

令和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))

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
総括研究報告書

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 大寺 祥佑 (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及び利用可能性の検証を行うことを目的とするものである。まず、NDB・介護DBの連結可能性の評価と、連結データの利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、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、新しいテーブルの作成を行った。次に、利用可能性の評価として、作成したデータマートをを用い、先行研究を再現する形式で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2題行った。

【研究方法】

今回の研究では、2014年度、2015年度分のNDBデータ、介護DB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。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】

介護に関連する各種データを用いた先行研究では、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を新たなテーブルとして作成した。こう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、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、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、得られる知見の幅が広がりうることが確認できた。

今後、NDBと介護DBの連結解析を行う基盤が整えられ、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。

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。レセプトデータには、病名情報や診療行為履歴情報が豊富に含まれるものの、疾患の重症度に関するデータが不足しているため、リスク調整が十分にできないことが多い。これまでに考案されてきた様々なリスク調整指標も、少数の変数を用いたロジスティック回帰に基づくものであり、それに加え、診療領域の専門的知見に基づいてモデルに投入する変数を決定するため、専門領域をまたいだ精緻なモデルの構築は難しかった。この課題に対し、深層学習モデルを用いて、レセプトデータに含まれる各個人の病名および診療行為履歴情報を網羅的にモデルに投入し、より高い精度で在院死亡を予測するモデルを開発し、予測精度と適合度からバリデーションを実施した。

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 2 題行った。1 つは、心筋梗塞入院患者の退院後の外来心臓リハビリテーションの全国の実施実態と関連要因についての分析、もう 1 つは心不全入院患者の退院後の在宅医療提供施設種別の違いとその後の再入院・死亡、及び往診利用状況との関連についての分析である。いずれも、社会問題となっている高齢者の心不全治療に関連する研究であり、将来

的に NDB・介護 DB の連結解析が定着すれば、介護 DB の情報を用いてさらに詳細な評価に繋げられる内容を有するものである。

B. 研究方法

今回の研究では、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。

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 (n=46,665,942) を用い、レセプトから得られる情報として年齢・性別、病名、過去の入院歴、初日の診療行為内容を予測変数として用いた。アウトカム変数は在院死亡とした。

なお、研究方法の詳細については、各分担研究報告書の記載を参照されたい。

C. 研究結果

NDB・介護 DB 連結解析基盤開発

介護 DB を構成する情報として、まず、各月の請求のたびに格納される情報として、介護給付実績情報、いわゆる介

護レセプト情報がある。介護給付費明細書の情報が格納されており、介護サービス事業者ごと、利用者ごとに毎月の情報が格納されている。ある意味で、NDB のレセプト情報と扱いはほぼ同様と考えてよい。また、匿名介護レセプト情報の各レコードには個人 ID が含まれているので、ここに NDB と同一の ID が付与されていれば、特段の加工をせずとも連結解析は可能である。

一方、要介護認定情報は、登録のたびに格納される情報、つまり毎月は入っていない情報である点で、介護レセプト情報等は仕様が異なっている。台帳情報も、変化が生じた際に国保連合会に異動情報の提出が行われ、その都度データが登録されることになっている。

日本における介護 DB に相当するデータを用いた研究を我々でいくつか確認したところ、大きく分けて以下 2 つの使用の態様が認められた。

- ① 得られた情報を患者のアウトカムとして使用する
- ② 得られた情報を対象者の状態像を示す情報として使用する

これらの情報は、介護レセプト情報だけでなく、要介護認定情報や台帳情報に含まれる受給者台帳情報に記載されていることが少なくない。レセプトは基本的には行った行為に対する請求情報であり、対象者の状態を示す情報はそれほど多く含まれていない。この傾向は NDB のレセプト情報についても言えることであ

り、NDB のレセプトデータにも要介護認定情報や台帳情報を円滑に連結できれば、患者の病態や ADL といった情報を付加した利用ができるようになり、データの価値が高まると考えられる。

一方で、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。このため、認定調査ごとのデータを毎月の匿名介護レセプト情報や NDB レセプト情報に連結させて分析しようとする、その都度、前後数か月、最長で合計 48 か月分の要介護認定情報を確認しなければならなくなる。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と考えられた。

こうした方針に沿って、今回、要介護認定情報が各月ごとに確認でき、調査内容も参照できるよう、新たなテーブル「NI (仮)」を作成した。その構成は以下のとおりである。

- ID
- 性、年齢階級
- 年月ごとの要介護度
- 要介護度の認定有効期間
(開始年月日、終了年月日)
- 認定通番
(本研究にて独自に導入)

NI テーブルに多くの情報を入れると使いづらくなるため、テーブルに含める情報は上記に限定しつつ、認定調査 1 回に

対して一意の ID となる「認定通番」を作成し、テーブルに加えた。これにより、該当する月の認定調査の詳細情報を参照できるようにした。

こう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 通り実施した。

(事例 1a：要介護度別の疾患の集計)

NDB の RE レコードで性別情報および患者を一意に定めることを行い、次に、SY レコードを使ってカテゴリー化した傷病名を特定する作業を行った。本来であれば SY レコードの傷病名コードを ICD10 などで臓器別に区分化して行うのが望ましかったが、今回は介護 DB における ID がランダムに付与されており、結果を解釈する余地がないため、集計に際してのパフォーマンスを確認することを優先課題とし、SY レコードにある主傷病フラグが 1 である傷病を用い、傷病名コードの下 1 けた (0~9) に基づいて 10 種類の傷病カテゴリーに分類する、原始的な試行的な傷病分類を作成した。また、介護 DB からは新規作成した NI を用いて、この期間のなかで最新の年月における要介護度を取得し、分析に使用した。

(事例 1b：傷病分類別の要介護度集計)

次に、事例 1a の分析とは逆の発想で、傷病分類ごとに要介護度の分布を明らかにすることにも意義があると考え、

連結集計を行った。今回は、NDBのSYレコードに高血圧、虚血性心疾患、脳卒中の傷病コードが含まれるレセプトを抽出し、それらの事例の要介護度を、最新の年月におけるNIテーブルの要介護度情報とかけあわせて試行集計を行った。事例1a、事例1bの集計いずれも、介護DBのIDはランダムな番号とし、集計結果の解釈ができないという前提において介護DBの全データの入手を了承されたという経緯があったことから、特段の傾向などを見出すことは出来なかったが、集計結果を得ることは成功した。

集計1の実施に際しては、一度傷病毎の分類を作成すれば、その後当研究で作成したテーブル(NI)と突合して年月ごと、患者ごとの処理を行う作業、IDで集計し傷病グループごとに集計表を作成する作業に要したのはそれぞれ数分程度であった。作業時間の減少については、介護情報を突合せせる段階で対象人数が大幅に減少することもその要因となっていると考えられた。

(事例2：認定調査情報を用いた集計)

もう1つの事例として、先行研究を参考に、施設入所者における避けられたかもしれない入院(Potentially Avoidable Hospitalizations, PAH)をアウトカムとし、PAHとそれ以外の入院、病院での死亡との関連要因について、集計を行った。アウトカムとして用いられている変数のひとつに認定調査項目に含まれている経管栄養の有無に関する情報を利用していることに着目し、経管栄養の有無

しのフラグを作成して、試行集計を行った。

深層学習モデルによる在院死亡予測精度

予測精度の計算にはDelongの方法によるArea under curve (AUC)を用い、初日の治療内容情報を用いて在院死亡を予測するメインモデルでの全症例のAUCを評価したところ、結果は0.954(95%信頼区間:0.9537-0.9547)であった。また、急性心筋梗塞・心不全・脳卒中・肺炎の4疾患については、既存の重症度指標を用いたモデルも作成して比較を行った。その結果、4疾患のいずれにおいても、メインモデルの方がAUCの値は高値であった。心筋梗塞に関するキャリブレーション曲線では、メインモデルでは予測死亡と実測死亡がほぼ一致しており、モデルの適合度が極めて良好である一方、既存の重症度指標を用いたモデルでは曲線が45度線の下方に偏位しており適合度はあまりよくないことが確認できた。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と、既存の疾患別重症度指標を用いたモデルを比較した場合、前者の方が在院死亡の予測精度が高いことが確認できた。

高齢者心不全に関するNDB個別研究

心臓リハビリテーション(CR)の実施道教を評価した研究では、最終的な研究対象者は39317人であった。全国の入院CR実施率は約60%で、外来CR実施率は約10%であった。外来CRの実施は5%未満の県から約20%の県まで幅が見

られた。入院 CR 実施者に限った全国の外来 CR 実施率は約 15%であった。外来 CR 実施率は総じて低い値となっており、日本においても心筋梗塞患者に対する外来 CR の更なる促進が求められる。

多変量ロジスティック回帰分析の結果、外来 CR 実施との間に、IABP 利用、ECMO 利用、非侵襲的陽圧換気利用、酸素投与、ACEI 処方、 β ブロッカー処方、抗血小板薬処方、抗凝固薬処方、男性の性別と正の関連がみられ、高齢（80-89 歳・90 歳以上、対照—40 歳未満）、チャールソン併存疾患指数と負の関連がみられた。外来 CR は軽症ではない心筋梗塞患者に対して実施されやすいことが考えられる。一方で高齢であることと併存疾患では外来 CR 実施と負の関連がみられた。これらは高齢で併存疾患を有するものには外来 CR が実施されない傾向があることを表している。

高齢者心不全入院患者の、退院後の在宅医療提供施設別の再入院率と死亡率、平均往診回数の違いを評価した研究では、機能強化型在支診/在支病を利用していた患者割合は約 50%、在支診/在支病を利用していた患者割合は約 40%、その他(一般診療所)を利用していた患者割合は約 10%であった。6 か月以内の全再入院者割合は約 60%、心不全再入院者割合は約 20%、死亡者割合は約 10%であり、在宅医療提供施設別のカイ二乗検定の結果では、全再入院はその他(一般診療所)が多かった一方、心不全再入院は差を認めなかった。在宅医療における患者で頻度が多い問題として感染症が挙げ

られているが、感染症などの治療に関しては在宅で継続し再入院を防げるのに対して、心不全の治療は入院が必要になる割合が高いため、差を認めなかった可能性が考えられる。

6 か月以内の再入院における多変量ロジスティック回帰分析の結果では、アウトカムが全再入院の場合、その他(一般診療所)と比べて在支診/在支病、機能強化型在支診/在支病が有意に低くなった一方で、アウトカムを心不全再入院とした場合は有意差を認めず、死亡に関する有意差を認めなかった。多変量線形回帰分析の結果では、その他(一般診療所)と比較して在支診/在支病、機能強化型在支診/在支病いずれも有意に 1 か月あたりの平均往診回数および時間外平均往診回数は多くなった。平均往診回数、時間外の平均往診回数は在支診・在支病、機能強化型在支診/在支病でどちらも有意に高くなる傾向にあった。このことから、再入院率が低下する機序として往診が関連している可能性が考えられるが、この点に関しては今後媒介分析などでさらなる検討が必要である。

D. 考察

本研究では、匿名要介護認定情報を各月の匿名介護レセプト等情報に新たなテーブルとして付与する手法を採用したが、これだと、「利用するデータの対象期間」の、特に開始時期を、匿名要介護認定情報と介護レセプト等情報で合わせることが難しい。何故なら、入手する初

月分のレセプト情報と一致する要介護認定情報は、多くの場合はそれ以前に登録されるため、介護 DB に含まれるデータの入手対象時期を同一にすると、捕捉できない要介護認定情報が発生してしまうためである。この課題を解決するには

- ① 匿名介護レセプト情報の入手初月分以前の匿名要介護認定情報を、参照用に入手する
- ② 匿名介護レセプト情報の月に対応する匿名要介護認定情報がない場合は、匿名介護レセプト等情報から確認できる要介護度情報を活用する

の 2 通りの解決策が考えられる。但し、①の場合には要介護認定情報を余分に入手し、処理を行わなければならない。一方、②の場合は、認定調査に関する項目の詳細情報を参照できない。今後、利活用が進むにつれて、登録型の匿名要介護認定情報と毎月の業務から発生する匿名介護レセプト等情報という、データ収集の仕組みが異なる 2 つのデータを以下に連結させて利用するか、という問題は顕在化していくのではないかと考えられた。

今回の試行集計では費用負担に関する項目は入っておらず、医療費と介護費の集計は実施していないが、今後連結分析が定着すれば、医療費と介護費の合計金額に焦点を当てた研究も進むと思われる。今回の NI テーブルに新たに合計金額の項目を付加するなどすれば、利用者

の利便性が高まることが想定された。今後の課題である。

E. 結論

NDB データと介護 DB データの連結解析を容易にする目的で、要介護認定情報から各月ごとの情報を入手できるテーブルを新しく導入し、介護 DB データにランダムに付与した ID で連結し、試行集計を実施した。これらの研究を通じて、NDB と介護 DB の全数レベルでの連結解析が実施可能であることを明らかにした。これによって今後の NDB・介護 DB 連結解析を実施できる基盤の有効性を確認することができた。

また、NDB を用い、入院初日の治療内容情報を用いて在院死亡を予測する深層学習モデルを構築し、既存の重症度指標に比べて在院死亡の予測精度が高いことを明らかにした。これは、今後の NDB・介護 DB 連結解析を推進する上で、利用者にとって有用なツールとなることが期待される。

また、個別研究の結果からは、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。

本研究は今年度が最終年度となるが、これからも 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。

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・取得状況

なし