

一般口演 | 第40回医療情報学連合大会（第21回日本医療情報学会学術大会） | 一般口演

一般口演19

データベース・データウェアハウス

2020年11月21日(土) 11:15 ~ 12:38 G会場 (イベントホール・特設会場2)

[4-G-1-05] レセプト情報・特定健診等情報データベース（NDB）における患者突合の精度向上に関する手法開発

*久保 慎一郎¹、野田 龍也¹、西岡 祐一¹、明神 大也¹、東野 恒之²、今村 知明¹（1. 奈良県立医科大学公衆衛生学講座, 2. (株)三菱総合研究所 ヘルスケア・ウェルネス事業本部）

*Shinichiro Kubo¹, Tatsuya Noda¹, Yuichi Nishioka¹, Tomoya Myojin¹, Tsuneyuki Higashino², Tomoaki Imamura¹（1. 奈良県立医科大学公衆衛生学講座, 2. (株)三菱総合研究所 ヘルスケア・ウェルネス事業本部）

キーワード：NDB, Rezept, matching, ID, number of patient

【目的】レセプト情報・特定健診等情報データベース(以下、NDB)とは、日本の保険診療の悉皆データである。本研究者らによってNDBの同一患者の複数レセプトを紐付ける変数「ID1」と「ID2」を用いた名寄せ手法により、「ID0」を開発し、様々な研究者が利用している。しかし、「ID1」の表記ゆれを解消した「ID1N」等が追加されたことや、一部年齢階層が人口を上回るID数であったことから、名寄せを効率化させた個人ID（ID0v2）を開発した。

【方法】本研究は医療計画策定に係る評価指標作成の一環として行った。6年分のNDBレセプト(医科、調剤、歯科)を用いた。従来の「ID0」で使用していた「ID1」、「ID2」、「診療年月」、「転帰区分」に加え「ID1N」を利用した名寄せアルゴリズムを開発した。また、「歯科」レセを加えたマージ方法を検討した。さらにIDの重複をみる期間を変えたときの患者数の変化を検証した。アルゴリズムの妥当性を検証するため、従来のID0による名寄せとID0v2による名寄せを、平成27年10月1日時点の推計人口と比較した。

【結果】総人口が127,094,745人、従来の集計方法(ID0)では112,872,483人であったのに対し、ID0v2では110,704,147人であった。年齢階級ごとに名寄せ率(総人口に占めるID数)を比較すると、例えば男性の80～84歳は1.02で変化が見られなかったものもあれば、男性の5～9歳は1.03から1.01に減っている階級も見られた。

【考察】ID1Nでの名寄せ率の向上や、レセプト重複期間の見直し、つなぐ際に最も近い診療年月から結合するなどの名寄せ方法を取ることによって、以前に比べて名寄せ率は向上した。年齢階層によって名寄せ率にばらつきがみられたのは、ライフイベントによって保険種別が代わりやすい世代の名寄せが行えたためと考えられる。

レセプト情報・特定健診等情報データベース(NDB)における 患者突合の精度向上に関する手法開発

久保 慎一郎^{*1}、野田 龍也^{*1}、西岡 祐一^{*1}、明神 大也^{*1}、東野 恒之^{*2}、今村 知明^{*1}

^{*1} 奈良県立医科大学 公衆衛生学講座、^{*2} (株)三菱総合研究所 ヘルスケア・ウェルネス事業本部

Development of a Method for Increasing the Accuracy of Patient Data Matching for the National Database of Health Insurance Claims and Specific Health Checkups of Japan

Kubo Shinichiro ^{*1}, Noda Tatsuya ^{*1}, Nishioka Yuichi ^{*1}, Myojin Tomoya ^{*1},
Higashino Tsuneyuki^{*2}, Imamura Tomoaki^{*1}

^{*1} Department of Public Health, Nara Medical University,

^{*2} Healthcare and Wellness Division, Mitsubishi Research Institute, Inc.

The National Database of Health Insurance Claims and Specific Health Checkups of Japan (hereinafter "the NDB") provides complete data on healthcare services covered by the National Health Insurance system in Japan. We previously developed the variable ID0 using a patient-matching method based on the variables ID1 and ID2 which linked multiple health insurance claims of the same patient recorded in the NDB. The variable ID0 is currently used by various researchers. We now have developed the individual ID variable ID0v2 based on a more precise patient-matching method. This study was conducted as part of work to prepare evaluation indicators for prefectural healthcare plans. We used NDB data covering 6 years of health insurance claims for medical care, pharmaceutical dispensing, and dentistry. Our patient-matching algorithm utilizes the variable ID1N in addition to ID1, ID2, date of healthcare service, and clinical outcome, which are used for ID0. We also examined a method to merge data taking into account health insurance claims for dentistry. Furthermore, we examined variations in the number of patients that resulted from changing the period considered when checking for multiple ID numbers assigned to a single individual. To examine the validity of our algorithm, we compared the estimated population of Japan as of October 1, 2015 with the number of individuals identified with ID0 and the number of individuals identified with ID0v2. The results were 127,443,493 (population), 112,164,891 (ID0), and 110,284,835 (ID0v2). The patient-matching rate is higher with the new patient-matching method as a result of using ID1N, reexamining the period considered for overlapping claims, and linking data starting with the most recent month of healthcare service.

Keywords: NDB, Rezept, matching, ID, number of patient

1. 目的

レセプト情報・特定健診等情報データベース(以下、NDB)は、病院や診療所から国に送信される電子レセプトデータと特定健診等のデータを個人が特定されないように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。国民皆保険制度をとる日本における保険診療の悉皆データであり、2009年4月～2019年12月診療分で約187億6,800万件(2020年3月末時点)¹⁾のレセプトデータが蓄積されるなど、世界最大級の健康関連データベースといえる。このデータを有効に活用することで日本全体の臨床研究(コホート研究等)や政策研究が強力に推進できると期待されている。

ただし、NDBにはさまざまな問題点や障壁が存在する。政策用途を目的としたデータであるため、利用するためには申請が必要である。申請し有識者の審査を受けてデータが切り出されるには早くても半年程度の時間がかかる。そのうえ、著者らが行った先行研究では、NDBの巨大なサイズによる処理の限界と、診療報酬請求のために設計されているレセプトの構造がそのままでは研究目的での利用に適さない形式となっていることから研究者が個人で分析をするのが困難である問題点が挙げられた²⁾。加えて、最も致命的なものとしてNDBに付与されている2種類の個人識別IDが経過とともに変化しやすいという問題が挙げられる。なお、NDBには2020年8月現在で一生変えられない個人IDは含まれていない。代わりに、個人紐付け用の匿名変数として、主に「ID1」と「ID2」が用意されている。ID1は保険者番号、被保険者証等記号・番号、

生年月日、性別から個人情報保護のためにハッシュ関数と呼ばれる関数を用いて変換された英数字列であり、ID2は氏名、生年月日、性別から同様に変換された英数字列である。ところが、同一患者でも、就職・転職等で保険者は変化し、医療機関での表記ゆれ(例:渡辺と渡邊)や結婚・離婚等のライフイベントで氏名表記は変化するため、ID1、ID2ともに容易に変わり得ることが分かった³⁾。

この問題を解決するためには、NDBに一生変えられない個人IDを付与する必要がある。現在の方向性としては被保険者番号に個人単位の番号付きの保険証(個人単位被保番)を一人一人発行することによる「医療等ID」が導入される予定である。この個人単位被保番の制度については、現在の保険証の番号に個人番号を付与する必要があり、2020年度内に登録が完了する予定である。しかし実際には被保険者への周知と医療機関の保険証確認の徹底から、2021年4月診療分の稼働に出揃うか不透明である。また、個人単位被保番の導入が完了した場合も保険証が変更になった場合やこれまでの個人単位になっていない被保険者番号を紐づける際には「名寄せシステム」が必要となり、このシステムの稼働について現在詳細を取り決めている最中である。さらに、不変な個人IDの導入が実現しても過去データについては現行のID1とID2を用いるほかない。医療等ID導入に先立ちNDBでの患者IDを確立するため、筆者らの先行研究により、ID1とID2を組み合わせて一つの個人ID「ID0」を作成することに成功した⁴⁾。

現在は「ID0」を、様々な研究者が利用している。しかし、

「ID1」の情報をハッシュに変換する際に被保険者番号の「全角」・「半角」等の表記ゆれを統一することで、レセプト情報と特定健診・特定保健指導情報の紐づけを可能となった「ID1N」が近年追加された。また、現在の「ID0」では、一部年齢階層が人口動態調査の人口を上回る ID 数であったことから、名寄せの精度向上は課題が残存していた。加えて、ID を突合する際に、レセプトの重複期間(つなぐ前の ID と後の ID がかぶる期間。本来は保険証の変更とともに切り替わると考えられていた)を 3 か月と設定していたが、実際には長期間レセプトの重複が発生することが分かったため、重複期間を変化させ適切な「のりしろ」を探る必要があった。そのため名寄せを効率化させた個人 ID「ID0 version2(ID0v2)」を開発した。

2. 方法

本研究は医療計画策定に係る評価指標作成の一環として行った。6 年分の NDB レセプト(医科、調剤、歯科)を用いた。従来の「ID0」で使用していた「ID1」、「ID2」、「診療年月」、「転帰区分」に加え「ID1N」を利用した名寄せアルゴリズムを開発した。外来では、医師が発行する処方箋に基づき院外の薬局で調剤が行われることが多いことを考えると、医科入院外と調剤についてはペアでレセプトが発生することが多く、一括処理を行うと患者の紐づけ情報が多対多となり、純正なマッチングが困難になると懸念される。そのため、「ID0」の時点では DPC 及び医科(入院／入院外)レセプトを用いた名寄せ中間データと調剤レセプトを用いた名寄せ中間データを別々に作成し、両者を突合して名寄せを完成させる仕組みをとった。しかし今回はさらに、「歯科」レセプトを合わせてマージする必要があり、どのようにマージすることで適切な名寄せが可能となるのかその方法を検討した。さらに ID の重複をみる期間を変えたときの患者数の変化を検証した。アルゴリズムの妥当性を検証するため、従来の ID0 による名寄せと ID0v2 による名寄せを、2018 年度の人口動態統計および住民基本台帳の結果と比較した。以下に「ID0」の名寄せ手法をさらに高度化させた「ID0v2」の名寄せ手法を示す。

2.1 名寄せ用中間データを作成

名寄せするレセプトデータの一例を図 1 に示す。「医科レセプト」「DPCレセプト」「調剤レセプト」「歯科レセプト」から ID1、ID2、診療年月を抽出する。「ID0」では転帰区分の死亡情報を抽出していたが「ID0v2」では除外した。複数月に渡って ID1 が同じ患者は同一人物とした。

ID1	ID2	医療機関コード	診療年月	都道府県コード	性別コード	年齢階級コード	入外区分
Xb0Vtu	aXxy1T	akdiiT	42505	45	1	220	1
Xb0Vtu	aXxy1T	akdiiT	42506	45	1	220	1
Xb0Vtu	aXxy1T	akdiiT	42507	45	1	220	1
b3zYx1	aXxy1T	akdiiT	42507	45	1	220	1
b3zYx1	ZzYyTz	qbmilc	42507	40	1	220	1
b3zYx1	ZzYyTz	qbmilc	42508	40	1	220	1
b3zYx1	ZzYyTz	qbmilc	42509	40	1	220	1
b3zYx1	ZzYyTz	qbmilc	42510	40	1	220	1
b3zYx1	ZzYyTz	qbmilc	42511	40	1	220	1
6YwxWV	ZzYyTz	qbmilc	42511	40	1	220	1
6YwxWV	ZzYyTz	qbmilc	42512	40	1	220	1
6YwxWV	ZzYyTz	qbmilc	42601	40	1	220	1
6YwxWV	ZzYyTz	qbmilc	42602	40	1	220	1
6YwxWV	ZzYyTz	qbmilc	42603	40	1	220	1

図 1 名寄せの事例

ID1 の変化により ID1 が途切れた場合は、途切れの前後で同一の ID2 を有するレセプトを探索して同一人物とした。探索

対象に候補となる ID2 が 2 つ以上見つかった場合、「ID0」では別人物が紐付けられてしまう事態を避けるため、紐付けずに名寄せ終了としていた。しかし、観察期間内に 2 回転職した人も除外されるため、「ID0v2」では最も近い年月の候補に名寄せを行った。

薬剤感受性試験は翌月請求となるケースがあるほか、資格喪失したレセプトにおいても審査支払機関を経由し保険者でチェックを行い再審査請求となるが、NDB では再審査請求レセプトは取り込まれないため、新旧の ID1 が併存しうる。この期間を 3 ヶ月間程度と考えられたため、まず ID1 が途切れた年月以降の ID2 を有するレセプトを探索し名寄せを行い、次に前月の 1 ヶ月間で同じ ID2 を有するレセプトを探索し名寄せし、最後に前々月の 1 ヶ月間で同じ ID2 を有するレセプトを探索し「ID0」の名寄せを行った。しかし、実際は長期間にわたって重複が発生していることがわかっており、このレセプトを探索する重複の期間(のりしろ)を1か月～6か月で変化させたときにどのように患者数が推移するか検証した。のりしろの判定で従来では、紐づける条件を先出 ID と後出 ID で表すと、先出 ID の最後と後出 ID の最初の年月のみで名寄せ判定をしていたが、循環参照を避けるため先出 ID の最初の年月が後出 ID の最初の年月より早くなるように、かつ先出 ID の最後の年月が後出 ID の最後の年月になるように条件を追加した。

これらの組み合わせをもとに、名寄せ中間データテーブルを 4 種類(医科・DPC・調剤・歯科)作成した。

2.2 中間テーブル同士の連結

医科・DPC・調剤・歯科の名寄せ用中間テーブルを連結し、「実際には同一人物である ID1 同士」の一対一対応表を作成した(対応セットの例:p8d89jss: ue8k22ue)。ここで、便宜上、対応セットの左の ID1 (例:p8d89jss)を「先出 ID1」、右の ID1 (例:ue8k22ue)を「後出 ID1」と呼ぶ。「先出 ID1」は、医科、DPC、調剤を問わず最初に出現した ID1、「後出 ID1」は二番目以降に出てきた ID1 である。「ID0」ではこのテーブルのマージ処理として、医科レセプトと DPC レセプトをマージさせたあと、調剤をマージしていた。この手法に関して、同様に「歯科」レセプトを追加でマージし順番に処理した手法(例:((医科+DPC)+調剤)+歯科)とすべてのレセプトを一括でマージした方法(例:医科+DPC+調剤+歯科)の両方を検討した。

2.3 名寄せ作業

2.2 の ID1 一対一対応表を利用し、対象となるレセプトに登場するすべての後出 ID1 を先出 ID1 に置き換えることで、同一人物の複数の ID1 を一種類に統合する。置き換えられた ID1 が別の対応セットの後出 ID1 であることもあるため、その場合は、さらに先出 ID1 へと置き換える。この作業を後出 ID1 がなくなるまで繰り返し、最終的に残った ID1 を新しい名寄せ変数「ID0」とした(図 2)。この手法は「ID0v2」でも同様の方法で行った。

後出ID1	先出ID1	ID2
ue8k22ue	p8d89jss	kerhu23y
Ajdke783	ue8k22ue	kerhu23y
78wmdjfg	Ajdke783	kerhu23y

ID1	ID0	年月From	年月To	ID2 From	ID2 To	転帰
p8d89jss	p8d89jss	201304	201306	382dhs87	dk328d87	
ue8k22ue	ue8k22ue	201307	201308	hs8ye726	ajd728uj	
Ajdke783	Ajdke783	201401	201403	la9js7d8	pq9e8eud	

ID1	ID0	年月From	年月To	ID2 From	ID2 To	転帰
p8d89jss	ue8k22ue	201304	201306	382dhs87	dk328d87	
ue8k22ue	Ajdke783	201307	201308	hs8ye726	ajd728uj	
Ajdke783	78wmdjfg	201401	201403	la9js7d8	pq9e8eud	

ID1	ID0	年月From	年月To	ID2 From	ID2 To	転帰
p8d89jss	Ajdke783	201304	201306	382dhs87	dk328d87	
ue8k22ue	78wmdjfg	201307	201308	hs8ye726	ajd728uj	
Ajdke783	78wmdjfg	201401	201403	la9js7d8	pq9e8eud	

ID1	ID0	年月From	年月To	ID2 From	ID2 To	転帰
p8d89jss	78wmdjfg	201304	201306	382dhs87	dk328d87	
ue8k22ue	78wmdjfg	201307	201308	hs8ye726	ajd728uj	
Ajdke783	78wmdjfg	201401	201403	la9js7d8	pq9e8eud	

図 2 名寄せの変容

これらによって作成された「ID0v2」の妥当性を検証するため、従来の「ID0」と NDB に用意されている「ID1」「ID1N」の性年齢階級別患者数を比較した。NDB の患者数については 2018 年 4 月 1 日～2019 年 3 月 31 日で一度でも医療機関に受診した者の数とした。人口と比較するために総務省統計局による人口動態調査(2018 年 1 月 1 日から 2018 年 12 月 31 日まで)と、住民基本台帳人口(2019 年1月1日現在の住民票に記載されている者の数)を比較した。NDB が年度集計で統計を暦年(年初め)とした理由は本研究が医療計画に資する目的で作成されており、医療計画は年度で評価されるためである。最大 3 か月に乖離はあるが、できるだけ近い時期を選定した。

なお、本研究の遂行にあたっては、「人を対象とする医学系研究に関する倫理指針」(2014 年 文部科学省、厚生労働省告示)を遵守し、奈良県立医科大学医の倫理審査委員会の承認(2019 年 9 月 26 日 受付番号 1123-5)を得た。

3. 結果

3.1 id の結合方法による患者数の検討

年齢階級別の人口・患者数分布を表 1 に示す。

表 1 年齢階級別の人口・患者数分布

年齢階級	住基 (H31.1.1)	人口動態統計 H30	id1	id1n	id0	一括マージ	個別マージ	個別マージ (転帰死亡除外)
0～4歳	4,917,358	4,762,514	5,234,327	5,153,993	4,867,147	4,808,873	4,773,351	4,770,357
5～9歳	5,330,568	5,120,025	5,520,613	5,441,449	5,228,710	5,153,264	5,117,770	5,116,550
10～14歳	5,510,498	5,338,375	5,329,645	5,261,479	5,109,093	5,043,135	5,018,728	5,017,731
15～19歳	5,933,818	5,803,056	5,197,404	5,140,469	4,993,289	4,936,217	4,917,460	4,916,250
20～24歳	6,421,424	5,963,102	5,510,317	5,459,788	5,125,520	5,055,526	5,021,165	5,019,469
25～29歳	6,434,689	5,893,367	5,774,144	5,728,491	5,317,604	5,247,349	5,204,928	5,203,059
30～34歳	7,152,129	6,690,447	6,357,534	6,298,752	5,923,902	5,849,292	5,807,212	5,804,785
35～39歳	7,859,598	7,483,085	6,869,096	6,804,620	6,467,608	6,389,512	6,348,570	6,345,345
40～44歳	9,152,072	8,925,616	7,760,731	7,685,114	7,352,874	7,263,896	7,220,299	7,215,434
45～49歳	9,831,545	9,506,041	8,332,107	8,248,719	7,908,314	7,812,278	7,766,062	7,758,343
50～54歳	8,477,896	8,228,969	7,418,313	7,336,736	7,030,250	6,942,308	6,900,336	6,890,099
55～59歳	7,662,101	7,554,378	6,910,902	6,825,409	6,522,620	6,438,700	6,397,050	6,383,057
60～64歳	7,574,556	7,521,408	7,135,942	7,064,546	6,564,617	6,512,497	6,454,408	6,431,760
65～69歳	9,157,379	9,314,416	8,621,206	8,453,805	8,070,963	7,953,976	7,894,496	7,851,713
70～74歳	8,245,416	8,195,221	8,039,788	7,938,222	7,747,923	7,681,502	7,640,719	7,577,709
75～79歳	6,941,222	6,904,067	7,918,346	7,889,328	6,739,673	6,764,664	6,665,248	6,577,223
80～84歳	5,236,368	5,330,212	5,170,074	5,157,620	5,138,652	5,132,796	5,122,850	5,000,495
85～89歳	3,454,586	3,505,515	3,606,830	3,598,661	3,577,714	3,575,733	3,567,319	3,410,055
90～94歳	1,631,884	1,672,000	1,841,077	1,837,824	1,823,836	1,824,323	1,819,690	1,677,656
95～99歳	449,396	438,000	562,376	561,664	557,491	557,774	556,566	486,491
100歳以上	68,990	69,000	97,726	97,634	97,091	97,113	96,972	78,503
合計	127,443,493	124,218,814	119,200,498	117,984,223	112,164,891	111,040,728	110,310,839	109,531,994

住民基本台帳の数が 127,443,493 人、人口動態統計に基づく人口は 124,218,814 人であった。これらの統計に差がみられることが分かった。NDB に用意されている ID1 の患者数は 119,208,498 人、ID1N の患者数は 117,984,223 人であり、人口こそ上回っていないが、一部の年齢階級で住民基本台帳

および人口を上回る結果となった(住民基本台帳と比較し、数が過大のものを色で示す)。従来の名寄せ(ID0)では 112,164,891 人であった。ID0v2 の候補である一括マージと個別マージでは、個別マージのほうが名寄せされる患者数が多かった(729,889 人)。住民基本台帳や人口動態に基づく集計は点集計のため、NDB のような期間集計と比較すると脱落者(死亡者など)が過大に集計されている可能性があった。そのため、NDB の転記死亡(病名に付与した死亡のフラグ)を参照し、統計に合わせ 12 月以前に死亡となっている患者は集計から除外したものを「個別マージ(転帰死亡除外)」として示した。多くの年齢階級でほぼ住民基本台帳に収まっており、高齢の区分のみ患者数が過大のところもみられたが、概ね台帳人数に近づいた。性年齢階級で分けた場合の結果を表 2、表 3 に示す。

表 2 年齢階級別の人口・患者数分布(男性)

年齢階級	住基 (H31.1.1)	人口動態統計 H30	id1	id1n	id0	一括マージ	id0v2	個別マージ (転帰死亡除外)
0～4歳	2,521,289	2,439,327	2,693,918	2,651,466	2,502,686	2,472,178	2,453,314	2,451,765
5～9歳	2,732,810	2,621,279	2,843,309	2,801,230	2,690,944	2,650,848	2,632,247	2,631,605
10～14歳	2,823,910	2,732,573	2,751,752	2,715,633	2,635,720	2,601,065	2,588,232	2,587,712
15～19歳	3,041,288	2,979,988	2,587,349	2,559,585	2,490,763	2,462,831	2,454,302	2,453,688
20～24歳	3,296,677	3,063,388	2,502,463	2,480,729	2,365,449	2,338,152	2,326,816	2,326,051
25～29歳	3,310,343	3,011,184	2,574,852	2,555,787	2,422,409	2,396,639	2,383,380	2,382,590
30～34歳	3,658,968	3,407,503	2,861,316	2,837,068	2,712,420	2,685,007	2,671,704	2,670,661
35～39歳	4,002,907	3,804,868	3,134,094	3,106,666	2,989,761	2,960,287	2,946,780	2,945,397
40～44歳	4,656,293	4,537,753	3,597,369	3,563,414	3,447,814	3,412,571	3,398,265	3,395,990
45～49歳	4,983,034	4,820,728	3,878,617	3,839,901	3,722,930	3,683,909	3,668,979	3,665,124
50～54歳	4,272,064	4,150,617	3,472,048	3,433,283	3,324,065	3,286,855	3,272,602	3,267,131
55～59歳	3,832,181	3,779,120	3,274,731	3,230,048	3,125,489	3,083,703	3,069,589	3,061,562
60～64歳	3,749,568	3,714,721	3,446,919	3,417,130	3,153,162	3,133,337	3,106,499	3,092,766
65～69歳	4,439,844	4,506,587	4,137,592	4,053,510	3,844,881	3,784,947	3,757,369	3,729,510
70～74歳	3,881,182	3,850,818	3,743,295	3,700,591	3,612,167	3,584,693	3,568,791	3,527,319
75～79歳	3,098,224	3,089,021	3,547,437	3,535,764	3,019,733	3,031,609	2,990,624	2,933,908
80～84歳	2,147,070	2,187,107	2,159,313	2,154,511	2,148,789	2,146,191	2,143,098	2,070,371
85～89歳	1,202,685	1,222,397	1,302,220	1,299,267	1,294,023	1,292,775	1,290,510	1,210,850
90～94歳	437,923	451,000	526,955	525,936	522,669	522,584	521,460	466,193
95～99歳	77,620	76,000	107,559	107,292	106,658	106,681	106,448	89,227
100歳以上	8,241	9,000	13,033	13,013	12,944	12,945	12,919	9,940
合計	62,174,121	60,454,979	55,156,141	54,581,904	52,145,476	51,649,807	51,363,928	50,969,354

表 3 年齢階級別の人口・患者数分布(女性)

年齢階級	住基 (H31.1.1)	人口動態統計 H30	id1	id1n	id0	一括マージ	id0v2	個別マージ (転帰死亡除外)
0～4歳	2,396,069	2,323,187	2,540,409	2,502,527	2,364,461	2,336,695	2,320,037	2,318,592
5～9歳	2,597,758	2,498,746	2,677,304	2,640,219	2,537,766	2,502,416	2,485,523	2,484,945
10～14歳	2,686,588	2,605,802	2,577,893	2,545,846	2,473,373	2,442,070	2,430,496	2,430,019
15～19歳	2,892,530	2,823,068	2,610,055	2,580,904	2,502,526	2,473,386	2,463,158	2,462,562
20～24歳	3,124,747	2,899,714	3,007,854	2,979,059	2,760,071	2,717,374	2,694,349	2,693,418
25～29歳	3,124,246	2,882,183	3,199,292	3,172,704	2,895,195	2,850,710	2,821,548	2,820,460
30～34歳	3,493,161	3,282,944	3,496,218	3,461,684	3,211,482	3,164,285	3,135,508	3,134,124
35～39歳	3,856,691	3,678,217	3,735,002	3,697,954	3,477,847	3,429,225	3,401,790	3,399,948
40～44歳	4,495,779	4,387,863	4,163,362	4,121,700	3,905,060	3,851,325	3,822,034	3,819,444
45～49歳	4,848,511	4,685,313	4,453,490	4,408,818	4,185,384	4,128,369	4,097,083	4,093,219
50～54歳	4,205,832	4,078,352	3,946,265	3,903,453	3,706,185	3,655,453	3,627,734	3,622,968
55～59歳	3,829,920	3,775,258	3,636,171	3,595,361	3,397,131	3,354,997	3,327,461	3,321,495
60～64歳	3,824,988	3,806,687	3,689,023	3,647,416	3,411,455	3,379,160	3,347,549	3,339,000
65～69歳	4,717,535	4,807,829	4,483,614	4,440,295	4,226,082	4,169,029	4,137,127	4,122,203
70～74歳	4,364,234	4,344,403	4,296,493	4,237,631	4,135,756	4,096,809	4,071,928	4,050,390
75～79歳	3,842,998	3,815,046	4,370,909	4,353,564	3,719,940	3,733,055	3,674,624	3,643,315
80～84歳	3,089,298	3,143,105	3,010,761	3,003,009	2,989,863	2,986,605	2,979,752	2,930,034
85～89歳	2,251,901	2,283,118	2,304,610	2,299,394	2,283,691	2,282,958	2,276,809	2,199,205
90～94歳	1,193,961	1,221,000	1,314,122	1,311,888	1,301,167	1,301,739	1,298,230	1,211,463
95～99歳	371,776	362,000	454,817	454,272	450,833	451,093	450,118	397,264
100歳以上	60,749	60,000	84,693	84,621	84,147	84,168	84,053	68,563
合計	65,269,372	63,763,835	64,052,357	63,402,319	60,019,415	59,390,921	58,946,911	58,562,640

どちらも表 1 の結果と同様の傾向を示していたが、id1 では結婚等で退職する時期である女性は保険が変わるため患者数が人口に収まらないことがあった。しかし id0、id0v2 はどちらも人口の範囲に収まった。また、男女比較で人口に占める id0、id0v2 の患者数は年齢階級に差があることが分かった。これは、死亡時年齢のピークの関係で生じているものと分かった。

年齢階級ごとに名寄せ比(住民基本台帳の人口に占める ID 数)を比較(表 4)すると、id0 と id0v2 では 100 歳以上のようにほとんど変化のない年齢階級もあれば、若年者を中心に名寄せ比が改善している年齢階級もみられた。

表 4 年齢階級別の名寄せ比

年齢階級	住基 (H31.1.1)	人口動態統計 H30	id1	id1n	id0	id0v2		
						一括マージ	個別マージ	個別マージ (転居死亡除外)
0～4歳	1.00	0.97	1.06	1.05	0.99	0.98	0.97	0.97
5～9歳	1.00	0.96	1.04	1.02	0.98	0.97	0.96	0.96
10～14歳	1.00	0.97	0.97	0.95	0.93	0.92	0.91	0.91
15～19歳	1.00	0.98	0.88	0.87	0.84	0.83	0.83	0.83
20～24歳	1.00	0.93	0.86	0.85	0.80	0.79	0.78	0.78
25～29歳	1.00	0.92	0.90	0.89	0.83	0.82	0.81	0.81
30～34歳	1.00	0.94	0.89	0.88	0.83	0.82	0.81	0.81
35～39歳	1.00	0.95	0.87	0.87	0.82	0.81	0.81	0.81
40～44歳	1.00	0.98	0.85	0.84	0.80	0.79	0.79	0.79
45～49歳	1.00	0.97	0.85	0.84	0.80	0.79	0.79	0.79
50～54歳	1.00	0.97	0.88	0.87	0.83	0.82	0.81	0.81
55～59歳	1.00	0.99	0.90	0.89	0.85	0.84	0.83	0.83
60～64歳	1.00	0.99	0.94	0.93	0.87	0.86	0.85	0.85
65～69歳	1.00	1.02	0.94	0.92	0.88	0.87	0.86	0.86
70～74歳	1.00	0.99	0.96	0.96	0.94	0.93	0.93	0.92
75～79歳	1.00	0.99	1.14	1.14	0.97	0.97	0.96	0.95
80～84歳	1.00	1.02	0.99	0.98	0.98	0.98	0.98	0.95
85～89歳	1.00	1.01	1.04	1.04	1.04	1.04	1.03	0.99
90～94歳	1.00	1.02	1.13	1.13	1.12	1.12	1.12	1.03
95～99歳	1.00	0.97	1.25	1.25	1.24	1.24	1.24	1.08
100歳以上	1.00	1.00	1.42	1.42	1.41	1.41	1.41	1.14
合計	1.00	0.97	0.94	0.93	0.88	0.87	0.87	0.86

3.2 id 重複期間による分布の検討

名寄せにおいては id1n をキーとしてグループ化したのち、id2 が同一のレセプトを結合している。その際診療年月を確認し、両グループの診療年月に重複がある場合、第一種過誤を防ぐため除外する必要がある。例えば、表 5 では、診療年月が 2018 年 7 月～2018 年 9 月まで重複しているとみなされる。その際、何か月の重複を許諾するか(のりしろ)について再度検証を行った。その結果を表 6 に示す。

表 5 レセプトの重複の一例

ノード	id1n	id2	診療年月	性別	年齢階級	医療機関 住所
親グループ	AbCdE	FgHiJ	201806	女性	10～14歳	大阪府
	AbCdE	FgHiJ	201809	女性	10～14歳	大阪府
	AbCdE	FgHiJ	201809	女性	10～14歳	大阪府
子グループ	KlMnO	FgHiJ	201807	女性	10～14歳	大阪府
	KlMnO	FgHiJ	201810	女性	10～14歳	大阪府
	KlMnO	FgHiJ	201903	女性	10～14歳	大阪府

表 6 レセプト重複期間別の患者数分布

タイプ	のりしろ					
	1ヶ月	2ヶ月	3ヶ月	4ヶ月	5ヶ月	6ヶ月
ID0			112,164,891			
ID0v2 個別マージ(1か月 を基準としたときの比率)	110,310,839 (100%)	110,294,469 (99.99%)	110,284,835 (99.98%)	110,277,493 (99.97%)	110,271,040 (99.96%)	110,264,967 (99.96%)

例えば、「同じ月の重複のみ許容する」ということであればのりしろは1か月として計算している。重複を許さないルール(1 か月)と重複を 6 か月まで許容するルールでは、患者数が 0.04%ほど少なくなった。これは、全レセプトに占める割合としては少ないものの実際には 5 万人程度の患者に影響することが分かった。

4. 考察

名寄せの検討の前に、ID1 と ID1N について確認すると 120 万人ほど名寄せが改善されていることが見受けられた。これは、表記ゆれを吸収した結果であり、従来の ID1 ではなく今後は ID1N を使うことが重要であろう。

名寄せにおいては教師データがないため、完全な再現は難しい。「第一種過誤を防ぎつつ、できるだけ多くの名寄せを行うにはどのような条件が適切であるか」、に焦点を当てて検証を行った。

日本は国民皆保険のため、日本に住むほぼすべての人が医療を受けることができる。そのため、日本在住の人口(=住民基本台帳)を大きく超えることがなければ名寄せにある程度成功しているものと考えられた。なお、住民基本台帳のほか人口動態統計においても比較検証のために用いたが、過去の文献⁶⁾によると、若年層は住民基本台帳のほうが国勢調査より多く、高齢者層は国勢調査のほうが多いという結果になっ

ている。理由としては、外国人の居住者が一部国勢調査の定義では対象とならないこと、個人情報保護意識の高まり、防犯意識の高まりなどが理由に挙げられている。このように、各種統計を用いるとき、今回の NDB 患者数との比較をする場合は、「住民基本台帳」と比較するほうが有用であるといえる。理由としては、3 か月を超える在留期間が決定されている場合は外国人に対して保険証の交付を行っていること、生活習慣の多様化や防犯意識の高まりから、国勢調査の調査員が面会できる世帯数が大幅に減少している等、である。

NDB 患者数と住民基本台帳の数を比較し、名寄せ比を算出すると、高齢者 90 歳代以上を除き概ね台帳の数の範囲内に収まった。ただし、1 年間で 1 度も病院・診療所を受診しない患者も一定数いると考えられる。しかし、何らかの形で年に 1 度は受診していると考えたほうが理解しやすい。例えば、子供はよく歯科を受診するであろうし、熱があれば小児科に行くと考えられる。また、予防接種等で受診することも考えられるであろう。一方で、生産年齢における分布はそれらの世代に比べて受診をしていない可能性もある。住民基本台帳に比べて数が少ないのはそのためであろう。高齢者では基礎疾患を持っていることが多く何らかの形で受診することがおおいであろう。このような考え方であれば、その分布に違和感はないと考えられる。なお、年齢階層によって名寄せ比にばらつきがみられたのは、ライフイベントによって保険種別が代わりやすい世代の名寄せが行えたためと考えられる。

マージ方法についてどれが適切であるか考えたとき、今回は個別マージが適切ではないかと考えた。理由としては、一度に複数のレセプトを一括で突合すると、順番が異なるため、レセプトの発行時期が重なる「医科・DPC」と「医科・調剤」が分散し、適切にマージできなくなるためであろうと考えられる。特に歯科においてはその可能性が高く、歯科で多くのレセプトがつながらない事例が散見された。このような状況から、レセプトをつなぎ合わせるには順番につなぎ合わせるということが必要であるとわかった。

重複期間(のりしろ)の検証については、どの期間まで重複を許容するかで明確な識別は難しいであろう。しかし、これまでの知見において死亡後 2 か月以内にレセプトが発生することがあるとわかっている。前述の通り、薬剤感受性試験は翌月請求となる可能性があるほか、死亡した場合だけでなく保険者が変わったときも、本人が来院しなければ、前月の保険者に請求が行く可能性が高い。レセプトの中には、さまざまな医療機関を受診するが、その際の ID1N の変更がばらつく(一定の月ではない事例)があった。医療機器の作成について診療報酬を請求できるのは完成して患者に提供した時点となる。インプラントのような体内人工物については、請求したのちに死亡することがあれば、その請求は死亡後に回る可能性がある。したがって少なくとも 3 か月ののりしろは必要と考え、従来の ID0 でのりしろを 3 か月とした。

6 か月まで延長した際に、名寄せできる患者数は軽微である。名寄せされた患者の実際のレセプトを確認してどのような患者が突合されたのか確認すると、3 か月までは明らかに説明がつく患者(同じ都道府県内である場合や、別の都道府県に異動になる場合でも生産年齢であったり、定年退職の世代であったり)が多かった。確認したレセプトにおいては疑わしい事例は確認されなかった。しかし、6 か月を対象としたレセプトは一部ではあるが、その結合に違和感がある状況が確認された(年齢階級が高齢者、複数の都道府県をまたいでいる、など)。第一種過誤を防ぐことが名寄せにおいて重要なことであるため、このような名寄せを行っていいのか判断に迷う事例が発見された場合、接続しない方が適切ではないかと考えた。そのため、ID0 のルールと同様に ID0v2 でものりしろは 3 か月を採用することとした。ただし、この 3 か月が 4 か月でも適切かどうかは個票を確認しても判断が難しい。限界では

あるが、名寄せ比の向上に貢献する適切な範囲は今後も検討が必要であろう。

5. 結論

ID1N での名寄せ比の向上や、レセプト重複期間の見直し、つなぐ際に最も近い診療年月から結合するなどの名寄せ方法を取ることによって、以前に比べて名寄せ比は向上した。今回は単年度における患者数の集計であったが、今後は多年度における名寄せの追跡率を算出し、名寄せを継続しなければ脱落する多くのレセプトの状況を把握することによって、その精度の確認を行っていく必要がある。

謝辞

本研究は厚生労働科学研究費補助金(地域医療基盤開発推進研究事業)「地域の実情に応じた医療提供体制の構築を推進するための政策研究(H30-医療-一般-013)」、科学研究費助成事業(科学研究費補助金)基盤研究(A)(一般)「データ科学・疫学・臨床医学の融合による日本の保険診療情報(NDB)の全解析(18H04126)」、科学研究費助成事業(科学研究費補助金)基盤研究(A)(一般)「疫学と臨床医学、情報工学の融合によるデータベース医学の標準的疾患定義の開発(20H00623)」、科学研究費助成事業(学術研究助成基金助成金)若手研究「NDB(レセプトデータ)を用いた難病患者の疫学調査の手法開発と治療実態の把握(19K19464)」の一環として実施したものである。

参考文献

- 1) 第三者提供の現状について(報告)(承諾案件集計), 2020.
[<https://www.mhlw.go.jp/content/12400000/000641055.pdf>]
(cited 2020-August-11)].
- 2) 久保 慎一郎, 野田 龍也, 明神 大也, 加藤 源太, 今村 知明. NDB(ナショナルデータベース)の課題および留意点と今後の展望. 医療情報学連合大会論文集, 36(1): 272-275, 2016.
- 3) 久保 慎一郎, 野田 龍也, 明神 大也, 東野 恒之, 松居 宏樹, 加藤 源太, 今村 知明. 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 2017;38:11-18.
- 4) 野田龍也、久保慎一郎、明神大也、西岡祐一、東野恒之、松居宏樹、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. 厚生指標. 64(12), 7-13, 2017-10.
- 5) Kubo S, Noda T, Myojin T, et al. National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB): outline and patient-matching technique. bioRxiv.
[<https://www.biorxiv.org/content/early/2018/04/02/280008.full.pdf>. (cited 2019-August-27)].
- 6) 菅桂太. 近年の「国勢調査」日本人人口の精度に関する一考察 人口学研究. 61, 61-73, 2007-11.