令和７年度　厚生労働行政推進調査事業費補助金　新興・再興感染症及び予防接種政策推進研究事業

公的医療及び社会の立場からのワクチンの費用対効果の評価法及び分析方法の確立のための研究

分担研究報告書

ワクチンの費用対効果評価における生成 AI の活用可能性と信頼性確保に向けた国際動向の整理

森脇 健介

立命館大学 生命科学部 生命医科学科 医療政策・管理学研究室

研究要旨

近年、生成 AI・大規模言語モデル（LLM）は、ワクチンを含む医療技術の有効性・安全性評価に係るシステマティックレビューや費用対効果モデルの構築、リアルワールドデータ解析など、ワクチンの HEOR 業務の効率化に寄与している。一方、透明性・再現性の不足、ハルシネーション、バイアスといった課題も指摘されており、ワクチン政策の科学的根拠として用いる際には特に慎重な取り扱いが求められる。本研究では、ワクチンの費用対効果評価を含む HTA 領域における AI 活用の信頼性確保に向け、国際的ガイドラインの動向を整理した。ISPOR は ELEVATE-GenAI フレームワークを提唱し、モデル特性や再現性、バイアス評価など 10 領域での報告を求めている。NICE は AI 活用をエビデンス生成、技術評価、業務効率化の 3 領域で推進し、方法論整備を進めている。CDA-AMC は AI を補助的手段と位置付け、透明性と人間の関与を重視している。これらの知見を踏まえ、日本のワクチン費用対効果評価制度においては、AI 利用の明示的な申告、リスクに応じた段階的運用、ワクチン分野における人材育成、国際整合的な方法論整備が必要である。AI は有用だが、ワクチン政策決定に直結する分析においては特に透明性と検証可能性を担保した慎重な導入が求められる。

１．はじめに

近年、生成 AI および大規模言語モデル（LLM）は、ワクチンを含む医療技術に関する医療経済・アウトカム研究（HEOR）において急速に活用され始めている。具体的には、ワクチンの有効性・安全性に関するシステマティックレビュー（SLR）、費用対効果モデルの構築、リアルワールドデータ（RWD）解析など、従来は多大な人的労力を要していた作業の自動化・効率化が可能となっている。

ワクチンの費用対効果評価は、定期接種への追加やワクチン接種プログラムの設計に直結する政策判断の根拠となるため、分析の透明性・再現性・正確性の確保が特に重要である。しかし、LLM の利用には以下の課題が存在する。すなわち、透明性の欠如、再現性の不足、ハルシネーション（誤情報生成）、バイアスや公平性の問題などである。HEOR・HTA 領域では LLM 利用の報告基準が未整備であり、研究の信頼性確保が課題となっている。

本年度は、ワクチン費用対効果評価を含む HTA 領域における生成 AI・LLM の「活用可能性」の段階を踏まえ、信頼性・透明性・再現性の確保を中心とした国際的ガイドラインの動向と、HEOR 実務への実装的展開を整理する。

２．国際的な指針の整理

2-1. ISPOR の状況

ISPOR のワーキンググループは、HEOR 研究における LLM 利用を透明かつ再現可能に報告するためのガイドライン ELEVATE-GenAI フレームワークを開発した[1]。このフレームワークは、LLM を用いた研究

の報告標準、研究の再現性確保、HTA や政策判断での信頼性向上を目的としており、ワクチンの有効性・安全性評価や費用対効果分析においても適用が期待される。

ELEVATE-GenAI フレームワークは、Bedi らによる AI 性能評価ドメイン、Holistic Evaluation of Language Models ベンチマークを基礎として構築されているが、ISPOR の Generative AI ワーキンググループは、HEOR 分野特有の方法論的および規制上の要件に対応するため、以下の 3 つの追加ドメインを導入した。

- モデル特性（Model characteristics）
- 再現性および一般化可能性（Reproducibility and generalizability）
- セキュリティおよびプライバシー（Security and privacy）

その結果、本ガイドラインは 10 の評価ドメインから構成される（表 1）。なお、各ドメインを以下の基準で評価し、すべてのドメインの点数を合計して総合スコアを算出することにより、各研究のガイドライン遵守度を半定量的に要約することもできる。

- 3 点：十分に報告されている（Clearly reported）
- 2 点：不明確または部分的（Ambiguous）
- 1 点：未報告（Not reported）

表 1. ELEVATE-GenAI チェックリスト

評価領域	チェック項目
モデル特性（Model characteristics）	モデル名、バージョン、開発者、公開日、ライセンス（オープンソースまたは商用）、モデルアーキテクチャが記述されているか。
	学習データの情報（例：PubMed などのドメイン特化データ）およびファインチューニングの詳細が記載されているか。
精度評価（Accuracy assessment）	タスクに応じた精度指標（例：Precision、Recall、F1 スコアなど）が報告されているか。
	出力結果が人間の評価やゴールドスタンダードデータセットと比較して検証されているか。
網羅性評価（Comprehensiveness assessment）	出力結果が既存の研究レビューや検証済みモデルなどのベンチマークと比較され、網羅性が確認されているか。
	重要な要素がすべて含まれていることについて専門家による評価が行われているか。
事実性検証（Factuality verification）	出力の事実性を確認する方法（例：文献との照合、専門家レビュー）が記述されているか。
	不一致があった場合の修正方法や対応が記録されているか。
再現性および一般化可能性（Reproducibility and generalizability）	再現性確保のための情報（例：プロンプト内容、ハイパーパラメータ）が公開されているか。

評価領域	チェック項目
	独立した研究者が検証できるワークフローが提示されているか。ワクチン有効性推定などのタスクへの一般化可能性が検討されているか。
ロバストネス検証（Robustness checks）	入力の誤字や曖昧な質問などへの耐性を評価するテストが実施されているか。その条件下でのモデル性能の変化が報告されているか。
公平性・バイアス評価（Fairness and bias monitoring）	出力が性別、年齢、人種などに関するバイアスや偏見を含んでいないか評価されているか。ワクチン接種対象集団の多様性に関するバイアスへの対応が示されているか。
実装環境と効率性（Deployment context and efficiency）	実装環境（ハードウェア、ソフトウェア、実行可能コードなど）が明確に記述されているか。処理時間、スケーラビリティ、計算資源などの効率性指標が報告されているか。
キャリブレーションと不確実性（Calibration and uncertainty）	モデルのキャリブレーション方法が記述されているか。ワクチン費用対効果分析における不確実性の閾値が定義されているか。
セキュリティとプライバシー（Security and privacy）	セキュリティ対策（暗号化、匿名化、アクセス制御など）が説明されているか。GDPR や HIPAA など関連法規への適合が示されているか（該当する場合）。知的財産権や著作権への配慮が記載されているか。

ELEVATE-GenAI の各ドメインの内容や報告事項は表 2 に要約される。なお、各ドメインには、現時点における確立された指標または報告基準の現状を反映した成熟度レベルが割り当てられているが、これらは今後の状況変化に応じて再検討・更新されるものである。ELEVATE-GenAI を公表されている SLR や費用対効果分析に適用し体系的に評価する試みも行われている[1]。

表 2. ELEVATE-GenAI のドメインの詳細

ドメイン	内容（概要）	報告すべき事項	指標成熟度
モデル特性	モデルの基本属性（名称、バージョン、開発者、アクセス方法、ライセンス、アーキテクチャ、学習データなど）を記述する。	モデル名、バージョン、開発者、公開日、ライセンス、アクセス方法を報告。モデル構造を説明。学習データやファインチューニングの内容を記載。	高
精度評価	モデル出力が正解や期待される結果とどれだけ一致しているかを評価する。	人間の評価やゴールドスタンダードデータとの比較。適切な指標（Precision、Recall、F1、AUC など）の報告。	中

ドメイン	内容（概要）	報告すべき事項	指標成熟度
網羅性評価	出力がタスクの全要素を網羅しているかを評価する。	既存レビューやモデルなどのベンチマークとの比較。専門家による評価。	高
事実性検証	出力内容が事実に基づいているかを評価する。	文献やデータソースとの照合。誤情報や不一致の修正方法の記録。	高
再現性・一般化可能性	研究結果が再現可能か、他の研究課題に適用可能かを評価する。	コード、プロンプト、ハイパーパラメータの公開。再現可能なワークフローの提供。他研究への適用可能性の説明。	高
ロバストネス	入力の変化（誤字、曖昧な質問など）に対するモデルの耐性を評価する。	誤字や曖昧入力などへのテスト。性能変化の報告。	高
公平性・バイアス	性別、年齢、人種などによる偏りがないか評価する。特にワクチン接種対象集団の偏りに注意。	バイアス検出。公平性指標（demographic parity など）の使用。	低
実装環境・効率性	モデルの実装環境や計算効率を評価する。	ハードウェア（GPU 等）とソフトウェア構成。処理時間、計算資源、スケーラビリティ。	高
キャリブレーション・不確実性	モデルが不確実性を適切に表現できているか評価する。費用対効果分析の感度解析との整合も重要。	キャリブレーション指標。手動レビューの閾値設定。	低
セキュリティ・プライバシー	個人情報やデータ保護、知的財産を適切に管理しているか評価する。	暗号化、匿名化、アクセス制御。GDPR/HIPAA 等への準拠。知的財産保護。	低

ELEVATE-GenAI ガイドラインは、HEOR における LLM の利用状況を報告するための基本的な枠組みを提供するが、下記のような課題に留意した上で、継続的に見直しを行う必要がある。

- 枠組みの根拠となった対象を絞った文献レビューは体系的ではなく、ワクチンの有効性・安全性評価に特化した情報源が欠落している可能性がある。

- スコアリングシステムが試験的に導入されたが、ワクチン政策分野での有用性についてはさらなる検証が必要である。
- 特定のドメイン定義は概念的に類似しているため、一貫して適用することが難しい場合がある。
- ワクチンの費用対効果評価など HEOR 特有のタスクでは、AI 研究で一般的に用いられる評価指標が十分に検証されていない。
- 反復的または半自律的なタスクを実行するエージェントアプローチが普及するにつれて、ワクチン評価への適用に関する追加ガイダンスが必要となる。

2-2. NICE の状況

NICE は、2024 年に「Statement of Intent for Artificial Intelligence (AI)」を公表し、医療分野および HTA における AI 活用の方針を示した[2]。同文書では、AI が医療におけるエビデンス生成、医療技術の評価、さらには HTA 機関自身の業務効率化において重要な役割を果たす可能性があることが指摘されており、ワクチン評価においても同様の活用が期待される。NICE は AI に関する取り組みを以下の 3 つの優先領域に整理している。

第一は、AI を用いたエビデンス生成方法の整備である。AI は有効性・安全性に関するシステマティックレビュー、エビデンス合成、費用対効果分析などの研究プロセスにおいて、文献検索やデータ分析の効率化を可能にする。しかし、AI を用いた研究では、アルゴリズムの透明性、バイアスの管理、再現性、倫理的配慮などの課題が存在する。そのため NICE は、AI を用いたエビデンス生成に関するベストプラクティスのガイダンスを整備し、研究の信頼性を確保することを目指している。

第二は、AI を組み込んだ医療技術そのものの評価である。AI を用いた医療技術の評価では、従来の評価枠組みに加え、アルゴリズムの妥当性、実臨床環境における性能、アルゴリズムバイアス、説明可能性、サイバーセキュリティといった新たな要素を考慮する必要があるとされている。NICE はこれらの課題に対応するため、AI 技術の評価基準や評価ツールの開発を進める方針である。

第三は、HTA 機関自身の業務における AI 活用である。NICE は、AI を用いることで関連の文献レビューやデータ解析、ガイダンス作成などのプロセスを効率化できる可能性があると指摘している。ただし、AI の利用にあたっては透明性や説明責任、サイバーセキュリティ、リスク管理などを確保する必要があり、また職員や委員会メンバーが AI 技術を適切に評価・活用できるよう AI リテラシーの向上も重要とされている。

さらに NICE は、AI 評価の方法論の整備には国際協力が不可欠であるとして、HTA 機関や規制当局、研究機関との連携を強化する方針を示している。評価基準の国際的な調和を進めることで、AI 技術の評価における重複を減らし、効率的な方法論開発を進めることが期待されている。

2-3. CDA-AMC の状況

Canada's Drug Agency (CDA-AMC) は、2025 年に「Position Statement on the Use of Artificial Intelligence in the Generation and Reporting of Evidence」を公表し、HTA における AI の利用に関する基本方針を示した[3]。同声明は、医療技術評価に提出されるエビデンスの生成および報告の過程における AI 手法の利用について、透明性、厳密性、信頼性を確保する必要があることを強調しており、ワクチンの費用対効果評価においても同様の姿勢が求められる。

CDA-AMC は、AI の利用が適切かつ価値を付加する場合に限り、エビデンス生成を補助する手段として活用されるべきであるとの立場を示している。同声明では、HTA における AI 利用の可能性として、主に以下の領域が示されている。

第一に、システマティックレビューおよびエビデンス統合の分野である。AI は有効性・安全性に関する文献検索戦略の作成、研究デザインによる分類、文献スクリーニング、データ抽出などのプロセスを自動化する可能性がある。大規模言語モデルを用いてメタ解析に必要なコード生成やデータ合成

を支援することも検討されている。

第二に、臨床エビデンスの生成である。AI は臨床試験の設計や対象集団選定、用量設定、試験期間の最適化などに活用される可能性がある。また、電子カルテなどのデータを自然言語処理によって解析し、副反応の検出や試験対象患者の特定に利用することも想定されている。

第三に、リアルワールドデータ（RWD）の解析である。AI は電子カルテ、レジストリ、行政データベースなどの大規模データを統合・標準化する過程において、データクレンジング、重複除去、欠損値補完などを自動化する役割を果たす可能性がある。アウトカムに関する機械学習を用いた因果推論手法の活用により、実臨床における有効性の推定精度を向上させることが期待されている。

第四に、費用対効果評価における経済モデルの開発である。AI はコスト構造や疾病負担の複雑なデータ解析を通じて、モデル構造の設計やパラメータ推定を支援する可能性がある。大規模言語モデルを用いて経済モデルのコード生成やモデル更新を自動化する可能性も指摘されており、費用対効果分析における構造的不確実性の検討に活用できる。

一方で CDA-AMC は、AI の利用に伴うリスクにも注意を払う必要があるとしている。具体的には、アルゴリズムバイアス、透明性の低下、サイバーセキュリティリスク、人間の監督の低下などが挙げられる。とりわけ、ワクチン政策は公衆衛生上の重大な判断を伴うため、AI は人間の判断を代替するものではなく、あくまで意思決定を補助する手段として利用されるべきであり、いわゆる「human-in-the-loop」の原則を維持する必要がある。

CDA-AMC は今後、AI を用いたエビデンス生成の実例を継続的にモニタリングし、HTA の方法論ガイドラインの更新に反映していく方針を示している。費用対効果評価への適用においても、同様の継続的モニタリングが望まれる。

３．日本のワクチン費用対効果評価制度への示唆

本研究では、HTA における AI 活用に関する国際的な指針の策定状況を整理した。AI の利用が今後拡大することを踏まえ、ワクチンを含む HTA における AI 活用の方法論や評価基準の整備が国際的に進展していくことが期待される。本検討の結果を踏まえ、日本のワクチン費用対効果評価において今後検討すべき事項として、以下の点を提言する。

（１）HTA 提出資料における AI 利用の透明性確保

ワクチンの費用対効果評価等の HTA 提出資料において、AI の利用有無および利用範囲を明示的に申告する仕組みを整備する。特に以下の工程における AI 利用について記載を求めることが望ましい。

- ワクチン有効性・安全性に関する文献検索・文献スクリーニング
- データ抽出および整理（ワクチン臨床試験データ、副反応データ等）
- RWD/RWE 解析（接種後の実臨床有効性・安全性評価）
- 統計解析および因果推論
- ワクチン費用対効果モデル構築・コード生成
- 報告書作成

また、AI ツール名、バージョン、使用方法、人間による確認手順、検証方法、既知の限界などを記載する様式の整備を検討する。

（２）AI 利用に関するリスクに応じた段階的運用

ワクチン HTA における AI 利用については、用途ごとのリスクに応じた段階的運用を行う。例えば以下のような区分が考えられる。

・ 低リスク用途（導入しやすい）
  - 文献検索補助・文献分類
  - 図表整形・要約作成
  - 疾病負担データの整理

・ 高リスク用途（厳格な検証が必要）
  - ワクチン有効性・安全性の推定
  - 接種後副反応の因果推論
  - RWD 分析（実臨床におけるワクチンの有効性）
  - ワクチンの費用対効果モデル構造の設計
  - ICER の算出

なお、高リスク用途では、独立再現、感度分析、代替手法との比較、人間による最終確認を必須とすることを検討する。

（3）AI 評価能力を有する人材育成

製造販売業者だけでなく、公的分析実施機関、費用対効果評価専門組織、事務局においても AI を適切に評価できる人材の育成が必要である。特にワクチン分野では以下の能力強化が求められる。

・ AI アルゴリズムの基礎理解
・ ワクチン有効性・費用対効果モデルに対する再現性確認
・ データガバナンス（ワクチン接種記録・副反応データの取り扱いを含む）
・ AI 倫理・公平性（ワクチン接種対象集団の偏りへの配慮）
・ サイバーセキュリティ

（4）国際動向を踏まえた方法論整備

NICE、CDA-AMC 等の HTA 機関では AI 利用に関する方法論整備が進められている。日本のワクチンの費用対効果評価においても、国際的な方法論やガイドラインとの整合性を確保しながら制度整備を進めることが重要である。特に、ワクチン政策に直結する HTA 制度においては、国際調和を通じた評価の効率化と信頼性向上が求められる。

４．さいごに

AI はワクチンの有効性・安全性・費用対効果評価を含む HEOR・HTA におけるエビデンス生成および評価プロセスの効率化・高度化に寄与する可能性を有する。特に、大規模なワクチン接種記録や疾病負担データの解析、接種後副反応の検出、費用対効果モデルの構築支援など、ワクチン評価の各フェーズにおいて AI の活用が期待される。

一方で、透明性、再現性、バイアス、セキュリティなど新たな課題も生じるため、日本のワクチン費用対効果評価制度においては AI を全面的に導入するのではなく、透明性と検証可能性を確保した上で段階的に活用を進めることが望ましい。ワクチン政策は国民の健康に直結する意思決定に関わるため、AI はあくまで人間の判断を補助する手段として位置付け、最終的な評価の責任は人間が担うべきである。

参考文献

1. Fleurence RL, et al. ELEVATE-GenAI: Reporting Guidelines for the Use of Large Language Models in Health Economics and Outcomes Research: An ISPOR Working Group Report. Value Health, 2025. 28(11): p. 1611-1625.
2. NICE statement of intent for artificial intelligence (AI) 2024. Available from: www.nice.org.uk/corporate/ecd12
3. Canada's Drug Agency Position Statement on the Use of Artificial Intelligence in the Generation and Reporting of Evidence. 2025. Available from: https://www.cadth.ca/