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

患者調査の傷病符号付けプロセスの最適化と効率化に資する研究

令和７年度　総括・分担研究報告書

研究代表者　荒井　康夫

（北里大学 未来工学部データサイエンス学科 准教授）

令和８（2026）年５月

目　　次

Ⅰ．総括研究報告書

患者調査の傷病符号付けプロセスの最適化と効率化に資する研究 ---------- 1
荒井　康夫

II. 分担研究報告書

１．標準病名マスターを照合源とする符号照合アルゴリズムを用いた
患者調査の傷病名符号化の最適化と効率化に資する並行照合手段の試作と評価 --------- 21
荒井　康夫　西本　寛

２．自動符号付けアルゴリズムと傷病符号付けチェックの要件
AI を用いた類似度判定についての試行 ---------- 35
阿南　誠

３．ICD 大分類別にみた手付けによる ICD 符号付与の出現要因に関する研究 ---------- 38
桜澤　邦男

４．アンマッチ病名における手付符号付け作業の効率化検討に関する研究 ---------- 44
來島　裕太

５．クラスタリングによる不一致パターン分析に基づく
既存符号付けアルゴリズムの改善に関する研究 ---------- 53
森藤　祐史

６．患者調査と標準病名マスターの符号化方針相違事例における
診療情報管理士による盲検下符号化の予備的検証 ---------- 61
荒井　康夫　　石割　大範

Ⅲ．研究成果の刊行に関する一覧
研究成果の刊行に関する一覧表 ---------- 71

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

患者調査の傷病符号付けプロセスの最適化と効率化に資する研究

研究代表者　荒井　康夫　　北里大学 未来工学部データサイエンス学科 准教授

研究要旨

本研究は、統計法（平成 19 年法律第 53 号）第 2 条第 4 項に規定する基幹統計である患者調査を対象に、傷病名から ICD-10 準拠統計基準のコード（ICD-10 コード）への変換（傷病名符号化）プロセスの最適化および効率化に資する基礎的知見を得ることを目的として、令和 5 年患者調査において得られた傷病名を試料として実施した。患者調査における傷病名符号化プロセスは、辞書照合（一次工程）、形態素解析を活用した手付け符号付け（二次工程）、ならびに手付け符号付け（三次工程）の 3 工程で構成される。

本研究は、当該プロセスを前提とし、その正確性、一貫性、再現性および判定根拠の明示性を高めるための補助的検証手段ならびに基礎的知見を提供することを志向して、5 課題の分担研究を相補的に実施した。具体的には、(1)標準病名マスターを照合源とする符号照合アルゴリズム（PCMA）および照合指数（PCMI）の試作と評価、(2)AI を用いた類似度判定の試行、(3)ICD 大分類別にみた手付けによる ICD 符号付与の出現要因の分析、(4)辞書照合で符号化されなかった病名の傾向分析、(5)クラスタリングによる不一致パターン分析に基づく既存符号付けアルゴリズムの論理的改善策の整理、(6)患者調査と標準病名マスターの符号化方針相違事例における診療情報管理士による盲検下符号化の予備的検証を実施した。

主要な結果として、PCMA の主指標対象 111,984 語における第 1 位候補と患者調査の確定符号との 3 桁分類単位の一致率は 85.09％であった。また、辞書照合で処理されなかった語 75,068 語を 3 層構造に層別化することにより、用手的確認における確認対象の重点化に資する構造を示した。

本研究は、患者調査の確定符号の自動決定を主張するものではなく、AI や機械学習による自動分類を主要方法とするものでもない。本研究の成果は、患者調査における傷病符号付けの適正化、辞書ファイル更新、傷病符号付けチェックプログラムの精緻化、ならびに次期統計処理システム更新方針の検討に資するものであり、基幹統計としての継続性と信頼性を損なうことなく、符号化判断の再現性、説明可能性、ならびに確認可能性を高めるための基礎資料を提供したものである。

研究分担者

西本　寛　　JA 長野厚生連　佐久総合病院
　　総合医療情報センター長

阿南　誠　　川崎医療福祉大学
　　医療福祉マネジメント学部

桜澤　邦男　国立大学法人東北大学
　　東北大学病院メディカル IT センター
　　医療秘書学科 特任教授

准教授
來島　裕太　山口県立総合医療センター
医療企画室　室長補佐
森藤　祐史　大阪府立病院機構
大阪急性期・総合医療センター
医療情報部診療情報管理室 総括主査
石割　大範　国立健康危機管理研究機構
システム基盤整備局 係長

研究協力者

小泉　麻美　北里大学病院　医療の質評価室
診療情報管理士
村上　沙織　北里大学病院　診療情報管理室
診療情報管理士
香取　麻紀子　北里大学病院　診療情報管理室
診療情報管理士

A．研究目的

（1）患者調査の位置づけ

患者調査は、統計法（平成 19 年法律第 53 号）第 2 条第 4 項に規定する基幹統計である患者統計を作成するための調査であり、患者調査規則第 1 条および第 2 条に基づき、医療施設を利用する患者について、その傷病の状況等の実態を明らかにすることを目的として実施される。本調査は、病院および診療所を利用する患者の傷病状況を明らかにする我が国唯一の網羅的な疾病統計であり、その集計結果は、保健医療政策の立案・評価における基礎資料として用いられる。

患者調査では、医療機関から無作為抽出された施設で調査日に受療した患者の傷病情報を、世界保健機関（WHO）が定める疾病及び関連保健問題の国際統計分類第 10 版（ICD-10）に基づき集計する。傷病名から ICD-10 準拠統計基準のコード（ICD-10 コード）への変換（以下「傷病名符号化」という。）の正確性、一貫性、再現性および判定根拠の明示性は、本基幹統計の品質を担保する基盤である。

（2）患者調査における傷病名符号化プロセスの現状と課題

患者調査における傷病名符号化プロセスは、3 つの工程で構成されている。一次工程は、過去調査の蓄積を反映した患者調査にもとづく辞書（以下「傷病名辞書」という。）との照合により ICD-10 コードを付与する工程（以下「辞書照合」という。）である。二次工程は、辞書照合で符号化されなかった語のうち、形態素解析により候補が提示された語に対して手付けを行う工程（以下「形態素解析を活用した手付け」という。）である。三次工程は、形態素解析により候補が提示されなかった語に対して手付け符号化を行う工程（以下「手付け符号化」という。）である。

令和 5 年患者調査において得られた傷病名の種類数約 11 万 2 千語のうち、辞書照合により約 33%（約 3 万 7 千語）、形態素解析を活用した手付けにより約 53%（約 6 万語）が符号化された一方、約 14%（約 1 万 6 千語）については手付け符号化を要した。

患者調査における傷病名は、医療機関が調査票の傷病名欄に自由記述形式で記入する。傷病名には、表記揺れ（異字・誤字・略語・全角半角混在）、修飾語、部位、病態、重症度および複合的記載、複数傷病名の併記が含まれる。また、修飾語のみ、部位のみ、診療科名、後置語、なし表現など、傷病名として成立しない語（非傷病名）も一定数含まれる。これらの多様性および曖昧性が、辞書照合で符号化されない語の発生要因として作用している。

手付け符号化を要する語の発生要因の分析、患者調査用辞書ファイルの要件整理、および自動符号付けアルゴリズムと傷病符号付けチェックの要件

整理は、傷病名符号化プロセスの基礎的検討において継続的に求められる課題である。

（3）国内外の動向と先行研究

傷病符号付けを支援するプログラムの研究開発は、国内外で進展しており、人工知能（AI）や機械学習の活用、クラウドベースのソリューション、自然言語処理、インターフェースの改善など、多様な技術的アプローチが提案されている。これらの動向は、医療情報の標準化と相互運用性の向上に寄与し得るものとして注目されている。

本邦においては、医療機関における傷病名の ICD-10 コード付与に関し、医療情報システム開発センター（MEDIS-DC）が管理する ICD10 対応標準病名マスター（厚生労働省標準規格 HS005、以下「標準病名マスター」という。）の利用が、事務連絡等により推奨され普及してきている。標準病名マスターは、レセプト電算処理システム用傷病名マスターと共通化が図られ、電子カルテおよび電子レセプトに使用されている。また、内閣府の戦略的イノベーション創造プログラム（SIP）「AI ホスピタル研究」では、約 40 万語の同義語を含む病名・症状の用語集が作成されている（標準病名集は約 3 万語）。同用語集の一部は商用利用されつつ、今後も拡張・修正が継続される予定であり、符号付け精度の向上に貢献することが期待されている。一方で、ICD-11 への移行に際しては、WHO が ICD-10 を ICD-11 に変換するテーブルを提供しているものの、付与病名そのものや変換テーブルの精度に課題があると考えられている。

先行研究（厚生労働科学研究費補助金 政策科学総合研究「患者調査の効率的な実施手法の確立に資する研究」、21AB1001）[1][2]では、効率化および精度向上の観点から、現状の符号化プロセスの課題分析を踏まえた情報通信技術（ICT）の活用手法が検討されている。同先行研究の検証によれば、ICT ツールは一定程度の表記揺れに対応可能である一方、一部の表記揺れには対応できず、結果として検索精度の低下が確認されている。また、同先行研究では、患者調査の本来の趣旨が臨床病名での調査であることに鑑み、電子カルテやレセプトの病名を ICT ツールに参考病名として取り込むことは可能であるが、調査票上の傷病名候補として直接表示することは望ましくないとの考察がなされている。

公的統計においては、再現性、説明可能性および判定根拠の明示性が、その品質を担保する基盤として求められる。生成 AI や教師あり機械学習は、学習データ、推論過程、出力の安定性および判定根拠の説明可能性の各観点において、公的統計実務への適用に際して検討すべき課題が残されている。

（4）本研究の目的

本研究の目的は、患者調査における傷病符号付けプロセスを前提として、当該プロセスの最適化および効率化に資する基礎的知見を得ることである。

本研究において「最適化」とは、患者調査における傷病名符号化プロセスを前提として、適合度の高い符号候補を選択するための判断材料を提示し、符号化の正確性、一貫性、再現性および判定根拠の明示性の向上に資することを指す。「効率化」とは、用手的確認における確認対象の重点化、および符号候補の選択に伴う判断負荷の軽減可能性を検討することを指す。なお、所要時間の短縮は本研究の検討対象に含めない。

本研究は、患者調査の傷病名符号化プロセスを変更・代替するものではなく、当該プロセスを尊重しながら、独立した照合ロジックを並行して提供することにより、明示性および再現性のある第二の判定根拠を提示する補助的検証手段としての位

置づけで設計した。本研究は、患者調査において最終的に確定された ICD-10 コード（以下「確定符号」という。）を否定するものではなく、当該符号の確認可能性を高めることに資することを意図している。本研究は、確定符号の自動決定を主張するものではなく、AI や機械学習による自動分類を主要方法とするものでもない。

これらの目的に対応するため、本研究班は 5 課題の分担研究を相補的に配置した。具体的には、（１）標準病名マスターを照合源とする符号照合アルゴリズム（Patient-survey Coding Match Algorithm；患者調査符号照合アルゴリズム、以下「PCMA」という。）および照合指数（Patient-survey Coding Match Index；患者調査符号照合指数、以下「PCMI」という。）の試作と評価、（２）AI を用いた類似度判定の試行、（３）ICD 大分類別にみた手付けによる ICD 符号付与の出現要因の定量的分析、（４）辞書照合で符号化されなかった病名の傾向分析、（５）クラスタリングによる不一致パターン分析に基づく既存符号付けアルゴリズムの論理的改善策の整理、（６）患者調査と標準病名マスターの符号化方針相違事例における診療情報管理士による盲検下符号化の予備的検証、を実施した。

これら 5 課題の分担研究は、令和 7 年度研究計画書に示された 3 つの整理項目――手付け符号化が必要となった傷病名の事例分類と発生要因の分析、患者調査用辞書ファイルの要件整理、自動符号付けアルゴリズムと傷病符号付けチェックの要件整理――を相補的に網羅する構造として設計されている。

（５）期待される効果

本研究の成果は、(1)コーディングの最適化・効率化に向けた基盤構築：PCMA の試作により、標準病名マスターを補助的参照基盤として用い、形態素解析とは異なる照合原理に基づく補助的判定根拠を提示する基盤を構築した。(2)辞書ファイル等を含む傷病符号付けに関する更新の方向性提示：傷病名の表記揺れにすべて辞書登録で対応することには限界がある。一方、近似照合により高い正答率が得られる語については、辞書登録だけに依存せずに符号候補を提示できる可能性が示された。(3)傷病符号付けチェックプログラム精緻化への貢献：PCMA で実装した概念点検の考え方を既存のチェックプログラムに応用し得ることを示すとともに、不一致病名の上位分類およびクラスタ構造を整理し、チェック対象を重点化するための基礎資料を得た。(4)統計調査公表の効率化に向けた基盤形成：PCMA により、手付け符号化を要する語の量的縮減可能性を示し、傷病符号付け作業の効率化を検討するための基礎資料を得た。なお、作業時間は測定していない。(5)次期統計処理システム導入への準備：PCMA を試行的に構築し、標準病名マスターを照合源とする傷病符号照合アルゴリズムの統計処理への適用可能性を示した。本研究は、患者調査における傷病符号付けの検証可能性と再現性を高める技術的基盤を提示した点に意義がある。今後の課題は、独立参照基準の構築および過年度データによる外的妥当性検証である。

以上の目的のもとで、本研究班は令和 5 年患者調査により得られた傷病名を試料として、5 課題の分担研究を実施した。次章以降、研究方法、研究結果、考察および結論を順に述べる。

Ｂ．研究方法

(1) 研究全体の基本方針

本研究は、患者調査における傷病名符号化プロセスを前提として、当該プロセスの正確性、一貫性、再現性および判定根拠の明示性を高めるため

の補助的検証手段ならびに基礎的知見を得る研究として設計した。本研究は、患者調査の既存プロセスを否定、変更または代替することを意図したものではない。

公的統計には、再現性、説明可能性および判定根拠の明示性が求められる。本研究班は、これらの要件と整合的な方法論として、ルールベースの並行照合手段を中核に据えた。生成 AI および教師あり機械学習は、学習データ、推論過程、出力の安定性および判定根拠の説明可能性の各観点において公的統計実務への適用に際して検討すべき課題が残されていることから、本研究では主要方法としては採用せず、補助的試行に限定して位置づけた。

本研究班は 5 課題の分担研究を相補的に配置し、令和 7 年度研究計画書に示された 3 つの整理項目――手付け符号化が必要となった傷病名の事例分類と発生要因の分析、患者調査用辞書ファイルの要件整理、自動符号付けアルゴリズムと傷病符号付けチェックの要件整理――にそれぞれ対応させた。

(2) 患者調査の傷病名符号化プロセスの位置づけ

本研究では、患者調査における傷病名符号化プロセスを、A 節(2)で述べた 3 工程――一次工程としての辞書照合、二次工程としての形態素解析を活用した手付け、三次工程としての手付け符号化――として固定的に位置づける。本研究は、これら 3 工程のすべてを尊重したうえで、各工程と相補的に並行する基礎的知見を提供する設計である。

PCMA および PCMI は、患者調査の確定符号を自動的に決定するものではなく、傷病名に対し、独立した照合ロジックに基づき適合度の高い ICD-10 符号候補を提示し、用手的確認を支援する第二の判定根拠として位置づける。

(3) 対象データ

本研究班は、令和 5 年(2023 年) 患者調査により得られた傷病名を主たる対象とした。各分担研究の対象データの範囲および分母の構成は、(5)各分担研究の方法において個別に示す。

(4) 5 課題の分担研究の構成

(4-1) 4 群構成

本研究班は、5 課題の分担研究を単に並列的に配置するのではなく、研究計画書の目的に沿った相補的構造として、以下の 4 群に整理した。

第 1 群は、標準病名マスターを参照軸とする補助的検証を担う。荒井・西本研究が、標準病名マスターを照合源とする並行照合手段の試作と評価を実施する。

第 2 群は、患者調査の傷病符号付けプロセスにおける課題の可視化を担う。桜澤研究、來島研究、森藤研究の 3 課題が、ICD 大分類別の傾向、ICD1 桁・3 桁別の分布、表記の特徴、ならびに辞書判定コードと手付けコードの不一致パターンの観点から、手付け符号化を要する語の発生構造を多面的に明らかにする。

第 3 群は、自動符号付けアルゴリズムと傷病符号付けチェックの要件整理を担う。荒井・西本研究、阿南研究、森藤研究の 3 課題が、ルールベース照合の試作、AI を用いた類似度判定の試行、ならびに既存符号付けアルゴリズムの論理的改善策の提示を通じて、要件整理に資する基礎的知見を提供する。

第 4 群は、患者調査の符号化方針と標準病名マスター側 ICD-10 コードとの関係整理を担う。荒井・石割研究が、診療情報管理士による盲検下符号化の予備的検証を実施し、両体系の符号化方針差を、文字列照合の結果と専門家評価の双方の観

点から検討する。

これら 4 群は相互に独立ではなく、複数の分担研究が複数の群にまたがる相補的構造となっている。

(4-2) 研究計画書の 3 整理項目との対応

令和 7 年度研究計画書に示された 3 つの整理項目と 5 課題の分担研究の対応関係は、以下のとおりである。

整理項目 1(患者調査用辞書ファイルの要件整理)には、荒井・西本研究、來島研究、森藤研究の 3 課題が対応する。

整理項目 2(手付け符号化が必要となった傷病名の事例分類と発生要因の分析)には、桜澤研究、來島研究、森藤研究の 3 課題が対応する。

整理項目 3(自動符号付けアルゴリズムと傷病符号付けチェックの要件整理)には、荒井研究、阿南研究、森藤研究の 3 課題が対応する。

これにより、5 課題が研究計画書の目的に沿って相補的に実施されたことを示す。

(5) 各分担研究の方法

(5-1) PCMA および PCMI の試作と評価(荒井・西本研究)

本分担研究は、本研究班の中核成果に位置づけられる。標準病名マスターを照合源とする PCMA を試作し、令和 5 年患者調査の傷病名符号化が完了したデータを対象に、患者調査の確定符号との一致率および工程別の振る舞いを評価した。

対象データは、令和 5 年患者調査において得られた種類数（type 数。重複を除いた傷病名の種類数）を単位とする。本分担研究では、評価対象を主指標対象 111,984 語と参考指標対象 110,670 語の 2 分母として構成した原データ 112,133 語からデータ例外 14 語および非傷病名 135 語を除外した 111,984 語を主指標対象とし、ここからさらに ICD 未設定または候補なし 116 語、参照基準不明 15 語(確定符号不明 9 語および索引一致・確定符号不明 6 語)、ならびに符号化方針差 1,184 語(索引一致・参照基準相違 746 語、左右差不問・参照基準相違 438 語)を除外した 110,670 語を参考指標対象とした[注 1]。

[注 1]：分担研究 1 原典の表 2 における左右関連の区分について、本総括報告書では「左右差不問・参照基準相違」に統一する。

照合源には、MEDIS-DC が管理する ICD10 対応標準病名マスター Ver.5.17 を用いた。標準病名マスターを照合源として採用した理由は 3 点ある。第一に、当該マスターは厚生労働省標準規格 HS005 に位置付けられた標準的病名資源であること、第二に、医療情報システムにおいて広く利用される標準的資源であること、第三に、患者調査の過去実績に基づく傷病名辞書とは独立した外部資源であり、第二の判定根拠を提供しうる位置にあることである。

PCMA は、Python によって実装したルールベースの照合アルゴリズムである。生成 AI および教師あり機械学習を主要方法として採用しなかった理由は、(1)に述べた説明可能性および再現性の要請による。PCMA の処理は、以下の 3 段階で構成される。第一段階は正規化である。傷病名に対し、NFKC による文字種正規化、全角・半角の統一、異字・誤字の変換、ならびに修飾語の識別と除去を行う。第二段階は索引用語との照合である。正規化された入力語と標準病名マスターの索引用語との間で照合を行い、上位候補を順位付けする。第三段階は概念点検である。疾病、傷害及び死因の統計分類提要 ICD-10 準拠 第 1 巻（内容例示表）に基づき、候補 ICD-10 コードの医学概念ならびに章・サブ分類との整合性を判定する。

複合的に記述された傷病名については、句読点等を区切りとして並列分割し、先頭パートを優先する簡略ルールを用いた。なお、第 18 章および第 21 章の症状・所見または保健サービス利用の理由が先頭にあり、後続に第 2 章悪性新生物等の特定章の傷病名を含む場合の例外規定を併設した。

PCMA は、傷病名ごとに上位 5 候補を ICD-10 コードを出力する設計とした。これは、用手的確認における確認範囲の絞り込みを支援する意図に基づく。PCMI は、候補の適合度を示す照合指数であり、0 から 1 の範囲の値をとる。PCMI は、8 観点の重み付き線形結合――共通文字率、非共通文字ペナルティ、語順整合度、索引用語長係数、2-gram 一致率、形態素類似度、修飾語タイプ整合度、ならびに修飾語除去後文字集合共通率――を正規化したうえで、ICD-10 章およびサブ分類との整合性を表す互換性係数を指数として適用し、補正項を加えて算出される。PCMI は候補順位付けを目的とした指標であり、候補が患者調査の確定符号と一致する確率を示すものではない。

PCMI の算出に用いる重み係数は、基本観点 8 個、互換性係数指数 1 個および補正項 2 個の計 11 次元である。これらの重み係数は、Optuna に実装された CmaEsSampler による CMA-ES 最適化により決定した。再現性を確保するため、最適化過程の乱数シードは 42 に固定した。なお、CMA-ES は重み係数のバランスの微調整に限定して用いており、アルゴリズム上の構造的問題の解決手段としては用いていない。

参照基準は、患者調査において最終的に確定された ICD-10 コード(以下、参照基準と確定符号を併用する)とした。評価単位は、ICD-10 基本分類表の 3 桁分類とした。PCMI 上位候補の ICD-10 コードと確定符号が同一の 3 桁分類に属する場合を、参照基準と一致した語と定義する。本分担研究では、分担研究 1 原典の用法に従い、本研究上の評価基準に対する一致を意味する語として「正答」「正答率」を限定的に使用するが、これは独立した医学的・分類学的絶対解との一致率を意味するものではない。統計手法として、形態素解析との比較には McNemar 検定(Yates 連続補正、有意水準 $\alpha=0.05$)を用い、各群の一致率に対する 95%信頼区間(以下「95%CI」という。)は Wilson 法により算出した。本分担研究では、文字列照合の結果として候補が選択されたことを意味する「一致」と、当該候補の ICD-10 コードが参照基準の 3 桁分類と同一であることを意味する「正答」を区別して用いる。

(5-2) AI を用いた類似度判定の試行(阿南研究)

本分担研究は、自動符号付けアルゴリズムと傷病符号付けチェックの要件整理を目的として、AI を用いた類似度判定を試行した。対象とした病態は、慢性 C 型肝炎、慢性 B 型肝炎およびアルコール性肝炎の 3 疾患である。これら 3 疾患は、いずれも慢性肝疾患であるという共通点を有しつつ、原因に基づく特有の特徴をも併せ持つことから、類似度判定の試行対象として選定した。

使用した解析ツールは、Gemini および MeCab である。MeCab による形態素解析の後、最多単語数の傷病名との類似度を比較することにより、コーディング精度の向上可能性を試行的に検証した。

(5-3) ICD 大分類別にみた手付けによる ICD 符号付与の出現要因(桜澤研究)

本分担研究は、令和 5 年度患者調査の傷病名について、辞書ファイル等により符号付けが行われた群と、手付けによる符号付けとなった群に区分したうえで、両群の傷病名リストを統合し、分析用データセットを構築した。なお、本分担研究では分析単位を「件」として集計している[注 2]。

[注 2]：本総括報告書では、PCMA 研究文脈での分

析単位を「語」(得られた種類数)とし、本分担研究および後述の來島研究の分析単位を「件」とする。

解析は、ICD 大分類別の集計、Chi-square test、ならびに単変量および多変量ロジスティック回帰分析により行った。解析ソフトウェアは IBM SPSS Ver.30 を用いた。

(5-4) 辞書照合で符号化されなかった病名の傾向分析(來島研究)

本分担研究は、令和 5 年患者調査において辞書照合で符号化されなかった病名 75,218 件を対象に、ICD1 桁(アルファベット)および 3 桁分類別の出現状況を集計し、出現頻度の高い分類への傾向を分析した。本分担研究の分析単位も「件」である。また、新型コロナウイルス感染症関連の傷病名については、表記揺らぎの実態を個別に検証した。

(5-5) クラスタリングによる不一致パターン分析(森藤研究)

本分担研究は、最新の傷病名辞書によるマッチングを経て、形態素解析による参考コードが付与された 59,436 語のうち、手付け符号付け作業により ICD コードが変更された 23,498 語を対象とした。辞書コードと手付けコードを 2 次元ベクトルとしてラベルエンコーディングし、K-means 法によりクラスタリングを行った。クラスタ数の決定には、エルボー法およびシルエット係数に基づく数量的判断と、実務上の診療領域分離状況を踏まえた解釈的判断を併用し、最終的にクラスタ数 8 を採用した。解析はレコード単位 30,020 レコードで実施した。

本分担研究におけるクラスタリングは、実務上の既知の不一致傾向を論理的に整理することを目的とするものであり、自動学習モデルの構築を目的とするものではない。

(5-6) 専門家による盲検下符号化の予備的検証(荒井・石割研究)

本分担研究は、本研究班の前報により抽出された符号化方針相違事例 1,184 件を母集団とし、ICD-10 章別の構成比を反映するように抽出した 124 症例(標本割合 10.6%)を検証対象とした。評価者は、診療情報管理士 4 名(3 組、うち 1 組は 2 名共同付与)である。評価用ワークシートには、入力語、索引用語、正規化後表記、病名表記、ならびに参考 ICD-A および参考 ICD-B を併記し、確認結果列に ICD-10 コードを記入する構成とした。参考 ICD-A と参考 ICD-B は、患者調査 ICD と標準病名マスターICD に対応するが、いずれが患者調査由来であるかは評価者に開示しない盲検評価とした。評価者は、標準病名マスターを参照しない条件下で ICD-10 コードを付与した。主要評価指標は、事例単位(n=124)における 3 組の多数決一致先である。3 組のうち 2 組以上が同一の参考符号と 3 桁レベルで一致した場合に多数決成立とみなし、95%CI は Wilson 法により算出した。副次評価指標として、4 桁完全一致、3 桁一致および ICD-10 章一致の 3 水準で延べ観測を行い、事例クラスタを考慮したノンパラメトリック・ブートストラップ法(2,000 回反復)により分布を推定した。専門家間一致度の指標として、ペアごとの Cohen の κ 係数および 3 組間の Fleiss の κ を算出した。本分担研究では、独立評価者対である A-B ペアの Cohen の κ を主要解釈の対象とし、A-C および B-C ペアならびに Fleiss の κ は参考値として扱う。これは、評価者群 C が 2 名共同付与であることに起因する一致度の偏りを考慮したものである。本分担研究は、仮説検定を行わず、効果量と信頼区間による記述統計的提示にとどめる予備的検証として位置づけた。

(倫理面への配慮)

本研究で取り扱うデータには、個人を直接識別できる情報は含まれていない。本研究は、人を対象とする生命科学・医学系研究に関する倫理指針との関係において、介入を伴わず、かつ個人を直接識別できる情報を含まない既存資料を用いる研究として実施したものであり、当該指針における倫理面の問題は生じない。MEDIS-DC が管理する標準病名マスター Ver.5.17 については、その利用条件に従って研究目的での参照に用いた。分担研究 6 においては、診療情報管理士 4 名（3 組）による盲検下符号化を実施したが、評価者個人を特定可能な形での結果記述を避けるため、専門家 A・B・C の匿名表記とした。

以上が、本研究班の研究方法である。次節「C. 研究結果」では、これら 5 課題の分担研究の結果を、第 1 群の PCMA/PCMI の試作と評価、第 2 群の課題可視化、第 3 群の AI 類似度判定の試行、第 4 群の診療情報管理士による盲検下符号化の順に統合して述べる。

C．研究結果

（1）患者調査の傷病符号付けプロセスにおける課題の可視化

桜澤研究、來島研究および森藤研究の 3 課題は、それぞれの観点から、辞書照合で符号化されなかった語の発生構造を明らかにした。本節では、これら 3 課題の結果を、規模および分布、ICD 大分類別の傾向、ICD1 桁および 3 桁分類別の分布、表記の特徴、ならびに辞書判定コードと手付けコードの不一致パターンの観点から、統合的に示す。

（1－1）規模および分布

桜澤研究は、令和 5 年患者調査の対象全体 112,133 件を、辞書照合により符号化された群（36,915 件、32.9%）と、辞書照合後に手付けを要した群（75,218 件、67.1%）に群別したうえで、両群間の分布特性を分析した。本分担研究では分析単位を「件」として集計した[注 1]。

[注 1]：桜澤研究および來島研究の分析単位は「件」、PCMA 研究および森藤研究の分析単位は「語」である。両者の対象範囲は近接するが、評価単位および抽出条件が異なるため、各分担研究の文脈で扱う。

辞書照合後に手付けを要した群の中で、出現頻度の高い分類は、新生物 19,858 件（17.7%）、損傷および中毒等 18,752 件（16.7%）、筋骨格系および結合組織の疾患 10,760 件（9.6%）、ならびに循環器系の疾患 9,747 件（8.7%）であった。

（1－2）ICD 大分類別の傾向

桜澤研究は、ICD 大分類別に辞書照合後に手付けを要した割合を集計し、Chi-square test および単変量・多変量ロジスティック回帰分析を行った。

対象全体における辞書照合後に手付けを要した割合（67.1%）よりも有意に高い分類として、U コード、Z コード、呼吸器系の疾患、皮膚および皮下組織の疾患、循環器系の疾患、ならびに消化器系の疾患が示された。一方、有意に低い分類として、周産期に発生した病態、先天奇形、耳および乳様突起の疾患、眼および付属器の疾患、神経系の疾患、損傷および中毒等、ならびに新生物が示された。

多変量ロジスティック回帰分析では、辞書照合後に手付けを要する確率を高める影響要因として、U コード（オッズ比 5.963）、Z コード（同 2.594）、呼吸器系の疾患（同 1.231）、皮膚および皮下組織の疾患（同 1.223）、ならびに循環器系の疾患（同 1.153）が導出された。周産期に発

生した病態のオッズ比は 0.601 であり、対象全体の中で最も低い値を示した。

（1－3）ICD1 桁および 3 桁分類別の分布

來島研究は、辞書照合で符号化されなかった病名（アンマッチ病名）75,218 件を対象に、ICD1 桁分類（アルファベット）および 3 桁分類別の出現状況を集計した。本分担研究の分析単位も「件」である。

ICD1 桁分類では、C コード（悪性新生物）が 9,395 件（12.5%）と最も多く、続いて S コード（損傷および中毒等のうち外傷に該当するコード、8,203 件、10.9%）、M コード（筋骨格系および結合組織の疾患、7,248 件、9.6%）、ならびに循環器系の疾患（6,867 件、9.1%）が上位を占めた。これら上位 4 分類で全体の 42.16%を構成した。

ICD3 桁分類では、I63（脳梗塞）が 1,442 件（1.9%）、C34（肺癌）が 1,198 件（1.6%）、C50（乳癌）が 865 件（1.1%）、L30（湿疹）が 760 件（1.0%）、ならびに D48（骨腫瘍）が 750 件（1.0%）の順に多く確認された。

「癌」を含む病名は 6,339 件確認され、そのうち 5,956 件（93.9%）が C コード（悪性新生物）に該当した。C コードに該当しなかった病名の内訳としては、上皮内癌・非浸潤癌 102 件（1.6%）、複数病名による別疾患への分類 177 件（2.8%）、ならびにその他 104 件（1.7%）が含まれた。

（1－4）表記の特徴

來島研究では、辞書照合でアンマッチ病名における側性表現の出現頻度として、「右」12,555 件（16.7%）、「左」12,239 件（16.3%）、「両」5,894 件（7.8%）、「両側」1,925 件（2.6%）、「左側」662 件（0.9%）、「右側」569 件（0.8%）、「左右」128 件（0.2%）、「ミギ」111 件（0.1%）、ならびに「ヒダリ」102 件（0.1%）が観察された（いずれも辞書照合でアンマッチ病名全体に対する割合、重複を許容して集計）。

新型コロナウイルス感染症に関連するアンマッチ病名表記は 916 件確認され、そのうち 715 件（78.1%）が U コードに符号化されていた。U コード以外への符号化は、感染後の病態が 54 件（7.6%）、関連する多系統炎症性症候群が 2 件（0.3%）であった。U コード内訳では、U07.1 が 641 件、U07.2 が 18 件であった。表記には、「コヴィッド 19」、「COVID」、「コロナウイルス感染症 2019」、「新型コロナウイルス感染症」など、略語、カタカナ表記、ならびに数字の組み合わせによる多様な揺らぎが認められた。

なお、標準病名マスター（Ver.5.16）における左右表記に起因する ICD コード変更事例は 24 件にとどまっていた。

（1－5）辞書判定コードと手付けコードの不一致パターン

森藤研究は、最新の傷病名辞書によるマッチングを経て参考コードが付与された 59,436 語のうち、手付け符号付け作業により ICD コードが変更された 23,498 語（約 39.5%）を対象に、辞書コードと手付けコードを 2 次元ベクトルとしてラベルエンコーディングし、K-means 法によるクラスタリングを行った。エルボー法およびシルエット係数（SC=0.52）に基づく数量的判断と、診療領域の分離状況を踏まえた解釈的判断により、クラスタ数 8 を採用した。レコード単位での解析対象は 30,020 レコードであった。

各クラスタの構成は以下のとおりであった。すなわち、Cluster 0（循環器疾患）5,363 語、Cluster 1（外傷）5,841 語、Cluster 2（感染

症）3,864 語、Cluster 3（泌尿器・産科関連）3,321 語、Cluster 4（眼科・精神疾患）2,807 語、Cluster 5（皮膚・筋骨格疾患）5,322 語、Cluster 6（妊娠・管理関連）1,800 語、ならびに Cluster 7（その他）1,702 語である。

クラスタ別の代表的な不一致パターンとしては、循環器疾患領域における「Z 付きコード」の基本コードへの置換、外傷領域における骨折部位の詳細分類による不一致、感染症領域における原因菌・ウイルス種の詳細差、産科領域における妊娠合併症の文脈依存的判断、眼科・精神疾患領域における年齢等の条件分岐を伴う詳細分類の選択、皮膚・筋骨格疾患領域における分類体系および脊椎詳細部位の選択差、妊娠・管理関連領域における R コード（症状・徴候）から特定疾患コードへの置換、ならびに、その他領域における呼吸器疾患の詳細差および COVID-19 関連への修正が観察された。

変更理由のコメントが付与された 784 語の分析からは、「年齢条件（16 歳未満等）によるコード分岐」、「初筆（最初に記載された傷病名）の優先判断」、「部位の明記不足による局所性判断の要否」、ならびに「『？』を含む不明瞭な記載に対する目視確認」が、不一致を生じさせる要因の上位として確認された。

（1-6）群間で共通する所見および相補的所見

3 課題の結果から、以下の所見が共通的および相補的に確認された。

第一に、U コードおよび COVID-19 関連の傷病名は、桜澤研究（多変量ロジスティック回帰における U コードのオッズ比 5.963）、來島研究（U07.1 単独で 641 件、関連表記 916 件）の双方において、辞書照合で符号化されにくい領域として一致して観察された。

第二に、Z コードについても、桜澤研究（オッズ比 2.594）と森藤研究（Cluster 0 における「Z 付きコード」の基本コードへの置換）の双方において、辞書判定と手付けの間に体系的な不一致が観察された。

第三に、悪性新生物（C コード）については、桜澤研究（出現件数 19,858 件、対象全体の 17.7%）、來島研究（辞書照合で符号化されなかった病名における上位、肺癌・乳癌）、ならびに森藤研究（Cluster 2 における腫瘍コード詳細化）の各観点から確認された。「癌」を含む病名の 93.9%が最終的に C コードに分類された一方で、上皮内癌・非浸潤癌の判別、複数病名による別疾患への分類、境界悪性病変、ならびに表記誤りなど、辞書照合に到達しない要因が並列的に存在する構造が示された。

第四に、側性表現（右・左・両等）については、來島研究で多数確認された一方で、標準病名マスター上での ICD コード変更事例は 24 件にとどまっており、左右情報の ICD コーディング全体への影響は限定的である可能性が示された。

これらの所見は、辞書ファイル更新方針および傷病符号付けチェックプログラム精緻化における重点領域として、D 節（考察）において論ずる。

（2）PCMA および PCMI の試作と評価

荒井・西本研究は、本研究班の中核成果として、標準病名マスターを照合源とする PCMA および PCMI の試作と評価を実施した。本節において、分担研究 1 原典の用法に従い「正答」「正答率」を用いる場合があるが、本研究における正答とは、参照基準（患者調査の確定符号）と 3 桁分類単位で一致した候補を指すものであり、独立した医学的・分類学的絶対解との一致を意味するものではない。

（2-1）主指標と工程別の参照基準との一致率

主指標対象 111,984 語における PCMA 第 1 位候補と確定符号との 3 桁分類単位での一致率は 85.09%（95,288／111,984、95%CI 84.88-85.30）であった。参考指標対象 110,670 語における同一致率は 86.10%（同 95%CI 85.90-86.30）であった。

工程別の参照基準との一致率は以下のとおりである。すなわち、辞書照合工程 92.03%（33,945／36,883、95%CI 91.75-92.31）、形態素解析を活用した手付け工程 83.48%（49,594／59,408、95%CI 83.18-83.78）、ならびに手付け符号化工程 74.87%（11,749／15,693、95%CI 74.18-75.54）であった[注 2]。

[注 2]：分担研究 1 原典における一次工程の表記について、本総括報告書では「辞書照合」に統一する。

（2-2）形態素解析との比較

形態素解析を活用した手付け工程の対象における比較を行うため、当該工程の対象 59,408 語のうち、形態素解析側で第 1 位候補が出力された 59,375 語を対象に、両者の出力 ICD-10 コードと確定符号との関係を集計した。

両者の出力が 3 桁分類単位で一致した語は 41,754 語（70.32%）であった。確定符号との一致率は、PCMA が 83.49%（49,575／59,375）、形態素解析が 72.02%（42,764／59,375）であり、その差は＋11.47 ポイントであった（McNemar 検定 $\chi^2 = 3,165.39$、$p < 0.0001$）。この結果は、標準病名マスターを照合源とするアプローチが、形態素解析とは異なる確認手がかりを提供し得ることを示すものである。

（2-3）辞書照合で処理されなかった語の 3 層構造

辞書照合で処理されなかった語は、形態素解析を活用した手付け工程の対象 59,408 語と、手付け符号化工程の対象 15,693 語の合計 75,101 語であり、主指標対象に対して 67.06%を占める。本研究では、これらの語のうち、形態素解析側で第 1 位候補が確認できた 59,375 語と、形態素解析が候補を提示しなかった語に対する手付け符号化対象 15,693 語を合わせた 75,068 語を、形態素解析と PCMA の出力関係に基づき 3 層に層別化した。

第 1 層は、形態素解析と PCMA が一致した符号を提示した語であり、41,754 語（辞書照合で処理されなかった語 75,101 語のうち 55.60%）であった。第 1 層における参照基準との一致率は 93.03%（95%CI 92.78-93.27）であった。第 1 層は、文字列照合と言語処理という異なる照合原理が同一の ICD-10 コードに到達した語であり、用手的確認における判断負荷の軽減可能性を示唆する層である。

第 2 層は、形態素解析と PCMA がいずれも符号を提示するが、両者が一致しない語であり、17,621 語（同 23.46%）であった。第 2 層は、両者の候補を手がかりとして手付け符号化に資する層である。

第 3 層は、形態素解析が候補を提示しなかった語に対する手付け符号化対象であり、15,693 語（同 20.89%）であった。このうち、PCMA が候補を提示した語は 15,652 語、PCMA が候補を提示しなかった語は 41 語であった。第 3 層は、形態素解析が手がかりを与えなかった語に対して、標準病名マスター照合に基づく候補情報を付加できる可能性が確認された層である。

第 1 層 41,754 語のうち、参照基準と一致しなかった語は 2,910 語（6.97%）であった。その内訳は、アルゴリズム改善余地として算出された語 2,171 語（74.60%）、符号化方針差に起因する語

696 語（23.92%）、ICD 未設定の語 28 語（0.96%）、ならびに参照基準不明の語 15 語（0.52%）であった。

（2-4）PCMI 閾値別の参照基準との一致率

PCMI 閾値別の該当語数と参照基準との一致率を集計したところ、閾値が高くなるにつれて該当語数（カバー率）は単調に減少する一方、一致率は閾値 0.50 から 1.00 の範囲で 88.6%から 89.2%の狭い帯域に留まり、単調関係を示す傾向は観察されなかった。この結果は、PCMI が候補順位付けを目的として設計された指標であることを反映しており、PCMI 値そのものを点検対象選別の主指標として用いることは適切でないことを示している。辞書照合で処理されなかった語の層別化は、PCMI 値ではなく、形態素解析と PCMA の出力関係という離散的根拠を主とする。

（2-5）標準病名マスターと患者調査の符号化方針差

PCMA が選択した索引用語と傷病名が完全に一致しながら確定符号と一致しなかった語は 1,184 語確認された。その内訳は、索引一致・参照基準相違 746 語、ならびに左右差不問・参照基準相違 438 語であった[注 3]。これらは文字列照合上の誤りではなく、標準病名マスターにおける ICD-10 付与方針と、患者調査における符号化方針との差に起因するものであり、参考指標対象から除外して扱った。

[注 3]：分担研究 1 原典の表 2 における左右関連の区分について、本総括報告書では「左右差不問・参照基準相違」に統一する。

（3）AI 類似度判定の試行から得られた要件

阿南研究は、自動符号付けアルゴリズムと傷病符号付けチェックの要件整理を目的として、慢性 C 型肝炎、慢性 B 型肝炎およびアルコール性肝炎の 3 疾患を対象に、Gemini および MeCab を用いた類似度判定の試行を行った。

MeCab による形態素解析の結果、「慢性 C 型肝炎」は「慢性／C 型／肝炎」、「慢性 B 型肝炎」は「慢性／B 型／肝炎」、ならびに「アルコール性肝炎」は「アルコール性／肝炎」に分割された。Gemini によるコサイン類似度の計算結果は、慢性 C 型肝炎と慢性 B 型肝炎の間で 0.853、慢性 C 型肝炎とアルコール性肝炎の間で 0.236、ならびに慢性 B 型肝炎とアルコール性肝炎の間で 0.211 であった。慢性 C 型肝炎、ならびに「慢性」を削除した C 型肝炎は、いずれも標準病名マスターに対応する ICD-10 コード B18.2 と整合する結果が得られた。

形態素解析単独では、用語および記述の類似性は把握可能であるが、ICD 分類条件（定義および意味）に基づく判断には限界がある。この限界は、原因菌、ウイルス種、病勢、病態あるいは原因に関する識別を必要とする傷病名において顕在化する。AI あるいは形態素解析は、傷病符号付けに対する補助的手がかりを提供し得る一方で、ICD 分類上の意味判断には限界があり、特に公的統計実務に求められる判定根拠の明示性および再現性の観点から、十分な配慮を要する。

実装上の課題として、(1) 使用する AI ツールの違いによる出力差が生じる可能性、(2) 形態素解析後の単語数が増加した場合の類似度低下の可能性、(3) 同一 ICD に対する複数の標準的傷病名（例えば、短い表現と長い表現）の準備の必要性、ならびに (4) 大量データ処理における時間的負担、の 4 点が確認された。

（4）符号化方針差事例に対する専門家検証

荒井・石割研究は、患者調査の確定符号と標準病名マスター側 ICD-10 コードとが相違する事例（符号化方針相違事例）に対し、診療情報管理士

4名（3組）が標準病名マスターを参照しない条件下で盲検下符号化を行う予備的検証を実施した[注4]。

[注4]：分担研究6では、原報告に従い、対象母集団を「符号化方針相違事例1,184件（索引一致・参照基準相違746件、左右の指定差異による438件）」として扱う。分担研究1（PCMA研究）の文脈における「符号化方針差1,184語」とは、評価単位および抽出条件が異なるため、各分担研究の文脈で扱う。

（4-1）多数決一致先

対象124症例の事例単位における多数決一致先は、標準病名マスター側ICD-10コードが80件（64.5%、95%CI 55.8-72.4）、患者調査の確定符号が34件（27.4%、同20.3-35.9）、多数決不成立が10件（8.1%、同4.4-14.2）であった。

この分布は、本予備標本における選択分布の記述にとどめ、両方針の優劣を示すものではない。なお、評価用ワークシートには参考ICD-Aおよび参考ICD-Bの双方が併記されており、評価者の付与結果の約90%が4桁レベルで参考ICD-AまたはBのいずれかに完全一致していたことから、本検証は実質的に2候補併記下での盲検選択評価としての性格を有する。

（4-2）評価者間一致度

独立評価者対であるA-BペアにおけるCohenのκ係数は、3桁レベルで0.577（中等度の一致）、章レベルで0.778（かなりの一致）であった。A-CおよびB-CペアのCohenのκ係数（3桁レベル0.661-0.694、章レベル0.829-0.837）、ならびに3組間のFleissのκ係数（3桁レベル0.647、章レベル0.819）は、A-Bペアより高い値を示したが、評価者群Cが2名共同付与であることに起因する一致度の偏りの可能性に留意し、参考値として扱う。

延べ観測（最大370観測）における標準病名マスター側ICD-10コードへの一致は246件（66.5%）、患者調査の確定符号への一致は108件（29.2%）であった。観測の独立性が成立しないため、事例クラスタを考慮したノンパラメトリック・ブートストラップ法による95%信頼区間（標準病名マスター側ICD-10コード：60.2-72.2%、患者調査の確定符号：23.5-35.0%）を併記した。

（4-3）探索的所見

事後的に同定された記述統計として、以下の所見が観察された。

第一に、術後状態に関する記述13件において、患者調査の確定符号は原疾患章（第Ⅱ章、第Ⅸ章、第Ⅺ章等）または第ⅩⅧ章に符号化されている一方、標準病名マスター側ICD-10コードは多くの場合、第ⅩⅪ章Z90-Z99に該当するコードに符号化されていた。具体例として、「冠動脈バイパス術後」におけるI25.1とZ95.1、「早期胃癌術後」におけるC16.9とZ08.0、ならびに「前立腺癌術後状態」におけるC61とZ98.8が観察された。

第二に、入力語「コロナウイルス感染症」が1件含まれ、患者調査の確定符号としてはU07.1（COVID-19、ウイルスが特定されたもの）が、標準病名マスター側ICD-10コードとしてはB34.2（コロナウイルス感染症、部位不明）が、それぞれ符号化されていた。本事例においては、3組すべてが標準病名マスター側ICD-10コードを選択した。

これらの探索的所見は、両方針の優劣を示すものではなく、両方針はそれぞれの目的（基幹統計としての疾病別医療需要・受療動向の把握、ならび

に電子カルテ等の臨床用途における手術後の継続的状態の表現）に照らしいずれも妥当な、運用目的の違いに由来する構造的差異として位置づけられる。

D．考察

（1）研究計画との整合性

本研究班は、令和 7 年度研究計画書に示された 3 つの整理項目を、5 課題の分担研究によって相補的に網羅した。すなわち、手付け符号化が必要となった傷病名の事例分類と発生要因の分析は桜澤研究、來島研究および森藤研究が担い、患者調査用辞書ファイルの要件整理は來島研究、森藤研究および荒井・西本研究が担い、自動符号付けアルゴリズムと傷病符号付けチェックの要件整理は荒井・西本研究、阿南研究、森藤研究および荒井・石割研究が担った。

研究計画書の【期待される効果】に示された 5 項目――傷病符号付けの適正化および効率化に向けた基盤構築、辞書ファイル等の更新の方向性提示、傷病符号付けチェックプログラム精緻化、迅速な統計調査公表への基盤形成、ならびに次期統計処理システム導入準備――に対しても、本研究班の成果は具体的に応答している。

（2）患者調査の意義を踏まえた研究の位置づけ

患者調査は、統計法に基づく基幹統計であり、医療施設を利用する患者の傷病状況を明らかにする我が国唯一の網羅的な疾病統計である。基幹統計には、継続性、信頼性、説明可能性および再現性が一貫して求められる。過去調査の蓄積、符号化担当者による判断、ならびに分類運用上の経験は、いずれも基幹統計としての品質を支える要素であり、本研究はこれらを尊重する立場をとった。

このため、本研究は、患者調査の既存プロセスを変更、代替あるいは置換するものではなく、当該プロセスに対する補助的検証手段の試作、ならびに改善に資する基礎的知見の提示として位置づけられる。本研究の成果は、患者調査の確定符号を否定するものではなく、当該符号の確認可能性および判定根拠の明示性を高めるための資料として活用されることを意図したものである。

（3）本研究が示す最適化

本研究における最適化は、単なる一致率の向上を意味するものではない。それは、符号候補の適合度、候補順位の妥当性、判定根拠の明示性、再現性、ならびに一貫性の向上に資するものとして定義される。

PCMA および PCMI は、患者調査の確定符号とは独立した照合ロジックに基づき、文字列照合（PCMI による候補順位付け）と概念点検（ICD-10 内容例示表[4]に基づく医学概念整合性確認）から構成される判定根拠を、明示的かつ再現的に提供する。PCMA の主指標対象における第 1 位候補の参照基準との一致率は 85.09％を示し、辞書照合工程における同一致率は 92.03％であった。これらの結果は、PCMA が、患者調査の確定符号と一定程度整合する形で機能し得ることを示している。

桜澤研究による ICD 大分類別の発生要因分析は、辞書ファイル機能改良の重点領域（U コード、Z コード、皮膚、呼吸器、循環器等）を定量的に提示することにより、辞書ファイル更新の優先順位判断に寄与する。

來島研究による ICD1 桁および 3 桁分類別の分析は、悪性新生物に関する表記、側性表現、ならびに新型コロナウイルス感染症関連表記の整理が、辞書ファイルの網羅性向上に資する具体的方向性を示した。

森藤研究によるクラスタリング分析は、既存の表記揺れ補正処理、目印コード抽出条件、ならびにアンマッチ基礎表における参考情報の充実といった、既存プロセスの論理的改善方向性を、AI や新規技術に依存しない形で提示した。

荒井・石割研究による専門家盲検下符号化は、患者調査と標準病名マスター側 ICD-10 コードとの符号化方針差を、専門家判断という第三の視点から検討する基盤を提供した。

阿南研究は、AI や形態素解析が提供し得る用語類似度の手がかりと、ICD 分類上の意味判断との差異を試行的に示し、公的統計における説明可能性および再現性の重要性を補強する位置づけとして機能した。

（4）本研究が示す効率化

本研究における効率化は、所要時間の短縮を意味するものではない。それは、用手的確認における確認対象の重点化と、符号候補の選択に伴う判断負荷の軽減可能性として定義される。

PCMA の 3 層構造は、用手的確認における確認対象の重点化に資する構造を具体的に示した。

第 1 層（41,754 語、参照基準との一致率 93.03%）は、文字列照合と言語処理という異なる照合原理が同一の ICD-10 コードに到達した語であり、用手的確認における判断負荷の軽減可能性が相対的に高い層として位置づけられる。

第 2 層（17,621 語）は、形態素解析と PCMA の双方が候補を提示するが両者が一致しない語であり、両者の候補を併存する手がかりとして、手付け符号化に資する層である。

第 3 層（15,693 語）は、形態素解析が候補を提示しなかった語であり、このうち 15,652 語に対しては PCMA が候補を提示できた。すなわち、形態素解析が手がかりを与えなかった語の大部分に対して、標準病名マスター照合に基づく候補情報の付加が可能であることが示された。PCMA が候補を提示できなかった 41 語は、標準病名マスターを照合源とする本手法の限界として整理する。

桜澤研究、來島研究および森藤研究の成果は、辞書ファイル機能改良およびチェック機能整備の重点領域を明らかにすることにより、用手的確認対象の重点化に資する基礎情報を提供する。

なお、本研究では作業時間の実測は行っておらず、効率化の効果は実測値ではなく、用手的確認の対象重点化および判断負荷の軽減可能性として解釈される。

（5）辞書ファイル更新方針への示唆

桜澤研究の結果は、辞書ファイル機能改良の優先領域として、U コード、Z コード、呼吸器系の疾患、皮膚および皮下組織の疾患、ならびに循環器系の疾患を提示する。

來島研究の結果は、悪性新生物（C コード）に関する表記整理――上皮内癌・非浸潤癌の判別、複数病名による別疾患への分類への対応、ならびに境界悪性病変の取扱いを含む――が、辞書ファイル更新における重点課題となることを示す。また、側性表現の前処理あるいは正規化が辞書照合に資する可能性が示される一方、片側性および両側性の違いが分類に影響する疾患（変形性膝関節症、鼠径ヘルニア等）における慎重な対応の必要性が併せて示された。新型コロナウイルス感染症関連表記の標準化および同義語統合についても、辞書ファイル更新における対応の方向性として整理される。

森藤研究の結果は、既存の表記揺れ補正処理における「Z 付きコード」の基本コードへの置換ルール、感染症等における原因菌差への対応、ならび

にアンマッチ基礎表における過去修正事例の自動提示等、ルールベースによる辞書ファイル機能改良の具体的方向性を示した。

なお、本研究は、標準病名マスター側 ICD-10 コードに対する是非を論じるものではない。標準病名マスターは、医療情報システムにおける臨床用途を主目的に整備された標準的病名資源であり、その付与方針は当該目的に照らして妥当なものである。本研究の成果は、患者調査の傷病名辞書の更新に資する基礎的知見の提供にとどまる。

（6）傷病符号付けチェックプログラムへの示唆

PCMA の第 1 位候補と確定符号との一致・不一致は、傷病符号付けチェックプログラムにおける確認対象の絞り込みに資する可能性を提供する。特に、PCMA の 3 層構造に基づく重点化は、形態素解析が手がかりを提供しなかった語に対するチェックの優先順位付けに資する。

森藤研究が提示した目印コード抽出条件の精緻化、ならびにアンマッチ基礎表における過去修正事例の自動提示は、チェックプログラム精緻化に直接資する実装方向性である。

荒井・石割研究が示した符号化方針差事例の構造的特徴――術後状態における原疾患章と Z90-Z99 の選択、コロナウイルス感染症事例における U07.1 と B34.2 の選択など――は、チェックプログラムにおける符号化方針差の事前識別および評価対象としての位置づけに資する。

阿南研究が示した AI 類似度判定の可能性と限界は、傷病符号付けチェックにおける AI 活用の位置づけ――補助的手がかりとしての活用と、ICD 分類上の意味判断における人間判断あるいは追加チェック機構との組み合わせ――を明確化する基礎的知見を提供する。

（7）標準病名マスターを照合源とする意義と限界

標準病名マスターは、厚生労働省標準規格 HS005 に位置付けられた、医療情報システムにおける標準的病名資源である。また、レセプト電算処理システム用傷病名マスターと共通化が図られ、電子カルテおよび電子レセプトに広く用いられている。これらの事実は、標準病名マスターを照合源として用いることが、患者調査の傷病名辞書とは独立した外部資源に基づく第二の判定根拠の提供を可能にすることを示している。

一方で、標準病名マスター側 ICD-10 コードは、修飾語の追加、語順の差、一語違い、多部位記載、ICD 未付与、ならびに符号化方針差により、そのまま患者調査の確定符号として適用できない場合がある。本研究において確認された 1,184 語の符号化方針差は、両体系における符号化方針差を検討する必要性を示すものである。

このため、本研究の方法論的立場は、標準病名マスター側 ICD-10 コードを絶対的な参照基準として扱うものではない。標準病名マスターを照合源とする意義は、独立した外部資源に基づく候補提示にあり、その限界は、患者調査の確定符号との関係を構造的に整理することの必要性にある。

（8）AI・機械学習を主要方法としない意義

公的統計には、再現性、説明可能性および判定根拠の明示性が一貫して求められる。生成 AI および教師あり機械学習は、学習データ、推論過程、出力の安定性、ならびに判定根拠の説明可能性の各観点において、本研究の段階では、公的統計実務への適用に際して検討すべき課題が残されている。

このため、本研究班は、AI および機械学習を主要方法としては採用せず、ルールベース照合を中核に据えた。本研究班における AI に関する分担研

究（阿南研究）は、AI や形態素解析が提供し得る用語類似度の手がかりと、ICD 分類上の意味判断との差異を明らかにする位置づけで実施された。また、森藤研究におけるクラスタリング分析は、自動学習モデルの構築を目的とせず、実務上の既知の不一致傾向を論理的に整理する目的で実施された。

これらの設計上の選択は、本研究の方法論的限界ではなく、公的統計実務に必要な説明可能性および再現性を重視した方針として位置づけられる。

（9）研究の限界

本研究の限界として、以下の事項を挙げる。

第一に、本研究は主として令和 5 年患者調査データを対象としており、過年度患者調査データによる外的妥当性検証は今後の課題である。

第二に、PCMA の第 1 位候補一致率は、患者調査の確定符号との一致率であり、独立した医学的・分類学的な絶対解との一致率ではない。患者調査の確定符号は、本研究における参照基準であり、医学的・分類学的な絶対解として扱うものではない。

第三に、効率化について、作業時間の実測は行っていないため、所要時間の短縮効果は本研究では示していない。

第四に、AI 類似度判定の試行は、慢性肝疾患領域の 3 疾患を対象とした限定的な範囲のものであり、一般化にはさらなる検証を要する。

第五に、専門家による盲検下符号化は、124 症例を対象とした予備的検証であり、本格的な独立参照基準の構築には、検証規模および評価設計の拡張を要する。

第六に、標準病名マスターを照合源とする方法には意義がある一方、標準病名マスター側 ICD-10 コードを患者調査の確定符号として単純に扱うことはできない。1,184 語の符号化方針差の存在は、両体系の関係を構造的に整理する必要性を示している。

（１０）今後の課題

今後の課題として、以下の事項を挙げる。

第一に、過年度患者調査データによる外的妥当性検証を進める。

第二に、診療情報管理士複数名による独立判定と評価者間一致度に基づく、独立参照基準の構築を進める。

第三に、ICD4 桁レベル等での要因分析の深掘り、ならびに入院・外来別、病院・診療所別、地域別の層別解析を進める。

第四に、本研究で明らかとなった重点領域に基づき、辞書ファイル更新方針の具体化を進める。

第五に、PCMA および森藤研究が提示した既存プロセスの論理的改善策に基づき、傷病符号付けチェックプログラムの実装要件の整理を進める。

第六に、ICD-11 への移行を見据えて、変換テーブルの整備、ならびに判定根拠の明示性を確保するための基盤整備を進める。

第七に、令和 8 年 1 月から導入予定の次期統計処理システムへの段階的適用可能性の検討を進める。

Ｅ．結論

本研究は、患者調査における傷病名符号化プロセスを前提として、令和 5 年患者調査により得られた傷病名を試料に、5 課題の分担研究により、傷病符号付けプロセスの最適化および効率化に資す

る基礎的知見を得た。

本研究を通じて、手付け符号化を要する語の発生要因（ICD 大分類別の傾向、ICD1 桁および 3 桁分類別の分布、悪性新生物および側性表現に関する表記上の課題、ならびに新型コロナウイルス感染症関連表記の揺らぎ）、辞書判定コードと手付けコードの不一致パターン（K-means 法による 8 クラスタへの整理と、クラスタ別の代表的な不一致パターン）、標準病名マスターを照合源とする並行照合手段（PCMA および PCMI）の有用性と限界（主指標対象における参照基準との一致率 85.09%、3 層構造による確認対象の重点化、ならびに符号化方針差 1,184 語の存在）、ならびに患者調査と標準病名マスターの符号化方針差（術後状態およびコロナウイルス感染症関連事例における方針差の予備的観察、ならびに専門家盲検下符号化における多数決一致先の分布）が明らかとなった。

本研究の成果は、患者調査における傷病符号付けの適正化、辞書ファイル更新、傷病符号付けチェックプログラムの精緻化、ならびに次期統計処理システム更新方針の検討に資するものである。

本研究は、患者調査の既存プロセスを置換するものではなく、基幹統計としての継続性と信頼性を尊重しながら、符号化判断の再現性、説明可能性、ならびに確認可能性を高めるための基礎資料を提示したものである。

Ⅰ．参考文献

1）厚生労働科学研究費補助金 政策科学総合研究「患者調査の効率的な実施手法の確立に資する研究（21AB1001）」令和 3 年度報告書.

2）厚生労働科学研究費補助金 政策科学総合研究「患者調査の効率的な実施手法の確立に資する研究（21AB1001）」令和 4 年度報告書.

Ｆ．健康危険情報

該当なし。

Ｇ．研究発表

1．論文発表

学術誌投稿準備中（荒井・西本研究）。

その他の分担研究については、本年度の論文発表は該当しない。

2．学会発表

阿南研究は、2026 年度日本医療マネジメント学会学術総会において発表予定である。

その他の分担研究については、本年度の学会発表は該当しない。

Ｈ．知的財産権の出願・登録状況（予定を含む。）

1．特許取得

該当なし。

2．実用新案登録

該当なし。

3．その他

該当なし。

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

標準病名マスターを照合源とする符号照合アルゴリズムを用いた患者調査の傷病名符号化の最適化と効率化に資する並行照合手段の試作と評価

研究分担者　荒井　康夫　北里大学 未来工学部データサイエンス学科　准教授
研究分担者　西本　　寛　JA長野厚生連 佐久総合病院 総合医療情報センター長

研究要旨

患者調査は、統計法に基づく基幹統計である患者統計を作成し、医療施設を利用する患者の傷病状況等を明らかにするための調査である。患者調査では、医療機関が調査票の傷病名欄に記入した傷病名に ICD-10（疾病及び関連保健問題の国際統計分類第 10 版）コードが付与される。傷病名は自由記述形式で記入されるため、表記揺れや構造的多様性を含む。このため、辞書照合で処理されず用手的確認を要する語が発生する。本研究は、患者調査の傷病名符号化プロセスを前提として、用手的確認を支援する補助的検証手段の試作・評価を目的とした。一般財団法人医療情報システム開発センター（MEDIS-DC）が公開する ICD10 対応標準病名マスターを照合源とし、符号照合アルゴリズム PCMA（Patient-survey Coding Match Algorithm；患者調査符号照合アルゴリズム）および照合指数 PCMI（Patient-survey Coding Match Index；患者調査符号照合指数）を試作した。令和 5 年患者調査で確定した ICD-10 準拠統計基準コードを参照基準とし、3 桁分類単位で一致した場合を本研究における正答と定義した。主指標対象 111,984 語における PCMA 第 1 位候補の確定符号との一致率は 85.09％であった。本研究の正答率は確定符号との一致率であり、独立した医学的正解との一致率ではない。形態素解析と比較可能であった 59,375 語では、PCMA は形態素解析とは異なる確認手がかりを提供し得ることが示された。辞書照合で処理されなかった語 75,101 語のうち、3 層化可能であった 75,068 語を形態素解析と PCMA の出力関係により整理し、用手的確認の対象重点化に資する構造を示した。PCMA は確定符号を自動決定するものではなく、候補提示、判定根拠の明示、判断負荷の軽減可能性に資する補助的検証手段として位置づけられる。独立参照基準の構築および過年度データによる外的妥当性検証は今後の課題である。

A. 研究目的

患者調査は、統計法に基づく基幹統計であり、医療機関から無作為抽出された施設で調査日に受療した患者の傷病情報を、世界保健機関（WHO）が定める『疾病及び関連保健問題の国際統計分類（ICD）第 10 版』に基づき集計する[1]。傷病名から ICD-10 コードへの変換（以下「傷病名符号化」）の正確性と再現性は、本基幹統計の品質を担保する。

患者調査において、医療機関が調査票（紙または電子）の傷病名欄に記入する傷病名は、自由記述形式である。患者調査では、紙調査票では傷病名欄に傷病名が自由記述され、電子調査票では電子カルテ情報等から得られた情報を登録する運用も想定さ

れている。記入される傷病名は、診療上の傷病概念を表現する自由文であるため、表記揺れ（異字・誤字・略語・全角半角混在）、構造的多様性（非傷病名・並列記載・因果関係記載）、否定・両側性・多発性等の医学的属性を含む。本研究において「非傷病名」は、修飾語のみ、部位のみ、診療科名、後置語、なし表現などで構成され、傷病名として成立しない語を指す。このため傷病名辞書との照合のみでは大部分の語を符号化できず、形態素解析または手付けによる符号化を要する。患者調査の傷病名符号化プロセスは、辞書符号付け、形態素解析が候補を提示した語に対する手付け、形態素解析により候補が提示されなかった語に対する用手的な手付け符号化の 3 工程で構成される。

手付け符号化は、ICD-10 索引表および疾病、傷害及び死因の統計分類提要 ICD-10 準拠第 1 巻（内容例示表）を参照しながら、自ら ICD-10 コードを探索する学術的かつ調査的な作業である。手付け符号化の対象範囲では、形態素解析の出力（候補提示の有無、提示された候補の妥当性）が当該作業の手がかりの有無を規定する。形態素解析が候補を提示しなかった語に対する手付けは、手がかりのない状態で行われる。

本研究の目的は、患者調査の傷病名符号化プロセスに対する並行照合の手段として、標準病名マスターを照合源とする符号照合アルゴリズム PCMA（Patient-survey Coding Match Algorithm）および照合指数 PCMI（Patient-survey Coding Match Index）を試作することである。試作後、令和 5 年患者調査の出現した傷病名を試料として、当該プロセスの最適化および効率化への資する程度を評価する。本研究において「最適化」は、患者調査における傷病名符号化プロセスを前提として、標準病名マスターとの照合結果に基づき、適合度の高い符号候補を選択するための判断材料を提示し、用手的な確認を支援することで、符号化の正確性、判定根拠の明示性および再現性の向上に資することを指す。一方、「効率化」は、目視作業を要する語の量的縮減可能性および判断負荷の軽減を指し、所要時間の短縮は本研究の検討対象に含めない。本研究は、患者調査プロセスを変更・代替するものではない。確定符号とは独立した照合ロジックを並行して提供することにより、第二の判定根拠を明示性・再現性のある形で付与する。これにより、手付け符号化が必要な語に対する調査・探索作業の軽減可能性を提供する。

B. 研究方法

1）データソースと評価分母

対象データは、令和 5 年患者調査の傷病名である[2]。語数の単位は得られた種類数（type 数。重複を除いた傷病名の種類数）であり、延べ出現数とは区別する。原データ 112,133 語から評価上の除外語を控除し、主指標対象 111,984 語、参考指標対象 110,670 語の 2 つの評価分母を定義した（除外項目の詳細は表 2）。PCMA の第 1 位候補とは、PCMI が最上位の候補を指す。

2）PCMA の構成と照合源

PCMA は、一般財団法人医療情報システム開発センター（MEDIS-DC）公開の ICD10 対応標準病名マスター（Ver.5.17[3]、以下「標準病名マスター」）を照合源とする Python 実装のルールベースアルゴリズムである。標準病名マスターは臨床現場で用いられる病名集であり、傷病名の臨床記載と ICD-10 を架橋する照合源として整合する。

標準病名マスターを照合源として採用した理由は、次の 3 点である。第一に、本マスターは厚生労働省標準規格 HS005 に位置づけられた公的標準であり、医療現場で広範に利用されている。第二に、各標準病名に ICD-10 準拠統計基準のコード（ICD-10 コード）が直接付与されている。本マスターを照合源

とすることにより、文字列照合の結果から ICD-10 コードの候補を直ちに提示できる構造とした。第三に、本マスターは臨床で用いられる病名表現を網羅している。これにより、本マスターを照合源とする照合の到達範囲を確保した。なお、本マスターは、基本病名に修飾語を付加することにより、多様な病態表現に対応する構造を有する。

本研究において、生成 AI および機械学習を主要な照合手段として採用せず、ルールベース照合を選択した理由は、次の 3 点に集約される。第一は再現性である。基幹統計は経年比較を前提とする。このため、同一入力に対して同一出力が得られる必要がある。第二は説明可能性である。基幹統計の符号化では、判定根拠が第三者により検証可能でなければならない。第三は判定根拠の明示性である。本研究のルールベース照合では、照合ルールと点検ルールが判定根拠の明文化に直接対応している。これらのルールは、規程化・公開・改訂が可能である。

PCMA の処理は、(1) 正規化（Unicode 正規化 NFKC、全角・半角統一、異字誤字変換、修飾語の識別と除去）、(2) 標準病名マスターの索引用語との照合および PCMI 算出、(3) 上位候補の順位付け、(4) 概念点検（ICD-10 内容例示表に基づく医学概念と章・サブ分類の整合性判定）、の 4 段階で構成される。複合病名は句読点等を区切りとして並列分割し、先頭パートを優先する。ただし先頭パートが第ⅩⅧ章（症状・徴候）または第ⅩⅪ章（保健サービス）であり、かつ後続に第Ⅱ章（悪性新生物）等の確定診断名が含まれる場合は、後者を優先する。各傷病名に対し、PCMI が高い順に上位 5 語を候補として出力する。本研究では、PCMA が上位 5 語までを提示する設計とした。これにより、最上位候補のみを返すのではなく、後段の手付け符号化における判定の選択肢を確保した。本研究の評価では、第 1 位候補に紐づく ICD-10 コードを PCMA の判定結果とした（表 3）。PCMA は原則として全評価対象に対して候補を提示できる設計であるが、(a) データ例外（記号・数字・空文字のみで構成される語、14 語）、(b) 非傷病名（135 語）、(c) 標準病名マスターに収載されるが当該病名単独では ICD-10 コードが一意に定まらない語または PCMA の照合で有効な候補が選定されなかった語（ICD 未設定、116 語）の 3 区分に該当する語については、構造上候補を提示しない。これら計 265 語のうち、データ例外および非傷病名は主指標対象から既に除外され、ICD 未設定は参考指標対象から追加除外される（表 2）。

標準病名マスターを照合源として用いる際の注意点は、次の 2 点である。第一に、修飾語が付加された病名表現に対して、基本病名に付与された ICD-10 コードが必ずしも妥当ではない場合がある。第二に、文字列近似照合は文字一致度を最大化する候補を選ぶ。これは、臨床概念上の整合性を保証するものではない。文字列の一部相違や語順差により、文字列上は近いが臨床的には別概念である候補が選ばれる余地が残る。本研究では、これらの注意点に対応するため、照合（PCMI による候補順位付け）と概念点検（ICD-10 内容例示表に基づく整合性確認）を別工程として配置した。両工程の分離により、文字列上は高一致であっても臨床概念上に矛盾を含む候補を識別する経路を確保した。

複数部位の記載を含む傷病名に対しては、単一部位の標準病名に紐づく標準病名マスター側ICD-10 コードでは妥当な符号を選択できない場合がある。本研究では、これらの語に対して多部位対応の符号への置換を試行的に導入した。

3）PCMI と重み係数の最適化

PCMI は、出現した傷病名と候補病名の類似度を 0〜1 の範囲で表す指標である。共通文字率、非共通ペナルティ、語順整合度、索引用語長係数、2-gram 一致率、形態素類似度の 6 観点の重み付き線形結合に、ICD-10 章・サブ分類の整合性を表す

互換性係数 f_compat を乗じ、理論最大値で正規化する。

$$PCMI = clamp(\Sigma w_i s_i / S_max \times f_compat^{w_7}, 0, 1)$$

重み係数 w_1〜w_8は、Optuna の CmaEsSampler を用いた共分散行列適応進化戦略（CMA-ES）[5][6]により、第 1 位候補の正答率を最大化する組合せを探索した（乱数シード固定）。PCMI の構成要素と算出例を表 4 に示す。

4）評価指標

参照基準は、患者調査の傷病名符号化プロセスで付与された確定 ICD-10 準拠統計基準コード（以下「確定符号」）である。正答は、PCMA の第 1 位候補の ICD-10 コードが確定符号と ICD-10 における 3 桁分類以上で一致した場合と定義した。ICD-10 における 3 桁分類は、厚生労働省「疾病、傷害及び死因の統計分類」の基本分類表と整合する公的統計の主要な表章単位である。本研究の正答率は確定符号との一致率であり、独立した医学的正解との一致を評価するものではない。

形態素解析が候補を提示した語に対する手付けの対象における PCMA と形態素解析との比較には、McNemar 検定（Yates 連続補正、α=0.05）を用いた。95%信頼区間は Wilson 法により算出した。実装は Python 3.13 を用い、主要ライブラリは pandas、numpy、SudachiPy、Optuna である。

本研究において「一致」と「正答」を次のように区別する。「一致」は、2 つの出力（PCMA の出力と形態素解析の出力、または PCMA の出力と患者調査の確定符号）が同一の ICD-10 コードを指す状態を表す。「正答」は、PCMA の第 1 位候補の ICD-10 コードが患者調査の確定符号と ICD-10 における 3 桁分類以上で一致した場合を指す。

5）評価の二段構成と当初設計

本研究は当初、手付け符号化を要する語（辞書符号付けで処理されなかった語、75,101 語）に対する PCMA の正答率評価を想定して設計した。そのなかで、辞書符号付け（患者調査の過去実績で正答として確定した範囲）に対する評価は、PCMA の性能を確認する明確な参照点として併せて試行した。研究実施中、PCMA の選択した索引用語と出現した傷病名が完全一致しながら、確定符号と一致しない事例を確認した。これにより、標準病名マスターと患者調査の符号付け基準の間に運用方針の差異（以下「符号化方針差」）が存在することが判明した。これは PCMA の文字列照合上の誤りではなく、両体系の符号付け基準の差異に起因する。

この差異への対処として、本研究では評価分母を二段構成とした。主指標対象（111,984 語）は出現した傷病名から評価上の除外語のみを控除したもの、参考指標対象（110,670 語）はそこから符号化方針差および ICD 未設定・参照基準不明を追加除外したものである（表 2）。符号化方針差の事例は PCMA の誤りでなく両体系の運用方針差として扱う。

（倫理面への配慮）

本研究は患者調査の出現した傷病名のみを対象とし、個人を直接識別できる情報を含まない集計済み統計データの二次的解析である。本研究は「人を対象とする生命科学・医学系研究に関する倫理指針」の研究定義に該当しないと判断し、所属機関倫理審査委員会への申請は行わなかった。標準病名マスターは、MEDIS-DC への利用申請および利用許諾の取得のうえで使用した。

C. 研究結果

1）主指標と工程別正答率

主指標対象（111,984 語）における PCMA 第 1 位候補の正答率は 85.09%（95,288／111,984、95%

CI：84.88–85.30）であった。工程別の正答率は、辞書符号付けで 92.03%、形態素解析が候補を提示した語に対する手付けで 83.48%、手付け符号化で 74.87%であった（表 1、表 5）。

2）形態素解析が候補を提示した語に対する手付けの対象における PCMA と形態素解析の比較

形態素解析が候補を提示した語に対する手付けの対象 59,408 語のうち、形態素解析側で第 1 位候補が出力された 59,375 語（差 33 語は比較不能のため除外）について、両者の出力 ICD-10 コードと確定符号との関係を集計した（表 6）。両者の出力が ICD-10 における 3 桁分類以上で一致した語は 41,754 語（70.32%）であった。確定符号との正答率は、PCMA が 83.49%、形態素解析が 72.02%であり、PCMA が 11.47 ポイント上回った（McNemar 検定 χ^2=3,165.39、p<0.0001）。

3）手付け符号化における判断負荷を軽減する可能性がある語の 3 層構造

辞書照合で処理されなかった語（形態素解析が候補を提示した語に対する手付けの対象 59,408 語と手付け符号化の対象 15,693 語の合計 75,101 語、主指標対象に対し 67.06%）を、「手付け符号化における判断負荷を軽減する可能性がある語」と定義した。当該語における形態素解析と PCMA の出力関係に基づき、3 層に層別化した（表 7）。

第 1 層は、形態素解析と PCMA が一致した符号を提示した語（41,754 語、当該語のうち 55.60%）である。形態素解析（言語処理アプローチ）と PCMA（標準病名マスター照合アプローチ）という異なるアプローチが同一の ICD-10 コードに到達した語であり、当該層の正答率は 93.03%であった。

第 2 層は、形態素解析と PCMA がいずれも符号を提示するが両者が一致しない語（17,621 語、当該語のうち 23.46%）である。両者を手がかりとして手付け符号化に資する語であり、PCMA 側は PCMI が候補病名の文字列上の近似度を表す。

第 3 層は、形態素解析が符号を提示せず、PCMA のみが符号を提示する語である。形態素解析が手がかりを与えなかった語にも PCMA が候補を提示できることにより、手がかりがなかった語における手付け符号化を軽減する可能性が生じる。PCMA 側は PCMI が候補病名の文字列上の近似度を表す。

4）当該語における不正答内訳

第 1 層 41,754 語における不正答は 2,910 語（6.97%）であった。当該不正答の内訳は、アルゴリズム改善余地（残余として算出）が 2,171 語（74.60%）、符号化方針差に起因する語が 696 語（23.92%）、ICD 未設定の語が 28 語（0.96%）、参照基準不明の語が 15 語（0.52%）であり、残存リスクの大部分は PCMA のアルゴリズム改善余地に帰属する（表 10）。

5）PCMI 閾値別の正答率

PCMI の閾値別の該当語数と正答率を集計した（図 1）。閾値が高いほど該当語数（カバー率）は単調に減少する一方で、正答率は閾値 0.50〜1.00 の範囲で 88.6〜89.2%の狭い帯域に留まり、明確な単調関係は確認されなかった。この結果は、PCMI が候補順位付けの指標として最適化されていることを反映する。したがって、PCMI 値そのものを点検対象選別の主指標として用いることは適切でない。手付け符号化における判断負荷を軽減する可能性がある語の層別化は、PCMI 値ではなく「形態素解析と PCMA の出力関係」という離散的根拠を主とする。

6）標準病名マスターと患者調査の符号付け

PCMA が選択した索引用語と出現した傷病名が完全一致しながら、確定符号と不一致であった語は 1,184 語（符号化方針差と総称）であった。内

訳は、索引一致・参照基準相違 746 語、左右差不問・参照基準相違 438 語である。これらは文字列照合上の誤りではなく、標準病名マスターと患者調査の符号付け基準の差異に起因する。これらの語は参考指標対象から除外した。

D. 考察

本研究の主要所見は次の 3 点である。第一に、PCMA が患者調査の確定符号に対し、主指標対象における第 1 位候補正答率 85.09%を示したこと。第二に、形態素解析が候補を提示した語に対する手付けの対象における直接比較で、PCMA の正答率が形態素解析を 11.47 ポイント上回ったこと。第三に、手付け符号化における判断負荷を軽減する可能性がある語（75,101 語）を 3 層に層別化し、形態素解析と PCMA が一致した符号を提示する第 1 層（41,754 語、当該層正答率 93.03%）を識別したことである。

（最適化への資する程度）

PCMA は、確定符号とは独立した照合ロジックに基づき、全評価対象に対して判定結果を提示できることが確認された。判定根拠は文字列照合（PCMI による候補順位付け）と概念点検（ICD-10 内容例示表に基づく医学概念整合性確認）から構成され、明示性・再現性のある形で第二の判定根拠を提供する。辞書符号付けで確定した語に対する PCMA の高い正答率は、PCMA が辞書による自動付与とほぼ整合する判定を再現し得ることを示す。一方で、標準病名マスターと患者調査の符号付け基準の差異（符号化方針差）が研究実施中に判明し、PCMA の文字列照合上の正しさが両体系の運用方針差により評価される構造的制約が確認された。

（効率化への資する程度）

手付け符号化は、ICD-10 索引表および ICD-10 内容例示表を参照しながら、自ら ICD-10 コードを探索する学術的かつ調査的な作業である。PCMA の候補提示は、当該作業に対する手がかりを与える。第 1 層（形態素解析と PCMA が一致した符号を提示する語）では、異なるアプローチが同一の符号に到達したことが手付け符号化を軽減する可能性の根拠となる。第 2 層では両者の候補が手がかりとして併存し、PCMA 側は PCMI が候補病名の文字列上の近似度を表す。第 3 層は、形態素解析が候補を提示しなかった語、すなわち手がかりがなかった語である。PCMA は当該語にも候補を提示でき、PCMI が候補病名の文字列上の近似度を表すことにより、手がかりがなかった語における手付け符号化を軽減する可能性が生じる。本研究は作業時間を測定していないため、効率化の効果は実測値ではなく、目視作業を要する語の量的縮減可能性および判断負荷の軽減として解釈する。

（辞書符号付けに対する本研究の立場）

辞書符号付けは患者調査の過去実績において正答として確定した範囲であり、本研究はこれに対して「別のアプローチ」を要するものとして提示するものではない。本研究において「異なるアプローチ」とは、形態素解析（言語処理アプローチ）と PCMA（標準病名マスター照合アプローチ）の対比のみを指す。当初は手付け符号化を要する語に対する PCMA の正答率評価を想定し、そのなかで辞書符号付けに対する評価は PCMA の性能を確認する参照点として併せて試行した。研究実施中、PCMA の選択した索引用語と出現した傷病名が完全一致しているにもかかわらず確定符号と一致しない事例から、標準病名マスターと患者調査の符号付け基準の符号化方針差が判明した。このため評価分母を二段構成とし、符号化方針差事例を主指標対象には含むが参考指標対象からは除外した。

（本研究の限界）

本研究の限界として、以下の 5 点を挙げる。

第一に、CMA-ES の目的関数と最終評価指標がいずれも令和 5 年患者調査の同一データに基づくものであり、hold-out 評価および過年度データを用いた外的妥当性検証は今後の課題である。

第二に、参照基準が本研究の対象とする傷病名符号化プロセスの出力そのものであり、本研究の正答率は再現性指標であって、独立した医学的正解との一致率ではない。

第三に、参考指標対象（110,670 語）の追加除外項目は研究結果から事後的に同定したカテゴリであり、研究者の解釈に依存する自由度を含む。主指標は 85.09%（111,984 語）に固定し、参考指標は感度分析の参考値として扱う。

第四に、形態素解析との直接比較は両者の第 1 位候補同士の対比であり、形態素解析が候補を提示した語に対する手付けの最終結果との比較ではない。

第五に、評価単位は得られた種類数であり、患者調査の集計（延べ出現数）への直接的影響を表現するものではない。延べ出現数による加重正答率の評価は今後の課題である。

（PCMA の試行的工夫の到達と残存する構造的課題）

本研究では、文字列照合の限界に対応するため、複合的な近似評価、候補の複数提示、照合と概念点検の分離、ならびに複数部位記載に対する多部位対応の符号への置換を試行的に導入した。これらの工夫により、出現した傷病名に対して補助的判定根拠を提示する基盤を試作した。一方で、標準病名マスターを照合源とする方法には構造的課題が残る。第一に、標準病名マスターに ICD-10 コードが付与されていない病名が存在する。第二に、修飾語付加により ICD-10 コードの妥当性が変動する。第三に、標準病名マスターと患者調査の符号化方針には差異がある。これらの構造的課題は、本研究の試行的工夫では完全には解消されない。

（今後の方向性）

今後の方向性として、(1) 診療情報管理士複数名による独立判定と評価者間一致度に基づく独立参照基準の構築、(2) 過年度データによる外的妥当性検証、(3) 標準病名マスターと患者調査の符号付け基準間の対応関係整備による符号化方針差の構造的理解、(4) 編集距離を考慮した近似照合による未登録誤字・略語への対応強化、(5) 形態素解析が候補を提示しなかった語のうち PCMA も候補を提示しなかった語に対する代替的照合手段の検討、を挙げる。

E. 結論

令和 5 年患者調査の主指標対象（111,984 語）における PCMA 第 1 位候補と確定符号との ICD-10 における 3 桁分類以上での一致率は 85.09%であった。辞書符号付けでの正答率は 92.03%、辞書照合で処理されなかった語（75,101 語）における PCMA の正答率は 81.68%であった。このうち形態素解析が候補を提示し PCMA と直接比較可能であった 59,375 語では、PCMA の正答率が形態素解析を 11.47 ポイント上回った（p<0.0001）。手付け符号化における判断負荷を軽減する可能性がある語（75,101 語）を 3 層に層別化し、形態素解析と PCMA が一致した符号を提示する第 1 層（41,754 語、当該層正答率 93.03%）を識別した。

以上より、PCMA は、傷病名符号化の最適化（正確性の担保と判定根拠の明示性・再現性）および効率化（辞書照合で処理されなかった語、とくに形態素解析が手がかりを与えなかった語に対する手付け作業の調査・探索負担の軽減可能性）に資する並行照合の手段として機能し得る。本指標は確定符

号との一致率であり、独立参照基準の構築および過年度データによる外的妥当性検証は今後の課題である。

F. 健康危険情報

特になし。

G. 研究発表

1）論文発表

学術誌投稿準備中。

2）学会発表

該当なし。

H. 知的財産権の出願・登録状況

特許取得、実用新案登録、その他いずれも該当なし。

I. 参考文献

1) World Health Organization. International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10). Geneva: WHO; 2019.

2) 厚生労働省政策統括官付参事官付保健統計室. 患者調査の概況(令和 5 年). 厚生労働省; 2024.

3) 一般財団法人医療情報システム開発センター (MEDIS-DC) . ICD10 対応標準病名マスター Ver.5.17. 2024.

4) 厚生労働省. 疾病、傷害及び死因の統計分類 (ICD-10 (2013 年版)) 内容例示表. 厚生労働省; 2016.

5) Hansen N, Ostermeier A. Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation. 2001;9(2):159–195.

6) Hansen N. The CMA evolution strategy: a tutorial. arXiv:1604.00772. 2023.

表 1．主指標対象における傷病名符号化工程別の内訳

工程	語数（n）	割合（%）	工程の内容
辞書符号付け	36,883	32.94	傷病名辞書との照合により符号が付与された語
形態素解析が候補を提示した語に対する手付け	59,408	53.05	辞書照合で不一致となった語のうち、形態素解析により候補が提示された語に対する手付け
手付け符号化	15,693	14.01	形態素解析により候補が提示されなかった語に対する用手的な手付け符号化
合計	**111,984**	**100.00**	

注：主指標対象（111,984 語）＝ 傷病名（112,133 語）− データ例外（14 語）− 非傷病名（135 語）。

表 2. 評価上の除外項目

区分	項目名	語数（n）	内容
主指標対象除外	データ例外	14	記号・数字・空文字のみで、傷病名として成立しない対象
主指標対象除外	非傷病名	135	修飾語のみ、部位のみ、診療科名、後置語、なし表現などで構成され、傷病名として成立しない語
参考指標対象 追加除外	ICD 未設定または候補なし	116	標準病名マスターに収載されるが、当該病名単独では ICD-10 コードが一意に定まらない傷病、または PCMA の照合で有効な候補が選定されなかった対象
参考指標対象 追加除外	確定符号不明系	15	患者調査側の確定符号が確認できない対象（確定符号不明 9 ＋ 索引一致・確定符号不明 6）
参考指標対象 追加除外	符号付け方針の差異	1,184	PCMA が選択した索引用語と病名が完全一致するが、標準病名マスターの ICD-10 コードと患者調査側の確定符号が一致しない対象（索引一致・確定符号相違 746 ＋ 左右不影響・確定符号相違 438）

注：主指標対象は、傷病名 112,133 語から、データ例外 14 語および非傷病名 135 語を除外した 111,984 語である。参考指標対象は、主指標対象 111,984 語から、ICD 未設定または候補なし、確定符号不明系、符号付け方針の差異に該当する語を除外した 110,670 語である。なお、ICD 未設定または候補なしと確定符号不明系の間に 1 語の重複があるため、各除外項目の単純合計を減じた値（110,669）とは 1 語の差がある。ICD 未設定または候補なし 116 語の符号化工程別内訳は、辞書符号付け 28 語、形態素解析を用いた手付け符号付け 47 語、手付け符号化（担当者による符号付け）41 語であった。

表 3. PCMA の処理段階と役割

段階	処理内容	出力
① 正規化	全角半角統一、小書き文字統一、 句読点・空白除去、修飾語識別	正規化された病名
② 照合	標準病名マスターの全索引用語との 文字類似度を PCMI として算出	候補リスト （PCMI が高い順、上位第 5 位候補）
③ 概念点検	悪性／良性、両側／単側、急性／慢性、 章・サブ分類等の医学概念を確認し、 不整合候補を抑制	PCMA 第 1 位候補

表 4. PCMI の主要な構成要素と算出例

記号	観点	算出例（「胃癌」vs「胃悪性腫瘍」）
s_base	共通文字率	「胃」が共通 → 1/2 ＝ 0.50
p_diff	非共通ペナルティ	候補側に「悪」「性」「腫」「瘍」 → ペナルティ加算
s_order	語順整合度	「胃」が両者の先頭 → 高値
s_len	索引用語長係数	候補語長＞入力長 → 適切と判定
s_gram	2-gram 一致率	共通の 2 文字連続なし → 低値
s_morph	形態素類似度	「胃」「癌」／「胃」「悪性」「腫瘍」 → 「胃」が一致
s_mod_coh	修飾語タイプ整合度	両者の修飾語（部位・側性・重症度等）タイプの整合度を評価
s_core	修飾語除去後文字集合共通率	傷病名から修飾語を除去した部分と候補索引用語との文字集合共通率を評価

表 5. 工程別および全体正答率

集計区分	語数（n）	正答数	正答率（%）	95%CI
辞書符号付け	36,883	33,945	92.03	91.75–92.31
形態素解析が候補を提示した語に対する手付け	59,408	49,594	83.48	83.18–83.78
手付け符号化	15,693	11,749	74.87	74.18–75.54
主指標対象 全体	**111,984**	**95,288**	**85.09**	**84.88–85.30**
参考指標対象	110,670	95,288	86.10	85.90–86.30

注：参考指標対象は感度分析の参考値（除外項目の詳細は表 2）。

表 6. PCMA と形態素解析の出力 ICD-10 コード（ICD-10 の 3 桁分類単位）と確定符号一致のクロス集計（n＝59,375）

出力 ICD-10 コードの関係	語数（n）	割合（%）	両方とも確定符号と一致	PCMA のみ確定符号と一致	形態素解析のみ確定符号と一致	両方とも確定符号と不一致
PCMA と形態素解析が同一の 3 桁分類を提示	41,754	70.32	38,844	—	—	2,910
PCMA と形態素解析が異なる 3 桁分類を提示	17,621	29.68	—	10,731	3,920	2,970
合計	**59,375**	**100.00**	**38,844**	**10,731**	**3,920**	**5,880**

注：PCMA 正答率 49,575／59,375 ＝ 83.49%、形態素解析正答率 42,764／59,375 ＝ 72.02%、差 ＋11.47 ポイント（McNemar 検定 χ^2＝3,165.39、p<0.0001）。

表 7. 辞書照合で処理されなかった語 75,101 語における 3 層構造

層	条件	語数（n）	辞書照合で処理されなかった語 75,101 語に占める割合（%）	層の役割
第 1 層	形態素解析と PCMA が一致した符号を提示	41,754	55.60	異なるアプローチが同一の符号に到達
第 2 層	形態素解析と PCMA が一致しないがいずれも符号を提示（PCMI が文字列上の近似度）	17,621	23.46	両者を手がかりに手付け
第 3 層	形態素解析が候補を提示しなかった語に対する手付け符号化対象（PCMA 候補提示あり 15,652／提示なし 41）	15,693	20.89	手がかりがなかった語に対する PCMA 候補提示の可否を整理
3 層化対象計	**—**	**75,068**	**99.95**	**—**

注：辞書照合で処理されなかった語は 75,101 語（形態素解析が候補を提示した語に対する手付けの対象 59,408＋同候補が提示されなかった語に対する手付けの対象 15,693）。辞書符号付けは傷病名辞書により確定符号が付与された範囲であり、当該語に含まれない。「異なるアプローチ」は形態素解析（言語処理）と PCMA（標準病名マスター照合）の対比を指す。第 1 層と第 2 層の合計 59,375 は形態素解析側で第 1 位候補が出力された語数であり、75,068 と 75,101 の差 33 語（第 1 位候補が出力されなかった語）は 3 層化の対象から除外した。各層の PCMA 第 1 位候補正答率は、第 1 層 93.03%（95%CI：92.78–93.27）、第 2 層 60.90%（同：60.18–61.62）、第 3 層 74.87%（同：74.18–75.54）であった。

表 8. 第 1 層 41,754 語における不正答 2,910 語の内訳

区分	語数（n）	割合（%）
アルゴリズム改善余地（残余として算出）	2,171	74.60
符号付け方針の差異に起因	696	23.92
ICD 未設定	28	0.96
確定符号不明系	15	0.52
合計	**2,910**	**100.00**

注：第 1 層は形態素解析と PCMA が一致した符号を提示した語（41,754 語、表 7）。「アルゴリズム改善余地」は不正答全体（2,910 語）から、符号付け方針の差異に起因する語、ICD 未設定、確定符号不明系を除いた残余として算出した。残余として整理された語については、今後のアルゴリズム改善余地として検討する必要がある。

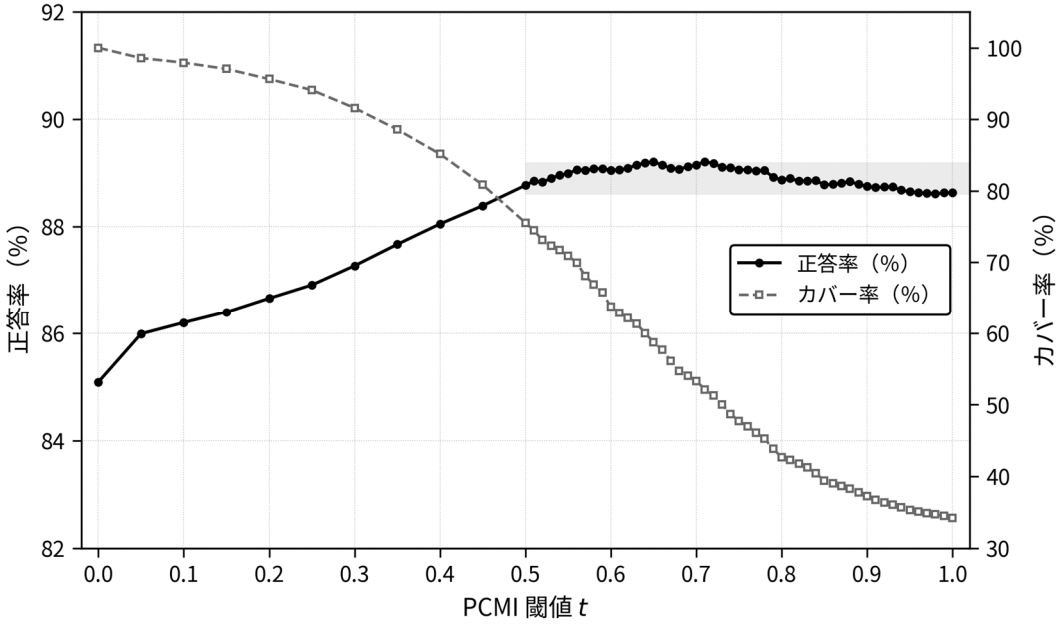


図 1. PCMI 閾値別の正答率とカバー率

注：横軸は PCMI 閾値 t、左縦軸は正答率（%）、右縦軸はカバー率（PCMI≥t の語数比率、%）。閾値を高めても、正答率は 88.6〜89.2%の狭い帯域に留まる。PCMI は候補順位付けを目的として設計された指標である。

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

自動符号付けアルゴリズムと傷病符号付けチェックの要件
AIを用いた類似度判定についての試行

研究分担者 阿南 誠 川崎医療福祉大学医療福祉マネジメント学部 医療秘書学科 特任教授

研究要旨

類似度判定について、類似度を計算する目的やどのようなデータと比較することによって適した計算方法は変わることになるが、具体的な方法として、標準病名マスターの中から、形態素解析によって単語に分割した後、最も単語数が多いものと類似度を比較することによって、コーディング精度は上げることが可能ではないかと考え試行を行った。対象は、慢性 C 型肝炎、慢性 B 型肝炎、アルコール性肝炎の 3 つの病態から AI ツールの Gemini を用いて判定が可能か否か試行を行い、正しい ICD コードが導かれることを確認した。

A 研究目的

本研究は、自動符号付けアルゴリズムと傷病符号付けチェックの要件を目的として、汎用的な AI ツールを用いて類似度判定が可能であるか試行することが目的である。

B 研究方法

1) 分析対象

慢性 C 型肝炎、慢性 B 型肝炎、アルコール性肝炎の 3 つの病態から検討することとした。

※前提として、慢性肝疾患という共通点、臨床的に共通する特徴があること。疾患の原因に基づく特有の特徴の両方があるという条件で選定した。

2)分析ツール(用いたツール)

研究に用いたツールは、AI ツールのうち、Gemini、形態素解析ツールとしては、MeCab である。

2) 分析方法

具体的な方法として、標準病名マスターの中から、形態素解析によって単語に分割した後、最も単語数が多いもの(ただし、実際には、不必要な単語は排除する必要があると思われるが、今回は配慮しない)、と類似度を比較する。この方法でコーディング精度は上げることが可能か検証する。

C 結果

1) 形態素解析の結果

※Gemini の他、MeCab でも同様の結果を示し、齟齬は発生しなかった。

「慢性 C 型肝炎」 →「慢性」「C 型」「肝炎」

「慢性 B 型肝炎」 →「慢性」「B 型」「肝炎」

「アルコール性肝炎」→「アルコール性」「肝炎」

※C 型は「C」と「型」が分離されることあり。

すなわち、この段階では、形態素解析として、単語の構成として類似しているかどうかは判定出来ても、ICD 分類の条件(定義や意味)は課されていないので、単純に似ているかどうかの判定しか出来ない。そこで、最後は人間か AI 等のチェックが必要となる。

つまり、Mecab を用いた形態素解析から計算出来るのは、見た目の表現「用語・記述の類似度」ということになり、ICD コーディング等の「意味の類似度」はこの段階では判定困難である。例えば、Gemini を用いた場合の解析ステップをみると、形態素解析以降のプロセスを加わることがわかる(表1参照)。

表1. Gemini における解析ステップ

ステップ	詳細
データ収集	各病態に関するテキストデータ（医学論文、教科書の記述、診断基準など）を集める。
形態素解析	集めたテキストを形態素（単語）に分割し、**単語の出現頻度**を数える。
特徴量抽出	**TF-IDF**（Term Frequency-Inverse Document Frequency: 特定の単語の重要度を示す指標）などの手法を用いて、各病態を特徴づける**単語のベクトル**を作成する。
類似度計算	作成されたベクトル間の**コサイン類似度**などの指標を計算し、類似度を数値で示す。

2）Gemini における類似度計算結果について

（1）方法、プロセス

慢性 C 型肝炎、慢性 B 型肝炎、アルコール性肝炎の類似度を、形態素解析で抽出された仮想的な特徴語の出現頻度に基づき、コサイン類似度で計算。

（2）類似度計算の結果

仮想的な医学的記述データに基づく計算結果は以下（3）の通り。

（3）比較ペアコサイン類似度

①慢性 C 型肝炎と慢性 B 型肝炎 {0.853}

②慢性 C 型肝炎とアルコール性肝炎{0.236}

③慢性 B 型肝炎とアルコール性肝炎 {0.211}

すなわち、①の類似度が顕著に高いということがわかる。

（4）計算プロセスについて

各病態に関する医学テキストを形態素解析によって処理し、特徴語とその出現頻度（重み）を抽出。これにより、各病態を多次元空間のベクトルとして表現（詳細な計算方法は割愛）。

表2. 計算プロセス（重み付け等）

形態素 (特徴語),	C型 (VC),	B型 (VB),	アルコール性 (VA)
ウイルス,	5,	7,	0
アルコール,	0,	0,	8
炎症,	3,	3,	4
線維化,	2,	2,	3
肝硬変,	2,	2,	2
治療薬,	4,	3,	0
飲酒,	0,	0,	6
DNA,	0,	3,	0
RNA,	3,	0,	0

3）条件設定を修正し（キーワードの追加、削除）、試みた結果

（1）「ウイルス性」というキーワードを追加した結果

慢性 C 型ウイルス性肝炎と慢性 B 型ウイルス性肝炎の類似度は、以前の計算に基づき、0.853→数値は変わらなかった。

即ち、C 型、B 型でウイルス性ということは認識され影響を与えない。

（2）「慢性」というキーワードを削除した結果

慢性 C 型肝炎と C 型肝炎の類似度は、極めて高い（ほぼ 1.0）という結果であった。このケースもキーワードの慢性（という意味）は認識され影響は与えない。

標準病名マスターにおける ICD コードは、どちらも B18.2 となっており、導き出された結果は正しくコーディングされるといえる。

D 考察

1） 形態素解析による類似度判定

結果の 1）に示すとおり、形態素解析によって、類似度はある一定の比較は可能である。

2） 形態素解析のみでは困難なケース

しかし、病勢、病態、原因等は、ICD コーディングに大きく影響するため、単語の意味を理解しない単純な形態素解析による単語抽出だけでは精度の改善は難しいケースがある。

3） AI ツールの可能性

したがって、標準化が十分ではない、死亡診断書のような記載レベル、入力レベル（必ずしも ICD コーディングのエキスパートの業務ではない）の違いが大きい場合は、その文章の内容（医学的知識や ICD の知識）を理解しなければ類似度を測ることは難しい。しかしながら、ここに AI ツールを導入することによって、精度改善は可能であることが示唆された。

例えば、結果の2）に示すとおり、Gemini による類似度計算結果から推察する限り、十分な満足のいく結果を示した。結果に示すように、慢性 C 型肝炎と慢性 B 型肝炎は非常に高い類似度を示しており。これは、両者が同じウイルス性肝炎であり、「炎症」「線維化」「肝硬変」といった病態の進行に関する共通語を多く含む（と AI が学習している）ためと推察される。一方、ウイルス性の 2 つとアルコール性肝炎の類似度は低い結果となっているが、これは、アルコール性肝炎は「アルコー

ル」「飲酒」といった原因に特有な語の頻度が高く、病因に関する記述が大きく異なり、併せてその語意（疾病への学習）を理解され区別が可能なため、類似度が低く計算されたと推察される。
すなわち、結果の3）で示したとおり、形態素解析を用いるテキストベースの類似度計算において、この二つの用語（ウイルス性、慢性というキーワードの有無）が医学的概念として実質的に同一であると認識されていると推察される。
4）　実装する場合の課題（AI ツールにおける課題）
（1）基本的に類似度を判定するためには、まず、比較する対象（正しい ICD が振られており、必要十分な条件を満たしている「標準的」傷病名であることが必要）を準備、決定する必要がある。
（2）上記の（1）で準備された傷病名とターゲットとした傷病名の類似度を検証することによって、ある程度、同一か否かは検証出来るが、いくつか留意すべき点が考えられる。それは、以下のとおりである。
①利用する AI ツールの違いが出現する可能性がある（Gemini と結果確認のため一部 ChatGPT で解析したが今回の対象について、差異はなかったが、偶然の可能性はある）。
②形態素分析した結果、単語が多くなった場合、類似度が下がる可能性を否定出来ない。すなわち、複雑な傷病名であった場合、専門性等の要因から特徴的な慣習があった場合等、学習が困難になる可能性がある。
③上記の②の影響を考えると、同一の ICD に対して複数の標準的な傷病名（例えば、短い表現と長い表現を準備しておく）を準備する必要があると思われる。改善のためには、汎用的な AI ではなく、医療、医学に適した教師学習を実施された AI が必要になろう。
④データが大量になった場合、解析に時間がかかるであろうことが推察される。

E 結論

今回の慢性肝炎という狭い範囲の試行例ではあったが、AI ツール Gemini で試行した結果を見る限り、現行の AI ツールでも、十分に傷病名における形態素解析結果の単語の医学的意味は理解可能であると思われ、複雑、例外的なケースでなければ、ICD コーディングの一括変換や内容確認には十分に実用レベルにあると考えられた。引き続き、サンプル数を増やした検証も行いたい。

F 健康危険情報

なし

G 研究発表

1.論文発表
なし
2.学会発表
2026 年度、日本医療マネジメント学会学術総会にて発表予定。

H知的財産権の出願・登録状況（予定を含む。）

1.特許取得
なし
2.実用新案登録
なし
3.その他
なし

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

ICD大分類別にみた手付けによるICD符号付与の出現要因に関する研究

研究分担者 桜澤邦男 国立大学法人東北大学 東北大学病院メディカルITセンター 准教授

研究要旨

ICD 大分類別の視点から、患者調査における手付け ICD 符号付与(アンマッチ病名)を高める要因の実態を把握し、手付けによらない辞書機能の改良に資する基礎情報を得ることを目的とした。アンマッチ病名を高める影響要因としてUコード、Zコード、皮膚、呼吸器、循環器が導出され、アンマッチ件数は新生物および損傷で相対的に多かった。本研究成果は今後の ICD 符号付け精度向上に向けた基礎資料として活用できる

A 研究目的

本研究は、ICD 大分類別の視点から、患者調査における手付け ICD 符号付与(アンマッチ病名)を高める要因を把握すること、および手付けによらない辞書機能の改良に資する基礎情報を得ることを目的とする。

B 研究方法

1)分析対象

本研究は、令和 5 年度患者調査で使用した辞書ファイル等を対象とした。掲載されている傷病名の数は約 11 万 2 千語である。これを統計処理システムのプログラム(厚生労働省担当職員独自作成)を用いて符号付けが行われているところ、同システムによる符号付けが完了した群(マッチ群)および手付けによる符号付けとなった群(アンマッチ群)に群別し分析を行った。

2)分析手法

マッチ群およびアンマッチ群の傷病名リストを統合し、群別に傷病名テキストと ICD を識別できるよう分析用データセットを構築した。続いて分析用データセット内の ICD を ICD 大分類別に集計し、群別件数とアンマッチ割合を記述した。ICD 大分類の内容を表 1 に示す。さらに Chi-square test を用いて、対象全体に占める ICD 大分類別のアンマッチ病名割合に統計的な有意差がみられるかを確認した。

次に解析として単変量および多変量ロジスティック回帰分析を行い、ICD 大分類別の視点からアンマッチとなる影響要因を推定した。解析モデルとして、多変量ロジスティック回帰分析は、同単変量解析で有意な影響要因として挙げられた ICD 大分類項目を投入した。統計的仮説検定は両側検定とし、p値が 0.05 未満の際に有意差があると判定した。解析ソフトは IBM SPSS (IBM Corp., Armonk, NY, USA)Ver.30 を用いた。

C 研究結果

対象全体の件数として 112,133 件が観察された。群別集計の結果、マッチ群は 36,915 件(32.9%)、アンマッチ群は 75,218 件(67.1%)の分布であった。

対象全体に占める ICD 大分類別の出現頻度として、新生物(n=19,858:17.7%)、損傷(n=18,752:16.7%)、筋骨格(n=10,760:9.6%)、循環器(n=9,747:8.7%)等が多く観察され、アンマッチの件数は、新生物、損傷、筋骨格で 1 万件を超えていた(表 2)。

ICD 大分類別のアンマッチ病名割合について、複数の項目で有意差がみられた。対象全体のアンマッチ割合(67.1%)よりも差が有意に高いものとして、U コード、Z コード、呼吸器、皮膚、循環器、消化器が、有意に低いものとして周産期、先天奇形、耳、眼、神経、損傷、新生物が、それぞれ取り上げられた(表 3)。

解析の結果では、まず単変量ロジスティック回帰分

析において前記と同様の結果がみられることを確認した。単変量解析で有意な影響要因を投入して展開した多変量ロジスティック解析分析では、U コードのオッズ比が最も高く 5.963 であり、以下 Z コード 2.594、呼吸器 1.231、皮膚 1.223、循環器 1.153 と続いた。周産期では 0.601 と最も低いオッズ比を示した(表 4)。

D 考察

本研究は ICD 大分類別の視点から、アンマッチ病名割合を高める影響要因の実態把握をねらいとした。アンマッチ病名を高める影響要因として U コード、Z コード、呼吸器、皮膚、循環器が導出され、アンマッチ件数は新生物、損傷、筋骨格そして循環器で相対的に多かった。

今後に向けた分析の拡張性として、ICD4 桁レベル等(傷病名レベル)での同様の分析でさらに深掘すること、および入院・外来別、病院・診療所別、地域別など様々な層別解析により実態をさらに詳しく明らかにすることが必要と考えられた。

E 結論

ICD 大分類別の視点からアンマッチ病名の要因を解析した結果、ICD 大分類で大きく異なることを導出し、実態を把握することができた。これらの研究成果は、今後の符号付け精度向上への辞書機能改良に資する基礎情報として活用できる。

F 健康危険情報

なし

G 研究発表

1.論文発表

なし

2.学会発表

なし

H知的財産権の出願・登録状況(予定を含む。)

1.特許取得

なし

2.実用新案登録

なし

3.その他

なし

表1、ICD大分類（章および名称）、本研究における略称一覧

章および名称	本研究における略称
Ⅰ　感染症および寄生虫症	感染症
Ⅱ　新生物	新生物
Ⅲ　血液および造血器の疾患ならびに免疫機構の障害	血液
Ⅳ　内分泌,栄養および代謝疾患	内分泌
Ⅴ　精神および行動の障害	精神
Ⅵ　神経系の疾患	神経
Ⅶ　眼および付属器の疾患	眼
Ⅷ　耳および乳様突起の疾患	耳
Ⅸ　循環器系の疾患	循環器
Ⅹ　呼吸器系の疾患	呼吸器
ⅩⅠ 消化器系の疾患	消化器
ⅩⅡ 皮膚および皮下組織の疾患	皮膚
ⅩⅢ 筋骨格系および結合組織の疾患	筋骨格
ⅩⅣ 腎尿路性器系の疾患	腎尿路
ⅩⅤ 妊娠,分娩および産じょく<褥>	妊娠
ⅩⅥ 周産期に発生した病態	周産期
ⅩⅦ 先天奇形,変形および染色体異常	先天奇形
ⅩⅧ 症状,徴候および異常臨床所見・異常検査所見で他に分類されないもの	症状
ⅩⅨ 損傷,中毒およびその他の外因の影響	損傷
ⅩⅩⅠ　健康状態に影響をおよぼす要因および保健サービスの利用	Zコード
ⅩⅩⅡ　特殊目的用コード	Uコード

表2、ICD大分類別、件数および構成割合

章	略称	n	%
Ⅰ	感染症	3,249	2.9%
Ⅱ	新生物	19,858	17.7%
Ⅲ	血液	742	0.7%
Ⅳ	内分泌	3,316	3.0%
Ⅴ	精神	2,718	2.4%
Ⅵ	神経	3,837	3.4%
Ⅶ	眼	4,766	4.3%
Ⅷ	耳	1,148	1.0%
Ⅸ	循環器	9,747	8.7%
Ⅹ	呼吸器	4,363	3.9%
ⅩⅠ	消化器	8,177	7.3%
ⅩⅡ	皮膚	7,524	6.7%
ⅩⅢ	筋骨格	10,760	9.6%
ⅩⅣ	腎尿路	3,405	3.0%
ⅩⅤ	妊娠	3,241	2.9%
ⅩⅥ	周産期	518	0.5%
ⅩⅦ	先天奇形	2,204	2.0%
ⅩⅧ	症状	2,740	2.4%
ⅩⅨ	損傷	18,752	16.7%
ⅩⅩⅠ	Ｚコード	280	0.2%
ⅩⅩⅡ	Ｕコード	773	0.7%
	該当なし	15	0.0%
	合計	112,133	100.0%

表3、ICD大分類別、群別件数およびアンマッチ割合

章	略称	マッチ群	アンマッチ群	合計	アンマッチ割合	p value
Ⅰ	感染症	1,060	2,189	3,249	67.4%	0.716
Ⅱ	新生物	6,753	13,105	19,858	66.0%	<0.001
Ⅲ	血液	264	478	742	64.4%	0.122
Ⅳ	内分泌	1,045	2,271	3,316	68.5%	0.080
Ⅴ	精神	920	1,798	2,718	66.2%	0.297
Ⅵ	神経	1,377	2,460	3,837	64.1%	<0.001
Ⅶ	眼	1,722	3,044	4,766	63.9%	<0.001
Ⅷ	耳	429	719	1,148	62.6%	0.001
Ⅸ	循環器	2,880	6,867	9,747	70.5%	<0.001
Ⅹ	呼吸器	1,231	3,132	4,363	71.8%	<0.001
ⅩⅠ	消化器	2,597	5,580	8,177	68.2%	0.020
ⅩⅡ	皮膚	2,133	5,391	7,524	71.7%	<0.001
ⅩⅢ	筋骨格	3,512	7,248	10,760	67.4%	0.514
ⅩⅣ	腎尿路	1,120	2,285	3,405	67.1%	0.972
ⅩⅤ	妊娠	1,017	2,224	3,241	68.6%	0.058
ⅩⅥ	周産期	231	287	518	55.4%	<0.001
ⅩⅦ	先天奇形	899	1,305	2,204	59.2%	<0.001
ⅩⅧ	症状	888	1,852	2,740	67.6%	0.564
ⅩⅨ	損傷	6,720	12,032	18,752	64.2%	<0.001
ⅩⅪ	Zコード	44	236	280	84.3%	<0.001
ⅩⅫ	Uコード	58	715	773	92.5%	<0.001
	該当なし	15		15	0.0%	
		36,915	75,218	112,133	67.1%	

Chi-square test

表4、ICD大分類別、アンマッチとなる影響要因（単変量・多変量ロジスティック回帰分析）

章	略称	単変量ロジスティック回帰分析			多変量ロジスティック回帰分析		
		OR	95% CI	p value	OR	95% CI	p value
Ⅰ	感染症	1.014	0.941 - 1.092	0.716		-	
Ⅱ	新生物	0.942	0.912 - 0.973	<0.001	0.939	0.904 - 0.975	<0.001
Ⅲ	血液	0.888	0.764 - 1.032	0.122		-	
Ⅳ	内分泌	1.069	0.992 - 1.151	0.080		-	
Ⅴ	精神	0.958	0.884 - 1.038	0.297		-	
Ⅵ	神経	0.873	0.816 - 0.933	<0.001	0.864	0.806 - 0.927	<0.001
Ⅶ	眼	0.862	0.811 - 0.916	<0.001	0.855	0.802 - 0.911	<0.001
Ⅷ	耳	0.821	0.728 - 0.926	0.001	0.811	0.718 - 0.916	<0.001
Ⅸ	循環器	1.187	1.135 - 1.242	<0.001	1.153	1.0971 - 1.212	<0.001
Ⅹ	呼吸器	1.259	1.178 - 1.347	<0.001	1.231	1.147 - 1.320	<0.001
ⅩⅠ	消化器	1.059	1.009 - 1.111	0.020	1.039	0.986 - 1.095	0.149
ⅩⅡ	皮膚	1.259	1.195 - 1.326	<0.001	1.223	1.156 - 1.292	<0.001
ⅩⅢ	筋骨格	1.014	0.972 - 1.058	0.514		-	
ⅩⅣ	腎尿路	1.001	0.931 - 1.077	0.972		-	
ⅩⅤ	妊娠	1.075	0.997 - 1.16	0.058		-	
ⅩⅥ	周産期	0.608	0.511 - 0.724	<0.001	0.601	0.505 - 0.716	<0.001
ⅩⅦ	先天奇形	0.707	0.649 - 0.771	<0.001	0.702	0.643 - 0.767	<0.001
ⅩⅧ	症状	1.024	0.944 - 1.111	0.564		-	
ⅩⅨ	損傷	0.856	0.828 - 0.884	<0.001	0.866	0.833 - 0.900	<0.001
ⅩⅩⅠ	Zコード	2.637	1.911 - 3.64	<0.001	2.594	1.879 - 3.583	<0.001
ⅩⅩⅡ	Uコード	6.099	4.665 - 7.972	<0.001	5.963	4.558 - 7.801	<0.001

従属変数：アンマッチ群1、マッチ群0

多変量解析は、単変量解析で有意な影響要因として推定された項目を投入した

OR：Odds Ratio、CI：Confidence Interval

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

アンマッチ病名における手付符号付け作業の効率化検討に関する研究

研究分担者 來島裕太（山口県立総合医療センター 医療企画室 室長補佐）

研究要旨

本研究は、統計法に基づく基幹統計である患者調査において、辞書ファイル適用後も符号付けができなかったアンマッチ病名の傾向と課題を明らかにし、今後の辞書ファイル改善に資することを目的とした。令和 5 年の患者調査では、全体の 14%にあたる約 1 万 6 千語が目視による手付け符号付けを要しており、その要因分析が課題であった。

今回、約 1 万 6 千語について、辞書ファイル適用後でも符号付ができなかったアンマッチ病名 75,218 件を対象に、ICD1 桁および 3 桁別の出現状況を集計・分析し、出現率の高い分類に着目した。また、新型コロナウイルス関連病名について表記揺らぎの検証を行った。

集計した結果、ICD1 桁では悪性新生物（C コード）、外傷（S コード）、筋骨格系疾患（M コード）、循環器系疾患が全体の約 4 割を占めた。ICD3 桁では脳梗塞、肺癌、乳癌などが多く、癌を含む病名の大半は C コードに該当していた。また、左右や両側といった側性表現を含む病名が多く確認された。悪性新生物に関する表記や側性表現の整理・除去など、クリーニング作業を追加実施することにより自動符号付け率の向上が期待できることが示唆された。

A 研究目的

患者調査は、統計法（第 2 条第 4 項）に基づく基幹統計であり、病院及び診療所を利用する患者の傷病状況を明らかにし、医療行政の基礎資料を得るための、日本における唯一の網羅的な疾病統計である。厚生労働省内における患者調査の傷病符号付けプロセスは三段階で構成されている。令和 5 年患者調査では、使用した辞書ファイル等約 11 万 2 千語のうち、厚生労働省の共通サービス及び職員独自作成プログラムによる「辞書ファイルを用いた自動符号付け」により 33%（約 3 万 7 千語）、患者調査オートコーディングシステムによる「形態素解析を活用した手付け符号付け」により 53%（約 6 万語）が符号付けされたが、14%（約 1 万 6 千語）は目視による手付け符号付けが必要であった。

このうち、辞書ファイル適用後に、符号付ができなかった病名（以下：アンマッチ病名）について、どのような傾向や課題があるかを分析し、今後の辞書ファイルのあり方について検討する材料とするため、検証した。

B 研究方法

令和 5 年患者調査で使用した辞書ファイル等のうち、アンマッチ病名 75,218 件を対象として分析を行った。

まず、ICD コードの第 1 桁（アルファベット）および第 3桁分類別に出現状況を集計した。その上で、出現頻度の高い分類に焦点を当て、傾向の分析を実施した。

また、令和5年患者調査の実施時期は、新型コロナウイルス感染症の国内初発例の報告から約 3 年が経過した時点で実施されている。この点を踏まえ、新型コロナウイルス感染症が付与された病名について、アンマッチの要因および表記揺らぎの実態について検証を行った。

（倫理面への配慮）

本研究で使用したデータには個人情報は含まれておらず、倫理的配慮に基づき適切に管理・利用した。

C 研究結果

アンマッチ病名の ICD1 桁目のアルファベット別の出現割合は、Cコード(悪性新生物)が 9,395 件(12.5%)と最も多かった。続いて S コード(損傷 中毒及びその他の外因の影響)のうち外傷に該当するコードが 8,203 件(10.9%)、M コード:筋骨格系及び結合組織の疾患が 7,248 件(9.6%)、循環器系の疾患が 6,867 件(9.1%)であり、これら上位4つは 42.16%を占めた。

アンマッチ病名の ICD3桁目別の出現割合の検証では、I63:脳梗塞が 1,442 件(1.9%)と最も多かった。続いて、C34:肺癌が 1,198 件(1.6%)、C50:乳癌が 865 件(1.1%)L30;湿疹が 760 件(1.0%)、D48:骨腫瘍が 750 件(1.0%)であった。

悪性腫瘍が多く含まれることから、アンマッチ病名に「癌」が含まれるものは、6,339 件であり、そのうち 5,956 件(93.9%)は C コード:悪性新生物であった。C コード(悪性新生物)にならない病名については、上皮内癌・非浸潤癌が 102 件(1.6%)、複数病名による別の疾患へのコーディングが、177 件(2.8%)が多く占めた。この複数病名については、急性肺炎、脳梗塞、狭心症などの疾患に加えて、癌の術後の病名付与していた。残りの 104 件(1.7%)については、「がん検診」など本来は病名ではない用語の付与や、「境界悪性卵巣腫瘍」などの境界病変、さらに「解離性大動脈瘤」や「右眼内レンズ挿入後」等における誤記・表記誤りなど、複数の要因により別のコードへ分類されていた。

アンマッチ病名に含まれる側性表現について、語単位での出現回数を集計した。「右」が 12,555 件(16.7%)、「左」が 12,239 件(16.3%)、「両」が 5,894 件(7.8%)、「右側」が 569 件(0.8%)、「左側」が 662 件(0.9%)、「両側」が 1,925 件(2.6%)、「左右」が 128 件(0.2%)、「ミギ」が 111 件(0.1%)、「ヒダリ」が 102 件(0.1%)であった(いずれもアンマッチ病名全体に対する割合)。なお、これらの語句は同一病名内に複数含まれる場合があるため、重複を許容して集計している。

新型コロナウイルス感染症に関係する病名表記は、916 件であり、そのうち 715 件(78.1%)は、U コードにコーディングされており、それ以外は、感染後の病態が 54 件(7.6%)、関連する多系統炎症性症候群が 2 件(0.3%)であった(表4)。病名表記には多様な揺らぎが認められ、「コヴィッド 19」、「COVID」、「コロナウイルス感染症 2019」、「新型コロナウイルス感染症」など、略語、カタカナ表記、数字の組み合わせによって多様な表現が用いられていた(図2)。

新型コロナウイルス感染症の活動性の状態である U07.1 は、641 件、U07.2 は 18 件であった(表5)。

D 考察

C コード(悪性新生物)、S コード(損傷 中毒及びその他の外因の影響)のうち外傷に該当するコード、Mコード:筋骨格系及び結合組織の疾患、循環器系の疾患が大半に占めることから、この4つの分類について改善されることが有効と考えられた。

Cコード(悪性新生物)については、病名に「癌」と記載されている場合、その多くが悪性新生物として分類され、上皮内新生物を除くカテゴリーに該当する。このことから、マッチング精度向上の観点では、C コード(悪性新生物)に分類されない病名表記のクリーニングが重要であると考えられる。

実際に、「癌」と診断された症例のうち、非浸潤性あるいは上皮内にとどまる病変(例:非浸潤性膀胱癌や結腸上皮内癌など)を除外することで、悪性新生物以外への誤分類は概ね整理可能であることが示唆された。さらに、それ以外のアンマッチ病名としては、境界悪性を示す病名や、「癌検診」など本来は病名ではない用語が付与されている事例も確認された。これらについても、一定のルールに基づくデータクリーニングを実施することで、対応可能であると考えられる。

I コードおよび S コードに関しては、左右情報の取扱いがマッチング結果に影響している可能性が示唆された。左右を含む病名について分析したところ、アンマッチ病名において「右」「左」「両」の順に語句を含む割合が高い傾向が認められた。この結果から、ICD コードに直接影響を及ぼさない左右情報については、前処理として除去または正規化を行うことで、マッチング件数の向上が期待される。

一方、標準病名マスター(ver.516)における左右表

記に起因するコード変更事例は24件にとどまっていた（表 3）。このことは、左右情報がコーディング全体に及ぼす影響は限定的である可能性を示している。

ただし例外として、変形性膝関節症や鼠径ヘルニアのように、左右情報ではなく、片側性あるいは両側性の違いが分類に影響する疾患も存在するため、左右情報の一律な除去には慎重な対応が求められる。

新型コロナウイルス感染症については、その多くがUコードに分類されることを踏まえると、全アンマッチ病名に占める割合は約 1%と低値であり、全体への影響は限定的であると考えられる。一方で、当該疾患に関する表記の揺らぎは相対的に大きいことが明らかとなった。英語表記とカタカナ表記やその混在が存在する多様な表記は、アンマッチの一因となっている可能性が高い。

また、新型コロナウイルス感染症は、臨床的・疫学的にはインフルエンザなど既存の呼吸器感染症に近似しつつあるものの、入院患者において主病名または併存症として一定数存在し、データの解析精度に影響を及ぼす可能性がある。

以上より、新型コロナウイルス感染症に関しては、影響の大きさ自体は限定的であるものの、マッチング精度向上の観点から、表記の標準化や同義語の統合といったデータクリーニングの必要性が示唆された。

E 結論

アンマッチ病名の傾向を把握することで、手付付与の ICD 作業を改善することに期待できる。さらに ICD 分類の性質や影響度合いから、クリーニング作業やマッチング率向上を優先的にすることが必要である。

F 健康危険情報

なし

G 研究発表

1.論文発表

なし

2.学会発表

なし

H知的財産権の出願・登録状況（予定を含む。）

1.特許取得

なし

2.実用新案登録

なし

3.その他

なし

表 1　アンマッチ 75,218 件に対する ICD1 桁目（アルファベット）分類による集計結果

A	感染症及び寄生虫症	711	0.9%
B	感染症及び寄生虫症	1,478	2.0%
C	悪性新生物	9,395	12.5%
D	上皮内・良性新生物／血液及び造血器の疾患並びに免疫機構の障害	4,188	5.6%
E	内分泌，栄養及び代謝疾患	2,271	3.0%
F	精神及び行動の障害	1,798	2.4%
G	神経系の疾患	2,460	3.3%
H	眼及び付属器／耳及び乳様突起の疾患	3,763	5.0%
I	循環器系の疾患	6,867	9.1%
J	呼吸器系の疾患	3,132	4.2%
K	消化器系の疾患	5,580	7.4%
L	皮層及び皮下組織の疾患	5,391	7.2%
M	筋骨格系及び結合組織の疾患	7,248	9.6%
N	腎尿路生殖器系の疾患	2,285	3.0%
O	妊娠，分娩及び産褥	2,224	3.0%
P	周産期に発生した病態	287	0.4%
Q	先天奇形，変形及び染色体異常	1,305	1.7%
R	症状，徴候及び異常臨床所見・異常検査所見で他に分類されないもの	1,852	2.5%
S	損傷 中毒及びその他の外因の影響	8,203	10.9%
T	損傷 中毒及びその他の外因の影響	3,829	5.1%
U	特殊目的用コード	715	1.0%
Z	健康状態に影響を及ぼす要因及び保健サービスの利用	236	0.3%
総計		75,218	100.0%

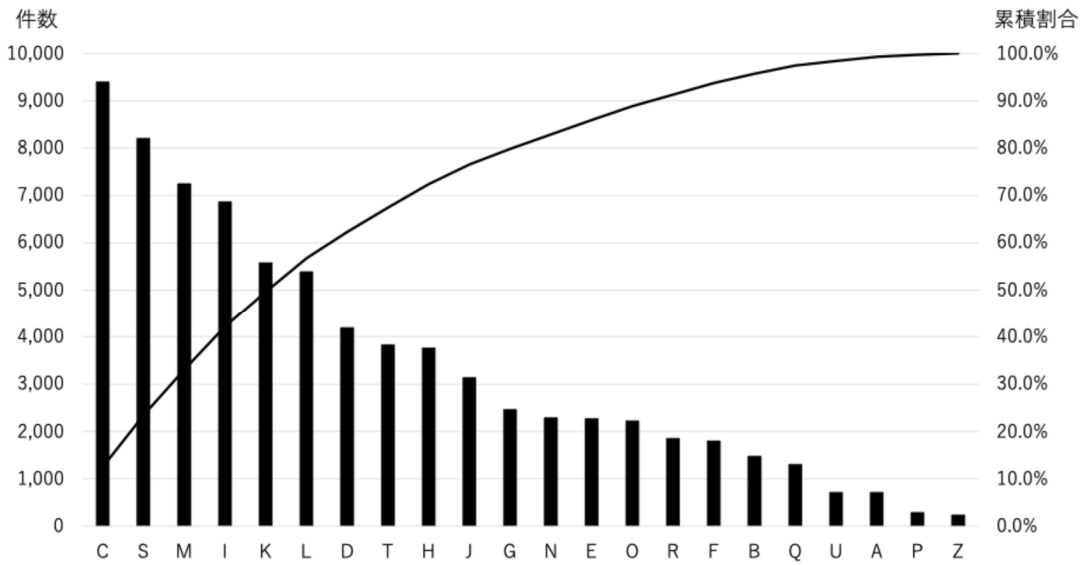


図 1：ノンマッチ75,218件に対するICD1桁目（アルファベット）分類による分布

表 2　アンマッチ 75,218 件に対する ICD3桁分類による集計結果（上位 25 位）

順位	ICD3 桁	病名	件数	割合	累積割合
1	I63	脳梗塞	1,442	1.9%	1.9%
2	C34	肺癌	1,198	1.6%	3.5%
3	C50	乳癌	865	1.1%	4.7%
4	L30	湿疹	760	1.0%	5.7%
5	D48	骨腫瘍	750	1.0%	6.7%
6	C16	胃癌	697	0.9%	7.6%
7	U07	新型コロナウイルス	659	0.9%	8.5%
8	S06	頭部外傷	649	0.9%	9.3%
9	C18	結腸癌	635	0.8%	10.2%
10	S32	腰椎損傷	607	0.8%	11.0%
11	S82	膝損傷	569	0.8%	11.7%
12	I61	脳出血	542	0.7%	12.5%
13	T02	多発骨折	525	0.7%	13.2%
14	S72	大腿骨折	517	0.7%	13.8%
15	L89	褥瘡	507	0.7%	14.5%
16	S62	指骨折	505	0.7%	15.2%
17	L03	蜂窩織炎	487	0.6%	15.8%
18	T81	術後合併症	481	0.6%	16.5%
19	I10	高血圧	457	0.6%	17.1%
20	I50	心不全	446	0.6%	17.7%
21	E11	２型糖尿病	442	0.6%	18.3%
22	M17	変形性膝関節症	431	0.6%	18.8%
23	H35	網膜症	428	0.6%	19.4%
24	E14	糖尿病	420	0.6%	20.0%
25	S92	足指骨折	410	0.5%	20.5%

表３：標準病名マスター(ver.516)における左右の記載がある病名一覧

病名表記（右）	ＩＣＤ（右）	病名表記（左）	ＩＣＤ（左）	比較
右心室粘液腫	D151	左心室粘液腫	D151	同一
右心房粘液腫	D151	左心房粘液腫	D151	同一
肝右葉腫瘍	D376	肝左葉腫瘍	D376	同一
右半球器質性感情障害	F078			片側のみ
右眼失明	H544	左眼失明	H544	同一
急性右室梗塞	I212			片側のみ
右室自由壁破裂	I219	左室自由壁破裂	I219	同一
不整脈原性右室心筋症	I428			片側のみ
		左脚前枝ブロック	I444	片側のみ
		左脚後枝ブロック	I445	片側のみ
		左脚ヘミブロック	I446	片側のみ
完全右脚ブロック	I451	完全左脚ブロック	I447	異なる
右脚ブロック	I451	左脚ブロック	I447	異なる
不完全右脚ブロック	I451	不完全左脚ブロック	I447	異なる
		左室駆出率が軽度低下した心不全	I501	片側のみ
		左室駆出率の保たれた心不全	I501	片側のみ
		左室駆出率の低下した心不全	I501	片側のみ
右室不全	I500	左室不全	I501	異なる
右心不全	I500	左心不全	I501	異なる
後天性左室右房交通症	I510	後天性左室右房交通症	I510	同一
		巨大左心房	I517	片側のみ
右室肥大	I517	左室肥大	I517	同一
右胃動脈瘤	I728	左胃動脈瘤	I728	同一
下顎右側偏位	K071	下顎左側偏位	K071	同一
上顎右側偏位	K071	上顎左側偏位	K071	同一
		潰瘍性大腸炎・左側大腸炎型	K515	片側のみ
両大血管右室起始症	Q201	両大血管左室起始症	Q202	異なる
右室型単心室症	Q204	左室型単心室症	Q204	同一
		左旋転位	Q205	片側のみ
右室流出路狭窄	Q208	左室流出路狭窄	Q208	同一
		大動脈左室トンネル	Q208	片側のみ
左室右房交通症	Q210	左室右房交通症	Q210	同一
右室低形成	Q226			片側のみ
右室低形成症候群	Q226			片側のみ
右心低形成症候群	Q226	左心低形成症候群	Q234	異なる
右胸心	Q240	左胸心	Q241	異なる
右心症	Q240	左心症	Q241	異なる
右室漏斗部狭窄	Q243			片側のみ
右冠動脈肺動脈起始症	Q245	左冠動脈肺動脈起始症	Q245	同一
右室二腔症	Q248			片側のみ
右側大動脈弓	Q254	先天性左室憩室	Q248	異なる
左肺動脈右肺動脈起始症	Q257			片側のみ
右肺動脈形成不全症	Q257	左肺動脈形成不全症	Q257	同一
右肺動脈欠損	Q257	左肺動脈欠損	Q257	同一
右肺動脈上行大動脈起始症	Q257	左肺動脈上行大動脈起始症	Q257	同一
		左上大静脈遺残	Q261	片側のみ
		左側胆のう	Q441	片側のみ
心電図右軸偏位	R943	心電図左軸偏位	R943	同一

表 4　アンマッチのうち、新型コロナウィルス感染症に関係する病名表記におけるコーディング状況

ｱﾙﾌｧﾍﾞｯﾄ	章	件数	割合
A	感染症及び寄生虫症	2	0.2%
B	感染症及び寄生虫症	0	0.0%
C	悪性新生物	4	0.4%
D	上皮内・良性新生物／血液及び造血器の疾患並びに免疫機構の障害	2	0.2%
E	内分泌，栄養及び代謝疾患	19	2.1%
F	精神及び行動の障害	4	0.4%
G	神経系の疾患	10	1.1%
H	眼及び付属器／耳及び乳様突起の疾患	0	0.0%
I	循環器系の疾患	12	1.3%
J	呼吸器系の疾患	53	5.8%
K	消化器系の疾患	10	1.1%
L	皮層及び皮下組織の疾患	1	0.1%
M	筋骨格系及び結合組織の疾患	4	0.4%
N	腎尿路生殖器系の疾患	10	1.1%
O	妊娠，分娩及び産褥	20	2.2%
P	周産期に発生した病態	12	1.3%
Q	先天奇形，変形及び染色体異常	0	0.0%
R	症状，徴候及び異常臨床所見・異常検査所見で他に分類されないもの	5	0.5%
S	損傷 中毒及びその他の外因の影響	11	1.2%
T	損傷 中毒及びその他の外因の影響	14	1.5%
U	特殊目的用コード	715	78.1%
Z	健康状態に影響を及ぼす要因及び保健サービスの利用	8	0.9%
総計		916	100.0%

- コヴイツト－１９
- コヴイツドナインテイーン
- コビツト
- コビツド１９
- ＣＯＶ－１９
- ＣＯＶ－１９感染症
- ＣＯＶＩＤ
- ＣＯＶＩＤ－１９
- コロナ感染
- コロナ感染者
- コロナ感染症
- コロナ陽性
- コロナウイルス
- コロナウイルス・２０１９
- コロナウイルス感染症
- コロナウイルス感染症２０１９
- 新型コロナウイルス
- 新型コロナウイルス感染症

など

図２：新型コロナウイルス感染症の表記例

表 5 アンマッチのうち、新型コロナウィルス感染症のコーディング状況

ICD	名　　称	件　数
U07.1	コロナウイルス感染症 2019，ウイルスが同定されたもの	641
U07.2	コロナウイルス感染症 2019，ウイルスが同定されていないもの	18
U08.9	コロナウイルス感染症 2019 の既往歴，詳細不明	0
U09.9	コロナウイルス感染症 2019 後の病態，詳細不明	54
U10.9	コロナウイルス感染症 2019 に関連する多系統炎症性症候群，詳細不明	2
U 11.9	コロナウイルス感染症 2019 に対する予防接種の必要性，詳細不明	0
U12.9	治療上の使用により有害作用を引き起こしたコロナウイルス感染症	0
	合計	715

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

クラスタリングによる不一致パターン分析に基づく既存符号付けアルゴリズムの改善に関する研究

研究分担者 森藤祐史 大阪府立病院機構大阪急性期・総合医療センター
医療情報部 診療情報管理室 総括主査

研究要旨

本研究は、患者調査における ICD 符号付けの精度向上を目的に、辞書判定コードと手付けコードの不一致パターンをクラスタリングにより分析した。分析対象とした不一致 23,498 語(約 39.5%)を K-means 法で 8 クラスタに分類した結果、診療領域ごとに体系的な不一致傾向が確認された。これらの知見に基づき、既存の表記揺れ補正プログラム(xx2.shf)における「Z 付きコードの基本コード置換ルール」の追加や、「目印コード(ZZZZZ 等)」の抽出条件精緻化、さらにはアンマッチ基礎表における参考情報の充実など、ルールベースによる既存プロセスの改善策を提示した。本提案は、高度な自動学習等を用いず、現行の作業アルゴリズムの論理的改善によって符号付け精度の向上と作業負担の削減を目指すものである。

A 研究目的

本研究は、令和 5 年度患者調査までの蓄積である「最新辞書」を用いてもなお、目視による手付け符号付け作業が多数発生している要因を分析することを目的とする。特に、AI 等の新規技術を導入するのではなく、既存の表記揺れ補正 PG や辞書管理、目印コードを用いた個別修正(ステップ 5)等の一連のアルゴリズムにおいて、どのプロセスを論理的に改善すべきかを考察し、実務に即した精度の最適化を図る。また、本研究では、既存の統計処理システムおよび ICD-10(2013 年版)準拠の枠組みの中で、いかに論理的な改善を図るかに焦点を当てている。

B 研究方法

1)分析対象

本研究は、最新辞書によるマッチングの処理を経て参考コードが付与された 59,436 語のうち、目視による手付け符号付け作業により ICD コードが変更された 23,498 語を対象とした。

・対象語数(辞書判定 ICD コードと手付け ICD コードのペア):59,436 件

・両コード不一致語数:23,498 件(約 39.5%)

クラスタリング前の対象語数の分布は図 1 のとおり。また、変更理由のコメント欄に符号付け作業時の記載がある 784 語から、上位グループは下表(表 1)のとおり。

表 1 不一致理由の上位グループ

確認済みの簡易コメント(235 語)
年齢条件(16 歳未満など)(58 語)
初筆(初回記載)に関する判断(24 語)
「?」のある確認コメント(20 語)
外傷性切断の術後か切断術後かの判断(18 語)
外傷関連(18 語)
年齢によるコード分岐(例:16 歳未満=L20.8、16 歳以上=L30.9)(15 語)
部位の明記不足で局所性と判断(13 語)
股関節の詳細不明疾患(9 語)

辞書コードは一般的な ICD コードを付与するが、手付けでは下表(表 2)のように、より詳細なコー

ドを選択されたケースが多い。

表２

傷病名	辞書 ICD	手付け ICD
サルモネラ症	A02.9	A02.0
ネズミチフス菌腸炎	A09.9	A02.0
腸管出血性大腸感染症	A09.0	A04.3
急性細菌性腸炎	A09.9	A04.4

そのほか、「NOS コード（詳細不明）への変更（部位や病態が不明確な場合に手付けで NOS を選択）」、「年齢や外傷の有無など、辞書では拾えない条件分岐が必要なケース」、「術後状態や後遺症など、文脈依存の判断が必要なケース」などのパターンがある。

これにより、ICD コードの不一致パターンは、類似性に基づくグルーピングが有効な可能性がある。

２）分析手法

辞書コードと手付けコードのペアを数値化し、K-means 法によるクラスタリングにより差異特徴として抽出した。患者調査の統計公表における最も詳細な単位は「傷病基本分類（ICD-10 の 3 桁項目および 4 桁細分類項目）」として設定されており、この分類体系に 5 桁目の情報は含まれない。しかし、符号付けの実務プロセスにおいては、部位詳細コードや条件判定のための目印コード等の 5 桁目の情報が実務上の解像度として介在している。本研究は自動符号付けアルゴリズムの最適化を主眼としており、4 桁以下に丸めた場合には、Cluster 0（Z コードの基本コードへの置換）や Cluster 1（部位詳細コードの付与）に代表される「実務上で発生している不一致パターン」が構造的に観察不能となる。そのため、本分析では実務上のボトルネックを精緻に特定する目的から、5 桁解像度での解析を採用した。

ICD コードは文字列のデータであるが、厳密なグループと階層の構造を持ち、コードの並び自体が医学的近接性を反映しているため、医学的類似度をユークリッド距離として近似的に抽出できる発見的手法として妥当であると判断した。各ペアを 2 次元の数値ベクトル（辞書コード、手付けコード）に変換したラベルエンコーディングにより数値化する。クラスタ数はエルボー法によって選定する。不一致の原因については、既存の「表記揺れ補正処理（xx2.shf）」や「患者独自ルール（ステップ 1～3）」との整合性の観点から整理した。分析にあたっては、ICD コード間の差異を客観的に評価するため、一時的に数値化した。これは実務上の既知の不一致傾向を論理的に整理する目的であり、自動学習モデルの構築を目的としていない。

C 研究結果

最新辞書によるマッチングを経て不一致となった語に対し、K-means 法を用いたクラスタリングを実施した結果、以下の通り診療領域ごとの体系的な特徴が抽出された。

１）クラスタリングの実施と分類状況

分析対象とした不一致語のうち、実運用における出現頻度を考慮するため、語彙単位のデータをレコード単位に再構成し、欠損値の除外や正規化処理を行なった。30,020 レコードが対象となった。シルエット係数（SC : Silhouette coefficient）はクラスタ数 6 で最大値を示したが、クラスタ数 8 においても SC =0.52 と高い値を維持していた。クラスタ数の決定にあたっては、クラスタ内分散（WCSS：Within-Cluster Sum of Squares）の減少傾向に加え、実務上の診療ドメイン（循環器、外傷等）の分離状況が最も整合的であったことから、クラスタ数 8 を採用した。（図２、表３…図表中の WCSS は 10^6単位で表示した）

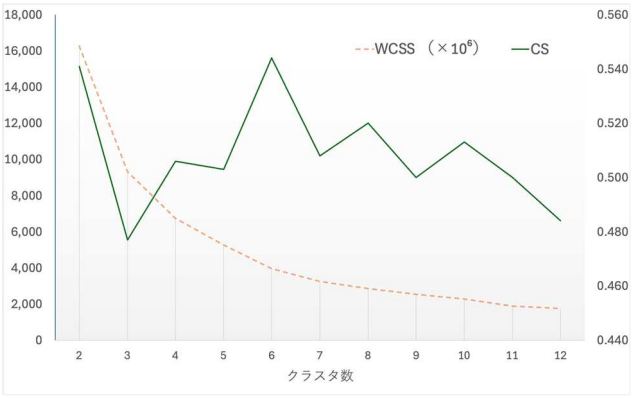


図２　エルボー法およびシルエット係数

表３　エルボー法およびシルエット係数

クラスタ数	WCSS	SC
2	16,265	0.541
3	9,313	0.477
4	6,742	0.506
5	5,259	0.503
6	3,960	0.544
7	3,251	0.508
8	2,856	0.52
9	2,530	0.5
10	2,269	0.513
11	1,885	0.5
12	1,758	0.484

各クラスタの名称および分類語数は下表（表４）。クラスタ名称は、当該クラスタにおいて最も頻出した不一致パターンに基づいて付与しており、必ずしも単一診療領域のコードのみで構成されるものではない。

表４　クラスタ名称および分類語数

Cluster 0：循環器疾患	5,363 語
Cluster 1：外傷	5,841 語
Cluster 2：感染症	3,864 語
Cluster 3：泌尿器・産科関連	3,321 語
Cluster 4：眼科・精神疾患	2,807 語
Cluster 5：皮膚・筋骨格疾患	5,322 語
Cluster 6：妊娠・管理関連	1,800 語
Cluster 7：その他	1,702 語

２）クラスタ別の代表的な不一致パターン

各クラスタにおいて、形態素分析による判定（R5_AC）と手付けコードの間で発生した代表的な不一致ペアとその差異の特徴は以下の通りである（表５より抜粋）

- Cluster 0（循環器疾患）：
  事例：I50.9Z（形態素分析）→ I50.9（手付け）
  特徴：循環器や呼吸器疾患における**「Z 付きコード」の基本コードへの置換**が目立つ。
- Cluster 1（外傷）：
  事例：S06.5B → S06.5、S32.0F → S32.0
  特徴：骨折部位の詳細な分類（部位コードの付与）による不一致が主である。
- Cluster 2（感染症）：
  事例：B02.9 → B02.2、D27.Z → D27
  特徴：原因菌やウイルス種の詳細差、および腫瘍コードの詳細化による差異が確認された。
- Cluster 3（泌尿器・産科関連）：
  事例：U99 → O34.2（妊娠管理関連）、M86.9B → M86.97（筋骨格）
  特徴：妊娠合併症や管理に関する文脈依存の判断、および筋骨格系の詳細部位差が見られる。
- Cluster 4（眼科・精神疾患）：
  事例：H40.9 → H40.1、F03 → G30.9
  特徴：白内障や緑内障、認知症関連における詳細分類の選択に差異が生じている。
- Cluster 5（皮膚・筋骨格疾患）：
  事例：L30.9 → L85.3、M48.0 → M48.0
  特徴：皮膚疾患の分類体系の差や、脊椎疾患の詳細部位の選択による不一致が多い。
- Cluster 6（妊娠・管理関連）：
  事例：R54 → H35.3、R58.Z → I61.0
  特徴：R コード（症状・徴候）から特定の疾患コード（眼疾患や循環器疾患等）への置換が多い。
- Cluster 7（その他）：

事例：J12.8F → U07.1（COVID-19 関連）
特徴：呼吸器疾患の詳細差や、特定の流行性疾患への修正が顕著である。

3）不一致要因の定性的分析

変更理由のコメントが付与されていた 784 語の分析からは、次の要因が上位に挙げられた。「年齢条件（16 歳未満等）によるコード分岐」、「初筆（最初に記載された傷病名）の優先判断」、「部位の明記不足による局所性判断の要否」、「？を含む不明瞭な記載に対する目視確認」。これらの結果は、単なる辞書の網羅性不足だけではなく、既存の「表記揺れ補正」や「目印コード」のロジック、および作業者の判断基準を精緻化する必要があることを示唆している。

D 考察

本研究のクラスタリング結果から、特定の診療領域において体系的な不一致傾向が確認された。形態素解析において、特定の診療領域（外傷の部位や感染症の菌名など）が未知語（辞書未登録語）として適切に処理されず、詳細不明コード（NOS）に流れてしまうという既存の形態素解析処理の限界が、クラスタ分析によって可視化された。これらの知見に基づき、既存の符号付けアルゴリズムおよび作業プロセスの改善に向けた考察を行う。

1）クラスタリングに基づく診療領域別の不一致要因分析

各クラスタの不一致パターンには、以下の通り診療領域固有の特徴が見られた。

- Cluster 0（循環器疾患）： 末尾に「Z」が付与されたコードが基本コードに置換されるケースが多く、詳細度よりも基本分類を優先する実務上の判断が反映されている。
- Cluster 1（外傷）： 骨折部位の詳細差異が主因である。これは傷病名に部位情報が不足している場合、既存のルールに従い詳細不明または代表的な部位を選択する手作業が発生しているためである。
- Cluster 2（感染症）： 原因菌やウイルス種の違いによるコード差異が頻出している。これらは辞書側の網羅性や、臨床病名と統計分類上の名称の乖離に起因する。
- Cluster 3（泌尿器・産科関連）： 妊娠合併症等の詳細分類において、患者の背景情報（妊娠週数等）を考慮した手付け判断が優先されている。
- Cluster 4（眼科・精神疾患）： 白内障の分類（老年性等）や認知症関連において、年齢等の条件分岐を伴うコード選択に差異が生じている。
- Cluster 5（皮膚・筋骨格疾患）： 皮膚疾患の分類体系や脊椎疾患の詳細部位において、辞書コードよりも手付けによる詳細化が図られている。
- Cluster 6（妊娠・管理関連）： 妊娠管理や産科合併症において、調査票の「受療の状況」等の文脈を反映したコード修正が行われている。
- Cluster 7（その他）： 複合診断や後遺症に関し、初筆優先ルール（原則として一番左側に記載の傷病名を選択する）に基づく修正が多く見られる。

本分析は ICD の章構造を一定程度再現しているが、これは診療領域ごとの対策を可能にする意図的な設計である。一方で、Cluster 7 における J コード（呼吸器）から U コード（特殊目的）への移行や、Cluster 3 における他領域から O コード（産科）への修正などは、文字列の近さを超えた「臨床的優先順位」や「実務上の判断ルール（初筆優先）」という章横断的な不一致構造を可視化している。

2）既存アルゴリズムおよびプログラムへの改善提案

分析で得られた頻出パターンを、既存の「職員独自作成プログラム（xx2.shf）」と「AC システム」の役割分担に基づき反映させることで、精度の向上と作業の効率化を図る。

- 表記揺れ補正 PG（xx2.shf）による前処理の拡張：Cluster 0 や 5 等で確認された「Z 付きコード」や「代表的な表記揺れパターン」を、既存プログラムの「漢字・記号変更ルール」や「誤字修正一覧」シートに直接定義することを提案する。特に、意味が通じる範囲での最小文字数による置換設定を精緻化することで、後工程の形態素解析に回る前の「完全一致マッチング率」を論理的に向上させることが可能である。
- 辞書および AC システムにおける参考情報付与の改善：xx2.shf を通過したデータに対し形態素解析を行う「AC システム」において、Cluster 2（感染症）等の原因菌による差異に対応するため、既存の辞書へのキーワード追加や優先順位の最適化を行う。また、既存の「アンマッチ基礎表（Excel）」作成時に、特定のキーワードが含まれる場合に「過去の修正事例」を自動提示する仕組みをマクロ等の処理工程に追加・改善する。
- 目印コード（ZZZZZ 等）の抽出条件の精緻化：Cluster 1（外傷）の部位不明や、Cluster 3（産科関連）の母子判定など、傷病名のみではコード決定が困難なケースに対し、特定のキーワード（「右上？骨」等）に基づき「目印コード（ZZZZZ、O コード等）」を自動付与する論理を強化する。これにより、ステップ 5（個別修正）において確認が必要な対象をより正確に抽出し、不適切なコードがそのまま計上されるリスクを低減できる。

3）手付け符号付け作業の標準化と教育への応用

AI 等による自動化に頼らずとも、不一致が発生しやすい領域を明確化することで、作業者の判断アルゴリズムを平準化することが可能と考えられる。

- 標準作業手順（SOP）の再定義：Cluster 0, 1, 5 等で抽出された詳細分類の不一致例を、実務担当者向けの「符号付け手順書」や「引き方のコツ」の具体的例示としてフィードバックする。
- 判断ルールの徹底： ステップ 4 における「詳細不明コード（NOS）の回避」や「初筆優先」といった既存の原則を、分析で判明したエラーリスクの高い具体例を用いて再徹底することで、人的なコーディング精度の向上を図る。

E 結論

本研究では、形態素解析による判定 ICD コードと手付け ICD コードの不一致パターンをクラスタリングにより分析し、診療領域ごとの特徴を明らかにした。その結果、循環器疾患、外傷、感染症など特定領域で体系的な不一致傾向が確認された。これらの知見は、既存の「表記揺れ補正 PG」の定義更新、目印コードの抽出ロジックの改善、および作業手順の標準化といったルールベースの対策に直接活用可能である。現行の手法を維持しつつも、一連の作業アルゴリズムを論理的に洗練させることで、符号付け精度の向上と実務の効率化が十分に可能であることが示唆された。

F 健康危険情報

なし

G 研究発表

1.論文発表

なし

2.学会発表

なし

H知的財産権の出願・登録状況（予定を含む。）

1.特許取得

なし

2.実用新案登録

なし

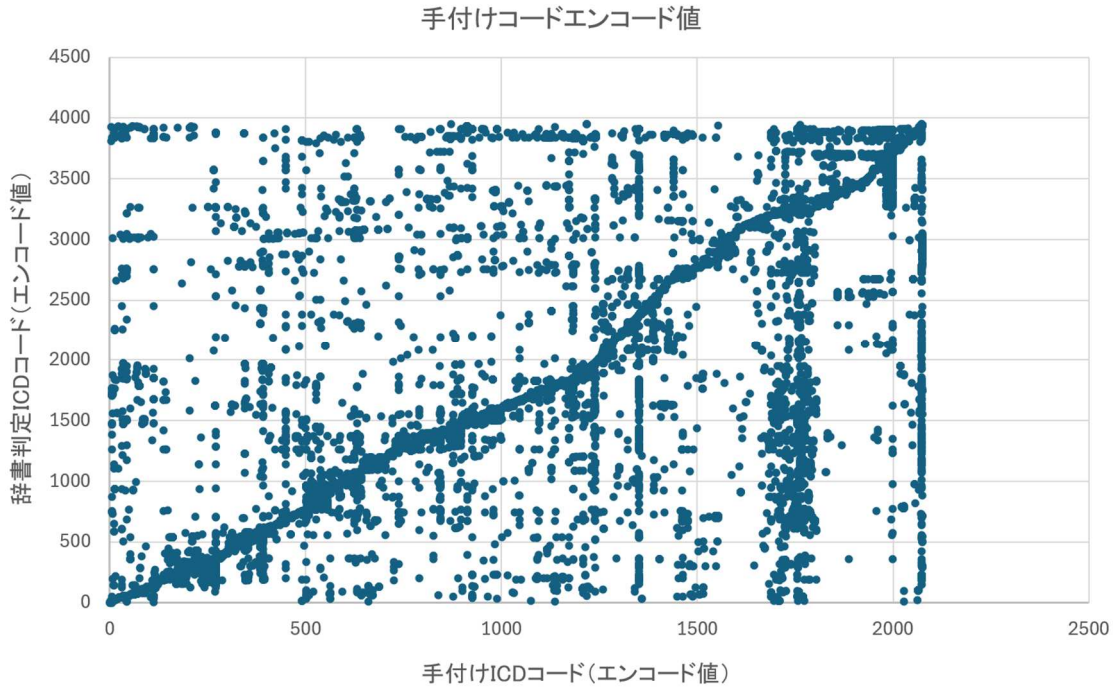


図１　クラスタリング前の対象語数

（形態素解析による判定 ICD コードと手付け ICD コードのペア）の分布

表５　クラスタ別の代表的な上位パターン（上位５例）

クラスタ No	クラスタ内の語数	代表ペア		語数	差異の特徴
		R5_AC	手付け		
Cluster0：循環器疾患	5363	I50.9Z	I50.9	222	循環器疾患の詳細差
		I63.9	I63.5	133	循環器疾患の詳細差
		J69.0Z	J69.0	133	呼吸器疾患の詳細差
		K63.5B	K63.5	85	その他
		I63.9	I63.3	79	循環器疾患の詳細差
Cluster1：外傷	5841	S06.5B	S06.50	192	外傷・損傷の部位差
		S32.0F	S32.00	151	外傷・損傷の部位差
		S72.0C	S72.00	100	外傷・損傷の部位差
		S72.1C	S72.10	78	外傷・損傷の部位差
		U97	R69	76	妊娠・管理関連
Cluster2：感染症	3864	B02.9	B02.2	85	感染症の原因差
		D27.Z	D27	70	腫瘍コードの詳細差
		D48.7	D48.1	53	腫瘍コードの詳細差
		D64.9Z	D64.9	48	腫瘍コードの詳細差
		D32.9	D32.0	44	腫瘍コードの詳細差
Cluster3：泌尿器・産科関連	3321	N18.9	N18.5	60	その他
		M86.9B	M86.97	58	筋骨格疾患の詳細差
		U99	O34.2	55	妊娠・管理関連
		U99	O42.9	39	妊娠・管理関連
		U99	O36.4	34	妊娠・管理関連
Cluster4：眼科・精神疾患	2807	H40.9	H40.1	57	その他
		H02.0Z	H02.0	55	その他
		H26.9	H25.0	41	その他
		H40.9	H40.5	41	その他
		F03	G30.9	39	その他
Cluster5：皮膚疾患・筋骨格疾患	5322	L30.9	L85.3	164	皮膚疾患の分類差
		M48.0	M48.06	130	筋骨格疾患の詳細差
		M17.9	M17.1	98	筋骨格疾患の詳細差
		M17.9	M17.0	83	筋骨格疾患の詳細差
		M81.9	M80.98	82	筋骨格疾患の詳細差
Cluster6：妊娠・管理関連	1800	R54	H35.3	79	その他
		R54	H25.0	35	その他
		R58.Z	I61.0	35	その他
		R56.8	H81.1	34	その他
		R54	H25.9	30	その他
	1702	J12.8F	U07.1	378	呼吸器疾患の詳細差

Cluster 7：その他		M62.5	U07.1	79	筋骨格疾患の詳細差
		J12.8H	U07.1	53	呼吸器疾患の詳細差
		J12.8H	U09.9	51	呼吸器疾患の詳細差
		J18.9	U07.1	33	呼吸器疾患の詳細差

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

患者調査と標準病名マスターの符号化方針相違事例における診療情報管理士による盲検下符号化の予備的検証

研究代表者　荒井　康夫　北里大学 未来工学部データサイエンス学科　准教授
研究分担者　石割　大範　国立健康危機管理研究機構 システム基盤整備局 係長

研究要旨

本研究班の前報により、令和 5 年（2023 年）患者調査で使用した辞書ファイル等と標準病名マスターの索引用語に対応する ICD コードとが相違する事例（符号化方針相違事例）の存在が確認された。本予備的検証では、ICD-10 章別の構成比を反映するように抽出した 124 症例を対象に、診療情報管理士 4 名（3 組）が、標準病名マスターを参照せず、患者調査 ICD と標準病名マスターICD の双方を併記したワークシート上で盲検下に ICD-10 コードを付与した。事例単位（n=124）の多数決一致先は、標準病名マスターICD が 80 件（64.5%、95%CI　55.8-72.4%）、患者調査 ICD が 34 件（27.4%、同：20.3-35.9%）、多数決不成立が 10 件 8.1%）であった。独立評価者対である A-B ペアの Cohen の κ 係数は 3 桁レベル 0.577（中等度）、章レベル 0.778（かなりの一致）であった。探索的所見として、術後状態 13 事例における原疾患章と Z90-Z99 との符号化方針の相違、および「コロナウイルス感染症」事例の方針差異が観察された。両方針は運用目的の違いに由来する構造的差異であり、本予備的検証は両方針の優劣を示すものではなく、今後検証規模および評価設計を拡張する必要がある。

研究協力者

小泉　麻美
　北里大学病院 医療の質評価室 診療情報管理士
村上　沙織
　北里大学病院 診療情報管理室 診療情報管理士
香取　麻紀子
　北里大学病院 診療情報管理室 診療情報管理士

A．研究目的

厚生労働省が実施する患者調査は、統計法（平成 19 年法律第 53 号）に基づく基幹統計のひとつであり、推計患者数や受療率の算出を通じて医療政策の立案・評価に用いられる重要な統計である[1]。患者調査では、医療機関が報告した自由記述の傷病名が WHO の「疾病及び関連保健問題の国際統計分類 第 10 回修正（ICD-10）」に準拠して符号化される[2]。一方、医療機関で広く用いられている標準病名マスター[3]は臨床用途を前提に整備されており、ICD-10 付与の判断基準が患者調査の符号化方針と一致しない場合がある。本研究班は前報により、令和 5 年（2023 年）患者調査で使用した辞書ファイル等と標準病名マスターの索引用語に対応する ICD コードとの突合により両者が一致しない事例の存在を確認した。

本予備的検証は、これら方針差異事例について、医療機関で符号化方針を担う専門家である診療情報管理士が標準病名マスターを参照しない条件下でいかなる符号を付与するかを実証的に検討することを目的とする。具体的には、（1）方針差異の

もとでの判定がいずれの方針側に偏る傾向にあるか、(2) 専門家間での判定の一致度はどの程度か、を明らかにする。本検証は、今後の本格的な検証研究に向けた実施可能性検討の位置づけで行った。

B．研究方法

1．対象事例の選定

令和 5 年患者調査で得られた傷病名の種類数(111,984 件)から、本研究班の前報の方法に従い、(i) 医療機関から報告された傷病名（以下、収集された傷病名）の正規化前文字列が標準病名マスターの索引用語と一致し、(ii) 患者調査において符号化された診断 ICD-10 準拠統計基準のコード（以下、符号化済み ICD コード）と当該索引用語に対応する標準病名マスター由来の ICD-10 コードとが相違する事例を「符号化方針相違事例」として抽出した。当該事例は前報において 1,184 件（索引一致かつ符号化済み ICD コードと ICD コード本体が相違する事例 746 件、および左右の指定差異により符号化済み ICD コードと ICD コード本体が相違する事例 438 件の合計）であった。

本検証では母集団 1,184 件の ICD-10 章別構成比を反映するように抽出した 124 症例を対象とし(標本割合 10.6%)、各事例には (ア) 患者調査 ICD (患者調査の符号化済み ICD コード) と (イ) 標準病名マスターICD (標準病名マスターの索引用語に対応する ICD-10 コード) の二つを併記したワークシートを作成した。なお、両者のいずれが患者調査由来かは、評価者には開示しなかった。

2．評価者と盲検性

診療情報管理士 4 名（うち 1 名は本分担研究者であるが、本検証用テストデータの作成には関与していない）が、標準病名マスターを参照しない条件下で、各事例に ICD-10 コードを手付けで付与した。担当区分は 3 組（うち 1 組は 2 名共同付与）とし、本報告では組の特定を避けるため、結果の記述では 3 組をそれぞれ「専門家 A」「専門家 B」「専門家 C」と表記する（割当順序は無作為)。

評価者には、各事例について「入力語」「索引用語」「正規化後表記」「病名表記」「参考 ICD-A」「参考 ICD-B」を併記したワークシートを配布し、別途設けた「確認結果」列に評価者が ICD-10 コードを記入する形式とした。参考 ICD-A と参考 ICD-B は、それぞれ患者調査の符号化済み ICD コードおよび標準病名マスター由来の ICD コードに対応するが、評価者にはいずれが患者調査由来かを開示しなかった（盲検評価)。評価者は標準病名マスターを参照せず、各自の判断により ICD-10 コードを付与することとした。なお、本ワークシート構成では参考 ICD-A・参考 ICD-B の双方が評価者に提示されているため、評価者の付与結果は両参考 ICD による選択の影響を受けうる。実際の集計でも、評価者の付与結果は対象 124 事例の約 90%（B：92.7%、C：91.9%、A：89.5%）が 4 桁レベルで参考 ICD-A または B のいずれかに完全一致しており、本検証は実質的に 2 候補併記下での盲検選択評価としての性格を有する。

3．評価指標と統計解析

主要評価指標は事例単位（n=124）の多数決一致先とした。各事例について 3 組の付与結果のうち 2 組以上が同一の参考符号と 3 桁レベルで一致した場合に多数決成立とみなし、対象 124 事例全体を分母として、患者調査 ICD 多数、標準病名マスター ICD 多数、多数決不成立の件数と割合および Wilson 法による 95%信頼区間を集計した。

本検証は予備的検証であり、仮説検定は行わない。多重比較の調整も行っていない。

副次評価指標として、各専門家による手付け ICD-10 コードと患者調査 ICD・標準病名マスター ICD との一致を 4 桁完全一致・3 桁一致・ICD-10 章一致の 3 水準で延べ観測（最大 370 観測）として集

計した。これらの観測は同一事例に対する反復測定であり独立性が成立しないため、95%信頼区間はWilson 法による粗推定に加え、事例クラスタを考慮したノンパラメトリック・ブートストラップ法（事例単位再抽出、評価者効果は固定、2,000 回反復）による区間を併記した。

専門家間の一致度はペアごとの Cohen の κ 係数および3組間の Fleiss の κ 係数で評価した[4][5]。ただし、共同付与の専門家 C は 2 名の合意ベース判定であるため、C を含む解析の κ 係数は独立評価者を仮定する前提に対し過大評価方向に偏りうる。本報告では独立評価者対である A-B ペアの Cohen の κ を主要解釈の対象とし、Fleiss の κ は参考値として併記する。κ 係数の 95%信頼区間はノンパラメトリック・ブートストラップ法（2,000 回反復）により算出し、解釈には Landis & Koch の基準[6]を参照した。一致率の 95%信頼区間は Wilson 法[7]により算出した。

感度分析として、共同付与の専門家 C を除外した A-B ペア（独立な評価者対）における両者一致事例の集計を併せて実施した。なお、本感度分析は両者一致事例における方針側分布の確認であり、全体集団における選好率を直接示すものではない。サブグループ別の集計（章別、術後状態、感染症章単一事例）は事後的に同定された記述統計であり、仮説検定は行わない。解析には Python 3.12（pandas、scikit-learn、statsmodels、NumPy、SciPy）を使用した。

（倫理面への配慮）本検証で用いたデータは令和 5 年（2023 年）患者調査で使用した辞書ファイル等であり、個人を特定し得る情報は含まれない。検証結果の集計にあたっては、専門家個人を特定可能な形での結果記述を避けるため、専門家 A・B・C の匿名表記とした。

C．研究結果

１．主要結果（事例単位、n=124）

対象 124 事例における事例単位の多数決一致先を表 1 に示す。多数決一致先が標準病名マスター ICD コードであった事例が 80 件（64.5%、95%CI：55.8-72.4%）、患者調査 ICD-10 準拠統計基準のコードであった事例が 34 件（27.4%、同：20.3-35.9%）、3 組すべてが 3 桁レベルで異なる付与（多数決不成立）が 10 件（8.1%）であった。本標本では多数決一致先として標準病名マスターICD コードが選ばれた事例の割合が患者調査 ICD の 2 倍を超える分布を示したが、これは予備標本における選択分布の記述にとどめ、両方針の優劣を示すものではない。

２．副次結果：評価者間の一致度

専門家ペアごとの一致率および Cohen の κ 係数を表 2 に示す。独立な評価者対である A-B ペアの Cohen の κ 係数は 3 桁レベル 0.577（中等度の一致）、章レベル 0.778（かなりの一致）であった。A-B ペアの一致率は 3 桁レベル 58.1%、章レベル 79.8%であった。専門家 C を含むペア（A-C、B-C）の κ 係数（3 桁レベル 0.661-0.694、章レベル 0.829-0.837）は A-B ペアより高かったが、専門家 C が 2 名共同による合意ベース判定であることに留意する必要がある。3 組間の Fleiss の κ 係数は 3 桁レベル 0.647、章レベル 0.819（参考値）であった。Fleiss の κ は独立な評価者を仮定するため、本検証では A-B ペアの Cohen の κ を主要解釈の対象とする。

専門家別の延べ観測（最大 370 観測）における標準病名マスターICD への一致は 246 件（66.5%）、患者調査 ICD への一致は 108 件（29.2%）であった。これらは同一事例に対する反復測定であり観測の独立性が成立しないため、Wilson 法による粗推定（標準病名マスターICD：61.5-71.1%／患者調査 ICD：24.8-34.0%）に加え、事例クラスタを考

慮したブートストラップ 95%信頼区間（標準病名マスターICD：60.2-72.2%／患者調査 ICD：23.5-35.0%）を併記する。クラスタを考慮した区間は粗推定よりも幅が広く、独立性の前提を置いた粗推定は不確実性を過小評価する。なお、本ブートストラップは事例クラスタのみを考慮し評価者効果は固定したため、評価者間の個性に伴う変動は反映されていない。

３．感度分析（副次解析：A-B ペアによる独立評価者対）

本感度分析は、独立評価者対である A-B ペアにおける主要結果の頑健性確認として位置づけられる副次解析である。A-B 両者の判定が 3 桁レベルで一致した事例は 72 件あり、両者が標準病名マスターICD で一致した事例は 57 件、両者が患者調査 ICD で一致した事例は 15 件であった。両者一致事例（n=72）における方針側分布の確認としては、両者標準病名マスターICD が 57/72 件（79.2%、95%CI: 68.4-86.9%）であった。なお、本値は両者一致事例における方針側分布の確認であり、両者不一致事例（52 件）の情報は反映していないため、全体集団における選好率を直接示すものではない。A・B の延べ観測（n=248）における標準病名マスターICD 一致は 161 件（64.9%）、患者調査 ICD 一致は 78 件（31.5%）であり、3 組合計の集計と整合する傾向を示した。

４．探索的所見

以下は事後的に同定された記述統計であり、仮説検定は行っていない。

＜対象事例の章別構成＞　対象 124 事例の患者調査 ICD に基づく ICD-10 章構成は、母集団 1,184 件の章別比率を反映するように抽出した結果に由来する。最も事例数の多い章は第Ⅱ章（新生物）28 件、第ⅩⅨ章（損傷・外因）23 件、第Ⅸ章（循環器系）10 件、第ⅩⅧ章（症状・徴候等）10 件、第ⅩⅠ章（消化器系）9 件であった。1-3 件の章が 8 章あり、第Ⅷ章（耳）、第ⅩⅥ章（周産期）、第ⅩⅩ章（外因）、第ⅩⅪ章（健康状態に影響を及ぼす要因）に該当する事例は対象に含まれなかった。各章の事例数の偏りは標本設計に由来し、本研究独自の発見ではない。章単位の統計的推測は本検証の標本設計の射程外である。

＜3 組の付与が分かれた事例＞　3 組の付与が 3 桁レベルですべて異なった事例は 10 件であった。これらには、（１）原疾患の章と症状の章の選択に幅があるもの（例：「肝門部癌」C22 と C24、「局所性痙攣」G40 と R25・R56）、（２）腫瘍の良悪性判定の差異（例：「椎間関節性腰痛症」M47 と M51）、（３）先天性と後天性の判定差異（例：「肺動脈弁下狭窄症」I37 と Q24、「副腎異常」E27 と Q89）、（４）損傷の章と臓器系の章の差異（例：「会陰部血腫」S30 と N30 と O71）が含まれた。専門家 C においては判定困難を理由とする自由記述記載が 3 件あり、いずれも判定不能の意思表示と解釈できる。

＜術後状態に関する事例＞　対象事例には、入力語に「○○術後」「○○術後状態」「○○術後遺症」「○○移植後」「○○留置（状態）」等の術後状態を記述する事例が 13 件存在した。これらの事例における患者調査 ICD は原疾患の章（第Ⅱ章 C 章、第Ⅸ章 I 章、第ⅩⅠ章 K 章等）または第ⅩⅧ章 R 章に符号化されている一方、標準病名マスターICD は多くの場合、第ⅩⅪ章 Z90-Z99（器官又は組織の後天的欠損及び植込み・置換術後の状態）に該当するコードに符号化されていた（例：「冠動脈バイパス術後」I251／Z951、「早期胃癌術後」C169／Z080、「前立腺癌術後状態」C61／Z988）。本検証における 3 組の付与は、これら術後事例の大半で標準病名マスターICD（Z 章）への一致を示した。

＜感染症章の単一事例＞　対象事例には、入力語「コロナウイルス感染症」が 1 件含まれた。当該事

例は、患者調査では新型もしくはCOVID-19の明記がなくともU071（COVID-19、ウイルスが特定されたもの）として符号化されていた。一方、標準病名マスターICDは入力文字列に対応する索引用語の登録に従いB342（コロナウイルス感染症、部位不明）であった。3組すべてが標準病名マスターICD側（B342系統）を選択した。

D．考察

１．方針差異事例における付与傾向

本予備的検証では、対象124事例における事例単位の多数決一致先のうち、標準病名マスターICDであった事例が64.5%（95%CI 55.8-72.4%）、患者調査ICDであった事例が27.4%（同：20.3-35.9%）、多数決不成立が8.1%であり、両割合の差は約37ポイントであった。本割合差は予備標本における選択分布の記述にとどまり、（１）一般集団における方針差異の存在を示すものではなく、（２）両方針のいずれが妥当であるかの判定材料となるものでもない。

本検証の評価者は標準病名マスターを参照しない条件下で判定したが、ワークシートには参考ICD-A（患者調査の符号化済みICDコード）と参考ICD-B（標準病名マスター由来のICDコード）が併記されていたため、評価者の判定はこれら2候補に強く影響された可能性がある。実際の付与結果では、評価者の付与の約90%が4桁レベルで参考ICD-AまたはBのいずれかと完全一致しており、本検証は実質的に2候補併記下での盲検選択評価としての性格を持つ。したがって、本検証で観察された割合差は、「専門家が独立に思考した結果としてB選好が出た」という解釈には直接つながらず、「2候補併記下において評価者がBを選びやすい傾向にあった」という限定的な解釈にとどめるのが妥当である。

患者調査の符号化方針は、統計法に基づく基幹統計として、WHOのICD-10分類原則に準拠しつつ、疾病別の医療需要および受療動向の把握を目的として整備されたものである。これに対し、標準病名マスターは電子カルテ等での臨床用途を主目的に整備された病名集であり、索引用語とICD-10との対応関係はその目的に沿って構築されている。両方針はそれぞれの目的に照らせばいずれも妥当であり、方針差異は優劣の問題ではなく、運用目的の違いに由来する構造的な差異である。

本検証で観察された標準病名マスターICDへの傾きは、診療情報管理士が日常業務において標準病名マスターを介してICD-10コードを取り扱うこととの関連で理解できる所見であり、いずれの方針が妥当であるかを示すものではない。

２．専門家間の一致度

本検証では、独立評価者対であるA-BペアのCohenのκ係数を主要指標とし、これは3桁レベル0.577（中等度の一致）、章レベル0.778（かなりの一致）であった[6]。ただし、本検証は独立評価者2名（A・B）に基づく予備的観察であり、評価者数および標本数の制約から、専門家間の一致度に関する本検証の評価は暫定的なものにとどまる。専門家Cを含むペア（A-C、B-C）のκ係数（3桁レベル0.661-0.694）はA-Bペアより高かったが、Cが2名の合意ベース判定であるため一致度が高めに出る方向の偏りが含まれうる。Fleissのκ係数（3桁レベル0.647、章レベル0.819）は独立な評価者を仮定する前提に対し過大評価方向に偏りうるため、参考値として位置づけられる。

章レベルの高い一致（A-Bペアκ=0.778）に対し3桁レベルでは0.577と相対的に低下しており、章の選択（疾患群の所属）には共通した判断枠組みが存在する一方、3桁分類への落とし込み（特定病態の選択）では複数の妥当な選択肢が併存し判断が分かれ得ることを示している。3組すべてが異なる

付与をした 11 事例には、原疾患か症状か、先天性か後天性か、損傷の章か臓器系の章かといった ICD-10 運用上の論点が含まれており、これは方針差異とは別の、原資料の情報不足や複数の合理的解釈の併存に由来する判定の困難性を示す。

これらの所見は、収集された傷病名のみからは適切な ICD-10 コードの選択が困難な事例が存在することを示すものであり、より適切な符号化のためには原資料となる診療情報を確認する必要性が示唆される。

３．標本設計と章別分析の解釈

対象事例の章別構成は、母集団 1,184 件の章別出現比率を反映するように抽出して決定した。これは母集団全体の傾向を縮約的に観察し、実態に近い条件下で予備的所見を得ることを優先した結果である。この設計により、第Ⅰ章、第Ⅲ章、第ⅩⅫ章等は事例数が 1 件にとどまり、第Ⅴ章・第Ⅵ章・第Ⅶ章・第Ⅹ章・第Ⅻ章は 2-3 件と少数となった。これらの章における選択件数は当該章の傾向を一般化する根拠としては不十分であり、章単位での統計的推測（章ごとの選好割合の信頼区間推定や章間比較等）は本検証の標本設計では行わない。章別の傾向を統計的に評価するには、章ごとに必要標本数を確保した次段階の検証が必要である。

４．術後状態に関する方針差異

対象事例には、入力語に術後状態を記述する事例が 13 件含まれた。これらの事例において、患者調査 ICD は原疾患の章（第Ⅱ章 C 章、第Ⅸ章Ⅰ章、第Ⅺ章 K 章等）または第ⅩⅧ章 R 章に符号化されていた一方、標準病名マスターICD は第ⅩⅪ章 Z90-Z99（器官又は組織の後天的欠損及び植込み・置換術後の状態）に該当するコードに符号化されていた。

患者調査における原疾患章への符号化は、基幹統計として疾病別の医療需要および受療動向を把握するという当該調査の目的に沿った方針運用と理解できる（例えば「早期胃癌術後」を C169 に符号化することで、胃癌に関する受療動向の集計に当該症例が反映される）。標準病名マスターにおける Z90-Z99 への符号化は、電子カルテ等の臨床用途における手術後の継続的な状態を表す概念分類として運用されている。両者はそれぞれ異なる目的・前提に基づく方針であり、いずれもその方針内で整合する符号化結果である。本検証における 3 組の手付け付与は、これら術後事例の大半で標準病名マスターICD への一致を示した。

５．コロナウイルス感染症の符号化

入力語「コロナウイルス感染症」の事例は、患者調査では新型もしくは COVID-19 の明記がなくとも U071（COVID-19、ウイルスが特定されたもの）として符号化されていた。一方、標準病名マスターICD は入力文字列に対応する索引用語の登録に従い B342（コロナウイルス感染症、部位不明）であった。本検証のワークシートでは U071 と B342 の双方が参考 ICD として併記された状態で評価者に提示され、3 組すべてが参考 ICD-B 側（B342）を選択した。

患者調査における U071 への符号化は、新興感染症の発生状況を統計上把握するという当該調査の方針運用に基づく。標準病名マスターにおける B342 への符号化は、入力文字列に対応する索引用語との関係に基づく方針運用である。両者はそれぞれの方針および目的に基づく符号化結果であり、本検証ではその相違が観察された。本事例は単一事例であり、検証規模を拡張した上で、同種事例の出現状況および各方針における取り扱いを引き続き観察することが、今後の検討項目となる。

６．本検証の限界と今後の展望

本検証は予備的なものであり、以下の限界がある。

＜標本に関する限界＞　対象は母集団 1,184 件のうち章別構成比を反映するように抽出した 124

件にとどまり、母集団全体の傾向を統計的に推定するには標本数が十分でない。事例数 1-3 件の章が 8 章を占めるため章別の統計的推測は本検証の標本設計の射程外であり、章単位での結論は導出しない。

＜評価者に関する限界＞　評価者は 3 組（4 名）に限られ評価者数による不確実性が大きく、Fleiss の κ 係数は評価者の代表性を仮定するが、本検証は便宜標本による評価者対であり、より広範な診療情報管理士集団への一般化可能性は本検証では確立されない。3 組のうち 1 組は 2 名共同付与のため、他の 2 組（各 1 名）と判定の独立性が完全には等価でなく、評価者単位の解析（Fleiss の κ）にも影響しうる。共同付与の専門家 C は 2 名の合意ベース判定であり、個別データが記録されていないため、4 者として分離した感度分析は本検証では不能であった。

＜統計処理に関する限界＞　本検証の主要結果は事例単位（n=124）の多数決一致先による割合と Wilson 法 95%信頼区間で構成されるが、専門家別の延べ観測（最大 370 観測）は同一事例に対する反復測定であり観測の独立性が成立しないため、独立性を仮定した Wilson 法による信頼区間は不確実性を過小評価する。本報告では延べ観測について事例クラスタを考慮したブートストラップ 95%信頼区間を併記したが、本ブートストラップは事例クラスタのみを再抽出し評価者効果は固定したため、評価者間の個性に伴う変動は反映されていない。より厳密には階層モデル（GLMM）による評価者効果と事例効果の同時推定が必要であり、これは次段階の課題である。

本検証では仮説検定は行わず、効果量（割合）と 95%信頼区間による記述統計的提示にとどめた。本検証は評価者が日常業務で標準病名マスターを介して ICD-10 コードを扱う診療情報管理士による評価であり、かつワークシート上に参考 ICD-A および B が併記されているため、「両方針への一致率が等しい」とする帰無仮説は構造上成立しにくく、形式的な仮説検定は実質的な意味を限定的にしか持たない。

感度分析（A-B ペアによる独立評価者対）は、両者一致した 72 事例における方針側分布の確認であり、両者不一致 52 事例の情報は反映していない。したがって、当該感度分析の数値（両者 B 79.2%）は両者一致事例における選好分布の記述であって、全体集団における選好率を直接示すものではない。Fleiss の κ 係数（0.647）は専門家 C が 2 名共同付与であるため独立評価者を仮定する前提に対し過大評価方向に偏りうる。本検証では独立評価者対である A-B ペアの Cohen の κ（3 桁 0.577、章 0.778）を主要指標とした。

本検証は予備的検証のため多重比較の調整は行っておらず、サブグループ別の集計（章別、術後状態、感染症章単一事例）は事後的に同定された記述統計として位置づけられる。多数決不成立事例（n=10）等の小数事例における信頼区間の精度は限定的である。

＜評価設計に関する限界＞　評価者は標準病名マスターを参照しない条件としたが、長年の業務を通じて獲得した符号化判断の枠組みそのものを統制することは方法論上困難である。また本検証では、ワークシートに参考 ICD-A／参考 ICD-B の二候補を併記し評価者にいずれが患者調査由来かを伏せた盲検評価を採用したが、二候補の併記自体が評価者の連想を誘導する可能性は残る。自由記述方式での独立評価の併用や、別形式での提示による感度分析は今後の課題である。

＜今後の展望＞　今後は対象事例数および評価者数を拡大し、独立評価を原則とする評価設計のもとで、章別の事例数バランスの改善、診療情報管

理士以外の関連職種との比較、方針差異事例における判定根拠の質的分析、術後状態およびコロナウイルス感染症等のように方針差異が生じる領域における観察の継続、を含む本格的な検証研究が必要である。

E．結論

令和 5 年（2023 年）患者調査で使用した辞書ファイル等と標準病名マスターの符号化方針が乖離する 124 事例を対象に、診療情報管理士 4 名(3 組)が標準病名マスターを参照しない条件下で、参考 ICD-A（患者調査 ICD）および参考 ICD-B（標準病名マスターICD）の双方が併記されたワークシート上で盲検下に ICD-10 コードを付与した結果、対象 124 事例における事例単位の多数決一致先は、標準病名マスターICD が 80 件（64.5%、95%CI 55.8-72.4%）、患者調査 ICD が 34 件（27.4%、同：20.3-35.9%）、多数決不成立が 8.1%であった。独立評価者対（A-B ペア）における Cohen の κ 係数は 3 桁レベル 0.577（中等度の一致）、章レベル 0.778（かなりの一致）であった。なお、評価者の付与結果は対象事例の約 90%が 4 桁レベルで参考 ICD-A または B のいずれかに完全一致しており、本検証は実質的に 2 候補併記下での盲検選択評価としての性格を有する。したがって本結果は、2 候補併記下で評価者が B を選びやすい傾向にあったという限定的解釈にとどめ、両方針の優劣を示すものではない。探索的所見として、術後状態に関する記述 13 件における原疾患章と Z90-Z99（第ⅩⅪ章）との符号化方針の相違、および入力語「コロナウイルス感染症」における符号化方針差異の事例（n=1）が観察された。これらは患者調査と標準病名マスターのそれぞれの方針および目的に沿った符号化結果である。両方針はそれぞれの目的に照らせばいずれも妥当であり、方針差異は優劣の問題ではなく、運用目的の違いに由来する構造的差異として位置づけられる。本検証は予備的なものであり、2 候補併記による選択誘導の可能性、観測の独立性、評価者の代表性、評価設計の限界などを踏まえ、今後さらに多数の専門家の協力を得て検証規模および評価設計を拡張する必要がある。

F．健康危険情報

該当なし。

G．研究発表

1．論文発表

該当なし。

2．学会発表

該当なし。

H．知的財産権の出願・登録状況

（予定を含む。）

1．特許取得

該当なし。

2．実用新案登録

該当なし。

3．その他

該当なし。

参考文献

[1] 厚生労働省. 令和 5 年（2023 年）患者調査の概況. 2024.

[2] 世界保健機関（WHO）. International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10). 厚生労働省 監訳. 疾病、傷害及び死因の統計分類提要 ICD-10（2013 年版）. 2016.

[3] MEDIS-DC. ICD10 対応 標準病名マスター. Ver.5.17.

[4] Cohen J. Educational and Psychological Measurement. 1960; 20(1): 37-46.

[5] Fleiss JL. Psychological Bulletin. 1971; 76(5): 378-382.

[6] Landis JR, Koch GG. Biometrics. 1977; 33(1): 159-174.
[7] Wilson EB. Journal of the American Statistical Association. 1927; 22(158): 209-212.

表 1　事例単位の多数決一致先（対象 124 事例を分母）

多数決一致先	件数	割合（%）	95%CI (%)
標準病名マスターICD	80	64.5	55.8–72.4
患者調査 ICD	34	27.4	20.3–35.9
多数決不成立（3 組とも異なる）	10	8.1	4.4–14.2
いずれとも不一致	0	06	0.0–3.0

割合は対象 124 事例を分母とした。CI は Wilson 法。

表 2　専門家ペアごとの一致率および Cohen の κ 係数

ペア	n	4 桁 一致率	3 桁 一致率	章 一致率	κ（3 桁）	κ（章）
A–B（独立対）	124	53.2	58.1	79.8	0.577	0.778
A–C	122	62.3	66.4	84.4	0.661	0.829
B–C	122	66.4	69.7	85.2	0.694	0.837

数値は一致率(%)。C は C のみ 2 名共同付与。A–B ペア（太字）は独立な評価者対。

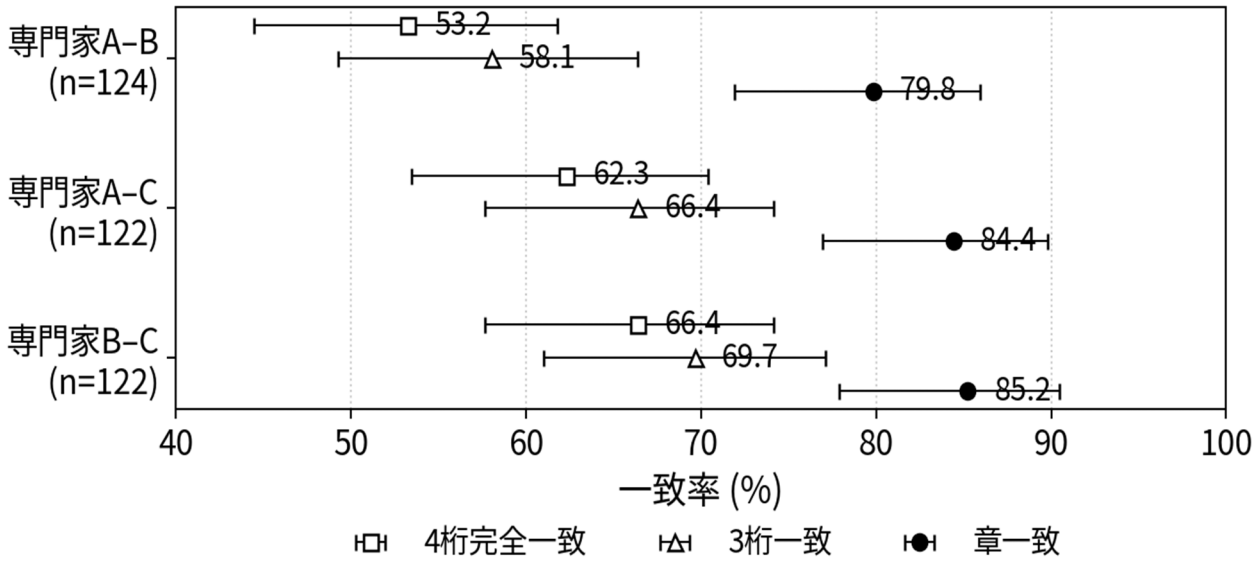


図1　ペアごとの一致率と 95%信頼区間（Wilson 法）

研究成果の刊行に関する一覧表

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					

令和8年5月25日

厚生労働大臣 殿

機関名　北里大学

所属研究機関長　職　名　学長

氏　名　砂塚　敏明

次の職員の令和７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　厚生労働行政推進調査事業費補助金（厚生労働科学特別研究事業）

２．研究課題名　患者調査の傷病符号付けプロセスの最適化と効率化に資する研究

３．研究者名　（所属部署・職名）　北里大学未来工学部データサイエンス学科・准教授

（氏名・フリガナ）　荒井　康夫（アライ　ヤスオ）

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）	□　■	□		□
遺伝子治療等臨床研究に関する指針	□　■	□		□
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	□　■	□		□
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：　　）	□　■	□		□

（※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）未審査の場合は、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 ■　未受講 □

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 ■　無 □（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 ■　無 □（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 ■　無 □（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 □　無 ■（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。

令和８年４月20日

厚生労働大臣　殿

機関名　長野県厚生農業協同組合連合会
佐久総合病院
所属研究機関長　職　名　統括院長
氏　名　渡辺 仁

次の職員の（元号）　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　：令和７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）

２．研究課題名　患者調査の傷病符号付けプロセスの最適化と効率化に資する研究

３．研究者名　（所属部署・職名）長野県厚生農業協同組合連合会佐久総合病院 総合医療情報センター長

（氏名・フリガナ）　西本　寛（ニシモト ヒロシ）

４．倫理審査の状況

	該当性の有無	左記で該当がある場合のみ記入（※1）		
	有　無	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）	□　■	□		□
遺伝子治療等臨床研究に関する指針	□　■	□		□
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	□　■	□		□
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：　）	□　■	□		□

（※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 ■　未受講 □

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 ■　無 □（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 ■　無 □（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 ■　無 □（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 □　無 ■（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。

令和８年３月３１日

厚生労働大臣　殿

機関名　川崎医療福祉大学

所属研究機関長　職　名　学長

氏　名　椿原　彰夫

次の職員の令和７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　：令和７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）

２．研究課題名　患者調査の傷病符号付けプロセスの最適化と効率化に資する研究

３．研究者名　（所属部署・職名）　川崎医療福祉大学医療福祉マネジメント学部　医療秘書学科特任教授

（氏名・フリガナ）　阿南　誠（アナン　マコト）

４．倫理審査の状況

	該当性の有無	左記で該当がある場合のみ記入（※1）		
	有　無	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）	□　☑	□		□
遺伝子治療等臨床研究に関する指針	□　☑	□		□
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	□　☑	□		□
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：　）	□　☑	□		□

（※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 ☑　未受講 □

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 ☑　無 □（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 ☑　無 □（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 ☑　無 □（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 □　無 ☑（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。

令和８年３月２６日

厚生労働大臣　殿

機関名　国立大学法人東北大学

所属研究機関長　職　名　総長

氏　名　冨永 悌二

次の職員の令和７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　患者調査の傷病符号付けプロセスの最適化と効率化に資する研究

３．研究者名　（所属部署・職名）　大学病院・准教授

（氏名・フリガナ）　桜澤 邦男　タラサワ　クニオ

４．倫理審査の状況

	該当性の有無	左記で該当がある場合のみ記入（※1）		
	有　無	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）	□　■	□		□
遺伝子治療等臨床研究に関する指針	□　■	□		□
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	□　■	□		□
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：　　）	□　■	□		□

（※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 ■　未受講 □

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■ 無□（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■ 無□（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■ 無□（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■ 無□（有の場合はその内容：　研究実施の際の留意点を示した。　）

（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。

令和８年４月20日

厚生労働大臣　殿

機関名　山口県立総合医療センター

所属研究機関長　職　名　院長

氏　名　杉野 法広

次の職員の（元号）　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　：令和７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）

２．研究課題名　患者調査の傷病符号付けプロセスの最適化と効率化に資する研究

３．研究者名　（所属部署・職名）　山口県立総合医療センター　医療企画室　室長補佐

（氏名・フリガナ）　來島　裕太（キジマ ユウタ）

４．倫理審査の状況

	該当性の有無	左記で該当がある場合のみ記入（※1）		
	有　無	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）	□ ■	□		□
遺伝子治療等臨床研究に関する指針	□ ■	□		□
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	□ ■	□		□
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：　）	□ ■	□		□

（※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 ■　未受講 □

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 ■　無 □（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 ■　無 □（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 ■　無 □（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 ■　無 □（有の場合はその内容：　研究実施の際の留意点を示した。）

（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。

令和８年４月２０日

厚生労働大臣　殿

機関名　地方独立行政法人大阪府立病院機構
大阪急性期・総合医療センター

所属研究機関長　職　名　　総長

氏　名　　嶋津　岳士

次の職員の（元号）　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　：令和７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）

２．研究課題名　患者調査の傷病符号付けプロセスの最適化と効率化に資する研究

３．研究者名　（所属部署・職名）大阪急性期・総合医療センター 医療情報部 診療情報管理室 総括主査

（氏名・フリガナ）森藤 祐史（モリトウ ヒロフミ）

４．倫理審査の状況

	該当性の有無	左記で該当がある場合のみ記入（※1）		
	有　無	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）	□　■	□		□
遺伝子治療等臨床研究に関する指針	□　■	□		□
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	□　■	□		□
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：　　）	□　■	□		□

（※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 ■　　未受講 □

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 ■　無 □（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 ■　無 □（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 ■　無 □（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 □　無 ■（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。

令和８年４月２０日

厚生労働大臣　殿

機関名　国立健康危機管理研究機構

所属研究機関長　職　名　理事長

氏　名　國土　典宏

次の職員の令和７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　：令和７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）

２．研究課題名　患者調査の傷病符号付けプロセスの最適化と効率化に資する研究

３．研究者名　（所属部署・職名）　国立健康危機管理研究機構　システム基盤整備局　係長

（氏名・フリガナ）　石割　大範（イシワリ　ヒロノリ）

４．倫理審査の状況

	該当性の有無	左記で該当がある場合のみ記入（※1）		
	有　無	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）	□　■	□		□
遺伝子治療等臨床研究に関する指針	□　■	□		□
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	□　■	□		□
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：　）	□　■	□		□

（※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 ■　未受講 □

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 ■　無 □（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 ■　無 □（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 ■　無 □（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 ■　無 □（有の場合はその内容：研究実施の際の留意点を示した。）

（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。