

一般口演

## 一般口演13 データ分析システム

2017年11月21日(火) 16:15 ~ 18:00 G会場 (10F 会議室1006-1007)

### [2-G-3-OP13-5] レセプト情報・特定健診等情報データベース（NDB）利用促進に向けた取り組み—患者突合(名寄せ)の手法開発と検証—

久保 慎一郎<sup>1</sup>, 野田 龍也<sup>1</sup>, 西岡 祐一<sup>1</sup>, 明神 大也<sup>1</sup>, 東野 恒之<sup>2</sup>, 松居 宏樹<sup>3</sup>, 加藤 源太<sup>4</sup>, 今村 知明<sup>1</sup> (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 ICT イノベーション事業本部, 3.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, 4.京都大学医学部附属病院 診療報酬センター)

【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、日本の保険診療受診者の悉皆調査であるが、個人IDが存在せず、保険者番号等を基にハッシュ化した「ID1」と、氏名等を基にハッシュ化した「ID2」の2つを組み合わせて名寄せを行わなければ、個人を追跡できない課題があった。本研究では名寄せにおける2種類の過誤を低減し、新たな個人IDとなる「ID0」を開発し、検証を行った。

【方法】名寄せの検証には平成25年4月～平成26年3月の計12か月分の医科入院、医科入院外、DPC、調剤レセプトを使用した。これらを基に、ID1、ID2の特性を確認した。また、ID1、ID2、診療年月、転記を使用して名寄せ「ID0」を作成した。ID0を基にして、患者数集計を行い、国勢調査の人口と比較し、その妥当性を検証した。

【結果】NDBの特性として1患者に複数のID1、ID2が存在する。1つのID1に対するID2の重複をみると、医科入院外と調剤レセプトで多くの重複が発生していた。また、受診した医療機関数ごとに同一ID1におけるID2数の平均を比較すると、受診する医療機関数が増えてもID2は概ね2程度に収束しており、ID2の変更理由である「患者氏名の表記ゆれ」は1人当たり2種類程度となることが分かった。また、従来名寄せが困難であったID1とID2が同時に代わる場合であっても、同年月で2枚のレセプトが発行されることがあるため、継続的な受診がある場合は名寄せが可能と分かった。従来のID1による名寄せでは、名寄せ後の人口が総人口を上回る欠点があったが、ID0による患者数はほぼ総人口の範囲内に収まり、名寄せ精度が改善されたことが示された。新生児から小児期と高齢者世代では、1年間にほぼ1回以上受診していた。名寄せ率（ID1数÷ID0数）を求めると、退職期（60～69歳）や後期高齢者医療制度への転換期（75～79歳）で名寄せ率向上を認めた。

【結論】ID0によって名寄せの精度が向上した。ID0を用いてNDBのコホート研究の利活用が求められる。

## レセプト情報・特定健診等情報データベース(NDB)利用促進に向けた取り組み

### - 患者突合(名寄せ)の手法開発と検証 -

久保慎一郎<sup>\*1</sup>、野田龍也<sup>\*1</sup>、西岡祐一<sup>\*1\*2</sup>、明神大也<sup>\*1</sup>、  
東野恒之<sup>\*3</sup>、松居宏樹<sup>\*4</sup>、加藤源太<sup>\*5</sup>、今村知明<sup>\*1</sup>

\*1 奈良県立医科大学公衆衛生学講座、

\*2 奈良県立医科大学糖尿病学講座、

\*3 (株)三菱総合研究所 ICT イノベーション事業本部、

\*4 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、

\*5 京都大学医学部附属病院診療報酬センター

## Promoting More Effective Use of the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB): Development & Validation of a New Patient-Matching Technique

Kubo Shinichiro <sup>\*1</sup>, Noda Tatsuya <sup>\*1</sup>, Myojin Tomoya <sup>\*1</sup>, Nishioka Yuichi <sup>\*1\*2</sup>, Higashino Tsuneyuki <sup>\*3</sup>,  
Matsui Hiroki <sup>\*4</sup>, Kato Genta <sup>\*5</sup>, Imamura Tomoaki <sup>\*1</sup>

\*1 Department of Public Health, Health Management and Policy, Nara Medical University ,

\*2 Department of Diabetology, Nara Medical University,

\*3 ICT Innovation Division, Mitsubishi Research Institute, Inc. ,

\*4 Department of Clinical Epidemiology and Health Economics, School of Public Health, The University of Tokyo,

\*5 Solutions Center for Health Insurance Claims, Kyoto University Hospital.

**Background:** The National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB) is a comprehensive database of health insurance claims data under Japan's National Health Insurance system. The NDB uses two types of personal identification variables (referred to in the database as “ID1” and “ID2”) to link the insurance claims of individual patients. However, the information entered against these ID variables is prone to change for a number of reasons, such as when claimants find or change employment, or due to variations in the spelling of their name. In the present study, we developed a new patient-matching technique that improves upon the existing system of using ID1 and ID2 variables. We also sought to validate a new personal ID variable (ID0) that we proposed in order to enhance the efficiency of patient matching in the NDB database.

**Methods:** Our study targeted data from health insurance claims filed between April 2013 and March 2016 for medical hospitalization, diagnosis procedure combinations, medical outpatient treatment, and dispensing of prescription medication. We developed a new patient-matching algorithm based on the ID1 and ID2 variables, as well as variables for treatment date and clinical outcome. We then attempted to validate our algorithm by comparing the number of patients identified by patient matching with the current ID1 variable and our proposed ID0 variable against the estimated patient population as of 1 October 2015.

**Results:** The numbers of patients in each sex and age group that were identified with the ID0 variable were lower than those identified using the ID1 variable. By using the ID0 variable, we were able to reduce the number of male and female patient matches by 5.8% and 6.4%, respectively. The numbers of infant, elderly, and young female patients identified using the ID1 patient-matching variable were all higher than their corresponding estimates. Conversely, the numbers of these patients identified with the ID0 patient-matching variable were all within their corresponding estimates.

**Conclusion:** Our study findings show that the proposed ID0 variable delivers more precise patient-matching results than the existing ID1 variable. The ID0 variable is currently the best available technique for patient matching in the NDB database. Future patient population estimates should therefore rely on the ID0 variable instead of the ID1 variable.

**Keywords:** health insurance claims in Japan, patient identification, personal identifiers

### 1. 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、病院等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～2016年12月診療分で約128億8,400万件(2017年3月末時点)<sup>1)</sup>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なっている。これ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。

しかし現状は、政策用途など部分的な活用に留まっており、

成果が十分に蓄積されているとまでは言いがたい。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>2)</sup>。その中でも、後述する「名寄せ」は大きな問題点である。

各医療機関のレセプトデータは、患者ごとに、毎月、診療報酬を請求するので、一人の患者に対し、医療機関単位、1ヶ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、追跡が必要な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分かっている<sup>3)</sup>。そのため、就職・転職等の前と後で、一人の患者による受診を異なる二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一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 に一生不変の個人 ID を付与する必要があるが、実現されるのはまだ先の、医療等 ID が本格導入される予定である 2020 年以降となる可能性が高い。また、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。なお、NDB の患者突合性を高める動きとして、ID3 を付与する動きがある<sup>4)</sup>。しかし、ID3 は特定健診の ID1 において前述の表記ゆれ(全角半角の違い、頭に 0 が付与されているか否か)を改善し、特定健診等のデータとレセプトのデータの患者突合性を改善するものである。ID3 を使用しても名寄せを行う必要性には変わりはない。

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。

## 2. 研究方法

本研究では、2013 年 4 月～2016 年 3 月の計 36 ヶ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 レセプトとは、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 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。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。

### 2.1 医科(入院/入院外)及びDPCレセプトで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	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が、NDB では再審査請求レセプトは取り込まれないため、3 ヶ月間程度は新旧の 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 ヶ月間で同じ 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 ヶ月間で同じ 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。

### 2.2 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、2.1 と同様の手法で ID2 を用いて名寄せを行った。

### 2.3 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: 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。先出と後出の ID1 は英数字が異なるが同一人物と考えられる。なお、「ID1: kwyr1s5T:Loi2g7Zx」と「ID1:Loi2g7Zx:

kwyrIs5T」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。

## 2.4 名寄せ作業

2.3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)。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がそのまま残したものも含まれている。

先出ID1と後出ID1の一対一対応テーブル

子ID1	親ID1
p8d89jss	ue8k22ue
ue8k22ue	Ajdke783
Ajdke783	78wmdjfg



名寄せテーブル(初期セット)

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	



名寄せテーブル(1回目更新後)

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	



名寄せテーブル(2回目更新後)

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	



名寄せテーブル(3回目更新後)

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

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2015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」(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(2015年10月8日 受付番号1123)を得た。

## 3. 研究結果

### 3.1 NDBにおけるID1、ID2の特性に関する検討

NDBのIDは、(1)ID1とID2が変化しない、(3)ID1、ID2のいずれかが変化する、(3)ID1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、一つのID2に対し2つのID1が発生する)。今回の観察期間(36ヶ月間)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。

表1 ID1とID2の重複

レセプト 種別	ID1に対するID2の重複			ID2に対するID1の重複		
	重複数	件数	構成比	重複数	件数	構成比
DPC	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	***	0.00%			
	6	***	0.00%			
合計	6,647,932	100%		6,670,937	100%	
医科 入院	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	***	0.00%
合計	5,713,500	100%		5,743,180	100%	
医科 入院外	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	***	0.00%	11	***	0.00%
	12	***	0.00%	12	***	0.00%
	14	***	0.00%			
	合計	117,048,101	100%		132,050,478	100%
調剤	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	***	0.00%	9	14	0.00%
	10	***	0.00%	10	***	0.00%
	14	***	0.00%	11	***	0.00%
				12	***	0.00%
	合計	94,937,633	100%		95,851,172	100%

(\*\*\*は10未満のため非公表、+++は10以上マスキング)

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見える)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の重複で見ると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た際に氏名の表記ゆれが多く発生するためであると考えられ、IDを用いた名寄せの品質に影響を及ぼす事象である。

### 3.2 ID0とID1の名寄せ精度の比較

2015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及び2015年10月の推計人口と比較した性年齢階級別の結果を表2と表3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5.8%、女性で6.4%であった。ID1により名寄せされた患者数は、男性は0~9歳や65~69歳、75歳以上、女性は加えて20~35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、0～9歳や男性80歳以上、女性85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よそ推計人口の範囲内に収まっていた。

表2 2015年における年齢階級別 ID0 患者数、ID1 患者数及び推計人口(男性)

年齢階級	ID1患者数	ID0患者数	ID1-ID0	追加名寄せ率	推計人口(平成27年)	年齢階級受診率(ID1)	年齢階級受診率(ID0)
0～4歳	2,902,841	2,701,346	201,495	6.9%	2,550,921	113.8%	105.9%
5～9歳	2,986,718	2,820,997	165,721	5.5%	2,714,591	110.0%	103.9%
10～14歳	2,862,608	2,735,321	127,287	4.4%	2,868,024	99.8%	95.4%
15～19歳	2,642,449	2,536,433	106,016	4.0%	3,085,416	85.6%	82.2%
20～24歳	2,469,556	2,330,268	139,288	5.6%	3,046,392	81.1%	76.5%
25～29歳	2,680,572	2,522,220	158,352	5.9%	3,255,717	82.3%	77.5%
30～34歳	3,031,674	2,873,256	158,418	5.2%	3,684,747	82.3%	78.0%
35～39歳	3,393,432	3,237,598	155,834	4.6%	4,204,202	80.7%	77.0%
40～44歳	3,928,757	3,766,946	161,811	4.1%	4,914,018	79.9%	76.7%
45～49歳	3,568,198	3,421,875	146,323	4.1%	4,354,877	81.9%	78.6%
50～54歳	3,313,382	3,170,192	143,190	4.3%	3,968,311	83.5%	79.9%
55～59歳	3,224,537	3,071,854	152,683	4.7%	3,729,523	86.5%	82.4%
60～64歳	3,808,634	3,451,475	357,159	9.4%	4,151,119	91.7%	83.1%
65～69歳	4,741,221	4,316,936	424,285	8.9%	4,659,662	101.8%	92.6%
70～74歳	3,319,841	3,212,377	107,464	3.2%	3,582,440	92.7%	89.7%
75～79歳	3,215,431	2,732,719	482,712	15.0%	2,787,417	115.4%	98.0%
80～84歳	2,053,758	2,043,227	10,531	0.5%	1,994,326	103.0%	102.5%
85～89歳	1,162,019	1,154,472	7,547	0.6%	1,056,641	110.0%	109.3%
90～94歳	417,574	414,334	3,240	0.8%	333,335	125.3%	124.3%
95～99歳	88,427	87,706	721	0.8%	63,265	139.8%	138.6%
100歳以上	12,620	12,552	68	0.5%	8,383	150.5%	149.7%
不明					828,411		
合計	55,824,249	52,614,104	3,210,145	5.8%	61,841,738	90.3%	85.1%

表3 2015年における年齢階級別 ID0 患者数、ID1 患者数及び推計人口(女性)

年齢階級	ID1患者数	ID0患者数	ID1-ID0	追加名寄せ率	推計人口(平成27年)	年齢階級受診率(ID1)	年齢階級受診率(ID0)
0～4歳	2,748,574	2,562,889	185,685	6.8%	2,436,785	112.8%	105.2%
5～9歳	2,819,348	2,667,262	152,086	5.4%	2,585,196	109.1%	103.2%
10～14歳	2,679,450	2,563,528	115,922	4.3%	2,731,293	98.1%	93.9%
15～19歳	2,679,371	2,561,397	117,974	4.4%	2,922,972	91.7%	87.6%
20～24歳	3,032,249	2,779,923	252,326	8.3%	2,921,735	103.8%	95.1%
25～29歳	3,413,184	3,096,815	316,369	9.3%	3,153,895	108.2%	98.2%
30～34歳	3,794,206	3,489,924	304,282	8.0%	3,606,131	105.2%	96.8%
35～39歳	4,059,079	3,784,893	274,186	6.8%	4,111,955	98.7%	92.0%
40～44歳	4,550,591	4,280,884	269,707	5.9%	4,818,200	94.4%	88.8%
45～49歳	4,095,073	3,858,235	236,838	5.8%	4,307,927	95.1%	89.6%
50～54歳	3,754,610	3,532,266	222,344	5.9%	3,961,985	94.8%	89.2%
55～59歳	3,620,039	3,372,870	247,169	6.8%	3,785,723	95.6%	89.1%
60～64歳	4,141,780	3,782,352	359,428	8.7%	4,303,891	96.2%	87.9%
65～69歳	5,137,773	4,787,259	350,514	6.8%	4,984,205	103.1%	96.0%
70～74歳	3,905,513	3,765,333	140,180	3.6%	4,113,371	94.9%	91.5%
75～79歳	4,002,735	3,409,676	593,059	14.8%	3,489,439	114.7%	97.7%
80～84歳	2,965,788	2,944,547	21,241	0.7%	2,967,904	100.0%	99.2%
85～89歳	2,135,719	2,116,096	19,623	0.9%	2,060,616	103.6%	102.7%
90～94歳	1,142,237	1,130,903	11,334	1.0%	1,015,785	112.4%	111.3%
95～99歳	374,850	371,411	3,439	0.9%	296,082	126.6%	125.4%
100歳以上	76,207	75,684	523	0.7%	53,380	142.8%	141.8%
不明					625,347		
合計	65,128,376	60,934,147	4,194,229	6.4%	65,253,007	99.8%	93.4%

### 3.3 3年の名寄せ ID 捕捉率の検討

ID0 で名寄せした 2013 年～2015 年のデータを用いて、ID の捕捉率を検証した。2013 年度を 100% としたとき、2014 年、2015 年に出現する ID0 の割合を男女別に示す(表 4、表 5)。死亡患者や年度内未受診者も含まれることに留意する必要はあるが、1 年で 10% 程度捕捉率が低下することが分かった。1 年で死亡する人や単年度限りの受診者も考慮すると、5～6% が見込まれるが今後の検証が必要な結果となった。

年齢階級別にみると、20～24 歳にピークが存在し、その後ゆるやかに減少するものの 85 歳以上でさらに増加する結果となった。

表4 2013年～2015年のID捕捉率(男性)

満年齢階級	捕捉患者数			脱落率	
	2013	2014	2015	2014	2015
0～4歳	2,771,836	2,605,080	2,492,733	6%	10%
5～9歳	2,817,459	2,645,331	2,502,248	6%	11%
10～14歳	2,793,074	2,593,271	2,376,648	7%	15%
15～19歳	2,493,326	2,138,046	1,846,311	14%	26%
20～24歳	2,291,451	1,783,726	1,545,973	22%	33%
25～29歳	2,612,312	2,125,291	1,856,644	19%	29%
30～34歳	2,916,629	2,462,412	2,177,310	16%	25%
35～39歳	3,388,503	2,931,083	2,606,639	13%	23%
40～44歳	3,624,797	3,168,144	2,832,335	13%	22%
45～49歳	3,183,234	2,822,016	2,558,981	11%	20%
50～54歳	2,982,234	2,687,216	2,475,269	10%	17%
55～59歳	3,048,311	2,772,717	2,575,662	9%	16%
60～64歳	3,832,411	3,436,340	3,210,298	10%	16%
65～69歳	3,747,865	3,480,790	3,321,085	7%	11%
70～74歳	3,319,978	3,136,271	2,979,843	6%	10%
75～79歳	2,642,922	2,481,235	2,359,388	6%	11%
80～84歳	1,905,220	1,761,457	1,621,891	8%	15%
85～89歳	1,054,270	925,399	805,457	12%	24%
90～94歳	343,818	278,434	222,183	19%	35%
95～99歳	78,689	55,938	39,185	29%	50%
100歳以上	11,215	6,823	4,100	39%	63%
総数	51,859,554	46,297,020	42,410,183	11%	18%

表5 2013年～2015年のID捕捉率(女性)

満年齢階級	捕捉患者数			脱落率	
	2013	2014	2015	2014	2015
0～4歳	2,631,496	2,455,561	2,342,469	7%	11%
5～9歳	2,667,387	2,488,503	2,337,878	7%	12%
10～14歳	2,606,693	2,406,947	2,211,559	8%	15%
15～19歳	2,545,157	2,266,765	2,043,680	11%	20%
20～24歳	2,815,416	2,300,540	2,028,828	18%	28%
25～29歳	3,295,039	2,759,453	2,443,648	16%	26%
30～34歳	3,613,670	3,148,750	2,847,355	13%	21%
35～39歳	4,014,149	3,568,381	3,253,591	11%	19%
40～44歳	4,180,832	3,738,287	3,421,031	11%	18%
45～49歳	3,626,668	3,277,676	3,027,346	10%	17%
50～54歳	3,365,422	3,073,068	2,861,927	9%	15%
55～59歳	3,397,284	3,095,614	2,891,806	9%	15%
60～64歳	4,247,141	3,884,053	3,661,480	9%	14%
65～69歳	4,185,212	3,959,582	3,813,201	5%	9%
70～74歳	3,901,843	3,730,497	3,575,450	4%	8%
75～79歳	3,340,996	3,191,305	3,097,005	4%	7%
80～84歳	2,807,870	2,680,704	2,553,906	5%	9%
85～89歳	1,983,260	1,830,867	1,677,370	8%	15%
90～94歳	1,016,985	878,040	747,606	14%	26%
95～99歳	322,954	250,191	190,097	23%	41%
100歳以上	67,295	44,162	28,454	34%	58%
総数	60,634,782	55,030,960	51,057,702	9%	16%

### 4. 考察

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 をひとつの ID1 へ統合する手法が本研究で提案した名寄せ手法「ID0」である。この名寄せ手法は、ID1 と ID2 を併用するだけでなく、転帰区分及び 3 ヶ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 となってしまうが、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に少ないため、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分かっている。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。資格喪失に起因するによる新旧 ID1 のレセプト発生

期間の重複については、ID1 が変化した場合の3ヶ月間のID2 探索を一度に行わず、1ヶ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 が記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分かった。

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。ただし、NDBの患者ID数は1年間に出現したIDが元となっているため、死亡患者もカウントされている。人口統計との比較では人口統計が10月1日の一時点の人口を示すことから、NDBの患者数は過大に算出されることが影響するであろう。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、各患者に一生不変のIDを割り当てるほかない。

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こりにくく、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。

また、今回3年分のNDBデータを用いて名寄せを行ったが、経年とともに一定の脱落患者が発生することに留意する必要がある。これは、ID0の名寄せが完全ではないことと、死亡患者を明確に定義することが困難であることから、脱落患者の実際がどこまでなのかは依然不透明な状況である。なお、転帰に「死亡」が含まれている患者を死亡情報として信用したとしても、全体の7割程度しか捕捉していない可能性がある。そのようなことから、転帰の精度を高めることが今後求められるといえよう。

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分かっていない。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)差異が出ることが分かっている<sup>5)</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」として認識されることが大きい。NDBの主な年齢層は高齢者であり、ID1のみを用いた患者数推計では実態から大きく乖離した結果となってしまう恐れがある。

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%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。一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。二つ目の理由として、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>6)</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るが、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、遠く離れた土地でまった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、逆に出張等で遠隔地での受診を繰り返す患者の事例や、治療内容がほぼ一致していても別人である事例の存在も否定できない。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の実現には多くの課題が残されている。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一つとしており、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。

## 5. 結語

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、本研究では検証した。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、29年度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)「病床機能の分化・連携や病床の効率的利用等のために必要となる実施可能な施策に関する研究」、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の一環として実施したものである。

## 参考文献

- 1) 中央社会保険医療協議会 総会(第356回) 議事次第 横断的事項(その2)について。レセプト情報・特定健診等情報データベース(NDB)の概要。厚生労働省保険局医療課, 2017。  
[<http://www.mhlw.go.jp/file/05-Shingikai-12404000-Hokenkyoku-Iryouka/0000170931.pdf> (cited 2017-Sep-08)].
- 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 2016; 36(1): 272-275.
- 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) 第38回レセプト情報等の提供に関する有識者会議 資料. 資料2 レセプトデータと特定健診等データの突合について(報告). 厚生労働省保険局医療介護連携政策課保険システム高度化推進室, 2017.  
[<http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000174510.pdf> (cited 2017-Sep-08)].
- 5) 山田 茂. 2005年国勢調査結果の精度の検討. 中央調査報 2008.
- 6) 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.