

1月17日(木)
13:00-17:00

資料4

全国レセプトデータ 利用上の注意点

大坪徹也

京都大学 大学院医学研究科 医療経済学分野

レセプトデータの利点

- レセプトデータには次の利点がある^[1]
 - 集団全体へのアプローチが可能
 - 多数の症例数の確保が可能
 - 長期追跡が可能
 - データ調達費用が比較的安い
- ただし、
データの不正確さは結果を歪める可能性
- 本日の話題: どのような点を配慮すべきか

本日の内容

特別抽出によるデータに関して

● データベースの構築

- 受領データのメディアへの格納状況
- サーバへの格納例
- データの精査

● データ利用上の注意事項

- 傷病コード等の入力状況
- 入院単位での集計における注意事項
- 年齢層別集計における注意事項

I . データベースの構築

1. 受領データの格納状況
2. サーバへの格納
3. データの精査

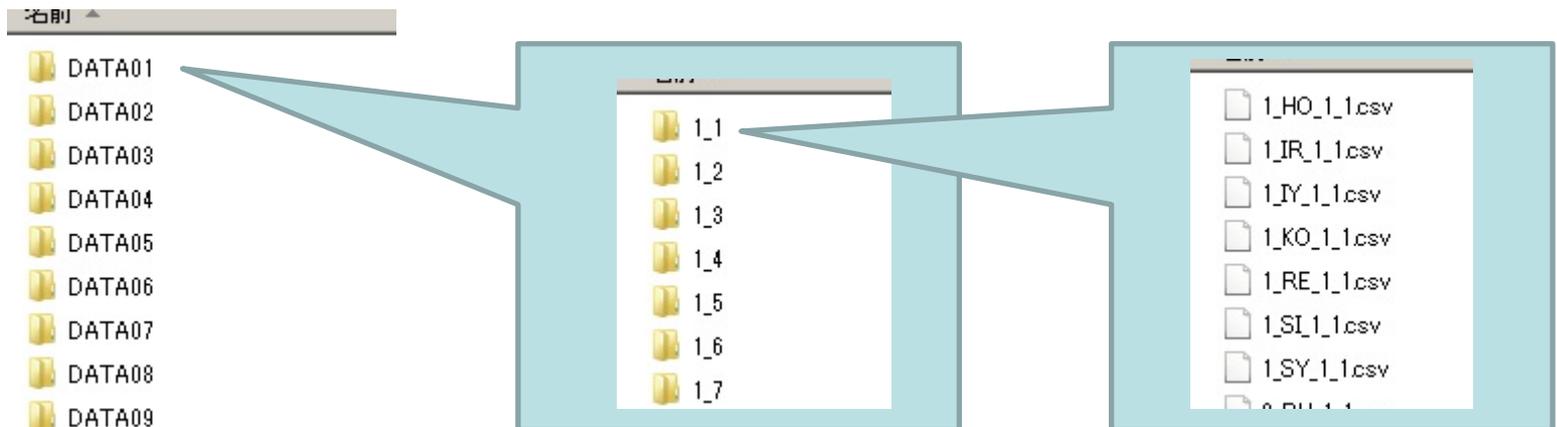
抽出依頼の概要

- 8府県の全被保険者に関する医科・DPCレセプト
- データの対象期間:2年分
- 傷病名・診療行為は対象を限定して抽出
- 抽出対象レコード識別:
 - DPC : IR,RE,HO,KO など計10種
 - 医科 : IR、RE、HO、KOなど計7種

1. 受領データの格納状況

- データは“レコード識別名”別に分割されて格納

上位		下位			
<u>34,816</u> =	<u>32</u> 個のフォルダ	×	<u>64</u> 個のフォルダ	×	<u>17</u> 種のCSV
総CSV数	DATA01~32		1_1~8_8		HO, IR など レコード識別名別



1レセプトに紐づく一連の“レコード識別”は、全て同一の下位フォルダに格納されている

※下線付き数値は、抽出依頼固有の結果である

2. サーバへの格納

- (量の例示として)RE(レセプト共通):
 - 医科 約2億6千8百万レコード
 - DPC 約5百万レコード
- SQL Server のバルクインサート機能を使用
 - 1. 取込み先のDBと同じサーバに, 取込み対象データを保存
 - 2. SQL Server のバルクインサートを実行する

3. データの精査

返戻データのDBへの登録状況

● 登録されるデータ

- 確定レセプトデータ(正常)
- 2009年以前の返戻データ(L3,L4)
 - 入力フォームはOKだが、存在しない摘要コードがあるなど
 - 同一審査月の再請求レセプトデータも登録される

● 登録されないデータ

- 入力フォーム違反に該当したレセプトデータ(L1,L2)
 - 欠測データとなる
- 2010年以降の返戻データ(L3,L4)
 - 確定するまでは当該レセプトのデータは欠測となる

→データ参照期間によって、データベースへの取込ルールが異なるため、参照期間によって重複データや欠測データの発生状況も異なることになる

3. データの精査

● 返戻データの検出

- レセプトを一意に決めるキーが同一のレセプト間で請求年月が異なる場合には、最新の請求年月のものを正とし、過去請求分は返戻とみなす
- 最新の請求年月が同一である再審査データが存在する可能性に注意

3. 重複データのケース

重複データ:

レセプトを一意に決めるキーが同一のデータ群

- ケース1: 入院 保険単独
- ケース2: 入院 公費単独
- ケース3: 入院 保険+一種公費併用と公費単独
- ケース4: 外来

重複データのサンプル

● 保険単独で重複しているパターン

(DPC)	RE				BU		HO		KO		IR	
通番2	レセプト種別	診療年月	個人番号	レセプト 総括区分	入院日	退院日	診療実日数	合計点数	診療実日数	合計点数	医療機関コード	請求年月
DPCA1	家族(保険単独)	42201	AAさん	0	4220107	4220118	12	41400	-	-	医療機関AA	42202
DPCA2	家族(保険単独)	42201	AAさん	0	4220107	4220120	14	47200	-	-	医療機関AA	42202
DPCA3	家族(保険単独)	42201	AAさん	0	4220107	4220118	12	41400	-	-	医療機関AA	42203
DPCA4	家族(保険単独)	42201	AAさん	0	4220107	4220120	14	47200	-	-	医療機関AA	42203

- 診療実日数を求める場合、単純に合算してしまうと、矛盾が発生してしまうため、重複したレセプトを単純に合算することには問題があると思われる。

例) 入院日:H22年1月7日 + 診療実日数計26(12+14)日 ⇒ 退院日:H22年1月33日(?)

※「通番2」とは、1レセプトに対して一意の固有番号であり、各テーブルのデータを連結するためのキー項目になる。

● 公費単独で重複しているパターン

(医科入院)	RE				HO		KO		IR	
通番2	レセプト種別	診療年月	個人番号	入院日	診療実日数	合計点数	診療実日数	合計点数	医療機関コード	請求年月
入院A1	公費単独	42210	Aさん	4221002	-	-	2	3,200	医療機関A	42212
入院A2	公費単独	42210	Aさん	4221002	-	-	2	3,500	医療機関A	42212
入院A3	公費単独	42210	Aさん	4221002	-	-	4	6,400	医療機関A	42212
入院A4	公費単独	42210	Aさん	4221002	-	-	1	1,700	医療機関A	42212

- 公費は適用されたものにだけ計上されるため、どれか一つだけを選んで使用することには問題があると思われる

重複データのサンプル

● 保険+一種公費併用と公費単独で重複しているパターン

(DPC)	RE				BU		HO		KO		IR	
通番2	レセプト種別	診療年月	個人番号	レセプト総括区分	入院日	退院日	診療実日数	合計点数	診療実日数	合計点数	医療機関コード	請求年月
DPCB1	家族(保険1種公費併用)	42104	BBさん	0	4210304	4210403	3	6,000	3	6,000	医療機関BB	42105
DPCB2	家族(保険単独)	42104	BBさん	0	4210304	4210403	3	8,000	-	-	医療機関BB	42105

- 診療実日数を求める場合、単純に合算してしまうと、矛盾が発生してしまうため、重複したレセプトを単純に合算することには問題があると思われる。

例) H22年4月1日 + 診療実日数計6(3+3)日 ⇒ 計算上退院日:H22年4月6日 ≠ 退院日:H22年4月3日

● 外来で重複しているパターン

(外来)	RE				HO	KO	IR	
通番2	レセプト種別	診療年月	個人番号	レセプト番号	診療実日数	診療実日数	医療機関コード	請求年月
外来A3	本人(保険単独)	42105	AAAさん	10	1	-	医療機関AAA	42106
外来A6	本人(保険単独)	42105	AAAさん	10	1	-	医療機関AAA	42106
外来A2	本人(保険単独)	42105	AAAさん	47	1	-	医療機関AAA	42106
外来A5	本人(保険単独)	42105	AAAさん	47	1	-	医療機関AAA	42106
外来A1	本人(保険単独)	42105	AAAさん	95	1	-	医療機関AAA	42106
外来A4	本人(保険単独)	42105	AAAさん	95	1	-	医療機関AAA	42106
外来A10	本人(保険単独)	42105	AAAさん	103	1	-	医療機関AAA	42106
外来A9	本人(保険単独)	42105	AAAさん	103	1	-	医療機関AAA	42106
外来A7	本人(保険単独)	42105	AAAさん	128	1	-	医療機関AAA	42106
外来A8	本人(保険単独)	42105	AAAさん	128	1	-	医療機関AAA	42106

- 医療機関が保険者に対しレセプトを提出する際に、1レセプト毎に「1」から昇順に採番される一連の番号
- 1提出毎に採番される為、返戻時に採番されるレセプト番号は、初回請求時のレセプト番号と一致するとは限らない

Ⅱ. データ利用上の注意事項

1. 傷病コード等の入力状況
2. 入院単位での集計における注意事項
3. 年齢層別集計における注意事項

集計単位

● レセプト件数

- 指定した傷病の診療に該当するレセプトの数

● 入退院件数

● 患者数

- 年齢層別患者数

● 回数

● 日数

- 一入院あたりの在院日数

● 医療費

- 一入院当たりの医療費

1. 傷病コード等の入力状況

- 傷病に関する情報の登録状況
 - 1レセプトには複数の傷病が記載
 - 主傷病フラグが全くないレセプトも存在
 - 例：入院で約3割，外来で約6割のレセプトで主傷病なし
 - 「疑い」は修飾語コードで判別可能
 - NDBでは修飾語コードは文字列で登録
 - 例：204916278002 左鎖骨疑い
 - 未コード化病名は病名欄がないNDBでは現在取扱えない
 - 例えば，病名コードの約5%が該当
- 上記を踏まえ，研究内容に沿った傷病の取扱を
 - 診療行為と組合わせる など

1. 傷病コード等の入力状況

● RE 「診療科コード」

- 平成22年3月まで 医科(個別診療科), DPC(診療科コード)が登録されている
→レセプトを一意に決定する際のキーとして活用

● 調剤レセプトの「医療機関コード」

- 平成22年10月診療年月以降に登録

2. 入院単位での集計における注意事項

入退院に関する情報

● 入院レセプトについて

- 入退院に関連する項目の入力状況は下表の通り

レセプト種別	入院年月日	退院年月日	診療実日数
医科	○	×	○
DPC	○	○ (退院月のみ)	○

● 1 入院で集計する際の注意事項

- ① いずれかの入院月におけるレセプトデータの欠測
- ② 退院年月日の把握
- ③ 同一月複数入院への配慮

①いずれかの入院月におけるレセプトデータの欠測

- 複数月にわたる医科入院においてレセプト欠測がある場合（特に2010年以降）



個人・医療機関・診療科・入院年月日が一致するレセプト群を1入院として扱う

1入院の医療費は、『レセプトが完備された入院かつ同月複数入院なし』のみ評価できる

※退院日は、退院月レセプトより特定可能

※入院日数は、入院日と退院日の差より取得可能

※医療費及び診療行為の日数・回数等は、レセプト欠測月分が過小評価となる

②退院年月日の把握方法

● 医科レセプトの場合

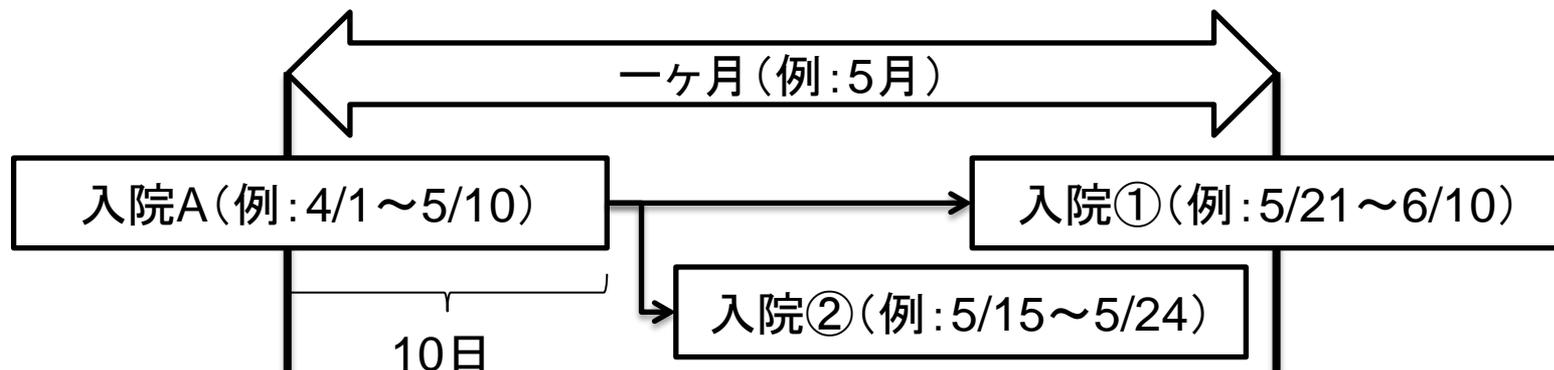
- 入院年月日・診療年月・診療実日数を用いて、退院年月日を算出
- 問題点：算出された退院日が、当該月の末日であった場合、退院したのかが判別できない
 - 次月レセプトが返戻となり、欠測となる可能性があるため
 - コメントデータにより把握可能ではあるが...
 - 翌月のレセプトを参照することで精度向上
 - 必要参照期間の1ヶ月先までのレセプトを申請したほうがよい

● DPCレセプトの場合

- 退院月に退院年月日が明記されている

③同月内複数入院への配慮

● 医科同月内複数入院(2回)



例1) 再入院の退院日が翌月以降の場合(入院①)

・5月レセプト

入院年月日 = 4月1日 (入院A)

診療実日数 = **21日** (入院A10日 + 入院①11日)

・6月レセプト

入院年月日 = 5月21日 (入院①)

例2) 再入院の退院日が同月内の場合(入院②)

・5月レセプト

入院年月日 = 4月1日 (入院A)

診療実日数 = **20日** (入院A10日 + 入院②10日)

医科レセプトで
入院単位の集計を
行なう場合、
退院翌月のレセプト
を確認すること

※「同月内同月内3回以上の入院」や「同一医療機関において、DPC入院を含む医科入院
例) 医科入院→DPC入院→医科入院」も取扱注意

3. 年齢層別の集計に関する注意事項

● 年齢層による人数マルチカウント

- 『診療月時点の年齢による年齢層を用いて、年間の年齢層別集計を行なう場合』に注意
- 同一被保険者が年度内に最大2つの年齢層に分かれる可能性がある
 - 必ず誕生日を迎えることになるため

例) 同一被保険者でレセプトの内容が左表の場合、集計結果は右表となる

診療年月	年齢層	医療費
4月	70～74歳	100
5月	70～74歳	200
6月	75～79歳	400
7月	75～79歳	500



年齢層	人数	医療費
70～74歳	1	300
75～79歳	1	900
計	2	1,200

その他注意事項

- 医療機関情報の変更に関するデータへの配慮
 - 医療機関コードの変更に伴う入院追跡
 - 入院日は同じ日付が登録されるため、個人で名寄せする等により同一入院での追跡は可能
- 患者情報の変更に伴うデータの限界 (対処不可能)
 - 保険者の異動に伴う重複データのマルチカウント
 - 新旧保険者のレセプトデータいずれもDBに登録される
→ 匿名化名寄せIDでは保険者の異動は把握できないため、当該重複データの検出は不可能
 - 改姓や保険者の異動に伴う個人追跡の限界

おわりに

まとめ

- 有用性の認識とともに、データの限界を知ること
 - そのためには、データ登録に関するクセを知る必要
- データ申請時には、データの精査を見据えた“抽出条件書”の作成が必要
- 集計単位によって注意点が異なる
 - 同一レセプトのマルチカウントに注意
 - 2009年のデータには返戻データも含まれる
 - 入院単位の集計では、欠測や再入院に注意
 - 集計対象(日数, 医療費など)に応じた除外基準を設定
 - 同一人物のマルチカウントに注意
 - 受診時の年齢層を用いた, 年齢層別の年間集計など