



Molecular Imaging CRO Network

Micron's ViewPoint

医療機器プログラム (SaMD) の開発 ～FDAガイダンスの紹介～

Clinical Performance Assessment:
Considerations for Computer-Assisted
Detection Devices Applied to Radiology
Images and Radiology Device Data in
Premarket Notification (510(k))



目次

はじめに	3
CADeシステムの臨床試験デザイン	4
CADeシステムの臨床試験デザインの留意事項	6
- エンドポイントの選択	7
- 読影試験とランダム化	8
- スコアリング	9
- トレーニング	9
- 研究対象集団	10
おわりに	12

免責事項

本文書に記載された内容は、予告無しに変更される場合があります。

株式会社マイクロンは、本文書に関していかなる種類の保証（商用性および特定の目的への適合性の黙示の保証を含みますが、これに限定されません）も致しません。

株式会社マイクロンは、本文書に含まれた誤謬に関しての責任や、本文書の提供、履行および使用に関して偶発的または間接的に起こる損害に対して、責任を負わないものとします。

執筆者から事前に書面による許可を得ることなく、本文書のいかなる部分も複製、再販売または改変することを禁じます。



はじめに

CTやMRI、超音波画像などの医用画像を用いて医師の診断を支援するシステム、すなわち医用画像診断支援 (CAD : Computer Aided Detection/Diagnosis) システムは、疾患の見落とし防止や医師の読影時間の短縮、医師間の診断のバラツキの減少など、医師の診断支援として期待されています。CADシステムを実臨床に導入する決定因子のひとつとして、CADシステムの客観的な診断性能があります。客観的な診断性能のデータはCADシステムが医療機器として認証／承認されるときに重要なデータとなります。

アメリカ食品医薬品局 (Food and Drug Administration、以下「FDA」) は、「Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data in Premarket Notification (510(k))」と題して最終改定ガイダンス (以下「FDAガイダンス」) を2022年9月28日に公表しました。

FDAガイダンスは、CADE (Computer-Assisted Detection) システムの客観的な診断性能を示すために、臨床試験 (臨床性能試験、性能評価試験、臨床研究、非臨床試験などがありますが、本文書では「臨床試験」と記載しています) をどのようにデザインし、実施すべきかについての推奨事項を示すものになっています。ここでCADEとは、患者の放射線画像 (または放射線データ) を意図する使用者 (医師または他の医療従事者など、以下「臨床医」) が解釈する際に、放射線画像の一部を特定し、マーク、ハイライト、または他の方法で注目を向けることを意図するものです。

本稿ではFDAガイダンスについて、CADEシステムの臨床試験デザインを構築する上で特に重要度が高いと思われるものを取り上げ、日本語で記載しています。

CADEシステムの開発や臨床試験を計画する企業や機関にとって、本稿が当該システムの臨床現場実装の一助になればと願っております。



CADeシステムの臨床試験デザイン

CADeシステムは、臨床医が患者の放射線画像（または放射線データ）を解釈する際に異常を示す可能性のあるデータを特定することを目的として設計されています。CADeシステムの出力と臨床医による診断プロセスとの間には複雑な関係があり、臨床性能は以下のような多様な要因に依存する可能性があるため、試験デザインにおいて考慮する必要があります。

- CADeの解釈プロセスでの適用タイミング（Concurrent ReaderまたはSecond Reader）
- CADeマークの特性（サイズ、形状、境界の種類（例：実線、点線、円）、異常との近さ）
- CADeが検出する疾患/異常に対する臨床医の知識
- CADeマークの数

臨床性能評価は、結論に影響を及ぼすかもしれないバイアスを排除または軽減するように、厳密に制御された試験デザインにする必要があります（表1, 2）。これは特に、臨床性能評価が実験室環境で行われる場合（臨床現場とは異なる環境）に推奨されます。

表1 CADeシステムの臨床試験デザインの例

研究デザイン	説明
フィールドテストまたは前向き臨床試験 (例：ランダム化対照試験)	実際の臨床環境でCADeシステムを評価する。有病率が低い場合には、過大な数の症例数が必要となり、実施が現実的でない場合がある。
後向き臨床試験	疾患例を拡充した後向きのケースコレクションからなる後向き臨床試験はフィールドテストの代替となる可能性がある。
ストレステスト	臨床で見られるよりも、読影が困難な画像の割合を高めた後向きの臨床試験。試験と臨床環境とでの有病率及び読影難易度の違いにより、パフォーマンスを変化させる可能性があることに注意する必要がある。

表2 CADeシステムの臨床試験での考慮事項

考慮事項	説明
研究対象集団	研究対象集団（疾患例と正常例の両方）が、目的母集団を適切に代表するものであるか、またはそこからサンプリングされたものであること
交絡因子	交絡を回避するように設計されていること
サンプルサイズ	パフォーマンスを示すためのサンプルサイズが十分であること
参照基準 (Reference Standard)	パフォーマンスを評価するためのTruthの定義が適切であり、参照基準の不確かさが分析されていること(該当する場合)
読影医	臨床医の集団を代表するよう選ばれていること
イメージングハードウェア	臨床現場と一致するよう選ばれていること



株式会社マイクロン

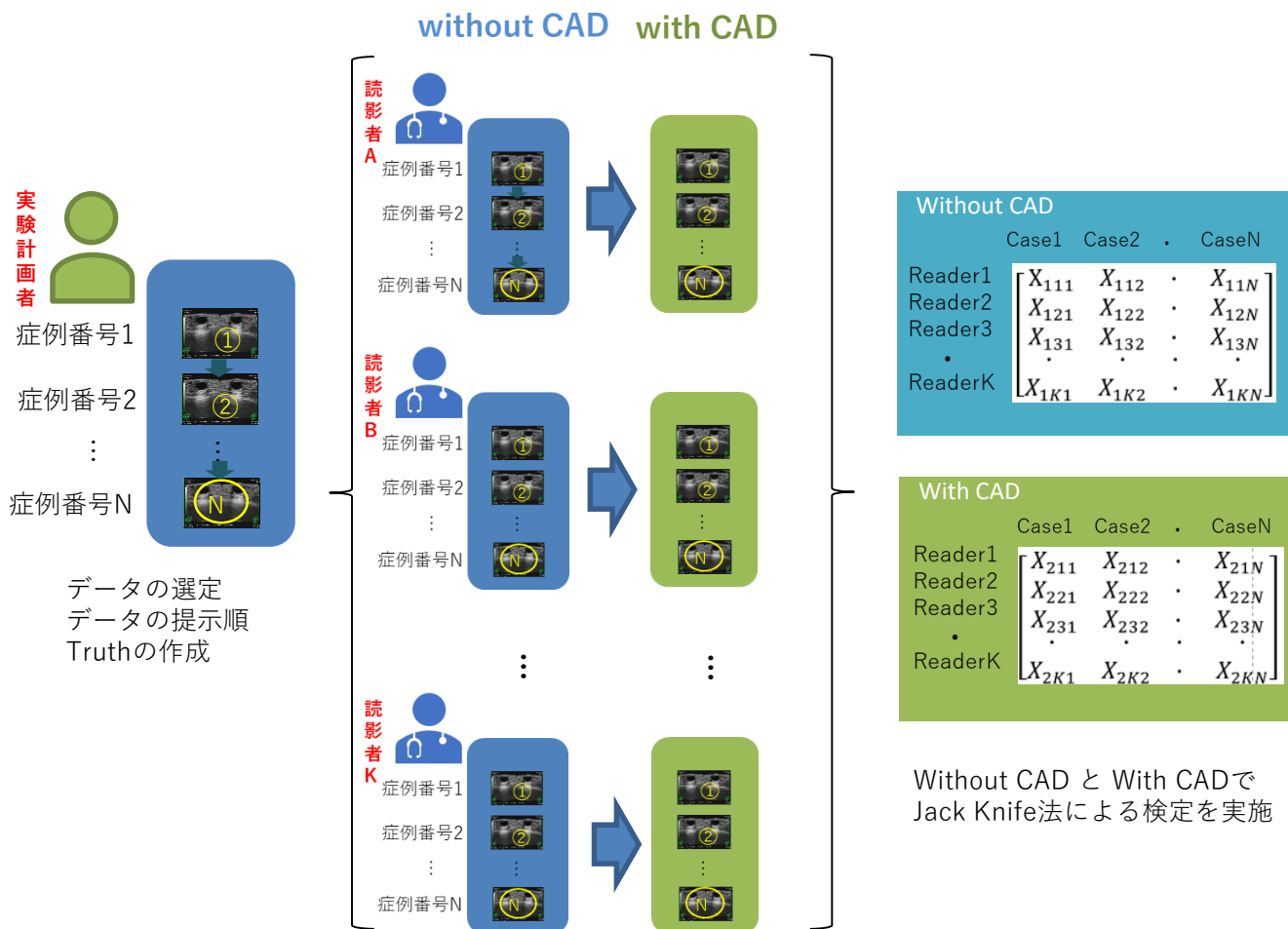
CADeシステムの臨床試験デザイン

CADeシステムの臨床性能評価は、複数の読影者が複数の症例を評価する、Multiple Reader Multiple Case (MRMC) デザインを用いることが一般的です。すべての読影者がすべての症例を独立して読む「完全交差」のデザインで実施することで、症例数あたりの統計的検出力が最大となります。ただし、「非完全交差」のデザインであっても、前向き研究などで、複数の読影者が同一の症例データを読影することが困難な場合には許容される場合があります。

完全交差デザインを選択するかどうかに関わらず、**MRMCデザインを使用して、CADeシステムの臨床性能を評価することが本ガイドラインで推奨**されています。

図1にSecond Reader型のCADeシステムの臨床試験で採用されることが多いSequential DesignのMRMCデザインの例を示します（FDAガイダンスからの抜粋ではありません）。

図1 MRMC実験の例(Sequential Design)



(白石順二, 医用画像の視覚評価法, オーム社から一部変更して引用)



株式会社マイクロン

CADeシステムの臨床試験デザインの留意事項

FDAガイダンスで示されている、CADeシステムの臨床試験デザインにおける留意事項の要約一覧を以下に示します。

以下に挙げた項目のうち、「エンドポイントの選択」、「読影試験とランダム化」、「スコアリング」、「トレーニング」、「研究対象集団」については次ページ以降詳細を記載しています。

エンドポイントの選択

- ・感度及び特異度に加えて、ROC又はその派生形（LROCやFROC等）に基づく性能指標は、エンドポイントとして候補となりうる

試験対照群

- ・CADeシステムの臨床性能を対照となる手法と比較して評価することを推奨
- ・従来の臨床解釈手法（例：CADeシステムを使用しない読影）を試験対照群として設定する場合、CADe性能評価において最も関連性の高いものを選択

読影試験とランダム化

- ・臨床試験における読影シナリオ（例：逐次読影デザイン、クロスオーバーデザイン）は、CADeシステムの意図した利用形態と一致するように選択する

読影試験の評価尺度

- ・ROCベースのエンドポイントでは、読影者が患者の病変の状態を評価する際に、評価尺度（例：7点又は100点スケール）を用いてデータを収集する
- ・評価尺度の使用方法について読影者にトレーニングを行うことを推奨する

スコアリング

- ・読影者の解釈と参照基準との関連付けを決定する手順のこと
- ・スコアリングの手順と定義は、CADeシステムの臨床試験において重要な要素であり、説明することを推奨する

トレーニング

- ・臨床試験に参加する読影者に対して、CADe システムの使用方法および臨床試験への参加方法の詳細について、指示を明記し、トレーニングを行うことを推奨する

研究対象集団

- ・症例データは、明確に定義されたInclusion・Exclusion基準に基づき、前向きまたは後向きに収集
- ・症例収集の手順をプロトコルに記載することを推奨する

参照基準 Reference Standard

- ・参照標準（Gold Standard, Ground Truth とも呼ばれる）は、Truthing processにより決定される、疾患/異常が存在するかどうか、及びその範囲や場所などを示す情報
- ・参照基準を決定するために利用される方法（例：病理診断など）を記述することを推奨する



エンドポイントの選択

CADeシステムの有効性を示すために、エンドポイントを選択する必要があります。エンドポイント（主要評価項目と副次評価項目）の選択は、CADeシステムの使用用途に基づく必要があり、評価を開始する前に固定しなければなりません。Receiver operative characteristic (ROC) やその派生形（LROCやFROC等）、または感度と特異度がエンドポイントの候補になります（表3、図2）。

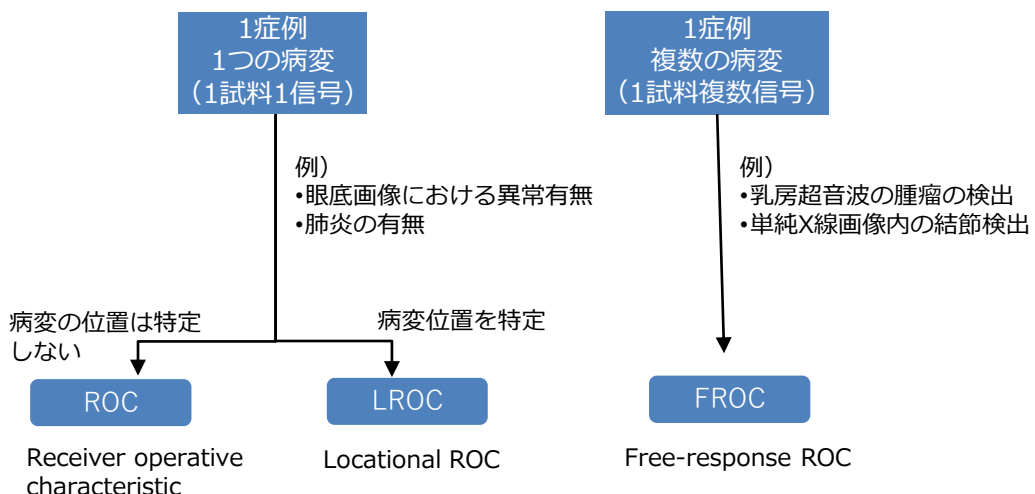
ROC、感度、および特異度のデータ収集は、1回の読影試験で同時に行うことができます。例えば、読影者は感度/特異度評価のための二値（陽性/陰性）判定を先に実施し、その後すぐにROC評価のための評価（二値判定の確信度など）を実施する方法があります。

FDAガイダンスでは、**主要評価項目としてROCを選択することを推奨**していますが、他のエンドポイントを選択することが適切である場合もあるとしています。主要評価項目としてROCを選択した場合は、上述のような方法で感度と特異度を副次評価項目として評価することを推奨しています。また、病変タイプや撮像条件等の交絡因子による層別解析結果も併せて報告することを推奨しています。

表3 CADeのエンドポイント例

指標	説明
感度 Sensitivity	参照基準が陽性であった場合の診断の結果が陽性（真陽性）である割合
特異度 Specificity	参照基準が陰性であった場合の診断の結果が陰性（真陰性）である割合
ROC Receiver Operating Characteristic	縦軸を真陽性率（感度）、横軸を偽陽性率（1 - 特異度）として閾値が取りうるすべての値においてプロットして得られた曲線

図2 ROC解析の種類
 (FDAガイダンスからの抜粋ではありません)





読影試験とランダム化

臨床試験における読影シナリオは、CADeシステムの意図した利用形態と一致するように選択します。表4は、利用形態ごとの読影シナリオの一例となります。Concurrent Reader型のような同じ画像を複数回読影するシナリオで、1回目の読影結果を覚えていることにより生じるバイアス (memory bias) を避ける必要がある場合、読影セッション間は少なくとも**4週間以上**隔てるようにすることが推奨されています。

また、性能評価のバイアスを減らすために、読影者、症例、読影順序をランダム化 (Latin Square Designなど) する必要があります。ランダム化の方法の説明、及び関連するフローチャートを示すことが推奨されています。

表4 読影シナリオの例

利用形態	デザイン	フロー
Second Reader: 読影者がCADなしで読影後、CADの結果を参照して再度読影	逐次読影 Sequential	<pre> graph LR Case[Case] --> withoutCAD[without CAD] withoutCAD --> withCAD[with CAD] </pre>
Concurrent Reader: 読影者がCADの結果を参照しながら読影	クロスオーバー Crossover	<pre> graph LR Case[Case] --> S1[Session 1] Case --> S2[Session 2] S1 --> W1[without CAD] W1 -- "≥4 weeks" --> S2 S2 --> W2[with CAD] </pre>

スコアリング

FDAガイダンスでは、読影者の解釈と参照基準との関連付けを決定する手順をスコアリングと呼んでいます。スコアリングの手順と定義は、CAdEシステムの臨床試験の重要な要素であり、説明することが推奨されています。スコアリング手順は、評価を開始する前に確定し、記述する必要があります。スコアリングプロセスに関する記述には、以下の情報を含めることが推奨されています。

- 手段は電子または非電子のどちらか
- 参照標準の境界、面積、体積に対する読影者評価による境界、面積、体積の重なり
- 参照標準の境界または空間的位置に対する読影者評価による重心の関係性
- 参照標準の重心に対する読影者評価による境界または空間的位置との関係性
- レビュアーによる決裁
- その他

レビュアーによる決裁を選択した場合、レビューに参加した読影者の数、資格、経験と専門性、事前の指示の内容、スコアリングに使用した特定の基準などを説明することが推奨されます。複数のレビュアーがスコアリングに関与する場合、評価の統合プロセスや、不一致への対処方法なども説明する必要があります。

トレーニング

FDAガイダンスでは、試験に参加する読影者に対して、以下の内容についてトレーニングを行うことが推奨されています。

- CAdEシステムの詳細や使い方の説明
- 臨床試験への参加方法に関する指示の説明
- 試験特有の読影指示またはルールに関する説明
- 評価尺度に関するトレーニング
- 代表的な正常/異常例に対する評価の実践

特殊な読影指示やルール（例：with CADで読影する際に、CADなしでの最初の解釈を変更するルール）については、読影作業、臨床ワークフロー、医療行為の観点から臨床的妥当性を示すことが推奨されています。なお、トレーニング時に使用する画像は試験本番では使用しない画像が推奨されます。

研究対象集団

症例データは、明確に定義されたInclusion・Exclusion基準に基づき、前向きまたは後向きに収集します。症例収集の手順をプロトコルに記載することが推奨されます。臨床試験用に収集した症例データは、システム開発時に使用した症例データとは独立している必要があります。症例収集の方法の例として、Inclusion基準内とExclusion基準外にある症例データを各参加施設から連続収集することが挙げられます。

症例収集時に留意すべき事項を以下に示します。

- CADeシステムの意図する使用対象集団から研究対象集団をサンプリングすること
- 疾患/異常例と正常例を適切な割合で含めること
- サンプルサイズは有意性を示すのに十分な検出力を持つように十分大きくすること
- 必要な場合は、臨床的に重要な因子を考慮したグループ（例えば、臨床的に重要な交絡因子、効果修飾因子、合併症）から十分な数の症例を含め、これらのサブグループごとに臨床性能の推定値と信頼区間を得られるようにすること
- サブグループごとに複数の性能の主張をする場合は、サブグループごとの試験のために統計的調整が必要となる

疾患/異常例を充実 (Enrichment) させることで効率的に症例データを集めることができますが、Enrichmentの程度は試験デザインにバイアスが発生する可能性を考慮して決める必要があります。また、研究対象集団を、臨床医にとって読影が困難な画像所見でEnrichmentすることも可能です。これは「ストレステスト」と呼ばれ、試験デザインにバイアスが発生することになるため、臨床試験にストレステストを適用する際は注意が必要です。

適切なストレステストの研究対象集団の例：

結腸ポリープを検出するCADeシステムで、幅広いポリープサイズを検出することを目的としている場合、小さなポリープサイズの症例でEnrichmentすることは適切と考えられる

なおFDAガイダンスには、CADeシステムの臨床試験においてデータの再利用を検討する場合についての注意事項も記載されており、テストデータの一部を再利用しても性能の推定に不合理なバイアスが生じないこと、およびテストデータの整合性が維持されることを説明する必要があります。



研究対象集団

研究対象集団を説明する際、必要に応じて以下のような情報を記述することが推奨されています。

表5 研究対象集団の記述事項

	情報	例
症例特性	人口統計データ	年齢、民族性、人種、性別
	CADe システムの適用に関連する病歴	
	患者の疾病状態及び放射線検査の適応症	
画像特性	撮像条件	造影剤を使用したか否か、造影剤の種類と患者あたりの投与量、患者の肥満度、放射線量、MRI 画像の T-weighting および撮像視野など
	画像データの詳細	撮像装置のメーカーとモデル、およびスキャンプロトコル
	収集施設	
	該当する場合、データ加工した施設	患者データのデジタル化
症例数	<ul style="list-style-type: none"> ・疾患/異常症例数 ・正常症例数 ・病気の状態、場所、範囲（参照基準）を決定するために使用される方法 	
交絡因子	以下のような交絡因子または効果修飾因子で層別した疾患群の分布 <ul style="list-style-type: none"> ・病変のタイプ（例：過形成性大腸ポリープと腺腫性大腸ポリープ） ・病変のサイズ、位置 ・病期、併発疾患 ・臓器の特徴（例；乳房の構成） ・撮像機器（例：メーカーとモデル、スキャンプロトコル） ・収集施設 ・該当する場合、データ加工した施設 	
比較	収集した症例データの臨床的、画像的、病理学的特徴を目的母集団と比較する	

おわりに

CADシステムの診断精度を適切な方法で評価するためには、適切な臨床試験の確立が重要であり、それが承認申請データとして利用されるならば、より一層の合理的で妥当な試験デザイン構築が望まれます。

本稿はCADEシステムの臨床試験を実施する際の留意事項が取りまとめられたFDAガイダンスを紹介いたしました。本邦において、CADシステムを開発する上での全般的な留意点についてはバックナンバー「[医用画像診断支援システムの臨床試験～医療AIの社会実装に向けて～](#)」にて解説しています。また、本稿で取り上げなかった「参照基準」の作成プロセスについてはバックナンバー「[医用画像診断支援システムの臨床評価：参照基準の重要性～医療AIの社会実装に向けて～](#)」にて解説していますので是非そちらも併せてご参照ください。

マイクロンは、国内はもとよりアジア圏のイメージングCROのリーディングカンパニーとして、画像を用いた臨床試験を多くサポートしています。本稿で紹介したようなSaMDのサポート実績も多数あり([プログラム医療機器 \(SaMD\) 承認取得支援実績のお知らせ](#)) 本ガイドラインに沿った臨床試験を提案しております。AI搭載CADや医用画像診断支援システムの認証・承認に関するコンサルテーションおよびご相談に関するお問い合わせは、下記連絡先までお問い合わせください。

会社概要

東京本社 : 〒108-0073 東京都港区三田三丁目13番16号 三田43MTビル9階
TEL.03-6631-3693

大阪支社 : 〒532-0003 大阪市淀川区宮原四丁目5-36 ONEST新大阪スクエア6階
TEL.06-6399-0007

名古屋事業所 : 〒474-8511 愛知県大府市森岡町七丁目430番地
国立研究開発法人 国立長寿医療研究センター 南病棟2階
TEL.0562-46-2105

事業内容 ①イメージング技法を活用した医薬品、診断薬、医療機器、バイオマーカーの開発支援
②臨床開発支援
(モニタリング、品質管理、イメージング・コアラボ業務、画像解析、読影支援等)
③PET薬剤の治験薬GMP製造支援
④臨床開発に係るコンサルティング

ホームページ : <https://micron-kobe.com>

LinkedIn : <https://www.linkedin.com/company/micron-imaging/>

お問合せEmail : imagingbiomarker@micron-kobe.com