

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^{*1}、野田龍也^{*1}、西岡祐一^{*1 *2}、明神大也^{*1}、
東野恒之^{*3}、松居宏樹^{*4}、加藤源太^{*5}、今村知明^{*1}

*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 経営イノベーション本部

*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

医療情報学会 COI開示

奈良県立医科大学 公衆衛生学講座
久保 慎一郎

演題発表に関連し、開示すべきCOI関係にある
企業などはありません。

- 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
- 日本における保険診療受診者の悉皆調査となりうる
- 各医療機関のレセプトは、**患者ごとに、毎月、診療報酬を請求するので、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。**
- **ただし、名寄せの困難さ**があり有病者の実人数は推計困難であった。
 - 生涯不変の個人ID(マイナンバー等)が用意されていない
 - レセプト同士を紐づけるために用意された2種類の個人ID「ID1」、「ID2」を用いて名寄せ行う必要がある
 - この2種類のIDはライフイベントに合わせて変更されてしまう

別人物にも関わらずIDが同じ（第一種過誤）

		ID 1（保険者番号（世帯）・生年月日・性別）が	
		同じ	異なる
ID 2 （氏名漢字・生 年月日・性別） が	同じ	X1 別人物同士で同一のID1, ID2が生成された場合	Y1 別人物同士で同一のID2が生成された場合 Y2 同姓同名で同一生年月日に生まれ、同じ性別だった場合
	異なる	Z1 別人物同士で同一のID1が生成された場合 Z2 同性・同一生年月日の複産(双子等) (複産全体は1万件/年 _{※1}) 過誤の発生がなくなるのは... 条件Y2：一方の姓・名が変更になった時点 条件Z2：両者が異なる保険者となった時点	(両ID不一致のため検討対象外) X～Zについては、条件が合致してしまった別人物のレセプトが、名寄せを行う期間内に発生した場合のみ過誤が発生する。 条件を潜在的に満たす別人物がいても、名寄せ期間内にレセプトが発生しない場合は問題とならない。

※1 平成26年 人口動態調査「上巻 出生 第4. 36表 単産－複産(複産の種類・出生－死産の組合せ)別にみた年次別分娩件数」

同一人物にも関わらずIDが合わない（第二種過誤）

		ID 1（保険者番号（世帯）・生年月日・性別）は	
		不変	変わる
ID 2 （氏名漢字・ 生年月日・性 別） は	不変	（両ID一致のため検討対象外）	A1-1 就職・転職 } 713万人/年 ^{※1} A1-2 離職 } A1-3 定年 } A1-4 就職・転職等で扶養が外れる 800万人/年 ^{※1} A2-1 扶養者の婚姻・離婚による保険変更 A2-2 養子縁組による扶養者変更に伴う保険変更 A2-3 扶養者の変更（家族：尊属・卑属・配偶者） A3 国保加入者における居住地変更 A4 後期高齢者制度への加入 122万人/年 ^{※2} A5 保険証番号が流出した患者 約18470人 ^{※3}
	変わる	B1 養子による改姓（労働者－大人） ^{※4} B2-1 結婚 ^{※5} 65万組/年・離婚による改姓 B2-2 扶養者の結婚・離婚等による改姓 B3 氏名変更 2万件/年 ^{※5} B4 記入ミス・氏名等の記載ゆれ B5 国籍取得に伴う氏名変更 千件/年 ^{※7}	AとBの掛け合わせ （例） A1-2 * B2-1 結婚に伴う退社・改姓 A1-1 * B2-1 離婚に伴う就職・改姓 A2-2 * B1 養子による改姓・保険者変更 ^{※4} C3 性転換 855人/年 ^{※6}

※1 厚生労働省「平成26年雇用動向調査結果の概況」注意)非正規雇用を含むため、実際の扶養についての数を示していない。

※2 総務省統計局人口推計（平成26年10月1日現在）-全国：年齢（各歳），男女別人口・都道府県：年齢（5歳階級），男女別人口-より日本人のみの75歳の人数を対象とした。

※3 厚生労働省保険局「被保険者証の記号番号を含む個人情報流出事案に係る調査結果の概要及び被保険者証の記号番号の変更対応等について」

※4 2014年における養子縁組は**83611件**（ただし、男女・大人子供含む）

※5 法務省 戸籍統計 2014年度時点

※6 司法統計 2015年時点

※7 法務省 帰化許可申請者数，帰化許可者数等の推移 平成27年 注意)国籍を取得する前から日本の氏名で保険証を取得している場合(氏名が変わらなかった場合)を含む

方法) 使用したレセプトデータと検証手法

Methods

- 使用データ：NDB
 - 内科入院レセプト、内科入院外レセプト、DPCLレセプト、調剤レセプト
- 期間：2013年4月～2016年3月(計36ヶ月分)
- 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**
- 作成したアルゴリズムをもとに名寄せを行い**ID0を作成**
- 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することで**ID0の妥当性を検証した**

方法) 名寄せアルゴリズム①

医科 (入院/入院外) ,DPCLレセプトで名寄せ用中間データを作成

Methods

- 「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DPCLレセプト」「DPC入院の総括対象DPCLレセプト」から**ID1、ID2、診療年月、転帰区分**を抽出。

このように同一診療月でID1とID2が入れ替わっている。
同時に変更された場合は捕捉が困難。

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	転帰区分
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

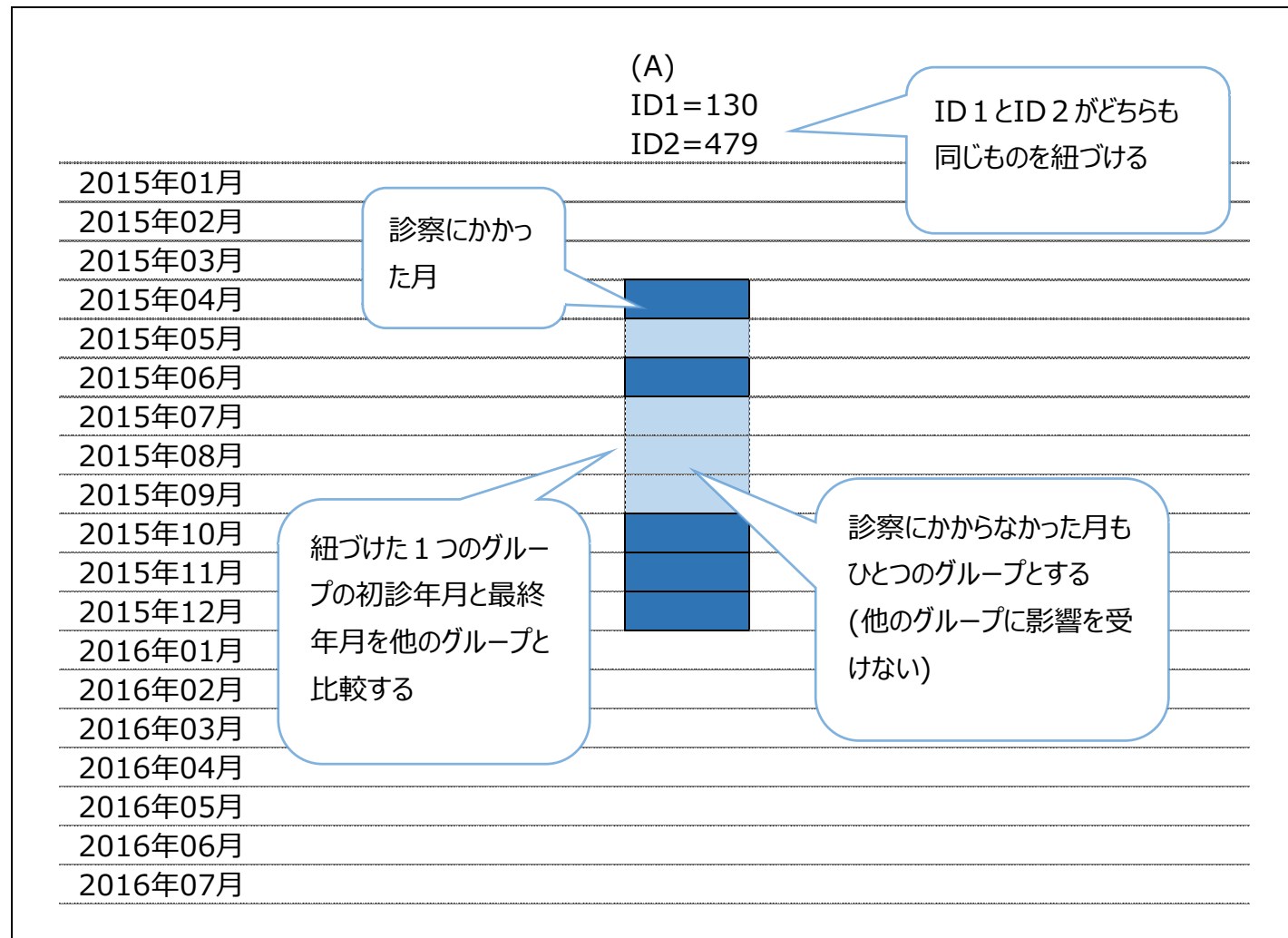
- 転帰区分**は死亡情報を用いた。複数月に渡ってID1が同じ患者は同一人物とした。
- 死亡転帰が示されている場合は、その月で名寄せ終了とした。

方法) 名寄せアルゴリズム①

名寄せの手法について (名寄せの前提条件)

Methods

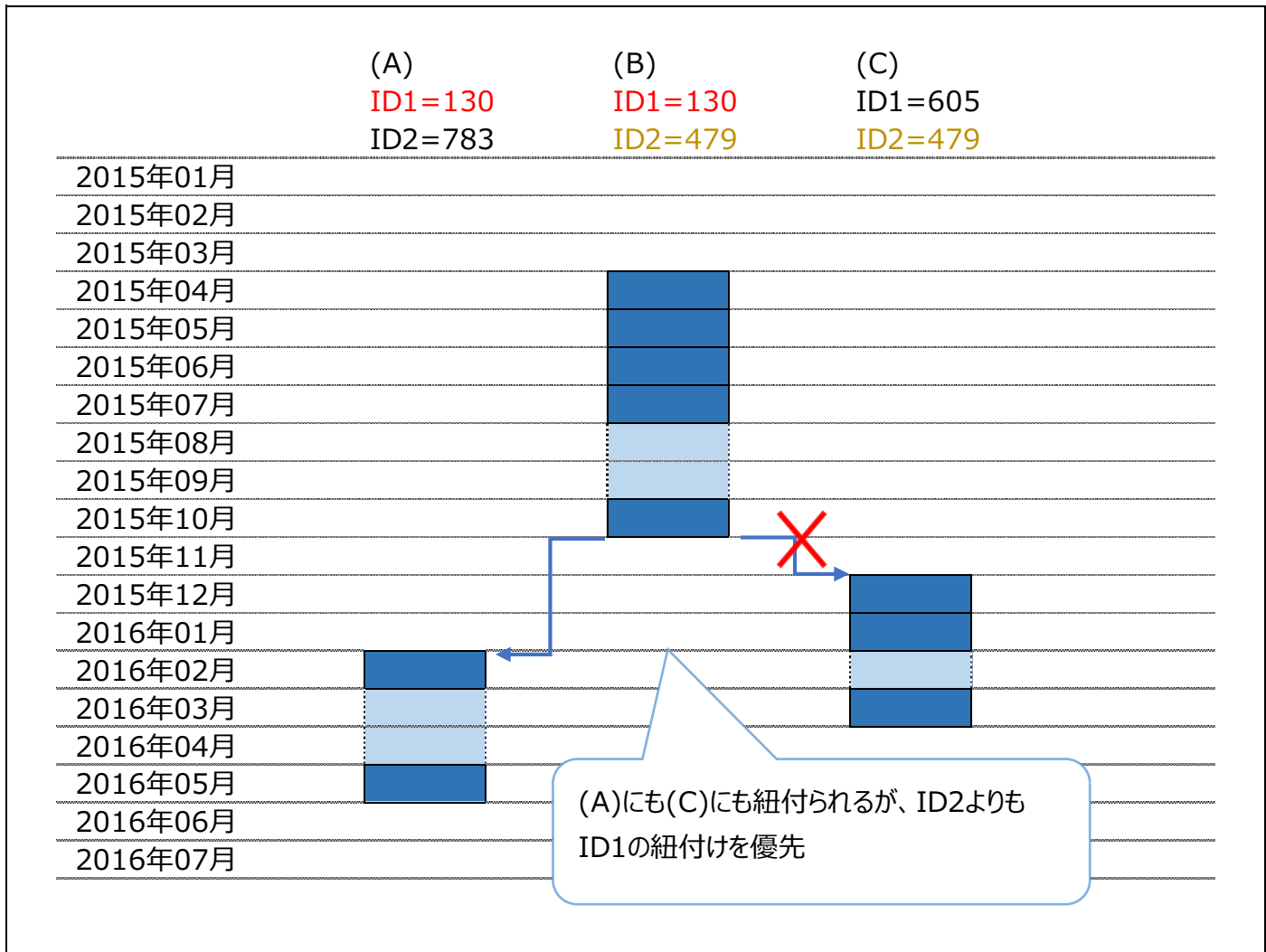
- ID1とID2がどちらも同一なものを1患者とする
- 診察にかかった月もかからない月も一つのグループとする
- 最初のレセプト発生月と最終の発生月を他のグループと比較する



方法) 名寄せアルゴリズム①

名寄せの手法について (名寄せルール①)

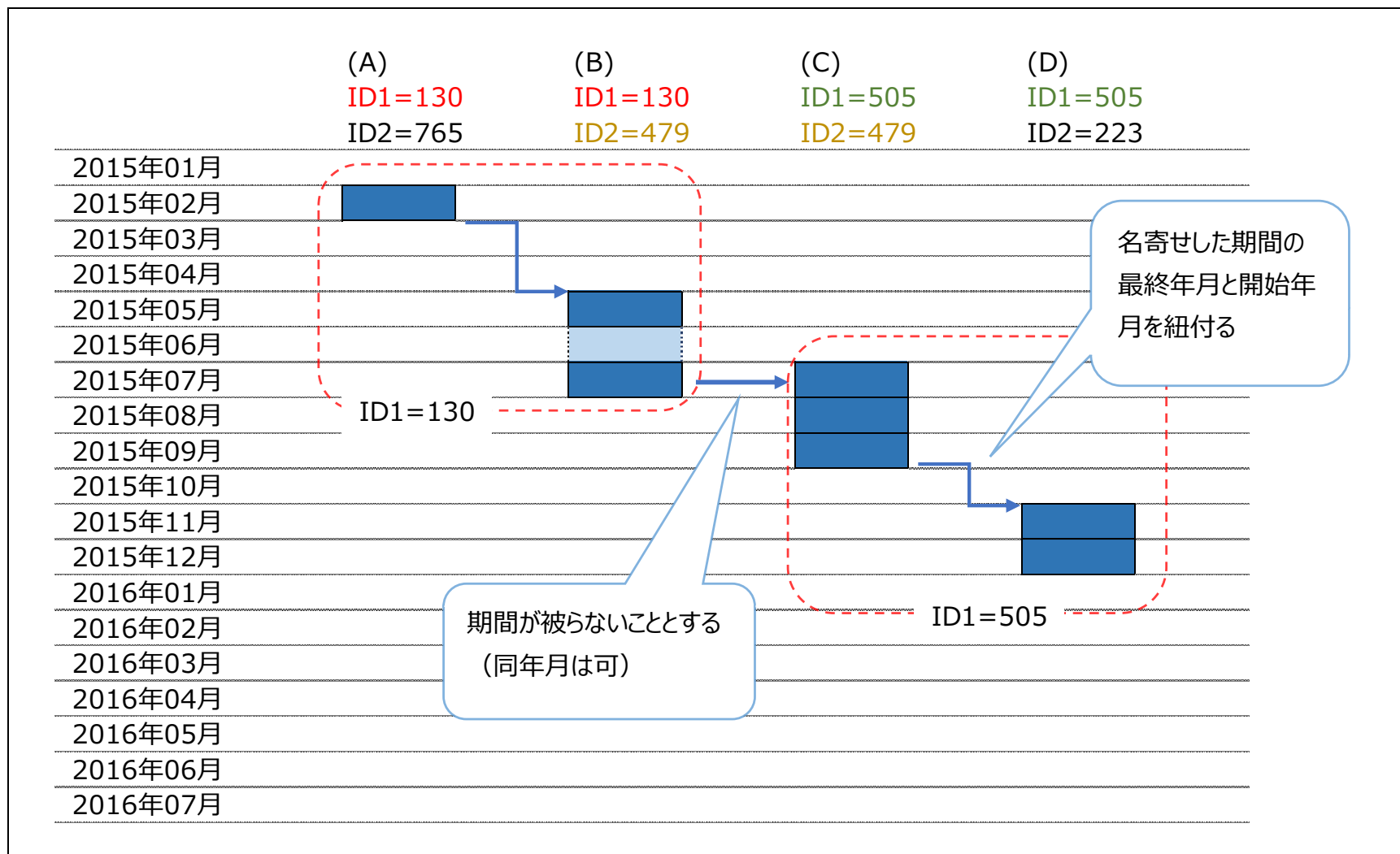
- ID 1 とID 2 のどちらにも紐づけ可能な場合ハッシュ 1 を優先する



方法) 名寄せアルゴリズム①

名寄せの手法について (名寄せルール②)

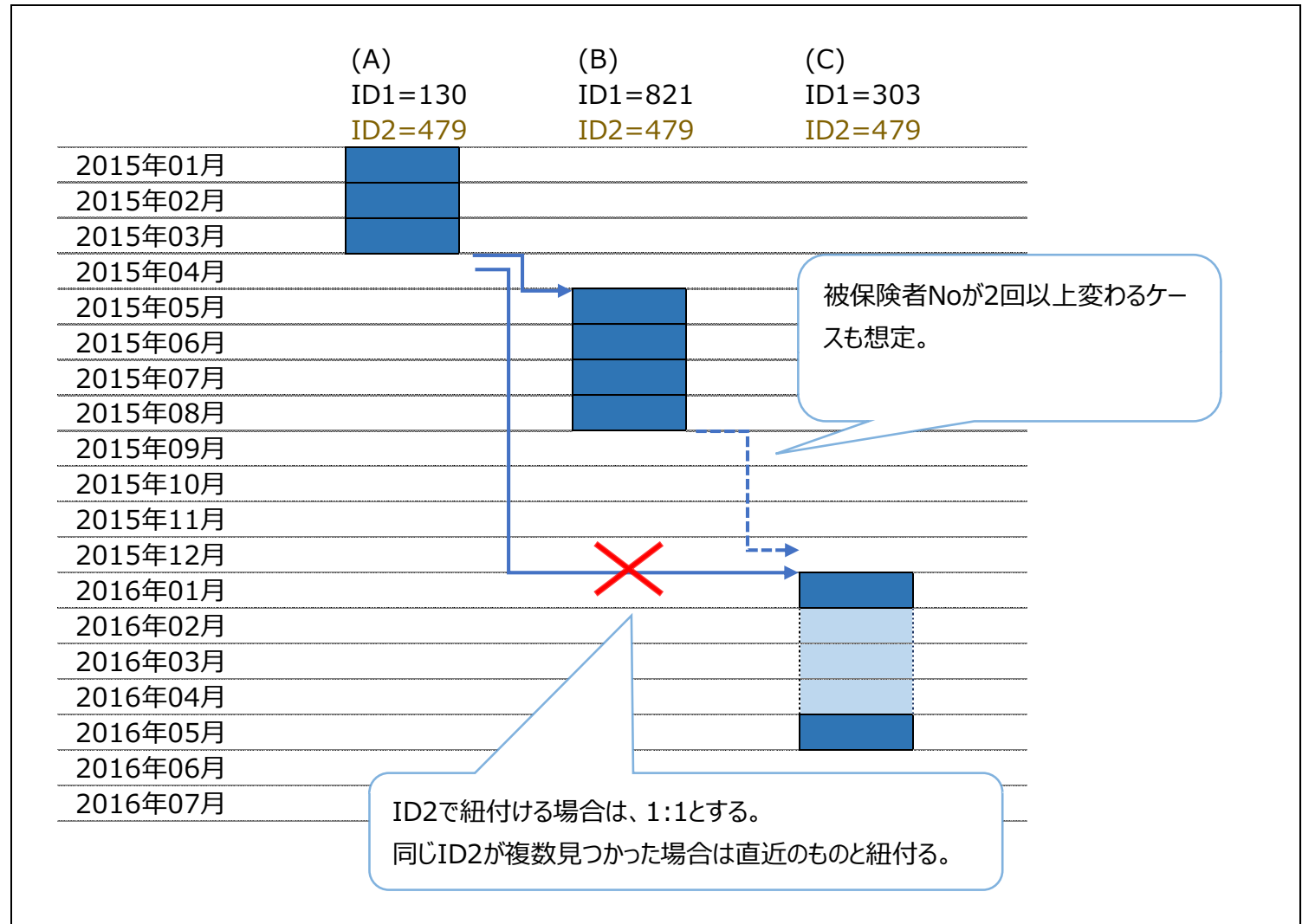
- レセプトの終了月と開始月が重ならないように紐づける



方法) 名寄せアルゴリズム①

名寄せの手法について (名寄せルール③)

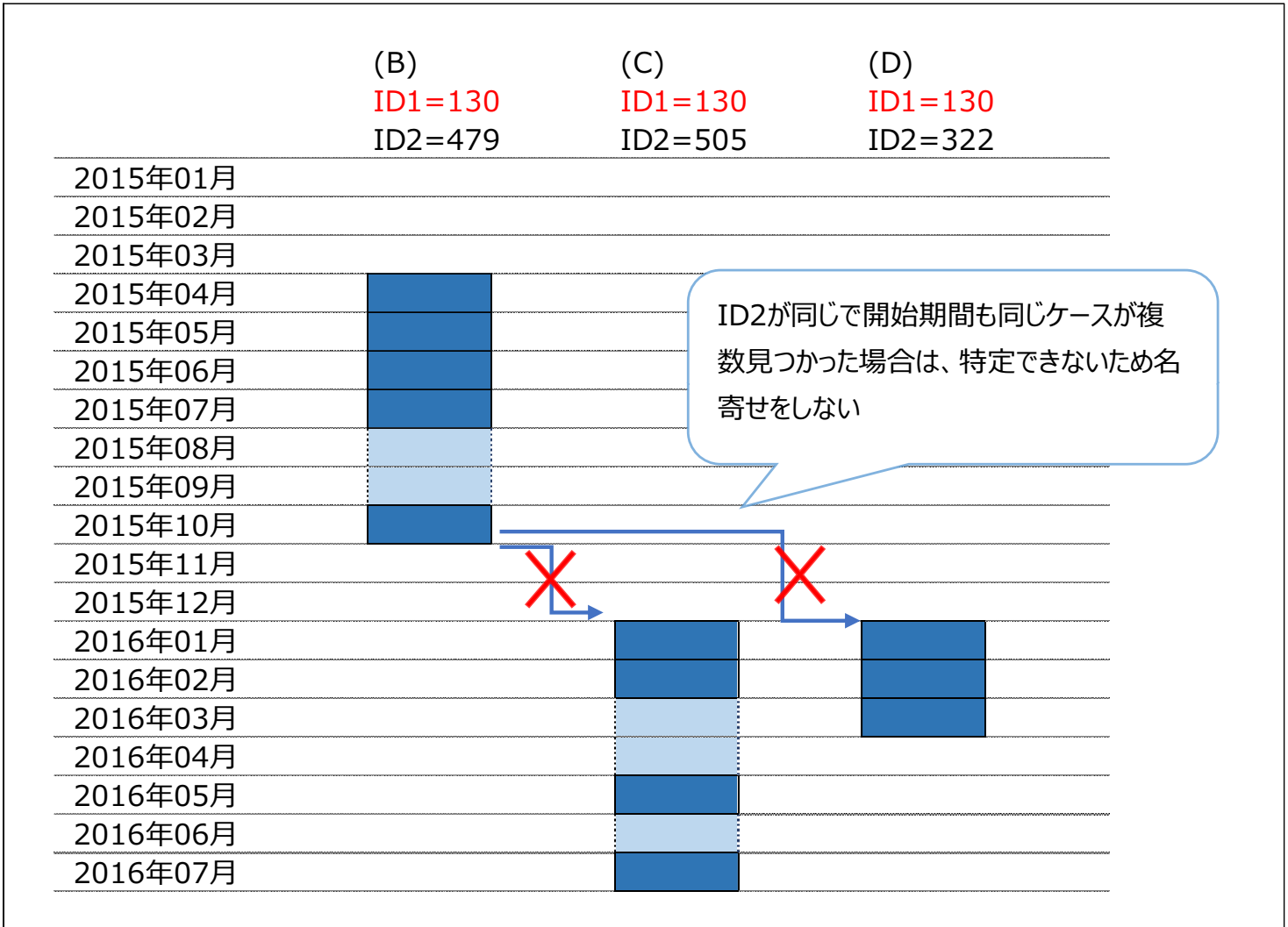
- 連結対象のレセプトグループが複数あった場合、もっとも近いものを紐づける



方法) 名寄せアルゴリズム①

名寄せの手法について (名寄せルール④)

- ID 1 が同じものが複数同じ期間にあるときは紐づけをしない



方法) 名寄せアルゴリズム②

調剤レセプトで名寄せ用中間データを作成

Methods

- 「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。
- 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、医科・DPCレセプトと同様の手法でID2を用いて名寄せを行った。

方法) 名寄せアルゴリズム③

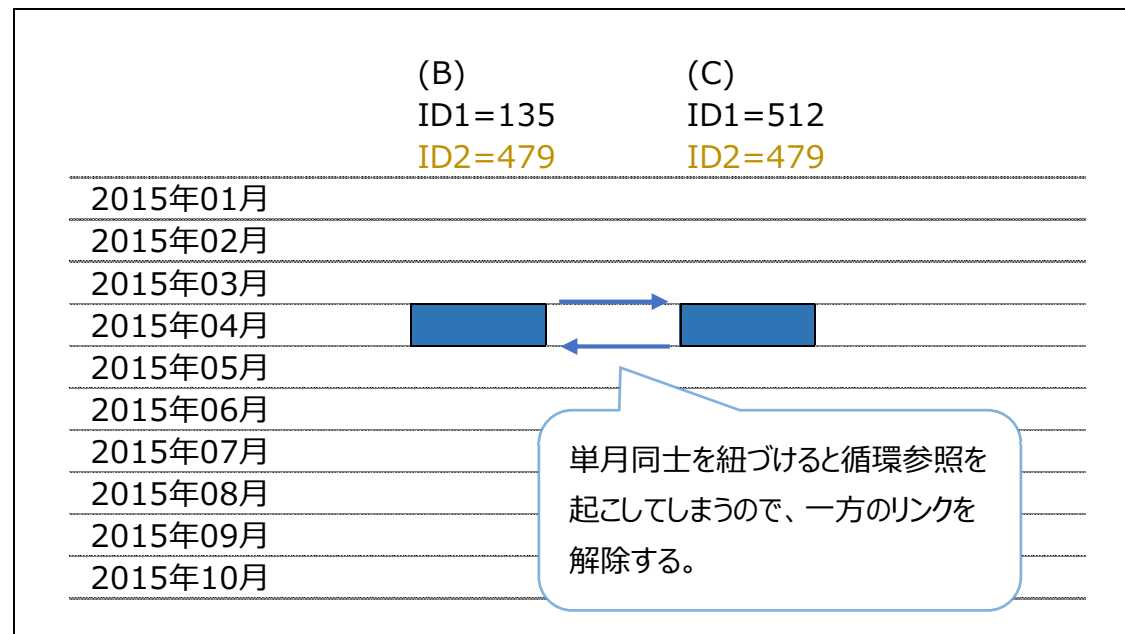
医科 (入院/入院外)、DPCレセプトと調剤レセプトを連結

Methods

- 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。
- 「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を指す。
- 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。

先出ID1と後出ID1の一対一対応テーブル

先出ID1	後出ID1
ue8k22ue	p8d89jss
Ajdke783	ue8k22ue
78wmdjfg	Ajdke783



方法) 名寄せアルゴリズム④

名寄せテーブルの作成

Methods

- 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。

先出ID1と後出ID1の一対一対応テーブル

先出ID1	後出ID1
ue8k22ue	p8d89jss
Ajdke783	ue8k22ue
78wmdjfg	Ajdke783

名寄せテーブル (3回目更新後)

ID1	ID0	年月From	年月To
p8d89jss	78wmdjfg	201304	201306
ue8k22ue	78wmdjfg	201307	201308
Ajdke783	78wmdjfg	201401	201403

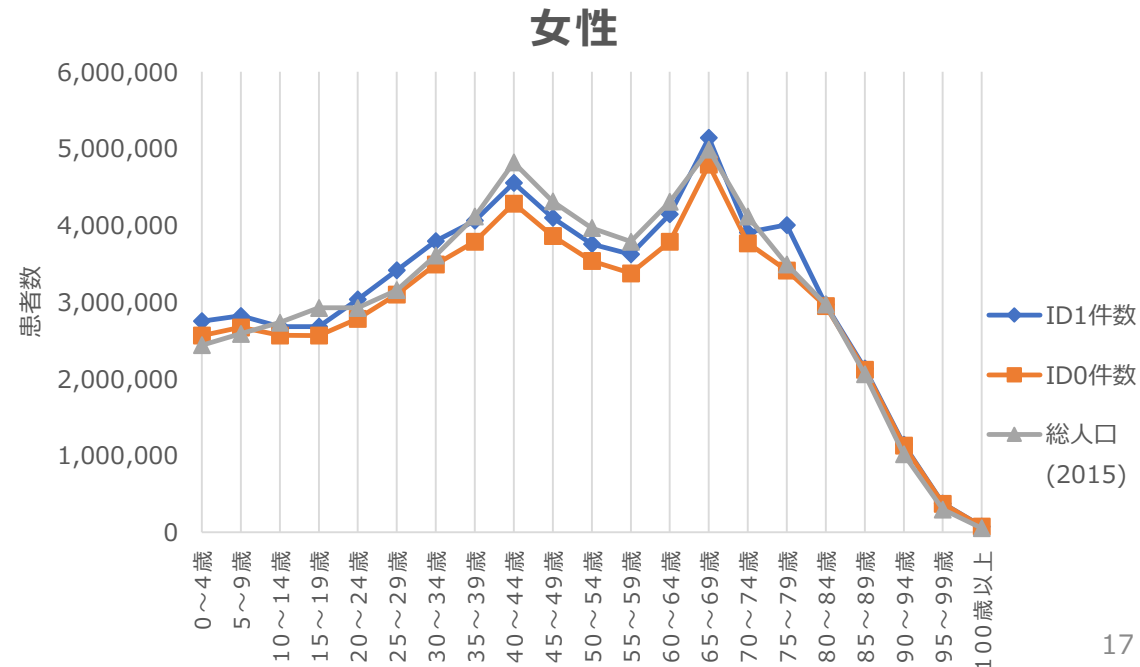
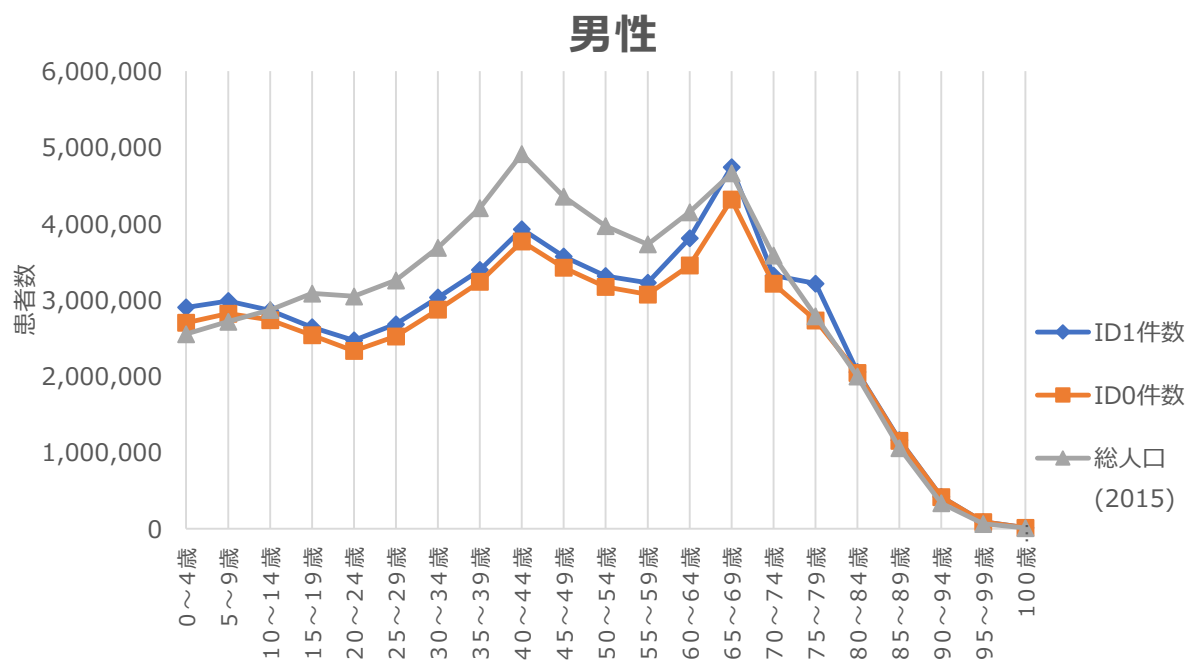
対応表を基に置き換える

コピー

結果) ID0とID1の名寄せ精度の比較

従来のID1名寄せとID0名寄せを推計人口と比較

- 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、9歳以下の乳幼児および65歳以上の老年世代で推計人口を上回っていた。名寄せ後はほぼ推計人口内に収まった。



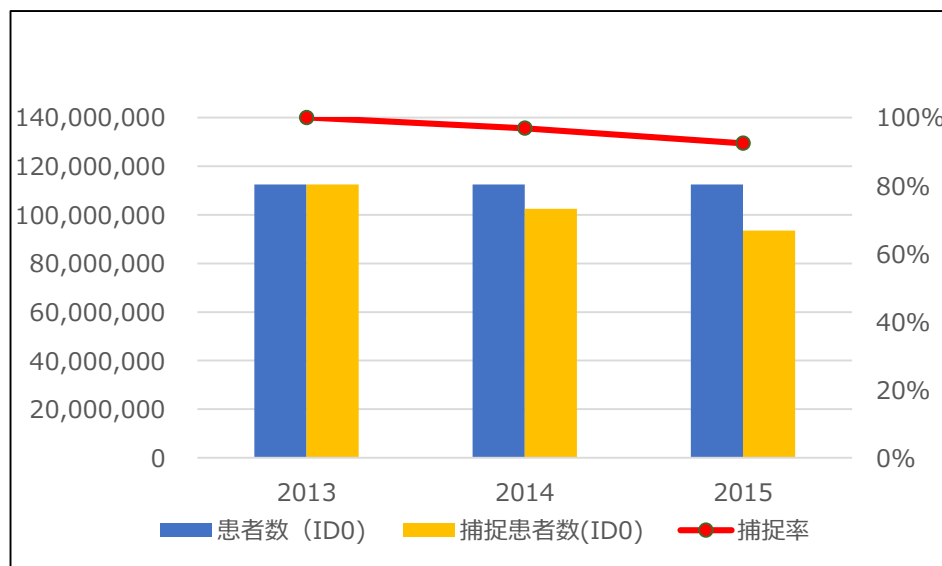
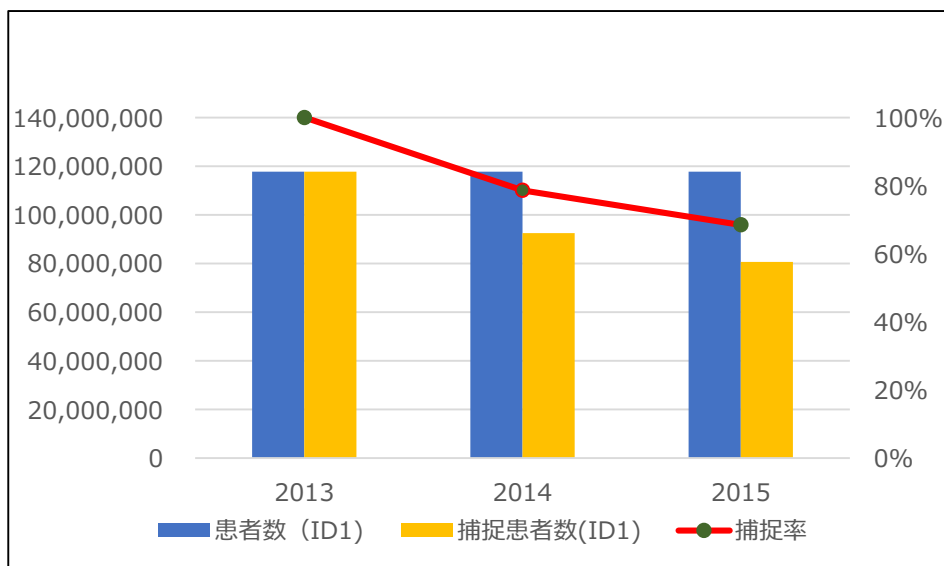
結果) ID0とID1の名寄せ精度の比較

3年分のデータを元にID0がどの程度脱落するか

- ID1では2年で32%の脱落が生ずるが、ID0では2年で8%に抑えられていることが想定される。

年度	患者数 (ID1)	捕捉患者数 (ID1)	捕捉率	脱落率
2013	117,760,974	117,760,974	100%	0%
2014	118,237,381	92,520,842	79%	21%
2015	119,328,548	80,657,315	68%	32%

患者数 (ID0)	追跡母数	年間死亡数	翌年受診歴あり	捕捉患者数 (ID0)	捕捉患者数補正 (ID0)	捕捉率	脱落率
112,492,323	112,492,323			112,492,323	112,492,323	100%	0%
116,917,899	112,492,323	1,269,000	5,205,045	102,436,642	108,910,687	97%	3%
113,051,576	112,492,323	1,302,000	9,198,818	93,465,870	103,966,688	92%	8%



考察) ID0の妥当性

- ID0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。
- これまでは、ID2が表記ゆれが大きく精度が低下するため、別人が同一IDとなる頻度が比較的少ない**ID1のみを用いて患者数を集計されてきた**。
- ID1では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**といった問題が存在した。
- ID0を用いるとID1のみを用いた名寄せに比し、男性で6.2%、女性で7.1%の追加的な名寄せに成功し、**実態に近い数値へと近づいた**。

考察) IDOの限界

- 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、**名寄せは完全ではない**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**している。
- 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。**転帰の精度を高めることが今後求められる。**
- 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を割り当てるほかない。**

考察) 年齢階級別の患者数(IDO)について

- IDO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が、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、**国勢調査は住民基本台帳人口と1%ほど差異が出ることが影響している可能性がある。**
- IDOによる名寄せの課題として、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。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い。
- 理由として「**出産に関連する医療保険の利用**」と「**氏名や保険者が変化しやすいこと**」が挙げられる。
 - 出産は自費診療だが異常分娩では周産期に医療保険が適用される。
例) 鉗子分娩や吸引分娩、帝王切開
 - 結婚と同時に退職する場合などでのID重複が20～30代の女性には多発する可能性がある。

考察) さらなる名寄せ率の向上にむけて

- ID0による名寄せの精度を上げるために、**受診地や病名などの連続性を用いる**方法が考えられる。
- 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が、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り、名寄せの実現には多くの課題が残されている。**

結論) ID0を用いたNDBの利用

Conclusions

- 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**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を用いるべき**である。

- 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研究」の一環として実施したものである。