

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	国立保健医療科学院 保健医療経済評価研究センター 主任研究官
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。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

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。2021年度において

①については、介護DBのデータについて網羅的な調査を行い、データブック作成用データウェアハウス(DWH)を構築、各種集計作業を実施した。

②については、2013年4月から2016年3月までのNDBに含まれるレセプト（歯科レセプト及び全額公費レセプトを除く）を対象とし研究を実施した。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とした。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、名寄せ精度を向上させた名寄せID（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。

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ン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。

結果：①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入院と外来を結合し、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」を行い、変換後のデータの信頼性を確認するため、外来患者総数および平均外来患者数を算出した。

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トンに対する分析を実施した。また、汎用性のあるデータセットの仕様検討のため、疾患群によるデータセットのプロトタイプを検討を行った。

結論：介護DBデータブックについては、サーバ故障による進捗遅延を認めたが、2022年度内には当初の目的を完了予定である。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、2022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」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。

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。

## 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3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データブックの作成を進めた。

### ②長期追跡性に優れた個人名寄せIDの検討

NDBデータの長期追跡性に優れた個人名寄せIDについて検討を行った。

### ③臨床系研究者でも利用可能なデータセットの開発

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の仕様について、

- ・想定される利用者
  - ・想定される利用形態
  - ・取り扱いが可能なデータサイズ
- 等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DWHの作成を行った。DWH作成において、提供された介護DBデータの項目等について調査し、データ定義、エンティティ一覧、エンティティ定義書、テーブルリストを作成した。また、DWHの作成と並行して、介護DBデータブックの仕様検討を行い、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データにおいて入院と外来を結合し、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」に対し、データの信頼性を調べるため、入院および外来の患者数、入院/受診日数を求めた。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。また、1患者1データ化に対し、技術的な1例を示すために、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示した。

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

- ・薬剤関連顎骨壊死（MRONJ）の有病割合、発生率、リスクファクターの探索
  - ・自治体医療介護連結データを用いた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
  - ・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
  - ・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（心リハ）の実施状況
- について分析を行った。

また、比較的汎用性の高いデータセットについて仕様検討を行い、プロトタイプを作成を進めた。

## D. 考察

介護DBのデータブック作成は2021年度で終了予定であったが、サーバの故障等、予定外の事象のため、作業の遅延を認めた。現在、サーバは復旧しており、2022年度前半での完成を予定している。完成後は公表に向けて手続きを進める予定である。NDBデータと同様に介護DBデータについても、データブックを作成、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、1患者1データ化により、NDBデータ内におけるの追跡調査・コホート調査が可能になった。

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、実際にいくつかのリサーチ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。）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対応が困難とな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も示唆された。今後、汎用性のあるデータセットについては、複雑な学術研究での使用ではなく、「臨床現場における比較的シンプルなクリニカルクエスチョン」や、「症例数の定期報告等現場の業務」に対応可能なものとするものの有用性について検討予定である。また、汎用性のあるデータセットの一案として、今年度は「疾患別」データセットについて検討を行った。乳癌と脳血管疾患を対象にプロトタイプを作成を試みたが、疾患名でデータを抽出すると「疑い病名」等を除外しても患者ID数が想定以上に膨大となることが分かった。データの取扱いの簡便さを考慮するとダウンサイジングが必要となる。ランダムサンプリング等によるダウンサイジングを検討しているが、NDBデータの悉皆性は失われることとなるため、分析結果の有用性や研究者のニーズを考慮した対応が必要となる。どのようなダウンサイジングが適当であるかについては今後さらに検討を進める予定である。

#### E. 結論

①介護DBデータのデータブックは介護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と考えられる。

②1患者1データ化がほぼ完了し、NDBデータにおける患者単位の追跡調査・コホート調査が可能になった。

③特定のリサーチ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。これらの結果をふまえて、「研究者個々のニーズに答えるデータセット」の仕様については、「汎用性のあるデータセット」とは分けて検討を進める必要があると考えられた。「汎用性のあるデータセット」の作成においては、利用目的や利用者像の設定を検討した上で、開発を進めることが必要と考えられた。

#### 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