

厚生労働科学研究費補助金 政策科学総合研究事業（統計情報総合研究）  
分担研究報告書

## ナショナルデータベース（NDB）データ分析における病名決定ロジック作成 のための研究

研究代表者 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

### 分担研究者

喜連川 優（国立大学法人東京大学 生産技術研究所）、合田 和生（国立大学法人東京大学 生産技術研究所）、横田 治夫（国立大学法人東京工業大学）、北川 博之（国立大学法人筑波大学）、櫻井 保志（国立大学法人熊本大学）、河原林 健一（国立情報学研究所）、豊田 正史（国立東京大学 生産技術研究所）、杉山 磨人（国立情報学研究所）、小笠原 克彦（北海道大学）、藤森 研司（東北大学大学院 医学系研究科）、東 尚弘（国立がん研究センター がん対策情報センター）、大江 和彦（国立大学法人東京大学 医学部附属病院）、山本 隆一（一般財団法人医療情報システム開発センター）、橋本 英樹（国立大学法人東京大学大学院）、今中 雄一（国立大学法人 京都大学）、伏見 清秀（国立大学法人 東京医科歯科大学）、宮田 裕章（慶應義塾大学）、津本 周作（島根大学）、松田 晋哉（産業医科大学）、中島 直樹（国立大学法人九州大学）、永井 良三（自治医科大学）、山縣 邦弘（筑波大学）、横手 幸太郎（千葉大学大学院医学研究院）、秋下 雅弘（東京大学医学部附属病院）、戸谷 義幸（横浜市立大学附属病院）、亀井 美和子（日本大学）、尾島 俊之（浜松医科大学）、三浦 克之（滋賀医科大学）、谷澤 幸生（山口大学）、森田 朗（津田塾大学）、野口晴子（学校法人早稲田大学）、飯島 勝矢（東京大学）、後藤 励（慶應義塾大学大学院）、佐方 信夫（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）、石川 智基（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）、酒井未知（京都大学大学院医学研究科）、野田光彦（埼玉医科大学 内分泌・糖尿病内科）

### 研究協力者

中山健夫（京都大学大学院医学研究科）、高橋由光（京都大学大学院医学研究科）、内海貴裕（京都大学大学院医学研究科）、西川佳孝（京都大学大学院医学研究科）、星野伸晃（京都大学大学院医学研究科）、堀松高博（京都大学大学院医学研究科）

### 研究要旨

本研究の目的は、レセプト情報を基にした発症群・合併症群・重症群の定義を確立し、臨床的知見に基づいた病名決定ロジックを作成することである。

レセプト情報・特定健診等情報データベース（NDB）の6年間データ及び保険者から収集した75万人規模の医療レセプトデータを申請者らが構築した「次世代超高速・超学際NDBデータ研究基盤」に保管・利用することにより、今回の病名決定ロジックを適用した患者数の算出と疾病別医療費の推計を行う。結果は、学会にフィードバックすることで妥当性についての検討を行う。これらのプロセスを経て、NDBデータ分析における病名決定ロジックについて、わが国の問題点が明らかになるとともに、患者数の算出、疾病別医療費の推計、効果的な公表のあり方について検討する。同時に、過去の研究者が開発してきた病名決定ロジックや厚生労働省の社会医療診療行為別調査等で活用されているロジックを調査する。

現在のNDBオープンデータは年々公開データの量と質が向上しているが、疾病別患者数データは公表されていない。当該研究の成果がNDBオープンデータに資するよう、今後も対象疾病を拡大する。

#### A. 研究目的

本研究では、レセプト情報・特定健診等情報データベース（NDB）の6年間データを保有する「次世代超高速・超学際NDBデータ研究基盤」を活用して、病名決定ロジックを適用した患者数の算出と疾病別医療費の推計を行うことが目的である。

病名決定ロジックは、複数の医学会および専門医の協力を得て、レセプト病名を基にした発症群・合併症群・重症群の定義を確立し、臨床的知見に基づいて決定する。算出した結果は、学会にフィードバックして妥当性について検討する。これらのプロセスを経て、NDBデータ利用における病名決定ロジックについて、わが国の問題点が明らかにしながら、これまで公表されてこなかった患者数、疾病別医療費の効果的な公表のあり方について検討する。

本研究では、妥当性の高い統計情報を得ることを目的として、本研究期間において病名、診療行為、医薬品の処方履歴を考慮した病名決定ロジックを作成する。

#### B. 研究方法

初年度は、生活習慣病を対象に、病名や処置等のレセプト情報の定義（ロジック）を作成した。日本糖尿病学会、日本動脈硬化学会、日本腎臓学会、日本高血圧学会との連携のもと、学会の理事・評議員を中心に、検討を行った。

2009-2014年のNDBにおける医科・調剤レセプトデータを活用し傷病名マスタより各疾患の病名を定義した。同様に処置や医薬品も定義をした。疾患によっては、都道府県別、医療機関規模別、性別、年齢階級別患者数や2014年の全国人口を基準人口として、男女ごとに年齢調整した都道府県別受療率（10万人対）を算出する。

ロジックを実装した解析を実施する前段階として、NDBによる抽出の実行可能性を検証する目的で糖尿病、急性心筋梗塞、高血圧、悪性新生物について病名のみ患者数の集計を実施した。さらに、悪性新生物の中から大腸がんを対象とし、公衆衛生学・疫学についての専門家、大腸がんに関する臨床専門家とエキスパートパネルを

作成し、病名決定ロジックを作成した。ロジックの作成において、診療行為、医薬品の特定には最新の診療ガイドラインを参照するが、ガイドラインにおける推奨と実臨床の診療行為・医薬品処方には乖離が存在する。こうしたエビデンス診療ギャップが、NDB 解析に影響を及ぼさぬようにエキスパートパネルで協議の上、該当疾患患者の可能性を示す診療行為・医薬品処方を特定する。具体的には、手術、放射線治療、薬物療法(化学療法)に関する治療行為を想定し、必要に応じて加算などもロジックに含むこととした。また、検査結果などの臨床的な情報が不足しているため、病名決定ロジックにおいて罹患状態である人を見落とし、過少報告する可能性が示唆される。そのため、「罹患である人を見落とすかもしれないが確実に罹患と考えられる」定義を「狭義の定義」とし、「罹患でない人が含まれるかもしれないが、見落とさない」定義を「広義の定義」として設定した。

上記ロジックに基づき、解析計画を作成し 2014 年度を対象にした集計を試みた。ロジックを活用するように効果を検証するために、病名のみを集計と、広義と狭義の定義によるロジックを活用した集計を行い、それぞれの結果を比較した。

### C. 研究結果

2014 年度において全国の糖尿病、悪性新生物、急性心筋梗塞、脳卒中、大腸がんの患者数はそれぞれ 15,001,988 人、12,895,963 人、7,650,226 人、5,065,805 人であった。広義の大腸がんについての患者数は 1,113,958 人(男性: 657,294 人、女性: 486,664 人)であり、同様に狭義の

大腸がん患者数では 1,081,466 人(男性: 613,081 人、女性: 468,385 人)であった。本研究でエキスパートパネルとの協働により作成したロジックの内、狭義の診療行為と医薬品処方とを合わせた集計では、広義の大腸がん患者数は 248,144 名(男性: 148,961 人、女性: 99,183 人)、狭義の大腸がん患者数は 239,658 人(男性: 145,070 人、女性: 94,588 人)であった。大腸がんにおいては、ロジックの使用によって、広義の場合で 22.3%、狭義の場合で 22.2% まで患者数が減少することが分かった。

### D. 考察

NDB 等、レセプトデータを使用した解析において、特定の疾患についての患者数、受療者数を集計する際に、レセプトで使用されている「傷病名」をそのまま使用した単純集計は妥当性について課題が指摘されている。第一に、レセプトに記載されている病名には、検査などを実施する際に付与される疑い病名を含んでおり、この疑い病名が付与されているレセプトを含んで患者数を集計した場合、患者数を過大評価すると考えられる。そのため、疑い病名フラグを使用し、単純集計から除外するなどして、この「疑い病名」を含んだレセプトデータについての処理を解析計画に含む必要がある。第二に、レセプト病名には複数の病名を記載されることが許容されるため、主病名フラグに対する処理を含んだ解析を実施しても、その集計が当該疾患の患者像を反映した値であるかについては検証ができないという課題がある。第三に、レセプトに記載されている病名の精度についての課題である。

NDB データ分析における病名決定ロジックを通して、NDB 利用の課題が明らかになるとともに、患者数の算出、疾病別医療費の推計、効果的な公表のあり方について検討することが必要である。

翌年度は参加学会を増やし、より精度の高い病名決定ロジックの研究開発を目指す。

海外調査は、希少疾患や長期観察が必要な疾患の検討を行う。例えば、米国ではメディケア加入者に関連が深い 21 種類の慢性疾患を対象に病名決定ロジックを医学会と共に開発し、経年追跡調査が可能となっている。台湾では、癌、糖尿病、精神疾患、高額な医療費を要する疾患といった 15 種類を対象にしている。そこで、米国 CMS(Centers for Medicare & Medicaid Services)の 21 種類の疾患に関連する診断コード、処置コード、診療請求タイプなどの詳細なデータソースに関しての情報収集を行うと同時に、それらを使ったロジ

ックに関するノウハウを得る予定である。

#### **E. 結論**

現在の NDB オープンデータは年々公開データの量と質が向上しているが、疾病別患者数データは公表されていない。本年度は、生活習慣病を中心として性別、年齢階級別、都道府県別の患者数算出を実施した。当該研究の成果が NDB オープンデータに資するよう、今後も対象疾病を拡大する。

#### **F. 健康危険情報**

該当なし

#### **G. 研究発表**

該当なし

#### **H. 知的財産権の出願・登録状況**

該当なし