

令和元年～2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))
総合研究報告書

医療・介護のデータの利活用の推進のための、
NDB・介護DBの連結可能性および
活用可能性の評価に関する研究

研究代表者 加藤 源太 (京都大学医学部附属病院 准教授)
研究分担者 黒田 知宏 (京都大学医学部附属病院 教授)
研究分担者 植嶋 大晃 (京都大学医学部附属病院 助教)
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)

研究要旨

【研究目的】

本研究は、匿名レセプト情報等データベース (NDB)、および匿名要介護認定情報・匿名介護レセプト等情報 (介護 DB) の両者のデータ連結について技術的検証を行うとともに、連結データが保健医療分野における政策や研究のエビデンスを導き出せる質を備えることができるのか、こういった課題に強みを発揮できるのかを評価するものである。この点について、NDB・介護DBの連結可能性の評価と、連結データの活用可能性の評価という2つの課題に取り組むべく、介護DBの概要と構成要素について整理を行い、これまでの医療・介護データの連結解析の議論や、両データを連結して行った先行研究を参考に、NDB データと介護DB データの連結解析を容易にするデータ構成を検討し、連結解析に使用できる新しいテーブルの作成を行った。次に、活用可能性の評価としては、作成したデータマートを用い、先行研究を再現する形式で非常に簡単なレベルの試行集計を行った。その際に、どの程度の作業時間を要したかについても、確認を行った。一方で、NDB そのものの価値を向上させることを目的として、NDB レセプトデータを用いて死亡を予測する深

層学習モデル(deep learning model)の開発及びバリデーションを行った。また、医療レセプトと要介護認定情報・介護レセプト情報とを連結解析した事例そのものがまだ非常に少ないという現状を鑑み、NDBのみで実施でき、かつ今後のNDB・介護DB連結解析で取り上げられて展開していく可能性の高い、高齢者の介護の課題に直結する個別研究も行った。

【研究手法】

今回の研究では、既存のNDBデータ分散処理基盤を活用し、すでに格納されている2014年、2015年度分NDBに追加する形で2014年度、2015年度介護DBデータを受領し、全く偽のハッシュ値を介護DBに付与した上でNDBと介護DBを形式的に連結できる仕組みを構築し、両データを連結した際の操作性、抽出速度、有用性等を調査して連結可能性を評価した。NDBを用いた個別研究もこのデータにて実施している。また、在院死亡予測モデルの開発に際しては、DPCデータ調査研究班のDPCレセプトデータにおいて2010年7月-2017年3月の期間に退院した全患者データを用いた。

【研究結果】

014～15年度分の介護DBデータを入手し、まず1か月分の介護DBデータにて、試行的なDWH構築及び課題の洗い出しを行った。そのうえで、既に構築済みのNDB基盤に付与されているID0を連結キーとして要介護認定情報・介護レセプト等情報の各テーブルの末尾に格納し、NDBとデータを連結する際に使用できるよう構成した。一方、介護DBに相当するデータを用いた研究をレビューした限りでは、得られた情報を患者のアウトカムとして使用したり、対象者の状態像を示す情報として使用する事例が多くみられた。これらの情報は要介護認定情報に含まれることが少なくないが、レセプト情報とは異なり、要介護認定情報は認定時にしか情報が格納されない。そこで、要介護認定情報を各月ごとに参照できるテーブルがあれば、介護レセプト情報のみならずNDBのレセプト情報とも連結した分析が行いやすくなるを考え、これら情報を新たなテーブルとして作成した。これらのデータにもID0を格納して作成したデータマートを用い、医療レセプト、要介護認定情報、介護レセプト情報間で連結解析を行っている2つの先行研究に着目し、その再現を試みた。集計は合計3通り実施し、いずれも簡略化した集計ではあったが、全数レベルでNDB・介護DBの連結解析を実施することができた。

深層学習モデルを用いたモデルでは、全症例のAUCを評価したところ、結果は0.954(95%信頼区間: 0.9537-0.9547)と高く、急性心筋梗塞・心不全・脳卒中・肺炎の4疾患について既存の重症度指標を用いたモデルも作成して比較したところ、4疾患のいずれにおいても、メインモデルの方がAUCの値は高値であり、レセプトデータの初日の治療内容情報を用いた深層学習モデルで、高い精度の在院死亡の予測を

得られることが確認できた。

NDB を用いた個別研究では、高齢者のリハビリ、在宅医療の実態を NDB だけでも相当程度明らかにできることを示したとともに、今後 NDB と介護 DB の連結解析が容易になれば、患者像の評価に資する情報が付加されることから、得られる知見の幅が広がりうることを確認することができた。

今後、NDB と介護 DB の連結解析を行う基盤が整えられていく見込みであることから、本研究で示したような連結解析に関する必要な知見が、利用者間で更に共有されていくことが期待される。

A. 研究目的

本研究は、匿名レセプト情報等データベース (NDB)、および匿名要介護認定情報・匿名介護レセプト等情報 (介護 DB) の両者のデータ連結について技術的検証を行うとともに、連結データが保健医療分野における政策や研究のエビデンスを導き出せる質量を備えることができるのか、その場合はどういった課題やテーマに対して強みを発揮できるのかを評価するものである。

当研究では、特に NDB・介護 DB の連結解析推進という観点から、以下の 2 点を評価すべき課題として、研究事業を進めた。

① 連結可能性の評価：NDB と介護 DB とを連結させることが技術的に可能か否か、どの程度の作業負荷が発生するのか

② 利用可能性の評価：NDB と介護 DB とが連結された場合、そのデータは、具体的にどういった分析が新たに可能となり、どういった政策の立案に寄与

できるのか

この 2 つの評価課題に対し、まず①連結可能性の評価として、介護 DB の概要と構成要素について整理を行った。そのうえで、これまでの医療・介護データの連結解析の議論や、両データを連結して行った先行研究を参考に、NDB データと介護 DB データの連結解析を容易にするデータ構成を検討し、連結解析に使用できるような、新しいテーブルの作成を行った。②の利用可能性の評価としては、作成したデータマートを用いて先行研究を踏まえつつ、それをトレースする形式で非常に簡単なレベルの試行集計を行った。その際に、どの程度の作業時間を要したかについても、確認を行った。

B. 研究方法

今回の研究の根幹をなす NDB・介護 DB 連結データマート作成にあたっては、傷病名・診療行為・医薬品に関する希少性の高い情報を秘匿化した 2014 年度、2015 年度分の全 NDB データ、および NDB との紐付けが物理的にしか実施

できないよう、ID をランダムに振り替えられた同年度の介護 DB の全データを利用し、連結解析を行うためのデータマート構築を実施した。NDB を用いた 2 題の個別研究も、この NDB データを用いて行った。NDB データの提供依頼申出を行い使用するにあたっては、「レセプト情報・特定健診等情報の提供に関するガイドライン」を、介護 DB データの提供依頼申出を行い使用するにあたっては、「要介護認定情報・介護レセプト等情報の提供に関するガイドライン」を遵守した。またデータ提供を受けるに際しては所属機関における倫理審査での承認が必要であるため、京都大学医の倫理委員会にて申請を行い、承認を得た (R1119-2)。

なお、分担研究も含めた研究方法の詳細については、主に令和 2 年度研究報告書の記載を参照いただきたい。

C. 研究結果

D. 考察

E. 結論

本研究 1 年目の令和元年度は、初年度は、すでに我々の研究班で整備している NDB データに連結できる介護 DB データを、申出手続きを経て入手し、1 か月分の介護 DB データを用いた、試行的な DWH 構築及び課題の洗い出しを行った。具体的には以下がそれに相当する。

- ・データ読み込み
- ・DWH 作成
 - ・連結キーの作成

- ・DM 作成
- ・データ出力
- ・残り 23 か月分のデータの NAS へのファイルコピー

ここでいう「連結キー」とは、既に構築した NDB 並列分散処理基盤に格納しておいた ID0 からランダムにピックアップし、介護 DB のそれぞれのデータの ID として使用する ID のことである。この連結キーを、要介護認定情報・介護レセプト等情報の各テーブルの末尾に、サービス提供年月とともに NDB との連結用に格納した。この「ID0」は、NDB の既存の「ID1」と「ID2」を組み合わせ、分担研究者である今村らにより、同一個人や別人物の識別性能を上げ、疫学的な観点で許容可能なレベルの第 3 の ID (ID0)として整備したものである。本研究では、この ID0 を介護 DB データと連結する際の連結キーとして使用した。

2 年目の令和 2 年度には、残り 23 か月分の介護 DB データ格納を行うとともに、請求情報として、診療および介護サービスが生じる際には毎月発生する NDB の匿名レセプト情報、介護 DB の匿名介護レセプト情報に、患者の状態像やアウトカムを規定する情報として有用性の高い要介護認定情報を連結しやすくするよう、新しいテーブルを作成した。こうして整備した連結データマートを用い、先行研究を参考に、簡易な試行集計を実施した。加えて、既存の NDB データの価値を橋上させて今後の利活用推進に資するべく、NDB レセプトデータを

用いて死亡を予測する深層学習モデル (deep learning model) の開発及びバリデーションを行った。また、今後の NDB・介護 DB 連結解析を見据えて、高齢者の介護に直結する課題に対して NDB 単独で個別研究を実施し、今後の連結解析に向けた知見を導き出した。

これら研究の結果、考察、結論については、各年度の総括研究報告書および分担研究報告書を参照いただきたい。各々の研究報告書は、本報告書に資料として掲載している。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

- 1) Kensuke Morris, Osamu Sugiyama, Goshiro Yamamoto, Manabu Shimoto, Genta Kato, Shigeru Ohtsuru, Masayuki Nambu, Tomohiro Kuroda, Towards a Medical Oriented Social Network Service: Analysis of Instant Messaging Communication among Emergency Physicians, *Advanced Biomedical Engineering*, 2020, 9, p35-42, <https://doi.org/10.14326/abe.9.35>
- 2) Tomohide Iwao, Genta Kato*, Shigeru Ohtsuru, Eiji Kondoh, Takeo Nakayama and Tomohiro Kuroda, An Optimum Data Warehouse for Epidemiological Analysis using the National Database of Health

Insurance Claims of Japan, *European Journal for biomedical Informatics*, 2019, 15(3), 31-42.

- 3) Iwao T, Kato G, Ito I, Hirai T and Kuroda T. Treatment of Mycobacterium avium–intracellulare complex lung disease in the real world: a retrospective big data analysis. *Drugs and Therapy Perspectives*. DOI:<https://doi.org/10.1007/s40267-019-00687-9>, p1-8.
- 4) Yuichi Nishioka Sadanori Okada Tatsuya Noda Tomoya Myojin Shinichiro Kubo Shosuke Ohtera Genta Kato Tomohiro Kuroda Hitoshi Ishii Tomoaki Imamura, Absolute risk of acute coronary syndrome after severe hypoglycemia: A population - based 2 - year cohort study using the National Database in Japan, *Journal of Diabetes Investigation*, p1-9. <https://doi.org/10.1111/jdi.13153>.
- 5) Tomohide Iwao Genta Kato Isao Ito Eiji Aramaki Tomohiro Kuroda, A survey of clarithromycin monotherapy and long - term administration of ethambutol for patients with MAC lung disease in Japan: A retrospective cohort study using the database of health insurance claims. *Pharmacoepidemiology and Drug Safety*, p1-6,

<https://doi.org/10.1002/pds.4951>

- 6) Shingo Fukuma, Tatsuyoshi Ikenoue, Sayaka Shimizu, Edward C. Norton, Rajiv Saran, Motoko Yanagita, Genta Kato, Takeo Nakayama, Shunichi Fukuhara and on behalf of BiDAME, Quality of Care in Chronic Kidney Disease and incidence of End-Stage Renal Disease in Older Patients ACohort Study, Medical Care, 2020, 58(7), 626-631.

2. 学会発表

- 1) 加藤源太、保険医療介護ビッグデータ研究の人材育成：京都大学の事例紹介、日本臨床疫学会 第3回年次学術大会、2019年9月28日
- 2) 加藤源太、大寺祥佑、明神大也、西岡祐一、久保慎一郎、野田龍也、患者調査におけるNDBデータの利用可能性に関する評価の－基本的な集計項目について－、第78回日本公衆衛生学会総会、2019年10月23日
- 3) 大寺祥佑、植嶋大晃、森由希子、加藤源太、黒田知宏、オンサイトリサーチセンター運用者の立場から、第39回日本医療情報学連合大会、

2019年11月24日

- 4) 植田彰彦、近藤英治、大寺祥佑、朝野美穂、中北麦、万代昌紀、加藤源太、黒田知宏、初学者による京都大学NDBオンサイトリサーチセンターの使用経験、第39回日本医療情報学連合大会、2019年11月24日
- 5) 朝野美穂、加藤源太、大寺祥佑、森由希子、植嶋大晃、黒田知宏、日本における保健医療ビッグデータの紹介：レセプト情報・特定健診等情報データベース（NDB）について、宮古島合同学術集会2019、2019年11月23日
- 6) Mizuki Watanabe, Shosuke Ohtera, Junya Kanda, Shusuke Hiragi, Tomohide Iwao, Tomohiro Kuroda, Akifumi Takaori-Kondo, Genta Kato, Cost analysis using Japanese National Database (NDB); How much does hematopoietic stem cell transplantation cost in the real world?, 第42回日本造血細胞移植学会総会

H. 知的財産権の出願・取得状況

なし