

レセプト情報における病名・病名開始日を用いた 名寄せロジック

竹下 沙希^{*1}、西岡 祐一^{*1}、久保 慎一郎^{*1}、明神 大也^{*1}、
野田 龍也^{*1}、今村 知明^{*1}

^{*1} 奈良県立医科大学 公衆衛生学講座

Identification by diagnosis-related codes and diagnosis-date using receipt-database

Saki Takeshita^{*1}, Yuichi Nishioka^{*1}, Shinichiro Kubo^{*1}, Tomoya Myojin^{*1},
Tatsuya Noda^{*1}, Tomoaki Imamura^{*1}

^{*1} Department of Public Health, Health Management and Policy, Nara
Medical University, Kashihara, Nara, Japan

レセプト情報の活用には正確な名寄せが重要である。本研究では、保険者が変化しても名前主体の ID を用いずに、正確な名寄せを可能にする新たなロジックを開発する。我々は、2013 年度から 7 年間の奈良県 KDB のデータを用い、生年月・性別・傷病名・診療開始日の組み合わせにより新名寄せ ID を生成した。奈良県国民健康保険団体連合会が有する被保険者台帳をもとに作成した台帳 ID を正解データとした。被保険者番号主体の個人 ID のうち台帳 ID によって他の個人 ID と紐づき、同一医療機関の外来レセプトが発生したものを対象とし、同一の異なる台帳 ID に対応する新名寄せ ID が紐づく場合を正/誤と判定した。対象の台帳 ID 数は 69,988 であった。真陽性(正しい紐付けの事例)は 62,643 件、偽陰性(紐付けられなかった事例)は 7,345 件、感度 0.90、特異度 1.00、陽性的中度 0.99、陰性的中度 1.00 であった。また、誤った紐づけは 100 万組あたり 0.047 組であった。

キーワード KDB, Patient-Matching Technique, Administrative Claims Database, Electronic Health Record, Patient Identification

1. はじめに

国保データベース(以下、KDB)とは、国民健康保険(以下、国保)と後期高齢者医療制度(以下、後期)のほぼ全患者のレセプト情報が格納されている。KDB では個人情報情報を削除し、なおかつ同一人物の情報を識別できるよう、ハッシュ化された個人 ID を提供できる仕組みが構築されている。[1]名寄せはレセプト分析最大の課題であるが、現在提供されている個人 ID 生成に用いる被保険者番号などのデータでは名寄せに不十分で、一年で総人口の約一割が変化するといわれている。データベースにおける名寄せとは、同一人物であるかどうかを判別して、同一人物には共通の ID を、別人には別の ID を付与することである。[2]我々は奈良県国民健康保険団体連合会が有する被保険者台帳をもとに作成した個人 ID(以下、台帳 ID)で名寄せを行うロジックを別途開発しているが、KDB を扱う全ての研究者がそれぞれ台帳 ID を実装するのは時間や費用の面から現実的ではないと考えられる。

レセプト情報の十分な利活用には、今ある情報から簡単に、十分な名寄せを行えるロジックの開発が必要である。汎用性の高い名寄せ手法の開発は、レセプト研究において重要な課題となっている。

2. 方法

2013 年度から 7 年間の奈良県 KDB(被保険者約 50 万人)のデータを用い、生年月・性別・傷病名・診療

開始日の組み合わせにより「新名寄せ ID」を生成した。以下に「新名寄せ ID」の名寄せ手法を示す。

(1) 名寄せ用中間データの作成

KDB に記載されている個人 ID/台帳 ID/レセプト年月/生年月/性別/傷病名/診療開始日を抽出した。

(2) 名寄せの除外対象

2013 年 4 月のレセプトと 2020 年 3 月のレセプトの両方に存在する個人 ID は、7 年間で ID の変化がなかった個人のものであるため名寄せ対象から除外した(以下、「名寄せ除外対象 ID」)。ただし、2013 年 4 月または 2020 年 3 月に 75 歳を迎えるものは国保から後期に原則移動するため名寄せ対象とした。名寄せ除外対象 ID に発生した傷病名と診療開始日の組み合わせは(3)名寄せ作業①に使用しなかった。

(3) 名寄せ作業①

ある傷病名と診療開始日の組み合わせ(N 組)をもつ生年月と性別の階級数が 1 でかつ個人 ID が複数あった場合に、生年月、性別、傷病名と診療開始日の組み合わせ(N 組)が全て一致した個人 ID を名寄せ候補 ID とした。なお、N=2, 3...について(4)(5)を名寄せ候補 ID が出力されなくなるまで繰り返した。

(4) 名寄せ用中間テーブルの作成

傷病名と診療開始日の組み合わせ数 N=1 の名寄せ用中間データを「基本テーブル」とする。N≥2 について、傷病名と診療開始日の組み合わせが(N-1)組

記載された中間テーブル(以下、「N - 1 テーブル」と基本テーブルを用い、両テーブルで個人 ID、性別、レセプト年月、生年月が全て一致した場合、その個人 ID について傷病名コード、診療開始日を追加することで、傷病名と診療開始日の組み合わせが N 組記載された中間テーブル(以下、「N テーブル」)を作成する。

(5) 名寄せ候補 ID の特定

N $\geq$ 2 について、ある傷病名と診療開始日の組み合わせ(N 組)をもつ生年月と性別の階級数が 1 でかつ該当する個人 ID が複数あった場合に、(4)で作成された N テーブルについて、生年月、性別、傷病名と診療開始日の組み合わせ(N 組)が全て一致する個人 ID を名寄せ候補 ID とした。N=3 における、名寄せ候補 ID の例を Table.1 に示す。Table.1 では、「A」と「B」、「D」と「E」は傷病名コードと対応する診療開始日の 3 つの組み合わせが他の性別・年齢階級で発生していないればそれぞれ名寄せ候補 ID の組となる。

Table.1 N = 3 における名寄せ候補 ID の例

個人ID	生年月	性別	レセプト年月	傷病名コード1	診療開始日1	傷病名コード2	診療開始日2	傷病名コード3	診療開始日3
A	198005	F	201708	1234567	20130501	2345678	20161001	3456789	20181222
B	198005	F	201803	1234567	20130501	2345678	20161001	3456789	20181222
C	199211	M	201507	4567890	20150625	5678901	20161001	6789012	20150303
D	196803	F	201810	2345678	20180404	3456789	20140824	4567890	20130929
E	196803	F	201812	2345678	20180404	3456789	20140824	4567890	20130929

※この例では個人IDについてAとB、DとEの組が名寄せ候補IDの組となる

(6) 個人 ID レセプト発生期間の重複による除外

一方のレセプト発生年月の最小値が他方の最小値以下でかつ一方のレセプト発生年月の最大値が他方の最大値以上である時は除外した。

(7) 名寄せ作業②

基本テーブルを参照し、傷病名と診療開始日と性別と生年月が一致する個人 ID どうして一方の最終受診年月が他方の最初の受診年月以下であるとき名寄せ候補 ID の組とした。

(8) 台帳 ID による名寄せとの比較 (検証作業)

台帳 ID を正解データ(reference standard)として検証作業を行った。同一医療機関の外来レセプトが発生した個人 ID のうち台帳 ID によって他の個人 ID と紐づき、同一医療機関の外来レセプトが発生したものを対象とした。また、同一医療機関の外来レセプトが発生しなかった場合についても検討した。

3. 結果

対象の台帳 ID 数は 69,988 (個人 ID の組み合わせの総数は 9,796,570,300 件でそのうち正しい紐づけは 69,988 件)であった。真陽性(正しい紐付けの事例)は 62,643 件、偽陰性(紐付けられなかった事例)は 7,345 件、感度 0.90、特異度 1.00、陽性的中度 0.99、陰性的中度 1.00 であった。また、誤った紐づけは 1 億組あたり 4.7 組であった。一方で、同一医療機関を受診しなかった個人 ID の組み合わせ(23,222 件)では、名寄せはほぼ不可能であった。

4. 考察

同一医療機関を受診した例では誤った紐づけを最小限に抑えつつ名寄せを行える手法の開発に成功した。同一医療機関を受診している場合は患者の保険制度が移行しても電子カルテ上の傷病名・診療開始日が引き継がれるために多くの個人について名寄せが可能であったと考えられる。偽陰性は 10%程度にみられた。同じ傷病名と診療開始日の組み合わせが異なる生年月を持つ ID どうしてみられ、特異的でなかったため名寄せされなかった。

本ロジックは、レセプト情報・特定健診等情報データ

ベース(以下、NDB)の名寄せにも使える可能性がある。NDB は、病院や診療所から保険者に請求された電子レセプトデータと保険者が保有する特定健診・特定保健指導のデータを、一部の情報を匿名化・削除した上で、格納・構築されているデータベースである。日本の全保険診療情報であり、1億人規模のコホート研究が実施できると期待されている[3]。NDB は保険者番号主体の ID1 と氏名主体の ID2 の 2 種類の ID が付与されたが[4]不完全であり、我々は ID1・2 などから ID0 を作成した。これによって同一人物の長期追跡性が向上したが [3] [5]、ライフイベントなどで保険と氏名が同時に変化すると名寄せは不可能である。その後、ID1 を改善した ID3 が作られ、ID0 を用いた名寄せ精度は向上したが、未だ十分ではない。さらに介護 DB と NDB の接続のためにカナ氏名と生年月日をハッシュ化し、ID4 が作られたが精度は不明である。今後の全てのデータを結合可能な ID5 が提供されようとしているが、付与割合自体が現状では不明である。本ロジックを使用すると、被保険者番号や氏名主体の ID と独立した名寄せを行え、名寄せの質向上に役立つ。

本手法は、既に提供されてきた情報のみから名寄せを実施できる大きな利点を持つ。今後、「医療等 ID」が付与されたとしても、過去のレセプトについて名寄せの問題は残存する。本手法を単体あるいは他の手法と組み合わせることで、今後あらゆるレセプト研究の名寄せに応用できる。

5. 結語

台帳 ID は奈良県と他の都道府県とで仕様が異なり、他の地域の KDB や NDB 等の他のデータベースへの展開は難しい。そこで本研究では、提供後のレセプトデータから個人を紐づける新たなロジックを開発した。

個人 ID が変化する前後に同一医療機関を受診していない事例での名寄せロジック開発や、NDB への本ロジックの適用が今後の課題である。

参考文献

- [1] 木村真也:レセプトにおける匿名化名寄せ技術と傷病名辞書, 公衆衛生 71(11)939-942, 2007.
- [2] 谷口学:JST の人名名寄せと, JDream II での活用方法, 薬学図書館. 56(4)292-262, 2011.
- [3] 西岡祐一, 野田龍也, 今村知明:世界最大の医療ビッグデータ NDB (ナショナルデータベース) の活用と今後の展望, 日本整形外科学会雑誌 93(11)959-979, 2019.
- [4] 久保 慎一郎, 野田 龍也, 明神 大也, 他:レセプト情報・特定健診等情報データベース(NDB)の臨床研究における名寄せの必要性和留意点, 日本健康開発雑誌 38 11-18, 2017.
- [5] 野田龍也, 久保慎一郎, 明神大也, 他:レセプト情報・特定健診等情報データベース(NDB)における 患者突合(名寄せ)手法の改良と検証, 厚生指針 64(12)7-13, 2017.