

アダプティブデザインに関する
FDA ガイダンスの邦訳

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

2021 年度 継続タスクフォース 2

2021 年 8 月

はじめに

アダプティブデザインは現在に至るまで、国内外の統計関連の学会や論文誌上などの場面で議論されてきました。アダプティブデザインに関する規制当局のガイダンスとしては、European Medicines Agency から「Reflection Paper on Methodological Issues in Confirmatory Clinical Trials with Flexible Design and Analysis Plan」が 2007 年に、米国食品医薬品局 (Food and Drug Administration、FDA) から「Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry」の案が 2010 年に公表されました。FDA が公表している他の企業向けガイダンス「Rare Diseases: Common Issues in Drug Development (2019 年 1 月)」と「Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products (2020 年 12 月)」においてもアダプティブデザインの適用について触れられています。また、数多くのアダプティブデザインを適用した臨床試験の報告例があり、最近では、COVID-19 治療薬開発のための臨床試験において、アダプティブデザインが適用されています。

さらに、アダプティブ臨床試験 (アダプティブデザインを適用する臨床試験) が ICH-E20 のトピックとして採択されています。ICH-E20 専門家作業部会では、アダプティブ臨床試験に関する用語の共通化、デザイン・実施・解析及び解釈に関する重要な原則、治験依頼者と規制当局との協議のための文書を明確にした、統一ガイドラインの策定を目的として、議論がなされています。

2019 年 11 月に、FDA からアダプティブデザインに関する企業向けガイダンス「Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry」の最終版が公表されました。本ガイダンスには、アダプティブデザインを適用する臨床試験の計画、実施、及び結果の報告に関する重要な原則や FDA と協議する際の考慮事項等がまとめられています。

アダプティブデザインは、安全でない又は効果が十分に期待できない治療への患者の曝露の制限、試験リソースの節約、開発プロセスを加速する等の潜在的な利点がある一方、試験デザインの選択、試験の実施、統計解析など臨床試験の一連の手順を適切に実施するためにはいくつかの留意事項があります。しかし、アダプティブデザインに関する国内のガイドラインが存在しないこと、FDA のガイダンスが最新かつ包括的な内容である等の理由から、アダプティブデザインの FDA ガイダンスの日本語訳を作成いたしました。

本翻訳がアダプティブデザインの理解や適切な使用の一助となれば幸いです。

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 2021 年度 継続タスクフォース 2

邦訳したガイダンス

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> [2021年7月8日アクセス]

アダプティブデザインを適用する COVID-19 治療薬開発のための臨床試験の文献

Stallard N, Hampson L, Benda N, Brannath W, Burnett T, Friede T, Kimani PK, Koenig F, Krisam J, Mozgunov P, Posch M, Wason J, Wassmer G, Whitehead J, Williamson SF, Zohar S, Jaki T. Efficient Adaptive Designs for Clinical Trials of Interventions for COVID-19. *Stat Biopharm Res.* 2020 Jul 29;12(4):483-497. doi: 10.1080/19466315.2020.1790415. PMID: 34191981; PMCID: PMC8011600.

目次

I. 緒言と適用範囲	6
II. アダプティブデザインの説明と動機	7
A. 定義	7
B. 重要な概念	7
C. 潜在的な利点と例	9
D. 限界	11
E. アダプティブデザインの選択	12
III. アダプティブデザインの原則	12
A. 誤った結論の可能性の制御	13
B. 治療効果の推定	13
C. 試験計画	14
D. 試験の実施と完全性の維持	15
IV. 比較を伴わないデータに基づくアダプティブデザイン	16
V. 比較データに基づくアダプティブデザイン	17
A. 群逐次デザイン	17
B. 症例数に対するアダプテーション	19
C. 患者集団に対するアダプテーション（例えば、アダプティブ・エンリッチメント）	21
D. 治療群の選択に対するアダプテーション	22
E. 患者割付に対するアダプテーション	23
F. 評価項目の選択に対するアダプテーション	24
G. 複数のデザイン要素に対するアダプテーション	25
VI. 特別な考慮事項とトピック	25
A. アダプティブデザイン計画におけるシミュレーション	25
B. ベイズ流アダプティブデザイン	27
C. イベント発現までの時間を評価項目とした場合のアダプテーション	28
D. 潜在的な代替評価項目又は中間評価項目に基づくアダプテーション	29
E. 副次評価項目	29
F. 安全性の考慮事項	30
G. 早期探索的試験におけるアダプティブデザイン	31
H. 中間解析の比較結果に基づく計画外のデザイン変更	31
I. 試験外部の情報源からの情報に基づくデザイン変更	31

VII.	試験の完全性の維持	32
VIII.	規制上の考慮事項	35
A.	FDA との協議	35
B.	アダプティブ臨床試験の実施前の文書.....	36
C.	完了した試験の評価と報告	38
IX.	参考文献	39

産業界のためのガイダンス¹

薬剤及び生物製剤の臨床試験のアダプティブデザイン

このガイダンスは、このトピックに関する食品医薬品局（FDA または政府機関）の現在の考え方を示すものである。これは、いかなる人に対してもいかなる権利も確立せず、FDA や一般市民を拘束するものではない。該当する法令や規制の要件を満たしている場合は、別のアプローチを使用することができる。別のアプローチについて議論するには、タイトルページに記載されている本ガイダンスを担当する FDA スタッフに連絡されたい。

I. 緒言と適用範囲

本文書は、試験依頼者、及び新薬臨床試験開始届（IND）、新薬承認申請（NDA）、生物学的製剤承認申請（BLA）、又は一変申請を提出する申請者に向けた、医薬品又は生物製剤の有効性及び安全性のエビデンスを提供するための臨床試験における、アダプティブデザインの適正使用に関するガイダンスである²。本ガイダンスには、アダプティブ臨床試験の計画、実施、及び結果の報告に関する重要な原則を記載している。また、コンピュータ・シミュレーションに依存するベイズ流アダプティブ試験や複雑なアダプティブ試験を含む、アダプティブデザインを用いた臨床試験の FDA による評価を円滑に進めるために提出すべき情報のタイプについても試験依頼者に助言している。

本ガイダンスは、医薬品の有効性と安全性を裏付けることを目的とした臨床試験のアダプティブデザインに主に焦点を当てる。本ガイダンスが含む概念は、早期臨床試験、探索的臨床試験、及び市販後のコミットメント又は要件を満たすために実施される試験にも有用である。

一般に、FDA のガイダンス文書は法的強制力のある責任を定めるものではない。代わりに、ガイダンスは特定の規制又は法的要件が引用されていない限り、トピックに関する FDA の最新の考え方が示しており、推奨事項として理解すべきである。FDA のガイダンスで「should」*という用語を使用する場合は、何かが提案又は推奨されているが、必須ではないことを意味する。

¹ このガイダンスは、Center for Drug Evaluation and Research の Office of Biostatistics と Food and Drug Administration の Center for Biologics Evaluation and Research の Division of Biostatistics によって作成された。

² 本文書で使用される医薬品という用語は、特に指定のない限り、ヒト用医薬品及び生物学的製剤の両方を指す。

* 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。

II. アダプティブデザインの説明と動機

A. 定義

本ガイダンスにおいて、「アダプティブデザイン (adaptive design)」を、臨床試験に参加した被験者の蓄積されたデータに基づいて、試験デザインの1つ以上の側面について、予め計画された変更を行うことができる臨床試験デザインと定義する。

B. 重要な概念

以下は、本ガイダンスで用いる重要な概念の説明である。

- 「中間解析 (interim analysis)」³とは、試験実施中に被験者から得られたデータを検討することであり、正式な群間比較がある場合に限定されない。中間解析で用いられる観測データには、ベースライン、安全性アウトカム、薬物動態学的、薬力学的又はその他のバイオマーカー、有効性アウトカム等、1つ又は複数のタイプのデータを含めることができる。
- 「比較を伴わない解析 (non-comparative analysis)」とは、被験者の治療群の割付情報を用いない、蓄積された試験データの解析のことである。「比較を伴う解析 (comparative analysis)」は、実際に割付けられた治療法又はコード (例えば、どちらが被験薬であるかを明らかにせずに、A と B のラベルが付けたもの) により特定された、蓄積された試験データの解析のことである⁴。「非盲検下の解析 (unblinded analysis)」と「盲検下の解析 (blinded analysis)」という用語は、治療の割付が特定されている解析か否かを区別するために用いることもある。本ガイダンスでは、治療の割付を知ることと、治療の割付をアダプテーションのアルゴリズムで用いることを混同させる恐れがあるため、「非盲検下の解析 (unblinded analysis)」と「盲検下の解析 (blinded analysis)」という用語の使用を避けている。被験者、試験責任医師、並びに試験依頼者やデータモニタリング委員会 (DMC) 等の他の担当者が、被験者

³ FDA の企業向けガイダンス「E9 臨床試験のための統計学的原則 (1998年9月)」では、中間解析を「試験が正式に完了する前に行われる有効性又は安全性に関する試験治療群間の比較を意図したすべての解析」と定義している。最新のガイダンスでは、試験のアダプテーションを決定するために用いることができる、蓄積されたデータの広範な解析に対応するために、中間解析をより広い意味で用いている。ガイダンスは定期的に更新している。ガイダンスの最新版については、FDA ガイダンスのウェブページ

(<https://www.fda.gov/RegulatoryInformation/Guidances/default.htm>) を参照のこと。

⁴ 比較を伴わない解析及び比較を伴う解析という用語の定義は、多群の臨床試験について言及している。単群の臨床試験では、蓄積された試験データの解析には割付けた治療を特定する情報が含まれるため、本ガイダンスの目的上、比較を伴う解析に相当するとみなされる。

に割付けられた治療を知っているか、群間比較の結果にアクセスできるかにかかわらず、中間解析は比較を伴う解析又は比較を伴わない解析になりうる。例えば、非盲検試験においても比較を伴わない解析に基づくアダプテーションを含めることができるが、アダプテーションが比較データを知ることによる影響を全く受けていないことを保証することはさらなる課題となる。中間解析に基づく比較結果へのアクセスを制限することの重要性については、本ガイダンスの VII 章で詳しく述べる。

- 本ガイダンスにおいて「事前に (prospective)」という用語は、蓄積された試験データの比較解析が行われる前にアダプテーションが計画され、詳細が規定されていることを意味する。ほとんどすべての状況において、試験開始前に、アダプティブなデザイン変更の可能性を計画し、試験実施計画書（及び統計解析計画書）に記載すべきである。
- 本ガイダンスでは、有効性の実質的なエビデンスを提供することを目的とする試験と、「探索的試験 (exploratory trials)」と呼ばれる他の試験を区別する⁵。この区別は、主要評価項目の臨床的関連性、試験実施の質、誤った結論が生じる可能性の厳密な管理、推定の信頼性等、臨床試験の複数の特徴に依存する。
- 「固定標本試験 (fixed sample trial)」とは、デザイン段階で目標症例数又は目標イベント数⁶を規定した臨床試験であり、事前に計画されたアダプテーションの対象とならない。
- 「非アダプティブ試験 (non-adaptive trial)」とは、試験デザインを変更する機会が事前に計画されていない臨床試験のことである。
- 「バイアス (bias)」とは、治療効果の推定が真の値からずれる系統的な傾向のことである。
- 「信頼性 (reliability)」とは、臨床試験から得られた統計的推測が、治療効果を精度よくかつ正確に評価する程度のことである。
- 医薬品の有効性、場合によっては安全性を実証する上で重要な要素は、臨床試験における仮説検定である。定められたある有意水準（通常は片側有意水準 0.025）で帰無仮説が棄却され、その薬剤に対する臨床的に意味のある効果が実証された場合、

⁵ 第 1 相、第 2 相、第 3 相 (21CFR 312.21)、ピボタル、レジストレーション、検証的 (FDA の企業向けガイダンス「E9 臨床試験のための統計学的原則 (1998 年 9 月)」等、異なるタイプの臨床試験を表すために様々な用語が用いられている。本ガイダンスではこれらの用語は用いない。

⁶ 関心のある主要アウトカムが、(死亡等の) イベント発現までの時間である場合、統計的検出力は症例数ではなく観察されたイベント数によって定まる。

一般にそのエビデンスは有効性の結論を裏付けるものとなる。ただし、薬剤が無効であっても、帰無仮説が棄却されることがある。これは「第一種の過誤 (Type I error)」と呼ばれる。通常、帰無仮説が正しいシナリオは複数あり、これらのシナリオにおける帰無仮説を棄却する最大確率という意味で、「第一種の過誤確率 (Type I error probability)」という用語を用いる。

C. 潜在的な利点と例

アダプティブデザインには、非アダプティブデザインと比べてさまざまな利点がある。これらの利点は、アダプティブデザインを伴う臨床試験の基本的な特性によるもので、試験開始時に利用できなかった情報に試験を適応させることが可能となる。利点の具体的な性質は、科学的背景と検討される1つ又は複数のアダプテーションのタイプによって異なり、潜在的な利点は、次の主要なカテゴリーに分類される。

- 統計的効率：アダプティブデザインは、同等の非アダプティブデザインと比べて、真の薬剤効果を検出する可能性が高くなる（つまり、統計的検出力が高くなる）場合がある⁷。例えば、群逐次デザイン (V.A.項) や症例数をアダプティブに変更するデザイン (V.B.項) がしばしば当てはまる。あるいは、アダプティブデザインは、同等の非アダプティブデザインと比べて、小さな期待症例数⁸又は短い期待試験期間で、等しい統計的検出力を確保する可能性がある。
- 倫理上の考慮事項：アダプティブデザインは、非アダプティブデザインと比べて、多くの点で倫理的に優れている。例えば、有効性を実証する可能性が低いことが明らかになった場合に試験を早期中止できることは、効果のない試験治療の不必要なリスクに晒される患者数を減らし、被験者により有望な代替治療を探索する機会を与えることができる。
- 薬剤効果の理解の向上：アダプティブデザインにより、非アダプティブデザインで通常できることよりも、幅広い問題に対処できる。例えば、ある患者集団の有効性を実証するためには、非アダプティブデザインであれば、実施不可能なほどの大規模な症例数を必要とする場合においても、アダプティブ・エンリッチメントデザイン (V.C.項) を用いると、その患者集団全体又は目標とする部分集団のいずれかで有効性を実証できる場合がある。アダプティブデザインは、試験治療がもつ効果への理解を向上させることもできる。例えば、アダプティブ用量選択を伴うデザイン (V.D.項) は、用量反応関係のより良い推定や、ひいてはより効率的な後続の試験につながる可能性がある。

⁷ 同等の非アダプティブデザインの例は、アダプティブデザインの期待症例数と等しい症例数をもつ固定標本デザインである。

⁸ 期待症例数とは、試験を何度も繰り返した場合の平均症例数のことである。

- 利害関係者への受容性：アダプティブデザインは、同等の非アダプティブデザインよりも柔軟性が増すことにより、利害関係者にとって受け入れやすいデザインであると考えられる場合がある。例えば、試験依頼者は、蓄積された情報に基づくデザインの変更が計画された試験に、より積極的に取り組むかもしれない。患者は、反応アダプティブランダム化を適用する試験（V.E.項）により有効な治療に割付けられる確率を高くなることから、より積極的に参加しようとするかもしれない。

以下のアダプティブ臨床試験の例は、いくつかの潜在的な利点を示している。

- 重度の頭部外傷患者に対する治療法として、エリプロゾルを評価するための臨床試験が行われた（Bolland et al. 1998）。有効性の主要評価項目は、治療開始 6 か月後の患者の機能状態を定義する 3 つのカテゴリーのアウトカムであった。デザイン段階では、プラセボ対照群における 3 つの異なる機能的アウトカムそれぞれを経験すると予想される割合は、かなり不確実であった。中間解析では、症例数の追加を想定し、併合した非比較データに基づいて、3 つのカテゴリーの割合の推定値を更新することが事前に計画されていた。このアプローチは、統計的検出力が不十分な試験を回避するために選択されたものであり、試験がその目的を効率的かつ確実に達成することを保証するのに有用であった。中間解析の結果、最終的に症例数を 400 人から 450 人に増加することとなった。
- PARADIGM-HF は、駆出率が低い慢性心不全患者を対象として、ネプリライシン阻害薬サクビトリルとレニン・アンジオテンシン系（RAS）阻害薬バルサルタンの配合薬である LCZ696 と RAS 阻害薬エナラプリルを、心血管死又は心不全による入院の複合評価項目のリスクに関して比較するようにデザインされた臨床試験であった（McMurray et al. 2014）。試験デザインには、総イベント数の 1/3、1/2、及び 2/3 が集積された時点で、3 回の中間解析が計画され、中間解析に基づく比較結果に基づいてエナラプリルに対する LCZ696 の優越性が示された場合に有効性による中止が可能とされた。有効性による中止規則を伴う中間解析を追加することにより、固定された総イベント数を観察された後に解析を 1 回実施する試験と比べて、試験の成功確率を同程度に維持しながら、期待症例数と期待試験期間を減少させることができた。3 回目の中間解析時に、エナラプリルに対する LCZ696 の明らかな優越性を示す、事前に規定した中止境界を越えたことから、PARADIGM-HF は中止された。したがって、群逐次デザインでは、固定標本デザインに比べて、より迅速にベネフィットの結論を導くことができた。
- 9 価のヒトパピローマウイルス（HPV）ワクチンの安全性と有効性を評価するために、アダプティブ用量選択を伴う臨床試験を実施した（Chen et al. 2015）。この試験では、9 価 HPV ワクチンの 3 つの用量製剤の 1 つ又は実薬対照である 4 価 HPV ワクチンのいずれかに被験者をランダム化した。3 つの用量製剤の 1 つを選択する中

間解析を実施し、試験の第 2 ステージに移行した。試験の目的は、適切な用量を選択し、その用量の安全性と有効性を適時に確認することであった。

- STAMPEDE は、標準的なアンドロゲン除去療法 (ADT) と、ADT を 1 つ又は複数の承認された治療法と組み合わせたいくつかの異なる治療レジメンと比較することにより、医療の現場に情報を提供し、前立腺癌における複数の治療法を同時に評価するようにデザインされた臨床試験であった (Sydes et al. 2012)。試験デザインには、比較結果に基づいて十分な効果が認められなかった治療群の除外を可能とした、複数回の中間解析が含まれていた。治療群の中止を可能とした逐次解析とともに、共通の対照群を用いることにより、複数の治療法を別々の試験で評価する場合と比べて、複数の治療法をより効率的に、同時に評価することが可能となった。
- PREVAIL II は、エボラウイルス病患者の治療として、ZMapp+標準治療と標準治療単独を比較評価するために実施された臨床試験であった (PREVAIL II Writing Group et al. 2016, Dodd et al. 2016)。本試験では、中間解析及び最終解析において有効性を結論付けるための決定規則は、標準治療に ZMapp を上乘せにすることにより 28 日死亡率が低下するベイズ流事後確率に基づいて行われる、ベイズ流アダプティブデザインを用いた。患者 2 例が完了する毎に中間解析を実施することとされ、最小症例数 (各群 12 例) が登録されるまでは何も行わずに試験を継続されることとされた。このデザインは、新たな治療群として試験治療を追加したり、試験実施中に有効であると判断された治療薬を標準治療に追加したり、標準治療に置き換えることも可能となった。

D. 限界

アダプティブデザインを用いる臨床試験では、以下のような限界が生じる可能性がある。

- アダプティブデザインは、誤った結論の可能性を高め、推定値にバイアスが生じることを避けるために、特別な解析方法を必要とする。複雑なアダプティブデザインに関しては、そのような方法は容易に用いることができない場合があり、しばしばシミュレーションが重要となる (VI.A.項)。
- ある側面における効率の向上は、別の側面での効率の低下によって相殺される可能性がある。例えば、アダプティブデザインでは、同等の非アダプティブデザインと比べて、最小症例数や期待症例数が減少するものの、最大症例数は増大する可能性がある⁹。さらに、アダプティブデザインの変更を事前に計画するには、デザイン段階でより多くの労力が必要となり、試験の計画から開始までのリードタイムが長く

⁹ 最小症例数と最大症例数は、試験を何回も繰り返した場合に、アダプティブデザイン下で選択される最も少ない症例数、最も多い症例数のことである。

なる可能性がある。

- アダプティブデザインを用いると、適切な試験実施と試験の完全性を維持するための試験運営上の課題が追加される。特に、中間解析に基づく比較結果へのアクセスを適切に制限するためのアプローチは複雑であり、試験の費用を増大させる場合がある。さらに、アダプテーションに関する意思決定が最新かつ信頼できる結果に基づいて行われるようにするために、質の高い中間データをタイムリーに入手できることを保証することが課題となる。
- アダプテーションによる効率向上の機会は、重要な科学的制約又は特定の臨床の状況によって制限される場合がある。例えば、信頼性の高い安全性評価を行うためには、最小症例数を設定することが必要な場合がある。また、関心のある主要なアウトカムの確認が、ほとんど又はすべての患者の登録に要する時間よりも長い期間をかけて行われる場合には、アダプテーションのタイプによっては、その有用性が限られる可能性がある。
- 試験デザインのアダプティブな変更は、アダプテーションの前後で結果が異なり、結果の解釈が困難につながる場合がある。

E. アダプティブデザインの選択

一般に、臨床試験のデザインにアダプティブな要素を用いるかどうかの決定は、前項までに説明した潜在的な利点と欠点を含む多くの要因に左右される。また、様々な科学的でない事項を考慮すべきかもしれない。すなわち、臨床試験のデザインは複雑なプロセスである。本ガイダンスは、アダプティブデザインの一般的な使用又は特定の状況での使用を要求若しくは制限することを意図していない。しかしながら、試験依頼者に、計画段階でさまざまな試験デザインの選択肢を検討し、その考慮事項について、End-of-Phase-2 (EOP2) 又は Type C meeting 等の規制当局との会議で、適切な FDA 審査部門と議論することを推奨する。

III. アダプティブデザインの原則

一般に、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アダプティブ臨床試験のデザイン、実施、解析は、四つの主要な原則を満たすべきである。すなわち、誤った結論を導く可能性を適切に制御すること、治療効果の推定を十分に信頼できるものとする、デザインの詳細を完全に事前に規定すること、及び試験の完全性を適切に維持することである。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すべての臨床試験は、これらの四つの原則を満たすべきであるが、以下の項では、アダプティブデザインに特有の考慮すべき事項を概説する。

A. 誤った結論の可能性の制御

臨床試験は市販前の意思決定において中心的な役割を果たすため、検討中の試験デザインが安全性又は有効性に関する誤った結論、安全性又は有効性が欠如しているという誤った結論、又はベネフィット・リスクの総合的な評価に寄与する推定に誤解を与える可能性を評価することが重要である。例えば、アダプテーションの特徴が試験の第一種の過誤確率を増大させる方法はいくつかあり、最も明らかな例は、複数の統計的仮説検定を行う場合である。計画被験者の50%が試験を完了した後に、有効性による試験を中止する可能性のある中間の検定を行う、群逐次デザインについて考える。試験が早期に中止しない場合、計画された被験者の全例が試験を完了した時点で、最終の検定が行われる。これら2回の検定をそれぞれ従来の片側有意水準0.025で行い、薬剤が有効でない場合、試験全体での第一種の過誤確率は2.5%を超えることになる。これはよく知られている問題であり、試験全体の第一種の過誤確率を2.5%に制御することを保証するとともに、中間及び最終解析の適切な有意水準を決定するためのさまざまな方法が存在する (Jennison and Turnbull 1999)。

明確な複数の仮説検定だけが、アダプティブデザインが誤った結論を導き得る特徴ではない。試験集団全体と部分集団のデータを試験途中で検討し、より大きな治療効果を有する集団を継続的に研究する集団として選択する、アダプティブ患者集団選択に対するナイーブなアプローチを考える。選択された集団を対象として、有意水準0.025で最終解析を行う場合に、患者集団の選択時に用いたデータも含めて解析を行うと、第一種の過誤確率は2.5%を超える。その他のアダプティブデザインの特徴によって、より繊細な第一種の過誤確率の増大が生じることもありえる。

したがって、帰無仮説に対する検定を組み込んだ試験にアダプティブデザインを提案する場合は、第一種の過誤確率の増大の可能性に対処すべきである。単純な群逐次デザイン (V.A.項) 等のいくつかの場合では、統計理論を用いて、第一種の過誤確率を目標とする水準に制御することを保証する有意水準を導出することができる。比較を伴わない中間解析結果に基づく症例数再推定 (IV章) 等の他の場合では、従来の0.025の有意水準で解析しても、第一種の過誤確率に影響がないか、影響が限定的であることを示すことができる。多くのベイズ流アダプティブデザイン (VI.B.項) 等の他の場合では、誤った結論を導く可能性を評価するためにシミュレーションを用いることが重要な場合がある (VI.A.項)。

B. 治療効果の推定

臨床試験では、ベネフィット・リスクの評価を円滑に進め、新薬の適切なラベリングのために、根拠に基づく医療の実践を可能にする、十分に信頼できる治療効果の推定値を算出することが重要である。アダプティブデザインの特徴は、治療効果とそれに関連する量の推定値に統計的バイアスをもたらす可能性がある。例えば、III.A項で述べた第一種の過誤確率の増大が生じる2つの場合では、推定値にバイアスが生じる可能性がある。具体的には、アダプテーションを考慮しない標本平均のような、試験終了時の治療効果の従来の

推定値は、真の集団治療効果を過大評価する傾向がある。これは、アダプテーションの基礎となる主要評価項目だけでなく、主要評価項目と関連する副次評価項目にも当てはまる。さらに、主要評価項目と副次評価項目の信頼区間は、真の治療効果の被覆確率が正確でない場合がある。

一部のアダプティブデザインでは、アダプテーションに関連するバイアスを低減又は除去するために推定値を調整する方法や、平均二乗誤差等の指標の性能を改善する方法が知られている¹⁰（例えば、Jennison and Turnbull 1999、Wassmer and Brannath 2016）。このような方法が利用可能である場合、事前に計画し、結果を報告するために用いるべきである。しかしながら、アダプティブデザインのバイアスの推定は、現在のところ、第一種の過誤確率の増大よりも十分に研究されていない事象であり、上記のような方法が利用できないデザインがある。それらのような他のデザインでは、推定値に含まれるバイアスの程度を評価し、治療効果の推定値とその信頼区間を、解釈に関する注意とともに提示すべきである。

C. 試験計画

一般に、他の臨床試験と同様に¹¹、アダプティブデザインの詳細は試験の開始前に完全に規定され、適切に文書化されることが期待される（VIII.B.項）。事前計画には、予定される中間解析の回数と時期、アダプテーションのタイプ、用いる統計的推論の方法、及びアダプテーションの決定に適用される特定のアルゴリズムの規定を含めるべきである。完全に事前規定することは、さまざまな理由から重要である。第一に、多くのタイプのアダプテーションにおいて、アダプテーションに関する意思決定の側面が計画されていない場合、一旦データを収集してしまうと、誤った結論を導く可能性を制御できず、信頼できる推定値を算出するための適切な統計手法を用いることができない可能性がある。第二に、完全な事前規定は、アダプテーションの決定が計画外の方法で、蓄積した知識に基づくものではないという信頼性を高める助けとなる。例えば、アダプテーションに関する意思決定に関与する担当者（例えば、モニタリング委員会）が中間解析に基づく比較結果にアクセスでき、中間の比較を伴わない併合分散の推定値に基づく症例数再推定（IV章）が計画された試験を考える。症例数を変更するための正確な規則を事前規定することにより、アダプテーションが比較結果の知見に影響されたのではないかという懸念が軽減され、中間解析の比較結果に基づく変更を考慮した統計的調整方法を用いずに済むことになる（V.B.項）。最後に、完全な事前規定は、デザイン段階での慎重な計画を動機付けることができ、中間解析に基づく比較結果への不要な試験依頼者のアクセスを排除することができる。また、DMC

¹⁰ 平均二乗誤差は、バイアスとばらつきの両方を組み込んだ推定値の性能の指標である。

¹¹ FDAの企業向けガイダンス「E9 臨床試験のための統計学的原則（1998年9月）」では、すべての臨床試験のデザイン及び解析計画の事前規定が推奨されている。

がアダプティブデザインの実施に関与している場合、患者の安全と試験の完全性を維持するという主要な責務に効果的に注力することを保証するのに役立つ（VII章）。

アダプテーションを決定する規則を事前に規定することが推奨されるが、モニタリング委員会の勧告は、データの総合的な評価に基づいて、想定されるアルゴリズムから逸脱する可能性がある。このタイプの柔軟性が望まれる場合、事前規定された計画は、想定されるアルゴリズムからの逸脱の可能性を認め、そのような逸脱につながる可能性のある要因の概要を示し、アルゴリズムの厳密な遵守に依存しない検定及び推定方法を提案すべきである。全く予期しない状況が発生した場合は、デザイン変更の可能性について、できるだけ早く FDA と協議することを推奨する。

D. 試験の実施と完全性の維持

アダプティブデザインは、試験実施上の複雑さを生み出す可能性がある。蓄積されたデータから得た知見は、試験の経過と実施、及び試験依頼者、試験責任医師、参加者の行動に、予測困難で調整不可能な方法で影響を与える可能性がある。したがって、すべての臨床試験（アダプティブ及び非アダプティブ）について、中間解析に基づく比較結果へのアクセスは、試験の実施又は管理を担当する者から独立した、適切な専門知識を持つ個人に制限することが強く推奨される¹²。中間解析に基づく比較結果の機密性を保持することは、試験デザインにアダプテーションの特性が含まれる場合、特に困難である。アダプティブ臨床試験で発生する可能性のある問題の例として、以下の2つの例が挙げられる。

- 試験責任医師が中間解析の比較結果へのアクセスを不適切に与えられた場合、信頼性の低いデータに基づいて治療効果の推定値が小さい又は好ましくないと知ることが、効果がないという信頼できるエビデンスとして誤って解釈されることがある。その結果、アドヒアランスの低下及び患者を試験に留める意欲の低下につながり、試験の残りにおける欠測データの量を増加させることになる。
- 比較結果に基づいて症例数再推定を伴うデザインの間中解析（V.B項）の後、目標症例数が追加されたという知見は、試験責任医師及び潜在的な試験参加者により、中間解析時の治療効果は期待よりも低いものとして暗に解釈され、その後の組み入れが遅くなり、試験の成功を危うくする可能性がある。

これらの問題や他の同様の問題は、データが収集された後に調整することが一般に不可能であるため、アダプティブ臨床試験の計画には、可能性のある試験実施上の問題の原因と結果を検討し、これらの問題を回避するための計画を含めるべきである。計画には、情報へのアクセスを制御し、試験を通じたアクセスを記録するプロセスを記述すべきである。これについては、VII章でより詳細に述べる。

¹² この勧告は、FDAの企業向けガイダンス「E9 臨床試験のための統計学的原則（1998年9月）」にも書かれている。

IV. 比較を伴わないデータに基づくアダプティブデザイン

本項では、アダプテーションが完全に比較を伴わないデータの解析に基づく、すなわち、治療の割付に関する情報を用いないアダプティブ臨床試験について考える。このような解析は、盲検下の解析又はマスク下の解析とも呼ばれる。一般に、比較を伴わないデータに基づいて適切に事前規定されたアダプテーションは、第一種の過誤確率に影響を与えないか、影響が限定的であり、特にイベントの発生確率や評価項目のばらつきに関する不確実性が高い場合等、多くの場面で魅力的な選択肢となる。

アウトカムデータを蓄積することは、試験のアダプテーションに有用な基礎となり得る。治療の割付情報を用いないアウトカムデータの解析は、併合解析と呼ばれることもある。併合されたアウトカムデータに基づく、最も広く用いられているアダプティブデザインのタイプには、症例数のアダプテーション（盲検下の症例数再推定とも呼ばれる）が含まれる。臨床試験における症例数の計算は、目標とする有意水準、目標とする検出力、治療の割付によるアウトカムの想定される又は目標とする群間差、及びその他の局外パラメータ（主要な関心はないが統計的な比較に影響を与えうる値）等、いくつかの要因に依存する。反応又は望ましくないイベント等の二値アウトカムの試験では、対照群の反応又はイベントの発生確率は、一般に局外パラメータとみなされる。症状スコア等の連続量アウトカムの試験では、スコアの分散が局外パラメータである。局外パラメータに関する蓄積された情報を利用することにより、事前に規定されたアルゴリズムに従って症例数を調整し、目標とする検出力を確保することができる。局外パラメータ自体が治療の割付の情報に依存するため、場合によっては、局外パラメータの推定のための統計的モデリングを伴う場合がある（Gould and Shih 1992）。一般に、これらのアダプテーションにより、第一種の過誤確率を増大することはない。しかしながら、非劣性又は同等性の仮説検定を含んだ試験では、第一種の過誤確率の限定的な増大が生じる可能性がある（Friede and Kieser 2003）。試験依頼者は、これらのシナリオでは第一種の過誤確率の増大の程度を評価すべきである。

併合されたアウトカムデータに基づくアダプテーションのもう1つの例は、予後エンリッチメント戦略（prognostic enrichment strategy）における、バイオマーカー又は他のベースライン特性の予後への影響の強さに対する計画された中間再評価である¹³。例えば、ある試験では、評価項目のイベント数を増やすために特定のバイオマーカーを有する患者をより多く登録することを目標とするかもしれない。このとき、中間のプールされたアウトカムデータから、バイオマーカーが併合されたイベント発生率に対して予想したほどの効果がないことが示唆される場合があり、これはおそらく登録方針の変更につながるだろう。

¹³ FDAの企業向けガイダンス「Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products（2019年3月）」におけるさらなる議論を参照のこと。

V. 比較データに基づくアダプティブデザイン

本項では、比較データの間中解析に基づく試験中止又はデザイン変更について事前規定された規則が存在する、様々なタイプの臨床試験デザインについて述べる。このような解析は、非盲検下の解析又は非マスク下の解析とも呼ばれる。以下の項にて、一般に適用できるいくつかの重要なコンセプトを示す。第一に、比較を伴わないデータに基づくアダプテーションとは対照的に、比較データに基づくアダプテーションにより、しばしば第一種の過誤確率が直接増大し、治療効果の推定にバイアスが生じる。したがって、アダプティブデザインを考慮した統計手法を用いるべきである。第二に、アダプテーションが比較を伴う中間解析に基づく場合、適切な試験実施を確実にするために追加の手順が不可欠である。これについては、VII章でより詳細に述べる。最後に、中止又はアダプテーションルールは、治療効果の推定値、固定標本の p 値、試験成功の条件付き確率、薬剤が有効であるベイズ流事後確率、又は試験成功のベイズ予測確率等、様々な異なる尺度で設定することができる。デザインの動作特性¹⁴が適切に評価されている限り、尺度の選択は比較的重要ではない。

A. 群逐次デザイン

群逐次試験では、事前規定された試験中止基準を用いて、事前に計画された比較データでの中間解析を1回以上行うことができる。逐次解析を含めることにより、臨床試験で期待される症例数と期間を減らし、安全かつ有効な新しい治療法の承認を早めることで、倫理的及び効率的な利点をもたらすことができる。例えば、1回の中間解析と一般に用いられる有効判定に関する中止境界を伴う群逐次デザインは、同等の固定標本試験と比較して、試験の期待症例数を約15%減少させることができる¹⁵。

群逐次デザインは、規制上の意思決定を支持するのに十分な有効性のエビデンスがある場合、又は試験が有効性を検証する可能性が低いというエビデンスがある場合に（これはしばしば無益性による中止と呼ばれる）、試験を中止するための規則を含めることができる。群逐次デザインの有効性に関する複数回の統計的仮説検定を、従来の片側有意水準0.025でそれぞれ行くと、第一種の過誤確率が増大し、その結果、誤った結論が導く可能性が高くなる。第一種の過誤確率が適切に制御するための、中間解析及び最終解析の適切な中止

¹⁴ 試験の動作特性は、所与のデザインでの試験の特性である。例えば、注目する特性には、第一種の過誤確率、検出力、期待・最小・最大症例数、治療効果の推定値のバイアス、信頼区間の被覆確率（すなわち、臨床試験が何度も繰り返された場合、信頼区間に真の治療効果が含まれる確率）が含まれる。

¹⁵ O'Brien-Fleming による有効性の中止境界を用いて、最大患者数の半分の患者でアウトカム情報が得られた時点で中間解析を伴う群逐次デザインでは、すべての患者が登録されてアウトカムが確認された時点で計画された1回の解析を伴うデザインと比較して、対立仮説（検出力90%）が真である場合、試験の期待症例数を約15%減少させる。

境界を決定する様々な方法が存在する。例えば、O'Brien-Fleming のアプローチでは、試験の有効中止のためには、非常に有力な早期の結果が必要となる傾向がある (O'Brien and Fleming 1979)。Pocock による提案などの別のアプローチは、O'Brien-Fleming のアプローチほどの有力な早期の結果を必要とせず、早期中止の確率がより高い (Pocock 1977)。これらの、あるいはその他のアプローチは、中間解析の回数、及び中間解析時の具体的な目標症例数又は目標イベント数双方の事前計画に依存する。

Lan-DeMets の α 消費¹⁶アプローチは、試験期間中に第一種の過誤確率をどのように消費するかを示す関数を指定することにより、早期中止の様々なレベルのエビデンスに対応すると同時に、中間解析の回数と時期を柔軟に決定することを可能にする (Lan and DeMets 1983)。解析時期の柔軟性は、モニタリング委員会のスケジュールを特定の間解析時における症例数や目標イベント数ではなく、特定の暦時間に調整するのに有用であり、解析回数の柔軟性は、登録率が予想よりも速い又は遅い場合に対応するのに有用である。しかしながら、蓄積されたデータに基づく比較結果に基づいて中間解析の回数を選択すると、第一種の過誤確率が增大する可能性がある。例えば、現在の間解析結果が中止境界に近い場合、次の中間解析が当初の計画よりも早く行われるように調整することは適切ではない。Lan-DeMets の α 消費アプローチではこのような潜在的な問題があるため、試験依頼者は、中間解析の目標回数及びそのおおよその実施予定日、並びに試験開始後の解析回数又は時期を変更するための決定の枠組み等、追加の予防手段を講じるべきである。決定の枠組みは、推定される治療効果から統計的に独立した情報 (例えば、登録率や一連の業務に関わるスケジュール) に基づくべきである。例えば、決定の枠組みでは、半年ごとの中間解析を規定し、登録が事前に定められた目標よりかなり遅い場合には、解析の追加を規定することもあり得る。

群逐次試験の適切なデザイン、実施、及び解析を確実にするためには、さらに多くの考慮事項がある。第一に、群逐次法が妥当であるためには、事前の解析計画を遵守し、中止基準が満たされた場合にのみ、有効性により試験を中止することが重要である。第二に、無益性による試験の早期中止に関するガイドラインを適切に実装すべきである。試験デザインは、制約のない無益性の規則をしばしば採用している。この無益性の規則は、得られた中間解析結果の総合的な評価に応じて、遵守してもしなくてもよいガイドラインを含めている。このような制約のない無益性ガイドラインを固定標本試験又は有効性に関する適切な群逐次中止規則を用いた試験に追加しても、第一種の過誤確率は増大せず、しばしば適切である。あるいは、群逐次デザインは、無益性の基準が満たされる場合には常に試験を中止すべきであるという、制約のある無益性の基準を含んでもよい。制約のある無益性の基準は有効性解析にいくつかの利点があるが (例えば、有効と判断するための閾値の緩和)、第一種の過誤確率は、中止規則に従う場合にのみ制御される。したがって、事前に

¹⁶ 臨床試験の第一種の過誤確率は、しばしばギリシャ文字 α (アルファ) で表される。

規定された制約のある無益性の基準を満たしているにもかかわらず試験を継続する場合、FDA は最終解析の結果にかかわらず、その試験が有効性のエビデンスを提供しなかったと判断する可能性が高いだろう。ただし、一部の DMC は、制約のない無益性ガイドラインの柔軟性を好む場合があることにも留意すべきである。

第三に、早期に有効中止した試験では、安全性及び潜在的に重要な副次的有効性評価項目の評価のための症例数が少なくなる。したがって、早期の有効中止は、通常、やむを得ない倫理的理由がある場合（例えば、主要評価項目が生存又は不可逆的な病的状態である）、又は推定される治療効果の大きさと効果のエビデンスの強さの両方の観点から、中止規則が非常に説得力のある結果を必要とする場合に限られる。場合によっては、信頼できる安全性評価のために必要な最低限の症例数により、どの程度早期に中間解析を実施すべきか、あるいはそもそも実施すべきかどうかには制限が生じるかもしれない。これは、例えば予防ワクチンの試験でしばしば当てはまる。

最後に、群逐次デザインを用いた場合、標本平均のような従来の固定標本下での点推定値は、真の値よりも大きな効果に偏る傾向があり、同様に、信頼区間は、目標とする名目上の被覆確率を有しない。そのため、群逐次デザインの中止規則を適切に調整する推定値と信頼区間を算出するための様々な方法が存在する（Jennison and Turnbull 1999）。試験結果の科学的及び統計的信頼性を保証し、重要なベネフィット・リスクの検討を容易にするために、群逐次デザインを適切に考慮した推定値及び信頼区間を算出するためのアプローチを事前に計画し、結果の報告に用いるべきである。

B. 症例数に対するアダプテーション

アダプティブなアプローチの 1 つは、治療の割付情報を用いた解析による局外パラメータの中間解析時の推定値に基づき、症例数の変更を事前に計画することである。例えば、連続的なアウトカムの分散を個々の治療群の分散の推定値から推定する方法や、対照群の二値のアウトカムの確率をその群のデータのみから推定する方法がある。これらのアプローチは、一般に第一種の過誤確率に影響を与えないか、影響があっても限定的である。しかしながら、治療群間の比較はせずに治療群を併合して推定した局外パラメータの推定値に基づくアダプテーション（IV 章）とは異なり、これらのアダプテーションには治療群の割付情報が含まれているため、試験の完全性を維持するために追加の手順が必要となる（VII 章）。

もう 1 つのアダプティブなアプローチは、中間解析に基づく比較結果（つまり、治療効果の中間推定値）に基づいた症例数の変更を事前計画することである。この方法は、「非盲検下での症例数のアダプテーション」又は「非盲検下での症例数再推定」と呼ばれる。症例数の決定は、対照群のイベント発現率、主要なアウトカムの分散、第一種の過誤確率、仮定した治療効果の大きさとその大きさを検出するため目標とする検出力等、多くの要因に依存する。IV 章では、試験の計画段階でのアウトカムの分散、又は対照群のイベント発

現率の不確実性に対処するための、比較を伴わない中間解析結果に基づくアダプテーションについて述べた。これに対し、真の治療効果の大きさについてかなりの不確実性がある場合は、中間解析に基づく比較結果に基づいて症例数を調整するデザインを用いる場合もある。群逐次試験と同様に、中間解析の比較結果に基づいて症例数を調整するデザインは、妥当な効果の大きさの範囲で十分な検出力を提供できるため、真の治療効果の大きさが仮定した値よりも小さいが、依然として臨床的に意味のある大きさであった場合に、試験が十分な検出力を維持することを保証するのに役立つ。さらに、症例数を変更するための事前規定を追加することにより、ある設定下では、特定の動作特性に関する効率上の利点を得ることができる。

適切な調整を行わずに試験の被験者数をむやみに変更することは、第一種の過誤確率を増大させる可能性がある。中間解析を1回実施し、その際の治療効果の推定値を用いて最終的な症例数を変更するデザインを考える。試験終了時に従来どおり、有意水準0.025で仮説検定を行った場合、第一種の過誤確率は2倍以上になりうる(Proschan and Hunsberger 1995)¹⁷。したがって、このようなタイプのアダプティブデザインでは、第一種の過誤確率を適切に制御するために、利用可能な様々な方法の1つを用いるべきである。例えば、仮説検定のアプローチとしては、事前規定した方法で異なる試験のステージから算出された検定統計量又はp値を併合する方法や条件付き第一種の過誤確率を制御することを通じて、第一種の過誤確率を維持する方法が開発されている。(例えば、Bauer and Köhne 1994、Fisher 1998、Cui et al. 1999、Denne 2001、Müller and Schäfer 2001、Chow and Chang 2011)。これらのアプローチは、将来の中間解析回数や解析間隔等、最大症例数以外のサンプリング計画の側面に対するアダプテーションにも対応している。

群逐次デザインに関するV.A.項で考察したアダプテーション計画の遵守、安全性評価、及び治療効果の推定に関する追加の考慮事項は、比較データに基づいた症例数のアダプテーションを伴う試験デザインにも当てはまる。ただし、事前計画には、用いる統計学的仮説検定手法だけでなく、症例数を変更する規則も規定すべきであることに注意が必要である。最後に、症例数のアダプテーションを伴う場合、試験の完全性を維持するためにはさらなる課題がある。例えば、症例数の変更規則は、しばしば試験終了時に統計学的に有意な治療効果が得られる条件付き確率(しばしば条件付き検出力と呼ばれる)を目標とする水準又はそれに近い水準に維持することに基づいている。このシナリオでは、症例数の変更規則とアダプティブに選択された症例数から、中間解析時の治療効果の推定値を比較的簡単に逆算することができる。したがって、試験の完全性を維持するためには、これらの詳細な情報を知り得る者が限定されるように追加の方策を講じるべきである。さらなる議論については、VII章を参照のこと。

¹⁷ これは、2つの解析を調整するためにボンフェローニ法を用いても十分ではないことを意味する。

本項で考察した原則は、アダプティブデザインによって、中間データの比較結果に基づく、事前に計画された総イベント数の変更が可能な、イベント発現までの時間を評価項目とした試験にも当てはまる。しかしながら、このような設定の試験に関しては、VI.C.項でより詳細に述べる、特別な追加の考慮事項がいくつか存在する。

C. 患者集団に対するアダプテーション（例えば、アダプティブ・エンリッチメント）

多くの場合、試験の対象集団のある特定の部分集団において、治療効果がより大きくなることが期待されるだろう。こうした部分集団は、例えば、人口統計学的特性や薬剤の作用機序に関連すると考えられる遺伝的又は病態生理学的マーカーによって定義することができる。このような状況では、中間解析に基づく比較結果に基づいて患者集団をアダプティブに変更することが可能なデザインを検討することができる。例えば、ある試験では、中間解析までは対象集団全体から被験者を組み入れるが、中間解析の時点で、事前に定めた基準に基づいて、その後の被験者も対象集団全体からの組み入れるのか、標的とすべき特定の部分集団のみから組み入れることにするのかを判断するかもしれない。標的とした部分集団に対する治療効果は、中間解析の前後に蓄積されたデータを統合して推定することができる。このようなタイプの試験デザインは、しばしばアダプティブ・エンリッチメントデザイン¹⁸と呼ばれ、非アダプティブデザインにはない利点を有する。特に、このようなアダプティブデザインは、全体集団を対象とした非アダプティブの固定標本デザインと同じ症例数でより大きな検出力¹⁹を有する。さらに、このようなアダプティブデザインは、初めから被験者の組み入れを標的部分集団に限定している試験とは異なり、標的でない部分集団（補集団）における試験治療の評価が可能である。

患者集団をアダプティブに変更することを可能にする試験デザインは、(1)中間解析に基づく比較結果に基づき、登録対象集団と主要解析で評価する集団等、デザインの特徴の変更、(2)標的部分集団と全体集団等、複数の集団に対する仮説検定を伴うことが多い。したがって、統計学的仮説検定の手法は、双方から生じる多重性の原因を考慮すべきである。例えば、1つのアプローチは、試験の異なるステージで得られた検定統計量又は p 値を事前に計画された方法で併合するとともに、適切な多重検定手順を用いるという方法である (Wassmer and Brannath 2016)。このようなアプローチであれば、症例数のアダプテーションや特定の部分集団から登録する患者割合を変更するようなアダプテーション（例えば、患

¹⁸ アダプティブ・エンリッチメントという用語は、例えば、FDA 企業向けガイダンス「Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products (March 2019)」において用いられている。

¹⁹ この文脈における検出力とは、例えば、標的部分集団又は全体集団のいずれかにおいて、真の治療効果を見極めることができる確率として定義することができる。

者の登録をある部分集団のみに限定するのではなく、その部分集団に登録される患者の割合を増加させる)にも適用できる可能性がある。

群逐次デザインや症例数の変更を伴うデザインに対して議論したこと以外にも、多くの重要な考慮事項がある。第一は、アダプティブ・エンリッチメントデザインの場合、患者集団をアダプティブに変更することの提案は、ある特定の部分集団においてベネフィット-リスクプロファイルが最も良好であると示唆される過去(例えば、早期段階)の試験結果、及び/又は強い生物学的妥当性に動機づけられるべきである。第二は、治療効果を修飾すると考えられるベースライン特性が本質的に二値ではない場合、部分集団を定義するための1つ又は複数の閾値は、適切に正当化されるべきである。第三は、標的とすべき部分集団を同定するために、*in vitro* の診断機器を利用することもあり得るが、この場合、診断機器は十分な性能を有するべきである²⁰。最後に、標的でない部分集団における治療効果の特徴を明らかにするために、どの程度試験をデザインすべきかは、例えば、エンリッチメントに対する病態生理学的又は経験的根拠、薬剤の毒性、部分集団を定義づけるベースラインマーカーの分布、部分集団の定義に用いる閾値の妥当性、及び承認が標的部分集団に限定された場合の標的でない部分集団における適応外使用の可能性等、多くの要因に依存するであろう。

D. 治療群の選択に対するアダプテーション

もう1つのアダプティブなアプローチは、中間解析に基づく比較結果に基づいて臨床試験に含める治療群の変更を事前に計画することである。治療群の変更には、治療群の追加又は除外が含まれる。この種のデザインは、早期の探索的な用量探索試験でしばしば用いられてきた。アダプティブ用量探索試験では、複数の用量群を設定して試験を開始し、比較データに基づく中間解析を実施して、継続して評価をするための用量群を選択する。その目的は、非アダプティブデザインよりも用量-反応関係の特徴を捉え易くするとともに、将来の検証的試験で評価するための最適な用量又は用量群の選択を可能にすることである。例えば、*continual reassessment method (CRM)* は、早期の相において、新薬の最大耐用量を確実かつ効率的に推定するために、観察された毒性に基づいて評価する用量をアダプティブに増量する手法である(Le Tourneau et al. 2009)。アダプティブ治療群選択は、有効性について実質的な根拠を提供することを意図した試験でも用いることができる。例えば、2つの用量のいずれか又は両方が好ましいベネフィット-リスクプロファイルを有すると考えら

²⁰ FDA の企業及びFDA スタッフ向けガイダンス「*In Vitro Companion Diagnostic Devices*(2014年8月)」及びFDA の企業及びFDA スタッフ向けドラフトガイダンス「*Principles for Codevelopment of an In Vitro Companion Diagnostic Device with a Therapeutic Product* (2016年7月)」を参照のこと。本ガイダンスが最終化されれば、本トピックに関するFDA の最新の考え方を示すことになる。ガイダンスの最新バージョンは、FDA ガイダンスのWeb ページ (<https://www.fda.gov/RegulatoryInformation/Guidances/default.htm>) を確認すること。

れる状況では、いずれかの用量を早期に中止できる逐次解析を含むアダプティブデザインは、非アダプティブデザインよりも効率的にその科学的な目的を達成することができる。このようなアダプティブデザインであれば、原理的には、登録される追加の患者数（症例数）や各治療群への割付比等、デザインのその他の側面を中間で変更することも可能かもしれない。

有効性について実質的な根拠を提供することを意図した試験では、統計学的仮説検定手法は、試験で評価された複数の用量の中から最良の用量又は用量群をアダプティブに選択することだけでなく、試験を早期に中止できる可能性や症例数の変更等のいくつかのアダプティブに変更する可能性を考慮すべきである。複数の用量群が設定され、有効性又は無益性のために特定の用量群への患者登録が中止可能な中間解析を伴う、単純な試験デザインの場合、複数の用量を評価する際に第一種の過誤確率を制御するための多重検定法とともに、典型的な群逐次検定法を用いることができる。もし、症例数の変更等、他のアダプテーションを許容しているデザインの場合には、症例数や対象集団のアダプテーションに関して述べた方法を用いるべきである。他のアダプティブデザインと同様に、事前の計画は重要であり、仮説検定の方法だけでなく、治療群を選択するための規則やそれ以外に生じ得るあらゆる中間での変更も事前に規定すべきである。一般に、III章で概説した原則に従えば、用量選択と（試験全体のデータに基づく）選択された用量の有効性の検証の両方を組み込んだシームレスデザインを検討することもできる。

アダプティブ治療群選択の特殊なケースは、ある1つの疾患を対象として、複数の試験治療を適切な1つの対照治療と比較するためにデザインされた、アダプティブプラットフォーム試験の状況で起こる（例えば、Woodcock and LaVange 2017）。効率向上のためにしばしば組み込まれるこれらの試験の2つの特徴は、共通の対照群を使用することと、中間解析時に有望な治療を選択するために事前に計画されたアダプテーションを使用することである。これらの試験には、複数の試験依頼者から提供される試験薬が含まれている可能性や、試験期間が未定のまま実施されている可能性があり、複雑なアダプテーションを伴うことが多い。このため、一般にFDAとの広範な協議を行うべきである。

E. 患者割付に対するアダプテーション

本項では、患者の割付に対する2種類のアダプテーションについて述べる。1つは、比較を伴うベースライン特性データに基づくアダプテーションであり、もう1つは比較を伴うアウトカムデータに基づくアダプテーションである。1つ目のタイプは、共変量に基づくアダプティブランダム化である。これは、患者に対する治療の割付が、その患者のベースライン特性、及びその患者よりも以前に登録された患者のベースライン特性と治療の割付に部分的又は完全に依存した割付方法である。このようなアプローチは、ベースライン共変量の分布が治療群間でバランスのとれた割付を促進するために用いられる。共変量に基づくアダプティブランダム化のよく知られた例として、最小化法がある（Pocock and Simon

1975)。これは、潜在的にアウトカムに影響する共変量の分布の治療群間差が最小となるようにそれぞれの患者を治療に割付ける方法である。共変量に基づくアダプティブランダム化の方法を用いても、適切な方法（一般には、ランダム化に基づく検定や並べ替え検定）で解析すれば、第一種の過誤確率は増大しない。これらの手法は、単純なランダム化の方法と比べて、割付けられる治療の予見可能性を高める恐れがあるが、この予見可能性は、（患者割付のプロセスに）ランダムな要素を追加することで、割付けられる治療を完全には特定できないようにすることが可能である。

2つ目のタイプは、反応に基づくアダプティブランダム化である。このアダプティブランダム化の特徴は、新たに登録された被験者がある治療群に割付けられる確率は、以前に登録された被験者の蓄積されたアウトカムデータに基づいて、試験の経過とともに変化するというものである。反応に基づくアダプティブランダム化の手法は様々なものがあり、そのうちのいくつかは **Play the winner** デザインなどと呼ばれるものである。ときには、反応に基づくアダプティブランダム化を用いる理由が、統計学的、倫理的、実用的なあらゆる側面から与えられることがある。統計学的には、反応に基づくアダプティブな手法は、ある状況下において検定統計量の分散を最小化することで、試験期間の短縮、症例数の低減、及び／又は統計学的検出力の増大をもたらすことができる。また、反応に基づくアダプティブランダム化の倫理的な理由は、このデザインが、有望な治療群により多くの被験者を割り付けられる可能性があるということである。最後に、実用面での理由は、このようなデザインの特徴を有する臨床試験は、潜在的な試験参加者にとって魅力的であり、これにより、被験者の登録を促進し、容易に集積できる可能性があることである。ただし、反応に基づくアダプティブランダム化にはまだ議論の余地があり、一部の研究者は、確定的でない中間解析結果を実施中の試験のランダム化の方法を変更するために用いるべきではないと考えていること、及び／又は2群の試験では、ランダム化における割付比の調整を正当化できるほど、統計的効率は改善されないと考えていることに留意すべきである (Hey and Kimmelman 2015 及び付随する解説)。

一般に、反応に基づくアダプティブランダム化のみでは、適切な統計解析手法を用いる限り、試験の第一種の過誤確率は増大しない。解析方法が試験デザインを適切に考慮していることを保証することが重要である。最後に、アウトカムデータに基づく他の多くのアダプティブな手法と同様に、反応に基づくアダプティブランダム化は、アウトカムが比較的短期間で確認できる試験で最もよく機能する。

F. 評価項目の選択に対するアダプテーション

これは、中間解析に基づく比較結果に基づき、主要評価項目をアダプティブに変更するデザインである。このようなデザインを用いる動機は、FDA が主要評価項目として受け入れ可能と考える、複数の患者アウトカムに対する治療効果の大きさが不確実であることかもしれない。他のアダプティブデザインと同様に、アダプテーションルールを事前に規定

すべきであり、統計学的仮説検定手法はアダプティブな評価項目の選択を考慮すべきである。評価項目の選択に関しては、重要な臨床的考慮事項が含まれるため、このようなデザインを検討する場合は、FDA の審査部門と早期に協議することが推奨される。

G. 複数のデザイン要素に対するアダプテーション

本ガイダンスで議論されている 2 つ以上のアダプティブデザインの特徴を組み合わせると、臨床試験がより複雑になる可能性がある。試験デザインが複雑になったとしても、単純なアダプティブデザインと同様の一般的原則が適用される。複数のアダプティブデザインの特徴を組み合わせた場合、第一種の過誤確率やそれ以外の試験の動作特性を推定することは、特に困難となるおそれがある。試験デザインの特徴を評価するためには、臨床試験のシミュレーション (VI.A.項) が、しばしば重要となる。

VI. 特別な考慮事項とトピック

A. アダプティブデザイン計画におけるシミュレーション

臨床試験シミュレーションは、一般に臨床試験の計画とデザインにおいて重要な役割を果たすことが多く、アダプティブ臨床試験では特に重要である。例えば、シミュレーションは中間解析の回数と時期の選択や、有効性又は無益性の検定統計量の適切な棄却限界値を決定するために用いることができ、代わりとなる別の試験デザインと性能を比較する場合にも有用であろう。アダプティブ臨床試験におけるシミュレーションの主な用途は、試験の動作特性を推定し、動作特性が目標とする水準を満たすことを示すことである。

従来の非アダプティブ臨床試験では、一般に第一種の過誤確率が目標とする水準に制御されることを保証し、試験の検出力の推定値を得るために、統計理論に依拠してきた。最も単純なケースでは、固定症例数の臨床試験デザインで単一の評価項目を検定する場合、通常は最終解析時の検定統計量が特定の確率分布に漸近的に従うことを示すことができ²¹、推論と動作特性はその分布の特性に基づきことができる。同様に、従来の群逐次デザイン等、多くのアダプティブデザインでは、漸近確率分布を数学的に導出し、それらの分布に基づいて推論と計画を行うことができる。

しかしながら、一部のアダプティブデザインでは、検定統計量の適切な分布を導出することができないか、検定統計量の分布自体が計算上扱いにくい。これは、複数の要素のアダプテーションを行うデザインや、解析時点を決定するために予測確率モデルを用いるデザイン等、より複雑なアダプティブデザインの場合に当てはまる傾向がある。このような場合、試験の動作特性は、しばしば臨床試験シミュレーションによって推定することができる。例えば、第一種の過誤確率と検出力の場合、このアプローチの基本的なロジックは、さまざまな仮定における多くの試験の例をシミュレーションにより発生させ、それぞれの

²¹ 検定統計量の漸近分布とは、症例数が大きくなった場合の検定統計量の漸近的な確率分布である。

仮定の下で、事前に定めた有効性の結論を支持する基準を満たしたシミュレーションの割合を評価することである。

第一種の過誤確率を推定することを目的としたシミュレーションでは、帰無仮説に合致した一連の仮定における臨床試験をシミュレーションにより発生させるだろう。このような一連の仮定における、誤って有効であると判断されたシミュレーションの割合は、第一種の過誤確率の推定値として扱われる。ほとんどの場合、帰無仮説に合致すると考えられるシナリオは無数にある。第一種の過誤確率を推定する際に考慮すべきシナリオの特定は、困難な場合があり、医学的考察と数学的考察の両面に依拠するかもしれない。

これらのシナリオには、局外パラメータに関するさまざまな仮定が含まれる場合がある。局外パラメータには、症状尺度の分散や対照群における反応確率等の統計的パラメータや、被験者の試験への組み入れ速度等の運用パラメータを含めることができる。例えば、生存率が非常に低い癌（例えば、生存期間の中央値 6 か月）を適応症とした、試験治療とプラセボの 2 年間の死亡率を比較する試験を考える。帰無仮説を 2 群における死亡率が等しいこととすると、この帰無仮説に合致した可能性のあるシナリオには、各治療群における死亡率が 5%、50%、99%、99.01%等の場合が含まれる。帰無仮説に合致するすべてのシナリオをシミュレーションにより発生させることは不可能であるが、可能性のある偽陽性の範囲を適切に表す、限定したシナリオを決めることは可能であるかもしれない。この例では、医学専門家は、文献や疾患の臨床経験に基づいて、例えば、2 年間のプラセボ死亡率が 75% 未満のシナリオを除外することに満足するかもしれない。第一種の過誤確率を推定するためにどのシナリオをシミュレーションにより評価する必要があるかを決定する際には、数学的な考察も重要な役割を果たす可能性がある。単調性に基づいて、ある特定のシナリオでは、他のシナリオよりも第一種の過誤確率が必ず小さいと言えるかもしれない。

多くの場合、臨症的及び数学的に考慮しても、すべての帰無仮説について第一種の過誤確率を推定することはできない。妥当な値のグリッド上でシミュレーションを実行し、シミュレーション結果から総合的に判断して、第一種の過誤確率の最大値が検討した範囲では目標とする水準を超えていないだろうと主張することが一般的である。上記の例では、プラセボ群と試験治療群の死亡率が 75、80、85、90、95、及び 99 パーセントとしてシミュレーションを実行するかもしれない。これらの各シナリオにおける第一種の過誤確率の推定値が 0.025 未満である場合、プラセボ群の死亡率が 75~99%であるすべてのシナリオについて、第一種の過誤確率が適切に制御されていることを示す十分な根拠と考えられるだろう。しかしながら、どのようなアプローチを用いたとしても、試験終了時の評価では、局外パラメータに関する蓄積された情報に照らして、適用する統計的推論が適切であり、結論を正当化されるかどうかを検討すべきである。この例では、観察されたプラセボ死亡率が予想に反して 50%であった場合、追加のシミュレーションが必要となるだろう。

別の複雑な要因は、複数の評価項目の存在である。試験依頼者が複数の臨床評価項目を検定し、これらすべての評価項目全体でファミリーワイズの第一種の過誤確率を制御した

い場合、帰無仮説下のシナリオにおける各被験者のすべての評価項目のシミュレーションを実行する必要があるが、このためには複数の評価項目の相関構造の知識が必要となる可能性がある。これは、臨床試験シミュレーションで対処するにはあまりにも複雑な問題になる可能性がある。しかしながら、場合によっては、複数の評価項目間の独立性を仮定すると、第一種の過誤確率の上限が得られるということもできる。これは、例えば、複数の検定を制御するために Bonferroni 又は Holm のアプローチを用いる場合が該当する²²。

シミュレーションの動作特性の精度を考慮することが重要であるが、これはシミュレーション回数（繰り返し回数）に依存する。シミュレーションの繰り返し回数は、提案された臨床試験デザインの理解とレビューを容易にするのに十分な回数にすべきである。例えば、シナリオごとに繰り返し回数を 100,000 回とすると、第一種の過誤確率の 95% 信頼区間の幅が約±0.1%になることが保証され、ほとんどの場合は十分な回数となる。これにより、第一種の過誤確率の推定値のごくわずかな差を特定でき、場合によっては重要となるかもしれない。一般に、シミュレーションのシナリオごとに異なる乱数シードを用いることも推奨され、これによりシナリオ間で異常な結果を常に回避することができる。また、第一種の過誤確率の評価は、より少ない繰り返し回数で十分な場合がある。例えば、シナリオの特に細かいグリッド上を探索し、すべてのシナリオで第一種の過誤確率の推定値が目標とする水準を下回っている場合は、10,000 回の繰り返し回数で十分であろう。また、第一種の過誤確率の推定値の 95% 信頼区間の上限が目標とする水準を下回っている場合は、一般により少ない回数のシミュレーションを行うことができる。

複雑なアダプティブデザインでは、検出力やその他の動作特性（期待症例数、期待試験期間、治療効果の推定値のバイアス等）を推定するために、臨床試験シミュレーションを用いることができる。これらの推定値には、第一種の過誤確率の推定値と同様の考え方が適用される。しかしながら、その他の動作特性では、第一種の過誤確率の推定に期待されるほどの精度は必要とせず、より少ない繰り返し回数でより広範なシナリオについて検討することが適切である。

B. ベイズ流アダプティブデザイン

ベイズ流アダプティブデザインという用語は、さまざまな方法でベイズ流の統計的推論及び／又は計算を用いる、多種多様な臨床試験デザインを指す用語として用いられている (Berry et al. 2010)。ベイズ流アダプティブデザインの特徴の例は以下の通りである。

²² Bonferroni 法、Holm 法、及びその他の複数の検定アプローチに関するさらなる議論は、FDA の企業向けドラフトガイダンス「Multiple Endpoints in Clinical Trials (2017 年 1 月)」にある。最終的に、このガイダンスは、このトピックに関する FDA の最新の考え方を表している。

- 中間解析の時期と決定規則を規定するために、場合によっては試験の外部情報を組み込んだ、予測統計モデリングの利用
- 用量漸増及び用量選択を規定するために、想定される用量反応関係の利用
- 試験の効率を向上させるために、情報のある事前分布を用いた、過去の試験、自然史研究、レジストリ等、外部の情報源からの情報の明示的な借用
- 試験の成功基準の設定するために、事後確率分布の利用

一般に、ベイズ流アダプティブデザインには、ベイズの特徴を持たないアダプティブデザインと同じ原則が適用される。ベイズ流アダプテーションを用いる試験デザインは、医薬品の有効性の結論を裏付けるために、頻度論流又はベイズ流の推測方法に基づく場合がある。頻度論流推論は、既知の検出力と第一種の過誤確率を用いて行われる仮説検定によって特徴付けられ、無情報事前分布に基づくベイズ流の計算手法に加えて用いられることが多い。ベイズ流推論は、医薬品が有効である事後確率に基づいて結論を直接導くという特徴があり、頻度論流推論と重要な違いがある（Berger and Wolpert 1988）。明確に外部情報を借用する試験等、情報のある事前分布を伴うベイズ流推論を用いる試験でのベイズ推測の特性は第一種の過誤確率よりも多くの情報をもたらす。FDAの企業向けドラフトガイダンス「Interacting with the FDA on Complex Innovative Clinical Trial Designs for Drugs and Biological Products（2019年9月）」は、ベイズ流推論を用いた試験デザイン案のレビューを円滑に進めるために、FDAに提出すべき情報に関する推奨事項を示している。

多くのベイズ流アダプティブデザインに共通する特徴の1つは、試験の動作特性を推定するために、シミュレーション（VI.A.項）を用いることである。多くのベイズ流の方法自体が広範な計算（マルコフ連鎖モンテカルロ（MCMC）法やその他の方法）に依拠しているため、ベイズ流アダプティブデザインでは、試験のシミュレーションのために、特に多くのリソースを要する可能性がある。

C. イベント発現までの時間を評価項目とした場合のアダプテーション

死亡までの時間や腫瘍反応までの時間等、特定のイベントが発生するまでの時間を主要評価項目とするアダプティブ臨床試験では、さらに考慮すべき事項がある。このような試験では、検出力は症例数ではなくイベント数に依存することから、ある定めた症例数ではなく、イベント数を目標とすることが一般的である。このような試験における症例数の変更は、イベント数の変更を目的とするため、症例数、各被験者の追跡期間又はその両方を変更する可能性がある。さらに、イベント発現までの時間を評価項目とした中間解析では、代替アウトカム又は中間のアウトカムに関する情報を利用する場合があり、このようなアプローチを用いるためには、解析時に適切に考慮する必要がある（さらなる議論は次項を参照）。

D. 潜在的な代替評価項目又は中間評価項目に基づくアダプテーション

ほとんどのアダプティブデザインは、1つ又は複数の主要評価項目の試験中のモニタリングに依拠する。しかしながら、主要評価項目と相関する潜在的な代替評価項目又は中間評価項目²³が存在し、主要評価項目自体の確認が困難である又は評価に時間を要する場合、アダプティブデザインは潜在的な代替評価項目又は中間評価項目に基づく可能性がある。例えば、癌の治療法に関する試験で、主要評価項目が全生存期間であり、生存期間の中央値が2年を大きく超え、腫瘍反応（例えば、完全奏効又は部分奏効）が臨床的ベネフィットを予測すると予想される場合を考える。このような場合、死亡率ではなく、腫瘍反応に基づいて、症例数再推定やその他のアダプテーションを計画することが賢明であるかもしれない。有効性の最終評価は、依然として主要評価項目（この例では全生存期間）に基づいて行われる。同様に、主要評価項目が6か月後の患者の症状である場合、アダプティブデザインは2か月後の患者の症状に基づくかもしれない。潜在的な代替評価項目又は中間評価項目と主要評価項目の関連性にある仮定を設定するアプローチがあるが、このようなアプローチを用いる場合は、関連性の誤指定に伴う第一種の過誤確率やその他の動作特性の評価への起こり得る影響を考慮すべきである。また、潜在的な代替評価項目又は中間評価項目と主要評価項目の関連性の仮定に依存しないアプローチも存在する（Jenkins et al. 2011、Irlle et al. 2012、Magirr et al. 2016）。

イベント発現までの時間又は経時的アウトカムを伴うアダプティブ臨床試験では、中間解析で代替評価項目又は中間評価項目の情報を用いた場合、適切な統計解析手法を用いない限り、有効性の誤った結論を導く可能性が高くなる可能性がある。例えば、イベント発現までの時間を用いた試験では、将来のイベント時間の予測に有用な、中間解析時におけるイベントが観察されていない被験者の代替情報を用いると、第一種の過誤確率の増大につながる可能性が指摘されている（Bauer and Posch 2004）。中間解析に基づく比較結果へのアクセス制限や、主要評価項目のみに依存するアダプテーションルールの事前規定等の追加の予防手段は、そのような計画外のアプローチが実行されなかったという信頼を高めるのに役立つ。さらなる議論は VII 章を参照。

E. 副次評価項目

ほとんどの臨床試験では、主要評価項目に加えて1つ以上の副次評価項目が規定されており²⁴、アダプティブデザインはこれらの副次評価項目の解析に影響を与える可能性がある。

²³ 本ガイダンスの目的では、潜在的な代替評価項目とは、候補となる代替評価項目、合理的な代替評価項目、又は検証済みの代替評価項目を指し、中間評価項目は中間の臨床評価項目を指す。これらの追加の用語の定義については、「The Biomarkers, EndpointS, and other Tools (BEST) Resource glossary for definitions of these additional terms」を参照。

²⁴ 臨床試験における複数の評価項目の評価における一般的な考慮事項の議論については、FDA の企業向けドラフトガイダンス「Multiple Endpoints in Clinical Trials (2017年1月)」を

群逐次デザインを考えると、主要評価項目に対する複数の解析により、第一種の過誤確率が増大し、主要評価項目の治療効果の推定値にバイアスが生じる可能性があることは広く理解されている。しかしながら、第一種の過誤確率が増大と推定値のバイアスが、主要評価項目と相関している任意の評価項目にも生じる可能性があることは、あまりよく理解されていない (Hung et al. 2007)。臨床試験のほとんどの副次評価項目は主要評価項目と相関しており、しばしば非常に高い相関がある。群逐次アプローチ等の一部のデザインでは、アダプテーションのために副次評価項目の解析を調整する方法が存在する (Glimm et al. 2010)。このような調整を行わない場合、副次評価項目の結果の解釈には十分注意すべきである。

F. 安全性の考慮事項

アダプティブ臨床試験の計画は、しばしば有効性の実証を目的としたアウトカムに焦点が当てられるが、安全性の目的としたアウトカムも重要な役割を果たす。第一に、アダプテーションが有効性評価項目ではなく安全性評価項目について計画される場合がある。1つの例は、CRM 又は他のアダプティブな手法を用いて最大耐用量の同定を目的とした、癌領域における早期段階の用量探索試験である。もう1つの例は、2006年にロタウイルスワクチンであるロタテックの承認の主要な根拠となった、ロタウイルスワクチンの有効性及び安全性の臨床試験 (REST) である (Heyse et al. 2008)。REST は、重度の胃腸障害である腸重積症のリスクを評価するために、最大 100,000 人の乳児を対象とした群逐次試験であり、そのうち一部の乳児が有効性評価の対象とされた。

第二に、通常、有効性評価項目に対してアダプテーションを行う試験において、医薬品の承認を裏付けるために十分な安全性情報を収集することは、主要な関心事項である。有効性に関する有力なエビデンスにより早期中止した試験においても、試験薬の信頼できるベネフィット・リスク評価を可能とし、ラベリングに記載するために、十分な安全性データを収集する必要がある。このため、中間解析の回数、時期及び中止境界を計画する場合、安全性データベースのサイズを考慮すべきである。特に、中間解析の時期は、信頼性の高い安全性評価を確実にするために、期待される最小の患者数と最小の曝露期間によって制限される可能性がある。

最後に、ある特定のアダプテーションにより、試験参加者を不必要なリスクに晒される可能性があるかどうかを検討することが重要である。これは、特に早期段階の用量漸増試験で懸念事項となる可能性がある。連続したコホートの被験者が、急速に増量した用量の投与を受けることを許容したアダプテーションルールは、より緩やかな増量を伴うデザインによって回避されたであろう、安全でない高用量の投与を被験者が受ける可能性がある。これは、試験薬が遅発性の作用をもち、重篤な有害事象の可能性がある場合に特に当ては

参照。最終的に、このガイダンスは、このトピックに関する FDA の最新の考え方を示している。

まる。このため、アダプティブ用量漸増試験で特定のアダプテーションルールを選択する際には、増量のスピードを考慮すべきである。

G. 早期探索的試験におけるアダプティブデザイン

医薬品開発における探索的試験は、医薬品を研究のための最善の方法（例えば、用量、レジメン、集団、併用治療法又は評価項目の選択）を後で決定する際の指針となる、医薬品使用の幅広い側面に関する情報を得ることを目的としている。これらは医薬品の効果の異なる側面を順次検討する連続した独立の早期の試験である場合や、複数のさまざまな側面を同時に評価しようとする、より複雑な試験となる場合もある。アダプティブデザインによってもたらされる柔軟性は、幅広い選択肢のある早期の評価を可能とすることにより、探索的な開発段階に特に有用であろう。開発早期の臨床試験にアダプティブデザインを適用して、用量、曝露、薬力学、患者の反応のばらつき又は反応の修飾因子のさまざまな側面について学ぶことは、デザインを改善し、開発後期の相の臨床試験が成功する機会を試験依頼者に提供することになる。

一般に、探索的試験では、統計的厳密さと動作特性に関して、有効性の実質的なエビデンスの提供を意図した試験と同様の規制からの期待は求められていないものの、探索的試験から誤った結論を導く可能性を認識することは依然として重要である。例えば、探索的に複数用量を比較する試験の不備により、後続の検証的試験に対して最適な用量が選択されず、結果として有効性が示されない、あるいは不必要に過剰な毒性が発現する可能性がある。したがって、探索的試験に関して、アダプティブ臨床試験デザインの原則に従うことにより、開発プログラムに悪影響を与えるリスクを低減することができる。

H. 中間解析の比較結果に基づく計画外のデザイン変更

中間解析に基づく比較結果で試験データを検討する場合、アダプテーションの根拠として事前に計画されていないデータ解析から、ある特定のデザイン変更（例えば、解析対象を特定の部分集団に限定、治療群の除外、症例数の調整、主要評価項目の変更又は解析方法の変更）が倫理的に重要であることや、最終解析結果で統計学的に有意な可能性が高くなりうることが思いがけず示唆される場合がある。例えば、複数の治療群のうちの1つの群における想定外の治療アドヒアランスの欠如は、その治療群の除外につながるかもしれない。このような事前に計画されていない解析に基づく改訂は、第一種の過誤確率の制御や試験結果の解釈を困難にする可能性がある。試験依頼者は、患者の安全性が損なわれないからといって、検討中している変更についてFDAに事前に相談することなく、そのような変更を実施することは全く推奨されない。

I. 試験外部の情報源からの情報に基づくデザイン変更

医薬品開発プログラムの実施中に、実施中の試験以外で発生する予測できないイベントは、実施中の試験に関連する重要な新しい情報を提供し、試験デザインの改訂の動機とな

る可能性がある。例えば、（おそらく異なる患者集団を対象とした）他の試験から得られた予期しない安全性情報、疾患の病態生理学又は疾患のサブタイプを識別する患者特性に関する新しい情報、医薬品の薬物動態又は薬力学的反応に関する新しい情報、あるいは試験デザイン時に知っていれば、別の試験デザインにつながったかもしれない、その他の情報が存在する場合がある。このような場合、例えば、既存の試験を終了して、変更したデザインで新しい試験を開始するのではなく、何らかの方法で試験デザインを改訂する理由があるかもしれない。安全性に重大な懸念がある場合、特に大規模な試験では、試験を継続させるために試験デザインを改訂することが重要となる場合がある。試験の外部情報のみに基づく、十分に動機付けられたデザイン変更は、統計的推論の妥当性に影響を与えず、多くの場合、FDA に受け入れ可能と考えられるだろう。実際は、試験依頼者が中間解析に基づく比較結果を完全に盲検化されている場合を除いて、試験変更の決定が完全に外部情報に基づいていることを保証することは非常に困難であり、このことは中間解析に基づく比較結果へのアクセス制限が非常に重要である理由の1つである。さらなる議論は VII 章を参照のこと。

VII. 試験の完全性の維持

一般に、中間解析に基づく比較結果へのアクセスは、試験の実施又は管理を担当する者から独立し、かつ結果を知っておく必要がある、適切な専門知識を有する者に制限することが強く推奨される。患者、試験責任医師及びそのスタッフ、並びに試験依頼者の関係者が中間解析に基づく比較結果にアクセスできないことを保証することは、2つの重要な目的を果たす。第一に、予定外のデザイン変更の可能性は、決して蓄積されたデータによるものではないという最大の信頼を提供する。例えば、試験管理担当者が中間解析に基づく比較結果を知ることにより、試験の外部情報が十分な動機づけとなっていると思われる試験実施計画書の改訂が、蓄積された比較データへのアクセスによって何らかの影響を受けているかどうか、規制当局が判断することを困難にする可能性がある。デザインの変更が中間解析に基づく比較結果の影響を受けた可能性があると考えられる場合には、誤った結論を導く可能性を制御し、かつ信頼できる推定値を導くための適切な統計手法が、知られていないか、実施が困難であるか、又は試験の効率を大幅に低下させる可能性がある。

第二に、中間解析に基づく比較結果へのアクセスの制限は、試験実施の質を最も高めることにつながる。試験責任医師が蓄積されたデータを知ることにより、患者の組み入れ、アドヒアランス、参加継続又は評価項目の評価に悪影響を及ぼし、試験の目的をタイムリーに確実に達成するための臨床試験の能力が損なわれる可能性がある(Fleming ら 2008)。試験実施に関する問題は予測が困難であり、統計解析で調整することは一般に不可能である。したがって、アダプティブ臨床試験では、厳密に計画を立案し、慎重に試験を実施するとともに、中間解析に基づく比較結果の機密性を維持し、試験の完全性を保証するためのアプローチを記載した包括的な文書を作成すべきである。

アダプティブ臨床試験において、試験依頼者に対する比較データへのアクセス制限を実装するための、複数の適用可能なモデルが存在する。DMC が存在する場合には、DMC とは別に、アダプテーション専門の独立した組織を設置することができる。あるいは、DMC の主な責務は依然として、患者の安全性と試験の完全性を保証することとし、アダプテーションに関する意思決定の役割を DMC に与えることもできる²⁵。この後者のモデルは、群逐次デザインや単純なアダプテーションアルゴリズムを有する、複雑でないアダプティブデザインに最適であろう。どちらのアプローチも、それぞれを支持する論拠がある。例えば、別々の組織を設置することにより、各委員会により関連した専門的意見を取り入れることが容易になり、DMC がその主な責務に最も効率的に注力することが可能になるであろう。一方、DMC のような単一の組織を双方の目的で設置することにより、複数のモニタリンググループ間で情報を共有し、協議するといった実施上の課題が回避される。

選択されたアプローチにかかわらず、アダプテーションに関する勧告の任務を負う委員会は、アダプテーションに関する方法論、データモニタリング計画、及び決定規則に精通した一人又は複数の統計家を含む、適切な専門知識を有する委員で構成されるべきである。さらに、本委員会の責務は、中間解析に基づく比較結果をレビュー後に、変更すべき潜在的なデザインの側面を特定することではなく、慎重にデザインされ、かつ事前に規定されたアダプテーションの計画を適切に実施した上で、アダプテーションに関する勧告又は決定を行うことである。したがって、DMC 及び/又はアダプテーション委員会は、デザイン段階において、仮想的なシナリオ及びアダプテーション計画に規定された施策がすべての関係者から妥当であると考えられるかどうかに関する、試験依頼者との広範な協議に関与することが重要である。

中間解析結果を準備し、DMC 又はアダプテーション委員会へ報告する責任をもつ者と、試験の管理及び実施を担当する者は、試験依頼者の組織の中か、開発業務受託機関（CRO）のような別の組織か、又はその双方に所属しているかどうかにかかわらず、物理的にかつ試験実施の手順上分離されていることを保証するための予防手段を講じておくべきである。これを実施することは、試験の日常的な管理及び実施を担当する者が、たとえ不注意であっても、試験治療の割付又は比較結果にアクセスできないことを保証するのに役立つであろう。同様に、上記の理由から、一般に中間解析結果の詳細は、DMC 又はアダプテーション委員会から試験依頼者が受け取る勧告から除かれるべきである。

一般に、すべての試験依頼者の代表者が中間解析に基づく比較結果にアクセスできないことが推奨されるが、特定の試験依頼者の関係者がアクセスできることについて、正当化できる場合がある。例えば、一部のアダプティブ臨床試験では、用量選択等の決定を伴う

²⁵ 臨床試験における DMC の役割、責任及び実施手順の詳細な議論については、FDA の試験依頼者向けガイダンス「Establishment and Operation of Clinical Trial Data Monitoring Committees (March 2006)」を参照のこと。

可能性があるが、このような決定は、非アダプティブ臨床試験の場合、通常、試験依頼者の責任で行われ、医薬品開発プログラムにおいて長期的に重要な影響を持つものである。試験依頼者の関係者への限定的なアクセスは、以下のような条件を満たせば正当化できるかもしれない。例えば、数名の試験依頼者の代表者が関与している場合、アクセスが許可された試験依頼者の関係者は、データにアクセスできる目的以外は試験の実施又は管理に関与しておらず、かつ患者、試験責任医師及び試験運営委員会等、他の主要な関係者に対して中間解析に基づく比較結果が知らされないことを保証するための適切な手順が整備されている場合である。しかしながら、試験の完全性に対するリスクは、試験依頼者の中間解析に基づく比較結果へのアクセスを完全に制限することで最も容易に最小化することができ、これはほとんどの場合、デザイン段階における試験依頼者と DMC 又はアダプテーション委員会との広範な計画と協議を通じて達成できる可能性が高い。

アクセスの適切な制限のためには、機密性を維持し、検証するための慎重に計画された手順、並びにモニタリングの文書化及び業務手順書の遵守が必要である。このアプローチには、通常、中間データにアクセスする者に対する秘密保持契約の利用、治療の割付を推測できる情報を含むあらゆるデータに、試験依頼者のアクセスを防止する試験実施の手順又は物理的なファイアウォールの利用、誰が機密性のあるデータにアクセスできるか、いつアクセスが生じるか、及びどのような種類のデータや結果が含まれるかを規定するデータアクセス計画の作成及び利用が含まれる。重要な文書については、第 VIII 章でより詳細に述べる。

アダプティブ臨床試験では、アダプテーションに関する決定を知ることにより、中間解析結果に関する情報が伝わる可能性がある。例えば、症例数変更アルゴリズム及びアダプティブに選択された症例数を知ることにより、中間解析における治療効果の推定値を逆算することが可能となる。したがって、知る者が推測できる情報を最小限に抑えるために、可能な限りの手段を講じるべきである。アダプテーションに関する規則の事前規定は依然として重要であるが、試験実施計画書は一般的なアプローチのみを概説し、具体的なアルゴリズムの詳細は、DMC charter や adaptive design charter 等、より少数の関係者が利用できる特定の文書に記述することが可能であろう。中間解析後に伝えられる情報の範囲に関して、慎重に検討し、計画することも重要である。一般に、試験責任医師及び試験参加者は、アダプティブな変更の内容を可能な限り知らされないようにすべきである。例えば、中間解析後に症例数が追加された場合、試験実施医療機関に、特定の目標とする最終症例数ではなく、目標登録例数に到達していないことを知らせることができるだろう。連続的なアダプテーション決定閾値ではなく、離散化したアダプテーション決定閾値を用いることは、試験の完全性に対するリスクを最小化するために推測できる知識を制限するための別の可能なアプローチである。

VIII. 規制上の考慮事項

A. FDA との協議

試験依頼者と FDA との協議の目的と内容は、開発の段階によって異なる。アダプティブ臨床試験の複雑さや動作特性に関する不確実性が増すことにより、通常よりも早期に、かつより広範な協議が必要となる場合がある。医薬品開発の早期段階では、FDA による試験実施計画書のレビューは、通常、薬理学的作用や有効性に関する推測の妥当性よりも、試験参加者の安全性に焦点を当てる。しかしながら、リソースが許される場合、FDA は、後続の試験のデザインの指針とするために収集される情報の適切性を検討するために、探索的試験の試験実施計画書をレビューする場合がある。早期の探索的試験におけるアダプティブデザインの要素に関する相談事項のある試験依頼者は、FDA にそれらの相談事項を検討するための会議（又は書面による回答のみ）を要望し、FDA に見解を求めるべきである。例えば、アダプティブ臨床試験の計画に関する議論は、Type C meeting を要望する根拠となり得る。開発の早期段階において、このような要望に対応できる FDA の能力は限られている可能性があり、競合する業務の優先順位や開発プログラムの詳細に左右される。

開発の後期段階では、FDA は、規制上の決定のための、十分に信頼性のある試験結果が得られることを保証するために、デザイン及び解析計画の評価においてより広範な役割を有する。後期臨床試験に関して、FDA から正式で実質的な見解を得るための規制上の仕組みは、十分に確立されており、例えば EOP2 meeting が含まれる。医薬品及びその使用目的に関する既存の知識、並びにアダプティブの要素の性質によっては、EOP2 meeting が、試験依頼者が見解を得るための適切な場となる場合もあれば、より早期の FDA との協議が望ましい場合もある（例えば、Type C meeting 又は EOP2A meeting）。より早期の協議により、医薬品の開発を遅らせることなく、議論を繰り返す時間を確保するのに有用となり得る。

複雑なアダプティブデザインに関する FDA のレビューでは、しばしばデザインの動作特性に関する困難な評価が含まれ、通常は広範なコンピュータ・シミュレーションが必要であり、専門家と FDA の審査部との間でその評価に関する議論が活発に行われる。このため、FDA が、このようなデザインを短期間で適切にレビューすることが困難になる可能性がある。Special protocol assessments (SPA) に関するタイムライン（45 日間の回答）及びコミットメントを考慮して、複雑なアダプティブデザインの試験については、提案された試験とそのデザインに関して FDA と試験依頼者との間で事前に広範な議論が行われた場合にのみ、SPA の提出を推奨する。

提案された後期アダプティブ臨床試験に関する FDA のレビューには、デザイン及び解析計画が本ガイドラインに概説されている主要な原則を満たしているかどうかの検討が含まれる。特に、試験依頼者は、アダプティブデザインの詳細を事前に規定し、誤った結論が得られる可能性をどのように適切に制御しているか、治療効果の推定が分に信頼できるものであること、試験の完全性が適切に維持されることを説明するべきである。さらに、試

験依頼者が、計画の際に様々なアダプティブデザイン及び非アダプティブデザインの選択肢を検討し、提案したアダプティブデザインを選択した際の検討事項を FDA と協議することは、よい対処法である。

(通常、試験実施計画書の改訂を通じて行われる) 試験実施中の計画外の試験デザインの変更の提案については、FDA に相談すべきである。一方、一般に、事前に計画されたアダプテーションに関する意思決定には、FDA は関与しない。これは試験依頼者の責任であり、通常アダプティブデザインを実施するために設置された委員会 (DMC 等) によって行われる。モニタリング委員会の公開審議の議事録は、FDA から試験実施中に要求されることがあるが、非公開審議の議事録又は中間解析に基づく比較結果に関するその他のコミュニケーション又は情報は、患者の安全性が脅かされる特別な状況を除き、試験の終了まで機密性を保持すべきである。

B. アダプティブ臨床試験の実施前の文書

FDA による徹底した評価を可能とするためには、アダプティブデザインを適用した臨床試験の計画書は、非アダプティブデザインの計画書よりも必然的に複雑になる。ICH ガイダンス E9「臨床試験のための統計的原則」で議論されているような、非アダプティブ臨床試験の試験実施計画書及び統計解析計画書の通常の構成要素に加えて、アダプティブ臨床試験の開始前に FDA に提出される文書には、以下の内容を含めるべきである。

- 選択したデザインの根拠。他項で述べたように、代替のアダプティブデザイン及び非アダプティブデザインと比較して提案するデザインの重要な動作特性を評価することはよい対処法であり、そのような情報を FDA に提出することは有用であろう。しかしながら、デザインの最終的な選択は試験依頼者の責任である。
- 中間解析の予定回数及び予定時期、変更される可能性のあるデザインの具体的な側面、及びアダプテーションの決定を行う際に用いられる規則を含む、アダプテーション計画の詳細な説明
- DMC 及び/又は専門のアダプテーション委員会 (設置する場合) 等、アダプティブデザインの実施に責任を負う組織の役割に関する情報
- 中間解析結果の作成、アダプテーションに関する決定の指針、仮説検定、治療効果の推定、試験終了時の治療効果の推定値の不確実性の推定に用いる統計手法の事前計画。中間解析及び最終解析を実施するためのソフトウェアを予め記載しておくこと。最新のソフトウェア又はカスタムソフトウェアを用いる場合は、実行される統計的手順に曖昧さが無いことを保証するために、試験前に十分な情報を FDA に提出すべきであり、場合によっては、コンピュータコードが含まれる可能性がある。
- デザインの動作特性の評価と考察。これには通常、第一種の過誤確率、検出力、期待/最小/最大症例数、治療効果の推定値のバイアス、信頼区間の被覆確率を含め

るべきである。このような評価は、解析的計算及び／又はコンピュータ・シミュレーションによって実現されるかもしれない。動作特性が解析的に評価される場合には、その方法論に関する適切な詳細（例えば、参考文献又は証明）を提出すること。

- シミュレーションが、上記で定義した試験の動作特性を評価するための主要な又は唯一の手法である場合には、以下の内容を含む詳細なシミュレーション報告書を提出すること。
 - － 試験デザインの全体的な説明
 - － 試験開始時の症例数で肯定的な結果が得られる試験、最初の中間解析で無益性中止となる試験、症例数が追加され、肯定的な結果が得られる試験等、異なる結論を導く、少数の仮想的試験による例示
 - － シミュレーションシナリオに用いられるパラメータ設定値の説明。選択したパラメータ設定値が妥当と考える根拠を含む。
 - － 各シナリオについて評価したシミュレーションの試行（反復）回数と、この回数が妥当と考える根拠
 - － 多様なシナリオにおける動作特性の推定値を詳述したシミュレーション結果
 - － シミュレーションコード。FDAの審査官は、試験の動作特性を評価するために用いたシミュレーション実験を検証する必要があるため、シミュレーションに用いたソフトウェアパッケージを文書化することが重要であり、カスタムソフトウェアを用いた場合は、シミュレーションに用いたコードを提供することが重要である。コードを提出する場合は、コードは読みやすく、適切にコメントを加えておくべきであり、シミュレーション結果を生成するために用いられた乱数シードを含めるべきである。また、広く用いられている統計プログラミング言語で記述されたコードを提供することも有用である。（通常は計算効率の理由から）別の言語を用いてシミュレーション結果を生成した場合であっても、シミュレーションのレビューを容易にするために、広く用いられている統計プログラミング言語で実行可能なバージョンのコードを提供することが有用である。場合によっては、シミュレーションコードの使用に関する数式や説明等、詳細な追加情報を含めることが重要となる。
 - － 全体的な結論を示す要約
- 計画されたアダプテーションが行われる場合に、試験の完全性がどのように維持されるかを定義する、包括的なデータアクセス計画書。この文書には、次の情報が含まれるべきである。(1)中間解析を実施する者、(2)中間解析結果にアクセスできる者、(3)中間解析結果へのアクセスがどのように制御されるか、(4)アダプテーションの決

定がどのようになされるか、(5)アダプテーションの決定後に、どのような種類の情報が誰に伝えられるか。データアクセス計画には、どのような状況下で、どのような情報を試験依頼者又は試験責任医師に伝えることが許されるかを記載すべきである。さらに、試験依頼者は、データアクセス計画の遵守状況を評価する手順や、アダプテーションに関する決定を任される委員会（すなわち、DMC 又はアダプテーション委員会）の試験中に実施されたすべての会議を文書として記録するための手順を作成することが推奨される。例えば、中間会議は、何が検討され、議論され、決定されたかを記載した議事録によって文書化されるべきである。

これらの文書は、試験実施計画書及び／又は別の文書（統計解析計画書、DMC charter、又は adaptation committee charter 等）に含めることができる。異なる種類の情報が別の文書に含まれる場合もあるが、上記のすべての重要な情報は、審査部門が試験開始前に見解を提供する十分な時間が確保できるように、デザイン段階で FDA に提出されるべきである。

C. 完了した試験の評価と報告

アダプティブデザインの試験に基づく FDA への製造販売承認申請には、FDA が結果を徹底的にレビューできるように十分な情報と文書を含めるべきである。特に、NDA 又は BLA²⁶の通常の内容に加えて、以下に示す内容を申請に含めること。

- すべての事前に規定した計画、あらゆる関連する委員会の charter（例えば、DMC 又はアダプテーション委員会の charter）、及び上記のあらゆる補足文書（例えば、参考文献、プログラミングコード、シミュレーション報告書）
- 計画されたアダプテーションルールの遵守状況及び試験の完全性を維持するためのデータアクセス計画に概説された手順の遵守に関する情報
- アダプテーションのプロセスに関与するすべての委員会による試験実施中における議論における、審議内容及び参加者の記録（例えば、非公開及び公開の DMC 又はアダプテーション委員会の議事録、運営委員会又は執行委員会の議事録）
- 中間解析又はアダプテーションの決定に用いた解析の結果
- 添付文書案の 14 項におけるアダプティブデザイン及び試験結果の適切な報告。例えば、試験の要約には、適用したアダプティブデザインを記載すべきである。さらに、治療効果の推定値は、デザインを適切に考慮するか、あるいは未調整の標本平均値のようなナイーブな推定値を用いる場合には、バイアスの程度を評価し、その解釈

²⁶ 例えば、FDA の企業向けドラフトガイダンス「Providing Regulatory Submissions in Electronic Format – Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications (2019 年 7 月)」を参照のこと。最終的には、このガイダンスは、このトピックに対する FDA の最新の考え方を示すことになる。

に関する適切な注意点を示した上で、推定値を提示するべきである。

IND 内での継続的な議論を補強するためには、開発の過程で FDA に提供された試験の要約として、より限定的な情報（例えば、データベースのコピーがない報告書やその他の側面に関してそれほど詳細でない情報）で十分な場合がある。

IX. 参考文献

Bauer, P and K Köhne, 1994, Evaluation of Experiments with Adaptive Interim Analyses. *Biometrics*, 50(4):1029–1041.

Bauer, P and M Posch, 2004, Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections, by H. Schäfer and H.-H. Müller. *Stat Med*, 23(8):1333–1334.

Berger, JO, RL Wolpert, 1988, MJ Bayarri, MH DeGroot, BM Hill, DA Lane, and L LeCam, The likelihood principle, *Institute of Mathematical Statistics, Volume 6*:iii–v, vii–xii, and 1–199.

Berry, S, BP Carlin, JJ Lee, and P Muller, 2010, *Bayesian Adaptive Methods for Clinical Trials*, CRC Press.

Bolland, K, MR Sooriyarachchi, and J Whitehead, 1998, Sample Size Review in a Head Injury Trial with Ordered Categorical Responses, *Stat Med*, 17(24):2835–2847.

Chen, YJ, R Gesser, and A Luxembourg, 2015, A Seamless Phase IIB/III Adaptive Outcome Trial: Design Rationale and Implementation Challenges, *Clin Trials*, 12(1):84–90.

Chow, SC and M Chang, 2011, *Adaptive Design Methods in Clinical Trials*, CRC Press.

Cui, L, HM Hung, and SJ Wang, 1999, Modification of Sample Size in Group Sequential Clinical Trials, *Biometrics*, 55(3):853–857.

PREVAIL II Writing Group; Multi-National PREVAIL II Study Team, Davey Jr, RT, L Dodd, MA Proschan, J Neaton, JN Nordwall, JS Koopmeiners, J Beigel, J Tierney, HC Lane, AS Fauci, MB Massaquoi, F Sahr, and D Malvy, 2016, A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection, *N Engl J Med*, 375(15):1448–1456.

Denne, JS, 2001, Sample Size Recalculation Using Conditional Power, *Stat Med*, 20(17-18):2645–2660.

Dodd, LE, MA Proschan, J Neuhaus, JS Koopmeiners, J Neaton, JD Beigel, K Barrett, HC Lane, and RT Davey, 2016, Design of a Randomized Controlled Trial for Ebola Virus Disease Medical Countermeasures: PREVAIL II, the Ebola MCM Study, *J Infect Dis*, 213(12):1906–1913.

FDA-NIH Biomarker Working Group, 2016. BEST (Biomarkers, EndpointS, and other Tools) resource.

Fisher, LD, 1998, Self-designing Clinical Trials, *Stat Med*, 17(14):1551–1562.

Fleming, TR, K Sharples, J McCall, A Moore, A Rodgers, and R Stewart, 2008, Maintaining Confidentiality of Interim Data to Enhance Trial Integrity and Credibility, *Clin Trials*, 5(2):157–167.

Friede, T, and M Kieser, 2003, Blinded sample size reassessment in non-inferiority and equivalence trials, *Stat Med*, 22(6):995-1007.

Glimm, E, W Maurer, and F Bretz, 2010, Hierarchical Testing of Multiple Endpoints in Group-sequential Trials, *Stat Med*, 29(2):219–228.

Gould, AL and WJ Shih, 1992, Sample Size Re-estimation Without Unblinding for Normally Distributed Outcomes with Unknown Variance, *Communications in Statistics – Theory and Methods*, 21(10): 2833–2853.

Hey, SP and J Kimmelman, 2015, Are Outcome-adaptive Allocation Trials Ethical? *Clin Trials*, 12(2):102-106.

Heyse, JF, BJ Kuter, MJ Dallas, P Heaton, and REST Study Team, 2008, Evaluating the Safety of a Rotavirus Vaccine: The REST of the story, *Clin Trials*, 5(2):131–139.

Hung, HMJ, S-J Wang, and R O’Neill, 2007, Statistical Considerations for Testing Multiple Endpoints in Group Sequential or Adaptive Clinical Trials, *J Biopharm Stat*, 17(6):1201–1210.

Irle, S, and H Schäfer, 2012, Interim design modifications in time-to-event studies, *Journal of the American Statistical Association*, 107(497):341-348.

Jenkins, M, A Stone, and C Jennison, 2011, An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints, *Pharm stat*, 10(4):347-356.

Jennison, C and BW Turnbull, 1999, *Group Sequential Methods with Applications to Clinical Trials*. CRC Press.

Lan, KG and DL DeMets, 1983, Discrete Sequential Boundaries for Clinical Trials, *Biometrika*, 70(3):659–663.

Le Tourneau, C, JJ Lee, and LL Siu, 2009, Dose Escalation Methods in Phase I Cancer Clinical Trials, *Journal of the National Cancer Institute*, 101(10):708–720.

Magirr, D, T Jaki, F Koenig, and M Posch, 2016, Sample size reassessment and hypothesis testing in adaptive survival trials, *PloS One*, 11(2):e0146465.

McMurray, JJ, M Packer, AS Desai, J Gong, MP Lefkowitz, AR Rizkala, JL Rouleau, VC Shi, SD Solomon, K Swedberg, MR Zile, and PARADIGM-HF Investigators and Committees., 2014, Angiotensin–Neprilysin Inhibition Versus Enalapril in Heart Failure, 2014, *N Engl J Med*, 371(11):993–1004.

Müller, HH and H Schäfer, 2001, Adaptive Group Sequential Designs for Clinical Trials: Combining the Advantages of Adaptive and of Classical Group Sequential Approaches, *Biometrics*, 57(3):886–891.

O’Brien, PC and TR Fleming. A Multiple Testing Procedure for Clinical Trials, 1979, *Biometrics*, 35(3):549–556.

Pocock, SJ, 1977, Group Sequential Methods in the Design and Analysis of Clinical Trials, *Biometrika*, 64(2):191–199.

Pocock, SJ and R Simon, 1975, Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial, *Biometrics*, 31(1):103–115.

Proschan, MA and SA Hunsberger, 1995, Designed Extension of Studies based on Conditional Power, *Biometrics*, 51(4):1315–1324.

Sydes, MR, MK Parmar, MD Mason, NW Clarke, C Amos, J Anderson, J de Bono, DP Dearnaley, J Dwyer, C Green, Jovic, AW Ritchie, JM Russell, K Sanders, G Thalmann, ND James, 2012, Flexible Trial Design in Practice-stopping Arms for Lack-of-benefit and Adding Research Arms Mid-trial in STAMPEDE: A Multi-arm Multi-stage Randomized Controlled Trial, *Trials*, 13(1):168.

Wassmer, G and W Brannath, 2016, Group Sequential and Confirmatory Adaptive Designs in Clinical Trials, Springer series in pharmaceutical statistics, New York: Springer.

Woodcock, J and LM LaVange, 2017, Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both, N Engl J Med, Jul 6; 377(1):62–70.

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
継続タスクフォース2 (2021年3月まではタスクフォース4)

タスクフォースメンバー

飯塚 政人	田辺三菱製薬株式会社	担当：I章、II章
大澤 志乃	中外製薬株式会社	担当：III章、IV章、V.A.項
小山 暢之*	第一三共株式会社	担当：V.B.項～V.G.項
中村 将俊	大日本住友製薬株式会社	担当：VI節
西田 朋由*	ノボ ノルディスク ファーマ株式会社	担当：VII節、VIII節

* 2021年3月まで

タスクフォースリーダー兼推進委員

角野 修司	武田薬品工業株式会社
棚瀬 貴紀	大鵬薬品工業株式会社

担当副部長

菅波 秀規	興和株式会社
-------	--------