

厚生労働行政推進調査事業費補助金
(政策科学総合研究事業(政策科学推進研究事業))
総括研究報告書

レセプト情報・特定健診等情報データベースの利活用の推進に関する研究
研究代表者 大江 和彦 東京大学医学部附属病院企画情報運営部 教授

研究要旨

レセプト情報・特定健診等情報データベース(NDB)について、平成 28 年度から提供が開始された NDB オープンデータとともに、その情報提供機能の充実と利用に関して潜在的に存在していると考えられる課題も含めてとりあげ、その解決方法についても検討する。具体的に検討すべき課題として 28 年度は、①ID による同一患者の名寄せ手法、②信頼できる傷病名情報の取得方法、③NDB 特別抽出データの取得と処理方法の効率化、④基本データセットの作成方法、⑤NDB オープンデータの役割および利用例としての国際比較データの作成、の 5 点について検討した。レセプトデータが患者ごとに連結されず、医療機関 x 保険者 x 個人 x 受診年月、に分断されたデータ単位として匿名化されているためにこれらの分断データ単位を再結合することが非常に困難になっている。この問題を解決するには保険者が厚労省にデータを匿名化して出す場合に、医療機関 x 個人 x 受診年月についてはその同一個人ごとに 1 匿名化 ID となるように処理するとともに、その匿名化 ID が保険者間での結合できるように、全レセプトデータを管理する機関が正確な被保険者台帳を管理運用することであろう。一方、データ提供にかかる時間の短縮、受け取ったデータの再 DB 化が効率的に実現できるように、研究者が利用するデータベースエンジンを想定したデータ提供も必要で、現状のテキストデータでの提供には限度がある。また傷病名の精度確保に関する研究をさらにすすめる必要もある。さらに今後も、NDB からのデータ提供としての特別抽出、サンプリングデータセット、基本データセット、NDB オープンデータをはじめとする情報提供機能の充実と利用が必要である。

以上のように課題は残っているが、患者 ID の取扱い、オープンデータの利用、特別抽出データの利用を含む NDB データ活用につき、具体的課題と今後の対応のあり方を示した。今後の NDB データの提供側およびデータの利用側にとっても活用推進に役立つと思われる。また NDB オープンデータから生成できるデータを使用した国際統計報告は、諸外国の行政関係者や研究者への波及効果も生み出し得るものと考えられた。

<研究分担者>

今中 雄一 京都大学大学院医学研究科医療経済学分野 教授
満武 巨裕 一般財団法人 医療経済研究・社会保険福祉協会 医療経済研究機構
副部長

<研究協力者>

國澤 進 京都大学大学院医学研究科医療経済学分野 講師
大坪 徹也 京都大学大学院医学研究科医療経済学分野 助教
佐藤 大介 東京大学医学部附属病院企画情報運営部 助教

A. 研究目的

日本におけるレセプト情報・特定健診等情報データベース（以下、NDB）の情報提供機能の充実と利用についての課題および解決方法について検討する。日本のNDBは、提供開始（2011年11月）から本年度までの利用承諾件数が100件を超えた。引き続き、研究利用として提供している特別抽出、サンプリングデータ、基本データセット、NDBオープンデータをはじめとする情報提供機能の充実と利用が必要であるが、潜在的に存在していると考えられる課題をとりあげ、その検討を行う。

B. 研究方法

具体的に検討すべき課題として今年度は、①IDによる同一患者の名寄せ手法、②信頼できる傷病名情報の取得方法、③NDB 特別抽出データの取得と処理方法の効率化、④基本データセットの作成方法、⑤NDB オープンデータの役割および利用例としての国際比較データの作成、の5点について取り上げた。①③⑤については分担研究者の今中が、④および⑤の国際比較データ作成については分担研究者の満武は、②について研究代表者の大江がそれぞれ分担して検討した。

C. 研究結果と考察

①IDによる同一患者の名寄せ手法に関する検討

「全患者ID1数」に占める「患者ID2を一つもつ患者ID1の数」は、63%であり、「全患者ID2数」に占める「患者ID1を一つもつ患者ID2の数」は、68%であった。これにより、氏名の入力ゆれや改名の発生頻度は、保険者番号・記号番号の入力ゆれや保険の変更の発生頻度よりも低いものの、小さいものではないといえる。開発した統合ID生成アルゴリズムを適用した結果、患者ID1と患者ID2の組合せ数に対する統合ID数は42%となった。

こうしたハッシュ化患者IDの問題に対する根本的な解決は、全レセプトデータを管理する機関が、正確な被保険者台帳を管理運用することであろう。

②信頼できる傷病名情報の取得方法の検討

2012年のある4ヶ月間での入院歴のないレセプトをランダムに40名分調査したところ、レセプト1件あたりの傷病名数は平均15.2個であり、レセプトに登録された診療行為と傷病名の適用関係データベースを利用し

た診療行為から傷病名にウエイトをつけて評価した先行研究では、ほぼ疑いなく患者に存在する傷病名の推定は93%の正確さであったとの報告があり、連続する数ヶ月のレセプトを処理することで性能が上がる可能性が指摘されている。

機械学習と上記手法とを組み合わせた傷病名推定を行うことにより、精度の高い傷病名推定ができる可能性があり、入院患者レセプトについて同一患者で連結した最長6ヶ月のレセプトの傷病名、診療行為、診療科、時間軸で出現情報などを入力とし、正解データとしてDPCの最も医療資源を消費した病名を用いて機械学習することが考えられた。

機械学習により傷病名推定を行うには、大量の正解データが必要で、入院レセプトではDPCの主たる医療資源を消費した病名以外には、退院時サマリの病名を電子的に収集する手法が考えられる。外来レセプトにおいて正解データである傷病名をどのように入手するかについてはさらに検討が必要であるが、特定の傷病名で実施される診療行為と、そうでない診療行為のリストを作成して分析することが必要であろう。

③NDB 特別抽出データの取得と処理方法の効率化に関する検討

ボトルネックになっている行程が、大きく3つあると考えられた。

i) 有識者会議の承諾から承諾通知発出の間

ii) データ返却、或いは承諾通知発出からデータ抽出作業開始の間

iii) データ抽出作業完了から提供データ格納用外付けハードディスク到着の間

これらに対し、i) は厚生労働省の事務処理、ii) は厚生労働省およびデータ抽出作業担当も含めた全体的なスケジュール調整・管理、iii) はデータ抽出作業担当からの連絡を改善することによって、それぞれ解消が図られると考えられた。

詳細は今中の分担研究報告書を参照されたい。

④基本データセットの作成方法に関する検討

基本データセットは、試行として2利用者にデータが提供されたが、課題となったのが、ユーザーが独自に指定する診療行為コードや医薬品コードの設定である。指定数は、厚生労働省・保険局との協議により256項目としていた。利用者は、診療行為、医薬品コードの指定にあたり、診療行為マスタは約6,700種類、医薬品マスタは

約20,000種類の中から設定しなければならなかった。このように、利用者が分析上条件設定として必要とする診療行為や医薬品は256程度よりはるかに多いことがわかる。これを解決するには、各診療行為や医薬品をカテゴリーに抽象化(指定粒度を意味的に粗くする)する必要がある。医薬品の場合、一般名化、WHO-ATC分類などの粒度に抽象化する、などの手法が必要である。

⑤NDB オープンデータの役割、および利用例としての国際比較データの作成に関する検討

都道府県ごとに、高齢者人口を医療需要総量を反映するものとみなして、どれだけの提供量があるかを示した。これらの指標は、治療へのアクセスの指標ともなりうると考えられる。一方で、都道府県内でも、地域間の格差が大きく、これらの指標は、各県内の格差の大きい地域を足し合わせた平均値である点に、特に留意すべきである。指標の解釈の例もコメントとして付記しているが、解釈も一例であることに留意が必要である。

NDB オープンデータを利用による国際比較データを作成は、CT の対 1000 人当たりの施行件数、MRI の対 1000 人当たりの施行件数について、国際比較を事例として OECD（経済協力開発機構）が公表している諸外国の医療の質データに関連した日本の新しい指標作成について検討した。OECD 加盟国間で比較可能なデータを作成できた。

現在、先進国は限りある財政状況の中、効率性と公平性を考慮しながら、医療および介護保険システムを運営している。我が国は、これまで、質の高い医療サービスを比較的少ない医療費で提供していると WHO(世界保健機関)等から評価されてきた。だが、増加傾向が続く医療費・介護費の対 GDP 比率は OECD(経済協力開発機構)の加盟国中で上位（第 3 位）となり、今後は一層、現在の医療・介護の質を確保しつつ、費用を適正化しなければならない課題に直面している。こうした状況は多くの先進国およびアジア諸国においても同様であり、このため OECD、WHO、国連等の国際機関は、医療及び介護分野における政策立案に資する国際統計報告として様々な HQ(ヘルスインディケータ:保健医

療指標)の迅速な提供を各国に求めているが、日本が国際機関に提出している厚生労働統計分野の項目数は少ない。

例えば、2015 年に OECD が提出を求めている 129 項目の HQ において、日本は 53 項目を提出しているが、加盟 35 カ国中 32 位と低い。ちなみに OECD 加盟国の平均提出件数は、89 項目である。そのため、日本はデータ提出状況を改善することが望ましい。NDB オープンデータから今回新しいエビデンスを作成することができたが、日本の全人口の保険医療記録から作成されたデータは、国内だけではなく諸外国に向けた情報としても価値のあるものと考えられる。

E.結論

レセプトデータが患者ごとに連結されず、医療機関 x 保険者 x 個人 x 受診年月、に分断されたデータ単位として匿名化されているためにこれらの分断データ単位を再結合することが非常に困難になっている。この問題を解決するには保険者が厚労省にデータを匿名化して出す場合に、医療機関 x 個人 x 受診年月についてはその同一個人ごとに 1 匿名化 ID となるように処理するとともに、その匿名化 ID が保険者間での結合できるようにインフラを整備すべきである。

データ提供にかかる時間の短縮、受け取ったデータの再 DB 化が効率的に実現できるように研究者が利用するデータベースエンジンを想定したデータ提供も必要で、テキストデータでの提供には限度がある。また傷病名の精度確保に関

する研究をさらにすすめる必要もある。

さらに今後も、NDB からのデータ提供としての特別抽出、サンプリングデータセット、基本データセット、NDB オープンデータをはじめとする情報提供機能の充実と利用が必要である。

以上のように課題はまだ多いが、患者 ID の取扱い、オープンデータの利用、特別抽出データの利用を含む NDB データ活用につき、具体的課題と対応を明確にした。当研究成果が公に広く提供されることにより、今後の NDB データの提供側、およびデータの利用側にとっても、活用推進に役立つと思われる。また NDB オープンデータから生成できるデータを使用した国際統計報告は、諸外国の行政関係者や研究者への波及効果も生み出し得るものであると考える。

厚生労働省が 2009 年から収集を開始した NDB は、ヘルスケア分野における最大規模のリアルワールド・データベースであり、そこから作成された二次データである NDB オープンデータとともに、国内だけではなく諸外国に向けた情報としても価値のあるものと考えられる。

F.研究発表

- 1) 満武巨裕、大江和彦、今中雄一：NDB オープンデータを研究利用に活用す

る：医療技術(CT,MRI,PET)の利用に関する国際比較の試み、社会保険旬報、第 2661 巻:12-16,2016 年

- 2) 満武巨裕（「諸外国の医療ビッグデータ」、第 2 回データヘルス時代の質の高い医療の実現に向けた有識者検討会（平成 28 年 5 月 23 日）
- 3) 大江和彦.医療情報データベースの基盤整備,情報管理.2016,vol.59,no.5, p.277-283.
- 4) 大江和彦,医療ビッグデータとこれからの医療,日本臨床検査自動化学会会誌(0286-1607)41 巻 4 号,371 (2016.08).
- 5) 松居宏樹,佐藤大介,大江和彦,レセプト情報等オンサイトリサーチセンターにおける、今後の第三者提供の方向性についてレセプト情報等オンサイトリサーチセンターにおける NDB データの利用システム環境と NDB の特性に関する報告,医療情報学連合大会論文集 36 回 1 号,138-140 (2016.11).
- 6) 佐藤大介,大江和彦,医療データベース利活用の国内基盤の最新状況レセプト情報等オンサイトリサーチセンターの試行的利用について,日本薬剤疫学会学術総会抄録集,巻 22,p51(2016.11).

G.知的所有権の取得状況

該当なし