

件下で、被験薬と実薬対照を比較することが可能となる（2.1.5.1.1節参照）。

2.0 各種の対照に対する詳細な検討

2.1 プラセボ対照

2.1.1 特徴（1.3.1参照）

プラセボ対照試験においては、被験者は、ほとんどの場合、ランダム化により被験治療又はプラセボに割り付けられる。プラセボは、色、重さ、味及び匂いといった物理的特性を可能な限り被験薬に似せた、試験薬を含まない「ダミー」の治療である。試験によっては、被験治療の複数用量を調べたり、実対照薬とプラセボの両方を含むこともある。このような場合、治験実施医師にとっては全ての治療の見かけを同じにしようとするよりも、複数のプラセボ（「ダブルダミー」）を用いる方が容易であろう。プラセボの使用により二重盲検（二重マスク）が容易となり、またほとんどの場合、プラセボを使用した試験は二重盲検下で行われる。実薬治療群とプラセボ群間の結果の差は、その試験条件の下での治療効果の測度となる。プラセボ対照試験と一般的に言っても、その中には、並行群間比較又はクロスオーバーデザイン（ICH E9 ガイドライン参照）、実薬群での固定単一用量又は用量調節、複数の固定用量等、効果的に利用可能な多くの種類のデザインが含まれる。ここでは、特に注意を払うべきいくつかのデザインについて説明する。なお、プラセボを含む試験が全てプラセボ対照試験というわけではないことに注意する。例えば、実薬対照試験では盲検性を容易にするために各薬剤に対応するプラセボを用いる場合があるが（ダブルダミー）、これは実薬対照試験であり、プラセボ対照試験ではない。プラセボ対照試験は、プラセボによる治療が被験薬による治療と比較される試験である。

すべてのプラセボが、まったく活性を持たないわけではないことにも注意しなければならない。例えば、局所用皮膚製剤の試験で用いられるある種の基剤対照は患者にとって有益な作用を示すことがあるが、被験薬自体の効果を測定するためのデザインの持つ力を損なうことはない。一方、選択された基剤対照が患者に悪影響をもたらす可能性がある場合には特別な問題が生じる。この場合には、「無治療」群を含めることで、被験薬にその基剤を加えた全体としての効果を測ることができる。

2.1.2 バイアスを最小にする力

プラセボ対照試験では、ランダム化と盲検化が用いられるため、被験者及び治験実施医師のバイアスは一般に最小となる。しかしそのような試験でも、一方の治療の薬

理作用が識別されることにより盲検性が破れることまでは防げない。そのような場合には、結果の評価を盲検下で行うことでバイアスを小さくできる。盲検性が破れる懸念は、クロスオーバー試験の場合に特に問題となる。

2.1.3 倫理上の問題

有効な治療が知られていない病態に対して新治療が試験される場合には、通常、新治療とプラセボを比較する試験に倫理上の問題は生じない。しかしながら、計画している試験で調べようとしている病態に有効な薬があるときには、プラセボ対照を使用すると倫理上の問題、試験が受け入れられるかどうかの問題、試験の実施可能性の問題が生じる。研究対象母集団において死亡や回復不能の障害のような重要な障害を防ぐ治療が利用できることが知られている場合には、プラセボの使用は一般に不適當である。しかし、例えば標準治療が非常に強い毒性を示すために多くの患者がその治療を受けることを拒否する場合等、場合によっては例外もある。

それ以外の場合、すなわち重大な障害が生じない場合には、たとえ患者が結果として不快・不便（discomfort）を経験する可能性があるとしても、プラセボ対照試験への参加を患者に依頼することは一般的に非倫理的ではないと考えられる。もちろんこれには、試験の実施状況が強制的、強圧的ではなく、全ての治療の選択肢と治療を遅らせることにより起こりうる結果が患者に十分に説明されているという前提が必要である。しかし、たとえ倫理的であっても、そのような試験では重大な実施上の問題が生じる可能性がある。例えば、痛みや他の症状の治療を遅らせることが患者や医師に受け入れられないかもしれず、その場合には彼らはそのようなことを求める試験への参加を望まない可能性がある。既知の有効な治療が知られている場合に、ある新薬のプラセボ対照試験が被験者及び治験実施医師に受け入れられるかどうかは、治験実施医師、患者及び治験審査委員会（IRB）／倫理委員会（IEC）の判断の問題であり、また、受け入れ可能性はICHの地域間で異なるかもしれない。以下で議論されるように（2.1.5節参照）、受け入れ可能性は試験ごとのデザイン、選択される患者母集団に依存する。

特定のプラセボ対照試験が倫理的か否かが、臨床的に証明されていると信じられている「事実」や試験の個別状況に依存することもある。例えば、末梢器官に合併症のない軽度の本態性高血圧症での新しい降圧薬の短期プラセボ対照試験は一般に受け入れられると考えられるが、一方、より長期の試験やより重症の患者を含む試験はおそらく受け入れられないであろう。

プラセボ又は無治療対照の使用が、患者が治療を全く受けないことを意味するわけではないことは強調されるべきである。例えば、承認された治療薬が存在しない腫瘍領域の試験では、プラセボ又は無治療群の患者も、被験薬群の患者も共に、必要な緩和療法（例：鎮痛薬）及び最良のサポーターティブ・ケアを受けることになる。多くのプラセボ対照試験は「上乘せ試験」として実施される。この場合、全ての患者が、特定の標準治療又は治療に当たる医師・施設に任せられた治療を受けることになる（2.1.5.2.1節参照）。

2.1.4 特定の条件でのプラセボ対照試験の有用性及び推測の妥当性

プラセボ対照試験は、治療の有効性を示すために用いられる場合、仮定や外部（試験外）情報へ依存する度合いが最も少ない試験である。プラセボ対照試験での試験のデザイン上又は実施上の問題点のほとんどは、治療差を示せない（それゆえ有効性を示せない）可能性を高めるため、プラセボ対照試験では優れた試験を行おうとするインセンティブが自ずと働く。試験の主目的が二種類の実薬の比較又は用量反応の評価である場合でも、プラセボ群を設置することにより、他の比較から導きうる推測をより確かなものとする内的な基準が与えられる。

また、プラセボ対照試験は、薬剤によって生じる副作用と基礎疾患又は合併症から生じる症状を最もよく区別する。しかし、例えば、二つの治療の類似性を確かめる試験を行い、ある薬剤で治療した患者とプラセボで治療した患者の間での特定の副作用の発生率が同程度であることを示すことにより、その薬剤が副作用を有さないことを証明する場合には、プラセボ対照試験であっても、同等性又は非劣性試験と同様の分析感度の問題が生じることに注意しなければならない（1.5.1節参照）。すなわち、仮に被験薬が有害事象を引き起こしたとして、その試験でその有害事象が間違いなく検出できたはずであることがわかっている場合のみ、得られた結果を正しく解釈できる。通常そのような試験には、問題とする有害事象を確実に引き起こす実薬対照による治療を含んでいる必要がある。しかし、特定の試験デザインにおける「有害な薬剤の作用に対する既存の感度」を文献的に示すことにより、有害作用に対する分析感度を持つと結論できる場合もあるだろう。

2.1.5 倫理上、実施上又は推測上の問題を解決しうるデザイン上の修正及び他の対照との組み合わせ

プラセボ対照試験の推測上の利点を保ちつつ、試験デザインに修正を加えることで、

倫理上又は実施上の制約に対処できることが多い。加えて、被験薬や既知の実対照薬の用量を複数にすること等、追加治療群を試験に加えることにより、プラセボ対照試験をより情報に富んだものとすることができる。

2.1.5.1 対照群の追加

2.1.5.1.1 三群比較試験；プラセボと実薬対照

1.5.1節で述べたとおり、プラセボ対照に実薬対照を加えた三群比較試験では、被験治療とプラセボとの差を見出せなかった場合に、被験治療に効果がなかったのか、それとも単に試験に実薬を区別する力がなかった結果なのかの評価が容易である。そのような試験でのプラセボと標準薬との比較は分析感度の内的な証拠を与える。実薬同士の比較の精度を高めることが重要と考えられる場合には、実薬群の被験者数をプラセボ群より多くすることも可能である。これはまた、被験者がプラセボ群にランダム化される可能性を減らすので、患者及び治験実施医師にとって試験をより受け入れ易いものにする。

2.1.5.1.2 用量の追加

プラセボに加えて、被験薬のいくつかの固定用量にランダム化を行うと、用量反応の評価が可能となる。このようなランダム化は、治療間の公平な比較を保証するための比較試験で特に有用なものになる（ICH E4ガイドライン参照）。

2.1.5.1.3 要因デザイン

要因デザインは、被験薬の複数用量を単独治療、及びその被験薬との併用が予定されている他の薬の複数用量と組み合わせて調べるために用いられる。この種の試験は、一つの試験で広い用量範囲にわたる組み合わせの性質を明らかにできる。新規の降圧薬の評価ではそのような試験は普通に行われているが、さらに、複数の治療が同時に使われる様々な状況の下で採用を考慮することができよう。例えば、心臓発作後の死亡の予防においてアスピリンとストレプトキナーゼそれぞれの効果が相加的であることが、要因デザイン試験で示されている。

2.1.5.2 その他の試験デザイン上の修正

2.1.5.2.1 プラセボ対照上乗せ試験；置換え試験

上乗せ試験は、標準治療を既に受けている患者で実施される新薬のプラセボ対照試

験である。利用可能な治療が死亡や回復不能の障害を減少させることが知られている場合、実薬対照として標準治療を用いた非劣性試験が実施できない場合、非劣性試験の結果の解釈が難しい場合（1.5節参照）に、上乘せ試験は特に重要である。抗腫瘍剤、抗てんかん薬、抗心不全薬の試験で、このような方法が採られることが多い。このデザインは標準治療が十分には有効でない場合にのみ有益であり（しかし、実際に、標準治療が十分有効でない場合は多い）、単なる非劣性ではなく臨床的な結果の改善の証拠を与えるという利点をもっている。もちろん、そのような試験で有効性が証明されるのは組合わせ治療についてのみであり、組合わせ治療で有効な用量は、単一治療での用量とは異なる可能性がある。一般に、この方法は、新治療と標準治療が異なった薬理的作用機序を持つ場合にのみ成功の可能性があると考えられるが、例外はある。例えば、AIDS患者に対する組合わせ治療は、耐性の発現を遅らせることによって、関連した薬理学的機序を有する薬剤の有益な効果を示すであろう。

このデザインの変形版であり、時に単独治療に関する情報を与えることができ、特に慢性疾患の場合に適用できるのが置換え試験（replacement study）である。この試験では、既存治療がその有効用量で用いられているところに、新薬又はプラセボをランダム化により上乘せし、次いで、通常は、既存治療の量を少しずつ減らしていく。そして、予め定めた判定基準に従って、被験者のベースラインの状態を維持できるかどうかを被験群及びプラセボ群で観察される。この方法は、ステロイド依存の患者でステロイドを減量するための代替薬の試験において、試験の最初にステロイドを中止する必要がなく、ウォッシュアウト期に症状の再発を生じない方法として用いられてきた。また、この方法は抗てんかん薬の単独治療の試験にも用いられてきた。

2.1.5.2.2 「早期離脱」；レスキュー治療

無効な治療からの早期離脱（early escape）を計画に折り込んだ試験をデザインすることも可能である。早期離脱とは、臨床症状が悪化したり、決められたレベルまで改善しない被験者（例えば、血圧が事前に決めた時期までにコントロールできない、てんかん発作の頻度が定められた値を超えた、血圧があるレベルまで上昇した、狭心症の頻度が予め定めたレベルを超えた、肝炎患者の肝酵素値が決められた時点までに正常化していない等）、その治療で防ごうとした事象（例えば、不安定狭心症の再発、てんかん大発作、発作性上室性不整脈等）が試験中最初に一回生じた被験者、又は他のレスキュー治療が必要となった被験者を、早急に試験から離脱させることをいう。そのような場合、治療を変更する必要が生じたことが試験のエンドポイントとなる。これらのエンドポイントが発生したことを決定する基準は適切に規定されていなければ

ばならない。また、エンドポイントの測定のタイミングは、疾患が十分にコントロールされていないのに患者が実薬を与えられぬまま放置されることがないように定められなければならない。この試験デザインの主な難点は、短期の有効性に関する情報が得られない可能性があることである。しかし、ランダム化治療中止試験(2.1.5.2.4節参照)からは、早期離脱の特徴も有しながら、長期の有効性に関する情報も得ることができる。

2.1.5.2.3 限定されたプラセボ期

長期のプラセボ治療が受け入れられない状況では、実薬対照試験の開始時点において、短期間プラセボ群を用いることで、(少なくとも短期効果について)分析感度が確立されることもあろう。その後、試験は、プラセボ群なしで継続されることになる。

2.1.5.2.4 ランダム化治療中止

ランダム化治療中止(randomized withdrawal)試験においては、一定期間被験治療を受けた被験者が、被験治療の継続又はプラセボ(実薬治療の中止)のいずれかにランダムに割り付けられる。そのような試験の被験者は、先行して実施された非盲検の単一群試験、現に治療を受けている患者集団(しかし、通常は開始時の治療下でのベースラインを得るために治験実施計画書に定められた「ウォッシュイン」期を持つ)、対照試験の実薬群、実薬対照試験の一方の群もしくは両群から得られる。継続治療を受ける群とプラセボにランダム化された群の間に生じる如何なる群間の差も、実薬の効果を示すものとなろう。ランダム化前の治療中の観察期間は、どのような長さでもよい。それゆえこの方法は、長期のプラセボ治療が受け入れられないような場合に、効果の長期の持続性を調べるために使用できる。中止後の観察期間については、固定期間とすることもあれば、早期離脱デザインの利用、又は事象発生時間(例えば、うつ病の再発等)の測定等の方法もありうる。早期離脱デザインと同様に、割り付けられた治療で効果が失われた患者が迅速に見出せるよう、患者の観察手順及び試験のエンドポイント評価手順に対して注意深い配慮が必要となる。

ランダム化治療中止アプローチは、例えば次のような状況下で有益である。第一に、再発性の疾患の症状を軽減する薬剤(例えば抗うつ薬)に適用できる場合がある。この場合、治療中止試験は事実上は再発予防試験となる。第二に、症候や徴候(慢性の痛み、高血圧、狭心症等)を抑えるような薬剤で長期のプラセボ対照試験が困難な場合に使える。この場合には長期の有効性を確立できる。第三に、このデザインは、ある治療をどの程度の期間続けるべきかの決定に特に有益である(例えば、心筋梗塞発

作後の β 遮断薬による治療)。

ランダム化治療中止デザインの一般的な利点は、症状の再発等、早期離脱を可能とするエンドポイントと組み合わせて用いると、患者が不十分な反応のままにプラセボを投与されなければならない期間を短くできることである。

用量の問題はこの種のデザインで扱うことができる。全ての患者に初期用量として同一用量を投与した後、中止期において複数の異なる用量（又はプラセボ）をランダムに割り付ける。これは、薬力学的な理由や、未変化体や活性代謝物の半減期が長いために実薬の蓄積が生じるといった理由から、初期用量と維持用量が異なると考えられる場合に特に有用な方法である。ランダム化治療中止デザインは、最初にプラセボ対照の用量調節試験を行った後、用量反応を調べるために使えることにも注目すべきである（ICH E4ガイドライン参照）。用量調節試験は有効性を証明するために効率的なデザインであるが、質の高い用量反応情報を与えない場合も多い。反応のあった被験者をいくつかの固定用量とプラセボにランダムに割り付けるランダム化治療中止試験は、試験の初期段階で用いられる用量調節試験の効率性を保ちつつ、さらに用量反応を厳密に調べることができる。

ランダム化治療中止デザインを用いる際には、離脱反応が起こる可能性を認識しておくことが重要である。場合によっては比較的ゆっくりと減量すべきとの結論が得られることもある。患者に耐性が生じ、薬剤が患者に利益を与えていないにもかかわらず、薬剤の中止が疾患の悪化を招き、それを薬剤の有効性が維持されていたことの根拠と誤解される場合もありうる。ランダム化治療中止試験は、反応者のみに「強化(enriched)」されており、また、耐薬性のない対象が除外されている。それゆえ、このような試験では、なんら選択されていない集団から得られる治療効果より大きな治療効果が見られる可能性があることを認識することも重要である。このような状況が生じるのは、過去に薬剤に反応したと思われる被験者だけを明示的に対象にしたり、試験の前相を完了した（このこと自体しばしば良好な反応の指標となり、また常に耐薬性の指標となる）人だけを対象にした場合である。ある治療をどのくらいの期間継続すべきかを決定するための試験の場合には、そのような組み入れ基準により試験集団が規定され、目的とする比較が行われる。

2.1.5.2.5 その他のデザイン上の考慮事項

いかなるプラセボ対照試験においても、非釣合型ランダム化（例えば、被験薬対プラセボを2:1に割り付ける）を行うことにより、安全性データベースが強化され、ま

た試験が患者や治験実施医師にとってより好ましいものになることがある。

2.1.6 プラセボ対照試験の長所

2.1.6.1 有効性を示す力

他の優越性試験と同様に、プラセボ対照試験には分析感度の証拠が内在している。差が証明された場合、外的な証拠を参照することなく、解釈可能である。

2.1.6.2 「絶対的」な有効性、安全性を測りうること

プラセボ対照試験は、薬理作用に基づく治療の総効果を測る。それに対して、実薬対照試験あるいは用量比較試験は、他の治療に対する相対的な効果を測る。また、プラセボ対照試験は、薬剤による有害事象と基礎疾患又は「背景ノイズ」による有害事象とを区別することを可能とする。試験の主目的が被験薬と実薬との比較にあるとしても、三群比較試験（被験薬、プラセボ、実対照薬）における絶対的な効果の大きさに関する情報は貴重である。

2.1.6.3 効率

プラセボ対照試験は、他のどの種類の同時対照試験よりも少ない被験者数で治療効果を検出できる点で効率的である。

2.1.6.4 被験者と治験実施医師の期待による影響の最小化

盲検化されたプラセボ対照を使用する場合は、被験者と治験実施医師の双方が実薬を投与されていない被験者がいることを知っているために、両者の期待の結果として生じる改善の程度を減らしうる。これにより、試験が真の薬剤効果を検出する力が増すことになる。

2.1.7 プラセボ対照試験の短所

2.1.7.1 倫理上の懸念（2.1.3節及び2.1.4節参照）

特定の集団に対して死亡あるいは回復不能な障害を防ぐことが知られている有効な治療が存在する場合には、通常、その集団でプラセボ対照試験を倫理的に実施することはできない。ただし、これが具体的にどのような条件、どのような集団に当てはまるのかは議論の余地があるかもしれない。倫理上の懸念があることから、試験の対象をより軽症の患者としたり、長期の結果により関心があるにも関わらず短期のエンド

ポイントを検討することになる可能性がある。プラセボ対照試験が非倫理的であり、実薬対照試験が信頼できないとされる場合、新薬の試験を実施することは非常に困難なものになろう。例えば、急性心筋梗塞の患者において血栓溶解剤のプラセボ対照試験を実施するのは倫理的であるとは考えにくい。しかし一方で、非劣性試験のために妥当な非劣性の限界値を過去のデータに基づいて決めることも、血栓溶解剤のもたらす利益の大きさを変えるかもしれない緊急血管再生術が出現したため、現状では容易ではない。このような場合には、2.1.5節の試験デザインが有用なことがある。

2.1.7.2 患者と医師の治験実施上の懸念

医師が治療を行わず、又は治療を遅らせても悪い結果に至らしめることはないことを一般的に合意している場合でさえも、医師と患者は、プラセボ治療に患者が割り付けられる可能性を受け入れることに躊躇するかもしれない。自分が改善していないと感じる被験者は、効果がないのはプラセボを投与されたからだと判断して治療から脱落するかもしれず、これは試験の解析を複雑にする。しかし、適切に注意を払うことにより、効果がないために起こる脱落をエンドポイントとして扱うことができる場合もある。被験者のこのような脱落は薬剤の効果に関する一定の情報を生むが、そのような情報は、割り付けられた治療を受けている被験者の臨床状態に関する直接的な情報ほど正確ではない。

2.1.7.3 一般化可能性

全ての対照試験について言えることだが、特にプラセボ対照試験は、「現実世界」の有効性と異なる結果を与える人工的な環境で実施されているということがときに論じられる。倫理上又は実施上の懸念のために、プラセボ対照試験の試験集団が目的集団を代表するものとなりえないのであれば、試験結果の一般化可能性の問題が起こりうる。例えば、より重症な患者は治験実施計画書、治験実施医師又は患者の選択によってプラセボ対照試験から除かれることもあろう。限られた数の患者や医療機関だけが試験に参加しようとする場合もある。これらの懸念が（理論的にではなく）実際に一般化可能性に制約を与えることになるかどうかに関しては結論は得られていない。

2.1.7.4 比較情報が得られないこと

実薬対照がないプラセボ対照試験からは、相対的な有効性に関する有用な情報はほとんど得られない。多くの状況下で、相対比較に関する情報への関心は高く、またかかる情報は重要なものである。試験条件が試験間で大きく異なっている可能性があるため、試験間の比較からは相対比較情報を十分に信頼できる形で得ることはできない。

2.2 無治療同時対照 (1.3.2節参照)

ランダム化無治療対照試験の一般的特徴、長所及び短所は、プラセボ対照試験に類似している。しかしプラセボ対照試験と異なるのは、完全には盲検化できない点である。この点は、被験者の保持（脱落させないこと）、患者の管理、観察に関する全ての側面を含め、試験のあらゆる側面に影響を与える（1.2.2節参照）。このデザインが適切なのは、プラセボを対照薬とすることは可能ではあるが、盲検化が不可能あるいは現実的ではないという状況においてである。このデザインを用いる場合には、被験者の適格性やエンドポイントの決定、プロトコル上の規定の変更等の重要な決定は、治療の割付けに関して盲検化された観察者によりなされることが望ましい。解析における患者の採否のようなデータ解析に関連した決定についても、治療の割付けにアクセスできない者が実施すべきである。これ以上の議論についてはICH E9ガイドラインを参照のこと。

2.3 用量反応同時対照 (1.3.3節参照)

2.3.1 特徴

用量反応試験は、二つ以上の用量群に被験者がランダムに割り付けられる試験であり、プラセボ群を含む場合と含まない場合がある。用量反応試験は、用量と有効性、安全性（副作用等）との間の関係を明らかにしたり、有効性を証明するために実施される。前者のための利用はICH E4ガイドラインで検討されており、有効性の証明のための利用が本ガイドラインの主題である。有効性の根拠は、用量群間もしくは用量群とプラセボ群との対比較における有意な差、又はたとえ二群間に有意差がない場合でも用量の増加に伴う傾きが有意に正の値を示すという証拠によって得られる。しかしながら、後者の場合には、低用量の有効性を評価するために別の試験が必要となることもある。ICH E9ガイドラインで述べられているように、主たる有効性解析の方法は事前に明示しておく必要がある。

治療群間で「用法」が異なる試験においては、用量反応試験と同様に考慮すべき点が数多く生じる。有効性の確立のために用法対照試験を利用することは一般的ではないので、ここでの議論は用量反応試験に焦点を絞る。

用量反応試験にプラセボ（ゼロ用量）群を含めることにはいくつかの利点がある。第一に、全ての用量が同程度の効果を示し、従って全ての用量が等しく有効なのか又

は等しく無効なのかを評価することができずに解釈不能となってしまう試験を避けることができる。第二に、プラセボ群が置かれていることにより、薬理作用に基づく治療の総効果を推定することができる。ただし、各用量群の被験者数が比較的小さい場合には推定値はそれほど正確ではないかもしれない。第三に、薬剤とプラセボの差は一般に用量間の差よりも大きいので、プラセボの使用により、より少ない被験者数で試験を実施することができる。各用量群の被験者数は同じである必要はない。例えば、低用量の効果に関するより正確な情報を得るため、又は最適な用量と期待されている用量が明らかな効果を示すよう試験の検出力を増すため、これらの用量群で被験者数を多くすることができる。用量反応試験には実対照薬治療の一つ又は複数の用量を含めることができる。ランダム化治療中止デザイン(2.1.5.2.4節参照)によってもまた、被験者を複数の用量レベルに割り付けることができる。

2.3.2 バイアスを最小にする力

用量反応試験が盲検化されて実施される場合には、他のランダム化・盲検化試験デザインと同様、被験者と治験実施医師のバイアスを最小にすることができる。一部の患者や治験実施医師における盲検性を破るような薬理効果を薬剤が有する場合、プラセボ対照試験よりも用量反応試験の方が盲検性を維持することが容易であろう。治療をマスク化するために、多種類のダミーや外観が同じいくつかの異なる用量の製剤を用意すべき場合もある。

2.3.3 倫理上の問題

用量反応試験に関する倫理上及び治験実施上の懸念は、プラセボ対照試験の場合におけるそれと似ている。死亡や回復不能の障害を防ぐために有効であることが知られている治療がある場合に、効果の不十分な対照治療に意図的にランダム化することは、プラセボにランダム化することと同様に倫理的に受け入れ難い。治療がより軽度の病態に用いられる場合や治療の利益に鑑みてその毒性がかなり高い場合には、用量反応試験で、効果も毒性も低いと考えられる低用量又はプラセボを使用することが、患者や治験実施医師に受け入れられることもありうる。

2.3.4 特定の条件での用量反応試験の有用性及び推測の妥当性

プラセボ対照試験が有用な状況では、一般に盲検化された用量反応試験は効果と安全性を決定するのに有用であり、プラセボ対照試験と同様に信頼できる(2.1.4節参照)。

2.3.5 倫理上、実施上又は推測上の問題を解決しうるデザイン上の修正及び他の対照との組み合わせ

倫理上、治験実施上、推測上の問題を最小にするためにプラセボ対照試験に対してなされるデザインの修正は、一般に、用量反応試験にも適用できる(2.1.5節参照)。

2.3.6 用量反応試験の長所

2.3.6.1 効率

十分に有効な高用量とプラセボとの比較は有効性を証明するには最も効率が良いであろうが、このデザインは安全性に許容できないほどの問題を生じさせる場合があり、また用量反応情報を与えない。用量反応が単調であれば、用量反応試験は有効性の証明に関して十分に効率的であり、また用量反応情報を与える。もし最適な有効用量が知られていなければ、ある範囲の用量を検討する方が、最適用量ではない単一用量、又は許容できない有害作用のある単一の用量を選択するよりもより堅実な方法となる。

2.3.6.2 倫理上の長所がありうること

特に、用量と関連した有効性や用量と関連した重要な毒性が予測されるような場合には、プラセボ対照試験が実施できないような場合でも、用量反応試験が倫理上又は実際上実施可能な差を示すための試験となりうる。なぜなら、患者や治験実施医師にとって安全性が高いことの代償として効果が低いことを受け入れることは十分ありうるからである。

2.3.7 用量反応試験の短所

対比較に有意な差が無くても、正の用量反応の傾向性(すなわち、用量と有効性に関する結果の間の有意な相関)により有効性の証明が可能であるが(2.3.1節参照)、(最大用量以外の)どの用量が実際に有効であるかが明らかにならない可能性があるという問題がありうることを認識すべきである。もちろん単一用量の試験でも、試験されている用量より下の用量に関する情報は一切得られないので、同様な問題が起こる。

用量反応試験で用量間に差が示せないことは稀ではないことも認識しておくべきである。この場合、プラセボ群が含まれていなければ、通常、情報のない結果となる。

全く治療用量範囲の見当がつかない場合には、多くの患者が無効な用量、あるいは効果が強すぎる用量に割り付けられる可能性があり、そのデザインは効率の悪いもの

となろう。

薬の有効性を示そうとする場合に、用量反応デザインはプラセボ対照の用量調節デザインよりも効率が悪い可能性がある。しかし、ほとんどの場合、用量反応デザインはより良好な用量反応情報を与えてくれる（ICH E4ガイドライン参照）。

2.4 実薬対照（1.3.4節参照）

2.4.1 特徴

実薬対照（陽性対照）試験は、被験薬を既知の実薬と比較する試験である。このような試験は、ランダム化が行われ、通常二重盲検下で行われる。デザイン上の最も重大な問題は、試験が二つの治療の間の差を示すことを目的としているのか、非劣性又は同等性を示すことを目的としているのかである。標準薬に対する被験薬の非劣性を示す試験により有効性を主張しようとする治験依頼者は、1.5節で議論されているとおり、その試験の分析感度の問題を検討する必要がある。非劣性試験あるいは同等性試験においては、使用される用量及び実施しようとしている試験条件下で、実対照薬による治療の有効性が確かめられている必要がある（ICH E9ガイドライン参照）。一般にこのことは、その試験成績が提出されようとする地域で、実対照薬がその試験で用いられるのと同じの効能、同一の用量で受入れ可能でなければならないことを意味する。一方、被験薬の方が優れているという結果を示す優越性試験は、たとえ実対照薬の用量が低すぎるか又は実対照薬の利益が不確かな時でさえ、有効性の証拠として容易に解釈できる（しかし、実対照薬が有害でありうる時はそうではない）。しかし、そのような結果（対照薬に対する被験薬の優越性）は、実対照薬が適切な用法・用量で適切な患者に用いられる場合にのみ、被験治療の対照治療に対する真の優越性を示すものとして解釈できる（1.4.3節参照）。対照治療が適切に用いられていない試験では、その試験の分析感度が保証されないことから、被験薬の優越性が示されない場合には非劣性試験としても使用できないであろう（1.5.2節参照）。

2.4.2 バイアスを最小にする力

ランダム化と盲検化のなされた実薬対照試験は、一般に被験者と治験実施医師のバイアスを最小にするが、一つ注意すべき点がある。非劣性試験においては、治験実施医師と被験者はどちらの薬を使用しているかは分からないが、全ての被験者が実薬を使用していることは分かっている。このことにより、例えば抗うつ薬の試験のように部分的に主観的な評価が実施される場合には、ボーダーライン上の症例を成功側と分類するといった傾向が生じうる。そのようなバイアスは観測された治療間差を小さく

する可能性があり、非劣性が示されたことが有効性の証拠とはならない可能性を高くする。

2.4.3 倫理上の問題

実薬対照試験は、全ての被験者が実薬を投与されるため、一般にプラセボ対照試験よりも倫理上及び実施上の問題は少ないと考えられている。しかし、新治療を受けている被験者は（ちょうどプラセボ群が標準治療を受けていないのと同様に）標準治療を受けておらず、無効又は有害な薬を投与されているかもしれないことを認識すべきである。実薬対照による治療が生存率を改善したり、回復不能な障害の発生を減少させることが知られている場合、すなわち、プラセボあるいは無治療対照が受け入れられない状況では、このことは重要である。従って、被験薬の使用については適切な根拠が必要である。新薬が少なくとも標準薬と同程度に良い結果を示すことを予想しうる強い証拠がない場合には、状況が許す限り、上乘せ試験デザイン（2.1.5.2.1節参照）がより適切であろう。

2.4.4 特定の条件での実薬対照試験の有用性及び推測の妥当性

新治療が実対照薬による治療に優るとき、実対照薬が実際に有害ではない（負の効果を示さない）という前提の下で、実薬対照試験は、その他の優越性試験と同じように、有効性を示していると直ちに解釈できる。実薬対照試験が、非劣性を示すことにより有効性を証明するために用いられる場合、1.5節で既に検討した分析感度への特別な配慮が必要となる。また、分析感度が確立されていれば、実薬対照試験を相対的な有効性の評価に用いることも可能である。

2.4.5 ~~2.4.5~~ 倫理上、実施上又は推測上の問題を解決しうるデザイン上の修正及び他の対照との組み合わせ

既に議論したように（2.1.5節参照）、実薬対照試験には、プラセボ群、被験薬の複数の用量群、実対照薬の他の用量群を含めても良い。被験薬と実対照薬双方の複数用量が用いられる用量反応試験は、鎮痛薬の試験では典型的なものである。実薬対照試験では、用量を固定することも調節することもでき、クロスオーバー・デザインと並行デザインのどちらも使える。試験の最後にランダム化されたプラセボ対照の中止期間を加えることによって（2.1.5.2.4節参照）、あるいは、試験の初期に短期のプラセボ比較期間を加えることにより（2.1.5.2.3節参照）、非劣性試験の分析感度を示せる場合がある。絞り込まれた集団（他の治療あるいは実薬対照への無反応者）での実薬対照優越性試験は、一般化できない可能性があるにせよ、非常に有用で、一般

に解釈が容易である。

2.4.6 実薬対照試験の長所

2.4.6.1 倫理上及び実施上の長所

試験の目的が、非劣性（同等性）を示すことであっても、あるいは優越性を示すことであっても、実薬対照試験は、重要な健康上の利益が証明されている薬を使用しないことに伴う倫理上の懸念を軽減する。また同様に、有効性の証明された治療を使用しないことに関する患者と医師の心配を軽減する。患者の募集や治験審査委員会/倫理委員会の承認が促進される可能性があり、また、より多くの被験者での試験が可能になるだろう。効果がないことによる脱落が少なくなる可能性もある。

2.4.6.2 得られる情報の内容

実薬治療に対する優越性が示されれば、実薬対照試験は有効性の証拠として容易に解釈できる。実薬対照試験においては、被験者数を多くすることがときとして可能であり、また、受け入れられ易い。その結果として、より多くの安全性情報を得ることも可能になる。デザインが適切であれば、実薬対照試験から相対的な有効性についての知見も得られる。

2.4.7 実薬対照試験の短所

2.4.7.1 得られる情報の内容

非劣性試験又は同等性試験における分析感度の問題及び有効性の結論を支持するための試験の力に関する議論については1.5節を参照のこと。たとえ分析感度が支持され、その試験が有効性を検出するのに適したものであっても、効果の大きさの直接的な評価は得られず、さらに、安全性に関する結果の定量的評価にはより一層の困難を伴う。

2.4.7.2 被験者数が多くなること

一般に、否定すべき非劣性の限界値は保守的に選ばれる。それは、限界値が、実薬対照が実際に持っていると期待される効果の大きさの最小値を超えないことに十分確実な保証を与えるためである。また、通常、ある合理的理由によって定められた対照薬の効果の大きさの一定の割合（1.5.1節参照）以上に劣っている場合を排除したいがため、さらに小さな限界値が選ばれることが多い。限界値はこのように保守的に選

扱われるため、被験者数は非常に多くなるだろう。2つの薬剤の差を示すための実薬対照優越性試験では、両薬間の差は薬とプラセボの間に期待される差よりも常に小さく、このことによっても被験者数が多くなる。

2.5 外部対照（既存対照も含む。1.3.5節参照）

2.5.1 特徴

外部対照試験は、被験薬を投与される群を含むランダム化比較試験には参加していない患者で対照群を構成する試験であり、同時にランダム化された対照群が存在しない試験である。それゆえ、対照群は、治療される集団と正確に同じ集団から得られるものではない。一般に、対照群は、以前に観察された患者のうち十分な資料のある集団である（既存対照）が、他の施設において同時期に観測された集団、あるいは同じ施設であってもその試験外で観察された集団である場合もある。外部対照試験は、優越性試験（例えば無治療群との比較）であることも、非劣性試験であることもありうる。時には、大規模な外部集団から治療群に類似するように患者特性に基づいて特定の患者を選んで対照群にすることがあり、さらに特定の対照群と治療群の患者をマッチ（対応）させようとする場合もある。

いわゆるベースライン対照試験では、経時的な患者の状態がベースラインの状態と比較される。このような試験は時折「自分を自分自身の対照として」用いていると考えられているが、実際には内的対照は存在しない。つまり、そのような試験では、ベースラインからの変化が、被験薬による治療が行われなかった場合にその患者群に生じたであろうと考えられる状態の推定値と比較される。ベースライン対照試験、及び「投薬—休薬—投薬（薬剤、プラセボ、薬剤）」のようにより複雑な連続デザインであっても、ランダム化した同時対照のない試験は、いずれも外部対照試験である。既に述べたように、これらの試験では、ベースラインからの変化や試験期間中の変化が、少なくとも暗黙のうちに、介入が行われなかった場合に起こったであろうと考えられる状態の推定値と常に比較される。このような推定値は、一般に、特定の対照集団を参照することなく、単に一般的な知識に基づいて得られる。例えば効果が劇的で治療後に迅速に起こり、また効果が自発的には起こりそうもない場合（例えば、全身麻酔、電氣的除細動、測定可能な程度の腫瘍縮小等）のように、このような方法が明らかに理にかなっている場合もあるが、多くの場合その妥当性はあまり明白とは言えず、個別の過去の経験を調べることを必要とする。外部対照試験のデザイン及び解析を担当する者は、この種の試験の限界を認識しておく必要があり、また、その使用を正当

とする根拠を示すことができなければならない。

2.5.2 バイアスを最小にする力

バイアスを制御できないことが外部対照試験の主たる、そして良く知られた限界である。多くの場合にこのデザインが不適當とされるのは、この限界のためである。治療群と対照群の比較可能性を確保し、対照群を置く主たる目的を達成することが、外部対照試験では常に困難であり、多くの場合不可能である(1.2節参照)。試験治療の使用の有無以外に、人口統計学的特性、診断基準、病期又は疾病の重症度、併用治療及び観察の条件(結果の評価法、研究者の期待)等、試験結果に影響しうる様々な因子が群間で異なる可能性がある。群間のそのような相違には、重要だが認識されていない予後要因が、観測されることなく、含まれている可能性がある。外部対照が用いられている場合には、これらのバイアスを最小限にするための盲検化やランダム化を用いることはできない。おそらく選択バイアスの影響により、無治療既存対照群の結果は、ランダム化比較試験において選ばれる明らかに類似した対照群より悪い結果となりがちであることはよく知られている。ランダム化比較試験における対照群は、試験に入るためのある種の基準、すなわち、一般的に、外部対照群に典型的なものよりも厳しく、より軽症な患者集団を規定することとなる基準を満たさなければならない。外部対照群はしばしばレトロスペクティブに定義され、その結果として選択バイアスが生じる可能性がある。バイアスを制御できない結果、外部対照試験の知見に説得力を持たせるためには、同時対照試験で必要とされるよりもはるかに厳しい統計学的有意性のレベル、そして非常に大きな治療間の差の推定値が求められることとなる。

バイアスを制御できないため、外部対照デザインの使用は、治療効果が劇的であり、疾患の通常の経過が十分に予測可能である場合に限定される。さらに、外部対照を採用するのは、エンドポイントが客観的であり、エンドポイントに対するベースラインや治療変数の影響の特徴が十分に分かっているような場合に限るべきである。

既に述べたように、ランダム化・盲検化されていないため、及びその結果として被験群と対照群との比較可能性が保証されないという問題のため、このデザインに固有かつ定量不能なかなりのバイアスが入る可能性がある。しかし、デザインと実施方法次第では、外部対照試験の説得力を高め、バイアスをより小さくできる可能性がある。対照群は、試験に関係する範囲において、例えば人口統計、ベースライン状態、併用治療及び試験の経過に関する個々の患者ごとの詳細なデータといった詳細な情報が得られている集団を選ぶべきである。対照群の患者は、試験で被験薬を投与される予定

の患者集団とできる限り類似させるべきであり、試験治療の有無以外は類似の状況及び方法で治療されているべきである。試験における観察は、対照群で用いられたものと同様の時期と方法を採用すべきである。選択バイアスを減らすため、比較解析の実施前に対照群を選定すべきである。ただし、これらの対照群の結果が公表文献になっていることもあり、必ずしもこのような対応が可能なのわけではない。集団の差に対応するために行われる選択基準に関するマッチング又は調整は、すべて、対照の選択と試験の実施の前に決められていなければならない。明らかに最適な単一の外部対照が存在しない場合には、推測を行うにあたって複数の外部対照それぞれをいかに保守的に扱うか（例えば、有効性を結論するために試験群は対照群の中で最も優れた成績の対照よりも十分に優る必要がある等）について解析計画に記載することを条件に、複数の外部対照と比較することが薦められる。また、独立な評価者グループを設け、盲検下で共通の基準に従って、対照群及び被験群のエンドポイントを再評価させることが有用な場合もあろう。

2.5.3 倫理上の問題

十分な治療法がない重篤な疾患の治療を目的とする薬剤の場合、特に新薬が理論上、又は動物実験やそれ以前の臨床試験成績から有望であると思われる場合に、新薬を投与されない患者からなる同時対照群を設けた比較試験を行うことを躊躇するのは理解できることである。同時に、治療の効果を十分に示す可能性が現実的にはないような試験を実施することは無責任であり、倫理的でない。対照試験を行って見たところ、有望と思われた治療法が期待したほど劇的な改善を示さなかったり、又は全く効果がないことが明らかになった例が多いことは認識しておく必要がある。上述の状況では、試験実施医師はきわめて厳しい判断を迫られることになる。例外的なケースかもしれないが、誰もが納得するほど劇的な効果を期待して外部対照試験を開始し、期待通りにならなかった場合は直ちにランダム化比較試験に切り替える方法を探ることもあろう。

むしろ一般的に好ましいのは、満足な治療法がない重篤な疾患であっても、疾患の経過が確実に予測できない場合には、開発初期の臨床試験においてもランダム化することである。この方法は、通常、その治療が有効との印象が根付く前に試験が行われるときに可能である。劇的な利益が早期に発見できるように、独立データモニタリング委員会が試験を監視することも可能である。同時対照試験では、大きな差を迅速に検出でき、さらに、外部対照試験で信頼に足る結果として証明されるほどに大きな差ではないが価値のある効果も検出できる。

2.5.4 特定の条件での外部対照試験の有用性及び推測の妥当性

外部対照試験を採用すべきか考慮するのは、一般に、被験治療が全ての既存の治療法より優れているとの事前の確信がきわめて強いため、他のデザインが受け入れ難いと考えられ、治療される疾患や症状の経過が文献上確立し十分予測できる場合のみとすべきである。その場合でも、外部対照試験の代わりにランダム化同時対照デザインを用いることが可能な場合も多い（2.1.5節参照）。

外部対照試験の説得力が最も高いと考えられる状況は、試験のエンドポイントが客観的で、治療群の結果が外部対照と著しく異なるため、治療と対照の比較で高度の統計的有意性が達成され、疾患の経過に影響を及ぼす変数の特徴が良く分かっており、ベースライン、（試験薬以外の）治療、及び観察変数といった既知の関連する要因が全て対照群と治療群で類似している場合である。しかしそのような場合でも、外部対照試験から導かれる結論が誤っている例が知られている。

外部対照を考慮する場合は、デザインと試験実施のあり方に適切な注意を払うことによってバイアスを減らせる場合がある（2.5.2節参照）。

2.5.5 倫理上、実施上又は推測上の問題を解決しうるデザイン上の修正及び他の対照との組み合わせ

既に説明したように、外部対照デザインには、早期離脱の形式をしばしば伴うランダム化プラセボ対照治療中止期を採用することにより、ランダム化と盲検化の要素を盛り込むことができる（2.1.5.2.4節参照）。この方法を採用すると、一見反応していると思われる被験者が特定され、治療が継続される治療開始期（プラセボ対照治療中止期に入る前の時期）における結果は、厳格で、仮定とバイアスのきわめて入り難い（プラセボ対照治療中止期における）試験により「検証」されることになる。

2.5.6 外部対照試験の長所

外部対照試験の主な長所は、全ての患者が有望な薬剤の投与を受けられることである。この長所があるため、外部対照試験は患者及び医師の双方にとって魅力的なものとなる。

このデザインでは、全ての患者が被験薬を投与されるため、ある意味で効率的（2.1.6.3節、2.3.6.1等の考え方参照）と言える。このことは希少疾病において特

に重要である。しかし、外部対照試験では治療群が一つで済むことは事実だが、一方で外部対照群における結果に関する推定値は常に保守的に設定しなければならないため、必要な被験者数はプラセボ対照試験に必要な数より多くなる可能性がある。さらに、治療群と対照群の間の違いの中には、特定されたもの、特定されていないものの両方があり、また、観測されていないものもあるが、これらが治療群に有利に働くことが多いので、慎重な注意（例えば、より厳しい有意水準を適用すること等）が必要となる。

2.5.7 外部対照試験の短所

外部対照試験は盲検化できないため、患者、観察者、解析者のバイアスの影響を受ける。これは重大な短所である。これらの問題はある程度まで減らすことはできるが、治療の割付けがランダム化されておらず、また、治療開始時の対照群と治療群の比較可能性及び試験中の治療の比較可能性は保証されず、十分評価されていないので、2.5.2節に示したステップを踏んでも、問題を完全に解決することはできない。外部対照試験では被験治療の有効性が過大評価される傾向があることもよく知られている。外部対照試験において実施された統計学的有意性検定は、ランダム化試験において実施されたものに比べ、信頼性が低いことを認識すべきである。

3.0 同時対照群の選択

表1は特定の種類の対照群の有用性を示し、図1は異なった種類の対照群を選択するためのディンジョンツリーを示す。この表と図は有効性を証明するための対照群の選択に焦点を絞っているが、被験薬と対照薬との比較に使えるデザインもある。対照の選択は、個別の地域における利用可能な治療と医療環境の影響を受けることがある。

表1には、主なタイプの対照（プラセボ、実薬、用量反応）の、特定の条件下での、そして特定の目的に対する潜在的な有用性が示されている。この表の利用に際しては、各試験の潜在的な有用性を実現可能とする具体的な条件について詳細に記述した本文を併せ読むこと。表では、全ての場合において試験は適切にデザインされていると仮定している。なお、外部対照はかなり異質なので、表には含めていない。

ほとんどの場合、最も強い説得力で有効性の証拠を示せるのは、同時対照治療に対する優越性を示すことによる。優越性試験が倫理上、もしくは実施上の理由から実施不可能、又は不適切であり、さらに、実対照薬の明確な治療効果が一貫して見られる場合（例えば、多くの状況下での抗生物質）には、非劣性又は同等性試験が

利用でき、またそれは説得力を持つものになる。