

科学研究費助成事業 研究成果報告書

令和元年6月17日現在

機関番号：14603

研究種目：若手研究(B)

研究期間：2017～2018

課題番号：17K17816

研究課題名(和文)疫学分析に適したデータモデル構築に関する研究

研究課題名(英文)Development of Data model for Epidemiology research

研究代表者

岩尾 友秀(Iwao, Tomohide)

奈良先端科学技術大学院大学・研究推進機構・研究員

研究者番号：60772100

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：NDBやJMDCデータベースは、我が国のレセプトデータベースとして良く知られているが、被保険者に関する情報が複数のファイルに分散して格納されている。一方で、データベースを疫学研究へ二次利用する際には、患者ごとの変数が1レコードに集約された構造が望ましいとされている。しかしながら、疫学研究で必要になる変数は、研究デザインにより異なり各自で作成する必要があるため、データハンドリングに不慣れな利用者は多大な労力を費やしている状況であった。本研究では、利用者のリサーチクエスチョン(要望)に合わせてデータウェアハウスを構築することで、患者ごとに集約された構造を定型的なSQLで実現する手法を考案した。

研究成果の学術的意義や社会的意義

本研究で考案した定型的なSQLは、ソフトウェアで生成することが容易であることに加え、利用者のリサーチクエスチョンを分割する作業は自然言語処理技術や機械学習と相性が良い。今後、これらの技術を本手法と組み合わせることで、将来的には半自動化が可能になると思われる。また、テーブル構造に依存しないため、レセプトデータベース以外の正規化されたデータベースにも適用できる可能性が高い。近年では複数の保健医療データベースを連結する動きも活発になってきており、製薬業界などでは市販後調査など二次利用する可能性が高まっているため、研究成果の社会実装が期待できる。

研究成果の概要(英文)：The NDB or JMDC databases are well-known as administrative insurance claims databases in Japan, but information regarding persons is distributed and stored in multiple files because these are stored using a RDB (Relational Database). On the other hand, when the database is used for epidemiological studies, a structure in which variables for each patient are integrated into one record is desirable. However, the variables needed for epidemiological research differ depending on a respective research design, and users who are not familiar with data handling have spent a lot of time and effort. In my study, I devised the pattern SQL, by using it, researchers can get needed variables in a patterned work.

研究分野：医療情報学

キーワード：データウェアハウス NDB JMDC レセプト分析 リアルワールドデータ データベース 薬剤疫学

1 . 研究開始当初の背景

近年では世界的にビッグデータの有効利用がにわかに注目されており、産業界をはじめとして多くの領域で活発な研究がなされている。ヘルスケア分野においてもビッグデータの有効利用が多く提案されつつあり、データベースを用いた疫学研究が盛んに行われるようになってきている。これに伴い、疫学研究に適したデータベースもいくつか構築されている。I2B2 や TranSMART はゲノム領域まで含んだ統合研究環境を提供している例として世界的に知られている。また、臨床研究をはじめとした治験の領域では、規制当局への申請用に CDISC と呼ばれるデータ収集、統計解析に適したデータベース構造が世界的に広まりつつある。

一方で、近年注目されている有用な医療系ビッグデータとして診療報酬系のデータベースが挙げられる。特に国民皆保険制度を導入している国では、ほとんどの国民に関するデータを時系列で蓄積しているためデータの質に関しては極めて悉皆性が高く、全国屈指のビッグデータである。しかしながら、これらのデータベースを用いて有用な疫学研究を実施するには、いくつかの課題が残されていた。

2 . 研究の目的

我が国ではレセプト情報等を格納した NDB(National Database of Health Insurance Claims and Specific Health Checkups of Japan)と呼ばれる診療報酬系のデータベースが 2008 年度から運用を開始している。

平成 18 年度の医療制度改革を受けて、医療費適正化を主目的として診療報酬明細書(レセプト)の電子請求が進められてきた。さらに平成 23 年度から、原則的にすべてのレセプトについてオンライン提出が義務化され、現在では大半のレセプトが電子請求に対応している。これらのレセプトと、特定健康診査・特定健康指導(以下、特定健診等)に関する情報を合わせた情報が NDB に格納されている。

[NDB の概要]

- (1)現在まで 100 億件以上のレセプト等情報が格納されている
- (2) NDB サンプリングデータセット、NDB 特別抽出という形式で一般公開されている

上記(2)のうち、の NDB サンプリングデータセットは、レセプト情報からあらかじめ一定の割合で抽出したデータに対して、さらに安全性に配慮した工夫を施した 1 ヶ月分のデータである。また、の NDB 特別抽出は、提供申出者が指定した抽出条件によって、NDB から抽出したデータでありレセプト、及び特定健診等情報から成るデータである。

現在、国家レベルで診療報酬系データベースを構築している代表的な国としては、デンマーク、スウェーデン、フランス、韓国、台湾、日本等が挙げられる。韓国と台湾では、研究者が SQL と呼ばれるデータベース操作言語を用いて直接データベースを操作する解析環境を構築している。

しかしながら、解析用のデータセット作成には、専門家であっても一定レベル以上のデータハンドリングスキルが要求されるという世界的な課題がある。

[データベースを用いた疫学研究の課題]

- (1)予め明確な仮説をたてて実施する臨床研究等と比較すると、仮説構築に時間と労力を要する
- (2)解析用のデータセット作成には、一定レベル以上のデータハンドリングスキルが要求される

上記(1)は、データベースを用いた疫学研究の特徴に起因する課題である。予め明確な仮説を構築したうえでデータを収集するような臨床研究や、研究機関や協力施設等から所望のデータのみ入手して実施する疫学研究とはデータの質や量の側面から大きく異なる。データベースを用いた疫学研究では、データ量が極めて多く、かつデータ構造も複雑である傾向が強くなる。このため、仮説を構築するための探索的な解析が注目されている。

次に上記(2)は、解析用のデータセットを作成するためのデータハンドリングの問題である。データ自体がデータベース化されているため、解析用データセット作成に関してもデータベース上で作成したほうが効率的である。

これらのデータ自体がリレーショナルデータベースで運用されており、かつデータ量が極めて多いという特徴がある。このため、解析用データセット作成に関してもデータベース上で作成したほうが効率的である。そこで、本研究ではリレーショナルモデル理論に立脚した技術を用いてこの課題を解決する。

3 . 研究の方法

本研究では NDB のサンプリングデータセットを入手する . それらのデータセットを用いて , 下記に述べる実施計画を立てた .

(1) NDB サンプリングデータセットに関して , 疫学分析に適したデータモデルを提案する
(2) 上記(1)で提案したデータモデルを実装し , 自動的にデータベース化する汎用的に利用できるシステムを構築する

上記(1) , (2)を実施することにより , NDB サンプリングデータセット用の分析環境構築ツールを開発する .

4 . 研究成果

本研究では , 利用者のリサーチクエスト (要望) を分割し , リサーチクエストに依存する部分の SQL コードを最小限に抑える工夫により , 患者ごとに集約された構造 (テーブル) を定型的な SQL で実現する手法を考案した . リサーチクエストが複雑な研究課題には対応できないケースも存在するが , データウェアハウスに変数を補うことで , 現実的には大半の研究がカバーできると考えられる .

定型的な SQL は , ソフトウェアで生成することが容易であることに加え , 利用者のリサーチクエストを分割する作業は自然言語処理技術や機械学習と相性が良い . 今後 , これらの技術を本手法と組み合わせることで , 将来的には半自動化が可能になると思われる . 本研究で構築した手法を利用し , 昨年度と今年度に本手法を使用して疫学研究を実施し , 得られた知見を論文誌として公表した .

本研究で考案した手法は , データベース構造に依存しないため , レセプトデータベース以外の正規化されたデータベース (例えば , 電子カルテなど) にも適用できる可能性が高い . 近年では複数の保健医療データベースを連結する動きも活発になってきており , 製薬業界などでは市販後調査など二次利用する可能性が高まっているため , 研究成果の社会実装が期待できる .

5 . 主な発表論文等

[雑誌論文] (計 4 件)

1. Tomohide Iwao: A Survey of Clarithromycin Monotherapy for Patients with MAC Lung Disease in Japan, *Advances in Pharmacoepidemiology and Drug Safety*, 2019, 8(1). No:224. pp.1-4. DOI: 10.24105/2167-1052.8.224. (査読あり)
2. Tomohide Iwao: The research attitude required for engineers who are involved in the development of data warehouses for epidemiological analysis: Let us provide value, not just a labor force. *Biomedical Journal of Scientific and Technical Research*. 11(4). pp.8619-8620, DOI: 10.26717/BJSTR.2018.11.002123. 2018 年 11 月. (査読あり)
3. 岩尾友秀 , 矢野憲 , 黒田知宏 . レセプトデータに基づく非結核性抗酸菌症患者への薬剤処方事例に関する調査 , *薬剤疫学*. 23(2). pp.89-94. 2018 年 7 月. (査読あり)
4. Sato Mai, Kondoh Eiji, Iwao Tomohide, Hiragi, Shusuke, Okamoto Kazuya, Tamura Hiroshi, Mogami, Haruta, Chigusa Yoshitsugu, Kuroda Tomohiro, Mandai, Masaki, Konishi Ikuo; Kato Genta : Nationwide survey of severe postpartum haemorrhage in Japan: an exploratory study using the national database of health insurance claims. *The Journal of Maternal-Fetal & Neonatal Medicine*, pp.1-6. DOI: 10.1080/14767058.2018.1465921. 2018. (査読あり)

〔学会発表〕(計 3件)

1. 中谷友香, 趙晃済, 大鶴繁, 岩尾友秀, 加藤源太: NDB データを用いた心肺蘇生後患者に対する診療の実態調査, 第45回日本救急医学会総会・学術集会, 大阪, Oct.24.2017(査読あり)
2. Tomohide Iwao, Shosuke Ohtera, Michi Sakai, Shusuke Hiragi, Shigeru Ohtsuru, Eiji Kondo, Hiroshi Tamura, Genta Kato, Tomohiro Kuroda: Research on the reconstruction method of health insurance claims database suitable for secondary use for epidemiological analysis, Proceedings of Symposium on Medical and Biological Engineering 2017, 23-23, Sep.15.2017(査読あり)
3. 岩尾友秀, 平木秀輔, 大寺祥佑, 酒井未知, 加藤源太, 田村寛, 黒田知宏: レセプト情報・特定健診等情報データベース(NDB)を対象とした疫学研究に適した分析用データベースの構築, ITヘルスケア学会第11回学術大会, 名古屋, May.28.2017(査読あり)

〔その他〕

ホームページ:

<https://datanalysis.work>

科研費による研究は, 研究者の自覚と責任において実施するものです。そのため, 研究の実施や研究成果の公表等については, 国の要請等に基づくものではなく, その研究成果に関する見解や責任は, 研究者個人に帰属されます。