

# 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

ノダ タツヤ      クボ シンイチロウ      ミヨウジン トモヤ      ニシオカ ユウイチ  
野田 龍也\*1      久保 慎一郎\*3      明神 大也\*4      西岡 祐一\*4\*5  
ヒガシノ ツネユキ      マツイ ヒロキ      カトウ ゲンタ      イمامラ トモアキ  
東野 恒之\*6      松居 宏樹\*7      加藤 源太\*8      今村 知明\*2

**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、わが国における保険診療の悉皆データである。NDBには、同一患者の複数レセプトを紐付ける変数として、「ID1」と「ID2」が用意されているが、就職・転職や氏名の表記ゆれ等で容易に変わり得る。本研究は、NDBにおける現行の名寄せ手法を改善し、種々の工夫により名寄せの効率を高めた新たな個人ID（ID0）を提案するとともに、その妥当性を検証することを目的としている。

**方法** 平成25年4月～平成26年3月の医科入院レセプト、医科入院外レセプト、DPCレセプト、調剤レセプト全体を対象とし、ID1、ID2、診療年月、転帰区分を利用した新しい名寄せアルゴリズムを開発した。アルゴリズムの妥当性を検証するため、従来のID1による名寄せとID0による名寄せを、平成25年10月1日時点の推計人口と比較した。

**結果** 男女とも、ID0による性年齢階級別患者数はID1による患者数を下回っており、ID1による患者数を男性で6.2%、女性で7.1%、圧縮することができた。ID1により名寄せされた患者数は、小児や高齢者、若年女性において推計人口を大きく上回っていたが、ID0により名寄せされた患者数はおおむね推計人口の範囲内に収まった。

**結論** 本研究で提案した新たな名寄せID（ID0）は、従来の名寄せで用いられてきたID1や、ID1とID2の併用に比し、名寄せの精度が向上した。現状ではID0が最善の名寄せ手法であり、今後、患者数の推計にはID1ではなくID0を用いることが望ましい。

**キーワード** NDB、レセプト、診療報酬データ、名寄せ、突合、データベース

## I はじめに

レセプト情報・特定健診等情報データベース（以下、NDB）とは、診療報酬請求のために病院等から審査支払機関と保険者に送信される電子レセプトデータおよび保険者の保有する特定健診等のデータを、匿名化処理を経て格納・構築したデータベースであり、国民皆保険制度をとるわが国における保険診療の悉皆データである。NDBは平成21年4月～平成28年12月診

療分で約128億8400万件（平成29年3月末時点）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であり、これを有効に活用することで各種の臨床研究、政策研究が強力に推進できると期待されている。

しかし現状は、疾患別、地域別等の部分的な活用に留まっており、成果が十分に上がっているとは言いがたい。その理由として、著者らが行った先行研究では、NDBの巨大なサイズと並び、診療報酬請求のために設定されているレ

\*1 奈良県立医科大学公衆衛生学講座講師      \*2 同教授      \*3 同修士課程      \*4 同博士課程  
\*5 奈良県立医科大学糖尿病学講座医員      \*6 (株)三菱総合研究所ICTイノベーション事業本部主席研究員  
\*7 東京大学大学院医学系研究科公共健康医学専攻臨床疫学・経済学助教  
\*8 京都大学医学部附属病院診療報酬センター准教授

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が主因として挙げられた<sup>1)</sup>。その中でも、後述する「名寄せ」は大きな問題点である。

各医療機関は、患者ごとに、毎月、診療報酬を請求するので、レセプトデータは、1人の患者に対し、医療機関単位、1カ月単位で送信されている。患者が複数月に渡って受診したり、同一月に複数の医療機関を受診したりすることは頻繁にあるため、同一患者の複数レセプトをつなぎ合わせる「名寄せ」作業を行わなければ、個人単位での分析を行うことはできない。そのため、NDBには、名寄せを可能とする個人紐付け用の匿名変数として、「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わかっている<sup>2)</sup>。そのため、就職・転職等の前と後で、1人の患者による受診を異なる2人の患者の受診と認識する等の可能性が生じ、NDBを用いた患者数の推計や患者1人あたりの推計値は大きな誤差を含むと考えられる。この問題を解決するためには、NDBに一生涯不変の個人IDを付与する必要があるが、その実現に向けては議論が緒に就いたばかりである。また、不変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。

本研究の目的は、NDBにおける現行の名寄せ手法を改善し、種々の工夫により名寄せの効率を高めた新たな個人ID (ID0) を提案するとともに、その妥当性を検証することである。

## Ⅱ 方 法

本研究では、平成25年4月～平成26年3月の

計12カ月分の医科入院レセプト、医科入院外レセプト、DPCレセプト、調剤レセプト全体を対象とした。DPCとは、平成15年に導入された診断群分類に基づく入院医療費の1日あたり包括払い制度であり、DPC対象の入院で発生するレセプトの内、包括支払いに関連するレセプトがDPCレセプトとなる（包括支払いに関連のない入院レセプトは医科入院レセプトとなる）。これらのレセプトデータを使用して、複数月に渡るID等の変化を観察し、ID1、ID2のほか、診療年月、転帰区分を利用した新しい名寄せアルゴリズムを作成した。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、DPCおよび医科（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。本研究で開発した新規の名寄せアルゴリズムの詳細を下記に示す。

### (1) 医科（入院／入院外）、DPCレセプトで名寄せ用中間データを作成

「医科入院レセプト」「医科入院外レセプト」「DPC入院の総括対象医科入院レセプト」「DPC入院のDPCレセプト」「DPC入院の総括対象DPCレセプト」からID1、ID2、診療年月、転帰区分を抽出する。転帰区分については死亡情報を抽出した。まず、複数月に渡ってID1が同じ患者は同一人物とした。死亡転帰が示されている場合は、その月で名寄せ終了とした。

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た。

なお、資格喪失したレセプトについては審査支払機関を経由し保険者でチェックを行い再審

査請求となるため、3カ月間程度は新旧のID1が併存しうる。この対応として、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った。次に前月の1カ月間で同じID2を有するレセプトを探索し名寄せを行った。最後に前々月の1カ月間で同じID2を有するレセプトを探索し名寄せを行った。以上を組み合わせて、名寄せ中間データ（医科・DPC）テーブルを作成した。

(2) 調剤レセプトで名寄せ用中間データを作成  
「調剤レセプト」からID1, ID2, 診療年月を抽出し、名寄せ中間データ（調剤）テーブルを作成した。複数月に渡ってID1が同じ患者を同一人物とし、ID1が途切れた前後では、(1)と同様の手法でID2を用いて名寄せを行った。

(3) 医科（入院／入院外）、DPCレセプトと調剤レセプトを連結

医科・DPCの名寄せ用中間テーブルと調剤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：Ajika783：33yud0fu）。ここで、便宜上、対応セットの左のID1（例：Ajika783）を「先出ID1」、右のID1（例：33yud0fu）を「後出ID1」と呼ぶ。「先出ID1」は、医科、DPC、調剤を問わず一番最初に出現したID1、「後出ID1」は二番目以降に出てきたID1である。先出と後出のID1は英数字は異なるが同一人物と考えられる。なお、「ID1：Pon89jss：w57Kurtz」と「ID1：w57Kurtz：Pon89jss」のように循環参照が発生する場合は対応セットの一方を削除した。

(4) 名寄せ作業

(3)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1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す。

返し、最終的に残ったID1を新しい名寄せ変数「ID0」とした。なお、このID0には、単回受診等で名寄せ対象とならなかったため元々のID1がそのまま残存したものも含まれている。

さらに、新しく作成された名寄せ変数ID0の妥当性を検証するため、ID0と従来のID1により名寄せした性年齢階級別の患者数を比較するとともに、総務省統計局による平成25年10月1日時点の推計人口とも比較し、大まかな性年齢階級別受診率を算出した。

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」（平成26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（平成27年10月8日受付番号1123）を得た。

### Ⅲ 結 果

(1) NDBにおけるID1, ID2の特性に関する検討

観察期間におけるNDBのID変化は、①ID1とID2が変化しない、②ID1, ID2のいずれかが変化する、③ID1とID2の両方が変化するのいずれかである。①ではID1とID2が一対一対応するが、②と③では一方のIDに対し他方のIDが複数個発生する（例えば、氏名表記が変わらず、転職等で医療保険が1回変わると、1つのID2に対し2つのID1が発生する）。今回の観察期間（12カ月間）における、一方のIDに対する他方のIDの重複割合を表1に示す。DPCレセプトや医科入院レセプトでは、ID1-ID2の一対一対応率は98%以上と高かった。しかし、医科入院外レセプト、調剤レセプトでは、それぞれ79.6%、92.3%と相対的に低かった。なお、ID1-ID2の一対一対応には、観察期間中に1回しか受診しなかった場合や、複数回受診しつつ③が発生した（両IDが変化して別人の初回受診のようにみえる）場合も含まれる。

医科入院外において、ID1-ID2の一対一対応率はID2に対するID1の重複でみると92.9%であるが、ID1に対するID2の重複では79.6%と著しく低下する。これは複数医療機関を受診し

表1 ID1とID2の重複

DPC

ID1に対するID2の重複			ID2に対するID1の重複		
	件数	構成比 (%)		件数	構成比 (%)
合計	6 647 932	100.00		6 670 937	100.00
1	6 551 879	98.60	1	6 598 453	98.90
2	95 502	1.40	2	71 336	1.10
3	500	0.00	3	1 125	0.00
4	48	0.00	4	23	0.00
5	xxx	0.00			
6	xxx	0.00			

医科入院

ID1に対するID2の重複			ID2に対するID1の重複		
	件数	構成比 (%)		件数	構成比 (%)
合計	5 713 500	100.00		5 743 180	100.00
1	5 600 307	98.00	1	5 661 120	98.60
2	112 202	2.00	2	79 723	1.40
3	974	0.00	3	2 221	0.00
4	17	0.00	4	+++	0.00
			5	xxx	0.00

医科入院外

ID1に対するID2の重複			ID2に対するID1の重複		
	件数	構成比 (%)		件数	構成比 (%)
合計	117 048 101	100.00		132 050 478	100.00
1	93 145 865	79.60	1	122 674 241	92.90
2	22 651 589	19.40	2	8 568 817	6.50
3	1 141 050	1.00	3	726 050	0.50
4	96 966	0.10	4	71 728	0.10
5	9 678	0.00	5	8 055	0.00
6	2 192	0.00	6	1 250	0.00
7	527	0.00	7	222	0.00
8	155	0.00	8	60	0.00
9	54	0.00	9	29	0.00
10	16	0.00	10	15	0.00
11	xxx	0.00	11	xxx	0.00
12	xxx	0.00	12	xxx	0.00
14	xxx	0.00			

調剤

ID1に対するID2の重複			ID2に対するID1の重複		
	件数	構成比 (%)		件数	構成比 (%)
合計	94 937 633	100.00		95 851 172	100.00
1	87 623 580	92.30	1	89 656 685	93.50
2	7 006 675	7.40	2	5 707 316	6.00
3	282 877	0.30	3	439 799	0.50
4	21 838	0.00	4	41 828	0.00
5	2 073	0.00	5	4 623	0.00
6	474	0.00	6	714	0.00
7	87	0.00	7	140	0.00
8	16	0.00	8	42	0.00
9	xxx	0.00	9	14	0.00
10	xxx	0.00	10	xxx	0.00
14	xxx	0.00	11	xxx	0.00
			12	xxx	0.00

注 xxxは10未満のため非公表, +++は10以上マスキング

実際に氏名の表記ゆれが多く発生するためであり、IDを用いた名寄せの品質に影響を及ぼす事象である。表2にID2の数と受診医療機関数の関係を示す。1年間の観察期間では、ID1に対するID2の数の標準的な重複数は2以内で98.9%を占め、おおむね3以内に収まっていた。1年間における平均受診医療機関数は2.5であった。

(2) ID0とID1の名寄せ精度の比較

平成25年度の1年分の患者（DPC、医科入院、医科入院外、調剤）を対象に、ID0にて名寄せを実施し、従来のID1による名寄せ患者数および平成25年10月の推計人口と比較した性年齢階級別の結果を表3と表4に示す。男女とも、ID0による性年齢階級別患者数はID1による患者数を下回っており、追加名寄せ率（ID1名寄せに比してID0名寄せで同一人物の特定に追加的に成功した割合）は男性で6.2%、女性で7.1%であった。ID1により名寄せされた患者数は、0

表2 医科入院外におけるID2重複数と平均受診医療機関数

	ID1に対するID2の重複		
	件数	累積割合 (%)	平均受診医療機関数
合計	117 048 101	100.0	2.5
1	93 145 865	79.6	
2	22 651 589	98.9	
3	1 141 050	99.9	
4	96 966	100.0	
5	9 678	100.0	
6	2 192	100.0	
7	527	100.0	
8	155	100.0	
9	54	100.0	
10	16	100.0	
11	xxx	100.0	
12	xxx	100.0	
13	xxx	100.0	
14	xxx	100.0	

注 xxxは10未満のため非公表

～9歳や75～79歳、90歳以上、男性の85～89歳、女性の25～29歳において推計人口を大きく上回っていた。一方、ID0により名寄せされた患者数は、男性の85歳以上や女性の90歳以上で推計人口を大きく上回ったが、それ以外の性年齢階級ではおおむね推計人口の範囲内に収まった。

### IV 考 察

本研究で提案した名寄せ手法「ID0」は、NDBにおけるID1を名寄せの基軸とし、ID2や転帰区分を補完的に用いて、同一人物の複数のID1を1つのID1へ統合する手法である。この名寄せ手法は、ID1とID2を併用するのみならず、転帰区分および3カ月間の名寄せ期間を考慮し

た点に大きな新規性がある。ID2では同姓同名・同一生年月日・同性患者が同一IDになってしまうが、ID1では（同じ扶養に入っている同性の双子等を除き）別人が同一IDとなる頻度が比較的少ないため、多くのNDB集計ではID1のみを用いて患者数を集計している。しかし、この手法は集計期間中のID1の変化に対応できない。また、ID1とID2を併用して名寄せを行う場合でも、死亡の転帰区分が発生しているのに別人物を名寄せしてしまったり、資格喪失による誤請求に起因する

表3 年齢階級別ID0患者数、ID1患者数および推計人口（男性）

	ID1患者数	ID0患者数	ID1-ID0	追加名寄せ率 (%)	推計人口 (平成25年)	年齢階級受診率 (ID1) (%)	年齢階級受診率 (ID0) (%)
合計	55 115 634	51 676 633	3 439 001	6.2	61 906 000	89.0	83.5
0～4歳	2 969 185	2 741 268	227 917	7.7	2 684 000	110.6	102.1
5～9	2 981 037	2 787 916	193 121	6.5	2 743 000	108.7	101.6
10～14	2 920 044	2 769 487	150 557	5.2	2 967 000	98.4	93.3
15～19	2 598 051	2 476 684	121 367	4.7	3 098 000	83.9	79.9
20～24	2 418 930	2 264 250	154 680	6.4	3 181 000	76.0	71.2
25～29	2 774 770	2 592 807	181 963	6.6	3 506 000	79.1	74.0
30～34	3 084 618	2 905 099	179 519	5.8	3 867 000	79.8	75.1
35～39	3 563 811	3 383 134	180 677	5.1	4 592 000	77.6	73.7
40～44	3 799 831	3 626 320	173 511	4.6	4 888 000	77.7	74.2
45～49	3 338 973	3 187 276	151 697	4.5	4 228 000	79.0	75.4
50～54	3 134 569	2 985 704	148 865	4.7	3 870 000	81.0	77.1
55～59	3 212 148	3 047 773	164 375	5.1	3 840 000	83.6	79.4
60～64	4 217 893	3 797 207	420 686	10.0	4 740 000	89.0	80.1
65～69	4 199 556	3 744 118	455 438	10.8	4 183 000	100.4	89.5
70～74	3 426 927	3 319 631	107 296	3.1	3 537 000	96.9	93.9
75～79	3 052 168	2 648 482	403 686	13.2	2 772 000	110.1	95.5
80～84	1 920 494	1 908 896	11 598	0.6	1 889 000	101.7	101.1
85～89	1 064 040	1 056 041	7 999	0.8	970 000	109.7	108.9
90～94	347 669	344 445	3 224	0.9	283 000	122.9	121.7
95～99	79 586	78 848	738	0.9	61 000	130.5	129.3
100歳以上	11 334	11 247	87	0.8	7 000	161.9	160.7

表4 年齢階級別ID0患者数、ID1患者数および推計人口（女性）

	ID1患者数	ID0患者数	ID1-ID0	追加名寄せ率 (%)	推計人口 (平成25年)	年齢階級受診率 (ID1) (%)	年齢階級受診率 (ID0) (%)
合計	64 753 273	60 138 764	4 614 509	7.1	65 390 000	99.0	92.0
0～4歳	2 812 626	2 603 322	209 304	7.4	2 554 000	110.1	101.9
5～9	2 818 198	2 639 709	178 489	6.3	2 618 000	107.6	100.8
10～14	2 721 514	2 584 830	136 684	5.0	2 823 000	96.4	91.6
15～19	2 659 334	2 522 597	136 737	5.1	2 949 000	90.2	85.5
20～24	3 042 012	2 746 277	295 735	9.7	3 023 000	100.6	90.8
25～29	3 606 108	3 225 659	380 449	10.6	3 365 000	107.2	95.9
30～34	3 916 328	3 553 636	362 692	9.3	3 757 000	104.2	94.6
35～39	4 295 486	3 964 045	331 441	7.7	4 467 000	96.2	88.7
40～44	4 440 369	4 140 521	299 848	6.8	4 779 000	92.9	86.6
45～49	3 847 923	3 597 548	250 375	6.5	4 177 000	92.1	86.1
50～54	3 578 405	3 343 092	235 313	6.6	3 863 000	92.6	86.5
55～59	3 648 627	3 372 150	276 477	7.6	3 892 000	93.7	86.6
60～64	4 658 286	4 208 224	450 062	9.7	4 925 000	94.6	85.4
65～69	4 541 557	4 181 569	359 988	7.9	4 517 000	100.5	92.6
70～74	4 041 274	3 895 613	145 661	3.6	4 060 000	99.5	96.0
75～79	3 852 981	3 347 160	505 821	13.1	3 529 000	109.2	94.8
80～84	2 837 376	2 813 254	24 122	0.9	2 874 000	98.7	97.9
85～89	2 008 486	1 987 802	20 684	1.0	1 955 000	102.7	101.7
90～94	1 031 102	1 020 070	11 032	1.1	933 000	110.5	109.3
95～99	327 251	324 145	3 106	0.9	282 000	116.0	114.9
100歳以上	68 030	67 541	489	0.7	48 000	141.7	140.7

失による誤請求に起因する期間ずれに対応できずに同一人物を名寄せし損ねたりといった問題があることがわかっている。本研究のID0は、これらの課題を克服した新しい名寄せ手法である。レセプト発生月の遅れについては、ID1が変化した場合の3カ月間のID2探索を一度に行わず、1カ月単位で繰り返すなど、名寄せの精度をできるかぎり高める工夫を行っている。なお、レセプトはID1、ID2の変更に伴い同月に2枚発行されることがあり、ID1が変更された場合でも、変更のないID2が収載されたレセプトが発行されることがあるため、入院している場合や継続的に通院している場合ではID1とID2が同時に変わるような場合（結婚退職や扶養の外れる離婚など）でも継続的な受診がある場合は名寄せが可能となることがわかった。

これらの工夫の結果、

ID1のみを用いた名寄せに比し、男性で6.2%、女性で7.1%の追加的な名寄せ（過剰な集計の圧縮）に成功した。ID1による名寄せでは、推計人口を超える実受診者数（延べではない）が観測される年齢階級があるなど、名寄せの精度に大きな疑義が生じていたが、ID0では実態に近い数値へと近づいた。一方、男女の0～9歳や男性の80歳以上、女性の85歳以上では受診率が100%を超えており、依然、名寄せが十全ではないことを示している。また、就職・転職・退職で保険者が変わりやすい年度末はID1が変化しやすいが、本研究は、対象期間が年度内に留まっている点も留意が必要である。NDBにおける名寄せ問題の解決には、各患者に一生涯不変のIDを割り当てるほかに、前述したようにその検討は緒に就いたばかりである。

医科入院レセプトやDPCレセプトにおけるID1-ID2重複度は医科入院外レセプトよりも高かった。これは、外来（医科入院外）に比べ、入院（医科・DPC）では1年間で複数の医療機関を受診することが少ないため、氏名の表記ゆれが起りにくく、同一人物に異なるID2が発生する可能性が低いことが影響していると考えられる。

一方、医科入院外レセプトにおけるID1に対するID2の重複数をみると、平均受診医療機関数は2.5であり、それと関連するID2重複数は、おおむね3以内であった。これは、わが国における2013年度の保険診療受診者が1年間あたり約2.5の医療機関を外来受診しているが、この期間における氏名変更や氏名の表記ゆれを考慮しても、同一人物において異なるID2が発生する場合、その回数は2回以内が標準であることを意味している。NDBにおいても、1人で数多くの医療機関を外来受診している患者は少数であり、医療機関ごとの氏名の表記ゆれもそれほど神経質に対応する必要はないと考えられる。しかし、単純にID1を用いた場合に比し、ID0では6ないし7%という大きな名寄せ精度の改善が認められているため、ID1が不正確な名寄せ手法であることには変わりがない。今後のNDB分析ではID1単独の名寄せは用いないこと

が望ましい。

なお、今回1年分のNDBデータを用いて名寄せを行ったが、名寄せする期間が長くなればなるほど、IDの紐付けは変わる可能性がある。理由は、期間の最初月と最終月に発生したレセプトはその前後のレセプトが不足するため、名寄せをする際に必要な情報が不足することが想定されるためである。よって、NDBデータが追加されるたびに名寄せを随時再構成することが望ましいといえる。

ID0による名寄せで受診率が100%を超えている年齢階級は、名寄せが不十分であることが明らかであるが、不十分となる理由は今のところはっきりとわかっていない。総務省の推計人口は直近の国勢調査をもとに算出された推計値であり、先行研究によると、国勢調査は住民基本台帳人口と1%ほど（100人以上）差異が出ることがわかっている<sup>3)</sup>。また、本研究に用いた推計人口は2013年度の年央人口であり、1年間を通じて名寄せされた人口ではない。推計人口自体の誤差の問題はあると思われるが、ID0側にも何らかの問題があると考えられ、今後の課題である。一方、ID1による名寄せでは、75～79歳においてその年齢階級での性別人口を大きく上回っており、ID0よりも13%ほど過剰に患者数を推計してしまっている。これは、75歳における後期高齢者医療制度への切り替えで変化したID1が「別人」として認識されているためである。この点でも、ID1のみを用いた患者数推計は望ましくない。

ID0による名寄せにおいても解決途上の課題として、性差がある。20～39歳の男性では年齢階級受診率が70%台前半であり、すべての年齢階級を通じて最も低い。青壮年期の受診率が低いことは当然であり、この結果は妥当である。ところが、20～39歳の女性では年齢階級受診率が90%前後に達しており、前後の年齢階級よりも高くなっている。今のところ、この理由として2つが考えられる。1つは出産に関連する医療保険の利用である。出産は原則として自費診療であり、NDBには登場しないが、異常分娩では周産期に医療保険が適用されることがある。

例えば、鉗子分娩や吸引分娩、帝王切開では医療行為の一部が医療保険の適応となるため、年齢階級受診率を押し上げている可能性がある。2つ目の理由として、この年代の女性で氏名や保険者が変化しやすいことが挙げられる。婚姻の際に「夫の氏」を選択する割合は2013年で96.2%であり、多くの女性は婚姻の際にID2が変更となる<sup>4)</sup>。また、婚姻に伴い夫の扶養に入るなどの理由で医療保険が変更となった場合はID1も変更となる。氏の変更や保険者の変更は男女ともに発生しうが、特に20代、30代の女性には多く発生するため、ID1とID2の両方が変更になる前後で複数回の受診を行うと、同一人物であるにも関わらず別人物としてカウントされる可能性が高くなっている。

ID0による名寄せの精度をさらに上げるためには、受診地や病名などの連続性を用いる方法が考えられる。同じIDであったとしても、遠く離れた土地で全く異なる病名で受診した患者は別人物と考えるのが妥当であるし、異なるIDであったとしても、受診地や年齢階級、病名、治療内容が一致する患者は同一人物である可能性が高まる。しかし、NDBは巨大であるため、出張等で遠隔地での受診を繰り返す患者が一定数存在し、治療内容が一致しても別人であることも多々ある。ある限定された状況においては、受診地や年齢階級、病名などを利用した精緻な名寄せが可能であると考えられるが、別人を同一人物として名寄せしてしまう第一種過誤が発生しやすくなるため、受診地等を利用した名寄せは全国規模では課題が大きい。なお、ID0による名寄せは、第一種過誤を発生させないことを原則の1つとしており、同一人物の可能性が高いレセプトであっても、IDによる照合が一致しない限りは別人として扱っている。

## IV 結 語

本研究では、NDBにおける現行の名寄せ手法を改善し、種々の工夫により名寄せの効率を高めた新たな個人ID (ID0) を提案するとともに、その妥当性を検証した。従来の名寄せで用いられてきたID1や、ID1とID2の相互利用に比し、ID0では名寄せの精度が向上した。ただし、小児や後期高齢者、再生産年齢の女性では過剰な集計が残存している。現状ではID0が最善の名寄せ手法であり、今後、患者数の推計にはID1ではなくID0を用いることが望ましい。

## 謝辞

本研究は、平成28年度国立研究開発法人日本医療研究開発機構 (AMED) 「地域横断的ICT活用医療推進研究事業 レセプト等の大規模電子診療情報を活用した薬剤疫学研究を含む医療パフォーマンス評価に関する研究」(研究開発代表者 今村知明) の支援で実施された。

## 文 献

- 1) 久保慎一郎, 野田龍也, 明神大也, 他. 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 2016; 36 (1): 272-5.
- 2) 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 2017; 38: 11-8.
- 3) 山田茂. 2005年国勢調査結果の精度の検討. 中央調査報 2008.
- 4) 厚生労働省. 平成28年度 人口動態統計特殊報告 「婚姻に関する統計」の概況. 2017.