

日本核医学会

分子イメージング臨床研究に用いるPET 薬剤の基準

III. 臨床評価基準

2011 年 10 月理事会承認

2011 年 11 月一部修正

Ⅲ. 臨床評価基準

1. 基本的考え方

本臨床評価基準では、PET薬剤について、その臨床適応を疾患診断及び疾患名を特定しない機能評価による診断の両側面からとらえて臨床適応を明確にするとともに、その臨床評価の各段階について、それぞれの段階における目的、試験デザイン、実施項目等について概説する。

本基準は、日本核医学会として、PET薬剤の臨床開発の段階を明らかにし、臨床試験デザインを検討する際の参考となる情報を提供することを目指して作成したものであって、ここに記載された基準の遵守を要求するものではない。主として「臨床研究に関する倫理指針」を遵守して行われる臨床研究を想定しているが、科学的なエビデンスを生成する手順と段階を示すものであるため、多くの部分は薬事法上の治験においても該当する。

本基準では、臨床開発の段階を、探索段階と検証段階とに分類し、探索段階をPhase I、IIと区分、検証段階をPhase IIIとする。1つのPhaseにおけるプロトコル数は1つのことも複数のこともあり、逆に複数のPhaseを組み合わせた1つのプロトコルをデザインすることもできる。他の施設や国外で得られたデータに基づいて、あるPhaseやプロトコルにおける特定の項目を除外することも可能である。

いずれの場合にも、当該プロトコルの実施にあたっては、「臨床研究に関する倫理指針」その他、本基準以外の適用される規則等に従って実施されなければならない。また、日本核医学会をはじめ諸学会が定めている他の関係する基準やガイドラインも十分尊重されるべきである。

2. PET 薬剤の臨床適応

2. 1. 臨床適応の考え方

PET 薬剤の臨床適応は、以下のような類型に分類できる。これらは順位や順序を示すものではなく、また同一の薬剤に複数の適応が与えられる場合もある。

2. 1. 1. 疾患の検出または評価（疾患診断用薬）

特定の疾患や病態の検出または評価、および局在診断を目的とする場合。

2. 1. 2. 生理学的、生化学的評価（機能診断用薬）

組織、器官、身体部位の正常な生化学的、生理学的、分子生物学的機能の低下または亢進の検出を目的とする場合。これらの機能の障害が複数の疾患や病態に共通して見られるような場合には、その機能を評価することを目的とし、何らかの特定の疾患や病態の診断を意味するものではない。すなわちバイオマーカーを検出又は評価する考え方である。

2. 1. 3. 患者の診療方針決定（患者選択用薬）

患者の診断または治療方針の選択という適応を目的とする場合。一定の臨床状況において使用された場合に、症状改善や転帰の予測、ある特定の治療薬への反応性に関する情報（ある特定の受容体の存在等）の付与が期待される。より侵襲的な診断の実施の適否の意思決定のための情報が得られることもある。疾患診断、機能診断のいずれも、この適応に応用できる可能性がある。

2. 1. 4. その他の目的（複合的イメージング薬）

上記分類に当てはまらない、あるいは複数の適応を目的とする場合。

2. 2. 試験デザインの特徴

上述の A～D の各種臨床適応についての臨床試験のデザインには、以下のような特徴が考えられる。

2. 2. 1. 疾患診断用薬の試験デザイン

特定の疾患や病態の検出または評価、および局在診断を目的とする場合の臨床試験計画においては、対象者の適格基準を明確に定義し、撮像の条件を統一し、画像の評価・解析の手法を定義しておく。

対象者中の当該疾患患者の割合（検査前確率）に影響する因子としては、①重症度、

②疾患が疑われる患者、既に臨床的に診断がついている患者の割合、③利用可能な確定診断方法、が挙げられる。

診断性能の評価は、感度・特異度・陽性的中率・陰性的中率・正診率などの統計量、ROC 曲線などを用い、さらに既存の標準的な診断方法との比較により評価することが望ましい。が、分析感度 (assay sensitivity=既存の方法と新しい方法との有意差を検出する力)によって、得られる診断性能の確からしさが異なることも考慮すべきである。適切に管理された症例集積研究、症例報告によって診断学的な意義を示すことも有用である。特に、標準的な方法が明確でない、または存在しない場合には、これらの情報がより重要となる。

2. 2. 2. 機能診断用薬の試験デザイン

機能診断用薬の探索的臨床試験では、標的とする機能に関わる生化学的・生理学的・分子生物学的な評価項目について検討を行う。

診断性能の評価は、疾患診断用薬の場合と同様、既存の標準的な診断方法との比較において、感度・特異度・陽性的中率・陰性的中率・正診率などの統計量、ROC 曲線などによって評価する。

検証的臨床試験では、疾患診断を適応とする場合と比べて小規模で、且つ、少数の症例で機能診断の検証が達成できることがある。

得られる結果は対象者のプロフィールによって影響を受けるため、検査前確率を評価しておく。標的とする生化学的・生理学的・分子生物学的機能の描出の証明を得ることが必要である。

2. 2. 3. 患者選択用薬の試験デザイン

探索的臨床試験においては、疾患診断または機能診断によって得られる所見と患者の診断または治療方針の選択との関連性を十分に検討しておくべきである。

検証的臨床試験においては、PETによる機能検査の結果に基いて診療方針を選ぶという事前に決定したアルゴリズムに従って行う治療群と、検査を行わず標準とされる現行の治療を行う治療群とを比較するランダム化試験を行ってもよい。また標準的治療に基づいた回復率、生存率、反応率などの患者転帰に関する既存のデータを収集し比較してもよい。新しい患者選択用薬の使用によって、現行のアルゴリズムに従う治療法に比べて、より良い診断または治療効果へつながるといふ仮説を検証できることが望ましい。

2. 2. 4. 複合的イメージング薬の場合

目的に応じて上に述べた点を考慮してデザインする。

たとえば疾患や病態の形態と機能の両面が、イメージングで同時に評価されることがありうる。このような場合には、臨床試験において、形態と機能の両面を評価する複合

的イメージング薬としての有効性を評価することが重要である。

機能診断用薬あるいは患者選択用薬として開発をスタートし、後に疾患診断用薬としての適応をも目的とする段階的開発方法もありうる。機能診断あるいは患者選択を適応とする有効性が実証された後に、疾患診断を適応として開発する場合には、治療薬の臨床試験に組み込んで、多数の被験者をもとに拡大臨床試験を実施し、疾患を診断する性能とその臨床的意義を証明する方法もありうる。

3. PET 薬剤の臨床試験デザイン

3. 1. 探索的臨床試験：Phase I

3. 1. 1. 目的

Phase I では、薬剤の安全性、薬物動態及び吸収線量の評価を主目的とする。加えて、有効性の予備的なエビデンスとして、分子イメージングの最適撮像条件の探索などを副次的目的に加えることもできる。

3. 1. 2. 対象者

原則として、健康な成人男性を対象とする。Phase I で未成年者、女性、患者、その他の特殊な集団を対象とする場合は、その必要性、リスクとベネフィットの比較考量に基づく妥当性を論証できなければならない。たとえば、評価する薬剤の使用ががん患者に特定される場合など、ある特定の集団に対してのみ投与が想定される場合は、その特定集団を対象とすることができる場合がある。

また、健常者を正常コントロールとし、患者その他の特定の集団における組織・臓器集積性、用量、放射能量、または投与方法が画像収集、安全性、有効性に与える影響について探索する試験も、正当な科学的論拠があれば Phase I における副次的目的として評価することも可能である。

A. 対象者登録前の確認事項

- ・問診等・理学的検査：年齢、性別、身長、体重、脈拍、血圧、体温、既往歴・病歴・アレルギー等、自覚症状、PS (ECOG)
- ・血液一般、血液生化学、凝固・線溶系、尿検査（定性）、生理機能検査

B 対象者選択基準（例）

- ① PS (ECOG Performance status) が 0 または 1
- ② 年齢、性別、既往歴、家族歴等の基準。患者の場合は、さらに臨床診断、類型、ステージ、重症度、治療歴、等の規定

③ PET 薬剤や目的に応じて主要臓器機能が十分保持されている被験者を対象とする

[例]

- ・好中球数 1,500 /mm³ 以上
- ・血小板数 75,000 /mm³ 以上
- ・ヘモグロビン 9.0 g/dL 以上
- ・AST および ALT 100 IU/L 以下
- ・総ビリルビン 1.5 mg/dL 以下
- ・結成クレアチニン 1.5mg/dL 以下
- ・心電図 正常もしくは治療を要しない程度の変化
- ・心エコー 左心%駆出率 60%以上

C. 除外基準 (例)

- ①重篤な薬剤過敏症の既往歴がある
- ②虚血性心疾患（狭心症、あるいは心筋梗塞）、心筋症、うっ血性心不全の合併、あるいは既往歴がある。薬物でコントロール不良な不整脈がある。
- ③活動性の感染症を合併している。間質性肺炎又は肺線維症である。
- ④コントロール不良の糖尿病、またはインスリン治療中である。
- ⑤同時性重複がんを有する。
- ⑥妊娠、又は授乳中、妊娠している可能性またはその意思がある。
- ⑦精神病、又は精神症状により治療が困難である。
- ⑧主治医や研究実施責任者により本臨床試験の対象として不適格と判断される。

3. 1. 3. 試験デザイン

3. 1. 3. 1. 用法・用量

被験者の不要な被ばくを避けるために、単一用量（放射エネルギー）、単回投与で実施することを基本とする。非臨床試験の毒性試験成績及び非臨床試験より推定されるヒトでの吸収線量に基づき、また、同一核種を用いた既承認放射性医薬品の投与量、吸収線量を参考として、安全で有効と想定される適切な第一義的適切用量を選択する。

Phase I の結果から得られた投与量（放射エネルギー）を[第一義的適切用量]と定める。

3. 1. 3. 2. 安全性と薬物動態および被ばく線量の評価

以下の項目は基本的な安全性評価項目としてプロトコルに含むことを推奨する。

- ・血液一般：(例) 白血球数、好中球数、白血球分画、赤血球数、ヘモグロビン、血小板数
- ・血液生化学血液生化学：(例) 総蛋白、アルブミン、総ビリルビン、AST、ALT、

BUN、Cr、ALP、LDH、Na、K、Cl

- ・凝固・線溶系：(例) フィブリノーゲン、FDP、PT、APTT
- ・尿検査：(例) 定性 蛋白、糖、潜血反応
- ・生理機能検査：(例) 安静時 1 2 誘導心電図、胸部単純 X 線
- ・有害事象・副作用の評価
- ・その他の重大な検討事項

また、以下の項目は、薬物動態と被ばく線量の評価のためにプロトコルに含むことを推奨する。

- ・放射能の血中濃度と尿中排泄量、それらの時間経過
- ・有効成分（血中未変化体）の血中濃度と時間経過
- ・全身臓器への放射能集積量と時間経過

安全性の評価は投与前後に問診、診察、およびこれらの検査を行い、報告された自覚症状、他覚的所見及び検査所見より有害事象を収集し（CTCAE: Common Terminology Criteria for Adverse Events v.3.0）、被検薬剤の安全性を評価する。

非臨床検査成績や先行する臨床試験成績等を参考にして、必要な場合は、検査項目を追加あるいは削減して差し支えない。

問診等の実施時期、回数、観察期間は、予想される薬剤の作用機序や薬剤特性を考慮して適宜設定する。

一般的な検査・観察項目は適宜実施する。

例) 対象者登録前 28 日以内：問診等・理学的検査

対象者登録前 14 日以内：血液一般、血液生化学、凝固・線溶系、尿検査（定性）、生理機能検査

薬剤投与当日、投与前と後：血液一般、血液生化学、問診等・理学的検査

薬剤投与 2 日後：血液一般、血液生化学、尿検査（定性）、自他覚症状（CTCAE）

薬物動態の評価のためには、放射能の血中濃度を経時的に測定し、さらに血中の有効成分（未変化体）や標識代謝物の割合を評価して、有効成分の血中濃度－時間曲線下面積（AUC）、クリアランス、分布容積、半減期等の薬物動態パラメータを求め、被検薬剤（有効成分）の薬物動態学的プロファイルを明らかにする。また、必要に応じて尿中排泄、糞便中排泄を測定する。

被ばく線量の評価のためには、PET カメラ等を用いて全身臓器への放射能集積を経時的に測定し、MIRD 法を用いて各臓器の吸収線量と（全身の）実効線量を計算する。

これらについても、先行する臨床試験の成績等を参考にして、適宜追加あるいは削減

させて差し支えない。

3. 1. 3. 3. 有効性の探索的評価

以下の項目は最適撮像条件の探索において重要なので、そのために必要なデータ収集を評価項目としてプロトコルに含むことができる

- ・ 標的とする組織・臓器への集積性
- ・ 定量法の確立
- ・ 用法用量の決定（投与方法（ボラス注射か持続投与かなど）、マスドーズを含む）
- ・ 画像収集の方法とタイミングの最適化
- ・ 画像の評価・解析方法と評価基準の構築

3. 2. 探索的臨床試験：Phase II

3. 2. 1. 目的

探索的臨床試験Phase IIの目的としては、安全性データの蓄積と開発コンセプト（有効性）のヒトでの探索、検証的臨床試験に向けた用法・用量の改善、撮像法の最適化、画像評価の方法と診断基準の設定、その他被検薬剤の問題点抽出、があげられる。試験結果の探索的な解析によって、検証的臨床試験Phase IIIの評価項目、対象患者群、対象患者数、診断基準などを設定することが可能である。探索的臨床試験Phase IIは、次の検証的臨床試験Phase IIIに進むべきかどうかを決定する段階であり、いかに効率よく手がかりを掴むかが重要である。

3. 2. 2. 対象者

通常は、Phase IIでは、予定される対象疾患あるいは病態の患者集団を対象とする。可能である場合は、有効性の予備的エビデンスとするために、対象疾患や病態を持たない集団を対照群として対象に含めることができる。

検証的臨床試験 Phase IIIの試験デザインを設定するため、Phase IIIで予定される集団を対象被験者として、診断性能を推定する。しかし、対象被験者に診断が未確定である症例（例えば、臨床症状等から予定される対象疾患を疑う段階）を予定している場合は、Phase IIで同様の被験者集団を対象とすることは困難な場合がある。この場合、予定される対象疾患において、診断すべき部位や機能の異常があること、又は、異常がないことが他の診断方法（組織診断、画像診断、追跡調査）などで確認できる集団を対象として診断能を推定し、Phase IIIにおける試験デザインの根拠とすることが可能である。

3. 2. 3. 試験デザイン

3. 2. 3. 1. 用法・用量

PET の画質は、標的や臓器に集積した放射性同位元素から放出される光子又は陽電子数に依存するため、収集データ量（収集カウント）を変えて作成した画像を評価することで、実際に複数の投与放射エネルギーを用いることなく投与放射エネルギーと画質や診断性能の関連性評価が可能である。

第 I 相試験の結果得られた投与量(放射エネルギー)を「第一義的適切用量」と定める。まず、第一義的適切用量を投与して、既存の核医学診療で用いられている最長時間で、画像データを収集する（収集カウント_{MAX}）。この収集カウント_{MAX}について、任意の時間で収集データを複数抽出し、画質が損なわれない適切な収集カウントと収集カウント_{MAX}との比と第一義的適切用量とから、最も適切な投与量を算出する。

探索的臨床試験 Phase II における用法・用量の評価では、被検薬剤について、含まれる成分の量や性状、放射エネルギー、放射能のないリガンドや担体の量、specific activity、含まれる放射性同位元素等について、必要に応じて PET 薬剤の基準を調整することを推奨する。

3. 2. 3. 2. 安全性

安全性評価は、一般的な評価項目である自覚症状、他覚所見、バイタル検査（血圧、脈はく、体温）、臨床検査（血液学的検査、血液生化学的検査、尿検査）を設定し、有害事象を調査する。これまでに被検薬剤に特有と思われる有害事象が発現したことが判明している場合には、当該事象を確認できる検査・観察項目を追加することを検討する。

3. 2. 3. 3. 有効性

検証的臨床試験 Phase III のサンプルサイズが設定できるよう、比較対照となる診断技術がある場合はその技術も含めて、被検薬剤の有効性について探索的な評価を行う。有効性評価の指標は、通常、病理所見や臨床転帰又は経過観察所見などの真のスタンダードを基にした診断能（感度、特異度）等を用いることが望ましい。

3. 3. 検証的臨床試験：Phase III

3. 3. 1. 目的

検証的臨床試験 Phase III の主要な目的は、先行する試験によって設定・開発された投与方法、撮像法、画像評価法あるいは診断基準を確固とすること、被検薬剤の使用が想定される患者集団において有効性を検証すること、及び安全性データを継続して蓄積することであり、それらを通じて、被検薬剤の使用方法を適正化することである。

3. 3. 2. 対象者

対象者は、原則として当該薬の使用が想定される患者集団とする。しかし、将来実際に使用が想定される対象が未確定である場合は、科学的妥当性と実施可能性を十分考慮して、診断すべき部位・機能の異常があること、又は、異常がないことが、他の診断方法で確認されている集団を対象として診断能を検証することが可能である。

3. 3. 3. 試験デザイン

検証的臨床試験 Phase III では、探索的臨床試験 Phase II の結果に基づき、そこで設定された主な仮説の確認、その有効性と安全性の実証、患者集団における使用方法および撮像方法の検証が含まれることが望ましい。試験のデザインには、例えば、用量、撮像方法および撮像時期、患者群、エンドポイントが含まれることが望ましい。臨床試験の主要エンドポイントに臨床的に有用な画像の解釈を組み込むことが望ましい。画像所見の臨床的意義が明確である場合には、客観的な画像所見を主要エンドポイントに組み込むことが可能である。しかし、画像所見の臨床的意義が明瞭にならない場合、画像所見はイメージングの副次的エンドポイントとすることが望ましい。

これら試験デザインは、薬剤の性質とそれぞれのデザインの利点と限界を考慮して決定する。

3. 3. 3. 1. 用法・用量

探索的臨床試験 Phase II で有効性と安全性が確認された用法・用量を用いる。

3. 3. 3. 2. 安全性

探索的臨床試験 Phase II に準じる。一般的な評価項目やこれまでに判明した被検薬剤特有と思われる有害事象がある場合には、それを確認できる検査・観察項目についてのデータを蓄積する。

3. 3. 3. 3. 有効性

探索的臨床試験 Phase II で得られた知見に基づく有効性の仮説を検証する。可能な限り第三者による画像評価（盲検化）を実施し、その結果に基づく指標を有効性の主要な評価項目とする。

3. 3. 3. 4. 比較対照

被検薬剤と同種の情報を与える既存の診断技術が存在し、当該技術の改良又は当該技術との置き換えを目指す場合には、当該技術との比較試験を行う必要がある。比較試験のデザインとしては、個体内比較試験及び並行群間比較試験がある。

4. 臨床的評価におけるその他の考察

以下の節では、PET 薬剤の臨床試験における有効性の評価において考慮すべき事項について述べる。以下に記載される事項は推奨する考え方を示すものである。

4. 1. 被験者の選択と症例数

有効性試験の対象となる被験者は、その画像診断薬の使用が意図されている集団を代表するものであるようにすることが望ましい。さらに、プロトコルと試験報告書には、選択バイアス（典型例ばかりが対象となるなど）の可能性を検討し易くするために、試験の参加者として患者を選択した方法（例えば、連続的被験者登録、無作為抽出など）や重症度・類型などのプロフィールについて記述することを推奨する。これは、被検薬と同様の診断目的をもつ他の検査結果に基づいて被験者を登録する場合には特に必要である。

有効性試験における症例数は、帰無仮説と対立仮説を設定し、本来無効である薬剤を誤って有効であると判断するエラー（ α エラー）を0.05-0.1、本来有効である薬剤を誤って無効であると判断するエラー（ β エラー）を0.1-0.2というように仮定して、必要症例数を設定する。

一般に、抗がん剤など治療薬では必要症例数はPhase I では15-30人、Phase II では100人未満、Phase IIIでは100人から数千人と設定されることが多い。しかしながら、PET薬剤は通常、検査の時に単回で微量（マイクロドーズ用量）が投与されるのみであるという特徴を考慮して、PET薬剤のPhase I 試験では治療薬のPhase Iで行われるdose escalationを行う必要はないとの考え方もある。その場合、PET薬剤のPhase Iの症例数の目安として、治療薬におけるdose escalation試験での一用量あたりの症例数である3-6例を行えばよい、との提案もある。

4. 2. 画像撮像条件の評価

関連する画像撮像条件(例えば、薬剤投与から撮像までの時間、撮影時間、機器の設定、患者の体位など)を変えることが画質や再現性に与える影響を、そうした条件の変更によって発生する問題も含めて、製剤開発の早い段階で評価しておくことが望ましい。その後の有効性試験では、こうした使用条件の妥当性を立証し、場合によっては修正することが望ましい。

4. 3. 画像評価の方法と基準

画像評価の方法と基準（画像解釈の基準も含む）は、製剤開発の早い段階で評価しておくことが望ましい。その後、臨床使用で予定される方法と基準を採用して、第III相有効性試験で実証することが望ましい。例えば、解析のための関心領域の選択方法などは、

早い段階の臨床試験で比較することが望ましい。同様に、画像の客観的な特徴（例えば、病変の明瞭性、相対的放射能濃度の測定値など）のどれがもっとも有効な指標となるか、たとえば腫瘍の良・悪性の鑑別といった画像解釈の際にもっとも有用であるかを、製剤開発の早い段階で評価することもある。

画像評価のためのこうした方法および基準のうちもっとも適切なものを、第Ⅲ相有効性試験のプロトコルの中に組み込むことが望ましい。

さらに、以下の項目を固定してから、有効性試験の被験者の登録を開始することが望ましい。

- 将来の臨床使用適応
- 有効性試験のプロトコル
- 臨床試験薬概要書
- 臨床試験担当医師用の症例報告書様式（CRF）
- 画像の盲検化評価の計画
- 盲検読影者用の症例報告書様式（CRF）
- 統計解析の計画
- 臨床試験実施施設での画像評価の方法（もし行うなら）と、その評価を当該患者の診療に使う方法（もし使うならば）

統計解析については、主要な有効性試験のそれぞれについて総合的な統計解析計画を提出することが望ましい。この統計解析計画は試験プロトコルの一部とし、盲検化した画像評価の計画が含まれており、統計解析計画がプロトコル上に規定されてから画像の収集を開始することが望ましい。

4. 3. 1. 画像評価

PET 薬剤の画像については、「客観的画像所見」、「画像解釈」、および、「主観的画像所見」により評価する。被検薬剤の適応や臨床現場で想定される利用法を勘案し、臨床試験の目的に応じて、これら所見は「主要評価項目」や「副次的評価項目」として用いることができる。

客観的画像所見とは、ターゲット／バックグラウンド比、摂取率、または、病変の大きさ・数等のように、客観的に数値で測定できる画像所見や、視覚によって定性的に得られる客観的所見をいう。客観的画像所見から臨床的意義が明らかである場合は、その所見を検証的臨床試験の「主要評価項目」とすることが可能であるが、更なる解釈が必要な場合は、「副次的評価項目」にすることが望ましい。

画像解釈とは、病変の有無、病変の性状（良性／悪性、虚血の程度等）など、客観的画像所見に基づいて、臨床的に判断されるものをいう。画像解釈は、それ自体が臨床

的意義を持つものであり、検証的試験の「主要評価項目」として取り入れることができる。

主観的画像所見とは、評価者の経験などから感じ取れる画像所見や診断の確信度をいう。検証的試験の「副次評価項目」として取り入れることができる。

4. 3. 1. 1. 客観的画像所見の評価

本ガイドンスにおいては、「客観的な画像所見」とは、視覚的に認知できる、または機器（コンピュータ）を用いて測定できる画像の特徴のことである。客観的な画像所見の例としては、信号対雑音比、描出の程度、不透明性の範囲、病変の大きさ・数・放射能濃度などがある。

客観的な画像所見は、連続値（例えば、腫瘍の直径など）、順序値（例えば、明らかに増加、おそらく増加、増加も減少もない、おそらく減少、明らかに減少など）、二項値（例えば、あり、なしで分類できる特徴など）といった尺度で得られる。定性的またはスコアなど半定量的の尺度に関しては、その基準（たとえば、縦隔より強く肝臓より低ければ中等度、それ以上なら高度）を示す必要がある。また、定量測定では、ROI（関心領域）の設定方法、病変の大きさの測定方法等、画像所見の測定値に影響しうる要因を予め規定しておく必要がある。

PET 薬剤は、客観的な画像所見によって望む効果を示そうとするものである。画像におけるそのような所見の特徴と位置付けについては、有効性の実証を目的とした臨床試験の画像評価の中ですべて記述することが望ましい。さらに、その記述には、目的とする望ましい所見だけでなく、意図しない望ましくない所見についての情報も含めることが望ましい。例えば、腫瘍のイメージングを目的とした PET 薬剤が肝臓にも集積することがあり、そのために腫瘍の一部の描出が曖昧になることがある。

可能ならば、機器を用いた画像の定量的分析だけでなく、視覚的な画像の定性的評価も行うことが望ましい場合が多い。機器を用いた定量的な画像分析は、例えば臨床現場で想定されていないか不可能の場合には、実際的な意味でそれ単独では画像診断薬の有効性の実証としては不十分だからである。

4. 3. 1. 2. 画像の解釈

本ガイドンスにおいては、「画像の解釈」とは、客観的な画像所見に基づく説明または意味づけのことである。画像所見の解釈は、画像から得られる客観的で定量的および/または定性的な情報によって支持されることが望ましい。例えば、腫瘍組織が画像上では炎症、壊死、もしくは正常に見えるという解釈は、PET 薬剤の腫瘍における局在分布や広がり（例えば、増加、正常、減少、消失、など）、その局在分布の経時的変化、運動や食事、薬剤負荷によってこれらの特徴が影響を受ける様子などの、客観的な画像所見に支持される。

4. 4. 試験におけるエンドポイント

PET 薬剤は、機能的、生理学的または生化学的な評価；疾患や病態の検出または評価；患者の診断や治療における方針決定；これらの複合またはその他の適応、を目的に開発される。主要エンドポイント（反応変数）は、その適応の臨床的有用性に関連したものでなければならない。

臨床的に有用な画像の解釈は、臨床試験の主要エンドポイントに組み込むことが可能である。

4. 4. 1. エンドポイントとしての客観的な画像所見

特定の客観的な画像所見の臨床的意義が明確で明らかである場合には、客観的な画像所見を主要エンドポイントに組み込むことが可能である。

しかし、場合によっては、特定の客観的な画像所見の臨床的意義が、さらなる解釈を加えないと容易には明瞭にならない場合がある。そうした場合、その客観的な画像所見はイメージングの副次的エンドポイントとすることが望ましい。

4. 4. 2. エンドポイントとしての主観的な画像評価

本ガイダンスにおいては、「主観的な画像評価」とは、読影者が感じ取る印象やそれに基づく判断のことである。こうした評価は明白であっても客観的に把握できないものをいう。

主観的な画像評価は、その評価の客観的な根拠が理解しやすいように、客観的な画像所見と直結されていることが望まれる。主観的な画像評価は、妥当性の検証と再現が難しい場合があり、そうした評価に大きなバイアスが入る可能性を除外することができない場合がある。したがって、主観的な画像評価は、イメージングの主要なエンドポイントとして用いるべきでない。

4. 4. 3. エンドポイントとしての臨床転帰

症状、臓器機能や QOL、生存率などといった臨床転帰は、臨床的有用性を測るものとも直接的な方法である。臨床転帰は、画像診断薬の試験において主要エンドポイントとすることができる。例えば、結腸癌患者の治療方針の選択を適応として意図された PET 薬剤の試験の主要エンドポイントは、症状、QOL、あるいは生存率の変化を測定する反応変数となるだろう。

4. 5. 症例報告書様式 (CRF)

PET 薬剤を含めた画像診断薬の試験における症例報告書様式 (CRF) には、担当医師が記録する観察や評価の項目をあらかじめ定めることが望ましい。CRF には、通常

の臨床試験で記入されるデータ（選択/除外基準、安全性の所見、有効性の所見など）に加えて、画像診断薬については検査担当医師による次の情報が含まれることが望ましい。

- 試験に用いた PET 薬剤の放射能投与量、および品質に関する特記事項（比放射能など、もしあれば）。
- 撮像装置などの機器の品質管理事項（例えば、点検日や校正結果）。
- PET 検査における画像収集方法、再構成条件、出力方法、表示条件、保管の方法。

これらのデータは将来当該 PET 薬剤の臨床使用ガイドラインを作成するにあたって重要になる場合がある。

4. 5. 1. 画像評価用の症例報告書様式（CRF）

画像評価のための CRF は、客観的な画像所見と、あらゆる所見の部位と解釈を記入するなど、イメージングのエンドポイントが得られるようにデザインされていることが望ましい。画像所見の解釈は、画像から得られる客観的な定量的もしくは定性的な情報で支持されていることが望ましい。画像の解釈は、客観的な画像所見の評価とは異なる項目として記録することが望ましい。さらに、画像評価のための CRF の項目は、求められる答えを誘導するようなバイアスを持ち込まずに情報が集められるように、注意深く設計されることが望ましい。将来の臨床使用指針に記載される適応は、CRF の特定の項目と、プロトコルの中であらかじめ記述されたエンドポイントと仮説から明瞭に導かれるものであることが望ましいからである。

4. 6. 盲検化された画像評価

画像評価は、PET 薬剤による特異的な画像所見が再現性を持って得られることが示されるようにデザインされることが望ましい。第Ⅲ相有効性試験においては、複数の独立した読影者による盲検化された画像評価を実施することが望ましい。

有効性を示すための主要な画像評価として、独立した読影者による「完全盲検化された画像評価」もしくは「転帰を盲検化した画像評価」のいずれかを用いることが望ましい。あるいは、両方の方法を用いることも可能で、その場合は段階的非盲検化によって実施することが可能である。イメージングの主要エンドポイントおよび副次的エンドポイントの両者が、こうした画像評価によって評価されることが望ましい。

次の節で概説する項目の他にも、盲検化画像評価の計画には次の要素を含めることが望ましい。

- プロトコルにおいて、読影者に対して盲検化する項目を明瞭に記述することが望ましい。
- すべてのエンドポイントの意義は、一貫性を持って明確に理解されることが望ましい。

い。画像の評価と分類の中で用いられる用語は、画像評価計画の中で明確に定義されていることが望ましい。盲検化された読影者は、第Ⅰ相および第Ⅱ相の試験で得られたサンプル画像を用いて、判定方法のトレーニングを受けることができる。

- 画像は、患者を特定できるものを全てマスクしておくことが望ましい。
- 盲検化された読影者は、画像の評価を無作為な順序で行うことが望ましい。画像の「無作為化」とは、試験で得られた画像を回収しその画像セットを実現可能なもつとも徹底した程度に無作為な順序で読影者に提示することを言う。

例えば、同じ基準で読影される何種類かの PET 薬剤の画像について、相対的な有効性を確立するために比較する場合(例えば、被検薬を既承認薬剤と比較するなど)には、読影者は、画像セットの中から画像を無作為の順序で評価することが望ましい。

4. 6. 1. 完全盲検化した画像評価

完全盲検化した画像評価の際には、読影者に対して以下のような情報を全く提供しないことが望ましい。

- 真のスタンダード、最終診断、患者の転帰の結果。
- 患者に特異的な情報(例えば、病歴、身体所見、臨床検査の結果、その他の画像検査の結果)。

患者の登録に関する一般的な選択・除外基準や、プロトコルの詳細、画像の解剖学的方向 (anatomic orientation) も読影者に開示しないことが望ましい場合がある。

複数の異なる処置で得られた画像を評価する試験において「完全盲検化した画像評価」の際には、読影者に対して、処置の内容を可能な限り開示しないことが望ましい。例えば、2種類以上の画像診断薬(もしくはある1つの画像診断薬の2種類以上の用量や投与方法)を比較する試験の中では、盲検化した読影者に対してどのイメージング剤(もしくはどの用量または投与方法)で得られた画像であるかを開示しないことが望ましい。

画像評価の方法が処置によって異なる場合には(例えば、PET 画像のみと PET/CT を比較する場合など)、読影者を処置法について盲検化することができない場合もある。

4. 6. 2. 転帰を盲検化した画像評価

「完全盲検化した画像評価」の場合のように、「転帰を盲検化した画像評価」を行う読影者は、真のスタンダードの評価、最終診断、患者の転帰の結果について知らされていないことが望ましい。

しかし、「転帰を盲検化した画像評価」では、読影者は患者に特異的な情報の特定の要素(例えば、病歴、身体検査、臨床検査の結果、その他の画像検査の結果など)につ

いては知っていてもよい。場合によっては、読影者は患者登録における全体的な選択・除外基準、プロトコルのその他の詳細、画像の解剖学的方向などを知っていてもよい。読影者に提供される情報の特定の要素については、全患者について標準化し、臨床試験プロトコルや統計解析計画、盲検化画像評価計画の中であらかじめ定義されることが望ましい。

異なる処置で得られた画像を比較評価するような試験の場合には、読影者は処置の内容に関して、可能な限り知らされていないことが望ましい。

4. 6. 3. 段階的非盲検化

本ガイダンスにおいては、「段階的非盲検化」は、通常、画像を評価するごとに読影者に与える情報(臨床的情報など)を徐々に増やして読影する方法である。段階的非盲検化は、ルーチン的な臨床現場で起こりうるいろいろな条件のもとでの画像評価を念頭に置いている(例えば、臨床的情報がない時、限られた臨床的情報がある時、十分な情報があるとき、など)。この方法は、試験薬剤が診断アルゴリズムの中でいつ、どのように用いられるべきかを決定する際に用いられる。典型的な「段階的非盲検化」画像評価は、以下の3段階の過程で実施することが望ましい。

- まず完全盲検化での画像評価を行う。この評価を記録し、妥当な方法によってデータセットを固定する。固定されたデータセットでは、後に新たな情報が利用できるようになっても、治験担当医師や他の読影者等から追加情報があった場合でも、評価を変更することができないようにする。
- 次に、PET 検査時点での患者と撮像に関する情報を開示して、転帰を盲検化した画像評価を行う。この評価を記録し、データセットを固定する。
- イメージング剤の診断性能を決定するために、上記2つの盲検化評価を真のスタンダード(または最終診断や患者の転帰)の評価の結果と比較する。

段階的非盲検化を複雑化すれば、読影者にきめ細かく新たな臨床情報を提供する方式も可能となる。段階的非盲検化が用いられる場合には、プロトコルにおいて各段階で評価すべき仮説を特定することが望ましい。また、有効性の決定の際に、どの段階の画像評価が主要評価となるかを、プロトコルで特定しておくことが望ましい。

4. 7. 非盲検化画像評価

「非盲検化画像評価」においては、読影者は、真のスタンダードによる患者評価、最終診断、患者の転帰などの結果を知っている。さらに盲検化されていない読影者は、患者に特異的な情報(病歴、身体検査、臨床検査の結果、その他の画像検査の結果など)、複数の処置で得られる画像を評価する場合の処置(無処置も含む)の内容、患者登録の際の選択・除外基準、プロトコルのその他の詳細、画像の解剖学的配置について知って

いる。

非盲検化画像評価は、完全盲検化した画像評価や、転帰について盲検化した画像評価の結果と一致していることを示すために用いることができる。これらの盲検化画像評価と非盲検化画像評価は、結果の比較が可能なように同一のエンドポイントを用いることが望ましい。しかし、非盲検化画像評価は有効性の証明における主要画像評価としては用いないことが望ましい。盲検化されない読影者は、新たな情報を得ることがあり、こうした追加情報によって、読影者の診断的評価が変化したり、その読影者による画像評価が攪乱され、バイアスがもたらされたりする可能性がある。

4. 8. 独立した画像評価

一つの事象を知っていても、もう一つの事象について何も言うことができないとき、その2つの事象は独立であるという。したがって、本ガイダンスにおいては、「独立した読影者」とは、他の読影者の所見（他の盲検化された読影者や撮像担当医師等の所見も含む）について何も知らず、他の読影者の所見によって影響を受けない読影者同士のことである。盲検化された読影者の評価を独立に保つために、盲検化されたそれぞれの読影者の評価は、評価が得られたら、別の種類の画像評価が行われるよりも前に、すぐにデータセットに入れて固定することが望ましい。

4. 9. コンセンサス画像評価

本ガイダンスにおいては、「コンセンサス画像評価（コンセンサス読影）」とは、複数の読影者が相談しながら一緒に画像を評価するために集合して行う画像評価のことである。コンセンサス画像評価は、個々の読影者の読影が完了して固定された後に実施する。コンセンサス読影では、読影者は互いに独立ではないと考えられ、それゆえにその種の読影は画像診断薬の有効性を示す主要画像評価としては用いないことが望ましい。コンセンサス読影は複数の読影者によって実施されるが、実際には単独の画像読影とみなすべきであり、盲検化された複数の独立した読影者による画像評価に代わるものではない。

個々の盲検化評価と同じように、コンセンサス読影は、一度評価が得られたら、その後には別の盲検化読影が実行される前に固定することが望ましい。

4. 10. 同一読影者による反復的画像評価

読影者が同じ画像を複数回（段階的非盲検化のように、あるいは、読影が読影者内変動を評価するようにデザインされている場合のように）評価する試験においては、その複数回の読影は可能な限り互いに独立して行うことが望ましい。これは、思い出しバイアスを最小限にするためである。さらに、読影者は可能な限りそれまでの自身の読影所見について知らされておらず、自分自身の以前の所見によって影響を受けないことが望

ましい。

同一読影者による反復的画像評価 2 つの異なる画像評価を行うためには、症例報告書様式 (CRF) の異なるページを用い、それぞれの画像評価は、以前の結果を思い出したり参照にしたりすることを低減させるために、読影の時間間隔を十分に与えることが望ましい。

4. 1 1. 施設外と施設内での画像評価

本ガイダンスにおいては、「施設外での画像評価」は、試験の実施に全く関係のない場所で、患者・検査担当医師・試験に関係するその他の者と全く接触のない読影者によって行われるものである。第Ⅲ相試験では、限られた数の場所(一箇所集中が望ましい)での施設外の画像評価を行うことが望ましい。このような施設外での評価においては、通常、盲検化画像評価の統一性を損ねる可能性のある要因を制御しやすく、盲検化された読影者がその他の画像評価とは独立して画像評価を実施することを保証しやすい。

本ガイダンスにおいては、「施設内での画像評価」とは、プロトコルの実施や患者の管理に関わった検査担当医師が行う画像評価のことである。この用語は、試験の実施に関係のある場所で行われる盲検化画像評価のことも指す。施設内の検査担当医師は、臨床試験プロトコルの中にはあらかじめ定められていなかった患者に関する追加情報を得ることがある。こうした追加情報は、検査担当医師の診断的評価を変化させたり、検査担当医師による画像評価を攪乱したりバイアスをもたらす可能性がある。したがって、施設内での画像評価は、通常、有効性を示すための主要画像評価として用いず、盲検化画像評価を支持するもの、とみなすことが望ましい。

しかし、施設内の検査担当医師が、転帰について盲検化した画像評価において患者の最終転帰を知らない場合などは、主要画像評価を実施できる可能性がある。そのような場合、検査担当医師が画像評価の時点で使用可能なすべての臨床情報が明確に特定され、すべて記録されていることが望ましい。また、そうした情報が読影にどの程度の影響を与えるかについて批判的評価を行うことが望ましい。さらに、有効性の所見を支持する独立した盲検化評価も行うことも有効である。

4. 1 2. 読影者内および読影者間の変動の評価

有効性を示すことを目的とする各試験について、少なくとも 2 名の盲検化された読影者 (3 名以上が望ましい) が画像を評価することが望ましい。複数の読影者が存在することにより、読影の再現性 (すなわち読影者内変動) の評価が可能になり、得られた所見の一般化に対してよりよい根拠が与えられる。理想的には、読影者間の一致性が測定できるように、各読影者が、有効性を示すことを目的とした全ての画像、すなわち被検薬の画像と参照画像のすべてを評価することが望ましい。大規模試験では、全ての画像を読影者全員が見ることは実行不可能な場合があり、適切に選定した部分集合について

そのような複数の読影者による画像評価を行うことが可能である。読影者間の一貫性については、定量的に測定することが望ましい（例えば、 κ 統計量など）。

読影者内の変動は、画像診断薬の開発段階で評価することが望ましい。これは、画像の全てもしくは一部について、個々の盲検化読影者が反復して画像評価を行うことで達成できる。

4. 1 3. プロトコル画像と非プロトコル画像

画像診断薬の臨床試験で得られた画像は、一般的に、プロトコル画像あるいは非プロトコル画像のいずれかとみなされる。

4. 1 3. 1. プロトコル画像

本ガイダンスにおいては、「プロトコル画像」とは、有効性を証明または支持するという目的の下に、プロトコルで特定された方法と条件と時点で得られた画像のことである。有効性の評価は、こうしたプロトコル画像の評価に基づくべきである。また、全てのプロトコル画像（評価対象と決定した画像のみではなく）を、盲検化された読影者によって評価することが望まれる。それには検査の対象患者の画像、対照被験者や健常者の画像が含まれる。さらに、プロトコル画像の評価は、その他の画像、すなわち非プロトコル画像などが読影者にレビューされるまでに完了させておくことが望ましい。

大量の画像や、画像テープ（エコー心電図など）が得られ場合に評価する画像を恣意的に選択することは、選択者のバイアスが入り込む可能性があるので勧められない。

欠測となった画像（または技術的に不適切、解釈不能、結果が確定できない、あるいはそのことに関して判定できないとされた画像）を解析時にどう扱うのかを有効性試験のプロトコルにあらかじめ定めておくことが望ましい。統計解析計画の中に、**intention-to-treat** の原則に基づきそれを診断の場面に適合させた解析を組み込むことが勧められる（例えば、**intention-to-diagnose** は診断薬の試験において、その診断薬を用いて撮像されたか否か、また、その画像の質に関わらず組み入れられた全ての被験者をさす）。画像が解析から除外されてしまうには多くの理由があり、患者の脱落、イメージングの技術的問題、プロトコル違反、画像の選択の問題などがある。主要な反応変数の解析において、値の欠損を取り扱うための適切な方法をあらかじめ構築することが望ましい。

4. 1 3. 2. 非プロトコル画像

本ガイダンスでは、「非プロトコル画像」とは、上述のプロトコル画像ではない画像を指す。このような画像は、時に探索的な目的で撮像されることがあるが、固定した第Ⅲ相データベースには含まれない。

4. 1 4. 個別の画像評価と組合せ画像評価

個別の画像評価を行っても、組合せた画像評価を行ってはいらないわけではなく、その逆も同じである。しかし、もし複数の画像評価が実行されるならば、どの画像評価が主要評価として扱われ、どの画像評価が副次的評価として扱われるかを、プロトコルで特定しておくことが望ましい。

4. 1 4. 1. 個別の画像評価

本ガイダンスにおいて、「個別の画像評価」とは、読影者による試験対象画像について、その患者から得られた他の検査画像と可能な範囲で最大限に独立して評価を行うことを指す。個別の画像評価においては、読影者は患者のそれぞれの検査画像を、可能な範囲で最大限に、同じ患者から得られた他のどの検査画像も参照にせず、また思い出すこともなく評価する。

個別の画像評価は、複数の画像を同時に見ることなく、かつ患者群の画像を時間経過順に評価することがないように、無作為の順序で個別に評価される。また、ある一つの条件下（または特定の時期）で得られた検査画像を、無作為の順序で個別に評価した後、複数の条件下（または異なる時期）で得られた個別の検査画像を無作為の順序で評価することもできる。

以下の例で述べるように、試験の目的が製剤の性能に関する比較である場合には、適切にデザインされた個別の画像評価を実施することが望ましい（例えば、ある画像診断薬の診断性能を他のものと比較するなど）。

例 製剤性能の比較推定

新規の画像診断薬の診断性能が既存薬よりも優れていることを証明するようにデザインされた比較試験では、主要画像解析として適切な個別の画像評価を実施することが望ましい。この場合での試験対象画像は、新規および既存の画像診断薬で得られる画像である。この2種類の薬剤は、実際の臨床現場では一緒に使用されることは意図されておらず、それゆえに、この様な場合の画像評価は、新薬で得られる情報が既存薬で得られる情報よりも臨床的・統計的に優位であることを示すことを目的とすることが望ましい。すべての患者において、実現可能な最大限の範囲で、新薬で得られる画像を、既存薬で得られる画像の評価と独立して評価することが望ましい。

希望するならば、副次的な画像解析として、新薬の画像と既存薬の画像の並列（対比）比較を行うことができる。しかし、このような並列比較では新薬の診断性能の評価にバイアスがかかる可能性がある。盲検化読影者は、この並列比較において新薬で得られた画像の中で腫瘍の存在を過大読影する傾向がある可能性がある。同様に、盲検化読影者は、既承認薬で得られた画像に明瞭に出現していない腫瘍は並列評価において新薬で得られた画像の中で過少読影する傾向がある可能性がある。

通常、画像評価におけるこれらの方法は非劣性を示すためにデザインされた試験にも適用できる。試験に被験者が登録される前に、試験デザインや解析計画について十分検討することが望ましい。

4. 14. 2. 組合せ画像評価

本ガイダンスにおいては、「組合せ画像評価」とは、異なる条件もしくは、投与時からの異なる経過時間に得られた複数の検査画像を、読影者が同時に評価することを指す。組合せ画像評価は、臨床で PET が用いられる条件に似ていることがある。例えば、ある臨床状況では、投与 1 時間後の検査と投与 2 時間後検査の両方が患者に対して行われることがある。その場合、そうした画像の評価は、対比の形で同時に行われることが多い。しかし、上述のように、組合せ画像評価によって、画像評価にバイアスが入り込む可能性が増大する場合がある（例えば、特定の画像所見に対する系統的な過大読影もしくは過小読影など）。

組合せ画像評価は、各患者について画像の組合せセットを作ることで実行できる。続いて、このセットを、無作為な順序で盲検化読影者に提示する。

しかし、この種の読影が行われるときには、組合せのうち少なくとも一つの画像について独立した「個別の画像評価」をさらに行うことが望ましい。個別の画像評価が現在の標準的手技で得られるものであれば、この方法で、「組合せ画像」と現在の標準的手技との差を評価することが可能である。新たな画像診断薬や追加撮像によって画像に情報が付加されることを示すことが目的ならば、組合せ画像からの情報が、個別の画像単独から得られる情報よりも臨床的にも統計的に優れていることを示すことができる。組合せ画像評価と個別の画像評価の結果は、対応のある比較によって統計的に分析することが可能である。

組合せ画像評価と個別の画像評価は、思い出しバイアスを軽減するために、互いに独立して行うことが望ましい。組合せ画像評価および個別の画像評価には、CRF の異なるページを用いること、また組合せ画像評価と個別の画像評価は、過去の結果を参照することなく、時期を変えて実施することが望ましい。

組合せ画像評価と個別の画像評価との間の差を評価したい場合には、組合せ評価の CRF と個別評価の CRF には、差が算出できるように、またバイアスを軽減することができるように、比較判定を要求する質問を使用せずに、同じ項目または同じ質問を含めることが望ましい。

4. 15. 真のスタンダード (ゴールドスタンダード)

真のスタンダードは、研究用の画像診断薬によって評価されるのと同じ項目を独立して「正しく」評価する方法を提供する。真のスタンダードは、患者の真の状態もしくは真の計測値であることが判っている、もしくはそう考えられているものである。真のス

スタンダードは画像診断薬によって得られた結果が妥当で信頼性のあることを示すために、また検査の要約統計量(例えば、感度、特異度、陽性的中度および陰性的中度など)を求めるために用いられる。次の一般的原則を画像診断薬の第Ⅲ相有効性試験のデザイン、実施、分析の中にあらかじめ組み込んでおくことが望ましい。

画像診断薬によって得られた検査結果は、真のスタンダードによって得られた結果や転帰が開示されていない状態で判定されることが望ましい。

被験者の真の状態(例えば、疾患がある、または、ないなど)は、画像診断薬を使用して得られた検査結果を開示されていない状態で、真のスタンダードによって判定されることが望ましい。

真のスタンダードには、画像診断薬を用いて得られたいかなる結果も要素として含めないことが望ましい(組込みバイアスを避けるため)。しかし、真のスタンダードの一部に、他の撮像モダリティの結果を含めてもよい。

真のスタンダードによる評価は、登録された全ての被験者について計画され、真のスタンダードで評価する被験者の決定は、試験対象の画像診断薬による検査結果の影響を受けないことが望ましい。例えば、被検薬剤で陽性の結果となった患者を(陰性の検査結果となった患者よりも)優先的に真のスタンダードで評価すると、試験の結果は、“**partial verification** バイアス”に影響される可能性がある。同様に、被検薬剤で陽性の結果となった患者を優先的に真のスタンダードで評価し、陰性の結果となった患者を優先的に厳密性に乏しいスタンダードで評価した場合には、試験の結果は、“**differential verification** バイアス”に影響される可能性がある。

かなりの割合の被験者が真のスタンダードで評価されない、もしくは、厳密性に乏しいスタンダードで評価されることが予測される場合、登録された被験者について臨床的転帰を評価するのが適切なことがある。

現実的観点から見ると、真実を得ることにおいて被検薬剤より確定的だと考えられる実際的方法で得られるものが、「診断スタンダード」である。例えば、病理組織診断や長期間の臨床転帰は、腫瘍が悪性かどうかを判断する診断スタンダードとして容認できるものである。診断スタンダードは誤差があるかもしれないが、臨床試験の目的からすれば確定的なものと思なされる。しかし真のスタンダードによって疾患の分類を誤ると診断性能の測定値に陽性バイアスや陰性バイアスが入り込む可能性がある(誤分類バイアス)。したがって、臨床試験の適切性を確保するため、真のスタンダードの選択について試験をデザインする際に十分検討することが望ましい。

真のスタンダードを決定した後、真のスタンダードに関して、検査の要約統計量の仮説を決定し、試験プロトコルにあらかじめ組み入れることが望ましい。その仮説および予測される要約統計量は、そのイメージング剤の使用が意図される臨床状況(例えば、

スクリーニング検査、段階的な評価、他のイメージング検査の代替 (alternative) あるいは置き換え (replacement) など) を反映したものとすることが望ましい。

4. 16. 比較対照群

臨床試験における対照群の選択に関する一般原則は ICH-E10 ガイドライン (臨床試験における対照群の選択とそれに関連する諸問題. 平成 13 年 2 月 27 日医薬審第 136 号.) で示されてされており、これらの原則は診断薬の試験にも適用可能である。

4. 16. 1. 同様の適応で承認された薬剤またはモダリティとの比較

既承認の薬剤や生物学的製剤、または他の診断モダリティよりも優れたものとして被検薬剤が開発されている場合には、承認されている比較対照と直接的に同時に比較する方法が望ましい。比較には、比較対照と被検薬剤の安全性と有効性の両者の比較を含めることが望ましい。疾患の多様性のために、このような比較は同一患者において実施されることが多い。試験対象の薬剤もしくはモダリティの画像評価は、承認された薬剤もしくはモダリティで得られた画像の結果を開示せずに行うことが望ましい。

試験対象の画像と比較対照の画像 (すなわち、新しい方法と古い方法) の両者から得られた情報は、画像同士の比較だけでなく、独立した真のスタンダードと比較することが望ましい。これによって、画像診断薬と比較対照との間にありうる差異が強調され、診断性能の比較評価が可能になる。そのような評価は、例えば、それぞれの診断薬の感度、特異度、陽性的中度および陰性的中度、尤度比、関連する測定値、受信者動作特性 (ROC) 曲線を比較することで得られる。2 種類の画像診断薬が、同じ患者群において得られた感度と特異度の値が類似していても、互いの一致率が低い場合がある点に注意すること。同様に、2 種類の画像診断薬で、一致率は高いが感度と特異度の値が低い場合がある。ROC 解析において、それぞれの薬剤で得られる全曲線下面積は同等でも、曲線のある範囲での曲線下面積は異なるということがある。同様に、ある診断薬の診断性能が ROC 曲線上のある 1 点において他のものより優れていても、別の点では診断性能が劣っているという場合がある。

既存の画像診断薬または診断モダリティと同じ適応について画像診断薬が開発される場合には、既存の薬剤や診断モダリティに対する直接的な同時比較が勧められる。しかし、特定の適応で既に承認されている画像診断薬があっても、その薬剤を単独で用いた検査が真のスタンダードとして使用できるとは限らない。例えば、ある画像診断薬の所見が、病理組織診断で確認される真の状態と十分に一致することを根拠に既に承認されている場合、開発している画像診断薬の評価の中には、病理組織診断による真の判定も含めることが望ましい。この場合、開発している画像診断薬と既承認薬とを真のスタンダードとしての病理組織診断により直接的かつ同時に評価することによって、2 剤の性能の差を最もよく示すことができる。

被検薬剤の効果を、他の画像診断薬またはイメージングモダリティと比較する試験においては、試験への登録前に、試験で評価しない薬剤を用いて得られた画像を患者の選択基準以外に用いないことが望ましい。こうした画像は、被検薬剤の性能を決定するために用いるデータベースの中に含めないことが望ましい。このようなベースラインの登録時画像は盲検化されておらず、検査の依頼や管理の意向に基づいているので、選択バイアスを有している。

また、被検薬剤の投与による検査は、疾患の経過が顕著に変化しないと予想される時間枠の中で実施されることが望ましい。これによって、被検薬剤と比較対照薬との公平かつ安定した比較が可能となる。

4. 16. 2. 非劣性試験

新しい薬剤が対照薬より劣性ではないことを示すために試験をデザインすることができる。一般的にこのような試験における要求は、優位性を示すためにデザインされる試験の要求より厳しい。特に、イメージングの試験は、不適切な対象集団、客観的なイメージングエンドポイントの欠落、真のスタンダードの誤りなどいくつかの理由により分析感度が評価できない可能性がある。さらに、イメージングの試験では薬剤の効果に対する分析感度について過去に根拠がないことが多いため、分析感度の評価の検証が困難であり、また、イメージングを行い、画像を評価することが、有効な治療と有効性が低い治療を区別する試験の能力を低下させなかったということがいつも明確であるとは限らない。これらの問題についてのさらなるガイダンスは ICH-E10 に述べられている。

非劣性試験は被験薬と対照薬との同時比較に基づくこと、そしてその試験は、適切な真のスタンダードによって検証された客観的に定義されたエンドポイントを用いることが望ましい。そのようなデザインにより、新薬と対照薬の診断的（あるいは機能的）性能の比較評価が可能になる。例えば、試験のエンドポイントが疾患の有無である場合、被験薬と対照薬の感度と特異度がそれぞれ比較できる。統計的仮説は優位性、非劣性またはその両者の可能性がある。被験薬が疾患の除外を第一の目的として使用される予定ならば、高い陰性的中度、したがって、高い感度が特異度よりも重要となると考えられる。そして、目的は、新薬が対照薬と比較して感度は優れていて、特異度は劣らないことを示すことになる。

試験デザインに真のスタンダードが設定されているが対照薬との比較がない場合、新薬の性能の値はある固定された閾値(例えば、事前に指定された感度および特異度の値)との比較のみになる。そして、統計上の目的は、閾値に対して優位性を示すことであるべきである。そのような値は、閾値を上回ることが被検薬剤の有効性を明確に示すという主張を支持するような実質的な臨床的エビデンスに基づくべきである。

対照薬に対して非劣性であるという主張を得るためには、被験薬が対照薬と類似した

性能特性を有することが示されていることと、厳密に定義された臨床状況の中で代替のモダリティとして使用できることを示すべきである。別の状況では、非劣性比較は被験薬の有効性を示す役割のみのこともある。一般に、非劣性試験は、新しい検査と比較対照の検査の性能の差が、最大でも臨床的に許容される範囲内であることを示すためにデザインされる。

4. 16. 3. 一致性試験

新しい試験薬と対照薬が一致して同じ結果となることを示すことによっても、両剤の類似性が示されることがある。その場合、真のスタンダードを用いることが不可能で、たとえその結果の妥当性(診断精度)が検証されていなくても、被験薬の結果と比較対照の結果の一致性を示すことが目的となる。新しい試験薬と対照薬の一致性が高いことにより、新しい試験薬は対照薬の代替として許容できることが支持される。

一致性試験では、分析感度が重要である。特に、転帰は客観的に定義され、また2剤は、適切なスペクトラムの疾患の状態(disease condition)にある患者において比較されなくてはならない。例えば、2つの診断検査が、被験者群の大多数において同じように陽性の診断となることを示しても十分ではない。陰性の診断が優勢の被験者群に対しても被験薬と対照薬が同様の反応を示し、結果が不一致の可能性はごく僅かであることを示す必要がある。結果が陽性陰性の2つではなく多数の値である場合、試験値の全てにわたって一致性が示されなくてはならない。

一致性の仮説は、単に被験薬と比較対照の結果の一致性が、比較対照間での結果の一致性を上回るという仮説にするべきではなく、検査の再現性読影者内・間変動をも考慮しなくてはならない。例えば、心筋血流の低下を評価する血流イメージングに用いられる新しい薬剤負荷用の薬剤について考えてみる。適用可能と考えられる1つのデザインは、すべての被験者が、1回目の検査は比較対照の検査を受け、2回目の検査は比較対照の検査か新しい試験薬の検査のどちらかを受けるように被験者をランダム化するという方法である。この方法で、非劣性仮説を用いて、検査内一致性と、比較対照との検査間一致性を直接比較することが可能になる。

一致性試験は新しい試験薬の妥当性のエビデンスを直接示すものではないため、デザインおよび効果的な実施が難しい。したがって、限られた状況の場合のみ一致性試験を行うこととし、許容される真のスタンダードを用いた代替のデザインを検討することが望ましい。

4. 16. 4. プラセボとの比較

画像診断薬の評価においてプラセボの使用が適切かどうかは、個々のイメージング剤、想定される臨床使用目的、イメージングモダリティによる。ある場合には、プラセボを使うことが、試験を実施する上で入り込む可能性のあるバイアスを軽減させるのに役立つ

ち、有効性や安全性の曖昧なデータを解釈しやすくなることもある。

4. 17. 統計解析

統計の手法と診断性能を評価するための手法は、各試験の統計解析計画の中にあらかじめ組み込んでおくことが望ましい。さらに、各試験のプロトコルに、検定される仮説が明確に述べられ、サンプルサイズの仮定と算出根拠が示され、また計画されている統計手法とその他のデータ解析の方法が記述されることが望ましい。

4. 17. 1. 統計の手法

画像評価における1つの要素は、その検査によって、測定しようとしているものがいかに適切に測定できるか（妥当性）を評価することである。その製剤の全体的な診断性能は、感度、特異度、陽性的中度、陰性的中度、および尤度比などの要素によって測定可能である。結果の妥当性は、その検査を用いることによって、臨床的な成績が改善したことを示すことによって証明することが可能である。

イメージング剤の信頼性は、検査結果の再現性を反映する（すなわち、同一患者で再現される測定値、異なる読影者による同一画像の評価の一致、あるいは同一読影者による同一画像の評価の一致など）。

イメージング剤の試験の多くが、二値尺度、順序尺度または分類尺度の結果を得るようデザインされている。そのような解析の中で適切な仮定と統計手法を用いることが重要であると考えられる。割合や比率についての統計学的検定は、分類尺度の結果が用いられるのが一般的であるが、順位尺度については順位付けに基づく手法も多く適用されている。試験の結果は、中心となるサブグループまたはそれ以外のサブグループといった自然な方法で層別されることが推奨され、Mantel-Haenszel 法も二値尺度や順位尺度のデータの解析に有効である。必要であれば、条件付き推論に基づいた正確な解析手法を用いることが望ましい。モデルに基づいた手法の使用も勧められる。そうしたモデルとしては、二値データにはロジスティック回帰モデル、順位データには比例オッズモデルが用いられる。分類尺度の結果に対しては、対数線形モデルが用いられることがある。

新薬と既存薬との比較対照試験においては、個人差を消去するためにクロスオーバー法がとられることが多い。被験者は処置の順序に対して無作為化されることが望ましい。被験者が処置の順序に対して無作為化されない場合は、画像を評価する順序を適切に無作為化することが望ましい。クロスオーバー試験の結果は常に、そのような試験に特化した方法を用いて解析することが望ましい。

4. 18. 診断性能

診断の妥当性は、いくつかの方法で評価できる。例えば、画像診断薬を使用した画像

を真のスタンダードと比較することができる。また、画像診断薬を使用した画像の感度と特異度を、真のスタンダードと比較することもできる。同じ方法で、異なる 2 種類の実薬と比較することもできる。診断の比較は、診断検査の結果が 3 つ以上の結果になる場合でも可能である。診断力の差を検定する一般的な方法としては、McNemar 検定と Stuart-Maxwell 検定がある。さらに、解析において感度、特異度および他の測定値の信頼区間を求めることが望ましい。閾値の範囲を超えて画像診断薬の診断性能を評価するには、ROC 解析も有用である。例えば、2 種類の画像診断薬について、陽性(または陰性)の検査結果を確定するいくつかの閾値を用いることができる場合には、相対的な診断性能を述べるのに ROC 解析を用いることができる。計画するすべての統計解析において、解析方法の詳細と検定する具体的な仮説は、統計解析計画の一部として、あらかじめプロトコルに定めておくことが望ましい。