

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、本ガイドンスにおいては、「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 本ガイドンスにおいては、「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、その後別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、読影者が読影者内変

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とることが望ましい。

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、試験の実施のある場所で行われる盲検化画像評価のことにも指す。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、 $\kappa$  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. 13. 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. 13. 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. 13. 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な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

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、このよう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瘤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瘤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

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. 1 6. 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. 1 6. 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、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。

研究成果の刊行に関する一覧表レイアウト（参考）

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
伊藤健吾	アミロイドイメージングの最近の知見と将来展望	Rad Fan	9	59-61	2011
栗原千絵子	臨床PET検査に関わる米国の規制に関する動向	Rad Fan	9	39-42	2011
栗原千絵子、米倉義晴	核医学研究における被験者の被ばく防護	PET journal	16	39-42	2011
Norinari Honda, et al.	Imaging of ventilation with dual-energy CT during breath-hold after a single vital-capacity breath of stable xenon.	Radiology	262	262-268	2011
倉田精二、井上登美夫他	NaF PET イメージングの臨床	Rad Fan	9	62-65	2011
井上登美夫、立石宇貴秀他	2. 新規PET製剤への期待	日獨医報	56	20-29	2011