

平成 30 年度厚生労働科学研究費補助金（政策科学総合研究事業
（臨床研究等 ICT 基盤構築・人工知能実装研究事業））
分担研究報告書

DPC データ研究の実践と人材育成プログラムの開発・検証

研究分担者 長瀬隆英 東京大学医学部附属病院 呼吸器内科 教授
研究代表者 康永秀生 東京大学大学院医学系研究科 臨床疫学・経済学教授

研究要旨

【目的】DPC データベースを用いた研究を実践し、人材育成プログラムを開発する。

【方法】DPC データベースを用いた研究の実践を通して、臨床医が大規模データベース研究に取り組む際の問題点を明らかにし人材育成プログラムの開発と検証を行う。

【結果】H30 年度には当研究班から、臨床医のクリニカルクエスチョンにもとづく DPC データベースを用いた研究の成果が 48 編、英文誌に掲載された。その内 6 編が呼吸器内科医師によるものであった。呼吸器内科では、国内学会に 2 編、国際学会に 2 編報告した。研究の遂行に際して、専門家のサポート・指導が不可欠であった。

【考察】クリニカルクエスチョンの構造化、データの抽出、解析、論文または学会報告は大規模データベースを用いた研究における主要な要素であるが、これらを臨床医のみで行うことは困難であった。臨床疫学、統計、データサイエンティスト等による効率的なサポートを行うためには、臨床医を対象とした人材育成プログラムの確立が必要である。

研究協力者

城大祐 東京大学大学院医学系研究科ヘル
スサービスリサーチ講座 特任准教授

A．研究目的

臨床医が DPC データベースを用いた研究
を実践するために必要と考えられる知識や
技能を身に着けるための人材育成プログラ
ムを、DPC データベース研究の実践を通
して開発・検証することを目的とする。

B．研究方法

DPC データベースを使用した研究の実践
を通して、人材育成プログラムの開発・検

証のための問題点を検討した。

臨床医が、臨床上の種々の疑問をクリニカ
ルクエスチョンとして持ち寄ったテーマ
を、リサーチクエスチョンに構造化した。
構造化されたりサーチクエスチョンに解答
するために最適な研究デザイン・統計手法
を検討し、研究を行うために必要な DPC
データの抽出を依頼した。DPC データベ
ースは 2010 年 7 月から 2017 年 3 月まで
を用いた。抽出されたデータを種々の統計
ソフトを用いて解析した。得られたデータ
の解釈について議論を行い、英語論文にま
とめ、投稿した。前述の過程において、デ
ータベース研究を行う臨床医にとって問題

となった点、困難であった点について、グループディスカッションによりまとめた。

C. 研究結果

呼吸器内科医による6つの研究について、以下にその内容を概説する。

1) パーキンソン症候群が高齢肺炎患者の転帰に及ぼす影響についての検討

【背景】

肺炎はパーキンソン病の患者さんが入院の原因で最も多いものである。本研究は、肺炎で入院した高齢者においてパーキンソニズムが及ぼす影響について評価することを目的とした。

【方法】

60歳以上の肺炎入院患者について、日本の入院データベースを用いて後ろ向きコホート研究を行った。年齢、性別について、パーキンソニズムがある症例とない症例を、1対4でマッチングした。多変量回帰分析により、在院死亡、在院日数、自宅退院について評価した。

【結果】

パーキンソニズムのある症例はない症例と比較して、有意に低い在院死亡率(オッズ比, 0.81; 95% 信頼区間, 0.74-0.89)を示した。パーキンソニズムのある症例では、在院日数は、8.1%長かった。パーキンソニズムのある症例は、自宅退院率が低かった(オッズ比, 0.62; 95% 信頼区間, 0.58-0.67)。

【結語】

パーキンソニズムは、肺炎の患者において、在院死亡の独立した危険因子ではなかったが、在院日数の延長と自宅以外への退

院と関連していた。

2) シスプラチンを含む肺がん化学療法とカルボプラチンまたはネダプラチンを含む化学療法の静脈血栓発症のリスクについての検討

【背景】

化学療法が担がん患者の静脈血栓症(VTE)の危険因子となることはよく知られており、その中でもシスプラチン(CDDP)は最もよく研究されている。しかしながら、このCDDPが、カルボプラチン(CBDCA)やネダプラチン(CDGP)と比較して、VTEのより大きな危険因子であるかどうかについてはよく分かっていない。この研究では、VTEの院内発症を呈した肺がん患者の特徴を解析し、関連する危険因子の同定、さらにCDDPとCBDCA/CDGPを使用した際のリスクの比較を行う。

【方法】

我々は、DPCデータベースにて、2012年4月から2015年3月までにプラチナ製剤ベースの化学療法を施行した肺がん患者をレトロスペクティブに抽出した。一般化推定方程式によるロジスティック回帰を用いて、化学療法の内容を含む様々な因子とVTE発症との関係を解析した。

【結果】

235,104名の患者のうち、675名(0.29%)の患者において、プラチナ製剤ベースの化学療法施行後にVTEを院内発症した。多変量解析では、年齢・ADL・侵襲的治療がVTEの有意な危険因子であった。さらに、CDDPベースの化学療法は、CBDCA/CDGPベースのものよりも、

VTE の発症頻度が高かった(調整後オッズ比 1.35、95%信頼区間 1.08-1.68、 $P<0.01$)。

【結語】

肺がん患者において、CDDP ベースの化学療法は、CBDCA/CDGP ベースの化学療法と比し、VTE のより大きな危険因子であると結論付けた。

3) プロペンシテスコア法と操作変数法を用いた、術後 stage II-III A 非小細胞肺がん患者における術後補助化学療法としての経口フルオロウラシル療法とシスプラチン+ナベルピン療法の比較検討

【背景】

術後 stage II-IIA 非小細胞肺がん患者における術後補助化学療法はシスプラチン+ナベルピン療法が標準治療であったが、近年経口フルオロウラシル療法も代替療法としての報告がみられる。両者についてプロペンシテスコア法と操作変数法を用いて、比較を行った。

【方法】

2010 年 7 月から 2015 年 3 月までに術後補助化学療法を受けた術後 stage II-III A 患者を DPC データより抽出した。シスプラチン+ナベルピン群が 384 例、経口フルオロウラシル群が 268 例であった。両群をプロペンシテスコアにて調整後、多変量解析にて無再発生存期間を比較した。未測定の変数因子に対応するため操作変数法も併用した。

【結果】

若年症例と N2 陽性症例にてシスプラチン+ナベルピン療法が実施される割合が多かった。高齢症例と T3N0 症例では経口フル

オロウラシル療法が実施される割合が多かった。プロペンシテスコア法にて調整した 172 組での検討では、経口フルオロウラシル群では術後補助化学療法が開始されるまでの期間がシスプラチン+ナベルピン群と比較して有意に短く、またその無再発生存期間が有意に延長していた。操作変数法でも同様の結果を得た。

【結論】

DPC データでの検討では、術後 stage II-III A 非小細胞肺がん症例において、経口フルオロウラシル療法はシスプラチン+ナベルピン療法と比較して無再発生存期間を延長させ、代替レジメンとして有用であると考えられた。

4) イリノテカン忍容性に関する半夏瀉心湯の有用性の検証

【目的】

イリノテカン(CPT-11)は、大腸がんや肺がんなど幅広い癌腫に用いられる抗癌剤であるが、副作用として重篤な下痢を呈することがある。半夏瀉心湯(TJ-14)は、CPT-11 誘発性下痢への有効性が報告されているが、少数例の報告でありその有効性は十分に検証されていない。我々は TJ-14 が CPT-11 の忍容性を改善するかを検討した。

【方法】

2010 年 7 月から 2016 年 4 月までの厚生労働科学研究 DPC 研究班データを用いた。CPT-11 が入院で初回投与された 90,437 症例を対象とし、CPT-11 投与日以前の TJ-14 開始を TJ-14 群、それ以外をコントロール群とした。傾向スコアを用いた逆確率重み付け(stabilized IPTW)を

行い両群間の背景バランスを調整し、CPT-11 の投与回数、投与量と在院死亡率を比較した。統計解析には、Mann-Whitney-U 検定、一般化線形モデルのロジステック回帰とポワソン回帰を行った。P<0.05 を有意とした。操作変数法により未測定の交絡について考慮した解析も行った。

【結果】

stabilized IPTW 後の TJ-14 群は 7,092 人、コントロール群は 82,019 人であった。TJ-14 群は CPT-11 投与回数 (incidence rate ratio, 1.34; 95% confidence interval, 1.31 - 1.38) と、投与量 (incidence rate ratio, 1.16; 95% confidence interval, 1.14 - 1.19) が有意に多かった。在院死亡率は TJ-14 群で有意に低かった (Odds ratio, 0.81; 95% CI, 0.71 - 0.93; P=0.002)。操作変数法でも同様の結果を得た。

【結語】

TJ-14 は、CPT-11 の忍容性を改善し、在院死亡率を低下させた。TJ-14 は CPT-11 の補助療法として推奨される。

5) 肺がん診断目的の気管支鏡入院における、Hospital volume と在院死亡の関連の検討

【背景】

超音波を用いた気管支鏡検査や分子標的薬の進歩により臨床医は肺がんの診断的気管支鏡を以前より全身状態の悪い患者にも行うようになっている可能性がある。肺がんの診断目的の気管支鏡の安全性についての知見は乏しい。肺がん患者における診断的気管支鏡検査の在院死亡と Hospital

volume との関連を検討した。

【方法】

2010 年 7 月から 2014 年 3 月までに入院して施行した肺がん患者の診断目的の気管支鏡検査を検討した。年間の Hospital volume は、非常に低い病院 (50 件/年以下)、低い病院 (51 - 100 件/年)、高い病院 (101 - 300 件/年)、とても高い病院 (300 件/年より多い病院) の 4 群に分けた。気管支鏡検査後 7 日以内の在院死亡を主たるアウトカムとし、一般化推定方程式を用いた多変量ロジステック回帰分析により患者の背景因子について調整したうえで、Hospital volume との関連を検討した。

【結果】

954 病院の計 77,755 人に肺がん診断目的の気管支鏡が施行されていた。気管支鏡検査後 7 日以内の在院死亡は 0.5% であった。Hospital volume の非常に低い病院と比較して、高い病院 (オッズ比 = 0.69, 95% 信頼区間: 0.52-0.92, p = 0.010) は有意に低い 7 日以内の在院死亡率を示し、非常に高い病院 (オッズ比 = 0.67; 95% 信頼区間: 0.43-1.05, p = 0.080) も同様の傾向を示した。超音波を用いた気管支鏡検査では、有意に低い 7 日以内の在院死亡率を示した。

【結語】

肺がんの診断目的の気管支鏡検査の高い在院死亡率は、Hospital volume と逆相関の関係を示した。肺がんの診断目的の気管支鏡のリスクを認識することは重要であり、超音波を用いた気管支鏡検査の導入は診断目的の気管支鏡の死亡リスクを減少させる可能性がある。

6) 後期高齢 COPD 患者における、漢方薬の大建中湯による増悪リスクの減少効果

【背景】

有症状の COPD 患者では、長時間作用型の抗コリン薬 (LAMA) を含む気管支拡張薬の使用が推奨されている。しかし、特に後期高齢の COPD 患者では、気管支拡張薬が抗コリン作用による副反応を引き起こす可能性がある。日本の漢方薬である大建中湯は日本で最も処方されており、主には腹部膨満や便秘に対して処方される。本研究の目的は、後期高齢の COPD 患者において、大建中湯の支持療法としての役割を検討することである。

【方法】

我々は、DPC データベースで、75 歳以上の COPD 増悪入院患者を同定した。大建中湯が処方されていた群と処方されていなかった群を 1 対 4 でマッチングし、再増悪入院または死亡のリスクを比較した。コックス比例ハザードモデルを用いて比較した。さらに、LAMA が処方されていた症例と処方されていなかった症例についてサブグループ解析を行った。

【結果】

大建中湯が処方されていた患者では、有意に再増悪入院または死亡のリスクが低く、1 対 4 マッチング後のハザード比は 0.82 (95%信頼区間; 0.67-0.99) であった。LAMA 処方患者のサブグループでは、大建中湯処方症例で有意に再増悪入院または死亡のリスクが低下していたが、LAMA 処方のない患者では差がみられなかった。

【結語】

我々の結果は、大建中湯の処方が、後期高齢の COPD 患者において LAMA の忍容性

を改善したことにより、再増悪による入院または死亡のリスクを減少させた可能性を示唆した。大建中湯の処方は、後期高齢の COPD 患者に対する有用な支持療法となる可能性がある。

D. 考察

今年度も、昨年に引き続き、DPC データベースを用いた研究の実践をとおして、臨床医が大規模データベース研究を行う上での問題点を検討した。今年度は、大学病院勤務の呼吸器内科医師だけでなく、DPC データを用いた関連病院の呼吸器内科医師との研究も開始した。

臨床医によるクリニカルクエストは、DPC データベースを用いた臨床疫学研究には不可欠である。他方、臨床医がデータベースを用いた臨床疫学研究に必要なリテラシーを独学で身に着けることは困難であるばかりでなく、誤った解析や解釈を行う危険をも伴う。今年度の研究で行った種々の検討においても、前年度と同様に専門的なサポートや指導が不可欠であった。以下に、臨床医に共通して認識・理解の不足がみられた点を列挙し、その後に対応について論じる。

1) DPC データベース: データベース一般についての知識 (パネルデータについて、解析に用いるデータテーブルについて、データベースの悉皆性についてなど) と DPC データベース固有の情報について (個々のデータ項目、個人の追跡は施設が変わると出来ないこと、他のデータベースとの結合が許容されないことなど) は、ほとんどの臨床医において理解が乏しい。関

連の論文から DPC 研究の概要はわかるが、解析に必要な知識を十分に得ることは容易ではない。

2) 研究デザイン：臨床医は、一般に関連文献の検索法、横断研究か記述的研究か分析的研究か、コホート研究かケースコントロール研究か、クリニカルクエスションからリサーチクエスションへの構造化などを系統的に学ぶ機会に乏しい。長所と短所を理解した研究デザインが重要であるが、そのための知識が不足している。

3) 統計解析の知識と手法：データの分布とばらつきの理解、適切な統計手法の選択、欠損値の扱い方、統計解析ソフト利用の実際、解析結果の解釈など専門家との協業が、効率よくかつ誤りのない研究を行うための解決策であると考えられた。

大規模データベース研究に関連した各領域の専門科と協業するためのコミュニケーションを円滑に行うには、臨床医に対する系統的かつ持続的な教育プログラムが必須と考えられた。その具体的な方法としては、臨床医が自身の都合に合わせて参照可能かつ作成者もアップデートが容易な電子媒体でのテキストやマニュアルの作成、e-learning の活用などが有効と考えられた。一方で、個々の臨床医が研究の実践によってのみ得られる気づきもあり、よりよいプログラムには研究の実践も必要不可欠な要素であると考えられた。

E . 結論

DPC データベースを用いた研究の実践は、人材育成プログラムの確立に必要な情報を得るために有効な方法であると考えら

れた。更に継続的に行う必要がある。

G . 研究発表

1. 論文発表

(1-6)

1. Jo T, Yasunaga H, Michihata N, Sasabuchi Y, Hasegawa W, Takeshima H, Sakamoto Y, Matsui H, Fushimi K, Nagase T, Yamauchi Y. Influence of Parkinsonism on outcomes of elderly pneumonia patients. *Parkinsonism Relat Disord.* 2018;54:25-9.
2. Mitani A, Jo T, Yasunaga H, Sakamoto Y, Hasegawa W, Urushiyama H, Yamauchi Y, Matsui H, Fushimi K, Nagase T. Venous thromboembolic events in patients with lung cancer treated with cisplatin-based versus carboplatin/nedaplatin-based chemotherapy. *Anticancer Drugs.* 2018;29(6):560-4.
3. Urushiyama H, Jo T, Yasunaga H, Michihata N, Matsui H, Hasegawa W, Takeshima H, Sakamoto Y, Hiraishi Y, Mitani A, Fushimi K, Nagase T, Yamauchi Y. Oral fluorouracil vs vinorelbine plus cisplatin as adjuvant chemotherapy for stage II-III non-small cell lung cancer: Propensity score-matched and instrumental variable analyses. *Cancer Med.* 2018;7(10):4863-9.
4. Urushiyama H, Jo T, Yasunaga H, Michihata N, Yamana H, Matsui H, Hasegawa W, Hiraishi Y, Mitani A,

- Fushimi K, Nagase T, Yamauchi Y. Effect of Hangeshashin-To (Japanese Herbal Medicine Tj-14) on Tolerability of Irinotecan: Propensity Score and Instrumental Variable Analyses. *J Clin Med*. 2018;7(9).
5. Hiraishi Y, Jo T, Michihata N, Hasegawa W, Sakamoto Y, Urushiyama H, Matsui H, Fushimi K, Nagase T, Yasunaga H, Yamauchi Y. Hospital Volume and Mortality following Diagnostic Bronchoscopy in Lung Cancer Patients: Data from a National Inpatient Database in Japan. *Respiration*. 2019;97(3):264-72.
6. Jo T, Michihata N, Yamana H, Sasabuchi Y, Matsui H, Urushiyama H, Mitani A, Yamauchi Y, Fushimi K, Nagase T, Yasunaga H. Reduction in exacerbation of COPD in patients of advanced age using the Japanese Kampo medicine Dai-kenchu-to: a retrospective cohort study. *Int J Chron Obstruct Pulmon Dis*. 2019;14:129-39.
2. 学会発表
1. Hiraishi Y, Jo T, Yamauchi Y, Urushiyama H, Nagase T, and Yasunaga H. Complications in bronchoscopy: Data from the Japanese Diagnosis Procedure Combination database. 20th WCBIP/WCBE World Congress. 2018
2. 城 大祐、道端伸明、山名隼人、漆山博和、笹淵裕介、松居宏樹、伏見清秀、康永秀生、長瀬隆英、山内康宏、大建中湯の慢性閉塞性肺疾患後期高齢患者における再増悪入院または死亡リスク低下効果の検討
日本臨床疫学会 第2回年次学術大会
2018年
3. 漆山博和、城 大祐、道端伸明、山名隼人、笹淵裕介、松居宏樹、伏見清秀、康永秀生、長瀬隆英、山内康宏
半夏瀉心湯のイリノテカン忍容性に対する改善効果の検証
日本臨床疫学会 第2回年次学術大会
2018年
4. Jo T, Yamauchi Y, Urushiyama H, Hiraishi Y, Mitani A, Tanaka G, Yasunaga H, and Nagase T. Effect of dai-kenchu-to on COPD exacerbations in elderly patients: A retrospective study using a nationwide database in Japan, 23rd Congress of the APSR (Asian Pacific Society of Respiriology). 2018
- H . 知的財産権の出願・登録状況
- 1.特許取得
なし
- 2.実用新案登録
なし
- 3.その他
なし

平成 30 年度厚生労働科学研究費補助金（政策科学総合研究事業
（臨床研究等 ICT 基盤構築・人工知能実装研究事業））
分担研究報告書

NDB データ研究の実践と人材育成プログラム開発・検証

研究分担者 中山健夫 所属 京都大学医学研究科 健康情報学分野
研究協力者 加藤源太 所属 京都大学医学部附属病院. 診療報酬センター

要旨: レセプト情報・特定健診等情報データベース(NDB)の利用にあたって、利用希望者があらかじめ準備・理解しておくべき事項を整理すると共に、NDB を含む医療ビッグデータ研究の推進のために必要とされる人材像の検討を行った。医療ビッグデータ研究に求められる人材像としては、まず研究シーズを有する臨床医、臨床研究者、疫学研究、公衆衛生学の専門家は従来型の各大学講座単独で育成できた人材と言える。医療ビッグデータ研究においては、それに加えて、データ入手の各種手続き、データ利用の根拠法等に通じた人材、一次データの解釈に長けた人材、情報解析基盤構築の専門家、情報分野の法制等にも造詣が深い医療倫理専門家なども必要とされる。適切かつ効率的に医療ビッグデータ研究を進めるには分野横断的な議論・連携が可能となる環境の整備と、他領域の専門家と協働できる研究者のコミュニケーション能力の涵養が期待される。

A. 目的

レセプト情報・特定健診等情報データベース(NDB)の利用にあたって何が必要か、利用希望者があらかじめ準備しておくべきことを明らかにする。

- ・データ構造がわかりにくく、解釈が難しい
- ・大学院生の単独利用は認められていない
- ・民間利用が制限されている

以上から NDB の利用に際して留意すべきこととして、以下が挙げられる。

B. 研究方法

既存資料による検討。

- ・NDB 提供のルール・スケジュール等の把握
スケジュール、ガイドライン、必要な手続きは厚労省のホームページ等を参照して把握

C. 研究結果

1. NDB 利用に際する一般的な認識(2018 年時点)

- ・厚生労働省の審査手続きが煩雑で、厳しい
- ・研究者への提供件数が限られている
- ・研究者への提供までに、長期間を要する
- ・東京大学・京都大学に設置されているオンサイトセンターが試行的利用段階に留まっている
- ・データ分析に高度な統計分析スキルが必要

- ・関係講演等で情報を収集し、分析能力を涵養
- ・利用者本人が一定の知識を修得した上で、NDB の利用経験者や、研究テーマに関する臨床家や病院事務担当者の実地の経験を聴取する。レセプトデータが発生する状況、それに伴う研究目的での利用に際する限界や落とし穴について理解を深めることが不可欠。

NDB の利用にあたって、研究者は二次利用ゆえの限界を十分踏まえる必要がある。NDB データの第一義は「診療報酬請求」であり、「項

目が不十分」「データが分かりにくい」などといった研究者目線での指摘は、逆に診療報酬請求の現場からすれば妥当とも言える。厚生労働省の所管部署である「保険局」の第一のミッションは医療保険の円滑な遂行であり、「研究目的でのNDBの精度向上」を主たる目的として予算を確保、運用することは容易ではないことと思われる。

2. NDB オンサイトリサーチセンターの動向

現在は試行期間中にあり、東大と京大で、模擬申出に対する研究を漸次進めている。京大では臨床医学系の教室など、レセプトデータを扱った経験は無いが、関心を持っている研究者らに試験的に利用環境を提供している段階である(2ブース)。今後は国とともに、学外研究者への公開を目指している。ただし厚労省からは、担当者を雇用する予算は充てられておらず、大学病院のスタッフが一定の-effortをオンサイトセンターの管理、利用者の支援に充てている状態である。

同様のシステムとして一つの目標となるのが台湾厚労省のオンサイトセンターである。台湾では衛生福利部内統計處に、端末は26台のオンサイトセンターが設置されている。このほかにも国立台湾大学をはじめとする複数の提携先機関に端末が66台設置されており、合計で92台の端末が管理されている。これらの端末のオンサイト端末設置空間、および各端末のインターフェイスはモニターされており、衛生福利部内にある端末から遡って監視することができる。

3. 医療ビッグデータ研究に必要な人材

研究シーズを有する臨床医、臨床研究者、疫学研究、公衆衛生学の専門家は従来型の各大学講座単独で育成できた人材と言える。医療ビッグデータ研究においては、それに加えて、データ入手の各種手続き、データ利用の根拠法等に通じた人材、一次データの解釈に

長けた人材(例:レセプトなら医療事務担当者)、情報解析基盤構築の専門家、安全な情報利用を推進する人材(例:情報分野の法制等にも造詣が深い医療倫理専門家)等、円滑な研究の推進、知見の蓄積に、これらの人材も不可欠と言える。適切かつ効率的に医療ビッグデータ研究を進めるには分野横断的な議論・連携が可能となる環境の整備と、他領域の専門家と協働できる研究者のコミュニケーション能力の涵養が期待される。

医療等ICT推進に資する人材育成の重要性について「第62回社会保障審議会医療部会(2018.06.02)」では以下のように議論されている。

“認定匿名加工医療情報作成事業者がデータ利活用基盤として適切に機能していくためには、こうした基盤を適切に構築・運営できる人材や、匿名加工医療情報を適切に利活用できる人材の養成・確保が必要である。”

“基盤の構築・運営を担う人材については、医療についての基礎的な知識を有し、医療情報の具体的内容や規格等を理解し、医療分野の研究開発に資する医療情報を適確に収集するとともに、収集した医療情報を安全に管理しつつ、利活用の目的に応じて適切な匿名加工を行うことが求められる。”

“...、医療情報の利活用により解決すべき研究の課題を適切に設定し、データ基盤構築人材と対話しつつ、適切に匿名加工医療情報を取得・分析できるようにする必要がある。”

“このため、データ利活用基盤を構築・運営する能力や医療情報を利活用する能力を育成する取組を通して、我が国全体として必要とされる人材を計画的かつ確実に養成・確保する観点から、キャリアパスの在り方の検討、育成の場としての大学や認定匿名加工医療情報作成事業者、学会等の連携を含め、人材の育成を継続・充実させる。”

4. 医療ビッグデータ・データベースの疫学的活用の課題

医療ビッグデータ・データベースは必ずしも仮説が事前に明確でない観察研究の、良い意味でも良くない意味でも宝庫と言える。事前に計画されていない、サブグループ解析、多重比較、交互作用など、さまざまな後付け解析 (post-hoc analysis) が可能となり、そこから得られた知見をどう扱うか、慎重な検討が不可欠である。「ビッグデータ」への過剰な期待、過信があると、「ビッグデータから得られた結論」が無批判に関係者に受け入れられ、大きな影響を持ってしまう懸念もある。同様に、「出版バイアス」「アウトカム報告バイアス」が臨床試験以上に発生する可能性も想定されるであろう。観察研究の事前登録は、議論のあるところであり、定着はしていないが、今後、その在り方の検討を進める必要があると考えられる。

得られた知見の解釈に関しては、N数が膨大なため、通常の統計的有意水準では有意となるが、関連の強さは弱い (weak association) または介入の効果サイズは小さい結果をどう解釈すべきか議論が必要であろう。重要な未測定交絡が存在する可能性が高ければ、「弱い関連」は、見かけのものに過ぎない場合もあり得る。また個々の臨床レベルから政策レベルまでの意思決定のエビデンスとして、ビッグデータから得られた知見をどのように位置づけるかも議論が必要である。ランダム化比較試験と対比してどのように考えるべきか、診療ガイドラインの推奨決定にどのように役立てるのが適切か、医薬品早期承認制度への活用の在り方など、重要な検討課題となるものと考えられる。

D. 考察 & E. 結論

NDBの利用希望者があらかじめ準備・理解しておくべき事項を整理すると共に、NDBを含む医療ビッグデータ研究の推進のために必要とされる人材像の検討を行った。医療ビッグデータ研究に求められる人材像として

は、従来型の専門家に加えて、データの2次利用に関わる経験を持つ専門家が必要であり、適切かつ効率的に医療ビッグデータ研究を進めるには分野横断的な議論・連携が可能となる環境の整備と、他領域の専門家と協働できる研究者のコミュニケーション能力の涵養が期待される。

F. 健康危機情報

なし

G. 研究発表

1. 中山健夫. 産官学シンポジウム医療データヘルス改革 医療ビッグデータ構築とデータが生み出す変革の可能性 座長基調講演. 医療科学研究所全 東京都渋谷 社協・灘尾ホール 2018年5月19日
2. 中山健夫. 第4回日本医薬品安全性学会特別講演 ビッグデータから見る医薬品安全性:現状と展望. 2018年8月18日 岡山県倉敷市芸文館
3. 中山健夫. 第51回日本薬剤師学術大会特別講演薬剤師がデータを正しく活用するために :ビッグデータ・AI時代の課題と期待. 石川県立音楽堂 2018年9月24日
4. 中山健夫. 京都大学における人材育成の試み 日本臨床疫学会第2回年次学術大会 シンポジウム「保健医療介護ビッグデータ研究の人材育成」 京都大学 2018年9月30日

H. 知的所有権の取得状況

なし

平成 30 年度厚生労働科学研究費補助金
(政策科学総合研究事業(臨床研究等 ICT 基盤構築・人工知能実装研究事業))
分担研究報告書

NDB データ研究の実践と人材育成プログラム開発・検証

研究分担者 小林廉毅 東京大学大学院医学系研究科公衆衛生学・教授

研究要旨:本分担研究では、レセプト情報・特定健診等情報データベース(National Database of Health Insurance Claims and Specific Health Checkup: 以下 NDB)から解析用データを抽出する作業を効率的に実施する方法を検討するため、高血圧・糖尿病・脂質異常症・認知症の各疾患で使われる計 10 薬剤について、性・年齢・都道府県等で層別化したジェネリック医薬品シェアの年度・月毎の推移を算出することを具体的な課題として取り上げ検証した。NDB という非常にデータサイズが大きいデータから必要なデータの抽出を行うためには、効率的なデータアクセスが求められる。本研究では、(1) NDB は Oracle Exadata Database Machine というデータウェアハウスに格納されており、通常の Oracle Database と挙動が異なる点があること、(2) NDB は非常にデータサイズが大きいにも関わらず、オンサイトセンターではユーザーの表領域の割り当て容量に厳しい制限があることの 2 点を考慮し検証を行った。前年度に構築した仮想環境を利用しながら、クエリの計画、テスト環境での実行によるクエリの妥当性の確認、オンサイトセンターでのクエリの実行、結果の評価、クエリの改善計画の検討、という PDCA サイクルを繰り返していきながら、クエリのパフォーマンスの改善を検討した。本分担研究で作成された資料は、NDB からデータを抽出、分析する作業を実行可能な高度人材を育成するプログラムにおいて、有用な教材になると考えられる。

研究協力者

伊藤佑樹(東京大学)

原 湖楠(東京大学)

宮脇敦士(東京大学)

舩島一樹(リクルートライフスタイル)

A. 研究目的

本分担研究では、NDB オンサイトセンターで NDB(National Database of Health Insurance Claims and Specific Health Checkup)からデータを抽出する作業を効率的に実施する方法を検討するため、高血圧・糖尿病・脂質異常症・認知症の各疾患で使われる計 10 薬剤

について、性・年齢・都道府県等で層別化したジェネリック医薬品シェアの年度・月毎の推移を算出することを具体的な課題として取り上げ検証した。

NDB という非常にデータサイズが大きいデータから必要なデータの抽出を行うためには、効率的なデータアクセスが求められる。本研究では、(1) NDB は Oracle Exadata Database Machine というデータウェアハウスに格納されており、通常の Oracle Database と挙動が異なる点があること、(2) NDB は非常にデータサイズが大きいにも関わらず、オンサイトセンターではユーザーの表領域の

割り当て容量に厳しい制限があること、の2点を考慮することが重要となった。

B. 研究方法

1. SQL チューニングの方針

SQL チューニングの方針は主に2つで、(1) Oracle Exadata Database Machine の性能を最大限活用するために、FULL SCAN をし、Smart Scan をかけること、(2)表領域が限られている、かつ扱うデータ量が多いため TEMP 落ちのエラー (PGA に入りきらず、一時表領域を使い、一時表領域にも入りきらない場合のエラー)にならないように出来るだけメモリを使用しないこと、であった。

(1)を実現するために、INDEX を用いず直接データにアクセスし可能な限りパーティションを利用することを心がけた。

(2)の点に対応するために、(a)ソートが必要とする UNION 等の関数の使用は最低限にする、(b)3 つ以上のテーブルの JOIN・サブクエリ・UNION ALL は用いずに複数のクエリに処理をわけ、必要に応じて、必要最小限の変数を中間テーブルに格納する、といった内容で対応した。

ただし、この方策だけでは今回の場合は対応しきれなかったため、年度毎にクエリを繰り返して、年度毎に最終生成物を作成していった。

前年度に構築した仮想環境を利用しながら、クエリの計画、テスト環境での実行によるクエリの妥当性の確認、オンサイトセンターでのクエリの実行、結果の評価、クエリの改善計画の検討、という PDCA サイクルを繰り返していきながら、クエリのパフォーマンスの改善を検討していった。

2. 本研究で扱った課題

本研究では、高血圧・糖尿病・脂質異常症・認知症といった慢性疾患で通院する患者について、後発医薬品と先発医薬品の使用割合の変化や差異、これらに関する要因を複数

年度の NDB データを用いて明らかにすることを目的としたコードを作成することとした。データの抽出条件と抽出した統計量を以下に記す。

2 - 1. 抽出条件 (NDB データ)

a. 期間

平成21年4月診療分から平成27年3月診療分

b. レセプトの種類

(1) 医科レセプト

(2) 調剤レセプト

c. レセプトの抽出条件

医科レセプト・調剤レセプトのうち、傷病名として、高血圧症・糖尿病・脂質異常症・認知症を有するもの

d. 薬剤の成分名

i. 高血圧症患者

アムロジピン、ロサルタン、カンデサルタン

ii. 糖尿病患者

ボグリボース、グリメピリド、ピオグリダゾン

iii. 脂質異常症患者

ピタバスタチン、アトロバスタチン、シンバスタチン

iv. 認知症患者

ドネペジル

2 - 2. 抽出する統計量

別表1 (別添資料1) の形での表が作成できる形での数値の抽出を目的とした。一薬剤に対して、後発医薬品と先発医薬品の使用割合を地域・月ごとに求めた。また、性別・年代別・地域別・自己負担割合別・公費補助の有無でも層別分析が可能な形での抽出を行った。

C. 研究結果

上記の方策を用いて試行錯誤を繰り返すことで、与えられた表領域で現実的な時間内 (高血圧症約39時間、糖尿病約144時間、脂質異常症約26時間、認知症約76時間、の合計約285時間) に処理を終えることが実現できた。最終的なクエリとその内容の詳細 (README ファイル) は、ならびに結果のテー

ブルを出力した CSV ファイルは当教室ホームページの研究プロジェクト「レセプト等データベース(NDB データ)を用いた後発医薬品の普及状況および関連要因に関する研究」の項に公開している。(<http://publichealth.m.u-tokyo.ac.jp/>)

CSV 出力された結果のうち、東京都の結果について別添資料2に図示した。別添資料2の図1 図10は各薬剤に対する、年齢層別ジェネリック置き換え率の推移である。ドネペジル以外の9つの薬剤については、30歳代～60歳代では80歳以上に比べ、一貫してジェネリック置き換え率が高い。70歳代は30歳代～60歳代と80歳以上との中間に位置している。20歳代では、今回解析の対象となった薬剤の処方頻度が多くないため、ジェネリック置き換え率にばらつきが生じているが、30歳代～60歳代と80歳以上との中間に位置していることが多い。

この年齢層によるジェネリック置き換え率の違いについてはいくつかの説明が考えられる。1つ目は、自己負担率の違いである。自己負担率が高いほど、先発薬とジェネリックとの価格差が大きくなるため、合理的な選択では、ジェネリックを選択するインセンティブが高くなる。観察期間中、わが国の30歳代から60歳代は公的医療保険での自己負担は3割、80歳以上は殆どで1割、70歳代は1割、3割が混在している。故に、高年齢になるほど、その他の条件が同じであれば、ジェネリックを選択するインセンティブが小さくなる。さらに、30歳代～60歳代、70歳代、80歳以上のそれぞれのグループ内で比較的揃っているジェネリック置き換え率は、自己負担割合の影響が強いことを示唆している。しかし、20歳代では30歳代～60歳代と同じ自己負担割合であるにもかかわらず、ジェネリック置き換え率が低いことや、男女では、やや男性の方がジェネリック置き換え率が高い傾向にあることは、自己負担割合だけで説明できない要因があることを示唆している。

2つ目は、各年齢層における先発品への選好の差である。高齢者ほど、健康状態が悪く、自分の体に対するリスクを心配して、使い慣れてきた先発薬からジェネリックに変更することを厭うかもしれない。これは、本解析期間中にジェネリックが発売された、アムロジピン、カンデサルタン、ロサルタン、ピオグリタゾン、ピタバスタチンの結果において、80歳代以上で立ち上がりが遅いことと矛盾しない。この選好の差は、所得の差、持っている情報の差、合併症の差、住んでいる地域の差(プロバイダーの影響を介して)などを反映していると考えられる。

他方、ドネペジルは年齢層による、ジェネリック置き換え率の違いがはっきりしない(40歳代以下については、ドネペジル処方数量が少ないため、結果を示していない)。これは、ドネペジルを内服している者は、認知症に罹患しているため、価格差をあまり意識しない、あるいは患者本人が選択をしていないことなどが原因と思われる。

時系列に注目すると、ジェネリックの発売から比較的時間のたっている薬剤においても、ジェネリック置き換え率は、少なくとも観察期間の間ではプラトーに達していない。先行研究では、先発薬に対する選好が、きわめて強い集団が一定数おり、ジェネリック置き換え率は一定期間後プラトーに達する可能性が指摘されているが、本研究ではそれは観察されなかった。しかし、本研究ではジェネリック置き換え率が高くてもせいぜい7割であり、まだジェネリックに置き換えられる余地があっただけかもしれない。より長期的な観察が期待される。

別添資料2の図11 図20は80歳代、90歳以上における、自己負担割合別のジェネリック置き換え率推移である。概ね、3割負担群のほうが1割負担群よりジェネリック置き換え率が低いことが見て取れる。これは、前述した自己負担率が高いほど、ジェネリック置き換え率が下がるという仮設とは一致しない結果となっているが、特に高齢者では3割負担群と1割負担群

で所得・資産を含む社会経済的因子が大きく異なることに注意しなければならない。例えば、3割負担群のほうが高所得であるため、所得効果により、先発品に対する選好は高くなる。そのため、3割負担群では、価格差によるジェネリック薬への選好と所得効果による先発薬への選好が逆向きに働いており、そのバランスによって、3割負担群と1割負担群のジェネリック置き換え率の差が現れていると考えられる。少なくとも、今回の薬剤に対する解析では、東京都においては、所得効果による(所得効果以外もあるかもしれないが)先発薬への選好が価格差によるジェネリックへの選好を上回っていると考えられる。

D. 考察

PDCA サイクルを繰り返すことで、最終的には与えられた制約の下で、現実的な時間内に日本の全人口をカバーするNDB データでの、本研究課題で意図したデータ処理作業を終えることができた。

試行錯誤の過程では、TEMP 落ちのエラーを解決するのに苦労した。当初はクエリの実行が終了するまでの時間を出来るだけ短くするために、中間テーブルを作成しないようにしていたが、TEMP 落ちのエラーが発生したことによって、徐々に中間テーブルを増やしていくこととなった。

最終的には、元テーブルにアクセスする場合には常に必要最小限の変数を一旦中間テーブルとして格納し、それらの中間テーブルをマージし、そこからまた必要最小限の変数を一旦中間テーブルに格納し、不要な中間テーブルは削除するなどのように非定型抽出によるデータ整形を少しずつ進めていきながら、TEMP 落ちのエラーは発生させることなく、目的の最終生成物を得ることに成功した。

また、かかる時間は長くなってしまふものの、TEMP 落ちのエラーを発生させないようにする最も単純な方策は、あるカテゴリー毎に分けて

クエリを実行していくことである。今回の場合、最終的なクエリでは、年度毎にクエリを繰り返して、年度毎に最終生成物を作成することにしたことが挙げられる。ただし、NDB オンサイトセンターが今後本格的に運用されることを考えると、クエリの実行時間の制限が必要になるので、この方策による対応には限界があり、基本的には、研究方法で述べた SQL チューニングの二つの方針によって、クエリの最適化を目指すべきである。

疾患による実行時間の差は、それぞれの疾患で扱うデータ量を鑑みると、予想以上に大きかった。NDB オンサイトセンターでは、他に使用者がいる場合にクエリの実行に待ち時間が発生するため、疾患による実行時間の差が大きかったのは、このような理由によるものであった可能性がある。

E. 結論

NDB オンサイトセンターでのデータの抽出は表領域が限られている、かつ扱うデータ量が多いものであったものの、本研究では、SQL 文の工夫によって、与えられた表領域で現実的な時間内に処理を終えることが実現できた。

本分担研究で作成された資料は、NDB データを抽出、分析する作業を実行可能な高度人材を育成するプログラムにおいて有用な教材になると考えられる。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

小林廉毅 . (特集:レセプト情報・特定健診等情報データベース(NDB)を用いた研究の動向) 1) NDB 利活用推進の取り組み NDB オンサイトリサーチセンターについて . Progress in Medicine 2018; 38(2): 119-122

2. 学会発表

なし

H. 知的財産権の出願・登録

なし

別添資料 1

別表 1

表. □□疾患治療における後発医薬品と先発医薬品の使用割合の変化（薬剤名：○○○○）

地域	2009年 4月	2009年 5月	・・・	△△年△月 ジェネリック ク薬上市	・・・	□□年 □月効能 追加	・・・	2015年 3月
□□県								
20歳代男								
20歳代女								
□□県								
30歳代男								
30歳代女								
・・・								
□□県								
90歳代以上男								
90歳代以上女								
○○県								
20歳代男								
20歳代女								
・・・								
◎◎県								
90歳代以上男								
90歳代以上女								

表. □□疾患治療における後発医薬品と先発医薬品の使用割合の変化（薬剤名：○○○○）

地域	2009年 4月	2009年 5月	・・・	△△年△月 ジェネリック ク薬上市	・・・	□□年 □月効能 追加	・・・	2015年 3月
自己負担割合								
3割								
(公費補助無)								
自己負担割合								
3割								
(公費補助有)								
自己負担割合								
1割								
(公費補助無)								
自己負担割合								
1割								
(公費補助有)								

上記について、性別、年齢別、地域別の層別分析を行う。

別添資料 2

図1 アムロジピンの年齢層別ジェネリック置き換え率推移



図2 アトルバスタチンの年齢層別ジェネリック置き換え率推移

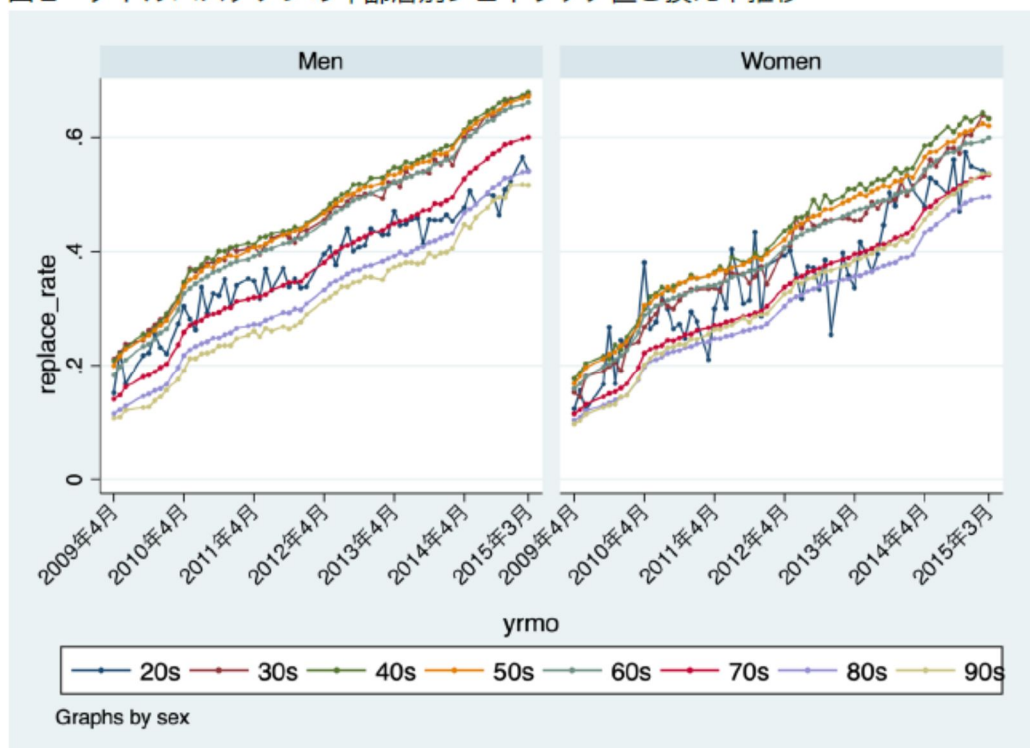


図3 カンデサルタンの年齢層別ジェネリック置き換え率推移

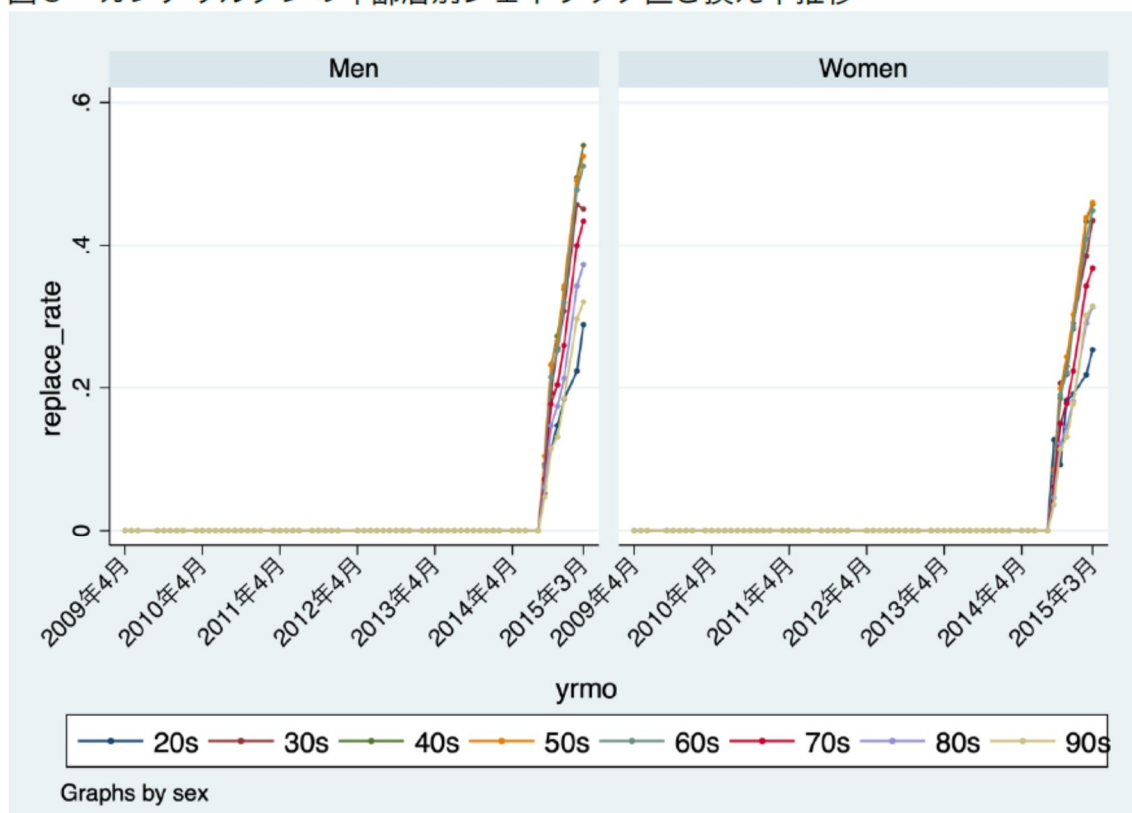


図4 ドネペジルの年齢層別ジェネリック置き換え率推移

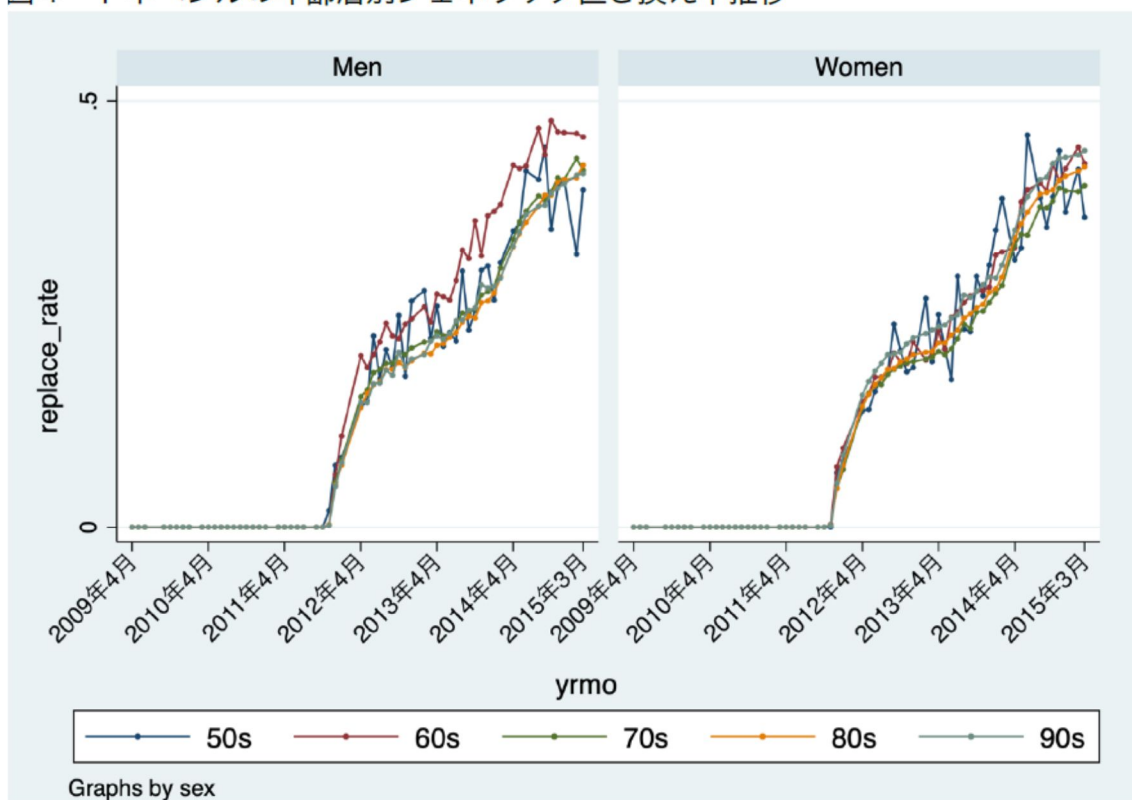


図5 グリメピロドのジェネリック置き換え率

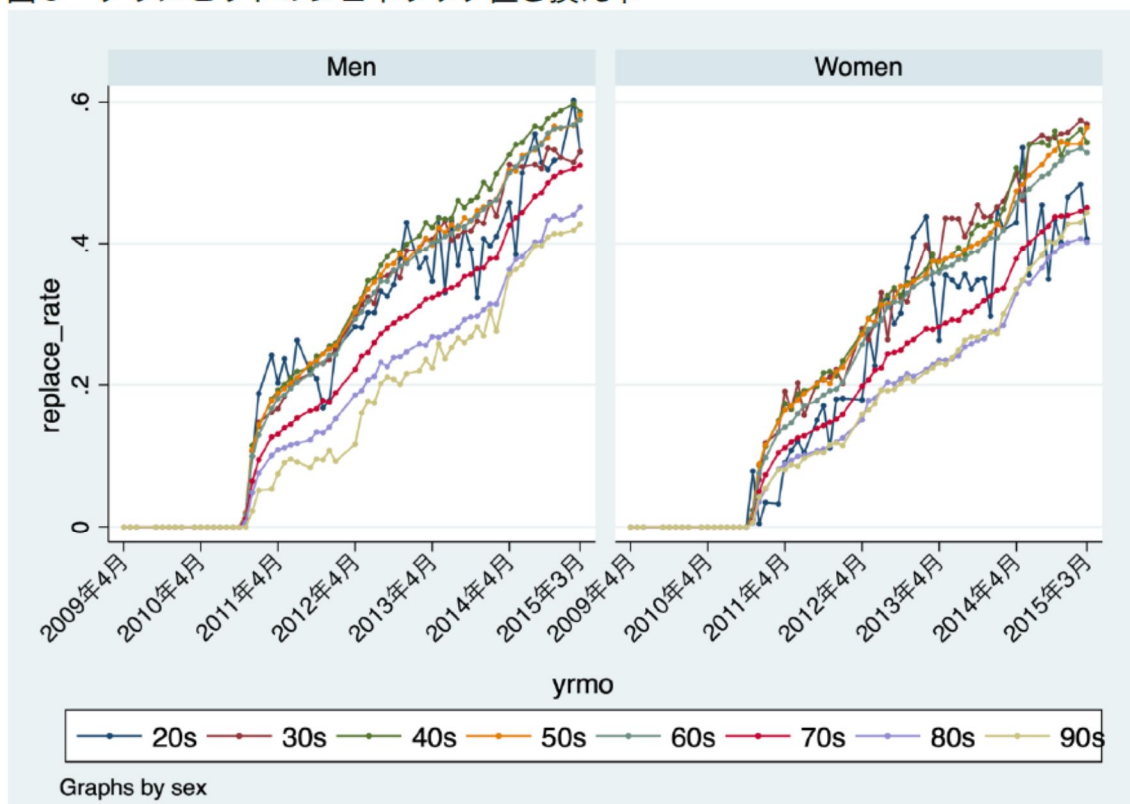


図6 ロサルタンの年齢層別ジェネリック置き換え率推移

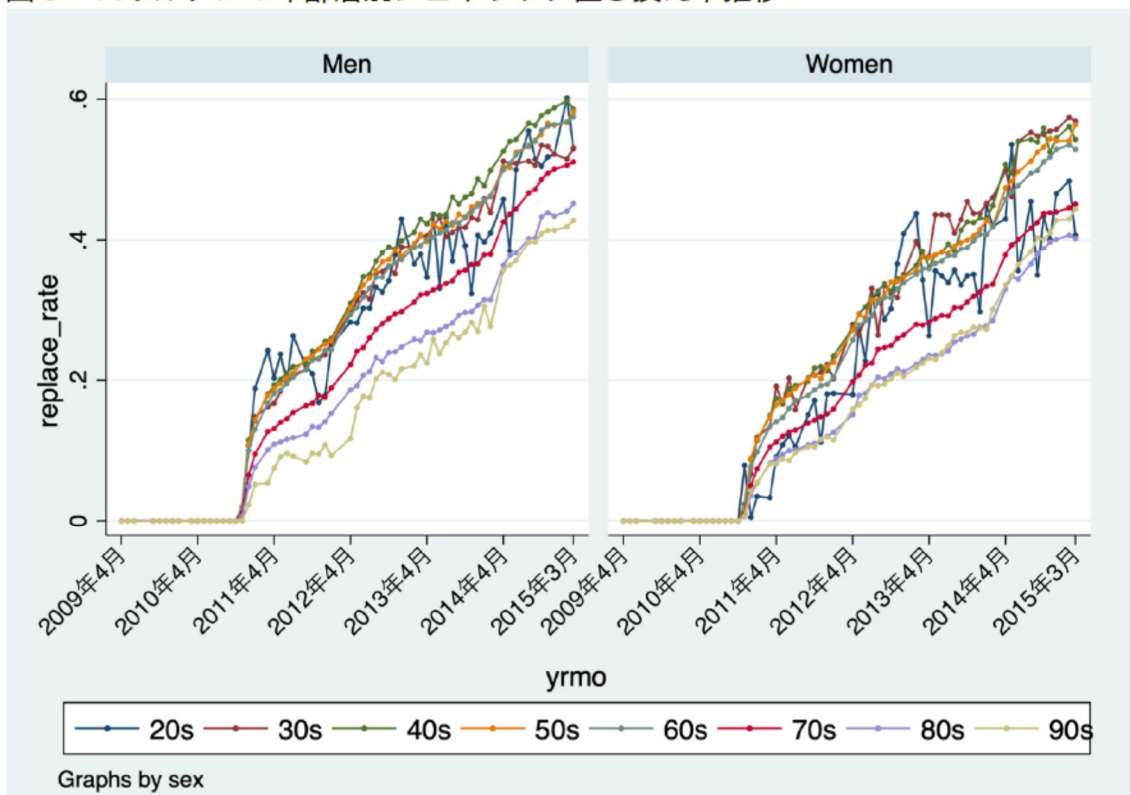


図7 ピオグリタゾンの年齢層別ジェネリック置き換え率推移

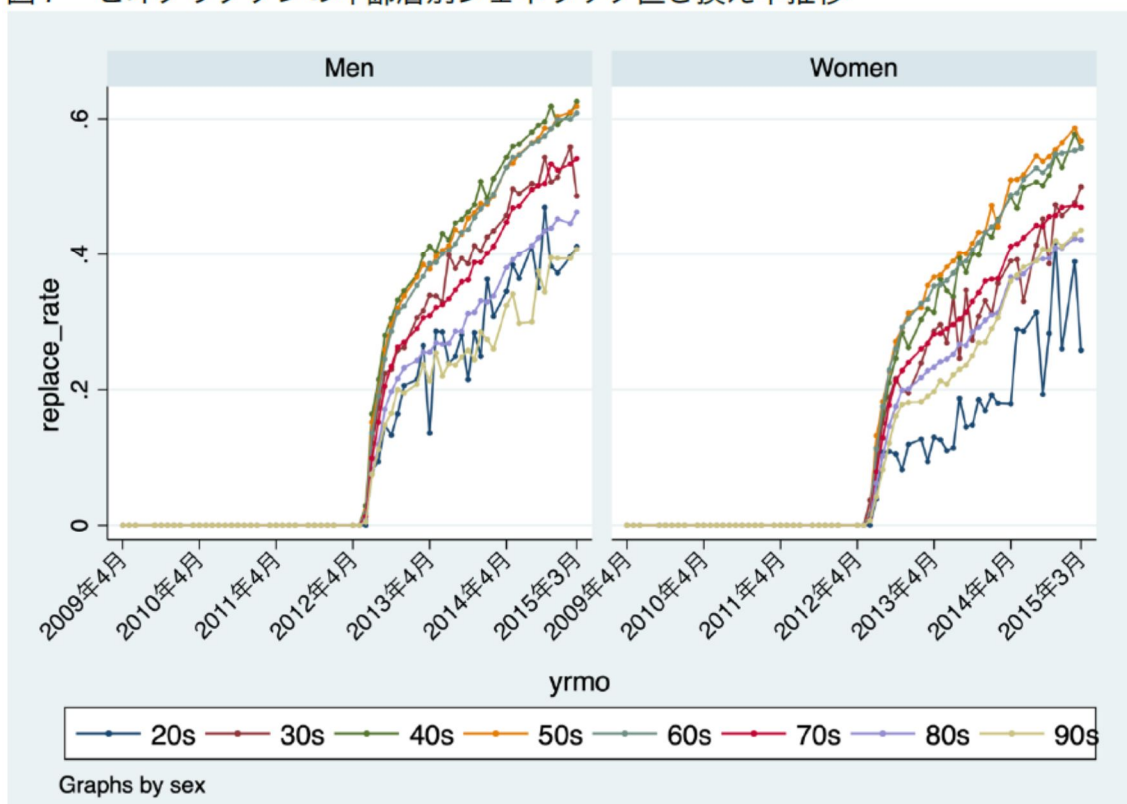


図8 ピタバスタチンのジェネリック置き換え率

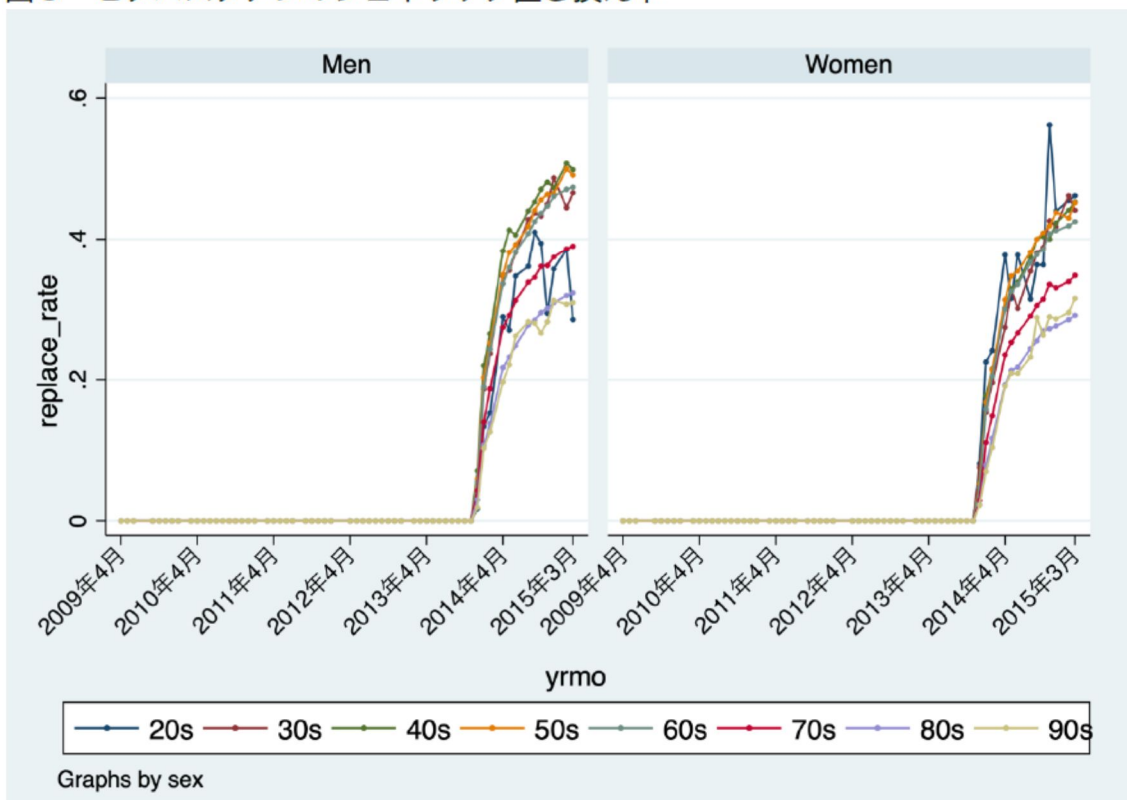


図9 シムバスタチンの年齢層別ジェネリック置き換え率推移



図10 ボグリボースの年齢層別ジェネリック置き換え率推移

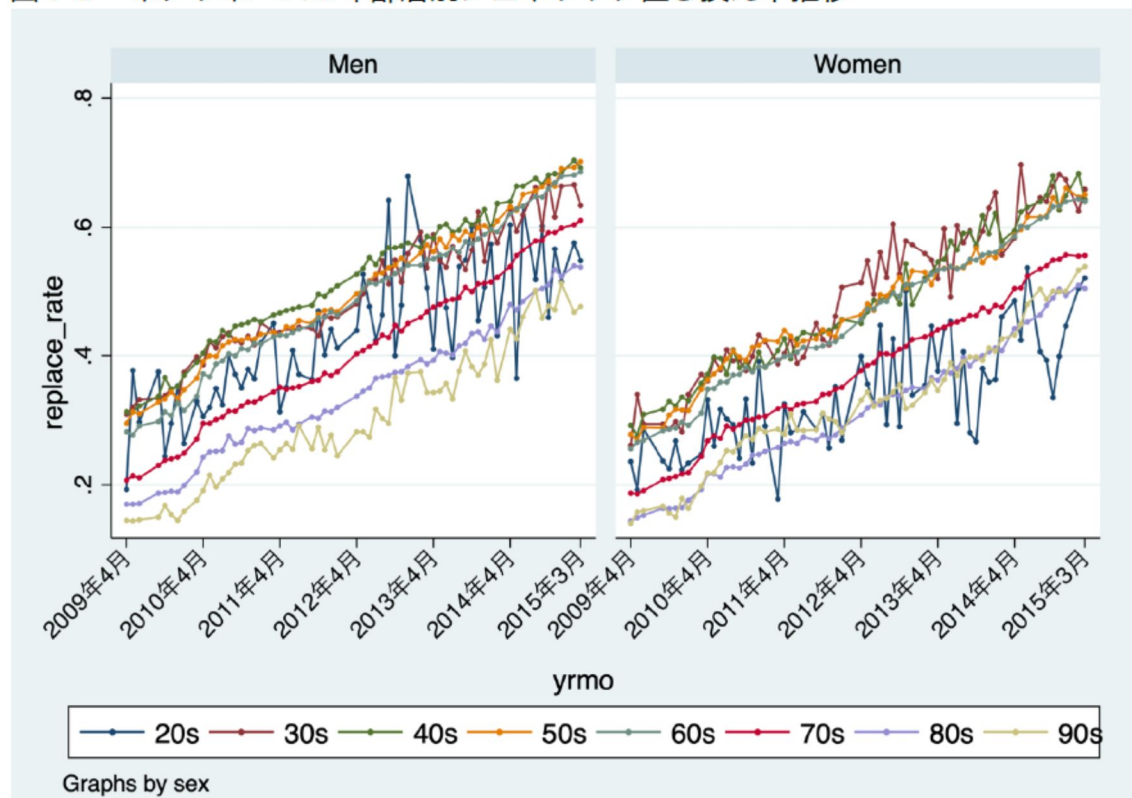


図 1 1 アムロジピンの自己負担割合別ジェネリック置き換え率推移

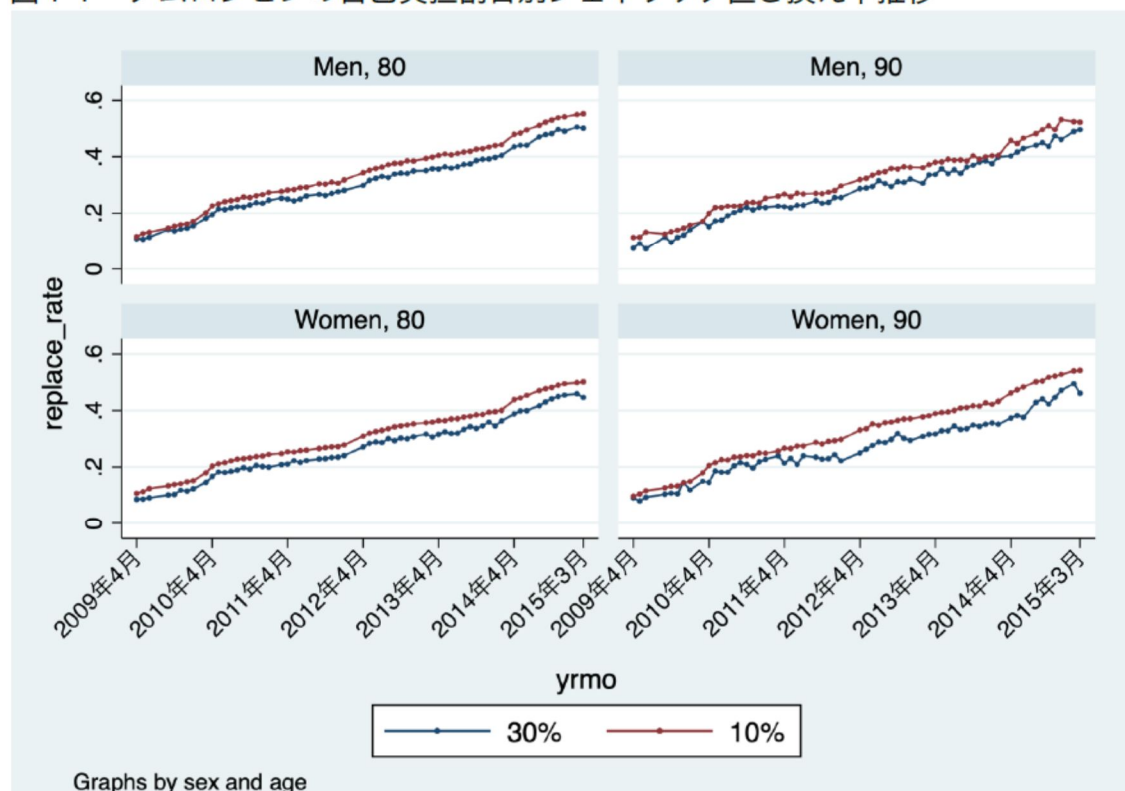


図 1 2 アトルバスタチンの自己負担割合別ジェネリック置き換え率推移

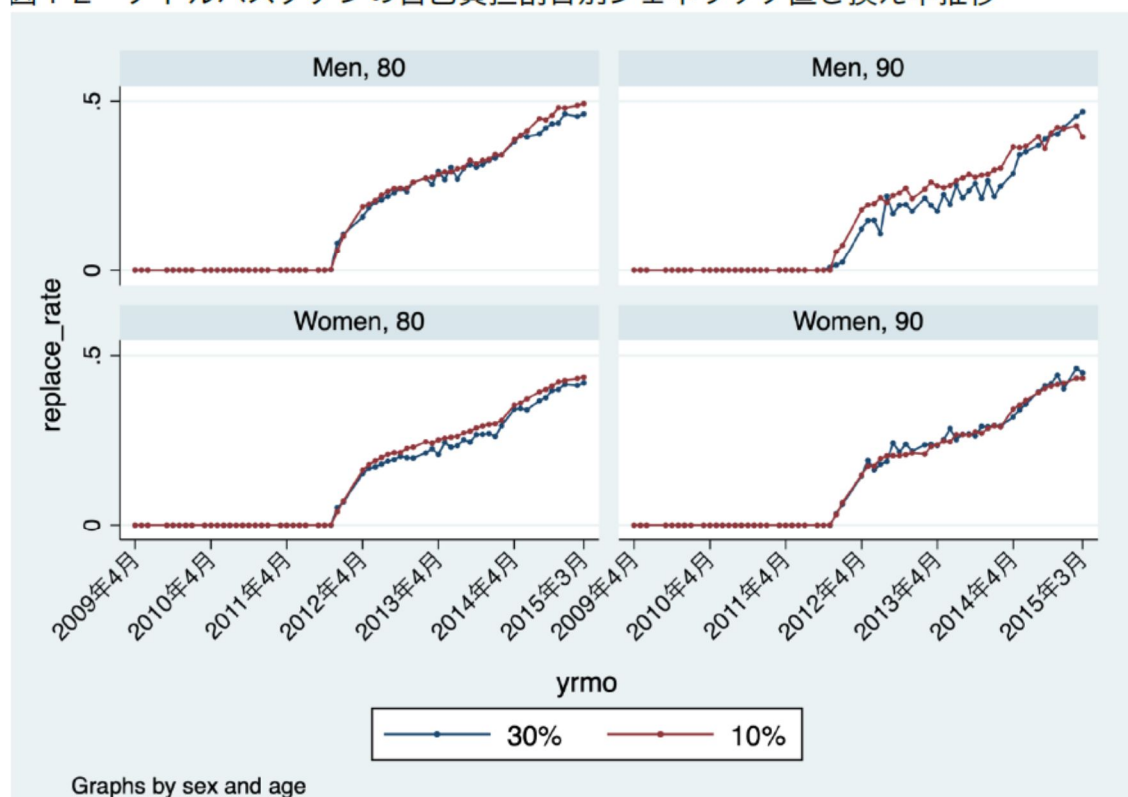


図13 カンデサルタンの自己負担割合別ジェネリック置き換え率推移

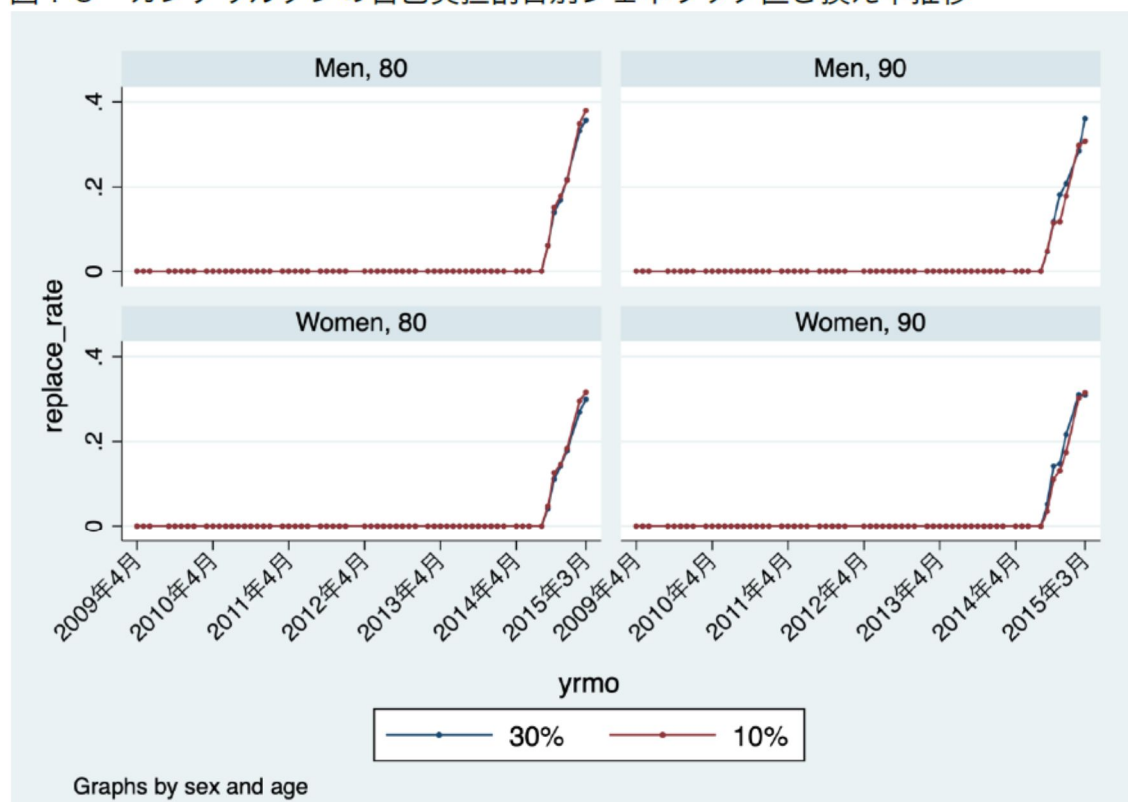


図14 ドネペジルの自己負担割合別ジェネリック置き換え率推移

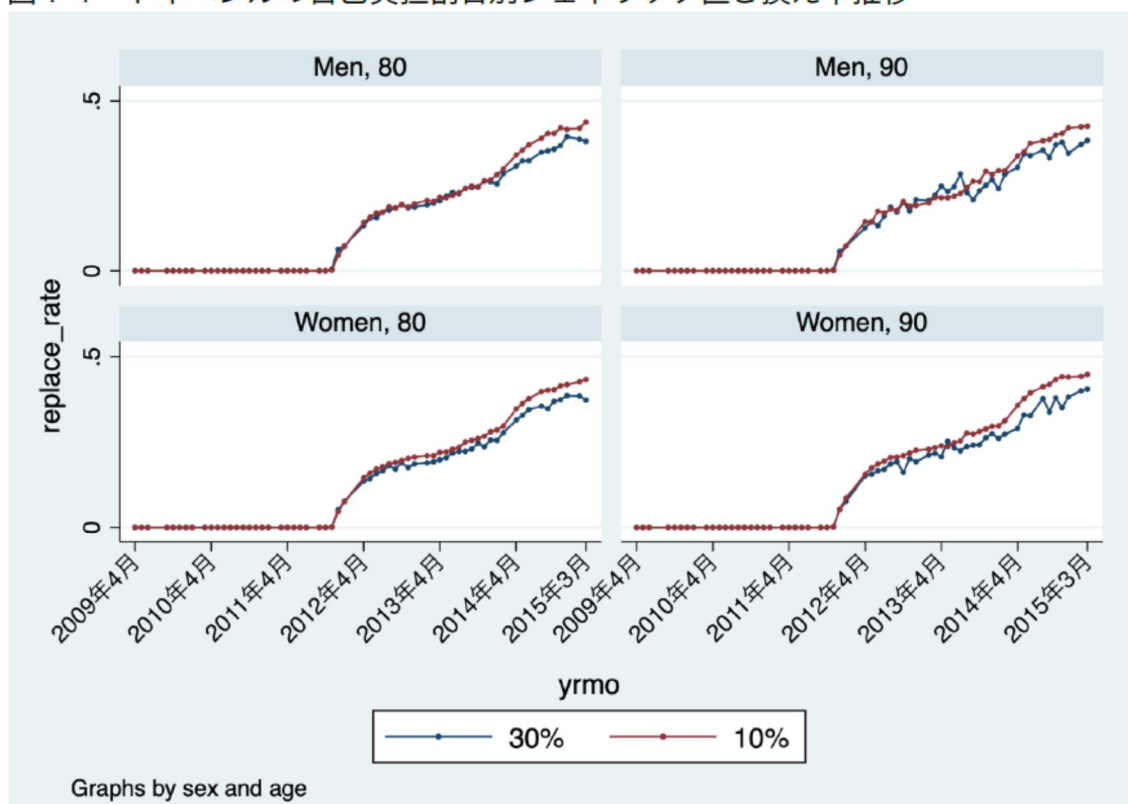


図15 グリメピリドの自己負担割合別ジェネリック置き換え率推移



図16 ロサルタンの自己負担割合別ジェネリック置き換え率推移



図17 ピオグリタゾンの自己負担割合別ジェネリック置き換え率推移



図18 ピタバスタチンの自己負担割合別ジェネリック置き換え率推移

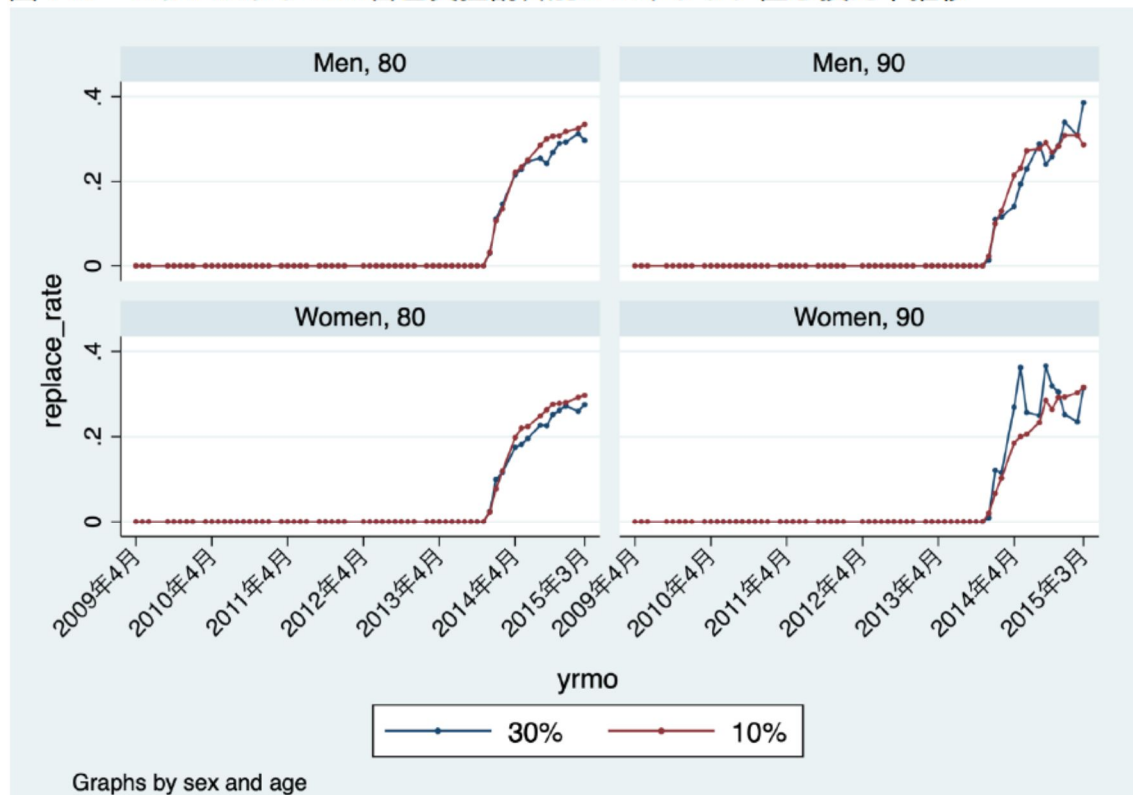
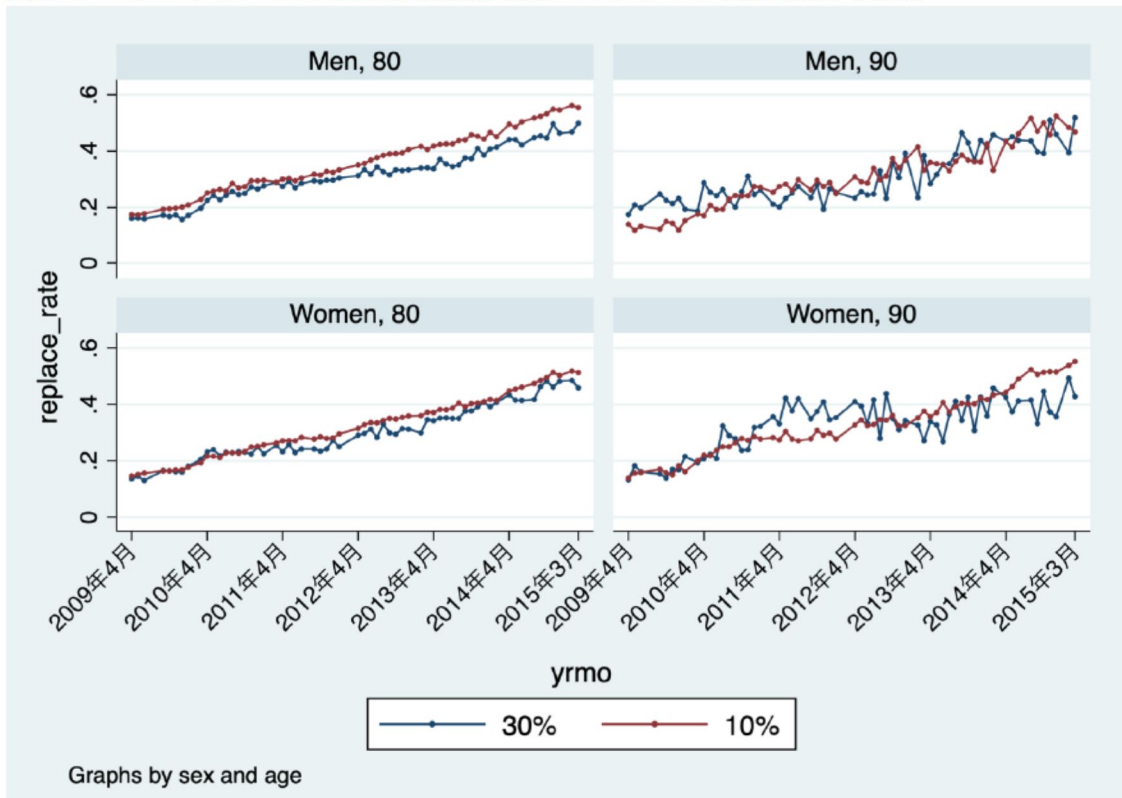


図19 シムバスタチンの自己負担割合別ジェネリック置き換え率推移



図20 ボグリボースの自己負担割合別ジェネリック置き換え率推移



平成 30 年度厚生労働科学研究費補助金（政策科学総合研究事業
（臨床研究等 ICT 基盤構築・人工知能実装研究事業））
分担研究報告書

ビッグデータの統計解析に関する人材育成

研究分担者 松山裕 東京大学大学院医学系研究科公共健康医学専攻教授

研究要旨 医療ビッグデータの統計解析に関わる人材育成プログラムに盛り込むべき統計解析手法を検討するため、以下のふたつの研究を行った。

代表的な医療ビッグデータのひとつである電子カルテデータベースを用いて、服薬状況を考慮した医薬品の安全性の評価が可能か検討した。大学病院の電子カルテデータベースを用いて、抗菌薬による肝障害発症リスクを評価した結果、ペニシリン系抗菌薬では、ベースライン以降の服薬状況を無視した intention-to-treat (ITT) 効果よりベースライン以降も継続して服薬したとする per-protocol (PP) 効果の方がリスクが大きく推定された。電子カルテデータベースのような医療ビッグデータを用いて医薬品の安全性を評価する際には、ITT 効果と PP 効果を区別して研究を実施し、結果を解釈すべきである。

医療ビッグデータ解析では、多くの交絡変数が想定される一方イベント数は限られるというデータにしばしば遭遇する。そのようなデータからリスク差で表される治療効果を推定する手法として、modified least-squares 回帰の統計的性能をシミュレーション実験で評価した。シミュレーション実験の結果では、modified least-squares 回帰は 1 交絡変数当たりのイベント数によらずバイアスのないリスク差の推定値を与え、曝露群と非曝露群ともに 5 イベント以上が期待されるときには 95%信頼区間の被覆確率も名目水準前後以上に保たれていた。Modified least-squares 回帰は交絡変数の数に比してイベント数が少ないデータからリスク差を推定する手法として有用である。

研究協力者

萩原康博 東京大学大学院医学系研究科 生物統計学 大学院生

するときに有用と考えられる modified least-squares 回帰の統計的性能をシミュレーション実験により評価した。

A. 研究目的

本分担研究の目的は、医療ビッグデータの統計解析に有用な手法を検討し、検討結果を医療ビッグデータ研究の人材を育成するプログラムに盛り込むことである。本年度は、(1) 電子カルテ情報中の服薬状況を考慮して医薬品の安全性を評価する統計解析手法、(2) 医療ビッグデータからリスク差を推定

B. 研究方法

1. 電子カルテデータベースを用いた服薬状況を考慮した医薬品の安全性評価

薬剤の安全性を観察研究で評価するとき、観察期間中の服薬状況を考慮せずに初期の服薬情報のみを用いる場合がある。このとき推定を試みているのは、変薬や中断も含めた実際に観測された服薬行動のもとでのリス

クであり、臨床試験での用語を借りて intent-to-treat (ITT) 効果と呼ばれる。一方、観察期間中の服薬状況を考慮してリスクを評価する場合もある。このとき、観察期間中に途切れることなく服薬した場合のリスクを評価する場合の効果は per-protocol (PP) 効果と呼ぶことができる。これら ITT および PP 効果を推定する手法は提案されているものの、薬剤疫学研究での適用事例は限られている。本研究では、電子カルテ情報を用いて、非曝露を対照とした抗菌薬による肝障害発症リスク評価を ITT および PP 効果の観点から実施した。

データソースとして、東京大学医学部附属病院の電子カルテデータ (2011~2015 年、約 2 万 4 千人) を使用した。曝露は抗菌薬 (マクロライド系、ペニシリン系、フルオロキノロン系) クラス単位の処方とし、それぞれのクラスごとに肝障害発症リスクを評価した。イベントである肝障害は肝関連臨床検査の異常値検出をもって発症と定義した。各患者の受診情報に基づいて定義された追跡期間を、抗菌薬への曝露期間と非曝露期間に分割した。各曝露エピソードに対して、最大 10 件の非曝露エピソードを、性別、年齢、曝露開始時点によってマッチングしたあと、傾向スコアをロジスティック回帰で推定した。傾向スコアの推定時の説明変数として、性別、マッチング時点の年齢、マッチング時点、入院有無に加え、高次元傾向スコア (high-dimensional propensity score) の変数選択アルゴリズムにもとづき、すべての併用薬および既往歴情報から合計 100 変数を選択した。

ITT 効果の推定では、ベースライン時点での交絡因子を調整する必要がある。そこで、離散比例ハザードモデルを傾向スコアに基づく inverse probability of treatment weighting (IPTW) 法で推定することで、曝露開始か

ら 30 日間の ITT ハザード比を推定した。PP 効果の推定には、ベースライン時点の交絡因子に加えて、ベースライン後のランダムでない治療変更 (非曝露から曝露あるいはその逆) を調整する必要がある。そのため、傾向スコアおよび同様の変数を用いて推定した服薬を継続する確率を用いる inverse probability of treatment and censoring weighting (IPTCW) 法により離散比例ハザードモデルを推定することで PP ハザード比を推定した。データ構造に関する識別可能条件 (一致性、条件付き交換可能性、正の治療・服薬継続確率) が満たされるとき、これらの方法はターゲットハザード比の一致推定量を与える。これらの方法では、重みを用いることに加えて、同一患者が複数エピソードに寄与しうるため、95% 信頼区間はロバスト分散を用いて構成した。

2. イベントがまれなときの modified least-squares 回帰の統計的性能の評価

観察コホート研究データを用いてアウトカムに対する曝露の効果を適切に推測するためには、交絡因子を調整した統計解析が求められる。多数の交絡因子を同時に調整する手法として、疫学研究では回帰分析が広く知られ、二値アウトカムに対してはロジスティック回帰が頻用されてきた。しかし、ロジスティック回帰は、交絡変数の数に比してイベント数が少ないとき、理論的にも経験的にも、その統計的性能は不良であることが知られている。

疫学研究論文では、ロジスティック回帰で推定されたオッズ比が報告されることが多い一方、リスク差も重要な効果指標のひとつである。交絡を調整したリスク差を推定する場合にも、二値アウトカムに対する回帰分析のひとつである線型二項回帰を用いること

ができる。線型二項回帰の一種に modified least-squares 回帰があり、交絡変数の数に比べてイベント数が少ないときでも理論的に優れた統計的性質（不偏性、明示解の存在、層別線型二項モデルでの Mantel-Haenszel リスク差推定量への帰着）を持つことがわかっている。本研究では、modified least-squares 回帰を用いる場合に、イベントが 1 交絡変数当たりどの程度存在すれば妥当なリスク差推定が可能か、シミュレーション実験により検討した

シミュレーション実験を行ったシナリオは、交絡変数の数（2、5、10、20 個）、1 交絡変数当たりの期待イベント数（2 から 12 まで 2 ずつ）、周辺リスク（0.5、2、5、10、20、40%）、曝露割合（20、50%）、周辺リスクに応じた曝露のリスク差（小、中、大）を変化させた 864 シナリオである。交絡変数は、カテゴリカル変数は Bernoulli 分布（成功確率 0.2 から 0.8）、連続変数はベータ分布（一様、ベル型、裾引き型）から発生させた。曝露変数は交絡変数の主効果のみを含むロジスティックモデルから発生させた。アウトカム変数は、曝露と共変量の主効果のみを含む線型二項モデルから発生させた。交絡変数と曝露・アウトカムの関連の強さは軽度から中等度となるように設定した。シミュレーション回数は各シナリオ 3000 回とした。

線型二項モデルを用いた modified least-squares 回帰を発生データに適用した。推定リスク差の標準誤差は HC3 サンドウィッチ分散推定量で推定し、Wald 信頼区間により 95% 信頼区間を構成した。

統計的性能の評価指標として、相対バイアス、Monte Carlo 標準誤差に対する平均推定標準誤差の比、95% 信頼区間の被覆確率を算出した。ロジスティック回帰に関する先行研究にもとづき、-15% 以下と 15% 以上の相対

バイアス、0.9 以下と 1.1 以上の標準誤差比、93% 以下と 97% 以上の被覆確率を「問題のある統計的性能」と定義した。

（倫理面への配慮）

東京大学医学部附属病院の電子カルテデータベースを用いた研究を行う際には、東京大学医学部・医学系研究科倫理委員会の承認を得た（承認番号 10675-1）。

C. 研究結果

1. 電子カルテデータベースを用いた服薬状況を考慮した医薬品の安全性評価

マクロライド系のコホートは 8,099 曝露エピソードと 80,990 非曝露エピソードで構成され、このうち 6,392 曝露エピソード（79%）および 281 非曝露エピソード（0.4%）で 30 日以内の治療変更が発生した。マクロライド系の 30 日間の肝障害発症リスクについての ITT ハザード比は 1.22（95% 信頼区間 0.75–1.98）であり、PP ハザード比は 1.22（95% 信頼区間 0.67–2.22）であった。ペニシリン系のコホートは 14,040 曝露エピソードと 140,400 非曝露エピソードで構成され、11,687 曝露エピソード（83%）および 1,014 非曝露エピソード（0.7%）で治療変更が発生した。ペニシリン系の ITT ハザード比は 4.01（95% 信頼区間 3.16–5.08）であり、PP ハザード比は 7.25（95% 信頼区間 5.58–9.41）であった。フルオロキノロン系のコホートは 16,020 曝露エピソードと 160,199 非曝露エピソードで構成され、14,233 曝露エピソード（89%）および 1,091 非曝露エピソード（0.7%）で治療変更が発生した。フルオロキノロン系の ITT ハザード比は 1.60（95% 信頼区間 1.27–2.03）であり、PP ハザード比は 1.69（95% 信頼区間 1.23–2.30）であった。

2. イベントがまれなときの modified least-squares 回帰の統計的性能の評価

図 1 に相対バイアスのシミュレーション実験結果を示す。1 交絡変数当たりのイベント数が 2 であったとしても、modified least-squares 回帰によるリスク差の推定値にバイアスは認められなかった。一部のシナリオで問題のある相対バイアスが認められたが、シミュレーション誤差の範囲内であった。

図 2 に Monte Carlo 標準誤差に対する平均推定標準誤差の比のシミュレーション実験結果を示す。周辺リスクが高く (20% または 40%) 1 交絡変数当たりのイベント数が少ないとき、つまり対象者数が少ないとき、HC3 サンドウィッチ分散推定量は標準誤差を過大評価する傾向があった。2 交絡変数かつ周辺リスクが低いとき (0.5% から 10%) 1 交絡変数当たりのイベント数が 4 以下になると、HC3 サンドウィッチ分散推定量は標準誤差を過小評価する傾向があった。

図 3 に 95% 信頼区間の被覆確率のシミュレーション実験結果を示す。2 交絡変数のシナリオでは、曝露割合が 20% のとき、1 交絡変数当たりのイベント数が 10 以下になると被覆確率が名目水準を下回ることがあった。5 交絡変数のシナリオにおいては、曝露割合が 20% のとき、1 交絡変数当たりのイベント数が 4 以下になると被覆確率が名目水準を下回ることがあった。被覆確率の低下は周辺リスクやリスク差の大きさには依存していなかった。10 または 20 交絡変数のシナリオにおいては、1 交絡変数当たりのイベント数によらず被覆確率は名目水準前後に保たれていた。被覆確率のシミュレーション実験結果は、1 交絡変数当たりのイベント数ではなくイベント数自体を用いて以下のように要約可能である。すなわち、20 イベント以上が期待されるシミュレーション実験シナリ

オにおいては、modified least-squares 回帰によるリスク差の 95% 信頼区間の被覆確率は名目水準前後を達成していた。

D. 考察

マクロライド系と比較してペニシリン系とフルオロキノロン系で高い肝障害発症リスクが推定された。この傾向はこれまでの薬剤疫学研究で得られている知見と一致している。マクロライド系およびフルオロキノロン系では、服薬状況を考慮してもハザード比の推定値はほとんど変化しなかった。これら抗菌薬クラスについては、研究開始から 30 日間抗菌薬を使用し続けても、30 日以内で抗菌薬を中断しても肝障害リスクは大きく変わらない可能性がある。一方、ペニシリン系では ITT ハザード比と比較して PP ハザード比で大きなリスク上昇が推定されたことから、服薬を中断することで肝障害リスクが低下する可能性が示された。本研究では、いずれの抗菌薬クラスでも曝露エピソードの 8 割程度で 30 日以内に服薬状況が変化しており、服薬状況を考慮しない ITT ハザード比と服薬状況を考慮した PP ハザード比を混同して解釈すると、誤った結論に至りうる。

Modified least-squares 回帰のシミュレーション実験では、相対バイアスの結果は理論通りであった。しかし、先行研究で報告されているロジスティック回帰のバイアスと比較すると、イベント数が少ないときの modified least-squares 回帰の不偏性は際立っている。被覆確率に関するシミュレーション結果は、曝露割合が 20% であることを考慮すると、曝露群と非曝露群でそれぞれ 5 イベント以上期待されれば被覆確率が名目水準前後に保たれると言いかえることができる。これは、よく知られた Mantel-Fleiss 基準に近い。

交絡変数の数に比してイベント数が少な

いデータに対する統計解析手法として、傾向スコア解析を推奨する研究者がいる。傾向スコア解析では、アウトカムではなく曝露をモデリングすることで、イベント数の少なさがもたらすアウトカムモデルの当てはめに伴う困難を回避できる。モデルの誤特定への感度も含めて、リスク差推定における傾向スコア解析と modified least-squares 回帰の統計的性能を比較する研究を行うと、手法の使い分けに関して示唆が得られるかもしれない。

E. 結論

医療ビッグデータのひとつである電子カルテデータベースを用いて、医薬品の安全性としての ITT 効果と PP 効果を評価することができた。医療ビッグデータを用いる際にも、ITT 効果と PP 効果を区別して統計解析を実施し、結果を解釈することが重要である。

Modified least-squares 回帰はイベント数に比して交絡変数が多いデータからリスク差を推定する手法として有用と考えられる。

本研究で扱った統計解析手法は、医療ビッグデータ解析を行う人材育成プログラムに含めるべき統計解析手法の候補となりうる。

F. 健康危険情報

分担研究報告書であるため、該当しない。

G. 研究発表

1. 論文発表

1. Takeuchi Y, Shinozaki T, Kumamaru H, Hiramatsu T, Matsuyama Y. Analyzing intent-to-treat and per-protocol effects on safety outcomes using a medical information database: An application to the risk assessment of antibiotic-induced liver injury. *Expert Opin Drug Saf.* 2018;17:1071–1079.

2. Hagiwara Y, Fukuda M, Matsuyama Y. The number of events per confounder for valid estimation of risk difference using modified least-squares regression. *Am J Epidemiol.* 2018;187:2481–2490.

2. 学会発表

1. Hagiwara Y, Shinozaki T, Matsuyama Y. G-estimation of structural nested restricted mean time lost models to estimate effect of time-varying treatment on survival outcome. 29th International Biometric Conference. Barcelona, Spain. July 2018.
2. 細井宏輝，竹内由則，柏原康佑，今井博久，松山裕．国民健康保険加入者における特定保健指導受診の継続的な効果の検証．第 29 回日本疫学会学術総会．東京都．2019 年 1–2 月．

H. 知的財産権の出願・登録状況（予定を含む）

1. 特許取得
なし。
2. 実用新案登録
なし。
3. その他
なし。

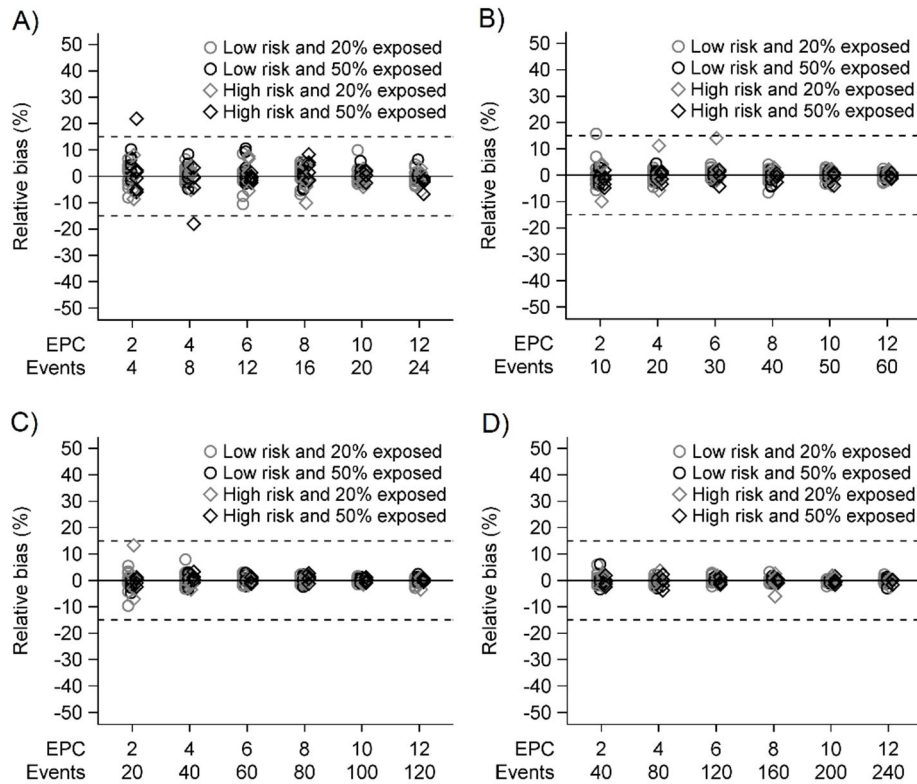


図1 相対バイアスに関するシミュレーション実験結果

A) 2 交絡変数、B) 5 交絡変数、C) 10 交絡変数、D) 20 交絡変数。low risk は 0.5% から 10% までの周辺リスク、high risk は 20% と 40% の周辺リスクを表す。EPC; events per confounder。

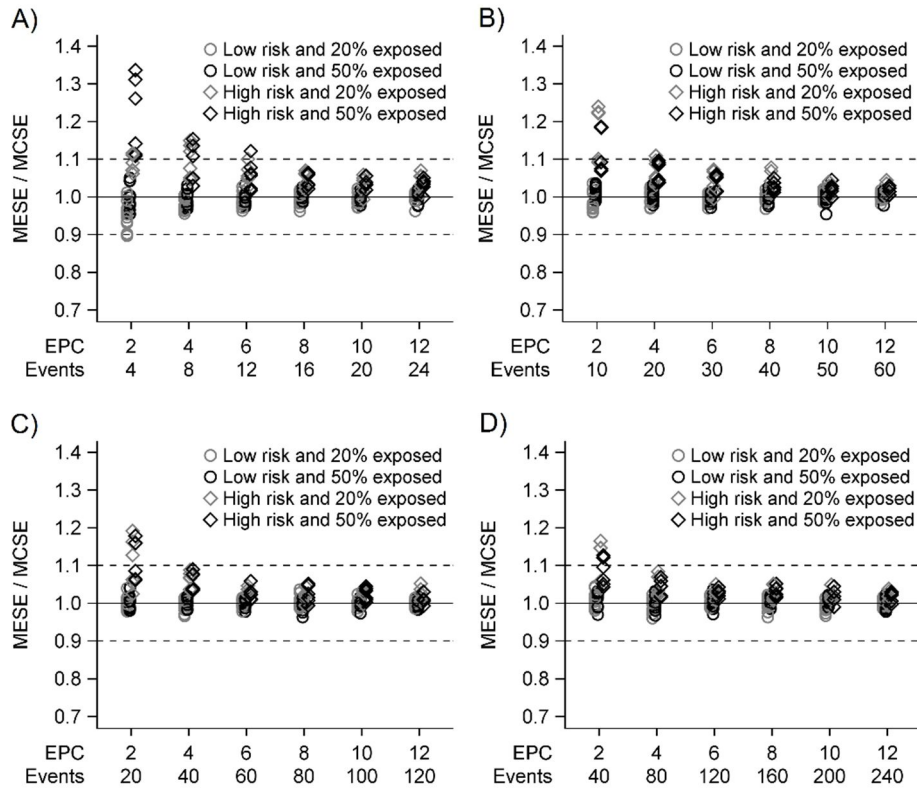


図2 標準誤差比に関するシミュレーション実験結果

A) 2 交絡変数、B) 5 交絡変数、C) 10 交絡変数、D) 20 交絡変数。low risk と high risk は図 1 と同様。EPC; events per confounder, MESE; mean estimated standard error, MCSE; Monte Carlo standard error。

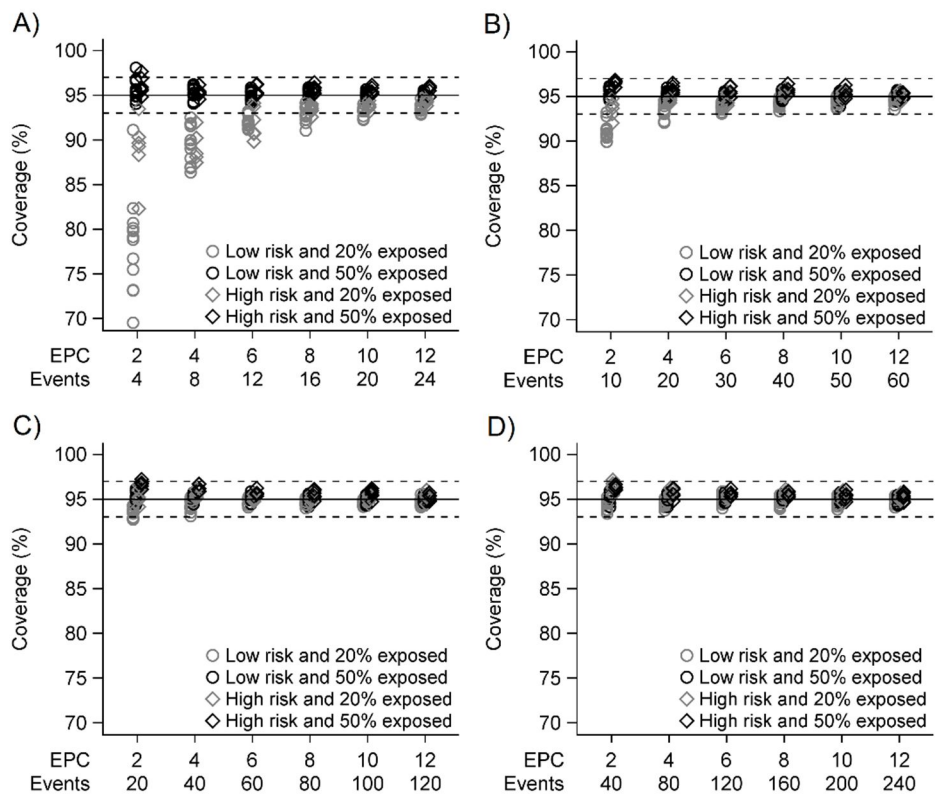


図3 95%信頼区間の被覆確率に関するシミュレーション実験結果

A) 2 交絡変数、B) 5 交絡変数、C) 10 交絡変数、D) 20 交絡変数。low risk と high risk は図 1 と同様。EPC; events per confounder。

平成 30 年度厚生労働科学研究費補助金
(政策科学総合研究事業(臨床研究等 ICT 基盤構築・人工知能実装研究事業))
分担研究報告書

I 介護保険レセプトデータ活用による研究の現状
II 情報工学と医学の協同によるビッグデータ解析 その可能性と課題
III 介護に関するデータベース研究の後進の育成

研究分担者 田宮菜奈子 筑波大学 医学医療系 ヘルスサービスリサーチ分野 教授
研究協力者 佐久間 淳 筑波大学 システム情報系 情報工学域 教授
研究協力者 Tran Quang Thien 筑波大学 システム情報工学研究科 コンピュータサイエンス
専攻 修士課程学生
研究協力者 Felipe Sandoval 筑波大学 医学医療系 ヘルスサービスリサーチ分野 研究員
研究協力者 金雪瑩 筑波大学 医学医療系 ヘルスサービスリサーチ分野 研究員
研究協力者 宇田和晃 東京大学大学院 公共健康医学専攻 臨床疫学・経済学 博士課程学生
研究協力者 佐々木健佑 筑波大学 システム情報工学研究科 社会工学専攻 修士課程学生

研究要旨

本研究分担班では、昨年度に引き続き介護保険レセプトデータ活用の現状の把握と情報工学との協同によるビッグデータ解析を推し進め、さらに本年度は後進の育成にも取り組んだ。まず、昨年度に開始していた介護保険レセプト(もしくは医療・介護突合レセプト)データ活用の現状についての文献レビューを継続し、英文の文献(計33文献)だけでなく和文の文献(計37文献)も含めたレビューを完遂した。その結果、米国等の海外と比較して、日本は介護においては全国レベルかつ細かいサービスについての情報収集が可能であり、これらをより有効に活用していく必要性が示唆された。次に、介護レセプトをもとに機械学習を用いて介護度の推移を予測する分析を昨年から継続し完遂した。結果として、介護レセプトデータだけでは介護度悪化の予測の能力は限られていることが示され、医療データとの連結等の方法により情報の幅を増やして介護についての予測を行う必要性が示唆された。最後に、本年度はNDB・DPCデータベース研究人材育成短期集中セミナーでの講演や、医学部生のビッグデータ解析の学会発表や論文化の指導を通じて、臨床疫学・統計・医療情報技術を持った後進の育成を進め、早い時期に医療介護データベースを用いた研究を指導するノウハウや意義を蓄積することができた。

I 介護保険レセプトデータ活用による研究の
現状：文献レビュー

A. 研究目的(背景)

世界で最初の国民皆保険による介護保険である我が国の介護保険のレセプトデータ

は、そのカバーする人数、サービスの種類・量の広さからみても世界唯一の介護に関するデータベースである。さらに、この介護保険レセプトは、介護給付費等実態統計(旧：介護給付費等実態調査)として国の統計に位置づけられてきたため、第三者(大学や研究機関)も、統計情報部を通じた統計法第33条

による二次利用が可能である。我々は、早くからこのデータを活用し研究に取り組んできた。しかし、2018年8月より、介護保険認定情報と突合したデータセット（介護DB）が、この統計法による利用とは別の枠組みとしての研究者への提供が開始された。これは、医療レセプトの研究者提供（NDB）と類似の枠組みであり、研究の意義や方法など、検討会等で審査された後に研究者に提供される。今後は介護保険レセプトのデータはこちらのルートに一本化される予定とのことである。このように、この1年で、我が国の全国介護レセプトの活用に対するしくみが大きく前進した。

一方で、介護に関するデータのみでは病名や医療の状況が不明であり、高齢者の実態とサービスの供給について総合的に理解するには不足がある。そこで、我々は、介護保険者である市町村と契約をすることによって、市町村内において介護レセプトと医療レセプトの突合されたデータを匿名加工していただき、共有を受けて分析を行ってきた。このような取り組みも、近年、全国の市町村で広がってきている。

さらに、上記の介護保険レセプトと医療レセプトの突合データ整備を全国レベルで国が実施し、公開する計画がデータヘルス計画に基づき、2020年目標に急ピッチで進められている。こうした中、公開後にすみやかに意義ある研究を推進するためには、現段階までの知見をある程度まとめておくことが有用であると考え、昨年度から介護保険レセプト（もしくは医療・介護突合レセプト）データ活用の現状についての文献レビューを実施した。昨年はまずは英文の原著論文にとどめたが、本年は、英文のアップデートに加え、和文論文も対象として実施した。

さらに、我々が研究を実施するにあたり参考にしてきたドイツ、米国などの介護関係のレセプトデータ研究の知見をもとに、それら

と我が国の介護レセプトデータ研究による違いも含めて、考察を行った。

B. 研究方法

医療そして/もしくは介護保険制度のレセプトデータの利用に関する英文原著論文をPubMedで、日文原著論文を医中誌ウェブで探し、系統的なレビューを行った。2018年8月までの出版論文を検索した。

選択基準として、原著論文に焦点を当て、PubMedで英語での要約、そして医中誌ウェブで日本語での要約のあるものにした。基本的に解説、編集ノート、政策のみの記述があるものは除外した。

検索用語としては、どの分野のインデックスからでも特定できるよう、“Japan” AND “long term care” AND “insurance OR claim OR administrative data”という3つのキーワードを組み合わせた。

データの選別は3つの段階に分けて行った。初めに、日本の介護保険を扱わない論文は除外した。次に、要約を読み内容が介護保険や介護保険レセプトに関するものであるかを確認した。最後に、内容をさらに明らかにする必要性のあるものについては、全文を読み評価した。図1に適格な英文原著論文の選定の流れについて、図2に日文原著論文の特定、選別、選出（著者順）の流れを示した。

最終的に選定された論文の中から、研究題目、著者、出版年、ジャーナル、介護保や険医療レセプトデータもしくはその両方を主要なデータソースとして使用しているか、アウトカムのタイプ、曝露要因のタイプ、限界が著者に言及されているかどうかの情報を抽出し、エクセルに表としてまとめた。

図 1. PubMed でレビューをした英文論文の特定、選別、適格性(2018 年 8 月時点)

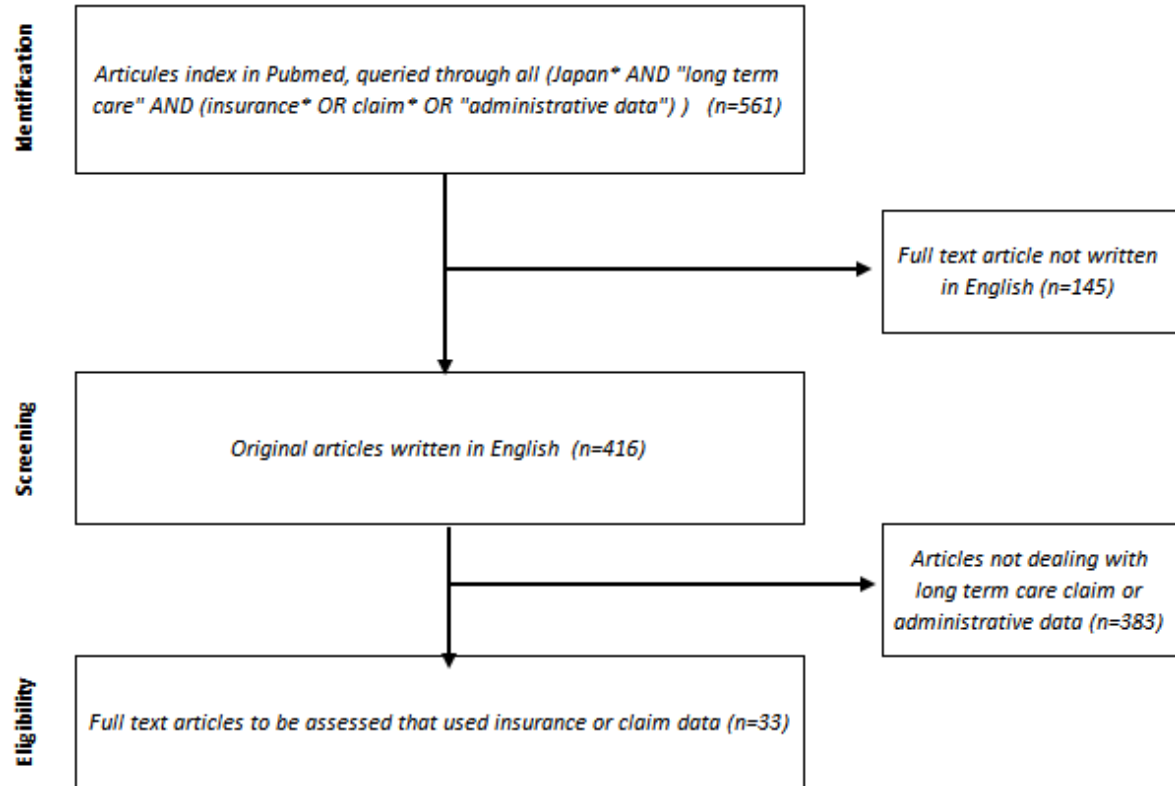
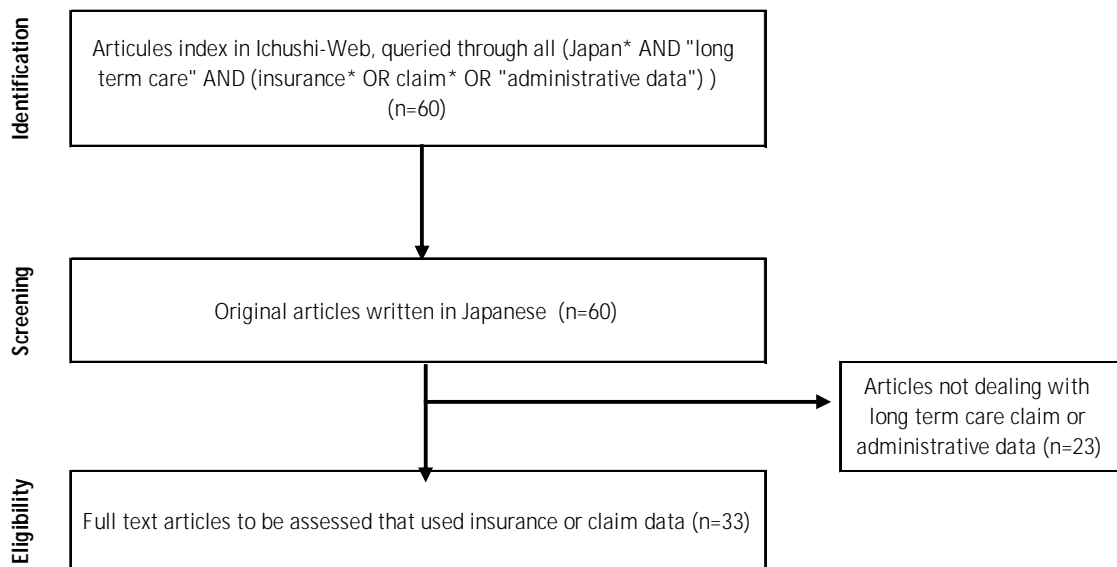


図 2. 医中誌ウェブのレビューをした和文論文の特定、選別、適格性(2018 年 8 月時点)



C. 研究結果

PubMed での検索では 561 件の検索結果が得られた。次に 145 件あった英語で書かれていない論文は除外した。そして、研究題目と要約の選別により 383 件の論文が除外された。この時点で、33 件の全文ある論文が調査対象となった。文献の出版年に制限は設けていないが、最も古い文献は 2006 年のものだった。

PubMed での介護保険や医療保険のデータ使用に関し、24 件の論文が介護保険のデータのみ、4 件の論文が医療保険のみを扱っていた。一方、5 件の研究が医療保険と介護保険のデータセットの両方を使用していた。この中で、9 件が過去 2 年に出版されている。表 1 に詳細を示した。最も多かった結果はサービス利用(13 件)、ケアのレベル(11 件)、健康アウトカム(5 件)、支出(3 件)そしてケアの場所(2 件)であった。もっとも多かった曝露要因のタイプはサービスの提供(21 件)、社会人

口統計学(14 件)、健康状況(13 件)、ケアのレベル(5 件)そして災害(3 件)であった。表 2 に詳細を示した。

医中誌ウェブでの検索では、60 件の検索結果が得られた。そして、23 件が研究題目と要約の選別により除外された。この時点で、37 件の全文ある論文が調査対象となった。今回は、原著論文かどうか明示されていないが、介護保険の行政データの重要な情報を提供している一連の特集論文(松田晋也氏による)があり、これも含めることにした。24 件の論文が介護保険データのみ、14 件の論文が医療保険のみを扱っていた。一方で 5 件の研究が医療保険、介護保険のデータセットの両方を扱っていた。この中で、14 件が過去二年に出版されている。表 3 に詳細を示した。

表 1. PubMed での文献レビュー (n = 33) の結果 1 (出版年、用いたデータ)

Study title		Year	Type of claim data used		
Query: all fields (Japan* AND "long term care" AND (insurance* OR claim* OR "administrative data"))			Long term care	Medical	Both
Study 1	Estimated expenditures for hip fractures using merged healthcare insurance data for individuals aged ≥ 75 years and long-term care insurance claims data in Japan.	2018	○	○	★
Study 2	Factors Affecting Discharge to Home of Geriatric Intermediate Care Facility Residents in Japan.	2018	○		
Study 3	Resident and facility characteristics associated with care-need level deterioration in long-term care welfare facilities in Japan.	2018	○		
Study 4	Survival analysis of increases in care needs associated with dementia and living alone among older long-term care service users in Japan.	2017	○		
Study 5	Measurement of special access to home visit nursing services among Japanese disabled elderly people: using GIS and claim data.	2017	○	○	★
Study 6	Comparative economic evaluation of home-based and hospital-based palliative care for terminal cancer patients	2017		○	
Study 7	Enhanced Oral Care and Health Outcomes Among Nursing Facility Residents: Analysis Using the National Long-Term Care Database in Japan.	2017	○		
Study 8	A Retrospective Claims Database Study on Drug Utilization in Japanese Patients with Crohn's Disease Treated with Adalimumab or Infliximab.	2016		○	
Study 9	Patterns of long-term care services use in a suburban municipality of Japan: a population-based study	2017	○		
Study 10	Cumulative number of hospital bed days among older adults in the last year of life: A retrospective cohort study.	2017		○	
Study 11	Long-term impact of the 2011 Great East Japan Earthquake and tsunami on functional disability among older people: A 3-year longitudinal comparison of disability prevalence among Japanese municipalities	2015	○		
Study 12	Diseases that precede disability among latter-stage elderly individuals in Japan.	2015	○	○	★
Study 13	Dialysis patients' utilization of health care services covered by long-term care insurance in Japan.	2015	○		
Study 14	The effects of dementia and long-term care services on the deterioration of care-needs levels of the elderly in Japan.	2015	○		
Study 15	Association between the disability prevention program "Secondary Preventive Services" and disability incidence among the elderly population: A nationwide longitudinal comparison of Japanese municipalities.	2016	○		
Study 16	Impact of the 2011 Great East Japan Earthquake on community health: ecological time series on transient increase in indirect mortality and recovery of health and long-term-care system	2014	○	○	★
Study 17	Impact of the 2011 Great East Japan Earthquake and Tsunami on functional disability among older people: a longitudinal comparison of disability prevalence among Japanese municipalities.	2014	○		
Study 18	Depression among family caregivers of community-dwelling older people who used services under the Long Term Care Insurance program: a large-scale population-based study in Japan.	2013	○		
Study 19	Long-term care-service use and increases in care-need level among home-based elderly people in a Japanese urban area.pdf	2013	○		
Study 20	Factors associated with the use of home-visit nursing services covered by the long-term care insurance in rural Japan: a cross-sectional study	2013	○		
Study 21	Factors that allow elderly individuals to stay at home with their families using the Japanese long-term care insurance system	2013	○		
Study 22	Effect of in-home and community-based services on the functional status of elderly in the long-term care insurance system in Japan	2012	○		
Study 23	Factors associated with the use of institutional long-term care in Japan	2012	○		
Study 24	Predictors of the highest long-term care expenditures in Japan	2011	○		
Study 25	Impact of home and community-based services on hospitalisation and institutionalisation among individuals eligible for long-term care insurance in Japan	2010	○	○	★
Study 26	Micro data analysis of medical and long-term care utilization among the elderly in Japan.	2010	○	○	★
Study 27	Should the provision of home help services be contained?: validation of the new preventive care policy in Japan	2010	○		
Study 28	Comparison of public and private care management agencies under public long-term care insurance in Japan: a cross-sectional study.	2010	○		
Study 29	Relationship between home care service use and changes in the care needs level of Japanese elderly	2009	○		
Study 30	In-depth descriptive analysis of trends in prevalence of long-term care in Japan	2008	○		
Study 31	Health-related quality of life and beneficiaries of long-term care insurance in Japan	2008	○		
Study 32	Use of home care services effectively reduces feelings of burden among family caregivers of disabled elderly in Japan: preliminary results	2006	○		
Study 33	The effect of a subsidy policy on the utilization of community care services under a public long-term care insurance program in rural Japan.	2006	○		

表 2. PubMed での文献レビュー (n = 33) の結果 2 (曝露要因、アウトカム)

	Year	Type of exposure	Type of outcome
Query: all fields (Japan* AND "long term care" AND (insurance* OR claim* OR "administrative data")			
Study 1	2018	health status	expenditure
Study 2	2018	sociodemographics	place of care
Study 3	2018	sociodemographics	care-level
Study 4	2017	health status and sociodemographics	care-level
Study 5	2017	service provision and sociodemographics	service use
Study 6	2017	service provision	expenditure
Study 7	2017	service provision	health outcome
Study 8	2016	service provision	service use
Study 9	2017	health status and sociodemographics	service use
Study 10	2017	health status service provision	service use
Study 11	2015	disaster	care-level
Study 12	2015	health status	care-level
Study 13	2015	service provision	service use
Study 14	2015	health status, service provision	care-level
Study 15	2016	service provision	care-level
Study 16	2014	disaster	health outcome and service use
Study 17	2014	disaster	care-level
Study 18	2013	service provision	health outcome
Study 19	2013	service provision	care-level
Study 20	2013	care-level, health status and sociodemographics	service use
Study 21	2013	care-level, health status and sociodemographics	service use
Study 22	2012	service provision	care-level
Study 23	2012	sociodemographics	service use
Study 24	2011	care-level, service provision, health status and sociodemographics	expenditure
Study 25	2010	service provision	place of care
Study 26	2010	sociodemographics, health status and service provision	service use
Study 27	2010	service provision	care-level
Study 28	2010	sociodemographics, health status and service provision	service use
Study 29	2009	service provision	care-level
Study 30	2008	sociodemographics and service provision	service use
Study 31	2008	care-level, service provision, health status and sociodemographics	health outcome
Study 32	2006	care-level, service provision, health status and sociodemographics	health outcome
Study 33	2006	service provision	service use

表3. 医中誌ウェブでの和文文献レビュー結果 (n = 37)

Study ID	Study title	Year	Type of claim data used		
			Long term care	Medical	Both
Query: all fields (Japan* AND "long term care" AND (insurance* OR claim* OR "administrative data"))					
Study 1	GISによる在宅医療受療率の地域差の要因分析	2018		○	
Study 2	医療・介護のビッグデータ分析 訪問診療の種類別にみた利用者の特性分析	2018	○	○	★
Study 3	医療・介護のビッグデータ分析 医療・介護レセプトデータを用いた胃腸造設患者の属性分析	2018	○	○	★
Study 4	福祉の現場から 在宅要介護者の要介護度とサービス給付実績及び主介護者の介護負担感の変化	2018		○	
Study 5	医療・介護のビッグデータ分析 誤嚥性肺炎患者の発症前の状況分析	2018	○	○	★
Study 6	医療・介護のビッグデータ分析 がん外来化学療法の実施曜日にに関する検討	2018		○	
Study 7	医療・介護のビッグデータ分析 薬学管理料算定状況からみた「かかりつけ薬局」の在り方に関する検討	2018		○	
Study 8	医療・介護のビッグデータ分析 薬剤師による居宅療養管理指導の現状分析	2018	○	○	★
Study 9	医療・介護のビッグデータ分析 訪問介護の頻度別にみた在宅要介護高齢者の特性分析	2018	○	○	★
Study 10	医療・介護のビッグデータ分析 訪問看護の種類別にみた利用者の特性分析	2018	○	○	★
Study 11	医療・介護のビッグデータ分析 要介護度別にみた傷病構造の分析	2018	○	○	★
Study 12	医療・介護のビッグデータ分析 医療・介護レセプト連結データを用いた脳梗塞患者の医療介護サービス利用状況の分析	2018	○	○	★
Study 13	地域包括ケアシステムの評価指標としての在宅期間 8年間の全国介護レセプトデータによる検討	2017		○	
Study 14	地域包括ケアシステムにおける医師と在宅薬剤師との連携 現状とこれからの課題	2017	○	○	★
Study 15	全国介護レセプトを用いた経口移行者実態把握の試み	2016	○		
Study 16	医療介護総合データベースを用いた大腿骨頸部骨折患者の医療介護サービス利用状況の分析	2016	○	○	★
Study 17	【介護予防・保健指導の効果はどのように評価されているか】 特定健診結果とレセプトデータを利用した腹囲と平均年間医療費の関係について	2016		○	
Study 18	要介護認定高齢者の施設入所に関連する要因について 医療と介護のレセプトデータを活用して	2015	○	○	★
Study 19	福岡県の認知症入院患者の在院日数に関する研究	2014		○	
Study 20	通所リハビリテーションの提供体制の整備が介護費に与える影響	2013	○		
Study 21	要介護認定データを用いた特別養護老人ホームにおけるケアの質評価の試み 11指標群の作成と施設間比較	2013	○		
Study 22	介護サービス資源の地域格差と要介護高齢者のサービス利用 介護保険レセプトデータに基づく実証分析	2013	○		
Study 23	保険料段階による在宅介護サービス費用の経時変化 特定自治体における2年間の介護給付実績情報より	2011	○		
Study 24	高齢者の要介護認定有無別医療費の比較分析	2011	○	○	★
Study 25	筋萎縮性側索硬化症患者における療養形態別の経済的自己負担	2011		○	
Study 26	【医療・介護政策に関する実証的検証】 新予防給付導入による介護サービス利用回数変化とアウトカム 検討会報告書と異なる分析手法に	2010	○	○	★
Study 27	【医療・介護政策に関する実証的検証】 医療資源の偏在が受診行動範囲、診療日数、医療費に与える影響について 国民健康保険レセプトデータに基づく実証的検証	2010		○	
Study 28	市町村における歯科保健事業とメタボリックシンドローム関係医療費との関連	2010		○	
Study 29	医療保険・介護保険レセプトと特定健診データの連結分析システムの開発	2010	○	○	★
Study 30	老人医療入院レセプトにおける医療費高騰要因 沖縄県の老人医療費が高い市と低い町の比較から	2009		○	
Study 31	住民健診受診者に対する訪問指導と医療費効果	2008		○	
Study 32	高齢者の大腿骨頸部骨折にかかる治療費の病院較差に関する研究	2006		○	
Study 33	ストレス、健康の自己評価と医療費の関係について	2006		○	
Study 34	高齢者本人による在宅福祉サービスの評価	2003		○	
Study 35	家族介護者による在宅福祉サービスの評価	2003		○	
Study 36	地域における初回脳卒中発症者に関する研究 保健師の家庭訪問に基づいた全数把握	2002		○	
Study 37	介護保険施行半年後の在宅サービス 利用率とそれに関する要因	2001	○		

D. 考察

本調査では、英文および和文で出版された我が国の介護に関するレセプトを活用した文献をレビューし、かつ海外の介護データを活用した同様な論文との比較から考察をした。

論文の扱っている内容については、アウトカムと曝露因子の両方でよく用いられた項目は「サービス内容」であった。サービス内容は予測因子（サービスの提供）やアウトカム（サービスの利用）として用いられている。これはヘルスサービスリサーチが進んでいる海外の他の行政のデータセットでは、早くから蓄積があったが、我が国の介護保険レセプトにおいては、とくにサービスの詳細が全国レベルで把握できる点が強みであり、近年蓄積されてきたことがわかる。一方、海外では、行政データが健康サービスの提供や利用の理解に不可欠だと考えられつつもこれらの情報が入手できる機会は限られている場合が多く、我が国の強みである。

サービス提供情報に加えて、日本の介護保険制度における全国一律の介護認定は、サービスへの適格性、すなわち認可された人のケアの必要性の程度に基づいている貴重なデータである。本レビューで分かるように、ケアのレベルはしばしば障害の状態の代替変数（proxy 変数）として使われている。海外の多くのデータセットでは障害や独立に関する統一された情報が欠如しているため、この情報は大変有用である。しかし、一方で、我が国の要介護度のレベルと欧米で確立されている ADL 尺度とを比較し妥当性を明らかにしていくことも重要であり、現在我々チームでも進めているところである。

また、医療保険と介護保険のデータを連結させることにより、研究の幅が広がっていることが示唆された。本レビューによれば、（日本

の介護保険制度の施行から 10 年後にあたる）2010 年頃から医療と介護の情報を同時に使用する試みが始まり、近年さらに増えている。海外でも異なる行政のデータセットの連結を行うことで、死亡率や、うつ病などの他の健康状況のような同様の情報の利用が盛んにおこなわれているが、前述のようにデータが全国レベルでは存在せず、商業的なデータベースに入っている場合も多く、こうしたデータは公的データとの連結が実現不可能である。

最後に、海外の介護データを用いた研究との比較からまとめると、一部記載したように、日本のデータは国家レベルで収集されていることが大きな違いであった。例えば米国は、行政データ利用による文献も多いが、データセットの内容は州ごとに異なっている。従って、研究も地域ごとに細分化されている。また、ドイツは、我が国に先立ち 1995 年から介護保険を実施しているが、我が国のように皆保険ではない。従ってデータが保険業者によって細分化されている。ドイツでは 118 社の異なる企業が法で定められた健康保険を 7200 万人のドイツ国民に提供しており、別の 50 社は 900 万人に民間の保険を提供している。また、アメリカ合衆国でも、データは払う側によっても細分化され、公的、私的、軍事的に分けられている。この点では、国家レベルでの統一的なレセプトデータの収集により、日本のデータは全国的、地域的な比較分析が行得る点で大変優れていると考えられる。

一方で、日本の研究の課題として、社会経済的な情報（教育や収入のような重要な情報）が現状考慮されていないことが挙げられる。特に、日本では貧困の指標を研究に使うことが普及しておらず、一部研究が推進されてきた貧困度（deprivation index）についての基礎研究の推進および、他のこうした社会的デ

ータ、さらには臨床データとの個人レベルでのリンクを可能にしていくことは、重要であろう。

E. 結論

介護保険レセプトデータ活用の現状についての文献レビューを行った。日本では国レベルの介護情報を集められていることが主な特徴として挙げられ、また医療保険と介護保険の両方のデータを合わせて研究に使用している例が近年多くみられるようになってきていることが明らかになった。また、サービス利用に関連した情報はしばしば曝露情報とアウトカムの定義の両方に使われ、介護保険を用いてヘルスサービスリサーチを実行するための重要な要素であることが示唆された。サービスへのアクセスを含む利用をどう捉えるか、またアウトカム評価におけるサービスの分析の仕方など、ヘルスサービスリサーチの方法論を必要とされるものが多いことがわかった。

また、公表された研究の数は近年になってますます増えており、介護分野の研究が急速に進んでいることがうかがえる。しかし、質の高い研究を行うためには、研究者の観点からレセプトデータの限界に関するより深い分析が必要であり、介護情報およびビッグデータ解析に精通した人材の育成が急務である。(なお、このレビューにあたって、研究室実習の一環として、筑波大学医学類生鈴木俊樹、照沼友里2名の協力を得た)

II 情報工学と医学の協同によるビッグデータ解析：介護保険請求データにおける Computational Phenotyping の試み

A. 研究目的

機械学習は、これまでの仮説検証のための疫学や統計手法に変わりうる方法として、近年ビッグデータの普及とともに発展してきた。

母集団を仮定におく従来の統計と異なり、豊富な変数の全数を基本にしたデータの分析には強みを発揮しうる。そこで、我々研究班では、早くから情報工学の専門家とコラボし、我々の整備した介護データを共有し、機械学習による様々なアプローチを推進してきた。

今年度は、まず、介護データが持つ、サービス加算等々の詳細かつ多様なデータをもとに、機械学習による類型化により、今後の分析をしやすくする方法を検討することとした。さらに、それらを活用した上で、要介護度を推定することを試みた。

本報告では、機械学習の手法を用い、介護データにおける Computational phenotyping (Phenotyping)問題に注目した。Phenotyping は、患者の状態を把握しやすくするために、患者の診断や治療の EMR データを、より解釈性のある表現に変換する方法を意味する。患者の状態がより把握できることで、治療行為に関する意思決定や、患者の管理などに役に立つ。本研究は、介護保険給付実績(介護レセプト)を用い、ユーザの状態をより解釈のできる表現に変換する phenotyping 問題に挑戦した。

さらに、これらの情報工学的アプローチにより明らかになった介護レセプトデータの特徴や今後の研究への課題についても議論することとした。

B. 研究方法

本研究は、にこの phenotyping 問題を将来予測問題に置き換えて考えた。具体的に注目する予測問題は要介護度の将来予測である。被保険者のある期間内の介護レセプトを用いて、数カ月後の要介護度を予測する問題に着眼した。このような将来予測が実現できれば、要介護度の悪化リスクで被保険者を分類したり、サービスの効果を明らかにしたりすることが可能になるからである。本研究はいくつの既存予測モデルを実装し、それらのモデ

ルの予測性能を評価し、予測問題によるアプローチの有効性を議論した。

a. サービス項目コードの利用に当たって

本研究が最初に扱った問題は、サービス項目コードの扱いである。サービス項目コードをそのまま利用するのは適切ではない。なぜなら、そのまま利用すると次元が非常に大きくなり、かつデータが非常にスパースになってしまうからである。このような問題は医療データにおける ICD コードの扱いと同様である。実際、ICD コードの数はおよそ 68000 個がある (ICD-10-CM の場合)。一般的に、ICD コードを学習に利用する際、何らかの方法で低次元な表現に変換するやり方が用いられる。また、ICD コードの場合、各コードの関係性が階層構造を持っていて、またコードからその階層構造が読み取れる。さらに、Charlson 併存疾患指数などエキスパートによる分類方法も存在する。一方、サービス項目コードの場合、コードの決め方には一律性が低い、またエキスパートによる既存の分類の方法も存在しない。(表 1)

コード 種類	コード サービス	サービス 内容略称
11	2733	身 6 生 2・II
11	2734	身 6 生 2・夜・II
11	8950	身 6 生 2・同・2 人・II
11	8951	身 6 生 2・同・2 人・夜・II
16	3936	通所リハ III251・延 6
16	3936	通所リハ III252・延 6

表 1: サービスコードの例

以上で説明した生のサービス項目コードのデメリットを踏まえて、本研究はサービス項目コードではなくサービス名を利用する別の表現方法を提案した。実際に、各サービスの内容略称には決まった構造で名付けられている。この情報を活用すれば、サービスコ

ードの数を多く減らせるだけでなく、サービス内容とサービスが行われる状況(サービス特徴)を区別して用いることができる。具体的に、「・」の文字で内容略称を分解し、一番目の要素をサービス内容とし、残りの要素をサービス特徴にした。このように、サービスコードよりも扱いが簡単なデータ表現が得られる。この用法で分解した結果、29608 個のサービス項目コードから 3633 種類のサービス内容と 59 種類のサービス特徴が得られた。

b. 予測モデル

この実験で用いられる予測モデルは主に 2 つに分けられる。1 つ目は一般的に時系列を考慮しないロジスティクス回帰とランダムフォレストの 2 つのモデルである。時系列を考慮しないこれらのモデルの場合、入力の特徴ベクトルは全ての時刻の特徴ベクトルの平均または最大値を用いて一つの特徴ベクトルを作成する。2 つ目は時系列を考慮する深層学習モデル Long short-term memory (LSTM) である。RNN とそのバリエーション、例えば GRU や LSTM は時系列データの分析に特化したモデルであり、自然言語処理や画像処理に多く用いられている。そして、EMR データのような医療データも時系列の関係を持つことから、近年ではこのような RNN モデルを医療データの予測に用いる研究が行われている。

c. 実験データと実験設定

本研究では、2006 年 4 月から 2014 年 4 月までの 8 年間における、日本全国の個人レベルの LTC 請求データを用いた。このデータは約 850 万人の被保険者のおよそ 170 千万レコードを含む。平均で、一人の被保険者が保有するデータ数は約 200 レコードである。実際に用いたデータは被保険者のデモグラ

フィックデータ(年齢・性別)サービスの利用履歴(種類・単位)である。このデータセットにある全ユーザから、少なくとも36ヶ月のデータがあり、そして連続する2つレコードの空き期間が6ヶ月未満であるという条件でユーザをフィルタリングして、66861人のユーザが実験対象となる。検証用データとテスト用データをそれぞれ2000ユーザを用い、残りのデータは学習に用いられる。また、本実験では18ヶ月の介護保険請求データを用い、現在の要介護度と6ヶ月後の要介護度を予測する問題に注目した。

	avg.Precision	avg.Recall	avg.F1-score
L.Regression	0.72	0.71	0.71
R.Forest	0.73	0.72	0.72
DeepCare	0.83	0.82	0.82

表2: 現在の要介護度の予測

	avg.Precision	avg.Recall	avg.F1-score
SameAsLast	0.82	0.81	0.81
L.Regression	0.61	0.60	0.60
R.Forest	0.63	0.61	0.62
DeepCare	0.68	0.66	0.66

表3: 6ヶ月後の要介護度の予測

C. 研究結果

実験の結果では、現在予測の場合、時系列を考慮する深層学習モデル LSTM が Logistic Regression と Random Forest より高い予測精度を達成できた。つまり、時系列の情報を利用できるモデルが要介護度の予測問題に有効であることが分かった。しかし、将来予測の場合、どのモデルもベースラインモデルより下回った結果が得られた。つまり、どの予測モデルも要介護度の変化を予測することが出来なかった。

D. 考察

本研究は、被保険者の健康状態が観測できない状況で行われた。しかし、実験結果でも示唆されたように、被保険者の健康状態の情報

なしでは有意義な Phenotyping や予測を得ることが極めて難しいと考えられる。なぜなら、要介護度の変化は、サービスの種類や質以外に、被保険者の持病や生活環境に強く影響されるからである。つまり、同じようなサービスを受けるとしても、被保険者の健康状態が異なることがある。しかし、本研究が用いているデータでは、このような違いを区別することができない。そのため、今後は医療データ(データが存在する地域がしばらくは限られるが)との併用が重要であることが確認できた。

また、予測問題に注目するのではなく、教師なし学習によるアプローチも検討する必要がある。

他に、研究を行う過程で、以下のような課題が挙げられた。

- ・本研究のデータの ground-truth の一つは要介護度である。しかし、この ground-truth ラベルを有効に利用することは容易ではない。具体的に、医療の場合では、適切な治療を施した場合、患者の状態がよくなることが期待できる。しかし、介護サービスの場合、適切なサービスを行ったとしても、年齢に伴う健康低下により、多くの被保険者の要介護度が悪化する。このように、観測できるデータと要介護度の関係性は明確ではない。
- ・平成29年4月で公開されたサービスコード表には合計23860個の異なるサービスコードが存在する。このようなサービスコードをそのまま利用するとデータが非常にまばら(sparse)になる。
- ・サービス請求が発生する際のみデータが出現するため、介護保険請求データは不規則に欠損する。また、欠損の理由は様々で、被保険者がサービスを必要としないか、単に入院中で介護保険ではなく医療保険に置き換えているか、見分けが難しい。
- ・各被保険者は異なるタイミングで介護サービス利用を始め、そして任意なタイミングに

において、死亡や入院などの様々な理由で介護保険から脱退する。そのように、各被保険者の異なるデータ数や期間を扱える手法が必要である。

・本データは非常に大規模なデータであり、一度にメモリに載せることができないため、バッチ学習を行うことができない。そのため、オンライン最適化が行えるアプローチが必要である。

E. 結論

本研究は機械学習の手法を用い、介護データにおける Computational phenotyping (Phenotyping) 問題に注目した。実験の結果から、現段階では予測精度の高い予測モデルを学習することが困難であることが分かった。今後は以上の課題を解決できるよう、より適切な予測モデル、データの性質にあった手法や問題設定への工夫を試みる必要がある。そのためにも、今回のように情報工学系の研究者と医学系の研究者が共同体制を組むことは有用であると考えられる。

III 介護に関するデータベース研究の後進の育成

本研究班の重要な目的の1つである「後進の育成」に関する活動の一環として、2018年8月のNDB・DPCデータベース研究人材育成短期集中セミナーにて講演を行った(参考資料1)。ここでは、介護レセプトを中心に、データの構造や意義、分析上の注意点、実際の研究例などを紹介した。この講演では、介護レセプト研究は医療ほど進んでいなかったこともあり、介護レセプトの可能性を感じた受講者からの問い合わせが複数あり、一部は今後の共同研究に発展する予定である。

また、2018年10月の第2回日本臨床疫学会学術総会のシンポジウムで講演を行う予

定となっていた(参考資料2)。ここでは、実際に医学生からの早期からの分析教育や情報系の学生指導の経験から、彼らにとって、医学系のデータへのアクセスが可能になったことの意義が大きいことなどを実体験とともにお伝えする予定であった。この講演は天候の影響により中止となったが、次回の2019年の第3回日本臨床疫学会学術総会で同内容の発表を行い後進の育成・教育に貢献する予定である

また、筑波大学医学医療系ヘルスサービスリサーチ分野に所属している医学部生および博士課程生、さらに他大学の大学院所属の学生も含め、保健医療介護現場の課題に即したビッグデータ解析に興味を持っている学生に対し、我々の整備した医療介護データベースをもとに、臨床疫学・統計・医療情報技術を教え、学会で発表し研鑽を積む機会を与えた。

ランセットに2010年に掲載された21世紀の医学教育(Frenk J, Chen L, et al. Health professionals for a new century: Lancet. 2010 Dec 4;376(9756):1923-58)においても、個別の臨床中心の視点からシステムの視点に医学教育が方向転換をするべきであると述べられており、医師国家試験の出題にも社会医学が増加している。ビッグデータ解析は、そうした社会システムとしての医療を考えるよいきっかけにもなると考える。学生としても、そのような視点を学べる機会を求めつつもこれまで得られなかったが、今回の実習をきっかけに取り組めたという感想もあった。

さらに、筑波大学では地域医療への貢献を担う、いわゆる地域枠の学生がいる。こうした学生も地域のニーズに目を向け、医療が何をすべきかを考えるきっかけにもなったようである。ビッグデータ研究は、実験研究等と異なり、僻地に勤務しながらも環境を整えることで研究が可能となる。むしろ僻地の医

療の課題を肌で感じながら、リサーチクエスチョンを組み立てることで意義ある研究が生み出される可能性も高い。今後しばらく増加していくこうした若手医師への研究指導と、それによる自己実現感、そして僻地医療へのやりがいや医師の定着にも貢献しうると考える。

具体的な学生の成果は、下記の通りである。

- ・一市町村の医療・介護連結レセプトを用いた尿路カテーテルについての研究を 2018 年度の第 2 回日本臨床疫学会学術総会にて発表した（参考資料 3）。現在論文投稿中である。
- ・国民生活基礎調査データを用いた高齢者の難聴と活動制限・不安・もの忘れの関連についての研究（参考資料 4）および日本の大規模入院データを用いた大腿骨頸部骨折手術後の早期リハビリが認知症高齢者の日常生活機能に与える影響を検討した研究（参考資料 5）を 2018 年度の Gerontological Society of America 2018 Annual Scientific Meeting に、大学院生や学部生を本経費で派遣して発表した。
- ・つくば市の高齢者に対するアンケート調査の結果を用いたポリファーマシーについての研究を 2018 年度の第 29 回日本疫学会学術総会にて発表した（参考資料 6）。現在論文

執筆中である。

以上の取り組みを通じて、介護に関するデータベース研究の後進の育成実績およびそのノウハウが蓄積されてきた。今後、これらの経験を生かし、それぞれの分野やレベルの違いに応じたビッグデータ教育の方法をまとめていくことが必要であると考えられる。

F．研究発表

1．論文発表

2 つの研究（参考資料 3 および 4）を論文化し投稿中、2 つの研究（参考資料 5 および 6）を論文化中

2．学会発表

4 回（参考資料 3-6）

G．知的財産権の出願・登録状況（予定を含む）

1．特許取得

なし

2．実用新案登録

なし

3．その他

なし

参考資料 1 : NDB・DPC データベース研究人材育成短期集中セミナーの講演スライド (計 65 枚中最初の 6 枚)

筑波大学
University of Tsukuba

ヘルスサービス開発研究センター
Health Services Research & Development Center

介護データベース研究

筑波大学医学医療系 ヘルスサービスリサーチ分野
筑波大学ヘルスサービス開発研究センター
教授・センター長 田宮 菜奈子

講義の内容

はじめに なぜ介護レセプトデータが必要か・・・

I. 全国介護レセプト

- I-1-a 介護給付費明細書 (介護レセ) について理解する
- I-1-b 全国介護レセに含まれる情報を理解する
- I-1-c 全国介護レセのみを用いた先行研究
- I-2-a 全国介護レセプトと介護サービス施設・事業所調査の突合
- I-2-b 介護レセと介護サービス・事業所調査データを突合して分析した先行研究
- I-3 全国介護レセ利用申請について理解する

II. 市町村レベルの医療・介護レセプト

- II-1 医療レセプトについて理解する
- II-2 医療・介護レセの突合
- II-3 医療・介護突合データを用いた先行研究

2

在宅医師としての原点
・自宅退院ゴールを達成した脳梗塞
女性の転院
・直腸がん術後の独居男性

病院では高い医療技術と共にありつつも、
その後の生活と幸福がある

Half Way Technology
by Lewis Thomas

医療
各種サービス
主な臨床研究対象

生活との乖離

生活との調和

介護 福祉サービス

入院治療開始
サービス利用開始
生活

Quality Of Life

社会経済状況 環境 地域 家族 etc

アウトカム

3

WHO事務局長(旧) Margaret Chan氏スピーチ

the Congress on Gerontology and Geriatrics 2012 より抜粋 (田宮が翻訳)

高齢者は、多くの病態を抱え、社会的ニーズを含めた多くのニーズを抱えている。人口の高齢化に伴い、医療と社会サービスとの緊密な協同が必須になる。

我々は、サービス提供のあり方の大きな転換をしなければならぬ。高齢者の健康な長寿を築くには、単一の疾病に対する適切な治療を提供することから、多くの疾病とともにあるよい健康状態をもたらすものへと、方向転換していくべきである。

保健医療制度や医学教育は、多くの疾病を抱えた状態への挑戦に向けて再構築されなければならない。これまでの長い間、感染症などの急性疾患に対する比較的短期間の対応をするべく、医療保険制度は設計され、医師は教育をされてきた。

WHOの視点として、高齢者のQOLを高め、彼らの尊厳、自立そして、社会との関わりを守るためのしくみは、薬剤やワクチンのような他の基本的な医療の産物と同じカテゴリーに属するものであり、それらと同様に取扱われるべきである。

http://www.who.int/dg/speeches/2012/ageing_20120330/en/index.html

これまで、データとして見えていた部分は・・・
これまでの病院主体の研究で把握できるのは・・・

高い医療技術と共にありつつ、この方たちは幸せなのか？

Half Way Technology
by Lewis Thomas

切り取った一部のみ

医療
各種サービス
主な臨床研究対象

生活との乖離

生活との調和

介護 福祉サービス

入院治療開始
サービス利用開始
生活

Quality Of Life

社会経済状況 環境 地域 家族 etc

アウトカム

予防・治療・看護・リハビリ・介護・福祉・・・
受けているのは一人の人間、家族・・・
→一連のサービスとしてみる必要性がある

各々の制度・サービスの狭間には？
必要とする人がアクセスできているのか？

これら一連の流れを把握し、隙間に光をあて、政策に生かす必要がある

5

I-1.介護給付費明細書 (介護レセプト) について理解する

6

参考資料 2：日本臨床疫学会シンポジウムの発表を予定していたスライド（計 22 枚中抜粋 6 枚）

University of Tsukuba
筑波大学

ヘルスサービス開発研究センター

日本臨床疫学会 シンポジウム
「保険医療介護ビッグデータ研究の人材育成」
生活と調和した医療のためのビッグデータ活用

筑波大学
医学・医療系 ヘルスサービスリサーチ分野
ヘルスサービス開発研究センター

田宮菜奈子

1

「健康の社会的決定要因」への医師の関与を求める

サー・マイケル・マーモット世界医師会講演会

サー・マイケル・マーモット世界医師会 (WMA) 会長の講演会が9月5日、約300名の参加者を集めて、日本医師会大講堂で開催された。

「健康の社会的決定要因 (Social Determinants of Health: SDH)」をテーマに講演したマーモットWMA会長は、健康格差は回避可能であるとして、それを引き起こす要因への医師の関与を求めた。

「健康の社会的決定要因 (SDH)」に関するマーモットWMA会長による一連の活動は、これまで、主に欧米、アフリカなど諸外国で行われてきたが、今回の講演会は、アジアにおいて健康の社会的決定要因に関する講演を行いたいとのWMA会長の希望を受けて開催されたものである。

「せっかく病気を治した人々を、なぜその病気にした状況に送り返してしまうのか」

もちろん医師は病気を治療するわけですが、**医師こそぜひ人々を病気にしてしまふ状況にも対応してもらいたい**

「医師は、病気を治すことに加えて、**人権の尊重、少年少女に対する教育、弱者の特定とその保護、司法、警察へのアクセスの確保**等にも関与できる私は考えています。」

<https://www.med.or.jp/nichionline/article/004708.html>

2

筑波大学 アドバンスドコース 情報医療・看護向上のためのヘルスサービスリサーチ 2018年8月 筑波大学 筑波大学大学院ヘルスサービスリサーチ分野 筑波大学 筑波大学

筑波大学
医学部4年生
アドバンスドコース
2日間

分析データ
つくば市ニーズ調査
米国のNHANES

医療系 学部教育
ビッグデータ解析体験
鉄は熱いうちに打て!

筑波大学 アドバンスドコース 情報医療・看護向上のためのヘルスサービスリサーチ 2018年8月 筑波大学 筑波大学大学院ヘルスサービスリサーチ分野 筑波大学 筑波大学

1. 一般学習目標
ヘルスサービスリサーチに関わる具体的な知識、技能、態度、態度を身に付けることにより、ヘルスサービスリサーチ分野の研究や実践に貢献し、健康増進・福祉向上に資する活動に参画する能力を育成する。

2. 参加目標 (本研修が終了後の学習成果に期待していること)
01. 「つくば市ニーズ調査」(基礎知識)を習得する。
02. 「つくば市ニーズ調査」(実践知識)を習得する。
03. NHANES データ解析のスキルを習得する。
04. NHANES データ解析のスキルを習得する。
05. NHANES データ解析のスキルを習得する。
06. NHANES データ解析のスキルを習得する。
07. NHANES データ解析のスキルを習得する。
08. NHANES データ解析のスキルを習得する。
09. NHANES データ解析のスキルを習得する。
10. NHANES データ解析のスキルを習得する。

3. その他
01. NHANES データ解析のスキルを習得する。
02. NHANES データ解析のスキルを習得する。
03. NHANES データ解析のスキルを習得する。
04. NHANES データ解析のスキルを習得する。
05. NHANES データ解析のスキルを習得する。
06. NHANES データ解析のスキルを習得する。
07. NHANES データ解析のスキルを習得する。
08. NHANES データ解析のスキルを習得する。
09. NHANES データ解析のスキルを習得する。
10. NHANES データ解析のスキルを習得する。

4. その他
01. NHANES データ解析のスキルを習得する。
02. NHANES データ解析のスキルを習得する。
03. NHANES データ解析のスキルを習得する。
04. NHANES データ解析のスキルを習得する。
05. NHANES データ解析のスキルを習得する。
06. NHANES データ解析のスキルを習得する。
07. NHANES データ解析のスキルを習得する。
08. NHANES データ解析のスキルを習得する。
09. NHANES データ解析のスキルを習得する。
10. NHANES データ解析のスキルを習得する。

3

筑波大学医学生研究室実習 (ビッグデータ解析)

新医学専攻 指導 5名 2012年～現在
研究室実習受け入れ 15名 2009年～現在 (2018) 原著論文3本

研究室実習 (全学年 随時)

2009	M5	2名	博士進学 (ダブルメジャー) →助教(米国・筑波大学MPH)
2013	M3	2名	ユタ大学 ハーバード大学(受け入れ Ichiro Kawachi 教授)へ
2014	M3	1名	
2015	M3	1名	
2016	M3	1名	日本公衆衛生学会優秀ポスター賞受賞
2017	M5	1名	
2018	M4	4名 M2 3名	

新医学専攻 (6年次半年)

2012	M5-6	1名	
2013	M5-6	1名	米国訪問
2016	M5-6	1名	米国訪問
2017	M5-6	1名	米国訪問、研修医として研究も継続中
2018	M5-6	1名	米国訪問、臨床疫学会(本学会)にて発表

3年次の時に研究実習を始めてから、学会発表の経験をし、アメリカでハーバードを見て、論文のアクセプトにまで至りました。先に発表されるがままに継続してきた結果、ここまで世界を駆けまわることができました。本当にありがとうございます。大学で最もお世話になり、影響を与えてくださった先生方ありがとうございます。(卒業生メール)

4

自主的な連絡を機にした、情報系の学生の学部～修士と継続的な関わり事例紹介
佐々木健佑氏 筑波大学 理工学群 社会学類 (修士課程)

情報系学生として医学研究室にきた経緯

- 経営工学の授業や実習などでは、分析のテーマや対象データが購買履歴データなど、マーケティング的なものに偏ってしまう部分があったこと
- 医療のデータ解析に関心があった中で、このままのカリキュラムでは医療データに関わるチャンスがなさそうだと感じたこと
- 「医療データ」は数層が高く感じており、直接そのようなデータを扱っている研究室に飛び込まなければチャンスがないと感じていたこと

5

情報系学生としての活躍の場

- 情報系の学生が身につけるプログラミングやデータ処理のスキルは、医療系研究室で必要とされる部分が多く、活躍の機会は少なくない
- 医療レセプトの特性(レセプト病名など)を把握している専門家の下でレビューを受けながら、データ整形を行っていく必要がある
- 情報系学生が活躍するための課題は医療分野の背景知識の補填と、人材不足

6

参考資料 3：一市町村の医療・介護連結レセプトを用いた尿路カテーテルについての研究 (2018)

P1-5

Long-term urinary catheterization and urinary tract infection among older people in the community

OMotohiko Adomi¹, Masao Iwagami^{2,3}, Takashi Kawahara⁴, Shota Hamada^{2,5}, Katsuya Iijima⁶, Satoru Yoshie^{2,6,7}, Tatsuro Ishizaki⁸, Nanako Tamiya^{2,3}

1. School of Medicine, University of Tsukuba, 2. Health Services Research and Development Center, University of Tsukuba, 3. Department of Health Services Research, Faculty of Medicine, University of Tsukuba, 4. Department of Urology, University of Tsukuba, 5. Institute for Health Economics and Policy, 6. Institute of Gerontology, University of Tokyo, 7. Department of Health Policy and Management, School of Medicine, Keio University, 8. Human Care Research Team, Tokyo Metropolitan Institute of Gerontology



Background

- Long-term urinary catheterization (LTUC: urinary catheterization for >28 days) is seen among people in the community.
- The association between urinary catheterization and urinary tract infection (UTI) has been well established in the hospital setting, but not in the community setting.

Objectives

- To examine the association between LTUC and incidence of UTI among older people living in the community.

Methods

- Data: Medical insurance and long-term care insurance claims data, the City of Kashiwa, Chiba
- Study period: April 2012 – September 2013
- Study population: Older people born before 1940 and living in the community
- Exposure: LTUC
Definition: Insertion/replacement of urinary catheter was claimed for more than 2 consecutive months.
- Outcome: First incidence of UTI requiring antibiotics after cohort entry
Definition: UTI was recorded in claims and antibiotics were administered in the same month.
- Covariates:
sex, age, charlson comorbidity index (CCI), history of hospitalization with/without use of urinary catheter, severe care-needs level (care-needs level ≥3)

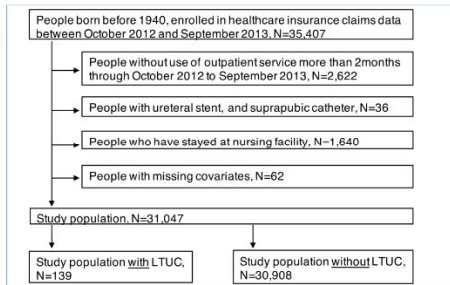


Fig. 1: Flow of population selection

	People with LTUC: n=139	People without LTUC: n=30,908	P-value
Sex, n(%)			
Men	97 (69.8)	13,338 (43.2)	<0.001 ¹
Age, n(%)			
72-76 (born in 1935-1939)	24 (17.3)	11,352 (36.7)	
77-81 (born in 1930-1934)	35 (25.2)	10,778 (34.9)	
82-86 (born in 1925-1929)	48 (34.5)	5,712 (18.5)	
87- (born before 1924)	32 (23.0)	3,066 (9.9)	<0.001 ²
CCI, n(%)			
0	26 (18.7)	12,330 (39.9)	
1-2	35 (25.2)	10,280 (33.3)	
3-	78 (56.1)	8,298 (26.9)	<0.001 ²
Past history of hospitalization, n(%)			
No	76 (54.7)	22,305 (72.2)	
Yes& without catheter	35 (25.2)	6,961 (22.5)	
Yes& with catheter	28 (20.1)	1,642 (5.3)	<0.001 ²
Severe care needs level, n(%)	40 (28.8)	1,979 (6.4)	<0.001 ¹

Table1: Characteristics of patients with and without LTUC

*1 Chi-square tests *2 Wilcoxon rank sum tests

Statistical Analysis

- Poisson regression analysis
- Propensity score matched analysis
- 1:3 ratio (LTUC patients: non-LTUC patients), nearest neighbor matching
- Propensity score generated from aforementioned covariates.

Results

- 139 patients (0.45%) with LTUC among 31,047 patients. (Flowchart shown in Fig. 1)
- Incidence rates of UTI were 32.9 and 4.7 per 100 person-years in people with and without LTUC.

	Follow-up period (person-years)	No. of patients having UTI and treated with antibiotics	Incidence rate of UTI (/100 person-years) (95% CI)	Rate ratio of UTI ¹ (95%CI)	
				Crude	Fully adjusted ²
Patients without LTUC: n=30,908	26,217	1,220	4.7 (4.3 – 4.9)	Reference	Reference
Patients with LTUC: n=139	64	21	32.9 (21.4 – 50.4)	7.06 (4.59 – 10.87)	2.39 (1.54 – 3.71)

Table2: Incidence rate ratio of UTI in patients with and without LTUC before propensity score matching

*1. Poisson regression analysis *2. Adjustment covariates: sex, age, CCI, past history of hospitalization, severe care-needs level

	Follow-up period (person-years)	No. of patients having UTI and treated with antibiotics	Incidence rate of UTI (/100 person-years) (95% CI)	Rate ratio of UTI ¹ (95%CI)	
				Crude	Fully adjusted ²
Patients without LTUC: n=417	329	48	14.6 (11.0 – 19.4)	Reference	Reference
Patients with LTUC: n=139	64	21	32.9 (21.4 – 50.4)	2.25 (1.35 – 3.76)	2.25 (1.35 – 3.76)

Table3: Incidence rate ratio of UTI in patients with and without LTUC after propensity score matching

*1. Poisson regression analysis

Discussion

- Medical. Long-term care insurance linked database enabled to perform analysis of people living in the community.
- Misclassification of incidence of UTI (defined as UTI was recorded in claims and antibiotics were administered in the same month) may be possible because exact reason for antibiotics prescription was unknown.
- Further research on clinical indication of catheterization and alternative methods (e.g., surgical intervention or intermittent catheterization) may need to be considered.

Conclusion

- We clarified prevalence rate of LTUC and incidence rate of UTI among older people in the community. The LTUC was significantly associated with the increased risk of UTI requiring antibiotics.

Conflict of Interest: The presenter has no conflict to disclose regarding my presentation.

についての研究(2018年度の Gerontological Society of America 2018 Annual Scientific Meetingにて発表)

Hearing Loss and Outdoor Activity Limitation, Psychological Distress, and Memory Loss among Elderly People in Japan

Masao Iwagami^{1,3}, Yoko Kobayashi², Eriko Tsukazaki¹, Taeko Watanabe¹, Takehiro Sugiyama^{1,4,5}, Tetsuro Wada⁶, Akira Hara⁶, Nanako Tamiya¹

1 Department of Health Services Research, University of Tsukuba 2 Research and Support Center on Higher Education for the Hearing and Visually Impaired, Tsukuba University of Technology
3 Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine 4 Diabetes and Metabolism Information Center, Research Institute, National Center for Global Health and Medicine 5 Department of Public Health/Health Policy, Graduate School of Medicine, the University of Tokyo 6 Department of Otolaryngology, University of Tsukuba

Introduction

- The prevalence of hearing loss is high and projected to increase further in most countries with population aging.
- However, national data on psychical and psychological outcomes associated with hearing loss are scarce.
- We aimed to investigate the associations between self-reported hearing loss and outdoor activity limitations, psychological distress, and memory loss in Japan.

Methods

Study design: Cross-sectional study
Data source: Nationally-representative questionnaire survey (2016 Comprehensive Survey of Living Conditions in Japan)
Exposure: Self-reported hearing loss (Yes/No)
Outcomes: Outdoor activity limitations (Yes/No), psychological distress (Kessler Psychological Distress Scale [K6] score of ≥ 5), and self-reported memory loss (Yes/No)

- Outdoor activity limitation means the limitation of work and job.
- K6 consists of six-item psychological distress questions, each of which is measured on a five-point scale 0-4.
- Memory loss was investigated using binary question.

Statistical analysis: Logistic regression analyses

Methods

Study population: people aged ≥ 65 years living at home, without a diagnosis of dementia

```

    graph TD
      A[289,470 households asked to participate in the survey] --> B[224,208 (77.6%) households, including 568,426 people]
      B --> C[20,545 people not living at home  
387,240 people aged <65 years]
      C --> D[160,393 people aged  $\geq 65$  years]
      D --> E[3,934 people with dementia diagnosis]
      E --> F[137,723 study participants]
      F --> G[Hearing loss (+)  
n = 12,389]
      F --> H[Hearing loss (-)  
N = 125,334]
    
```

Results

Fig 1. Prevalence of self-reported hearing loss (%)

Results(Continued)

Fig 2. Prevalence of outdoor activity limitations, psychological distress, and memory loss by self-reported hearing loss status

Results(Continued)

Table 1. Associations between self-reported hearing loss and outdoor activity limitations, psychological distress, and memory loss

	Number of outcomes (%)	Unadjusted	Age- and sex-adjusted	Fully-adjusted
Outdoor activity limitations (Yes/No):				
No hearing loss	11,846 (9.5)	1 (ref.)	1 (ref.)	1 (ref.)
Hearing loss	3,577 (28.9)	3.89 (2.75-4.98)	2.78 (2.65-2.91)	2.56 (2.45-2.69)
Psychological distress (Kessler Psychological Distress Scale [K6] ≥ 5):				
No hearing loss	24,212 (19.3)	1 (ref.)	1 (ref.)	1 (ref.)
Hearing loss	4,917 (39.7)	2.75 (2.64-2.86)	2.53 (2.41-2.63)	2.42 (2.32-2.52)
Self-reported memory loss (Yes/No):				
No hearing loss	6,480 (5.2)	1 (ref.)	1 (ref.)	1 (ref.)
Hearing loss	4,674 (37.7)	11.08 (10.62-11.59)	9.42 (9.00-9.85)	8.67 (8.17-9.20)

Confounding factors: smoking, alcohol, educational status, number of household members, average household expenditure, and chronic illnesses (DM, HT, CAD, stroke, thyroid disease, Parkinson's disease, and cancer)

Discussion & Conclusion

- Older people with hearing loss were more likely to experience outdoor activity limitations, psychological distress, and memory loss than those without hearing loss in Japan.
- Early identification of and intervention in hearing loss (e.g. hearing aid) may potentially reduce the risk of these poor health outcomes

参考資料 5 : 日本の大規模入院データを用いた大腿骨頸部骨折手術後の早期リハビリが認知症高齢者の日常生活機能に与える影響を検討した研究 (2018年度の Gerontological Society of America 2018 Annual Scientific Meetingにて発表)

Early Intensive Rehabilitation after Hip Fracture Surgery and Activities of Daily Living in Patients with Dementia

Kazuaki Uda MPH PT¹, Hideo Yasunaga PhD MD¹, Nanako Tamiya PhD MD²

1. Department of Clinical Epidemiology and Health Economics, School of Public Health, The University of Tokyo
2. Department of Health Services Research, Faculty of Medicine, Research and Development Center for Health Services, University of Tsukuba

Objective

To examine the effects of early intensive postoperative rehabilitation on activities of daily living (ADL) based on the Barthel Index (BI) after hip fracture surgery in patients with dementia.

Methods	Results	Discussion																																																																																																																														
<ul style="list-style-type: none"> Design: Retrospective cohort study Data source/Period: The Diagnosis Procedure Combination database(na nationwide inpatient administrative claims and discharge database in Japan)/1 April 2014 to 31 March 2016. Participants: aged ≥ 65 years with a diagnosis of dementia at admission who underwent hip fracture surgery and received postoperative rehabilitation. Exclusion: those who were discharged within 7 days after surgery, and those who died in the hospital. Scale for dementia: scale for dementia provided by the Ministry of Health, Labour and Welfare in Japan. This includes the 6 categories: (I) no dementia; (II) level I, some dementia but almost independent in ADL; (III) level II, mild dementia with some difficulty in communication, but independent in ADL with minimal observation; (IV) level III, dementia with some difficulty in communication and requiring partial care; (V) level IV, severe dementia; and (VI) level V, dementia with severe psychiatric symptoms, behavioral and psychological symptoms, or physical disorders. We identified patients with dementia as recorded with level I or greater. ADL: We categorized these into 4 groups: < 1.0 units/day; 1.0-1.9 units/day; 2.0-2.9 units/day; ≥ 3.0 units/day. One rehabilitation unit was 20 minutes. Outcome Measure: BI score at discharge from acute-care hospitals. Statistical Analysis: A multivariable linear regression analysis was conducted to determine whether early intensive postoperative rehabilitation was associated with better BI at discharge from acute care hospitals with adjustment for the patient's and surgical characteristics, while also accounting for clustering within hospitals using generalized estimating equations. 	<ul style="list-style-type: none"> Eligible patients: 42,468 patients (at 1,048 acute care hospitals). Table: Baseline characteristics of patients grouped by the early rehabilitation intensity. <table border="1" style="font-size: small;"> <thead> <tr> <th>Characteristic</th> <th><1.0 units/day (n=14,736)</th> <th>1.0-1.9 units/day (n=4,351)</th> <th>2.0-2.9 units/day (n=2,742)</th> <th>≥ 3.0 units/day (n=17,639)</th> <th>P</th> </tr> </thead> <tbody> <tr> <td>Age, years</td> <td>75.64</td> <td>75.64</td> <td>75.64</td> <td>75.64</td> <td>0.25</td> </tr> <tr> <td>Sex</td> <td>50.2</td> <td>50.2</td> <td>50.2</td> <td>50.2</td> <td><0.001</td> </tr> <tr> <td>Male</td> <td>17.5</td> <td>17.5</td> <td>17.5</td> <td>17.5</td> <td>0.99</td> </tr> <tr> <td>Female</td> <td>52.9</td> <td>52.9</td> <td>52.9</td> <td>52.9</td> <td><0.001</td> </tr> <tr> <td>Body mass index, kg/m²</td> <td>23.2</td> <td>23.2</td> <td>23.2</td> <td>23.2</td> <td>0.16</td> </tr> <tr> <td><18.5</td> <td>8.3</td> <td>8.3</td> <td>8.3</td> <td>8.3</td> <td>0.2</td> </tr> <tr> <td>18.5-24.9</td> <td>5.7</td> <td>5.7</td> <td>5.7</td> <td>5.7</td> <td>0.2</td> </tr> <tr> <td>25.0-29.9</td> <td>10.4</td> <td>10.4</td> <td>10.4</td> <td>10.4</td> <td>0.002</td> </tr> <tr> <td>≥ 30</td> <td>12.8</td> <td>12.8</td> <td>12.8</td> <td>12.8</td> <td>0.001</td> </tr> <tr> <td>ADL</td> <td>58.3</td> <td>58.3</td> <td>58.3</td> <td>58.3</td> <td>0.001</td> </tr> <tr> <td>I-II</td> <td>41.7</td> <td>41.7</td> <td>41.7</td> <td>41.7</td> <td>0.004</td> </tr> <tr> <td>III-IV</td> <td>5.0</td> <td>5.0</td> <td>5.0</td> <td>5.0</td> <td><0.001</td> </tr> <tr> <td>V</td> <td>50.4</td> <td>50.4</td> <td>50.4</td> <td>50.4</td> <td><0.001</td> </tr> <tr> <td>VI</td> <td>55.3</td> <td>55.3</td> <td>55.3</td> <td>55.3</td> <td><0.001</td> </tr> <tr> <td>Type of fracture</td> <td>41.7</td> <td>41.7</td> <td>41.7</td> <td>41.7</td> <td><0.001</td> </tr> <tr> <td>Intertrochanteric</td> <td>71.4</td> <td>71.4</td> <td>71.4</td> <td>71.4</td> <td><0.001</td> </tr> <tr> <td>Subtrochanteric</td> <td>28.6</td> <td>28.6</td> <td>28.6</td> <td>28.6</td> <td><0.001</td> </tr> <tr> <td>Distal femoral</td> <td>51.2</td> <td>51.2</td> <td>51.2</td> <td>51.2</td> <td><0.001</td> </tr> <tr> <td>Proximal femoral</td> <td>31.5</td> <td>31.5</td> <td>31.5</td> <td>31.5</td> <td><0.001</td> </tr> <tr> <td>Distal humeral</td> <td>31.5</td> <td>31.5</td> <td>31.5</td> <td>31.5</td> <td><0.001</td> </tr> </tbody> </table>	Characteristic	<1.0 units/day (n=14,736)	1.0-1.9 units/day (n=4,351)	2.0-2.9 units/day (n=2,742)	≥ 3.0 units/day (n=17,639)	P	Age, years	75.64	75.64	75.64	75.64	0.25	Sex	50.2	50.2	50.2	50.2	<0.001	Male	17.5	17.5	17.5	17.5	0.99	Female	52.9	52.9	52.9	52.9	<0.001	Body mass index, kg/m ²	23.2	23.2	23.2	23.2	0.16	<18.5	8.3	8.3	8.3	8.3	0.2	18.5-24.9	5.7	5.7	5.7	5.7	0.2	25.0-29.9	10.4	10.4	10.4	10.4	0.002	≥ 30	12.8	12.8	12.8	12.8	0.001	ADL	58.3	58.3	58.3	58.3	0.001	I-II	41.7	41.7	41.7	41.7	0.004	III-IV	5.0	5.0	5.0	5.0	<0.001	V	50.4	50.4	50.4	50.4	<0.001	VI	55.3	55.3	55.3	55.3	<0.001	Type of fracture	41.7	41.7	41.7	41.7	<0.001	Intertrochanteric	71.4	71.4	71.4	71.4	<0.001	Subtrochanteric	28.6	28.6	28.6	28.6	<0.001	Distal femoral	51.2	51.2	51.2	51.2	<0.001	Proximal femoral	31.5	31.5	31.5	31.5	<0.001	Distal humeral	31.5	31.5	31.5	31.5	<0.001	<ul style="list-style-type: none"> Strength of the study <ul style="list-style-type: none"> Focusing on patients with dementia in a real-world setting using a nationwide inpatient database. Our results suggest that 20 to 40 additional minutes for each daily rehabilitation may provide better functional outcomes for patients with dementia. Study limitations <ul style="list-style-type: none"> Unmeasured confounders may have biased the results. Lack of data on longer-term outcomes. Detailed information about the contents of the rehabilitation programs were unclear. <h4 style="text-align: center;">Conclusion</h4> <p style="text-align: center;">The results suggest that early intensive postoperative rehabilitation is associated with a better recovery of ADL in acute care setting among patients with dementia after hip fracture surgery.</p>
Characteristic	<1.0 units/day (n=14,736)	1.0-1.9 units/day (n=4,351)	2.0-2.9 units/day (n=2,742)	≥ 3.0 units/day (n=17,639)	P																																																																																																																											
Age, years	75.64	75.64	75.64	75.64	0.25																																																																																																																											
Sex	50.2	50.2	50.2	50.2	<0.001																																																																																																																											
Male	17.5	17.5	17.5	17.5	0.99																																																																																																																											
Female	52.9	52.9	52.9	52.9	<0.001																																																																																																																											
Body mass index, kg/m ²	23.2	23.2	23.2	23.2	0.16																																																																																																																											
<18.5	8.3	8.3	8.3	8.3	0.2																																																																																																																											
18.5-24.9	5.7	5.7	5.7	5.7	0.2																																																																																																																											
25.0-29.9	10.4	10.4	10.4	10.4	0.002																																																																																																																											
≥ 30	12.8	12.8	12.8	12.8	0.001																																																																																																																											
ADL	58.3	58.3	58.3	58.3	0.001																																																																																																																											
I-II	41.7	41.7	41.7	41.7	0.004																																																																																																																											
III-IV	5.0	5.0	5.0	5.0	<0.001																																																																																																																											
V	50.4	50.4	50.4	50.4	<0.001																																																																																																																											
VI	55.3	55.3	55.3	55.3	<0.001																																																																																																																											
Type of fracture	41.7	41.7	41.7	41.7	<0.001																																																																																																																											
Intertrochanteric	71.4	71.4	71.4	71.4	<0.001																																																																																																																											
Subtrochanteric	28.6	28.6	28.6	28.6	<0.001																																																																																																																											
Distal femoral	51.2	51.2	51.2	51.2	<0.001																																																																																																																											
Proximal femoral	31.5	31.5	31.5	31.5	<0.001																																																																																																																											
Distal humeral	31.5	31.5	31.5	31.5	<0.001																																																																																																																											

参考文献 6：つくば市の高齢者に対するアンケート調査の結果を用いたポリファーマシーについての研究（2018年度の第29回日本疫学会学術総会にて発表）（計13枚中最初の6枚）

第29回日本疫学会学術総会 2019年1月31日
一般口演(1) 社会疫学・近隣環境・高齢者

筑波大学
University of Tsukuba

高齢者における通院中の病院数と多剤併用の関連

鈴木俊輝¹ 岩上将夫^{2,3} 田宮菜奈子^{2,3}

¹筑波大学医学群医学類
²筑波大学医学医療系ヘルスサービスリサーチ分野
³筑波大学ヘルスサービス開発センター

筑波大学
University of Tsukuba

背景：多剤併用 (polypharmacy)

- 6種類以上の服薬で、薬物有害事象が増加¹
- 高齢者の処方薬剤数は外来患者 平均4.4剤² 入院患者 平均6.4剤³
- 日本はフリーアクセス → 複数の医療機関に受診 → 多剤併用のリスク？

薬物有害事象の割合 (%)

処方薬剤数	薬物有害事象の割合 (%)
1~3	~6
4~5	~9
6~7	~13
8~9	~12
10~	~14

¹Kojima T: GGI:2012;12:761. ²Suzuki Y: GGI:2006;6:244. ³Fushiki Y: General medicine: 2014;15:110 1

筑波大学
University of Tsukuba

目的

つくば市の65歳以上の高齢者において、現在通院中の医療機関数と多剤併用の関連を検討する。

曝露因子 → アウトカム

通院中の医療機関数 → 多剤併用

交絡因子

年齢・性別・疾患・飲酒・喫煙・経済状況

筑波大学
University of Tsukuba

方法

データ：つくば市高齢者福祉計画策定のためのアンケート調査
調査期間：2016年12月
調査対象：つくば市在住の要介護認定を受けていない高齢者

	配布件数 (ランダム抽出)	回収件数	回収率
前期高齢者 (65~74歳)	1500	759	50.6%
後期高齢者 (75歳~)	1500	798	53.2%
合計	3000	1557	51.9%

3

筑波大学
University of Tsukuba

方法

説明変数：通院中の医療機関数

問11	現在、病院・医院（診療所、クリニック）に通院していますか。	1. はい 2. いいえ
問11-1	【問11で「1. はい」の方のみ】何か所の病院・医院に通院していますか。また、掛かっている診療科の種類は何種類ありますか。	病院・医院 (2) か所 掛かっている診療科 () 種類

目的変数：多剤併用 (6種類以上)

問10	現在、医師の処方した薬を何種類飲んでますか。	1. 1種類 2. 2~3種類 3. 4~5種類 4. 6種類以上 5. 飲んでいない
-----	------------------------	---

4

筑波大学
University of Tsukuba

方法：

治療中の疾患 (重要な交絡因子の1つ)

問9	現在治療中、又は後遺症のある病気はありますか。(いくつでも)	1. ない 2. 高血圧 3. 脳卒中 (脳出血・脳梗塞等) 4. 心臓病 5. 糖尿病 6. 高脂血症 (脂質異常) 7. 呼吸器の病気 (肺炎や気管支炎等) 8. 胃腸・肝臓・胆のうの病気 9. 腎臓・前立腺の病気 10. 筋骨格の病気 (骨粗しょう症、関節症等) 11. 外傷 (転倒・骨折等) 12. がん (悪性新生物) 13. 血液・免疫の病気 14. うつ病 15. 認知症 (アルツハイマー病等) 16. パーキンソン病 17. 目の病気 18. 耳の病気 19. その他 ()
----	--------------------------------	--

5

厚生労働科学研究費補助金
政策科学総合研究事業（臨床研究等 ICT 基盤構築・人工知能実装研究事業）
分担研究報告書

大規模データマネジメント手法開発と人材育成に関する研究

研究分担者 笹淵裕介 自治医科大学データサイエンスセンター

研究要旨：医療ビッグデータハンドリング教育プログラムを作成し、日本臨床疫学会 NDB・DPC データベース研究人材育成＜短期集中セミナー＞において提供した。また、自治医科大学データサイエンスセンターにおいても継続的に研究者へ教育プログラムを提供した。その結果、短期集中セミナーにおいては高い評価を受け、また自治医科大学データサイエンスセンターの受講者は熊本県国民健康保険および後期高齢者医療制度における医療レセプトデータを研究者自身で加工からデータの解析までを行うことができるようになり、医療ビッグデータ研究の促進につながった。今後も継続的に本教育プログラムの提供を通して、医療ビッグデータを用いた研究へつなげていきたい。

A．研究目的
近年医療ビッグデータを利用した研究が爆発的に増加しており、臨床判断や医療政策策定に必要なエビデンスの創出に大きな役割を占めていることはすでに周知の事実である。医療ビッグデータを利用した研究を行うためには通常の臨床データを用いた臨床研究と異なり、臨床の知識・臨床研究の知識に加えて、データベース、統計学、機械学習、プログラミングの知識や技術を要する。しかしながら、これまでに臨床家がこれらの知識や技術を習得するための体系だった教育プログラム等はほとんど存在しないため、研究者が自己学習によってこれらの知識や技術を習得することは非効率である。本研究の目的は 医療ビッグデータを利用した研究を行うにあたり必要なデータハンドリング技術である SQL 言語、統計解析や機械学習に必要な R、SPSS、python 等の

統計ソフトやプログラミング言語の習得を目指す教育プログラムの作成 作成したプログラムを用いた教育の提供を行うことである。

B．研究方法

国内外にデータサイエンティストを養成する目的の書籍、オンライン講座、あるいは講習会等から医療ビッグデータの研究に必要なエッセンスを抽出し、さらに臨床家と疫学・統計学専門家の両方とコミュニケーションを取るための知識や技術を加えることで医療ビッグデータに特化した教育プログラムを作成した。作成した教育プログラムを 2018 年 8 月 6 日(月)～8 月 10 日(金)に行われた日本臨床疫学会 NDB・DPC データベース研究人材育成＜短期集中セミナー＞(以下サマーセミナー)において提供した。これに加えて、作成した教育プログラムを自治医科大学データサイエンスセンタ

ーにおいて臨床家へ提供した。

C. 研究結果

様々な関連書籍、オンライン講義等からエッセンスを抽出し、SQL, R, SPSS, python に関する教育プログラムを作成した。サマーセミナーにおいてこれらのプログラムを受講者に提供した。

各教育プログラムの作成および提供

(1) SQL によるデータベースハンドリング

複数のテーブルから SELECT 文により必要な情報を抽出・集計し、これらを JOIN により統合することを基本として、サブクエリを利用したやや複雑なクエリなどを自分自身で書くことを目的とした。SQL 習得プログラムにより統計解析・機械学習に利用するためのデータセットを抽出することができるようになる。

サマーセミナーでは実際に受講者が SQL によってサンプルデータセットから統計解析や機械学習を行うために必要なデータセットへの抽出作業をハンズオン形式で行った。受講後のアンケートでは 90%以上の受講者がわかりやすかった、満足したと回答した。

自治医科大学ではデータサイエンスセンターでの研究に参加している研究者に対して同じプログラムを提供した。受講者は全員自身の研究計画に沿ったデータの抽出を自身で行うことができるようになった。

(2) SPSS/R による統計解析

SQL によって抽出したデータを利用し、(i)データのクリーニング、(ii)各変数の集計及び可視化、(iii)重回帰分析、ロジスティック回帰分析、生存時間分析、(iv)傾向スコア分析を自分自身で行うことが可能となることを目的とする。統計初心者でも比較的扱

いやすいことから SPSS、EZR を利用した。また、より高度な統計解析が必要になった際には R を用いることとした。統計解析習得プログラムにより、臨床疫学研究で利用される一般的な統計手法を行うことができるようになる。

サマーセミナーでは R の基本的操作から傾向スコア分析までの一通りのデータ分析の手法についてのハンズオン形式の講義を提供した。講義後のアンケートでは 70%以上の受講者がわかりやすいと評価し、約 90%の受講者が満足したと回答した。

自治医科大学ではデータサイエンスセンターでの研究に参加している研究者に対して R に加えて EZR、SPSS での同様のプログラムを提供した。SQL で抽出したデータを EZR、R、SPSS を用いて、研究者自身で基本的な統計解析を行うことができるようになった。

(3) Python による機械学習

データ分析に有用なライブラリである Numpy 及び Pandas の基本的な使い方を学び、ついで機械学習ライブラリである scikit-learn を利用して回帰・k 近傍法・サポートベクターマシン・ランダムフォレスト等、機械学習の基礎を学ぶ。機械学習プログラムにより、これらの基本的な機械学習を行うことができるようになる。

サマーセミナーにおいて講義形式の教育プログラムを提供した。受講後のアンケートでは約 70%の受講者がわかりやすかった、満足したと回答した。

ビッグデータを利用した研究

自治医科大学では、今回開発した教育プログラムで身につけるべき知識・技術を用いて、実際に医療ビッグデータ(熊本県の国民

健康保険及び後期高齢者医療制度レセプトデータ)を用いた研究を行った。

熊本県地震が住民の健康に与えた影響

【目的】2016年4月、熊本県において二度の震度7を観測する地震が発生した。この地震後、プライマリ・ケアへのアクセス障害により地域住民の健康が損なわれていた可能性があるものの、これまで詳細な検討がなされていない。そこで、熊本県地震後短期的に避けられる入院が増加しているかどうかを検討した。【方法】本研究では熊本県国民健康保険レセプトデータを用いて熊本県地震と避けられる入院の関連を検討した。避けられる入院の定義としてイギリスのNational Health Serviceにより外来診療へのアクセスの評価指標として利用されているAmbulatory care sensitive condition (ACSC)による入院を利用した。2016年3月15日-4月13日を地震前、2016年4月17日-5月16日を地震後として地震前後の避けられる入院数の変化を2012年-2015年の同時期の変化と比較する差の差分析を行った。【結果】研究期間において、「ワクチンによって避けられる入院」、「早期介入によって避けられる入院」、「適切な外来での管理によって避けられる入院」はそれぞれ7,921名、18,763名、85,436名であった。地震前後7日間の比較では「ワクチンによって避けられる入院」、「早期介入によって避けられる入院」、「適切な外来での管理によって避けられる入院」はそれぞれ32.6% (10.2-59.5, $p=0.003$), 44.1% (27.0-63.5, $p<0.001$), 27.7% (20.2-35.6, $p<0.001$)の増加を認めた。しかしながら、地震前後30日間の比較ではいずれの入院についても統計学的な有意差を認めなかった。【考察】熊本

県地震では震災後早期に避難所から帰宅できた住民が多かったことから、この様な災害の場合震災直後の避けられる入院を減らすために、可能な限り早期に介入が必要であることが示唆された。

レセプトデータを用いた外来経口抗菌薬使用実態の疫学解析

【背景】抗菌薬適正使用のためには、どの感染症にどの程度の抗菌薬が処方されているかの把握が肝要である。【方法】熊本県の国民健康保険加入者と後期高齢者医療制度対象者の外来レセプト(2012年度, 777万レセプト)から、受診年月と受診医療機関が共通の感染症病名と経口抗菌薬を抽出し、疾患・薬剤ごとの抗菌薬処方レセプト数・抗菌薬処方率を算出した。また、急性気道感染症(上気道炎, 咽頭炎, 気管支炎, 副鼻腔炎)と急性下痢症に対する抗菌薬処方率を医療機関規模別に解析した。【結果】1県・1年間で69万レセプトの経口抗菌薬が処方され、第3世代セフェム(35%), マクロライド(32%), キノロン(21%)が大部分を占めた。病名と使用抗菌薬が突合できた50万レセプトの解析では、抗菌薬の59%が急性気道感染症に、9%が急性下痢症に処方された。最大の処方対象疾患は上気道炎であった(10万レセプト, 21%)。上気道炎, 咽頭炎, 気管支炎, 副鼻腔炎における抗菌薬処方率はそれぞれ39%, 59%, 57%, 66%であった。急性気道感染症に対する医療機関規模別の抗菌薬処方率は、診療所55%, 20-199床33%, 200-499床29%, 500床以上20%であった。【考察】広域抗菌薬が極端に多く、最多の処方対象は本来抗菌薬の不要な上気道炎(感冒)で、咽頭炎や気管支炎でも約6割に使用された。医療機関規

模が小さいほど抗菌薬 の処方率が高かった.AMR 対策啓発の際には考慮を要する。これ以外にも現在複数の研究者が教育プログラムを受け、研究者自身の研究テーマに沿ってデータの抽出から分析を行い、発表へ向けて準備を進めている。

(倫理面への配慮)

倫理的な問題はない。

D . 考察

既存の教育プログラムでは不十分であった医療ビッグデータを用いた研究のための教育プログラムを作成し、試行した。このプログラムの受講者の評価は非常に高く、また実際に医療ビッグデータを利用した研究へ繋がった。

医療ビッグデータを用いた研究を行うにあたり、最も一般的に利用される言語はRDBMS 操作のためのSQL、統計解析のためのR やSPSS、機械学習のためのpythonなどが挙げられる。従って、医療ビッグデータを用いた研究を行う研究者はこれらの言語を習得する必要がある。しかしながら、例えばSQL を学ぶにあたって、一般的な書籍や講座等ではデータ定義言語、データ操作言語、データ制御言語を学ぶことになるが、実際に医療ビッグデータを用いた研究にはデータ定義言語及びデータ操作言語が必要である。さらに、この中でも特によく使われるクエリは限られているが、どのクエリが研究に必要なのかなどは初学者にはわからないため、効率的な学習が困難である。本研究において作成した教育プログラムでは特に医療ビッグデータを用いた研究に特化し、頻出するクエリを中心に学習できることから一般的な書籍や講座と比較して圧倒的に効率的に学習することが可能である。これ

はSPSS やR、python についても同様である。実際に医療ビッグデータを利用した研究を行いたいと考える研究者が本教育プログラムを受講し、高い評価を受けていること、自治医科大学で実際に研究に結びついていることから効率的に身につけることができていると考えられる。

E . 結論

医療ビッグデータ研究の為に必要な知識・技術を養成するための教育プログラムを作成・提供した。本教育プログラムはわかりやすく満足度も高いと評価を受け、実際の研究にも結びつくことが明らかとなった。今後も本教育プログラムを通して多くのより研究者が医療ビッグデータを利用した研究を行うための知識と技術を身につけられるよう、継続的に提供していきたい。

F . 健康危険情報

なし

G . 研究発表

1.論文発表

1-1. Sasabuchi Y, Matsui H, Kotani K, Lefor AK, Yasunaga H. Effect of the 2016 Kumamoto earthquakes on preventable hospital admissions: a retrospective cohort study in Japan. *BMJ Open*. 2018;8:e021294.

1-2. Hashimoto H, Matsui H, Sasabuchi Y, Yasunaga H, Kotani K, Nagai R, Hatakeyama S. Antibiotic prescription among outpatients in a prefecture of Japan, 2012-2013: a retrospective claims database study. *BMJ Open*. 2019;9:e026251

2.学会発表

2-1. Y Sasabuchi, K Kotani, H Matsui, AK

Lefor, H Yasunaga. Effect of the 2016 Kumamoto Earthquakes on Preventable Hospital Admissions: A Retrospective Cohort Study. Academy Health Annual Research Meeting, Seattle, the US, June 24-26, 2018.

2-2. 橋本英樹, 畠山修司, 松居宏樹, 笹淵裕介, 康永秀生. レセプトデータを用いた外来経口抗菌薬使用実態の疫学解析. 第92回日本感染症学会総会・学術講演会. 2018/5/31-6/2 岡山.

H . 知的財産権の出願・登録状況
なし