

令和 7 年度 厚生労働科学研究費
(医薬品・医療機器等レギュラトリーサイエンス政策研究事業)

AI を用いた医療情報の医薬品安全への活用に向けた諸要件の調査研究
分担研究報告書

個別症例評価 AI シミュレーションモデルの検討：
生成 AI を用いた副作用既知・未知判断支援と段階的症例報告の因果関係評価

研究分担者 岡本里香
神戸医療産業都市推進機構 医療イノベーション推進センター 学術研究推進室
室長

研究要旨

海外では、薬事規制への AI 技術導入に向けたガイダンス作成やファーマコビジランス (PV) 分野における AI 活用に関する議論が進みつつある。本研究では、日本における PV 分野の AI 活用に関するガイダンス作成に資する研究として、PV 分野における AI 活用上の留意点を幅広く整理・抽出することを目的とした。2023 年～2024 年度にかけて、国内製薬企業及び国内 IT 企業を対象に、PV 業務における AI 活用状況に関するアンケート調査及び個別ヒアリング調査を実施し、PV 分野での AI 活用の共通課題、PV プロセスごとの課題、並びに規制上の留意点を整理した。2024 年度には、これらの調査結果を踏まえ、AI 活用シミュレーションモデルとして副作用因果関係評価に関するモデル作成に着手した。具体的には、副作用の既知・未知判断を支援することを目的として、部分一致検索処理を活用した副作用既知・未知判断支援モデルを作成した。最終年度である 2025 年度には、当該モデルの出力精度向上を目的として生成 AI 技術を適用し、副作用既知・未知判断支援における AI 活用の可能性及び課題を整理した (1.副作用既知未知判断支援モデル作成)。さらに、新たなシミュレーションモデルとして、報告初期段階から追加情報の蓄積に伴って更新される段階的症例報告を対象に、生成 AI を用いた副作用因果関係評価を探索的に検討し、段階的な症例情報に基づく因果関係評価における生成 AI 活用の可能性、課題等を抽出した (2. 生成 AI を用いた段階的症例報告の因果関係評価)。

1. 副作用既知未知判断支援モデル作成

協力者：太田実紀（東京大学）

A. 研究目的

2025 年度は 2024 年度に作成した既知・未知判断支援システムの有用性、品質、運用

上の課題を検討し、当該課題に対して、生成 AI の活用可能性を検討した。

なお、シミュレーションモデルの追加機能設計は株式会社 CMC エクスメディカに外部委託した。

B. 研究方法

データソースは 2024 年度と同様に、独立行政法人医薬品医療機器総合機構「医薬品副作用データベース」(JADER) を用いた。対象は、医薬品副作用報告データセット (2004 年 4 月～2024 年 12 月) に含まれる表 1 の 8 薬剤 (*詳細は令和 6 年度報告書) とした。

表 1 解析対象薬剤一覧

一般的名称 / 販売名
ニボルマブ (遺伝子組換え) / オブジーボ点滴静注 20mg, 100mg, 120mg, 240mg
ペムブロリズマブ (遺伝子組換え) / キイトルーダ点滴静注 100mg
イピリムマブ (遺伝子組換え) / ヤーボイ点滴静注液 20mg 50mg
リバビリン / レベトールカプセル 200mg
アビキサバン / エリキュース錠 2.5mg, 5mg
ラモトリギン / ラミクタール錠 25mg・100mg/小児用 2mg・5mg, 他後発 5 社
リツキシマブ (遺伝子組換え) / リツキサン点滴静注 100mg, 500mg, 他後発 2 社
バンコマイシン塩酸塩 / 塩酸バンコマイシン散 0.5g, 他, 剤・静注・点眼あり, 後発あり

2024 年度に作成した AI 技術を用いない部分一致検索処理技術による既知・未知判断支援モデルをモデル v1.0 とした。本年度は、モデル v1.0 の結果を精査し、添付文書

から検出できなかった記載パターン (複合語) に注目して抽出ルールを検討した。その上で大規模言語モデル (LLM) である ChatGPT (GPT-5 系モデル) (以下、GPT) を用いた複合語分解・抽出機能をモデル v1.0 に機能追加し、モデル v2.0 として構築した。

GPT による複合語分解結果の精度評価は、機械学習や情報検索などで一般的に用いられる適合率、再現率及び F 値に基づいて行った。図 1 に示すとおり、添付文書から人手で検出した複合語を人間抽出群 (正解群)、モデルが抽出した結果をシステム抽出群とした。評価指標の定義は以下のとおりとした。

- True Positive (TP) : 添付文書に存在する複合語構造 (表 2) を含む記載があり、かつシステムで正しく検出された事象
- False Positive (FP) : 複合語構造を有さない、又は対象としない記載を複合語として誤って検出された事象
- False Negative (FN) : 添付文書に複合語構造を有する記載があるにもかかわらず検出できなかった事象

再現率は $TP / (TP + FN)$ 、適合率は $TP / (TP + FP)$ として評価した。

使用した添付文書は、モデル v1.0 との比較可能性を確保するため、2024 年度の検討時点で取得した版を用いた。辞書データは、意味的に包含される事象も適合範囲に含めるため、2024 年度と同様に、ティ辞書企画社「医薬用語シソーラス辞書 T 辞書」の病名 T 辞書 (以下、T 辞書) に収載された上位語、下位語、同義語等を使用した。

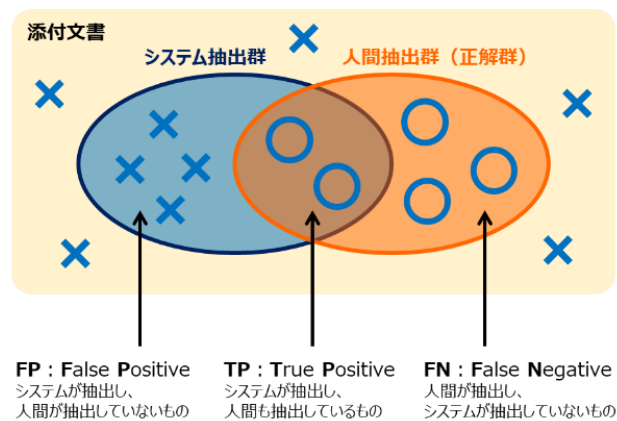


図1 GPTによる複合語分解結果の評価区分 (TP/FP/FN) の定義

C. 結果・考察

モデル v1.0 について、2024 年度の結果を用いて意味的包含事象も含めた適合性を確認したところ、図 2 のように、JADER の報告事象 (オリジナル) 「脳出血」に対し、添付文書の「脳出血」に加え、同義語、上位語、下位語に相当する記載を検出すべきところ、「頭蓋内又は脊髄内出血」に含まれる「頭蓋内出血」が検出されていなかった (図 2 の赤矢印)。そこで、検出漏れが生じるパターン

表 2 未検出となる複合語パターン

「・」 (中黒)
[A]・[B]致死作用, [A]・[B]痛, [A]・[B]炎, [A]・[B]死亡, [A]・[B]のびらん, [A]・[B]や[C]のびらん, 発疹は[A]・[B], [A]・[B]・[C]等の低下, 消化管[A]・[B]
「、」 (読点)
[A], [B]上昇, [A], [B]の上昇, [A], [B], [C], [D], [E]等の上昇, [A], [B], [C]の低下
「及び」
[A]及び[B]ミオグロビン上昇
「又は」
[A]又は[B]出血, [A]若しくは[B] (文章) 又は[C] (文章) 増加
「や」
[A]や[B]の上昇, [A]や[B]の低下

副作用 既知・未知判定ツール

販売名 エリキュース錠2.5mg/エリキュース錠5mg
 一般的名称 アピキサバン 薬効分類名 経口FXa阻害剤 製造販売業者等 製造販売元/ブリistol・マイヤーズ スクイブ株式会社
 販売元/ファイザー株式会社

添付文書情報

基本情報

1.警告

2.禁忌 (次の患者には投与しないこと)

3.組成・性状

3.1組成

3.2製剤の性状

4.効能又は効果

5.効能又は効果に関連する注意

6.用法及び用量

7.用法及び用量に関連する注意

8.重要な基本的注意

9.特定の背景を有する患者に関する注意

9.1合併症・既往歴等のある患者

9.2腎機能障害患者

9.3肝機能障害患者

9.5妊婦

9.6授乳婦

11.2 その他の副作用

	1%以上	1%未満	頻度不明
免疫系障害			過敏症 (皮疹等の薬物過敏症、アレルギー性浮腫等のアナフィラキシー反応等)
神経系障害		味覚異常、 くも膜下出血 、三叉神経痛	脳出血 、頭蓋内又は脊髄内出血 (硬膜下血腫 及び脊髄血腫等)
眼障害	眼出血	眼充血	
血管障害		血腫	腹腔内出血
呼吸器、胸部及び縦隔障害	鼻出血	喀血、咳嗽	気道出血 (肺動脈出血、喉頭出血、及び咽頭出血等)
胃腸障害	歯肉出血、胃腸出血、消化不良、便潜血陽性	口腔内出血、便秘、腹部不快感、上腹部痛、血便排泄、下痢、逆流性食道炎、悪心	直腸出血、痔出血、後腹膜出血、吐血、マロリー・ワイス症候群、出血性消化性潰瘍
肝胆道系障害		血中ビリルビン増加、γ-GTP増加、肝機能異常	

検索キーワード:
 脳出血 (オリジナル) ピンク
 上位語, 同義語 黄色

図2 部分一致検索処理技術における意味的包含事象の未検出例

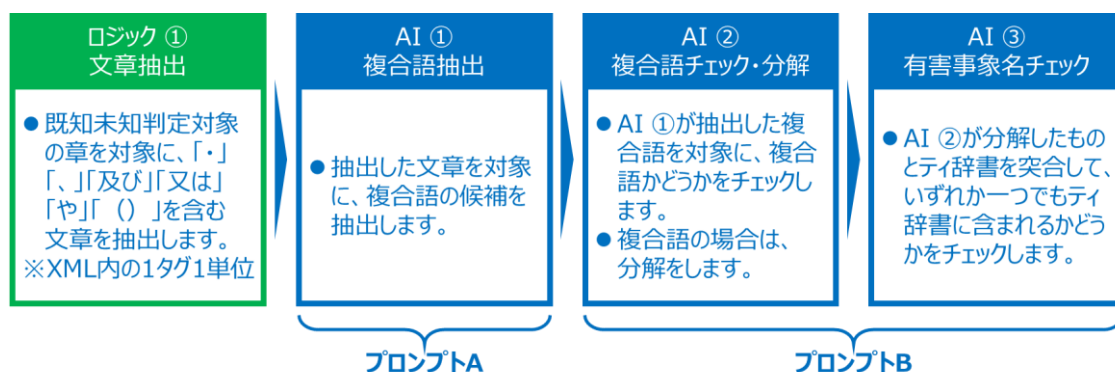


図3 複合語抽出及びGPTによる複合語分解処理の概要

ンを調査したところ、表2に示す複合語パターンが、部分一致検索処理技術のみでは抽出できていないことが明らかとなった。

部分一致検索処理技術は、表記ゆれや複合語構造を考慮できず、意味的に同一又は包含関係にある事象を十分に検出できない点が課題であると考えられた。そこで、複合語抽出を補完するため、自然言語の文脈理解及び構文分解に有用と考えられる生成AI（ここでは、GPT）を用いて、複合語の分解を試みた。図3に示すとおり、まず、添付文書内の複合語パターンを含む文章をルールベースで抽出し（ロジック①）、当該文章から複合語の例を5件提示したプロンプトAにより複合語を抽出した（AI①）。しかし、AI①の結果には、意図しない抽出が多く含まれていたため、続いて、プロンプトAで抽出した複合語を分解し、分解後の用語

が複合語であるかを確認する処理（AI②）、さらに、分解後の用語がT辞書に収載されているかを確認する処理（AI③）をプロンプトBにより実行した。結果を別紙1（アピキサバンのみ）に示す。

複合語分解の精度評価結果を表3に示す。再現率は、多くの薬剤において比較的高値を示し、複合語に起因する事象の検出漏れを一定程度補完できる可能性が示された。一方で、適合率は薬剤間でばらつきが大きく、一部の薬剤では0.2程度と低値を示し、誤検出（FP）の増加が認められた。F値はニボルマブ（0.83）、アピキサバン（0.82）で高く、イピリムマブ（0.29）、リツキシマブ（0.32）は低く、薬剤によってモデルの性能差が大きいことが示された。

以上より、GPTは網羅的な候補抽出には有効である一方、出力にはノイズが含まれ

表3 GPTによる複合語分解の精度評価

一般名	TP	FP	FN	FNFP*	再現率	適合率	F値
アピキサバン	7	1	2		0.78	0.88	0.82
イピリムマブ（遺伝子組換え）	1	4	1		0.50	0.20	0.29
ニボルマブ（遺伝子組換え）	17	6	1		0.94	0.74	0.83
バンコマイシン塩酸塩散	0	1	0				
ペムブロリズマブ（遺伝子組換え）製剤	10	10	2		0.83	0.50	0.63
ラモトリギン	4	2	3		0.57	0.67	0.62
リツキシマブ（遺伝子組換え）製剤	3	11	0	1	0.75	0.20	0.32
リバビリン	4	15	2		0.67	0.21	0.32
* 同一文書にFNとFPが発生した事例である。各1件として評価対象とする。							

るため、辞書ベースのフィルタリングや人手による確認を組み合わせた運用が必要であると考えられた。また、薬剤によって精度に差が認められたことから、複合語構造や添付文書における記載様式の違いがモデル性能に影響を与える可能性が示唆された。

False (FN、FP 及び FNFP) の内訳は表 4 のとおりで、FN 事象 11 件の内容を別紙 2 に示す。

表 4 False の内訳

False内訳	合計 / 件数
FN	11
抽出なし または 複合表現チェックがFALSE	4
抽出範囲が広い (FN)	3
有害事象表現チェックがFALSE	4
FNFP	1
有害事象表現チェックがFALSE	1
FP	50
過剰分解	4
人間系のチェック漏れ	17
抽出範囲が広い (FP)	1
複合表現でない	27
有害事象表現チェックがFALSE	1
総計	62

FN の分類として「抽出なし または 複合表現チェックが FALSE」4 件では、AI①による検出ができなかった。このうち 1 件は抽出の失敗、3 件は、AI①が複合表現として認識しなかったことが要因と考えられた。次に「抽出範囲が広い」3 件は、AI①による抽出はされたものの、分解後の用語が適切でなかった。例えば、「胚・胎児死亡率」の「・」に対して「胚死亡率、胎児死亡率」に分割できたものの、最終的に抽出された用語には「率」が含まれていた。また「AST 増加、ALT 増加」として抽出すべきところ、「基準値上限の 3~5 倍」が「増加」の前に記述されていることにより、適切な分解ができていないケースが認められた。AI②では、複合語と判断した理由を求めるプロン

プトを含めており、分解結果の説明を「解説」として出力している。当該解説は、FN 事例を低減するためのプロンプトチューニングに活用可能であると考えられる。「有害事象表現チェックが FALSE」4 件は、T 辞書には「・」で接続した用語として存在しているため、プロンプト A で「・」を分解した結果に対しては、「T 辞書にない」と判断されたケースである。また、そもそも T 辞書にない有害事象のため、FALSE となったケースも認められた。これらに対しても、プロンプトチューニングや辞書の使用方法などで、改善可能と考えられる。FNFP が 1 件あり、これは 1 文中の異なる箇所有害事象表現 TRUE となる箇所と、複合表現 FALSE となる箇所が含まれるケースであった。このケースについてもプロンプト A のチューニングにより適正化できる可能性があると考えられる。一方、FP 事象では、複合語ではない記載を複合語として誤認するケースが 27 件と多く認められた。これらに対しては、構文情報（係り受け関係）を考慮したフィルタリングや生成 AI による再判定処理等の導入が必要であると考えられた。一方で、人手によるチェック漏れをモデル v2.0 で 17 件抽出できており、人的確認の補助としての有用性が示唆された。

既知・未知判断支援システムでは、業務効率化が期待されるが、本システムの適用により FP が増加することは、業務効率化の観点では負の結果である。しかし、FPの中には、人手による抽出見落としを補完し得る事例があり、真の誤検出と人手抽出結果の見直しを要する事例に分類して評価する必要がある。一方、FN は本来抽出すべき事象を検出できていない事例であり、安全性評

価業務においては重要なリスクとなる。今回検出された FN 事象 11 件の内容から、前述の通り、プロンプトチューニングなどでの改善の可能性は示唆されたが、より対象薬剤を拡大した精査が必要と考える。また、プロンプト B の AI②において、AI①で抽出した複合語に対して、複合語と認識した理由を「解説」として出力する機能を、FN 事象に対しても同様に、プロンプト A にロジックで抽出した文章に対して、なぜ複合語がないと判断したのか、その理由を出力する指示を含めることにより、処理ロジックの透明性及び説明可能性を高めることができる可能性がある。

薬剤別の FN 及び FP 事象数については、表 3 に示すとおり、FN は全体数も少ないため、傾向分析は難しいが、薬剤間で件数の差異はみられなかった。一方、FP 数については、リバビリン 15 件、リツキシマブ（遺伝子組み換え）11 件、ペムブロリズマブ（遺伝子組換え）10 件が高値であった。今後、各薬剤の添付文書での記載の特徴、例えば、「又は」の頻度が高いなどを整理し、プロンプト設計に反映する必要があると考える。

以上より、GPT の導入は事象の検出漏れを補完する可能性を有する一方、誤検出の

抑制及び出力の再現性確保が課題として示された。

さらに、既知・未知判断支援機能の高度化を目的として、報告事象と添付文書記載との近接性に基づく分類を作成した。本分類では、報告事象が添付文書に同一事象で記載されているもの、同義語として記載されているもの、上位語または下位語として記載されているもの等を、添付文書記載との近接性に応じて優先度 10～1 に分類した（図 4）。本分類においては、優先度 10 を最も既知性が高い事象、優先度 1 を添付文書記載との近接性が最も低く、追加確認を要する事象と定義した。ただし、本分類における「優先度」は、添付文書記載との近接性に基づく便宜的な既知性分類であり、安全対策上の緊急度や臨床的重要性を直接示すものではない。

モデル v2.0 の検出結果を優先度分類した結果を図 5 に示す。意味的包含事象も含めて、添付文書に該当する記載が確認されなかった事例、すなわち優先度 1 に分類された事例の割合は 8 薬剤において 47.1%（リバビリン）～68.9%（バンコマイシン塩酸塩）であり、いずれの薬剤でも最も高かった。一方で、オリジナル及び同義語として添付文

優先度	オリジナル	同義語	上位語-1	上位語-2	上位語-3	下位語-1	下位語-2	下位語-3	同階層
取扱い提案		"有"の場合、オリジナルが"有"の場合と同義	"有"の場合、オリジナルが包含されている可能性が高いが、広義であることに注意			"有"の場合、オリジナルが包含されている可能性はあるが、狭義のため、要注意			"有"の場合でも、オリジナルとは別物である可能性が高く、"無"と扱う方がよい
1	無	無	無	無	無	無	無	無	無
2	無	無	無	無	無	無	無	無	有
3	無	無	無	無	無	無	無	有	無
4	無	無	無	無	無	無	有	無	無
5	無	無	無	無	無	有	無	無	無
6	無	無	無	無	有	無	無	無	無
7	無	無	無	有	無	無	無	無	無
8	無	無	有	無	無	無	無	無	無
9	無	有	無	無	無	無	無	無	無
10	有	無	無	無	無	無	無	無	無

図 4 優先度基準

書に記載がある事例、すなわち優先度 9 及び 10 に分類された事例の割合は、8.8% (バンコマイシン塩酸塩) ～16.9% (ニボルマブ (遺伝子組換え)) であった。モデル v2.0 での検出数とモデル v1.0 での検出数を比較したところ、多くの薬剤でモデル v2.0 の方が、添付文書上の関連事象をより広く抽出できており、GPT による複合語分解・抽出機能の追加が、部分一致検索処理技術の限界を補完する可能性が示された。

以上より、GPT を追加することにより、部分一致検索処理技術のみでは検出が困難であった複合語や意味的包含事象を補完的に抽出できる可能性が示された。一方で、文脈依存表現や複数文にまたがる記述、否定・条件付き表現への対応、ならびに誤抽出の抑制や出力の再現性確保は依然として課題である。また、添付文書の記載様式の違いがモデル性能に影響する可能性があることから、添付文書記載の標準化やデータソースの調整も、今後の検討課題と考える。

D. 結論

本研究では、JADER 及び添付文書を用いた既知・未知判断支援シミュレーションモデルに対し、生成 AI を活用した複合語の分解及び抽出機能を追加することで、従来の部分一致検索処理技術のみでは検出が困難であった複合語や意味的包含事象を補完的に抽出できる可能性を検討した。その結果、GPT を追加したモデル v2.0 では、複合語に起因する検出漏れを一定程度補完できる可能性が示され、既知・未知判断における網羅性向上に資することが示唆された。

一方で、GPT の導入により再現率は向上したものの、適合率には薬剤間でばらつきがあり、一部の薬剤では誤検出の増加が認められた。このことから、本手法は確定的な既知・未知判断をモデル単独でおこなうものではなく、網羅的な候補抽出やスクリーニングを目的とした支援的利用に位置付けることが適切であると考えられた。今後は、構文情報を活用したフィルタリングや生成

分類	優先度	アピキサバン	イビリムマブ (遺伝子組換え)	ニボルマブ (遺伝子組換え)	バンコマイシン 塩酸塩	ベムプロリスマ ブ (遺伝子組換え)	ラモトリギン	リツキシマブ (遺伝子組換え)	リハビリン
有害事象数		1151	1475	1982	717	1657	1290	1305	1241
なし	1	735(63.9) ▼7	754(51.1)	913(46.0) ▼45	494(68.9)	862(52.0) ▼37	779(60.4) ▼15	739(56.6) ▼19	584(47.1) ▼1
同階層	2	89(7.7) ▼4	104(7.1)	96(4.8) ▼35	37(5.2)	101(6.1) ▼9	106(8.2) △4	60(4.6) ▼15	92(7.4) ▼1
下位語-3	3	0	0	0	0	0	0	0	0
下位語-2	4	3(0.3)	2(0.1)	3(0.2)	1(0.1)	2(0.1)	4(0.3)	1(0.1)	0
下位語-1	5	8(0.7) ▼1	14(0.9)	16(0.8) ▼6	5(0.7)	20(1.2) △1	13(1.0) △1	6(0.5)	12(1.0)
上位語-3	6	25(2.2)	56(3.8)	120(6.1) △26	22(3.1)	73(4.4) △5	33(2.6)	57(4.4) △9	68(5.5) ▼2
上位語-2	7	62(5.4) △6	105(7.1)	200(10.1) △35	36(5.0)	152(9.2) △16	72(5.6) △3	126(8.8) △11	113(9.1)
上位語-1	8	98(8.5) △3	195(13.2) ▼1	300(15.1) △22	59(8.2)	220(13.3) △19	141(10.9) △3	156(9.7) △11	179(14.4) △2
同義語	9	32(2.8) △2	36(2.4)	53(2.7)	27(3.8)	37(2.2) △2	58(4.5) △1	55(4.2) △3	62(5.0)
オリジナル	10	99(8.6) △1	209(14.2) △1	281(14.2) △3	36(5.0)	190(11.5) △3	84(6.5) △3	105(8.0)	131(10.6) △2

図 5 優先度分類結果 (件数 (各薬剤における割合%), v2.0-v1.0 の増減)

AIによる再判定処理、プロンプトチューニング等により、誤検出の抑制と検出感度のバランスを改善する必要がある。

また、FPと分類された事例の一部には、人手による抽出時の見落としを補完し得る事例が含まれており、モデルv2.0は人的確認を補助するスクリーニングツールとして有用である可能性が示された。一方、FNは本来抽出すべき事象の検出漏れであり、安全性評価業務において重要なリスクとなるため、FN事例の要因分析に基づく処理ロジックの改善が不可欠である。

さらに、報告事象と添付文書記載との近接性に基づく分類を導入することで、安全対策検討における事例の候補を整理できる可能性が示された。ただし、本優先度分類は添付文書記載に基づく暫定的な既知性指標であり、安全対策上の緊急度、重篤性、発現頻度、因果関係評価等の臨床的重要性を直接示すものではない。実運用においては、これらの要素を統合した評価体系の構築する必要がある。

規制科学の観点からは、実運用時に個別症例情報を扱うことを踏まえた、個人情報保護及びデータガバナンスの確保、処理ロジックの透明性及び説明可能性の担保、ならびに出力結果に対する責任所在の明確化が不可欠である。また、従来のコンピュータ化システムバリデーションの枠組みに加え、生成AIの非決定性、プロンプト依存性およびモデル更新による出力変動を踏まえた評価・検証手法の構築が必要と考える。

以上より、生成AIの導入は既知・未知判断支援の高度化に寄与する可能性を有する一方で、実運用にあたっては、技術的課題への対応とともに、個人情報保護、品質確保、

信頼性担保、バリデーション及びガバナンスに関する包括的な検討を行い、ガイダンスとして体系的に整理することが必要である。

E 研究発表

1. 論文発表

該当なし

2. 学会発表

(講演) 荒川憲昭、松永雄亮、笠原 成晃、水野 光、(パネラー) 佐井 君江、太田 実紀、(座長) 岡本 里香、守田 真: シンポジウム「ファーマコビジランス領域におけるAI活用のガイダンス案作成に向けて」第15回 レギュラトリーサイエンス学会学術大会, 2025年9月5日

別紙 1 複合語抽出及び GPT による複合語分解処理の結果（アピキサバンのみ）

ロジック ①	ロジック ①	AI ①	人間系	AI ②	AI ②	AI ②	AI ③	集計用フラグ	集計用フラグ	目検による判定	目検による判定	目検による判定	目検による判定
チャックタイトル	抽出元原文	抽出範囲	抽出範囲(ラベル)	個別副作名	解説	複合表現	有害事象表現	対象ラベル	ラベル種別	判定	結果ラベル	要因ラベル	要因
2. 禁忌	＜非弁膜症性心房細動患者における虚血性脳卒中及び全身性塞栓症の発症抑制＞	虚血性脳卒中及び全身性塞栓症		虚血性脳卒中, 全身性塞栓症	「虚血性脳卒中及び全身性塞栓症」は共通の修飾語（ここでは「の発症抑制」）にかかる2つの被修飾語が「及び」で並列になっている合成された表現です。それぞれ「虚血性脳卒中」、「全身性塞栓症」に分解されます。	TRUE	TRUE	TRUE	other	x	FN	複合表現でない	抽出対象は単純な並列表現
8. 重要な基本的注意	凝固能検査（プロトロンビン時間（PT）、国際標準比（INR）、活性化部分トロンボプラスチン時間（aPTT）等）は、本剤の抗凝固能をモニタリングする指標とはならないため、本剤投与中は出血や貧血等の徴候を十分に観察すること。また、必要に応じて、血算値（ヘモグロビン値）、便潜血等の検査を実施し、急激なヘモグロビン値や血圧の低下等の出血徴候を確認すること。臨床的に問題となる出血や貧血の徴候が認められた場合には、本剤の投与を中止し、出血の原因を確認すること。また、症状に応じて、適切な処置を行うこと。[1.1 参照][2.2 参照][11.1.1 参照]	急激なヘモグロビン値や血圧の低下等の出血徴候	ヘモグロビン値や血圧の低下	ヘモグロビン値低下、血圧低下	「急激なヘモグロビン値や血圧の低下等の出血徴候」は「急激なヘモグロビン値の低下」「急激な血圧の低下」といった複数の低下事象が共通の被修飾語「出血徴候」にかかっている表現であり、分解するとヘモグロビン値低下および血圧低下に対応する徴候に分けられます。	TRUE	TRUE	TRUE	target	○	TP		
10.2 併用注意	これらの薬剤がCYP3A4及びP-糖蛋白を同時に強力に阻害するため、本剤の代謝及び排出が阻害されと考えられる。		代謝及び排出が阻害			FALSE	FALSE	TRUE	target	x	FN	抽出なし または 複合表現チェックがFALSE	代謝・排出阻害の抽出失敗
10.2 併用注意	これらの薬剤のCYP3A4及び/又はP-糖蛋白の阻害作用により、本剤の代謝及び排出が阻害されと考えられる。		代謝及び排出が阻害			FALSE	FALSE	TRUE	target	(DUP)	(DUP)		
10.2 併用注意 (併用に注意すること)	これらの薬剤又はセイヨウオトギリソウがCYP3A4及びP-糖蛋白を同時に強力に誘導するため、本剤の代謝及び排出が促進されと考えられる。	本剤の代謝及び排出	代謝及び排出が促進	**本剤の代謝**、 **本剤の排出**	「本剤の代謝及び排出」は共通の修飾語「本剤の」が「代謝」と「排出」にかかっている表現であり、2つの用語に分解されます。分解すると「本剤の代謝」「本剤の排出」となります。	TRUE	FALSE	TRUE	target	x	FN	有害事象表現チェックがFALSE	代謝・排出阻害の抽出失敗
項目の中身 (データ)	AST、ALTの上昇等を伴う肝機能障害があらわれることがある。	AST、ALTの上昇等	AST、ALTの上昇	A S T の上昇、A L T の上昇	「AST、ALTの上昇等」は「AST」「ALT」という複数の項目に共通する修飾語「の上昇」がまとめてかかっている表現であり、分解すると各酵素の上昇を示す個別の用語に分けられます。辞書には全角表記の「A S T の上昇」「A L T の上昇」が含まれているため、それらを用いて分解しています。末尾の「等」はその他の類似する上昇を含むことを示しています。	TRUE	TRUE	TRUE	target	○	TP		
	脳出血、頭蓋内又は脊髄内出血（硬膜下血腫及び脊髄血腫等）	頭蓋内又は脊髄内出血	頭蓋内又は脊髄内出血	頭蓋内出血、脊髄内出血	「頭蓋内又は脊髄内出血」は「頭蓋内」と「脊髄内」が共通の被修飾語「出血」を持ち、又は接続された2つの用語からなる合成された表現です。それぞれ頭蓋内出血、脊髄内出血に分けられます。	TRUE	TRUE	TRUE	target	○	TP		
	呼吸器、胸郭及び縦隔障害	呼吸器、胸郭及び縦隔障害	呼吸器、胸郭及び縦隔障害	呼吸器系障害**、 胸郭障害**、**縦隔障害**	「呼吸器、胸郭及び縦隔障害」は「呼吸器」「胸郭」「縦隔」という複数の修飾語が共通の被修飾語「障害」にかかっている合成された表現です。それぞれ「呼吸器障害」、「胸郭障害」、「縦隔障害」に分解できます。辞書（入力の words）には「呼吸器系障害」が含まれるため、呼吸器にかかる障害は辞書語である「呼吸器系障害」を用いています。	TRUE	TRUE	TRUE	target	○	TP		
	腎及び尿路障害	腎及び尿路障害	腎及び尿路障害	腎障害**、**尿路障害**	「腎及び尿路障害」は「腎」と「尿路」が「及び」でつながり、共通の被修飾語「障害」をもつ2つの用語に分解できる合成された表現です。分解すると「腎障害」「尿路障害」になります。	TRUE	TRUE	TRUE	target	○	TP		
	生殖系及び乳房障害	生殖系及び乳房障害	生殖系及び乳房障害	**生殖系障害**、 乳房障害	「生殖系及び乳房障害」は「及び」でつながれた修飾語（生殖系、乳房）が共通の被修飾語「障害」をもつ合成された表現です。分解すると生殖系障害、乳房障害の2つに分けられます。	TRUE	TRUE	TRUE	target	○	TP		
	皮膚及び皮下組織障害	皮膚及び皮下組織障害	皮膚及び皮下組織障害	皮膚障害**、**皮下組織障害**	「皮膚及び皮下組織障害」は「皮膚」と「皮下組織」が共通の被修飾語「障害」を持つ合成された表現です。それぞれ「皮膚障害」「皮下組織障害」に分けられます。	TRUE	TRUE	TRUE	target	○	TP		

別紙2 FN (11件) 及びFNFP (1件)

	ロジック①	AI①	人間系	AI②	AI②	AI②	AI③	目検による判定	目検による判定	目検による判定
一般名2	抽出元原文	抽出範囲	抽出範囲(ラベル)	個別副作用名	解説	複合表現	有害事象表現	結果ラベル	要因ラベル	要因
アピキサバン	これらの薬剤がCYP3A4及びP-糖蛋白を同時に強力に阻害するため、本剤の代謝及び排出が阻害されると考えられる。	抽出結果なし	代謝及び排出が阻害	抽出結果なし	抽出結果なし	FALSE	FALSE	FN	抽出なしまたは 複合表現チェックがFALSE	代謝・排出阻害の抽出失敗
イビリマブ（遺伝子組換え）	AST若しくはALTが基準値上限の3倍超かつ5倍以下又は総ビリルビンが基準値上限の1.5倍超かつ3倍以下に増加した場合	抽出結果なし	総ビリルビン // 増加	抽出結果なし	抽出結果なし	FALSE	FALSE	FN	抽出なしまたは 複合表現チェックがFALSE	複合表現として抽出されていない。
ベムプロロズマブ（遺伝子組換え）製剤	AST若しくはALTが基準値上限の3～5倍又は総ビリルビンが基準値上限の1.5～3倍に増加した場合	抽出結果なし	総ビリルビン // 増加	抽出結果なし	抽出結果なし	FALSE	FALSE	FN	抽出なしまたは 複合表現チェックがFALSE	複合表現として抽出されていない。
ラモトリギン	レボノルゲストレルの血中濃度が減少し、血中卵胞ホルモン（FSH）及び黄体形成ホルモン（LH）が上昇し、エストラジオールが僅かに上昇したとの報告がある。	抽出結果なし	血中卵胞ホルモン（FSH）及び黄体形成ホルモン（LH）が上昇	抽出結果なし	抽出結果なし	FALSE	FALSE	FN	抽出なしまたは 複合表現チェックがFALSE	複合表現として抽出されていない。
ニボルマブ（遺伝子組換え）	妊婦又は妊娠している可能性のある女性には、治療上の有益性が危険性を上回ると判断される場合にのみ投与すること。妊娠サルを用いた出生前及び出生後の発生に関する試験において、10mg/kgの週2回投与（AUC比較で臨床曝露量の約6倍に相当する）により妊娠末期における胚・胎児死亡率あるいは出生児死亡率の増加が認められたが、催奇形性は認められなかった。また、出生児の成長及び発達に影響は認められなかった。なお、本剤は出生児の血清中で認められている。[9.4]	胚・胎児死亡率	胚・胎児死亡	**胚死亡率**、**胎児死亡率**	「胚・胎児死亡率」は「胚」と「胎児」が「・」で接続され、共通の被修飾語「死亡率」をもつ合成された表現です。それぞれ「胚死亡率」と「胎児死亡率」に分けられます。なお、語彙リストには「胎児死亡」が含まれますが、本表現と完全一致する「胎児死亡率」は含まれていません。	TRUE	FALSE	FN	抽出範囲が広い（FN）	末尾に率を含むため、範囲が適切でない
ベムプロロズマブ（遺伝子組換え）製剤	AST若しくはALTが基準値上限の3～5倍又は総ビリルビンが基準値上限の1.5～3倍に増加した場合	AST若しくはALTが基準値上限の3～5倍	AST若しくはALT // 増加	**ASTが基準値上限の3～5倍に増加**、**ALTが基準値上限の3～5倍に増加**	「AST若しくはALTが基準値上限の3～5倍」は「若しくは」でつながれた2つの検査項目（AST、ALT）に共通の修飾語「基準値上限の3～5倍に増加」がかかっている合成された表現です。それぞれ「ASTが基準値上限の3～5倍に増加」、「ALTが基準値上限の3～5倍に増加」に分解されます。入力の辞書には「AST増加」「ALT増加」等の関連語が含まれていますが、基準値上限の倍率まで含めた上記の分解後表現は辞書に含まれていません。	TRUE	FALSE	FN	抽出範囲が広い（FN）	症病名判定ミス
ラモトリギン	動物を用いた生殖発生毒性試験において催奇形性作用は認められなかったが、本剤はジヒドロ葉酸還元酵素に対し弱い阻害作用を有するため、妊娠中に本剤を投与した場合、胎児奇形を誘発する危険性が考えられる。また、ラットでヒト最大用量である400mg/日の0.12倍以上の投与量【体表面積換算（mg/m ² ）に基づく】において母動物の一般状態の悪化に関連した胎児体重の低値、着床後胚・胎児死亡率及び死産児数の増加、胎児骨格変異の発現頻度増加、出生児における神経行動学的異常、出生児回収率（哺育中の巣から出生児を離し、5分以内に母動物が巣内に出生児を連れ戻す）の低下又は出生後の生存率低下がみられた。	着床後胚・胎児死亡率	着床後胚・胎児死亡	**着床後胚死亡率**、**胎児死亡率**	「着床後胚・胎児死亡率」は「着床後胚」と「胎児」が共通の被修飾語「死亡率」を持つ合成された表現です。各々「着床後胚死亡率」「胎児死亡率」に分解されます。これらは複数の修飾・被修飾関係が1語に集約されているため合成された表現に該当します。なお、入力の辞書(words)には「胎児死亡」は含まれますが、分解後の語形である「着床後胚死亡率」「胎児死亡率」は含まれていません。	TRUE	FALSE	FN	抽出範囲が広い（FN）	末尾に率を含むため、範囲が適切でない
アピキサバン	これらの薬剤又はセイウオトギリソウがCYP3A4及びP-糖蛋白を同時に強力に誘導するため、本剤の代謝及び排出が促進されると考えられる。	本剤の代謝及び排出	代謝及び排出が促進	**本剤の代謝**、**本剤の排出**	「本剤の代謝及び排出」は共通の修飾語「本剤」が「代謝」と「排出」にかかっている表現であり、2つの用語に分解されます。分解すると「本剤の代謝」「本剤の排出」となります。	TRUE	FALSE	FN	有害事象表現チェックがFALSE	代謝・排出阻害の抽出失敗
ラモトリギン	本剤の投与による発疹は斑状・丘疹状にあらわれることが多く、重篤な皮膚障害の発現率は、本剤投与開始から8週間以内に高く、また、バルプロ酸ナトリウムと併用した場合、あるいは小児において高いことが示されているので、本剤の投与にあたっては十分に注意し、異常が認められた場合には、投与を中止し適切な処置を行うこと。[1.3 参照],[7.1 参照],[9.7.1 参照],[11.1.1 参照],[17.3.1 参照],[17.3.2 参照]	斑状・丘疹状	発疹は斑状・丘疹状	**斑状発疹**、**丘疹状発疹**	「斑状・丘疹状」は「斑状」と「丘疹状」という2つの修飾語が共通の被修飾語（文中では「発疹」）にかかっている合成された表現です。それぞれ「斑状発疹」、「丘疹状発疹」に分けられます。	TRUE	FALSE	FN	有害事象表現チェックがFALSE	斑状丘疹状発疹 がT辞書にある
リバビリン	妊婦、妊娠している可能性のある女性又は授乳中の女性【動物実験で催奇形性作用及び胚・胎児致死作用が報告されている。】[1.1 参照],[9.5 参照],[9.6 参照]	胚・胎児致死作用	胚・胎児致死作用	**胚致死作用**、**胎児致死作用**	胚・胎児致死作用は「胚」と「胎児」が「・」で接続され、共通の被修飾語「致死作用」をもつ2つの用語の合成された表現です。それぞれ胚致死作用、胎児致死作用に分けられます。なお、分解後の用語はいずれも入力の辞書（words）には含まれていません。	TRUE	FALSE	FN	有害事象表現チェックがFALSE	「胚致死作用」「胎児致死作用」とともにT辞書にない
リバビリン	妊婦又は妊娠している可能性のある女性には投与しないこと。動物実験で催奇形性作用（ラット及びウサギ：1mg/kg/日）及び胚・胎児致死作用（ラット：10mg/kg/日）が認められている。[1.1 参照],[2.1 参照],[9.4.1 参照],[9.4.2 参照]	胚・胎児致死作用	胚・胎児致死作用	**胚致死作用**、**胎児致死作用**	「胚・胎児致死作用」は「胚」と「胎児」が「・」で接続され、共通の被修飾語「致死作用」をもつ2つの用語の合成された表現です。それぞれ「胚致死作用」「胎児致死作用」に分けられます。	TRUE	FALSE	FN	有害事象表現チェックがFALSE	「胚致死作用」「胎児致死作用」とともにT辞書にない
リツキシマブ（遺伝子組換え）製剤	細菌、真菌、あるいはウイルスによる重篤な感染症（敗血症、肺炎等）があらわれることがあるので、本剤の治療期間中及び治療終了後は患者の状態を十分に観察すること。[8.6 参照],[9.1.4 参照]	重篤な感染症（敗血症、肺炎等）	細菌、真菌、あるいはウイルスによる重篤な感染症	**重篤な感染症**、敗血症、肺炎	「重篤な感染症（敗血症、肺炎等）」は主語である「重篤な感染症」と、その具体例を括弧内で列挙している表現であり、括弧内は具体例の列挙にあたるため合成された表現ではありません。抽出可能な語としては主表現の「重篤な感染症」と、括弧内の具体例である「敗血症」「肺炎」が該当	FALSE	TRUE	FNFP	有害事象表現チェックがFALSE	抽出すべき範囲が誤っている

2. 生成 AI を用いた段階的症例報告の因果関係評価の検討

A. 研究目的

医療現場から規制当局へ提出される副作用報告 (Individual Case Safety Report: ICSR) は、因果関係評価が確定する前段階から情報収集が開始され、追加情報の取得に応じて逐次更新される。企業及び医療機関は、規制要件に基づき迅速な報告を行う必要があるため、症例情報が十分に整っていない段階であっても第 1 報として報告を行い、その後、情報更新の有無にかかわらず、報告日から規定された期間内に第 2 報、第 3 報を経て最終報告に至る運用がなされている。

一方、規制当局においては、特に第 1 報の段階では症例情報の量や質が限定的である場合が多いにもかかわらず、安全性確保の観点から、入手可能な情報に基づき副作用の評価及びリスク判断が実施されている。このように、副作用因果関係評価は、十分な

情報が揃った症例のみを前提として行われるものではなく、むしろ不確実性を伴う情報環境下での迅速な判断が求められることが多い。

しかしながら、実務上、因果関係評価に必要とされる情報が充足された症例は限定的であり、報告初期段階においては、評価の前提となる情報が断片的であることが一般的である。にもかかわらず、リスク評価及びリスク最小化策を適時に講じるためには、情報の不完全性を許容した上で、暫定的な因果関係判断を行うことが不可避である。

以上の背景を踏まえ、副作用報告において時系列的に蓄積される症例情報の量及び内容に応じて、生成 AI を用いた因果関係評価が実施可能であるかを探索的に検討し、副作用因果関係評価に生成 AI を活用する際の課題を抽出することを目的とした。

なお、プロンプト設計は株式会社 CMC エクスメディカに外部委託した。

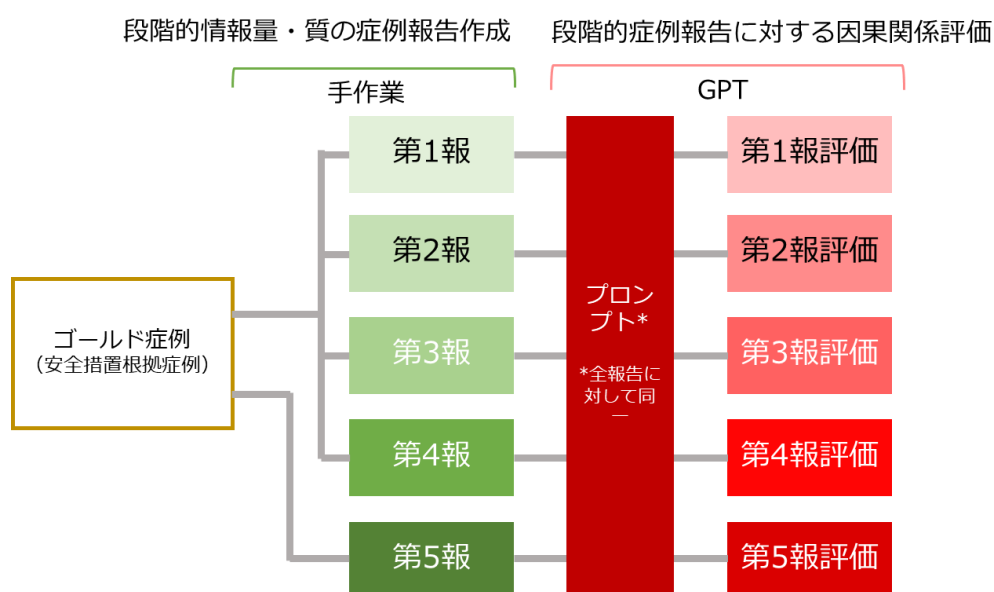


図 1 本研究の全体図

B. 研究方法

本研究の全体図を図1に示す。

段階的な情報量・質で構成される症例報告を作成するため、医薬品安全性情報において因果関係評価の根拠として示された症例情報をゴールド症例として用いた。具体的には、医薬品・医療機器等安全性情報 No.421 2025 年 7 月*「2.重要な副作用等に関する情報」においてチアマゾールによる急性膵炎に対する安全対策措置の根拠症例として掲載された2症例の症例概要のうち図2に示す症例概要を用いた。

- *<https://www.pmda.go.jp/files/000276332.pdf>
- 症例概要から段階的に情報量・質を有する症例報告を作成するにあたり、次に示す因果関係評価上、重要な項目を時系列に沿って段階的に除外し、第1報から第4報までを手作業で作成した。第5報は元の症例概要を用いた。
- 時間関係（投与開始日・中止日・発現日

- 中止/再投与に対する症状変化（中止後改善、再投与後再発）
- 患者背景（原疾患、既往歴、背景疾患、併用薬）
- アウトカム（転帰、重篤性、検査値、画像初見等の客観的情報）
- 用量・曝露情報（投与量、投与経路、服薬状況、血中濃度等）

段階的症例報告に対する因果関係評価には大規模言語モデル(LLM)として ChatGPT (GPT-5 系モデル) (以下、GPT) を使用し、また Thinking5.5 及び Instant3.5 の2種のモードを適用した。

段階的症例報告に対する因果関係評価を GPT に実施させるため、第1報から第5報までの報告では、時間経過に伴い情報量が増えていること、及び各報告の有害事象と被疑薬の因果関係を Naranjo スコアに基づいて評価することを指示するプロンプトを設計した。また、急性膵炎に関する注意喚起

No.	患者		1日投与量 投与期間	副作用					
	性・ 年齢	使用理由 (合併症)		経過及び処置					
1	女 70代	甲状腺中毒症、バセドウ病(高血圧症、シェーグレン症候群、骨粗鬆症、心房細動、心不全)	15mg 16日間	急性膵炎	以前より、体重減少、動悸あり。 動悸、疲労感のため受診。心不全、心房細動が判明。バセドウ病の診断にて入院。 本剤及びヨウ化カリウム投与開始。				
				投与12日目	38.3℃の発熱と急性腹痛、下痢が発現。 CRPがやや高値を示した以外は特記すべき所見なし。腹部単純CTにて憩室炎が疑われ、セフメゾール投与するも改善せず。				
				投与16日目 (投与中止日)	膵酵素が上昇し、急性膵炎と診断。 本剤投与中止。ヨウ化カリウム増量。 補液、ウリナスタチン、イミベネム水和物・シラスタチンナトリウムを投与開始。				
				投与中止8日後	腹痛、発熱は改善傾向。イミベネム水和物・シラスタチンナトリウムを中止し、ウリナスタチンをカモスタットメシル酸塩に変更。				
				投与中止16日後	甲状腺ホルモン上昇のため、プロピルチオウラシルを投与開始し、ヨウ化カリウムを減量。その後、増悪なし。 全身状態改善したため退院。				
臨床検査値									
検査項目 (単位)			投与 開始日	投与 12日目	投与16日目 (投与中止日)	投与中止 1日後	投与中止 8日後	投与中止 15日後	投与中止 25日後
アミラーゼ (IU/L)			56	76	214	336	134	195	120
リパーゼ (IU/L)			—	—	923	1519	269	—	214
エラスターゼ 1 (ng/dL)			—	—	1537	3133	1703	—	1050
CRP (mg/dL)			—	0.98	9.97	13.06	0.61	0.05	0.06
併用薬：メトプロロール酒石酸塩、オルメサルタン メドキシミル、ニフェジピン、リセドロン酸ナトリウム水和物、エソメプラゾールマグネシウム水和物、モメタゾンフランカルボン酸エステル水和物、ビスプロロールフマル酸塩、トラセミド、アピキサパン、トルバプタン、ジゴキシン、ヨウ化カリウム									

図2 症例概要

が追加される前の評価状況を模擬するため、添付文書情報から急性膵炎に関する記載を除外した資料を用いた。

C. 結果・考察

症例概要から作成した第1報から第4報を別紙3に示す。第5報は症例概要とした。第1報から第5報をテキスト化し、図3に示すプロンプトを設計した。なお、検討時点では、急性膵炎の安全措置が施行された後の情報が、当該安全対策措置検討時点から更新されている可能性がある。そのため、出力形式1で指示している「1.過去の報告」については、当時の情報に基づく評価と異なる可能性があることに留意が必要である。この点は、生成AIを用いて過去時点の因果関係評価を再現・検証する際の課題であると考えられた。

```
#役割
あなたは、臨床現場で発現した有害事象と、その被疑薬の因果
関係を評価する担当者です。

#出力いただきたいこと
以下は、時系列ごとの有害事象報告です。
第1～第5報まで、時間経過に伴い、情報量が増えていきます。
第1～第5報それぞれの、有害事象と被疑薬の因果関係を
Naranjoスコアに基づいて評価してください。
また、その評価結果の根拠を出力してください。

#出力形式1
- スコア
- 評価:Definite/Probable/Possible/Doubtful
- 内訳と根拠:以下の表で出力してください。

| 項目 | 評価 | 点数 | 根拠 |
| ---- | ---- | ---- | ---- |
| 1. 過去の報告 | | | |
| 2. 時系列 | | | |
| 3. 中止時の反応 | | | |
| 4. 再投与時の反応 | | | |
| 5. 代替原因 | | | |
| 6. プラセボ反応 | | | |
| 7. 血中濃度 | | | |
| 8. 用量反応 | | | |
| 9. 過去の類似経験 | | | |
| 10. 客観的証拠 | | | |

#出力形式2
- 総合評価として、第1～5報までのスコアと判定を時系列の表
にまとめてください。

#参考情報
被疑薬の添付文書情報XMLを添付いたします。
```

図3 プロンプト

Naranjo 判定基準による分類は、9 点以上を Definite、5～8 点を Probable、1～4 点を Possible、0 点以下を Doubtful とした。

本プロンプトを用いて、第1報から第5報までの段階的症例報告に対する因果関係評価を実施した結果を別紙4（モード Thinking5.5）及び別紙5（モード Instant3.5）に示す。Thinking5.5では、Instant3.5に比べて根拠説明が詳細であり、Naranjo 評価も保守的であった。また、時系列に追加される情報よりも、代替原因の有無やその影響を重視する傾向が認められた。一方、Instant3.5では、第1報から第5報にかけて総評価の点数が増加しており、一見すると追加情報に応じた段階的判断が行われているように見えた。しかし、出力された総評価の点数と項目別点数の合計点との間に不一致が認められたため、評価結果をそのまま採用するのではなく、点数計算の整合性を確認する必要があると考えられた。以上より、生成AIを用いた因果関係評価では、使用するモードによって、根拠説明の詳細さ、評価の保守性、代替原因の扱い、点数計算の整合性に差異が生じることが示された。の特性も考慮すべき課題であることがわかった。したがって、生成AIを副作用因果関係評価に活用する際には、モード特性を考慮するとともに、出力結果の妥当性を確認するため、評価点だけでなく、その根拠及び項目別評価結果を併せて出力させる必要がある。

D. 結論

本研究では、副作用報告（ICSR）において時系列的に蓄積される症例情報の量及び内容に応じて、因果関係評価を段階的に更新するアルゴリズムの枠組みを検討した。

具体的には、チアマゾールによる急性膵炎の症例概要をもとに、報告初期段階から最終報告に至るまでの情報充足過程を模擬した段階的症例報告を作成し、生成 AI を用いて各段階における Naranjo スコアに準じた因果関係評価を実施した。その結果、症例情報の追加に伴い、因果関係評価が変化し得ることは示された。一方で、その変化の方向性や妥当性は、使用するモード、プロンプト設計、代替原因の扱い、点数計算の整合性に影響されることが分かった。特に Thinking5.5 では根拠説明が詳細で、代替原因を重視した保守的な評価を行う傾向が認められたのに対し、Instant3.5 では段階的な合計点の増加が認められたものの、出力された総合点と項目別点数の合計点との間に不一致が認められた。

このことから、生成 AI を用いた因果関係評価では、出力された評価結果をそのまま採用するのではなく、項目別点数、合計点、総合評価、及び判断根拠の整合性を確認することが不可欠であると考えられた。また、各段階における評価結果とその根拠を可視化することで、評価の透明性及び説明可能性の向上に資する可能性が示された。特に、報告初期段階のように情報が限定的な状況においても、一定の前提条件のもとで暫定的な因果関係評価を行い、その後の情報更新に応じて評価を見直すという枠組みは、実務における ICSR 評価プロセスと整合的であり、リスク評価及び安全対策の迅速化に寄与する可能性がある。

一方で、本研究により、生成 AI を用いた因果関係評価支援の実運用に向けた課題も明らかとなった。第一に、症例情報には個人情報や機微情報が含まれる可能性があるこ

とから、データの取得、利用、処理にあたっては、個人情報保護及びデータガバナンスへの十分な配慮が不可欠である。特に、匿名化・仮名化処理の適切な実施や、外部システム利用時の情報管理体制、入力データの管理範囲を明確化する必要がある。第二に、生成 AI を含むシステムの品質確保の観点から、入力データの適切性、プロンプト及び処理ロジックの妥当性、出力結果の正確性を担保する仕組みが必要である。特に、生成 AI は出力のばらつき、計算不整合、誤生成のリスクを内在するため、評価結果のレビュー、再計算、根拠確認、品質管理プロセスの構築が重要である。第三に、システムの信頼性及び責任所在の明確化が必要である。因果関係評価は規制上重要な判断に関わるため、AI の出力をそのまま意思決定に用いるのではなく、人による確認及び最終判断を前提とした運用設計が求められる。また、システム利用における責任範囲について、開発者、運用者、評価者、最終判断者の役割を明確化する必要がある。第四に、バリデーションの実施が不可欠である。従来のコンピュータ化システムバリデーション (CSV) の考え方に加え、生成 AI 特有の非決定性やプロンプト依存性、モデル更新による出力変動を踏まえた評価枠組みを構築し、用途に応じた性能評価、再現性確認、リスクベースでの適用範囲の明確化が求められる。また、生成 AI を用いて過去時点の因果関係評価を再現・検証する場合、評価時点以降に蓄積・更新された情報が出力に影響する可能性がある。特に「過去の報告」や既知・未知判断では、安全対策措置検討時点に実際に利用可能であった情報と、生成 AI が参照・反映し得る現在の情報との間に差異が生じ

る可能性があるため、評価時点及び参照情報源を明確化、記録し、時間的情報バイアスを制御することが必要である。

以上を踏まえ、過去の安全性評価症例、規制当局による評価事例、添付文書・ラベル情報等を知識基盤として活用し、過去の人による評価結果を Retrieval-Augmented Generation (RAG) として蓄積し、知識基盤と組み合わせることで、精度の高い因果関係評価支援モデルの構築が期待される。その際には、個人情報保護、品質確保、信頼性担保及びバリデーションに関する要件を体系的に整理し、ガイダンスとして明確化することが重要であると考えられる。

E. 研究発表

1. 論文発表

該当なし

2. 学会発表

該当なし

別紙3 症例概要から作成した第1報から第4報

第1報

第1報は、未知・重篤事象が発現し、15日以内の規制当局報告を実施するにあたり、「投薬開始時」「SAE発現時」のみの情報とした。

☐ チアマゾールとする

[症例概要]		副作用	
No.	患者 性・年齢 使用理由 (合併症)	1日投与量 投与期間	経過及び処置
1	女 70代 甲状腺中毒症、パセドウ病、 [不明]	15mg 16日間	急性肺炎 投与開始日 以前より、体重減少、動悸あり。動悸、疲労感のため受診。心不全、心房細動が判明。パセドウ病の診断にて入院。 [本剤及びヨウ化カリウム投与開始。] [不明] 投与16日目 解熱薬が上昇し、急性肺炎と診断。 [投与中止日] 本剤投与中止。ヨウ化カリウム増量。 [不明]
臨床検査値			
[不明]			
併用薬：メトプロロール酒石酸塩、オルメサルタン・メドキシミル、ニフェジピン、リセドロン酸ナトリウム水和物、エソメプラゾールマグネシウム水和物、モメタゾンフランカルボン酸エステル水和物、ビソプロロール fumarate、トラセミド、アジキサン、トルバプタン、ジゴキシン、ヨウ化カリウム			

第2報

第2報は、以下を追加
☒ 「投薬開始時」と「SAE発現時」間の情報
☒ 「SAE発現時」の対応内容
☒ 「SAE発現時」までの臨床検査値
☒ 合併症情報
☒ 併用薬

[症例概要]		副作用																					
No.	患者 性・年齢 使用理由 (合併症)	1日投与量 投与期間	経過及び処置																				
1	女 70代 甲状腺中毒症、パセドウ病、 [不明]	15mg 16日間	急性肺炎 投与開始日 以前より、体重減少、動悸あり。動悸、疲労感のため受診。心不全、心房細動が判明。パセドウ病の診断にて入院。 [本剤及びヨウ化カリウム投与開始。] 投与12日目 38.3℃の発熱と急性腹痛、下痢が発現。CRPがやや高値を示した以外は特記すべき所見なし。腹部単純CTにて憩室炎が疑われ、セフメタゾール投与するも改善せず。 投与16日目 解熱薬が上昇し、急性肺炎と診断。 [投与中止日] 本剤投与中止。ヨウ化カリウム増量。 補液、ウリナスタチン、イミベネム水和物・シラスタンナトリウムを投与開始。 [不明]																				
臨床検査値																							
<table border="1"> <thead> <tr> <th>検査項目 (単位)</th> <th>投与開始日</th> <th>投与12日目</th> <th>投与16日目 (投与中止日)</th> </tr> </thead> <tbody> <tr> <td>アミラーゼ (IU/L)</td> <td>56</td> <td>76</td> <td>214</td> </tr> <tr> <td>リパーゼ (IU/L)</td> <td>-</td> <td>-</td> <td>923</td> </tr> <tr> <td>エラスターゼ1 (ng/dL)</td> <td>-</td> <td>-</td> <td>1537</td> </tr> <tr> <td>CRP (mg/dL)</td> <td>-</td> <td>0.98</td> <td>9.97</td> </tr> </tbody> </table>				検査項目 (単位)	投与開始日	投与12日目	投与16日目 (投与中止日)	アミラーゼ (IU/L)	56	76	214	リパーゼ (IU/L)	-	-	923	エラスターゼ1 (ng/dL)	-	-	1537	CRP (mg/dL)	-	0.98	9.97
検査項目 (単位)	投与開始日	投与12日目	投与16日目 (投与中止日)																				
アミラーゼ (IU/L)	56	76	214																				
リパーゼ (IU/L)	-	-	923																				
エラスターゼ1 (ng/dL)	-	-	1537																				
CRP (mg/dL)	-	0.98	9.97																				
併用薬：メトプロロール酒石酸塩、オルメサルタン・メドキシミル、ニフェジピン、リセドロン酸ナトリウム水和物、エソメプラゾールマグネシウム水和物、モメタゾンフランカルボン酸エステル水和物、ビソプロロール fumarate、トラセミド、アジキサン、トルバプタン、ジゴキシン、ヨウ化カリウム																							

第3報

第3報は、投与中止後の15日以内の情報を追加

☐ チアマゾールとする

☐ 追加

[症例概要]		副作用																															
No.	患者 性・年齢 使用理由 (合併症)	1日投与量 投与期間	経過及び処置																														
1	女 70代 甲状腺中毒症、パセドウ病、 [不明]	15mg 16日間	急性肺炎 投与開始日 以前より、体重減少、動悸あり。動悸、疲労感のため受診。心不全、心房細動が判明。パセドウ病の診断にて入院。 [本剤及びヨウ化カリウム投与開始。] 投与12日目 38.3℃の発熱と急性腹痛、下痢が発現。CRPがやや高値を示した以外は特記すべき所見なし。腹部単純CTにて憩室炎が疑われ、セフメタゾール投与するも改善せず。 投与16日目 解熱薬が上昇し、急性肺炎と診断。 [投与中止日] 本剤投与中止。ヨウ化カリウム増量。 補液、ウリナスタチン、イミベネム水和物・シラスタンナトリウムを投与開始。 投与中止8日後 腹痛、発熱は改善傾向。イミベネム水和物・シラスタンナトリウムを中止し、ウリナスタチンをカモスタットメシル酸塩に変更。 [投与中止16日後] 甲状腺ホルモン上昇のため、プロピルチオウラシルを投与開始し、ヨウ化カリウムを減量。その後、増量なし。 [不明]																														
臨床検査値																																	
<table border="1"> <thead> <tr> <th>検査項目 (単位)</th> <th>投与開始日</th> <th>投与12日目</th> <th>投与16日目 (投与中止日)</th> <th>投与中止1日後</th> <th>投与中止8日後</th> </tr> </thead> <tbody> <tr> <td>アミラーゼ (IU/L)</td> <td>56</td> <td>76</td> <td>214</td> <td>336</td> <td>134</td> </tr> <tr> <td>リパーゼ (IU/L)</td> <td>-</td> <td>-</td> <td>923</td> <td>1519</td> <td>269</td> </tr> <tr> <td>エラスターゼ1 (ng/dL)</td> <td>-</td> <td>-</td> <td>1537</td> <td>3133</td> <td>1703</td> </tr> <tr> <td>CRP (mg/dL)</td> <td>-</td> <td>0.98</td> <td>9.97</td> <td>13.06</td> <td>0.61</td> </tr> </tbody> </table>				検査項目 (単位)	投与開始日	投与12日目	投与16日目 (投与中止日)	投与中止1日後	投与中止8日後	アミラーゼ (IU/L)	56	76	214	336	134	リパーゼ (IU/L)	-	-	923	1519	269	エラスターゼ1 (ng/dL)	-	-	1537	3133	1703	CRP (mg/dL)	-	0.98	9.97	13.06	0.61
検査項目 (単位)	投与開始日	投与12日目	投与16日目 (投与中止日)	投与中止1日後	投与中止8日後																												
アミラーゼ (IU/L)	56	76	214	336	134																												
リパーゼ (IU/L)	-	-	923	1519	269																												
エラスターゼ1 (ng/dL)	-	-	1537	3133	1703																												
CRP (mg/dL)	-	0.98	9.97	13.06	0.61																												
併用薬：メトプロロール酒石酸塩、オルメサルタン・メドキシミル、ニフェジピン、リセドロン酸ナトリウム水和物、エソメプラゾールマグネシウム水和物、モメタゾンフランカルボン酸エステル水和物、ビソプロロール fumarate、トラセミド、アジキサン、トルバプタン、ジゴキシン、ヨウ化カリウム																																	

第4報

第4報は、投与中止後の15日×2以内の情報を追加

☐ チアマゾールとする

☐ 追加

[症例概要]		副作用																																				
No.	患者 性・年齢 使用理由 (合併症)	1日投与量 投与期間	経過及び処置																																			
1	女 70代 甲状腺中毒症、パセドウ病、 [不明]	15mg 16日間	急性肺炎 投与開始日 以前より、体重減少、動悸あり。動悸、疲労感のため受診。心不全、心房細動が判明。パセドウ病の診断にて入院。 [本剤及びヨウ化カリウム投与開始。] 投与12日目 38.3℃の発熱と急性腹痛、下痢が発現。CRPがやや高値を示した以外は特記すべき所見なし。腹部単純CTにて憩室炎が疑われ、セフメタゾール投与するも改善せず。 投与16日目 解熱薬が上昇し、急性肺炎と診断。 [投与中止日] 本剤投与中止。ヨウ化カリウム増量。 補液、ウリナスタチン、イミベネム水和物・シラスタンナトリウムを投与開始。 投与中止8日後 腹痛、発熱は改善傾向。イミベネム水和物・シラスタンナトリウムを中止し、ウリナスタチンをカモスタットメシル酸塩に変更。 [投与中止16日後] 甲状腺ホルモン上昇のため、プロピルチオウラシルを投与開始し、ヨウ化カリウムを減量。その後、増量なし。 [不明]																																			
臨床検査値																																						
<table border="1"> <thead> <tr> <th>検査項目 (単位)</th> <th>投与開始日</th> <th>投与12日目</th> <th>投与16日目 (投与中止日)</th> <th>投与中止1日後</th> <th>投与中止8日後</th> <th>投与中止15日後</th> </tr> </thead> <tbody> <tr> <td>アミラーゼ (IU/L)</td> <td>56</td> <td>76</td> <td>214</td> <td>336</td> <td>134</td> <td>195</td> </tr> <tr> <td>リパーゼ (IU/L)</td> <td>-</td> <td>-</td> <td>923</td> <td>1519</td> <td>269</td> <td>-</td> </tr> <tr> <td>エラスターゼ1 (ng/dL)</td> <td>-</td> <td>-</td> <td>1537</td> <td>3133</td> <td>1703</td> <td>-</td> </tr> <tr> <td>CRP (mg/dL)</td> <td>-</td> <td>0.98</td> <td>9.97</td> <td>13.06</td> <td>0.61</td> <td>0.05</td> </tr> </tbody> </table>				検査項目 (単位)	投与開始日	投与12日目	投与16日目 (投与中止日)	投与中止1日後	投与中止8日後	投与中止15日後	アミラーゼ (IU/L)	56	76	214	336	134	195	リパーゼ (IU/L)	-	-	923	1519	269	-	エラスターゼ1 (ng/dL)	-	-	1537	3133	1703	-	CRP (mg/dL)	-	0.98	9.97	13.06	0.61	0.05
検査項目 (単位)	投与開始日	投与12日目	投与16日目 (投与中止日)	投与中止1日後	投与中止8日後	投与中止15日後																																
アミラーゼ (IU/L)	56	76	214	336	134	195																																
リパーゼ (IU/L)	-	-	923	1519	269	-																																
エラスターゼ1 (ng/dL)	-	-	1537	3133	1703	-																																
CRP (mg/dL)	-	0.98	9.97	13.06	0.61	0.05																																
併用薬：メトプロロール酒石酸塩、オルメサルタン・メドキシミル、ニフェジピン、リセドロン酸ナトリウム水和物、エソメプラゾールマグネシウム水和物、モメタゾンフランカルボン酸エステル水和物、ビソプロロール fumarate、トラセミド、アジキサン、トルバプタン、ジゴキシン、ヨウ化カリウム																																						

別紙4 段階的症例報告に対する GPT (Thinking5.5 使用) 副作用因果関係評価

項目	第1報		第2報		第3報		第4報		第5報	
	点数	根拠	点数	根拠	点数	根拠	点数	根拠	点数	根拠
1. 過去の報告	1	チアマゾール/メチマゾール誘発急性膵炎の既報あり。(PubMed)	1	既報症例あり。(PubMed)	1	既報症例あり。(PubMed)	1	既報症例あり。(PubMed)	1	チアマゾール/メチマゾール誘発急性膵炎の症例報告あり。(PubMed)
2. 時系列	2	2020/09/23にチアマゾール開始、2020/10/08に急性膵炎診断。投与後発現として時間的関連あり。	2	投与開始後約15日で急性膵炎診断。	2	チアマゾール開始後に腹痛・膵酵素上昇・急性膵炎診断。	2	投与後に急性膵炎が発現。	2	2020/09/23投与開始後、2020/10/08に急性膵炎発現。
3. 中止時の反応	0	第1報時点では被疑薬中止後の改善情報なし。	0	2020/10/08にチアマゾール中止されたが、第2報時点では改善経過がまだ示されていない。	1	2020/10/08中止後、2020/10/16に腹痛・発熱は改善傾向、CRPも13.06から0.61へ低下。	1	中止後に全身状態が改善し、2020/10/24に退院。CRPも0.05まで低下。	1	チアマゾール中止後、症状・炎症反応が改善し、最終的に転帰は「回復」。
4. 再投与時の反応	0	再投与情報なし。	0	再投与なし。	0	再投与なし。	0	再投与なし。	0	再投与なし。
5. 代替原因	0	併用薬・他疾患・画像所見などの情報が不足。	-1	多数の併用薬、ヨウ化カリウム併用、腹部CTで憩室炎疑いなど、他原因の可能性が残る。	-1	併用薬多数、憩室炎疑い、感染・炎症の関与などが完全には除外されていない。	-1	併用薬・合併症・憩室炎疑いがあり、他原因を否定しきれない。	-1	多数の併用薬、基礎疾患、憩室炎疑いなどがあり、代替原因の完全除外には不足。
6. プラセボ反応	0	プラセボ投与なし。	0	プラセボ情報なし。	0	プラセボ情報なし。	0	プラセボ情報なし。	0	プラセボ情報なし。
7. 血中濃度	0	チアマゾール血中濃度情報なし。	0	血中濃度情報なし。	0	チアマゾール血中濃度なし。	0	血中濃度情報なし。	0	チアマゾール血中濃度なし。
8. 用量反応	0	増量・減量に伴う反応情報なし。	0	用量変更に伴う反応なし。	0	用量反応情報なし。	0	用量反応情報なし。	0	増量・減量による反応情報なし。
9. 過去の類似経験	0	本人の過去のチアマゾール曝露・膵炎歴なし。	0	過去の類似反応情報なし。	0	本人の過去の類似反応情報なし。	0	過去の同様反応なし／不明。	0	本人の過去の類似反応情報なし。
10. 客観的証拠	1	「膵酵素が上昇し、急性膵炎と診断」と記載あり。	1	アミラーゼ214、リパーゼ923、エラスターゼ1 1537、CRP 9.97と客観的検査異常あり。	1	膵酵素上昇後、アミラーゼ・リパーゼ・CRPの推移が示されている。	1	膵酵素、CRP、臨床経過により客観的な裏付けあり。	1	アミラーゼ、リパーゼ、エラスターゼ1、CRPの経時推移があり、2020/11/02にはアミラーゼ120、リパーゼ214、CRP 0.06まで改善。
総評価	4	Possible	3	Possible	4	Possible	4	Possible	4	Possible
主な変化	投与開始後に発現し、膵酵素上昇による診断あり。ただし中止後経過・併用薬・代替原因情報が不足		併用薬・合併症・憩室炎疑いが判明し、代替原因ありとして減点。中止直後で改善確認はまだ不可		中止後に腹痛・発熱・CRP・膵酵素が改善傾向となり、dechallenge陽性として加点		全身状態改善・退院により改善経過は補強。ただし再投与・代替原因除外なし		転帰「回復」と追加検査値改善で客観性は補強。ただしNaranjo上は追加加点点目が限定的	

別紙 5 段階的症例報告に対する GPT (Instant3.5 使用) 副作用因果関係評価

項目	第1報		第2報		第3報		第4報		第5報	
	点数	根拠	点数	根拠	点数	根拠	点数	根拠	点数	根拠
1. 過去の報告	0	添付文書に急性膵炎の記載なし	0	前報と同じ	0		0		0	
2. 時系列	2	投与開始後約2週間で発現	2	投与→症状発現	2		2		2	
3. 中止時の反応	0	中止情報なし	0	改善確認なし	1	中止後改善傾向	1	症状改善・退院	1	回復確認
4. 再投与時の反応	0	情報なし	0	情報なし	0		0		0	
5. 代替原因	-1	甲状腺中毒症・心不全など	-1	憩室炎疑い・多剤併用	-1	併用薬・感染	-1		-1	
6. プラセボ反応	0	情報なし	0		0		0		0	
7. 血中濃度	0	測定なし	0		0		0		0	
8. 用量反応	0	情報なし	0		0		0		0	
9. 過去の類似経験	0	情報なし	0		0		0		0	
10. 客観的証拠	1	膵酵素上昇	1	アミラーゼ・リパーゼ上昇	1	検査値推移	1	数値改善	1	酵素正常化
総評価	3	Possible	4	Possible	6	Probable	7	Probable	7	Probable
主な変化	記載なし		客観データ追加で信頼性上昇 ただし代替原因が強く残る		デチャレンジ陽性（改善） 因果関係が一段強化		臨床経過がより明確（回復） 因果関係の一貫性が強化		回復確定により評価安定 これ以上の加点要素（再投与など） がないため頭打ち	