

INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL  
REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN  
USE

**ICH HARMONISED TRIPARTITE GUIDELINE**

**CHOICE OF CONTROL GROUP AND RELATED  
ISSUES IN CLINICAL TRIALS**

Recommended for Adoption  
at Step 4 of the ICH Process  
on 20 July 2000  
by the ICH Steering Committee

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.

<<< ICH E10 >>>

## ICH E10

# 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

医 薬 審 発 第 1 3 6 号  
平成 1 3 年 2 月 2 7 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長

「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 E U 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。

(別添)

## 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み合わせ
- 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っただろう」といったことを仮定することはできる。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うる被験者の管理、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

果を期待するが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。効果や忍容性に基つ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・試験実施医師の期待、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ない）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、被験者と試験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

いずれの種類の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、ある特定の实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も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「対照治療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ろう。

「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、症候性うっ血性心不全、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験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ろう。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、試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ない。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र्ट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、上乗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乗せし、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しか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る。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る。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、固定期間とするのもあれば、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 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ろう。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、また試験が患者や試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さう。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う。

### 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るう。

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ない、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る。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に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った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ろう。

#### 2.3.6.2 倫理上の長所がありうること

特に、用量に関連した有効性や用量に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しての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）がより適切である

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 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択されるため、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いなければならない。

### 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当とされるのは、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る。

### 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で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。対照試験を行ってみたところ、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、治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だろう。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採用することもある。

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シ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ってである。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。

表1．種々の状況における各同時対照の有用性

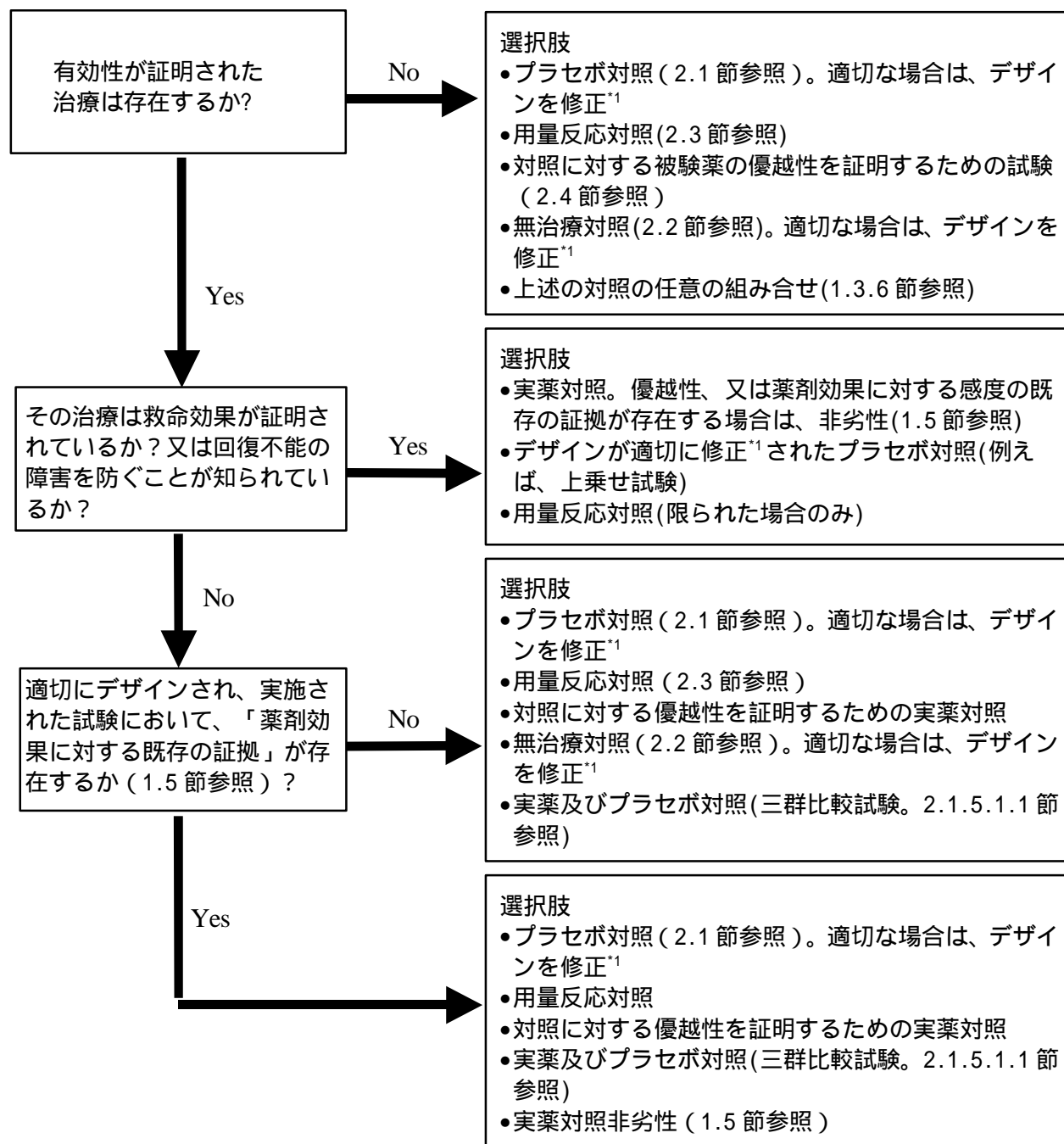
試験の目的	対照の種類							
	プラセボ	実薬 (非劣性)	実薬 (優越性)	用量反応	プラセボ +実薬	プラセボ +用量反 応	実薬+ 用量反応	プラセボ +実薬 +用量反応
「絶対」的な 効果の大き さの測定	Y	N	N	N	Y	Y	N	Y
効果の存在 の証明	Y	P	Y	Y	Y	Y	Y	Y
用量反応関 係の証明	N	N	N	Y	N	Y	Y	Y
治療間の比 較	N	P	Y	N	Y	N	P	Y

Y = Yes, N = No, P = 可能

(「薬剤効果に関する感度の既存の証拠」があるか否かによる)

## 図１．有効性を証明するための同時対照の選択

この図は対照群を選択する際の基本的な考え方を示すものである。実際の決定は個別の地域において利用可能な薬剤や医療に依存する。



\*1. 上乗せ、置換え、早期離脱、短期のプラセボ期及びランダム化治療中止試験 (2.1.5.2 節参照)。

(参考)

## **「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」 に関する質疑応答集**

**Q1. 本ガイドラインの位置付けを明らかにして頂きたい。**

**A1.**

本ガイドラインは、「1.1 ガイドラインの構成の概略と目的」に述べられているとおり、医薬品の承認審査資料を収集する目的で実施される臨床試験、特に有効性を証明するための試験において、適切な対照群を選択するための一般的な考え方を、科学的、倫理的及び試験の実務的な観点から示したものである。

本ガイドラインは、個々の臨床試験における対照群の選択に関するものであり、臨床データパッケージ全体の構成のあり方に関する指針を示すものではない点に注意が必要である。

本ガイドラインに示された考え方は、医薬品の有効性評価において従来からきわめて重要とされているものであり、どの時点で実施された試験についても適用可能である。

**Q2. 本ガイドラインを実際の試験に適用する際に必要な対応等があれば挙げて頂きたい。**

**A2.**

本ガイドラインは、対照群選択の「考え方」を示すものであり、試験実施の手続き等を具体的に定めることを目的とするものではない。しかし、承認審査の段階で、本ガイドラインに基づいた検討が試験のデザイン、実施、解析等の各段階で適切に行われたことを実際の記録等により確認できるような形で試験関連文書（総括報告書等）を作成することにご配慮頂きたい。

具体的には、例えば、「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）」における「9 治験の計画」、「10 治験対象患者」、「11 有効性の評価」に、必要に応じて、本ガイドラインに係る記載を盛り込むべきである。また、今後、承認申請の際に提出されるコモンテクニカルドキュメントにおいても、本ガイドラインに対応する記載が求められている箇所があることにご注意頂きたい。

**Q3. 「分析感度」についての平易な解説をお願いしたい。特に非劣性の証明により有効性を主張しようとする場合の分析感度を保証する方法を示して頂きたい。**

**A3.**

本ガイドラインで「分析感度」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（1.5 節）」とされている。本ガイドラインでは、承認審査における有効性の証明の重要性及びそれに伴う困難さを念頭に置いた上で、適切な臨床試験のデザイン・実施のあり方が論じられており、「分析感度」はその意味合いにおける概念的な用語として使用されている。

「分析感度」は、様々な領域（例えば臨床検査法の評価等）において固有の意味で用いられてきた用語であるが、本ガイドラインにおける「分析感度」もまた、本ガイドライン領域に固有の意味で使用されていることにご注意頂きたい。

非劣性の証明により有効性を主張する際の分析感度を保証するためには、本ガイドライン 1.5.1 節に示された 4 段階のステップを踏まえることが重要である。

**Q4. 「ある種の試験の誤り」が治療群間で観察される差を小さくする場合と、ばらつきを増加させて信頼区間を広くする場合のそれぞれの例示と解説を願いたい。**

**A4.**

まず、「ある種の試験の誤り」が治療群間の差を小さくすることと、ばらつきを増加させて信頼区間を広くすることは排他的な概念ではないことに注意が必要である。治療群間の差を小さくする「試験の誤り」が同時に信頼区間を広くする要因であることもあり、全ての場合に非劣性検証を容易とする、あるいは逆に検証を困難にするとは限らない。臨床的に不適切な評価変数の採用、評価基準の不遵守、許容できる併用薬・併用治療からの逸脱、組み入れ基準違反等は状況によってその影響が変わりうる。

治療群間の差を小さくする「試験の誤り」の例としては、臨床的な差を捉えられない尺度による評価、不適切な評価時期等が挙げられる。このような問題がある場合には、両群間の差が小さくなることで、結果として本来は対照薬に劣っている薬剤が誤って対照薬に対し劣らないとの判断を下してしまう可能性が生じる。信頼区間を広くする「試験の誤り」の例としては、目標とした被験者数が集積できない場合はもちろんのこと、評価変数が計量値の場合に、施設間の差が十分検討されていない測定機器を用いること、測定精度が適切でないこと等が挙げられる。これらによって、予め想定していないばらつきが生じることにより、目的とした仮説を検証できない可能性が生じる。

いずれの場合であっても「ある種の試験の誤り」が試験結果に与える影響の大きさについて一般的な目安を設けることは不可能である。しかし、試験計画時には、既存の情報から各種の誤りが試験結果にどの程度の影響を生じさせるか十分検討することが重要である。



**Q5. 本ガイドラインには「薬剤効果に対する感度の既存の証拠」が保証されているとは言い難い領域が例示されているが（1.5.1.1 節）これらの領域では、我が国においても優越性試験（特にプラセボ対照試験）でなければ有効性の検証試験とみなされないのか。**

**A5.**

ガイドラインの当該箇所に併せて記載されているように、「薬剤効果に対する感度の既存の証拠」の有無は、「ある程度までは判断の問題」である。1.5.1.1 節に記されたのは、「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」疾患の例であって、かかる疾患領域において試験が実施された全ての薬剤について「薬剤効果に対する感度の既存の証拠」が存在しえないという意味ではない。

また、医薬品の有効性に関する主張・判断が、単独の試験のみに基づき行われることは稀である。通常、そのような主張・判断は、複数の探索的・検証的試験からなる臨床データパッケージに基づいて行われており、また、行われるべきものである。

しかし、本ガイドラインにおいて繰り返し述べられているように、優越性試験を採用することにより、薬効評価における非劣性試験の短所を回避できるのは事実である。臨床データパッケージにいかなる試験を盛り込むかを慎重に検討した上で、可能な場合には、プラセボ対照試験を含む優越性試験を積極的に採用すべきである。

**Q6. 各種対照における試験デザイン上の工夫として挙げられている例を説明して頂きたい。**

**A6.**

本ガイドラインでは、臨床試験の倫理性、科学性、実施可能性等の観点から、試験の目的に応じてそれぞれの対照の長所が十分に活かされ、また短所がカバーされるよう、試験デザイン上の工夫がいくつかの具体例とともに説明されている。

例えば、プラセボ対照試験では、対照群の追加、標準治療への上乗せ、既存治療の置換え、無効な治療からの早期離脱、プラセボ期を限定すること、ランダム化治療中止等のデザイン上の修正、工夫が挙げられている（2.1.5 節参照）。実薬対照試験では、複数の用量群の設置、試験の初期や最後に短期間のプラセボ対照期を設けること等の例が挙げられている（2.4.5 参照）。

**Q7. 「相対的な有効性、安全性（1.4.2 節）」とはどういう意味か。特に承認取得という目的にどのように寄与すると考えられるのか。**

**A7.**

本ガイドラインの 1.4.2 節は、試験の目的が、被験薬の有効性（安全性）の検証自体ではなく、被験薬とそれ以外の治療の有効性（安全性）の比較である場合を指している。比較には、全般改善度（有効率）といった総合的な評価項目が用いられることもあれば、直接に計測される数値化された指標が用いられることもあるであろう。なお、本節で説明されているのは薬剤（治療）間の相対比較であり、ある薬剤（治療）自身のリスクベネフィットの比較考量ではない点に注意が必要である。

「相対的な有効性、安全性」に関する試験の必要性・有用性は、各地域において承認申請の際に求められる臨床データパッケージの内容によって異なる。