

臨床試験の質の管理に対するAI適用の予備的検討
～ AIも仲間にしたQuality Cultureの醸成を目指して ～

日本製薬工業協会
データサイエンス部会
2024年度タスクフォース 1・2 「AIを活用したデータマネジメントの変革」

2025年12月

目次

1. はじめに.....	4
2. AI 活用の検討対象の候補の抽出および選定	5
3. 選定した AI 活用の検討を実現するための試験計画およびデータの入手	6
4. CTQ 要因の特定支援.....	7
4.1. CTQ 要因特定の意義と AI 活用の可能性	7
4.2. 目的	8
4.3. 予備的試行（AI に CTQ 要因を特定させるワークフローの検討）	8
4.4. 方法	8
4.5. CTQ 要因素案等の生成結果.....	13
4.5.1. 生成結果 1	13
4.5.2. 生成結果 2	21
4.6. 考察	26
4.6.1. 全体的な所感	26
4.6.2. 対話型 AI とシステムプロンプトの有用性について	27
4.6.3. AI をより有効活用するための留意事項	27
4.6.4. 実業務における AI の活用方法の検討.....	29
5. 試験実施計画に基づいた CRITICAL な疑義事項の検出	30
5.1. RBA を適用したデータレビュー	30
5.2. 目的	32
5.3. 予備的試行（テストデータでの検討）	32
5.4. 方法	34
5.4.1. レビューアーのロール設定	34
5.4.2. 臨床試験データでの検討	35
5.5. 疑義事項の検出の結果	38
5.5.1. 疑義事項の検出結果	38
5.5.2. 重要度分類の判定結果.....	48
5.6. 考察	54
5.6.1. 人間のインプットの重要性	54
5.6.2. Human-in-the-Loop の重要性	55
5.6.3. AI によるデータレビューと今後の展望	57
6. まとめ	57
7. おわりに.....	58

参考文献	60
APPENDIX 1: プロンプトの作成	63
APPENDIX 2: 余禄（AI による CTQ 要因特定会議のシミュレーション）	64

1. はじめに

様々な業種で AI が本格的に活用されてきていることは明らかである。例えば、製造業における自動運転技術とモビリティ AI 基盤の開発 [1]、小売業における需要予測と発注業務の省力化 [2]、金融業における経営改善支援必要取引先候補の特定 [3]などがその一例であるが、その他多くの分野で AI の利用が確実に進んでいる [4]。これらの事例では、AI を用いた革新によって大規模な業務効率の改善ならびに新たなビジネス価値の創出を実現しようとしていることがわかる。

医薬品開発におけるデータマネジメント領域でも日本製薬工業協会（以下「製薬協」）のデータサイエンス部会（以下「DS 部会」）が公表している資料 [5]で紹介されているように、「データレビューならびにクエリ文作成」、「WHODrug コーディング」等の AI を用いたプロセスが実務に採用されている。また、複数のベンダーが AI を活用した臨床試験の支援サポートサービスを既に提供しているか今後提供していく予定があることも知られている [6] [7]。しかし、データマネジメント領域では、AI を用いた革新による大規模な業務効率の改善ならびに新たなビジネス価値の創出の未来像は具体化されていないように思われる。ベンダーが AI を活用した臨床試験の支援サポートサービスを開発してくれるとしても、実務でデータマネジメントに携わっている者が、現状の非効率や非合理に対する問題意識を強く持ち、それに対する解決策を自ら考えて提案していかなければ、真に有用な革新はベンダーによって実用化され難い。

製薬協の DS 部会の 2023 年度成果物「データマネジメント業務における AI 活用検討のためのモデル試作」 [8]にかかる活動においては、データマネジメント領域における身近な業務の中から AI の活用によって効率化やミスの削減が期待されるプロセスをいくつか抽出し、それらのプロセスに適用可能な AI モデルの試作が行われた。そして、それらの AI モデルの有用性や実際の業務で利用する上での課題や留意すべき事項について検討した結果が成果物としてまとめられている。つまり、この活動においては当初から「従来データマネージャーがマニュアル作業で行ってきたことを AI にやらせてみるとどうなるか？」ということに主眼をおいて検討が進められてきた。

そこで、2024 年度の活動としては、日々発展が目覚ましい様々な AI 技術を活用して、データマネジメント領域においてよりビジネスインパクトの大きな変化を起こさせる可能性を探索することを目標に掲げてみた。注意した点は、「AI に何ができるか」を先に考えるのではなく、どのようなプロセスや業務に変更が必要かをまず先に考えようとしたことである。その上で、利用可能と思われる具体的な AI 技術/モデルに関する調査/精査、特定の AI 技術/モデルの予備的利用、またこれらを用いたプロトタイピングを実施し、大きなプロセス変更を包含するデータマネジメントの「変革」の可能性について考察を行うことを考えた。

製薬協 DS 部会では、2022 年度および 2023 年度を通じて以下の 4 種類の AI のプロトタイピングを実施され、その結果が公表されている。

- 有害事象等の安全性情報を検出する AI
- 論理チェック仕様からクエリメッセージを作成する AI
- 併用薬剤の適応の妥当性を予測する AI
- 疾患名の辞書コーディング結果を予測する AI

今回の成果物はこれらに続くタスフォース独自の研究の結果をまとめたものである。日本のデータマネジメントに関わる公表物には文献レビュー的なものが少なくないと思われる。その意味で、このタスフォース独自の研究計画とその結果をまとめた本書は、筆者らが自ら悩みながら汗を流しながら検討を進めていった様子を伺い知ることができるユニークな資料になっていると思われる。時間の都合上、当初の予定どおりに進まなかったところもあるが、是非ご一読いただき、本成果物をまとめるまでに筆者らの実施してきたこと、考えたことなど追体験していただきたいと思っている。

なお、タスフォースにて作成・公開したプログラムおよびプロンプトについては、AI 技術の急速な進展に伴い、一部に陳腐化が見られる可能性がある。読者各位におかれては、プロンプトやその出力結果そのものよりも、取り組みにおけるアプローチや思考過程に着目いただければ幸いである。

2. AI 活用の検討対象の候補の抽出および選定

今年度の活動としては大幅な効率化や顕著な品質向上が必要と思われる以下のプロセスや業務等をまずブレインストーミング的に取り上げ、AI 活用によるビジネスバリュー、AI の適用方法、AI を適用する上での課題等について検討を進めていった。

- ① 業務上必要な規制当局や関連団体から発信される外部情報のタイムリーな収集
- ② 試験実施計画書に基づいたデータチェック仕様の作成
- ③ デジタルバイオマーカーや eCOA 等種々の様々なソースから逐次発生するデータに対するタイムリーなレビュー
- ④ 前相試験や Real World Data 等を用いて対象疾患等の臨床検査やバイタルサイン変動モデルを作成、逸脱パターンを検出
- ⑤ 試験計画立案時または変更時におけるその試験に特有な潜在的リスクの特定
- ⑥ 自然言語で記載された仕様の実行可能なプログラムへの翻訳
- ⑦ データリスティングのレビュー実施
- ⑧ 個別参加者の臨床データ要約（「Patient Profile」とも言われるもの）のレビューおよびデータの不整合・懸念点の指摘
- ⑨ 監査証跡レビュー（Audit Trail Review）の実施
- ⑩ 通常と異なるデータパターンの検出
- ⑪ 監査または規制当局の信頼性調査の模擬的实施による潜在的問題点の把握および是正措置
- ⑫ ICH E6 (R3)で重要視されている Critical-to-Quality Factor（CTQ 要因）の特定

候補の個々について、実現に必要なデータの準備や法規制等の非技術的な課題を踏まえ、タスクフォースで取り組む対象を議論していった。例えば上述⑨、⑩については、テストデータに各社の実際の試験のデータを利用するわけにはいかず、これらの目的に合った特性を持つデータを人工的に作成するのは困難と判断された。また①については技術的には可能と思われビジネスバリューも高いと判断されたが、たとえ更新時にのみ特定の URL へ情報を取得しに行く仕組みを構築しても、自動取得する仕組みそのものがサーバーに過大な負荷をかけ、システムの動作に支障を与える可能性があり、検討の対象とするには不適切と判断された。

その他、技術的な難易度等も踏まえて検討を進めて行ったが、検討当時 ICH E6 (R3) Step 4 が間近に迫っているという状況であったために、ICH E8 (R1) [9]や ICH E6 (R3) [10]に関連したプロトタイピングに取り組むことができないかと考えるに至った。そこで「Critical to Quality」や「Risk-proportionate」などの試験の質の管理に関するキーワードに強く引き付けられた。そのような経緯で上述の⑫と⑧をもとにして以下の2つを行うAIのプロトタイピングを本年度の検討の対象とすることにした。

- CTQ 要因の特定支援 (⑫から派生)
- 試験実施計画に基づいた Critical な疑義事項の検出 (⑧から派生)

次いで今回のプロトタイピングを実施するにあたり、適していかつ利用可能なAI技術/モデルの検討を行った。2023年度活動の成果物「データマネジメントにおけるAIの活用～AIをより身近な存在に～」[5] (2024年6月公表) ではファインチューニングの活用によってAIを開発することが主であった。この資料からファインチューニングによるアプローチの場合、相当量のバランスの良い学習データが必要となり、また新たに作成するとなるとかなりの負荷がかかるであろうということが読み取れる。一方、利用可能な大規模言語モデルをそのまま使用するアプローチの場合は、学習データの用意の必要もなく、ある程度の性能を発揮できたということが述べられている。今回の二つのプロトタイピングを実施するにあたりファインチューニングによるアプローチは到底現実的とは思われなかった。幸いAI技術の急速な進展により、OpenAI や Google などの企業は、ユーザー拡大とデータ収集を目的に無料で使用できる対話型大規模言語モデルを公開してきた。これらは各社により不定期に更新されるため同一モデルの継続利用は難しいという側面はあるものの、その有用性に対する期待は極めて高かった。そこで今回のプロトタイピングでは、機密情報に十分配慮しつつ、利用可能な大規模言語モデルをそのまま使用するアプローチを取ることにした。

3. 選定したAI活用の検討を実現するための試験計画およびデータの入手

上述の「CTQ 要因の特定支援」の検討を行うためには具体的で詳細な試験実施計画が必要であり、また「試験実施計画に基づいた Critical な疑義事項の検出」の検討を行うために

は具体的な試験実施計画に基づき収集されたデータが必要であった。しかし、タスクフォースメンバーが所属する会社において実際に使用された試験実施計画書やその試験で実際に収集されたデータを AI に入力することは、機密情報保護等の観点から選択肢にはならなかった。とはいえ、架空の試験実施計画書を作成して、それに沿ったデータを作成することもリソースの負荷という観点で現実的ではなかった。

幸い CDISC の `sdtm-adam-pilot-project` で使用された試験実施計画書とデータが公開されていた¹のでそれを活用することを考えたが、「試験実施計画に基づいた Critical な疑義事項の検出」のためにはデータにエラーを含ませなければならなかった。しかし、`sdtm-adam-pilot-project` のデータの利用条件には「改変をしないで使用すること」が明記されていたため、タスクフォースは `sdtm-adam-pilot-project` の試験実施計画書およびデータの権利所有者である CDISC に本タスクフォースでの検討の意図を伝え、特別にデータを改変して使用することの許可を得て、本タスクフォースでの検討に使用した。

`sdtm-adam-pilot-project` のデータは使用するにあたりプログラムにて書き換えを行っており、書き換えた後のデータ、書き換えに使用したプログラムおよびアプリケーションは GitHub に公開した²。必要なアカウントを取得すれば誰でも無料で実行可能であるため、実際の挙動や検討すべき事項の確認など、今後の AI 活用を検討する際の参考となれば幸いである³。

4. CTQ 要因の特定支援

4.1. CTQ 要因特定の意義と AI 活用の可能性

ICH E8(R1)や ICH E6 (R3)で重要視される CTQ 要因とは、臨床試験の質を確保するための重要な要因であり、参加者の保護、試験結果の信頼性と解釈の可能性および試験結果に基づく意思決定に極めて重大な影響を与えるものである [11]。CTQ 要因は ICH E8(R1)で提唱された Quality by Design (QbD) の理念を具体化する手段として位置づけられ、CTQ 要因の明確化はリスクベース品質管理 (RBQM) の出発点となる。また CTQ 要因の特定は、試験固有の重要事項に焦点を当て、リソースの最適配分や効率的なモニタリングを可能にするものである。しかし、その特定には試験実施計画、疾患特性、治験薬の特徴など多様な情報を総合判断する必要がある、さらに新たな概念であることから、抜け漏れや過剰管理のリスクも懸念される。

AI が CTQ 要因という概念を適切に整理した上で、与えられた試験計画の対象集団、評価項目、治験薬の特徴、手順やデザイン等に基づいてその試験に特有な CTQ 要因の候補および CTQ 要因に関連したリスクに対するリスク低減策をどの程度提案できるか、人間は CTQ 要因の特定に AI を利用する場合にどのように関わっていくべきかに興味があった。

¹ <https://github.com/cdisc-org/sdtm-adam-pilot-project>

² <https://github.com/Takumi173/JPMA2024TF1-2>

³ プログラムの作者および製薬協は、本プログラムの実行によって生じたいかなる損害に対して、一切の責任を負いません。

本タスクフォースでは、この新たな概念に素早く適応できるよう、CTQ 要因の特定を支援する対話型 AI を構築し、AI と人間が協力して進めるプロセスを検討した。なお、実際に AI を使って作業をしていた時期は 2024 年 10 月から 2025 年 2 月までである。

4.2. 目的

AI により生成された素案および低減策素案の妥当性を評価するとともに、AI を用いて CTQ 要因素案等を作成させることの有用性、より効果的な活用方法、ならびに利用時の留意点について考察を行う。

4.3. 予備的試行（AI に CTQ 要因を特定させるワークフローの検討）

まず予備的試行として、AI にすべての CTQ 要因を特定させるワークフローの構築を検討した。具体的には、試行当時に利用可能であった大規模言語モデル「Gemini 1.5 Pro」を使用して、試験実施計画書と Clinical Trials Transformation Initiative (CTTI) が作成した CTQ 要因に関する指針文書「CTTI QUALITY BY DESIGN PROJECT - CRITICAL TO QUALITY (CTQ) FACTORS PRINCIPLES DOCUMENT⁴」を読み込ませ、当該試験の CTQ 要因を特定するように指示を与えた。その結果、AI は CTQ 要因となりうる情報を提示したものの、試験実施計画書に記載されている情報のみを元にした一般的な回答の提示にとどまった。これを踏まえてより具体的な内容にすべく CTQ 要因を引き出せるようプロンプトの調整を行ったが、タスクフォースメンバーも CTQ 要因に精通しておらず、試験実施計画書の対象疾患領域に詳しいメンバーもいなかったため、AI に実施してもらいたい内容を明確にプロンプトに記述することは困難であると判断した。そこで、最初から AI に CTQ 要因の素案作成させるのではなく、AI から CTQ 要因を特定するために不足している情報を投げかけてもらうワークフローに変更した。具体的には、CTQ 要因を特定するための知識となる参考資料を AI から参照可能にし、CTQ 要因の特定に必要な情報を AI に質問してもらうことでユーザーが追加の情報を提供するという、AI と協業しながら CTQ 要因の特定を行う形式を目指した。

4.4. 方法

以下の手順で、AI に CTQ 要因に関する文献を参照させ、指定された試験実施計画を読み込ませた上で、さらに必要と思われる情報を人間が会話形式で与えることにより、具体的な CTQ 要因素案およびその CTQ 要因素案に関するリスクとそのリスク低減策案を生成した。最後にタスクフォースでレビューを行った。(生成結果 1)。タスクフォースが考察した改善点を AI に伝え、修正した CTQ 要因素案を生成させ、考察した。(生成結果 2)

⁴ https://ctti-clinicaltrials.org/wp-content/uploads/2021/07/CTTI_QbD_Workshop_Principles_Document.pdf

方法の概要

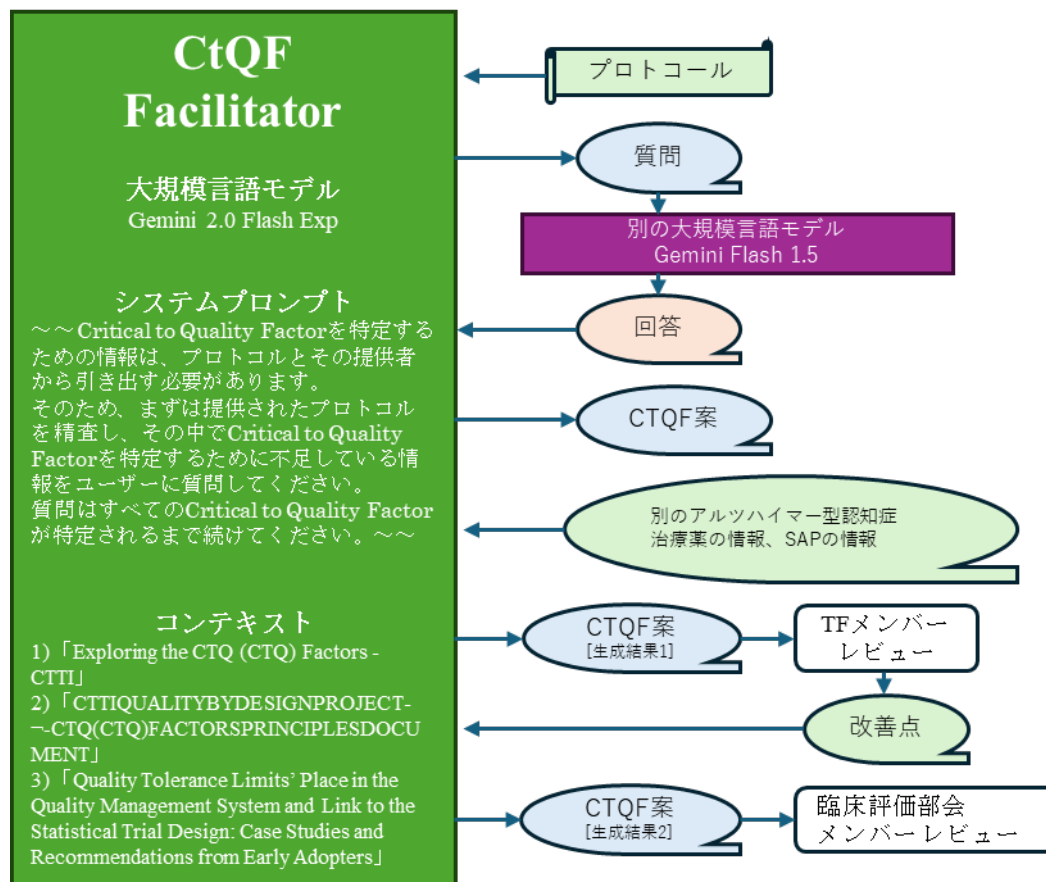


図 4-4-1 方法の流れ

方法の詳細は、以下の①～⑪に示す。

① 対話型 AI「CtQF Facilitator」の作成

オープンソースプラットフォーム「Dify⁵」を用いて「CtQF Facilitator」を作成した。

- 使用した大規模言語モデル：「Gemini 2.0 Flash Exp」
- 設定したシステムプロンプト：

あなたはファシリテーターとしてユーザの提供するプロトコールの Critical to Quality Factor を特定していきます。

Critical to Quality Factor を特定するための情報は、プロトコールとその提供者から引き出す必要があります。

そのため、まずは提供されたプロトコールを精査し、その中で Critical to Quality Factor を特定するために不足している情報をユーザーに質問してください。

質問はすべての Critical to Quality Factor が特定されるまで続けてください。

なお、ユーザーはこれまでにいくつもの薬剤を世界に送り出している製薬メーカーであ

⁵ <https://dify.ai/>

り、評価項目の設定に疑義は無く、試験を実施する上での課題となる Critical to Quality Factor を考慮したリスクを特定するようにしてください。

与えられたプロトコル以外でも起こりうる要因は Critical to Quality Factor として考慮する必要があります。

以下の知識を参考にして、ユーザーをファシリテートしてください。

知識

[コンテキスト]

- コンテキストとして持たせた資料：
 1. Exploring the Critical to Quality (CTQ) Factors⁶
 2. CTTI QUALITY BY DESIGN PROJECT - CRITICAL TO QUALITY (CTQ) FACTORS PRINCIPLES DOCUMENT⁴
 3. Quality Tolerance Limits' Place in the Quality Management System and Link to the Statistical Trial Design: Case Studies and Recommendations from Early Adopters⁷

② CtQF Facilitator へ試験実施計画書をアップロード

試験実施計画書の PDF をアップロードした。

- タイトル：「Safety and Efficacy of the Xanomeline Transdermal Therapeutic System (TTS) in Patients with Mild to Moderate Alzheimer's Disease」
- 有効成分：Xanomeline
- 剤型：経皮パッチ製剤
- 対象疾患：軽～中等度アルツハイマー病
- 主要評価項目：ADAS-Cog, CIBIC+, 安全性
- デザイン：無作為化・二重盲検・並行 3 群プラセボ対象
- 期間：26 週間

詳細は CDISC の sdtm-adam-pilot-project を参照

③ CtQF Facilitator からの質問

試験実施計画書のアップロードを受けて、CtQF Facilitator から質問が返された。

抜粋)

「・・・Critical to Quality Factor を特定するために、以下の質問をさせてください。

質問 1: TTS 製剤の貼付手順について、患者自身または介護者が行う際に、特に間違い

⁶ <https://ctti-clinicaltrials.org/about/ctti-projects/quality-by-design/qbd-quality-by-design-toolkit/exploring-the-critical-to-quality-ctq-factors/>

⁷ https://pmc.ncbi.nlm.nih.gov/articles/PMC10276776/pdf/43441_2023_Article_504.pdf

やすい、または困難になりやすいと考えられる具体的な手順はありますか？

質問 2: 評価項目（ADAS-Cog、CIBIC+、DAD、NPI-X）の実施において、評価者によるばらつきを最小限にするために、どのようなトレーニングや手順が計画されていますか？また、評価者が複数いる場合、評価者間の一貫性をどのように確保しますか？

質問 3: 心血管系の安全性評価（ECG、24 時間ホルター心電図）において、データ収集や解析の際に、特に注意すべき点はありますか？例えば、ホルター心電図の装着時間や解析の際のアーチファクトの処理などについて教えてください。

質問 4: 患者の組み入れ基準と除外基準について、スクリーニング時に特に判断が難しい、または誤って組み入れてしまう可能性のある項目はありますか？

質問 5: 併用が禁止されている薬剤について、患者が誤って使用してしまう可能性が高いと考えられる薬剤はありますか？また、その場合、どのように対応しますか？」

- ④ CtQF Facilitator からの質問を、別の大規模言語モデル（Gemini Flash 1.5）へ質問し、回答を得た。（このモデルに関しては Google が公開しているインターフェースを利用した。）

- ⑤ ④で得た回答を、CtQF Facilitator へ転用し質問した。

抜粋)

「以下が質問 1 の回答です・・・以下が質問 2 の回答です・・・」

- ⑥ CtQF Facilitator からの CTQ 要因素案の提示

CtQF Facilitator から CTQ 要因素案が提示された。しかしながら、最初に提示された CTQ 要因素案が、「CTQ 要因素案、CTQ 要因素案に対するリスク」という構成でなかったため、構成変更を依頼する旨のプロンプトを送り、構成変更された CTQ 要因素案を最終的に得た。5つの項目に対して、それぞれ、具体的な課題、CTQ 要因素案、リスク、リスク低減策案が提示された。

- ⑦ 類薬の情報、SAP の情報を追加提供

アルツハイマー型認知症を適応に持ちかつ同剤型の治療薬の添付文書の「基本的な注意事項」の情報、および審査報告書の「解析対象集団から除外された事例」の情報、および解析計画（Statistical Analysis Plan、SAP）の情報を CtQF Facilitator へ提供した。

⑧ CtQF Facilitator による CTQ 要因素案の再生成

CtQF Facilitator により CTQ 要因素案を再生成させた。：生成結果 1

⑨ タスクフォースメンバーによる、生成された CTQ 要因素案のレビュー

タスクフォースメンバーにて、生成された CTQ 要因素案の妥当性・有用性を検討した。

⑩ タスクフォースメンバーにより見出された改善点を、CtQF Facilitator へフィードバック

⑪ CtQF Facilitator による CTQ 要因素案の再生成

⑫ CtQF Facilitator へ、構成変更依頼

「リスク」として記載されていることはそのリスクが顕在した場合の帰結であって、「課題」として書かれていることが CtQ 要因に対応するリスクと考えられます。

このことを踏まえ Critical to Quality Factor 等を以下の要領で記載しなおしてください。

・Critical to Quality Factor

・リスク

・リスク低減策

⑬ CtQF Facilitator による CTQ 要因素案の再生成

「具体的な課題」の内容がそのまま「リスク」に置き換わり、元々の「リスク」の内容が削除された CTQ 要因素案が作成された。

⑭ 臨床評価部会有志による生成結果に対するレビュー：生成結果 2

なお、③から⑬までの CtQF Facilitator との応答記録は GitHub で公開している⁸。

⁸ <https://github.com/Takumi173/JPMA2024TF1-2>

4.5. CTQ 要因素案等の生成結果

4.5.1. 生成結果 1

表 4.5.1-1 生成結果 1

項目	具体的な課題	CTQ 要因素案	リスク	リスク低減策素案
1. TTS 製剤の貼付手順の遵守	<ul style="list-style-type: none"> 貼付部位の選択ミス（同じ場所への繰り返し貼付、毛が多い部分や傷のある部分への貼付） 保護フィルムの剥がし方ミス（薬剤への接触） 空気の混入 貼付後の圧着不足 貼付頻度・時間の誤り 高齢者、小児、関節可動域制限のある患者における貼付困難 貼付済み製剤の除去忘れ 	<ul style="list-style-type: none"> TTS 製剤の貼付手順の標準化 貼付済み製剤の除去確認 	<ul style="list-style-type: none"> 薬物動態の変動、有効性の低下 皮膚トラブル（刺激、かぶれ、剥がれ） 患者の不快感、治療への不信感 過量投与による重篤な副作用 ITT 解析と PP 解析の結果の乖離 As Treated 解析の信頼性低下 	<ul style="list-style-type: none"> 貼付部位のローテーション、貼付方法、貼付時間、剥がし方、剥がれた場合の対応などを詳細に記載したマニュアルを作成し、患者と介護者に配布する。 実演形式での貼付トレーニングを実施し、理解度を確認する。 貼付状況を定期的に確認し、問題があれば早期に対応する。 貼付カレンダーや鏡などの補助具を提供する。 介護者への説明を徹底する。 貼り替え時に貼付済み製剤を除去したことを確認する手順を徹底する（チェックリスト、写真撮影など）。 貼付状況の記録を徹底し、As Treated 解析の精度を向上させる。 投与量の変更があった場合の解析方法を事前に検討する。
2. 評価項目の実施手順の標準化と評価者	<ul style="list-style-type: none"> 評価者による評価のばらつき 	<ul style="list-style-type: none"> 評価項目の実施手順の標準 	<ul style="list-style-type: none"> 評価結果の信頼性低下 	<ul style="list-style-type: none"> 評価項目の定義、実施手順、スコアリング基準などを詳細に記述した評価マニュアルを作成す

項目	具体的な課題	CTQ 要因素案	リスク	リスク低減策素案
の一貫性確保	<ul style="list-style-type: none"> ・評価環境の不統一 ・評価順序の不統一 ・評価ツールの不統一 ・心理評価者の変更 	化 <ul style="list-style-type: none"> ・評価者の一貫性確保 ・評価者の変更管理 	<ul style="list-style-type: none"> ・試験結果の解釈困難 ・統計解析におけるバイアス ・評価結果の比較困難 ・主要評価項目、副次評価項目の解析結果の信頼性低下 ・交互作用の評価への影響 	る。 <ul style="list-style-type: none"> ・模擬評価やビデオ教材を用いたトレーニングを実施し、評価者のスキルを向上させる。 ・評価者間の定期的なミーティングを実施し、評価基準の共有を図る。 ・カルキュレーションチェックを実施し、計算ミスやスコアリングの誤りを検出する。 ・統計学的手法を用いて評価者間の一致度を定量的に評価する。 ・可能な限り盲検法を導入する。 ・評価者の選定基準を明確化し、関連分野の知識や経験を持つ者を採用する。 ・評価者変更時の手順を標準化し、変更理由を記録する。 ・変更後の評価者に対して、十分なトレーニングを実施する。 ・評価データに影響を与える評価者の変更を最小限にする。 ・交互作用の評価に影響を与える可能性のある評価者の偏りを最小限にする。
3. 心血管系の安全性 評価の正確性	<ul style="list-style-type: none"> ・ECG 電極の装着ミス ・ノイズの混入 	<ul style="list-style-type: none"> ・心血管系の安全性評価の実 	<ul style="list-style-type: none"> ・心血管系イベントの見逃し 	<ul style="list-style-type: none"> ・ECG 電極の正しい位置、装着方法、ノイズ対策などを詳細に規定した手順書を作成し、実施者

項目	具体的な課題	CTQ 要因素案	リスク	リスク低減策素案
	<ul style="list-style-type: none"> ・ホルター心電図の装着時間不足 ・電極の脱落 ・アーチファクトの処理ミス ・解析ソフトウェアの選択ミス 	<ul style="list-style-type: none"> ・施手順の標準化 ・データ解析の正確性確保 	<ul style="list-style-type: none"> ・患者の安全性の脅威 ・データ解析における誤った結論 ・安全性解析の結果の信頼性低下 	<ul style="list-style-type: none"> ・にトレーニングする。 ・ホルター心電図の装着時間、電極の脱落防止策、日常生活記録の重要性などを詳細に規定した手順書を作成し、患者に説明する。 ・アーチファクトの処理、心拍数解析、不整脈検出、QT 間隔測定などの解析手順を標準化し、解析ソフトウェアの選択基準を明確にする。 ・心電図やホルター心電図の解析結果は、医師や臨床工学技士などの専門家が解釈する。 ・安全性解析に影響を与えるデータ欠損を最小限にする。
4. 患者の組み入れ基準と除外基準の厳格な適用	<ul style="list-style-type: none"> ・疾患の鑑別困難（うつ病、血管性認知症、レビー小体型認知症など） ・病歴の申告漏れ、虚偽申告 ・診断の遅延、誤診 ・スクリーニングプロセスにおけるエラー ・病状の変化 ・薬物依存の定義の違い ・地理的要因による選択バイアス 	<ul style="list-style-type: none"> ・スクリーニング手順の標準化 ・患者情報の徹底的な確認 ・アルツハイマー型認知症の診断の確認 	<ul style="list-style-type: none"> ・不適切な患者の組み入れ ・試験結果の信頼性低下 ・患者の安全性への影響 ・解析対象集団からの除外 ・ITT 解析と PP 解析の結果の乖離 ・層別解析の解釈への 	<ul style="list-style-type: none"> ・組み入れ基準と除外基準を詳細に規定した手順書を作成し、スクリーニング担当者にトレーニングする。 ・患者へのヒアリングだけでなく、過去の医療記録や家族からの情報収集を徹底する。 ・必要に応じて、精神科医や神経内科医などの専門医による評価を実施する。 ・インフォームド・コンセントを徹底し、患者が治療の内容を十分に理解した上で参加できるようにする。 ・介護者の負担を軽減するためのサポート体制を

項目	具体的な課題	CTQ 要因素案	リスク	リスク低減策素案
	<ul style="list-style-type: none"> ・介護者の負担増加 		影響	整備する。 <ul style="list-style-type: none"> ・アルツハイマー型認知症の診断を確定するための追加検査（画像検査、バイオマーカー検査など）を検討する。 ・スクリーニング時の情報収集を徹底し、ITT 解析と PP 解析の結果の乖離を最小限にする。 ・層別解析の結果に影響を与える可能性のある患者の偏りを最小限にする。
5. 併用禁止薬の使用防止	<ul style="list-style-type: none"> ・患者による併用禁止薬の誤使用 ・有効性評価に影響を与える併用療法 ・オピオイド製剤の使用 	<ul style="list-style-type: none"> ・併用禁止薬リストの徹底的な周知 ・併用薬の確認 ・併用療法、オピオイド製剤の使用状況の確認 	<ul style="list-style-type: none"> ・薬物相互作用による副作用 ・試験結果への影響 ・有効性評価のバイアス ・解析対象集団からの除外 ・ITT 解析と PP 解析の結果の乖離 	<ul style="list-style-type: none"> ・併用禁止薬リストを患者と介護者に分かりやすく説明し、理解度を確認する。 ・治験期間中に患者が使用している薬剤を定期的に確認し、併用禁止薬の使用を早期に発見する。 ・併用禁止薬を使用した場合の対応手順を明確化し、実施者に周知する。 ・併用療法、オピオイド製剤の使用状況を定期的に確認し、記録する。 ・併用療法、オピオイド製剤の使用が判明した場合の対応手順を明確化する。 ・併用薬の使用状況を記録し、ITT 解析と PP 解析の結果の乖離を最小限にする。
6. その他	<ul style="list-style-type: none"> ・中止基準違反 ・治験薬投与違反 	<ul style="list-style-type: none"> ・中止基準の遵守 	<ul style="list-style-type: none"> ・解析対象集団からの除外 	<ul style="list-style-type: none"> ・中止基準を詳細に規定し、実施者に周知する。 ・治験薬の投与手順を標準化し、実施者にトレー

項目	具体的な課題	CTQ 要因素案	リスク	リスク低減策素案
	<ul style="list-style-type: none"> ・脱水症状 ・体重減少 ・欠測値の増加 ・データの正規性の逸脱 	<ul style="list-style-type: none"> ・治験薬投与手順の遵守 ・脱水症状の早期発見と対応 ・体重変化のモニタリング ・欠測値の最小化 ・データの正規性の確認 	<ul style="list-style-type: none"> ・試験結果の信頼性低下 ・患者の安全性の脅威 ・統計解析におけるバイアス ・ランク変換による解析の必要性 	<ul style="list-style-type: none"> ・モニタリングする。 ・患者と介護者に脱水症状の兆候を説明し、早期に報告するように指導する。 ・体重変化を定期的にモニタリングし、異常があれば医師に報告する。 ・脱落患者の追跡調査を行い、欠測値を最小限にする。 ・LOCF 法以外の欠測値補完方法も検討する。 ・データの正規性を確認し、必要に応じてランク変換による解析を行う。

対象疾患（アルツハイマー型認知症）と薬剤の剤形（TTS 製剤）の特徴を捉えた結果が出力されていると判断できた。具体的には、対象疾患の鑑別を CTQ 要因素案として生成した上で、診断の正確性を担保するための方法（例：患者へのヒアリングだけでなく、過去の医療記録や家族からの情報収集を徹底）をリスク低減策案として生成したり、薬剤の剤形が貼付薬であることに基づき特定された CTQ 要因を損なうリスクの低減策として、患者と介護者へのマニュアル配布の提案と、その具体的な内容（例：貼付部位のローテーション、貼付方法、貼付時間、剥がし方、剥がれた場合の対応）の例示が適切に提示されたものと判断できた。

また、情報として与えた試験実施計画、類似薬の「基本的な注意事項」、審査報告書の「解析対象集団から除外された事例」、SAP、および CtQF Facilitator からの質問に対する回答を踏まえ、多角的な視点から網羅的な内容が出力されていると判断できた。

例えば、本試験の評価項目には複数の臨床医による評価ツール（チェックリストやスケールなど）を用いた主観的な評価（CIBIC+など）や介護者による報告に基づいた評価（NPI-Q など）が多く設定されている。これに基づき「評価項目の実施手順の標準化と評価者の一貫性確保」を課題に挙げ、その CTQ 要因を損なうリスクの低減策として、以下のように試験の各段階にて実施依頼者と施設でできる対応が AI により提示されたものと判断できた（注：提示されたリスク低減策の一覧に基づき、タスクフォースで分類）。

表 4.5.1-2 試験の各段階にて、実施依頼者と施設でできる対応

	試験実施依頼者	試験実施施設
試験準備段階	<ul style="list-style-type: none"> ● 評価項目の定義、実施手順、スコアリング基準などを記述した評価マニュアル作成 ● 模擬評価やビデオ教材を用いたトレーニングの実施 ● 盲検法の導入 	<ul style="list-style-type: none"> ● 評価者の選定基準を明確化し、関連分野の知識や経験を持つ者を採用
試験実施中	<ul style="list-style-type: none"> ● カルキュレーションチェックによる、計算ミスやスコアリングの誤りの検出 ● 変更後の評価者に対し、十分なトレーニングの実施 	<ul style="list-style-type: none"> ● 評価者間の定期的なミーティングの実施 ● 評価者変更時の手順を標準化し、変更理由を記録 ● 評価データに影響を与える評価者の変更を最小限にする ● 交互作用の評価に影響を与える可能性のある評価者の偏りを最小限にする
解析時	<ul style="list-style-type: none"> ● 統計学的手法を用いた評価者 	NA

	間の一致度の定量的評価	
--	-------------	--

さらに、その他の課題として「脱水症状の早期発見と対応」を CTQ 要因素案として挙げている点も注目すべきと考える。対象試験の薬剤はムスカリン受容体のアゴニスト、すなわちコリン作動薬であることが対象試験の試験実施計画にて明記されている。また一般的にコリンエステラーゼ阻害剤は、ムスカリン受容体刺激を伴う作用機序により、嘔吐あるいは下痢の持続により脱水が現れることがあり、ゆえに、類似薬の「基本的な注意事項」にもこの点を踏まえて記載がされている。一方、後者の注意事項については特に対象試験の試験実施計画には明記されていないことから、読み込ませた情報を組み合わせ複合的に考えた上で CTQ 要因素案を生成していることは注目に値する。

一方、回答結果には対象の試験実施計画や薬剤に特化しない一般的な情報も含まれていた。例えば、対象試験に限らず多くの試験で採用されている心電図検査に焦点を当てた「心血管系の安全性評価の正確性」を課題とした CTQ 要因素案も提示されたが、それを損なうリスクと低減策含め、対象試験において特に注意すべき項目とは思われない内容であった。また、「併用禁止薬の使用防止」を課題とした CTQ 要因素案も提示されたが、併用禁止薬は試験実施計画に含まれる一般的な記載項目であり、またそのリスクや低減策も本試験に特異的でない一般的な内容が多く含まれていた。また、オピオイド製剤について言及されているものの、類似薬の審査報告書の「解析対象集団から除外された事例」にオピオイド製剤の服用した参加者が含まれたことが要因と考えられ、対象薬剤の課題として含めるには根拠が乏しいと判断された。

加えて、試験実施計画に沿わない内容や、CTQ 要因素案やそれを損なうリスクの低減策案としては提示が不要と思われるものも見受けられた。特にリスク低減策案については、試験実施計画に対するものは具体的に示されていた一方、あとから SAP を追加情報として与えたことで追加された解析に関するものは抽象的な内容が多かった。

表 4.5.1-3 出力結果 1 に対する、タスクフォースメンバーから挙げた疑義事項

特定された課題	出力結果	疑義事項
TTS 製剤の貼付手順の遵守	具体的な課題にて、小児への薬剤貼付にかかる留意事項あり	本試験の対象は 50 歳以上である
	リスクにて、「患者の不快感、治療への不信感」を列挙	読み込ませた情報からは読み取れない内容であり、記載根拠が不明
評価項目の実施手順の標準化と評価者の一貫性確保	リスク低減策にて、盲検法導入を推奨	本試験は二重盲検比較試験である
	リスク低減策にて、「交互作用の評価に影響を与える可能性のあ	SAP を追加情報として加えたことで追加されたが、具体性がなく

	る評価者の偏りを最小限にする」ことを推奨	かつ現実的には対応が困難である
心血管系の安全性 評価の正確性	リスク低減策にて、「日常生活記録の重要性などを詳細に規定した手順書を作成し、患者に説明する」ことを推奨	本課題に関連する内容ではあるものの、やや派生した事項であり重要性は他と比較して低い
患者の組み入れ基準と除外基準の厳格な適用	リスク低減策にて「層別解析の結果に影響を与える可能性のある患者の偏りを最小限にする」ことを推奨	SAP を追加情報として加えたことで追加されたが、具体性がなくかつ現実的には対応が困難である
併用禁止薬の使用 防止	CTQ 要因にて「オピオイド製剤の使用」に言及	類似薬で Per Protocol Set から当該製剤の服用者が除外されたことが記載根拠と推察されるが、発生頻度や機序などが考慮されたものではない
	リスク低減策にて、「併用薬の使用状況を記録し、ITT 解析と PP 解析の結果の乖離を最小限にする」ことを推奨	リスク低減策としてあえて記載されなくとも、一般的に留意されている事項である
その他	CTQ 要因にて「治験薬投与手順の遵守」を列挙	既に課題として記載された「TTS 製剤の貼付手順の遵守」と重複

4.5.2. 生成結果 2

表 4.5.2-1 生成結果 2

項目	CTQ 要因素案	リスク	リスク低減策素案
1. TTS 製剤の貼付手順の遵守	<ul style="list-style-type: none"> ・ TTS 製剤の貼付手順の標準化 ・ 貼付済み製剤の除去確認 	<ul style="list-style-type: none"> ・ 貼付部位の選択ミス ・ 保護フィルムの剥がし方ミス ・ 空気の混入 ・ 貼付後の圧着不足 ・ 貼付頻度・時間の誤り ・ 高齢者、関節可動域制限のある患者における貼付困難 ・ 貼付済み製剤の除去忘れ 	<ul style="list-style-type: none"> ・ 貼付部位のローテーション、貼付方法、貼付時間、剥がし方、剥がれた場合の対応などを詳細に記載したマニュアルを作成し、患者と介護者に配布する。 ・ 実演形式での貼付トレーニングを実施し、理解度を確認する。 ・ 貼付状況を定期的に確認し、問題があれば早期に対応する。 ・ 貼付カレンダーや鏡などの補助具を提供する。 ・ 介護者への説明を徹底する。 ・ 貼り替え時に貼付済み製剤を除去したことを確認する手順を徹底する（チェックリスト、写真撮影など）。 ・ 貼付状況の記録を徹底し、ITT 解析の精度を向上させる。 ・ 投与量の変更があった場合の解析方法を事前に検討する。
2. 評価項目の実施手順の標準化と評価者の一貫性確保	<ul style="list-style-type: none"> ・ 評価項目の実施手順の標準化 ・ 評価者の一貫性確保 ・ 評価者の変更管理 	<ul style="list-style-type: none"> ・ 評価者による評価のばらつき ・ 評価環境の不統一 ・ 評価順序の不統一 ・ 評価ツールの不統一 ・ 心理評価者の変更 	<ul style="list-style-type: none"> ・ 評価項目の定義、実施手順、スコアリング基準などを詳細に記述した評価マニュアルを作成する。 ・ 模擬評価やビデオ教材を用いたトレーニングを実施し、評価者のスキルを向上させる。 ・ 評価者間の定期的なミーティングを実施し、評価基準の共有を図る。 ・ カルキュレーションチェックを実施し、計算ミスやスコアリングの誤りを検出する。 ・ 統計学的手法を用いて評価者間の一致度を定量的に評価する。 ・ 評価者の選定基準を明確化し、関連分野の知識や経験を持つ者を採用

項目	CTQ 要因素案	リスク	リスク低減策素案
			<p>する。</p> <ul style="list-style-type: none"> ・評価者変更時の手順を標準化し、変更理由を記録する。 ・変更後の評価者に対して、十分なトレーニングを実施する。 ・評価データに影響を与える評価者の変更を最小限にする。 ・交互作用の評価に影響を与える可能性のある評価者の偏りを最小限にする。
3. 患者の組み入れ基準と除外基準の厳格な適用	<ul style="list-style-type: none"> ・スクリーニング手順の標準化 ・患者情報の徹底的な確認 ・アルツハイマー型認知症の診断の確認 	<ul style="list-style-type: none"> ・疾患の鑑別困難（うつ病、血管性認知症、レビー小体型認知症など） ・病歴の申告漏れ、虚偽申告 ・診断の遅延、誤診 ・スクリーニングプロセスにおけるエラー ・病状の変化 ・薬物依存の定義の違い ・地理的要因による選択バイアス ・介護者の負担増加 ・アルツハイマー型認知症でない患者の組み入れ 	<ul style="list-style-type: none"> ・組み入れ基準と除外基準を詳細に規定した手順書を作成し、スクリーニング担当者にトレーニングする。 ・患者へのヒアリングだけでなく、過去の医療記録や家族からの情報収集を徹底する。 ・必要に応じて、精神科医や神経内科医などの専門医による評価を実施する。 ・インフォームド・コンセントを徹底し、患者が治験の内容を十分に理解した上で参加できるようにする。 ・介護者の負担を軽減するためのサポート体制を整備する。 ・アルツハイマー型認知症の診断を確定するための追加検査（画像検査、バイオマーカー検査など）を検討する。 ・スクリーニング時の情報収集を徹底し、ITT 解析の精度を向上させる。 ・層別解析の結果に影響を与える可能性のある患者の偏りを最小限にする。
4. 併用禁止薬の使用防止	<ul style="list-style-type: none"> ・併用禁止薬リストの徹底的な周知 	<ul style="list-style-type: none"> ・患者による併用禁止薬の誤使用 	<ul style="list-style-type: none"> ・併用禁止薬リストを患者と介護者に分かりやすく説明し、理解度を確認する。

項目	CTQ 要因素案	リスク	リスク低減策素案
	<ul style="list-style-type: none"> ・併用薬の確認 ・併用療法の使用状況の確認 	<ul style="list-style-type: none"> ・有効性評価に影響を与える併用療法 	<ul style="list-style-type: none"> ・治験期間中に患者が使用している薬剤を定期的に確認し、併用禁止薬の使用を早期に発見する。 ・併用禁止薬を使用した場合の対応手順を明確化し、実施者に周知する。 ・併用療法の使用状況を定期的に確認し、記録する。 ・併用療法が判明した場合の対応手順を明確化する。 ・併用薬の使用状況を記録し、ITT 解析の精度を向上させる。
5. その他	<ul style="list-style-type: none"> ・中止基準の遵守 ・脱水症状の早期発見と対応 ・体重変化のモニタリング ・欠測値の最小化 ・データの正規性の確認 	<ul style="list-style-type: none"> ・中止基準違反 ・脱水症状 ・体重減少 ・欠測値の増加 ・データの正規性の逸脱 	<ul style="list-style-type: none"> ・中止基準を詳細に規定し、実施者に周知する。 ・患者と介護者に脱水症状の兆候を説明し、早期に報告するように指導する。 ・体重変化を定期的にモニタリングし、異常があれば医師に報告する。 ・脱落患者の追跡調査を行い、欠測値を最小限にする。 ・LOCF 法以外の欠測値補完方法も検討する。 ・データの正規性を確認し、必要に応じてランク変換による解析を行う。

生成結果 2 (表 4.5.2-1)として得られた出力結果の構成について、指摘に基づき元々具体的な課題として記載されていた事項が、リスクとしての記載に適切に修正されたと判断できた。また、初回の結果に対して誤りまたは記載不要であることを指摘した事項につき各所で該当の記載が削除されていることを確認した。一方、その他の出力内容に特に変化は見られず、引き続き、対象疾患と薬剤の剤形の特徴を捉え、かつ情報として与えた各種資料および CtQF Facilitator からの質問に対する回答を踏まえた、多角的な視点からの網羅的な内容が出力されたと判断できた。

また、生成結果 1 (表 4.5.1-1)と同様に「併用禁止薬の使用防止」が課題として含まれたが、指摘に基づき「オピオイド製剤の使用」にかかる記述が CTQ 要因を初め本課題に関する記載から全体的に削除されたことに伴い、初回よりも一般的な内容の記載にとどまった。その一方で、追加の指摘に基づき初回で誤りまたは記載不要と思われた記述は既に削除されており、一般的な記載については削除のみで容易に対応できることから許容することとし、タスクフォースメンバーの視点から追加で修正すべき箇所はないと判断した。

本タスクフォースのメンバーはデータマネジメント業務に従事している者であり、試験実施計画の作成や医療機関におけるオペレーション等には不案内であったことを鑑み、タスクフォース内での検討の後に、AI が生成した CTQ 要素案の評価に日本製薬工業協会医薬品評価委員会 臨床評価部会の有志に協力を要請した。

臨床評価部会有志による生成結果に対するレビューにおいて以下の点が指摘された。

表 4.5.2-2 臨床評価部会有志による、生成結果 2 に対する疑義事項

特定された課題	出力結果	疑義事項
NA	(生成された CTQ 要因全体について)	システムプロンプトで「ユーザーはこれまでにいくつもの薬剤を世界に送り出している製薬メーカーであり」という前提を入れたためか、全体に生成された CTQ 要因に医療機関または患者視点がありあまり考慮されていない印象を持った。特に試験の初期は来院感覚が短い。患者および介護者の負担はエンrollmentあるいはリテンションに大きな影響を与えるためその観点が CTQ 要因の特定の際により考慮されるべきである。
TTS 薬剤の貼付手	リスクとしての「貼付部位の選	認知症患者の場合、患者が手の届

順の遵守	択ミス」	く部位では剥がしてしまう可能性があるため、認知症の程度によってはできるだけ簡単に剥がせない部位に患者自身ではなく介護者が貼るように試験実施計画書の規定を変更することが「貼付部位の選択ミス」のリスクの低減策となる。
	リスクとしての「保護フィルムの剥がし方ミス、空気の混入、貼付後の圧着不足」	よほど特別な貼付剤でなければ市販後に想定される取り扱いを考慮した製剤を用いて治験を実施すると思われるので「保護フィルムの剥がし方ミス、空気の混入、貼付後の圧着不足」のような事項がリスクになるとは考えにくい。
	リスク低減策としての「貼付部位のローテーション、貼付方法、貼付時間、剥がし方、剥がれた場合の対応などを詳細に記載したマニュアルを作成し、患者と介護者に配布」	認知症患者を対象とする本治験では、介護者を設定した上で貼付は介護者が実施するように試験実施計画書の規定を変更することが貼付手順の順守に対するリスク低減策になると考える。
	リスク低減策としての「実演形式での貼付トレーニング」	よほど特殊な製剤でなければ、実演トレーニングまでは不要と考える。
	リスク低減策としての「貼付カレンダーや鏡などの補助具を提供する」	リスク低減策としては認知症である患者ではなく介護者が貼付するように試験実施計画書の規定を変更することがリスクの低減策になる。
	リスク低減策としての「貼り替え時に貼付済み製剤を除去したことを確認する手順を徹底する（チェックリスト、写真撮影など）。」	使用済み製剤を廃棄せずに、来院時に持参させて確認するような手順を設定することはあるかもしれないが、通常写真撮影までは不要と思われる。
	評価項目の実施手	リスク低減としての「評価者間
		治験開始前には参加医師等に評

順の標準化と評価者の一貫性確保	の定期的なミーティングを実施し、評価基準の共有を図る」	価者トレーニングを実施することはあるが、治験開始後に定期的トレーニングなどは一般には実施しない。
患者の組み入れ基準と除外基準の厳格な適用	リスクとしての「診断の遅延」	「診断の遅延」が何らかの治験の質に対するリスクになるとは考えにくい。
	リスクとしての「スクリーニングプロセスにおけるエラー」	具体的に何に言及しているのか不明である。
	リスクとしての「地理的要因による選択バイアス」	地理的要因はあまり想定していないように思われる。
	リスク低減策としての「必要に応じて、精神科医や神経内科医などの専門医による評価を実施」	神経心理学的検査は検査に数時間を要し、他の医師や心理士等とも分担しながら実施するのが一般的である。治験を実施する上での実施体制の確立や複数の医師間での評価結果の一貫性の確保が重要である。一方、このような実施体制の確立が可能な医療機関に限定して実施することは、症例組み入れの足かせになる可能性がある。
	リスク低減策としての「インフォームド・コンセントを徹底し、患者が治験の内容を十分に理解した上で参加できるようにする」	患者には可能な範囲で説明すると共に、法的な同意者となりうる介護者の同意が必須になる治験と思われる。

4.6. 考察

4.6.1. 全体的な所感

CTQ 要因の概念を適格に捉えた上で、与えた情報を正確に認識し、多角的な視点からの網羅的かつ具体的な結果が出力されており、CTQ 要因の特定とそれを踏まえた対応の検討を進める土台として役立つと考えられる素案が生成された。必要最低限のシステムプロンプトと文書を与えるのみで、実業務での実用性を感じられる素案が生成されたことは想定以上であり、AI の進化を肌で感じる結果となった。

4.6.2. 対話型 AI とシステムプロンプトの有用性について

まず特筆すべき事項として、CTQ 要因特定における対話型 AI の有用性である。CtQF Facilitator からの質問に回答した後に生成された CTQ 要因素案、リスク、リスク低減策案に対して、不足していると思われた情報、視点等の追加および不要または誤って生成されたと思われる情報の削除など対話を通じて指示することができたことが、想定以上の CTQ 要因素案の生成に大きく貢献したことは間違いない。

次に強調したい点としてはシステムプロンプトの有用性である。対話型 AI にて設定したシステムプロンプトとして、結果出力に向けた前提事項に加え、不足情報がある場合はユーザーに質問を投げかける工程を踏む形式としたことが、上述の結果に繋がったと考えられる。この工程により AI が初めに得られた情報のみに基づき必要以上に解釈せず、十分な情報が揃った段階で結果を出力させたことで、出力内容の網羅性と具体性を高めることに繋がった可能性が高い。例えば、対象疾患に関する詳細な説明文書を初めの情報として与えなかったが、質問 4（患者の組み入れ基準と除外基準について、スクリーニング時に特に判断が難しい、または誤って組み入れてしまう可能性のある項目はありますか？）への回答内容を踏まえ、最終的な出力結果には疾患特異的な情報として「レビー小体型認知症」が含まれた。また AI からの質問内容より、どのような情報が初回に与えた文書に不足しているかをユーザー側でも認識することができた。

4.6.3. AI をより有効活用するための留意事項

一方、出力結果を踏まえ、ユーザーが留意すべき事項もいくつか顕在化した。

まず、特定された課題として、読み込ませた試験実施計画特有とは考えにくい一般的な内容である「併用禁止薬の使用防止」が含まれた。これは 2 つの薬剤の添付文書の「重要な基本的注意」を情報として読み込ませたことによる併用禁止薬に関する情報量の増加が、原因の一つと考えられる。偏った情報を読み込ませるとバイアスのかかった結果が出力されることは AI の特性の一つであり、与える情報量のバランスに留意する、あるいは特定の事項について多くの情報を与える場合、ユーザー側で複数の情報間の優先順位を設定した上で AI にその文書の位置づけを伝えるなどの対応も必要と考えられる。

AI に読み込ませる情報の種類が結果に与える影響、という観点では、SAP を読み込ませたことで追加されたリスク低減策案は、試験実施計画単体に基づき提示されたものよりも内容の抽象度が高かったことも注目すべきである。要因としては試験実施計画と SAP の間では、文書内の試験特異的な情報の粒度に違いがあったことが可能性として考えられる。この点においては、AI から出力された結果を踏まえ、回答として求める情報の粒度を、例えば回答の想定事例を提示するなどして AI へ追加で指示し、結果を再出力することも有用と考える。

また、初回に AI から得られた出力結果として、与えた試験の情報に沿わない内容や不要な記載が複数含まれた。これは、結果の出力前に行われた AI からユーザーへの質問に対し、

別の AI が作成したものを転用して回答したが、その内容は対象の試験実施計画に特化しない一般的なものであったことが要因として考えられる。初めに強調すべき前提として、回答作成を AI に一任したことで、回答対応の負荷を軽減できたことはメリットとして大きかったと言える。一方、上述のような回答の一般化が見られたことも踏まえ、これらから得られる学びとして、**Human-in-the-Loop** の重要性とプロンプト作成時の工夫の必要性の 2 つに注目したい。

AI を用いる上で **Human-in-the-Loop** の考え方は欠かせず、各所から発表されている AI 規制関連文書の中でも、AI に一任するのではなく適宜人が介在することの重要性が強調されている。例えば今回の事例では、別の AI により作成された回答をそのまま **CtQF Facilitator** へ追加情報として与えるのではなく、対象試験に特化した内容を厳選してから与えることで、対象試験実施計画と整合しない内容が最終成果物に出力される状況を回避できた可能性がある。

CTQ 要因の特定における **Human-in-the-Loop** の意義は、リスク低減策案の効果とそれに対する負荷の評価に目を向けたときに極めて明確になる。ここでは本タスクフォースで経験した一例をあげて説明をしたい。

対象とした試験計画ではパッチ製剤を 12-14 時間貼付した後に除去することが定められていた。そのためか **DS** 部会タスクフォースが行った AI による **CTQ** 要因の特定とリスク低減策の生成の試行では、AI はリスク低減策として「製剤除去の確認手順の徹底（チェックリスト、写真撮影など）」を取り上げてきたことがあった。確かに製剤除去の度に当該箇所を写真を撮影することは試験計画順守の確認という意味では有用と言えるだろうが、参加者または介護者にとっては大きな負担となり得る。**GCP Renovation** に深く長くかかわってきたある識者は様々な場面で、本来 **Risk-proportionate Approach** が取られるべき **ICH E6 (R3)** の実践において、雪だるま式にルールやプロセスが増えていく傾向があることに懸念を示している [12]。CTQ 要因の特定においてリスクの大きさとそれに対する低減策の効果を見誤ると同様な問題が生じる。



図 4.6.3-1 雪だるま式に増えていきがちなルールやプロセス

患者または介護者にとっての負担は、患者の脱落にも繋がる可能性もある。脱落数が増えれば試験結果の妥当性に **Critical** な影響を与えかねない。そもそも、そのような患者または介護者に負担の大きい臨床試験であれば、参加者の組み入れが難航することも十分考えられる。

同様な事例として、主治医と評価者の分離が挙げられる。評価の一貫性を保つという意味では主治医と評価者を分けることは有用だと思われるが、一方でそれに対応できる医療機関の選定が足かせとなり、参加者の組み入れが進まなくなるリスクもある。

「品質」とは「対象に本来備わっている特性の集まりが要求事項を満たす程度」と定義されている。臨床試験を「対象」として品質を考えるのであれば、臨床試験の特性としては、科学的特性、方法論的特性、倫理的特性等に加えて、実施可能性（**feasibility**）も含めた運用的特性も考慮されるべきものであり、その場合の「要求事項」には「妥当な期間内での試験完了」も含まれるべきであろう。リスクの大きさ、リスク低減策の効果、リスク低減策が運用的特性に与える影響、その結果の要求事項（例、許容できる試験実施期間）の充足度などの複雑な問題を検討し最適解を導き出すことは「人間」、それも必要な専門家を含むチームでなければできないことだと思われる。

またプロンプト作成時の工夫の必要性につき、ユーザーとして AI に持っていて欲しい考え方をプロンプトに仕込むことも、AI からより有用な結果を引き出す上で必要な工夫と考える。これは AI の特性として、人が与えた情報に重きを置き、そこに忠実に結果を出力する傾向があるためである。この傾向は 2 回目の出力結果にて、誤りと伝えた箇所がそのまま削除された以外、特に出力内容に変化が見られなかった、すなわち疑義事項に追加の疑問を持つことなく指摘が結果へ直接的に反映されたことから読み取れる。具体的な工夫を加えるならば、例えば今回のように AI に対してユーザーがフィードバックを行い、指摘箇所だけでなく他箇所も再考した結果の出力を望む場合、ユーザープロンプトにて再考の依頼を含める、あるいは、ユーザーが与える情報自体に誤りがある可能性を予めシステムプロンプトに仕込むことなどが有効と考えられる。

4.6.4. 実業務における AI の活用方法の検討

以上を踏まえた上で、今回作成した CtQF Facilitator のような AI を実業務でどのように活用するのが効果的か合わせて考察したい。CTQ 要因を特定するタイミングとしては、試験の立ち上げ段階や、試験実施期間中に試験に大きな影響を与えうる重要な計画変更が生じた時などが考えられる。特にこれからの臨床試験は、試験の立ち上げ段階で各種の準備が並行して進む中で CTQ 要因の特定も行っていくことが求められてくる。そして、その内容によっては作成途中の計画書や仕様書などにリスク低減策として盛り込むことで試験の質や参加者の安全性などの向上に繋がる可能性があるため、試験の立ち上げ段階で CTQ 要因

を特定する意義は大きく、CTQ 要因を検討するための土台として網羅的な結果を即座に出力する AI の寄与は大きいと考えられる。その際、AI が生成した結果の確からしさを人が一つひとつレビューするというよりむしろ、その結果を参考にして、関係者が CTQ 要因の特定に向けた議論を進めることがより適切な進め方と思われる。国際共同試験のように標準治療が異なる複数の国が参加する試験において標準治療の差が CTQ 要因にどのように影響するかを検討する際にこのような AI を活用するという使い方も考えられる。留意点として、試験立ち上げ段階の場合、試験実施計画を含む各種計画書が定まっていないことが多い。そのため、まずは試験実施計画の骨子を読み込ませ、後から SAP など各種計画書のドラフト版を追加で情報として与える、既存の文書だけでは情報が不足する場合は、適宜他のデザインが類似した試験や類似薬の情報を与えるなど、段階的な使用も視野に入れることで活用の幅が広がると考えられる。

更に、CtQF Facilitator へ与える情報として医薬品開発の各部署に特化した文書を含めることで、また、本成果物「5.試験実施計画に基づいた Critical な疑義事項の検出」のプロトタイピングの検討結果から各 Role（試験計画者、オペレーション、試験参加医師、患者など）の視点で CTQ 要因案を作成させることで、より専門的な事項も踏まえた、具体的かつ網羅的な結果の出力に繋がる可能性がある。このことから、CtQF Facilitator の有用性を高めるためには医薬品開発に関わる複数部署の連携と関与も重要になると考えられ、また同時に部署横断的な CTQ 要因の検討が促進されることも期待ができる（巻末の「余録」を是非参照いただきたい）。

今回の反省点として、AI に CTQ 要因素案を生成させる際に、それを CTQ 要因として取り上げた理由、すなわち、試験目的の達成や患者の保護にどのように影響すると判断したのについても併せて出力させるようにしておいた方がよかったという点が挙げられる。そうしておけば AI の「思考過程」をより深く理解できた可能性がある。タスクフォースに割けるメンバーの時間にも限りがあり、これらを含めて CTQ 要因素案等を再生成することはしなかったが、同様な検討をされる方にはこれらを含めることをお勧めする。

5. 試験実施計画に基づいた Critical な疑義事項の検出

5.1. RBA を適用したデータレビュー

近年の臨床試験においては、種々の検査データに加え、ウェアラブルデバイスや ePRO（電子患者報告アウトカム）など、多様かつ複数のデータソースから情報が取得されるようになってきている。その結果、データは断続的かつ大量に発生・収集される傾向にあり、データマネジメント担当者には、これらのデータをより正確かつ迅速に処理する能力が求められている [13]。

さらに、分散型臨床試験（Decentralized Clinical Trials : DCT）の導入が進む中で、モニタリング業務およびデータマネジメント業務には、従来とは異なる観点からの柔軟性と即応性が求められるようになってきている。

このような背景を踏まえると、これまで一般的に行われてきた「症例ごとに単一なデータレビューを実施し、不備があればクエリを発行し、施設と個別にやり取りを行いながらデータ固定を進める」という一律的・手動中心のプロセスでは、これからの試験環境においては効率性と即応性に欠けるため、**Risk-Based Approach (RBA)** の考え方を取り入れ、より効果的・効率的なデータチェック手順を構築する必要があると考える [14]。

また、データマネージャーから発出されるクエリは、参加者の安全性や主要解析に直結するような重要な問い合わせから、念のため確認を行うといった軽度な問い合わせまで、さまざまなレベルのものが含まれている。そのため、すべての疑義事項を同一の優先度で処理するのではなく、内容に応じて適切に分類し、効率的に対応を進めていくことが望まれる。

本タスクフォースでは、疑義事項の重要度に応じて「**Critical**」「**Major**」「**Minor**」の3段階に分類する運用を提案する（図 5.1-1 疑義事項の重要度分類例）。「**Critical**」に該当する疑義事項は、参加者の安全性や主要解析結果に直接関わる重大なものであり、迅速な対応が求められるもの、「**Major**」は **Critical** ほど緊急性が高くないものの、何らかの対応が必要と判断されるもの、「**Minor**」は、念のためにクエリが発行された確認的な疑義事項を含み、内容によっては必ずしも修正対応等を要さないものとした。

このように疑義事項を3段階で分類することで、限られたリソースの中でも優先度の高い事項から効率的に対応を進めることが可能となり、全体としてのデータレビューの効率化と品質の向上が期待できる。

各重要度の定義は本タスクフォースでは「図 5.1-1 疑義事項の重要度分類例」のとおりとしてプロンプトを作成した。なおプロンプトでは以下に加え、各重要度の具体的な状況例も記載しているため、詳細は **GitHub** の公開データを参照していただきたい。なお、実際に AI を使って作業をしていた時期は 2024 年 10 月から 2025 年 5 月までである。

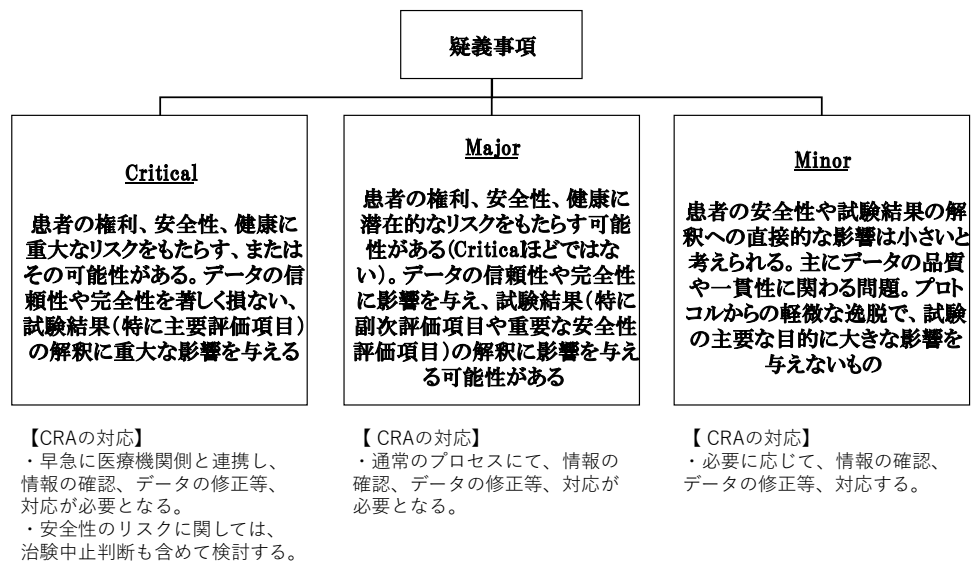


図 5.1-1 疑義事項の重要度分類例

5.2. 目的

データレビューを効率的かつ迅速に実施するためには、従来の一律的な手順では限界がある。本タスクフォースでは解決策として「5.1.RBA を適用したデータレビュー」に記載の考え方と、AI を組み合わせた新しいデータレビュープロセスにより、ますます多様化、複雑化するデータマネジメント業務をより正確かつ迅速に行い、断続的かつ大量に発生・収集されるデータの品質の向上に加え、参加者の安全性に寄与できるのではないかと考えた。

大規模言語モデルでは、標準化された治療知識やガイドライン、医学文献の内容に基づく、広範な医学知識の整理や要約、推論の迅速な処理が可能であることが知られている。実際に GPT-4 では、米国の医師国家試験に合格可能なレベルの知識が備わっていることが報告されている [15]。当然ながら、人間の医師が行うような患者の状態およびそれに影響を与える外的要因を踏まえた臨床的な総合判断はできないものの、データマネジメント担当者が行うデータレビューにおける医学的な点検に AI を利用することは有用であると考えた。本項では RBA の考え方を活かしつつ、AI を活用した効率的かつ迅速な次世代のデータレビューについて、その有用性と実現可能性を検討する。

5.3. 予備的試行（テストデータでの検討）

AI を活用したデータレビューの具体的なテストの方法を検討するため、まずは多くの臨床試験で収集されると考えられる以下の参加者情報について、10 症例分のテストデータを本タスクフォースにて作成し、このテストデータを用いて AI による疑義事項の検出精度の確認を行った。いくつかの症例には一般的に論理チェックでの検知は難しく、データマネジ

メント担当者が目視確認で検知するような疑義事項を設定した。

- 患者背景
- 併用薬剤
- 有害事象
- 臨床検査値
- 患者日誌

用意した疑義事項は以下の内容であった。

- 「臨床検査値」で白血球減少がみられるが、「有害事象」に該当データ無し
- 「併用薬剤」でアムロジピンの投与経路が外用になっている
- 「有害事象」に血小板減少、好中球減少があるが、「臨床検査値」の好中球、血小板数は微増している
- 「患者日誌」にインフルエンザとあるが、「有害事象」は COVID-19 となっている
- 「併用薬剤」でエンシトレルビルの投与理由は新型コロナ感染だが、有害事象の COVID-19 の発現日より前に投与を開始している
- 「患者日誌」にだいぶ症状が落ち着いてとあるが、各症状のスコアは悪化している
- 「併用薬剤」でダパグリフロジンプロピレングリコールを投与しているが、データ内に糖尿病に関する疾患無し
- 「併用薬剤」で薬剤名がダパグリプロシンプロピレングコリールとあるが正しくはダパグリフロジンプロピレングリコールである
- 「患者日誌」に食欲が無かったとあるが、関連する「有害事象」の発現日と日誌の評価日が異なる
- 「患者日誌」に発熱で治験薬は 2 月まで服薬していない旨の情報があるが、1 月の「有害事象」の発熱の因果関係がありになっている
- 「併用薬剤」でロキソプロフェンナトリウムを有害事象の発熱に対して使用しているが、本症例は小児であり、ロキソプロフェンナトリウムは小児の適応無し
- 「併用薬剤」でクエン酸第一鉄ナトリウムを貧血に使用しているが、該当する有害事象は再生不良性貧血で適応外
- 「有害事象」に月経不順があるが、患者背景の性別が男性

このテストデータを Microsoft 365 Copilot のチャット機能（以下「Copilot」）に読み込ませ、以下のプロンプトでデータのレビューを指示した。

添付の JSON ファイルは臨床試験の症例データです。このデータには誤りが含まれている可能性があります。

臨床試験の専門医として、以下の観点から JSON データを確認し、回答してください。時系列を考慮に入れて回答することが非常に重要です。また、症例本人の主訴も重要な

情報です。

有害事象は、被験薬との因果関係を問わないあらゆる好ましくない事象や兆候を指します。

JSON データをレビューする観点

1. 臨床の経過から見て疑わしい情報

臨床の経過から見て疑わしい情報をすべて列挙し、その理由を説明します。

2. 有害事象として報告されている情報の整合性

報告された有害事象が、関連するその他の情報と整合が取れない場合、その理由を説明します。

3. 有害事象以外のデータにある安全性上の懸念

安全性上の懸念が認められる情報をすべて列挙し、それが有害事象として記載されているかどうかを評価します。

4. その他の不整合

上記以外に不整合が疑われるデータをすべて列挙し、その理由を説明します。

回答フォーマット

- **疑義事項**：理由

- **疑義事項**：理由

-

Copilot によるデータレビューの結果、2024 年 12 月時点では用意した疑義事項をすべて検知することはできなかったものの、一定数の疑義事項は検知できたため Copilot によるデータレビューが有用であることが示唆された。なお、本施行を含めて、後述のデータを用いて AI によるデータレビューを実行する際も、Excel 形式、CSV 形式、Word 形式、JSON 形式などの様々なファイルの形式でテストを行い、また、複数症例分のデータレビューを一度に実行する方法と 1 症例ずつ実行する方法を試した。その結果、コンピューターで処理しやすいシステムフレンドリーな JSON 形式で 1 症例ずつ実行する方法が最も検出精度が高いように思われたため、最終的には JSON 形式を用いて 1 症例ずつレビューを実行することとした。

5.4. 方法

5.4.1. レビューアーのロール設定

実際の臨床試験では、データの品質や妥当性を担保するために、さまざまな専門的視点からデータレビューが行われる。AI によるデータレビューにおいても専門性が発揮されるように、プロンプト設計において AI に対してロールを設定し、ロール毎に役割を定義した。

具体的には、以下の 3 つのロールとそれに対する役割を AI に割り当て、それぞれの専門性に応じた観点からデータレビューを行う構成とした。

1) メディカルモニター

医学的知識に基づいてデータを評価する役割を担っている。具体的には、治療経過や検査結果などに臨床的な矛盾がないか、医学的妥当性が保たれているかといった点を中心にレビューを行う。臨床試験の安全性／有効性評価や医学的解釈に関わる観点でのレビュー担当者である。

2) クリニカルデータマネージャー

データの整合性や一貫性、欠測値の有無、不自然なデータ変動など、主に「データの信頼性・完全性」に着目してチェックを行う。異なるデータセット間での不整合や入力ミスなど、データ処理上の問題を検出する観点でのレビュー担当者である。

3) 臨床開発モニター

試験実施計画書に記載された選択除外基準や、条件・基準・手順などと実データとの整合性を確認する。試験実施計画書に準拠していない手順や逸脱がないかを確認することで、試験の品質および適格性を確認する観点でのレビュー担当者である。

上記のような役割を AI に明確に設定することで、1つの視点に依存したレビューでは見落とされがちな点を多角的にカバーすることを期待した。また、ルールを明確に分けることで、AI が事実と異なる情報を提示するハルシネーションの発生も抑えられ、より正確かつ実用的なデータチェックが実現出来ると考えた。また、プロンプトの作成を始めた当初は1つのプロンプトで1つのルールのみを指定して AI にレビューを行わせていた。しかしながら、この手法では3つのルールで同じようなクエリが生成されることから、確認すべき結果が3倍に増え、クエリの重複確認などを追加で行う必要性が生じた。そこで、3つルールを1つのプロンプトに含めて、それぞれの観点でのレビューをするように指示することを試みたところ、それぞれ別のプロンプトでレビューを指示した時とほとんど変わらない結果が得られ、かつクエリの重複はほぼ発生しないことが分かったため、この手法を取ることにした。これにより、実際の臨床試験における3人で行うデータレビューを1つのプロンプトで再現することを可能とした。

5.4.2. 臨床試験データでの検討

次に実際の臨床試験データからの AI による疑義事項の検出、および検出された疑義事項に対する「図 5.1-1 疑義事項の重要度分類例」の分類が可能か検討を行った。

事前準備

臨床試験データは「3.選定した AI 活用の検討を実現するための試験計画およびデータの入手」に記載した公開されている試験実施計画書およびデータを活用した。また、AI による疑義事項の検出精度の確認のため、その臨床試験データのうち30症例に対して、一部の

データをあらかじめ疑義事項のあるデータに変換した⁹。

AIによる疑義事項の検出（1回目）

JSON 形式のデータで AI への読み込みを試みたところ、2025 年 3 月時点では以下の理由で GPT4o と Copilot では読み込みができなかった。

GPT4o : 1 ファイル 50MB まで読み込めるが、行数が多い場合は読み込めない

Copilot : 1 ファイル 1MB までしか読み込めない

そのため、JSON 形式のファイルを CSV に変換した上で複数回に分けて GPT4o にて読み込みを行ったところ、一部のファイルを除き、読み込みを行うことができた。一部のファイルの読み込みができなかった原因の究明には至っていないが、原因の一つとして 2 バイト文字がファイルに含まれていた点が考えられた。プロンプトは「5.3.予備的試行（テストデータでの検討）」で使用したものと同じものを使用した。

GPT4o で実行した結果、事前にタスクフォースで設定した疑義事項を含め、一定数の疑義事項を検出することができた。ただし、本テストを行った時点では GPT4o の無料版は 1 日の読み込みファイル数に制限があり、複数回の実行が難しい状況であった。

AIによる疑義事項の検出（2回目）

これまでの検討結果より、症例を絞って実行した方が精度が上がると考えられたため、疑義事項を設定した 30 症例に対してそれぞれ 1 例ずつ疑義事項の検出を試みた。なお、この時点で Gemini 2.5 Pro が試験的に利用可能であったため、Gemini 2.5 Pro を利用した¹⁰。

実行した結果、1 回目の実行結果よりもより精度の高い結果を得ることができた。実行結果の確認時に、タスクフォース側であらかじめ設定したテストデータの疑義事項に一部誤りが見つかったため、テストデータの修正およびプロンプトの見直しを行い、さらなる検出精度の向上を試みることにした。全体的なプロンプトの作成や見直しの経緯は巻末の「Appendix 1: プロンプトの作成」もあわせて参照いただきたい。

AIによる疑義事項の検出（3回目）

テストデータの修正およびプロンプトの見直しを行い、2025 年 4 月に、疑義事項を設定した 30 症例に対する Gemini 2.5 Pro を用いた疑義事項の検出および検出された疑義事項に対する「図 5.1-1 疑義事項の重要度分類例」の分類を行った。また、想定どおりに疑義事項が検出されなかった事例については、Gemini 2.5 Pro を用いて出力結果に対する見解を個別に生成させた。なお、本タスクフォースでは今回の 3 回目の検出結果をもとに考察を行

⁹ データの更新はプログラムで行った。更新内容は実行ログから確認可能であるが、更新後のデータも CSV ファイルに変換して公開した：

<https://github.com/Takumi173/JPMA2024TF1-2>

¹⁰ バージョンは 2.5-pro-exp-03-25 を使用した。

った。プロンプトも実行ごとに更新がされているため、最終的に実行した 3 回目の実行プロンプトのみを GitHub で公開している。

プロンプトの概略

本タスクフォースでは、作成した疑義事項検出のための実行プロンプトを Smart Clinical Data Validator と名付けた。

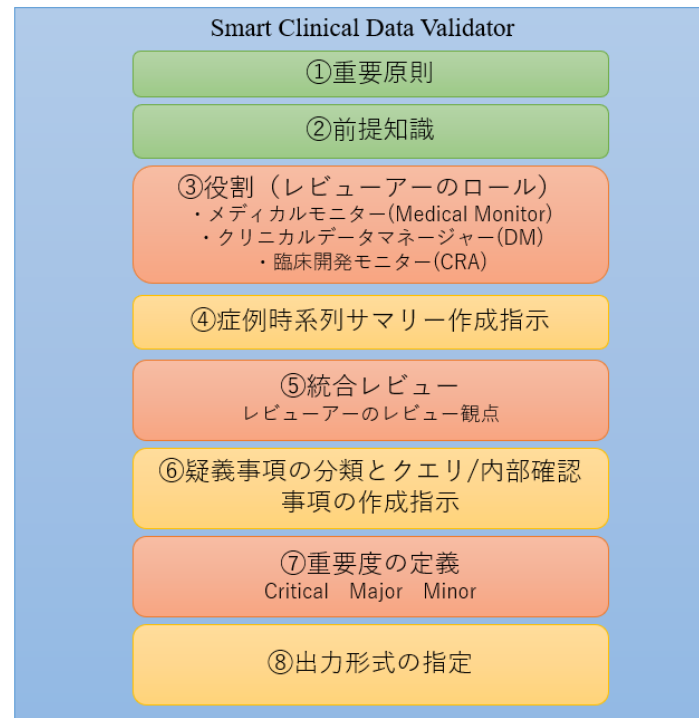


図 5.4.2-1 プロンプトの構成

図 5.4.2-1 は、Smart Clinical Data Validator の全体構成を概略的に示したものである。以下に各構成要素の内容は以下に述べる。詳細については、GitHub で公開しているプロンプトを参照されたい。

(1) 重要原則の設定

最初に、AI が出力を生成する際の基本的な判断基準として重要原則（図 5.4.2-1 ①）を明示的に設定した。人間のレビューアーが日常的に行っている判断の中には、経験や常識による「暗黙知」に基づいたものがある。本プロンプトの作成においては、重要原則に限らず、この「暗黙知」を明文化することで、結果の出力を我々の期待する人間のレビュー結果に近づけることを試みた。

(2) 専門知識の明示

AI が参照すべき専門知識領域を前提条件 (図 5.4.2-1 ②) として設定した。具体的には、医学的知識、データクリーニングの基礎、SDTM および Define.xml などデータ標準に関する知識を参照するよう指示した。

これにより、AI が臨床試験特有の用語、構造又は要件等を理解した上で出力できるようにした。

(3) レビューアーのロール設定

AI が複数の視点からレビューを行うために、レビューアーのロールを役割 (図 5.4.2-1 ③) に定義した。

設定したロールは、5.4.1 章「レビューアーのロール設定」で述べた通り、メディカルモニター、クリニカルデータマネージャーおよび臨床開発モニターの 3 種類である。

(4) 各ロールにおけるレビュー観点の明確化

統合レビュー (図 5.4.2-1 ⑤) では、各ロールのレビュー観点を明確に設定した。

具体的には、メディカルモニターは、医学的妥当性、安全性および有効性に関する判断を行うように設定し、クリニカルデータマネージャーでは、データ間の整合性を中心に確認を行うよう設定し、臨床開発モニターでは、プロトコル遵守の観点を重視し、逸脱等を検出する役割を担うよう設定した。

(5) 重要度分類

重要度の定義 (図 5.4.2-1 ⑦) では「図 5.1-1 疑義事項の重要度分類例」の基準に基づき、AI が「Critical」「Major」「Minor」に自動的に分類できるよう規定した。

(6) 出力形式の統一化

最後に、プロンプトの出力形式の指定 (図 5.4.2-1 ⑧) にて、症例時系列サマリー作成指示 (図 5.4.2-1 ④) および疑義事項の分類とクエリ/内部確認事項の作成指示 (図 5.4.2-1 ⑥) で設定した症例サマリーやクエリを人間が見やすく表示できるよう出力形式を明確に指示した。

5.5. 疑義事項の検出の結果

5.5.1. 疑義事項の検出結果

30 症例の中に設定した疑義事項は 38 件であり、38 件に対する Smart Clinical Data Validator による疑義事項の検出精度を確認した。正確に検出できた件数は 23 件であり、一部検出できた件数は 4 件であった。また 11 件は検出できなかった。疑義事項の詳細および Smart Clinical Data Validator の検出結果は以下「疑義事項の詳細および Smart Clinical Data Validator の検出結果(表 5.5.1-1)」のとおりである。

表 5.5.1-1 疑義事項の詳細および Smart Clinical Data Validator の検出結果

USUBJID	疑義事項の詳細	Smart Clinical Data Validator の検出結果
01-701-1015	(SAE の) Start Date が End Date よりも後	○
01-701-1034	重篤性は Non-SAE だが、疾患名 (MALIGNANT HYPERTENSION) より SAE が疑われる	△
01-701-1047	重篤性は Non-SAE だが、バイタル (Diastolic Blood Pressure=121,124、Systolic Blood Pressure=185,183) より SAE が疑われる	△
01-703-1258	AE (DIZZINESS) の Severity が SEVERE→MODERATE と推移しているが、Seriousness は N→Y と推移している	×
01-703-1258	SAE 発現時に試験薬を減量し、減量後に SAE が回復しもとの DOSE に戻しているが、Causality は NONE となっている (関連ありの疑い)	○
01-704-1009	有害事象に SAE 「Chronic kidney disease」があるが、短期間で回復している	○
01-704-1010	有害事象に SAE 「Diabetes mellitus」があるが、短期間で回復している	×
01-704-1017	重篤性は Non-SAE だが、疾患名 (Late effects of cerebral infarction) より SAE が疑われる	×
01-704-1017	重篤性は Non-SAE だが、疾患名 (Brain death) より SAE が疑われる。また転帰が「RECOVERED/RESOLVED」となっている	△
01-704-1017	SAE 「MYOCARDIAL INFARCTION」の転帰日が「Brain death」の後の日付になっている	×
01-701-1153	事象名、発現日、転帰日、重症度等が同一の AE 「INCREASED APPETITE」があるが、一方の「Action Taken with Study Treatment」が「DRUG WITHDRAWN」となっている	○
01-701-1180	事象名 (SUDDEN DEATH) に対して、Severity (MILD)、Serious (N; 非該当)、Outcome (RECOVERED/RESOLVED) となっており、齟齬がある	△
01-703-1096	「Inclusion Criteria」に「Males and postmenopausal females at least 50 years of age」とあるが、49 才	○

01-701-1023	Inclusion Criteria に合致しない ・ Hachinski Ischemic Scale score of ≤ 4	○
01-702-1082	「Inclusion Criteria」に「Males and postmenopausal females at least 50 years of age」とあるが、性別が Female であり、MH に「Premenstrual pain」あり	×
01-701-1028	現病歴に Exclusion Criteria となっているパーキンソン病あり	○
01-704-1008	「Exclusion Criteria」に [12] Diagnosis of serious neurological conditions, including d) Parkinson's disease とあるが、試験開始前よりパーキンソン病の 3 大症状があり Tremor in hands, legs Muscle stiffness Slowness of movement	×
01-703-1279 「	「Exclusion Criteria」の [14] A history within the last 5 years of the following: a) Schizophrenia に該当する可能性のある疾患 (SCHIZOPHRENIFORM DISORDERS) あり	○
01-703-1299	「Exclusion Criteria」の [14] A history within the last 5 years of the following: b) Bipolar Disease に該当する可能性のある疾患 (CYCLOTHYMIC DISORDER) あり	×
01-703-1335	「Exclusion Criteria」に [12] Diagnosis of serious neurological conditions, including e) Multiple sclerosis	○

	とあるが、有害事象に「Multiple sclerosis relapse」あり（基礎疾患に多発性硬化症がある可能性）	
01-703-1403	「Exclusion Criteria」に [12] Diagnosis of serious neurological conditions, including g) Myasthenia gravis とあるが、有害事象に「Myasthenia gravis aggravated」あり（基礎疾患に重症筋無力症がある可能性）	○
01-701-1097	「Exclusion Criteria」に「A history of syncope within the last 5 years」とあるが、MH に「LOSS OF CONSCIOUSNESS (PASSED OUT)」あり	○
01-703-1042	「Exclusion Criteria」の [18] A history within the last 5 years of a serious gastrointestinal disorder, including e) Hepatitis f) Cirrhosis of the liver に該当する可能性のある肝機能検査値異常あり	○
01-703-1076	「Exclusion Criteria」の [16b] Evidence from ECG recording at screening of any of the following conditions d) Second or third degree heart block unless treated with a pacemaker に該当する可能性のある現病歴あり	○
01-703-1086	「Exclusion Criteria」の [27b] Laboratory test values exceeding the Lilly Reference Range III for the patient's age in any of the following analytes: ↑↓ white blood cell count に該当する可能性のある検査値異常あり	○

	※Lilly Reference Range III の詳細は不明だが、WBC を概ね低値に設定した	
01-701-1111	基礎疾患に聴覚不全があるが、Visit3 の「COMMANDS」および「COMPREHENSION OF SPOKEN LANGUAGE」のスコアが低い（＝聴覚不全と矛盾）	×
01-701-1118	Visit10 で Marked improvement だが、Visit12 で Marked worsening となっており、一般的な経過としては不自然である	○
01-701-1148	過量投与の可能性	○
01-701-1148	試験実施計画書上ありえない投与量の記載	○
01-701-1363	（胃腸障害かつ Non-SAE の）Start Date が End Date よりも後	○
01-701-1383	「Increased blood pressure (Non-SAE)」の Outcome が RECOVERED/RESOLVED ですが、血圧が高値のまま	○
01-701-1387	「Increased blood pressure (Non-SAE)」が発現しているが、関連する血圧の Systolic Blood Pressure と Diastolic Blood Pressure の値が逆転	○
01-701-1111	基礎疾患に聴覚不全があるが、Visit4 の「COMMANDS」および「COMPREHENSION OF SPOKEN LANGUAGE」のスコアが低い	×
01-701-1118	Visit8 と Visit10 の評価日が逆転	○
01-701-1180	基礎疾患に DEPRESSION (ANXIETY) があるが、NPI-X の ANXIETY および DEPRESSION/DYSPHORIA がいずれも ABSENT である	×
01-701-1181	DISABILITY ASSESSMENT FOR DEMENTIA (DAD)の MEAL PREPARATION のみ、Visit がすすむにつれ明らかに悪化しており、他の評価項目と比べ不自然な変化である	×

01-701-1146	以下の併用禁止薬が使用されている。 ・抗うつ薬 ・抗精神病薬 ・抗不安薬 ・精神安定剤	○
01-701-1047	(Non-SAE の) Start Date が End Date よりも後	○

※「Smart Clinical Data Validator の検出結果」は「○＝正確に検出」「△＝一部検出」「×＝未検出」としている

<未検出事例の Gemini 2.5 Pro の見解>

「疑義事項の詳細および Smart Clinical Data Validator の検出結果(表 5.5.1-1)」について一部検出 (△) や未検出 (×) であった事例については Gemini 2.5 Pro の見解を生成している。これらの見解の内、特筆すべきと考える見解は以下のとおりである。

<未検出事例①: Seriousness が Non-SAE だが SAE が疑われる事例>

「疑義事項の詳細および Smart Clinical Data Validator の検出結果(表 5.5.1-1)」の内、以下の「一部検出および未検出事例 1 (表 5.5.1-2)」は Non-SAE と判定された有害事象について SAE が疑われる事例であるが、いずれも Smart Clinical Data Validator の検出結果は一部検出 (△) や未検出 (×) であった。

表 5.5.1-2 一部検出および未検出事例 1

USUBJID	疑義事項の詳細	Smart Clinical Data Validator の検出結果
01-701-1034	重篤性は Non-SAE だが、疾患名 (MALIGNANT HYPERTENSION) より SAE が疑われる	△
01-701-1047	HYPERTENSION の重篤性は Non-SAE だが、バイタル (Diastolic Blood Pressure=121,124、Systolic Blood Pressure=185,183) より SAE が疑われる	△
01-701-1180	事象名 (SUDDEN DEATH) に対して、Severity (MILD)、Serious (N ; 非該当)、Outcome (RECOVERED/RESOLVED) となっており、齟齬がある	△
01-704-1017	重篤性は Non-SAE だが、疾患名 (Late effects of cerebral infarction) より SAE が疑われる	×
01-704-1017	重篤性は Non-SAE だが、疾患名 (Brain death) より SAE が疑われる。また転帰が「RECOVERED/RESOLVED」となっている	△

一部検出 (△) や未検出 (×) と判定した理由について、Gemini 2.5 Pro の見解は、USUBJID が 01-701-1034、01-701-1047、01-704-1017 (Late effects of cerebral infarction (脳梗塞の後遺症) の事例) についてはいずれも試験実施計画書記載の Serious Adverse Event の定義 (Results in death~Is significant for any other reason) に該当するデータがいずれも “N (非該当)” となっている点から、試験実施施設にて Non-SAE と判定されたものと判断したため、不一致ではないとの見解であった。

しかし、USUBJID が 01-701-1180 (SUDDEN DEATH) および 01-704-1017 (Brain death) については、事象名と関連するデータの不一致に関して早急に確認する必要があるとの回答で、疑義事項を設定した 30 症例に対する疑義事項の検出時と、出力結果のデータ

に対する症例個別の見解生成時とで、Smart Clinical Data Validator の検出結果が異なっていた。

なお、Seriousness に関しては上述のとおりであるが、USUBJID が 01-701-1034、01-701-1047 の Severity に関して、いずれのデータも「MILD」であったが、01-701-1034 の Severity (MILD) については、Smart Clinical Data Validator が「Malignant Hypertension は通常、緊急治療を要する重篤な高血圧緊急症であり、臓器障害を伴うことが多い。バイタルサインデータでは実際に著しい高血圧が記録されており、Severity=Mild との評価は医学的に不適切である可能性が高い。」と指摘をしており、01-701-1047 の Severity (MILD) については「同日の立位時血圧は 185/121 mmHg、183/124 mmHg と著しい高値を示しており、医学的には中等度以上の評価が妥当と考えられる。本有害事象が試験中止理由とされていることから、Severity 評価の妥当性について確認が必要である。85 歳という高齢であり、顕著な血圧上昇は心血管イベントのリスクを高めるため、参加者の安全性確保の観点から重要である。」と、事象名と Severity の齟齬について適格かつ高度な指摘をしていた。

そのため、Smart Clinical Data Validator の Seriousness に対する評価は試験実施計画書記載の Serious Adverse Event の定義および関連するデータに強く影響を受けた可能性が高く、一方「Severity」については、試験実施計画書に具体的な定義の記載が無いため、一般的な判断基準のもと「Severity」の評価がなされたことが示唆された。

<未検出事例②：Inclusion / Exclusion Criteria への抵触が疑われる事例>

「疑義事項の詳細および Smart Clinical Data Validator の検出結果(表 5.5.1-1)」の表の内、以下の「一部検出および未検出事例 2 (表 5.5.1-3)」は Inclusion Criteria または Exclusion Criteria への抵触が疑われる事例であるが、いずれも Smart Clinical Data Validator の検出結果は未検出 (×) であった

表 5.5.1-3 一部検出および未検出事例 2

USUBJID	疑義事項の詳細	Smart Clinical Data Validator の検出結果
01-702-1082	「Inclusion Criteria」に「Males and postmenopausal females at least 50 years of age」とあるが、性別が Female であり、MH に「Premenstrual pain」あり	×
01-703-1299	「Exclusion Criteria」の [14] A history within the last 5 years of the following: b) Bipolar Disease に該当する可能性のある疾患 (CYCLOTHYMIC DISORDER) あり	×

01-704-1008	「Exclusion Criteria」に [12] Diagnosis of serious neurological conditions, including d) Parkinson's disease とあるが、試験開始前よりパーキンソン病の3大症状があり Tremor in hands, legs Muscle stiffness Slowness of movement	×
-------------	---	---

未検出（×）と判定した理由について、Gemini 2.5 Pro の見解は、それぞれ以下のとおりであった。

USUBJID が 01-702-1082 の「Premenstrual pain」（月経前の痛み）について、Gemini 2.5 Pro の見解は、「84 歳という年齢を考慮すると、通常は閉経後であると強く推測され、また臨床試験の適格性評価においてスクリーニング時に閉経状態が確認されているはずである。これより Medical History に記録された軽度の"Premenstrual pain"は、過去の既往歴である可能性が高く、Inclusion Criteria との整合性に大きな問題はないと考えられる。ただし、この既往歴が記録された背景（例：スクリーニング時の問診でどのように聴取・判断されたか）や、治験責任医師による適格性評価の詳細な根拠を確認することが、より確実な評価のためには望ましい」とのことであった。

USUBJID が 01-703-1299 の「CYCLOTHYMIC DISORDER」（気分循環性障害）について、Gemini 2.5 Pro の見解は、「医学的診断分類において「CYCLOTHYMIC DISORDER」と「Exclusion Criteria」に該当する「Bipolar Disease」（双極性障害）は関連性があるものの、異なる診断名として区別されます。試験実施計画書に「Bipolar Disease」と明確に記載している場合、その用語の厳密な解釈が優先されるべきです。したがって、被験者の既往である「CYCLOTHYMIC DISORDER」は、除外基準 14b で規定されている「Bipolar Disease」とは厳密には一致しないため、本被験者は除外基準 14b に該当しないと判断されます。ただし「CYCLOTHYMIC DISORDER」は「Bipolar Disease」のスペクトラムに含まれる疾患であり、臨床的な判断によっては、試験の安全性評価の観点から慎重な検討が必要となる場合があります。」とのことであった。

USUBJID が 01-704-1008 のパーキンソン病の3大症状について、Gemini 2.5 Pro は、「パーキンソン病の診断に関する記録は見当たらないものの、治験薬投与前からパーキンソン病の3大症状を呈しており、これは試験実施計画書で定められた除外基準「パーキンソン病を含む重篤な神経疾患の診断」に抵触する可能性が非常に高いと考えられ、この被験者の組み入れは、試験実施計画書違反であった可能性があり、その適格性には重大な疑義が生じます。」とのことであった。

他にも Inclusion Criteria または Exclusion Criteria への抵触が考えられるデータは事前にタスクフォース側で設定していたが、それらは「疑義事項の詳細および Smart Clinical Data Validator の検出結果(表 5.5.1-1)」のとおりいずれも適切に検知がされていた。

<未検出事例③：Seriousness と Severity に矛盾がある事例>

「疑義事項の詳細および Smart Clinical Data Validator の検出結果(表 5.5.1-1)」の内、以下「一部検出および未検出事例 3 (表 5.5.1-4)」は Seriousness と Severity の変化に矛盾がある事例であるが、Smart Clinical Data Validator の検出結果は未検出 (×) であった

表 5.5.1-4 一部検出および未検出事例 3

USUBJID	疑義事項の詳細	Smart Clinical Data Validator の検出結果
01-703-1258	AE (DIZZINESS) の Severity が SEVERE→MODERATE と推移しているが、Seriousness は N→Y と推移している	×

表 3.2.5.1-4 一部検出および未検出事例 3 (※該当データ)

Term	Severity	Serious	Outcome	Start Date	End Date
DIZZINESS	SEVERE	N	NOT RECOVERED/NOT RESOLVED	2012-08-01	2012-10-01
DIZZINESS	MODERATE	Y	RECOVERED/RESOLVED	2012-10-02	2012-12-31

未検出 (×) と判定した理由について、Gemini 2.5 Pro の見解は、以下のとおりであった。

「臨床試験における有害事象の評価において、Severity と Seriousness は異なる概念であることを踏まえ、DIZZINESS の経過における Severity と Seriousness の変化は以下のように解釈できる。2012 年 8 月 1 日に発生しためまいは、症状の程度としては「SEVERE (重度)」と評価された。しかし、この時点では、試験実施計画書で定義されている「重篤な有害事象」の基準 (死亡、生命を脅かす、入院、永続的な障害など) には該当しなかったため、重篤性は「N (非重篤)」と判断された。

最初の DIZZINESS (浮動性めまい) が一旦終了した翌日の 2012 年 10 月 2 日に、再度 (または継続して) DIZZINESS が発生した。この時の症状の程度は「MODERATE (中等度)」であり、最初のイベント時よりも軽減していた。しかし、この 2 番目のイベントに関連して、参加者は「入院または入院期間の延長」を必要とした。試験実施計画書の定義に基づき、「治療のための入院または入院期間の延長が必要となるもの」は重篤な有害事象に該当するため、このイベントの重篤性は「Y (重篤)」と判断された。これは重症度と重篤性がそれぞれ独立した基準で評価されるために起こりうる状況である。」とのことであった。

5.5.2. 重要度分類の判定結果

正確に検出できた 23 件の疑義事項に対して、Smart Clinical Data Validator が判定した重要度分類は以下の「Smart Clinical Data Validator が判定した重要度分類(表 5.5.2-1)」のとおりであった。

表 5.5.2-1 Smart Clinical Data Validator が判定した重要度分類

USUBJID	疑義事項の詳細	重要度分類
01-701-1015	(SAE の) Start Date が End Date よりも後	Major
01-701-1023	Inclusion Criteria に合致しない ・ Hachinski Ischemic Scale score of ≤ 4	Major
01-701-1028	現病歴に Exclusion Criteria となっているパーキンソン病あり	Major
01-701-1047	(Non-SAE の) Start Date が End Date よりも後	Minor
01-701-1097	「Exclusion Criteria」に「A history of syncope within the last 5 years」とあるが、現病歴に「LOSS OF CONSCIOUSNESS (PASSED OUT)」あり	Major
01-701-1118	Visit10 で Marked improvement だが、Visit12 で Marked worsening となっており、一般的な経過としては不自然である	Major
01-701-1118	Visit8 と Visit10 の評価日が逆転	Major
01-701-1146	以下の併用禁止薬が使用されている。 ・ 抗うつ薬 ・ 抗精神病薬 ・ 抗不安薬 ・ 精神安定剤	Major
01-701-1148	過量投与 (216 mg) の可能性	Critical
01-701-1148	プロトコル上ありえない投与量 (82 mg) の記載	Minor
01-701-1153	事象名、発現日、転帰日、重症度等が同一の AE 「INCREASED APPETITE」があるが、一方の「Action Taken with Study Treatment」が「DRUG WITHDRAWN」となっている	Major
01-701-1363	(胃腸障害かつ Non-SAE の) Start Date が End Date よりも後	Major

01-701-1383	「Increased blood pressure (Non-SAE)」の Outcome が RECOVERED/RESOLVED ですが、血圧が高値のまま	Major
01-701-1387	「Increased blood pressure (Non-SAE)」が発現しているが、関連する血圧の Systolic Blood Pressure と Diastolic Blood Pressure の値が逆転	Minor
01-703-1042	「Exclusion Criteria」の [18] A history within the last 5 years of a serious gastrointestinal disorder, including e) Hepatitis f) Cirrhosis of the liver に該当する可能性のある肝機能検査値異常あり	Major
01-703-1076	「Exclusion Criteria」の [16b] Evidence from ECG recording at screening of any of the following conditions d) Second or third degree heart block unless treated with a pacemaker に該当する可能性のある現病歴あり	Major
01-703-1086	「Exclusion Criteria」の [27b] Laboratory test values exceeding the Lilly Reference Range III for the patient's age in any of the following analytes: ↑↓ white blood cell count に該当する可能性のある検査値異常あり ※Lilly Reference Range III の詳細は不明だが、WBC を概ね低値に設定した	Critical
01-703-1096	「Inclusion Criteria」に「Males and postmenopausal females at least 50 years of age」とあるが、49 才	Major
01-703-1258	SAE 発現時に試験薬を減量し、減量後に SAE が回復しもとの DOSE に戻しているが、Causality は NONE となっている（関連ありの疑い）	Major

01-703-1279	<p>[14] A history within the last 5 years of the following:</p> <p>a) Schizophrenia</p> <p>に該当する可能性のある疾患 (SCHIZOPHRENIFORM DISORDERS) あり</p>	Major
01-703-1335	<p>「Exclusion Criteria」に</p> <p>[12] Diagnosis of serious neurological conditions, including</p> <p>e) Multiple sclerosis</p> <p>とあるが、有害事象に「Multiple sclerosis relapse」あり（基礎疾患に多発性硬化症がある可能性）</p>	Critical
01-703-1403	<p>「Exclusion Criteria」に</p> <p>[12] Diagnosis of serious neurological conditions, including</p> <p>g) Myasthenia gravis</p> <p>とあるが、有害事象に「Myasthenia gravis aggravated」あり（基礎疾患に重症筋無力症がある可能性）</p>	Critical
01-704-1009	有害事象に SAE 「Chronic kidney disease」があるが、短期間で回復している	Critical

上記「Smart Clinical Data Validator が判定した重要度分類(表 5.5.2-1)」のとおり、23 件中「Critical」と判定された疑義事項は 5 件、「Major」は 15 件、「Minor」は 3 件であった。有害事象データや治験薬の投与量など参加者の安全性に直接かかわる疑義事項が「Critical」と判定される傾向にあり、Start Date と End Date の逆転や、血圧の Systolic Blood Pressure と Diastolic Blood Pressure の値が逆転など単純なデータの入力ミスが疑われるような事例は「Minor」と判定される傾向があった。

ただし、単純なデータの入力ミスが疑われる事例であっても、たとえば有害事象の Start Date と End Date の逆転はあらかじめ 3 例に対して設定していたが、01-701-1015 の「DIARRHOEA (SAE)」は Major、01-701-1047 の「UPPER RESPIRATORY TRACT INFECTION (Non-SAE)」は「Minor」と判定され、同じく Non-SAE の 01-701-1363 の「NAUSEA (Non-SAE)」は「Major」と判定されていた。これは、同じパターンの入力ミスであっても、SAE に関わるものや、治験薬であるムスカリン受容体作動薬の特徴である消化管関連の有害事象に関わるものを「Major」と判定し、有害事象自体が試験の結果への影響が大きいと考えられるものは「Minor」と判定されたと考えられ、試験への影響範囲や薬剤特性などを踏まえた高度な判定結果である可能性が推察された。

また、治験薬の投与量について 01-701-1148 に「216 mg」と「82 mg」のデータをあらかじめ設定したところ、「216 mg」は「Critical」、「82 mg」は「Minor」と判定された。試験実施計画書より治験薬は 1 パッチあたり「54 mg」もしくは「27 mg」であるため、「54 mg」製剤 4 枚分または「27 mg」製剤 8 枚分の「216 mg」は投与量としては可能性のある値であり、かつ試験実施計画書から大きく逸脱した値であるため、参加者の安全性に影響があるとの判断で「Critical」となり、一方「82 mg」は治験薬の用量を考慮するとあり得ない値であるため、入力ミスとの判定で「Minor」と判定された可能性も考えられる。こちらもデータだけでなく試験実施計画書の内容も考慮された判定結果である可能性も考えられた。

一方 Inclusion / Exclusion Criteria に関して、01-701-1023、01-701-1028、01-701-1097、01-701-1146、01-703-1042、01-703-1076、01-703-1086、01-703-1096、01-703-1279 の 9 例にあらかじめ関連する疑義事項を設定したところ、01-703-1086 のみ「Critical」の判定で、他はすべて「Major」の判定であった。プロンプトの「重要度の定義 ; Critical」の「具体的な状況例」に「主要な選択/除外基準の明確な違反で、参加者への重大なリスクや主要評価項目の信頼性への重大な影響が懸念されるもの。」と記載していたため、上記の 9 例はいずれも「Critical」と判定されてもおかしくはないが、01-703-1086 のみ「Critical」と判定された根拠は不明であった。重要度の定義はプロンプトに詳細に記載したものの、明確な判定基準があるわけではないため、このような結果となったものと考えられる。明確な基準が無いものを様々な情報から判定できるのは AI の大きな特徴のひとつではあるが、一方で今回のケースのように実行の都度判定結果が異なる可能性が高いと推察されるため、実際に AI によるデータレビューを行う際にはたとえば複数回実行する、適宜プロンプトや基

準を見直す等の対応が必要になると考えられる。

そのほか特筆すべき事例

これまで述べた結果は、タスクフォースにて事前に設定した 38 件の疑義事項に対する Smart Clinical Data Validator による検討結果であるが、タスクフォースで設定した疑義事項以外にも想定していなかった新たな疑義事項が Smart Clinical Data Validator により検出された。これらの中には、医学的観点から見ても重要と考えられる内容が含まれており、AI によるレビューの有効性が示唆された。検出されたものの中から、注目すべきケースを 3 つのパターンに分けて紹介する。

1) 医学的知見から安全性に重要な指摘

医学的知識を活用し、主に有害事象や、検査結果（臨床検査、バイタルサイン等）の情報から、参加者の安全性に関わる重大な疑義事項が検出された。これらは安全性の観点から試験中止判断も検討する必要がある、Criticality は「Critical」や「Major」と判断されている。

具体的な指摘内容は、“有害事象「ATRIOVENTRICULAR BLOCK SECOND DEGREE」(第二度房室ブロック) (Mild) が報告されている。本剤 (Xanomeline) はコリン作動薬であり心伝導系への影響が懸念される薬剤クラスであるため、プラセボ群であっても心血管系の安全性評価は重要である。発現時の詳細な心電図所見、関連症状の有無、実施された処置、および試験中止判断への関与について確認が必要。プロトコル 3.9.4 章では心血管系の安全性モニタリングが強調されている”である。

この疑義事項は、AI が治験薬情報として、コリン作動薬を試験実施計画書から読み取り、コリン作動薬の特性である心疾患への懸念を医学的な情報から指摘している。試験中止判断を含めた確認を実施するように助言しており、薬剤の知識や医学的知識を駆使して高度な分析を行っていると感じた。

2) 臨床経過に伴う臨床上の矛盾

試験の時間的経過に伴い、検査または症状の変化から臨床的な矛盾や安全性上の懸念を自動的に抽出している。

例えば、“ベースライン時から一貫して赤血球恒数 (MCH, MCV) が高値を示しており、試験後半には形態異常 (ANISO, MACROCY, POLYCHR) も出現している。これは大球性貧血を示唆する所見であり、臨床的な評価（原因検索、関連症状の有無）および治験薬との関連性評価が必要。患者の安全性に関わる可能性がある”という指摘では、単一時点での検査値の異常を捉えるのではなく、時間を追って複数の指標の変化を統合的に解析しており、AI が時系列的なデータ把握を行っていると考えられる。特に本症例は高齢者であり、血球異常が見られることは決して稀ではないものの、その異常が試験期間中にどのように推移

したかという臨床的経過を詳細に捉え、安全性上のリスクとして浮き彫りにしている。

現在の試験関連業務では、このような時系列での変化は、可視化ツールなどを使用して目視で確認していることが多いが、複数の項目の時間経過に伴う確認は、目視での実施は見落としのリスクがある。AI を使用し、時間軸に沿った異常変化の兆候を機械的かつ網羅的に捉えることが出来る様になれば、今後のデータレビューの質が向上されるものと期待できる。

3) 複数のデータセットにまたがる整合性チェック

単一のデータセットにとどまらず、3つ以上のデータセットのデータから横断的に照合・統合し、整合性の観点から安全性上の問題の指摘を行っている事例である。具体的な例として、“治験薬投与開始後、皮膚関連の有害事象（紅斑、全身性そう痒症、適用部位小水疱、適用部位そう痒症）が頻発している。特に適用部位そう痒症は持続し、Day 141 には軽度から中等度に悪化している。Day 57 および Day 85 には好酸球増多（基準値上限の約 1.5 倍）も認められており、治験薬に対するアレルギー反応の可能性が示唆される。Hydrocortisone 外用薬が使用されているが、適用部位そう痒症は持続している。プロトコル 3.9.3.4 には皮膚反応と好酸球増多に関する記載があり、既知の事象と考えられるが、持続・悪化している点、および好酸球増多のフォローアップ状況について確認が必要”という疑義事項が検出された。

この指摘では、有害事象ドメイン、臨床検査ドメイン（Eosinophils）、併用薬ドメイン（Hydrocortisone）のデータ、試験実施計画書 3.9.3.4 章および、薬剤性皮膚反応および好酸球増多に関する一般的な医学知識により総合的分析し、指摘している。

データマネージャーによる目視チェックでは、複数のドメインにまたがる異常を網羅的かつ時系列的に評価することは難しい事項だと考える。このような多変量的・多ソース的な観点からの分析においては、AI を導入することは重要な医学的懸念を見逃さず、効率的かつ網羅的なデータレビューを実現する上で非常に有益であると感じた。

5.6. 考察

5.6.1. 人間のインプットの重要性

通常 Copilot や Gemini のような AI は、人間からのインプットをインターネットから収集された膨大なテキストデータ等より推論しアウトプットを作成する。一般的に人間からのインプットが具体的であればあるほどアウトプットの精度も上がると考えられるが、一方で今回アウトプットが、人間がインプットした内容に強く影響されすぎて、齟齬を検出できないケースも見受けられた。たとえば、上述の<未検出事例①>; Seriousness が Non-SAE だが SAE が疑われる事例>の USUBJID が 01-701-1034、01-701-1047、01-704-1017 のケースは、Serious が疑われるような有害事象が、試験実施計画書に記載されている Serious Adverse Event の定義と、取り込んだデータの当該有害事象の Serious Adverse Event の

関連データである「Results in death」～「Is significant for any other reason」がいずれも「N（非該当）」となっている点から、試験実施施設にて Non-**SAE** と判定されたと判断し、疑義事項と考えなかったとの **Smart Clinical Data Validator** の見解であったが、既存のデータマネージャーによる目視チェックでは、同様のケースは **Seriousness** の妥当性を再確認するクエリが上がるケースも多いのではないかと考えられる。しかし一方で試験実施計画書に明確な判定基準の記載がない「**Severity**」に関しては当該有害事象のデータが「**MILD**」となっている点を **Smart Clinical Data Validator** が疑義として指摘をあげていた。

これらは人間のインプットがより具体的であったがために、アウトプットに対するバイアスや制限がかかった一例と考えられ、人間が **AI** を利用する際はインプットがアウトプットに対する影響を理解し、慎重にインプットする情報を選択し、時にはあまり多くの情報をインプットせず、**AI** に一般的な見解を求めるような運用も検討する必要があると考えられた。

ただし、同時に<未検出事例①>；**Seriousness** が Non-**SAE** だが **SAE** が疑われる事例>の **USUBJID** が 01-701-1180 および 01-704-1017 のケースは、**Smart Clinical Data Validator** による疑義事項の検出時には検出ができなかったものの、未検出に対する **Smart Clinical Data Validator** の見解を求めた際には不一致であったと判断された、つまり同じ試験実施計画書およびデータを読み込ませても、プロンプトの記載によりアウトプットが変動したケースである。またそのほかの検討結果からも仮に同じプロンプトを複数回実行したとしても、結果が変動性することが分かる。実運用を想定した場合、これらは一貫性や再現性に疑問が残ることになるため、変動幅をできる限り小さくするために、**AI** のパラメータを調整したり、より具体的で明確なプロンプトに仕上げるといった対応で制御していくことが必要である。また、同じプロンプトを複数回実行し、結果を比較したり統合したりすることで、より信頼性の高いアウトプットが得られるようなプロセスを考えることも重要である。

5.6.2. Human-in-the-Loop の重要性

今回 **Copilot** や **Gemini** は、膨大な **SDTM** 形式の **csv** ファイルおよび **DATASET-JSON** 形式の **JSON** ファイルを、それらが **SDTM** 形式であることも含め短時間で推論できる能力を示した。また自然言語処理により試験実施計画書の内容を推論し、取り込んだデータから参加者の背景情報や併用薬等の試験への影響を考慮に入れつつ、医学的な観点も加えた高度な疑義事項を検出することができた。これらの疑義事項の迅速な検出と報告は、治験や製造販売後調査等における参加者の保護に大きく寄与する可能性を秘めている。例えば、特定の背景疾患を持つ参加者に対する副作用リスクや、特定の併用薬剤に対する治験薬との相互作用等の可能性を **AI** が早期に特定し試験担当者にフィードバックすることで、試験データの質の向上に加え、何より参加者により早く適切な予防措置を検討し講じることができる可能性があると考えられる。これらは従来の手法でも実現可能ではあったが、多くの時間

と労力が必要であったところ、AI の登場によって 2・3 分で結果が得られる可能性が示された。これは膨大なデータを扱う上で非常に有効な手段になりうると考える。

しかし今回の検討結果は、同時に AI による検出精度の限界や品質の限界を感じる部分もあった。たとえば、「5.3.予備的試行（テストデータでの検討）」に記載した、10 症例分のテストデータの疑義事項の検出を一度に実行するよりも 1 症例ずつ症例番号を指定して実行する方が、検出精度が上がることを示唆されたケースや、「5.5.1. 疑義事項の検出結果」に記載した「未検出事例」のようなケースである。また、今回は神経領域の疾患に対する試験実施計画書やデータを用いた検討であったが、希少疾患などの情報の少ない疾患領域で同等の結果が得られるとは限らない。臨床試験のデータは日々更新されており、多くの場合は不完全な状態のデータを AI に読み込ませて結果を得ることになる。そのような状況において AI から求める回答を得るためには、AI を利用するための目的を明確にし、それを明確なプロンプトに落とし込み、AI に正しく指示をする能力が求められてくるものと考えられる。AI の「賢さ」は日進月歩で進化しているが、それを適切に扱えるか否かは人の指示にかかっているのである。

人間のインプットの重要性に加え、アウトプットの吟味も同様に重要である。インプットの質については上で述べたとおりだが、一方で、ハルシネーションが含まれる可能性を理解せず、AI のアウトプットをそのまま受け入れることは危険であり、かえって業務の質の低下や非効率化を招く要因となりうる。AI のアウトプットはあくまでも「参考情報」であり、最終判断は人間が責任を持って行うべきであると考え。人間がアウトプットを吟味し、必要に応じて裏付けを取る姿勢が求められる。

AI は人間の代替ではなく、業務全体の中で補完的な役割を果たす存在であり、重要なのは AI と競うのではなく、AI の仕組みや特徴を理解し、人間として押さえておくべきポイントがどこなのかをよく考え、AI との役割分担を考えることである。例えば、創造的な判断や倫理的な意思決定は人間が担い、AI には大量データの処理やパターン抽出を任せる、といった協働の形が理想であると考え。このような共存の姿勢は、業務効率化だけでなく、結果として人間の知識の深化や仕事人としての成長にもつながる。

今回の本タスクフォースの検討結果から見ても AI による疑義事項の検出や重要度分類の能力は、人間のモニタリング担当者やデータマネジメント担当者に匹敵するレベルに近づいていると感じる。しかしながら、一部の結果においては検出精度や検出結果にバイアスが加わっていると思われる点があり、また、明らかな疑義事項が見落とされるケースもあった。そして、時にはハルシネーションが含まれる可能性があること、最終的な倫理的な判断や結果に対する説明責任が人間にあることより、現時点ではすべての工程を AI に任せるのは難しく、AI の出力した結果を最終的に人が確認する、いわゆる **Human-in-the-Loop** のプロセスによって結果の精度や品質を担保し、透明性を持たせた上で人間が説明責任を果たしていくことが重要であると考え。

5.6.3. AI によるデータレビューと今後の展望

本プロトタイプリングでは、AI がメディカルモニター、クリニカルデータマネージャー、臨床モニターといった異なる専門的視点を一括して担うことが可能であるという一定の手応えを得ることが出来た。

従来、人間により実施されてきたこれらのレビュー業務は、医学的知識、臨床的判断力、試験実施計画書遵守の理解、さらにはデータ構造の理解といった多様な能力を必要とし、専門人材による目視での確認がされてきた。

今回の試行では、一部において未検出や過検出といった課題は残ったものの、AI が高度な言語認識能力と医学的知識を活用し、一定レベル以上の精度で疑義事項の検出ができたことは有益であった。とりわけ、AI が標準的な治療ガイドラインや文献に基づいた疑義事項の検出を短時間で行えるという点は、目視による確認に依存してきた従来の作業プロセスに対する有効な代替手段となる可能性がある。

さらに、AI を活用することで、専任のメディカルモニターを配置できない企業や小さな組織による試験実施体制でも、一定水準のレビュー品質を担保できる可能性がある点は、非常に有益である。

ただし、これまで述べたように現時点では AI の判断が必ずしも完璧ではなく、最終的には人間によるレビューが必要となる。特に、文脈依存的な情報や試験実施計画書の微妙な解釈といった部分では、AI の出力を補完する人による判断が不可欠となる。

この問題については、AI の判断根拠を AI に明示させ、レビューワーが納得可能な説明性を持たせることも重要である。

今回試行した AI と RBA を融合したデータレビューでは、従来のプロセスに比べて飛躍的な効率化と高速化をもたらす可能性を秘めており、臨床試験データの品質管理業務の大きな転換を促すことを期待している。

6. まとめ

今回実施した 2 つのプロトタイプリングでは、いずれも実用化に向けた可能性を十分に感じさせる成果を得ることができた。特に注目すべき点は、ハルシネーションを抑制し、生成 AI の回答の信頼性を向上させるためには、現時点ではプロンプトの設計が極めて重要であるという知見である。プロンプトの記載に際しては、質問の目的や期待される回答の範囲をあらかじめ明確にし、曖昧さを極力排除することにより、AI の出力が焦点を逸らさず、一貫性と精度の高い内容となることが確認された。これは、他業務においても生成 AI を活用する際には、欠かせない重要なポイントである。さらに、AI の進化は非常に速く、その時々で最適な活用方法を見極めることが重要である。技術の進歩に伴い、プロンプト設計のベストプラクティスや AI の特性も変化する可能性があるため、固定的な方法に依存せず、常に最新の知見を取り入れながら運用を改善していく姿勢が求められる。

今回のプロトタイプリングにおいて当初直面した現実的な大きな課題として、試験実施計

画書や症例データなどの社外秘な情報を、外部のオープンな生成 AI に直接読み込ませることが、情報漏えいリスクや社内規定の観点から困難であるという点が挙げられた。しかし本タスクフォースの検討中にセキュアな環境下で構築・運用されるクローズドな生成 AI プラットフォームも登場してきて [16] [17]、現場ニーズに即した実用が可能となることが期待される。

本検討を通じて AI が人間の役割を完全に置き換えるのではなく、適切に役割分担された“協働体制”を構築することが、実務的かつ現実的な活用方法であると示唆された。つまり、AI が一次段階で広範な情報の中から候補となる事象を迅速かつ網羅的に抽出し、その後に人間が専門的な観点からその妥当性を精査するという“協働型フロー”を確立することで、業務の効率性と適格性の両立が実現されることが考えられる。

この AI による分析力と人間による判断力を組み合わせた“協働型フロー”を築くことで、試験の計画段階から実施中にかけて、リスクを早期に検出し、迅速かつ的確に対応策を講じることが可能であると考えられる。

これまで述べてきたように AI は膨大な情報を高速で処理し、潜在的なリスクや異常パターンを見逃すことなく網羅的に抽出することが可能であることが示唆された。したがって、AI は治験業務全体における更なる効率化と品質確保の両立を実現するためのゲームチェンジャーとしての可能性を感じている。

7. おわりに

本書の表紙に記載されているタスクフォース名「AI を活用したデータマネジメントの変革」からもお分かりいただけるように、筆者らが当初目指していたのは、「AI を活用して従来のデータマネジメントに変革をもたらすこと」であった。したがって、当初 AI の適用対象として想定していたものは、データマネジメントにおける各種プロセスやタスクであった。

しかしその後、ICH E8 (R1)および ICH E6 (R3)で重視されている CTQ という概念に強く影響を受けたことで、我々の関心は次第に、臨床試験の結論の信頼性や参加者の安全性に重大な影響を及ぼし得る要因、ならびにそうした要因に関連するデータそのものやデータから得られるシグナルへと移っていった。そのために従来のデータマネジメントという概念を超えて、AI が質の管理にどのように貢献できるかに焦点を当てることになった。

本研究では、臨床試験における生成 AI の活用可能性を探るため、2つのプロトタイプングを通じて、実務に即した検討を行った。その結果、AI が試験関連業務において実用化に向けた十分な可能性を秘めていることが強く示唆された。特にプロンプト設計の工夫が AI の出力の精度と信頼性に大きく寄与するということがあらためて認識された。

また、本研究は、AI と人間の役割分担による“協働体制”の有効性を示した点でも意義深い。AI が膨大な情報を高速で処理し、候補を提示する段階を担い、人間が専門的な観点からその妥当性を精査することで、業務の効率性と適格性の両立が可能となる。このような協

働体制は、CTQ 要因および Risk-proportionate Approach を意識した臨床試験の実施計画およびその実施において、CTQ 要因の特定、それに関するリスクに見合った対応を支援するものとして、実務への貢献が期待できるであろう。

機密性の高い情報の AI への入力に関する懸念は、セキュアな環境下で運用可能な特定のクローズド AI プラットフォームサービスは提供されてきており、今後は機密性の高い情報を扱う実務への AI の適用可能性が大きく広がっていくと考えられる。

今後の展望として、今回使用した CtQF Facilitator のような対話型 AI は単なるチャットボットとしての役割を超え、より高度な分析力と知識を備えたエージェントへと進化していくものと思われる。将来、より成長した AI は、臨床試験の開発チームにおいて、広範で深い知識を瞬時に提供できる”チームメンバー”の一人としての役割を果たし、専門家と協働しながら臨床試験を支援する存在となるだろう。ただし、本書で繰り返し述べてきたように、今後とも責任ある AI の利用を確保するためには、人間による適切な介在が不可欠であることを忘れてはならない。AI という”チームメンバー”が誕生することによって起こり得る医薬品開発における様々な「変革」を実現すべく、現状の非効率や非合理に対する問題意識を強く持ち、それに対する解決策を導き出すことが人間の役割と言えるのではないだろうか。こうした未来像を見据え、AI と人間が具体的にどのようにしてお互いの強みを活かし合っていくかが、今後の重要な研究テーマの一つにさえなるように思われる。

なお、「はじめに」でも述べたとおり、AI 関連技術は驚異的な速度で進展しており、今回のタスクフォースで利用した Copilot や Gemini もいずれ最新の技術に置き換えられるだろう。そのため、今回作成また公開した出力結果、プログラム、プロンプト等は、早晚陳腐化し、その価値を減じることになるだろう。本成果物において読者に参考としていただきたいのは、出力結果やプログラム、プロンプトそのものではなく、むしろ、現状に対して強い問題意識を持ち、その問題をその時点で利用可能な AI 技術でいかに解決し得るかを考え、悩み、可能な解決策を試行するという取り組み方ではないかと考えている。本検討が、成長していく AI を臨床試験業務に取り入れていくことの有用性を示した実践的な一歩となったとしたら望外の喜びである。AI による「変革」により、医薬品開発における質の管理と大幅な業務効率化が実現し、アンメットメディカルニーズに対する新たな治療選択肢が早期に患者の下へ届けられることを心から願っている。

参考文献

- [1] 経済産業省 製造産業局 自動車課 モビリティ DX 室, “「モビリティ DX 戦略」2025 年のアップデート,” 6 2025. [オンライン]. Available: <https://www.fsa.go.jp/common/about/research/20240329/20240329.html>.
- [2] 農林水産省・経済産業省, “省力化投資促進プランー小売業ー,” 6 2025. [オンライン]. Available: https://www.cas.go.jp/jp/seisaku/atarashii_sihonsyugi/shouryokukatousi/03.pdf.
- [3] 金融庁, “AI 技術を活用した経営改善支援の効率化に向けた調査・研究」に係る最終報告書等の公表について,” 3 2024. [オンライン]. Available: <https://www.fsa.go.jp/common/about/research/20240329/20240329.html>.
- [4] 内閣府, “田中構成員提出資料（進化する日本での AI の利活用）,” 5 2024. [オンライン]. Available: https://www8.cao.go.jp/cstp/ai/ai_senryaku/9kai/9kai.html.
- [5] 日本製薬工業協会 データサイエンス部会, “データマネジメントにおける Artificial Intelligence の活用 ～これから始める AI ～,” 5 月 2023. [オンライン]. Available: https://www.jpma.or.jp/information/evaluation/results/allotment/g75una00000024me-att/DS_202305_2022TF1_1_AI_DM.pdf.
- [6] Saama, “Accelerate Your Trials Intelligently,” Saama, 2025. [オンライン]. Available: <https://www.saama.com/>.
- [7] Medidata, “AI Everywhere,” Medidata, 2025. [オンライン]. Available: <https://www.medidata.com/en/clinical-trial-products/medidata-ai/#data-management-cl>.
- [8] 日本製薬工業協会 データサイエンス部会, “データマネジメント業務における AI 活用検討のためのモデル試作,” 6 2024. [オンライン]. Available: https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202406_AI_Prototype.html.
- [9] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, “GENERAL CONSIDERATIONS FOR CLINICAL STUDIES E8(R1),” 10 2021. [オンライン]. Available: https://database.ich.org/sites/default/files/ICH_E8-R1_Guideline_Step4_2021_1006.pdf.
- [10] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, “GUIDELINE FOR GOOD CLINICAL PRACTICE E6(R3),” 1 2025. [オンライン]. Available: https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf.

- [11] 日本製薬工業協会 医薬品評価委員会 臨床評価部会, “クオリティ・バイ・デザインの考え方に基づく効果的・効率的な臨床試験の質の確保～計画から実施まで一貫した品質マネジメントを目指して～,” 7 2023. [オンライン]. Available: https://www.jpma.or.jp/information/evaluation/results/allotment/g75una00000035d5-att/CL_202307_TF2.pdf.
- [12] 小宮山靖, “GCP リノベーションが目指すもの,” 20 7 2024. [オンライン]. Available: https://www.qms-amed.jp/material_d/2024_007.pdf.
- [13] Society for Clinical Data Management, “The Evolution of Clinical Data Management into Clinical Data Science,” 9 2022. [オンライン]. Available: <https://scdm.org/wp-content/uploads/2024/03/SCDM-Position-Paper-Evolution-into-Clinical-to-Data-Science-V9.0.pdf>.
- [14] C. Rich Davies, “The Risk-Based Data Management (RBDM) Revolution,” [オンライン]. Available: [https://www.clinicalresearchnewsonline.com/news/2024/05/24/the-risk-based-data-management-\(rbdm\)-revolution](https://www.clinicalresearchnewsonline.com/news/2024/05/24/the-risk-based-data-management-(rbdm)-revolution).
- [15] V. S. A. V. A. S. B. S. G. A. W. C. G. N. & E. K. Dana Brin, “Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments,” 2023. [オンライン]. Available: <https://www.nature.com/articles/s41598-023-43436-9>.
- [16] Copilot, “Data, Privacy, and Security for Microsoft 365 Copilot,” 8 2025. [オンライン]. Available: <https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy>.
- [17] Google, “Generative AI in Google Workspace Privacy Hub,” 8 2025. [オンライン]. Available: <https://support.google.com/a/answer/15706919?hl=ja>.

タスクフォースメンバー（執筆者）

会社名	氏名	役割
中外製薬株式会社	淡路 直人	推進委員
ファイザーR&D 合同会社	三沢 秀敏	推進委員
EA ファーマ株式会社	橘田 由香里	メンバー
ノバルティス ファーマ株式会社	井上 祐介	メンバー
ファイザーR&D 合同会社	田中 拓海	メンバー
アステラス製薬株式会社	柴田 奈津実	メンバー
大塚製薬株式会社	京坂 侑土	メンバー
ファイザーR&D 合同会社	小宮山 靖	担当副部長

謝辞

本成果物の公表に際し、査読いただいたデータサイエンス部会運営幹事会ならびに医薬品評価委員会副委員長に御礼申し上げます。

また、CTQ 要因素案の評価にご協力いただいた以下の臨床評価部会有志の皆様には改めて感謝の意を表します。

会社名	氏名
アステラス製薬	松澤寛
エーザイ株式会社	青柳充顕
エーザイ株式会社	田中利知
バイエル薬品株式会社	實雅昭
大鵬薬品工業株式会社	篠原貴
帝國製薬株式会社	藤岡慶壮

Appendix 1: プロンプトの作成

「5.3.予備的試行（テストデータでの検討）」でも述べたとおり、初期のプロンプトは以下のようなシンプルなものであった。

添付の JSON ファイルは臨床試験の症例データです。このデータには誤りが含まれている可能性があります。

臨床試験の専門医として、以下の観点から JSON データを確認し、回答してください。時系列を考慮に入れて回答することが非常に重要です。また、症例本人の主訴も重要な情報です。

有害事象は、被験薬との因果関係を問わないあらゆる好ましくない事象や兆候を指します。

JSON データをレビューする観点

1. 臨床の経過から見て疑わしい情報

臨床の経過から見て疑わしい情報をすべて列挙し、その理由を説明します。

2. 有害事象として報告されている情報の整合性

報告された有害事象が、関連するその他の情報と整合が取れない場合、その理由を説明します。

3. 有害事象以外のデータにある安全性上の懸念

安全性上の懸念が認められる情報をすべて列挙し、それが有害事象として記載されているかどうかを評価します。

4. その他の不整合

上記以外に不整合が疑われるデータをすべて列挙し、その理由を説明します。

回答フォーマット

- **疑義事項**: 理由

- **疑義事項**: 理由

-

タスクフォースで作成した簡易的な JSON データを対象とした初期検討では、上記のプロンプトにより一定の成果が得られた。しかし、より実務に近い多様なデータを含む「sdm-adam-pilot-project」の臨床試験データでは、期待したほどの結果は得られなかった。これはデータ量の増加に伴って解釈の幅が広がったこと、そして、プロンプトの指示が曖昧であったことが原因であると考えた。たとえば、JSON データの構造や定義が明示されておらず、また、「臨床の経過から見て疑わしい情報」や「安全性上の懸念」が何を意味するのかが明確に記述されていない。これらは、メディカルモニターやクリニカルデータマネージャー、臨床開発モニターであれば経験的に理解している暗黙知であり、通常業務では不足情報を自ら補いながら判断する。しかし、大規模言語モデルによるタスク処理には作業の文脈が無く、与えたプロンプトに強く依存した結果を返すため、暗黙知の明文化と必要情報の明示が不可欠となる。そこで、作業の文脈を補うために、JSON のデータ定義情報 (Define.xml)

や試験実施計画書をプロンプトに埋め込み、さらにレビューの役割と観点を指定して、立場に基づく判断基準や知識を明文化することを試みた。

このプロンプトの作成は、判断基準を書き出して、実行結果を確認し、不足観点を補うという地道な反復作業であった。しかし、暗黙知となっている判断基準をすべて書き出してプロンプトを補うことは容易ではない。まず、無意識下で判断していることを明示していく作業に困難があり、さらにその判断基準をプロンプトにしていく点にも困難があった。大規模言語モデルの出力はプロンプトの細やかな記載に影響を受けることもあり、詳細に指示すると出力が硬直化して細部に偏ったり、逆に指示を緩めると重要と判断してほしい内容が表出してこないという状況にあった。

こうした課題を踏まえ、プロンプト作成のアプローチを転換した。人力でプロンプトを構成していく方法から、大規模言語モデルと対話しながら、次のような問いに答えていくことでプロンプトを構成していく方法に切り替えた。具体的には、以下のような問いに対する回答を大規模言語モデルに与え、本質的なタスクが実行できるように目的を明確化し、矛盾なくプロンプトを構成してもらうアプローチをとった。

- このプロンプトの最終的な目的は何か？本質的に実施したいことは何か？
- やってほしくないことは何か？
- 判断や出力において重要な点はどこにあるか？

このプロンプト作成においても一度で完成するものではなく、試行を繰り返して最終的なプロンプトに至っている。まずは作成したプロンプトでレビューを実行させ、得られた結果から不足していた観点を追加してプロンプトを再構成する、という試行を繰り返した結果である。

大規模言語モデルの発展と共にこのようなデータレビューサポートは現実的になっていくことが期待される。しかしながら、期待する結果を得るためには背景情報の提示だけでは不足、人が行っている判断プロセスを再現可能な形でプロンプトに落とし込んでいくこと、すなわち、「本質的に何をやっているのか」を意識してプロンプトを構成していくことが重要になるであろうと強く感じた。

Appendix 2: 余禄（AI による CTQ 要因特定会議のシミュレーション）

臨床評価部会有志のあるメンバーは自ら CtQF Facilitator を使用して CTQ 要因素案の生成を試みた。その後、CtQF Facilitator に医療機関の医師やスタッフの立場、またアルツハイマー型認知症患者または介護者の立場からそれらの CTQ 要因について検討をさせ、さらに臨床試験を実施する立場、また臨床試験に参加する立場から特定した CTQ 要因素案に対応するリスクの軽減策を取ることで臨床試験の実施や参加が難しくなる可能性の検討をさせた。それらの検討を AI に行わせた上で、CTQ 要因を特定するための会議のシミュレーションを行い、その会議の内容をまとめるように AI に指示をした。以下、AI が行ったシミュレーションの記録である。

ICH E8 (R1) では、患者や医療者を含む幅広い利害関係者からの情報が臨床試験のデザインに最も有用であることが述べられており、この会話風会議記録は一見ユーモラスながら幅広い利害関係者からの情報を試験デザインの検討に活かすという点において非常に興味深いと思われたので「余録」として収載することとした。

プロンプト

ここまで議論した内容を基に、実際に CTQ 要因を特定するための会議を開催したことをシミュレーションし、その会議の内容を会話風にまとめていただけないでしょうか。会議の参加者は、①製薬企業の臨床試験計画者、②製薬企業のモニタリング担当者、③製薬企業のデータマネジメント／統計解析担当者、④製薬企業の安全性管理担当者、⑤医療機関の試験担当医師／臨床試験コーディネーター、⑥臨床試験への参加対象となりえるアルツハイマー型認知症患者／介護者、とします。

会議記録

会議名: Xanomeline TTS 臨床試験 (H2Q-MC-LZZT(c)) CTQ 要因特定会議

参加者:

臨床試験計画者 (企): 製薬企業、臨床試験の全体計画を担当

モニタリング担当者 (モ): 製薬企業、臨床試験の実施状況を監視

データマネジメント／統計解析担当者 (デ): 製薬企業、データの収集・管理・解析を担当

安全性管理担当者 (安): 製薬企業、安全性情報の収集・評価を担当

試験担当医師 (医): 医療機関、臨床試験の実施責任者

臨床試験コーディネーター (コ): 医療機関、臨床試験のサポート

アルツハイマー型認知症患者の介護者 (介): 臨床試験への参加を検討

会議の冒頭 (臨床試験計画者):

企: 皆さん、本日はお忙しい中、Xanomeline TTS の臨床試験 (H2Q-MC-LZZT(c)) における Critical to Quality Factor (CTQ 要因) 特定会議にご参加いただき、ありがとうございます。本日は、この試験の成功に不可欠な要素を明確にし、リスクを特定し、適切な対策を講じることを目的としています。これまでの議論を踏まえ、各々の立場から意見を出し合い、実効性のある CTQ 要因を特定していきましょう。

評価項目の実施とデータの信頼性について

企: まず、評価項目の実施とデータの信頼性について議論しましょう。

医: 認知機能評価は、アルツハイマー型認知症患者さんにとって負担が大きい場合があります。長時間の評価は疲労を招き、結果に影響を与える可能性があります。

コ: 評価者間での評価のばらつきも懸念されます。特に、CIBIC+のような主観的な評価項目では、評価者の経験や判断によって結果が異なることがあります。

デ: データマネジメントの観点からは、評価項目の実施手順が標準化されていない場合、施設間でのデータ比較が困難になる可能性があります。

モ: モニタリングの立場からは、評価者のトレーニング状況や、評価手順の遵守状況を定期的に確認する必要があります。

介: 患者の立場からは、評価項目が難解で理解しにくい場合、不安やストレスを感じる場合があります。

安: 安全性評価の観点からは、認知機能評価の結果が、有害事象の判断に影響を与える可能性があるため、データの信頼性は重要です。

企: ありがとうございます。これらの意見を踏まえ、CTQ 要因を以下のように定義します。「ADAS-Cog、CIBIC+、NPI-X、DAD などの認知機能評価が、プロトコルに定められたスケジュール通りに、かつ標準化された方法で実施され、評価者間の一貫性が確保されること」。リスクとしては、評価者の熟練度不足、評価方法のばらつき、患者の疲労などが挙げられます。

企: 対策として、評価者に対する十分なトレーニングを実施し、評価手順を標準化すること、評価実施前に患者の体調を確認し、必要に応じて休憩を挟むこと、評価結果の品質管理を徹底することなどを検討します。

治験薬の適切な投与と患者アドヒアランスについて

企: 次に、治験薬の適切な投与と患者アドヒアランスについて議論しましょう。

医: TTS パッチは、貼付忘れや貼付部位の間違いが起こりやすい可能性があります。特に、認知機能が低下している患者さんでは、自己管理が困難な場合があります。

コ: 介護者の協力が不可欠ですが、介護者も高齢である場合や、多忙な場合、十分なサポートができない可能性があります。

モ: モニタリングの立場からは、TTS パッチの貼付状況や副作用に関する情報を定期的に収集し、患者アドヒアランスをモニタリングする必要があります。

介: 患者の立場からは、TTS パッチの貼付方法が複雑な場合、負担を感じる場合があります。また、貼付部位のローテーションが煩雑な場合、混乱を招く可能性があります。

安: 安全性評価の観点からは、TTS パッチの貼付部位に皮膚刺激が生じる可能性があるため、注意が必要です。

デ: データマネジメントの観点からは、TTS パッチの貼付状況や副作用に関する情報を正確に記録し、管理する必要があります。

企: ありがとうございます。これらの意見を踏まえ、CTQ 要因を以下のように定義します。「Xanomeline TTS が、プロトコルに定められた用法・用量で、かつ適切な方法で患者に投与され、患者アドヒアランスが維持されること」。リスクとしては、TTS パッチの貼付忘れ、貼付部位の間違い、貼付時間の逸脱、皮膚刺激などが挙げられます。

企: 対策として、患者および介護者に対して、TTS パッチの使用方法に関する詳細な説明と教育を行うこと、TTS パッチの貼付状況や副作用に関する情報を定期的に収集し、

患者アドヒアランスをモニタリングすること、TTS パッチの貼付を容易にするための補助具や、皮膚刺激を軽減するための対策を検討することなどを検討します。

安全性情報の正確な収集と報告について

企: 続いて、安全性情報の正確な収集と報告について議論しましょう。

医: アルツハイマー型認知症患者さんの場合、有害事象の自覚症状を正確に伝えることが難しい場合があります。介護者からの情報収集が重要になります。

コ: 有害事象の定義、重症度評価、因果関係評価に関する明確な基準を設ける必要があります。

モ: モニタリングの立場からは、有害事象の報告が遅延しないように、報告手順を明確化する必要があります。

安: 安全性管理の立場からは、重篤な有害事象が発生した場合、迅速に対応できる体制を整える必要があります。

デ: データマネジメントの観点からは、有害事象に関する情報を正確に記録し、管理する必要があります。

企: ありがとうございます。これらの意見を踏まえ、CTQ 要因を以下のように定義します。「有害事象、臨床検査値異常、バイタルサイン変動などの安全性情報が、プロトコルに定められた方法で正確に収集され、適切に報告されること」。リスクとしては、有害事象の過小報告、因果関係の誤認、報告遅延などが挙げられます。

企: 対策として、有害事象の定義、重症度評価、因果関係評価に関する明確な基準を設けること、臨床検査の標準化、品質管理、および基準範囲に関する適切なトレーニングを実施すること、重篤な有害事象の報告手順を明確化し、迅速な報告を徹底することなどを検討します。

盲検性の維持について

企: 次に、盲検性の維持について議論しましょう。

モ: モニタリングの立場からは、治験実施者、患者、および評価者が、治療割り当てを推測できるような情報が漏洩しないように注意する必要があります。

医: 緊急時の盲検解除手順を厳格に管理する必要があります。

企: ありがとうございます。これらの意見を踏まえ、CTQ 要因を以下のように定義します。「治験実施者、患者、および評価者が、治療割り当てについて盲検化されていること」。リスクとしては、治療割り当ての推測、意図的な情報漏洩、緊急時の盲検解除などが挙げられます。

企: 対策として、TTS パッチの外観をプラセボと実薬で完全に同一にすること、緊急時の盲検解除手順を厳格に管理すること、治験関係者に対して、盲検性の重要性に関する教育を徹底することなどを検討します。

倫理的配慮と患者の権利保護について

企: 最後に、倫理的配慮と患者の権利保護について議論しましょう。

医: アルツハイマー型認知症患者さんの場合、インフォームドコンセントの説明内容を理解することが難しい場合があります。家族や介護者への説明も重要になります。

コ: 患者さんのプライバシーと個人情報を保護するための対策を徹底する必要があります。

介: 患者の立場からは、試験への参加が強制ではなく、いつでも自由に中止できることを知っておく必要があります。

企: ありがとうございます。これらの意見を踏まえ、CTQ 要因を以下のように定義します。「患者のインフォームドコンセントが適切に取得され、患者のプライバシーと個人情報が保護されること」。リスクとしては、インフォームドコンセントの説明不足、個人情報の漏洩などが挙げられます。

企: 対策として、インフォームドコンセントの説明内容を平易な言葉で記述し、患者の理解度を確認すること、患者の個人情報の取り扱いに関する規定を明確化し、セキュリティ対策を徹底することなどを検討します。

会議のまとめ（臨床試験計画者）：

企: 本日は、活発な議論をいただき、ありがとうございました。皆様の意見を参考に、CTQ 要因を特定し、リスクを特定し、適切な対策を講じることで、臨床試験の品質を向上させ、患者さんの安全を確保し、信頼性の高いデータを得ることができると確信しています。今後とも、ご協力をお願いいたします。

（会議終了）