

業界、治験責任医師、 施設内治験審査委員会、 食品医薬品局スタッフ向け指針

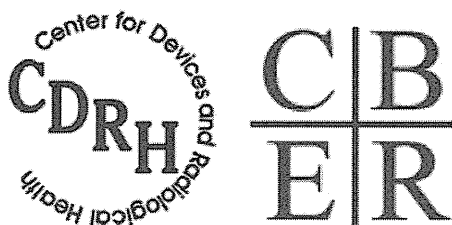
医療機器のピボタル臨床試験の デザインに関する事項

文書発行日：2013年11月7日

本文書の草案は2011年8月15日に発行された。

本書に対しCDRHの規制機器に関する質問がある場合は、Gregory Campbell, PhD（電話番号301-796-5750 又はeメール greg.campbell@fda.hhs.gov）までお問い合わせ下さい。

本書に対してCBERの規制機器に関する質問がある場合は、Stephen Ripley（電話番号301-827-6210）までお問い合わせ下さい。



U.S. Department of Health and Human Services（米国保健福祉省）
Food and Drug Administration（食品医薬品局）
Center for Devices and Radiological Health（医療機器・放射線保健センター）
Center for Biologics Evaluation and Research（生物製剤評価研究センター）

前文

パブリックコメント

書面によるコメントおよび提案は、FDAが検討するためにいつでも以下の部署に提出することができます。Division of Dockets Management, Food and Drug Administration, 5630 Fishers Lane, Room 1061 (HFA-305), Rockville, MD, 20852。

追加コピー

追加コピーはインターネットで入手できます。Eメールで dsmica@fda.hhs.gov に請求して本ガイダンスの電子コピーを受け取るか、ファックス番号301-847-8419に請求してハードコピーを受け取ることもできます。文書番号1776を用いて、請求するガイダンスを特定してください。

本指針の追加コピーは以下からも入手できます：

Center for Biologics Evaluation and Research（生物製剤評価研究センター）、

Office of Communication, Outlook and Development, HFM-40、

1401 Rockville Pike, Suite 200N, Rockville, MD 20852-1448、

さらに電話番号1-800-835-4709 又は 301-827-1800、eメール ocod@fda.hhs.gov、インターネット <http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>からも入手できます。

目次

1.	緒言.....	4
2.	適用範囲.....	5
2.1	本指針で取り上げる試験の種類.....	5
2.2	本指針で取り上げない試験の種類.....	5
3.	エビデンスの水準と試験デザインに対する規制的枠組み.....	6
3.1	PMA承認のための法的基準：安全性と有効性の合理的保証.....	7
3.2	妥当な科学的エビデンス.....	8
3.3	ベネフィット-リスクの評価.....	9
3.4	臨床試験のエビデンスの水準と規則.....	9
3.5	最小負担のコンセプトと試験デザインの原則.....	10
3.6	治験医療機器の適用免除の承認.....	11
4.	医療機器の種類.....	11
4.1	使用目的に基づく機器の種類.....	11
4.2	機器の臨床試験のための特別な事項.....	12
5.	ピボタル試験デザインにおける探索試験の重要性.....	13
6.	臨床試験デザインを選択するためのいくつかの原則.....	14
6.1	試験の種類.....	14
6.2	一般的な事項：機器の性能におけるバイアスと変動性.....	15
6.3	試験の目的.....	16
6.4	被験者の選択.....	16
6.5	被験者選択の層別化.....	17
6.6	治験実施施設の選択.....	18
6.7	比較試験デザイン.....	19
7.	臨床転帰試験.....	19
7.1	臨床試験の評価項目.....	19
7.2	臨床転帰試験のインターベンションの割付（無作為化）.....	21
7.3	盲検化（マスキング）.....	22
7.4	比較臨床転帰試験の対照.....	23
7.5	プラセボ効果とその他の現象.....	25
7.6	非比較臨床転帰試験.....	26
7.7	診断臨床転帰試験.....	28
7.8	臨床転帰試験の長所と短所.....	29
7.9	IDE申請に関する事項.....	31
8.	診断臨床性能試験.....	31
8.1	使用目的.....	32
8.2	標的疾患の臨床参照基準.....	33
8.3	診断性能の評価のための試験集団.....	33
8.4	試験の計画、被験者の選択、検体の収集.....	34
8.5	診断臨床性能比較試験.....	35
8.6	診断性能試験における盲検化（マスキング）.....	35
8.7	機器を使用する人のスキルと行動（全検査コンセプト）.....	35
8.8	診断臨床性能試験によくあるバイアスの種類.....	36
9.	臨床試験の品質の維持.....	37
9.1	臨床データの取り扱い.....	37
9.2	試験の実施.....	38
9.3	試験の解析.....	39
9.4	ピボタル試験に予測される変更.....	40
10.	治験プラン又は治験実施計画書.....	40
11.	用語集.....	41

医療機器のピボタル試験の デザインに関する事項

業界、治験責任医師および 食品医薬品局スタッフ向け指針

本指針は食品医薬品局 (FDA) のこのテーマに関する最新の考えを表すものである。これは何人に対しても如何なる権利を生じるもしくは付与するものではなく、またFDAまたは国民を拘束するために機能するものではない。適用される法的要件および規制要件を満たすものであれば、代替的アプローチを用いて構わない。代替的アプローチについて協議したい場合は、本指針の施行を担当するFDAスタッフに連絡されたい。該当するFDAスタッフが分からない場合は、本指針の表紙に記載された該当する電話番号に電話されたい。

1. 緒言

本書は、医療機器の市販前届出を裏付ける臨床試験のデザインに携わる者、その届出を審査するFDAスタッフ向けの指針である。当局は、特定の機器の裏付けとなる試験のデザインに関する方針、さらに規制要件に合わせて実証義務を調整する一般的な方針を体系化しているが、機器の市販前届出を裏付けるのに適切な、様々な臨床試験デザインの説明や、治験依頼者が特定の機器の届出を裏付けるために、どのピボタル臨床試験を使用すべきかを決定する方法を明確にしていなかった。本指針文書は、市販前臨床データの要件を満たすことができる医療機器臨床試験の作成に関する、様々な試験デザインの原則を説明する。本指針は、医療機器の治験に対し、最善の臨床業務や統計的手法を教授するためのものではない。

医療機器には、一般的に3つの臨床開発の段階がある。これらの段階は互いに極めて依存しており、1つの段階の評価を綿密に行うと、次の段階がはるかに簡単になる。医療機器に最初に行われるのは、探索的臨床評価である。この段階で、その医療機器の限界と利点が評価される。この段階にはfirst-in-human試験とフィージビリティ試験がある。次の段階であるピボタルステージでは、特定の使用目的に対する機器の安全性と有効性を評価するために必要な情報が作成される。通常、1つ又は複数のピボタル試験が行われる。最後の段階は市販後であり、追加試験や、まれな有害事象や長期有効性のように、機器の安全性をさらによく知るための試験が行われる。本指針はピボタル臨床試験のデザインに関する事項を解説し、他の段階はここでは詳細に取り扱わない。

医療機器のピボタル試験は、その使用目的に対する医療機器の安全性と有効性の評価を裏付けるためのエビデンスを集める、決定的な試験である。1つ又は複数のピボタル臨床試験のエビデンスは、通常、市販前承認申請 (PMA) を行う医療機器の安全性と有効性の合理的な保証を決定し、FDAが全体的なベネフィット-リスクを決定するための主要な根拠となる。1つのPMAに、複数の科学的問題に対応した複数のデザインの試験が含まれる場合もある。

本指針の目的は、PMAの裏付けとするために、臨床試験をどのようにデザインするかについて、治験依頼者に推奨を提供することである。しかし、市販前届出 (510(k)) と新規届出の裏付けに臨

床試験を行う治験依頼者も、本指針文書の原則をよりどころにできる。

本指針を含めFDAの指針文書は法的強制力のある責務を確定するものではない。むしろ諸指針はあるテーマに関するFDAの最新の考えを述べたものであり、特定の規制要件または法的要件が引用されていない限り、あくまで推奨事項（勧告-recommendations）とみなされるべきである。FDAの指針で使用される“*should*（すべき）”という言葉は、何かを提案（*suggest*）もしくは推奨（*recommend*）することを意味し、義務付けている（*require*）ものではない。

2. 適用範囲

本指針は、医療機器の安全性と有効性の立証の中核（*pivotal*）となる、市販前臨床試験1のデザインの原則を解説する。ピボタル臨床試験デザインの実際的な問題と陥りやすい事項、安全性と有効性に関する試験から導き出される結論に与えるそれらの影響を論じる。本指針で論じる原則は、販売申請を行う治験依頼者と、機器に関する安全性と有効性の妥当な確実性を示す適切な試験をデザインする治験責任医師を支援することを目的としている。

2.1 本指針で取り上げる試験の種類

医療機器に関する使用目的とリスクの範囲、さらに臨床試験を実施する際の制約があるため、本指針では、ピボタル臨床試験を一般的な方法で取り扱う：

本指針は、次のような2つの医療機器の広いカテゴリーに基づいてFDAの勧告をまとめている：

- 治療機器と美容機器
- 診断機器

本指針から、機器の開発者は、各カテゴリーの機器に対する重要なピボタル試験デザインについての情報を得ることができる。同時に、（例えばPre-Submissionのやりとりを介した）FDA審査スタッフとのコミュニケーションは、実際的で適切なピボタル臨床試験デザインに到達する上で、有益であることが多い。

本指針には、21 CFR Part 3 に定義される併用製品（例、機器-医薬品、機器-生物製品）の機器固有の問題に該当する原則も含まれている。ただし、併用製品にとってやはり重要である特定の薬物や生物に関する問題は、本指針では取り上げない。

本指針は他の既存の指針を補足するものであり、他の指針文書に説明されている方針に代わるものではない。質問がある場合は、しかるべきFDAの審査部門に直接問い合わせるか、機器の審査の担当に応じて、医療機器・放射線保健センター（CDRH）のDivision of Small Manufacturers, International and Consumer Assistance and Consumer Assistance、又は生物製剤評価研究センター（CBER）のOffice of Communication, Outreach and Development（OCOD）に問い合わせること。

2.2 本指針で取り上げない試験の種類

本指針では以下の試験については取り上げないが、ここで論じる原則のいくつかはこれらの多くに当てはまる：

- 非臨床試験（例えば、ベンチ、動物、測定試験、インビトロ診断機器の場合の分析バリデーション試験）；

拘束力のない推奨事項を記載

- 人道機器適用免除（HDE）の申請2を裏付けるための試験；
- 市販前のフィージビリティ臨床試験、その他のピボタル段階にない市販前臨床試験；
- 付随診断機器の臨床的妥当性を立証する試験（すなわち、それに対応する治療製品を安全かつ有効に使用するために必要な情報を提供する*in vitro*診断試験）。付随診断機器の臨床開発プログラムは、通常、それに対応する治療製品の臨床開発プログラムの一部である；
- 市販後臨床試験。市販後臨床試験の必要性は、市販前臨床試験の結果の解釈によって生じるが、市販後試験によって最初の安全性や有効性の判断がゆらぐことはなく、そのデザインについては本指針では取り上げない。しかし、本指針で論じる原則は、このような試験のデザインに有用であると思われる；
- ドナースクリーニング試験のように、新薬治験許可申請や生物製剤ライセンス申請を要するCBERの規制製品の試験は、本指針の範囲に含まれない。

本指針は主としてPMAの裏付けに使用される臨床試験に対して作成されたものであるが、本指針の推奨事項は、一部の510(k)や、該当する場合は臨床データのある新規届出の裏付けに使用される臨床試験のデザインにも使用できる。

3. エビデンスの水準と試験デザインに対する規制的枠組み

医療機器の臨床試験は、ある一定の法的要件を満たしていなければならない。本項では以下のことを説明する：

- 医療機器の市販前届出を裏付ける臨床試験のデザインに適用される規制的枠組み；
- PMA承認のための法的基準；
- PMA承認の法的基準を満たすために使用される臨床データ及びその他のデータに適用される規制要件；
- 機器のリスクとベネフィットを査定するために、FDAはどのようにデータを評価するか；
- 治験医療機器の適用免除（IDE）の申請に関する基本情報；
- 最小負担のコンセプトに示されるgood regulatory practice（良き規制慣行）についてのFDAの現在の見解

本指針は、ヒト被験者が参加する試験のデザイン、実施、記録、報告に対する当局の基準に関する考慮を反映している。これに関する国際的文書に「Declaration of Helsinki（ヘルシンキ宣言）」があり、*International Standards Organization (ISO) 14155:2011, Clinical investigation of medical devices for human subjects - Good clinical practice*（ヒト被験者に対する医療機器の臨床試験-医薬品の臨床試験の実施に関する基準）、さらにInternational Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)（医薬品許可のための技術要件の調和に関する国際会議）*E6 Good Clinical Practice: Consolidated Guidance*（医薬品の臨床試験の実施に関する基準：統合指針）で、詳しく説明されている。21 CFR Part 50、54、56及び812のFDA規則で、医療機器の臨床試験に該当する医薬品の臨床試験の実施に関する基準（GCP）の要件が組織化されている。さらに、FDAの指針文書に、GCPと臨床試験の実施に関するFDAの現在の見

拘束力のない推奨事項を記載

解が説明されている3。医療機器に該当するGCPの遵守は、ヒト被験者の権利、安全、福祉を保護し、臨床試験の適切で科学的な実施と結果の信頼性を保証し、治験依頼者と治験責任医師の責務を明確にし、治験依頼者、治験責任医師、施設内治験審査委員会（IRB）その他の倫理委員会、規制当局、その他医療機器の開発と審査に携わる団体を支援するものである。

機器の臨床試験が米国内で実施される場合、それは、21 CFR Part 56（IRB）、Part 50（インフォームド・コンセント）、Part 812（治験医療機器の適用免除：IDE）の当該規則を遵守しなければならない。機器の臨床試験が米国外で実施され、PMAの裏付けとして届出がなされ、IDEに基づいて実施される場合、その試験は21 CFR Part 812を遵守する。21 CFR 814.15(a)。ある機器の臨床試験が完全に米国外で行われ、IDEに基づいて行われない場合、そのデータが妥当であり、治験責任医師が「ヘルシンキ宣言」又は試験が行われる国の法律及び規則のいずれかヒト被験者をより保護できる方に従って試験を行った場合、FDAは21 CFR 814.15に従い、届出られた試験をPMAの裏付けとして容認する。海外の臨床データのみに基づくPMAは、その海外のデータが米国の集団と医療方法に適用でき、認められた能力を持つ治験責任医師によって試験が実施されている場合、検査が必要でFDAが現場の検査その他の適切な方法でデータの妥当性確認ができる場合、それ以外に申請が承認基準を満たしている場合は、承認される。21 CFR 814.15(d)。治験依頼者は、海外の臨床データが、当該基準に従って米国内の機器の承認を裏付けるに十分であることを証明できなければならない。連邦食品医薬品化粧品法（FD&C法又は本法）のSection 569B。海外のデータや海外と米国の併合データに基づいて承認を求めたい場合、我々は、Pre-Submission会議でFDAスタッフと話し合うことを推奨している。

3.1 PMA承認のための法的基準：安全性と有効性の合理的保証

FD&C法のSections 513(a)(1)(C)で定められているように、PMAは、機器の安全性と有効性の合理的保証を示さなければならない。FD&C法Section 513(a)(2) に次のように記載されている：

- [T] 以下に関して機器の安全性と有効性を判断すべきである---
- (A) 機器が主に使用される人、またはその使用を意図されている人
 - (B) 機器のラベルに規定、推奨、提案される使用の条件
 - (C) 機器の使用による健康への予想されるベネフィットと、その使用による傷害や疾患の予想されるリスクの比較考量

さらに、規則によって、FDAは、PMA承認の法的基準を次のように解釈している：

21 CFR 860.7(d)(1)。妥当な科学的エビデンスに基づいて判断したときに機器が安全であり、適切な指示と安全でない使用に対する警告がある場合に、その使用目的と使用条件に即して機器を使用することによる健康へのベネフィットがリスクを上回るという合理的保証がある。機器の安全性を判断するための妥当な科学的エビデンスが、その使用目的と使用条件に即して機器を使用することに関し、疾患や傷害の不合理なリスクがないことを適切に実証する。

21 CFR 860.7(e)(1)。妥当な科学的エビデンスに基づいて判断したときに機器が有効であり、標的集団の相当部分において、適切な指示と安全でない使用に対する警告がある場合に、その使用目的と使用条件に即して機器を使用することにより臨床的に有意な結果がもたらされるという合理的保証がある。

これらの法律や規則の条項は、安全性と有効性の合理的保証の発見が対象集団に関するデータによって裏付けられ、機器のラベルに照らして評価されなければならない、と定めている。さらに、PMA承認の基準が満たされているかどうかは、予想される健康へのベネフィットとリスクのバランスに基づいて判断される。

3.2 妥当な科学的エビデンス

機器の安全性と有効性は妥当な科学的エビデンスに基づいて決定される、と規則に記載されている。21 CFR 860.7(c)(1)。妥当な科学的エビデンスは、規則で次のように定義されている：

21 CFR 860.7(c)(2) 妥当な科学的エビデンスとは、適切な対照を設けた治験、部分的対照を設けた試験、マッチした対照がない試験及び客観的試験、資格を有する専門家によって行われた文書の裏付けがあるケースヒストリー、ヒトによる市販機器の重要な使用経験の報告によるエビデンスで、そのエビデンスから、その使用条件下で機器の安全性と有効性について合理的保証があると資格を有する専門家が公正かつ確実に結論できる。必要とされるエビデンスは、機器の特性、その使用条件、警告その他の制限の有無と適切性、使用経験の程度によって異なるかもしれない。個別の症例報告、無作為の使用経験、科学的評価のできる十分な内容が欠けている報告書、根拠のない意見は、安全性や有効性を証明する妥当な科学的エビデンスと見なされない。しかしこのような情報は、その安全性と有効性に疑問がある機器を特定する際に検討することができる。

どのような種類のエビデンスが安全性と有効性の合理的な保証を裏付けるのかについても、FDAの規則で論じられている。

21 CFR 860.7(d)(2) 例えば機器が安全であるという合理的保証があると判断するために必要なエビデンスの種類には、実験動物を使った試験、ヒト被験者による試験、*in vitro* 試験などの非臨床試験がある。

21 CFR 860.7(e)(2) 機器の有効性を判断するための妥当な科学的エビデンスは、原則的に、[21 CFR 860.7](f)に定義される適切な対照を設けた試験から成る。ただし、[FDA]が他の妥当な科学的エビデンスに信頼が置けると認め、適切な対照を設けた試験がなくても、それが機器の有効性を判断するに十分なエビデンスであると判断した場合を除く。[FDA]は、[21 CFR 860.7](f)の適切な対照を設けた試験の要件がその機器に合理的に当てはまらない場合に、そのような判断をすることができる。

このように、21 CFR 860.7で明白な1つの主要原則は、医療機器の有効性のエビデンスは、(21 CFR 860.7(f)の説明にあるように)一般的に適切な対照を設けた試験から得なければならないということである。しかし、規則では、医療機器の安全性を実証する妥当な科学的エビデンスと見なすことができるエビデンスの種類を決定するに当たって、FDAにある程度の柔軟性が与えられている。

ほとんどの場合、最初のPMAで審査される機器の有効性を実証するには、臨床データが必要であるとFDAは考えている。本指針の第6、7、8項に、治験依頼者が適切な試験デザインを決定する際の原則をいくつか示している。FDAが上記の規則に定める安全性と有効性の合理的保証について判断するためには、試験の結果によって十分なエビデンスが提示されなければならない。治験依頼者とのPre-Submissionの話し合いに基づき、FDAは、別の試験デザインの方が適切なデータが出て、当局が安全性と有効性を判断できると決定するかもしれない。

十分計画されたデザインであっても、その試験が、予測されたような結果や安全性と有効性を実証するために必要な結果をもたらさない場合もある。治験依頼者はその医療機器の目標を再度評価し、安全性と有効性を実証するために必要なエビデンスが得られる試験を別途実施する必要があるかもしれない。

3.3 ベネフィット-リスクの評価

FDAは、市販前審査プロセスでベネフィット-リスクの判断をする際に、いくつかの要因を検討する。FDAは、標的集団など、その機器の使用目的の範囲内でそれらの要因を検討する。

FD&C法のSection 513(a)に基づき、FDAは、重要な要因の中でも特に、PMA申請が「機器の使用による予想される健康へのベネフィットと、その使用による予想される傷害や疾患のリスクとを比較考量」することによって、「安全性と有効性の合理的保証」を示しているかどうかを判断する。21 CFR 860.7(b)(3)には、ある機器の安全性と有効性を判断する際に、FDAは、「機器の使用による予想される健康へのベネフィットと、その使用による予想される傷害や疾患のリスク」を比較考量しなければならない、と記載されている。

FDAは、その機器が安全かつ有効であるという合理的保証があるかどうかを判断するために、妥当な科学的エビデンスに頼る。21 CFR 860.7(c)(1)。妥当な科学的エビデンスに基づいて、予想されるベネフィットが予想されるリスクを上回ると判断できる時、その機器が安全であるという合理的保証がある。21 CFR 860.7(d)(1)。妥当な科学的エビデンスに基づいて、その使用目的に対する機器の使用が臨床的に重要な結果をもたらすと判断できる時、その機器が有効であるという合理的保証がある。21 CFR 860.7(e)(1)。

医療機器の予想されるベネフィットとリスクを比較考量する際には、FDAが検討するいくつかの要因がある。市販前承認申請でベネフィットとリスクの判断を行う際に、FDAが検討する要因を説明した、「Guidance for Industry and Food and Drug Administration Staff; Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approval and *De Novo* Classifications (業界及び食品医薬品局スタッフのための指針；医療機器の市販前承認と新規分類においてベネフィット-リスクを決定する際に検討する要因)」(2012)という指針がFDAによって発行されている⁵。データを十分審査した後にFDAが他の要因と共に、機器の予想されるベネフィットとリスクを評価するのは、安全性と有効性の合理的保証があることを確認するためである。予想されるベネフィットの程度を検討する際にFDAが考慮する要因には、以下のものがある：

- ベネフィットの種類
- ベネフィットの大きさ
- 患者が1つもしくは複数のベネフィットを経験する確率
- 効果の期間

予想されるリスクや危害の程度を検討する際にFDAが考慮する要因には、以下のものがある：

- その機器の使用に関連する有害な事象の重症度、種類、数、発生率
- 有害な事象の確率
- 有害な事象の期間
- 診断機器の場合は偽陽性や偽陰性の結果によるリスク

販売承認申請時にこれらのベネフィット-リスクの判断をするための適切な情報が試験で集められるように、臨床試験をデザイン、実行する際はこれらの要因を検討すべきである。製造業者は、ベネフィット-リスクの枠組みで臨床試験の計画書を作成することが望ましい。

3.4 臨床試験のエビデンスの水準と規則

21 CFR Part 812 の規則に、臨床試験を開始する前にいつIDEの承認が必要かについて説明されている。治験依頼者や治験依頼者のIRBは (FDAとの相談の有無にかかわらず)、まず、提案する試験が高リスク機器のものであるか低リスク機器の試験であるかを判断しなければならない。ただし、最終的にこの決定を裁定するのはFDAである。21 CFR 812.2(b)を参照。試験が高リスク機器に関するものである場合、治験依頼者は試験開始前にFDAにIDEを申請し、承認を得なければなら

拘束力のない推奨事項を記載

ない。同上。試験が低リスク機器に関するものである場合、治験依頼者が審査を行うIRBにその機器が高リスクでない理由を簡単に説明した後、IBRの承認を得たら、その試験はIDEの承認がなされたと見なされ、その承認と812.2(b)のIDE簡略規則の遵守を維持する。同上。この場合、812.20(a)により申請の承認が必要である旨をFDAが治験依頼者に通知している場合を除き、FDAへのIDE申請は必要ない。同上。

いかなる場合も、臨床試験に必要な科学的厳正さと収集するエビデンスの堅牢性は、試験を開始するのにIDEが必要かどうかにかかわらず左右されない。さらに、厳正さと堅牢性は、高リスク機器の試験、低リスク機器の試験、IDEの規則が免除される機器の試験といった、臨床試験のカテゴリーによって影響されるべきではない。治験依頼者は、治験実施計画書を完成させる前に、Pre-Submissionプロセスで治験実施計画書の主要要素についてFDAと協議するために、市販前申請の裏付けとなる適切な試験デザインについて、FDAにガイダンスを請求することが望ましい。IDEの承認が必要ない試験でも、FDAは、Pre-Submissionで治験実施計画書の草稿を提出し、試験開始前にFDAのフィードバックを得ることを推奨している。

3.5 最小負担のコンセプトと試験デザインの原則

適切な臨床試験のデザインを考える際に、FDAは、立証の負担がしかるべき規制要件や科学的要件と釣り合っていないからではないという原則にも従っている。この原則は、PMAと510(k)の届出に必要な臨床データと非臨床データに適用される、2つの法条項に反映されている。以下の2つの条項が「最小負担」について言及している。

本法のSection 513(a)(3)(D) の関連部分に、以下のように規定されている：

- (ii) 機器の有効性の合理的保証を実証するためにFDA長官が文書で定める、1つ又は複数の適切な対照を設けた試験などの臨床データは、そのデータが機器の有効性を実証するために必要であると長官が判断した結果として規定される。長官は、申請者と相談し、承認の妥当な可能性があると思われる、最小負担の適切な有効性の評価方法を検討する。
- (iii) 第(ii)項において、「必要」という語句は、ある申請が機器の有効性を合理的に保証しているという長官の決定を裏付ける最小限必要な情報を意味する。
- (iv) 本項の条項によって、機器の市販前承認申請の評価基準が変わることはない。

同じく、本法のSection 513(i)(1)(D) に次のように定められている：

- (i) 異なる技術的特性を持つ機器が実質的に同等であることを実証するために長官が情報を要求する場合、長官は必ず、実質的同等性の判断に必要な情報のみを要求する。このような要求をする際に、長官は実質的同等性の実証の最小負担の方法を考慮し、それに従って情報を要求する。
- (ii) 第(i)項において、「必要」という語句は、新しい機器と既承認機器の実質的同等性の決定を裏付ける最小限必要な情報を意味する。
- (iii) 本項の条項によって、実質的同等性の判断の基準が変わることはない。

FDAは、最小負担の規定をどのように適用するつもりかについて説明した指針「The Least Burdensome Provisions of the FDA Modernization Act of 1997: Concept and Principles; Final Guidance for FDA and Industry (FDA近代化法1997の最小負担の規定：コンセプトと原則；FDA及び業界のための最終指針)」(2002) (最小負担の指針)を発行している。最小負担の指針では、「最小負担」を、市販前の課題を対処するにあたって、企業とFDAの双方が時間、労力、資源を最適に投資で

きる方法、と解釈している。この指針は、2012年のFood and Drug Administration Safety and Innovation Act（食品医薬品局安全及びイノベーション法）によって最小負担の規定に追加された修正事項で支持した原則どおり、最小負担の規定によって、機器の市販前審査の法的基準が変わることはないとしている。さらに指針では、臨床試験デザインについて、別の対照に関するバイアスを最小限に抑えることができる場合、FDAと企業は無作為化臨床試験に代る方法を検討すべきであると定めている。本指針で論じる試験デザインの原則は、最小負担の指針で論じる原則と同じであり、市販前の承認や許可の基準を満たすために必要なエビデンスの水準に影響を及ぼす可能性がある事項を論じることによって、それを拡大している。

3.6 治験医療機器の適用免除の承認

本指針で論じる原則は、販売申請を行う治験依頼者と、機器に関する安全性と有効性の合理的な保証を実証するためのデータを適切に提供する試験をデザインする治験責任医師を支援することを目的としている。これらの原則に従い、ピボタル臨床試験のデザインを機器の市販前審査基準に合わせて調整すれば、FDA審査者からの追加情報、データの再解析、全く新しいピボタル試験の要求を回避することができ、承認プロセスが能率的になると思われる。しかし、FDAは、試験が販売申請を十分裏付けていないと言う理由でIDEを不承認としない。FDASIAのsection 601 で修正されたFD&C法のSection 520(g)(4)(C)に、FDAは以下の理由から[IDE]を不承認としないことが規定されている：

- (i) その臨床試験で、実質的同等性、新規分類の決定、又は機器の承認を裏付けることができない；
- (ii) その臨床試験が、機器の承認や許可に関し、データ要件などの要件を満たすことができない；
- (iii) その機器の許可や承認を裏付けるために、別の異なる臨床試験が必要である。

したがって、IDEの承認は、必ずしも、その試験が販売申請の承認を裏付けるに十分であるとFDAが考えていることを意味しない。

本指針の焦点は、IDE承認の要件ではなく、安全性と有効性の合理的な保証を示す臨床試験デザインの一連の原則である。

4. 医療機器の種類

本指針は、その使用目的に基づいた医療機器の2つの大きなカテゴリーに当てはまる：(1) 治療機器及び美容機器と (2) 診断機器である。ある機器が治療、美容、診断のどの機器として使用されるかは、機器のラベルに表示されている使用目的によって異なる。本項ではこれらの種類の機器と共に、ピボタル臨床試験をデザインする際に考慮すべき各種の機器に固有の特性を説明する。本指針はあらゆる状況におけるあらゆる機器の種類を網羅するものではない。

4.1 使用目的に基づく機器の種類

治療機器と美容機器

治療機器は一般に、特定の症状や疾患の治療に使用される。美容機器は、身体構造を物理的に修正して外観を望むように変えるために使用される。

診断機器

本指針では、診断機器を、単独でまたは他の情報と共に結果を示し、被験者の健康状態や標的疾患の評価を助ける機器として、広く説明する。標的疾患とは、被験者の過去、現在、今後の健康

状態、特定の疾患、疾患のステージ、その他の特定できる状態、または治療の開始、修正、中止のような臨床措置が必要な健康状態である。本指針において、診断機器には、人体から採取する検体の収集、準備、検査に使用される機器 [*in vitro*診断 (IVD) 機器]、診断用イメージングシステム (例えばデジタルマンモグラフィ)、*in vivo*診断機器 (イメージング以外)、解剖学的測定 (例えば骨密度、脳容量、網膜の厚さ) を行う機器、被験者の機能の測定 (例えば心臓駆出率、被験者の反応時間) を行う機器、被験者のデータを組み合わせて被験者ごとの出力を出すアルゴリズム (例えば分類、スコア、指標) を含む。診断という用語を疾患の有無を評価することと理解している者もいるが、診断検査には、例えば、妊娠の検出、特定の疾患への免疫性の評価、疾患の進行に関する情報の提供、病原体に関する情報の提供、治療薬を患者に適合させるための遺伝子型情報の提供に使用される機器、病理システムに対する染色機能のように様々な機能を自動化して検査者を助ける機器が含まれる。

複数の使用目的がある機器

多くの機器は、単純に治療用、美容用、診断用に分類できるが、症状の診断と、あると判断された症状に対する治療の両方を行う機器のように、このカテゴリーの複数の機器に該当する機器がある。このうちの1つのカテゴリーに該当するが、そのカテゴリーの中で複数の使用目的がある機器もある。例えば、2つの非常に異なる患者集団の2つの非常に異なる症状を治療する治療機器や、最初に診断を行い、その症状の進行もモニターする診断機器などである。いずれの場合も、複数の臨床試験、おそらくベンチ試験や分析試験のように複数の裏付け試験が必要になる可能性がある。

4.2 機器の臨床試験のための特別な事項

機器の臨床試験をデザインする際は、医療機器特有の事項を考慮すべきである。これらの事項は、治療、美容、診断用の全ての機器に当てはまるが、機器の種類が試験デザインの決定に影響を及ぼす。医療機器に特有の以下のような特性や機能は、FDAによる機器の評価方法に影響を及ぼす可能性があり、臨床試験デザインの中で対応しておくべきである：

- 機器がどのように機能するか：全ての機器において、機器の機能の基礎にある科学的原理、作用機序を理解することは、提案する試験デザインの性能と適切性を評価する上で重要である。このことは、特に、治験依頼者が臨床試験の作成においてFDAの助言を依頼するPre-Submissionの一環として重要である。
- ユーザーのスキル水準とトレーニング：安全かつ有効な方法で使用するために、相当なトレーニングとスキルを要する機器がある。これは明らかに埋め込み機器に当てはまる。特にその処置が複雑な場合、ユーザーは高度なトレーニングを受けた外科の専門医でなければならない。適切に機器を使用するには、複数の人 (担当者) やスキルが必要な場合もある。例えば、あるIVDの場合、あるスキルを持った人が検体を集め、別のスキルを持った人が検体を処理し、3つ目のスキルを持った別の人が検査結果を解釈する。機器の試験をデザインする際に、機器の安全で有効な使用に必要なスキルを検討すべきである。治験責任医師と担当者のスキルは、販売承認後の使用現場でその使用目的のために機器を使用する可能性がある人のスキルの範囲を反映すべきである。機器の適切な使用に関して治験責任医師と担当者に与えられるトレーニングは、機器が販売された時にユーザーに与えられるトレーニングの指針となるべきである。市販機器にトレーニングが提供されない場合、その試験が目的の使用条件を反映するようにするには、試験担当者に機器の使用を特にトレーニングすべきではない。
- 学習曲線：いくつかの機器は非常に新しいので、機器の使用に関して学習曲線がある。新しいテクノロジーの場合は、臨床試験で機器を使用するまで、ステップをマスターするのに時間を要する可能性がある。インプラントの場合は、通常これに、埋め込み処置専用の外科的手技が含まれる。一部の機器は、パイロット試験を含めた探索段階で学習

曲線を決定できる。市販前ピボタル試験で治験依頼者がオペレータに機器の実践的トレーニングを行う場合、そのトレーニングが市販後の現場でも行われることが予測される。急な学習曲線がある機器は、在宅使用のようなある種の状況には適さない可能性がある。その状況で安全かつ有効に使用できないからである。ピボタル試験で学習曲線が明白である場合、この学習曲線期間で集められた情報を（例えば試験のどの被験者が学習曲線期間に属しているかを明確に定義することによって）治験実施計画書でどのように検討するか、この結果を統計解析プラン（SAP）でどのように報告するかを考えることが重要である。学習曲線が急な場合、このことはユーザーに対するラベル表示やトレーニングの要件にも波及する可能性がある。

- ヒューマンファクターに関する事項：医療機器の開発において、ヒューマンファクターが非常に重要な役割を果たす可能性がある⁸。機器の開発過程のどの時点でも、機器の使用に関するヒューマンファクターの試験によって、被験者や医療専門家がより安全かつ有効で、簡単に使用できるように、機器のデザインや使用方法の変更が必要になる場合がある。ソフトウェアが組み込まれている機器や、ユーザーインターフェースが設けられた機器は、ユーザーエラーを最小限に抑えるようにデザインすべきである。ユーザーの側により多くのマニュアルインターベンションや客観的判断が必要とされる機器は、より多くのユーザースキルが必要とされる。ヒューマンファクター評価の一環として、現実世界における機器-ユーザーインターフェースの明確なドキュメンテーションが必要である。

5. ピボタル試験デザインにおける探索試験の重要性

医療機器は開発中にデザインがしばしば改善され、初期の研究段階から治験での使用、製品の最初の販売承認や販売許可を経て、その後承認、許可されて市販されるライフサイクルの間に、改良される。

新しい医療機器や市販機器の重要な変更の場合、臨床開発は多くの場合次の3つの段階に区分される：探索段階（first-in-human、フィージビリティ）、ピボタル段階（機器の安全性と有効性を決定する）、市販後段階（デザインの改善、機器の安全性と有効性のさらなる理解、新しい使用目的の開発）。このような段階に区分できるが、機器の開発は、継続し、繰り返されるプロセスであり、新しい情報が得られ、新しい使用目的や適応症が開発されるごとに更なる探索やピボタル試験が必要である。開発後期（例えばピボタル試験）で得られた洞察から、臨床試験や非臨床試験を含めた追加の試験が必要になる可能性がある。

この項では、（非臨床試験及び臨床試験の）探索作業からピボタル試験デザインプランの作成にいたる重要性を中心に述べる。非臨床試験（ベンチ試験、死体試験、モデリング、動物実験）は、これによって作用機序がよくわかることが多々あり、被験者にリスクをもたらすかもしれない機器に対し基本的な安全性情報を提供できる。臨床機器開発の探索段階はfirst-in-human試験とフィージビリティ試験で構成され、機器の作用機序とその安全性の理解を進め、ピボタル試験の基盤を作るために、機器のデザインを繰り返し改善できるようにすることが目的である。

探索段階で機器を徹底的に評価しておくこと、機器とその機能方法がよくわかる。これによって、機器の使用目的が治験依頼者の期待に添うかどうかを確認することができる。適切なピボタル試験デザインの選択にも有用である。堅牢な探索段階で、機器をピボタル試験と市販後に使用される形にできるだけ近づけるべきである。これにより、ピボタル試験で予測しない結果が出て、そのために変更しなければならなくなる可能性が低くなる。進行中のピボタル試験を変更するとコスト、時間、患者の資源が増えるので、これは重要なポイントである。試験が無効になったり、放棄する結果になったりするおそれもある。一般に、科学的に妥当かつ確証的な推断を下すには、

探索段階で詳細なプランニングを行わうことなく、フィージビリティ試験データとピボタル試験データを統合すべきではない。

診断機器の場合、分析の特異性、精度（反復性／再現性）、検出限界のような性能特性を立証するための機器の分析バリデーションは、探索段階の一部であることが多い。さらに、そのような機器の場合、その探索段階を使って、アルゴリズムの作成、臨床判断の閾値の決定、ピボタル臨床試験で使用される機器のバージョンの開発を行うことができる。*In vivo*と *in vitro*診断機器のいずれの場合も、早期臨床試験の結果から機器の改良が急務になり、ヒトやヒトの検体による小規模の試験を追加することが必要になるかもしれない。これらの試験を開始する前にFDAに相談すべきである。

臨床機器開発のピボタル段階が始まって、探索試験を継続することができる。例えば、FDAは、埋め込みの6ヶ月後、2年後、3年後まで埋め込み機器の動物実験を続けるよう要求する可能性がある。6ヶ月後のデータを入手した後にピボタル試験を開始できるが、追加データも収集する必要があるかもしれない。別の例として、小児への使用を目的とする場合、追加の動物試験が必要になるかもしれない。*In vitro*診断機器の場合、ピボタル試験を実施しながら（又はその後であっても）機器の（例えば有効期間の）安定性試験を継続することは珍しいことではない。

ピボタル段階は一般的に、医療機器の使用目的に対する主要な安全性と有効性の評価を裏付けるための妥当な科学的エビデンスが集められる、最終的な段階であるが、探索段階では、機器のデザインとピボタル試験のための適切な評価項目を最終決定すべきである。これは、治験機器が21 CFR 860.7(f)(2)の記載に従って標準化されるようにするためである。ここには以下のように記載されている：

治験結果の信頼性を確実にするために、適切な対照を設けた治験に、構成やデザイン及び性能を標準化した被験機器を使用する。

6. 臨床試験デザインを選択するためのいくつかの原則

FDAは販売承認申請の一部として申請された医療機器のピボタル臨床試験を審査し、それらが機器の安全性と有効性を合理的に保証しているかどうかを判断する。FDAの認識では、この期待に答えることができるいくつかの種類の実験がある。したがって、申請者がしかるべきFDAの審査部門と会い、試験を開始する前に機器の安全性と有効性を立証するための試験デザインの選択について、話し合うことを強く勧めている。

本書では、臨床試験を大きく2つの種類に分ける。臨床転帰試験と診断臨床性能試験である。以下の考察は、臨床的意義があり、統計的に適切な試験評価項目を用いた試験ごとに、回答すべき適切な質問を選んで書かれている。

この項では、医療機器のピボタル臨床試験全てに該当する問題点を取り上げる。医療機器の臨床試験をデザインする際はバイアス、変動性、妥当性のような一般的な事項、さらに試験の目的、被験者の選択、層別化、実施施設の選択、比較試験デザインなど試験ごとの事項を含め、様々な要因が重要である。これらの一つ一つを以下のように定義し、論じる。

6.1 試験の種類

臨床転帰試験

臨床転帰試験では、被験者をあるインターベンションに割り当て、臨床転帰パラメータ（又はそのバリデートされた代理パラメータ）を評価するバリデートされた評価ツールを使って予定の間

隔で被験者を検証し、そのインターベンションの安全性と有効性を判断する。この試験について第7項で詳しく説明する。機器の臨床性能（すなわちその機器が臨床現場で目的の効果を発するかどうか）も検証する場合があるが、この治験の主要ポイントは、1つ又は複数の臨床転帰である。本書において、「インターベンション」という用語は、治験機器または対照機器の使用を指す。この治験機器は治療用又は美容用である。診断機器の場合、「インターベンション」という用語は、診断機器によって出た結果に基づく被験者管理の方法に関連する。臨床転帰試験は、その目標が、機器の結果によって被験者のその後の治療経過や医療提供者による管理方法がどのように変わるかという影響を評価することである場合、診断機器の評価に使用される。

診断臨床性能試験

診断機器のほとんどの場合、ピボタル臨床評価は臨床転帰試験でなく、診断臨床性能試験である。性能試験にも臨床転帰がある場合もあるが、これらの転帰は試験の主要ポイントではない。この試験を第8項で詳しく説明する。診断臨床性能試験において、診断検査の結果は被験者から得るが、被験者の管理には使用されない。すなわち、第8項で詳しく説明するように、ある検査の診断臨床性能の特徴は、診断機器の出力が標的疾患の被験者の評価に使用する臨床参照基準とどれくらい一致するかを各被験者に対して量化する、性能測定である。

一般的な事項

疾患を検出し、治療をするというように、診断と治療の両方の機能を持った機器は、診断臨床性能試験及び／又は診断性能要素のある臨床転帰試験で評価できる。

本書では臨床転帰試験と診断臨床性能試験を別々に論じる。臨床転帰試験の詳細は第7項、診断機器の臨床性能試験の詳細は第8項を参照のこと。診断と、治療又は美容のいずれかのコンポーネントがある製品は、第7項と第8項を参照のこと。第9項は、臨床試験のエビデンスの水準を持続させるプランとテクニックについて書かれており、臨床転帰試験と診断臨床性能試験の両方に該当するものである。

6.2 一般的な事項：機器の性能におけるバイアスと変動性

単により多くのデータを集めるよりも、正しいデータを集めるための試験をデザインすることが重要である。試験デザインは、バイアスと変動性の両方を検討しなければならない。試験デザインを評価して、そのデザインがPMAの承認の裏付けになるかどうかを判断する際に重要な事項は、バイアスの統計的コンセプトである。バイアスは真実からの系統的な誤差の混入である。被験者の選択、試験デザイン、試験の実施、データ解析の手順に、バイアスが混入する可能性がある。臨床試験では、バイアスによって安全性と有効性が誤って判断される可能性がある。バイアスはいくつかの理由から臨床試験に入り込む可能性があり、バイアスをコントロールしないデザインより、ほとんどバイアスの混入しない試験デザインが望ましい。いくつかのバイアスを、デザインフェーズでバイアスをなくす又は最小限に抑える戦略と共に、以下にレビューする（第7項から9項も参照）。

バイアスは試験転帰の解釈をゆがめる可能性がある。機器の性能が優れているときは、中程度のバイアスがあっても、全体の有効性についての結論を裏付ける試験の能力は損なわれないかもしれない。性能がぎりぎりであることがわかっている（あるいはそう思われる）時は、性能がバイアスに負けてしまう試験デザインもある。特に、探索段階の試験が不十分で、機器の効果が十分理解されていない場合、機器の効果がバイアスに負けないような適切な試験デザインを選ぶことは難しいかもしれない。バイアスが最終的な試験結果を無効にするリスクを抑えるために、試験のバイアスの可能性を検討することが試験デザインにおいて重要な要因である。

試験デザインのエビデンスの水準を評価する際の2番目の一般的な事項は、サンプリングの変動性であり、これは試験の標本数によってコントロールされる。一方で、標本数が多いほどデータが

多くなり、性能の推定においてサンプリングの変動性が少なくなり、より正確となる。他方で、標本数が大きいほど、臨床的に重大でない転帰が解析され、統計的に有意であると実証される可能性もある。試験は、臨床と統計の両方の有意性が証明されるようにデザインすべきである。さらに重要な点は、標本数が大きいほど、バイアス、不適切な転帰の評価、臨床意義を証明できない転帰のわずかな改善といった問題、その他試験デザインの問題に必ずしも対応できるとは限らない、ということである。

6.3 試験の目的

試験の目的とは、その試験が何故行われているかという科学的な根拠を示すことである。目的は、必要なラベルの強調表示を含め、機器の使用目的を裏付けなければならない。

強調表示は、正規仮説検定や点推定と信頼区間で統計的に裏付けることができる。科学的仮説を検証するためにデザインされたピボタル試験は、試験の目的に、必要な強調表示に対応した帰無仮説と対立仮説の記述を入れるべきである。推定目標がある試験（例；一部の診断性能試験）の場合、仮説検定よりも、機器の性能を説明した点推定と信頼区間で強調表示を裏付けることができる。

6.4 被験者の選択

21 CFR 860.7(f)(1)(ii) では、試験プランや治験実施計画書に、以下のことを含めなければならない、と記載されている：

以下のような被験者の選択方法：

(a) 被験者が試験の目的に適していることを十分保証している、治療や診断を行う疾患の診断基準が示されている、場合によっては確認検査を行う、疾患や症状を予防する機器の場合は、感受性と予防法が必要な症状への曝露のエビデンスを示す；

臨床試験に選択される被験者は、特定の登録基準と確認検査その他の検査に基づいて、機器の標的集団（その機器が使用される集団）を適切に反映しているべきである。標的集団を代表しない被験者を登録する場合、その試験結果に被験者選択のバイアスが混入する可能性がある。

臨床試験の被験者が望ましい標的集団を反映するためには、標的集団の主要な特徴にマッチした適格基準を、治験実施計画書の中で具体的に定めるべきである。治験を実施する際、治験依頼者はこの基準を満たす者だけが含まれるようにする。これらを、試験への被験者登録の選択／除外基準という。

標的集団を検討するときに、FDAは、年齢、性別、人種及び民族に関し、当該集団の背景因子を反映した被験者を登録するよう、治験依頼者に推奨している^{10,11}。母集団のある部分からの参加が不十分であると、重要な部分集団の機器の安全性と有効性に関し、情報が不十分になるおそれがある。場合によっては、その機器が使用される疾患の罹患率、診断と治療のパターンの背景的考察を入れることを勧めている。この考察には、性別、人種別の罹患率；標的適応症に対する過去の治験に参加した女性とマイノリティの割合；場合によっては女性とマイノリティの参加不足の原因になるかもしれない要因に対するプラン、などを含めるべきである。この情報のサマリーを、治験実施計画書と治験責任医師のトレーニング資料に入れることを推奨している。試験に必要な集団の登録が容易な治験実施施設の登録を検討すべきである。患者集団 [21 CFR 812.25(c)] と海外データの使用[21 CFR 814.15(d)(1)]を説明する際に、それぞれが米国の集団と米国の医療現場にどのように該当するかを検討を試験デザインに含めるべきである。

拘束力のない推奨事項を記載

臨床試験に、小児、囚人、妊婦、身体障害者や精神障害者、金銭や教育に恵まれない人のような弱者集団が含まれる場合、治験依頼者は、21 CFR 56.111(b) と 21 CFR Part 50を遵守するために、試験に先立ってFDAと問題点を話し合う準備をしなければならない。

試験に先立って、試験の実施を改善し、成功の機会を増やすことができる情報があるかもしれない。試験を計画するときは、ユーザー／外科医のスキル、疾患の重症度、被験者の性別や年齢など、転帰に関連する可能性のある要因を検討することが重要である。性別、年齢、民族のような重要なサブグループと、現行の試験に特に重要なグループ全てを適切に示すことに関し、注意が必要である。対象の疾患がまれな場合、診断試験の被験者選択が難しい場合は、別のアプローチを検討してもよい。

治験実施計画書に、無作為選択、連続選択、系統的選択、及び／又は便宜的選択など、いくつか考えられる被験者の選択方法の1つを入れる。

- **無作為選択**は、被験者全員に、臨床試験に選択されるという既知の（通常は同じ）機会がある。実行可能な場合、臨床試験の被験者に無作為選択を使用すると、選択された集団に対してバイアスのない推定ができる。無作為選択は記号論理的に難しいため、実現性のないことが多々ある。
- 臨床試験は一般的に、選択／除外基準を満たした者の中から、**連続選択**（すなわち、治験実施施設に来た順番に被験者を選択する）又は**系統的選択**（例えば10番目の被験者ごとに選択する）を使用する。これらの選択方法は、転帰に影響を及ぼす被験者に関する他の変数によってその試験期間に交絡が生じない限り、バイアスのない推定ができる可能性が高い。例えば、試験が朝まで続く場合、早朝に来院する被験者はクリニックに来院する全被験者の一般的な例ではないので、試験被験者は標的集団の代表者になれない。
- **便宜選択**は、研究者の所に来やすいという理由から被験者を選択する方法である。この方法は、標的集団に一般化できない結果が出る可能性がある。

6.5 被験者選択の層別化

複数の実施施設で被験者を登録する場合、標的集団を適切に代表する施設から被験者を選択することが必要である。代表的な実施施設を選択するだけではこれができない場合がある。機器の性能や使用方法の違いが予測される重要なサブグループにおいて、機器の性能を適切に特徴化する必要がある。例えば、男性と女性の両方が使用する機器の試験に男性を多く登録すべきでない。全成人に使用される機器の試験は、全成人の年齢群を適切に網羅すべきである。

被験者の選択には大きく分けて2種類の方法がある：層別化選択と、選択／除外基準に基づく選択である。

層別化は、標的集団を、既定の重複しない被験者のサブグループ、すなわち層に分けることである。被験者の層別化選択とは、被験者を各サブグループ（層）から個別に選択することである。例えば、ある被験者集団を性別（男性、女性）と年齢（ある年齢より下か上か）のグループで層別化すると、4つの層ができ、それぞれ別の性別と年齢の組み合わせになる。これらの特徴は被験者が試験に登録されるときに記録されるものであり、インターベンションの結果ではない。被験者の層別化選択は、重要なサブグループを適切に表すだけでなく、統計的により正確な機器の性能を推定できる。サブグループごとに機器の性能が異なると考える理由がある場合は、FDAと相談して容認できるデザインを決定するとよい。

特定のベースラインの特徴や層に関係なく、単に規定の選択／除外基準に従って被験者が募集さ

拘束力のない推奨事項を記載

れることがよくある。この種の選択は、機器が被験者のどのサブグループにも同じように機能することが予測される場合に適しているかもしれない。

このような臨床転帰試験の場合、試験が行われる複数の実施施設のような重要なサブグループや層、共存症（例えば糖尿病）の有無のような被験者の転帰に対して高い予測性があると思われる共変量を試験することが決まったら、既定の層の中で別々に無作為化する、層別無作為化を検討することも賢明かもしれない。

6.6 治験実施施設の選択

治験依頼者は、機器の使用目的に適した被験者登録施設（実施施設）を選択する。診断機器の場合、検査施設は通常被験者登録施設とは異なる。

新しい機器の最初のフェージビリティを評価するには、単一施設の治験責任医師による試験が有用なスタートポイントかもしれない。ロジスティック面で連携しやすく、少ない資源を集中でき、一般的に交絡変数の少ない均一な被験者集団に集中して試験が行えるからである。より大規模な多施設共同試験のプランニングにも役立つ。単一施設の試験は、ピボタル臨床試験にはほとんど適さない。

むしろ、治験機器の安全性と有効性の評価は通常、大規模な多施設共同試験の施設全体で概ね一致した結果を実証できるかどうかにかかっている。多施設共同試験のメリットは、被験者を募集しやすく、必要な標本数に早く達するという点である。

多施設共同試験は、標的集団のより代表的な標本が得られ、そのため試験の所見を一般化しやすい。実施施設間の転帰の差は、被験者の選択、手術手技、臨床医のスキル、学習曲線の違いを反映している可能性があり、これら全てが試験結果の解釈を偏らせるおそれがあるので、医療機器の試験転帰を評価する上で非常に重要である。同様に、診断臨床性能試験では、被験者や検体が他の施設から紹介されてくることがあり、検査を行う人や結果を解釈する人のスキルが異なる可能性がある。臨床転帰試験と診断臨床性能試験はどちらも、実施施設によって試験結果が大きく異なる可能性があるため、複数の施設の試験結果を統合するための特別な統計手法が必要かもしれない（poolingを裏付けるための解析など）。

場合によっては、性別、年齢、人種、民族、社会経済学的状況、共存症のような変数に関して、治療する疾患の疫学的分布を反映している被験者を治験実施施設に組み入れるように、特別な注意が必要である。さらに、使用対象の集団と異なる範囲の疾患を持つ被験者を組み入れると（例えば被験者のスクリーニングに使用する機器に別のまれな疾患の被験者を組み入れる）、機器の性能の推定にバイアスが生じる可能性がある（紹介バイアス）。診断試験において、実施施設ごとに高リスクの疾患や平均リスクの疾患を持った被験者がいる場合があり、これらの結果を個別に検討する必要がある。（第6.5項を参照）

同様に、機器に応じて、実施施設ごとの治験責任医師やオペレータの経験の違いを考慮することが重要である。例えば、臨床転帰試験の場合、三次医療施設の外科医は地域病院の外科医よりも専門的な経験が多いかもしれない。したがって、臨床試験に三次医療施設だけを選択すると、機器の性能の評価にバイアスが生じるおそれがある。

米国の市販前届出の裏付けとする試験は、被験者の背景因子、標準治療、医学的方法に関し、米国の患者に使用したときの機器の安全性と有効性を知る上で意義がなければならない。これは、米国内と国外の両方で行われる試験にとって重要である。これらの基準を満たしていない試験は、機器の承認の裏付けには不適切である。

治験依頼者は、その機器の試験に必要なトレーニングと経験を有する治験責任医師を選択する責任がある。21 CFR 812.43(a)を参照。ある種の新しい機器は、機器を安全に使用できるように、臨床試験計画の一環として追加トレーニングが必要かもしれない。治験依頼者は、治験機器のユーザーが必要なトレーニングを受け、経験できるように徹底すべきである。治験責任医師は、臨床試験の実施に関する当該規則と指針を知っておくことも望ましい。

6.7 比較試験デザイン

複数の治療法や複数の診断検査の性能を比較する試験は、比較試験デザインと呼ばれる。比較デザインにはいくつかの種類がある。

- **並行群間比較デザイン**: 各被験者や標本を、比較する治療法や検査法の1つだけに割り付ける。異なる被験者群（又は標本群）に各治療法（又は各診断検査）が割り付けられるため、比較は被験者群間で行われる。この種の試験では、治療や検査の前のベースライン時に両群が同等であり、公正な比較ができるように、被験者をインターベンション群に無作為化することが望ましい。
- **一対比較デザイン**: 各被験者や標本が同時に全ての治療法又は検査法を受ける。したがって、同じ被験者で治療法や検査法を比較することができる。一例として、顔の左右を異なる器具で治療する、**split-face**デザインがある。一般的に、一対比較デザインの方が並行群間デザインよりも比較が正確であることが多い。並行群間デザインは、被験者群全体比較に被験者間の変動性が混入するからである。一対比較デザインは一部の治療や美容治療にはあまり一般的ではないが、同じ被験者や標本に色々な診断機器を使用する診断試験には非常に多い。このデザインの短所は、局所以外の有害事象が2つのインターベンションの1つに関連するのか両方に関連するのかの見極めが難しいことである。
- **クロスオーバーデザイン**: 各被験者や標本が、異なる時期に、ただし既定の順番で複数の治療（又は診断検査）を受ける。各被験者は既定の順番で治療（又は診断検査）を受け、インターベンションが複数の順番で試験されることが多い。クロスオーバーデザインは、全ての治療法（又は診断検査）を同時に1人の被験者に割り当てることができない場合、さらに1つの治療法（又は診断検査）の効果が次の段階まで持ち越されない場合に、適切かもしれない。この種のデザインは、特に理由がない限り、順番を無作為化すべきである。クロスオーバーデザインは治療法、美容法、診断法の試験に使用できる。

7. 臨床転帰試験

臨床転帰試験をデザインする際は、様々な重要要因を検討する必要がある。本項ではこれらの要因を考察する、例えば：

- 被験者の評価項目、インターベンションの割付（無作為化）、盲検化、プラセボ効果、対照、非比較試験の特別な事項
- バイアスの原因とバイアスの最小化のための一般的な事項

7.1 臨床試験の評価項目

どの臨床試験でも、機器の性能と臨床転帰を実証する主要な変数が選択される。臨床転帰試験の

拘束力のない推奨事項を記載

場合、これらの変数は主要及び副次臨床評価項目である。全ての主要評価項目と副次評価項目が、ピボタル臨床試験のデザイン段階で既に決められていることが重要である。

機器の性能と臨床転帰を最小のバイアスで客観的に測定する必要がある。以下のようないくつかの考察点がある：

- 評価項目、転帰、測定値が、機器の使用目的に対する（安全性と有効性の両方の）臨床効果を分析する上で、十分なエビデンスを提示するべきである。
- どのような評価項目や転帰を測定するか、それらをどのようにして、いつ測定するか、それらを統計的にどのように解析するかを、治験実施計画書に規定する。統計解析の詳細を別の統計解析プランに盛り込んでもよい。
- 評価項目、転帰、測定値に臨床的意義があり、記載の試験の目的や使用目的と関連性がなければならない。ピボタル試験は、単に機器がどのように機能するかを実証するのではなく、既定の被験者集団に対する臨床ベネフィットを実証するようにデザインすべきである。
- できる限り、評価項目は客観的で、国内外的に妥当であり、最小のバイアスで判断されるべきである。例えば、治療評価項目に脳卒中や心筋梗塞の有無のような診断評価が入っている場合、主観的な評価より診断の臨床参照基準が望ましい。より客観的な評価方法があるときは、臨床転帰試験の評価項目の決定に主観的臨床評価に頼ることは通常不適切である。このような場合、いくつかの懸念への対応策として、評価項目の裁定に臨床専門家委員会を使用してもよい。
- 評価項目、転帰、測定値を誰が評価するか、被験者及び／又は治験責任医か、例えばインターベンションの割付を盲検化された評価者か治験責任医師かなどを、治験実施計画書に明記する。
- 例えば、レントゲン写真の読解のように、客観的な評価がなく、主観的な評価が使用される場合に、独立した裁定委員会が評価項目を裁定することが勧められる試験もある。評価項目、転帰、測定値の裁定のルールを、ピボタル試験の治験実施計画書にあらかじめ決めておく。
- 対象の転帰や必要な使用目的を被験者の観点から測定するのが最善である場合（例えば疼痛の軽減）は、被験者の転帰報告のための道具を使用できる。このような場合、当該被験者集団と治療する疾患に対してバリデートされている、目標の使用目的と一致した採点評価方法を選択することが重要である。このため、試験デザインの段階で早期にFDAと話し合うことが重要である。複合評価項目の一部として、これらの主観的な測定項目が客観的な評価方法と一緒に使用されることが多々ある。多国籍の治験で主観的な評価尺度を使用する場合、治験依頼者は、その尺度が文化や言語を超えて解釈可能であり、有効であることを確認すべきである。被験者報告転帰の使用とそのバリデーションについての詳細は、FDAの指針を参照のこと12。
- 複合評価項目とは、複数の評価項目をあらかじめ規定して組み合わせた評価項目である。複合評価項目の使用は難しく、FDAと慎重に早期の話し合いをし、適切な評価項目と解析プランを作成する必要がある。複合評価項目を使用する場合、複合評価項目全体に対する機器の影響の解析に加えて、FDAは、その複合評価項目がいずれかのコンポーネントに支配されていないか、個々のコンポーネントの結果に一貫性が欠如していないかを評価できるように、複合評価項目のそれぞれに対する機器の影響も評価する。