

Table 2

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
Xp	1,643	1,003	16	143	218	152	142
CT・MRI	2,814	2,083	132	513	230	233	107
核医学	5,917	3,324	1,880	2,618	30	18	5
造影							

【再掲】

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
CT	1,096	607	114	375	114	81	70
MRI	1,453	261	25	141	127	61	98
二回目～	1,133	606	139	410	153	68	27

医
療
費

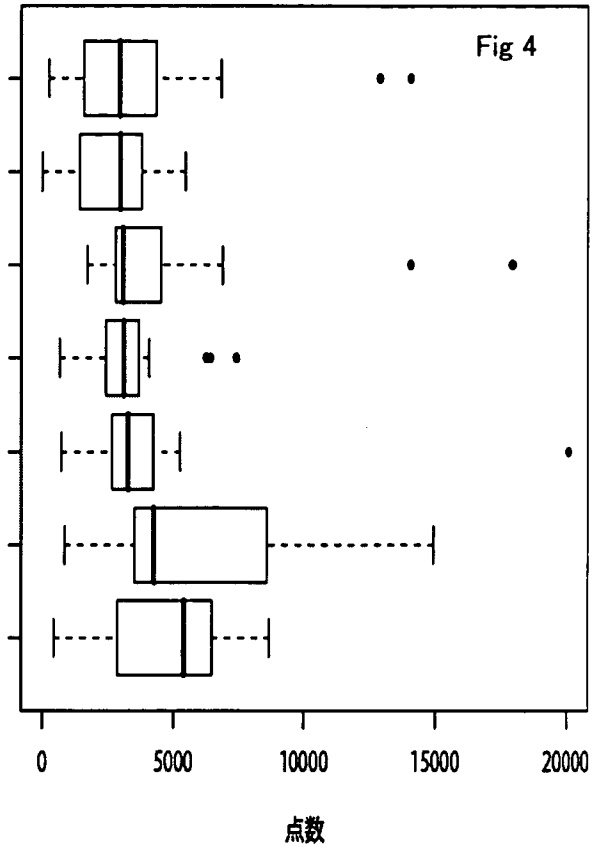


Fig 5

1606903x99xxxx

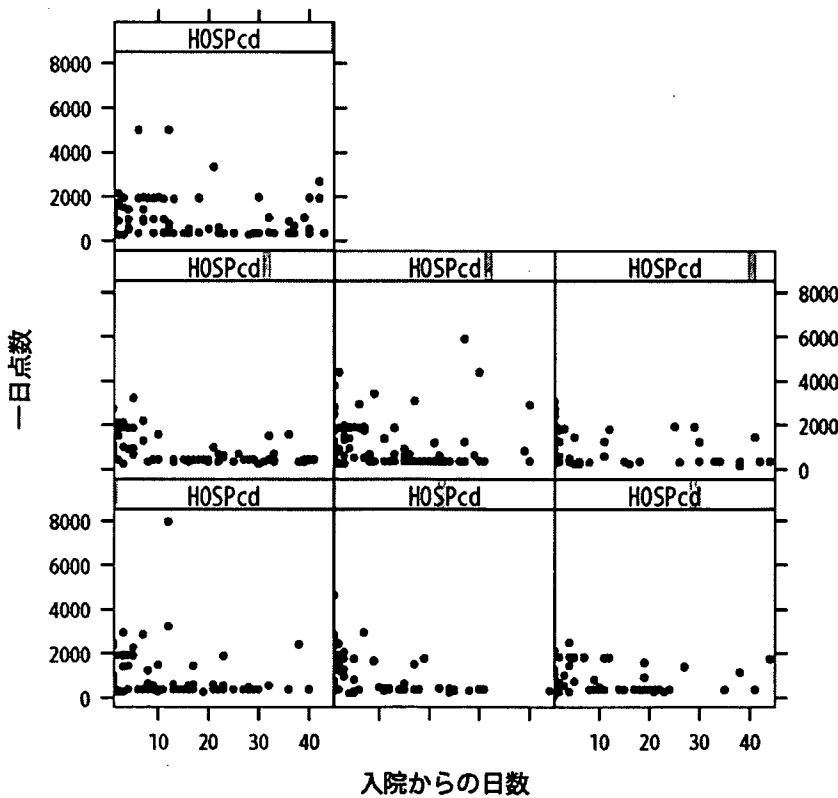
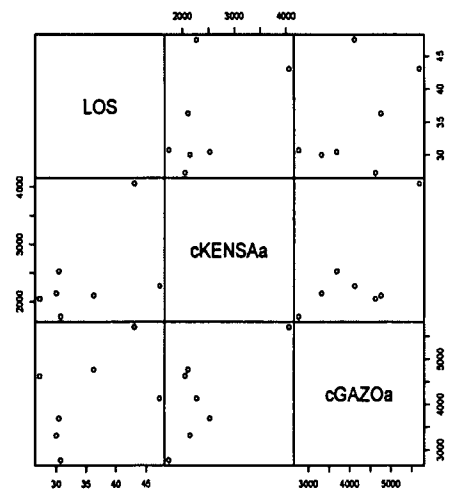


Fig 6



1606903x99xxxx

胸椎、腰椎以下骨折損傷(胸・腰髄損傷を含む。)、手術なし

患者数: 1,853

Table 1

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
検体	1,472	893					530
細菌	291	252					112
病理	894	574					28
生理	527	595	1	7		10	540
超音波	1,077	462					289
内視鏡	1,568	808	89	126	47	67	30
AG	10,559	887	2,790	488	2,369	400	2
穿刺	60	85		1			178
RI							
骨塩	360						7
判断料	1,290	1,123					555
心理	280						1

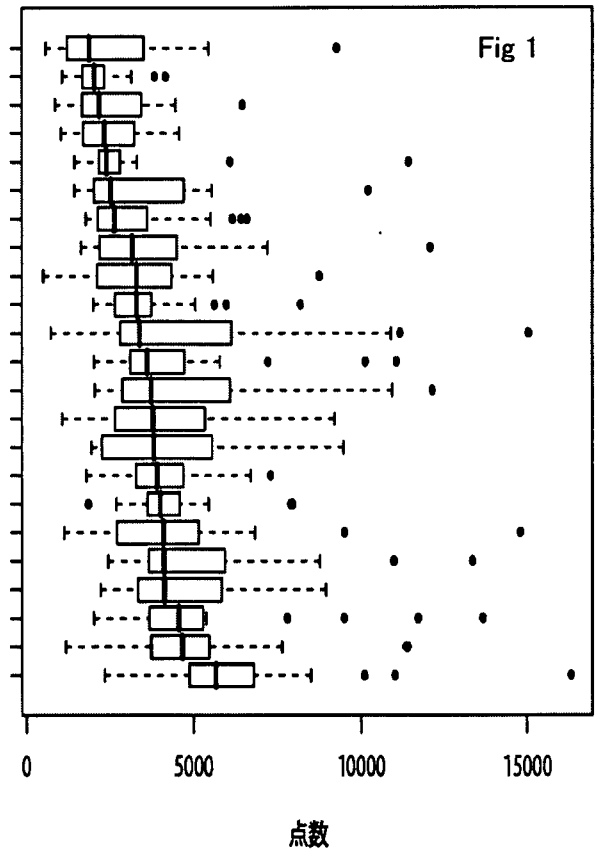


Fig 2

1608003x02xx0x

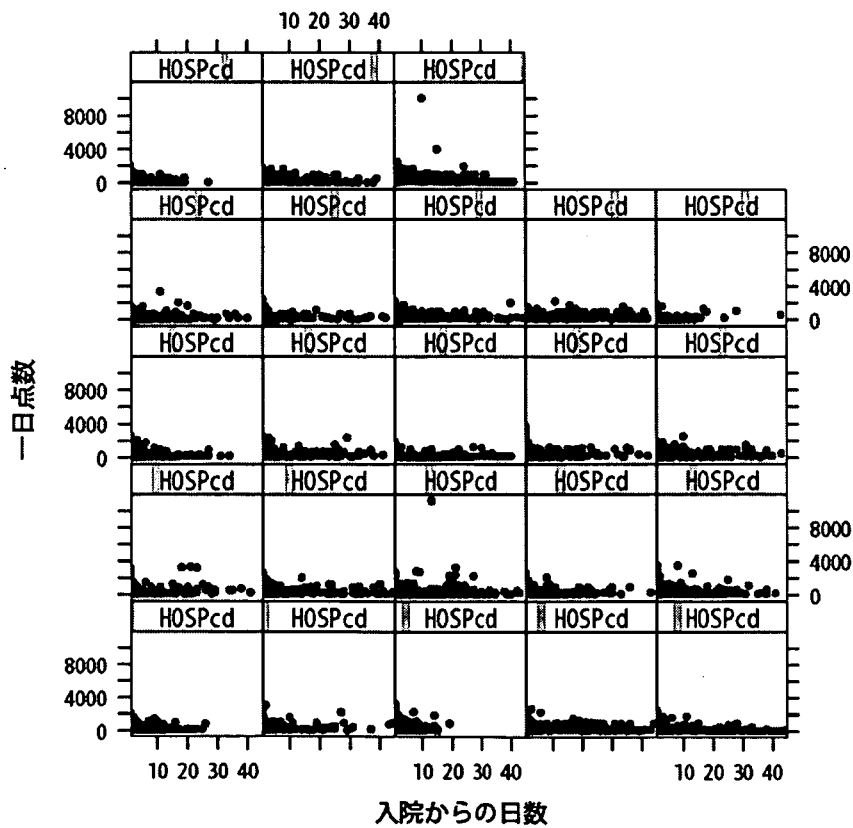


Fig 3

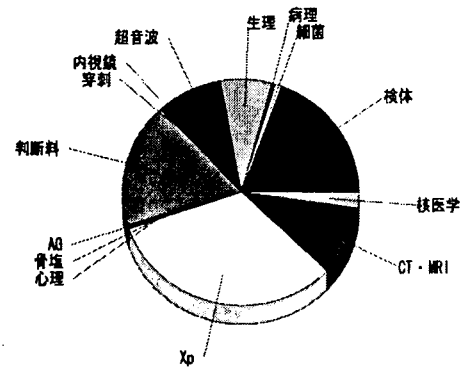


Table 2

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
Xp	2,248	1,379	1	23	286	216	583
CT・MRI	2,265	2,026	319	811	154	193	176
核医学	6,211	2,698	3,525	2,146	55	30	15
造影							

【再掲】

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
CT	1,205	763	217	489	102	89	198
MRI	1,459	650	148	403	141	79	40
二回目～	1,416	990	432	635	80	99	17

病院

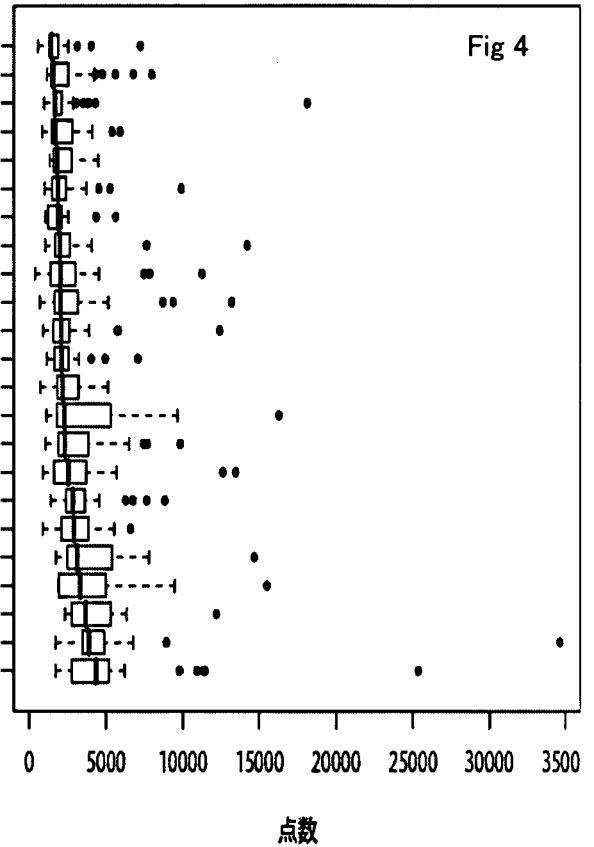


Fig 4

Fig 5

1608003x02xx0x

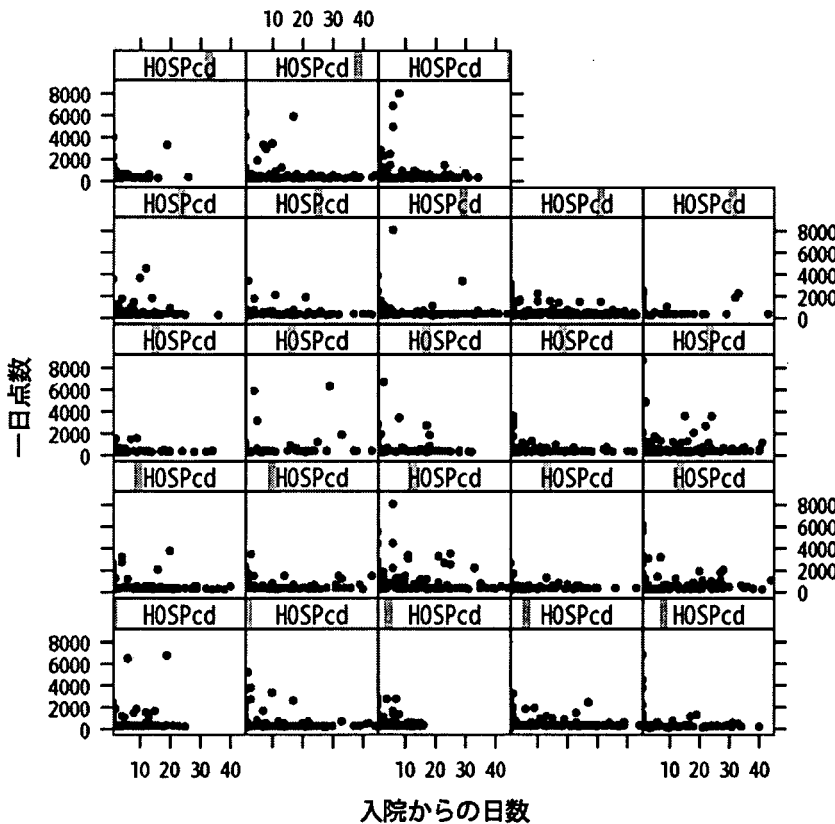
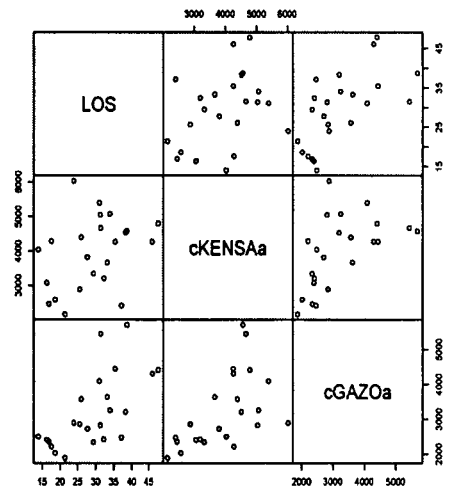


Fig 6



1608003x02xx0x

股関節大腿近位骨折、手術関節脱臼観血的整復術 肩、股、膝等、副傷病なし

患者数: 2,549

Table 1

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
検体	930	1,935					838
細菌	461	558					211
病理	1,474	1,211					131
生理	652	991	1	6	7	80	817
超音波	1,085	1,696	8	115	84	1,192	200
内視鏡	1,457	962	64	132	33	67	52
AG	296	960					24
穿刺	192	479	2	8			176
RI							
骨塩	336	73					9
判断料	2,287	3,094					946
心理	730						1

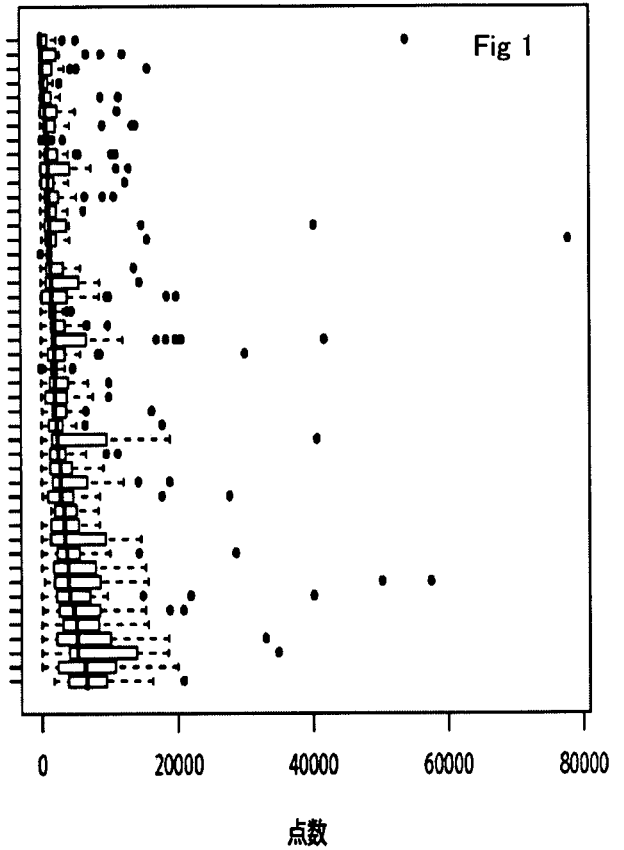


Fig 2

1610603x97x0xx

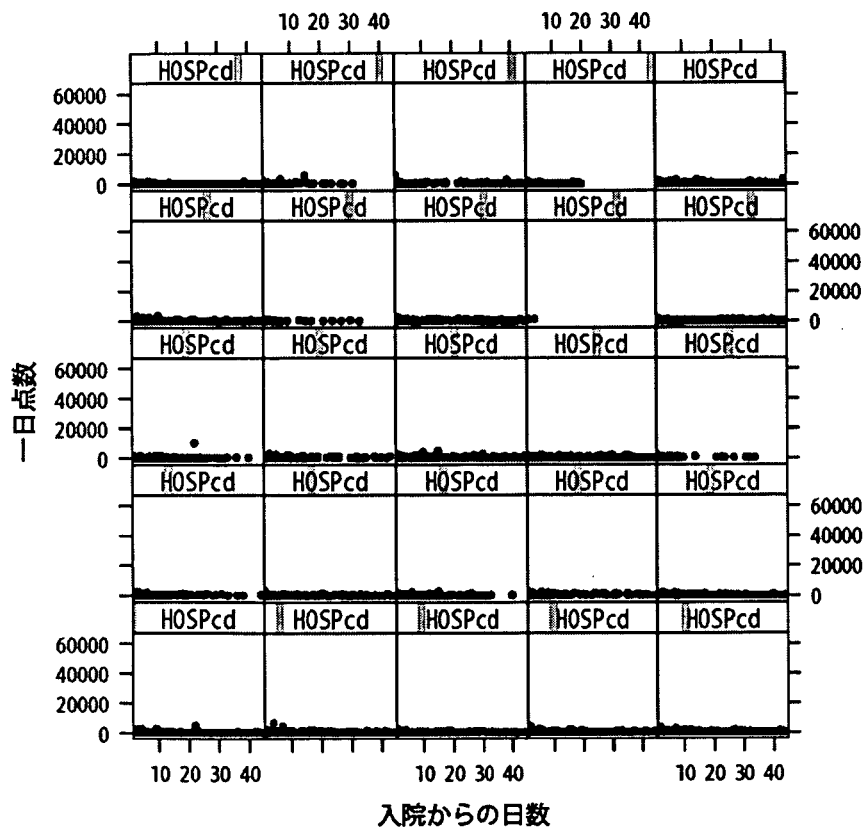


Fig 3

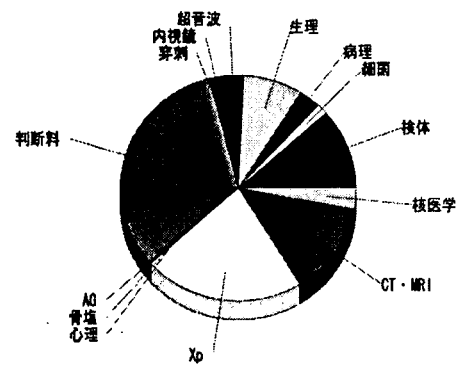


Table 2

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
Xp	1,697	2,117	141	464	222	570	865
CT・MRI	3,177	2,321	628	1,014	232	262	276
核医学	6,234	1,612	3,369	1,206	39	39	34
造影							

【再掲】

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
CT	1,375	961	349	605	116	113	387
MRI	1,705	698	277	491	161	91	57
二回目～	1,071	747	208	501	87	91	98

自然

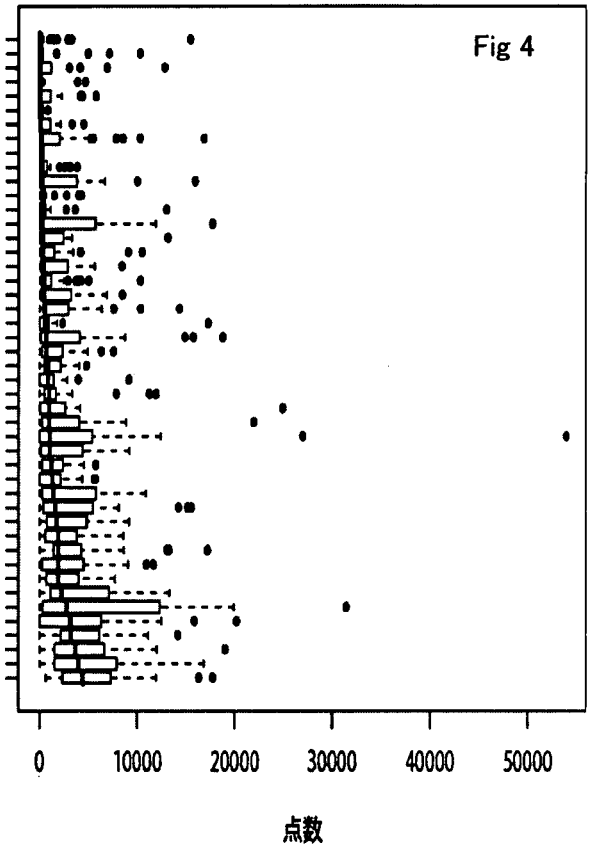


Fig 5

1610603x97x0xx

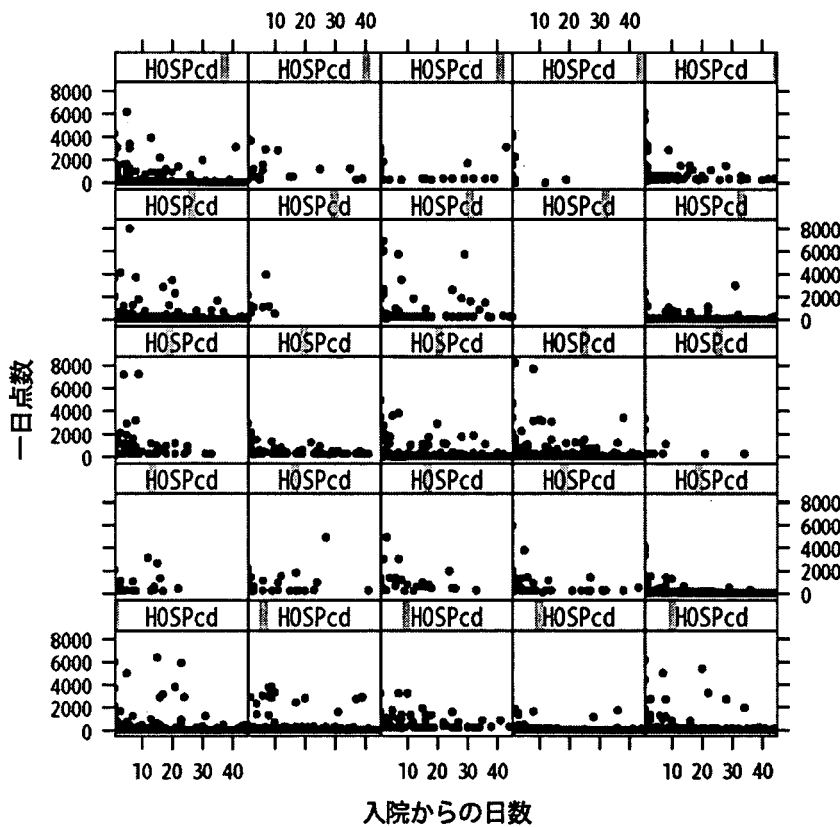
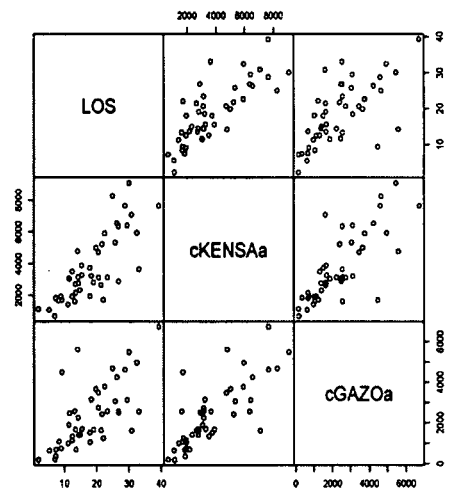


Fig 6



1610603x97x0xx

詳細不明の損傷等、手術あり、手術・処置2なし

患者数: 1,781

Table 1

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
検体	1,157	1,850		1			206
細菌	463	535					42
病理	2,364	1,984	15	20			44
生理	550	685	3	12			180
超音波	951	513					38
内視鏡	1,275	582	31	50	111	502	52
AG							
穿刺	605	1,305	2	6			57
RI							
骨塩	213	127					3
判断料	1,642	2,079					248
心理							

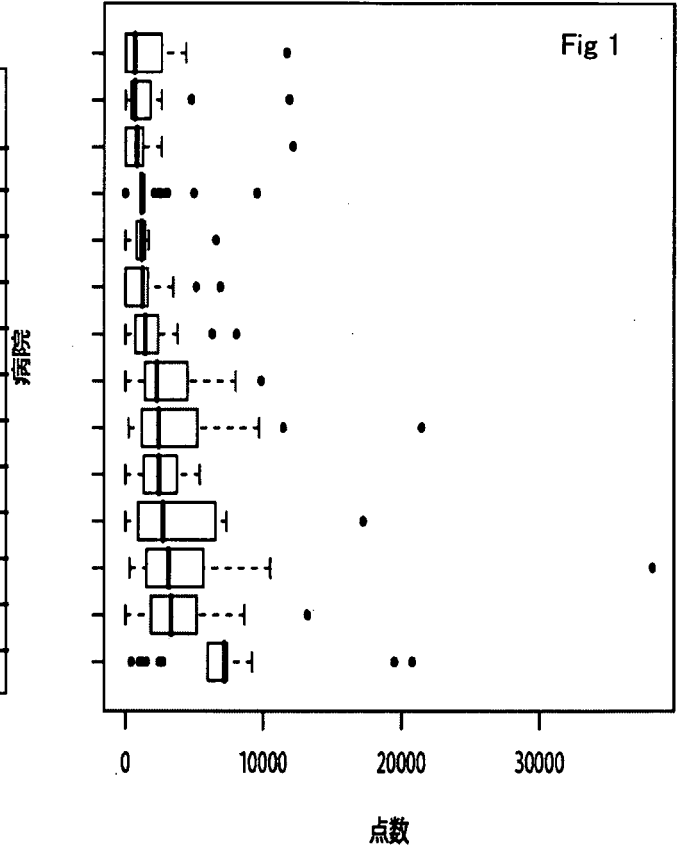


Fig 2

1610603x99x0xx

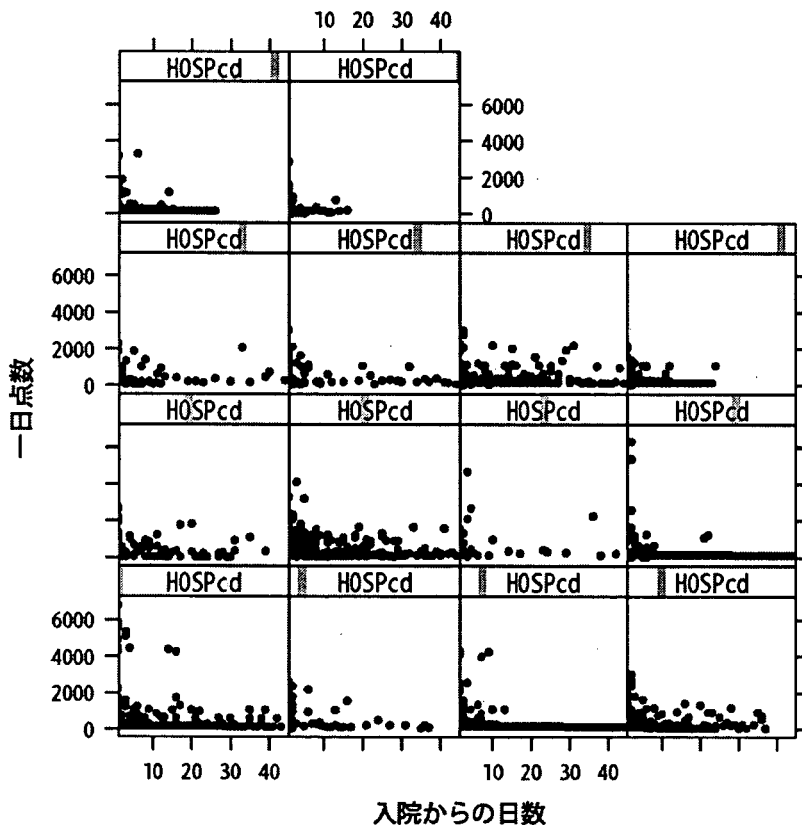


Fig 3



Table 2

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
Xp	1,114	1,817	72	400	163	569	214
CT・MRI	3,106	2,441	580	935	292	274	81
核医学	7,856	3,821	4,144	2,613	45	20	6
造影							

【再掲】

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
CT	1,551	956	402	630	159	108	89
MRI	1,333	436	53	247	160	57	22
二回目～	1,376	838	303	568	163	100	33

病院

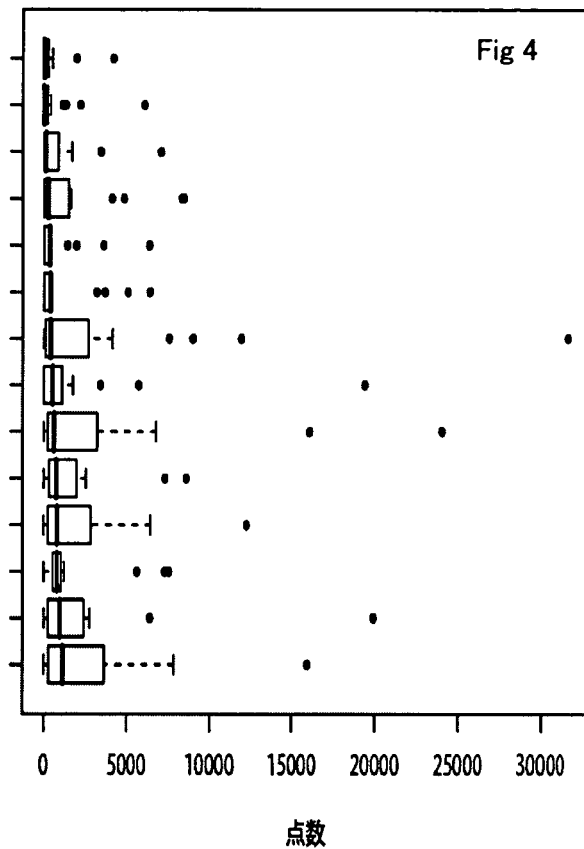


Fig 5

1610603x99x0xx

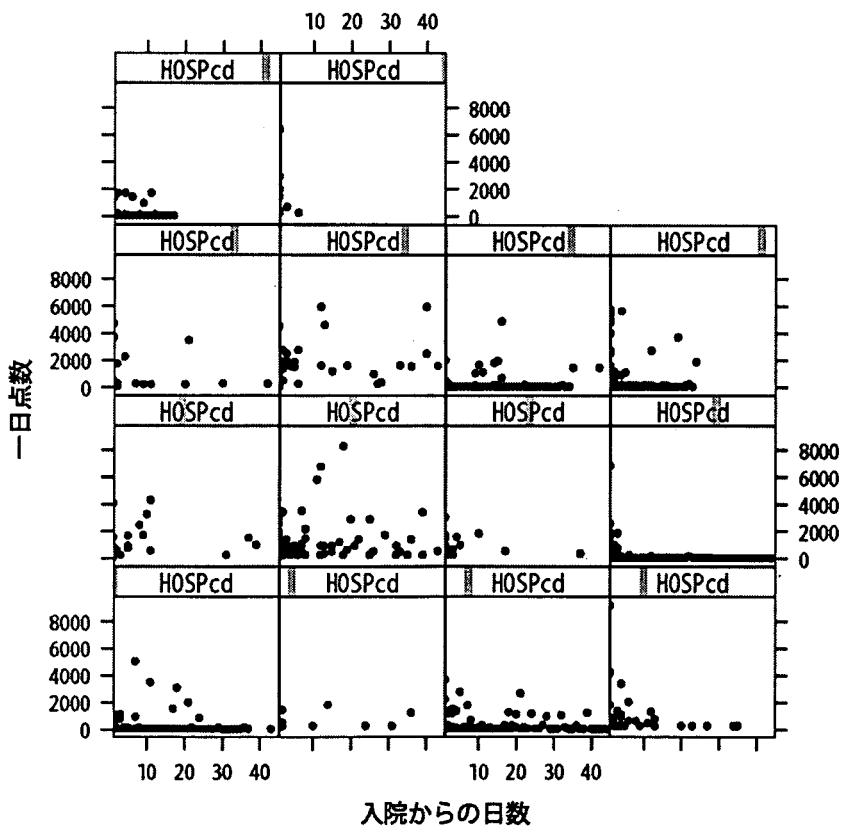
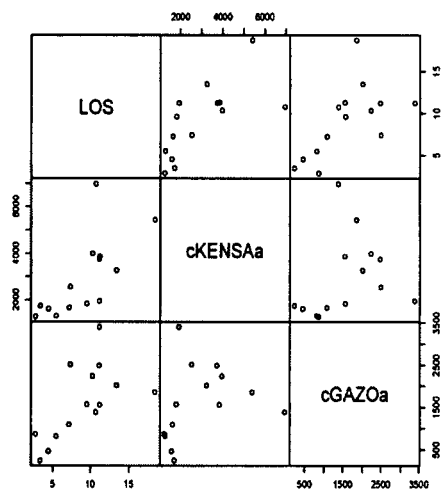


Fig 6



1610603x99x0xx

詳細不明の損傷等、手術なし、手術・処置2なし

患者数: 2,483

Table 1

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
検体	383	491			1	8	540
細菌	74	151					47
病理	799	418					5
生理	309	413					503
超音波	726	325					29
内視鏡	520	673	10	38	26	53	65
AG							
穿刺	26	62					256
RI	110						1
骨塩							
判断料	859	410					1,013
心理	80						1

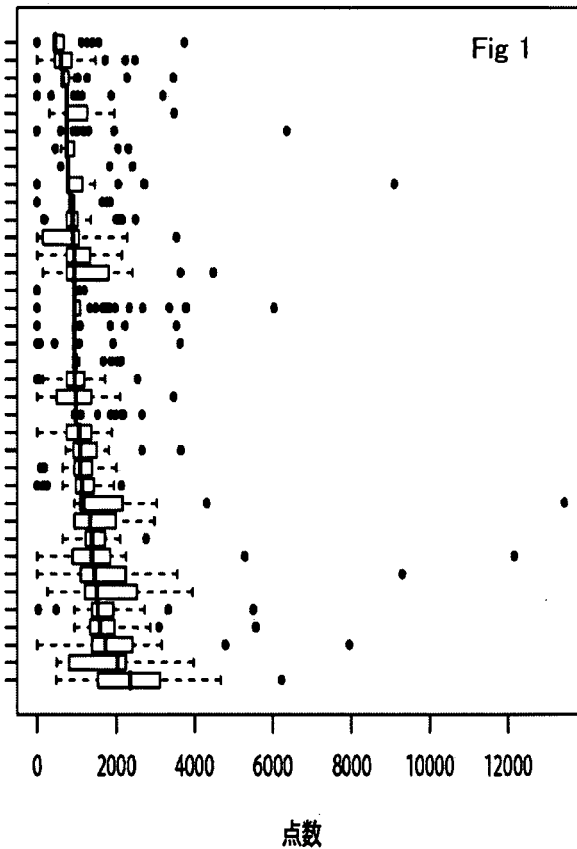


Fig 2

1610703x99x00x

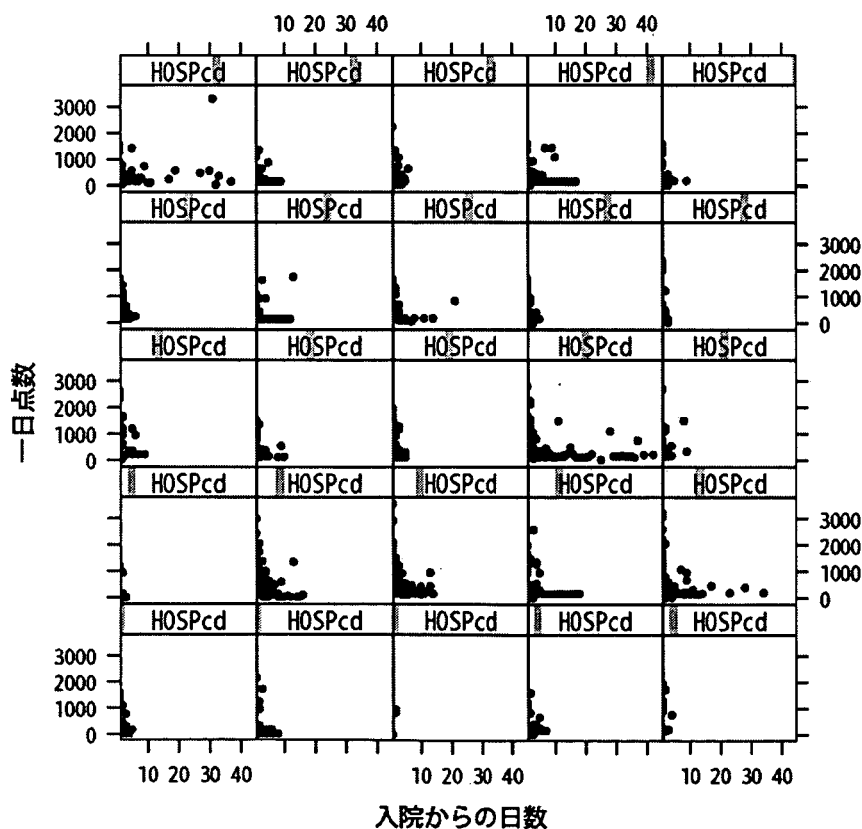


Fig 3

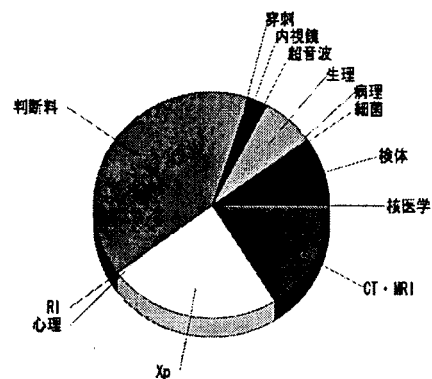


Table 2

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
Xp	552	367			58	55	914
CT・MRI	1,470	785	84	328	98	119	238
核医学	375						1
造影							

【再掲】

	行為		薬剤		材料		n
	mean	SD	mean	SD	mean	SD	
CT	871	494	69	304	79	90	239
MRI	1,166	434	61	266	168	81	19
二回目～	981	732	177	438	78	49	13

病院

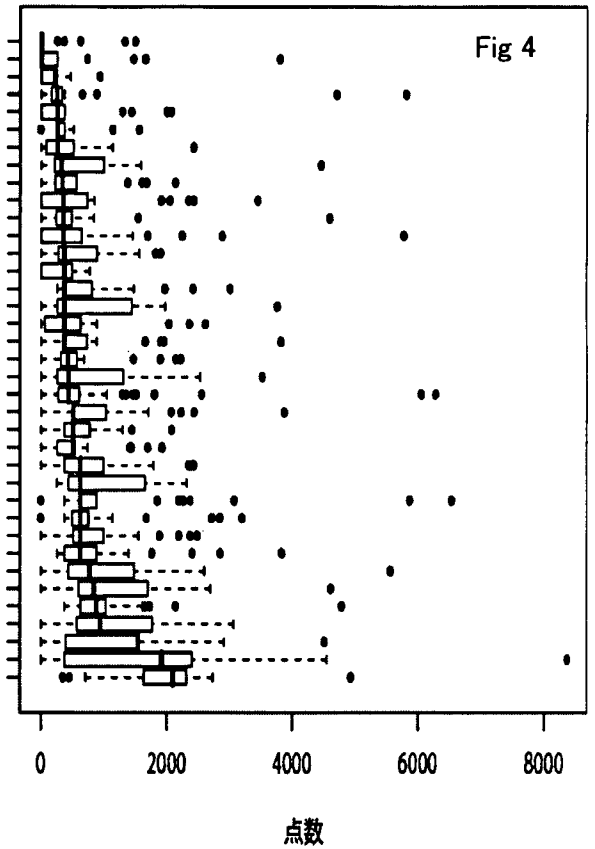


Fig 5

1610703x99x00x

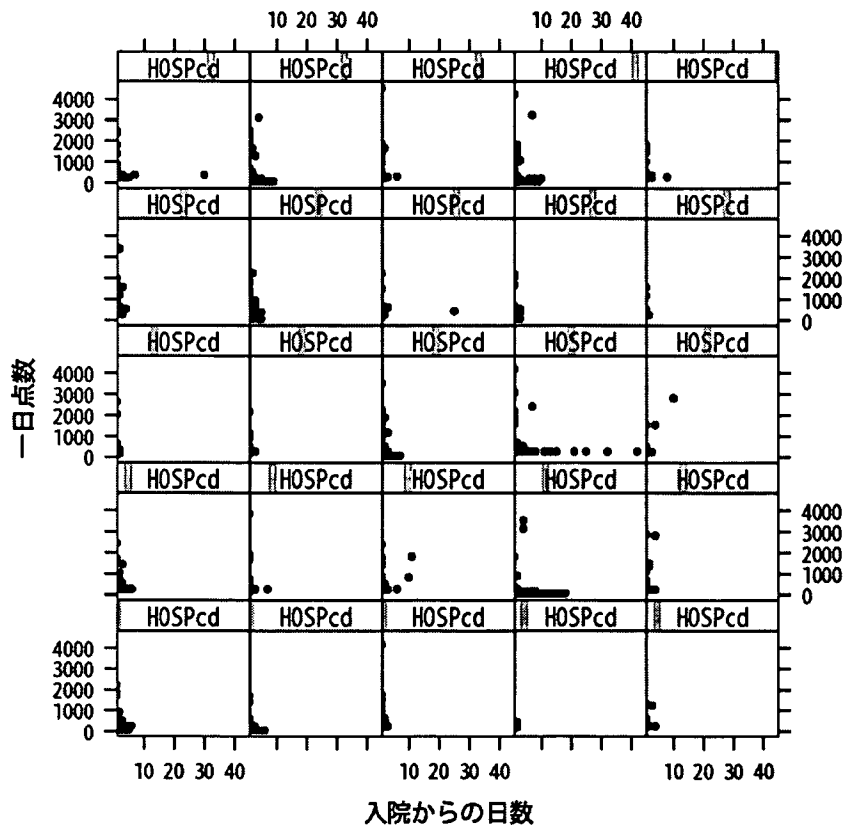
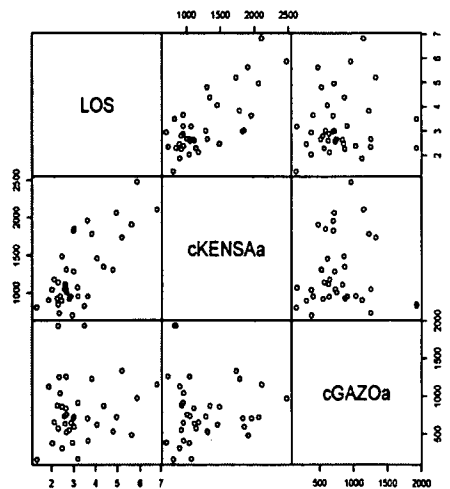


Fig 6



1610703x99x00x

薬物中毒(その他の中毒)、手術なし、手術・処置2なし、副傷病なし

医療機関におけるDPCデータ分析の実際(上)

— MS-Excel、MS-Access を使用して —

札幌医科大学 放射線医学講座・医療情報企画室 藤森 研司

はじめに

平成18年に入り、DPC実施病院は360、DPC準備病院は371余となり、一定数の急性期病院が様式1、E、F-file等のDPCデータを厚生労働省の本体調査に提出するようになってきた。DPCデータは自院の診療情報の宝庫であり、自ら分析する力を身につけることがこれからの病院マネジメントには欠かせない。

そこで本稿では、①平成18年8月21日～23日に産業医科大学で行われたDPC研究班主催の「DPC夏季セミナー」での筆者の実習コースと、②北海道内のDPC実施病院・準備病院を対象に10月14日に当院で行った第1回「DPCセミナーin札幌医大」での実習をたたき台とし、自院におけるDPCデータの分析手法について2回にわけて解説する。

前編は基礎編として、MS-Excel、MS-Accessによるデータの取り扱いについて、実際の手順をstep by stepで解説し、後編は応用編として、OLAPツールと呼ばれるデータマイニングソフトによる課題解決型のデータ分析の手法について述べる。

DPCの導入により、自院の病歴情報、診療内容が、医療従事者・病院事務職員が扱える形で電子化された。しかもその形式は全国共通であり、分析のノウハウを共有することができる。DPCに係わるすべての病院が強力なマネジメントツールを手中にしているが、それを生かすかどうか問われているのである。是非ともパソコンの前に座り、本稿を読みながら実際に手を動かしてみたい。どの病院でもかならずこのレベルまでの分析はできる。なぜなら、持っているデータ構造はみな同じなのだから。

本稿は医療機関における自院のDPCデータの分析法を示したものであるが、保険者でも同様な分析がより広範囲に可能である。保険者における応用を後編に示したのであわせて参照して欲しい。

I データの種類と概要

本年度に各病院から厚生労働省のDPC本体調査に提出されているデータの概要を表1に示す。

様式1は患者ID、入退院日、主傷病名、術式、化学療法の有無などの患者の臨床像を表すファイルである。このファイルだけで、どのような患者がどのような診療を受け、どのように退院したかが分かる。診療情報管理の基本データである。各種の分類情報もあり、癌のTNM分類やステージなども分かる(但し、今年度は必須ではないものも多い)。

E-fileはオーダー単位のレコードで、医療行為ごとに診療区分、代表名称、点数、実施回数、実施日などが記載される。このレコードの入院ごとの点数合計は、出来高レセプトの合計と一致する。それぞれのレコードは各行為(オーダー)の小計であり、たとえば処方では個々の薬剤名ではなく、代表薬剤として一つが表示されるのみである。注射、処置、手術、検査、画像診断なども同様で、詳細についてはF-fileを参照することになる。

F-fileはE-fileの明細であり、詳細レベルの行為、薬剤名称や材料が厚労省マスターコード、点数、使用量などとともに記載されており、医療内容を評価するには欠かせないファイルである。検査ではすべての検査項目が記載され、その点数合計は丸めがあるために実際の請求額とは異なる。F-fileは実施日や実施回数を持たないので、必ずE-fileとあわせて使うことが必要である。E、F-fileは外来診療も扱えるデータ形式であり、今年度は試行として外来E、F-fileの提出も始まっている。

D-fileはE-fileに近い構造を持ち、実際のDPCレセプトの内容をあらわすデータである。包括対象となる処方、注射、処置、検査、画像診断、入院料などの情報はなく、DPC1日点数や医療機関別係数、算定開始日・終了日などの支払いに必要な情報

を持つ。DPCで支払いをしていないDPC準備病院では、提出は義務化されていない。

これらのDPCデータに、自院の医療情報システムのマスターや各種の厚労省マスター、診断群分類点数表などをあわせて使用することで、自院のDPCごとの平均在院日数と診断群分類点数表上の入院期間Ⅱとの比較、DPCごとの診療区分別点数の集計、あるいはDPCごとの使用薬剤や材料の集計など、DPCにかかわる各種の集計・分析を行うことができる。

この段階までのデータの取り扱いにあたっては特別な環境は不要であり、通常のPCとMS-Excel、MS-Accessで十分である。高機能なテキストエディタもあることが望ましいが、フリーウェアやシェアウェアでも十分な機能がある。さらに課題解決型のデータ分析を行う場合は、データマイニング用のソフトウェアが有用である。その詳細は後編で詳しく述べるが、本稿ではデータマイニングに持ってゆくまでのデータの基本的加工のノウハウについて解説する。

Ⅱ データの用意

まず、自院の様式1、様式4、E-file、F-fileを用意する。DPC実施病院ではD-fileも用意すると良いだろう。これらのデータは本体調査のデータ提出先である厚生労働省保険局医療課からエラーがないことが確認されたデータを用いよう。良い分析には、正しいデータが必要である。

様式1にはDPCコードがないので、自院のDPCコード付きの退院患者リストを別に用意する必要がある。「DPCコード付き退院患者リスト」では患者ID、入院日、DPCコードが必須である。DPC準備病院ではデータを提出するだけで実際のコーディングまでに至らない病院もあるようだが、自院にとっての主要領域だけでもDPCコーディングを行うことを強くお勧めする。自院で分析力がついてはじめてデータ提出の労力が報われるだろう。「ふくろう君」という優れたコーディング用ソフトウェアがあるのでそれを利用して欲しい¹⁾。

さらに、自院の診療科コードもあると診療科別の分析もできる(様式1の診療科コードは調査用に割り振られた共通コードであるため、複数の内科や外科がある場合は様式1では識別できない可能性がある)。その他にも、分析目的に合わせて必要なデータを用意すると良いだろう。

表1 本体調査に提出しているデータ

ファイル名称	概要
様式1	退院患者の基本情報。患者ID、入院日、傷病名、手術名・手術日、各主要治療法の有無、癌のステージ、NYHA分類等。
様式3	施設調査票で、各病床区分の平均在院日数、稼働率等。
様式4	患者ごとの医科診療保険以外の情報(自費、労災等)。
E-file	患者ごと、実施日ごと、診療行為ごとのレコード。オーダごとの回数と点数を持つ。合計は出来高点数に一致する。
F-file	E-fileの診療行為の詳細。個々の薬剤や材料も記載される。“丸め”は考慮しない。
D-file	DPC実施病院のみが作成。E-fileに似るが実際の請求内容である。算定開始・終了日、医療機関別係数もある。

● 診断群分類点数表も

次に、全国との在院日数比較のために診断群分類点数表を用意しよう²⁾。加えて、最新の厚労省の診療マスター、医薬品マスター、材料マスターなども用意しておくといいたい³⁾。自院の医事マスター等も手元においておくとよい。

E、F-fileは、診療行為や点数分析の基本であるが、退院数の多い病院では1年間分のすべてのデータをMS-Accessに読み込ませることは難しい。MS-Access 2003では、一つのファイルサイズの上限が2GBのためである。退院数の多い病院で1年間分を一つのファイルで扱う場合は、行為名称の空白を削除する、必要な項目のみを読み込む、行為名称は読み込まずコードだけにするなど、ファイルサイズ縮小化のための工夫が必要である。あるいはOracle、SQLserver、DB2などの大規模データベースを使用し、必要な部分だけをMS-Accessで切り取ってくるのも良い手法である。データの蓄積は規模に合わせて取り組んで欲しい。

診断群分類点数表はMS-Excel形式で提供されているので、必要な部分を切り出してMS-Accessにも読み込ませるとよい。厚労省の診療マスター、医薬品マスター、材料マスターを使うのは特殊な場合であるが、診療行為の区分記号(手術はKコードなど)や薬効分類が必要な場合は、これらのファイルを参照することになる。ある程度分析スキルが上がってくれば、是非とも手元においておきたいファイルである。

データ分析の前に、それぞれのファイルの構造と各項目の意味を十分に知っておくことが重要である。厚生労働省の説明会資料(DPC導入の影響評価に係る調査に関する説明会資料6/13版)も参照して欲

しい4)。

Ⅲ MS-Excelで様式1を分析する

はじめに入門編としてMS-Excelを使用して、DPCごとの自院の平均在院日数と全国のそれを比較してみよう。診断群分類別点数表の入院期間Ⅱは50%タイル値であり、点数表の算出に使われた病院群の全国平均にほぼ一致する。

様式1の各患者レコードに相当するDPCの入院期間Ⅱを割り当て、DPCごとの平均値を集計することが、この章の目標である。

ステップとして、

- ① 各レコードに「DPCコード付き退院患者リスト」からDPCコードを引き当てる
 - ② DPCコードに診断群分類別点数表から入院期間Ⅱを引き当てる
 - ③ 各レコードの在院日数を計算する
 - ④ 粒度調整を行う
 - ⑤ 「ピボットテーブル」機能を使用して、平均値を得る
- を順次行ってゆく。

① DPCコードを引き当てる

まず、様式1をMS-Excelに読み込ませ、自院で作成した「DPCコード付き退院患者リスト」も同じファイルの次のシートに読み込ませる。

次に、様式1の各レコードに「DPCコード付き退院患者リスト」のDPCコードを割り当てる。ただし、同じ患者が何度も入退院するため、複数回入院のケースではIDだけでは一意には関連付けができない。データを入院ごとに識別するためには、常に患者IDと入院日のペアで取り扱うことが必要である。

MS-Excelはlookup関数と呼ばれる一連の参照関数を持っており、あるデータをキーとして、他のデータを引用することができる。ただし、参照できるキーは1カラムのみであり、患者IDと入院

日を同時に参照することはできない。

従って、事前に患者IDと入院日をあわせた項目を追加しておく。患者IDも入院日も整数なので、"+"記号で結んでしまうと単純な足し算となってしまうため、"&"記号で結び文字列とする。

まず、=患者ID&入院日として、それぞれのシートに項目を追加する作業を行う。vlookup関数で参照されるデータ群は、キーとなる列が参照範囲の一番左でかつ昇順にソートされていることが必要なので、シート全体を「患者ID&入院日」列で昇順にソートしておく。上記の操作が終了後、様式1の各レコードにvlookup関数を使用してDPCコードを引き当てる(図1)。

これで①のDPCコードを引き当てる作業が終わるが、vlookupの式を下方にコピーする際に、lookup対象のセルの範囲指定で、行数字の前に"\$"をつけるのを忘れないように。

② 入院期間Ⅱを引き当てる

次に、②の入院期間Ⅱを引き当てるために、診断群分類別点数表を同じファイルの別なシートに読み込ませ、全体をDPCコードで昇順にソートする。①で得られたDPCコードから診断群分類別点数表のDPCコードを同様にvlookup関数で参照し、入院期間Ⅱを様式1の各レコードに引き当てる。検索の型は"0"としておく。不一致なものは"#N/A"が返されるが、これは診断群分類別点数表に該当するものがないという場合で、出来高支払いの群である。

以上、①と②の操作で様式1の各レコードにDPCコードと平均在院日数Ⅱが割り当てられた。

③ 在院日数の計算

次は③の在院日数の計算であるが、これは退院

図1 vlookup関数によるDPCコードの引き当て

シート「様式1」		シート「DPCコード付き退院患者リスト」		
患者ID&入院日①	参照DPCコード	患者ID & 入院日②	DPCコード	診療科コード
001010101020060401	=vlookup()	001010101020060401	010010x01xx0x	4
001010301020060412	"	001010103020060418	010030x97xx0x	2
以下続く	"	以下続く	以下続く	7

vlookup関数で、退院患者リスト全体を参照し、キーが一致したものを持ってくる

シート「様式1」でDPCコードを参照するカラムには、
=vlookup(患者ID&入院日①、シート「退院患者リスト」の範囲、DPCコード列番号、0)
と入力する。患者ID & 入院日①の項目には、本例では一つ左のカラムを指定することになる。DPCコード列番号は本例では2(左から二番目の意味)。関数最後の項目は検索の型を指定するもので、“0”は不一致レコードを“#N/A”として返す。

日-入院日+1である。

ただし、入院日も退院日も8桁のyyyymmdd形式の整数であるため、単純に引き算をしても実際の日数にはならない。MS-ExcelではLEFTやMIDなどの文字列関数を使用して日付型に変換を行ったうえで引き算をすることもできるが、ここではTEXT関数で一気に型変換を行いDATEDIF関数で日数を求める方法(図2)を紹介しておく。

④ 粒度調整

以上でDPCごとの在院日数の比較ができるが、この作業は実際には次章で解説するMS-Accessを使用したほうがスマートで手早い。しかしながらそれぞれのデータの所在を意識していただき、なにかと便利なlookup関数を紹介するために、あえてMS-Excelで行う方法を述べた。

MS-Excelで行うべきは、むしろこれから示す粒度の調整と集計である。

各病院単位ではDPC14桁コードそのままでは症例数が少ない群が多く、統計値としてはもの言にくいケースがあるだろう。自院での分析では、DPC上6桁に手術の有無、化学療法の有無等で、大枠の分類を用いるほうが有用性が高いと思われる。この作業を粒度調整と呼ぶ。

ここでは手術を有り・無しに二分して見よう。様式1に手術フラグの列を追加し、フィルター機能を使用して手術1~5に一つもないものを“手術なし”とする。手術1が最も主要な手術であると担保されていれば、手術1を見るだけでよいだろう。

手術の項にはK920*として輸血も入っているが、これは手術とは見なさない方がよいかもしれない。フィルター機能を使用してK920*しかないレコードも、手術フラグを“手術なし”とする。残りを“手術あり”とするが、さらには、胃瘻や気切などの処置的な手術(DPCの手術区分が"97")を、主たる手術とは別に扱った方がよいと考える場合は、フラグ区分を増やすとよいだろう。

図2 8桁整数を日付に変換する数式

①MS-Excelでは：

=DATEDIF(TEXT(入院年月日,"####-##-##"),TEXT(退院年月日,"####-##-##"),"D")+1

8桁の整数型の年月日データを、TEXT関数で日付型に変換し、さらにDATEDIF関数を使用して入院日と退院日の間の日数を求める。DATEDIF関数は、隠し機能のようであるので、問題があれば引き算で求める。

②MS-Accessでは：

DateDiff("d",CDate(Format([入院年月日],"@@@-@@-@@")),CDate(Format([退院年月日],"@@@-@@-@@")))+1

8桁の整数型の年月日データを、Format関数とCDate関数により日付型に変換し、さらにDateDiff関数で日数を求める。上記と似ているが関数名に“f”が一つ多い。

図3 MS-Excelピボットテーブルで集計する

DPC	平均 / 在院日数	平均 / 入院期間	患者数
010000	27.0	50.2	39.8
020000	10.7	42.7	29.3
030000	3.0	11.0	20.0
040000	1.9	7.8	28.5
050000	4.0	18.5	8.5
060000	7.0	4.0	11.0
070000	5.0	8.5	6.0
080000	9.0	11.0	10.3
090000	1.0	2.0	3.0
100000	3.6	14.4	11.1
110000	14.1	21.7	17.8
120000	18.0	7.0	23.0
130000	10.1	29.0	15.3
140000	12.0	32.0	17.0
150000	3.0	7.0	4.0
160000	18.0	18.0	18.0
170000	18.0	18.0	18.0
180000	1.0	1.0	1.0

化学療法は様式1にフラグ化されているので、そのまま用いることもできる。その他にも自院の分析に必要な項目を選び適宜粒度調整を行い、使用しない項目は削除してコンパクトなファイルにしておく。

⑤ 集計

いよいよ⑤の集計であるが、MS-Excelにはピボットテーブルという優れたクロス集計機能がある。DPC上6桁の手術の有無ごとに、自院の平均在院日数、全国の平均在院日数、症例数などが簡単に集計できる。図3はDPC6桁に手術の有無で自院の平均在院日数、全国の入院期間Ⅱの平均と患者数を集計したものである。さらに入院から手術までの日数や手術から退院までの日数の集計、診療科別などの

分類項目の追加も容易である。

IV MS-Accessで様式1を分析する

MS-Excelの限界は、1シートで扱えるレコード数が65,536と限られていることであり、様式1の分析には向くが、E,F-fileの分析には向かない。退院数1,200名/月程度の当院では、1ヶ月のE-fileは22万レコード、F-fileは50万レコード程度になる。従ってE,F-fileの分析にはMS-Access等のデータベースが必須となる。

様式1、E,F-file等をMS-Accessに読み込ませる際には、いくつかの注意点がある。

まず、データチェックプログラムを介して保存されたデータは、1行目がヘッダーで項目名称になっているが、先頭のエラーコード欄のみ項目名称に日時がついており、ファイルごとに名称が異なる。MS-Accessは読み込むファイルのヘッダー名称が異なるとエラーとなるので、この部分をテキストエディタで修正しておく必要がある。

また、複数のテーブルにリレーションをはる場合、項目のデータ型が一致している必要がある。特に、患者ID、入院日はデータに関連付けるための基本であるので、すべてのテーブルでデータ型をあわせておくことが必要である。インポート機能を使用した場合、自動的に型が振られるため、意図しない型になってしまう場合がある（一般には長整数型かテキスト型）。データ型は読み込んだあとでも修正可能なので、必ずデータ型を確認しておこう。

F-fileは実施日情報を持っていないため、複数の月を同じテーブルに読み込むと、どの月のものか区別がつかなくなってしまう。それを回避するために、E-file、F-fileのテーブルを拡張して診療月の項目を用意し、既定値を読み込ませる月にあわせて変えておくが良い。

なお、前記の在院日数比較の作業をMS-Accessで行うと、前章①～③の作業はきわめて容易にできる。これはDBにはリレーションという機能があるため、複数のテーブルを複数のキーで結び、あたかも一つのテーブルであるがごとく扱うことができる。一方、④、⑤の作業はMS-Excelの方が簡便に行えるので、MS-AccessとMS-Excelのそれぞれ得意なところを使い、両者を巧みに行き来することで作業が効率化できる。

では、MS-Accessを使用して、DPCごとの平均在院日数を求めてみよう。

● 様式1をインポート

まず、csv形式の様式1をインポートする。複数の月のデータがあればどんどん追加していこう。ただし、エラーコードの項目名称を編集することを忘れずに。

次に、自院の「DPCコード付き退院患者リスト」も読み込む。MS-Excel形式であっても読み込みは可能である。往々にして患者IDと入院日のデータ型が様式1とは異なってしまうので、デザインビューでデータ型をあわせておく。診断群分類点数表も読み込むが、これは少し凝ったデザインのMS-Excelファイルなので、一度適宜整形してcsv形式に保存しインポートすると良い。

● クエリで三つのテーブルを結合

以上のインポートが終了したらクエリを使用して三つのテーブルをつなげる（図4）。

様式1と「DPCコード付き退院患者リスト」は患者IDと入院日でリレーションをはり、「DPCコード付き退院患者リスト」と診断群分類点数表は14桁のDPCコードでリレーションをはる。この作業だけで、①、②は終了である。

さらに在院日数を求めるが、MS-Accessでは8桁のyyyymmdd形式の整数データを文字列および日付に変換する型変換関数が用意されているのでこれを使う（図2、下段）。この数式をクエリに追加すると③が終了である。実際には背後でSQL言語の式が作成されているわけだが、クエリというビジュアルなインターフェイスでユーザにSQLを意識させない敷居の低さがMS-Accessの魅力である。

④、⑤の作業はMS-Excelの方が効率的なので、このクエリをMS-Excelにエクスポートして、後半の作業を進めるのも一法である。

なお、MS-Accessでもif文を使って粒度調整は可能なので、こちらも試して欲しい。MS-Accessにもピボットテーブルビューというクロス集計機能があるので、④で粒度調整を行ったものを再びMS-Accessにインポートしクロス集計を試みよう。MS-Excelと似たようなことができる。

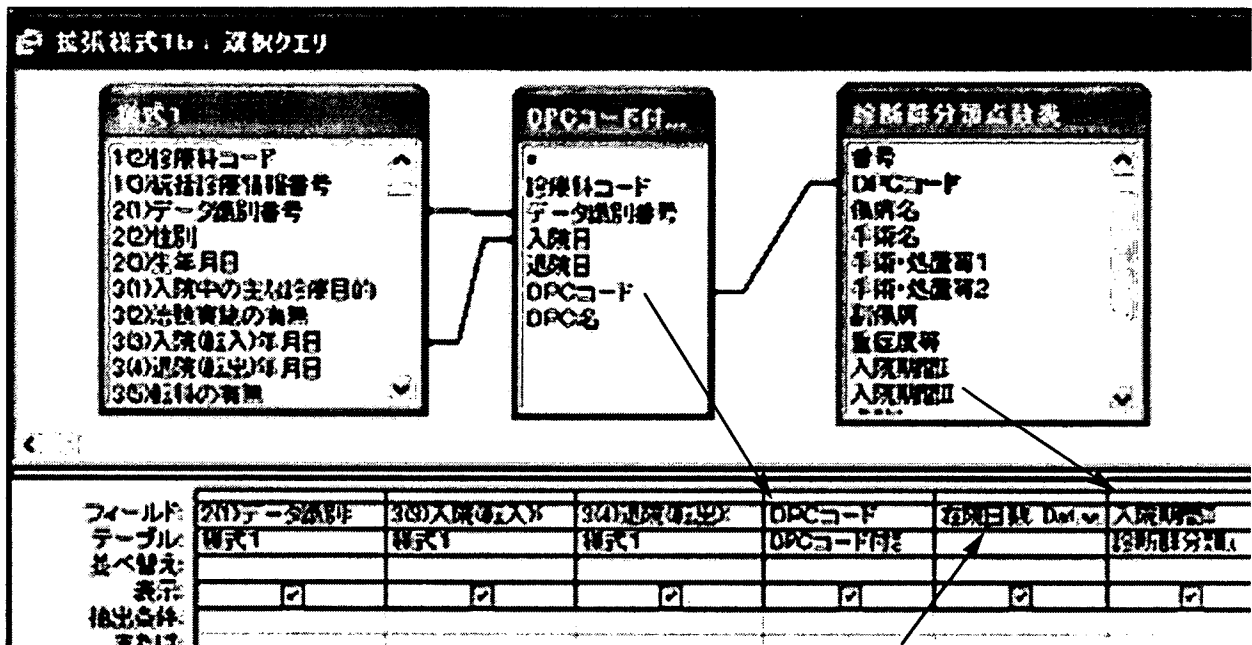
V MS-AccessでE、F-fileを分析する

さて、MS-Accessの本領はレコード数の多いE,F-file等の取り扱いにおいてである。

① E-fileを使って診療区分別点数集計

E-fileは診療区分、実施日、点数、回数を持つ

図4 MS-Accessのクエリで平均在院日数を求める



在院日数は、図2の下段の式によって求める。

で、患者別、DPC別の診療区分別点数集計も行える。入院から退院までの間に投入された医療内容を時系列で保険点数として捉えることで、一入院の診療密度の評価にも利用できる。

本稿では、入院ごと、保険診療区分ごとの出来高点数の集計方法を例示する。時系列の情報は使用していないが、実施日を追加するだけで時系列情報も扱うことはできる。適宜応用してみたい。

診療区分別点数集計のステップは、

- ① クエリのデザインビューにて前章で作成した拡張様式1、様式4とE-fileを患者IDと入院日でのリレーションをはる。クエリは他のクエリを参照することができる。様式4で保険種別を1（医科レセプトのみ）に限定し、入院日はデータのある月以降とする
- ② 診療区分をやや大きな分類に取りまとめるテーブルを作り、①に追加する
- ③ 患者、入院日、診療大区分でグループ化し、点数x回数を合計する
を行う。

①で様式4を追加しているのは、E-fileは保険診療のデータのみで自費、治験、労災や自賠責の部分は含まない。これらの場合でもDPCコードは付いているので、これらの患者を除外し保険診療のみの患者に絞り込むためである。また、この集計は一入院ごとの集計を目的としているので、入院日をE-fileの存在する最も古い月のもの以降とする。

②の診療区分の粒度調整は、E-fileの診療区分は21、22などと細かいので、20の処方、30の注射などと大枠で集計するための作業である。テーブルを追加して細かい診療区分と見たい大区分の対応表を作っておく。

③では集計機能を使用し、患者ID、入院日、診療区分等でグループ化する。実際の点数は点数x回数なので、これを計算式として記述し、かつ合計としておく。この集計のためのクエリを図5（次頁）に示す。

これらの作業で、患者ごと・入院ごと・診療区分ごとの点数集計ができる。ピボットテーブルビューで表示すると、患者ごと診療区分ごとの点数集計が見やすい形で得られる。DPCごとなどの集計もできるので、適宜工夫してみたい。

② F-fileとD-fileを活用

さらにF-fileを加えると詳細な診療内容を分析することができる。F-fileはE-fileを親とするもので、患者ID、入院日、行為区分、順序番号でリレーションをはる。加えて複数の月を扱う場合は、独自に導入した診療月にもリレーションをはることがポイントである（図6）。

薬剤や材料について使用量を得たい場合は、F-fileの使用量にE-fileの行為回数をかける。薬剤料も同様である。DPCごとの使用薬剤の分析は、このレベルで可能となる。その詳細は後編で述べる。

またMS-Accessでは、患者ごとの簡易な日計表も作成できるので、是非試してみたい(図7)。時系列で医療の投入内容の詳細が分かり、クリニカルパスの微調整には有用な情報である。

D-fileがある病院では、患者単位で出来高点数と包括点数の比較が可能となるが、これは前編の卒業試験としてDPC実施病院の皆様にはトライしてもらうことにしよう。

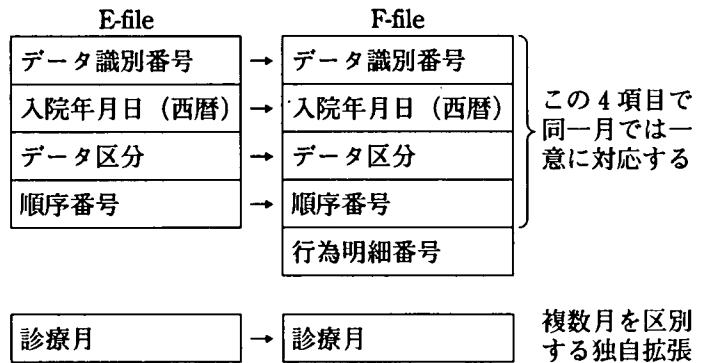
ここまでデータ加工ができるようになると、上記のような定型帳票的な作業だけではなく、課題解決型の分析が可能となる。そのためには、データマイニング用のアプリケーションを利用することになるが、詳細に関しては後編で述べる。

おわりに

以上、様式1、E、F-fileをMS-Excel、MS-Accessで取り扱う方法を概説した。様式1だけでも多くの有用な情報があり、たとえばDPC別の平均在院日数と患者数で、自院に最適な疾患別の必要病床数が分かる⁵⁾。あるいは二次医療圏ごとの患者数と比較して、自院のシェアを把握することもできる⁶⁾。ファイルの構造と各項目の意味を理解していただき、工夫を重ねて欲しい。

本稿で示したものはほんの一例であり、いろいろ

図6 E-fileとF-fileの対応



な解法と応用があるだろう。

当院では「DPCセミナーin札幌医大」として、道内のDPC実施病院・準備病院の実務担当者を対象に、DPCデータの分析手法についての実習型のセミナーをシリーズで行っている。皆様の地域でも横のつながりによってノウハウを交換し、スキルの向上を目指して欲しい。

自院で分析を行わず、コンサルタント会社に一任している病院もあるだろう。その費用対効果を必ずしも否定するものではないが、まず自分たちで取り組んで分析の仕組みを理解し、その上でコンサルタント会社が提供する結果を吟味するスキルの獲得が重要である。ブラックボックスのものをそのまま信用してはいけない。すべての病院が同じ構造のデータを持っているのだから、他院とのベンチマーク以

図5 E-fileから診療区分別集計をする

ファイル名	データ識別番号	入院日	退院日	DPCコード	データ区分	行為明細番号	医療保険外との組み合わせ
テーブル	DPCコード別集計	DPCコード別	DPCコード別	DPCコード別	診療区分別		医療保険外との組み合わせ
集計	グループ化	グループ化	グループ化	グループ化	グループ化	合計	Where 条件
並べ替え							
表示	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
抽出条件	>=20060401						

診療区分の集計は、行為点数*行為回数を合計する。「医療保険外との組み合わせ」を“1”とし、医科レセプトに限定する。入院日はデータの存在する範囲と限定する。

外はどの病院でもコンサルタント会社と同じ分析は可能である。是非とも自院で取り組んでいただき、DPCデータを病院マネジメントに役立てて欲しい。

後編ではOLAPツールを用いた、課題解決型のデータ分析の手法について解説する。

【参考資料】

- 1) <http://www.dis.h.u-tokyo.ac.jp/byomei/>
- 2) <http://www.prrism.com/dpc/dpc.html>

3) <http://www.iryohoken.ne.jp/>

4) <http://www.prrism.com/dpc/setsumeikai>
20060613.pdf


5) 藤森研司、新たな病院マネジメントの確立を目指して「DPCと病院マネジメント」(松田晋哉編)じほう、2005,188-198.

6) 伏見清秀「DPCデータ活用ブック」じほう、2006

図7 F-fileから簡易日計表を作る

日計表の作成 双数クリ

実施年月日		20060724	20060725	20060726	20060727	20060728
中	診療科目名	行高	診療科目名	行高	診療科目名	行高
10	悪性腫瘍科手術	¥400				
20			フランドンテープ ¥290 麻酔科(入院) ¥70			フランドンテープ ¥696 麻酔科(入院) ¥70
30	フルカニク2号 ¥2844 ファーストシメチ ¥2232 プロスタルモンF ¥2106 ウーゾ注 0.2 ¥1926 中心静脈注射 ¥1400 ウーゾ注 50 ¥591 ヘリクナトリウム ¥366 生理食塩液 20 ¥264 ヘリクナトリウム ¥366 生理食塩液 10 ¥194 ヒューマリン剤 ¥31	¥4464 ¥2353 ¥2106 ¥1400 ¥591 ¥366 ¥330 ¥194 ¥77	ファーストシメチ ¥4464 フルカニク2号 ¥2353 プロスタルモンF ¥2106 中心静脈注射 ¥1400 ウーゾ注 50 ¥591 ヘリクナトリウム ¥366 ザンタック注射 ¥330 ヒューマリン剤 ¥77	ファーストシメチ ¥4464 フルカニク2号 ¥2353 プロスタルモンF ¥2106 中心静脈注射 ¥1400 ウーゾ注 50 ¥591 シペノール注射 ¥1037 ウーゾ注 0.2 ¥963 精密内視鏡透視 ¥800 ウーゾ注 50 ¥708 ウーゾ注 0.2 ¥642 ヘリクナトリウム ¥366 ザンタック注射 ¥330 生理食塩液 10 ¥194 ヒューマリン剤 ¥124 生理食塩液 50 ¥97	ファーストシメチ ¥4464 フルカニク2号 ¥2353 プロスタルモンF ¥2106 中心静脈注射 ¥1400 ウーゾ注 50 ¥708 精密内視鏡透視 ¥800 ウーゾ注 0.2 ¥642 ヘリクナトリウム ¥366 ザンタック注射 ¥330 生理食塩液 10 ¥194 ヒューマリン剤 ¥124 生理食塩液 50 ¥97	フルカニク2号 ¥2353 プロスタルモンF ¥2106 中心静脈注射 ¥1400 ウーゾ注 50 ¥965 精密内視鏡透視 ¥800 ウーゾ注 0.2 ¥642 ヘリクナトリウム ¥266 ザンタック注射 ¥330 生理食塩液 10 ¥194 ヒューマリン剤 ¥163 ウーゾ注 0.2 ¥97 生理食塩液 50 ¥97 生理食塩液 20 ¥64
40			レクス麻酔 ¥41100 ミラーアセット ¥1500		ドレーン注(フレ) ¥210	
50						
60	超音波(UCG) ¥7800 ECG12 ¥3000 パルスドプラズマ ¥2000 基準的胸部検査 ¥1400 末梢血液一般 ¥230 糖 ¥190 Tcho ¥170 P&U-PO4 ¥170 CRP(定量) ¥170 GOT ¥170 GPT ¥170	¥7800 ¥3000 ¥2000 ¥1400 ¥230 ¥190 ¥170 ¥170 ¥170 ¥170	呼吸器拍動機 Z ¥1500 基準的胸部検査 ¥1400 末梢血液(血) ¥110	ECG12 ¥1500 基準的胸部検査 ¥1400 末梢血液一般 ¥230 糖 ¥190 GPT ¥170 CRP(定量) ¥170 GOT ¥170 Tcho ¥170 P&U-PO4 ¥170 TG ¥110 UA ¥110	基準的胸部検査 ¥1400	基準的胸部検査 ¥1400 末梢血液一般 ¥230 糖 ¥190 CRP(定量) ¥170 GPT ¥170 GOT ¥170 Tcho ¥170 P&U-PO4 ¥170 LOH ¥110 CRP ¥110 TTT ¥110



レセプト請求用傷病名による 特定疾患早見表 (18年4月版)

B5判 400頁/定価**4,200**円(本体4,000円+税)

発売中

◆「保険請求時に原則使用すべき傷病名」のうち、特定疾患療養指導料やその他難病に関する指導料・診療料等の算定対象疾患(特定疾患)を50音順に配列した実用早見表です。同義語等からも検索が可能です。ICD-10の改正に対応した改訂版です。

株式会社 **社会保険研究所** 〒101-8522 東京都千代田区内神田2-4-6 WTC内神田ビル
電話 03-3252-7901(代) ファックス 03-3252-7977

医療機関におけるDPCデータ分析の実際(下)

— OLAPツールを使用して —

札幌医科大学 放射線医学講座・医療情報企画室 藤森 研司

はじめに

前編(上)ではMS-Excel、MS-Accessを使用して、様式1、E-file、F-file等の基本的なデータ分析手法について解説を行った。後編にあたる本編では、それらの技法によって編集・加工されたDPCデータを利用して、課題解決型の分析手法について解説を行う。

大量のデータから課題解決型の分析を行うにはOLAPツールが有効だが、OLAPツールはデータ加工そのものは不得意なので、MS-Access等を用いて分析に使用する最終的な形にまで加工したものを、OLAPツールで分析をすることになる。従って、OLAPツールを使用する前提として、MS-Access等で必要な加工ができていなければならない。

2006年8月21日～23日に産業医科大学で行われたDPC研究班主催の「DPC夏季セミナー」では、アシスト社の協力を得てOLAPツールの一つであるPowerPlayを、IT教室のPCに参加人数分導入し実習を行った。4回のセッションで200名程度の参加をいただいたが、「OLAPツールの威力はよく分かったが、どうやって様式1、E、F-fileからここまで持ってくるかを知りたい」との声が多く聞かれた。前編はその問に答えるものなので、前編を参考に基本的なデータ加工のスキルを獲得して欲しい。

I OLAPツールとは何か

DPCの導入により自院でも分析できる詳細な診療データを得たわけであるが、大量のデータの中からマネジメント上の課題を発見し、傾向を把握し、解決のための道筋をつける手段を持つことが求められる。MS-Excelは病院でもよく利用する表計算ソフトウェアであると思われるが、一度に扱えるデータ

の量が限られている。MS-Accessではより大量のデータを扱うことができ、集計や定型帳票作成にも優れるが、多角的にデータを眺め、大枠から詳細に・詳細から大枠にと、適宜視点を切り替えて未知の課題や傾向を把握するような業務には向かない。

OLAPとはOnline Analytical Processingの略であり、日々更新されているデータベース(以下、DB)にアクセスして、解析的に集計・分析を行うものである。販売系の業界では売れ筋や地域特性などを早期に見抜くような業務にも使用されている。

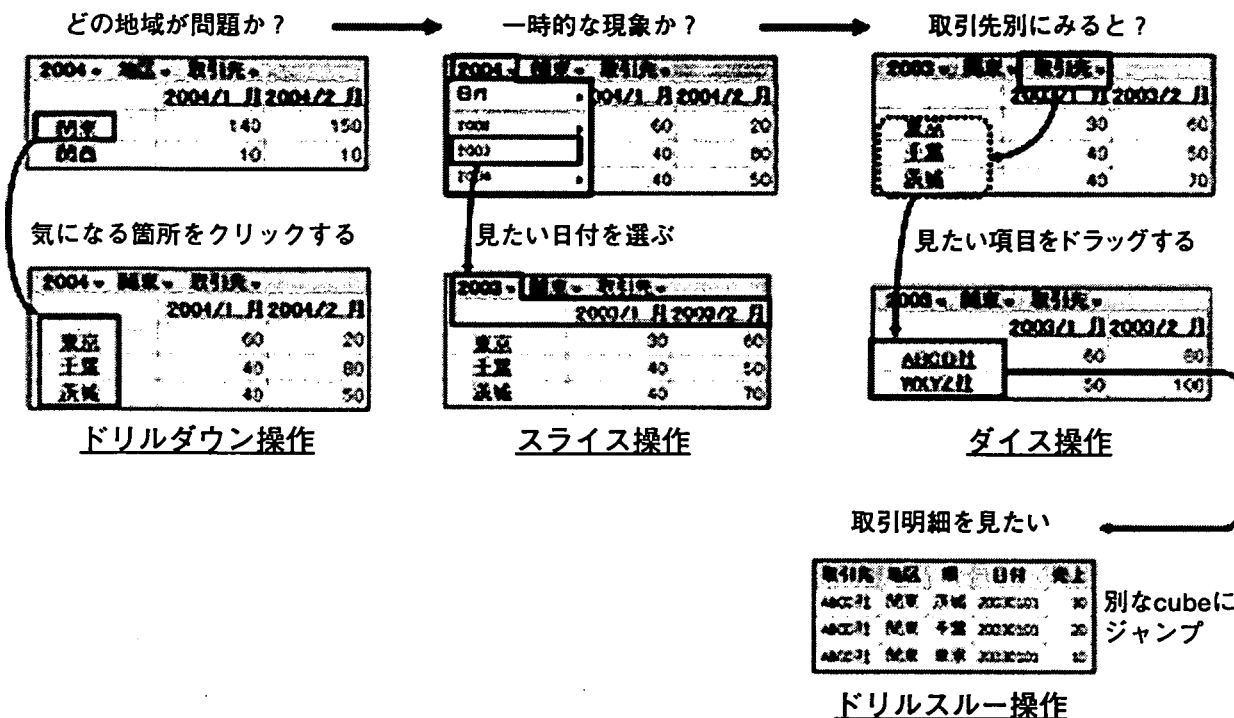
データを多数の次元(見たい視点)に分割し、各次元で大枠から詳細へと数段階のレベルを用意する。次元の入れ替えや組み合わせを行い(ダイス操作)、データを絞り込み(スライス操作)、レベルを自由に行き来し(ドリルダウン操作)、グラフ機能も駆使してデータの特徴を発見してゆく。一つのファイルからさらに別な詳細なファイルにジャンプすることもできる(ドリルスルー操作)。総じてデータマイニングと呼ばれる分析手法である(図1)。

病院も日々データが蓄積されてゆき、その中からマネジメント上の課題を早期に発見し、解決してゆくことが重要である。従来の経営指標と言えば、稼働率や外来数、患者単価などの定型的なデータのみで、OLAP手法までは必要なかった。今日では医療情報システムから大量の診療データを得ることができるようになり、病院でもOLAPツールの普及が進みつつある。E、F-fileは月に1回作成されるファイルなので、必ずしもオンラインである必要はないが、大量の規格化されたデータなのでデータマイニングの手法が生きる。また、大量のデータから定型的切り口の出力をデータ更新の都度に得たい場合も、OLAPツールは強力な手段である。

● MOLAPとROLAP

OLAPは大きくMOLAPとROLAPに大別される。MOLAPはDBのデータをコピーして独自の専用ファ

図1 データマイニングの基本操作



イルを作成しデータマイニングを行う。ROLAPは分析の都度DBにアクセスしてデータマイニングを行う。MOLAPではデータが更新される都度に分析用の専用ファイルも更新する必要があるが、分析に最適化された専用ファイルを用いるので分析速度は速い。ROLAPは日々更新されるDBのデータそのものを使用するので専用ファイル作成の時間が不要であるが、最適化はされないため分析速度は劣る。扱うデータの種類と量と分析内容によって、適宜選択するのが良いだろう。またスタンドアロンのものとネットワーク対応のものがあるので、ユーザーの数に応じて製品を選ぶと良い。

当院ではMOLAPツールの一つであるPowerPlayを使用しているため、本稿ではPowerPlayによる分析について述べるが、他のOLAPツールでも基本は変わらないだろう。PowerPlayはCognos社の製品戦略の変更により、最近ではCognos 8 mobile editionと呼ばれるようになってきたが、本編では従来の呼称「PowerPlay」を用いる¹⁾。中身は同一である。

なお、OpenOLAPというオープンソースによるOLAPツールもあり、すべてオープンソースでデータマイニング環境の構築が可能である(図2)²⁾。開発の難易度は高いが、院内のイントラネット環境で広く使用する場合は、オープンソース環境は検討の価値は高い。一方、担当者数名だけが使用する場合は、スタンドアロン型であるPowerPlayは生産性が

高い。特に試行錯誤が続く初期段階では、PowerPlayのビジュアルなインターフェイスは重宝する。

II Transformerによるcubeの作成

E,F-file等からPowerPlayに持ってくるまでの全体の流れを図3に示す。MS-Excel、MS-Access等で必要な加工を施し、複数のファイルをほぼ最終形のフラットな一つのファイルとする。次いで、cubeの設計を行い、加工済みデータを読み込んで分析用のcubeを作成する。

OLAPでは、分析する次元(ディメンジョン)とレベルの設定を行うことから始める。

PowerPlayではTransformerと呼ばれるアプリケーションで分析モデルを構築し、実際のデータを読み込んで「cube」と呼ばれる専用ファイルを作成する。そのcubeを専用のアプリケーション(PowerPlay本体)でデータマイニングを行う。ROLAPでもモデルを作成することは同様であるが、直接DBをデータマイニングすることになる。

ここではPowerPlayのcube作成アプリケーションであるTransformerを使用して、モデル作成の各ステップを解説する。例として、

- ① DPCごとの平均在院日数の分析
- ② DPCごとの使用薬剤の分析

を行ってみよう。

① DPCごとの平均在院日数の分析

DPCごとの平均在院日数の分析には、前編でMS-Excelで作成した様式1にDPCコード、在院日数、入院期間Ⅱがついたファイル（以下、拡張様式1と呼ぶ）を使用する。これらをcsvファイルに書き出しておくか、MS-Excelのデータベースとしておく。TransformerはMS-Excelのシートそのものは読まないのが難点だろうか。

良いデータマイニングをするためには良いcubeを設計することが重要であり、そのためにはMS-ExcelやMS-Accessで如何に最適な事前加工をしておくかに尽きる。MS-Accessのファイルはテーブル、クエリともに読めるが、Transformerで複雑なクエリを読み込むのは時間がかかるので、csvファイルに書き出しておくのが効率的である。試行錯誤の段階ではクエリを、形が決まってデータ量が増えてきたらcsv

ファイルに書き出したものを読み込むことをお勧めする。

さて、①のDPCごとの平均在院日数の分析では、ディメンジョンとしてDPC、診療科、手術の有無、化学療法の有無などを用意する。各ディメンジョンが表形式での軸に相当する。DPCの次元ではMDCレベル、DPC6桁レベル、DPC14桁レベル、患者レベルなどを階層として用意すると良いだろう。下にいくのがドリルダウンである。一つのレベルについて、七つ程度のカテゴリ数が最適といわれている。なかなかそのようにはならないが、適切なディメンジョンとレベルの設定が、cubeの使いやすさのポイントでもある。

集計する数値は、患者数、在院日数、入院期間Ⅱを用意してみた。

Transformerに拡張様式1を「データソースの挿入」で読み込ませ、Left関数を使ってMDCとDPC6桁を作る。コードだけでは分かりにくいので、コードと日本語名称が対になったテーブルも用意して読み込ませる。Transformerでは一つの項目であれば関係付けることができるが制限事項もあるので、このような作業はMS-ExcelやMS-Access側で行っておいた方が良いかも知れない。

次元やレベルの設定はビジュアルなインターフェイス環境で行うことができ、試行錯誤がしやすい（図4）。集計数値は合計や平均などの内部関数を持っているので、適宜設定する。ここでは、在院日数と入院期間Ⅱは自動集計を“平均”とした。患者数の集計も、カテゴリーカウントという機能を持っているので簡単に求めることができる。入院年月日や退院年月日を使用して月・四半期・年といった集計もきわめて容易であるので、試してみたい。並び順やラベル名称もここで設定する。ダイアグラム機能を用いて、新たなカテゴリの導入も可能である。

すべての設定が終わったら、「PowerCubeの作成」を実行すると専用の分析用ファイルとしてcubeが作成される。即座にPowerPlayを起動してcubeを試し、不十分な部分があればモデルの変更を行い、再度cubeを作成する。この繰り返しの作業がストレスなくできるのがPowerPlayの魅力である。

② DPCごとの使用薬剤の分析

次に、②のDPCごとの使用薬剤の分析では、前編の「V MS-Accessで様式1を分析

図2 OpenOLAPによるデータマイニング

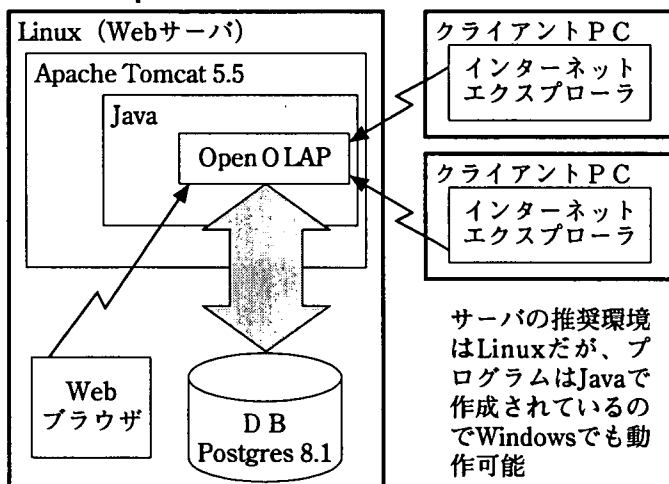


図3 データ加工の流れ

