

*Contains Nonbinding Recommendations  
Draft – Not for Implementation*

人工知能対応医療機器ソフトウェア機能：ライフサイクル管理および販売許可申請に関する推奨ガイダンス  
(ドラフト) 2025 年 1 月 7 日

**Artificial Intelligence-Enabled Device Software Functions: Lifecycle  
Management and Marketing Submission Recommendations**

**DRAFT GUIDANCE**

**This draft guidance document is being distributed for comment purposes  
only.**

**Document issued on January 7, 2025.**

このドキュメント案に関するコメントと提案は、ガイダンス案の利用可能性を公表する通知が連邦官報に掲載されてから 90 日以内に提出する必要があります。電子コメントを <https://www.regulations.gov> に送信します。書面によるコメントを、Dockets Management Staff, Food and Drug Administration, 5630 Fishers Lane, Room 1061, (HFA-305), Rockville, MD 20852-1740 に提出してください。連邦官報に掲載された利用可能性の通知に記載されているすべてのコメントを識別します。

CDRH 規制デバイスに関するこのドキュメントに関する質問については、[digitalhealth@fda.hhs.gov](mailto:digitalhealth@fda.hhs.gov) の Digital Health Center of Excellence にお問い合わせください。CBER 規制デバイスに関するこのドキュメントに関する質問については、1-800-835-4709 または 240-402-8010 の Office of Communication, Outreach, and Development (OCOD)、または [ocod@fda.hhs.gov](mailto:ocod@fda.hhs.gov) に電子メールでお問い合わせください。CDER 規制製品に関するこのドキュメントに関する質問については、[druginfo@fda.hhs.gov](mailto:druginfo@fda.hhs.gov) にお問い合わせください。コンビネーション製品に関するこのドキュメントに関する質問については、[combination@fda.gov](mailto:combination@fda.gov) の Office of Combination Products にお問い合わせください。



米国保健福祉省  
米国食品医薬品局  
医療機器・放射線保健センター  
生物学的製剤評価研究センター  
医薬品評価研究センター  
局長室配合剤室

**Note)** This document is for purpose of providing reference information for the use of the original document, and does not have the same effect as the original document.  
If you have any doubts about this document, please refer to the original Regulatory Authority (FDA) information.  
HiroPharmaConsulting assumes no responsibility for any inconvenience caused by the use of this document.  
Only valid the "The FDA published the Draft Guidance: Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations. January 6, 2025".

## はじめに

### その他の写し

#### **CDRH**

追加のコピーはインターネットから入手できます。[CDRH-Guidance@fda.hhs.gov](mailto:CDRH-Guidance@fda.hhs.gov) に電子メールでリクエストを送信して、ガイダンスのコピーを受け取ることもできます。リクエストには、ドキュメント番号 GUI00007028 とガイダンスの完全なタイトルを含めてください。

#### **CBER**

追加のコピーは、生物学的製剤評価研究センター (CBER)、Office of Communication, Outreach, and Development (OCOD) (10903 New Hampshire Ave., Bldg.71, Room 3128, Silver Spring, MD 20993-0002)、または電話 1-800-835-4709 または 240-402-8010、電子メール [ocod@fda.hhs.gov](mailto:ocod@fda.hhs.gov)、またはインターネット <https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidances> で入手できます。

#### **CDER**

追加のコピーは、医薬品評価研究センター (CDER)、Office of Communication, Division of Drug Information (10001 New Hampshire Ave., Hillandale Bldg.,4th Floor, Silver Spring, MD 20993-0002)、または電話 855-543-3784 または 301796-3400、電子メール [druginfo@fda.hhs.gov](mailto:druginfo@fda.hhs.gov)、またはインターネット <https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs> で入手できます。

## 目次

I. はじめに .....	4
II. 適用範囲 .....	5
III. TPLC アプローチ: 一般原則.....	7
IV. このガイダンスの使用法:AI 対応デバイスマーケティングの提出コンテンツの推奨事項の概要 .....	8
A. Quality System ドキュメント.....	10
V. デバイスの説明 .....	11
VI. ユーザーインターフェースとラベル付け .....	13
A. ユーザーインターフェース.....	15
B. ラベリング.....	16
VII. リスク評価 .....	20
VIII. データ管理.....	22
IX. モデルの説明と開発 .....	28
X. 検証 (Validation) .....	30
A. パフォーマンスの検証.....	31
XI. デバイスのパフォーマンスモニタリング .....	36
XII. サイバーセキュリティ .....	38
XIII. 公開申請の概要.....	41
付録 A: 推奨されるドキュメントの表 .....	44
付録 B: Transparency_Design の考慮事項.....	45
付録 C: パフォーマンス検証に関する考慮事項 .....	48
付録 D: ユーザビリティ評価の考慮事項 .....	52
付録 E: モデルカードの例.....	54
付録 F: 例 510 (k) の概要とモデルカード .....	57

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

人工知能対応医療機器ソフトウェア機能：ライフサイクル管理および販売許可申請に関する推奨ガイダンス  
(ドラフト) 2025年1月7日

業界および食品医薬品局のスタッフ向けのガイダンス案

このガイダンス案は、最終決定された時点で、このトピックに関する食品医薬品局 (FDA) の現在の考えを表しています。本ガイドラインは、いかなる個人の権利も確立するものではなく、FDA または一般の人々を拘束するものではありません。適用される法令の要件を満たす場合は、代替アプローチを使用できます。代替アプローチについて検討するには、タイトルページに記載されているこのガイダンスを担当する FDA スタッフまたは Office にお問い合わせください。

## I. はじめに

FDA は長い間、人工知能 (AI) 対応機器を含む医療機器の監視に対する、トータル製品ライフサイクル (TPLC) アプローチを推進しており、そのようなアプローチのためのガイダンスとリソースの開発に取り組んできました。最近の取り組みとしては、安全で効果的かつ高品質な機械学習モデルの促進を支援するために、優れた機械学習の実践 (GMLP)<sup>1</sup>と機械学習対応機器の透明性<sup>2</sup>のための指針原則の開発が挙げられます。また、ユーザーのための機器の透明性に関する議論を含む、AI 対応機器に対する患者中心のアプローチを促進するための公開ワークショップも開催されました<sup>3</sup>。本ガイダンスでは、AI 対応機器の TPLC アプローチと整合性のあるライフサイクル管理と販売申請の推奨事項を提供することで、これらの取り組みを継続することを目的としています。

本ガイダンスでは、FDA の審査を支援する文書や情報を含む、AI 対応機器のソフトウェア機能を含む機器の販売申請の内容に関する推奨事項を提供します。FDA による機器の評価のための適切な文書の作成を支援するために、本ガイダンスでは、製造業者が TPLC 全体で使用を検討できる AI 対応機器の設計と開発に関する推奨事項も提供します。

<sup>1</sup> See FDA's website on [Good Machine Learning Practice for Medical Device Development: Guiding Principles](#).

<sup>2</sup> See FDA's website on [Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles](#).

<sup>3</sup> See FDA's website on [Artificial Intelligence and Machine Learning \(AI/ML\) Software as a Medical Device Action Plan](#), the Executive Summary for the "[Patient Engagement Advisory Committee Meeting on Artificial Intelligence \(AI\) and Machine Learning \(ML\) in Medical Devices](#)," and the website on the [Virtual Public Workshop - Transparency of Artificial Intelligence/Machine Learning-enabled Medical Devices](#).

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

この推奨事項は、TPLC 全体にわたる AI 対応機器のライフサイクル管理に対する包括的なアプローチを反映しています。さらに、このガイダンスには、AI 対応機器の TPLC 全体にわたる透明性とバイアスに対処するための戦略に関する FDA の現在の考え方が含まれています。これには、機器がすべての関連する人口統計学的グループ(人種、民族、性別、年齢など)に同様に利益をもたらすかどうかを評価するための証拠を収集することも含まれており、これらの機器が意図された用途に対して安全かつ有効であることを保証するのに役立ちます。

ソフトウェアに関連するコンセンサス基準の出現は、特にリスク評価や管理などの活動に関して、ソフトウェア開発と文書化の一貫性と品質を向上させるのに役立っています。可能な場合、FDA はこのガイダンスの用語と推奨事項をソフトウェア関連のコンセンサス基準と調和させました。FDA は、AI 対応機器を開発し、市販前の文書を準備する際に、このような FDA 承認のコンセンサス基準を考慮することを奨励します。この文書で参照されている FDA 承認のコンセンサス基準の最新版については、[FDA 承認のコンセンサス基準データベース](#)を参照してください。承認された基準に適合宣言を提出する場合は、適切な補足文書を含めることをお勧めします。規制当局への提出におけるコンセンサス基準の使用に関する詳細については、「[医療機器の市販前の提出における自主的なコンセンサス基準の適切な使用](#)」および「[生物学的製剤評価研究センターで検討された規制当局への提出における基準の開発と使用](#)」と題する FDA ガイダンスを参照してください。一般に、FDA のガイダンス文書は法的強制力のある責任を確立するものではありません。

代わりに、ガイダンスはあるトピックに関する FDA の現在の考え方を記述したものであり、特定の規制要件または法定要件が引用されていない限り、推奨事項としてのみ見るべきです。FDA のガイダンスにおける should という言葉の使用は、何かが提案または推奨されているが、必須ではないことを意味します。

## II. 適用範囲

本ガイダンスでは、FDA はデバイスの定義を満たすソフトウェア機能を「デバイスソフトウェア機能」と呼びます。「デバイスソフトウェア機能」とは、連邦食品医薬品化粧品法 (FD&C 法) のセクション 201 (h) のデバイス定義を満たすソフトウェア機能です。<sup>4</sup> FDA の他のガイダンスで説明されているように、「機能」という用語は製品の明確な目的であり、製品の意図された使用または意図された使用のサブセットである可能性があります。<sup>5</sup>

AI 対応デバイスとは、1 つ以上の AI 対応デバイスソフトウェア機能 (AI-DSF) を含むデバイスです。AI-DSF は、意図された目的を達成するために 1 つ以上の「AI モデル」(本ガイダンスでは「モデル」と呼びます) を実装するデバイスソフトウェア機能です。モデルとは、新しい入力データに基づいて推論または予測を生成する数学的構成要素です。本ガイダンスでは、「AI 対応デバイス」を使用する場合はデバイス全体を指し、「AI-DSF」を使用する場合は AI を使用する機能のみを指します。本ガイダンスでは、「モデル」を使用する場合は数学的構成要素のみを指します。

<sup>4</sup> Device software functions may include Software as a Medical Device (SaMD) and Software in a Medical Device (SiMD). See FDA's website on [Software as a Medical Device \(SaMD\)](#).

<sup>5</sup> See FDA's guidance titled "[Multiple Function Device Products: Policy and Considerations](#)."

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

AI 対応デバイスの開発を引き続きサポートするために、このガイダンスでは、AI-DSF を含むデバイスの FDA の審査をサポートするために、市販申請に含める必要がある文書と情報に関する推奨事項を提供します。このガイダンスでは、「市販申請」という用語は、市販前通知 (510 (k) ) の申請、De Novo 分類要求、市販前承認 (PMA) 申請、人道的デバイス適用除外 (HDE)、または生物学的製剤ライセンス申請 (BLA) を指します。<sup>6</sup>このガイダンスで提案されている推奨事項の一部は、治験用デバイス適用除外 (IDE) 申請にも適用される場合があります。510 (k) 要件の対象となる AI 対応デバイスの場合、AI 対応デバイスは、特に、AI 対応デバイスが非 AI 対応デバイスと比較して安全性および有効性に関する異なる問題を導入しておらず、FD&C 法のセクション 513 (i) に従って実質的同等性を決定するための他の要件を満たしている場合に、同じ使用目的で AI 対応デバイスと実質的に同等であると判断できます。

一般に、本ガイダンスの推奨事項は、デバイス構成部品 <sup>7</sup>に AI-DSF が含まれている場合には、配合剤のデバイス構成部品 <sup>8</sup>にも適用されます。AI-DSF を開発する際、治験依頼者は、AI-DSF が配合剤全体に及ぼす影響を考慮する必要があります。AI-DSF を含む配合剤については、FDA の主審査部門に早期に関与することを強く推奨します。<sup>9</sup> FDA の主審査部門は、センター間協議プロセスに従い、適切な主題の専門家と協議します。<sup>10</sup> FDA は、配合剤についての治験の検討事項および販売申請の他の側面に関する推奨事項について、治験依頼者が他のガイダンスを参照することを推奨します。<sup>11</sup>

本ガイダンスで提案されている推奨事項は、さまざまな AI 対応デバイスの審査における FDA の経験と、現在の規制科学研究に基づいています。提案されている推奨事項は、AI 対応デバイスに広く適用されることを意図していますが、これらの推奨事項の多くは、機械学習として知られる AI のサブセット、特にディープラーニングとニューラルネットワークを組み込んだデバイスに特に関連する可能性があります。その他の形態の AI には、追加の考慮事項が適用される場合があります。場合によっては、本ガイダンスでは、メーカーがこれらの推奨事項を AI 対応デバイスに適用するのを支援するために、他のガイダンスの推奨事項を強調しています。本ガイダンスに特定の推奨事項が含まれていても、含まれていない可能性のある他のガイダンスの適用可能な推奨事項が無効になるわけではありません。本ガイダンスは、FD&C 法、その施行規則、およびその他のガイダンスドキュメントとの関連で考慮する必要があります。

<sup>6</sup> Certain devices are subject to review through a BLA under section 351 of the Public Health Service Act.

<sup>7</sup> See 21 CFR 4.2.

<sup>8</sup> See 21 CFR 3.2(e).

<sup>9</sup> See FDA's guidance titled "[Principles of Premarket Pathways for Combination Products.](#)"

<sup>10</sup> See FDA's Staff Manual Guide titled "[Combination Products: Inter-Center Consult Request Process.](#)"

<sup>11</sup> See FDA websites titled "[Combination Products Guidance Documents](#)" and "[Search for FDA Guidance Documents.](#)" See also FDA's website on [Combination Products](#) for additional policy information regarding combination products.

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

本ガイダンスは、AI 対応デバイスの販売申請に含める必要がある内容の完全な説明を提供することを意図したものではありません。特に、本ガイダンスでは、AI 対応デバイスに関する重要な追加の考慮事項を含む FDA ガイダンス [“Content of Premarket Submissions for Device Software Functions”](#) (以下、[“Premarket Software Guidance”](#)) のセクションを参照していますが、そのガイダンスのすべてのセクションへの参照は含まれていません。さらに、本ガイダンスでは、AI 対応デバイスの特定の適応症を裏付けるために提出するデータおよび情報のすべてを扱っているわけではありません。FDA は、治験依頼者が特定のデバイスに適用される他のガイダンスも参照することを推奨しています。市販申請のその他の側面に関する推奨事項。<sup>12</sup> 特定の技術に関連するガイダンスの例としては、[「放射線機器の市販前申請における定量的イメージングの技術的性能評価」](#)および[「生理的閉ループ制御技術を備えた医療機器の技術的考慮事項」](#)と題する FDA ガイダンスがあります。FDA はさらに、治験依頼者が販売申請を準備する際に、コンセンサス基準や一般に公開されている情報など、他の利用可能なリソースを考慮することを奨励しています。すべてのデバイスと同様に、FDA は、AI 対応デバイスの販売申請をサポートするために、特定のテストと適用可能な推奨事項を決定するためにリスクベースのアプローチを採用する予定です。

FDA との早期の関与は、製品開発と申請準備の指針となる可能性があります。特に、デバイスの開発または設計に新しい新興技術が使用されている場合、またはデバイスの検証中に新しい方法が使用されている場合に、早期の関与が役立つ可能性があります。FDA は、治験依頼者が Q 提出プログラム (Q-Submission Program) を介して FDA とこれらの計画について話し合うことを検討するよう奨励しています。<sup>13</sup>

### **III. TPLC アプローチ: 一般原則**

このガイダンスでは、AI 対応デバイスの管理に対する TPLC アプローチの重要性を認識しています。TPLC 全体のリスク管理に対する包括的なアプローチを反映した、販売申請に含めるべき文書と情報に関する推奨事項に加えて、このガイダンスで提供されるリソースは、AI 対応デバイスのデバイス開発とライフサイクル管理を支援することも目的としており、これらのデバイスの安全性と有効性をサポートする必要があります。このガイダンスでは、AI 対応デバイスの販売申請をサポートするための情報と文書に関する具体的な推奨事項と、パフォーマンス管理を含む AI 対応デバイスの設計、開発、展開、および保守に関する推奨事項の両方を提供しています。<sup>14</sup>

<sup>12</sup> For other guidances with digital health content, please see FDA’s website on [Guidances with Digital Health Content](#). For all other guidances, please see FDA’s website on [Guidance Documents \(Medical Devices and Radiation-Emitting Products\)](#).

<sup>13</sup> See FDA’s guidance titled, [“Requests for Feedback and Meetings for Medical Device Submissions: The Q Submission Program”](#) (hereafter referred to as the “Q-Submission Program”).

<sup>14</sup> This guidance is not intended to provide recommendations on reporting to FDA when a device has or may have caused or contributed to a death or serious injury as required by section 519 of the FD&C Act, the Medical Device Reporting (MDR) Regulation in 21 CFR Part 803, or the Medical Device Reports of Corrections and Removals Regulation in 21 CFR Part 806. For an explanation of the current reporting and recordkeeping requirements

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

このガイダンスには、AI 対応デバイスの TPLC 全体で透明性とバイアスに対処する戦略に関する FDA の現在の考え方も含まれています。これらの相互に関連する考慮事項は、TPLC 全体で重要であり、デバイスの透明性とバイアスの制御を設計し、その安全性と有効性を確保するために、デバイス設計の最も早い段階から廃止まで組み込む必要があります。透明性とは、重要な情報がアクセス可能であり、機能的に理解可能であり、情報の共有とデバイスのユーザビリティの両方に関連していることを保証することです。AI バイアスとは、体系的に、しかし時には予測できない方法で誤った結果を生成する潜在的な傾向であり、意図された使用集団(例えば、異なる医療環境、異なる入力装置、性別、年齢等)のすべてまたはサブセット内でデバイスの安全性と有効性に影響を与える可能性があります。透明性とバイアスに対する包括的なアプローチは、AI 対応デバイスで特に重要です。これは、多くのモデルが不透明であり、生物学的に妥当な作用メカニズムに直接マップされない可能性のあるデータ相関にモデルが依存しているため、ユーザーが理解するのが難しい場合があります。透明性に対する設計アプローチの推奨事項については、付録 B (透明性設計の考慮事項) を参照してください。AI 対応デバイスのバイアスの制御に関しては、製品ライフサイクル全体にわたる開発、テスト、および監視のためのデータ収集における代表性への対処、および意図された使用のサブグループ間でのパフォーマンスの評価を含めることができます。

最後に、このガイダンスには、市販後の設定を含む TPLC 全体で AI 対応デバイスのパフォーマンスに対処する推奨事項が含まれています。たとえば、AI 対応デバイスは、実際の導入時の入力データと比較して開発時に使用される入力データなど、入力データの違い (データドリフトとも呼ばれます) に敏感になる可能性があります。さらに、AI 対応デバイスの入力を生成するシステムが、デバイスのパフォーマンスに影響を与える可能性があるがユーザーにはわからない方法で時間の経過と共に変化した場合に発生するデータドリフトに加えて、AI 対応デバイスは他の要因によるパフォーマンスの変化の影響を受ける可能性もあります。また、FDA ガイダンス [“Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions”](#) で説明されています、事前に定められた変更管理計画 (PCCP) の使用を検討することも推奨されます。このガイダンスでは、PCCP に準拠した変更を実施する前に、製造業者が AI-DSF (例えば、デバイスの性能を向上させるため) に対する意図された変更について、追加のマーケティング資料を提出したり、さらに FDA の承認を取得したりすることなく、事前に特定し、市販前の承認を求めるアプローチについて説明しています。

#### **IV. このガイダンスの使用法:AI 対応デバイスマーケティングの提出コンテンツの推奨事項の概要**

このガイダンスでは、AI-DSF を含むデバイスの FDA の審査をサポートするために、マーケティングの提出に含める必要がある文書と情報に関する推奨事項を提供します。

---

applicable to manufacturers of medical devices, please refer to FDA’s guidance titled [“Medical Device Reporting for Manufacturers.”](#)



***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

FDAとAIコミュニティがAI対応デバイスTPLCと特定の用語を検討する方法には、いくつかの違いがあります。したがって、本ガイダンスでは、本ガイダンスの推奨事項の理解を深めるために、これらの違いを明確にしています。たとえば、AIコミュニティでは、モデルの選択を最適化するためにモデルのトレーニングフェーズと組み合わせることができるデータキュレーションまたはモデルチューニングを指すために、「検証（Validation）」という用語を使用することがよくあります。<sup>15</sup> ただし、検証は21 CFR 820.3 (z) <sup>16</sup>で「...特定の意図された用途の特定の要件を一貫して満たすことができるという客観的な証拠の調査と提供による確認」と定義されています。本ガイダンスでは、21 CFR 820.3 (z) の定義を使用しています。特に、意図された用途のモデルのパフォーマンスの評価に対処する場合に使用します。明確にするために、トレーニングおよびチューニングプロセスを指すために「検証」という用語を使用することは、医療機器の販売申請のコンテキストでは避ける必要があります。また、「開発（Development）」という用語は、トレーニング、チューニング、およびチューニング評価（AIコミュニティでは「内部テスト」と呼ばれることが多い）を指すために本ガイダンス全体で使用されています。本ガイダンスでは、「テストデータ」は、テストプロセスとも呼ばれる検証および検証活動に使用される可能性のあるデータを指すために使用されており、開発プロセスの一部を説明するためには使用されていません。“[FDA Digital Health and Artificial Intelligence Glossary–Educational Resource](#)”には、人工知能および機械学習の分野で一般的に使用される用語とその定義がまとめられています。

本ガイダンスのセクション V から XIII では、AI対応デバイスに関する販売承認提出コンテンツの推奨事項について説明します。具体的には、各セクションの「AI対応デバイスの提出書類に含める必要がある理由」の下に、特定の情報を販売申請提出書類に含める必要がある理由が説明されています。販売申請提出書類に含める必要がある文書と情報の説明については、「治験依頼者が提出書類に含める必要があるもの」を参照してください。最後に、治験依頼者が販売申請提出書類の各セクション内の情報を含める必要がある場所に関する推奨事項については、「治験依頼者が提出書類に提供する必要のある場所」を参照してください。ライフサイクルの考慮事項に関する推奨事項および販売申請提出書類の例に関する情報は、本ガイダンスの付録に記載されています。

販売申請提出書類に関連する推奨事項は、提出書類に記載する方法に従って編成されています（付録 A（推奨される文書の表）を参照）。これは、必ずしもTPLCの活動の順序と直接一致するわけではありません。参照されている提出書類のセクションはすべて、市販前審査中にFDAに提供されますが、デバイスの開発と検証のために既に行われたこと、およびデバイスの継続的な安全性と有効性を確保するために治験依頼者が将来行う予定のことに関する情報が含まれています。ガイダンスの一部のセクションでは、TPLCの複数のステップに関連する情報についても説明しています。

<sup>15</sup> See International Medical Device Regulators Forum Technical Document N67 titled “[Machine Learning-enabled Medical Devices: Key Terms and Definitions.](#)”

<sup>16</sup> On February 2, 2024, FDA issued a final rule amending the device Quality System Regulation (QSR), 21 CFR Part 820, to align more closely with international consensus standards for devices (89 FR 7496). This final rule will take effect on February 2, 2026. Once in effect, this rule will withdraw the majority of the current requirements in Part 820 and instead incorporate by reference the 2016 edition of the International Organization for Standardization (ISO) 13485, Medical devices – Quality management systems – Requirements for regulatory purposes, in Part 820. As stated in the final rule, the requirements in ISO 13485 are, when taken in totality, substantially similar to the requirements of the current Part 820, providing a similar level of assurance in a firm’s quality management system and ability to consistently manufacture devices that are safe and effective and otherwise in compliance with the FD&C Act. When the final rule takes effect, FDA will also update the references to provisions in 21 CFR Part 820 in this guidance to be consistent with that rule.

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

本ガイダンスのセクションが TPLC とどのように整合するかの一例を以下に示します。

- **開発 (Development)** – リスク評価、データ管理、およびモデルの説明と開発
- **検証 (Validation)** – データ管理と検証
- **最終的なデバイスの説明** – デバイスの説明、モデルの説明と開発、ユーザインターフェースとラベリング、公開申請の概要
- **市販後管理** – デバイスのパフォーマンスモニタリングとサイバーセキュリティ

このガイダンスでは、一般に、ソフトウェアの開発、検証、および検証中に生成され、文書化される情報について説明します。ただし、市場承認をサポートするために必要な情報は、各 AI 対応デバイスの詳細に基づいて異なります。また、FDA は市販前審査中に、申請を評価するために必要な追加情報を要求することがあります。

## A. Quality System ドキュメント

本ガイダンスの第 V 章から第 XIII 章の推奨事項を検討する際には、「治験依頼者が申請書に含めるべき内容」に記載されている、販売申請書に含めるべき文書および情報が、Quality System ドキュメントに存在する可能性があるかどうかを検討することが役立つ場合があります。特定の AI 対応デバイスの販売申請において、実質的な同等性または安全性と有効性の合理的な保証を示す一部として使用できる文書のソースの 1 つは、品質システム (QS) 規則の継続的な要件に関連する文書です。<sup>17</sup> 本ガイダンスでは、一般的に医療機器の QS 規制遵守に関連する可能性のあるいくつかの文書を市販前に提供して、治験依頼者又は製造業者が AI 対応機器に関連するリスクに具体的にどのように対処しているかを示す方法について説明します。

たとえば、QS 規則では、製造業者が特定の完成デバイスの設計管理を確立することが求められています (21 CFR 820.30 参照)。具体的には、設計管理の一部として、製造業者は「デバイスの設計を検証するための手順を確立し、維持する」必要があります。この手順には、「デバイスが定義されたユーザーのニーズと意図された用途に適合していることを確認し、実際またはシミュレートされた使用条件での生産ユニットのテストを含める」(21 CFR 820.30 (g)) が含まれます。さらに、21 CFR 820.30 (i) では、製造業者は、ソフトウェアで自動化されたものを含むすべてのデバイスについて、実装前の設計変更 (「設計変更」) を特定、文書化、検証、または適切な場合は検証、レビュー、承認するための手順を確立し、維持しなければなりません。

<sup>17</sup> In the postmarket context, design controls may also be important to ensure medical device performance and maintain medical device safety and effectiveness. FDA recommends that device manufacturers implement comprehensive performance risk management programs and documentation consistent with the QS Regulation, including but not limited to management responsibility (21 CFR 820.20), design validation (21 CFR 820.30(g)), design changes (21 CFR 820.30(i)), nonconforming product (21 CFR 820.90), and corrective and preventive action (21 CFR 820.100). While FDA generally does not assess QS Regulation compliance as part of its review of premarket submissions under section 510(k) of the FD&C Act, this guidance is intended to explain how FDA evaluates the performance of the device performance-related outputs of activities that are part and parcel of QS Regulation compliance, and explain how the QS Regulation can be leveraged to demonstrate these performance related outputs.

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

同様に、不適合製品の管理の一部として、製造業者は、「不適合製品の既存及び潜在的な原因。」を特定するために、状況によってはユーザー要件を含む「特定の要件に準拠していない製品を管理する」手順を確立し、維持し、「苦情」や「品質データの他のソース」を含む是正および予防措置を実施する必要があります (21 CFR 820.90 (a) および 820.100 (a) (1) )。さらに、製造業者は、品質目標が満たされていることを確実にするために、「定められた間隔で、定められた手順に従って十分な頻度で」品質システムの適切性と有効性をレビューすることを含め、品質システムを管理し、デバイスの品質を維持する継続的な責任を負っています。<sup>19</sup>

## V. デバイスの説明

AI 対応デバイスの申請に含める必要がある理由：次のセクションでは、FDA が AI 対応デバイスの一般的な特性を理解するのに役立つように、治験依頼者が販売申請のデバイスの説明セクションで提供する必要があります。次の推奨事項は、該当する場合に、[市販前ソフトウェアガイダンス](#)に記載されているデバイス固有の推奨事項と推奨事項を補足するものです。

デバイスの説明は、意図された用途、デバイスの予想される動作シーケンス(例えば、デバイスの臨床ワークフロー)、使用環境、モデルの特徴、および AI 対応デバイスの設計に関する FDA の理解をサポートします。この情報は、FDA がデバイスの安全性と有効性を評価するために必要です。デバイスの説明は、デバイスがどのように動作するか、ユーザーがどのように操作するか、どのような状況でデバイスが意図したとおりに使用される可能性があるかなど、デバイスの機能に関する重要なコンテキストを提供します。

モデルの技術的特性およびモデルの開発方法に関する情報を販売申請に含める方法に関する推奨事項については、本ガイダンスのセクション IX (モデルの説明と開発) を参照してください。

治験依頼者が申請に含める必要があるもの：一般に、治験依頼者は、AI 対応製品のデバイスの説明の一部として、次の種類の情報を含める必要があります。

- デバイスで AI が使用されていることの説明。
- デバイスの入力と出力の説明 (入力が手動で入力されるか自動で入力されるかを含む)、および互換性のある入力デバイスと取得プロトコルの一覧 (該当する場合)。
- デバイスの使用目的を達成するために AI がどのように使用されるかの説明。複数の機能を持つデバイスの場合、この説明には、AI-DSF が相互にどのように対話するか、および AI-DSF 以外の DSF とどのように対話するかを含めることができます。
- 対象となるユーザー、その特性、およびそのユーザーが受けることが予想されるトレーニングのレベルと種類の説明。ユーザーには、出力を解釈する人も含まれます。必要に応じて、出力を解釈するユーザーの資格や臨床的役割を記載します。ユーザーには、設置、使用、保守など、デバイスを操作するすべての人も含まれます。たとえば、ユーザーには、技術者、医療提供者、患者、介護者、管理者など、医療機器の導入方法や医療機器の臨床ケアへの適合性に関する決定に関与する人も含まれます。

<sup>18</sup> See 21 CFR 820.20.

<sup>19</sup> 21 CFR 820.20(c).

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

- 意図された使用環境の説明(例えば、臨床環境、家庭環境)。
- 以下を含む、デバイスを使用するための意図されたワークフローの説明(例えば、意図された意思決定の役割)。
  - 現在の標準治療のワークフローと比較して、装置が提供する自動化の程度の説明;
  - 使用につながる可能性のある臨床状況の説明;および
  - 出力が臨床ワークフローでどのように使用されるかの説明。
- インストールおよびメンテナンス手順の説明。
- パフォーマンスを維持するためにユーザーが定期的に行う必要がある調整および/または構成手順の説明。これには、調整を実行する必要がある場合や、必要に応じて調整が再度必要かどうかをユーザーが識別する方法が含まれます。

さらに、治験依頼者は、ユーザーが構成できる要素を持つ AI 対応デバイスのデバイス説明の一部として、次の種類の情報を含める必要があります。

- AI 対応デバイスのすべての構成可能な要素の説明。例:
  - ユーザーがオン/オフにできる視覚エフェクト(オーバーレイ、品質インジケータ、ヒートマップなど);
  - ソフトウェア入力;
  - 使用中に構成された場合のモデルパラメータ;および/または
  - アラートしきい値。
- 以下を含む、これらの要素とその設定方法の説明。
  - 必要に応じて、これらの決定を行うために必要な資格とトレーニングを含む、設定の決定を行うユーザの説明(例えば、臨床使用者、管理使用者、患者);
  - どの選択が行われたかをユーザーが知る方法の説明;
  - 患者、臨床現場、病院のネットワークレベルなど、設定が定義されているレベルの説明;および
  - 必要に応じて、カスタマイズ可能な定義済みの動作ポイント、その出力、およびパフォーマンス範囲の説明。また、デバイスの使用目的に基づいて、動作ポイントまたは動作ポイント範囲がどのように選択されたかを指定することも重要です。
- ユーザーの意思決定に対する構成可能な要素の潜在的な影響の説明。

最後に、デバイスに個別のインターフェイスを持つ複数の接続されたアプリケーションが含まれている場合、デバイスの説明では、デバイス内のすべてのアプリケーションに対応する必要があります。例えば、患者向けのアプリケーション、介護者向けのアプリケーション、および医療提供者向けのデータポータルがある場合、デバイスの説明には、アプリケーション全体にわたるすべての機能の詳細を含め、それらがどのように接続されているかを説明する必要があります。治験依頼者は、グラフィックス、図、イラスト、画面キャプチャ画像、または画面キャプチャビデオを含むビデオデモを使用して、デバイスの説明を強化することを検討することもできます。販売申請でユーザインターフェースの要素を共有する方法の詳細については、セクション VI.A (ユーザインターフェース) を参照してください。

治験依頼者が申請で提供する箇所: AI 対応デバイスの説明情報は、販売申請の「**デバイスの説明**」セクションに含める必要があります。

## VI. ユーザインターフェースとラベル付け

ユーザインターフェースには、ユーザーが操作するデバイスのすべての要素(例えば、ユーザが見たり、聞いたり、触ったりする装置の部分)を含む、ユーザーとデバイス間のすべての対話ポイントが含まれます。また、必要に応じて、デバイスによって送信されるすべての情報ソース (パッケージとラベル付けを含む <sup>20</sup>)、トレーニング、およびすべての物理的な制御と表示要素 (アラームと、各デバイスコンポーネントおよびユーザインターフェースシステム全体の操作のロジックを含む) も含まれます。ユーザインターフェースは、ユーザーがデバイスをセットアップするとき(開梱、セットアップ、校正など)、デバイスを使用するとき、またはデバイスのメンテナンスを実行するとき(例: 清掃、バッテリー交換、部品修理)など、インストールと使用のさまざまなフェーズを通じて使用される可能性があります。<sup>21</sup> ユーザーにとってのデバイスの安全性と有効性をサポートする方法の 1 つは、デバイスが定義されたユーザーのニーズに準拠していることを確認するために、使用中に重要な情報が提供されるようにユーザインターフェースを設計することです。ユーザインターフェース全体に重要な情報を統合するアプローチは、デバイスのユーザーが、デバイスの意図された使用と一貫して、安全で効果的な使用をサポートするために、適切なタイミングと場所で情報にアクセスできるようにするのに役立ちます。<sup>22</sup> ソフトウェアまたはモバイルアプリケーションの場合、製造業者は、デバイスのラベリングに加えて、画面上の情報や他の製品に送信されるアラートなどのユーザインターフェース要素を活用して、デバイスに関するリスクを伝達し、必要な情報が適切なタイミングで提供されるようにすることができます。

デバイスがどのように機能するかについての当局の理解を支援するために、販売申請においてユーザインターフェースの全体的な理解を提供することが重要です。治験依頼者がリスクを管理するためにリスク分析または申請の別のセクションでユーザインターフェースの設計を参照している場合、ユーザインターフェースを含めることで、それらのリスク管理の説明をサポートすることもできます。ただし、リスクコントロールの有効性の実際の分析は、ユーザインターフェースの説明とは別に配置する必要があります。このトピックに関する詳細は、セクション VII (リスク評価) および付録 D (ユーザビリティ評価の考慮事項) に記載されています。

特にラベリングに関しては、デバイスのユーザインターフェースには、ラベリングが含まれるが、これに限定されません。さらに、ユーザインターフェース内では、ラベリングは特定の規制の対象となります。例えば、デバイスが処方用であるかどうかに応じて、メーカーは、素人、または処方用デバイスの場合は、デバイスを管理するために法律で認可された医療従事者が「デバイスを安全に、意図された目的のために使用できる」ことを保証する適切な使用方法を含むラベリングを提供する必要があります。<sup>23</sup>

<sup>20</sup> See section 201(m) of the FD&C Act which defines labeling as “all labels and other written, printed, or graphic matter (1) upon any article or any of its containers or wrappers, or (2) accompanying such article.” 21 U.S.C. § 321(m).

<sup>21</sup> See FDA’s guidance titled “[Applying Human Factors and Usability Engineering to Medical Devices.](#)”

<sup>22</sup> See 21 CFR 820.30(g).

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

AI 対応デバイスに関するこれらの要件を満たす方法の一つは、ラベリングにおいて、モデル、そのパフォーマンス特性、およびモデルがデバイスにどのように統合されているかについての明確な情報を提供することです。モデル、そのパフォーマンス特性、およびモデルがデバイスにどのように統合されていますか。たとえば、ユーザーは、モデルがトレーニングされたデータの性質など、モデルに関する特定の情報を知る必要がある場合があります。これらの技術的特性は、デバイスがどのように動作することが期待されるか、およびどのような要因がパフォーマンスに影響を与える可能性があるかについてのユーザーの理解を支援することができるため、デバイスの安全かつ効果的な使用に不可欠です。

以下のセクションでは、ユーザーに伝達される内容とユーザーが操作するデバイスの要素に関する FDA の理解を支援するために、販売申請で提供する必要があるユーザーインターフェース(第六節-A: Section VI-A)およびラベリング(第六節 B: Section VI-B)に関する推奨情報について詳しく説明します。

本ガイダンスの付録 B (Appendix-B: Transparency Design Considerations) では、AI 対応医療機器のユーザーインターフェース (ラベリングを含む) を設計する際に考慮すると役立つ可能性のある情報の種類、通信モード、および通信スタイルの例を含む、透明性に対する推奨アプローチを概説しています。また、AI 対応機器に関する情報を明確に伝達するために、機器のラベリングにモデルカードを統合することも役立つ場合があります (付録 E (Appendix-E: Example Model Card) を参照)。

UDI 要件の対象となる機器 (AI 対応機器を含む) には、ラベリングに一意的デバイス識別子 (UDI) を含める必要があることに注意してください<sup>24</sup>。新しいバージョンやモデルがある場合、および新しいデバイスパッケージには、新しい UDI が必要です<sup>25</sup>。詳細については、FDA の Web サイト「[Unique Device Identification System](#)」を参照してください。

*Contains Nonbinding Recommendations*  
*Draft – Not for Implementation*

## A. ユーザーインターフェース

AI 対応デバイスの申請に含めるべき理由: FDA は、デバイスの使用方法を理解するために、デバイスのユーザーインターフェースを理解することが重要です。ユーザーインターフェースは、デバイスが意図していること、およびユーザーがどのように操作することを意図しているかについての重要な情報を伝えることができます。ユーザーインターフェースを見ることで、FDA はデバイスがどのように操作され、臨床ワークフローにどのように適合するかを理解することができます。これは、デバイスのレビューを支援し、デバイスが安全で有効であるかどうかを判断するのに役立ちます。

ユーザーインターフェースの表現は、ユーザーインターフェースがこれらのセクションの要素として参照されている場合に、治験依頼者のリスク評価やその他の文書をサポートするのにも役立ちます。たとえば、ユーザーインターフェースは、デバイスの安全かつ効果的な使用をサポートする重要な情報をユーザーに伝達することができ、ユーザーインターフェースの設計は、デバイスの安全かつ効果的な使用に不可欠な情報を知らなかつたり誤解したりすることに関連するリスクを制御または排除する上で重要な役割を果たす可能性があります。必須ではありませんが、治験依頼者がリスク評価におけるリスクコントロールの一部としてユーザーインターフェースの要素を使用することを選択した場合、ユーザーインターフェースを含めることで、レビューをさらに容易にすることができます。このトピックの詳細については、「セクション VII (Section VII: リスク評価)」および「付録 D (Appendix D: ユーザビリティ評価の考慮事項)」を参照してください。

<sup>23</sup> See 21 CFR 801.5; 21 CFR 801.109(d); FD&C Act section 502(f), 21 U.S.C. § 352(f). Device labeling must comply with the requirements in 21 CFR part 801 and any device specific labeling requirements such as for hearing aids or in special controls.

<sup>24</sup> See 21 CFR 801.20.

<sup>25</sup> See 21 CFR 830.50.

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

ユーザインターフェースには印刷されたラベリング(パッケージやユーザーマニュアルなど)が含まれており、ユーザインターフェースのすべての要素は、ユーザーがデバイスの使用方法を総合的に理解できるように設計されている必要がありますが、治験依頼者はセクション VI.B (Section VI-B ラベリング) に記載されているように、ラベリングを個別に提出する必要があります。このセクションでは、治験依頼者がユーザインターフェースの残りの要素について FDA に理解を提供する方法について説明します。

治験依頼者は、何を申請に含める必要があるか: 治験依頼者は、ユーザに提供される情報、情報が提供されるタイミング、および情報の表示方法など、デバイスワークフローを明確にするユーザインターフェースの情報と説明を提供する必要があります。ユーザインターフェースに関するこの種類の情報を提供する方法には、次のものがあります。

- デバイスとそのユーザインターフェースのグラフィカル表現(例えば、写真、イラスト、ワイヤーフレーム、線画)。これには、デバイス全体と、ユーザが操作するユーザインターフェースのすべてのコンポーネント(ディスプレイおよびファンクションスクリーン、アラームスピーカ、コントロールなど)の説明が含まれる場合があります。
- デバイスのユーザインターフェースの記述。
- デバイスの操作シーケンスと、ユーザがユーザインターフェースとの間で予想される操作の概要。これには、デバイスを使用するために実行されたユーザアクションのシーケンスと、必要に応じてデバイスの応答が含まれる場合があります。
- 出力形式の例 (予想される結果の範囲を表すレポートの例を含む)。
- 例えば、録画されたビデオを提供することによる装置のデモンストレーション。

治験依頼者が申請で提供できる場合: ユーザインターフェース情報は、販売申請のソフトウェアドキュメントセクションの「**ソフトウェアの説明**」に含める必要があります。

## **B. ラベリング**

AI 対応デバイスの申請に含める必要がある理由: 販売申請には、提案されたラベリングが販売申請の種類に適用される要件を満たしているかどうかを FDA が判断できるように、十分に詳細なラベリング情報を含める必要があります。<sup>26</sup> デバイスのラベリングは、前述の 21 CFR Part 801 を含むがこれに限定されない、すべての該当する FDA ラベリング要件を満たす必要があります。<sup>27</sup> ガイダンスのこのセクションには、これらの要件への準拠をサポートするための AI 対応デバイスのラベリングに関する考慮事項が含まれています。

治験依頼者が申請に含める必要があるもの: AI 対応デバイスのラベリングでは、ユーザーが重要な情報にすばやくアクセスできるように、目的のユーザー(例えば、年齢、教育または識字レベル、感覚障害または身体障害、職業上の専門性などの特性を考慮します。) に適した形式と読み取りレベルで、次の種類の情報を扱う必要があります。この情報を伝達するために、表とグラフィックスを使用することができます。

<sup>26</sup> See e.g., 21 CFR 807.87(e) or 21 CFR 814.20(b)(10).

<sup>27</sup> Generally, if the device is an in vitro diagnostic device, the labeling must also satisfy the requirements of 21 CFR 809.10.



*Contains Nonbinding Recommendations*  
*Draft – Not for Implementation*

### AI の包含

- デバイスで AI が使用されているという記述。
- デバイスの使用目的を達成するために AI がどのように使用されているかの説明。
  - 複数の機能を持つデバイスの場合、この説明には、AI-DSF が互いにどのように相互作用するか、および AI-DSF 以外の DSF とどのように相互作用するかを含めることができます。

### モデル入力

- モデル入力の説明(例えば、他の互換性のある装置から取得された信号又はパターン、取得システム(例、MRI)からの画像又は患者由来のサンプルであって、手動又は自動で入力することができるもの)。考慮すべき関連側面は次のとおりです。
  - 電子インターフェイスからの入力を組み込んだシステムの場合、入力が AI 対応デバイスの設計および検証と一致していることを確認するために必要なシステム構成に関する情報。<sup>28</sup>
  - 他の医療機器(例えば X 線装置)からの入力を必要とするシステムの場合は、特定の互換性のある機器または機器仕様の一覧と、該当する場合は許容される取得プロトコル。
  - モデル入力の喪失によって AI 対応装置が出力を生成できなくなる可能性があるシステムについては、喪失した入力が AI 対応装置の性能に及ぼす潜在的な影響の説明。
- デバイスによる処理のために入力データを準備するためにユーザーが実行すると予想される手順について、これらの手順を実行するユーザーの予想される特性(例えば、機能的な能力、経験と知識のレベル、訓練のレベル)を含めて説明します。

この情報は、デバイスの検証で検討された使用目的と一致している必要があります。

### モデル出力

- モデル出力の意味と使用方法について説明します。

### オートメーション

- デバイスが示す自動化の目的の程度について説明します。

### モデルアーキテクチャ

- デバイスに実装されているモデルを開発するために使用される方法とアーキテクチャの概要について説明します。

### モデル開発データ

- 以下を含む開発データの説明。
  - ソース (s);
  - 研究施設;
  - サンプルサイズ;
  - 人口統計学的分布;および
  - 臨床参照基準 (ground truth) を決定するために使用される基準/専門知識。

<sup>28</sup> For more information, please see FDA guidance titled "[Design Considerations and Premarket Submission Recommendations for Interoperable](#)

**Contains Nonbinding Recommendations**  
**Draft – Not for Implementation**

[Medical Devices.”](#)

### パフォーマンスデータ

- 以下を含むパフォーマンス検証データの説明。
  - データのソース;
  - 研究サイト;
  - サンプルサイズ;
  - その他の重要な研究デザインおよびデータ構造情報(例えば、ランダム化スキーム、反復測定、臨床参照基準);
  - 事前に指定されたパフォーマンス基準を含む検証研究の主要エンドポイント;および
  - 臨床参照標準データの決定に使用される基準/専門知識。

### デバイスのパフォーマンスメトリック

- デバイスのパフォーマンスメトリックの説明。
- パフォーマンス指標の例としては、受信者動作特性曲線下面積 (AUROC)、感度及び特異度、真/偽陽性及び真/偽陰性の数(例えば、混同行列)、
- 陽性/陰性的中率 (PPV/NPV)、陽性/陰性診断尤度比 (PLR/NLR) などの指標が挙げられます。すべてのパフォーマンス推定値は、信頼区間と共に提供されるべきです。
- 重要なサブグループにわたるデバイスパフォーマンスの説明。一般に、患者特性(例:性別、<sup>29</sup>性別、年齢、人種、民族、<sup>30</sup>疾患の重症度)、地理的な場所、およびデータ収集機器によるサブグループ分析が適切です。
- 必要に応じて、各操作ポイントのサブグループ分析を含む、さまざまな操作ポイントに対応するパフォーマンスの説明。

### パフォーマンスモニタリング

- デバイスの安全かつ効果的な使用のために、ユーザーによる継続的なパフォーマンスの監視と管理が必要であると考えられる場合に、該当する場合は、そのようなツールの使用方法を含む、デバイスのパフォーマンスを監視および管理するための方法またはツールの説明。

### 制限事項

- AI 対応デバイス、AI-DSF (s)、またはモデルのすべての既知の制限事項の説明。
  - モデルの制限の中には、禁忌、警告、または注意を保証するほどの重大度に達しないものもありますが、それでもラベルに含めることが重要な場合があります。たとえば、トレーニングデータセットには、病気や状態のまれな症状を示す少数の患者しか含まれていない可能性があります。そのまれな症状がモデルによって診断として示唆されている場合、ユーザーはデータの制限を知ることによって利益を得ることができます。

<sup>29</sup> For more information regarding sex-specific data, please see FDA guidance titled “[Evaluation of Sex-Specific Data in Medical Device Clinical Studies.](#)”

<sup>30</sup> For more information regarding the reporting of age, race, and ethnicity related data, please see FDA guidance titled “[Evaluation and Reporting of Age-, Race-, and Ethnicity-Specific Data in Medical Device Clinical Studies.](#)”

*Contains Nonbinding Recommendations  
Draft – Not for Implementation*

### インストールと使用

- 以下を含む、インストールと実装の手順に関する情報。
  - AI 対応デバイスをサイトのデータシステムと臨床ワークフローに統合する手順;および
  - 入力データがデバイスと互換性があり、適切であることを確認する手順<sup>31</sup>
    - 用語を明示的に定義する必要がある場合があります。たとえば、医療システムとメーカーの両方に「性別」というラベルの付いたデータがある場合、一方は出生時に性別を使用し、もう一方は自己申告の性別を使用している可能性があります。

### カスタマイズ

- 次のようなカスタマイズ可能な機能の説明と手順
  - ユーザまたは医療システムがデバイスの動作ポイントを設定できる場合;
  - 異なる構成を選択することが適切な場合;および
  - 動作ポイントが設定可能な場合、エンドユーザがデバイスを現在動作している動作ポイントを識別する方法。

### メトリックと視覚化

- モデル出力にコンテキストを追加するために使用される追加のメトリックまたは視覚化の説明。

### 患者と介護者の情報

患者又は介護者が使用することを意図した人工知能（AI）対応機器については、製造業者は、患者及び介護者向けに設計された、使用方法、機器の適応、使用目的、リスク及び制限事項を記載した表示資料を提供すべきです。患者及び介護者は、機器を操作し、結果を解釈し、又は結果に基づいて意思決定を行う場合には、機器の唯一の使用者又は主要な操作者でない場合でも、使用者とみなされます。この資料は、対象とする読者にとって適切な読解力のあるものでなければなりません。患者及び介護者向けの資料が提供されていない場合、治験依頼者は、患者及び介護者が機器の使用方法をどのように理解するかについての説明を提供すべきです。これには、機器を使用するかどうかの決定方法及び機器の出力をどのように使用するかについての決定方法が含まれます。

治験依頼者が申請書に記載すべき箇所: AI 対応機器のラベリングに関する情報は、市販申請書の「**ラベリング**」の項に記載すべきです。

---

<sup>31</sup> For more information, please see FDA guidance titled, "[Design Considerations and Pre-market Submission Recommendations for Interoperable Medical Devices.](#)"

*Contains Nonbinding Recommendations  
Draft – Not for Implementation*

**追加のリソース:**

- 付録 B (Appendix B: Transparency Design Considerations) では、デバイスの使用適応症とモデルカードを理解するための可能性のあるアプローチについて概説しており、ユーザインターフェースの開発に役立つ可能性があります。
- モデルカードは、ラベリングやユーザインターフェースに関する情報を提示するためには必要ありませんが、情報を整理するのに役立つツールとなる場合があります。一般に、モデルカードは、各AI対応機器の特定のニーズと状況に合わせて調整することができます。
  - 付録E (Appendix E: モデルカードの例) には、ユーザーと医療提供者を対象とした基本的なモデルカードのフォーマットの例が含まれています。このフォーマットには、モデルの使用目的と対象ユーザーの概要、安全性と有効性を裏付ける証拠などの情報が記載されています。
  - 付録F (Appendix F: モデルカードを含む例510 (k) 提出概要) には、完成した基本的なモデルカードの例が含まれています。
- FDAのガイダンス「[Device Labeling Guidance#G91-1 \(Blue Book Memo\)](#)」には、機器のラベリングに含めるべき情報に関する提案が含まれています。

## VII. リスク評価

AI 対応デバイスの申請に含めるべき理由: 包括的なリスク評価は、デバイスの安全性と有効性を確保するのに役立ちます。包括的なリスク評価を販売申請に含めると、FDA は適切なリスクが特定されているかどうか、およびそれらがどのように管理されているかを理解するのに役立ちます。[市販前ソフトウェアガイダンス](#)のセクション VI.C で、FDA は、デバイスソフトウェア機能を含む販売申請に、リスク管理計画、リスク評価、およびリスク管理レポートで構成されるリスク管理ファイルを含めることを推奨しています。これに沿って、AI 対応デバイスの販売申請には、他の適用可能なガイダンスに加えて、[市販前ソフトウェアガイダンス](#)の推奨事項とこのガイダンスの推奨事項を考慮したリスク管理ファイルを含める必要があります。

治験依頼者は、AI 対応デバイスにも適用されるリスク管理ファイルの開発と適用に関する追加情報について、FDA が承認した ANSI/AAMI/ISO 14971 医療機器-医療機器へのリスク管理の適用についても参照する必要があります。FDA はまた、AI 対応デバイスが、一般的にデバイスソフトウェア機能とは異なる新しいリスクに関連付けられる可能性があることも認識しています。そのため、FDA は、AI 対応デバイスに固有の AAMI CR34971 人工知能および機械学習への ISO 14971 の適用に関するガイダンスの FDA が承認した自主的なコンセンサス基準に記載されている考慮事項を治験依頼者に組み込むこともお勧めします。

### TPLC 全体のリスク

リスク分析を行う際には、医療機器; 現行の適正製造基準 (CGMP) の最終規則(1996年10月7日、[61 FR 52602](#))に「メーカーは、正常状態と故障状態の両方において設計に関連して起こり得る危険を特定することが期待されます。次に、使用者の過誤に起因するものを含め、危険有害性に関連するリスクを、通常状態と故障状態の両方で計算すべきです。リスクが許容できないと判断された場合には、適切な手段により許容可能なレベルまで低減されるべきです。」と記載されています。このリスク評価では、このガイダンスのセクション VI (Section VI: ユーザインターフェースとラベル付け) で説明されているように、TPLC 全体のすべてのユーザーを考慮する必要があります。FDA は、製造業者が TPLC 全体で AI 対応デバイスに対してこのアプローチに従うことを推奨しています。

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

AI 対応デバイスの情報に関連するリスク

AI 対応デバイスにとって特に重要となり得るリスク管理の 1 つの側面は、情報の不足や不明確な情報に関連するリスクなど、デバイスの使用や解釈に必要な情報の理解に関連するリスクの管理です。誤解、誤用、または利用できない情報は、デバイスの安全かつ効果的な使用に影響を与える可能性があります。たとえば、AI 対応デバイスを含む複雑なアルゴリズムを利用するデバイスでは、さまざまな疾患サブタイプのパフォーマンスがユーザーに明らかでない場合や、出力情報の基になるロジックが簡単に理解できない場合があり、ユーザーの理解とデバイスの使用に悪影響を与える可能性があります。情報の不足や不明確な情報は、デバイスが期待どおりに動作していないかどうかや、指示に正しく従う方法をさまざまなユーザーが理解するのを困難にすることもあります。FDA は、情報の理解に関連するリスクの考慮は、AI 対応デバイスのリスク管理に対する包括的なアプローチの一部である必要があることを推奨します。

**追加リソース:**

- ANSI/AAMI HE75 *Human factors engineering*-医療機器の設計には、リスクを管理するためにラベリングで情報を使用することに関する推奨事項が含まれています。

治験依頼者が提出する内容：治験依頼者は、リスク評価を含むリスク管理計画を含む「リスク管理ファイル」を提供する必要があります。その他の考慮事項に加えて、リスク評価では、デバイスのインストールプロセス、時間の経過に伴うパフォーマンスの維持、デバイスの結果のユーザー解釈に関連するリスクなど、デバイスの使用全体を通して発生するユーザータスクと知識タスクを考慮する必要があります。FDA が承認した自主的なコンセンサス基準に記載されている考慮事項に加えて<sup>33</sup>、FDA は治験依頼者がリスク評価中に情報を理解することに関連するリスクを考慮することを推奨しています。特定されたすべてのリスクと同様に、治験依頼者は、特定されたリスクに対処するラベリングなどのユーザーインターフェースの要素を含む、リスクコントロールの説明を提供する必要があります。このようなリスクとそのコントロールについて説明するのに役立つ可能性のある情報については、「付録 D (Appendix D: ユーザビリティ評価の考慮事項)」を参照してください。

治験依頼者が申請書に記載すべき箇所：AI 対応機器のリスク評価に関する情報の多くは、[市販前ソフトウェアガイドライン](#)で推奨されているように、市販申請書のソフトウェア文書セクションの「**リスク管理ファイル**」に記載すべきです。

<sup>32</sup> For more information, see the [FDA Recognized Consensus Standards Database](#).

<sup>33</sup> For more information regarding use of consensus standards in regulatory submissions, refer to the FDA guidances titled “[Appropriate Use of Voluntary Consensus Standards in Premarket Submissions for Medical Devices](#)” and “[Standards Development and the Use of Standards in Regulatory Submissions Reviewed in the Center for Biologics Evaluation and Research](#).”

**追加リソース:**

- 付録B (Appendix B: 透明性設計の考慮事項) は、リスクの特定とリスクコントロールの開発に役立つ可能性のある、デバイス開発におけるユーザー中心の設計アプローチに関する推奨事項を概説しています。
- 付録D (Appendix D: ユーザビリティ評価の考慮事項) は、ユーザビリティテストに関する推奨事項を示しています。これは、スポンサーが情報関連リスクに対するコントロール案の有効性を評価するのに役立つ可能性があります。

## VIII. データ管理

AI 対応デバイスの申請に含めるべき理由: AI 対応デバイスの場合、モデルは作用メカニズムの一部です。したがって、FDA は、デバイスがどのように開発され、検証されたかを理解するために、データ管理のプラクティス(すなわち、データがどのように収集され、処理され、注釈が付けられ、保存され、管理され、使用されてきたか、またはどのように使用されますか)や、AI 対応デバイスの開発と検証に使用されたデータの特徴付けを含む、データ管理についての明確な説明が不可欠です。この理解は、FDA が AI 対応デバイスの安全性と有効性を評価できるようにするのに役立ちます。

AI システムのパフォーマンスと動作は、トレーニングと調整に使用されるデータの質、多様性、および量に大きく依存します。AI 対応デバイスの検証の精度と有用性は、テストに使用されるデータの質、多様性、および量にも依存します。そのため、FDA の審査官は、AI 対応デバイスが安全で有効かどうかを理解するためにデータ管理を評価します。これには、トレーニングデータとテストデータの収集と管理を、意図された使用と結果として生じるデバイス要件と整合させることが含まれます。

データ管理は、バイアスを特定して軽減するための重要な手段でもあります。AI 対応デバイスにおける AI バイアスの可能性を評価するには、バイアスの発生源を特性化する必要があります。AI バイアスとは、トレーニングデータの制限や機械学習プロセスにおける誤った仮定のために、体系的ではあるが予測できない方法で誤った結果を生成する潜在的な傾向です。AI バイアスは十分に文書化されています<sup>34</sup>。たとえば、トレーニング中に、特定のスキャナー、患者サブ集団、または臨床現場に固有の画像の特徴を認識するためにモデルが過剰にトレーニングされる可能性があります。一般化可能な患者の解剖学、生理学、または状態とはほとんど関係がありません。これにより、結果のモデルで AI バイアスが発生する可能性があります。別の例では、データセット内の特定の集団の表現が不足していると、人口統計学的特性に基づいた過剰適合(すなわち、訓練データの潜在的バイアスにあまりにも密接に適合するデータ)につながる可能性があります。これは、表現が不足している集団における AI 対応デバイスのパフォーマンスに影響を与える可能性があります。

モデルにバイアスのない代表的なトレーニングデータを使用すると、意図した使用集団への一般化が促進され、データ自体からのバイアスや特異性の永続化が回避されます。たとえば、画像認識タスクでは、すべての症例が同じ機器で画像化されている場合、または定規が含まれている場合に交絡が発生する可能性があります(例えば黒色腫の臨床画像)。

<sup>34</sup> See Karen Hao, "This is how AI bias really happens—and why it's so hard to fix," MIT Technology Review 2019.

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

潜在的な交絡因子のうち 1 つの例は、トレーニングで米国外（OUS）で収集されたデータを使用することであり、人口統計学、医療行為、または標準治療の違いにより OUS 集団が米国集団を反映していない場合、モデルにバイアスがかかる可能性があります。トレーニングデータ内のこのような交絡因子は、特定して軽減しないと、モデルによって不注意に学習され、無関係な特性に基づいて一見正確な（しかし誤解を招く）予測につながる可能性があります。

過小評価は、過小評価された集団のパフォーマンスを理解するなど、パフォーマンスの問題を特定する能力に影響を与える可能性があるため、検証データセットに代表的なデータを含めることが重要になる場合があります。バイアスを完全に排除することは困難な場合がありますが、FDA は、製造業者が出発点として、検証データが医療機器の意図された使用（ターゲット）集団を十分に表していることを確認することを推奨します。

年齢、人種、民族、性別に固有のデータに関する詳細については、FDA のガイダンスを参照してください。

「[FDA 規制医薬品の臨床試験および臨床研究における人種および民族データの収集](#)」。<sup>35</sup> 「[医療機器臨床試験における年齢、人種及び民族に特有のデータの評価及び報告](#)」および「[医療機器臨床試験における性別に固有のデータの評価](#)」。

検証データに開発データと同じ交絡因子が見つかった場合、正しい予測につながっていると思われる偽の相関関係を特定することは特に困難になる可能性があります。したがって、AI 対応デバイスの開発と検証に使用されたデータセットの代表性に関する情報は、FDA が実質的な同等性を判断するのに役立つか、またはデバイスが意図された用途に対して安全で効果的であるという合理的な保証があるかどうかを判断するために重要です。<sup>36</sup> AI バイアスに対処するだけでなく、データ管理の詳細はデバイスの意図された用途をサポートする必要があります。

デバイスのパフォーマンスを客観的に評価するために、FDA の審査官は、テストデータがトレーニングデータから独立している（例えば、全く異なる臨床現場から採取されたもの）、モデル開発者とモデル開発段階から隔離されているかどうかを理解することも重要です。開発データセットとテストデータセットを適切に分離することは、AI 対応デバイスの真のパフォーマンスを評価するのに役立ちます。検証データセットと開発データセットの間のデータ漏洩により、AI 対応デバイスの真のパフォーマンスに関する不確実性が生じる可能性があります。<sup>37</sup>

治験依頼者が申請に含める必要がある内容:申請では、治験依頼者はトレーニングデータとテストデータの両方について、適切な販売申請セクションで次の種類の情報を提供する必要があります。以下に説明するセクションごとにデータ管理情報を整理すると役立つ場合があります。一般に、データ収集、開発およびテストデータの独立性、参照基準、および代表性に関する情報を提供する必要があります。また、治験依頼者は、開発段階と検証段階との間のデータ管理手法及びデータの特性的の違いについて説明する必要があります。提出書類には、その違いとその正当性についての説明を含める必要があります。

<sup>35</sup> When final, this guidance will represent FDA's current thinking on this topic.

<sup>36</sup> For more information, see FDA guidance titled "[Acceptance of Clinical Data to Support Medical Device Applications and Submissions: Frequently Asked Questions](#)."

<sup>37</sup> Robert F Wolff, Karel GM Moons, Richard D Riley, et al. "PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies." *Annals of Internal Medicine* 170, no. 1 (2019): 51–58 <https://doi.org/10.7326/M18-1376>; Altman, Douglas G., and Patrick Royston. "What Do We Mean by Validating a Prognostic Model?" *Statistics in Medicine* 19, no. 4 (2000): 453–73 [https://doi.org/10.1002/\(sici\)1097-0258\(20000229\)19:4<453::aid-sim350>3.0.co;2-5](https://doi.org/10.1002/(sici)1097-0258(20000229)19:4<453::aid-sim350>3.0.co;2-5).

*Contains Nonbinding Recommendations  
Draft – Not for Implementation*

## データ収集

- 以下を含む、データがどのように収集されたかの説明(例えば、選択/除外基準を有する臨床試験プロトコル)。
  - 関係する臨床施設または施設の名称。
    - 施設は一意に識別されるべきであり、申請書全体を通じて一貫して参照されるべきです。
  - データが取得された期間。
  - データが既存のデータベースから使用された場合は、このデータベースの使用の適切性。
  - 実世界データ (RWD) を使用する場合は、この証拠のソースと収集。
    - RWD を使用する場合、FDA は治験依頼者が、RWD の関連性と信頼性の両方を評価する、選択したデータソースの目的適合データの評価を提供することを推奨します。FDA は、治験依頼者が RWD の使用案に関する FDA のフィードバックを得るために [Q 提出プログラム](#) を活用することを推奨します。RWD に関する詳細については、「[医療機器の規制上の意思決定を支援するための実世界の証拠の使用](#)」というタイトルの FDA ガイダンスを参照してください。
- データセットの制限の説明。
- データに関連する品質保証プロセスの説明。該当する場合は、データ取得中の人為的ミスから保護するために実施された管理を含みます。
- 各データセットのサイズの説明。
- 研究の範囲内で登録の多様性を向上させるために使用されたメカニズムの説明と、患者集団および臨床現場全体で研究結果の一般化可能性を確保する方法の説明。38 このトピックの詳細については、「[臨床試験における人種および民族データの収集](#)」というタイトルの FDA ガイダンスを参照してください。
- 合成データの使用に関する説明。39 規制当局への提出をサポートするために使用される合成データには、データがどのように生成されたか、およびデータが目的に適している理由についての包括的な説明を添付する必要があります。

## データのクリーニング/処理

最適なトレーニング結果を提供するには、開発に使用されるデータをクリーンアップすることが重要な場合があります。

たとえば、不正確なデータ、重複したデータ、不完全なデータを削除します。これらの処理手順を記述する必要があります。これには、使用されるデータ品質要因、データの包含/除外基準、欠落データの処理、手順が AI-DSF の内部または外部のどちらであるかなどが含まれます。

---

38 Sponsors may be required to develop or submit information regarding the enrollment of clinical study participants to help improve the strength and generalizability of the study results. For example, the FD&C Act, as amended by section 3601(b) of the Food and Drug Omnibus Reform Act of 2022 (FDORA), enacted as part of the Consolidated Appropriations Act, 2023 (P.L. 117-328)), requires sponsors to submit to FDA diversity action plans for studies of certain devices. See FD&C Act section 520(g)(9), 21 U.S.C. § 360j(g)(9).

39 For the purposes of this guidance, “synthetic data” is defined as data that have been created artificially (e.g., through statistical modeling, computer simulation) so that new values and/or data elements are generated. Generally, synthetic data are intended to represent the structure, properties and relationships seen in actual patient data, except that they do not contain any real or specific information about individuals. For more information, please see [FDA Digital Health and Artificial Intelligence Glossary – Educational Resource | FDA](#).



***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

一方、テストデータは、モデルが意図した使用で遭遇する RWD を代表する方法でのみ処理される必要があります。このようなデータ処理、使用されるデータ品質要因、データの包含/除外基準、および欠落データの処理は、最終的な AI-DSF で実装されている前処理に合わせて正当化される必要があります。

### 参照標準

本ガイダンスの目的において、参照基準は、各患者/症例/記録の真の状態を定義するために使用できる、利用可能な最良の代表的な真実です。<sup>40</sup> 参照基準は、デバイスのトレーニング、デバイスの検証、またはその両方で使用される可能性があります。参照基準は、臨床タスクに関する患者の真の状態を確立するために、医療および規制コミュニティ内の現在の実践から得られた証拠によって検証されます。参照基準は、臨床タスクを反映する必要があります。臨床タスクは、例えば、疾患または状態の分類、医療画像上の輪郭のセグメンテーション、境界ボックスによる検出、またはマーキングによる位置確認で構成される可能性があります。選択した参照基準に関して、以下の種類の情報を提供する必要があります。

- 参照基準がどのように確立されたかの説明。
- 選択した参照基準に内在する不確実性の説明。
- 参照基準を使用して得られた結果があいまいまたは欠落している可能性がある場合に対処するための戦略の説明。
- 参照基準が臨床医からの評価に基づいている場合は、以下を記載します。
  - 使用したグレード決定プロトコル。
  - これらの臨床医にどのようなデータが提供されますか。
  - 臨床参照基準を決定するために臨床医の評価がどのように収集/判断されますか。
    - 盲検プロトコル;および
    - 参加臨床医の数とその資格。
  - 必要に応じて、各タスクの臨床医内および/または臨床医間の変動の評価、および観察された変動が特定の測定タスクについて一般的に受け入れられている標準内にあるかどうかの評価。

### データ注釈

- データ注釈を使用する場合は、データ注釈のアプローチに関して以下の種類の情報を提供する必要があります。
  - データ注釈を行う担当者の専門知識の説明。
  - データ注釈に提供される特定のトレーニング、指示、またはガイドラインの説明。注釈が相互にブラインドされているかどうかを含め、注釈の決定の指針となるように。

<sup>40</sup> For an illustrative example of a reference standard, see FDA guidance titled "[Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data in Premarket Notification \(510\(k\)\) Submissions.](#)" This guidance addresses the reference standard for this device. Other device specific guidances and special controls note the appropriate reference standard to be used. For questions about what the appropriate reference standard may be for a device and proposed intended use, consult the appropriate review division via the Q-Submission Program.

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

- データ注釈の品質/一貫性を評価し、不一致を裁定する方法の説明(コンセンサス評価、サンプリング)。FDA は、客観的で高品質なデータ注釈を確保するために、他の注釈作成者の決定を知ることなく、各注釈作成者による独立した評価を使用することを推奨しています。および
- 不適切なデータ注釈に対処するための詳細な計画。

### データストレージ

トレーニングデータとテストデータの両方のデータストレージの説明。説明では、データセットのバージョン管理に対処し、このガイダンスのセクション XII (Section XII: サイバーセキュリティ) で説明されている項目に対処することによって、データのセキュリティを確保する必要があります。

### データの管理と独立性

- 開発データの説明 (開発データがトレーニング、チューニング、チューニング評価、および追加のサブセットに分割された方法、および各データセットを使用して実行されたモデル開発アクティビティの仕様を含む)。
- テストに使用されるデータが開発プロセスから隔離されていることを確認するために実施されているコントロールの説明。
- 検証に使用されるデータが堅牢な外部検証を提供する理由の正当性。
- たとえば、テストデータが生成されたサイトの説明。一般に、テストデータは AI-DSF の開発に使用されたサイトとは異なるサイトから取得する必要があります。

### 代表性

- データが意図された使用集団<sup>41</sup>をどのように代表しているかの説明と、以下を含む使用の適応症。
  - 可能な場合は、以下を含む関連する集団特性の説明。
    - 疾患状態(例:陽性/陰性症例、疾患重症度、疾患サブタイプ、併存疾患、疾患スペクトルの分布);
    - 患者集団の人口統計(例:性別、<sup>42</sup>性別、年齢、人種、民族、<sup>43</sup>身長、体重);
    - データ収集装置および条件(例えば、データが収集される場所、データ収集装置/方法、画像及び再構成プロトコル)、データ収集中に分析されたシグナルに影響を与える可能性のあるすべての要因を含む(例えば、患者が歩行しているか、安静にしているか、立っているかなどの患者の活動;または集中治療室、MRI などのデータ収集環境);そして

<sup>41</sup> Sponsors may be required to develop or submit information regarding the representativeness of clinical study participants. For example, the FD&C Act, as amended by section 3601(b) of FDORA, enacted as part of the Consolidated Appropriations Act, 2023 (P.L. 117-328)), requires sponsors to submit to FDA diversity action plans for studies of certain devices. See section 520(g)(9) of the FD&C Act, 21 U.S.C. § 360j(g)(9).

<sup>42</sup> For more information regarding sex-specific data, please see FDA guidance titled "[Evaluation of Sex-Specific Data in Medical Device Clinical Studies](#)."

<sup>43</sup> For more information regarding age-, race-, and ethnicity-related data, please see FDA guidances titled "[Evaluation and Reporting of Age-, Race-, and Ethnicity-Specific Data in Medical Device Clinical Studies](#)," and "[Collection of Race and Ethnicity Data in Clinical Trials](#)."

***Contains Nonbinding Recommendations  
Draft – Not for Implementation***

- 試験データ収集サイト(例:臨床現場、施設)。一般に、単一のデータ収集サイトは、初期のデータ評価段階では有用な出発点となり得るが、単一のサイトに依存することは一般に適切ではありません。データが意図された使用集団および使用適応を代表しているかどうかを理解するために適切ではありません。複数のデータ収集サイトを使用することは、例えば、多様な臨床診療環境（例えば、大規模大学病院と地域病院）のサイトのよう、意図された使用集団のより代表的なサンプルを保証することができます。例えば、少なくとも 3 つの地理的に多様な米国の臨床サイト（または医療システム）を使用することは、AI 対応デバイスを臨床的に検証するために適切である可能性がある<sup>44</sup>。
  - 上記の集団特性に対応するものを含め、重要な共変量に沿ったデータの分布の特徴。
  - 上記の関連する集団特性のいずれかがデータに利用できなかった場合、その理由の説明、およびこの情報なしでデータを使用したことの正当性。FDA は、トレーニングおよびテストデータで使用された患者および/またはサンプルのソースによっては、関連する患者特性情報の一部が利用できない可能性があることを理解しています。
  - 特定された共変量によって層別化されたサブグループ分析。
  - 検証時に OUS データを使用する場合は、患者の人口統計学的特性だけでなく、一般的な医療行為、疾患の症状、有病率、進行の観点から、データが米国の集団および米国の医療行為とどのように比較されているかに関する説明。<sup>45</sup>
    - 典型的なモデルはデータ駆動型であり、エンドユーザーにとってそのアルゴリズムが不明瞭であるため、検証データの大部分が OUS データである場合、米国の対象集団における一般化されたパフォーマンスを臨床試験で適切に把握できない可能性があります。また、AI 対応機器は、従来の医療機器よりもトレーニングデータやテストデータの特異パターンに敏感である可能性があります。これらの理由から、臨床検証ではより高い割合の米国データが必要になる可能性があります。FDA は治験依頼者に対し、OUS データの使用案に関する FDA のフィードバックを得るために Q-Submission プロセスを活用することを奨励します。

治験依頼者が申請書で提供する必要がある場合：モデルの開発に使用されるデータのデータ管理情報は、[市販前ソフトウェアガイダンス](#)に記載されているように、市販申請書のソフトウェアドキュメントセクションの「**ソフトウェアの説明**」に含める必要があります。

<sup>44</sup> For more information regarding site selection, please see FDA guidance titled "[Design Considerations for Pivotal Clinical Investigations for Medical Devices.](#)"

<sup>45</sup> For more information on the use of OUS data, please see FDA guidance titled "[Acceptance of Clinical Data to Support Medical Device Applications and Submissions: Frequently Asked Questions.](#)"

<sup>46</sup> For more information on the Q-Submission program, please see FDA guidance titled "[Requests for Feedback and Meetings for Medical Device Submissions: The Q-Submission Program.](#)"

*Contains Nonbinding Recommendations  
Draft – Not for Implementation*

性能検証(すなわち、臨床的検証)文書で使用されるデータのデータ管理情報は、市販申請の「**性能試験**」セクションに記載する必要があります。モデルトレーニングと検証に使用されるデータの特性が異なる場合、治験依頼者は、性能試験文書要素の性能検証データ管理セクションとともに、その違いを強調し、正当化する必要があります。

**追加リソース:**本ガイダンスの考慮事項に加えて、開発への TPLC アプローチをサポートするために、FDA はスポンサーと研究者に対し、臨床試験の研究デザイン、実施、および報告段階で AI 対応デバイスの固有の特性を考慮することを推奨しています。研究者は、Investigational Device Exemption (IDE)、Protection of Human Subjects and Institutional Review Board regulations,<sup>47</sup>、および Good Clinical Practice (GCP) regulations<sup>48</sup> がデバイスにどのように適用されるかを理解する必要があります。リソースには、コンセンサスガイドライン<sup>49</sup>、および以下のタイトルの FDA ガイダンスが含まれます。

- [“Significant Risk and Nonsignificant Risk Medical Device Studies”](#)
- [「治験審査委員会、治験責任医師及び治験依頼者に対するインフォームド・コンセントの指針」](#)
- [「医療機器の申請及び申請を支援するための臨床データの受入れ:よくある質問」](#)

年齢、人種、民族に固有のデータ、および性別に固有のデータに関する詳細については、以下のタイトルの FDA ガイダンスを参照してください。

- [「臨床試験における人種および民族に関するデータの収集」](#)
- [「医療機器臨床試験における年齢、人種、民族特異的データの評価と報告」](#)
- [「医療機器臨床試験における性特異的データの評価」](#)

## IX. モデルの説明と開発

AI 対応デバイスの申請に含めるべき理由: バイアスや制限を含むモデル (およびデバイス) 設計に関する情報は、AI 対応デバイスの安全性と有効性を評価し、デバイスのパフォーマンステスト仕様を決定する FDA の評価能力をサポートします。

[市販前ソフトウェアガイダンス](#)のセクション VI.B には、モデルの説明を含め、販売申請におけるソフトウェアの説明の一部として含めるべき情報が記載されています。デバイスの説明はより広範で、デバイス全体、ユーザーがそれを操作する方法、およびそれが臨床ワークフローにどのように適合するかについての情報を提供しますが、モデルの説明は、ソフトウェアの説明の一部として、モデル自体の技術的な特徴と、その開発に使用されたアルゴリズムと方法に関する詳細な情報を具体的に提供します。この情報は、FDA が AI 対応デバイスの機能の基礎を理解するのに役立ちます。モデルの開発に使用された方法を理解することは、FDA が潜在的な制限、AI バイアスの原因、適切なデバイスのラベリングに関する考慮事項を特定するのにも役立ちます。

<sup>47</sup> See 21 CFR Parts 50 and 56.

<sup>48</sup> See FDA’s website on [Regulations: Good Clinical Practice and Clinical Trials](#).

<sup>49</sup> See Liu, Xiaoxuan et al “Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension,” *Natural Medicine* (2020) <https://doi.org/10.1038/s41591-020-1034-x>; Rivera, Samantha C. et al “Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension,” *Lancet* (2020) [https://doi.org/10.1016/S2589-7500\(20\)30219-3](https://doi.org/10.1016/S2589-7500(20)30219-3) Vasey, Baptiste et al Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI,” *BMJ* (2022)<https://doi.org/10.1136/bmj-2022-070904>.

*Contains Nonbinding Recommendations*  
*Draft – Not for Implementation*

治験依頼者が申請に含めるべき内容：申請には、AI 対応デバイスの各モデルについて以下に説明する情報を含める必要があります。

AI 対応デバイスの一部として複数のモデルが使用されている場合は、モデルの出力を組み合わせてデバイスの出力を作成する方法の図を含めると特に役立ちます。アルゴリズムとモデルの説明は、有能な AI 担当者が同等のモデルを作成できるように、十分に詳細である必要があります。わかりやすくするために、テキストの説明に加えて図を使用することをお勧めします。

#### モデルの説明

- AI 対応デバイスの一部として使用される各モデルの説明。次のものが含まれますが、これに限定されません。
  - モデルの入力と出力;
  - モデルアーキテクチャの説明;
  - 機能の説明;
  - 必要に応じて、モデルの設計と最適化に使用される機能選択プロセスと損失関数の説明;および
  - モデルパラメータ。
- AI 対応デバイスに、モデルに関連するカスタマイズ可能な機能がある場合（事前に定義された複数の操作ポイントで動作するようにカスタマイズできる場合や、入力の数が増える場合など）、カスタマイズを可能にして制御するモデルの技術要素の説明。
- AI ベースおよびサードパーティのものを含む、入力データの品質管理基準またはアルゴリズムの説明。これには、品質評価メトリックがデバイスの使用目的とどのように一致するか(対象患者集団及び使用環境等)が含まれます。
- 以下を含む、入力および/または出力データに適用されるメソッドの説明。
  - 入力データの前処理(正規化など);
  - 出力データの後処理;および
  - データの補強または合成。

#### モデルの開発

- モデルがどのようにトレーニングされたかの説明：
  - 最適化方法;
  - トレーニングパラダイム(例えば、教師あり学習、教師なし学習、半教師あり学習、連合学習、アクティブ・ラーニング);
  - 使用された正則化技術;
  - 必要に応じて、トレーニングハイパーパラメータ(損失関数学習率など);および
  - さまざまなデータサブセットの損失関数収束曲線など、トレーニングパフォーマンスの概要(トレーニング、チューニング、チューニング評価など)。
- チューニング評価が実施された場合は、メトリックと得られた結果の説明。
- 必要に応じて、使用された事前トレーニング済みモデルの説明。

**Contains Nonbinding Recommendations**  
**Draft – Not for Implementation**

- 事前トレーニング済みモデルが使用された場合は、事前トレーニングに使用されたデータセットと、事前トレーニング済みモデルがどのように取得されたかを明記。
- 必要に応じて、アンサンブル法(例えば、袋詰めまたはブースト)の使用の説明。
- 閾値(動作点など)がどのように決定されたかの説明。
- モデル出力のキャリブレーションの説明。

治験依頼者が申請書に記載すべき場合:モデルの説明及びモデルの開発方法を含むモデル開発に関する情報は、[市販前ソフトウェアガイダンス](#)に記載されているとおり、申請書のソフトウェア文書の「**ソフトウェアの説明**」の一部として記載すること。

**追加資料:**製造業者が自動または継続的に更新されるモデルの開発を検討したい場合、FDA は製造業者に対し、[Q 提出プログラム](#)を利用して、開発プロセスの早い段階でこれらの AI モデルに関連する考慮事項について議論し、FDA ガイダンス「[Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions](#)」を確認することを奨励します。

## X. 検証 (Validation)

AI 対応デバイスの場合、検証には、ユーザーが利用するデバイスが意図した用途を安全かつ効果的に実行できることを保証することと、デバイスの関連する性能仕様を一貫して満たすことができることを確立することが含まれます。AI 対応デバイスの場合、メーカーは、デバイス自体が関連する性能仕様を満たしていることを保証することに加えて、ユーザーが意図したとおりにデバイスと対話し、理解する能力を実証する必要があります。この目的のために、パフォーマンスの検証(人的要因の検証を含む)とユーザビリティの評価の両方を検討することが役立ちます。本ガイダンスの目的(人的要因の検証がない場合のリスク管理の文脈において)では、ユーザビリティとは、ユーザーが AI 対応デバイスに関連する情報を一貫して正しく受け取り、理解し、解釈し、適用できるかどうかを含め、意図したユーザーがデバイスを安全かつ効果的に使用できるかどうかを説明することに注意してください。

FDA ガイダンス「[医療機器への人的要因とユーザビリティエンジニアリングの適用](#)」(以下、「[人的要因ガイダンス](#)」と呼ぶ)は、デバイスに関する推奨事項と要件を説明し、人的要因の検証テストには「患者またはユーザーに重大な害をもたらす、またはもたらす可能性のある使用上のエラーを特定するために、デバイスのユーザインターフェースとのユーザー対話を評価するためにデバイス開発プロセスの最後に実施されるすべてのテスト」が含まれ、「リスク管理措置の有効性を評価するため」にも使用されることを確立しています。[人的要因ガイダンス](#)は、重要なタスクを持つデバイスに対する人的要因の検証に関する具体的な推奨事項と要件を概説していますが、同じまたは類似のプロセスの適用は、他のリスクの適切な管理を示すのにも役立ちます。付録 D (Appendix D: ユーザビリティ評価の考慮事項)には、ユーザビリティテストがリスクの管理を支援するのに役立つ場合を治験依頼者が理解するのに役立つ推奨事項が含まれています。付録には、ヒューマンファクターの検証に加えて、またはヒューマンファクターの検証が必要ない場合に、治験依頼者が特定の種類のユーザビリティテストを開発して説明するのに役立つ推奨事項も含まれています。付録は、デバイス固有の推奨事項と、該当する場合は[ヒューマンファクターガイダンス](#)に記載されている推奨事項を補足します。パフォーマンス検証とヒューマンファクターの検証(または必要に応じてユーザビリティの評価)は、デバイスが実際の環境でどのように使用され、実行され

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

るかを理解するための情報を FDA に提供するのに役立ちます。パフォーマンス検証では、テスト条件下でモデルの統計的パフォーマンスを評価するために、さまざまなテストおよびモニタリング方法を採用することがあります。ヒューマンファクターの検証テストには、さまざまなユーザーが状況に応じてデバイスをどのように使用する可能性があるかを理解することが含まれます。言い換えると、パフォーマンス検証は、デバイスの仕様がユーザーのニーズと意図した用途に準拠しており、実装されているパフォーマンス要件を一貫して満たすことができることを確認することを目的としています。一方、ヒューマンファクターの検証とユーザビリティの評価は、意図したすべてのユーザーがデバイス使用中に特定の目標を達成できるかどうか、およびユーザーが一貫してデバイスと安全かつ効果的に対話できるかどうか具体的に対処することを目的としています。

#### ソフトウェアバージョン履歴

市販前ソフトウェアガイダンスのセクション VI.I (Section VI. I) では、ソフトウェアバージョン履歴に関する情報を含め、市販申請のソフトウェア説明の一部として含める必要がある情報について説明しています。AI 対応デバイスの場合、ソフトウェアバージョン履歴には、モデルのバージョン、モデルのテストされたバージョンとリリースされたバージョンの違い、およびデバイスの安全性と有効性に対する違いの潜在的な影響の評価が含まれます。FDA は、すべての検証活動が客観的であり、FDA の同意なしにテストデータ(事後調整)に照らしてモデルが便宜的に調整されていないことを確認するために、どのバージョンのモデルがテストされたかを理解することが重要です。

新しいバージョンやモデルがある場合にラベルに UDI を表示する必要があるデバイス、および新しいデバイスパッケージには、新しい一意のデバイス識別子 (UDI) が必要です。<sup>50</sup>

## **A. パフォーマンスの検証**

AI 対応デバイスの申請に含める必要がある理由 :AI 対応デバイスのパフォーマンスの検証では、デバイスがその使用目的に応じてターゲット母集団で予測可能かつ確実に動作するという客観的な証拠が提供されます。以下の推奨事項は、デバイス固有の推奨事項や、「医療機器の重要な臨床試験の設計に関する考慮事項」、「診断テストを評価する研究の結果報告に関する統計的ガイダンス」、「臨床試験における電子ソースデータ」など、該当する場合に他の FDA ガイダンスで提供されている推奨事項を補足することを目的としています。

FDA によるデバイスの安全性と有効性の評価の一環として、FDA は、対象となる使用集団および対象となるサブグループにおいて、デバイスが全体的にどのように動作するかを理解することが重要です。特定のサブグループで許容されるパフォーマンスは、評価が全集団に対してのみ行われる場合、他のサブグループの低いパフォーマンスを覆い隠す可能性があります。特定のサブグループでのパフォーマンスが低いと、デバイスがそれらのグループでの使用に安全でなくなり、対象となる使用集団の潜在的な範囲に影響を与える可能性があります。セクション VIII (Section VIII: データ管理) では、FDA による安全性と有効性の評価において、対象となるサブグループの層別化と分析が重要である理由について概説します。また、想定される使用対象集団全体での安全かつ有効な使用をサポートするサブグループのパフォーマンスの分析は、デバイスが対象となるすべての患者に使用できることを保証するのに役立ちます。

<sup>50</sup> See 21 CFR 830.50.

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

サブグループ間のパフォーマンスの差は AI 対応デバイスに固有のものではありませんが、大量のデータから学習した関係にモデルが依存していること、およびユーザーに対するモデルの相対的な不透明性により、AI 対応デバイスはパフォーマンスの予期しない差の影響を特に受けやすくなります。モデルの開発に使用されるデータがトレーニング中に代表的なものであっても、モデルが過剰にトレーニングされて、研究データセットの特定の特性に固有のデータの特徴を認識する可能性があります。病気や状態の識別または治療には誤っている可能性があります。誤った学習は、特に異なるグループの研究参加者からのデータが異なるサイトで収集される傾向がある場合に、病気のサブタイプや患者の人口統計学などの関心のある特性間でパフォーマンスに異なる影響を与える可能性があります。たとえば、モデルでは、ある人口統計学的情報または人口統計学的情報と相関する別の変数を、モデルの関心のある変数として誤って使用する可能性があります。これは、トレーニングデータセットでは、ある人口統計学的情報の患者の方が病気にかかっている可能性が高いためです。これは、関心のある変数が人間に理解できない可能性がある複雑なモデルでは、特定が特に困難になる可能性があります。このため、AI 対応デバイスの評価の精度と有用性は、テストに使用されるデータの質、多様性、および量にも依存します。

サブグループ分析は、特定の集団におけるデバイスのパフォーマンスを評価するためのツールを提供し、デバイスのパフォーマンスが全体的なパフォーマンスよりも低いシナリオを特定するのに役立ちます。さらに、サブグループ分析は、デバイスの潜在的な制限を特定するのに役立ち、エンドユーザーに追加の有用な情報を提供することで、効果的なラベリングに貢献できます。

デバイス出力の不確実性に関する情報は、審査官がデバイス出力をどのように解釈するかを理解するのに役立つため、重要です。法令、規制、またはガイダンスでデバイスの種類が指定されていない場合でも、再現性および/または再現性の研究は、提供された場合にデバイス出力に関連する不確実性を FDA が理解し、定量化するのに役立ちます。

このガイダンスの付録 C (Appendix C: パフォーマンス検証の考慮事項) には、パフォーマンス検証の一般的なアプローチに関する追加の推奨事項が含まれています。さらに、FDA は治験依頼者に対し、AI 対応デバイスの開発と検証に対する提案されたアプローチに関する FDA のフィードバックを得るために、[Q 提出プログラム](#)を活用することを奨励します。特に、RWD の使用、新規および新興の研究方法の使用、または新技術の検証について議論する際に、早期の関与が役立つ可能性があります。

#### ヒューマンデバイスチームのパフォーマンスの評価

治験依頼者は、パフォーマンステストの適切な方法を特定する際に、ユーザーとデバイス間の相互作用を考慮することが重要です。このドキュメントでは、「[Good Machine Learning Practice: Guiding Principles](#)」に焦点を当てることについて説明しています。この原則では、モデルに「単独でのモデルのパフォーマンスではなく、人間と AI のチーム」がある場合に、「人間」のパフォーマンスを理解することが重要であることを説明しています。AI 対応デバイスの意図された使用と臨床ワークフローは、より自律的なシステムから、特定のユーザーを支援するが、AI の出力を解釈し、最終的に臨床的な意思決定を行うために人間に依存する支援（援助）ツールまで、一連の意思決定の役割にわたっています。



***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

デバイスがこのスペクトルに沿って移動するにつれて、AI ベースの医療デバイスのパフォーマンス評価をサポートするために適切な臨床研究またはその他の研究(例:ヒト因子バリデーション試験)の性質は、モデルの意図された使用によって異なります。デバイスによっては、モデルのスタンドアロンのパフォーマンス(つまり、実際の出力は期待された出力と一致しましたか?) に重点が置かれる場合があります。また、単独でのモデルのパフォーマンスだけでなく、人間と AI のチームのパフォーマンスを評価することに重点を置く場合もあります(すなわち、新しいデバイスを使用して作業する予定のユーザーは、オペレーター単独または別のデバイスを使用した場合と同じか、またはそれ以上のパフォーマンスを発揮しましたか?)。治験依頼者は、特定のシナリオでは、スタンドアロンと人間とデバイスのチームのパフォーマンス評価の両方が、AI 対応デバイスの全体的なパフォーマンス評価をサポートする可能性があることを考慮する必要があります。

AI ベースの医療画像分析システムのパフォーマンス評価は、デバイスの使用目的が人間とデバイスの相互作用の範囲に沿って変化するにつれて、臨床研究のアプローチがどのように変化するかを示す例です。スタンドアロンの評価では、人間との相互作用とは無関係にモデルのパフォーマンスが測定されます。一方、リーダーの研究では、AI 対応デバイスを使用している場合と使用していない場合の両方の対象ユーザーのパフォーマンスが比較されます(すなわち、人間対人間デバイスチームのパフォーマンスの比較)。<sup>51</sup>リーダーの研究は、通常、医療画像アプリケーションの臨床的意思決定に役立つ AI 対応デバイスの主要なパフォーマンス評価として機能します。これは、治験依頼者が対象ユーザーの手の中でツールの臨床的利点を評価できるためです。<sup>52</sup>

治験依頼者が申請に含める必要がある内容: 検証テストでは、目的の使用に関してモデルのパフォーマンスを特徴付けるための客観的な情報を提供する必要があります。検証では、独立したデータセットでモデルのパフォーマンスを評価します。必要に応じて、これらの変更に関連するリスクに基づいて、入力データと使用条件の合理的に予測可能な変更を予測するためのモデルの堅牢性の評価も含める必要があります。

検証方法は、デバイスの目的の使用によって異なります。例:

- 一般に認められている参照方法で実行される定義済みの測定値を推定するデバイスでは、再現性と再現性を適切に評価するために精密検査が必要になる場合があります。
- 患者の時系列データをモニタリングし、定期的な再校正が必要な機器は、動的応答を評価するために安定性試験および変化追跡試験が必要となる場合があります。
- 明確に定義されていない患者パラメータを測定する調査機器に類似した機器は、構成概念妥当性の追加的な証拠(すなわち、テストが測定しようとしているものをどの程度測定しますか)が必要となる場合があります。
- 予後臨床意思決定支援機器は、特に生存分析、校正分析、および/または識別分析(リスク層別化分析など)による縦断的データを必要とする場合があります。

<sup>51</sup> For more information, see FDA's guidance titled, "[Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data Premarket Notification \[510\(k\)\] Submissions.](#)"

<sup>52</sup> For more information on computer-assisted detection devices, please see FDA guidance titled, "[Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data in Premarket Notification \(510\(k\)\) Submissions.](#)"

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

特定の AI 対応デバイスに応じて、このエビデンスは非臨床ベンチまたは分析研究、前臨床動物試験、臨床パフォーマンス研究、臨床アウトカム研究、またはそれらの組み合わせから得られる可能性があります。

### 研究プロトコル

パフォーマンス検証をサポートするために、治験依頼者は統計分析計画を含むすべての研究プロトコルに関する情報を含める必要があります。統計解析計画書には、試験デザイン及び解析の詳細を記載すること。これらの文書に記載すべき重要な事項は以下のとおりです。

- 以下を含む試験デザインの詳細。
  - 試験デザインの説明(例えば、重要な臨床成績の指標を実証するのに十分な統計的検出力を有する前向き比較試験デザイン)。
    - 前向き試験の場合、従う手順と方法、これらの手順と方法に關与するオペレーターの説明、使用するツールまたは機器。
    - レトロスペクティブ試験の場合、アーカイブされたデータまたは材料の取り扱い、準備、処理、および選択方法に関する計画。
  - 特定の患者の試験中に使用された AI 対応機器のバージョンまたは状態を記録するために使用されるデータ記録メカニズムの説明。
    - 正確性を確保するために、電子症例報告書 (eCRF) <sup>53</sup>または電子データキャプチャ (EDC) システムで実装されたこれらのデータの自動収集が適切な場合があります。
  - 必要に応じて、臨床参照基準の決定プロセスからの機器出力の盲検化、機器出力のユーザー/インタープリターからの臨床参照基準のマスキング、およびモデル開発者と臨床チームからの試験データのマスキング (試験デザインにおける日和見的な調整やバイアスを回避するため) の手順と方法の説明。
  - 試験中に AI 対応機器が患者またはユーザーにもたらすリスクに対処するために実施されているコントロールの説明。
  - 臨床試験の実施中に治験実施計画書が変更された場合には、申請者は、その変更について説明し、軽微な変更と重大な変更を特定し、反復試験又は事前に規定された計画から逸脱した試験について、その妥当性を示すこと。治験実施計画書に従うべきであり、軽微とみなされるものを含め、あらゆる種類の治験実施計画書からの逸脱を最小限に抑えるべきです。
  - 登録されたすべての被験者の完全な説明 (説明表を含む)。
  - 試験集団のベースライン分布及びデータセット内の他の重要な因子 (データ収集装置、装置構成、疾患の状態又は状態など) の説明、及びそれらの代表性の正当性。AI 対応医療機器における代表性の詳細については、本ガイダンスのセクション VIII (データ管理) を参照のこと。
- 以下を含む統計解析計画:

<sup>53</sup> For more information, see FDA guidance titled “Electronic Source Data in Clinical Investigations.”

***Contains Nonbinding Recommendations  
Draft – Not for Implementation***

- 主要評価項目またはアウトカムの説明。これは試験の主要目的を反映したものでなければなりません。
- 臨床的に正当化される複数の主要評価項目(パフォーマンス目標など)のそれぞれに関して、事前に規定された試験の成功/失敗基準(例えば、文献または以前の研究によって裏付けられています)。
- 帰無仮説及び代替(作業)仮説などの統計的仮説の説明。
- 十分な調査力を確保するためのサンプルサイズの妥当性。
- サンプルサイズの計算を正当化するための情報を含む、主要評価項目の統計的分析の説明。
- 多重度の問題に対処し、全体的なタイプ I エラー率を制御するための、事前に規定された適切な統計的アプローチの説明。
- 適切な統計的方法論の説明。
- サブグループ分析計画。
  - 適切なサブグループは、デバイスの使用目的によって通知されますが、一般的には、患者の性別、<sup>54</sup> 性別、年齢、人種、民族、<sup>55</sup> 疾患変数、臨床データサイト、データ収集装置
  - (カメラのブランドなど)、および該当する場合は、使用条件(該当する場合はユーザーのスキルレベルを含む)、デバイスの構成、およびデバイスのパフォーマンスに影響を与える可能性のあるその他の関連する交絡因子を含める必要があります。
  - サブグループに関して特定のパフォーマンスを主張する場合、サブグループ分析は、適切に検出されたサブグループを含めることを含め、統計的に有意である必要があります。ただし、特定の
  - サブグループのパフォーマンスに関する主張が行われていない場合、サブグループのパフォーマンスは各サブグループに対して統計的に検出される必要はありませんが、報告された結果に意味と文脈があるように、各サブグループに妥当な数の患者を含めるように努力する必要があります。

## 研究結果

パフォーマンスの検証をサポートするために、治験依頼者は研究結果に関する情報を含める必要があります。これらの文書で取り上げるべき重要な側面は次のとおりです。

- サブグループ分析を含む、各テストについて事前に指定された結果の説明。
- 上記のように、関連するサブグループの適切なサブグループ分析を含む結果の説明。
  - 研究データに人口統計学的情報が利用できない場合は、利用できない理由、人口統計学的サブグループ分析なしでパフォーマンス評価をサポートできる理由、および人口統計学的サブグループ分析の欠如に関連するリスクがどのように管理されているかの説明。
- 実行可能で適切な場合は、デバイスの再現性と再現性の評価。これらの研究がどのように実施されるかの詳細は、評価される特定のデバイスによって異なり、ファントム、シミュレーション、意図的なものが含まれる可能性があります。または臨床データ。

<sup>54</sup> For more information on sex-specific data, please see FDA guidance titled "[Evaluation of Sex-Specific Data in Medical Device Clinical Studies.](#)"

<sup>55</sup> For more information on age-, race-, and ethnicity-specific data, please see FDA guidance titled "[Evaluation and Reporting of Age-, Race-, and Ethnicity-Specific Data in Medical Device Clinical Studies.](#)"

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

治験依頼者が提出する場所：デバイスの非臨床または臨床試験に関する情報は、販売申請提出書類の適切なセクションに含める必要があります。たとえば、臨床試験の結果は、販売申請提出書類の臨床セクションに含める必要があります。モデルのソフトウェア検証及びソフトウェアバリデーションに関する情報は、[市販前ソフトウェアガイダンス](#)に記載されているとおり、販売承認申請書のソフトウェア文書の「**検証及びバリデーションの一部としてのソフトウェア試験**」に記載すること。

**追加資料:**付録 C (パフォーマンス検証に関する考慮事項) には、パフォーマンス検証試験とそのデータの開発と分析に役立つ推奨事項が記載されています。付録 D (Appendix D: ユーザビリティ評価に関する考慮事項) には、スポンサーが AI 対応デバイスの申請に関するユーザビリティリスクコントロールを評価するのに役立つ情報が記載されています。

FDA は、スポンサーが [AI 対応デバイスの開発及びバリデーションに関する提案されたアプローチについて FDA のフィードバックを得るために Q-Submission Program](#) を利用することを推奨しています。実世界のエビデンスを使用する場合、スポンサーは「[医療機器の規制上の意思決定を支援するための実世界のエビデンスの使用](#)」というタイトルの FDA ガイダンスを参照することもできます。

## **XI. デバイスのパフォーマンスモニタリング**

AI 対応デバイスの申請に含める必要がある理由: 実際の環境 (すなわち、承認または認可後に市販された AI 対応機器) に展開されている AI 対応デバイスのパフォーマンスは、時間の経過とともに変化または低下し、患者にリスクをもたらす可能性があります。一般に、AI 対応デバイスを含む医療機器の品質システムの一部として、製造業者は、市販後の設定でのパフォーマンスの変化を特定して対応するために、市販後のパフォーマンスモニタリング計画を持つ必要があります。製造販売承認申請にパフォーマンスモニタリング計画を含めることは、不確実性を軽減し、FDA によるリスク管理の評価を支援するのに役立つ可能性があります。

AI 対応デバイスの継続的な管理の一環として、製造業者は、デバイスのパフォーマンスの変化だけでなく、デバイスのパフォーマンスの変化につながる可能性のあるデバイス入力およびデバイスが使用される状況の変化を事前に監視し、特定し、対処する必要があります。さらに、治験依頼者は、AI 対応デバイスのデバイスパフォーマンスの望ましくない変化に関連するリスクを管理するために、品質システム規則 (21 CFR Part 820) に準拠した包括的なリスク分析プログラムと文書化の計画を策定および実装する必要があります。<sup>56</sup> これらの規則には、管理責任 (21 CFR 820.20)、設計検証 (21 CFR 820.30 (g) )、設計変更 (21 CFR 820.30 (i) )、不適合製品 (21 CFR 820.90)、是正措置および予防措置 (21 CFR 820.100) が含まれますが、これらに限定されません。さらに、製造業者は、21 CFR Part 803 および 806 に従って、デバイスのパフォーマンスを監視し、死亡、重傷、および不具合に関する情報を FDA に報告しなければなりません。

<sup>56</sup> When the final rule amending the device QSR, 21 CFR Part 820, takes effect on February 2, 2026, the term “risk analysis” will be replaced with “risk management.”

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

FDA は通常、FD&C 法のセクション 510 (k) に基づく販売申請の審査の一環として、品質システム規制への準拠を評価しません。ただし、適切な継続的パフォーマンスを確保するために、FDA が販売申請における治験依頼者の品質システムの詳細を審査することが適切な場合もあります。このような審査は、実質的な同等性の決定を支援することができます。

AI 対応デバイスでは、継続的なパフォーマンス監視が重要です。これは、前述のように、モデルがトレーニングに使用されるデータの特性に大きく依存しているため、そのパフォーマンスがデータ入力の変化に特に敏感になる可能性があるためです。デバイスのパフォーマンスの変化は、患者数の経時的な変化、疾患パターン、他の変更によるデータのずれなど、多くの要因によって発生する可能性があります。パフォーマンスの変化が発生した場合、たとえば、デバイスが高度に自動化されたプロセスの一部であり、継続的な人間の介入が制限されている場合、または出力が予測的なものであり、デバイスの使用や患者との確認的なフォローアップのやり取りに異なる医療従事者が関与する可能性がある場合は、ユーザーが AI 対応デバイスでパフォーマンスの変化を特定する可能性が低くなります。AI 対応デバイスのパフォーマンスは、使用が承認または許可されている環境の側面が時間の経過とともに変化する可能性があるため、市販前（デバイスの承認と導入の前）に実施される開発およびテスト活動でリスクを完全に制御することはできない可能性があります。

FDA は、医療デバイスが導入される環境をデバイスメーカーが完全に制御することはできないことを認識しています。さらに、デバイスのパフォーマンスの変化につながる可能性のある要因の存在が、常に患者への害に関する懸念を引き起こすとは限りません。むしろ、継続的なリスク管理の一環として、デバイスメーカーがこれらの要因（例えばデータドリフト）がデバイスの安全性と有効性に及ぼす影響を考慮することが重要です。パフォーマンス管理プロセスに関する追加情報は、FDA がリスクを適切に特定され、対処され、管理されているかどうかを判断するのに役立つ可能性があります。

治験依頼者が申請に含めるべき内容: リスク管理の手段としてプロアクティブなパフォーマンスモニタリングを採用し、デバイスの安全性と有効性を合理的に保証することを選択した AI 対応デバイスの治験依頼者は、市販前申請の一部としてパフォーマンスモニタリング計画に関する情報を含める必要があります。治験依頼者は、[Q-申請プログラム](#)を通じて計画に関する FDA のフィードバックを得ることをお勧めします。510 (k) 申請の場合、FDA は通常、特定の状況を除いて、パフォーマンスモニタリング計画が特定のデバイスタイプ(すなわち、適用される分類規則)に対する特別な管理ではないデバイスについては、そのような計画を要求しません。De Novo 分類要求の場合、このような計画は、特定のデバイスの種類によってもたらされるリスクを管理するために必要になる可能性があります。そのため、FDA は今後、デバイスの種類に対して特別な管理を確立することができます。PMA の場合、パフォーマンス監視計画が承認の条件になることがあります。<sup>57</sup>ただし、治験依頼者は、AI 対応デバイスの申請にパフォーマンス監視計画に関する情報を含めることを選択できます。

---

<sup>57</sup> See 21 CFR 814.44 and 21 CFR 814.82.

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

パフォーマンス監視計画では、パフォーマンスの変更または低下につながる可能性のあるパフォーマンスの変更または条件を適時に特定し、対応する必要があります。堅牢なパフォーマンス監視計画には、展開後にデバイスのパフォーマンスをキャプチャするための積極的な取り組みが含まれます。このような計画のコンポーネントには、次のものが含まれます。

- 次のためのデータ収集および分析方法の説明。
  - 安全性と有効性に関するパフォーマンス監視の結果の評価を含む、モデルパフォーマンスの変更の特定、特徴付け、および評価。
  - 次のような、パフォーマンスの望ましくない変更の潜在的な原因の監視。
    - 患者の人口統計や疾患の有病率の変化;
    - 入力データの変化;
    - データパイプラインの破損 (入力データの整合性) による入力データの変化 (値の欠落、重複レコード、データ型の不一致など);および
    - ユーザーの行動またはユーザーの人口統計の変化。
- 展開環境での監視メカニズムを含む、堅牢なソフトウェアライフサイクルプロセスの説明。
- パフォーマンスの変化にタイムリーに対処する更新、緩和、および修正措置を展開するための計画。
- FDA は、パフォーマンスの変化に対処するために実施される措置の中には、実施前にマーケティングへの提出または承認 (21 CFR 807.81 (a) (3) および 21 CFR 814.39 (a) ) を必要としないものがあることに注意してください。特定の変更が FDA への市販前の提出を必要とするかどうかを評価するには、[「既存のデバイスへの変更のために 510 \(k\) を提出するタイミングの決定」](#)および[「既存のデバイスへのソフトウェア変更のために 510 \(k\) を提出するタイミングの決定」](#)というタイトルの FDA ガイダンスを参照してください。治験依頼者は、必要に応じて PCCP の使用を検討することもできます。<sup>58</sup>
  - 本計画は、21 CFR Part 803 および 806 に基づく特定の有害事象、修正および削除に関する情報を FDA に報告する要件を含む、適用される法令または規制要件に代わるものではありません。
- パフォーマンスモニタリングの結果および軽減策をデバイスユーザーに伝達するための手順の説明。

治験依頼者が提出する必要がある場合:適切な場合、デバイスパフォーマンスモニタリング計画は、販売申請提出書類の「ソフトウェアドキュメント」セクションの「**リスク管理ファイル**」に含める必要があります。

## **XII. サイバーセキュリティ**

AI 対応機器の申請に含めるべき理由:医療機器に組み込まれたあらゆるデジタルコンポーネントまたはソフトウェアコンポーネントと同様に、AI はサイバーセキュリティリスクをもたらす可能性があります。サイバーセキュリティの設計と維持に関する FDA の一般的な推奨事項および関連する販売申請文書は、[「医療機器におけるサイバーセキュリティ:品質システムの考慮事項と市販前申請の内容」](#)と題するガイダンス文書 (以下、[2023 年市販前サイバーセキュリティガイダンス](#)) に記載されています。[2023 年市販前サイバーセキュリティガイダンス](#)では、AI 対応機器を含む医療機器に関連する可能性のあるセキュリティ目標を特定しています。;認可 ;可用性; 機密性; およびセキュアでタイムリーな更新性とパッチ適用性。

<sup>58</sup> See FDA's guidance titled "Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions."

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

FD&C 法のセクション 524B (c) に基づく「サイバーデバイス」の定義を満たす AI 対応デバイスの場合、ガイダンスのこのセクションの推奨事項は、製造業者が FD&C 法のセクション 524B に基づく義務を果たすのに役立つことを目的としています。サイバーセキュリティの脅威によって影響を受ける可能性のある AI リスクの例には、次のようなものがあります。

- データポイズニング：サイバー脅威は、不正なデータまたは悪意を持って変更されたデータを意図的に注入することによってデータポイズニングを引き起こし、医療診断などの分野で結果を危険にさらす可能性があります。
- モデルの逆転/盗用：サイバー脅威は、意図的に偽造または変更されたデータを使用して、モデルから詳細を推測したり、モデルを複製したりする可能性があります。これらは、知的財産やプライバシーの侵害だけでなく、モデルのパフォーマンスの継続にもリスクをもたらします。
- モデルの回避：サイバー攻撃者は、意図的に入力サンプルを作成または変更してモデルを欺き、誤った分類に導く可能性があります。これらは、モデル予測の信頼性と整合性にリスクをもたらし、AI 対応デバイスの信頼性が損なわれ、悪意のある悪用にさらされる可能性があります。
- データの漏洩：サイバー脅威は、脆弱性を悪用して、モデル内の機密性の高いトレーニングデータや推論データにアクセスする可能性があります。
- 過剰適合：サイバー脅威は、意図的にモデルを「過剰適合」させ、AI コンポーネントを敵対的な攻撃にさらす可能性があります。これらのコンポーネントは、変更された患者データに効果的に適応するのに苦労しています。
- モデルバイアス：サイバー脅威は、バイアスを導入または強調するためのトレーニングデータの操作につながる可能性があります。敵対的な例を使用して既知のバイアスを悪用したり、トレーニング中にバックドアを埋め込んでバイアスのある動作を後でトリガーしたり、固有のバイアスを持つ事前トレーニング済みモデルを活用して、歪んだ微調整データで増幅したりする可能性があります。
- パフォーマンスのドリフト：サイバー脅威は、基になるデータ分布を変更することによってモデルのパフォーマンスのドリフトにつながる可能性があります。これにより、モデルのパフォーマンスが低下します。サイバー脅威は、時間の経過とともに入力データをわずかにシフトさせたり、動的環境の脆弱性を悪用したりして、モデルが不正確な予測を行ったり、敵対的な攻撃を受けやすくなったりする可能性があります。

治験依頼者が提出に含める必要がある内容：AI コンポーネントまたは機能に関連するサイバーセキュリティのコントロールとセキュリティリスク管理に関する [2023 年のプレマーケットサイバーセキュリティガイダンス](#) で推奨されている提出ドキュメントに沿って、治験依頼者は次の種類の情報を含める必要があります。

- サイバーセキュリティリスク管理レポート、脅威モデリング、サイバーセキュリティリスク評価、ラベリング、およびその他の成果物の追加要素（AI サイバーセキュリティに関連する独自の考慮事項がある場合）。
- モデルに関連するリスクに対処するためにサイバーセキュリティテストがどのように適切であるかについての説明。少なくとも次のテストが含まれます。
  - 不正な入力（ファジー）テスト;および
  - 侵入テスト。
- デバイスの AI 対応の考慮事項をカバーするセキュリティユースケースビュー。
- データの脆弱性に対処し、データ漏洩を防止するために実装された、次のようなコントロールの説明。

**Contains Nonbinding Recommendations**  
**Draft – Not for Implementation**

- アクセスコントロール;
- データの暗号化;および
- 機密データの匿名化または非識別化。

治験依頼者は、上記の特定のリスクに対処する方法について、[2023 年市販前サイバーセキュリティガイダンス](#)の付録 1 の管理推奨事項を参照してください。AI 対応デバイスに関連するサイバーセキュリティリスクを制御するアプローチの例には、次のものがあります。

- データポイズニング攻撃については、次のことを考慮してください。
  - データの検証、認証、およびクレンジング。
  - 異常検出とデータ整合性チェック(暗号ハッシュなど)の採用。
  - 敵対的トレーニングの適用。これは、モデルの堅牢性とセキュリティを向上させるために使用される方法です。
- 偽造データを使用してオーバーフィッティング、モデルバイアスなどを導入するサイバー脅威については、次のことを検討してください。
  - 差分プライバシーの採用。データセット内の個々のデータポイントのプライバシーを保護する手法です。差分プライバシーを利用する場合、治験依頼者はプライバシーとモデルの精度、実用性、効率性などの要因との間の潜在的なトレードオフを認識し、そのトレードオフに対処する方法に関する情報を提供する必要があります。
  - セキュリティで保護されたマルチパーティ計算 (MPC) の使用。これは、複数のパーティがローカルデータセットを互いに公開することなく、モデルを共同でトレーニングできるようにする手法です。
  - データ認証と整合性保護の採用。
  - 所有権を証明するために AI モデルに非表示の透かしを埋め込むウォーターマークの導入。
  - 継続的なモデルパフォーマンス監視の適用。
- モデル回避については、モデルの堅牢性を高めるために敵対的トレーニングを検討し、データが予想されるパターンに準拠していることを確認するために厳格な入力検証チェックを実装します。敵対的トレーニング手法を展開する場合、治験依頼者は、攻撃に対する堅牢性の強化とモデルパフォーマンスへの潜在的な悪影響(精度など)の間に生じる可能性のあるトレードオフを認識し、トレードオフを管理する方法に関する情報を提供する必要があります。

治験依頼者が申請で提供する必要がある場所:サイバーセキュリティ情報は、[2023 プレマーケットサイバーセキュリティガイダンス](#)に記載されているように、販売申請の「**サイバーセキュリティ/相互運用性**」セクションに含める必要があります。

**追加リソース:** 治験依頼者は、サイバーセキュリティに関連する追加の推奨事項について、他の FDA ガイダンス文書を参照することもできます。

- [医療機器におけるサイバーセキュリティ:品質システムに関する考慮事項と市販前申請](#)の内容
- 医療機器におけるサイバーセキュリティの市販後管理
- 既製 (OTS) ソフトウェアを含むネットワーク化された医療機器のサイバーセキュリティ
- 医療機器における市販ソフトウェアの使用



*Contains Nonbinding Recommendations  
Draft – Not for Implementation*

### XIII. 公開申請の概要

AI 対応デバイスの申請に含めるべき理由:透明性は市販前承認の重要な要素であり、患者ケアにとって重要です。これは、高度にデータ駆動型で、ある程度の不透明性を示すアルゴリズムが組み込まれている AI 対応デバイスにとって特に重要です。2021 年 10 月 14 日に開催された AI 対応デバイスの透明性に関するオンライン公開ワークショップ「[Transparency of Artificial Intelligence/Machine Learning-enabled Medical Devices \(人工知能/機械学習対応医療デバイスの透明性\)](#)」を含む公開ワークショップとコメントでは、患者はケアにおける AI の使用に関する懸念を指摘しました。一般の人々は一貫して、FDA が AI 対応デバイスに関する承認決定を行う方法に関する追加情報、およびこれらのデバイスの設計と検証に関するより多くの情報を求めてきました。公開提出要約には、透明性をサポートするために、これらのデバイスの特徴を説明する具体的な情報を含める必要があります。これは、AI 対応デバイスに対する理解を深め、一般の信頼を醸成することによって公衆衛生に貢献することができます。

公開提出要約 (Public submission summaries) は、ほとんどの販売承認決定に必要であり、FDA のウェブサイトで入手できます <sup>59</sup>。公開要約が必要な場合は、FDA による実質的な同等性の決定、またはデバイスの安全性と有効性の合理的な保証のユーザーに対する透明性をサポートするために、一般向けの文書に AI 対応デバイスに関する詳細を十分に詳細に含める必要があります <sup>60, 61, 62</sup> 重要な情報への一般のアクセスを確保するために承認された AI 対応デバイスに関する情報については、このセクションでは、治験依頼者が公開申請要約に含めるべき情報の種類と、そのような情報の可能な形式について説明します。

PMA、HDE、De Novo、BLA、または 510 (k) の経路を通じて提出される AI 対応デバイスについて、FDA は、このセクションで説明する情報を、該当する公開申請要約、または 510 (k) 要約 (21 CFR 807.92 (a) (4) に従って作成されたセクション) に含めることを推奨します。治験依頼者は、患者の識別情報、企業秘密、および商業上の機密情報を除いた推奨情報を提供する必要があります。510 (k) ステートメント (21 CFR 807.93) を提出する治験依頼者に対して、FDA は、患者の識別情報、企業秘密、および商業上の機密情報を除いた同じ情報を提出することを推奨しています。<sup>63</sup>

<sup>59</sup> See FDA's website titled "[CDRH Transparency: Premarket Submissions](#)." See 21 CFR 807.92 for requirements on the form and content of a 510(k) Summary. See 21 CFR 807.93 for requirements on the content and format of a 510(k) Statement. See 21 CFR 814.9(e) for requirements on a PMA decision summary.

<sup>60</sup> In accordance with 21 CFR 807.92, "a 510(k) summary shall be in sufficient detail to provide an understanding of the basis for a determination of substantial equivalence." See 21 CFR 807.92 for requirements on the content and format of a 510(k) Summary. If a sponsor chooses to submit a 510(k) Statement rather than 510(k) Summary, the sponsor should provide information that supports FDA's determination of substantial equivalence. See 21 CFR 807.93 for requirements on the content and format of a 510(k) Statement.

<sup>61</sup> In accordance with 21 CFR 814.9(e), "FDA will make available to the public ... a detailed summary of information submitted to FDA respecting the safety and effectiveness of the device that is the subject of the PMA and that is the basis for the order." See 21 CFR 814.9(e) for requirements on a PMA decision summary.

<sup>62</sup> The De Novo decision summary is intended to present an objective and balanced summary of the scientific evidence that served as the basis for the FDA's decision to grant a De Novo request. For more information on De Novo decision summary documents, please see FDA's website on [De Novo Classification Request](#).

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

必須ではありませんが、モデルカードの使用は、AI 対応デバイスに関する情報を伝達する 1 つの方法である可能性があります。なぜなら、モデルカードは AI 対応デバイスの主要な側面を一貫して要約する手段であり、その特徴、パフォーマンス、および制限事項を簡潔に説明するために使用できるからです。付録 E (Appendix E: モデルカードの例) には、モデルカードの内容と書式に関する推奨事項が記載されています。研究では、モデルカードの使用がユーザーの信頼と理解を高めることができることが実証されています。特に、公開申請の概要の一部としてモデルカードを使用することは、一般の利害関係者や、患者、臨床医、規制当局、研究者などのユーザーに対して、AI 対応デバイスに関する明確で一貫したコミュニケーションをサポートする 1 つの方法です。モデルカードを使用すると、AI 対応デバイスに関する重要な情報を伝達するための最適なアプローチを決定することに関連する課題に対処できます。

治験依頼者が申請書に記載すべき事項： 治験依頼者は、それぞれの申請に関する申請規則を遵守する必要があります<sup>64</sup>。さらに、治験依頼者は、関連する販売申請の種類に関する FDA の推奨事項を考慮する必要があります。治験依頼者は、患者の識別情報、企業秘密、および機密の商業情報を除く、以下の種類の情報も提供する必要があります。

- デバイスで AI が使用されているという記述;
- デバイスの使用目的の一部として AI がどのように使用されているかの説明。複数の機能を持つデバイスの場合、この説明には、AI-DSF が互いにどのように相互作用するか、および AI-DSF 以外の DSF とどのように相互作用するかを含めることができます。
- モデルのクラス(例えば、畳み込みニューラルネットワーク、リカレントニューラルネットワーク、サポートベクターマシン、トランスフォーマー)の説明と、デバイスの説明内のモデルの制限事項;
- トレーニングおよび検証データの人口統計学的特性に関する情報、および使用目的の母集団の人口統計学的特性に関する情報を含む、開発および検証データセット(サイズ、データソース)の説明。また、トレーニングデータセットと検証データセット、および使用目的で想定されるモデルデータ入力を比較する必要があります。比較では、トレーニングデータからのテストデータの独立性がどのように確保されたかを説明する必要があります。
- 必要に応じて、統計的信頼性と不確実性を説明するその他の説明またはメトリックを含む、予測の統計的信頼レベルの説明;および
- 該当する場合は、モデルが時間の経過と共に更新および維持される方法の説明。

治験依頼者は、情報を整理するためにモデルカードの使用を検討する必要があります。付録 E (Appendix E: モデルカードの例) には、モデルカードに含めることができる要素に関する推奨事項が含まれています。モデルカードの例には、モデルカードの推奨要素と形式が含まれていますが、治験依頼者は追加情報を含めることも、異なる形式に従うこともできます。モデルカードの構造がない場合でも、治験依頼者はモデルカードに含まれる情報を含めることを検討する必要があります。

<sup>63</sup> For more information, see FDA guidance, "[The 510\(k\) Program: Evaluating Substantial Equivalence in Premarket Notifications \[510\(k\)\]](#)."

<sup>64</sup> For more information regarding the requirements for PMA, see 21 CFR Part 814. For more information regarding the requirements for 510(k), see 21 CFR 807.81 – 807.100. For more information regarding the requirements for De Novo, see 21 CFR 860.200 – 860.260. For more information regarding the requirements for HDE, see 21 CFR 814.100 – 814.126. For more information regarding the requirements for BLA, see 21 CFR Part 600 – 680.

*Contains Nonbinding Recommendations  
Draft – Not for Implementation*

治験依頼者が提出する必要がある場所: 公開提出の概要は、マーケティング提出の「**管理文書**」セクションに含める必要があります。

**追加情報:** 本ガイダンスの付録 E (Appendix E: モデルカードの例) には、モデルカードの形式の一例が記載されています。本ガイダンスの付録 F (モデルカードを含む例 510 (k) の概要) には、完成したモデルカードを含む製品の公開提出の概要の例が記載されています。

DRAFT

*Contains Nonbinding Recommendations  
Draft – Not for Implementation*

**付録 A: 推奨されるドキュメントの表**

本ガイダンスのセクション V~XIII では、AI 対応機器の販売申請に含めることができる文書に関する推奨事項を示します。以下の表は、議論されている文書を提供するために、販売申請内で推奨される場所をまとめたものです。この文書を提出する方法の 1 つとして、eSTAR プログラムがあります。具体的には、eSTAR はインタラクティブな PDF フォームであり、申請者が包括的な医療機器申請を準備するプロセスを案内します。<sup>65</sup> eSTAR は無料であり、免除されない限り、すべての 510 (k) 申請に必要です。

ガイダンスセクションと推奨される情報	治験依頼者の販売申請の推奨セクション
セクションV デバイスの説明	デバイスの説明
セクションVI.A ユーザ	ソフトウェアの説明
セクションVI.B ラベリング	ラベリング
セクションVII リスク評価	ソフトウェアドキュメントのリスク管理ファイル
セクションVIII データ管理	開発用データ:ソフトウェアソフトウェアドキュメントの説明 テスト用:パフォーマンステスト
セクションIX モデルの説明とテスト	ソフトウェア
セクションX.A パフォーマンスの検証	臨床および非臨床テスト:パフォーマンステスト ソフトウェアの検証およびソフトウェアの検証:ソフトウェアドキュメントの検証および検証の一部としてのソフトウェアテスト
セクションXI デバイスパフォーマンスの監視	ソフトウェアドキュメントのリスク管理ファイル
セクションII サイバーセキュリティ	サイバーセキュリティ
セクションXIII 公開申請の概要	管理情報

<sup>65</sup> For more information on eSTAR, please see FDA's website on [eSTAR Program](#).

## 付録 B: Transparency\_Design の考慮事項

この付録には、ユーザーを中心とした透明なデバイスを開発するための推奨事項が含まれています。これらの推奨事項は、治験依頼者が安全で効果的な医療機器と高品質の販売申請を開発するのに役立つことを目的としています。治験依頼者は、FDA の安全性と有効性の評価をサポートする代替アプローチを特定することができますが、販売申請をサポートする情報の可用性を確保するために、TPLC の設計段階から透明性の考慮事項を統合する必要があります。デバイスの変更追加のテストが必要になる可能性がある場合、TPLC の後の段階でデバイスに透明性を統合することは困難な場合があります。このガイダンスでは、透明性とは、適切な利害関係者が理解して行動できる方法で、デバイスの状況に応じた性能と設計情報を明確に伝えることを指します。透明性とは、重要な情報がアクセス可能であり、機能的に理解可能であり、情報の共有とデバイスのユーザービリティの両方に関連していることを保証することです。そのため、透明性設計に対するユーザー中心のアプローチは、AI 対応デバイスの安全かつ効果的な使用をサポートするのに役立ちます。適切な透明性情報を含めることで、デバイスの使用意欲が 2 倍以上になることも示されています。

### TPLC 全体の設計による透明性

治験依頼者は、どのような情報を伝達すべきかを決定する際に、デバイスの使用に関連するコンテキスト要因と、それらの要因がデバイスのパフォーマンスにどのように影響するかを特定するために、総合的なアプローチを取る必要があります。治験依頼者は、AI 対応デバイスの使用、メンテナンス、および使用停止を通じた実装の完全な連続全体を通じて透明性を考慮し、最初から透明性を考慮してデバイスを設計する必要があります。

ユーザインターフェースは、必要に応じて透明性の原則を使用する必要があるもう 1 つの領域です。ユーザインターフェースの他の要素の情報は、印刷されたラベリング(パッケージやユーザーマニュアルなど)を補完して、セクション VI (Section VI: ユーザインターフェースとラベリング) で説明されているように、使用プロセス全体でタイムリーかつコンテキストに関連する情報を提供することによって、デバイスの使用方法に関するユーザーの理解を支援することができます。対話ポイントの例としては、デバイスによって生成され、デバイスに表示されるか、別の製品にプッシュされるアラート、関連するハードウェアのコンポーネント、およびディスプレイ画面があります。効果的な透明性の計画では、目的のユーザーと使用のコンテキストに必要な情報だけでなく、必要な情報の伝達を成功させるための最適な媒体、タイミング、および戦略を特定します。

一般に、透明性の設計プロセスは、製品が使用されるコンテキストを理解するための全体的なアプローチから開始し、ユーザーのタスクと、それらのタスク中の情報の伝達に関連する可能性のあるリスクを特定する必要があります。これは、情報が必要な方法とタイミングを決定し、コンテキストに応じた適切なリスクコントロールを製品の設計に統合し、最終的に、目的のユーザーが受信し、関連する使用コンテキストで重要な情報を機能的に理解できることを検証することによって実現できます。このプロセスは反復的であり、直線的ではない可能性があります。

透明性はコンテキストに依存するため、適切な情報は AI 対応デバイスの範囲によって異なり、そのベネフィット/リスクプロファイルと対象ユーザーのニーズによって異なります。この付録の考慮事項は、AI 対応デバイスの透明性のニーズを理解するための一貫したアプローチを開発するために、デバイスが使用されるコンテキストとユーザーのニーズに関する情報を治験依頼者が特定できるようにすることを目的としています。また、透明性は特定のデバイスリスク、特にデバイスによって出力された情報の誤解や誤用に関連するリスクに対処するのに役立ちますが、対象ユーザーのサブグループにおけるパフォ

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

ーランスに関連する重大なリスクを含む、重大な臨床リスクの存在に関する透明性を提供するだけでは、十分なリスクコントロールではない可能性があることに注意してください。

適切なタイミングでの適切な情報

意思決定を容易にするためにユーザーが必要とする可能性のある情報、およびユーザーが適切なタイミングで適切な情報をまったく持っていない場合、または誤解された場合の潜在的なリスクを考慮します。各ユーザーが実行する必要があるタスクと、意図した使用に合わせてそれらを実行するためにユーザーが知っておく必要があることに焦点を当てることが重要です。取得する必要があり、ユーザーにとって重要な情報を特定するには、次のような質問を使用して、意図した使用を包括的に検討します。

- 誰が情報を必要としており、最も効果的なコミュニケーション方法は何ですか？
- 安全で効果的な使用をサポートするために、ユーザーが情報を理解する必要があるのはどのような場合ですか？
- 使用のコンテキストは何ですか？使用のコンテキストに関する質問の例には、次のようなものがあります。
  - デバイスはどこで使用され、その場所の条件は何ですか？
  - ユーザーは同時に他に何をしている可能性がありますか？
  - 情報の適用はどのくらいタイムリーですか？
  - デバイスの出力はどのような設定で表示されますか？
  - 出力を解釈して適用するユーザーは、デバイスを操作するユーザーと同じですか？

情報は、必要ときに伝達される必要があります。透明性情報の伝達に使用できるユーザインターフェースの要素の例には、次のものがあります。

- パッケージング、
- ラベル付け、
- ユーザトレーニング
- コントロール、
- 表示要素
- 出力/レポート、
- アラーム/警告、および
- 各デバイスコンポーネントおよびユーザインターフェースシステム全体の動作ロジック。

ユーザー特性とニーズの理解

ユーザーがAI対応デバイス进行操作できるかどうかは、ユーザーの特性やデバイスの使用環境によって異なります。また、AI対応デバイスの使用環境がユーザインターフェースの設計に影響する場合があります。設計入力の一部として、使用のコンテキストでユーザーのニーズを考慮します。ユーザーとそのニーズと制限を理解することは、AI対応デバイスの開発プロセスの早い段階で行う必要があり、設計プロセスの継続中に繰り返すことができます。ユーザーには、次のようなものが含まれます。

- 患者、
- 購入者、
- 管理者、
- 医療従事者、

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

- 介護者、および
- メンテナンス技術者。

コミュニケーションの適切なコンテンツや形式など、ユーザーのニーズに影響を与える可能性のある各ユーザーの特性を考慮することが重要です。考慮事項は次のとおりです。

- ユーザーの機能的な能力（認知、身体、および感覚の能力を含む）；
- ユーザーの経験と知識のレベル（学歴を含む）；
- ユーザーがデバイスと対話する頻度；
- ユーザーが受けると予想されるトレーニングのレベル；および
- ユーザーが過去に利用した情報と比較した新しい情報の類似点と相違点。

コミュニケーションのスタイルと形式

コミュニケーションに使用する形式を考慮することも重要です。形式は、各ユーザータスクの各ユーザーに明確で適切である必要があります。次のような要因が考えられます。

- ユーザーの読み取りレベル。
- 情報の場所。
- 次のような設計要素。
  - 階層、
  - 視覚化、および
  - 動的ラベル付け。

反復設計を可能にするために、通信のタイミング、モード、および形式の選択を早期に組み込む必要があります。

説明可能性情報と視覚化

追加情報が理解を深めるのではなく、理解を損なう可能性がある場合を考慮することも重要です。たとえば、説明可能性ツールまたは視覚化は、モデルの透明性を高め、モデルの出力に対するユーザーの信頼性を高める上で有用であり、ユーザーインターフェースの一部として開発することができます。ただし、ターゲットユーザーグループに対して適切に設計および検証されていない場合、説明可能なツールまたは視覚化もユーザーに大きな誤解を与える可能性があります。

そのため、治験依頼者は、適切なテストを通じて説明可能なメトリックと視覚化を開発および検証する必要があります。

## 付録 C: パフォーマンス検証に関する考慮事項

この付録には、AI 対応デバイスにおける臨床パフォーマンス検証のいくつかの側面に関する推奨事項が含まれています。これは、治験依頼者が安全で効果的な医療デバイスを開発するのに役立つことを目的としています。治験依頼者は、FDA の安全性と有効性の評価をサポートする代替アプローチを特定することができますが、デバイスのパフォーマンスを確立するためにデバイスを厳密にテストし、デバイスの使用目的をサポートするための適切なデータを確実に収集するために、設計および開発プロセスの早い段階でその計画を統合する必要があります。たとえば、重要な臨床試験の完了後に追加の裏付けデータを収集することは困難な場合があります。治験依頼者は、特定の臨床試験の考慮事項に関する他の FDA ガイダンスに記載されている推奨事項にも従う必要があります。たとえば、AI 対応デバイスの評価と結果の報告に関する追加情報については、FDA ガイダンス「[医療機器の重要な臨床試験の設計に関する考慮事項](#)」、「[診断テストを評価する研究の結果報告に関する統計的ガイダンス](#)」、および「[臨床試験における電子ソースデータ](#)」を参照してください。これらの推奨事項は、すべてのデバイスタイプに適用されるわけではありません。

### 治験実施計画書及び統計解析計画書の事前規定

事後分析は、性能評価にバイアスをかける可能性があります。そのため、デバイスの性能を正確に評価するには、治験実施計画書及び統計解析計画書を事前に規定する必要があります。データが前向きに収集されるか遡及的に収集されるかにかかわらず、検証研究を開始する前に治験実施計画書の要素(サンプルサイズの妥当性、アーカイブされたデータまたは資料の処理、準備、処理、選択方法に関する計画など)を指定する必要があります。

### 治験報告書

すべての性能及び使い勝手の評価は客観的であるべきであり、試験データの結果(すなわち、事後調整なしです)を考慮して日和見的にモデルを調整すべきではありません。一般に、健全な検証計画(治験実施計画書及び統計分析計画書)が文書化され、最終決定された後のみ治験実施計画書を実行することは、これらの事後調整を回避するのに役立ちます。計画の実行には、必要なデータの収集、事前に規定された分析の実施及び治験結果の報告が含まれます。検証治験報告書には、関連する治験実施計画書のバージョンを明記すべきであり、反復試験又は事前に規定された計画から逸脱した試験については、適切な正当性を示すべきです。

### マスキング治験実施計画書

診断機器の場合、臨床試験におけるマスキングプロトコルは、検査の使用者が臨床参照標準の結果に対して「盲検/マスク」され、臨床参照標準の結果の提供者が検査結果に対して「盲検/マスク」されることを保証します。マスキングプロトコルはまた、モデル開発プロセス中にモデル開発者と臨床チームが検査データから完全にマスクされることを保証します。

治療機器の場合、マスキングは、非侵襲的な診断機器など倫理的に適切な場合には、2 群(例えば、プラセボ/偽のデバイス群と被験者のデバイス群)のランダム化比較試験を通じて実施されることがあります。これにより、患者と医療提供者は実際の治療割り当てに対して盲検化されます。プラセボ群には測定値がなく、盲検化ツールとしてのみ機能する場合があります(例えば、介護者が異なる治療群で差別化されたケアを提供しないようにします)。このような 2 群試験デザインが実行可能でない場合は、プラセボ効果によるパフォーマンス評価にバイアスが生じる可能性があります。



## *Contains Nonbinding Recommendations Draft – Not for Implementation*

### モデルの精度:再現性と再現性

多くの場合、AI 対応デバイスは、デバイスを特定の解剖学的位置に置いたときに生理学的信号を測定することを目的としています。測定システム(例えば、ユーザによるテストを繰り返すと、オペレータの違いや信号の変動により、著しく異なるデバイス出力が生成されるかどうか)の潜在的な変動によるデバイス出力の堅牢性を知ることが重要です。精密検査では、同じ患者に対して同じオペレーターとデバイスを使用して繰り返し測定を行う場合(再現性)、または異なるオペレーターとデバイスを使用して測定を行う場合(再現性)、デバイス出力のばらつきを測定します。より一般的には、再現性とは、同じ条件下で行われた繰り返し測定の一致度の近さです。再現性とは、あらかじめ指定された異なる条件下で行われた繰り返し測定の一致度の近さです。

臨床的および実現可能性の考慮により、すべての診断機器が精密検査を必要とするわけではないことに注意することが重要です。例えば、機器が繰り返し使用されることで患者に過度に有害である可能性がある場合(例えば、放射線または侵襲的装置の場合)、実現可能性の懸念があります。もう 1 つの例は、患者の生理的状態(血行動態パラメータなど)の変化をリアルタイムで追跡するモニタリング機器であり、同じ真実を繰り返し観察することは不可能です。

モデルの連続メトリック(例えば、確率スコア)を使用した分散成分分析に基づいて、再現性と再現性を要約する重要な統計は、被験者レベルの標準偏差(SD)と%変動係数(%CV)です。SDまたは%CVを減らすためにモデルを改善することは、製品の品質と将来の重要な臨床試験の成功確率を改善するための低コストの方法を提供する可能性があります。これは、臨床参照標準(すなわち、目標条件の有無を確立するための利用可能な最良の方法)が精密試験で測定されないことが一因です。製品によっては、精密試験で追加の要因が考慮されることがあります。画像分類タスクでは、モデルはデータの摂動(例えば、画像の移動/回転、光強度の変化、ランダムノイズ)に敏感である可能性があります。この現象は、カメラを使用して測定データをキャプチャする汎用のスマートフォンで実行されている AI 対応医療機器ソフトウェアに多く発生する可能性があります(例えば皮膚病変分析装置)。

### 試験の評価項目と判定基準

主要評価項目は通常、統計的仮説検定の枠組みの中で事前に規定された判定基準を用いて評価されます。このアプローチでは、十分な研究力を確保するために適切なサンプルサイズが必要である(すなわち、許容可能なタイプ II エラー率)。二次エンドポイントと探索的エンドポイントは、デバイスの有効性を知らせるためにも使用される可能性があり、規制当局の決定に情報を提供する証拠全体の一部です。主要評価項目の評価は、通常、95%の両側信頼区間(両面テストでタイプ I のエラーを 5%で保護できるように;片側テストでは 2.5%)に基づいて行われます。すべての出力の検証は、種類(例えば、継続的、カテゴリー的、リスクスコア)ごとに適切に対処する必要があります。

AI 対応医療機器は、診断/予後予測、治療トリアージ/優先順位付け/選択/計画など、さまざまな出力を生成できます。これらの出力の検証には、分析研究(例:精度、ベンチ、シミュレーションスタディ)、文献レビュー、診断パフォーマンス研究、読者研究(例えば、マルチリーダー、マルチケースイメージング研究)、または臨床アウトカム研究(例えば、研究またはランダム化比較試験デザインに基づきます)が含まれる場合があります。

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

AI 対応診断機器を具体的に検討する場合、主要なパフォーマンス評価は診断精度です。これは、主要な診断パフォーマンス研究で評価されます。サンプリングの変動により、精度推定の不確実性は通常、95%両側信頼区間の形式で定量化されます。試験の判定基準は、仮説検定法(例えば、信頼限界の下限/上限を事前に指定したパフォーマンス目標と比較します)を用いた統計的推論に基づることができます。点推定に基づく推論は、推定の統計的不確実性を無視し、一次分析では一般的に受け入れられないことに注意します。デバイスと同じ患者/データでテストおよび評価できる比較対象と常に比較されます。この比較対象は、臨床医、同じ使用目的に対して適切に検証されている別のデバイス、または標準治療である可能性があります。同じ患者/データでの評価は、サンプリングのばらつきによるタスクの難易度レベルと疾患スペクトルの違いを軽減するために重要です。

診断出力の性質(すなわち、二元性、多毛性、連続性)に応じて、異なる評価メトリックが可能です。

- バイナリ診断出力の場合、評価は、感度、特異度、陽性/陰性的中率 (PPV/NPV)、および陽性/陰性診断尤度比 (LR+/LR-)に基づることができます。
- 患者を複数のリスクグループの 1 つに分類し、予後モデルにしばしば見られるリスク層別化出力の場合、いくつかの評価メトリックは、検査前/検査後リスクと尤度比です。
- 連続スコアで患者の疾患リスクを評価する出力の場合、いくつかのリスク評価方法は、校正プロット、受信者動作特性 (ROC) 曲線、および決定曲線分析です。バイオマーカー評価の文脈では、予測曲線分析が使用されることがあります。
- 連続スコアについては、MAE (mean absolute error)、RMSE (root mean squared error)、散布図、デミング回帰、Bland-Altman 分析を用いた一致研究法が用いられることが多いです。

検査データが患者ごとに複数回の観察で構成される場合、信頼区間の計算では患者内の相関を考慮する必要があります。統計解析で繰り返し測定を適切に考慮しないと、推定値が偏ったり、信頼区間が狭くなったりする可能性があります。機器の性能の客観的評価を妨げる可能性があります。患者レベルの繰り返し測定を考慮する統計手法には、ブートストラップ再サンプリング法やクラスター化されたデータの分析法などがあります。

#### AI ベースの前処理ステップの検証

一部のモデルには、品質が「低い」ケースを以降の処理から除外する品質管理アルゴリズムが含まれている場合があります。ただし、このような品質が低いケースは、実際には非常に困難/困難なケースである可能性がある。たとえば、ランダムではないミス (MNAR) である。このようなケースでは、精度メトリック(例えば、感度と特異度)の診断パフォーマンスが偏っているだけでなく、バイアスがかかっている可能性もある(例えば、保証されているよりも多くの患者が、宣言された質の低い事象のために結果を得られない可能性があるという意味で)。品質が低いと見なされるケースの分析を実施して、品質管理アルゴリズムが困難なケースを除外していないことを検証する必要があります。

***Contains Nonbinding Recommendations  
Draft – Not for Implementation***

たとえば、特定の皮膚疾患の検出に携帯電話のカメラを使用する 2 つの仮想的な AI 対応診断デバイス (A と B) を比較します。A が B よりも積極的な品質管理 (QC) アルゴリズムを使用して品質が低いケースを宣言する以外は、同じ診断モデルを使用すると仮定します。QC アルゴリズムに失敗したケースを除外した後、A が B よりも優れた診断パフォーマンスを持つことは驚くことではありません。これは、A によってドロップされたが B によってドロップされなかった低品質の画像の多くが、実際には A のパフォーマンス評価に含まれていない高品質である可能性があるためです。

したがって、低品質のドロップアウトの割合を確認し、最悪のシナリオを想定した感度分析の結果を評価することによって、QC アルゴリズムの影響を調べることをお勧めします(すなわち、QC の失敗事例はすべて、モデルがうまく分類できない困難な事例であると仮定します)。

DRAFT

## 付録 D: ユーザビリティ評価の考慮事項

このガイダンスのセクション X 検証で説明されているように、治験依頼者は、市販前の提出を必要とするすべての医療機器について、設計管理 (21 CFR 820.30) の一部としてヒューマンファクター評価を実施する必要があります。[ヒューマンファクターガイダンス](#)では、この評価に対する分析アプローチと、1 つ以上の重要なタスクが特定された場合のデバイスのヒューマンファクター検証の具体的な要件について説明します。ヒューマンファクターエンジニアリングプロセスは通常、すべてのタスクの予備的な分析と評価から始まります。これにより、誤って実行された場合、またはまったく実行されなかった場合に、重大な害を引き起こす可能性のある重要なタスクが特定されます。<sup>66</sup> 治験依頼者は、デバイスに重要なタスクがあるかどうかを特定するために、この分析を実行する必要があります。重要なタスクが特定された場合、治験依頼者は[ヒューマンファクターガイダンス](#)を参照し、Human Factors Validation を実行する必要があります。重要なタスクがないデバイスの治験依頼者は、Human Factors Guidance のテストレポートを提出する必要はありませんが、デバイス設計をテストし、リスクコントロールの有効性をサポートするために、[ヒューマンファクターガイダンス](#)に概説されているプロセス、またはユーザーが選択した別のアプローチを使用してユーザビリティを評価することを選択できます<sup>67</sup>。この付録では、Human Factors Validation のテストレポートが必要ない場合に、リスクコントロールをサポートするためのユーザビリティの評価に焦点を当てています。ユーザビリティでは、すべての意図されたユーザーがデバイス使用中に特定の目標を達成できるかどうか、およびユーザーが一貫して安全かつ効果的にデバイスと対話できるかどうか扱われます。これには、ユーザーが AI 対応デバイスに関連する情報を一貫して正しく受信、理解、解釈、および適用できるかどうかが含まれますが、これに限定されません。

FDA の[ヒューマンファクターガイダンス](#)は、ユーザビリティを評価するための推奨される分析アプローチを概説していますが、治験依頼者は、そのガイダンスの範囲外のユーザータスクの評価に代替アプローチを利用することを選択できます。このテストがリスクコントロールをサポートするために使用される場合 (セクション VII (リスク評価) で説明されています)、治験依頼者は、事前に指定されたテストプロトコルと分析計画の説明、および評価方法の適切性の正当性を含める必要があります。

AI 対応デバイスの場合、治験依頼者は、情報の解釈と使用、および新しいユーザーインターフェースとの相互作用に関するユーザータスクに関連するリスクコントロールを特定し、評価することが特に重要な場合があります。AI 技術によって開発されたモデルは説明可能性と解釈可能性が異なるため、この情報の適用は AI 対応デバイスのユーザーにとって特に困難です。たとえば、一部のモデルは単純なデシジョンツリーを使用して説明できます。これは、一般にユーザーがモデルの推奨事項の基礎をフォローして理解するのが簡単です。また、複雑なディープニューラルネットワークを使用するモデルもあります。この場合、入力、ノード、および重みに関する包括的な情報があっても、ユーザーが推奨事項の基礎を完全に理解できるように説明することはできません。これは、ユーザーが AI 対応デバイスによって行われた推奨事項や決定が適切かどうかを簡単かつ独立して検証できない可能性があることを意味します。そのため、AI 対応デバイスは、デバイスの使用と情報の解釈のエラーが発生しやすい可能性があります。

<sup>66</sup> For more information regarding critical tasks, please see FDA guidance titled "[Applying Human Factors and Usability Engineering to Medical Devices.](#)"

<sup>67</sup> See Section X (Validation) for context regarding "usability" for the purpose of this guidance.

**Contains Nonbinding Recommendations**  
**Draft – Not for Implementation**

意図されたユーザーがモデルの出力を解釈するためのトレーニングが限られている場合、意図された使用が緊急のアクションを必要とする状況にある場合、モデルに明確な生物学的作用メカニズムがない場合、および反復的な更新によってモデルが変更される場合、解釈可能性と説明可能性に関する課題が増加します。これらのエラーは、害(遅れた又は誤った臨床的介入の影響を含む健康への傷害又は損害、財産又は環境への損害)<sup>68</sup>を引き起こし、デバイスの安全かつ効果的な使用に影響を与える可能性があります。

セクション VII (リスク評価) に記載されているように、治験依頼者が情報に関連するリスクの管理をサポートするためにユーザビリティの評価を含めることを選択する場合、評価は、ユーザーが情報を見つけて適用できることを実証するために適切である必要があります。このような場合は、影響評価を使用して、デバイスの情報の把握、理解、および適用に悪影響またはプラスの影響を与える可能性があるユーザータスクを判断することができます。AI デバイスに応じて、この評価には、例えば、リスク管理を目的としたトレーニングプログラムの評価を含めることができます。複数の逐次的なリスク管理を伴うより複雑な AI デバイスの場合、評価アプローチの例として、ユーザーが AI の出力をどのように解釈し、どのようなアクションが取られたかの包括的な評価を含む、臨床的実現可能性研究でのデバイスの使用を含めることができます。最終的には、ユーザーがデバイスを操作し、解釈できるかどうかを評価することが重要です。これには、ユーザーがデバイスの使用とその出力に関する重要な情報を理解し、臨床的意思決定の実際のコンテキストで適用できることを示すことも含まれます。

治験依頼者は、[ヒューマンファクターガイダンス](#)のセクション 6.3.1 (タスク分析) で概説されている一般的な構造を利用することを望むかもしれません。このセクションでは、デバイスの使用を個別のユーザータスクに体系的に分割する分析手法の例を提供しています。ただし、[ヒューマンファクターガイダンス](#)は「重大な害」に焦点を当てているが、治験依頼者は、デバイスが意図された使用に対して安全かつ有効であることを保証するために、誤った解釈を含む誤用に関連する潜在的なリスクを評価し、対処する文書を提供する必要があるかもしれないことを理解することが重要です。

本ガイダンスの付録 B (Appendix B: Transparency Design Considerations) では、ユーザー中心の透明性に関する推奨事項も概説しています。これは、ユーザビリティと情報の解釈に関連するユーザータスクとリスクの特定に役立つ可能性があり、治験依頼者がこれらのリスクを制御するための設計アプローチを開発するのに役立ちます。

<sup>68</sup> ANSI/AAMI/ISO 14971 Medical devices—Application of risk management to medical devices.

## 付録 E: モデルカードの例

モデルカードは、FDA が必要とする可能性のある種類の情報と一致する可能性のあるデバイスに関する情報を伝達するための一般的な形式です。たとえば、一般に公開されている 510 (k) の概要<sup>69</sup> やラベリング<sup>70</sup> などです。以下で説明するモデルカードの形式と内容は、モデルと AI 対応デバイスに関する情報を公開申請の概要や、治験依頼者がこの情報を共有できるその他の場所で伝達するために治験依頼者が使用できる可能性のある形式の例として機能することを目的としています。はモデルカードや特定のモデルカード形式を含めることを要求しておらず、この例はテンプレートと見なされないことに注意することが重要です。以下のモデルカードの例は、ユーザー中心の研究に基づいて設計されています。これは、技術者以外のユーザーにとって便利で理解しやすい順序と形式でデータを提示し、モデルカードの追加を容易にするために治験依頼者に提供されます。

一般に、モデルカードは、各 AI 対応デバイスの特定のニーズとコンテキストに適合させることができます。ただし、公開要約については、治験依頼者がこのガイダンスに記載されているモデルカードを作成するための一般的な原則に従うことをお勧めします。一部の要素は、デバイスによっては使用できない場合があります。

モデルカードがデジタル形式で提供されている場合、ユーザーが必要に応じてセクションを個別に展開できるようにする動的なアプローチを使用すると、情報を消化しやすくなることが研究によって実証されています。公開提出の要約は PDF ドキュメントとして提供され、形式は静的ですが、治験依頼者は可能な場合は動的ラベル付けの使用を検討する必要があります。

### デバイス名-モデルカード

#### デバイス情報:

- デバイスの名前
- デバイスのバージョン
- デバイスが作成（または最終更新）された日付
- モデルアーキテクチャ

#### Regulatory Status (公開申請の概要以外で使用されているモデルカードの場合) :

- Authorization status (認証ステータス)
- ファイル番号

#### 説明:

- 対象ユーザー(例えば、医療従事者、介護者、患者)。
- 使用目的-デバイスまたはその機能の一般的な目的。これには、使用の適応症が含まれます。
- 使用の適応症-デバイスが診断、予防、治療、または軽減する疾患または状態を記述します。これには、デバイスの対象となるターゲット患者集団および意図された使用環境(例えば、集中治療室、降圧室、在宅)の説明が含まれます。

---

<sup>69</sup> See 21 CFR 807.92.

<sup>70</sup> See 21 CFR Part 801.

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

- 使用上の指示-モデルを最適に使用するための指示と推奨事項。
- デバイスが医療提供者による使用を目的としているか、または医療提供者の監督下で使用されるかを含み、臨床上の利点(例えば、診断、治療の割り当て、モニタリング、または病状の予防、リスク評価を改善するために、個別化された患者情報を分析します)と制限事項。
- 臨床ワークフローフェーズ(例:患者の事前登録、フォームまたは臨床スケールのデジタル化、患者のトリアージ、遠隔医療およびバーチャラウンド、臨床意思決定支援システム、ワークフローの最適化、医療介入を最適化するためのエビデンスに基づく方法、ユーザーからのフィードバック)。
- モデルの入力と出力、および医療に関する意思決定またはアクションへの貢献度。
- デバイスが意思決定をサポートするか自動化するかを含み、現在の標準ケアと比較した自動化の程度。

**パフォーマンスと制限事項:**

- 精度(例えば、感度、特異度、陽性/陰性適中率及びそれらの 95%両側信頼区間)。
- 既知のバイアスまたは障害モード。
- 提供される出力に関連付けられている精度 (再現性) 。
- 開発(例えば、訓練)またはテストデータセットで十分に表現されていないため、バイアスのリスクがある可能性がある患者集団など、データの特性評価における既知のギャップ。
- モデル開発またはパフォーマンス評価の制限。
- デバイス入力が開発および検証で使用されるデータと一致しない既知の状況。
- 証拠(例えば、臨床試験番号又は裏付けとなる試験の公表結果については、デジタルオブジェクト識別子等の固有の参照 ID 又は PubMed 識別子情報)。
  - デバイスのテストに使用されるデータの特徴:
    - データの収集に使用されるデバイスの詳細を含むデータソース(例えば、臨床試験、公的または専有データベース);
    - 使用されるデータの種類(例えば、構造化された数値データ、構造化されたカテゴリカルデータ、構造化されていないテキスト、画像、時系列データ、又はこれらの組合せ);および
    - サンプルサイズ、効果サイズ、データ品質、参照標準、多様性、および代表性を含む関連する詳細。
- モデルが意図された用途とユーザーの要件を満たしていることを確立し、確認するために使用される方法(例:ヒューマンファクターのバリデーション/ユーザビリティ評価、ユーザー受容性試験、臨床バリデーション、事前トレーニング済みモデルの特定、その他)。

**リスク管理:**

- モデル、データ、および出力に関連する潜在的なリスク(例えば、禁忌、副作用、データプライバシーリスク、サイバーセキュリティリスク、バイアスリスク、情報ギャップ)。
- 製品ライフサイクル全体で、リスクと患者のアウトカムに影響を与える可能性のある情報の説明。

*Contains Nonbinding Recommendations*  
*Draft – Not for Implementation*

- 相互作用、展開、および更新。必要に応じて、次の情報を提供します。
  - 計算リソースが必要です。
  - モデルの展開および更新方法に関する詳細（次を含む）。
    - ローカルサイト固有の受け入れテストまたは検証の実施方法;
    - 継続的なパフォーマンス監視;
    - 成功と失敗の透明なレポート;
    - 変更管理戦略;および
    - 脆弱性に対処するためのプロアクティブなアプローチ。
  - 関係者への必要に応じた情報の伝達。
  - ソフトウェアの品質(仕様、基準及び規制遵守の問題、知的財産の問題、使用するリスク管理及び保護措置、その他)。

**開発:**

- データ装置の開発に使用されるデータの特徴:
  - データの収集に使用される装置の詳細を含むデータソース(例えば、臨床試験、公的または専有データベース)。
  - 使用されるデータタイプ(例えば、構造化された数値データ、構造化されたカテゴリカルデータ、構造化されていないテキスト、画像、時系列データ、又はこれらの組合せ)。
  - サンプルサイズ、効果サイズ、データ品質、参照標準、多様性、および代表性を含む関連する詳細。



**Contains Nonbinding Recommendations**  
**Draft – Not for Implementation**

**付録 F: 例 510 (k) の概要とモデルカード**

一般に公開されている概要は、特定の販売申請(例:510 (k)、<sup>71</sup> De Novo、<sup>72</sup> PMA <sup>73</sup>)に適用される要件に従う必要があります。以下の項目は、製造業者がカバーすることが期待されるトピックの完全なリストではなく、すべてのトピックがすべての販売申請に適用されるとは限りません。同様に、FDA はこの概要に追加情報を含めるよう要求することがあります。本付録は、完成した基本モデルカードの例を含め、一般的に治験依頼者が 510 (k) 要約で提供すべき情報の種類の一例にすぎません。モデルカードと公開要約の他のセクションとの間で情報を繰り返す必要はないが、治験依頼者が代替フォーマットに有用な文脈を提供すると考える場合は、情報を繰り返すことができます。

**使用適応:**

疾病 X スクリーニングモデルは、互換性のある ECG デバイスから記録された 12 誘導心電図 (ECG) を分析することにより、22 歳以上の患者における疾病 X のスクリーニングを支援することを目的としたソフトウェアです。疾病 X の独立した診断デバイスとして意図されているわけではないが、陽性の結果は疾病 X の診断を確立するためにさらなる臨床評価の必要性を示唆する可能性があります。患者が疾病 X のリスクが高い場合、陰性の結果はさらなる非侵襲的評価を除外すべきではありません。疾病 X の診断のための現在の標準治療法に代わるものとして使用すべきではなく、臨床医の判断と併用すべきです。

**デバイスの説明:**

独立したソフトウェアには、畳み込みニューラルネットワークを使用して、4 つの互換性のある心電図デバイス(A、B、C、および D)から取得した 12 誘導安静時心電図の 10 秒間を解釈および分析し、患者が疾病 X であり、さらなる臨床評価が必要であるかどうかの可能性に関する出力を提供する機械学習モデルが含まれています。ソフトウェアには、提供された心電図データがモデル出力を生成するための心電図入力要件を満たしているかどうかをエンドユーザーに通知する品質チェックも含まれています。要件を満たしていない場合は、エラーメッセージが表示されます。

**技術的特徴の概要:**

	対象医療機器	既承認医療機器	比較
アプリケーション番号	KXXXXXXX	KXXXXXXX	-
製品コード	XXX	XXX	-

<sup>71</sup> See 21 CFR 807.92. For more information, please see FDA guidance titled “The 510(k) Program: Evaluating Substantial Equivalence in Premarket Notifications [510(k)].”

<sup>72</sup> See 21 CFR 860.220. For more information, please see FDA guidance titled “De Novo Classification Process (Evaluation of Automatic Class III Designation).” <sup>73</sup> See 21 CFR Part 814.9(e).

**Contains Nonbinding Recommendations**  
**Draft – Not for Implementation**

	対象医療機器	既承認医療機器	比較
レギュレーション番号	21 CFR XXXX	21 CFR XXXX	-
Rx/OTC	Rx	Rx	同じ
使用目的 (使用適応症)	<p>疾病Xスクリーニングモデルは、互換性のある心電図装置から記録された12誘導心電図を解析することにより、22歳以上の患者における疾病Xのスクリーニングを支援することを目的としたソフトウェアです。疾病Xの単独の診断装置として意図されたものではありません。しかし、陽性の結果は、疾病Xの診断を確立するためにさらなる臨床評価の必要性を示唆する可能性があります。患者が疾病Xのリスクが高い場合、陰性の結果はさらなる評価を除外すべきではありません。疾病Xの診断のための現在の標準治療法に代わるものとして使用すべきではなく、臨床医の判断と併用すべきです。</p>	<p>18歳以上の患者が疾病Xであるかどうかを判断するための補助として使用されることを目的としたソフトウェアです。このソフトウェアは、互換性のあるデバイスからの12誘導心電図を分析するものであり、単独の診断装置として使用すべきではありません。</p>	<p>類似しています。どちらのデバイスも、疾病Xの補助およびスクリーニングツールとして使用されます。述語デバイスの使用適応は18歳以上の患者であり、対象デバイスは22歳以上の患者です。</p>
動作モード	スポットチェック/診断デバイスとして使用しません。	スポットチェック/診断デバイスとして使用しません。	同じ

***Contains Nonbinding Recommendations  
Draft – Not for Implementation***

	対象医療機器	既承認医療機器	比較
ハードウェア入力	次の互換性のあるデバイスからの 12リードECG: A B C D	次の互換性のあるデバイスからの 12リードECG: A B	類似しています。どちらも12リードECGからの入力を必要としますが、対象のデバイスはより互換性のあるECG入力デバイスを可能にします。
出力	ソフトウェアの出力は次のとおりです。 1. 疾病Xの存在。 疾病Xの診断を確定するために、さらなる臨床評価を求めてください。 2. 疾病Xの存在の可能性は低いですが、しかし、臨床的判断を用いて、さらなる評価が必要かどうかを判断してください。 3. エラーメッセージ: 12誘導ECGは適切な品質チェックに合格しません。	ソフトウェアは、疾病Xの可能性と、さらに評価が必要かどうかに関する出力を提供します。	類似しています。どちらのデバイスも、疾患Xの存在とさらなる評価が必要かどうかを識別します。どちらのデバイスも、単独で使用すべきではなく、疾患Xの診断のためにさらなる評価が必要な場合は臨床的判断を使用すべきであることを識別します。
モデルトレーニングのための臨床基準	心エコー図	心エコー図	同じ

**Contains Nonbinding Recommendations**  
**Draft – Not for Implementation**

	対象医療機器	既承認医療機器	比較
性能	感度:87% (83%、89%) 特異度:83% (81%、85%) 正の予測値 (PPV) :56%	感度:82% (78%、85%) 特異度:81% (79%、84%) ポジティブ予測値 (PPV) :53%	類似しています。対象医療機器は、感度が既承認医療機器よりも優れています。 特異度とPPV

**モデルトレーニング説明:**

モデルはテストデータセットから独立したデータセットからトレーニングされました。モデルは、互いに 30 日以内に実施された心電図および心エコー図を受けた 30,000 人の患者でトレーニングされました。心エコー図は、患者における臨床基準 (ground truth) を確立するために使用されました。データセットは、2 つの多様な病院ネットワーク (病院 A と病院 B) の臨床データベースから収集されました。疾患 X は、心エコー画像に基づいて左心室壁厚が 15 mm 以上の患者と定義されました。

トレーニングデータセットには、疾患集団を代表する以下の人口統計学的内訳が含まれていました。

人種	割合(%)
白人	75.5
黒人またはアフリカ系アメリカ人	13.6
アメリカンインディアンまたはアラスカ先住民	1.3
ネイティブハワイまたは太平洋諸島	6.3
ハワイ先住民または太平洋諸島	0.3
2つ以上の人種	3.0

男性 49.5%、女性 50.5%でした。平均年齢は 62 歳で、その内訳は以下のとおりでした。

年齢 (歳)	割合(%)
40歳未満	10
40-49	10
50-59	25
60-69	30
70-79	15
79歳以上	10

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

疾患 X を有する患者はコホート全体の 20%であり、疾患 X を有さない患者（対照群）はコホート全体の 80%でした。両群をトレーニングデータセット（50%）、チューニングデータセット（20%）、チューニング評価データセット（30%）に分けました。チューニング評価データセットからモデルの感度と特異度を計算しました。モデルは以下を達成できました。

- 感度： 87%（83%、89%）
- 特異度： 83%（81%、85%）
- 陽性適中率（PPV）： 56%

### **非臨床パフォーマンスデータの要約**

モデルは、FD&C 法および施行規則の適用要件を考慮して評価されました。これには以下のテストが含まれました。

- Human Factors and Usability テストは、FDA のガイダンス文書「[医療機器への Human Factors and Usability Engineering の適用](#)」で推奨されているとおりに実施され、文書が提供されました。
- サイバーセキュリティ試験を実施し、FDA のガイダンス文書「[医療機器におけるサイバーセキュリティ:品質システムの考慮事項と市販前申請の内容](#)」で推奨されている文書を提供しました。
- ソフトウェア検証および検証試験を実施し、[市販前ソフトウェアガイダンス](#)で推奨されている文書を提供しました。

### **臨床検証の概要:**

#### **研究デザイン**

このモデルは、米国内の 5 つの異なる多様な医療システムにおける 25, 000 人の患者とその患者記録のレトロスペクティブ研究で検証されました。この研究の目的は、疾患 X の存在のスクリーニングにおけるモデルの性能を確立することでした。

- この研究に登録された患者は 22 歳以上であり、少なくとも一回の安静時心電図と心電図の日付から 30 日以内の心エコー図を有していました。最新の心エコー図は、心エコー図の前にその患者の最新の心電図と対になっていました。
- 12 誘導安静時心電図データを収集し、モデルへの入力として以下のモデルの心電図装置(A、B、C および D)を使用しました。
  - 12 誘導心電図の持続時間は 10 秒でなければなりません。

モデルの除外基準は以下のとおりでした。

- 本研究に登録された患者は 22 歳未満でした。
- 必須データが欠落していた(すなわち、ECG の技術的パラメータ、年齢又は人種、実施された ECG 及び心エコー図に関する情報)。
- 心電図データの収集には、12 誘導心電図の異なるデバイスモデルが使用されました。
- 12 誘導心電図の持続時間は 10 秒ではありません。
- 患者はペースメーカーを装着しています。

**Contains Nonbinding Recommendations**  
**Draft – Not for Implementation**

5つの施設のそれぞれが、25,000人2,114人の心電図ペアの最終プールに約5,000人の患者-心電図ペアを提供しました。研究サンプルは、疾患集団を代表する以下の人口統計学的内訳を有していました。

人種	割合(%)
白人	75.5
黒人またはアフリカ系アメリカ人	13.6
アメリカインディアンまたはアラスカ先住民	1.3
アジア系	6.3
ハワイ先住民または太平洋諸島系	0.3
2つ以上の人種	3.0

研究サンプルの病院サイトの内訳は以下のとおりでした。

病院サイト	割合(%)
A	19.64
B	21.36
C	20.1
D	18.4
E	21.5

対象は男性49.5%、女性50.5%でした。平均年齢は65歳で、その内訳は以下の通りでした。

年齢(歳)	割合(%)
40歳未満	10
40-49	10
50-59	16
60-69	23
70-79	22
79歳以上	19

研究サンプルのECGペアは、以下のECG収集デバイスによって収集されました。内訳は以下のとおりです。

ECG収集デバイス	パーセンテージ(%)
A	26.6
B	25.1
C	24.9
D	23.3

***Contains Nonbinding Recommendations  
Draft – Not for Implementation***

主要評価項目

本研究の主要評価項目は、95%両側信頼区間の下限を以下としました。

- 感度:75%以上
- PPV:50%以上

研究結果

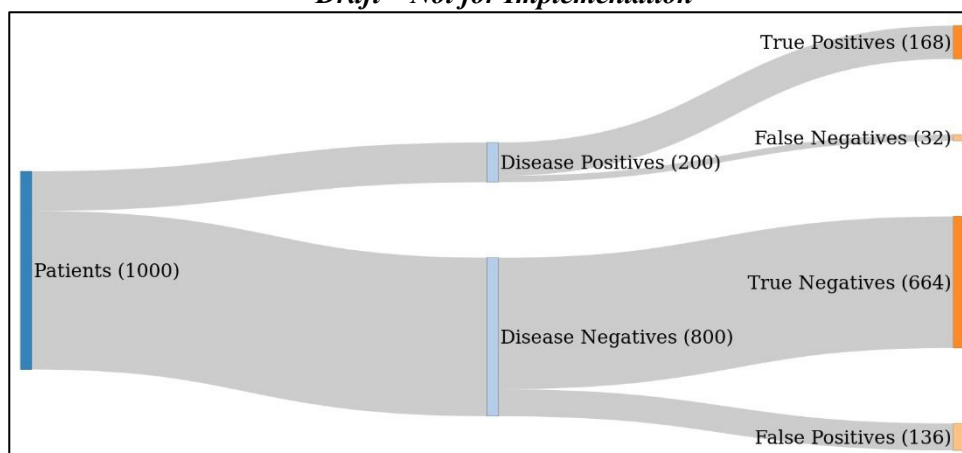
本モデルは、感度 84%、特異度 83%、PPV 55%、陰性的中率 (NPV) 95%を達成しました。点推定値とその95%両側信頼区間、および混同行列は、次の例に示すように表で報告できます。

	Ref. Pos	Ref. Neg	SUm	尤度比	パフォーマンス
Test. Pos	4200	3400	7600	4.9 (4.8, 5.1)	PPV=55.3% (4.9%、654.1%、56.4%)
テスト. Neg	800	16600	17400	0.2 (0.2, 0.2)	NPV=95.4% (95.1% 95.7%)
Sum	5000	20000	2500	1 (1, 1)	有病率20% (19.5%、20.5%)
Performance	感度=84% (82.9%、85%)	特異度=83% (82.5%、83.5%)			

ベネフィット・リスク考慮のための試験結果の平易な解釈

デバイスの意図された使用集団における疾患 X の有病率が 20%であると仮定します。対象集団からの 1000 人の患者のうち、約 168 人 (1000×有病率×感度) の患者が疾患 X を有すると正しく分類され(すなわち、全参照陽性患者 200 人のうち 168 人がデバイス真陽性でした)、約 136 人 (1000× (1-有病率) × (1-特異度) ) の患者が疾患 X を有すると誤って分類される(すなわち、全参照陰性患者 800 人のうち 136 人がデバイスの偽陽性でした)。したがって、真の陽性患者 1 人につき 0.8 人 (136/168) の偽陽性患者が犠牲になる(は、標準的なケアからの Y、またはすべての患者が陽性と呼ばれる最悪のシナリオからの 4 (800/200) と比較しています。)。さらに、NNP (Number Needed to Predict) = 1/ (PPV+NPV-1) = 1.97 (NNP が Z の標準治療、または NNP が 1 の完全なデバイスと比較して)であるため、1 人の余分な真の陽性患者を特定するには、約 2 人の患者を評価する必要があります。

**Contains Nonbinding Recommendations  
Draft – Not for Implementation**



各人口統計のサブグループ分析を以下に示します。

この架空の例では信頼区間を生成できませんでしたが、治験依頼者は報告されたすべての結果に信頼区間を含める必要があることに注意してください。各セルには、信頼区間を表すためのプレースホルダーが含まれています。(XII、IXul)。ll は下限、ul は上限を表します。

人種	パーセンテージ(%)	感度	PPV
白人	75.5	85.3 (XII, Xul)	57.3%(XII, Xul)
黒人またはアフリカ系アメリカ人	13.6	82.9 (XII, IXul)	54.4%(XII, Xul)
アメリカインディアンおよびアラスカ先住民	1.3	81.6 (XII, Xul)	54.8%(XII, Xul)
アジア系	6.3	83.9 (XII, Xul)	56.1%(XII, Xul)
ハワイ先住民その他の太平洋諸島系	.3	83.6 (XII, Xul)	56.5%(XII, Xul)
二つ以上の人種	3	84.1 (XII, Xul)	55.4%(XII, Xul)
<b>年齢</b>	<b>割合(%)</b>	<b>感度</b>	<b>PPV</b>
40歳未満	10	84.9 (XII, Xul)	55%(XII, Xul)
40-49	10	85.1 (XII, Xul)	55.4%(XII, Xul)
50-59	16	84.1 (XII, Xul)	55.4%(XII, Xul)
60-69	23	84.5 (XII, Xul)	56%(XII, Xul)
70-79	22	83.6 (XII, Xul)	55.4%(XII, Xul)
79歳以上	19	82.1 (XII, Xul)	52.7%(XII, Xul)



**Contains Nonbinding Recommendations**  
**Draft – Not for Implementation**

各 ECG 収集装置のサブグループ解析を以下に示します。

心電図取得デバイス	パーセンテージ (%)	感度	PPV
A	26.6	84.7 (XII, Xul)	56.5% (XII, Xul)
B	25.1	83.6 (XII, Xul)	54.3% (XII, Xul)
C	24.9	85.4 (XII, Xul)	57.9% (XII, Xul)
D	23.3	84.6 (XII, Xul)	55.1% (XII, Xul)

各施設のサブグループ解析は以下のとおりです。

施設	割合 (%)	感度	PPV
A	19.64	83.6 (XII, Xul)	54.3% (XII, Xul)
B	21.36	85.1 (XII, Xul)	51.4% (XII, Xul)
C	20.1	84.1 (XII, Xul)	55.4% (XII, Xul)
D	18.4	85.4 (XII, Xul)	57.9% (XII, Xul)
E	21.5	84.7 (XII, Xul)	56.5% (XII, Xul)

**モデルカード:**

デバイス情報

- モデル名: 疾患 X スクリーニングモデル
- モデルのバージョン: バージョン 1.0.1
- モデルのリリース日: 2023 年 12 月
- モデルのアーキテクチャ: 畳み込みニューラルネットワーク

デバイスの説明

- 対象ユーザー: 医療従事者。
- 使用の適応: このモデルは、互換性のある ECG デバイスで作成された 12 誘導 ECG の記録を分析することによって、22 歳以上の患者の疾患 X のスクリーニングを支援することを目的としたソフトウェアです。疾患 X のための独立した診断デバイスとして意図されているわけではありません。しかし、陽性の結果は、疾患 X の診断を確立するためにさらなる臨床的評価の必要性を示唆する可能性があります。患者が疾患 X のリスクが高い場合、陰性の結果はさらなる非侵襲的評価を除外すべきではありません。これは、疾患 X の診断のための現在の標準的なケア方法を置き換えるために使用されるべきではありませんが、臨床医の判断と一緒に適用されます。

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

- 臨床ワークフローの段階: 疾病 X の診断を確立するために、さらなる臨床的フォローアップ(例えば、心エコー図)の補助およびスクリーニングツールとして使用されます。
- 臨床的利益: 心臓画像が利用できない可能性のある疾病 X のポイントオブケアスクリーニングを提供します。

性能と限界

- データ型: 12 誘導心電図 (ECG)
  - 説明: 以下の四つの互換性のある ECG 装置(A、B、C、および D)から得られた 12 誘導心電図 (ECG) の 10 秒間の持続時間。互換性のある ECG 装置のサンプリングレートは 500 Hz です。
- 臨床参照標準: 臨床参照標準を確立するために ECG から 30 日以内に得られた心エコー図。
- モデル検証:
  - データのサイズと種類: 米国の 5 つの異なる多様な医療システムにおける 25, 000 人の患者とその患者記録の後ろ向き研究。5 施設はそれぞれ 5, 000 対の患者-ECG ペアを提供し、最終的な 25, 000 対の患者-ECG ペアのプールとなりました。
  - 除外基準:
    - 本研究に登録された患者は 22 歳未満でした。
    - 必須データが欠落していた(すなわち、ECG の技術的パラメータ、年齢又は人種、実施された ECG 及び心エコー図に関する情報)。
    - 心電図データに破損または欠落したリードが含まれていました。
    - 心電図データの収集には、異なるモデルの 12 誘導心電図が使用されました。
    - 12 誘導心電図の持続時間が 10 秒ではありません。
    - 患者はペースメーカーを装着しています。
  - データ結果 (テストデータセットから計算):
    - 感度: 84% (82.9%、85%)
    - 特異度: 83% (82.5%、83.5%)
    - PPV: 55.3% (54.1%、56.4%)
    - NPV: 95.4% (95.1%、95.7%)
- 非臨床試験:
  - FDA のガイダンス文書「医療機器へのヒューマン・ファクタおよびユーザビリティ・エンジニアリングの適用」で推奨されているように、ヒューマン・ファクタおよびユーザビリティ・テストが実施され、文書が提供されました。
  - FDA のガイダンス文書「医療機器におけるサイバーセキュリティ: 品質システムの考慮事項と市販前申請の内容」で推奨されているように、サイバーセキュリティ・テストが実施され、文書が提供されました。
  - 市販前ソフトウェアガイダンスで推奨されているように、ソフトウェアの検証および検証テストが実施され、文書が提供されました。

*Contains Nonbinding Recommendations  
Draft – Not for Implementation*

**リスク管理:**

市販前ソフトウェアガイダンスで推奨されているように、ANSI/AAMI/ISO 14971 医療機器に準拠して、リスク管理が実施され、文書が提供されました。

-医療機器へのリスク管理の適用。

- モデル、データ、および出力に関連する潜在的なリスク(例えば、禁忌、副作用、データプライバシーリスク、サイバーセキュリティリスク、バイアスリスク、情報ギャップ):モデルに関連する潜在的なリスクには、偽陽性または偽陰性の出力による不適切なフォローアップが含まれます。これは、(1) モデルのバイアス、または(2) サポートされていない患者集団またはサポートされていない入力/ハードウェアでモデルを使用することによって発生する可能性があります。さらに、情報のギャップにより、フォローアップのためにデバイスの出力に過度に依存する可能性があります。特定されたリスクの管理には、臨床検証テスト、ソフトウェアの検証および検証テスト、ヒューマンファクターのテストおよびラベリングが含まれます。
- 製品ライフサイクル全体でリスクと患者のアウトカムに影響を与える可能性のある情報の説明:モデルの開発と臨床検証には、40歳未満の参加者のわずか10%が含まれていました。これは、そのサブグループでのモデルのパフォーマンスが完全に特徴づけられていないことを意味している可能性があります。
- 相互作用、展開、および更新:展開されたモデルを継続的に監視して、サイト固有のパフォーマンスを評価し、脆弱性を特定し、パフォーマンスの透明性とサイトおよびエンドユーザーに対する継続的なメンテナンスを確保する、品質システム規則(21 CFR Part 820)に準拠した包括的なデバイスパフォーマンス監視計画が用意されています。
  - 必要な計算リソース。
  - モデルの展開方法と更新方法に関する詳細:
    - ローカルサイト固有の受け入れテストまたは検証の実施方法:サイトの既存のシステムへの統合に関する問題を理解し、そのサイトの患者集団のサブセットでパフォーマンスを測定するために、サイトの母集団全体でモデルを使用する前に、モデルを展開し、データを一か月間収集します。このプロセスを通じて、母集団全体に公開する前に展開に関する問題に対処でき、モデルのパフォーマンスと追加のトレーニングと開発の必要性を特徴付けるのに役立ちます。または、サイトで予想されるパフォーマンスを評価するために使用できる履歴データを提供することもできます。
    - 継続的なパフォーマンス監視:自動化されたパフォーマンス計算がモデルと共に展開され、6か月ごとに計算されます。パフォーマンスが予想範囲外の場合は、サイト管理者と治験依頼者に自動電子メールが送信されます。これにより、パフォーマンスの問題を理解するためのプロセスが開始され、これに対処するための軽減計画が実施されます。
    - 成功と失敗の透明なレポート:すべてのサイトは、さまざまなサイトで展開されたモデルの成功と失敗、およびこれらの成功と失敗をコンテキスト化するためのサイトの特性を含む匿名のレポートにアクセスできます。
    - 変更管理戦略:変更を必要とする問題が発生した場合、またはサイトとユーザーによって機能が要求された場合は、確立された品質システムの手順に従って変更管理が実施されます。
    - 脆弱性に対処するための積極的なアプローチ:サイトとユーザーは、問題が発生してから48時

***Contains Nonbinding Recommendations***  
***Draft – Not for Implementation***

間以内に問題を報告することをお勧めします。その後、苦情処理手順に従って、これらの手順に従って修正プログラムが発行されます。

- 必要に応じた情報の関係者への伝達: 前述のように、パフォーマンスが予想範囲外の場合は、デバイスによって自動電子メールが生成されます。
- ソフトウェアの品質(仕様、基準及び規制遵守の問題、知的財産の問題、使用するリスク管理及び保護措置、その他):

**開発:**

- モデルトレーニング:
  - データサイズ: 30 日以内に心電図と心エコー検査を受けた 3 万例。2 つの異なる病院ネットワーク (病院 A と病院 B) の臨床データベースから収集されたデータセット。
  - 疾患 X を有する患者はコホート全体の 20% であり、疾患 X を有さない患者 (対照群) はコホート全体の 80% でした。両群をトレーニングデータセット (50%)、チューニングデータセット (20%)、チューニング評価データセット (30%) に分けました。
  - データ結果 (チューニング評価データセットから算出):
    - 感度: 87% (83%、89%)
    - 特異度: 83% (81%、85%)

**結論:**

なお、本システムと既承認機器の技術的特性に差異はあるものの、安全性や有効性について異なる問題を提起するものではありません。本申請において提供された情報に基づき、本デバイスは、臨床性能の結果及び非臨床での検証・検証の結果を通じて、既承認機器と実質的に同等であることを実証しています。