



# **Complex Innovative Trial Design**

## **の現状と事例の整理**

**- FDA ガイダンス邦訳付き**

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

2023 年度 TF3-2

2024 年 4 月

## 目次

1.	はじめに.....	3
2.	CID に関する FDA Public Meeting.....	5
2.1	Session I: Complex Adaptive Designs .....	6
2.2	Session II: General Considerations for Other Innovative Designs Including External/Historical Control Subjects, Bayesian Designs, and Master Protocols .....	9
2.3	Session III: Clinical Trial Simulations for Confirmatory Trial Design and Planning .....	12
2.4	Session IV: Complex Innovative Design Pilot Program .....	15
3.	CID に関するガイダンスおよびプログラム .....	19
3.1	FDA ガイダンスの概要.....	19
3.2	CID Pilot Meeting Program .....	19
3.3	CID Paired Meeting Program .....	20
4.	CID の事例.....	25
4.1	エボラウイルス病のプラットフォーム試験.....	26
4.2	敗血症性ショックの第 IIb/III 相試験.....	34
4.3	病原性細菌感染症のプラットフォーム試験.....	45
4.4	小児多発性硬化症の第 III 相試験.....	47
4.5	慢性疼痛の第 II 相試験.....	49
4.6	全身性エリテマトーデスの第 IIb 相試験.....	52
4.7	びまん性大細胞型 B 細胞リンパ腫の第 III 相試験.....	58
4.8	ミオクロニー脱力発作を伴うてんかんの臨床試験.....	65
4.9	その他 .....	68
5.	おわりに.....	69
	補遺.....	70
1.	Interacting with the FDA on Complex Innovative Clinical Trial Designs For Drugs and Biological Products の邦訳 .....	70
I.	はじめに.....	70
II.	適用範囲.....	70
III.	CID 提案に関する FDA との協議.....	71
IV.	CID の事例と協議を促進するために推奨される情報 .....	75

## 1. はじめに

近年、医薬品開発の領域において Complex Innovative Trial Design（以下 CID と呼ぶ）が注目を集めている。CID には新薬申請や生物学的製剤の認可申請などにおいてこれまでほとんどまたはまったく使用されていない試験デザインが含まれ、現時点では複雑なアダプテーション、ベイズ流アプローチおよび外部対照の利用などその他の新しい臨床試験デザインを指すと考えられている。しかしながら、革新的および新規性のあるデザインは時代と共に変化し、CID には固定された概念はないとされている。CID を用いることにより、特に小児患者や希少疾患患者を対象とする治療法の開発など、通常のランダム化比較試験の実施が困難な状況においても適切にデザインされた試験の実施及びデータ解析が可能になり、医薬品開発を加速させることが期待されている。一方で CID は従来と比べると試験デザインが複雑になるため、包括的に試験デザインの特性を評価すること（シミュレーションによる動作特性の評価、デザインに関する各種パラメータの設定に関する十分な説明、結果の解釈可能性の検討など）が求められる。

FDA は2018年3月20日、CID の利用推進を目指した公開ミーティングとして「Public Meeting on Promoting the Use of Complex Innovative Designs in Clinical Trials」を開催した。このミーティングには、主に製薬会社、規制当局、アカデミアの統計家が参加し、臨床試験デザインにおける革新的な手法の利用についての議論や意見交換が行われた。ミーティングでは、CID の利点として効率性、病態理解の進展、個別化医療の可能性などが強調され、CID の実施に関連する課題やリスクに対する解決策、規制要件に関する意見交換がなされた。その後、FDA は革新的な試験デザインの採用を進めるための方策を更に検討し、2018年から CID Pilot Meeting Program を開始した。このプログラムは CID に関連した臨床試験の設計や実施における規制上の課題を解決することを目的としており、CID の適用に関する具体的な課題や疑問について、製薬会社やアカデミアなどの相談者が FDA と協議できる機会を提供している。相談者はこのプログラムを活用することで、提案する CID について FDA から助言や見解、試験実施の支援を得ることができる。また、FDA は2020年12月に「Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products」のガイダンスを発出し、相談者が CID の提案について FDA と議論する際の大きな指針を提示している。CID Pilot Meeting Program は2022年まで継続され、2023年より CID Paired Meeting Program として引き継がれている。

現在本邦では CID に関する特定のガイダンスやプログラムは存在していないものの、CID に該当する試験デザインの議論がなされている。CID が適用される状況としては、希少疾患や小児患者を対象とした臨床試験、アダプティブデザインやベイズ流アプローチ、外部対照を用いた臨床試験などが挙げられる。日本製薬工業協会は報告書「Rare disease の治療効果の推測法」、「アダプティブデザインに関する FDA ガイダンスの邦訳」、「アダプティブデザインの統計的推測に関する検討」、「薬事申請に Real World Data を外部対照として利用する際の留意点」を公表しており、これらも参考にされたい。また、AMED 医薬品等規制調和・評価研究事業にて平川班から「希少疾患領域の臨床試験におけるベイズ流アプローチの適用に関する基本指針」が公開されている。さらには、2023年の計量生物セミナー「ベイズ推

測と臨床研究への応用」にて、CID を含めた議論がされた。今後更に CID を活用した、より革新的な試験の計画・実施が増加することが想定される。本報告書は CID の理解を促進することを目的として、FDA の CID に関連した取り組みおよび CID を採用した臨床試験の事例の調査を通して、革新的な試験デザインの整理を試みる。

本報告書を次の構成で作成した。2章で、2018年3月20日に開催された Public Meeting の詳細をまとめた。本章では当日パネリストから出た意見と共に、後に PhRMA から発出されたコメントも併せて紹介し、CID 活用における専門家や規制当局の議論を俯瞰する。3章では2020年12月に発出された FDA ガイダンスおよび CID Pilot/Paired Meeting Program について紹介する。FDA ガイダンスは Appendix に全邦訳を添付したので参考にされたい。4章では CID Pilot/Paired Meeting Program に採択された臨床試験に加え、計画された試験デザインから CID に該当すると考えられる臨床試験の事例についてまとめた。FDA との議論ポイント、それを踏まえたシミュレーション計画などを調査可能な範囲でまとめた。マスタープロトコルの事例として、エボラウイルス病 (EVD) (4.1節)、薬剤耐性をもつ病原性細菌感染症 (4.3節)、慢性疼痛 (4.5節) の臨床試験、アダプティブデザインの事例として、敗血症性ショック (4.2節) と全身性エリテマトーデス (SLE) (4.6節) の臨床試験を、外部データから情報を借用するベイズ流アプローチの事例として、小児多発性硬化症 (MS) (4.4節)、びまん性大細胞型 B 細胞リンパ腫 (DLBCL) (4.7節)、ミオクロニー脱力発作を伴うてんかん (EMAS) (4.8節) の臨床試験を取り上げた。補遺には、CID に関する FDA ガイダンスの邦訳を載せた。

本報告書の読者としては、臨床試験の計画や実施に携わる統計解析担当者、臨床試験を計画する臨床メンバー、規制関連業務を担う薬事メンバーを想定している。本報告書が、本邦においてもより革新的な試験デザインを検討するための一助となれば幸いである。

## 2. CID に関する FDA Public Meeting

CID Pilot Meeting Program が開始される前の2018年3月20日に FDA の医薬品評価研究センター（CDER）および生物製剤評価研究センター（CBER）は、「Public Meeting on Promoting the Use of Complex Innovative Designs in Clinical Trials」を開催した。本ミーティングは4つのセッションで構成されており、主に米国および欧州の産官学に所属する統計家によって CID に関して議論された[1][2][3][4][5][6]。

各セッションにおける論点について、ミーティングでの議論の内容および Regulations.gov<sup>1</sup> に投稿された PhRMA の意見について紹介する[7]。Friends of Cancer Research、Biotechnology Innovation Organization、medidata からのコメントも投稿されているが、PhRMA からの文書が唯一、本ミーティングでの論点に対するコメントとして、表形式でまとめられている。次節以降では、本ミーティングのパネリストのコメントと、それらに対する PhRMA からのコメントをまとめて表示している。

本ミーティングの参加者を以下に示す。

氏名	所属
Deborah Ashby (PhD)	Imperial College London
Julie Beitz (MD)	Office of Drug Evaluation III, CDER, FDA
Robyn Bent	Office of Biostatistics, CDER, FDA
Scott Berry (PhD)	Berry Consultants, LLC
Frank Bretz (PhD)	Novartis
Aloka Chakravarty (PhD)	Office of Biostatistics, CDER, FDA
Shein-Chung Chow (PhD)	Office of Biostatistics, CDER, FDA
Ivan Chan (PhD)	AbbVie Inc.
Scott Emerson (MD, PhD)	University of Washington
Steven Goodman (MD, PhD)	Stanford University School of Medicine
Frank Harrell, Jr. (PhD)	Vanderbilt University School of Medicine
Telba Irony (PhD)	Office of Biostatistics and Epidemiology, CBER, FDA
Laura Lee Johnson (PhD)	Office of Biostatistics, CDER, FDA
Stefanie Kraus	Office of Regulatory Counsel, CDER, FDA
Lisa LaVange (PhD)	University of North Carolina at Chapel Hill

<sup>1</sup> Regulations.gov は、米国政府が発行する連邦規制およびその他の関連文書の策定に関する情報の管理システムである。

氏名	所属
J. Jack Lee (DDS, MD)	MD Anderson Cancer Center
Gracie Lieberman	Genentech
Gregory Levin (PhD)	Office of Biostatistics, CDER, FDA
Roger Lewis (MD, PhD)	University of California, Los Angeles; Harbor-UCLA Medical Center
Olga Marchenko (PhD)	Bayer
Cyrus Mehta (PhD)	Cytel Corporation; Harvard University
Will Meurer (MD)	University of Michigan Health System
Thomas Permutt (PhD)	Office of Biostatistics, CDER, FDA
Dionne Price (PhD)	Office of Biostatistics, CDER, FDA
Karen Price (PhD)	Eli Lilly and Company
John Scott (PhD)	Division of Biostatistics, CBER, FDA
Rajeshwari Sridhara (PhD)	Office of Biostatistics, FDA
Joseph Toerner (MD)	Division of Anti-Infective Products, CDER, FDA
Z. John Zhong (PhD)	Biogen

## 2.1 Session I: Complex Adaptive Designs

Session I では、複雑なアダプティブデザインにおいて一般的に考慮すべき事項について焦点を当てて議論された[1][2][6]。アダプティブデザインとは、試験に参加した被験者の蓄積されたデータに基づき、試験デザインの1つ以上の側面に対して予め計画された変更を行うことを許容した試験デザインを示す。詳細については FDA ガイダンス[8]やその邦訳[9]等を参照のこと。なお、本セッションでは特に事前に計画されるアダプテーションに焦点が当てられた。

本セッションの前半では、アダプティブデザインについての基本的な概念や一般的な考慮事項について、以下のような内容が紹介された。

アダプティブデザインを用いることの利点としては、統計的検出力が高くなることや、倫理上の考慮事項の解決、薬剤効果の理解の向上などがあげられる。一方で限界もあり、誤った結論を導く可能性や推定の信頼性への問題といった方法論における課題、試験の機密性や完全性を維持するための試験運営上の課題、関心のある Estimand が変更されることによる解釈可能性や一般化可能性への影響といった潜在的な課題などがあげられる。また、アダプティブデザインを用いる試験は、場合によっては複数の種類のアダプテーションが含ま

れ、非常に複雑になることがある。このような複雑なアダプティブデザインには、Estimandなどの科学的側面に対するアダプテーションが含まれる場合もあり、また多くの場合、計画段階において動作特性を評価するためのシミュレーションを要する。複雑なアダプティブデザインを用いた試験の一例として、EVD患者を対象としたPREVAIL II試験や、早期の乳がん患者を対象としたI-SPY 2試験などがある。各試験の詳細については、それぞれ本報告書の4.1節およびデータサイエンス部会より2022年12月に報告されている「Rare diseaseの治療効果の推測法」[10]を参照のこと。

アダプティブデザインにおいて考慮すべき事項としては、上述の課題にもあるように、誤った結論を導く可能性の制御、治療効果の推定における信頼性、試験の完全性の維持などがあげられる。想定される中間解析の回数や時期、中間解析および最終解析の統計的手法、アダプテーションの決定規則など、事前に規定する試験デザインの範囲も、重要な考慮すべき事項の一つである。また、アダプティブデザインのための文書は、一般的な文書よりも包括的な内容となり得るため、考慮が必要である。これにはデザインの理論的根拠、重要な動作特性の評価、アダプテーション、モニタリングおよびデータアクセス計画、包括的なシミュレーションレポートなどが含まれるだろう。

前半の内容を踏まえ、本セッションの後半では論点として下記3点が議論された。論点に対するパネリストから出た意見の一部と、ミーティング後に発出されたPhRMAからのコメント[7]を紹介する。

複雑なアダプティブデザインを適切かつ効果的に用いるために最も重要な原則は何か。

#### パネリスト

- 目的および Estimand を明確にするべきである。アダプテーションを実施するのであれば、それが試験の目的を達成するためになぜ必要であるかを考えるべきである。(Dr. Ashby, Imperial College London)
- 必要以上に複雑にせず、実用的なデザインを検討することが重要である。(Dr. Chan, AbbVie Inc., 他)
- 統計家、企業の臨床専門家、治験責任医師、独立データ委員会などの関係者の協力体制、そのための教育が重要である。(Dr. Zhong, Biogen)
- デザインの質、完全性、妥当性を考慮することも重要な原則である。(Dr. Chow, CDER)
- 導入したデザインが科学的に評価され、臨床試験をどのように特徴づけ、潜在的バイアスをどのように調整するか確認するためにはアダプテーションの規定の背景に統計学的根拠があるかを確認することも重要である。(Dr. Chan, AbbVie Inc.)

#### PhRMA

- FDA 審査部門全体にわたる規制上の意思決定を支持するような効果的で革新的な試験デザインに対する、FDAによる受け入れと合意を確保することが重要である。

- 複雑で革新的なデザインに対する固有の課題への対処方法を明確にする FDA からのガイドラインも必要である。ただし、ガイドライン内には柔軟性の余地も確保すべきである。
- これらのガイドラインには効率性の欠如により受け入れられない可能性があるデザインの例を含める必要がある[11][12]。

### アダプティブデザインの事前規定はどの程度すべきか。

#### パネリスト

- 決定規則は事前に定めることが重要である。決定規則のアルゴリズムの規定は、何が試験における最優先目的であるかを考慮する上で重要となる (Dr. Zhong, Biogen; Dr. Ashby, Imperial College London)。
- すべてを事前に規定しきることは難しいため、何が事前規定で何が追加要因か、また何故そうなったのかという点についても透明化すべきである。(Dr. Ashby, Imperial College London)
- 事前のデザインの検討の際には、もし異なる決定規則を用いた場合にはどうなるか、という点についてもシミュレーションなどで確認しておくことも重要であると考えられる。(Dr. Goodman, Stanford University School of Medicine)
- アルゴリズムに関する情報が漏洩することでバイアスが生じることもあるため、事前に規定はしても公開はしない、ということも場合によっては必要ではないかと思われる。(Dr. Meurer, University of Michigan Health System)

#### PhRMA

- 現実的な場合に限り事前に規定することを推奨するが、適切に正当化できる場合には事前に定めたアダプテーションのルールの変更を許容するガイドラインを作成する。このガイドラインには、データアクセスおよび中間モニタリング計画に関するガイダンスも含める必要がある。
- FDA の同意を得たアダプテーションのガイドライン、データアクセス、中間モニタリング計画の事前規定が必要である。
- 様々な潜在的な試験結果に対するシナリオ計画を立てる必要もある。

治療効果の推定におけるバイアスについては第一種の過誤確率の問題ほど議論されていないのが現状であるが、点推定および区間推定の特性の評価はどの程度重要か。

#### パネリスト

- バイアスの調整について考えることは非常に重要であるが、それを行っている人はほとんどいないのが現状である。(Dr. Emreson, University of Washington)
- 有効性が極めて高い場合の早期中止では平均値に偏りが生じるといった特性があるが、ベイズ流の事後平均を用いると正しい方向に導くことができる場合がある。

(Dr. Harrell, Vanderbilt University School of Medicine)

- バイアスについて評価する上では、バイズの特徴について考えることが非常に重要である。また、バイズを用いて正しい結論を得るためには事前分布の母集団を得ることも非常に重要である。(Dr. Emreson, University of Washington)
- 区間推定については、特に頻度論的な考え方をを用いる場合や非劣性試験<sup>2</sup>の場合には重要となる。(Dr. Emreson, University of Washington)

#### PhRMA

- 複雑なアダプテーションを用いることの一番の目的は、効率的なデザインと解析を実現すること、例えば試験のコストを考慮して検出力を最大化することである。多くの複雑なアダプテーションを伴う試験デザインにおいては、不偏推定に対応する検定を用いる場合、検出力が大幅に低下する<sup>3</sup>ため、点推定の特性の評価は重要となる。
- 二番目の目的は、様々なサブグループや共変量により調整された集団についてベネフィットとリスクを評価することである。これはより複雑なプロセスであり、適切な推定が重要であるが、一般にバイアスを無くすることはできない<sup>4</sup>。
- 点推定値に加えて区間推定値も重要である。

## 2.2 Session II: General Considerations for Other Innovative Designs Including External/Historical Control Subjects, Bayesian Designs, and Master Protocols

Session I で焦点が当てられているように、アダプティブデザインの中でも複数の側面に試験計画の変更が許容されるような試験デザイン、または動作特性の確認のためにシミュレーションを伴う試験デザインであれば、CID として取り扱われる可能性は高い。Session II では、上記に加えて外部データを対照群やヒストリカルコントロールとして設定するデザイン、意思決定の判断根拠となる閾値を設け、バイズ流アプローチによって検証するデザインや、マスタープロトコルを用い、複数の疾患もしくは複数の候補医薬品を評価していくデザインも CID に該当する可能性がある事が紹介されている[1][3][6]。このセッションではこれ

---

<sup>2</sup> タスクフォースによる補足：非劣性の検証では検定ではなく信頼区間を用いる場合があるが、アダプティブデザインを用いる場合、従来のデザインよりも被覆確率が低くなることもあり、その際に信頼区間を用いて非劣性を検証すると第一種の過誤確率が高くなる。

<sup>3</sup> タスクフォースによる補足：複雑なアダプテーションを伴う試験では、未調整の推定方法ではバイアスが生じることがある。第一種の過誤確率を適切に制御し、かつ不偏推定に対応する頻度流の検定方法を用いると、対立仮説下にて保守的になり検出力が低下する場合がある。

<sup>4</sup> タスクフォースによる補足：主たる解析だけでなく、サブグループ解析や共変量調整でもバイアスが生じてしまう可能性がある。さらに、サブグループや共変量がアダプテーションと関連している場合、バイアスの性質はより複雑になると考えられる

らの革新的な試験デザインが必要な状況・考慮すべき一般的事項として下記が紹介されている。

1) 希少疾患など試験に参加できる症例数が限られている場合であって、以下の事を検討する場合

- 疾患レジストリや自然歴研究などの外部データを使用するとき
- 早期相の対照群データや治療群データなどの外部データを使用するとき
- 解析モデルを構築するために自然歴研究から取得した疾患情報を使用するとき

外部データを使用した具体的な例としてはセルリポナーゼ アルファ（プリニューラ）が挙げられる。本薬剤の Integrated Summary of Efficacy (ISE) 解析の中で、第 I/II 相試験と別途実施された自然歴研究を比較する際に1:1マッチングによる被験者の抽出を行い、その集団に対する有効性の評価を行っている。詳細な内容はデータサイエンス部会より2022年12月に報告されている「Rare Disease の治療効果の推測法」[10]を参照のこと。

2) ベイズ流アプローチを利用する場合

- ベイズ流アダプティブデザインを用いて、他のデータソースと新規試験の対照群データとの類似度に応じてデータソースから情報を借用するとき
- がん領域の早期試験における、用量毒性反応関係に統計モデルを仮定し、得られたデータに基づきモデルを更新する事で用量探索を行う試験デザイン（Continual Reassessment Method (CRM) など）
- 希少疾患の小児試験で、効率性を高めるために成人臨床試験からの事前情報を使用するとき

ベイズ流アプローチを利用した具体的な例としては敗血性ショックを対象としたセレプレシンの第 IIb/III 相試験が挙げられる。この試験ではセレプレシン群4用量とプラセボ対照群を設定し、用量探索パートにおいて反応適応的無作為化（response-adaptive randomization: RAR）と無益性による中止を含むデザインを採用している。詳細は4.2節を参照のこと。

3) マスタープロトコルを使用する場合

- 複数の疾患、バイオマーカー等で定義された複数の患者サブグループ、複数の治療法に対して一つの包括的なプロトコルで試験を実施する場合

マスタープロトコルデザインの利用により、医薬品の輸送や保管等のロジスティクスを合理化し、データ共有と収集を促進するためのインフラを備えた臨床試験ネットワークを確立することができる。

該当する具体的な例としては EVD を対象とした ZMapp の第 II 相試験（PREVAIL II）が挙げられる。この試験では、EVD の高い致死率に対する緊急性を考慮し、プラットフォーム試験デザインを採用している。詳細は4.1節に記載する。

Session II の後半では論点として下記3点が議論されている。論点に対するパネリストからの意見とミーティング後に発出された PhRMA からのコメントを示す[7]。

特に希少疾患や抗菌剤などのアンメットメディカルニーズのある領域において、どのような革新的な臨床試験デザインが医薬品開発を進展させるのか。

#### パネリスト

- 疾患のサブグループ間で情報の共有を認める事が必要であり、ベイズ流階層モデルを用いた情報借用が利用できる。またこのようなデザインを設定した際には、起こり得る様々なシナリオ（欠測やバイアスの程度など）を想定し、シミュレーションでそのデザインの動作特性を確認する必要がある。（Dr. Lewis, University of California, Los Angeles）

#### PhRMA

- 第 II/III 相のシームレスデザイン、アダプティブデザイン、ベイズ流アプローチ、マスタープロトコル、プラットフォーム試験、バスケット試験、複数の主要・副次評価項目、ヒストリカルコントロール/ヒストリカル治療効果を用いる試験、その他リアルワールドエビデンスやより臨床に近い状況を想定した臨床試験であるプラグマティック試験に由来するデータソースを使用する試験が該当する。

スポンサーや規制当局に革新的なデザインが受け入れられるためには、どのような要因が必要か。

#### パネリスト

- 障壁の1つには学术界・産業界・規制当局内のスキルや経験不足がある事は否めない。また審査グループ間で柔軟性に対するばらつきがある事も確か。そのためにも複雑な設計をシミュレートするソフトウェアの開発・利用が必要であり、受け入れられた前例を積み重ねていく事が重要。（Dr. Lewis, University of California, Los Angeles）

#### PhRMA

- 革新的デザインを促進するために、スポンサーと FDA 双方の協力関係をより強化するための柔軟性と機会が重要。

革新的な試験デザインが更に使用・許容されていくために必要なアウトリーチ活動、研究活動などはあるか。

#### パネリスト

- プラットフォーム試験では個々の治療法の効果を正確に検証するというよりは、疾患に対して最も効果的な治療法を見つける事に焦点を当てており、その点においては患者支援団体によって推進される場合がある。（Dr. Lewis, University of California, Los Angeles）

#### PhRMA

- バイズ流アプローチの下での革新的な試験デザインの動作特性を理解するための更なる検討
- 検証的試験における事例を蓄積していくこと
- DIA ワーキンググループやアメリカ統計協会 (American Statistical Association: ASA) 等の団体と協力体制を築くこと

### 2.3 Session III: Clinical Trial Simulations for Confirmatory Trial Design and Planning

ここでは、Session III: Clinical Trial Simulations for Confirmatory Trial Design and Planning において議論された内容を簡単に紹介する[1][4][6]。臨床試験には、様々な重要な動作特性があり、多数の臨床試験の条件についてシミュレートし、その結果を考察することが重要である。特にシミュレーションが必要となる場合として、複数のアダプテーションを持つ複雑な臨床試験、バイズ流の臨床試験、小規模な臨床試験が挙げられる。動作特性として考え得る要素として第一種の過誤確率、検出力、期待症例数、推定特性 (バイアスや平均二乗誤差等)、バイズ流の設定 (バイズ流の検出力、バイズの平均誤差) などがある。また、シミュレーションにおいて、以下5つの問題がある。

#### 1) シミュレーション空間の定義

シミュレーション空間の定義とは想定されうる対照群、試験群の治療効果の範囲 (境界線) である。例えば、非常に悪性度の高い癌に対する薬剤を考える。生存期間中央値は約1年とする。新しい試験では、対照群は予想通り1年の生存期間中央値を持っているかもしれない。数値的には生存期間の中央値は、1000年である可能性もある。その際、試験デザインの動作特性を評価するシミュレーションで可能性のあるすべてのシミュレーション空間、または単に臨床的に妥当な空間でシミュレートするのか、その境界線はどのように設定するべきかという問題である。

#### 2) シミュレーションのスコープの範囲 (どれくらいの回数等)

シミュレーションのスコープ範囲とは治療効果に加えて、臨床的パラメータ (例: 対照群の有効率)、統計的局外パラメータ (例: 分散)、オペレーション上のパラメータ (例: 症例登録速度) など、通常他の多くのパラメータについても仮定する必要がある。シミュレーションから適切な推定値が得られることを合理的に保証するために、どのくらい異なる組み合わせのパラメータを探索する必要があるかを検討することは重要である。例えば、対照群の有効率の20%と50%の間を探索する場合、どのくらいの間隔でシミュレーションを行うべきかという問題である。

#### 3) 多重検定/複数の仮説

多重検定を行った場合、誤った結論を導く可能性を検討する必要がある。その際、シミュレーションの焦点は、典型的には有効性の主要評価項目の主要解析に当てられる。しかし、解釈が複雑になるが複数の主要および、副次評価項目、有効性以外の安全性の解析項目等の試験の解釈に至る全ての関連可能性を考慮し、シミュレーションする必要があるか、どこま

で実施すべきかという問題がある。

#### 4) 情報事前分布を持つベイズの設定での適用

ベイズの設定として、事前分布の作成に用いたデータまたはその他の外部情報について、外部情報の情報源、妥当性、データの質、および信頼性について適切に説明できる必要がある。また、多くのベイズ流の臨床試験デザインは、頻度論的な動作特性に基づいて検討されている（第一種の過誤確率が制御され、少なくとも第二種の過誤確率のおおよその推定値が分かっているような決定規則に依存している試験デザインの提案が多数ある）という問題がある。

#### 5) リソースとレビュー

シミュレーションの計算が多い状況では、FDA にとってもレビューには膨大なリソースを必要とする。例えば、複数のエンドポイント、中間解析の結果に基づき、各群への被験者の割付割合を変更する **Adaptive allocation**、統計モデルのパラメータを推定する際に用いられるアルゴリズムであるマルコフ連鎖モンテカルロ (MCMC) 法および/または2群それぞれのデータ数を固定したまま、データをランダムに入れ替え、それぞれの結果とオリジナルの結果を比較することで、2つの集団に差があるかどうかを検定する **Permutation test** の組み合わせ等の全てのシナリオを検討することが困難で真にシミュレートできないいくつかの問題を持っているような試験である。

また、シミュレーションの報告は、既定のフォーマットがないため、ベストプラクティスが活発に公開されており、一般的に概要、試験デザイン、仮想的な試験結果、シミュレートする予定のシナリオ、結果（各シナリオからの動作特性）、治験デザインについて得られる結果の全体的な要約、シミュレーションコード、技術的詳細、統計的導出が含まれる。ただし、FDA によるシミュレーションの審査に関しては、明確な方針はなく、全ての審査部門や審査員毎に必ずしも同じとは限らない開発プラクティスがあり、レビューには申請者のシミュレーションコードまたは既製ソフトウェアによる検証、審査官が作成したコードまたは他のソフトウェアによる検証、追加シナリオの探索が行われることがある。

Session III の後半では論点として下記3点が議論されている。論点に対するパネリストからの意見と、ミーティング後に発出された PhRMA からのコメントを示す[7]。

#### 第一種の過誤確率のシミュレーションの範囲について

薬剤が無効である数（学）的に可能なすべてのパラメータ値を含めるべきか、それともある意味で臨床的に妥当な値だけを含めてシミュレーションを行うべきか。

##### パネリスト

- 第一種の過誤確率のシナリオに関しては考え得る全ての値をシミュレーションすることはできない。(Dr. Emerson, University of Washington)
- シミュレーションには臨床医が洞察できるような側面、試験や手技の臨床的特徴や患者の状況が組み込まれていないことがある。(Dr. Emerson, University of Washington)

**PhRMA**

- シミュレーションとシナリオの計画は、臨床的にもっともらしいシナリオ (数 (学) 的に可能なすべてのパラメータ値ではない) の下での第一種の過誤確率等の評価するために実施されるべきであり、これには生物学的および、疾患メカニズムの理解が必要となる。
- 多くの状況において、仮定は臨床試験データが得られた際にチェックされるかもしれない。
- ランダムな変動性を考慮した場合のシミュレートされた第一種の過誤確率の許容範囲や、試験終了後にパラメータを再調整する必要がある<sup>5</sup>かどうかを決定するためには、より明確な説明が必要である。

ベイズの枠組みの中で、形式的に事前情報を借用する場合、**error rate** のシミュレーションはどのように行うべきか。

試験の対象集団に近い集団を外部データから選択し、傾向スコアなどの手法により単群試験患者集団と比較可能な集団として構築される外部対照である **Synthetic control arm** のベイズの枠組みの中で、形式的に事前情報を借用する場合、第一種の過誤確率とは何を意味するのか。その代わりに他の **error rate** を考えるべきか。また、ベイズの枠組みにおける多重性の調整とは何で、どのような意味か。

**パネリスト**

- 閉手順における多重性検定の観点からは、多群間試験において、各群のすべての組み合わせと効果量のシミュレーションについて検討する必要がある。アダプティブ・エンリッチメントデザインを考えるのであれば、治療効果の全ての組み合わせ、サブグループと全体集団の組み合わせについてシミュレーションする必要がある。(Dr. Bretz, Novartis)

**PhRMA**

- 前提条件を明確にすることが基本であり、結果の頑健性は、仮定の様々な妥当なバリエーションの下でチェックされるべきである。
- 本質問では **Synthetic control arm** について言及しているが、過去のデータを使用する他のアプローチもあり、使用可能であると考えられるため、何が適用範囲内または適用範囲外となり得るのかは不明である。他のアプローチと比較して、1つのアプローチの選択の背後にある理由を学ぶことは有益であろう。

---

5 タスクフォースによる補足：試験終了後に得られた情報が試験前にシミュレーションで想定した値の範囲と異なり、重要な動作特性に影響を与えることが考えられる場合に再校正が必要となる可能性がある。

**臨床試験のシミュレーションを実施するための実践的な提案について**

反復回数、計算の詳細、ドキュメントの詳細など臨床試験のシミュレーションを実施するための実践的な提案は何かあるか。

パネリスト

- 前もって、目的、仮想患者の生成方法、データの分析方法、シミュレーションの収束方法を明確にしたシミュレーション戦略を立てる必要がある。ベストプラクティスについては、何年か前のアダプティブデザイン・ワーキンググループや最近のベイズ・サイエンティフィック・ワーキンググループが参考になる。(Dr. K. Price, Eli Lilly and Company)
- 今後改善できる分野の一つは、シミュレーションのサマリーをグラフィカルにまとめ、検討することである。(Dr. K. Price, Eli Lilly and Company)
- 他の提案として、シミュレーションされた特定の臨床試験とその試験がどうなったかを知ることはとても重要である。(Dr. Permutt, CDER)

PhRMA

- シミュレーションプログラムは検証され、レビューしやすいように文書化されるべきである。また、可能な限り数学的導出、数値計算、近似値を使用すること、シミュレーションの不確実性を反映した信頼区間を常に提供することを推奨する。
- 計算時間が問題となる場合には、分散を低減させるシミュレーション技術 (Ross, 1992, [13]) を考慮し、厳密なコーディングプロセス (アルゴリズムのフローチャート、コード中の広範なコメント、各手順の単体テスト、説明的な変数名、グローバル変数の回避、独立した検証など) を含む最善のシミュレーション手法 (O'Kelly et al, 2017, [14]) に沿って開発することが重要である。

## 2.4 Session IV: Complex Innovative Design Pilot Program

Session IV では、CID Pilot Meeting Program について以下の3つの論点が議論された[1][5][6]。各論点について、ミーティングでの議論の内容およびミーティング後に発出された PhRMA の意見について紹介する[7]。

**CID Pilot Meeting Program の初回提案に重要なデザインおよび解析の具体的な要素について**パネリスト

- 解析の詳細、すべてのシミュレーション情報が必要である。(Dr. Lieberman, Genentech)

PhRMA

- 少なくとも以下3点が初回提案に必要である。

- 当該の治療領域で、どのような革新的試験デザインが用いられているかとその種類
- クリティカルな問題に対処する性能と一般化可能性への順応性
- 検証的試験のデザイン

### CID Pilot Meeting Program の選択が優先されるべき試験デザインの種類について

#### パネリスト

- シームレスタイプのデザイン、情報借用するデザイン、マスタープロトコルを用いた試験デザインがパイロットプログラムの対象となる。(Dr. K. Price, Eli Lilly and Company)
- 治療領域としては、オンコロジー、希少疾患以外でも CID を利用可であるが、評価の基準を甘くするためのものではないことに留意しなければならない。(Dr. Marchenko, Bayer)

#### PhRMA

- 以下のような医薬品の申請または新しい適応症に影響を与える可能性のある試験が選択される候補である。
  - 第 III 相試験、シームレスな第 II/III 相試験デザイン、用量探索試験、ベイズ流デザイン、マスタープロトコル、ヒストリカルデータを外部対照とした試験、プラグマティック試験、規制当局の意思決定をサポートする早期臨床試験など

### CID Pilot Meeting Program への応募を阻害または奨励する可能性のある要因について

#### パネリスト

- 利用可能なプログラムコードの例が必要である。(Dr. Lewis, University of California, Los Angeles)

#### PhRMA

- 阻害する可能性のある要因
  - スポンサーの機密情報の開示
  - FDA の部門間の連携不足
  - FDA 審査担当者とパイロットプログラム審査担当者との間の知識のギャップ
  - 長期にわたる提案レビュープロセス、透明性の欠如
  - 固定された対面会議のタイミング
- 奨励する可能性のある要因

- スポンサーへのインセンティブ（FDA とのより頻繁なやり取りや規制上のサポート等）
- 共有すべき情報の明確化や黒塗りすべき機密情報の明確化
- パイロットプログラムとして承認された試験に対して、最初から最後まで同じ審査員チームを維持すること
- 明確で標準化されたプロセス、柔軟性、スポンサーと FDA のタイムリーなやり取り
- 採択または拒否された提案の明確化
- その他にも必要に応じて追加の会議を行えるプロセスを設定すること等、スポンサーの参加を促進するための考慮事項を提案している。

<参考文献>

- [1] Meeting Slide “Promoting the Use of Complex Innovative Designs in Clinical Trials PowerPoint Presentation” (fda.gov). <https://www.fda.gov/media/111551/download>
- [2] Meeting Recordings “Promoting the Use of Complex Innovative Designs in Clinical Trials\_0” (fda.gov). <https://collaboration.fda.gov/p3reuvfzx1m/>
- [3] Meeting Recordings “Promoting the Use of Complex Innovative Designs in Clinical Trials\_1” (fda.gov). <https://collaboration.fda.gov/p6k2yo7g8ig/>
- [4] Meeting Recordings “Promoting the Use of Complex Innovative Designs in Clinical Trials\_2” (fda.gov). <https://collaboration.fda.gov/p5dwetrzo9m/>
- [5] Meeting Recordings “Promoting the Use of Complex Innovative Designs in Clinical Trials\_3” (fda.gov). <https://collaboration.fda.gov/p8c8oxa981f/>
- [6] Transcript “Transcript: Public Meeting: Promoting the Use of Complex Innovative Designs in Clinical Trials” (fda.gov). <https://www.fda.gov/media/113320/download>
- [7] Regulations.gov “Promoting the Use of Complex Innovative Designs in Clinical Trials; Public Meeting; Request for Comments”. <https://www.regulations.gov/document/FDA-2018-N-0049-0001/comment>
- [8] Food and Drug Administration. “Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry”. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry>
- [9] 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 “アダプティブデザインに関する FDA ガイダンスの邦訳”（2021）
- [10] 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 “Rare disease の治療効果の推測法”（2022）

- [11] Jennison, C. and Turnbull, B. W., “Adaptive and nonadaptive group sequential tests,” *Biometrika*, Volume 93, Number 1.
- [12] Burman, C-F, Sonesson, C., “Are flexible designs sound?,” *Biometrics* 2006; 62: 1–6.
- [13] Ross, Shedon, “Simulation,” 1992
- [14] O'Kelly, M., Anisimov, V., Campbell, C. and Hamilton, S. (2017), “Proposed best practice for projects that involve modelling and simulation,” *Pharmaceutical Statistics*, 16(2), pp. 107-113.

### 3. CID に関するガイダンスおよびプログラム

#### 3.1 FDA ガイダンスの概要

2020年12月に FDA は、CID に関するガイダンス「Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products」を公刊した。このガイダンスは医薬品または生物学的製剤に関する CID の提案に対して FDA と議論をする際の治験依頼者および申請者のためのものである。本ガイダンスでは開発および規制審査における CID の使用、モデリング&シミュレーションに関連する技術的問題について治験依頼者が FDA からフィードバックを取得する方法、および FDA での審査のために提出すべき定量的および定性的手法の種類について説明されている。

本ガイダンスは以下の4つの章で構成されている。

##### I. INTRODUCTION

##### II. SCOPE

##### III. INTERACTING WITH FDA ON CID PROPOSALS

##### IV. EXAMPLES OF CID AND INFORMATION RECOMMENDED TO FACILITATE INTERACTIONS

##### V. REFERENCES

特にIII章では、FDA とのミーティングの利用や各種ミーティングでの相談内容、FDA よりフィードバックを得るための資料の作成およびその構成要素、CID にベイズ流アプローチが含まれる場合に考慮すべき特有の要素、および試験の動作特性などを評価するためのシミュレーションについて記載されている。また、IV章では現時点で CID と考えられる例として、Master Protocols、Leveraging Data From Phase 2 to Phase 3、および Sequential Multiple Assignment Randomized Trials (SMARTs)の事例が取り上げられている。詳細については本報告書の補遺を参照されたい。

なお、本報告書作成時点（2024年4月）において、CID に関する各種ミーティング・相談等は CID Paired Meeting Program（3.3節参考）に基づき実施されることとなっている。従って本ガイドラインに記載されている CID の各種ミーティング・相談方法等（III. A. Meeting Availability 等）は CID Paired Meeting Program の枠組みにも注意されたい。

#### 3.2 CID Pilot Meeting Program

Price and Scott (2021)[1]では、FDA が CID の使用を推進する取り組みの一環として開始した CID Pilot meeting program（2018年8月30日から2022年9月30日まで実施）について紹介されている。本項では、Price and Scott (2021)[1]の内容を概説する。

CID Pilot meeting program 実施の背景には、FDA 側の CID の受入れ可能性が不明瞭であったことや NDA 承認までの間に CID について公に議論できないといった課題があった。CID Pilot meeting program は、CID の計画と実施のための枠組みおよびインフラを提供した。ま

た CID Pilot meeting program には、FDA が CID Pilot meeting program に採択された試験デザインの側面を議論し、明確さ、一貫性、および透明性を高めるための開示要素も含まれていた。この開示機能により、ステークホルダーの知識の拡大にも貢献する（ステークホルダー向けの情報がウェブサイト[2]で公開されている）。

CID Pilot meeting program の相談案件は、医療ニーズおよび試験デザインの革新性で順位付けされ、要件を満たし、情報公開に関する FDA との合意が得られた場合に採択された（四半期毎に最大2つまで採択）。採択された案件は、約120日毎に開催される2回の相談を通して試験デザインについて議論すること可能であった。

論文執筆時点（2021年10月以前）で5つの試験が CID Pilot meeting program に採択されており、採択された試験の疾患領域は、デュシェンヌ型筋ジストロフィー、小児 MS、慢性疼痛、SLE、および DLBCL である。（Price and Scott (2021)[1]では、これら5つの試験の事例が紹介されている。本報告書の4章では、小児 MS（4.4節）、慢性疼痛（4.5節）、SLE（4.6節）、および DLBCL（4.7節）の事例を紹介する。デュシェンヌ型筋ジストロフィーの事例の詳細は、データサイエンス部会から2022年12月に報告されている「Rare disease の治療効果の推測法」[6]の3.6.1.2節を参照されたい）。

これまでの CID Pilot meeting program での経験から、FDA がシミュレーションの計画書および報告書に関する要件を明確にする必要があることが示された。更に、CID は繰り返し試験計画が変更される可能性を有しており、そのデザインに関する各分野による総合的な議論が重要であるといったことが明確になった。採択された5つの試験はいずれもベイズ流の手法が使用されている。これは、ベイズ流の手法が柔軟な試験デザインおよび解析を可能とするため、CID に適しているからである。また、複数の情報源を利用する場合にも適切であるかもしれない。

### 3.3 CID Paired Meeting Program

FDA は2018年度から2022年度で実施された CID Pilot Meeting Program を引継ぎ、2023年度より CID Paired Meeting Program を開始している。CID Paired Meeting Program に関するウェブサイト[2]にはこのプログラムの目標、申請手順、FAQ、CID のケーススタディ等が記載されている。本項では本ウェブサイトに基づき、CID Paired Meeting Program の概要を説明し、また FAQ について抜粋して紹介する。本項に記載していない CID Paired Meeting Program の概要やその他の FAQ については FDA のウェブサイト[2][3]を参照されたい。

本ウェブサイトでは、当該プログラムの目標について、「後期医薬品開発に重点を置いた CID アプローチの利用促進」および「FDA がまだ承認していない医薬品の試験デザインも含め、CID Paired Meeting Program で受け入れられた試験デザインについて FDA が公の議論を可能にすることによるイノベーションの促進」であると述べている。CID Paired Meeting Program のミーティングリクエストは四半期ごとに審査され、承諾されたリクエストについては翌四半期に最初のミーティングを行う。スケジュールや参加基準、選考等の詳細はウェブサイト上の「Procedures and Submission Information」を参照されたい。また、ウェブサイトでは CID のケーススタディを紹介しており、こちらの詳細は本報告書の4章 CID 事例でも

紹介されている。以下で、本ウェブサイトの FAQ を抜粋し紹介する。

#### **CID Paired Meeting Program とは何か？**

- CID Paired Meeting Program は、第6次処方薬ユーザーフィー法（PDUFA VI）の下で制定された CID Pilot Meeting Program から第7次 PDUFA（PDUFA VII）に引き継がれたプログラムである。
- PDUFA VI（2018年度から2022年度）において、FDA は、試験特性（第一種の過誤など）を決定するために解析的な導出をするのが困難な場合があり、シミュレーションを必要とする高度で革新的な試験デザインについて、CID Pilot Meeting Program を実施することに合意した。
- PDUFA VII に基づく Meeting Program の継続は、2022年10月20日の連邦官報[4]で発表された。Meeting Program は、複雑なアダプティブデザイン、ベイズ、その他の新規臨床試験デザインの使用を促進し、推進させるという目標を支援するものである。このプログラムは、CDER、CBER の FDA 専門家と医薬品開発における CID や解析について議論する機会を提供する。

#### **CID は医薬品開発プロセスをどのように合理化できるか？**

- CID は、臨床試験の効率を高めることで医薬品開発プロセスの合理化に役立てることができる。
- 例えばアダプティブデザインを用いた臨床試験では、効果の低い治療群の中止を可能にすることで、想定する症例数で薬効を検出できる可能性を高めることができる。
- 別のアダプティブデザインでは、薬効が高いという説得力のある証拠があれば、早期に試験を中止することができる<sup>6</sup>。

#### **FDA は CID を使用した医薬品を承認したことがあるか？**

- CID は単一プロトコルの下での EVD の臨床試験（PREVAIL II 試験）で使用された。
- この試験は EVD の流行が沈静化したため早期に中止され、この試験のデータを用いた医薬品は承認されなかった。
- 規制当局の意思決定を支援することを目的とした臨床試験で CID が使用された例が少ないことが、CID Pilot Meeting Program を作成した動機の一つである。

<sup>6</sup> タスクフォースによる補足：試験期間中に中間解析等を実施することを想定

同じ試験について、CID Paired Meeting のリクエストと並行して Type C Meeting<sup>5</sup> のリクエストを提出することは可能か？

- FDA は、CID Paired Meeting のリクエストと並行して Type C Meeting のリクエストを提出する目的は、医薬品の開発プログラムのタイミングや CID Paired Meeting が許可されるかどうかの不確実性に基づくためであると理解している。
- FDA は、FDA リソースの非効率的な使用であるため、並行して提出することを推奨していない。医薬品部門と統計審査部門のスタッフを含む CID 選定委員会がミーティングリクエストを評価し選定を行う。選定されなかったリクエストは、その後従来のパス<sup>7</sup>を辿ることを選択できる。

企業単位、化合物単位、異なる集団や試験デザイン単位等、リクエストの提出数に制限はあるか？

- 企業や化合物による提出数の制限はない。

CID Paired Meeting のリクエストが認められなかった場合、FDA はその情報を公表するか？

- FDA は、CID Paired Meeting Program に採択されなかったリクエストに関する情報を公表しない。

FDA はスポンサーが CID Paired Meeting Program に採択されたことを公表するか？

- FDA は、スポンサーが CID Paired Meeting Program に採択されたことを公表しない。
- スポンサーが CID Paired Meeting Program に受け入れられたことを公表することを選択した場合、FDA はプログラムに関する議論の際に、スポンサーが CID Paired Meeting Program の参加者であることに言及する可能性がある。
- FDA は、CID のミーティングリクエストの全体数およびプログラムに選ばれたリクエストの数を公表する。
- FDA はまた、スポンサーとの開示契約に基づいて許可された試験デザインの要素を公表する。

<sup>7</sup> タスクフォースによる補足：従来の承認申請のパスに戻ることが可能という意図である。例えば Type C Meeting での相談等を指している

**CID Paired Meeting Program に選ばれるために、スポンサーは CID についてどのような情報<sup>8</sup>を開示する必要があるか？**

- 具体的に開示されるべき情報は、提案された CID の内容によって異なるが、FDA は、CID の使用を促進するために有益な情報、CID の理解に関連する要素、および規制当局の承認をサポートすることを意図した臨床試験での使用の可能性に焦点を当てたいと考えている。
- 詳細については、CID Paired Meeting Program のウェブサイト[2]を参照されたい。CID の領域におけるイノベーションを促進するため、まだ FDA で承認されていない医薬品の試験デザインを含め、プログラムを通じて開発された試験デザインはケーススタディとして（例えば、ガイダンスや公開ワークショップで）FDA によって発表されることがある。

**CID Paired Meeting Program の CID 試験が進むにつれて、一般に開示される情報量が変わることはあるか？**

- FDA とスポンサーは、FDA が CID のミーティングリクエストを許可する前に、開示すべき情報について協議し、合意しなければならない。
- CID Paired Meeting Program に受け入れられた試験デザインについて開示される要素は、提案されたデザインの革新的な側面を考慮し、ケースバイケースで決定される。開示される情報についてはミーティングが進む中で複数回協議されることがありえる。

**初回とフォローアップの CID Paired Meeting はどのように違うのか、また推奨される構成はあるか？**

- FDA は反復的なプロセスを想定しており、最初のミーティングで1)提案されたデザインと解析の根拠、2)統計解析計画、3)シミュレーションの議論に焦点を当てる。
- 実りある議論ができるよう、初回会議の資料には十分な詳細を記載する。初回会議の目的は、重要な論点、整合性、明確化された事項、およびアクションアイテムを要約することである。
- フォローアップミーティングは、フォローアップミーティングパッケージを受け取ってから約90日以内に行われる。初回会議のフィードバックを取り入れた結果得られた新しい情報に焦点を当てる。

<sup>8</sup> タスクフォースによる補足：ケーススタディ等として FDA に公開される可能性がある情報

最終的な試験デザインの選択をサポートするために、FDA はデザインのオプションや関連するシミュレーション結果の検討過程をどの程度知りたいか？

- スポンサーは、試験デザインのオプションの変遷と関連するシミュレーションの簡単な要約をミーティング資料に含めるべきである。このような要約はスポンサーが検討したオプションについて FDA がよりよく理解するのに役立つであろう。
- 試験デザインの検討段階において、スポンサーは時間の経過とともに状況が変化する中で多くのデザインのオプションを検討し、その後、最適な動作特性・動作機能を持つ最終的な試験デザインを選択するのが一般的である。

シミュレーション報告書において、FDA は選択されなかった試験デザインのシミュレーション結果を確認したいか？

- シミュレーション報告書は主に選択された試験デザインに焦点を当てるべきである。ただし、他の候補デザインの概要や主要な結果、またそのデザインを選択しなかった理由を示すことで、開発戦略をより包括的に理解することが可能になる。

<参考文献>

- [1] Price D, Scott J. The US Food and Drug Administration’s Complex Innovative Trial Design Pilot Meeting Program: Progress to date. *Clinical Trials* 2021; 18(6): 706-710.
- [2] Food and Drug Administration “Complex Innovative Trial Design Meeting Program.” <https://www.fda.gov/drugs/development-resources/complex-innovative-trial-design-meeting-program>
- [3] Food and Drug Administration “Complex Innovative Trial Design Paired Meeting Program Frequently Asked Questions.” <https://www.fda.gov/drugs/development-resources/complex-innovative-trial-design-paired-meeting-program-frequently-asked-questions>
- [4] Food and Drug Administration, “Complex Innovative Design Paired Meeting Program.” Federal Register, Notices, Vol.87, No.202 (2022) <https://www.govinfo.gov/content/pkg/FR-2022-10-20/pdf/2022-22794.pdf>
- [5] 医薬産業政策研究所 “日米の新医薬品の承認審査に関する比較” 医薬産業政策研究所 リサーチペーパー・シリーズ No. 25 (2005) [https://www.jpma.or.jp/opir/en/issue/en\\_rs\\_025/s5nhtn0000000az1-att/paper\\_25.pdf](https://www.jpma.or.jp/opir/en/issue/en_rs_025/s5nhtn0000000az1-att/paper_25.pdf)
- [6] 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 “Rare disease の治療効果の推測法” (2022)

#### 4. CID の事例

4章では CID Pilot/Paired Meeting Program に採用された事例を含む8試験について取り上げる。各試験の革新的要素や利点は以下のとおりである。いずれの事例においても、従来の試験デザインと比較した場合にどのような利点があるのかという観点を含め、動作特性を示すための広範なシミュレーションが求められている。

事例	カテゴリー	主な革新的要素	利点
4.1 PREVAIL II	Adaptive design protocols/ Bayesian statistical design and simulation protocols/ Master protocol design	<ul style="list-style-type: none"> <li>試験中に比較対照である標準治療を更新することや、新たな治療群として試験治療を追加することを許容</li> <li>主要解析は BBA の考え方を適用</li> </ul>	<ul style="list-style-type: none"> <li>BBA は極端な状況下で試験を早期に中止可能（無益性による早期中止に加えて、治療効果が大きい場合に少ない例数で有効中止が可能）</li> </ul>
4.2 SEPSIS-ACT	Adaptive design protocols	<ul style="list-style-type: none"> <li>中間解析による早期の無益性中止や割付確率変更を検討する、2段階（第 IIb/III 相）アダプティブデザイン</li> </ul>	<ul style="list-style-type: none"> <li>より有効な用量群に割付けられる症例数を増加可能</li> </ul>
4.3 ADAPT	Master protocol design	<ul style="list-style-type: none"> <li>有効な症例数を増やすために感染症疾患間での情報借用</li> <li>早期の有効/無効中止基準の設定</li> <li>一度に複数の薬剤を同時に評価できるプラットフォーム試験</li> </ul>	<ul style="list-style-type: none"> <li>左記の方法を組み合わせることで、標準的な試験と比較して被験者を 45-60% 低減</li> </ul>
CID Pilot Meeting Program として採用された事例			
4.4 小児 MS	Bayesian statistical design and simulation protocols	<ul style="list-style-type: none"> <li>対象の適応症では前例のない実薬対照を用いた非劣性デザイン</li> <li>検出力向上のため外部データからの情報借用</li> <li>モデルに基づく成人から小児への外挿</li> </ul>	<ul style="list-style-type: none"> <li>実薬対照が患者登録の促進と途中脱落の防止に寄与</li> <li>外部データを活用することで、試験を効率化</li> <li>RMAP 事前分布を適用するため、新規試験データと乖離のある外部データから情報借用するリスクを軽減</li> </ul>
4.5 CPMP	Master protocol design/ Bayesian statistical design and simulation protocols	<ul style="list-style-type: none"> <li>同一の疾患で実施された、他の化合物のプラセボ群の情報借用</li> <li>異なる疾患で実施された、同じ化合物の治療群情報の情報借用</li> </ul>	<ul style="list-style-type: none"> <li>経時的な階層ベイズモデルを用い、各 ISA のプラセボデータの類似度によってプラセボ群情報の借用度合いを調整</li> </ul>
4.6 SLE	Adaptive design protocols/ Bayesian statistical design	<ul style="list-style-type: none"> <li>実薬用量群間の割付比を変更可能とする RAR の使用</li> <li>ベイズ流階層モデルによる用</li> </ul>	<ul style="list-style-type: none"> <li>より有効な用量群に割付けられる症例数を増加可能</li> <li>観察された用量群間の治療効</li> </ul>

	and simulation protocols	量群間の情報借用	果の類似性に基づいて、用量群間で動的に情報を借用することが可能
4.7 DLBCL	Bayesian statistical design and simulation protocols	<ul style="list-style-type: none"> <li>OS の比較に外部対照データを借用 (Commensurate prior を利用)</li> <li>OS の解析にベイズ流パラメトリックモデルを利用</li> </ul>	<ul style="list-style-type: none"> <li>名義水準の検出力を満たしつつ対照群に登録される症例数を低減</li> </ul>
4.8 EMAS	Bayesian statistical design and simulation protocols	<ul style="list-style-type: none"> <li>異なる適応症を対象とした同一薬剤の過去の試験からの情報借用</li> <li>症例数を決定するためにベイズ予測確率を用いた Goldilocks 法の適用</li> </ul>	<ul style="list-style-type: none"> <li>情報借用と適応的症例数により、症例数を削減</li> <li>動的な借用アプローチにより、類似度の低い外部データを借用するリスクを軽減</li> </ul>

#### 4.1 エボラウイルス病のプラットフォーム試験

PREVAIL II 試験 (ClinicalTrials.gov Identifier: NCT02363322) は、エボラウイルス病 (EVD) 患者を対象とした臨床試験である。CID Pilot Meeting Program および CID のガイダンスの公刊よりも前に開始された試験であるが、プラットフォームデザインや複雑なアダプティブデザインが適用された試験であり、CID に該当すると考えられるため、本項で紹介する。

##### 【エボラウイルス病】

EVD は、フィロウイルス科のエボラウイルスを原因とする疾患であり、非常に高い致死率を伴う重篤なウイルス性出血熱として知られている。2014年から2016年に西アフリカで発生した EVD の流行は、規模、期間、死傷者の多さにおいて前例のないものであり、この流行により28,000人以上の EVD の疑いまたは確定患者が発生し、11,000人以上が死亡した[1]。流行の開始時、候補となるいくつかの治療法が開発の初期段階にあったものの、EVD に対する臨床試験で評価された治療法はまだなかった。多くの実証されていない治療法がコンパッショネートユースのポリシーの下で実施されていたが、それらの結果は潜在的な有効性についての理解を助けるものにはならなかった[2]。

##### 【PREVAIL II 試験概要】

2015年3月、ポリメラーゼ連鎖反応 (PCR) 法で診断された西アフリカの EVD 患者を対象に、最適化された標準治療 (oSOC) に加えて候補となる新たな治療法を組み合わせた治療と、oSOC 単独とを比較する無作為化比較試験を開始した。試験計画段階で、EVD に対して利用可能とされる治療法は前臨床試験のデータに基づきランク付けがされており、これらの候補となる治療法を順に試験薬群に含めることとしていた (ただし、治験薬の供給上の問題が生じる場合にはこの限りではない)。なお、同等のランク付けがされている治療法の場合には、多群比較として同時評価を許容することとしていた[2]。試験開始時には、候補となる治療法の中で最も有望であった ZMapp (後述【薬剤】の項を参照) が oSOC と組み合わせ

る試験薬として選択された。oSOC は登録された地域のベストプラクティスを反映した標準治療として定義され、点滴、血行動態および電解質のモニタリング、および抗マalaria薬や抗生物質などの補助薬が最小限必要とされていた。なお、ギニアでは試験開始前に経口ファビピラビルが oSOC の一部として認められていた。あらゆる年齢の適格患者を、oSOC を受ける群と、oSOC+ZMapp の点滴静注を3回 (50 mg/kg を3日ごと) 受ける群に、1:1の割合で無作為に割り付けた。患者は、ベースラインのウイルスに対する PCR サイクルの閾値 (≦ 22または>22) と登録国に基づいて層別化された。主要エンドポイントは28日時点の死亡率とした[1][2]。

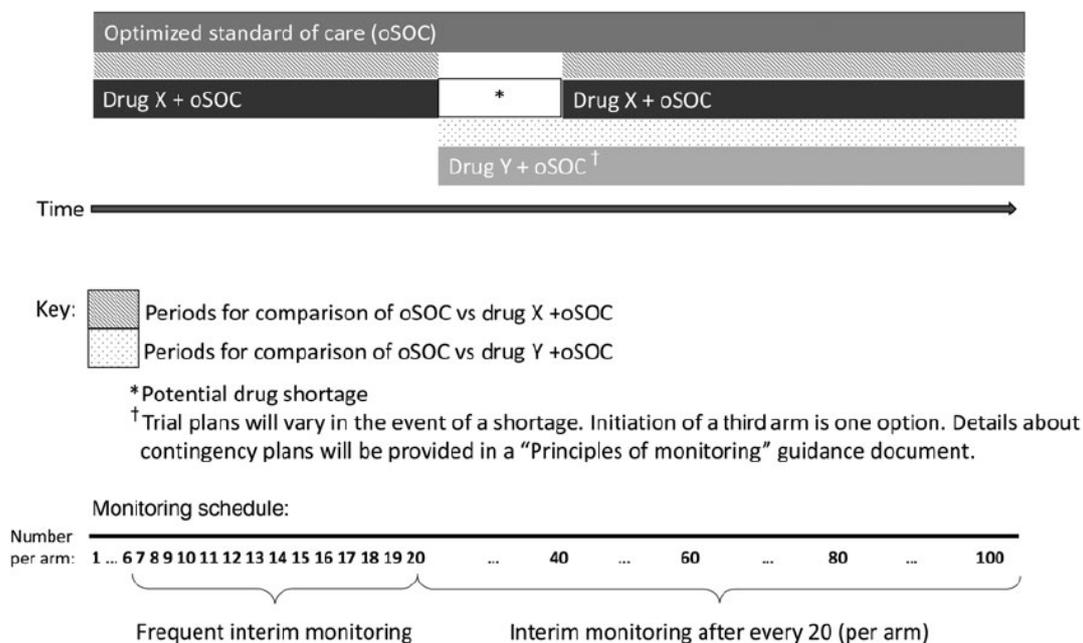
#### 【薬剤】

ZMapp はエボラウイルス表面に発現する糖タンパク質に結合する3種類のモノクローナル抗体の混合であり、非ヒト霊長類を対象とした研究データから EVD の有望な免疫系療法であることが示唆されており、候補となる治療法の中で最も有望な治療法であった[1][2]。

#### 【試験デザイン】

本試験のデザインとして、適応的な無作為化対照比較試験が選択された。このデザインの特徴は、試験中に EVD に対する活性があることが示された治療を現在の標準治療に追加して比較対照である標準治療を更新することや、新たな治療群として試験治療を追加することを許容する点にあった[1]。なお、比較には同時期に各群に無作為化された患者のみを含めることとしていた。例えば、図 4.1.1に示した Potential drug shortage の期間 (白いバーの期間に相当) に oSOC に割り付けられた患者は、薬剤 X+oSOC 群と oSOC 群の比較には含まれず、薬剤 Y+oSOC 群と oSOC 群の比較には含まれる。本試験では中間モニタリングを頻繁に行うことで、早期の治療結果が試験薬群で著しく有利な場合には当該薬剤を中止することを許容した。その場合、試験薬剤は更新された oSOC の一部として組み込まれ、更新された oSOC+新しい試験薬剤と比較される。その時点で更新された oSOC に既知の効果的な治療法が含まれている場合、群間差は試験当初に計画していた群間差よりも小さくなるため、症例数の再計算が必要となる可能性があった[2]。

図 4.1.1 PREVAIL II 試験における症例組み入れデザイン



[出典：Dodd et al. (2016) [2]の Figure 1]

## 【統計的デザイン】

主要評価項目は28日時点の死亡率であった。必要症例数は各群100例としたが、これはoSOC群の死亡率を0.4、試験治療+oSOC群の死亡率を0.2とした場合、すなわち相対的な死亡率の減少が50%であると仮定した場合、88%の検出力をもって差が検出できる例数であった[1]。

主要解析は barely Bayesian approach (BBA)の考え方をを用いて、「28日時点の生存率がoSOC群よりも試験治療+oSOC群の方が高い」という確率の事後確率の計算に基づくとした。試験終了時の事後確率が十分に高ければ(例：97.5%超)、有効であるという結論を下すことができるとした。しかしながら、EVDの発生という差し迫った状況を考慮すると、頻繁かつ早期の中間解析が必要であると考えられていた。したがって、治験の進行中に指定された時点(図4.1.1参照；各群6~20例までは各群に1例追加ごと、以降は各群20例が完了するごとに中間解析を実施する)として事後確率の推定が行われることとした。最終解析に先立つ中間解析では、試験治療+oSOC群の生存率がoSOC群の生存率を超える確率が99.9%以上ある場合、試験治療+oSOC群が優れていると宣言されるとした。最終的な解析では、この確率が97.5%を超えた場合に試験治療+oSOC群の優越性が宣言されるとした[2]。

表4.1.1は、本試験の設定において各群の症例数が6~10例の場合の中止境界を示している。たとえば、各群に8例の患者が登録された後、(1)試験治療+oSOC群で死亡が0例、oSOC群で7例または8例が死亡した場合、または(2)試験治療+oSOC群で死亡が1例、oSOC群で8例が死亡した場合には、早期中止が宣言される[2]。

表 4.1.1 各群における例数別の中止境界

Stopping Boundary, No. of Observed Deaths									
6 Subjects per Arm		7 Subjects per Arm		8 Subjects per Arm		9 Subjects per Arm		10 Subjects per Arm	
oSOC	Arm X	oSOC	Arm X	oSOC	Arm X	oSOC	Arm X	oSOC	Arm X
6	0	6	0	7	0	7	0	7	0
...	...	7	(1 or 0)	8	(1 or 0)	8	(1 or 0)	8	(1 or 0)
...	...	...	...	...	...	9	(2, 1, or 0)	9	(2, 1, or 0)
...	...	...	...	...	...	...	...	10	(3, 2, 1, or 0)

Abbreviation: oSOC, optimal standard of care.

[出典：Dodd et al. (2016) [2]の Table 2]

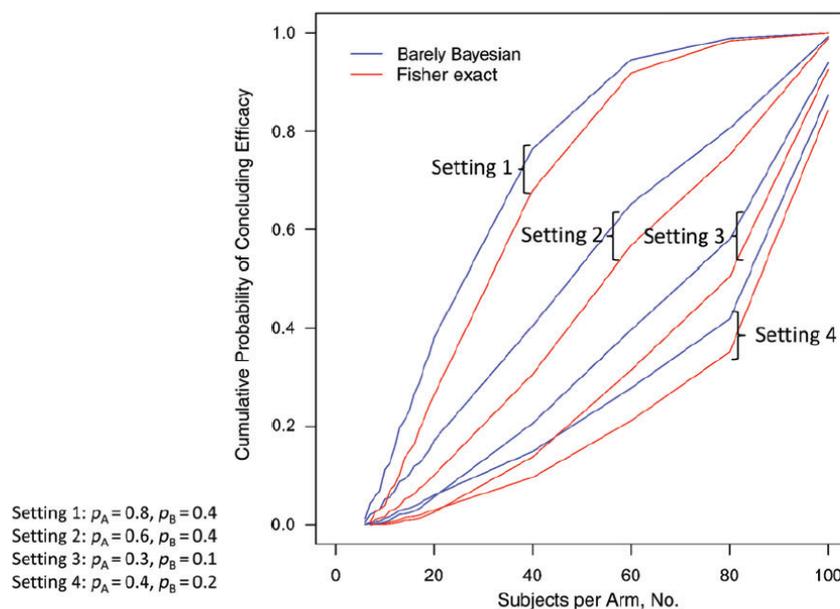
このデザインでは、新たな試験治療法が有効である場合に、早期に特定することを期待して頻繁に中間解析を行うとしていたが、新たな治療法の効果が不十分であったり、有害であったりする可能性もあった。他の潜在的に有望な治療法も利用できる場合に、効果のない治療法に被験者を無作為に割りつけることには大きなリスクが内在しており、無益であるという圧倒的な証拠がある場合には、無益性により試験を早期に終了するという倫理的配慮が求められる。これに対しては、事後確率を用いた有効性の評価とは異なる基準として、条件付き検出力を使用して無益性を評価するとした。条件付き検出力は、これまでに観察されたデータが与えられた場合に、試験終了時に統計的有意性が達成される確率のことであり、将来の治療効果についてさまざまな仮定の下で計算できる[2]。残りの患者における相対的な死亡率の減少が50%であると仮定したとしても、条件付き検出力が20%未満となる場合には、無益性による中止が検討されるとしていた。

#### 【従来の方法（頻度流）との比較】

BBA の重要な特徴は、従来のアプローチと比較して、極端な状況下で研究を早期に中止できることである[2]<sup>9</sup>。これは、標準的な中止境界と比較した場合に大きな違いとなる。たとえば、O'Brien-Fleming の境界を考えてみる。試験治療+oSOC 群で全員が生存し、oSOC 群で全員が死亡した場合、各群に15例が集まるまで O'Brien-Fleming の中止境界は越えられない。つまり、各群に14例が割り付けられたとき、その100%が試験治療+oSOC 群では生存し、oSOC 群の下では死亡した場合でも、境界は越えられない。一方、BBA に基づく場合、各群に6例が割り付けられた時、一方の群で6例全員が生き残り、もう一方の群で6例全員が死亡した場合に中止境界を越える。Haybittle-Peto の境界とフィッシャーの直接確率検定を併用する方法もあるが、これは BBA に比べやや保守的で、中止境界を越えるには合計14例の被験者の割り付けられたとき、oSOC 群の7例全員が死亡し、試験治療+oSOC 群では1例も死亡しないことが必要である。図 4.1.2は、治療効果がさまざまな設定の下で、Haybittle-Peto 境界とフィッシャーの直接確率検定を併用する方法と BBA を用いた方法における、帰無仮説を棄却する確率を比較したものである。BBA による最大の利点は、治療効果が大きい場合に少ない例数で有効中止ができる可能性があることとも言える。

<sup>9</sup> タスクフォースによる補足：従来の群逐次デザインのような頻度流の中止計画よりも、1例ごとに中止を判断できる柔軟な計画をしている、という意図と理解した。

図 4.1.2 治療効果の設定別の検出力と各群の例数の関係



\* $p_A$ : 試験治療 + oSOC 群における死亡率、 $p_B$ : oSOC 群における死亡率

[出典 : Dodd et al. (2016) [2] の Figure 4]

#### 【シミュレーションによるデザインの評価】

無作為生成された結果によるコンピューターシミュレーションを用いて、さまざまな死亡率を設定したシナリオの下で BBA を評価した[2]。表 4.1.2は、治療効果がない場合（帰無仮説が真の場合）および治療効果がある場合における、さまざまな死亡率の設定下での第一種の過誤確率および検出力をまとめたものである。なお、無益性の中間解析は、各群あたり40例の被験者の登録後に実施されるとした。

oSOC 群における死亡率の妥当な初期推定値の1つは0.50であった。oSOC 群の真の死亡率が0.5であり、かつ試験治療が効果を持たない場合、第一種の過誤確率は0.033となり、多くの試験で設定されている片側有意水準0.025よりも高くなるが、今回の状況では許容できると考えられた。oSOC 群の死亡率が0.50から離れるほど、第一種の過誤確率は若干低くなる。oSOC 群の死亡率に応じた、試験終了時の各群の平均症例数は50～90例の範囲であったが、条件付き検出力が低い場合には早期に無効中止となるため、いずれも100例未満であった。

治療が効果的であり、oSOC 群の死亡率が試験治療+oSOC 群の死亡率の約2倍である場合、検出力は妥当であった。たとえば、oSOC 群の死亡率が0.4、試験治療+oSOC 群の死亡率が0.2であるというシナリオは、研究を設計するとき最も合理的であると考えられた。このシナリオでは、平均症例数が1群あたり80例で検出力は約87%となった。oSOC 群の死亡率が0.8、試験治療+oSOC 群の死亡率が0.4というより極端なシナリオでは、各群の平均症例数は36例で検出力は100%に近づいた。

なお、このシミュレーションは各群の症例数を100例としており、oSOC の更新に基づくエ

フェクトサイズの変更に伴う症例数の再計算などは考慮していなかった。

表 4.1.2 真の死亡率の各設定に対する第一種の過誤確率および検出力

Mortality Rate, oSOC	Mortality Rate, Arm X	Rejection Rate	Mean Sample Size per Group	Proportion of Times Group Sample Size Is		
				≤40	≤60	≤80
Under the Null Hypothesis (Type I Error Rate)						
0.1	0.1	0.018	49.7	0.69	0.88	0.94
0.2	0.2	0.023	58.9	0.41	0.73	0.91
0.3	0.3	0.027	67.0	0.19	0.58	0.87
0.4	0.4	0.029	73.9	0.07	0.42	0.81
0.5	0.5	0.033	79.9	0.02	0.24	0.74
0.6	0.6	0.028	85.2	<0.01	0.09	0.63
0.7	0.7	0.027	90.3	<0.01	0.02	0.46
0.8	0.8	0.025	88.9	<0.01	0.03	0.51
0.9	0.9	0.023	76.3	0.08	0.33	0.77
Under a Treatment Effect (Power)						
0.10	0.05	0.200	63.4	0.47	0.63	0.72
0.20	0.10	0.474	80.1	0.17	0.33	0.48
0.30	0.15	0.708	84.5	0.11	0.23	0.40
0.40	0.10	>0.999	48.9	0.55	0.82	0.94
0.40	0.20	0.873	80.1	0.15	0.29	0.47
0.40	0.28	0.432	87.8	0.05	0.16	0.37
0.50	0.25	0.958	70.9	0.25	0.44	0.62
0.50	0.30	0.834	82.8	0.13	0.25	0.40
0.50	0.35	0.590	89.6	0.06	0.14	0.29
0.60	0.30	0.993	58.2	0.41	0.65	0.81
0.80	0.40	>0.999	36.0	0.76	0.95	0.98
0.80	0.20	>0.999	15.5	0.998	0.999	1.000

Abbreviation: oSOC, optimal standard of care.

[出典 : Dodd et al. (2016) [2]の Table 3]

#### 【最終解析結果[1]】

本試験は、新規 EVD 症例の減少により早期中止となり、試験治療としては ZMapp のみが評価されることとなった。試験終了までに、リベリア、シエラレオネ、ギニア、米国の施設で計72例が登録された。評価可能であった71例のうち21例が死亡し、致死率は試験全体で30%であった。内訳は、oSOC 群35例中13例 (37%)、oSOC+ZMapp 群36例中8例 (22%) であった。oSOC+ZMapp 群の oSOC 群に対する優越性は、事後確率が91.2%で、事前に設定した閾値の97.5%に達しなかった。頻度論的解析でも同様の結果であった (死亡率の絶対差 : -15%, 95%CI : -36%~7%)。

#### 【デザインに関する議論[2]】

本試験では、効果的な EVD 治療法の欠如、さまざまな臨床現場での死亡率に影響を与える要因の理解が限られていること、コンパッショネートユースの結果に関する報告しか存在しないことから、このような複雑で適応的なデザインが採用された。

高い死亡率が予想される治験では、効果を得るためにできるだけ早く中止して治療をすぐに受けられるようにすることが不可欠である。本試験では、oSOC 群で6例の死亡 (6例の被

験者中)が観察されたのに対し、試験治療+oSOC 群では死亡なし(6例の被験者中)という観察が得られた場合、有効中止とすることとした。しかしながら、より極端な結果(oSOC にランダム化された被験者10例中10例が死亡したのに対し、実験的治療にランダム化された被験者10例中死亡なしなど)の方がより適切だったと主張する人もいるかもしれないことには注意が必要である。この問題に関しては必ずしも合意に達するとは限らないが、採用される具体的な基準は状況に応じて変わる可能性が高く、治験開始前に徹底的に議論する必要がある。

西アフリカにおける新たな EVD 症例の減少を受けて、2016年1月に本試験は終了した。したがって、この試験デザインの全てについては、現在の感染拡大ではおそらく完全には検証されないだろうと考察された。しかしながら、本試験デザインに組み込まれた概念的な柔軟性は、今後の流行の際の新たな EVD の研究や、効果的な予防と制御手段が欠如している世界の地域で発生し続けるその他の新興感染症の流行においては、それでも必要となる可能性がある。

#### 【結果に対する議論[1]】

急性 EVD の新しい治療法に関するこのランダム化比較試験では、ZMapp と oSOC が oSOC 単独よりも優れている事後確率は91.2%であった。全体の28日までの死亡率は、ZMapp + oSOC 治療を割り当てられた患者の方が、oSOC のみを割り当てられた患者よりも15%低く(22%対37%)これは相対的な死亡率が40%減少したことに相当する。ただし、この結果は事前に指定された優越性の確率97.5%には届かなかった。

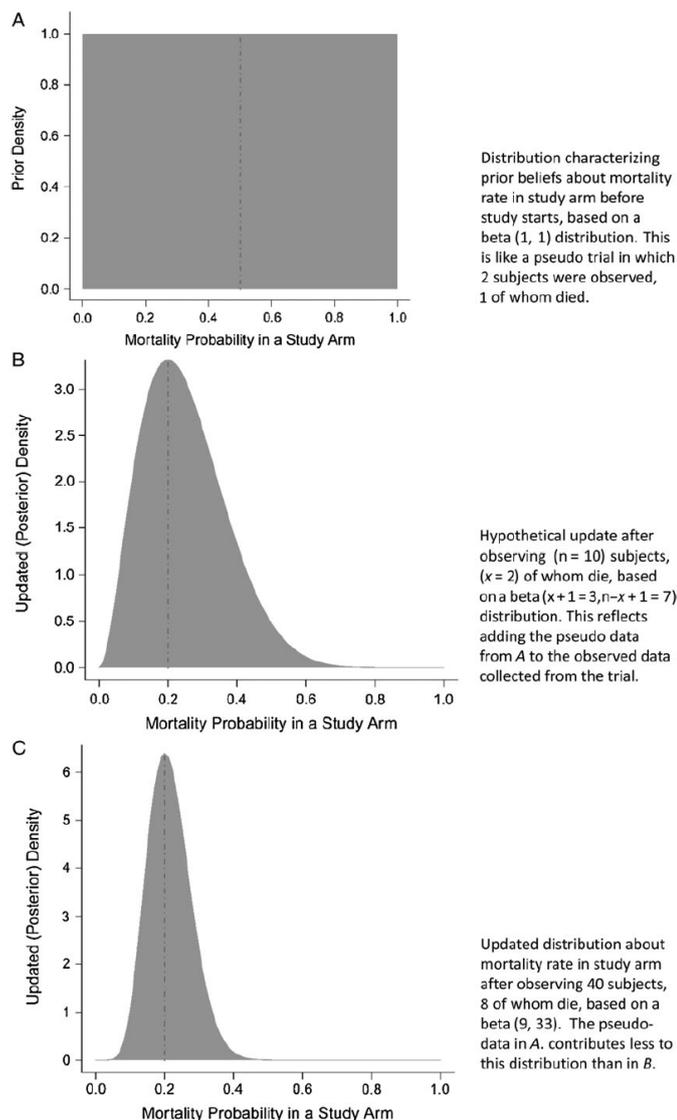
本試験は、新規 EVD 症例の減少により早期中止となったが、EVD の適格な新規症例数の急速な減少は、いかなる試験設計でも予想できなかった要因であり、最終的な結論に影響を与えた。

しかし別の意味では、この試験は、重大な公衆衛生上の緊急事態の状況下で、課題はあるものの、複雑なデザインを伴う無作為化対照試験を実施することが実際に実行可能であることを証明することに成功した。一連の課題と相まって、実験的介入はできる限り厳密な方法で評価され、公衆衛生政策が求める信頼をもってその成功または失敗を宣言できることが重要である。

#### 【ベイズ流の考え方】

図 4.1.3はベイズ流デザインにおける思考プロセスを示す[2]。図 A は28日死亡率の初期分布を示しており、本試験ではパラメータ( $\alpha, \beta$ )=(1,1)としたベータ分布に従うとした。これは、『2例の被験者が観察され、うち1例が死亡した』というデータが観察されたことを意味するが、実質的な不確実性が大きいため、図 A の様な一様分布を表すことになる。図 B と C は、それぞれ  $n=10$  および  $n=40$  の観測からの仮想的なデータに基づく28日死亡率の分布の更新を示している。それぞれ10例中2例、40例中8例が死亡した、という仮想的なデータを想定しており、図 C の分布は事後分布が図 B よりも多くのデータに基づいているため、死亡率0.2付近により集中している。

図 4.1.3 ベイズ流デザインにおける思考プロセス



[出典 : Dodd et al. (2016) [2]の Figure 2]

【事後確率の計算式】

以下の式は、A 群での死亡が $n_A$ 例中 $x_A$ 件、B 群での死亡が $n_B$ 例中 $x_B$ 件である場合に、A 群が B 群よりも優れているという事後確率を計算する方法を示している。ここで、 $p_A$ および  $p_B$ はそれぞれ A 群、B 群における死亡率を示す。 $G_A(p|x_A)$ は A 群で $x_A$ 件の死亡が観察された際のベータ累積分布関数であり、 $g_B(p|x_B)$ は B 群で $x_B$ 件の死亡が観察された際のベータ密度関数である。 $\binom{n}{k}$ は「n から k 個を選ぶ」ことを示す関数である。詳細は Proschan et al. (2016) [3]を参照されたい。

$$P(p_A < p_B | x_A, x_B) = \int_0^1 G_A(p|x_A) g_B(p|x_B) dp = \sum_{k=x_A+1}^{n_A+1} \frac{\binom{n_A+1}{k} \binom{n_B+1}{x_B} (n_B - x_B + 1)}{\binom{n_A+n_B+2}{x_B+k} (n_A + n_B - x_B - k + 2)}$$

<参考文献>

- [1] PREVAIL II Writing Group et al. A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection. N Engl J Med 2016;375:1448-56.
- [2] Lori E. Dodd, Michael A. Proschan, Jacqueline Neuhaus, et al. Design of a Randomized Controlled Trial for Ebola Virus Disease Medical Countermeasures: PREVAIL II, the Ebola MCM Study. The Journal of Infectious Diseases 2016;213:1906-13
- [3] Michael A. Proschan, Lori E. Dodd and Dionne Price. Statistical Considerations for a Trial of Ebola Virus Disease Therapeutics. Clin Trials 2016;13(1):39-48.

## 4.2 敗血症性ショックの第 IIb/III 相試験

敗血症性ショック患者を対象とした SEPSIS-ACT 試験 (ClinicalTrials.gov Identifier: NCT02508649) は、CID Pilot Meeting Program の開始および CID のガイダンスの公刊よりも前に開始された臨床試験である。しかしながら、複雑なアダプティブデザインが適用されており、CID に該当すると考えたため、ここで紹介する。なお、Protocol に盲検化維持のための手順や体制が記載されている。本報告書の主な目的ではないが、オペレーション上の留意事項として参考になると考えたため、同様に紹介する。

### 【敗血症性ショック (Septic Shock)】

敗血症性ショックは、感染症への反応が制御不能に陥ることで生命を脅かす臓器機能障害が生じる臨床症候群が重症化する病態である。要因は組織灌流の危機的な減少であり、肺、腎臓、および肝臓をはじめとした急性多臓器不全を招くことがある。

敗血症性ショックが発生した状況では、生命維持のために速やかな灌流圧の回復が必要であり、ノルアドレナリンやエピネフリンなどの昇圧剤の利用が不可欠であると考えられている。しかし、同時に、これらの薬剤は不整脈、心筋虚血/梗塞、急性腎障害等の重篤な有害事象を引き起こす可能性がある。このような合併症は、敗血症性ショックからの回復を妨げるか、回復期間を延長させ、長期の後遺症を起こす可能性がある。そこで、昇圧剤の使用期間を短縮する介入は、合併症のリスクを低下させ、患者の転帰を改善することが期待された。

### 【治験薬セレプレシン (Selepressin)】

セレプレシンは新規の選択的バソプレシン V1a 受容体アゴニストで、昇圧剤依存性の敗血症性ショック患者の治療のための血管収縮剤として、Ferring Pharmaceuticals 社が開発していた。

### 【SEPSIS-ACT 試験の概要】

SEPSIS-ACT は、昇圧剤依存性の敗血症性ショック患者の治療薬としてセレプレシンの有効性と安全性を検証する多施設・二重盲検・無作為化・プラセボ対照・2段階 (第 IIb/III 相) アダプティブ臨床試験 (ClinicalTrials.gov Identifier: NCT02508649) である。本試験は、2015

年7月に開始され、パート1の中間解析で無効中止と判断され、2018年2月に中止された。5か国52施設で患者6377名をスクリーニングし、適格基準を満たした868名をセプレシン群（計585名 [Arm1: 197名、Arm2: 189名、Arm3: 199名]）、またはプラセボ群283名に無作為に割り付けた。各 Arm は用量別コホートとして設定された（表4.2.1参照）。世界敗血症診療ガイドライン（SSCG2021）[3]では本試験結果を含むメタアナリシスの結果、セプレシンはノルアドレナリンに対して致死率で統計学的有意差はなく（セプレシン群: 41.8% vs ノルアドレナリン群: 40.45%; RR 0.99 (95% CI 0.84–1.18)）、セプレシンのノルアドレナリンに対する臨床上の意義は示されなかったことから、第一選択療法としてのセプレシンの使用は weak recommendation となっている。

#### 【用量別コホートの設計】

用量別コホートを最大で4種類設定し、パート1で最適用量を探索し、パート2では最適用量でプラセボ対照臨床試験を行った。

表 4.2.1 セプレシン治療群の用量

	Starting Dose (ng/kg/min)	Maximum Dose (ng/kg/min)	Range (ng/kg/min)
Arm 1	1.7	2.5	0-2.5
Arm 2	2.5	3.75	0-3.75
Arm 3	3.5	5.25	0-5.25
Arm 4	5.0	7.5	0-7.5

[出典：Laterre PF et al. (2019) [4] の Supplement の Protocol (Page 5)]

#### 【試験デザイン】

本試験は、2段階（パート1：第 IIb 相、パート2：第 III 相）のアダプティブ臨床試験である。パート1は治療を受けた評価可能患者300名～800名で構成された。パート2の症例数は、パート1およびパート2の合計症例数が評価可能患者1800名となるまで十分な症例数を組み入れ、パート2の症例数が最低1000名となるようにした。最終解析の対象は、本試験の両パートの全患者とした。

主要評価項目は、患者の治療開始から30日目までの昇圧剤を使用しないおよび機械換気を行わない日数（Vasopressor- and Mechanical Ventilator-free Days [PVFDs]）として、①生存期間（死亡した場合は0）、②昇圧剤を使用しない期間（連続する24時間での昇圧剤使用が合計60分間未満である日数）、③機械換気を行わない期間（連続する24時間での機械換気が合計60分間未満である日数）を組み合わせた複合エンドポイントとした。

パート1は患者200名を対象とした「burn-in」期間から開始し、その期間中は、患者の1/3がプラセボ群に、患者の2/3がセプレシン群に（2/9ずつ表 4.2.1に記載した Arm 1、Arm 2および Arm 3に）割り付けられるように無作為化する。「burn-in」期間終了後は、RAR を用いて主要評価項目である PVFDs に関して最大のベネフィットが得られると考えられるセプレシン群に優先的に患者を割り付ける（図 4.2.1①）。パート1全体で患者の1/3がプラセボ群

に割り付けられるように無作為化して、治療群と同じ時期の対照患者が組み入れられるようにする。Arm 3が Arm 2よりも PVFDs が長いという事後確率が50%以上でかつ低用量のデータ (Arm 1~3すべてのデータ) から重大な安全性の懸念は示唆されない場合には、中間解析時点での治療を受けた評価可能患者が200名~600名のときに Arm 4をオープンし割り付けを行う (図 4.2.1①②)。

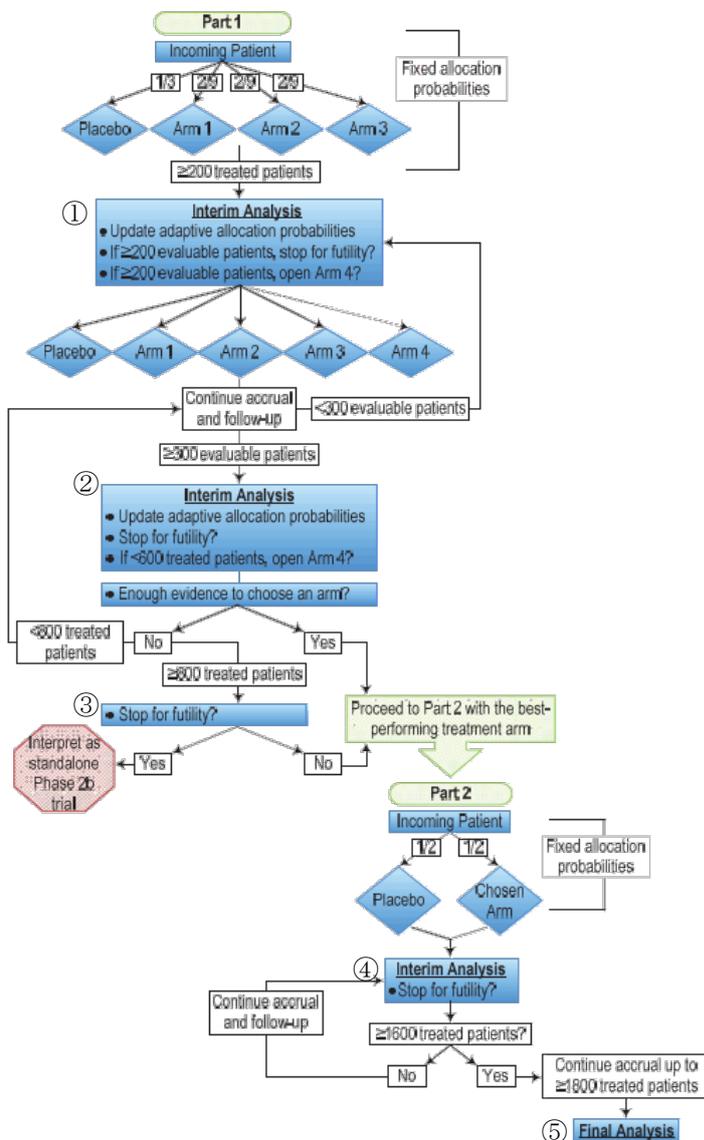
パート1の結果をうけてパート2を実施することを決定した場合には、パート2ではプラセボ群とセレプレシン最適用量群に1:1で割り付けられるように無作為化する。パート1の終了時に最適用量を特定する (図 4.2.1③)<sup>10</sup>。

事前に計画された中間解析を行うが、中間解析の実施中も患者の組み入れは中断せずに継続する。パート1で治療を受けた評価可能な患者数が300名~800名に達した後の各中間解析の時点で Arm 1~3の試験成功の予測確率 (当該用量でパート2を実施した場合にプラセボ群に対して統計学的有意に優越性を示す確率) を計算し、その確率が一番高い Arm で90%以上の場合に後にパート2にシームレスに移行する (図 4.2.1②)。評価可能な患者数が800名に到達した時点の中間解析で無効中止にならず、一番高い Arm の予測確率も90%未満だった場合には、予測確率が25%以上の場合はパート2に移行し、そうでない場合は試験を中止する (図 4.2.1③)。

---

<sup>10</sup> タスクフォースによる補足: Arm4の結果が示されていないため Arm4はオープンしなかったと思われる。

図 4.2.1 SEPSIS-ACT の Design Flow Chart



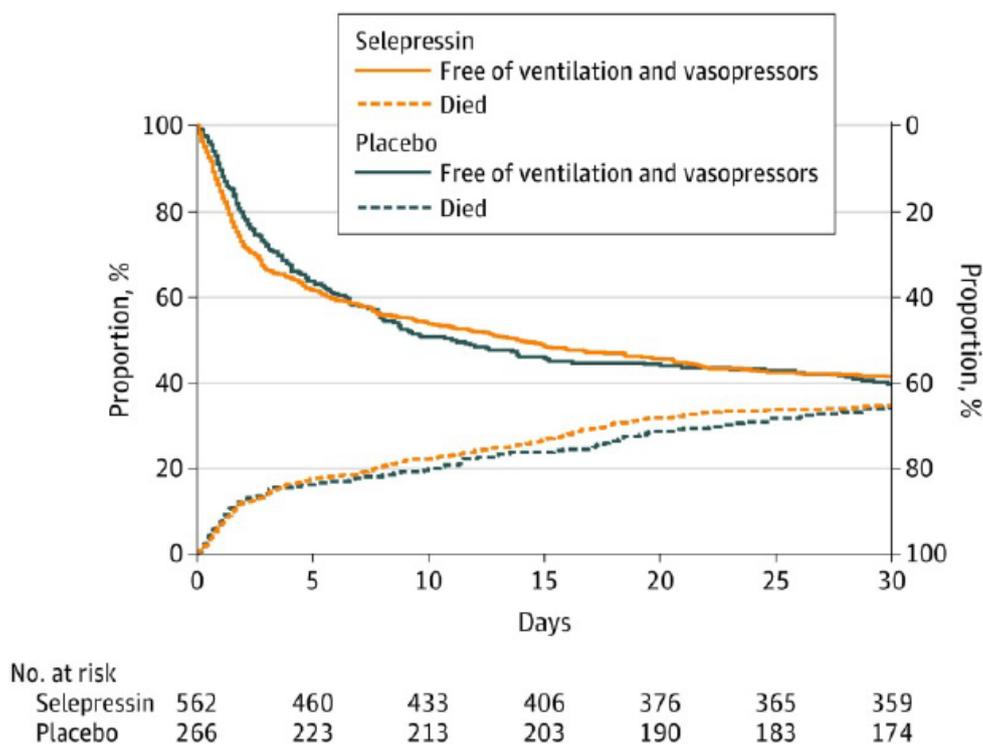
[出典 : Laterre PF et al. (2019) [4] の Supplement の Protocol (Page 6)]

【主解析 (図 4.2.1③)】

適格基準を満たし治療薬投与を受けた患者を解析対象集団として主解析を行った。主解析では、第一種の過誤を制御するために全てのセレプレシンの Arm を併合し、セレプレシン群562名とプラセボ群266名を比較した。有意水準は両側5%として、ファン・エルターレン検定 (層別 Wilcoxon 順位和検定) を用いた。層別因子は、機械換気の必要性 (有無)、敗血症ショックの発生から介入 (セレプレシン投与またはプラセボ投与) の開始までの時間 (6 時間未満または 6 時間以上)、ベースラインでのノルエピネフリン/ノルアドレナリン必要量 (<math>< 30 \mu\text{g}/\text{分}</math>または <math>\geq 30 \mu\text{g}/\text{分}</math>) とした。主解析の結果、統計学的有意差は認められなかった ( $p$  値=0.3015)。負の二項分布を仮定したモデル解析の結果、PVFDs 最小二乗平均 [95% 信頼区間]はセレプレシン群14.45日間 [12.82, 16.09]、プラセボ群15.00日間 [13.77, 16.23]、

群間差0.55日間 [-1.34, 2.43]と推定された。セレプレシン群およびプラセボ群における無作為化後から30日間の機械換気を行っておらず、かつ昇圧剤を使用しなかった累積確率、累積死亡発生率を図 4.2.2に示した。

図 4.2.2 セレプレシン群およびプラセボ群における機械換気を行っておらず、かつ昇圧剤を使用しなかった累積確率、累積死亡発生率



[出典：Laterre PF et al. (2019) [4] の Figure 2]

#### 【症例数の設計と第一種の過誤の制御】

最終評価可能症例数が最大で1800例と想定すると、パート1とパート2の結果を統合した場合の統計学的有意差を得るための全体検出力は、「プラセボ群と比較してセレプレシン群（4群全て）で真となる死亡率が1.5%低く、PVFDsが1.5日間延長する（PVFDs 1.5日間分の全治療効果に相当する）」という対立仮説の下で91%である。また、この対立仮説は、25%と想定された死亡率が相対的に6%減少することとPVFDsの延長に関するエフェクトサイズ（変化量を標準偏差で除した値）が0.30になることを仮定している。なお、セレプレシン群4群全てで死亡率を2%減少させ、PVFDsを2日間延長させる（PVFDs 2日間分の全治療効果に相当する）エフェクトサイズが真の場合、検出力が98%となる。このときパート2に進む確率は約99%となる。シミュレーションの結果、最終的な主解析（図 4.2.1⑤）に対する未調整の名目上の有意水準2.5%を用いた場合、片側検定での過誤確率は0.021であり、目標の過誤確率0.025未満に制御できることが示された。

#### 【盲検/非盲検の体制について】

本試験における盲検/非盲検の体制について紹介する。以下に示す3つの独立した当事者が

本試験のアダプティブな部分を保証した。

### 1) Randomization company

アダプティブなアルゴリズムを実施するための専門知識を提供する。

### 2) Statistical consulting company

Data Safety and Monitoring Board (DSMB) のレポート作成に関する専門知識を提供し、DSMB に統計およびプロジェクトマネジメントのサポートを行う。中間解析のレポートを DSMB に提供する。

### 3) Adaptive design company

アダプティブアルゴリズムのパフォーマンスを監督するためにデザインの専門知識を提供し、DSMB にアダプティブデザインに関する統計的なサポートを行う。

また、Trial Steering Committee (TSC) が設置され、盲検状況を監視した。TSC は、Ferring Pharmaceutical (治験依頼者)、Clinical Coordinating Center、trial investigator、臨床ケア/救急用薬剤/臨床試験の方法論のエキスパートで構成されている。

盲検/非盲検の体制については、Protocol 5.5 Blinding/Unblinding に図示されているため、参考にされたい ([2]および[4]から閲覧可能な Protocol 参照)。

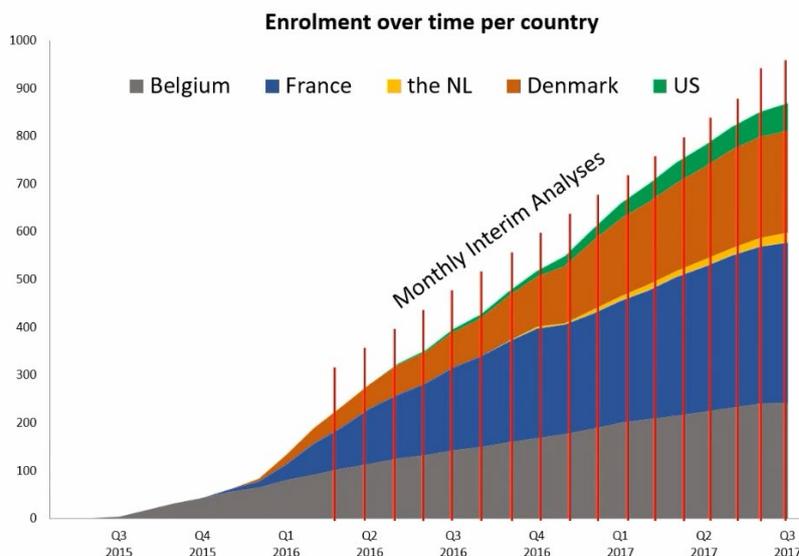
#### 【中間解析】

本試験では1ヵ月ごとに計17回、中間解析を実施した (図 4.2.3)。

パート1では中間解析で有効性の欠如による試験の中止 (無益中止)、割付確率の変更、Arm 4の追加、パート2の Arm 選択およびパート2への早期の移行を検討した。また、パート2では中間解析により有効性や安全性の欠如による試験の中止を検討した。(図 4.2.3)。

Statistical analysis plan の Appendix 5の Example Trial に当該デザインの動作例があり、規制当局への説明に参考になると想定されるため、適宜参照されたい。[2]

図 4.2.3 SEPSIS-ACT の中間解析数



[出典：Laterre PF et al. (2019) [4]から閲覧可能な European Society of Intensive Care Medicine (ESICM) 32nd Annual Congress、LIVES 2019 の Video]

表 4.2.2 SEPSIS-ACT の中間解析における検討内容

症例数	検討内容	図 4.2.1 の対応箇所
200 名以上 (治療例)	割付確率の変更 (PVFDs が最も多いと予想される Arm を大きい割付確率とする)	【パート 1】①
200 名以上 (評価対象例)	無益中止 (パート 2 で有意な結果が得られる予測確率が 5%を超える Arm がない場合)	【パート 1】①、②
200 名 (評価対象例) ~ 600 名 (治療例)	Arm 4 の追加 (Arm 3 の PVFDs Arm 2 よりも高い確率が少なくとも 50%の場合および Arm 1~3 のデータが重要な安全性のシグナルを示唆していない場合)	【パート 1】①、②
300 名 (評価対象例) ~ 800 名 (治療例)	パート 2 の Arm 選択および早期移行 (一部の Arm の試験成功の予測確率が少なくとも 90% の場合。同等の Arm があれば、投与量が最も低い Arm を選択)	【パート 1】①、②
800 名 (治療例) ~	パート 2 の Arm 選択および移行 (一部の Arm の試験成功の予測確率が少なくとも 25% である場合にパート 2 へ移行する。パート 2 の成功予測確率が 25%を超える治療群がない場合、決定的な結果が得られず試験は終了し、独立した第 IIb 相試験として解釈する。試験成功の予測確率が最も大きい Arm をパート 2 に選択するが、同等の Arm があれば、投与量が最も低い Arm を選択)	【パート 1】③
~1600 名 (治療例)	無益中止 (全体的に有意な結果の予測確率が 5%未満またはプラセボと比較して実薬治療群で 2% 以上高い死亡率が観察される予測確率が 90%を超える場合。1600 名以上となった最初の中間解析が最後の中間解析となる)	【パート 2】④

[出典：ClinicalTrial.gov に公開されている protocol[2]の 11 章 Interim analyses を参考に作成]

#### 【解析モデル】

死亡および生存者の PVFDs およびファン・エルターレン検定の期待値は以下のようにモデリングされた。モデル導出の詳細は、[6]supplement、[2]または[4] supplement から閲覧可能な Statistical analysis plan の Appendix を参照されたい。

#### 死亡および生存者の PVFDs

死亡

$$P_r\{X_i = 1 | d_i = d, s_i = s\} = \frac{\Delta}{\{\Delta + (1 - \Delta) \exp(\Phi_d + \psi_s)\}}$$

生存者の PVFDs<sup>11</sup>

$$P_r\{Y_i = k | d_i = d, s_i = s, X_i = 0\} = \frac{\pi_k e^{k(\theta_d + \omega_s)}}{\sum_{j=0}^{300} \pi_j e^{j(\theta_d + \omega_s)}}$$

$d_i$  : 被験者  $i$  に割り当てられた治療レジメン、プラセボは  $d_i = 0$

<sup>11</sup> 観察期間は30日間であるが、昇圧剤を使用しないおよび機械換気を行わない日数 (PVFDs) は小数第1位まで収集し、解析時にスケールアップしているため、分母は300までの和となっている。

$Y_i$  : 被験者  $i$  の PVFDs

$s_i$  : 被験者  $i$  が該当する層  $s_i \in \{0, 1, \dots, 7\}$

$X_i$  : 被験者  $i$  が死亡した場合  $X_i = 1$ 、それ以外は  $X_i = 0$

$\Phi_d$  : 用量  $d$  の死亡への効果

$\theta_d$  : 用量  $d$  の生存者の PVFDs への効果

$\psi_s$  : 層  $s$  の死亡への効果

$\omega_s$  : 層  $s$  の生存者の PVFDs への効果

なお、 $\Delta$  および  $\pi$  は以下のように定義された。

$$P_r\{X_i = 1 | d_i = 0, s_i = 0\} = \Delta$$

$$P_r\{Y_i = k | d_i = 0, s_i = 0, X_i = 0\} = \pi_k, \text{ with } \sum_{j=0}^{300} \pi_j = 1$$

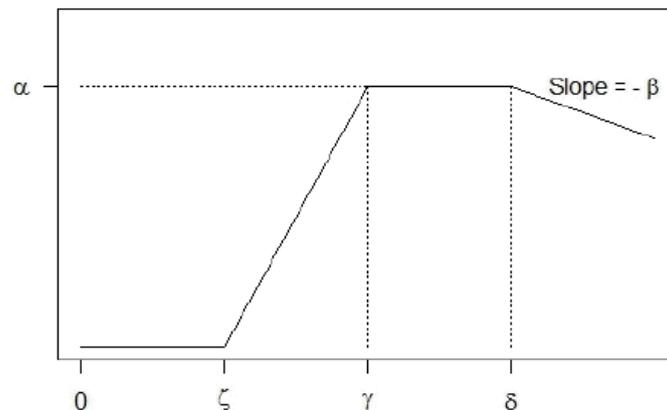
$\psi_s$  と  $\omega_s$  の事前分布は以下のように定義された。

$$\psi_s \sim N(0, \tau_\psi^2) \text{ and } \omega_s \sim N(0, \tau_\omega^2)$$

$$\tau_\psi \sim \text{Uniform}(0,1) \text{ and } \tau_\omega \sim \text{Uniform}(0,1)$$

ここで、用量  $d$  の死亡への効果  $\Phi_d$ 、用量  $d$  の生存者の PVFDs への効果  $\theta_d$  の推定のために、以下の Inverted-U dose-response models が定義された。

図 4.2.4 Inverted-U dose-response models



縦軸：最大の治療効果、横軸：治療レジメン

[出典：ClinicalTrial.gov に公開されている Statistical analysis plan [2] の Appendix 4]

上記のモデルについて、以下のような事前分布が定義された。

$$\delta - \gamma \sim \text{Uniform}(1,5)$$

$$\gamma | \delta \sim \text{Uniform}(0, \delta - 1)$$

$$\zeta | \gamma \sim \text{Uniform}(0, \gamma)$$

$\Phi_d$  および  $\theta_d$  は、以下のように定義された。

$\Phi_d = IU(d | \alpha_1, \zeta_1, \gamma_1, \delta_1, \beta_1)$  and  $\theta_d = IU(d | \alpha_2, \zeta_2, \gamma_2, \delta_2, \beta_2)$  for  $d = 1, 2, 3, 4$

$$IU(d | \alpha, \zeta, \gamma, \delta, \beta) = \alpha \left\{ \left( 1 - \left[ 1 - \frac{(d - \zeta)_+}{(\gamma - \zeta)} \right]_+ \right) - \beta (d - \delta)_+ \right\}, \text{ where } x_+ = \max(x, 0)$$

必要なパラメータの事後分布は以下のように定義され、Metropolis-Hastings アルゴリズム

によりサンプリングされた。

$$\begin{aligned}\Delta &\sim \text{Beta}\left(\frac{1}{3}, \frac{1}{3}\right) \\ (\pi_0, \pi_1, \dots, \pi_{300}) &\sim \text{Dirichlet}\left(\frac{1}{30}, \frac{1}{30}, \dots, \frac{1}{30}\right) \\ \alpha_1 &\sim \text{Uniform}(-0.35, 0.35) \\ \alpha_2 &\sim \text{Uniform}(-0.01334, 0.01334) \\ \beta_1 &\sim \text{Uniform}(0, 10) \\ \beta_2 &\sim \text{Uniform}(0, 1)\end{aligned}$$

#### ファン・エルターレン検定統計量の条件付き期待値

ファン・エルターレン検定統計量の posterior sample  $i$  の条件付き期待値は以下のように計算された。モンテカルロ法を使って、当該指標が正規分布の97.5パーセント点を超えたデータセットの数を数えることで試験の成功確率を判断した。

$$\begin{aligned}E(T|i) &= \frac{\sum_{s=1}^8 \{E(T_s|D, i) - E(T_s|H_0)\} / (N_s + 1)}{\{\sum_{s=1}^8 E\{\text{Var}(T_s|H_0)|D, i\} / (N_s + 1)^2\}^{1/2}} \\ \text{Var}(T|i) &= \frac{\sum_{s=1}^8 \{\text{Var}(T_s|D, i)\} / (N_s + 1)^2}{\sum_{s=1}^8 E\{\text{Var}(T_s|H_0)|D, i\} / (N_s + 1)^2}\end{aligned}$$

$T_s$  : 層  $s$  の Wilcoxon test statistic

$N_s$  : 層  $s$  の症例数

$E(T_s|H_0)$  : 帰無仮説における  $T_s$  の期待値

$\text{Var}(T_s|H_0)$  : 帰無仮説における  $T_s$  の分散

$i$  :  $i$  番目の posterior sample

$D$  : current data

#### **【シミュレーション】**

PVFDs の推定値のバイアス、検出力、各 Arm の選択の可能性、合計症例数、パート2へ進む確率について、シミュレーションが実施された。以下に概要を示すが、Berry Consultant (Adaptive Clinical Trial のツールである FACTS (Fixed and Adaptive Clinical Trial Simulator) の開発企業) によって作成された[6]supplement に掲載されているため、詳細は当該文書を参考にされたい。なお、[6]supplement における各種シミュレーションは、一部を除き[2]および[4]から閲覧可能な Statistical analysis plan の Appendix となっている。

#### PVFDs の推定値のバイアスについて

すべての実薬治療群の効果が同一である状況下でデータを発生させ、すべての実薬治療群を用いたシナリオおよびパート2で選ばれた Arm を用いたシナリオにおける、死亡率および PVFDs の推定値を求めることで評価された。シミュレーション回数は3000回とした。いずれのシナリオも治療バイアスが無視できる程度であったため、治療の推定値に対してバイアスは調整しないと判断した。([2]および[4]Statistical analysis plan の Appendix 8)

#### 検出力、各群の選択の可能性、合計症例数、パート2へ進む確率について

4つのケース（①死亡率、生存者の PVFDs がポジティブで、すべての実薬治療群の治療効果が同一、②死亡率、生存者の PVFDs がポジティブで、実薬治療群の治療効果が異なる、③実薬治療群の死亡者と生存者の PVFDs が異なる、④実薬治療群の治療効果は一定で、プラセボ群の治療効果が異なる）について、シミュレーションされた。なお、①～③は、プラセボ群の治療効果は一定であるもとでシミュレーションされている。

本報告書では、①のケースを紹介する。プラセボ群の死亡率は25%、生存者の PVFDs の平均を24とし、死亡者と生存者の PVFDs の平均は18とした。本解析時の8つの層は考慮されず、ファン・エルターレン検定の代わりに Wilcoxon 順位和検定が使用された。

プラセボ群に対する実薬治療群の治療効果（群間差）が0、0.5、1、1.5、2、3である表 4.2.3 の計6シナリオの下で、表 4.2.4に示す試験が成功する確率「P(Success)」、パート2の移行確率「P(Part2)」、1800例まで試験が実施される確率「P(1800)」、症例数の平均「Mean Sample size」が計算された。また、各シナリオにおいてパート2の用量として選ばれる確率「P(Target)」、パート2の用量として選ばれて試験が成功した確率「P(Target and Success)」、症例数の平均「Mean Sample Size Per Dose」が計算された。本書では、プラセボ群に対する実薬治療群の治療効果が0のシナリオを表 4.2.5に示した。

今回紹介した①の6シナリオを含む計27シナリオ（②の7シナリオ、③の6シナリオ、④の8シナリオ）において、検出力、各群の選択の可能性、合計症例数、パート2へ進む確率等が評価することで事前に本試験デザインの検討が行われた。([6]supplement)

表 4.2.3 ケース①の6シナリオ

Effect	Arm	P(Die)	Mean P&VFDs Survivors	Combined Mean P&VFDs
-	Placebo	0.25	24.0	18.0
<b>0</b>	Active (All dose)	0.25	24.0	18.0
<b>0.5</b>		0.245	24.5	18.5
<b>1</b>		0.24	25.0	19.0
<b>1.5</b>		0.235	25.5	19.5
<b>2</b>		0.23	26.0	20.0
<b>3</b>		0.22	26.5	20.7

Combined Mean P&VFDs は、 $1 - P(\text{Die})$  と Mean P&VFDs Survivors の積である。Effect は、Active と Placebo の Combined Mean P&VFDs の群間差であり、表 4.2.4 の Scenario に該当する。

[出典：Lewis RJ et al. (2018) [6] の Supplement 中の表を本タスクフォースの解釈で改変]

表 4.2.4 ケース①の6シナリオにおけるパート2の移行確率、1800例まで試験が実施される確率、平均症例数

Scenario	P(Success)	P(Part 2)	P(1800)	Mean Sample Size			
				Total	Part 1	Part 2	Target Dose
0	0.021	0.381	0.034	957.9	720.8	237.1	714.9
0.5	0.144	0.674	0.211	1275.9	715.8	560.1	715.7
1	0.590	0.902	0.651	1621.8	635.1	986.7	754.2
1.5	0.915	0.983	0.930	1760.7	524.9	1235.8	781.2
2	0.977	0.997	0.977	1784.2	452.7	1331.5	794.6
3	0.988	1.000	0.988	1792.8	440.0	1352.8	809.0

[出典 : Lewis RJ et al. (2018) [6] の Supplement]

表 4.2.5 ケース①のシナリオ0における各 Arm においてパート2の用量として選ばれる確率、パート2の用量として選ばれて試験が成功した確率、平均症例数

Arm	True Effect	P(Target)	P(Target and Success)	Mean Sample Size Per Dose		
				Total	Part 1	Part 2
Placebo	0			315.2	197.1	118.1
Dose 1	0	0.044	0.003	247.5	233.7	13.8
Dose 2	0	0.178	0.012	206.5	148.6	57.9
Dose 3	0	0.144	0.005	146.2	103.0	43.2
Dose 4	0	0.015	0.001	42.5	38.4	4.1

[出典 : Lewis RJ et al. (2018) [6] の Supplement]

## 【FDA との協議】

最終的な治験デザインは FDA CDER によって審査され、2015年7月14日に Special Protocol Assessment (SPA) の適用が合意された。SPA は、FDA と申請者との間で第3相臨床試験前に、対象疾患、目的、試験デザイン（評価項目、用法・用量、症例数）、解析方法などの詳細な取り決めに関して事前合意し、試験終了後は合意内容を変更せずにそのまま承認審査での承認要件として認めるという制度である。

## 【まとめ】

本事例では、生命を脅かす敗血症性ショックの臨床試験において、Monthly Based の中間解析により早期の無益性中止や割付確率変更を検討する、2段階（第 IIb/III 相）アダプティブデザインが用いられた。主要評価変数の推定値のバイアス、検出力、各群の選択の可能性、合計症例数、パート2へ進む確率について、シミュレーションが実施された。最終的な治験デザインは、FDA CDER によって審査され、SPA の適用が合意された。

<参考文献>

[1] [MSD マニュアル プロフェッショナル版 敗血症および敗血症性ショック](#)

[2] ClinicalTrials.gov Identifier: [NCT02508649](#) (Protocol と SAP へのリンクあり)

- [3] Evans L et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. Crit Care Med. 2021 Nov 1; 49(11): e1063-e1143.
- [4] Laterre PF et al. Effect of Selepressin vs Placebo on Ventilator- and Vasopressor-Free Days in Patients With Septic Shock. JAMA. 2019 Oct 15; 322(15): 1476–1485. (Supplement に Protocol と SAP あり)
- [5] Evans L et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med. 2021; 47(11): 1181–1247.
- [6] Lewis RJ et al. Rationale and Design of an Adaptive Phase 2b/3 Clinical Trial of Selepressin for Adults in Septic Shock. Selepressin Evaluation Programme for Sepsis-induced Shock-Adaptive Clinical Trial. Ann Am Thorac Soc. 2018 Feb;15(2):250-257.
- [7] FDA. Special Protocol Assessment Guidance for Industry. April 2018.  
<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/special-protocol-assessment-guidance-industry>

#### 4.3 病原性細菌感染症のプラットフォーム試験

ADAPT 試験は、Antibacterial Resistance Leadership Group (ARLG) が資金提供するプロジェクトで、医療ニーズが満たされていない領域における抗菌薬を評価するためのベイズ流適応的プラットフォーム試験である。

本タスクフォースが ADAPT 試験について調査した結果、関連する資料として見つかったのは ARLG のウェブサイト[1]および Berry Consultants のウェブサイト[2]に掲載されている情報のみであった。本項は、主に Berry Consultants のウェブサイトに掲載されている情報に基づいて構成している。

##### 【ADAPT 試験概要】

ADAPT 試験は、さまざまな感染症疾患 (院内肺炎/人工呼吸器関連肺炎 [HAP/VAP: hospital-acquired pneumonia/ventilator-associated pneumonia]、複雑性尿路感染症 [cUTI: complex urinary tract infection]、腹腔内感染症 [IA: interval appendectomy]) に対する薬剤耐性の治療を目的とした抗菌薬の効果をより効率的に評価するプラットフォーム試験である。ADAPT の主な特徴は、一度に複数の薬剤を同時に評価できることであり、対照群への割り付けの効率化、薬剤間のより適切な比較、および長年にわたって複数の薬剤を評価するための安定したネットワークを提供することが述べられている。この試験のデザインには、感染症疾患間での情報借用や、パフォーマンスが悪い (または良好な) 感染症疾患での薬剤の早期中止など、革新的な特徴が組み込まれている。これらの特徴を組み合わせると、各感染症疾患で薬剤ごとに別々の試験を行う場合と比較して評価対象の薬剤集団に対して必要な症例数を大幅に (場合によっては最大50%) 削減できる。

##### 【標準的なデザインの下での評価】

標準的なデザインは、各感染症疾患で対照群と治療群の1:1のランダム化比較であり、感受

性菌患者に対する非劣性、耐性菌患者に対する優越性、および全体患者に対する非劣性が評価される。しかし、耐性菌を有している患者は一般的に少ないことや感受性菌に対する非劣性マージンは比較的狭いことから、ある程度の検出力を保つために試験規模が大きくなる傾向にある。薬剤耐性に関する優越性について、標準治療に耐性がある被験者で標準治療を受けて治癒する割合は感染症疾患（HAP/VAP、cUTI、IA）ごとにそれぞれ40%、45%、70%と想定される。一方、標準治療に耐性がある患者で新規薬剤を投与して治癒する割合は、以下の3つのシナリオに分類される。

- **Alternative:** 新規薬剤が標準治療に取って代わる状況
- **Join Null:** 標準治療と効果が変わらない状況
- **UTIIA:** cUTI と IA には効果があるが、HAP/VAP は標準治療と変わらない状況

各シナリオで想定される治療成功割合を表 4.3.1に示す。例えば、HAP/VAP、cUTI、IA のすべての感染症疾患を合わせて300名/群（標準治療薬または新規治療薬）を登録する場合を考える。HAP/VAP、cUTI、IA の患者分布はそれぞれ30%、40%、30%とし、さらにその中で薬剤耐性菌を有している患者はそれぞれ19%、20%、25%とする。このときの HAP/VAP の薬剤耐性菌を有している患者は17名/群になる。

この状況において、各シナリオで検出力を計算した結果を表 4.3.2に示す。Alternative シナリオでは IA の患者集団における検出力が最も低く、HAP/VAP の検出力もそれほど高くない。標準的な試験デザインでは、薬剤耐性菌患者の症例数を増やすために全体の症例数を増やすことになる。

表 4.3.1 治療成功率の仮定

Scenario	HAPVAP	cUTI	IA
Alternative	0.82	0.88	0.89
Joint Null	0.40	0.45	0.70
UTIIA	0.40	0.88	0.89

[出典: ADAPT Summary Presentation (2016)]

表 4.3.2 各シナリオの検出力または第一種の過誤確率 (joint null シナリオ)

Power	HAPVAP	cUTI	IA
Alternative	0.762	0.913	0.419
Joint Null	0.035	0.029	0.036
UTIIA	0.035	0.913	0.419

[出典: ADAPT Summary Presentation (2016)]

#### 【ADAPT 試験のデザインの特徴】

ADAPT 試験では従来の方法ではなく革新的な方法を導入している。考えられる革新的な点は3つある。

- 有効な症例数を増やすために感染症疾患間で情報を借用する

- 効果が弱い（または良好な）感染症疾患に対して早期中止基準を設定する
- 一度に複数の薬剤を同時に評価できるプラットフォーム試験とする

これらの方法を組み合わせることで、標準的な試験と比較して被験者の45-60%を節約できる。詳細は、引用元の資料を確認されたい。

<参考文献>

- [1] Antibacterial Resistance Leadership Group “ADAPT” <https://arlg.org/adapt/>
- [2] Berry Consultants “Antibiotic Platform Design”  
<http://www.berryconsultants.com/antibiotic-platform-design/>

#### 4.4 小児多発性硬化症の第 III 相試験

本項では、CID のウェブサイトでの CID pilot meeting program に採択された試験の1つとして公開されている、小児多発性硬化症（MS）患者を対象とした第 III 相試験（ClinicalTrials.gov Identifier: NCT04926818）の事例[1]を紹介する（本事例の詳細は Schmidli et al. (2020) [2]を参照されたい）。

##### 【試験デザイン】

CID Pilot meeting Program に採択された小児 MS 患者を対象とした試験は、実薬対照、無作為化、二重盲検、バイズ流、群逐次、非劣性試験である。主要評価項目は、年間再発率（ARR: annualized relapsed rate）であり、有効性判定のための1回の中間解析が計画されている。

利用可能な外部データには、小児 MS 患者を対象とした1試験と、成人 MS 患者を対象とした複数試験が含まれている。これらの外部データからの情報は、統計モデルのパラメータに設定する有情報 meta-analytic predictive (MAP) 事前分布に活用される。更に、この MAP 事前分布は、無情報事前分布と結合され、観察されたデータとの整合性に基づき情報借用量を調整する robust meta-analytic predictive (RMAP) 事前分布に利用される。

##### 【試験デザインの革新性／新規性】

FDA が革新的または新規とみなした試験デザインの要素は以下の通りである。

- 対象の適応症で過去に使用されたことがない実薬対照を用いた非劣性デザイン
- 検出力および中間解析での有効中止確率の向上のための外部データからの情報借用
- モデルに基づく成人から小児への外挿

##### 【試験デザインの潜在的な利点】

- 提案された非劣性試験は、FDA が承認した薬剤を対照群に設定しているため、患者とその家族にとって魅力的であり、患者登録の促進と途中脱落の防止に寄与する。
- バイズ流アプローチを使用することで、外部データを活用し試験を効率化できる。
- 小児を対象とした外部データが限られているため、新規試験データとの乖離がある（交換可能性が成立しない）可能性（新規試験と乖離のある外部データから情報借用するリスク）がある。このリスクを軽減するために、外部データは、2つのコンポー

ネットの正規分布とロバストな無情報分布のコンポーネントからなる RMAP 事前分布により活用される。

#### 【試験デザインに対する検討事項】

- 提案された CID の利点
- 提案された非劣性マージンの妥当性
- 事前分布と非劣性マージンの両方に同じ外部データの情報を使用する影響
- 有情報 MAP 事前分布により情報借用するアプローチの適切性と解釈

#### 【シミュレーションによる動作特性の評価】

スポンサーは、5つのパラメータ（対照群の ARR、ARR-ratio（被験薬／対照薬）、登録期間、最大追跡期間、および RMAP 法を使用する際の有情報 MAP コンポーネントの重みの大きさ）を変化させ、動作特性を評価するためのシミュレーションを実施した。ARR が外部データと類似していると仮定した主要シナリオのもとで、提案デザインと同じ検出力を確保するためには、情報借用しない試験は、提案デザインよりも34%多い症例数が必要であることが示された。主要シナリオからの乖離の影響を評価することを目的に、追加シナリオのもとでの評価も実施された。そのシミュレーションの結果、ARR-ratio の主解析および感度解析で約-30%から+10%の差に相当するバイアスが認められた（頻度流解析のバイアスが最小であった）。ARR-ratio の推定精度（ARR-ratio の推定値の平均二乗誤差を使用）も追加シナリオのもとで検討され、多くのシナリオでベイズ流デザインが頻度流解析よりも優れていることが示された。これにより、新規試験データと外部データとの間に乖離があった場合でも、外部データの利用が ARR-ratio の推定値の精度を向上させることが示された。

事前分布と非劣性マージンの両方に外部データの情報を使用することによる影響についてもシミュレーションにより評価された。

#### 【FDA との議論】

FDA との議論では、小児 MS が希少疾患であり、アンメットニーズがあるという状況を踏まえると、非劣性試験が優越性試験と比較して優れている（既承認薬が存在するため、非劣性試験では、プラセボを使用する必要がない）という点が触れられた。スポンサーと FDA は、小児科医と患者にとって非劣性試験が魅力的であり、患者の負担を最小限に抑えられる可能性があることに同意した。

実施可能性および非劣性マージンについても議論された。FDA は、試験間の異質性を考慮した非劣性マージンを正当化するための包括的かつ体系的な文献のレビューを要求した。小児 MS を対象とした試験は1つしかなく、成人対象の試験結果を外挿することについての不確実性があるため、FDA は慎重な非劣性マージンの選択を勧めた。また FDA は、年齢による治療効果の違いを考慮し、追加の関連データを組み込んだモデリングの検討を提案した。更に、対照薬の効果の推定では、比較試験のデータを活用すべきであり、また試験間のバラツキもモデル化すべきであると助言した。FDA は、試験間の効果の一貫性の重要性を指摘した。更に FDA は、非劣性マージンと事前分布の両方に同じ外部データを使用する統計的意義を十分に説明する必要があると述べ、提案された事前分布と計画されたデザイン

の動作特性に関する広範なシミュレーションを要求した。

<参考文献>

- [1] Food and Drug Administration “CID Case Study: A Study in Pediatric Patients with Multiple Sclerosis”. <https://www.fda.gov/media/172313/download?attachment>
- [2] Schmidli et al. Beyond randomized clinical trials: Use of external controls. *Clinical Pharmacology & Therapeutics* 107.4 (2020): 806-816.

#### 4.5 慢性疼痛の第 II 相試験

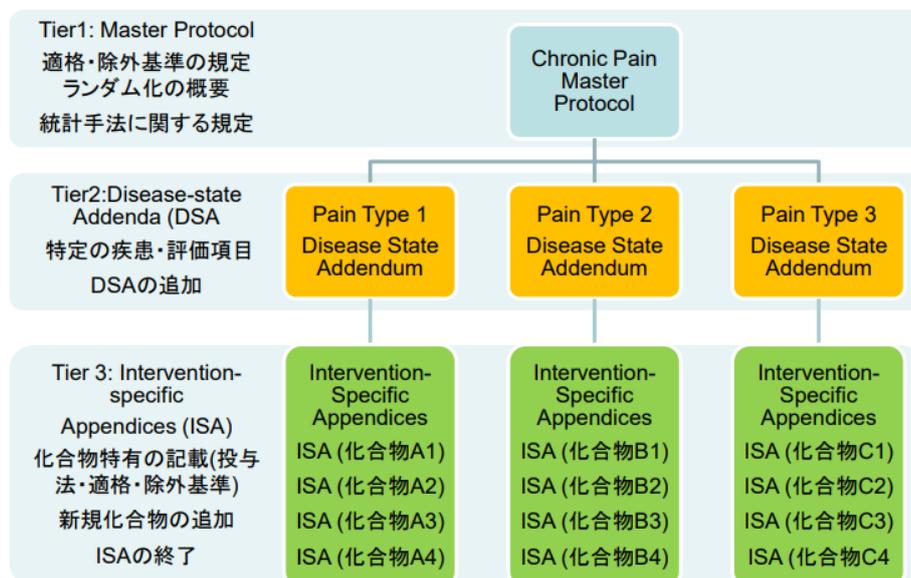
本項では、CID のウェブサイトでの CID pilot meeting program に採択された試験の1つとして公開されている、慢性疼痛を対象とした臨床試験の事例[1]を紹介する。

##### 【Chronic Pain Master Protocol の試験デザイン】

慢性疼痛は「治癒に要すると予測される時間を超えて持続する痛み、あるいは進行性の非がん性疾患に関連する痛み」と定義される。整形外科疾患や術後に遷延する痛み、帯状疱疹や糖尿病に関連する神経障害性疼痛などがある。

CID Pilot Program に採択された Chronic Pain Master Protocol (CPMP)は、慢性疼痛の変形性膝関節症 (OA)、慢性腰痛症 (CLBP) および糖尿病性末梢神経障害疼痛 (DPNP) の3疾患を対象とした第 II 相試験 (ClinicalTrials.gov Identifier: NCT05986292) である。複数の適応に対し、複数の化合物のシグナルを効率的に検出することを目的に、疾患および化合物を組み合わせたマスタープロトコルを採用している (図 4.5.1)。Tier 1: Master Protocol では、試験全体での共通の選択除外基準やランダム化の概要、統計手法等を規定している。Tier 2: Disease-state Addenda (DSA)では、OA、CLBP、DPNP の疾患の種類ごとに、疾患特有な試験の要素や評価項目を規定している。Tier 3: Intervention-specific Appendices (ISA)では、投与方法や各化合物特有の選択除外基準等を規定している。2024年3月時点で ClinicalTrials.gov に公開されている ISA を表 4.5.1に示す。

図 4.5.1. CPMP の試験デザイン



[出典：日本製薬工業協会シンポジウム (2023) [4]のスライド 10]

表 4.5.1 ClinicalTrials.gov に公開されている ISA の一覧

ClinicalTrials.gov Identifier	化合物	対象疾患
NCT05620563	LY3857210	OA
NCT05630196	LY3857210	CLBP
NCT05620576	LY3857210	DPNP
NCT04707157	LY3556050	OA
NCT04874636	LY3556050	CLBP
NCT04627038	LY3556050	DPNP
NCT05080660	LY3526318	OA
NCT05086289	LY3526318	CLBP
NCT05177094	LY3526318	DPNP
NCT04456686	LY3016859	OA
NCT04529096	LY3016859	CLBP
NCT04476108	LY3016859	DPNP

[出典：ClinicalTrials.gov で Funding = Industry とし、”Chronic Pain Master Protocol”を検索]

本試験の革新的または新規とみなされた要素として以下の2点が挙げられる。

- 1) 同一の疾患で実施された、他の化合物のプラセボ群情報を借用する
- 2) 異なる疾患で実施された、同じ化合物の治療群情報を借用する

他の ISA のプラセボ群情報を借用するために、すべての ISA はランダム化、二重盲検、プラセボ対照と共通の試験デザインで実施されている。また、各 ISA の主要評価項目や評価

時点を共通にし、標準化されたデータ収集がなされている。主要解析では経時的な階層ベイズモデルを用い、各 ISA のプラセボデータの類似度によってプラセボ群情報の借用度合いを調整することとしている。

さらに、試験の後半に実施される ISA では、既に完了した ISA のプラセボ群情報の借用を前提に症例数設計を行うことで、試験全体での症例数の低減が期待できる。

#### 【FDA との議論】

FDA は CPMP を計画するにあたり、考慮すべき点として下記内容を挙げている。

- 借用データの影響はどのようなものがあるか。
- 借用データに対する重み付けの手法の違いは何か。
- プラセボ効果が時間的に変化した場合の問題は何かがあるか。
- より類似性が高いと思われるデータをより強く借用するアプローチはあるか。例えば、古い研究よりも最近の研究に重みを付けたり、同一タイプの疾患に重みをつけることである。

上記懸念に対し、スポンサーは複数のシミュレーションを実施している。特に、プラセボ群情報を借用した際の推定値の正確さ（バイアス）と精度（標準誤差）を評価することに焦点を当てた。

他の ISA から借用したプラセボ群情報と、評価を行う ISA のプラセボ群情報の分布が一致している場合、情報を借用することで推定値の精度が向上する。一方、借用した情報と現在の情報に差がある場合、情報借用は推定にバイアスをもたらす、平均二乗誤差で評価した場合の全体的な正確さと精度を低下させる。

また、スポンサーは適切な情報借用の手法を検討するため、情報の借用程度を調整する手法である階層ベイズモデルおよび Commensurate prior についてシミュレーションを行った。シミュレーションの詳細は[1]を参照されたい。

FDA は、上記に示した借用データの影響に対する懸念を考慮し、CPMP を CID Pilot Meeting Program に採用した。

#### <参考文献>

- [1] Food and Drug Administration. “CID Case Study: Master Protocol to Study Chronic Pain”  
<https://www.fda.gov/media/155403/download?attachment>
- [2] 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 “Rare disease の治療効果の推測法” (2022)
- [3] 日本製薬工業協会. “Rare disease の治療効果の推測法 - CID パイロットプログラムとその実例 -”. 「Rare disease の臨床開発における治療効果推測の現状と課題」シンポジウム(2023)  
[https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS\\_202302\\_cid.pdf](https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS_202302_cid.pdf)
- [4] Saptarshi Chatterjee and Phebe Kemmer, “Learnings from Chronic Pain Master Protocol:

FDA Complex Innovative Designs Pilot Program” [http://www.bayesianscientific.org/wp-content/uploads/2022/10/DIA-BSWG-KOL-CPMP-21OCT2022\\_final.pdf](http://www.bayesianscientific.org/wp-content/uploads/2022/10/DIA-BSWG-KOL-CPMP-21OCT2022_final.pdf)

#### 4.6 全身性エリテマトーデスの第 IIb 相試験

本項では、CID のウェブサイトでの CID pilot meeting program に採択された試験の1つとして公開されている全身性エリテマトーデス (SLE) を対象とした臨床試験の事例[1]を紹介する。

##### 【全身性エリテマトーデス】

SLE は、自己免疫疾患であり、多臓器の関与と免疫学的異常が特徴で、組織損傷の多くは自己抗体の生成と免疫複合体の沈着によって生じると考えられている。日本では難病に指定されており、令和3年度の衛生行政報告例によると、SLE 患者数は64,304人であり、日本での有病率は、他のアジア諸国や西洋諸国とほぼ同じと報告されている。症状としては、患者間での異質性があり、発熱、全身倦怠感、およびその他の全身症状、関節痛、発疹、光線過敏、脱毛、アフタ性潰瘍が含まれる。患者は、生活の質の著しい悪化、ステロイド関連の毒性による副作用の悪化、重篤な血栓や感染症、臓器損傷、死亡のリスクがある[2][7]。

SLE 患者を対象とした試験では厳格な適格基準を設定しているため、適格患者数は限られており、患者の競合により、患者登録が遅くなる可能性がある。患者の募集は、治療の有効性を証明するのに十分な疾患活動性を有し、経口コルチコステロイドまたはその他の強力な免疫抑制療法の過剰使用を必要とするほど重篤ではない患者に限定されることが多い。また、患者がプラセボ対照ランダム化比較試験に安全に参加できるように、適格基準が年々更新されている。これにより、例えばプラセボを投与される患者や被験薬に反応しない患者が危険にさらされたり、過剰な治療を受けたりしないように、適切で過剰ではない標準治療が必要となる。これらの課題に加えて、新しい治療オプションの安全性と有効性に関する科学的な結論に達するために、参加した被験者の長期間の協力を必要とする。さらに、SLE 患者では、疾患活動性が経時的に変動し、複数の器官系が関与しているため、登録の適格性および治療に対する反応を評価する疾患活動性の測定は特に複雑となる。そのため、疾患活動性の測定の正確性を確保するためには、経験豊富な評価者が必要である。したがって、治験責任医師のトレーニングが不適切であり、データの解釈に大きな影響を及ぼす可能性がある治験実施施設を使用する必要性が生じた場合、治験データに問題が生じることになる。このような SLE を対象とした医薬品開発で直面するいくつかの深刻な課題に対処するために、アダプティブデザイン、ベイズ流の統計モデルが提案された。

##### 【試験デザイン】

Efavaleukin alfa (AMG 592) は、インターロイキン (IL) 2 ミュータントフラグメント結晶化 (Fc) 融合タンパク質であり、選択的に制御性 T 細胞の増殖を誘導する薬剤である。AMG 592 に対して安全性と有効性を評価する多施設共同ランダム化二重盲検プラセボ対照用量設定第 IIb 相試験 (ClinicalTrials.gov Identifier: NCT04680637) が実施された。この第 IIb 相試験の主要目的は、3用量の被験薬の有効性をプラセボと比較評価することである (図 4.6.1)。この試験は、後続の第 III 相試験のための最適な投与量を特定するために設計されており、

標準治療（経口ステロイド剤および他の免疫抑制剤および/または免疫調節剤）を受けているにも関わらず、活動性 SLE のある18～75歳の患者を320例登録する計画であった。主要評価項目は、52週時点の SLE Responder Index 4 (SRI4) を達成した被験者の割合 (SRI4反応率) である。SRI4は、ベースラインから4点以上の Safety of Estrogens in Lupus Erythematosus National Assessment SLE Disease Activity Index (SELENA SLEDAI) の改善（減少）、Physician Global Assessment (PGA) の悪化なし（スコアの増加が0.3点未満）、および British Isles Lupus Assessment Group (BILAG) でカテゴリーAに悪化した臓器系がない、かつカテゴリーBに悪化した臓器系が2つ以上ない、の4つから構成される複合評価項目である。従来の第 II 相用量設定試験では、通常24週間の治療を評価するが、この第 IIb 相試験では、主要評価項目が規制で求められる52週間で評価される。これにより、この試験は承認に必要な2つの検証的試験のうちの1つとして位置づけられる可能性があり、開発プログラムに必要な試験数を、1つの第 II 相試験および2つの第 III 相試験の3試験から2試験に減らすことができる可能性がある。



図 4.6.1. AMG 592（実薬）の第 IIb 相試験の試験デザイン

[出典：日本製薬工業協会シンポジウム (2023) [8]のスライド 7]

#### 【アダプティブデザインおよびベイズ流階層モデルによる情報借用】

本試験では、24週時の評価を完了した症例数が規定に達した時点で、独立データモニタリング委員会 (IDMC) を用いた中間解析が計画された。初回の中間解析 (40例) では反応適応的無作為化 (response-adaptive randomization: RAR) のみ、2回目以降の中間解析 (32例増えるごと) では RAR および無益性の評価、最終の中間解析 (320例) では administrative success (第 III 相試験の準備開始) の評価のみが計画されていた。

##### 1) RAR

試験開始時の割付比は1:1:1:1であるが、より有効な用量群に割付けられる症例数を増加させるために、中間解析時点の有効性の結果に基づいて、事前に規定したアルゴリズムによる実薬用量群間の割付比の変更が計画された (プラセボ群の割付比は固定)。変更後の割付比は、ベイズ流のモデルに基づく各用量群の52週時の SRI4反応率が最大である確率に比例し、有効性が高い用量群への割付比が増加する。用量反応関係が線形であるという仮定のもとでのシミュレーション結果の例示では、中間解析ごとに高用量群への割付比が増加し、5回目の中間解析後では、ほとんどの被験者が高用量群に割付けられるような割付比に変化し

ている様子が示されている[3]。有効性が高い用量群への割付比が増えることは、当該用量とプラセボ群との比較に対する検出力を高めるとともに、安全性評価の例数も増やすことができる。また、より有効な用量群に割付けられる確率が高くなることは患者にとってのベネフィットと考えられ、試験への登録を促進することにつながる可能性もある。一方、RARは有効性の低い用量群に割り付けられる症例数が少なくなる可能性があり、用量反応関係を特定することが難しくなる可能性があることに留意する必要がある。

## 2) 無益性の評価

被験薬の有効性がない場合に試験を早期中止する判断を行うために、中間解析での無益性の評価が計画された。無益性の評価は、患者が有効ではない治療を受けるリスクを低減するとともに、試験が失敗した場合のコストを削減し、患者やSLEの治験経験を有する施設が、有効性を示す可能性のある他の開発プログラムに参加できるようにすることで、SLEに対する治療の開発効率を高めることができる[3]。

無益性の基準は、52週時のSRI4反応率の各用量群とプラセボ群との差が目標とする治療効果を上回る確率の最大値がある閾値未満になる場合と設定された[4]。主要評価項目である52週時のデータを十分に蓄積するためには時間を要することから、これまでSLEの用量設定試験で用いられてきた16、20、24週時のデータから経時モデルによって52週時のデータが得られていない患者の反応を予測することにより、中間解析の時期が約6ヵ月前倒しされた[3]。

## 3) Administrative success

最終の中間解析では、第IIb相試験の実施には影響を及ぼさないが、第IIb相試験と第III相試験の間の期間を削減するために、第III相試験の準備開始を判断するための評価が実施された[3]。Administrative successの基準は、モデルに基づいた仮想的な第III相試験において、最終解析で片側有意水準2.5%の検定を行った場合に成功する確率があるカットオフ値よりも大きくなる場合と設定された[4]。

## 4) ベイズ流階層モデルによる用量群間の情報借用

無益性の中間解析および最終解析での52週時のSRI4反応率の各用量群とプラセボ群との比較には、ベイズ流階層モデルが用いられた。一般的に、用量が高くなるほど大きな有効性が得られると仮定されることが多いが、SLEをはじめとする炎症性疾患では、単調な用量反応関係が観察されないことがある[3]。ベイズ流階層モデルでは、用量反応に関する事前の仮定を置くことなく、観察された用量群間の治療効果の類似性に基づいて、用量群間で動的に情報を借用することができる。用量群間の反応率が類似していればいるほど、より多くの借用が行われ、逆に用量群間で反応率が大きく異なる場合は、借用はほとんど行われぬ。したがって、各用量群が同じような反応率である場合、ある用量群の治療効果の推定に、同様の効果を持つ他の用量群のデータの情報も活用されることで、治療効果の推定精度および検出力が向上する。これにより、試験全体の被験者を減らすことができる可能性がある。

本試験で用いられたベイズ流階層モデルでは、用量群 $d \in \{low, medium, high\}$ の52週時のSRI4反応例数 $X_d$ が、以下の二項分布でモデル化された[4]。

$$X_d \sim \text{Binomial}(p_d, N_d)$$

ここで、 $p_d$ は用量群 $d$ の52週時の SRI4反応率、 $N_d$ は用量群 $d$ の例数である。そして、各用量群の対数オッズに対応するパラメータ $\alpha_d$ に対して、階層事前分布が用いられた。

$$\log\left(\frac{p_d}{1-p_d}\right) = \alpha_d$$

$$\alpha_d \sim N(\alpha_{\text{treatment}}, \sigma^2)$$

ここで、 $N(\alpha_{\text{treatment}}, \sigma^2)$ は、平均 $\alpha_{\text{treatment}}$ 、分散 $\sigma^2$ の正規分布である。 $\sigma^2$ が用量群間の反応の類似性の程度を表しており、その値が小さいほど反応が類似していることを意味する。

#### 【シミュレーションによる動作特性の評価】

アダプティブデザインに関する FDA ガイドラインでは、有効性の実質的なエビデンスを提供することを目的とした臨床試験が満たすべき、4つの主要な原則（1.誤った結論を導く可能性を適切に制御すること、2.治療効果の推定を十分に信頼できるものとする、3.デザインの詳細を完全に事前に規定すること、および4.試験の完全性を適切に維持すること）が述べられている[5][6]。1つ目および2つ目の原則が満たされていることを確認するため、シミュレーションによる試験デザインの動作特性の評価が行われた。以下に、実施されたシミュレーションの概要を説明する。なお、3つ目および4つ目の原則に関しては、アダプティブデザインの詳細は事前規定され、プロトコル、統計解析計画書、IDMC チャーター、シミュレーション報告書に詳述された。また、中間解析の実施、割付比の変更、中間解析結果のレビューなどは、治験依頼者を含む治験関係者とは独立な IDMC を用いて盲検性を維持したまま実施することによって、中間解析結果の機密性および試験の完全性が維持された[3]。

1つ目の原則に対して、第一種の過誤確率が5%以下に制御されていることを確認するために、いずれの用量群の反応率もプラセボ群と同じというシナリオのもとで、シミュレーションによる評価が行われた。プラセボ群の反応率、登録スピードなどの複数の試験結果に影響を及ぼす項目のもっともらしい設定を組み合わせた場合の広範なシミュレーションが実施され、提案された試験デザインは、これらの設定の範囲において、第一種の過誤確率を5%以下に制御できることが確認された。また、用量反応関係は線形で高用量群のみ目標の治療効果があるといったシナリオや、低用量群および中用量群はプラセボ群と差がなく、高用量群のみ目標の治療効果があるといったシナリオなど、7通りの用量反応関係のシナリオのもとで、第二種の過誤確率の評価が行われた。提案された試験デザインの成功確率（検出力）は、中間解析や RAR を行わない従来の試験デザインと比べて高く、より少ない症例数で試験を実施可能であることが確認された[3]。2つ目の原則に対して、シミュレーションの各シナリオで仮定した真の治療効果とベイズ流階層モデルによって推定された治療効果を比較することにより、バイアスおよび平均二乗誤差が評価された。シミュレーションの結果から、十分に信頼できる治療効果の推定値が得られることが確認された[3]。その他、最終的に各用量群にランダム化された平均症例数、試験完了までの時間、正しい用量を選択する確率、無益性の基準を満たす確率、administrative success の基準を満たす確率など、様々な試験の動作特性がシミュレーションによって評価された[3][4]。

**【FDA との議論】**

当初に計画されていた試験デザインの以下の特徴が革新的であり、CID pilot meeting program のもとでレビューすることが適切と判断された（RAR 以外は FDA との協議で削除）。

- 中間解析に基づいて主要評価項目の変更を可能とするアダプテーションルールの使用
- 主解析のプラセボ群との比較において、異なる用量のデータをプールすることを可能とするアダプテーションルールの使用
- 実薬用量群間の割付比を変更可能とする RAR の使用

また、用量設定試験と検証的試験の両方として位置付けられる可能性があり、これらを別々に行う場合に比べて、時間とリソースが節約できること、より多くの患者を最も効果がある用量群に割付ける傾向があること、最も効果がある用量群とプラセボ群との比較における検出力を増加させることができること、などが試験デザインの利点として評価されている。

提案された試験デザインに関する主な論点として、以下が挙げられている。

- 治療群間の患者背景の比較可能性に対する RAR の影響
- 真の用量反応関係の仮定およびそれ以外の試験結果に影響を及ぼす項目のもっともらしい設定を組み合わせた場合における試験デザインの動作特性の適切性
- Arm-dropping アプローチと比較した RAR の性能
- 主解析において、競合する従来の方法と比較したベイズ流階層モデルの性能
- 中間解析結果の機密性および試験の完全性を維持するための手順

上述したように、広範なシミュレーションによる試験の動作特性の評価が行われた。FDA との議論を通じて、シミュレーションの変更および追加が行われた。例えば、初期のシミュレーションでは、真の用量反応関係以外の試験結果に影響を及ぼす項目（真のプラセボ群の反応率、登録スピード、反応の被験者内相関、データカットオフからアダプテーションまでの時間ラグ）の数が多く、包括的な評価を行うことが困難であったため、主要評価項目を変更するアダプテーションルールの削除、データカットオフからアダプテーションまでの時間ラグの固定といった変更が行われた[1]。

従来の試験デザインとの比較によって CID の実用性を示すことは有益であるとされ、RAR と arm-dropping アプローチとの比較、ベイズ流階層モデルと多重性を制御する方法（用量が高いほど有効性も高いことを仮定した階層的検定手順や Dunnett 検定など）との比較が行われた[1]。

統計的な問題に対処することに加え、臨床的／科学的な問題も適切に考慮することが重要であるとされ、主要評価項目の変更および主解析のプラセボとの比較において異なる用量のデータをプールすること許容するアダプテーションルールは最終的に削除された[1]。

登録期間の初期と後期に参加した患者の疾患の重症度などといった特徴が異なる場合、RAR は治療群間の患者背景の不均衡を生じさせる可能性が懸念されたが、アンメットニー

ズの高い希少疾患であることから、このリスクは許容できると判断された[1]。

#### 【まとめ】

本事例では、適切な治験実施施設や適格な患者が限られている SLE の臨床試験において、試験の効率および成功確率を向上させ、早期の意思決定を行うためにアダプティブデザインが用いられた。FDA との協議を経て、本試験デザインは適切な動作特性を有し、また従来の早期臨床試験のような24週間ではなく、52週間のエンドポイントを採用したことなどから、承認に必要な2つの検証的試験のうちの1つとして位置づけられる可能性があると考えられた[3]。本事例から、アダプティブデザインは、適切に導入すれば試験の完全性を損なうことなく、効率的な開発を行える可能性があると考えられる。また、アダプティブデザインに関する規制当局との議論では、試験の動作特性の適切性、結果の臨床的意義、試験の完全性の維持を示すことが重要であると考えられた。

#### <参考文献>

- [1] Food and Drug Administration. "CID Case Study: A Study in Patients with Systemic Lupus Erythematosus" <https://www.fda.gov/media/155404/download?attachment>
- [2] Kaul A, Gordon C, Crow MK, Touma Z, Urowitz MB, van Vollenhoven R, Ruiz-Irastorza G, Hughes G. Systemic lupus erythematosus. Nat Rev Dis Primers. 2016 Jun 16;2:16039.
- [3] Garces Sandra et al. "Improving resource utilisation in SLE drug development through innovative trial design." *Lupus Science & Medicine* 10.2 (2023): e000890.
- [4] Amy Xia, May Mo, and Tony Jiang. "Delivering Transformative Medicines to Patients by Leveraging Complex Innovative Designs AMG 592 SLE Case Study." Presentation [http://www.bayesianscientific.org/wp-content/uploads/2022/08/CID\\_BSWG\\_AUG2022.pdf](http://www.bayesianscientific.org/wp-content/uploads/2022/08/CID_BSWG_AUG2022.pdf)
- [5] Food and Drug Administration. "Adaptive designs for clinical trials of drugs and biologics-Guidance for Industry." (2019).
- [6] 日本製薬工業協会. "アダプティブデザインに関する FDA ガイダンスの邦訳." (2021) [https://www.jpma.or.jp/information/evaluation/results/allotment/adaptive\\_design.html](https://www.jpma.or.jp/information/evaluation/results/allotment/adaptive_design.html)
- [7] 厚生労働省. 令和3年度 衛生行政報告例 <https://www.nanbyou.or.jp/wp-content/uploads/2023/02/koufu20221.pdf>
- [8] 日本製薬工業協会. "Rare disease の治療効果の推測法 - CID パイロットプログラムとその実例 -". 「Rare disease の臨床開発における治療効果推測の現状と課題」シンポジウム(2023) [https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS\\_202302\\_cid.pdf](https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS_202302_cid.pdf)

#### 4.7 びまん性大細胞型 B 細胞リンパ腫の第 III 相試験

本項では、CID のウェブサイトにて CID pilot meeting program に採択された試験の1つとして公開されているびまん性大細胞型 B 細胞リンパ腫（DLBCL）を対象とした臨床試験の事例 [1]を紹介する。

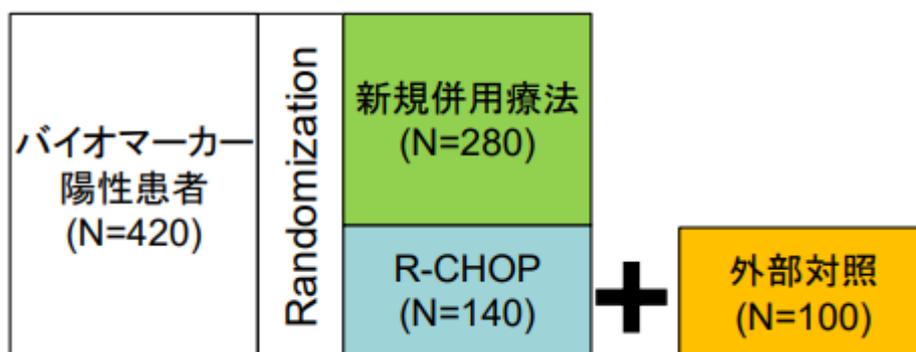
##### 【びまん性大細胞型 B 細胞リンパ腫】

DLBCL は全ての非ホジキンリンパ腫のうち3-4割を占める最も発生頻度の高いリンパ腫病型であり、米国では年間約25000名の新規患者がいる [2]。本例数は米国では希少疾患に該当し、大規模臨床試験での開発が困難であることが推察される。DLBCL の一次標準治療は開発されて20年以上経過しており、治療ニーズの高い疾患領域である。

##### 【試験デザイン】

CID pilot meeting Program に採択された DLBCL 試験は、特定のバイオマーカーが陽性である DLBCL の一次療法を対象とする第 III 相非盲検ランダム化比較試験である（図 4.7.1） [3]。ランダム化は試験群と対照群に対して2:1の比率で実施され、主要評価項目である無増悪生存期間（PFS）が検証されるように、ハザード比0.6、有意水準5%、検出力80%のもとで症例数が設計されている。ただし、PFS イベントが80%集まった時点で1回の中間解析を実施予定である。本試験デザインでは、全生存期間（OS）を重要な副次評価項目として設定している [4]。

図 4.7.1 試験デザイン



[出典：日本製薬工業協会 (2023) [3]のスライド 8 の図]

本試験の革新的または新規とみなされた要素として以下の3点が挙げられる [4]。

- 重要な副次評価項目である OS の比較に外部対照データを利用
- 外部データの借用に Commensurate prior を利用
- 重要な副次評価項目である OS の解析にベイズ流パラメトリックモデルを利用

1点目の外部対照データの利用は、PFS の主解析時点にて OS を早期に解析することに伴う、OS のイベント数の不足を補うことを目的としている。ただし、新規試験と外部対照データとの患者背景の不均一性を考慮して、傾向スコアマッチングにより比較に用いる外部対照データ（集団）を選択している。2点目の Commensurate prior の利用は、傾向スコアマ

ッチングにて選択された外部対照データを対照群と併合して用いるのではなく、新規試験と外部対照データの情報量の共通性に応じて、外部対照データの利用度を調整することを目的としている[4]。そして3点目のベイズ流パラメトリックモデルの利用は、Commensurate prior を利用するために、Time to Event データの解析にて使用される Cox 回帰モデルではなく、OS がワイブル分布に従うと仮定したパラメトリックモデルを用いてベイズ推定を行うことが計画されている。以上の3点から、FDA は当該試験を革新的なデザインと評価している。

上記の革新的な要素を含めた試験デザインを用いることで、OS がワイブル分布に従うというベイズ流パラメトリックモデルを利用できる仮定が満たされる場合に、通常のランダム化比較試験と比較して、名義水準の検出力を満たしつつ対照群に登録される症例数を減らす利点が挙げられる。加えて、Commensurate prior を利用し、新規試験の結果に応じて外部対照データの情報量の利用度を調整することで、本試験と互換性のない外部データを借用するリスクを軽減させることもデザイン上の工夫として挙げられる[1]。

#### 【数理的背景】

本試験の OS の解析で用いられたベイズ流パラメトリックモデルでは、各被験者*i*の OS を  $y_i$  ( $i = 1, \dots, N$ )と表記し、以下のワイブル分布でモデル化することを計画している[5]。

$$y_i \sim \text{Weibull}(r, \exp(\alpha + \beta z_i))$$

$r$ 、 $\alpha$ 、 $\beta$ はワイブル分布を構成するパラメータである。 $z_i$ は被験者*i*に割り付けられた群を表すダミー変数であり、新規併用療法群の時に1、対照群の時に0をとる。本試験では、新規試験と外部対照の二つのデータを用いてワイブル分布を構成するパラメータを推測することを計画している。一方で、新規試験と外部対照データのそれぞれの被験者の OS の従う分布が異なることを想定している。そこで、外部対照データの各被験者の OS を  $y_{1i}$ 、新規試験の各被験者の OS を  $y_{2i}$  と表現し、 $y_{1i}$ 、 $y_{2i}$ のそれぞれを以下のワイブル分布でモデル化する。

$$y_{1i} \sim \text{Weibull}(r, \exp(\alpha_1))$$

$$y_{2i} \sim \text{Weibull}(r, \exp(\alpha_2 + \beta z_i))$$

パラメトリックモデルに基づく推測では、ワイブル分布の確率密度関数と各被験者の OS の打ち切りの情報を用いて部分尤度関数を導出することで、最尤推定にてワイブル分布を構成するパラメータを推定する。本試験デザインでは、ベイズ流パラメトリックモデルを用いることから、ワイブル分布を構成するパラメータの事後分布に基づきパラメータの推測を行う。各パラメータの事前分布を  $P(r)$ 、 $P(\alpha_1)$ 、 $P(\alpha_2)$ 、 $P(\beta)$ と表現すると、各パラメータの事後分布  $P(\alpha_1, r | Y_1)$ 、 $P(\alpha_2, \beta, r | Y_2)$  は以下の式で表現される。ただし、 $L(\alpha_1, r | Y_1)$ 、 $L(\alpha_2, \beta, r | Y_2)$ は外部対照データ、新規試験データに基づく部分尤度関数を表し、 $\alpha$ は左辺が右辺に比例することを表現する。

$$P(\alpha_1, r|Y_1) \propto L(\alpha_1, r|Y_1)P(\alpha_1, r)$$

$$P(\alpha_2, \beta, r|Y_2) \propto L(\alpha_2, \beta, r|Y_2)P(\alpha_2, \beta, r)$$

上記の式では、各試験データに基づきそれぞれのパラメータを推測している。一方で、本試験デザインでは、新規試験データに基づくパラメータの推測の際に、外部対照データを利用して効率的に推測することを目指す。外部対照データを利用してパラメータを推測する Commensurate prior を利用した場合の各パラメータの事後分布は以下の式で表現される。

$$P(\alpha_1, \alpha_2, \beta, r, \tau^2|Y_1, Y_2) \\ \propto L(\alpha_1, r|Y_1)L(\alpha_2, \beta, r|Y_2)P(\alpha_2|\alpha_1, \tau^2)P(\alpha_1)P(\beta)P(r)P(\tau^2)$$

$\tau^2$ は新規データとヒストリカルデータの共通性の乖離度を表す指標であり、 $P(\alpha_2|\alpha_1, \tau^2)$ の部分が新規試験と外部対照データの違いを考慮している部分である。 $\tau^2$ の値が大きいほど新規試験と外部対照データが大きく異なると仮定したもとでパラメータが推測され、 $\tau^2$ の値が小さいほど新規試験と外部対照データが似ていると仮定したもとでパラメータが推測される。 $\tau^2$ の大きさによって推定される結果が異なることから、 $\tau^2$ の事前分布 $P(\tau^2)$ の設定に応じた結果の変動の程度を確認する必要がある。

#### 【FDA との議論】

FDA の報告に基づく、本試験のデザインに対する懸念として以下の点を挙げている[1]。

- 新規試験の実施時期と外部対照データが収集された時期の違いに伴う集団の違い(治療状況、Covid-19の発生状況など)
- ワイブル分布を仮定したパラメトリックモデルの当該疾患領域の過去データへの適合性
- 外部対照データの借用情報量を調整する Commensurate prior の利用の適切性、解釈可能性
- 傾向スコアモデルの利用の適切性、解釈可能性
- 統計解析手法の頑健性、第一種の過誤確率への影響の程度

いくつかの懸念点に対しては、試験デザインによる制御または臨床論文等で報告された情報に基づき説明できる。例えば、外部対照データの利用では、Pocock の基準を参照して利用する外部対照データの適切性を説明することが考えられる。本試験の場合、同一のスポンサーによる実施中のランダム化比較試験の対照群データを利用し、同一の適格・除外基準や類似した施設や医師の選択を行うことで、外部対照データを利用できるとの説明がされている。FDA のコメントでも、専門家の臨床的な意見、ヒストリカルデータ等の情報に基づき説明できる点があると述べられている。本試験デザインの例では、傾向スコアモデルの選択された共変量は、専門家の臨床的意見と文献に基づく臨床的根拠が必要であった。ワイブル分布の仮定は、この疾患領域の試験結果によってサポートされ、ワイブル分布と比較的適合することがシミュレーション実験以外の情報から推察されている。一方で、これらの情報に基づき説明が困難な状況については、シミュレーション実験に基づく評価および FDA と

の協議が必要になる。

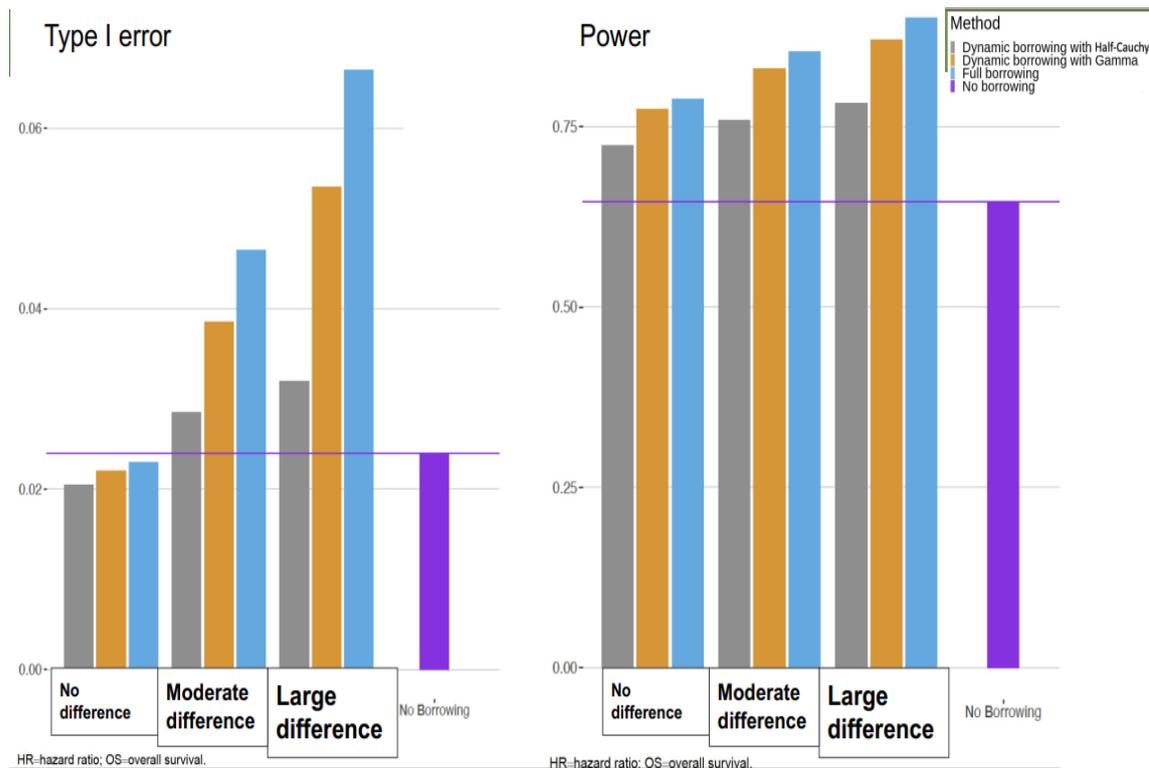
#### 【シミュレーション実験】

一つ目のシミュレーション実験では、対照群と外部データの OS 曲線の振る舞いに様々な差異がある場合の評価への影響を評価している[1]。具体的には、予後に影響する患者背景について、新規試験と外部対照データの患者背景分布の違いに伴う第一種の過誤確率および検出力への影響を評価している。ただし、新規試験と外部対照データの患者背景の違いの程度や違いの程度に応じた OS のハザードに関するシミュレーションシナリオの明記はなかった。加えて、外部データの利用の程度に応じて第一種の過誤確率および検出力が変化することから、以下の4つの場合の外部データの利用方針を評価している[4]。

- 選択された外部対照データを利用しない場合
  - 事後分布： $P(\alpha_2, \beta, r | Y_2) \propto L(\alpha_2, \beta, r | Y_2) P(\alpha_2) P(\beta) P(r)$
- 借用程度を調整する Commensurate prior で利用する  $\tau^2$  の事前分布 ( $P(\tau^2)$ ) を変えた場合
  - 半コーシー分布を利用する場合： $\tau^2 \sim \text{Half-Cauchy}(x_0, \gamma)$
  - ガンマ分布を利用する場合： $\tau^2 \sim \text{Gamma}(a, b)$
- 選択された外部対照データのすべての情報を借用する場合
  - 事後分布： $P(\alpha, \beta, r | Y_1, Y_2) \propto L(\alpha, \beta, r | Y_1, Y_2) P(\alpha) P(\beta) P(r)$

シミュレーション結果を図 4.7.2に示す[4]。新規試験と外部対照データの患者背景に違いがない場合 (No difference) には、外部対照データを利用しない場合と比較して、第一種の過誤確率はいずれの借用方法においても同等であり、検出力はいずれの借用方法も上回る傾向である。患者背景の違いがある場合 (Moderate difference、Large difference) では、違いの大きさに比例して、外部対照データを利用する手法では検出力が大きくなる一方、第一種の過誤確率は名義水準を大きく超える結果であった。本傾向は、外部対照データの情報を借用する程度が大きくなることで、同様な傾向が観察されている。

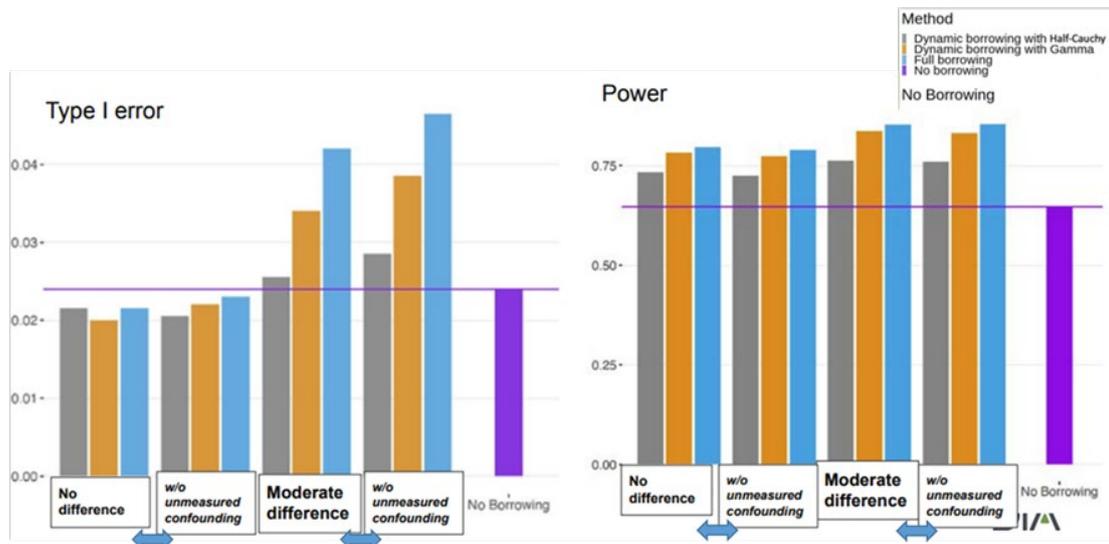
図 4.7.2 患者背景の違いの影響を評価したシミュレーション実験結果



[出典：Jiawen Zhu et al. (2022) [4]のスライド 26 の図]

上記のシミュレーション実験の結果に基づき FDA と協議した結果、複数の仮定からの逸脱を想定したもとの追加のシミュレーション実験が求められたことも報告されている。具体的には、OS の分布の違い、未測定の変数因子の有無の影響および傾向スコアマッチングで利用した傾向スコアモデルのモデルの形状（線形、非線形、Non-Additive model）の違いによる、第一種の過誤確率および検出力の影響をシミュレーション実験にて評価することが求められている。上記のうち、未測定の変数因子の有無の影響を評価するためのシミュレーション実験結果を図 4.7.3 に示す[4]。図 4.7.3 の「w/o unmeasured confounding」は未測定の変数因子がない場合（図 4.7.2 の結果と一致）を示し、「No difference」や「Moderate difference」は未測定の変数因子がある場合を示している。本シミュレーション実験では、新規試験と外部対照データの患者背景の違いがない場合（No difference と w/o unmeasured confounding の比較）と違いが中程度ある場合（Moderate difference と w/o unmeasured confounding の比較）における、未測定の変数因子の有無による第一種の過誤確率と検出力の比較が行われた。シミュレーション結果より、未測定の変数因子の有無に寄らず第一種の過誤確率と検出力は同程度であった。

図 4.7.3 未測定の変絡因子の有無の影響を評価したシミュレーション実験結果



[出典：Jiawen Zhu et al. (2022) [4]のスライド 29 の図]

以上のシミュレーション結果から、患者背景の違いを含む仮定からの逸脱の程度に応じて、第一種の過誤確率の考慮が必要であることが明らかになった。そこで、外部データの利用方法の決定に際して、想定される仮定からの逸脱の発生頻度に基づく第一種の過誤確率に関するシミュレーションが実施され、そのシミュレーション結果を表 4.7.1 に示す。結果より、第一種の過誤確率の平均および重み付き平均の値が名義水準以下に抑えられ、かつ外部データを利用しない場合よりも検出力の高い Commensurate prior に半コーシー分布を利用する手法が外部データの利用方法として望ましいと評価されている[4]。

表 4.7.1 手法間の第一種の過誤確率の要約

手法		第一種の過誤確率の平均	第一種の過誤確率の重み付き平均*	第一種の過誤確率の最大値
外部対照データを利用しない場合		0.024	0.024	0.024
Commensurate prior で利用する場合	半コーシー分布を利用する場合	0.023	0.023	0.032
	ガンマ分布を利用する場合	0.028	0.026	0.054
外部対照データのすべての情報を借用する場合		0.033	0.029	0.067

※重み付き平均は、仮定した患者背景の違いの発生確率をもとに計算する。患者背景に違いがない場合：62.5%、中程度の違いの場合：20%、大きな違いの場合：5%、逆比率<sup>12</sup>の中程度の違いの場合：10%、逆比率の大きな違いの場合：2.5%。

[出典：Jiawen Zhu et al. (2022) [4]のスライド 30 の表の改変]

その他の追加のシミュレーション実験にて、傾向スコアを OS の解析に用いるワイブル分布の共変量として使用できるかどうかを検討された。FDA は傾向スコアを共変量として使用した場合に結果の解釈および公表が難しくなると考えた。スポンサーは、FDA の意見を踏まえて、患者背景の違いを調整するための方法として、傾向スコアマッチングの利用に至った[1]。

#### 【まとめ】

希少疾病である DLBCL において、外部対照データを利用し、OS にワイブル分布を仮定したベイズ流パラメトリックモデルで解析する事例を紹介した。本事例では、新規試験と外部対照データの患者背景に違いがある状況を基本とし、OS の解析にて使用したベイズ流パラメトリックモデルの仮定からの逸脱を考慮したシミュレーション実験が実施されている。試験デザインによって必要なシミュレーション実験は変化するため、規制当局との事前の協議にて解決すべき課題を見出す必要がある。さらには、合理的な試験デザインの根拠とデザインに対応するシミュレーション実験は、試験デザインの妥当性の説明に重要であるが、得られる推定値等の結果の検討も試験デザインの利点と制約を理解するために必要である。

<参考文献>

[1] Food and Drug Administration “CID Case Study: External Control in Diffuse B-Cell

<sup>12</sup> タスクフォースによる解釈：逆比率に関するシナリオは、中程度の違いや大きな違いがある場合を仮定したシナリオに対して、新規試験と外部対照データにおける予後に影響する患者背景の分布の比率を逆転させた場合を示している。

- Lymphoma” <https://www.fda.gov/media/155405/download?attachment>
- [2] Susanibar-Adaniya S, Barta SK. 2021 Update on Diffuse large B cell lymphoma: A review of current data and potential applications on risk stratification and management. Am J Hematol. 2021 May 1;96(5):617-629
- [3] 日本製薬工業協会.“Rare disease の治療効果の推測法 - CID パイロットプログラムとその実例 –“. 「Rare disease の臨床開発における治療効果推測の現状と課題」シンポジウム(2023)  
[https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS\\_202302\\_cid.pdf](https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS_202302_cid.pdf)
- [4] Jiawen Zhu and Herb Pang. “Case Study of a Phase 3 Study Design with Hybrid Control in 1L DLBCL – FDA Complex Innovative Designs Pilot Program”  
<http://www.bayesianscientific.org/wp-content/uploads/2022/09/Bayesian-KOL-session.pptx.pdf>
- [5] J. Normington, J. Zhu, F. Mattiello, S. Sarkar, B. Carlin. “An efficient Bayesian platform trial design for borrowing adaptively from historical control data in lymphoma” Contemp. Clin. Trials, 89 (2020)

#### 4.8 ミオクロニー脱力発作を伴うてんかんの臨床試験

本項では、CID のウェブサイトにて CID Paired Program Trial Design Case Studies の一つとして公開されているミオクロニー脱力発作を伴うてんかん (EMAS) 患者を対象とした臨床試験の事例[1]を紹介する。

##### 【ミオクロニー脱力発作を伴うてんかん】

EMAS は、乳幼児期に発症する小型運動発作 (ミオクロニー発作、失立発作、脱力発作など) を伴うてんかん群のうち、Doose らが提唱した、正常発達幼児における遺伝性素因を背景としたミオクロニー脱力発作を主徴とする疾患である。日本では難病に指定されており、痙攣性疾患の家族歴が高頻度で遺伝性疾患が想定されているが、原因はまだ不明である[2]。

##### 【試験デザイン】

本試験は、EMAS を有する小児および青年患者を対象とした多施設共同二重盲検ランダム化プラセボ対照並行群間比較試験<sup>13</sup>である。主要評価項目は、投与期間中の EMAS 関連発作頻度である。

---

<sup>13</sup> タスクフォースによる補足：レノックス・ガストー症候群、ドラベ症候群、または結節性硬化症に対して2018年に FDA で承認されたカンナビジオール内用液を用いた試験と考えられる[4]。

EMAS 集団における治療の必要性、登録の困難性、および関連する適応症<sup>14</sup>間での一貫した治療効果に基づき、治験依頼者は EMAS を対象集団とした本試験の主要解析に、関連する適応症を対象集団とした試験結果を組み込むためのベイズ法を提案した。全体の治療効果について階層ベイズモデルを規定し、各試験固有の治療効果は全体の治療効果を中心とする正規分布を示すとした。関連する対象集団の過去の試験からのデータを動的借用するために、分布の分散パラメータに対する事前分布を規定する。過去のデータと比較して観察された EMAS データの類似性が借用の程度に影響する。

症例数は事前に規定したアルゴリズムに従い、ベイズ予測確率を用いた Goldilocks 法に基づく非盲検中間解析<sup>15</sup>により更新する。

- Goldilocks 法 [3]

「この症例数は大きすぎるのか、小さすぎるのか、それともちょうどいいのか?」という質問を常に投げかけているため、Goldilocks<sup>16</sup>法と呼ばれている本手法は、蓄積されたデータに基づいて試験の症例数を選択するベイズ流の適応的アルゴリズムである。

Goldilocks 法の試験デザインでは、無益性による早期中止だけでなく、成功が予測される場合の早期有効中止も可能となる。このアルゴリズムは2つの予測確率  $P_n$  および  $P_{max}$  に基づいている。 $P_n$  は現在の症例数で試験成功となる予測確率であり、 $P_{max}$  は最大の症例数まで試験が継続された場合の試験成功となる予測確率である。

症例数の最小値と最大値、症例数選択の解析頻度、早期有効中止基準 ( $S_n$ )、および早期無益性中止基準 ( $F_n$ ) は事前に指定する必要がある。 $n$  例の被験者を用いた症例数選択の解析で、試験成功の予測確率  $P_n$  が早期有効中止基準  $S_n$  より高い場合 ( $P_n > S_n$ )、新たな被験者登録は中止され、 $n$  例で主要解析が実施される。一方、最大症例数での試験成功の予測確率  $P_{max}$  が早期無益性中止基準  $F_n$  より低い場合 ( $P_{max} < F_n$ )、試験は無益であると判断され早期中止に至る。いずれの早期中止基準も満たさない場合、被験者の登録は継続され、次の症例数選択の解析のために追加情報が収集される。

【試験デザインに対する検討事項】

- 情報を借用した場合どのような影響があるか
- 提案された手法の借用は適切で解釈可能か
- 提案するデザインはモデルの仮定からの逸脱に対して頑健であるか
- モデルの仮定からの複数の妥当な逸脱の下での統計的特性およびデザインの性能は

---

<sup>14</sup> タスクフォースによる補足：レノックス・ガストー症候群、ドラベ症候群、結節性硬化症が関連する適応症と考えられる[4]。

<sup>15</sup> タスクフォースによる補足：データサイエンス部会「アダプティブデザインに関する FDA ガイドランスの邦訳」の V. 比較データに基づくアダプティブデザイン B. 症例数に対するアダプテーションに該当する。

<sup>16</sup> タスクフォースによる補足：英国の童話「ゴルディロックスと3匹のくま」に登場する少女ゴルディロックス (Goldilocks) がくまの家で飲んだ熱すぎず冷たすぎない、ちょうど良い温かさのスープにちなむ。

どのようなものか

- Goldilocks 法を用いたアダプテーションが動的借用の設定において動作特性に及ぼす影響はどのようなものか

#### 【シミュレーションによる動作特性の評価】

- 1) 広範囲の事前情報分布を調査し、許容可能な動作特性を示す単一の事前分布を事前を選択する。
- 2) 選択したデザインの動作特性を、小さい症例数で外部データをベイズ借用するという状況において、治療効果とエンドポイントのばらつきに関する広範な仮定の下で評価する。
- 3) 単一の仮想研究のためのベイズ借用の挙動を調べる、すなわち、単一の観察された EMAS データセットにおける結果の解釈に、借用がどの程度影響し得るかを理解する。
- 4) 提案した Goldilocks 法の性能を評価する。

#### 【FDA との議論】

FDA は本試験のデザインにおいて以下の特徴が革新的であると考えている。

- 異なる適応症を対象とした同一薬剤の過去の試験から得られた治療効果を情報借用する点
- 症例数を決定するためにベイズ予測確率を用いた Goldilocks 法を適用する点

またデザインの潜在的利点として、①情報借用と適応的症例数により、治験に必要となる症例数を減らすことができ、プログラムをより実現可能なものにすることができ、②動的な借用アプローチにより、本試験で観察されたデータと一致しない患者データを借用するリスクを軽減できる可能性があることが挙げられている。

治験依頼者が過去の試験で観察されたデータに基づく一連の仮定を用いて予備的なシミュレーションを実施したところ、FDA は多様なシナリオおよびモデルの仮定からの逸脱の下で動作特性を評価するための追加のシミュレーションを要請している。また FDA は、①治療効果の分散パラメータに対する事前分布がどのようにして借用量を動的に決定するか、②Goldilocks 法の動作特性についても明確化を求めている。

ベイズ流階層モデルの基礎となる重要な仮定は、異なる試験間の治療効果の交換可能性である。すなわち、ある一連の治療効果がこれらの試験で観察される確率は、試験の順序を変えても不変ということである。この仮定は臨床的根拠および過去の試験で認められた結果の一貫性に基づき妥当であると考えられる。

FDA は事前データに矛盾がある場合に、事前情報分布の異なる選択肢や借用量を減らすためのその他のアプローチを探索することを治験依頼者に提案している。治験依頼者は、より大きな名目有意水準での頻度論的検定、Bayesian power prior および階層ベイズモデルの異なるバージョンを含む他の方法を検討し、提案している2段階階層モデル（関連する適応症間で治療効果が一貫していると仮定したモデル[4]）の動作特性がそれらの動作特性と類似することを示した。加えて治験依頼者は、事前データの矛盾がある場合におけるベイズ流の有効症例数を探索し、最悪のシナリオの下で情報を借用した場合の影響を示した。その際、

FDA は結果解釈の頑健性を検討可能にする感度分析を事前に規定する必要があることを強調している。また治験依頼者は、提案する Goldilocks 法の動作特性への影響を理解する一助とすべく、ベイズ流解析と適応的症例数の影響を分離する代替戦略を検討した。

<参考文献>

- [1] Food and Drug Administration “CID Case Study: A Study in Patients with Epilepsy with Myoclonic-Atonic Seizures” <https://www.fda.gov/media/172312/download?attachment>
- [2] 難病センター: <https://www.nanbyou.or.jp/entry/4417>
- [3] Kristine R. Broglio, Jason T. Connor & Scott M. Berry (2014) Not Too Big, Not Too Small: A Goldilocks Approach to Sample Size Selection, Journal of Biopharmaceutical Statistics, 24:3, 685-705, DOI:10.1080/10543406.2014.888569
- [4] Margaret Gamalo (2023) Extrapolation and Complex Innovative Designs, 計量生物セミナー

#### 4.9 その他

その他の CID 事例として、優先遺伝アルツハイマー・ネットワーク試験ユニット (DIANTU) が実施するプラットフォーム試験、マスタープロトコルを用いた進行期非小細胞肺癌患者を対象とする Lung-MAP 試験、ベイズ流アプローチを活用した早期乳がん患者を対象とする I-SPY 2試験などが挙げられる。いずれの事例についても、データサイエンス部会から2022年12月に報告されている「Rare disease の治療効果の推測法」のそれぞれ3.4.4節、補遺2で詳しく取り上げられているので参照されたい。

## 5. おわりに

本報告書では、FDA の CID に関連した取り組みおよび CID を採用した臨床試験の事例の詳細をまとめることで、革新的な要素や規制当局と議論となりうるポイントをまとめた。

従来の方法では実施が難しい臨床試験（例えば、希少疾患や小児を対象）に CID を採用することで、合理的な期間や症例数にて試験を実施できるようになることが期待される。一方で、CID は従来と比較すると試験デザインが複雑になるため、選択したデザインの動作特性をシミュレーションにより評価することが重要である。CID の計画と実施のためには、統計だけでなく、臨床、データマネジメント、薬事など、複数の部門の連携がより重要である。さらには、規制当局との議論が円滑に進められることも重要である。

本報告書の内容は作成時点での情報であり、まさにいま進捗している状況である。最新の情報は、CID のウェブサイトや各事例に関連する論文などを参照し、随時アップデートを頂きたい。本報告書が CID を知るきっかけとなり、理解を深める一助となれば幸いである。

## 補遺

### 1. Interacting with the FDA on Complex Innovative Clinical Trial Designs For Drugs and Biological Products の邦訳

#### I. はじめに

この文書は、医薬品または生物学的製剤の複雑な革新的試験デザイン (complex innovative trial design : CID)の提案に関して FDA とやり取りする際の治験依頼者および申請者へのガイダンスを提供する。FDA は、「21世紀の治療に関する法律」(治療に関する法律) 第3021条に基づく義務を部分的に満たすためにこのガイダンスを発行している。治療に関する法律の義務に従って、このガイダンスでは、医薬品および生物学的製剤の開発および規制審査における新規試験デザインの使用、モデリング&シミュレーションに関連する技術的問題について治験依頼者がフィードバックを取得する方法、および審査のために提出すべき定量的および定性的手法の種類について説明している。治療に関する法律の第3021条に定められた義務に関連する追加の推奨事項は、FDA ガイダンス : Adaptive Designs for Clinical Trials of Drugs and Biologics [1]で取り上げられている。

このガイダンスは、2019年9月付けの同タイトルのガイダンスのドラフトを最終化するものである。

このガイダンスを含む FDA のガイダンス文書は、法的強制力のある責任を持たない。代わりに、これらのガイダンスは、あるトピックに関する FDA の現在の考え方を説明するものであり、特定の規制または法的要件が引用されていない限り、推奨事項としてのみ見なされるべきである。FDA のガイダンスにおける「Should」\*という言葉の使用は、何かが提案または推奨されるが、必須ではないことを意味する。

#### II. 適用範囲

CID は、複雑なアダプティブ、ベイズ流、およびその他の新しい臨床試験デザインを指すと考えられているが、革新的または新規とみなされるものは時間の経過とともに変化する可能性があるため、CID に固定された定義はない。このガイダンスの目的には、CID には新薬申請または生物学的製剤認可申請において有効性の十分なエビデンスを提供するために、これまでほとんどまたはまったく使用されていない試験デザインが含まれている。CID には、複雑な試験デザインの特徴がすでに他の適応症で使用されていて、所与の適応症にそれらのデザインの特徴を新たに適用する場合も含まれる。多くの CID に共通する特徴は、試験的な動作特性を推定するために数式よりもシミュレーションが必要であることである (このガイダンスのセクション III)。新規または CID とみなされる可能性のある試験デザインのいくつかの例は、このガイダンスのセクション IV で説明されている。これには、試

---

\* 訳者注：本邦訳では、「should」と記載されている箇所は「～すべき」と訳した。

験内の対照群を拡張するためにヒストリカルまたは過去の情報を正式に借用するものや、先の試験から対照群のデータを借用するもの、連続多段階ランダム割付試験 (Sequential Multiple Assignment Randomized Trials)、またはマスタープロトコルが含まれる。

複雑で革新的なデザインは臨床開発のすべての段階で適用できるが、このガイダンスの主な焦点は、有効性の実質的なエビデンスを提供することを目的とした試験の CID 提案に関する FDA と治験依頼者の やり取りにある。ほとんどの場合、CID 提案に関連するやり取りは、新薬臨床試験開始申請 (IND) に関連して発生する。新しい臨床試験デザインでは、デザインの目的、実行、運用上の特徴 (誤った結論を生み出す可能性など)、試験データの解析やその提示方法など、デザインの側面について治験依頼者と FDA の間で明確なコミュニケーションが必要である。このガイダンスでは、そのような対話に関する推奨事項を提供する。また、このガイダンスでは FDA が CID とみなす可能性のある臨床試験デザインの例を示しており、治験依頼者と FDA の間の生産的な議論を円滑に進めるために、FDA が提案書とともに提出することを推奨する情報の種類についても説明している。さらに、このガイダンスでは臨床試験のデザインと計画におけるシミュレーションの役割についても取り上げている。

このガイダンスは、特定の新規デザインが規制上の使用に適しているかどうかを示すものではない。そのような決定は、デザインが提案されている理由、特定の設定における妥当性および特定の開発プログラムに依存してケースバイケースで行われる。ある適応症のある製品クラスには適切な CID 提案が、別の製品クラスや別の適応症には適切ではない可能性がある。このガイダンスでは、CID 提案に関する治験依頼者と FDA 間の効果的な対話のために推奨される要素に重点を置いている。

### III. CID 提案に関する FDA との協議

#### A) ミーティングの利用<sup>17</sup>

CID 提案には新たな科学的審査の考慮事項が含まれる可能性があるため、FDA は CID 提案を行った治験依頼者に対し、CID 計画の詳細に関して FDA との早期の対話を求めることを奨励する。一般に、治験依頼者は、臨床開発プログラムの過程で FDA とやり取りするための既存の相談形式を使用すべきである。これには、Type B End of Phase 2 meeting、Type C meeting、場合によっては新規デザインの要素を持つ初期臨床試験の pre-IND meeting などの正式な会議や IND 修正案の提出に関するルーチン的なフィードバックも含まれる[2]。FDA による CID 提案の審査には、多くの場合、専門分野や FDA Office の全体にわたる詳細な議論だけでなく広範囲のコンピューターシミュレーションを含む、デザインの動作特性のチャレンジングな評価も含まれる。このようなデザインに対する FDA 審査を効果的なも

---

<sup>17</sup> タスクフォースによる補足：現在は CID Paired Meeting Program が存在するため、ミーティングの利用に関してはその Program の枠組みも参照されたい(本ガイダンスは CID paired meeting program が制定される前に施行されている)。<https://www.fda.gov/drugs/development-resources/complex-innovative-trial-design-meeting-program>

のにするためには、ミーティングパッケージの審査に使用できる時間を考慮するべきである。たとえば、End of Phase 2 meeting には第 III 相試験での CID の使用に関する全体的な戦略についての議論が含まれる可能性がある一方で、Type C meeting はシミュレーション計画などの統計的な詳細についての議論に適しているかもしれない。

CID 提案がデザインの重要な部分を決定するために大規模なコンピューター シミュレーションを含む場合、FDA はより詳細なフィードバックを提供するために追加の Type C meeting の場を提供する可能性がある。治験依頼者は、ミーティングパッケージ内で、CID 提案が臨床開発プログラム全体にどのように適合するか、また、より単純なまたは従来の臨床試験デザインと比較して、プログラムや試験の効率、またはその結果の一般化可能性をどのように向上させることができるかを示すことが推奨される。

治験依頼者は、提案された CID について FDA の審査部門と追加の会議を行うために、複雑で革新的な試験デザインのための FDA のパイロットプログラム (CID Pilot meeting program) を検討することもできる[3]。CID Pilot meeting program は、プログラムに参加している治験依頼者に、CID 提案について FDA と面談し、審査チームや規制当局の上級意思決定者から詳細なフィードバックを得る追加の機会を提供する。このプログラムは2018年8月30日から2022年9月30日まで実施され、CID の使用を促進することを目的としている。CID Pilot meeting program への参加申請に関する詳細と、Pilot meeting program の全体的なプロセスについては、FDA のウェブサイト[4]を参照されたい。

## B) CID 提案にて推奨される共通事項

FDA が新規デザインの提案を徹底的に検討し、フィードバックを提供するには、詳細な文書が重要である。文書には、組み込まれる予定の新しい特徴、計画された実装のタイミングとその詳細、基本となる科学的な目的にデザインがどのように対応するかを含めるべきである。具体的な文書は提出される提案の種類によって異なるが、ほとんどの提案に含めるべき共通の要素がいくつかある。このセクションでは、該当する場合は常に、CID 提案に含めるべきである一般的な要素の例を示す。このガイダンスのセクション IV では、特定の種類の CID 提案にとって重要となる可能性のある追加情報の例を示す。

- 試験デザインの選択と、それが全体的な医薬品開発計画にどのように適合するかについての説明。多くの場合、特定の製品および適応症について、新規のデザインが従来の試験デザインに比べてどのような利点があるのかを説明することに役立つ。この説明には、動作特性の詳細な比較や、提案されたデザインが関連する従来のデザインよりも開発プログラムの目標をどのように満たしているかについての説明が含まれるであろう。
- 考えられるアダプテーションの計画、中間解析の実施の詳細、決定基準など、デザインの重要な側面の詳細な説明。
- 複雑で革新的なデザインの特徴に関連する統計的考慮事項についての説明。
- 事前情報が正式に借用されている場合は、事前情報の情報源と選択に関する詳細、提案された試験計画との関連性、および関連するすべての事前情報が考慮されている

ことを保証するために講じられた手順の説明。これにより正式に借用する事前情報が誤解を招く結果につながることをないようにする。

- ベイジアン フレームワークなど、外部情報がデザインに明示的に借用される場合、借用の理論的根拠と、事前情報から事前分布がどのように構築されたかについての説明。
- 誤った結論を生み出す可能性や治療効果の推定の信頼性など、デザインの動作特性の詳細な評価。該当する場合は、第一種の過誤確率の制御と検出力に対処すべきである。第一種の過誤確率が適用できない場合（外部情報を借用する一部のベイズ流デザインなど）には、それ以外の適切な試験の特性を考慮すべきである。
- ベイズ推定の場合、事前分布の選択に対する試験の動作特性の感度を評価することは有益である。
- 研究の動作特性を評価するためにシミュレーションが使用された場合のシミュレーション報告書。
- データモニタリング委員会または CID の重要な側面の実施を担当するその他の機関に割り当てられた決定事項、およびその決定に付随する指示事項（該当する場合およびデザインの理解に関連する場合）
- データアクセスを適切に制限し、試験の完全性（**trial integrity**）をどのように維持するかを説明するための包括的な計画。これについては、FDA の **Guidance for Industry on Adaptive Design for Clinical Trials of Drugs and Biologics**[1]で詳細に考察されている。
- 試験デザインまたは解析に患者の意見を利用する試験（例えば、エンドポイントの選択、許容できるリスク・ベネフィットのトレードオフ、最小限許容できるベネフィット、最大限許容できるリスク、安全性および有効性の治療効果の推定において目標とする確実性のレベル）については、患者の意見を評価するために用いた試験の治験実施計画書を協議のために提出すべきである[6]。

### C) ベイズ流の CID 提案での推奨事項

ベイズ流アプローチは、特に複雑なアダプテーションや予測モデルが使用される場合に、試験のデザインと解析に柔軟性を提供できるため、有効性の実質的なエビデンスを提供することを目的とした CID に適している可能性がある。さらに、ベイズ推定は、成人データを小児集団に外挿したり、第 II 相試験から対照データを借用して第 III 相試験を強化したりするなど、複数のエビデンスの情報源を体系的に組み合わせることが有利な状況で適切である可能性がある（このガイダンスのセクション IV を参照）。

治験依頼者がベイズ流の CID 提案の提出を選択した場合、FDA による提案の評価は、事前分布および主要評価項目・主要副次評価項目に関する試験での決定基準という2つの領域に関する治験依頼者と FDA 間の明確なコミュニケーションに依存する。試験外部の情報を形式的に借用するベイズ流アプローチは、すべての開発プログラムに適切であるとは限らない。外部情報の借用に関するベイズ流の CID 提案を提出する場合、治験依頼者は、借用した情報の選択においてどのようにバイアスを回避したかに関する具体的な詳細といった

借用の根拠を含めるべきである。このガイダンスの焦点は CID に関する FDA とのやり取りにあるため、次のセクションでは、治験依頼者と FDA の間の議論をより良いものにするためにベイズ流の CID 提案に含めるべきベイズ流 CID に特有の要素について説明する。

#### 事前分布

ベイズ推定は、誤った結論が生じる可能性を制御するために、十分な情報に基づいた事前分布の仕様に依存する。事前分布を慎重に設定しないでベイズ流手法を使用すると、誤った結論が得られる可能性が高くなる。そのため、事前分布に関する議論は、FDA によるベイズ流の提案の評価にとって特に重要である。事前分布は、先行する試験データや他の臨床データを含む様々な情報源から形成される可能性があり、また、同時データと比較して先行データの重みを軽くしたり、試験治療の大きな効果の可能性を最初から否定したりといった特徴も組み込まれる。一部の設定では、関心のあるパラメータに関する一般的な不確実性を反映するために、無情報事前分布または参照事前分布が使用される場合がある。

一般に、ベイズ流の CID 提案には、事前分布の詳細な議論が含まれているべきである。事前分布の形成に使用されたデータまたはその他の外部情報は、FDA が外部情報の情報源、完全性、その関連性、データの品質および信頼性を理解できるように詳細に提示されるべきである。外部データの関連性の1つの側面は交換可能性の問題に関連しており、ベイズ流の CID 提案ではこれに対応するべきである。交換可能性の概念については、FDA の *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials* [5] で詳細に説明されている。正確な推定を得るために、情報のある事前分布は一般に、治療効果の存在または大きさに関する初期の懐疑を示唆する可能性のあるエビデンスを含め、関連するエビデンスの徹底的な評価に基づくべきである。このため、ベイズ流の提案には、情報が選択的に取得または使用されないことを保証するために治験依頼者が講じた手順を説明する議論も含めるべきである。過去のデータの直接的な重み付けやその他の非データドリブンな特徴が事前分布に組み込まれている場合、提案にはこれらの特徴の使用と規模の理論的根拠が含まれているべきである。

#### 決定規則

頻度論的な統計的アプローチを使用して実施される多くの臨床試験では、製品の有効性の裏付けを確立するために、片側有意水準0.025での仮説検定が行われる。ベイズ流アプローチを使用する場合は、代替の決定基準を指定することが重要である。決定基準の選択は、試験デザインと推定の質の両方に大きな影響を与える。

決定基準を指定するにはさまざまな方法が考えられる。たとえば、治験実施計画書には、グループ A の奏効率がグループ B の奏効率よりも高い確率が99%を超えた場合に有効性の結論が支持されると記載されている場合がある（数学的表記では、 $\Pr(\pi_A > \pi_B) > 0.99$ ）。別の治験実施計画書では、有効性を検証するには、グループ A の奏効率がグループ B の奏効率より少なくとも10%高い確率が95%を超える必要があると記載されている場合もある（ $\Pr(\pi_A - \pi_B > 0.10) > 0.95$ ）。

治験依頼者は、製品が承認された場合、製品のラベル表示に含めることを意図したすべての主要評価項目および副次評価項目について、治験実施計画書で決定基準を提案するべき

である。これらの提案には、基準を選択する根拠を含めるべきである。FDA は IND 審査中にこれらの提案を評価し、最終決定は審査部門との合意に基づいて行われる。

一部のベイズ流デザインでは、シミュレーションを使用して、検出力と第一種の過誤確率の頻度論的な動作特性を推定することができる。このような場合、指定されたレベルで第一種の過誤確率の制御を提供するための決定基準を選択できる。

#### D) シミュレーション

CID 提案で共通する特徴は、試験の動作特性を推定するために、また中間解析の回数やタイミングなどのデザインパラメータを最適化するためにシミュレーションを使用することである。FDA の Guidance for Industry: Adaptive Design for Clinical Trials of Drugs and Biologics [1]では、特にアダプティブデザインの状況においてそのようなシミュレーションについて詳しく説明しており、申請書に含まれるシミュレーション報告書の内容と形式に関する推奨事項を提供している。一般的に、同じ推奨事項がさまざまな CID 提案に適用される。

有効性の実質的なエビデンスを提供することを目的とした試験の場合、FDA との会議で試験シミュレーションの計画について議論することが有利である。具体的には、End of Phase 2 meeting は、第 III 相試験の予備的な計画について検討できる開発プロセスの段階で行われることが多く、これらの予備計画は試験シミュレーションの必要性を示唆する場合がある。これらのシミュレーションを中心に End of Phase 2 の議論の一部を構成することは、治験依頼者と FDA がシミュレーションで検討するシナリオや、発生率や想定される対照群の結果などの基礎となる試験の仮定を検討するのに役立つ可能性がある。

### IV. CID の事例と協議を促進するために推奨される情報

このセクションでは、いくつかの特定の CID 要素について説明する。その目的は、治験依頼者と FDA が会議や申請審査の中でこれらのデザインに関して生産的な対話を行うことを促進するために、推奨情報の例を示すことを意図している。例は包括的なものではない。他にも多くの CID 要素が提案される可能性があり、そのうちのいくつかはやがて考案されるかもしれない。さらに、現在 CID と考えられているものが、将来的には通常の試験デザインになる可能性がある。

#### A) マスタープロトコール

マスタープロトコールデザインは、1つの包括的なプロトコールが複数の質問に答えるようにデザインされている状況で使用される[7][8]。CID Pilot meeting program (セクション III.A.) に基づいて審査のために受理された試験提案の1つは、複数の慢性疼痛症状に対して複数の介入を評価するようにデザインされた試験である。主要解析では、ベイジアン混合モデル反復測定アプローチを使用して、実薬療法とプラセボ間の痛みの数値スケールの差を評価する。この解析により、ある疼痛状態内のプラセボ群から患者情報を借用したり、同じ実薬群に対する異なる疼痛状態間の治療差に関する情報を借用したりすることが可能になる。さらに、外部の情報源からのデータを使用して、一部のモデル パラメータに関する情報のある事前分布を作成することもできる。

このデザインの最初の考慮点には、プラセボ効果の変動の可能性、異なる疼痛状態を持つ患者間での交換可能性の仮定、慢性疼痛試験でしばしば遭遇する欠測データの影響などが含まれていた。

#### B) 第 II 相試験のデータを第 III 相試験に活用

CID Pilot meeting program (セクション III.A.) に基づく審査に受理された複数の提案に共通する特徴は、先行試験からの情報を新しい試験のデザインと解析に活用していることである。これには、第 II 相試験からの対照データを活用して第 III 相試験の効率を向上させる試みが含まれる可能性がある。このシナリオでは、第 II 相試験からの対照群の結果のデータが、その後の第 III 相試験からの治療の差の推定に組み込まれる。第 III 相試験で対照群を増加するために第 II 相の対照群のデータを使用することは、第 III 相試験の合計症例数の減少に繋がる可能性がある。

第 II 相試験のデータを第 III 相試験に組み込むという CID 提案は、多くの複雑な問題を提起しており、治験依頼者と FDA の間で生産的なやり取りを確保するために徹底的な文書化と議論が求められる。このような提案には、試験手順、治療法、エンドポイントの類似性や、結果に影響を与える可能性のある患者の治療基準の経時的変化の有無など、第 II 相試験の患者集団と予想される第 III 相患者集団の臨床的比較可能性に関する議論が含まれるべきである。第 II 相試験のデータは、FDA がその品質を評価できるように十分に詳細に提示されるべきである。他の情報源（医学文献など）からの他のデータが第 II 相試験データと一致するかどうかを判断するための背景情報の評価を提案に含めることが重要である。借用のリスクは、第 II 相試験と第 III 相試験のデザインまたは実施の間の何らかの不均一性により、前相の対照群の奏効率が同時に実施している対照群より系統的に低い場合、先行試験のデータを借用することで誤って有効と結論する可能性が高まることである。このため、階層モデルの使用や、異質性が存在する場合に自動的に借入を軽減する他のアプローチなど、先行試験のデータと同時の第 III 相のデータ間の異質性を評価して対処するための戦略を含めるべきである。上述で議論したように、ベイズ流アプローチが使用される場合、提案には必要に応じて借用される第 II 相試験のデータの有効症例数とその借用方法を含む、決定基準と事前分布に関する詳細な議論が含まれるべきである。

同様のアプローチは、第 II 相試験以外の情報源から情報を借用する場合にも適用できる。CID パイロット プログラムの例には、デュシェンヌ型筋ジストロフィーの外來患者を対象とした低用量および高用量の試験治療に関するランダム化二重盲検プラセボ対照試験、小児多発性硬化症患者集団を対象とした被験薬と実対照薬を比較するランダム化二重盲検群逐次非劣性試験が含まれる。

#### C) 連続多段階ランダム割付試験 (SMARTs)

SMART は、アダプティブな介入の開発に情報を提供するようにデザインされている。アダプティブな介入とは、臨床的に重要な結果を最適化するために、患者の特性および/または継続的なパフォーマンス（反応、遵守など）に応じて、治療の種類および/または強度をいつ、どのように変更するべきかを規定する一連の決定規則である。SMART は複数の介入

段階で構成され、各段階はアダプティブな介入に含まれる重要な決定の1つに対応する。SMART では、患者は複数の段階に沿って進み、各段階でいくつかの治療選択肢のうちの1つにランダムに割り付けられる。

SMART 提案または同様のデザインにおいて FDA と議論および伝達すべき要素には、統計的な質問/仮説、段階の数、デザインに組み込まれた介入、中間反応カテゴリー、フローの明確な図示、および多重性を調整する方法（該当する場合）が含まれるべきである。

<参考文献>

- [1] Guidance for Industry: Adaptive Designs for Clinical Trials of Drugs and Biologics, November 2019, <https://www.fda.gov/media/78495/download>.
- [2] Draft Guidance for Industry: Formal Meetings Between the FDA and Sponsors or Applicants of PDUFA Products, December 2017\*, <https://www.fda.gov/media/109951/download>.
- [3] Food and Drug Administration, Federal Register Notice, Complex Innovative Designs Pilot Meeting Program, 83 FR 44274, August 30, 2018, <https://www.federalregister.gov/documents/2018/08/30/2018-18801/complex-innovative-designs-pilot-meeting-program>.
- [4] Complex Innovative Trial Designs Pilot Program, <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/ucm617212.htm>.
- [5] Guidance for Industry and FDA Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials, February 2010, <https://www.fda.gov/media/71512/download>.
- [6] Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders: Patient-Focused Drug Development: Collecting Comprehensive and Representative Input, June 2020, <https://www.fda.gov/media/139088/download>.
- [7] Woodcock, J and LaVange, L. M. (2017). Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. *New England Journal of Medicine*, 377: 62 – 70.
- [8] Draft Guidance for Industry: Master Protocols: Efficient Clinical Trial Design Strategies To Expedite Development of Oncology Drugs and Biologics, October 2018\*, <https://www.fda.gov/media/120721/download>.

\*最終化された時点で、当ガイダンスは当トピックにおける FDA の現在の考え方を表すことになる。

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

2023年度タスクフォース3-2

“STAT 新時代 効率的試験デザイン法の整理”

タスクフォースメンバー

氏名	所属
相澤 愛奈	ファイザーR&D 合同会社
伊庭 克拓	大塚製薬株式会社
今井 翔悟	旭化成ファーマ株式会社
加倉井 靖之	第一三共株式会社
こう けんせい	日本ベーリンガーインゲルハイム株式会社
高戸 俊一	杏林製薬株式会社
たぎし 和隆	千寿製薬株式会社
田中 優一	大鵬薬品工業株式会社
東川 正晃	EA ファーマ株式会社
兵頭 洋平	住友ファーマ株式会社
平井 岳大	アストラゼネカ株式会社
松田 裕也	中外製薬株式会社
棚瀬 貴紀*	大鵬薬品工業株式会社
白石 亜矢子*	ヤンセンファーマ株式会社
菅波 秀規**	興和株式会社

\* タスクフォース推進委員、\*\* データサイエンス部会担当副部長