

医療事務担当者のための データマイニング入門 ～DPCデータが教えてくれること～

石川 ベンジャミン 光一
国立がん研究センター がん対策情報センター
がん統計研究部 がん医療費調査室長

診療報酬の請求業務

- ▶ 業務で利用しているデータ：
オンライン請求のため、大量の電子データが蓄積
- ▶ 記載されている情報：
レセプト電算処理マスターコードにより、
コンピュータで集計可能な形式になっている
↓
- ▶ 支払基金では(9/30記者会見発表)、
コンピュータチェックの寄与率：55.2%(←昨年47.4%)
 - ▶ チェックマスターの活用、突合点検、縦覧点検
←既存のルールに従ったデータの処理が中心
- ▶ 今後は…
データから教わりながら、その取扱いを考える局面へ

データマイニングとは…

- ▶ (大量の) **データ** から (貴重な) **知識** を見つけ出すこと
 - “データベースからの知識発見”
Knowledge Discovery in Databases (KDD) と呼ばれる
- ▶ 最近では、“**ビッグデータ**”分析というキーワードで語られることが多くなっている
 - ▶ 電子データを集めるコストが安くなっている
← 業務のオンライン化、自動測定データ、ソーシャルネットワーク
 - ▶ 大規模なデータの保管・集計が容易になっている
← 大容量ディスク/SSD、マルチコアCPU、分散処理、Hadoop
- ▶ その時々で“**ホット**”な技術は入れ替わるものの、
重要なのは…使っている道具ではなく、
生み出された **知識** と、その **知識** を現場で生かす **知恵**

データが教えてくれること…

例えば、コンピュータチェックの活用場面を考えると、

- ▶ **傷病名と医薬品** のチェック
 - 医療費へのインパクトが大きい医薬品、傷病は？
 - ▶ **医薬品と用量** のチェック
 - 添付文書にある標準量・極量ではなく、実際の使用量は？
 - ▶ **傷病名と診療行為** のチェック
 - 手術に際して標準的な輸血量の範囲は？
- 単純なチェックを超えて、よりきめ細かな対応をする上では、
- ▶ **症状詳記** のチェック
 - 適応と 使用量/回数 の観点から、必要性を判断

標準的な診療内容
を知ることにより
特殊な症例を発見

適切な症状詳記を
添付する

医薬品の使用に関するチェック

- ▶ yes/noルールで済む処理
 - 医薬品の使用あり、適応する傷病名なし → 査定
- ▶ 使用量に関する処理
 - 添付文書等に**極量**が記載されている
 - ▶ 極量を超えた処方 → 査定/返戻
 - 添付文書に**標準量**が示されている
 - ▶ 標準量として**1日量**が記載されている → 使用量に基づいて、査定
 - ▶ 標準量は**体重あたり**で記載されている
 - レセプトに体重の記載はないので、データの分布に基づいて基準を定める → 極端に多い/少ないレセプトを査定/返戻 **データマイニング→閾値**
 - **“疾患や年齢・症状により増減” / 内服薬の複数日処方**
 - ▶ 傷病名等の背景情報も含めた判断が必要 **データマイニング+個別処理**

「DPC制度」とは…

- ▶ “DPC調査”で収集される、**標準化された診療データ**
 - ▶ 様式1 …… 退院サマリ=傷病名(副傷病)、入院日数、重傷度
 - ▶ E/Fファイル … 入院中の診療行為・資源投入量
 - ▶ “DPC”という**患者分類手法**(casemix分類)
 - ▶ DPC = **D**iagnosis **P**rocedure **C**ombination
 - ▶ 分類ごとの資源投入量の分析に基づいて設計される**包括支払のルール**(PDPS = **P**er **D**iem **P**ayment **S**ystem)
 - ▶ (入院1日あたり額×日数×調整係数) + 出来高払い
- ↓
- ▶ これら3つの要素から構成される複合体(DPC/PDPS)です。

「DPCデータ」とは…

医療施設情報 + 患者情報 + 診療行為情報の複合体

- ▶ **医療施設**：機能、基準…→**structure**
 - 様式3/様式6
 - ▶ **患者**：「入院」についての要約… →**case mix**
 - 様式1 = 入退院情報、傷病名、手術 ← 「退院サマリ」
 - ▶ DPCを決定するため(コーディング)に必要な情報 + 補足情報
 - ▶ **診療**：いつ、どこで、誰が、誰に、何を…→**process**
 - 患者に提供したサービスを(全て)リストアップ→**E/Fファイル** (出来高払いの請求範囲で…)
 - DPCで請求した通りに記述→**Dファイル** (DPC対象病院のみ)
 - ▶ 「レセプトデータダウンロード方式」
- 医療機関から提出の時点で匿名化済み

一般病床の過半数、入院件数の6割以上をカバーするデータ

DPCデータが得意とすること

- ▶ **診療プロセスの解析**
 - DPC分類別、病院別の診療実態の把握
 - DPC分類別の入院日数(全体/術前/術後)、診療区分別点数
 - 病院間の比較
 - 薬剤利用の分析(化学療法など)、手術術式別の分析
 - より詳細な診療内容の分析…
 - ▶ 入院中の診療行為のバリエーション解析
 - ▶ 開腹 vs. 腹腔鏡手術の比較分析 など
 - ただし、出来高請求情報の精度に起因する制約がある
 - ▶ CT/MRIの撮影部位・方向、同月2回目以降のCT/MRIの別
- ▶ **病院の機能と診療圏**についての分析
 - 厚生労働省の公開データ
 - 様式1：患者住所地の7桁郵便番号を利用

DPCデータが不得手とすること…

▶ 治療成績の評価

- outcomeについてのデータは限定的（退院時死亡のみ）
 - ▶ 長期予後の分析のためには、別途臨床研究データが必要
- “adverse outcome”のコーディング
 - ▶ ICD10による情報の収集には限界がある…

▶ 詳細な臨床病期・重傷度、部位別の解析

- 様式1の重傷度等は、最低限度
 - ▶ UICC TNMのデータあり / ただし精度評価は未実施
 - ▶ 別途臨床研究データが必要（例：組織型、狭窄度、検査値…）
- 入院前の経緯・状態についての評価
 - ▶ 発症からの時間、前医での治療内容等は一部を除いて不明

しかしながら…

DPCデータを核として、必要なデータを追加・補完することで分析の効率化・迅速化を図ることが可能

EFファイルの分析にどう取り組むか…

▶ E/Fファイルは巨大

- H24研究班データ：入院・外来をあわせて40億件
 - ▶ 様式1：退院数-685万件、患者ID数480万人

▶ 取り回しを良くするために工夫する

- (多施設データでは)都道府県単位で分割
- データ区分(診療識別コード)で分割
 - ▶ 入院基本(90)/特定入院(92) / 食事療養(97) と それ以外のもの
- レセ電コードの先頭1桁で分割
 - ▶ 行為(1) / 薬剤(6) / 材料(7) / コメント(8)
- データの“粒度”をまとめる
 - ▶ レセ電 → 薬剤の成分(薬価基準コードの先頭7桁) → 薬効分類(4桁)
- 時間軸をまとめる
 - ▶ 1日複数行 → 1日1行 → 1入院1行 → 1患者1行…

時間軸の粒度

統合EF	Fファイル由来	Eファイル由来	時間軸の粒度				集計の視点
			キー	日	入院	患者	
EF-1 施設コード	F1施設コード	E1施設コード	E/F	●	●	●	施設別 / 施設の種類の …
EF-2 データ識別番号	F2データ識別番号	E2データ識別番号	E/F	●	●	●	患者属性別 …
EF-4 入院年月日(西暦)	F4入院年月日	E4入院年月日	E/F	●	●		年度別 / 季節別 …
EF-5 データ区分	F5データ区分	E5データ区分	E/F				
EF-6 順序番号	F6順序番号	E6順序番号	E/F				
EF-7 行為明細番号	F7行為明細番号		F				
EF-9 レセ電算コード	F9レセプト電算コード	E8レセプト電算コード	(F)				点数表コード別 / 薬剤別 …
EF-10 解釈番号	F10解釈番号	E9解釈番号					
EF-11 診療明細名称	F11診療明細名称	E10診療行為名称					
EF-12 使用量	F12使用量						
EF-13 基準単位	F13基準単位						
EF-14 明細点数	F14点数 / F15薬剤料 / F16材料料						
EF-15 円・点区分	F17円点区分	E14円点区分					
EF-16 出来高実績点数	F18出来高実績点数						
EF-17 出来高・包括フラグ	F19出来高包括フラグ						
EF-21 行為回数		E15行為回数					
EF-22 保険者番号		E16保険者番号					保険者別
EF-23 レセプト種別コード		E17レセプト種別コード					
EF-24 実施年月日		E18実施年月日	(E)	●			日数別 / 術前・術後日数別 …
EF-25 レセプト科区分		E19レセプト科区分					診療科別(全国共通)
EF-26 診療科区分		E20診療科区分					診療科別(病院内での分析用)
EF-27 医師コード		E21医師コード					医師別
EF-28 病棟コード		E22病棟コード					病棟別
EF-29 病棟区分		E23病棟区分					病棟区分別
EF-30 入外区分		E24入外区分					入院・外来別

統合EFファイルに含まれている情報

がん化学療法の分析

▶ 薬剤の成分別の分析が必要

- ▶ レセ電コード → 薬価基準コード → 化学療法薬剤の成分別
- ▶ 477種類のレセ電算コードを140の成分(AgentCD)に読み替えて分析

▶ 薬剤の組み合わせに関する分析が必要

- ▶ 140の成分の組み合わせを[0/1]で表現 → RegimenCD : 001010001…

▶ 薬剤の投与計画(時間軸)に関する分析が必要

- ▶ 日単位 での薬剤組み合わせ
- ▶ 入院単位 での薬剤組み合わせ
- ▶ 患者単位 での薬剤組み合わせ、組み合わせの変化

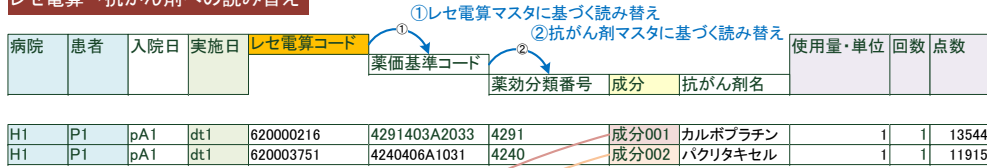
▶ 抗がん剤の投与に伴う事象に関する分析が必要

- ▶ 副作用対策として利用される薬剤：制吐剤、造血薬/輸血
- ▶ 副作用として報告される病名、転帰

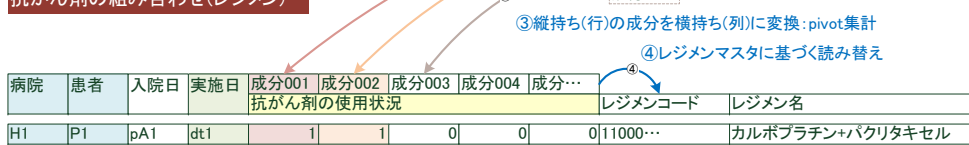
事前準備：E/Fファイルの連結



レセ電算→抗がん剤への読み替え

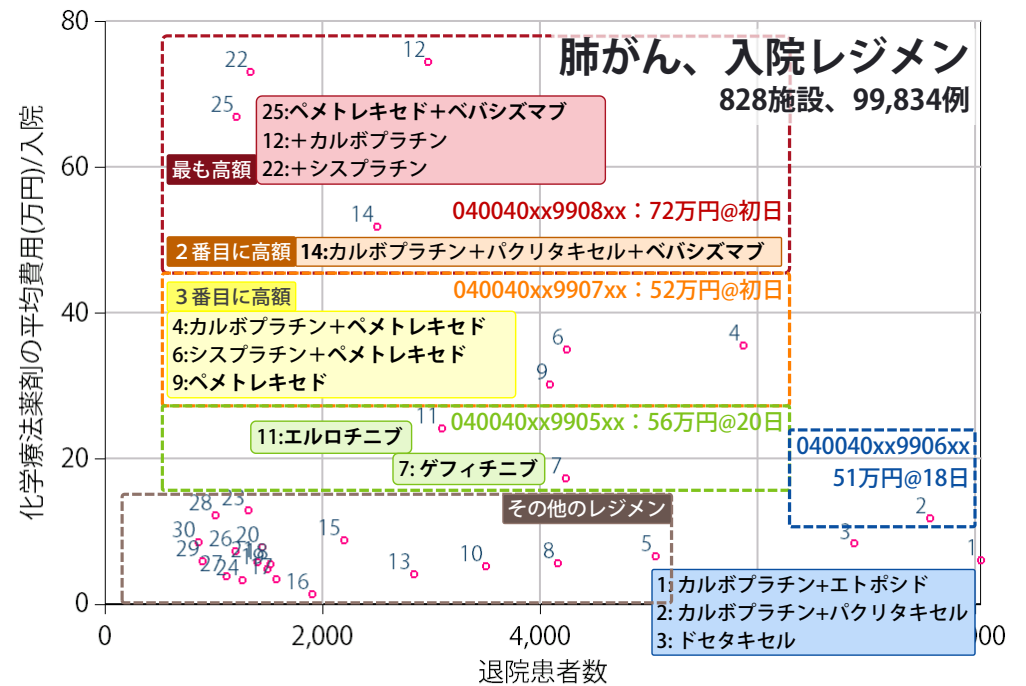


抗がん剤の組み合わせ(レジメン)

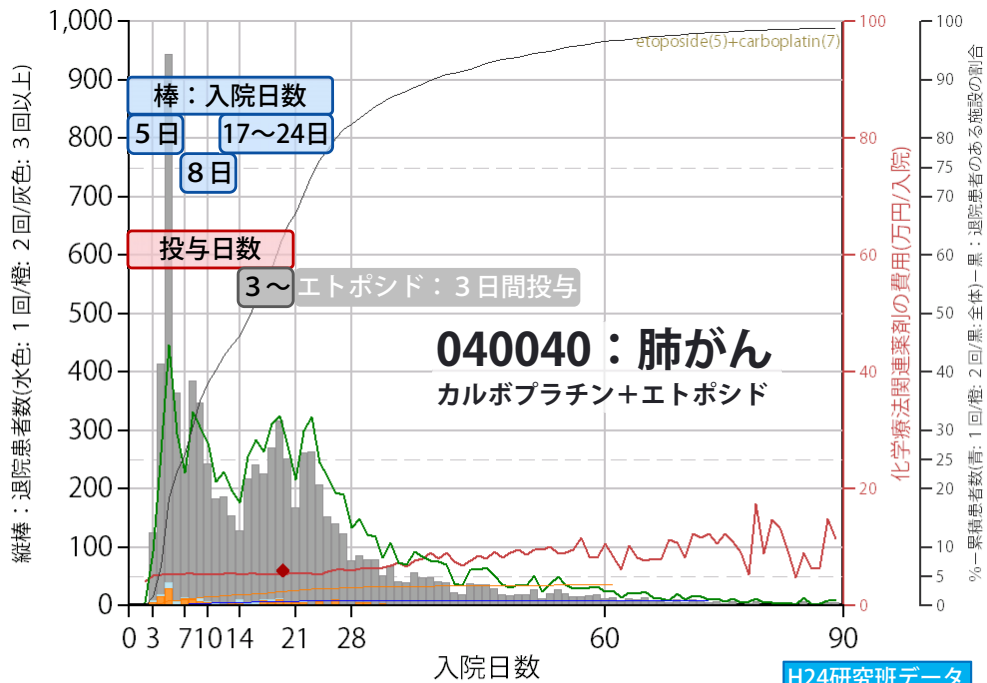


化学療法レジメンの集計法

kishikaw@ncc.go.jp/20131214/伏見班高知セミナー

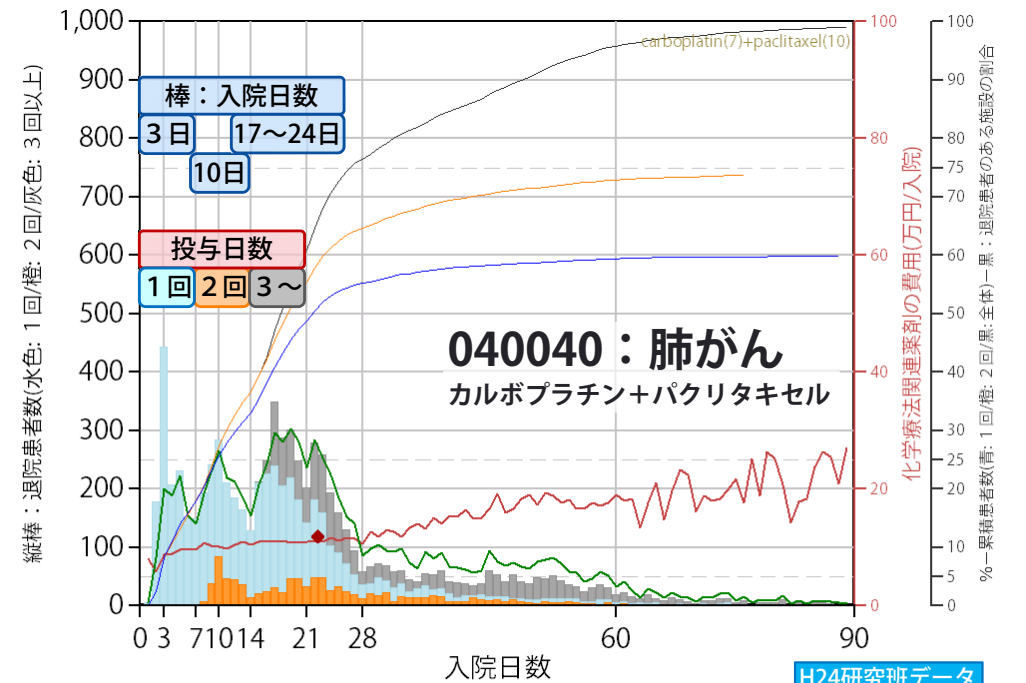


注)入院レジメン：1入院中に使用された薬剤の組み合わせ 040040 / 99834例 / 828施設 / H2 H24研究班データ



040040：Rank(01)：Regimen(018) / 退院患者数：8051(8.1%) / 施設数：541(65.3%)

棒：退院患者数(水色：1回/橙：2回/灰色：3回以上) / 青&橙&黒：累積患者数 / 緑：退院施設の割合(%) / 赤：費用(万円/入院)・点は平均値

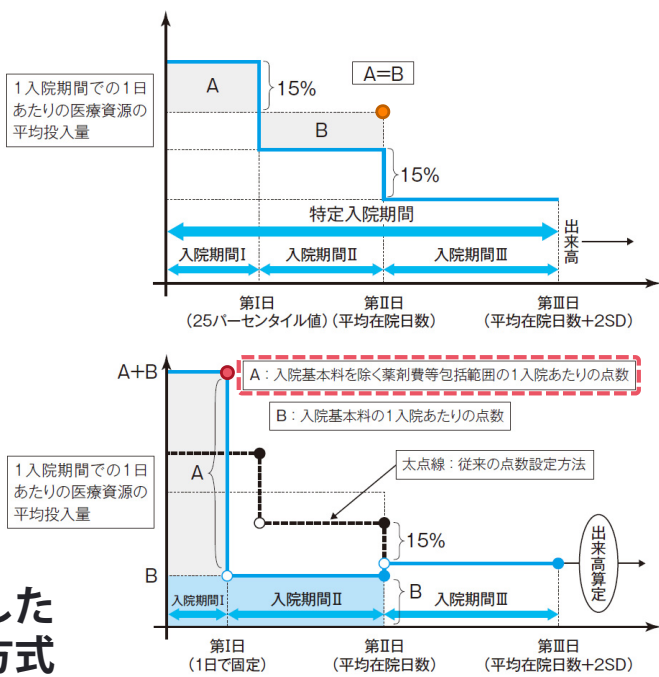


040040：Rank(02)：Regimen(001) / 退院患者数：7584(7.6%) / 施設数：564(68.1%)

棒：退院患者数(水色：1回/橙：2回/灰色：3回以上) / 青&橙&黒：累積患者数 / 緑：退院施設の割合(%) / 赤：費用(万円/入院)・点は平均値

▶ 従来

- 1日あたりの平均資源投入量を基準として、
- 入院期間IIまでに支払額を適正化



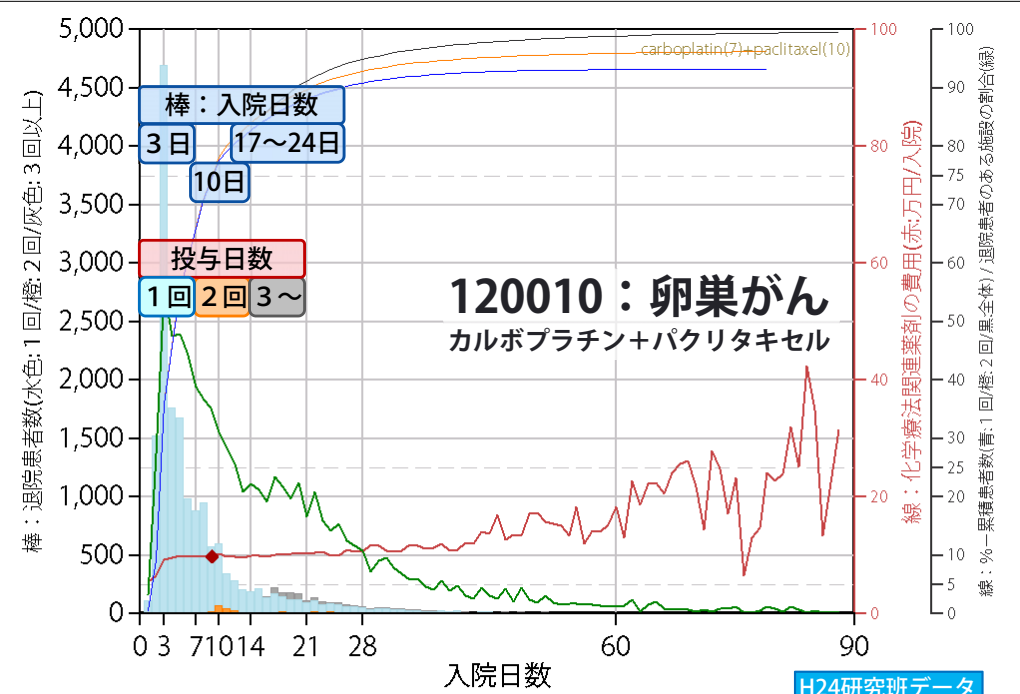
▶ 点数設定方式(D)

- 1入院あたりの資源投入(総)量(入院基本料を除く)を基準として、
- 入院初日までに支払額を適正化

高額薬剤に対応した新たな点数設定方式

kishikaw@ncc.go.jp/20131214/伏見班高知セミナー

17



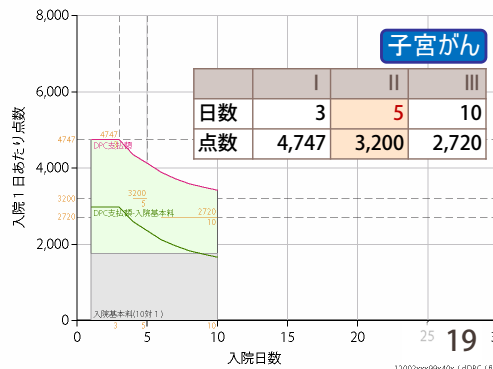
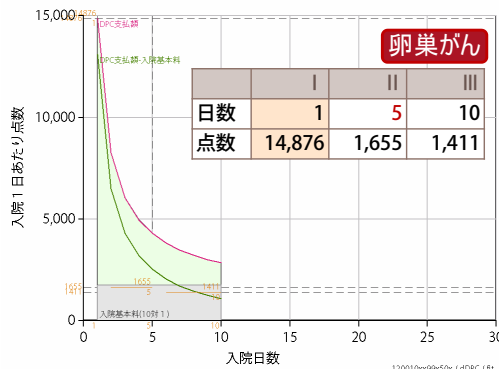
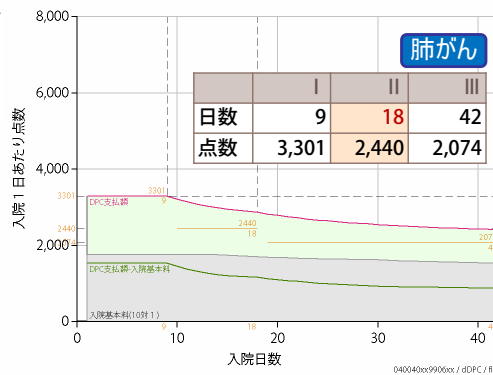
120010: Rank(01): Regimen(001) / 退院患者数: 17605(45.9%) / 施設数: 507(90.1%)

棒: 退院患者数(水色: 1回/橙色: 2回/灰色: 3回以上) / 青&橙&黒: 累積患者数 / 緑: 退院施設の割合(%) / 赤: 費用(万円/入院)-点は平均値

カルボプラチン+パクリタキセル H24DPC/PDPSでの支払

傷病名	処置等2	価格設定の方法
卵巣がん	独立(5)	入院初日に一括(設定方式D)
子宮がん	なし	従来通り(入院日数II=5)
肺がん	独立(6)	従来通り(入院日数II=18)

- ▶ 卵巣がん・子宮がんでの価格設定は異なる
- ▶ 肺がんの入院期間は長い

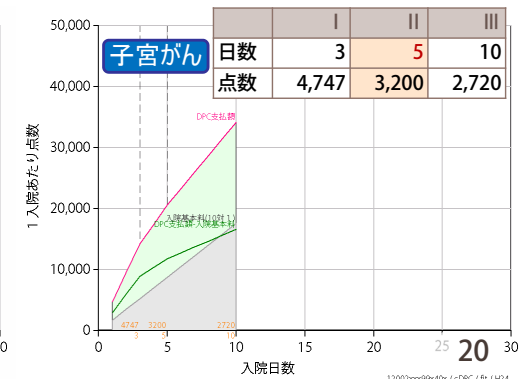
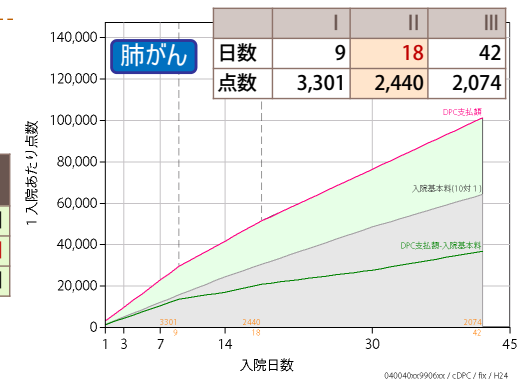
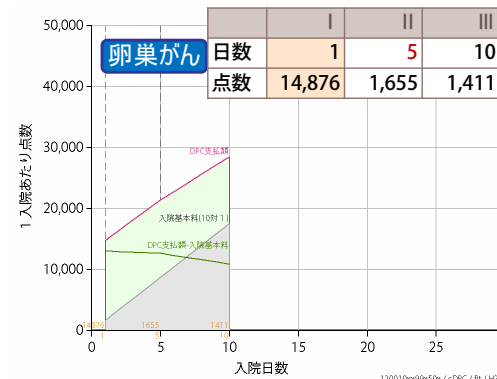


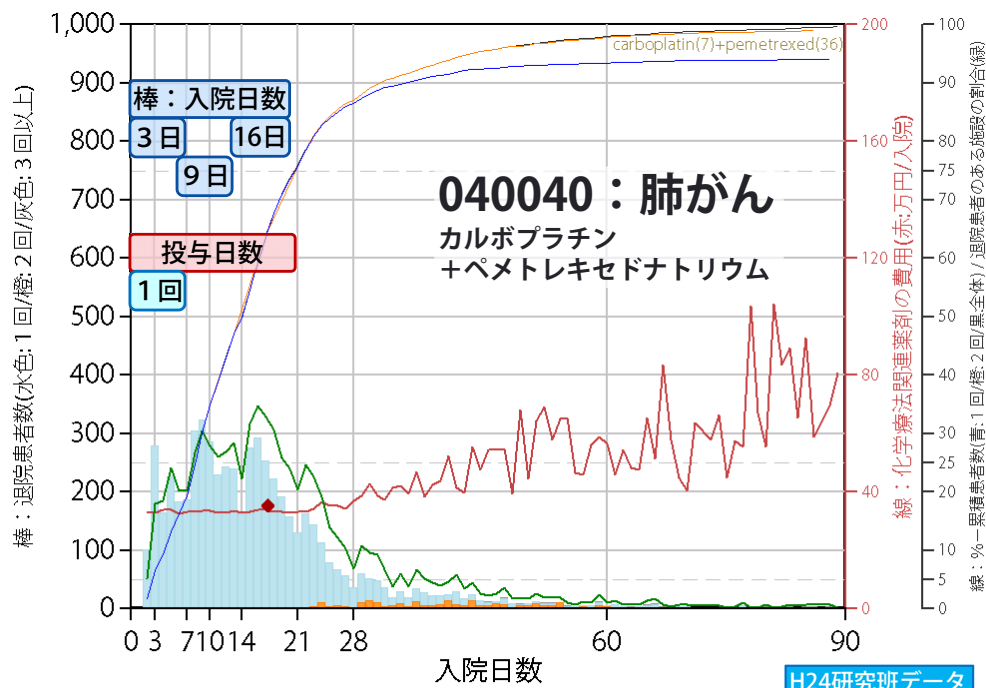
点数設定方式(D)

1入院あたりの薬剤費用と包括払い額

傷病名	処置等2	抗癌剤の平均費用	入院基本料を除く包括払い額(II)
肺がん	独立(6)	11.8万円	21.0万円 @ 18日
卵巣がん	独立(5)	9.7万円	12.7万円 @ 5日
子宮がん	なし	10.0万円	11.8万円 @ 5日

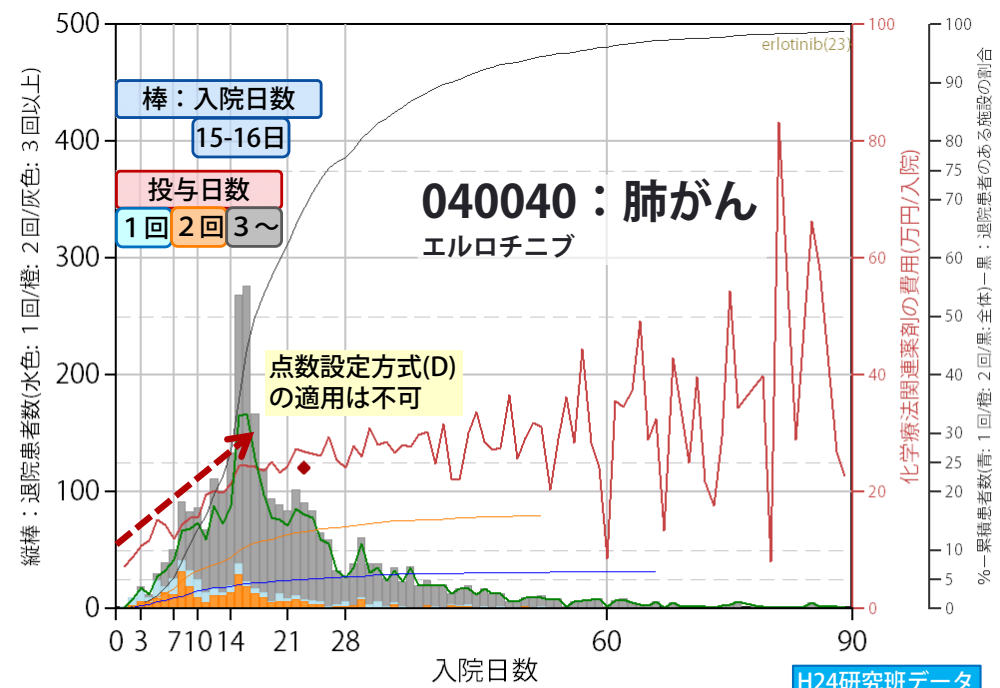
→ CCPマトリックスでは、どう考えるか





040040 : Rank(04) : Regimen(025) / 退院患者数 : 5874(5.9%) / 施設数 : 495(59.8%)

棒：退院患者数(水色: 1回/橙: 2回/灰色: 3回以上) / 青&橙&黒: 累積患者数 / 緑: 退院施設の割合(%) / 赤: 費用(万円/入院)-点は平均値



040040 : Rank(11) : Regimen(041) / 退院患者数 : 3103(3.1%) / 施設数 : 500(60.4%)

棒：退院患者数(水色: 1回/橙: 2回/灰色: 3回以上) / 青&橙&黒: 累積患者数 / 緑: 退院施設の割合(%) / 赤: 費用(万円/入院)-点は平均値

データマイニングを活用したチェックの進化

- ▶ 症例単位の IF ... THEN ... ルールによる処理
 - 根拠となるルールが明確であれば、容易に実行できる
 - ルールが明確でない場合には、現実のデータを元に、例外的な値をスクリーニングする
 - ▶ 閾値の設定方法により、確認が必要な件数をコントロールする
 - 上位 5% の症例の使用量を閾値に → 確認作業も 5% 発生
 - 5% では多すぎるので、平均値 + 2SD などの方法で “はずれ値” を区別
 - +
 - ▶ 医療機関ごとの特徴を加味した処理
 - 他の施設と比較して、処方実績は異なるのか？
 - ▶ 使用量の平均値が高い、処方されている患者の割合が高い
 - 特徴的な値を持つ施設 → 重点的な審査の対象
- 単純なルールによる確認は時と共に効果が目減りする
- 適正な診療と請求の水準を維持することが重要

行為ごとの出来高評価から施設の包括評価へ

データの見える化が重要

- ▶ 落とし穴：すべてのルールをデータマイニングで決められる？
- ▶ より失敗の少ないアプローチ：データマイニングで実態を知る
 - 問題症例に注目した分析
 - 査定・返戻症例の特徴を知る = 問題とならなかった症例との判別
 - クラス分類、決定木(Decision Tree)、ベイズ分類器、機械学習
 - ← “迷惑メール”の判別で使用されている技術
 - 診療の品質管理の一環として、標準的な診療内容を知る
 - DPC分類別にデータを集計し、CP(クリニカル/クリティカルパス)の形式でとりまとめる
 - EFファイルの分析、BI(ビジネスインテリジェンス)ツールの利用

院内での請求業務の高度化に向けて

- ▶ **使いやすいデータ** は準備できていますか？
 - 基本は “売り上げ明細書”
 - 病院 > 患者 > 入院日 > 実施日 > 診療項目 > 数量 + 査定結果
 - ▶ DPC調査の EFファイル の活用が効果的
 - 外来EFファイルを作成し、入院データとともに分析
- ▶ **包括払いにも対応した分析機能** はありますか？
 - 膨大なデータを細かく処理するための工夫
 - ▶ DPC分類別の分析：外来データにも仮想的にDPCコードを割り振る
 - ▶ 診療科・部門単位での分析
 - 細かな診療項目を分析しやすくするための工夫
 - 薬剤のレセ電算コード
 - 薬価基準コード → 薬効分類コードに読み替えて処理

包括評価が拡大することで…

- ▶ 明細書の内容確認が中心であった、出来高払い
 - ▶ 包括評価では、DPCに代表される患者分類を活用して、
 - 診断と治療内容の組み合わせに注目した、症例単位での管理
 - ▶ 正しいDPCコードに分類されているか？ → コーディング
 - ▶ 入院日数は長くないか？ → 効率性
 - 医療機関の機能に注目した管理
 - ▶ 対応できる 傷病・治療 の範囲は？ → カバー率
 - ▶ 入院日数が長く・1日あたり点数の高い症例は？ → 複雑性
 - ▶ 地域医療 / 救急医療 への貢献は？
- ↓
- ▶ 単なる請求事務から、病院マネジメントへの深化が求められている

出来高払いとは別に、
機能評価係数IIによる
評価が拡大していく

データに基づいた改革を進める必要がある