

厚生労働行政推進調査事業補助金  
政策科学総合研究事業（政策科学推進研究事業）

NDBをはじめとする医療ビッグデータを利活用できる人材の裾野を広げるための研究  
総括研究報告書

研究代表者 康永秀生 東京大学大学院医学系研究科 教授  
研究分担者 田宮菜奈子 筑波大学医学医療系/ヘルスサービス開発研究センター 教授  
研究分担者 中山健夫 京都大学大学院医学研究科 教授  
研究分担者 山名隼人 自治医科大学データサイエンスセンター 講師

研究要旨：

NDB・DPC等の大規模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。東京大学は、令和5年度研究において、先行研究で開発した人材育成プログラムのコンテンツを更新・改善し、新たな「NDB・DPC等データベース研究人材育成セミナー」を開発した。令和6年度のセミナーの実施と評価に向けて準備中である。NDB、DPCなどの大規模データベースを用いた研究を引き続き実施し、2023年は英文原著論文が124編出版された。

京都大学では、2019年度より文部科学省が管轄する「関西広域・医療データ人材育成拠点形成事業」を開始しており、主コースである修士課程追加履修コースのほかに、民間向けに展開する2つのインテンシブコースを併設している。2023年度は本事業の成果を第43回医療情報学連合大会ワークショップ「未来を担う医療データ取扱人材の育成～関西広域 医療データ人材教育拠点形成事業（KUEP-DHI）の取り組み～」で報告した。アカデミアにおける研究者養成とは異なる形式での保健医療介護ビッグデータ研究人材育成の可能性について課題と可能性について検討を進めている。

筑波大学では、昨年度に引き続き、医療・介護保険レセプトデータやDPCデータを用いた研究を推し進め、特に博士課程・研修医・若手教員を対象にした人材育成に力を入れている。リアルワールドデータを用いた臨床医学・社会医学研究に関心を持つ医療者や若手教員を対象に医学教育を行い、研究者の裾野を広げ、本事業の目標である幅広い医療ビッグデータ研究にかかわる人材育成に貢献することができた。

自治医科大学では、自治体職員など様々な利活用者と目的に合わせたプログラムの改良点を検討し、今後求められる新規データの検討と更なる利活用のための基盤強化を実施した。このことにより、医療ビッグデータを利活用できる人材の裾野を広げることに貢献した。

## A. 研究目的

近年、保健医療介護のリアルワールドデータを収集した大規模データベース研究が拡大している。

平成23年度からレセプト情報・特定健診等情報データベース（以下、NDB）の研究への第三者提供が始まり、令和元年度からNDBオンサイトリサーチセンターが一般開放された。令和2年度から第三者提供が法制化され、民間事業者等への提供も可能となった。令和3年度には提供情報の範囲が拡大され、さらにNDBのリプレイス及びNDBデータ等をクラウド上で解析できる「医療・介護データ等の連結解析基盤」（以下、HIC）の開発が進められた。令和4年度からオンサイトリサーチセンターのクラウド化、HICの試行的利用が開始された。このように、保健医療介護データベース研究の素地は整いつつある。

データベース研究に精通した研究者は、本研究に先行する研究班のこれまでの取り組みによって徐々に増加している。医学系研究者を中心にNDB利用者は徐々に増加しつつあるものの、十分とはいえない。さらに他の領域（経済学系や工学系など）の研究者や民間事業者のNDB利用はいまだ極めて限定的である。NDB利用者の裾野を広げるための人材育成が引き続き求められている。保健医療介護の実態を理解し、なおかつデータ解析に精通した研究者を育成することは喫緊の課題である。そのために、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の教育が必須である。

### 1. ビッグデータ研究実践能力およびデータハンドリング技術養成プログラムの実践（東京大学）

東京大学では、NDB・DPC等の医療ビッグデータの解析に精通した研究者を育成するための教育プログラム開発等を行ってきた。上記研究をさらに発展させ、令和5年度は先行研究にて開発された人材育成プログラムを改良し、NDB・DPC等データベース研究人材育成セミナーを開発した。これらを用いて、令和6年度にはセミナーをWebで実施し、参加者によるプログラム評価も実施する予定である。

さらに、人材育成セミナーでの教育を活かして、多数の研究者によるデータベース研究を推進し、これまで通り論文実績を積み重ねていくことも目的としている。

### 2. NDB データ研究の実践と人材育成プログラム開発・検証（京都大学）

京都大学では、大学院課程における研究者養成とは異なる、企業の勤務者等社会人を対象とする保健医療介護ビッグデータ研究人材育成プログラムの可能性と課題を検討する。

### 3. 医療・介護データ活用による研究・人材育成（筑波大学）

リアルワールドデータの有効活用に基づく

evidence based policy making (EBPM)の重要性が広く一般世間・社会に認識されるようになってきた。この時に重要なことは、時間と研究意欲に溢れる若手研究者が、経験・実績の豊富な指導者の下で、素早くデータを解析し、報告書や論文を執筆して世に発信することである。そのような教育体制・環境を日頃から構築することが重要である。

筑波大学では、これらの取り組みが加速でき、過去2年間に引き続き、今年度も研修医・博士課程生・若手教員が実際に国際誌掲載に成果をあげることができた。

#### 4. NDBをはじめとする医療ビッグデータを利活用できる人材の裾野を広げるための研究（自治医科大学）

医療ビッグデータを利活用する目的は臨床疫学研究に限定されず、今後は地域医療分析や政策への活用も期待されている。このため、既存の教育プログラムを様々な利用者と目的に合わせ更に改良することが求められている。

自治医科大学では、これまでに開発した医療ビッグデータ活用のための教育プログラムを継続して実施することで更なるエビデンス創出を図ることに加え、様々な利活用者と目的に合わせて教育プログラムの改良点を探るとともに、今後の利活用を見据えて新規データの検討と更なる利活用のための基盤構築を行った。

## B. 研究方法

### 1. ビッグデータ研究実践能力およびデータハンドリング技術養成プログラムの

## 実践（東京大学）

### （1）人材育成プログラムの開発

先行研究において開発した人材育成プログラムをベースに、新たに「NDB・DPC等データベース研究人材育成セミナー」のコンテンツを開発した。これには、ビッグデータ研究実践能力養成プログラムおよびビッグデータハンドリング技術養成プログラムが含まれる。

ビッグデータ研究実践能力養成プログラムは、日常臨床からクリニカル・クエスチョンを紡ぎ出し、検証可能なりサーチ・クエスチョンに構造化し、既存のビッグデータを用いてリサーチ・クエスチョンを解き明かす臨床研究・疫学研究・ヘルスサービスリサーチの実践能力を養成するプログラムである。

ビッグデータハンドリング技術養成プログラムは、テラバイト級のデータベースから個別の研究目的に沿うデータセットを抽出するデータハンドリング技術、膨大なテキストデータを含む大規模かつ複雑な構造のデータを研究用の扱いやすいデータベースに再構築するなどのデータベースマネジメント技術を養成するプログラムである。

令和6年度にオンラインによるセミナーを実践し、その評価を行う予定である。

### （2）大規模データベース研究の実践

先行研究における人材育成プログラム受講者をはじめ、多施設から多くの研究者

と共同研究体制を構築し、NDB, DPC 等のデータベースを用いた臨床研究、疫学研究、医療経済・政策研究、ヘルスサービスリサーチを継続的に実施した。

## 2. NDB データ研究の実践と人材育成プログラム開発・検証(京都大学)

京都大学は2019年度より文部科学省の助成を受け、関西広域の計13大学とコンソーシアムを構築して「関西広域・医療データ人材育成拠点形成事業(Kansai Union / Kyoto University Education Program for Digital Health Innovation: KUEP-DHI. 責任者:黒田知宏・京都大学医学部附属病院医療情報企画部教授)」に取り組んでいる。本事業は医療現場から大規模に収集される多様なデータの利活用を推進し、質の高い医療を実現するため、医療データの活用基盤を構築・運営する人材、医療データを利活用できる人材を育成することを目指すものである。KUEP-DHIの根幹は、大学院生(主に修士学生)を対象とした「医療データ取扱専門家育成コース」であり、京都大学大学院医学研究科と同大学院情報学研究科に追加履修コースとして設置されている。

## 3. 医療・介護データ活用による研究・人材育成(筑波大学)

医療・介護保険レセプトデータ活用による研究・人材育成をすすめた。若手研究者による医療・介護ビッグデータ活用による論文投稿を支援した。解析及び論文執筆は、研究

分担者(田宮)および研究協力者(岩上)の指導のもと、筑波大学医学医療系ヘルスサービスリサーチ分野に所属する博士課程学生および客員研究員が行った。

4. NDBをはじめとする医療ビッグデータを利活用できる人材の裾野を広げるための研究(自治医科大学)

既存のビッグデータハンドリング教育プログラムのうち、改良が期待される内容についてヒアリングを行い、継続して重視すべき項目や、追加・修正が必要な事項について検討する。

利活用者のニーズを踏まえ、今後取得を目指すデータについて検討し、利活用を更に進めるために必要なインフラ面の強化を検討した。

## C. 研究結果

### 1. ビッグデータ研究実践能力およびデータハンドリング技術養成プログラムの実践(東京大学)

#### (1) 人材育成プログラムの開発

先行研究では撮影済みの動画のオンデマンド配信形式であったところ、今回はウェビナー形式の講義+eLearningとし、さらにウェビナーの録画をオンデマンド配信することとする。

受講対象は保健医療介護ビッグデータ研究に興味のある方とし、これまでの主な対象である医療・介護従事者、大学・研究機関に所属する研究者の他に、企業に所属する者、

自治体等の医療政策の担当者にも門戸を広げる。

コンテンツは以下の通り。

1. NDB データ等研究に関連する法律や制度と倫理申請
2. NDB・DPC 等公的データベースの動向
3. NDB のデータ申請とオンサイトセンターの利用
4. SQL の活用：基礎編
5. SQL の活用：中級編
6. リアルワールドデータの研究デザイン
7. 医療・介護連結データを用いた研究のデザインと実践
8. NDB 研究のデザインと実践
9. DPC データ研究のデザインと実践
10. DeSC データ研究のデザインと実践
11. JMDC データ研究のデザインと実践
12. バリデーション研究
13. 因果推論(1)傾向スコア分析の応用
14. 因果推論(2)高次元傾向スコア
15. 因果推論(3)操作変数法
16. 因果推論(4)不連続回帰分断デザイン・差の差分析
17. 因果推論(5)時間依存性交絡
18. 競合リスク分析
19. 多重代入法
20. マルチレベル分析
21. 自己対照ケースシリーズ
22. 機械学習・深層学習

各講義の後に無記名アンケートを実施する。アンケートの内容を踏まえて次年度のコンテンツの改良に繋げる。

## (2) 大規模データベース研究の実践

2023 年に 124 編の原著論文を出版した。

### 2. NDB データ研究の実践と人材育成プログラム開発・検証(京都大学)

KUEP-DHI では大学院生という立場に依らず、企業の勤務者等、社会人を想定して次の2種類のインテンシブコースを提供している。

#### (1)DHIEP Program(ディープ プログラム)

Data-driven Healthcare Innovation Evangelist Promotion Program 分析だけを請け負う従来型データサイエンティストではなく、ライフサイエンス・ヘルスケア領域に深い造詣を有し、社会変革を実現しうる人材育成を目指すプログラム。対象は医療データの利活用を志す個人であり、個人参加、登録等も個人で行う。定員は 20 名。カリキュラムの修業期間は半年間であり、大きく経営・社会変革、医学・生命科学、情報・データサイエンスの 3 領域で基礎知識を e-learning で学び、必修科目 45 時間、選択科目 21 時間、実習(必修)20 時間で構成されている。

(2)医療データ人材育成拠点形成事業ビジネス特化型インテンシブコース Kansai Union / Kyoto University Education Program for Digital Health Innovation directly on themes of business (KUEP-DHI dot.b.):民間(企業)のニーズを踏まえつつ、医療データの活用による造詣のある人材を育成していくことを目的とするプログラム。対象は医療データ利活用に造詣のあるスタッフを確保したい企業であり、個人参加ではあるが、契約は企業－大学間で

締結する。

育成する人材像は参加者が所定のプログラムを終えられた後には、各企業に戻り、医療データを企業内で有効かつ安全に利用する方針を立案したり、使用できるデータを適切に活用したりすることで有用な商品やビジネス等の開発に繋げ、社会に貢献できる人材を想定。

### 3. 医療・介護データ活用による研究・人材育成（筑波大学）

若手研究者による以下の2論文を刊行した。

Characteristics and in-hospital mortality of patients with COVID-19 from the first to fifth waves of the pandemic in 2020 and 2021 in the Japanese Medical Data Vision database (Heliyon. 2023 Aug 25;9(9):e19490.)

Factors associated with non-participation in breast cancer screening: analysis of the 2016 and 2019 comprehensive survey of living conditions in Japan (Breast Cancer . 2023 Nov;30(6):952-964.)

また、日本臨床疫学会および日本薬剤疫学会の年次学術総会で教育講演を実施した。

4. NDBをはじめとする医療ビッグデータを利活用できる人材の裾野を広げるための研究（自治医科大学）  
既存の教育プログラムのうち、幅広く利活用可能であり、引き続き重視すべきコンテ

ントは下記(1)-(5)のとおり整理された。

#### (1) データ構造の理解

- ・レセプトデータ等の種類と仕様
- ・データ抽出のためのコード等

#### (2) 研究デザインの検討

- ・クリニカルクエスション (CQ) からリサーチクエスション (RQ) への作り変え
- ・FINER による RQ の評価
- ・文献検索

- ・抽出変数の定義と抽出計画の策定

#### (3) SQL によるデータの切り出し

- ・複数のテーブルから情報を抽出する
- ・解析用テーブルに加工する

#### (4) R による統計解析

- ・データの成型
- ・各変数の集計及び可視化
- ・検定、回帰分析

#### (5) 論文執筆

- ・論文の構造の理解
- ・データの適切な解釈

また利用者へのヒアリングからは、系統的学習（データに関する説明資料、講義動画、ダミーデータ等）に加えて、解析に慣れていない初学者を中心に、個別指導をより重視するニーズが明らかになった。更に下記(6)(7)が要望として挙げられた。

#### (6) 様々な解析ソフトウェアの活用

Stata, SPSS, TreeAge, ArcGIS 等、利用者に応じて様々なソフトウェアを使用するニーズがあることが明らかになった。特に統計解析ソフトウェアについてはグラフィカルユーザーインターフェースを希望する利活用

者が多かったことから、Rに加えて Stata を用いた統計解析の支援を導入した。

#### (7) データ可視化のためのビジネスインテリジェンス(BI)ツールの活用

自治体職員を中心として、BI ツールを用いた可視化の希望が挙げられた。Tableau を導入し分析を行うとともに、BI ツールに利用できる汎用的な中間テーブルを作成することとした。

### D. 考察

#### 1. ビッグデータ研究実践能力およびデータハンドリング技術養成プログラムの実践 (東京大学)

先行の人材育成プログラムを改良し、双方向型の教育が可能なオンラインプログラムを開発した。令和6年度のリリース、ウェビナーの実践と評価を行う。

また、大規模データベース研究の実践をさらに進めた。2019年の約50編、2020年の約70編、2021年の約100編、2022年の117編に引き続き、2023年は124編の原著論文を出版した。コンスタントに増加しており、本研究の人材育成プログラムの効果を示唆するものである。

#### 2. NDB データ研究の実践と人材育成プログラム開発・検証(京都大学)

「関西広域・医療データ人材育成拠点形成事業」における2種類の社会人向けインテンシブコースの経験から、データサイエンスの人材育成に関するいくつかの知見を得た。下

記に抜粋して記述する。

- ・医療者でない参加者の場合、医療データ以前の「医療そのもの」への距離感が大きいいため、プログラム提供に当たっては、そのギャップを十分意識することが必要。
- ・NDB 等の実際のデータに触れることは学習者のモチベーションを大きく高め得る。
- ・「公益性のある研究」として NDB 申出書を書くことの意義を体感することは大いに重要。
- ・医師の診断プロセスや、電子カルテ入力体験を通じた、「データが絶対とは限らない」「バリデーションが重要」という感覚の共有はきわめて重要。
- ・レセプトでの傷病名同定における投薬や処置の情報を理解することの重要性。
- ・医療の場で入力されている情報が必ずしも正確かつ網羅的とは言えないことの実感。
- ・模擬的なデータ分析体験を通じた、臨床研究への意識の高まり。
- ・「企業が求める人材育成」と「大学教育が想定する人材育成」の違いの認識。
- ・受講者ニーズと提供するコンテンツとのマッチングの重要性。
- ・人材育成プログラムに参加することで、その者が通常業務に就いていたことで得られたはずの派遣元への利益を上回る貢献を派遣元にもたらす可能性を、できる限り可視化することの重要性。
- ・人材育成プログラムに参加した者が、派遣元に戻った後も継続的に医療データサイエンティストとして活躍できるために、医療データに関連する具体的な研究事業が派遣元に存在することの重要性。

### 3. 医療・介護データ活用による研究・人材育成（筑波大学）

「後進の育成」に関する活動の一環として、筑波大学の若手に限らず、日本全国を対象にした啓蒙教育活動に関わることができた。医療・介護ビッグデータ研究の後進の育成実績およびそのノウハウが順調に蓄積されている。これは、今後の医療人材のあるべき方向にも沿ったものであり、かつ学生や若手自身が自ら意欲的に取り組める内容でもあり、適切な指導により高い成果につながることが示唆された。

### 4. NDBをはじめとする医療ビッグデータを利活用できる人材の裾野を広げるための研究（自治医科大学）

本研究では、医療ビッグデータを利活用できる人材の裾野を広げるにあたり、研究者や大学院生に加えて自治体職員等とも意見交換を行い、既存の教育プログラムの改良を図るとともに、更なる利活用のための検討を行った。

今後は地域医療分析や医療政策の立案など、様々な目的でビッグデータを利活用することが求められている。本研究では、増加する新たなニーズに合わせて利活用者から意見を集めることで、ビッグデータを用いて実現可能な分析について整理するとともに、既存の教育プログラムのうち継続して重視すべき内容や、改良が必要な点について明らかにした。

本研究では、当センターにおいてデータと研究の機会を臨床医を中心とした若手研究者に引き続き提供することで、多数のエビデンスを持続的に創出した。このことは一連の教育プログラムの有効性を示していると考えられる。一方で、新たに様々なバックグラウンドを持つ利活用者を迎え入れ、データの種類を増やし、求める分析内容に対応した教育プログラムを作成することが必要である。

### E. 結論

本研究は、厚生労働省が進める NDB 等データベースの高度利活用に直接反映される研究である。

近年、保健医療介護データベースのインフラ整備が進められている。それらのインフラを活用できる人材の育成にさらに貢献し、現在整備中の統合データベースの稼働の際には全省的な政策課題に関する研究・知見の提供にも貢献できる。さらに、日常臨床のクリニカル・クエスチョンをデータベース活用により解明する研究実践能力を持つ研究者を多数育成し、データハンドリング技術と臨床研究実践能力の両方に長けた人材も多数育成することにより、わが国の大規模データベース研究の進歩を加速できる。わが国の大規模データベース研究の技術水準を世界トップレベルに向上させ、わが国発のエビデンスを量産できることが期待される。

また、様々な利活用者と目的に合わせて教育プログラムの改良を検討し、今後求めら

れる新規データの検討と更なる利活用のため  
の基盤を強化する必要がある

**F. 健康危険情報**  
なし

**G. 研究発表**  
(別添4および各分担研究報告書に記載)

**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