

一般社団法人 日本薬剤疫学会 薬剤疫学とデータベースタスクフォース
実務者のためのデータベース研究講座 その2



2019年7月10日 13:30-17:30

NDB を用いた研究の実際(2):
特別抽出を用いた研究について

NPO 日本医薬品安全性研究ユニット
久保田潔



2019/7/12

NPO DSRU Japan

セミナーの趣意書から

- 「生データから「解析用データセット」を作成するまでのプロセスで留意すべき事項やコツなど、実装に活かせる情報を提供する」
 - － NDB特別抽出に関しては「実装に活かせる」のは当面、アカデミア・行政などに限られる・・

本発表の背景・目指すところ

- 「乾癬の疫学研究」承認番号0004(東大、久保田潔)*と
- 「レセプト情報を用いた抗凝固薬の使用実態と使用実態が抗凝固薬の効果に与える影響に関する研究」承認番号: 0118(日大、大場延浩)*

*「第三者提供の現状について(報告)(成果物集計・承諾案件)

<https://www.mhlw.go.jp/content/12401000/000520125.pdf> 参照

の利用者としての経験で感じたことを基に

- NDB(特別抽出)の薬剤疫学研究への利用で
 - どのように「解析用データセット」を作成していくか
 - 「コツ」を含めて実務に役立つ情報を提供

疾患の疫学と薬剤疫学研究では解決すべき問題が相当異なる。

AD

Analysis Dataset: 解析用データセット

以下の2つに区分(正式名称ではありません)

PLAD

Patient Level Analysis Dataset: 1患者1レコード

MRAD

Multi-Record Analysis Dataset: 1患者複数レコード

CDISCでは？

- CDISC (FDAが求める臨床試験データの基準) における解析用データセットとは
 - ▲ Analysis-readyである
 - ▲ メタデータと関連づけられている
 - 以下を含む
 - ADSL: Subject-Level Analysis Dataset
 - ▲ 1患者1レコード
 - BDS: Basic Data Structure
 - ▲ 1患者複数の各時点ごとのデータ
 - その他
- 薬剤疫学研究における解析でも役に立つ・・・
 - PLAD←ADSLに対応
 - MRAD←BDSに対応

NDBによる薬剤疫学研究

- 薬とイベント発生との関係を研究するコホート研究を想定
- どんな「解析用データセット」が必要か？
- 「解析用データセット」を作成していく具体的な手順は？
 - － 様々ありうる
 - － 個人的な「流儀」を参考までに紹介します
 - － SASと「あまり強力ではないマシン」使用を前提

作成されるべきPLAD

1患者1レコードのデータ

- 含まれるべき変数(の一部)
 - ID
 - Sex
 - 生年月日(DOB)またはAge
 - 死亡日(DOD)
 - 観察開始日
 - ▲ New User Designなら薬の使用開始日
 - 観察(最終)終了日
 - ▲ 薬の中断などで打ち切りとする場合(条件により打ち切り日は複数)でも、「その日以後は観察対象になりえない日」
(例:異なる類薬へのswitchの日)は通常定義可能
 - 曝露:コホート研究の観察開始日における薬

作成されるべきMRAD:

1患者複数レコード

- 薬(曝露)の処方日を単位とするAD
 - ー 複数イベントの解析(感度解析を含む)で対象のイベントを一つに絞りこめない(絞りこまない方がよい)場合で、イベントごとに累積処方期間、累積投与量などを推定する必要がある場合
- イベント発生日を単位とするAD
 - ー 感度解析を含む複数の解析を予定していて、単一のイベント発生日を一意に指定できない(しない方がよい)などの場合
 - ー 観察開始日が(デザインによりマッチングされるなど)動的に決まる場合には、観察開始日が特定されるまで「開始前のイベント」と「開始後のイベント」を識別できないなど
- 交絡因子のうちいくつか: データ取得日を単位とするAD
 - ー 交絡因子取得の起点の日が研究デザイン上変化するなどの理由で動的に情報を取得できるようにしておいた方がよい場合

準備作業

提供されるNDBデータ

- 「多数の」フォルダ構造に含まれる「多数の」csvファイル
- 数年前に比べて大分整理されてきた

DPC
MED
PHA

201504
201505
201506
201507
201508
201509
201510
201511
201512
201601
201602
201603
201604
201605
201606
201607

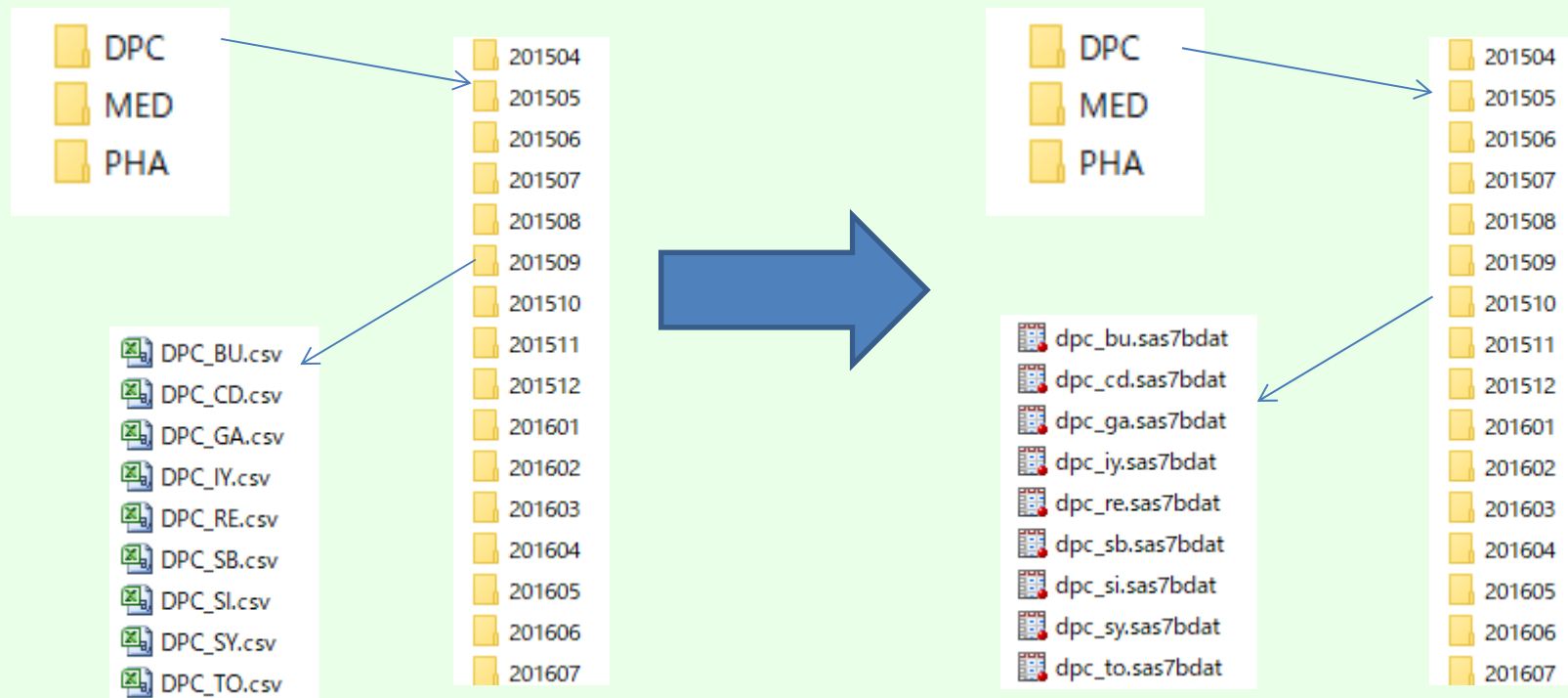
DPC_BU.csv
DPC_CD.csv
DPC_GA.csv
DPC_IV.csv
DPC_RE.csv
DPC_SB.csv
DPC_SI.csv
DPC_SY.csv
DPC_TO.csv

準備作業

CSVをSASデータセットに変換

● Originalと同一構造の空フォルダへ

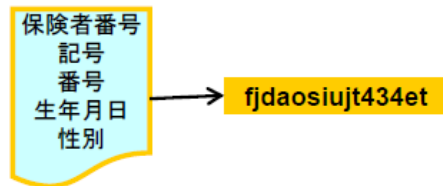
- E:¥A¥DPC_201504¥DPC_RE.csvを読み込みSASデータセット
- E:¥B¥DPC_201504¥DPC_RE.sas7bdat に変換・・・繰り返し



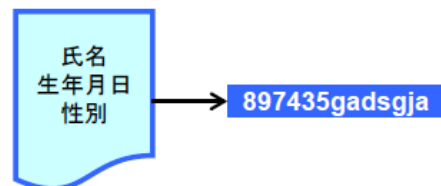
患者ID(newID)の生成

- 乾癬の疫学(BMJ Open 2015)で用いた方法
 - 「ID1(ハッシュ値1)とID2(ハッシュ値2)のいずれかが同一であれば同一人とみなす」(newIDを作成)
 - “We identified 565 903 patients – from a total of 681 827 ID1s and 778 767 ID2s assuming that identical ID1 and ID2 pairs represented the same patient.” (BMJ Open 2015;5:e006450).
 - ID1は**17.0%**、ID2は**27.3%**圧縮

① 保険者番号・記号番号・生年月日・性別からハッシュ値①を生成させる。



② 氏名・生年月日・性別からハッシュ値②を生成させる。



ID1とID2の問題については厚労省の最近(2018年)の資料

(<https://www.mhlw.go.jp/content/12401000/000350567.pdf>) でも言及されている

野田らのID0

厚生の指標 第64巻2017年10月:7-13

- ID0:「・・平成25年4月～平成26年3月の・・(各種)レセプト全体を対象とした」
- 「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、ID1による患者数を男性で**6.2%**、女性で**7.1%**、圧縮することができた。」
 - newIDによる圧縮(17.0%)の半分以下(扱うレセ発行年が2年異なること、対象集団が異なることも関連?)
- 「ID1により名寄せされた患者数は、小児や高齢者、若年女性において推計人口を大きく上回っていたが、ID0により名寄せされた患者数はおおむね推計人口の範囲内に収まった。」
- 「(年齢層によっては)・・受診率が100%を超えており、依然、(ID0は)名寄せが十全でないことを示している。」

ID0: ID1を信頼

- 「・・・ID1が途切れた場合は、途切れの前後で同一のID2を・・・探索して同一人物とした」
 - － 「探索対象に候補となるID2が2つ以上見つかった場合は、別人物が紐付けられてしまう事態を避けるため・・・終了した。」
- 「ID1が途切れた年月以降のID2を・・・探索・・・次に前月の1ヵ月間で・・・最後に前々月の1ヵ月間で同じID2を・・・探索し名寄せを行った。」
- 「ID0による名寄せは、第一種過誤を発生させないことを原則の1つとしており、同一人物の可能性が高いレセプトであっても、IDによる照合が一致しない限りは別人として扱っている。」
 - － 異なる人が同一人と認識されないことを優先させ、同一人が異なる人と認識される可能性はある程度許容(久保田コメント)

ID1

保険者番号
記号
番号
生年月日
性別

ID2

氏名
生年月日
性別

ID0とnewIDの違い

- ID0:「・・・平成25年4月～平成26年3月の・・・(各種)レセプト全体を対象とした」
 - － 全レセプトを対象!!
- newID: 特定条件(特定の傷病名(乾癬)、薬剤(抗凝固薬))の患者のID1とID2から生成。
- “Overmatching”(ここでは「異なる人を同一人とみなすこと」と定義)の確率は
おそらく[①ID2が同一(同姓同名同一生年月日)+②「同一の疾患/薬剤使用」]の場合(newID)、
[①だけ]の場合(ID0)より小さい。
- ID0はID1が途切れた時点前後のID2(“近傍ID2”)を利用するが、「特別抽出」では“近傍ID2”が、提供されるデータに含まれていない可能性がある。
 - － 特別抽出データからのID0生成は困難?

“Zombie Claims”と“Overmatching”

- “Zombie Claims”: 「転帰死亡」のレセの後の年月に発行されるレセ
 - 2013年のJMDCデータの研究では3.1% (13/413) (95%CI: 1.7-5.3%)
 - “Of those (413), 13 had ‘zombie’ claims during 1 to 36 months following the index claim indicating death”

Ooba N etc (2013) Claims-Based Definition of Death in Japanese Claims Database: Validity and Implications. PLoS ONE 8(5): e66116.

ID1とID2の名寄せの評価

- 理想は氏名、記号・番号をゴールドスタンダードとするバリデーション研究(実施はおそらく困難)。
- 「特別抽出」では一般人口の性・年齢分布との比較は役立たない。
- “Zombie Claims”の割合(3%程度)による評価が有用？
 - ー 同姓同名同一生年月日だが異なる人が同年月に死亡することはマレ(結果“Zombie Claims”発生につながる)と考えられる。
- “Zombie Claims”<3%なら「ID2が同一なら同一人」でよい？
- いずれにしても、
「よりよいID」(世帯内の個人の識別、医療のマイナンバー)が近く登場する(ようだ)が、ID1とID2の名寄せ問題は(過去のデータに関しては)今後相当期間続く。

どのような名寄せが必要か？

- 研究の目的にもよる？
- 疫学研究では「誤った名寄せをしない」こと自体より「結果に与える影響」が重要
 - － 同一人を異なる人とすることが与える影響 VS
 - － 異なる人を同一人とすることが与える影響

AD作成プロセス

newIDの生成と全ファイルへの付与

- 全フォルダのレセプト共通レコード("RE")のID1とID2の組み合わせを集め、単一ファイルを作成。
- newID生成: "sort"を何度かかけて、ID1かID2のいずれかが一致するpairsに同一番号(newID)。
- ID1-ID2-newIDリストを各フォルダ内の
"RE"の「レセプト通番」(claimN)-ID1-ID2の組み合わせと照らしあわせて、(そのフォルダ内の)
claimNにnewIDを付与。
- 各フォルダ内の全ファイル(claimNを必ずもつ)に、
claimNに対応するnewIDを付与する。

コツ1

サンプルデータの抽出

- 一定割合（例：5～20％）のnewIDをランダムサンプルし、サンプルデータを同一構造の空フォルダ構造に保存
- 解析コードのバグ発見/処理時間を短縮
 - E:¥B¥DPC_201504 ¥ DPC_RE.sas7bdatから5～20％のサンプルnewIDのレコードを抽出
 - E:¥C¥DPC_201504 ¥ DPC_RE.sas7bdatとして保存

201504
201505
201506
201507
201508
201509
201510
201511
201512
201601
201602
201603
201604
201605
201606
201607

dpc_bu.sas7bdat
dpc_cd.sas7bdat
dpc_ga.sas7bdat
dpc_iy.sas7bdat
dpc_re.sas7bdat
dpc_sb.sas7bdat
dpc_si.sas7bdat
dpc_sy.sas7bdat
dpc_to.sas7bdat

201504
201505
201506
201507
201508
201509
201510
201511
201512
201601
201602
201603
201604
201605
201606
201607

dpc_bu.sas7bdat
dpc_cd.sas7bdat
dpc_ga.sas7bdat
dpc_iy.sas7bdat
dpc_re.sas7bdat
dpc_sb.sas7bdat
dpc_si.sas7bdat
dpc_sy.sas7bdat
dpc_to.sas7bdat

サンプルを使って適切に作動することを確認した後
解析コードの"E:¥C"を"E:¥B"に書き換える。

コツ2

大きなファイルを作らない

- マシンの「力量」も考慮し、なるべく大きなファイルを作らない方針をとった

```
%macro abcd(ym);  
libname sasdp "E%B%DPC%&ym";  
data d1; set sasdp.dp_cd;
```

....

必要最小限の情報を抽出

....

```
%if &ym=201504 %then %do; data d3; set d2; %end;
```

```
%else %do; data d3; set d3 d2; %end;run;
```

```
%mend;
```

```
%abcd(201504); %abcd(201505); %abcd(201506); run;
```

「各フォルダから抽出した情報のまとめ(d3)を最後に保存」など

生年月日 (DOB)

- もともとの電子レセプトの「共通レコード」(RE)には7桁の生年月日が含まれる
- NDBではレセプト発行年月における年齢(多くは「5歳間隔、85歳以上は一括」)
- DOBの暦区間: 5年よりは狭められることが多い
 - － 例: 2015年4月に年齢区分9(40-44歳)だが、2015年5月に年齢区分10(45-49歳)なら、2015年の4月か5月に45歳の誕生日

死亡日 (DOD)

- 先行研究 (PLoS ONE 2013; 8(5): e66116) ではレセプトで特定される死亡の感度は61.6%
- DODについては、年月まで特定可能
 - 医科レセのSY(傷病名)の転帰区分「3」
 - DPCレセのBU(診断群分類)のDPC転帰区分「6」
 - 医科・DPCレセプトのSI(診療行為)
 - ▲ 死亡診断加算(コードとしては3種類)

PLADとMRAD

医薬品に関する情報

- 以下で必要
 - － 観察開始日 (PLAD)
 - ▲ New User Designなら薬の開始日
 - － 観察終了日 (の一要素) (PLAD)
 - ▲ 例: 同系統の異なる薬へのswitchで(常に)打ち切りとする場合
 - － 処方日と処方開始日からの累積投与量などの情報 (MRAD)
 - ▲ 複数のイベントを検討するときに、あるイベント発生日までの累積投与日数や累積投与量の計算
 - ▲ 動的に交絡因子の情報を取得するときに、「特定の日に先立つ●ヵ月の薬の使用」
- 準備作業が必要

準備作業

医薬品に関する準備作業

- 形式の統一
 - ▲ 1処方日1レコードの形式へ
 - － 調剤レセプト: CZとIY
 - － 医科・DPCレセプト: IY
 - ▲ 2012年4月以後は「1日の情報」・・・「31日の情報」
 - － DPCのCD
- ダブリ情報の整理
 - － DPCのCDとIY
- 同処方日同一薬の処理
- 前回処方日+処方日数 > 今回処方日の処理
 - － 主たる曝露である医薬品について(のみ)

調剤レセプトの医薬品



CZ

IY

調剤報酬明細書										都道府 薬局コード	
令和 年 月 分 13										県番号	
公費負担者番号 (1)										公費負担者番号 (2)	
公費負担者番号 (3)										公費負担者番号 (4)	
氏名										特記事項	
1男 2女 1男 2大 3男 4平 5令										1 職務上 2 下船後3月以内 3 通勤災害	
保険医機関の所在地及び名称										保険医氏名	
都道府 県番号										点数表番号	
医師番号										処方月日	
処方番号										医薬品名・規格・用量・剤形・用法	
単位薬剤料										点	

- CZ: 処方月日、調剤月日、調剤数量(通常IYの「回数」に相当)
- IY: 医薬品コード、使用量(1日量)
- CZとIYを重ねて(SASでは"set")「レコード順序」順に並べ替える
- CZの調剤年月日、調剤数量(日数)をその下にあるIYに書き写す(SASでは"retain"を使う)

- 「処方番号」でCZとIYを結合(SQLのinner joint)することも可能。
- 「レコード順序」「処方番号」:もともとの電子レセには存在しない情報。NDB特別抽出作成時に新たに付与されると考えられる。

準備作業

医科・DPCレセプトYの「日」の処理

- 研究の「曝露」である場合には丁寧な処理が必要
 - ー 2012年3月まで:「日」に関する情報ないので、「月の15日」に全てが処方されたとして扱う

項 目	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(14)	(15)	(16)			(43)	(44)
	レコード識別情報	診療識別	負担区分	医薬品コード	使用量	点数	回数	コメントコード	1日の情報	2日の情報	3日の情報	～		30日の情報	31日の情報
モード	英数	数字	英数	数字	数字	数字	数字	数字	数字	数字	数字			数字	数字
最大 バイト数	2	2	1	9	11	7	3	9	3	3	3			3	3
項目形式	固定	可変	固定	固定	可変	可変	可変	可	可変	可変	可変			可変	可変
記録必須	※		※	※			※								

算定日情報の
1日の情報から
31日の情報の
合計値は、回数
と一致する。

(14)～(44) 算定日情報1日の情報～算定日情報31日の情報
ア 回数を記録します。

- 2012年4月以後:1処方日1レコードの形式に変換
- (14)から(44)の合計が(7)と一致することを確認する
 - ー 一致しない場合、「(7)回数」-「(1-31日の)合計」を最後の日に追加

準備作業

DPCレセプトのCDの薬の情報

- もともと1日1レコードの形式
- CD(coding data)には、薬、診療行為、特定器材のデータが記載される
 - － 薬のデータ("6"から始まる9桁コード)を抽出

準備作業

ダブリ情報の整理

- DPCのCDの一部がDPCのIYとダブっていることがある

- 包括対象外の(出来高請求の)術中の薬剤、退院時処方など*(IYに記載)がCDに残っている(残す?)。

*詳細は「診療報酬の算定方法の一部改正に伴う実施上の留意事項について(通知)」参照

- DPCの医薬品のうちCDとIYでダブっていればIY(あるいはCD?)から削除

- 2012年3月までについては、同月の同一レセプトのCDとIYでダブっている薬が対象

- 2012年4月以後については、同一レセプトの、同日においてCDとIYでダブっている薬が対象

準備作業

同処方日同一薬の処理

- 基本は「ケースバイケース」

- 以下を参照に処理の方法を決定

- ▲そもそも、どのような使われ方をする薬か？

- ▲データ上どのような特徴があるか？

- 各処方日1レコードに整理する(とよいことが多い?)。

- ▲次の作業(「前回処方日+処方日数>今回処方日」)が必要な時は、その前に実施しておいた方がよい

アムロジピン 7.5mg 1X 28日

レセ上は2レコード

アムロジピン 5mg錠 28日

アムロジピン 2.5mg錠 28日



「7.5mg 28日」の1レコードに整理

メトレキサート 2mg 1X 4日
(土曜夕)

メトレキサート 4mg 2X 4日
(日曜朝夕)

レセ上も2レコード

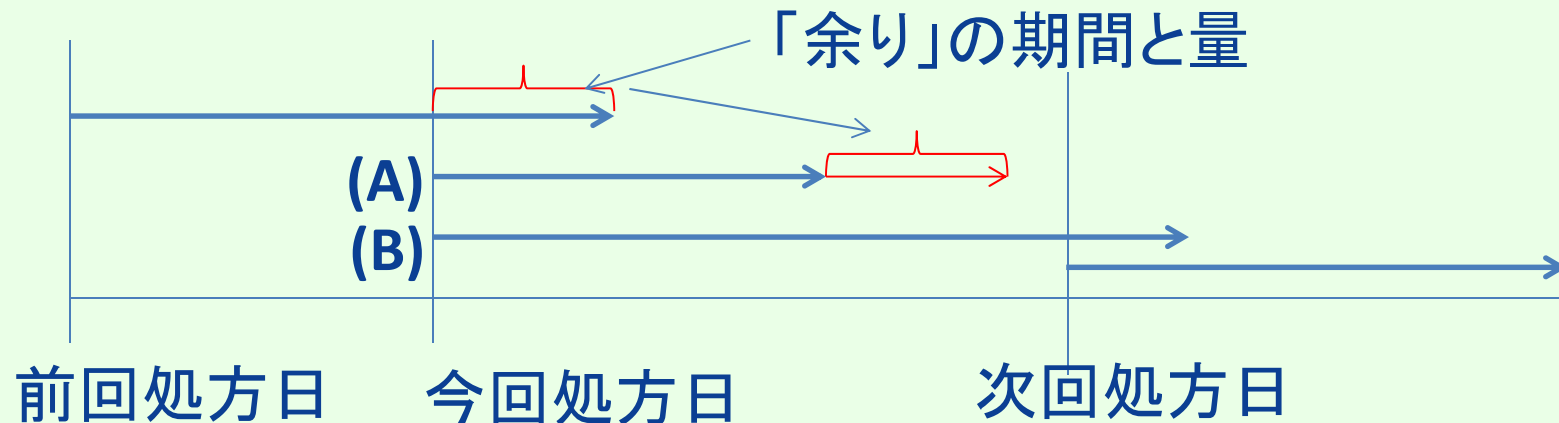


「6mg 4週」の1レコードに整理？

MRAD(曝露情報)



前回処方日+処方日数>今回処方日



毎日長期に使用する薬では問題となりうる

- アプローチ1:「余り」の期間と量を(常に)捨てる
- アプローチ2:「余り」を次の処方に「足す」

データ処理が比較的容易な方法(ほかにもありうる)

- (A)の場合: 今回処方日+処方日数<次回処方日の時には「余り」の期間と量を足す(足すと次回処方日を超えるときには超える分は捨てる)
- (B)の場合: 今回処方日+処方日数>次回処方日の時には前回の「余り」は捨てる



準備作業

診療行為(SI)に関する準備作業

- 医科・DPCレセのSI
 - 「1日の情報」・・・「31日の情報」
 - IYと同様の処理で1日1レコードの形式に変更
- DPCのCDとSIのダブリ処理
 - CDから診療行為のデータ("1"から始まる9桁コード)を抽出
 - IYと同様、CDにあるレコードはSIから削除

MRAD(アウトカム)

「・・・による入院」

- 急性疾患の「発生」に関して薬剤疫学研究でしばしば使用される定義。
- 入院日ごとの1患者複数レコードのMRADとすることが感度解析などで有用
 - DPC/非DPC入院
 - 傷病名の属性(傷病名区分、主傷病など)が異なる
 - 治療開始前/後を区別する必要
 - マッチング時点以前/以後の入院を区別する必要
- 長期入院では同一入院日のレセプトが複数月にわたって発行されることに留意

「・・・による入院」のアウトカム バリデーション研究における”claims-based definition”に含まれうる要素

- 傷病名区分(DPCレセプト)
- 主傷病(医科/DPCのSY)
- 主傷病決定フラグ(医科レセプト)
- 「疑い」(SY・SBの「修飾語コード」8002)
 - － 最近のNDBでは「疑い」フラグとして「再掲」
- 診療開始日(入院日との前後関係)
- その他
 - － 検査、治療
 - － 傷病名の特定期間内の出現頻度

傷病名区分

- DPCレセプト: 診断群分類 (BU) に紐づく ICD10 レベルの傷病名 (SB)
 - ― 「傷病名区分コード」をもつ
 - ▲ 1、11、21 は必須。
 - ▲ 多く同一の傷病名

別表 2 3 傷病名区分コード

コード名	コード	内 容
傷 病 名 区 分 コー ド	0 1	傷病名 (医療資源を最も投入した傷病名)
	0 2	副傷病名
	1 1	主傷病名
	2 1	入院の契機となった傷病名
	3 1	医療資源を 2 番目に投入した傷病名
	4 1	入院時併存傷病名 (1)
	4 2	入院時併存傷病名 (2)
	4 3	入院時併存傷病名 (3)
	4 4	入院時併存傷病名 (4)
	5 1	入院後発症傷病名 (1)
	5 2	入院後発症傷病名 (2)
	5 3	入院後発症傷病名 (3)
	5 4	入院後発症傷病名 (4)

1、11、21の重要性

Positive predictive value of ICD-10 codes for acute myocardial infarction in Japan: a validation study at a single center

Takashi Ando¹, Nobuhiro Ooba², Mayumi Mochizuki¹, Daisuke Koide³, Koichi Kimura⁴, Seitz L. Lee⁵, Soko Setoguchi⁶ and Kiyoshi Kubota^{7*} 

Ando et al. *BMC Health Services Research* (2018) 18:895
<https://doi.org/10.1186/s12913-018-3727-0>

Table 3 PPV of condition codes for AMI in DPC and non-DPC claims

Definition	N	True AMI	PPV (95% confidence interval) ^a
Condition code for AMI in DPC or non-DPC claims	200	165	82.5 (76.5–87.5)%
Condition code for AMI in DPC claims	178	159	89.3 (83.8–93.4)%
Condition code for AMI in non-DPC claims only	22	6	27.3 (10.7–50.2)%
Condition code for AMI in either of type B-1, B-2 or B-3 in DPC claims ^b	161	151	93.8 (88.9–97.0)%
Condition code for AMI in either of type B-5, B-6 or B-7 but not in B-1 B-2 or B-3 in DPC claims ^{bc}	17	8	47.1 (23.0–72.2)%

- B-1、B-2、B-3: 傷病名区分の1、11、21

主傷病決定フラグ

- 「・・・による入院」の発生はDPC病院以外でも起こる
 - － 「よいアウトカム指標」はDPCレセと非DPCレセで異なりうる？
- 医科レセの「主傷病」: 項目としてはあるが・・・
 - － まったく指定されていないこともある
 - － 複数の傷病名に指定されていることもある
- 「主傷病決定フラグ」
 - － 主傷病フラグと「診療開始日」をもとにしたもの
 - － 最近のNDBには含まれている

【改善点2】疑い病名・主傷病決定フラグを導入する

＜旧システム＞

疑い病名の判定

SYレコードの項目「修飾語コード」に疑い病名コード"8002"という文字が含まれていた場合、該当レコードが疑い病名となる。

主傷病の判定

一連の医療行為の中で主傷病区分=01となるレコードが複数存在する場合があるため、どれが主傷病か判断できない。

＜新システム＞

項目「疑い病名フラグ」を導入し、疑い病名の場合に1を設定する。

項目「主傷病決定フラグ」を導入し、主傷病の場合に1を設定する。

平成27年12月9日	資料2
第26回レセプト情報等の提供に関する有識者会議	

準備作業

DPCレセの入院日

- DPCの傷病名
 - SB: 1BU(診断群分類)に対し与えられる傷病名
(傷病名区分で区別された複数の傷病名)
 - SY: 総括対象医科レセプトにおける傷病名
- DPCの入院日
 - REの「入院年月日」: 総括対象医科入院レセプトの場合(傷病名はSY)。それ以外は通常空白
 - BUの「今回入院年月日」: DPCレセプトまたは総括対象DPCレセプトの場合(傷病名はSB)

準備作業と留意点

DPCレセの入院日と「転帰」

- SBの傷病名の入院日と「DPC転帰区分」
 - － RE、SB自体ではなく(同一レセの)BUに含まれる
 - ▲ SYでは入院日はREに、「転帰区分」はSY自体に含まれる
 - － NDB利用申請にあたっては、傷病名を指定するだけではなく、「SBの場合は・・・」
 - ▲ 「同一レセのBUの情報(全て)を含めてほしい」または(せめて)
 - ▲ 「同一レセのBUの「今回入院年月日」「DPC転帰区分」を含めてほしい」旨を指定すべきことに留意。
- 同一DPCレセに複数のBUがある場合
 - － BUとSBを重ね「レコード順序」でソート
 - － SBの直上にあるBUの「今回入院年月日」「DPC転帰区分」がそのSBに関する情報

MRAD(交絡因子)

交絡因子情報

- ▲現段階で演者はこの点については検討未完了
 - －それほど詳細な情報は不要？
- 医薬品のうち該当するもの、レセ年月ごと
- 傷病名のうち該当するもの、レセ年月ごと
 - －交絡因子情報の場合「診療開始日」「今回入院年月日」ではなく、レセ年月を単位
 - ▲「いつ医療サービスを開始したか、いつ入院したか？」よりも「ある時点より前●ヶ月に受診/入院しているか？」がより重要なことが多い。
 - ▲「診療開始日」「今回入院年月日」はMRADのフィールドに含める

まとめに変えて

- 解析用データセット
 - CDISCに倣い、PLAD(1患者1レコード)、複数のMRAD(1患者複数レコード)に区分することが有用
 - 準備作業が必要
 - ▲ 現段階では、どのような準備作業をどのように進めるかに関する探索的な調査(謎解き)から開始する必要
 - ▲ ノウハウの共有が必要(レセプトデータベースベンダーの協力、on-siteにおいて経験されたことの共有にも期待)
 - ID1とID2の名寄せの問題
 - ▲ 特に「特別抽出」では“Zombie Claims”の割合の評価などで適切なら「ID2が同一なら同一人」としてよいのでは？
- コツ: “なるべく身軽に!!” (マシンの性能等にも依存?)
 - 5~20%サンプルデータで解析プログラムコードを検討
 - 大きなファイルを避ける

ご清聴 ありがとうございました。

NDB関連の最近の情報を提供いただいた大日本住友製薬の兼山達也様、
CDとIYのダブリのメカニズムについてご教示をいただいた東京薬科大学の
今井志乃ぶ先生とJMDC社木村真也会長に感謝します。