

科学研究費助成事業 研究成果報告書

令和元年6月18日現在

機関番号：12601

研究種目：研究活動スタート支援

研究期間：2017～2018

課題番号：17H06629

研究課題名(和文)医療ビッグデータを用いた慢性疾患発症及び重症化の早期予測モデル構築

研究課題名(英文) Establishing prediction models for onset of chronic diseases using medical receipt database

研究代表者

大野 幸子(Ono, Sachiko)

東京大学・大学院医学系研究科(医学部)・特任助教

研究者番号：20797237

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：本研究では、レセプト・健診データを用いて、甲状腺機能亢進症発症、糖尿病患者の腎機能低下を複数の機械学習手法および従来の疫学手法検討を用いて検討した。後者の研究は国際学会で発表を行った。さらに、予測モデル構築における検査値データの有用性を検討するため、協力施設から得られたデータを用いて、腹部悪性手術後の創部感染、抗菌薬の再開、術後入院期間の予測を行った。創部感染、抗菌薬再開は症例数が少なく予測不能であったものの、Random Forestを用いた術後入院期間の予測値と実測値の差は10分位それぞれで2日以内であった。

研究成果の学術的意義や社会的意義

現状の研究利用可能な医療レセプトデータは、サンプル数、説明変数、追跡期間が限られており、機械学習の手法で重症化や疾患の発症予測は困難であった。一方、術後入院期間は一定の精度で予測することが確認されたため、既存の医療レセプトデータは機械学習を用いた短期的アウトカムの予測に利用できる可能性が示唆された。

研究成果の概要(英文)：In this study, we investigated the onset of hyperthyroidism and renal dysfunction in diabetic patients by using multiple machine learning methods and conventional epidemiologic methods. The latter study was presented at an international conference. In addition, in order to examine the usefulness of laboratory data in the construction of prediction models, we used data obtained from cooperating institutions to predict surgical site infection after abdominal surgery, resumption of antibiotics, and postoperative length of stay. Although the number of cases of surgical site infection and resumption of antibiotics were small and unpredictable, the difference between the predicted value and the observed value of postoperative length of stay using Random Forest was within 2 days in each deciles.

研究分野：臨床疫学

キーワード：医療ビッグデータ 予測

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

生活習慣病を始めとする慢性疾患では、その発症や重症化の危険性がある集団を早期に特定し、行政や保険者が対策を講じる事が公衆衛生の観点から重要である。既存の研究から糖尿病のリスクスコアなどが知られているものの、予測可能時期は疾患の発症や重症化と近接しており、早期介入は困難である。近年急速に発達してきた機械学習は膨大なデータを分析し、有用な規則や判断基準などから、アルゴリズムを発展させる手法である。その予測能力から医学領域では、疾患発症・重症化の早期シグナル検出への応用が期待されているものの、研究利用可能な医療データベースの不足、機械学習の手法に精通する臨床研究者の不足から十分な研究がなされていなかった。

2. 研究の目的

既存の医療ビッグデータに機械学習手法を適用し、種々の慢性疾患発症・重症化をより早期に予測するモデルを構築し、疾患発症・重症化リスクのある集団の早期特定に寄与すること。

3. 研究の方法

本研究では、既存の医療ビッグデータに対して機械学習の手法を適用し慢性疾患の発症、重症化等のアウトカムを予測することができるかについて検討を行った。

- (1). JMDCデータを用いた甲状腺疾患の発症予測
- (2). JMDCデータを用いた糖尿病患者の腎機能低下予測
- (3). 協力施設から得られた検査値データを用いた腹部悪性腫瘍患者の術後感染、抗生剤再開、術後入院期間の予測

4. 研究成果

- (1). JMDCデータを用いた甲状腺疾患の発症予測
SQL 言語を用いてJMDCより取得したレセプト・健診データの整形、結合を行い、甲状腺機能亢進症の予測モデル構築に必要なテーブル作成を行った。データに機械学習の各手法を適応し、Random forestでは重症バセドウ病の新規発症を陽性的中率100%、特異度83%で予測するモデルを構築した。一方、感度は30%未満となっており、今後の精度向上が課題として残された。
- (2). JMDCデータを用いた糖尿病患者の腎機能低下予測
JMDCデータを用いて新規に血糖降下薬を開始した糖尿病患者において、ガイドラインアドヒアランスが腎機能の低下に与える影響について検討を行った。複数の機械学習手法および従来の疫学手法において同様の結果が得られ、ガイドラインアドヒアランスは腎機能の低下と関連しなかった。(表1)当該研究については国際学会で発表を行った。

表1. ガイドライン記載項目実施と腎機能低下(2年間のeGFR変化)

	ガイドライン記載項目				
	実施	未実施	差	95%信頼区間	P 値
全対象者	-2.62	-2.73	0.11	-2.74 to 2.95	0.941
正常血圧者	-0.54	-2.68	2.14	-1.61 to 5.88	0.263

- (3). 協力施設から得られた検査値データを用いた腹部悪性腫瘍患者の術後感染、抗生剤再開、術後入院期間の予測

予測モデル構築における検査値データの有用性を検討するため、協力施設から得られた日毎の検査値を含む医療データベースを用いて、腹部悪性手術後の創部感染、抗菌薬の再開、術後入院期間の予測を行った。創部感染、抗菌薬再開は症例数が少なく予測不能であったものの、Random Forestを用いた術後入院期間の予測値と実測値の差は10分位それぞれで殆どが2日以内であった。(表2, 図1)

表2. 10分位数ごとの術後入院日数

術後入院日数		
分位数	予測値	観測値
1	9.47	10.5
2	11.3	11
3	12.7	11.3
4	13.9	12.5
5	15.1	14.7
6	16.4	15.8
7	17.7	18.7
8	18.9	18.6
9	21.5	18.8

10 27.7 30.3

図1.術後入院日数の予測値と観測値



横軸：観測値，縦軸：予測値，単位：日

5. 主な発表論文等

〔雑誌論文〕(計 0 件)

〔学会発表〕(計 1 件)

Ono S, Ono Y, Koide D, Yasunaga H. Association between Albuminuria Testing and Subsequent Change in Estimated Glomerular Filtration Rate: A Japanese Non-Elderly Cohort Study 2018 AcademyHealth 2018.06.23-25. Seattle.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称:

発明者:

権利者:

種類:

番号:

出願年:

国内外の別:

取得状況(計 0 件)

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕
ホームページ等

6. 研究組織

(1)研究分担者

研究分担者氏名：
ローマ字氏名：
所属研究機関名：
部局名：
職名：
研究者番号(8桁)：

(2)研究協力者

研究協力者氏名：
ローマ字氏名：

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。