

平成25年6月26日  
レセプト情報等の提供に関する事前説明会資料

# レセプト情報を利用した研究について

名古屋市立大学大学院薬学研究科  
頭金正博



# レセプトデータの構造

# レセプトの構造説明

## レコード名

## レコード内容

### 医科レセプト

REレコード(レセプト共通)

通番1、通番2、年齢、性別、ハッシュ値...

IRレコード(医療機関)

通番1、通番2、都道府県...

SYレコード(傷病名)

通番1、通番2、**傷病名**

SILレコード(診療行為)

通番1、通番2、検査、**診療行為**...

IYレコード(医薬品)

通番1、通番2、**医薬品**、使用量、回数...

### 調剤レセプト

REレコード(レセプト共通)

通番1、通番2、年齢、性別、ハッシュ値...

YKレコード(医療機関)

通番1、通番2、都道府県...

IYレコード(医薬品)

通番1、通番2、**医薬品**...

一枚のレセプトはレコードと呼ばれるユニットによって構成されている。レコードは左記のように機能的に分かれている。また、赤字部分は**コードと呼ばれる半角数字の羅列で記録されている**。(次ページで説明)

# コードとは

全角文字が入るとデータ量が多くなってしまいうため、基本的に全ての情報は半角の数字(コード)の状態記録されている。

各診療行為、医薬品、傷病名等に該当するコードの一覧がマスター表である。マスター表は医科診療行為、医薬品、傷病名等の種類があり、機能的に分かれている。

診療行為名	診療行為コード
初診	111000110
再診	112007410

医薬品名	医薬品コード
オイグルコン錠1. 25mg	613960002
オイグルコン錠2. 5mg	613960003
グリベンクラミド錠1. 25mg「EMEC」	610412056

傷病名	傷病名コード
1型糖尿病	2500014
2型糖尿病	2500015

# 申出から解析の流れ

# 申出から解析の流れ

## 提供申出

- 解析施設や各種規定を整備する。
- 特別抽出では必要なデータ種類、対象期間を指定し厚労省に抽出を依頼する。この申出段階で具体的な結合方法や解析方法を申出書別添8にて示しておく必要がある。

## データ結合

- 解析上の必要に応じてデータベースの加工形成をする必要がある。

## クリーニング

- 提供を受けたらまず、自分たちの指定したデータがもれなく抽出されているかを確認する。
- その後、解析に応じて使えないデータに対してデータクリーニングを行なう。

## 抽出解析

- 実際に解析を進めていく
- 基本的に初めの研究計画通りに解析を進める。

提供申出（特別抽出）

# 提供申出

提供申出のためにはまず研究機関の設備や基準を整えることが必要。

- 各種規定の整備
- 解析室の設置(次ページ)

各種規定等の整備について、以下を作成した。(マニュアルP15別添2)

- セキュリティー基本方針
- 実施フロー図(別添2-1)
- リスク分析・対応表(別添2-2)
- 運用管理規定(別添2-3)
- 内部監査(自己点検)規定(別添2-4)



# 実施フロー図

厚生労働省

DVDの返却  
レセプトデータ原本 (DVD)

名古屋市立大学大学院薬学研究科  
医療情報解析室

(施錠、入退室管理)

外付けHD  
レセプトデータ  
(1回複写原本)  
中間解析データ

物理的フォーマット

(チェーン固定)

解析用PC

LANは接続しない(OSおよびウイルス定義ファイル更新時を除く、ポート管理)

USBメモリ  
集計データ

(チェーン固定)  
(スクリーンセーバー)

プリンター  
帳票データ

外部発表

施錠ロッカー  
レセプトデータ原本(DVD)、  
USBメモリ、帳票データ等

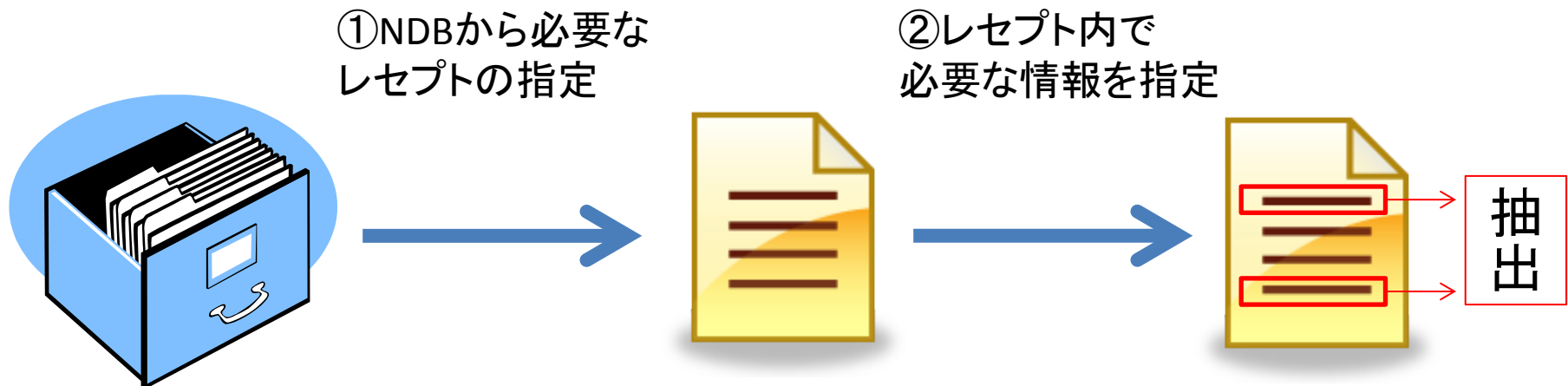
シュレッダーによる裁断

# 特別抽出のロジック

イメージとして以下のような手順で行なう。

※必要最小限のデータ提供であることを忘れてはならない。

- ① 抽出レセプトの指定(医科レセプトと調剤レセプトでそれぞれ指定)
- ② 抽出レセプトのうち必要な情報のみを指定。



# 申出書別添8記入例

特別抽出においてはどのようなデータを受け取りたいかについては、申出書に添えて提出する必要がある。(マニュアルP15、別添8)  
別添資料を用いて不必要なデータが含まれないかなどを厳しく調査される。そのため、申出者は必要なデータのそれぞれに対して必要性を説明できなければならない。

例) 平成24年の一年間での糖尿病患者における糖尿病治療の医薬品使用実態を調べたい場合。

## 1. 期間

平成24年1月から平成24年12月診療分

## 2. レセプトの種類

医科レセプト、調剤レセプト

## 3. レセプト抽出条件

糖尿病患者の全ての医科レセプト。

糖尿病治療薬を処方された全ての調剤レセプト。

①抽出レセプトの指定

# 申出書別添8記入例

## 4. 必要な項目

②抽出レセプトのうち必要な情報の指定

医科レセプトREレコード

通番1、通番2、診療年月、調剤年月、匿名化氏名、年齢階層、男女区分

医科レセプトIRレコード

通番1、通番2、都道府県

医科レセプトSYレコード

通番1、通番2、傷病名コード(糖尿病、別表1参照)

医科レセプトIYレコード

通番1、通番2、医薬品コード(糖尿病治療薬、別表2参照)

調剤レセプトREレコード

通番1、通番2、診療年月、調剤年月、匿名化氏名、年齢階層、男女区分

調剤レセプトYKレコード

通番1、通番2、都道府県

調剤レセプトIYレコード

通番1、通番2、医薬品コード(糖尿病治療薬、別表2参照)

上記別表については次ページを参照。

# 申出書別添8記入例(別表)

別表1	
1型糖尿病	2500014
2型糖尿病	2500015
...	...

これらの例の他にも糖尿病の病名は多く存在する。その全てを全てこの別表に書き出す。

別表2	
オイグルコン錠1. 25mg	613960002
オイグルコン錠2. 5mg	613960003
グリベンクラミド錠1. 25mg 「EMEC」	610412056
...	...

病名と同様に、集計したい医薬品は全て表にまとめる。  
また、剤型の違いや、含有量の違い、ジェネリックなどでそれぞれのコードが異なる。それを全て書き出す必要がある。

# データ結合

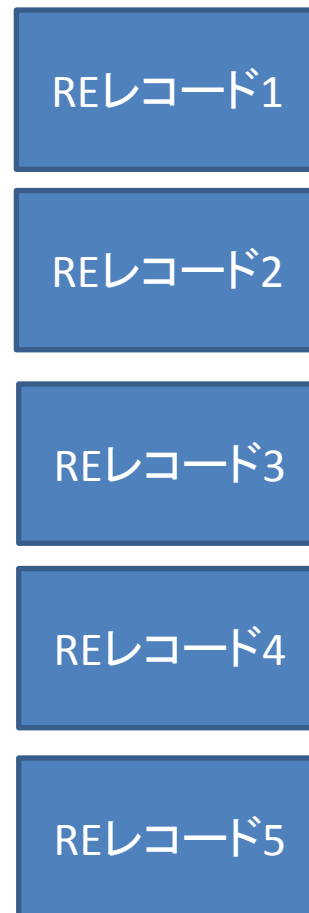
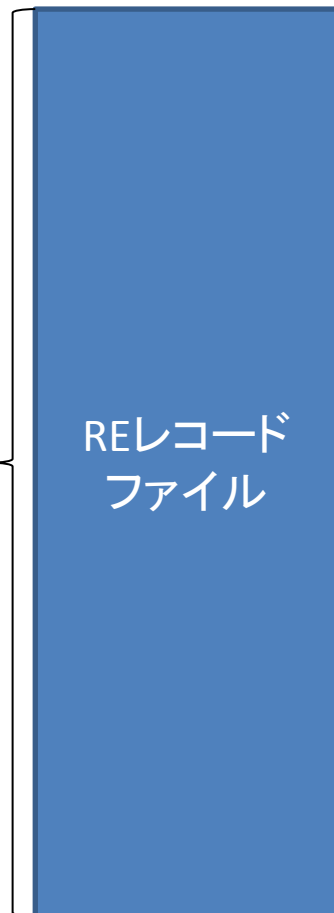
# 提供される状態について

研究に適したデータの形式



実際に提供された形式

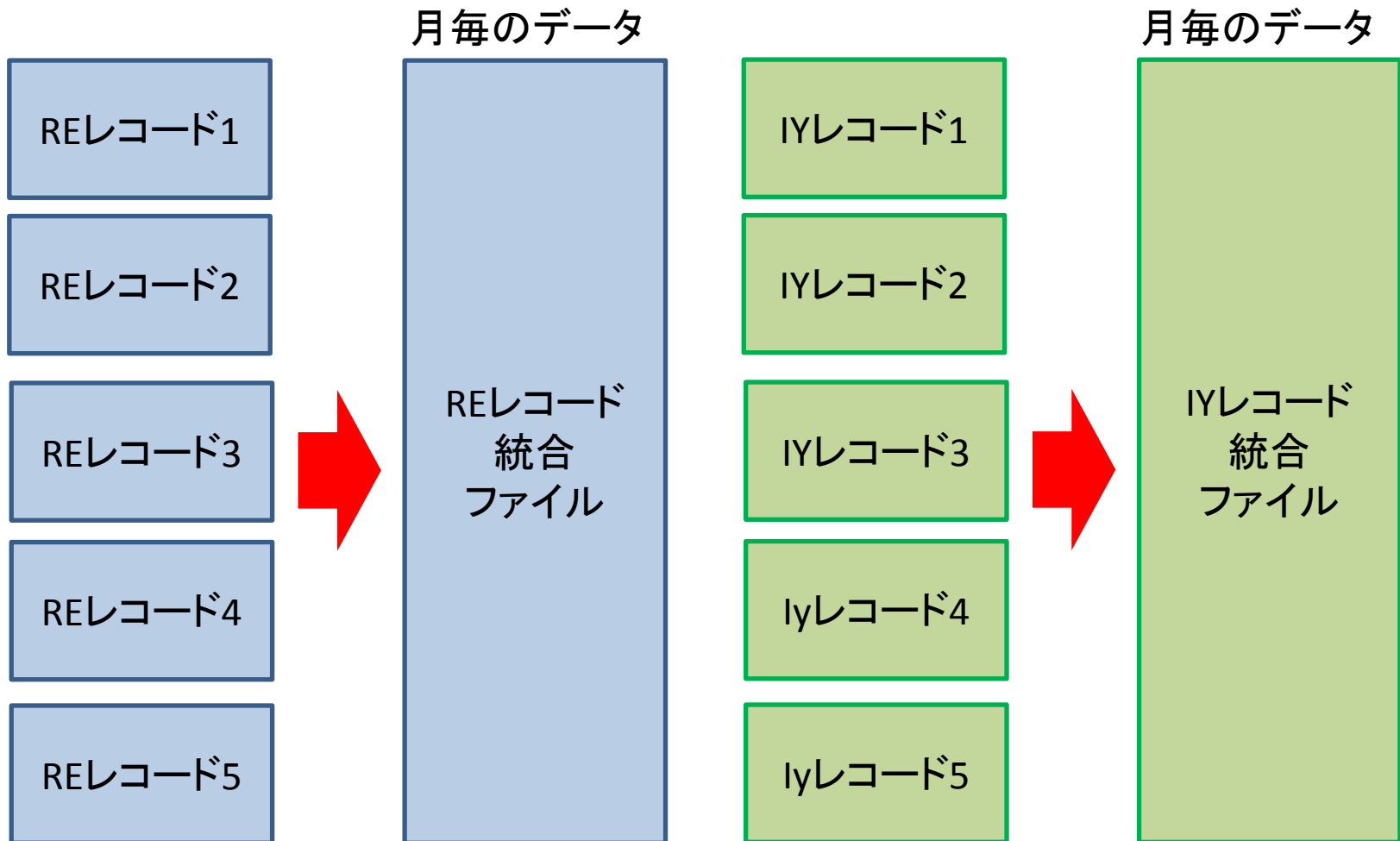
100万人分のデータがあると仮定すると、その全てが1ファイルにまとまっている状態。



実際の提供では20万人分×5ファイルというように分割された状態であった。

# 統合ファイルの作成

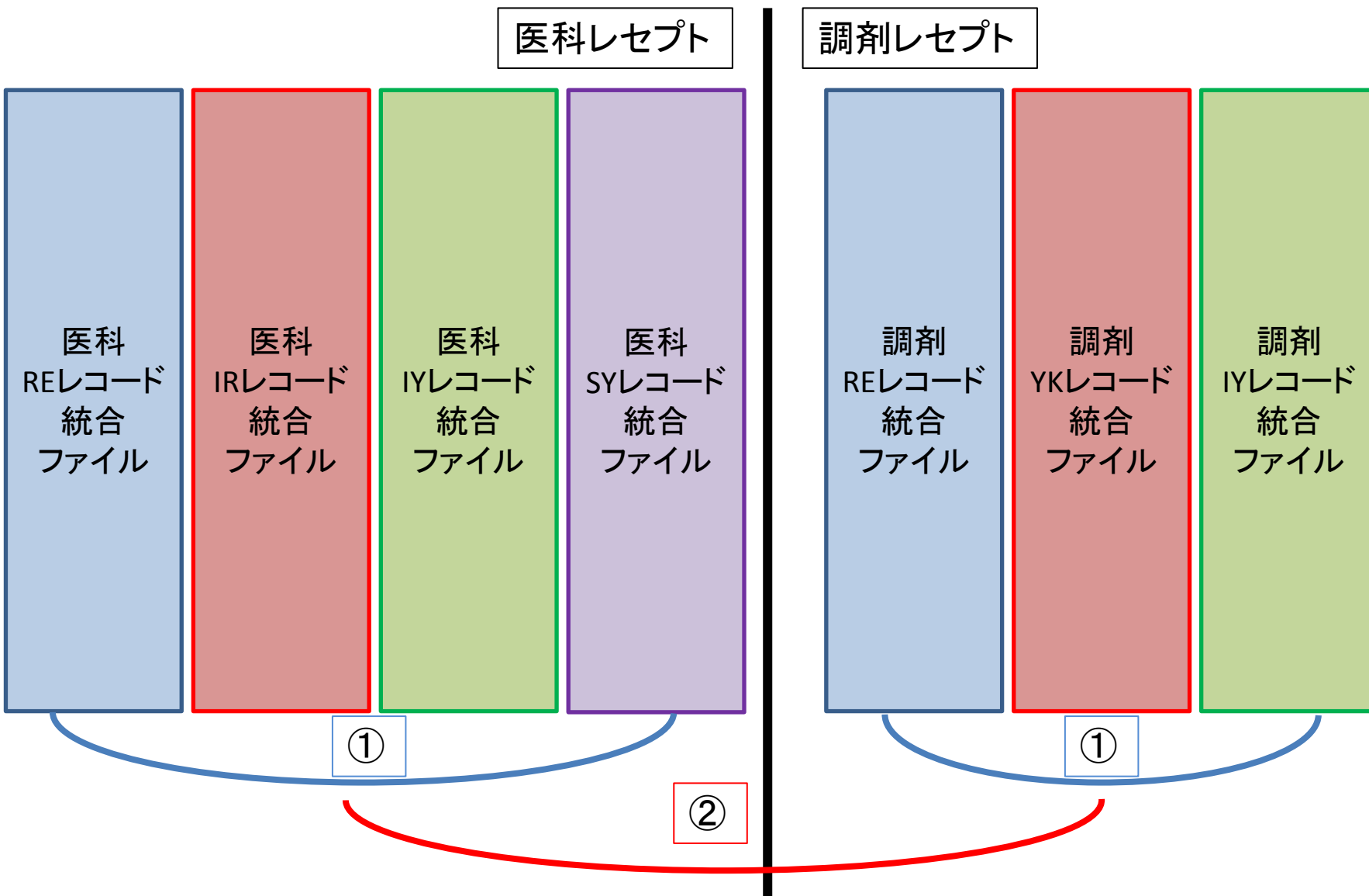
毎月のレコードごとにデータをまとめて一つの統合ファイルを作成した。





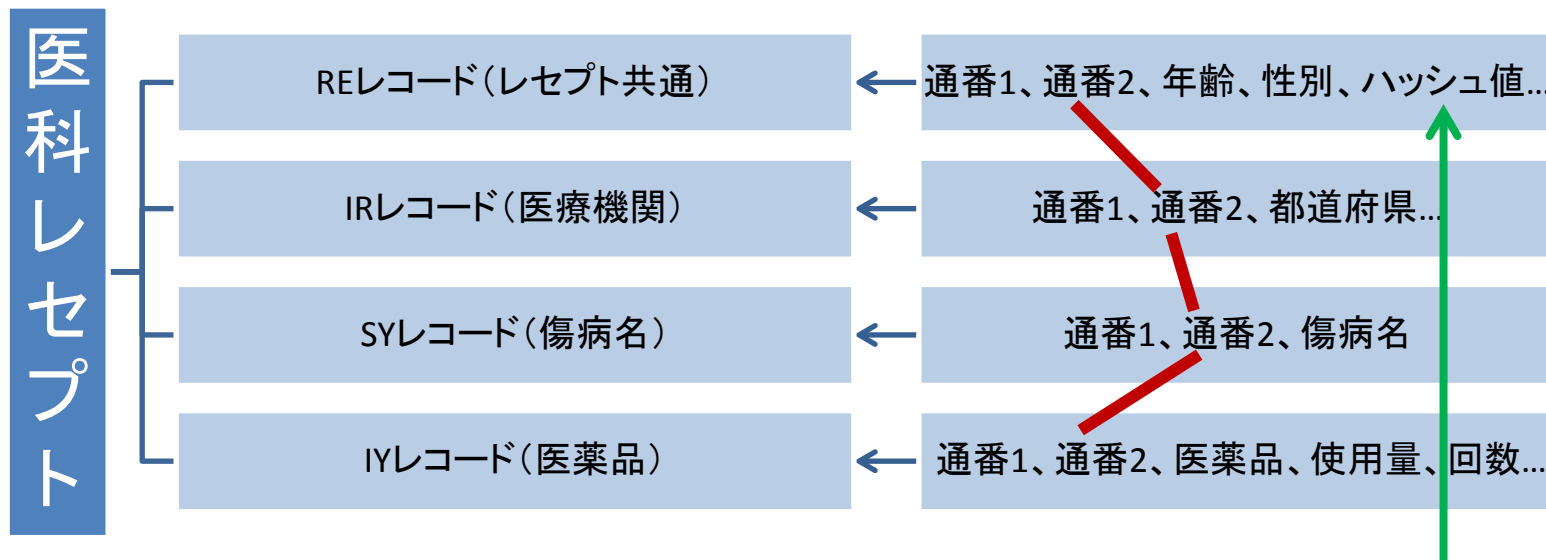
# 結合手順

- ① 同一レセプト内結合
- ② 別のレセプト同士結合(下の例では医科レセプトと調剤レセプトだが、別月の医科レセプト同士もこれに含まれる。)



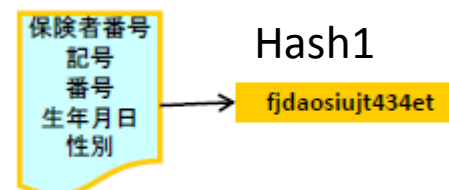
# ①同一レセプト内結合

データベース上では、通番2と呼ばれる同一レセプト由来のレコードにおいての共通番号が付与されている。それを用いて各レコードを名寄せできる。

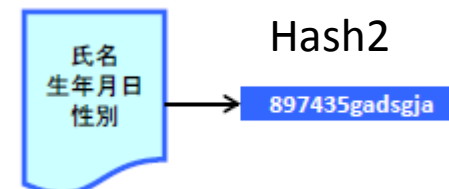


## <ハッシュ値の説明> (利用例は次ページ)

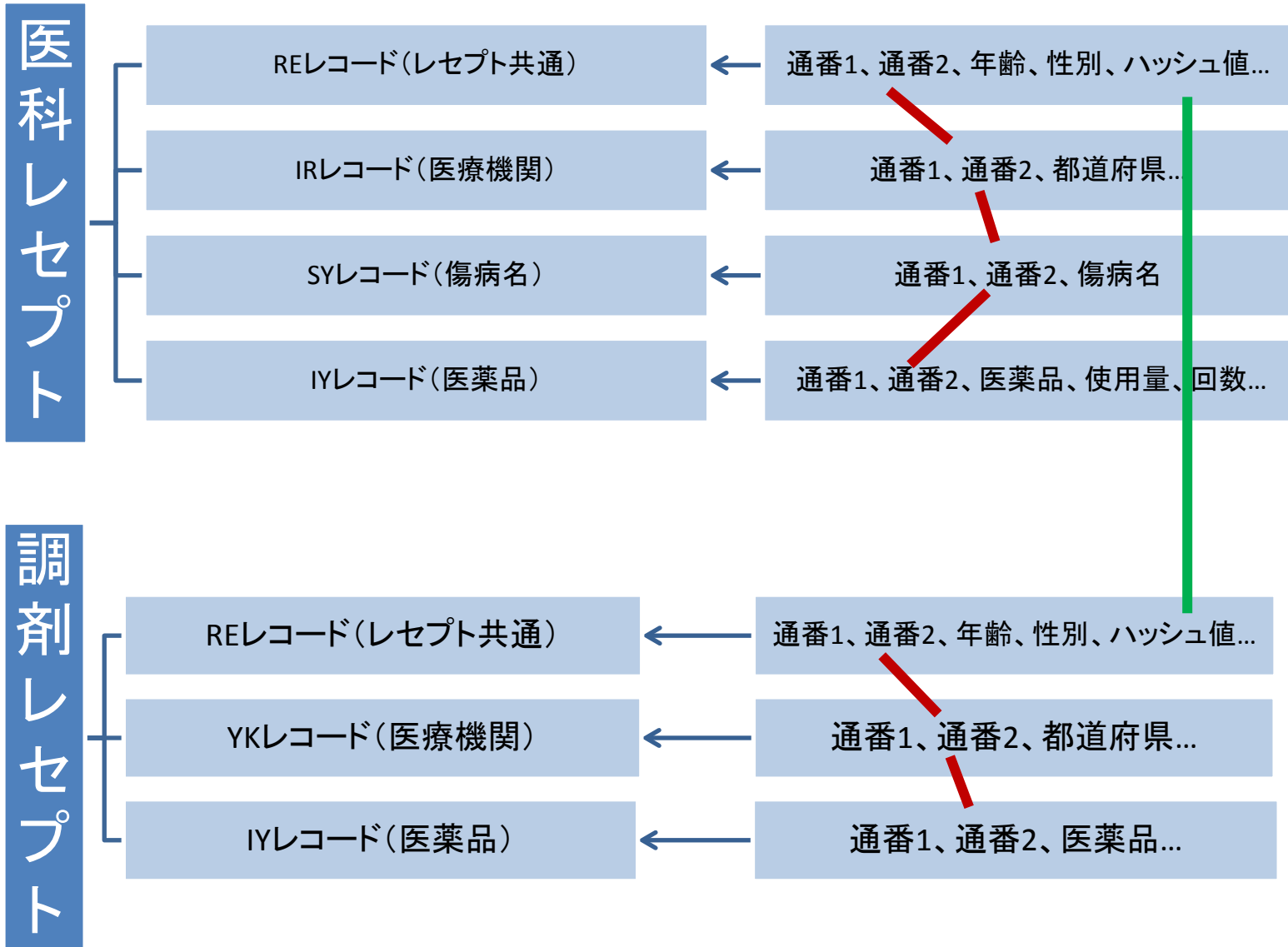
ハッシュ値は匿名化氏名であり、別レセプトとの結合に利用できる。右図のように、ハッシュ値は2種類あるため、保険者番号や名前が同時に変わらない限り、ある個人のレセプトを結合できる。同じ医療機関で受けた診療を時系列で結合することはもちろん、別々の医療機関での診療情報を結合できるなど、非常に有用な匿名化氏名であると考えられる。(マニュアルP11, P12参照)



元データ



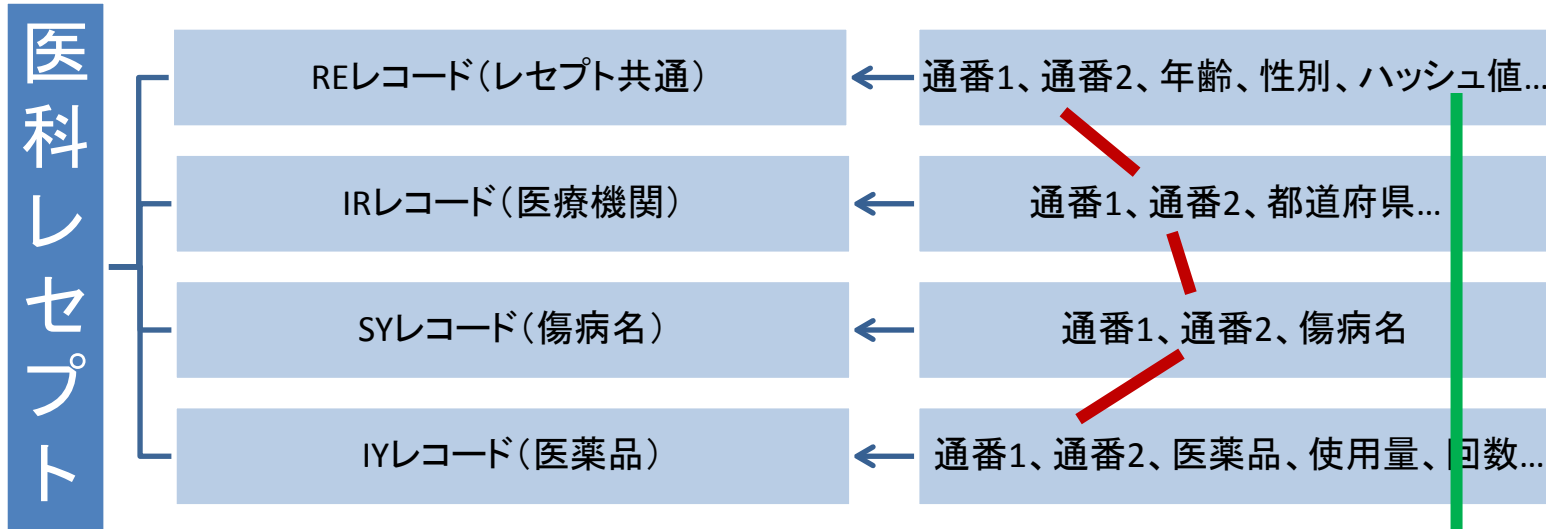
## ②別のレセプト同士の結合



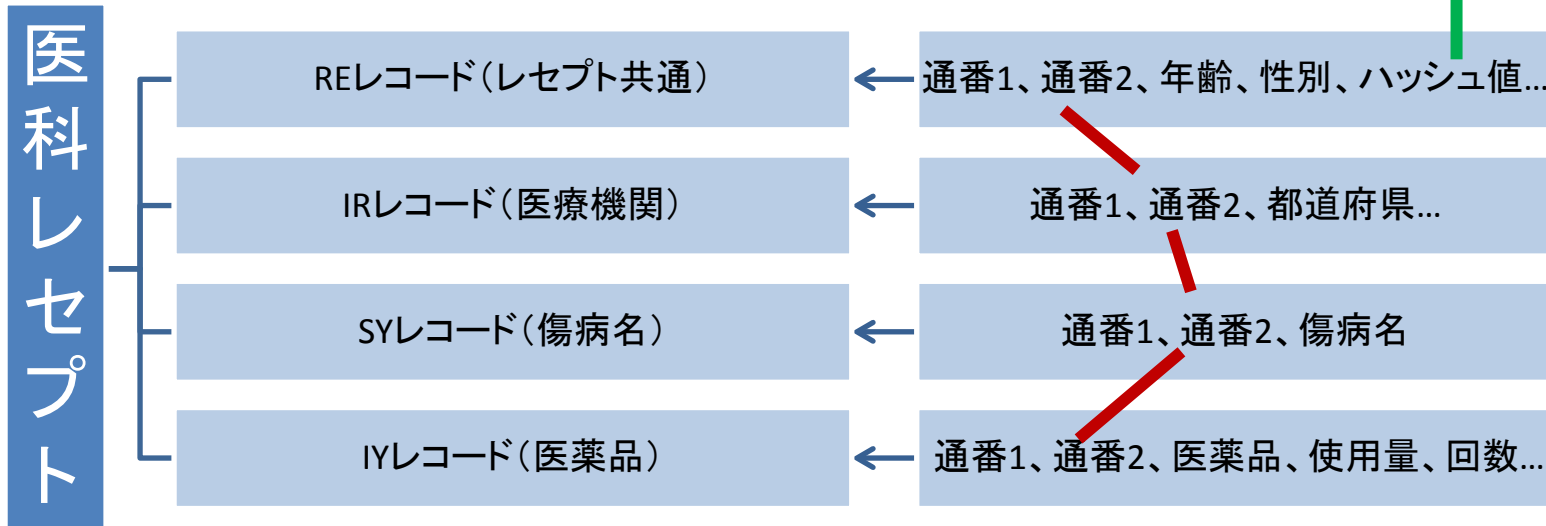
先ほどのステップで同一レセプトの結合をした。(赤)  
その上でそれぞれのレセプトのREレコード由来のハッシュ値を用いて結合することが出来る。(緑)

## ②別のレセプト同士の結合

1月



2月



同様に同一人物でハッシュ値が変わらなければ、月またぎの結合が可能。

## 実際のレコード結合例

RE レコード	通番1	通番2	性別	年齢階層	ハッシュ値
	1	100	1	30	1000
	11	200	2	55	2000

IR レコード	通番1	通番2	都道府県
	2	100	13
	12	200	26

IY レコード	通番1	通番2	IYコード
	3	100	4646868
	13	200	3468467

SY レコード	通番1	通番2	SYコード	修飾語コード
	4	100	354564	8002
	14	200	864436	.



通番2をもとに結合することで一つのレセプト情報を  
まとめ上げたラインリストが作成できる。

通番2	性別	年齢階層	ハッシュ値	都道府県	IYコード	SYコード	修飾語コード
100	1	30	1000	13	4646868	354564	8002
200	2	55	2000	26	3468467	864436	.

RE レコード	通番1	通番2	性別	年齢階層	ハッシュ値
	1	100	1	30	1000
	11	200	2	55	2000

IR レコード	通番1	通番2	都道府県
	2	100	13
	12	200	26

IY レコード	通番1	通番2	IYコード
	3	100	4646868
	4	100	5464886
	13	200	3468467

➡ 実際は同じ通番で二つ以上のIYレコードがあることも多い。

SY レコード	通番1	通番2	SYコード	修飾語コード
	5	100	354564	8002
	14	200	864436	.



必要に応じてこのように変数を増やしてまとめることもある。

通番2	性別	年齢階層	ハッシュ値	都道府県	IYコード1	IYコード2	SYコード	修飾語コード
100	1	30	1000	13	4646868	5464886	354564	8002
200	2	55	2000	26	3468467	.	864436	.

信頼性確認、クリーニング

# 信頼性確認・クリーニング

## データ信頼性の確認

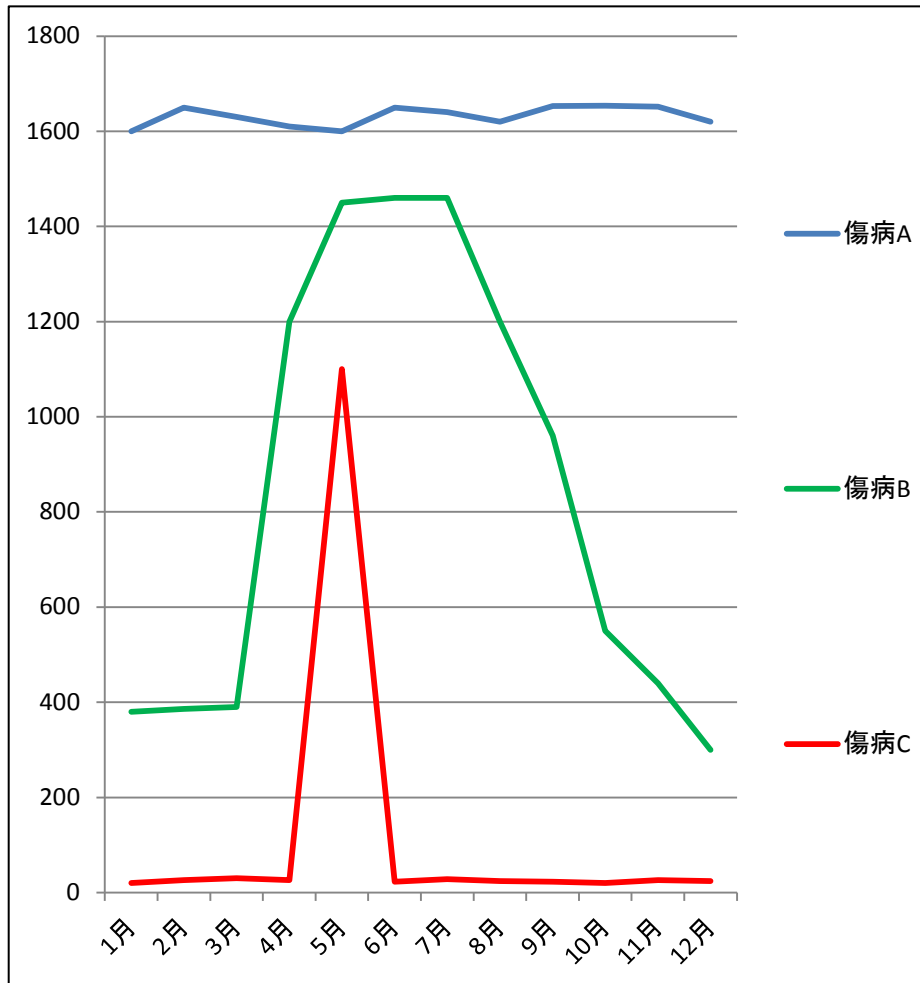
- レセプトの収集過程や抽出過程においては、何らかのミスがある可能性を否定できない。研究に使うデータとして十分な信頼性があるかは各自が必要に応じて確認する必要がある。

## データクリーニング

- 結果の信頼性を上げたいときに一部のデータを除外する必要がある場合がある。ただし、これはバイアスがかかる可能性があるため、必要に応じて慎重に行なう必要がある。



# データの信頼性確認

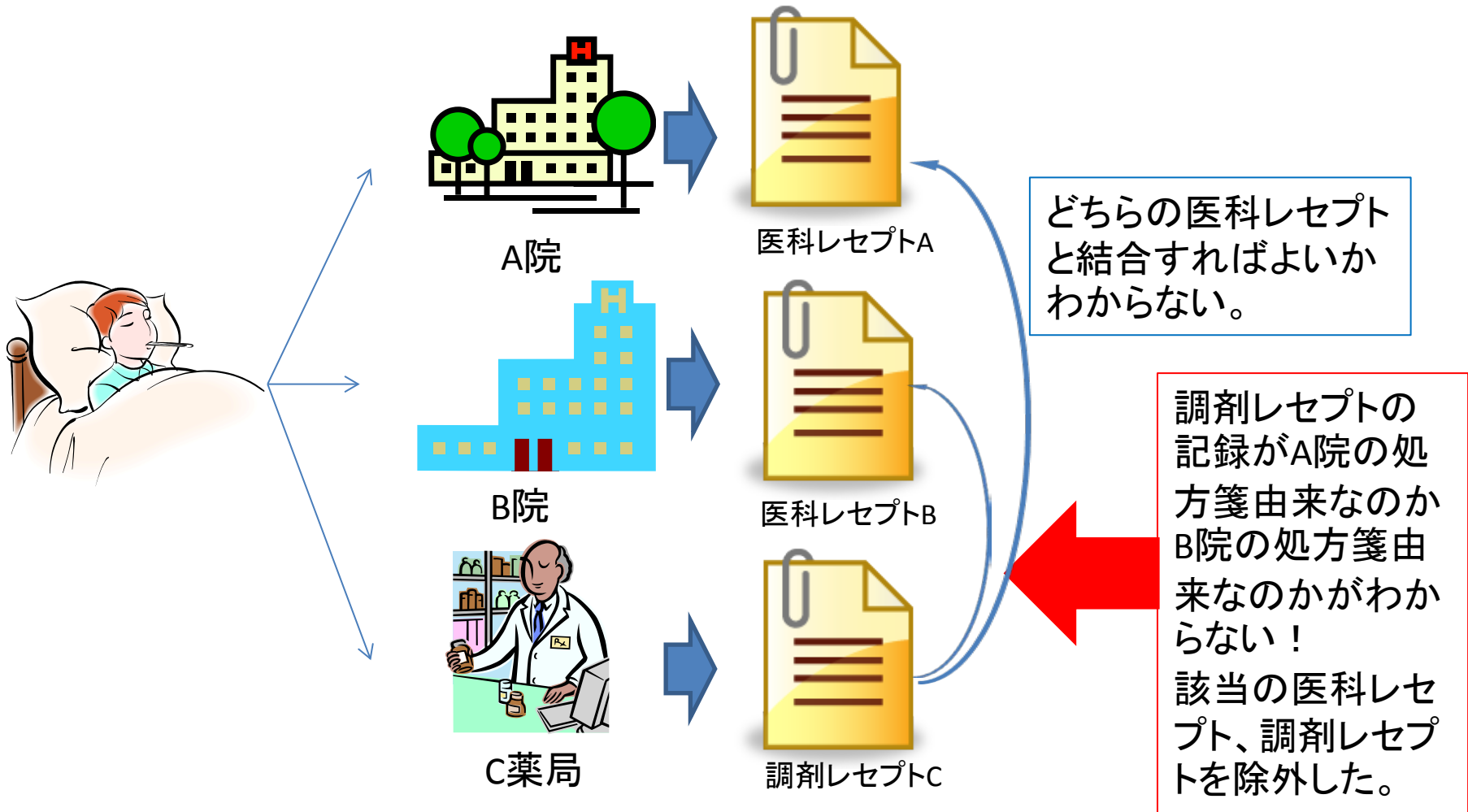


年間を通して患者数が変化しない。  
傷病Aが、季節性なく年間で一定の患者数がある疾患であれば自然。

年間を通して大きく件数が変動している。  
傷病Bが、夏に多くなるという季節性のある疾患であれば自然。

5月のみ突然大幅な増加が見られる。合理的な説明がつかなければ不自然。扱いの際には注意が必要。

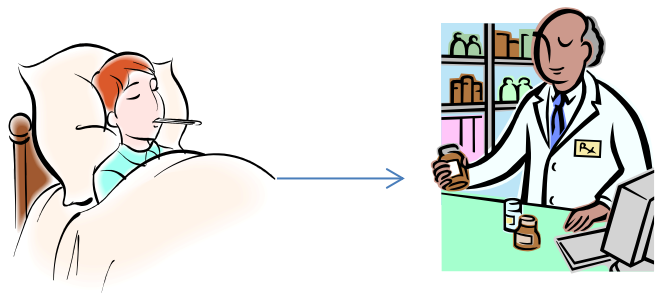
# データクリーニング例①



# データクリーニング例②

日本の保険制度では下記のような無診療調剤は認められていない。そのため、医科レセプトがない患者の調剤レセプトは存在しないことが考えられる。しかし実際にはこのようなデータが存在した。実態として考えられる例を挙げる。

- 病院か薬局のどちらかで情報の入力ミスがありハッシュ値が異なっているため、見かけ上調剤レセプトに対応する医科レセプトがなくなっている。
- 1月31日に病院に行き処方箋をもらったが、当日は忙しかったため、翌2月1日に調剤薬局で処方を受けた。



薬局



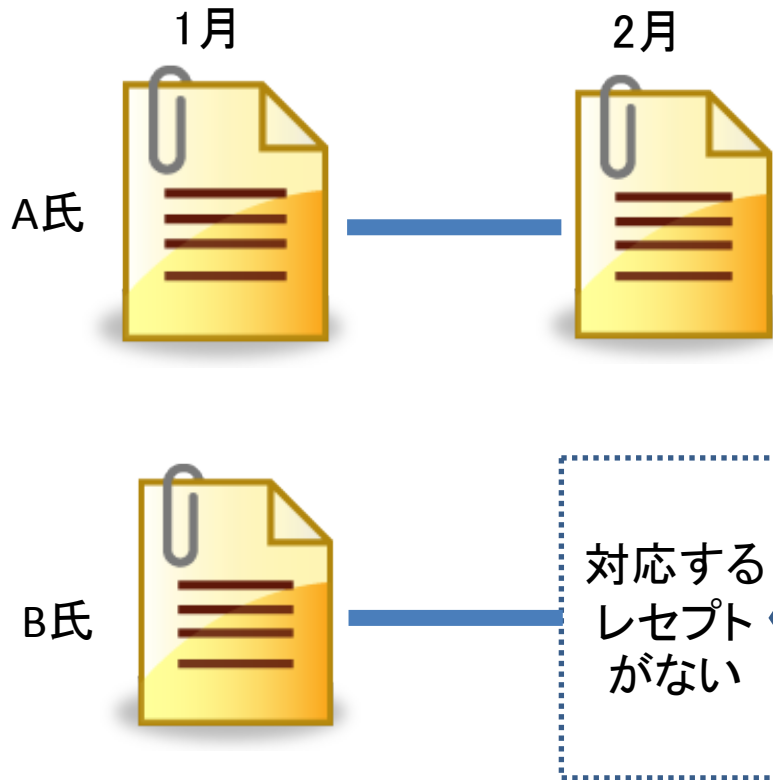
調剤レセプト

該当する調剤レセプトについては除外した。

# データクリーニング例③

例えば医薬品の処方中止を見たいとき。

一月では不可能なので下記のように翌月のデータを結合する必要がある。



このような場合、

- 完治し、病院に行かなくなった。
- ハッシュ値が変わった。

等の実態が考えられる。前者は処方中止と考えていいかもしれないが、後者を中止と考えることは不適切である。よってこのような場合はB氏のデータを除外して解析するのが最適と考えられる。

# 抽出解析

# 抽出解析

申出時の計画に沿ってデータ解析を行なう。  
申出書の別添8の公表形式をあらかじめ計画しておく必要があり、基本的に解析もそれに沿って行なう。

通番2	性別	年齢階層	ハッシュ値	都道府県	IYコード1	IYコード2	SYコード	修飾語コード
100	1	30	1000	13	4646868	5464886	354564	8002
200	2	55	2000	26	3468467	.	864436	.



傷病名(SY)が「864436」で、かつ医薬品「3468467」の投与を受けたレセプト数 = 1

# レセプト情報利用上の注意

# レセプト情報利用上の問題点

- 傷病名が実際の疾患名と一部異なっている場合がある。
- 医療行為の順番がわからない。
- 検査結果がわからない。
- ハッシュ値を用いた結合を繰り返すと標本数が徐々に減少する。
- 処方量データの利用には注意が必要。
- マスター表を用意する必要がある。



# 傷病名が実際の疾患名と一部異なっている可能性がある。

傷病名情報は重要な情報であるが、いわゆるレセプト病名と呼ばれるものがある。

研究において傷病名情報の確実性が必要な場合は、レセプト病名を出来るだけ回避する必要があるため、傷病名情報に診療年月や検査、投薬等の情報を組み合わせるなどして傷病名情報の信頼性を独自の方法で担保していく必要があると考えられる。

# 医療行為の順番がわからない

カルテ等を用いた簡単な副作用検出アルゴリズムの例

医薬品  
処方

検査値  
変動

医薬品  
休薬

アルゴリズムにおいては医療行為の順番が必要と考えられるが、レセプトデータでは同月内の医療行為がどの順番で行なわれたかわからない

平成24年以降は日付が入ったため、順番を確認できるようになった。  
それ以前のデータを利用する場合において注意が必要であると考えられる。

# 検査結果がわからない

カルテ等を用いた簡単な副作用検出アルゴリズムの例

医薬品  
処方

検査値  
変動

医薬品  
休薬

レセプト情報を用いたときの副作用検出アルゴリズムの例

医薬品  
処方

検査あり

医薬品  
休薬



レセプトでは検査結果をアルゴリズムに入れられない

# ハッシュ値を用いた結合を繰り返すと 標本数が徐々に減少する。

レセプト情報を用いたときの副作用検出アルゴリズムの例



このアルゴリズム作成のためには二回の結合が必要



ただ、前述のようにハッシュ値は結婚や転職等で変わることがある。(仮に5%変わるとする)



結合が二回あるので三ヶ月分のデータが結合できる割合は $0.95 \times 0.95 \doteq 0.90$ つまり約90%程度の標本数となる。

# 処方量データの利用には注意が必要

医薬品の処方を絡めた研究では処方量が重要な要素であるが、我々の研究で提供されたデータの中には、処方量等のデータが曖昧ではないかということが示唆されるものが存在した。

以下はその模擬データである。

	使用量	回数
処方記録例	31	1

1日で31錠もの量を服用するとは考えがたい。

# 研究対象期間に対応する マスター表を用意する必要がある。

診療報酬の改訂に伴いマスター表も改訂されるので、研究期間に対応するマスター表を用いる必要がある。



最新のマスター表と更新情報と合わせることで  
目的のマスター表を作成する必要がある。

詳細はマニュアルP30を参照。

# ご清聴ありがとうございました。

詳細については、下記までお問い合わせ下さい。

名古屋市立大学大学院薬学研究科医薬品安全性評価学分野  
頭金 正博

Email; [tohkin@phar.nagoya-cu.ac.jp](mailto:tohkin@phar.nagoya-cu.ac.jp)