

レセプト情報・特定健診等情報データベース (NDB)の活用拡大の取り組み

明神 大也^{1),2)} 野田 龍也¹⁾ 久保 慎一郎¹⁾
西岡 祐一^{1),3)} 今村 知明¹⁾

- 1) 奈良県立医科大学 公衆衛生学講座
- 2) 奈良県立医科大学附属病院 病理診断科
- 3) 奈良県立医科大学 糖尿病学講座

本発表で伝えたいこと

- NDBには日本のほぼ全てのレセプトデータが入っており、いわゆる医療ビッグデータ
- しかし産業界と比べると、データ量はずっと少ない
- ただ普通のPCでは操作できない

- もともとレセプトの構造が複雑
- さらにNDBで匿名化されたり、レセプトにあった情報が抜かれたりすると余計に複雑になる
- しかもクラウドには置けない
- そのためほとんど誰も手付かずだったし、ほぼベンダーロックオン状態だった
- これらを今解決しようとしている

Agenda

- NDBとは？
- NDBの課題
- NDBの課題克服の現状

3

Agenda

- NDBとは？
- NDBの課題
- NDBの課題克服の現状

4

用語説明: レセプトとは?

レセプトとは医療機関が保険者に対して医療費を請求するために発行する請求明細書



- いずれも
- 患者ごと
 - 月ごと
 - 医療機関ごと に1枚(1セット)

紙レセプト(外来)のサンプル

第5回 レセプト情報等の提供に関する事前説明会資料を一部改変

レセプト共通レコード (RE レコード)

医療機関情報レコード (IR レコード)

保険者レコード (HO レコード)

傷病名レコード (SY レコード)

診療行為レコード (SI レコード)

医薬品レコード (IY レコード)

02 ディオバン錠 80 mg 1錠
 ノルバスクOD錠 5mg 1錠
 アペマイド錠 250mg 1錠
 0.5錠 1.9 × 3.5
 03 ジバトス錠 50 mg 2錠 2 × 3.5

回数・日数が同じ薬剤を一つにくっつけて (*印でひとくくり) 点数×日数 で書いている

この明細書は、社会保険診療報酬支払基金が、保険医療機関・保険薬局から提出された電子レセプトについて算出された後、その請求情報に基づき作成したものです。 Ver:000108285372a8465107d0474a974418

6

電子レセプトのデータ (CSV)

いわゆる「生レセ」

```

2,1,0,MN,910000213,東京都港区新橋,1.31E+16,,,,,,,,,,,,,
1,2,0,IR,1,13,1,9999913,,サンプル医科クリニック1,42205,0,,,,,,,,,
1,3,0,RE,23,1118,42204,サンプル 79,1,3120628,,,,,,,,sample-ika-079,,,,,
1,4,0,HO,6132013,1234567,79,1,1619,,,,,,,,,,,,,
1,5,0,SY,2500013,4140619,1,,,1,,,,,,,,,,,,,
1,6,0,SY,5739014,4140619,1,,,,,,,,,,,,,
1,7,0,SY,8833421,4150716,1,,,1,,,,,,,,,,,,,
1,8,0,SI,12,1,112007410,,69,1,,,,,,,,,,,,,
1,9,0,SI,12,1,112011010,,52,1,,,,,,,,,,,,,
1,10,0,SI,13,1,113001810,,225,1,,,,,,,,,,,,,
1,11,0,SI,21,1,120000710,,9,1,,,,,,,,,,,,,
1,12,0,IY,21,1,610443044,1,,,,,,,,,,,,,
1,13,0,IY,,1,620008041,1,,,,,,,,,,,,,
1,14,0,IY,,1,613960041,0.5,19,35,,,,,,,,,,,,,
1,15,0,IY,21,1,620004502,2,2,35,,,,,,,,,,,,,
1,16,0,SI,25,1,120001210,,42,1,,,,,,,,,,,,,
1,17,0,SI,25,1,120003170,,65,1,,,,,,,,,,,,,
1,18,0,SI,27,1,120001810,,8,1,,,,,,,,,,,,,
1,19,0,SI,60,1,160000310,,26,1,,,,,,,,,,,,,
1,20,0,SI,60,1,160010010,,50,1,,,,,,,,,,,,,
1,21,0,SI,60,1,160022510,,,,,,,,,,,,,
1,22,0,SI,,1,160022610,,,,,,,,,,,,,
1,23,0,SI,,1,160020410,,,,,,,,,,,,,
1,24,0,SI,,1,160019410,,56,1,,,,,,,,,,,,,
2,28,1,EX,,,,,,,,,1:42205:00000000000000,,,,,
2,29,0,RC,Ver00001df061727252faec47486f785da58f351,,,,,,,,,,,,,
    
```

電子レセプトのデータ (表形式に変更)

IR	1	13	1	9999913	サンプル医科クリニック1
RE	23	1118	42204	サンプル 79	1
HO	6132013	1234567	79	1	1619
SY	2500013	4140619	1		1
SY	5739014	4140619	1		
SY	8833421	4150716	1		1
SI	12	1	112007410		69
SI	12	1	112011010		52
SI	13	1	113001810		225
SI	21	1	120000710		9
IY	21	1	610443044	1	
IY		1	620008041	1	
IY		1	613960041	0.5	19
IY	21	1	620004502	2	2
SI	25	1	120001210		42
SI	25	1	120003170		65
SI	27	1	120001810		8
SI	60	1	160000310		26
SI	60	1	160010010		50
SI	60	1	160022510		
SI		1	160022610		
SI		1	160020410		
SI		1	160019410		56
SI	60	1	160095710		13

・ディオバン
・ルバスク
は情報が空白
(必要に応じ、按
分による空白補
充が必要)

各数値の詳細等は、「オンライン又は光ディスク等による請求に係る記録条件仕様(医科用)」を参照。

<http://www.iryohoken.go.jp/shinryohoshu/receMenu/doReceInfo>

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）

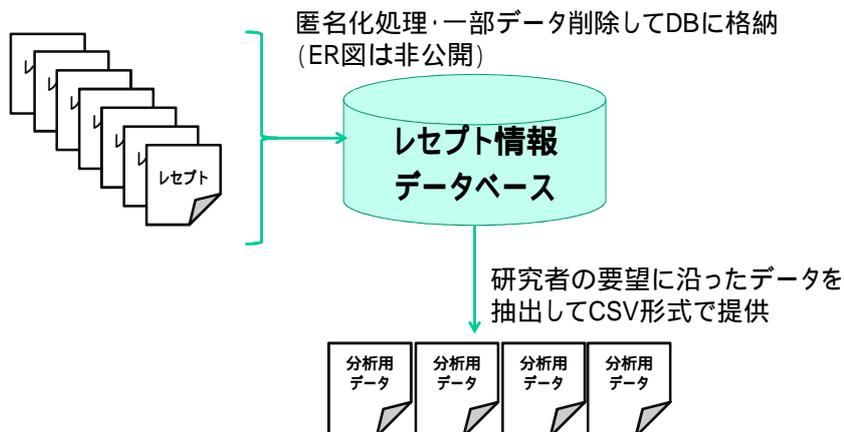
約1億9,800万件を格納

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

9

NDB本体と切り出し方



10

Agenda

- NDBとは？
- NDBの課題
- NDBの課題克服の現状

11

NDBの課題 ~ どうしたらNDBで解析できるか？ ~

- 入手前
 - 申請が複雑
- 入手後
 - データ量が多い
 - データが複雑
 - ・ 生レセプトとNDBデータの違い(特にDPC)
 - ・ 名寄せが困難
 - ・ 入退院情報が把握できない
 - ・ コストの算定が困難

12

NDBの課題～申請が複雑～

第三者提供申請までの流れ



事前相談からデータ提供されるまで、6ヶ月以上を要する。
1年といった単位で探索的な研究開発に取り組むには不向きなデータともいえる。

NDBの提供申請には、希望する提供指定期間の全レセプトは提供できないルールとなっており、提供希望する一部のマスターコード（傷病名、診療行為、医薬品）を申請時に提出する必要がある。

13

NDBの課題～データ量が多い～

- Windows Server 2012 R2 Standard等でCPUはXeon E5-2643v2 3.5GHz等 計10台程度使用
- データ量は1年CSVで3TB程度

処理内容	処理時間
レセプト復元処理(保険別、診療年月別、レセプト種別、都道府県別の単位で分割)	61日
データベースを扱いやすい形式に整える処理(入院と外来に分割、EF化)	83日
DPCLレセプトのデータベース化	9日
医科外来レセプトと調剤レセプトをDPCファイル形式に変換、主傷病を決定する処理	55日

14

NDBの課題～生レセとNDBデータの違い(DPCの例)

(複数月にまたがり特定入院期間(包括期間)を超過し入退院が完結する場合)



(入院期間と出来高期間が混在するN月のレセプトデータ)

NDBでは、総括対象DPCレセプト(SIレコードの包括部分)と総括対象医科入院レセプト(SIレコードの出来高部分)をまとめて、DPCレセプトのSIレコードに取り込んでいる。

上記のN月のケースでは、

「特定入院期間(包括対象)の診療行為」と「出来高期間の診療行為」を分離してデータ抽出することが困難(そもそも気づきにくい)。

NDBの課題～生レセとNDBデータの違い(DPCの例)

(DPCレセプトに存在しないデータ)

基本診療料 入院料等

- (入院基本料)
 - A100 一般病棟入院基本料(1日につき)
 - A104 特定機能病院入院基本料(1日につき)
 - A105 専門病院入院基本料(1日につき)
- (入院基本料等加算)
 - A200 総合入院体制加算(1日につき)
 - A204 地域医療支援病院入院診療加算(入院初日)
 - A204-2 臨床研修病院入院診療加算(入院初日)
- A207 診療録管理体制加算(入院初日)
 - A207-2 医師事務作業補助体制加算(入院初日)
 - A207-3 急性期看護補助体制加算(1日につき)
 - A207-4 看護職員夜間配置加算(1日につき)
- A214 看護補助加算(1日につき)
 - A218 地域加算(1日につき)
 - A218-2 離島加算(1日につき)
- A234 医療安全対策加算(入院初日)
 - A234-2 感染防止対策加算(入院初日)
- A244 病棟薬剤業務実施加算(週1回)
 - A245 データ提出加算(入院中1回)

- 左に示す基本診療料の区分番号はDPCの包括評価項目である。
- これら項目はNDBのDPCレセプトのCDレコードにデータが一切記録されない。
(この部分の診療報酬が把握できない)

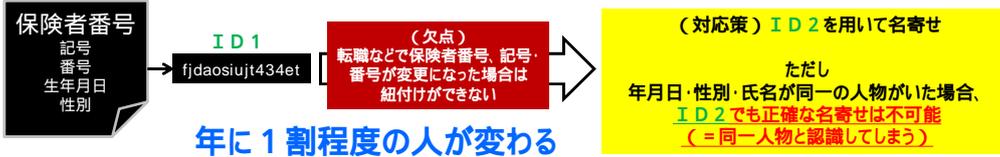
**電子レセプトの仕様により
医療機関出力のレセプトデータに含まれず
結果、NDBにもデータが存在しない。**

DPCレセプトを対象とする研究では、DPC調査データベース(厚生労働省)の「DPC調査データ」が有用である。

NDBの課題 ~ 名寄せが困難 ~

- ある個人を表すIDとして、2つのID値 (= 暗号化された英数字) が付けられている
 - NDBでは、異なる人物が同じID値を持ってしまうことがある (= 名寄せができなくなる) (下記)

(1) ID1・「保険者番号, 中心型ID」記号・番号, 「生年月日, 「性別」からID1を生成させる



(2) ID2・「氏名, 中心型ID」生年月日, 「性別」からID2を生成させる



17

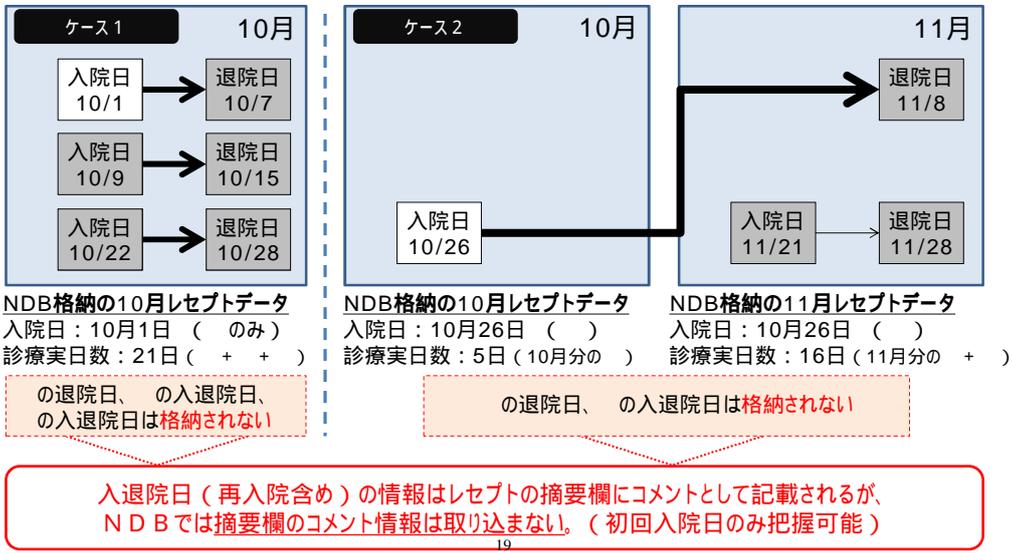
NDBの課題 ~ 名寄せが困難 ~



18

NDBの課題～入退院情報が把握できない～

- 同月内の複数回入退院を把握することができない。



19

NDBの課題～コストの算定が困難～

- レセプトごとの合計点数はわかるが、内訳はすぐにはわからない
- 入院
 - 大きく分けて、医科入院・DPC包括・DPC出来高に分類
 - 麻酔・手術の加算や、DPC包括外の処置・投薬などの処理
 - 医科入院・DPC出来高では処置や投薬の処理を毎回実施
- 外来・調剤
 - 診察日・調剤日ごとに分割
 - 時間外加算の処理

20

Agenda

- NDBとは？
- NDBの課題
- NDBの課題克服の現状

21

NDBの課題～申請が複雑～

第三者提供申請までの流れ



少しずつ慣れて、早くなってきた
それでも半年は必要

ベンダー(ニッセイ情報テクノロジー)が代行

有識者会議

ベンダーが代行

22

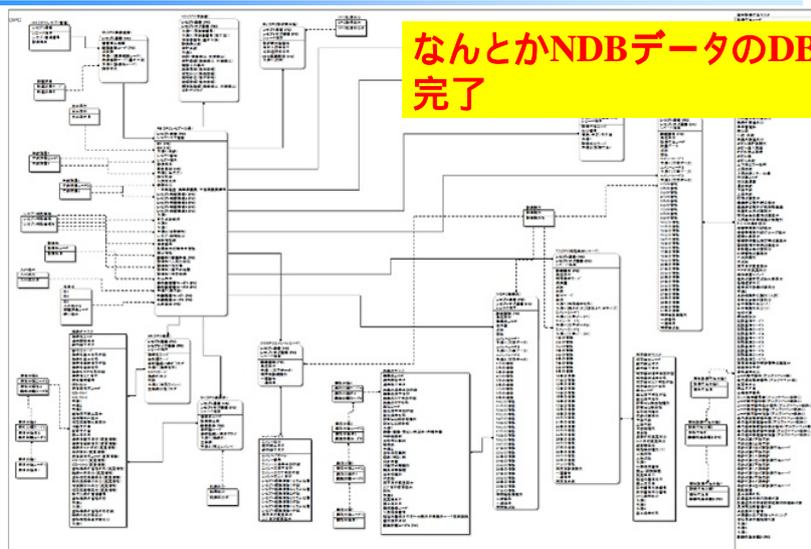
NDBの課題～データ量が多い～

- 並列型DBを用いたマシンに変更
- EF化・主傷病決定処理を行わず、別処理でDB化
- 所要期間はおよそ1か月に短縮

- その後、以下のスペックに変更
 - Windows Server Standard / SQL Server Enterprise
 - CPU: 2CPU 20Core(インテル XeonR E5-2640 v4 2.4GHz)
 - メモリ: 1536GB (64GB LRDIMM x24枚)
 - HDD: 2TB NLSAS x26本 (RAID1/0)
- 3年間のデータを2週間程度で処理(データクリーニングを除く)

23

NDBの課題～生レセとNDBデータの違い(DPCの例)



24

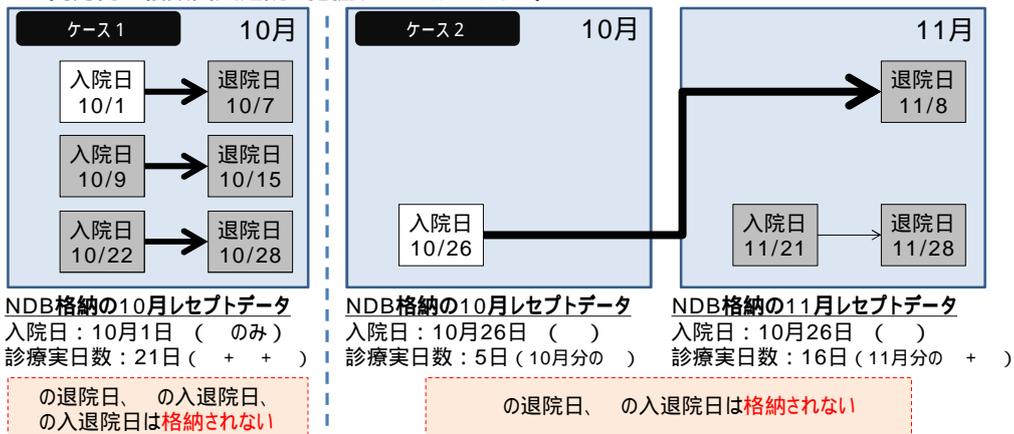
NDBの課題～名寄せが困難～

- NDBでは複数のファイルやレコードに分置された同一人物等のデータを、キー変数を用いて一つに紐付けすることが重要
- 匿名化された2つのIDから「1患者1ID」を振りなおし、患者単位でレセプトを抽出するための**統合IDルール (ID0)**を作成、論文公開
- ID1をベースに、途切れた際にID2を使って追跡したり、医科入院・外来・DPC・調剤のレセを結合したりして、精度を向上させた

25

NDBの課題～入退院情報が把握できない～

- 同月内の複数回入退院を把握することができない。

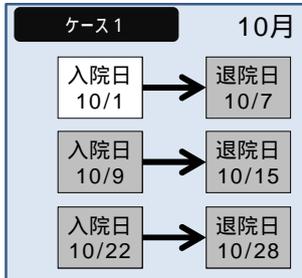


入退院日（再入院含め）の情報はレセプトの摘要欄にコメントとして記載されるが、NDBでは摘要欄のコメント情報は取り込まない。（初回入院日のみ把握可能）

26

NDBの課題～入院情報が把握できない～

- 同月内の複数回入院退院を把握することができない。



1入院期間ごとにデータを分割

- 入院期間の把握
- 入院ごとの処置の把握
(CT撮影や消化管内視鏡など)
- 入院ごとのコストの把握

最近完了

27

NDBの課題～コストの算定が困難～

- レセプトごとの合計点数はわかるが、内訳はすぐにはわからない

■ 入院

- 大きく分けて、医科入院・DPC包括・DPC出来高に分類
- 麻酔・手術の加算や、DPC包括外の処置・投薬などの処理
- 医科入院・DPC出来高では処置や投薬の処理を毎回実施

最近完了

■ 外来（医科入院外）

- 外来診察日ごとに分割
- 時間外加算の処理 **DB化の時点でほぼ完了**

28

まとめ

NDBの分析に必要なスキル一覧

1. 臨床医学
2. 疫学
3. 医療事務
4. NDBの構造
5. 行政経験者
6. 情報システム・データベース

29

まとめ

1. 巨大な請求書の束であるNDBを 一患者一データ化しコホート化を目指している
2. 「巨大すぎて扱いづらい」点は技術的に解決できると思われるが大変！

うまくいけば一億人分の全数超巨大コホート

今 3年分 やっています
NDBは宝の山だということはわかってきた

30