事　務　連　絡
令和８年 11 月５日

各都道府県衛生主管部（局）　御中

厚生労働省医薬局医薬品審査管理課

「医薬品の製造・品質管理（ＣＭＣ）領域におけるＡＩ・機械
学習を含むモデル活用時の考え方」について

医薬品の製造現場において、機械学習を含む人工知能（ＡＩ）技術の検討が進んでいる中、利用されるモデルの性能や信憑性の評価が課題とされているところです。

今般、令和６年度ＡＭＥＤ医薬品等規制調和・評価研究事業「先進的手法を用いた医薬品の製造・管理及び評価手法におけるライフサイクルマネジメントに関する研究」において、「医薬品の製造・品質管理（ＣＭＣ）領域におけるＡＩ・機械学習を含むモデル活用時の考え方」が別添のとおりとりまとめられましたので、貴管下関係業者等においてご活用いただきますよう、周知をお願いします。

また、本事務連絡の写しについて、別記の関係団体等宛てに発出するので、念のため申し添えます。

（別記）

日本製薬団体連合会
一般社団法人日本製薬工業協会
米国研究製薬工業協会在日執行委員会
一般社団法人欧州製薬団体連合会
独立行政法人医薬品医療機器総合機構

別添

医薬品の製造・品質管理（CMC）領域における AI・機械学習を含むモデル活用時の考え方

1. はじめに

近年、医薬品の製造現場において、また将来的な製造販売承認申請も見据え、機械学習（Machine Learning: ML）を含む人工知能（Artificial Intelligence: AI）技術の検討が進んでいる。これらの技術は、開発の効率化や製品品質の高度な保証に寄与する一方で、そのモデルの性能や信憑性をどのように評価し、規制当局へ説明すべきかとの課題も顕在化している。

本文書は、ICH M15 ガイドライン「General Principles for Model-Informed Drug Development」（2026 年 1 月 29 日付 Step4 文書）に示された一般原則を踏まえつつ、医薬品の製造・品質管理（CMC）領域においてモデルを具体的に適用する際の評価の考え方をまとめたものである。本文書が対象とするモデルには、ML 等のデータ駆動型モデルのみならず、物理法則や化学反応機構等に基づくメカニスティックモデル（本文書では「物理モデル」という。）及びこれらを組み合わせたハイブリッドモデルを含む。

本文書の目的は、開発者/申請者が自身のモデルについて、その使用目的（Context of Use: COU）に応じた適切なリスク評価を行い、必要な検証（Verification & Validation: V&V）計画を立案・実行するための標準的な検討フローを提供することにある。これにより、医薬品の製造現場及び製造販売承認申請において、具体的にどのようなデータセットで何を検証すればよいかという予見可能性を提供し、医薬品の製造・品質管理（CMC）領域における ML を含む AI 技術の活用促進に繋げたい。

2. 適用範囲と位置付け

本文書は、主に医薬品の CMC 領域（製造プロセスの開発・管理、品質試験等）で使用されるモデルを対象とする。ICH M15 ガイドラインがモデル評価の包括的なフレームワークを提供しているのに対し、本文書はそのフレームワークをベースに、CMC の実務に落とし込んだ際の具体的な考え方を簡略化し、示すものである。

CMC 領域では、これまでもデータ駆動型モデル（例：ケモメトリックモデル）や物理モデル（例：滞留時間分布（Residence Time Distribution: RTD）モデル）、あるいはこれらを組み合わせたハイブリッドモデルが使用されてきており、特に確立された手法に基づくモデルに対する薬事上の評価すべき事項については、規制当局との共通認識が進んでいる部分もある。そのため、ICH M15 ガイドラインに従ったモデル解析計画書（Model Analysis Plan: MAP）に基づいた事前の規制当局との協議が必ずしも必要とされないケースも考えられる。

本文書はモデル評価の円滑化及び規制当局との協議の簡素化を目指し作成されたものである。規制当局との相談のタイミングについては本文書では触れられていないものの、規制

当局への相談を阻害するものではない。またコンピュータ化システムバリデーション（Computerized System Validation: CSV）やデータの完全性等の GMP 上の要求事項については関連する通知やガイドラインを別途参照する必要がある。

本文書の記載事項は、個々のモデルの特性や技術進歩に応じて柔軟に解釈されるべきであり、全てのケースに一律に適用される硬直的なルールとして意図されたものではない。なお、本文書はデータ駆動型モデル（ML 等）、物理モデル及びハイブリッドモデルを対象とする。ただし、運用期間中にモデルパラメータが固定されているモデルを対象とする。運用中に学習等によりモデルパラメータが継続的又は自動的に更新されるモデル、出力の内容や挙動をあらかじめ十分に規定・検証することが困難なモデルは、本文書の対象外とする。

3. 用語の定義

本文書で用いるモデルの類型及び検証・妥当性確認に関する用語を以下の通り定義する。

＜モデルの類型＞

- データ駆動型モデル

  観測データに基づき入出力間の関係を経験的に学習するモデル。物理法則や化学反応機構に関する事前の仮定を必ずしも前提としない。PLS 回帰、ランダムフォレスト、深層学習等がこれに含まれる。特に、内部構造の解釈が困難であり、入出力関係の透明性が低いモデルをブラックボックス型モデルと呼ぶ。

- 物理モデル（メカニスティックモデルとも呼ばれる。）

  物理法則、化学反応機構、物質収支等の科学的原理に基づいて構築されるモデル。滞留時間分布（RTD）モデル、反応速度モデル等がこれに含まれる。

- ハイブリッドモデル

  物理モデルとデータ駆動型モデルを組み合わせたモデル。例えば、物理モデルのパラメータを機械学習で推定する構成等が含まれる。

- サロゲートモデル（代理モデル）

  本文書におけるサロゲートモデル（代理モデル）とは、計算負荷又は評価時間等の観点から直接反復使用することが困難な参照物理モデルを用いたシミュレーションにより得られるデータセットを学習データとして構築されるデータ駆動型モデル又は参照物理モデルの振る舞いを模倣するように学習させたデータ駆動型モデルをいう。参照物理モデルの入出力関係を近似することにより、より低い計算負荷又は短い評価時間での予測を可能とする。メタモデル又はエミュレータとも呼ばれる。サロゲートモデルは単独で用いられる場合のほか、ハイブリッドモデルの構成要素として定義された COU

の範囲内で、参照物理モデルの全部又は一部を代替する目的で用いられる。

＜検証・妥当性確認に関する用語＞

・ モデルバリデーション

モデルの頑健性及び性能の適切性を評価することを目的とする作業。

・ モデルベリフィケーション

データ処理及び解析実施のためのユーザー生成コード（すなわち、プログラミング言語又はソフトウェアのユーザーによって記述された指示）にエラーが含まれず、式がモデルの仮定を反映しており、プログラミング言語又はソフトウェアでの表記に間違いがなく、かつ計算が正確であることを保証することを目的とする作業。

・ 分析法バリデーション

分析法が使用目的にかなっていることを立証すること。

・ 分析法ベリフィケーション

日常的に使用される試験室で分析法が分析法開発の目的・目標（目標プロファイル）を常に満たすことを確認すること。

・ プロセスバリデーション

工業化研究の結果、既存の類似製品の製造実績等に基づく製品品質への影響要因（例：原料等の物性、操作条件等）を考慮して設定した許容条件の下で工程が稼働し、求められる品質の製品が恒常的に得られる妥当な工程である旨を検証し、文書とすること。

・ プロセスベリフィケーション

製造手順等が期待される結果を与えたことを確認し、これを文書とすること。

＜データに関する用語＞

・ 学習データ

モデルの学習に使用されるデータであり、テストデータには含まれないもの。

・ チューニングデータ

モデルを選択及び最適化するために使用されるデータであり、テストデータには含まれないもの。

・ テストデータ

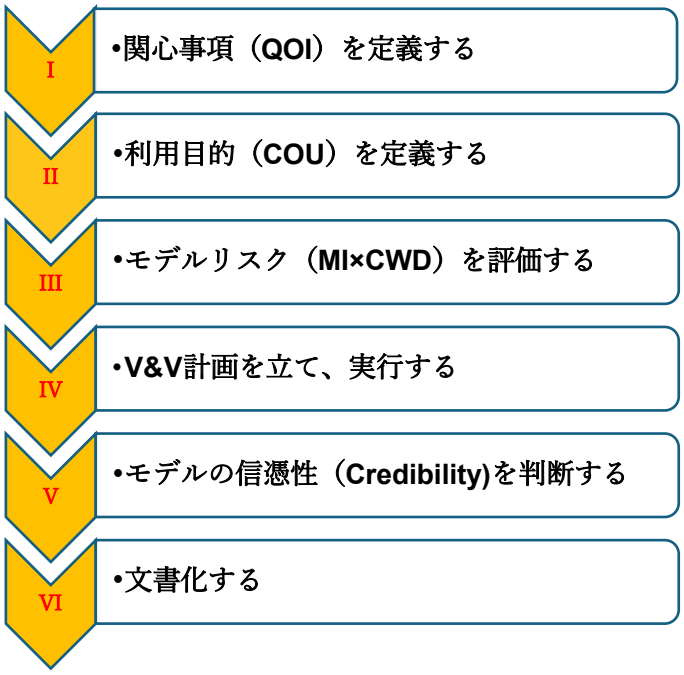
モデルの学習/チューニングには使用されないデータであり、学習/チューニング後にモデルの性能を推定するために使用されるデータ。

4. モデル評価フロー

モデル評価フローは以下の通りとなる。

① 関心事項（Question of Interest: QOI）を定義する。
② 利用目的（Context of Use: COU）を定義する。
③ モデルリスク（Model Influence: MI × Consequence of Wrong Decision: CWD）を評価する。
④ V&V 計画を立て、実行する。
⑤ V&V 及び不確かさの定量化（Uncertainty Quantification: UQ）の結果から、モデルが COU に対して適切であるか（モデルの信憑性（Credibility））を判断する。
⑥ 文書化する。

モデル評価フロー



I. QOI の定義

モデルにより回答しようとしている疑問。

II. COU の定義

QOI への回答に使用するモデルの役割及び範囲。

【役割】何をモデル化するのか、モデルの出力がどのように利用されるか及び意思決定のワークフロー（モデルの出力を受けて誰がどのように次のアクションを決定するか）を説明する。

【範囲】モデルの適用範囲を、モデル構築に用いたデータの範囲（入力変数の範囲、製品・プロセスの種類等）に基づき明示する。また、モデルの出力と併せて他の情報（例：検証実験データ、製造実績データ、分析法による実測値）が使用されるのか、モデルの出力が唯一の情報となるのかを、他のエビデンスとの相対的寄与の観点から説明する。

III. モデルリスクの評価

モデルリスクは、ある意思決定に関するエビデンス全体に占めるモデルの出力の寄与（Model Influence: MI）と、誤った決定によって生じる結果（Consequence of Wrong Decision: CWD）を組み合わせることで、決定される。

MI（Model Influence）：追加データ又はエビデンスの寄与を考慮した上で、意思決定に占めるモデルの出力の意図された重み。

High	モデルの出力が、意思決定において他の寄与する証拠資料との相対的な関係において支配的な根拠となり、対抗できる独立したエビデンスが存在しない。
Medium	High 又は Low に分類されないもの。
Low	モデルの出力は複数の独立した証拠資料のひとつとして参照される補完的情報であり、他の証拠資料がモデルの出力と同等以上の重みで意思決定に寄与する。

（注）MI の評価においては、モデルの出力のみならず、当該意思決定に寄与する他の独立したエビデンス（例：工程内管理試験、最終製品試験等）の有無及びその寄与の程度を考慮する。これらの独立した検証手段が存在する場合、モデルの出力は意思決定における唯一の根拠ではなくなるため、MI は相対的に低く評価され得る。

CWD（Consequence of Wrong Decision）：入手可能な全ての情報に基づく誤った決定から生じる潜在的な悪影響。有害な結果の潜在的な重大性（Severity）と、有害な結果が発生する可能性（Likelihood）の両方を考慮して評価する。

High	誤った決定が、患者の安全性（健康被害）又は有効性に直接的な悪影響を及ぼす可能性がある、若しくは重要品質特性（CQA）を損なう結果として重大な規制措置（製品回収等）に繋がりうる。このような結果に至ることが、入手可能な全ての情報に照らして十分にあり得る。
Medium	High 又は Low に分類されないもの。

Low	誤った決定が、CQA や重要工程パラメータ（CPP）への影響も患者への直接的な影響も生じないか又はその可能性が低い。

（注）Severity が高い場合であっても、Likelihood が低い場合は、CWD を Medium と評価できる場合もある。

モデルリスク分類

	CWD High	CWD Medium	CWD Low
MI High	**High**	**High**	**Medium**
MI Medium	**High**	**Medium**	**Low**
MI Low	**Medium**	**Low**	**Low**

（注 1）上記マトリクスは通常の状況における判定の目安を示すものである。各区分の判定に際しては、その根拠を文書化することが求められる。

（注 2）誤判断の検出可能性（Detectability）は、モデルリスクを事後的に補正する要素ではない。ただし、独立した検証手段（例：工程内管理試験、最終製品試験等）が存在する場合、モデルの出力は意思決定における唯一の根拠ではなくなるため、当該検証手段の存在は COU の設定及び MI の評価において考慮され得る。なお、Detectability に関連する管理手段は、モデルの信憑性及び実運用上の受容性を判断する際のエビデンスとしても考慮される。この場合、その判断の正当化を文書化すること。

（注 3）リスク評価で決定したモデルリスクをモデル固有の特性であるモデルの複雑性に基づき、更にモデルリスクの補正が必要かを判断する。

モデルの複雑性（Model Complexity）

複雑性	モデルの例	モデルリスクの補正
High	ブラックボックス型モデル（深層学習等）	モデルリスク評価の上方補正検討（Low→Medium→High 向きの補正）
Medium	High 又は Low に分類されないもの	モデルリスク評価の維持
Low	PLS・PCR 等の線形モデル又は第一原理に基づく物理モデル	モデルリスク評価の下方補正検討可（High→Medium→Low 向きの補正）

（注）上記のクライテリアは判定の一貫性を高めるための例示的な基準であり、個々のモデルの特性や利用文脈によって柔軟に解釈されるべきものである。ハイブリッドモデルの複雑性は物理成分とデータ駆動型成分のうち、より支配的な成分の方の判定を採用する。

IV. モデルの信憑性（Credibility）評価に必要なモデルベリフィケーション&モデルバリデーション（V&V）計画

V&V の目的はモデルの信憑性（Credibility）を判断することであり、この信憑性評価はモデルリスクの如何に関わらず全てのモデルで実施する。モデルリスクが低い場合は、V&V の範囲・深度を縮小することができるが、評価自体を省略すべきではない。なお、以下に示した事項は、V&V 計画を立てる上で考慮すべき事項を例示したものであり、全ての項目が全てのモデルに必須となるわけではない。適用すべき項目はモデルの種類（データ駆動型/物理/ハイブリッド）、COU、モデルリスク及びモデルの複雑性に応じて選択される。

【一般項目】

- モデルの仮定の説明
- 「モデルの種類と構造」（モデルのアルゴリズムやアーキテクチャの概要）と「モデル選択の科学的根拠」（対象とする現象やデータの特性を踏まえた説明）の説明
- モデルの学習方法
- モデル構築に用いた学習/チューニングデータの種類・分布、データサイズ及びハイパーパラメータの説明
- データの前処理及び変数選択の妥当性の説明
- モデルの頑健性（robustness）及び限界事項（limitations）
- COU で定義したモデルの適用範囲に対するデータの網羅性（該当する場合にはデザインスペースを含む）

【モデルベリフィケーション項目】

- ソフトウェア品質保証
- ユーザー生成コード及び数値アルゴリズムが、モデルの数式、仮定及び仕様を正しく反映し、意図したとおりに機能することの検証（適用可能な場合には、既知解を用いた検証、ベンチマーク比較、独立した実装との比較、単体テスト、再計算等。）
- 計算誤差の評価。数値近似による解法を用いる場合には、その近似の精度の評価を含む。

【モデルバリデーション項目】

- テストデータの適切性（学習/チューニングデータとは独立）の説明
- 学習/チューニング/テストデータが COU を代表しているかの説明（分布、バイアス等）
- モデル評価手法（性能指標）
- モデル予測と観測データとの一致度

・　モデル予測の不確実性

【モデリング手法の種類に応じた項目】

データ駆動型モデルで重要な項目:

・　交差検証（Cross-validation）等の手法による汎化性能（モデルがテストデータに対してどの程度正確な予測又は意思決定を行うことができるか）の評価
・　モデルの解釈可能性及び説明可能性の評価

物理モデルで重要な項目:

・　モデルの基礎となる物理法則や化学反応式等の正当性の説明
・　モデル内パラメータを実験データから推定した際の方法と精度の説明
・　シミュレーションにおける計算誤差や安定性（値が発散せず（無限大又はマイナス無限大に向かわず）特定の値に収束することや許容できない振動をしないこと）の評価

ハイブリッドモデルで重要な項目:

上記のデータ駆動型モデル及び物理モデルで重要な項目の他、以下の通り。

・　データ駆動型モデル及び物理モデルの各々の役割
・　両方のモデルパーツが、同じデータセットでトレーニングされているのか、それとも異なるデータセットなのかの説明
・　物理モデルの全部又は一部をサロゲートモデルで代替する場合、サロゲートモデルが、COU に照らして参照物理モデルを適切な精度で再現していることの評価。

（注）バリデート済みのコンピュータ化システム（CSV 対応済みシステム）を使用する場合は、当該 V&V 計画におけるソフトウェア品質保証の要件は当該 CSV によって充足されているものとみなすことができる。その場合、開発者は希望のモデルが当該システムに正しく実装されたことを示すこと。また、オープンソースソフトウェアやライブラリについては、その使用実績等を考慮しつつ、使用したプログラミング言語、実行環境、ライブラリ名及びそれらのバージョン並びに適用した関数等を明示することで、手法の透明性及び再現性を確保すること。

V.　モデルが COU に対して適切であるか（モデルの信憑性）の判断

COU に対するモデルの予測結果及びその寄与の適切性について、あらかじめ設定した技術基準、V&V 結果及び UQ に基づき、当該モデルが当該 COU に対して適切であるかを判断するため、少なくとも以下の事項を説明すること。

・　あらかじめ設定した技術基準（V&V において確認すべき判定基準又は受入基準）

- V&V 結果の要約と上記技術基準への適合状況
- UQ の方法、前提、結果、予測への影響及びモデルの寄与への影響
- V&V 及び UQ により得られた証拠の、当該 COU への適用可能性（Applicability）の評価

以上を踏まえ、当該モデル及びそのモデルの予測が、当該 COU に照らして適切かつ受容可能であり、QOI への回答に利用し得るかに関する最終判断

（注）本文書における UQ とは、モデルの予測に内在する不確かさの源泉（例：測定誤差、パラメータ推定誤差、数値計算誤差等）となるものを考慮し、それらがモデル出力へ伝播した際の影響を定量的に評価することをいう。なお、必要とされる UQ の厳密さはモデルリスク及び COU に依存する。

VI. 文書化

上記の内容をまとめて、文書化する。その際、モデルの薬事規制上の位置付けを明確にする。

分類	内容の例
モデルが製造・意思決定に直接使用されず、プロセス理解・条件検討のみに使用される（薬事には結果として関わりうる）	開発段階での条件最適化、実験計画の効率化
モデルの出力が承認事項（制御戦略・デザインスペース等）の根拠に直接含まれる	リアルタイムリリース試験、リアルタイム工程管理（承認申請段階・承認後導入）
モデルが製造現場で GMP 管理下で稼働するが、承認事項には直接含まれない	逸脱調査支援、社内品質管理ツール（Pharmaceutical Quality System・GMP 範疇）

- モデルの出力が承認事項（制御戦略・デザインスペース等）の根拠に直接含まれる場合又は管理戦略の正当化を裏付けるために不可欠な場合は承認申請資料（Common Technical Document: CTD）に含める。
- モデルが製造現場の GMP 管理下で稼働するものの、承認事項には直接含まれない場合は、CTD に含める必要性は低いと考えられるものの、規制当局の求めに応じて当該資料を提示できるようにしておくこと。
- モデルリスクが High の場合は、以下の内容も含める：
  ① モデルの予測精度が時間経過と共に劣化していないか（モデルドリフト）を監視する計画。監視計画には、例えばモデル性能の継続的モニタリング頻度・方法及びモデルの異常又は性能劣化の検知方法が含まれ得る。モデルドリフトの監視はモデルリスク High に限らず、モデルリスク Medium においても製造工程への適用期間が長い場合には考慮することが望ましく、モデルドリフト

が確認された場合には、当該モデルの性能が COU に対して引き続き適切であるかを再評価する。評価結果に基づき、モデルの継続使用の可否、再学習又は再検証の要否を判断する。なお、ドリフトへの対応として COU の変更を伴う場合には、変更後の COU に基づきモデルリスクの再評価を行う。

② モデルの再学習や更新を行う際のトリガー、手順及び再検証の基準。

③ モデルの異常又は想定外の性能劣化が検出された場合の対応計画。対応計画には、例えば以下の要素が含まれ得る：

- モデルの使用停止の判断基準及び手順
- 代替判断又は従来法への切替え手順（フォールバック手順）
- GMP における品質システム（逸脱管理、変更管理等）との連携方法

Considerations for the Use of Models, Including AI and Machine Learning, in Pharmaceutical Manufacturing and Quality Control (CMC)

1. Introduction

In recent years, artificial intelligence (AI) technologies, including machine learning (ML), have increasingly been explored in pharmaceutical manufacturing, with a view also to their future use in marketing authorization applications. While these technologies may contribute to more efficient development and advanced assurance of product quality, questions have emerged regarding how to evaluate the performance and credibility of such models and how to explain them to regulatory authorities.

This document summarizes considerations for evaluating models when they are applied in practice in the pharmaceutical manufacturing and quality control (CMC) area, taking into account the general principles set forth in the ICH M15 Guideline "General Principles for Model-Informed Drug Development" (Step 4 document dated 29 January 2026). The models covered by this document include not only data-driven models such as ML, but also mechanistic models based on physical laws, chemical reaction mechanisms, or other scientific principles (referred to in this document as “physics-based models”), as well as hybrid models combining these.

The purpose of this document is to provide a standard workflow for consideration that enables developers/applicants to perform an appropriate risk assessment of their models in accordance with the Context of Use (COU), and to plan and execute the necessary Verification & Validation (V&V) activities. Through this, we aim to provide predictability regarding what should be verified and with what datasets in pharmaceutical manufacturing and in marketing authorization applications, and thereby to promote the utilization of AI technologies, including ML, in the pharmaceutical manufacturing and quality control (CMC) area.

2. Scope and Positioning

This document mainly covers models used in the pharmaceutical CMC area, such as manufacturing process development, process control, and quality testing. Whereas the ICH M15 Guideline provides a comprehensive framework for model evaluation, this document builds on that framework and presents simplified, practice-oriented considerations for applying it in CMC settings.

In the CMC area, data-driven models (e.g., chemometric models), physics-based models (e.g., Residence Time Distribution (RTD) models), and hybrid models combining these have long been used. In particular, for models based on well-established methodologies, a common understanding with regulatory authorities has been established in some areas regarding the regulatory points to be assessed.

Therefore, there may be cases where prior consultation with regulatory authorities based on a Model Analysis Plan (MAP), as described in the ICH M15 Guideline, is not necessarily required.

This document has been prepared with the aim of facilitating model evaluation and simplifying consultation with regulatory authorities. Although this document does not address the timing of consultation with regulatory authorities, it does not preclude such consultation. In addition, GMP requirements such as computerized system validation (CSV) and data integrity should be addressed by referring separately to the relevant notifications and guidelines.

The contents of this document should be interpreted flexibly according to the characteristics of individual models and technological progress, and are not intended as rigid rules to be uniformly applied to all cases. This document covers data-driven models (such as ML), physics-based models, and hybrid models. This document covers only models whose parameters remain fixed throughout operation. On the other hand, models whose model parameters are continuously or automatically updated during operation through learning or other means, or models whose outputs or behavior cannot be adequately specified and verified in advance, are outside the scope of this document.

3. Definitions of Terms

The terms used in this document for model classifications and for verification and validation are defined as follows.

＜Model Classifications＞

- Data-driven model

  A model that empirically learns the relationship between inputs and outputs based on observed data. It does not necessarily rely on prior assumptions about physical laws or chemical reaction mechanisms. Examples in this category include PLS regression, random forest, deep learning, etc. In particular, models whose internal structure is difficult to interpret and whose input-output relationship lacks transparency are referred to as black-box models.

- Physics-based model (also called a mechanistic model)

  A model constructed based on scientific principles such as physical laws, chemical reaction mechanisms, and material balances. Residence Time Distribution (RTD) models, reaction kinetics models, etc., are included in this category.

- Hybrid model

  A model that combines a physics-based model and a data-driven model. For example, this

category includes configurations in which parameters of a physics-based model are estimated using machine learning.

- Surrogate model

  In this document, a surrogate model refers to a data-driven model that is built using, as training data, a dataset obtained through simulations with a reference physics-based model that is difficult to use directly and repeatedly because of its computational load or evaluation time, or a data-driven model trained to mimic the behavior of a reference physics-based model. By approximating the input-output relationship of the reference physics-based model, it enables predictions at a lower computational load or in a shorter evaluation time. It is also referred to as a metamodel or an emulator. A surrogate model may be used on its own, or as a component of a hybrid model for the purpose of replacing the reference physics-based model, in whole or in part, within the defined scope of the Context of Use (COU).

＜Terms Related to Verification and Validation＞

- Model validation

  A task aimed at evaluating the robustness of the model and the adequacy of its performance.

- Model verification

  A task aimed at ensuring that user-generated code for processing the data and conducting the analysis (i.e., instructions written by the user of a programming language or software) is free of errors, that equations reflect the model assumptions, that the equations are correctly represented in the programming language or software, and that calculations are accurate.

- Analytical procedure validation

  Demonstrating that the analytical procedure is fit for its intended use.

- Analytical procedure verification

  Confirming that the analytical procedure consistently meets the objectives and goals (target profile) of analytical procedure development in the laboratory where it is routinely used.

- Process validation

  Verifying and documenting that the process operates under the acceptance conditions established by considering factors affecting product quality (e.g., physical properties of raw materials, operating conditions) based on the results of scale-up/industrialization studies and manufacturing experience with similar existing products, and that it is an appropriate process that consistently

yields products of the required quality.

・ Process verification

Confirming and documenting that the manufacturing procedures and related activities have produced the expected results.

＜Terms Related to Data＞

・ Training data

Data that is used to train the model and is not part of the test data.

・ Tuning data

Data that is used to select and optimize the model and is not part of the test data.

・ Test data

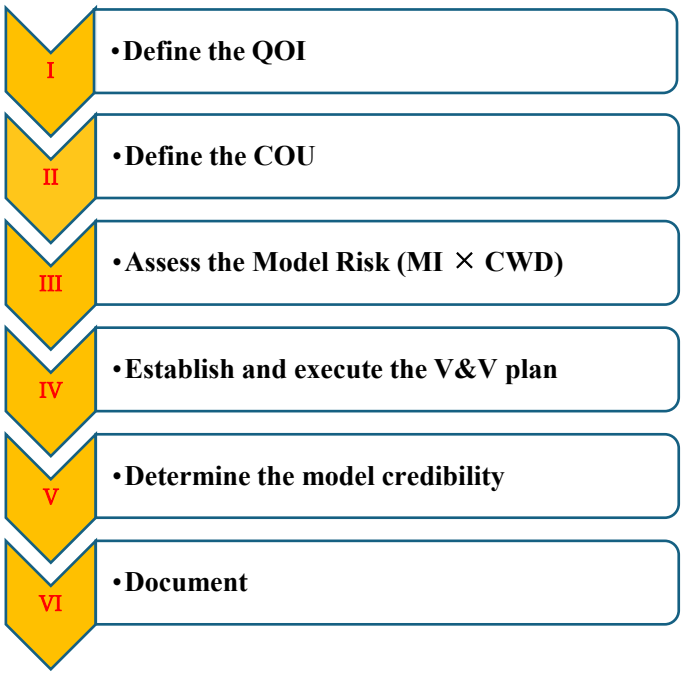
Data that is not used to train or tune the model, and is used after training and tuning to estimate the model's performance.

4. Model Evaluation Flow

The model evaluation flow is as follows:

① Define the Question of Interest (QOI).
② Define the Context of Use (COU).
③ Assess model risk based on Model Influence (MI) and Consequence of Wrong Decision (CWD).
④ Establish and execute a Model Verification and Validation (V&V) plan.
⑤ Based on the results of V&V and Uncertainty Quantification (UQ), determine whether the model is appropriate for the COU, i.e., assess model credibility.
⑥ Document the results.

Model Evaluation Flow



I. Definition of QOI

The question that the model is intended to answer.

II. Definition of COU

The role and scope of the model used to answer the QOI.

【Role】Describe what is to be modeled, how the model outputs are to be used, and the decision-making workflow, including who decides the next action and how, based on the model outputs.

【Scope】Specify the scope of application of the model based on the range of data used to build the model (range of input variables, types of products and processes, etc.). Also describe, from the perspective of the relative contribution of other evidence, whether other information (e.g., verification experiment data, manufacturing experience data, measured values obtained using analytical procedures) will be used together with the model outputs, or whether the model outputs will be the sole source of information.

III. Model Risk Assessment

Model risk is determined by combining the contribution of the model outputs to the overall body of evidence for a given decision (Model Influence: MI) and the consequence resulting from a wrong decision (Consequence of Wrong Decision: CWD).

MI: The intended weight of the model outputs in decision-making, after taking into account the contribution of additional data or evidence.

High	The model outputs serve as the dominant basis in decision-making in relation to other contributing evidence, and no independent evidence exists that can counter them.
Medium	Not classified as either High or Low.
Low	The model outputs are complementary information referenced as one of multiple independent pieces of evidence, and other evidence contributes to the decision with weight equal to or greater than that of the model outputs.

（Note）In the assessment of MI, consideration should be given not only to the model outputs but also to the presence of other independent evidence (e.g., in-process control tests, final product tests) and the extent to which such evidence contributes to the decision in question. When such independent means of verification exist, the model outputs are no longer the sole basis for the decision, and therefore MI may be evaluated as relatively low.

CWD: The potential adverse consequences of a wrong decision made based on all available information. It is assessed by taking into account both the potential severity of an adverse outcome and the likelihood that an adverse outcome will occur.

High	A wrong decision may directly cause adverse effects on patient safety (health hazards) or efficacy or may impair a Critical Quality Attribute (CQA), resulting in major regulatory actions (such as product recall). Such outcomes are sufficiently plausible considering all available information.
Medium	Not classified as either High or Low.
Low	A wrong decision does not affect, or has only a low likelihood of affecting, CQAs, Critical Process Parameters (CPPs), or patients directly.

（Note）Even when severity is high, CWD may be assessed as Medium in some cases if likelihood is low.

Model Risk Classification

	CWD High	CWD Medium	CWD Low
MI High	**High**	**High**	**Medium**
MI Medium	**High**	**Medium**	**Low**
MI Low	**Medium**	**Low**	**Low**

（Note 1）The above matrix provides a general guide for judgment under normal circumstances. When

making determinations for each category, the rationale should be documented.

（Note 2）Detectability of an incorrect decision is not a factor for retrospectively adjusting model risk. However, when independent means of verification (e.g., in-process control tests, final product tests) exist, the model outputs are no longer the sole basis for the decision, and therefore the existence of such means of verification may be considered in the setting of the COU and in the assessment of MI. Furthermore, control measures related to Detectability may also be considered as evidence when assessing the credibility of the model and its acceptability in actual operation. In such cases, the justification of such judgment should be documented.

（Note 3）Based on the model risk determined in the risk assessment, it should be determined whether model risk should be further adjusted based on model complexity, which is an inherent characteristic of the model.

Model Complexity

Complexity	Examples of Models	Adjustment of Model Risk
High	Black-box models (such as deep learning)	Consider upward adjustment of model risk assessment (in the direction of Low → Medium → High)
Medium	Not classified as either High or Low	Maintain the model risk assessment
Low	Linear models such as PLS/PCR, or physics-based models based on first principles	Downward adjustment of model risk assessment may be considered (in the direction of High → Medium → Low)

（Note）The above criteria are illustrative standards intended to enhance the consistency of judgment, and should be flexibly interpreted according to the characteristics of individual models and the context of their use. For the complexity of a hybrid model, the complexity rating of the dominant component, whether physics-based or data-driven, should be adopted.

IV. Model Verification and Validation (V&V) Plan for Assessing Model Credibility

The objective of V&V is to assess the credibility of the model, and this credibility assessment should be performed for all models regardless of the model risk. When the model risk is low, the scope and depth of V&V may be reduced, but the assessment itself should not be omitted. The items shown below are examples of matters to be considered when establishing a V&V plan, and not all items are mandatory for all models. Applicable items should be selected according to the type of model (data-driven / physics-based / hybrid), the COU, the model risk, and the complexity

of the model.

【General Items】

- Description of the model assumptions.
- Description of the "type and structure of the model" (an overview of the model's algorithm and architecture) and the "scientific rationale for model selection" (explanation based on the characteristics of the target phenomenon and data).
- Description of the model training method.
- Description of the type and distribution of the training and tuning data used to build the model as well as the data size and hyperparameters.
- Justification of data pre-processing and variable selection.
- Robustness and limitations of the model.
- Data coverage with respect to the model's scope of application defined in the COU, including the design space, where applicable.

【Model Verification Items】

- Software quality assurance.
- Verification that user-generated code and numerical algorithms correctly reflect the model equations, assumptions, and specifications and function as intended (including, where applicable, verification against known solutions, benchmark comparisons, comparisons with independent implementations, unit testing, recalculation, etc.).
- Assessment of computational error. When a numerical approximation method is used to solve the model, this includes assessment of the accuracy of the approximation.

【Model Validation Items】

- Description of the appropriateness of the test data (independent of the training and tuning data).
- Description of whether the training, tuning, and test data are representative of the COU (distribution, bias, etc.).
- Model evaluation methods (performance metrics).
- Degree of agreement between model predictions and observed data.
- Uncertainty of model predictions.

【Items According to the Type of Modeling Methodology】

<u>Important items for data-driven models:</u>

- Evaluation of generalization performance (the extent to which the model can make accurate

predictions or decisions on test data) using methods such as cross-validation.

- Assessment of model interpretability and explainability.

Important items for physics-based models:

- Justification of the physical laws, chemical reaction equations, etc., that form the basis of the model.
- Description of the method used to estimate model parameters from experimental data, and the accuracy of the estimates.
- Evaluation of computational error and stability in simulations to confirm that the numerical solution converges to a specific value without diverging (i.e., tending toward positive or negative infinity) or exhibiting unacceptable oscillations.

Important items for hybrid models:

In addition to the important items for data-driven models and physics-based models described above, the following also apply:

- Respective roles of the data-driven model and the physics-based model.
- Description of whether both model components are trained using the same dataset or different datasets.
- When the whole or a part of the physics-based model is replaced with a surrogate model, assessment of whether the surrogate model reproduces the reference physics-based model with adequate accuracy for the COU.

（Note）When a validated computerized system (a system for which computerized system validation (CSV) has been performed) is used, the software quality assurance requirements in the V&V plan may be considered to be satisfied by that CSV. In such a case, the developer should demonstrate that the intended model has been correctly implemented in that system. In addition, for open-source software and libraries, taking into account their track record of use, transparency and reproducibility of the methodology should be ensured by clearly specifying the programming language, execution environment, library names and their versions, and the functions used.

V. Determination of Model Appropriateness for the COU (Model Credibility)

To determine whether the model is appropriate for the COU with respect to the model predictions for the COU and the appropriateness of the model's contribution to the decision, based on the pre-defined technical criteria, V&V results, and UQ, at least the following matters should be described.

- Pre-defined technical criteria (acceptance criteria or decision criteria to be evaluated in V&V).
- Summary of V&V results and conformance to the above technical criteria.
- The UQ methods, assumptions, and results, including their impact on predictions and on the contribution of the model.
- Assessment of the applicability of the evidence obtained through V&V and UQ to the COU.

Based on the above, a final determination as to whether the model and its predictions are appropriate and acceptable for the COU and can be used to answer the QOI.

（Note）UQ in this document refers to quantitatively evaluating how sources of uncertainty in model predictions (e.g., measurement error, parameter estimation error, and numerical computation error) propagate to and affect model outputs. The required rigor of UQ depends on the model risk and the COU.

VI. Documentation

The above information should be compiled and documented. In doing so, the regulatory positioning of the model should be clearly stated.

Classification	Example of Contents
The model is not used directly in manufacturing or decision-making, and is used only for process understanding and condition exploration (it may be related to regulatory matters as a result)	Optimization of conditions during the development stage, efficiency improvement of experimental design
The model outputs are directly included as the basis of approved matters (control strategy, design space, etc.)	Real Time Release Testing, real-time process control (at the marketing authorization application stage or upon post-approval implementation)
The model operates under GMP control at the manufacturing site but is not directly included in approved matters	Deviation investigation support, internal quality control tools (within the scope of Pharmaceutical Quality System/GMP)

- When the model outputs are directly included as the basis of approved matters (control strategy, design space, etc.), or when they are indispensable to support the justification of the control strategy, they are to be included in the marketing authorization application dossier (Common Technical Document: CTD).
- When the model operates under GMP control at the manufacturing site but is not directly included in approved matters, there is generally less need to include it in the CTD; however, the relevant materials should be made available to be presented upon request by regulatory

authorities.

- When model risk is High, the following information should also be included:
  - ① A plan to monitor whether the model's predictive performance has deteriorated over time (model drift). The monitoring plan may include, for example, methods and frequency for continuously monitoring model performance and for detecting model anomalies or performance deterioration. Monitoring of model drift is desirable not only for models with High model risk but also for models with Medium model risk when the model is applied to the manufacturing process over an extended period. When model drift is identified, whether the model's performance remains appropriate for the COU should be re-evaluated. Based on the evaluation results, whether the model can continue to be used and whether retraining or revalidation is necessary should be determined. If the response to drift involves a change in the COU, model risk should be re-assessed based on the revised COU.
  - ② Triggers, procedures, and revalidation criteria for retraining or updating the model.
  - ③ A response plan to be followed if model anomalies or unexpected performance deterioration are detected. The response plan may include, for example, the following elements:
    - Criteria and procedures for deciding whether to discontinue use of the model.
    - Procedures for switching to alternative decision-making or conventional methods (fallback procedures).
    - Methods for coordination with the GMP quality system, including deviation management and change management.