

令和 2 年 6 月 9 日現在

機関番号：37116

研究種目：基盤研究(C)（一般）

研究期間：2017～2019

課題番号：17K00423

研究課題名（和文）健診データによる健康状態遷移予測モデルの開発

研究課題名（英文）Research in modeling of the health state transition with physical checkup data

研究代表者

浅川 剛（Asakawa, Takeshi）

産業医科大学・情報管理センター・准教授

研究者番号：90419515

交付決定額（研究期間全体）：（直接経費） 1,300,000円

研究成果の概要（和文）：本研究では、国内で広く実施されている健康診断のデータと病院を受診した際に発生する明細記録などの医療ビッグデータを用いる事により、ある疾患を持つ人がその健康状態維持に必要な平均的な医療費が幾らになるのかを推定する手法の開発を行いました。既に持病を持つ人が、新たに別の疾病に罹患した際におよそどれほど医療費が上がるかの試算も可能である汎用性の高い手法となっています。開発した手法の新規性と実施可能性が認められ国内で特許として認められました。

研究成果の学術的意義や社会的意義

従来、病気の状態や合併症などの非常に複雑な個々の患者背景の相違から、既に蓄積された膨大な医療データにも関わらず、ある疾病の治療に平均幾らの医療費が費やされているのかという問いに十分に答えられていなかった。これが健康状態の遷移を考える上で最も早急に解決されるべき課題だと感じ、本研究では統計科学・機械学習を援用して傷病別の医療費推定方法の開発を行った。この手法を用いると、ある基本疾患を持った人が、合併症併発時の医療費の上昇を試算することも可能となり、健康状態の悪化を医療費で数値化して現実的に把握出来るようになる。これは今後の社会のデータヘルス事業の礎となるべき手法である。

研究成果の概要（英文）：In this study, using the medical big data such as medical examinations which are widely conducted and receipt records generated when visiting hospitals, we developed a method to estimate the averaged medical cost with which a person with a certain disease can maintain their health condition. It is a highly versatile method that can be used to estimate how much medical costs will increase if the person who already has the disease has another disease. Our method was already registered as a patent in Japan because of its novelty and feasibility.

研究分野：統計科学、情報学

キーワード：医療ビッグデータ 健診データ レセプトデータ 医療費

1. 研究開始当初の背景

研究を開始した平成29年は、政策方針として「データヘルス計画」が策定された平成25年から数年経った時点であり、健診や診療明細(レセプト)データを積極的に活用した研究活動が散見されるようになっていた。健診やレセプトについては社会的インフラとして概ね標準化された形で電子データ化されており、当時でもその活用の必要性が示唆されてきたにもかかわらず、扱われている情報の限界やデータ構造の煩雑さなどから、実務的にも利用は限定的であり、散見されるようになったとはいえ研究的にはまだ緒に就いたばかりであった。

健診データは主に年一回計測された経時データであって、電子化されたものとしては最長で約10年分のデータとして蓄積されていた。紙ベースや統一されたデータ形式でなければ更に長い年限が蓄積されているようであったが、利用可能なデータとして成型する必要もあり、現実的には利用が困難と言わざるを得ないのが実情であった。また、入手した最長10年のデータであったとしても、リアルワールドでのデータらしく個人が毎年健診にかかっているわけでもないため、欠測だらけであり非常に使いづらい形式として敬遠されてもいた。

2. 研究の目的

本研究における当初の目的は、特定の疾病に着目した上で健診データとそれと連結可能な他の医療関連データから健康状態の変遷過程を見ていく事を想定していた。健診データは主として年一回計測されており、経時的には十分な年数を蓄えられてきた定量的なデータであったため、それを入手できるのであれば経時的に推移していく「健康状態」の追跡が可能だと考えられていた。経年データをより効果的に利用していくために、入手可能であればレセプトも積極的に連結させる事で、多様な健康状態の遷移を表現可能となり、それによるデータヘルス事業の推進から社会実装までを主目的としていた。

3. 研究の方法

(1) データの入手と前処理

研究開始当初は健診データのみで実施可能な健康状態の遷移モデルから着手する事にして、健診データ10年分を幾つかの健診機関から入手した。しかしながら、健康状態という抽象的な概念は少なからず解析の方向性を不明瞭していくことが判明した。理由として考えられたのは、例えば10年分のデータであったとしてもほとんどのデータが欠測していることと、例えば毎年のデータであっても劇的な変化はあまりないという「状態」であった事があげられる。一方、多くても一年に一度しか取得されないデータであるため、詳細な健康変化を追跡するには粗すぎる経時データということも当初の計画にそぐわないことがわかった。

そこで視点を変えて、健康状態が現実社会へと反映されるものとして医療費に着目した。健診データには医療費という項目は存在しないが、診療報酬(レセプト)データには詳細な記載がなされている。現実的に考えて健康な状態であっても、年に一つ二つの疾病に罹患したり、持病なりを持っているケースが大半であり、健康状態が医療費として反映されるのは社会実装を考える上でも非常に意味のある方向性だと考えるに至った。このような理由から、最終的には健診データとレセプトデータを連結させて利用するという方向性を固めたが、全く異なるデータを同一個人間で連結させるのはそれ相応の労力と時間と費用をかけて慎重に行う必要がある。データが連結されるまで本来の目的である解析手法の開発を保留しておく事は望ましくないため、連結後データの形式を念頭に置きつつ、ひとまずは医療費という点からレセプトデータだけから解析手法を開発することにした。

次にデータ成型についてだが、最終的に健診データと連結することを前提にして、データでは1年間でのタイムスパンで考え、一人につきどの診療行為等の項目が何度発生したかをカウントする形に成型し、性別や年齢など基本項目などはその年を通じて固定させた。医療費に関しては指定した期間内で発生した金額の総計とするようなデータ成型を行った。一人につき項目が2~3万という形になり、ほとんどが「該当なし」の0回というデータとなった。それが数万から数十万人まで集まった暫定データを、健診データとの連結が出来るまでの当面のデータとして用いて、以下の学会発表には用いた。健診データと連結されても項目数が増えるだけのデータとなり、解析手法に関しては特段の変更が必要ないという事で、このようなデータ形式を採用した。

(2) 解析手法の開発

本研究では、研究当初の狙いは健康状態の遷移という事にあったが、健康状態という概念が非常に抽象的であったため、次のように解析手法開発の方向性を修正した。すなわち、ある疾患に罹患した人の健康状態を現実世界に反映したのが医療費という対象であり、その平均的な変遷を追跡することに焦点を当てて解析を行うことにした。

本研究で用いるデータの構造が、一人につき非常に多項目でスパースなデータであり、かつ現実社会で発生した観察データであるため、解析手法にはその点で注意を要した。そこで全データを、着目する疾病に罹患した群としていない群にレセプトの診断名に応じて分けて、その群間の背景を調整した上で医療費の平均値を出すことを考えた。こうすることで着目する疾患で一年

間にどれだけの医療費が使われることになるのかが見えることになる。

具体的手法としては、調査観察データの背景を調整する事のできる傾向スコアを用いて、逆確率重みづけ法で計算することで求めた。ここで重要なのは傾向スコアの推定であるが、これは機械学習で頻繁に用いられているL1正則化法を採用する事で情報量基準の観点からも望ましい推定が行えた。

4．研究成果

(1) 傷病別医療費の推計手法

上記のように、リアルワールドからの医療データを扱うことの難しさがあったが、それでもそこから有意義な指標を見出すことを目的として若干の方針変更をした結果、指定した傷病毎に年間一人当たり幾らの医療費が使われたのかという推定方法を提案する事が出来た。これまでも傷病別で医療費を推計するという手法は提案されてはいたが、診療コードの発生を集計したような概算であり、科学的、統計的な手法を用いての推定ではなかった。それでも、一般的な単純平均とは異なる概算なので、それらの先行研究では特許が権利化されている。本研究で行った手法は、国内ではじめて統計学的な基盤に即した傷病別の医療費推定の手法である。ビッグデータであっても単純平均された傷病別医療費しか計算されていなかった場合と比べると、背景調整をされた本研究の手法は、より真の値に近づいていると言える。また、異なる年のデータであっても同じ傷病ならば、その年の医療費推計値はほぼ同じで推移することも確認することができ、信頼性も高いと言える。それにより、その研究成果で特許申請を行い、権利化された。その中で学会発表を行った。今後の展望としては、本手法をより簡素な形で定型化して社会実装を目指していくのが一つの方向である。

(2) マルチレベル分析の応用先への提言

本研究においては、本来は健診データの時系列変化に対しての解析を行うことを主眼に置いていた。しかしながら、前述のように方向転換を行ってレセプトデータと合わせることで医療費へと解析対象をシフトしたのは上述の通りである。しかしながら、そのシフトチェンジの過程で、混合効果モデル(マルチレベル分析)を用いての解析手法も検討していたことから、ソーシャルキャピタル研究で頻繁に用いられているマルチレベル分析の深刻な誤用に気づくに至った。それを学会で報告すると共に、別件で依頼されていたマルチレベル分析の解説を行ったソーシャルキャピタル関連書籍にもそのことを明記した。このような副次的な成果で統計学の分野からは既知の事であったが、ソーシャルキャピタル研究においては全世界的な誤用された論文を確認しているため、この分野の今後の統計手法に関して大きな提言になったと言える。研究開始当初は全く予期しなかった分野への貢献となったが、今後はこの分野での正しい用いられ方が浸透していくことを期待したい。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 浅川剛, 大谷誠, 藤本賢治, 永田昌子, 藤野善久, 松田晋哉
2. 発表標題 統合データベースに基づく傷病別医療費推計の手法開発
3. 学会等名 第1回 日本メディカルAI学会
4. 発表年 2019年

1. 発表者名 浅川剛, 藤野善久
2. 発表標題 ソーシャルキャピタル研究におけるマルチレベル解析について
3. 学会等名 第29回 日本疫学会学術集会
4. 発表年 2019年

〔図書〕 計1件

1. 著者名 近藤 克則編	4. 発行年 2020年
2. 出版社 ミネルヴァ書房	5. 総ページ数 288
3. 書名 ソーシャル・キャピタルと健康・福祉	

〔出願〕 計0件

〔取得〕 計1件

産業財産権の名称 傷病別医療費推計装置および方法並びにプログラム	発明者 松田晋哉, 藤野善久, 藤本賢治, 浅川剛	権利者 同左
産業財産権の種類、番号 特許、6537121	取得年 2020年	国内・外国の別 国内

〔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
連携研究者	松田 晋哉 (Matsuda Shinya) (50181730)	産業医科大学・医学部 公衆衛生学・教授 (37116)	
連携研究者	藤野 善久 (Fujino Yoshihisa) (80352326)	産業医科大学・産業生態科学研究所・教授 (37116)	