

令和 3 年 5 月 31 日現在

機関番号：12601

研究種目：若手研究(A)

研究期間：2017～2020

課題番号：17H05077

研究課題名（和文）大規模医療・介護関連データベースを用いた臨床疫学・経済学研究基盤構築

研究課題名（英文）Development of a platform for clinical epidemiological and economic research using large-scale medical and nursing care related databases

研究代表者

松居 宏樹 (Matsui, Hiroki)

東京大学・大学院医学系研究科（医学部）・助教

研究者番号：70608794

交付決定額（研究期間全体）：（直接経費） 19,000,000円

研究成果の概要（和文）：大規模医療データの二次利用はその重要度を増している。本研究の目的は、NDBを解析して疫学研究を実施する事、そして、NDBと他のデータベースを連携して解析する事である。NDBを用いた研究を実施するうえで、そのデータ構造が複雑であるという課題があった。そこで、本研究では研究者が利用しやすいデータ構造の定義と、NDBから研究データを抽出変換するシステムを独自に開発し、疫学研究を行った。さらに近年注目されている深層学習等の手法を用い、データベース間をID連結せずとも解析を行う手法の開発を行った。さらに、実際にID連結解析したデータベースの構築を行った。

研究成果の学術的意義や社会的意義

本研究により、今まで十分に実施されてこなかったNDBの長期追跡データを用いた疫学研究の実施が可能となった。

さらに、今までは制度上の問題で実施できなかった、データベース間に患者属性が散らばる場合でのデータ解析を、データベース突合を行わずに実施する方法を考案した。加えて、実際にデータベース間を突合したデータの取得にも成功した。今後、今回の研究を基盤として、さらなる疫学研究の実施が可能となった。

研究成果の概要（英文）：The secondary use of large-scale medical data is becoming more and more important. The purpose of this study was to conduct epidemiological research by analyzing the NDB. And to analyze the NDB in conjunction with other databases.

In order to conduct research using the NDB, there was a problem that the data structure was complicated. Therefore, in this study, we defined a data structure that is easy for researchers to use, developed our own system for extracting and converting research data from NDB, and conducted epidemiological research. In addition, we developed a method for analyzing the data without linking the databases by using methods such as deep learning, which has been attracting attention in recent years. In addition, we constructed an actual database with ID-connected analysis.

研究分野：医療情報・臨床疫学

キーワード：大規模医療データベース NDB 介護レセプト 深層学習 データフュージョン 臨床疫学

1. 研究開始当初の背景

近年、大規模な Administrative Claim Database の臨床疫学・経済学分野へ二次利用はその重要度を増している。米国では、CMS の有する Medicare Claim database 用いた研究が盛んに行われ、韓国・台湾では、国民の医療実施レセプトを用いた疫学研究の報告数が増加している。本邦では、全国のレセプトと特定健診情報等データベース (National receipt database:NDB) の研究利用を促す体制が整備され研究利用が開始されている。また、介護領域の介護保険レセプトデータベースを整備したデータベースも整備されつつある。Administrative Claim database を臨床疫学・経済学研究に用いる価値は、その悉皆性と症例追跡力にある。さらに、本邦の Administrative Claim Database は診療行為を実施日単位で追跡出来る等、国外のデータと比して情報の粒度が細かい。対して、本邦の Administrative Claim Database の他国と共通する課題は、あくまで医療・介護実績情報のみが記録されており、患者の重症度に関する情報は取得不可能である点である。さらに、本邦では他国と異なり、国民を一意的に識別する ID 体系が整備されていないため、他のデータベースの情報 (死亡状況や介護の状況) と情報の突合が困難である点も重要な課題である。さらに、保険者毎の被保険者番号と生年月日性別を用いた ID1 と氏名と生年月日性別を用いた ID2 という不完全な ID を用いた形でしか、個人を識別・追跡出来ない点に大きな課題が残る。本邦の Administrative Database の課題を克服するためには、i)新たな長期追跡 ID 体系を利用した解析を行うこと、ii)他のデータベースと連携した解析を行う事が重要である。しかし、現状 NDB には不完全な ID 体系しか作成されておらず、NDB と他のデータベースを組み合わせて解析も行った事例も未だない。

2. 研究の目的

本研究の目的は、1) 新たな長期追跡可能な ID 体系を構築した NDB データを解析して長期追跡の強みを活かした疫学研究を実施する事、そして、2) NDB と介護保険レセプトデータベース・DPC データベース等を連携して解析する事である。

3. 研究の方法

1) 新たな長期追跡可能な ID 体系を構築した NDB データを解析して長期追跡の強みを活かした疫学研究を実施する。

1-1) NDB を利用するうえで必要となる ID 体系の構築と、それを基にした研究データの抽出

NDB を用いた疫学研究を実施するためには、まず、DB 内の長期追跡 ID 体系を構築する必要がある。我々は研究開始時点で、ID1 と ID2 を用い ID の再帰的探索を行うことで、ID0 と呼ばれる長期追跡 ID 体系を構築することに成功していた。しかし、NDB はレセプト情報が全 73 テーブル (医科, 歯科, 調剤, DPC: 18, 15, 14, 26) に格納されており、研究利用を行うには、複雑すぎる構造をしている。そのため、NDB 本体データから、研究テーマに沿った症例群を抽出し、症例群の情報を解析しやすいデータ構造に変換する必要があった。我々は、当該研究期間中に既存の ID0 の構築手法の改善と、構築された ID 体系を基盤とした研究データを NDB 本体から抽出するシステムを構築した。

1-2-1) 術前口腔ケアの主要癌術後合併症に対する効果 (NDB を利用した疫学研究の実践)

患者の口腔衛生状態を改善することは、口腔や咽頭の分泌物の吸引によって引き起こされる可能性のある術後の肺炎を予防するための選択肢の一つである。しかし、歯科医による術前の口腔ケアが術後の合併症を減少させるかどうかについては、まだ議論がある。そこで、大規模ながん手術を受けた患者を対象に、術前の口腔ケアと術後合併症の関連性を評価するため、NDB データを用いたレトロスペクティブコホート研究を実施した [1]。

NDB データから、2012 年 5 月から 2015 年 12 月の間に、頭頸部がん、食道がん、胃がん、大腸がん、肺がん、肝がんの切除術を受けた患者を特定した。主要アウトカムは、術後の肺炎と術後 30 日以内の全死因死亡とした。患者背景は、傾向スコアを用いた治療の逆確率による重み付けで調整した。

1-2-2) 糖尿病患者における sodium-glucose cotransporter-2(SGLT2 阻害剤)と尿路感染症のリスクについて- 日本の Administrative Claims Database を用いたターゲット・トライアルエミュレーション (NDB を利用した疫学研究の実践)

糖尿病における SGLT2 阻害剤の使用に関連する尿路感染症 (UTI) の発生リスクを、ターゲット・トライアルエミュレーションを用いて評価した [2]。

NDB を用いて、2014 年 4 月から 2015 年 3 月の間に SGLT2 阻害薬、dipeptidyl peptidase-4(DPP-4) 阻害薬、ピグアナイドのいずれかを処方された 40 歳以上の患者のコホートを構築した。計算を容易にするために、SGLT2 阻害薬ユーザーの 100%、DPP-4 阻害薬ユーザーの 3%、ピグアナイド薬ユーザーの 20% を無作為に抽出し、抗糖尿病薬の New User を分析した。UTI の intention-to-treat (ITT) ハザード比 (HR) は、その後の治療変更を無視した IPT (inverse probability

of treatment) 加重 Cox の比例ハザードモデルを用いて推定した。治療の重み付けは、患者の性別、年齢、投薬、病歴、入院歴を用いて計算した。また、非ランダムな治療変更を調整した IPT および逆打ち切り確率加重 Cox モデルを用いて、per-protocol (PP) HR を推定した。

2) NDB と介護保険レセプトデータベース・DPC データベース等を連携して解析

NDB と介護保険レセプトデータベースや DPC データベース等の大規模医療データベースとの連携解析を行う上で複数の方法を検討した。

2-1) データ・フュージョンを用いた方法(2017 年度公衆衛生学会にて発表)

複数データベース間を突合する「ID 情報」が存在しない状況で、多重代入法(multiple imputation:MI)を用いたデータ・フュージョン(data fusion)により、個人情報突合を伴わずに未測定交絡因子の影響を補正できるかどうかを DPC データに適用可能かを検証した。

PC データを元にしたモンテカルロ・シミュレーション(n =50)を行った。平成 22 年 7 月-平成 23 年 12 月の DPC データより、COPD の緊急入院症例を抽出した。アウトカムを退院時 ADL、治療を入院 48 時間以内のリハビリテーション開始とし、年齢と入院時 ADL を交絡因子とした。また、DPC データから得られる治療実績及び、重症度情報を欠測補完に用いる因子として用意した。アウトカムのみをマスクしたデータ A、交絡因子のみをマスクしたデータ B を作成した。データ A、B を統合し、アウトカムが欠測した行と交絡因子が欠測した行が混在するデータ C を作成し、欠測補完に用いる変数を元に多重代入法を実行した。作成された 10 個の代入後データそれぞれから、データ A に由来する行を削除したデータを用いて、治療効果をロジスティック回帰モデルにより解析した。アウトカムをマスクする前のデータ A の治療効果(真値)と、代入後データで推定された治療効果(検証値) およびデータ C での結果(偽値)とを比較した。

2-2) 深層学習を用いた方法(2018 年度日本臨床疫学会にて発表,2021 年英語論文投稿中)

入院時点で推定される患者予後はリスクスコアとして用いることで、臨床疫学研究における患者のリスク補正を行うことができる。臨床現場において種々のリスクスコアが利用されているが、近年数多くの臨床疫学研究で利用されているレセプトデータベースにおいては、リスクスコア算出に必要な重症度を把握することが困難であるという強い限界が存在する。近年、深層学習を始めとする機械学習手法の発達により、大量の予測因子を用いたモデル構築が可能となった。本研究では、DPC データを用いて、レセプトデータから取得可能な大量の処方情報を用い作成したリスクスコアの妥当性を検証し、DPC 様式 1 に含まれる既存重症度との対応を検討した。さらに、既存重症度から算出されるリスクスコアも作成し、二つのリスクスコアの精度を比較し妥当性を検証した。

2-3) 医療介護連結済みデータベースの構築

医療と介護と特定健診のデータは地方自治体が国民健康保険を運用するうえで蓄積される大規模データベースである。今回、革新的研究開発推進プログラム ImpACT 社会リスクを低減する超ビッグデータプラットフォーム(原田 PM)の一環として、自治医科大学データサイエンスセンターで自治体から直接、匿名化した医療介護特定健診連結データベースデータ収集が行われた。データ取得と解析を行うためには、取得に必要な匿名化プロセス及び、データベース構造の設計が必要であった。取得プロセスと、データベースの設計に、本研究の 1-1) で開発された NDB 本体から研究データを抽出するシステムを継承した。

医療・介護・特定健診とそれらを連結する自治体国保加入者台帳と複数の ID 体系のデータベースを連結するため、1-1) で構築した ID0 を拡張子、新たな ID 体系を構築した。

4. 研究成果

1-1) NDB を利用するうえで必要となる ID 体系の構築と、それを基にした研究データの抽出

研究データの構造として、73 テーブルを 15 テーブルに集約した。

研究データの自動抽出過程は、適切なインデックスを参照し処理を逐次的に進めることで、高速かつ大規模な研究データの取得が可能となるようにシステムを構築した。

研究者は、研究に必要な症例を定義するクエリを作成し、研究に用いる情報(必要となる病名情報と診療プロセス情報)を用意したうえで、必要項目を GUI に入力することで、NDB 本体データから必要となる研究データを取得するスクリプトを自動生成することが可能となった。

1-2-1) 術前口腔ケアの主要癌術後合併症に対する効果(NDB を利用した疫学研究の実践)

調査対象となった 509179 名の患者のうち、81632 名(16.0%)が歯科医による術前の口腔ケアを受けていた。合計 15724 人(3.09%)の患者が術後肺炎にかかり、1734 人(0.34%)が術後 30 日以内に死亡した。潜在的な交絡因子を調整した結果、歯科医による術前の口腔ケアは、術後の肺炎(3.28 対 3.76%、リスク差-0.48 (95%CI: -0.64 to -0.32)%) および術後 30 日以内の全死亡率(0.30 対 0.42%、リスク差-0.12 (-0.17 ~ -0.07)%)の減少と有意に関連していた。

1-2-2) 糖尿病患者における sodium-glucose cotransporter-2(SGLT2 阻害剤)と尿路感染症のリスクについて- 日本の Administrative Claims Database を用いたターゲット・トライアル

ミュレーション (NDB を利用した疫学研究の実践)

分析対象は、SGLT2 阻害剤の投与開始者 11,364 人、DPP-4 阻害剤の投与開始者 9,035 人、ビッグアナイドの投与開始者 10,359 人であった。SGLT2 阻害剤投与群とビッグアナイド投与群を比較すると、SGLT2 阻害剤投与群の粗 HR は 1.14 (95%CI:1.05-1.24)、ITT HR は 0.94 (95%CI:0.86-1.03)、PP HR は 0.90 (95%CI:0.78-1.03) であった。DPP-4 阻害剤の開始者は、粗 HR が 1.13 (95%CI:1.04-1.23)、ITT HR が 0.85 (95%CI:0.77-0.94)、PP HR が 0.83 (95%CI:0.71-0.95) であった。SGLT2 阻害剤や DPP-4 阻害剤の使用は、ビッグアナイド系薬剤の使用と比較して UTI のリスクを増加させなかった。治療法の変更を考慮しても、推定効果に大きな影響はなかった。

2-1) データ・フュージョンを用いた方法(2017 年度公衆衛生学会にて発表)

DPC データより 14393 症例を抽出した。真値のオッズ比は 0.97 (SD: 0.05) であった。検証値・偽値のオッズ比はそれぞれ 0.85 (0.05)・0.80 (0.04) で、検証値が真値に近かった。多重代入法を用いた data fusion により、実データでも未測定交絡因子の解析結果への影響を低減できる可能性が示唆された。

しかしながら、利用できる情報量が少ないと、検証値と真値の間の差が縮まりきらない可能性が示された。

2-2) 深層学習を用いた方法(2018 年度日本臨床疫学会にて発表,2021 年英語論文投稿中)

対象となった患者のうち、2,005,035 人 (4.3%) が死亡していた。モデルの判別能とキャリブレーションは満足のいくものであった (Fig.1)。Validation data におけるモデルの AUC は 0.954 (95%CI:0.9537-0.9547) であった。モデルは、疾患別モデルよりも高い識別能力を有していた。

これにより、精度の高い入院時死亡リスクスコアの算出が可能となった。このリスクスコアは既存のリスクスコアと同等以上の疾患に対する影響を有していることから、このリスクスコアを用いてリスク調整を行ったうえでレセプト情報のみでも精度の高い臨床疫学研究が実施可能となった。

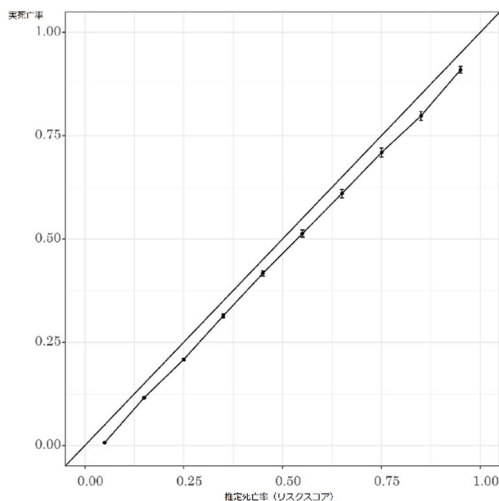


図 1: キャリブレーションカーブ

2-3) 医療介護連結済みデータベースの構築

新たな ID 体系により、医療にかかった後の長期的な介護予後等の新たな臨床疫学研究の実施が可能となった。

1. Ishimaru M, Matsui H, Ono S, Hagiwara Y, Morita K, Yasunaga H. Preoperative oral care and effect on postoperative complications after major cancer surgery. *British Journal of Surgery*. 2018.
2. Takeuchi Y, Kumamaru H, Hagiwara Y, et al. Sodium-glucose cotransporter-2 inhibitors and the risk of urinary tract infection among diabetic patients in Japan: Target trial emulation using a nationwide administrative claims database. *Diabetes, Obes Metab*. 2021;23(6):1379-1388. doi:10.1111/dom.14353

5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 9件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 Okui Jun, Ueno Ryo, Matsui Hiroki, Uegami Wataru, Hayashi Hiroshi, Miyajima Toru, Kusanagi Hiroshi	4. 巻 -
2. 論文標題 Early prediction model of organ/space surgical site infection after elective gastrointestinal or hepatopancreatobiliary cancer surgery	5. 発行年 2020年
3. 雑誌名 Journal of Infection and Chemotherapy	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.jiac.2020.04.009	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Goto Tadahiro, Jo Taisuke, Matsui Hiroki, Fushimi Kiyohide, Hayashi Hiroyuki, Yasunaga Hideo	4. 巻 16
2. 論文標題 Machine Learning-Based Prediction Models for 30-Day Readmission after Hospitalization for Chronic Obstructive Pulmonary Disease	5. 発行年 2019年
3. 雑誌名 COPD: Journal of Chronic Obstructive Pulmonary Disease	6. 最初と最後の頁 338 ~ 343
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/15412555.2019.1688278	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Uda Kazuaki, Matsui Hiroki, Fushimi Kiyohide, Yasunaga Hideo	4. 巻 100
2. 論文標題 Intensive In-Hospital Rehabilitation After Hip Fracture Surgery and Activities of Daily Living in Patients With Dementia: Retrospective Analysis of a Nationwide Inpatient Database	5. 発行年 2019年
3. 雑誌名 Archives of Physical Medicine and Rehabilitation	6. 最初と最後の頁 2301 ~ 2307
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.apmr.2019.06.019	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Takano Ayumi, Ono Sachiko, Yamana Hayato, Matsui Hiroki, Matsumoto Toshihiko, Yasunaga Hideo, Kawakami Norito	4. 巻 9
2. 論文標題 Factors associated with long-term prescription of benzodiazepine: a retrospective cohort study using a health insurance database in Japan	5. 発行年 2019年
3. 雑誌名 BMJ Open	6. 最初と最後の頁 e029641 ~ e029641
掲載論文のDOI（デジタルオブジェクト識別子） 10.1136/bmjopen-2019-029641	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Hashimoto Hideki, Matsui Hiroki, Sasabuchi Yusuke, Yasunaga Hideo, Kotani Kazuhiko, Nagai Ryoza, Hatakeyama Shuji	4. 巻 9
2. 論文標題 Antibiotic prescription among outpatients in a prefecture of Japan, 2012?2013: a retrospective claims database study	5. 発行年 2019年
3. 雑誌名 BMJ Open	6. 最初と最後の頁 e026251 ~ e026251
掲載論文のDOI (デジタルオブジェクト識別子) 10.1136/bmjopen-2018-026251	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Hiroki Matsui , Soichi Koike , Kiyohide Fushimi , Tomoki Wada , Hideo Yasunaga	4. 巻 1
2. 論文標題 Effect of neurologic specialist staffing on 30-day in-hospital mortality after cerebral infarction	5. 発行年 2019年
3. 雑誌名 Annals of Clinical Epidemiology	6. 最初と最後の頁 86 ~ 94
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Matsui Hiroki, Jo Taisuke, Fushimi Kiyohide, Yasunaga Hideo	4. 巻 18
2. 論文標題 Outcomes after early and delayed rehabilitation for exacerbation of chronic obstructive pulmonary disease: a nationwide retrospective cohort study in Japan	5. 発行年 2017年
3. 雑誌名 Respiratory Research	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12931-017-0552-7	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Ishimaru M, Matsui H, Ono S, Hagiwara Y, Morita K, Yasunaga H	4. 巻 105
2. 論文標題 Preoperative oral care and effect on postoperative complications after major cancer surgery	5. 発行年 2018年
3. 雑誌名 British Journal of Surgery	6. 最初と最後の頁 1688 ~ 1696
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/bjs.10915	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Takeuchi Yoshinori, Kumamaru Hiraku, Hagiwara Yasuhiro, Matsui Hiroki, Yasunaga Hideo, Miyata Hiroaki, Matsuyama Yutaka	4. 巻 23
2. 論文標題 Sodium glucose cotransporter 2 inhibitors and the risk of urinary tract infection among diabetic patients in Japan: Target trial emulation using a nationwide administrative claims database	5. 発行年 2021年
3. 雑誌名 Diabetes, Obesity and Metabolism	6. 最初と最後の頁 1379 ~ 1388
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/dom.14353	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計6件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 松居 宏樹
2. 発表標題 深層学習をレセプト記載情報に適用した入院予後予測モデルの構築と妥当性検証
3. 学会等名 日本臨床疫学会第3回年次学術大会
4. 発表年 2019年

1. 発表者名 松居宏樹, 伏見清秀, 康永秀生
2. 発表標題 肺炎症例における深層学習をレセプト記載情報に適用した予後予測モデルの構築と妥当性検証
3. 学会等名 日本臨床疫学会第2回年次学術大会
4. 発表年 2018年

1. 発表者名 松居宏樹, 伏見清秀, 康永秀生
2. 発表標題 肺炎症例におけるレセプト記載を基にした深層学習リスクモデルの構築と妥当性検証
3. 学会等名 第29回日本疫学会学術総会
4. 発表年 2019年

1. 発表者名 松居宏樹、佐藤大介、大江和彦
2. 発表標題 レセプト情報等オンサイトリサーチセンターの現況および今後について- これまでの進捗の報告、および今後の第三者利用について -
3. 学会等名 第37回医療情報学連合大会
4. 発表年 2017年

1. 発表者名 松居宏樹、康永秀生
2. 発表標題 レセプトデータベースを用いた心不全症例における入院中リハビリテーションの効果の検証
3. 学会等名 日本臨床疫学会 第1回年次学術大会
4. 発表年 2017年

1. 発表者名 松居宏樹、康永秀生、伏見清秀
2. 発表標題 臨床データにおける多重代入法を用いたデータ・フュージョンの使用例
3. 学会等名 第76回日本公衆衛生学会総会
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------