

Small Clinical Trials による 薬効評価の考え方

2013 年 5 月

日本製薬工業協会
医薬品評価委員会 データサイエンス部会
タスクフォース 2

目 次

1	はじめに	3
1.1	Small Clinical Trials の定義と取り巻く環境	3
1.2	本報告書の構成	3
1.3	本タスクフォースの提言	4
2	米国, 欧州及び日本の指針・ガイドライン	5
2.1	米国	5
2.2	欧州	5
2.2.1	Guideline on Clinical Trials in Small Populations	6
2.2.2	Concept paper on extrapolation of efficacy and safety in medicine development.....	7
2.3	日本	8
3	Small Clinical Trials を用いたエビデンスの主張	9
3.1	評価変数	9
3.1.1	評価変数の種類	9
3.1.1.1	真のエンドポイントの利用	9
3.1.1.2	代替エンドポイントの利用	9
3.1.2	評価変数の数	11
3.1.3	評価変数の精度	12
3.2	比較対照	12
3.2.1	ヒストリカル・コントロールを用いた外部対照試験	12
3.2.2	ベースライン対照	13
3.2.3	無対照	13
3.3	エビデンスの質と量	13
3.4	情報の示し方	15
3.4.1	統計モデルの利用	15
3.4.2	検証試験における α の調整	16
3.4.3	ベイズ法の利用	16
4	少数例でエビデンスを確立させるための基本的な考え方	18
5	まとめ	20
	参考文献.....	22

1 はじめに

1.1 Small Clinical Trials の定義と取り巻く環境

医薬品開発の標準的な方法は第 I 相試験, 第 II 相試験, 第 III 相試験から構成される. そして, 第 III 相試験では, 統計学的仮説検定によって仮説が検証される [1, 2]. この方法により仮説の検証を成功させるためには適切な規模の試験が必要となる. しかしながら, 被験者組入れのあらゆる努力を行っても, 合理的な時間枠の中では, 適切な規模での試験実施が不可能な場合がある [3].

一例として, N-アセチルグルタミン生成酵素 (NAGS) 欠乏による高アンモニア血症治療を挙げる. 同疾患治療薬は, 厚生労働省により医療上の必要性の高い未承認薬・適応外薬として開発企業の募集が行われたものであり, その罹患例は世界で約 50 例が報告されているのみである. この場合には十分な被験者を収集することは困難であり, また被験者の組み入れのためのコスト (費用, 時間) も大きい. これほどではなくとも, 被験者の組み入れが困難であったり, コストが大きかったりする場合もあり得る.

被験者の組み入れが極めて困難な状況において仮説に対する回答を与えようとする試験を, 本報告書では, Small Clinical Trial (以下, SCT; 複数の試験を指す場合は SCTs) と呼ぶ. SCT は, 対象患者数が少ない場合と, 対象患者数は少なくはないが被験者の組み入れが困難な場合がある. 前者は, 希少疾病用医薬品 (対象者の数が本邦において 5 万人未満の医薬品) とこれに近いほど対象患者が少ない場合であり, 後者は, 被験者負担の大きい測定が必要な場合や小児のように同意が得られにくい場合などである.

従来, 多くの製薬企業が開発のターゲットとしていた患者数の多い生活習慣病などの疾患領域では, 治療法が充実してきたため, 承認のためのベネフィット・リスク評価の水準が高まり, 必要な臨床試験の規模が拡大し, 新薬開発に巨額な研究開発投資が必要になっている. また, 医科学の進歩による疾病原因の解明や創薬技術の発達が進んできていることから, アンメット・メディカルニーズに対する製薬企業の取組みが盛んになってきている. これらの状況から, 患者数の少ない, もしくは, 収集することが困難な領域に対する医薬品開発への興味が高まっている. しかしながら, 先ほど示した SCTs による医薬品開発では, 薬効評価法が十分には整理されていない. そこで, 本タスクフォースは, SCTs による薬効評価法を整理することを目的として召集された.

まず, 我々は, IOM (institute of medicine) が 2001 年と 2011 年に公表した 2 つのレポート [4, 5] や NIH と FDA によるトレーニングコース [6, 7] から, SCTs 用の効果的な試験デザイン及び統計解析手法を調査したが, 革新的な方法は見られなかった. また, Kianifard (2011) [8] でも, これらの方法以外に発展的な方法は示されていなかった.

そこで, FDA で承認された希少疾病用医薬品の事例調査を中心に, どのような場合に, 既存の方法 (統計学的仮説検証を用いた薬効の証明) とは異なる方法で, 医薬品が製造販売承認されるかを調査した. その結果, 当初の我々の予想を超えて, 既存の方法が利用されていることがわかった. 即ち, 希少疾病であるということだけをもって, 承認の基準が緩められるということではなかった. また, 単一遺伝子の変異によって引き起こされるような比較的単純なメカニズムによって引き起こされている疾患の場合には, 疾患メカニズムに対する理解を積極的に利用した開発に可能性があることもわかった. 例えば, unvalidated 代替エンドポイントを利用したり, 影響を与える可能性のある変数を経時的に観察したりする方法. 本報告書では, これらの検討を元に, 新たな枠組みについて, 一つの考え方をまとめた.

1.2 本報告書の構成

第 1 章の最後に, 提言を示す. この提言の根拠として, 第 2 章では, 現状を理解するために, SCTs による薬効評価に関係する三極のガイドライン・通知について示す. 第 3 章では, SCTs を用いて有効性を示すための方法について示す. 第 4 章では, 少数例でエビデンスを確立するための基本的な考え方を示す. そして第 5 章で, SCT に関する我々の考えを総合的に述べる.

Appendix に, 我々が調査した FDA で承認された希少疾病用医薬品の事例を添付したので参照

して頂きたい。なお、本文中や appendix で紹介している事例では、理解のためインターネットから得た疾患情報なども記載しているため、自身の開発計画で参考とする場合には、根拠文書の確認をお願いしたい。

1.3 本タスクフォースの提言

本タスクフォースの提言を以下にまとめる。

提言 1. 製薬企業及び規制当局は、ベネフィット評価法を柔軟に検討すべきである。

- ・ 疾患メカニズムと開発候補化合物の作用機序を考慮し、評価変数、比較対照、エビデンスの質と量、情報の示し方の 4 点を整理することで効率の良い開発・試験計画を選択できる。
- ・ 従来 of 検証の枠組みの実施可能性は最大限、検討されるべきである。

提言 2. SCTs に関するガイドラインを作成する。

- ・ PMDA が公開している「新医薬品承認実務に関わる審査員のための留意事項」 [9] には、柔軟な審査を行うことが示されている。柔軟な開発・試験を行うために、規制当局と製薬企業は協力し、ガイドラインを整理すべきである。

提言 3. 質の高い試験を実施する。

- ・ 従来と異なる方法を用いるということで、データの品質を犠牲にしてはならない。個々のデータには常に高い品質が求められる。
- ・ 評価変数の標準誤差を小さくするには、症例数を増やす方法と標準偏差を小さくする方法がある。SCT では、被験者数を増やすことが困難であるため、評価変数の精度の向上に対する工夫を行うべきである。

2 米国、欧州及び日本の指針・ガイドライン

本章では、SCTsによる薬効評価に関係する各地域のガイドラインや通知を紹介する。米国では、SCTに特化したガイドラインは存在しないが、既存のガイドラインを改訂することによって、ニーズに対応している。一方、欧州ではSCTのためのガイドラインが整備されている。日本では、ガイドラインは存在しないが、承認審査の確認事項が公開されており、その中で審査の柔軟性について触れている部分があるので、それらについて示す。

2.1 米国

米国FDAガイダンス「Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products」(1998) [10] (以下FDAガイダンス(1998))には有効性のエビデンスに関するFDAの考え方が記載されている。1962年にFederal Food, Drug and Cosmetic Act (FDC Act) [11]が改訂され、医薬品の製造承認取得のためにそれまでの安全性に加えて有効性のエビデンスも必要となった。以来FDAは、承認のためには、適切かつよく管理された試験(adequate and well-controlled investigations)によって確立された十分な有効性エビデンスを求めるという姿勢をとっており、これは通常2本以上のPivotal Studyによる有効性の検証が必要であることと理解されている。FDAガイダンス(1998)では複数の適切かつよく管理された試験が必要な理由の一つとして、統計学的な不確実性などから、薬剤の有効性を確立するために一試験の結果では不十分であることを挙げている。

適切かつよく管理された試験とは、21 CFR part 314.126 [12]によれば、「薬剤効果の定量的な評価のために対照群との適切な比較ができるようにデザインされているものであり、試験の目的、観察期間、対照群、症例数などの試験デザインがプロトコールと報告書に記載されなければならない」である。また、同規定は「ヒストリカル・コントロールとの比較は通常、プラセボ群、無治療群、あるいは実薬群などを用いた同時対照群との比較に比べて妥当な評価ではない」としている。ただし、ヒストリカル・コントロールについては「高い死亡率であることが分かっている疾患が対象で、薬剤の効果が自明であるような特別な場合には用いられる場合もある」とも記されている。

エンドポイントについては、FDA Actやより詳細な規則・規定集であるCFR Title 21にも明確な記載はない。しかし、1992年に交付された、重篤で生命を脅かす疾患に対する薬剤の迅速な承認を目的とした21 CFR 314 Subpart H [12] (および生物製剤に対する21 CFR 601 Subpart E [13])、あるいは1997年に同じく重篤で生命を脅かす疾患に対する薬剤の開発を早め審査の迅速化を図るために設けられたFast Trackの規定 [14]では、臨床的利益を合理的に予測できるような、疫学的、治療学的、病態生理学的あるいは他のエビデンスを基にした代替エンドポイント、あるいは生存時間または不可逆的な罹患率以外の臨床エンドポイントに対する効果に基づく承認が可能であることが定められている。

1997年のFood and Drug Administration Modernization Actの改訂では、FDA長官の決定により、一つの適切かつよく管理された試験とそれを補強する証拠(confirmatory evidence)で有効性を立証するのに十分と判断することもできるとされ、21 USC Section 355(d)に追記された [11]。FDAガイダンス(1998)では一つの試験結果で有効性を立証できる可能性がある場合として、他の対象集団(年齢、人種、性、合併症など)に対する試験結果がある場合、近い関係にある疾患やそれほど近くはないが治療の一般的な目的が類似しているような疾患に対する試験結果があり、PK/PD試験などを用いて外挿できる場合、また、遺伝的要因により酵素が欠損していることにより生じる疾患のように、病態と治療の機序が良く理解されている場合、統計学的に非常に説得力のある結果(P値が非常に小さい結果)がある場合などを挙げている。

2.2 欧州

欧州ではEMAのCHMPから「Guideline on Clinical Trials in Small Populations」(2006) [15] (以下CHMPガイドライン(2006))が公開されている。また、限られた数の患者を対象としなければならない

い場合に検討すべき方法として他の対象集団での結果を外挿することが挙げられるが、これに関して「Concept paper on extrapolation of efficacy and safety in medicine development」が draft 版 [16] として 2012 年に EMA より公開されている。以下に要約を紹介する。

2.2.1 Guideline on Clinical Trials in Small Populations

CHMP ガイドライン(2006)では、「多くの希少疾病用医薬品や小児適用で、一般的に受け入れられる規則やガイダンスに従った無作為化比較試験を基に承認申請がされており、それが不可能な場合に限り原則から外れる方法を考慮すべき」と前置きしたうえで、以下のようにいくつかの観点から検討すべき考え方などをまとめている。

Pharmacological considerations

疾患の病態生理や薬剤の薬理の詳細な理解は、効率的な臨床試験設計を容易にし、必要となる臨床データの量を決定するのに役立つ。遺伝子欠損に由来する疾患に対する補充療法においては、欠損に伴う症状がよく理解され、薬剤の PK/PD が臨床試験によって明らかにされていれば、患者内の投与前後の比較は薬剤のベネフィット・リスク評価を支持するのに十分なデータとなりうる。用量反応関係が見られる、あるいはイベントの連鎖(e.g., 薬剤が作用部位に存在し、薬力学的な測定がなされ、臨床的なアウトカムがみられる)が明らかにされたような場合には試験結果の信頼性が増す。

非常にまれな疾患では、ベネフィット・リスク評価を可能にするために、試験に参加した患者からできるだけ多くの情報を得ることが重要である。

Choice of Endpoints

理想的には、確か(hard)で臨床的に適切なエンドポイントが望ましい。疾患が進展するまでの時間は中間レベルのエンドポイントで、理想的にはエンドポイントとして validate されているべきだが、患者数が少ないためにそれが困難であることは認識されている。

腎不全のような臨床エンドポイントは、臨床的に適切なエンドポイントの良い例である。症状の緩和や患者の評価(preference)は価値があるエンドポイントかもしれないが、これらは疾患の進行や死亡に対する治療効果を真に示したものではないので、疾患に対する治療に依存していることを確認することが常に必要である。

時には、最も適切な臨床エンドポイントが知られていなかったり、幅広く支持されていなかったり、あるいは validate された臨床エンドポイントが存在しない場合がある。試験薬により影響される結果を予測するには、試験薬の作用機序が十分に知られていない場合もある。そのような場合にはプライマリーエンドポイントを前もって決めておくという通常のアプローチは保守的過ぎるかもしれないので、可能なすべてのエンドポイントを集め、最終報告書ですべてのデータを提示することでより多くの知見を得ることも考えられる。希少疾患では、所与の臨床エンドポイント、あるいは validate された代替エンドポイントに対して十分な数の被験者を集めることが困難である場合や、このエンドポイントを用いて結果を明らかにするには合理的でない長さの時間がかかる場合がある。そのような場合には臨床エンドポイントの代わりに他の代替マーカーの使用が考慮される場合もある。疫学データや患者登録から得られたデータが代替マーカーの選択の助けになると考えられる。

代替マーカーによる検証は臨床的有用性や長期の benefit の最終証明ではないので、規制当局の審査や承認の根拠にしようとするならば、クリニカル・ベネフィットや安全性、ベネフィット・リスク評価のサポートとなるさらなるエビデンスをもたらす追加試験の計画の必要があるだろう。

Choice of Control Groups

偏りのない薬剤効果を推定するためには同時比較の内部対照が望ましい。例外的な状況においてはヒストリカル・コントロールが許容される場合もある。疾患の自然経過が非常によく理解されている場合にのみ非対照が受け入れられる可能性がある。患者登録のデータは疾患の自然経過に関する重要な情報源になりうるとともに、ヒストリカル・コントロールとしての可能性がある。

Methodological and Statistical Considerations

試験デザイン段階で考慮すべき事項として、層別化といったランダム化の方法、Response-adaptive methods, Sequential designs, n-of-1 trials などを紹介している。

Data analysis の観点からは、statistical modeling を仮定しての感度分析、ベースライン値による調整や予後因子による層別解析、repeated measurements に対する Generalized Estimated Equations (GEE) method, Hierarchical Linear Models, Mixed-effects models の利用や Bayesian Methods などについての説明がある。

2.2.2 Concept paper on extrapolation of efficacy and safety in medicine development

Introduction

外挿とは、既に実施された他の集団（以下、外挿元の母集団）の部分集団、あるいは関連する疾患や関連する医薬品の試験から得られる情報や結論を拡張して、他の部分集団（以下、外挿目標の母集団）、他の疾患や他の医薬品の結果を推論することである。外挿により、結論を得るために必要な追加情報を減らすことができる。

外挿を行う主な理由は、外挿目標の母集団への不必要な試験を回避できる倫理的な面と、試験を本当に必要とする領域へリソースを割り当てられる効率性の面がある。さらに、試験の実施可能性が限られる場合に、他の外挿元のデータを再検討することで、外挿目標の母集団での限られたエビデンスを合理的に解釈することができるかもしれない。

大人から子供への外挿は代表的な例であるが、それ以外にも多くの領域で外挿が適用できる場合もある。以下が例である。

1. 集団のサブセット間：年齢、成長、成熟、性別、妊娠、合併症、臓器障害、内因性・外因性の民族差
2. 疾病サブタイプ間又はステージ間、異なる疾病間や症状間
3. 薬剤クラスの内部・外部の薬剤間
4. 動物試験から人へ
5. 健康人から患者へ

Discussion

FDA は、大人から小児へ有効性データを外挿するフレームワークのアルゴリズムを次のように示している (FDA 1994)。

1. 外挿元の母集団と外挿目標の母集団で、1) 疾病進行及び 2) 介入への反応の両方が十分類似している場合、小児試験デシジョンツリー (Pediatric study decision tree) が利用可能。
2. 医薬品の曝露とその反応関係 (exposure-response) が類似していると仮定される場合、用量を決めるための PK 試験と安全性試験だけが必要。ただし、曝露とその反応関係が不明瞭な場合、追加の PK/PD 試験が必要。
3. 合理的な確実性で類似性を仮定することができない場合、外挿目標の母集団での PK 試験・PK/PD 試験及び有効性・安全性試験など、十分な試験プログラムが必要。
4. 安全性プロファイルが小児と大人で異なる可能性があるため、安全性試験は常に必要。

同様の指針は、「ICH E11 Clinical Investigation of medicinal products in the paediatric population (CPMP/ICH/2711/99)」及び「Role of Pharmacokinetics in the development of medicinal products in the Paediatric Population (CHMP/EWP/147013/2004)」の EMA ガイドラインで議論されている。

EMA 外挿ワーキンググループは、大人から小児への有効性の外挿に限らず、医薬品開発のすべての領域の外挿に関するアルゴリズム開発の可能性について議論する必要があるとし、次のフレームワークを提案する。

A. 臨床的背景

- 前向き研究をする代わりに、何故外挿をする必要があるのか正当化する必要がある。これは

科学的な妥当性のもとに議論されるべきである。

B. 外挿コンセプトの開発

- 生物学・薬学の論理的根拠として、以下の項目を外挿元の母集団と外挿目標の母集団で評価する
 - 疾病（病因，病態生理学，臨床症状，進行に基づいたサブタイプ）の類似性。
 - 薬物の体内動態と効果（PK/PD）の類似性。
 - 臨床での有効性と安全性のエンドポイントの類似性及び妥当性。
- 定量的なエビデンス
 - グループ間の疾病進行の差を特徴づけるために疾病モデルを利用。
 - PK/PD と年齢及び他の重要な共変量の関係を調べるために、既存のデータ及び生理学に基づいた PK（又は PD）モデリング及びシミュレーションを利用。
 - すべての既存のデータ（*in vitro*，前臨床，臨床）を用いてモデル化し，外挿元の母集団と外挿目標の母集団に対して，臨床的なレスポンス（有効性，いくつかの安全性状況）で類似性の程度を予測する。外挿戦略の選択。
 - 外挿目標の母集団と外挿元の母集団の間で予想される薬剤への反応の違いに基づき，外挿戦略を次のうちから選択する。
 - 外挿しない：外挿目標の母集団で完全な開発プログラムを実施する。
 - 部分的な外挿：外挿目標の母集団と外挿元の母集団の間で予測される差の大きさ及び仮定の確実性に応じて，外挿目標の母集団での開発プログラムを減らす。
 - 完全な外挿：外挿コンセプトを妥当にするデータが存在する場合。

C. 外挿計画

大きく二つのシナリオが考えられる。

- 有効性を外挿するために，外挿目標の母集団の PK 又は PK/PD 情報を利用してブリッジングする。
- 外挿元の母集団の事前情報を利用して，ベイズ流の統計アプローチを利用する。

2.3 日本

日本では SCTs のための試験デザインや解析方法などに関する指針は示されていないが，医薬品医療機器総合機構（PMDA）の審査員向けの文書である「新医薬品承認審査実務に関わる審査員のための留意事項」[\[9\]](#) が PMDA のホームページで公開されている。その中には承認の可否について判断する際に留意する事項として，「有効性に関し，プラセボ又は多用量等に対する優越性が検証されているか」，あるいは，「有効性に関し，標準薬に対する非劣性/優越性が検証されているか」，また，「結果の信頼性を担保するため，原則として「2 本以上の無作為化比較試験」において，有効性が検証されていることが望ましい」といった原則の記載がある。一方で，「特に，希少疾病医薬品又は既存治療が確立していない重篤な疾患領域の医薬品については，その臨床的意義等を踏まえ，個別に判断すべき事項が多いと考えられることから，必ずしも本文書における各事項の適用のみによって最終判断を行うものではないが，本文書の趣旨を十分に理解し，適切なデータのもとに科学的な評価を行う必要がある。」という記載もあり，SCTs しか実施できない状況においては，考慮すべき事項があることも述べられている。

3 Small Clinical Trials を用いたエビデンスの主張

SCTs は、統計学的仮説検定によって仮説を検証するために十分な被験者を合理的な時間枠の中では集めることができないことを背景としている。この状況において有効性の存在を主張し、認めるためのデータパッケージを構築するための考え方を、評価変数、比較対照、エビデンスの質と量、情報の示し方という4つの観点から、米国で承認された希少疾病用医薬品の事例とともに示す。

3.1 評価変数

3.1.1 評価変数の種類

評価変数として、真のエンドポイントと代替エンドポイントを利用することができる。ただし、SCT では被験者の組み入れが困難であるため、真のエンドポイントによってベネフィットを直接評価するか、代替エンドポイントを用いて、直接的な評価ではないが、情報量を増やすかの選択を検討する必要がある。

3.1.1.1 真のエンドポイントの利用

非常に重篤な疾患では、死亡や疾患に起因するイベントのハザードが大きいいため、これらの発現を短期かつ小規模の試験で評価することが可能な場合がある。被験者の組み入れが困難であるというSCTの状況では、対象の選択とエンドポイントの選択とのバランスを検討する必要がある。

- 真のエンドポイントが用いられた事例：アルグルコシダーゼアルファ

アルグルコシダーゼアルファ（商品名：Myozyme）は、ポンペ病治療薬として開発された。ポンペ病は酸性 α グルコシダーゼ（GAA）の欠乏によって細胞内にグリコーゲンが過剰蓄積し、重度の筋機能障害を引き起こす。ポンペ病の有病率は4万人に1人であり、幼児発症においては生後1年以内に殆どの患者が死に至る。アルグルコシダーゼアルファは遺伝子組換え技術で作られたヒト型GAAであり、蓄積されたグリコーゲンを分解・代謝する。

この事例では、真のエンドポイントによる評価が行われた。臨床試験は人工呼吸器離脱状態での生存率を主要評価項目として、幼児発症例18例を対象として実施された。外部対照（ヒストリカル・コントロール）62例の全生存率と比較が行われ、有意な生存率の上昇が認められた。なお、ポンペ病は成人発症においては病態が異なり、幼児発症と較べて筋機能障害の進行が遅く、生存期間も10年単位で延長される。アルグルコシダーゼアルファでは、6分間歩行試験・努力肺活量を主要評価として成人発症を対象とした試験も行われた [17] が、こちらについては十分な有効性を示す事が出来ず、「成人型糖原病2型患者に対する本剤の有効性及び安全性は確立していない」とされている。

3.1.1.2 代替エンドポイントの利用

代替エンドポイントについて、ICH E9 [2]では証拠の強さとして次の3点が挙げられている。（注：原文では代替変数と表現されている。）

- (i) 代替エンドポイントと臨床的結果の関連の生物学的合理性
- (ii) 代替エンドポイントが臨床的結果の予後を予測する上で有益であると疫学研究によって示されていること
- (iii) 試験治療の代替エンドポイントに対する効果が臨床的效果と対応しているという臨床試験の結果

(iii) が成立している代替エンドポイントは、効率の観点から積極的に利用すべきである。この代替エンドポイントのことを validated 代替エンドポイントと呼ぶ。一方で、(iii) の成立が確認されていなくても、(i)、(ii) から (iii) の成立が強く示唆される場合には、(i)、(ii) のみから定義されるエンドポイントを利用できると考えられる。この場合のエンドポイント

を **unvalidated** 代替エンドポイントと呼ぶ。

SCT の対象となる疾患には、疾患メカニズムが良く解明されているものがある。例えばある種の生天性代謝異常症である。米国での承認事例では、a) 特定酵素の欠乏、b) 疾患原因物質の過剰蓄積、c) バイオマーカーの変動、d) 疾患イベントの発現、e) 合併症・重篤なイベントの発現、といった疾患経過が病理学的に解明されていた。また、開発化合物の薬理作用（欠乏酵素の補酵素や疾患原因物質の除去）も明確化されていた。このような病態と治療の機序の明確化は、(i)、(ii) から (iii) の成立を強力に後押しするものである。

実際に米国で承認された希少疾病用医薬品では、**validated** 代替エンドポイントにより承認された事例のほか、**unvalidated** 代替エンドポイントにより承認された事例があり、以下に両事例を示す。

- **validated** 代替エンドポイントが用いられた事例：カルグルミン酸

カルグルミン酸（商品名：Carbaglu）は、N-アセチルグルタミン生成酵素（NAGS）欠乏による高アンモニア血症治療薬として開発された。カルバミル・リン酸合成酵素は尿素サイクルにおける本質的な酵素であり、NAGS はカルバミル・リン酸合成酵素を活性化させる補助因子である。NAGS の欠乏によって尿素サイクルが機能不全に陥るとアンモニアが代謝されず、高アンモニア血症を発症する。症状の進行によっては死に至る、重篤な疾患である。

この事例では **validated** 代替エンドポイントによる評価が行われた。カルグルミン酸は NAGS の生成物である N-acetylglutamate に類似した構造（末端のアンモニア基をメチル基に置き換えただけの違い）であることより、N-acetylglutamate と同様の役割を果たすと考えられた。血漿中アンモニア濃度を主要評価項目として、後ろ向きの症例集積研究 13 例で評価された。血漿中アンモニア濃度はカルグルミン酸投与後に正常域まで改善し、無治療の NAGS 欠乏患者における自然経過としては考えられないとして、臨床的に有意である、と FDA 審査報告書にある。

- **unvalidated** 代替エンドポイントが用いられた事例：アガルシダーゼベータ

アガルシダーゼベータ（商品名：Fabrazyme；ファブラザイム点滴）はファブリー病治療薬として開発された。ファブリー病は α ガラクトシダーゼ (α -Gal A) の酵素活性の低下によって細胞内にグロボトリアオシルセラミド (GL-3) が過剰蓄積する。ファブリー病の症状は多岐に渡り、症状の発現には性差がある。典型的ファブリー病の男性では主要臓器（腎、心血管、脳血管）の機能低下や機能不全を引き起こし、死に至る。ファブリー病はその病型によって、古典型ファブリー病と亜型ファブリー病に分類される。古典型ファブリー病では初期に四肢末端の疼痛、知覚異常を発症し、年齢に応じて角膜混濁、低汗症及び被角血管腫を認める。さらに心、腎不全及び脳血管障害を起こし、これらが死因となる。一方、亜型ファブリー病は知覚異常や被角血管腫等の症状は示さず、腎不全だけを発現する腎臓型、軽症で発症年齢が遅く主に心障害だけを示す心臓型がある。

この事例では **unvalidated** 代替エンドポイントによる評価が行われた。アガルシダーゼベータは遺伝子組換え技術で作られたヒト型 α -Gal A であり、蓄積されたグロボトリアオシルセラミド (GL-3) を代謝する。臨床試験は腎臓毛細血管内皮の GL-3 蓄積に関する形態学的評価を主要評価項目として、ファブリー病患者 58 名を対象として、プラセボ対照二重盲検比較試験が実施された。この結果、プラセボに対する有意な改善が認められた。また、副次評価で行われた皮膚毛細血管内皮細胞及び心組織についても同様に有意な改善が認められた。これらの試験結果を元に米国では「Fabrazyme（アガルシダーゼベータ）はファブリー病患者への使用に適応する。Fabrazyme は、腎臓の毛細血管内皮及び他の細胞における GL-3 蓄積を低減させる」とされ、病型に関わらず承認されている。一方、本邦においては、効能効果はファブリー病とされているものの使用上の注意として「心臓にのみ病変が認められる亜型のいわゆる心ファブリー病患者での安全性及び有効性は確立していない」とされた [17]。

3.1.2 評価変数の数

通常の試験では、一つの評価変数の最終時点だけを用いた推測が行われることが多い。これは被験者あたりの情報量が少ないが、ほぼ十分に薬効が発現しているであろうということ以外に時間-応答関係に対する仮定が無く、広く受け入れられている。しかし、SCTsによって開発を行う（つまり被験者数が少ない）場合には、疾患メカニズムと薬理作用を深く理解し、複数の変数による推測や、経時的な変化を測定するなどして、被験者あたりの情報量を増やすことを検討すべきである。

経時測定データの場合には時間を含む統計的モデルを利用することも可能であろうし、多次元データの場合には、例えば、MANOVA、Hotelling T²、O'Brien 法など多変量解析の方法を利用して1次元に縮約する方法も利用可能である。

しかし、重みを考慮しない縮約は、妥当性に疑問が持たれることがあり、また、特定の変数の良さが条件次第で変化する場合には、単純には情報の縮約ができない場合がある。例えば、肺高血圧症では、一般的には肺動脈圧の低下が病態の改善を意味するが、心拍出量が低下している場合には肺動脈圧の低下は病態の悪化を意味することもある。このような場合には、新たに評価尺度を開発することが考えられる。

評価尺度を用いて評価を行う場合には、その評価尺度の妥当性及び信頼性の根拠を明らかにする必要があり [2]、そのための臨床試験が実施されることもある。しかし、SCTsによって開発を試みる場合には、そのための試験を計画することや、その評価に必要な被験者数を確保することも困難であると考えられる。

この場合には、専門家による判定委員会の結果を評価変数とすることが考えられる。即ち、評価指標は経時的多変量測定データであり、評価項目が委員会判定である。規制当局による承認審査の段階では、測定されたデータだけでなく委員会判定の結論と根拠が議論の対象となる。

この他にも、開発化合物が影響しうる全ての変数を測定し、それらの結果をまとめることによって知見を得る方法も欧州ガイドラインに示されている。

以下に、米国で承認された希少疾病用医薬品の事例を以下に示す。

- **総合評価のエンドポイントが用いられた事例：遺伝子組換え血液凝固第 IX 因子**

遺伝子組換え血液凝固第 IX 因子（商品名：ベネフィクス）は、血友病 B 患者における出血抑制と予防（手術中の出血抑制と予防を含む）を対象として開発された。血友病は主に遺伝性の血液凝固異常症であり、症状としては関節内や筋肉内で内出血が起こる。有効性の評価は、被験薬を投与する度に、出血予防について 5 段階の主観的な評価（Excellent, Good, Moderate, No Response, Failure）によって実施された。評価は、医師又は被験者本人によって行われた。

この事例については、単に 5 段階の主観的な評価のデータを収集するだけでなく、出血に関する詳細な情報（例えば、出血の部位、内出血の範囲、持続時間）を収集することにより、被験者一人あたりの情報を充実させることが可能であろう。

- **複数の評価指標を同時に評価した事例：オプレルベキン**

オプレルベキン(商品名：Neumega®)は、化学療法による重度の血小板減少症の予防を対象として開発された。骨髄抑制作用のある抗悪性腫瘍薬による化学療法では、重度の血小板減少症が発現することがある。一般的には抗悪性腫瘍薬の減量及び血小板輸血で対応するが、抗悪性腫瘍薬の減量は原疾患の治療に対して効果を減弱させ、血小板輸血は感染症及びアレルギーのリスクがある。主要な試験において、有効性は血小板輸血の回数により評価された。

この事例については FDA の審査報告書において、化学療法の治療レジメンにより骨髄抑制作用の大きさが異なる点、輸血を実施する際に好中球数の減少がトリガーとして考えられる点が指摘されている。また、好中球数の回復期間や、血小板の回復期間について審査官が考察している。このことより、血小板輸血の回数のみならず、治療レジメン、血小板数の推

移、トリガーとなる好中球数の推移、及び回復期間を測定し評価することによって、被験者一人あたりの情報量を増やすことが可能であったと考えられる。

3.1.3 評価変数の精度

十分な人数の被験者の組み入れが困難であることが、評価変数の測定の精度を下げた良い理由にはならない。SCT では、通常の開発方法に比較し、データパッケージにおける 1 例に重みがある。このため、なるべく多くの情報を個別の症例から得ようとし、症例毎に測定の方法が変わってしまうかもしれない。これはデータパッケージを総合的に判断することを困難にするかもしれない。

また、通常の開発方法では、被験者数を増やすことによって、標準誤差を小さくすることができるが、SCT では、被験者数を増やすことに限界があるので、それにより十分に標準誤差を小さくすることができないことが多くある。このため、通常の開発方法よりも、個々のデータの精度を上げる工夫を行うべきである。例えば、通常の開発では、コスト面から採用が躊躇されるような変数であっても、精度良い測定が期待できるなら、積極的な採用を検討すべきである。

CHMP ガイドライン (2006) [15] では、「非常に稀な疾患では、複数の個別症例研究の統合結果だけが、エビデンスを示す唯一の方法になるかもしれない。そのような状況では、治療状況やデータ収集は標準化されるべきであり、データは GCP 基準を満たし高品質である必要がある。」とある。また FDA-NIH のトレーニングコース(2012) [7] の中で、FDA 統計 Director の LaVange は、「品質管理は試験の結果の妥当性を確保するための重大な要素である。高い精確さが不可欠である少数例の試験ではより重要になる。試験実施にあたっては、データの質、トレーニングそしてモニタリングを適切に実施することが重要である。」と述べている。

3.2 比較対照

通常の試験では同時対照を設定する。それは、治療効果と疾患の自然経過を分離することが目的である。検出しようとする効果が疾患の自然経過の変動を大きく上回る場合には、同時対照を設定しなくとも有効性の存在を主張できることがあり得る。同時対照を置かないものとして、ヒストリカル・コントロールを用いた外部対照試験、ベースライン対照試験あるいは無対照試験の利用が考えられる。

しかしながら、少数例の試験では適切にランダム化や盲検化がされていないためバイアスを回避できず、少数例の試験の方が大きい治療効果が得られる傾向があるというメタ・アナリシスの報告がある [18]。このため、疾患の自然経過に対して、大きい治療効果が期待出来る場合であっても、可能な限り同時対照の設定を検討すべきである。

3.2.1 ヒストリカル・コントロールを用いた外部対照試験

ヒストリカル・コントロールを用いた外部対照試験では、同時対照を置かずに試験を実施する。過去の試験結果から対照群の効果を推定し、それを対照群のデータとする。信頼できる十分な量の過去のデータが利用できる場合に限られ、また同時対照ではないため、留意すべき事項がある [7]。

米国で承認された希少疾病用医薬品で、外部対照試験が用いられた事例を以下に挙げる。

- **事例：アルグルコシダーゼアルファ**

アルグルコシダーゼアルファ（商品名：マイオザイム）は、ポンペ病（II 型糖原病（先天性代謝異常症））を対象として開発された。ポンペ病の有病率は 4 万人に 1 人であり、重篤な疾患である。臨床試験は幼児発症例 18 例を対象として実施された。試験デザインは、非盲験、外部対照試験（ヒストリカル・コントロール）であった。被験薬群 18 例の人工呼吸器離脱状態での生存期間と、外部対照群 62 例の全生存期間と比較した。

- **事例：遺伝子組み換えヒトアンチトロンビン**

遺伝子組み換えヒトアンチトロンビンは、先天性アンチトロンビン欠損症（HD）患者の血栓塞栓予防を対象として開発された。当該疾患の米国内の患者数は約6万人である。有効性データは、第II相試験（評価対象例13例）と第III相試験（同18例）の統合解析で評価された。2試験共に非盲験単群試験である。疾患の希少性から、FDAはヒストリカル・コントロールを対照とした非劣性、広いマージン及び統合解析による評価を認めた。対照には、血漿由来ヒトアンチトロンビンで治療された血栓症誘発要因を持つHD患者（35例）が選ばれた。有効性の主要評価項目は血栓症イベント発現率であり、非劣性マージンは20%とされた。各群のイベント発現率を10%とした時、目標症例数は各群32例となった。主要評価である血栓症イベント発現率は被験薬群3.2%（1/31例）、対照群0%（0/35例）であり、被験薬群のイベント発現率の両側差の95%信頼区間上限は16.7%となることから、非劣性が検証された。

3.2.2 ベースライン対照

ベースライン対照試験は、被験者の治療中の状態を治療前の状態と比較する試験デザインであり、疾患の自然経過のばらつきに対して、薬剤の効果が大きい場合に利用できる。米国で承認された希少疾病用医薬品で、ベースライン対照試験が用いられた事例を以下に挙げる。

- **事例：アナグレライド**

アナグレライドは、本態性血小板血症を対象として開発された。本態性血小板血症の年間発症率は、10万人あたり2～3人である。2本の試験がPivotal試験として実施された。試験デザインはいずれも、非盲験、ベースライン対照試験であった（n=44及び498）。主要評価項目は血小板数のベースライン時点からの変化であり、主要な解析時点である4週間においてベースラインからの改善が認められた。

3.2.3 無対照

無対照試験は、比較される群やデータが存在しない試験デザインである。ヒストリカル・コントロールに用いるべき信頼できる情報が利用できず、疾患の状態を定義できない場合に利用される。疾患の状態を定義できないとは、変化については評価できるが、ある時点の重症度について定義することが困難なことを意味する。

- **事例：遺伝子組換え血液凝固第IX因子**

効性の評価は、被験薬を投与する度に、出血予防について5段階の主観的な評価（Excellent, Good, Moderate, No Response, Failure）によって実施された。臨床試験においては、対照群をおかず、投与前後の症状の比較も行われていない。出血予防については、36例に対して治療薬が計488回投与され、その内437回（90%）については、Excellent又はGoodと評価された。ただし本薬剤は、既存の血液由来製剤よりも感染症のリスクが低いという点で優れていることにより承認された点に留意が必要である。

3.3 エビデンスの質と量

医薬品のベネフィット・リスクバランスの評価は、その医薬品が世に存在し続けている限り行われ続けるものである。製造販売承認は、世の中に新規化合物を提供するに足りるベネフィット・リスクバランスが成立しているかどうかによって決定される。ベネフィット・リスクの評価には、新規化合物が製造販売後に使用される対象集団の大きさと特性も加味されるべきである。

1本の無作為化臨床試験の結果の利用、他の適応症に対する臨床試験結果の活用、あるいは臨床研究の活用など、個々の疾患の病理や医薬品の作用機序に応じ、データパッケージを柔軟に検討することが重要である。2.3にも示したが、PMDAが公開している「新医薬品承認審査実務に関わる審査員のための留意事項」（平成20年4月17日）でも、審査において柔軟に対応するということが示されている。

以下に、米国での承認事例を挙げる。

- **一本の無作為化試験で承認された事例：アガルシダーゼベータ**

アガルシダーゼベータ（商品名：ファブラザイム）は、ファブリー病を対象として開発された。ファブリー病の有病率は4万人に1人であり、重篤な疾患である。臨床試験は58例を対象として実施された。試験デザインは1群29例の、ランダム化、プラセボ対照、二重盲検比較試験であった。主要評価項目は、腎臓毛細血管内皮中のグロボトリアオシルセラミド（GL-3）の量を「ゼロ」又は「ゼロ以上」の2値に区分したカテゴリーであり、カイ2乗検定により有意差が得られた（ $P < 0.001$ ）。
- **一本の無作為化試験で承認された事例：ガルスルファーゼ**

ガルスルファーゼ（商品名：ナグラザイム）は、ムコ多糖症VI型における酵素補充療法薬として開発された。日本における患者数は5名である（衛藤班による平成13年の全国調査）。臨床試験は39例を対象として実施された。試験デザインはランダム化、プラセボ対照、二重盲検比較試験であった。主要評価項目は12分間歩行であり、ベースラインを共変量とした共分散分析により有意差が得られた（ $P = 0.025$ ）。
- **他の適応症での臨床試験を活用して承認された事例：シナカルセト**

シナカルセト（商品名：レグパラ）は、甲状腺がん患者における高カルシウム血症を対象として開発された。甲状腺がん患者における高カルシウム血症は極めて稀な疾患である。試験デザインは10例の非盲検ベースライン対照試験であった。主要評価項目は、血漿カルシウム量であった。10例中7例については改善が認められた。

しかし、同時に二次性の副甲状腺機能亢進症を有する慢性腎障害患者に対する臨床開発も実施されており、二次性の副甲状腺機能亢進症については3本のRCT（無作為化、二重盲検、プラセボ対照比較試験）が実施され、1,000例以上のデータが存在する。従って副甲状腺機能亢進に伴う高カルシウム血症に対するシナカルセトの血清カルシウム低減効果については、十分なエビデンスが存在する。
- **新たに臨床試験を実施せず、過去の臨床研究の結果を利用して承認された事例：コルヒチン**

コルヒチンは、家族性地中海熱を対象として開発された。患者数は世界で約10万人である。コルヒチンは痛風薬として古くから使用されてきており、アメリカでも未承認薬として数十年使用されてきた。家族性地中海熱に対するコルヒチンは、過去の文献情報のみで承認された。なお、別の適応症である痛風については臨床試験が実施され、また健康成人を対象とした薬物動態試験が6試験実施された。申請者は包括的なデータベース検索（Biosis Previews, EMBASE, JICST-Eplus, MEDLINE）を通じて74試験の論文を参考文献として提出した。その中で、FDAは有効性をサポートする試験として19試験を特定し、その中でプラセボ対照、クロスオーバー、二重盲検ランダム化比較試験である3試験（全48例）が評価資料に採用された。評価資料の3試験については、いずれも有意差が認められた。残り16試験はオープンラベル試験であり参考資料として利用された。
- **新たに臨床試験を実施せず、後ろ向きの症例集積研究の結果で承認された事例：カルグルミン酸**

カルグルミン酸は、NAGS欠乏による高アンモニア血症を対象として開発された。患者数は世界で約50例が報告されているのみである。申請においては、23名の患者からなる後ろ向きの症例集積研究が提出された。有効性は血漿中アンモニア濃度により評価された。23名中13名の血漿中アンモニア濃度について、カルグルミン酸の投与前後の比較が可能であり、投与3日目までにデータが存在する全ての被験者は正常なアンモニア濃度に達した。

3.4 情報の示し方

3.4.1 統計モデルの利用

3.1 において、SCT の場合には、臨床試験に組み入れ可能な被験者数が少ないことから、個々の被験者からの情報を多く収集すべきであると述べた。個々の被験者から同時に得られた複数の情報は統計モデルを利用することによって、効率良く示すことができる場合がある。

以下に、統計モデルとして、経時測定型線形混合効果モデル (MMRM) を利用した例を示す。図 1 に、被験者ごとの時間-応答関係のデータの仮想データを示す。時点 0 はベースライン、時点 1 から 5 が投与後の測定時点であり、時点 5 が最終評価時点である。

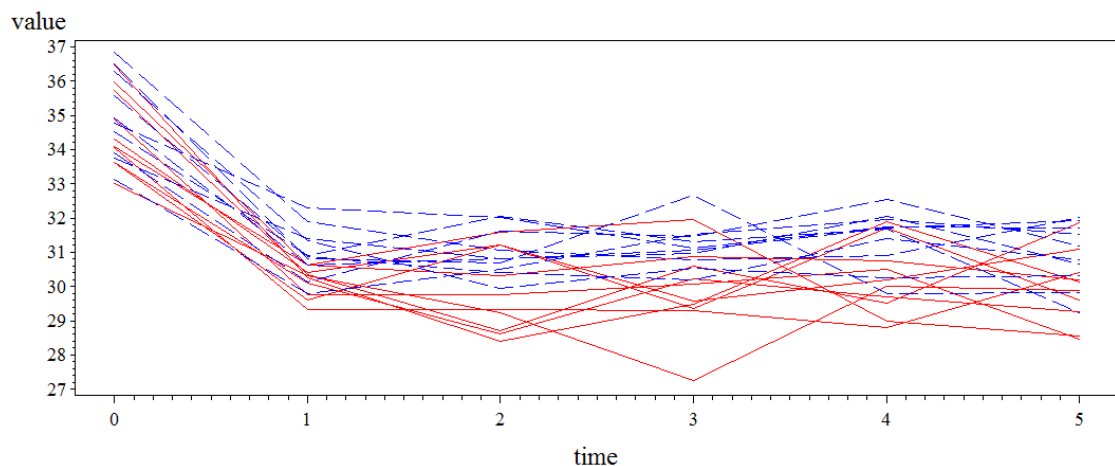


図 1 被験者ごとの時間-応答関係データ (赤：治療 A 群，青：治療 B 群)

このデータに対して、最終時点だけを用いた解析結果を表 1 に、時点 1 以降の 5 時点を用い、時点によらず薬効差とばらつきが一定という仮定を置いた MMRM での解析結果を表 2 に示した。

表 1 最終時点における評価

群	n	平均(S.D.)	群間差(治療 A 群-B 群)		
			点推定値	95%信頼区間	P 値
A	10	30.0(1.05)	-0.96	-1.90 ~ -0.03	0.0439
B	10	30.9(0.93)			

表 2 5 時点を用いた統計モデル*による評価

群	N	平均	群間差(治療 A 群-B 群)		
			点推定値	95%信頼区間	P 値
A	10	30.0	-1.1	-1.48 ~ -0.71	<0.001
B	10	31.1			

*Repeated ANOVA モデル

最終時点における評価に比べて、5 時点を用いた統計モデルによる評価の方が、信頼区間の幅が狭いことから、推定精度が良いことがわかる。この例の場合では、1 群 10 例の 5 時点分のデータを用いることは、1 群 16.7 例の最終時点のデータを用いることと、同程度の精度があると言える。つまり、測定時点を増やすことで、被験者数を 1.67 倍したことと同程度の精度が得られたことになる。

5時点のデータを用いた統計モデルによる解析では、「時点によらず薬効差とばらつきが一定」という仮定を置いている。もし、この仮定が正しく無い場合には、統計モデルによる解析結果の妥当性が疑わしくなる

ここで重要なのは、仮定の確からしさと、少ない情報から結論を導くことの不確実性のバランスを考慮し、判断のための推測の効率を向上させることを考慮すべきであるということである。ここでは時間に対して疾患の状態も薬効も一定であるという単純な仮定をおいた例を示したが、それぞれが時間に応じて状態が変わるような統計モデルを利用することも考えられる。どの程度まで仮定の確からしさを許容するかに対しては、利用する統計モデルの複雑さなどに依存し、個別の議論が必要になるが、今後、統計モデルの利用の積極的な検討が望まれる。

3.4.2 検証試験における α の調整

SCTsが想定される状況で、通常の検証の枠組みで試験を実施するならば、十分な検出力を確保することができない。実際に米国で承認された事例でも、不十分な検出力の下で検定を行い、統計的有意差が得られていないにも関わらず承認されたものがある。事後的に基準を緩める場合には、データに依存してエビデンスの強さを決定することとなり、客観性が著しく失われる。

検出力が不十分である状況に対する一つの対処方法が、適切な α を設定することである。一般に、有意水準 α は5%に、検出力 $1-\beta$ は80%から90%に設定されるが、これらは慣習的に選択されてきた値であり、絶対的な数値ではない。Bakerらは、 α 及び β の設定は両者のバランスから決定されるべきであると述べている [19]。また、Mudgeらは、新薬と比較対照との薬効差のエフェクトサイズと症例数を固定した際に、 α と β が非線形の関係性を示すことを明らかにした [20]。その上で、 α と β の相対的な重要性に基づいて重み付けを行った平均エラーを最小とすることで、最適な α 及び β の設定が可能である、としている。これらの論文は2012年に発表されたこともあり、臨床試験に適用された事例はない。しかし、不十分な検出力の下で機械的に有意水準5%の評価を行うよりも、 α 及び β の関連性を示した上で、調整された α に基づく評価を行うことは有益であると考えられる。このような判断は試験実施後に行われるべきものではなく、試験計画段階で検討され、当局との合意を計るべきものである。以下の米国での承認事例は、事前に α について議論されたものではなく、事後的な解析によって認められたものである。このような解析では用いられた検定の有意水準が分からなくなる。事前に議論されることによって適切なサンプルサイズの議論も可能となる。

● 主要な解析で非劣性を検証できなくても承認された事例：デフェラシロクス

デフェラシロクス（商品名：エクジェイド）の効能・効果は2歳以上の患者における輸血に伴う慢性鉄過剰症である。既存薬は注射剤一剤のみで、本剤は錠剤である。Subpart H, Fast trackにより2005年に米国で承認された。本剤の承認審査でpivotalとなった試験は総症例591例の第Ⅲ相試験である。本試験は実薬を対照とした非劣性試験であり、主要評価項目は肝臓鉄濃度の値から判定する有効率であったが、主要な解析で非劣性を検証する事は出来なかった。

本剤が承認に至ったポイントの一つは、事後的に行ったより少数例の部分集団での解析結果が、主要評価項目、副次評価項目ともに良好であった事である。本試験は無作為化試験であるがオープンラベル試験であり、実際には前治療を見て割り付け後の治療（投与量）が調整されていた。こうした調整の影響を受けにくかった部分集団で非劣性を示す事ができた。このような事後的な解析では有意水準の議論は困難である。厳密に統計的仮説検定を適用すべきであるか否かについては別に議論があるが、用いる場合には有意水準に対して事前に検討しておくことで適切な試験デザインを検討することが可能となる。

3.4.3 ベイズ法の利用

統計学的仮説検定は、帰無仮説（通常、効果が無い）の下での検定統計量の分布を利用して、仮説に対して結論を下す。即ち、真には差が無い場合に、当該試験から得られたデータに関する

る確率を考えており、差があるとした場合の差の大きさに関する情報は与えない。P 値が有意水準を下回った場合には、定性的に効果があるとして議論を進める。有意水準を緩めるということは、効果があると判断する水準を緩めることである。しかし、ベネフィット・リスクを総合的に判断するためには、効果があると定性的に判断するのではなく、効果がありそうな蓋然性を評価する方が相応しい。

ベイズ法を利用すれば、母数に対する分布を持った推測が可能となるため、効果が存在しない確率や、ある一定以上効果が存在する確率を得ることができる。SCTs では、これらの確率を利用したベネフィット・リスクの評価方法を検討する価値があると考えられる。

なお、米国でベイズ法により有効性を示し承認された事例は見つからなかった。

4 少数例でエビデンスを確立させるための基本的な考え方

SCTs は、統計学的仮説検定によって仮説を検証するために十分な被験者を合理的な時間枠の中では集めることができないため、個別の試験結果から高いエビデンスが期待できない。ここでは、少数例でエビデンスを確立するための基本的な考え方を示す。

エビデンスを高めるためのアプローチは、「同じ方向を指し示す手がかりが多いほどエビデンスは高まっていく」という考え方を基礎とする[21]。日米欧で示されている関連する指針の根底にある考え方とも共通していると考えられる。手がかりとは、データのパターンや、複数のエンドポイントの関連性という形で表現される所見 (findings) であり、1つの試験全体のデータに基づく所見もあれば、一部の部分集団で観察される所見、個別の症例で観察される所見まで、さまざまなレベルの所見があり得る。「同じ方向を指し示す」とは、複数の手がかりの再現性、一致性、整合性、合理的な説明が可能であるかなどを意味する。

最強の手がかりを与えるのは、統計的な検証の枠組みで示されたポジティブな結果である。2つの検証試験を求める考え方は、最強の手がかりの再現性まで確認し、エビデンスを最大限高めることを目指したものである (もっとも高いエビデンスを与えるのはもっと多くの検証試験に対するメタ・アナリシスによって得られたエビデンスであるが、製造販売承認の時点でこれが行える状況は限られている)。これは理想的であるが、実現できない場合の次善の方法が、1つの検証試験の結果と、これと同じ方向を指し示す外部情報で補う方法である。SCTs に近い状況であっても、検証試験を追求するアプローチは、このレベルのエビデンスを目指したものであると言える。

本報告が焦点を当てた検証試験が行えない状況下では、統計的な有意差を追求するのではなく、効果の大きさとその不確実性を推定することが目標となり、同じ方向を指し示す他の試験の情報、すでに得られている知見をモデル化し効果の大きさを推定した結果などにより、エビデンスを補強することとなる。さらに不確実性があまりにも高い (推定が十分な情報を与えない) 状況下では、一例一例の症例の体内で起きている現象をストーリーとして捉えるアプローチが重要になる。有効性、安全性を含め時間反応関係を念頭にデータのパターンや関連性をストーリーとして構築する。複数の評価項目を隔離された形で個別に検討する場合には1症例の情報は、散布図中の1つのプロット、分割表の1つのカウントの意味しかないが、ストーリーとして捉えられた1症例の情報のエビデンスレベルは高くなる。極少数例でしか実施できない状況下では、これら一例一例の情報を積み上げていくことにより、エビデンスを少しずつ高めていくしかない。考え方は n-of-1 試験に近づいていく (図2参照)。

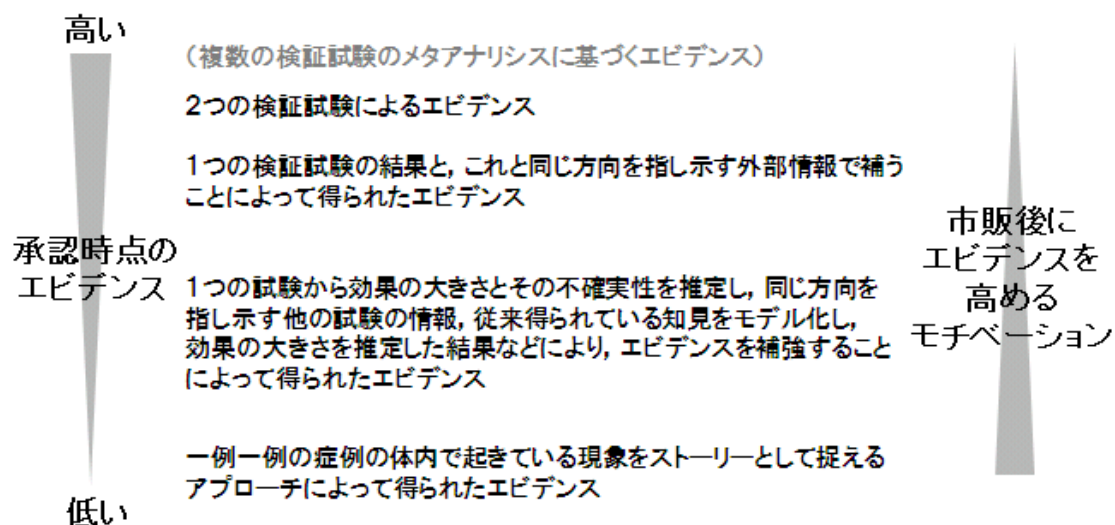


図 2 製造販売承認時点のエビデンスと市販後にエビデンスを高めるモチベーションのトレードオフ

審査基準，承認基準に柔軟性を導入することは，それぞれの状況によって，どのようなエビデンスレベルをもってよしとするか，市場に出してよいという「仮免許」を与えるかを，変化させることを意味する．1 つ以上の検証試験によって検証された有効性が確立され，製造販売承認が得られた場合には，市販後に有効性を確認するモチベーションは高くない．一方，有効性のより低いエビデンスをもって製造販売承認が与えられた場合には，市販後にも有効性を確立するための学習過程は続くものと考えられるべきであろう．医薬品リスク管理計画指針（薬食安発 0411 第 1 号，薬食審査発 0411 第 2 号，平成 24 年 4 月 11 日）に記されている「有効性に関する調査・試験」が重要な意味を持つてくる．従来の使用成績調査等の枠組みに囚われることなく，前向きコホート研究や，利用可能な場合には疾患レジストリなどを用いたデータベース研究でエビデンスを創出し，必要に応じて検証的な臨床試験を行うような柔軟な対応が期待される．つまり，製造販売承認の時点で有効性の十分なエビデンスが得られていない場合，製造販売承認をもって有効性の検討に終止符が打たれるとは考えるべきでなく，開発から市販後までを学習過程とみなす考え方が，審査基準，承認基準に柔軟性を導入することを可能にする．

5 まとめ

医薬品のベネフィットの評価方法は永きに亘り開発されてきた。そのための 20 世紀最大の発明はランダム化である。二重盲検法と組み合わせることによって、少ない仮定の下であらゆるバイアスを避けることができる。そして、事前設定した仮説に対して、十分な検出力を持ち、第一種の過誤を制御した統計学的仮説検定により仮説を検証することによって、ベネフィットのエビデンスとする方法が確立された。

21 世紀になり、生活習慣病のような疾患のニーズに対する科学的エビデンスをもつ医薬品が格段に充実してきたため、製薬企業は、アンメット・メディカルニーズと呼ばれるまだ医薬品によるベネフィットが不十分な疾患領域の医薬品開発に着手するようになってきた。

アンメット・メディカルニーズの中には希少疾病に相当するものがある。希少疾病であっても、従来の方法によってベネフィット評価が可能なものも多数ある一方で、今まで培って来た評価法を用いて医薬品を開発するには、コスト、例えば費用と時間が掛かりすぎる場合も存在する。この様な領域において、製薬企業が医薬品を開発する、①公の資金を投入して費用負担を軽くする方法、②税制優遇や特許期間の延長による方法、③コストそのものを軽減する方法などが考えられる。①、②については、既にいくつかの制度が存在するが、今もって海外で承認されている多くの希少疾病治療薬が本邦で開発されていない現状を鑑みると、①、②をさらに拡充するか、③について検討しない限り、この領域で新薬を待ちわびている患者さんを助けることができない。

本タスクフォースは、現実的な時間の枠内では被験者の組み入れが極めて困難な状況におけるベネフィットの評価方法について検討した。従来の方法ではコストが掛かりすぎる原因は、ベネフィットの評価に柔軟性を持たせる基準が曖昧な点である。即ち、社会が受け入れるリスクの大きさに応じて、ベネフィット評価の方法を変えるべきところを、審査の段階まで、どのような評価方法が適切であるかがはっきりとせず、保守的に伝統的な方法が要求されると考えるために、SCT を利用しなければならない領域では、コストが高すぎると判断されるという問題を引き起こしている。

そこで、新しい枠組みを検討するために、評価変数、比較対照、エビデンスの量、情報の示し方について調査し整理した。その結果、疾患メカニズムが良く解明されており、開発化合物の薬理作用がはっきりしている場合には、これらの情報を利用して新たなベネフィット評価法が利用できる可能性を示すことができた。また、医薬品のベネフィット・リスクバランスの評価は、その医薬品が世に存在し続けている限り継続して行われるものであるため、製造販売後のベネフィット・リスク評価で、さらに有益な情報を得ることも忘れてはならない。

最後に、SCT における新たなベネフィット評価法を実現可能とするための本タスクフォースの提言を示し、まとめとしたい。

提言 1. 製薬企業及び規制当局は、ベネフィット評価法を柔軟に検討すべきである。

- ・ 疾患メカニズムと開発候補化合物の作用機序を考慮し、評価変数、比較対照、エビデンスの質と量、情報の示し方の 4 点を整理することで効率の良い開発・試験計画を選択できる。
- ・ 従来の検証の枠組みの実施可能性は最大限、検討されるべきである。

提言 2. SCTs に関するガイドラインを作成する。

- ・ PMDA が公開している「新医薬品承認実務に関わる審査員のための留意事項」[\[9\]](#)には、柔軟な審査を行うことが示されている。柔軟な開発・試験を行うために、規制当局と製薬企業は協力し、ガイドラインを整理すべきである。

提言 3. 質の高い試験を実施する。

- ・ 従来と異なる方法を用いるということで、データの品質を犠牲にしてはならない。個々

のデータには常に高い品質が求められる。

- 評価変数の標準誤差を小さくするには、症例数を増やす方法と標準偏差を小さくする方法がある。SCT では、被験者数を増やすことが困難であるため、評価変数の精度の向上に対する工夫を行うべきである。

参考文献

1. ICH E8 ガイドライン, 臨床試験の一般指針 (1998)
2. ICH E9 ガイドライン, 臨床試験のための統計的原則 (1998)
3. 小宮山靖, 日本製薬工業協会 医薬品評価委員会, 第 119 回医薬品評価委員会総会「医療イノベーション 5 か年戦略後の医薬品開発の姿」-研究開発から製造販売後までの医療イノベーションへの挑戦- 講演要旨集, 2012/11/16 ; 58-59
4. IOM, Small Clinical Trials: Issues and Challenges (2001)
5. IOM, Rare Diseases Research and Orphan Product Development, Accelerating Research and Development (2011)
6. NIH and FDA training course, Science of small clinical trials (2010), <http://videocast.nih.gov/podcasts.asp?c=88>
7. NIH and FDA training course, Science of small clinical trials (2012), https://events-support.com/events/FDA-NIH_Science_Small_Clinical_Trials
8. Farid Kianifard, M. Zahur Islam, A guide to the design and analysis of small clinical studies. (2011) Pharmaceutical Statistics, 10, 363-368,
9. PMDA, 新医薬品承認審査実務に関わる審査員のための留意事項, 平成 20 年 4 月 17 日, <http://www.pmda.go.jp/topics/file/h200417kohyo.pdf>
10. FDA, Guidance for Industry, Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products, (1998) <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm078749.pdf>
11. United States Code, 2010 Edition Title 21 (21 U.S.C.)- Food and Drugs, Chapter 9, Federal Food, Drug, and Cosmetic Act, Subchapter V Drugs and Devices, Part A Drugs and Devices, http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAAct/ChapterVDrugsandDevices/default.htm#Part_A
12. Code of Federal Regulations, Title 21, Food and Drugs, Chapter I Food and Drug Administration Department of Health and Human Services, Subchapter D Drugs for Human Use, Part 314 Applications for FDA Approval to Market a New Drug, <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=314&showFR=1>
13. Code of Federal Regulations, Title 21, Food and Drugs, Chapter I Food and Drug Administration Department of Health and Human Services, Subchapter F Biologics, Part 601 Licensing, <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=601&showFR=1>
14. United States Code, 2010 Edition Title 21 (21 U.S.C.)- Food and Drugs, Chapter 9, Federal Food, Drug, and Cosmetic Act, Subchapter V Drugs and Devices, Part A Drugs and Devices, Sec. 356 - Fast track products, <http://www.gpo.gov/fdsys/pkg/USCODE-2010-title21/pdf/USCODE-2010-title21-chap9-subchapV-partA-sec356.pdf>
15. CHMP, Guideline on Clinical Trials in Small Populations (2006) http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf
16. EMA, Concept paper on extrapolation of efficacy and safety in medicine development, Draft (2012) http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129285.pdf
17. アガルシダーゼベータ審査報告書, http://www.info.pmda.go.jp/shinyaku/P200400006/34053100_21600AMY00008_A100_1.pdf
18. Nuesch E., Trelle S., et. al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. (2010) BMJ, 341. c3515.
19. L. F. Baker, J. F. Mudge. Making statistical significance more significant (2012) significance, p29-30. <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2012.00574.x/abstract>
20. J. F. Mudge, L. F. Baker, et al. Setting an optimal alpha that minimizes errors in null hypothesis

significance tests. (2012) PLoS ONE, 7(2) e32734.
<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0032734>

21. くすりの安全性を科学する, 第7章エビデンスの重みを判断するーパターンと関連性, サイエンティスト社 (2012)

資料作成者

日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース2

菅波 秀規	興和株式会社	(タスクフォースリーダー)
土屋 悟	大日本住友製薬株式会社	(タスクフォースリーダー)
鹿野 哲司	旭化成ファーマ株式会社	
松嶋 優貴	大塚製薬株式会社	
五月女 想	キッセイ薬品工業株式会社	
野村 真功	株式会社三和化学研究所	
工藤 健太郎	武田薬品工業株式会社	
岩崎 倫久	田辺三菱製薬株式会社	
庄司 晃	帝人ファーマ株式会社	
佐々木 紀幸	日本アルコン株式会社	
平井 隆幸	日本化薬株式会社	
笹川 裕次	Meiji Seika ファルマ株式会社	
吉田 篤人	株式会社ヤクルト本社	
廣岡 秀樹	アステラス製薬株式会社	(担当副部長)