

令和3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 准教授
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系研究科 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。レセプト情報等データベース（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、レセプト情報・特定健診等情報（NDB）、および要介護認定情報・介護レセプト情報（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の（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。2022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た。
- ② については、NDBと介護DBの連結可能なサンプルデータの作成を見据えた上での、主にKDBを用いたデータベースの試行的な処理を実施した。
- ③ については、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とともに、複数のリサーチクエストに対応可能な汎用性のあるデータセットの仕様について検討を行った。
- ④ について実際のHICの利用を開始するとともに、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、調査結果としてコードブックを作成した。
- ② 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。
- ③ 実際の分析に必要なデータセットの検討のため、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チョンに対する分析を実施した。上記結果を踏まえて汎用性のあるデータセットの仕様検討を行った。
- ④ HICにおける大規模データ利用の課題を検討した。

結論：介護DBコードブックについては、2022年度内には完了した。また、長期追跡可能なIDの検討及びデータセットの検討についてはほぼ当初の予定通り検討を進めており、2023年度も引き続き継続予定である。

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、レセプト情報等データベース（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3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）」が稼働される予定であることから、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。

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4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

NDBデータと介護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た。

②長期追跡性に優れた個人名寄せIDの検討

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、サンプル抽出を行って課題を整理した。

③臨床系研究者でも利用可能なデータセットの開発
実際のリサーチクエスチョンに対してどのようなデータセット作成が必要になるかについて、実際にいくつかの分析を行って検討した。また、比較的簡便かつ汎用性のあるデータセットについて、

- ・対象疾患
- ・対象となるデータ抽出項目
- ・想定される分析内容

等について仕様の検討を行った。

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比較検討を行った。

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。

②NDBと介護DBの連結可能なサンプルデータの作成を見据えた上での、主にKDBを用いたデータベースの試行的な処理を実施した。

③実際の分析に必要なデータセットの検討のため、NDBおよび自治体医療介護データを用いて

- ・誘因イベント後の交感性眼炎の発生率の推計
- ・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
- ・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
- ・心不全退院後の在宅医療を受ける患者において医療機関種別による再入院のリスクの比較
- ・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

について分析を行った。
また、汎用性の高いデータセットについて仕様検討を進めた。

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オンサイトリサーチセンターでの利用との比較検討を行った。

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）。2022年度は介護DBデータとのデータ連結を念頭に、KDBデータを用いて介護データと連結する際の課題の整理を実施した。

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、実際にいくつかの研究クエスチョンについて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。（個々の分析については分担研究報告書内に記載。）また、実際の分析経験を踏まえて「汎用性のあるデータセット」の仕様について検討を行った。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。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。

HICの試行利用においては、オンサイトリサーチセンターとの比較検討を実施した。両者ともDaaS（Data as a Service）環境下でのデータ利用であるが、オンサイトリサーチセンターとHICではアクセス経路と利用可能な環境が異なっており、HICで大規模データ解析を実施するには拡張機能のインストール等、様々な工夫が求められることが示唆された。

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について検討を行い、連結データを利用する際の知見が広がった。

③特定の研究クエスチョンを解明するためには目的に沿ったデータセットの作成が必要になることが改めて確認された。その結果を踏まえて「汎用性のあるデータセット」について仕様検討が進んだ。

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 あり
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