

第110回日本精神神経学会学術総会

教育講演

臨床試験のデザイン，データの確認ポイント，そして読み解き方

中林 哲夫（独立行政法人医薬品医療機器総合機構新薬審査第三部）

治療環境の向上には，既存の治療法のエビデンスを蓄積し治療戦略を確立するとともに，新たな治療法の開発を行っていくことが重要である。近年，精神神経疾患領域では，数多くの新薬が本邦でも承認されるなど，治療法の選択肢は拡大した。数ある治療法の中で最適な治療方法を合理的に選択するには，より多くの情報を収集し，エビデンスを適切に理解することが必要になる。このため本稿では，有効性や安全性の直接的な証拠となる臨床試験に焦点をあて，試験実施計画やデータの確認ポイントについて概説する。そして，精神神経疾患を対象とした臨床試験の特徴についても説明し，臨床医にとって必要な情報を中心に，結果の解釈を行う上での留意点を解説する。

<索引用語：精神神経疾患，臨床試験，頑健性，プラセボ反応性，バイアス>

はじめに

精神神経疾患の代表的な治療方法には，薬物療法と心理社会的治療がある。そして，治療環境の向上には，既存の治療法のエビデンスを蓄積し治療戦略を確立するとともに，新たな治療法の開発を行っていくことも重要である。

薬物療法については，精神神経疾患領域の医薬品開発は国際的にも活発^{1,12)}であり，近年は数多くの向精神薬が本邦でも承認^{26,28)}された。これにより，欧米で標準治療薬に位置づけられる抗精神病薬^{2,31)}，抗うつ薬^{7,29)}，そして抗てんかん薬^{8,30)}の大半が本邦でも使用可能となった。すなわち，当該領域のドラッグ・ラグ（欧米で承認されている医薬品が日本においては未承認であり，国民に提供されていない状態）はおおむね解消し，薬物療

法の選択肢は広がった。

では日々の臨床では，数ある選択肢の中でどの治療法が選択されるのか？ これには，臨床的疑問を明確化し，最適な治療方法を合理的に選択することが望まれる。そしてエビデンスの適切な理解には，検定統計量（ p 値）の確認のみでなく，研究目的と評価方法，そして結果の十分な吟味が重要となる。このため本稿では，有効性や安全性の直接的な証拠となる臨床試験に焦点をあて，試験実施計画やデータの確認ポイント，そして結果の解釈を行う上での留意点を，以下について考慮して概説する。

- ・精神神経疾患を対象とした臨床試験に重点を置く。
- ・多忙な臨床医でも効率的に臨床試験の結果を理

第110回日本精神神経学会学術総会＝会期：2014年6月26～28日，会場：パシフィコ横浜

総会基本テーマ：世界を変える精神医学——地域連携からはじまる国際化——

教育講演：臨床試験のデザイン，データの確認ポイント，そして読み解き方 座長：森信 繁（高知大学医学部神経精神科学教室）

表1 試験実施計画の構成要素

1. 目的および仮説
2. 試験デザイン
・無作為化
・対照の設定
・盲検化
3. 対象集団
・選択基準 (対象疾患)
・除外基準
・目標例数
4. 試験治療
・被験薬
・用法・用量
・投与期間
・前治療および併用療法
5. 評価項目
・有効性評価項目 (主要評価項目, 副次評価項目)
・安全性評価項目
6. 解析計画
・主要解析
・副次解析

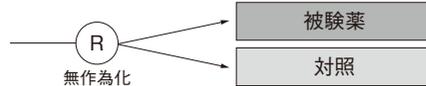
解できるよう、特に確認が必要と思われる点に限定する。

本稿は、論文を読む際の留意点を整理するものであり、統計学的な検定方法の詳細を知りたい場合や臨床試験を計画する場合には成書を参照いただきたい。なお、意図を明確にするために薬物療法に関連した用語を中心に用いるが、心理社会的治療や医療機器の臨床試験でも読み替えが可能な内容とする。また、本稿は筆者個人の意見に基づいた概説であり、医薬品医療機器総合機構 (Pharmaceuticals and Medical Devices Agency : PMDA) の見解を示すものではない。

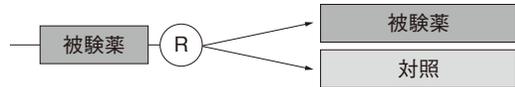
I. 臨床試験の計画の確認ポイント

臨床試験データの確認は、試験実施計画の把握から始まる (表1)。試験実施計画は複数の項目で構成され、全項目の設定内容の把握が基本的には必要となる。なかでも、基本骨格である試験デザインはその種類ごとに特徴が異なるため、必要な情報に応じて適した試験デザインの試験データを確認する。

①プラセボ対照試験, 実薬対照試験



②ランダム化治療中止試験



③クロスオーバー試験



④非盲検非対照試験 (オープン試験)



図1 精神神経疾患領域で用いられる代表的な試験デザイン

1. 臨床試験のデザイン (図1, 表2)

精神神経疾患領域で用いられる代表的な試験デザインと各デザインの特徴を説明する。

無作為化比較試験 (randomized controlled trial : RCT) は、各被験者に、評価対象である治療薬 (被験薬) と対照 (control) となる比較薬剤のいずれかを無作為に割付け (randomization)、比較する試験である。精神神経疾患領域の代表的な RCT には、プラセボ対照試験、実薬対照試験、ランダム化治療中止試験、クロスオーバー試験などがある (図1)。非盲検非対照試験 (オープン試験) は、対照と無作為割付けが設定されず、RCT と区別される。

1) プラセボ対照試験と実薬対照試験

プラセボ対照試験と実薬対照試験は、RCT の中で最も一般的な試験デザインである。

実薬対照試験の結果から、他治療薬との比較情報、つまり「相対的な効果」の大きさに関する情報が得られる。そして、有効性の検証を目的とした実薬対照試験は、非劣性試験と優越性試験などに分けられる。非劣性試験は、被験薬の効果が比較薬剤 (対照) に対して臨床的に劣らないことを

表2 代表的な試験デザインの特徴（利点と欠点）

		特徴		結果の意義など
		利点	欠点	
無作為化比較試験 (RCT)	プラセボ対照試験	<ul style="list-style-type: none"> 絶対的な効果の大きさを測定できる 分析感度の証拠が内在している 必要となる被験者数が，実薬対照試験と比べて少ない 期待バイアスを最小化できる 	<ul style="list-style-type: none"> プラセボ群に割り当てられることへの躊躇 	医薬品の基本情報を検討でき，検証的試験に用いられる <ul style="list-style-type: none"> 有効性プロファイル 安全性プロファイル 用量反応関係 など
	実薬対照試験	<ul style="list-style-type: none"> 相対的な効果の大きさを測定できる 中止・脱落が少なくなる可能性がある 対照（実薬治療）に対する優越性が示されれば，有効性の証拠として容易に解釈が可能である 	<ul style="list-style-type: none"> 必要な被験者数は多くなる 非劣性試験では，対照群の効果に対する信頼性が十分に保証されないことがある（無効同等の可能性） 	他治療薬との比較情報を検討できる
	ランダム化治療中止試験	<ul style="list-style-type: none"> プラセボの投与期間を短くすることができる 	<ul style="list-style-type: none"> 被験薬に効果を示した被験者のみに強化(enriched)され，有効性が過大評価される可能性がある 離脱症状が発現する可能性がある 	再燃の抑制効果（効果の持続性）を検討できる
	クロスオーバー試験	<ul style="list-style-type: none"> 精度を最大化できる 必要な被験者数は比較的少ない 被験者ごとの試験治療の効果についての情報が得られる 	<ul style="list-style-type: none"> 持越し効果や時期効果の影響を受けることがある 前提条件（症状が安定している慢性疾患，被験薬の効果発現が速い，被験薬の効果が可逆的，治療中止後に被験者は基準値の状態に戻る）が多い 試験期間が比較的長い 1人の被験者の中止の影響が大きい 	検証的試験には通常は用いられない 薬物動態の検討などで用いられることが多い
非盲検非対照試験 (オープン試験)		<ul style="list-style-type: none"> 全ての被験者に被験薬投与を行うことができる 	<ul style="list-style-type: none"> 期待バイアスや平均への回帰現象により，有効性が過大評価される可能性がある 	長期投与時の安全性の検討で用いられることが多い

示すことが目的の試験である。また，優越性試験は，被験薬の効果が対照である比較薬剤に対して臨床的に優れていることを示すことが目的の試験である。実臨床では，治療薬間の比較情報が得られる実薬対照試験の情報は有益かもしれない。しかし，非劣性試験では，被験薬や比較薬剤（対照）とプラセボの効果が同程度，つまり無効同等の可

能性³⁸⁾が排除できないことに注意を要する。

精神神経疾患を対象とした臨床試験の最大の特徴は，プラセボ反応性が小さくはなく，そして一定しないこと^{4,18,19)}である。このため，医薬品開発での有効性の検証（confirmation）は，有効な治療法と無効な治療法を区別する力，つまり分析感度（assay sensitivity）が内在するプラセボ対照試

験が国際的にも必要と考えられている^{3,4,19,20}。プラセボ対照試験は医薬品の基本情報を把握するのに有益であり、有効性や安全性のプロファイル、そして用量群が複数の試験では用量反応関係を評価することができる。

2) ランダム化治療中止試験

ランダム化治療中止試験は、被験薬を一定期間投与した後に、被験薬の継続またはプラセボ（被験薬の中止）投与のいずれかを無作為に割付ける試験デザインである。

プラセボ対照試験や実薬対照試験が、主に急性期症状に対する効果を検討するデザインであるのに対して、ランダム化治療中止試験は再燃の抑制効果、つまり効果の持続性を検討する試験デザイン^{4,18,19}である。ランダム化治療中止試験は、離脱症状を評価する試験デザインと一部で誤解されることがある。当該デザインは、プラセボの投与期間を短くすることができる利点を有するが、被験薬に効果を示した被験者のみに強化（enriched）された結果が得られるため、有効性が過大評価される可能性もある。

3) クロスオーバー試験

クロスオーバー試験は、同一の被験者に被験薬と比較薬剤（対照）を投与し、投与順序を無作為に割付けする試験デザインである。当該試験の利点は、少ない被験者数で実施可能なことだが、欠点は持越し効果や時期効果の影響を受ける可能性があることである。2群2期のクロスオーバー試験を例（図1）にとると、持越し効果とは、第1投与期の効果が第2投与期に残存することである。そして時期効果とは、同じ被験者でも個体内変動により第1投与期と第2投与期の初期条件が異なることである。クロスオーバー試験では、前提条件（症状が安定している慢性疾患が対象であること、被験薬の効果発現が速いこと、被験薬の効果が可逆的であること、治療中止後に被験者は基準値の状態に戻る）が多く、当該デザインによる有効性評価は失敗も多いため、一般的には検証的試験のデザインとして用いられない。

4) 非盲検非対照試験（オープン試験）

非盲検非対照試験（オープン試験）は、投与前後の比較により有効性評価が行われる。投与前後の比較では、薬剤投与による期待バイアスや平均への回帰現象（regression to the mean）により有効性がしばしば過大評価されるため、当該デザインによる有効性の検証（confirmation）は、基本的には困難である。平均への回帰現象とは、慢性疾患を対象とした臨床試験では被験者は状態が悪い時期に参加することが多く、参加後にランダムな変動で改善に向かい、試験開始後に見かけ上の改善が認められる現象^{6,18,35}である。

2. 主要評価項目（表3）

臨床試験の評価項目には主要評価項目と副次評価項目があり、通常は1つの試験で1つの主要評価項目が設定される。殊に、主要評価項目は、臨床試験の主要な目的に直結し、試験データの確認では最も重要な評価項目となる。大うつ病性障害を対象とした臨床試験では Montgomery-Åsberg Depression Rating Scale (MADRS) とハミルトン大うつ病性障害評価尺度 (Hamilton Depression Rating Scale: HAM-D) が、そして統合失調症を対象とした臨床試験では Positive and Negative Syndrome Scale (PANSS) が主要評価項目として汎用（図2）されている。主要評価項目の適切性の吟味が必要な場合は、臨床的な適切性、信頼性および妥当性の確立、そして使用実績を確認する。

3. サンプルサイズの設定（表3）

サンプルサイズは、その設定根拠を確認する。有効性の検証（confirmation）を目的とした臨床試験のサンプルサイズは、第一種の過誤（ α ）、第二種の過誤（ β ）、予想される群間差（ Δ ）とバラツキの大きさ（SD）により、数学的に設計^{9,23}される。被験者数が事前に計画したサンプルサイズより小さすぎれば、臨床的に意味のある効果が見逃される可能性が高くなり、逆に大きすぎれば意味のない効果を検出する可能性が高くなる。第一

表3 臨床試験の計画の確認ポイント

	確認内容
臨床試験の目的・位置づけ	<ul style="list-style-type: none"> ・試験の目的として，探索（exploration）と検証（confirmation）の区別 ・探索的試験で，統計学的に有意な結果が得られた場合，有効性を結論づけていないこと
試験デザイン	<ul style="list-style-type: none"> ・必要な情報に適した試験デザインであること
用法・用量	<ul style="list-style-type: none"> ・診療での使用方法と乖離がないこと
主要評価項目	<ul style="list-style-type: none"> ・主要評価項目の適切性について吟味が必要な場合 ・臨床的に最も適切で説得力のある証拠を与える変数であること ・信頼性および妥当性が確立している変数であること ・先行研究などで使用実績があること
サンプルサイズの設定根拠	<ul style="list-style-type: none"> ・サンプルサイズの設計条件の適切性 ・α：第一種の過誤（偽陽性）の確率（検定の有意水準） 原則として，片側仮説では2.5%，両側仮説では5% ・β：第二種の過誤（偽陰性）の確率 慣例として10～20% ・Δ（delta）：予想される群間差 ・SD：バラツキの大きさ
実際の被験者数	<ul style="list-style-type: none"> ・計画より極端に被験者数が少ないこと ＝検出力（1－第二種の過誤の確率）の不足 ・必要以上に被験者数が多くないこと ＝臨床的に意味のない微妙な効果を検出する可能性/倫理的問題

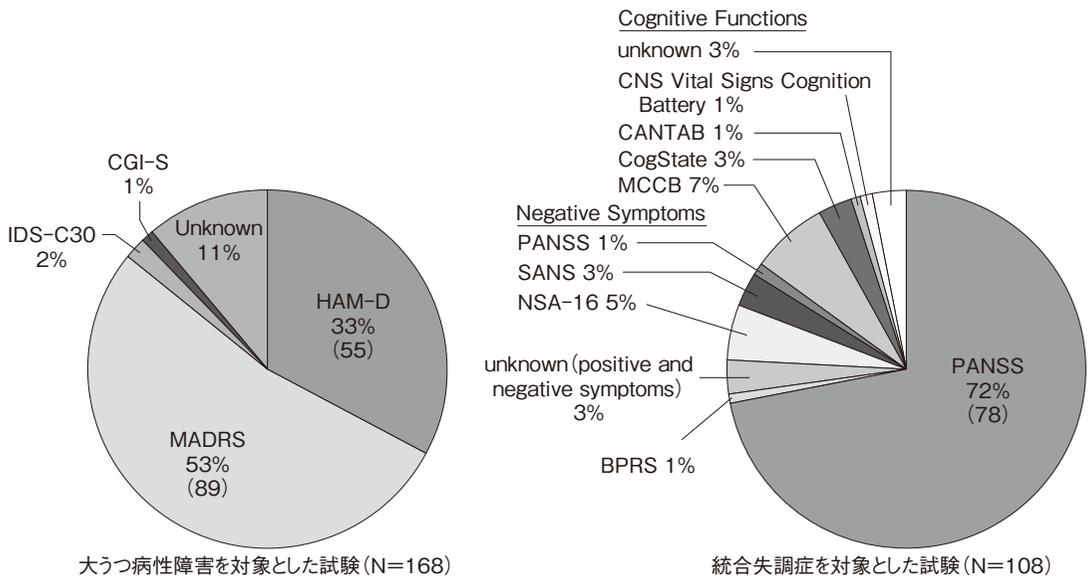


図2 主要評価項目に用いられる症状評価尺度（大うつ病性障害，統合失調症）

情報源：ClinicalTrials.gov

検索条件：Condition：schizophrenia and major depressive disorder phase II or phase III

Study design：randomized, double-blind, parallel-group, placebo-controlled study (without withdrawal study)

Age：adult (18-65 years old) or senior (over 65 years old)

Primary endpoint：efficacy assessment Sponsored by industry

Start year：2004-2014

Locations：all countries

表 4 臨床試験データの確認ポイント

		確認内容
単一の試験	対照集団の群間での比較可能性	群間での偏りの有無 ・被験者背景 (診断, 年齢, 性別, 体重, 重症度, 病型, 罹病期間など) ・中止・脱落の割合 ・中止理由 (効果不十分, 有害事象による中止)
	主要解析と副次解析の結果	・主要解析での統計学的に有意な差が得られていること ・副次解析の結果のみによる有効性の主張が行われていないこと ・結果の頑健性 (主要解析と副次解析, 感度分析)
	得られた結果の臨床的意義	・主要評価項目 (エフェクトサイズ) と副次評価項目 (レスポンドー解析) の総合的判断
複数の試験	臨床試験間の結果比較	・1つの治療薬でも複数の臨床試験結果がある場合, 全ての試験結果を確認し, 結果の一貫性 (頑健性) があること ・複数の臨床試験のエフェクトサイズを直接比較し, 異なる治療薬の有効性を比較できないこと

種の過誤 (片側仮説で 2.5%, 両側仮説で 5%), 第二種の過誤 (10~20%) の設定²¹⁾に加え, 群間差 (Δ) とバラツキの大きさ (SD) の見積もりで, 適切な臨床試験成績が参照されているかも確認する (表 3)。

全く異種の治療方法の成績を参照してサンプルサイズが設計された臨床試験の報告も散見されるが, この場合は有効性を結論づけることには限界があることも少なくない。小規模の臨床試験で統計学的に有意な結果が認められた場合, その治療法の有効性は大きいと誤解されることがある。しかし, 実際の被験者数が目標例数にはるかに達していないにもかかわらず, 統計学的に有意な結果が得られた場合は, バイアスが原因した可能性も考慮する。

II. 臨床試験データの確認ポイント (表 4)

臨床試験には疾患領域に特異的な特徴^{25,27)}があり, この特徴の把握は試験データの確認ポイントを整理するのに役立つ。精神神経疾患を対象とした臨床試験では, 症状評価尺度を用いた有効性評価, つまり主観的評価が一般的³⁾であり, この評価方法に関連した特徴がみられる。この特徴を考慮して臨床試験データの確認ポイントを説明する。

1. 対象集団の群間での比較可能性

対象集団については, 被験者背景, 被験者の中止・脱落の割合やその理由が, 群間 (被験薬群と対照群) で偏りがいないかを確認する。

対象集団の重症度は, 選択基準において症状評価尺度のカットオフ値により規定される (図 3)。臨床試験での有効性評価は, エフェクトサイズ (症状評価尺度の合計スコアの変化量の群間差) に基づいて行われるが, このエフェクトサイズはベースラインの重症度の影響を受ける。ベースラインの重症度が高い試験では低い試験と比較して, エフェクトサイズは大きくなる傾向^{5,15~17,32)}がある。つまり, 群間でベースラインの重症度に差異がある場合は, 有効性の比較が難しいこともある。

中止・脱落の理由の確認も行う。試験治療の用量が高すぎると有害事象による中止が増加し, 逆に低すぎると効果不十分による中止が増加するかもしれない。このため, 実薬対照試験で, 中止・脱落が対照群に多く認められる場合は対照群に不利な用量設定となっていなかったか, つまり用量不足または過剰の有無を吟味する。

2. 主要解析と副次解析の結果

臨床医にとって, どのような患者で治療効果が

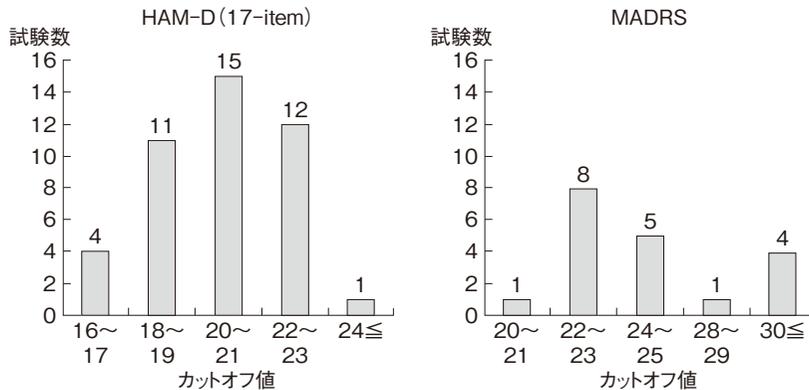


図3 選択基準における症状評価尺度のカットオフ値（大うつ病性障害を対象とした臨床試験）

情報源：ClinicalTrials.gov

検索条件：Condition：schizophrenia and major depressive disorder phase II or phase III

Study design：randomized, double-blind, parallel-group, placebo-controlled study (without withdrawal study)

Age：adult (18-65 years old) or senior (over 65 years old)

Primary endpoint：efficacy assessment Sponsored by industry

Start year：2004-2014

Locations：all countries

期待できるかは重要な情報である。このため、背景因子（年齢、病型、罹病期間など）による部分集団解析の結果は関心が深いかもしれない。しかし有効性を結論づけられるのは、基本的には検証的試験（confirmatory study）の主要解析の結果である。副次解析で行われる部分集団解析は、多くの場合、効果の一樣性を確認する探索的解析²¹⁾であり、部分集団解析の結果のみで有効性を主張できない。

精神神経疾患を対象とした試験での被験者の中止・脱落の割合は、他疾患と比較して高いことが経験される。そしてプラセボ対照試験での中止・脱落の割合は、実薬対照試験と比較して高い^{13,34)}。このため、感度分析（sensitivity analysis）の結果が掲載されている場合は、その結果も確認する。

感度分析では、主要解析と異なる欠測値の取扱いや異なる解析対象集団などで解析を行い、結果の一貫性（頑健性）が検討される。欠測値の取扱い、例えば LOCF（last observation carried forward）と OC（observed cases）で異なる結果が得られた場合は、被験者の中止・脱落が結果に影響した可能性があり、中止・脱落の理由を確認す

る。早期脱落の影響を受けやすい LOCF 解析で、症状評価尺度スコアの変化量が小さくなる投与群があれば、投与初期の忍容性不良や効果の発現の遅いことでの中止がなかったか吟味する必要があるかもしれない。

解析対象集団により異なる結果が得られた場合、例えば、FAS（Full Analysis Set：最大の解析対象集団）、PPS（Per Protocol Set：試験実施計画書に適合した対象集団）で異なる結果が得られた場合は、PPS での除外理由の確認を行う。

3. 臨床試験間の結果の比較

臨床医にとってより確実な治療薬を選択することは重要である。このため、1つの治療法でも複数の試験結果がある場合、多くの情報を収集し、試験結果の一貫性（頑健性）を吟味する。

また臨床医にとってより有効な治療薬を選択することも重要であり、複数の治療薬の比較情報には関心が深い。治療薬の比較情報は、実薬対照試験の結果を参照するのが基本であり、エフェクトサイズなどの試験データを、試験間で直接比較することは原則行えない。これは、臨床試験の結果

は、臨床試験の実施時期、試験デザインや試験環境など、多くの変動因子が関係し、各試験の対象集団間の比較可能性が保証されないためである。プラセボに対する反応性が経年的に大きくなっていることも指摘^{14,16,22,39)}されており、実施時期が異なれば、症状評価尺度スコアの変化量などが異なる可能性も否定できない。試験デザインは、試験結果に影響を与えること²⁷⁾からも、各試験の実施計画を確認し、結果を十分に吟味することが重要と考える。

おわりに

精神神経疾患領域では、数多くの向精神薬が本邦でも承認されたように、治療の選択肢は拡大した。最後に当該疾患領域の医薬品開発の現況にふれ、臨床試験での有効性と安全性の評価、つまり臨床評価における課題について私見を加える。

これまでの医薬品開発は、非定型抗精神病薬、選択的セロトニン再取り込み阻害薬 (selective serotonin reuptake inhibitor : SSRI) やセロトニン・ノルアドレナリン再取り込み阻害薬 (serotonin-noradrenaline reuptake inhibitor : SNRI) のように、リスク・ベネフィットバランスの改善が主体²⁷⁾であった。しかし、現在の標準治療にも臨床上の課題 (unmet needs) は残されている。例えば大うつ病性障害については、大規模臨床研究である STAR*D (Sequenced Treatment Alternatives to Relieve Depression Study) の結果から、複数の抗うつ薬への切り替えでも十分な効果が得られない患者が一定数存在することが示されており、標準治療薬の効用に限界があることも示されている^{33,36,37)}。このため、うつ病治療薬では SSRI や SNRI で効果不十分な患者に対する第二選択薬の開発に移行している。そして統合失調症治療薬では認知機能障害に対する治療薬が、アルツハイマー型認知症では疾患修飾薬が開発されているように、今日では新たな臨床的位置づけの医薬品開発が行われている²⁷⁾。新規性の高い治療法はエビデンスや実臨床での使用経験も少なく、確実に医療現場に定着させるには、臨床評価に対する

専門家の育成と教育体制の強化が引き続き必要と考える。

①臨床評価に対する専門家の育成

日米 EU 医薬品規制調和国際会議 (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use : ICH) は、臨床試験の計画と実施に関する基本的考え方をガイドライン¹¹⁾として取りまとめている。この ICH ガイドラインは総論的考え方であり、各論として精神神経疾患領域に特異的な課題を抽出し、より適切な臨床評価方法を今後も継続的に検討していく必要がある。

PMDA では、次世代審査・相談体制準備室¹⁰⁾を発足し、臨床データなどを活用した解析や研究を進める体制構築²⁴⁾の準備を進めている。ドラッグ・ラグがおおむね解消された当該領域では新たな世代の治療薬の開発が望まれるが、関連学会でも橋渡し研究に関する活動を開始している。精神神経疾患領域でも、Academic Research Organization (ARO) 機能を有する研究機関ができており、これらの機関は臨床試験の計画から実施までを経験でき、方法論の蓄積が期待される。

②臨床評価に対する教育体制の充実化

これまでは生物統計学の専門書はあっても、臨床評価の方法論に関する講座や解説書は限られ、エビデンスの解釈に必要な知識を体系的に得ることができる機会は少なかった。今後も科学性の高い臨床試験を適切に計画し実施することは必要であり、適切なエビデンスの普及のために、臨床評価に対する教育体制の充実も望まれる。

現在の標準治療に残された臨床上の課題を解決していくために、既存の治療法のエビデンスを蓄積し治療戦略を確立するとともに、新たな治療方法の開発が必要となる。そしてエビデンスが適切に普及され、今後も医療環境が向上していくことに期待したい。

なお、本論文に関連して開示すべき利益相反はない。

文 献

- 1) Berggren, R., Moller, M., Moss, R., et al.: Outlook for the next 5 years in drug innovation. *Nat Rev Drug Discov*, 11 (6); 435-436, 2012
- 2) Dixon, L., Perkins, D., Calmes, C.: Guideline watch (September 2009): Practice guideline for the treatment of patients with schizophrenia, 2009. American Psychiatric Association. available at: <http://psychiatryonline.org/content.aspx?bookid=28§ionid=1682213> (Accessed October 1, 2014).
- 3) Downing, N. S., Aminawung, J. A., Shah, N. D., et al.: Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA*, 311 (4); 368-377, 2014
- 4) European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP): Guideline on clinical investigation of medicinal products in the treatment of depression (EMA/CHMP/185423/2010 Rev. 2). 30 May 2013
- 5) Fournier, J. C., DeRubeis, R. J., Hollon, S. D., et al.: Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*, 303 (1); 47-53, 2010
- 6) Gallin, J. I. 編: NIH 臨床研究の基本と実際. 丸善, 東京, 2004
- 7) Gelenberg, A. J.: Practice guideline for the treatment of patients with major depressive disorder, third edition. American Psychiatric Association, 2010. available at: <http://psychiatryonline.org/content.aspx?bookid=28§ionid=1667485> (Accessed October 1, 2014)
- 8) Glauser, T., Ben-Menachem, E., Bourgeois, B., et al.: Updated ILAE evidence review of antiepileptic drug efficacy and effectiveness as initial monotherapy for epileptic seizures and syndromes. *Epilepsia*, 54 (3); 551-563, 2013
- 9) 浜田知久馬: 学会・論文発表のための統計学. 統計パッケージを誤用しないために. 真興交易医書出版部, 東京, 2012
- 10) 医薬品医療機器総合機構: 次世代審査・相談体制について (申請時電子データ提出). available at: <http://www.pmda.go.jp/operations/shonin/info/iyaku/jisedai.html> (Accessed October 1, 2014)
- 11) 医薬品医療機器総合機構: 日米 EU 医薬品規制調和国際会議. available at: http://www.pmda.go.jp/ich/ich_index.html (Accessed October 1, 2014)
- 12) Karlberg, J. P.: Trends in disease focus of drug development. *Nat Rev Drug Discov*, 7 (8); 639-640, 2008
- 13) Kemmler, G., Hummer, M., Widschwendter, C., et al.: Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. *Arch Gen Psychiatry*, 62 (12); 1305-1312, 2005
- 14) Kemp, A. S., Schooler, N. R., Kalali, A. H., et al.: What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it? *Schizophr Bull*, 36 (3); 504-509, 2010
- 15) Khan, A., Leventhal, R. M., Khan, S. R., et al.: Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol*, 22 (1); 40-45, 2002
- 16) Khin, N. A., Chen, Y. F., Yang, Y., et al.: Exploratory analyses of efficacy data from schizophrenia trials in support of new drug applications submitted to the US Food and Drug Administration. *J Clin Psychiatry*, 73(6); 856-864, 2012
- 17) Kirsch, I.: Antidepressant drugs 'work', but they are not clinically effective. *Br J Hosp Med (Lond)*, 69 (6); 359, 2008
- 18) 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. 医薬審発第 136 号, 平成 13 年 2 月 27 日
- 19) 厚生労働省医薬食品局審査管理課長: 「抗うつ薬の臨床評価方法に関するガイドライン」について. 薬食審査発 1116 第 1 号, 平成 22 年 11 月 16 日
- 20) 厚生労働省医薬食品局審査管理課長: 「睡眠薬の臨床評価方法に関するガイドライン」について. 薬食審査発 1213 第 1 号, 平成 23 年 12 月 13 日
- 21) 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. 医薬審第 1047 号, 平成 10 年 11 月 30 日
- 22) Mallinckrodt, C. H., Zhang, L., Prucka, W. R., et al.: Signal detection and placebo response in schizophrenia: parallels with depression. *Psychopharmacol Bull*, 43 (1); 53-72, 2010
- 23) 永田 靖: サンプルサイズの決め方. 朝倉書店, 東京, 2003
- 24) 内閣官房長官, 情報技術通信 (IT) 政策担当大臣, 内閣府特命担当大臣 (科学技術政策), 総務大臣, 外務

大臣, 文部科学大臣, 厚生労働大臣, 経済産業大臣, 国土交通大臣申合せ: 健康・医療戦略, 平成 25 年 6 月 14 日

25) 中林哲夫, 矢花直幸, 宇山佳明: 第 3 章第 1 節 [1] 中枢神経系疾患における指摘事項の傾向と対策. 成功・失敗の傾向, 各疾患の特徴からつかむ臨床試験計画とデザインの設定 (坂下祥深, 天願ルイス編). 科学技術情報協会, 東京, p.105-114, 2012

26) 中林哲夫: 臨床試験の有効性評価項目に関連した知見とその解釈 大うつ病性障害における抑うつ症状の軽減, 反応, 寛解, そして社会機能を中心に. 精神科治療学, 28 (4); 413-420, 2013

27) 中林哲夫: 精神神経疾患領域の臨床試験の特徴と課題—Study design, operation and outcome—lessons learned—. 臨床精神薬理, 17 (6); 783-794, 2014

28) 中林哲夫: 精神神経疾患領域における臨床試験の現状, 動向, そして課題. 臨床精神医学, 43 (12); 1705-1715, 2014

29) National Institute for Health and Clinical Excellence: Depression: Treatment and management of depression in adults (update) (NICE clinical guideline 90), 2009. available at: <https://www.nice.org.uk/guidance/cg90> (Accessed October 1, 2014)

30) National Institute for Health and Clinical Excellence: The epilepsies: the diagnosis and management of the epilepsies in adults and children in primary and secondary care (NICE clinical guideline 137), 2012. available at: <https://www.nice.org.uk/guidance/CG137> (Accessed October 1, 2014)

31) National Institute for Health and Clinical Excellence: Psychosis and schizophrenia in adults: treatment and management (NICE clinical guideline 178), 2014. available at: <https://www.nice.org.uk/guidance/CG178>

(Accessed October 1, 2014)

32) Papakostas, G. I., Fava, M.: Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur Neuropsychopharmacol*, 19 (1); 34-40, 2009

33) Rush, A. J., Trivedi, M. H., Wisniewski, S. R., et al.: Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med*, 354 (12); 1231-1242, 2006

34) Rutherford, B. R., Cooper, T. M., Persaud, A., et al.: Less is more in antidepressant clinical trials: a meta-analysis of the effect of visit frequency on treatment response and dropout. *J Clin Psychiatry*, 74 (7); 703-715, 2013

35) 丹後俊郎, 上坂浩之編: 臨床試験ハンドブック—デザインと統計解析—. 朝倉書店, 東京, 2006

36) Trivedi, M. H., Fava, M., Wisniewski, S. R., et al.: Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*, 354 (12); 1243-1252, 2006

37) Trivedi, M. H., Rush, A. J., Wisniewski, S. R., et al.: Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*, 163 (1); 28-40, 2006

38) Vieta, E., Cruz, N.: Head to head comparisons as an alternative to placebo-controlled trials. *Eur Neuropsychopharmacol*, 22 (11); 800-803, 2012

39) Walsh, B. T., Seidman, S. N., Sysko, R., et al.: Placebo response in studies of major depression: variable, substantial, and growing. *JAMA*, 287 (14); 1840-1847, 2002

What are the Specific Characteristics of Clinical Trials for Neuropsychiatric Disorders?

Tetsuo NAKABAYASHI

Office of New Drug III, Pharmaceuticals and Medical Devices Agency

To increase medical standards, not only the establishment of therapeutic strategies to accumulate evidence for existing therapies but also the development of new therapies are required. In recent years, as a number of new drugs have been approved, treatment options for neuropsychiatric disorders have expanded. It is significant for clinicians to interpret the evidence adequately in order to select the optimal therapeutic method. This paper aims to describe the important points in interpreting the results of clinical trials, which involve direct evidence for efficacy and safety.

There are several specific characteristics of clinical trials for neuropsychiatric disorders. As typical features, the following can be mentioned: 1) a high and variable response in placebo groups, 2) impact on the effect size of the baseline severity of disorders, 3) high dropout rates, and 4) biases related to subjective measures of clinical symptoms. In this paper, the key points to be evaluated in the results of clinical trials for neuropsychiatric disorders are discussed based on these features.

< Author's abstract >

< **Keywords** : neuropsychiatric disorders, clinical trials, robustness, placebo response, assessment biases >
