

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)
「臨床疫学に活用可能な NDB 等データセットの作成に関する研究」
分担研究報告書 (令和 5 年度)

(長期追跡性に優れた個人名寄せ ID の検討)

研究分担者 今村知明 (奈良県立医科大学 公衆衛生学)
研究分担者 野田龍也 (奈良県立医科大学 公衆衛生学)
研究協力者 明神大也 (奈良県立医科大学 公衆衛生学)

研究要旨

本研究では「健康・医療・介護分野の大規模データの利活用を推進する」ことの一環で、NDB に含まれる個人単位被保険者番号に基づく個人名寄せ ID (id5) が付与されたデータの長期追跡性について検討及び評価を行う。さらに、id5 が付与される前のデータについても、既存の id とのより精緻なデータ連結の可能性について検討を行う。2013 年 4 月診療分から 2023 年 3 月診療分までの NDB 特別抽出データのうち、RE レコードの id1n, id2, id5, 診療年月を用いた。そして、id1n と id2 と診療年月から id0 を作成した。さらに、作成された id0 に対して対応する id5 が存在する場合は id0 を id5 で上書きして、それを id0n とした。

id5 の設定状況を調査したところ。全レセプトに含まれる有効な id5 の割合は DPC では 95% 程度、医科入院では 92% 程度、医科入院外では 97% 程度であった。id5 が 9 の割合は 1% 未満、id5 が null だった比率は 2~8% 程度であった。また、2022 年 4 月に存在した id0, id0n, id5 に対し、2023 年 3 月まで存在する id を追跡した結果、id0, id0n, id5 いずれも、9 ヶ月後で約 90% の捕捉率を示し、id0n が最も捕捉率が高かった。

本研究において id5 の設定状況を把握したとともに、id0 および id5 の長期追跡の妥当性評価を行い、id0 と id5 を組み合わせた新たな id0n の開発に踏み込んだ。今後、死亡者を除外するなど、id5 及び id0n の精度をより詳細に検証していく。

研究協力者

東野恒之 (三菱総合研究所)

A. 研究目的

昨今のコロナ禍に伴う社会状況や、近年の急速な少子高齢化を背景とする社会構造の変化に伴い、より適正な医療を提供するための基礎資料として、健康・医療・介護分野のビッグデータ解析の必要性が高まっている。実際 NDB 及び介護 DB の情報はすでに医療・介護それぞれの分野での政策利用や研究利用が開始されており、提供件数は徐々に増加している。一方データの容量が膨大であること、

データの構造が複雑であること等から、これらデータの構造や操作を知悉した研究者でなければ解析を行うのは容易ではなく、医療者等いわゆる臨床系の研究者だけでは取り扱いが困難なデータとなっている。一方で臨床系の研究者は、日々医療・介護の現場で医療を担っており、適正な医療の提供・実施にあたっての課題にも直面していることから、こうした研究者が自ら NDB や介護 DB 等の大規模データを簡便に分析できるようになれば、それら課題の解決に貢献することが期待される。

加えて、NDB は 2020 年 10 月から介護 DB 及び 2022 年 4 月から DPC データベースとの連結が可能となり、さらに「医療・介護データ

等の解析基盤」の実運用が開始される予定であることから、将来的には現在よりも大規模なデータ解析が可能になる見込みである。これらの大規模データを広く研究者が利用できるようにするためには、より簡便にデータ分析できる環境が整備される必要がある。

本研究は「健康・医療・介護分野の大規模データの利活用を推進する」ことを目的に、匿名医療保険等関連情報データベース（NDB）、および介護保険総合データベース（介護DB）さらに2020年から提供が開始されている医療・介護の連結情報から、より利用者にとって使いやすいデータセットの開発を目指すものである。その一環で、長期追跡性に優れた個人名寄せIDの検討として、個人単位被保険者番号に基づく個人名寄せIDが付与されたデータを受領し、長期追跡性について検討及び評価を行う。さらに、後ろ向きのデータ連結性（2021年10月以前のデータとの連結）についても、既存のIDとのより精緻なデータ連結の可能性について検討を行う。

B. 研究方法

2013年4月診療分から2023年3月診療分までのNDB特別抽出データのうち、REレコードのid1n, id2, id5, 診療年月を用いた。そして、id1nとid2と診療年月からid0を作成した。さらに、作成されたid0に対して対応するid5が存在する場合はid0をid5で上書きして、それをid0nとした。

id5は2022年3月診療分から含まれていたが、処理の関係で大部分を2022年4月診療分から使用した。

- ① 2022年4月診療分から2023年3月診療分までの各月のid5の格納状況を、DPC・医科入院・医科入院外に分けてレセプト枚数ベースで集計した。
- ② 2022年4月診療分から2022年12月診療分までの患者数としてid0及びid5の件数を、性別・年齢階級別に集計した。年齢は

2022年12月末時点の年齢とした。また参考として2023年1月1日時点の住民基本台帳の人数を示した。

- ③ id5とid1nの対応割合を調べた。具体的には、2022年4月から2023年3月までのid1nとid5の対応表を作成しdistinctをとった。
- ④ ③と比較するため、2022年4月から2023年3月までのid1nとid0の対応表を作成しdistinctをとった。
- ⑤ 2022年4月に存在したid0, id0n, id5に対し、2023年3月まで存在するidを追跡した。
- ⑥ 2013年4月に存在したid0, id0nに対し、2023年3月まで存在するidを追跡した。

（倫理面への配慮）本研究は、奈良県立医科大学医の倫理審査を受けて実施された。また、NDBデータを用いた分析結果は、厚生労働省及び社会保険診療報酬支払基金の公表物確認を受けて承認された。

C. 研究結果

研究方法①～⑥の結果を以下に示す。可読性の関係上、一部考察も含めている。

- ① 2022年4月診療分から2023年3月診療分までの各月のid5の格納状況を表1に示した。id5はid1nやid2と同様、16進数64桁のハッシュ値であったが、一部にnullまたは9が含まれていた。16進数64桁のハッシュ値が含まれるid5を有効値とみなした。その結果、全レセプトに含まれる有効なid5の割合はDPCでは95%程度、医科入院では92%程度、医科入院外では97%程度であった。id5が9の割合は1%未満、id5がnullだった比率は2～8%程度であった。id5が9は正しく何らかの理由でid5が取得できなかった場合、id5がnullは全額公費であった場合と考えられる。全額公費を除くと、全体の99%近くにid5が含まれており、id5の格納率の高さが示され

た。

- ② 2022年4月診療分から2022年12月診療分までのid0及びid5のユニークな件数を、性別・年齢階級別に表2及び図1に示した。id5=9またはid5=nullとなっている値は、それぞれid5に9またはnullが入っていたレセプトの、ユニークなid0数を意味する。id5の数はid0の数よりやや少なくなっている。id0で名寄せできなかったものもid5に含まれている可能性はあるものの、id5の数とid5がnullの数を合わせるとid0の数とほぼ合致することから、id0もid5とほぼ同等の名寄せ精度を示している可能性も示唆された。なお、参考に掲載した住民基本台帳の人数と比べると、0-4歳と75歳以降は100%近くのレセプト発生率であった。85歳以上では住民基本台帳の人数よりid0,id5の数のほうが多かったが、これは集計対象期間内に死亡した人も含まれているためと考えられる。
- ③ 2022年4月診療分から2023年3月診療分までのid1nとid5の対応件数を表3に示す。id1n:id5=1:1はid5に対応するid1nが1つしかなかった場合である。id1n:id5=n:1は1つのid5に対応するid1nが複数あった場合であり、12ヶ月の間に保険証が変わったことが示唆される。id1n:id5=1:nは1つのid1nに対応するid5が複数存在した場合で、id1n:id5=n:mはid1nとid5が多対多になった場合である。id1n:id5=1:nまたはn:mだった場合は、同性の双子の可能性もあるがそれ以外にid5の誤りの可能性も考えられる。具体的なケースをNDB抽出データから取り上げると、以下の2つのパターンであることが分かった。1つ目はid1nとid2が一致してid5が異なるケース、これが118,590件（レセプト枚数にして1,533,250枚）であった。2つ目はid5が一致しているが性別又は生年月が異なるケース、これが327,830件（レセプト枚数にして

4,307,520枚）であった。これらはid5の異常と言えるかもしれないが、全体の0.5%未満であり、NDB分析には大きな支障がないと考えられる。

次に、id1n:id5=n:1(id5=5/null)はid5に9又はnullが入っていたレセプトのユニークなid1nの数である。id5=nullが全額公費とした場合、日本で約250万人の医療扶助等による受診者がいると推計された。

- ④ 2022年4月診療分から2023年3月診療分までのid1nとid5の対応件数を表4に示す。id1n:id0=1:1はid0に対応するid1nが1つしかなかった場合である。id1n:id0=n:1は1つのid0に対応するid1nが複数あった場合であり、id0の作成過程で名寄せされたものである。この結果によると、1700万程度のid1nが800万程度のid0に名寄せされたことになる。なお、医療保険から全額公費に移行したid0の数は170,560件であった。
- ⑤ 2022年4月に存在したid0, id0n, id5に対し、2023年3月まで存在するidを追跡した結果のID数と捕捉率を表5及び図2に示す。id0,id0n,id5いずれも、9ヶ月後で約90%の捕捉率であったが、id0nが最も捕捉率が高かった。
- ⑥ 2013年4月に存在したid0, id0nに対し、2023年3月まで存在するidを追跡した結果のID数と捕捉率を表6及び図3に示す。id0,id0nともに、9年6か月後で約80%の捕捉率であったが、id0に比べてid0nのほうが捕捉率は高かった。

D. 考察

本研究で、id5の設定状況を把握したとともに、id0およびid5の長期追跡の妥当性評価を行い、id0とid5を組み合わせた新たなid0nの開発に踏み込んだ。

本研究を通して日本で初めてNDBに含まれる

id5 の設定状況を公表した。その結果、id5 は 1 年間で 0.5% 程度誤結合している可能性はあるものの、大部分は正しく連結されていることが分かった。それとともに我々が従来作成していた id0 の正確性も高いことが裏付けられた。

医療扶助は 250 万人と結果に示したが、厚生労働省が発表している生活保護受給者数は約 200 万人で、本研究の名寄せが不十分だと思われる。名寄せされた id0 に対して対応する id5 が存在する場合は id5 で上書きする id0n を作成した。id0n のメリットは id0 で名寄せできなかった id を id5 で名寄せが可能になったこと、2022 年 2 月以前の id5 が存在しなかったレセプトについても名寄せが可能になったこと、公費に移行した患者についても名寄せが可能になったことである。しかしながら id0 をベースにしている (id0 は id1n をベースにしている) ため、id5 で可能な同性の双子の切り分けは不可能である。

今後、死亡者を除外するなど、id5 及び id0n の精度をより詳細に検証していく。

E. 結論

本研究において、id5 の追跡を行ったとともに id0 の妥当性を検証した。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

なし

2. 学会発表

なし

H. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

1) 令和 4 年 4 月 28 日 第 5 回 医療扶助に関する検討会資料 1

<https://www.mhlw.go.jp/content/12002000/000934593.pdf>

表1 2022年度のid5設定状況（レセプト枚数ベース）

診療年月	DPC						
	id5有効値	id5=9	id5=null	合計	id5有効値の比率	id5=9の比率	id5=nullの比率
202204	954,350	7,078	39,688	1,001,116	95.3%	0.7%	4.0%
202205	957,595	6,918	40,666	1,005,179	95.3%	0.7%	4.0%
202206	1,033,643	7,452	43,515	1,084,610	95.3%	0.7%	4.0%
202207	989,852	7,658	41,404	1,038,914	95.3%	0.7%	4.0%
202208	918,731	7,412	37,578	963,721	95.3%	0.8%	3.9%
202209	938,990	7,281	39,025	985,296	95.3%	0.7%	4.0%
202210	1,000,254	7,727	41,658	1,049,639	95.3%	0.7%	4.0%
202211	981,715	6,964	39,981	1,028,660	95.4%	0.7%	3.9%
202212	949,990	6,290	38,479	994,759	95.5%	0.6%	3.9%
202301	901,028	6,111	37,291	944,430	95.4%	0.6%	3.9%
202302	971,558	6,407	40,263	1,018,228	95.4%	0.6%	4.0%
202303	1,045,278	6,588	43,372	1,095,238	95.4%	0.6%	4.0%

診療年月	医科入院						
	id5有効値	id5=9	id5=null	合計	id5有効値の比率	id5=9の比率	id5=nullの比率
202204	1,144,251	6,386	97,943	1,248,580	91.6%	0.5%	7.8%
202205	1,129,549	6,161	97,728	1,233,438	91.6%	0.5%	7.9%
202206	1,158,550	6,283	99,319	1,264,152	91.6%	0.5%	7.9%
202207	1,165,103	6,662	99,414	1,271,179	91.7%	0.5%	7.8%
202208	1,161,232	6,585	99,504	1,267,321	91.6%	0.5%	7.9%
202209	1,146,171	6,519	98,559	1,251,249	91.6%	0.5%	7.9%
202210	1,143,703	6,771	97,776	1,248,250	91.6%	0.5%	7.8%
202211	1,140,398	6,142	96,945	1,243,485	91.7%	0.5%	7.8%
202212	1,155,245	5,918	98,105	1,259,268	91.7%	0.5%	7.8%
202301	1,129,150	5,695	97,588	1,232,433	91.6%	0.5%	7.9%
202302	1,137,444	5,560	97,173	1,240,177	91.7%	0.4%	7.8%
202303	1,165,063	5,474	98,381	1,268,918	91.8%	0.4%	7.8%

診療年月	医科入院外						
	id5有効値	id5=9	id5=null	合計	id5有効値の比率	id5=9の比率	id5=nullの比率
202204	82,525,970	496,171	2,203,448	85,225,589	96.8%	0.6%	2.6%
202205	80,548,455	472,157	2,169,132	83,189,744	96.8%	0.6%	2.6%
202206	82,777,730	462,745	2,227,036	85,467,511	96.9%	0.5%	2.6%
202207	85,290,190	501,145	2,202,035	87,993,370	96.9%	0.6%	2.5%
202208	84,850,299	494,256	2,208,168	87,552,723	96.9%	0.6%	2.5%
202209	82,276,028	458,019	2,192,986	84,927,033	96.9%	0.5%	2.6%
202210	83,525,939	523,328	2,220,060	86,269,327	96.8%	0.6%	2.6%
202211	83,612,710	502,293	2,210,009	86,325,012	96.9%	0.6%	2.6%
202212	86,954,462	492,012	2,246,400	89,692,874	96.9%	0.5%	2.5%
202301	80,577,072	451,461	2,134,000	83,162,533	96.9%	0.5%	2.6%
202302	80,757,258	439,526	2,138,218	83,335,002	96.9%	0.5%	2.6%
202303	90,541,374	471,948	2,278,397	93,291,719	97.1%	0.5%	2.4%

表2 性別・年齢階級別の患者数 (id0,id5) の分布 (2022年4月診療分～12月分)

性別	年齢階級	id0件数	id5件数	id5 = 9	id5 is null	住基(R5.1.1)
男性	0～4歳	2,141,056	2,072,462	100,103	17,942	2,192,250
	5～9歳	2,463,379	2,416,437	43,589	24,444	2,589,182
	10～14歳	2,492,087	2,446,049	33,495	30,676	2,769,436
	15～19歳	2,273,831	2,225,341	36,291	28,808	2,863,895
	20～24歳	2,189,483	2,120,515	60,790	14,478	3,242,208
	25～29歳	2,249,746	2,177,476	54,317	17,295	3,324,472
	30～34歳	2,315,005	2,247,208	48,238	20,120	3,347,657
	35～39歳	2,596,962	2,524,723	48,441	25,736	3,731,475
	40～44歳	2,853,432	2,772,087	48,902	35,794	4,078,212
	45～49歳	3,423,644	3,318,743	53,362	56,343	4,808,345
	50～54歳	3,564,774	3,434,291	54,012	83,068	4,827,997
	55～59歳	3,236,545	3,110,133	33,222	96,608	4,085,447
	60～64歳	3,147,469	2,994,849	37,064	110,348	3,701,790
	65～69歳	3,234,084	3,073,172	40,756	119,801	3,631,126
	70～74歳	4,035,120	3,828,185	54,011	172,959	4,336,773
	75～79歳	3,114,322	2,964,465	10,107	130,279	3,191,907
	80～84歳	2,393,880	2,300,854	1,872	92,932	2,375,135
	85～89歳	1,486,600	1,438,625	1,295	47,483	1,413,673
	90～94歳	605,614	591,950	619	13,200	542,360
	95～99歳	137,069	134,850	145	2,071	112,356
100歳以上	13,937	13,722	17	204	10,052	
	小計	49,968,039	48,206,137	760,648	1,140,589	61,175,748
女性	0～4歳	2,028,062	1,963,782	94,079	16,216	2,087,071
	5～9歳	2,313,236	2,270,965	40,877	21,624	2,461,529
	10～14歳	2,320,590	2,278,475	31,518	27,979	2,632,354
	15～19歳	2,262,144	2,215,783	31,205	30,033	2,724,879
	20～24歳	2,525,778	2,437,172	60,879	18,484	3,068,216
	25～29歳	2,697,997	2,593,007	55,474	23,186	3,124,431
	30～34歳	2,750,794	2,656,544	48,585	27,709	3,166,865
	35～39歳	3,050,078	2,955,687	45,383	35,288	3,571,272
	40～44歳	3,308,510	3,209,185	43,834	44,080	3,935,243
	45～49歳	3,895,322	3,774,230	47,999	61,637	4,656,040
	50～54歳	3,994,955	3,858,297	47,557	77,709	4,720,154
	55～59歳	3,494,477	3,371,809	41,546	73,443	4,058,198
	60～64歳	3,313,474	3,189,930	44,680	71,185	3,759,265
	65～69歳	3,447,809	3,322,537	47,512	81,621	3,814,948
	70～74歳	4,534,891	4,388,427	70,112	139,227	4,841,620
	75～79歳	3,796,330	3,622,142	12,680	148,562	3,907,207
	80～84歳	3,284,111	3,135,292	2,649	149,551	3,310,601
	85～89歳	2,527,512	2,421,282	2,406	106,684	2,480,888
	90～94歳	1,463,838	1,415,613	1,609	47,897	1,380,195
	95～99歳	524,794	509,325	553	15,274	462,073
100歳以上	98,189	95,193	96	2,976	78,033	
	小計	57,632,891	55,684,677	771,233	1,220,365	64,241,082
	合計	107,600,930	103,890,814	1,531,881	2,360,954	125,416,830

図1 性別・年齢階級別の患者数 (id0,id5) の分布 (2022年4月診療分～12月分)

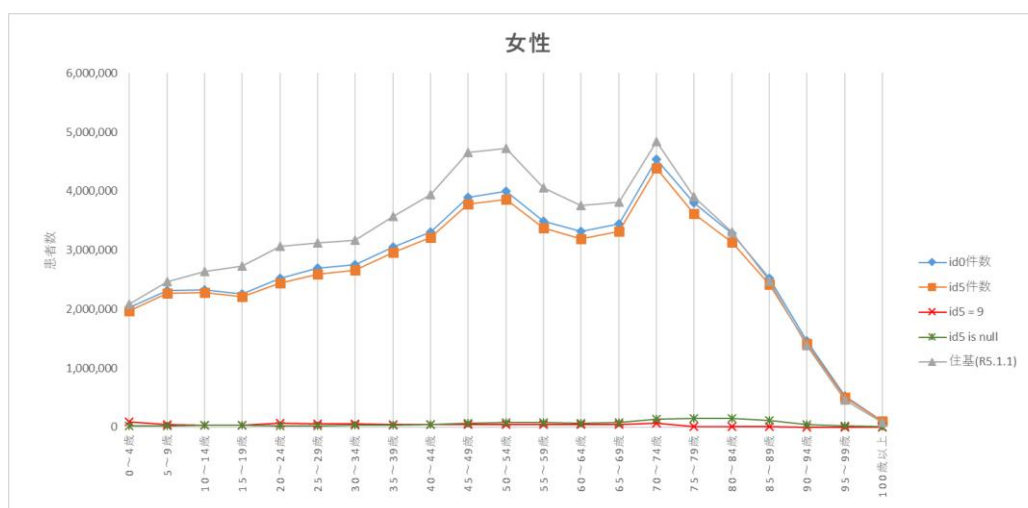
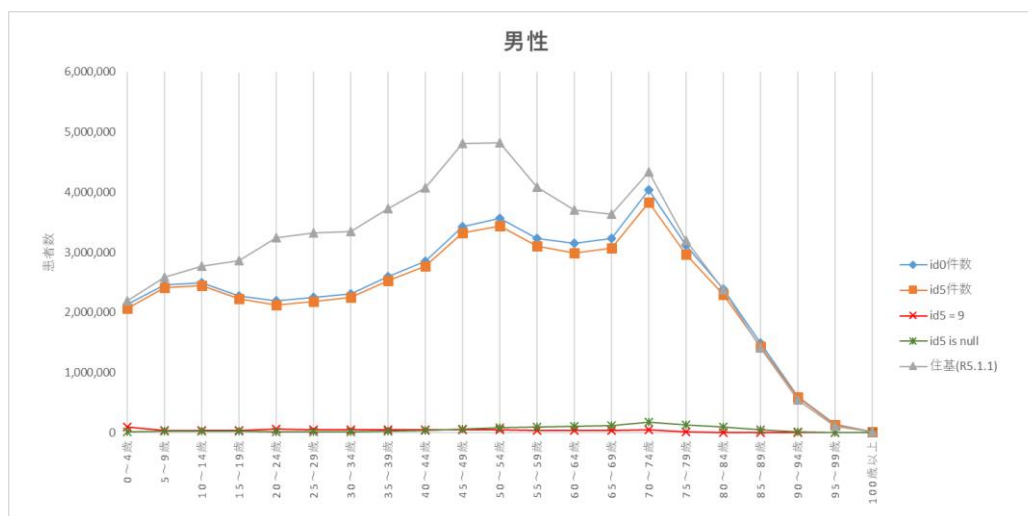


表 3 id5 設定状況

	id1n		id5	
	件数	構成比	件数	構成比
id1n : id5 = 1 : 1	98,671,350	81%	98,671,350	91%
id1n : id5 = n : 1	17,928,280	15%	8,512,100	8%
id1n : id5 = 1 : n	137,530	0%	276,090	0%
id1n : id5 = n : m	485,500	0%	416,530	0%
id1n : id5 = n : 1 (id5=9)	1,274,040	1%	10未満	
id1n : id5 = n : m (id5=9を含む)	512,960	0%	513,170	0%
id1n : id5 = n : 1 (id5=null)	2,513,440	2%	0	0%
id1n : id5 = 1 : 2 (id5=null : 9)	3,320	0%	10未満	
id1n : id5 = 1 : 2 (id5=hash値 : 9)	10未満		10未満	
全件数	121,526,410		108,389,240	

表 4 id0 設定状況

	id1n		id0	
	件数	構成比	件数	構成比
id1n : id0 = 1 : 1	104,949,470	86%	104,949,470	93%
id1n : id0 = n : 1	16,576,940	14%	7,955,120	7%
全件数	121,526,410	100%	112,904,600	100%

表 5 2022 年度における id 数とその追跡率

診療年月	id数			追跡率		
	id0	id0n	id5	id0	id0n	id5
202204	59,463,613	59,386,609	57,565,542	100.0%	100.0%	100.0%
202205	58,230,825	58,287,823	56,514,833	97.9%	98.1%	98.2%
202206	57,874,469	57,987,698	56,222,195	97.3%	97.6%	97.7%
202207	57,536,571	57,705,305	55,945,574	96.8%	97.2%	97.2%
202208	57,115,762	57,336,802	55,580,649	96.1%	96.5%	96.6%
202209	56,609,145	56,882,161	55,126,021	95.2%	95.8%	95.8%
202210	56,081,255	56,429,315	54,672,134	94.3%	95.0%	95.0%
202211	55,398,825	55,788,122	54,039,741	93.2%	93.9%	93.9%
202212	54,530,313	54,952,089	53,214,228	91.7%	92.5%	92.4%
202301	53,101,544	53,562,102	51,840,702	89.3%	90.2%	90.1%
202302	50,453,276	50,924,177	49,244,545	84.8%	85.8%	85.5%
202303	43,726,451	44,165,299	42,616,545	73.5%	74.4%	74.0%

図 2 2022 年度における id 数とその追跡率

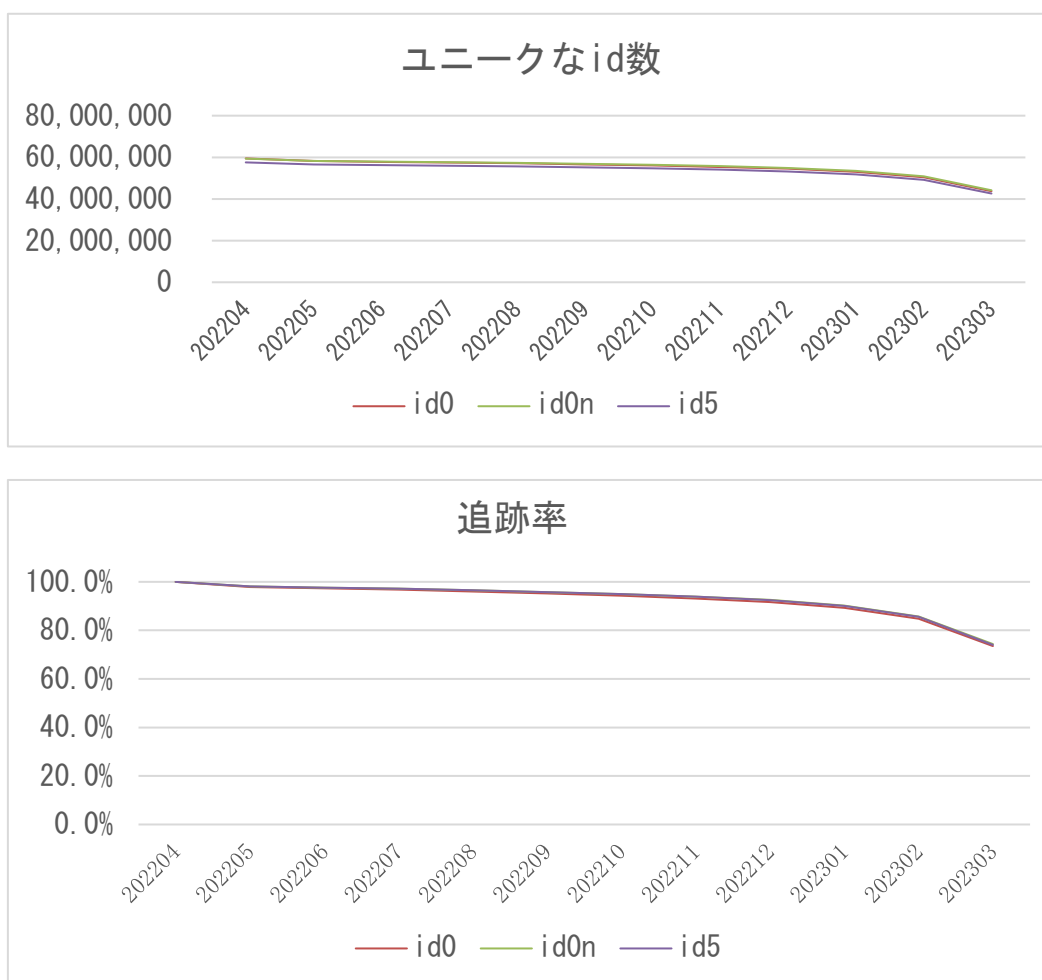


表 6 2013 年度～2022 年度の
Id 数とその追跡率

診療年月	id数		追跡率	
	id0	id0n	id0	id0n
201304	50,116,547	50,106,238	100.0%	100.0%
201305	49,922,975	49,912,666	99.6%	99.6%
201306	49,851,840	49,841,531	99.5%	99.5%
201307	49,794,025	49,783,716	99.4%	99.4%
201308	49,735,951	49,725,642	99.2%	99.2%
201309	49,684,151	49,673,842	99.1%	99.1%
201310	49,634,294	49,623,985	99.0%	99.0%
201311	49,580,576	49,570,267	98.9%	98.9%
201312	49,527,576	49,517,267	98.8%	98.8%
201401	49,468,795	49,458,486	98.7%	98.7%
201402	49,413,561	49,403,252	98.6%	98.6%
201403	49,357,013	49,346,704	98.5%	98.5%
201404	49,279,202	49,268,893	98.3%	98.3%
201405	49,227,248	49,216,939	98.2%	98.2%
201406	49,179,968	49,169,659	98.1%	98.1%
201407	49,134,118	49,123,809	98.0%	98.0%
201408	49,086,231	49,075,922	97.9%	97.9%
201409	49,043,207	49,032,898	97.9%	97.9%
201410	48,997,935	48,987,626	97.8%	97.8%
201411	48,949,994	48,939,685	97.7%	97.7%
201412	48,904,303	48,893,994	97.6%	97.6%
201501	48,847,068	48,836,759	97.5%	97.5%
201502	48,791,988	48,781,679	97.4%	97.4%
201503	48,740,439	48,730,130	97.3%	97.3%
201504	48,669,218	48,658,909	97.1%	97.1%
201505	48,621,387	48,611,078	97.0%	97.0%
201506	48,578,094	48,567,785	96.9%	96.9%
201507	48,532,739	48,522,430	96.8%	96.8%
202101	44,965,130	44,954,821	89.7%	89.7%
202102	44,893,607	44,883,298	89.6%	89.6%
202103	44,816,046	44,805,737	89.4%	89.4%
202104	44,704,680	44,694,371	89.2%	89.2%
202105	44,615,221	44,604,912	89.0%	89.0%
202106	44,527,366	44,517,057	88.8%	88.8%
202107	44,426,440	44,416,131	88.6%	88.6%
202108	44,318,012	44,307,703	88.4%	88.4%
202109	44,200,465	44,190,156	88.2%	88.2%
202110	44,084,082	44,073,773	88.0%	88.0%
202111	43,958,468	43,948,159	87.7%	87.7%
202112	43,823,526	43,813,217	87.4%	87.4%
202201	43,671,892	43,661,583	87.1%	87.1%
202202	43,504,754	43,494,445	86.8%	86.8%
202203	43,321,495	43,311,186	86.4%	86.4%
202204	43,063,037	43,052,728	85.9%	85.9%
202205	42,823,252	42,831,227	85.4%	85.5%
202206	42,560,005	42,584,556	84.9%	85.0%
202207	42,258,174	42,300,110	84.3%	84.4%
202208	41,809,592	41,870,055	83.4%	83.6%
202209	41,229,531	41,309,869	82.3%	82.4%
202210	40,702,302	40,806,280	81.2%	81.4%
202211	40,058,773	40,179,764	79.9%	80.2%
202212	39,189,538	39,325,902	78.2%	78.5%
202301	37,827,450	37,982,742	75.5%	75.8%
202302	35,685,727	35,850,180	71.2%	71.5%
202303	30,720,862	30,882,352	61.3%	61.6%

図 3 2013 年度～2022 年度の
Id 数とその追跡率

