

令和 2 年 4 月 29 日現在

機関番号：14301

研究種目：若手研究

研究期間：2018～2019

課題番号：18K14950

研究課題名（和文）医療用ビッグデータを用いた薬剤有害事象検出：機械学習によるデータ駆動型アプローチ

研究課題名（英文）Anomaly detection of drug-adverse events from large-scale database: data-driven analysis with machine learning

研究代表者

竹内 正人 (Takeuchi, Masato)

京都大学・医学研究科・准教授

研究者番号：80598714

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：当該研究は、医療費請求データなどの大規模既存データより、機械学習による薬剤有害事象の異常検知が可能であるかを探索的に検討した。題材として血糖降下薬の一つであるSGLT2阻害薬と、骨粗鬆症治療薬の一種であるビスフォスフォネート製剤の二つを選択し、有害事象として前者は腎機能低下や脱水など、後者には顎骨壊死を用いた。機械学習の手法により異常点の検知自体は可能であったが、手法の選択やそれぞれの事象の特定に用いる病名や薬剤のコードの選択に大きく依存する結果であった。

研究成果の学術的意義や社会的意義

病名の入力に関しては医療機関・医師レベルにより方針が異なるのが日本の実情と思われる。機械学習を用いた検出においても、その精密さはこれら診療プラクティスの違いに大きく依存した。したがって、この技術を実用面で用いる前提として、病名入力の方針について何らかの統一的な指針を整備することが必要であることを示唆した研究である。

研究成果の概要（英文）：We explored whether, using machine learning techniques, automated anomaly detection of drug-adverse events was feasible from large-scale database (eg, claims data). We selected two case examples; the first example was SGLT-2 inhibitor and its adverse events such as renal insufficiency or dehydration, and the second one was bisphosphonate and osteonecrosis of the jaw. Machine learning techniques successfully detected the anomaly points from a time-series data of large-scale database. However, the detected anomaly points were sensitive to the model or the administrative code selection of adverse event.

研究分野：薬剤疫学

キーワード：薬剤疫学 機械学習 時系列データ

## 様式 C-19、F-19-1、Z-19 (共通)

### 1. 研究開始当初の背景

薬剤による有害事象の早期発見、および迅速な対応は公衆衛生上の課題の一つである。これまでのところ薬剤有害事象の検出は、医療機関などからの自発的な有害事象報告に依存し、記述疫学的な検討、あるいはシグナル検知と呼ばれる統計解析手法などが用いられてきた。しかし、自発報告には (1) 分母に相当する被疑薬剤を使用している人数が不明である、(2) 有害事象の必ずしも全例が報告されない、(3) 併存疾患や併用薬剤に関する情報が十分ではない、および (4) 自由記載などの非構造化データが多い、などの問題点が指摘されている。これらのため、現状では被疑薬剤と有害事象との因果関係の推定、あるいは有害事象が疑われた場合の正確かつ迅速な対応が困難な状況にある

近年の様々な医療データベース (DB) の整備に伴い、これら DB を薬剤有害事象の検出に使用できないか模索する動きもあるものの、その適切な活用法は定まっていない。推察される理由として、従来からある統計手法が必ずしも医療 DB に最適化されていないこと、あるいは未知のパターンに関しての検出力が弱いなどが挙げられる

医療以外の分野に目を向けると、機械学習の手法が、様々な異常パターンの自動的な検知に用いられ、日常業務に組み入れている産業領域もある。このよう異常点検知の手法が、医療 DB において薬剤有害事象シグナルを検知可能かは検証されていない。例えば、新規薬剤が上市された後で、ある特異的なイベントの数が急激に上昇した際には、その上昇が当該薬剤の (予期されない) 有害事象の初期シグナルを捉えている可能性がある。このようなシグナルが、薬剤の有害事象であるかの因果関係に関しては別の面からの調査検討が必要である。しかし、シグナルを捉えられないことには以降の対策に動き出すことができないことも一方で事実である

### 2. 研究の目的

異常点検知と呼ばれる機械学習の手法が、医療用 DB を用いた薬剤有害事象の検出に応用できるかを探索的に検証する。

### 3. 研究の方法

本研究では2つの異なる医療 DB を用い、既知の薬剤—有害事象の組み合わせが機械学習の手法により検出可能かを探索的に調べた。

#### (1) SGLT2 阻害薬

データソース：株式会社 JMDC が保有する 100 を超える健康保険組合データ (累計加入者数 500 万人) を用い、2014 年に発売された新しいクラスの 2 型糖尿病薬である SGLT2 阻害薬よその有害事象を例とした。SGLT2 阻害薬はそれまでの糖尿病薬の有害事象である低血糖などの他に、脱水・腎機能低下・骨折などのこのクラスに特有の有害事象があることが知られている。

対象：2015—2018 年の JMDC データより、以下の病名が付された患者数を年次ごとに分けて数え上げた：骨折 (部位を問わず)、脱水、低血糖、膝炎、悪性腫瘍 (固形腫瘍のみ：部位問わず)、肝障害、腎機能障害、静脈血栓症、ケトアシドーシス、尿路感染症、高カリウム血症、四肢切断。これらは既知の SGLT2 阻害薬の有害事象として報告されているものを文献報告より事前に特定した。データベースからの抽出には病名コードである ICD10 コードを用いたが、四肢切断のみは手技コードである K コードを用いた。また、一人の人が複数のイベントを経験する可能性を考慮して、人数ではなく病名件数での報告とした。

解析方法：統計ソフト R とそのパッケージである anomalize を用いた。このパッケージでは異常点検出の方法として、Generalized Extreme Studentized Deviate 検定が採用されている。上記対象で述べられた有害事象の総数を月別に分類した時系列データとして取り扱い、この検定によりその前後の時点からの異常な増加・減少が統計的な特異点として検出可能かを検定した。有害事象に関しては季節性 (夏に多くなる脱水など) が見られるものから、時系列データは decomposition 処理、つまりはトレンドと季節性と誤差の 3 つの成分に分解し、トレンド成分のみを解析対象とした。

#### (2) ビスフォスフォネート (BP) 製剤

データソース：国内で最も悉皆性の高い医療データベースであるレセプト情報・特定健診等情報データベース (NDB) を用い、BP 製剤とその特異的な有害事象である顎骨壊死との関連を調べた。NDB へは京都大学病院オンサイトセンターにてアクセスを行なった。このため、データ解析環境はオンサイトセンターで実装されているものを使用し、また厚生労働省から事前に許可された情報のみこの報告書への記載が可能であることを付け加えておく。

対象：2010 年から 2018 年までの間に (1) BP 製剤を使用、かつ (2) 使用より後の時点で顎骨壊死あるいはその関連病名が付された患者数を数え上げた。この有害事象に関しては、その出現時期が使用開始直後から数年、あるいは中止後など、一定の window 期間を設定することが難しいことから、特に使用開始と病名付与までの interval の設定を行わなかった。病名抽出にあたっては、二つの定義を用いた。一つは、特異度を重視した定義で、この定義では「ビスフォスフォネート関連顎骨壊死」「骨吸収抑制剤関連顎骨壊死」など有害事象との関連が強く疑われる 4 つの病名を用いた。しかし、その薬剤との関連が明らかではない (あるいは薬剤がありながら、注視されていない) ものに関しては、単に「顎骨炎」「顎骨骨髓炎」「顎骨壊死」などの病名が付与

されている可能性を考慮し、幅広く計 16 個の病名により定義づけを行なった。一人に複数回の顎骨壊死に関連する病名が付されている場合には、その初回のみを解析対象とした。

解析方法：統計ソフト R とそのパッケージである strucchange を用いた。このパッケージでは変化点検知の手法として、Generalized Fluctuation Tests を採用しており、この手法は anomaly detection 手法としては変化点検知と呼ばれる手法に該当し、時系列データが線形であるという仮定を必要とする。上と同じく、データは decomposition 処理を行い、トレンド成分を解析対象とした

#### 4. 研究成果

##### (1) SGLT2 阻害薬

対象となった約 600 万件が有害事象定義を満たし、内訳としては脱水が最も多く、ついで肝障害の順であった。

decomposition 処理後に、3 つの異常点のクラスターが観察され、SGLT2 阻害薬発売後の 2014 年以降は、そのうちの 1 つのみであった。

95%信頼区間と共に表示した結果、2014 年以降の異常点は統計的に有意な結果ではないものと判断された。

図 1

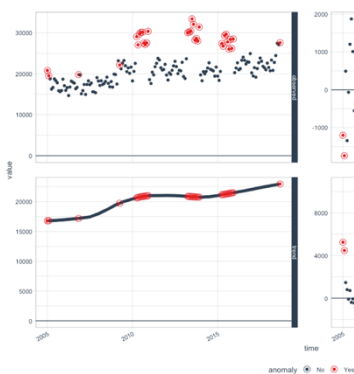
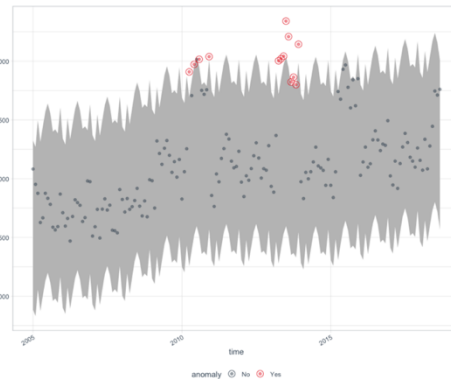


図 2



##### (2) BP 製剤

得られた患者背景を表 1 に示す。

表 1：患者背景

定義	患者数 (女性割合)	年齢 (歳) <sup>a</sup>	暴露から診断までの日数 <sup>a</sup>
特異度重視	14490 人 (88%)	81 (74-86)	1675 (823-2526)
感度重視	85722 人 (87%)	77 (72-84)	1216 (577-2009)

a：中央値とカッコ内は四分位範囲

それぞれの定義において検知された時点を表 2 に示す。

表 2：検知された変化点

定義	変化点
特異度重視	<ul style="list-style-type: none"> <li>2015 年 5 月</li> <li>2017 年 8 月</li> </ul>
感度重視	<ul style="list-style-type: none"> <li>2011 年 2 月</li> <li>2012 年 7 月</li> <li>2013 年 12 月</li> <li>2015 年 5 月</li> <li>2017 年 2 月</li> </ul>

##### (3) 考察

医療 DB においても異常点の自動的な検知が可能であったが、(1) SGLT2 阻害薬の例においてはその異常点は SGLT2 阻害薬の発売前のみであり、(2) BP 製剤においては病名抽出の定義 (アルゴリズム) により、有害事象の数が変わったのみならず、異常点の同定にも大きく影響した。

この原因を情報学の観点から考察を行うと、アルゴリズムの問題以上に、医療 DB には真の有意義な情報であるシグナルと、一見シグナルに似ているが偽陰性情報であるノイズとの識別が困難であることが考えられる。BP 製剤の例においては、「ビスフォスフォネート関連顎骨壊死」などの病名を用いることで、おそらくシグナルを正確に捉えることが可能であったものと思われる。

る。しかし、このような病名づけは薬剤と有害事象との関連が明らかになって初めて、病名コードの体系に盛り込まれる。したがって、このような精密な病名づけが行えないような初期段階においては、感度を重視した幅広い病名にて抽出を行う必要があるが、これにより検出される異常点の時点数自体が異なることが本研究において明らかになった。シグナルとノイズとの関係性に関しては、SGLT2 阻害薬の事例においても同様の考察が行える。知られている有害事象は（BP 製剤と異なり）SGLT2 阻害薬に特異的なものはなく、このため「SGLT2 阻害薬関連 XX」のような病名づけは困難であり、また仮にコードとして体系化したにしても、臨床の現場においては判断が難しければ、そのようなシグナル情報となるような病名づけがなされるかは疑問である。つまりは、シグナルとノイズが混じった情報のままであり、利活用が困難な状況が予測される。

シグナルとノイズの混じった情報から、シグナル情報のみを取り出す方法に関して、薬剤疫学においては複数の性質の異なる情報を結びつける、という方法が用いられることがある。例えば糖尿病患者を DB から抽出する際に患者定義を、「糖尿病の病名があり、かつ糖尿病に対して使用される薬剤が 2 回以上処方されている患者」などとするのがその一例である。しかし、全ての疾患・有害事象において、このような組み合わせを用いる方法が可能とは限らない。例えば今回のケースである顎骨壊死においては、うがい（軽症例）、抗生剤、手術（重症例において行われ、術式は複数ある）などが挙げられるが、一方で経過観察という患者もいる。うがい薬や抗生剤については、顎骨壊死の治療薬として処方されたかの判別が困難という問題もあり、今回は病名のみを患者抽出要件として用いた。

アルゴリズムの観点から簡単な考察を行う。今回使用した以外にも、異常点検知の手法は複数ある。しかし、情報・DB の構造とそれに適した手法の組み合わせに関してはあまり議論されていない。したがって、今回使用した Generalized Extreme Studentized Deviate 検定や Generalized Fluctuation Tests 以外の手法でより良いパフォーマンスが発揮される可能性は残る。しかしながら上に述べたように、シグナルとノイズの混在した情報での使用においてはいずれの手法においても困難が予想される。

## 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 Takeda Chikashi, Takeuchi Masato, Mizota Toshiyuki, Yonekura Hiroshi, Nahara Isao, Kuwauchi Aki, Joo Woo Jin, Kawasaki Yohei, Kawakami Koji	4. 巻 33
2. 論文標題 Utilization of arterial pulse waveform analysis during non-cardiac surgery in Japan: a retrospective observational study using a nationwide claims database	5. 発行年 2019年
3. 雑誌名 Journal of Anesthesia	6. 最初と最後の頁 159 ~ 162
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s00540-018-02609-x	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 Takeda Chikashi, Takeuchi Masato, Kawasaki Yohei, Yonekura Hiroshi, Nahara Isao, Kuwauchi Aki, Yoshida Satomi, Tanaka Shiro, Kawakami Koji	4. 巻 33
2. 論文標題 Prophylactic sivelestat for esophagectomy and in-hospital mortality: a propensity score-matched analysis of claims database	5. 発行年 2019年
3. 雑誌名 Journal of Anesthesia	6. 最初と最後の頁 230 ~ 237
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s00540-018-2602-9	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 Takeuchi Masato, Kawakami Koji	4. 巻 7
2. 論文標題 Association between Hemoglobin and Hemoglobin A1c: A Data-Driven Analysis of Health Checkup Data in Japan	5. 発行年 2018年
3. 雑誌名 Journal of Clinical Medicine	6. 最初と最後の頁 539 ~ 539
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/jcm7120539	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 Masato, Nakano Sayuri, Tanaka-Mizuno Sachiko, Nishiyama Chika, Doi Yuko, Arai Masaru, Fujii Yosuke, Matsunaga Toshiyuki, Kawakami Koji	4. 巻 42
2. 論文標題 Adherence and Concomitant Medication Use among Patients on Warfarin Therapy: Insight from a Large Pharmacy Dispensing Database in Japan	5. 発行年 2019年
3. 雑誌名 Biological and Pharmaceutical Bulletin	6. 最初と最後の頁 389 ~ 393
掲載論文のDOI (デジタルオブジェクト識別子) 10.1248/bpb.b18-00576	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 Kawano Sumihiro, Takeuchi Masato, Tanaka Shiro, Yamashita Takehiro, Sakamoto Taiji, Kawakami Koji	4. 巻 63
2. 論文標題 Current status of late and recurrent intraocular lens dislocation: analysis of real-world data in Japan	5. 発行年 2018年
3. 雑誌名 Japanese Journal of Ophthalmology	6. 最初と最後の頁 65 ~ 72
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s10384-018-0637-y	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 Takeuchi Masato, Ito Shuichi, Nakamura Masaki, Kawakami Koji	4. 巻 20
2. 論文標題 Changes in Hemoglobin Concentrations Post-immunoglobulin Therapy in Patients with Kawasaki Disease: A Population-Based Study Using a Claims Database in Japan	5. 発行年 2018年
3. 雑誌名 Pediatric Drugs	6. 最初と最後の頁 585 ~ 591
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s40272-018-0316-y	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 Seki Tomotsugu, Takeuchi Masato, Miki Ryusuke, Kawakami Koji	4. 巻 34
2. 論文標題 Follow-up tests and outcomes for patients undergoing percutaneous coronary intervention: analysis of a Japanese administrative database	5. 発行年 2018年
3. 雑誌名 Heart and Vessels	6. 最初と最後の頁 33 ~ 43
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s00380-018-1224-3	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 Kinoshita Takuya, Tanaka Shiro, Inagaki Mieko, Takeuchi Masato, Kawakami Koji	4. 巻 17
2. 論文標題 Prescription pattern and trend of oral contraceptives in Japan: A descriptive study based on pharmacy claims data (2006?-?2014)	5. 発行年 2018年
3. 雑誌名 Sexual & Reproductive Healthcare	6. 最初と最後の頁 50 ~ 55
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.srhc.2018.06.004	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 Takeuchi Masato, Ogura Masahito, Minoura Takaaki, Inagaki Nobuya, Kawakami Koji	4. 巻 95
2. 論文標題 Comparative Effectiveness of Sodium-Glucose Cotransporter-2 Inhibitors Versus Other Classes of Glucose-Lowering Medications on Renal Outcome in Type 2 Diabetes	5. 発行年 2020年
3. 雑誌名 Mayo Clinic Proceedings	6. 最初と最後の頁 265 ~ 273
掲載論文のDOI (デジタルオブジェクト識別子) doi:10.1016/j.mayocp.2019.12.004	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2件(うち招待講演 0件/うち国際学会 0件)

1. 発表者名 竹内正人、川上浩司
2. 発表標題 慢性腎臓病を合併した2型糖尿病患者における腎障害の進展：実臨床データに基づくSGLT2阻害薬と他クラス薬剤の比較
3. 学会等名 第24回日本薬剤疫学会学術総会
4. 発表年 2018年

1. 発表者名 竹内正人、川上浩司
2. 発表標題 SGLT2 阻害薬の 2 型糖尿病患者に対する腎保護作用：日本人実臨床データを用いた予備的解析
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

〔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
---------	---------------------------	-----------------------	----