

令和6年度厚生労働行政推進調査事業費補助金

厚生労働科学特別研究事業

診療行為の構造化と生成 AI 等を活用した標準化されたレセプト作成機能開発  
の為に基礎的調査研究（24CA2031）

令和6年度 総括・分担研究報告書

研究代表者 藤林 和俊

令和7年 5月



厚生労働大臣 殿

機関名 順天堂大学  
所属研究機関長 職 名 学長  
氏 名 代田 浩之

次の職員の令和6年度厚生労働科学研究費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

1. 研究事業名 厚生労働科学特別研究事業
2. 研究課題名 診療行為の構造化と生成 AI 等を活用した標準化されたレセプト作成機能開発の為の基礎的調査研究
3. 研究者名 (所属部署・職名) 革新的医療技術開発研究センター・准教授  
(氏名・フリガナ) 藤林 和俊・フジバヤシ カズトシ

4. 倫理審査の状況

	該当性の有無 有 無	左記で該当がある場合のみ記入 (※1)		
		審査済み	審査した機関	未審査 (※2)
人を対象とする生命科学・医学系研究に関する倫理指針 (※3)	<input checked="" type="checkbox"/> <input type="checkbox"/>	<input checked="" type="checkbox"/>	順天堂大学	<input type="checkbox"/>
遺伝子治療等臨床研究に関する指針	<input type="checkbox"/> <input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	<input type="checkbox"/> <input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
その他、該当する倫理指針があれば記入すること (指針の名称: )	<input type="checkbox"/> <input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>

(※1) 当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他 (特記事項)

(※2) 未審査に場合は、その理由を記載すること。  
(※3) 廃止前の「疫学研究に関する倫理指針」、「臨床研究に関する倫理指針」、「ヒトゲノム・遺伝子解析研究に関する倫理指針」、「人を対象とする医学系研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について

研究倫理教育の受講状況	受講 <input checked="" type="checkbox"/> 未受講 <input type="checkbox"/>
-------------	---

6. 利益相反の管理

当研究機関におけるCOIの管理に関する規定の策定	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合はその理由: )
当研究機関におけるCOI委員会設置の有無	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合は委託先機関: )
当研究に係るCOIについての報告・審査の有無	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合はその理由: )
当研究に係るCOIについての指導・管理の有無	有 <input type="checkbox"/> 無 <input checked="" type="checkbox"/> (有の場合はその内容: )

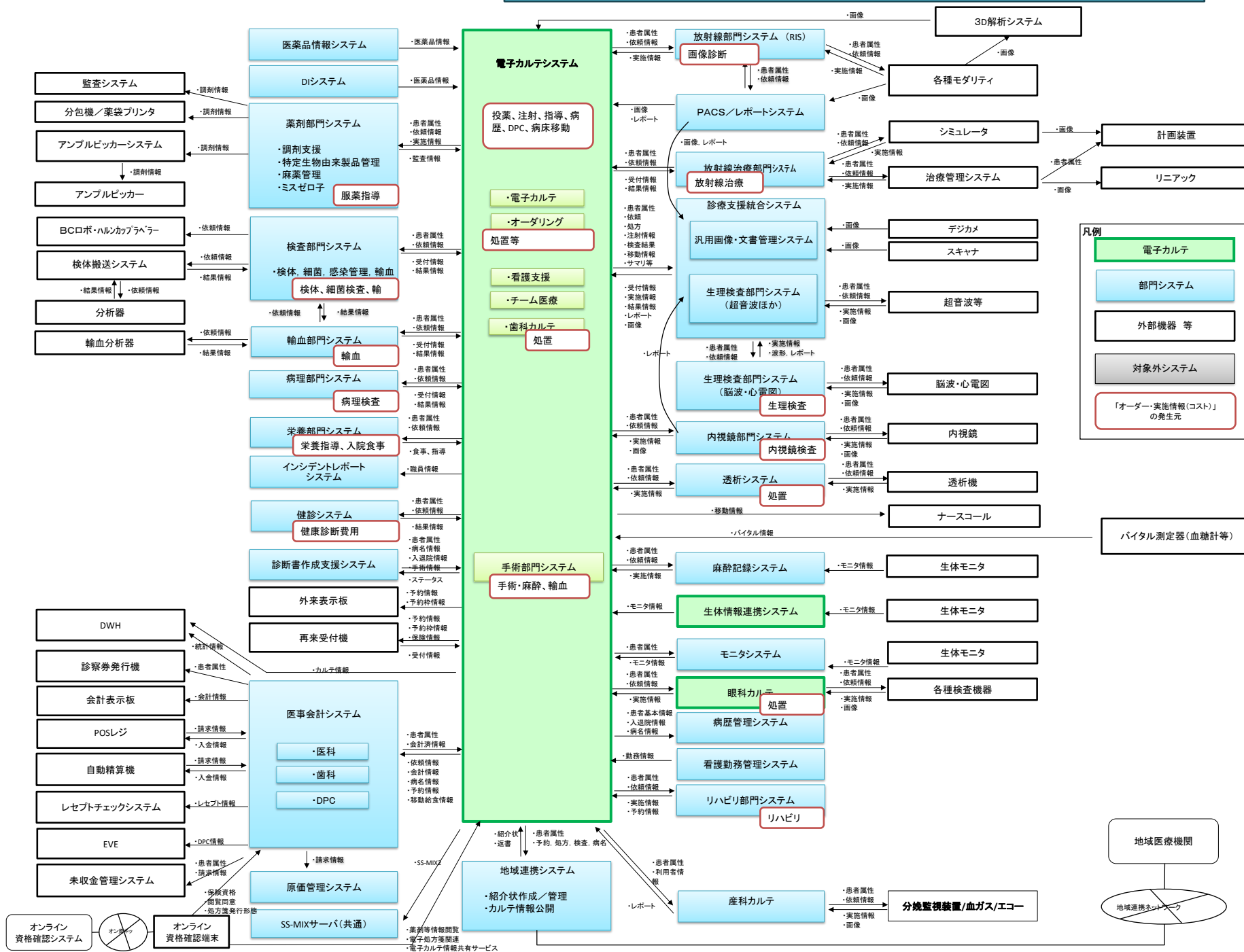
(留意事項) ・該当する□にチェックを入れること。  
・分担研究者の所属する機関の長も作成すること。



# システム連携概念図(大規模病院:500床程度をモデル)

※県立病院を参考

基本的には各部門に個別システムが存在しており、オーダーに対する実施入力を部門システムで行われる。  
※機器が存在しない部門システム(手術、リハビリなど)は電子カルテシステムより一体型として提供される事例もあり。  
部門システムにて実施入力された情報が電子カルテシステムを経由して医事会計システムへコスト等の情報として伝達される。  
投薬、指導、および処置(即実施の場合)などはオーダー時に医事会計システムへコスト等の情報として伝達される。

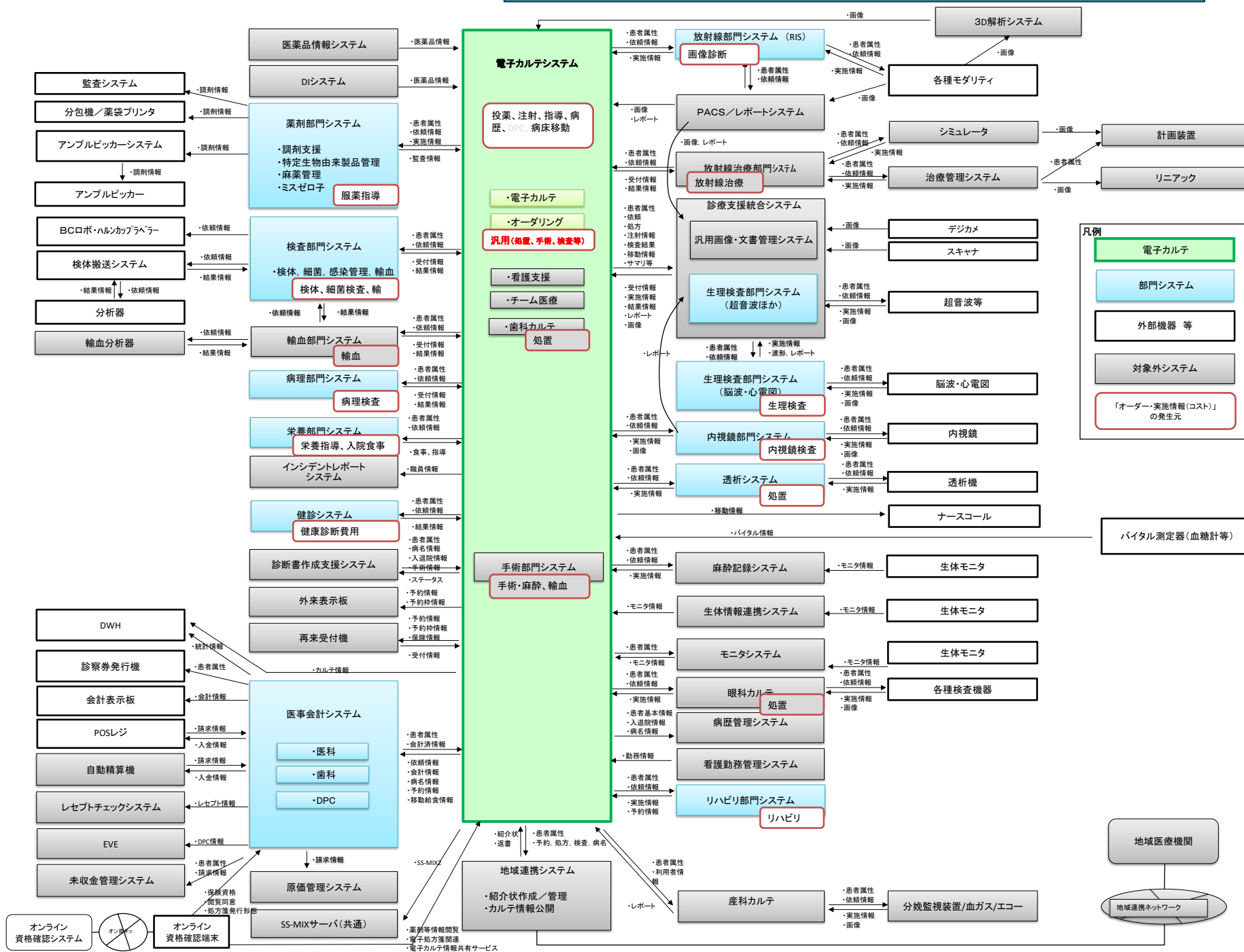




# システム連携概念図(中規模病院:精神科250床程度をモデル)

※県立A病院を参考

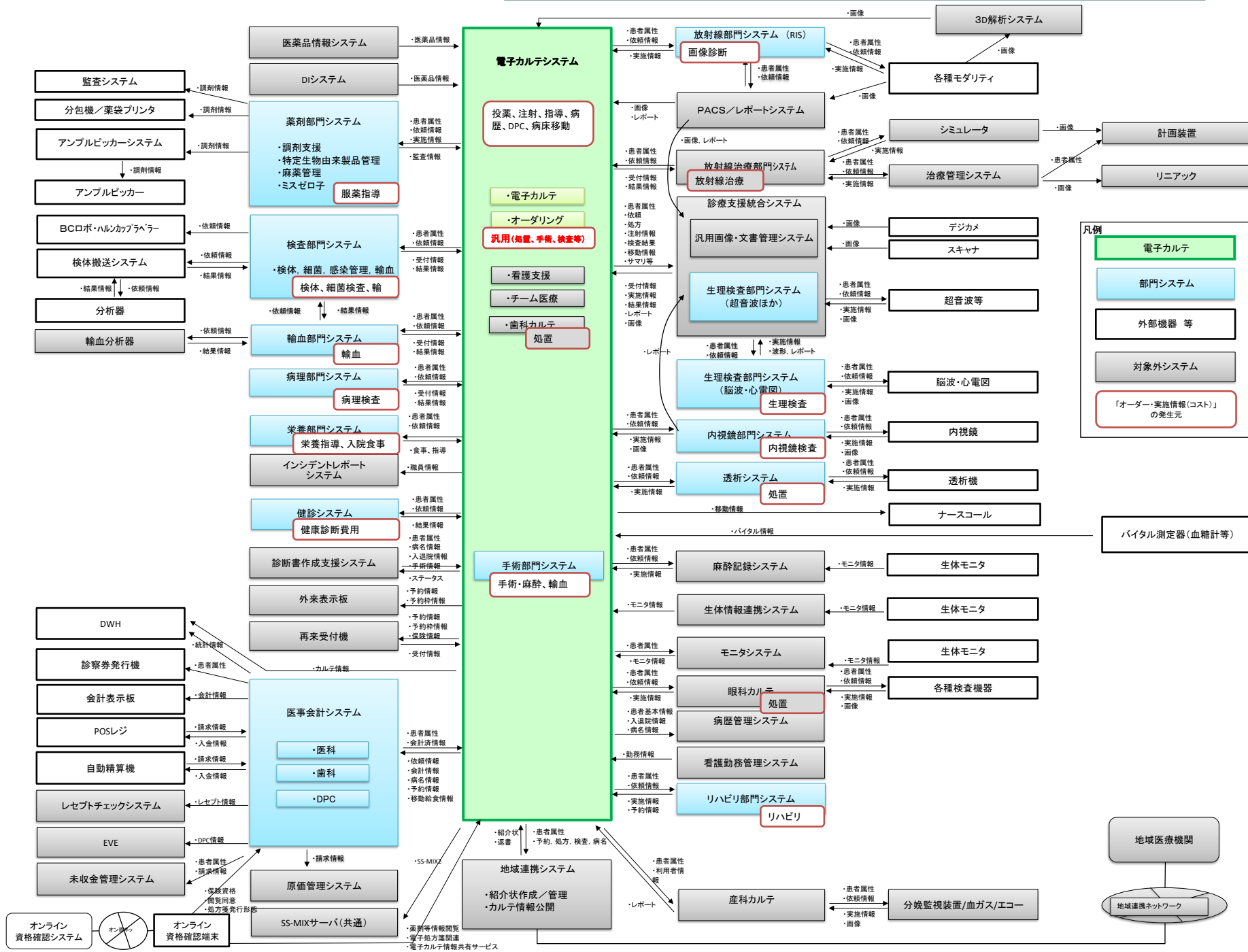
大規模モデルから一部の部門システムが存在しないケース(カルテ記載のみ行い伝票運用など)増える。もしくは小規模モデルから手術、麻酔、輸血部門、DPCを削いだ構成となるイメージ。  
ただし、昨今の「電子カルテ化」に伴い「結果の伴う部門システム」については何らかの部門システムが導入され、結果値を電子カルテ側へ連携している傾向あり。  
部門システムが存在しないオーダー種については電子カルテ側でオーダー入力(EX:汎用オーダー機能など)された情報を医事システムへペストとして連携する。  
その他に患者様による操作を行う機器(自動精算機)の導入が避けられる傾向あり。





※民間A病院を参考

大規模モデルから一部の部門システムが存在しないケース(カルテ記載のみ行い伝票運用など)増える。  
ただし、昨今の「電子カルテ化」に伴い「結果の伴う部門システム」については何らかの部門システムが導入され、結果値を電子カルテ側へ連携している傾向あり。  
部門システムが存在しないオーダー種については電子カルテ側でオーダー入力(Ex.汎用オーダー機能など)された情報を医事システムへコストとして連携する。

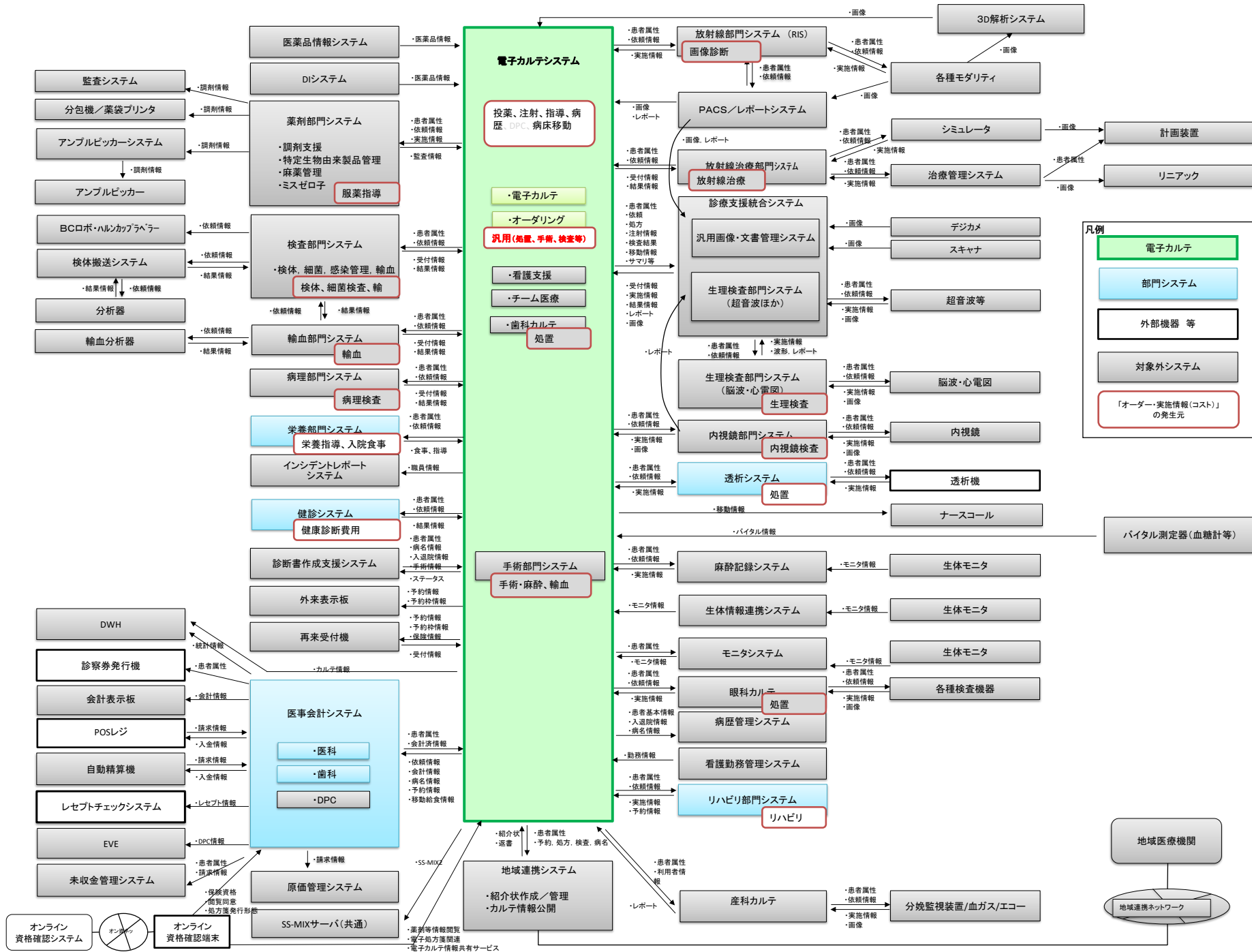




# システム連携概念図(クリニック:0~19床の有床診療所をモデル)

※民間Kクリニック、民間Yクリニックを参考

大規模モデルから一部を除いた部門システムが存在しないケースとなる。  
ただし、その診療所の「特色に沿った部門システム(検査系、遠隔、リハビリ、健診など)」が導入されているケースあり。  
基本的には電子カルテ側でオーダー入力(EX:汎用オーダー機能など)された情報を医事システムへコストとして連携する。  
小規模になることでリアルタイムレセプトチェック機能(会計登録時に即時レセプトチェックを実施する機能)を導入する事例あり。

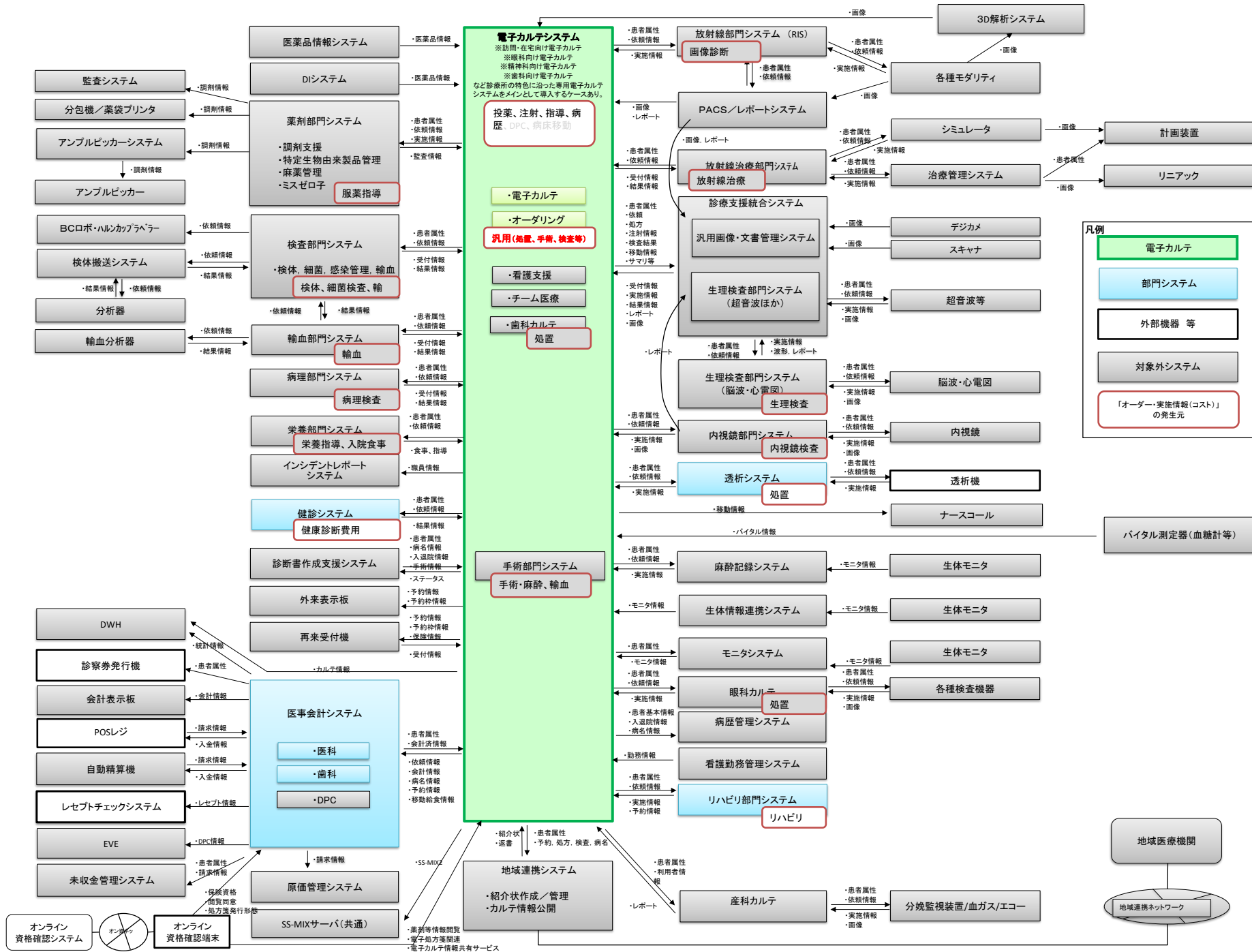




# システム連携概念図(クリニック:無床診療所をモデル)

※民間クリニック、民間クリニックを参考

大規模モデルから一部を除いた部門システムが存在しないケースとなる。  
ただし、その診療所の「特色に沿った電子カルテ、部門システム(検査系、透析、リハビリ、健診など)」が導入されているケースあり。  
基本的には電子カルテ側でオーダー入力(Ex.汎用オーダー機能など)された情報を医事システムへコストとして連携する。  
小規模になることでリアルタイムレセプトチェック機能(会計登録時に即時レセプトチェックを実施する機能)を導入する事例あり。無床診療所の為「病棟管理関連(病床移動、食事)」が不要となる。





別添 5

研究成果の刊行に関する一覧表レイアウト

書籍

著者氏名	論文タイトル名	書籍全体の 編集者名	書 籍 名	出版社名	出版地	出版年	ページ
該当なし							

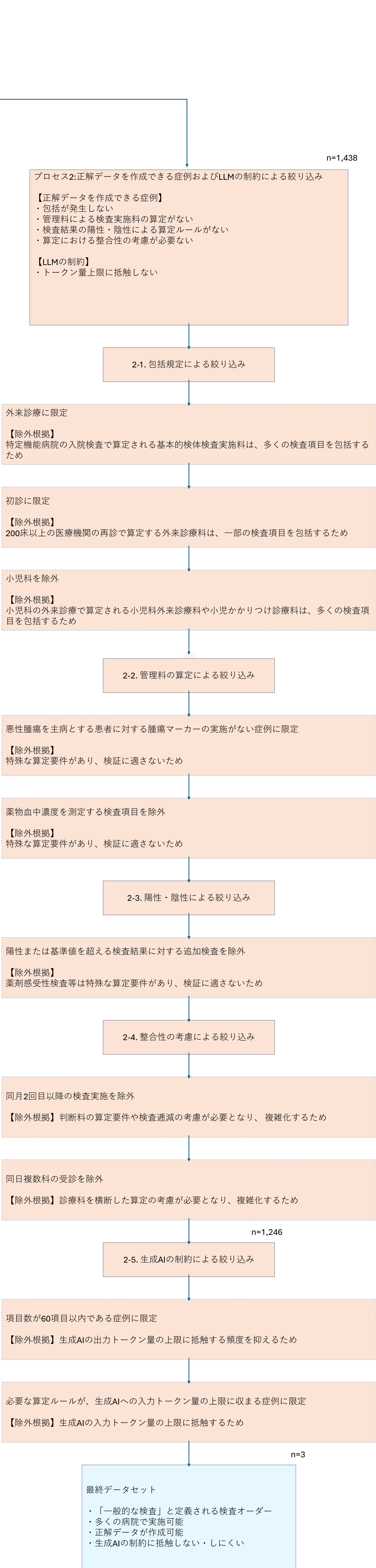
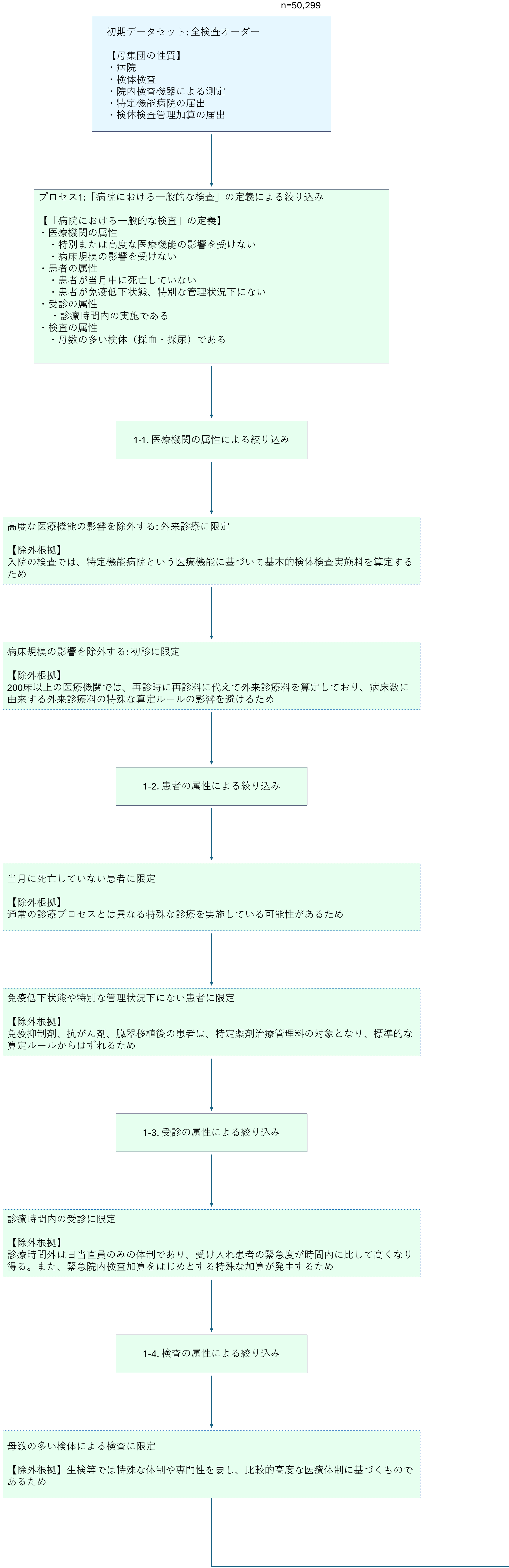
雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
該当なし					









フロー左側・緑色の枠組みは、「病院における一般的な検査」の定義に合致する症例への絞り込みのプロセスを示している。



## 目 次

### I. 総括研究報告

- 診療行為の構造化と生成 AI 等を活用した標準化されたレセプト作成機能開発  
の為に基礎的調査研究（24CA2031）----- 1  
藤林和俊

### II. 分担研究報告

1. 診療行為の構造化と生成 AI 等を活用した標準化されたレセプト作成機能開発  
の為に基礎的調査研究：AS-ISモデルの調査と整理 ----- 3  
齋藤智明
2. 診療行為の構造化と生成 AI 等を活用した標準化されたレセプト作成機能開発  
の為に基礎的調査研究：CANBEモデル検討の為に基礎的調査・実験 -- 4  
八木摂子



令和6年度厚生労働行政推進調査事業費補助金（厚生労働科学特別研究事業）  
総括研究報告書

診療行為の構造化と生成 AI 等を活用した標準化されたレセプト作成機能開発  
の為に基礎的調査研究（24CA2031）

研究代表者 藤林 和俊 順天堂大学革新的医療技術開発研究センター准教授

研究要旨：本研究は、生成AIを診療報酬算定業務に適用する可能性と課題を明らかにした。現状では完全自動化は難しいが、技術の発展と運用方法の最適化により、医療関係者の負担軽減と算定精度向上が見込まれる。

順天堂大学・健康総合科学先端研究機構バイオリソースリサーチセンター・非常勤助教：齋藤智之  
株式会社 FIXER・エンタープライズ部門・メンバー：八木摂子

### A. 研究目的

診療行為の多様性やデータの整合性、請求ルールの複雑さにより、診療報酬請求は医療従事者の大きな負担となっている。本研究では、電子カルテや医事会計システムを構造化し、生成AIと既存の請求システムを組み合わせること

### B. 研究方法

令和6年度は、AI活用による事務改善のユースモデル構築の為に、以下の調査・研究を実施した。

#### A. AS-ISモデルの調査と整理

- 複数規模医療機関のシステム構造調査：
- 規模に関わらず使用しているHISの項目まとめ
- 診療オーダーから診療報酬請求フロー作成
- 既存の自動レセプトチェックシステムのまとめ
- 現状の課題の整理

#### B. CANBEモデル検討の為に基礎的調査・実験

- 生成AI活用シーン探索の為に基礎実験：検体検査における診療報酬請求算定ルール（＝診療報酬点数表）をつかって算定項目を自動生成
1. を基にした、レントゲン、手術等診療行為における、生成AI活用の机上検証
- 表記ゆれ等の補正の精度検証
- 実験に使用された計算資源・通信資源量の計算（倫理面への配慮）

上記、1.生成AI活用シーン探索の為に基礎実験等は、順天堂大学医学部医学系研究等倫理委員会で審査承認を受けて実施された。実施許可番号:E23-0372-H03

### C. 研究結果

A. AS-ISモデルの調査と整理：分担研究報告書1. 参照（202406031A-buntan1）

B. CANBEモデル検討の為に基礎的調査・実験：分担研究報告書2. 参照（202406031A-buntan2）

### D. 考察

本調査研究では、1.現状の診療・診療報酬算定システムのフローと課題の構造化・概念化を図った。

次に、2.生成AI利活用の可能性を検証するため、範囲を限定して算定ルールに基づく生成AIによる診療報酬計算の自動化の可能性を検証した。

本調査研究で、生成AI利活用の可能性を検証の為に実施した検体検査における実験では、生成AIが診療報酬算定ルールを理解し、検査オーダー項目から診療行為への変換ステップで73.7%～97.1%の正答率を示した。この結果は、限定的な条件ではあるが、生成AIが医療専門用語と診療報酬ルールの関連性を理解できることを示したと考えられる。この自然言語処理能力は、診療報酬改定対応に要する人的リソースと時間的コストの削減に貢献する可能性がある。

医療現場では、2年ごとの診療報酬改定に対応するため、病院システム担当職員や関係ベンダーは①改定内容の理解、医事マスターへの影響範囲の調査・特定、②医事マスターとオーダーコードとの紐付けの見直し、マスターの追加・修正という大きな負担を抱えている。加えて、これらの作業には専門知識が必要であり、限られた期間内に集中して作業する事が求められている。

診療報酬改定時に公開される新旧対照表により変更点が明確に特定されるため、これをLLMに入力することでルールの理解を促進できる可能性がある。具体的には、診療報酬点数表の新旧対照表と医療機関が使用する改定前の医事マスターをLLMに入力することにより、①影響を受ける可能性のある医事マスターの抽出、②影響を受ける具体的根拠の明示、③医事マスターの変更案の提案をLLMに行わせることで、医療機関およびベンダーの診療報酬改定対応作業の負担軽減に寄与することが期待される。



#### **E. 結論**

本研究は、生成AIを診療報酬算定業務に適用する可能性と課題を明らかにした。完全自動化は現状では困難だが、人間と生成AI、従来の機械的プログラムの協働による効率化・自動化の可能性は感じられ、今後の技術発展と運用方法の最適化により、医療関係者の負担軽減と算定精度向上の両立が期待できる。

#### **F. 健康危険情報**

該当なし

#### **G. 研究発表**

該当なし

#### **H. 知的財産権の出願・登録状況**

(予定を含む。)

該当無し



令和6年度厚生労働行政推進調査事業費補助金（厚生労働科学特別研究事業）  
分担研究報告書1

診療行為の構造化と生成 AI 等を活用した標準化されたレセプト作成機能開発  
の為の基礎的調査研究（24CA2031）

研究分担者 齋藤 智之 順天大学健康総合科学先端研究機構バイオリソースリサーチセンター  
非常勤助教

研究要旨：総括報告書参照

C. 結果-1

A. AS-ISモデルの調査と整理

1.複数規模医療機関のシステム構造調査  
(詳細は202406031A-sonota1-5参照)

大規模病院では各部門に個別のシステムがあり、オーダーに対する実施入力には部門システムで行われる。部門システムで入力された情報は電子カルテを経由して医事会計システムに伝達されるが、投薬や指導、処置などはオーダー時に医事会計システムにコスト情報として伝達される。

中規模病院や小規模病院では、大規模病院に比較して一部の部門システムが存在しない場合が増える。有床クリニックや無床クリニックでは、診療所の特色に応じた部門システムが導入されている場合がある。基本的には電子カルテでオーダーが入力され、医事システムにコストとして連携される。無床診療所では病床管理関連が不要となっている。

2. 規模に関わらず使用しているHospital Information System : HIS (病院情報システム) の項目まとめ

医療機関の規模・科に関わらず共通して使用している項目について、本研究では以下の項目と定義した。オーダー等区分:病名・処方・注射・処置・検体検査・生理検査・画像診断・指導料パターン1・指導料パターン2、レセプト対応箇所: 傷病情報・20:投薬・30:注射・40:処置・60:検査・70:画像診断・13:指導管理

3. 診療オーダーから診療報酬請求フロー作成

処方オーダーを例に、各プロセスを以下のように整理した。1.オーダー入力⇒2.医事送信⇒3.医事受信⇒4.診療計算⇒5.レセプト編集⇒6.レセプト確認(詳細は様式A8参照)

4. 既存の自動レセプトチェックシステムのまとめ

A社マニュアルを参考に作成  
生成されたレセプト情報について、正解データ(診療報酬算定ルール等)を基にヒトが作成したマッピングマスターと照合し、プログラムで挙動するチェック機能を備えたアプリケーション。マスターとレセプト情報に齟齬があった場合、エラーを返すシステム。

5. 現状の課題の整理

a. システムベンダーの負荷：  
診療報酬・薬価・保険制度改定等（年数回）に応じ、診療行為・医薬品マスター整備や各種システム（電子カルテ・医事会計等）の改修が必要。DPC・外来調査の仕様変更にも対応。

b. 医師の負荷：  
電子カルテ入力や症状詳記の記載作業が発生。

c. 医事会計系の負荷：  
電子カルテ入力内容と算定の整合確認、病名と検査・処方内容の整合確認のため、医師への問い合わせが必要。



令和6年度厚生労働行政推進調査事業費補助金（厚生労働科学特別研究事業）  
分担研究報告書2

診療行為の構造化と生成 AI 等を活用した標準化されたレセプト作成機能開発  
の為に基礎的調査研究（24CA2031）

研究代表者 八木 摂子 株式会社 FIXER・エンタープライズ部門

研究要旨：総括報告書参照

## C. 結果-2

### B. CANBEモデル検討の為に基礎的調査・実験

1. 生成AI活用シーン探索の為に基礎実験：検体検査における診療報酬請求算定ルール（＝診療報酬点数表）をつかって算定項目を自動生成する試み

#### 【A. 研究目的】

診療報酬請求算定ルールを予め学習させた生成AIが、診療報酬算定ルールに基づき、診療行為等データの標準化（表記ゆれの補正）が可能か、そして、電子カルテ検体検査オーダーから請求点数が計算可能か、実験した。

#### 【B. 研究方法】

##### A. 対象データ

1. 電子カルテ情報：令和4年6月分

株式会社富士フイルム：統合診療支援プラットフォーム：Clinical Intelligence Technology & Architecture (CITA Clinical Finder：電子カルテと連携しているシステム)より抽出

2. 電子レセプト情報：上記期間令和4年6月に対応するデータを準備

3. 検証症例の選定(以下の条件を同時に満たす症例)

・診療報酬算定を踏まえた「病院における一般的な検査」を定義し、これに合致するもの

・レセプト結果(正解データ)と電子カルテオーダー項目がマッチング可能な症例

・診療報酬の算定ルールのひとつである、包括規定の影響を受けないもの

・検査項目が検査実施料として算定されるもの

・トークン制約に抵触しないもの

##### 4. 算定ルール

診療報酬の算定ルールは厚生労働省による「令和4年度診療報酬改定について」にて公開されている、次の告示および通知を基にした。

・令和4年厚生労働省告示第54号「診療報酬の算定方法の一部を改正

・令和4年3月4日保医発0304第1号「診療報酬の算定方法の一部改正に伴う実施上の留意事項について（通知）」

株式会社ケアネットによる運営Webサービス「しろぼんねっと」は以上の公開情報を簡便に参照できるサービスであり、「しろぼんねっと」内の各種資料を参考に、プロンプト入力可能な構造に加工した。

5. レセプト電算処理システム医科診療行為マスター 社会保険診療報酬支払基金ホームページにて公開されている、2024年5月17日版レセプト電算処理システム医科診療行為マスター 全件マスターファイルを使用した。

B. 生成AI：使用モデル：Claude Sonnet 3.5 V1

C. プロンプト設計

Retrieval-Augmented Generation(RAG)を使用すると生成AIのブラックボックス部分にふれる可能性が懸念されている。加えて、生成結果に影響する変数としてプロンプトにRAGを加えることで、生成精度に対する評価や判断、対策が難しくなることが指摘されている。

ファインチューニングは、安定して精度上昇等の効果を見込めるかどうかは不透明とされる。また、ファインチューニングの使用は、コスト等との兼ね合いから、モデルの選択肢が限定されてしまい、採用し得るモデルは現行の主流または最新のモデルと比べ、性能が低い傾向がある。

AIによる生成に必要な全ての情報は、プロンプトに直接入力する事で精度上昇が期待されることもあり、本実験では現状のLLMの精度を測ることを目指し、本実験ではプロンプト設計のみを実行する事で検証を実施した。



#### D. 解析・正答判定

正答判定は以下を同時に満たす場合に、「正答」と判定した。

- 算定された診療行為がすべて正解データと一致すること
  - 正解データに存在しない診療行為が算定されていないこと
  - 各包括計算グループを構成する診療行為が正解データと一致すること
  - 各グループの点数が、出来高計算グループであれば構成診療行為の点数の合計に、包括計算グループであれば構成する診療行為の数から導かれる所定点数に一致すること
  - 合計点数が正解データと一致すること
- 正答解の評価条件に含まないもの: 出来高計算グループはグルーピングに明確な基準はないため、次の点は無問とした。
- 算定されているグループの数が正解データと一致すること
  - 各出来高計算グループを構成する診療行為が正解データと一致すること

#### 【C. 結果】

##### A. 実験対象

実験対象は、補足資料2. 対象選定フローに則って、以下の3症例が選択された(詳細は202406031A-sonota6参照)。

糖尿病ケース: 項目数59・検体種類2

潰瘍性大腸炎ケース: 項目数51・検体種類1

術前検査ケース: 項目数49・検体種類1

また、プロンプト実行の予備調査の為、サンプルケースを設定した。その為、サンプルケースに使用された検体種類は1つとし、検体項目も、外来診療で測定される事が多い項目を20個選択して設定した。サンプルとして選択・使用した項目は以下に記載する: 総蛋白定量・アルブミン・総ビリルビン・GOT・GPT・LDH・ALP・ $\gamma$ -GTP・Na・K・Cl・尿素窒素・クレアチニン・CRP・白血球数・赤血球数・血色素量・ヘマトクリット・血液像・血小板数。

##### B. プロンプト設計

プロンプト設計は、ヒトによる検査オーダーを診療報酬として算定するまでの作業工程を複数のステップに分解し、ステップ単位に具体的な指示を実施するプロンプトを組み立てた(詳細は様式A8参照)。  
1巡目(ステップA～ステップG): 算定ルールに基づき、診療行為および請求点数を算出する。以下に詳細を示す。

◆ステップA: 指示項目⇒診療行為マッピング  
検査オーダーの各指示項目を、AIによる推論を加えながら点数表上の該当する診療行為に変換する。

◆ステップB: 診療行為重複項目の除去  
変換した診療行為に重複がある場合、重複をとりのぞく。

◆ステップC: 診療行為グルーピング  
点数を算出するため、各診療行為をグルーピングする。

◆ステップD: 請求点数計算  
各診療行為のグループについて、点数を決定する。項目数によって算定する検査グループについては、項目数に対応する所定点数を算出する。それ以外のグループについては、構成される診療行為の点数を合計する。

◆ステップE: 算定ルールに照らして、追加算定できる診療行為を算定する。  
検査判断料や検体採取料、加算など、オーダーでは直接指示されない診療行為の追加算定をおこなう。点数表記載の算定要件を参照し、要件を満たしているものを算定する。

◆ステップF: 算定ルールに照らして、算定できない診療行為を消去する。  
ルール上、算定要件を満たしていない診療行為や、互いに同時算定ができない診療行為の組(=背反の関係)がある場合、点数の低い一方を消去する。

◆ステップG: 各グループの明細と点数を示し、最後に合計点数を示す。  
合計点数を算出することで、一連の処理が完了する。



<p>2巡目（ステップH～ステップI）：電子フォーマットによる算定内容の構造化をおこなう（電子レセプトとして診療報酬を審査支払機関に請求するため、電子レセプトファイル(拡張子UKE)のフォーマットで電送をおこなう必要がある為。）</p>	<p>◆ステップD.点数計算結果 ケース1. 潰瘍性大腸炎：項目数算定：100%(正答数233/検証回数233) ケース1. 潰瘍性大腸炎：出来高算定：99.6%(正答数2278/検証回数2288)</p>
<p>◆ステップH: ステップGを経た算定結果と医科診療行為マスター、診療日といった背景情報を入力とし、算定した診療行為をレセプト電算処理コードによって符号化する。</p> <p>◆ステップI: 診療行為の算定日や回数とともに、電子レセプトの構造を反映したYAML形式として出力する。</p>	<p>潰瘍性大腸炎では、項目数算定であれば項目数から点数を算定ルールに基づいて導出し、出来高算定であれば単に点数を合計するという作業は、高い精度で正答が可能であった。</p> <p>◆ステップE. 追加算定結果 検証した検体検査管理加算の算定・検体採取料の算定・外来迅速検体検査加算・判断料の算定では、プロンプト最適化後は概ね正確に追加算定された。</p>
<p>C. 正答率 総合結果を以下に示す。 1巡目（ステップA～ステップG）正答率 ケース1. 潰瘍性大腸炎：79.9%(正答数195/試行回数244) ケース2. 術前検査：6.5%(正答数55/試行回数852) ケース3. 糖尿病：0.6%(正答数2/試行回数314) 参考：サンプル：63.2%(正答数156/試行回数247)</p> <p>2巡目（ステップH～ステップI）正答率 ケース1. 潰瘍性大腸炎58.2%(正答数110/試行回数189) (1巡目で正答が得られた195件から2巡目のプロンプトに不備のあった6件を省いた数) ケース2. 術前検査 74.5%(正答数41/試行回数55) ケース3. 糖尿病：1巡目の正答率が低く未実施 参考：サンプル：73.0%(正答数114/試行回数156) 各ステップ正答率を以下に示す。</p>	<p>◆ステップF. 消去の結果 不規則抗体検査の消去：術前検査症例における「不規則抗体検査」は、一般的な算定要件チェック指示ではほぼすべてのケースで誤って算定された。不規則抗体検査の算定要件を明示的に確認するよう具体的な指示を追加した後は、改善が見られ、多くのケースで適切に消去された。</p>
<p>◆ステップA.検査オーダー項目から診療行為への変換結果 ケース1. 潰瘍性大腸炎：73.7%(正答検査項目数7933/回答検査項目数10761) ケース2. 術前検査：79.9%(正答検査項目数7592/回答検査項目数9506) ケース3. 糖尿病：94.5%(正答検査項目数11681/回答検査項目数12355) 参考：サンプル：97.1%(正答検査項目数4180/回答検査項目数4060)</p> <p>◆ステップC: 診療行為グルーピングと項目数の集計結果 •CRPのグルーピング誤り：出来高算定となるべき「CRP（C反応性蛋白）」を、少なくとも頻度で項目数算定グループに含むという誤判断をした。 ケース1. 潰瘍性大腸炎：98.0%(正答数239/1巡目試行回数244) ケース2. 術前検査：98.6%(正答数710/1巡目試行回数720) (1巡目の試行回数852件からトークン溢れにより出力が途切れた132件を除く) ケース3. 糖尿病(CRP指示無)：正解できず。 (複数検体にまたがるグルーピング：糖尿病症例における「クレアチニン」と「クレアチニン（尿）」を同一グループに含めるべきケースで、ほとんどのケースで誤って別グループに分類した) 参考：サンプル：68.4%(正答数169/1巡目試行回数247)</p>	<p>【D. 考察】 ・実験結果要約 本実験では、ケース間で正答率79.9-0.6%とばらつきがあった。これは、LLMの計算対応能力が症例の複雑性に強く影響を受けていることを示唆していると考えられた。正答率が低かった症例は、より複雑な手続を要すると考えられた。具体的には、算定ルール上の変数が多い症例であり、各ルールについての個別正答は見られても、総合的にすべてを正答できたケースは極めて少なかった。</p> <p>本実験結果は、LLMによるルール判定は、単一の算定ルールに対しては一定の精度が保たれている一方で、複合的・統合的なルールの判定は精度が落ちる可能性を示唆している。</p> <p>・実験結果解釈 本実験におけるLLMによるルール判定では、検査項目から診療行為への基本的変換における高い柔軟性を示していた。これは、ベースとなるLLMに既に一般的な医学情報を計算可能な状態で有しており、プロンプトからインプットされた診療報酬の基本体系を基に適切な計算・判定を導き出せる可能性を示唆している。一方で、「クレアチニン」と「クレアチニン（尿）」の区別など、項目名だけでは決定できない状況では精度が低下した。これは、LLMによる暗黙的な文脈の推測には限界があることを示している。また、「血液化学検査グループ」と「血液学的検査判断料」に代表される、類似した専門用語の文脈依存的な区別についてもLLMの限界が示されたと考えられた。加えて、LLMを診療報酬選定過程全体に適応する事に対して、幾つかの課題が示された。具体的な課題例として、糖尿病ケースにおいて、LLMは血液・尿といった検体横断の判断を正しく解釈できなかった。横断的なカウントの根拠となる算定ルールは、点数表上、複数箇所において断片的に記載されており、適切な関連づけ・統合しての理解が必要となる事が、その理由と考えられる。</p>



二つ目の課題例として、術前検査ケースにおける「不規則抗体検査」の算定要件判断の失敗が挙げられる。ステップA: 検査オーダー指示項目⇒点数表収載診療検査の臨床的背景（術前検査）と算定要件（特定条件行為変換結果を再掲）

下での手術当日のみ算定可）の関連性判断の失敗は、LMの臨床文脈理解の限界を示していると考えられた。

・実験の限界と今後の展望

1. 検証範囲の限定: 単一医療機関における3ケースのみの検証であり、一般化には多様な症例での検証が必要。

2. 実験環境制約: Claude Sonnet 3.5 V1のみを使用し、他LLMとの比較が未実施。プロンプト最適化や代替手法（RAG・ファインチューニング）の検証も検討の余地がある。

3. 限定された適用範囲: 実験範囲が検体検査算定に限定され、診療報酬全般の適用可能性は未検証。加えてトークン制約やコスト・処理速度の評価も不十分。

以上を踏まえ、LLMによる判定・計算精度の改善の為に、次のような検討点が挙げられる。

1. オーダー情報の標準化・構造化など入力情報の質向上

2. 現在の行政文書形式の診療報酬算定ルールを、LLMによる解析が容易な体系的なデータ構造化する。

3. 従来のプログラムによる機械的な構造化データと、LLMによる非構造化データの自然言語処理を組み合わせたハイブリッドアプローチの検討

【実験結語】

1. LLMは検査オーダーからの診療行為導出は基本的に可能だが、追加コンテキスト情報がないと精度が低下し、特に複数解釈可能な検査項目では補足情報が不可欠と考えられた。算定ルールに基づく点数計算は単純なケースで可能だが、複雑なグルーピングや特殊ルールでは困難で、複雑な条件では明示的な指示が必要。

2. 今年度実験を基にした、レントゲン、手術等診療行為における、生成AI活用の机上検証

・画像診断

画像診断は、原則として「撮影料+診断料+部位コメント」の構成で算定する。撮影枚数のように実施時に定まる情報もあるため、指示オーダーに加えて実施結果が算定のためのインプットとして必要となる。

部位コメントについては、診療報酬では「選択的コメント」と呼ばれる仕組みで実現されるが、本研究で検証していないものの、生成AIへの入力に選択的コメントの一覧やマスターを加えることで、生成AIによる適切なコメントの選択および必要事項の追記は十分可能であると想像される。以上より画像診断は、今回の検体検査の検証結果が概ね適用できると考える。

・手術等(手術、輸血)

必要な情報としては、術式や薬剤、材料、監視内容といった実施記録に加え、麻酔記録や輸血記録も必要になり多岐に渡る。加えて、診療報酬上は手術として算定するが、臨床的には手術そのものではない行為があり、それらの判定やインプットは検討が必要と思われる。最後に、手術料は概して高単価であり、更に高額な薬剤や特定保険医療材料の算定が多い。その為、高額な請求漏れや査定減にも繋がりがねないため、他分野に比して高い精度が求められる。以上より、手術等(手術、輸血)は、算定における難度が比較的高い領域であると考えられる。

3. 表記ゆれ等の補正の精度検証

ステップA: 検査オーダー指示項目⇒点数表収載診療行為変換結果を再掲

ケース1. 潰瘍性大腸炎: 73.7%(正答検査項目数7933/回答検査項目数10761)

ケース2. 術前検査: 79.9%(正答検査項目数7592/回答検査項目数9506)

ケース3. 糖尿病: 94.5%(正答検査項目数11681/回答検査項目数12355)

参考: サンプル: 97.1%(正答検査項目数4180/回答検査項目数4060)

LLMによる表記ゆれの補正については、一定の精度を示していると考えられた。

4. 実験に使用された計算資源・通信資源量の計算

Claude 3.5 Sonnet V1は、入力トークン量の上限は20万トークンで、出力トークン量の上限は8,132トークンである。検証に使用した実行環境では、出力トークン上限が4,096トークンであった。

1. サンプルケース: 合計入力トークン量: 159,971・合計出力トークン量: 3,240

2. 潰瘍性大腸炎ケース: 合計入力トークン量: 162,141・合計出力トークン量: 5,269

3. 術前検査ケース: 合計入力トークン量: 165,669・合計出力トークン量: 5,404

4. 糖尿病ケース: 入力トークン量: 37,903・出力トークン量: 2,836

潰瘍性大腸炎ケースで、トークン当たり文字数を算出した。

トークン当たり文字数は、入力0.92字・出力0.97字-1.34字

・入力:  $65,168 \text{ 字} / 71,086 \text{ トークン} = 0.92 \text{ 字}$

・出力:  $3,472 \text{ 字} / 3,683 \text{ トークン} = 0.97 \text{ 字}$

次に関科における算定ルール(診療報酬点数表)の全文をプロンプトとして掲載するために必要なトークン数を概算する。

令和4年度 診療報酬点数表は全4章15部で構成されており、1巡目のプロンプトには第2章第3部「検査」から、主に検体検査に関する算定ルール(注と通知を含む)を部分的に抜粋した。検査の部は点数表コードが264あり、プロンプトに掲載したものはうち11であった。プロンプトの消費トークン量71,086に対し、検査の部全体の所要トークン量は、71,086トークン/プロンプト抜粋点数表コード数11×部全体点数表コード数264=1,706,064と試算できる。続いて、令和4年度 医科診療報酬点数表において、点数表全体930ページ・検査の部146ページ(全体の15.6%)であるので、点数表全体のトークン量は  $1,706,064 / 0.156 = 10,936,308$  と試算された。