

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))
総合研究報告書

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

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

研究要旨

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。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」ことを目的に、

- 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の検討
- ③ 臨床系研究者でも利用可能なデータセットの開発

を実施した。また研究の一部を医療・介護データ等の解析基盤（HIC）上で実施し、

- ④ HICの安全性と操作性について検討を行った。

①については、すでに公表されているNDBコードブックを参考に、介護DBのデータについて網羅的な調査を行い、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、介護DBコードブックとして公表した。②についてはまずNDBにおける入院データと外来データを結合し、1患者の入院・外来の情報を追跡できる「1患者1データ化」を行った。次に医療レセプトと介護レセプトの連結可能性と連結時の課題を整理する基礎として、KDBデータを用い、サンプル抽出を行って課題を整理した。さらに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い、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についてはまず学術研究に対応可能なデータセットの仕様について検証するためにNDBおよび自治体医療介護データを用いて学術研究領域の複数のリサーチクエストに対する分析を実施した。その後、上記検討結果を踏まえてNDBと介護DBの連結データの分析練習に使用可能なNDB・介護DB連結データのダミーデータセットを作成した。④については研究の一部をHICの試行利用として実施し、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。

本研究の成果により、以下の効果が期待される。

- ・ NDB・介護DBデータの基本構造、データ項目、データ格納状況に関する情報を公開することにより、利用者は利用するデータの内容を理解したうえで研究計画の立案・分析が可能となり、効率的なデータ利用の一助となることが期待される。
- ・ 長期追跡性に優れた個人名寄せIDについて、id5の長期追跡性を評価し、既存のidとのより精緻なデータ連結の可能性を検討した。この結果は異なるデータベース間の連結や長期追跡を要するデータ分析の支援となることが期待される。
- ・ 学術研究に対応可能なデータセットの仕様について実際の分析を通して検討を行った。また検討結果をふまえて分析練習に使用可能なNDB・介護DB連結データのダミーデータセットを作成した。今後、利用者支援の一環となることが期待される。
- ・ HIC利用における課題の抽出及び対応法について検討を行った。今後、研究者がHICを利用する際の一助となることが期待される。

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

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。コロナ禍に伴うパンデミック時の対応や、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。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の検討
 - 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
- を行うことを目的とした。

B. 研究方法

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4つの課題について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
- NDBデータと介護DBデータの連結解析を想定し、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・集計を行った。NDBデータの詳細についてはNDBコードブックとしてすでに公開されていることから、連結データ分析に円滑に進めるため、同様のコードブックを介護DBデータに対しても作成することとした。介護DBデータの詳細検討を行うにあたり、まず以下の手順でデータウェアハウス(DWH)の構築を行った。
- 1)DWH の設計について検討
 - 2)介護DB データ内容の確認、データ読み込み
(2014年4月から2016年3月までの介護DB データを利用)
 - 3)DWH 作成
- その後、構築したDWHを用いて、介護DB データブック作成した。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。

②長期追跡性に優れた個人名寄せIDの検討

【1患者1ID化】

NDBデータの構造を再構築し、個人識別変数をより高精度化し、患者のレセプト情報を追跡する方法を検証した。第1段階は入院データのトレース、第2段階は外来データのトレース、第3段階は入院と外来の連結を実施した。第1段階ではDPC出来高・医科入院データをCDファイル化し、DPC包

括と結合し、連続入院を1つのシーケンスに変換した。入院は、id0、医療機関コード、病棟分類別に変わるものとした。第二段階では調剤レセプトと医科外来レセプトを、id0、医療機関コード、処方日をキーに結合した。第3段階では入院と外来をid0で結合した。これで1患者1データ化が完了した。そして、1患者1データ化後のデータベースをもとに、日本の保険診療の実態を調査した。

【id5の長期追跡性の検討】

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。また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。具体的には、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。その上で以下について検討を行った。

①2022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。

②2022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。

③id5とid1Nの対応割合を調べた。

④③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。

⑤2022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。

⑥2013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。

③臨床系研究者でも利用可能なデータセットの開発

【臨床系研究に利用可能なデータセットの要件検討】

臨床系研究者でも利用可能なデータセットの開発にあたり、実際のデータ分析の実施を通してデータ分析上の課題の整理を行った。具体的には以下の臨床疫学研究を実施した: (詳細は各年度の報告書にて報告)

A) 薬剤関連顎骨壊死の有病割合、発生率、リスクファクターの探索: NDBを用いて、後ろ向きコホート研究を実施した。2015年4月から2018年12月に新たに骨吸収抑制剤を使用した患者を対象とした。主要アウトカムはMRONJの発症である。時間依存性Cox比例ハザード回帰分析を行い、MRONJ発症のリスク因子を検討した。

B) 要介護状態の高齢者に対する人工呼吸管理の発生率と1年後転帰: 栃木県医療介護レセプトデータベースを用いて、2014年6月から2018年2月までの65歳以上登録者のうち、挿管下人工呼吸管理を受けた患者を同定した。人工呼吸管理の発生率と1年後転帰を人工呼吸管理前の要介護度によって層別して評価した。

C) 日本における脆弱性大腿骨近位部骨折の医療費の分析: NDBを用いて日本における脆弱性

大腿骨近位部骨折に係る医療費の推定を行った。

D) 誘因イベント後の交感性眼炎の発生率の推計: NDBデータを用いて2012年から2019年の間に誘因イベント(inciting event)を経験した患者を対象とし、Cox比例ハザード回帰分析により、交感性眼炎発症のリスク因子を検討した。

E) 栃木県医療介護連結データを用いた要介護高齢者における心肺蘇生後の1年後の機能的転帰の推計: 2014年6月から2018年2月までに心肺蘇生を受けた65歳以上患者を同定した。アウトカムは1年の死亡率および要介護度とした。

F) 在宅医療を提供する医療機関種別と往診、入院、看取りの関連: NDBデータを用いて在宅医療を提供する医療機関種別の医療機能のアウトカムを評価した。65歳以上で、2014年7月から2015年9月の間に新たに訪問診療を開始し、一般診療所、従来型在宅療養支援診療所/病院(在支診/病)、機能強化型在支診/病を利用した160,674人を特定した。追跡機関を6か月間とし、緊急往診、入院、在宅死をアウトカムとして共変量を調整した上で多変量ロジスティック回帰分析を行った。

G) 在宅療養支援診療所/病院の心不全再入院に対する予防効果の分析: NDBデータを用いて、2014年7月から2015年9月の間に心不全で入院し、退院後1か月以内に一般診療所、従来型在支診/病、機能強化型在支診/病のいずれかによる訪問診療を受けた65歳以上12393人を特定した。主要アウトカムを退院後6か月以内の再入院(初回)、副次アウトカムを緊急往診(初回)とし、交絡因子を調整した上でFine and Gray法を用いた競合リスク回帰分析を行った。

H) 日本の医療機関における心臓リハビリテーション(CR)の実施実態と関連要因: NDBデータを用いて日本の病院における心筋梗塞患者に対する心臓リハビリテーション(CR)の実施状況について記述研究を行なった。

【汎用性のあるデータセットの検討】

汎用性と簡便性の観点から

- 1)一定の患者数が報告されている疾患であること
- 2)傷病名コードの選定の容易であること
- 3)診療行為(検査含む)コード選定が容易であること

を前提に、対象疾患を選定した。選定した疾患を対象に、データセットを作成する際に使用する

- ・傷病名コード
- ・疾患特有の診療行為(検査含む)コード
- ・疾患特有の医薬品コード

について検討を行った。

【NDBと介護DB連結分析練習用データセットの検討】

NDBと介護DBの連結について、医療レセプトと介護レセプトの連結可能性と連結時の課題を整理する基礎として、某県のKDBデータを用いて課題を整理した。

さらに、NDBと介護DB連結データ分析の練習を想定したダミーデータセットの仕様を検討し作成した。

データセットの利用方法については

- 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

を中心に検討を行った。

その後、以下の手法でNDB・介護DB連結ダミー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
- 3)ID0で連結したデータセットを作成

④HIC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以下の観点から比較検討を行った。

- 1)利用者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

さらに特別抽出データをHIC上で分析する上での課題と対策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。

②長期追跡性に優れた個人名寄せIDの検討

【1患者1ID化】

1患者1データ化による変換後のデータの信頼性を確認するため、1患者1データ化後のデータベースをもとに、日本の保険診療の実態を調査した。2015年4月から2016年3月の全患者数、入院患者数/平均入院日数/平均入院回数/平均再入院回数、外来患者数/外来総日数/平均外来日数/時間外患者数/時間外日数/平均時間外外来日数を算出した。さらに急性心筋梗塞に対するステント留置術後の入院外来日数を算出した。

【id5の長期追跡性の検討】

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。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。

③臨床系研究者でも利用可能なデータセットの開発

【臨床系研究に利用可能なデータセットの要件検討】において実施した臨床疫学研究の結果の概要は以下の通りであった。（詳細は各年度の報告書にて報告）

- A) 研究対象患者は2,819,310人であり、2,664,104人（94.5%）が骨粗鬆症、155,206人ががんであった。骨粗鬆症患者では1,603人（0.6%）、がん患者では2,274人（1.47%）がMRONJを発症した。MRONJの発症は、口腔内の状態不良（抜歯を含む）、年齢、男性、薬剤の種類、併用薬剤、併存疾患、がんの種類、地域と関連していた。
- B) 65歳以上登録者593,990人のうち、6,352人（1.1%）が挿管下人工呼吸管理を受けた。人口10万人当たりの人工呼吸管理の年間発生率の平均は、全体で322人、要支援1では659人、要介護5では1685人であった。人工呼吸管理から1年以内の死亡率は、支援なし、要支援1-2、要介護度1-3、要介護度4-5の患者で、それぞれ54.7%、56.5%、70.6%、76.8%であった。人工呼吸管理後に要介護度が改善した症例はほとんどいなかった。
- C) NDBデータから脆弱性大腿骨近位部骨折の患者142,361人（男性28,868人、女性113,493人）を特定した。1人当たりの平均医療費は、男性で2,550,000円、女性で2,494,000円であった。脆弱性大腿骨近位部骨折の年間総医療費は3,292億円であり、男性では679.6億円、女性では2,612.4億円（23.7億米ドル）であった。これらの結果から、股関節脆弱性骨折は、日本社会に大きな医療経済的負担を課していることがわかった。
- D) 合計888,041件の誘因イベント（704,717人）が同定された。交感性眼炎の総症例数は263例であり、60ヶ月間の交感性眼炎の累積発生率は0.044%であった。40歳から49歳のグループは、最も高い0.104%の発症率を示した。
- E) 65歳以上登録者594,092人の対象者のうち、5,389人（0.9%）が心肺蘇生術を受けた。要介護度なし、要支援1~2、要介護度1、要介護度2~3、要介護度4~5の患者における心肺蘇生後の1年死亡率はそれぞれ89.8%、89.2%、89.8%、89.9%であった。
- F) 一般診療所：13,477人、従来型在支診/病：64,616人、機能強化型在支診：82,851人を含む160,674人が解析対象となった。一般診療所と比較して、従来型在支診/病、機能強化型在支診では、追跡調査中に往診を受けた患者の割合が高かった一方（一般診療所：36.1%、従来型在支診/病：48.5%、機能強化型在支診/病：53.1%）、入院した患者の割合は低くなって

いた（一般診療所：27.0%、従来型在支診/病：24.6%、機能強化型在支診/病：25.7%）。

- G) 一般診療所群1,009人、従来型在支診/病群4,681人、機能強化型在支診/病群6,703人を含む12,393人が分析対象となった。6ヶ月の追跡調査終了時に、一般診療所の43.0%（434/1,009）、従来型在支診/病の38.1%（1,784/4,681）、機能強化型在支診/病の38.1%（2,556/6,703）の患者が少なくとも一度は再入院していた。6ヶ月間の死亡率は、一般診療所群で24.0%（242/1,009）、従来型在支診/病群で27.0%（1,267/4,681）、機能強化型在支診/病群では29.6%（1,982/6,703）であった。
- H) NDBデータから入院CRは813施設から35,298人、外来CRは799施設から33,328人の患者を特定した。医療機関レベルの入院および外来CR実施割合の中央値は、それぞれ73.3%および1.8%であった。入院CR実施割合の分布は二峰性であり、入院・外来CR実施割合のジニ係数は、それぞれ0.37および0.73であった。医療機関レベルのCR実施割合にはいくつかの要因で統計的な差があったが、CR実施の分布に影響を与える視覚的に明らかな要因は、CR算定認証だけであった。入院および外来CR実施は医療機関レベルでの実施不均等がみられ、外来CRの実施割合は低い医療機関が大半であった。

【汎用性のあるデータセットの検討】

汎用性の観点から、乳がん、消化器がん、感染症を対象疾患として検討を行った。乳がん、消化器がん、感染症は対象id数が多く、診療行為、医薬品の情報等様々なデータ項目を維持したデータセットを作成した場合、データサイズが非常に大きくなることが分かった。このため様々な検討を可能とする汎用性のあるデータセットを作成する場合、簡便さにおいて課題があることが分かった。

【NDBと介護DB連結分析用サンプルデータセットの検討】

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連結ダミー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)介護度、経管栄養の有無毎に集計用テーブルを作成

④ HICの安全性及び操作性の検証

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、その特性の違いを比較し、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cituz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。

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。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られる。

長期追跡可能なIDの検討においては、まずNDBデータにおいてid0を用いた1患者1データ化の手法が開発した。このことによって、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た。さらに個人単位被保険者番号に基づく個人名寄せIDである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。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。また、汎用性のあるデータセットについても検討を実施したが、一定の患者数が報告されている疾患を対象に様々な分析を想定したデータ項目（診療行為や薬剤）を有するデータセットを作成した場合、データサイズが非常に大きくなることは明らかなため比較的「簡便な」データセットにするためには取り扱いやすいデータサイズにダウンサイジングする必要がある。ランダムサンプリングによるダウンサイジングは比較的簡便であるが、NDBの特徴である悉皆性は失われることになる。分析項目に一定の条件を付けた場合（対象診療行為を薬物療法のみとする、検査のみとする等）についても検討を行ったところデータ項目を限定することで、分析できる内容は制限されるが、

- ・乳がん患者における薬剤Aの使用状況
- ・大腸がん患者におけるロボット手術の都道府県別実施件数

等の簡単な分析であれば可能と考えられた。汎用性を考慮した場合は、データサイズが巨大になり、簡便性を考慮した場合は、分析できる内容には制限が生じるため、ある程度利用目的を設定したうえでの設計が必要になると考えられた。これらの検討結果をふまえて、NDB・介護DB連結データの利用者支援を想定した連結データの分析練習ができるダミーデータセットの開発を行った。簡便に使用可能なダミーデータセットにより分析練習が可能になれば、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。今回分析環境を構築する際の課題としてストレージ不足の対応に時間を要した。データ圧縮等である程度の対応は可能であるが、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、今回の試行利用を通して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。

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個人単位被保険者番号由来の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、NNDB・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

1. 論文発表

- ① 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 (MRONJ) in patients with osteoporosis and cancer: a nationwide population-based study in Japan. *J Oral Maxillofac Surg.* 2021;S0278-2391(21)01511-1.
- ② Mori T, Komiyama J, Fujii T, Sanuki M, Kume K, Kato G, Mori Y, Ueshima H, Matsui H, Tamiya N, Sugiyama T. Medical expenditures for fragility hip fracture in Japan: a study using the nationwide health insurance claims database. *Arch Osteoporos.* 2022 Apr 11;17(1):61.
- ③ Hashimoto Y, Matsui H, Michihata N, Ishimaru M, Yasunaga H, Aihara M, Kaburaki T. Incidence of sympathetic ophthalmia after inciting events: a national database study in Japan. *Ophthalmology.* 2022;129(3):344-352.
- ④ Ohbe H, Nakajima M, Miyamoto Y, Shibahashi K, Matsui H, Yasunaga H, Sasabuchi Y. One-year functional outcomes after cardiopulmonary resuscitation for older adults with pre-existing long-term care needs. *Age and Aging.* 2023;52(5):afad072
- ⑤ Sun Y, Iwagami M, Komiyama J, et al. The Effect of Home Care Support Clinics on Hospital Readmission in Heart Failure Patients in Japan. *J Gen Intern Med.* 2023; 10.1007/s11606-023-08030-9.
- ⑥ Sun Y, Iwagami M, Komiyama J, et al. Association between types of home healthcare and emergency house calls, hospitalization, and end-of-life care in Japan. *J Am Geriatr Soc.* 2023;10.1111/jgs.18268.
- ⑦ Komiyama J, Sugiyama T, Iwagami M, et al. Hospital-Level Variation in Cardiac Rehabilitation After Myocardial Infarction in Japan During Fiscal Years 2014–2015 Using the National Database, *Circulation Reports,* 2023 May 11; 5 (5): 177-186

2. 学会発表

- ① 小宮山潤、杉山雄大、岩上将夫、石丸美穂、松居宏樹、小山照幸、田宮菜奈子：医療機関における急性心筋梗塞術後の心臓リハビリテ

ーション実施の不平等と関連要因、第80回日本公衆衛生学会総会、2021.12

- ② 孫瑜、岩上将夫、小宮山潤、杉山雄大、井口竜太、佐方信夫、伊藤智子、吉江悟、讃岐勝、久米慶太郎、松居宏樹、植嶋大晃、森由希子、加藤源太、田宮菜奈子：全国レセプトデータを用いた在宅療養支援診療所の心不全再入院に対する予防、第81回日本公衆衛生学会総会、2022
- ③ 松居宏樹.NDBオンサイトセンターと臨床疫学研究基盤の構築～NDBダミーシステムの作成～. 第5回NDBユーザー会. 2022年9月. 東京.
- ④ 植嶋大晃、森由希子、加藤源太、中山健夫、今村知明、康永秀生、田宮菜奈子、杉山雄大、大寺祥佑、山本喜一、松岡歩、岩淵奈穂子、黒田知宏：介護DBを用いた基礎的集計および「介護DBコードブック」の作成、第82回日本公衆衛生学会総会、2023

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