

令和 6 年 4 月 23 日現在

機関番号：11101

研究種目：若手研究

研究期間：2020～2023

課題番号：20K18954

研究課題名（和文）電子レセプト情報から傷病名を明らかにする確率計算式の開発

研究課題名（英文）Development of a probability calculation formula to identify the disease name from the electronic medical record information

研究代表者

田中 里奈（Tanaka, Rina）

弘前大学・医学研究科・助教

研究者番号：40709713

交付決定額（研究期間全体）：（直接経費） 2,500,000円

研究成果の概要（和文）：厚生労働省は平成23年5月頃から「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」に基づいてナショナルデータベースから抽出された電子レセプト情報を提供している。得られる電子レセプト情報には傷病名は含まれているものの、その傷病名はカルテを正確に反映していないことが多い。そのため電子レセプト情報の提供を受けても正確な傷病名別の解析を行うことは現状困難である。本研究では、電子レセプト情報データを各種疾病登録データと照合し、疾病別の電子レセプト情報上の特徴を明らかにすることで提供を受けた電子レセプト情報の特徴から傷病名を明らかにする確率計算式を開発することを目的とした。

研究成果の学術的意義や社会的意義

電子レセプト情報には傷病名は含まれているものの、その傷病名はカルテ上の病名を正確に反映していないことが多く、例えばがん患者の罹患数をがん登録データと比較した報告では、レセプト情報上のがんは実際の患者数の最大2.9倍多かった。本研究により構築した確率計算式は、疾病登録以外のデータへの応用が可能であることから、さらに予測精度を向上させることで将来的にレセプト情報による解析が容易となり、ナショナルデータベース活用による研究の可能性が広がることが期待される。

研究成果の概要（英文）：In Japan, since 2011, the Ministry of Health, Labour and Welfare has been providing electronic medical record information extracted from the National Database (NDB) based on the "Guidelines for the Provision of Receipt Information and Specific Health Checkup Information." While the electronic medical record information obtained contains disease name, these names often do not accurately reflect the patient's medical history. Therefore, despite receiving electronic medical record information from NDB, conducting accurate analysis based on specific disease names is currently more challenging. This study aims to develop a probability calculation formula to identify disease names from the provided electronic medical record information by comparing it with various disease registration data and elucidating the characteristics of electronic medical record information.

研究分野：がん疫学

キーワード：がん登録

1. 研究開始当初の背景

平成 20 年 4 月より施行されている「高齢者の医療の確保に関する法律」に基づき、平成 21 年度からすべての電子レセプト（医科、DPC、歯科、調剤）は匿名化され、厚生労働省保険局が所有するナショナルデータベースに集積されている。そして、平成 23 年 3 月に公表された「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」に基づき、同年 5 月頃からレセプト情報等の都道府県や研究者への提供が開始された。ナショナルデータベースには毎年自動的に 16 億件余りの電子レセプトが一定の様式で集積されているため、提供される電子レセプト情報は日本国内の医療状況のほぼすべてを反映した全数調査に近いデータであるといえる。

しかし、電子レセプト情報にはいくつかの問題点が存在するため、単純に分析をすることはできない。ナショナルデータベースでは、個人を特定しやすい情報を最小限化するため、テキスト記述部分はすべて削除されている。電子レセプトデータは傷病名が未コード化病名である場合もあり、医療機関では未コード化病名をテキストによって傷病名として管理している。このような未コード化病名は全体の 8% 程度存在し、これらは傷病名別の分析の対象外となる。さらに、レセプトデータ上の傷病名は、カルテにおける最終的な傷病名と異なっている場合がある。例えば、「〇〇がん」を疑われたため検査等の診療行為を行ったが、後から振り返れば「がんではない」と判明するケースが存在する。その場合、実施した診療行為を保険請求するために傷病名は「〇〇がんの疑い」となることがある。これらのデータについて、電子レセプト情報から、実際にがんであったのかを判断することは不可能であり、また、「がん」病名が付いている症例の中に本当のがん症例がどの程度存在するのかを把握することはできない。柴田らは、電子レセプト情報とがん登録データからそれらの患者数を比較し、電子レセプト情報は総患者数を最大 2.9 倍上回っていたと報告した(柴田亜希子他, 厚生指針: 61; 6-12. (2014))。以上の点から、現状では電子レセプト情報の提供を受けても、傷病名ごとの分析を正確に行うことは困難である。

レセプトの電子化率が 98.6% (平成 27 年 5 月時点) となった今日では、電子レセプト情報から日本の医療状況のほぼ全てを把握することが可能であり、今後、全国民を対象とするコホート研究などの情報源となり得る。そのため、電子レセプト情報で傷病名が正確に把握できるようになれば、さらに用途が広がることが期待される。

2. 研究の目的

本研究は、電子レセプト情報を疾病登録データと照合し、その疾病別の電子レセプト情報上の特徴を明らかにして、電子レセプト情報の各項目の組み合わせごとの電子レセプト上の傷病名の確率、すなわちベイズ統計を用いて傷病名を明らかとする確率計算式を開発する。本研究では、日本人の死因の過半数を占める悪性新生物(がん) 心疾患、脳血管疾患などの生活習慣病に関する傷病名を対象とする。本研究成果によってレセプト確率計算式の開発法を構築し、ナショナルデータベースの活用による研究の可能性を広げることを最終的な目的とする。

3. 研究の方法

(1) 確率計算式の開発

弘前大学医学部附属病院の2013～2019年におけるレセプトデータから、病名（ICD-10）で生活習慣病（がん、急性心筋梗塞など）に該当するコードがある症例を病名以外の情報とともに AppLink®から抽出した。また、同一期間に診断された生活習慣病（がん、急性心筋梗塞など）を院内がん登録データや放射線部の心臓カテーテル検査実施記録データなどから抽出し、これらを照合した。AppLink®とは、弘前大学医学部附属病院が独自開発した電子レセプトファイルから診療情報を抽出して保存するシステムである。これにより、レセプト病名に生活習慣病（がん、急性心筋梗塞など）がある症例の中で生活習慣病（がん、急性心筋梗塞など）症例と非生活習慣病（がん、急性心筋梗塞など）症例を分類し、AppLink®から抽出した診療情報ごとに事後確率を算出した。

(2) 確率計算式の精度向上

ベイズ更新 (Bayesian updating)

サンプルを追加することで事後確率を修正し、全体の精度を向上させるのがベイズ統計の考え方である (Bayesian updating)。本研究では、2013～2019年における確率計算式を開発した後、2021～2022年の院内がん登録データや放射線部の心臓カテーテル検査実施記録データが完成し次第モデルに投入して予測（判別）の精度を向上させた。

問題点の修正

自動的に生活習慣病（がん、急性心筋梗塞など）症例であるか否かを判別する機能 (Native Bayesian filtering) を設け、カットオフ値の検討はこの段階で行い、確率計算式を完成させる。

4. 研究成果

本検討では、生活習慣病のうち「がん」を対象とした。弘前大学医学部附属病院の2013～2019年におけるレセプトデータから、ICD-10コードのC00-C97（悪性新生物）、D00-D09（上皮内がん）、D32-D33およびD42-D43（脳腫瘍）、D12-D13およびD37（GIST）、D39（卵巣境界悪性腫瘍）について、病名以外の情報とともに AppLink®から抽出した。レセプトデータと照合するがん情報は、弘前大学医学部附属病院院内がん登録より、2013～2019年にがんと診断された症例の提供を受けた。これらを照合し、レセプトデータでの治療内容について、それぞれの事後確率を算出した。使用した治療内容は、レセプト病名、抗がん剤・内服薬（商品名）、抗がん剤・点滴薬（商品名）、処置（抗がん剤胸腔内注入など）、手術の術式名（〇〇切除術など）、腫瘍マーカーとした。

確率計算式の構築の際、いくつかの問題点が明らかとなった。

(1) データ特性による把握漏れ

レセプトデータ入院・外来の違い

レセプトデータは、入院・外来に分かれているが、入院患者のデータと外来患者のデータでは、予測精度に大きな差が生じた。これは、入院患者の多くは手術などのがん標準治療を受けている可能性が高く、外来患者ではがん標準治療を満たさない患者（前医術後の患者フォロー、再発の有無についての経過観察、検査のみの実施など）が多いためと考えられた。そのため、予測式は入院

患者データと外来患者データに分けて構築する必要があると考えられた。

院内がん登録情報の特性

院内がん登録では、入院・外来を問わず自施設で「がん」と初めて診断された症例、もしくは他施設で既に診断されたのちに初診した症例を登録している。そのため、自施設で再発した症例や、転移した症例については、診断日が過去の年月日となっているため、照合の期間（2013～2019年）より以前に登録された症例はレセプトデータと突合されなかった。予測精度をさらに向上させるためには、使用する院内がん登録情報の期間を使用するレセプトデータの期間よりも長くする必要があったと考えられた。

（２）がんの部位による違い

がんの部位による予測精度の差が明らかとなった。弘前大学医学部附属病院を受診する患者数の部位による差が大きいためと考えられた。

今後、明らかとなった問題点を改善し、新たな確率計算式を構築する予定である。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔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
---------	---------