

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

研究要旨

本研究の目的は、レセプト情報を基にした病名決定ロジックの定義を作成すると同時に、そのプロセスを示すことである。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、公衆衛生学・疫学についての専門家、各がんに関する臨床専門家とエキスパートパネル、各連携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も行う。

今回、各種がん（子宮頸がん、前立腺がん、大腸がん、乳がん、肺がん）、糖尿病（1型糖尿病）、脂質異常症、脳卒中、心不全・虚血性心疾患・急性心筋梗塞、関節リウマチ、透析、高血圧等の病名ロジックを作成・公開することができた。一方で、研究者間での病名ロジックの完成度、検証が十分ではない等の理由から、報告書に掲載できなかった疾患も存在する。また、我々のIDの名寄せ技術により、正確な患者数を算出する上で有効であった。NDBの識別子のユニーク数はID1が2.73億件、ID2が2.28億件（NDB6年分）となり人口を大きく超えている。そのため、識別子遷移グラフ処理による識別子併合を行うことで、1.38億件にまで識別子のユニーク数を併合することができる。個人追跡可能性は、約70%から99%と大幅に向上することになり、患者数のダブルカウントの可能性も低いと考えられた。

各学会のワーキンググループおよびエキスパートパネルに共通する課題は、電子レセプトマスタにおける網羅性および各種病名や医薬品コードとの対応である。医薬品の場合は、国際分類コード（解剖治療化学(ATC)分類等）とレセプト電算コードとの対応であるが、新薬の追加などにも対応して網羅性を確保しなければならない。

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し、今回作成したロジックや定義書（病名や医薬品に対応する各種コードと電子レセプトコードの）を共有し、更新していくことが望ましい。

A. 研究目的

本研究の目的は、レセプト情報を基にした病名決定ロジックの定義を作成すると同時に、そのプロセスを示すことである。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。

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。例えば、我々の班における解析では、妊娠や分娩に関連する病名等で、男性に付与されるとは考えにくい

病名が男性に付与されているレコードが年間で数百件確認されるなどのエラーと思われる事象を確認している。本研究では、上記の課題から発生するバイアスを排除し、妥当性の高い統計情報を得ることを目的として、診療行為と医薬品の処方履歴を考慮した病名決定ロジックを作成し解析を実践することで、これまで公表されてこなかった患者数について検討をおこなう。

B. 研究方法

初年度は、生活習慣病を対象に、糖尿病、高血圧、脂質異常症の病名や処置等のレセプト情報の定義を作成した。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）を算出した。

次年度は、各種がん（子宮頸がん、前立腺がん、大腸がん、乳がん、肺がん）、糖尿病（1型糖尿病）、脂質異常症、脳卒中、心不全・虚血性心疾患・急性心筋梗塞、関節リウマチ、透析、高血圧等の患者数の集計を実施した。各種がんにおいては、公衆衛生学・疫学についての専門家、各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。なお、前者は特異度を重視しており、後者は感度を重視したデザインと言うことができる。上記ロジックに基づき、解析計画を作成し集計した。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。その他の疾患は、日本糖尿病学会、日本高血圧学会、日本動脈硬化学会、日本腎臓学会、日本リウマチ学会と連携し、各学会内にワーキンググループを設置する等して、病名コード、処置コード、医薬品コードなどを設定し、検討を進めた。

C. 研究結果

各種がん(子宮頸がん、前立腺がん、大腸がん、乳がん、肺がん)、糖尿病(1型糖尿病)、脂質異常症、脳卒中、心不全・虚

血性心疾患・急性心筋梗塞、関節リウマチ、透析、高血圧等について試行的に患者数を算出した。そのロジックや集計結果は、資料1を参照されたい。一方で、研究者間での病名ロジックの完成度、検証が十分ではない等の理由から、報告書に掲載できなかった疾患も存在する。

D. 考察

初年度は、NDB を用いたわが国の悉皆データを用いて、病名と医薬品情報をもとに各疾患の人数の算出過程から、NDB データ分析における問題点が明らかになる。特に、患者数の算出、疾病別医療費の推計を通して、効果的な公表のあり方について検討する。初年度は、特に都道府県別に集計する際の患者数の算出が課題となった。一人の患者が異なる都道府県の医療機関で受診している場合は、どこの都道府県とするか、あるいは一定のロジックを作成して按分するのが問題となった。

次年度は、本研究の各学会のワーキンググループおよびエキスパートパネルに共通する課題が、電子レセプトマスタにおける網羅性およびレセ電コードと対応する形での各種病名や医薬品などの国際分類コード(ICD10 や解剖治療化学(ATC)分類等))の付与であった。網羅性については、新薬の追加等により数か月で更新されることが原因の一つであった。我々は、医薬品マスタのレセ電コードと対応する形での国際分類コード(解剖治療化学(ATC)分類等))を準備・提供することで、エキスパートパネルおよび学会のワーキンググループのメンバーの医薬品のコード指定の労力が軽減する措置を講じた。しかし、

これらコードとの対応は労力と専門知識が必要とされ、病名決定ロジック作成時には継続的に労力が必要となる課題である。そのため、エキスパートパネルや学会のワーキンググループの労力が軽減されるために、病名や医薬品に対応するコードを公開・共有し、多数の研究者が共同し更新していくことが望ましい。

ID の名寄せ技術を活用することも、正確な患者数を算出する上で必要であった。現在、NDB の識別子のユニーク数は ID1 が 2.73 億件、ID2 が 2.28 億件 (NDB 6 年分) となる。そのため、識別子遷移グラフ処理による識別子併合を行うことで、日本国民の数に近い 1.38 億件 (人) にまで識別子のユニーク数を併合することができた。個人追跡可能性は、約 70% から 99% と大幅に向上することになり、患者のダブルカウントの可能性も極めて低いと考えられた。

米国ではメディケア加入者に関連が深い 23 種類の慢性疾患 (Chronic Conditions) 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、日本の NDB データの可視化ツールの具体的なイメージを得るために、米国 CMS (Centers for Medicare & Medicaid Services) の 23 種類の疾患に関連する診断コード、処置コードなどの情報収集を行い、日本においては米国 CMS の Chronic Conditions の可視化指標 (有病率、医療費、入院率、再入院率、入院死亡率等) に加えて、都道府県別、年齢階級別、性別の表示に加えて、年齢調整済受療率 (都道府県別)、時系列の可視

化機能を実現することで、国民に対する情報提供機能としての展開もあると考えられる。

E. 結論

各種がん (子宮頸がん、前立腺がん、大腸がん、乳がん、肺がん)、糖尿病 (1 型糖尿病)、脂質異常症、脳卒中、心不全・虚血性心疾患・急性心筋梗塞、関節リウマチ、透析、高血圧等の病名ロジックを作成・公開することができた。一方で、研究者間での病名ロジックの完成度、検証が十分ではない等の理由から、報告書に掲載できなかった疾患も存在する。また、ID の名寄せ技術の開発は、正確な患者数を算出する上で有効であった。各学会のワーキンググループおよびエキスパートパネルに共通する課題は、NDB システムを運用する上で利用されており電子レセプトマスタにおける網羅性および各種コードとの対応である。そのため、本報告書の資料 1 で示した様に病名や医薬品に対応するコードおよび推計ロジック等を公開・共有することが望ましい。

F. 健康危険情報

該当なし

G. 研究発表

論文発表 (査読有)

1. Hideki Hashimoto, Makoto Saito, Jumpei Sato, Kazuo Goda, Naohiro Mitsutake, Masaru Kitsuregawa, Ryoza Nagai, Shuji Hatakeyama, International Journal of Infectious Disease, Vol91, 1 - 8, 2020

Indications and classes of outpatient antibiotic prescriptions in Japan: A descriptive study using the national database of electronic health insurance claims, 2012-2015, *International Journal of Infectious Diseases*, 91 (2020), 1-8

2. 満武巨裕, 石川智基, 佐藤淳平, 合田和生, 喜連川優、レセプト情報等データベースの利活用により作成した国際統計報告に関する検討、*医療情報学*、Vol.39, No.4, pp.189-194 (2020.4.20)

学会発表

1. Jumpei Sato, Kazuo Goda, Masaru Kitsuregawa, Naoki Nakashima, Naohiro Mitsutake, Novel Analytics Framework for Universal Healthcare Insurance Claims Database, *MEDINFO 2019: Health and Wellbeing e-Networks for All*, 264:1578-1579, 2019

2. Jumpei Sato, Hiroyuki Yamada, Kazuo Goda, Masaru Kitsuregawa, Naohiro Mitsutake, Enabling Patient Traceability Using Anonymized Personal Identifiers in Japanese Universal Health Insurance Claims Database, *AMIA 2019 Informatics Summit*, 345-352, 2019

3. Kazutoshi Umemoto, Kazuo Goda, Naohiro Mitsutake, and Masaru Kitsuregawa, A Prescription Trend Analysis using Medical Insurance Claim Big Data, *35th IEEE International Conference on Data Engineering (ICDE 2019)*, 1928-1939, 2019

4. 満武巨裕, 石川智基, 佐藤淳平, 合田和生, 喜連川優、レセプト情報等データベースの利活用により作成した国際統計報告に関する検討、第23回日本医療情報学会春季学術大会 シンポジウム、熊本、2019年6月

5. 佐藤淳平, 山田浩之, 合田和生, 喜連川優、満武巨裕、健康医療データベースに於ける暗号化された識別子に基づく患者の追跡方法の検討と実験に基づく考察、第10回データ工学と情報マネジメントに関するフォーラム、第16回日本データベース学会年次大会、2018.03.04-6

6. 石川智基, 満武巨裕, 佐藤淳平, 合田和生, 喜連川優、レセプト情報等データベースを活用した放射線画像診断における利用状況可視化の試み、第47回日本放射線技術学会秋季学術大会、大阪、2019年10月

7. 和氣宗, 三浦克之, 田中佐智子, 大屋祐輔, 野出孝一, 伊藤裕, 佐藤淳平, 合田和生, 喜連川優、石川智基、満武巨裕、わが国における高血圧治療薬の使用実態: レセプト情報・特定健診等情報データベースを用いた分析、第42回日本高血圧学会総会、東京、2019年10月

8. 三浦克之, 田中佐智子, 和氣宗, 大屋祐輔, 野出孝一, 伊藤裕, 佐藤淳平, 合田和生, 喜連川優, 石川智基, 満武巨裕、我が国の全国および都道府県別の高血圧受療者数および受療率: 厚生労働省レセプト情報・特定健診等情報データベースを用いた分析、第42回日本高血圧学会総会、東京、2019年10月

9. 高橋由光, 内海貴裕, 西川佳孝、

星野伸晃、堀松高博、佐藤淳平、合田和生、喜連川優、石川智基、満武巨裕、中山健夫、NDB を活用した大腸がんにおける患者数・医療費推計の試み、第 57 回日本医療・病院管理学会、新潟 2019 年 11 月

10. 石川智基、満武巨裕、佐藤淳平、高橋由光、中山健夫、内海貴裕、西川佳孝、星野伸晃、堀松高博、合田和生、喜連川優、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:大腸がん、第 57 回日本医療・病院管理学会、新潟、2019 年 11 月

11. 石川智基、満武巨裕、佐藤淳平、合田和生、喜連川優、放射線画像検査の利用実態における地域差分析、第 39 回医療情報学連合大会、千葉、2019 年 11 月

12. 満武巨裕、医療政策・経営とビッグデータ解析: NDB 大規模データ解析と活用、第 57 回日本医療・病院管理学会 学術総会 日本医学会連合 合同シンポジウム、新潟・2019

13. 山縣 邦弘、角田 亮也、満武 巨裕、中島 直樹、NDB ビッグデータにおける CKD および腎代替療法の検討、第 62 回

日本腎臓学会、名古屋・2019.06.23

14. 満武巨裕、エビデンスの飛躍的創出を可能とする超高速・超学際次世代 NDB データ研究基盤構築を活用した糖尿病関連エビデンスと課題、日本医療情報学会合同シンポジウム:世界最大級の医療 Real World Data、NDB を用いた糖尿病研究、第 61 回日本糖尿病学会年次学術集会(2018 年 5 月 24 日)

15. 後藤 温、野田光彦、満武巨裕、NDB からみた糖尿病とその併存疾患のパノラマビュー、日本医療情報学会合同シンポジウム:世界最大級の医療 Real World Data、NDB を用いた糖尿病研究、第 61 回日本糖尿病学会年次学術集会(2018 年 5 月 24 日)

16. 満武巨裕、日本の保健医療分野における国際統計報告の現状と展望、第 77 回日本公衆衛生学会総会(2018 年 10 月 24 日~26 日)

H.知的財産権の出願・登録状況

該当なし