

NDBを用いた研究の実際（1）

サンプリングデータセットを用いた研究について

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（水）、東京大学 鉄門記念講堂



徳島文理大学香川薬学部 飯原 なおみ

1. サンプルングデータセットの概要
2. 受領したサンプルングデータセット
3. サンプルングデータセットの利点・弱点

* 参考資料

1. サンプルングデータセットの概要
2. 受領したサンプルングデータセット
3. サンプルングデータセットの利点・弱点

* 参考資料

サンプリングデータセットとは

高額レセプト除去後に、
一定の割合で抽出をかけたうえで、
出現頻度の低いコード^aを匿名化（特定のコードで代替）

^a 出現回数少ないコードから順に、総出現回数の0.1%^bに達するまで匿名化。

医薬品、傷病名、医科診療行為、DPC診断群分類のみ。

例えば、医薬品コードの場合「999999999」。

^b 医科診療行為については、0.01%以下。レセプト出現回数が100~200程度まで。

サンプリングデータセットは4種類

- 医科入院外
- 医科入院
- DPC
- 調剤

歯科レセプトのサンプリングデータセットは用意されていない。

サンプリングデータセットの特徴

○ 医科入院外

- 5万点以上のレセプト除去後、患者（ハッシュ値1）の1%を、性別、年齢層別（5歳刻み）に抽出。
 - 患者単位の分析可。
 - 調剤（同一月と翌月）を、ハッシュ値1で紐づけ可。

○ 医科入院

- 70万点以上のレセプト除去後、レセプト数の10%抽出。
 - 患者IDに相当する情報はなく、患者単位の分析できず。

○ DPC

- レセプト数の10%抽出。
 - 患者IDに相当する情報はなく、患者単位の分析できず。

○ 調剤

- 5万点以上のレセプト除去後、レセプト数の1%抽出。
 - 調剤単体の場合は、患者IDに相当する情報はなく、患者単位の分析できず。
 - 医科入院外との紐づけの場合は、患者単位の分析可能。

(参考)ある月における高額レセプトの状況

累計点数	出現頻度	
	医科入院	DPC
100,000～	4.5%	14.04%
200,000～	0.50%	2.57%
300,000～	0.16%	0.80%
400,000～	0.072%	0.33%
500,000～	0.036%	0.16%
600,000～	0.016%	0.063%
700,000～	0.008%	0.022%

累計点数	出現頻度	
	医科入院外	調剤
30,000～	0.40%	0.038%
40,000～	0.15%	0.021%
50,000～	0.042%	0.014%

用意されているデータセット

✓ 平成23年1月以降、毎年4回分*

*提供開始当初は、年1回、10月診療分のみ提供

H23: 1月診療分（医科入院外、医科入院、DPC、調剤）

4月診療分（医科入院外、医科入院、DPC、調剤）

7月診療分（医科入院外、医科入院、DPC、調剤）

10月診療分（医科入院外、医科入院、DPC、調剤）

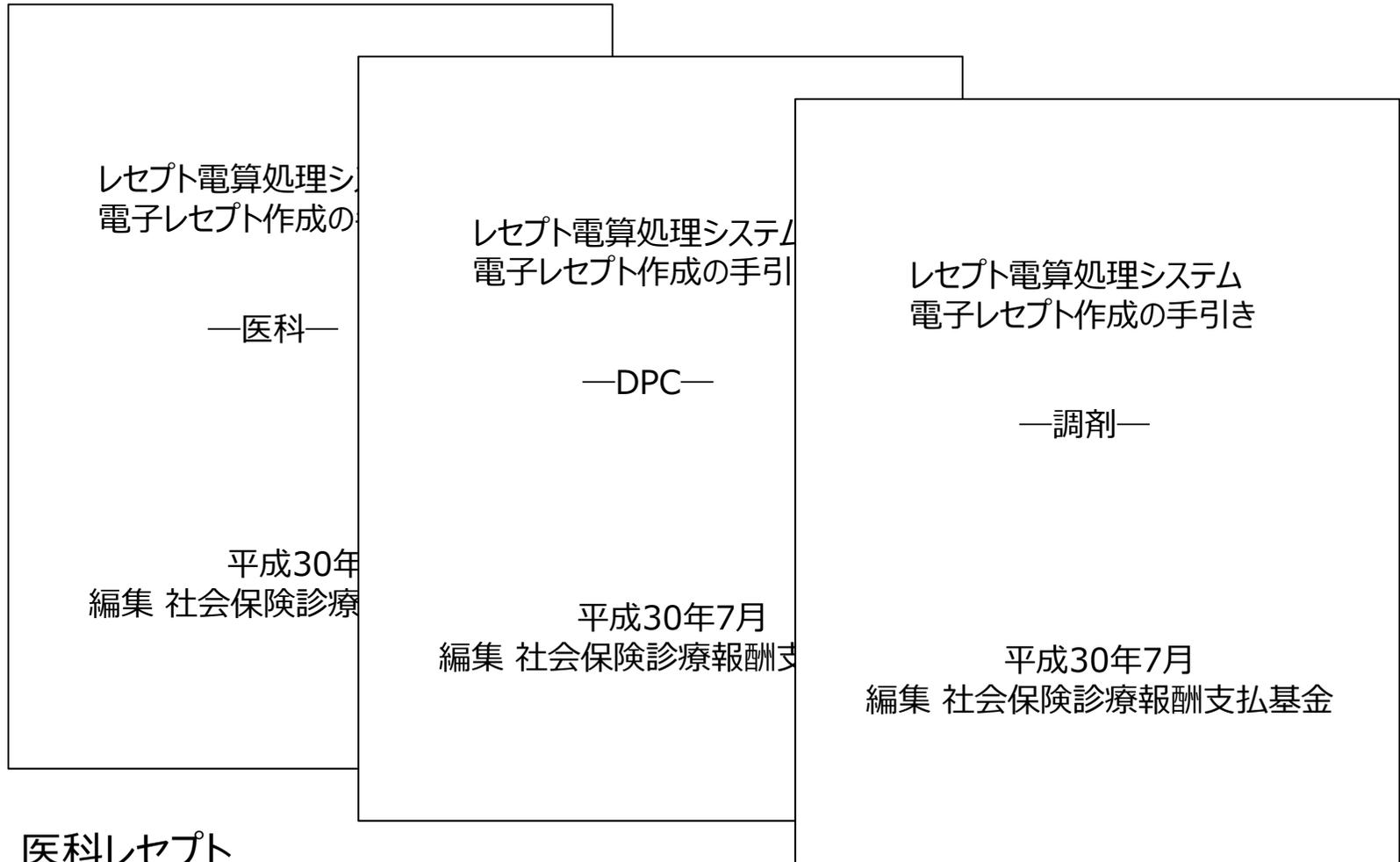
H24: 1月診療分（医科入院外、医科入院、DPC、調剤）

4月診療分（医科入院外、医科入院、DPC、調剤）

⋮ ⋮

→ 申し出時に指定（例えば「H24年4月診療分のDPCのみ」）。

一度に、全てのデータセットを依頼してもよい。



医科レセプト

IR 医療機関情報

RE レセプト共通 → レセプト種別（奇数は入院、偶数は入院外）

HO 保険者

IY 医薬品

SY 傷病名 → 修飾語コード（8002は疑い）

⋮ ⋮

医科レセ・レコードフォーマット：RE（レセプト共通）

レコード識別名：レセプト共通レコード (I_RE.csv)

項番	データ項目名（日本語）
1	通番1
2	通番2
3	レコード識別情報
4	空欄
5	空欄
6	診療年月
7	空欄もしくはID
8	男女区分
9	生年月日
10	空欄
11	入院年月日
12	病棟区分
13	一部負担金・食事療養費・生活療養費標準負担
14	空欄
15	病床数
16	空欄
17	空欄
18	予備
19	予備
20	旧診療科
21	検索番号
22	記録条件仕様年月情報
23	空欄
24	診療科名（診療科1）
25	人体の部位等（診療科1）
26	性別等（診療科1）
27	医学的処置（診療科1）
28	特定疾病（診療科1）
29	診療科名（診療科2）
30	人体の部位等（診療科2）
31	性別等（診療科2）
32	医学的処置（診療科2）
33	特定疾病（診療科2）
34	診療科名（診療科3）
35	人体の部位等（診療科3）
36	性別等（診療科3）
37	医学的処置（診療科3）
38	特定疾病（診療科3）

✓ 通常のレセプト項目に、以下が追加

1 通番1：1レセプト内での並び順

2 通番2：全レセプトの中で一意となる通番

→ 通番2でファイルを連結

✓ 空欄の項目、加工されている項目あり（赤字）

4・5・10・14・16・17・23空欄

レセプト番号、レセプト種別、給付割合。レセプト特記事項、カルテ番号、割引点数
単価、請求情報

参考資料④

医科レセ・レコードフォーマット：RE（レセプト共通）

レコード識別名：レセプト共通レコード (I_RE.csv)

項番	データ項目名（日本語）	属性	最大バイト数	項目形式	説明	備考
1	通番1	英数	10	固定	ファイル内の行番号	
2	通番2	英数	13	固定	レセプト単位の番号	
3	レコード識別情報	数字	2	可変		
4	空欄					
5	空欄					
6	診療年月	数字	5	固定		
7	空欄もしくはID				医科入院外のみハッシュ値IDを提供、その他の場合は空欄	
8	男女区分	数字	1	固定		
9	生年月日				年齢階級コードをセット	年齢階級コードは別シート参照
10	空欄					
11	入院年月日	数字	7	可変		
12	病棟区分	英数	8	可変		
13	一部負担金・食事療養費・生活療養費標準負担額区分	数字	1	可変		
14	空欄					
15	病床数	数字		可変	数字が記載されている場合はカテゴリー化した値をセット	カテゴリーコードは別シート参照
16	空欄					
17	空欄					
18	予備	数字	1	可変		
19	予備	数字	1	可変		
20	旧診療科	数字	2	可変		
21	検索番号	数字	30	可変		
22	記録条件仕様年月情報	数字	5	可変		
23	空欄					
24	診療科名（診療科1）	数字	2	可変		
25	人体の部位等（診療科1）	数字	3	可変		
26	性別等（診療科1）	数字	3	可変		
27	医学的処置（診療科1）	数字	3	可変		
28	特定疾病（診療科1）	数字	3	可変		
29	診療科名（診療科2）	数字	2	可変		
30	人体の部位等（診療科2）	数字	3	可変		
31	性別等（診療科2）	数字	3	可変		
32	医学的処置（診療科2）	数字	3	可変		
33	特定疾病（診療科2）	数字	3	可変		
34	診療科名（診療科3）	数字	2	可変		
35	人体の部位等（診療科3）	数字	3	可変		
36	性別等（診療科3）	数字	3	可変		
37	医学的処置（診療科3）	数字	3	可変		
38	特定疾病（診療科3）	数字	3	可変		

4・5・10・14・16・17・23空欄

レセプト番号、レセプト種別、給付割合、レセプト特記事項、カルテ番号、割引点数
単価、請求情報

参考資料④

医科レセ・レコードフォーマット：RE（レセプト共通）

レコード識別名：レセプト共通レコード（I_RE.csv）

項番	データ項目名（日本語）	属性	最大 バイト 数	項目 形式	
1	通番1	英数	10	固定	ファ
2	通番2	英数	13	固定	レセ
3	レコード識別情報	数字	2	可変	
4	空欄				
5	空欄				
6	診療年月	数字	5	固定	
7	空欄もしくはID				医科
8	男女区分	数字	1	固定	
9	生年月日				年齢
10	空欄				
11	入院年月日	数字	7	可変	
12	病棟区分	英数	8	可変	
13	一部負担金・食事療養費・生活療養費標準負担額区分	数字	1	可変	
14	空欄				
15	病床数	数字		可変	数字が
16	空欄				
17	空欄				

7 空欄もしくはID：医科入院外のみハッシュ値1がハッシュ値IDとして提供。

医科入院は空欄。

9 生年月日：年齢階級コード（5歳刻み）

15 病床数：病床数（200床刻み、800床以上）

医科レセ・レコードフォーマット：IY（医薬品）

レコード識別名： 医薬品レコード(I_IY.csv)

項番	データ項目名（日本語）	属性	最大 バイト 数	項目 形式	
1	通番 1	英数	10	固定	ファ
2	通番 2	英数	13	固定	レセ
3	レコード識別情報	英数	2	固定	
4	診療識別	数字	2	可変	
5	空欄				
6	医薬品コード	数字	9	固定	希少
7	使用量	英数	11	可変	
8	点数	数字	7	可変	希少
9	回数	数字	4	可変	
10	コメントコード1（コメント				
11	空欄				
12	コメントコード2（コメント				
13	空欄				
14	コメントコード3（コメント				
15	空欄				
16	1日の情報				
17	2日の情報				
18	3日の情報				
19	4日の情報				

→ 通番 2 あり

5 空欄：負担区分

6 医薬品コード：希少例では匿名化处理

8 点数：希少例では削除

11・13・15 空欄：コメントに対する文字データ

16以降：○日の情報

（H23年度は空白、H24年度以降は算定日あり）

医科レセ・レコードフォーマット：SY（傷病名）

レコード識別名： 傷病名レコード (I_SY.csv)

項番	データ項目名（日本語）	属性	最大 バイト 数	項目 形式	
1	通番 1	英数	10	固定	コ
2	通番 2 → 通番 2 あり	英数	13	固定	シ
3	レコード識別情報	英数	2	固定	
4	傷病名コード	数字	7	固定	病
5	診療開始日	数字	7	固定	
6	転帰区分	数字	1	固定	
7	修飾語コード	英数	80	可変	
8	空欄				
9	主傷病	数字	2	可変	
10	空欄				

- 4 傷病名コード：希少例では匿名化处理
- 8 空欄：傷病名コード「000999」の時の傷病名称
- 10 空欄：補足コメント

医科レセ・レコードフォーマット：IR（医療機関情報）

レコード識別名：医療機関情報レコード（I_IR.csv）

項番	データ項目名（日本語）	属性	最大 バイト 数	項目 形式	
1	通番 1	英数	10	固定	ファイル レセプト
2	通番 2	英数	13	固定	
3	レコード識別情報	英数	2	固定	
4	審査支払機関	数字	1	固定	
5	空欄				
6	点数表	数字	1	固定	
7	空欄				
8	予備	数字	2	可変	
9	空欄				
10	請求年月	数字	5	固定	
11	マルチボリューム識別情報	数字	2	固定	
12	空欄				

5 空欄：都道府県

7 空欄：医療機関コード

9 空欄：医療機関名称

12 空欄：電話番号

4 審査支払機関

1 社会保険診療報酬支払基金

2 国民健康保険団体連合会

空欄・加工された項目の一覧

医療レセプト(入院、入院外とも)					
レコード 識別情報	項番	項目名	匿名化内容		
IR	5	都道府県	空欄		
	7	医療機関コード	空欄		
	9	医療機関名称	空欄		
	12	電話番号	空欄		
RE	4	レセプト番号	空欄		
	5	レセプト種別	空欄		
	7	DPCレセプト			
	9	レコード 識別情報	項番	項目名	匿名化内容
	10	IR	5	都道府県	空欄
	14		7	医療機関コード	空欄
	15		9	医療機関名称	空欄
	16		12	電話番号	空欄
	17		4	レセプト番号	空欄
	23	5	レセプト種別	空欄	
HO	4	RE	7	氏名	医療入院外レセプトに同梱される場合はハッシュ値IDを提供、その他の場合は空欄
	5		9	生年月日	年齢階級コードをセット
	6		10	給付割合	空欄
	12		11	レセプト特記事項	空欄
	13		12	都道府県(医療機関名称・所在地)	空欄
	14		14	医療機関(医療機関名称・所在地)	空欄
	15		15	名称(医療機関名称・所在地)	空欄
	16		16	所在地(医療機関名称・所在地)	空欄
KO	4	HO	17	保険医氏名1	空欄
	5		6	負担	空欄
	6		7	受給	空欄
	9		8	任意	空欄
	10		9	任意	空欄
SY	4	KO	4	負担	空欄
	5		5	受給	空欄
	6		6	任意	空欄
			20	保険医氏名4	空欄
			21	保険医氏名5	空欄

1. サンプルングデータセットの概要
2. 受領したサンプルングデータセット
3. サンプルングデータセットの利点・弱点

* 参考資料

受領したサンプリングデータセット H23.10 医科入院外＋調剤

医科入院外 (H23.10)
1.41GB
105万件のレセプト



1_IY.csv … 通番 2
1_RE.csv … 通番 2、氏名 (ハッシュ値ID)
1_SI.csv … 通番 2
1_SY.csv … 通番 2

調剤レセプト (H23.10)
680MB
63万件のレセプト



4_CZ.csv … 通番 2
4_IY.csv … 通番 2
4_RE.csv … 通番 2、氏名 (ハッシュ値ID)
4_SH.csv … 通番 2

調剤レセプト (H23.11)
557MB
50万件のレセプト

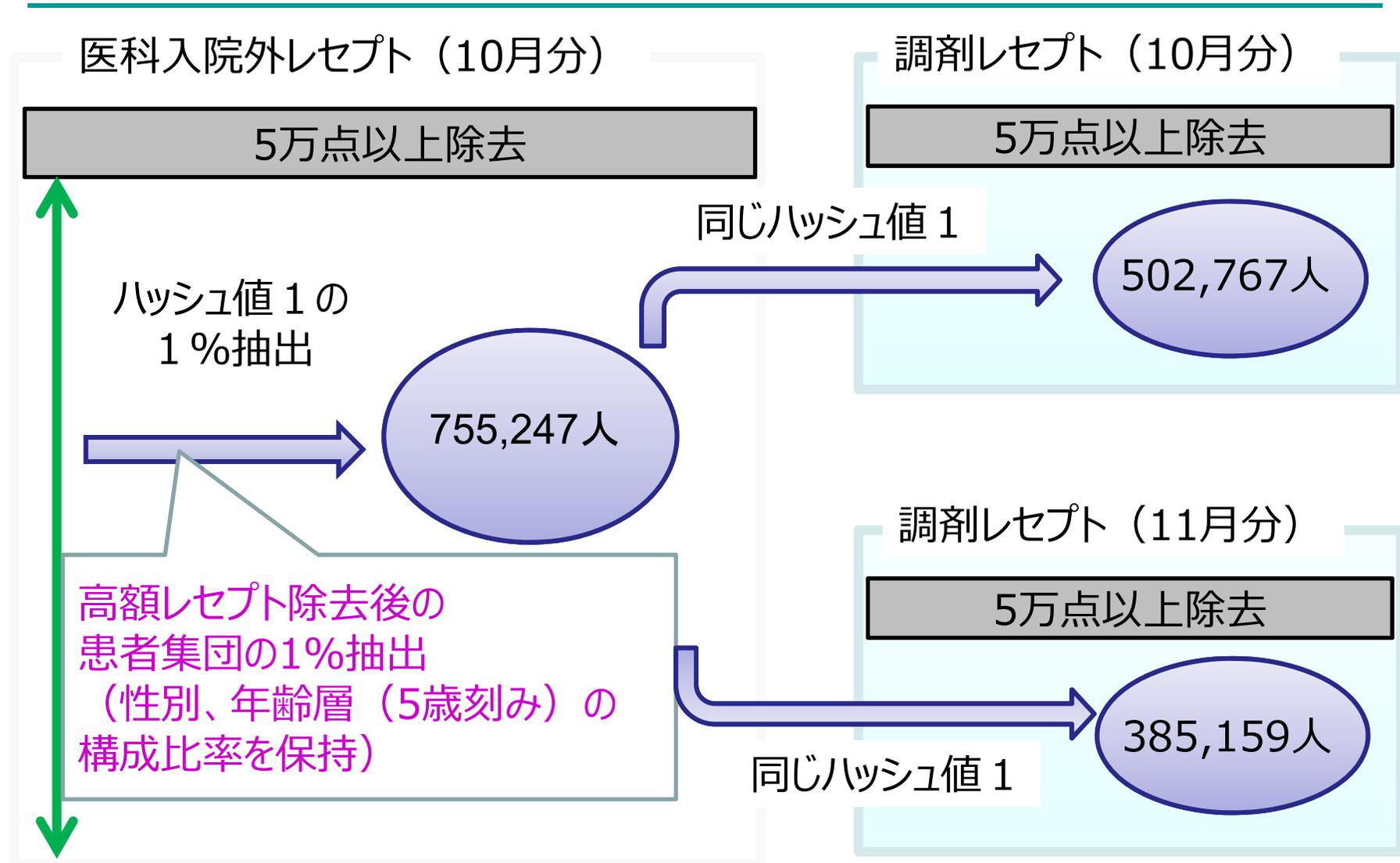


4_CZ.csv … 通番 2
4_IY.csv … 通番 2
4_RE.csv … 通番 2、氏名 (ハッシュ値ID)
4_SH.csv … 通番 2

→ 通番 2 で、各ファイルを連結

→ ハッシュ値IDで、医科入院外と調剤とを連結

受領したサンプリングデータセット：医科入院外＋調剤



→ わが国の外来患者のほぼ1%サンプリング

755,247人はどんな患者？ 保険

年齢層	社会保険診療報酬支払基金 n=337,370人	国民健康保険団体連合会 n=417,877人
0-4歳	11.0%	2.0%
5-9歳	7.7%	1.5%
10-14歳	5.6%	1.2%
15-19歳	4.1%	0.9%
20-24歳	4.1%	0.9%
25-29歳	5.3%	1.2%
30-34歳	6.3%	1.4%
35-39歳	7.6%	1.8%
40-44歳	7.5%	1.9%
45-49歳	7.4%	1.8%
50-54歳	8.1%	2.2%
55-59歳	9.0%	3.5%
60-64歳	9.4%	9.4%
65-69歳	4.0%	12.4%
70-74歳	2.9%	15.7%
75-79歳	<0.1%	18.0%
80-84歳	—	13.5%
85歳以上	—	10.6%

→ 両審査支払機関あり

→ 保険種別の影響を
受けない

→ サンプリングデータセットは
後期高齢者も対象

各ファイルの内容 医科入院外 RE

通番1	通番2	レコード識別	レセプト番号	レセプト種別	診療年月	氏名	男女区分	生年月日	給付割合	...
123	1_01	RE	空欄	空欄	42310	53	1	3	空欄	
567	1_02	RE	空欄	空欄	42310	79	2	11	空欄	
...										

ハッシュ値ID

「データベース内の通し番号」ではなく、「1レセプト内の通し番号」

4 (平成) 23年10月

1 (男性)
2 (女性)

3 (10-14歳)
11 (50-54歳)

「診療科」空白はレセプト数の73%

「病床数」空白はレセプト数の73%

→ 「診療科」「病床数」は使えず

各ファイルの内容 医科入院外 IY

回数記入は処方最後の医薬品のみ（H23年度*）

通番1	通番2	レコード識別	診療識別	負担区分	医薬品コード	使用量	点数	回数	...
136	1_01	IY	21	空欄	620425801	3	空欄	空欄	
137	1_01	IY	21	空欄	620421201	3	4	2	
138	1_01	IY	22	空欄	620002038	1	1	2	
...									

21（内服）
22（頓用）

620425801（フェロベリン配合錠） 1日3錠 2日分
620421201（ビオスリー配合錠） 1日3錠 2日分
620002038（コカール錠200mg） 1回1錠 2回分（頓服）

「22（頓用）」において、
デゾラム錠0.5mg 56錠、セルシン錠2mg 20錠あり
→「22（頓用）」の使用量は【1回量】【全量】の両者あり

* H24年度以降は、回数は全ての行にあり。

各ファイルの内容 医科入院外 IY「処方日数」

処方日数	処方件数*1 n=548,077件	処方全件数*1に 占める割合(累積)
1-10日	143,472件	26.2%(26.2%)
11-20日	62,252件	11.4%(37.5%)
21-30日(~1カ月)	254,732件	46.5%(84.0%)
31-60日(~2カ月)	75,081件	13.7%(97.7%)
60-90日(~3カ月)	10,696件	2.0%(99.7%)
91-120日(~4カ月)	1,685件	0.31%(99.97%)
130-150日(~5カ月)	54件	0.01%(99.98%)
151-180日(~6カ月)	100件	0.02%(100.0%)
181-210日(~7カ月)	1件	0.0002%(100.0%)
211-270日(~9カ月)	1件	0.0002%(100.0%)
271-360日(~1年)*2	3件	0.0005%(100.0%)

*1 診療識別 21 (内服) の処方の件数

*2 処方日数の最大値は360日

→ 1カ月で
84.0%の処方をカバー

→ 4か月で
99.97%の処方をカバー

→ 併用薬を評価するには
1年間必要

サンプリングデータセットは代表性に優れており
自分の研究の位置づけや限界を教えてくれる

各ファイルの内容 調剤RE, SH, CZ, IY

通番1	通番2	レコード識別	レセプト番号	レセプト種別	調剤年月	氏名	男女区分	生年月日	...
336	1_41	RE	空欄	空欄	42310	75	1	13	

通番1	通番2	レコード識別	No	剤形コード	用法コード	特別指示	...
338	1_41	SH	1	1	3	空欄	

通番1	通番2	レコード識別	医師番号	処方月日	調剤月日	処方せん受付回	調剤数量	...
339	1_41	CZ	空欄	4231005	4231005	1	7	

通番1	通番2	レコード識別	負担区分	医薬品コード	使用量	...
340	1_41	IY	空欄	611240248	3	

→ 処方内容（赤字）を、医薬品毎に1行のレコードに

→ 「氏名：75」、男、60-64歳、「通番1：340」、「通番2：1_41」、処方No.1、内服、ミオナール錠50mg、3錠/日、7日分、1回目、処方日：H23年10月5日、調剤日：H23年10月5日

「抽出対象集計表」と「匿名化対象コード一覧」も提供

✓ レセプトファイル

- ・ 医科入院外 (CO, HO, IR, IY, KO, NI, RE, ...)
- ・ 調剤 10月診療 (CO, CZ, HO, IY, KI, KO, RE, ...)
- ・ 調剤 11月診療 (CO, CZ, HO, IY, KI, KO, RE, ...)

✓ 抽出対象集計表

医科外来				調剤(10月診療)				調剤(11月診療)		
年齢	性別	母数	抽出後件数	年齢	性別	母数	抽出後件数	年齢	性別	母数
0-4歳	男性	〇〇〇	〇〇〇	0-4歳	男性		〇〇〇	0-4歳	男性	
	女性	〇〇〇	〇〇〇		女性		〇〇〇		女性	
5-9歳	男性	〇〇〇	〇〇〇	5-9歳	男性		〇〇〇	5-9歳	男性	
	女性	〇〇〇	〇〇〇		女性		〇〇〇		女性	
10-14歳	男性	〇〇〇	〇〇〇	10-14歳	男性		〇〇〇	10-14歳	男性	
	女性	〇〇〇	〇〇〇		女性		〇〇〇		女性	
⋮	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	

✓ 匿名化対象コード一覧 (医薬品コード、傷病名コード、診療行為コード)

1. サンプルングデータセットの概要
2. 受領したサンプルングデータセット
3. サンプルングデータセットの利点・弱点

* 参考資料

サンプリングデータセットの利点

1. 医科入院外レセ：患者単位で処理でき、代表性に優れている
2. 多くの項目あり → 疑問点の確認が容易

(特別抽出：申出者が指定した項目のみ)

3. 通常PCで処理可能 (Win64bit 32GBメモリ)
4. 特別抽出に比べて
 - ・セキュリティ要件緩い
 - ・準備すべき申請書類少ない

(「倫理委員会承諾書」「詳細な抽出条件」「詳細な公表形式」不要)

- ・ 探索的研究内容で申請できる
5. すぐに提供される (抽出済みのデータセットを複写して提供)
 6. 実地監査は原則なし

サンプリングデータセットの弱点

1. 高額レセプトは除去

→ どういった疾病の患者で高額レセプトが多いのか、レセプトを日常扱っていないとわからない。

2. 出現頻度の少ない傷病名や医薬品などは匿名化

→ 匿名化された傷病名や医薬品名が何かは、データ提供後にわかり、申請時にはわからない。

3. 規定の項目のみの提供

→ 例えば、地域単位の分析には不向き。

4. 1か月分のデータ

→ 時系列の分析には不向き。

サンプリングデータセットの弱点

Q. 同じ人物が複数の病院にかかり、
A病院から高額レセプトが提出され、
B病院からは高額ではないレセプトが提出された場合、
B病院からのレセプトは抽出対象となるのか？

A . 高額レセプトでないものは抽出対象
(この例だとB病院からのレセプトは抽出対象) 。

ただし、サンプリングデータセットは、抽出対象となったデータを母集団とし
一定の割合で絞り込んだ結果の提供なので、この例において必ずしもB
病院のレセプトが提供データに含まれるとは限りません。

サンプリングデータセットに適した研究

1. ある時点の患者像、治療、医薬品使用についての概要把握
 - これら概要の推移の把握
 - 医科入院外は、性、年齢層別の1%抽出であり、
外来患者の特徴やその推移の把握
2. 医科入院外と調剤とを紐づけできるので、
外来患者の医薬品使用実態

参考資料

① レセプト情報・特定健診等情報の提供に関するホームページ

a 「利用を検討している方々へのマニュアル PDF」

b 「レセプト情報等の第三者提供の手続き等について（第2部） YouTube」

c 「サンプリングデータセットの使用経験から YouTube」

d 「FAQ 4. 「サンプリングデータセット」に関するご質問 PDF」

② 第9回有識者会議 資料1 基本データセットについて

<https://www.mhlw.go.jp/stf/shingi/2r98520000024d12-att/2r98520000024d5x.pdf>

③ 第1回提供サンプリングデータセットの仕様

<https://www.mhlw.go.jp/stf/shingi/2r9852000002911e-att/2r9852000002912v.pdf>

④ レコードフォーマット

<https://www.mhlw.go.jp/stf/shingi/2r98520000029hp2-att/2r98520000029hso.pdf>

⑤ 社会保険診療報酬支払基金ホームページ/電子レセプトの作成

https://www.ssk.or.jp/seikyushiharai/rezept/iryokikan/iryokikan_02.html

ご清聴、有難うございました

