

令和5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）

臨床疫学に活用可能な NDB 等データセットの作成に関する研究

(21AA2006)

令和5年度 総括・分担研究報告書

研究代表者 森 由希子

(京都大学医学部附属病院 医療情報企画部)

令和6(2024)年 5月

令和5年度報告書 目次)

目 次

I. 総括研究報告

臨床疫学に活用可能な NDB 等データセットの作成に関する研究	-----	1
森由希子、加藤源太、黒田知宏、植嶋大晃、大寺祥佑、今村知明、野田龍也、 康永秀生、田宮菜奈子、杉山雄大、中山健夫		

II. 分担研究報告

1. NDB・介護 DB 連結データセットの仕様検討	-----	5
森由希子、加藤源太、黒田知宏、植嶋大晃、大寺祥佑、中山健夫		
2. 長期追跡性に優れた個人名寄せ ID の検討	-----	17
野田龍也、今村知明、明神大也		
3. NDB オンサイトリサーチセンターと比較した HIC	-----	27
松居宏樹、康永秀生		
4. HIC 試行利用における課題検討	-----	39
森由希子、加藤源太、黒田知宏、植嶋大晃、大寺祥佑、田宮菜奈子、杉山雄 大、小宮山潤		

III. 研究成果の刊行に関する一覧表	-----	44
---------------------	-------	----

令和5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))

臨床疫学に活用可能なNDB等データセットの作成に関する研究に関する研究
統括研究報告書

研究代表者	森 由希子	京都大学医学部附属病院医療情報企画部 講師
研究分担者	加藤 源太	京都大学医学部附属病院診療報酬センター 特定教授
研究分担者	黒田 知宏	京都大学医学部附属病院医療情報企画部 教授
研究分担者	植嶋 大晃	京都大学国際高等教育院附属データ科学イノベーション教育研究センター 特定講師
研究分担者	大寺 祥佑	国立研究開発法人 国立長寿医療研究センター 老年学・社会科学研究センター 医療経済研究部 副部長
研究分担者	今村 知明	奈良県立医科大学公衆衛生学講座 教授
研究分担者	野田 龍也	奈良県立医科大学公衆衛生学講座 准教授
研究分担者	康永 秀生	東京大学大学院医学研究科 教授
研究分担者	田宮 菜奈子	筑波大学医学医療系/ヘルスサービス開発研究センター 教授
研究分担者	杉山 雄大	国立国際医療研究センター 医療政策研究室長
研究分担者	中山 健夫	京都大学大学院医学研究科 教授

背景：近年、社会情勢や人口構造の変化に伴い、健康・医療・介護分野のビッグデータ解析の必要性が高まっている。匿名医療保険等関連情報データベース（NDB）及び介護保険総合データベース（介護DB）のデータについても利活用が期待されているが、データの容量が膨大であること、データの構造が複雑であること等から、データの構造や操作を知悉した研究者でなければ解析を行うのは容易ではない。NDBや介護DB等の大規模データを簡便に分析できるようになれば、健康・医療・介護ビッグデータの利活用推進に貢献することが期待される。

目的：「健康・医療・介護分野の大規模データの利活用を推進する」ことを目的に、NDB、および介護DB、さらに令和2年から提供が開始されている医療・介護の連結情報から、より利用者にとって使いやすいデータセットを開発する。また、医療・介護データ等の解析基盤（HIC）の試行利用を行い、安全性、操作性の検証を行う。

方法：本研究では、利用者によるデータセット設計を補助するために「①既存の大規模データベース（NDB・介護DB）のデータ構造の評価」を実施する。また、利用者にとって使いやすいデータセットの設計のために「②長期追跡性に優れた個人名寄せIDの検討」および「③臨床系研究者でも利用可能なデータセットの開発」について検討を行う。②および③については、研究の一部をHIC上で実施し、④HICの安全性と操作性について検討を行う。2023年度においては

- ① については、介護DBのデータ項目、データの格納状況について集計を実施し、結果を公表した。
- ② については、NDBに含まれる個人単位被保険者番号に基づく個人名寄せID（id5）が付与されたデータの長期追跡性について検討及び評価を行った。さらに、id5が付与される前のデータについても、既存のidとのより精緻なデータ連結の可能性について検討を行った。
- ③ については、今年度はNDBと介護DBの連結データの分析練習に使用可能なサンプルデータセットの仕様検討を行った。
- ④ については、実際のHICの利用を実施し、オンサイトリサーチセンターでのNDBデータ利用との比較検討を行った。

結果：

- ① 介護DBのデータについて網羅的な調査を行い、調査結果として作成したコードブックを学会およびホームページにおいて公表した。
- ② id5の設定状況を把握したとともに、id0およびid5の長期追跡の妥当性評価を行い、id0とid5を組み合わせた新たなid0nの開発に踏み込んだ。
- ③ NDBと介護DBの連結データの分析練習に使用可能なサンプルデータセットの開発を行った。
- ④ HIC利用における課題について、オンサイトリサーチセンター利用との比較およびHICにおいて比較的大規模な特別抽出データを取り扱う際の課題についての検討を行った。

結論： NDBにおける長期追跡可能なidに関する評価を行った。またNDB・介護DB連結データの練習用データセットを開発した。さらにHICにおけるデータ分析に関する課題抽出および解決法の検討を行った。これらの成果は今後利用者支援の一環となることが期待される。

A. 研究目的

本研究は「健康・医療・介護分野の大規模データの利活用を推進する」ことを目的に、匿名医療保険等関連情報データベース（NDB）データ及び介護保険総合データベース（介護DB）データ、さらに令和2年から提供が開始されている医療・介護の連結情報から、より利用者にとって使いやすいデータセットの開発を目指すものである。

昨今のコロナ禍に伴う社会状況や、近年の急速な少子高齢化を背景とする社会構造の変化に伴い、より適正な医療を提供するための基礎資料として、健康・医療・介護分野のビッグデータ解析の必要性が高まっている。実際NDB及び介護DBの情報はずでに医療・介護それぞれの分野での政策利用や研究利用が開始されており、提供件数は徐々に増加している。一方データの容量が膨大であること、データの構造が複雑であること等から、これらデータの構造や操作を知悉した研究者でなければ解析を行うのは容易ではなく、医療者等いわゆる臨床系の研究者だけでは取り扱いが困難なデータとなっている。一方で臨床系の研究者は、日々医療介護の現場で医療を担っており、適正な医療の提供・実施にあたっての課題にも直面していることから、こうした研究者が自らNDBや介護DB等の大規模データを簡便に分析できるようにすれば、それら課題の解決に貢献することが期待される。

加えて、NDBは令和2年10月から介護DB及び令和4年4月からDPCデータベースとの連結が可能となり、さらに「医療・介護データ等の解析基盤（HIC）」が稼働されていることから、将来的には現在よりも利用の利便性が向上する見込みである。NDBデータを広く研究者が利用できるようにするためには、より簡便にデータ分析できる環境が整備される必要がある。

以上のような背景から、本研究班では、

①既存の大規模データベースの（NDB・介護DB）のデータ構造の評価

②長期追跡性に優れた個人名寄せIDの検証

③臨床系研究者でも利用可能なデータセットの開発の3つの課題について検討を行うことを目的とした。また、②および③の一部をHIC上で実施（試行利用）することにより、④HICの安全性及び操作性の検証を行った。

B. 研究方法（詳細については各分担研究報告書に記載）

本研究は、NDBと介護DBさらに医療・介護の連結情報の利活用の推進を目的に、より多くの研究者が利用可能なデータセットの開発を行うものである。このため前述の3つの課題について令和5年度は以下のような方法で検討を行った。

①既存の大規模データベースの（NDB・介護DB）のデータ構造の評価

介護DBデータを用いて、データ構造、テーブル構造、データ形式等の評価を行い、集計した情報について介護DBコードブックを作成し、公表した。

②長期追跡性に優れた個人名寄せIDの検証

NDBに含まれる個人単位被保険者番号に基づく個人名寄せID（id5）が付与されたデータの長期追跡性について検討及び評価を行った。

③臨床系研究者でも利用可能なデータセットの開発
NDBと介護DBはそれぞれ特有のデータ項目や構造があるため、連結データの分析にはそれらを理解した上での操作が必要となる。連結データ分析の練習を想定したサンプルデータセットの仕様検討を行い、作成した

⑤ HICの安全性及び操作性の検証

オンサイトリサーチセンターとHICにおけるNDBデータ利用について比較検討を行った。さらに特別抽出データを分析する上での課題について検討を行った。

C. 研究結果（詳細については各分担研究報告書に記載）

①介護DBデータのコードブックの作成を行った。
NDBとの連結データの利活用に資することを念頭に、NDBデータのコードブックの仕様をひな型とし、各項目の要約統計量（最大値、最小値、平均値、標準偏差など）による粗集計及び項目値のサンプル（最大値から上位50位までの項目値の実例）を算出する方針とし、集計可能なデータ項目、データ内容の精査を行い、実際の集計を行った。集計結果は第82回公衆衛生学会で報告し、京都大学医学部附属病院医療情報企画部ホームページ（<https://medinfo.kuhp.kyoto-u.ac.jp/document/kaigo-db-codebook/>）に公開した。

②id5の設定状況を調査した。全レセプトに含まれる有効なid5の割合はDPCレセプトでは95%程度、医科入院レセプトでは92%程度、医科入院外レセプトでは97%程度であった。またid0とid5から新たにid0nを作成した。2022年4月に存在したid0、id0n、id5に対し、2023年3月まで存在するidを追跡した結果、id0、id0n、id5いずれも、9ヶ月後で約90%の捕捉率を示し、id0nが最も捕捉率が高かった。

③。NDBと介護DBの連結データの分析練習に使用可能なサンプルデータセットの開発を行った。また、ユースケースの検討を行った。

④2023年2月からHICの試行利用を開始し、データの解析環境への取り込み、データベース構築を開始した。HICの構造およびデータアクセスとセキュリティ、解析環境の構築と大規模データの取扱いについて検討を行った。また、オンサイトリサーチセンターでの利用との比較についても検討を行った。

D. 考察

介護DBに格納されているデータに関する基礎データを集計し、コードブックを作成した。NDBデータと同様に介護DBデータについても、コードブックを作成、公開することにより、今後介護DBデータに関する利用者の理解が深まり、データ利活用の一助となることが期待される。

長期追跡可能なIDの検討においては、NDBデータにおいてはすでにid0を用いた1患者1データ化の手法が開発されており、NDBデータ内における追跡調査・コホート調査が可能となっている（2021年度報告書にて報告）。今年度はid5の設定状況を把握したとともに、id0およびid5の長期追跡の妥当性評価を行い、id0とid5を組み合わせた新たなid0nの開発に踏み込んだ。この取り組みにより、すでに蓄積されている膨大な過去データとの連結解析が可能になり、様々な課題に対するデータの有効利用が期待される。

データセットの検討においては、研究者のニーズ

に対応できるデータセット作成に必要な条件を同定するために、2021年度、2022年度において実際にいくつかのリサーチクエストionsについてNDBデータおよび自治体医療介護データを用いた分析を実施した（報告済み）。その結果、学術研究に資するようなデータセットの作成には、研究内容に即したデータセットの設計が必要であり、レディメイドのデータセットでは研究内容に制限がある可能性が示唆された。一方で、データセットの基本構造（エンティティ定義）等については共有できる可能性が改めて示唆された。今年度はいままでの検討結果をふまえて、NDB・介護DB連結データの分析練習ができるデータセットの開発を行った。簡便に使用可能なサンプルデータセットにより、連結データ利用に対するハードルが下がり、利用推進が期待される。

HICの試行利用においては、実際のデータを用いて、HIC上に分析用データベース構築を行い、HIC利用における課題検討を行った。また、HICとオンサイトリサーチセンターとの比較検討を行い、その特性の違いを説明した。

E. 結論

- ①介護DBデータのコードブックは介護DBデータ及びNDB-介護DB連結データ利活用推進の一助となることが期待される。
- ②新規idであるid5のNDBにおける設定状況や長期追跡性を把握するとともにすでに開発されているid0とid5を組み合わせた新たなid0nの開発に踏み込んだ。
- ③「汎用性のあるデータセット」の一つとして、NDB・介護DB連結データの練習用データセットを作成した。
- ④HICにおけるデータ利用の課題と対応策について検討を行った。

F. 健康危険情報 なし

G. 研究発表（別添4および各分担研究報告書に記載）

1. 論文発表 なし
2. 学会発表 あり

H. 知的財産権の出願・登録状況 なし

1. 特許取得
なし
2. 実用新案登録
なし
3. その他
なし

令和4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)) 分担研究報告書

臨床疫学に活用可能なNDB等データセットの作成に関する研究に関する研究
NDB・介護DB連結データセット仕様の検討

研究代表者	森 由希子	京都大学医学部附属病院医療情報企画部 講師
研究分担者	加藤 源太	京都大学医学部附属病院診療報酬センター 特定教授
研究分担者	黒田 知宏	京都大学医学部附属病院医療情報企画部 教授
研究分担者	植嶋 大晃	京都大学国際高等教育院附属データ科学イノベーション 教育研究センター 特定講師
研究分担者	中山 健夫	京都大学大学院医学研究科 社会健康医学系専攻 健康情報学分野 教授
研究分担者	大寺 祥佑	国立研究開発法人 国立長寿医療研究センター 老年学・社会科学研究センター 医療経済研究部 副部長

研究要旨

1. 背景：匿名医療保険等関連情報データベース (NDB)は、「高齢者の医療の確保に関する法律」に基づき、2009年から診療報酬請求書 (レセプト) 情報および特定健康診査・特定保健指導情報を収集している世界でも有数の規模と悉皆性を誇るデータベースである。また、介護保険総合データベース (介護DB) は「介護保険法」に基づき、介護レセプト及び要介護情報を収集しているデータベースであり、2013年から運用が開始されている。2020年度からNDBと介護DBの連結データの第三者提供が開始されたが、データを有効に活用するにはそれぞれのデータベースに格納されている情報の内容や特性について十分理解する必要がある。NDBについてはコードブックが作成、公開されており、NDB利用者、研究者に活用されているが、介護DBのデータについては同様のコードブックが存在せず、今後の介護DBデータおよび連結データの利活用推進には介護DBデータの詳細な情報の共有が必要である。
2. NDB・介護DBの連結データの分析において、それぞれのDBの保有するデータ項目やデータの特性、データ間の関連性の理解に時間を要し、これらのことがデータ利活用推進の壁となっている。データ項目についてはNDBコードブック、介護DBコードブックの作成により、一定の成果を得たと考えるが、実際の分析を行うにあたり、ある程度一般的な分析手法に対応可能な練習用データセットが存在すれば、NDB・介護DB連結データ利用に対するハードルが下がり、利用者の拡大に有用と考える。

目的：

1. 本研究の目的は、介護DBのデータの構造や内容について調査し、研究者等の利活用に有用なコードブックを作成することである。
2. 本研究の目的は、NDB・介護DB連結データを初めて利用する研究者 (医療者や行政関係者含む) を想定し、比較的一般的な分析手法の練習に利用可能な汎用性のあるデータセットの仕様検討を行うことである。

方法：

1. 本研究では、2014年4月から2016年3月までの介護DBに含まれる介護レセプト・要介護情報について、格納 (提供) されているデータの項目、データ構造、データ数等について検討した。
2. 本研究ではNDB・介護DB連結データにおける分析手法の練習に利用可能な「汎用性のあるデータセット」について、仕様検討を行い、実際に作成を行った。

結果：

1. 介護DBのデータ項目、格納されているデータの概要についての調査・集計を行い、各項目に入力されているデータのコード内容や件数を明らかにした。また、データの格納状況についても明らかにした。調査・集計結果を第82回公衆衛生学会および京都大学医学部附属病院医療情報企画部ホームページ(<https://medinfo.kuhp.kyoto-u.ac.jp/document/kaigo-db-codebook/>)において公表した。

2. 連結データ分析練習用データセットの設計を行った。連結データ分析に必要なデータ項目の検討を実施し、サンプルデータセットを作成した。

結論：

1. 本研究では、介護DB格納されているデータのデータ構造やデータ項目の詳細を明らかにし、コードブックを作成した。
2. 連結データ分析練習用データセットの設計を行い、作成されたデータセットを用いてできる分析（実習）内容の検討を進めた。

A.

研究目的

1. 介護DBデータブック作成

社会情勢や人口構造の変化に伴い政策立案、行政・民間サービス提供等、様々な場面、状況において健康医療介護データの利活用のニーズは高まっている。特に日本においては、急速な人口の高齢化に伴い、安定した医療、介護の提供体制の確立・維持が、喫緊の課題であり、医療・介護現場におけるリアルワールドデータの利活用とデータに基づく施策やサービス提案が必要とされている。

厚生労働省が保有する匿名医療保険等関連情報データベース（NDB）には「高齢者の医療の確保に関する法律」に基づき、保険診療における診療報酬請求書（レセプト）の情報が2009年から収集、蓄積されており、2011年から研究者等に対する第三者提供が開始されている。また、介護保険総合データベース（介護DB）には

「介護保険法」に基づき、介護レセプト及び要介護情報が収集、格納されており、2018年度から研究者等に対する第三者提供が開始されている。これらのデータの利活用に対するニーズは高く、データ利用申出は年々増加しており、NDBのデータ利用申出は2023年3月の時点で453件が承諾されている。また、共通のIDが付与された、NDBと介護DBの連結データの提供も2020年度から開始されている。一方でこれらのデータの利用者からは、二次利用データであるが故のデータ分析の困難さが以前から問題視されている。NDBに格納されているレセプトデータは本来、診療報酬請求のために作成、提出されたデータのため、必ずしも分析に適した形式・構造となっておらず、膨大なデータ量と使用されているコードの複雑さもあいまって、データ提供を受けてもすぐに分析を解析することは難しい状態である。

この状態を改善するために奈良県立医科大学公衆衛生学講座において、NDBに格納されているデータについて医科の全項目を対象に、各項目にどのような値がセットされているかについて、要約統計量と項目値サンプルを集計したコードブックが作成され、ホームページ上に公開された。（奈良県立医科大学 今村知明 野田龍也

2018 https://www.naramed-u.ac.jp/~hpm/res_document_ndbcodebook.html）この中において、データ分析の障害となる、NDBに項目としては存在しているものの、非公表や未収集等の理由によりすべて欠損値となっている項目や項目名からは想定されづらい数値が入っている項目等についても明らかにされており、NDBの全体像を把握し、特定部分を参照する辞書のような役割

として、NDBの概要を把握したい研究者やNDBの利用申請において抽出依頼項目の選定を行う研究者、データ到着後に分析結果の妥当性を検討する研究者を中心に利用されている。

一方、2020年から介護DBのデータ提供が開始されたが、介護DBについては、同様の“辞書”が作成されていないため、以前のNDBデータと同様に、利用者はデータが手元に来てから、データの精査に着手する必要がある、意味ある分析にたどり着くまでに大変な手間を要している。介護DBには要介護認定情報と介護レセプト等情報が収集されており、要介護認定情報としては要介護認定一次判定、基本調査74項目、主治医意見書のうち、短期記憶、認知能力、伝達能力、食事行為、認知症高齢者の日常生活自立度の項目、要介護認定等基準時間、一次判定結果および要介護認定二次判定（認定有効期間・二次判定結果等）等の情報が格納されている。

（<https://www.mhlw.go.jp/content/12301000/000922065.pdf>）また、介護レセプト情報としては性別、生年月、要介護状態区分、認定有効期間等の属性情報や提供されたサービスの種類や単位数、日数、回数等の情報が格納されている。これらの情報は医療レセプトには含まれていない患者の“状態”や“転帰”に係る情報が含まれており、NDBデータと連結することにより、今までNDBデータだけでは実施できなかった分析が可能となることが期待される。

今後介護DBデータやNDBとの連結データの利活用を推進するためには、介護DBデータについてもNDBと同様のコードブックの整備が必要と考える。このため本研究では第三者提供において研究者に提供される介護DBデータについてコードブックの作成を行うことを目的とした。

2. NDB・介護DB連結データセット仕様検討

前述の通りNDB及び介護DBの情報はすでに医療・介護それぞれの分野での政策利用や研究利用が開始されており、提供件数は徐々に増加している。一方データの容量が膨大であること、データの構造が複雑であること等から、これらデータの構造や操作を知悉した研究者でなければ解析を行うのは容易ではなく、医療者等いわゆる臨床系の研究者だけでは取り扱いが困難なデータとなっている。NDB・介護DBの連結データの分析においてはさらにそれぞれのDBの保有するデータ項目やデータの特性、データ間の関連性の理解に時間を要し、これらのことがデータ利活用推進の壁となっている。一方で臨床系の研究者は、日々医療介護の現場で医療を担っており、適正な医療の提供・実施にあたっての課題にも直面していることから、こうした

研究者が自らNDBや介護DB等の大規模データを簡便に分析できるようになれば、それら課題の解決に貢献することが期待される。このため、NDB・介護DB連結データの分析練習が行える簡便かつ安全なサンプルデータセット仕様検討を行い、作成することを目的とした。

B. 研究方法

1. 介護DBデータブック作成

本研究では、2014年4月から2016年3月までの介護DBデータを用いた。コードブック作成用のDWHを構築し分析・集計作業を行った。コードブックの仕様はNDBデータのコードブックをひな型とし、各項目の値を連続値とみなした要約統計量（最大値、最小値、平均値、標準偏差など）による粗集計及び各項目に含まれる値の度数分布（最大値から上位50位までの項目値の実例）を算出する方針とし、集計可能なデータ項目、データ内容の精査を行い、実際の集計を行った。

2. NDB・介護DBデータセット作成

本研究では、まずデータセットの利用方法について

- 1) 対象疾患の検討
- 2) 対象レコードの検討
- 3) 対象項目の検討
- 4) 実施可能な研究内容を

を中心に検討を行った。その後、以下の手法でNDB・介護DB連結データセットを作成した

- 1) 本研究用の介護DBデータマート（介護DBの認定情報のデータマート）を作成
- 2) 共通IDを保有していないNDBデータと介護DBデータに仮想の共通ID（ID0）を付与して連結
- 3) ID0で連結したデータからデータセットを作成
- 4) 作成したデータセットからサンプリング

C. 研究結果

1. 介護DBコードブック作成

NDBとの連結データの利活用に資することを念頭に、すでに公表されているNDBデータコードブックと同様に使用ができることを目指して、介護DBコードブックの作成を行った。集計結果については第82回公衆衛生学会で発表し、京都大学医学部附属病院医療情報企画部ホームページで公表を行った。

（表1①-⑦ データが膨大なため、本報告書には一部を添付）

2. NDB・介護DBデータセットの仕様検討

本研究では共通IDを保有していないNDBデータと介護DBデータ（連結不可能データ）を使用して、以下の手法でNDB・介護DB連結データ

セットを作成した。

- 1) 共通IDを保有していないNDBデータと介護DBデータに仮想の共通ID（ID0）を付与して連結（この際、年齢層、性別についてはマッチングを実施）

【傷病毎の要介護認定集計用】

- 2) NDBデータ（REファイル（外来）、SYファイル、マスター）を対象に以下のデータを抽出
 - ・ ID0
 - ・ 診療年月
 - ・ 年齢階級コード
 - ・ 傷病コード（主傷病名）
- 3) 傷病グループ（高血圧、虚血性心疾患、脳卒中、該当なし）に分類
- 4) 個人匿名性を下げるため、100IDをランダムサンプリング
- 5) 本研究用の介護DBデータマートのNIデータと2) 3) で作成したデータマートを診療年月とID0で連結し、年月ごと、患者ごとの連結データマートを作成（図1）
- 6) 傷病名グループ毎に集計用テーブルを作成（図2）

【介護施設入所者（特養/老健）の入院有無、経管栄養有無の集計用】

- 7) 介護DBのD1ファイルから介護老人福祉施設サービスを利用したID0を抽出（対象者情報）
- 8) 個人匿名性を下げるため、100IDをランダムサンプリング
- 9) 介護DBのH1ファイルから対象者の入所年月日を抽出
- 10) NDB医科REファイル、DPCBUファイルから対象者の入院年月日を抽出
- 11) 9)と10)をID0で連結
- 12) 介護DBのNIファイルおよびNINTEIファイルから要介護度と経管栄養有無の情報を取得しID0で10)と連結（図3）
- 13) 要介護度、経管栄養の有無毎に集計用テーブルを作成
- 14) 作成されたデータセットで実施可能な分析内容について検討を行った。

E. 結論

本研究で行った、第三者提供で提供される介護DBデータのデータ格納形式、データ項目、データ内容等のコードブック作成により、介護DBデータを利用する際に利用者が多大な労力を要していた、データ構造の理解、データクレンジングに要する時間を短縮し、データ利用におけるハードルが下がることが期待され、介護DBの行政利用、研究利用の推進の一助となると考えられる。また、NDB、介護DBのコードブックが整備されることにより連結データの利活用推進にも有用であると考ええる。

NDB・介護DB連結データの汎用性のあるデータセットについて、対象疾患、データ項目お

表1 介護DBコードブック (R4年度報告書より再掲)

① 要介護認定情報

項番	データ項目名 (日本語)	データ項目名 (英名)	型	カラムの型	カラムの長さ	最大値	最小値	平均値	標準偏差	欠損値の数 (概数)	欠損割合	テーブル番号	カラム番号
1	識別コード	SHIKIBETSU_CD	文字	VARCHAR	3	—	—	—	—	0	0%	1	1
2	保険者番号	HOKEN_NO	文字	VARCHAR	6	—	—	—	—	0	0%	1	2
3	認定申請日	SHINSEI_DATE	文字	VARCHAR	8	20180925	20090531	20140994.89	9834.95	0	0%	1	3
4	枝番	SUBKEY	数字	DECIMAL	1	8	0	0.33	0.51	0	0%	1	4
5	申請区分(法令)コード	SHINSEI_CD	文字	VARCHAR	1	4	1	1.78	0.57	0	0%	1	5
6	申請区分(申請時)コード	SHINSEI_S_CD	文字	VARCHAR	1	6	1	1.90	0.58	14230	0%	1	6
7	取下区分コード	TORISAGE_CD	文字	VARCHAR	1	4	1	1.01	0.19	0	0%	1	7
8	被保険者区分コード	HIHOKEN_CD	文字	VARCHAR	1	2	1	1.03	0.16	0	0%	1	8
9	申請代行区分コード	DAIKOU_CD	文字	VARCHAR	1	9	1	3.94	3.01	0	0%	1	9
10	年齢階級コード	NENREI_CD	文字	VARCHAR	1	8	1	5.08	1.61	0	0%	1	10
11	性別コード	SEIBETSU_CD	文字	VARCHAR	1	2	1	1.67	0.47	0	0%	1	11
12	意見書依頼日	IK_IRAI_DATE	文字	VARCHAR	8	20180925	19160415	20141581.89	9704.31	1269730	6%	1	12
13	意見書入手日	IK_NYUSHU_DATE	文字	VARCHAR	8	74250718	19490326	20141632.93	28307.24	746860	4%	1	13
14	意見書「短期記憶」	IK_TANKI_CD	文字	VARCHAR	1	9	1	1.62	0.64	63280	0%	1	14
15	意見書「認知能力」	IK_NINCHI_CD	文字	VARCHAR	1	9	1	2.09	1.10	63280	0%	1	15
16	意見書「伝達能力」	IK_DENTATSU_CD	文字	VARCHAR	1	9	1	1.88	1.04	63280	0%	1	16
17	意見書「食行為」	IK_SHOKUJI_CD	文字	VARCHAR	1	9	1	1.12	0.58	63280	0%	1	17
18	意見書「認知症高齢者の日常生活自立度」	IK_JIRITSU_CD	文字	VARCHAR	1	9	1	3.31	2.03	63580	0%	1	18
19	調査依頼日	CH_IRAI_DATE	文字	VARCHAR	8	20620703	20100401	20141888.98	9353.00	1574850	8%	1	19
20	調査実施日	CH_JISSHI_DATE	文字	VARCHAR	8	20780428	20100401	20141174.52	9948.35	87910	0%	1	20
21	委託区分	CH_ITAKU_CD	文字	VARCHAR	1	9	1	1.88	1.78	16776830	91%	1	21
22	指定居宅介護支援事業者番号	CH_JIGYOSYA_NO	文字	VARCHAR	10	—	—	—	—	1586610	8%	1	22
23	認定調査員資格コード	CH_SHIKAKU_CD	文字	VARCHAR	2	99	1	34.06	38.03	14442660	78%	1	23
24	一次判定結果	H_ICHIJI_CD	文字	VARCHAR	2	25	1	19.12	5.48	93710	0%	1	24
25	一次判定結果(認知症加算)	H_ICHIJI_KASAN_CD	文字	VARCHAR	2	25	1	19.15	5.50	93710	0%	1	25
26	要介護認定等基準時間	H_KJN_TOTAL	数字	DECIMAL	4	2253	0	596.76	317.53	93710	0%	1	26
27	要介護認定等基準時間(食事)	H_KJN_SHOKUJI	数字	DECIMAL	4	714	0	88.09	105.85	93710	0%	1	27
28	要介護認定等基準時間(排泄)	H_KJN_HAISETSU	数字	DECIMAL	4	280	0	84.11	89.56	93710	0%	1	28
29	要介護認定等基準時間(移動)	H_KJN_IDOU	数字	DECIMAL	4	214	0	69.91	66.64	93710	0%	1	29
30	要介護認定等基準時間(清潔保持)	H_KJN_SEIKETSU	数字	DECIMAL	4	243	0	93.23	59.23	93710	0%	1	30
31	要介護認定等基準時間(間接ケア)	H_KJN_KANSETSU	数字	DECIMAL	4	113	0	64.17	30.38	93710	0%	1	31
32	要介護認定等基準時間(BPSD関連)	H_KJN_BPSD	数字	DECIMAL	4	212	0	67.94	22.35	93710	0%	1	32
33	要介護認定等基準時間(機能訓練)	H_KJN_KINOU	数字	DECIMAL	4	154	0	63.03	28.57	93710	0%	1	33
34	要介護認定等基準時間(医療関連)	H_KJN_IRYOU	数字	DECIMAL	4	985	0	66.25	80.02	93710	0%	1	34
35	要介護認定等基準時間(認知症加算)	H_KJN_KASAN	数字	DECIMAL	4	400	0	5.95	34.66	93710	0%	1	35
36	中間評価項目得点第1群	H_TOKUTEN1	数字	DECIMAL	4	1000	0	698.32	229.56	93710	0%	1	36
37	中間評価項目得点第2群	H_TOKUTEN2	数字	DECIMAL	4	1000	0	755.44	301.17	93710	0%	1	37
38	中間評価項目得点第3群	H_TOKUTEN3	数字	DECIMAL	4	1000	0	860.17	219.14	93710	0%	1	38
39	中間評価項目得点第4群	H_TOKUTEN4	数字	DECIMAL	4	1000	0	915.24	130.89	93710	0%	1	39
40	中間評価項目得点第5群	H_TOKUTEN5	数字	DECIMAL	4	1000	0	458.87	275.15	93710	0%	1	40
41	一次判定警告(配列)コード	H_KEIKOKU_CD	文字	VARCHAR	62	—	—	—	—	93710	0%	1	41
42	状態の安定性	H_ANTEI_CD	文字	VARCHAR	1	2	1	1.66	0.47	93710	0%	1	42
43	認知症自立度II以上の蓋然性	H_GAIZEN_P	数字	DECIMAL	4	9999	-1	97.27	342.33	93310	0%	1	43
44	認知機能及び状態安定性から推定される給付区分	H_KYUFU_CD	文字	VARCHAR	1	2	1	1.71	0.45	93710	0%	1	44
45	合議体番号	H_GOUGITAI_NO	文字	VARCHAR	6	15383	1	2209.85	4373.79	90550	0%	1	45
46	二次判定日	H_HANTEI_DATE	文字	VARCHAR	8	20541202	20090811	20141861.14	10100.51	260	0%	1	46
47	二次判定結果	H_NIJI_CD	文字	VARCHAR	2	99	1	19.59	4.85	0	0%	1	47
48	認定有効期間(開始)	H_START_DATE	文字	VARCHAR	8	20160301	20100601	20142055.42	10048.28	0	0%	1	48
49	認定有効期間(終了)	H_END_DATE	文字	VARCHAR	8	20180228	20140430	20154540.86	10144.80	0	0%	1	49
50	特定疾病コード	H_SHIPPEI_CD	文字	VARCHAR	2	99	1	10.42	6.83	17850750	97%	1	50
51	要介護Iの場合の状態像	H_JOTAI_CD	文字	VARCHAR	2	2	1	1.31	0.46	14824450	80%	1	51
52	現在のサービス区分コード	H_SERVICE_CD	文字	VARCHAR	1	9	1	4.25	3.52	93710	0%	1	52
53	現在の状況	H_JOUKYOU_CD	文字	VARCHAR	1	9	1	2.58	2.72	93710	0%	1	53
54	訪問介護(ホームヘルプサービス)	K_SERVICE_01	数字	DECIMAL	3	999	0	2.70	11.43	93330	0%	1	54
55	訪問入浴介護	K_SERVICE_02	数字	DECIMAL	3	925	0	0.05	0.80	93330	0%	1	55
56	訪問看護	K_SERVICE_03	数字	DECIMAL	3	908	0	0.24	1.84	93330	0%	1	56
57	訪問リハビリテーション	K_SERVICE_04	数字	DECIMAL	3	994	0	0.10	1.09	93330	0%	1	57

② 給付実績情報（基本情報レコード）

項番	データ項目名（日本語）	データ項目名（英名）	型	コラムの型	コラムの長さ	最大値	最小値	平均値	標準偏差	欠損値の数（概数）	欠損値割合	テーブル番号	コラム番号
1	交換情報識別番号	H1_01	文字	VARCHAR	4	1118	1118	1118.00	0.00	0	0%	2	1
2	入力識別番号	H1_02	文字	VARCHAR	4	—	—	—	—	0	0%	2	2
3	レコード種別コード	H1_03	文字	VARCHAR	2	—	—	—	—	0	0%	2	3
4	給付実績情報作成区分コード	H1_04	文字	VARCHAR	1	3	1	1.02	0.15	0	0%	2	4
5	証記載保険者番号	H1_05	文字	VARCHAR	6	—	—	—	—	0	0%	2	5
6	提供先番号（被保険者番号）	H1_06	文字	VARCHAR	64	—	—	—	—	—	—	2	6
7	サービス提供年月（対象年月）	H1_07	文字	VARCHAR	6	201603	201504	201531.58	40.80	0	0%	2	7
8	給付実績区分コード	H1_08	文字	VARCHAR	1	2	1	1.00	0.04	0	0%	2	8
9	事業所番号	H1_09	文字	VARCHAR	10	—	—	—	—	0	0%	2	9
10	整理番号	H1_10	文字	VARCHAR	10	9831110846	1	2251349.83	81884277.59	0	0%	2	10
11	公費1負担者番号	H1_11	文字	VARCHAR	8	—	—	—	—	147713320	93%	2	11
12	公費1受給者番号	H1_12	文字	VARCHAR	7	—	—	—	—	—	—	2	12
13	公費2負担者番号	H1_13	文字	VARCHAR	8	—	—	—	—	158167520	99%	2	13
14	公費2受給者番号	H1_14	文字	VARCHAR	7	—	—	—	—	—	—	2	14
15	公費3負担者番号	H1_15	文字	VARCHAR	8	—	—	—	—	158169650	99%	2	15
16	公費3受給者番号	H1_16	文字	VARCHAR	7	—	—	—	—	—	—	2	16
17	被保険者情報生年月日	H1_17	文字	VARCHAR	8	8	1	5.20	1.66	0	0%	2	17
18	被保険者情報性別コード	H1_18	文字	VARCHAR	1	2	1	1.68	0.47	0	0%	2	18
19	被保険者情報要介護状態区分コード	H1_19	文字	VARCHAR	2	25	1	20.52	4.28	28260	0%	2	19
20	被保険者情報旧措置入所者特例コード	H1_20	文字	VARCHAR	1	2	0	1.01	0.13	151215410	95%	2	20
21	被保険者情報認定有効期間開始年月日	H1_21	文字	VARCHAR	8	20160401	20130401	20146786.04	6119.29	10	0%	2	21
22	被保険者情報認定有効期間終了年月日	H1_22	文字	VARCHAR	8	20600716	0	20153387.62	382766.04	10	0%	2	22
23	被保険者情報老人保健市町村番号	H1_23	文字	VARCHAR	8	—	—	—	—	—	—	2	23
24	被保険者情報老人保健受給者番号	H1_24	文字	VARCHAR	7	—	—	—	—	—	—	2	24
25	被保険者情報保険者番号（後期）	H1_25	文字	VARCHAR	8	—	—	—	—	—	—	2	25
26	被保険者情報被保険者番号（後期）	H1_26	文字	VARCHAR	8	—	—	—	—	—	—	2	26
27	被保険者情報保険者番号（国保）	H1_27	文字	VARCHAR	8	—	—	—	—	—	—	2	27
28	被保険者情報被保険者証番号（国保）	H1_28	文字	VARCHAR	20	—	—	—	—	—	—	2	28
29	被保険者情報宛名番号	H1_29	文字	VARCHAR	10	—	—	—	—	—	—	2	29
30	居宅サービス計画居宅サービス計画作成区分コード	H1_30	文字	VARCHAR	1	8	0	1.46	0.87	19334970	12%	2	30
31	居宅サービス計画事業所番号（居宅介護支援事業所等）	H1_31	文字	VARCHAR	10	—	—	—	—	186710	0%	2	31
32	開始年月日	H1_32	文字	VARCHAR	8	20160331	0	1842878.47	5806625.63	177680	0%	2	32
33	中止年月日	H1_33	文字	VARCHAR	8	20871231	0	127305.94	1596718.55	178020	0%	2	33
34	中止理由・入所（院）前の状況コード	H1_34	文字	VARCHAR	1	8	0	2.80	2.25	140313230	88%	2	34
35	入所（院）年月日	H1_35	文字	VARCHAR	8	20160331	0	2773677.82	6939062.07	180070	0%	2	35
36	退所（院）年月日	H1_36	文字	VARCHAR	8	20280206	0	667940.28	3607618.49	180120	0%	2	36
37	入所（院）実日数	H1_37	文字	VARCHAR	2	89	0	3.09	9.10	0	0%	2	37
38	外泊日数	H1_38	文字	VARCHAR	2	31	0	0.04	0.83	0	0%	2	38
39	退所（院）後の状態コード	H1_39	文字	VARCHAR	1	8	0	3.28	1.77	157566740	99%	2	39
40	保険給付率	H1_40	文字	VARCHAR	3	100	0	91.69	9.01	0	0%	2	40
41	公費1給付率	H1_41	文字	VARCHAR	3	100	0	6.61	24.84	0	0%	2	41
42	公費2給付率	H1_42	文字	VARCHAR	3	100	0	0.00	0.37	0	0%	2	42
43	公費3給付率	H1_43	文字	VARCHAR	3	100	0	0.00	0.01	0	0%	2	43
44	合計情報（決定前）保険サービス単位数	H1_44	文字	VARCHAR	8	58407785	0	5976.00	9905.74	0	0%	2	44
45	合計情報（決定前）保険請求額	H1_45	文字	VARCHAR	9	10105557	0	54984.20	80018.94	0	0%	2	45
46	合計情報（決定前）保険利用者負担額	H1_46	文字	VARCHAR	8	30996123	-173088	5667.16	10325.03	0	0%	2	46
47	合計情報（決定前）保険緊急時施設療養費請求額	H1_47	文字	VARCHAR	9	83808	0	0.02	25.63	0	0%	2	47
48	合計情報（決定前）保険特定診療費請求額	H1_48	文字	VARCHAR	9	193716	0	85.94	1738.55	0	0%	2	48
49	合計情報（決定前）保険特定入所者介護サービス費等請求額	H1_49	文字	VARCHAR	8	103850	0	2253.92	9828.90	0	0%	2	49
50	合計情報（決定前）公費1サービス単位数	H1_50	文字	VARCHAR	8	101815	0	439.38	2816.14	0	0%	2	50
51	合計情報（決定前）公費1請求額	H1_51	文字	VARCHAR	8	503860	-32004	674.75	7195.75	0	0%	2	51
52	合計情報（決定前）公費1本人負担額	H1_52	文字	VARCHAR	8	260743	0	13.76	410.31	0	0%	2	52
53	合計情報（決定前）公費1緊急時施設療養費請求額	H1_53	文字	VARCHAR	8	910	0	0.00	0.21	0	0%	2	53

③ 給付実績情報（明細情報レコード）

項番	データ項目名（日本語）	データ項目名（英名）	型	カラムの型	カラムの長さ	最大値	最小値	平均値	標準偏差	欠損値の数（概数）	欠損値割合	テーブル番号	カラム番号
1	交換情報識別番号	D1_01	文字	VARCHAR	4	1118	1118	1118.00	0.00	0	0%	3	1
2	入力識別番号	D1_02	文字	VARCHAR	4	—	—	—	—	0	0%	3	2
3	レコード種別コード	D1_03	文字	VARCHAR	2	—	—	—	—	0	0%	3	3
4	証記載保険者番号	D1_05	文字	VARCHAR	6	—	—	—	—	0	0%	3	4
5	提供先番号（被保険者番号）	D1_06	文字	VARCHAR	64	—	—	—	—	—	—	3	5
6	サービス提供年月	D1_07	文字	VARCHAR	6	201603	201504	201531.60	40.80	0	0%	3	6
7	事業所番号	D1_09	文字	VARCHAR	10	—	—	—	—	0	0%	3	7
8	整理番号	D1_10	文字	VARCHAR	10	9000032794	1	9313.96	4676603.77	0	0%	3	8
9	サービス種類コード	D1_11	文字	VARCHAR	2	—	—	—	—	0	0%	3	9
10	サービス項目コード	D1_12	文字	VARCHAR	4	—	—	—	—	0	0%	3	10
11	単位数	D1_14	文字	VARCHAR	4	9965	-925	233.49	386.52	0	0%	3	11
12	日数・回数	D1_15	文字	VARCHAR	2	99	0	13.70	13.10	0	0%	3	12
13	公費1対象日数・回数	D1_16	文字	VARCHAR	2	99	0	1.03	5.11	0	0%	3	13
14	公費2対象日数・回数	D1_17	文字	VARCHAR	2	87	0	0.00	0.05	0	0%	3	14
15	公費3対象日数・回数	D1_18	文字	VARCHAR	2	30	0	0.00	0.00	0	0%	3	15
16	サービス単位数	D1_19	文字	VARCHAR	6	584000	-9982	2055.87	5151.39	0	0%	3	16
17	公費1対象サービス単位数	D1_20	文字	VARCHAR	6	72168	-8711	158.92	1451.82	0	0%	3	17
18	公費2対象サービス単位数	D1_21	文字	VARCHAR	6	24556	-2632	0.01	11.12	0	0%	3	18
19	公費3対象サービス単位数	D1_22	文字	VARCHAR	6	308	0	0.00	0.02	0	0%	3	19
20	摘要	D1_23	文字	VARCHAR	20	—	—	—	—	—	—	3	20
21	決定後単位数	D1_24	文字	VARCHAR	4	9965	-925	233.49	386.52	0	0%	3	21
22	決定後日数・回数	D1_25	文字	VARCHAR	2	99	0	13.70	13.10	0	0%	3	22
23	決定後公費1対象日数・回数	D1_26	文字	VARCHAR	2	99	0	1.03	5.11	0	0%	3	23
24	決定後公費2対象日数・回数	D1_27	文字	VARCHAR	2	87	0	0.00	0.05	0	0%	3	24
25	決定後公費3対象日数・回数	D1_28	文字	VARCHAR	2	30	0	0.00	0.00	0	0%	3	25
26	決定後サービス単位数	D1_29	文字	VARCHAR	6	72168	-9982	2055.84	5151.29	0	0%	3	26
27	決定後公費1対象サービス単位数	D1_30	文字	VARCHAR	6	72168	-8711	158.92	1451.81	0	0%	3	27
28	決定後公費2対象サービス単位数	D1_31	文字	VARCHAR	6	24556	-2632	0.01	11.12	0	0%	3	28
29	決定後公費3対象サービス単位数	D1_32	文字	VARCHAR	6	308	0	0.00	0.02	0	0%	3	29
30	再審査回数	D1_33	文字	VARCHAR	2	16	0	0.01	0.09	0	0%	3	30
31	過誤回数	D1_34	文字	VARCHAR	2	5	0	0.00	0.06	0	0%	3	31
32	審査年月	D1_35	文字	VARCHAR	6	201901	201505	201541.87	46.84	0	0%	3	32
33	給付実績情報連結キー	dt1111Key	文字	VARCHAR	108	—	—	—	—	—	—	3	33

④ 給付実績情報（緊急時施設療養・緊急時施設診療情報レコード）

項番	データ項目名（日本語）	データ項目名（英名）	型	カラムの型	カラムの長さ	最大値	最小値	平均値	標準偏差	欠損値の数（概数）	欠損値割合	テーブル番号	カラム番号
1	交換情報識別番号	D2_01	文字	VARCHAR	4	1118	1118	1118.00	0.00	0	0%	4	1
2	入力識別番号	D2_02	文字	VARCHAR	4	—	—	—	—	0	0%	4	2
3	レコード種別コード	D2_03	文字	VARCHAR	2	—	—	—	—	0	0%	4	3
4	証記載保険者番号	D2_05	文字	VARCHAR	6	—	—	—	—	0	0%	4	4
5	提供先番号（被保険者番号）	D2_06	文字	VARCHAR	64	—	—	—	—	—	—	4	5
6	サービス提供年月	D2_07	文字	VARCHAR	6	201603	201504	201530.18	39.89	0	0%	4	6
7	事業所番号	D2_09	文字	VARCHAR	10	—	—	—	—	0	0%	4	7
8	整理番号	D2_10	文字	VARCHAR	10	1	1	1.00	0.00	0	0%	4	8
9	緊急時施設療養情報レコード順次番号	D2_11	文字	VARCHAR	2	99	1	96.85	14.32	0	0%	4	9
10	緊急時傷病名1	D2_12	文字	VARCHAR	80	—	—	—	—	—	—	4	10
11	緊急時傷病名2	D2_13	文字	VARCHAR	80	—	—	—	—	—	—	4	11
12	緊急時傷病名3	D2_14	文字	VARCHAR	80	—	—	—	—	—	—	4	12
13	緊急時治療開始年月日1	D2_15	文字	VARCHAR	8	20160330	0	3067212.66	7239143.40	0	0%	4	13
14	緊急時治療開始年月日2	D2_16	文字	VARCHAR	8	20160330	0	330510.53	2559554.60	0	0%	4	14
15	緊急時治療開始年月日3	D2_17	文字	VARCHAR	8	20160330	0	224754.88	2116358.91	0	0%	4	15
16	往診日数	D2_18	文字	VARCHAR	2	5	0	0.10	0.44	0	0%	4	16
17	往診医療機関名	D2_19	文字	VARCHAR	80	—	—	—	—	—	—	4	17
18	通院日数	D2_20	文字	VARCHAR	2	13	0	0.75	1.01	0	0%	4	18
19	通院医療機関名	D2_21	文字	VARCHAR	80	—	—	—	—	—	—	4	19
20	緊急時治療管理単位数	D2_22	文字	VARCHAR	6	511	0	71.29	177.05	0	0%	4	20
21	緊急時治療管理日数	D2_23	文字	VARCHAR	2	3	0	0.25	0.70	0	0%	4	21
22	緊急時治療管理小計	D2_24	文字	VARCHAR	7	1533	0	127.38	358.74	0	0%	4	22
23	リハビリテーション点数	D2_25	文字	VARCHAR	7	0	0	0.00	0.00	0	0%	4	23
24	処置点数	D2_26	文字	VARCHAR	7	4000	0	2.35	65.38	0	0%	4	24
25	手術点数	D2_27	文字	VARCHAR	7	4606	0	1.01	68.10	0	0%	4	25
26	麻酔点数	D2_28	文字	VARCHAR	7	0	0	0.00	0.00	0	0%	4	26
27	放射線治療点数	D2_29	文字	VARCHAR	7	0	0	0.00	0.00	0	0%	4	27
28	摘要1	D2_30	文字	VARCHAR	128	—	—	—	—	—	—	4	28
29	摘要2	D2_31	文字	VARCHAR	128	—	—	—	—	—	—	4	29
30	摘要3	D2_32	文字	VARCHAR	128	—	—	—	—	—	—	4	30
31	摘要4	D2_33	文字	VARCHAR	128	—	—	—	—	—	—	4	31
32	摘要5	D2_34	文字	VARCHAR	128	—	—	—	—	—	—	4	32
33	摘要6	D2_35	文字	VARCHAR	128	—	—	—	—	—	—	4	33
34	摘要7	D2_36	文字	VARCHAR	128	—	—	—	—	—	—	4	34
35	摘要8	D2_37	文字	VARCHAR	128	—	—	—	—	—	—	4	35
36	摘要9	D2_38	文字	VARCHAR	128	—	—	—	—	—	—	4	36
37	摘要10	D2_39	文字	VARCHAR	128	—	—	—	—	—	—	4	37
38	摘要11	D2_40	文字	VARCHAR	128	—	—	—	—	—	—	4	38
39	摘要12	D2_41	文字	VARCHAR	128	—	—	—	—	—	—	4	39
40	摘要13	D2_42	文字	VARCHAR	128	—	—	—	—	—	—	4	40
41	摘要14	D2_43	文字	VARCHAR	128	—	—	—	—	—	—	4	41
42	摘要15	D2_44	文字	VARCHAR	128	—	—	—	—	—	—	4	42
43	摘要16	D2_45	文字	VARCHAR	128	—	—	—	—	—	—	4	43
44	摘要17	D2_46	文字	VARCHAR	128	—	—	—	—	—	—	4	44
45	摘要18	D2_47	文字	VARCHAR	128	—	—	—	—	—	—	4	45
46	摘要19	D2_48	文字	VARCHAR	128	—	—	—	—	—	—	4	46
47	摘要20	D2_49	文字	VARCHAR	128	—	—	—	—	—	—	4	47
48	緊急時施設療養費合計点数	D2_50	文字	VARCHAR	8	4606	0	3.36	94.38	0	0%	4	48
49	決定後往診日数	D2_51	文字	VARCHAR	2	5	0	0.10	0.44	0	0%	4	49
50	決定後通院日数	D2_52	文字	VARCHAR	2	13	0	0.75	1.01	0	0%	4	50
51	決定後緊急時治療管理単位数	D2_53	文字	VARCHAR	6	511	0	71.29	177.05	0	0%	4	51
52	決定後緊急時治療管理日数	D2_54	文字	VARCHAR	2	3	0	0.25	0.70	0	0%	4	52
53	決定後リハビリテーション点数	D2_55	文字	VARCHAR	7	0	0	0.00	0.00	0	0%	4	53
54	決定後処置点数	D2_56	文字	VARCHAR	7	4000	0	2.35	65.38	0	0%	4	54
55	決定後手術点数	D2_57	文字	VARCHAR	7	4606	0	1.01	68.10	0	0%	4	55
56	決定後麻酔点数	D2_58	文字	VARCHAR	7	0	0	0.00	0.00	0	0%	4	56
57	決定後放射線治療点数	D2_59	文字	VARCHAR	7	0	0	0.00	0.00	0	0%	4	57
58	再審査回数	D2_60	文字	VARCHAR	2	1	0	0.01	0.12	0	0%	4	58

⑤ 給付実績情報（特定診療費・特別療養費情報レコード）

項番	データ項目名（日本語）	データ項目名（英名）	型	カラムの型	カラムの長さ	最大値	最小値	平均値	標準偏差	欠損値の数（概数）	欠損値割合	テーブル番号	カラム番号
1	交換情報識別番号	D3_01	文字	VARCHAR	4	1118	1118	1118.00	0.00	0	0%	5	1
2	入力識別番号	D3_02	文字	VARCHAR	4	—	—	—	—	0	0%	5	2
3	レコード種別コード	D3_03	文字	VARCHAR	2	—	—	—	—	0	0%	5	3
4	証記載保険者番号	D3_05	文字	VARCHAR	6	—	—	—	—	0	0%	5	4
5	提供先番号（被保険者番号）	D3_06	文字	VARCHAR	64	—	—	—	—	—	—	5	5
6	サービス提供年月	D3_07	文字	VARCHAR	6	201603	201504	201530.73	40.36	0	0%	5	6
7	事業所番号	D3_09	文字	VARCHAR	10	—	—	—	—	0	0%	5	7
8	整理番号	D3_10	文字	VARCHAR	10	2015004109	1	2514.02	2250275.57	0	0%	5	8
9	特定診療情報レコード順次番号	D3_11	文字	VARCHAR	2	99	1	28.51	42.80	0	0%	5	9
10	傷病名	D3_12	文字	VARCHAR	200	—	—	—	—	—	—	5	10
11	識別番号	D3_13	文字	VARCHAR	2	55	1	25.39	17.64	1790	0%	5	11
12	単位数	D3_14	文字	VARCHAR	4	480	0	57.32	76.72	0	0%	5	12
13	保険回数	D3_15	文字	VARCHAR	2	93	0	18.81	11.88	0	0%	5	13
14	保険サービス単位数	D3_16	文字	VARCHAR	6	12180	0	476.03	643.97	0	0%	5	14
15	保険合計単位数	D3_17	文字	VARCHAR	8	28650	0	476.06	1285.20	0	0%	5	15
16	公費1回数	D3_18	文字	VARCHAR	2	93	0	1.48	6.07	0	0%	5	16
17	公費1サービス単位数	D3_19	文字	VARCHAR	6	10962	0	38.24	227.96	0	0%	5	17
18	公費1合計単位数	D3_20	文字	VARCHAR	8	20736	0	38.24	395.10	0	0%	5	18
19	公費2回数	D3_21	文字	VARCHAR	2	1	0	0.00	0.00	0	0%	5	19
20	公費2サービス単位数	D3_22	文字	VARCHAR	6	5	0	0.00	0.00	0	0%	5	20
21	公費2合計単位数	D3_23	文字	VARCHAR	8	10	0	0.00	0.01	0	0%	5	21
22	公費3回数	D3_24	文字	VARCHAR	2	0	0	0.00	0.00	0	0%	5	22
23	公費3サービス単位数	D3_25	文字	VARCHAR	6	0	0	0.00	0.00	0	0%	5	23
24	公費3合計単位数	D3_26	文字	VARCHAR	8	0	0	0.00	0.00	0	0%	5	24
25	摘要	D3_27	文字	VARCHAR	200	—	—	—	—	—	—	5	25
26	決定後単位数	D3_28	文字	VARCHAR	4	480	0	57.32	76.72	0	0%	5	26
27	決定後保険回数	D3_29	文字	VARCHAR	2	93	0	18.80	11.89	0	0%	5	27
28	決定後保険サービス単位数	D3_30	文字	VARCHAR	6	12180	0	475.43	643.15	0	0%	5	28
29	決定後保険合計単位数	D3_31	文字	VARCHAR	8	21454	0	475.42	1283.39	0	0%	5	29
30	決定後公費1回数	D3_32	文字	VARCHAR	2	93	0	1.48	6.07	0	0%	5	30
31	決定後公費1サービス単位数	D3_33	文字	VARCHAR	6	10962	0	38.19	227.67	0	0%	5	31
32	決定後公費1合計単位数	D3_34	文字	VARCHAR	8	20736	0	38.19	394.63	0	0%	5	32
33	決定後公費2回数	D3_35	文字	VARCHAR	2	1	0	0.00	0.00	0	0%	5	33
34	決定後公費2サービス単位数	D3_36	文字	VARCHAR	6	5	0	0.00	0.00	0	0%	5	34
35	決定後公費2合計単位数	D3_37	文字	VARCHAR	8	10	0	0.00	0.01	0	0%	5	35
36	決定後公費3回数	D3_38	文字	VARCHAR	2	0	0	0.00	0.00	0	0%	5	36
37	決定後公費3サービス単位数	D3_39	文字	VARCHAR	6	0	0	0.00	0.00	0	0%	5	37
38	決定後公費3合計単位数	D3_40	文字	VARCHAR	8	0	0	0.00	0.00	0	0%	5	38
39	再審査回数	D3_41	文字	VARCHAR	2	2	0	0.00	0.02	0	0%	5	39
40	過誤回数	D3_42	文字	VARCHAR	2	2	0	0.00	0.06	0	0%	5	40
41	審査年月	D3_43	文字	VARCHAR	6	201901	201505	201540.97	47.32	0	0%	5	41
42	給付実績情報連結キー	dt111Key	文字	VARCHAR	108	—	—	—	—	—	—	5	42

図1

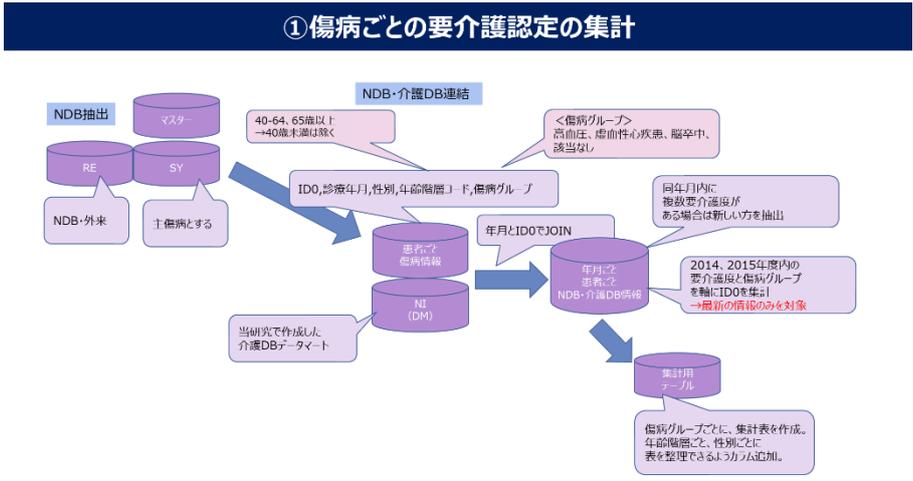


図2

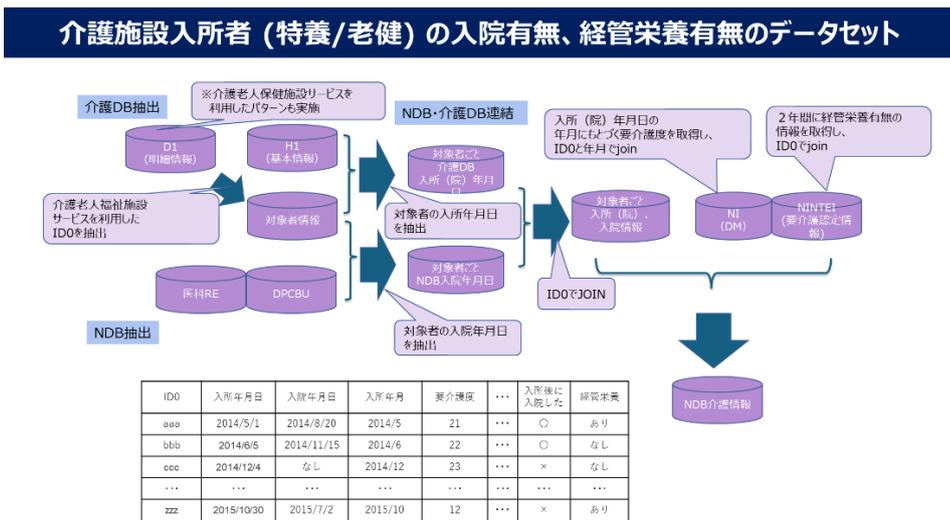
傷病ごとの要介護認定の集計

年齢階層	性別	要支援1	要支援2	要介護1	...	要介護4	要介護5
40-64歳	男性	16	26	106	...	356	406
40-64歳	女性	20	30	110	...	360	416
65歳以上	男性	30	40	120	...	370	426
65歳以上	女性	40	50	130	...	380	436

年齢階層	性別	要支援1	要支援2	要介護1	...	要介護4	要介護5
40-64歳	男性	13	23	103	...	353	403
40-64歳	女性	20	30	110	...	360	413
65歳以上	男性	30	40	120	...	370	423
65歳以上	女性	40	50	130	...	380	433

年齢階層	性別	要支援1	要支援2	要介護1	...	要介護4	要介護5
40-64歳	男性	100	200	1000	...	3500	4000
40-64歳	女性	200	300	1100	...	3600	4100
65歳以上	男性	300	400	1200	...	3700	4200
65歳以上	女性	400	500	1300	...	3800	4300

図3



厚生労働科学研究費補助金(地域医療基盤開発推進研究事業)
「臨床疫学に活用可能な NDB 等データセットの作成に関する研究」
分担研究報告書 (令和 5 年度)

(長期追跡性に優れた個人名寄せ ID の検討)

研究分担者 今村知明 (奈良県立医科大学 公衆衛生学)
研究分担者 野田龍也 (奈良県立医科大学 公衆衛生学)
研究協力者 明神大也 (奈良県立医科大学 公衆衛生学)

研究要旨

本研究では「健康・医療・介護分野の大規模データの利活用を推進する」ことの一環で、NDB に含まれる個人単位被保険者番号に基づく個人名寄せ ID (id5) が付与されたデータの長期追跡性について検討及び評価を行う。さらに、id5 が付与される前のデータについても、既存の id とのより精緻なデータ連結の可能性について検討を行う。2013 年 4 月診療分から 2023 年 3 月診療分までの NDB 特別抽出データのうち、RE レコードの id1n, id2, id5, 診療年月を用いた。そして、id1n と id2 と診療年月から id0 を作成した。さらに、作成された id0 に対して対応する id5 が存在する場合は id0 を id5 で上書きして、それを id0n とした。

id5 の設定状況を調査したところ。全レセプトに含まれる有効な id5 の割合は DPC では 95% 程度、医科入院では 92% 程度、医科入院外では 97% 程度であった。id5 が 9 の割合は 1% 未満、id5 が null だった比率は 2~8% 程度であった。また、2022 年 4 月に存在した id0, id0n, id5 に対し、2023 年 3 月まで存在する id を追跡した結果、id0, id0n, id5 いずれも、9 ヶ月後で約 90% の捕捉率を示し、id0n が最も捕捉率が高かった。

本研究において id5 の設定状況を把握したとともに、id0 および id5 の長期追跡の妥当性評価を行い、id0 と id5 を組み合わせた新たな id0n の開発に踏み込んだ。今後、死亡者を除外するなど、id5 及び id0n の精度をより詳細に検証していく。

研究協力者

東野恒之 (三菱総合研究所)

A. 研究目的

昨今のコロナ禍に伴う社会状況や、近年の急速な少子高齢化を背景とする社会構造の変化に伴い、より適正な医療を提供するための基礎資料として、健康・医療・介護分野のビッグデータ解析の必要性が高まっている。実際 NDB 及び介護 DB の情報はすでに医療・介護それぞれの分野での政策利用や研究利用が開始されており、提供件数は徐々に増加している。一方データの容量が膨大であること、

データの構造が複雑であること等から、これらデータの構造や操作を知悉した研究者でなければ解析を行うのは容易ではなく、医療者等いわゆる臨床系の研究者だけでは取り扱いが困難なデータとなっている。一方で臨床系の研究者は、日々医療・介護の現場で医療を担っており、適正な医療の提供・実施にあたっての課題にも直面していることから、こうした研究者が自ら NDB や介護 DB 等の大規模データを簡便に分析できるようになれば、それら課題の解決に貢献することが期待される。

加えて、NDB は 2020 年 10 月から介護 DB 及び 2022 年 4 月から DPC データベースとの連結が可能となり、さらに「医療・介護データ

等の解析基盤」の実運用が開始される予定であることから、将来的には現在よりも大規模なデータ解析が可能になる見込みである。これらの大規模データを広く研究者が利用できるようにするためには、より簡便にデータ分析できる環境が整備される必要がある。

本研究は「健康・医療・介護分野の大規模データの利活用を推進する」ことを目的に、匿名医療保険等関連情報データベース（NDB）、および介護保険総合データベース（介護DB）さらに2020年から提供が開始されている医療・介護の連結情報から、より利用者にとって使いやすいデータセットの開発を目指すものである。その一環で、長期追跡性に優れた個人名寄せIDの検討として、個人単位被保険者番号に基づく個人名寄せIDが付与されたデータを受領し、長期追跡性について検討及び評価を行う。さらに、後ろ向きのデータ連結性（2021年10月以前のデータとの連結）についても、既存のIDとのより精緻なデータ連結の可能性について検討を行う。

B. 研究方法

2013年4月診療分から2023年3月診療分までのNDB特別抽出データのうち、REレコードのid1n,id2,id5,診療年月を用いた。そして、id1nとid2と診療年月からid0を作成した。さらに、作成されたid0に対して対応するid5が存在する場合はid0をid5で上書きして、それをid0nとした。

id5は2022年3月診療分から含まれていたが、処理の関係で大部分を2022年4月診療分から使用した。

- ① 2022年4月診療分から2023年3月診療分までの各月のid5の格納状況を、DPC・医科入院・医科入院外に分けてレセプト枚数ベースで集計した。
- ② 2022年4月診療分から2022年12月診療分までの患者数としてid0及びid5の件数を、性別・年齢階級別に集計した。年齢は

2022年12月末時点の年齢とした。また参考として2023年1月1日時点の住民基本台帳の人数を示した。

- ③ id5とid1nの対応割合を調べた。具体的には、2022年4月から2023年3月までのid1nとid5の対応表を作成しdistinctをとった。
- ④ ③と比較するため、2022年4月から2023年3月までのid1nとid0の対応表を作成しdistinctをとった。
- ⑤ 2022年4月に存在したid0, id0n, id5に対し、2023年3月まで存在するidを追跡した。
- ⑥ 2013年4月に存在したid0, id0nに対し、2023年3月まで存在するidを追跡した。

（倫理面への配慮）本研究は、奈良県立医科大学医の倫理審査を受けて実施された。また、NDBデータを用いた分析結果は、厚生労働省及び社会保険診療報酬支払基金の公表物確認を受けて承認された。

C. 研究結果

研究方法①～⑥の結果を以下に示す。可読性の関係上、一部考察も含めている。

- ① 2022年4月診療分から2023年3月診療分までの各月のid5の格納状況を表1に示した。id5はid1nやid2と同様、16進数64桁のハッシュ値であったが、一部にnullまたは9が含まれていた。16進数64桁のハッシュ値が含まれるid5を有効値とみなした。その結果、全レセプトに含まれる有効なid5の割合はDPCでは95%程度、医科入院では92%程度、医科入院外では97%程度であった。id5が9の割合は1%未満、id5がnullだった比率は2～8%程度であった。id5が9は正しく何らかの理由でid5が取得できなかった場合、id5がnullは全額公費であった場合と考えられる。全額公費を除くと、全体の99%近くにid5が含まれており、id5の格納率の高さが示され

た。

- ② 2022年4月診療分から2022年12月診療分までのid0及びid5のユニークな件数を、性別・年齢階級別に表2及び図1に示した。id5=9またはid5=nullとなっている値は、それぞれid5に9またはnullが入っていたレセプトの、ユニークなid0数を意味する。id5の数はid0の数よりやや少なくなっている。id0で名寄せできなかったものもid5に含まれている可能性はあるものの、id5の数とid5がnullの数を合わせるとid0の数とほぼ合致することから、id0もid5とほぼ同等の名寄せ精度を示している可能性も示唆された。なお、参考に掲載した住民基本台帳の人数と比べると、0-4歳と75歳以降は100%近くのレセプト発生率であった。85歳以上では住民基本台帳の人数よりid0,id5の数のほうが多かったが、これは集計対象期間内に死亡した人も含まれているためと考えられる。
- ③ 2022年4月診療分から2023年3月診療分までのid1nとid5の対応件数を表3に示す。id1n:id5=1:1はid5に対応するid1nが1つしかなかった場合である。id1n:id5=n:1は1つのid5に対応するid1nが複数あった場合であり、12ヶ月の間に保険証が変わったことが示唆される。id1n:id5=1:nは1つのid1nに対応するid5が複数存在した場合で、id1n:id5=n:mはid1nとid5が多対多になった場合である。id1n:id5=1:nまたはn:mだった場合は、同性の双子の可能性もあるがそれ以外にid5の誤りの可能性も考えられる。具体的なケースをNDB抽出データから取り上げると、以下の2つのパターンであることが分かった。1つ目はid1nとid2が一致してid5が異なるケース、これが118,590件（レセプト枚数にして1,533,250枚）であった。2つ目はid5が一致しているが性別又は生年月が異なるケース、これが327,830件（レセプト枚数にして

4,307,520枚）であった。これらはid5の異常と言えるかもしれないが、全体の0.5%未満であり、NDB分析には大きな支障がないと考えられる。

次に、id1n:id5=n:1(id5=5/null)はid5に9又はnullが入っていたレセプトのユニークなid1nの数である。id5=nullが全額公費とした場合、日本で約250万人の医療扶助等による受診者がいると推計された。

- ④ 2022年4月診療分から2023年3月診療分までのid1nとid5の対応件数を表4に示す。id1n:id0=1:1はid0に対応するid1nが1つしかなかった場合である。id1n:id0=n:1は1つのid0に対応するid1nが複数あった場合であり、id0の作成過程で名寄せされたものである。この結果によると、1700万程度のid1nが800万程度のid0に名寄せされたことになる。なお、医療保険から全額公費に移行したid0の数は170,560件であった。
- ⑤ 2022年4月に存在したid0, id0n, id5に対し、2023年3月まで存在するidを追跡した結果のID数と捕捉率を表5及び図2に示す。id0,id0n,id5いずれも、9ヶ月後で約90%の捕捉率であったが、id0nが最も捕捉率が高かった。
- ⑥ 2013年4月に存在したid0, id0nに対し、2023年3月まで存在するidを追跡した結果のID数と捕捉率を表6及び図3に示す。id0,id0nともに、9年6か月後で約80%の捕捉率であったが、id0に比べてid0nのほうが捕捉率は高かった。

D. 考察

本研究で、id5の設定状況を把握したとともに、id0およびid5の長期追跡の妥当性評価を行い、id0とid5を組み合わせた新たなid0nの開発に踏み込んだ。

本研究を通して日本で初めてNDBに含まれる

id5 の設定状況を公表した。その結果、id5 は 1 年間で 0.5% 程度誤結合している可能性はあるものの、大部分は正しく連結されていることが分かった。それとともに我々が従来作成していた id0 の正確性も高いことが裏付けられた。

医療扶助は 250 万人と結果に示したが、厚生労働省が発表している生活保護受給者数は約 200 万人で、本研究の名寄せが不十分だと思われる。名寄せされた id0 に対して対応する id5 が存在する場合は id5 で上書きする id0n を作成した。id0n のメリットは id0 で名寄せできなかった id を id5 で名寄せが可能になったこと、2022 年 2 月以前の id5 が存在しなかったレセプトについても名寄せが可能になったこと、公費に移行した患者についても名寄せが可能になったことである。しかしながら id0 をベースにしている (id0 は id1n をベースにしている) ため、id5 で可能な同性の双子の切り分けは不可能である。

今後、死亡者を除外するなど、id5 及び id0n の精度をより詳細に検証していく。

E. 結論

本研究において、id5 の追跡を行ったとともに id0 の妥当性を検証した。

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

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

1) 令和 4 年 4 月 28 日 第 5 回 医療扶助に関する検討会資料 1

<https://www.mhlw.go.jp/content/12002000/000934593.pdf>

表1 2022年度のid5設定状況（レセプト枚数ベース）

診療年月	DPC						
	id5有効値	id5=9	id5=null	合計	id5有効値の比率	id5=9の比率	id5=nullの比率
202204	954,350	7,078	39,688	1,001,116	95.3%	0.7%	4.0%
202205	957,595	6,918	40,666	1,005,179	95.3%	0.7%	4.0%
202206	1,033,643	7,452	43,515	1,084,610	95.3%	0.7%	4.0%
202207	989,852	7,658	41,404	1,038,914	95.3%	0.7%	4.0%
202208	918,731	7,412	37,578	963,721	95.3%	0.8%	3.9%
202209	938,990	7,281	39,025	985,296	95.3%	0.7%	4.0%
202210	1,000,254	7,727	41,658	1,049,639	95.3%	0.7%	4.0%
202211	981,715	6,964	39,981	1,028,660	95.4%	0.7%	3.9%
202212	949,990	6,290	38,479	994,759	95.5%	0.6%	3.9%
202301	901,028	6,111	37,291	944,430	95.4%	0.6%	3.9%
202302	971,558	6,407	40,263	1,018,228	95.4%	0.6%	4.0%
202303	1,045,278	6,588	43,372	1,095,238	95.4%	0.6%	4.0%

診療年月	医科入院						
	id5有効値	id5=9	id5=null	合計	id5有効値の比率	id5=9の比率	id5=nullの比率
202204	1,144,251	6,386	97,943	1,248,580	91.6%	0.5%	7.8%
202205	1,129,549	6,161	97,728	1,233,438	91.6%	0.5%	7.9%
202206	1,158,550	6,283	99,319	1,264,152	91.6%	0.5%	7.9%
202207	1,165,103	6,662	99,414	1,271,179	91.7%	0.5%	7.8%
202208	1,161,232	6,585	99,504	1,267,321	91.6%	0.5%	7.9%
202209	1,146,171	6,519	98,559	1,251,249	91.6%	0.5%	7.9%
202210	1,143,703	6,771	97,776	1,248,250	91.6%	0.5%	7.8%
202211	1,140,398	6,142	96,945	1,243,485	91.7%	0.5%	7.8%
202212	1,155,245	5,918	98,105	1,259,268	91.7%	0.5%	7.8%
202301	1,129,150	5,695	97,588	1,232,433	91.6%	0.5%	7.9%
202302	1,137,444	5,560	97,173	1,240,177	91.7%	0.4%	7.8%
202303	1,165,063	5,474	98,381	1,268,918	91.8%	0.4%	7.8%

診療年月	医科入院外						
	id5有効値	id5=9	id5=null	合計	id5有効値の比率	id5=9の比率	id5=nullの比率
202204	82,525,970	496,171	2,203,448	85,225,589	96.8%	0.6%	2.6%
202205	80,548,455	472,157	2,169,132	83,189,744	96.8%	0.6%	2.6%
202206	82,777,730	462,745	2,227,036	85,467,511	96.9%	0.5%	2.6%
202207	85,290,190	501,145	2,202,035	87,993,370	96.9%	0.6%	2.5%
202208	84,850,299	494,256	2,208,168	87,552,723	96.9%	0.6%	2.5%
202209	82,276,028	458,019	2,192,986	84,927,033	96.9%	0.5%	2.6%
202210	83,525,939	523,328	2,220,060	86,269,327	96.8%	0.6%	2.6%
202211	83,612,710	502,293	2,210,009	86,325,012	96.9%	0.6%	2.6%
202212	86,954,462	492,012	2,246,400	89,692,874	96.9%	0.5%	2.5%
202301	80,577,072	451,461	2,134,000	83,162,533	96.9%	0.5%	2.6%
202302	80,757,258	439,526	2,138,218	83,335,002	96.9%	0.5%	2.6%
202303	90,541,374	471,948	2,278,397	93,291,719	97.1%	0.5%	2.4%

表2 性別・年齢階級別の患者数 (id0,id5) の分布 (2022年4月診療分～12月分)

性別	年齢階級	id0件数	id5件数	id5 = 9	id5 is null	住基(R5.1.1)
男性	0～4歳	2,141,056	2,072,462	100,103	17,942	2,192,250
	5～9歳	2,463,379	2,416,437	43,589	24,444	2,589,182
	10～14歳	2,492,087	2,446,049	33,495	30,676	2,769,436
	15～19歳	2,273,831	2,225,341	36,291	28,808	2,863,895
	20～24歳	2,189,483	2,120,515	60,790	14,478	3,242,208
	25～29歳	2,249,746	2,177,476	54,317	17,295	3,324,472
	30～34歳	2,315,005	2,247,208	48,238	20,120	3,347,657
	35～39歳	2,596,962	2,524,723	48,441	25,736	3,731,475
	40～44歳	2,853,432	2,772,087	48,902	35,794	4,078,212
	45～49歳	3,423,644	3,318,743	53,362	56,343	4,808,345
	50～54歳	3,564,774	3,434,291	54,012	83,068	4,827,997
	55～59歳	3,236,545	3,110,133	33,222	96,608	4,085,447
	60～64歳	3,147,469	2,994,849	37,064	110,348	3,701,790
	65～69歳	3,234,084	3,073,172	40,756	119,801	3,631,126
	70～74歳	4,035,120	3,828,185	54,011	172,959	4,336,773
	75～79歳	3,114,322	2,964,465	10,107	130,279	3,191,907
	80～84歳	2,393,880	2,300,854	1,872	92,932	2,375,135
	85～89歳	1,486,600	1,438,625	1,295	47,483	1,413,673
	90～94歳	605,614	591,950	619	13,200	542,360
	95～99歳	137,069	134,850	145	2,071	112,356
100歳以上	13,937	13,722	17	204	10,052	
	小計	49,968,039	48,206,137	760,648	1,140,589	61,175,748
女性	0～4歳	2,028,062	1,963,782	94,079	16,216	2,087,071
	5～9歳	2,313,236	2,270,965	40,877	21,624	2,461,529
	10～14歳	2,320,590	2,278,475	31,518	27,979	2,632,354
	15～19歳	2,262,144	2,215,783	31,205	30,033	2,724,879
	20～24歳	2,525,778	2,437,172	60,879	18,484	3,068,216
	25～29歳	2,697,997	2,593,007	55,474	23,186	3,124,431
	30～34歳	2,750,794	2,656,544	48,585	27,709	3,166,865
	35～39歳	3,050,078	2,955,687	45,383	35,288	3,571,272
	40～44歳	3,308,510	3,209,185	43,834	44,080	3,935,243
	45～49歳	3,895,322	3,774,230	47,999	61,637	4,656,040
	50～54歳	3,994,955	3,858,297	47,557	77,709	4,720,154
	55～59歳	3,494,477	3,371,809	41,546	73,443	4,058,198
	60～64歳	3,313,474	3,189,930	44,680	71,185	3,759,265
	65～69歳	3,447,809	3,322,537	47,512	81,621	3,814,948
	70～74歳	4,534,891	4,388,427	70,112	139,227	4,841,620
	75～79歳	3,796,330	3,622,142	12,680	148,562	3,907,207
	80～84歳	3,284,111	3,135,292	2,649	149,551	3,310,601
	85～89歳	2,527,512	2,421,282	2,406	106,684	2,480,888
	90～94歳	1,463,838	1,415,613	1,609	47,897	1,380,195
	95～99歳	524,794	509,325	553	15,274	462,073
100歳以上	98,189	95,193	96	2,976	78,033	
	小計	57,632,891	55,684,677	771,233	1,220,365	64,241,082
	合計	107,600,930	103,890,814	1,531,881	2,360,954	125,416,830

図1 性別・年齢階級別の患者数 (id0,id5) の分布 (2022年4月診療分～12月分)

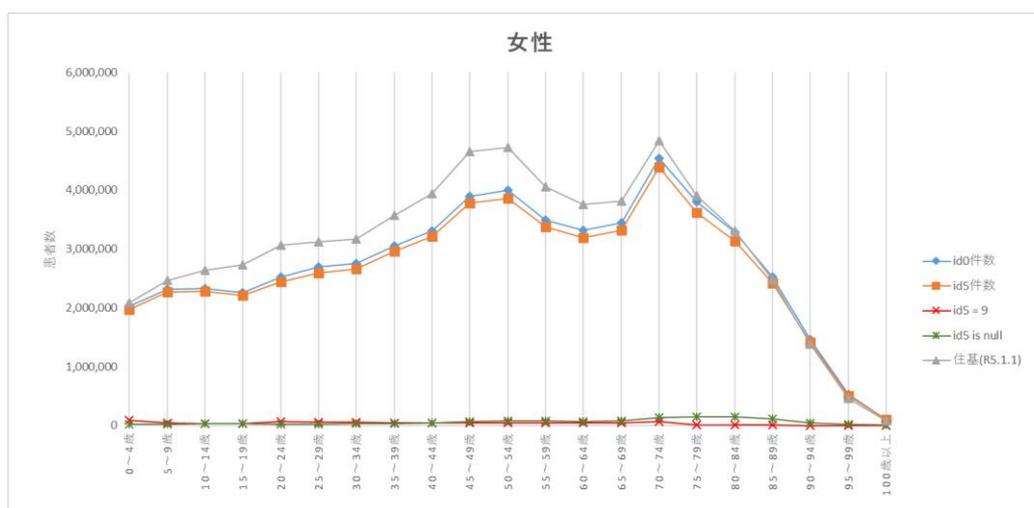
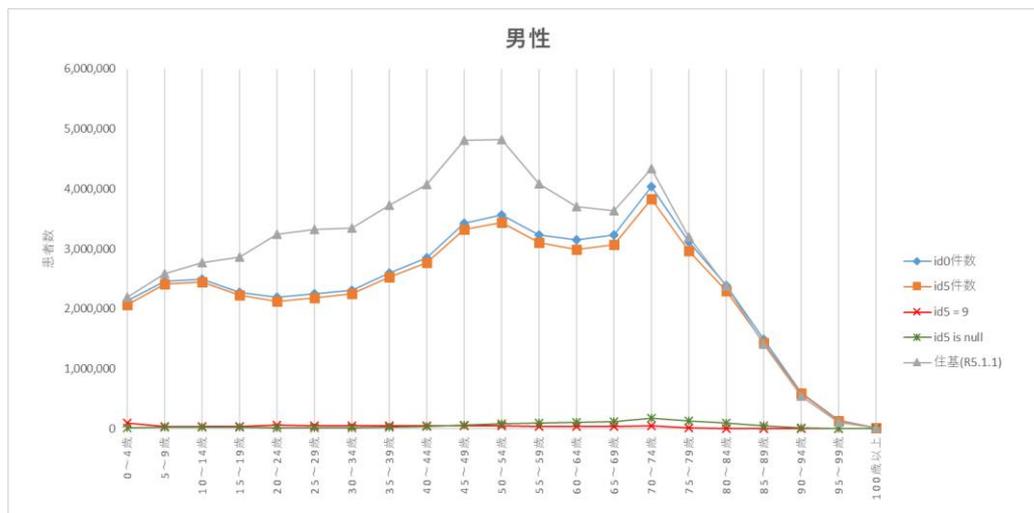


表 3 id5 設定状況

	id1n		id5	
	件数	構成比	件数	構成比
id1n : id5 = 1 : 1	98,671,350	81%	98,671,350	91%
id1n : id5 = n : 1	17,928,280	15%	8,512,100	8%
id1n : id5 = 1 : n	137,530	0%	276,090	0%
id1n : id5 = n : m	485,500	0%	416,530	0%
id1n : id5 = n : 1 (id5=9)	1,274,040	1%	10未満	
id1n : id5 = n : m (id5=9を含む)	512,960	0%	513,170	0%
id1n : id5 = n : 1 (id5=null)	2,513,440	2%	0	0%
id1n : id5 = 1 : 2 (id5=null : 9)	3,320	0%	10未満	
id1n : id5 = 1 : 2 (id5=hash値 : 9)	10未満		10未満	
全件数	121,526,410		108,389,240	

表 4 id0 設定状況

	id1n		id0	
	件数	構成比	件数	構成比
id1n : id0 = 1 : 1	104,949,470	86%	104,949,470	93%
id1n : id0 = n : 1	16,576,940	14%	7,955,120	7%
全件数	121,526,410	100%	112,904,600	100%

表 5 2022 年度における id 数とその追跡率

診療年月	id数			追跡率		
	id0	id0n	id5	id0	id0n	id5
202204	59,463,613	59,386,609	57,565,542	100.0%	100.0%	100.0%
202205	58,230,825	58,287,823	56,514,833	97.9%	98.1%	98.2%
202206	57,874,469	57,987,698	56,222,195	97.3%	97.6%	97.7%
202207	57,536,571	57,705,305	55,945,574	96.8%	97.2%	97.2%
202208	57,115,762	57,336,802	55,580,649	96.1%	96.5%	96.6%
202209	56,609,145	56,882,161	55,126,021	95.2%	95.8%	95.8%
202210	56,081,255	56,429,315	54,672,134	94.3%	95.0%	95.0%
202211	55,398,825	55,788,122	54,039,741	93.2%	93.9%	93.9%
202212	54,530,313	54,952,089	53,214,228	91.7%	92.5%	92.4%
202301	53,101,544	53,562,102	51,840,702	89.3%	90.2%	90.1%
202302	50,453,276	50,924,177	49,244,545	84.8%	85.8%	85.5%
202303	43,726,451	44,165,299	42,616,545	73.5%	74.4%	74.0%

図 2 2022 年度における id 数とその追跡率

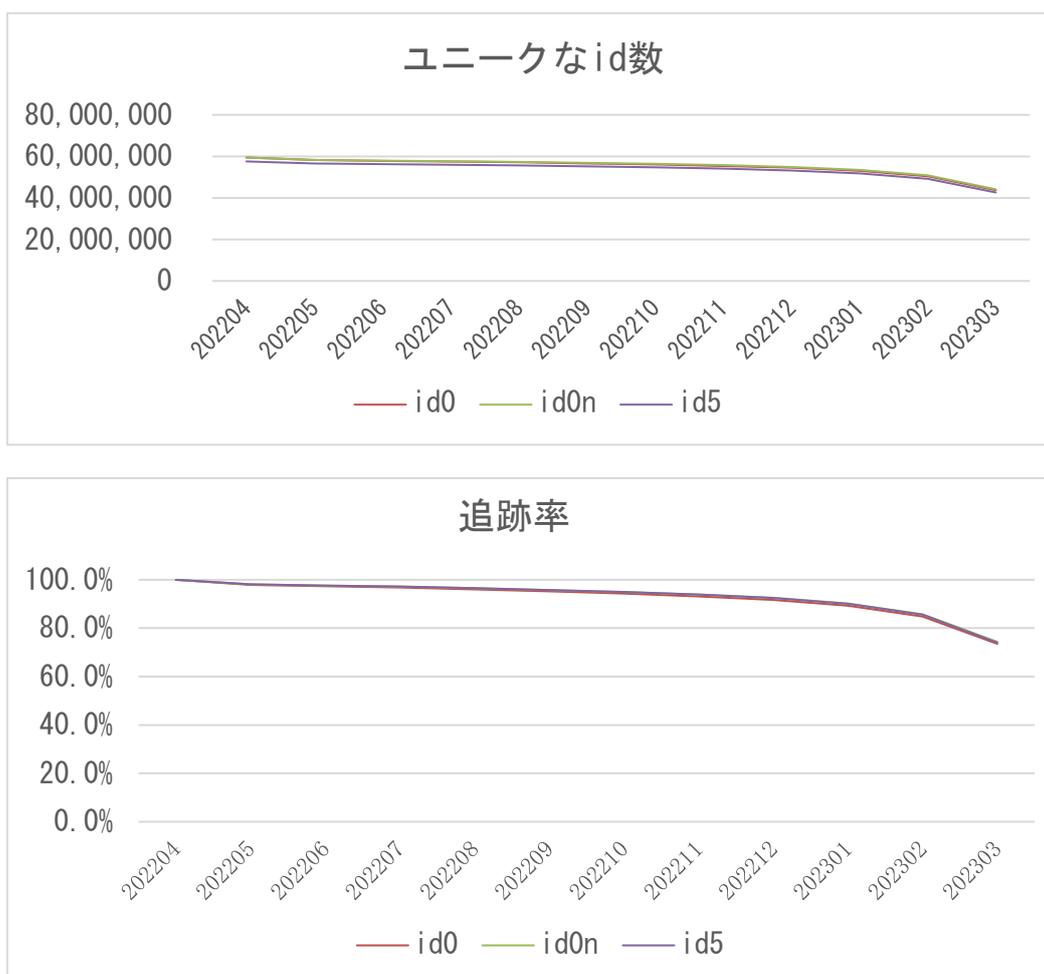
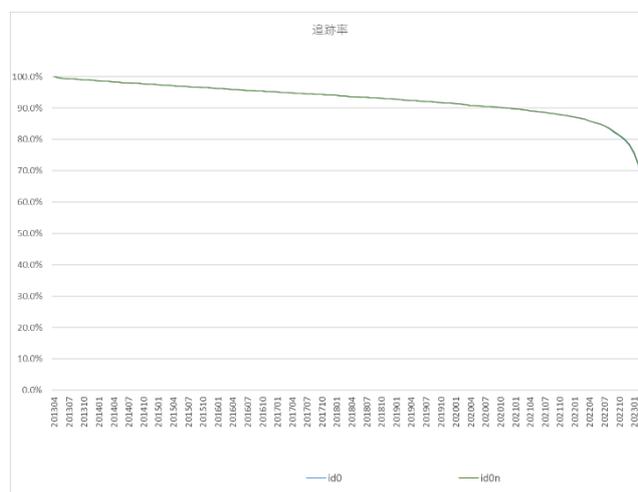
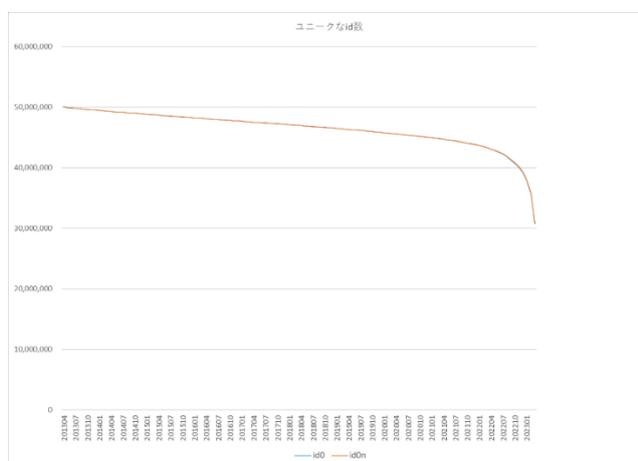


表 6 2013 年度～2022 年度の
Id 数とその追跡率

診療年月	id数		追跡率	
	id0	id0n	id0	id0n
201304	50,116,547	50,106,238	100.0%	100.0%
201305	49,922,975	49,912,666	99.6%	99.6%
201306	49,851,840	49,841,531	99.5%	99.5%
201307	49,794,025	49,783,716	99.4%	99.4%
201308	49,735,951	49,725,642	99.2%	99.2%
201309	49,684,151	49,673,842	99.1%	99.1%
201310	49,634,294	49,623,985	99.0%	99.0%
201311	49,580,576	49,570,267	98.9%	98.9%
201312	49,527,576	49,517,267	98.8%	98.8%
201401	49,468,795	49,458,486	98.7%	98.7%
201402	49,413,561	49,403,252	98.6%	98.6%
201403	49,357,013	49,346,704	98.5%	98.5%
201404	49,279,202	49,268,893	98.3%	98.3%
201405	49,227,248	49,216,939	98.2%	98.2%
201406	49,179,968	49,169,659	98.1%	98.1%
201407	49,134,118	49,123,809	98.0%	98.0%
201408	49,086,231	49,075,922	97.9%	97.9%
201409	49,043,207	49,032,898	97.9%	97.9%
201410	48,997,935	48,987,626	97.8%	97.8%
201411	48,949,994	48,939,685	97.7%	97.7%
201412	48,904,303	48,893,994	97.6%	97.6%
201501	48,847,068	48,836,759	97.5%	97.5%
201502	48,791,988	48,781,679	97.4%	97.4%
201503	48,740,439	48,730,130	97.3%	97.3%
201504	48,669,218	48,658,909	97.1%	97.1%
201505	48,621,387	48,611,078	97.0%	97.0%
201506	48,578,094	48,567,785	96.9%	96.9%
201507	48,532,739	48,522,430	96.8%	96.8%
202101	44,965,130	44,954,821	89.7%	89.7%
202102	44,893,607	44,883,298	89.6%	89.6%
202103	44,816,046	44,805,737	89.4%	89.4%
202104	44,704,680	44,694,371	89.2%	89.2%
202105	44,615,221	44,604,912	89.0%	89.0%
202106	44,527,366	44,517,057	88.8%	88.8%
202107	44,426,440	44,416,131	88.6%	88.6%
202108	44,318,012	44,307,703	88.4%	88.4%
202109	44,200,465	44,190,156	88.2%	88.2%
202110	44,084,082	44,073,773	88.0%	88.0%
202111	43,958,468	43,948,159	87.7%	87.7%
202112	43,823,526	43,813,217	87.4%	87.4%
202201	43,671,892	43,661,583	87.1%	87.1%
202202	43,504,754	43,494,445	86.8%	86.8%
202203	43,321,495	43,311,186	86.4%	86.4%
202204	43,063,037	43,052,728	85.9%	85.9%
202205	42,823,252	42,831,227	85.4%	85.5%
202206	42,560,005	42,584,556	84.9%	85.0%
202207	42,258,174	42,300,110	84.3%	84.4%
202208	41,809,592	41,870,055	83.4%	83.6%
202209	41,229,531	41,309,869	82.3%	82.4%
202210	40,702,302	40,806,280	81.2%	81.4%
202211	40,058,773	40,179,764	79.9%	80.2%
202212	39,189,538	39,325,902	78.2%	78.5%
202301	37,827,450	37,982,742	75.5%	75.8%
202302	35,685,727	35,850,180	71.2%	71.5%
202303	30,720,862	30,882,352	61.3%	61.6%

図 3 2013 年度～2022 年度の
Id 数とその追跡率



令和5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））

臨床疫学に活用可能なNDB等データセットの作成に関する研究（21AA2006）

分担研究報告書

NDB オンサイトリサーチセンターと比較した HIC

研究協力者 松居宏樹 東京大学大学院医学系研究科臨床疫学・経済学准教授

研究分担者 康永秀生 東京大学大学院医学系研究科臨床疫学・経済学教授

研究要旨 この報告書では、NDB オンサイトリサーチセンターにおける大規模医療データを用いた研究の成果について説明する。さらに、NDB データを取り扱う際の利用方法である、NDB オンサイトリサーチセンターと HIC について、その特性の違いを説明する。そのうえで、NDB オンサイトリサーチセンターを基準とした場合に HIC が抱えている課題として、データ記憶領域の不足と、データベース応答速度の問題を明らかにした。それを解消する糸口となる技術として、PostgreSQL の citus 拡張を用いたデータの圧縮、および partitioning について検証を行った。その結果、データの圧縮と partitioning の併用はデータベース応答速度を上げることがわかった。HIC は今後 NDB の利用形態として重要な位置を占めると考えられるものの、NDB オンサイトリサーチセンターと比較すると、現状では HIC 特有の課題が存在する。HIC を用いてデータベース研究を行うためには、ユーザー側の十分な技術的工夫が必要であると考えられる。

A. 研究目的

匿名医療保険等関連情報データベース（NDB）が研究目的で国内研究者に提供され始めて 10 年以上が経過した。一部の研究者の間では、NDB を用いた臨床疫学研究が実施され、ハイインパクトジャーナルに掲載されるなど、その存在価値は増している。しかし、多くの研究者にとっては、そのデータ利用に伴う制約の多さから利用が難しい側面があった。提供開始当初、NDB は特別抽出が主要なデータ利用方法

として活用されてきた。特別抽出は、ユーザーがデータ利用環境を独自に用意し、その環境でデータを解析する。データ利用環境には、十分なセキュリティ要件を満たす必要がある。しかし、一部の研究者にとってそのような環境を確保することは困難であった。そういった研究者に向けて、NDB オンサイトリサーチセンターが国内数か所開設された。現在、データ利用環境がクラウド上に構築され、各 NDB オンサイトセンターに向けて利用環境が提供されてい

る。我々は、東京大学 NDB オンサイトリサーチセンターにおいて、複数の臨床疫学研究を実施し、成果を上げてきた。

また、データ利用環境をクラウド上に構築し、各ユーザーに向けて利用環境を提供する Healthcare Intelligence Cloud (HIC)が、2024 年から運用を開始する予定とされている。HIC は、ユーザーがオンサイトセンターに出向かなくとも、ユーザー個人の端末からクラウド上の仮想環境に接続することで、NDB の解析を可能とするものである。しかしながら、それぞれのシステムにおける制約について利用者に十分な理解が得られていないため、その環境を十分に活用できず、成果が上がらない事例も散見される。この報告書では、まず、我々が NDB オンサイトリサーチセンターを活用して挙げってきた成果を紹介する。次に、HIC 本格利用に先立ち HIC の試行利用を行い、NDB オンサイトリサーチセンターでの利用方法を基準として、HIC のシステム上の制約下で成果を得るために必要となる工夫について検討したので報告する。

B. 研究方法

(1) NDB オンサイトセンターでの成果

2022 年以前は、NDB データは厚生労働省のオンプレミス環境に構築された Oracle 社製データベースに保存されていた。ユーザーはオンサイトセンター内から、専用回線を用いて Oracle 社製データベースを操作し、データをハンドリングし解析を行っていた。2022 年以前にオンサイトセンターを用いて実施した研究には、以下がある。

1. Ishimaru M, Matsui H, Ono S, Hagiwara Y, Morita K, Yasunaga H.

Preoperative oral care and effect on postoperative complications after major cancer surgery. *British Journal of Surgery*. 2018 Oct 12;105(12):1688–96.

2. Hashimoto Y, Matsui H, Michihata N, Ishimaru M, Yasunaga H, Aihara M, et al. Incidence of sympathetic ophthalmia after inciting events: a national database study in Japan. *Ophthalmology*. 2021 Sep 21;S0161-6420(21)00719-3.
3. Takeuchi Y, Kumamaru H, Hagiwara Y, Matsui H, Yasunaga H, Miyata H, et al. Sodium-glucose cotransporter-2 inhibitors and the risk of urinary tract infection among diabetic patients in Japan: Target trial emulation using a nationwide administrative claims database. *Diabetes Obes Metab*. 2021 Jun;23(6):1379–88.
4. Ishimaru M, Ono S, Morita K, Matsui H, Hagiwara Y, Yasunaga H. Prevalence, Incidence Rate, and Risk Factors of Medication-Related Osteonecrosis of the Jaw in Patients With Osteoporosis and Cancer: A Nationwide Population-Based Study in Japan. *Journal of Oral and Maxillofacial Surgery*. 2022;80(4):714–27.
5. Kasajima M, Eggleston K, Kusaka S, Matsui H, Tanaka T, Son BK, et al. Projecting prevalence of frailty and dementia and the economic cost of care in Japan from 2016 to 2043: a

microsimulation modelling study. The Lancet Public Health. 2022;7(5):e458–68.

2022年4月から、NDBシステムはAWSクラウド上に移行され、オンサイトセンターもクラウド上のシステムへの接続に変更となった。ユーザーの利用形態も、オンサイトセンター内から専用回線を用いてクラウド環境にあるRedshift（AWS社製高速分散DWH）を操作し、データをハンドリングし解析を行う形へ変更となった。2022年以降にオンサイトセンターを用いて実施した研究には、以下のようなものがある。

6. Kimura Y, Suzukawa M, Jo T, Hashimoto Y, Kumazawa R, Ishimaru M, et al. Epidemiology of severe childhood asthma in Japan: A nationwide descriptive study. Allergy. 2024 Jan 7;all.16008.
7. Kimura Y, Jo T, Hashimoto Y, Kumazawa R, Ishimaru M, Matsui H, et al. Epidemiology of patients with lymphangioleiomyomatosis: A descriptive study using the national database of health insurance claims and specific health checkups of Japan. Respir Investig. 2024 May;62(3):494–502.

この報告書では特に、上記6の研究についてそのデータ特性の観点から紹介する。

(2) HICとNDBオンサイトセンターの比較軸

研究者が利用可能なNDBの提供方法には、

サンプリングデータセット、特別抽出、集計表、オンサイトリサーチセンター利用(i)、オンサイトリサーチセンター利用(ii)、HICの6種類の方法がある。

サンプリングデータセットは、年間複数時点のレセプト横断データからあらかじめ一定割合でランダムに抽出した個票レベルレセプトを提供する方法である。

特別抽出は、レセプトデータから、利用者の指定する特定条件で抽出した個票レベルレセプトを提供する方法である。

集計表は、レセプトデータを、利用者の指定する特定条件で集計した集計表レベルの情報を提供する方法である。

オンサイトリサーチセンター利用(i)は、オンサイトリサーチセンターでNDBデータに直接アクセスし、データの解析をオンサイトセンター内で終える方法である。

オンサイトリサーチセンター利用(ii)は、オンサイトリサーチセンターでNDBデータに直接アクセスし抽出した個票レベルレセプトを、オンサイトセンターから取り出し、利用者が用意した自身の環境で解析する方法である。

それぞれの方法で特性が異なっているが、今回は特にオンサイトセンター利用(i),(ii)とHICを対象を絞り、以下のポイントで整理行う。

- 1) 利用者に求められるセキュリティ要件の違い
- 2) データ利用可能になるまでの時間
- 3) 使用可能データの違い
- 4) システム上の制約の違い
 - ・システムのデフォルトのスペック
 - ・システムの拡張性

- ・ソフトウェア
- ・利用可能なクラウドリソース

また、上記のポイントで各提供方法を整理したうえで、NDB オンサイトリサーチセンターと HIC のシステム上の制約で特に着目すべき点を挙げる。

(3) HIC のシステムを利用するうえでの工夫とその使用感

我々は当研究班の昨年の報告書において、HIC でデータを解析する場合に必要な工夫を以下のように提言した。

- ・ Linux 環境が用意できる場合、適切な拡張機能のインストールが望ましい。
- ・ S3 上に保存されたデータを直接 PostgreSQL に load するためには、PostgreSQL 拡張である、s3csv_fdw のインストールが望ましい。
- ・ PostgreSQL において、テーブルデータの圧縮を行うことが望ましい。NDB オンサイトリサーチセンターで実績のある citus database のインストールが望ましい。
- ・ Docker 環境を立ち上げたうえで解析環境を整備することが可能であるならば、そうすべきである。

今年度は、本研究班代表者の尽力により、PostgreSQL の拡張である citus (2)を用いて、テーブルデータを圧縮した NDB データベースが作成された。このデータベースの課題と、その解決のための工夫を提案する。

C. 研究結果

(1) NDB オンサイトリサーチセンターでの成果

上記研究 6 において、我々は、NDB を用いて小児重症喘息患者の全国的有病率の記述研究を行い、有病率の経時的变化を明らかにした。

この研究において、診断コードと喘息関連薬の処方状況を組み合わせて喘息児を定義した。

その結果、2019 年には、0~5 歳の喘息児 253,684 人のうち、253,052 人 (99.8%)、632 人 (0.2%) がそれぞれ軽度~中等度、重度のグループに分類された。

重症小児喘息の全国有病率は、調査期間中に 0~5 歳 (10 万人当たり 19.7 人から 11.0 人) および 6~11 歳 (44.4 人から 22.8 人) の両年齢層でほぼ半減したが、軽度から中等度の喘息はより緩やかな減少傾向を示した。

(2) NDB オンサイトリサーチセンターでのデータハンドリング

この研究のデータハンドリングにおいて重要なポイントは、診断コードと喘息関連薬の処方状況を組み合わせて喘息児を定義した点である。図 1 に示すように、症例を検索する際に、「診断コードを有するレセプト」と「喘息関連薬の処方があるレセプト」は単一レセプト内で完結しない場合がある。例えば、患者の診断名と処方状況が同一以下レセプトではなく、医科・調剤で分かれて処方される場合や、診断がついた翌月にたまたま処方薬を調剤した場合である。そのため、この研究においては、まず、条件に該当するレセプトを特定した。

次に、そのレセプトを算定した症例のレセプトデータを、時系列に沿って追跡した。また、時系列の追跡精度を向上させる目的

で、図 2 にあるように、再帰的な ID 追跡を行った。再帰的 ID 追跡は、抽出条件を絞って行う特別抽出データでは実施された報告がない。そのため、条件を絞らずにレセプトデータを取得する特別な研究か、オンサイトリサーチセンターでのみ行うことが可能な抽出方法である。

さらに、我々のチーム内の研究者がデータを取り扱いやすくするために、データを時系列に整理するプログラム（NDB データ抽出プログラム）を作成し、研究者に整理されたデータを共有している。

NDB データ抽出プログラムは、Redshift 上に保存された NDB オンサイトリサーチセンターからアクセス可能な NDB データに対して動作する、Postgresql の Stored Procedure として構築されている。NDB データは規模が大きいため、そのデータを整形する際に適切な処理手順を踏まなければ、データの処理が不可能になる。

NDB オンサイトリサーチセンターからアクセス可能な NDB データは、月単位（PRAC_YM 単位）でパーティショニングがなされている。そのため、抽出プログラムのアルゴリズムを各 PRAC_YM に対して独立して動作するプログラムとして構築した。プログラムのプロセスは以下のとおりである。

1. キーレセプトの取得

(1) 特定条件に合致するキーレセプトのレセプト ID(i) を取得

2. ID 追跡

(1) レセプト ID(i) に紐づく ID1N(i) を取得

(2) ID1N(i) とペアとなる ID2(i) を取得

(3) ID1N(i)ID2(i) から紐づく ID1N(ii), ID2(ii) をすべて取得しする（再帰的処理）。

(4) ID1N(ii), ID2(ii) の中で、ID の接続を確認して ID0 と ID1N(ii) の対応を取得する（再帰的処理）。

3. 患者詳細情報の取得

(1) 任意の PRAC_YM(i) に対して、ID1N(iii) に紐づくレセプト ID(ii) を取得する。

(2) レセプト ID(ii) に紐づく傷病情報・診療情報を取得する。

一連の患者情報を整理取得する際には、上記のプロセスを経て抽出を実施している。特に 3 のプロセスはすべて、任意の PRAC_YM に対して逐次的に実行できる。そのため、対象となる PRAC_YM(n) が増えても処理時間は n に比例して増えるだけで、指数関数的に処理時間が延びることはない。

(4) NDB オンサイトリサーチセンターと HIC の比較

1. 利用者に求められるセキュリティ要件の違い

NDB オンサイトリサーチセンターも HIC も、データ本体は AWS 上の閉じられた VPC(<https://aws.amazon.com/jp/vpc/>) に構築されており、外部からアクセスすることはできない。NDB オンサイトリサーチセンターは HIC に比べ、物理的に隔離されたセキュリティレベルの高い環境からのデータへのアクセスが可能である。これに対して、HIC は物理的に隔離された環境を整備するわけではない。そのため、若干 HIC のほうがセキュリティの厳重さは低く設計がなされている。（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

<https://www.mhlw.go.jp/content/12400000/001174849.pdf>

2.利用開始までの時間

NDB 利用方法によって、承認を受けてからデータ利用可能となるまでの時間が異なる。例えば、特別抽出では利用承諾を受けてからデータ利用可能となるまでに長大な時間が必要となる。令和5年6月の規制改革実施計画においては、利用申請から申請者が実際にデータの利用を開始するまでに平均で1年以上の時間を要することが課題としてあがっている。この原因として、研究に必要な NDB データのみを抽出したデータセットを研究者に提供していることが挙げられる。

NDB オンサイトリサーチセンターでは、承認後、事務処理を経てデータの利用を開始できる。これは、オンサイト向けデータセットがあらかじめ構築されており、利用者がそこから必要となる情報を選択して解析をするためである。

これに対し、HIC では、承認後、研究に必要な NDB データのみを抽出したデータセットを環境内に配置する形でデータ提供が行われる。そのため、NDB オンサイトリサーチセンターのように、あらかじめ構築されたデータセットを環境内に設置しておく対応をとらなければ、特別抽出と同様に利用開始まで長大な時間を要するものと考えられる。

3.使用可能データの違い

利用形態によって、利用可能なデータは異なる。

まず、NDB 本体に含まれるレセプト、特定健診データは、NDB オンサイトリサーチセンターでも HIC でも利用可能である。しかし、NDB 本体データに紐づけることができる介護データや DPC データ等のその他データは NDB オンサイトリサーチセンターでは利用することができない。

次に、NDB 本体データについても、NDB オンサイトリサーチセンターと HIC で利用可能な情報が異なる。NDB オンサイトリサーチセンターでは、直近 10 年分の NDB 本体データがデータベースとして保管されている。これに対し、HIC では、上述のように研究者が指定した条件で抽出した研究用データセットが環境内に設置される。

4.NDB 利用方法によるシステム上の制約の違い

昨年の報告書でもふれたが、NDB オンサイトリサーチセンター、HIC とともに、解析環境として m5.4xlarge (vCPU:16, memory:64GB, Strage: 3TB) か、m5.2xlarge (vCPU:8, memory:32GB, Strage: 1TB) が利用可能である。NDB オンサイトリサーチセンター、HIC とともに、OS は Windows か Linux を選択可能である。

NDB オンサイトリサーチセンター、HIC とともに、仕様で決定しているシステムの拡張は原則としてできない。そのため、NDB オンサイトリサーチセンター利用(ii) では、ユーザーが特別抽出と同等の環境を独自に用意することで、個票レベルのデータを独自システムで解析する仕組みが存在している。

また、システム仕様として決定しているハードウェア仕様に対して、ソフトウェアイ

インストールは厚生労働省と協議のうえ、柔軟に実施できる。NDB オンサイトリサーチセンターでは、Linux 上に Docker 環境が用意されており、Docker コンテナを環境内に持ち込んで運用することができる。HIC においては、ソフトウェア持ち込みを申請して、厚生労働省の確認を経て、ソフトウェアのインストールをユーザー自身が行うことができる。この際、環境をインターネット上にオンラインにすることはできないため、ソフトウェアによっては利用できない場合がある。

最後に、利用可能なクラウドリソースが HIC と NDB オンサイトリサーチセンターで異なっている。いずれの利用環境でも解析環境から S3 のバケットを参照できる。しかし、NDB オンサイトリサーチセンターでは、Redshift 上の NDB データにアクセスできるのに対し、HIC 環境では Redshift へのアクセスはできない。そのため、HIC では、大規模なデータをハンドリングするためには十分な工夫が必要である。

(5) HIC のシステムを利用する上での工夫

HIC を利用する上での課題は、データ記憶領域のサイズ問題が第一に挙げられた。そこで、この研究班では、昨年の報告書の提案に基づき Postgresql の拡張である、citus を用いたデータの圧縮を行った。citus はもともと、データの分散配置を行う事で大規模データの高速処理を行う目的で開発が進んでいた拡張である。Citus Data

(<https://www.citusdata.com/>) が開発を行っており、2019 年に Microsoft に買収されたものの、オープンソース版が提供されてい

る。また、分散配置を行う機能以外にも columnar table という機能が追加されており、テーブルを列単位で圧縮することで、データサイズの圧縮と全件検索の高速化を可能としている。今回の研究班においても、テキスト形式でおおよそ 21TB の容量がある 4 年分のレセプトデータを、2.6TB まで圧縮することができた。

しかしながら、columnar table では、INDEX を利用することができない制約がある。そのため、大規模なデータを Full Scan してしまいデータベースの応答速度が致命的に低下することがある。今回作成したデータベースにおいても、特定の診療年月の MED_SI レコードをスキャンし 10 行表示する以下のクエリを実行するだけで 1 時間以上の時間が必要であった。

```
SELECT * FROM med_si WHERE prac_ym = '202204' LIMIT 10;
```

データベースの検索速度を向上するための工夫として、NDB オンサイトリサーチセンターの Redshift 上のデータベースでは、partitioning を行っていることはすでに述べたとおりである。今回、columnar table と partitioning を併用することで、ディスク記憶領域の課題を回避しながら、データの検索速度を上げることができるか検討した。検討プロセスは以下 4 つテーブルを作成したうえで、同一データを挿入し、ディスク容量と検索速度を比較した。データは 48 ヶ月分の dpc_si テーブルから、レセプト ID(seq2_no),レセプト電算コード (prac_act_code),prac_yo(診療年月)を選択し、格月 25 万行を挿入した。

1. partitioning を行わず、圧縮も行わないテーブル test_nonp_heap
2. prac_ym で partitioning し、圧縮を行わないテーブル test_p_heap
3. partitioning を行わず、圧縮を行うテーブル test_nonp_col
4. prac_ym で partitioning し、圧縮を行うテーブル test_p_col

Table 1 に、データの圧縮効率と検索速度（テストクエリ実行時間）を示した。わずかなオーバーヘッドはあるものの、partitioning を行った圧縮テーブル test_p_col でも、partitioning を行わない圧縮テーブル test_nonp_col と同等のテーブルサイズにデータを圧縮できた。

さらに、prac_ym を用いて検索範囲を絞ったテストクエリでは、Partitioning を行わない圧縮テーブル test_nonp_col に比べ、Partitioning を行った圧縮テーブル test_p_col の検索速度が大幅に向上した。

D. 考察

この報告書では、NDB オンサイトリサーチセンターにおける大規模医療データを用いた研究の成果について説明した。さらに、NDB データを取り扱う際の利用方法である、NDB オンサイトリサーチセンターと HIC について、その特性の違いを説明した。そのうえで、NDB オンサイトリサーチセンターを基準とした場合の HIC が抱えている課題として、データ記憶領域の不足と、データベース応答速度の問題を明らかにした。それを解消する糸口となる技術として、pgstgresql の citus 拡張を用いたデータの圧縮、および partitioning について検証を行った。その結果、データの圧縮と partitioning

の併用はデータベース応答速度を上げることがわかった。

NDB オンサイトリサーチセンターにおける抽出プログラムは、prac_ym で partitioning されたデータベースへの適用を前提として作成した。今回構築した圧縮データベースは実用に耐えるものではなかったが、適切な partitioning を行うことで既存資産を利用しながら、臨床疫学研究を実施することが可能であると考えられる。

NDB データの提供については、HIC での提供の実現で利用者の初期投資を大幅に削減できることが期待されている。実際に使用可能なデータの種類を考えると、HIC は重要な利用方法であると思われる。しかしながら、技術的な工夫を行わなければ、データが得られても現実的な時間で解析を行うことは不可能であることが、今回の試行利用から明らかとなった。このことから、ユーザー側も十分な技術的工夫を行う能力が求められる。また、ユーザー側の技術的工夫を補うための対応として、厚労省側が検討することができるオプションは複数残されている。

例えば、HIC において Redshift 等のリソースへのアクセスを許可することは、一つの方策であろう。ただし、Redshift は従量課金のシステムであり、ユーザー側の不手際で多大なリソースを消費ことも危惧される。NDB の維持コストとの兼ね合いも検討する必要である。他にも例えば、NDB オンサイトリサーチセンターでデータの抽出処理を行い、その個票レベル生成物を HIC 環境で利用する方策も有用な方法であると考えられる。すでに我々はプロジェクトごとの研究用標準データセットを抽出するプロ

グラムを NDB オンサイトリサーチセンターで運用していることは上述の通りである。これらのプログラムを有償で維持していくような形をとることも今後検討してよいのではないかと考える。

また、今年度は検証ができなかったものの、Postgresql から S3 に保存した csv ファイルを直接参照する s3csv_fdw のような拡張機能が存在しており、適切に利用することで、保存領域の課題を解消できる可能性がある。

E. 結論

HIC は今後 NDB の利用形態として重要な位置を占めると考えられる。しかし、NDB オンサイトリサーチセンターと比較すると、なし

HIC 環境で臨床疫学研究を行うためには、ユーザー側の十分な技術的工夫が必要であることが明らかとなった。

F. 健康危険情報

なし

G. 研究発表

なし

H. 知的財産権の出願・登録状況

なし

図 1 : オンサイトセンターでの症例抽出 (第 6 回 NDB ユーザー会資料より)

オンサイトセンターでの研究の実際

症例抽出計画の策定

- Inclusion Period の決定
 - 許諾された全期間で症例を検索する。
 - 容量削減や抽出プロセス短縮のために、期間を絞ることはある。
- キーレセプト条件は単一レセプト内で完結する条件にする。
 - 他のレセプトを参照して決定する条件は好ましくない。
 - 例えば、“脳梗塞病名のついた症例のレセプト”や“rtPA実施症例のレセプト”は検索できる。
 - “脳梗塞発症後”、“リハビリテーション”を行ったレセプトの検索は危険
- 研究に利用する情報を決める。

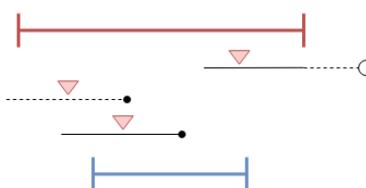


図 2 : オンサイトセンターでの症例追跡 (第 6 回 NDB ユーザー会資料より)

オンサイトセンターでの症例の実際

症例の追跡

- 症例追跡期間の決定
 - 許諾された全期間でキーレセプト前後情報を追跡する。
- 症例の追跡
 - キーレセプトを基にID1, ID2を再帰的に追跡する。
 - ‘ID1 = OR ID2 = ’の条件一回で追跡を打ち切るのはダメ
- 追跡したID1, ID2を基にID0を作成する。
 - 別人を誤って追跡した場合は後でその症例を除くか諦める。
 - ランダム抽出する場合は、ID0単位で抽出が望ましい。

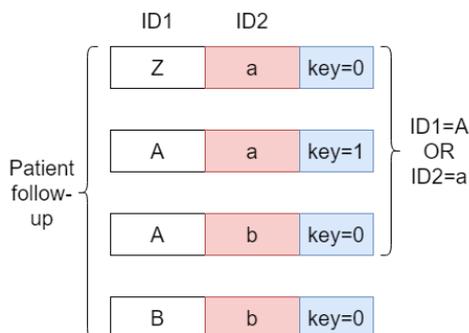


図3：オンサイトセンターでの抽出プログラム（第6回 NDB ユーザー会資料より）

環境整備について

抽出プログラムの作成

- データ抽出プロセスの仕様を策定
 - オンサイト環境で追加のプログラムなしで動くPostgresql 上に Stored procedure として構築
 - Redshift には Select 権限のみが付与されている
 - TEMP TABLE は作れる。
 - 適切に sequential key の付与を行う
 - あまり大きなテーブルは作らない。
 - NDB 本体データは PRAC_YM 単位で partition が切られている
 - PRAC_YM 毎に処理を逐次的に行う。

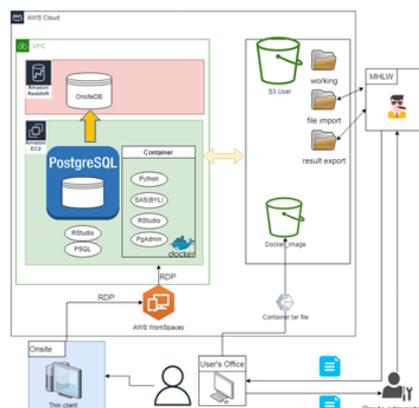


Table 1: 列圧縮と Partitioning test

	テーブルサイズ	テストクエリ 1*	テストクエリ 2**
test_nonp_heap	690Mb	865.776ms	119.406 ms
test_p_heap	691Mb	0.025ms	43,816 ms
test_nonp_col	23Mb	939.189ms	1043.778 ms
test_p_col	24Mb	0.811ms	34.084 ms

** テストクエリ 1 は `Select * FROM table where prac_ym = '202204' Limit 10;`

** テストクエリ 2 は `Select * FROM table where prac_ym = '202204' and prac_act_cd = '190101770' Limit 10;`

Reference

1. Kimura Y, Suzukawa M, Jo T, Hashimoto Y, Kumazawa R, Ishimaru M, et al. Epidemiology of severe childhood asthma in Japan: A nationwide descriptive study. Allergy. 2024 Jan 7;all.16008.
2. Citus Documentation - Citus 12.1 documentation [Internet]. [cited 2024 May 8]. Available from: https://docs.citusdata.com/en/v12.1/?_gl=1*_2oenlh*_ga*MTYyODQ0MTI5NC4xNzE1MDk2MzE2*_ga_DS5S1RKEB7*MTcxNTA5NjMxNi4xLjAuMTcxNTA5NjMxOS4wLjAuMA..

付録: Postgresql の操作

Partitioning Table の作成と columnar 圧縮の適用

まずは、以下のように partitioning したテーブルを作成する。

```
CREATE TABLE test_p_col (  
column1 data_type,  
column2 data_type,  
...  
prac_ym character(6)  
) partitioned by prac_ym;
```

次に、partitioning された子テーブルを各月分用意しデータ圧縮設定(using columnar)を適用する。

```
CREATE Table test_p_col_202204 PARTITION OF test_p_col FOR VALUES IN ('202204') using  
columnar;
```

test_p_col に対してデータを挿入すれば、partition で分けられた圧縮テーブルが完成する。

令和5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))
臨床疫学に活用可能なNDB等データセットの作成に関する研究(21AA2006)

分担研究報告書

HICにおける分析用データベース構築における課題検討

研究代表者	森 由希子	京都大学医学部附属病院医療情報企画部 講師
研究分担者	加藤 源太	京都大学医学部附属病院診療報酬センター 特定教授
研究分担者	黒田 知宏	京都大学医学部附属病院医療情報企画部 教授
研究分担者	植嶋 大晃	京都大学国際高等教育院附属データ科学イノベーション教育研究センター 特定講師
研究分担者	大寺 祥佑	国立研究開発法人 国立長寿医療研究センター 老年学・社会科学研究センター 医療経済研究部 副部長
研究分担者	田宮菜奈子	筑波大学医学医療系/ヘルスサービス開発研究センター 教授
研究分担者	杉山 雄大	国立国際医療研究センター 医療政策研究室長
研究協力者	小宮山 潤	筑波大学医学医療系 助教

研究要旨

NDBは令和2年10月から介護DB及び令和4年3月からDPCデータベースとの連結が可能となった。さらに「医療・介護データ等の解析基盤(Healthcare Intelligence Cloud: HIC)」がクラウド環境で構築・運用開始されており、今後利用者の利便性が向上することが期待される。本研究ではHICの試行利用を行い、HICの利用における課題抽出および解決法の検討を行った結果を報告する。

A. 研究目的

レセプト情報等を活用した分析の体制整備の推進および保健医療に関するビッグデータの効果的な利活用を推進するため、NDBや、NDBと介護DBの情報の連結解析を可能とするために医療介護連結解析基盤(Healthcare Intelligence Cloud: HIC)が構築された。現在NDBデータの提供方法として

・特別抽出(審議会による提供承諾を得た研究において、研究者の依頼に応じたデー

タ抽出を行い、個人単位のデータを提供)

・集計表(審議会による提供承諾を得た研究において、研究者の依頼に応じたデータ抽出を行い、作表して提供)

・サンプリングデータセット(トライアルデータセット)(1カ月分の匿名レセプトデータに対し抽出・匿名化処理等を行ったレディメイドのデータを提供)

・オンサイトリサーチセンター(厚生労働省が指定する情報セキュリティ対策が講じられた施設において、厚生労働省が管理す

る NDB と通信回線で結ばれた端末の利用環境)

がある。オンサイトリサーチセンターでのデータ利用以外では、利用者は研究計画に基づいて抽出されたデータの提供を受け、自身で準備した分析環境において解析を行う。

HIC はオンサイトリサーチセンターと同様、クラウド上に設置された NDB データの分析環境であるが、従来のオンサイトリサーチセンターでのデータ利用と異なり、申請に基づき抽出されたデータが HIC 上に提供され、利用者はすでに抽出されたデータを HIC 上で分析するため、いわゆる特別抽出データやサンプリングデータセットの利用環境としての側面が強い。今回特別抽出データを HIC 上で利用するケースにおける課題及びその解決法について検討した。

B. 研究方法

利用承諾を受けた NDB・介護 DB 連結データを用いて、HIC 上での分析環境構築し、以下の課題について検討を行った。

- ・データ格納
- ・windows 環境における分析環境構築
- ・Linux 環境における分析環境構築

今回利用したデータは以下の通り。

- ・NDB データ (医科入院・医科外来・DPC・歯科・調剤) 4 年分 (HIC 利用開始時は 3 年分)
- ・介護 DB データ 3 年分 (HIC 利用開始時のデータ量は全 csv ファイルで約 11TB)

C. 研究結果

(1) データ格納

HIC は AWS Cloud 上に環境が構築されており、提供されるデータは s3 領域に格納される。利用者は s3 からデータを EC2 instance にダウンロードし、EC2 instance に与えられた領域で分析を行う。今回 HIC 利用開始時に提供を受けたデータは全 csv ファイルで約 11TB。試行利用において、EC2 instance に与えられたストレージは約 16TB であった。

(2) EC2 instance におけるデータベース構築 (windows 環境)

EC2 instance に以下の手順で PostgreSQL データベース作成を試みた。

- ① NDB DWH を作成
- ② 介護 DB DWH を作成
- ③ ①と②を名寄せしデータセットを作成

しかし、Postgres に提供データを登録した時点でデータ量が約 1.2 倍増加し、さらに名寄せ情報付与用の領域が必要であり、この時点で研究者作業用の領域が確保できないことが判明した。このため、ストレージ容量不足に対して以下の検証を行った。

(図 1)

(ア) 対受領データに対するアプローチ

【不要カラムの削除の検証】

- ・提供不可カラムの削除 →効果なし
- ・NULL のカラムの削除 →効果なし
- ・不要となる ID 系カラムの削除→1,434GB (約 11%)

【データの型変更の検証】

- ・ハッシュ値を数字で振りなおす→174GB (約 1%)
- ・フラグ系カラムの最適化→848GB

(約 7%)

- ・年月系カラムの最適化→903GB (約 8%)
- ・数値系カラムの最適化→51GB (約 0.5%)

検証結果として全約 12TB のデータに対して total で約 2.4-3.4TB の削除が可能であったが、十分な作業領域の確保は困難であった。

(イ) 対利用環境に対するアプローチ

【PostgreSQL 圧縮機能の利用の調査】

PostgreSQL におけるデータ圧縮機能である TOAST 圧縮 (PGLZ、lz4) の活用を検討したが、適用可否の判断が 1 レコードの長さに依存するため、NDB データは適用対象とならず、対応不可であった。

【s3 をネットワークドライブ利用する方法の調査】

ネットワークドライブのように、s3 を”マウント”するアプリは存在したが、

PostgreSQL などでファイルを参照する際、”s3 からダウンロードして開く”という挙動のため、ストレージ容量不足に対しては効果がなかった。

【s3 上のファイルを外部参照する方法の調査】

PostgreSQL のアドオンツール (s3csv_fdw (フリーソフト)) で、s3 上の CSV ファイルをテーブルとして参照する機能が存在したが、Windows 非対応のため、Windows 環境では利用できなかった。

(3) EC2 instance におけるデータベース構築 (Linux 環境)

【s3 に対する SQL 実行 (S3_fdw)】

Postgres の拡張機能である Foreign Data Wrapper (fdw) を利用し、s3 に直接アクセ

ス可能かを検証したが、権限なしエラーが発生し実行できなかった。

【PostgreSQL によるデータ圧縮 (citus)】

Postgres の拡張機能である Citus を用いて、実現性と圧縮可否について検証を行った。テストデータにおいて、圧縮有と無のテーブルを比較し、問題なく圧縮されていることを確認した後、実際のデータを用いて圧縮率の検証を行った。Citus を用いて 4 年分の NDB について圧縮を行ったところ、全体で約 20TB のデータを Postgres 上において約 4TB までの圧縮が可能になることが分かった。この作業により、分析用の領域の確保が可能になった。一方で、INDEX については圧縮を行うことができないため、複数の利用者が同一領域を利用する際には特に注意が必要なことも分かった。

D. 考察

【ストレージの課題について】

今回 HIC の試行利用にあたり、当初 Windows 環境での分析環境構築を試みたが、ストレージの問題から EC2 instance における PostgreSQL データベース作成ができなかった。このため容量不足に対するいくつかの検討を実施したが、Windows 環境下では分析に十分な容量の確保はできなかった。次いで Linux 環境での検討を行った。Citus を利用することで、PostgreSQL のデータを圧縮することが可能となり、一定の分析用領域の確保が可能となったが、実際に分析を行う場合には、INDEX については圧縮を行うことができないことなど、圧縮後も容量超過に対していくつかの注意が必要なが分かった。今回は試行利用としてストレージについても通常より多い 16TB の

環境において検討を行ったが、データベース構築および分析用の領域の確保には様々な工夫を要した。実際に HIC で分析を行う際は提供されるストレージの容量内で分析可能なデータサイズを認識しておく必要がある。また、今回のように複数チームで領域をシェアすることや、解析中に発生する中間生成物の規模を考えると、ストレージについてはある程度余裕があることが望ましい。

【解析ソフトウェア環境について】

今回、データベース構築に時間を要し、実際のデータ分析は十分に行えなかったが、今後実際に分析を行うことを想定すると、解析ソフトウェア環境についても、singularity 等のコンテナが使える環境が望ましいと考える。(R のパッケージのインストールについてはパッケージ間の依存関係があるため)

【利用者への情報提供について】

今回の試行利用を行った経験から、現状において、このシステムを適切に運用するためには、作業にコミットできる情報システムの高度な知識を持った者が 1 人は必要になると考える。また、今回行ったデータ圧縮等、HIC 利用におけるノウハウについても利用者に対する一定の情報提供が必要と考える。

HIC は AWS 上に構築されており、利用状況に応じてクラウド利用料が発生するシス

テムである。このため、利用者に対してはどの処理に対して費用がかかるかなどの説明がある方が良いと考える。この知識がないと利用方法によっては多大な費用が生じてしまい、HIC 運用が立ち行かなくなる可能性が懸念される。利用者リテラシー向上の観点からも費用に関する情報提供も重要と考える。

E. 結論

HIC における NDB 特別抽出データ分析環境構築の課題を抽出し、対応策の検討を実施した。HIC においてある程度大きなデータを分析する場合はストレージ容量を考慮した対応等が必要となり、情報システムの高度な知識を持った利用者の参加が必要になると考える。

F. 健康危険情報

なし

G. 研究発表

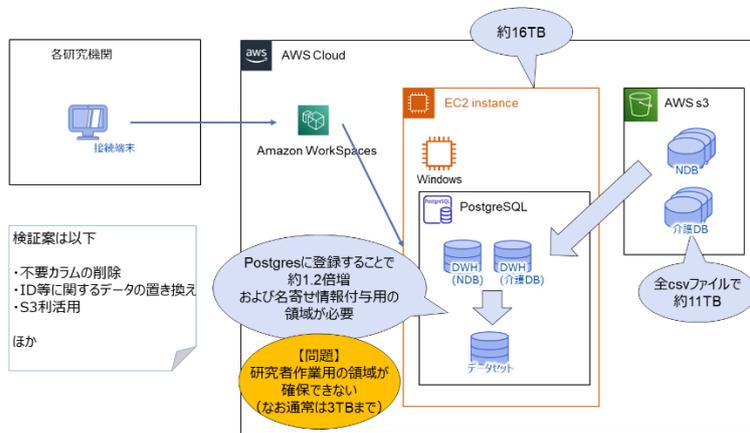
なし

H. 知的財産権の出願・登録状況

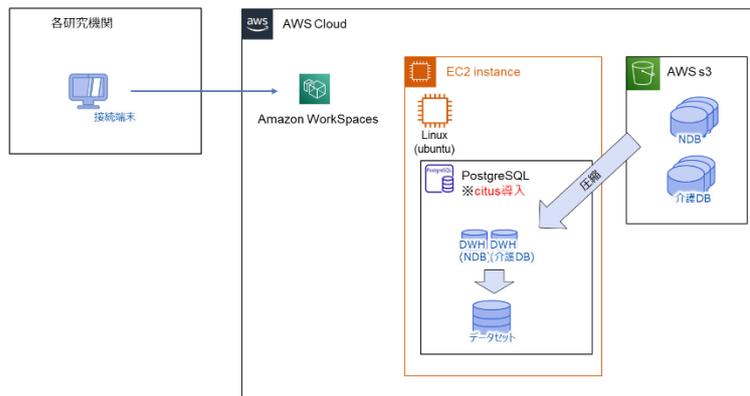
なし

図1 HIC 利活用に関する問題（Windows 環境）

HIC利活用に関する問題（Windows環境）



PostgreSQLによるデータ圧縮（citius）



別添5

研究成果の刊行に関する一覧表

書籍：対象なし

雑誌：対象なし

厚生労働大臣
—(国立医薬品食品衛生研究所長)— 殿
—(国立保健医療科学院長)—

機関名 京都大学
所属研究機関長 職名 医学研究科長
氏名 伊佐 正

次の職員の令和5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 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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究
3. 研究者名 (所属部署・職名) 医学部附属病院・医療情報企画部・講師
(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 checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	京都大学	<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
—(国立医薬品食品衛生研究所長)— 殿
—(国立保健医療科学院長)—

機関名 京都大学

所属研究機関長 職名 医学研究科長

氏名 伊佐 正

次の職員の令和5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 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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究

3. 研究者名 (所属部署・職名) 医学部附属病院・病床運営管理部・病院教授

(氏名・フリガナ) 加藤 源太 ・ カトウ ゲンタ

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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	京都大学	<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
—(国立医薬品食品衛生研究所長)— 殿
—(国立保健医療科学院長)—

機関名 京都大学

所属研究機関長 職名 医学研究科長

氏名 伊佐 正

次の職員の令和5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 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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究

3. 研究者名 (所属部署・職名) 医学部附属病院・医療情報企画部・教授

(氏名・フリガナ) 黒田 知宏 ・ クロダ トモヒロ

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 checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	京都大学	<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。

令和6年5月15日

厚生労働大臣
(国立医薬品食品衛生研究所長) 殿
(国立保健医療科学院長)

機関名 京都大学

所属研究機関長 職名 国際高等教育院長

氏名 大嶋 正裕

次の職員の令和5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 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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究

3. 研究者名 (所属部署・職名) 京都大学国際高等教育院 付属データ科学イノベーション教育研究
センター・特定講師

(氏名・フリガナ) 植嶋 大晃 ・ ウエシマ ヒロアキ

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"/>	京都大学	<input type="checkbox"/>
遺伝子治療等臨床研究に関する指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
その他、該当する倫理指針があれば記入すること (指針の名称:)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>

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

その他 (特記事項)

(※2) 未審査に場合は、その理由を記載すること。

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

5. 厚生労働分野の研究活動における不正行為への対応について

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

6. 利益相反の管理

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

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

令和6年5月31日

厚生労働大臣 殿

機関名 国立研究開発法人
国立長寿医療研究センター

所属研究機関長 職 名 理事長

氏 名 荒井 秀典

次の職員の令和5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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究
3. 研究者名 （所属部署・職名） 老年学・社会科学研究センター 医療経済研究部 副部長
（氏名・フリガナ） 大寺 祥佑 （オオテラ ショウスケ）

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"/>	国立長寿医療研究センター	<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。

令和6年2月1日

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5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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究

3. 研究者名（所属部署・職名） 公衆衛生学講座・教授

（氏名・フリガナ） 今村 知明・イマムラ トモアキ

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"/>	奈良県立医科大学	<input type="checkbox"/>
遺伝子治療等臨床研究に関する指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
その他、該当する倫理指針があれば記入すること (指針の名称：)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>

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

その他（特記事項）

（※2）未審査に場合は、その理由を記載すること。

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

5. 厚生労働分野の研究活動における不正行為への対応について

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

6. 利益相反の管理

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

（留意事項） ・該当する□にチェックを入れること。

・分担研究者の所属する機関の長も作成すること。

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5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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究

3. 研究者名（所属部署・職名） 公衆衛生学講座・准教授

（氏名・フリガナ） 野田 龍也・ノダ タツヤ

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 checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	奈良県立医科大学	<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。

令和6年4月10日

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5年度 厚生労働行政推進調査事業費補助金 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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究

3. 研究者名 (所属部署・職名) 大学院医学系研究科・教授

(氏名・フリガナ) 康永 秀生・ヤスナガ ヒデオ

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)	<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。

令和 6年 4月 10日

厚生労働大臣 殿

機関名 国立大学法人筑波大学

所属研究機関長 職名 学長

氏名 永田 恭介

次の職員の令和5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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究 (21AA2006)
3. 研究者名 (所属部署・職名) 医学医療系/ヘルスサービス開発研究センター・教授/センター長
(氏名・フリガナ) 田宮菜奈子・タミヤナナコ

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 checked="" type="checkbox"/>	筑波大学・京都大学	<input type="checkbox"/>
遺伝子治療等臨床研究に関する指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
その他、該当する倫理指針があれば記入すること (指針の名称:)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>

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

その他 (特記事項)

(※2) 未審査に場合は、その理由を記載すること。

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

5. 厚生労働分野の研究活動における不正行為への対応について

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

6. 利益相反の管理

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

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

令和6年4月22日

厚生労働大臣 殿

機関名 国立研究開発法人国立国際医療研究センター

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5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 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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究

3. 研究者名 （所属部署・職名）研究所 糖尿病情報センター・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
—(国立医薬品食品衛生研究所長) 殿
—(国立保健医療科学院長)—

機関名 京都大学

所属研究機関長 職名 医学研究科長

氏名 伊佐 正

次の職員の令和5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 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 臨床疫学に活用可能な NDB 等データセットの作成に関する研究

3. 研究者名 (所属部署・職名) 大学院医学研究科 ・ 教授

(氏名・フリガナ) 中山 健夫 ・ ナカヤマ タケオ

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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	京都大学	<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。