

厚生労働行政推進調査事業費補助金
(政策科学総合研究事業 (政策科学推進研究事業))

NDB のユーザビリティ向上を通じてクラウド上での
データ二次利用を推進するための研究 (25AA2002)

令和 7 年度 総括・分担研究報告書

研究代表者 明神 大也
(浜松医科大学 健康社会医学講座)

令和8 (2026) 年 5 月

目次

[令和7年度 報告書]

I. 総括研究報告

1. NDB のユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究

明神大也

研究要旨	1
A. 背景と目的	2
B. 研究方法	2
C. 研究結果	3
D. 考 察	5
E. 結 論	6
F. 健康危険情報	6
G. 研究発表	6
H. 知的財産の出願・登録状況（予定含む）	6

II. 分担研究報告書

2. NDB のユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究

ユースケース調査と汎用データセットの検討

明神大也、森由希子、牧戸香詠子、西岡祐一

研究要旨	7
A. 背景と目的	8
B. 研究方法	8
C. 研究結果	9
D. 考 察	11
E. 結 論	13
F. 健康危険情報	13
G. 研究発表	13
H. 知的財産の出願・登録状況（予定含む）	13
(図・表)	14

3. NDB のユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究

HIC のユーザビリティテストと改善提案

明神大也、村松圭司

研究要旨	18
A. 研究目的	18
B. 研究方法	18
C. 研究結果	20
D. 考 察	21
E. 結 論	21
F. 健康危険情報	21
G. 研究発表	21
H. 知的財産の出願・登録状況（予定含む）	21
別紙「課題及び改善事項一覧」	22

4. NDB のユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究

標準化された SQL データハンドリングフレームワークの提示

松居宏樹

研究要旨	24
A. 研究目的	24
B. 研究方法	25
C. 研究結果	27
D. 考 察	28
E. 結 論	30
F. 健康危険情報	30
G. 研究発表	30
H. 知的財産の出願・登録状況（予定含む）	30

5. NDB のユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究

利便性の高いクラウド環境の検討・提案

柏木公一、杉山雄大

研究要旨	31
A. 研究目的	32
B. 研究方法	33
C. 研究結果	34
D. 考 察	35
E. 結 論	36
F. 健康危険情報	37
G. 研究発表	37
H. 知的財産の出願・登録状況（予定含む）	37
（ 図 ）	38

6. NDB のユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究

諸外国の医療関連データ利活用・提供状況の調査

明神大也、牧戸香詠子

研究要旨	39
A. 研究目的	39
B. 研究方法	40
C. 研究結果	40
D. 考 察	45
E. 結 論	46
F. 健康危険情報	46
G. 研究発表	46
H. 知的財産の出願・登録状況（予定含む）	46

III. 研究成果の刊行に関する一覧表

著 書	48
雑 誌	48

IV. 倫理審査等報告書の写し

49

厚生労働行政推進調査事業費補助金（政策科学総合研究事業（政策科学推進研究事業））

総括研究報告書（令和7年度）

NDBのユーザビリティ向上を通じて
クラウド上でのデータ二次利用を推進するための研究

研究代表者 明神 大也（浜松医科大学 健康社会医学講座）

研究要旨

背景と目的: NDBは、レセプト情報や特定健診等情報を格納した公的データベースであり、第三者提供の法制化、死亡情報等の収載拡大、HICの運用開始により、利活用の基盤整備が進められている。一方で、プリセットデータがCSV形式で提供されることや、解析環境でコマンドライン操作が必要であることなど、ユーザビリティには課題が残る。本研究では、ユースケース調査、プリセットデータ改善、クラウド環境の操作性向上、技術供与等を通じて、NDB/HICの利便性を高め、医療等情報の二次利用を推進することを目的とした

方法: NDB-β、通年パネルデータセット等の利用申請を行い、HIC上での活用可能性を検証した。具体的には、過去のNDB提供申出成果物を対象としたユースケース調査、HIC/NDBガイドラインおよび利用者マニュアルに基づくユーザビリティテスト、標準化されたSQLデータハンドリング手法の整理、S3・Parquet・DuckDBを用いたクラウド分析基盤の検証を行った。さらに、韓国・フィンランド・スウェーデン・フランスの医療関連データ提供状況を調査し、研究用マスタ情報の共有を試みた。

結果: 論文47件、報告書16件を分析した結果、論文24件、報告書6件はHICで提供可能なデータ形式により実施可能と考えられた。一方、希少疾患や詳細な層別解析では特別抽出が必要であった。HICの解析環境やポータル等には改善課題が確認された。SQL処理では、解析単位、時間的アンカー、観察ウィンドウ等に基づく標準ワークフローを整理した。また、1年分620GBのデータをS3上にParquet化し、低コストで実用的な分析基盤の可能性を確認した。

考察: 本研究により、NDB/HICの利活用拡大には、汎用データセットの拡充、申請様式やポータルの操作性改善、再利用可能なSQL処理手法、低コストで柔軟なクラウド分析基盤、マスタ共有の促進が重要であることが示された。特に通年パネルデータセットの複数年化は、既存研究の一定割合をHIC上で代替可能にする可能性がある。また、S3・Parquet・DuckDBを組み合わせた基盤は、従来のRedshift依存によるコストや柔軟性の課題を補完し得る。

研究分担者

村松 圭司（千葉大学医学部附属病院）
森 由希子（京都大学医学部附属病院）
牧戸 香詠子（東京大学大学院 医学系研究科生物統計情報学）
松居 宏樹（東京大学大学院医学系研究科）
杉山 雄大（国立健康危機管理研究機構・国立国際医療研究所糖尿病情報センター）
柏木 公一（国立健康危機管理研究機構・国立看護大学校）
西岡 祐一（奈良県立医科大学）

研究協力者

川田 加奈子（浜松医科大学 健康社会医学講座）
古野 孝志（国立健康危機管理研究機構・国際医療協力局グローバルヘルス政策研究センター）
市瀬 雄一（国立健康危機管理研究機構・国際医療協力局グローバルヘルス政策研究センター）
加藤 源太（京都大学医学部附属病院 診療報酬センター／病床運営管理部）

A. 背景と目的

匿名医療保険等関連情報データベース（以下「NDB」という。）は、厚生労働省が法令に基づき収集・提供しているレセプト情報や特定健診・特定保健指導情報などを格納したデータベースである。2020年より民間事業者を含む第三者への提供が法制化されて以降、他の医療・介護データ等との連結解析、死亡情報等収載情報の拡大が進んでいる。また、医療DXの推進に関する工程表（令和5年6月2日医療DX推進本部決定）、規制改革実施計画（令和5年6月16日閣議決定）「NDBや公的統計データの利活用の円滑化・迅速化」に従って、オンライン上で利用申請や審査が可能となる二次利用ポータルおよびクラウド上の解析基盤として「医療・介護データ等解析基盤（HIC）」の運用も開始され、データの迅速提供・利活用促進のための提供体制整備が進められてきた。HICで迅速提供できるデータ形式は、サンプリングされた一年間のデータである「通年パネルデータセット」、サンプリングされた単月毎の断続データである「トライアルデータセット」、機微コード等をマスクしたレセプト全件を含む「NDB-β」の3種類とされている（2026年5月時点）。しかしプリセットデータがクラウド上にも関わらずCSVファイルで提供される、解析環境内ではコマンドラインの操作が必要である等、これまでのオンプレミス環境における解析環境と比較しても、ユーザビリティに関して多くの課題が残っている。

こうした中で、規制改革推進に関する中間答申（令和6年12月25日）においては、公的DBの仮名化情報について安全であるのみならず迅速かつ円滑に利用・解析を行うことができるクラウド環境について、HICとの関係を整理しつつ実現することとされている。

そこで本研究では、ユースケース調査、プリセットデータの改善提案、クラウド環境の操作性への提言、技術供与等を通じて、

NDB/HICのユーザビリティを向上し、医療等情報の二次利用を推進することを目的とする。

具体的には①ユースケース調査、②現行のHICのユーザビリティテストと改善提案、③汎用性の高いマスタやSQLクエリ等のプログラム作成、④汎用性の高いデータセットの検討・具体的な仕様作成、⑤利便性の高いクラウド環境の検討・提案、⑥HICを利用して解析可能なリサーチクエスチョンやその手法の例示、⑦研究者支援体制への協力それぞれの分野で研究を進める。

B. 研究方法

本研究は、上記を満たすために、以下のA～Fの研究を実施した。その前提として、NDB-βのオンプレミス環境での提供申出と、通年パネルデータセット及びNDB-βのHIC上での利用申請を行った。

A) ユースケース調査と汎用データセットの検討

HICで利用可能なデータセットの幅を広げるために、ユースケース調査をおこなった。具体的にはNDBデータの第三者提供に係る二次利用ポータルで公開されている「過去の提供申出の成果物」から、発表形式が論文または報告書（厚労科研に限る）に分類されている既存研究を対象とした。対象研究について、利用目的、分析に必要なデータ項目、対象疾患等を整理し、HICで提供可能なデータ形式に基づくユースケースの分類を行った。さらに、分類結果を踏まえ、HIC利用申請時に必要となる別添8「申出依頼テンプレート（抽出）」の記載例を作成し、具体的なデータ仕様の検討を行った。これは背景にある①と④に該当する。

B) HICのユーザビリティテストと改善提案

「匿名医療保険等関連情報データベース（NDB）の利用に関するガイドライン」及び「医療・介護データ等解析基盤（HIC）の利

用に関するガイドライン」(以下「HIC/NDBガイドライン」という。)、及び二次利用ポータルに掲載されている利用者マニュアルの記載に従って HIC のユーザビリティを確認し、HIC 解析環境・二次利用ポータル・関連ドキュメントに改善余地のある課題を発見した。これは背景にある②に該当する。

C) 標準化された SQL データハンドリングフレームワークの提示

データの取り扱いには一般に SQL を用いたデータハンドリングが行われる。しかし、SQL は記述の自由度が高いゆえに、研究者ごとに独自のクエリが書かれる傾向がある。そのため、クエリ設計の方針が担当者の経験と裁量に依存し、コードが引き継がれる際に意図が伝わらず、研究の再現性が損なわれることも少なくない。そのため、SQL のクエリを共有しても、そこには膨大な量のレビュープロセスが発生し、第三者によるコードレビューが困難である。また、エラーが潜在しても発見が難しい。また、数テラバイトに及ぶ元テーブルを不必要に全件スキャンしたり、同一内容のテーブルを重複してコピーしたりするクエリが散見され、計算資源を浪費する原因ともなっている。

大規模時系列医療データを疫学研究に使用しやすい形に整理するための、標準化されたデータハンドリングプロセスを一般化可能なように提示することである。具体的には、(1) 解析単位、(2) 観察のタイムライン、(3) マスタ整備という三要素を軸としたクエリ設計の考え方を示し、各ステップを体系的に解説した。これは背景にある③に該当する。

D) 利便性の高いクラウド環境の検討・提案

NDB における HIC では、Redshift・ローカル Postgres 等の解析環境が提供されている。一方、同環境においては、Redshift 永続テーブル作成権限の欠如、ローカル解析サーバのメモリ・ストレージ不足、外部参照データ取

込の困難、計算コストの増大といった複合的課題が存在する。

そこで HIC 内で提供されるクラウドストレージ(Amazon S3)と列指向ファイル形式 Parquet、軽量分析エンジン DuckDB を組み合わせた分析基盤を構築し、その実用性を検証した。これは背景にある⑤に該当する。

E) 諸外国の医療関連データ利活用・提供状況の調査

海外の医療データ提供の運用状況について調査(韓国・フィンランド・スウェーデン・フランス)を行った。その状況を踏まえ、解析環境の改善点、データ提供形式、マニュアル等の在り方等について検討し、利便性向上に資する提案を取りまとめる。これは背景にある①④⑤に該当する。

F) マスタ情報の提供

背景③にある汎用性の高いマスタが求められている事情として、NDB に含まれるレセプト情報は、医療機関が原則審査支払機関に提出したレセプトが、審査を経て、原則匿名化処理のみ行ったものであることが挙げられる。そのためいわゆる「コードの塊」になっており、研究テーマに沿った処理を研究者自身で行う必要がある。しかしながら研究テーマは研究者によって異なることから、多種多様なマスタが必要となるが、汎用性を示す根拠は乏しい。そこで二次利用ポータル上にある [コンテンツ] - [マスタ共有] を使いマスタ情報等の共有ができたり、研究者同士で誤りを指摘したりすることを目指すために、本研究担当者が NDB 等の研究で用いているマスタをアップロードすることとした。

C. 研究結果

上記方法に示した結果を下記に示す。なお各種 NDB の申出については、第 29 回匿名医療情報等の提供に関する専門委員会にて NDB-β のオンプレミス環境の提供承諾を、第 34 回匿名医療情報等の提供に関する専門委員会にて

通年パネルデータセット及び NDB-β の HIC 上での利用承諾（いわゆる迅速提供）を得た。

A) ユースケース調査と汎用データセットの検討

本研究の分析対象は論文 47 件、報告書 16 件であった。論文では処方動向、介入効果（治療）、疾患疫学、地域差分析に関する研究が多く、報告書では地域差分析、医療提供体制調査、処方動向に関する研究が多かった。HIC で提供可能なデータ形式に基づき分類した結果、論文 24 件（51.1%）、報告書 6 件（37.5%）が、通年パネルデータセット、トライアルデータセット、NDB-β のいずれかを用いた研究が実施可能であると考えられた。一方で、希少疾患や詳細な層別解析を伴う研究についてはサンプリングデータセットでは十分な分析対象数の確保が困難であり、特別抽出による提供が妥当と考えられた。（分担研究報告書 1）

B) HIC のユーザビリティテストと改善提案

HIC/NDB ガイドライン及び利用者マニュアルの記載に従って、HIC にて通年パネルデータセット及び NDB-β を利用する際のユーザビリティの確認を行った。そして解析環境・二次利用ポータル・ドキュメントにかかる課題及び改善提案を一覧化した。（分担研究報告書 2）

C) 標準化された SQL データハンドリングフレームワークの提示

「解析単位」「時間的アンカー」「観察ウィンドウ」「マスタテーブル整備」を基盤概念とし、研究デザインと SQL 構造を体系的に接続する方法を整理した。具体的には、解析単位テーブル作成、可変アンカー付加、時系列データ抽出、観察ウィンドウ集計、Long-to-Wide 変換という標準ワークフローを構築し、効率的かつ再利用可能なデータ処理を可能とした。この手法により、数 TB 規模の医療データに対しても計算効率を高めつつ、研究品質と教育的価値を向上でき

る。さらに、本フレームワークは複数の医療データベースへ横展開可能であり、将来的な研究基盤標準化や自動化にも資する汎用的手法である。（分担研究報告書 3）

D) 利便性の高いクラウド環境の検討・提案

2024 年 4 月から 2025 年 3 月の医科・DPC・調剤の全テーブル（1 年分、計 620GB）を Redshift から S3 上に Parquet として生成し、MD5 ハッシュの加算突合により 684 テーブルすべてについてデータの一致を確認した。加えて、全テーブルの各カラムについて、NULL 数、ユニーク値数、出現頻度上位 50 件、最大・最小・平均・標準偏差からなるデータサマリを生成した。HIC の甲区分（メモリ 64GB）の環境において、本構成が大規模レセプトデータに対する実用的な処理性能を有することを確認した。ストレージ料金も 1 年分 620GB で月額約 2,100 円と低コストで運用でき、従来の Redshift 依存の分析環境におけるコスト増大・柔軟性の不足といった課題を解消できる可能性を示した。（分担研究報告書 4）

E) 諸外国の医療関連データ利活用・提供状況の調査

韓国では、NHID (Korean National Health Information Database) や HIRA (Health Insurance Review and Assessment Service) といった大規模データベースが整備され、オンサイトセンターを中心とした運用のもと、NHID では年間約 1,000 件のデータ提供の実績がある。フィンランドでは、Findata を通じた一元的な提供体制およびリモート解析環境が整備され、レディメイドデータセットの提供により迅速化とコスト削減が図られている。フランスでは、Health Data Hub を介して、レセプト等のデータベースである SNDS と希少疾患のレジストリである BNDMR の連結が進められている。スウェーデンでは個人番号制度を基盤として、行政・医療・福祉分野のレジストリが多層的に整備

されている。(分担研究報告書5)

F) マスタ情報の提供

本研究の代表・分担研究者が NDB 研究等で用いており、他の研究者等の了解を得た薬価コードの時系列マスタ、新型コロナウイルス感染症(以下、Covid-19)における医療需要検討マスタ、がんにおける医療需要マスタ、糖尿病薬マスタ、病棟区分マスタの5マスタを提供した。前3つは GitHub 上に公開している。

薬価コードの時系列マスタは、社会保険診療報酬支払基金が公開しているマスタを再帰的にダウンロードし、日にちに応じた薬価、診療報酬を与える辞書型の構造として再構成している。Covid-19 における医療需要検討マスタは、日本で Covid-19 の医療資源利用状況を分析した研究で用いたものである。がんにおける医療需要マスタは、日本におけるがん治療のための医療資源利用状況を分析した研究で用いたものである。糖尿病薬マスタは 2026 年 3 月末時点のもので、日本でこれまでに販売された糖尿病関連治療薬の医薬品コード・商品名・一般名・成分・YJ コードである。病棟区分マスタはレセプトで算定できる入院基本料・管理料等の診療行為コードから、医療法上または診療報酬上での区分に該当するかを示したものである(2013年~2025年)。

いずれも研究者が作成したものなので、信頼性を完全には保証できないが、一定のオーソライズは取られている。

D. 考察

本研究はユースケース調査、プリセットデータの改善提案、クラウド環境の操作性への提言、技術供与等を通じて、NDB/HIC のユーザビリティを向上し、医療等情報の二次利用を推進することを目的として実施された。その中で、本報告書では①~⑤を示した。

A) ユースケース調査と汎用データセットの検討については、ユースケース調査の結果から、

現在1年間で提供されている通年パネルデータセットを5年以上に拡張することで、論文発表においては半数程度、厚労科研においても3分の1程度はHICで提供可能なデータ形式で代用可能と考えられた。また、NDBの提供申出書様式1別添8は多数の項目が入っているため、研究者にとって操作が困難である。そこで、一定の条件下で妥当なテンプレート作成を行った。決してこれが全ての研究に適応できるわけではないが、今後の研究者等にとって参考になることを期待したい。また現在〇×で選択することになっているが、チェックボックスの方がユーザビリティは向上すると考えられる。

B) HIC のユーザビリティテストと改善提案については、本分担報告書に記載された内容を、今後の HIC の改善または情報連携基盤での仕様検討に反映されることを期待したい。

C) 標準化された SQL データハンドリングフレームワークの提示に関しては、一定の限界があるものの、研究デザイン・デザインダイアグラム・SQL クエリ構造を一貫した論理構造として接続できることが強みである。将来的な SQL テンプレート自動生成、共通データモデルとの接続、AI 支援型クエリ生成、研究プロトコル自動実装などへの発展可能性を有していると考えられる。

D) 利便性の高いクラウド環境の検討・提案に関しても、従来の Redshift 依存の分析環境におけるコスト増大や柔軟性の不足といった課題を解消し、低コストかつ再現性の高い大規模レセプトデータ分析基盤を実現できたと考えられる。セキュリティと利便性、コスト、瑕疵担保責任の4方向からバランスをとるのは困難であるが、より良い環境での分析に期待したい。

E) 諸外国の医療関連データで活用提供状況の調査に関しては、今回、韓国・フィンランド・フランス・スウェーデンを調査した。特に、韓国とフィンランドに関しては、2024年に調査を実施した際より、データ利活用環境が整備され

ていた。世界的にリアルワールドデータの整備の流れに沿って、レセプト関連データの利活用が進んでいることが実感した。フランスは、Health Data Hub を介してデータ連結を進めるなど、日本と類似の状況である可能性はあるが、活用状況が不明瞭のため引き続き調査を継続する。

F) マスタ情報の提供に関しては、今回のマスタ提供を機に、二次利用ポータル上での研究者等のマスタの共有が進むことを期待する。それとともに、研究班としても、また情報の提供を継続して行っていきたい。

本報告書で触れることなかった、⑥HIC を利用して解析可能なリサーチクエスションやその手法の例示、⑦研究者支援体制への協力についても進めており、次年度の報告書にて記載する。

E. 結論

本研究により、NDB/HIC の利活用拡大にむけてデータセット拡充の必要性を示すとともに、操作性改善案や標準化された SQL 処理、低コストなクラウド分析基盤等を提示した。今後各種調査・分析を進めるとともに、引き続き必要なデータ提供・支援を継続していく予定である。

F. 健康危険情報 なし

G. 研究発表

1. 論文発表

- 1) 松居宏樹. 匿名医療保険等関連情報データベースの利用経験および今後の期待と課題. 統計. 2025 ; 76 (8) : 14-22.

2. 学会発表

- 1) 牧戸 香詠子、明神大也、森由希子 村松圭司 我が国およびフィンランドの医療データベースの最近の動向 第84回日本公衆衛生学会総会 2025年10月30日

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

参考

Fukuyama K, Mori Y, Ueshima H, et al. Medical resource usage for COVID-19 evaluated using the National Database of Health Insurance Claims and Specific Health Checkups of Japan. PLoS One. 2024 May 13;19(5):e0303493.

厚生労働行政推進調査事業費補助金（政策科学総合研究事業（政策科学推進研究事業））

「NDB のユーザビリティ向上を通じて
クラウド上でのデータ二次利用を推進するための研究」
分担研究報告書（令和7年度）

ユースケース調査と汎用データセットの検討

研究代表者 明神 大也（浜松医科大学 健康社会医学講座）

研究分担者 森 由希子（京都大学医学部附属病院）

研究分担者 牧戸 香詠子（東京大学大学院 医学系研究科生物統計情報学）

研究分担者 西岡 祐一（奈良県立医科大学 公衆衛生学講座）

研究協力者 川田 加奈子（浜松医科大学 健康社会医学講座）

研究要旨

目的：本研究は、NDB を用いた既存研究について、利用目的、分析に必要なデータ項目、対象疾患等を整理し、「医療・介護データ等解析基盤（HIC）」で提供可能なデータ形式を用いた研究のユースケースおよび課題を明らかにすることを目的とした。

方法：「NDB データの第三者提供に係る二次利用ポータル」で公開されている「過去の提供申出の成果物」のうち、発表形式が論文または報告書（厚労科研に限る）に分類されている既存研究を対象とした。対象研究について、利用目的、分析に必要なデータ項目、対象疾患等を整理し、HIC で提供可能なデータ形式に基づくユースケースの分類を行った。さらに、分類結果を踏まえ、HIC 利用申請時に必要となる別添 8「申出依頼テンプレート（抽出）」の記載例を作成し、具体的なデータ仕様の検討を行った。

結果と考察：分析対象は論文 47 件、報告書 16 件であった。利用目的別の分類において、論文では処方動向や介入効果（治療）、疾患疫学、地域差分析に関する研究が多く、報告書では地域差分析、医療提供体制調査、処方動向に関する研究が多かった。HIC で提供可能なデータ形式に基づき分類した結果、論文 24 件（51.1%）、報告書 6 件（37.5%）が、通年パネルデータセット、トライアルデータセット、NDB- β のいずれかを用いた研究が実施可能であると考えられた。一方で、希少疾患や詳細な層別解析を伴う研究についてはトライアルデータセットでは十分な分析対象数の確保が困難であり、特別抽出による提供が妥当と考えられた。本研究では、厚生労働省のマニュアルに示された想定ユースケースを参考に分類を行ったが、通年パネルデータセットおよび NDB- β では想定ユースケース以外である「その他」に分類される研究も一定数存在した。同じ利用目的でも、必要となる情報の粒度により活用できるデータ形式が異なることから、既存研究で確認された利用目的に基づき、必要となる情報の粒度と併せてユースケースを整理することが重要である。また、分類結果を踏まえ、処方動向に関する研究を想定した別添 8 記載例として、抽出レコードおよび各項目の要否について「必須」「必要に応じて選択」「原則不要」「不要」の区分を付与しており、申請支援資料の提供も有用と考えられた。

A. 背景と目的

匿名医療保険等関連情報データベース（以下「NDB」という。）は、厚生労働省が法令に基づき収集・提供しているレセプト情報や特定健診・特定保健指導情報などを格納したデータベースである。2020年より民間企業を含む第三者への提供が法制化されて以降、他の医療・介護データ等との連結解析、死亡情報等収載情報の拡大が進んでいる。また、データの迅速提供を図るためオンライン上で利用申請や審査が可能となる二次利用ポータルおよびクラウド上の解析基盤である「医療・介護データ等解析基盤（HIC）」の運用も開始されており、利活用のための提供体制整備が進められている。HICで迅速提供できるデータ形式は、サンプリングされた一年間のデータである「通年パネルデータセット」、サンプリングされた単月毎の断続データである「トライアルデータセット」、機微コード等をマスクしたレセプト全件を含む「NDB-β」の3種類とされており（2026年5月時点）、医療情報等の利活用促進のためには、これらのデータ形式で解析可能な研究課題をユースケースとして提示することが有効と考えられる。

本研究では、医療情報等のさらなる利活用促進を図るため、NDBを用いた既存研究の利用目的、分析に必要なデータ項目、対象疾患等を整理し、HICで迅速提供可能なデータ形式を用いた研究のユースケースや課題を明らかにすることを目的とする。

B. 研究方法

本研究は、NDBデータの第三者提供の申請や問い合わせ、教材の閲覧などを行うことができるシステム「NDBデータの第三者提供に係る二次利用ポータル」で公開されている「過去の提供申出の成果物」（以下「成果物一覧表」という。）のうち、発表形式が「論文」または「報告書」（厚労科研）に分類されている成果物を対象にした。

対象研究は以下のとおり選定した（図1）。

1. 対象研究の選定

① 論文

成果物一覧表で発表形式が「論文」に分類されている成果物を抽出した。なお、同一承諾番号で複数の論文が発表されている場合は代表一件を無作為に抽出した。

成果物一覧表では、特別抽出で提供し、データ容量が算出可能であったもののみデータ容量が表示されているが、HICで提供可能なデータ容量が限られていることから、本研究ではデータ容量が不明なもの（「N/A」または「-」と表記されているもの）を除外した。さらに、抄録から分類の判断ができない論文および抄録にアクセスできない論文は除外した。

② 報告書

成果物一覧表で発表形式が「報告書」に分類されている成果物のうち、発表媒体欄より厚労科研（厚生労働科学研究、厚生労働行政推進調査事業）の報告書である旨明記されている成果物を抽出し、NDBを利用した分担研究報告書を対象とした。なお、同一承諾番号かつ同一研究課題の成果物のうち、成果物一覧に総括研究報告書と分担研究報告書がある場合、一方のみ採用した。さらに、同一承諾番号の成果物のうち、研究期間が複数年度にわたるものは最新年度の報告書を対象とした。

また、データ容量が不明なもの（「N/A」または「-」と表記されているもの）のほか、NDBの利用を確認できなかった報告書は除外した。

2. 対象研究の分類

① 利用目的およびユースケース分類

本研究では、今後どの程度の割合でHICの利用が可能か明らかにするため、過去にNDBを利用した研究について、HICの利用が可能と考えられる研究があるかどうか検討を行った。

始めに、各研究の利用目的を 11 種類（疾患疫学、診断動向、手術動向、処方動向、費用分析、介入効果（政策）、介入効果（治療）、因果関係分析、医療提供体制調査、地域差分析、データの妥当性評価）に分類した。その際、2 種類まで選択可能とした。

次に、HIC で提供できるデータ形式ごとに収録される情報が異なることから、各研究における日・使用量等・医療機関（医療機関コード／病院・診療所フラグ／病床階級コード）・地域（都道府県コード／二次医療圏コード）・機微な情報（所得等）・ID にかかる情報の可否を確認した。併せて、通年パネルデータセットまたはトライアルデータセットによる分析可否確認のため、各研究の分析対象数及び層別解析有無について確認した。

これらの利用目的分類と分析に必要なデータ項目、分析対象数等に基づき、ユースケースの分類を行った。ユースケース分類については、厚生労働省が提供している NDB の利用を検討している方へのマニュアル 令和 7 年 11 月版¹⁾（以下「利用マニュアル」という。）においてデータ形式（通年パネルデータセット、トライアルデータセット、NDB-β）ごとに提示している「想定されるユースケース」を参考にし、2 種類まで選択可能とした。

なお、HIC における各種データ形式で提供できる対象期間が限られていることから、各研究の対象期間との合致有無は分類の条件に含めていない。また、現行の通年パネルデータセットで提供されるレセプトは 1 年間（令和 4 年度分）を対象としているが、同一個人の追跡可能期間を拡大するため、3 年間（令和 4～6 年度）に延長したデータセットを作成する方針である²⁾。このことから、複数年度の追跡を要する研究においても、提供対象期間が延長した場合という条件付きで、通年パネルデータセットの利用が可能と考えられるとの分類にした。

② 疾病分類

各研究の主な対象疾病について、ICD-10（2013 年版）に準拠して分類した。その際、2 種類まで選択可能と研究者が判断した。

3. 汎用性の高いデータセット・具体的なデータ仕様の検討

対象研究の分類結果に基づき、HIC の活用可能性が高いと考えられるユースケースについて、分析に必要なデータ項目および提供方法を検討した。具体的には、HIC 利用申請時にデータ抽出項目の記載・提出を要する様式・別添 8「申出依頼テンプレート（抽出）」の記載例を作成した。作成にあたり、別添 8 と併せて厚生労働省が公開している「申出依頼テンプレートの注意点」および診療報酬情報提供サービスが公開している「オンライン又は光ディスク等による請求に係る記録条件仕様」（令和 6 年 6 月版）を参照し、各抽出レコードおよび項目の可否を検討した。

C. 研究結果

1. 論文の分類結果

成果物 453 件のうち、発表形式が「論文」に分類されているものは計 124 件（全成果物の 27.4%）であった。同一承諾番号（56 件）、データ容量が不明（17 件）、全文にアクセス不可かつ抄録からは分類の判断ができなかった論文（1 件）、全文および抄録にアクセスできない論文（3 件）を除外し、計 47 件（全成果物の 10.4%）が分析対象となった。

利用目的分類（2 種類まで選択可能）については、処方動向 12 件（分析対象 47 件の 25.5%）、介入効果（治療）10 件（21.3%）、疾患疫学および地域差分析 9 件（19.1%）、費用分析および介入効果（政策）5 件（10.6%）、手術動向・診断動向・医療提供体制調査・因果関係分析がいずれも 3 件（6.4%）、データの妥当性評価 2 件（4.3%）となった（図 2）。

これらの利用目的分類、分析に必要なデータ項目、分析対象数等に基づき、ユースケースを検討した結果、15件(31.9%)が通年パネルデータセットを、7件(14.9%)がトライアルデータセットを、6件(12.8%)がNDB-βを活用することが可能と考えられた。このうち4件(8.5%)については、通年パネルデータセットとNDB-βのいずれの活用も可能と考えられることから、最終的に分析対象47件の約5割に値する計24件(51.1%)がHICで提供可能なデータ形式を利用可能と分類された。また、通年パネルデータセットの活用が可能と考えられる研究のうち8件(17.0%)は、当該データセットが複数年度分提供される場合という条件付きで、活用が可能と判断された。なお、残りの23件(48.9%)は、日・使用量等・医療機関(医療機関コード/病院・診療所フラグ/病床階級コード)・地域(都道府県コード/二次医療圏コード)・機微な情報(所得等)・IDが必要であり、通年パネルデータセットまたはトライアルデータセットでは十分な分析対象数が確保できない懸念が生じることなどから特別抽出の利用が妥当と考えられた(表1-1)。

疾病分類(2種類まで選択可能)については、論文の性質上いずれにも該当しなかった11件(分析対象47件の23.4%)に次いで、「第9章 循環器系の疾患(I00-I99)」が9件(19.1%)、「第4章 内分泌、栄養及び代謝疾患(E00-E90)」が5件(11.4%)となった(図3)。第9章分類のうち「I60-I69:脳血管疾患」に該当するのは5件、次いで「I20-I25:虚血性心疾患」の4件、「I30-I52:その他の型の心疾患」の3件、「I70-I79:動脈、細動脈及び毛細血管の疾患」の1件となった。なお、うち1件は「I60-I69:脳血管疾患」と「I20-I25:虚血性心疾患」の両方に該当しており、うち1件は「I20-I25:虚血性心疾患」、「I30-I52:その他の型の心疾患」、「I60-I69:脳血管

疾患」、「I70-I79:動脈、細動脈及び毛細血管の疾患」に該当していた。また、第4章分類5件のうち、「E10-E14:糖尿病」が2件、「E70-E90:代謝障害」が2件(うち1件は併せて循環器系の疾患も対象)、「E65-E68:肥満(症)及びその他の過栄養<過剰摂食>」が1件となった(図3)。

2. 報告書の分類結果

成果物453件のうち、発表形式が「報告書」に分類されている成果物は80件(全成果物の17.7%)であり、うち発表媒体欄において厚労科研(厚生労働科学研究、厚生労働行政推進調査事業)の報告書である旨明記されている成果物は計42件(9.3%)であった。同一承諾番号(13件)、NDBを利用していない報告書(1件)、データ容量が不明(12件)な報告書を除外し、計16件(全成果物の3.5%)が分析対象となった。

利用目的分類については、地域差分析9件(分析対象16件の56.3%)、医療提供体制調査8件(50.0%)、処方動向6件(37.5%)、疾患疫学4件(25.0%)、データの妥当性評価2件(12.5%)、診断動向および因果関係分析がそれぞれ1件(6.3%)となった(図2)。これらの利用目的分類、分析に必要なデータ項目、分析対象数等に基づき、ユースケース(2種類まで選択可能)を検討した結果、5件(分析対象16件の31.3%)が通年パネルデータセットを、1件(6.3%)がトライアルデータセットを活用することが可能と考えられ、分析対象16件の約4割に値する計6件(37.5%)がHICで提供可能なデータ形式を利用可能であることが示唆された(表1-2)。また、通年パネルデータセットの活用が可能と考えられる研究のうち4件(25.0%)は、当該データセットが複数年度分提供される場合という条件付きで、活用が可能と判断された。なお、残りの10件(62.5%)は、日・使用量等・医療機関(医療

機関コード／病院・診療所フラグ／病床階級コード）・地域（都道府県コード／二次医療圏コード）・機微な情報（所得等）・ID）が必要であり、トライアルデータセットでは十分な分析対象数が確保できない懸念が生じることなどから特別抽出の利用が妥当と考えられた。

疾病分類については、報告書の性質上いずれにも該当しなかった5件（分析対象16件の31.3%）に次いで、「第9章 循環器系の疾患（I00-I99）」が3件（18.8%）、「第04章 内分泌、栄養及び代謝疾患（E00-E90）」が2件（12.5%）となった（図3）。以上のように、分析対象となった論文47件のうち24件（51.1%）、報告書16件のうち6件（37.5%）が、通年パネルデータセット、トライアルデータセット、NDB- β のいずれかを用いた研究が実施可能であると考えられた。一方で、論文の23件（48.9%）、報告書の10件（62.5%）は、HICで提供可能なデータ形式では再現が困難であり、特別抽出による提供が妥当と考えられた（表2）。

3. 利用目的分類別にみた想定データ形式

論文および報告書を合わせた分析対象63件について、利用目的分類別に想定データ形式を確認した（表3）。なお、利用目的分類および想定データ形式はいずれも1研究につき2種類まで選択可能としたため、表3の件数は延べ件数である。

利用目的分類別にみると、地域差分析が19件と最も多く、このうち12件が特別抽出、7件が通年パネルデータセットに分類された。処方動向は18件であり、特別抽出9件、通年パネルデータセット6件、トライアルデータセット3件に分類された。疾患疫学は13件であり、特別抽出6件、NDB- β 3件、通年パネルデータセット2件、トライアルデータセット2件に分類された。医療提供体制調査は11件であり、このうち8件が特別抽出、3件が通年パ

ネルデータセットに分類された。

4. 汎用性の高いデータセット・具体的なデータ仕様の検討結果

論文における対象研究を分類した結果、利用目的として処方動向に関する研究が12件と最も多かったため、処方動向に関する研究を想定した別添8「申出依頼テンプレート（抽出）」の記載例を作成した（添付資料1）。また、別添8作成プロセスの簡素化を図るため、参照資料に基づき各抽出項目の説明（概要）を整理のうえ記載するとともに、項目要否の判断を補助するため「必須」「必要に応じて選択」「原則不要」「不要」の区分を付与した。

D. 考察

本研究では、NDBを用いた既存研究の利用目的、対象疾患、分析に必要なデータ項目等を整理し、HICで提供可能なデータ形式を用いた研究の実施可能性について検討した。その結果、分析対象となった論文のうち約5割、報告書のうち約4割が、通年パネルデータセット、トライアルデータセット、NDB- β のいずれかを用いた研究が実施可能であると考えられた。特に、罹患者数の多い疾患を対象とした記述疫学研究や処方動向の把握などは、通年パネルデータセットやトライアルデータセットの活用可能性があると考えられた。また、慢性疾患の長期追跡や医療費の全国的推移の把握など、一部の研究ではNDB- β の活用可能性が示唆された。

一方で、HICで提供可能なデータ形式では再現が困難と考えられる研究も一定数確認された。特に、希少疾患や詳細な層別解析を伴う研究については、通年パネルデータセットやトライアルデータセットなどのサンプリングされたデータセットでは十分な分析対象数を確保できない懸念がある。このような研究については、引き続き特別抽出による提供が必要と考えられた。

また、医療機関（医療機関コード）・地域（都

道府県コード／二次医療圏コード)にかかるとの情報、HICで提供可能なデータ形式のうち通年パネルデータセットでのみ提供されている。そのため、これらの情報を必要とする医療提供体制調査や地域差分析では、活用可能なデータ形式が通年パネルデータセットに限定されやすいと考えられた。さらに、HICで提供可能なデータ形式の活用が可能と分類された研究のうち、2～3割は通年パネルデータセットが複数年度分提供される場合という条件付きでの判断であった。このことから、当該データセットの提供対象期間が拡大されることにより、HICの活用可能性がさらに高まることが示唆された。

利用目的分類については、論文では処方動向や介入効果(治療)、疾患疫学、地域差分析に関する研究が多かった。一方、報告書では医療提供体制調査や処方動向、地域差分析に関する研究が多く、論文と報告書で利用目的に違いがみられた。これは、論文では特定の疾患、治療、薬剤等に注目した分析が多い一方、厚労科研等の報告書では政策的課題への対応や医療提供体制の把握を目的とした分析が多いことを反映していると思料される。なお、報告書については、論文と比較してHICで提供可能なデータ形式の活用が可能と分類された割合が低く、特別抽出が妥当と考えられる研究の割合が高かったが、これは医療機関(医療機関コード等)・地域(道府県コード／二次医療圏コード)にかかるとの情報が必要とする医療提供体制調査や地域差分析が多いことが一因と考えられる。このことから、政策や制度の評価・改善等に資する分析では、地域単位・医療機関単位の情報や、複数年度にわたる追跡情報が求められる場合が多く、HICで提供可能なデータ形式の活用には一定の制約があると考えられた。

本研究では、厚生労働省のマニュアルに示された想定ユースケースを参考に分類を行ったが、通年パネルデータセットおよびNDB-βでは「その他」に分類される研究も一定数存在した。

例えば通年パネルデータセットの「その他」には、処方動向や地域差分析(処方動向の患者追跡・時系列追跡や地域差)、薬剤等の安全性評価(患者の長期追跡を伴う)、医療提供体制調査(医療資源の適正配置や地域差)などが含まれていた。このことから、HICの利活用促進にあたっては、利用マニュアル上の代表的ユースケースに加え、既存研究で確認された利用目的に基づいてユースケースを整理し、必要となる情報の粒度と併せて提示することが有用と考えられる。

特に処方動向に関する研究では、処方数量、処方日数、処方日、分析目的によって必要となる情報の粒度は異なる。例えば、全国的な処方動向の把握や一定規模以上の対象集団における処方状況の概略把握であれば、トライアルデータセットや通年パネルデータセットを活用できる可能性がある。一方で、詳細な量反応評価、長期継続処方の把握、希少疾患を対象とする分析では、より詳細な抽出条件や十分な対象数の確保が必要となり、特別抽出が妥当となる場合がある。このように、同じ処方動向であっても、分析目的と必要となる情報に応じて適切なデータ形式を選択する必要がある。

また、本研究では分類結果を踏まえ、処方動向に関する研究を想定した別添8「申出依頼テンプレート(抽出)」の記載例を作成した。抽出レコードおよび各項目の要否を整理し、「必須」「必要に応じて選択」「基本的に不要」の区分を付与することは、申請者が分析目的に応じて必要なデータ項目を検討する際の負担軽減につながる可能性がある。今後は、処方動向以外の主要なユースケースについても同様の記載例やデータ仕様の整理を進めることで、HICの円滑な利用に資することが期待される。

本研究にはいくつかの限界がある。第一に、本研究は二次利用ポータルで公開されている成果物一覧表に基づく既存研究を対象としており、公開されていない成果物や、分類に必要な情報

が十分に得られない研究は対象外となっている。第二に、HIC 利用可能性の判断は、公開情報から把握可能な研究目的、対象疾患、分析内容、必要データ項目等に基づくものであり、実際の申請時に必要となる詳細な抽出条件やデータ仕様を完全に再現したものではない。第三に、HIC で提供可能なデータ形式や収載項目は今後変更される可能性があり、本研究の分類結果は主に検討時点における提供方針を前提としたものである。したがって、今後の HIC の仕様変更や提供対象期間の拡大に応じて、ユースケース分類や利用可能性の判断を更新していく必要がある。

E. 結論

HIC で提供可能なデータ形式は、既存 NDB 研究の一定割合に活用可能であり、特に罹患者数の多い疾患を対象とした記述疫学研究等において有用性が期待される。一方で、地域差分析や医療提供体制調査、長期追跡、希少疾患研究等では、現在 HIC で提供可能なデータ形式では研究が困難な場合がある。HIC の利活用促進にあたっては、特別抽出との役割分担を含め、既存研究で確認された利用目的に基づいてユースケースを整理し、必要となる情報の粒度と併せて提示することが重要である。

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

参考：

1) 厚生労働省「NDB の利用を検討している方へのマニュアル 令和 7 年 11 月版」

https://www.mhlw.go.jp/stf/seisakunit-suite/bunya/kenkou_iryuu/iryuuhoken/rese-puto/index.html

(令和 8 年 5 月 14 日アクセス)

2) 厚生労働省「第 26 回 匿名医療情報等の提供に関する専門委員会」(令和 7 年 3 月 5 日開催)資料 1 「通年パネルデータセットの今後の作成方針について(案)」

<https://www.mhlw.go.jp/content/12400000/001430216.pdf>

(令和 8 年 3 月 28 日アクセス)

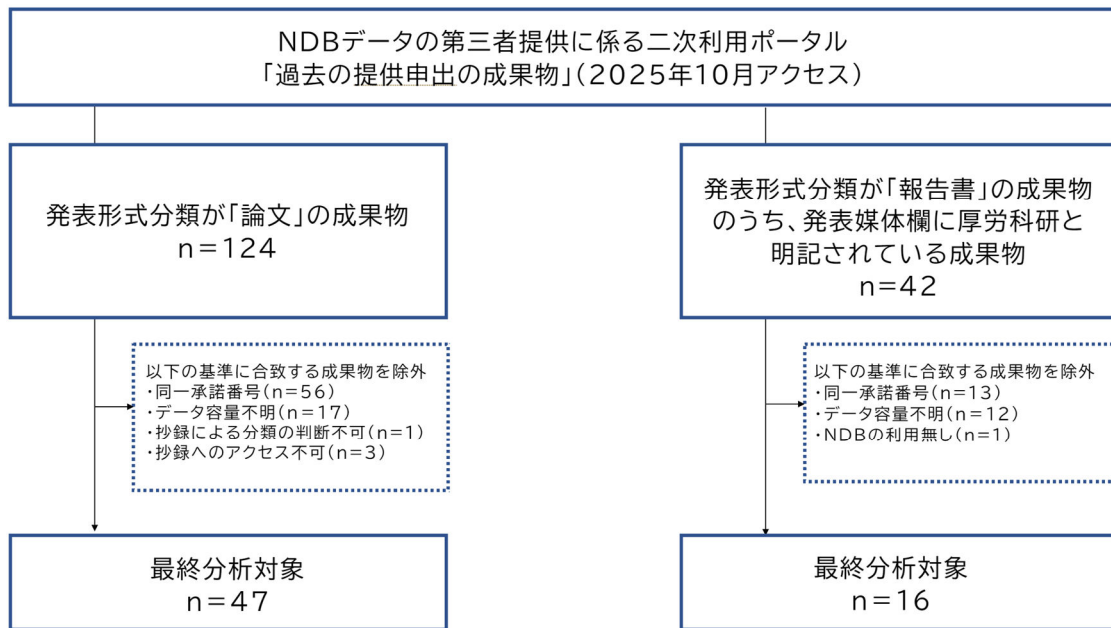


図1 対象研究の選定フロー

表 1-1 想定されるユースケース分類（論文）

データ形式	サブカテゴリ	小計 (件)	合計 (件)	割合 (%)
通年パネルデータセット	(罹患者数の多い疾患の記述疫学研究)	4	15	31.9
	(急性疾患の短期追跡 (縦断的分析))	2		
	(手術件数の都道府県別推移の調査)	0		
	(薬剤効果の用量反応評価)	0		
	(その他)	9		
トライアルデータセット	(罹患者数の多い疾患の概略調査)	1	7	14.9
	(処方動向のサンプリング調査)	2		
	(疾患の横断的把握)	2		
	(検査の全国的推移の調査)	1		
	(その他)	1		
NDB-β	(慢性疾患の長期追跡 (縦断的分析))	1	6	12.8
	(薬剤等の安全性評価)	0		
	(希少疾患の記述疫学研究)	0		
	(医療費の全国的推移の調査)	1		
	(その他)	4		
特別抽出	-	23	23	48.9

※想定されるユースケース分類は1研究につき2種類まで選択可能としたため、件数は延べ件数。

表 1-2 想定されるユースケース分類（報告書）

データ形式	サブカテゴリ	小計 (件)	合計 (件)	割合 (%)
通年パネルデータセット	(罹患者数の多い疾患の記述疫学研究)	1	5	31.3
	(急性疾患の短期追跡（縦断的分析）)	0		
	(手術件数の都道府県別推移の調査)	0		
	(薬剤効果の用量反応評価)	0		
	(その他)	4		
トライアルデータセット	(罹患者数の多い疾患の概略調査)	0	1	6.3
	(処方動向のサンプリング調査)	1		
	(疾患の横断的把握)	0		
	(検査の全国的推移の調査)	0		
	(その他)	0		
NDB-β	(慢性疾患の長期追跡（縦断的分析）)	0	0	0.0
	(薬剤等の安全性評価)	0		
	(希少疾患の記述疫学研究)	0		
	(医療費の全国的推移の調査)	0		
	(その他)	0		
特別抽出	-	10	10	62.5

※想定されるユースケース分類は 1 研究につき 2 種類まで選択可能としたため、件数は延べ件数。

利用目的分類の該当件数（論文・報告書別）

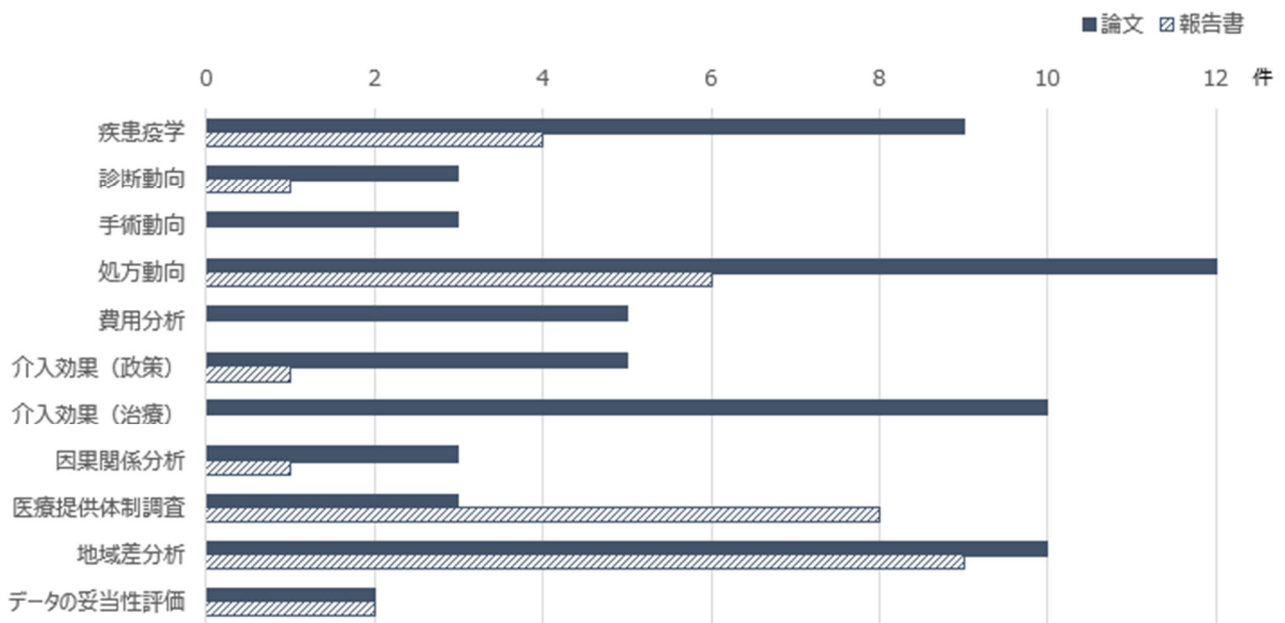


図 2 利用目的分類の該当件数（論文・報告書別）

※利用目的分類は 1 研究につき 2 種類まで選択可能としたため、件数は延べ件数。

ICD-10大分類の該当件数（論文・報告書別）

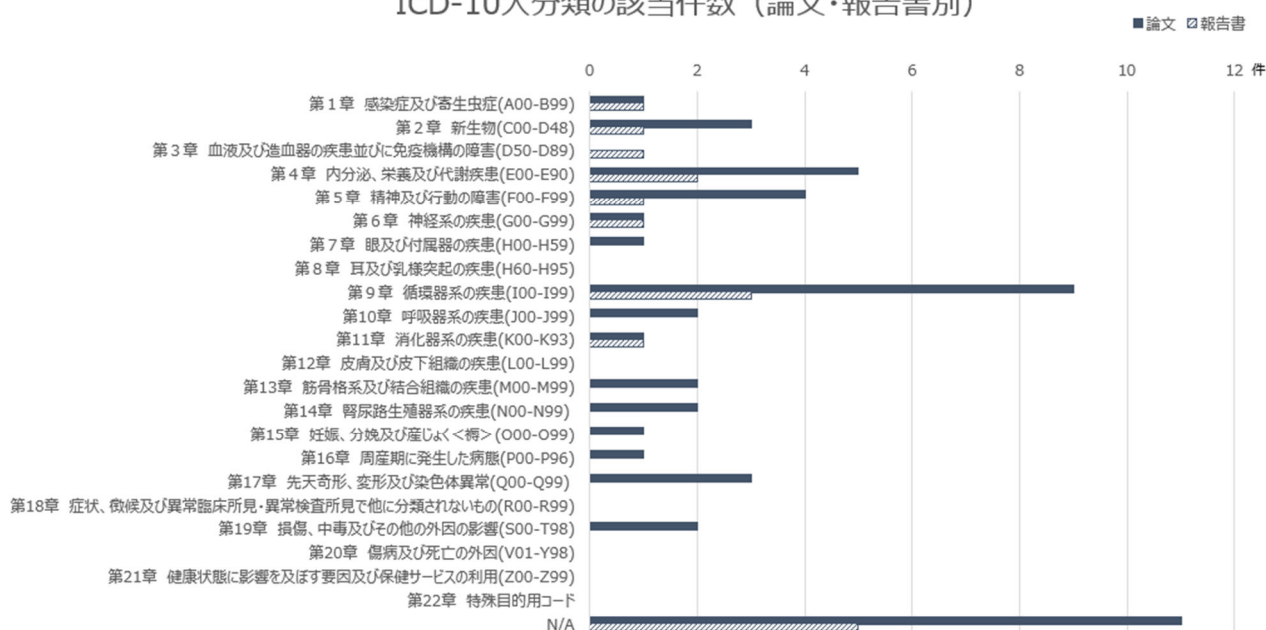


図3 ICD-10大分類の該当件数（論文・報告書別）

※疾病分類は1研究につき2種類まで選択可能としたため、件数は延べ件数。

表2 HIC利用可能性の分類結果（報告書）

対象	分析対象 件数	HIC利用可能件数（割合）	うち通年パネルデータセットの 複数年度提供を要する件数 （割合）	特別抽出 （割合）
論文	47	24 (51.1%)	8 (17.0%)	23 (48.9%)
報告書	16	6 (37.5%)	4 (25.0%)	10 (62.5%)
計	63	30 (47.6%)	12 (19.0%)	33 (52.4%)

表3 利用目的分類別の想定されるユースケース分類

利用目的分類	通年パネルデータセット	トライアルデータセット	NDB-β	特別抽出	計
疾患疫学	2	2	3	6	13
診断動向	2	2	0	0	4
手術動向	0	0	0	3	3
処方動向	6	3	0	9	18
費用分析	2	1	1	1	5
介入効果（政策）	2	0	1	3	6
介入効果（治療）	5	0	0	5	10
因果関係分析	1	1	0	2	4
医療提供体制調査	3	0	0	8	11
地域差分析	7	0	0	12	19
データの妥当性評価	1	0	1	2	4
計	31	9	6	51	97

※利用目的および想定されるユースケース分類は1研究につき2種類まで選択可能としたため、件数は延べ件数。

厚生労働行政推進調査事業費補助金（政策科学総合研究事業（政策科学推進研究事業））
「NDB のユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究」
分担研究報告書（令和7年度）

HIC のユーザビリティテストと改善提案

研究代表者 明神 大也（浜松医科大学 健康社会医学講座）

研究分担者 村松 圭司（千葉大学医学部附属病院）

研究要旨

【目的】本研究は、厚生労働省の NDB（匿名医療保険等関連情報データベース）の利用に関するホームページに掲載されている「匿名医療保険等関連情報データベース（NDB）の利用に関するガイドライン」・「医療・介護データ等解析基盤（HIC）の利用に関するガイドライン」（以下「HIC/NDB ガイドライン」という。）、及び二次利用ポータルに掲載されている利用者マニュアルの記載に従って HIC のユーザビリティを確認し、HIC の利用上の課題を発見することを目的とする。

【方法と結果】通年パネルデータセット及び NDB-β を申請し、HIC での NDB データの利用を開始した。その過程で気づいた課題や改善提案を HIC 解析環境・二次利用ポータル・関連ドキュメントに分けて記載した。

【考察】今後は、複数の研究課題を立て、HIC の探索的利用環境において現行配置されているデータセットを扱い解析し、これらのデータセットの研究可能性を調査する。

A. 研究目的

【背景】

医療DXの推進に関する工程表（令和5年6月2日医療DX推進本部決定）、また規制改革実施計画（令和5年6月16日閣議決定）

「NDB や公的統計データの利活用の円滑化・迅速化」に従って、HIC を活用して、NDB データを迅速に提供する取組が進められてきた。しかし、プリセットデータがクラウド上にも関わらず CSV ファイルで提供される、解析環境内ではコマンドラインの操作が必要である等、これまでのオンプレミス環境における解析環境と比較しても、ユーザビリティに関して多くの課題が残っている。

こうした中で、規制改革推進に関する中間答申（令和6年12月25日）においては、公的 DB の仮名化情報について安全であるのみならず迅速かつ円滑に利用・解析を行うことができるクラウド環境について、HIC との関

係を整理しつつ実現することとされている。

【目的】

本研究は、「匿名医療保険等関連情報データベース（NDB）の利用に関するガイドライン」及び「医療・介護データ等解析基盤（HIC）の利用に関するガイドライン」（以下「HIC/NDB ガイドライン」という。）、及び二次利用ポータルに掲載されている利用者マニュアルの記載に従って HIC のユーザビリティを確認し、HIC の利用上の課題を発見することを目的とする。

B. 研究方法

通年パネルデータセット及び NDB-β を申請し、第 34 回匿名医療情報等の提供に関する専門委員会にて通年パネルデータセット及び NDB-β の HIC 上での利用承諾（いわゆる迅速提供）を得た。その後、Windows 環境で利用開始した。本研究ではその過程で気づいた課題や改

善提案を、HIC 解析環境と二次利用ポータル、関連ドキュメントの3つに分けてまとめた。

C. 研究結果

HIC/NDB ガイドライン及び利用者マニュアルの記載に従って、HIC にて通年パネルデータセット及びNDB-βを利用する際のユーザビリティの確認を行った。課題及び改善提案は、HIC 解析環境・二次利用ポータル・関連ドキュメントに分け、下記のとおり整理した。なお、課題及び改善提案は一覧化し、「別添_課題及び改善事項一覧」（以下、別紙）に整理した。

1. HIC 解析環境に関する課題及び改善提案

【課題 1-①】

HIC を利用した研究では、複数のユーザーが同一のサーバーで解析する必要がある一方で、仮想サーバーや Docker の利用が想定されておらず、ユーザー間の調整コストが高いほか、情報連携不足による誤操作のリスクが存在する。

また、Windows OS では Hyper-V が利用不可であり、利用者全員が同じ OS を利用する必要がある。

【改善提案 1-①】

仮想サーバーの構築や Docker の活用を可能とすることで、複数のユーザーが同一のサーバーで解析する際の調整コストやリスクが低減できるほか、利用者ごとに習熟した OS が利用可能となるなど、研究全体の生産性が向上することが想定される。

なお、上記の設計が難しい場合も、その旨をあらかじめ周知しておくことにより、利用者が当該前提のもと運用方針を検討できる。

【課題 1-②】

複数の利用者が同一のサーバーを利用する場合は、ある利用者の操作に起因したトラブルが利用者全体に波及し、データの取得や前処理、分析をやり直す必要性が生

じるなど、研究が停滞する可能性がある。

【改善提案 1-②】

ある利用者の操作に起因するトラブルが他の利用者に波及しうること、及びその防止策（案）を利用者マニュアルにより周知することで、トラブルの発生を未然に防げる可能性がある。

また、利用者による環境の巻き戻しを可能とすることにより、トラブル発生時の研究への影響を最小限にとどめられると想定される。

なお、利用者マニュアルの事前準備編（2026/2/27 に取得）には、最新のバックアップイメージでのリストア（再構築）作業は「通常運用中に発生した、解析基盤の動作不具合時のみ対応」することとされており、また、「利用者様の操作に起因する、トラブル発生時の環境の巻き戻しを目的としたリストア依頼などにつきましては、対応致しかねます」と記載されている。

【課題 1-③】

HIC を利用した端末においては、利用端末に機微な情報が保存されない一方で、特別抽出等の利用形態と同様に、申請した利用場所以外での利用が不可能である。

【改善提案 1-③】

クラウド環境での分析であることを活かし、例えば HIC 環境での利用に限り、窃視対策を条件に利用場所の制限を緩和することにより、利用者がより柔軟に研究を推進可能となることが想定される。

【課題 1-④】

解析環境上にコミュニケーションツールが存在せず、複数の組織での共同研究を推進する際にコミュニケーションコストが発生する。

【改善提案 1-④】

解析環境にコミュニケーションツールを導入することで、解析環境内で、解析環境上の情報を基に相談でき、コミュニケー

ションコストが低減すると想定される。

2. 二次利用ポータルに関する課題及び改善事項

【課題 2-①】

HIC を利用して NDB の研究を行う場合、複数の研究者の関与が想定される一方で、二次利用ポータルの公表前確認の申請や、最新の申請書の確認は申請の担当者のみが実施可能であるため、担当者を除く利用者は担当者を通して依頼する必要があり、コミュニケーションコストが発生する。

【改善事項 2-①】

二次利用ポータルにおいて、担当者のアカウントに加え、利用者のアカウントを研究と紐づけて管理し、利用者による公表前確認や最新の申請書の確認を実施可能とすることで、コミュニケーションコストを削減できると想定される。

また、二次利用ポータルログイン時にログインパスワードを誤った場合、「ユーザ認証に失敗しました」の画面ではなく「セッションが既に終了しています」の画面となったため、cookieなどを削除して等で何度かログイン試していると最終的にアカウントロックになった。パスワードを間違っても必ずしもセッション終了の画面が出るわけではなく再現ができない（規則性の判断がつかない）ため、正式な仕様漏れやバグとも言えないが本報告書には記載した。

3. 関連ドキュメントに関する課題及び改善事項

【課題 3-①】

利用者マニュアルでは、Amazon Redshiftをはじめとする専門的な知識が必要とされる箇所の情報量が相対的に少ないため、特にAWS等の習熟度が低い研究者において、分析に用いるデータの前処理に時間を要すことや、意図しない操作によりトラブル

が発生する可能性がある。

【改善事項 3-①】

利用者マニュアルに記載されている操作の前提条件の記載を充実化させることや、Research Questionを基に、Amazon Redshiftからデータを抽出し、解析結果をS3に格納するまでの一連の流れをユースケースとして記載することにより、NDB-βを用いて分析する研究者の障壁を削減できる可能性がある。なお、前者の具体的な改善案は別紙に記載している。

【課題 3-②】

HIC/NDB ガイドライン及び利用者マニュアルに関して、ドキュメント間での記載の不統一や、URLのリンク切れ等が存在する。

【改善提案 3-②】

ドキュメントを更新する。なお、具体的な改善案は別紙に記載している。

D. 考察

今年度はHIC解析環境、二次利用ポータル、関連ドキュメントの3つの観点で課題及び改善事項の検討を行った。HIC解析環境や二次利用ポータルに関連する課題には、複数人が利用者となり、研究を推進するという特性から生まれる課題が多く見つかった。

複数人が利用者となることを前提に、HIC解析環境や二次利用ポータル、ドキュメントの改善を行うことにより、利用者のユーザビリティが向上すると考えられる。また、ドキュメントに関しては、HIC上での操作に想定以上の時間を要することや、意図しないトラブルが発生することを避けるため、HICを操作する際に必要な情報を充実化することが重要であると考えられる。

今後は、複数の研究課題を立て、HICの探索的利用環境において現行配置されているデータセット(NDB-β)を扱い解析し、NDB-βの研究可能性を調査する。

E. 結論

本研究では、HIC/NDB ガイドライン及び利用者マニュアルの記載に従って、HIC のユーザビリティを確認し、HIC 解析環境・二次利用ポータル・関連ドキュメントに改善余地のある課題を発見した。これらの課題の改善によりユーザビリティ上の向上が期待される。今後は、複数の研究課題を立て、HIC の探索的利用環境において現行配置されているデータセット（通年パネルデータセット及びNDB-β）を扱い解析し、これらの研究可能性を調査する。

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

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

(参考)

厚生労働省 HP 【NDB】匿名医療保険等関連情報データベースの利用に関するホームページ
https://www.mhlw.go.jp/stf/seisakunit-suite/bunya/kenkou_iryuu/iryuhoken/rese-puto/index.html, 2026/2/27 アクセス
二次利用ポータル
<https://www.hic.mhlw.go.jp/imart/top>, 2026/2/27 アクセス

No	カテゴリー	課題	改善に向けた対応方針	改善事項	本紙上のNo.
1	解析環境	医療・介護データ等解析基盤（Healthcare Intelligence Cloud、以下HICという。）を利用した研究では、複数のユーザーが同一のサーバーで解析する必要がある一方で、仮想サーバーやDockerの利用が想定されておらず、ユーザー間の調整コストが高いほか、情報連携不足による誤操作のリスクが存在する。また、Windows OSではHyper-Vが利用不可であり、利用者全員が同じOSを利用する必要がある。 なお、解析環境へのミドルウェアのインストールを試みる場合、二次利用ポータルへのアップロード機能（ファイル容量は500MBまで）を活用するか、保守事業者へ依頼する必要がある。	・システム改修 ・利用者への周知	仮想サーバーの構築やDockerの活用を可能とすることで、複数のユーザーが同一のサーバーで解析する際の調整コストやリスクが低減できるほか、研究者ごとに習熟したOSが利用可能となるなど、研究全体の生産性が向上することが想定される。 なお、上記の設計が難しい場合も、その旨をあらかじめ周知しておくことにより、利用者が当該前提のもと運用方針を検討できる。	1-①
2	解析環境	複数の利用者が同一のサーバーを利用する場合は、ある利用者の操作に起因したトラブルが利用者全体に波及し、データの取得や前処理、分析をやり直す必要性が生じるなど、研究が停滞する可能性がある。	・利用者への周知 ・制限緩和	ある利用者の操作に起因するトラブルが他の利用者に波及しうること、及びその防止策（案）を利用者マニュアル（二次利用ポータルに掲載されている。）により周知することで、トラブルの発生を未然に防げる可能性がある。 また、利用者による環境の巻き戻しを可能とすることにより、トラブル発生時の研究への影響を最小限にとどめられると想定される。 なお、利用者マニュアルの「事前準備編」（2026/2/27に取得）には、最新のバックアップイメージでのリストア（再構築）作業は「通常運用中に発生した、解析基盤の動作不具合時のみ対応」することとされており、また、「利用者様の操作に起因する、トラブル発生時の環境の巻き戻しを目的としたリストア依頼などにつきましては、対応致しかねます」と記載されている。	1-②
3	解析環境	HICを利用した端末においては、利用端末に機微な情報が保存されない一方で、特別抽出等の利用形態と同様に、申請した利用場所以外での利用が不可能である。	・制限緩和	クラウド環境での分析であることを活かし、例えばHIC環境での利用に限り、窃視対策を条件に利用場所の制限を緩和することにより、利用者がより柔軟に研究を推進可能となることが想定される。	1-③
4	解析環境	解析環境上にコミュニケーションツールが存在せず、複数の組織での共同研究を推進する際にコミュニケーションコストが発生する。	・ツールの導入	解析環境にコミュニケーションツールを導入することで、解析環境内で、解析環境上の情報を基に相談でき、コミュニケーションコストが低減すると想定される。	1-④
5	2次利用ポータル	HICを利用してNDBの研究を行う場合、複数の研究者の関与が想定される一方で、二次利用ポータルの公表前確認の申請や、最新の申請書の確認は申請の担当者のみが実施可能であるため、担当者を除く利用者は担当者を通して依頼する必要があり、コミュニケーションコストが発生する。	・システム改修	二次利用ポータルにおいて、担当者のアカウントに加え、利用者のアカウントを研究と紐づけて管理し、利用者による公表前確認や最新の申請書の確認を実施可能とすることで、コミュニケーションコストを削減できると想定される。	2-①

No	カテゴリー	課題	改善に向けた対応方針	改善事項	本紙上のNo.
6	ドキュメント	利用者マニュアルでは、Amazon Redshiftを始めとする、専門的な知識が必要とされる箇所の情報量が相対的に少ないため、特にAWS等の習熟度が低い研究者において、分析に用いるデータの前処理に時間を要すことや、意図しない操作によりトラブルが発生する可能性がある。	・ドキュメントの見直し	利用者マニュアルに記載されている操作の前提条件の記載を充実化させることや、Research Questionを基に、Redshiftからデータを抽出し、解析結果をS3に格納するまでの一連の流れをユースケースとして記載することにより、NDB-βを用いて分析する研究者の障壁を削減できる可能性がある。 なお、前者の具体的な改善案は下記の通り。 ・利用者マニュアルの「事前準備編」から「トライアルデータセット・通年パネルデータセット操作編」に記載されている内容を一覧として整理し、特定の操作を実施する際にどのマニュアルを参照すればよいか分かるようにする。 ・利用者マニュアルの「事前準備編」の「2.1.3.ログイン（解析サーバ（EC2））」を実施するに当たり、事前にEC2を起動する必要がある旨、及びEC2の起動方法は利用者マニュアルの「基本操作編」に記載されていることを記載する。 ・Redshiftでのスキャン実施等、高額な利用料が発生しうるケース及びその対策の記載を充実させる。 ・NDB利用者マニュアルの別紙「別紙_コマンド一覧.xlsx」の情報をHICのPowerShellにコピーする方法を記載する。 ・通年パネルデータセットのダウンロードの考え方が利用者マニュアルごとに異なるため、どのような時にどの考え方でダウンロードするのが望ましいのかを記載する（利用者マニュアルの「S3～EC2操作編」を基に、「別紙_コマンド一覧」のコマンドNo.10を実行すれば、指定した年度の通年パネルデータセットがダウンロードされる一方で、利用者マニュアルの「トライアルデータセット・通年パネルデータセット操作編」のp.10やp.11に記載されている操作を行うと、S3上に保管されている通年パネルデータセットのうち、一部のddl及びcsvがダウンロードされる）。	3-①
7	ドキュメント	「匿名医療保険等関連情報データベース（NDB）の利用に関するガイドライン」・「医療・介護データ等解析基盤（HIC）の利用に関するガイドライン」、及び利用者マニュアルに関して、ドキュメント間での記載の不統一や、URLのリンク切れ等が存在する。	・ドキュメントの見直し	ドキュメントを更新する。 ・「解析環境利用に係る通知書」のQRコードを手打ちせずとも表示できるようにする（表示するURLを押下すると404エラーが表示され、URLを全て手打ちする必要があるため）。 ・一般的なWindows製品において二要素認証を実現するための具体的な対策例をHIC/NDBガイドラインに記載する（HIC利用端末には二要素認証が求められるが、一般的なWindows製品に搭載されている認証方法（Windows Hello）だと顔認証した時点でログインできてしまい、二要素認証とならないため）。 ・利用者マニュアルに記載されているURLのリンク切れを解消する（「S3～EC2操作編」の1.3章に記載の「匿名医療保険等関連情報データベース（NDB）の利用に関するガイドライン」のリンクが切れているため）。 ・利用者マニュアルの別紙「別紙_コマンド一覧.xlsx」に記載されているコマンドを直接ペーストできるようにする（セキュリティ対策によりマクロ付Excelを扱わない方針としている可能性はあるものの、「別紙_コマンド一覧.xlsx」に記載されているコマンドは関数により作成されており、直接ペーストできないため）。 ・利用者マニュアルの「トライアルデータセット・通年パネルデータセット操作編」と「別紙_論理テーブル仕様書_通年パネルデータセット（NDB）（2022年度）」の「v1」版と「v1なし」版との違いを記載する。 ・ドキュメントの更新時は、関連するドキュメントを横並びで更新する（記載が統一されていないことにより、誤った理解を促す可能性があるため。例えば、2025/11/1に第3.1版に更新された「匿名医療保険等関連情報データベース（NDB）の利用に関するガイドライン」では、「なお、オンサイト環境におけるリモート用全量 NDB、若しくは HIC におけるリモート用全量 NDB 又は NDB-β を利用する場合、利用開始時点から遡って最大で 10 年分のデータが参照可能であるため、この範囲内の期間を抽出対象期間として記載すること。」と記載されている一方で、利用者マニュアルのRedshift操作編では、「Redshiftによるレポート情報取得時の性能劣化防止の観点から利用可能範囲を10年分と規定しております。ただし、ご利用期間中にRedshiftにてデータの再抽出を行うことを想定し、利用開始月から起算して10年(120か月)前までの診療月のデータは、常に参照可能となるよう設定しております。※なお、仕様上最大10年分のレポート情報が利用可能ですが、申請いただいた抽出期間を超えてデータ参照を行うことはできませんのでご注意ください。※やむを得ない理由により利用期間延長の変更申請を行われた場合は、この限りではありません。」と記載されている）。	3-②

令和7年度 厚生労働行政推進調査事業費補助金(政策科学総合研究事業
(政策科学推進研究事業))

「NDB のユーザビリティ向上を通じて
クラウド上でのデータ二次利用を推進するための研究」
分担研究報告書

標準化された SQL データハンドリングフレームワークの提示

研究分担者 松居宏樹 東京大学大学院医学系研究科准教授

研究要旨

本報告では、NDB、DPC、JMDC、DeSC などの大規模時系列医療データベースを疫学研究へ活用するため、標準化された SQL データハンドリングフレームワークを提示した。従来、SQL クエリ設計は研究者ごとの経験に依存し、再現性や保守性、レビュー容易性に課題があった。本研究では、「解析単位」「時間的アンカー」「観察ウィンドウ」「マスタテーブル整備」を基盤概念とし、研究デザインと SQL 構造を体系的に接続する方法を整理した。具体的には、解析単位テーブル作成、可変アンカー付加、時系列データ抽出、観察ウィンドウ集計、Long-to-Wide 変換という標準ワークフローを構築し、効率的かつ再利用可能なデータ処理を可能とした。この手法により、数 TB 規模の医療データに対しても計算効率を高めつつ、研究品質と教育的価値を向上できる。さらに、本フレームワークは複数の医療データベースへ横展開可能であり、将来的な研究基盤標準化や自動化にも資する汎用的手法である。

A. 研究目的

大規模な時系列医療データベースの研究利用が急速に拡大している。わが国においては、匿名医療保険等関連情報データベース (NDB)、DPC 参加病院の退院患者情報を収録した DPC データベース、民間保険者の医療情報を収録した JMDC・DeSC データベース等、いずれも数テラバイト規模に達するデータが研究者に提供されるようになった(表 1)。これらのデータベースは、「誰が、いつ、どのような医療を受けたか」を記録した時系列情報の集積であり、疫学研究のきわめて有力なリソースである。

データの取り扱いには一般に SQL を用いたデータハンドリングが行われる。しかし、SQL は記述の自由度が高いゆえに、研究者ごとに独

自のクエリが書かれる傾向がある。そのため、クエリ設計の方針が担当者の経験と裁量に依存し、コードが引き継がれる際に意図が伝わらず、研究の再現性が損なわれることも少なくない。そのため、SQL のクエリを共有しても、そこには膨大な量のレビュープロセスが発生し、第三者によるコードレビューが困難である。また、エラーが潜在しても発見が難しい。また、数テラバイトに及ぶ元テーブルを不必要に全件スキャンしたり、同一内容のテーブルを重複してコピーしたりするクエリが散見され、計算資源を浪費する原因ともなっている。

本報告の目的は、大規模時系列医療データを疫学研究に使用しやすい形に整理するための、標準化されたデータハンドリングプロセスを一般化可能なように提示することである。具体的には、(1)解析単位、(2)観察のタイムライン、(3)マスタ整備という三要素を軸としたク

エリ設計の考え方を示し、各ステップを体系的に解説する。

B. 研究方法

B.1 大規模時系列医療データベース

わが国で研究利用される主要な大規模医療データベースを表 1 に示す。いずれのデータベースも、原則として診療の都度生成されるレセプト請求データを基盤としており、医薬品の処方・調剤、診療行為、傷病名、検査結果等の情報が時系列順に記録されている。

表 1. 主要な大規模医療データベース

データベース	収録対象	収録期間の目安	データ規模 (概算)
DPC データベース	DPC 参加病院の入退院サマリ情報 (全国)	2004 年～	年間約 6TB
JMDC データベース	一部保険者 (社保) の医療レセプト・特定健診	約 17 年分	約 1.5TB
DeSC データベース	一部保険者 (社保・国保・後期高齢) の医療レセプト・特定健診	約 8 年分	約 2.5TB

データベース	収録対象	収録期間の目安	データ規模 (概算)
NDB データベース	全保険者の医療レセプト請求情報・特定健診	約 10 年分	約 10TB

B.2 NDB データベースの情報構造と時系列整理

NDB に収録される情報には、レセプト情報 (医科・DPC・調剤・歯科) と特定健診情報が含まれる (表 2)。近年では、死亡個票情報、匿名介護保険等関連情報データベースなどの公的データベース、次世代医療基盤法に基づく医療機関 EHR 情報等との連結も制度的に可能となり、活用可能な情報範囲は拡大している。

オンサイト・特別抽出データでは診療行為・医薬品等の実施年月日まで取得可能であり、日単位での詳細な時系列解析が可能である。一方、HIC 環境下の NDB-β では日付情報がマスクされ、月単位での解析に制限される。そのため、研究計画段階で利用環境を明確にし、利用可能な最小時系列粒度を前提にデータ整理を行う必要がある。

表 2. NDB データに含まれる主要情報区分と時系列解像度

医科レセプト

レコード種別	主な内容	時系列粒度
RE / HO / IR	レセプト共通ヘッダー(患者属性、診療年月、保険情報)	月単位
SY	傷病名、主傷病、修飾語、診断開始年月日	月単位 + 診断開始日
SI	診療行為、処置、検査、手術	日単位
IY	投薬情報(薬剤、用量、投与日数)	日単位
TO	手術・麻酔等詳細	日単位

DPC レセプト

レコード種別	主な内容	時系列粒度
RE	レセプト共通ヘッダー	月単位
SY / SB	傷病情報、診断開始年月日	月単位 + 診断開始日 / 入院単位
SI	出来高診療行為	日単位
IY	医薬品情報	日単位
TO	手術・処置	日単位
CD	DPC 包括評価関連コード	日単位

調剤レセプト

レコード種別	主な内容	時系列粒度
RE	レセプト共通ヘッダー	月単位
IY(SH)	調剤薬剤情報	日単位
CZ(SH)	調剤行為詳細	日単位
TO(SH)	特定加算・補足	日単位

歯科レセプト

レコード種別	主な内容	時系列粒度
RE	レセプト共通ヘッダー	月単位
SY / SB / SK	歯科傷病情報、開始日	月単位 + 診断開始日
SI	歯科診療行為	日単位
IY	歯科薬剤情報	日単位
TO	歯科処置・手術	日単位
SS	歯式情報	月単位または処置単位

B.2.1 NDB データの最小時系列粒度でのデータ整理

NDB データを最小時系列粒度で整理するためには、データ保存構造そのものを事前に設

計したうえで処理を進める必要がある。我々は「NDB Local PostgreSQL 最終出力テーブル仕様書」(添付1)に基づき、RE レコードを起点として患者ID、レセプトID、診療年月を整理する receheadertbl を作成している。

この設計により、「だれが」「いつ」「どうしたか」という疫学研究に必要な基本構造を保持しつつ、各診療行為情報を統合可能となる。さらに、SI、IY、TOレコードなど日単位情報を持つデータについては、実施年月日を付加することで、診療報酬請求コード単位で再整理を行う。

この一連の整理により、NDB データ全体を統一された最小時系列粒度で管理可能となる。HIC 環境では日付情報が制限されるため、月内代表日を設定して時系列表現を補完する。

B.3 大規模時系列医療データの特性

疫学研究において重要なのは、「だれが、いつ、どうしたか」を明確に記述することである。大規模医療データベースは本質的にその情報を蓄積しているが、研究利用には単純保存だけでは不十分であり、研究デザインに応じたデータハンドリングが必要となる。

したがって、大規模時系列医療データの活用には、時系列構造を維持しながら研究目的に即して再整理する体系的プロセスが求められる。

B.4 疫学的時系列構築の概念的フレームワーク

疫学研究デザインに対応した時系列構築には、解析単位、時間的アンカー、観察ウィンドウ、マスタテーブル整備の4要素が必要となる。

解析単位とは、解析テーブル上の1行に対応する観察対象であり、個人単位、入院単位、手術単位、ICU 滞在単位など研究目的に応じ

て設定される。すべての解析単位は固有のタイムラインと Index date を持つ。

時間的アンカーとは、タイムライン上の基準日であり、Index date、イベント日、打ち切り日などが該当する。観察ウィンドウは、これらアンカーを起点として定義される時間区間であり、washout、baseline、exposure、follow-up など研究目的別に設定される。

また、診療コード・薬剤コード・病名コード等を研究カテゴリへ変換するため、マスタテーブルを整備する必要がある。これにより、コード管理の保守性と再利用性が確保される。

B.5 標準化された SQL ワークフロー

本研究では、概念的フレームワークを具体的な SQL 処理として実装するため、解析単位テーブル作成、可変アンカー付加、マスタ結合、ウィンドウベース集計、Long-to-Wide 変換という標準化ワークフローを採用した。

まず、Index date を基準とする解析単位テーブルを作成し、各観察単位に必要な可変アンカーを付加する。その後、マスタテーブルを用いて時系列データへカテゴリ情報を付与し、定義された観察ウィンドウごとに必要な情報を集計する。最終的に Long 形式の集計結果を Wide 形式へ変換し、解析可能な最終テーブルを構築する。

この構造化プロセスにより、再現性、可読性、保守性、計算効率を兼ね備えた標準的なデータハンドリングが可能となる。

C. 研究結果

C.1 標準化 SQL データハンドリングフレームワークの構築

本研究では、大規模時系列医療データを疫学研究へ効率的かつ再現性高く利用するため、

解析単位、時間的アンカー、観察ウィンドウ、マスタテーブル整備の 4 要素を基盤とした標準化 SQL データハンドリングフレームワークを構築した。

このフレームワークにより、研究デザイン概念図(デザインダイアグラム)と SQL クエリ構造を一对一で対応させることが可能となった。すなわち、「だれが」「いつ」「どうしたか」という疫学研究の本質的構造をそのままデータベース設計とクエリへ反映できるようになった。

また、解析単位テーブルを起点として、個々の研究デザインに応じた可変アンカーを付加し、必要な患者・必要な観察期間のみを対象とした時系列データ抽出が可能となった。その結果、不必要なフルスキャンを回避しつつ、研究目的に即した効率的なデータ抽出が実現した。

C.2 SQL ワークフローの標準化

提案フレームワークに基づき、以下の一連の標準化ワークフローが整理された。

まず、Index date を持つ解析単位テーブルを作成し、研究ごとに異なる観察開始日・終了日・イベント日などの可変アンカーを付加した。次に、マスタテーブルを用いて時系列データへカテゴリ情報を付与し、観察ウィンドウごとの曝露・共変量・アウトカム情報を抽出した。さらに、Long 形式で集計した情報を Wide 形式へ変換することで、統計解析へ直接投入可能な最終解析テーブルを構築した。

この構造により、複雑な時系列データ処理が明確なモジュール単位に分割され、各工程の独立性と保守性が大きく向上した。

C.3 再現性およびレビュー容易性の向上

本フレームワーク導入により、SQL クエリの再現性は大幅に向上した。従来は担当者依存であったクエリ設計が、標準化された設計思想

に基づいて構築されるため、研究者間での引継ぎや第三者レビューが容易となった。

特に、デザインダイアグラムとクエリ設計を対応させることで、コードレビュー時に研究デザインから逸脱した処理や不適切なウィンドウ設定を発見しやすくなった。これにより、ブラックボックス化した長大な SQL クエリの問題が大幅に軽減された。

C.4 計算効率および保守性の改善

解析単位テーブルを中心とした JOIN 構造により、必要最小限のデータ抽出が可能となり、大規模データベースに対する計算効率が改善した。さらに、コードのハードコーディングを排除し、マスタテーブル管理へ統一したことで、コード体系改定時や研究条件変更時の修正範囲が限定され、保守性が向上した。

加えて、中間テーブルの活用や CLUSTERED COLUMNSTORE INDEX の導入を前提とすることで、数 TB 規模のデータベースに対しても現実的な計算時間で運用可能な設計となった。

C.5 横展開可能性

本フレームワークは NDB データベースのみならず、DPC、JMDC、DeSC 等の他の大規模時系列医療データベースにも適用可能であることが確認された。各データベース固有のコード体系やテーブル構造の違いは存在するものの、解析単位、アンカー、観察ウィンドウ、マスタという基本構造は共通しており、マスタテーブルおよび一部テーブル定義を差し替えることで同一フレームワークを横展開できる。

この結果、本研究で提案する標準化 SQL データハンドリングフレームワークは、時系列医療データ研究全般に共通する基盤技術として利用可能であることが示された。

D. 考察

D.1 本フレームワークの意義

本報告で提示した標準化 SQL データハンドリングフレームワークは、従来属人的であった大規模医療データ解析を、疫学研究デザインに基づく体系的な設計へ転換する点に大きな意義がある。

最大の強みは、研究デザイン、デザインダイアグラム、SQL クエリ構造を一貫した論理構造として接続できることである。従来の SQL 設計では、担当者個人の経験や記述習慣に依存しやすく、同一研究テーマであってもクエリ構造が大きく異なり、再現性やレビュー容易性に課題があった。本フレームワークでは、「解析単位」「時間的アンカー」「観察ウィンドウ」「マスタテーブル」という共通概念を用いることで、研究目的とクエリ構造の整合性が担保される。

さらに、解析単位テーブルを中心とするモジュール型構造は、可読性・保守性を大幅に向上させる。各工程が独立して管理されることで、研究条件変更時にも全体を書き換える必要がなく、限定的修正で対応可能となる。これは長期的な研究基盤整備において重要である。

D.2 再現性・教育的価値

標準化された構造は、単に個別研究の効率化にとどまらず、研究教育にも有用である。大規模医療データ解析は高度な SQL 技術を要求する一方、その教育体系は十分に整備されていないことが多い。

本フレームワークを用いることで、初学者であっても「どの順序で」「どの概念に基づき」データハンドリングを行うべきかが明確になる。これは教育コストを低減し、組織全体での解析品質向上に寄与する。

また、第三者レビューにおいても、研究デザインから SQL クエリへの変換過程が透明化されるため、疫学的方法論担当者、統計担当者、データエンジニア間の協働が容易になる。

D.3 パフォーマンスと実装面での重要性

数 TB 規模のデータベースを扱う現代の医療データ研究において、計算効率は研究実現可能性そのものに直結する。本フレームワークでは、解析単位テーブルを起点とする対象限定型 JOIN、マスタテーブル活用、中間テーブル管理、列指向インデックス活用などにより、パフォーマンス面でも合理的設計となっている。

これにより、不必要な全件スキャンや冗長なコピーを抑制し、計算資源消費を大幅に削減できる。したがって、本手法は単なる方法論的整備だけでなく、実務的運用基盤としても有用である。

D.4 限界

一方で、本フレームワークにはいくつかの限界も存在する。

第一に、データベースごとの固有仕様への対応は依然として必要である。NDB、DPC、JMDC、DeSC ではテーブル構造、コード体系、時間粒度が異なるため、個別マスタ整備やデータ理解は不可欠である。

第二に、コード体系の改定や制度変更に応じた継続的マスタ更新が必要となる。マスタテーブルを導入しても、その維持管理コストは残る。

第三に、研究デザインそのものが不適切であれば、SQL 標準化のみでは研究妥当性は保証されない。したがって、本フレームワークはあくまで「適切な研究デザインを高品質に実装する基盤」として位置づけられる。

D.5 将来的展望

本フレームワークは、将来的な SQL テンプレート自動生成、共通データモデルとの接続、AI 支援型クエリ生成、研究プロトコル自動実装などへの発展可能性を有する。

研究デザインを構造化データとして保持できれば、将来的にはプロトコルから半自動的にSQL生成を行う研究基盤も構築可能となる。その意味で、本研究は単なるクエリ標準化にとどまらず、大規模疫学研究インフラ整備の第一歩と位置づけられる。

E. 結論

本報告では、大規模時系列医療データを疫学研究に利用するための標準化SQLデータハンドリングフレームワークを提示した。

本フレームワークは、「だれが(解析単位)」「いつ(タイムライン)」「どうしたか(マスタ整備)」という疫学研究の本質的三要素を体系化し、解析単位テーブル作成、可変アンカー付加、時系列データ結合、観察ウィンドウ集計、Long-to-Wide変換という一貫したワークフローとして整理した。

これにより、従来の属人的かつ非標準的であった大規模医療データ解析に対し、再現性、可読性、保守性、計算効率を兼ね備えた汎用的研究基盤を提供できる。

本手法は、NDB、DPC、JMDC、DeSCなど多様な大規模時系列医療データベースへ適用可能であり、個別研究の品質向上のみならず、将来的な研究基盤標準化、教育体系化、半自動化研究環境整備にも寄与する。

大規模医療データ研究の発展に伴い、SQLデータハンドリングの標準化は、今後ますます重要な基盤技術となると考えられる。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表(著作)

松居宏樹. 匿名医療保険等関連情報データベースの利用経験および今後の期待と課題. 統計. 2025;76(8):14-22.

H. 知的財産権の出願・登録状況

(予定を含む。)

なし

令和7年度厚生労働行政推進調査事業費補助金
(政策科学総合研究事業(政策科学推進研究事業))
「NDBのユーザビリティ向上を通じて
クラウド上でのデータ二次利用を推進するための研究」
分担研究報告書

利便性の高いクラウド環境の検討・提案

研究分担者

柏木 公一 国立健康危機管理研究機構・国立看護大学校・准教授
杉山 雄大 国立健康危機管理研究機構・国立国際医療研究所糖尿病情報センター

研究協力者

市瀬 雄一 国立健康危機管理研究機構・国際医療協力局グローバルヘルス政策研究センター
古野 孝志 国立健康危機管理研究機構・国際医療協力局グローバルヘルス政策研究センター

研究要旨

NDB(匿名医療保険等関連情報データベース)におけるHIC(医療・介護データ等解析基盤)では、Redshift・ローカルPostgres等の解析環境が提供されている。一方、同環境においては、Redshift永続テーブル作成権限の欠如、ローカル解析サーバのメモリ・ストレージ不足、外部参照データ取込の困難、計算コストの増大といった複合的課題が存在する。

本分担研究では、HIC内で提供されるクラウドストレージ(Amazon S3)と列指向ファイル形式Parquet、軽量分析エンジンDuckDBを組み合わせた分析基盤を構築し、その実用性を検証した。

2024年4月から2025年3月の医科・DPC・調剤の全テーブル(1年分、計620GB)をRedshiftからS3上にParquetとして生成し、MD5ハッシュの加算突合により684テーブルすべてについてデータの一致を確認した。加えて、全テーブルの各カラムについて、NULL数、ユニーク値数、出現頻度上位50件、最大・最小・平均・標準偏差からなるデータサマリを生成した。

HICの乙区分(メモリ64GB)の環境において、本構成が大規模レセプトデータに対する実用的な処理性能を有することを確認した。ストレージ料金も1年分620GBで月額約2,100円と低コストで運用でき、従来のRedshift依存の分析環境におけるコスト増大・柔軟性の不足といった課題を解消できる可能性を示した。

A. 研究目的

本研究で対象とする NDB(匿名医療保険等関連情報データベース)は、日本の公的医療保険における診療報酬明細書(以下「レセプト」)および特定健診・特定保健指導等の情報を収集・格納した匿名化された大規模医療データベースであり、クラウド基盤上の医療・介護データ等解析基盤(HIC: Healthcare Information Center)が提供されている。本分担研究は、厚生労働科学研究「NDBのユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究」の一環として、HIC内の分析基盤の柔軟性およびコスト効率を改善するための技術的検証を担うものである。

既存の Redshift 環境には、研究用途における柔軟なデータ処理を妨げる技術的困難が複数存在する。まず、研究者に対して永続的なテーブル作成権限が付与されていないため、段階的に中間成果物を生成しながらデータ処理を進めることができず、複雑な集計・加工処理を単一のクエリとして一括実行する必要がある。例えば、1) ある医療行為が含まれるレセプトの患者 ID を取得し、2) この患者 ID について前後月の医療行為を取得しようとする場合、通常は、前段の 1 のクエリ結果を中間テーブルとして保存し、その後 2 以降の処理を行う。しかし提供されている Redshift 環境では中間テーブルを作成することができず、毎回 1 の集計を再実行せざるを得ない(一時テーブルであっても、セッションごとに再作成が必要なため、セッションをまたぐ分析では同様の制約が残る)。

Redshift のクエリ実行は計算資源の利用量に応じたコストが発生するため、大規模な試行錯誤を伴う探索的分析を行う際には費用も無視できない課題であった。さらに、既存の Redshift 環境には CSV 等の外部データを直接取り込むためのインポート機能が整備されておらず、傷病名コードや薬剤コードの条件として数千件規模の参照データを利用する場合、条件として IN 句を用いるか、INSERT 文や UNION ALL を連結した長大な SQL を作成する必要があった(図 1)。このような方法は可読性・保守性に乏しく、分析作業の効率を低下させる要因となっていた。代替手段としてローカル環境に CSV 形式でデータをダウンロードし、PostgreSQL 等に取り込む方法も提供されている。HIC は用途別に甲・乙・丙の 3 種類の解析サーバ環境を提供しており、利用可能なメモリ容量はそれぞれ 128GB、64GB、32GB、ストレージ容量は 16TB、3TB、1TB である(厚生労働省, 2026)。このため、甲以外の環境ではローカルストレージに大規模データを保持することが困難であり、メモリ制約からも大規模データ分析は現実的ではない。特に、糖尿病、高血圧症、脂質異常症等の生活習慣病のように、数千万人規模の患者を対象とする分析では、ローカル環境での処理は事実上不可能であった。

以上のように、既存の分析環境では、権限制約、参照データ利用の非効率性、ならびにローカル計算資源の不足が複合的に存在し、大規模レセプトデータを対象とした実用的な分析基盤の構築は困難な状況にあった。本研究は、これらの制約を解消する代替アーキテクチャの技術検証を目的とする。

B. 研究方法

本研究では、既存の分析環境における権限制約、計算コスト、およびローカル計算資源の制限といった課題を解決するため、クラウドストレージ、列指向ファイル形式 Parquet、および軽量分析エンジン DuckDB を組み合わせた新たな分析基盤を構築した。

1. Parquet について

Parquet は、Apache Software Foundation が開発するオープンソースの列指向ファイルフォーマットであり、大規模データに対する分析処理を主な用途として設計されている。CSV 等の行指向形式が 1 レコード分の値を行単位で連続的に格納するのに対し、Parquet は同一カラムの値をひとまとまりとして格納する列指向方式を採用している。

この構造には主に 3 つの利点がある。第一に、分析時に必要なカラムのみを選択的に読み込むことができ、不要なカラムに対する I/O が発生しない。第二に、同一カラム内では値の型および分布が類似するため、汎用の圧縮アルゴリズム(本研究では ZSTD を採用)を適用した際に高い圧縮率が得られ、ストレージ容量およびネットワーク転送量を削減できる。第三に、カラム単位の最小値・最大値等の統計値がメタデータとして格納されており、検索条件に該当しないデータブロックを読み飛ばす最適化(predicate pushdown)が可能である。レセプトデータのように、同一スキーマのレコードが多数存在し、かつ分析時には特定カラムのみが参照される用途においては、Parquet 形式の特性が効果的に作用する。

3

2. DuckDB について

DuckDB は、分析用途(OLAP : OnLine Analytical Processing)に特化したオープンソースの組込型データベースエンジンである。サーバ構築や事前のデータベース構築を必要とせず、コマンドラインからの利用や、R や Python のスクリプトからも直接利用できる。実行エンジンは列指向かつベクトル化処理を採用しており、集計・結合等の分析クエリを高速に処理する。また、Parquet ファイルを外部テーブルとして SQL から直接クエリする機能をネイティブに備えており、httpfs 拡張を読み込むことで、ローカルディスク上のファイルだけでなく Amazon S3 等のクラウドストレージ上のファイルに対しても同一の構文で操作可能である。さらに、利用可能なメモリを超える大規模データを対象とする場合には、中間結果の一部をディスクに退避させながら処理を継続する機構(spill-to-disk)を備えており、限られたメモリ環境における大容量データの分析に適している。本研究では、DuckDB version 1.4.2 と httpfs エクステンションを解析サーバ上にインストールした。

3. 実装

まず、HIC で一時データ保存領域として利用可能な Amazon S3 に、Redshift 上のレセプトデータを Parquet 形式として保存した。この構成により、大容量データをローカルストレージに保持する必要がなくなり、解析サーバにおけるストレージ容量の制約を実質的に解消することが期待される。

次にデータの検証を行った。Redshift から分割ダウンロードを行った際には、正確にデータが取得できたかを検証する

仕組みが必要である。一般的には取得行数の確認が用いられるが、本研究では100万行単位でダウンロードを行っており、同一ブロックが重複ダウンロードされた場合に行数集計では、重複を検知できない場合がある。そこで、RedshiftとParquetファイルの両方でデータが一致していることを検証する仕組みとして、件数(行数)の突合に加え、ハッシュ値に基づく照合手法を考案し、実行可能性を検証した。

さらに、格納データの分析可能性を検証するため、全テーブルの各カラムに対して、ユニーク値の数、NULLの数、出現頻度が高い上位50件、最大値、最小値、平均値、標準偏差からなるデータサマリを算出した。

C. 研究結果

1. S3上のParquetファイルの生成

実行可能性の検証として、2024年4月から2025年3月の1年間の医科、DPC、調剤の全テーブル情報をRedshiftからS3上にParquetファイルとして格納した。

使用したS3ストレージは、「一時フォルダ格納領域」として割り当てられているworkを用いた。Redshiftから多量データをロードする際のローカルメモリ容量の制約、およびParquetファイルを適切なサイズに留める観点から、1ファイルあたり100万行単位でファイル分割を行った。実装したRスクリプトでは、追加ライブラリとしてRedshiftへのアクセスにPostgres、Parquetファイルの生成にarrowを使用した。実行時間は1ヶ月分あたり平均5時間34分、12ヶ月合計で66時間53分であっ

た。データ容量は1年分で620GB(内訳:医科399GB、DPC25.5GB、調剤196GB)であった(Parquetファイルの圧縮アルゴリズムはZSTD、圧縮レベル1を採用した)。

2. RedshiftとParquetのデータ検証

Redshiftから分割ダウンロードしたParquetファイルが一致することを確認するために、検証用カラムとして、いずれのレセプトにも存在し、かつレセプト単位で一意的な値を持つseq2_noを対象に、各行のMD5ハッシュ値(128ビット、16進32文字)を算出した。これをそのまま加算すると桁あふれを起こすため、本手法では32文字のうち上位4文字(16ビット相当、10進換算で0~65,535の整数)を抽出し、テーブル内の全レコードにわたって64ビット整数として合計(SUM)した値を、Redshift側とDuckDB(Parquet読込)側の双方で算出して突合する方式とした。

1つのテーブルを複数の分割ファイルとして保存する場合にも対応できるよう、加算によって全体集計値を得られるハッシュ合成を採用している点が本手法の要点である。上記手法で検証した結果、684テーブルすべてについてデータの一致を確認できた。

3. DuckDBによるS3上のparquetファイルの操作

DuckDBからS3上のファイルを読み込むには、httpfs拡張のインストールとロードを行った後、AWS認証が必要である。HIC環境では、標準的な認証情報探索ルールであるcredential_chainを指定するだけでよい。

1回のみ

```
INSTALL httpfs;
# 以下はセッションごとに必要
LOAD httpfs;
CREATE SECRET (
TYPE s3,
PROVIDER credential_chain
);
```

DuckDB を使用した典型的なクエリは以下のようなになる。

```
SELECT * FROM read_parquet(
'S3://[work フォルダ]/path/file.parquet');

SELECT * FROM read_parquet(
'S3://[work フォルダ]/fy=2024/**/*.parquet')
```

パスの指定に* (アスタリスク) を使用できるため、年度や月を指定したクエリが指定できる。なお、DuckDB は、標準で複数ファイルを同時に処理するため、ファイルが分割されていることは高速化に寄与しやすい。

パスに/name=value/という形式(HIVE形式)が含まれる場合 SQL 内で name の値を利用することができる。一般に年月単位でテーブルを JOIN する場合、次のクエリのように S3 のパスとして ym=202404 が含まれれば、ym カラムとして SQL 内で値を利用できる。

```
SELECT
med_re.id5,
med_si.rcpt_cd,
ym
FROM
read_parquet('S3://[work]/med_si/*
*/ym=202404/*.parquet') AS med_si
```

```
JOIN
read_parquet('S3://[work]/med_re/*
*/ym=202404/*.parquet') AS med_re
ON med_si.req2_no = med_re.req2_no
```

4. データサマリ生成

各カラムのデータサマリを DuckDB を用いて作成した。ID 関連などユニーク値が多いカラムについては、12 ヶ月分をまとめて処理することは解析サーバのメモリ不足によって実行できなかったため、1 ヶ月ごとに各カラムのユニーク値と件数を求めた後に 1 年分を集計した。集計したデータサマリはいったん JSON 形式で保存した後、テーブル別・カラム別でアクセス可能な HTML 形式に加工した。実行時間は以下のとおりである。

- 1) 月ごとに各カラムのユニーク値と件数を集計: 6 時間 17 分 53 秒
- 2) 月別集計をまとめて年集計へ: 6 時間 16 分 55 秒
- 3) 各カラムの固有値の数、最小値、最大値、NULL の数、平均、標準偏差の集計: 31 分 34 秒
- 4) 各カラムの出現頻度上位 50 位を集計: 31 分 24 秒
- 5) 全集計データをまとめる: 32 秒

D. 考察

S3 上の Parquet ファイルは、HIC の乙スペック(メモリ 64GB)において問題なく処理できた。これによって、解析サーバのストレージ容量の制約や、中間テーブルを作成できないといった既存環境の問題が実質的に解消される。

Parquet 形式は列指向でデータが格納され、かつ列単位で圧縮が施されるため、クエリ実行時には必要なカラムのみを読み出すことができる。この特性により、ネットワーク転送量を抑制しつつ効率的なデータ処理が可能となる。本研究で使用した HIC 内の EC2 インスタンスにおいて、S3 との通信スループットは単一接続で約 100~500 Mbps の範囲にあり、複数ファイルへの並列アクセスを組み合わせることで、実効で 1 Gbps を超えるデータ読込性能が得られ、大規模データに対する実用的なクエリ実行性能を確保できた(なお、これらの数値は本研究環境における実測値であり、EC2 のインスタンスタイプやネットワーク構成に依存する)。

Redshift から S3 上への Parquet ファイル生成には、医科・DPC・調剤の 1 年分で約 67 時間を要したが、並列数を 4 まで増やす限りにおいては処理速度が線形に向上することが確認できた。つまり、同時実行数を 4 とした場合は約 17 時間で生成できると考えられる。また、生成した Parquet ファイルは、全利用者で共通に使用可能な性質のものであり、あらかじめ作成したものを読み取り専用として共有する運用も可能と考えられる。この場合は、Redshift は使用せず、解析用 PC を準備し、620GB×年数分の S3 ストレージを共有できればよい。

料金について、Amazon S3 (Standard クラス)の東京リージョンにおけるストレージ料金は GB あたり月額約 0.024 米ドル(本報告時点の為替換算で約 3.5 円/GB/月)と比較的低コストであり、1 年分の医科・DPC・調剤の全データ 620GB を保持しても月額約 14 米ドル(約 2,100

円)である。また、同一リージョン内における S3 と EC2 間のデータ通信は追加料金が発生しないため、大規模データを継続的に利用する研究用途において費用効率の高い構成となる(別途、S3 API 呼出しに対するリクエスト課金が発生するが、GET リクエスト 10 万件あたり約 6 円程度と軽微である)。さらに、S3 は単一アカウントのストレージ総量に事実上の上限がないため、数十億レコード規模のレセプトデータを対象とした長期的な分析基盤としても持続的に利用可能である。

本研究で使用した Redshift データベースと EC2 分析サーバの単位時間あたりの利用料金を比較すると、同等の稼働時間において Redshift のほうがおおむね 10 倍程度高い試算となった(ノードタイプ・インスタンスタイプに依存するため、環境による差異はある)。この点からも、EC2 分析サーバの計算資源のみで分析可能な S3 上の Parquet ファイルを用いる構成は、費用面での優位性が高いと考えられる。

一方、DuckDB と Parquet を用いた分析手法には限界もある。大量データの分析においては解析サーバのメモリが最も強い制約となり、本分析手法では、メモリ不足を回避するために 1 ヶ月ごとに処理した中間結果を後段で統合するといった追加の手間がかかっている。その点、Redshift では潤沢なメモリを前提とした解析が可能であり、たとえば 64 文字で提供される ID を INT32 に置き換えるような処理は、Redshift で実行する方がはるかに手間がかからず高速である。

E. 結論

HICの実環境において、684テーブルに対するデータ一致検証、および全テーブル・全カラムのデータサマリ生成という、大規模データを対象とした代表的な集計処理を実施し、本構成が研究用途において実用的な処理性能を有することを確認した。また、本研究で採用した DuckDB はメモリ効率に優れた列指向分析エンジンであり、64GB のメモリ環境においても大規模データに対する実用的なクエリ処理が可能であることが確認された。

以上により、従来の Redshift 依存の分析環境におけるコスト増大や柔軟性の不足といった課題を解消し、低コストかつ再現性の高い大規模レセプトデータ分析基盤を実現できたと考える。

2026年3月現在、DuckDBはHIC環境の標準ツールとなっておらず、別途申請が必要である。コマンドライン版

(CLI)、Python、Rのライブラリについては標準ツールとしての採用が望まれる。

F. 健康危険情報

特記すべき事項はない。

G. 研究発表

なし

H. 知的財産権の出願・登録状況

なし

引用文献

厚生労働省 (2026). 匿名医療保険等関連情報データベース (NDB) の第三者提供よくあるご質問 (FAQ) 2026年3月版.

<https://www.mhlw.go.jp/content/1240000/001679224.pdf>

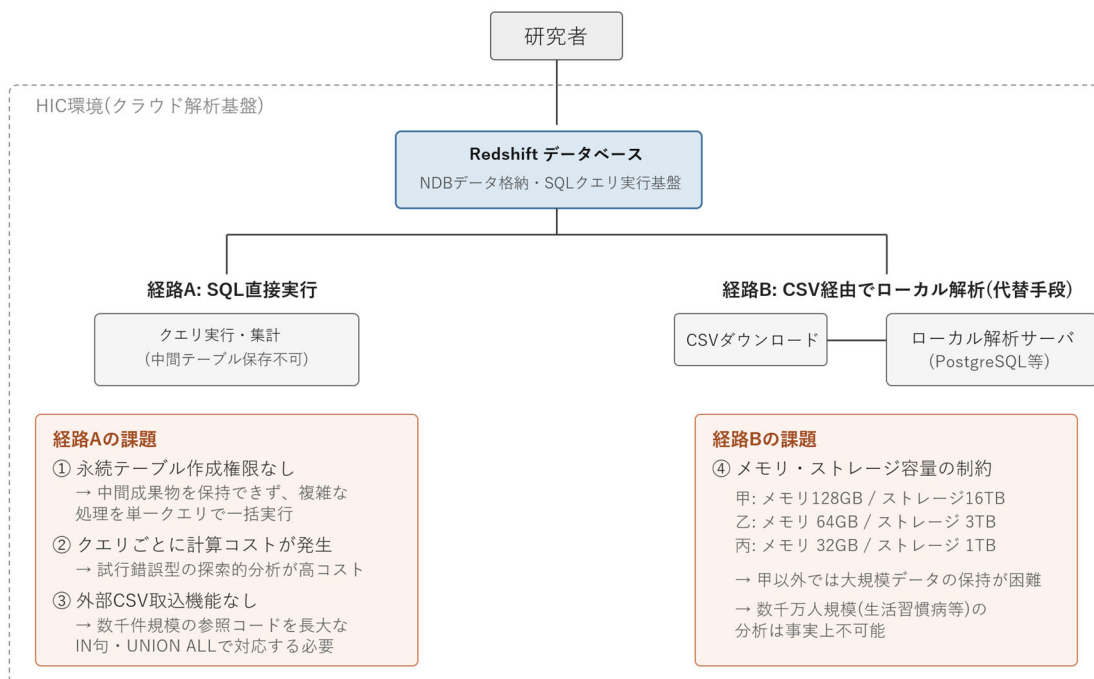


図 1 既存環境におけるレセプトデータ解析の構成と課題

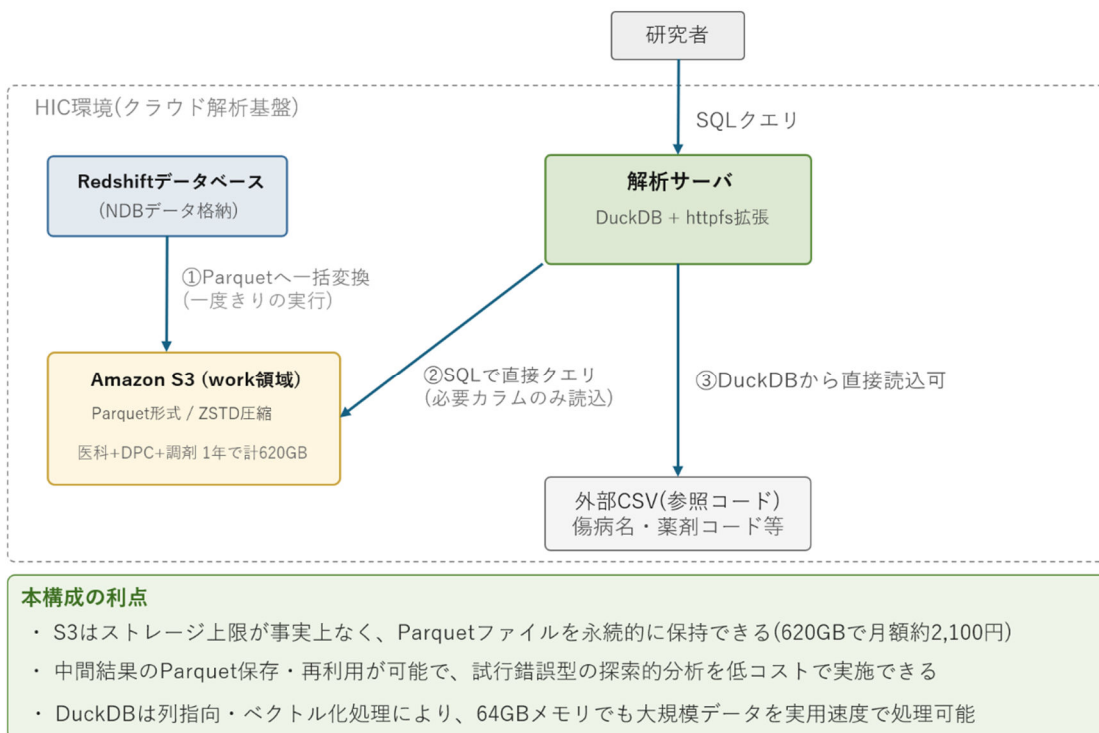


図 2 提案アーキテクチャ (S3 + Parquet + DuckDB)

厚生労働行政推進調査事業費補助金（政策科学総合研究事業（政策科学推進研究事業））

「NDB のユーザビリティ向上を通じて
クラウド上でのデータ二次利用を推進するための研究」
分担研究報告書（令和7年度）

諸外国の医療関連データ利活用・提供状況の調査

研究代表者 明神 大也（浜松医科大学 健康社会医学講座）

研究分担者 牧戸 香詠子（東京大学大学院 医学系研究科生物統計情報学）

研究要旨

本研究は、医療・介護データ等解析基盤（HIC）における NDB データのユーザビリティ向上を目的とし、諸外国の医療データ提供体制を調査し、改善提案を検討した。韓国では、NHID や HIRA といった大規模データベースが整備され、オンサイトセンターを中心とした運用のもと、NHID では年間約 1,000 件のデータ提供の実績がある。フィンランドでは、Findata を通じた一元的な提供体制およびリモート解析環境が整備され、レディメイドデータセットの提供により迅速化とコスト削減が図られている。フランスでは、Health Data Hub を介して、レセプト等のデータベースである SNDS と希少疾患のレジストリである BNDMR の連結が進められている。スウェーデンでは個人番号制度を基盤として、行政・医療・福祉分野のレジストリが多層的に整備されている。日本の NDB においても HIC 上で多様なデータセット提供が進む一方、研究可能性の明確化や操作性の改善が課題である。今後は海外事例を踏まえ、データセットごとの活用可能性の整理やユーザビリティ向上を図り、医療情報の二次利用の促進が求められる。

研究協力者：

加藤 源太（京都大学医学部附属病院 診療報酬センター／病床運営管理部）

市瀬 雄一（国立健康危機管理研究機構）

A. 研究目的

医療 DX の推進に関する工程表（令和 5 年 6 月 2 日医療 DX 推進本部決定）および規制改革実施計画（令和 5 年 6 月 16 日閣議決定）において、クラウド解析基盤である医療・介護データ等解析基盤（Healthcare Intelligence

Cloud : HIC）を用いた NDB データの迅速提供が進められている。しかし、プリセットデータが CSV 形式で提供されることやコマンドライン操作が必要であることなど、ユーザビリティには課題が残る。こうした中、規制改革推進に関する中間答申（令和 6 年 12 月 25 日）では、公的データベースの仮名化情報を安全かつ迅速に利用できるクラウド環境の整備が求められている。本研究では HIC における NDB データのユーザビリティ向上を目的として、①ユースケース調査、④汎用性の高いデー

タセットの検討・具体的な仕様作成、および⑤
利便性の高いクラウド環境の検討・提案を行
う。本報告書では、その前提として諸外国の医
療データベースの運用状況を概説する。

B. 研究方法

NDB/HIC 解析環境の改善点、データ提供形
式、マニュアル等の在り方等について検討し、
利便性向上に資する提案を取りまとめること
を目指し、韓国・フィンランド・フランス・ス
ウェーデンにおける医療データ提供の運用状
況について、ヒアリング及び文献調査を行っ
た。

C. 研究結果

【韓国のデータベース】

① 概要

韓国の人口は約 5,175 万人（2024 年 7
月 1 日時点）である。出生率の低下と平均
寿命の延長に伴い、人口の高齢化が急速に
進行している。韓国の社会保険制度は、国

民健康保険（National Health Insurance）、
国民年金、雇用保険、労災保険、高齢者介
護保険の 5 つの制度から構成される。

韓国は日本と同様に国民皆保険制度を
採用しており、国民健康保険公団（NHIS:
National Health Insurance Service）がほぼ
全国民に医療保険を提供している。国民健
康保険制度は 1963 年の医療保険法の制定
により導入され、1989 年に全国民を対象
とする皆保険体制が達成された。その後、
2000 年には公的医療保険制度の下で単一
保険者体制が確立されるとともに、医療費
の審査・評価機能を担う健康保険審査評価
院（HIRA: Health Insurance Review and
Assessment Service）が設立された。韓国で
は医療費の電子請求率がほぼ 100%に達し
ており、医療機関から提出された医療費請
求は HIRA によって審査され、その結果が
NHIS および医療機関に通知される仕組み
となっている。¹（図 1）



図 1. 韓国における医療保険制度

※HIRA 提供資料「HIRA のビッグデータの現状と活用」を基に改変²

② 保有する医療保険関連データベース

韓国における公的な医療データベースとしては、単一保険者である NHIS が提供する NHID (Korean National Health Information Database) と、HIRA が提供する HIRA Big Data の 2 つが存在する。

NHID には、20 年以上にわたる全国規

模の医療・健診データに加え、社会経済的地位 (SES) などの情報が個人単位、追跡可能な形式で収録されている。また、複数の運用システムから収集された膨大なデータを統合し、研究支援、PHR サービスの提供、医療政策立案の支援などに活用されている。(図 2)

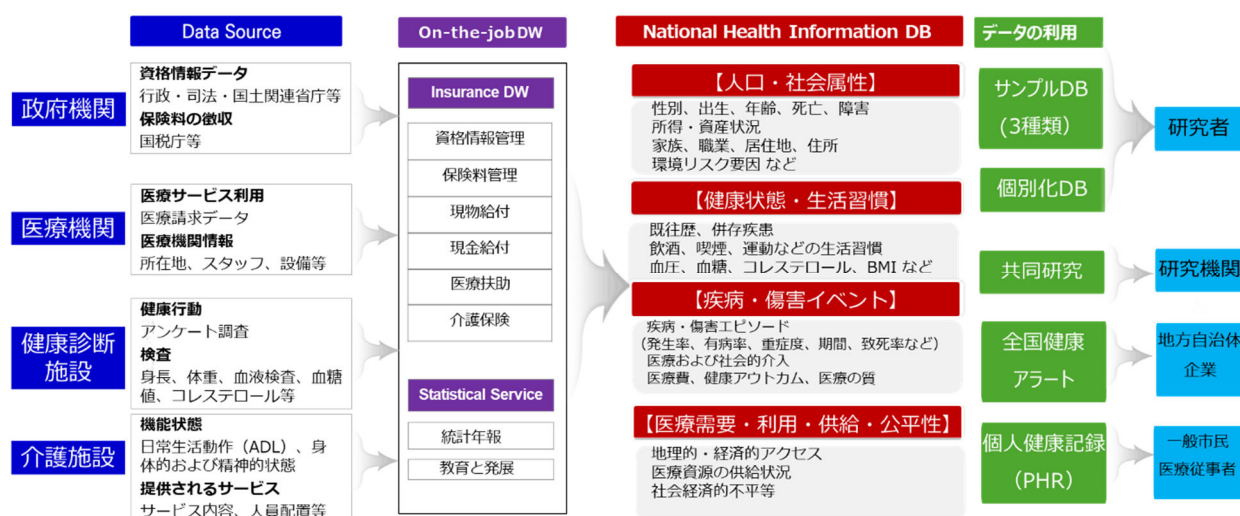


図 2. National Health Information DB (NHID) の概要

※NHIS 提供資料「Big Data Driven Evidence Based Policymaking in the Korean Context: Characteristics and Cases」および「2025 National Health Insurance & Long-Term Care Insurance System Figure3-6」を改変、統合^{1,3}

データは、研究者以外に、ヘルスケアプラットフォーム企業や臨床研究機関を含む民間企業に対しても提供されており、健康管理アプリやウェブサイト、医療機器の開発などに活用されている。データ提供件数は年間 1,000 件以上にのぼる。また、研究者向けには、研究テーマに応じて作成される個別化 DB と、国民健康情報データベースに基づく標準化サンプルコホート DB (3 種類) が整備されている。⁴データの利用は、サンプルコホート DB を除き、基本的に韓国全土にある 14 カ所の分析センターで行う。

一方、HIRA が保有するデータベースは、全国民の診療情報を基盤とし、医療機関、製薬企業、関連機関等から収集された診療行為情報、医薬品情報、医療の質評価情報、医療資源情報等を統合・整備したものである。健康保険診療費請求明細書の審査件数は、2024 年時点で約 15 億 699 万件にのぼる。NHID と同様に、個別化 DB に加えて約 100 万人規模のサンプルデータ (4 種類) が研究者向けに提供されている。データの利用は NHID と同様に全国にある分析センターにて解析するが、一部の個別化 DB に関しては研究者の自施設からリモート

にて解析が可能である。(図3)

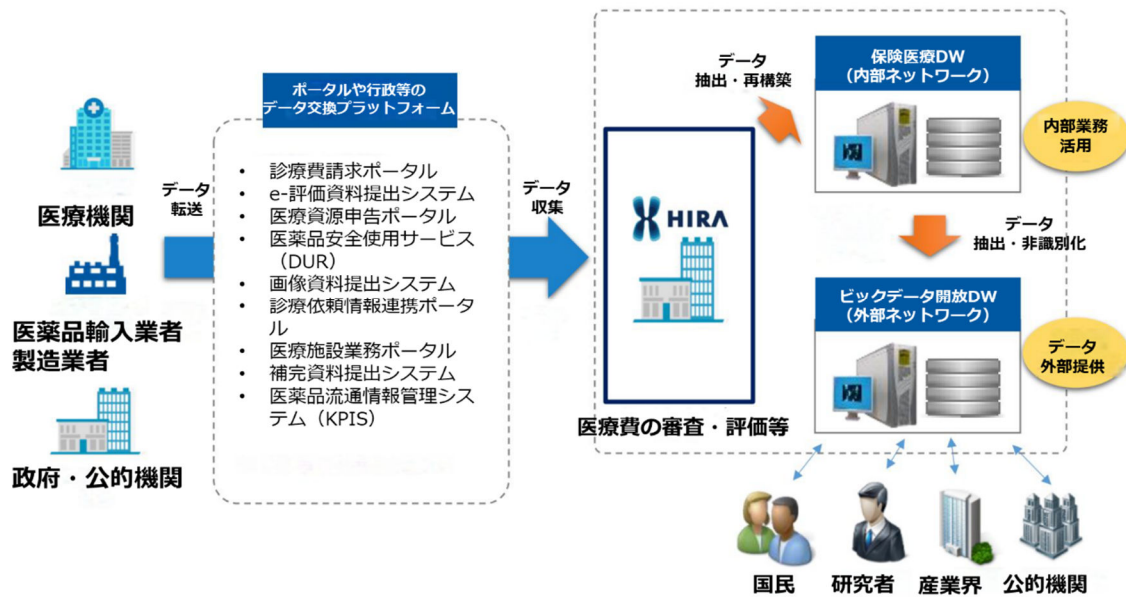


図3. HIRA の構築するデータベースの概要

※HIRA 提供資料「HIRA のビッグデータの現状と活用」を基に改変²

③ NHID と HIRA のデータの違い

NHID と HIRA のデータベースについて、研究目的に応じた適用の違いを整理する。NHID は、死亡情報、健康診断データ、所得情報、検査結果などの情報を含む点で HIRA よりも多様な情報を有しており、長期間にわたる縦断的解析が可能である。このため、NHID を用いた研究報告数は HIRA と比較して多い。⁴一方、患者の長期追跡を必要としない疾病の動向、処方傾向、

医療費に関する分析などの横断的研究では、HIRA データの利用が適していると考えられる。(図4)

また、HIRA および NHID のデータは、含まれる変数、分析センターへのアクセス方法、リモート解析の可否、利用コスト、データ提供までの期間などが異なる。そのため、研究者は研究目的や研究実施期間を踏まえ、適切なデータベースを選択することが重要となる。⁵

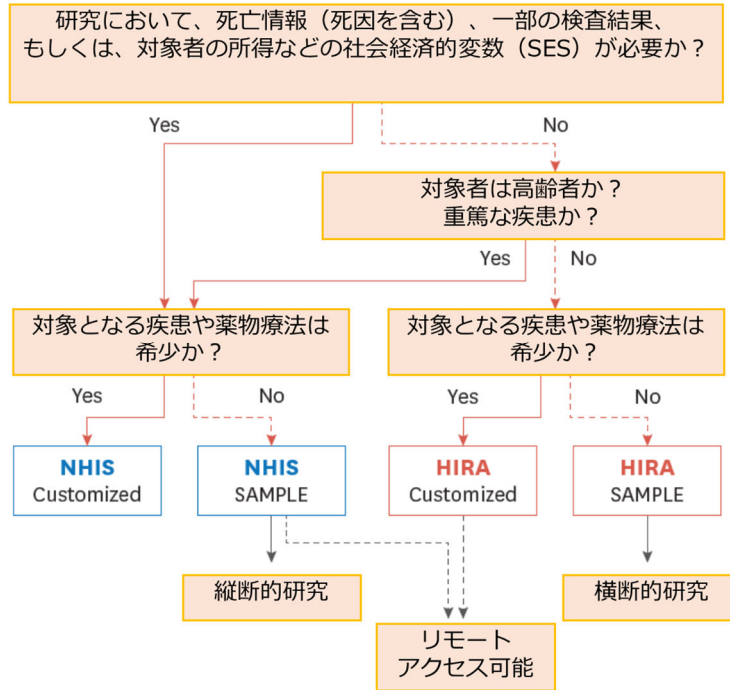


図 4. 研究目的に応じた適切な請求データの選択⁴

【フィンランドのデータベース】

① 概要

フィンランドの人口は約 550 万人(2023 年時点)である。⁶1964 年には医療保険制度の導入とともに個人識別番号制度が導入され、1968 年までにすべての国民および永住権取得者に付与された。個人識別番号制度を基盤として、1980 年代以降、プライバシー保護を前提としたレジストリ利用に関する法制度が整備された。⁷2009 年には Finnish Institute for Health and Welfare (THL: フィンランド健康福祉研究所) が設立され、保健福祉分野のレジストリの維持・管理を担う体制となった。⁸また、全国規模の電子医療記録システム (Kanta) の構築が進められ、2010 年代には患者が自身の医療情報を閲覧できる My Kanta Pages が導入された。⁹2019 年には Act on the Secondary Use of Health and Social Data に基づきデータ利用許可機関 Findata が THL 内に設立され、健康・社会

データの二次利用に関する許可手続きが一元化された。¹⁰これらの制度整備により、フィンランドでは適切な許可手続きの下でレジストリ等のデータを研究に利用することが可能となっている。

② THL (フィンランド健康福祉研究所)

THL は、専門家、政策決定者、研究者等と連携しながら、国民の健康、福祉および安全に関するモニタリングおよび研究を実施する機関である。THL は、これらの知見に基づき、健康および福祉の向上を目的とした施策や解決策の開発を行っている。¹¹また、行政機関 Ministry of Social Affairs and Health (Finland) の一つとして位置づけられている。¹²

THL は、レジストリ以外に健康・福祉に関する質問紙のデータ、バイオバンクデータ、健康診断などのさまざまな健康関連データを保有している。¹³

これらのデータの二次利用に関する許

可や管理は、Findata が担っている。Findata は社会・保健分野のデータ利用許可を一元的に管理する機関であり、複数の機関が保有するデータを研究目的で利用する際の調整を行う役割を持つ。社会保障や医療に関するデータは、Social Insurance Institution of Finland (Kela)、Finnish Centre for Pensions、Finnish Medicines Agency(Fimea)等、人口基本情報は Digital and Population Data Services Agency や Statistics Finland によって管理されている。このようにフィンランドでは、複数の公的機関が保有するデータを Findata を通じて統合的に利用できる仕組みが整備されており、連結されたデータを用いた研究が進められている。¹⁴

③ フィンデータの迅速提供の取り組み

フィンランドでは、リモート接続経由 (Kapseli) でデータにアクセス可能である。¹⁴ 2024年3月にヒアリング調査を行った際は、提供するデータは個票以外に、データセットとして、コロナ感染症データセットがあった。しかし、今回のヒアリングでは、コロナ感染症データセット以外にも個人の背景情報、がん登録、医療・薬剤・福祉関連の電子記録等を収集したレディメイドデータセットの提供を開始し、データ提供の迅速化やコスト削減を図っている。¹⁵

【フランスのデータベース】

① 概要

フランスの人口は約 6,910 万人 (2026年1月時点)である。¹⁶フランスの医療制度は、公的医療保険を基盤とする国民皆保険型であり、Protection universelle maladie (PUMa)により、フランスで安定的かつ適法に居住する者または就労する者に医療費給付の権

利が保障されている。受診にあたっては médecin traitant (かかりつけ医)を中心とした受診経路が推奨されており、公的医療保険が一定割合を償還し、残りを自己負担または補足保険で補う形をとっている。¹⁷

② SNDS と HDH

フランスでは、全国的な医療保険制度を基盤として、Système national des données de santé (SNDS) が整備されている。¹⁸これは SNIIRAM をもとに、全国民をほぼカバーする匿名化された医療行政データ基盤で、日本の NDB に類似している。SNDS には、医療保険の保険償還データ、入院データ (PMSI)、死亡関連データ

(CepiDC)、補完保険データ、障害関連データ (CNSA) などが含まれており、臨床研究や政策評価等への活用が進められている。一方で、その利用は厳格な法的枠組みの下で管理されており、定められた手続に従ってアクセスが認められる。利用できるデータの粒度としては、SNDS には日付情報が含まれるが、研究者等へデータ提供する場合は年月のみに絞られる。生年月日についても同様であるが、2歳未満の乳幼児の場合に限り、生年月日の提供が許可されている。また、居住地情報は県 (デパルタマン) 単位で提供される。SNDS の利用は学術目的では原則無料である。

SNDS の分析を補完するデータベースの1つとして BNDMR (Banque Nationale de Données Maladies Rares ; フランス国立希少疾患データバンク)がある。BNDMR は、難病・希少疾患政策の中核を担うデータ基盤であり、フランスの希少疾患患者データを全国規模で収集・統合するレジストリ型の国家データベースである。BNDMR には SNDS では把握しにくい医療行政的・

臨床的情報を含んでいる。具体的には、最初の症状が現れた年齢や診断時年齢などの診断過程に関する情報、Orpha コードによる確定診断情報、出身地や人種に関する情報、出生前診断を含むゲノム情報も含まれる。

そして Health Data Hub (HDH) が SNDS と BNDMR を連結する仕組みとして構築が進められている。

【スウェーデンのデータベース】

スウェーデンの人口は約 1,059 万人(2024 年時点)である。¹⁹ スウェーデンでは、住民登録された者に対して personal identity number (個人番号) が付与され、当該番号は生涯を通じて用いられる識別子となっている。個人番号は行政機関や民間事業者とのやり取りに広く用いられ、人口レジストリを通じて各種行政情報の基盤となっている。²⁰

保健医療・福祉分野では、National Board of Health and Welfare (Socialstyrelsen) が中央行政機関として広範な役割を担っており、患者レジストリ、がんレジストリ、死亡原因レジストリ、処方薬レジストリ、出生レジストリ等の全国レジストリを管理している。加えて、スウェーデンでは National Quality Registries と呼ばれる疾患別レジストリ群が整備されており、患者の病状、医療介入、治療後の転帰に関する個別データが日常診療の中で収集されている。²¹

研究利用に当たっては、Swedish Ethical Review Authority による倫理審査が重要な前提となる。同機関は、人を対象とする研究や生体試料、要配慮個人情報を含む研究について倫理審査を行う政府機関であり、スウェーデンにおけるレジストリ研究の基本的なガバナンスを担っている。

データは媒体提供とリモート上での提供の 2 パターンがある。研究機関の分析インフラストラクチャーがない場合も、MOVA 上でリモートアクセス環境を通じてデータ解析を行うことができる。

D. 考察

本研究では諸外国のデータ収集・利活用状況を調査した。

韓国におけるデータ提供の特徴は、NDB と比較して、検査値等の情報を収集している点 (NHID)、オンサイト形式の解析環境にて運用している点、年間約 1,000 件程度のデータ提供実績がある点などが挙げられる。一方、NDB では、通年パネルデータセット等の多様なデータセットを HIC 上で提供するなど、迅速なデータ提供に向けた取組が進められている。しかし、これらのデータセットを用いてどのような研究が実施可能であるかについては、必ずしも明確ではない。そのため、今後は文献 4 に示されているような形式で、各データセットに応じた研究の実現可能性を整理・提示することが必要かもしれない。

フィンランドの Findata においても、コスト削減や迅速提供を目的としてレディメイドデータセットの提供が開始されており、この点は NDB の取組と類似している。一方で、NDB においては、コスト低減への寄与は限定的であると考えられる。

フランスでは文献調査や BNDMR にヒアリングを行った結果を報告したが、SNDS や HDH やそれぞれのユーザーに直接ヒアリングを行うことでより理解が深まると考えられる。

スウェーデンでは個人番号制度を基盤として、行政・医療・福祉分野のレジストリが多層的に整備されており、適切な倫理審査と安全管理の下で、これらのデータを研究や医療の

質改善に活用することが可能となっている。

日本におけるオンサイトリサーチセンターや全量 NDB のように、全データを研究者自身が直接触る仕組みは確認できなかった。

利便性の高いクラウド環境の整備と安全性の確保は相反すると言われているが、各国同じ悩みを抱えていた。

来年度、これらの結果と追加調査・ヒアリングを踏まえ、NDB/HIC 解析環境の改善点やデータ提供形式、マニュアル等の在り方等について検討し、利便性向上に資する提案の取りまとめを目指す。

E. 結論

HIC を活用した多様なデータセットの提供等は、迅速なデータ提供や利便性向上に対する重要な取組である。一方で、データセットごとの研究可能性の整理や、利用者の視点に立ったユーザビリティ向上が課題として残されている。本研究で得られた海外の知見を踏まえ、NDB の利活用を一層促進するための方策を検討する必要がある。

謝辞：ヒアリングにご協力いただきました韓国 HIRA/NHIS の皆さま、Finland THL の Gissler Mika 教授、フランス BNDMR の Alban Lermine 局長と Nabila Elarouci 課長、オレブロ大学日吉綾子准教授に心より感謝申し上げます。

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

- 1) 牧戸 香詠子、明神大也、森由希子 村

松圭司 我が国およびフィンランドの医療データベースの最近の動向 第 84 回
日本公衆衛生学会総会 2025 年 10 月 30 日

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

参考：

- 1) 2025 National Health Insurance & Long-Term Care Insurance System page3,8-9
- 2) HIRA 提供資料「HIRA のビッグデータの現状と活用」
- 3) NHIS 提供資料「Big Data Driven Evidence Based Policymaking in the Korean Context: Characteristics and Cases」
- 4) Park JS, Lee CH. Clinical Study Using Healthcare Claims Database. J Rheum Dis. 2021;28(3):119-125.
- 5) Kyoung DS, Kim HS. Understanding and Utilizing Claim Data from the Korean National Health Insurance Service (NHIS) and Health Insurance Review & Assessment (HIRA) Database for Research. J Lipid Atheroscler. 2022;11(2):103-110.
- 6) 外務省 基礎データ
<https://www.mofa.go.jp/mofaj/area/finland/data.html> Accessed in March 3, 2026
- 7) Mika Gissler, Jari Haukka Finnish health and social welfare registers in epidemiological research Norsk Epidemiologi 2004; 14 (1): 113-120
- 8) Finnish Institute for Health and Welfare

- History <https://thl.fi/en/about-us/about-thl/history>
- 9) Kanta MyKanta service description
<https://www.kanta.fi/documents/20143/120102/Omakannan+palvelukuvaus+EN.pdf/b2e9ff44-27cb-0461-bf1f-4ffe661bda95?t=1740386736150>
- 10) FINDATA Legislation
<https://findata.fi/en/services-and-instructions/legislation/>
- 11) THL 提供資料「So that we all could live a healthy life」
- 12) Finnish Institute for Health and Welfare About THL <https://thl.fi/en/about-us/about-thl>
- 13) THL 提供資料「Health and social welfare information system in Finland」
- 14) FINDATA Kapseli
<https://findata.fi/en/kapseli/>
- 15) FINDATA Data
<https://findata.fi/en/data/>
- 16) 外務省 基礎データ <https://www.mofa.go.jp/mofaj/area/france/data.html> Accessed in May 15, 2026
- 17) l'Assurance Maladie
<https://www.ameli.fr/assure/droits-demarches/principes/protection-universelle-maladie> Accessed in May 15, 2026
- 18) Health Data Hub
<https://www.health-data-hub.fr/snds> Accessed in May 15, 2026
- 19) 外務省 基礎データ
<https://www.mofa.go.jp/mofaj/area/sweden/data.html> Accessed in May 15, 2026
- 20) Skatteverket
<https://www.skatteverket.se/privat/folkbokforing/personnummer.4.3810a01c150939e893f18c29.html>
- 21) Socialstyrelsen
<https://www.socialstyrelsen.se/en/statistics-and-data/registers/>

Ⅲ 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
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
松居宏樹	匿名医療保険等関連情報データベースの利用経験および今後の期待と課題	統計	76巻 8号	14-22	2025

倫理審査等報告書の写し

令和8年2月13日

厚生労働大臣 殿

機関名 浜松医科大学

所属研究機関長 職名 学長

氏名 渡邊裕司

次の職員の令和7年度厚生労働行政推進調査事業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 政策科学総合研究事業（政策科学推進研究事業）
2. 研究課題名 NDBのユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究
3. 研究者名（所属部署・職名） 健康社会医学講座 准教授
 （氏名・フリガナ） 明神大也（ミヨウジン トモヤ）

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）	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<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。

厚生労働大臣 殿

機関名 国立大学法人千葉大学

所属研究機関長 職 名 学長

氏 名 横手 幸太郎

次の職員の令和7年度厚生労働行政推進調査事業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 政策科学総合研究事業（政策科学推進研究事業）
2. 研究課題名 NDBのユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究(25AA2002)
3. 研究者名 (所属部署・職名) 医学部附属病院 特任教授
(氏名・フリガナ) 村松 圭司・ムラマツ ケイジ

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)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<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。

厚生労働大臣 殿

機関名 国立大学法人 京都大学

所属研究機関長 職 名 学長

氏 名 湊 長博

次の職員の令和7年度厚生労働行政推進調査事業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 政策科学総合研究事業（政策科学推進研究事業）

2. 研究課題名 NDBのユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究

3. 研究者名 (所属部署・職名) 医療情報企画部 准教授

(氏名・フリガナ) 森 由希子 ・モリ ユキコ

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)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<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。

厚生労働大臣 殿

機関名 国立大学法人東京大学

所属研究機関長 職 名 学長

氏 名 藤井 輝夫

次の職員の令和7年度厚生労働行政推進調査事業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 政策科学総合研究事業（政策科学推進研究事業）
2. 研究課題名 NDBのユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究
3. 研究者名 （所属部署・職名）大学院医学系研究科生物統計情報学講座・特任助教
（氏名・フリガナ）牧戸 香詠子・マキト カナコ

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）	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/> 未審査の理由：調査研究主体で、NDBデータの利用に関しても倫理審査不要のデータを用いる研究であるため倫理審査不要
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 <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。

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 国土典宏

次の職員の令和7年度厚生労働行政推進調査事業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 政策科学総合研究事業（政策科学推進研究事業）
2. 研究課題名 NDBのユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究
3. 研究者名 （所属部署・職名）国立国際医療研究センター・医療政策研究室長
（氏名・フリガナ）杉山雄大・スギヤマタケヒロ

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）	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<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。

厚生労働大臣 殿

機関名 公立大学法人奈良県立医科大学

所属研究機関長 職 名 理事長

氏 名 細井 裕司

次の職員の令和7年度厚生労働行政推進調査事業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 政策科学総合研究事業（政策科学推進研究事業）
2. 研究課題名 NDB のユーザビリティ向上を通じてクラウド上での データ二次利用を推進するための研究
3. 研究者名 (所属部署・職名) 公衆衛生学講座・助教
(氏名・フリガナ) 西岡 祐一・ニシオカ ユウイチ

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)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<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。

厚生労働大臣 殿

機関名 国立大学法人東京大学

所属研究機関長 職 名 学長

氏 名 藤井 輝夫

次の職員の令和7年度厚生労働行政推進調査事業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 政策科学総合研究事業（政策科学推進研究事業）

2. 研究課題名 NDBのユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究

3. 研究者名 （所属部署・職名）大学院医学系研究科・准教授

（氏名・フリガナ）松居 宏樹・マツイ ヒロキ

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）	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	東京大学	<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 <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。

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年度厚生労働行政推進調査事業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 政策科学総合研究事業（政策科学推進研究事業）
2. 研究課題名 NDBのユーザビリティ向上を通じてクラウド上でのデータ二次利用を推進するための研究
3. 研究者名 (所属部署・職名) 国立健康危機管理研究機構・国立看護大学校・准教授
(氏名・フリガナ) 柏木公一・カシワギキミカズ

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)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<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。