

平成 29 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・DPC 等データベース研究人材育成プログラムの開発

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

NDB・DPC 等の医療ビッグデータ解析に精通した研究者を育成することは急務である。平成 29 年度研究では、（1）医療ビッグデータ研究の実践、（2）ビッグデータ研究実践能力養成プログラム案の開発、（3）ビッグデータハンドリング技術養成プログラム案の開発を実践した。多数の若手研究者を指導し、医療ビッグデータ研究で H29 年度以降に 58 編の英文原著論文が掲載・受理された。その過程で実施した教育プロセス（データハンドリング技術、観察研究における統計解析技術、など）を体系化・一般化し、既存の知識（NDB の落とし穴等）と合わせて、上記の（2）（3）に活かした。開発した人材育成プログラム案を H30 年度に実施し、その効果を検証し、必要な修正を加え、恒久プログラム化する。

### 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  
松居宏樹 東京大学大学院医学系研究科 臨床疫学・経済学 助教  
加藤源太 京都大学医学部附属病院、診療報酬センター 准教授  
篠崎智大 東京大学大学院医学系研究科 生物統計学 助教  
山名隼人 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 特任助教

道端伸明 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 特任助教  
大野 幸子 東京大学生物統計情報学 特任助教  
麻生将太郎 東京大学大学院医学系研究科臨床疫学・経済学 特任研究員  
森田 光治良 東京大学大学院医学系研究科臨床疫学・経済学 特任研究員  
宇田和晃 東京大学大学院医学系研究科臨床疫学・経済学 特任研究員  
石丸美穂 東京大学大学院医学系研究科臨床疫学・経済学 特任研究員

## A. 研究目的

NDB データの解析に精通した研究者は非常に限られている。

DPC データの研究利用は着実に進んでいるものの、さらに研究者の裾野を広げ、世界トップレベルの研究実績を挙げていくことが期待される。NDB、DPC、介護データベースやその他のデータベースを含めて、医療ビッグデータ解析に精通した研究者を育成することは急務である。

本研究の全体の目的は以下の3つである。

- (i)医療ビッグデータ解析に精通する人材を養成するプログラムを開発すること
- (ii)上記(i)を通じて当該プログラムを実施できる体制を確立すること
- (iii)プログラム開発と同時にビッグデータ研究実践をさらに加速すること

H29 年度には、以下を行うことを目的とした。

- (1) 医療ビッグデータ研究の実践
- (2) ビッグデータ研究実践能力養成プログラム案の開発
- (3) ビッグデータハンドリング技術養成プログラム案の開発

- (1) の過程で実施した個別の教育プロセスを体系化・一般化し、(2) (3) の開発に活かすこととした。
- (2) については、日常臨床のクリニカル・クエスチョンを既存のビッグデータを用いて解明する研究実践能力を養成するプログラム案を開発することとした。
- (3) についてはデータハンドリング、データベースマネージメントに関する技術を養成する養成するプログラム案を開発することとした。

## B. 研究方法

### ◆研究を遂行するための体制

研究班は日本を代表する保健医療介護ビッグデータ研究者、統計学者、データサイエンティストで構成され、臨床家も参画している。東大・京大の NDB オンサイトセンターの運営にもあたっている。

NDB データを研究用のデータセットに転換する際の落とし穴の洗い出しや NDB データの処理方法などは、東大・京大オンサイトセンターでの研究で既に明らかにしている。先行の AMED 黒田班(康永が分担)で開発された一般向け NDB 活用 e-learning を、本研究で実用化するも計画している。

### ◆医療ビッグデータ研究の実践

総勢約 160 名の臨床家を動員し、日常臨床からクリニカル・クエスチョン (CQ) を紡ぎ出し、リサーチ・クエスチョン (RQ) に構造化し、ビッグデータを用いて解明する恒常的なサイクルを築き、英文原著論文を量産する体制を強化した。

これらの過程で実施した教育プロセスを体系化・一般化した。

### ◆若手研究者コンソーシアム

H29 度にはビッグデータ利活用のための若手研究者コンソーシアムを立ち上げ、疫学・統計学・医療情報学、臨床医学、健康科学・看護学などの若手研究者も多数参画させ、そこで開発される個別技術(データハンドリング技術、観察研究における統計解析技術など)を体系化・一般化した。

### ◆人材育成プログラム案の作成

上記の医療ビッグデータ研究の実践および若手研究者コンソーシアムでの実践を通じて得た個別の知見や技

術を体系化・一般化し、「ビッグデータ研究実践能力養成プログラム案」、「ビッグデータハンドリング技術養成プログラム案」を開発した。

◆プログラム評価判定の計画  
人材育成プログラム案の策定に加えて、当該プログラムの評価を実施するための評価判定方法について検討した。

## C. 研究結果

### 1. 全体の流れ

(1) 医療ビッグデータ研究の実践  
多数の若手研究者を指導し、医療ビッグデータ研究で H29 年度以降に 58 編の英文原著論文が掲載・受理された。

その過程で実施した教育プロセスを体系化・一般化し、以下の (2)

(3) に活かした。

(2) ビッグデータ研究実践能力養成プログラム案の開発

日常臨床の CQ を既存のビッグデータを用いて解明する研究実践能力を養成するプログラム案を策定した。

(3) ビッグデータハンドリング技術養成プログラム案の開発

データハンドリング、データベースマネジメントに関する技術を養成するプログラム案を策定した。

上記のプログラム案は、平成 30 年度中に短期集中セミナーを実践し、その効果を検証し、必要な修正を加えることとした。

具体的には、日本臨床疫学会との共催で、平成 30 年 8 月 6 日—8 月 10 日に東京大学で一週間短期集中セミナーを開催し、全国から参加者を募り、開発したプログラムを一般向け

に公開・実践することとした。(別添資料 1、2 を参照)

短期的効果(受講生の知識・技術の向上)の検証のために、受講者に試験・ハンズオン課題・研究プロトコル作成を課し、それらを検査することとする。

また、平成 30 年 9 月 30 日に京都大学で開催される日本臨床疫学会においてシンポジウムを行い、その成果を一般公開する予定である。

それらの成果を応用し、さらに本研究終了後に実施予定のビッグデータ人材育成の長期履修プログラムを策定する。

### 2. 個別の具体的成果

(1) 医療ビッグデータ研究の実践  
NDB、DPC、JMDC、介護給付実態調査データなどを用いて、H29 年度以降に当研究班から英文誌に掲載・受理された研究は 58 編であった。それらは「G.研究発表」に示すとおりである。このうち NDB は文献番号 58 の 1 編であり、NDB で IF5 を超えるメジャー・ジャーナルに受理された最初の例である。

(2) ビッグデータ研究実践能力養成プログラム案

1) ビッグデータ研究デザイン  
ビッグデータ研究デザインについては、以下の重点 6 教育項目が挙げられた。

- (i) 日常臨床から CQ を生み出す
- (ii) CQ→RQ の構造化
- (iii) 観察研究のデザイン
- (iv) 内的妥当性と外的妥当性、系統誤差と偶然誤差
- (v) 統計的因果推論
- (vi) ヘルスサービスリサーチ

これらの教育を実践するために、講義形式とワークショップによる双方向型教育によることとした。

8月の短期集中セミナーでも講義とワークショップを実践する。本研究終了後の長期履修プログラムの計画においては、これらの内容をさらに充実した長期の講義とワークショップが中心となると見込まれる。

## 2) データベースの理解

データベースの理解については、以下の重点2教育項目が挙げられた。

(i)各データベースの構造およびコーディングの理解

(ii) データ抽出依頼書の作成

これらの教育を実践するために、8月の短期集中セミナーでは、講義形式および e-learning による対話型教育

(AMED 黒田班で開発した教材を活用) とすることとした。

短期集中セミナーではさすがに演習形式では時間が不足するために導入しないこととした。本研究終了後の長期履修プログラムの計画においては、長期の講義と演習が中心となると見込まれる。

## 3) ビッグデータ統計学

ビッグデータを分析するための統計技術については、以下の重点3教育項目が挙げられた。

(i)SPSS, STATA, R を用いた統計分析

(ii)観察研究の統計手法(傾向スコア、操作変数法、Difference in Difference、Regression discontinuity、時間依存性交絡、自己対照デザイン、など)

(iii)不十分なデータの下でのリスク補正、不完全追跡、欠損値補完

これらの教育を実践するために、8月の短期集中セミナーでは、講義形式

およびハンズオンによる実践型教育とすることとした。

なお短期集中セミナーでは時間的制約のため、講義は傾向スコア、操作変数法、時間依存性交絡にしぼり、ハンズオンは傾向スコア分析、および R の基礎に絞ることとした。

本研究終了後の長期履修プログラムの計画においては、全項目を網羅する長期の講義とハンズオンが中心となると見込まれる。

## 4) 論文の書き方

ビッグデータ研究における論文の書き方については、以下の重点2教育項目が挙げられた。

(i)STROBE、RECORD に沿った論文の記載

(ii)ビッグデータ研究の背景・方法・結果・考察の記載

これらの教育を実践するために、講義形式と演習による双方向型教育によることとした。

なお短期集中セミナーでは時間的制約のため、講義のみとした。本研究終了後の長期履修プログラムの計画においては、長期の講義と演習が中心となると見込まれる。

## (3) ビッグデータハンドリング技術養成プログラム案

データベースハンドリングについては、以下の重点5教育項目が挙げられた。

(i)研究計画からデータ抽出・解析までの流れの理解

(ii) NDB の落とし穴の理解

(iii)データ抽出依頼書に沿ったデータベースからのデータ抽出

(iv)患者・エピソード・実施日単位のデータマートの作成

(v)SQL、Python などの習熟

これらの教育を実践するために、講義形式およびハンズオンによる実践型教育とすることとした。

8月の短期集中セミナーでは、時間的制約のため講義を主体とし、一部ハンズオンを導入することとした。具体的には、NDB、DPC、JMDC、介護データなど各データベースの概要に関する講義に加え、初級者向けのSQLによるデータ抽出技術の基礎をハンズオンで教育することとした。また、中級以上者向けの講義として、UNIXやGitの知識を前提にNDBをOracle Databaseの仮想環境でSQLを用いて処理する作業についての講義を用意した。さらにプログラミング言語Pythonの基本についての講義を用意した。

本研究終了後の長期履修プログラムの計画においては、全項目を網羅する長期の講義と演習が中心となると見込まれる。

#### (4) プログラム評価判定の計画 ＜短期的評価＞

8月の短期集中セミナーにおけるプログラム評価判定方法は以下のとおりとする。

- 1) NDB データハンドリングのe-learningによるテスト
- 2) ハンズオンセミナーにおける課題の達成度
  - ①統計ハンズオンセミナーにおけるデータ解析課題
  - ②SQLハンズオンセミナーにおけるデータ抽出課題
- 3) レポート
  - ①ビッグデータを用いた研究計画書の作成

仮説の立案、対象者の選定、要因・アウトカムのデータの定義、統計解析計画、利用するデータベースとの

親和性、研究の限界、実施可能性、新規性、独創性、臨床的・政策的含意、などをチェック

#### ②データ抽出依頼書の作成

データベースの構造およびコーディング、計画書との整合性などをチェック

#### ③プログラム内容に関する意見聴取（質問票調査）

#### ＜中期的評価＞

受講者が実際に研究に実施・参画し、論文出版の成果を挙げることが、中期的な評価の指標とすることとした。

## D. 考察

H29年度における多数のビッグデータ研究・論文執筆を通して用いられてきた個別技術（データハンドリング技術、観察研究における統計解析技術、など）を体系化・一般化し、既存の知識（NDBの落とし穴等）と合わせて、種々のビッグデータに応用可能な人材育成プログラム案を平成29年度に開発した。

これをH30年度に実施し、その効果を検証し、必要な修正を加える。プログラムの完成度をさらに上げるとともに、H30年度も継続的にビッグデータ研究を実践し、その過程を通じて得られた知識・技術をもとに、さらに長期履修プログラムに開発する。また、今回のプログラムそのものの短期的効果を評価するとともに、受講者の知識・技術向上を通じてさらに論文発表実績が上がっていくことをもって中期的評価とする。

研究期間中に実践・検証する短期集中セミナー等々の成果は、研究終了

後に長期履修プログラムへの展開を想定している。

## E. 結論

本研究は、これまでわが国の臨床研究・ビッグデータ研究において決定的に不足している「人材育成」という視点を最重要視し、これまで各研究者によって散発的・断片的に行われてきたビッグデータ研究のための種々の技術を一般化し体系的なプログラムを構築した上でその社会実装を行い、広く一般に公開・普及するという点で独創的である。

## F. 健康危険情報

なし

## G. 研究発表

### 1.論文発表

1. Wada T, Yasunaga H, Doi K, Matsui H, Fushimi K, Kitsuta Y, Nakajima S. Impact of hospital volume on mortality in patients with severe torso injury. *J Surg Res.* 2018;222:1-9.
2. Usui T, Hanafusa N, Yasunaga H, Nangaku M. Association of dialysis with in-hospital disability progression and mortality in community-onset stroke. *Nephrology (Carlton).* 2018 in press
3. Tsuchiya A, Yasunaga H, Tsutsumi Y, Matsui H, Fushimi K. Mortality and Morbidity After Hartmann's Procedure Versus Primary Anastomosis Without a Diverting Stoma for Colorectal Perforation: A Nationwide Observational Study. *World J Surg.* 2018;42(3):866-75.
4. Suzuki S, Yasunaga H, Matsui H, Fushimi K, Ando M, Yamasoba T. Postoperative mechanical bowel obstruction after pharyngolaryngectomy for hypopharyngeal cancer: Retrospective analysis using a Japanese inpatient database. *Head Neck.* 2018 in press.
5. Shinkawa H, Yasunaga H, Hasegawa K, Matsui H, Fushimi K, Michihata N, Kokudo N. Mortality and morbidity after hepatic resection in patients undergoing hemodialysis: analysis of a national inpatient database in Japan. *Surgery.* 2018 in press.
6. Sawada Y, Sasabuchi Y, Nakahara Y, Matsui H, Fushimi K, Haga N, Yasunaga H. Early Rehabilitation and In-Hospital Mortality in Intensive Care Patients With Community-Acquired Pneumonia. *Am J Crit Care.* 2018;27(2):97-103.
7. Okubo Y, Morisaki N, Michihata N, Matsui H, Fushimi K, Yasunaga H. Dose-dependent relationships between weight status and clinical outcomes among infants hospitalized

- with respiratory syncytial virus infections. *Pediatr Pulmonol.* 2018;53(4):461-6.
8. Okubo Y, Michihata N, Uda K, Morisaki N, Miyairi I, Matsui H, Fushimi K, Yasunaga H. Dose-response relationship between weight status and clinical outcomes in pediatric influenza-related respiratory infections. *Pediatr Pulmonol.* 2018;53(2):218-23.
  9. Okubo Y, Michihata N, Morisaki N, Uda K, Miyairi I, Ogawa Y, Matsui H, Fushimi K, Yasunaga H. Recent trends in practice patterns and impact of corticosteroid use on pediatric *Mycoplasma pneumoniae*-related respiratory infections. *Respir Investig.* 2018;56(2):158-65.
  10. Okubo Y, Michihata N, Morisaki N, Hangai M, Matsui H, Fushimi K, Yasunaga H. Recent trends in practice patterns and comparisons between immunoglobulin and corticosteroid in pediatric immune thrombocytopenia. *Int J Hematol.* 2018;107(1):75-82.
  11. Oichi T, Oshima Y, Chikuda H, Ohya J, Matsui H, Fushimi K, Tanaka S, Yasunaga H. In-hospital complication rate following microendoscopic versus open lumbar laminectomy: a propensity score-matched analysis. *Spine J.* 2018 in press.
  12. 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 in press.
  13. Koizumi C, Michihata N, Matsui H, Fushimi K, Yasunaga H. In-Hospital Mortality for Hepatic Portal Venous Gas: Analysis of 1590 Patients Using a Japanese National Inpatient Database. *World J Surg.* 2018;42(3):816-22.
  14. Isogai T, Matsui H, Tanaka H, Kohyama A, Fushimi K, Yasunaga H. Clinical features and peripartum outcomes in pregnant women with cardiac disease: a nationwide retrospective cohort study in Japan. *Heart Vessels.* 2018 in press.
  15. Isogai T, Matsui H, Tanaka H, Fushimi K, Yasunaga H. Clinical characteristics of patients with Takotsubo syndrome diagnosed without coronary artery evaluation: A retrospective nationwide study. *J Cardiol.* 2018;71(3):268-76.

16. Funakoshi H, Matsui H, Fushimi K, Yasunaga H. Variation in Patient Backgrounds, Practice Patterns, and Outcomes of High-Risk Pulmonary Embolism in Japan. *Int Heart J*. 2018 in press.
17. Yamana H, Moriwaki M, Horiguchi H, Kodan M, Fushimi K, Yasunaga H. Validity of diagnoses, procedures, and laboratory data in Japanese administrative data. *J Epidemiol*. 2017;27(10):476-82.
18. Wada T, Yasunaga H, Yamana H, Matsui H, Matsubara T, Fushimi K, Nakajima S. Development and validation of a new ICD-10-based trauma mortality prediction scoring system using a Japanese national inpatient database. *Inj Prev*. 2017;23(4):263-7.
19. Wada T, Yasunaga H, Yamana H, Matsui H, Fushimi K, Morimura N. Development and validation of an ICD-10-based disability predictive index for patients admitted to hospitals with trauma. *Injury*. 2018; 49(3):556-563
20. Wada T, Yasunaga H, Doi K, Matsui H, Fushimi K, Kitsuta Y, Nakajima S. Relationship between hospital volume and outcomes in patients with traumatic brain injury: A retrospective observational study using a national inpatient database in Japan. *Injury*. 2017;48(7):1423-31.
21. Urushiyama H, Jo T, Yasunaga H, Yamauchi Y, Matsui H, Hasegawa W, Takeshima H, Hiraishi Y, Mitani A, Fushimi K, Nagase T. Adjuvant chemotherapy versus chemoradiotherapy for small cell lung cancer with lymph node metastasis: a retrospective observational study with use of a national database in Japan. *BMC Cancer*. 2017;17(1):613.
22. Uda K, Matsui H, Fushimi K, Yasunaga H. Preoperative short-term plus postoperative physical therapy versus postoperative physical therapy alone for patients undergoing lung cancer surgery: retrospective analysis of a nationwide inpatient database. *Eur J Cardiothorac Surg*. 2018 in press.
23. Takeuchi M, Yasunaga H, Matsui H, Fushimi K. Pediatric urolithiasis associated with acute gastroenteritis: an inpatient database study in Japan. *Eur J Pediatr*. 2017;176(4):501-7.
24. Tagami T, Matsui H, Moroe Y, Fukuda R, Shibata A, Tanaka C, Unemoto K, Fushimi K, Yasunaga H. Antithrombin use and 28-day in-hospital mortality among severe-burn patients: an observational nationwide



- study. *Ann Intensive Care*. 2017;7(1):18.
25. Sugihara T, Yasunaga H, Matsui H, Nagao G, Ishikawa A, Fujimura T, Fukuhara H, Fushimi K, Ohori M, Homma Y. Accessibility to surgical robot technology and prostate-cancer patient behavior for prostatectomy. *Jpn J Clin Oncol*. 2017;47(7):647-51.
  26. Shakya S, Matsui H, Fushimi K, Yasunaga H. In-hospital complications after implantation of cardiac implantable electronic devices: Analysis of a national inpatient database in Japan. *J Cardiol*. 2017;70(5):405-10.
  27. Sato M, Tateishi R, Yasunaga H, Matsui H, Horiguchi H, Fushimi K, Koike K. Mortality and hemorrhagic complications associated with radiofrequency ablation for treatment of hepatocellular carcinoma in patients on hemodialysis for end-stage renal disease: A nationwide survey. *J Gastroenterol Hepatol*. 2017;32(11):1873-8.
  28. Sasabuchi Y, Yasunaga H, Matsui H, Lefor AK, Fushimi K, Sanui M. Epidural analgesia is infrequently used in patients with acute pancreatitis : a retrospective cohort study. *Acta Gastroenterol Belg*. 2017;80(3):381-4.
  29. Sako A, Yasunaga H, Matsui H, Fushimi K, Hamasaki H, Katsuyama H, Tsujimoto T, Goto A, Yanai H. Hospitalization with hypoglycemia in patients without diabetes mellitus: A retrospective study using a national inpatient database in Japan, 2008-2012. *Medicine (Baltimore)*. 2017;96(25):e7271.
  30. Sakamoto Y, Yamauchi Y, Yasunaga H, Takeshima H, Hasegawa W, Jo T, Sasabuchi Y, Matsui H, Fushimi K, Nagase T. Development of a nomogram for predicting in-hospital mortality of patients with exacerbation of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*. 2017;12:1605-11.
  31. Sakamoto Y, Yamauchi Y, Yasunaga H, Takeshima H, Hasegawa W, Jo T, Matsui H, Fushimi K, Nagase T. Guidelines-concordant empiric antimicrobial therapy and mortality in patients with severe community-acquired pneumonia requiring mechanical ventilation. *Respir Investig*. 2017;55(1):39-44.
  32. Okubo Y, Michihata N, Yoshida K, Morisaki N, Matsui H, Fushimi K, Yasunaga H. Impact of pediatric obesity on acute asthma exacerbation in Japan. *Pediatr Allergy Immunol*.

- 2017;28(8):763-7.
33. Okubo Y, Michihata N, Morisaki N, Sundel RP, Matsui H, Fushimi K, Yasunaga H. Association between dose of glucocorticoids and coronary artery lesions in Kawasaki disease. *Arthritis Care Res (Hoboken)*. 2018 in press.
  34. Oichi T, Chikuda H, Ohya J, Ohtomo R, Morita K, Matsui H, Fushimi K, Tanaka S, Yasunaga H. Mortality and morbidity after spinal surgery in patients with Parkinson's disease: a retrospective matched-pair cohort study. *Spine J*. 2017;17(4):531-7.
  35. Ohya J, Chikuda H, Oichi T, Kato S, Matsui H, Horiguchi H, Tanaka S, Yasunaga H. Seasonal Variations in the Risk of Reoperation for Surgical Site Infection Following Elective Spinal Fusion Surgery: A Retrospective Study Using the Japanese Diagnosis Procedure Combination Database. *Spine (Phila Pa 1976)*. 2017;42(14):1068-79.
  36. Odagiri H, Yasunaga H, Matsui H, Matsui S, Fushimi K, Kaise M. Hospital volume and adverse events following esophageal endoscopic submucosal dissection in Japan. *Endoscopy*. 2017;49(4):321-6.
  37. Odagiri H, Yasunaga H. Complications following endoscopic submucosal dissection for gastric, esophageal, and colorectal cancer: a review of studies based on nationwide large-scale databases. *Ann Transl Med*. 2017;5(8):189.
  38. Nagata N, Yasunaga H, Matsui H, Fushimi K, Watanabe K, Akiyama J, Uemura N, Niikura R. Therapeutic endoscopy-related GI bleeding and thromboembolic events in patients using warfarin or direct oral anticoagulants: results from a large nationwide database analysis. *Gut*. 2018 epub ahead of print.
  39. Nagano M, Ichinose J, Sasabuchi Y, Nakajima J, Yasunaga H. Surgery versus percutaneous transcatheter embolization for pulmonary arteriovenous malformation: Analysis of a national inpatient database in Japan. *J Thorac Cardiovasc Surg*. 2017;154(3):1137-43.
  40. Morita K, Matsui H, Yamana H, Fushimi K, Imamura T, Yasunaga H. Association between advanced practice nursing and 30-day mortality in mechanically ventilated critically ill patients: A retrospective cohort study. *J Crit Care*. 2017;41:209-15.
  41. Morita K, Matsui H, Fushimi K, Yasunaga H. Association between Nurse Staffing and In-Hospital Bone Fractures: A Retrospective Cohort

- Study. *Health Serv Res.* 2017;52(3):1005-23.
42. Matsui H, Jo T, Fushimi K, Yasunaga H. Outcomes after early and delayed rehabilitation for exacerbation of chronic obstructive pulmonary disease: a nationwide retrospective cohort study in Japan. *Respir Res.* 2017;18(1):68.
  43. Maeda T, Sasabuchi Y, Matsui H, Ohnishi Y, Miyata S, Yasunaga H. Safety of Tranexamic Acid in Pediatric Cardiac Surgery: A Nationwide Database Study. *J Cardiothorac Vasc Anesth.* 2017;31(2):549-53.
  44. Kishimoto M, Yamana H, Inoue S, Noda T, Myojin T, Matsui H, Yasunaga H, Kawaguchi M, Imamura T. Sivelestat sodium and mortality in pneumonia patients requiring mechanical ventilation: propensity score analysis of a Japanese nationwide database. *J Anesth.* 2017;31(3):405-12.
  45. Kawata M, Sasabuchi Y, Inui H, Taketomi S, Matsui H, Fushimi K, Chikuda H, Yasunaga H, Tanaka S. Annual trends in knee arthroplasty and tibial osteotomy: Analysis of a national database in Japan. *Knee.* 2017;24(5):1198-205.
  46. Jo T, Yasunaga H, Sasabuchi Y, Michihata N, Morita K, Yamauchi Y, Hasegawa W, Takeshima H, Sakamoto Y, Matsui H, Fushimi K, Nagase T. Association between dementia and discharge status in patients hospitalized with pneumonia. *BMC Pulm Med.* 2017;17(1):128.
  47. Iwagami M, Yasunaga H, Matsui H, Horiguchi H, Fushimi K, Noiri E, Nangaku M, Doi K. Impact of end-stage renal disease on hospital outcomes among patients admitted to intensive care units: A retrospective matched-pair cohort study. *Nephrology (Carlton).* 2017;22(8):617-23.
  48. Isogai T, Yasunaga H, Matsui H, Tanaka H, Hisagi M, Fushimi K. Factors affecting in-hospital mortality and likelihood of undergoing surgical resection in patients with primary cardiac tumors. *J Cardiol.* 2017;69(1):287-92.
  49. Isogai T, Matsui H, Tanaka H, Yokogawa N, Fushimi K, Yasunaga H. Treatments and in-hospital mortality in acute myocardial infarction patients with rheumatoid arthritis: a nationwide retrospective cohort study in Japan. *Clin Rheumatol.* 2017;36(5):995-1004.
  50. Isogai T, Matsui H, Tanaka H, Fushimi K, Yasunaga H. In-hospital

- management and outcomes in patients with peripartum cardiomyopathy: a descriptive study using a national inpatient database in Japan. *Heart Vessels*. 2017;32(8):944-51.
51. Isogai T, Matsui H, Tanaka H, Fushimi K, Yasunaga H. Seasonal variation in patient characteristics and in-hospital outcomes of Takotsubo syndrome: a nationwide retrospective cohort study in Japan. *Heart Vessels*. 2017;32(10):1271-6.
  52. Hasegawa W, Yamauchi Y, Yasunaga H, Takeshima H, Sakamoto Y, Jo T, Sasabuchi Y, Matsui H, Fushimi K, Nagase T. Prognostic nomogram for inpatients with asthma exacerbation. *BMC Pulm Med*. 2017;17(1):108.
  53. Aso S, Matsui H, Fushimi K, Yasunaga H. Resuscitative endovascular balloon occlusion of the aorta or resuscitative thoracotomy with aortic clamping for noncompressible torso hemorrhage: A retrospective nationwide study. *J Trauma Acute Care Surg*. 2017;82(5):910-4.
  54. Okubo Y, Michihata N, Morisaki N, Kinoshita N, Miyairi I, Urayama KY, Yasunaga H. Recent patterns in antibiotic use for children with group A streptococcal infections. *Journal of Global Antimicrobial Resistance* 2017;13;13:55-59.
  55. Ono S, Ono Y, Matsui H, Yasunaga H. Impact of clinic follow-up visits on body weight control in people with prediabetes or diabetes mellitus: Japanese nonelderly cohort study. *Family Practice* 2017;34(5):552-557
  56. Ono S, Ono Y, Matsui H, Yasunaga H. Enhanced Oral Care and Health Outcomes among Nursing Facility Residents: Analysis Using the National Long-term Care Database in Japan. *Journal of the American Medical Directors Association* 2017;18(3):277.e1-277.e5
  57. Morita K, Ono S, Ishimaru M, Matsui H, Naruse T, Yasunaga H. Factors affecting discharge to home among geriatric intermediate care facility residents in Japan. *Journal of the American Geriatrics Society* 2018 in press
  58. Ishimaru M, Matsui H, Ono S, Hagiwara Y, Morita K, Yasunaga H. Preoperative oral care by a dentist and postoperative complications after major cancer surgery: a nationwide administrative claims database study. *British Journal of Surgery* 2018 in press.

## 2.学会発表

1. 岡田 寛之, 松本 卓巳, 道端 伸明, 小林 寛, 松原 全宏, 廣瀬 旬, 康永 秀生, 田中 栄.病的骨折の治療を考える 骨転移に伴う病的骨折に対する手術はADLを改善するか DPCデータベースを用いた検討.骨折 39 卷 Suppl. Page S26(2017.07)
2. 杉原 亨, 康永 秀生, 松居 宏樹, 石川 晃, 本間 之夫.全国ビックデータからみる高齢者悪性腫瘍手術の術後ADL低下リスク.日本泌尿器科学会総会 105 回 Page SKH02-2(2017.04)
3. 津田 祐輔, 康永 秀生, 小林 寛, 篠田 裕介, 澤田 良子, 池上 政周, 平井 利英, 河野 博隆.骨転移による大腿骨病的骨折患者の術後合併症発生率とその危険因子 DPC データベースによる検討.日本整形外科学会雑誌 91 卷 6 号 Page S1404(2017.06)
4. 木下 善隆, 杉原 亨, 康永 秀生, 松居 宏樹, 南 哲司, 山田 幸央, 石橋 由孝, 石川 晃, 本間 之夫.腹膜透析カテーテル留置術の周術期成績と年間手術件数の関連 DPC データを用いた 2505 例の検討.日本透析医学会雑誌 50 卷 Suppl.1 Page767(2017.05)
5. 平石 尚久, 城 大祐, 長谷川 若恵, 坂本 幸世, 竹島 英之, 漆山 博和, 三谷 明久, 山内 康宏, 長瀬 隆英, 康永 秀生.DPC データを用いた肺癌に対しての診断的気管支鏡の周術期死亡率の検討.気管支学 39 卷 Suppl. Page S222(2017.05)
6. 加藤源太, 酒井未知, 大寺祥佑, 下垣 徹, 松居宏樹, 野田龍也, 康永秀生, 今村知明, 黒田知宏.新たなエビデンス創出のための次世代 NDB データ研究基盤構築に関する研究:疫学研究への活用可能性について.日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日.
7. 田上隆, 康永秀生. DPC データを臨床疫学研究に活用するためのアプリケーションの開発. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
8. 本田千可子, 山名隼人, 松居宏樹, 康永秀生, 永田智子. 乳児の外傷・異物・火傷による外来受診の時期. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
9. 大谷隼一, 筑田博隆, 尾市健, 田中栄, 松居宏樹, 康永秀生. 脊椎固定術の術後創部感染リスクは季節によって変動する: DPC データベース研究日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
10. 山本博之, 橋本英樹, 松居宏樹, 康永秀生. DPC データベースを用いた腹膜透析関連腹膜炎発症率の推計. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
11. 田所史野, 森田光治良, 道端伸明, 伏見清秀, 康永秀生. 小児におけるスガ

- マデクスとアナフィラキシーの関連：DPC データを用いた nested casecontrol study. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
12. 宇田和晃、康永秀生. 大腿骨頸部骨折術後の歩行自立獲得と術後リハビリテーション介入との関連. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
  13. 松居宏樹、康永秀生. レセプトデータベースを用いた心不全症例における入院中リハビリテーションの効果の検証. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
  14. 平塚義宗、道端伸明、城大祐、春日俊光、小野浩一、村上晶、康永秀生. 介護施設入所者の白内障手術アウトカムの検討. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
  15. 麻生将太郎、松居宏樹、伏見清秀、康永秀生. 特発性肺線維症急性増悪に対するサイクロフォスファミドと副腎皮質ホルモン大量療法の併用療法の効果の検討. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
  16. 尾市健、大島寧、筑田博隆、大谷隼一、松居宏樹、伏見清秀、田中栄、康永秀生. 内視鏡手術は腰椎椎弓切除術の周術期リスクを減らさうるか？—propensity scorematched analysis—. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
  17. 道端伸明、城大祐、康永秀生. 妊娠前母体 Body mass index と児のアウトカムとの関連：診療報酬請求データを用いた解析. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
  18. 笹渕裕介、松居宏樹、Alan Lefor、伏見清秀、康永秀生. 高齢者に置ける股関節骨折に対する手術タイミングと短期予後. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
  19. 菅沼大、笹渕裕介、松居宏樹、伏見清秀、康永秀生. 大腿骨骨折術後の活動型せん妄発症率に関するデスフルランとセボフルランの比較：傾向スコア法を用いた解析. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
  20. 山名隼人、松居宏樹、野田龍也、伏見清秀、今村知明、康永秀生. ハイリスク分娩における母体合併症発生と施設および地域要因の関連. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
  21. 阿部博昭、住谷昌彦、内田寛治、松居宏樹、伏見清秀、康永秀生、山田芳嗣. 予定帝王切開における麻酔法と母体重症合併症の関係 —DPC データを用いた populationbased study—. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
  22. 大野幸子、大野洋介、松居宏樹、康永秀生. 健診時糖尿病受診勧奨者にお

- ける医療機関受診と Body Mass Index 変化の関連. 日本臨床疫学会第 1 回年次学術大会. 2017 年 9 月 30 日
23. 和田智貴、康永秀生、山名隼人、森村尚登. ICD-10 コーディングを用いた外傷患者の重度身体機能障害予測スコアモデルの作成と妥当性の検討. 日本臨床疫学会第 1 回年次学術大会. 2017 年 10 月 1 日
24. 森田光治良、松居宏樹、道端伸明、伏見清秀、康永秀生. ステイブンス・ジョンソン症候群 (SJS) / 中毒性表皮壊死症 (TEN) に対するステロイド療法の有効性. 日本臨床疫学会第 1 回年次学術大会. 2017 年 10 月 1 日
25. 土谷飛鳥、山名隼人、川原拓也、堤悠介、松居宏樹、康永秀生. 重症熱傷患者に対する気管切開術と死亡との関連; 周辺化構造モデルを用いた検討. 日本臨床疫学会第 1 回年次学術大会. 2017 年 10 月 1 日
26. 岸本美和、山名隼人、野田龍也、松居宏樹、康永秀生、今村知明. 安定冠動脈疾患患者に対する PCI 術前の虚血検査施行率の評価と、未施行の背景要因の検討. 日本臨床疫学会第 1 回年次学術大会. 2017 年 10 月 1 日
27. 杉原亨、康永秀生、松居宏樹、伏見清秀、本間之夫. 手術支援ロボットの近隣の出現が非ロボット前立腺癌手術に与える影響 - 手術日間隔を 頻度指標とする試み. 日本臨床疫学会第 1 回年次学術大会. 2017 年 10 月 1 日.
28. 石丸美穂、大野幸子、松居宏樹、康永秀生. 高次元傾向スコアと従来の傾向スコアマッチングの比較: 周術期口腔機能管理の術後肺炎予防の効果分析. 日本臨床疫学会第 1 回年次学術大会. 2017 年 10 月 1 日

## 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

## 別添資料 1

### NDB・DPC データベース研究人材育成<短期集中セミナー>のご案内

保健医療介護のデータベース研究に必要な知識と技能を学びませんか？

#### ◆開催概要

---

名称	NDB・DPC データベース研究人材育成<短期集中セミナー>		
期間	2018年8月6日(月)～8月10日(金) 13:00～17:50 (6日のみ 16:50で終了)		
場所	東京都文京区本郷7-3-1 東京大学医学部2号館(本館)3階大講堂		
主催・共 催	主催：厚生労働省科学研究・保健医療介護ビッグデータ人材育成研 究班 (研究代表者：東京大学 康永秀生) 共催：日本臨床疫学会		
申し込み 期間	日本臨床疫学会	<u>2018年5月21日(月)正午～</u>	<u>先行予約開始</u>
	会員	<u>2018年6月4日(月)正午～</u>	<u>予約開始</u>
	非会員	<u>申込締切：2018年7月6日(金)18:00</u>	
料金	無料(事前登録制)		

---

#### ◆セミナー概要

近年、NDB・DPCなどの保健医療ビッグデータや介護ビッグデータを用いた大規模データベース研究が拡大しています。これらの研究をさらに発展させるためには、臨床医学・疫学・統計学の知識に加えて、データベースに関連する知識や技能に精通した人材の育成が必須です。

本セミナーでは、保健医療介護ビッグデータ研究に興味のある方々を対象とし、各種大規模データベースの概要や研究計画の立案、データハンドリング、統計解析、論文報告内容について短期集中の学習機会を提供します。



◆本セミナーの特徴

- ・演習やハンズオン形式により研究計画立案やデータハンドリング、統計解析を習得できる
- ・保健医療介護ビッグデータ研究で実績のある講師陣による講義・演習を受けることができる
- ・短期集中で大規模データベース研究の計画立案から統計解析まで学習できる

◆プログラム

・日程表

Time	8月6日 (月)	8月7日 (火)	8月8日 (水)	8月9日 (木)	8月10日 (金)
12:30- 13:00	受付	受付	受付	受付	受付
13:00- 13:50	NDB データの概要 (講義) 松居宏樹	介護データベース研究 (講義) 田宮菜奈子	DPC データを利用した研究レビュー(講義) 康永秀生	R を用いた統計解析基礎 I (ハンズオン) 笹渕裕介・松居宏樹	ビッグデータ研究における統計解析～傾向スコア分析(講義) 山名隼人
14:00- 14:50	NDB データの落とし穴(講義) 奥村泰之	JMDC データを利用した研究(講義) 康永秀生	DPC データの利用方法(講義) 康永秀生	R を用いた統計解析基礎 II (ハンズオン) 笹渕裕介・松居宏樹	SPSS を用いた傾向スコア分析(ハンズオン) 山名隼人・森田光治良

15:00-15:50	NDB利用のための e-learning(講義) 加藤源太	レセプトデータ利用のための SQL セミナー I (ハンズオン) 大野幸子・松居宏樹	データベース研究で求められる論文報告内容(講義) 森田光治 良・康永秀生	R を用いた統計解析基礎 III (ハンズオン) 笹渕裕介・松居宏樹	STATA を用いた傾向スコア分析(ハンズオン) 道端伸明・麻生将太郎
16:00-16:50	NDB を世の中の役に立てるには～利用者の立場で政策を前進させる方法～(講義) 吉村健佑	レセプトデータ利用のための SQL セミナー II (ハンズオン) 大野幸子・松居宏樹	レセプトデータベース研究計画立案 I (演習) 康永秀生・城大佑、他	Python の基礎 (講義) 笹渕裕介	ビッグデータ研究における統計解析～操作変数法(講義) 篠崎智大
17:00-17:50		レセプトデータ利用のための SQL セミナー III (ハンズオン) 大野幸子・松居宏樹	レセプトデータベース研究計画立案 II (演習) 康永秀生・城大佑、他	Oracle SQL を用いた NDB からのデータ抽出 (講義) 原湖楠	ビッグデータ研究における統計解析～時間依存性交絡(講義) 松山裕

・各プログラムの詳細

詳細は[こちら](#)に掲載。←リンク 1<サマーセミナー企画説明>

・定員

演習・ハンズオン：各 30 名、 講義：各 200 名

◆応募要件

保健医療介護ビッグデータ研究に興味のある方。

- ・医療・介護従事者
  - ・大学関係者
  - ・各種研究機関に勤務する研究者
  - ・企業に所属する者
  - ・医療政策の担当者
- など

◆受講費

無料

◆受講申し込みの注意事項

- ・すべて事前登録制になります。

※日本臨床疫学会の会員は先行申し込みが可能です。

- ・講義の受講申し込みについて

事前登録のみですべての講義が受講できます(プログラムを登録時に指定する必要はありません)。定員に達した次第、事前申し込みを締め切らせていただきますので、予めご了承ください。

- ・演習とハンズオンの受講申し込みについて

受講するプログラムを指定してお申し込みください(複数のプログラムを申し込みできます)。

各プログラムは先着順となります、予めご了承ください。当日受講可能な方のみお申込みください。

できるだけ多くの方に参加していただくため、お申し込み後に受講できなくな

った方は、7月27日（金）18：00までに参加登録の修正(キャンセル)を行って  
ください。

お申し込みいただいたにもかかわらず参加登録の修正(キャンセル)を行わずに当日欠席された方は、次回以降同様のセミナー開催時に、演習とハンズオンのお申し込みをお断りすることがありますので、予めご了承ください。

以下のプログラムは複数の単元でセットとなります(すべて受講可能な方のみ申し込み可能です)。

レセプトデータ利用のための SQL セミナー I-III
レセプトデータベース研究計画立案 I-II
R を用いた統計解析基礎 I-III

・セミナーのプログラム評価を目的として、各講義の後に無記名によるアンケートあるいは理解度テストを実施することがありますので、アンケート用紙や答案の回収にご協力をお願いします。

・演習・ハンズオンでは課題を課されることがありますので、講師の指示に従い課題の提出（無記名）をお願いします。

#### ◆受講申し込み

受講を申し込む

#### ◆お問い合わせ

東京大学大学院医学系研究科公共健康医学専攻 臨床疫学・経済学分野

113-0033 東京都文京区本郷 7-3-1 東京大学医学部 2号館本館 3階

E-mail: [clinepi\\_adm](mailto:clinepi_adm)★[www.heer.m.u-tokyo.ac.jp](http://www.heer.m.u-tokyo.ac.jp)

※ 送信時は★を半角アットマークに変更して下さい

**別添資料 2** サマーセミナー企画説明

**NDB・DPC データベース研究人材育成<短期集中セミナー>**

日程：2018年8月6日～2018年8月10日（月～金）

時間：13:00～17:50（一部 16:50 で終了）

場所：東京大学医学部本館大講堂

共催：日本臨床疫学会

**<お願い>**

・セミナーのプログラム評価を目的として、各講義の後に無記名によるアンケートあるいは理解度テストを実施することがありますので、アンケート用紙や答案の回収にご協力をお願いします。

・演習・ハンズオンでは課題を課されることがありますので、講師の指示に従い課題の提出（無記名）をお願いします。

1日目：2018年8月6日（月）	
13:00 ～ 13:50	<b>テーマ：NDBデータの概要(講義)</b> <b>&lt;講師&gt;東京大学大学院医学系研究科臨床疫学・経済学 助教 松居 宏樹</b> <b>&lt;目標&gt;</b> NDBデータに含まれる情報を理解する。 NDBデータの構造を理解する。 NDBデータの申請方法について理解する。 ビッグデータを用いる際に必要なサーバー構築方法の基礎を理解する。 データを安全に利用するために必要なデータ管理方法を理解する。 オンサイトリサーチセンターについて理解する。 <b>&lt;想定する対象者&gt;</b> NDBデータを利用したことがなく、今後利用を検討している方々 NDBデータを利用したことがあるものの、うまくいかずに悩んでいる方々
14:00 ～ 14:50	<b>NDBデータベース研究の落とし穴(講義)</b> <b>&lt;講師&gt;東京都医学総合研究所 奥村 泰之</b> <b>&lt;目標&gt;</b> NDBの落とし穴・欠点を克服するために行われている工夫を理解する。 <b>&lt;想定する対象者&gt;</b> NDBデータを利用したことがなく、今後利用を検討している方々 NDBデータを利用したことがあるものの、うまくいかずに悩んで

	いる方々
15:00 ~ 15:50	<p><b>NDB 利用のための e-learning (講義)</b></p> <p>&lt;講師&gt; 京都大学医学部附属病院診療報酬センター 准教授 加藤 源太</p> <p>&lt;目標&gt; NDB の研究利用を円滑化するための学習支援ツールである e-learning について利用方法や内容を理解する。</p> <p>&lt;想定する対象者&gt; NDB データを利用したことがなく、今後利用を検討している方々 NDB データを利用したことがあるものの、うまくいかずに悩んでいる方々</p>
16:00 ~ 16:50	<p><b>NDB を世の中の役に立てるには～利用者の立場で政策を前進させる方法～ (講義)</b></p> <p>&lt;講師&gt; 千葉大学医学部附属病院 病院経営管理学研究センター 特任講師 吉村 健佑</p> <p>&lt;目的・目標&gt; NDB の活用を検討する際に心がけることがわかる。 NDB オープンデータの活用推進など、NDB 制度と利用を広げるために重要な視点を理解する。</p> <p>&lt;想定する対象者&gt; NDB データを利用したことがなく、今後利用を検討している方々 NDB データを利用したことがあるものの、うまくいかずに悩んでいる方々</p>

<b>2 日目 : 2018 年 8 月 7 日 (火)</b>	
13:00 ~ 13:50	<p><b>介護データベース研究 (講義)</b></p> <p>&lt;講師&gt; 筑波大学医学医療系ヘルスサービスリサーチ分野 教授 田宮 菜奈子</p> <p>&lt;目標&gt; 介護診療報酬データベースについて理解する。 データベースに含まれる情報を理解する。 データベース利用申請について理解する。</p> <p>&lt;想定する対象者&gt; 介護データを利用したことがなく、今後利用を検討している方々 介護データを利用したことがあるものの、うまくいかずに悩んでいる方々</p>
14:00 ~ 14:50	<p><b>JMDC データと JMDC を利用した研究レビュー(講義)</b></p> <p>&lt;講師&gt; 東京大学大学院医学系研究科臨床疫学・経済学 教授 康永秀生</p> <p>&lt;目標&gt;</p>

	<p>JMDC データに含まれる情報を理解する。  JMDC データの構造を理解する。  JMDC データを用いた様々な先行研究について学ぶ。  &lt;想定する対象者&gt;  JMDC データを利用したことがなく、今後利用を検討している方々  JMDC データを利用したことがあるものの、うまくいかずに悩んでいる方々</p>
15:00 ~ 15:50 16:00 ~ 16:50 17:00 ~ 17:50	<p><b>レセプトデータ利用のための SQL セミナー I~III (ハンズオン)</b>  &lt;講師&gt;東京大学生物統計情報学 特任助教 大野 幸子  東京大学臨床疫学・経済学 助教 松居 宏樹</p> <p>&lt;目標&gt;  SQL 言語の基礎を理解する。  JMDC データベースから自分の研究デザインに応じたデータセットを作成する。  JMDC データのデータクリーニングを行う。  &lt;想定する対象者&gt;  JMDC データを利用したことがなく、今後利用を検討している方々  JMDC データを利用したことがあるものの、うまくいかずに悩んでいる方々  ※事前参加登録した 30 名に限定</p>

<b>3 日目 : 2018 年 8 月 8 日 (水)</b>	
13:00 ~ 13:50	<p><b>DPC データを利用した研究レビュー(講義)</b>  &lt;講師&gt;東京大学大学院医学系研究科臨床疫学・経済学  教授 康永秀生</p> <p>&lt;目標&gt;  DPC データに含まれる情報を理解する。  DPC データの構造を理解する。  DPC データを用いた様々な先行研究について学ぶ。  &lt;想定する対象者&gt;  DPC データを利用したことがなく、今後利用を検討している方々  DPC データを利用したことがあるものの、うまくいかずに悩んでいる方々</p>
14:00 ~ 14:50	<p><b>DPC データの利用方法(講義)</b>  &lt;講師&gt;東京大学大学院医学系研究科臨床疫学・経済学  教授 康永秀生</p> <p>&lt;目標&gt;  DPC データを利用した研究の方法論を理解する。  DPC データを利用した研究計画書、データ抽出依頼書の作成方法</p>

	<p>を理解する。          &lt;想定する対象者&gt;          DPC データを利用したことがなく、今後利用を検討している方々          DPC データを利用したことがあるものの、うまくいかずに悩んでいる方々</p>
15:00 ~ 15:50	<p><b>データベース研究で求められる論文報告内容(講義)</b>          &lt;講師&gt; 東京大学大学院医学系研究科臨床疫学・経済学          特任研究員 森田 光治良</p> <p>&lt;目標&gt;          観察研究、データベースを用いた研究で求められる報告のポイントを理解する。          STROBE, RECORD について理解する。          &lt;想定する対象者&gt;          各種の保健医療介護データベースを利用して論文発表を志している方々</p>
16:00 ~ 16:50 17:00 ~ 17:50	<p><b>レセプトデータベース研究計画立案 I・II (演習)</b>          &lt;講師&gt; 東京大学大学院医学系研究科臨床疫学・経済学          教授 康永秀生</p> <p>&lt;目標&gt;          NDB,DPC,JMDC レセプトデータベース研究に適した研究デザインを立案する。          講師が提示する 3-4 つ程度の CQ から一つを選び、データベースを利用した研究をデザインする。グループディスカッション形式で、CQ を RQ に落とし込み、FINER を検討し、発表を行う。          &lt;想定する対象者&gt;          各種の保健医療介護データベースを利用して論文発表を志している方々</p> <p><u>※事前参加登録した 30 名に限定</u>  <u>※上記の 30 名は、同日の講義「DPC データを利用した研究レビュー」「DPC データの利用方法」「データベース研究で求められる論文報告内容」への参加を必須とする。</u></p>

<b>4 日目 : 2018 年 8 月 9 日 (木)</b>	
13:00 ~ 13:50 14:00 ~ 14:50 15:00 ~ 15:50	<p><b>R を用いた統計解析基礎 I~III (ハンズオン)</b>          &lt;講師&gt; 自治医科大学データサイエンスセンター講師 笹渕 裕          介          東京大学臨床疫学・経済学 助教 松居宏樹</p> <p>&lt;目標&gt;          R の基礎的な使い方について、ハンズオンで演習する。          &lt;想定する対象者&gt;</p>



	R を使用したことがなく、これから R にチャレンジしたい方々。 ※事前参加登録した 30 名に限定
16:00 ~ 16:50	<b>Python の基礎 (講義)</b> <講師>自治医科大学データサイエンスセンター講師 笹渕 裕介 <目標> Python の様々な活用方法のデモを通じて、Python の機能を理解する。 Python を用いた機械学習について基本を理解する。 <想定する対象者> Python を使用したことがなく、これから Python にチャレンジしたい方々。
17:00 ~ 17:50	<b>Oracle SQL を用いた NDB からのデータ抽出 (講義)</b> <講師>東京大学公衆衛生学 原湖楠 <目標> NDB オンサイトセンターでデータ抽出する上で必要な作業について理解する。 データベースにアクセスするための SQL 言語、特に Oracle SQL の概要について理解する。 <想定する対象者> NDB データを利用したことがある方々、または本セミナーの関連する講義を受講する方々

**5 日目 : 2018 年 8 月 10 日(金)**

13:00 ~ 13:50	<b>ビッグデータ研究における統計解析～傾向スコア分析(講義)</b> <講師>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 特任助教 山名隼人 <目標> 傾向スコア分析の基礎理論を理解する。 <想定する対象者> 臨床研究・疫学研究の初級者以上
------------------	---

<p>14:00 ～ 14:50</p>	<p><b>SPSS を用いた傾向スコア分析(ハンズオン)</b>  <b>&lt;講師&gt;</b>東京大学ヘルスサービスリサーチ講座 山名隼人  東京大学臨床疫学・経済学 森田 光治良</p> <p><b>&lt;目標&gt;</b>  SPSS を用いた傾向スコア分析の実践的な方法を理解する。  <b>&lt;想定する対象者&gt;</b>  臨床研究・疫学研究の中級者以上  <u>※事前参加登録した 30 名に限定</u></p>
<p>15:00 ～ 15:50</p>	<p><b>STATA を用いた傾向スコア分析(ハンズオン)</b>  <b>&lt;講師&gt;</b>東京大学ヘルスサービスリサーチ講座 道端伸明  東京大学臨床疫学・経済学 麻生将太郎</p> <p><b>&lt;目標&gt;</b>  STATA を用いた傾向スコア分析の実践的な方法を理解する。  <b>&lt;想定する対象者&gt;</b>  臨床研究・疫学研究の中級者以上  <u>※事前参加登録した 30 名に限定</u></p>
<p>16:00 ～ 16:50</p>	<p><b>ビッグデータ研究における統計解析～操作変数法 (講義)</b>  <b>&lt;講師&gt;</b>東京大学大学院医学系研究科生物統計学  助教 篠崎智大</p> <p><b>&lt;目標&gt;</b>  近年、臨床研究・疫学研究でも利用される場面が増えている「操作変数法」について、概要を理解する。  <b>&lt;想定する対象者&gt;</b>  臨床研究・疫学研究の中級者以上</p>
<p>17:00 ～ 17:50</p>	<p><b>ビッグデータ研究における統計解析～時間依存性交絡 (講義)</b>  <b>&lt;講師&gt;</b>東京大学大学院医学系研究科生物統計学  教授 松山裕</p> <p><b>&lt;目標&gt;</b>  近年、臨床研究・疫学研究でも利用される場面が増えている「時間依存性交絡」とその対処法について、概要を理解する。  <b>&lt;想定する対象者&gt;</b>  臨床研究・疫学研究の中級者以上</p>

平成 29 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、また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】H29 年度には当研究班から、臨床医のクリニカルクエストにもとづく DPC データベースを用いた研究の成果が 53 編、英文誌に掲載された。その内 5 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、国際学会に 1 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】大規模データベースを用いた研究における各ステップ(クリニカルクエストの構造化、データの抽出、解析、論文または学会報告)の全て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の専門家が効率よくサポートを行うには、臨床医に対する教育のための人材育成プログラムの確

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トとして持ち寄ったテーマを、リサーチクエストに構造化した。構造化されたリサーチクエスト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2015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) 増悪入院した慢性閉塞性肺疾患患者の在院死亡リスクを予測するためのノモグラムの作成

【背景】COPD患者はしばしば増悪を起こし、入院治療を必要とすることもあり死亡率の増加にも関連している。これまでにCOPD増悪による死亡率については報告されてきたが、個々の在院死亡率を予測するには情報が限られていた。そのため、本研究では日本の入院患者データベースを用いて入院時の患者背景から在院死亡率を予測するノモグラムを作成することを目的とした。

【方法】COPD増悪によって入院し2010年7月から2013年3月の間に退院した患者を後ろ向きに解析した。ロジスティック回帰分析によって在院死亡率に関連する因子を検証し、それらの因子を用いて在院死亡率を予測するノモグラムを作成した。

【結果】解析対象者は3,064人、在院死亡者は209人（在院死亡率6.8%）だった。高い死亡率は高齢、男性、BMIが低いこと、意識障害、呼吸困難症状の強さ、機械的人工呼吸の使用歴、肺炎を併発していること、喘息を併発していないことと関与していた。これらの因子を用いて在院死亡率を予測するノモグラムを作成した。Concordance index は0.775であり、Bootstrap法を用いてInternal validation

を行いcalibration plotsを作成したところ、予測される在院死亡率とよく一致していた。

【結語】COPD増悪によって入院した患者の在院死亡率を予測するノモグラムを作成した。このノモグラムは臨床医がより良い治療を行う助けとなり、また、患者及びその家族に適切な情報を伝えるために役立つと考える。

### 2) 気管支喘息発作にて入院した患者の予後予測ノモグラムの検討

【背景】気管支喘息発作は救急外来受診や入院を要し、時に致死的な疾患である。しかし入院を要する気管支喘息発作の予後を予測する因子の知見は限られており、入院患者の予後を予測する方法は現在までに知られていない。そこで我々は本邦の入院患者データベースを用いて、気管支喘息発作にて入院した患者の臨床的特徴と入院中死亡に関連する因子を調査し、入院中死亡を予測するノモグラムを作成した。

【方法】本邦のDPCデータベースを用いて2010年7月から2013年3月までに気管支喘息発作にて入院した患者のデータを後ろ向きに収集した。患者背景を調査し、Cox比例ハザード解析を用いて全ての原因の入院中死亡に関連する因子を検討した。

【結果】総患者数は19,684人、平均年齢は58.8歳(標準偏差19.7歳)、入院日数の平均値は8日(四分位範囲5-12日)、入院中死亡は118名(0.6%)であった。高い入院中死亡率に関連する因子は高齢、男性、意識低下、肺炎および心不全の併存であった。これらの入院時の変数を使用してノモグラムを作成した。結果に基づいた

calibration plotは妥当であり、このノモグラムにより入院中死亡の確率を予測可能と言えた。

【結語】今回我々が作成したノモグラムにより、気管支喘息発作にて入院した患者個人の入院中死亡が予測可能であった。

### 3) リンパ節転移を伴う小細胞肺癌患者に対する術後の化学療法と放射線化学療法の比較

【背景】小細胞肺癌、とくにリンパ節転移症例の最適な術後の治療戦略は明らかでない。我々はリンパ節転移を伴う術後小細胞肺癌に対する化学療法と化学放射線療法を比較した。

【方法】DPCデータベースを用いて、2010年7月から2015年3月までのN1とN2リンパ節転移のある小細胞肺癌症例を抽出した。年齢、性別、併存疾患、手術時のTNM分類、手術法、化学療法薬剤、入院中の放射線治療、退院状況を検討した。Cox回帰分析を用いて、化学療法と放射線化学療法の無再発生存期間を比較した。

【結果】無増悪生存期間の中央値は、化学療法群 (n=489) と放射線化学療法群 (n=75) でそれぞれ、1,146日 (95%信頼区間885-1407) と873日 (95%信頼区間464-1,282) であった。患者背景で調整したのちには2つの治療法に差はみられなかった (ハザード比 1.29、95%信頼区間 0.91-1.84)。

【結語】N1、N2リンパ節転移を伴う小細胞肺癌患者において化学療法と放射線化学療法とで無再発生存期間に有意な差を認めなかった。無作為盲検試験が必要と考えられた。

### 4) 肺炎入院患者における認知症と

### 自宅退院の関連についての検討

【背景】認知症患者において肺炎はもっとも一般的な死因であるが、入院を要した肺炎患者の予後については十分に検討されていない。我々は、肺炎で入院を要した患者において、認知症と在院死亡および退院先の関連について検討した。

【方法】日本国の入院データベースである、DPC (Diagnosis Procedure Combination) データベースを用いて2010年5月1日から2014年3月31日までの間に入院した60歳以上の肺炎患者を後ろ向きに検討した。性別、年齢、BMI (body mass index)、肺炎の重症度、認知症を含む併存疾患を検討した。アウトカムは在院死亡と自宅退院とした。自宅退院との関連の検討は、多変量COX回帰分析にて行った。

【結果】470,829の肺炎入院患者のうち45,031(9.6%)は認知症を伴っていた。在院死亡は、認知症のありとなしでそれぞれ13.1%と13.4%であった。

(P=0.63) 自宅退院した患者の割合は、認知症ありとなしでそれぞれ52.9%と71.3%であった。(P<0.001) 自宅退院についての調整されたハザード比は、認知症のある患者で0.68 (95%信頼区間0.67-0.69、P<0.001) であった。

【結語】肺炎患者の在院死亡率は、認知症のありとなしとでは有意な差を認めなかった。しかしながら、認知症のある患者は自宅退院できる可能性が低かった。

### 5) シスプラチンを含む肺癌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。

## 6) 肺がん診断気管支鏡入院における、在院死亡の検討について

【背景】気管支鏡検査の技術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きている可能性がある。肺がんの診断的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について検討した。

【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気管支鏡検査を検討した。組織診や細胞診を伴わない気管支鏡検査、挿管人工呼吸器管理下の気管支鏡検査、治療目的の気管支鏡検査は除外した。気管支鏡検査後7日以内の在院死亡をアウトカムとし、患者のADL、併存疾患や肺がん病期で調整した。

【結果】954病院の計77,755人に診断的気管支鏡が施行されていた。気管支鏡検査後7日以内の高い在院死亡は、男性、低いADL、3期以上の肺がん病期、慢性心不全、間質性肺炎と有意な関連をしめした。

【結語】肺がんの診断的気管支鏡後の高い在院死亡は、ADLスコア、肺がん病期、併存疾患と関連していた。上記の6つの研究の全てにおいて、臨床疫学の専門家の支援は不可欠であった。

## D. 考察

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。

DPCデータベースを用いた臨床疫学研究には、臨床医によるクリニカルクエスチョンが不可欠である。一方で、データベース研究に必要なリテラシーを臨床の合間に独学で身につけることには相当の困難を伴うとともに、誤った解析や解釈を行う危

険をも伴うと考えられる。本研究で行った種々の検討においても、専門的なサポートが必須であった。

特に以下のように、基本的な点も含めて、十分な認識・理解の不足がみられた。

1)DPCデータベース:一般的なデータベースの知識 (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特性について (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)

2) 研究デザイン: 関連文献の検索法、記述的観察研究か分析的観察研究か、コホート研究かケースコントロール研究か、クリニカルクエスチョンからリサーチクエスチョンへの構造化など

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るが、協業の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、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の活用などが考えられる。実践と検証を継続して行い内容をアップデートすることで、よりよいプログラムの作成が可能となると考えられた。

## 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. Sakamoto Y, Yamauchi Y, Yasunaga H, Takeshima H, Hasegawa W, Jo T, Sasabuchi Y, Matsui H, Fushimi K, Nagase T. Development of a nomogram for predicting in-hospital mortality of patients with exacerbation of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis.* 2017;12:1605-11.
2. Hasegawa W, Yamauchi Y, Yasunaga H, Takeshima H, Sakamoto Y, Jo T, Sasabuchi Y, Matsui H, Fushimi K, Nagase T. Prognostic nomogram for inpatients with asthma exacerbation. *BMC Pulm Med.* 2017;17(1):108.
3. Urushiyama H, Jo T, Yasunaga H, Yamauchi Y, Matsui H, Hasegawa W, Takeshima H, Hiraishi Y, Mitani A, Fushimi K, Nagase T. Adjuvant chemotherapy versus chemoradiotherapy for small cell lung cancer with lymph node metastasis: a retrospective observational study with use of a national database in Japan. *BMC Cancer.* 2017;17(1):613.
4. Jo T, Yasunaga H, Sasabuchi Y, Michihata N, Morita K, Yamauchi Y, Hasegawa W, Takeshima H, Sakamoto Y, Matsui H, Fushimi K, Nagase T. Association between dementia and discharge status in patients hospitalized with pneumonia. *BMC Pulm Med.* 2017;17(1):128.
5. 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.

## 2. 学会発表

1. Hiraishi Y, Jo T, Yamauchi Y, Nagase T, and Yasunaga H. All cause in-hospital mortality of

diagnostic bronchoscopy in lung cancer patients: Data from the Japanese Diagnosis Procedure Combination database, 18<sup>th</sup> World Conference on Lung Cancer. 2017

## 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 29 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

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

研究要旨

本研究では、まず、介護保険レセプト（もしくは医療・介護突合レセプト）データ活用によるヘルスサービスリサーチ(以下HER)を中心とする研究の現状を把握することとし、介護保険レセプトにより算出可能な項目をドナベディアン<sup>®</sup>の3概念にあてはめて整理し、かつ最近のこれらを活用した研究の文献レビューを実施し、現状および今後の可能性を整理した。次に、保健・医学分野の研究に工学との連携をはかってきた経験から、医学と工学を融合することによる、付加価値点及びこれから解決すべき課題について考察することとした。具体的には、機械学習を用いた介護レセプトデータ分析の試みと工学部出身の若手が保健医学研究に携わることから得た知見及び、韓国医療介護関連機構に訪問調査を行った結果を取りまとめた。

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

高橋秀人 国立保健医療科学院 保健・医療・福祉サービス研究分野 統括研究官

中里佳央 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

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

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

佐々木健佑 筑波大学理工学群社会学類

I. 介護レセプト研究の実態把握と可能性

1. 介護レセプトで算出可能な項目の整理

介護レセプトは、現行では行政が主に利用しているが、第三者（大学や研究機関）も、統計情報部を通じた二次利用が可能である。しかし、公的資金に基づく研究であることなどの限定的条件もあり、まだあまり活用実績がないようである。介護レセプトの活用により、介護の質の向上に資する研究等が可能であり、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、未だその有用性はあまり理解されていないようである。現在、データヘルス計画に基づく介護保険レセプトと医療レセプトの突合データの公開が急ピッチで進められており、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た。

そこで、本研究では、まず、我々のこれまでの経験に基づき、介護保険レセプトにより算出可能な項目をドナベディアン<sup>®</sup>の3概念にあてはめて整理した。

我々は、介護保険サービスのHSRを推進するために、ドナベディアン<sup>®</sup>の「構造」(ストラクチャ)、「過程」(プロセス)、「結果」(アウトカム)に着目し、項目を整理し発表してきた(田宮ら 2017)。ここでは、各項目につき、介護保険レセプトで算出されるものを確認し整理した。

その結果が図1である。ここでは、上記で出版された既存のものに改変を加え、介護レセプトで得られる情報を、赤字下線とした。このように、かなりの項目を介護保険レセプトで把握することが可能である。しかも、全国悉皆に近いデータであることは大きな強みである。

次に、介護レセプトで算出できるアウトカムについて、我々の研究経験を通じたまとめを以下に示す。

#### 個人レベルのアウトカム指標

a. 要介護度悪化：要介護度悪化により、医療介護費が大きく増加するだけではなく、本人や介護側の負担やQOLの低下にもつながる。国の決めた基準であり、費用にも直結するため、有用なHSRに活用できる指標であると考え。定期的な認定頻度は少ない場合も多いが、悪化した場合には、いつでも認定申請が可能であり、要介護度が高くなることは給付が増えるというドライブも働き、悪化状態の反映は比較的早いと考えている。

主な研究としては、個人の介護度の推移とサービス利用の関係などを調べることにより、介護度悪化に関連する要因などの把握が可能である。最近の例では、施設利用者の要介護悪化に関連する要因を分析した結果、要介護度が低いほど、多床室に住む利用者の方が悪化しやすい結果が得られた。

- b. 再入院の発生：退院し地域での療養になってからの再入院は、地域包括ケアの指標として重要なアウトカムである。一方、介護保険レセプトでは、転帰(転居か入院か)が明らかでないのが弱点である。しかし、厚生労働省の報告によると、特別養護老人ホームの利用者の退所の理由の97%が死亡或は入院だそうである。これにより、介護レセプトから、記録がなくなった利用者はほぼ死亡或は入院だと判断できる。そこで、我々は、記録がなくなってから1年の間にまだ記録が再現した利用者は、ほぼ入院であることが判断できるとして、研究を進めてきた。
- c. 在宅期間：住み慣れた地域で自分らしい暮らしを人生の最後まで続ける目的で、要介護認定で重度の高齢者に着目し、在宅期間に影響する高齢者の特徴(例えば、どのようなサービスを主に受けているか)を明らかにすることは、介護レセプト情報で分析できる。
- d. 施設入所：今後、高齢者の「住まい」の問題が顕著化していると予測されている。(厚生労働省2006)2014年の調査では、入所希望の待機者が52万人まで

増え、待機者半分以上が待機に不満を感じると報告されている(井上ら 2011)。そのため、施設入居を予防する政策が至急に必要となっている。そこで、要介護度認定高齢者の施設入所に関連する要因を分析した結果、要介護度悪化、居宅サービスの利用が関連していることが分かった。(秋山ら 2015)

#### 施設単位のアウトカム指標

- a. 施設の在宅復帰率: 在宅復帰率とは、老健を退所された方のうち、在宅(自宅・グループホーム・有料老人ホーム・サービス付き高齢者住宅等)に移られる方の割合であり、老健が中間施設の役割をどの程度遂行しているかを図る指標と考える。介護レセプト情報で、在宅復帰支援機能加算が付く施設を算出することができる。さらに、施設調査と突合することにより、在宅復帰率が高い施設の特徴を明らかにすることができる。
- b. 事業所評価加算: 事業所評価加算は、予防介護通所サービス事業所を評価する加算である。選択的サービスを行うことにより身体機能が維持・向上されているかを評価したもので、要支援状態の状態が維持・向上されている人が一定割合以上いた場合に算定される。これも、介護レセプト情報で抽出することができ、事業所評価加算が付く予防事業所に関連する要因を分析することができる。
- c. 要介護度悪化率: 施設の要介護度悪化率が高い施設は、利用者全体の QOL と施設職員の介護負担を増やすことにもつながる。施設レベルの要介護悪化率は施設の質

の評価の指標と考えられる。しかし入所者の身体状況、疾病などの施設ごとのケースミックスを考慮すべきであり、介護レセプトだけではこれらの情報が限られているのが課題である。

- d. 社会参加支援加算: リハビリテーションの利用により ADL・IADL が向上し、社会参加に資する取組に移行するなど、質の高いリハビリテーションを提供する事業所の体制を評価する。ここで、社会参加に資する取組とは、指定通所介護、小規模多機能型居宅介護、一般介護予防事業などへ移行することである。この加算はリハビリテーションの出口だけではなく、就労などさらに進んだ社会参加も考えられる。(厚生労働省 社保審一介護給付費分科会 第 150 回) 介護レセプトと施設調査を突合することで、社会参加支援加算が付く施設に関連する施設特徴を明らかにすることができる。

#### 地域のアウトカム指標

全国ほぼ悉皆のデータにより市町村レベルまで算出可能であることが大きなメリットであるが、市町村名の表記は許可されていない。本研究班では、以下を算出し、研究を実施した。

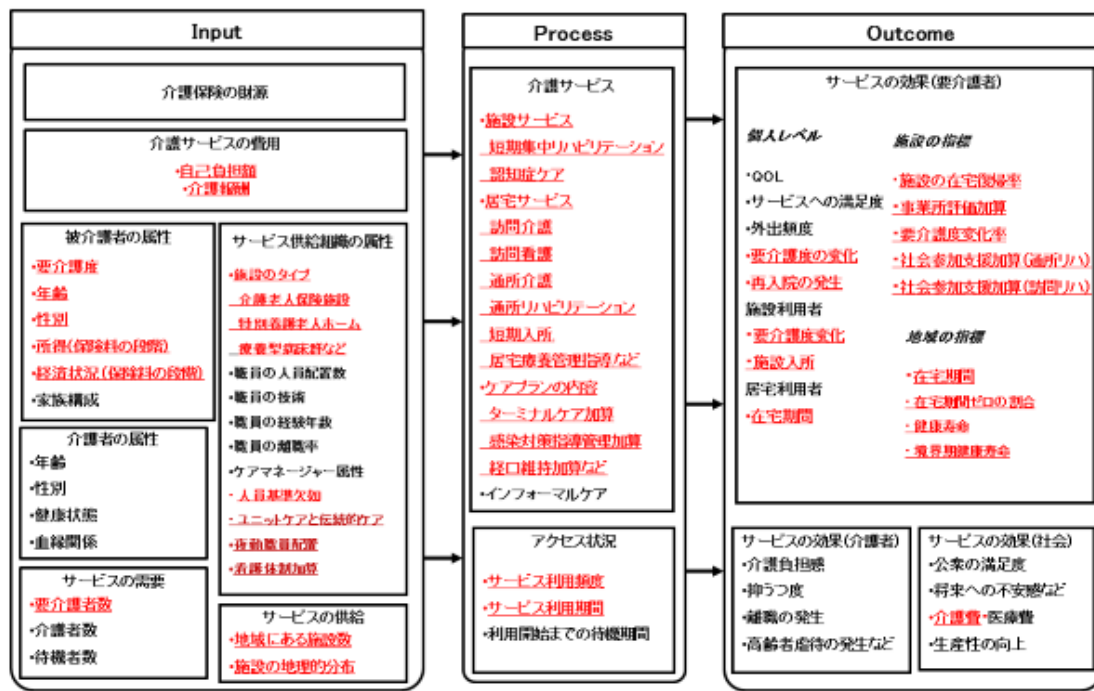
- a. 在宅期間: これは「住み慣れた地域で生活を・・・」という地域包括ケアの目標そのものであり、主要アウトカムとして位置づけられる。実際に算出すると、打ち切りや、要介護になってから一度も在宅にいない利用者(b.在宅ゼロ者割合)が少なからずいることなどがわかり、組み合わせで地域の指標とすることを提言した。(植嶋ら 2017)

b. 在宅ゼロ者割合：上記分析により「在宅期間0日」という対象者（介護保険利用全期間が在宅以外）が50%程度もいることがわかった。aと併用し解釈する必要がある。

c. 要介護度推移：全国介護レセプトの経年蓄積があればこそ算出できる重要なアウトカムである。前の個人レベルでも述べたが、要介護度の厳密な妥当性については議論のあるところである。カテゴリ変数として大きな変化のみを見るなどの工夫により、有用なアウトカムになりうる。また、費用とも直結しているため、政策研究上も有用であると考えられる。しかし、アウトカム指標においては

常にケースミックスを考慮しなければならない。

d. 健康寿命および境界期健康寿命：これまで、一部のサンプルデータ（国民生活基礎調査など）を用いて推定されてきた健康寿命（要介護2になる年齢）が、ほぼ全国の市町村別に実算出できる。また、あらたな概念として境界期健康寿命（要支援1から要介護2までの市町村別平均）を考案し、現在、別経費で分析を進めている（健康寿命及び地域格差の要因分析と健康増進対策の効果検証に関する研究。厚生科学研究（H28-循環器等-一般-009））。



下線の赤字:介護レセプト情報により算出できる指標

図1. 日本の介護保険サービスにおける Input、Process、Outcome の指標の分類。および介護保険レセプトから算出できる指標の整理（赤下線）

（田宮 2017 ヘルスサービスリサーチ入門：生活と調和した医療のためにから改変）

## 2. 介護保険レセプト（もしくは医療・介護突合レセプト）データを活用した研究の文献レビュー

次に、最近5年間の介護保険レセプト（もしくは医療・介護突合レセプト）データを活用した研究の文献レビューを実施し、介護レセプト研究の実態と可能性を考察することとした。

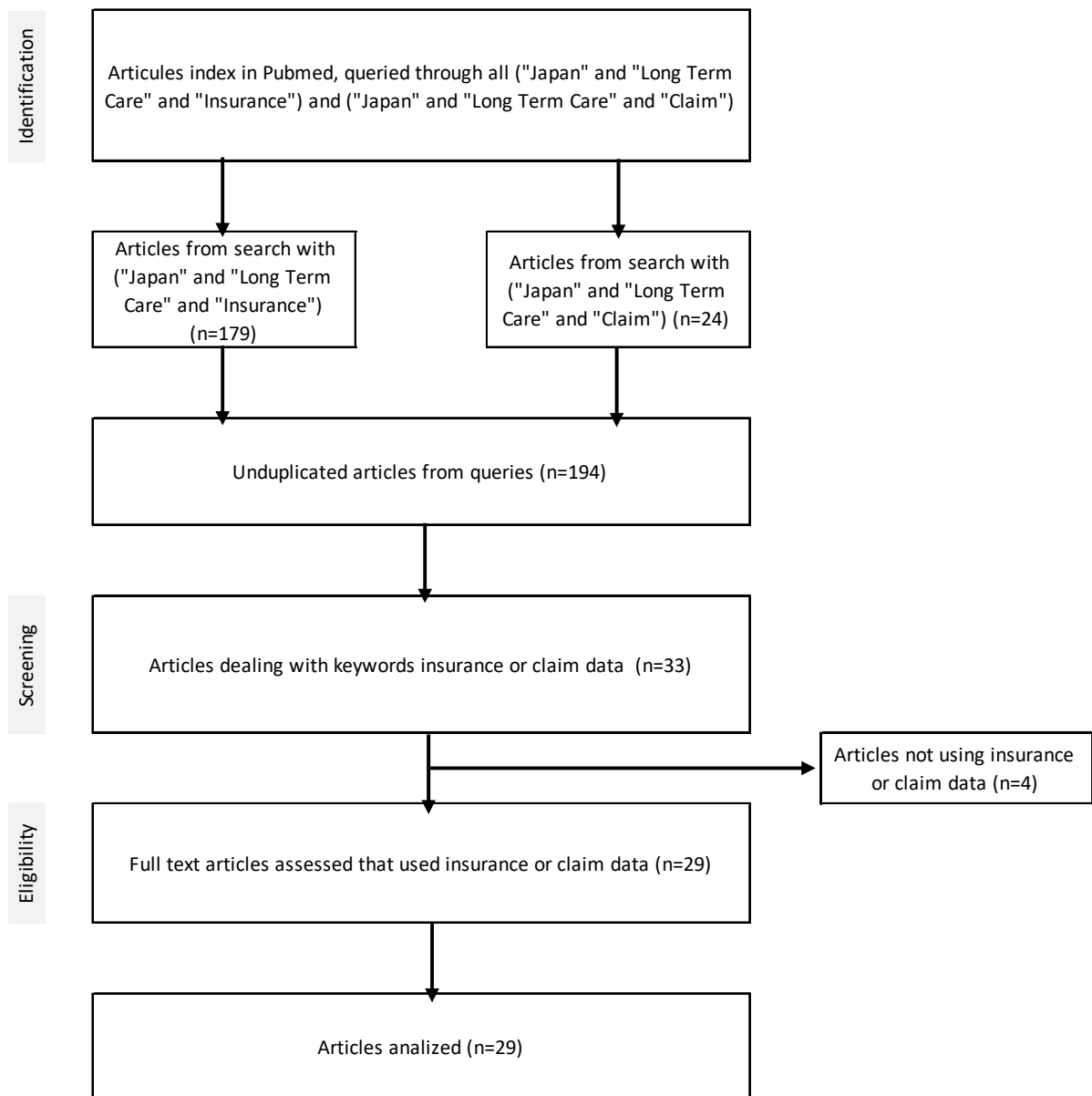
【背景】近年、日本の介護保険レセプト（もしくは医療・介護突合レセプト）の研究利用が進められている。しかし、これらのデータがどのように使われているのか、その取扱いにどのような課題があるかに関する知識はいまだ普及していない。

【方法】過去5年間に英語で出版された要約付きの英文原著論文のうち、PubMedのキーワード検索により“Japan”, “long term care” and “insurance” or “claim”を用いて検索される文献に絞り込んだ。編集ノート、解説、政策の記述、日本では行われていない研究は除外した。タイトルと要約の両方を用いて該当する文献をさらに絞り込んだ。レセプトデータの使用がタイトルと要約で判別できない場合は、フルテキストを取得した。タイトル、ジャーナル、出版日、レセプトデータのタイプ、データの詳細、アウトカムおよび曝露因子のタイプを含むデータを抽

出した。トピックの類似性に関するデータ分析は基本的に定性的である。

【結果】194件の論文が検索され、そのうち組み入れ基準に合致する文献は29件であった（図2）。このうち、過去2年間に31%（9件）が出版された。よく用いられるアウトカムは、サービス利用（14件）、障害（8件）、健康アウトカム（5件）、支出（4件）であった。よく用いられる曝露要因はサービス提供（14件）、健康アウトカム（11件）、社会的背景因子（10件）、災害（2件）、および種々の要因（1件）であった。（表1）。

【考察】アウトカムと曝露要因の両方でよく用いられる項目はサービスであり、サービス提供は予測因子として、サービス利用はアウトカムとして用いられていた。この点はレセプトデータを利用する上での強みのひとつであることが強調されている。もう一つの強みは、死亡率やうつ病などの障害や健康アウトカムを、アウトカムと曝露要因の両方で利用できる点である。支出はアウトカムとしてのみ使用され、社会的背景因子と災害は曝露要因としてのみ使用された。これらの研究のデザインおよび推論における制約を探求するためには、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る。



☒ 2. Identification, screening and eligibility of reviewed articles.

表 1. Included studies by year, outcome type, and exposure type by type of long-term care insurance or claim data in Japan. n=29

	Study title	Year	Type of outcome	Type of exposure	Type of data used	
<b>Query: all fields ("long term care" AND "claim" AND "Japan")</b>					<b>Long term care</b>	<b>Medical long term care</b>
1	Study 1	2018	expenditure	health outcomes		○
2	Study 2	2018	discharge	sociodemographics	○	
3	Study 3	2018	disability	sociodemographics	○	
4	Study 4	2018	disability	health outcomes and sociodemographics	○	
5	Study 5	2017	service use	service provision and sociodemographics	○	○
6	Study 6	2017	expenditure	service provision		○
7	Study 7	2017	health outcome	service provision	○	
8	Study 8	2016	service use	service provision		○
9	Study 9	2017	service use	sociodemographics, health outcomes, service provision		○
10	Study 10	2015	disability	health outcomes	○	○
11	Study 11	2016	expenditure	health outcomes		○
12	Study 12	2015	service use	service provision	○	
13	Study 13	2014	service use	sociodemographics	○	
14	Study 14	2015	disability	sociodemographics, health outcomes, service provision	○	
15	Study 15	2016	disability	service provision	○	
16	Study 16	2014	health outcome and service use	disaster	○	○
17	Study 17	2014	disability	disaster	○	○
18	Study 18	2014	health outcome	service provision	○	
<b>Best match Query: all fields ("long term care" AND "claim" AND "Japan")</b>					<b>Long term care</b>	<b>Medical long term care</b>
1	Study 19	2010	service use	sociodemographics, health outcomes	○	○
2	Study 20	2010	service use	sociodemographics and service provision	○	
3	Study 21	2006	service use	service provision	○	
4	Study 22	2016	service use	various	○	
5	Study 23	2012	disability	health outcomes	○	
6	Study 24	2016	health outcomes	health outcomes	○	
7	Study 25	2016	service use	service provision		○
8	Study 26	2015	disability	health outcomes	○	○
9	Study 27	2016	service use	sociodemographics and service provision	○	
10	Study 28	2014	health outcomes	service provision		○
11	Study 29	2016	expenditure	health outcomes		○

## II. 情報工学と医学の協同によるビッグデータ解析の可能性

### A. 研究目的

本報告書では、「保健医療介護現場の課題に即したビッグデータ解析を実践するための臨床疫学・統計・医療情報技術を磨く高度人材育成プログラムの開発と検証に関する研究」という研究課題に応じるため、情報工学と医学の融合をめざし、工学の方法論を介護レセプトに応用した事例を挙げて説明する。さらに、人材育成に焦点を当てるため、工学出身である若手が、分野を渡り、医学研究に貢献できる点や考えるべき課題をまとめる。最後に、工学部出身で現在医療介護のビッグデータ解析に取り組んでいる若手が平成28年度科研費基盤C（代表高橋秀人）「全国介護保険レセプトデータ・国民生活基礎調査の整備と代表性に関する研究」による韓国視察に同行し、得た知見を取りまとめる。

### B. 研究方法

#### 1. 時間横断的特徴選択に基づく介護レセプトサンプルデータのモデリングとその分析

機械学習を用いた時系列データに対する分析手法の可能性を探るために、既存のデータからランダムサンプリングデータを作成し、スパース正則化を用いたマルチタスク学習において時間横断的な特徴選択を導入し、得られた統計モデルを分析した。さらに、機械学習と統計解析との比較・レビューを行った。

#### 2. 工学出身若手当事者から見た知

見

1) 筑波大学ヘルスサービスリサーチ研究室では2015年4月から工学部と積極的な知識および技術交流を目指し、情報工学専門の若手と共同研究をしている。本報告書では、その共同研究の可能性及び課題について、若手の経験から整理する。

2) 2018年3月7日から8日において、工学部出身である筑波大学ヘルスサービスリサーチ研究室の佐々木健佑と金雪瑩の2名がNational Health Insurance Service と Korean National Health Nutrition Examination Survey へ訪問し、聞き取り調査で全国レベルの医療介護データの管理及び調査方法についての議論を行った。

### C. 研究結果及び考察

#### 1. 時間横断的特徴選択に基づく介護レセプトデータのモデリングとその分析

提案手法を用いた横断的特徴選択の結果、以下の結果を得た。

- ① 各サービス利用と要介護度との関係
- ② サービス利用後の時系列的介護度の変化
- ③ 年齢など基本属性と介護度との関係

マルチタスク学習によるスパース推定により、要介護度の高低と介護サービスや背景情報の関連の強さを定量的に可視化することができることが明らかになり、今後の介護レセプトを用いた行政に役立つ分析手法として、従来の疫学的手法のみでなく



こうした方法も取り入れていく意義が示された。今後、ビッグデータの活用においては、因果推論の分析のみでなく、可視化してすぐに政策決定に役立つ結果を出すことが求められ、機械学習による本方法はそれらを担える可能性があると考えられた。

これらの比較表をもとに、介護の分析に携わった視点から考えると、統計学的方法では「どの要因が要介護状態に影響するか」を説明することが主題であり、機械学習では「要介護状態になる人をどの程度の精度で予測できるか」が主題になるケースが多いようである。

医学公衆衛生系とシステム情報系で同じデータを共有しつつ、さらに両者を比較し、利点欠点を議論しつつ、これまでの学問的背景による特色の生かしながら、よりよい分析方法を見出していくことが、今後重要であろうと考える。

### 1-2 機械学習と統計解析との比較(レビュー)

ここでは、両者について記述してあった3文献をもとに、比較考察をした。(表2)

表 2 統計学と機械学習の比較表

	伝統的統計学	機械学習
方法論的な違い (Zhou, Z.H. 2003)より引用 (Hand, D. J. も含む)	理論的妥当性を重視 (theoretical soundness) 統計学的関連性について、 偶然誤差との比較からの 評価	理論的妥当性と実験的有効 性の両方を重視 (theoretical soundness and experimental effectiveness) データに基づく経験を基に した予測情報の提示
重要な指標 (寺野 隆雄 2001)より作 成	データとモデルの適合度 (作成したモデルによる未 知データへの予測精度 は、議論の対象にならな いことが多い)	(予測・分類問題における) 処理できるデータ量、与え られた未知データに対する 予測の精度
重要な課題 (寺野 隆雄 2001)より作 成	データ全体の傾向を把握 すること	専門家と一致した、もしく は類似した予測・分類を実 際に行うこと (例：脳画像から疾病の有無 を判別、スパムメールの分 類など)

出典：[1] Hand, D. J. (1999). Statistics and data mining: intersecting disciplines. ACM SIGKDD Explorations Newsletter, 1(1), 16-19.

[2] Zhou, Z. H. (2003). Three perspectives of data mining.

[3] 寺野 隆雄. (2001). データマイニングの展望. 情報処理, 42, 5.

[https://www.jstage.jst.go.jp/article/sicej11962/41/5/41\\_5\\_315/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/sicej11962/41/5/41_5_315/_pdf/-char/ja)

## 2. 若手の視点からの考察

### 1. 当事者として考えたこと

#### ① 情報工学からレセプト分析やビッグデータ分析に携わって

まず、情報工学分野の学生が医学に貢献できる可能性の1つとして、レセプトデータ加工・抽出におけるプログラミングスキルやデータベース操作に関するスキルの活用が挙げられる。医療介護の研究において、様々なリサーチクエスション (RQ) があり、その RQ 検証のためのデータセットを加工・抽出するにはプログラミングスキルが必要である。ここで、データ加工や抽出の経験がない医療介護関連研究者にとって、0 からプログラミングやデータベース操作を習得することのコストは少なくない。この点に、情報系の学生が医学に貢献できる可能性があると考えられる。

一方で、情報系学生及び研究者が医療・介護レセプトデータを扱う上で課題となり得るのが、診療報酬制度への理解である。いわゆる「レセプト病名」の問題や、疑い病名、転記区分の取扱など、一見ある事象を表しているように思えるデータであっても、診療報酬制度特有の事象を反映したデータであるケースは少なくない。例えば、ある傷病名の患者を抽出する際に、レセプトの病名を用いて抽出した場合、真にその疾患をもった患者だけではなく、ある処方をするために都合上その傷病名が記録された患者も同時に抽出されてしまう場合などがそれにあたる。レセプトを活用した研究の質を保つために、データ加工・抽出に関する依頼者とのコミュニケーションは、上記のようなレセプトの基礎知識を持った上

で行われるべきである。この点が、今後情報系学生が医学への貢献を果たすための課題の1つであると考えられる。

#### ② 医学と工学融合の現状及び課題：筑波大学を例として

筑波大学理工学群が用意している通常のカリキュラムや研究活動の中では、医療分野で蓄積された情報・データに直接的に関わるチャンスが少ない、もしくはほとんど無い。経営工学主専攻のカリキュラムでは、経営工学やデータ解析などを実践の場で身につける講座が展開されており、学生はデータ解析に要するプログラミング言語やデータハンドリングのスキルを習得することができる。図3は、経営工学主専攻で解説される授業の一覧である。

ここでは、店舗のPOSデータ(Point of sales)や株価のデータなどが対象として扱われる。一方で、医療レセプトデータや健診データなど、医療分野の二次データを対象として扱うことのできる講座は現状では用意されていない。そのような点もあり、2017年度経営工学主専攻の卒業研究論文全54件の中で、医療分野のデータを用いたテーマが扱われているのは3件のみであった。これらのことから、医療分野に関するデータ活用に関心がある場合でも、関連のスキル・経験を蓄積するチャンスが少ない事がわかる。

医療分野でのビッグデータ活用のニーズ・注目度の高まっている状況の中で(日経新聞 2017)、それらを活用するスキルを持った人材をより良く育成するためには、我々のような理工系・情報系の学生に対して、医療

分野の二次データ等に触れる機会(授業・研究等)をより多く提供することが必要ではないかと考える。

## ② 工学部で学んだ数学を医学分野へ応用

医学の世界における数学は、社会科学と同様に、生体现象を理解するための道具に他ならない。しかし、医学部では、道具となる数学への理解を深める、線形代数、微積分、確率統計の科目は選択科目となっている。少なくとも、レセプト等のビッグデータ分析には、理論の基礎が薄い場合、結果に間違いがあっても発見できないことが生じてしまう。

社会工学分野の学生は、必修科目として、線形代数、微積分、確率統計を学ぶ。(図3)そして、データ解析やシミュレーションなどにおける計算の道具の理論的基礎から経営工学における応用の実例までを視野に科目を提供している。そして、数理モデルや計量経済モデルを使うことには、数学的証明は欠かせない。これにより、データ分析の結果に対しての理解を深めることができる。

## 2. 韓国のビッグデータシステムを見学して

### 1. National Health Insurance Service (NHIS) へ訪問して得た知見

韓国における国民健康保険に関する情報の中で、「国民健康保険データ共有サービス」ではサンプルコホートデータを提供している。図4は実際のwebサービスのページである。ウェブサイトにアクセスすると、韓

国の研究者や学術機関に向けて、韓国国民健康保険のサンプルコホートデータの利用申請をオンライン上で行うことができる。また、一部項目(資格と保険料、出生と死亡、診療、健康診断、療養機関)については、個票ベースのデモデータをダウンロードすることが可能である。日本では、レセプトを用いた研究に関心がある臨床医の先生であっても、レセプトデータがどのような形式のデータで、どう扱えるのかに関してのイメージを持つ機会は少ないように思われる。2016年にはNDBのオープンデータが公表され、集計済みのレセプトデータにはアクセスすることができる体制ができていたが、個票ベースのデータセットに関しては、利用申請に何種類もの書式を用意する必要がある(藤森 2016)。韓国では、サンプルのデモデータではあるが、個票ベースのレセプトデータに関してそのイメージが得られる体制が整っている点が非常に魅力的である。

谷原(2012)は、レセプト情報への理解が不足している点について、今後我が国においても、個票ベースのサンプルデータをオープンに提供し、医療レセプト情報を研究者にとってより身近に感じられるような体制の整備が有効であるとしている。

### 2. Korean National Health Nutrition Examination Survey (KNHNES) への訪問で得た知見

—日本の国民健康栄養調査と韓国の国民健康栄養調査の比較から

韓国国民栄養調査は1998年から毎年行われている。調査内容は血

液、糖尿病や歯などの検診と喫煙飲酒や生活習慣などの質問紙調査と食品及び栄養の摂取などの栄養調査が含まれている。日本でも国民健康栄養調査があり、調査内容はほぼ同じである。本報告書では、日本と韓国の健康栄養調査(表5)を比較することで、今後の課題や展望をまとめる。

#### ① 公的データの応用

韓国の国民健康栄養調査のデータは、国の健康増進計画の設立及び評価、OECDの国での比較や予防プログラム開発などに使われている。また、データはホームページに公開され、暗号化された個票データをすぐダウンロードすることができる。データの公開により、多くの研究者が手軽にアクセスすることができ、アカデミックで効率的に活用している。しかし、日本では、研究者が公的個票データを利用する際は、申請からデータの入手まで大変工夫が必要であり、時間がかかってしまう。谷原(2012)では、我が国における医療レセプトを用いた学術研究は、韓国や台湾に遅れを取っている現状が示されている。その原因として、我が国ではレセプト情報を研究に利用する際に、制度面での制約が大きいこと、医療レセプトに記載される情報の特性に対する理解が不十分であることが指摘されている。今回の視察から、韓国において医療レセプトを用いた研究が先進的に行われている背景として、医療レセプト等のデモデータが広くオープンになっていることが1つの要因である可能性が見えてきた。

#### ② データの突合

韓国国民栄養調査で得たデータは他の公的データ(例えば国民健康保

険データ)を暗号化された個人番号で突合することができる。医療介護が突合することで、医療介護連携の評価や医療介護でのサービスの重複利用による非効率利用を避けることができる。

#### ③ アカデミアの声

韓国では、調査票の質問設計においてアカデミアからの意見を積極的に受け取っており、眼科学会、耳鼻咽喉学会など7つの学会と協力している。国民健康栄養調査の調整委員会では、アカデミアから収集した意見を公的意義や質問の妥当性などに基づいて調査票を修正または追加しながら更新を行う。しかし、日本の政府はアカデミアの視点を受け取ることには消極的である。今後日本政府から行われる調査も、アカデミックセクターと協力し、質問紙の質や評価の正確性を上げることを期待する。

#### ④ 質問紙の質

韓国では、調査票の各質問項目に対応した定義書では妥当性検討の説明や定義が整理しており、公表結果の正確性や信頼性を上げている。しかし、日本の公的調査票の質問項目は定義が明確でなく、妥当性の検討も乏しい。今後は、各質問項目のガイドラインが必要であると考えられる。

#### E. 結論

今後、保健医療介護現場の課題に即したビッグデータ解析を実践するためには、医学系の専門知識に工学を媒介として学際的に融合する研究が期待される。また、多国の比較により、より効率的な医療介護ビッグデータの整備へ可能性や方法を明らかにす

る必要があると考える。

F. 研究発表

1. 論文発表

レビュー論文および総説を投稿  
予定

2. 学会発表

なし

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

図3 筑波大学社会工学類 経営工学主専攻シラバス (筑波大学理工学群 2018)

1年次	2年次	3年次	4年次
体育			
社会工学実習 情報リテラシー・演習 プログラミング実習	社会調査実習 社工専門英語	問題発見と解決	卒業研究
線形代数 微積分 統計	<b>マネジメントエリア</b> <ul style="list-style-type: none"> <li>マネジメント実習</li> <li>産業・組織心理学</li> <li>ファイナンス</li> <li>国際企業論</li> <li>経営学</li> <li>マーケティング工学</li> <li>生産品質管理</li> <li>保険数理</li> </ul>		
ミクロ経済学 経営工学概論 都市計画原論	マクロ経済学 会計学概論 都市計画の歴史	<b>情報技術エリア</b> <ul style="list-style-type: none"> <li>情報技術実験</li> <li>経営情報システム</li> <li>情報ネットワーク</li> <li>計算機科学</li> <li>シミュレーション</li> <li>データ解析</li> </ul>	
総合科目 外国語	<b>数理工学モデル化エリア</b> <ul style="list-style-type: none"> <li>数理工学モデル化実習</li> <li>数理最適化法</li> <li>応用確率論</li> <li>数理解析</li> <li>数理統計学</li> <li>応用確率過程</li> </ul>		
フレッシュマン・セミナー			

图4 国民健康保険データ共有サービス (Data providing website)  
<https://nhiss.nhis.or.kr/bd/ay/bdaya001iv.do> (Accessed on 23 March 2018)

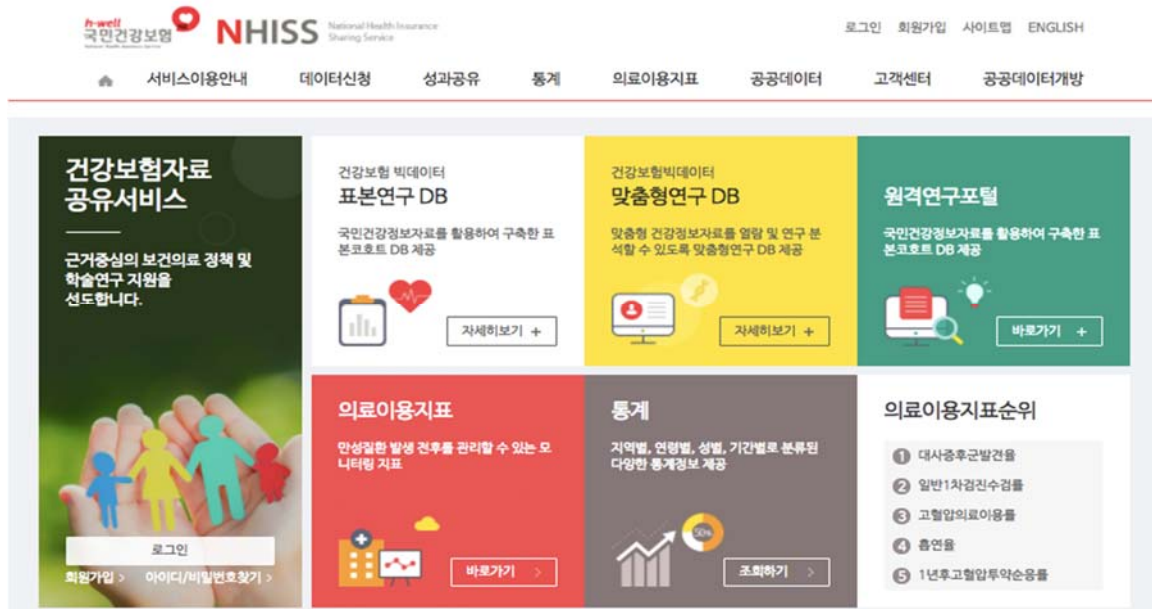


表 4 日本国民健康栄養調査と韓国国民健康栄養調査の概要

	日本国民健康栄養調査	韓国国民健康栄養調査 (KNHNES)
調査の目的	国民の身体の状態、栄養摂取量及び生活習慣の状態を明らかにし、国民の健康の増進の総合的な推進を図るための基礎資料を得ることを目的として、毎年実施するものとする。	国民健康栄養調査は国民の健康及び栄養の状態を明らかにし、政策上優先にすべき健康虚弱集団を抽出し、保健政策が効率的に届いているかの評価に必要である。また、WHO や OECD で求めている喫煙、飲酒、身体活動などの統計を提供している。
調査の根拠法令	健康増進法	健康増進法
調査の対象及び抽出方法	調査年の国民生活基礎調査において設定された単位区から、層化無作為抽出した 300 単位区内の世帯(約 6,000 世帯)及び世帯員	192 調査地点に住んでいる入院を除いた 19 歳以上の成人。
調査事項	身体状況調査票 栄養摂取状況調査票 生活習慣調査票	健康インタビュー 健康診断 栄養状態調査票
調査の時期	11 月中の 1 日	1 年を掛けて行う
調査の方法	調査地区を管轄する保健所	移動検診車両が調査地域に訪問
回答率 (2011 年)	62.9%	健康インタビュー (76.1%) 健康診断 (76.1%) 栄養状態調査票 (82.4%)

出典：厚生労働省 国民健康・栄養調査

<http://www.mhlw.go.jp/toukei/itiran/gaiyo/k-eisei.html>

Korea center for disease control and prevention. About Korea national health and nutrition examination survey <https://knhanes.cdc.go.kr/knhanes/eng/index.do>



参考文献：

田宮 菜奈子, 小林 廉毅 (2017) 「ヘルスサービスリサーチ入門: 生活と調和した医療のために」 東京大学出版会

秋山 直美, 白岩 健, 福田 敬, 村嶋 幸代 (2015) 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— 日本医療・病院管理学会誌 52 巻 2 号

厚生労働省, 第 1 回 介護施設等の在り方委員会  
<http://www.mhlw.go.jp/shingi/2006/09/dl/s0927-8e.pdf>

井上信次, 岡本信夫, 特別養護老人ホームの待機者調査にみる待機者及びその介護者へのアウトリサーチの必要性, 川崎医療福祉学会誌, 20 (2), 331-345, 2011

植嶋 大晃, 高橋 秀人, 野口 晴子, 川村 颯, 松本 吉央, 森山 葉子, 田宮 菜奈子: 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-、厚生 の 指 標、64 (15) 8-18 頁 2017.12

厚生労働省(2016). 「保険診療の理解のために【医科】(平成28年度版)」(Accessed on 23 March 2018)  
[http://www.mhlw.go.jp/seisakunitsuite/bunya/kenkou\\_iryuu/iryuuhoken/dl/shidou\\_kansa\\_01.pdf](http://www.mhlw.go.jp/seisakunitsuite/bunya/kenkou_iryuu/iryuuhoken/dl/shidou_kansa_01.pdf)

厚生労働省 (2016). 「第 1 回 NDB オープンデータ 【解説編】」  
<http://www.mhlw.go.jp/file/06-Seisakujouhou-12400000-Hokenkyoku/0000141549.pdf>

National Health Insurance Service HP (Accessed on 23 March 2018)  
<https://nhiss.nhis.or.kr/bd/ay/bdaya001iv.do>

藤森研司. (2016). レセプトデータベース (NDB) の現状とその活用に対する課題. 医療と社会, 26(1), 15-24.

平成 29 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 生物統計学 教授

研究要旨 日常診療データから治療効果を推定する手法として、動的治療レジメを用いた統計解析の有用性を検討した。日常診療に近いと考えられる転移・再発乳がん患者を対象にした SELECT BC 試験を動的周辺構造 Cox モデルの IPW 推定法で統計解析した結果、1 次化学療法としてタキサン系薬剤を投与された患者が 2 次化学療法として TS-1 を投与されると、その他のプロトコル推奨化学療法と比較して、全生存期間が延長した。加えて、マッチドコホート研究での層別 Cox 回帰の統計的性能および自己対照デザインの統計学的性能をシミュレーション実験で評価した。マッチドコホート研究での層別 Cox 回帰は、追跡からの脱落がマッチドペアや曝露に依存していたとしても、バイアスなく共通ハザード比を推定できることが示された。自己対照デザインである自己対照ケースシリーズデザイン、ケースクロスオーバーデザイン、sequence symmetry analysis を用いて曝露とアウトカムに関連を妥当に推定できるかどうかは、解析に必要な仮定の妥当性に依存することが示された。

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

A. 研究目的

本分担研究の目的は、大規模診療データ（いわゆる、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、経時日常診療データから治療効果を推定する手法のうち、特に動的治療レジメに対する統計的推測法を研究した。また、医療ビッグデータの統計解析に有用と考えられるマッチドペアコホート研究での層別 Cox 回帰と自己対照デザインの統計的性能をシミュレーション実験により評価した。

B. 研究方法

1. 経時日常診療データからの治療効果推定

多くの医療ビッグデータに当てはまる特徴のひとつは、日常診療データから構成されていることである。実験的側面が強い臨床試験では治療が研究プロトコルで厳密に規定されている一方、日常診療では治療への反応などの患者個人のデータに応じて治療が変更されていく。このようなデータから経時的治療の効果を正しく推測するためには、アウトカムに影響を与え、かつ次の治療も予測する時間依存性交絡因子が惹起する時間依存性交絡を適切に調整しなければならない。しかし、時間依存性交絡因子が以前の治療の影響を受ける場合、古典的交絡調整法である層別解析や回帰分析では時間依存性交絡を適切に調整でき

ない。そのような時間依存性交絡因子は調整すべき交絡因子である一方、層別解析や回帰分析は治療効果の一部をブロックしてしまうだけでなく、アウトカムと時間依存性交絡因子の共通原因が測定されていない場合、合流点バイアスを引き起こす可能性もあるためである。そのため、時間依存性交絡を適切に調整できる方法として、g-formula、周辺構造モデルの IPW (inverse probability weighted) 推定法、構造ネストモデルの g 推定法などが提案されてきた。

周辺構造モデルの IPW 推定法を中心に、時間依存性交絡を適切に調整できるこれらの手法を実データへ適用した研究は増加している。その一方、臨床検査値や臨床所見をもとにした適応的治療変更を織り込んだ治療方針（動的治療レジメ）の治療効果を推定する手法を用いた解析事例は依然として少ない。どのような順序で治療を行うと望ましいアウトカムを得ることができるのかという疑問や、どのような臨床検査値が得られた場合に治療を開始すべきかという疑問に答えられるのは、動的治療レジメに関する統計手法である。動的治療レジメに対して統計的に推測する際にも時間依存性交絡を適切に調整しなければならないが、g-formula、周辺構造モデルの IPW 推定法、構造ネストモデルの g 推定法のいずれでも、動的治療レジメを扱う手法が提案されている。

本年度は、将来的に医療ビッグデータに適用することを念頭に、動的治療レジメに対する IPW 推定法を比較的小規模な実データで実行し、その実行可能性と結果の解釈可能性について検討した。

今回用いたデータは、HER2 (human epidermal growth factor receptor 2) 陰性

ホルモン療法抵抗性転移・再発乳がん患者を対象に本邦で行われたランダム化第 III 相 SELECT BC 試験データである。本分担研究テーマから見た SELECT BC 試験の特徴は、1 次治療（タキサン系薬剤または TS-1）はランダム化されているものの、2 次治療に関しては日常診療を反映した種々の化学療法レジメンが試験プロトコルで許容されていることである。また、SELECT BC 試験は多数の施設（154 施設）で行われ、適格基準や治療・検査計画も比較的日常診療に近いプラグマティックな側面が強い臨床試験である。

本研究では、SELECT BC 試験データに動的周辺構造 Cox モデルの IPW 推定法を適用し、各 1 次治療（タキサン系薬剤および TS-1）に関して最適な 2 次治療を推定した。今回用いる動的周辺構造 Cox モデルは、1 次治療を中止または終了したあとに 2 次治療を開始する動的治療レジメンのもとの全生存期間のハザードを比例ハザードモデルでモデル化したものである。交絡調整に用いた変数のうちベースライン変数はホルモン受容体有無、肝転移有無、リンパ節転移個数、再発までの時間、術後補助療法歴とし、時間依存性変数は 2 次治療開始までの時間、ECOG (Eastern Cooperative Oncology Group) パフォーマンスステータス、腫瘍増悪有無、有害事象を理由にした 1 次治療の中止とした。2 次治療はクロスオーバー（1 次治療がタキサン系薬剤の患者が TS-1 を 2 次治療として受けること、もしくはその反対）、クロスオーバー以外のプロトコル推奨化学療法（アントラサイクリン系薬剤、カペシタビン、ビノレルビン、シクロフォスファミド+メソトレキセート+フルオロウラシル、イリノテカン）、そ

の他の化学治療に分類した。多項ロジスティック回帰により推定した2次治療を受ける確率を用いて動的治療レジメ用の重みを算出し、重み付き Nelson–Aalen 推定量と動的周辺構造 Cox モデルの IPW 推定法を用いて全生存期間に関する生存関数とハザード比を推定した。

## 2. 医療ビッグデータ解析に有用と考えられる統計解析手法の性能評価

医療ビッグデータの統計解析に有用と考えられる手法の統計的性能をシミュレーション実験で評価した。本年度は(1) マッチドペアコホート研究における層別 Cox 回帰および(2) 自己対照デザインの統計的性能を評価した。

### 2.1 マッチドペアコホート研究における層別 Cox 回帰

マッチドペアコホート研究は、マッチング因子に含まれる交絡因子を研究デザインで調整できる。これに加えてマッチドペアコホート研究における層別 Cox 回帰は、マッチング因子に依存した追跡からの脱落も調整可能という特徴を持つ。対象者が十分追跡されない(たとえば、レセプトにもとづくデータセットではレセプトが発生しない限り追跡が行われないことを意味する)こともある医療ビッグデータの統計解析において、マッチドペアコホート研究における層別 Cox 回帰は脱落による選択バイアスの調整に有用と考えられる。

本研究では、マッチドペアコホート研究における層別 Cox 回帰と通常(周辺) Cox 回帰の統計的性能を、層別 Cox 回帰に従うシミュレーションデータを用いて検討した。シミュレーション実験では以下のシナリオを検

討した。すなわち、共通ハザード比の大きさ(2.0、1.0、0.5)、打ち切りメカニズム(独立な打ち切りとマッチドペアおよび曝露に依存した打ち切り)、曝露と打ち切りの関連の強さ、サンプルサイズ(50層100名および250層500名)である。生存時間変数は正規変量切片を持つ指数フレイルティモデルから、打ち切り時間変数は独立な打ち切りの場合は指数モデルから、ペアおよび曝露に依存した打ち切りの場合は正規変量切片を持つ指数フレイルティモデルから発生させた。発生させたシミュレーションデータは、層別 Cox 回帰、モデル分散を用いた(非層別) Cox 回帰、サンドウィッチ分散を用いた(非層別) Cox 回帰で解析した。評価指標はバイアス、95%信頼区間の被覆確率、平均2乗誤差の平方根とし、それぞれの評価指標を対数共通ハザード比と対数周辺ハザード比の両者に関して算出した。シミュレーション回数は2000回とした。

### 2.2 自己対照デザイン

自己対照デザインは、交絡因子が測定されていなかったとしても、その交絡因子が時間非依存性であれば、対象者個人を調整因子とすることで交絡を調整できるという特徴を持つ。この特徴は、交絡因子に関するデータに乏しいこともある非研究用の医療ビッグデータの統計解析において、交絡によるバイアスを低減するのに有用と考えられる。これまで、いくつかの自己対照デザインと付随する統計解析法が提案されており、各研究デザインが妥当な推定値を与える条件は異なっていることが知られている。しかし、仮定からの逸脱が自己対照デザインの統計的性能にどのように影響するかは十分検討されていない。

本研究では、自己対照デザインとして、自己対照ケースシリーズデザイン、ケースクロスオーバーデザイン、sequence symmetry analysis の3つの研究デザインに焦点を当て、仮定を変えたシミュレーション実験により、3つの研究デザインの統計的性能を評価した。今回のシミュレーション実験で扱ったシナリオは以下の通りである。すなわち、時間依存性交絡因子の有無、時間依存性交絡因子と時間非依存性交絡因子の交互作用の有無、曝露およびアウトカムにおける時間傾向の有無、アウトカム発生による追跡終了の量である。交絡変数の個数は、時間非依存性交絡変数が2つ、時間依存性交絡因子が1つとした。曝露発生までの時間、およびアウトカム発生までの時間は指数分布に従うとし、曝露またはアウトカムが発生した場合には、再度曝露またはアウトカムの発生までの時間を再発生させ、対象者内で複数回の曝露とアウトカムが発生しうる状況とした。以上のデータ発生を100,000名の対象者に関して1800日の範囲で行った。発生させたシミュレーションデータに、条件付きポアソン回帰を用いる自己対照ケースシリーズデザイン、条件付きロジスティック回帰を用いるケースクロスオーバーデザイン、調整 sequence 比を算出する sequence symmetry analysis を適用し、バイアス、平均二乗誤差、95%信頼区間の被覆確率を算出した。シミュレーション回数は各シナリオ 2000 回とした。

(倫理面への配慮)

SELECT BC 試験はヘルシンキ宣言ならびに当時の臨床研究に関する倫理指針を遵守する形で実施された。試験プロトコルならびに同意文書が参加施設の倫理審査委員会で承認され

たあと、患者から同意文書による同意を得て実施された。

### C. 研究結果

#### 1. 経時日常診療データからの治療効果推定

SELECT BC 試験における最大の解析対象集団であるタキサン群 286 名、TS-1 群 306 名のうち、タキサン群では 110 名がクロスオーバーし、76 名がクロスオーバー以外のプロトコル推奨化学療法を受け、16 名がその他の化学療法を受け、TS-1 群では 114 名がクロスオーバーし、68 名がプロトコル推奨化学療法を受け、36 名がその他の化学療法を受けた。

図 1 に 1 次治療と 2 次治療で定義された動的治療レジメンのもとでの全生存期間に関する推定生存曲線を示す。1 次治療がタキサン系薬剤の場合、全生存期間中央値はクロスオーバー (TS-1) で 39.6 か月、プロトコル推奨化学療法で 35.7 か月、その他の化学療法で 36.9 か月であった。1 次治療がタキサン系薬剤の場合、プロトコル推奨化学療法と比較するとクロスオーバーで有意に全生存期間が延長した (ハザード比 0.72、95%信頼区間 0.52–0.98、 $P=0.037$ )。クロスオーバーとその他の化学療法の比較 (クロスオーバーのハザード比 0.71、95%信頼区間 0.43–1.18、 $P=0.183$ ) およびプロトコル推奨化学療法とその他の化学療法の比較 (プロトコル推奨化学療法のハザード比 0.99、95%信頼区間 0.61–1.62、 $P=0.971$ ) では全生存期間に有意差はなかった。1 次治療が TS-1 の場合、全生存期間中央値はクロスオーバー (タキサン系薬剤) で 34.3 か月、プロトコル推奨化学療法で 38.9 か月、その他の化学療法で 42.0 か月であった。1 次治療が TS-1 の場合、クロスオーバーのハザード比は、

プロトコル推奨化学療法との比較で 1.16 (95%信頼区間 0.85–1.60、 $P = 0.350$ )、その他の化学療法との比較で 1.06 (95%信頼区間 0.74–1.53、 $P = 0.739$ )であった。プロトコル推奨化学療法とその他の化学療法を比較した場合でも有意な全生存期間の延長はなかった (プロトコル推奨化学療法のハザード比 0.91、95%信頼区間 0.62–1.34、 $P = 0.644$ )。

## 2. 医療ビッグデータ解析に有用と考えられる統計解析手法の性能評価

### 2.1 マッチドペアコホート研究における層別 Cox 回帰

マッチドペアおよび曝露に打ち切りが依存したシナリオの結果を表 1 および表 2 に示す。表 1 および表 2 は 250 層 500 名のシナリオの結果である。打ち切りがマッチドペアのみに依存するシナリオでは、曝露効果がなければいずれの解析方法にもバイアスは認められなかった。しかし、サンドウィッチ分散を用いない非層別 Cox 回帰では標準誤差を過大評価し 95%信頼区間の被覆確率は名目水準以上になった。曝露効果がある場合には、打ち切りが曝露に依存していなくても非層別 Cox 回帰には共通ハザード比および周辺ハザード比のいずれでもバイアスがあった。曝露にも打ち切りが依存しているシナリオでは、曝露効果に関わらず非層別 Cox 回帰は共通ハザード比および周辺ハザード比のいずれにもバイアスがあった。いずれのシナリオにおいても層別 Cox 回帰は共通ハザード比に対してはバイアスがなく、95%信頼区間の被覆確率も名目水準に保たれていた。50 層 100 名のシナリオにおいても同様の結果が得られた。

### 2.2 自己対照デザイン

時間依存性交絡因子、時間依存性交絡因子と時間非依存性交絡因子の交互作用、曝露およびアウトカムにおける時間傾向、アウトカム発生による追跡終了がいずれもない場合には、3つの研究デザインによる点推定値にはバイアスはなく、95%信頼区間の被覆確率は名目水準前後に保たれていた。これらのシナリオで最も推定精度が高かったのは、自己対照ケースシリーズデザインであった。

表 3 に時間依存性交絡因子が存在する場合のシミュレーション結果を示す。時間依存性交絡因子が存在する場合には、自己対照ケースシリーズデザインおよびケースクロスオーバーデザインにバイアスが認められた。このバイアスは、曝露効果が大きいほど、また時間依存性交絡因子の影響が長期間におよぶほど大きくなる傾向があった。曝露効果が大きい場合や時間依存性交絡因子の影響が長期間におよぶ場合には、95%信頼区間の被覆確率の低下が自己対照ケースシリーズデザインおよびケースクロスオーバーデザインには認められた。時間依存性交絡因子とアウトカムとの関連が時間非依存性交絡因子によって強められる場合にも、自己対照ケースシリーズデザインおよびケースクロスオーバーデザインでバイアスと被覆確率に関する性能が低下した。

曝露およびアウトカムにおける時間に関する傾向に関しては、両者が同時に存在する場合に、自己対照ケースコントロールデザインでバイアスと被覆確率の低下が認められた。アウトカムが発生した場合に一部の対象者で追跡が終了する場合には、sequence symmetry analysis に大きなバイアスと被覆確率の低下があり、これらの統計

的性能の低下は、追跡終了割合が高いほど顕著であった(表4)。曝露のリスク期間として真実より短い期間を用いた場合、自己対照ケースシリーズデザインおよび **sequence symmetry analysis** では大きなバイアスが認められなかった一方、曝露のリスク期間として真実より長い期間を用いた場合、すべてのデザインで大きなバイアスと被覆確率の低下が認められた。

#### D. 考察

SELECT BC 試験データに動的周辺構造 Cox モデルの IPW 推定法を適用したことで、各1次治療に対して最適な2次治療に関して示唆を得ることができた。このような示唆は、時間依存性交絡因子を適切に調整できる手法の中でも、動的治療レジメに対する推測法を用いることではじめて得ることができた。日常診療データから患者データに依存した治療方針に関して知見を得たいときに、動的治療レジメによる定式化とそれに対する推測法は優れた道具となりえる。本研究では動的治療レジメに対する推測法として、動的周辺構造モデルの IPW 推定法を用いた。動的周辺構造モデルの IPW 推定法による結果は、疫学研究でも適用例が増加している周辺構造モデルの IPTW 推定法による結果と同様に解釈可能であり、非統計家でも実行しやすく、また容易に解釈可能と考えられる。

通常のコホート研究にくらべて、マッチドペアコホート研究は用いられることが少ない研究デザインである。本研究で行ったシミュレーション実験では、マッチドペアや曝露に打ち切りが依存している状況下で、マッチドペアを層とした層別 Cox 回帰は、通常の Cox 回帰とくらべて、優れた統計的

性能を有していた。そのため、マッチドペアコホート研究における層別 Cox 回帰は、追跡からの脱落が共変量に依存していることが疑われる研究において有用な統計解析法と考えられる。しかし、層別 Cox 回帰はハザード比が層間で共通という制約的仮定をおいている。さらに、層間のハザード比が共通であったとしても、ハザード比の **non-collapsibility** により、共通ハザード比を周辺ハザード比と解釈することはできない。層別 Cox 回帰を適用する際には、このような解釈の点に注意する必要がある。

自己対照デザインは、個人内で不変な交絡因子が十分測定されていない可能性がある医療ビッグデータの統計解析で有用な可能性がある。本研究で行ったシミュレーション実験により、自己対照ケースシリーズデザイン、ケースクロスオーバーデザイン、**sequence symmetry analysis** の統計的性能は条件により大きく影響することが示された。実際に自己対照デザインを用いる場合には、どのような仮定が(近似的に)成り立っているか十分吟味し、適切な研究デザインを選択することが重要である。

本年度の研究成果を、ビッグデータの統計解析に関する人材育成に活用する予定である。医療ビッグデータ解析に精通する人材の養成プログラムの一部として、ビッグデータの統計解析技能を養成するプログラムを提供する。開発したプログラムは平成30年8月に開催予定の短期集中セミナーやその他の講習会で試行し、より優れたプログラムへ改善することを目指す。

平成30年度の課題として、医療ビッグデータ解析において、バイアス解析をどのように実施すればいいのか十分検討できていない。特に、経時治

療の効果を推定する文脈において、バイアス解析に関する研究は限定的である。医療ビッグデータは本来研究用データではないため、十分な調整変数が得られないことがしばしばある。また、医療ビッグデータ解析は十分なサンプルサイズにより偶然誤差が小さくなる傾向にあるため、系統誤差（バイアス）の重要性がより大きいと考えることができる。そのため、医療ビッグデータを用いて経時的治療の効果を推定する際に、バイアスを定量的に評価するバイアス解析をどのように実施すればよいのか検討が必要である。

#### E. 結論

動的治療レジメは医療ビッグデータなどの日常診療データを用いて経時治療に関する研究仮説を検討する際に有用な手法となりうる。シミュレーション実験から、層別 Cox 回帰と自己対照デザインは、仮定に注意して用いることで、医療ビッグデータ研究における研究デザイン・統計解析法として有用となりうる。

#### 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. Mukai H, Hagiwara Y, Imi K, Isaka H, Watanabe K, Matsuyama Y for the SELECT BC Study Group. The impact of treatment preferences in second-line chemotherapy on the prognosis of HER2-negative metastatic breast cancer. *Oncology*. 2017;93:315–322.
2. Shinozaki T, Hagiwara Y, Matsuyama Y. Re. Biases in randomized trials: a

conversation between trialists and epidemiologists (Letter). *Epidemiology*. 2017;28:e40–e41.

3. Shinozaki T, Mansournia MA, Matsuyama Y. On hazard ratio estimators by proportional hazards models in matched-pair cohort studies. *Emerg Themes Epidemiol*. 2017;14:6.
4. Takeuchi Y, Shinozaki T, Matsuyama Y. A comparison of estimators from self-controlled case series, case-crossover design, and sequence symmetry analysis for pharmacoepidemiological studies. *BMC Med Res Methodol*. 2018;18:4.

##### 2. 学会発表

1. Takeuchi Y, Shinozaki T, Hiramatsu T, Matsuyama Y. The risk of liver injury associated with antibiotics adjusted for change of treatments by inverse probability of censoring weighting. 33rd International Conference on Pharmacoepidemiology & Therapeutic Risk Management. Montreal. Aug 2017.
2. 竹内由則, 篠崎智大, 隈丸拓, 平松達雄, 松山裕. 薬剤疫学研究における Intent-to-treat effect と Per-protocol effect: 抗菌薬による肝障害発症リスク評価への適用事例. 日本薬剤疫学会第 23 回学術総会. 東京. 2017 年 11 月.
3. Takeuchi Y, Shinozaki T, Matsuyama Y. Stratified proportional hazards model in self-controlled design: validity and extension of sequential symmetry analysis. Society for Epidemiologic Research 50th Annual Meeting. Seattle. Jun 2017.
4. 萩原康博, 福田武蔵, 松山裕. Modified Least-Squares 回帰による妥当なリスク差推定に必要な 1 交絡変数当たりのイベント数. 2018 年度日本計量生物学会年会. 東京.



2018年3月29日.

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。

表 1 打ち切りがマッチドペアおよび曝露に依存したシナリオにおける共通ハザード比に関するシミュレーション実験結果 (Shinozaki et al. Emerg Themes Epidemiol. 2017;14:6.)

	打ち切り時間 に 関する曝露率 比	バイアス	被覆確率 (%)	RMSE
生存時間に関する曝露ハザード比 2.0				
層別 Cox 回帰	0.25	0.00	94.95	0.16
Cox 回帰 (モデル分散)		-0.04	96.50	0.10
Cox 回帰 (サンドウィッチ分散)		-0.04	91.55	0.10
層別 Cox 回帰	1.0	0.00	94.85	0.18
Cox 回帰 (モデル分散)		-0.15	78.80	0.18
Cox 回帰 (サンドウィッチ分散)		-0.15	68.55	0.18
層別 Cox 回帰	4.0	0.01	95.35	0.22
Cox 回帰 (モデル分散)		-0.39	22.85	0.41
Cox 回帰 (サンドウィッチ分散)		-0.39	17.95	0.41
生存時間に関する曝露ハザード比 1.0				
層別 Cox 回帰	0.25	0.00	95.40	0.16
Cox 回帰 (モデル分散)		0.17	71.00	0.20
Cox 回帰 (サンドウィッチ分散)		0.17	59.65	0.20
層別 Cox 回帰	1.0	0.00	95.50	0.18
Cox 回帰 (モデル分散)		0.00	97.75	0.11
Cox 回帰 (サンドウィッチ分散)		0.00	95.20	0.11
層別 Cox 回帰	4.0	0.00	95.80	0.24
Cox 回帰 (モデル分散)		-0.33	52.20	0.36
Cox 回帰 (サンドウィッチ分散)		-0.33	47.00	0.36
生存時間に関する曝露ハザード比 0.5				
層別 Cox 回帰	0.25	0.00	95.50	0.18
Cox 回帰 (モデル分散)		0.35	15.30	0.37
Cox 回帰 (サンドウィッチ分散)		0.35	9.60	0.37
層別 Cox 回帰	1.0	0.00	95.10	0.21
Cox 回帰 (モデル分散)		0.11	90.35	0.17
Cox 回帰 (サンドウィッチ分散)		0.11	85.85	0.17

層別 Cox 回帰	4.0	-0.01	95.35	0.28
Cox 回帰 (モデル分散)		-0.29	77.05	0.35
Cox 回帰 (サンドウィッチ分散)		-0.29	74.15	0.35

---

RMSE; root mean squared error。

表 2 打ち切りがマッチドペアおよび曝露に依存したシナリオにおける周辺ハザード比に関するシミュレーション実験結果 (Shinozaki et al. Emerg Themes Epidemiol. 2017;14:6.)

	打ち切り時間 に 関する曝露率 比	バイアス	被覆確率 (%)	RMSE
生存時間に関する曝露ハザード比 2.0				
層別 Cox 回帰	0.25	0.26	65.90	0.30
Cox 回帰 (モデル分散)		0.21	52.90	0.23
Cox 回帰 (サンドウィッチ分散)		0.21	37.75	0.23
層別 Cox 回帰	1.0	0.26	70.50	0.31
Cox 回帰 (モデル分散)		0.11	89.05	0.15
Cox 回帰 (サンドウィッチ分散)		0.11	82.65	0.15
層別 Cox 回帰	4.0	0.26	80.00	0.34
Cox 回帰 (モデル分散)		-0.13	87.40	0.19
Cox 回帰 (サンドウィッチ分散)		-0.13	83.60	0.19
生存時間に関する曝露ハザード比 1.0				
層別 Cox 回帰	0.25	0.00	95.40	0.16
Cox 回帰 (モデル分散)		0.17	71.00	0.20
Cox 回帰 (サンドウィッチ分散)		0.17	59.65	0.20
層別 Cox 回帰	1.0	0.00	95.50	0.18
Cox 回帰 (モデル分散)		0.00	97.75	0.11
Cox 回帰 (サンドウィッチ分散)		0.00	95.20	0.11
層別 Cox 回帰	4.0	0.00	95.80	0.24
Cox 回帰 (モデル分散)		-0.33	52.20	0.36
Cox 回帰 (サンドウィッチ分散)		-0.33	47.00	0.36
生存時間に関する曝露ハザード比 0.5				
層別 Cox 回帰	0.25	-0.26	72.70	0.31
Cox 回帰 (モデル分散)		0.10	90.70	0.14
Cox 回帰 (サンドウィッチ分散)		0.10	84.75	0.14
層別 Cox 回帰	1.0	-0.26	77.85	0.33
Cox 回帰 (モデル分散)		-0.15	85.50	0.20
Cox 回帰 (サンドウィッチ分散)		-0.15	79.80	0.20

層別 Cox 回帰	4.0	-0.26	86.80	0.39
Cox 回帰 (モデル分散)		-0.54	25.40	0.58
Cox 回帰 (サンドウィッチ分散)		-0.54	22.20	0.58

---

RMSE; root mean squared error。

表 3 時間依存性交絡因子が存在する場合のシミュレーション実験結果  
(Takeuchi et al. BMC Med Res Methodol. 2018;18:4.)

	曝露効 果	曝露の リスク期 間	交互作用*	バイアス	平均二乗誤 差	被覆確率 (%)
SCCS	log 1.0	15	log 1.0	0.0903	0.0318	88.6
CCO				0.1017	0.0649	93.1
SSA				0.0145	0.0655	96.1
SCCS	log 3.0	15		0.1012	0.0185	77.0
CCO				0.1101	0.0506	92.3
SSA				0.0227	0.0471	94.8
SCCS	log 10.0	15		0.1036	0.0133	46.4
CCO				0.1269	0.0475	91.0
SSA				0.0193	0.0349	96.1
SCCS	log 3.0	5		0.0171	0.0092	94.5
CCO				0.0367	0.0418	94.8
SSA				0.0213	0.0529	95.1
SCCS		10		0.0501	0.0108	90.0
CCO				0.0643	0.0420	94.4
SSA				0.0286	0.0503	95.1
SCCS		20		0.1552	0.0312	55.0
CCO				0.1658	0.0610	88.5
SSA				0.0203	0.0426	95.1
SCCS		15	log 0.2	0.0195	0.0099	93.7
CCO				0.0267	0.0409	94.6
SSA				0.0174	0.0521	94.3
SCCS		15	log 0.5	0.0528	0.0117	89.1
CCO				0.0685	0.0438	94.1
SSA				0.0340	0.0511	95.1
SCCS		15	log 2.0	0.1843	0.0413	40.4
CCO				0.2010	0.0750	82.7
SSA				0.0281	0.0410	95.3
SCCS		15	log 5.0	0.3587	0.1445	18.7
CCO				0.3691	0.1754	53.6
SSA				0.0432	0.0461	94.5

SCCA; self-controlled case series (自己対照ケースシリーズ)、CCO; case-crossover (ケースクロスオーバー)、SSA; sequence symmetry analysis。時間依存性共変量のアウトカムに対する発生率比は 5.0。\*時間依存性共変量と時間非依存性共変量の交互作用。

表 4 アウトカム発生による追跡終了が存在する場合のシミュレーション実験結果 (Takeuchi et al. BMC Med Res Methodol. 2018;18:4.)

	追跡終了確率	バイアス	平均二乗誤差	被覆確率 (%)
SCCS	0.3	0.0055	0.0104	93.9
CCO		0.0053	0.0426	94.2
SSA		0.3799	0.2206	72.8
SCCS	0.6	0.0202	0.0113	93.1
CCO		0.0053	0.0426	94.2
SSA		0.9647	1.0630	0.1
SCCS	1.0	0.0242	0.0136	90.3
CCO		0.0053	0.0426	94.2
SSA		NA	NA	NA

SCCA; self-controlled case series (自己対照ケースシリーズ)、CCO; case-crossover (ケースクロスオーバー)、SSA; sequence symmetry analysis。

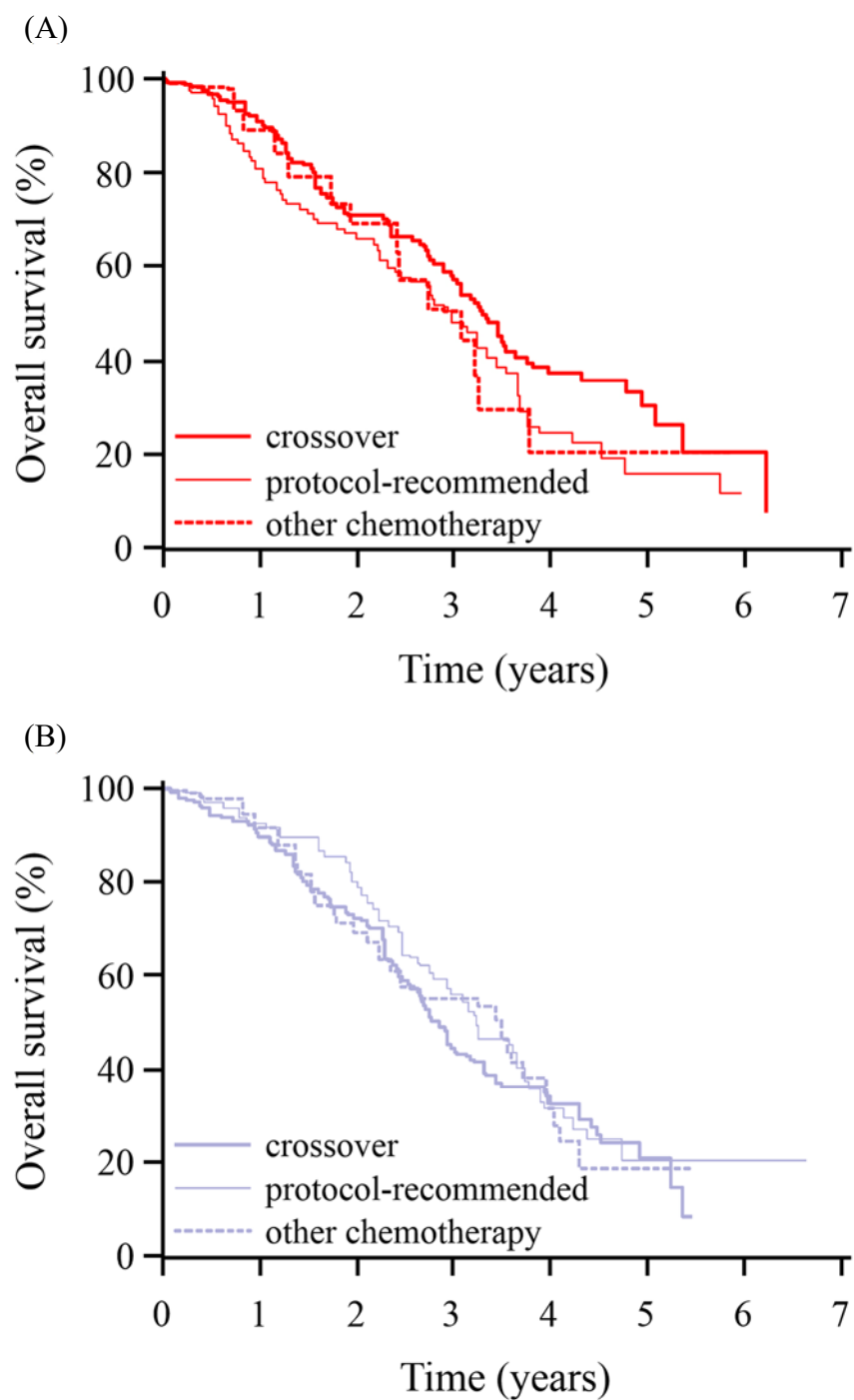


図1 1次治療と2次治療で定義された動的治療レジメンのもとでの推定生存曲線 (Mukai et al. *Oncology*. 2017;93:315-322.)。 (A) 1次治療がタキサン系薬剤の場合、 (B) 1次治療がTS-1の場合。



平成 29 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

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 笹渕裕介 自治医科大学データサイエンスセンター 講師

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、一部少人数に対して試行した。試行の結果、受講者は自分自身でデータベースの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った。

A. 研究目的

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を作成する。

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。

C. 研究結果

教育用データベースを作成し、これを用いた教育プログラム案を作成、少人数に対して試行した。受講者は自身の研究データの加工・統計解析が可能となった。

具体的なプログラム案

(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する。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。

(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 を、また EZR から R へと移行することでより高度な分析を行うことが可能となるという理由から、R (EZR) を利用する。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。

(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。

(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データサイエンティスト教育プログラムの案を作成し、一部試行した。このプログラムにより臨床家と疫学・統計学専門家の両方とコミュニケーションが取れる人材を育成するという点において、十分に機能すると考えられる。

E. 結論今年度、医療ビッグデータ研究の為に必要な技術を養成するため

の教育プログラムを作成した。少人数への試行によりプログラムが機能することが明らかとなった。

F. 健康危険情報  
なし

#### G. 研究発表

1.論文発表

なし

2.学会発表

なし

#### 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 29 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 京都大学医学研究科 健康情報学分野 教授

要旨： 医療科学におけるビッグデータ解析においてレセプト、すなわち診療報酬明細の活用はその大きな柱の一つとなっている。厚生労働省は特定健診・特定保健指導情報（2008 年度～）とレセプト情報（2009 年度～）を含むレセプト情報・特定健診等情報データベース（NDB: National Database of Health Insurance Claims and Specific Health Checkups of Japan）を構築している。NDB は 2017 年時点で日本における保険診療の 95%以上の請求情報が含まれる。2011 年より医療サービスの質の向上等を目指した正確なエビデンスに基づく施策の推進等を目的とした分析・研究に対し、有識者会議での審査を経て政策立案者や研究者が用いることが可能となった。初年度の本課題では、NDB の利用経験に基づき、その活用を巡る諸課題の概観を行った。

**研究協力者**

加藤源太 京都大学医学部附属病院.  
診療報酬センター 准教授

**A. 目的**

レセプト情報・特定健診等情報データベース（NDB: National Database of Health Insurance Claims and Specific Health Checkups of Japan）は、特定健診・特定保健指導情報（2008 年度～）とレセプト情報（2009 年度～）から構成されている厚生労働省管理のデータベースである。2017 年時点で日本における保険診療の 95%以上の請求情報が含まれ、2011 年より医療サービスの質の向上等を目指した正確なエビデンスに基づく施策の推進等を目的とした分析・研究に対し、有識者会議での審査を経て政策立案者や研究者が用いることが可能となった。

初年度の本課題では、NDB の利用経験に基づき、その活用を巡る諸課題

の概観を行った。

**B. 研究方法**

文献的検討と厚生労働省・健康医療分野のデータベースを用いた戦略研究（2014～16 年度）における NDB 活用の事例検討（高齢者医療の適正化推進に向けたエビデンス診療ギャップの解明：既存データベースを利用した京都大学オンサイトセンターにおけるレセプト情報等データベース（NDB）の活用方策の検討 [代表・中山健夫]）。

**C. 研究結果**

**1. データベースに格納されている情報**

NDB は大きく分けて「レセプト情報」と「特定健診・特定保健指導情報」によって構成されており、レセプトには氏名や生年月日・性別などといった患者情報、医療機関や保険者、被保険者等の情報、保険診療に関連する傷病名や治療内容、投薬等の情報が含

まれている。また個人を特定する ID として、主に保険者、被保険者関連情報を元にしたハッシュ値 (ID1) と、主に氏名情報を元にしたハッシュ値 (ID2) の、2 種類のハッシュ値が個々のレセプト情報に付与されている。この ID1,2 を用いることで、医療機関、各月ごとに発行されているレセプト情報を同一人物の情報として特定することが可能となる。特定健診情報は、40 歳以上 75 歳未満の被保険者・被扶養者を対象とする内臓脂肪型肥満に着目した問診結果や生活習慣病に関連した測定項目の結果、あるいは特定保健指導については保健指導レベルや支援形態などの情報が含まれており、レセプト情報と同様に ID1,2 が付与され格納されている。格納されているデータの詳細については、厚生労働省のホームページから参照できる。

## 2. 利用形式

NDB データは様々な利用形式が設けられているが、実際に利用者がデータを操作する形式としては、大きく分けて主に「特別抽出」、「サンプリングデータセット」の 2 つがある。研究者が必要と考えるデータすべての提供を申出する形式は、「特別抽出」と呼ばれている。有識者会議（審査分科会）において承諾されれば、申出者は希望する NDB データを入手できる。事前に限定された目的に対する完成度の高い研究プロトコルの準備と、合理的な根拠に基づいたデータ項目の指定が必須となっている。また縦断研究を行えるのは特別抽出のみであるが、ID で統合し、分析できるデータ形式を再構築するなど分析の前処理が必要なため、データ分析に関する人材や物理的分析環境が整っていることが不可欠であり、現実的にはレセプトデータの分析に相当習熟した研究者、研究チームに利用

者は限られよう。

特別抽出では探索的な研究が原則として認められていないことから、そうしたニーズに対応出来るよう整備されたデータが、「サンプリングデータセット」である。これは、単月分のレセプトに対し、性別および 5 才刻み年齢別に、入院で 10%、外来および調剤で 1%の抽出を行い、出現回数の少ない傷病名や診療行為、医薬品情報のダミー化など、匿名性を強化する一定の処理が行われたデータである。サンプリングデータセットは単月分の情報しか含まれていないため、縦断研究が出来ないという欠点があるが、特別抽出に比べるとデータ量が限られており、ID で紐付けする作業も不要であることから、比較的現実的な環境下で操作することが可能である。厚生労働省ホームページにこれまでの利用者による体験報告も紹介されており、各疾患の出現頻度や治療実態を横断的に評価することを検討している研究には比較利用しやすいと言える。

## 3. NDB オンサイトリサーチセンター

NDB の利用にあたっては十分なセキュリティ環境の確保も求められているが、その環境を誰しも確保できるとは限らない。そこでセキュリティが確保された環境下で多様な研究者が NDB データにアクセスできることを目的として、NDB オンサイトリサーチセンター（以下「センター」）が 2015 年に設置された。センターは東京大学と京都大学に設置されるとともに、両機関は NDB データの利活用の推進・普及・啓発に関し厚生労働省との連携協力機関として協定を締結することとなり、NDB データの利活用を推進する役割を担うこととなった。

センターでの NDB データ利用はま

ず両機関において試行的に利用が開始され、2016年は主にセンターが備えるシステムの性能を評価するパフォーマンステストが行われた。センターの端末には様々なアプリケーションや統計解析ソフトが用意されているが、目的とする集計・分析によってアプリケーションの適性が異なっていたり、ローカル環境で統計解析する場合とサーバー上で統計解析する場合とで所要時間等に大きな差が出たりするなど、システム性能に様々な特性があることが明らかとなった。現在、具体的な研究課題の実行可能性を評価中であり、今後は両機関以外の研究者等もセンターを利用できるようにするため、活用方策の検討や諸規程の整備が行われる予定となっている。

#### 4. NDB データを研究に使用する際の課題

NDB データの第三者利用は2011年以降徐々に進んでおり、申出件数も2017年3月時点で125件の研究申出に対してデータ提供が承諾されている。一方で、NDB データの更なる利活用を考えた場合、様々な課題が浮かび上がってくる。まず、レセプトは月ごと、医療機関ごと、レセプト種類ごとに発行されるため、特別抽出で研究を行う場合、患者単位でデータを統合する作業が必要となるが、この作業に際しての定形化されたID統合ロジックが未だ確立されていない。現在のID1とID2は、就職先の変更や氏名の変更、記入の揺れなどによって変化が生じるため、結婚や就職などといったライフイベントのたびにNDBデータのIDの精度が減じる。ID統合ロジックが確立されていないため、研究者によって得られる結論が異なるという懸念が付き纏っている。このことは、縦断研究のハードルを高くしていると共に、デー

タ分析に多大な年月を必要とする事となるため、理論上診療後2~3か月でデータベース化されるはずのNDBデータの鮮度を結果的に落としてしまうことになっている。

またレセプトデータの構造や特性、効果的な分析手法や活用方法も、研究目的での二次利用という観点から見て、未だ知見が網羅的に蓄積されているとは言い切れない。患者単位でのデータ統合ロジック以外にも、いわゆる「レセプト病名」の存在など、傷病名の精度を揺るがせる要因もあり、適切にデータを抽出する手法が未だ確立されていない。特定の疾患の患者を抽出する際には、傷病名情報以外に、行われた診療行為の情報や処方された医薬品の情報を活用することで一定程度バリデートすることが可能ではあるが、それら手法についても確立されたものは存在しておらず、各研究者が手探りでやっているところである。研究目的でNDBデータを利用してきた者が、今後切り拓いていかなければならない課題であろう。

米国ではMedicare, Medicaidデータの研究利用が盛んに行われているが、データ利用時は、データを管理するCMSと契約関係にあるResearch Data Assistance Center (ResDAC) という組織が、利用申請手続きに関する相談や研究者からの学術的質問を受け付けている。日本では一部業務は外注されているものの、利用者に向けたデータ利用を支える学術的支援体制が十分に確立されているとは言い難く、現時点では一部公開されている関連資料を頼ったり、あるいは既に研究利用の知見を有している研究者と共同でデータ利用を行ったりするなど、現実的には分析のノウハウを何らかの形で確保しなければ、NDBデータの十全な利活用は容

易ではない。

## 5. 厚生労働科学・戦略研究における活用事例

本課題で行ったがん治療領域と死亡に関するレセプトのバリデーションと、NDB を利用した潜在的不適切処方 (PIM) , 慢性腎臓病 (CKD) 診療, 高齢者の終末期の研究について, その概要と抽出された課題は以下の通りであった。

### ① 消化器がん患者における治療実態調査に関する研究

使用データベース：2014年4月から2017年3月までのレセプト情報 (医科入院外, 医科入院, 調剤, DPC)

データ抽出の条件・方法：2014年6月から2015年3月の10ヶ月間に新たに消化器癌 (食道がん, 胃がん, 小腸がん, 大腸がん, 肝臓がん, 膵臓がん, 胆道がん) の病名が付与された患者の, その後2年間の全国のレセプト情報を調査する。主な調査項目は, 患者背景, 診断・治療試行施設状況, 治療内容 (手術, 放射線療法, 抗がん薬治療) で, 特に年齢 (高齢者と非高齢者) や病院機能 (がん拠点病院と非がん拠点病院) による治療内容の差異に注目して解析を行う。

主な結果：全国のレセプト情報を解析するにあたり, 事前検討として, 院内のレセプトデータと診療録データを用いてがん病名および薬剤 (抗がん薬) についてのバリデーションを行った。レセプト病名は保険病名や類似病名, 疑い病名などの問題が生じるため, 実際の診療情報を反映していない可能性がある。そこで, 院内レセプトデータに記載されているがん病名について, 院内がん登録に登録された病名をゴールドスタンダードとし, 病名一致率について検討をおこなった。結果, 病名一致率はがん種により差が

あることがわかり, がん種に占める疑い病名の割合や解剖学的な問題, 対象期間外におけるレセプト病名の登録などが影響していると考えられた。次に, 抗がん薬についてレセプトデータと診療録データを用いて突き合わせを行ったところ, 一致率が低いことがわかった。この原因として, 診療録データから得られる薬剤情報が YJ コードやレセ電コードといった汎用の薬剤コードと対応していないケースが多数あることがあげられた。これらの結果もふまえ, 現在全国のレセプト情報の解析を開始している。

考察：レセプト情報は, 莫大な医療情報が得られることから, がんの診療実態の把握には有用なデータベースであると考えられる。一方で, 臨床データなど患者の病態を表すデータが含まれていないこと, 転帰が不明のため治療効果がわからないこと, 実際の診療情報を反映していない可能性があることなど, データの性質上の制限があり, これらの問題を解決するためには診療情報の二次利用を活性化する体制の整備が望まれる。

### ② 高齢患者における高齢者の安全な薬物療法ガイドライン (STOPP リスト) ならびにビアーズリスト収載薬の使用実態に関する研究

使用データベース：NDB (特別抽出)

データ抽出の条件・方法：日本における処方薬との紐付けを実施したビアーズリスト、高齢者の安全な薬物療法ガイドライン (STOPP リスト) に掲載されているいずれかの薬剤1か月以上の処方について、厚生労働省より収受した NDB データを用いた分析を行った。

主な結果：リスト掲載薬の処方に関して、高齢患者における1年間での

期間有病割合の推定に成功した

(STOPP リスト：59.9%、ビアーズリスト 32.5%)。STOPP リストに関しては、ベースライン (2010 年 10 月) 時点で STOPP リスト掲載の薬剤の処方が無い者を対象として、その後 4 年 6 ヶ月追跡した新規発生率は 100 人年あたり 22.2 例であった。

考察：今回はリスト掲載薬処方に関する分析のみであり、結果の解釈には細心の注意が必要であるが、NDB データを用いた、高い外的妥当性が担保された状況において、これらの薬剤に関する基礎的な記述疫学を明らかにすることに成功した。今後も医療の現状をも加味した上で、判断の精度を高めるなどの更なる検証が望まれる。

苦労した点：65 歳以上の悉皆データ 5 年分を用いた ID1 と ID2 の突合作業が一番困難を極めた。

③ 高齢者における慢性腎臓病診療の質に関する研究

使用データベース：NDB (特別抽出)

データ抽出の条件・方法：

1) 対象患者

\*CKD 有病者集団

60 歳以上で、2010 年 4 月から 9 月までの 6 か月間のベースライン期間で、糸球体疾患、腎尿細管間質性疾患、腎不全、腎尿細管機能障害、処置後腎不全、腎障害を伴う糖尿病、高血圧性腎疾患の CKD 関連病名の ICD10 コードが付与されている患者を CKD 有病者集団として定義した。また、本研究は保存期の CKD 患者を対象としているため、ベースライン期間中に人工腎臓維持透析管理料、在宅事故腹膜還流指導料の透析療法を示す処置管理コードが付与されている患者は除外した。

\*CKD 新規診断集団

60 歳以上で 2010 年 4 月から 2014 年 9 月までの間に、新規で CKD 関連病名の ICD10 コード (前述) が付与された患者を CKD 新規診断集団として定義した。

2) 説明変数

CKD 診療の質指標を用いて患者単位で受けている CKD 診療の質をスコア化した高齢者に適応可能でクレイムデータのみで測定可能な以下の 3 項目を利用した。

定期的な尿検査：6 か月間で 2 回以上の尿検査実施コード付与あり

栄養指導：6 か月間に処置管理コードで栄養指導管理料の付与あり

NSAIDs 常用の回避：NSAIDs 処方日数が 1 か月あたり 14 日未満を 6 か月継続

CKD 診療の質スコアは患者単位の遵守項目割合として、遵守項目数/3 項目で算出した。

CKD 診療の質の定義は、CKD 有病者集団の場合、ベースライン期間終了後の 6 か月間とした。CKD 新規診断集団の場合、新規診断を受けてから 6 か月間とした。

主な結果：

\*CKD 有病者集団

診療の質指標間で順守割合には違いがあり、DM 患者と非 DM 患者の間でも違いを認めた。診療の質スコアと末期腎不全発生の関連は非調整 Cox、調整 Cox では糖尿病では診療の質指標が防御的に働いていたが、非糖尿病ではこの傾向が見られなかった。一方で、未測定交絡の影響を考慮した操作変数法では糖尿病、非糖尿病でも診療の質指標が防御的に働いていた。

\*CKD 新規診断集団

診療の質スコアと末期腎不全発生の関連は CKD 有病集団と同様の結果で

あった。

考察：CKD 診療の質評価の実施は、患者のアウトカムを改善に寄与する可能性がある。

苦勞した点：データセットの作成と固定に多大な時間と労力がかかった。特に、縦断的な解析を行うにあたっての ID の突合方法を考案、実行するのに時間がかかった。末端においてこの突合を行うにはハード面での十分な準備が必要である。また、提供されている ID がもとの情報に処理を加えずに匿名化を行っていることから、誤字や入力ミス等に対応できないものとなっている。これについては、Levenshtein 距離などを用いて匿名化前に確率的に個人を同定する ID を付与すべきである。今後、医療 ID の導入を行うことが予測されているとしても、既存データの活用を行うという観点では検討すべき事項と考える。

#### ④ 高齢がん終末期医療の実態に関する研究

使用データベース：NDB（サンプリングデータセット）

データ抽出の条件・方法：2012 年 10 月に死亡した 65 歳以上のがん患者を対象とし、先行研究にしたがって皮膚がんは除外した。がん患者は、国際疾病分類第 10 版 ICD10（C00-C96のうち C43-C44 を除外）をレセプト電算処理システムで使用する傷病コードと突合することにより同定した。死亡前における医療の利用実態を調べるために、死亡 30 日前に入院した患者を対象とした。

主な結果：2012 年 10 月の NDB サンプリングデータセットに含まれる延べ 142270 名の入院患者のうち、延べ 1873 名が病院で死亡していた。この集団の中で、延べ 1023 名（54.6%）

は死亡する 30 日以内に入院し、平均在院日数は 13.6 日（標準偏差 8.3 日）であった。

考察：死亡 30 日以内に入院したがん患者は、他の先進国に比べて在院日数が長く、一人あたりの入院費用が高いことが示唆された。

苦勞した点：本研究では解析対象集団の設定に苦勞した。サンプリングデータセットからは、ひと月分のレセプト情報しか得られないので、入院中の診療行為や請求点数を調べたい場合には入院期間に月の変わり目を含むケースは解析の対象外とせざるを得ない。さらに、月の変わり目を含まない入院レセプトであっても、入院日が月末に近づくほど在院日数の短いレセプトの情報しか得られない。つまり、すべての入院日（本研究では 10 月 1 日から 31 日まで）を対象とすると、入院期間が短いレセプトが実際よりも多い集団となってしまう。そのため本研究では、診療行為と請求点数の解析対象を入院日が 10 月 1 日および 2 日のレセプトに限定した。サンプリングデータセットでも、ひと月分だけでなく、2 ヶ月分以上のデータが提供されるようになるとデータ活用の可能性が広がると思われる。

#### ⑤ 高齢者終末期医療の実態とレセプト上での死亡情報の妥当性検証

使用データベース：

1) レセプトの死亡転帰情報の妥当性検証

日本医療データセンターのレセプトデータベース

2) 終末期高齢者の診療内容の記述 NDB（サンプリングデータセット）

データ抽出の条件・方法：

1) レセプトの死亡転帰情報の妥当性検証

対象は 65～74 歳の高齢者、2012 年 9 月



～2015年8月診療分の医科入院・入院外・DPCレセプト、健保組合加入者台帳データ（台帳）を用いた。台帳に記録された、死亡による資格喪失情報をゴールドスタンダードとし、レセプトの死亡転帰情報の感度、特異度を算出した。

2) 終末期高齢者の診療内容の記述  
対象は85歳以上の高齢者、2013年10月、2014年10月、2015年10月診療分の医科入院・DPCレセプトを用い、死亡前7日間の積極的治療の実施割合を算出した。

主な結果：

(1)レセプトの死亡転帰情報の妥当性検証

分析対象数は入院患者1,541例、外来患者90,780例。レセプトの死亡転帰の感度・特異度は、入院で93.0%・87.3%、外来で40.3%・99.9%。

(2)終末期高齢者の診療内容の記述

分析対象数はDPC1,362例、医科入院3,671例、合計5,033例。死亡前7日間の心肺蘇生は7.4%、挿管は2.8%、中心静脈栄養は6.7%、経鼻栄養は8.6%、胃瘻は1.3%実施されていた。

考察：高齢者の入院レセプトの死亡転帰情報の妥当性は高いが、対象は65～74歳に限定され、一般化できない。外来レセプトの死亡転帰の妥当性は低く、NDB外来レセプトを用いた外来診療の実態評価に限界があると考えられ、今後、NDBと外部データを突合可能な仕組みの検討が課題と考えられる。

苦労した点：NDBは外部データと直接突合した妥当性検証が許可されていない

ため、職域保険加入者のレセプトデータを用いたバリデーションが必要であった。

## D. 考察 & E. 結論

現時点ではさまざまな課題を持つNDBであるが、日本のほぼ全人口を対象としたレセプトの最大規模のデータが集積されている貴重な情報源であることは確かである。2017年度にはAMED臨床研究等ICT基盤構築研究事業において、「新たなエビデンス創出のための次世代NDBデータ研究基盤構築に関する研究」（代表・黒田知宏）が単年度の研究事業として実施され、人材育成プログラムの雛形が作成されている。その成果の活用も視野に入れて、次年度は、これまで蓄積されたNDBの使用経験に基づき、新たな研究者の参入を進め、多様な研究を活性化していくための教育プログラムの開発を進めたい。

F. 健康危機情報  
なし

## G. 研究発表

中山健夫. National Database：厚生労働科学・戦略研究の経験から. 日本臨床疫学会第1回年次学術大会 シンポジウム1 ビッグデータを用いた臨床疫学研究（東京大学）2017年9月30日

H. 知的所有権の取得状況  
なし

平成 29 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、オンサイトセンターで National Database (NDB)からデータを抽出する上で必要な作業に関連して、具体的な 4 つの課題を取り上げ検証した。課題（1）として、NDB データベースにアクセスするための SQL 言語、ここでは特に Oracle SQL を学ぶ手順を検証した。課題（2）として、オンサイトセンターでは通常のプログラミング作業でイメージされる、現場でコーディングを検討しながら、実際のデータにアクセスし、トライ&エラーで作業を進めて行くことは現実的ではないことから、NDB のサンプルデータの入った ODB の仮想環境をフリーのアプリケーションのみでローカルに構築することを試みた。課題（3）として、NDB からデータを抽出する SQL コードを、後発医薬品使用割合の分析を実例としてテストデータを用い解析を行った。課題（4）として、実行時間を短くすることを目的とした SQL のパフォーマンスを改善する方法を検討した。本分担研究で構築した仮想環境においては、オフラインで利用可能であり、各利用者が独立した環境を利用することが出来、さらに、簡単に環境の立ち上げや削除が可能になるため試行錯誤がし易く、無料で環境構築が可能であるという利点がある。今後、仮想環境のもとで、解析目的に即したコードの作成、パフォーマンスの改善の検討などを行う予定である。本分担研究で作成された資料や仮想環境は、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。

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、オンサイトセンターで National Database (NDB)からデータを抽出する上で必要な作業を検証し、実例を示す。ここで作成された資料は、NDB からデータを抽出する作業を実行可能な高度人材を育成するプログラムの基礎となる。NDB

からデータを抽出するに当たって、具体的な 4 つの課題を取り上げた。課題（1）NDB は Oracle Database (ODB)に格納されており、SQL\*Plus でデータベースにアクセスする仕様になっているため、SQL 言語、ここでは特に Oracle SQL、を学ぶ必要がある。そこでまず始めに、どのように SQL 言語を学ぶ事が出来るかを簡単にまとめる。課題（2）オンサイトセンターでは、通常のプログラミング作業でイメージされる、現地でコーディングを検討しながら、実際のデータにアクセスし、トライ&エラーで作業を進め

て行くことは現実的ではない。これには、(a)インターネットでコーディングの方法を検討しながらの作業は出来ない、(b)設置されているパソコンにコードを入力する際に、慣れたテキストエディタを利用する事が出来ない、(c) オンサイトセンターは共有の施設であるため、長時間の独占的な利用は望ましくない、という3つの理由が挙げられる。そこで、NDBのサンプルデータの入った ODB の仮想環境を、フリーのアプリケーションのみで、ローカルに構築する必要がある。AWS などを利用して NDB のサンプルデータの入った ODB のサーバーの準備をすることは可能だが、その場合と比較して、ローカルに仮想環境を構築するほうが、(a)オフラインでも利用可能である、(b)各人が独立した環境を利用することが出来る、(c) 簡単に環境の立ち上げ、削除が可能になるので試行錯誤がし易い、(d)無料で環境構築が可能である、という利点がある。そこで、SQL コードの実行のテストまで行うことが可能で、コードをそのままオンサイトセンターで利用するスキーマとする仮想環境を構築する。

課題(3) 実際に NDB からデータを抽出する SQL コードを、後発医薬品の使用割合の推移を求めることをテーマに、実例を示す。ここでは、上述した仮想環境において、テストデータを用い解析を行った経過を記す。

課題(4) NDB はデータサイズが非常に大きく、全てのデータを用いた作業は時間の制約上難しい可能性が高い。全てのデータを用いず、ランダム抽出でのスキーマとしたとしても、可能な限りランダム抽出の比率を上げ、実行時間を短くすることは非常に重要である。そこで最後に、SQL の

パフォーマンスを改善する方法について検討を行う。

## B. 研究方法

(1) 課題(1)のため、SQL 言語、特に Oracle SQL を学ぶための有用な資料、ネット上のサイトを収集、探索した。

(2) 課題(2)のため、SQL コードの実行のテストまで行い、コードをそのままオンサイトセンターで利用出来るようにするための仮想環境を構築した。今回は、macOS High Sierra で環境を構築した。

(3) 課題(3)のため、実例として、後発医薬品の使用割合の推移を求めることを目的として、上記の仮想環境においてコードの実装を行った。

(4) 課題(4)のため、SQL のパフォーマンスを改善する方法を検討した。

## C. 研究結果

(1) Codecademy というウェブサイトでインタラクティブに SQL coding の基礎を学ぶことのできるサービスが無料で提供されていた。また、「スッキリわかる SQL 入門」(中山清喬・飯田理恵子著)で、SQL の基本的なコーディングから、テーブルの設計の基礎までを網羅的に理解することができる。次に、「SQL パズル 第2版」(ジョー・セルコ著)は問題形式で SQL を学ぶ参考書であり、有用であった。その問題を解く際には「達人に学ぶ SQL 徹底指南書」(ミック著)が参考となった。

OS が macOS High Sierra であれば、後述する、ODB の仮想環境を NDB のサンプルデータ挿入の手前まで実行することで、Oracle SQL の練習環境を構築する事ができる(別添資料1

参照)。また、Oracle 社が提供している Web 上のサービスで Oracle Application Express というサービスがあり、自らの PC に環境構築せずとも、SQL code を web 上で実際に走らせることが可能である。なお、本報告では扱わないが、Oracle SQL 以外の SQL では、MySQL (<https://www.mysql.com/jp/>) が無料かつ容易に環境を構築でき、Oracle SQL とある程度のコードの互換性がある。

(2) ローカル仮想環境の構築のための作業マニュアルを別添資料 1 に示す。作業マニュアルに従っていけば、特に前提知識を必要とせず環境を構築出来るように設計したが、利用するアプリケーションのバージョン変更等に伴う不具合への対応には UNIX や Docker 等の知識を必要とする。

ここからは、作業マニュアル(別添資料 1)の手順を概説する。Docker と呼ばれる、コンテナ型の仮想化環境を提供するオープンソースソフトウェアを利用した。Docker では、コンテナと呼ばれる、アプリケーションを独立したプロセスとして実行する環境を整え、仮想化環境を実現する。このコンテナを作成する際に必要となるファイル群のテンプレートであるイメージを、公開レポジトリから取ってくることで、仮想環境上に ODB を構築する。ODB の構築のために必要なイメージは、Oracle が公式にリリースしているものがあるので、それを利用する。このイメージをローカルにダウンロードすることを、Docker では、ビルドと呼び、まずはその作業が必要になる。このイメージに加えて、ODB の Linux 版(フリ

ーで入手可能)のダウンロードが必要となる。このビルドしたイメージを起動させれば、コンテナが構築され、ローカルで ODB が動き始める。

起動した ODB へのアクセスは、オンサイトセンター同様、SQL\*Plus を利用する。Oracle の公式サイトから、Oracle Instant Client (OIC、sqlplus コマンド)を利用可能にするために Oracle が提供しているアプリケーション)をダウンロードし、データベースに接続するためのいくつかの初期設定を行う。sqlplus コマンドで接続し、データベースの設定やサンプルデータの挿入を行う。この作業は多くのコマンドを必要とするが、1つのシェルスクリプトにまとめたので、簡易に実行出来る。これでテスト環境が整ったので、sqlplus コマンドでアクセスし、SQL を実行する事が可能である。終了や再起動もシェルスクリプトにまとめてあるので、少ないコマンドで簡易に実行出来る。

(3) 上述した仮想環境において、テストデータを用い解析を行った。

#### 1. 解析の目的

SQL のコーディングの実践として、高血圧症・脂質異常症・糖尿病・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して、テストデータを対象としたものを次項に記す。

#### 2-1. 抽出条件 (NDB テストデータ)

##### a. 期間

平成 28 年 5 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、傷病名を有するもの
- d. 薬剤の成分名  
ドネペジル

## 2-2. 抽出条件に関する実データを取り扱う場合との差異

・今回の解析では、テストデータとして平成28年5月診療分に設定したものをういたため、期間を限定しているが、実データを用いた解析では、複数年度にわたる解析を行う予定である。

・今回の解析では、全ての傷病名コードが認知症を示すと仮定して、解析を行った。実データを用いた解析では、高血圧症・脂質異常症・糖尿病・認知症を示す傷病名を有するレセプトのみを取り扱う予定である。

・今回の解析では、認知症患者におけるドネペジルの処方について解析を行った。実データを用いた解析では、高血圧症患者におけるロサルタン・カンデサルタン・アムロジピンの処方、脂質異常症患者におけるシンバスタチン・アトルバスタチン・ピタバスタチンの処方、糖尿病患者におけるグリメピリド、ボグリボース、ピオグリタゾンの処方、認知症患者におけるドネペジルの処方について解析を行う予定である。

## 3. 抽出する統計量

別表1（別添資料2）の形での表が作成できる形での数値の抽出を目的とする。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・月ごとに求める必要がある。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う。

## 4. SQL コードのコーディング

コーディングの概要は以下のとおりである。第一に、テストデータを生成した。次に、抽出条件をみたすレセプトを抽出し、テーブル間のマージを行った。最後に、抽出したレセプトから統計量を計算した。コーディングについては、結果の整合性等の検証のため、3つのコード（コードA, B, C）を作成した。

3つのコード（コードA, B, C）の比較結果については以下のとおりである。コードAでは、まず、抽出条件をみたすレセプトに対して、性別・年代別・地域別・レセプト種別でGROUP BY をかけることによって、目的とする統計量の計算のもととなるVIEWを作成した。そのVIEWを用いて、目的とする統計量をさらにGROUP BY することにより計算した。

コードBをコードAと比較した場合の特筆すべき違いは、(1)サブクエリを、VIEWを利用して記載する事でコードの可読性を高めている、(2)最終的なVIEWが、1タプルを処方箋/医薬品単位として構成している事、の2点である。

コードCをコードAと比較した場合の特筆すべき違いは、対象疾患をもつ個人の呼び出しにおいて、

「GROUP BY」ではなく、DISTINCT句を用いたことである。これにより、同一個人が複数の病名を持つ場合は複数回カウントされることになる（例えば、糖尿病と脂質異常症がある場合は2回）が、サマライズで対象薬剤ごとにGROUP BY する際に（今回は対象とした薬剤と病名が1対1対応になるようなコード作成をしているため）単回カウントになり問題ない。この違いが処理速度にどのような違いを生み出すかは今後の検討

課題である。また、CASE 句を用いたために冗長ではあるが、70歳-74歳の自己負担額が年度により10%もしくは20%になっている例を反映させることができている。

(4) 実行時間の短縮には、スケールアップ・スケールアウトといったシステムリソースの増強という手段もあるが、リソースの増強には費用も時間もかかるため、まずは効率の最大化を図るチューニングを行うことが必要である。実行時間は扱うデータ量にも大きく左右されるが、本研究では相当量のデータを用いるため、パフォーマンスのチューニングによる分析時間の短縮が特に求められる。しかし、本研究環境では、リソースのチューニングや Oracle インスタンスのチューニング、また表や索引の設計の変更などが難しかったため、実行計画に基づく SQL の見直しや、「基本的には EXISTS 句を使う」と言ったような SQL のベストプラクティスに従う、という最低限のパフォーマンスチューニングに留めた。

#### D. 考察

課題(1)の SQL 言語については、標準 SQL という規格があるものの、個々の SQL 言語により、記法に若干の違いがある事に注意しなければならない。Oracle SQL は練習環境を準備するために、Oracle Database を何らかの形で用意しなければならないため、他のメジャーな SQL 言語(MySQL、PostgreSQL、SQLite など)と比較して、オフラインで練習環境を準備するのが難しく、課題(2)を検討する必要があった。

課題(2)については、複雑なコマンドもシェルスクリプトにまとめる

ことで、作業マニュアルに従っていけば、特に前提知識を必要とせず環境を構築出来るように設計した。本研究で実施した環境構築に当たっては、複数のファイルをダウンロードする必要があり、現時点では、本プロジェクト用の GitHub のプライベートレポジトリで管理しているため、プロジェクト外の個人がアクセスすることは出来ないが、NDB のサンプルデータを除けば、全てフリーでアクセス出来る資料を用いて作成されているため、NDB のサンプルデータが利用可能であれば、本研究成果を用いて利用可能である。本研究で利用した Docker は、VM 系の仮想化と比べ軽量なため、環境作成にかかるコンピューショナルなコストが低い。また、Windows や Linux にも対応しているので、本研究と同様の手順で Mac 以外の OS への応用も可能と考えられるが、これは今後の課題である。

課題(3)については、データ抽出を目的とした相異なる3つのコードを作成した。同一の最終生成物を得ることを目的としたが、個人の呼び出しの方法・集計のタイミング・VIEW を作成するタイミングなどにおいて、さまざまな差異のあるコードとなった。NDB のような大規模のデータを取り扱う場合には、細かなコーディングにおける差異が実行時間を大幅に変える可能性がある。今後、各コードのパフォーマンスの比較を検討する予定である。

課題(4)については、今後、与えられたデータベースの再設計なども視野に入れ、Oracle インスタンスの初期化パラメータの調整や、表、索引の設定、設計を再検討による更なるパフォーマンスの向上を目指す。

#### E. 結論

第一に、NDB が格納されている、ODB にアクセスするために、SQL 言語、特に Oracle SQL を学ぶための方法を検討した。第二に、ODB のローカル仮想環境を、macOS High Sierra で Docker を利用して構築した。第三に、実例として、後発医薬品の使用割合の推移を求めることを目的として、上記の仮想環境においてコードの実装を行った。最後に、SQL のパフォーマンスを改善する方法を検討した。

#### F. 健康危険情報

なし

#### G. 研究発表

#### 1. 論文発表

なし

#### 2. 学会発表

伊藤佑樹、原湖楠、小林廉毅. 先発／後発医薬品の差額通知書が個人の後発医薬品の選択に与える影響. 日本公衆衛生学会総会、鹿児島、2017年11月1日

原湖楠、伊藤佑樹、小林廉毅. 先発／後発医薬品の差額通知書送付政策が後発医薬品シェアの推移に与える影響の検討. 日本公衆衛生学会総会、鹿児島、2017年11月1日

#### H. 知的財産権の出願・登録

なし

# Oracle DBの環境構築 (Mac OSX)

## 0. Overview

- Oracle DBはMac OSXに対応していないので、Dockerを使ってOracle DBのimageを入手し、Containerを起動する。
- sqlplusを使ってOracle DBに接続し、サンプルデータを挿入する。
- 必要なスペック: ストレージの空き容量30GB以上、メモリ16GB以上での動作は確認済み。
- 2018/02/09にmacOS High Sierra version 10.13.3、Docker version 17.12.0-ce-mac49 (21995)での動作を確認。

## 1. Docker for Macのインストール、設定

### Reference

- [Get started with Docker for Mac | Docker Documentation \(https://docs.docker.com/docker-for-mac/\)](https://docs.docker.com/docker-for-mac/)
- [Docker入門 \(全11回\) - プログラミングならドットインストール \(http://dotinstall.com/\)](http://dotinstall.com/)

### インストール、設定

- [Install Docker for Mac | Docker Documentation \(https://docs.docker.com/docker-for-mac/install/\)](https://docs.docker.com/docker-for-mac/install/)  
でStable channelをインストールする。
- Docker -> Preferences -> AdvancedでMemoryを8.0GBに変更する。

## 2. OracleDBのdocker-imageのbuild

### Reference

- [docker-images/OracleDatabase at master · oracle/docker-images \(https://github.com/oracle/docker-images/tree/master/OracleDatabase\)](https://github.com/oracle/docker-images/tree/master/OracleDatabase) (公式)
- [OS X 10.11で、Oracle Database 12cのDocker Imageを使ってみた - Qiita \(http://qiita.com/lethe2211/items/0bb493fa93a0088cfac9\)](http://qiita.com/lethe2211/items/0bb493fa93a0088cfac9)
- [公式 Oracle Database の Docker イメージを構築 - #chiroito's blog \(http://chiroito.hatenablog.jp/entry/2016/12/28/235627\)](http://chiroito.hatenablog.jp/entry/2016/12/28/235627)

### 事前準備

- 作業ディレクトリの作成

WorkspaceにOracle/product/databaseというフォルダを作成し、そのフォルダにworking directoryを設定する。

Workspaceは各自好きなディレクトリを設定する。ここでは、例としてホームディレクトリにWorkspaceというフォルダを作成する。



```
$ cd $HOME
$ mkdir Workspace
$ cd Workspace
$ mkdir -p Oracle/product/database
$ cd Oracle/product/database
```

- Oracleの公式docker-imageのgitをクローン

```
$ git clone https://github.com/oracle/docker-images.git
$ cd docker-images/OracleDatabase/dockerfiles
```

gitコマンドを初めて使う場合には、Xcodeのインストールを求められるので、指示に従う。

- Oracle DBのファイルをダウンロード  
[Oracle Database 12c Release 2 for Linux x86-64 Downloads \(http://www.oracle.com/technetwork/database/enterprise-edition/downloads/oracle12c-linux-12201-3608234.html\)](http://www.oracle.com/technetwork/database/enterprise-edition/downloads/oracle12c-linux-12201-3608234.html)  
から、linuxx64\_12201\_database.zipをダウンロードしておく。  
ダウンロードの際に、Oracleのアカウントを作成する必要がある。  
ダウンロードフォルダにそのまま置いておき、zipファイルは解凍しないように注意する。
- Oracle DBのファイルをダウンロードフォルダから指定の場所に移動

```
$ cp $HOME/Downloads/linuxx64_12201_database.zip 12.2.0.1/
```

## Imageをbuild

```
$ ./buildDockerImage.sh -v 12.2.0.1 -e -i
# PCのスペックによってはかなり時間がかかる
# 以下のように表示されたら終了
Oracle Database Docker Image for 'ee' version 12.2.0.1 is ready to be extended:
--> oracle/database:12.2.0.1-ee
Build completed in 395 seconds.
```

- 以下のコマンドで無事imageがbuildされていることを確認できる。下のように表示されれば成功。

```
$ docker images
```

REPOSITORY	TAG	IMAGE ID	CREATED
oracle/database	12.2.0.1-ee	5b3356b0fe40	3 weeks ago
oraclelinux	7-slim	08a01cc7be97	2 months ago

## 3. Oracle Instant Clientの設定

- sqlplusコマンドを用いてOracle DBへの接続する

## Reference

- [Oracle Instant Client Downloads](http://www.oracle.com/technetwork/jp/database/features/instant-client/index-097480.html)  
(<http://www.oracle.com/technetwork/jp/database/features/instant-client/index-097480.html>)
- [Mac OS XからOracle Databaseへ接続する方法あれこれ | 正論紅茶](https://www.tank-sakurai.com/mac-os-x-oracle-database/) (<https://www.tank-sakurai.com/mac-os-x-oracle-database/>)
- [OS X 10.11で、Oracle Database 12cのDocker Imageを使ってみた - Qiita](http://qiita.com/lethe2211/items/0bb493fa93a0088cfac9)  
(<http://qiita.com/lethe2211/items/0bb493fa93a0088cfac9>)

## Oracle Instant Clientをダウンロード

[公式サイト \(http://www.oracle.com/technetwork/topics/intel-macsoft-096467.html\)](http://www.oracle.com/technetwork/topics/intel-macsoft-096467.html)から、Instant Clientをダウンロードする。

- instantclient-basic-macos.x64-12.1.0.2.0.zip
- instantclient-sqlplus-macos.x64-12.1.0.2.0.zip

の二つ

解凍すると、instantclient\_12\_1という名称のフォルダになるので、二つのフォルダをマージし、Workspace/Oracle/product/に格納する。

## PATHを通す

\$HOME/.bash\_profileにPATHを通す。

PATHの内容

```
# Setting PATH for Oracle Instant Client
export ORACLE_HOME=$HOME/Workspace/Oracle/product
export SQLPATH=$ORACLE_HOME/instantclient_12_1/
export PATH="$HOME/.rbenv/bin:$PATH:$SQLPATH"
export NLS_LANG=Japanese_Japan.AL32UTF8 # 文字化けを防ぐ
```

このPATHを通す作業には、一般的にvimを使うのが多いので、以下のサイトを参考にする。

[vim入門 \(全18回\) - プログラミングならドットインストール](https://dotinstall.com/lessons/basic_vim)  
([https://dotinstall.com/lessons/basic\\_vim](https://dotinstall.com/lessons/basic_vim))

普段使っているテキストエディタがある場合には、それを使う。

vimを使う場合の具体例を以下に示す。

```
# vimの起動
$ sudo vi $HOME/.bash_profile
```

起動するとインサートモードになっているので、

```
# Setting PATH for Oracle Instant Client
export ORACLE_HOME=$HOME/Workspace/Oracle/product
export SQLPATH=$ORACLE_HOME/instantclient_12_1/
export PATH="$HOME/.rbenv/bin:$PATH:$SQLPATH"
export NLS_LANG=Japanese_Japan.AL32UTF8 # 文字化けを防ぐ
```

と記載する。

escキーでコマンドモードに移り、:wqで保存して終了する。

ターミナルの再起動か

```
$ source $HOME/.bash_profile
```

で設定が反映される。  
以下の作業で、vimを使ってファイルを編集・保存する時も同様。

## sqlplusのuser-nameの設定

sqlplusでは、初めにuser-nameの設定を行う必要がある。  
まず、

```
$ uname -n
```

でlocalのホスト名を確認する。  
ここでは仮に、username.localというホスト名であったとする。

```
$ sudo vi /etc/hosts
```

で、

```
127.0.0.1 localhost
```

となっている部分を

```
127.0.0.1 localhost username.local
```

と編集して保存する。

## sqlplusについて便利な設定

sqlplusを実行すると、\$HOME/oradiag\_<username>というディレクトリが自動生成されるので、それを自動的に削除するよう設定する。

\$ORACLE\_HOMEの直下にnetwork/adminというディレクトリを作る。  
このディレクトリにsqlnet.oraというファイルを作り、

```
DIAG_ADR_ENABLED=off  
LOG_FILE_CLIENT=/dev/null
```

と記載して保存する。  
コマンドラインでは例えば以下のようにする。

```
$ mkdir -p $ORACLE_HOME/network/admin  
$ vi $ORACLE_HOME/network/admin/sqlnet.ora
```

## 4. Docker containerの起動、sqlplusでの接続

### 事前準備

- 作業ディレクトリの作成

WorkspaceにNDB/NDB-UTDPHというフォルダを作成し、そのフォルダにworking directoryを設定する。

```
$ cd $HOME/Workspace  
$ mkdir -p NDB/NDB-UTDPH  
$ cd NDB/NDB-UTDPH
```

- 必要なファイルをクローン

```
$ git clone https://github.com/NDB-UTDPH/ndb-sql.git
$ cd ndb-sql/TestDBSetup
```

- Dockerのclean up

```
$ docker ps
# もし起動しているcontainerがあったら...
$ docker stop CONTAINER_ID

$ docker ps -a
# もしcontainerがあれば...
$ docker rm CONTAINER_ID

# docker内の容量整理
$ docker system prune
```

- スクリプトの実行権限付与

```
$ chmod 755 run_image.sh
$ chmod 755 setup.sh
```

このシェルスクリプトの中身は以下のようにになっている。

run\_image.sh

```
# !bin/bash

docker run --name ndb-dummy -p 1521:1521 -p 5500:5500 -e ORACLE_SID=NDBdummyCDB
-e ORACLE_PDB=NDBdummy -e ORACLE_PWD='password' oracle/database:12.2.0.1-ee
```

'password'は自分で設定する。

setup.sh

```
# !bin/sh

SCRIPT_PATH="$ORACLE_HOME/../../NDB/NDB-UTDPH/ndb-sql/TestDBSetup"

# dcl.sqlでは、権限の設定などを行なっている（詳細は以下に示す）
sqlplus system/'password'@//localhost:1521/NDBdummy @$SCRIPT_PATH/sql/dcl.sql

# ddl_create.sqlでは、テーブルやビュー等の構造を定義している
sqlplus NDB/'password'@//localhost:1521/NDBdummy
@$SCRIPT_PATH/sql/ddl_create.sql

# dml.sqlでは、サンプルデータの挿入を行なっている
sqlplus NDB/'password'@//localhost:1521/NDBdummy @$SCRIPT_PATH/sql/dml.sql

# 00_REFRESH_ALL.sqlでは、マテリアライズビューのリフレッシュ等を行なっている
sqlplus NDB/'password'@//localhost:1521/NDBdummy
@$SCRIPT_PATH/ddl/12_CB_MASTER_VIEW/nengetsu/REFRESH/00_REFRESH_ALL.sql
```

dcl.sql (ddl\_create.sql、dml.sql、00\_REFRESH\_ALL.sqlの詳細は略)

```
create tablespace NDB_USERS1 datafile
'/opt/oracle/oradata/NDBdummyCDB/NDBdummy/ndb_user01.dbf' size 1000M;
create user NDB identified by 'password' default tablespace NDB_USERS1
temporary tablespace temp;
grant create sequence to NDB;
grant create session to NDB;
grant create table to NDB;
grant create view to NDB;
grant create materialized view to NDB;
grant create procedure to NDB;
grant create any directory to NDB;
alter user NDB quota unlimited on NDB_USERS1;
exit;
```

## Imageからcontainerを起動

```
$ ./run_image.sh
```

以下のように表示されれば成功。

```
...
#####
DATABASE IS READY TO USE!
#####
```

起動すると、これまで使っていたターミナルは使えなくなるので、別のターミナルを開いて（Cmd+T or Cmd+N）以下の作業を行う。

## Databaseの設定、サンプルデータの挿入

```
$ ./setup.sh
```

- 途中で失敗した場合  
./drop.shを実行することで、作りかけのテーブルなどを全て削除することができる。  
実行前にdrop.shにも実行権限を付与する必要がある。
- SQLの練習のためにDatabaseの設定だけを行いたい場合  
./dcl.shでDatabaseの設定のみ行える。  
実行前にdcl.shにも実行権限を付与する必要がある。

## sqlplusでの接続

- sqlplusでの接続の基本構文

```
$ sqlplus username/password@//hostname:port/SID
```

このコマンドを実行したディレクトリが、SQL内でもカレントディレクトリになる。

- 構成したOracleDBへの接続

```
$ sqlplus NDB/'password'@//localhost:1521/NDBdummy
```

とすると、DBに接続できる。  
sqlplusを使って、AWS等にもアクセスすることが可能である。

## テーブルの確認

```
> select table_name from user_tables;
```

でテーブルの一覧が出力されれば、成功。

## 5. 終了、再起動

### 終了

- sqlplusの接続終了  
^D or exitでsqlでの接続を終了できる。
- Docker containerの停止

```
$ docker stop ndb-dummy
```

ここで、docker start ndb-dummyとすれば、すぐに再開可能だが、PCをシステム終了や再起動してしまうとdockerのprocessは終了してしまうので、状態は保存されない。

- Docker containerの削除

```
$ docker rm ndb-dummy
```

### 再起動

- TestDBSetupフォルダをworking directoryにする

```
$ cd $HOME/Workspace/NDB/NDB-UTDPH/ndb-sql/TestDBSetup
```

- Dockerのclean up

```
# docker内の容量整理
$ docker system prune

$ docker ps
# もし起動しているcontainerがあったら...
$ docker stop CONTAINER_ID

$ docker ps -a
# もしcontainerがあれば...
$ docker rm CONTAINER_ID
```

- Imageからcontainerを起動

```
$ ./run_image.sh
```

起動すると、これまで使っていたターミナルは使えなくなるので、別のターミナルを開いて（Cmd+T or Cmd+N）以下の作業を行う。

- Databaseの設定、サンプルデータの挿入

```
$ ./setup.sh
```

SQLの練習のためにDatabaseの設定だけを行いたい場合は./dcl.sh。

- sqlplusでの接続

```
$ sqlplus NDB/'password'@//localhost:1521/NBDummy
```

これで初期状態から作業を再開することが出来る。

## 6. Dockerのclean re-install

何らかの理由でDocker自体のclean re-installが必要になった時に行う。

- dockerfilesフォルダをworking directoryにする

```
$ cd $HOME/Workspace/Oracle/product/database/docker-images/OracleDatabase/dockerfiles
```

- Imageをbuild

```
$ ./buildDockerImage.sh -v 12.2.0.1 -e -i
```

Buildが成功すれば、後は[再起動](#)のセクションと同じ。

# 別添資料 2

## 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 夕薬上市	...	□□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歳以上男								
90歳以上女								
○○県								
20歳代男								
20歳代女								
...								
◎◎県								
90歳以上男								
90歳以上女								

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 夕薬上市	...	□□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。