

共同研究集会「官民オープンデータ利活用の動向及び人材育成の取組」(29-共研-5010)
統計数理研究所 2017.11.17

レセプト情報・特定健診等情報データ ベース(NDB)の活用： 血友病患者の医療体制の把握

奈良県立医科大学公衆衛生学講座
野田龍也 西岡祐一 久保慎一郎
明神大也 今村知明

本発表の目次

前半：
「そもそもNDBは何者か」

後半：
「NDBによる血友病患者数集計」

本発表に関連するCOI(利益相反)はありません。

NDBの威力(事前予想)

- レセプト情報・特定健診等情報データベース(NDB)は、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 (ただし、生活保護等の公費医療は抜け落ちているものがある)
- 1億2700万人の受診情報がおそらく9割以上は捕捉されている
- 疑いなく、世界最大のヘルスケアデータ、市販後調査データ、医療コホート…である
- 国民皆保険 + 定率負担 + 高額療養費制度 = 他国と異なり、選択バイアスも小さそう

3

レセプト情報・特定健診等情報データベース(NDB)のデータ件数 (平成29年3月末現在)

○レセプトデータ(平成21年度～平成28年12月診療分)

約128億8,400万件を格納

○特定健診・特定保健指導データ(平成20年度～平成27年度実施分)

約1億9,800万件を格納

ここでの「件」はレセプト枚数のこと、年間15億枚ほどのレセプトが発生していることになる。
(人数ベースでは、年間約1億人)

レセプトデータ		特定健診データ		特定保健指導データ	
—	—	H20年度	約2,000万件	H20年度	約39万件
H21年度	約12億1,700万件	H21年度	約2,200万件	H21年度	約58万件
H22年度	約15億1,100万件	H22年度	約2,300万件	H22年度	約61万件
H23年度	約16億1,900万件	H23年度	約2,400万件	H23年度	約72万件
H24年度	約16億8,100万件	H24年度	約2,500万件	H24年度	約84万件
H25年度	約17億2,800万件	H25年度	約2,600万件	H25年度	約84万件
H26年度	約18億0,800万件	H26年度	約2,600万件	H26年度	約86万件
H27年度	約18億9,200万件	H27年度	約2,700万件	H27年度	約87万件
H28年度※	約14億2,800万件				
計	約128億8,400万件	計	約1億9,300万件	計	約571万件

※H28.4-H28.12診療分

厚生労働省 平成29年5月17日 第37回レセプト情報等の提供に関する有識者会議 資料より

この発表はH25年度を使用

現在の当講座は太枠の3年分を使用

4

患者数 (ID0数) (レセプト種別)

H25年度NDBデータ

(2013/4~2014/3)

	医科(入院)	医科(入院外)	DPC	調剤	合計
レコード数	1,254,000,041	20,079,107,485	2,280,404,994	9,402,647,616	33,016,160,136
構成比	4%	61%	7%	28%	100%
レセプト数(レセプト通番の数)	15,400,802	939,633,615	11,351,253	592,079,015	1,558,464,685
構成比	1%	60%	1%	38%	100%
レコード数/レセプト	81	21	201	16	
レセプト数2(総括区分別レセプトの数)			14,104,128		
レコード数/レセプト			162		
ID1数 (ID1による患者実人数)	5,713,500	117,048,101	6,647,932	94,937,633	119,868,907
構成比	5%	98%	6%	79%	(合計>100%)
レコード数/ID1	219	172	343	99	
レセプト数/ID1	3	8	2	6	
ID0数 (ID0数による患者実人数)	5,642,905	109,780,160	6,586,835	89,722,055	112,133,984
構成比	5%	98%	6%	80%	(合計>100%)
レコード数/ID0	222	183	346	105	
レセプト数/ID0	3	9	2	7	

5

従来の「レセプト分析」とは一線を画する

	NDB		従来のレセプト分析		(参考)
	今回	従来	JMDC	都道府県/企業職保等	DPC調査データ
標本サイズ(1年あたり)	1億人	1億人	数百万人	数千~100万人	数百万人
悉皆性の高さ			×	×	×
全国代表性			×	×	×
個人識別性	ID0	高くない		DBによる	
外来の情報					一部
地域間比較			制限あり	困難	制限あり
プロセス指標					
アウトカム指標					
使いやすさ	ひどい↑	ひどい↑		DBによる	
対象年齢	全年齢	全年齢	75歳未満	DBによる	全年齢

「レセプト分析」とあった場合、どれに該当するかを見る

↑ どのように「ひどい」(とても分析にくい)かの良い解説としては、以下の文献など:

奥村泰之, 佐方信夫, 清水沙友里, 松居宏樹: ナショナルデータベースの学術利用促進に向けて: レセプトの落とし穴. Monthly IHEP 268: 16-25, 2017.

6

NDBの法的位置づけ

- 「高齡者の医療の確保に関する法律」
(高確法)による行政記録情報
(法16条「全国医療費適正化計画の作成等に資するため、
医療保険者から提出された情報の調査・分析を行う」)
- 統計法による統計とは異なる
(「マイクロデータの二次的利用」ではないかも…)

7

匿名化された個人の追跡(紐付け)

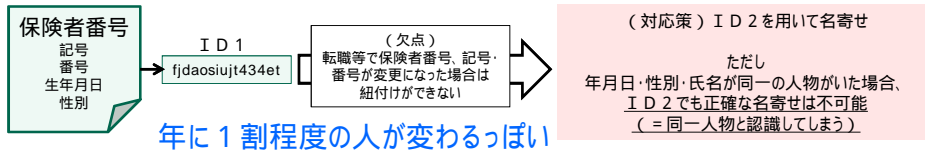
- 匿名化された個人を追跡する変数
NDBではID1とID2の2つが用意されている
- ID1: 保険者番号、記号、番号、生年月日、性別から作成
- ID2: 氏名、生年月日、性別から作成
- 上記下線部はライフイベントに合わせて可変的
- 一方のIDが変化した場合は他方のIDで個人をつないでほしい、という設計意図 (しかし…)

8

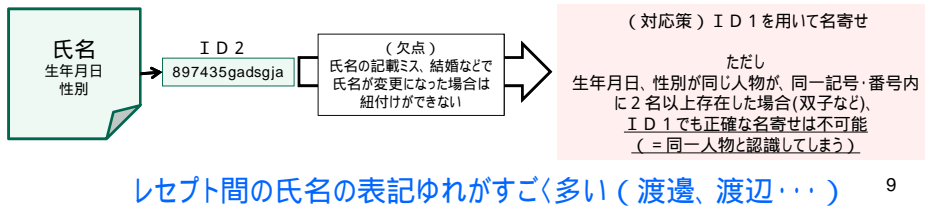
ID1とID2

- ある個人を示す匿名IDとして、2種のID (=ハッシュ化後の英数字) が付けられている

(1) ID1・「保険者番号」に依拠したID



(2) ID2・「氏名」に依拠したID



別人物にも関わらずIDが同じ (第一種過誤)

		ID1 (保険者番号(世帯)・生年月日・性別)が	
		同じ	異なる
ID2 (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) 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4.3.6表 単産・複産(複産の種類・出生・死産の組合せ)別みた年次別分母件数

同一人物なのに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 A1-2 離職 } 1 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年雇用労働者調査結果の概況」(注)改正職業区分を含むため、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・性別)・男女別人口・都道府県(年齢)・男女別人口 より日本人のみの75歳以上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8841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)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(氏名が変更された)は含まれない。


患者数 / ID1・ID2 / ID0

- 人の数の把握は医療統計や疫学分析の一丁目一番地
- レセプト件数は、医療機関ごと、患者ごと、月ごとの発行のため、そのままでは患者数はわからない。
- NDBには個々の患者を追跡する匿名変数(「ID1」「ID2」)がある(ものの、生涯不変ではない¹⁾)
- ID1,2に代わる新しい個人追跡用の匿名キー変数「ID0」を開発 → より精緻に患者数を算定可能²⁾)
- 今回研究はID0を使用した初の患者数推計となる。

1) 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と留意点. 日本健康開発雑誌 2017年

2) 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, 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)における患者突合(名寄せ)手法の改良と検証. 厚生指針 2017年

今回集計の経緯

- レセプトは請求書の束(検査結果などは載っていない)
- NDBで各傷病の患者を特定する作業は、店のレシートから「カレーライスを作った人」を特定する作業に似ている(カレールー+玉ねぎ+人参・・・カレーライスという蓋然性) 
- NDBの名寄せは複数のレシートから同一の購入者を特定する作業に似ている(→従来より少し精緻な名寄せ技術(ID0)を開発)
- ID0を使って保険診療受診者数を算出してみた

13

NDBによる傷病名特定(患者数推計)の肝

- 傷病名だけではあまり当てにならない(レセプト病名)
- A-1.「疾患特異的な治療法(医療行為や処方)」がある場合は傷病名は比較的正確(例:血友病)。また、A-2.「疾患特異的な検査」も使えるが、検査結果が不明なのでやや妥当性に欠ける(例:HIV-RNA定量)。
- B.「高頻度に受診する疾患」は、集計条件を臨床的に整えることで、病名の正確さが高くなる(例:糖尿病)。
- 上記2条件(A,B)を満たす傷病は、NDBによる集計が向いている(例:人工透析、血友病)。
- ただし、未受診患者はNDBに登場しない→悩ましい¹⁴

**NDBには、ここに書ききれない
種々の落とし穴がたくさんある**

**それらを乗り越えて、
やっと「真のNDB分析」に
たどり着ける**

(逆に言えば、これらの落とし穴が考慮・記載されていない「NDB分析」は何かを見落としているか、何かを無視している可能性が高い。)

混沌から秩序をもたらす試み

NDBの開拓に必要なと思われる人材

- 臨床医学の専門家 (疾患ごとに必要)
- NDBの構造、落とし穴をよく知っている人
- SE (特にデータベース)
- 医療事務 (診療報酬請求の実務に詳しい人)
- 統計家
- 疫学者 (ここでは、医学部で統計ふうのことをしている人の意)
- 行政経験者 (申請書等の作成に慣れた人)
- 上記の職種を取りまとめる人

一人でこなすのは不可能。チームで行う必要性

患者数集計の例 傷病名 医科入院外（日本の外来病名ランキング）

傷病名コード	傷病名名称	レセプト件数	患者数 (ID0)	レセプト件数/ 患者数(ID0)*
4779004	アレルギー性鼻炎	106,762,949	30,054,953	3.6
8833421	高血圧症	228,902,280	25,512,220	9.0
4660009	急性気管支炎	47,630,481	24,603,197	1.9
4659007	急性上気道炎	37,071,096	20,237,394	1.8
2500013	糖尿病	105,341,287	19,900,491	5.3
3672003	近視性乱視	42,264,305	16,598,483	2.5
8840042	便秘症	106,269,671	15,212,012	7.0
2724007	高脂血症	104,817,429	14,456,058	7.3
5351003	慢性胃炎	86,661,038	14,343,636	6.0
3721002	アレルギー性結膜炎	42,731,756	13,809,332	3.1
8840829	腰痛症	82,399,915	13,662,071	6.0
6923002	湿疹	43,399,472	13,063,641	3.3
4939008	気管支喘息	56,951,594	13,013,057	4.4
8839792	不眠症	105,393,279	12,620,240	8.4
4871001	インフルエンザ	13,924,701	11,623,739	1.2
5319009	胃潰瘍	58,882,689	10,512,058	5.6
8830417	胃炎	41,544,847	10,351,768	4.0
8832281	急性咽頭喉頭炎	19,232,780	10,275,300	1.9

* 年間受診月数 × 年間医療機関受診数。単一医療機関受診の場合は年間受診月数

17

患者数集計の例 医薬品 医科入院外（日本の外来処方ランキング）

医薬品コード	医薬品名称	レセプト件数	患者数 (ID0)	レセプト件数/ 患者数(ID0)
620098801	ロキソニン錠60mg	28,336,554	15,020,509	1.9
620002023	カロナール錠200 200mg	17,535,465	12,960,191	1.4
620452801	ムコスタ錠100mg	24,639,738	10,617,336	2.3
610411058	フロモックス錠100mg	11,663,903	8,965,387	1.3
620160501	PL配合顆粒	13,063,028	8,313,050	1.6
620003852	キシロカインゼリー2%	9,719,846	7,229,684	1.3
610407447	ムコダイン錠500mg	12,007,970	7,158,695	1.7
612220325	メジコン錠15mg	8,996,612	6,519,083	1.4
662640418	リンデロン-VG軟膏0.12%	11,270,316	6,468,858	1.7
620009048	ヒルドイドソフト軟膏0.3%	14,262,368	6,460,293	2.2
620004080	メイアクトMS錠100mg	7,718,416	6,012,310	1.3
616140105	クラリス錠200 200mg	8,510,867	6,004,006	1.4
620004961	ガスコンドロップ内用液2%	6,550,426	5,904,975	1.1
620509501	SPTローチ0.25mg「明治」	7,787,076	5,579,791	1.4
620006516	ベストロン耳鼻科用1% 10mg(溶解後の液として)	12,067,948	5,412,615	2.2
621925801	クラビット錠500mg (レボフロキサシンとして)	6,943,977	5,398,605	1.3
620007804	モーラステープL40mg 10cm×14cm	21,126,292	5,396,586	3.9
610453119	ムコダイン錠250mg	9,399,611	5,363,395	1.8
610443040	アレグラ錠60mg	10,443,776	5,350,597	2.0

18

NDBによる血友病患者数集計

19

背景

- 「標準的治療」として多くの疾病にガイドラインが整備されているが、施設間や地域によって実施率が異なることが指摘。
- 血液凝固異常症は薬害エイズ等の歴史的背景もあり、標準的治療の全国均てん化への社会的な要望が強い。

20

目的

- レセプト情報・特定健診等情報データベース (NDB) を用いて、血液凝固異常症の患者数、地域差などを詳細に把握すること
- 従来 of 調査と比較し、結果の信頼性を検討

21

用いたデータ

- 使用データ: NDB
- 期間: 2013年4月 ~ 2014年3月 (1年間)
- 名寄せとして、個人追跡用の匿名個人変数「ID0」を使用

(ID0はID1とID2の一方が途切れた場合に他方で紐付けしつつ、死亡情報や診療報酬の請求遅れなども加味するなどし、名寄せを精緻化したもの)

22

方法

血液凝固異常症の患者として、以下の条件を両方満たす患者を抽出し、集計

- 血友病A・血友病B・von Willebrand病(以下vWD)の病名がついている
- 血液製剤(10品目30剤・別表1)を使用している

いわゆる教師データとして、平成25年(2013年)血液凝固異常症全国調査の患者数と比較

また、集計患者数と各都道府県の男性人口との比を求め、おおまかな有病率を算出

23

結果 血液凝固異常症患者数

	血友病A		血友病B		vWD	
	本研究	既存	本研究	既存	本研究	既存
北海道	155	220	33	49	25	47
東北	306	374	57	76	24	40
関東	1,467	1,635	311	373	139	445
甲信越	154	211	26	32	13	38
北陸	66	77	15	19	***	4
中部	530	559	114	136	55	111
近畿	762	817	150	172	72	236
中国	245	257	42	35	30	50
四国	109	145	17	20	13	59
九州沖縄	533	575	100	122	56	99
男性	3,813	4,835	781	1,020	220	505
女性	45	35	14	14	188	624
合計	3,858	4,870	795	1,034	408	1,129

既存:血液凝固異常症全国調査 *** 患者数10人未満のためマスキング

24

結果 受診地別患者数

受診地別患者数(上位5都道府県)

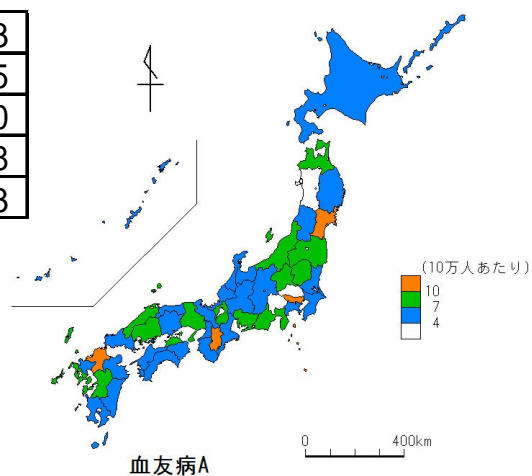
血友病A		血友病B		vWD	
東京都	757	東京都	169	東京都	71
愛知県	308	愛知県	68	愛知県	39
神奈川県	250	神奈川県	49	福岡県	27
福岡県	239	大阪府	48	北海道	25
大阪府	229	福岡県	43	大阪府	21

25

血友病Aの受診地別有病率

上位5都道府県

奈良県	24.3
東京都	11.5
福岡県	10.0
宮城県	9.8
愛知県	8.3

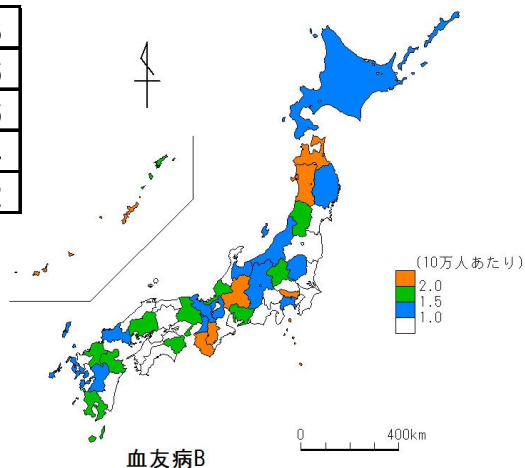


26

血友病Bの受診地別有病率

上位5都道府県

奈良県	3.8
和歌山県	2.6
東京都	2.6
岐阜県	2.4
青森県	2.2

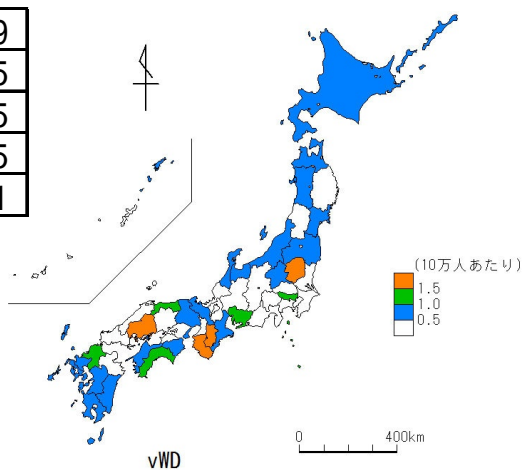


27

vWDの受診地別有病率

上位5都道府県

奈良県	2.9
栃木県	1.5
和歌山県	1.5
広島県	1.5
高知県	1.1



28

考察

- 既存の全国調査は重症～軽症を対象としている。本研究では「血液製剤を使わない軽症者」が含まれない。軽症者が多いとされるvWDでは特に集計数が少ない。本研究の患者数が全国調査より少ない原因の1つと考えられる。
- 本研究では患者数を保険病名で集計しており、疑い病名もカウントしている。そのため、血友病A/血友病B/vWDの重複があり得る。
(= 本当は血友病AなのにBやvWDも病名付与される)

29

考察(つづき)

- 奈良県と東京都で、血友病A/Bの男性人口あたり受診者数が全国平均を大幅に上回っている。また、奈良県は、vWDの男性人口あたり受診者数が全国平均を大幅に上回っている。これらの都県に、血友病の中核的な医療施設があるためと考えられる。

30

全体のまとめ

NDBを適切に用いると、さまざまな疾病の全国患者数推計を短時間で何度でも種々の切り口*で行える！

* 切り口の例として、ある疾患の患者さんがどのような併存疾患(合併症)を有しているか、どのような治療を受けているか、1年間にどれくらい医療費がかかっているかなど。これらの分析が数日の早さで可能となっている。

NDBは落とし穴が多いが、世界最大の健康関連データベースであり、開拓のしがいはある。チームで取り組むための有為な人材*が待ち望まれています。

* さまざまな分野の専門家であり、かつ、チームプレーヤーでもある方

31

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」(研究代表者:今村知明)及び平成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)の一環として実施した。

別表1 抽出条件にした血液製剤一覧

品目名	補充因子	メーカー	製剤数
PPSB	IX	日本製薬	2
アドベイト	VIII	バクスアルタ	4
クリスマシン	IX	日本血液製剤機構	2
クロスエイト	VIII	日本血液製剤機構	3
コージネイト	VIII	バイエル	4
コンファクト	VIII	アステラス	3
ノバクト	IX	アステラス	3
ノボセブン		ノボ	3
ファイバ	VIII	バクスアルタ	2
ベネフィクス	IX	ファイザー	4 ³⁾