

研究報告書

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

診療現場の実態に即した医療ビッグデータ（NDB 等）の解析の精度向上及び高速化を可

能にするための人材育成プログラムの実践と向上に関する研究

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

研究要旨：「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」において医療ビッグデータハンドリング教育プログラムを作成、運用した。本研究では①この教育プログラムの更なる改良、②e-learning 用プログラムの開発を通してより広い対象者へ医療ビッグデータハンドリング教育を提供することを目的とする

A. 研究目的

近年医療ビッグデータを利用した研究が爆発的に増加しており、臨床判断や医療政策策定に必要なエビデンスの創出に大きな役割を占めていることはすでに周知の事実である。医療ビッグデータを利用した研究を行うためには通常の臨床データを用いた臨床研究と異なり、臨床の知識・臨床研究の知識に加えて、データベース、統計学、機械学習、プログラミングの知識や技術を要する。しかしながら、これまでに臨床家がこれらの知識や技術を習得するための体系だった教育プログラム等はほとんど存在しないため、研究者が自己学習によってこれらの知識や技術を習得することは非効率的である。本研究の目的はすでに「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」で開発した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 等の統計ソフトの習得を目指す教育プログラムの改良と e-learning 用教育プログラムの開発を行い、これをより多くの対象者へ提供することである。

B. 研究方法

①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において作成した医療ビッグデータハンドリング教育プログラムを改良するために、これまで日本臨床疫学会 NDB・DPC データベース研究人材育成<短期集中セミナー>（以下サマーセミナー）、自治医科大学データサイエンスセンターにおける臨床家の教育を行った。これらのプログラム提供後にフィードバックを得たことに加えて、再度プログラムを精査し

た上で改良を加えた。さらに改良した教育プログラムを自治医科大学データサイエンスセンターにおいて研究を行う臨床家へ提供した。②これまでに作成した R に関する教育プログラムをもとに e-learning 用教育プログラムの開発を行った。

C. 研究結果

①サマーセミナーや自治医科大学で本プログラムの提供を受けた臨床家より受けたフィードバックをもとにプログラムを精査し、R、SPSS に関する教育プログラムをそれぞれ改良した。具体的には、プログラムを初級プログラム、発展プログラムと分けることで①基礎的な分析を習得する、②必要に応じて発展的な内容を習得するという 2 段階構成にすることで研究者の必要に合わせたきめ細やかなプログラムを提供できるようになった。

②R に関する教育プログラムとして e-learning 用の動画を作成し臨床家へ提供した。

各教育プログラムの作成および提供

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

自治医科大学ではデータサイエンスセ

ンターで臨床家に対して改良したプログラムを提供した。受講者は全員自身の研究計画に沿ったデータの抽出を自身で行うことができるようになった。

(2) SPSS/R による統計解析

SQL によって抽出したデータを利用し、(i)データのクリーニング、(ii)各変数の集計及び可視化、(iii)重回帰分析、ロジスティック回帰分析、生存時間分析、(iv)傾向スコア分析、(v)操作変数法 (R のみ)、(vi)自己対象ケースシリーズ分析 (R のみ) 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習得できる。また、必要に応じて傾向スコアを始めとした発展的な統計手法も習得することが可能となった。

自治医科大学ではデータサイエンスセンターでの研究に参加している研究者に対して R、SPSS での同様のプログラムを提供した。プログラムの提供を受けた臨床家は SQL で抽出したデータを R、SPSS を用いて、自身で基本的な統計解析を行うことができるようになった。更に傾向スコア分析などの発展的な分析も必要に応じて実施できるようになった。

R の基本操作から一般的に利用される統計解析を学ぶ e-learning 用教育プログラムの開発を行った。プログラムの提供を受けた臨床家は e-learning によって R の基本操作や基本的な分析を自身で行うことができるようになっている。

(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った。

(倫理面への配慮)

倫理的な問題はない。

D. 考察

既存の教育プログラムでは不十分であった医療ビッグデータを用いた研究のための教育プログラムを作成・改良し、試行した。また、一部 e-learning 用教育プログラムを開発した。このプログラムの受講することで実際に医療ビッグデータを利用した研究に繋がった。特に初級プログラムを習得した後、研究に合わせて必要な発展プログラムを受講することができるため、目的が明確でわかりやすいと評価を得た。

医療ビッグデータを用いた研究を行うにあたり、最も一般的に利用される言語は RDBMS 操作のための SQL、統計解析のための R や SPSS、機械学習のための python などが挙げられる。従って、医療ビッグデータを用いた研究を行う研究者はこれらの言語を習得する必要がある。しかしながら、例えば SQL を学ぶにあたって、一般的な書籍や講座等ではデータ定義言語、データ操作言語、データ制御言語を学ぶことになる

が、実際に医療ビッグデータを用いた研究にはデータ定義言語及びデータ操作言語が必要である。さらに、この中でも特によく使われるクエリは限られているが、どのクエリが研究に必要なのかなどは初学者にはわからないため、効率的な学習が困難である。本研究において作成した教育プログラムでは特に医療ビッグデータを用いた研究に特化し、頻出するクエリを中心に学習できること、更に受講者からのフィードバックを得て改良を行っていることから一般的な書籍や講座と比較して圧倒的に効率的に学習することが可能である。実際に医療ビッグデータを利用した研究を行いたいと考える研究者が本教育プログラムを受講し、高い評価を受けていること、自治医科大学で実際に研究に結びついていることから効率的に身につけることができていると考えられる。

また、本研究で開発した R の e-learning 用教育プログラムはより多くの対象者への教育の提供が見込まれる。実際に e-learning 用教育プログラムの提供を受けた臨床家が基本的な R の操作や分析を行うことができるようになったことから、医療ビッグデータを利用した研究につながることを期待できる。

E. 結論

医療ビッグデータ研究の為に必要な知識・技術を養成するための教育プログラムを改良・提供した。本教育プログラムは実際の研究に結びつくことが明らかとなった。今

後も本教育プログラム及び、開発中の e-learning 用教育プログラムを通して多くのより研究者が医療ビッグデータを利用した研究を行うための知識と技術を身につけられるよう、継続的に改善・提供していく。

F. 健康危険情報

なし

G. 研究発表

論文発表

1. Hayashi K, Sasabuchi Y, Matsui H, Nakajima M, Ohbe H, Ono K, Yasunaga H. Blood Purif. Clinical Effect of the Acrylonitrile-Co-Methallyl Sulfonate Surface-Treated Membrane as a Cytokine Adsorption Therapy for Sepsis due to Acute Panperitonitis: A Retrospective Cohort Study. 2020;49:364-371.
2. Suzuki J, Sasabuchi Y, Hatakeyama S, Matsui H, Sasahara T, Morisawa Y, Yamada T, Yasunaga H. The effect of antithrombin added to recombinant human-soluble thrombomodulin for severe community-acquired pneumonia-associated disseminated intravascular coagulation: a retrospective cohort study using a nationwide inpatient database. J Intensive Care. 2020;8:8.
3. Suzuki J, Sasabuchi Y, Hatakeyama S, Matsui H, Sasahara T, Morisawa Y, Yamada T, Yasunaga H. Histamine-2

receptor antagonists versus proton pump inhibitors for septic shock after lower gastrointestinal tract perforation: a retrospective cohort study using a national inpatient database. J Intensive Care. 2020;8:56.

4. Ookawara S, Ito K, Sasabuchi Y, Hayasaka H, Kofuji M, Uchida T, Horigome K, Imai S, Akikawa T, Wada N, Kiryu S, Imada S, Shindo M, Miyazawa H, Hirai K, Onishi Y, Shimoyama H, Watanabe A, Tabei K, Morishita Y. Associations of cerebral oxygenation with hemoglobin levels evaluated by near-infrared spectroscopy in hemodialysis patients. PLoS One. 2020;15:e0236720.
5. Ookawara S, Ito K, Sasabuchi Y, Ueda Y, Hayasaka H, Kofuji M, Uchida T, Horigome K, Aikawa T, Imada S, Minato S, Miyazawa H, Shimoyama H, Hirai K, Watanabe A, Shimoyama H, Morishita Y. Association between Cerebral Oxygenation, as Evaluated with Near-Infrared Spectroscopy, and Cognitive Function in Patients Undergoing Hemodialysis. Nephron. 2021 Online ahead of print.

学会発表

1. 室岡由紀恵. ICU入室患者の早期リハビリテーションによる長期予後への影響. 第48回日本集中治療医学会学術集会.

2021/2/12-14

H. 知的財産権の出願・登録状況

なし

