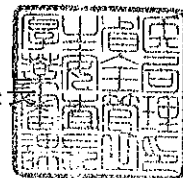




医薬審発第136号  
平成13年2月27日

各都道府県衛生主管部（局）長 殿

厚生労働省医薬局審査管理課

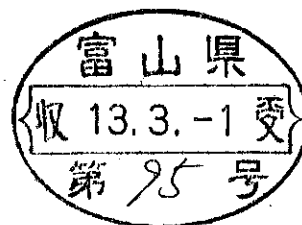


「臨床試験における対照群の選択とそれに関連する諸問題」について

今般、日米EU医薬品規制ハーモナイゼーション国際会議における合意に基づき、臨床試験における対照群の選択に関する指針を、別添の通り、「臨床試験における対照群の選択とそれに関連する諸問題」として取りまとめたので、ご了知の上、貴管下関係者に対し周知徹底方ご配慮願いたい。

なお、本指針の理解を深めるために参考として添付した質疑応答集については、今後得られる知見に基づき、修正及び拡充する予定である。

49年書き修正の時



## 臨床試験における対照群の選択とそれに関連する諸問題

### 目次

1.0	はじめに	1
1.1	ガイドラインの構成の概略と目的	1
1.2	対照群の目的	2
1.2.1	ランダム化	3
1.2.2	盲検化	3
1.3	対照の種類	4
1.3.1	プラセボ同時対照	5
1.3.2	無治療同時対照	5
1.3.3	用量反応同時対照	6
1.3.4	実薬（陽性）同時対照	6
1.3.5	外部対照（既存対照を含む）	6
1.3.6	複数の対照群	7
1.4	臨床試験の目的とそれに関連する諸問題	7
1.4.1	有効性の証拠	7
1.4.2	相対的な有効性、安全性	8
1.4.3	比較の公平性	8
1.4.3.1	用量	8
1.4.3.2	患者集団	9
1.4.3.3	エンドポイントの選択と評価時期	9
1.5	分析感度	10
1.5.1	非劣性試験又は同等性試験における分析感度	10
1.5.1.1	「薬剤効果に対する感度の既存の証拠」と非劣性の限界値の選択	12
1.5.1.2	試験の適切な実施	14
1.5.2	優越性の証明を目的とした試験における分析感度	16

2.0	各種の対照に対する詳細な検討	17
2.1	プラセボ対照	17
2.1.1	特徴	17
2.1.2	バイアスを最小にする力	17
2.1.3	倫理上の問題	18
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	19
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	19
2.1.5.1	対照群の追加	20
2.1.5.1.1	三群比較試験；プラセボと実対照薬	20
2.1.5.1.2	用量の追加	20
2.1.5.1.3	要因デザイン	20
2.1.5.2	その他の試験デザインの修正	20
2.1.5.2.1	プラセボ対照上乗せ試験；置換え試験	20
2.1.5.2.2	「早期離脱」；レスキュー治療	21
2.1.5.2.3	限定されたプラセボ期	22
2.1.5.2.4	ランダム化治療中止	22
2.1.5.2.5	その他のデザイン上の考慮事項	23
2.1.6	プラセボ対照試験の長所	24
2.1.6.1	有効性を示す力	24
2.1.6.2	「絶対的」な有効性、安全性を測りうること	24
2.1.6.3	効率	24
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	24
2.1.7	プラセボ対照試験の短所	24
2.1.7.1	倫理上の懸念	24
2.1.7.2	患者と医師の治験実施上の懸念	25
2.1.7.3	一般化可能性	25
2.1.7.4	比較情報が得られないこと	25
2.2	無治療同時対照	26
2.3	用量反応同時対照	26
2.3.1	特徴	26

2.3.2	バイアスを最小にする力	27
2.3.3	倫理上の問題	27
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	27
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	28
2.3.6	用量反応試験の長所	28
2.3.6.1	効率	28
2.3.6.2	倫理上の長所がありうること	28
2.3.7	用量反応試験の短所	28
2.4	実薬対照	29
2.4.1	特徴	29
2.4.2	バイアスを最小にする力	29
2.4.3	倫理上の問題	30
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	30
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	30
2.4.6	実薬対照試験の長所	31
2.4.6.1	倫理上及び実施上の長所	31
2.4.6.2	得られる情報の内容	31
2.4.7	実薬対照試験の短所	31
2.4.7.1	得られる情報の内容	31
2.4.7.2	被験者数が大きくなること	31
2.5	外部対照（既存対照を含む）	32
2.5.1	特徴	32
2.5.2	バイアスを最小にする力	33
2.5.3	倫理上の問題	34
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	35
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	35
2.5.6	外部対照試験の長所	35
2.5.7	外部対照試験の短所	36
3.0	同時対照群の選択	36

表 1	種々の状況における各同時対照の有用性 .....	38
図 1	有効性を証明するための同時対照の選択 .....	39

## 1.0 はじめに

臨床試験のデザインに際して、どのような対照群を選択するかは、いかなる場合においてもきわめて重要である。対照群の選択は、試験から引き出しうる推測、試験が倫理的に受け入れられるかどうか、試験の実施及び解析におけるバイアス（偏り）を小さくできる程度、組入れうる被験者のタイプと組入れの速さ、検討可能なエンドポイントの種類、結果の社会的・科学的信頼性、結果が規制当局に受け入れられるかどうか、その他の多くの試験の特徴、試験実施のあり方及び解釈に影響を与える。

### 1.1 ガイドラインの構成の概略と目的

本ガイドラインの目的は、ある治療の有効性を証明することを目的とした臨床試験に関し、対照群の選択において考慮すべき一般的原則を記述すること、そして、関連する試験デザインや実施上の問題について議論することである。本ガイドラインは、各地域での規制上の要求事項に言及するものではなく、各デザインを用いた試験により何を示すことができるのかを述べるものである。本ガイドラインに記述された一般的原則は、どのような比較試験にも適用できるものであるが、特に、薬剤の開発段階で有効性を証明するために実施される臨床試験においては、対照群の選択は極めて重要である。対照群の選択においては、利用可能な標準治療、選択されたデザインの妥当性を支持する証拠の適切性、そして倫理上の配慮といった観点からの考慮が求められる。

本ガイドラインでは、まず、対照群の目的と、有効性の証明のために一般的に採用される対照群の種類を述べる。次に、対照群に対する非劣性あるいは同等性を示すことにより有効性を証明しようとする実薬対照試験の利用に関する重要なデザイン上・解釈上の問題について議論する（1.5節参照）。非劣性を見出したことが有効性の証拠とは解釈できない状況が存在する。つまり、非劣性を有効性の証明と解釈するためには、有効な治療と有効性がそれより低い治療又は無効な治療とを区別する力を実施された試験が有していたことが前提となる。

次いで本ガイドラインでは、各種の対照群を用いる試験をより詳細に説明し（2.0節 - 2.5.7節参照）、それぞれの対照群について次のような点を考察する。

- バイアスを最小にする力
- その対照群の使用に関連した倫理上及び実施上の問題
- 特定の状況における有用性と推測の質
- 倫理上、実施上又は推測上の問題を解決するための試験デザイン上の修正又は他

の対照との組合わせ

- 全般的な長所と短所

本ガイドラインは、次のICHガイドラインと特に関連が深い。

- 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
- 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
- E5：外国臨床データ受け入れにおける人種的要因
- E6：医薬品の臨床試験の実施の基準
- E8：臨床試験の一般指針
- E9：臨床試験のための統計的原則

本ガイドラインで記述、議論された対照群のいずれについても、それが採用されたときに当該試験が承認の根拠の一部として有用とされ、受け入れられる状況は確かにあるだろう。しかし、こうした試験が、あらゆる状況下で等しく適切、有用というわけではない。対照群の選択に関する一般的なアプローチの概略を、3.0節の図1及び表1に示す。

本ガイドラインは、治療の「有効性」の評価を目的とした試験に主たる焦点を当てているが、「安全性」に関する特定の仮説の評価や、二種類の治療の「安全性・有効性の相対比較」にも議論の多くは適用できる。

## 1.2 対照群の目的

対照群を置く大きな目的の一つは、被験治療により患者に起こった結果 (outcomes。例えば、症状、徴候、その他の病態の変化) と、疾患の自然の進行、観察者・患者の期待、他の治療等の要因により引き起こされた結果との弁別を可能にすることである。対照群を用いることにより、被験治療が行われなかった場合に患者に起こったであろう結果、又は有効であることが知られている他の治療を受けた場合に起こったであろう結果を知ることができる。

もし、特定の患者集団における疾病の経過が一様であったり、特定の被験者 (群) の結果が患者の特徴を用いて確実に予測可能だとすれば、治療を行った結果は、治療を行わなかった場合の既知の結果と単純に比較できるであろう。例えば「痛みは一定時間持続したであろう」、「血圧は変化しなかったであろう」、「うつ状態は一定の期間持続したであろう」、「腫瘍は進行したであろう」、「急性心筋梗塞後の死亡率

は以前に観察された死亡率と同じであつたらう」といったことを仮定することはできない。特殊な場合には、ある定義された患者集団における疾病の経過が実際に予測でき、過去に調べられた類似の患者群を「既存対照」として用いることが可能なこともある（1.3.5節参照）。しかしながら、ほとんどの場合、十分な正確さ・確実さで結果を予測することは不可能であり、従って同時対照群が必要となる。

同時対照群は、被験群と同じ母集団から選ばれ、被験治療が検討されている同じ試験の一部として、あらかじめ定められた方法で、同じ時期に治療される群である。被験群及び対照群は、試験の検討対象となっている治療を除き、結果に影響を与えうる全てのベースライン変数及び治療変数（併用薬や観察等）について類似していなければならない。この類似性が実現できなかった場合、試験にバイアスが入りうる。ここでいうバイアスとは、ICH E9 ガイドラインにおいて用いられているとおり、ある治療の効果についての推定値と真の値の間に差を生じさせる、臨床試験のデザイン、実施、解析、結果の解釈におけるあらゆる観点からの系統的な傾向を意味する。バイアスの発生の可能性を最少限にし、被験群と対照群が試験の開始時点で類似し、また試験の経過の中でも同じように治療されていることを保証するために通常用いられる二種類の手法が、ランダム化と盲検化である（ICH E9ガイドライン参照）。試験デザインがこれらの特徴を備えているか否かで、その試験の質と説得力は大きく違ってくる。

#### 1.2.1 ランダム化

被験群と対照群の被験者集団の類似性を保証する最良の方法は、単一標本集団を被験治療を受ける群と対照治療を受ける群にランダムに分けることである。結果に影響を与える可能性のある既知の、あるいは未知のベースライン変数についての系統的な群間差を、ランダム化により回避できる。治療群間の系統的な差を排除できないことがランダム化同時対照の存在しない試験の大きな問題である（1.3.5（外部対照）節参照）。また、ランダム化は統計的推論のための適切な基礎を与える。

#### 1.2.2 盲検化

被験薬、対照薬のいずれが投与されるかという点を除き、各群は、ベースラインが類似しているだけではなく、試験中に実施される治療や観察においても類似していなければならない。臨床試験は、しばしば「二重盲検」（又は「二重マスク」）下で行われる。これは、各被験者に割り付けられた治療を、被験者及び治験実施医師だけでなく、治験依頼者、被験者の治療や臨床評価に関係する治験実施医師のスタッフも知らないことを意味する。盲検化は、被験者や治験実施医師が割り付けられた治療を知



っていることにより生じる被験者の管理、治療、評価の違いや結果の解釈の違いから生じる可能性のあるバイアスを最小にすることを意図したものである。例えば：

- 実薬群の被験者は、もし自分に実薬が投与されていると知った場合には、良い結果を期待するために実際より良い結果を報告し、また、試験に留まろうとする可能性がある。
- 観察者は、無治療群での反応の確認や報告を怠り、実薬を投与されている患者で見られた好ましい結果や有害事象にはより敏感になる可能性がある。
- 割り付けられた治療を知っていることが、試験中・試験後の追跡データを得ようとする意欲に影響を与えうる。
- 割り付けられた治療を知っていることが、被験者を試験治療に留まらせるべきか否かや、併用薬又は他の補助的な治療を行うべきか否かについての決定に影響を与えうる。
- 割り付けられた治療を知っていることが、特定の被験者の成績を解析に含めるべきか否かの決定に影響を与えうる。
- 割り付けられた治療を知っていることが、統計解析法の選択に影響を与えうる。

盲検化の目的は、主観的な評価と決定が、割り付けられた治療を知っていることによる影響を受けないことを保証することである。

### 1.3 対照の種類

臨床試験における対照群は、二種類の重要な特性、すなわち、用いられる治療の種類及び誰が対照群に入るかを定める方法に基づいて分類することができる。対照治療の種類は次の四種類のいずれかになる。すなわち、(1)プラセボ、(2)無治療、(3)異なった用量又は用法の被験治療、(4)被験治療とは異なる実薬による治療である。誰が対照群に入るかを決定する主な方法としては、ランダム化を用いるか、試験に参加して治療を受ける集団とは別の対照集団（外部対照又は既存対照）を選択するかのいずれかがある。本ガイドラインでは対照群を五種類に分類している。最初の四種類は、通常は治療へのランダム化による同時対照（対照群と被験群が同じ集団から選ばれ、同時に治療される）であり、それは先に挙げたどの種類の対照治療を受けるかによりさらに区別される。外部（既存）対照群を用いる試験では、対照治療の種類にかかわらず、そのような試験の被験群と対照群の比較可能性を保証する力と、影響の大きいバイアスを最小限にする力について大きな懸念が抱かれる。それゆえ、このデザインは例外的な状況下でしか使用できず、したがって、外部（既存）対照群を用いる試験はまとめて五番目の種類とみなす。

複数の異なる種類の対照群を含む試験を実施することが次第に一般的になりつつある。それぞれの対照が、ある条件のもとでは妥当であるにしても、いかなる条件でも利用可能で妥当であるというわけではない。五種類の対照は次のとおりである。

### 1.3.1 プラセボ同時対照

プラセボ対照試験では、被験者は被験治療か、それと外観が同じで被験治療（薬）を含まない治療かにランダムに割り付けられる。効果や忍容性にに基づき用量が調節される場合もあれば、単一用量又は複数の固定用量が投与される場合もある。そういった試験は、ほとんど常に二重盲検試験である。プラセボ対照（placebo control）という呼び方は、その目的が「プラセボ」効果（薬を使用していると考えることによって被験者に改善が見られること）をコントロールすることを示すが、そのことだけが唯一の利点、主たる利点ではない。むしろ、プラセボ対照試験の利点は、盲検化とランダム化を可能とし、また、不活性な治療を受ける群を試験に組み込むことにより、実際の、又は見かけの疾病の進行に対する被験薬の薬理作用以外の全ての潜在的な影響をコントロールする点にある。ここで考えられる影響としては、自然変動（疾病の自然経過、平均への回帰）、被験者・治療実施医師の期待、試験に参加していることによる効果、他の治療の使用、診断・評価の主観的要素等が挙げられる。プラセボ対照試験では、有効性を調べようとする場合には治療間の差を見出すことが目的となるが、安全性に関する測定値の評価では（一定の大きさの）差がないことを示そうとすることもあろう。その場合には、もし差が存在したのであれば、実施された試験がその存在する差を示し得たのか否かという問題がきわめて重要なものとなる（1.5節参照）。

プラセボ対照群を使用することは、対照群が治療されないことを意味するのではない。多くのプラセボ対照試験においては、一般的な標準治療に、新しい治療又はプラセボがそれぞれ上乗せされる（「上乗せ試験」と呼ばれる。2.1.5.2.1節参照）。

### 1.3.2 無治療同時対照

無治療対照試験では、被験者は被験治療群又は無治療（すなわち、被験治療、対照治療のいずれも行われぬ）群のどちらかにランダムに割り付けられる。このデザインがプラセボ対照試験と大きく異なるのは、被験者と治療実施医師が割付けられた治療を知っている点である。二重盲検デザインの長所を考えると、無治療対照試験が必要かつ適切なデザインといえるのは、二重盲検とすることが困難又は不可能であり（例

えば、容易に識別しうる毒性を伴う治療の場合)、かつ試験のエンドポイントが客観的で、試験結果が1.2.2節に挙げられた要因の影響を受けないだろうとの合理的な自信が持てる場合に限られるだろう。たとえ試験全体が二重盲検化されていない場合でも、盲検化された評価者がエンドポイントの評価を行うことはしばしば可能であることに注意を要する。これは有用な手段であり、盲検化できない試験では常に採用を考慮しなければならないが、割り付けられた治療を知っていることに関連したその他の問題を解決できるわけではない(1.2.2節参照)。

### 1.3.3 用量反応同時対照

ランダム化された固定用量の用量反応試験では、被験者はいくつかの固定用量群の中の一つにランダムに割り付けられる。最初から用量を固定する方法と、その用量に至るまで次第に漸増する方法があるが、目的とする比較は最終の用量間で行われる。用量反応試験は、通常、二重盲検下で実施される。用量反応試験はプラセボ(ゼロ用量)や実薬対照を含む場合もある。血中濃度対照試験においては、いくつかの固定された血中濃度範囲に入るよう治療群の用量が調節される。この種の試験は、概念的には固定用量の用量反応試験に類似したものである。用法対照試験においては、被験者は、試験薬の2つ以上の用法のいずれかにランダム化される(例えば、1日1回投与対2回投与、3日間投与対7日間投与)。

### 1.3.4 実薬(陽性)同時対照

実薬対照(陽性対照)試験では、被験者が、被験治療か実対照薬による治療かにランダムに割り付けられる。そのような試験は、通常、二重盲検試験として実施されるが、二重盲検化が常に可能とは限らない。例えば、多くの腫瘍領域の試験では、用法、投与経路も異なる上に(1.3.2節参照)、毒性も異なるために盲検化は困難、あるいは不可能と考えられる。実薬対照試験には、有効性を示すことに関し、二つの異なった目的がありうる。すなわち、既知の有効な治療と同程度に良いことを示すことにより被験治療の有効性を示すこと、及び、実対照薬による治療に対する被験治療の優越性を示すことにより被験治療の有効性を示すことである。実薬対照試験はまた、主目的が二治療間の有効性・安全性の比較である場合にも用いることができよう(1.4節参照)。試験の目的が、新しい治療の有効性の証明であれ、二つの治療の比較であれ、その試験が有効な治療と有効性の低い治療、無効な治療とを区別しえたのか否かがきわめて重要な問題である(1.5節参照)。

### 1.3.5 外部対照(既存対照を含む)

外部対照試験では、被験治療を受けた患者群と当該試験外の患者群が比較される。比較の対象は、同一集団の中で異なる治療に割り付けられた患者から構成される内部的な対照群ではない。外部対照には、その試験の実施前に治療された患者からなる群（既存対照）の場合と、同時期であるが他の条件で治療される群の場合がある。外部対照は、特定の患者群として明確に定義される場合もあれば、結果に関する一般的な医学的知識に基づいた比較対照群のように明確に定義されない場合もある。一般的印象というものはしばしば不正確であり、後者の比較対照群の使用は特に誤った結論を与え易い（そのような試験には対照がないと普通は考える）。被験者の治療中の状態（例えば血圧や腫瘍のサイズ）が治療前の状態と比較されるいわゆるベースライン対照試験は、内部に対照群を持たず、それゆえ、無対照試験、あるいは外部対照試験である（2.5節参照）。

#### 1.3.6 複数の対照群

後に説明するとおり（1.5.1節参照）、例えば実薬対照とプラセボ対照の同時使用など、一つの試験で二種類以上の対照を用いることが可能で、かつ利点がある場合が多い。同様に、被験薬と実薬対照の双方で複数の用量を試験に用いることもできる。この場合、さらにプラセボを入れる場合もあれば、入れない場合もある。このデザインは、二つの薬剤の相対効力が十分に確立されていない場合又は試験の目的が相対効力の確立にある場合の実薬同士の比較に有用であろう。

### 1.4 臨床試験の目的とそれに関連する諸問題

臨床試験の二つの目的、すなわち、(1)治療の有効性、安全性の評価と、(2)二つの治療の相対的な有効性、安全性、リスクベネフィットの関係、有用性の評価とは、区別する必要がある。

#### 1.4.1 有効性の証拠

いずれの種類に対照を用いた試験でも、被験治療が対照（プラセボ、無治療、被験薬の低用量、実薬）に優ることを示すことにより、その有効性を示すことができよう。さらに、実薬対照試験では、新治療の効果が既知の有効な治療と同程度であることを示すことにより有効性を示すことができる場合もある。しかし、ある試験において両治療とも効果がない場合にもやはり同程度に見えることから、こういった類似性が試験治療の有効性の証明となるのは、実薬対照による治療が当該試験の条件下で有効であると仮定できる場合だけである（1.5節参照）。

効果が標準薬と同程度であることを示すことにより新薬の有効性を証明するようデザインされた臨床試験は、これまで「同等性」試験と呼ばれてきた。これらの試験のほとんどは、実際には、新薬が対照に対して一般にマージンと呼ばれる予め決められた大きさ以上劣ることはないことを証明しようとする非劣性試験にあたる。

#### 1.4.2 相対的な有効性、安全性

試験の焦点が、被験薬自体の有効性そのものではなく、ある治療と他の治療の比較に向けられる場合もある。治療分野によっては、これらの試験は、相対的なリスクベネフィット評価のために重要な情報を生み出すものと考えられよう。その場合の実比較対照薬は、データを提出する予定の地域において受け入れられるものでなければならない。必ずしも実比較対照薬に対する優越性の証明は必要ではなく、また状況によっては、非劣性を示す必要もない場合もあるかもしれない。例えば、有効性の劣る薬が安全面では利点があり、それゆえ有用であると考えられる場合である。

たとえ非劣性試験の主目的が有効性の証明ではなく治療間の比較にあるとしても、非劣性試験の実施と解釈に関する注意点について同様に考察する必要がある（1.5節参照）。つまり、実際には差がある二つの治療を弁別できない試験からは有用な比較情報が得られないのであり、治療間に差がある場合はその差を検出できる力が試験に備わっている必要がある。

#### 1.4.3 比較の公平性

相対的な安全性・有効性に関する十分な情報を比較試験から得るためには、試験が「公平」である必要がある。すなわち、試験の条件が、他方の治療に比べ一方に不当に有利なものであってはならない。試験で用いられた対照がその通常の効果を発揮したかどうかに関して何らかの疑問があれば、その試験に分析感度（1.5節参照）があったことの保証を失いかねない。それゆえ、実際には、有効性の証拠として提出される実薬対照同等性あるいは非劣性試験では、ほとんどの場合、対照との公平な有効性の比較が行われる必要がある。一方の治療を不当に有利にしうる試験デザインの側面として、用量、患者集団の選択、エンドポイントの選択及びその評価時期が挙げられる。

##### 1.4.3.1 用量

被験薬と実対照薬の比較において、対照薬と被験薬の適切な用法・用量を選ぶことは重要である。二つの治療の比較結果を検討する際には、見かけ上効果の劣っている治療についてはその用量が低すぎた可能性を、一方、見かけ上忍容性の劣っている治

療についてはその用量が高すぎた可能性を検討することが重要である。有効性、安全性の優越性を説得力をもって示すために、対照治療の複数の用量、そしておそらくは被験治療も複数の用量を試験することが必要となる場合もある。

#### 1.4.3.2 患者集団

実薬対照試験における被験者の選択は結果に影響を与えうるので、試験が示した結果を評価する際には、試験の対象となった集団を注意深く検討すべきである。例えば、試験に組み入れられた多くの被験者が、過去に対照治療に反応していなかった場合には、新治療に有利なバイアスが入りうる。そのような試験結果は、前治療を受けていない患者集団全体へ一般化することはできないだろう。しかし、そのような場合であっても、新治療の優越性が証明されたのであれば、試験の結果は、試験されている集団において新治療が有効であることの証拠にはなるであろう。実際、（患者を危険な状態に置かない限りにおいて）ある治療への無反応者（nonresponder）を新治療か反応がなかった治療かにランダム化することにより、明らかな無反応者における新治療の効果を調べる試験は、そのような無反応者における新治療の価値を証明できるのであり、臨床的に価値ある観察といえる。

同様に、特定の薬により反応しやすい（反応しにくい）患者部分集団、あるいは、より副作用の出やすい（出にくい）患者部分集団を見出すことができる場合もある。例えば、一般に、黒人はベータ遮断薬やアンジオテンシン変換酵素阻害薬の血圧に対する作用に反応しにくいとされており、このような患者で新たな降圧薬とベータ遮断薬等の降圧薬とを比較すると新薬の優越性を示す方向に傾く可能性がある。このような比較で新薬が一般に優れていると結論することは不適切である。しかし、その限界と、正しく導きうる結論は何かを認識した上で、部分集団で計画された試験からは、やはり有用な情報を得ることができる。

#### 1.4.3.3 エンドポイントの選択と評価時期

二種類の治療が同じ疾病・病態に用いられるとき、特に、両治療の治療クラスや治療様式が異なる場合には、これらの治療は、その疾患の検討対象である様々な結果に異なった影響を与えうる。それゆえ、これらの治療を臨床試験で比較する場合には、エンドポイントの選択と評価時期によって一方の治療が有利になる可能性がある。例えば、急性心筋梗塞の患者に対する抗血栓薬は、死亡を減らしうるが、出血性の脳卒中中のリスクを増やす可能性がある。新規の、より活性の強い抗血栓薬が古い抗血栓薬と比べられるとき、もしエンドポイントを死亡にした場合には、活性の強い治療の方が優るが、死亡と障害を残すような脳卒中を合わせた複合エンドポイントを用いる場

合には、活性の強い治療の方が劣るということもありうるのである。同様に、歯痛のコントロールに関して二種類の鎮痛薬を比較するとき、早期の時点での痛みに特に評価の重きを置けば、より早く効果が発現する薬に有利になり、遅い時点により重きをおけば、効果のより持続する薬に有利になるだろう。

### 1.5 分析感度

「分析感度」(assay sensitivity)とは、有効な治療と有効性の低いあるいは無効な治療とを区別する力として定義される臨床試験の特性である。分析感度はいかなる試験においても重要であるが、その意味するところは、治療間の差を証明することを目的とした試験(優越性試験)と、非劣性を証明することを目的とした試験では異なる。もし、対照治療に対する試験治療の優越性を証明することにより有効性を証明しようとする試験が分析感度を欠くならば、試験治療が優れることを証明できず、したがって有効であるとの結論を得られないことになる。一方、実薬対照に対する被験治療の非劣性を証明することにより有効性を証明しようとする試験に分析感度がない場合には、無効な治療が「劣っていない」と判断され、誤って有効であるとの結論に至ることがありうる。

ある試験において二つの治療が異なる有効性を示す(すなわち、一方の治療が優れている)場合、その結果自体が、当該試験が分析感度を有することの証しになる。一方、「成功」した非劣性試験(すなわち、非劣性を示した試験)、あるいは「成功」しなかった優越性試験からは、一般に、分析感度の直接的な証拠は得られない。

#### 1.5.1 非劣性試験又は同等性試験における分析感度

非劣性試験又は同等性試験における分析感度の存在は、次の二つが満たされているかどうかから推測されることになる。

##### 1) 「薬剤効果に対する感度の既存の証拠」

すなわち、同様にデザインされた過去の試験では、ほぼ一貫して有効な治療と有効性の低いあるいは無効な治療とを区別できたこと。

##### 2) 「試験の適切な実施」

すなわち、試験の実際に行われ方が、有効な治療と有効性の低いあるいは無効な治療を区別する力を低下させなかったこと。

「薬剤効果に対する感度の既存の証拠」は、非劣性試験の開始前に評価可能であり、

実際に評価されなければならない。具体的に言えば、試験を実施しようとしている特定の治療領域において、ある特定の実薬による治療あるいは同様の効果を持つ他の治療を用い、適切にデザインされ実施された試験が、信頼性をもって一定の効果を示したと判断されなければならない。これを最も適切に証明するのは、実薬対照として使おうと計画している実薬治療が、プラセボに対して確実に優越性を示していることである。その場合には、同様にデザインされた実薬対照の試験には、「薬剤効果に対する感度の既存の証拠」があることになる（1.5.1.1節参照）。

「試験の適切な実施」は、実薬対照非劣性試験が終了した後にのみ十分に評価可能となる。非劣性試験のデザイン（例えば、組入れ基準、許容される併用療法）が、「薬剤効果に対する感度の既存の証拠」があるとの判断に用いられた過去の試験と類似しているべきであるだけでなく、さらに、その試験が実際にどのように行われたかがこうした過去の試験と類似したものであったことを保証するため、実際に組み入れられた集団、実際に用いられた併用療法等について評価が行われなければならない。さらに、試験の行われ方は質の高いものでなければならない（例えば、良好なコンプライアンス、追跡不能例がほとんどないこと等）。「薬剤効果に対する感度の既存の証拠」とともに、「試験の適切な実施（1.5.1.2節参照）」が、新たに実施される実薬対照試験に分析感度の保証を与える。

以上を踏まえると、非劣性試験のデザイン及び実施は、次の重要な四段階を経ることになる。

1. 「薬剤効果に対する感度の既存の証拠」が存在することを確認する。

この判断がなければ、非劣性を示すことにより有効性を証明することは不可能であり、そのような証明を試みるべきではない。

2. 試験をデザインする。

例えば試験対象集団、併用療法、エンドポイント、導入期間（run-in period）といった試験デザイン上の重要な点は、「薬剤効果に対する感度の既存の証拠」が存在することの判断の根拠となった試験のデザインをできる限り踏襲しなければならない。

3. 非劣性の限界値を設定する。

既存のデータ、妥当な臨床的・統計学的考察に基づき、許容可能な非劣性の限界値



を設定しなければならない。

#### 4. 試験を実施する。

試験の行われ方は、過去の試験にできるだけ近いものでなければならず、また、質の高いものでなければならない。

##### 1.5.1.1 「薬剤効果に対する感度の既存の証拠」と非劣性の限界値の選択

既述のように、大部分の実薬対照同等性試験は、実際には新治療の有効性の確立を意図した非劣性試験である。非劣性試験の結果の解析についてはICH E9ガイドライン及びICH E3ガイドラインで議論されている。簡単に言うと、このような試験では被験治療と既知の有効な治療が比較される。その試験に先立ち、ときに「デルタ」と呼ばれる同等性又は非劣性の限界値（マージン）が選択される。この限界値が、その試験で統計的に否定しようとしている「対照~~照~~に対して被験治療が劣っている程度」である。被験治療と対照治療との差の信頼区間が、被験治療の非劣性の限界値を含まない場合には、被験治療は対照治療に対して非劣性であると宣言できる。一方、限界値が信頼区間に含まれるならば、被験薬は非劣性であると宣言できない。

非劣性試験のために選ばれる限界値は、計画している試験の設定条件で、実対照薬とプラセボを比べた場合に、「確実に期待できる実対照薬の効果の大きさの最小値」より大きな値であってはならない。もし、実対照薬と新薬の差について、この限界値と同程度以上に対照の方が良いという結果が得られたのであれば、新薬には全く効果がない可能性もある。「薬剤効果に対する感度の既存の証拠」がある場合にのみ、実対照薬に対して確実に期待できる効果の大きさの最小値を決定でき、実際、その証拠に基づいて限界値が決定される。一般に、限界値は、新しい試験で計画されている条件と類似の条件の下で適切にデザインされた過去のプラセボ対照試験の経験に基づいて決定されるが、用量反応試験あるいは実薬対照の優越性試験から決定することもできる。このような既存の試験でどのような種類の対照群が用いられていようと、限界値の決定の基礎となる値は、（既存の試験での）対照群に対する実対照薬の優越性の程度であり、ベースラインからの変化のような比較に基づかない測度ではない。ただし、本ガイドラインは厳密な限界値の計算法を説明するものではなく、計算の方法についてはあまり公表文献がないことには留意すべきである。

非劣性試験における限界値の決定は、統計的推論と臨床判断の双方に基づくものであり、その選択の基となった証拠の不確実性を反映するとともに、適度に保守的であ

るべきである。もしこの限界値の決定が適切に行われれば、新薬と実対照薬の差の信頼区間が適切に選ばれた限界値を含まないという結果から、その被験薬がゼロより大きな効果を持っているとの確証が得られる。実際には、臨床的に受け入れることのできるある一定の大きさの効果（又は対照薬の効果の何割か）が保持されていたことを保証することに関心があるために、選択される非劣性の限界値は、通常、実対照薬の最小の期待される効果の大きさから示唆される値よりも小さなものとなる。例えば、死亡率に関する非劣性試験においては、被験治療がゼロより大きい効果を有することを保証するだけでは、一般には十分なものとは考えられないであろう。その場合には、通常、対照薬の死亡率に対する効果の大きさの相当の割合が保持されていることが求められよう。これはまた、その主目的が被験薬と実薬対照との相対的な有効性を示すことにある試験についても言えるだろう（1.4.2節参照）。この場合は、単に新薬が何らかの効果を持っているということだけではなく、被験薬と対照薬がよく似ているということの保証が通常求められよう。

否定すべき限界値の選択が既存の証拠に基づくという事実から、非劣性試験には既存対照（外部対照）試験と共通の要素が存在することになる。非劣性試験のデザインが適切で信頼できるものとなるのは、対照薬の効果の大きさに関する既存の推定値が、その薬剤に関する過去の試験の結果に照らして、十分に裏付けられる場合だけである。これらの過去の試験からは、新たに計画している試験と類似のデザインで適切な規模の試験を仮に行った場合には、その実対照薬とプラセボを一貫して区別しうるとの結論が得られなければならない。また、これらの過去の試験から、その対照が確実に示すと期待できる薬剤効果の最小値という意味での薬効の大きさが決められなければならない。もし現在計画している試験と同様のデザインのプラセボ対照試験において、予定している実対照薬とプラセボとの間の差が高い頻度で見出せず、このことを試験の特性により説明できなければ、被験薬の（実対照薬等に対する）優越性のみが解釈可能な結論となる。

「薬剤効果に対する感度の既存の証拠」があるか否かは、ある程度まで判断の問題である。薬剤効果に対する感度は、既存のプラセボ対照試験の結果の一貫性から明らかであることもあれば、疾病を治療した場合と治療しなかった場合の結果が大きく異なることから明らかなることもある。例えば、多くの感染症において、短期間の試験中に、有効な治療における治癒率は、自然経過による治癒率をはるかに上回る。しかしながら、有効と考えられている薬剤が適切に計画・実施された対照試験で必ずしも毎回プラセボ対照に優ることを示すことができず、従って特定の試験条件でその薬剤が

有するであろう最小の効果を十分信頼できるほどに確定できないような疾患は数多く存在する。そのような疾患の例としては、うつ病、不安神経症、痴呆、狭心症、症候性うつ血性心不全、季節性アレルギー、症候性逆流性食道疾患のように、プラセボ群で大きな改善や変動が認められたり、治療効果が小さかったり大きくばらつくようなものが挙げられる。

これら全ての場合において、標準治療が有効であることは疑いない。なぜなら、標準治療に用いられる各薬剤については、その効果を支持する数多くの適切に計画・実施された対照試験があるからである。しかし、これまでの経験から、その薬剤が少なくとも最小限の効果を確実に有し（すなわち「薬剤効果に対する感度の既存の証拠」が存在し）、従って適切な限界値を決定するために使えるような試験条件を記述することは困難であろう。「薬剤効果に対する感度の既存の証拠」がよりどころとしている経験の妥当性に疑問が持たれる場合もあろう。例えば、治療の標準や診断基準が時間とともに大きく変化した場合（例は2.1.7.1節参照）である。実薬対照あるいは非劣性試験デザインを計画する場合に、選択した非劣性の限界値を用いた計画中の試験の「薬剤効果に対する感度の既存の証拠」を十分に裏付けることができないのであれば、そのような試験で仮に非劣性が証明されたとしても、有効性について十分な情報を提供するとは考え難い。

先に注意を喚起したように、「薬剤効果に対する感度の既存の証拠」に関する判断は、特定のデザインの試験に対してのみ適用できる。計画された非劣性試験が、薬剤効果に対して同様に感度を持つためには、その試験の重要なデザイン上の特徴が既存の試験の特徴と類似していることが、本質的に重要である。重要なデザイン上の特徴には、例えば、組入れ基準（疾病自体の重症度、合併症、診断方法）、対照薬の用法・用量、併用療法、観測されるエンドポイントと評価の時期、そして特定の患者を除外するためのウォッシュアウト期間の採用が含まれる。（例えば、技術や治療の進歩により）試験デザイン上の特徴に違いが生じることが避けられない場合やそのような違いが望まれる場合、「薬剤効果に対する感度の既存の証拠」の存在についての判断や限界値の選択に関してその相違がどのような意味を持つかを慎重に考慮しなければならない。

#### 1.5.1.2 試験の適切な実施

たとえ「薬剤効果に対する感度の既存の証拠」があり、新しい試験のデザインが過去の試験と類似していたとしても、分析感度はその試験の実際に行われ方により損な

われることがある。ある試験の分析感度を保証するためには、その試験の実際に行われ方の質が高くなければならず、また、実際に組み入れられた患者、（試験治療以外の）実際に行われた治療、実際に行われた評価が、薬剤効果に対する感度の存在の根拠とした試験におけるそれらと類似していなければならない。

試験の実施上、有効な治療と、より効果の低いあるいは無効な治療との間に観察される差を小さくし、そのために試験の分析感度を低下させうる様々な要因がある。例としては次のようなものがある。

- 1) 服薬（治療）不遵守
- 2) 薬剤の作用に反応しにくい試験集団の組み入れ
- 3) 試験薬の作用を妨げたり、示しうる反応の大きさを小さくするような治療実施計画書外の併用薬又は他の治療の使用
- 4) 自然回復する傾向があり、薬剤によるそれ以上の改善の余地がない患者集団の組み入れ
- 5) 診断基準の不適切な適用（試験の対象となる疾患を持たない患者）
- 6) 全ての患者が活性を持つ可能性のある薬を投与されていることを知っていることにより生じる偏りのあるエンドポイントの評価。例えば、血圧に対する効果を正常化したものとして読む傾向により、被験薬と対照薬間の差が縮まる可能性がある。

臨床研究者や治験依頼者は質の高い試験を実施したいと考えており、GCP（ICH E6 ガイドライン）の施行が今後も試験の質を高めていくであろう。試験の質を高めるこれらの要因を考慮してもなお、治療間の差を見出すことを目的とした試験の場合には、良い試験デザインを用い、試験の誤りを出来るだけ少なくしようとする強い動機が生じることを認識すべきである。なぜなら、多くの場合、試験に不完全な点があれば、治療間に差が存在していてもその差の検出に失敗する可能性が増すためである。プラセボ対照試験の場合には、有効な治療とプラセボを区別することを保証するために、実施計画書の遵守状況を改善し、また、患者集団が薬剤効果に反応する可能性を高めるための多くの努力が払われる。しかし、多くの臨床環境において、試験の質の高さと分析感度を保証するための強い動機があり多くの努力がなされたにもかかわらず、臨床試験が有効な薬剤とプラセボを確実に区別できないことはしばしば起こる。

それに対し、二つの治療間に一定の大きさの差がないこと（非劣性）を示すことを目的とする試験では、差が検出されるであろうことの保証、すなわち分析感度の保証

の助けとなるであろう試験の質を保証するために様々な努力をしようとする動機ははるかに弱いだらう。例えば、服薬不遵守、高いプラセボ反応率、ある種の併用療法、結果の誤分類等の、治療間で観察される差を小さくする種類の誤りは、分析感度を保つ上で特に重要な問題となる。しかし、新薬が対照薬に比べ実際に優れているとの確信がある場合には、非劣性の限界値だけの差があるという仮説を否定し易くする（非劣性を検証する）ために質の高い試験を実施しようとする強い動機が生じるかもしれない。ある種の試験の誤りはバラツキを増加させ、信頼区間を広くして、限界値を超える被験治療と対照治療との間の差を否定できなくするため、非劣性を示す可能性を下げることには注意すべきであり、それゆえ、バラツキをもたらす原因（例えば未熟な測定技術）を減らそうとする強い動機は、非劣性試験において働きうる。

既に述べたように、ある非劣性試験が適切に実施されたかどうかを判断するためには、その試験の実施に関して、治療間の差を曖昧にする可能性のある要因の存在、及び、非劣性の限界値の決定根拠を与えた試験と当該試験の間に違いを生じさせるような要因の存在についても検討しなければならない。特に、組み入れられた集団、併用療法の利用、治療のコンプライアンス、そして脱落の程度とその理由等において見られる（決定根拠を与えた試験との）何らかの違いによる影響で、分析感度が低下する可能性があるかどうかを判断すべきである。ある試験のデザイン及び実施のあり方が、非劣性の限界値の決定根拠を与えた試験と非常に類似しているように見える場合でも、実薬対照による治療の結果が明らかに通常のそれと異なれば（例えば、ある抗生物質の試験で異常に高い又は低い治癒率が見られた場合等）、それは重要な相違が存在したことを示す可能性もある。

### 1.5.2 優越性の証明を目的とした試験における分析感度

分析感度の問題は、特に非劣性試験において重要であるが、実際には、治療間の差を検出することに失敗した全ての試験において生じるものであり、そこにはプラセボ対照試験や用量反応試験も含まれる。例えば、ある治療のプラセボに対する優越性が証明されなかった場合に、証明されなかったという事実は、その治療が無効であったか、あるいは、そのようにデザインされ実施された試験が、有効な治療とプラセボを区別する力を持たなかったことを意味している。

実薬対照試験およびプラセボ対照試験の分析感度を評価する有用な方法の一つが、プラセボと既知の実薬の双方を含む「三群比較試験」である。この試験デザインにはいくつかの長所がある。三群比較試験では効果の大きさを測定することができ（被験薬対プラセボ）、さらに実薬対照とプラセボの比較によって分析感度が確立された条