

平成24年度～平成25年度厚生労働科学研究  
「汎用性の高いレセプト基本データセット作成に関する研究」  
の進捗状況について

平成25年12月19日

医療経済研究機構 満武 巨裕

# 研究の概要について

- 本研究「汎用性の高いレセプト基本データセット作成に関する研究」は、第8回有識者会議において検討された「傷病や診療行為等、さまざまな研究目的に対して最適化されたデータセットの整備」を受けた研究で、平成24年度に採択されたものである。
- 本研究では、おおむね以下の研究を行うこととしている。
  - ✓ 匿名化等のセキュリティ対策を施した汎用性の高いレセプト基本データセットの設計・作成を目的とする。
  - ✓ 分析を行いやすい抽出率の検討、アウトライヤーの処理等、統計的検証も行う。
  - ✓ 一定期間のレセプトを個人ID毎に統合し、最小限の情報しか含まないデータセットから、相当程度の情報を含むデータセットまで、複数の水準で基本データセットを整備する。
  - ✓ 特定健診・特定保健指導情報とレセプト情報を紐付けたデータセットの作成も行う。
- 基本データセット作成は保険局が所有するレセプト情報等データを用いて行っているが、全データを使って分析するのはデータセンターでの抽出作業に時間を要し他業務への影響が大きくなることから、以下の要領で抽出したデータを分析対象とした。
  - 2009年度、2010年度、2011年度の全レセプト・特定健診・特定保健指導データから下記を分析対象とする。
    - 都道府県での区分が可能な保険者（協会けんぽ、市町村国保など）のデータは、47都道府県を8ブロックに分け、それぞれのブロックから都道府県をランダムに人口比約25%となるように選定し、その県のすべての保険者のレセプト情報を抽出する。
    - 全国にまたがる保険者（組合健保、共済組合など）のデータは、全保険者番号から20%をランダムに選定し、その保険者番号を有する情報をすべて抽出する。
    - 都道府県のランダム選定、保険者番号のランダム選定は、第三者に依頼する。
    - 医療機関コード、保険者番号は全く異なる一意の番号を付与する。ただし保険者番号に付与された一意の番号からは、都道府県情報および法別番号情報が確認できる形式とする。
  - 加えて、抽出データとの比較対照用に通年の全数データが必要であることから、2010年度分データを分析対象とする。
- データ利用にあたっては、第三者提供で求められている水準と同等のセキュリティ環境を準備し、データの安全な運用に細心の注意を払いつつ、研究を行っているところである。

## 昨年度の結果：レセプト情報等解析システムの構築とデータ検証

- NDBデータの医科(入院、外来)・調剤の集計値と国民医療費(厚生労働省・統計情報部)を比較したところ、同様の傾向を示した。疾患別医療費も同様であった。
- IDは、被保険者証番号に半角・全角や氏名に関しては漢字表記の揺れ等の理由で、実際には同一の被保険者にも関わらず、複数が発生している可能性がある。そのために、試行的に、ID1とID2の特性を生かし、複数紐づくものに関しては同一人物とみなし、新しいIDを検討した。
- 調剤レセプトは、外来診療に付随して発生するものであるが、外来レセプトに対応しない件数がID1では5.6%、ID2では12.9%存在した。

## 現時点の成果：基本データセットの試作(次頁参照)

- 一定期間のレセプトを個人ID毎に統合する等の処理をし、診療行為も含むデータセット
- 特定健診・保健指導データとレセプトとリンクさせたデータセット

# 傷病名の特定化:「主傷病名」の使用について

- 基本データセットでは、各レセプトに対し記載されている傷病名から「**主傷病名**」をもとにした傷病名を入院・外来それぞれ**1つを付与**した、サマリー的なデータセットが提供可能である。
- 具体的な傷病名の選択方法としては、レセプトに付与されている傷病名のうち、**主傷病名として格納されているもののなかから最も新しい傷病名であり、複数レセプトの中で一番高い点数の病名**を付与することを想定している。
- レセプトに付与されている傷病名については、いわゆる「レセプト病名」の存在など、その精度がしばしば問われる(第5回有識者会議資料 <http://www.mhlw.go.jp/stf/shingi/2r98520000013uoy-att/2r98520000013uvi.pdf> も参照)ことがあるが、本データセットにおいて、入院・外来それぞれの傷病名1つを付与することとした理由は以下のとおりである。
  - ✓ 他の各種統計調査においても、レセプトに記載されている傷病名から傷病名を1つに限定したうえで調査が行われていることから、主傷病名を中心とした代表性の高い傷病名を1つ抽出して患者の傷病状況を代表させることは、一定程度可能であると考えられた。
    - 社会医療診療行為別調査においては、レセプトに主傷病が複数記載されている場合は、記載順により傷病名が選択されている。  
(政府統計の総合窓口e-Statウェブサイトの社会医療診療行為別調査の項目「用語の解説等」(<http://www.e-stat.go.jp/SG1/estat/List.do?lid=000001106262>)も参照)
    - 国民医療費の概要に関する調査(入院-入院外・年齢階級・性・傷病分類別にみた医科診療医療費)においては、患者調査のデータをもとに推計している。患者調査の主傷病名は、“入院患者においては、調査日現在、入院の理由となっている傷病、外来患者においては、調査日現在、主として治療又は検査をしている傷病、退院患者においては、退院時に入院の理由となっていた傷病をいう。”
  - ✓ 傷病名に関する情報をすべて提供することとなると、傷病名に関する細かい情報が提供されることとなり、それをもとに患者個人の特定可能性が高まる恐れがあると考えられた。

## 国民医療費(厚生労働省)と研究班データの 疾病分類別医療費の構成割合の比較(2010年度)

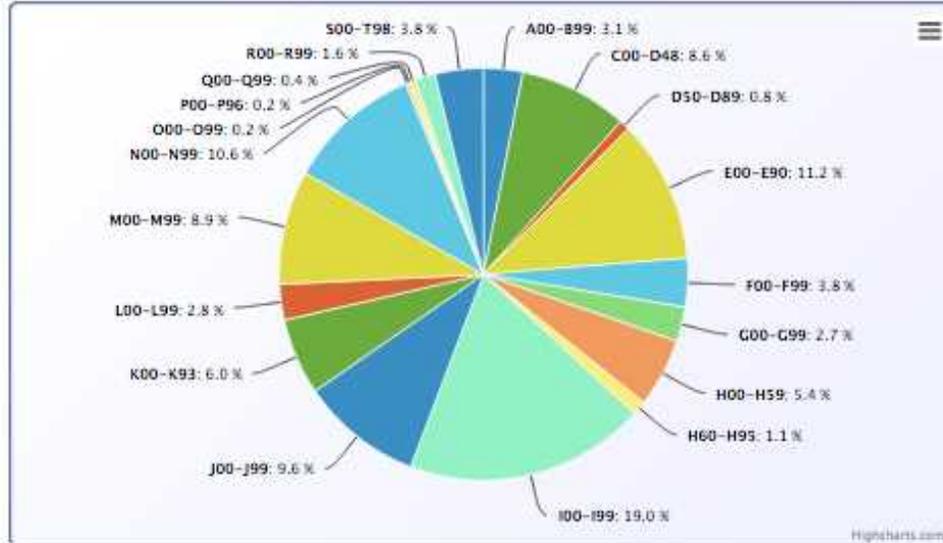
			厚生労働省 国民医療費公表値	研究班データ
1	<b>A00-B99</b>	感染症および寄生虫症	2.5%	2.3%
2	<b>C00-D48</b>	新生物	12.8%	8.7%
3	<b>D50-D89</b>	血液および造血器の疾患ならびに免疫機構の障害	0.8%	0.6%
4	<b>E00-E90</b>	内分泌,栄養および代謝疾患	7.3%	7.7%
5	<b>F00-F99</b>	精神および行動の障害	7.2%	7.7%
6	<b>G00-G99</b>	神経系の疾患	4.3%	4.7%
7	<b>H00-H59</b>	眼および付属器の疾患	3.5%	3.3%
8	<b>H60-H95</b>	耳および乳様突起の疾患	0.7%	0.7%
9	<b>I00-I99</b>	循環器系の疾患	20.8%	19.3%
10	<b>J00-J99</b>	呼吸器系の疾患	7.8%	7.2%
11	<b>K00-K93</b>	消化器系の疾患	6.1%	4.6%
12	<b>L00-L99</b>	皮膚および皮下組織の疾患	1.7%	1.5%
13	<b>M00-M99</b>	筋骨格系および結合組織の疾患	7.4%	5.0%
14	<b>N00-N99</b>	尿路性器系の疾患	7.1%	7.8%
15	<b>O00-O99</b>	妊娠,分娩および産じょく<褥>	0.8%	0.3%
16	<b>P00-P96</b>	周産期に発生した病態	0.6%	0.2%
17	<b>Q00-Q99</b>	先天奇形,変形および染色体異常	0.6%	0.4%
18	<b>R00-R99</b>	症状,徴候および異常臨床所見・異常検査所見で他に分類されないもの	1.5%	1.5%
19	<b>S00-T98</b>	損傷,中毒およびその他の外因の影響	6.6%	4.5%
20	<b>V00-Y98</b>	傷病および死亡の外因		0.0%
21	<b>Z00-Z99</b>	健康状態に影響をおよぼす要因および保健サービスの利用		0.3%
22	<b>U00-U99</b>	特殊目的用コード		0.0%
		分類不能		11.7%

研究班データの病名は、主病名フラグがついているレセプトの最初の傷病名コードとした。

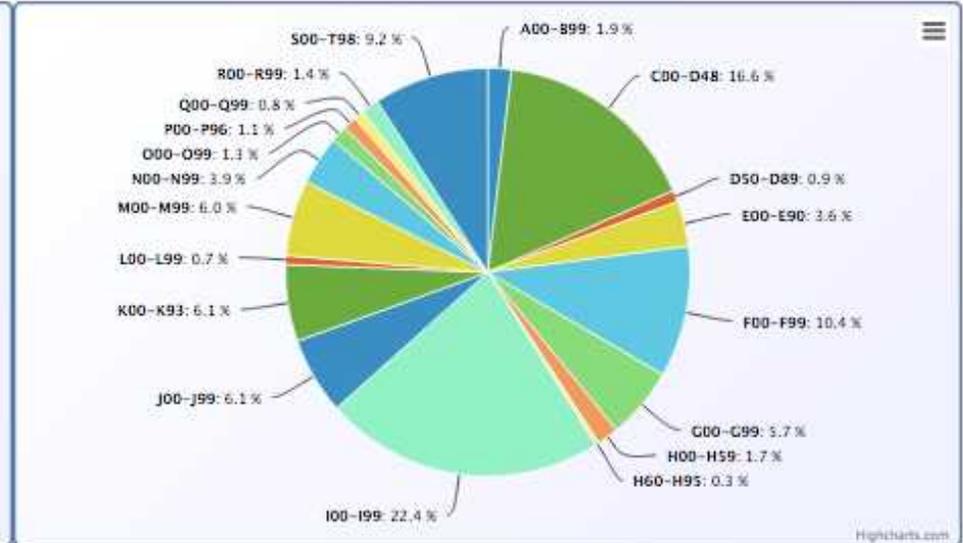
# 参考：基本データセット「主傷病名」と国民医療費の比較

## 「国民医療費の概況」

疾病大分類ごとの医療費（外来）

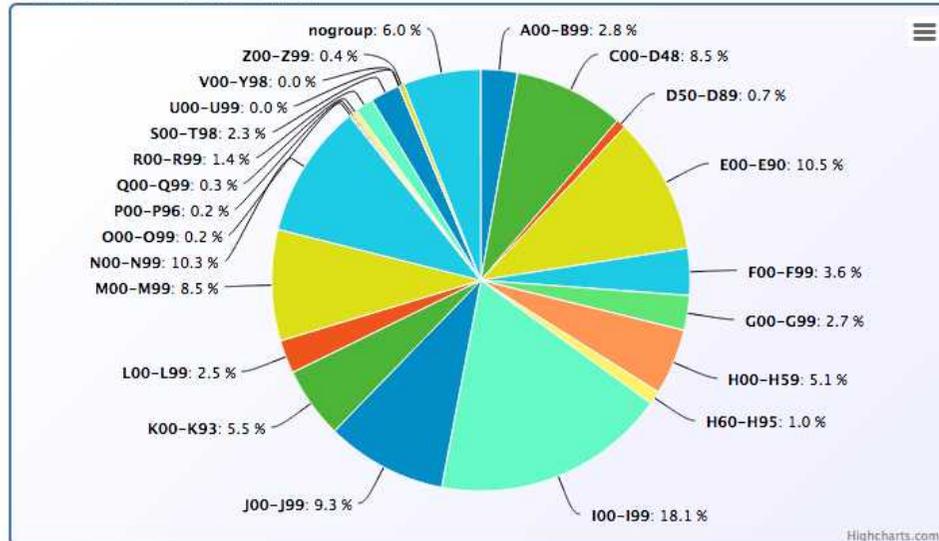


疾病大分類ごとの医療費（入院）

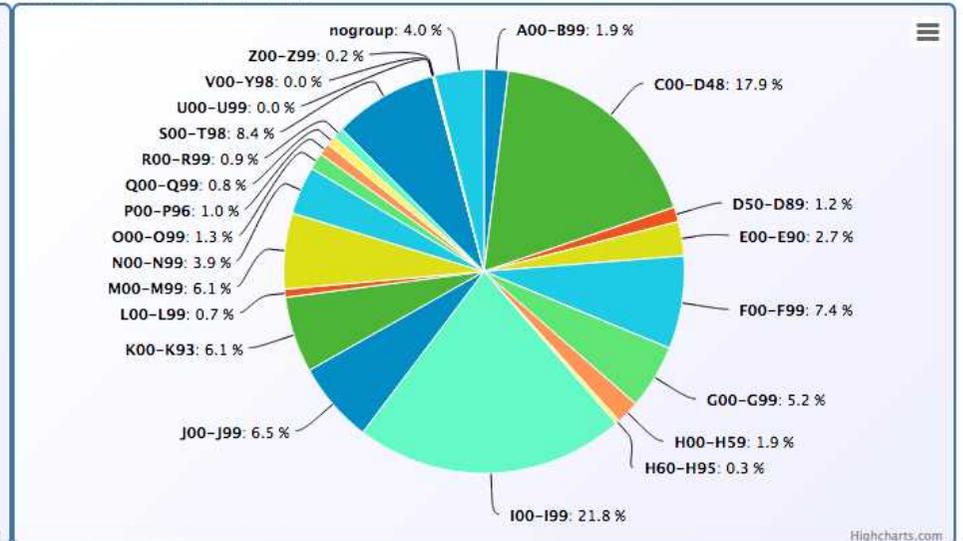


## 「研究班データ」

疾病大分類ごとの医療費（外来）

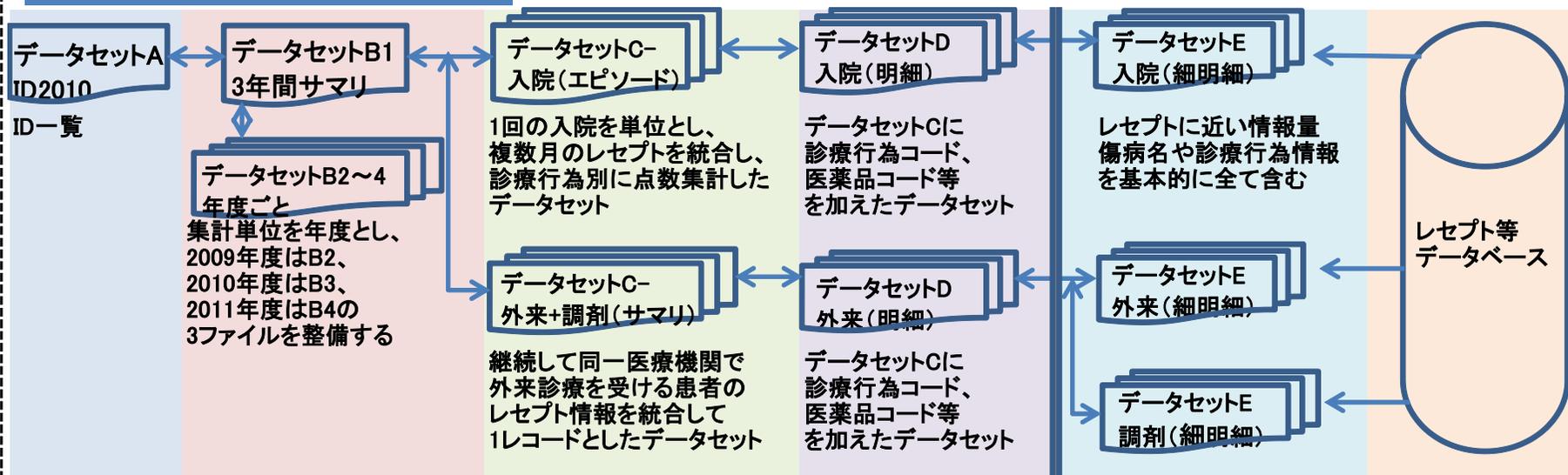


疾病大分類ごとの医療費（入院）



# 基本データセット(案)の構成

## NDBデータセット構成(案)



## 個人の特定期間性の高いデータ

### データセットA

- ID、性別、年齢階級のみ
- 提供データの管理用データセットであり、第三者提供は行わない

### データセットB

- 患者毎に紐付けを行い、3年分のデータ、および単年度分毎のデータからなるデータセット
- 属性に加え医療費、傷病名など限られた情報のみのデータセット

### データセットC

- 1回の入院、あるいは一連の外来診療をひとまとまりの単位として整備したデータセット
- データセットBと同水準の情報を含むデータセット

### データセットD

- 診療をひとまとまりの1回の入院、あるいは一連の外来単位として整備したデータセット
- データセットCの情報に加え、ユーザーの指定した診療行為コード、医薬品コード等の情報を含むデータセット

### データセット細明細

- データセットDに加え、傷病名や診療行為、医薬品等のデータを全て盛り込んだデータセット
- 現在のレセプト情報等データベースに近いデータセット

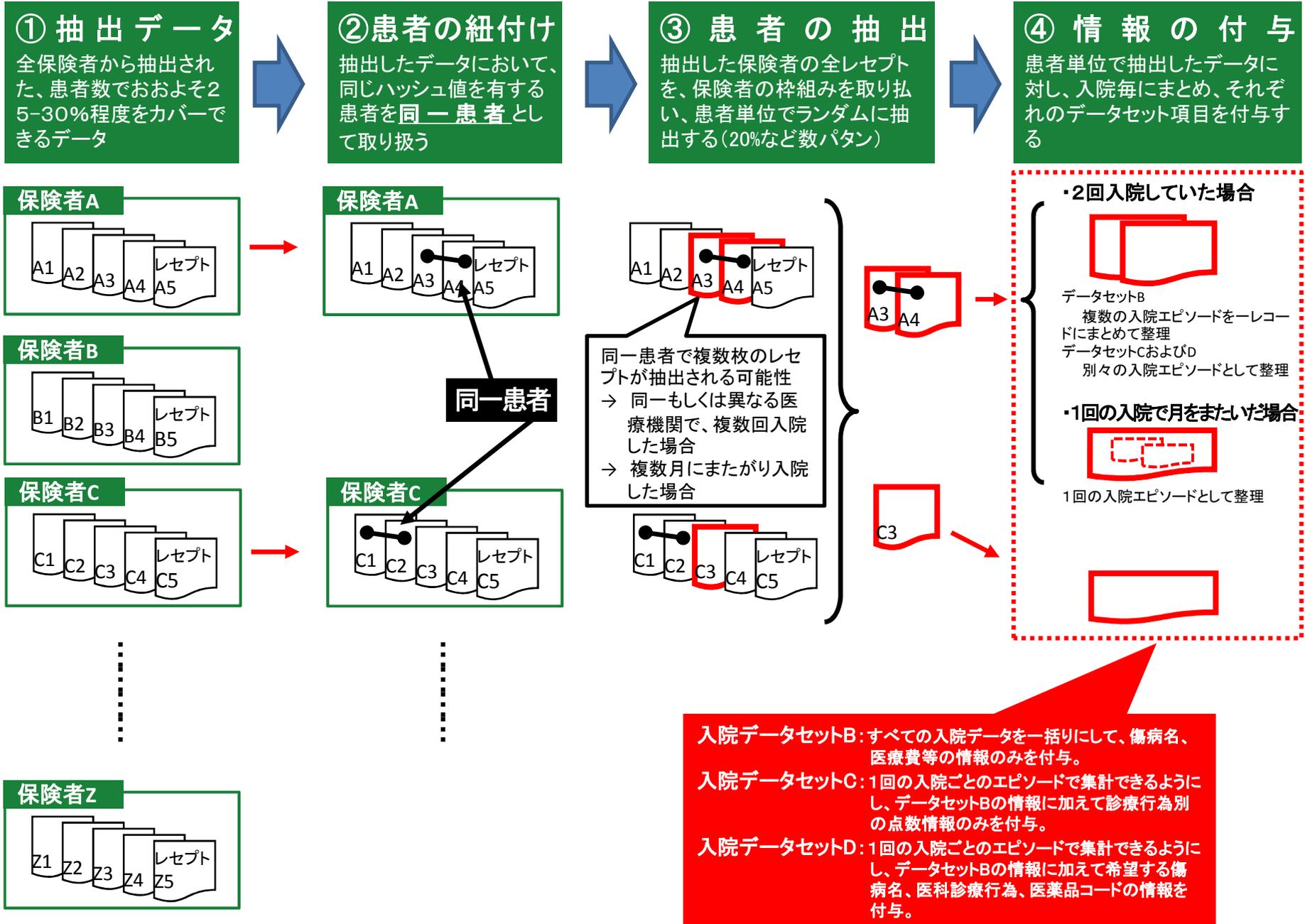
現在のNDB特別抽出

# 入院データセット B, C, Dについて

	入院データセットB	入院データセットC	入院データセットD
ファイル容量	<ul style="list-style-type: none"> <li>約500MB</li> </ul>	<ul style="list-style-type: none"> <li>約500MB</li> </ul>	<ul style="list-style-type: none"> <li>約1GB</li> </ul>
レコード数 (=患者数)	<ul style="list-style-type: none"> <li>保険者単位で抽出した約534万件の患者レコードから、さらに20%をランダムに抽出</li> </ul>	<ul style="list-style-type: none"> <li>左記患者の入院エピソード全て</li> </ul>	<ul style="list-style-type: none"> <li>左記患者の入院エピソード全て</li> </ul>
抽出患者	<p>【保険者の抽出方法】</p> <ul style="list-style-type: none"> <li>都道府県単位の保険者（協会けんぽ等）は、全国8ブロックから1つずつ、人口25%分の保険者を抽出</li> <li>全国にまたがる保険者（組合健保、共済等）は、全保険者から20%をランダムに抽出</li> </ul> <p>【抽出された保険者間での、患者抽出方法】</p> <ul style="list-style-type: none"> <li>抽出された保険者の全患者から20%分に該当する患者をランダムに抽出</li> </ul>		
対象期間	<ul style="list-style-type: none"> <li>1年</li> </ul>	<ul style="list-style-type: none"> <li>入院イベントごと</li> <li>データセットBと併用することで、年間の入院イベントを全て把握できる。</li> </ul>	<ul style="list-style-type: none"> <li>入院イベントごと</li> <li>データセットBと併用することで、年間の入院イベントを全て把握できる。</li> </ul>
キーになるID	<ul style="list-style-type: none"> <li>ID1とID2を用いて試行的に作成したID3を用いて、レセプト間、データセット間の紐付けが可能である。</li> </ul>		
患者属性情報	<ul style="list-style-type: none"> <li>IDおよび性、5歳刻み年齢階級</li> </ul>		
項目	<ul style="list-style-type: none"> <li>入院／外来合計点数</li> <li>入院／外来合計診療実日数</li> <li>入院／外来受診医療機関数</li> <li>主傷病名</li> </ul>	<p>【入院データセットBに加え】</p> <ul style="list-style-type: none"> <li>診療識別情報</li> </ul>	<p>【入院データセットBに加え】</p> <ul style="list-style-type: none"> <li>申出者の要望により、診療行為、医薬品情報、傷病名等を提供</li> </ul>
データ構造	<ul style="list-style-type: none"> <li>データセットは固定</li> </ul>	<ul style="list-style-type: none"> <li>データセットは固定</li> </ul>	<ul style="list-style-type: none"> <li>データセットは要望に応じ可変</li> </ul>
研究の例	<ul style="list-style-type: none"> <li>年間の医療機関受診動向</li> <li>年齢階級別、性別、入院／外来別、傷病別の医療費</li> <li>等</li> </ul>	<p>患者別の診療行為別医療費分析等 (社会医療診療行為別調査の個票のイメージ)</p>	<p>患者別の診療行為・医薬品・特定器材の利用別分析等 (研究に必要なデータ項目を含み個票)</p>

- 提供依頼申出者は、申出前に各ファイルのデータ項目を確認し、あらかじめ用意された上記データセットから、研究目的に応じて必要となるデータセットを全て申出する。

# 入院データセットの大まかなイメージ



# データセットDにおける医科診療行為、医薬品情報等の提供イメージ

## 【申出方法】

- 提供依頼申出者が、分析に必要とする医科診療行為、医薬品コード等を抽出条件に書き加える。

## 【データ提供方法】

- 依頼を受けたコードに関して、入院エピソード毎に、どれだけの回数、およびどれだけの点数が算定されたのかが合算された情報を提供する。
- 提供対象とするのは、「傷病名コード」「医科診療行為コード」「医薬品コード」「特定器材コード」の4種類。
- 最大でも、統計処理ソフトで簡便に処理が可能な**256行**までの範囲で、データを提供することとする。
- コードは1つだけ指定しても構わないし、具体的な指定があれば、それに対応したデータをひとまとまりにして指定しても構わない。

・在宅自己注射指導管理料(複雑な場合): 114021010  
を指定し、下記のインスリンの処方実態も確認することで、インスリンポンプ療法を受けている人がどの程度いるか、医療費はどれくらいか、等を評価する。

・ノボリンR注フレックスペン300単位: 620000265  
・ノボリンN注フレックスペン300単位: 620000266  
等々、インスリンの医薬品コードを指定し、インスリンの処方状況を確認する。

心臓バイパス手術に関連のあるコードをまとめて指定し、バイパス術の受療動向が確認する。

糖尿病治療薬であるスルホニル尿素剤の医薬品コードをまとめて指定し、処方の実態について確認する。

データセットB、Cに含まれない病名等の指定が可能。

行番号	データ項目名	記載事項
1	ID番号	
2	性別	男女区分を示す
3	年齢情報	5歳刻み
...	...	
	傷病名	傷病名を示す
	...	
...	114021010	○回、△点
...	インスリン	○回、△点
...	冠動脈バイパス術	○回、△点
256	糖尿病SU剤	○回、△点
	傷病名	

1回分の入院エピソード

レコードフォーマット		ファイル名	基本データセット	版数	作成日	作成者
				初版	2013/12/19	厚労科研 研究班
基本データセットB1（年度単位、「備考」欄は平成22年度版データセットを想定している）						
項番	データ項目名（日本語）	項目仕様	備考			
1	ID1					
2	ID2					
3	ID3(研究班試作)					
4	男女区分					
5	年齢階級コード(H22年次)		途中で変更があった場合は、最初のレコードをとる			
6	合計点数		項番9+13			
7	合計診療実日数		項番10+14			
8	受診医療機関数		項番11+15			
9	外来合計点数		H22年度の外来合計点数			
10	外来合計診療実日数		H22年度の外来合計日数			
11	外来医療機関数		H22年度のID毎の外来医療機関のDistinct			
12	調剤点数		H22年度の調剤合計点数			
13	入院回数		入院月日のDistinct			
14	入院合計点数		H22年度の入院合計点数			
15	入院合計診療実日数		H22年度の入院合計日数			
16	入院医療機関数		H22年のID毎の入院医療機関のDistinct			
17	主病名(外来)		エピソードでの最高点数レコードの主病名(外来)			
18	主病名(入院)		エピソードでの最高点数レコードの主病名(入院)			

基本データセットC1-入院（エピソード単位：診療行為点数ファイル）

項番	データ項目名（日本語）		条件等記入欄
1	ID1		
2	ID2		
3	ID3(研究班試作)		
4	男女区分		
5	年齢階級コード(H22年次)		
6	入院回数		
7	合計点数		
8	合計診療実日数(入院)		
9	医療機関数		
10	入院回数シーケンス番号		
11	入院-DPCフラグ		
12	合計点数		
13	診療実日数(入院)		
14	初診(1)		
15	医学管理		
16	在宅		
17	投薬（内服）		
18	投薬（頓服）		
19	投薬（外用）		
20	投薬（調剤）		
21	投薬（麻毒）		
22	投薬（調基）		
23	投薬（その他）		
24	注射（皮下筋肉内）		
25	注射（静脈内）		
26	注射（その他）		
27	薬剤料減点		
28	処置		
29	手術		
30	麻酔		
31	検査・病理		
32	画像診断		
33	その他		
34	入院（入院基本料）		
35	入院（特定入院料・その他）		
36	食事療養・生活療養・標準負担額		
37	主病名1		SYの主傷病(10)にフラグが付いていた傷病名コード(5)
38	診療開始日1		SYの主傷病(10)にフラグが付いていた診療開始日コード(6)
39	転帰区分1		SYの主傷病(10)にフラグが付いていた転帰区分(7)
40	主病名1(ICD10_大分類)		
41	主病名1(ICD10_中分類)		

基本データセットD1-入院（エピソード単位：診療行為明細ファイル）

項番	データ項目名（日本語）		条件等記入欄
1	ID1		C1と同じ
2	ID2		
3	ID3(研究班試作)		
4	男女区分		
5	年齢階級コード(H22年次)		
6	入院回数		
7	合計点数		
8	合計診療実日数(入院)		
9	医療機関数		
10	入院回数シーケンス番号		
11	入院-DPCフラグ		
12	合計点数		今回の入院の合計点数
13	診療実日数(入院)		今回の入院の入院日数
14	主病名1		SYの主傷病(10)にフラグが付いていた傷病名コード(5)
15	診療開始日1		SYの主傷病(10)にフラグが付いていた診療開始日コード(6)
16	転帰区分1		SYの主傷病(10)にフラグが付いていた転帰区分(7)
17	主病名1(ICD10_大分類)		
18	主病名1(ICD10_中分類)		
...	傷病名コード	オプション	傷病名コードが存在するかのフラグ等が設定可能
...	診療行為コード・フィールド	オプション	診療行為レコード(SI)の情報格納：データセットE
...	医薬品コード・フィールド	オプション	医薬品レコード(SI)の情報格納：データセットE
...	特定器材コード・フィールド	オプション	特定器材レコード(SI)の情報格納：データセットE
...	...	オプション	...
256	その他要望されたコード、コード群など	オプション	行数は最大256行とする