

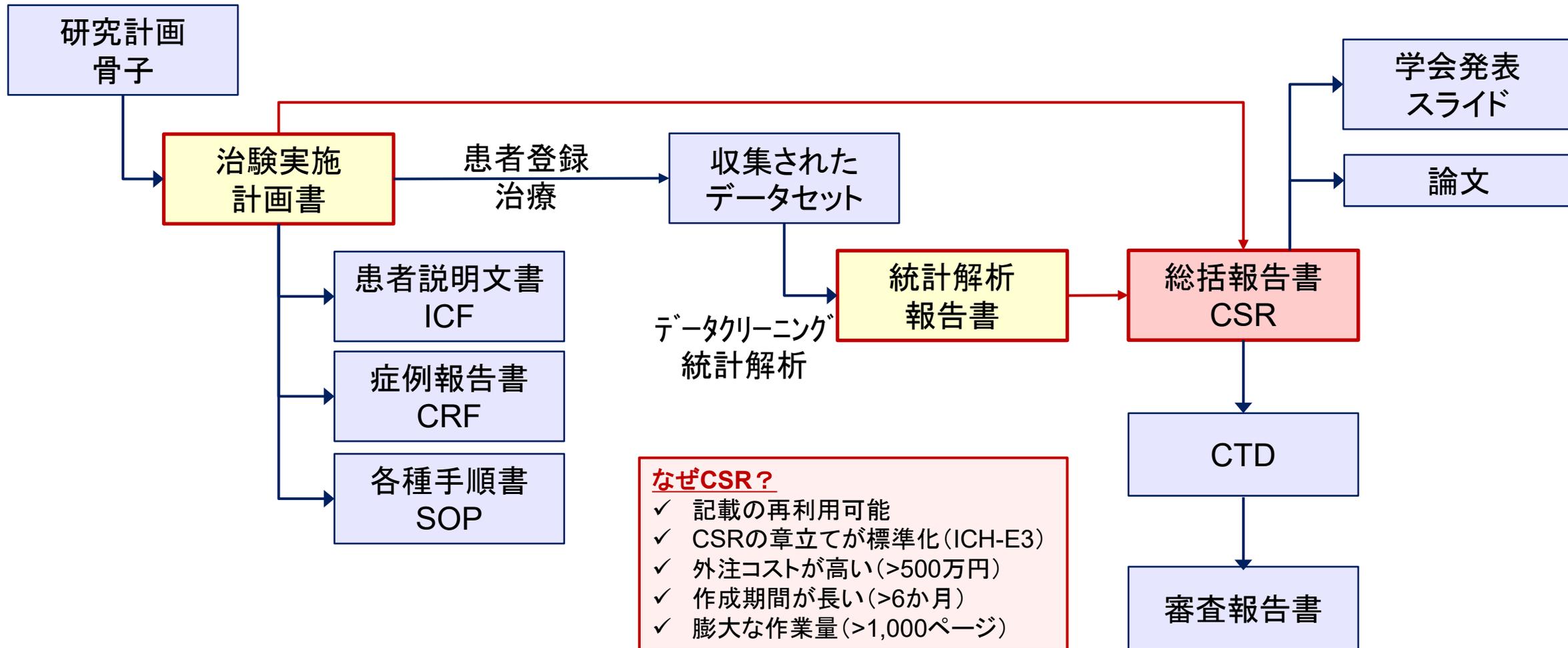
# 臨床試験プロセスにおける 生成AIの利活用

国立がん研究センター中央病院  
国際開発部門/臨床研究支援部門  
中村 健一

# 医薬品開発におけるAI/DX



# 臨床試験プロセスは文書のつながり



# 治験関連文書の自動作成プロジェクト

## ■ 第1期(2024年度)

- 治験総括報告書(CSR)の自動作成プロジェクトを開始
- ベンダーとの共同研究で、生成AIを用いたCSR作成ツールを開発
- CSRの第9章についてAIによる自動作成の精度を評価

## ■ 第2期(2025年度)

- 第1期の結果を踏まえ、CSR作成用プロンプトの改善と汎用化を実施
  - CSRの第1章、第5～8章、第9章について自動作成
- PMDAと共同で厚労科研を獲得(2025年11月～)
  - 「生成AIを活用した薬事承認申請・審査関連文書作成の推進のための研究」
    - NCCではCSR作成のさらなる精度向上が主目的
    - PMDAでは審査報告書の一部の自動作成に着手
- 並行してNCCではプロトコールの自動作成にも着手

# 文書自動作成ツールの仕組み

例：CSR



\*RAG: Retrieval-Augmented Generation

# 実際のCSR自動作成のイメージ

Reference Link:  
Click here to see the data source

Generate Options

Initial Content:  
1回の投与を1コースとして「9.3.3.2 プロトコール治療中止規準」に該当しない限り治療を継続した。

Option 1:  
1回の投与を1コースとして治療を継続した。

Option 2:  
1回の投与を1コースとして「9.3.3.2 プロトコール治療中止規準」に該当しない限り治療を継続した。

エディタ画面

9. 治験の計画

9.1 治験の全般的デザインおよび計画 - 記述

9.1.1 試験デザイン

9.1.1.1 プロトコール治療

本試験では、[redacted] を2週間間隔で30分以上かけて点滴静注にて投与した。

1回の投与を1コースとして「9.3.3.2 プロトコール治療中止規準」に該当しない限り治療を継続した。

テスト エディタで編集可能

本試験では [redacted] の減量は行われなかった。よって本試験におけるプロトコール治療中の治療変更は「9.1.1.1.2 コース開始 [redacted] による「延期」のみであった。

9.1.1.1 プロトコール治療の開始

二次登録後14日以内にプロ [redacted] 開始した。14日以内に投与を開始できなかった場合は、開始が遅れた理由をCRFに [redacted] する。

2コース以降の [redacted] 日～+3日を許容した。ただし、年末年始が [redacted] することとした。

なお、点滴の方法 [redacted] する。

9.1.1.1.2 コース開始

各コースのday1は [redacted] 上、 [redacted] の投与を行った。なお、1コース開始時は二次登録時の所見、検査値を用いてよいが、登録後にも検査を行った場合は、投与直近の所見、検査値を用いた。

「コース開始規準」のいずれかを満たさない場合、 [redacted] の投与を延期し、当該規準をすべて満たすことが確認できた時点で投与を再開した。4コース以降でコース開始前に臨床検査、12誘導心電図検査および胸部X線検査を実施しないコース（5コース、6コース、8コース、9コースなど）においては、各コース開始前に実施した評価項目に基づき、「コース開始規準」を満たしていると担当医が判断した場合、 [redacted] の継続投与を可成とした。なお、コース開始規準を満たさないことを理由に、day1の許容範囲内に投与が行われなかった場合は、治験実施計画書からの逸脱として取り扱われなかった。

チャットボット

Data source

参照箇所を黄色ハイライトで表示

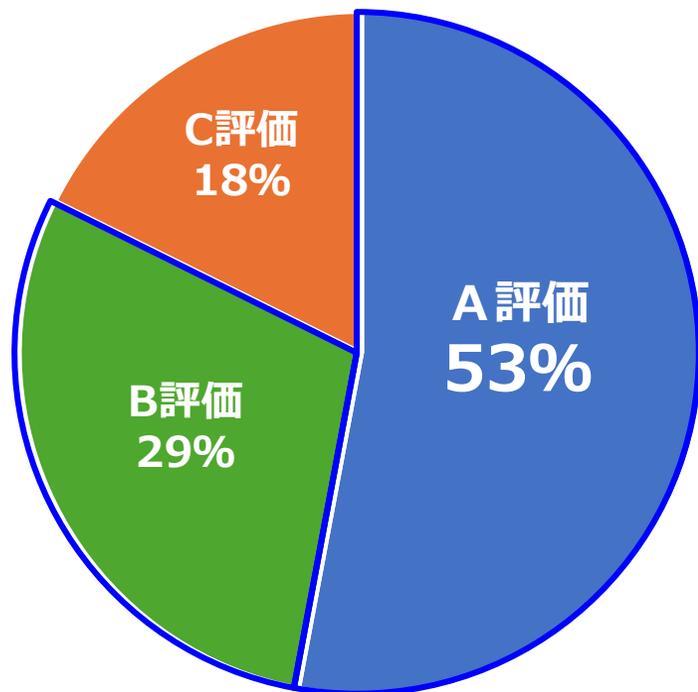
エディタ画面で追記・修正が可能 (赤枠：追記した箇所)

自動作成された複数の文書候補が表示されるため、好みに応じて選択可能

自動生成されたCSR：赤枠部分に複数の文書候補が生成され、好みの候補を選択可能。  
内容は青枠のエディタ画面で自由に編集することが可能。  
緑枠のチャットボットで一般的な質問などが可能。

# 自動作成したCSRの精度 (第9章)

## 第1期 (評価項目数119)



**Phase1 (試験固有)**  
 A + B = **約80%**が実用レベル  
 (ほぼそのまま使用可能)

**Phase2 (汎用化)**  
 A + A- + B = **約97%**が実用レベル  
 (修正最小限で使用可能)

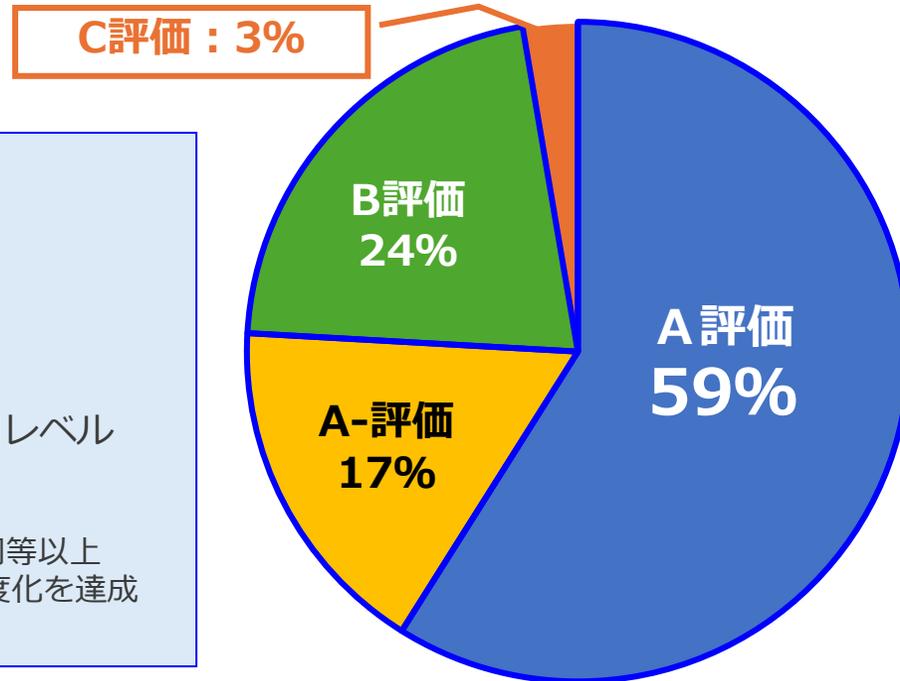
➡汎用化後も生成精度はPhase 1と同等以上  
 実用性を損なうことなくツールの高度化を達成

評価項目	評価内容	項目数	割合
A	ほぼそのまま利用可能	63	53%
B	微修正で利用可能	35	29%
C	大幅な追加作業が必要	21	18%
D	利用は困難	0	0%

**【C評価の例】**

- 表の形式が崩れる
- 検査の評価期間のハルシネーション
- 実施していない「薬物濃度の測定」の記載を生成

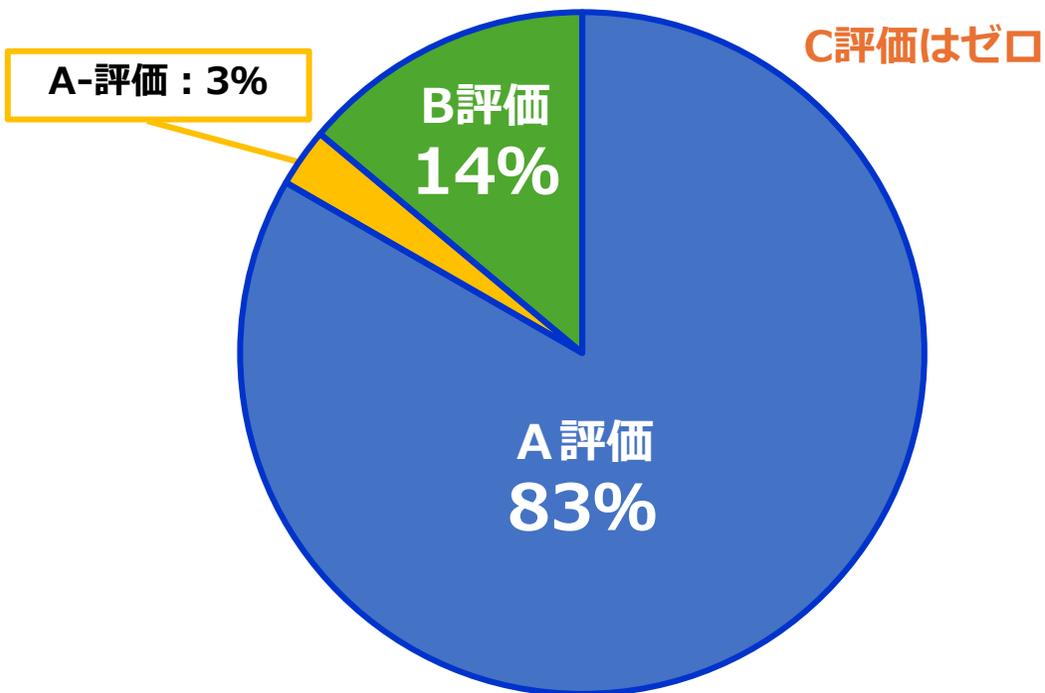
## 第2期 (評価項目数112)



評価項目	評価内容	項目数	割合
A	そのまま利用可能	66	59%
A-	軽微な記載整備で利用可能	19	17%
B	微修正で利用可能	24	24%
C	大幅な追加作業が必要	3	3%
D	利用は困難	0	0%

# 自動作成したCSRの精度 (第1章・第5～8章)

## 第2期 (評価項目数36)



評価項目	評価内容	項目数	割合
A	そのまま利用可能	30	83%
A-	軽微な記載整備で利用可能	1	3%
B	微修正で利用可能	5	14%
C	大幅な追加作業が必要	0	0%
D	利用は困難	0	0%

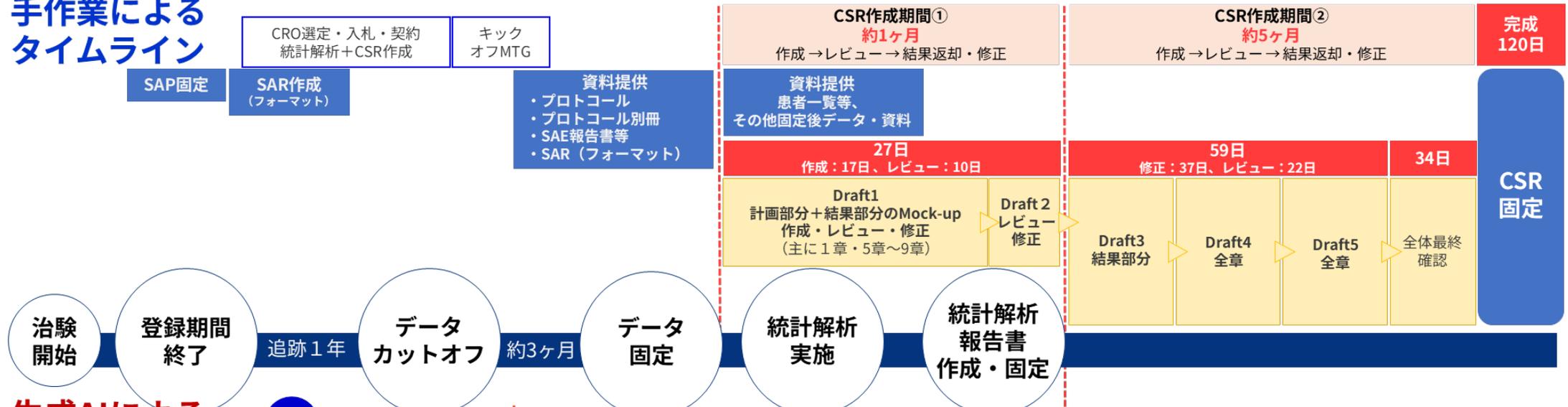
### ■ 第2期の結果

- 第1章・第5～8章は**100%**、第9章でも**97%**が実用可能と評価され、最小限の微修正で利用可能であることを確認
- 今後、複数の試験での再現性を評価しつつ、がん以外の臨床試験での精度を評価予定
- インプット側のプロトコール等の章立ての標準化を進めることも精度向上のためには必要

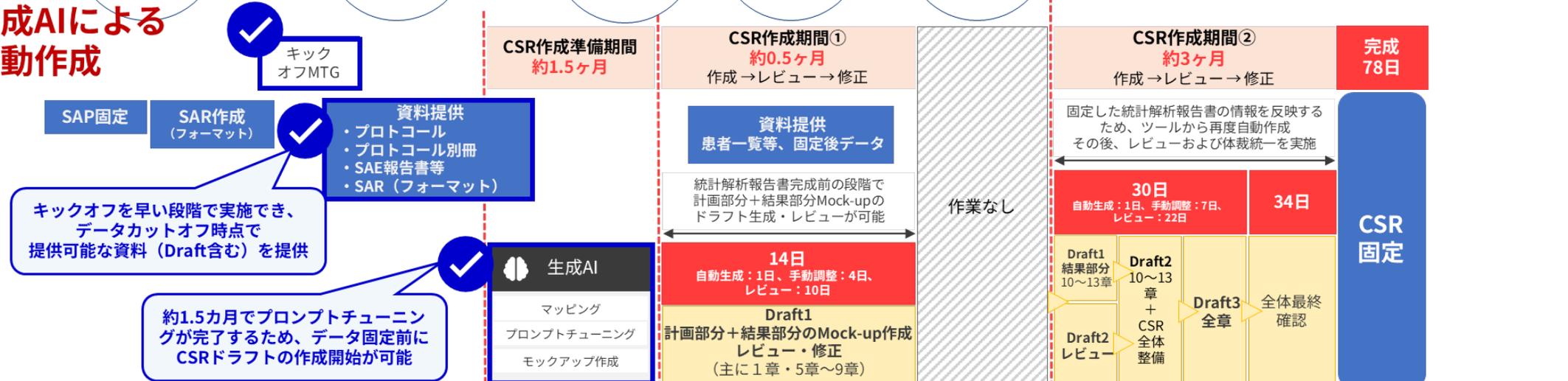
# 作業時間の短縮効果

生成AIの活用により営業日換算で120日→78日に短縮

## 手作業によるタイムライン



## 生成AIによる自動作成



# 生成AIの利活用事例：患者-試験マッチング “Patient to Trial” type

## ■ 個別患者に適合する臨床試験の自動マッチング

### □ TrialGPT(患者-試験マッチングLLM)

■ Retrieval: 大規模フィルタリングによる候補試験の抽出

■ Matching: 電子カルテ内の患者情報と適格規準との関連性を評価

■ Ranking: 試験レベルで関連性をスコア化し、総合スコアで試験をランキング

□ 患者-試験マッチング判定の精度は87.3%(専門家の水準に近いレベル)

□ 試験ごとのランク付けの精度は43.8%(従来モデルを上回る)

□ 患者1人あたりの試験照合に要する時間が42.6%短縮

nature communications

Article

<https://doi.org/10.1038/s41467-024-53081-z>

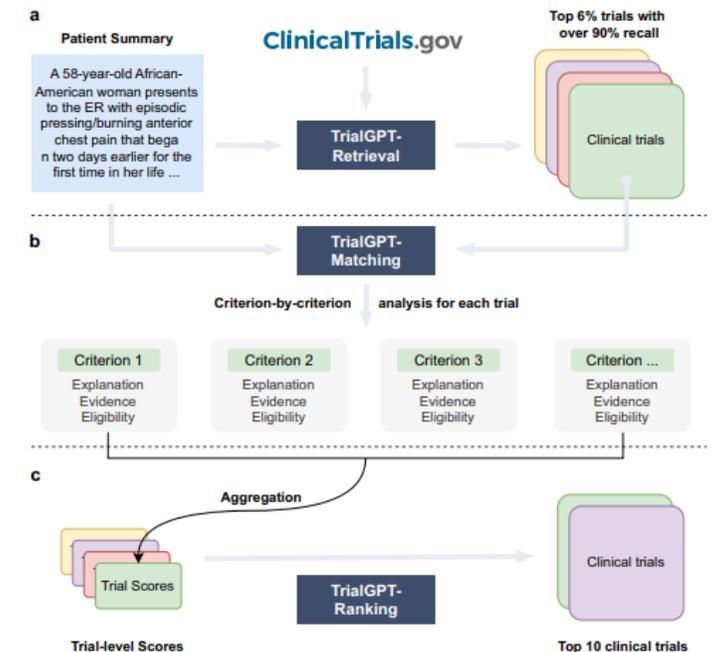
## Matching patients to clinical trials with large language models

Received: 18 January 2024

Accepted: 1 October 2024

Qiao Jin<sup>1</sup>, Zifeng Wang<sup>2</sup>, Charalampos S. Floudas<sup>3</sup>, Fangyuan Chen<sup>4</sup>, Changlin Gong<sup>5</sup>, Dara Bracken-Clarke<sup>3</sup>, Elisabetta Xue<sup>3</sup>, Yifan Yang<sup>1,6</sup>, Jimeng Sun<sup>2</sup> & Zhiyong Lu<sup>1</sup>✉

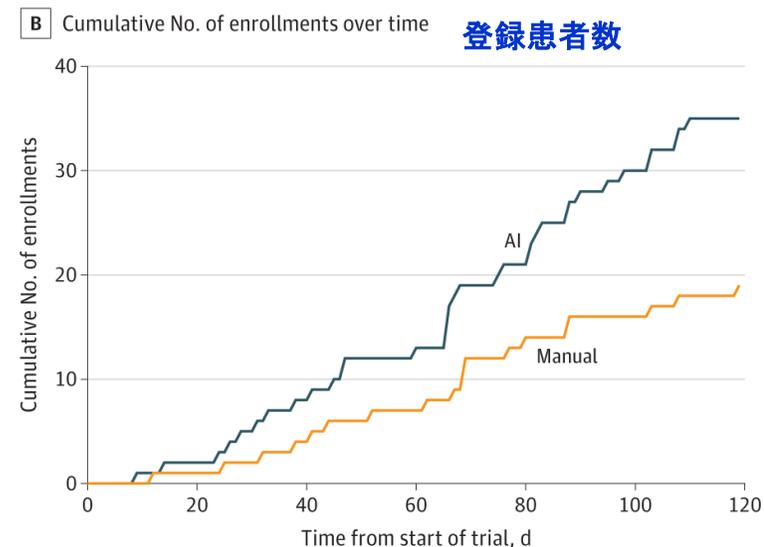
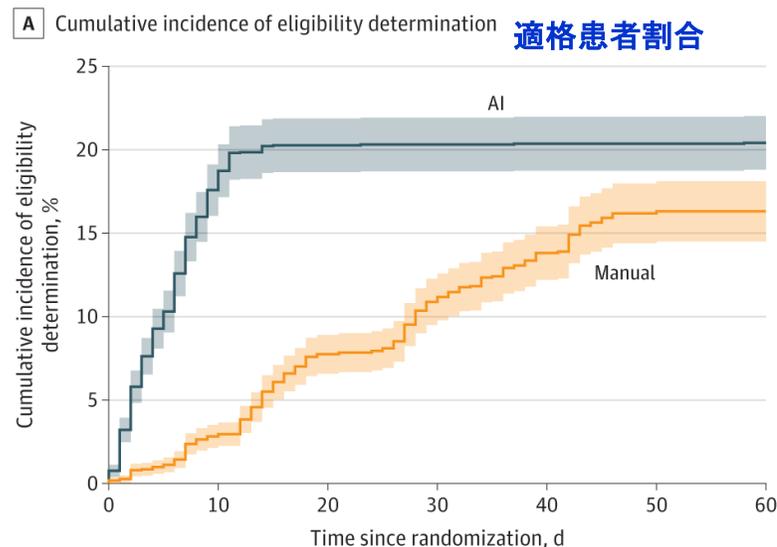
Jin Q, et al. Nat Commun 15, 9074. 2024



# 生成AIの利活用事例：患者-試験マッチング “Trial to Patient” type

## ■ 特定試験における適格患者のスクリーニング

- 電子カルテ上の非構造化診療メモから、臨床試験の適格規準に合致する患者を自動抽出
- 心不全治療の臨床試験(COPILOT-HF試験)において、AIツール使用群と従来の手作業群にランダム化して適格性判定までの時間を比較
  - AI群で適格性判定を有意に迅速化(ハザード比1.78,  $p < 0.001$ )
  - 適格と判定された患者の割合: AI群 20.4% > 手作業群 12.7% ( $p < 0.001$ )
  - 実際の試験への登録患者数: AI群 35例(1.6%) > 手作業群 19件(0.9%) ( $p = 0.04$ )



# 生成AIの利活用事例：予測系

## ■ 臨床試験成功確率の予測

- GPT-4を用いた臨床試験結果予測モデル
  - 430件の過去試験データと過去論文からモデルをファインチューニング
  - 強化学習で専門家のフィードバックを組み込み
- 試験成功/不成功を70%以上の精度で的中(AUC=0.84)
  - 人間では気付かない試験設計上の欠陥を指摘したケースも

<https://www.panda-int.com/insights/ai-in-clinical-trials/>

## ■ 施設選定の最適化

- 多発性骨髄腫の試験を題材に、全米1,246施設・約15,881人の患者データをベースに、目標登録ペースと患者の多様性の規準に合致した施設選定案を提示
- 3施設のみを選定することで、目標登録ペース(10例/月)を上回りつつ、患者の多様性を確保できることを予測
- 従来の手作業よりも迅速・高精度に施設選定が可能であることを示唆

Ruchlin I, et al. J Clin Oncol 42, 2024 (suppl 16; abstr 1558)

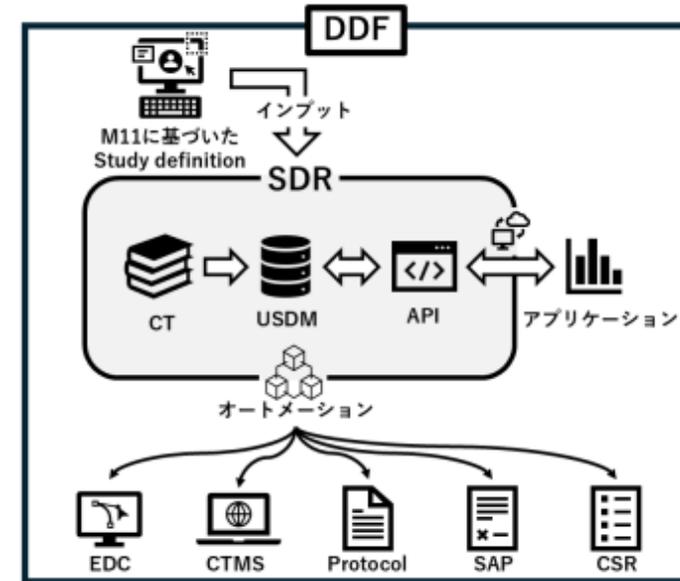
# ICH-M11とDigital Data Flow

## ■ ICH-M11による標準化

- なぜプロトコルの国際標準化？
  - 企業別、団体別に異なるプロトコルテンプレート
  - プロトコル情報を下流のCRFやCSRへ転記する際のエラー
  - 手作業によるプロトコル作成で生じる遅延
- これらを解決するためプロトコルを国際標準化 (CeSHaRP)
  - プロトコル情報のデータ互換を可能とする技術仕様を定義
    - 2025年11月にICH-M11としてStep 4に到達

## ■ Digital Data Flow (DDF)

- TransCelerate主導の、臨床試験情報の構造化・標準化プロジェクト
- Study Definitions Repository (SDR)を最上流とする、プロトコルやEDC、CTMS、あらゆる下流の文書への自動的な流し込み
- 構造化されたデータはAPIを通じて外部連携も可能に
- DDFと生成AIの組み合わせにより、試験期間の飛躍的な短縮とコスト削減が期待される



製薬協医薬品評価委員会データサイエンス部会  
「ICH M11導入後の医薬品開発への影響」

✓ 生成AI利活用を推進する上でも、プロトコルの標準化と臨床試験情報の構造化は極めて重要

# 最後に

- 医薬品開発のあらゆるプロセスでAI利活用が進む
  - 治験関連文書の自動化は最もハードルが低い利活用事例
  - CSRをはじめとした文書自動作成により作業時間短縮、コスト削減が進む
  - 文書作成のみならずデータマネジメント、統計解析等、あらゆるプロセスにAIの適用が進む
  - 米国、中国等ではプロセスの「一部」ではなく「全体」のautomationがなされつつある
  - 結果的に医薬品開発サイクルの飛躍的な迅速化が進む
- 生成AIの利活用にはデータの標準化・構造化・集約化が必要
  - AIによるアウトプット向上には、インプット側（プロトコール等）の標準化が必要
  - 臨床試験の情報に加えて、作業プロセスも生成AI利活用を前提とした見直しが迫られる
  - 生成AIの基礎が学習データ量であることを踏まえると、日本の課題である施設間データ連携の遅れを解決する試みも重要