

OECD 試験 / 評価シリーズ
No. 405

定量的構造活性相関モデル、予測および複数の予測から導かれる結果の
規制評価のためのガイダンス
第 2 版

OECD (2024), (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions - Second edition, OECD Series on Testing and Assessment No. 405, OECD Publishing, Paris.

この翻訳は日本動物実験代替法評価センター（JaCVAM）が情報提供のみを目的として作成したものであり、OECD はその正確性を保証するものではありません。唯一の公式バージョンは以下の OECD のウェブサイトで購入できる英語の文書です。

https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/11/q-sar-assessment-framework-guidance-for-the-regulatory-assessment-of-quantitative-structure-activity-relationship-models-and-predictions-second-edition_cc89955e/bbdac345-en.pdf

はじめに

2004 年 11 月、OECD の the Chemicals Committee（化学品委員会）と the Working Party on Chemicals, Pesticides and Biotechnology（化学・農薬・バイオテクノロジー作業部会）の第 37 回合同会議において、「規制目的のための(Q)SAR モデルの検証に関する OECD 原則」が合意された。

2007 年、OECD の同作業部会は、行政適用における(Q)SAR の使用に関して、「（定量的）構造活性相関 [(Q)SAR]モデルの検証に関するガイダンス文書」を公表した（OECD, 2007）。

これ以降、(Q)SAR の利用可能性については、化学物質に関する様々な規制に取り入れられてきた。これにより、様々な利害関係者が、様々な規制目的のために(Q)SAR を評価する実務経験を得ることができた。

これらの経験において、（訳注：規制目的ごとに要件が異なることから、）科学的妥当性が検証された(Q)SAR モデルによる予測が必ずしも、あらゆる規制目的に受け入れられるわけではないことが浮き彫りになった。このことから、「(Q)SAR 予測」または「複数の予測から導かれる(Q)SAR 結果」を意図する規制目的で使用する場合には、特定の規制へ適用するという要件ごとに確認する必要がある。しかし、(Q)SAR モデルの評価に関する原則はすでに合意されているものの、規制評価のための「(Q)SAR 予測」および「複数の予測から導かれる(Q)SAR 結果」について、一般的に合意された原則を確立する必要があった。

2020 年終わり、OECD Working Party on Hazard Assessment（WPHA）において、ISS（Istituto Superiore di Sanita、イタリア国立衛生研究所）がリードする(Q)SAR 評価フレームワーク（QAF）を作成するプロジェクトが提案され、2021 年初めに OECD のワークプランに追加された。プロジェクトには、チェックリスト形式による(Q)SAR モデルと「(Q)SAR 予測」の評価についての原則、チェックリストにおける要素の評価基準の策定が含まれる。さらに、ISS に加えて、ECHA（European Chemical Agency、欧州化学機関）がプロジェクトを共同してリードすることとなった。

WPHA 代表への専門家の指名要請を経て、QAF への意見提供とレビューを行うために 40 名以上の専門家が招集された。2021 年から 2023 年にかけて QAF 専門家グループによる電話会議を複数回行い、そのサブグループが草案の作成を行った。2022 年 11 月には、草案の最終化に向け QAF 専門家グループの対面会議が開催された。

QAF 第 1 版は 2023 年 8 月に公開された。本文書において、OECD QAF 専門家グループは、「(Q)SAR 予測」および「複数の予測から導かれる(Q)SAR 結果」の評価に関する OECD の原則を確立し、(1) 「(Q)SAR モデル」、(2) 「(Q)SAR 予測」、および (3) 「複数の予測から導かれる(Q)SAR 結果」の評価を実際に行うチェックリストについて合意した。

2024 年に、ECHA と EFSA（European Food Safety Authority、欧州食品安全機関）の主導で OECD QRRF 専門家グループが招集され、QAF 第 2 版に添付するための(Q)SAR 結果報告様式（(Q)SAR Result Reporting Format、QRRF）を作成した。

本文書は、OECD の the Chemicals and Biotechnology Committee（化学・バイオテクノロジー委員会）により出版される。

(Q)SAR Assessment Framework (Q)SAR 評価の枠組み：

定量的構造活性相関モデル、予測および複数の予測から導かれる結果の規制評価のためのガイダンス 第 2 版

OECD 2024

目次

はじめに	1
エグゼクティブ・サマリー（要旨）	3
ビジュアル・アブストラクト(概要図)	5
1.(Q)SAR モデルの評価（モデルチェックリスト）	6
1.1 エンドポイントの定義	6
1.2 曖昧さのないアルゴリズム	8
1.3 適用範囲の定義	9
1.4 適合度、頑健性、予測性の適切な評価	9
1.5 メカニズムに関する解釈	10
1.6 (Q)SAR モデルの評価結果	11
1.7 (QSAR モデル評価の結論	11
2. 個々の(Q)SAR 予測の評価（予測チェックリスト）	12
2.1 モデルへの入力正しいこと	12
2.2 モデルの適用範囲内にある物質であること	13
2.3 信頼できる予測であること	14
2.4 規制目的に適した出力であること	17
2.5 個々の予測に関する評価の結論	18
3. 「複数の予測から導かれる(Q)SAR 結果 ¹⁾ 」の評価（結果チェックリスト）	20
3.1 結果チェックリストを使用する場面	20
3.2 (Q)SAR 結果の不確実性と評価結果	21
4.最終的な考察	22
付録Ⅰ (Q)SAR モデル報告様式 (QMRF) v.2.1	23
付録Ⅱ (Q)SAR 予測報告様式 (QPRF) v.2.0	24
付録Ⅲ (Q)SAR 結果報告様式 (QRRF) v.1.0	25
用語集	26
図	
図 1 個々の予測に基づく(Q)SAR 評価フレームワーク (QAF)	6
図 2 複数の予測に基づく(Q)SAR 評価フレームワーク (QAF)	6

エグゼクティブ・サマリー（要旨）

(Q)SAR 評価フレームワーク (QAF) の目的は、(1) 「(Q)SAR モデル」、(2) 「(Q)SAR 予測」、および (3) 「複数の予測から導かれる (Q)SAR 結果¹」についての規制評価のための、体系的で調和のとれたフレームワークを作成することである。提案される評価法は、(Q)SAR モデルを作成する際に用いられたモデル化の方法、予測エンドポイントおよび規制目的に関係なく適用されることを意図している。本文書の主な対象は、規制当局およびその利害関係者である。加えて、その他の(Q)SAR 使用者が規制目的で(Q)SAR を用いる際にも、このフレームワークを参照することが推奨される。

科学的妥当性が検証された(Q)SAR モデルであっても、ある条件下では受け入れられない予測を行うことがありうるため、規制目的の(Q)SAR 評価は、(Q)SAR モデルの検証に限定すべきではなく、個々の「(Q)SAR 予測」および「複数の予測から導かれる (Q)SAR 結果¹」それぞれに専用の評価が必要である。この目的を満たすため、QAF は OECD の(Q)SAR モデルの検証に関する原則²（OECD、2007：以降は「OECD の(Q)SAR モデル原則」と記載）に基づく。また、「(Q)SAR 予測」と「複数の予測から導かれる (Q)SAR 結果¹」を評価するための原則を、新たに定義した。1) モデルへの入力正しいこと、2) モデルの適用範囲（Applicability Domain: AD）内にある物質であること、3) 信頼できる予測であること、4) 規制目的に適した出力であることの 4 原則（以降は「OECD の(Q)SAR 予測原則」と記載）を確立した。評価を効率化するため、各原則を、評価において考慮すべき要素（評価要素、Assessment Element: AE）に分割した。AE は 3 種類のチェックリスト（モデルチェックリスト、予測チェックリスト、結果チェックリストとして個別文書として添付）に盛り込まれ、実際に(Q)SAR を使用する際の受け入れ可能性評価のために使用できる。各々の AE は、満たす、満たさない、記載なし（文書化されていない）、非該当/評価しない、のいずれかとなる。またチェックリストの別表では、個々の AE についての詳細や評価例も提供されている。

モデルチェックリストは、OECD の(Q)SAR モデル原則に従って(Q)SAR モデルを評価する AE のリストで構成される。個々の「(Q)SAR 予測」あるいは「複数の予測から導かれる (Q)SAR 結果¹」（訳注：原文では単に「(Q)SAR 結果」として定義、記載しているが、本文書中では「複数の予測から導かれる (Q)SAR 結果」と記載する。）を評価する際には、モデルチェックリストを他のチェックリストとともに使用すべきである。この場合、受け入れ可能な(Q)SAR モデルの使用が評価の最初のステップと考えられる。(Q)SAR モデルが受け入れ可能と判断されるときは、他のチェックリストをさらに考慮する必要がある。(Q)SAR モデルが受け入れられないと判断されるときは、予測および結果を考慮せずに評価を終了できる。あるいは、例えば(Q)SAR がデータベースのスクリーニングに使用され、個々の予測を一つずつ評価しない場合、または将来再度利用される(Q)SAR モデルを評価する個別記録として保存する場合は、モデルチェックリストは独立したツールとして利用できる。(Q)SAR モデルの評価はその規制目的に特化しており、同じ(Q)SAR モデルを他の規制目的で評価するときには改めて評価を実施すべきである。

¹ 複数の予測から導かれる (Q)SAR 結果: 複数の (Q)SAR 予測を統合して導出した、物質の特性についての評価結果。

² OECD principles for the Validation, for Regulatory Purposes, of (Q)SAR Models, ENV/JM/MONO(2007)2 <https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/assessment-of-chemicals/oecd-principles-for-the-validation-for-regulatory-purposes-of-quantitative-structure-activity-relationship-models.pdf> (OECD, 2007)

予測チェックリストは個々の「(Q)SAR 予測」を、結果チェックリストは「複数の予測から導かれる(Q)SAR 結果¹」を評価するために使用される。これらのチェックリストは、OECD の(Q)SAR 予測原則に基づく AE から構成され、これらの AE は、評価における重要度によって異なる重み付けがなされている。各 AE について提案されているデフォルトの重みは、予測チェックリストおよび結果チェックリストに示されているが、評価者が個々の規制の枠組みあるいは方法論に適した重み付けに変更できる。さらに、本ガイダンスおよびチェックリスト別表の評価例に準じて、各 AE の不確実性に半定量的な値（低・中・高）を割り当てることができる。最終的に、予測全体の不確実性は、個々の AE の不確実性と評価における重み付けを考慮して決定される。使用目的および個々の予測の不確実性レベルに基づいて、評価者は評価結果について結論づける。

「複数の予測から導かれる(Q)SAR 結果¹」では、各予測の個別評価に加え、結果チェックリストにおいて検討される AE が 1 つ追加され、最終的な結果を確定するために個々の予測が正しく統合されているかについて評価する。最終的な結果の不確実性のレベルは個々の予測および追加の AE の不確実性の重みづけから割り当てられる。この場合も、受け入れ可能な不確実性のレベルを決定し、結果の使用目的に応じた評価結果、つまり、意図する使用目的に受け入れ可能かを評価者が結論づける。

QAF を補完するために、(Q)SAR モデル報告様式 (QMRF) および(Q)SAR 予測報告様式 (QPRF) を更新した。QMRF では、フィールドの名前および順番は変更せずに、各フィールドに期待される情報の記載のみを更新した。一方、QPRF は、新規に確立した OECD の(Q)SAR 予測原則を反映して大幅に更新された。OECD QRRF プロジェクトでは、複数の予測から導かれる(Q)SAR 結果¹のための(Q)SAR 結果報告様式が専門家により作成された。QMRF、QPRF の最新版および QRRF のテンプレートは、本文書の付録として添付した。

さらに、チェックリストの使用を解説する例も作成され、別文書として提供されている。専門家グループは、(Q)SAR アプローチを含むケーススタディが実施する IATA ケーススタディプロジェクトにおける QAF 原則の更なる使用あるいは応用を推奨している。

最後に、グループの一部の専門家により、(Q)SAR モデルの外部予測性の計測方法に関する更なるガイダンスの必要性が指摘され、QAF プロジェクトの終了に向けて討議された。しかし、これは、外部予測性の計測方法について科学的な合意が得られなかったことから QAF の最終版にこのトピックについて詳述していない。これは、このプロジェクトの目的範囲外の広範な文献調査が必要なためである。さらに、(Q)SAR モデルの検証はモデル開発者の作業であり、規制当局が既存の(Q)SAR モデルを評価するための手順の提供を主目的とする QAF の範囲外である。

評価チェックリスト、QMRF、QPRF、QRRF は QAF³の“Supporting Material”タブからダウンロードすることができる。

³ https://www.oecd.org/en/publications/q-sar-assessment-framework-guidance-for-the-regulatory-assessment-of-quantitative-structure-activity-relationship-models-and-predictions-second-edition_bbdac345-en.html

ビジュアル・アブストラクト(概要図)

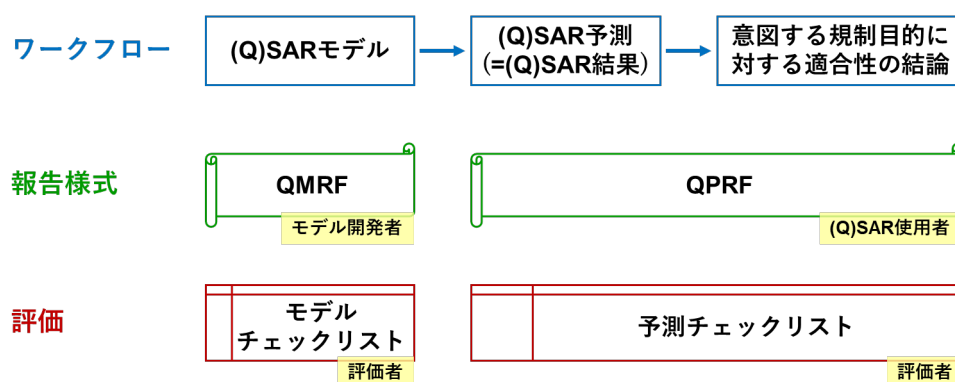


図1 個々の予測に基づく(Q)SAR 評価フレームワーク (QAF)

注：個々の予測に基づく評価における OECD (Q)SAR 評価フレームワーク (QAF) に従う (Q)SAR 情報ワークフロー。(Q)SAR モデルの情報はモデル開発者が作成した (Q)SAR モデル報告様式 (QMRF) により報告され、規制当局が QAF モデルチェックリストを使用して評価する。「(Q)SAR 予測」の情報は、(Q)SAR 使用者が (Q)SAR 予測報告様式 (QPRF) により報告し、規制当局が QAF 予測チェックリストを使用して評価する。評価者の作業を容易にするため、(Q)SAR 使用者がチェックリストを事前に作成することもできる。

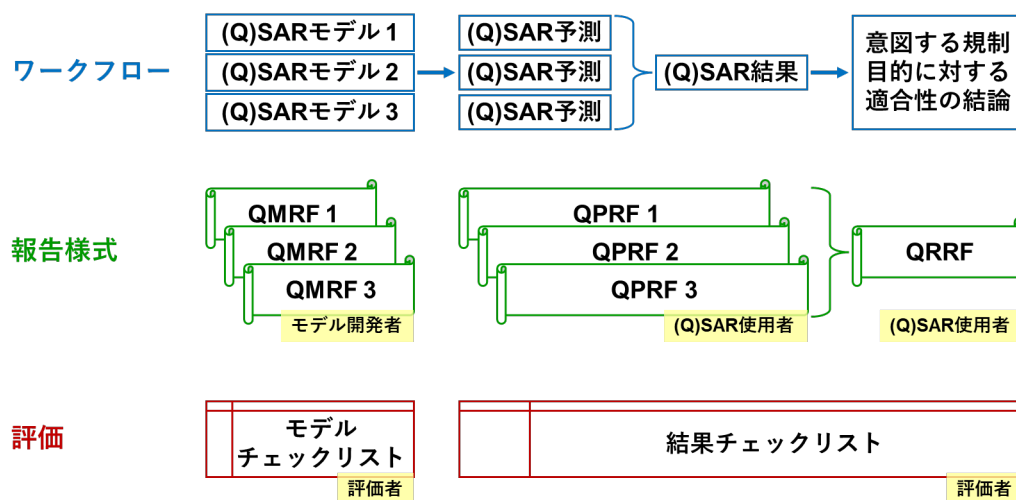


図2 複数の予測に基づく(Q)SAR 評価フレームワーク (QAF)

注：複数の予測に基づく評価における OECD (Q)SAR 評価フレームワーク (QAF) に従う (Q)SAR 情報ワークフロー。(Q)SAR モデルの情報は、モデル開発者が作成した (Q)SAR モデル報告様式 (QMRF) により報告され、規制当局が QAF モデルチェックリストを使用して評価する。「(Q)SAR 予測」および「複数の予測から導かれる (Q)SAR 結果¹⁾」(訳注：図中「(Q)SAR 結果」と同義)については、(Q)SAR 使用者が (Q)SAR 予測報告様式 (QPRF) により報告し、規制当局が QAF 結果チェックリストを使用して評価する。

(Q)SAR Assessment Framework (Q)SAR 評価の枠組み：

定量的構造活性相関モデル、予測および複数の予測から導かれる結果の規制評価のためのガイダンス 第2版

OECD 2024

1.(Q)SAR モデルの評価（モデルチェックリスト）

本章では、モデルチェックリストを完成させるための実践的なアドバイスを示す。これは、より理論を中心とした OECD の(Q)SAR モデル原則に関するガイダンス（OECD、2007）に基づき、それを補完するものである。

OECD の(Q)SAR モデル原則では、以下のように述べられている：

「(Q)SAR モデルを規制目的で検討することを促進するためのもので、以下の項目と関連付けなくてはならない：

1. エンドポイントの定義
2. 曖昧さのないアルゴリズム
3. 適用範囲（AD）の定義
4. 適合度、頑健性、予測性の適切な評価
5. 可能ならば、メカニズムに関する解釈」

以降の項およびモデルチェックリストでは、これらの原則のさらなる詳細および、(Q)SAR モデルがこれらを順守していることの確認方法について述べる。評価は、(Q)SAR モデル報告様式（QMRF）の情報に基づくべきである。モデルチェックリストの別表には評価要素（AE）と QMRF 項目の対応表が示されており、評価に関連する情報と照合できる。以下の各節の末尾で示される通り、OECD の(Q)SAR モデル検証の各原則は、予測チェックリストおよび結果チェックリストのいずれかの AE においても考慮される。

1.1 エンドポイント⁴の定義

OECD の(Q)SAR モデル原則 1（OECD, 2007）によると、(Q)SAR は「定義されたエンドポイント」に関連付けられるべきである。エンドポイントとは測定可能な物理化学的、生物学的または環境影響に関する特性で、モデル化できるものを指す。多様な実験プロトコルや条件の下でエンドポイントが決定されることから、この原則では与えられた(Q)SAR モデルで予測されるエンドポイントの透明性を確保することを目的としている。

モデルチェックリストでは、エンドポイントが明確に定義されていることを確認するために、次の AE が含まれる。

1. 科学的/規制上の明確な目的
2. 基盤となる実験データの透明性
3. 基盤となる実験データの品質

科学的/規制上の明確な目的（モデルチェックリスト AE 1.1）

⁴ 用語集で説明しているように、本文書では、モデルが予測するエンドポイントを指す用語として「特性」を使用する。ただし、「OECD の(Q)SAR モデル原則 1：エンドポイントの定義」に精通している読者の混乱を避けるため、この章では元の用語「エンドポイント」をそのまま使用する。

科学的目的が明確であるためには、予測される特性が正確に記述されていなければならない。規制目的が明確であるためには、(Q)SAR モデルは特定の試験方法や試験ガイドラインと関連付けられる特定の規制要件に対応するか、規制要件を裏付ける情報（例えば、メカニズムに関する情報）を提供すべきである。予測される特性の記述は可能な限り詳細であるべきで、検討したすべての要素（例えば、測定単位、タイムスケール、成長や死亡率といった観察値など）を含む。予測される特性の複雑さは必要な文書化の程度に影響する。すなわち、Ames 試験における *in vitro* 変異原性のようなより単純な特性を予測する(Q)SAR モデルと比較して、発生毒性のようなより複雑な特性を予測する(Q)SAR モデルでは、より詳細な特性についての定義が要求される。

基盤となる実験データの透明性（モデルチェックリスト AE 1.2）

この AE は、基盤となる実験データの透明性および関連するデータの選択およびキュレーション（訳注：収集したデータを目的に合わせて整理し、解釈を行うこと）手順の透明性に関するものである。実験データの情報源は適切に報告されるべきで、実験データの選択基準、データ処理および試験された物質の化学識別子（InChI/InChIKey や（正規表現された（Canonical））SMILES のような少なくとも 1 つの化学構造をコードする分子識別子、および CAS 登録番号のような一般的に報告される情報を含む）も同様である。データ選択における潜在的なバイアス（例えば、予測されるエンドポイントに関係しない試験法ガイドラインで測定されたデータがトレーニングセットに体系的に含まれる、など）についても調査すべきである。元（オリジナル）の研究結果（またはアクセス可能な文献）が参照可能であれば、透明性は最も高くなるが、そのようなケースは稀である。逆に、機密保持や文書化が不十分なために、基盤となるデータがまったく参照できない(Q)SAR モデルもある。多くの既存の(Q)SAR モデルでは、透明性レベルはこれら両極端なケースの間にあり、(Q)SAR モデル構築に使用した実験研究のすべてではないが、一部の詳細が参照可能である。

評価を担当する規制当局は、いくつかの(Q)SAR モデルは、例えば商業的な理由から参照可能な情報が限られることなどを理解した上で、特定の規制目的に必要な透明性の最低レベルを決定することができる。一般に、基盤となる実験データまたはデータキュレーション手順について、データの品質の評価に十分な情報があるべきである。

基盤となる実験データの品質（モデルチェックリスト AE 1.3）

(Q)SAR モデルは十分な品質のデータで構築すべきである。しかし、各データポイントを個別に評価することは不可能なことが多い。このような場合、データキュレーション手順の説明に基づいて実験データの品質を評価できる。例えば、評価者は、実験結果に影響を与える可能性がある実験パラメーター（例：性別、生物種、温度、曝露期間、プロトコル）が、モデル構築のためのデータ選択に際して、どのように考慮されているかを確認できる。評価者は、(Q)SAR モデルの開発および検証に用いられたすべてのデータポイントが、1) 同じアッセイプロトコル、2) 最新アッセイプロトコルで行われたものか、信頼性が及ぼす影響についても検討できる。個別データの品質も、可能な範囲で評価されるべきである。

OECD の(Q)SAR モデル原則 1 は、予測チェックリストおよび結果チェックリストの要素「4.2 予測される特性は規制要件となる特性と対応しているか」においてさらに検討される。

1.2 曖昧さのないアルゴリズム

OECD の(Q)SAR モデル原則 2 (OECD, 2007) によれば、(Q)SAR モデルは曖昧さのないアルゴリズム（曖昧さのない明確に記載されたアルゴリズムを意図する）の形で表現されるべきである。この原則が意図するものは、その予測値を独立して再現できるように、(Q)SAR モデルのアルゴリズムの記載の透明性を保証することである。

モデルチェックリストでは、曖昧さのないアルゴリズムの原則を確認するために、以下の AE が含まれる。

1. アルゴリズムおよび/またはソフトウェアの説明
2. 入力とその他のオプション
3. モデルのアクセス性

アルゴリズムおよび/またはソフトウェアの説明（モデルチェックリスト AE 2.1）

最初にチェックすべき要素は、アルゴリズムに関する透明性の高い説明が入手可能かである。モデル式を用いる(Q)SAR モデルの場合、モデル式のすべての記述子とその選択に用いたアプローチを含めて、説明が提供されるべきである。さらに、フラグメントベースの(Q)SAR モデルであれば、フラグメント/構造アラートのリスト（例：活性の有無や（訳注：活性を阻害する）マスク作用など）とその説明が提供されるべきである。それらの特定に至った論理的根拠も含めるべきである。

計算で値を求めた記述子は、その計算に使用したソフトウェア名とバージョンが示されなくてはならない。さらに、(Q)SAR モデルについて、バージョン、開発者の連絡先およびソフトウェアに関するすべての入手可能な説明が提供されるべきである。アルゴリズム自体の記述が公開されていない場合（例えば市販の(Q)SAR モデル）、すべての入手可能な関連情報について評価すべきである。

入力とその他のオプション（モデルチェックリスト AE 2.2）

次に、評価者は、（訳注：(Q)SAR モデルへの）入力および(Q)SAR モデルの設定に関する記述が文書に含まれるかを確認すべきである。適用可能な前処理手順（例えば、塩や互変異性体について）を含む、化学構造とその記述子に関する許容（または推奨）される入力形式が記載されているべきである。さらに、カスタマイズ可能なオプション/設定についての報告および説明もなくてはならない。特に正当な理由がない限り、推奨される入力形式とオプションは、モデル開発者が(Q)SAR モデルの開発および性能評価に使用したものと同一であることが期待される。

モデルのアクセス性（モデルチェックリスト AE 2.3）

最後に、その(Q)SAR モデル（バージョン）が公にアクセス可能かを確認すべきである。QMRF の記述には、(Q)SAR モデルにアクセスまたはダウンロードするための有効なリンクの記載が期待される。評価者が評価対象の(Q)SAR モデルの異なるバージョン（より新しいバージョン、など）にアクセスできる場合には、アウトプットの最終的な差異について調査すべきである。

OECD の(Q)SAR モデル原則 2 は、予測チェックリストおよび結果チェックリストの要素「3.1 再現性があるか」においてさらに検討される。留意点として、評価者がアクセスできるソフトウェアプログラムに(Q)SAR モデルが実装されている場合には、アルゴリズムの説明の完全な開示がなくとも予測結果の再現性を確認できる。評価者は、一定の規制上の使用目的によっては結果の受け入れが可能と判断す

ることもある。

1.3 適用範囲の定義

OECD の(Q)SAR モデル原則 3 (OECD, 2007) により、科学的妥当性が検証された(Q)SAR モデルは定義された適用範囲 (Applicability Domain: AD) に関連付けられる。OECD ガイダンス文書 (OECD, 2007) に記載されるように、(Q)SAR モデルの AD は「(Q)SAR モデルが既定の信頼性で予測できる応答空間および化学構造空間」である。この AD の定義を詳細に説明するには、AD は(Q)SAR モデルのパラメーター、構造、メカニズム、代謝および応答について考慮しなくてはならない。しかしながら、複数の科学的に妥当な方法論が使用できうるため、QAF では AD を定義する特定の方法を規定しない。ガイダンス文書 (OECD, 2007) では AD についてさらに科学的な観点から記載されているが、ここでは QAF 内での評価に関する実際的な観点到焦点を当てる。

モデルチェックリストには、適用範囲に関連する以下の 1 つの AE が含まれる。

1. 適用範囲の明確な定義およびモデルの制約

適用範囲の明確な定義およびモデルの制約 (モデルチェックリスト AE 3.1)

評価者は、予測チェックリストおよび結果チェックリストにおいて、対象物質と(Q)SAR モデルの AD との関係性 (例えば、その物質は(Q)SAR モデルの AD 内にあるか) について評価する上で、モデル開発者が提示した定義が十分に詳細であることを確認すべきである。評価を容易にするために、新しい(Q)SAR モデルの開発者には、入力物質と AD との関係性が予測レポートに自動的に含まれる機能の実装が推奨される。さらに、予測チェックリストおよび結果チェックリストには、AD に含まれるか否かに関わらず、結果の信頼性に影響しうる重要な点について検討する別の信頼性評価も含まれる。これらは、2.3 項で後述する。

(Q)SAR モデルの説明には、AD の定義に加え、使用が推奨されない物質クラスなど、(Q)SAR モデルの制約事項のリストが含まれる場合がある。この場合も、QAF ではそれらの制約の定義方法について規定しないが、入力物質とこれらの制約との関係性を評価するために、追加的な制約の記載が十分に詳細であることを確認すべきである。

OECD の(Q)SAR モデル原則 3 は、予測チェックリストおよび結果チェックリストの要素「2.1 物質が適用範囲内にあるか」および「2.2 モデルに関するその他すべての制約が考慮されているか」において、さらに検討される。

1.4 適合度、頑健性、予測性の適切な評価

OECD の(Q)SAR モデル原則 4 (OECD, 2007) によると、(Q)SAR は「適合度、頑健性、予測性についての適切な評価」と関連付けられる。この原則は、内部検証で決定される(Q)SAR モデルの適合度と頑健性および外部検証で決定される(Q)SAR モデルの予測性に関する情報を提供する必要性を示している。AD とモデル性能は関連する。モデル性能は開発者が定義した適用範囲内において計測すべきである。一般に、AD が狭いほど高いモデル性能が得られる。では、原則 4 に関するさらに科学的な観点について、OECD ガイダンス文書 (OECD, 2007) を参照することができる。

モデルチェックリストには、(Q)SAR モデルの適合度、頑健性、予測性の評価法が適切かを確認するた

(Q)SAR Assessment Framework (Q)SAR 評価の枠組み：

定量的構造活性相関モデル、予測および複数の予測から導かれる結果の規制評価のためのガイダンス 第 2 版

OECD 2024

め、以下の AE が含まれる。

1. 適合度と頑健性
2. 予測性

適合度と頑健性 / 予測性 (モデルチェックリスト AE 4.1 および 4.2)

評価者は、(Q)SAR モデルの内部性能と外部性能に関する入手可能な情報が、新しい物質の予測の際に (Q)SAR モデルに期待される正確度の指標として使用できることを確認すべきである。トレーニングセット、テストセットおよび外部検証セットのサイズ、使用する統計手法および統計指標、統計指標の値、性能測定手順の透明性は、(Q)SAR モデルの性能を評価する際に考慮すべき重要な点である。トレーニングセット、テストセットおよび外部検証セットは独立しているべきである。精査対象の (Q)SAR モデルがどのように選択されたか、および予測性能がどのように推定されたかを示す情報について評価しなければならない。これにより (Q)SAR モデルが、外部予測性推定用の情報を用いず正しくトレーニングされたかをチェックする。例えば、本来は外部予測性評価専用のはずの外部検証セットの物質の情報を、(Q)SAR モデルの開発に使用する ((Q)SAR モデルの調整や選択のためのテストセットからの「データ漏出」) と、(Q)SAR モデルの外部予測性の過大評価に繋がる可能性が非常に高い。

OECD の (Q)SAR モデル原則 4 は、予測チェックリストおよび結果チェックリストの要素「3.2 モデルの全体的な性能がどの程度か」において、さらに検討される。

1.5 メカニズムに関する解釈

OECD の (Q)SAR モデル原則 5 (OECD, 2007) によれば、(Q)SAR は、「可能ならばメカニズムに関する解釈と関連付けられるべきである」。化学構造と活性の相関を記述する統計的手法は、化学および毒理学における他の知識がある場合、その知識を置き換えることを意図してはいない。評価者は、(Q)SAR モデルの背後にある理論的根拠が、予測される特性に関連する知識とどのように整合またはそれを説明するか (例えば、予測される特性に関連する既知の有害性発現経路 (AOP))、すなわちメカニズムに関する解釈 (訳注: 既知のメカニズムに関する知識による、予測の説明や裏付け) を (Q)SAR モデルの文書化に含めることを要求することができる。さらにトキシコキネティクスが対象とする特性に関連するのであれば、これもメカニズムに関する解釈に内包される。

モデルチェックリストには、メカニズムの解釈に関連する以下の AE が含まれる。

1. メカニズムに関する解釈の尤もらしさ

メカニズムに関する解釈の尤もらしさ (モデルチェックリスト AE 5.1)

フラグメント/アラートベースの (Q)SAR モデルでは、アラートに関連付けられたフラグメントによって惹起あるいは阻害される化学反応性または分子間相互作用に基づいて、メカニズムについての解釈を説明することができる。数式ベースの (Q)SAR モデルでは、各記述子の物理化学的な解釈およびメカニズムとの関連性に基づいて、メカニズムについて説明することができる。(Q)SAR モデルにおけるメカニズムの根拠が、先に決められた (例えば、特定のメカニズムに合わせて予め記述子やフラグメントを選択した) のか、あるいは後から決められた (例えば、モデル化後に最終的なトレーニングセットの構造および/または記述子から解釈した) のかについても言及すべきである。

OECD の(Q)SAR モデル原則 5 は、予測チェックリストおよび結果チェックリストの要素「3.5 メカニズムおよび/または代謝が考慮されているか、ある場合どのような内容か」において、さらに検討される。

1.6 (Q)SAR モデルの評価結果

各 AE と(Q)SAR モデルの評価は、全般に、意図する規制目的に依存する（かつ特異的である）。

各 AE が「満たす」場合、あるいは予測チェックリストおよび結果チェックリストで(Q)SAR モデルに関連する AE を評価するために、(Q)SAR モデルに関連する十分な情報が得られている場合には、その(Q)SAR モデルは受け入れ可能と判断する。特定の規制目的や、満たさない要素の正当性について根拠がある場合には、すべての AE が満たされなくても受け入れ可能とすることもできる。

1.7 (Q)SAR モデル評価の結論

モデルチェックリストの記入は、(Q)SAR 予測や「複数の予測から導かれる(Q)SAR 結果¹」を評価する最初のステップである。また、それは予測チェックリストおよび結果チェックリストで(Q)SAR モデルの科学的妥当性に関係する AE を評価する際の補助となる。モデルチェックリストは再利用可能であり、新規予測が提出される度に評価者が(Q)SAR モデルを再評価する必要はない。さらに、予測を個別に評価することが意図されないデータベースのスクリーニングに使用される(Q)SAR モデルを評価する際には、モデルチェックリストは独立したツールとして使用できる。

最後に、モデルチェックリストは、別表として QMRF の各欄との対応表を含む。この対応表が公開されれば、モデルチェックリストにおける評価結果とともに、モデル開発者へのフィードバックとして、(Q)SAR モデルと関連文書の改良に繋がる。

2. 個々の(Q)SAR 予測の評価（予測チェックリスト）

(Q)SAR 予測および「複数の予測から導かれる(Q)SAR 結果¹⁾」を評価するためには、(Q)SAR モデル原則に加えて、予測の規制評価のための原則を確立する必要がある。QAF では、規制当局による(Q)SAR 予測および「複数の予測から導かれる(Q)SAR 結果¹⁾」の評価のための 4 原則を定める。

1. モデルへの入力正しいこと
2. モデルの適用範囲内にある物質であること
3. 信頼できる予測であること
4. 規制目的に適した出力であること

予測チェックリストおよび結果チェックリストにはこの 4 原則が含まれる。本章と予測チェックリストおよび結果チェックリストにより、これらの原則についての詳細や、この原則の順守についての確認方法が示されている。評価は QMRF および QPRF の情報に基づいて行われるべきである。本章では予測チェックリストと結果チェックリストに共通する AE について述べ、次章において結果チェックリストで検討する追加の要素を取り上げる。

2.1 モデルへの入力正しいこと

(Q)SAR モデルでは、予測のために 1 つ以上の入力を行う必要がある。入力が物質の構造情報のみか、または、さらにいくつかの記述子を含むかなどは(Q)SAR モデルごとに異なる。また一部の(Q)SAR モデルでは、ソフトウェア内の設定をカスタマイズすることも可能である。

入力内容が明確で完全に記述されており、それが評価対象物質を代表的に表し（訳注：当該物質の評価に適した予測が得られる入力を用いることを意図する）、信頼性できるパラメーター（例えば、予測で用いられる記述子のための数値）を使用していれば、正しい入力が行われている。入力が正しいことを確認するために必要な文書化のレベルは、(Q)SAR モデル（または実装するソフトウェア）および評価対象物質の複雑さによって異なる。一般的には、入力は注意深く準備されるべきで、可能であれば(Q)SAR モデルあるいはソフトウェアの開発者の指示に従う。

予測チェックリストには、入力が正しいことを確認するための次の AE が含まれる。

1. 入力とモデル設定を明確かつ完全に説明しているか
2. 評価対象物質を代表的に表す入力か
3. 信頼できる入力（パラメーター）か

入力とモデル設定を明確かつ完全に説明しているか（予測/結果チェックリスト AE 1.1）

最初にチェックすべき評価要素は、明確で完全な入力についての説明である。もっとも単純なケースでは、構造情報（SMILES など）を唯一の(Q)SAR モデルへの入力とし、カスタマイズ可能なオプションがない。このケースでは、予測に使用した正確な構造情報と(Q)SAR モデル/ソフトウェアのバージョンが明示されれば十分である。より複雑なケースでは、化学構造に関する三次元情報、カスタマイズ可能なオプション（「設定」）、入力に必要なソフトウェアのパラメーター（例えば手動で入力した記述子の値やそ

の情報ソース)を含む、すべての情報の提示が要求される。

評価対象物質を代表的に表す入力か (予測/結果チェックリスト AE 1.2)

次に、入力が評価対象物質を代表的に表し、評価に適切なものであることを確認することが重要である。物質が明確に定義される単一成分で構成される場合は、物質名、構造および数値識別子の一致を確認すれば十分である。三次元モデルにおいては、入力した立体構造を選択した根拠に関する情報が期待される。複雑な組成の物質については、(Q)SAR 結果は構成成分や不純物をカバーする複数の予測から導かれる。事実、(Q)SAR の利点の 1 つは、より多くの成分や代謝物について予測することで、少ない追加コストで物質の毒性全体への寄与を調査できることである。

加えて、一部の(Q)SAR モデルでは、使用前に入力構造についてキュレーションする必要があるかもしれない。塩、イオン性物質または互変異性構造などがしばしばこれに該当する。これらの場合、様々なアプローチが存在し、どのアプローチにするかはケースバイケースで決定すべきである。その際、(Q)SAR モデル開発者がトレーニングセット物質をどのように前処理したかや、必要があれば、対象とする規制の推奨条件をよく考慮すべきである。

信頼できる入力 (パラメーター) か (予測/結果チェックリスト AE 1.3)

最後に、化学構造以外の物理化学的特性のような記述子を直接入力する(Q)SAR モデルでは、その記述子の値が実験的に測定されたものか、他の(Q)SAR モデルによって予測されたものかに関わらず、(訳注：その記述子を用いて)別の特性を予測する前にその値の信頼性を評価する必要がある。適切に正当化されない限り、モデル開発者が(Q)SAR モデルの開発/性能評価の間に適用したのと同じアプローチを行うべきである。(Q)SAR モデルが多く物理化学的記述子に依存し、すべての入力の信頼性を個別に評価することが不可能な場合は、最も影響力のあるいくつかの記述子に焦点を当てるべきである。

2.2 モデルの適用範囲内にある物質であること

第 2 原則は、(Q)SAR モデルが評価対象物質に適用可能であることを求めるものである。物質に対する(Q)SAR モデルの適用可能性は、適用範囲およびモデル開発者によって定義された(Q)SAR モデルの制約に、評価対象物質がどのように関連するかにより評価される。モデル開発者が適用範囲の定義時に考慮しておらず、予測の信頼性に影響を与える点があった場合は、次項の「信頼できる予測であること」の原則において評価する。適用範囲は、予測の信頼性の情報を与えるものである。本文書においては、評価手順を効率化するために、AD と信頼性を別々に評価する。

予測チェックリストには、物質が(Q)SAR モデルの適用範囲内にあることを確認するための以下の AE が含まれる。

1. 物質が適用範囲内にあるか
2. モデルに関するその他すべての制約が考慮されているか

物質が適用範囲内にあるか / モデルに関するその他すべての制約が考慮されているか (予測/結果チェックリスト AE 2.1 および 2.2)

一般に、AD の定義には、評価対象物質が AD 内にあるかを評価するための基準を明示しなければならない。場合によってモデル開発者は、(Q)SAR モデルの既知の制約や、除外基準（例えば、予測が不正確になると判明しているケミカルクラスや記述子値の範囲など、使用すべきでないケースのリスト）も提供す

る。これらは AD の定義とは別に提供されるかもしれないが、重要性に違いはない。最近のソフトウェアでは物質に対する(Q)SAR モデルの適用可能性を自動的に評価し、予測結果とともにその結果を提供することも多い。このような場合では、自動的な評価の結果を独立して確認できることが重要である。

物質が(Q)SAR モデルの AD 内にあり、特定の制約にも該当しないのが望ましいシナリオである。

逆に、物質が(Q)SAR モデルの AD 外あるいは特定の制約に該当することが明らかな場合がある。そのようなケースでは、予測を必要とする理由（例えば、試験が技術的に不可能であるが、数値が必要とされるなど）がない限り、その予測はそれ以上考慮すべきではない。例として、ある物理化学的特性の範囲が適用範囲として定義される(Q)SAR モデルで、ある特定のケミカルクラスの有機化学物質を予測するために開発されたことが明示されているとする。この場合、特定のケミカルクラス以外の有機化学物質の予測については、物理化学的特性の AD に関わらず、それ以上考慮すべきではない。

以上の 2 つのわかりやすいシナリオの間には中間的なケースが存在する。一例としては、AD との関係を一義的に確立できないものの、(Q)SAR モデルを使用すべきでないとする除外基準には当てはまらない物質がある。この場合、(Q)SAR モデルの正確度が不明または低い可能性はあるが、すべての規制目的で必ずしも受け入れられないわけではない。別の例として、AD の境界付近に位置する物質の予測あり、それ故に高い不確実性を伴う可能性が挙げられる。これらのケースでは、次項で述べる予測の信頼性の評価が、特定の規制目的での予測の全体的な受け入れを決定する。

2.3 信頼できる予測であること

予測の信頼性を評価する際には、AD に加えていくつかの点が考慮される。これらが既に AD の定義に含まれている場合は、評価を繰り返す必要はない。

予測チェックリストには、予測の信頼性を確認するための次の AE が含まれる。

1. 再現性があるか
2. モデルの全体的な性能がどの程度か
3. モデルのトレーニングセットの物理化学的、構造および応答空間への適合性があるか
4. 類似物質に対するモデルの性能がどの程度か
5. メカニズムおよび/または代謝が考慮されているか、ある場合どのような内容か
6. 情報に一貫性があるか

再現性があるか（予測/結果チェックリスト AE3.1）

最初に、記載されている入力と(Q)SAR モデルを用いて予測の再現（すなわち、同じ予測結果を得る）ができるかを確認すべきである。評価者が予測に使用された同一モデルの同一バージョンにアクセスできる場合、同じ予測を行い、同じ予測が得られることを確認できる。バージョン違いなど正当な説明がない場合、予測が異なれば、それは信頼性が低い兆候である。

モデルの全体的な性能がどの程度か（予測/結果チェックリスト AE 3.2）

次に、(Q)SAR モデルの全体的な性能について考慮する。これは信頼性評価のベースラインであり、(Q)SAR モデルの説明文書（QMRF 参照）に記載されていると想定される。例としては、定量的な予測結果を得る(Q)SAR モデルについては同様の標準誤差が（カテゴリーで予測結果を得る(Q)SAR モデルについては同様の正確度が）、評価対象物質の予測においても期待される。このため、性能が優れたモデルによる予測

は、性能の低い(Q)SAR モデルによる予測より信頼性があると考えるべきである。

モデルのトレーニングセットの物理化学的、構造および応答空間への適合性があるか（予測/結果チェックリスト AE3.3）

トレーニングセットの物質では、物理化学的記述子、分子記述子などの値と、対象となる特性値（応答値、予測される特性に相当）が定義されている。これらを使用して、範囲または分布に基づいた記述子空間と応答空間を定義できる。評価者は、評価対象物質の記述子と応答の値を、トレーニングセットで定義された空間と比較することができる。

同様に、トレーニングセットの物質を用いて、官能基や構造フラグメントの観点から構造空間を定義できる。物質のすべての構造的特徴が(Q)SAR モデルにとって既知であれば（トレーニングセットに内包されていれば）、予測の信頼性がより高くなると考えられる。(Q)SAR モデルにとって未知の構造的特徴は、予測の信頼性に影響を与える可能性がある。この点は、フラグメント/アラートベースの(Q)SAR モデルが「生物学的活性がない（訳注：影響なし）」と予測した際に特に重要である。その(Q)SAR モデルの「影響なし」という予測が、あるフラグメントの影響を予測するトレーニングがされていない(Q)SAR モデルの知識の欠落によりもたらされている可能性がある。

評価対象物質が、トレーニングセットの物質が少ない、または予測性能が低い空間に位置する場合、予測の信頼性は低くなる。この点に特に注意を払うべきである。

類似物質に対するモデルの性能がどの程度か（予測/結果チェックリスト AE 3.4）

(Q)SAR モデルにおける局所的な性能の評価には、予測対象となる特性についての信頼できる実験データがある類似物質の特定が必要である。ソフトウェアが自動的に類似物質を特定しない場合には、(Q)SAR 使用者がこれを行い、評価者が確認すべきである。類似物質は、構造的、物理化学的および/またはメカニズム的類似性など、複数の側面を考慮して特定する。類似物質は、OECD QSAR Toolbox⁵のようなアナログ同定のための計算ツール使用の有無に関わらず、専門家判断により、使用者が手動で特定できる可能性がある。一部の(Q)SAR モデルでは、入力物質に類似した物質のリスト、その実験値および予測の正確度を、予測とともに自動的に提供する。

しかし、分子類似性の指標だけでは類似性を定義するのに必ずしも十分ではない。これは、一部の分子フラグメントはある特性には非常に重要だが、別の特性には重要でない場合があり、場合によっては複数のフラグメントの共存が非常に重要になることもあるからである（例：架橋剤）。

一方、評価対象物質と類似する物質について信頼性のある実験データが得られる場合には、これらの物質を予測した時の(Q)SAR モデルの性能を用いて、より適切に予測の信頼性を評価することができる。この評価に使用した類似物質がトレーニングセットに含まれない場合、(Q)SAR モデルを実際に新しい物質の予測に適用するケースをより客観的に模倣することになるため、よりよい裏付けとなる。特に予測に最も関係する記述子、フラグメント、またはその他の特性に関して、評価対象物質と類似しているほど、評価対象物質の予測の正確度の推定に、より有益となる。場合によっては、評価対象物質あるいは類似物質が形式上は(Q)SAR モデルの AD 外であっても、信頼性が向上する。(Q)SAR モデルの開発時には、そのケ

⁵ OECD と欧州化学機関 (ECHA、European Chemical Agency) が共同開発した、無料で利用可能なソフトウェア。詳細は www.qsartoolbox.org

ミカルスペースにおける実験データ不足のために、評価対象物質が AD 外になった可能性もある。もし類似物質の新規のデータが、そのケミカルスペースにおいて(Q)SAR モデルが良好な性能を示せば、評価対象物質の予測の信頼性は付随して向上する。

フラグメント/アラートベースの(Q)SAR モデルの場合、特定の陽性/陰性アラートの性能（例えば Cooper 統計 (OECD, 2007) で表される）は、(Q)SAR モデルの局所的な性能に関する情報も提供することができる。

類似物質のデータがない場合やその特定が（実現）不可能な場合には、この AE は「非該当/評価しない」と記載する。

メカニズムおよび/または代謝が考慮されているか、ある場合どのような内容か（予測/結果チェックリスト AE 3.5）

この AE は、評価する特性に関連するトキシコキネティクスおよびトキシコダイナミクスの考慮を対象とする。AOP などのメカニズムの観点や ADME⁶特性（生物学的または非生物学的変換、つまり代謝や加水分解、自己酸化や光分解など）に関する考慮が、予測にどのように合致するかという考察は信頼性評価に役立つので、可能な場合には行うべきである。評価する特性に関連のある特定のメカニズムを持つことが既知または予測される構造の場合、または特定の代謝物や変換生成物が予想される場合には、それらの予測との関係性や、(Q)SAR モデルにおいてどのように考慮されているかを説明することが望ましい。一例として、タンパク質との結合によって起こる皮膚感作性のようなメカニズムが既知の特性についての予測を考える。評価対象物質は、(Q)SAR モデルでは「非感作性」と予測されたが、この物質がタンパク質に結合することがわかっている、または予測されている。この場合、陰性予測の信頼性を裏付けるために追加の説明が必要である。同様に、評価対象物質が有害な代謝物を生成する代謝を受けることがわかっている場合、最終的に有害性がないと予測するには、(Q)SAR モデルの予測がその代謝と生成物の影響をどのように考慮しているかを考察し、その正当性に関して追加説明する必要がある。

情報に一貫性があるか（予測/結果チェックリスト AE 3.6）

最後に、対象とする特性の予測において、複数の(Q)SAR モデルが利用可能であることや、同一または関係した特性についての追加情報（測定値または計算値）が利用可能であることがよくある。複数の(Q)SAR 予測および/またはそのほかに利用可能な信頼できる情報があれば、それらの一貫性について検討する必要がある。独立した(Q)SAR モデル（トレーニングセット、モデル化手法、および/または使用される記述子/アラートに関して）による予測では、予測がお互いに一貫（すなわち、一致）している場合にはさらに信頼性が高くなる。矛盾した情報があり、正当な説明ができない場合には、予測の信頼性が低下する傾向がある。「複数の予測から導かれる(Q)SAR 結果¹」については、結果チェックリストを使用する（第3章参照）。証拠の重みをさらに検討することについては、本文書の範囲外とする。

信頼性の評価は、評価における最も複雑な部分と考えられる。最新の(Q)SAR ソフトウェアでの包括的なレポートによる補助があっても、詳細な解析と専門家判断が要求される。しかし、予測に特化した徹底した評価を行えば、(Q)SAR モデルの適用範囲に関する一般的な情報より多くの知見が得られる可能性がある。

⁶ Absorption, Distribution, Metabolism, Elimination：吸収、分布、代謝、排出

る。

2.4 規制目的に適した出力であること

この原則は、特定の規制上の問いに答えるための(Q)SAR 予測の有用性の評価に関するものである。たとえば非常に信頼性が高くとも、予測あるいは「複数の予測から導かれる(Q)SAR 結果¹⁾」は、特定の規制目的に適合しなければ、その規制目的には使用できない。目的への適合性評価に考慮すべき要素は、規制の枠組みによって異なる。目的への適合性（および全般的な受け入れ可能性）に関する最終的な決定は、その規制の枠組みに責任を負う規制当局が行う。

予測チェックリストには、評価結果が規制目的に適合することを確認するための以下の AE が含まれる。

1. その他の規制要件への適合性があるか
2. 予測される特性は規制要件となる特性と対応しているか
3. 使用する枠組みにおける決定可能性（があるか）

その他の規制要件への適合性があるか（予測/結果チェックリスト AE 4.1）

規制の枠組みで、(Q)SAR の使用が受け入れ可能かについての特定の基準を直接的に示していることがある。この基準に本文書で扱っていない要素が含まれている場合、（訳注：(Q)SAR の）使用を受け入れるには、これらの追加基準に適合していることが求められる。例えば、複雑な組成を持つ物質や混合物についての規制の場合、評価すべき重要な要素は、その規制で指定された基準に従って物質全体の組成または混合物の成分を考慮することである（例：特定の閾値以上の濃度成分すべてを考慮する、濃度加算/用量加算アプローチを使用する、あるいは合理的なワーストケースを選択する）。場合によっては、各成分の潜在的な拮抗作用や相乗作用を考慮する必要があるかもしれない。別の例としては、特性を予測するために 2 つの異なるタイプの(Q)SAR モデルの使用が要求される規制もある（例えば、ICH M7 ガイドライン（潜在的発がんリスクを低減するための医薬品中 DNA 反応性（変異原性）不純物の評価および管理）⁷⁾はルールベースと統計ベースの 2 種類の(Q)SAR の使用が要求されている）。このような場合、単一の予測のみを使用することは目的に適合しない。

予測される特性は規制要件となる特性と対応しているか（予測/結果チェックリスト AE 4.2）

(Q)SAR モデルで予測する特性が、規制要件となる特性に一致することが重要である。(Q)SAR モデルで予測される特性は、トレーニングセットとなる実験データと、(Q)SAR モデル化の前に行われるデータキュレーションに依存する（例：包含/除外基準、手動の評価と外れ値除外の透明性、陽性と陰性を定義する閾値）。例としては、細菌の復帰突然変異アッセイにおける *in vitro* 変異原性を評価する際、規制で要求されるのであれば、(Q)SAR モデルにおいて、特定の細菌株や代謝活性化の有無について明示的に考慮する必要があるかもしれない。規制が特定の試験ガイドラインを参照している場合、トレーニングセットには試験法ガイドラインに従って得られた実験結果を含めるべきである。しかし、これは常に必須とは限らない。例えば、Ames 変異原性試験の(Q)SAR モデルに、現在要求されているすべての菌株を使用して実施されたものではない過去のデータが含まれている可能性があったとしても（訳注：一つの株で陽性を示せば陽性と判断することを考えれば）、陽性の予測は適切となる場合がある。別の例として分解半減期

⁷⁾ https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-m7r1-assessment-control-dna-reactive-mutagenic-impurities-pharmaceuticals-limit_en.pdf

について、一次分解半減期のデータでトレーニングした(Q)SAR モデルからの予測は、究極分解半減期を評価する目的には適合しない。

使用する枠組み内における決定可能性があるか（予測/結果チェックリスト AE 4.3）

使用する枠組み内で規制上の決定を下すことができる場合に、予測は目的に適合するといえる。規制上の閾値がある場合、予測値が閾値と比較可能（つまり、同じ単位または変換可能な単位で表される）で、閾値の片側にあること、あるいは懸念を排除できない程度に閾値に近いことについて十分な信頼水準が確保されるときに、決定可能となる。例えば、ある物質の対水溶解度について、(Q)SAR モデルによる予測値 0.09 mg/L の不確実性が高く、規制上の決定が 0.10 mg/L を上回るか下回るかによる場合、その予測は決定不可能であろう。必要な信頼水準およびその設定方法は、規制上の使用目的によって異なる。

2.5 個々の予測に関する評価の結論

予測チェックリストの結論には、評価の不確実性スコアとその不確実性に基づく評価結果が含まれる。

2.5.1 不確実性

予測チェックリストを用いる際、各 AE が満たされているかを判断する必要がある。ほとんどの場合、この決定には一定の不確実性⁸が含まれる。

個々の AE の不確実性は、半定量的なスコア（低、中、高）を使用して表現できるが、「非該当/評価しない」とされた AE には不確実性スコアを付与しない。「満たす」となった AE については、不確実性スコアの割り当て方を予測チェックリストの別表で説明している。また、「満たさない」または「記載なし」となった AE には、正当性のある根拠が示されない限り、デフォルト値として高い不確実性スコアが割り当てられる。そして、各 AE の不確実性スコアを使用して、予測全体の不確実性を決定することができる。

以下のとき、予測チェックリストの全体的な不確実性は「低」と見なされる。

- すべての要素を満たし、個々の要素の不確実性が低い。 または、
- ほとんどの要素の不確実性が低く、不確実性のより高い（訳注：「中」または「高」の）要素（満たさない、または記載なしの要素がある場合は、それらを含む）が全体的な評価において重みが大きくないと考えられ、それに対して説得力のある正当な考察が示されている。

以下のとき、予測チェックリストの全体的な不確実性は、「中」と見なされる。

- すべてのまたはほとんどの要素を満たし、それらの不確実性は「中」である。より高い不確実性の要素がもしあっても、全体的な評価で重みが大きくないと考えられる。 または、
- ほとんどの要素を満たし、それらの不確実性は低い、全体的な評価で重みが大きいいくつかの要素の不確実性が「中」である。

以下のとき、予測チェックリストの全体的な不確実性は、「高」と見なされる。

- すべてのまたはほとんどの要素を満たすが、それらの不確実性は「高」である。 または、
- ほとんどの要素を満たし、不確実性は「低」または「中」だが、全体的な評価で重みの大きいいくつかの要素の不確実性は「高」である。

⁸ 不確実性は、予測が、あるポイントの値、範囲または確率分布のいずれで表現されるかに関係なく、評価そのものを対象とする。不確実性の定義については、用語集を参照。

かの要素で、不確実性が「高」である。

2.5.2 予測全体の評価結果

予測全体の評価結果は、予測チェックリストでの個々の AE の評価結果と不確実性を統合したものに基づく。一般に、重みの大きい AE は評価にとって極めて重要であり、受け入れ可能な予測とするには、それらすべての AE を「低」または「中」の不確実性で満たすべきである。重みの小さい AE も重要であるものの、適用する場面によっては、それらの一部の AE については満たすものの不確実性が高い、または満たさないといった結果も受け入れることができる。2.5.1 項で述べたロジックを適用する場合、不確実性が「低」または「中」の予測を受け入れ可能とみなすことに相当する。

本文書は、評価に関する一般的な指針を示し、各 AE において使用できる「重み」について提案している。しかし、各規制当局は、その権限が及ぶ用途に関して、受け入れ可能性について本文書とは異なる独自の要件を確立・通知することができる。

個々の予測についての評価は、この段階で終了する。

3. 「複数の予測から導かれる(Q)SAR 結果¹」の評価（結果チェックリスト）

3.1 結果チェックリストを使用する場面

同じ特性や関係する特性について、「複数の予測から導かれる(Q)SAR 結果¹」を評価するときには、結果チェックリストを使用すべきである。複数の予測を考慮するケースには、以下が含まれる。

- a. 同じ構造についての異なるモデルでの予測
- b. 異なる構造についての同じモデルでの予測（ある物質の複数の成分、評価対象物質とその代謝物や変換生成物、など）
- c. 上記の組み合わせ

まず、結果チェックリストに内包される予測チェックリストにより、個々の予測を評価する必要がある。c のような複雑なケースでは、最初に一つの構造についての複数の結果を処理してから、次に異なる構造についての予測を検討することが推奨される。

（結果チェックリスト内の）各予測チェックリストを完成させる必要がある。各予測の評価は、以下の AE を除き独立して行うことができる。以下の AE については、すべての予測からの情報を考慮する必要がある。

- 評価対象物質を代表的に表す入力か（訳注：AE 1.2）
- メカニズムおよび/または代謝が考慮されているか、ある場合どのような内容か（訳注：AE 3.5）
- 情報の一貫性があるか（訳注：AE 3.6）
- 「規制目的に適した出力であるか」の原則に関連するすべての AE（訳注：AE 4.1/4.2/4.3）

同じ物質の 2 つの成分を個別に予測する場合を例として挙げる。予測ごとに、結果チェックリスト内の予測チェックリストを 1 つずつ作成する。「1.2 評価対象物質を代表的に表す入力」に関する AE を評価する際には、両成分を考慮し、結果チェックリスト内の個々の予測チェックリストに同じ評価結果を記載することができる。

「複数の予測から導かれる(Q)SAR 結果¹」を評価する際に考慮すべき追加の AE が 1 つある。（訳注：この AE は OECD の(Q)SAR 予測原則に明示的には含まれないが、「複数の予測から導かれる(Q)SAR 結果¹」の評価に対応する追加原則「複数の予測から最終結果が導かれ決定されていること」に紐づけられる。）

1. 個々の予測に基づき最終結果が正しく決定されているか

「複数の予測から導かれる(Q)SAR 結果¹」の評価は、QMRF、QPRF および(Q)SAR 結果報告様式(QRRF)に記載された情報に基づくべきである。

評価者は、個々の予測値から導出された最終結果が正しく決定されたことを確認すべきである。エンドポイントの種類と規制要件に応じて、多数決、ワーストケース、平均値またはさらに複雑な手法により、最終結果が決定される。この計算では、例えば予測の信頼性などに応じて、予測に異なる重み付けが行われるかもしれない。いずれの場合でも、評価者は、その決定が文書化され、適切に正当化されていることを確認すべきである。

3.2 (Q)SAR 結果の不確実性と評価結果

3.2.1 不確実性

最後に、結論に到達するために、すなわち、最終結果の受け入れ可能性を決定するために、各予測の重み付けを目的とした追加のステップがある。「複数の予測から導かれる(Q)SAR 結果¹⁾」の不確実性を見積もるためのロジックは、以下のとおりである。

- 異なる独立した(Q)SAR モデルから同一構造に対する一貫した（つまり、一致する）予測が得られ、それを統合して最終結果が生成される場合、最終結果の不確実性は個々の予測の不確実性と同等またはそれ以下になる。個々の予測における不確実性の高い AE が異なるときには、複数の予測からの結果の不確実性はさらに低くできる。すなわち、不確実性の要素を相互に補うことができる。複数の予測で同じ AE の不確実性がより高いときはこの限りではない。
- 異なる独立した(Q)SAR モデルが同一構造に対して矛盾する予測を行い、それを統合するときは、最終結果の不確実性は個々の予測の不確実性と同等かそれ以上になる。通常、矛盾した予測から導かれる結果は、（訳注：先に矛盾した結果を与えた複数の予測とは異なる予測であり、先の片方の予測と一致する）1 つ以上の予測が、矛盾した予測より重みで勝る場合にのみ、許容される。それには、適用範囲がよりよく合致する、信頼性が高い、あるいは不確実性がより低いなどの理由が必要となる。(Q)SAR 使用者には、統合したことについて論理的根拠の正当性ある説明が期待され、一方、(Q)SAR 評価者は、AE「3.6 情報に一貫性があるか」のもとでこの側面に対する見解を述べるができる。
- 最終結果が、異なる複数の構造（例：同じ物質の異なる成分および/または親物質とその代謝物など）の予測に依存するときは、その不確実性は、個々の予測の不確実性と評価対象物質の組成を考慮すべきである。例えば対水溶解度を予測する場合、主成分の予測の不確実性は、微量不純物の予測の不確実性に比べ、最終結果の不確実性に大きく影響しうる。この例では、主成分の予測の不確実性が「中」で、微量不純物の予測では「高」の場合、最終結果の不確実性は「中」となる。一方、主成分の予測の不確実性が「高」で、微量不純物の予測では「中」の場合には、最終結果の不確実性は「高」になりうる。一方、特性によっては、最終結果の不確実性を決定する個々の予測の重みづけにおいて、成分や不純物の濃度がほとんど、または全く寄与しない場合もある。これには、変異原性不純物が存在するだけでその物質全体の変異原性が決定づけられるといった例がある。

3.2.2 結果全体の評価結果

予測結果の不確実性は、評価結果の決定に使用される。2.5.2 項で述べた個々の予測のアプローチと同様に、不確実性が「低」か「中」の結果であれば受け入れ可能と判断する。最終結果の評価結果を決定して、評価は終了する。

4.最終的な考察

本文書は、(Q)SAR モデル、(Q)SAR 予測および「複数の予測から導かれる(Q)SAR 結果¹」を評価するために必要な要素を包括的に列挙することを目的としている。特定のケースや使用する状況によっては、適用されない/要求されない AE がある場合もありうる。各規制当局は、特定の規制の枠組みと状況において、評価のために体系的に検討する必要がある要素を決定し、伝達すべきである（例えば、データベースのスクリーニングを目的として(Q)SAR を使用する場合、個々の物質の評価に使用する場合と比較して信頼性に関する要求が低くなる可能性がある）。規制当局によっては、AE を満たすかどうかを判断する閾値を導入する可能性もあるが、以下の少なくとも 2 つの理由から、本文書には閾値を記載しない。(1) 科学界において広く合意された閾値は存在しない。(2) 対象とする特性、規制目的、使用状況などの多数の要因に依存するため。

通常、科学的妥当性が検証された(Q)SAR モデルを使用することが期待されている。これは、モデルチェックリストを使用して確認することができる。

使用目的に関わらず、常に正しい入力が要求される。入力が正しくなければ、以降の規制目的での検討において、(Q)SAR は有用ではない。

物質と(Q)SAR モデルの AD との関係も重要である。理想的には、物質はその(Q)SAR モデルの AD 内に明確に収まるべきである。そうでない場合でも、その他の信頼性の観点から、特定の目的のためであれば予測あるいは結果の使用が受け入れられる可能性がある。

最終的に、あくまで予測あるいは結果は目的に適している場合にのみ使用できるのであり、他のどんな点で正確で信頼できるとしても、それ以外では使用できない。(Q)SAR 結果の使用についての規制要件が明確であれば、この原則の評価が容易になる。

複数の予測を統合して 1 つの結果を導出、決定するときは、最終結果に到達するために使用したアプローチについての説明とその正当化が不可欠である。

本文書には、QPRF の大幅な更新が含まれる。QMRF についてもより包括的な更新の必要性が指摘されており、今後の取り組みとして検討される予定である。さらに、新たに作成した「複数の予測から導かれる(Q)SAR 結果¹」を報告するテンプレートが、この文書の付録 III として添付されている。

QAF チェックリストの別表には、実際の評価に役立つ詳細な説明と実施例が記載されている。本文書の発行後、予測や規制固有の要件を明確にするために、規制に特化したガイダンス文書またはケーススタディが作成されることが予想される。

付録 I (Q)SAR モデル報告様式 (QMRF) v.2.1

編集可能な様式は、QAF³の“Supporting Material”タブからダウンロード：

https://www.oecd.org/en/publications/q-sar-assessment-framework-guidance-for-the-regulatory-assessment-of-quantitative-structure-activity-relationship-models-and-predictions-second-edition_bbdac345-en.html

QMRF v.2.1 は、QMRF フィールドの説明のみに関係する、バージョン 2.05 QMRF テンプレートのマイナー更新版である。唯一の例外は、セクション10を完全に削除したことである。このセクションは、JRC QSARモデルデータベースを参照しているが、当該データベースは現在更新されていない。

付録 II (Q)SAR 予測報告様式 (QPRF) v.2.0

編集可能な様式は、QAF³の“Supporting Material”タブからダウンロード：

https://www.oecd.org/en/publications/q-sar-assessment-framework-guidance-for-the-regulatory-assessment-of-quantitative-structure-activity-relationship-models-and-predictions-second-edition_bbdac345-en.html

付録 III (Q)SAR 結果報告様式 (QRRF) v.1.0

編集可能な様式は、QAF³の“Supporting Material”タブからダウンロード：

https://www.oecd.org/en/publications/q-sar-assessment-framework-guidance-for-the-regulatory-assessment-of-quantitative-structure-activity-relationship-models-and-predictions-second-edition_bbdac345-en.html

「複数の予測から導かれる(Q)SAR 結果 1」の文書化にあたり(Q)SAR 使用者が情報を整理できるように、(Q)SAR 結果報告様式 (QRRF) が作成され QAF に追加された。QRRF は、（事前定義されたアルゴリズムではなく）(Q)SAR 使用者が複数の予測を組み合わせて「複数の予測から導かれる(Q)SAR 結果 1」を決定する場合に、記入しなくてはならない。QRRF は、個々の QPRF および QMRF に追加して提出する。QPRF の特定の項目（セクション 2「物質」および 8「使用目的」の項 8.1、8.2、8.3、8.5）は、QRRF と重複している。これらの欄は QRRF に記入済みの場合、各 QPRF に記入する必要はない。報告様式の作成を容易にするために、QRRF の Excel ファイルには、QPRF のコピーと、複数の予測を組み合わせて「複数の予測から導かれる(Q)SAR 結果 1」を決定する場合にどのように入力するかについての情報も含まれる。(Q)SAR 使用者が入力しやすいように、実用的な形式で記入欄を配置している。

用語集

Assessment element (評価要素、AE) : (Q)SAR モデル、予測、全体の結果の評価において考慮すべき極めて重要な要素。AE は、OECD の (Q)SAR モデル原則と (Q)SAR 予測原則に関連付けられる。

Assessor (評価者) : (Q)SAR モデルおよび/または結果について、意図した規制目的での受け入れ可能性を評価する者。

Conclusion (結論) : 個々の「(Q)SAR 予測」または「複数の予測から導かれる (Q)SAR 結果¹⁾」の評価についての不確実性および評価結果。

Model checklist (モデルチェックリスト) : QAF 原則に則った (Q)SAR モデル評価を補助する別文書。考慮すべき評価要素のリスト、評価結果の記録欄、実践的アドバイス、実施例が含まれる。

Outcome of the assessment (評価結果) : 「(Q)SAR 予測」や「複数の予測から導かれる (Q)SAR 結果¹⁾」が意図した規制目的において受け入れ可能かについての決定。

Prediction checklist (予測チェックリスト) : QAF 原則に則った「(Q)SAR 予測」評価を補助する別文書。考慮すべき評価要素のリスト、評価結果の記録欄、実践的アドバイス、実施例が含まれる。

Property (特性) : 物理化学、毒性、生態毒性または環境中運命、化学反応性または生物学的相互作用を含む。「エンドポイント」という用語は人によってその理解が異なるため、本文書においては「エンドポイント」よりも「特性」という用語を選択している。

(Q)SAR model ((Q)SAR モデル) : 構造を入力情報として、物質の特性を予測するモデル。

(Q)SAR prediction ((Q)SAR 予測) : (Q)SAR モデルの個々の出力結果（すなわち、特性についての予測値）。（訳注：数値のような）連続したもの、あるいはカテゴリー分類（訳注：陽性/陰性など複数カテゴリーのいずれに属するか、など）の形で出力される。

(Q)SAR result ((Q)SAR 結果) : 複数の (Q)SAR 予測に基づき（訳注：統合して導出した）物質の特性についての評価結果。（訳注：本文書中では「複数の予測から導かれる (Q)SAR 結果¹⁾」と記載している。）

Regulatory framework (規制の枠組み) : 規制要件を満たすために使用される特定のガイドライン、規則あるいは仕組み。

Regulatory purpose/use/purpose (規制目的 / 用途 / 目的) : (Q)SAR 結果が使用される、規制の枠組み内での特定の用途。

Result checklist (結果チェックリスト) : QAF 原則に則った、「複数の予測から導かれる (Q)SAR 結果¹⁾」評価を補助する別文書。考慮すべき評価要素のリスト、評価結果の記録欄、実践的アドバイス、実施例が含まれる。

Substance (物質) : 評価対象の化学物質。1 つあるいは複数の成分から成り、したがって 1 つあるいは複数構造に関連付けられる。

Uncertainty (不確実性) : 科学的評価における不確実性に関する EFSA ガイダンス (2018)⁹によると「評価のための質問に対して想定されうる回答の範囲と確率に影響を与える、利用可能な知識におけるあらゆる種類の制約を指す、一般的な用語」とされる。結果チェックリストでは、個々の評価要素、「(Q)SAR 予測」または「複数の予測から導かれる(Q)SAR 結果¹」に対して、低/中/高の不確実性の値が割り当てられる。

Internal and external validation of a (Q)SAR model ((Q)SAR モデルの内部検証/外部検証) : トレーニングデータセットや(独立した)テストセットを使用した、(Q)SAR モデルの性能を評価する統計的手法。

Weight of the assessment element (評価要素の重み) : 「(Q)SAR 予測」の全体的な評価における各評価要素の重要度。規制目的に依存し、低/中/高の値を取りうる。

⁹ EFSA (European Food Safety Authority) Scientific Committee, Benford, D, Halldorsson, T, Jeger, MJ, Knutsen, HK, More, S, Naegeli, H, Noteborn, H, Ockleford, C, Ricci, A, Rychen, G, Schlatter, JR, Silano, V, Solecki, R, Turck, D, Younes, M, Craig, P, Hart, A, Von Goetz, N, Koutsoumanis, K, Mortensen, A, Ossendorp, B, Martino, L, Merten, C, Mosbach-Schulz, O and Hardy, A, 2018. Guidance on Uncertainty Analysis in Scientific Assessments. EFSA Journal 2018;16(1):5123, 39 pp. <https://doi.org/10.2903/j.efsa.2018.5123>