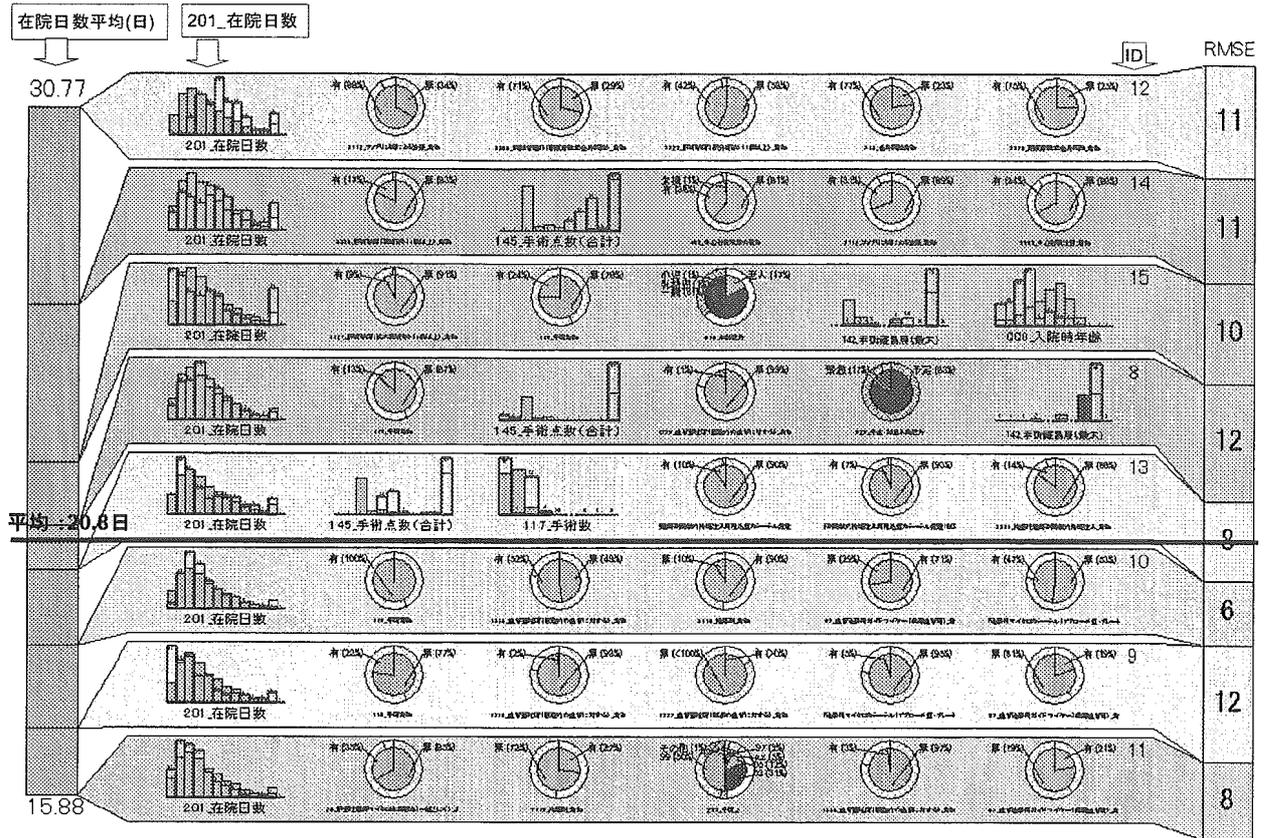


MDC6=060050 肝・肝内胆管の悪性腫瘍(続発性を含む。) 在院日数 RBF

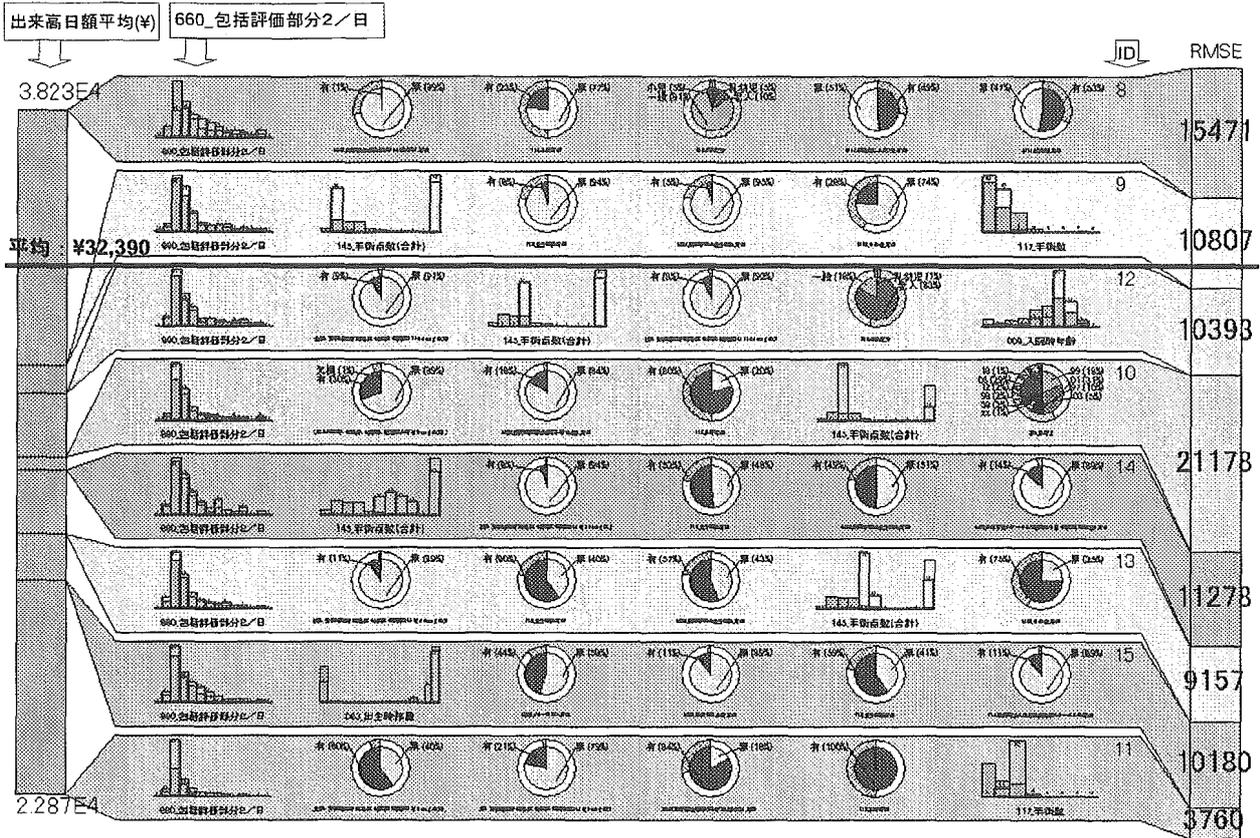


MDC6=060050 肝・肝内胆管の悪性腫瘍(続発性を含む。) 在院日数 RBF

各クラスターごとの在院日数平均およびサイズ

ID	サイズ(n)	在院日数平均(日)
12	6.6	30.8
14	7.2	26.5
15	9.7	23.1
8	13.3	21.7
13	11.3	20.8
10	31.7	19.2
9	9.8	17.4
11	10.4	15.9

MDC6=070040 骨軟部の悪性腫瘍(脊椎脊髄を除く。) 出来高換算額/日 RBF

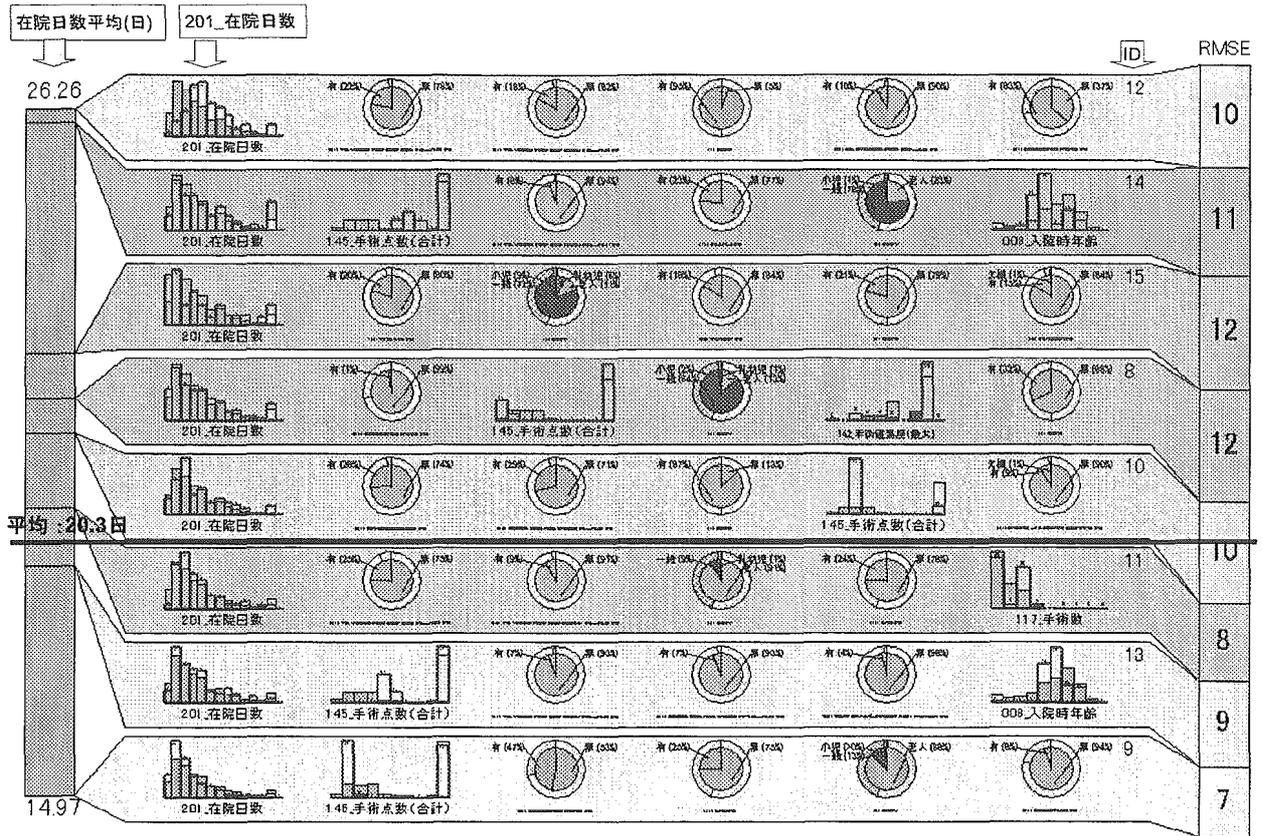


MDC6=070040 骨軟部の悪性腫瘍(脊椎脊髄を除く。) 出来高換算額/日 RBF

各クラスターごとの出来高日額平均およびサイズ

ID	サイズ%	出来高日額平均(¥)
8	24.6	38,232
9	25.4	32,522
12	15.6	31,907
10	11.4	30,468
14	6.7	30,199
13	6.7	29,743
15	2.8	27,720
11	7.0	22,873

MDC6=070040 骨軟部の悪性腫瘍(脊椎脊髄を除く。) 在院日数 RBF

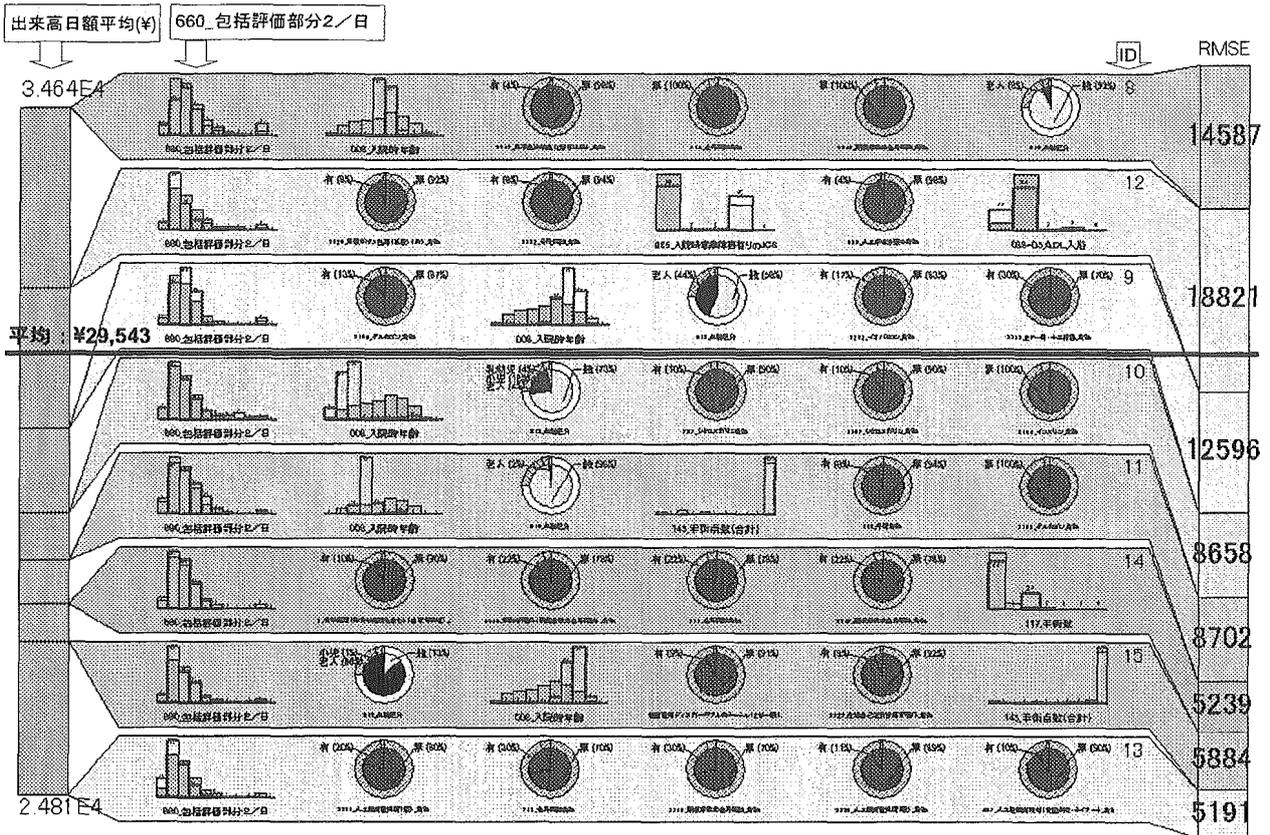


MDC6=070040 骨軟部の悪性腫瘍(脊椎脊髄を除く。) 在院日数 RBF

各クラスターごとの在院日数平均およびサイズ

ID	サイズ(%)	在院日数平均(日)
12	7.2	26.3
14	8.9	26.0
15	10.3	22.2
8	15.6	21.5
10	10.5	20.9
11	15.9	19.7
13	10.4	18.8
9	21.3	15.0

MDC6=070560 リウマチ以外の全身性臓器障害を伴う自己免疫性疾患 出来高換算額/日 RBF

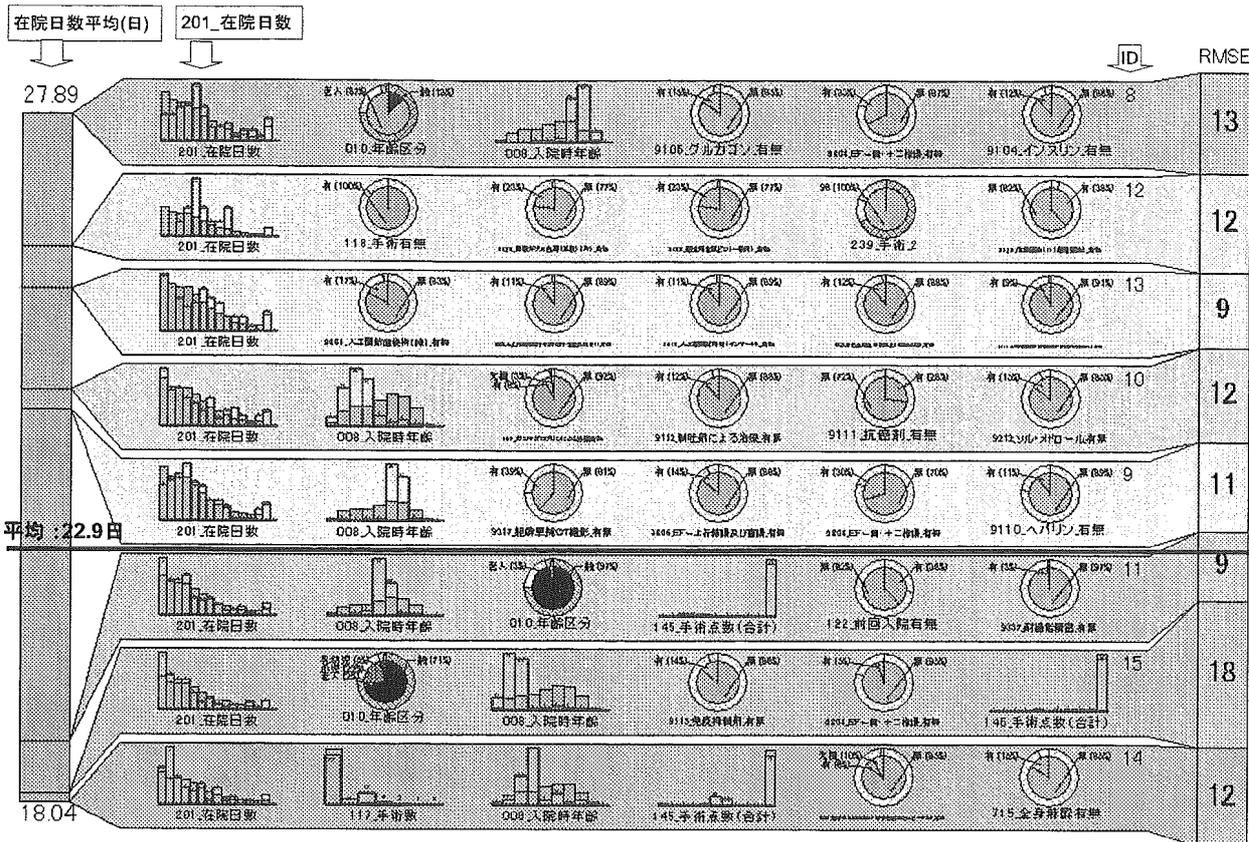


MDC6=070560 リウマチ以外の全身性臓器障害を伴う自己免疫性疾患 出来高換算額/日 RBF

各クラスターごとの出来高日額平均およびサイズ

ID	サイズ(%)	出来高日額平均(¥)
8	17.6	34,639
12	6.7	32,077
9	13.5	30,072
10	21.2	28,857
11	16.8	28,186
14	5.5	27,552
15	10.6	27,016
13	8.2	24,809

MDC6=070560 リウマチ以外の全身性臓器障害を伴う自己免疫性疾患 在院日数 RBF

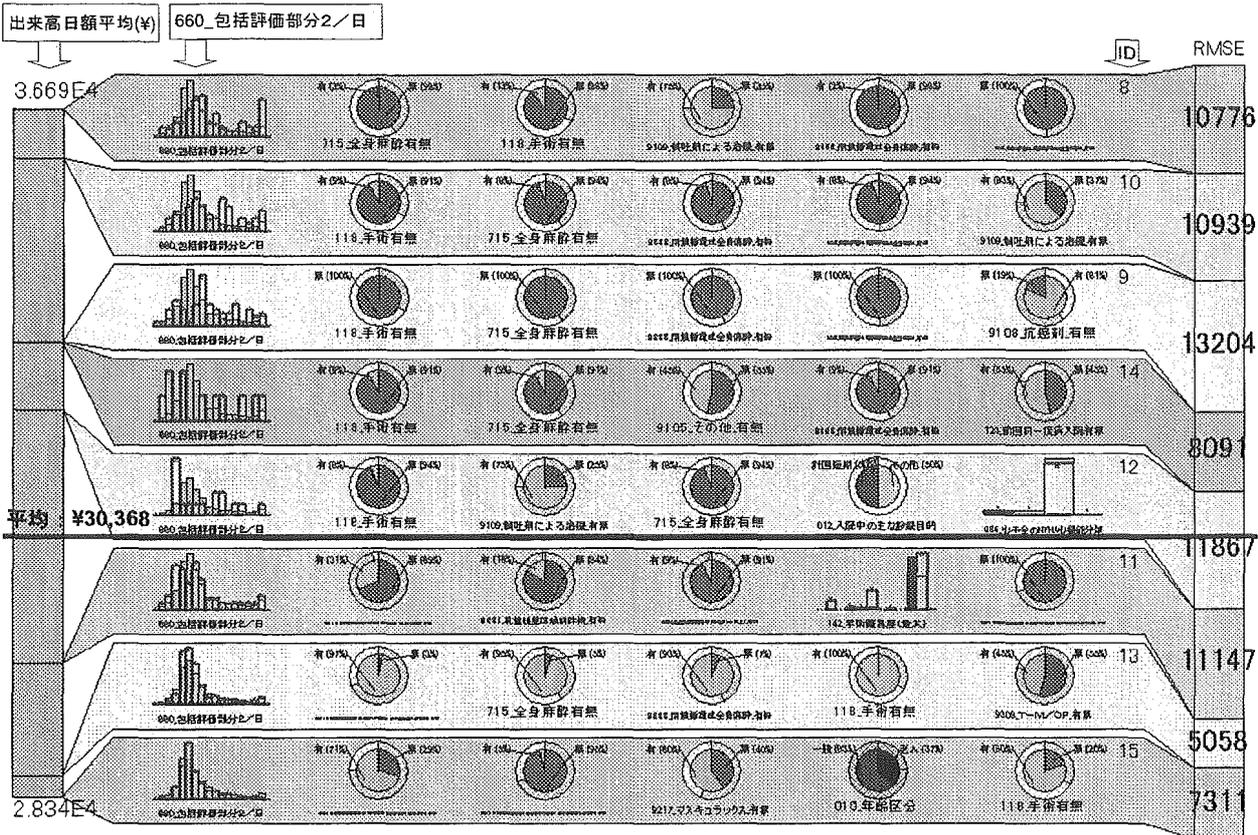


MDC6=070560 リウマチ以外の全身性臓器障害を伴う自己免疫性疾患 在院日数 RBF

各クラスターごとの在院日数平均およびサイズ

ID	サイズ(人)	在院日数平均(日)
8	18.5	27.9
12	1.7	26.0
13	10.0	25.4
19	15.6	24.0
9	20.1	23.7
11	15.2	18.9
15	12.2	18.2
14	6.6	18.0

MDC6=090010 乳房の悪性腫瘍 出来高換算額/日 RBF

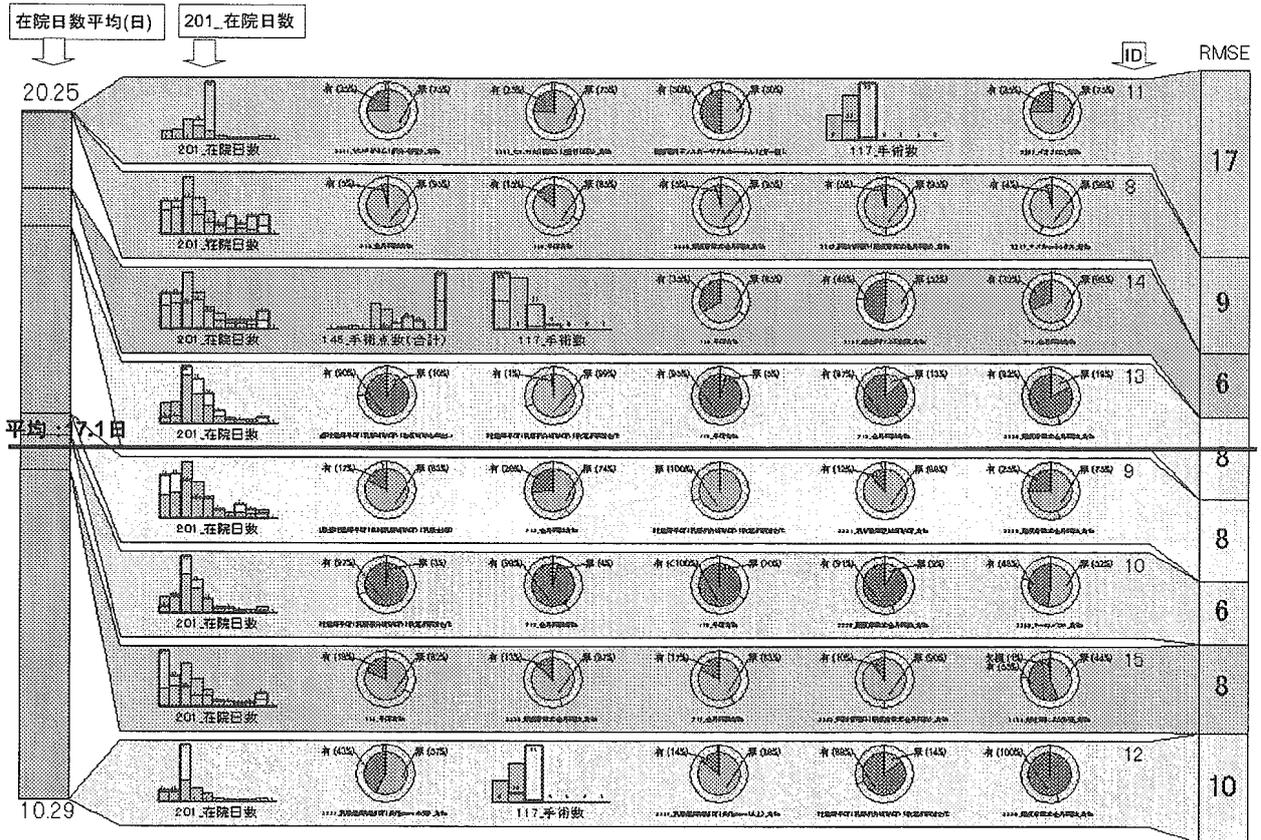


MDC6=090010 乳房の悪性腫瘍 出来高換算額/日 RBF

各クラスターごとの出来高日額平均およびサイズ

ID	サイズ(%)	出来高日額平均(¥)
8	7.1	36,692
10	7.9	36,096
9	7.6	33,882
14	1.3	33,862
12	1.9	33,048
11	9.5	29,978
13	30.6	28,606
15	34.1	28,341

### MDC6=090010 乳房の悪性腫瘍 在院日数 RBF

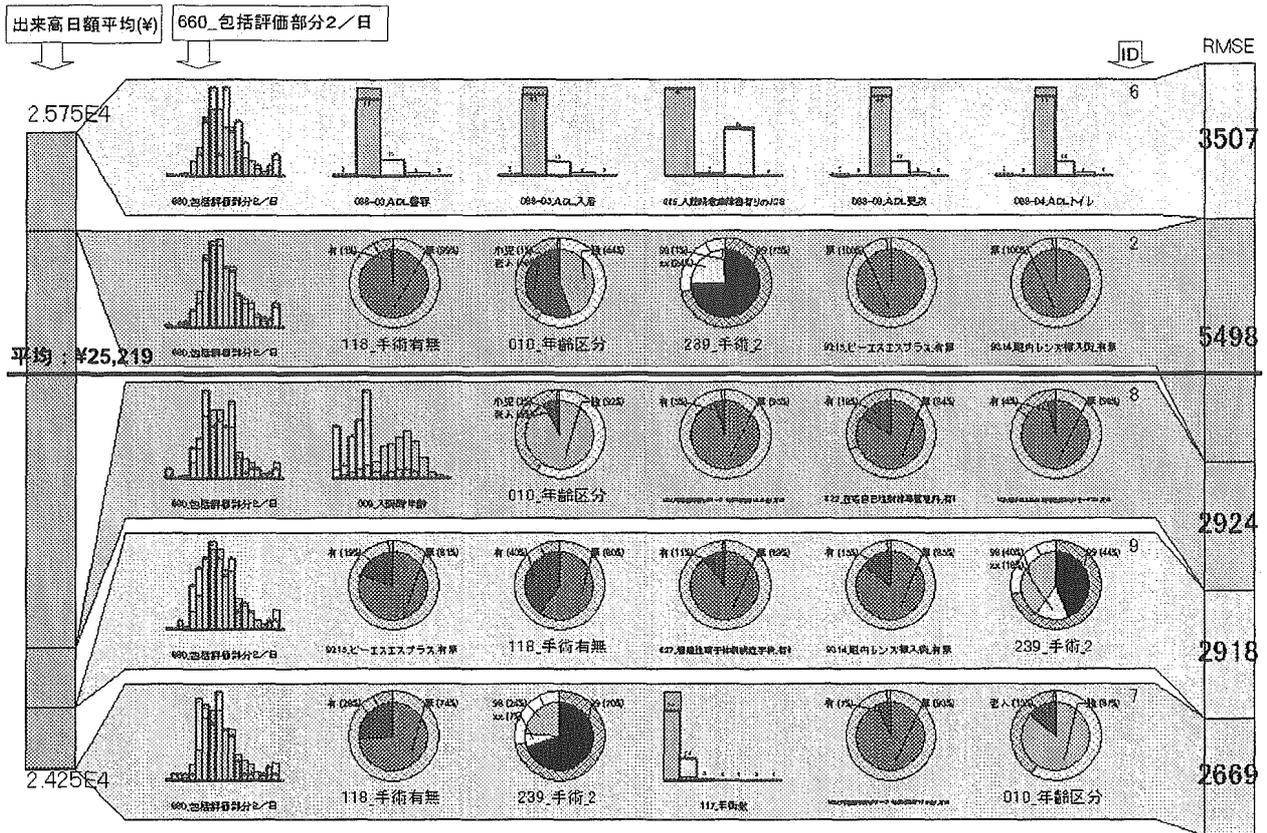


### MDC6=090010 乳房の悪性腫瘍 在院日数 RBF

各クラスターごとの在院日数平均およびサイズ

ID	サイズ%	在院日数平均(日)
11	0.5	20.3
8	12.0	20.2
14	7.7	19.1
13	24.1	18.6
9	17.6	15.9
10	29.5	15.6
15	7.9	15.1
12	0.8	10.3

MDC6=100070 2型糖尿病(糖尿病性ケトアシドーシスを除く。) 出来高換算額/日 RBF

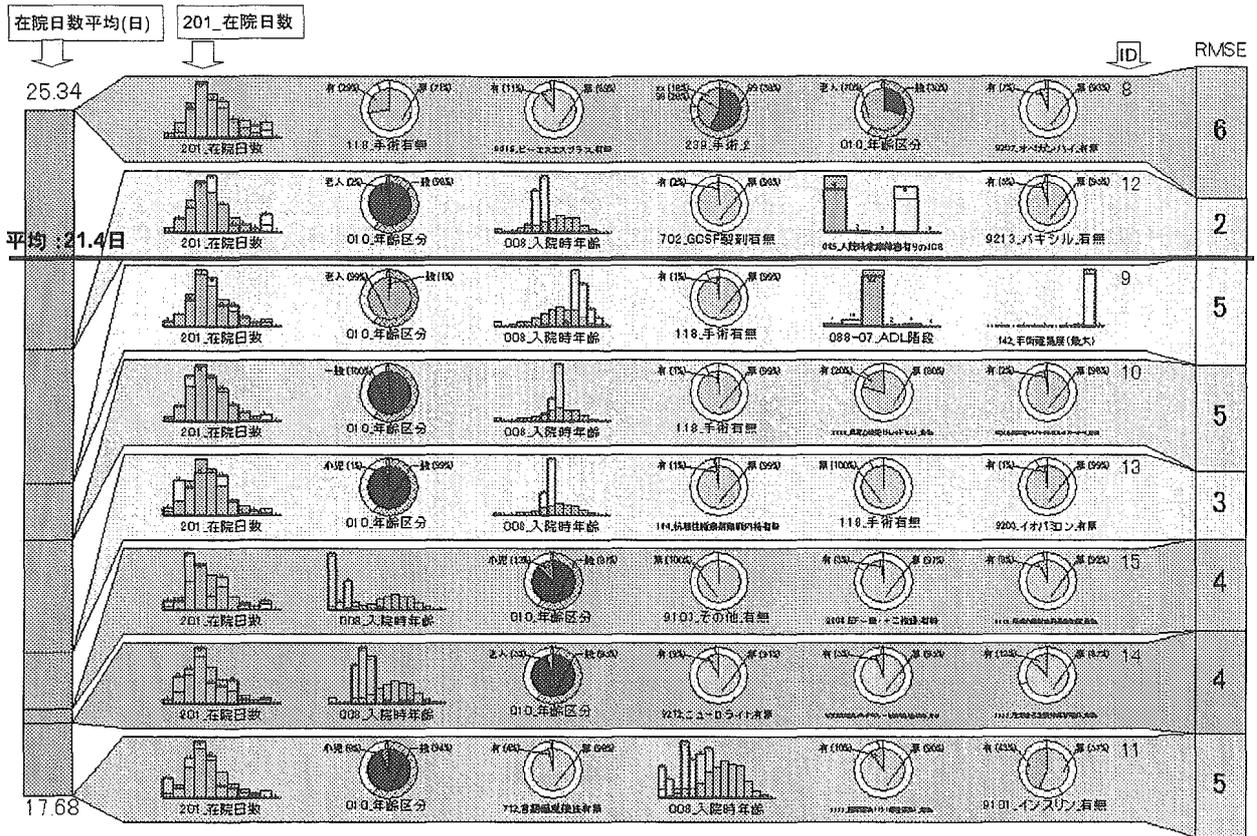


MDC6=100070 2型糖尿病(糖尿病性ケトアシドーシスを除く。) 出来高換算額/日 RBF

各クラスターごとの出来高日額平均およびサイズ

ID	サイズ%	出来高日額平均(¥)
6	4.9	25,745
2	67.5	25,515
8	11.0	24,533
9	7.5	24,390
7	9.2	24,245

MDC6=100070 2型糖尿病(糖尿病性ケトアシドーシスを除く。) 在院日数 RBF

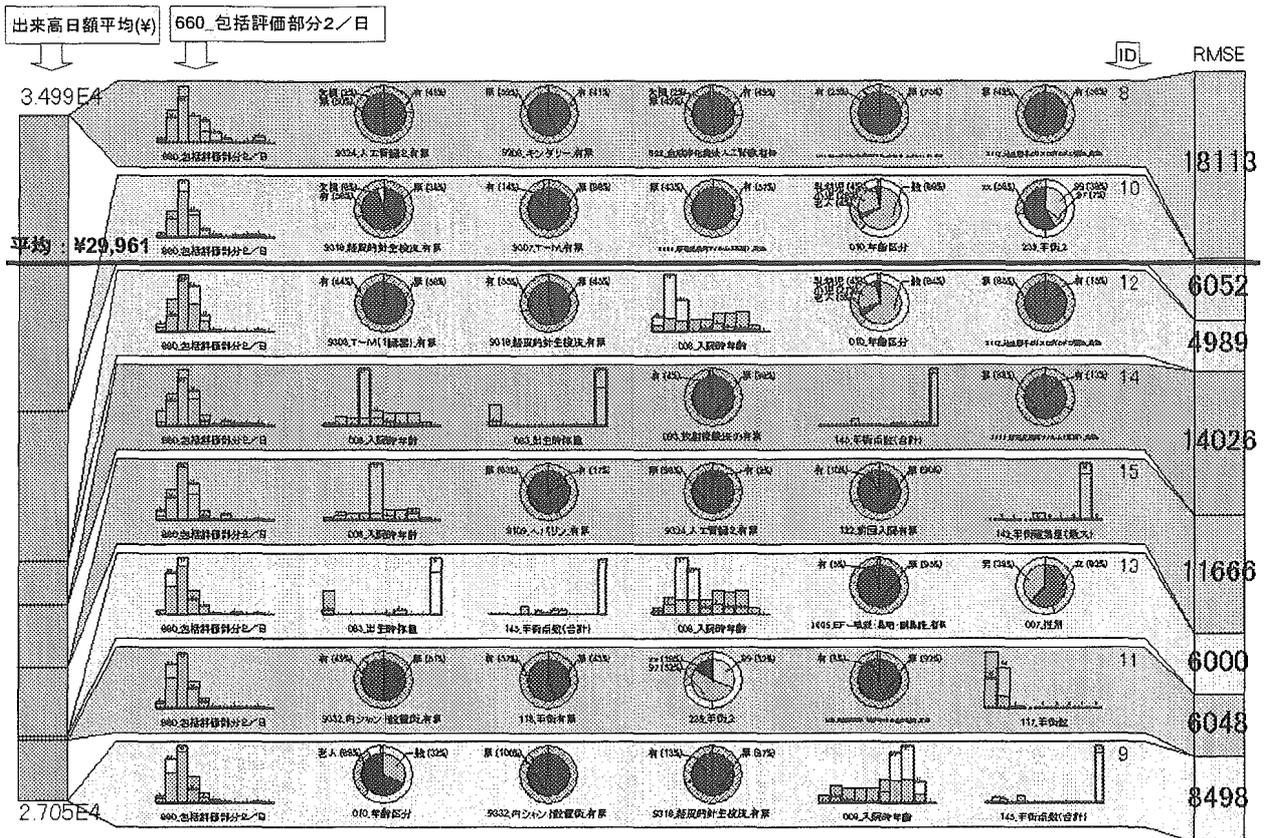


MDC6=100070 2型糖尿病(糖尿病性ケトアシドーシスを除く。) 在院日数 RBF

各クラスターごとの在院日数平均およびサイズ

ID	サイズ(%)	在院日数平均(日)
9	23.8	25.3
12	6.8	22.7
8	22.2	21.2
5	17.8	20.6
10	9.6	19.3
13	4.5	18.7
14	6.3	18.5
11	9.1	17.7

MDC6=110280 慢性腎炎症候群・慢性間質性腎炎・慢性腎不全 出来高換算額/日 RBF

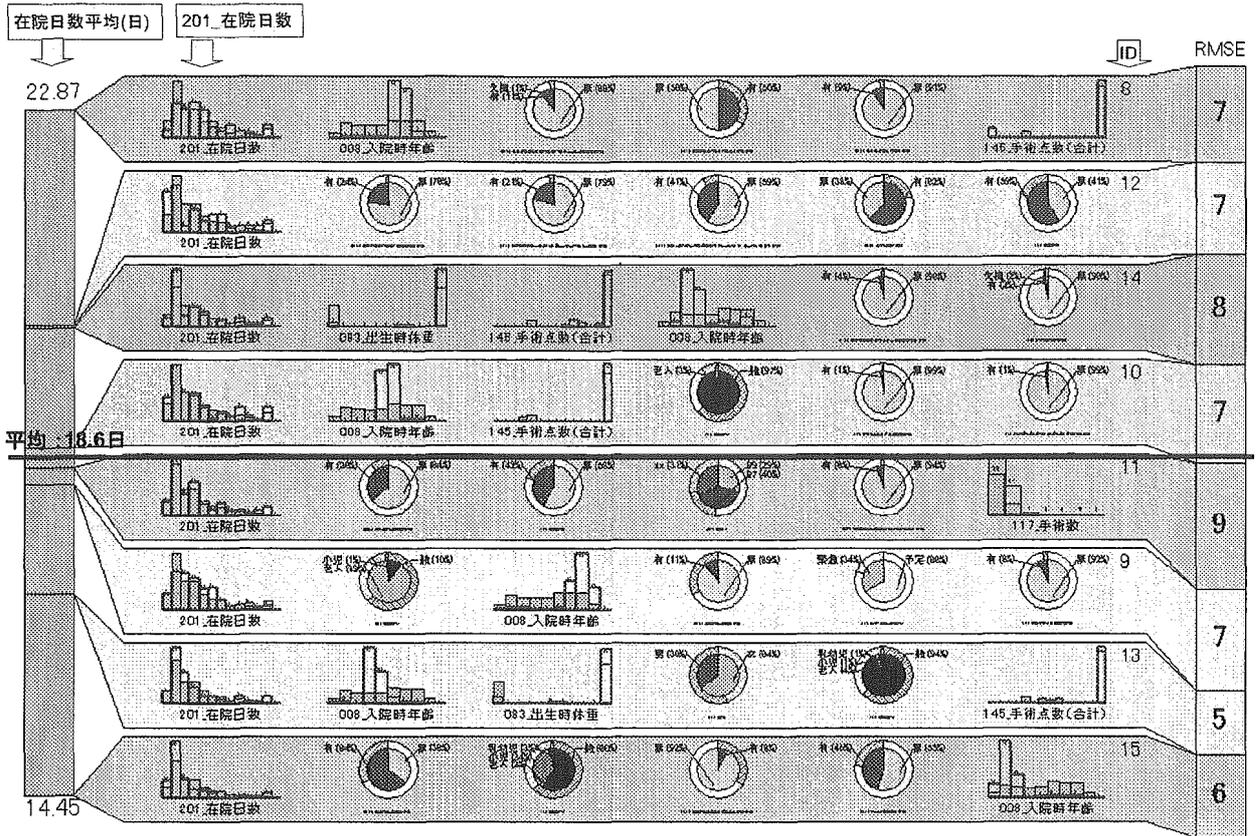


MDC6=110280 慢性腎炎症候群・慢性間質性腎炎・慢性腎不全 出来高換算額/日 RBF

各クラスターごとの出来高日額平均およびサイズ

ID	サイズ(%)	出来高日額平均(¥)
8	18.5	34,993
10	13.5	31,580
12	14.6	29,835
14	7.0	29,333
15	5.2	28,598
13	5.3	27,791
11	21.0	27,749
9	15.0	27,047

MDC6=110280 慢性腎炎症候群・慢性間質性腎炎・慢性腎不全 在院日数 RBF

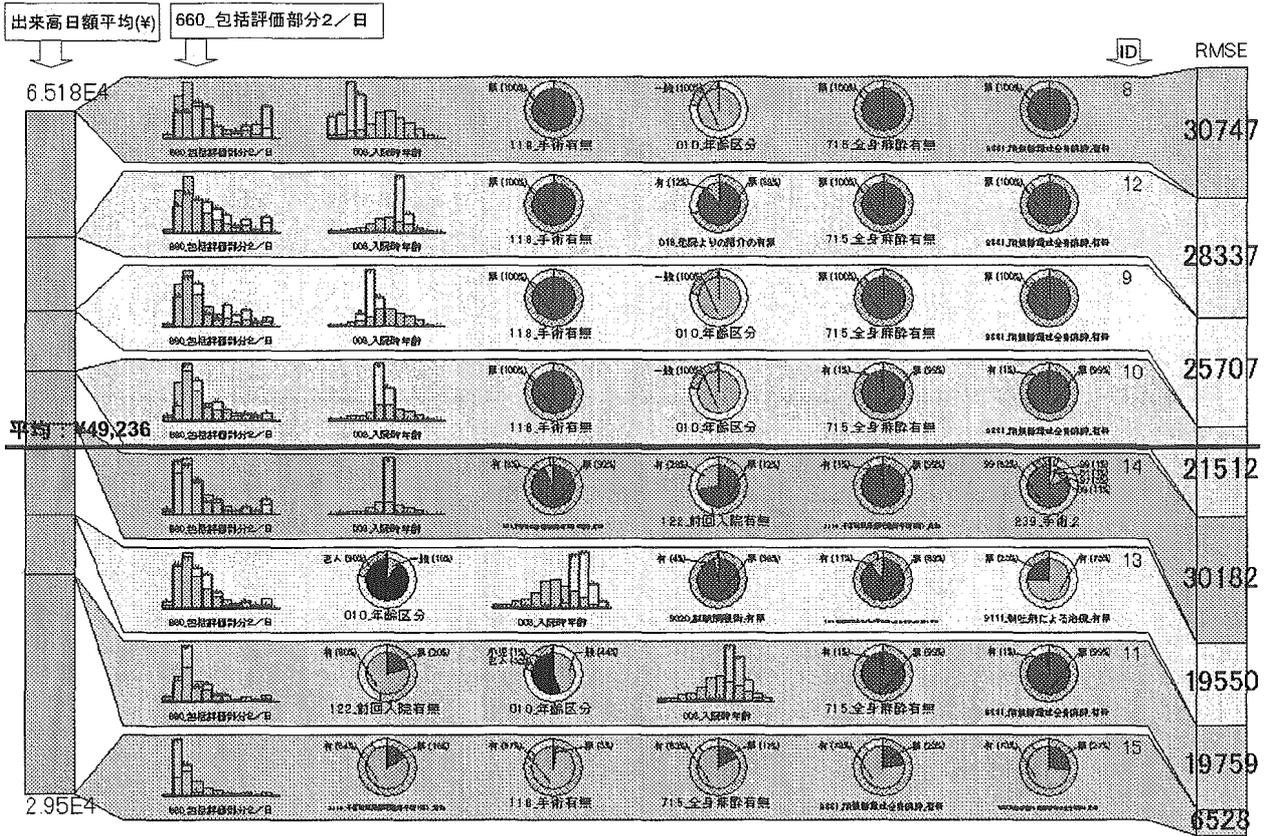


MDC6=110280 慢性腎炎症候群・慢性間質性腎炎・慢性腎不全 在院日数 RBF

各クラスターごとの在院日数平均およびサイズ

ID	サイズ(N)	在院日数平均(日)
7	15.3	22.9
12	4.1	20.2
14	5.7	20.2
10	8.8	18.7
13	22.9	18.5
9	19.9	18.3
11	9.3	16.9
15	14.0	14.5

MDC6=120010 卵巣・子宮附属器の悪性腫瘍 出来高換算額/日 RBF

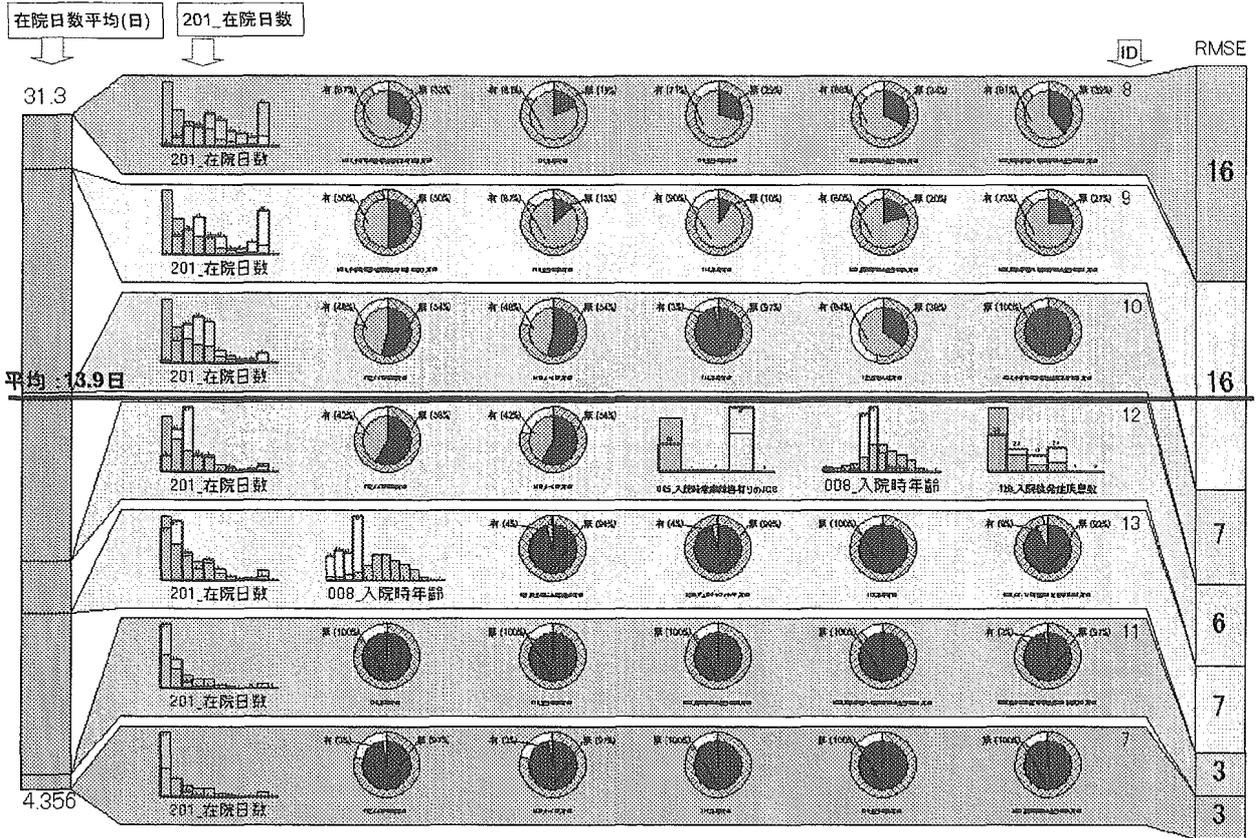


MDC6=120010 卵巣・子宮附属器の悪性腫瘍 出来高換算額/日 RBF

各クラスターごとの出来高日額平均およびサイズ

ID	サイズ(%)	出来高日額平均(円)
8	10.8	65,177
12	10.0	58,689
9	15.3	54,786
10	16.5	61,695
14	11.6	48,906
13	10.9	44,157
11	13.6	41,080
15	11.2	29,503

MDC6=120010 卵巣・子宮付属器の悪性腫瘍 在院日数 RBF



MDC6=120010 卵巣・子宮付属器の悪性腫瘍 在院日数 RBF

各クラスターごとの在院日数平均およびサイズ

ID	サイズ%	在院日数平均(日)
8	14.1	31.3
9	3.0	29.2
10	23.6	20.2
12	7.5	13.5
13	4.9	11.4
11	30.7	5.0
7	16.2	4.4

## 参考資料2 データ・マイニング・セミナー資料

# 長寿科学総合研究事業

「国立大学病院医療に及ぼすDPC導入の影響－  
主要老年病の医療・治療の標準化・充実を目指して」研究班主催

## データマイニング基礎講座

於 東京医科歯科大学歯学部特別講堂

2004. 8. 23～24

日本アイ・ビー・エム株式会社  
ライフサイエンス事業部  
河村佳洋  
eb04987@jp.ibm.com

※当資料は二次配布無用にてお願いします※

© 2004 IBM Corporation.

### 目 次

1. 基礎	
1) データマイニングの概説	3
(1) データマイニングの概念	
(2) データマイニングの概要	
(3) 統計解析とデータマイニング	
2) データマイニングの理論	27
(1) リスク分析	
(2) クラスタリング	
(3) モデリング	
2. 応用	
1) データマイニングの実践	96
(1) データマイニングの手順	
(2) 統計解析とデータマイニングの補完関係	
2) データマイニングの事例	108
※当配布資料にては一部割愛させていただきます※	
3. ツール	111
1) データマイニング・ツール	
2) IBM DB2 Intelligent Miner	

## データマイニングの概説

従来の解析手法とは大きく概念を異にしている。

3

© 2004 IBM Corporation.

## A社によるIntelligent Minerの評価

1. 自分で有効性を納得できねば、上司へ提案できないと徹底査定を宣言された。  
備考]幸運であったのはツール査定者としてトップ技術者のアサイでした。
2. 新たな発見が期待できるとの売り込みも新発見はなかったと苦言を呈された。  
注]同社では、該当ケースについて長期(10年以上)に亙り徹底説明されていた。
3. 重回帰分析では多大な時間を要してたが、予想を超え短期に問題解明できた。  
備考]短期解明は非常に価値があると高い評価を頂いた。
4. 生データによる試行解析により、データマイニングの威力は十分に確認できたが、前例がないから今回は無難を期し、導入資金は研究開発費から出したい。  
備考]当時は技術系分野へのデータマイニング導入実績が皆無でした。
5. 斯くも有効な結果が得られるなら、設備投資として処理しても良いが、...。  
備考]結果報告を聞かれた所長が、裁定を求められた折りの所感として。

備考]技術系に対して、汎用データマイニング・ツールの提供を開始した折りに遭遇したプロスペクトの反応を提示し、データマイニングが如何に受け留められていたか、ご参考までに紹介する。

4

© 2004 IBM Corporation.

## B社によるIntelligent Minerの評価

1. 最初に結果の妥当性について確認したい。
2. 今回の結果には知見との不整合は見当たらない。
3. 全く専門知識のないツールにして、何故このような結果が出せるのか。  
注]長年の解析に基づいてハンドブック(数百頁)を作成されていた。
4. 新しいケースのため知見も少なく、一部に知見にはない新しい結果も見られるが、妥当か不当か判断できない。ただ、何か新たな示唆の可能性も予感させるから、現場技術者に伝え実地追及させたい。  
備考]追試の結果を聞き漏らしたが、実際にご導入を頂いたことから、有効な結果であったものと推察している。

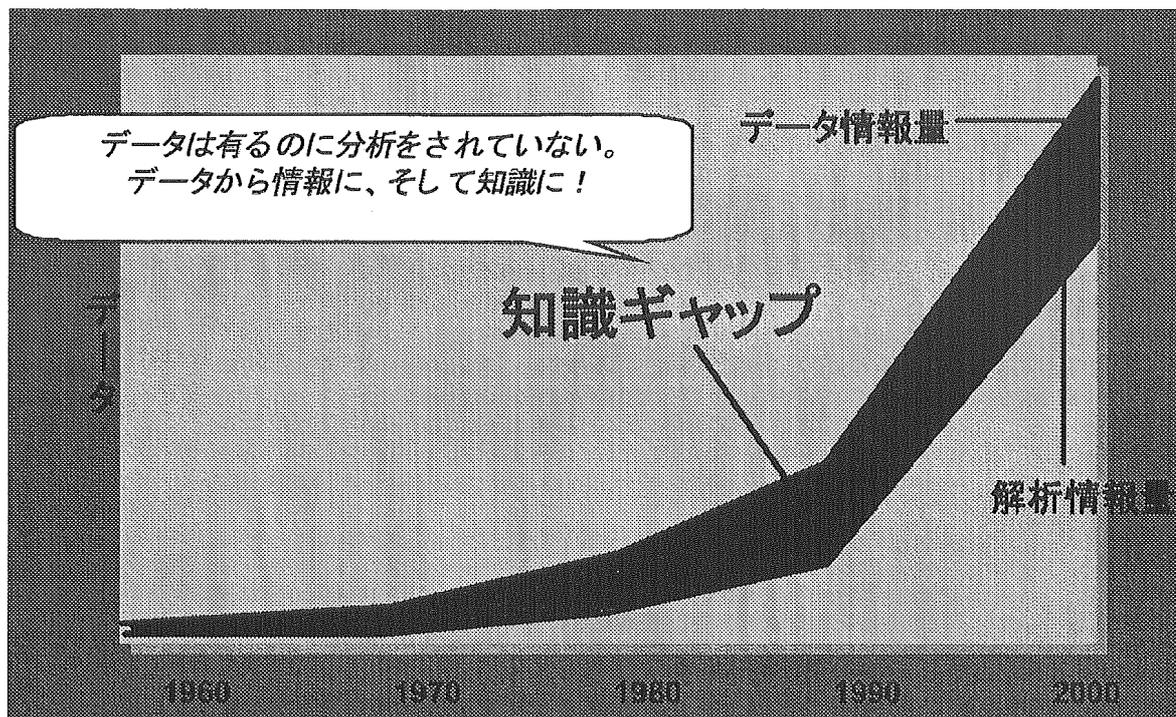
備考]データマイニングにより数百億円とも推定される損失危機を回避された。  
蓋し、世界で最も効果を挙げた適用事例かも知れない。

## C社によるIntelligent Minerに対する評価

1. 生データによる試行解析に立ち合わせて下さい。  
注]セリング・フェーズでの要望でした。
2. 試行結果について、知見通りの結果であるとの評価を頂いた。  
備考]熾烈な競合を展開したが、最終的にご導入を決定頂いた。  
補足]同社系列会社が競合ツールの特約店でした。
3. 難解な問題を抱えているが、実習テーマには出来ないか。  
備考]導入時の講習会において、製品を低コストで実現する添加物と原料の投入比率に関する最適化問題でした。
4. 誠に素晴らしく低コストで良質の結果が得られたとの評価でした。  
備考]講習会で突然提示されたテーマとデータでしたが、僅か2日間の短期コースながら、前準備に多くの時間を割く必要はなかった。  
注]従来手法による分析に向け十分なデータ整備がされていた。

備考]以上は、実際に経験した顧客反応の列挙ですが、3社は何れも技術系の会社です。

## 知識ギャップの増大



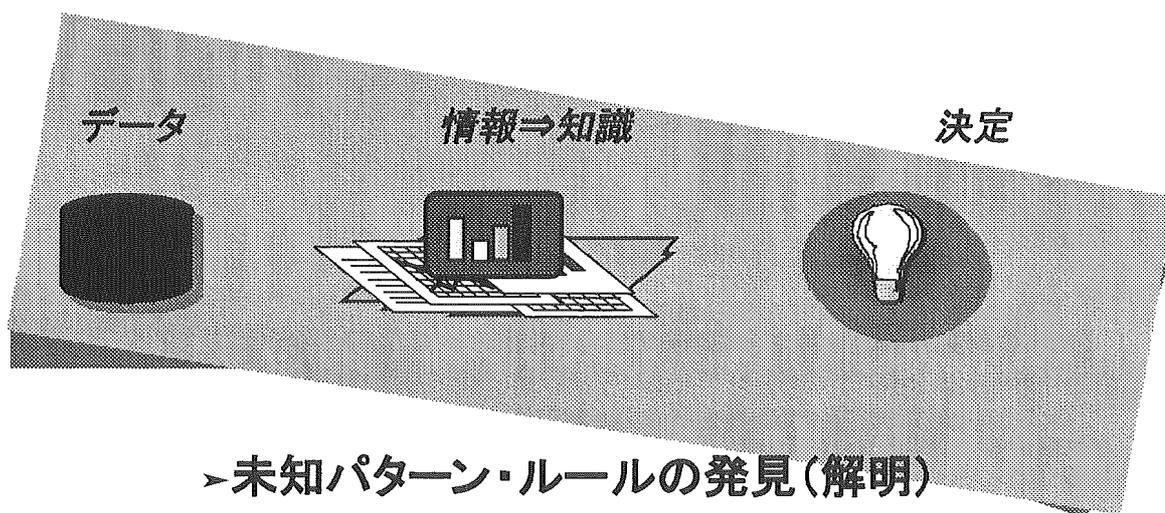
Source: Gartner Group

注] 今日では、当時(1990)のギャップ予測を遥かに上回る増大傾向を辿っている。

7

© 2004 IBM Corporation.

## データを意思決定へ活用



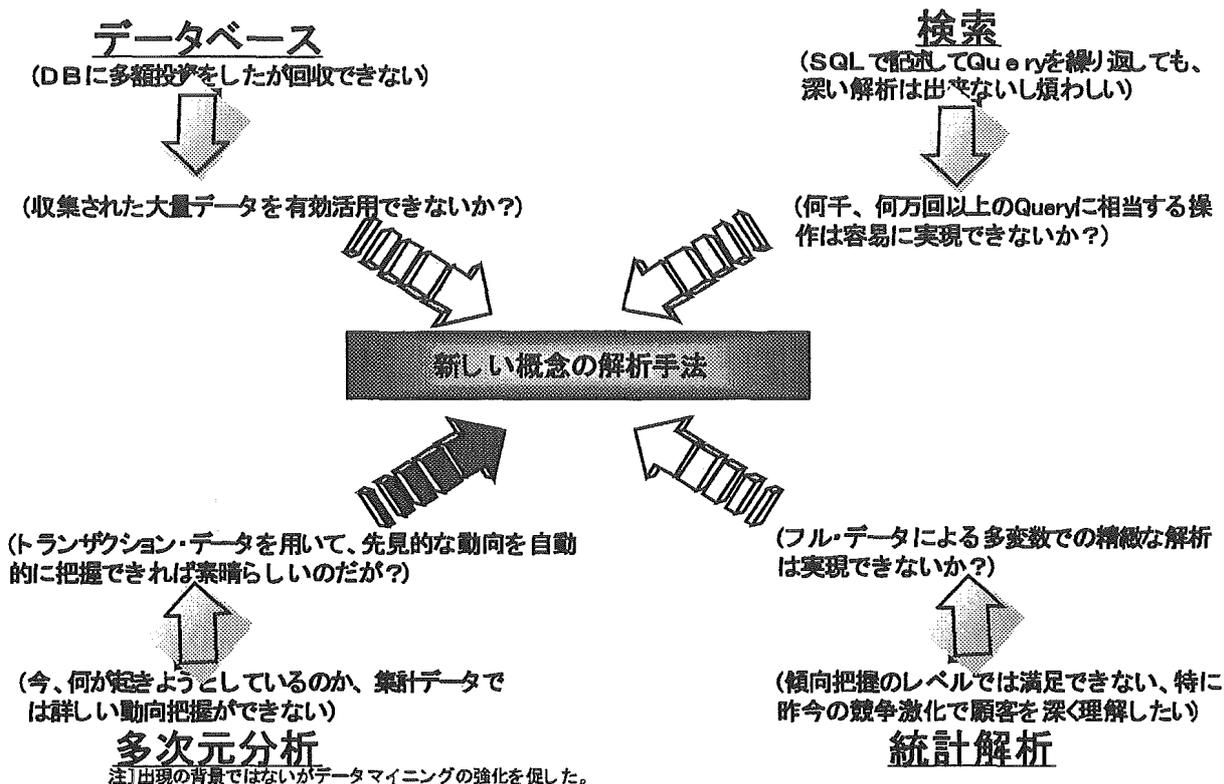
- ▶ 未知パターン・ルールの発見(解明)
- ▶ 履歴データに基づく未知予測(推定)

補記1] 不当なデータや不適正な処理によって得た情報に価値はなく寧ろ危険である。

補記2] 価値ある情報も正しく解釈されなければ知識に成らず役に立たない。

8

© 2004 IBM Corporation.



備考]コンピュータ技術と数理技術の飛躍的な発展がデータマイニングの出現を可能にしたことは言うまでもない。

## データマイニングの概念

共通の概念として、巨大データの中に埋もれた未知の価値ある知識を自動的に発見するプロセスのことである。

### 1. 新しい概念の登場

種々の違いは見られるが、主要な差異は次の点である。

- 1) 統計解析、多変量解析、... → 仮説検証型、線形解析、...
- 2) データマイニング → 知識発見型、非線形解析、...

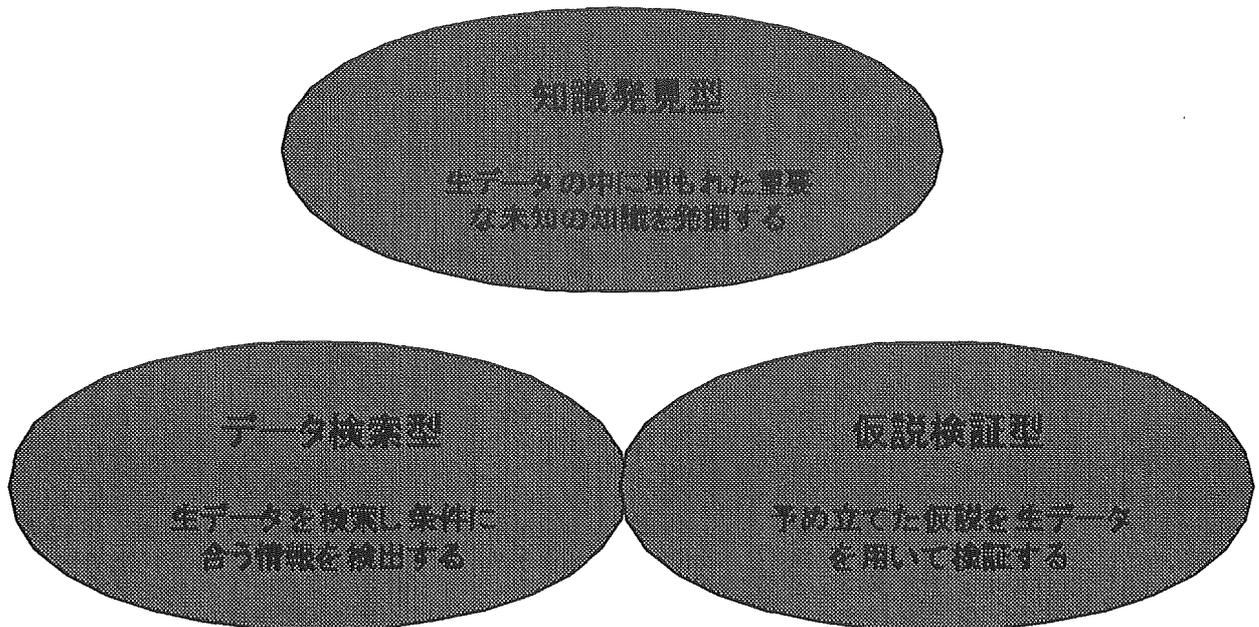
備考]巨大データを想定しており、小規模データの場合は、相応の結果が得られても、仮説発見の域は出ないため、マイニング結果の追試を要する。

### 2. 出現から普及へ

進化した人工知能と伝統的な統計理論に基づくデータマイニングは、1993年頃からDr. Agrawalなどにより呼称され始めた造語であるが、1980代末には登場しており、データベース・マイニング等と称された。現在では、1996年に発表されたKDDプロセスの概念が、データマイニングの目指すところと見做されている。我が国でデータマイニングの普及開始は1990年代末頃のことである。

補足]実質的には、世界初の汎用データマイニング・ツールとして、当初はIBM社内専用に開発のNNU(Neural Network Utility)が挙げられる。(1980年代末)

## データ解析手法の分類

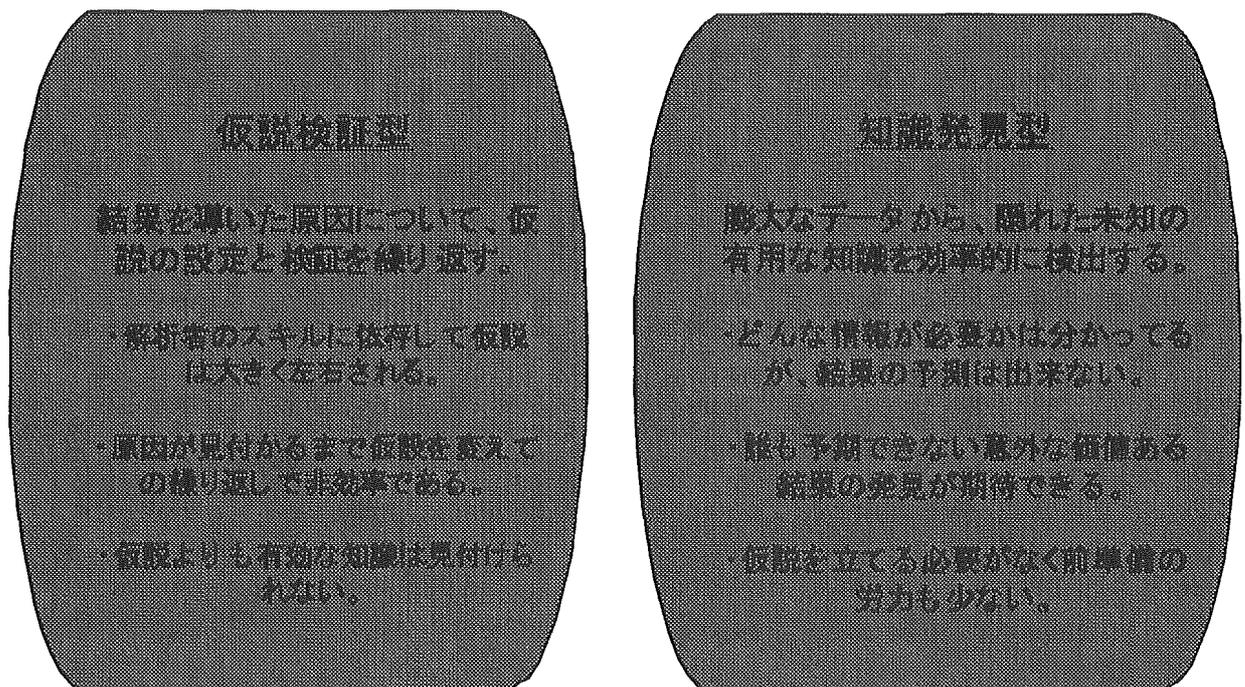


備考]大枠ではデータ検索型も仮説検証型の範疇に入る。

11

© 2004 IBM Corporation.

## 仮説検証型と知識発見型



隠れているのはこれかな?…やっぱり。 何が隠れているのかな?…あっ、こんな!

12

© 2004 IBM Corporation.