析及び結果の解釈と関連した因子の影響により、試験治療の効果（用語集参照）の推定値と真の値に系統的な差が生じること」という意味で用いる。偏りを低く抑えるためには、偏りの潜在的な原因を可能な限り明らかにすることが重要である。偏りの存在により、臨床試験から妥当性のある結論を導くことが困難になるおそれがある。

偏りの原因が試験実施計画にある場合がある。例えば試験治療の割付が不適切で、リスクの低い患者が一方の試験治療に系統的に割付けられる場合がそうである。偏りの原因は、臨床試験の実施や解析の際に生じることもある。例えば、試験実施計画書違反及び個々の被験者の結果を知った後で被験者を解析から除外することは偏りの原因となり得るものであり、試験治療効果の正確な評価に影響を及ぼすおそれがある。偏りが起こる理由は明確にはとらえられない場合があり、偏りの影響は直接測定できないため、試験の結果と主要な結論の安定性を評価することは重要である。安定性（ロバストネス）とは、データ、仮定及び解析方針についての様々な制限に対して全体の結論がどの程度変わり易いかに関連した概念である。安定性は、異なる仮定又は異なる解析方針に基づいて解析を行った場合でも、試験治療の効果と試験の主要な結論は大きく影響されないことを意味している。試験治療の効果と試験治療の比較における不確実さに関する統計的指標の解釈には、 p 値、信頼区間又は推測に偏りが与える影響を考慮に入れるべきである。

臨床試験の計画と解析においては、頻度論的立場からの統計手法に基づく方法が主流になっていることから、本ガイドラインは仮説検定や信頼区間を議論する場合、主として頻度論的手法（用語集参照）を念頭に置いている。これは、他の方法が適切でないと主張するものではない。ベイズ流の手法や他の手法の使用も、それらの使用の理由が明らかであり、異なる仮定の下でも結果として得られる結論が十分に安定している場合には検討することができる。

II. 臨床開発全体を通して考慮すべきこと

2.1 試験の性格

2.1.1 開発計画

新しい医薬品を臨床開発する過程全体を通しての目的は、臨床上の利益との兼ね合いで

リスクが許容できる限度において、医薬品が安全性と有効性を兼ね備えている用量範囲と使用スケジュールが存在するかどうかを知ることにある。医薬品から利益を受ける対象となる集団を明確にし、医薬品の適応症も定める必要がある。

これら臨床開発全体としての目的を達成するためには、通常それぞれが特定の目的を持った臨床試験の順序だったプログラムが必要である（ICH E8 参照）。このプログラムは、判断をくだす適切な時点と、知識の蓄積に伴う修正を認める柔軟性を持つ、一つ又は一連の臨床開発計画として明示されるべきである。承認申請では、臨床開発計画の趣旨及び個々の試験がどのように寄与するかについて明確に述べるべきである。試験プログラム全体から得られる証拠の解釈と評価は、個々の試験からの証拠を総合する過程を伴うものである（7.2節参照）。これは、医学用語の辞書、主要な測定の見解と時期、試験実施計画書からの逸脱の取り扱い、といった試験のいくつかの特徴について、共通の基準を採用することで容易になる。医学上の問題に複数の試験によって対応する場合、統計的要約、総括（overview）又はメタアナリシス（用語集参照）が有益であろう。可能であれば、このように対応することは臨床開発計画の中で明確にすべきであり、そうすることによって関連する試験が把握され、試験間で共通にすべき計画上的特徴を前もって特定することができる。共通の臨床開発計画によって実施される複数の試験に影響することが予測されるその他の主要な統計的問題があるならば、臨床開発計画の中に述べるべきである。

2.1.2 検証的試験

検証的試験とは、事前に定められた仮説を評価するための、適切に計画・実施された比較試験である。一般に、検証的試験は有効性又は安全性の確固たる証拠を提示するために必要である。検証的試験では、関心のある重要な仮説は試験の主要な目的から直接導かれ、常に事前に設定され、試験完了後に検証される。検証的試験では、関心のある試験治療に由来する効果の大きさを十分な精度で推定すること、及びそれらの効果を臨床的な意義と関連付けることも同様に重要である。

検証的試験は、承認に関わる主張の裏付けとなる確固たる証拠を提示することを目的としているため、試験実施計画書と標準業務手順書に従うことは特に重要である。このため、やむをえない変更については説明を文書として残すべきであり、変更の影響も検討しておくべきである。個々の検証的試験の計画及び予定している解析の主要な特徴等の重要な統計的側面について、それを正当とする理由は、試験実施計画書に記述すべきである。個々の試験は限られた少数の問題のみを扱うべきである。

承認に関わる主張の裏付けとなる確固たる証拠としては、被験薬が臨床上的利益を持つことを、検証的試験の結果で示す必要がある。したがって検証的試験は、有効性又は安全性の主張に関連した個々の主要な臨床的問題に、明確かつ確定的な回答を与えるのに十分なものであるべきである。更に、意図している患者集団への一般化（用語集参照）の根拠が理解でき、説明できることは重要である。このことによっても、必要な施設（センター）の数と型式（例えば、専門医か一般臨床医か）及び試験の数と型式が影響を受けるであろう。検証的試験の結果は、条件によらず安定しているべきである。場合によっては、一つの検証的試験からの証拠だけで十分であることもありうる。

2.1.3 探索的試験

検証的試験の正当性の根拠と計画は、事前に実施された一連の探索的研究の臨床成績に

ほとんど常に依存している。すべての臨床試験と同様に、これらの探索的研究も明確で精密な目的を持つべきである。しかし、検証的試験とは対照的に探索的試験の目的は、必ずしも事前に設定した仮説の単純な検定に帰着するわけではない。更に、探索的試験では、結果の蓄積に応じて変更が可能となるような、より柔軟な方法を必要とするときもある。その解析は探索的なデータ解析を伴ってもよい。すなわち仮説検定を行う場合、その仮説の選択は得られたデータに基づいて行うことが有り得る。このような試験は、有効性を証明する証拠全体には貢献するといってもよいが、その正式な証明の根拠にはならない。

個々のいかなる試験も、検証的側面と探索的側面の両方を持つものであるといってもよい。例えば、検証的試験であってもほとんどの場合、結果を説明する又は裏付けする根拠を与えるため、また後の研究の仮説を引き出すためにデータの探索的解析も行われる。治験実施計画書には、各試験について検証的な証明として用いられる側面と、探索的解析のためにデータを提供する側面とを、明確に区別しておくべきである。

2.2 試験で扱う範囲

2.2.1 対象集団

医薬品開発の初期の相では、臨床試験の被験者の選択は、関心のある特定の臨床効果が観察できる見込みを最大にしたいという要求に大きく影響されるであろう。したがって、被験者は、最終的にその薬が使用される全患者集団のうちごく限られた部分集団から選ばれる可能性がある。しかし、検証的試験が実施される時までには、被験者を目標集団に十分近いものにしておくべきである。したがって、検証的試験では一般に、試験治療の効果を精度よく推定するために十分な均質性を維持する一方、目標集団の範囲内でできるだけ広い患者を対象とするような選択基準及び除外基準を定めることが有益である。単一の臨床試験の被験者をもって、将来の使用者を完全に代表させることは期待できない。なぜならば、地理的な位置、試験が実施された時期、個々の治験責任医師や診療所で行われる実地の医療内容などが影響する可能性があるからである。しかし、これらの要因の影響は可能な限り小さくすべきであり、その上で試験の結果を解釈する際にその影響を議論しておくべきである。

2.2.2 主要変数と副次変数

主要変数（「目標」変数、主要評価項目ともいう）は、試験の主要な目的に直結した臨床的に最も適切で説得力のある証拠を与えうる変数であるべきである。主要変数は通常ただ一つにすべきである。ほとんどの場合、検証的試験の主要な目的は有効性に関して科学的に説得力のある証拠を提示することにあるため、主要変数は、通常有効性に関する変数となる。安全性及び忍容性は常に重要な問題であり、ときには主要変数となりうるものである。生活の質（QOL）及び保健経済に関する測定値も、主要変数となる可能性がある。主要変数の選択には、開発に関連した研究領域で一般に認められている規範と基準を反映させるべきである。先行研究又は公表論文で使用された実績のある、信頼性及び妥当性の確立した変数を使用することが薦められる。主要変数は、選択基準と除外基準によって規定される患者集団において、臨床的に適切で重要な治療上の利益に関する妥当で信頼のかける指標であることが十分に証拠づけられているべきである。被験者数の見積もりに用いる変数は、通常は主要変数であるべきである（3.5節参照）。

多くの場合、被験者の結果を評価する方法はあまり単純ではないであろうから、主要変

数は慎重に定義する必要がある。例えば、主要変数を明確に定めずにただ死亡とするのでは不十分である。死亡の評価にも、定められた時点における生存割合の比較、又は特定期間の生存時間分布全体の比較がある。別の例として、事象が繰り返し起こる場合が挙げられる。この場合試験治療の効果の指標としては、単純な二分類の変数（一定期間に一度でも起きたかどうか）、初発までの時間、生起率（単位観察時間あたりの事象数）などがある。慢性疾患のための治療の研究で、経時的に機能の状態を評価する場合も、主要変数の選択に関して別の問題が生じる。可能な対処法としては、観察期間の最初と最後になされた評価の比較、全期間を通じたすべての評価から求めた傾きの比較、定めた閾値を超える若しくは下回る被験者の割合の比較、又は繰り返し測定データのための方法に基づいた比較といった多くのものがある。事後的に定義することから生じる多重性の問題を回避するために、統計解析で用いる主要変数の正確な定義を治験実施計画書に明記することは重要である。更に、選択した特定の主要変数の臨床的な適切さ及びその測定手順の妥当性は、通常治験実施計画書に記載し正当性を示すことが必要である。

主要変数は、その変数を選択した理由とともに治験実施計画書に明記すべきである。割付が明らかになった後に主要変数を定義し直すことは、ほとんどの場合許容できない。なぜならば、それによって生じる偏りの評価が難しいからである。主要な目的として定めた臨床効果を二通り以上の方法で測定しようとしている場合、臨床的な適切さ、重要性、客観性、その他関連する特徴に基づいて、治験実施計画書にはできる限り一つの測定値を主要変数として指定すべきである。

副次変数は、主要な目的に関連した補足的な測定値又は副次目的に関連した効果の測定値のどちらかである。治験実施計画書において副次変数を事前に定義し、試験結果の解釈の際に副次変数が果たす相対的な重要性と役割を説明することも重要である。副次変数の数は、試験で答えるべき限られた少数の問題と関連して制限すべきである。

2.2.3 合成変数

主要な目的に関する複数の測定値の中から、主要変数として一つを選ぶことができない場合、それに代わる有用な戦略は、事前に定められたアルゴリズムを用いて、複数の測定値を単一の変数、つまり「合成」変数に統合又は結合することである。実際、主要変数を複数の臨床測定値の組み合わせとして定める例がみられる（例えば、関節炎、精神障害その他で用いられる評価尺度）。この方法は、多重性の問題に対処するに当たり、第一種の過誤の調整を必要としない。複数の測定値を結合する方法は、治験実施計画書に明記すべきであり、得られた尺度について、臨床的な利益の大きさを適切にあらわしているかどうかという観点から説明がなされるべきである。合成変数が主要変数として用いられる場合、合成変数の成分に臨床的意義がありかつ妥当性が示されているならば、その成分を個別に解析することがある。評価尺度を主要変数として用いる場合、内容的妥当性（用語集参照）、評価者内信頼性（用語集参照）及び評価者間信頼性（用語集参照）並びに疾病重症度の変化を検出するための反応性といった特性を説明しておくことは、特に重要である。

2.2.4 総合評価変数

場合によっては、試験治療の全体的な安全性、全体的な有効性、全体的な有用性を測定するために、「総合評価」変数（用語集参照）が作られることもある。この種の変数は、被験者の状態又はその変化についての客観的な変数と治験責任(分担)医師の全体的な印象を統合した、通常は順序カテゴリの評価尺度となる。全体的な有効性の総合評価は、神経科、

精神科など一部の治療領域で確立されている。

総合評価変数は、一般的に主観的な要素を併せ持っている。総合評価変数を主要変数又は副次変数として使用する際は、尺度に関する以下の事項についての詳細を治験実施計画書に記述する必要がある：

- 1) 試験の主要な目的に対する尺度の適切さ
- 2) 尺度の妥当性及び信頼性の根拠
- 3) 個々の被験者を尺度のカテゴリの内の一つに判定するための、その被験者から集められたデータの利用法
- 4) 欠測データがある被験者を尺度のカテゴリの内の一つに判定するための方法又は被験者を評価する方法

総合評価を実施する際、治験責任医師によって考慮されている客観的な変数があれば、それらの客観的な変数は主要変数又は少なくとも重要な副次変数として追加することを検討すべきである。

有用性の総合評価は、医薬品の使用による利益とリスクの両方の要素を統合するもので、医薬品の使用を決定するために、その使用による利益とリスクの比較検討をしなければならぬ臨床医の意思決定過程を反映したものである。有用性の総合評価尺度の問題点の一つは、二つの試験治療が有益な効果と有害な作用に関する全く異なるプロファイルを持っているにもかかわらず、有用性の総合評価を用いることで同等と示す結果を導く場合があることである。例えば、試験治療の有用性総合評価が他方の試験治療に対して同等である、又は優っているとの結果であっても、それは有害な作用が少ないだけで、試験治療には有効性がほとんど又は全くないという事実を意味しているおそれがある。したがって、有用性の総合評価を主要変数とすることは薦められない。有用性の総合評価を主要変数とする場合には、有用性の総合評価に用いた特定の有効性及び安全性の結果を主要変数として別個に追加して考慮することが重要である。

2.2.5 複数の主要変数

治療の及ぼす効果の範囲を一つ一つの変数（又は変数の組）が包含するような、複数の主要変数を用いることが望ましい場合がある。この型式の証拠を解釈する方法は、事前に慎重に説明しておくべきである。試験の目的を達成するために必要と考えられるのは、複数の変数のいずれかに対する効果なのか、ある一定数の変数に対する効果なのか、又は変数すべてに対する効果なのかを明らかにすべきである。一つ又は複数の主要な仮説と関心のあるパラメータ（例えば、平均、割合、分布の状況）は、指定した複数の主要変数に関して明確に決められているべきであり、また統計的推測の方法も明確に述べるべきである。このとき、多重性の問題が起こり得るため、治験実施計画書では、第一種の過誤に与える影響を説明し（5.6節参照）、それを制御する方法を定めるべきである。第一種の過誤への影響を評価するためには、提案した複数の主要変数間の内部相関の程度を考慮する必要がある。指定した主要変数のすべてにおいて有効性を示すことが試験の目的である場合、第一種の過誤を調整する必要はないが、第二種の過誤及び必要な被験者数への影響は慎重

に考慮すべきである。

2.2.6 代替変数

実際の臨床的有効性を観察することにより被験者の臨床的な利益を直接評価することが実際的でない場合には、間接的な基準（代替変数、用語集参照）を考慮することができる。代替変数は、それが臨床的利益の信頼できる予測因子であると信じられている多くの領域において、一般的に容認されたものとして用いられている。代替変数を提案し導入する際には、大きな問題が二つある。一つめは、代替変数が関心のある臨床結果の真の予測因子ではないおそれがあることである。例えば、代替変数はある特定の薬理作用と関連した試験治療の作用を測定しているだけで、肯定的であろうと否定的であろうと、試験治療の作用範囲と最終的な効果の範囲に関する完全な情報はもたらさないおそれがある。提案された代替変数においては非常に有効であることを示している試験治療が、結局は被験者にとって臨床上有害であると示された例は数多い。それとは逆に、提案された代替変数には何の影響もないが、臨床的には利益をもたらした試験治療の例もある。二つめは、提案された代替変数が、有害作用に対して直接比較考量することのできる臨床的利益の定量的な指標とは必ずしもならないことである。代替変数の妥当性を確認する統計的基準は提案されているが、その基準を使用した経験は限られている。実際には、代替性の証拠の強さは、(i) 代替変数と臨床的結果の関連の生物学的合理性、(ii) 代替変数が臨床的結果の予後を予測する上で有益であると疫学研究によって示されていること及び (iii) 試験治療の代替変数に対する効果が臨床的效果と対応しているという臨床試験の結果、に依存している。ある医薬品における臨床的変数と代替変数との関係は、同じ疾患の治療に用いる医薬品であっても、作用機序の異なる医薬品について当てはまるとは限らない。

2.2.7 カテゴリ化した変数

連続変数若しくは順序変数の二分化又はその他のカテゴリ化を行うことが、望ましい場合もある。「成功」又は「反応あり」という基準は二分化のよくある例であり、例えば、連続変数では（基準となる時点での値から）何パーセント以上の改善か、又は順序評価尺度であれば、ある閾値レベル（例えば「良好」）以上に分類されるか、という観点から二分化の方法を正確に記述することが要求される。拡張期血圧が90mmHgを下回るまで低下するかどうかを基準にして二分化することは、その例である。カテゴリ化が最も有用なのは、それが明確な臨床的意味を持つ場合である。試験の結果を知った後でカテゴリ化の基準を決めることは偏りを生じやすいため、カテゴリ化の判定基準は前もって定め、治験実施計画書中に明記すべきである。カテゴリ化は一般に情報の損失となるため、結果として解析での検出力の低下を招く。被験者数の計算では、この点を考慮すべきである。

2.3 偏りを回避するための計画上の技法

臨床試験で偏りを回避するための最も重要な計画上の技法は、盲検化及びランダム化（無作為化）であり、これらは承認申請に利用することを目的とするほとんどの比較臨床試験で標準的に採用すべきである。そのような試験は、ほとんどの場合二重盲検法によるものであり、治験薬を適切なランダム割付表に従って事前に箱詰めし、試験の実施に係る者全てが個々の被験者に割付けられた試験治療を、その試験治療のコードのみであっても、知ることのないよう被験者番号と治療期間のみを表示して治験実施施設に供給するというものである。2.3.1節全体と2.3.2節のほとんどでは、上記の方法を前提とし、2.3.2節最後にその他の方法を述べる。

試験の実施に伴って発生する十分な解析が損なわれる可能性のあるあらゆる変則的な事例について、予想されるもの全て(様々な種類の治験実施計画書違反、試験治療の中止及び欠測値など)の発生頻度を最小にするための手段を治験実施計画書に明記することにより、計画段階で偏りを減じることにも可能である。治験実施計画書では、そういった問題の発生頻度を減じる方法と、データ解析において起こる問題の対処法の両方を考慮すべきである。

2.3.1 盲検化

盲検化又はマスク化は、臨床試験の実施及び解釈における意識的、無意識的な偏りの発生を制限するために行われる。割付けられた試験治療を知ることが、被験者の募集と割付、それに引き続き行われるケア、被験者の試験治療に対する態度、評価項目の評価、試験治療を中止した被験者の取り扱い、解析からのデータの除外等に影響を及ぼし、偏りを発生させるためである。盲検化の本質的な目的は、割付けられた試験治療を知ることにより偏りが生じる可能性のある間は、試験治療が同定されるのを防ぐことである。

二重盲検試験は、被験者並びに被験者の試験治療又は臨床評価を行う治験責任(分担)医師及び治験依頼者のスタッフのすべての者が被験者に割付けられた試験治療を知ることができないものである。これには、被験者の適格性の判断、評価項目の評価及び治験実施計画書遵守状況の評価に関わる者すべてが含まれる。この盲検化のレベルは治験実施中維持され、許容できる質にまでデータクリーニングが行われてはじめて、適切な関係者に割付が明らかにされる。被験者に対し試験治療又は臨床評価を行っていない治験依頼者のスタッフ(例えば、生体試料分析者、監査担当者、重大な有害事象報告にたずさわる者)に試験治療コードの割付を明らかにする必要性が生じた場合のために、治験依頼者は、試験治療コードの不適切な開示を防ぐための標準業務手順書を持つべきである。単盲検試験とは、治験責任医師若しくはそのスタッフのどちらか又は両方が割付けられた試験治療を知っているが、被験者が知ることはない試験であり、逆の場合もありうる。非盲検試験とは、どの試験治療が割付けられたかが全員に知られている試験である。この中で二重盲検試験が最適な方法である。二重盲検試験においては、試験中実施される複数の試験治療が、実施前にも実施中にも区別できないこと(外見、味、その他)及び全試験期間中盲検を適切に維持することが必要である。

二重盲検という理想の実現が困難な場合も生じる可能性がある。比較する試験治療が、例えば外科的療法と薬物療法のように、完全に異なる性質を持つ場合である。治験薬剤型が異なる場合もある。カプセルの使用によって識別不能にできたとしても、剤型の変更が薬物動態と薬力学の性質のどちらか又は両方を変える可能性もあるため、この場合は剤型間の生物学的同等性を立証する必要がある。二つの治験薬の毎日の服薬パターンが異なる場合もある。こういった状況で二重盲検の状態を達成する一つの方法は、「ダブルダミー」(用語集参照)技法を用いることである。しかし、この技法は、被験者の服薬に対する動機づけを低下させ服薬遵守に悪影響を与えるような通常はありえない服薬計画を強いる場合がある。また、例えば二重盲検を保つために偽の手術を必要とする場合のように、倫理的問題により、ダブルダミーの使用が妨げられることがある。それでも、これらの問題を克服するための十分な努力をすべきである。

明らかに試験治療に由来する効果により、一部の臨床試験では部分的に二重盲検が破れるおそれがある。そのような場合、ある種の検査結果(例えば、臨床検査項目の一部)を治験責任(分担)医師及び関連する治験依頼者スタッフが知ることのないようにすることに

より、盲検性が改善されるであろう。試験治療特有の又は特定できるような効果により、個々の被験者の割付が明らかになるおそれのある試験では、後述する非盲検試験で偏りを最小にする手法と同様の手法を検討すべきである。

二重盲検試験が実施できない場合は、次に単盲検を選択することを検討すべきである。非盲検試験のみが实际的に又は倫理的に可能な場合もある。単盲検試験と非盲検試験は二重盲検試験に比べて柔軟性があるが、次の試験治療が何であるか治験責任(分担)医師が知ることにより、被験者の登録の決定に影響を与えないようにすることが重要である。登録の決定は、常に割付けられる試験治療を知る以前になされるべきである。これら単盲検又は非盲検試験では、試験治療のランダム割付を管理するために、電話による割付のような一カ所でランダム化を行う方法を検討すべきである。更に、臨床評価は被験者の処置を行っておらず、試験治療が盲検化された状態の医療スタッフによってなされるべきである。単盲検試験又は非盲検試験では、様々な既知の偏りの原因を最小にするために、あらゆる努力をなすべきであり、主要変数は可能な限り客観的にすべきである。採用した盲検化の程度について、それを選択した理由は、偏りを減じるために用いた他の手段による処置とともに、治験実施計画書に述べるべきである。例えば治験依頼者は、解析のためにデータベースを公開する前のデータベースクリーニングの段階では、試験治療コードへのアクセスが適切に制限されることを保証するために十分な標準業務手順書を用意すべきである。

(被験者個人について)割付を明らかにすることは、割付けられた試験治療を知ることが被験者のケアのため本質的であると主治医が考える場合に限り検討すべきである。故意であろうとなかろうと、割付を明らかにした場合は、開示の理由にかかわらず、治験終了時に報告し、説明すべきである。割付けられた試験治療を明らかにするための手続きと時期は記録しておくべきである。

本ガイドラインでは、試験完了(最後の被験者の最終観察)から割付を明らかにするまでの間にデータをチェックすることを、データの盲検下レビュー(用語集参照)と呼ぶ。

2.3.2 ランダム化無作為化

ランダム化は、臨床試験において、被験者への試験治療の割付に意図的に偶然の要素を取り入れており、後に試験データを解析する際に、試験治療の効果に関する証拠の定量的な評価のための正しい統計的根拠を与える。また、ランダム化は予後因子が既知であるか未知であるかにかかわらず、予後因子の分布が類似した試験治療グループを作るために役立つものである。ランダム化は、盲検化と組み合わせることで、試験治療の割付が予見可能な場合に、被験者の選択的割付によって生じる可能性のある偏りを回避することに役立つものである。

臨床試験のランダム割付表は、被験者への試験治療のランダム割付を記録するものである。最も単純な状況では、割付表は一連の試験治療のリスト又は(クロスオーバー試験では、試験治療の順序のリスト)被験者番号に対応するコードである。スクリーニングの段階がある試験のような、一部の試験の実施手順は問題をより複雑なものにすることがあるが、被験者に対する試験治療又は試験治療の順についての事前に予定された唯一通りの割付は明確にすべきである。試験計画が異なれば、ランダム割付表の作成のためにも異なる手順が必要である。(必要な場合に備えて)ランダム割付表は、再現可能なものとすべきである。

制約をおかないランダム化は受け入れ可能な方法ではあるが、一般にはブロック別にランダム化の方が有利な点がある。このブロック別ランダム化の方法は、試験治療グループ間の比較可能性を高めることに役立つものである。これは例えば募集方針の変更により、結果として被験者の特徴が時間的に変化する可能性があるような場合である。また、ブロック別ランダム化は試験治療グループの被験者数をほぼ等しくすることを保証する。クロスオーバー試験では、ブロック別ランダム化を採用することが、効率が高く解釈しやすいバランスのとれた計画を得る手段となる。ブロックの長さは、被験者数のバランスが崩れる可能性を制限できる程度に短く、しかしブロックの終りの方での予見可能性を回避できる程度に十分長くするように注意すべきである。治験責任医師及びその他関係するスタッフは、一般にブロックの長さを知ることがないようにすべきである。二つ以上の異なるブロックの長さを用いて、それぞれのブロックで異なる長さをランダムに選ぶことにより、同じ目的を達成できる。（理論的には、二重盲検試験において予見可能性は問題とならない。しかし、治験薬の薬理作用により、知的な当て推量をする機会を提供してしまうおそれがある。）

多施設共同治験（用語集参照）では、ランダム化の手続きは中央で準備すべきである。施設で層別する、又は各施設にブロック全体を数個割付けるといった、各施設毎にランダム化する計画を立てることが推奨される。より一般的には、基準となる時点で測定された重要な予後因子（例えば、疾患の重症度、年齢、性、など）で層別することは、層内でバランスのとれた割付を促進するために有益といってよい。これは小規模な試験では大きな潜在的利益となる。三つ又は四つ以上の層別因子を用いる必要はほとんどない。そのうえ、例数のバランスがとりにくく、実施が煩雑になる。後述する動的割付法を用いることは、多数の層別因子のバランスを同時にとることに役立つであろう。そのためにはその後の試験手続きをこの型式の方法に応じて調整する必要がある。層別ランダム化に用いた因子は、後の解析の際にも考慮すべきである。

ランダム割付けされ試験に組み入れられる被験者は、常に適切なランダム割付表（層別ランダム化の場合には、各層ごと）の中の、まだ割付けられていない最初の番号に対応した試験治療を受けるべきである。次の被験者の該当番号とそれに対応する試験治療は、被験者が試験のランダム割付段階へ登録されたことを確認してから、はじめて割付けるべきである。予見可能性を高めてしまうようなランダム化に関する内容の詳細（例えば、ブロックの長さ）については治験実施計画書に含めるべきではない。ランダム割付表自体は、全試験期間において、盲検性が適切に維持されることを保証する方法で、治験依頼者又は第三者によって安全に保管されなければならない。どの被験者についても緊急の場合には割付を明らかにしなければならないことがあるため、試験中ランダム割付表にアクセスする可能性を考慮すべきである。その際に従うべき手順、必要な証拠資料及び割付を明らかにした後の被験者の治療と評価の仕方はすべて治験実施計画書に記述すべきである。

動的割付は割付方法の選択肢の一つであり、被験者への試験治療の割付に、現在までに割付けられている試験治療の例数バランスを反映させ、層別された試験では被験者が属する層での試験治療の例数バランスを反映させるものである。決定論的な動的割付法は避けるべきであり、試験治療の一つ一つの割付にランダム化の要素が適切に取り入れられるべきである。動的割付法を取り入れた試験の二重盲検性を保つためには、あらゆる努力がなされるべきである。例えば、通常は電話登録を用いることにより動的割付を管理している

中央治験事務局以外は試験治療コードを知ることができないようにする方法が採られるであろう。この方法は結果として、(電話登録時に)適格基準の追加確認を可能とし、試験への登録を確定する。これらの特徴はある種の多施設共同治験では有益である。二重盲検試験において通常用いられる治験薬を事前に箱詰めし、ラベルを貼り供給するシステムを、動的割付でも用いることができる。ただし、使用する順番は番号順ではなくなる。中央治験事務局の職員に、試験治療コードを知られることがないようにするため、適切なコンピュータアルゴリズムを用いることが望ましい。動的割付を検討する場合には、実施手順の複雑さと解析に与える潜在的な影響を慎重に評価すべきである。

III. 試験計画上で考慮すべきこと

3.1 試験計画の構成

3.1.1 並行群間比較計画

検証的試験で最もよく用いられる試験計画は並行群間比較計画である。被験者はそれぞれ異なる試験治療が割り当てられている二つ以上の群の一つにランダムに割付けられる。これらの試験治療は、一つ又は複数の用量の被験薬を含み、プラセボ若しくは実薬など一つ又は複数の対照治療も含むものであろう。並行群間比較計画の基礎となる仮定は、他のほとんどの計画における仮定に比べ複雑ではない。しかし、他の計画と同様に試験の解析と解釈を複雑にする別な側面（例えば、共変量、経時的繰り返し測定、要因間の交互作用、治験実施計画書違反、脱落（用語集参照）及び試験治療の中止）が存在することがある。

3.1.2 クロスオーバー計画

クロスオーバー計画では、各被験者は二つ又はそれ以上の試験治療を行う順序をランダムに割付けられる。したがって被験者自身を対照として試験治療比較が行われることになる。この単純な計画上の工夫は、主として、それを行うことが、定められた検出力の達成に必要な被験者数と通常は評価件数を劇的に減少させることがあるという理由から魅力的である。最も単純な 2×2 クロスオーバー計画では、各被験者は、多くの場合ウォッシュアウト期間をはさんで連続した二つの試験治療期間に、二種の試験治療のそれぞれをランダム化された順番で受ける。この計画の最も一般的な拡張では、 $n (>2)$ 個の試験治療を n 期間で各被験者がすべて受け、試験治療を比較することになる。計画の拡張には、各被験者が $n (>2)$ 個の試験治療の一部だけを受け、同じ試験治療を繰り返し行う、といった様々な変法が存在する。

クロスオーバー計画は、結果の妥当性を損なうおそれのある多くの問題を抱えている。最大の問題は持ち越し効果に関するものである。持ち越し効果とは、先行する試験治療が次に続く試験治療期間において及ぼす残存効果である。加法モデルでは、不均等な持ち越し効果が試験治療の直接の比較を偏らせる。 2×2 クロスオーバー計画では、持ち越し効果が試験治療と時期間の交互作用から統計的に分離できず、どちらの効果の検定も対応する対比が「被験者間」であるため検出力に欠ける。この問題は、試験治療の数が多い、より高次の計画ではそれほど深刻なものではないが、完全に排除することはできない。

したがって、クロスオーバー計画を用いる場合は、持ち越し効果を回避することが重要である。このためには、疾患の領域及び新しい薬剤の双方に関する十分な知識に基づき、クロスオーバー計画を選択的かつ慎重に使用することが最善の結果を生むことになる。対

象とする疾患は慢性的で症状が安定しているべきである。薬剤に関連のある効果はその試験治療期間中に完全に現れなければならない。ウォッシュアウト期間は薬剤効果が完全に消失するよう十分長くすべきである。これらの条件が満たされていると考えてよいについて、試験に先だって事前情報及びデータから確認しておくべきである。

クロスオーバー試験には、他にも注意を払わなければならない問題がある。その中で最も注意すべきことは、被験者の減失に起因する解析と解釈の複雑さである。また、持ち越し効果が存在する可能性があることから、後続の試験治療期間に発生した有害事象に対応した試験治療がどちらであるか特定することは困難である。ICH E4には、これらの問題と共に、他の問題も含め記述されている。クロスオーバー計画は、一般に試験からの被験者の減失が少ないと期待できる場合に限定すべきである。

2×2クロスオーバー計画が一般的に使用されており、通常その妥当性が満たされているのは、同一薬剤の異なる二つの剤型間で生物学的同等性を示す場合である。特にこの健康志願者への適用の場合、二つの投与期間の間のウォッシュアウト時間が十分長ければ、持ち越し効果が、それに関連のある薬物動態変数に関して、発生することはほとんど考えられない。しかし、得られたデータに基づいて、例えば各試験薬の投与開始時に薬物が検出できなかったことを示すことによって、解析時にこの仮定をチェックすることはやはり重要である。

3.1.3 要因計画

要因計画は、複数の試験治療の異なる組み合わせを複数用いて、二つ以上の試験治療を同時に評価するものである。最も単純な例は、2×2要因計画である。被験者は二つの試験治療で可能な四つの組み合わせの一つにランダムに割付けられる。その四つとは、試験治療をA, Bとすると、「A単独」、「B単独」、「AB同時」、「どちらもなし」である。多くの場合、要因計画はAとBの交互作用を調べるという特定の目的のために用いられる。主効果の検定に基づいて必要な被験者数が計算されている場合には、交互作用の統計的検定はその検出力に欠けるであろう。要因計画がAとBの併用効果を調べるために用いられる場合、特に二つの試験治療が併用される見込みのある場合、この配慮は重要である。

要因計画の別の重要な利用法は、試験治療CとDを同時に使用する場合の用量-反応特性を立証することである。それは特に、先行する試験から、それぞれ単剤での有効性がある用量において立証されている場合である。Cについて、通常ゼロ用量（プラセボ）を含んだmの用量が選択され、Dについてもそれに近い数であるnの用量が選択される。全ての組み合わせを含んだ計画はm×n個の試験治療グループからなり、各グループはCとDの異なる用量の組み合わせのうちの一つを受ける。このようにして得られた反応曲面を用いることが、臨床適用のために適切な、CとDの用量の組み合わせを同定するのに役立つ場合がある(ICH E4参照)。

2×2要因計画は、一方の試験治療の有効性を評価するための被験者数で両方の試験治療の有効性を評価することにより、臨床試験の被験者を効率よく利用するために用いられる場合がある。この戦略は、死亡に関する大規模試験で特に有益であることが示されている。この方法の効率と妥当性は、試験治療AとBとの間に交互作用がないことに依存している。交互作用がなければ、主たる有効性変数に関するAとBの効果は加法モデルに従い、このためA単独の効果とBの効果に上乗せされたAの効果はほとんど等しくなる。クロスオーバー

試験と同様に、この条件が満たされていると考えてよいことを示す根拠は、試験に先だって事前情報及びデータから確認しておくべきである。

3.2 多施設共同治験

多施設共同治験が実施されることには、主として二つの理由がある。第一の理由は、多施設共同治験が、新医薬品の有効性をより効率よく評価するための方法として認められていることである。場合によっては多施設共同治験が、妥当と考え得る範囲の期間内に試験の目的を満たすための十分な被験者を登録する唯一の実用的な手段となることがある。このような特徴を持つ多施設共同治験は、原則として、臨床開発のあらゆる段階で実施してよい。多施設共同治験は、施設当たりの被験者数が多い数力所の施設において行う場合があり、まれな疾患の試験の場合は施設当たりの被験者数が少ない多数の施設において行うこともある。

多施設（かつ多治験責任医師）共同治験が計画される第二の理由は、得られた結果を一般化するためのより適切な根拠を与えるためであろう。第二の理由は、多施設共同治験がより広い患者集団から被験者を募集できる可能性があること、及び臨床の広い範囲の現場でこの医薬品が使用される可能性があることに基づいている。したがって、多施設共同治験は、将来使用される状況により近い実験状況を提供するものである。この場合、多数の治験責任医師が参加することは、医薬品の有益性に関して広範囲にわたる臨床的判断がなされる可能性をも生み出すことになる。一般化を目的とした多施設共同治験は、医薬品開発の後期の相での検証的試験となり、多数の治験責任医師と施設が参加することになる。一般化可能性（用語集参照）を更に高めるために、多施設共同治験が複数の異なる国にまたがって実施されることもあり得る。

多施設共同治験の結果の意味が十分に解釈され、外挿されるためには、治験実施計画書を実行する方法は明確で、すべての施設で同様のものであるべきである。更に、通常行われる必要な被験者数と検出力の計算は、施設が異なる場合でも比較する試験治療間の差の大きさは同じであるという仮定に依存している。このような背景を考慮して、共通の治験実施計画書を作成し、以下のように多施設共同治験を実施することが重要である。実施手順はできるかぎり徹底して標準化すべきである。評価基準及び評価体系のばらつきは、治験責任医師会議、試験前に行う関係者の訓練、試験実施中の慎重なモニタリングを通して小さくすることができる。適切な計画を立てるためには、一般に施設内で試験治療毎に被験者の分布が同じになるよう心掛けるべきであり、適切な運営管理により、この計画の目的を維持すべきである。後に施設間での試験治療効果の不均一性を考慮する必要性が判明した場合には、施設当たりの被験者数が過度に異なることがないようにしている試験及び極端に小規模な施設を含まない試験が有利である。それは、施設毎の重みを変えた場合でも、試験治療効果の重み付き推定値がそれほど異ならないからである。（この点は、すべての施設が小規模で、施設の特徴が解析には現れない多施設共同治験には当てはまらない。）これらの予防策を採用しないことは、結果の均一性が疑わしいことと併せると、深刻な場合には承認に関わる治験依頼者の主張に対して説得力のある根拠を示すものとは見なせない程度まで多施設共同治験の価値を減じるおそれがある。

最も単純な状況の多施設共同治験では、個々の治験責任医師が一つの病院で募集された被験者に対し責任をもつため、「施設」は治験責任医師又は病院に対して一つに特定される。しかし、多くの場合、状況はもっと複雑である。恐らく一人の治験責任医師が数力所

の病院で被験者を募集するということもあり、一人の治験責任医師が、一つ又は複数の関連病院における自身の診察室で被験者を募集する臨床家（治験分担医師）のチームの代表であることもある。統計モデルにおける施設の定義に疑問の余地がある場合、治験実施計画書の統計の部（5.1節参照）には、その多施設共同治験の枠組みにおける施設という用語の定義を明確にすべきである（例えば、治験責任医師毎なのか、場所なのか、又は地域なのか）。ほとんどの場合、施設は治験責任医師により定義することが十分可能で、ICH E6はこれに関連した指針を示している。施設の定義が疑わしい場合には、主要変数の測定及び試験治療に影響を与える重要な要因が施設内で均一になるように施設を定義すべきである。解析の際に施設を併合するためのルールはすべて、可能な限り前もって治験実施計画書中にその正当性も含めて記述すべきである。しかしどのような場合でも、とるべき手段に関する決定は、常に試験治療について盲検下で、例えば盲検下レビューの際に行うべきである。

試験治療の効果の推定と検定に用いる統計モデルは治験実施計画書に記載すべきである。試験治療の主効果は、最初に施設と試験治療の交互作用を含まず施設間差を考慮に入れるモデルを用いて調べることができる。モデルに常に交互作用を含めると、試験治療の効果が施設間で均一な場合、主効果の検定の効率が低下する。試験治療効果の不均一性が真に存在する場合には、主効果の解釈には様々な議論がある。

例えば死亡を評価する大規模試験で施設当たりの被験者数が少ない試験にみられる例として、施設が臨床的に重要な影響を反映するとは考えにくいために施設が主要変数又は副次変数に影響を与えると考える理由はないといって差し支えないような場合がある。別の試験では、施設当たりの被験者数が少ないために、統計モデルに施設の効果を含めることが実行不可能であることが、あらかじめ認識できる場合もある。それらの場合、モデルに施設の項を含めることは適切ではなく、施設で層別したランダム化を行うことも必ずしも必要ではない。

施設当たりの被験者数が不均一性を評価しうる規模の試験で、試験治療の肯定的な効果が判明した場合、結論の一般化可能性に影響する可能性があるため、通常は施設間における試験治療効果の不均一性を探索すべきである。著しい不均一性は、個々の施設の結果を図示すること又は試験治療と施設間の交互作用の有意性検定などの解析手法によることでも確認される場合がある。交互作用の統計的有意性検定を用いる場合、試験治療の主効果を検出することを目的に計画した試験では、一般に交互作用の検定の検出力は低いことを認識しておくことが重要である。

試験治療効果の不均一性がみられた場合、その解釈には注意すべきであり、試験の運営管理面又は被験者の特徴といったそれとは別の面から説明できるかどうかを、積極的に調べるべきである。通常はその説明によって、適切な追加解析と解釈が示唆される。説明ができない場合、例えば著しい量的交互作用（用語集参照）から試験治療効果の不均一性の存在が明らかとなることは、施設に異なる重みを与えて試験治療効果の推定値を複数求め、試験治療効果の推定値の安定性を実証する必要があることを意味する。不均一性が著しい質的交互作用（用語集参照）により特徴付けられるものであれば、その理由を解明することは更に重要であり、説明ができない場合は、試験治療効果を確実に予測するために、追加の臨床試験を必要とするであろう。

これまで、多施設共同試験に関する議論は、固定効果モデルを用いることを前提としてきた。混合モデルも試験治療効果の不均一性を探索するために利用できる。混合モデルでは、施設及び試験治療と施設の交互作用を変量効果として扱っており、特に施設数が多い場合に用いることが適切である。

3.3 比較の型式

3.3.1 優越性を示すための試験

科学的には、有効性を立証するには、プラセボ対照試験でプラセボに優ることを示すこと、実対照薬に優ることを示すこと又は用量-反応関係を示すことが最も説得力がある。この型式の試験を「優越性」試験（用語集参照）と呼ぶこととする。本ガイドラインでは、特に断らない限り優越性試験を前提としている。

重篤な疾患に対して優越性試験により有効であることが示されている治療法が存在する場合、プラセボ対照試験は非倫理的と考えられることがある。その場合、実治療を対照として科学的に正しく用いることを考慮すべきである。プラセボ対照と実薬対照のどちらが適切であるかは、個々の試験ごとに判断すべきである。

3.3.2 同等性又は非劣性を示すための試験

優越性を示す目的以外にも、被験薬と標準治療とが比較される場合がある。この型式の試験は目的に応じて二つの主要なカテゴリに分けられる。一つは「同等性」試験（用語集参照）で、もう一つは「非劣性」試験（用語集参照）である。

生物学的同等性試験は前者のカテゴリに属している。また、例えば化合物が吸収されずそのために血中に現れない場合に、後発医薬品と先発医薬品との臨床的同等性を示すといった規制側の理由から、臨床的同等性試験が要求されることがある。

多くの実薬対照試験は、被験薬の有効性が実対照薬の有効性よりも劣らないことを示すために計画され、したがって後者のカテゴリに属している。実薬対照試験のもう一つの例は、被験薬の複数の用量と標準薬の推奨用量又は複数の用量が比較される試験である。被験薬の用量-反応関係を示すこと及び被験薬と実対照薬とを比較することを同時に行うことがこの試験計画の目的である。

実薬対照同等性試験又は非劣性試験には、プラセボを組み込んでよく、そうすることで一つの試験で複数の目標を達成できる。例えば、プラセボに対する優越性の立証とその結果として試験計画の妥当性を確認できると同時に、実対照薬に対する有効性及び安全性がどの程度類似しているかについても評価できる。プラセボを含まない又は被験薬の複数用量を用いない実薬対照同等性試験（又は非劣性試験）には、よく知られた問題点がある。その問題点とは、（優越性試験とは対照的に）内部妥当性を示すいかなる指標も必然的に存在していないことであり、このため外部情報による妥当性の確認を必要とする。同等性試験（又は非劣性試験）は本質的に保守的でないため、試験の計画上又は実施上の多くの不備が、同等であると結論づける方向へ結果を偏らせる傾向がある。これらの理由から、このような試験ではその計画上の特徴に特に注意すべきであり、慎重に実施する必要がある。例えば、登録基準違反、服薬不遵守、試験治療の中止、追跡不能、欠測データ及び試験実施計画書からのその他の逸脱を最小限に抑えることは特に重要であり、またこれらが

解析に与える影響も抑えることが重要である。

実対照薬は慎重に選択すべきである。適切な実対照薬の例としては、広く使用されている治療法で、十分に計画され記録されている一つ以上の優越性試験によって適切な適応に対する有効性が明確に立証され定量的に示されており、現在計画している実薬対照試験においても同様の有効性を示すことが十分に期待できるものがあげられる。このためには、新たに行う試験に関連した、医学又は統計学の進歩を考慮した上で、新たに行う試験計画上の重要な特徴（主要変数、実対照薬の用量、適格基準など）を、実対照薬が臨床的に適切な有効性を明確に示した過去の優越性試験と同じにすべきである。

同等性又は非劣性を証明するために計画された試験では試験実施計画書に同等性又は非劣性を示すために計画されたということを明確に述べることが不可欠である。試験実施計画書には同等限界を明示しておくべきである。同等限界とは、臨床的に許容できると判断しうる最大の差であり、実対照薬の有効性を立証した優越性試験において観測された差よりも小さいものであるべきである。実薬対照同等性試験では、上側及び下側両方の同等限界が必要であり、実薬対照非劣性試験では下側同等限界のみが必要である。同等限界の大きさの選択には、十分な臨床的根拠を示すべきである。

統計解析は、通常信頼区間に基づいて行われる（5.5節参照）。同等性試験では、両側信頼区間を用いるべきである。信頼区間全体が同等限界内に含まれる場合、同等であると推論する。両側信頼区間の使用は、試験治療の差は同等限界の外側にあるという（複合）帰無仮説に対し、試験治療の差は同等限界の内側にあるという（複合）対立仮説を検定する、二つの片側検定を同時に行う方法と実際上同じものである。二つの帰無仮説には重なりがないため、第一種の過誤は適切に制御される。非劣性試験では、片側信頼区間を用いるべきである。信頼区間を用いた方法は、（被験薬から対照を引いた）試験治療間の差は下側同等限界に等しいという帰無仮説に対して、試験治療間の差は下側同等限界よりも大きいという対立仮説を検定する片側仮説検定に対応する。第一種の過誤の大きさの選択は、片側検定又は両側検定のどちらを選択するかとは別に検討すべきである。被験者数の計算は、これらの方法に基づくべきである（3.5節参照）。

被験薬と実対照薬に差がないという帰無仮説の検定結果が有意でないことから、同等性又は非劣性が示されたと結論することは不適切である。

解析対象集団の選択にも特別な問題が生じる。試験治療グループ又は対照治療グループにおいて、試験治療を中止した被験者又はそれらのグループから脱落した被験者は、効果が現れにくいことから、最大の解析対象集団（用語集参照）を用いた結果は同等性を示す方向に偏るおそれがある（5.2.3節参照）。

3.3.3 用量-反応関係を示すための試験

被験薬がどのような用量-反応関係を示すかは、開発のすべての相から、様々な方法によって解答が得られる可能性がある問題である（ICH E4 参照）。用量-反応試験は多くの目的に役立つであろう。中でも次に示すものは特に重要である。有効性の確認、用量-反応曲線の形状と位置の研究、適切な開始用量の推定、個人毎の用量の調整に最適な戦略の同定、それ以上臨床上の利益を見込むことができない最大用量の決定。プラセボ（ゼロ用量）を含めることが適切な場合にはプラセボを含め、多くの用量について集められたデータを用

いて、これらの目的に対応する必要がある。そのためには、用量-反応関係の推定に信頼区間の構成及びグラフ表示を用いた手法を適用することが、統計的検定を使用することと同程度に重要である。仮説検定を用いる場合は、用量の順序関係又は用量-反応曲線の形状に関する個々の問題（例えば単調性）に対応した方法を用いる必要があろう。予定している統計的な手続きに関する内容の詳細は、治験実施計画書に述べるべきである。

3.4 逐次群計画

逐次群(群逐次)計画は、中間解析(4.5節及び用語集参照)を実施するために用いられる。中間解析を可能にする計画がいくつかある中で、逐次群計画が受け入れ可能な唯一の型式というわけではないが最もよく適用されている。なぜならば、被験者の結果を試験期間中定期的にまとめて評価することは、個々の被験者の結果が利用可能になる都度評価するよりも実際のだからである。逐次群計画での統計手法は、試験治療の結果及び試験治療の割付に関する情報が利用可能となる(盲検解除、4.5節参照)前に、完全に明記しておくべきである。独立データモニタリング委員会(効果安全性評価委員会)(用語集参照)は、逐次群計画から得られるデータの中間解析の実施又は検討のために利用される(4.6節参照)。逐次群計画は、死亡又は重大な非致死性の評価項目を調べる大規模で長期にわたる試験で広く用いられ成功してきたが、その他の状況でも利用されることが増えてきている。特にすべての試験で安全性をモニターしなければならないことは共通の認識となっていることから、安全性の理由から早期中止を行うことも含めて正式な手続きの必要性を常に検討しておくべきである。

3.5 必要な被験者数

臨床試験の被験者数は、提示された問題に信頼のおける解答を与えられるよう常に十分多くすべきである。試験に必要な被験者数は、通常試験の主要な目的により決められる。被験者数がその他の理由から決定される場合には、その理由を明確にし正当化しておくべきである。例えば、安全性に関する問題若しくは要求に基づいた試験又は重要な副次目的に基づいて被験者数が決定される試験では、主要な有効性の問題に基づいて被験者数が決定される試験よりも多くの被験者数を必要とするであろう(例えば、ICH E1A 参照)。

適切な被験者数を決定するために用いられる通常の方法を利用するためには、以下の項目を定めておくことが必要である。それらは、主要変数、検定統計量、帰無仮説、選択された用量での対立(「作業」)仮説(その用量と選ばれた対象集団で検出すべき又は棄却すべき試験治療の差を考慮することも含めて)、誤って帰無仮説を棄却する確率(第一種の過誤)及び誤って帰無仮説を棄却できない確率(第二種の過誤)であり、更に試験治療を中止した被験者及び治験実施計画書違反を取り扱う方法も定める必要がある。検出力の評価のために、イベント発生率が主要な関心事項となる場合には、試験に必要なイベント数から最終的な被験者数を外挿するための仮定も置くべきである。

被験者数を計算する方法は、計算に用いる見積値(分散、平均値、反応割合、イベント発生率、検出すべき差)とともに、治験実施計画書に定めておくべきである。また、これらの見積値の根拠も示すべきである。これらの仮定からの様々なずれに対して、必要な被験者数がどの程度変わり易いか調べることは重要であり、このためには実際に起こりうるずれの範囲に対応する被験者数の範囲を示すことで実施することが容易になるであろう。検証的試験では、通常これらの仮定は公表されたデータ又は先行する試験の結果に基づくべきである。検出すべき試験治療の差は、患者の治療管理を行う上で臨床的に意味をもつ

最小限度の効果に関する判断又は新しい試験治療の予想される効果の方が大きい場合にはその効果に関する判断に基づいて決まるものであろう。慣例的に、第一種の過誤は5%以下に設定され、多重性を考慮するために必要な調整がなされる場合はそれに従って設定される。検証すべき仮説のもっともらしさ及び検定結果に望む影響力の強さにより、第一種の過誤の的確な選択に影響が及ぼされるであろう。第二種の過誤は、慣例として10%～20%に設定される。第二種の過誤を実施可能な範囲でできる限り小さくすることは、特に繰り返すことが困難又は不可能な試験の場合、治験依頼者の利益となる。慣例として用いている第一種の過誤の値及び第二種の過誤の値とは異なる値を用いることも許容される場合があり、むしろそれが好ましいこともあり得る。

被験者数の計算は、主要な解析で用いる解析対象集団に基づくべきである。解析対象集団が「最大の解析対象集団」である場合、効果の大きさに関する見積値は、治験実施計画書に適合した対象集団（用語集参照）の場合に比べて小さくする必要があるだろう。これは、試験治療を中止した被験者又は服薬遵守状況の悪い被験者を解析に含めることにより、試験治療の効果が薄められることを考慮するためである。このときばらつきに関する仮定も再検討する必要があるだろう。

同等性試験又は非劣性試験の被験者数（3.3.2節参照）は、通常試験治療の差の信頼区間を用いて、試験治療間の差が最大でも臨床的に許容できる範囲であることを示すという目的に基づいて計算すべきである。同等性試験での検出力が真の差をゼロとして設定されている場合、真の差がゼロでなければ、この検出力を達成するために必要な被験者数よりも少なく見積もられることになる。非劣性試験での検出力が差をゼロとして設定されている場合、被験薬の効果が実対照薬の効果よりも小さければ、この検出力を達成するための必要な被験者数よりも少なく見積もられることになる。「臨床的に許容できる」差は、その選択に当たり市販後使用される患者に対してどのような意味を持つかに関する正当な理由が必要であり、差が存在することを立証するために計画する優越性試験において参照した前述の「臨床的に適切な」差よりも小さくなるであろう。

逐次群試験での正確な被験者数は、選択した中止指針と真の試験治療の差に依存する上に、偶然の動きにも左右されるため、事前には固定できない。中止指針の設計には、試験を続けた際の被験者数の分布を考慮すべきであり、通常これは期待被験者数及び最大被験者数により具体的に示される。

イベント発生率が予想よりも低い場合、又はばらつきが予想よりも大きい場合は、割付を明らかにすること又は試験治療間の比較を行うことなく被験者数を見直すことができる（4.4節参照）。

3.6 データの獲得と処理

治験責任医師から治験依頼者へのデータの収集と転送は、症例記録用紙、遠隔地モニタリングシステム、医療コンピューターシステム、電子的転送等の様々な媒体で行うことができる。どのようなデータ獲得の手段を用いても、収集する情報の様式及びその内容は治験実施計画書と完全に一致させるべきであり、臨床試験の実施前に確定しておくべきである。収集する情報の様式及びその内容は予定した解析の実施に必要なデータに合わせて考えるべきである。必要なデータには、治験実施計画書遵守状況の確認又は重要な治験実施計画書からの逸脱を明らかにするために必要な背景情報（服薬に対応した評価の時期等）

が含まれる。「欠測値」は「ゼロ」又は「該当せず」と区別できるようにすべきである。

データベースの確定までのデータ獲得の手順は、GCPに従って実行すべきである(ICH E6、5節参照)。特に、質の高いデータベースの引き渡しを確かにし、予定した解析の履行を通じた試験目的の達成を確実にするためには、データの記録並びに誤り及び無記入の訂正のために、適切なタイミングで信頼できる処理を実施することが、必要である。

IV. 試験実施上で考慮すべきこと

4.1 治験モニタリングと中間解析

治験実施計画書に従って臨床試験が慎重に実施されているかどうかは、結果の信憑性に重要な影響を与える(ICH E6 参照)。慎重なモニタリングによって、実施上の問題の所在を早期に発見するとともに、問題の発生又は再発を最小限に抑えることが保証できる。

モニタリングには、製薬企業が依頼する検証的試験の性格を一般に左右するような二つの異なる型式が存在する。一方の型式は試験の質の監視と関係したものであり、もう一方の型式は試験治療の比較のために割付を明らかにすることを伴うものである(中間解析)。治験モニタリングにおける二つの型式はどちらも、異なるスタッフの責任を伴ううえ、異なる型式の試験データ及び情報へのアクセスを必要とすることから、このための異なる原則が、潜在的な統計的及び運営上の偏りの制御に適用される。

試験の質を監視するためには、治験モニタリングで、治験実施計画書が守られているか、集積されたデータが受け入れ可能か、予定している集積目標が達成されているか、計画時に用いた仮定は適切か、患者の試験への継続的参加に成功しているか、などをチェックする必要がある(4.2節から4.4節を参照)。この型式のモニタリングは、試験治療効果の比較に関する情報へのアクセスを必要としないだけでなく、割付を明らかにしたデータを必要としないため、第一種の過誤への影響を与えるものではない。この目的での治験モニタリングは治験依頼者の責任であり(ICH E6 参照)、治験依頼者又は治験依頼者によって任命された独立したグループが行うことができる。この型式のモニタリングの期間は、通常試験実施施設が選択されたときに始まり、最後の被験者のデータが収集されクリーニングされたときに終了する。

治験モニタリングのもう一つの型式(中間解析)は、比較のため試験治療の結果の集積を必要とする。中間解析は、割付を明らかにして(キーコードの開示)試験治療グループにアクセスすること(実際の試験治療の割付、又は割付グループの同定)を必要とし、比較を行う試験治療グループ間の要約情報を必要とする。このため、ある種の偏りを防ぐ目的で、中間解析のための統計解析計画を治験実施計画書中(又は最初に解析を行う前に目的にあった改訂を行った場合、その改訂中)に含める必要がある。これに関しては4.5節と4.6節で議論する。

4.2 選択基準と除外基準の変更

選択基準及び除外基準は、被験者募集期間を通じて、治験実施計画書に明記されているとおり一定に保つべきである。ときには基準を変更することが適切な場合もある。例えば、長期にわたる試験において、その試験以外又は中間解析による医学知識の蓄積により、登録基準の変更が示唆される場合である。登録基準の違反が日常的に起こること又は募集率

の低さが深刻であることが、登録基準の制限が強すぎたためであることをモニタリング担当者が発見することにより、登録基準の変更がなされる場合もある。登録基準の変更は割付を明らかにしない状態で行うべきであり、治験実施計画書の改訂に常に記述すべきである。治験実施計画書の改訂には、例えば、イベント発生率が異なることにより行われなければならない必要な被験者数の調整などの統計的変更の内容、又は修正された選択/除外基準に従った解析の層化など、予定した解析の修正を含めるべきである。

4.3 集積率

被験者の集積が長期にわたる試験では、集積率をモニターすべきである。もしそれが予定している水準を大きく下回る場合には、その理由を確認すべきであり、試験の検出力を保ち、選択的登録及びその他試験の質に関する別の側面についての懸念を和らげるための対応策をとるべきである。多施設共同治験では、個々の施設において、これらの配慮が適用される。

4.4 必要な被験者数の調整

長期にわたる試験では、通常、当初の計画で用いた被験者数の計算根拠となる仮定を確認するための機会がある。この確認は、試験計画の詳細が予備的情報若しくは不確実な情報、又はその両方に基づいている場合、特に重要であろう。盲検下のデータを用い中間での確認を行うことにより、それまでの試験全体での、反応の分散、イベント発生率又は生存状況が予期していた状況と異なることが明らかにされる場合がある。その場合、適切に修正した仮定に基づいて被験者数の再計算を行うこととなるが、その正当性を明らかにし、治験実施計画書の改訂及び総括報告書に記録しなければならない。盲検性を維持するために行う手続きと共に、可能であれば、第一種の過誤と信頼区間の幅に対する被験者数の変更による影響を説明すべきである。被験者数の再見積もりが必要になる可能性がある場合には、そのことを可能な限り治験実施計画書に述べるべきである（3.5節参照）。

4.5 中間解析と早期中止

中間解析とは、試験が正式に完了する前に行われる有効性又は安全性に関する試験治療群間の比較を意図したすべての解析を指す。中間解析の回数、方法及び結果が試験の解釈に影響するため、実施するすべての中間解析は前もって慎重に計画し、治験実施計画書に記述すべきである。特別な状況では、試験開始当初には予定していなかった中間解析が必要となる場合がある。中間解析が必要となる場合、割付が明らかにされ試験治療を比較するデータにアクセスする前に、治験実施計画書の改訂に中間解析について記述しておくべきである。試験を継続すべきか中止すべきかの決定を目的として中間解析を計画する場合、統計的モニタリング計画を指針とする逐次群計画を用いるのが通常である（3.4節参照）。このような中間解析の目的は、研究中の試験治療の優越性が疑いなく立証された場合、適切な試験治療の差を示す見込みのないことが判明した場合又は許容できない有害作用が明らかになった場合に試験を早期に中止することにある。一般に、有効性モニタリングのための棄却限界値は、安全性モニタリングのための棄却限界値よりも、試験を早期に中止するための証拠をより多く必要とする（つまり、より保守的とする必要がある）。治験実施計画とモニタリングの目的により、複数の評価項目が用いられる場合、それによる多重性にも注意を払う必要がある。

治験実施計画書には中間解析のスケジュール又は、例えばアルファ消費関数を用いる柔軟な方法を予定している様な場合には、少なくとも中間解析の方針について述べるべきで

ある。より詳細な内容については最初の中間解析時までには治験実施計画書の改訂に示してもよい。中止の指針及びそれらの特質を治験実施計画書又はその改訂に明確に述べるべきである。早期中止が中止の判断に用いた変数以外の重要な変数に及ぼす影響と、それらの変数の解析方法についても考慮すべきである。この内容は、独立データモニタリング委員会が存在する場合には、その委員会により記述されるか又は承認されるべきである（4.6節参照）。予定した手順からの逸脱により常に試験結果の妥当性が損なわれる可能性がある。試験に変更を加える必要性が生じた場合、変更により必要となる統計手法の変更のすべてを、できるだけ早い機会に治験実施計画書の改訂に明記すべきである。特にそのような変更が原因となるおそれがあるすべての解析及び推測への影響を議論しておかなければならない。選択した中間解析の方法が、全体の第一種の過誤の確率を制御していることを保証すべきである。

中間解析では、割付を明らかにしたデータと結果が必要となりうるため、内容が全く漏れない手順により実施しなければならない。試験に対するスタッフの態度の変化及び募集される患者の特徴の変化が起こり、試験治療の比較に偏りをもたらす原因となりかねないため、試験の実施に関係しているすべてのスタッフが、中間解析の結果を知ることがないようにすべきである。この原則は、中間解析の実施に直接関係している者を除く、治験責任医師に関係するスタッフ全員及び治験依頼者に雇用されているスタッフにも適用されるといいよい。治験責任医師には、試験の継続若しくは中止の決定、又は試験手順の変更の決定のみを知らせるべきである。

被験薬の有効性及び安全性を裏付けることを意図した臨床試験は、ほとんどの場合、予定した被験者数の集積が完全に完了するまで継続すべきである。試験は、倫理的な理由又は検出力が容認できない場合に限り、早期に中止すべきである。しかし、医薬品開発計画には、他の試験計画の立案など様々な理由から、比較のための試験治療データに治験依頼者がアクセスする必要があることが認識されている。また、集積していく試験治療効果の比較に関して継続的なモニタリングが倫理的な理由から必要となるような、生命を脅かす重篤な結果に関する研究又は死亡に関する研究があるが、そのような試験は全体のごく一部であることも認識されている。どちらの状況でも中間統計解析の計画は、生じるおそれのある潜在的な統計的及び運営上の偏りに対処するため、割付を明らかにして試験治療の比較データにアクセスする前に、治験実施計画書又はその改訂の適切な箇所に記載すべきである。

被験薬に関する臨床試験では多くの場合、特に公衆衛生上意義のある試験の場合は、有効性若しくは安全性の一方、又はその両方の比較についてのモニタリングに対する責任は、外部の独立したグループに委ねられるべきである。このグループは、独立データモニタリング委員会、データ及び安全性モニタリング委員会、又はデータモニタリング委員会と呼ばれることがあり、その責任は明確にしておく必要がある。

治験依頼者が有効性又は安全性を比較するためのモニタリングの役割を担い、割付を明らかにした情報へのアクセス権を持つ場合、試験の完全性を維持し、情報の共有を適切に管理し制限するために特別の注意が払われるべきである。治験依頼者は、内部モニタリング委員会が文書で書かれた標準業務手順書に従っていること及び意思決定を行った会議の議事録が中間解析の結果とともに保持されていることを保証し記録に残すべきである。

適切に計画されていない中間解析はすべて（試験の早期中止の結論によらず）、試験の結果を損なう恐れがあるとともに導いた結論の信憑性を低下させる可能性がある。したがって、計画されていない中間解析は行うべきではない。もし予定外の中間解析を実施するならば、その解析の必要性及びどの程度割付を明らかにしたかを総括報告書に説明すべきであり、生じるおそれのある偏りの大きさ及び結果の解釈への影響を評価すべきである。

4.6 独立データモニタリング委員会の役割

（ICH E6、1.25節及び5.52節参照）

独立データモニタリング委員会は、臨床試験の進行状況、安全性データ及び重要な有効性変数を何回かにわたり評価するとともに治験依頼者に試験の継続、変更、又は中止を勧告するために治験依頼者が設立できる。独立データモニタリング委員会は、文書で書かれた業務手順書を持つべきであり、すべての会議の議事録と中間解析の結果を保持すべきである。また、これらは試験完了時には審査可能な状態にしておくべきである。独立データモニタリング委員会の独立性は、比較を目的とした重要な情報の漏洩を管理すること及び臨床試験の情報にアクセスすることによる悪影響から試験の完全性を守ることを目的としたものである。独立データモニタリング委員会は治験審査委員会又は倫理委員会とは別の組織であり、統計学を含む適切な学識を持った臨床試験の専門家から構成されるべきである。

独立データモニタリング委員会に治験依頼者を代表する者が参加する場合、（例えば、主要な問題の採決に参加できるかどうか等）その役割を委員会の業務手順書に明確に定めるべきである。委員会に参加した治験依頼者のスタッフは割付を明らかにした情報に対しアクセス権を持つと考えられることから、業務手順書には治験依頼者の組織に対し試験の中間結果の開示を制限することについても述べるべきである。

V. データ解析上で考慮すべきこと

5.1 解析の事前明記

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。統計の部には、主要変数について行うこととした検証的解析の主要な特徴のすべてと、予想される解析上の問題に対処する方法を含めるべきである。探索的試験の場合には、統計の部の記述をより一般的な原則と方向性にとどめることができる。

統計解析計画書（用語集参照）は治験実施計画書完成後に別の文書として作成することができる。統計解析計画書には、治験実施計画書に述べた解析の主要な特徴について、より技術的で詳細な記述を含めることができる（7.1節参照）。統計解析計画書には、主要変数、副次変数及びその他のデータに関する統計解析を実行するための詳細な手順を記載してもよい。統計解析計画はデータの盲検下レビュー（7.1節で定義する）の結果に基づいて再検討すべきであり、場合によっては更新し、割付を明らかにする前に完成させるべきである。割付を明らかにした日付だけでなく、統計解析計画書を完成した日付も公式な記録として保持すべきである。

盲検下レビューから、治験実施計画書に述べた解析の主要な特徴となる事項の変更を提案しようとする場合、変更の内容は治験実施計画書の改訂に記録すべきである。変更が主

要事項でない場合は、盲検下レビューにより示唆された検討事項に基づいて統計解析計画書を更新すれば十分であろう。治験実施計画書（その改訂を含む）に予定された解析の結果のみが検証的であるとみなされる。

総括報告書の統計の部には、一連の臨床試験の過程において方法論に関する決定がいつなされたかも含めて、統計的方法論を明確に記述すべきである（ICH E3 参照）。

5.2 解析対象集団

主要な解析でデータが用いられる被験者の集団は、治験実施計画書の統計の部で定義すべきである。更に、試験の手続きが開始された（例えば、治験導入期の）すべての被験者の記録は有用といってよい。この被験者の記録の内容は、個々の試験の詳細に依存するが、少なくとも人口統計学的なデータ及び疾患の状態に関する基準となる時点のデータは、可能な限り収集すべきである。

ランダム化が行われて臨床試験に参加したすべての被験者が、すべての登録基準を満たし、追跡不能となることもなくすべての試験手順に従い、そしてデータの記録が完全であれば、解析に含める被験者の集団は自明であろう。試験を計画し、実施するに当たっては、この理想に可能な限り近づけることを目指すべきであるが、実際には、これまでにこの理想が完全に達成されたことがあるかどうかは疑わしい。したがって治験実施計画書の統計の部には、予想される問題について、それらが被験者及び解析されるデータにどのように影響を与えるかについて前もって述べておくべきである。更に治験実施計画書には、十分な解析を損なうと予想される研究実施中の変則的な事例のすべて（様々な種類の治験実施計画書違反、試験治療の中止、欠測値を含む）を最小にするための手続きも明記すべきである。治験実施計画書では、そのような問題の発生する頻度を少なくするための方法と、データ解析中に起こる問題へ対処する方法とを考慮しておくべきである。治験実施計画書違反に対処する解析の方法を修正する必要性があれば、盲検下レビューの際に確認しておくべきである。すべての重要な治験実施計画書違反は、違反が起きた日時、違反が起きた理由及び試験結果へ与える影響を明らかにすることが望ましい。治験実施計画書違反及び欠測値などの問題の発生頻度とその種類は総括報告書に記録すべきであり、それらが試験結果に対し、どのような影響を与える可能性があるかを記述すべきである（ICH E3 参照）。

解析対象集団の選択に関する判断は、次の二つの原則に従うべきである：1) 偏りを最小にすること、及び 2) 第一種の過誤の増大を回避すること。

5.2.1 最大の解析対象集団

Intention-to-treatの原則（用語集参照）は、ランダム化が行われた全被験者を主要な解析に含めるべきであると主張するものである。この原則を遵守するには、臨床結果を得るためにランダム化が行われた全被験者を完全に追跡することを必要とするであろう。しかし実際には、後に述べる理由からこの理想を達成することは難しいであろう。本ガイドラインでは「最大の解析対象集団」という用語を用いて、ランダム化が行われた全被験者を含める intention-to-treat という理想に可能な限り完全に近づけた解析対象集団を表すことにする。解析時にも最初のランダム化を維持することは、偏りを防ぎ、統計的検定の強固な基盤を与える点で重要である。多くの臨床試験において、最大の解析対象集団を用いることは保守的な戦略となる。また多くの状況で、最大の解析対象集団により得られる試験治療の効果の推定値は、後の日常診療での効果を反映する可能性がより高いといってよい。

ランダム化が行われた被験者を最大の解析対象集団から除外することになる状況は限られている。それらには、主要な登録基準を満たしていない場合（適格基準違反）、試験治療を一回も受けていない場合、ランダム化後のデータがない場合などがある。そのような除外については常に理由を示すべきである。登録基準を満たしていない被験者は、以下の条件下でのみ偏りを導入する可能性なく除外できるであろう。

- (i) 登録基準はランダム化以前に評価されている
- (ii) 除外の対象となる適格基準違反の発見は完全に客観的になされる
- (iii) すべての被験者が適格基準違反について同様の綿密さで調べられている（非盲検試験においてはこの保証は困難であり、二重盲検試験であっても割付を明らかにした後では難しい。このことは盲検下レビューの重要性を強調している。）
- (iv) 特定の登録基準違反が発見された場合、それに関するすべての違反が除外される

ランダム化が行われた全被験者から試験治療を一度も受けていない被験者を除くことが適切な場合もある。これらの患者を除外しても、例えば試験治療を開始するかどうかの判断が、割付けられた試験治療が何であるかに影響されない場合には、intention-to-treat の原則は保たれるであろう。ランダム化後のデータが全くない被験者をランダム化が行われた全対象集団から除外しなければならない場合もある。これらの個々の除外から生じる可能性のある偏り又はその他の偏りを明示しない限り、どのような解析も完全とはいえない。

最大の解析対象集団を用いる場合、ランダム化後に起こる治験実施計画書違反は、特にそれらの発生が試験治療の割付と関連しているときに、データ及び結論に影響するおそれがある。ほとんどの場合、ランダム化後に違反があった被験者のデータを解析に含めることが適切であり、intention-to-treat の原則とも一致する。試験治療を一度以上受けた後で試験治療を中止し、その後のデータがない被験者及び中止ではないが追跡不能となった被験者の扱いに関連して特別の問題が起きる。なぜなら、これらの被験者を最大の解析対象集団に含めないことにより、この方法が大きく損なわれるおそれがあるからである。どのような理由にせよ被験者が追跡不能となった時点で測定された主要変数の測定値、又は追跡不能となった後に治験実施計画書に従って集められた主要変数の測定値は、この観点から重要である。主要変数を追跡不能となった後に収集することは、死亡又は重大な疾患の発生を主要変数とする研究では特に重要である。このようにしてデータを収集するということは、治験実施計画書に記述すべきである。最終観察値をそれ以降の値に外挿する方法から複雑な数理モデルの使用まで、欠測値を埋め合わせる試みとして様々な技法を用いることができる。最大の解析対象集団の被験者全てについて、主要変数の測定値が利用できることを保証するために用いられる別の方法では、被験者の臨床結果又はより単純化した臨床結果（例えば、成功か失敗か）について何らかの仮定が必要であろう。これらのうちの戦略を用いる場合でも、治験実施計画書の統計の部にその理由とともに記述し、数理モデルで用いる前提は明確に説明すべきである。対応する解析結果の安定性を示すことも重要であり、用いる戦略自体が試験治療の効果の偏った推定値を与える可能性がある場合には特に重要である。

前もって予見できない問題もあるため、それらの変則的な事例を取り扱う方法に関する詳細な検討は、時期を遅らせて研究終了前に行うデータの盲検下レビュー時に行った方がよい場合もある。そのような場合は、治験実施計画書に述べるべきである。

5.2.2 治験実施計画書に適合した対象集団

ときには「'valid case'(妥当例)」、「'the efficacy' sample(有効性サンプル)」、又は「'evaluable subjects' sample(評価可能被験者サンプル)」と表される「治験実施計画書に適合した」対象集団は、最大の解析対象集団の被験者のうち治験実施計画書をより遵守している一部であり、以下のような基準から特徴づけられる。

- (i) 事前に定められた最低限の試験治療規定を完了していること
- (ii) 主要変数の測定値が利用可能であること
- (iii) 登録基準違反などの重大な治験実施計画書違反がないこと

治験実施計画書に適合した対象集団から被験者を除外する理由の詳細は、個々の試験の状況に合わせた適切な方法により、割付を明らかにする前にすべて明確にし、文書として記録すべきである。

治験実施計画書に適合した対象集団を用いることは、解析において新しい試験治療が有効性を示す可能性を最も大きくするといってもよく、治験実施計画書の基礎となっている科学的なモデルを最もよく反映することになる。しかし、対応する仮説検定と試験治療の効果の推定値は試験により保守的となる場合もならない場合もある。治験実施計画書の遵守状況が試験治療及び臨床結果と関連している場合には、生じる偏りは重大なものとなるおそれがある。

治験実施計画書に適合した対象集団を作り出すために被験者を除外しなければならないような問題及びその他の治験実施計画書違反は、十分に把握し要約すべきである。ここで言う治験実施計画書違反には、試験治療の割付間違い、禁止薬剤の使用、低い服薬遵守状況、追跡不能及び欠測値などがある。試験治療グループ間で、これらの問題の発生頻度及び発生時間のパターンを評価することはよい対処法である。

5.2.3 二つの異なる解析対象集団の役割

一般に、解析に用いる被験者集団の選択の変更を行っても、主要な試験結果が変わらないことを示すことは有益である。検証的試験では、最大の解析対象集団の解析と治験実施計画書に適合した対象集団の解析との相違を明示的な議論と解釈の対象にできるよう、通常両方の解析を計画することが適切である。解析対象集団の変更により結論がどの程度変わり易いかを、更に探索する計画を立てた方が望ましい場合もある。最大の解析対象集団の解析と治験実施計画書に適合した対象集団の解析が本質的に同じ結論に達する場合、試験結果の信用度は高くなる。しかしその場合でも、最大の解析対象集団の相当の割合を治験実施計画書に適合した解析で除かなければならない場合には、試験全体の妥当性が疑われることに注意する必要がある。

最大の解析対象集団と治験実施計画書に適合した対象集団は、（被験薬が優っていることを示そうとする）優越性試験と（被験薬が匹敵しうることを示そうとする、3.3.2節参照）

同等性試験又は非劣性試験とで異なる役割を果たす。優越性試験では、最大の解析対象集団から得られる有効性の推定値は、治験実施計画書に適合した解析ほど過度に楽観的ではないため、最大の解析対象集団が（例外的な状況を除いて）主要な解析に用いられる。なぜならば、最大の解析対象集団に含まれる遵守状況の悪い被験者は、推定される試験治療の効果を一般に小さくするからである。しかし、同等性試験又は非劣性試験では、最大の解析対象集団を使用することは一般に保守的ではないため、その役割は十分慎重に考慮すべきである。

5.3 欠測値と外れ値

欠測値は、臨床試験において偏りを起こし得る代表的な原因である。したがって、データ収集及びデータマネジメントに関しては、治験実施計画書が要求する事項について、すべてを満たすようあらゆる努力をすべきである。しかし、実際には、ほとんど常に欠測値がいくつか存在するであろう。そうであったとしても、欠測値を扱う方法が適切で、特にそれらの方法が治験実施計画書の解析計画中に事前に定められている場合には、試験は妥当であるとみなすことができる。この観点から盲検下レビューの間に統計解析計画書を見直すことにより、欠測値を扱う方法が洗練されるであろう。残念ながら、欠測値に対処する方法で、普遍的に適用可能と薦められる方法はない。欠測値に対処する方法により解析結果がどの程度変わり易いかを、欠測の数が多い場合には特に、検討すべきである。

同様の手法は、外れ値(その統計的定義はある程度恣意的だが)の影響を探索するためにも用いるべきである。特定の値が外れ値であると明確に判断するのに最も説得力があるのは、統計的に正当であるだけでなく医学的にも正当な場合であり、医学的な背景によりしばしば適切な対応が決定される例がみられる。治験実施計画書又は統計解析計画書に記載する外れ値対策は、どのような場合でも先験的にどの試験治療グループの利益にもならないようにすべきである。この観点から、盲検下レビュー中に再度解析を見直すことが有益である。治験実施計画書中に外れ値に対処する方法が記載されていない場合、実際の値を用いた解析のほかに、外れ値の影響を除くか小さくする別の解析を少なくとも一つ行うべきであり、それらの結果の間の差異を議論すべきである。

5.4 データ変換

重要な変数を変換するために必要な判断は解析の前に行い、先行する臨床試験での類似データに基づいて治験実施計画立案時に行うのが最善である。変換（例えば、平方根、対数）を行うことは、主要変数については特に治験実施計画書に明記すべきであり、その理論的根拠を述べるべきである。統計手法の前提を満たすことを保証するための変換の一般的な原則は、標準的な教科書に書かれている。また、特定の変数についての変換の慣例的方法は、多くの臨床領域別に開発されてきている。変数を変換するかどうか、変換するのであればどのように変換するかという判断は、臨床的な解釈を容易にする尺度を選択するという観点も含めて行われるべきである。

同様な配慮は、基準となる時点での値からの変化、基準となる時点での値からの変化割合、繰り返し測定「曲線下面積」、又は二つの異なる変数の比、といった新たな変数の導出の際にも行うべきである。新たな変数の臨床的解釈は注意深く検討されるべきであり、その正当性も治験実施計画書に述べるべきである。密接に関連した話題が2.2.2節に述べられている。

5.5 推定、信頼区間及び仮説検定

治験実施計画書の統計の部には、試験の主要な目的に対応した検定すべき仮説及び推定すべき試験治療の効果のどちらか又は両方を明記すべきである。主要変数については（望むべくは副次変数についても）、これらの目的を果たすための作業に用いる統計手法を記述すべきであり、基礎となる統計モデルを明確にすべきである。試験治療の効果の推定値は可能な限り信頼区間とともに提示すべきものであることから、治験実施計画書の統計の部にはこれらを求める方法を示すべきである。例えば共分散分析を用いて、精度の向上のため又は存在する可能性のある基準となる時点での差について推定値を調整するために、基準となる時点での値を使用するいかなる目的に関しても説明を加えるべきである。

片側検定を用いるか両側検定を用いるかを明確にすることは重要であり、特に片側検定の使用については前もって理由づけることが重要である。仮説検定が適切ではないと考えられる場合には、統計的結論を導くための別な手続きを提示すべきである。推測を片側と考えるか両側と考えるかには議論があり、統計学の文献にも様々な見解がみられる。承認申請のための試験では、片側検定の第一種の過誤を両側検定で慣例的に用いている値の半分に設定する方法が好ましい。このように設定することにより、試験治療間の差の大きさを推定するために通常用いられる両側信頼区間との整合性を図ることができる。

選択する個別の統計モデルには、試験の計画段階での統計的な配慮を反映させると同時に、解析する変数に関する現在の医学的及び統計的知識を反映させるべきである。（例えば、分散分析モデルで）解析に含める効果のすべてを十分に明記すべきであり、予備的な結果からこの効果の組を変更する可能性があるのであれば、そのやり方を説明すべきである。共分散分析に用いる共変量の組に関しても同様の考慮が当てはまる（5.7節も参照のこと）。統計的方法を選択する際には、主要変数及び副次変数両方の統計的分布に対して十分な注意を払うべきである。（例えばパラメトリック手法を用いるかノンパラメトリック手法を用いるかの）選択の際には、試験治療効果の大きさに関する統計的推定値を、（有意性検定に加えて）信頼区間とともに示すことが必要だと心得ておくことが重要である。

主要変数の主要な解析は、その裏付けとして行う主要変数又は副次変数の解析とは明確に区別すべきである。治験実施計画書の統計の部又は統計解析計画書には、主要変数及び副次変数以外のデータをどのように要約し報告するかについての概要も記述すべきである。その際、一連の試験にまたがる解析、例えば安全性データの解析を一貫させるためにとつたすべての手段についても言及すべきである。

既知の薬理パラメータ、個々の被験者の治験実施計画書遵守の程度又はそのほかの生物学的な根拠に基づくデータをモデル化する方法により、真の有効性又は潜在的な有効性について、とりわけ試験治療の効果の推定の点から、有益な理解が得られるであろう。そのようなモデルの前提条件は常に明確にしておくべきであり、モデルから得られる結論の限界についても慎重に記述すべきである。

5.6 有意水準と信頼水準の調整

多重性が存在する場合、臨床試験データの解析に対する通常の頻度論的立場からは、第一種の過誤の調整を必要とするであろう。多重性は、例えば主要変数が複数ある場合（2.2.2節参照）、試験治療間の多重比較、時間に伴う繰り返し評価、中間解析（4.5節参照）から生じるであろう。多重性を回避する又は減じる方法は、それが利用できる場合には好まし

いこともある。例えば、変数が複数ある場合に主要な変数の中でも重要な変数を指定すること、多群比較の場合に試験治療間の重要な対比を選択すること、繰り返し測定の場合に「曲線下面積」といった要約指標を使用すること、などである。検証的解析では、このようにして多重性を減じた後の段階でまだ残っている多重性のすべての側面について試験実施計画書において明らかにすべきである。調整は常に考慮すべきであり、調整方法の詳細、又はなぜ調整は必要ないと考えるのかという説明は、統計解析計画書に述べるべきである。

5.7 部分集団、交互作用及び共変量

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。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、探索的な部分集団別解析のみに基づいては受け入れ難い。

5.8 データの完全性の維持とコンピュータソフトウェアの妥当性

数値で表された解析結果の信憑性は、データマネジメント（データ入力、保存、確認、訂正及び復元）及びデータの統計処理の両方に用いられる方法とソフトウェア（自作、外注、市販）の質及び妥当性に依存するものである。したがって、データマネジメントは綿密で効果的な標準業務手順書に基づいて行うべきである。データマネジメント及び統計解析に用いるコンピュータソフトウェアは信頼できるものを使用すべきであり、ソフトウェアの適切な検証手順を資料として用意すべきである。

VI. 安全性及び忍容性評価

6.1 評価の範囲

すべての臨床試験において、安全性及び忍容性（用語集参照）の評価は重要な要素である。初期の相では、この評価の大部分は探索的な性質のものであり、敏感にとらえられるのは明らかな毒性の出現のみである。しかし、後期の相では、被験薬の安全性及び忍容性のプロファイルを、より多くの被験者により十分に特徴づけて確立することができる。後期の比較試験は、一般にこの点での検出力を欠いているとしても、新たな潜在的有害作用のすべてを偏りなく探索するための重要な手段を提供するものである。

ある種の試験は、他の医薬品又は被験薬の別な用量と比較して、安全性及び忍容性に関する優越性又は同等性についての具体的な主張のために計画される場合がある。このような承認に関わる具体的な主張は、対応する有効性の主張に対し要求されるのと同様に、検証的試験による適切な証拠によって確認されるべきである。

6.2 変数の選択とデータ収集

どのような臨床試験でも、医薬品の安全性及び忍容性を評価するために選ばれる方法と測定値は、多くの要因に依存する。その要因には、関連医薬品の有害作用についての知識、非臨床試験及び初期の臨床試験からの情報、個々の医薬品の薬力学的/薬物動態的特質から起こると考えられる結果、使用方法、研究対象となる被験者の特徴並びに試験の期間といったものがある。臨床化学と血液学に関する臨床検査値、バイタルサイン及び臨床的有害事象（疾患、徴候及び症状）は、通常、安全性及び忍容性データの主要部を形成する。重篤な有害事象の発生及び有害事象による試験治療の中断については、登録することが特に重要である（ICH E2A と E3 参照）。

更に、異なる臨床試験からのデータを結びつけることを容易にするために、試験プログラム全体を通して一貫したデータ収集及び評価の方法論を用いることが薦められる。共通の有害事象の辞書の使用は特に重要である。有害事象の辞書は、器官分類、基本語又は慣用語（用語集参照）という、三つの異なる水準で有害事象データを要約できるように構成されている。有害事象を要約する通常の水準は基本語であり、同一の器官分類に属している基本語は、データの記述的提示の際にまとめることができる（ICH M1 参照）。

6.3 評価される被験者集団とデータの提示

全体的な安全性及び忍容性を評価するのに用いられる被験者集団は、通常被験薬を少なくとも一回服用した被験者の集団である。安全性及び忍容性の変数は、これらの被験者から可能な限り包括的に有害事象の種類、重症度及び発現時と持続期間を含めて収集されるべきである（ICH E2B 参照）。女性、高齢者（ICH E7 参照）、重症者、又は共通の併用治療を受けた被験者といった特定の属性別集団については、恐らく更なる安全性及び忍容性評価が必要となるであろう。これらの評価では、より個別の問題に答えることが必要であろう（ICH E3 参照）。

評価の際には、すべての安全性及び忍容性変数に注意を払う必要があるため、広範な方法を治験実施計画書に示すべきである。試験治療と関係していると考えられるか否かにかかわらず、すべての有害事象を報告すべきである。評価の際には、研究対象集団の利用できるデータのすべてを用いるべきである。測定単位と臨床検査変数の参照範囲は注意深く

定義すべきである。もし異なる単位又は異なる参照範囲を同一の試験で用いるのであれば（例えば、二つ以上の検査機関が入っている場合）、統一的な評価を可能にするために測定値を適切に標準化すべきである。毒性評価尺度の使用については、事前に定め、正当化しておくべきである。

ある有害事象の発現は、通常有害事象を経験した被験者数とその有害事象を発現する可能性のある被験者数との関係を示す割合の形で表現される。しかし、発現の評価の仕方はいつも自明というわけではない。例えば状況に応じて、試験治療が使用された被験者数、又は使用の程度（人年）を分母とすることが考えられる。計算の目的がリスクの推定であるか、試験治療グループ間での比較であるかにかかわらず、その定義を試験実施計画書に示すことは重要である。この定義は、試験治療が長期にわたることが予定され、かなりの割合で試験治療の中止又は死亡が起こると予想される場合、特に重要である。そのような状況では、生存解析の方法を考慮すべきであり、過小評価を避けるために累積有害事象発現率を計算すべきである。

徴候や症状に相当の背景ノイズが存在する状況では（例えば、精神科での試験）、異なる有害事象に対するリスクの推定に背景ノイズを考慮する方法を考えるべきである。そのような方法の一つは、「試験治療下での発現」（用語集参照）という概念を用いることである。「試験治療下での発現」では試験治療前の基準となる発現状況と比べて、新たに発現又は悪化した有害事象のみを記録する。

軽度の有害事象は無視する、又は分子に加えるための基準として、繰り返しの来院で事象が観察し続けられることを要求するような背景ノイズの効果を減らすための別の方法もまた適切な場合がある。そのような方法は試験実施計画書に正当性を説明しておくべきである。

6.4 統計的評価

安全性及び忍容性の研究は多次元的な問題である。どのような被験薬についても、何らかの特定の有害作用は通常予測でき特定してモニターできるが、起こりうる有害作用の幅はたいへん広く、新しく、予想もされない作用が常に生じうる。更に、禁止薬の使用のような試験実施計画書違反の後で発生した有害事象は恐らく偏りの原因となるであろう。このような背景があることが、被験薬の安全性及び忍容性の解析的評価が統計的に困難となる原因となり、検証的試験から結論を確定するような情報を得ることをむしろ例外としている。

ほとんどの試験で、安全性及び忍容性関連事項を扱うには、記述統計の手法でデータを整理し、信頼区間が解釈の助けとなる場合にはその計算を加えることが最善である。試験治療グループ内と被験者個人内両方で有害事象のパターンが示されるようなグラフ表示を利用することもまた有益である。

p 値の計算は、関心のある特定の差を評価する補助として、又は多数の安全性及び忍容性変数に対して注目するだけの価値のある差を際立たせるための目印として、有用な場合がある。これは検査データに特に有用であり、この方法以外で検査データを適切に要約することは難しい。検査データには、例えば試験治療ごとの平均の評価のような定量的な解析と、ある閾値を超える又は下回る数を数える定性的な解析の両方を行うことが薦められ

る。

仮説検定を用いる場合、第一種の過誤を勘案して多重性を統計的に調整することは適切ではあるが、通常は第二種の過誤により注意を払うべきである。多重性の調整を行っていない場合、統計的に有意となった結果の解釈には注意すべきである。

大多数の試験で、治験責任医師たちは、実対照薬又はプラセボに比べて安全性及び忍容性に関して臨床的に許容できない差はないことを立証しようとしている。有効性に関する非劣性又は同等性評価の場合と同様に、この状況では仮説検定よりも信頼区間を使用することが望ましい。信頼区間を用いると、生起数が少ないことが原因となってみられることの多い、はなはだしい精度の低さを明確に示すことができる。

6.5 統合した要約

被験薬の安全性及び忍容性に関する特質は、一般に被験薬を開発する過程で逐次的に、複数の試験を通して要約され、特に承認申請時には必ず要約されるものである。しかし、この要約の有用性は高い質のデータを伴い適切に計画・実施された個々の比較試験に依存する。

被験薬の全体的な有用性は、常にリスクと利益のバランスの問題であり、リスクと利益の評価は通常全臨床試験プログラムを要約して行われるものであるが、単一の試験でも、有用性が見込まれるかについて検討することは可能である（7.2.2節参照）。

安全性及び忍容性に関連する報告の要求事項についての詳細は、ICH E3 12章を参照すること。

VII. 報告

7.1 評価と報告

「I. はじめに」で述べたように、総括報告書の構成と内容は、ICH E3 での主題である。ICHガイドラインE3は、臨床とその他の資料を適切に統合する統計作業の報告について十分に網羅している。したがって、この節は比較的簡潔にとどめる。

試験の計画段階では、解析の主要な特徴は5節に述べたように治験実施計画書に明記すべきである。試験の実施が終了し、データが集積されて予備的な点検ができるようになると、やはり5節で述べたように、予定した解析の盲検下レビューを実施することは有益である。この解析前に行う検討では、試験治療を盲検化した状態で、例えば被験者又はデータを解析対象集団から除外することに関する判断を行うべきである。また、変数変換の可能性の検討と外れ値の定義、最近の研究で明らかになった重要な共変量をモデルに加えること、パラメトリック手法を用いるかノンパラメトリック手法を用いるか等について再検討して差し支えない。この時点で下された判断は、報告書に記述されるべきである。一般に盲検下での判断は偏りをもたらす可能性が小さいので、統計家が試験治療のコードを知った後の判断と区別しておくべきである。割付を明らかにして行った中間解析に従事した統計家及びその他のスタッフは、盲検下レビュー又は統計解析計画の変更に参加すべきではない。また試験治療に由来する効果がデータ上明らかなことによって盲検が破れる可能性がある場合、盲検下レビューには特別の注意を必要とするであろう。

提示と作表のより詳細な内容の多くは、盲検下レビュー時、又はその前後までに固定すべきである。その結果、実際の解析時には、被験者の選択、データ選択と変換、データの要約と作表、推定と仮説検定を含むすべての解析内容についての完全な解析計画が存在することになる。データの妥当性の立証が完了したら、解析は事前に定められた計画に従って進めるべきである。事前の計画が守られればそれだけ結果の信憑性は増すことになる。治験実施計画書、治験実施計画書の改訂、又はデータの盲検下レビューに基づいて更新された統計解析計画に記述されている、予定した解析と実際の解析の間のどんな差にも特別な注意を払うべきである。予定した解析からの逸脱に対しては、慎重な説明を行うべきである。

試験に登録されたすべての被験者は、解析に含まれているかどうかにかかわらず、報告書で明らかにすべきである。解析から除外した理由はすべて記録すべきである。最大の解析対象集団には含まれるが治験実施計画書に適合した対象集団には含まれない被験者については、治験実施計画書に適合した対象集団から除外された理由も記録すべきである。同様にして、ある解析対象集団に含まれる被験者すべてについて、すべての重要な変数のすべての適切な時点での測定値も明らかにしておくべきである。

被験者又はデータの減失、試験治療の中止及び重大な治験実施計画書違反が主要変数の主な解析に及ぼす影響について、慎重に考慮すべきである。追跡不能、試験治療の中止、又は重大な治験実施計画書違反があった被験者は明らかにすべきであり、減失の理由、減失と試験治療及び結果との関係を含めた、記述的解析を行うべきである。

記述統計は報告書に欠くことのできない部分である。適切な表若しくはグラフ表示、又はその両方により、主要変数及び副次変数並びに主な予後変数及び人口統計学的変数の重要な特徴を明確に説明すべきである。試験の目的に関連する主な解析結果には、特別に慎重な記述的報告を行うべきである。有意性検定の結果を報告する際には、限界値を越えたかどうかではなく、正確なp値（例えば「 $p=0.034$ 」）を報告すべきである。

臨床試験の解析の主たる目標は、その主目的から生じた問題に答えることであるべきだが、割付を明らかにした後の解析で、観察されたデータに基づいた新たな問題が生じるおそれがある。追加解析、そして恐らく複雑な統計解析がその結果として必要になるであろう。この追加解析は、治験実施計画書に予定していた解析の報告とは厳密に区別すべきである。

偶然により、多少とも予後に重要な意味を持つが、共変量として事前に定めていなかった基準となる時点の測定値に関して、試験治療グループ間の予期しないバランスのくずれが起こる可能性がある。このバランスのくずれに対処するには、それを考慮する解析を追加し、予定した解析と本質的に同じ結論に達することを示すことが、最善である。しかし両者が同じ結論に達しない場合には、バランスのくずれが結論に与える影響を議論すべきである。

一般に、予定していない解析は最小限にとどめるべきである。予定外の解析は、試験治療の効果が何か別の要因に応じて変化する可能性があると考えられる場合に実施される例がよくみられる。その場合、被験者集団の中で試験治療効果が特に有益な部分集団を明らかにする試みがなされるであろう。よく知られているように、予定していない部分集団別

解析を拡大解釈することは潜在的に危険なことであり（5.7節を参照）、慎重に避けるべきである。試験治療が被験者のある部分集団に何の利益もないと思われる場合、又は有害作用があるように思われる場合にも同様の問題が生じるが、それらの可能性について正しい評価を行うべきであり、したがって常に報告すべきである。

最後に、臨床試験の結果の解析、解釈及び提示には統計的判断が下されるべきである。このためには、試験統計家は総括報告書に責任を持つチームの一員であるべきであり、総括報告書を承認する存在であるべきである。

7.2 臨床データベースの要約

報告されているすべての臨床試験での安全性及び有効性の証拠を、全体的に要約して総合することは、承認申請の要求事項である（EUでの「専門家報告書(Expert Report)」、米国での「総合要約報告書(Integrated Summary Reports)」、日本での「資料概要」）。これは、適切であれば、結果の統計的結合を伴ってもよい。

要約の中には、以下のような多くの領域において特定の統計的興味が発生する。一連の臨床試験プログラムにおいて試験治療を受けた集団の人口統計学的及び臨床的特徴を記述すること。適切な（通常、比較を伴う）試験の結果を考慮にいれて有効性に関する主要な問題に答えること及び治療が相互に補強しあう、又は矛盾しあう程度を際立たせること。承認申請の根拠となるすべての治療を結合したデータベースから利用できる安全性情報をまとめること及び安全性に関する潜在的な問題を確認すること。臨床プログラムの設計段階では、特に試験を通じてよく結合される測定値に対して、後になって一連の試験の解釈が容易になるように測定値を統一して定義し収集するよう、慎重な注意を払うべきである。薬物治療、医療歴及び有害事象の詳細を記録するために、共通の辞書を選んで使用すべきである。主要変数と副次変数を共通して定義することはほとんど常に有益であり、メタアナリシスのためには必須である。主要な有効性変数の測定方法、ランダム化時/登録時から評価を行うまでの期間、治療実施計画書違反例と逸脱例の取り扱い及び恐らく予後因子の定義についても、そのようにしない妥当な理由がない限り、矛盾なく保つべきである。

いくつかの試験を通してデータを結合するために用いた統計手法はすべて、詳細に記述すべきである。結合の際には、試験を選択することに伴う偏りの可能性、結果の一様性及びばらつきの様々な原因を適切にモデル化することに注意を払うべきである。また、用いた仮定及び試験の選択に対して結論がどの程度変わり易いかを探索すべきである。

7.2.1 有効性データ

一つ一つの臨床試験は、常にその目的を果たせるだけの十分な規模で行うべきである。本質的に同一であるような主要な有効性の問題を扱っている一連の臨床試験を要約することで、更なる有益な情報が得られるであろう。そのような一連の試験の主たる結果は、通常は推定値と信頼限界を中心とする表又は図として、比較可能な同一形式で提示すべきである。これらの推定値を結合するためのメタアナリシス技法の使用が有用な補足となる例がよくみられる。なぜならば、メタアナリシス技法は試験治療効果の大きさに関するより精度の高い包括的な推定値を与え、試験の結果の完備した簡潔な要約を与えるからである。例外的な状況下ではあるが、全体的な仮説検定を行うメタアナリシス手法が、有効性の十分な全体的証拠を与える最も適切な方法、又は唯一の方法となる場合もある。このような目的でメタアナリシス技法を用いる場合には、メタアナリシスを実施するための研究計画

を前もって準備すべきである。

7.2.2 安全性データ

安全性データの要約では、潜在的な毒性を示すいかなる徴候に対しても徹底的に安全性データベースを調べることが重要であり、裏づけるパターンを探索してその徴候を追跡することが重要である。医薬品のあらゆる人体曝露に関する安全性データを結合することが、重要な情報源となる。なぜならば、そのようにして被験者数が多くなることによりまれな有害事象を検出し、恐らく有害事象のおおよその発現数を推定する可能性が最も高くなるからである。しかし、このようなデータベースからの有害事象発現データは、比較するグループを欠いていることから評価が困難であり、この困難さを克服するためには比較試験からのデータが特に有益である。それぞれの対照薬について十分なデータを提供するため、共通の対照薬（プラセボ又は特定の実対照薬）を用いている試験の結果は結合し、別々に提示すべきである。

データの探索から判明した毒性を持つ可能性を示す徴候はすべて報告すべきである。これらの潜在的有害作用がどれだけ現実に起こりうるかの評価には、多数の比較の実施によって生じる多重性の問題を考慮すべきである。評価には、有害事象の発生に曝露期間若しくは追跡期間又はその両方が潜在的に関連しているかどうかを探索するため、生存解析手法を適切に使用すべきである。確認された有害作用に関連するリスクは、リスクと利益の関係を正しく評価するために適切に定量化すべきである。

用語集

ベイズ流の方法 Bayesian Approaches

一部のパラメータ（例えば、試験治療の効果）について、観察されたデータとパラメータの事前確率分布から事後確率分布を与えるデータ解析の方法。得られた事後分布を、統計的推測の基礎とする。

偏り（統計的及び運営上の）Bias (Statistical & Operational)

臨床試験の計画、実施、解析及び結果の評価と関連した因子の影響により、試験治療の効果の推定値と真の値に系統的な差が生じること。試験実施中の逸脱によって起こる偏りを「運営上の」偏りとよび、それ以外の原因によって起こる偏りを「統計的」偏りとよぶ。

盲検下レビュー Blind Review

試験完了（最後の被験者の最終観察）から割付を明らかにするまでの間に予定した解析を固定する目的で実施するデータの評価と点検。

内容的妥当性 Content Validity

ある変数（例えば、評価尺度）で測定しようとしているものを、どれだけ間違いなく測定しているか、その程度。

ダブルダミー Double-Dummy

臨床試験で二つの試験治療の区別がつく場合に、医薬品投与時の盲検を維持する技法。試験治療Aについて、実際の製剤と共にそれと区別不能なプラセボを用意し、また試験治療Bについても、実際の製剤と共にそれと区別不能なプラセボを用意する。被験者は二組の

試験治療(一つはAの実際の製剤とBのプラセボ、もう一つはAのプラセボとBの実際の製剤)のどちらかを受ける。

脱落 Dropout

試験実施計画書が要求する最終観察以前に、何らかの理由で臨床試験の継続ができない被験者。

同等性試験 Equivalence Trial

二つ以上の試験治療に対する反応が、臨床的に重要な意味を持つほど異ならないことを示すことが主要な目的の試験。このことは、通常臨床的に許容できる差である上側同等限界と下側同等限界の間に、試験治療間の真の差が存在する可能性が高いことを示すことにより証明される。

頻度論的方法 Frequentist Methods

有意性検定及び信頼区間といった統計的方法であり、その意味は同一実験状況下という仮説的な繰り返しのもとで起こるある結果の頻度という観点から解釈できる。

最大の解析対象集団 Full Analysis Set

Intention-to-treat の原則に可能な限り近づけた被験者集団。最大の解析対象集団は、ランダム化が行われた全被験者から、除くべき理由のある最低限の被験者を除外した集団である。

一般化可能性、一般化 Generalisability, Generalisation

臨床試験で得た知見を、その試験に参加した被験者からより広い患者集団とより広い医療現場へ外挿することが信頼をもってできる程度。

総合評価変数 Global Assessment Variable

被験者の疾患の状態又は疾患の状態の変化についての客観的変数と治験責任(分担)医師の全体的な印象を統合した、通常、順序カテゴリの評価尺度である単一の変数。

独立データモニタリング委員会効果安全性評価委員会(データ及び安全性モニタリング委員会、モニタリング委員会、データモニタリング委員会) Independent Data Monitoring Committee (IDMC) (Data and Safety Monitoring Board, Monitoring Committee, Data Monitoring Committee)

臨床試験の進行状況、安全性データ及び重要な有効性評価項目を何回かにわたって評価するとともに、治験依頼者に試験の継続、修正、又は中止を勧告するために、治験依頼者が必要に応じて設立する委員会。

Intention-To-Treat の原則 Intention-To-Treat Principle

治療に用いる治療方針により得られる効果は、実際に受けた試験治療ではなく、被験者を治療しようとする意図(予定した試験治療規定)に基づくことにより最もよく評価できる、ということを主張する原則。この原則から、一つの試験治療グループに割付けられた被験者は、予定した試験治療のコースを遵守したかどうかにかかわらず、割付けられたグループのまま追跡され、評価され、解析されるべきであることが導かれる。

交互作用(質的及び量的) Interaction (Qualitative & Quantitative)

試験治療の対比（被験薬と対照薬との差など）が、他の要因（施設など）により変わる状況。量的な交互作用とは、要因のレベルが異なるとそれに応じて対比の大きさが変わることであり、質的な交互作用とは、要因のレベルの少なくとも一つにおいて、対比の方向までもが変わることである。

評価者間信頼性 Inter-Rater Reliability

異なる評価者が異なる機会に評価をする場合、同じ結果を与える特性。

評価者内信頼性 Intra-Rater Reliability

同一評価者が異なる機会に評価をする場合、同じ結果を与える特性。

中間解析 Interim Analysis

試験の正式な完了以前に、有効性又は安全性に関して試験治療群間を比較することを意図して行われるあらゆる解析。

メタアナリシス Meta-Analysis

同じ問題を扱う二つ以上の試験から得られる定量的な証拠について形式に則って行う評価。最も一般的なメタアナリシスでは、様々な試験の要約統計量を統計的に結合するが、生データを結合する場合もメタアナリシスと呼ぶ場合がある。

多施設共同試験 Multicentre Trial

単一の試験実施計画書に基づいて、二つ以上の施設で、したがって二人以上の試験責任医師によって実施される臨床試験。

非劣性試験 Non-Inferiority Trial

被験薬への反応が比較薬剤（実薬又はプラセボ）よりも臨床的に劣らないことを示すことが主要な目的の試験。

基本語及び慣用語 Preferred and Included Terms

MedDRAのような階層的医学辞書では、慣用語とは試験責任医師の記述がコード化される最下層の辞書用語である。基本語とは、発生頻度を報告するために用いられるものであり、慣用語をグループ化するレベルである。例えば、「左腕に痛み」という試験責任医師の記述は、慣用語では「関節痛(Joint Pain)」としてコード化され、基本語レベルでは「関節痛(Arthralgia)」として報告される。

試験実施計画書に適合した対象集団(妥当例、有効性サンプル、評価可能被験者サンプル) Per Protocol Set (Valid Cases, Efficacy Sample, Evaluable Subjects Sample)

データの集合であり、そのデータは基礎となる科学的モデルに従い試験治療の効果をよく示すと十分考えられる程度に試験実施計画書を遵守した部分集団から得られる。遵守には、試験治療への曝露、測定値の利用可能性及び大きな試験実施計画書違反がないことが含まれる。

安全性及び忍容性 Safety & Tolerability

医療用医薬品の安全性は、臨床試験では通常臨床検査（臨床化学、血液学を含む）、バイタルサイン、臨床的有害事象（疾患、徴候、症状）、その他特別な安全性検査（心電図、眼科学など）によって評価される、被験者の医療上のリスクに関するものである。医療用

医薬品の忍容性とは、明白な有害作用が被験者にとってどれだけ耐えうるかの程度を示す。

統計解析計画書 Statistical Analysis Plan

統計解析計画書とは、治験実施計画書に記されている解析の主要な特徴のより技術的な詳細を述べた文書であり、主要変数、副次変数、その他のデータに関する統計解析を実行するための詳細な手順を含むものである。

優越性試験 Superiority Trial

被験薬への反応が比較薬剤（実薬又はプラセボ）よりも臨床的に優れることを示すことが主要な目的の試験。

代替変数 Surrogate Variable

臨床的効果を直接測定することが实际的でない場合に、効果の間接的な測定値を示す変数。

試験治療の効果 Treatment Effect

臨床試験における試験治療の効果。ほとんどの臨床試験では、関心のある試験治療の効果は、二つ以上の試験治療間の比較（又は対比）である。

試験治療下での発現 Treatment Emergent

試験治療前には存在しておらず試験治療期間に出現した事象、又は試験治療前の状態に比べて悪化した事象。

試験統計家 Trial Statistician

本ガイドライン中の原則を実行するために、十分な理論又は実地の教育及び経験を併せ持ち、かつ当該試験の統計的側面に責任を持つ統計家。

「臨床試験のための統計的原則」に関する質疑応答

Q1. 本ガイドラインで定義される試験統計家の要件とは具体的にどのようなものか。

(答)

現時点では、ガイドラインの「十分な理論又は実地の教育及び経験を合わせ持ち、臨床試験における統計的側面に責任をもつことができる者」という以上に具体的な要件は定められていない。治験依頼者は、ガイドラインの趣旨を踏まえ、生物(医学)統計学に関連して、大学等の教育機関において受けた教育の内容、卒業後の研修・訓練の状況、研究業績、実際の臨床試験に対してどのような立場でどの程度関わったかといった経験等を総合的に判断して個々の試験での試験統計家の選定に当たっていたきたい。

また、「治験の総括報告書の構成と内容に関するガイドライン(平成8年5月1日薬審第335号)」では、試験に関与した者の履歴等の文書を作成することを求めている。審査の際には必要に応じ、どのような者が試験統計家として選定されたか確認するため、これらの資料の提出を求めることがある。

Q2. 片側検定又は両側検定のどちらを用いるか、またそこでの有意水準をいくらにすべきかを、優越性試験と非劣性試験のそれぞれで説明願いたい。

(答)

ガイドラインでは、同等性を示す場合には両側信頼区間、非劣性試験では片側信頼区間による解析を行うことが記載されているが、一般には推測を片側と考えるか両側と考えるかには議論があり一概に決められるものではないとされている。また、有意水準についても、個々の試験において適切な基準を設定すべきである旨の記載がある。

しかしながら、推論を片側とするか両側とするかにより統計的な判断に大きな差異が生じることは規制上の観点から望ましくない。また、一方で、臨床試験における有効性の評価では、検定により有意差があるか否かを判断するだけでなく、試験治療効果の大きさ(比較群間の差の大きさ)がどの程度であるかを推定することも重要である。そこで、今後は、検証的試験においては、仮説の検定においてどちらの方法を用いる場合であっても、効果の推定には95%信頼係数の両側信頼区間を用い、検定の際の有意水準は、これによる判断との整合性を図るため、優越性試験、非劣性試験のいずれにおいても、片側2.5%又は両側5%とすることを原則とする。用量反応試験についても、用量反応性を示すことにより薬剤の有効性を検証するような試験においては上記と同様である。ただし、適切な説明ができるのであれば、より強固な有効性の根拠を示すために有意水準を厳しくする、稀少疾病用医薬品にみられる例のように十分な被験者を集めることが困難な場合は有意水準を緩くする、などの措置をとってもよい。なお、生物学的同等性試験については、「後発医薬品の生物学的同等性試験ガイドライン(平成9年12月22日医薬審第487号)」により、90%信頼係数の両側信頼区間を用いるとされているが、臨床効果を指標に標準製剤との同等性を検証しようとする場合(臨床的同等性試験)は、上記と同様に95%信頼係数の両側信頼区間を用いることを原則とする。

Q3. 2.1.2には「一つの検証的試験からの証拠だけで十分であることもあり得る」とあるが、申請までに通常複数の検証的試験が必要であると考えなのか。

(答)

承認申請に当たっては、検証的試験により承認に関わる主張の裏付けとなる十分な証拠を提示する必要があるとともに、実際に薬剤が使用される状況や適用の範囲等を考慮して、意図している患者集団に対する一般化の根拠を説明できることが重要である。したがって、検証的試験が一つの場合には、承認の根拠となる証拠が十分に認められるか又は他の資料から一定の根拠が得られること等が必要であろう。また、一般化の根拠についても十分な説明が必要であり、例えば、少数の施設で限定された状況において行われた一つの検証的試験のみで広範囲での使用を予定した申請を行う場合には、特にその根拠を十分説明できることが必要である。

Q4. 多施設共同試験においては、どのような考え方で施設当たりの被験者数を設定すればよいか。

(答)

多施設共同試験に関しては、ガイドラインでは主に固定効果モデルを前提に議論されている。被験者数の設定については、施設効果及び試験治療と施設の交互作用を推定し、不均一性がみられた場合、試験の運営管理や被験者の特徴といった面から説明できるか十分に調べるべきであることから、施設当たりの被験者数を多くすることが原則であり、同時に施設間で被験者数に大きな差が生じないように施設の選定に十分配慮する必要がある。具体的な被験者数は対象となる疾患や試験治療により異なるが、一群10例以上が一つの目安である。また、施設当たりの被験者数が少なくなる場合であっても、施設当たりの被験者数に大きな差が生じることのないよう配慮する必要がある。施設当たりの被験者数が少なくなる場合には、施設数が多くなることにより、混合モデルを用いることが適切な場合があるが、施設当たりの被験者数が少ないと試験の質及び盲検性を確保することが難しくなることが指摘されており、解析の結果不均一性がみられた場合であっても施設の運営管理や被験者の特徴と結びつけることは一般には困難であることに注意すべきである。

いずれにしても、これらの検討結果を一般化の議論と結びつけるためには、実際の臨床現場ではどのような治療が行われるかを念頭に置き、どのような施設を選定するかが重要である。

なお、ガイドラインで記載しているように、ここでいう施設は必ずしも一つの医療機関を指すものではない。特定の複数の医療機関について、試験実施の観点からは一つの医療機関とみなすことが可能とする適切な根拠があるときは、これらの医療機関をまとめて一施設として取り扱うことができる。

Q5. 「臨床試験の統計解析ガイドライン(平成4年3月4日薬新薬第20号)」では第 相試験において2群比較を推奨していたが、3.3.2の記述は、この方針を変更して3群以上の比較試験を推奨していると考えるべきか。

(答)

3.3.2で記載されているとおり、例えば実薬を対照として行われる同等性試験又は非劣性試験においては、プラセボを加え3群の比較試験を行うことにより、重要な情報を得られる場合がある。同等性試験や非劣性試験に限らず、試験計画を立案する際には比較の型式についての十分な検討を行うことが重要であり、必要に応じてどのような計画を用いるか決定すべきである。したがって、2群比較のみを推奨するということはない。

なお、臨床試験における対照群の選択の問題は、ICHにおいて別途設けられた専門家委員会(E10)で検討が行われているので、そちらも参照されたい。

Q6. 同等性又は非劣性試験での同等限界はどのように設定すればよいか。

(答)

同等限界は、疾患の領域や薬剤の性質、評価変数が計数値か計量値かなどを考慮し、臨床的な見地から、それぞれ設定すべきであり、領域毎に専門的な合意が得られていることが望ましい。そうでない場合には、申請者が個々の臨床試験において臨床的に適切と考えられる値を設定することになるが、承認申請においては、設定した同等限界の妥当性の根拠を明示し、説明できることが重要である。実薬対照の非劣性試験を行う場合には、少なくともプラセボとの差が明確となる範囲を設定する必要がある。

Q7. 3.2では「試験治療の主効果は、最初に施設と試験治療の交互作用を含まず施設間差を考慮に入れるモデルを用いて調べることができる」とあるが、交互作用の有無を最初に検討しないのはなぜか。

(答)

本ガイドラインは、試験治療効果をまず調べるという立場で記載されており、このような考え方に立って、施設と試験治療の交互作用を含めず施設間差を考慮に入れるモデルを用いて調べる方法が記載されている。これは、交互作用を考慮しなくてよいという意味ではない。むしろ、このようなモデルで解析を行った場合には、仮に交互作用が存在する場合であっても、その影響を上回る差が試験治療間に存在することが必要である。

主効果の存在が確認された場合は、交互作用を調べ、交互作用が認められた場合には結果の安定性を確認することが必要である。特に質的な交互作用が認められた場合にはその理由を考察すべきである。理由が十分に説明できない場合は、追加の試験が必要となる。

また、主効果が認められないときには、仮に交互作用の検討の結果、試験治療がある条件の下に有効であることが示唆されるような場合であっても、新たにそのような条件を考慮した検証的試験を行うことが必要である。

Q8. 本ガイドラインにある二つの解析対象集団（最大の解析対象集団と治験実施計画書に適合した対象集団）のどちらを主とするかをどのように選択すればよいか。

(答)

基本的には本ガイドラインでは最大の解析対象集団を主要な解析対象集団とすることを推奨しているが、同等性試験又は非劣性試験においては、最大の解析対象集団を用いることが必ずしも保守的であるとは言えないことから、その役割を十分慎重に考慮した上で判断すべきである。

一般に、二つの解析集団が著しく乖離するのは、治験実施計画書が遵守されていない場合、治験実施計画書に不備がある場合などが考えられる。どちらも試験の信頼性を損なうものであることから、試験の計画、実施に当たっては二つの解析集団ができるだけ一致するよう努力すべきである。このためには、すべての被験者を可能な限り追跡することも重要である。また、何らかの原因により二つの解析対象集団に乖離が生じた場合には、乖離の原因を明らかにしてそれが結果の偏りをもたらす可能性を吟味し、さらに二つの集団で解析結果がどのように異なるかを検討する必要がある。

Q9. 被験者の解析上の取扱いはどの時点までに決定しておくべきか。

(答)

被験者の解析上の取扱いは、原則として事前に治験実施計画書に記載しておくべきである。しかし、計画書の作成段階では取扱いを定めることができない事項、又は実施中の情報により取扱いを見直さなければならない事項は、盲検下で検討を行い、その取扱いを定めることになる。

なお、従来症例検討の際に慣例的に行われてきたように、試験計画書に記載された取扱いの基準を盲検下レビューの際に緩和することは望ましくない。試験開始後に、変更又は新たに定める取扱い事項が多いことは試験の妥当性を大きく損なうことに注意すべきである。

非盲検比較試験の場合であっても、被験者への安全上の対策を損なわない範囲で、割付方法、症例報告書への割付薬剤の記載方法、モニタリング方法等を工夫し、可能な限りモニター、データマネジメント担当者、解析担当者等が試験治療の割付を知ることがないようにするとともに、被験者の解析上の取扱い及び試験計画の見直しを行う場合には、可能な限り盲検下で行うべきである。

Q10. G C P では統計解析計画書に関する記載はないが、統計解析計画書は申請資料の中でどのように位置づけられるのか。

(答)

統計解析計画書は承認申請の必須文書ではないが、治験実施計画書を補足し、解析の詳細を記述したものであるから、要求があれば提出できる形で作成しておくべきである。統計解析計画書の改訂を行っている場合はその履歴も同様である。

なお、外国臨床試験データが申請資料とされている場合には、統計解析計画書の翻訳及び説明を要求することがある。