

令和6年度厚生労働科学研究費補助金
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(令和6年度)

ワクチンの医療技術評価における生成 AI・大規模言語モデル(LLM)の応用

研究分担者 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

立命館大学 総合科学技術研究機構 医療経済評価・意思決定支援ユニット(CHEERS)

近年、生成 AI (Generative AI) および大規模言語モデル (Large Language Models: LLM) は、様々な分野における情報処理と意思決定支援において革新的な役割を果たしつつある。特に医療技術評価 (Health Technology Assessment: HTA) の分野では、システムティックレビュー、リアルワールドエビデンス (RWE) 分析、費用効果分析モデルの構築支援など、多岐にわたる応用が模索されている。本研究では、Chat-GPT を用いたワクチン評価に関する費用効果分析モデルの構築事例を通じて、生成 AI・LLM の HTA 領域での活用の可能性と限界を実証的に検討した。具体的には、日本の高齢者を対象とした RS ウイルス (RSV) 感染症ワクチンの導入効果を評価するためのマルコフモデルを Chat-GPT によって構築し、その過程・出力・精度・再現性・実装上の課題を多面的に分析した。その結果、一定の精度と再現性を持つモデルが構築可能である一方で、専門的知識による検証の必要性、透明性・説明責任・倫理的配慮といった課題も明らかとなった。

A. 研究目的

人工知能 (AI: artificial intelligence) の明確な定義はないが、一般に、「大量の知識データに対して、高度な推論を的確に行うことを目指したもの」(人工知能学会より) と理解される (図 1) [1]。生成 AI (Generative AI) とは、入力データ (プロンプト) に基づいてテキスト、画像、またはその他のコンテンツを生成することができる AI

システムをさす。基礎モデル (Foundation Models) とは、様々な目的を果たす機械学習モデルである。こうしたモデルは大規模なデータで訓練され、微調整の有無にかかわらず、幅広いタスクに適応可能である。大規模言語モデル (LLM: Large Language Models) は、この両者の条件を満たすものであり、膨大なテキストデータで学習された特定のタイプの基礎モデ

ルで、膨大なデータセットから得られた知識に基づいて、テキストやその他のコンテンツを認識、要約、翻訳、予測、生成するものである。なお、GPT(Generative Pre-trained Transformer)は、OpenAIが開発したLLMであり、特に人間のようなテキストを生成するのに適している。近年、医学研究や医療実践において、こうした生成AIやLLMの利活用が急速に進んでいる。こうした動きは、医療技術評価(HTA:Health Technology Assessment)の領域でも同様であり、本研究ではその一端を報告する。

B. 研究方法

本研究では、以下のアプローチで実証的な検討を行った。

文献レビュー:生成 AI・LLM と HTA に関連する国内外の主要文献を精読し、活用状況・技術的特徴・制度的対応等を整理した(例: Florence et al., 2025; Reason et al., 2024)。

Chat-GPT を用いたマルコフモデル構築:日本における高齢者向け RSV ワクチン(RSVPreF3 OA)を対象とし、接種群と非接種群の比較による費用効果分析を想定したマルコフモデルを構築。Chat-GPT への一連のプロンプト入力を通じ、モデル構造・パラメータ設定・Rコード生成を行った。

出力の解析と検証:Chat-GPT が生成したモデルをRで実行し、ベースケースのICER算出結果を報告。モデル構造やコードに含まれる潜在的エラー、再現性の有無、感度分析の可

否などを確認した。

課題の整理:Chat-GPT 利用における妥当性・信頼性・バイアス・倫理的配慮に関する論点を抽出し、HTA 領域における実装上の留意点として整理した。

(倫理面への配慮)

公表文献のレビューであり、倫理的な問題は無い。

C. 研究結果

1. Chat-GPT について

2022 年に Chat-GPT が開発・公開されて以降、誰もが生成 AI・LLM の恩恵に預かることとなっている。Chat-GPT の仕組みは、実はシンプルであり、基本的には一定のルールの下、1 つずつ単語を足しているだけである[2]。ルールとは、現在の出力内容の「順当な続き」を出力しようと試みることである。ここでいう「順当」とは、「億単位の WEB ページに書かれている内容を見たらうで、人間が書きそうだと予測される」という意味である。例えば、Chat-GPT が「The best thing about AI is its ability to…」と出力しているときに、次に来る単語は、学習済みのデータから同じような文脈で出現確率の高いものを出力させるのである。ただし、このルールだとありきたりなものばかり生成されることになる。このため、Chat-GPT のモデル内では、Temperature(温度)というパラメータの設定により、ときどきランクの低い単語をランダムに選ぶのである。これより創造性のある応答の

生成が可能とされる。

なお、質問の言語によって「使用言語モデル」が変わるようである。どの言語を使用すべきかであるが、「使用言語モデル」によって、質問に対する情報量は異なるため、質問に対して、情報をより多く持っていそうな言語モデルを使うと「より精度の高い回答」が期待できる。例えば、日本に独特のローカルな内容については、英語のモデルより日本語のモデルの方が、その情報量が多いため、日本語でプロンプトを作成した方がよいということになる。

2. HTA における活用状況

生成 AI・LLM は、HTA のあらゆる文脈での活用が期待されている。その中でも、現在、具体的な活用が活発になっているのが、以下の 3 領域である[1]。

① システマティックレビューとエビデンス統合への応用

追加的有用性の評価では、当該医療技術の臨床研究のエビデンスについてのシステマティックレビューを実施し、必要に応じてメタアナリシス等の統合解析が行われる。例えば、リサーチクエスチョンに対応する最適な検索式の構築やタイトル・アブストラクトのスクリーニングの自動化、アウトカム情報の抽出、ネットワークメタアナリシス等の統計モデルのプログラム構築といった業務において、生成 AI・LLM の利用が進められている。いくつかの研究事例では、人間と同等の精度でデータ抽出やスクリーニング可能であることを示唆している。ただし、

誤りや捏造(幻覚)が含まれる可能性があるため、人間による監視と検証が必要であり、また、異なる LLM 間での再現性が課題となっている。

② リアルワールドエビデンス(RWE)への応用

生成 AI・LLM の利点として、データ処理と分析の効率性の向上、ヒューマンエラーを最小限に抑えエビデンス生成プロセスを標準化することによる精度と一貫性の向上が挙げられる。例えば、RWE で活用する際のメリットとして、電子カルテ等に含まれるメモといった構造化されていないデータを抽出、変数化し、臨床研究に活用できることが期待されている。一方で、記述データから国際疾病分類や診療行為コードにマッピングする際の分類精度は依然として課題があるとされている。

③ 医療経済評価のモデリングへの応用

生成 AI・LLM は、費用効果分析のモデルの概念化、パラメータ化、モデルの実装、モデル結果の評価と検証など、モデル開発のさまざまな段階をサポートする可能性がある。Reasonらは、GPT-4 を使って、公表されている 2 つの費用効果分析(腎癌・肺癌)を自動的にプログラムした[3]。R でモデルをプログラムするよう指示するプロンプトを開発し、各モデルの方法、仮定、パラメータ値の説明を提供した。生成されたスクリプト(15 回試行)の結果を、人間がプログラムしたオリジナルモデルの公表値と比較したところ、肺癌モデルの 93%(14/15)は腎癌モデルの 60%(9/15)は完全にエラーがなかったとされ

る。また、エラーのないモデルスクリプトは、公表された増分費用効果比を 1%以内で再現した。こうした利用事例がある一方で C 型肝炎のマルコフモデルの構築を、2 種類の基礎モデルで実行したところ、モデル間で疾患進行の概念化に大きなばらつきがあり、専門家による指導が不可欠であることが考察されている[4]。また、パラメータ誤りやコーディングエラーのリスクがあるため、解析者は慎重に利用する必要がある。

3. Chat-GPT のカスタマイズ

Chat-GPT に役割を与え、所定の作業を自動化することが可能である。Turgay Ayer 氏は、昨年の国際学会 ISPOR において、一連のプロンプトを入力し、Chat-GPT に医療経済評価のためのモデル構築をサポートするアシスタントとして機能させるデモンストレーションを行った[5]。ここでは、そのときのプロンプトの記載を日本語に翻訳し、アレンジしたものを紹介する。

①最初に Chat-GPT に、1)を入力する。

1) 以下、一連のプロンプトを順番に提示します。各プロンプトの後、「プロンプトの続きを入力してください」とだけ言ってください。私が「プロンプトは以上です」と言うまで、これを続けてください。私が「プロンプトは以上です」と言ったら、すべての情報を一度に処理してください。いいですか？

②次に、2)~9)を入力する。

2) HEOR モデラーのアシスタントとして行動してください。あなたのタスクは、マル

コフ連鎖の HEOR モデルを構築することです。まず、疾患、介入、国について質問します。不足しているものがあれば、再度質問します。

3) 3 つの項目すべてが提供されたら、以下の作業を行います。

- 選択された疾患、介入、および国の概要を提示します。

- 関連するマルコフ費用対効果研究を検索し、健康状態、介入、および引用を一覧表に記載します。

- ユーザーがさらに研究を追加したいかどうかを尋ねます。該当する研究が見つからない場合は、「この特定の疾患、介入、および国に関する関連研究は見つかりませんでした」と伝えます。

4) 次に、ユーザーにモデルパラメータを提示します。

- 健康状態：公表された研究に基づいていくつ提案します。

- 人口規模：特に指定がない限り、デフォルトは 100,000 です。

- タイムホライズン：ユーザーに月単位の値を尋ねます。

5) 各パラメータについて、ユーザーの確認を待ってから次に進みます。すべて確認されたら、モデル構造を要約し、「確認」または「変更」を求めます。

6) 次に、調査に基づいて遷移確率を推定します。表形式で値を表示し、ユーザーに変更の可否を尋ねます。

7) ユーザーがすべてを確認した後、以下を表示します。

- 「パラメータ」と「値」の列を持つ「**モデルパラメータ表**」

- 「From」、「To」、および「Probability」の列を持つ「**遷移確率表**」

8) その後、適切なモデルを構築します。モデルが構築されたら、モデルの R コードを表示し、ベースケースの結果を表示します。その後、ユーザーにこのモデルで何をしたいかを尋ね、それらのタスクを実行します。

9) 常にユーザー入力を待ち、次のステップを想定しないようにします。正確かつ精密に、無関係な質問には回答しないようにします。

③最後に以下を入力する。

「プロンプトは以上です」

そうすると、Chat-GPT はプロンプトにある通り、モデル構築のアシスタントとして、使用者のリサーチクエストに沿ったモデルの概念化、仮想的なパラメータの設定、Python によるモデル構築を実行してくれる。また、追加の指示を与えることで R のコーディングを行わせることも可能である。

4. Chat-GPT によるモデル構築の支援

現在、特に費用効果分析のモデリングにおいて生成 AI・LLM 活用の検討が進んでいる。今回は、日本の高齢者に対する RS ウィルス感染症重症化予防のためのワクチンの

費用対効果をワクチン未接種群と比較するためのマルコフモデルを Chat-GPT のプロンプトを使用して構築した。まず、本テーマに関する関連文献の検索が行われ、健康状態の検討が行われた。

資料 1 に Chat-GPT による出力結果、資料 2 に文献値等を参考に Chat-GPT が設定したパラメータ値を示す。

基本分析の結果、ICER は 59474 円/QALY と報告された。(資料 3)

その他、コホートトレースや感度分析の結果、使用者の依頼に応じて様々な図表を出力することが確認された。(資料 4)

しかしながら、Chat-GPT による計算過程の詳細やエラーの有無を確認するためには、R のコード等を出力させた上で、経験のあるモデラーによる精査が必要不可欠と考えられた。(資料 5)

D. 考察

Chat-GPT を活用した HTA モデリングは、以下の観点から一定の有用性が確認された。

1. 実装効率の向上: モデル構造の設計、パラメータ抽出、R コードの自動生成など、従来人手に依存していた作業の一部を自動化できた。
 2. 研究設計支援: 関連文献の要約、アウトカム抽出、パラメータ化といった初期設計において、プロンプトベースで反復的に仮説構築を行える点は有益であった。
- 一方で、利用における以下のような限界や課

題を認識しておく必要がある。

① 妥当性、信頼性、透明性、説明責任

・WEB から一般に入手できる大規模データで学習されるため、医療など専門分野に応用した場合のエラー、幻覚(捏造)のリスクがある。

・従来の統計ツールよりも複雑で、ユーザーやLLMに依存したばらつきが生じる。

② バイアスと公平性・公正性

・LLMの開発で混入したバイアスが伝播・増幅し、個人や社会に害をもたらすリスクがある。

③ 規制と倫理的配慮

・絶対的な非識別化は達成不可能であり、再識別化のリスクもゼロではないため、保護された健康情報を含むデータの使用は避けること。

なお、代表的なHTA機関であるイギリスのNICEは、AIを使用したエビデンスの生成に関連して、いち早く声明を公表しており、HTAの各種タスクにおける生成AI・LLMの利用可能性と留意点について整理を行っている[6]。

E. 結論

これまで紹介してきた通り、生成AI・LLMは、SRの実施や経済評価モデル構築を支援する有力なツールとなる。作業過程の一部を自動化し、エビデンスの統合、パラメータ化、レポート作成などの作業に必要な時間と労力を削減することができる。一方で、生成AI・LLMは、人間の専門家に完全に取って代わるのではなく、それを補強しサポートする立場にあることを理解する必要がある。今後、医療者・研究者は、生成AI・LLM使用の限界に留意すべきで、規

制・ルール、教育環境の整備に継続して取り組む必要がある。

F. 健康危険情報

特になし。

参考文献

1. Fleurence, Rachael L. et al. Generative Artificial Intelligence for Health Technology Assessment: Opportunities, Challenges, and Policy Considerations: An ISPOR Working Group Report. Value in Health, Volume 28, Issue 2, 175 – 183. 2025
2. スティーヴン・ウルフラム(著), 高橋聡(訳). ChatGPTの頭の中. ハヤカワ新書.2023
3. Reason, T., Rawlinson, W., Langham, J. et al. Artificial Intelligence to Automate Health Economic Modelling: A Case Study to Evaluate the Potential Application of Large Language Models. PharmacoEconomics Open 8, 191–203 (2024).
4. J. Chhatwal, I.F. Yildirim, D. Balta, et al. Can large language models generate conceptual health economic models? ISPOR <https://www.ispor.org/heor-resources/presentations-database/presentation/intl2024-3898/139128>
5. November 17: AI-Powered HEOR: Advancing Insights and Decisions with Large Language Models - In Person at ISPOR Europe 2024

<https://www.ispor.org/conferences-education/event/2024/11/17/default-calendar/november-17-ai-powered-heor-advancing-insights-and-decisions-with-large-language-models---in-person-at-ispor-europe-2024>

6. Use of AI in evidence generation: NICE position statement.

https://www.nice.org.uk/about/what-we-do/our-research-work/use-of-ai-in-evidence-generation--nice-position-statement?utm_medium=social&utm_source=linkedin&utm_campaign=aiposition

人工知能 (AI : artificial intelligence)

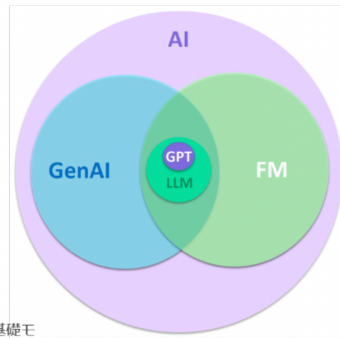
「大量の知識データに対して、高度な推論を的確に行うことを目指したもの」
(人工知能学会)

生成AI (Generative AI)

入力データ(プロンプト)に基づいてテキスト、画像、またはその他のコンテンツを生成することができるAIシステム

大規模言語モデル(Large Language Models)

膨大なテキストデータで学習された特定のタイプの基礎モデルで、膨大なデータセットから得られた知識に基づいて、テキストやその他のコンテンツを認識、要約、翻訳、予測、生成する



基礎モデル(Foundation Models)

様々な目的を果たす機械学習モデル。大規模なデータで訓練され、微調整の有無にかかわらず、幅広いタスクに適用可能。

GPT (Generative Pre-trained Transformer)

OpenAIが開発したLLM。特に人間のようなテキストを生成するのに適している。

図 1. AI 関連用語

[1]より引用改変

The best thing about AI is its ability to…
(AIの一番の長所としてあげられるのは)



次に来る単語	確率
Lean	4.5%
Predict	3.5%
Make	3.2%
Understand	3.1%
Do	2.9%

Temperature(温度)というパラメータ設定により、
ときどきランクの低い単語をランダムに選ぶ。

図 2. Chat-GPT による生成のイメージ

[2]より引用改変

関連研究の概要

日本における高齢者（60歳以上）を対象としたRSVワクチン（RSVPreF3 OA）の費用効果分析が行われています。[Taylor & Francis Online+3ResearchGate+3PubMed+3](#)

- **研究名：**「Cost-effectiveness analysis of respiratory syncytial virus vaccination with the adjuvanted prefusion F protein vaccine (RSVPreF3 OA) for adults ≥ 60 years old in Japan」
- **著者：**Akiko Mizukami et al.
- **掲載誌：**Expert Review of Vaccines, 2024年
- **モデル：**静的マルコフモデル
- **タイムホライズン：**3年間
- **割引率：**2%
- **視点：**医療支払者および社会的視点

[ResearchGate+7PubMed+7ResearchGate+7ResearchGate](#)

この研究では、RSVワクチン接種により以下の効果が示されています：

- 約100万件のRSV感染症の予防
- 約6,840人の死亡の回避
- 109,119 QALYの獲得
- ICER（増分費用効果比）は、医療支払者の視点で約4,180,084円/QALY、社会的視点で約4,041,917円/QALY

[IQVIAPubMed+1ResearchGate+1](#)

これらの結果は、感度分析においても頑健であり、RSVワクチン接種が日本の高齢者にとって費用効果的な介入であることを示唆しています。

モデル構造の提案

上記の研究に基づき、以下のようなマルコフモデル構造を提案します：

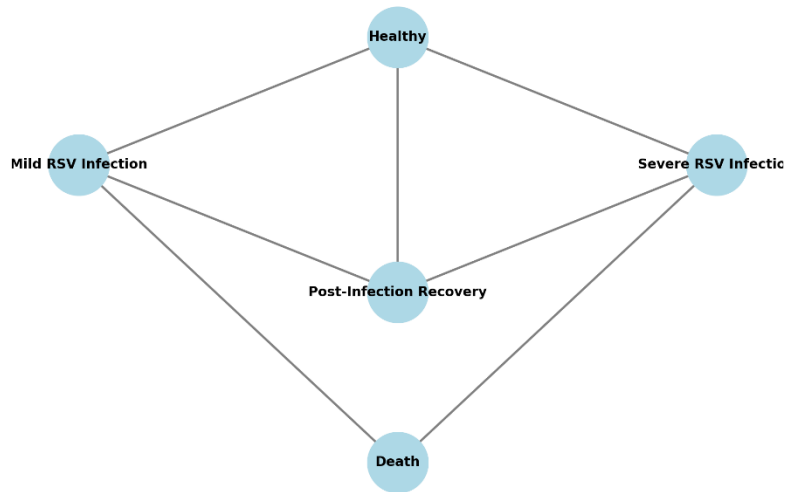
健康状態（Health States）

1. **Healthy**（健康）
2. **Mild RSV Infection**（軽度のRSV感染症）
3. **Severe RSV Infection**（重度のRSV感染症）
4. **Post-Infection Recovery**（感染後の回復）
5. **Death**（死亡）

[PMC+1SpringerLink+1ResearchGate](#)

この構造は、RSV感染症の進行と回復、及び死亡リスクを反映しています。

Markov Diagram for RSV in Older Adults (Japan)



人口規模とタイムホライゾン

- 人口規模: 1,000 人 (デフォルト)
- タイムホライゾン: 36 ヶ月 (3 年間)

これらの設定は、先行研究に基づいており、日本の高齢者における RSV 感染症の影響を評価するのに適しています。

資料2 文献値等を参考に、Chat-GPT が設定したパラメータ値

遷移確率

From	To	Probability (per month)
Healthy	Mild RSV Infection	0.0042
Healthy	Severe RSV Infection	0.0008
Mild RSV Infection	Post-Infection Recovery	0.9500
Severe RSV Infection	Post-Infection Recovery	0.7500
Mild RSV Infection	Death	0.0020
Severe RSV Infection	Death	0.0200
Post-Infection Recovery	Healthy	1.0000

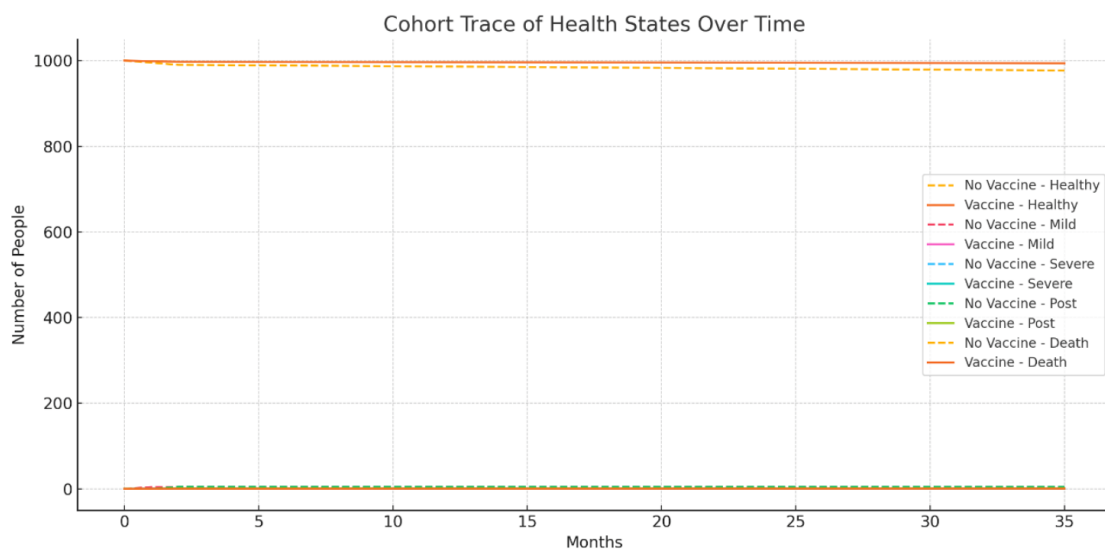
各種パラメータ

パラメータ	値	備考
Population size	1,000	初期の高齢者集団
Time horizon	36 ヶ月	月次マルコフサイクル
Discount rate (cost/QALY)	2% 年率	日本のガイドライン準拠
Cost of RSV vaccination	¥12,000	1 回接種、推定値
Cost of mild RSV infection	¥5,000	外来診療費＋投薬等
Cost of severe RSV infection	¥500,000	入院費用込みの推定
QALY (Healthy)	0.85/年 → 0.0708/月	高齢者の基礎効用
QALY (Mild RSV)	0.75/年 → 0.0625/月	軽度感染の期間的効用
QALY (Severe RSV)	0.50/年 → 0.0417/月	入院など重症時の低下
QALY (Post-infection)	0.80/年 → 0.0667/月	回復後一時的な QOL 低下あり
Vaccine efficacy (mild)	70%	ワクチンで軽症 RSV 予防
Vaccine efficacy (severe)	80%	ワクチンで重症 RSV 予防

資料3 Chat-GPTによる基本分析の結果

Arm	Cost (¥)	QALY	ICER (¥/QALY)
No Vaccine	14080856	2457.02	-
Vaccine	14919303	2471.12	-
Incremental	838447	14.1	59474

資料4 Chat-GPTによるコホートトレースの結果



資料5 デモンストレーションで Chat-GPT が出力した R のコード

```
# 必要なパッケージ
library(dplyr)

# モデル設定
cycles <- 36
population <- 1000
discount_rate <- 0.02 / 12 # 月割引率

# 状態名
states <- c("Healthy", "Mild", "Severe", "Post", "Death")
n_states <- length(states)

# 初期状態
state_vector <- rep(0, n_states)
state_vector[1] <- population # 全員 Healthy から

# QALY (月)
qaly <- c(Healthy = 0.0708, Mild = 0.0625, Severe = 0.0417, Post = 0.0667, Death = 0.0)

# コスト (月)
cost <- c(Healthy = 0, Mild = 5000, Severe = 500000, Post = 0, Death = 0)
vaccine_cost <- 12000 # ワクチン初期費用

# 遷移確率 (非接種)
tm_no_vaccine <- matrix(c(
  1 - 0.0042 - 0.0008, 0.0042, 0.0008, 0, 0,
  0, 0, 0, 0.95, 0.002,
  0, 0, 0, 0.75, 0.02,
  1, 0, 0, 0, 0,
  0, 0, 0, 0, 1
), nrow = n_states, byrow = TRUE)

# ワクチン有効性適用 (軽症 70%、重症 80%)
mild_eff <- 0.70
severe_eff <- 0.80
tm_vaccine <- matrix(c(
  1 - 0.0042 * (1 - mild_eff) - 0.0008 * (1 - severe_eff),
  0.0042 * (1 - mild_eff),
  0.0008 * (1 - severe_eff),
  0, 0,
  0, 0, 0, 0.95, 0.002,
  0, 0, 0, 0.75, 0.02,
  1, 0, 0, 0, 0,
  0, 0, 0, 0, 1
), nrow = n_states, byrow = TRUE)

# マルコフモデル関数
run_markov <- function(transition_matrix, with_vaccine = FALSE) {
  state <- state_vector
  total_cost <- 0
  total_qaly <- 0
  for (cycle in 0:(cycles - 1)) {
    discount <- 1 / ((1 + discount_rate) ^ cycle)
    total_cost <- total_cost + sum(state * cost) * discount
    total_qaly <- total_qaly + sum(state * qaly) * discount
  }
}
```

```

    if (with_vaccine && cycle == 0) {
      total_cost <- total_cost + population * vaccine_cost
    }
    state <- state %*% transition_matrix
  }
  return(list(cost = total_cost, qaly = total_qaly))
}

# 実行
res_no_vaccine <- run_markov(tm_no_vaccine, FALSE)
res_vaccine <- run_markov(tm_vaccine, TRUE)

# 結果計算
inc_cost <- res_vaccine$cost - res_no_vaccine$cost
inc_qaly <- res_vaccine$qaly - res_no_vaccine$qaly
icer <- inc_cost / inc_qaly

# 結果出力 (サウザンドセパレータ付き)
cat("【ベースケース結果】 ¥n")
cat(sprintf("非接種群: 費用 = ¥%s, QALY = %.2f¥n", format(round(res_no_vaccine$cost), big.mark=","),
res_no_vaccine$qaly))
cat(sprintf("接種群   : 費用 = ¥%s, QALY = %.2f¥n", format(round(res_vaccine$cost), big.mark=","),
res_vaccine$qaly))
cat(sprintf("増分費用: ¥%s¥n", format(round(inc_cost), big.mark=",")))
cat(sprintf("増分 QALY: %.2f¥n", round(inc_qaly, 2)))
cat(sprintf("ICER   : ¥%s/QALY¥n", format(round(icer), big.mark=",")))

```