

厚生労働科学研究費補助金

医療技術評価総合研究事業

電子カルテシステムにより集積したデータの
診療情報解析（データマイニング）に関する研究

平成15年度 総括研究報告書

主任研究者 秋山 昌範

平成16年3月

目 次

I. 総括研究報告書

- 電子カルテシステムにより集積したデータの診療情報解析（データマイニング）に関する研究…………… 1
（資料）国立国際医療センターにおける電子カルテ管理会計
情報に関するデータマイニング報告書

秋山 昌範

II. 分担研究報告書

- 肝疾患ネットワークにおける知識発見型大規模診療自動解析（データマイニング）システムの臨床応用に関する研究…………… 45

八橋 弘

- 腫瘍外科学における知識発見型大規模診療情報自動解析システムの
応用に関する研究…………… 50

大内 憲明

- 国立病院における病院業務分析に関する検討 ―電子カルテシステム
を利用した IgA 腎症に関するデータマイニングの試み―…………… 53
（資料）国立病院総合情報ネットワークシステムによる IgA
腎症に関するデータマイニングの試み

吉村 光弘

- 中核病院における病院業務分析に関する検討…………… 84

鈴木 明彦

- III. 研究班会議発表資料…………… 86

- IV. 研究成果の刊行に関する一覧表…………… 108

- V. 研究成果の刊行物・別刷…………… 113

厚生労働科学研究費補助金（医療技術評価総合研究事業）

総合研究報告書

電子カルテシステムにより集積したデータの診療情報解析（データマイニング）に関する研究

主任研究者：秋山 昌範

研究趣旨

21世紀型医療開拓推進研究事業の対象疾病であるがん、心臓病、脳卒中、痴呆及び骨折のみならず、各政策医療ネットワーク対象疾患を横断的に包括してその対象とし、より効果的かつ効率的な予防、診断、治療、リハビリテーション等を確立するために、根拠に基づく医療（Evidence-based Medicine）の基盤となる日本人の特性に配慮した根拠（Evidence）を、網羅的かつ自動的に即時提供できる知識発見型大規模診療情報自動解析（データマイニング）システムを開発する。乱数発生させた各種疾患患者模擬データ群を作成してデータウェアハウス（通常のデータベースとは異なり、サブジェクト中心型に統合・最適化された、時間変化する不揮発性データの集合体）を構築し、これを本研究成果物で解析する。基礎検証後に、インフォームドコンセント取得済み登録患者の診療情報自動集積システムを構築しつつある HIV 診療支援ネットワーク（A-net）・肝疾患ネットワーク（L-net）・今後構築予定の内分泌代謝性疾患ネットワーク（EM-net）、がんネットなどと協力し、これらのネットワークで自動集積された膨大かつ詳細な実際の臨床データを本研究成果物で解析することによって、1) 従来からの医学的常識を自動解析結果で裏付けることで本研究成果の妥当性を検証すると共に、2) 本来の目的である「未知の知識の発見を示唆」出来るかどうかで有用性を検証する。検証終了後は、国立国際医療センターで開発中の標準化仕様電子カルテシステムとの接続を試み、同電子カルテシステムの後方支援システムとしての機能付与を併せて行う。

分担研究者

- 岡 慎一（国立国際医療センターエイズ医療研究開発センター 臨床研究開発部長）
島津 章（国立京都病院臨床研究部／内分泌・代謝性疾患センター 臨床研究部長）
八橋 弘（国立病院長崎医療センター臨床研究部／消化器科 臨床研究部長）
大内 憲明（東北大学大学院医学系研究科病態学講座腫瘍外科学分野
腫瘍外科、乳線・内分泌外科、癌の分子生物学、癌の疫学 教授）
吉村 光弘（国立金沢病院第二内科医長）
鈴木 明彦（盛岡赤十字病院検査部長）

A. 研究目的

医療における IT 革命は、コンピュータ性能と記憶媒体の飛躍的増大とアクセス速度の飛躍的

向上、データベース管理技術の飛躍的向上といった技術的背景によるオーダーエントリーシステムの普及によって急速に到来した。膨大な量

の診療情報が蓄積され始め、その有効活用による、より効果的かつ効率的な予防、診断、治療、リハビリテーション等の確立が切望されている。しかしながら、1980年代後半の流通業界におけるIT革命と同様に、蓄積される情報量に対して統計・解析が追いつかず、未解析情報(knowledge gap)が急速に増大するという、新たな課題が浮き彫りとなっているのが現状である。従来より医学研究においては、臨床データをリアルタイムで収集して検索し、分析する「データ収集型」ないし、予め立てた仮説に対して、臨床データを用いて検証する「仮説検証型」統計解析が用いられて来た。これらの手法では、膨大なデータの膨大な理論的組合せの中から、ごく一部を仮説として設定し検証するために、①結果の質(重要度)は解析者の洞察能力などに大きく依存し、②従って、常識の範囲を越えるような相関ルールの発見は難しく、③また、全ての仮説を網羅的に検証することは到底困難であるがゆえに、knowledge gapを増大させるに至った実状がある。本研究は、このknowledge gapを埋めるべく、膨大な倫理的組合せに対して仮説を立てることなく網羅的に解析することで、有用な知識を自動生成・発見する「知識生成型」解析システムの開発を目標とする。従って、本研究は「メディカル・フロンティア戦略」の中核的研究として位置付けられ、本研究成果物により、21世紀医療開拓推進研究事業の対象疾患であるがん、心臓病、脳卒中、痴呆及び骨折のみならず、各政策医療ネットワーク対象疾患を横断的に包括してその対象とし、より効果的かつ効率的な予防、診断、治療、リハビリテーション等を確立するための根拠に基づく医療(Evidence-based Medicine)の基盤となる日本人の特性に配慮した根拠(Evidence)を、網羅的かつ自動的に即時提供することが期待できる。

さらに、提供可能なEvidenceは、純粋な医学的見地における統計解析のみならず、診療に際して得られた診療報酬情報や単品管理などの物流情報も包括することで、稼働済みのクリティカルパスの妥当性検証ツールやDRG/PPS導入などにおける基礎検証ツールとしての応用も可能であり、医療経済改革に対しても大きく貢献することが期待できる。

B. 研究方法

本研究において既に終了している、流通業界や金融業界で既に実績のあるデータマイニングアプリケーションのアルゴリズムを応用・改良し、これを本研究のデータマイニングエンジンとすべき基礎研究による手法を用いてA-netのデータを解析した。これをデータマイニングの基本的アルゴリズム複合体で、機械学習技術の一つである決定木導出法(decision tree induction methods: 参照1)および数値技術のひとつである回転対象基底関数法(RBF法: radial basis function methods: 参照2)によって解析し、その妥当性を検証した。また、これまでの成果であるA-netに登録されたHIV感染患者における治療データに対する決定木導出法、コホーネン・クラスタリング法(参照3)を用いた解析と服用効果分析に加えて、L-netに登録された肝疾患における治療データにおいても決定木導出法、コホーネン・クラスタリング法などを持ちいて解析を行った。

(参照1)決定木導出法(decision tree induction methods): ノード(レコードの属性のテスト)とリーフ(レコードが収集される終点)を結びつけた、クラス判別モデルを作成するためのデータマイニング手法。

(参照2)回転対象基底関数法(RBF法: radial basis function methods): より複雑な関数の近

似を得るために、ある特定点からの距離を表す基礎関数を用いて値を予測する。

(参照 3) コホーネン・クラスタリング法：雑多に分布した要素の集合の中の偏りを検出し、指定した制約条件に合うように要素を複数のグループに分類する方法。

(倫理面への配慮)

①本研究成果産物の開発時の妥当性評価には乱数発生模擬患者データ群を使用する。②本研究での実診療情報自動解析の対象は、インフォームドコンセント取得済み登録患者の診療情報自動集積システムを構築中の A-net・L-net と宮城県対がん協会の集団検診データのみとする。③他診療情報、管理会計データからのマイニングに際しては、原始データを個人特定不可能な状態に加工する。

以上のように対象を限定したうえで、上記を遵守することにより倫理的配慮とする。

C. 研究結果

まず、本研究に最適化した診療情報データウェアハウスを検討し構築した。診療情報データウェアハウスへのデータ入力は即時性・正確性などを考慮して全自動を基本とする。その仕組みについては、A-netやL-netなどと協力し、既に両政策医療ネットワークで構築中のインフォームドコンセント取得済み登録患者の診療情報自動集積システムを応用し、自動集積されたXML形式診療情報を本データウェアハウスへ自動転送されるように設計した。これまでの研究で、A-netにおける抗HIV薬投与中AIDS患者の体重減少の予測因子として、血小板増加が有意な因子であることを発見した。昨年度の研究では肝疾患の予後解析モデル作成により、PIVKAが有効な予後判定因子であることを発見した。その詳細は、肝疾患の診断で最も重要な診断項

目である治療効果を含む予後予測診断であり、その正診率が情報解析精度と言い替えてもよく、この分野での解析モデルを策定した。1990年から2002年の期間、国立長崎中央病院で死亡した肝疾患患者456名(男性325名、女性131名、平均年齢64歳の25-92歳)を対象とした。うち死亡時診断は肝癌346名、肝硬変慢性肝不全59名、急性肝不全14名、その他37名であった。患者情報と血液検査所見(A1b,ALT,LDH,CHO,PIVKAなど35項目、一項目あたり約25000件)の情報をIBM Intelligent Minerを用いて分析し検査時点から1年生存できるか否かの1年余命モデルを決定木(decision tree)手法で求めた。その結果、検査時点で(PIVKA>8255mAU/ml)の条件を満たすと93.9%、(1034<PIVKA<8255)(AFP>1215ng/ml)の2条件では91.7%の確立で1年以内に死亡するのに対し、(PIVKA<1034)(<CHO>102mg/dl)(AFP<531.5)の3つの条件を満たすと85.5%の確立で1年以上生存するモデルが作成された。即ち、肝癌腫瘍マーカーの絶対値と肝予備機能が肝疾患患者の生存期間に寄与することが確認された。Decision tree以外に回転対象基底関数やニューラルネットワークを用いての解析では、随時の検査値を用いて、後何年生存できるかの予後予測モデルの作成も可能であることが確認された。

今年度は国立国際医療センターにおいて、管理会計データを使用し、病院経営の改善に結びつく因子を抽出することを目的とし、データマイニング解析を試行した。対象は、2003年4月から2003年9月までの6か月間に国立国際医療センターに入退院した3,215件(実患者数2,623名)であり、このうち不正確なデータを排除した1,587件(実患者1,433名)について解析を行った。今回の解析では、損益分岐点を境に、「黒字グループと似た特徴を持つ赤字グループ」に着目

し、この違いを明確にすることによって、改善に結びつく施策を導くことを目標とした。全件データを対象とした決定木分析では、救急であるか否かにより損益が分かれるが、救急の場合でも黒字化が可能なグループが存在し、病棟経費・注射・検体検査の原価回収率を向上させることで改善しうるとの結果を得た。また、診療科別の決定木分析では、救急部、小児科、呼吸器科において、特徴的な所見を見いだした。さらに RBF 解析により、全件データを原価回収率別に 8 グループに分け、それぞれのグループ間の比較検討を行った。その結果、赤字グループでは救急患者が対象となっていることが特徴であった。小児患者では、赤字になりやすい傾向があることが判明した。小児・救急のグループ以外では、手術の有無が関与しており、手術がない場合は原価回収率が高いことがわかった。

D. 考察

医療情報を対象としたデータマイニングシステム開発を目的とした本研究は、その対象規模を考慮すると国内・国外に類や比較対象が存在せず、世界初の試みである。(流通業界においては、1980 年代後半に本研究で応用したデータマイニング基本手法が実用化され、1999 年の「Fortune」誌ランキングで上位 500 社のうちの 80% がデータマイニングのパイロットプロジェクトを進行中か、すでに 1 つ以上のデータマイニングを配備し活用している。) 本研究で開発したデータマイニングシステムは、従来の医学研究における「データ収集型」ないしは、「仮説検証型」統計解析と比較して、以下の特色・独創的な点を持つ。

①知識発見型：すなわち仮説の設定不要で意外な発見の可能性がある。②意外性の発見：予期範囲を越えた知識の検出の可能性がある。③完

全制：全ての組合せに対して網羅的に検証可能である。④自動的：自動的に道の知識を検出可能である。⑤高速性：項目数に制限はなく、巨大データを処理可能である。⑥網羅性：総当たりに全件データを処理可能である。⑦データ・タイプ：記号と数値の混在を容易に許容可能である。⑧非線形モデル：多変量解析の限界を打破し、厳密な分析が可能である。⑨定式化が不要：学習により関数/モデルを自動生成可能である。⑩容易性：統計解析の取得は不要で誰でも使える。⑪具体性：結果が具体的で分かり易い。⑫即応性：対話的に迅速出力可能で、分析効率の向上が期待できる。⑬拡張性：拡張可能な並列処理(準線形)である。以上の特色・独創的な点により、本研究は、医療情報における knowledge gap を埋め、本研究事業の目標である、より効果的かつ効率的な予防、診断、治療、リハビリテーション等を確立するための根拠に基づく医療(Evidence-based Medicine)の基盤となる日本人の特性に配慮した根拠(Evidence)を、網羅的かつ自動的に即時提供することが期待でき、純粋な医療的見地における統計解析のみならず、医療経済改革に対しても大きく貢献することが期待できる。

今年度までの解析によって HIV 感染症疾患のみならず肝疾患にも有用性が証明されたことから、本解析手法は医学一般に応用できると期待され、今後は他分野における応用が期待できる。従って、次年度以降はがんネット等と協力し、がん検診等の大規模なデータベースのでの分析の検討を加えたい。

E. 結論

本研究は、医療情報における knowledge gap を埋め、本研究事業の目標である、より効果的かつ高率的な予防、診断、治療、リハビリテー

ション等を確立するための根拠に基づく医療 (Evidence-based Medicine) の基盤となる日本人の特性に配慮した根拠 (Evidence) を、網羅的かつ自動的に即時提供することが出来、純粋な医学的見地における統計解析のみならず、医療経済改革に対しても大きく貢献することが期待できるシステムである。

昨年度までの解析によって HIV 感染症疾患のみならず肝疾患にも有用性が証明されたことから、本解析手法は医学一般に応用できると期待された。また、本年度の研究を通し、医療の経営分析、経営改善にも応用が可能であることが示唆された。今後はさらに他分野における応用が期待できる。従って、次年度以降はがんネット等と協力し、がん検診等の大規模なデータベースの分析の検討を加えたい。

F. 研究発表

1. 論文発表

1. 秋山昌範. 病院管理を行うための ERP (Enterprise Resource Planning) システム. 医療情報学 23,3-13.2003.
2. 秋山昌範, 斎藤澄. 遠隔病理診断におけるデジタルマイクロスコープの有用性. 遠隔医療研究会論文集 7,78-79.2003.
3. 秋山昌範. 電子タグのネットワーク利活用に関する検討. 医療情報学 23(Suppl.),103-106.2003.
4. 秋山昌範. 薬事法改正に対応した医療材料・医薬品のトラッキング. 医療情報学 23(Suppl.),317-319,2003.

2. 学会発表

第 23 回 医療情報学連合大会
HOSPITALOG ASIA 2003. (Bangkok, Thailand)

Inaugural Symposium of the Seoul National University Bundang Hospital. (Seoul, Korea) など、国際学会等 2 件、国内学会 23 件

G. 知的所有権の取得状況

1. 特許取得

特願 2003-118496 疾病予後モデルの作成方法、このモデルを用いた疾患予後予測方法、このモデルによる予後予測装置、ならびにプログラム、記憶媒体。

2. 実用新案登録

なし

3. その他

なし

国立国際医療センター

厚生労働科学研究費補助金「医療技術評価総合研究事業」
電子カルテシステム管理会計情報に関するデータマイニング解析
報告書

平成16年3月
情報システム部

秋山 昌範



はじめに

管理会計データを使用し、病院経営の改善に結びつく因子を抽出することを目的とし、データマイニング解析を試行しましたのでご報告いたします。

【ご報告内容】

- ①本解析内容について
 - ②管理会計データマイニング概要
 - ③データの前処理
 - ④分析アプローチ
 - ⑤マイニング結果
- 補足資料(添付)



目次

- 1. 本解析内容について
- 2. 管理会計システムデータマイニング概要
- 3. データの前処理
 - ▶ 原始データ一覧
 - ▶ マイニングデータ作成手順 (2003年4月-9月分)
 - ▶ データ前処理の内訳
 - ▶ マイニングデータ
- 4. 分析アプローチ
 - ▶ 分析の着眼点
 - ▶ 分析のステップ
- 5. マイニング結果
 - ▶ 1. 全件データによる分析(決定木手法)
 - ▶ 2. 診療科別分析(決定木手法)
 - ▶ 3. 全件データによる分析試行(RBF手法)

1. 本解析内容について

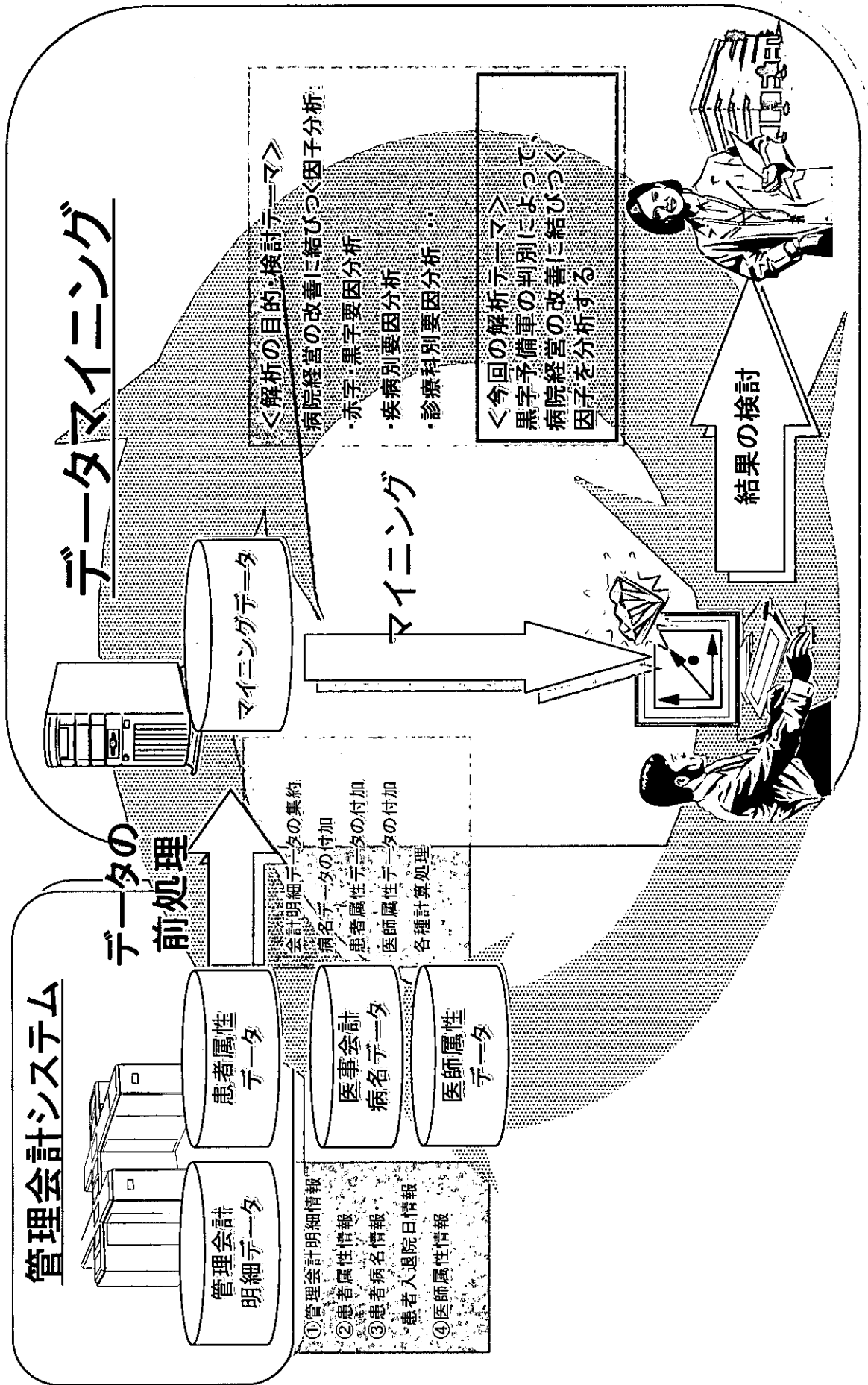
□ 目的:

国立国際医療センターにおける管理会計システムに蓄積された管理会計明細データを活用してデータマイニング解析を行い、病院経営の改善に結びつく因子の抽出を試行する。

□ 使用したデータ:

- ▶ ①管理会計明細データ（2003年4月～2003年9月 の6ヶ月分）
- ▶ ②患者属性データ（管理会計システム）
- ▶ ③患者病名データおよび入院日データ（医事会計システム）
- ▶ ④医師属性データ（手動作成データ）

2. 管理会計データマイニング概要



3. データの前処理

原始データ一覧

マイニング用の原始データとして以下を使用しています。

① 管理会計システム：管理会計明細データ

番号	内容	必要：○ 不要：×
1	病院ID	○
2	年月日	○
3	オーダID	○
4	ACTID	○
5	患者ID	○
6	入院日	○
7	オーダ種	○
8	収益/原価区分	○
9	構成要素	○
10	単価	○
11	単価単位	○
12	計算数値	○
13	金額	○
14	入外区分	○
15	管理会計部門コード	○
16	部門区分	○
17	診療区分	○
18	診療区分名称	○
19	管理会計診療科コード	○
20	管理会計治療コード	○
21	管理会計科目コード	○
22	診療科コード	○
23	診療科名称	○
24	病棟コード	○
25	病棟名称	○
26	オーダ依頼科コード	○
27	オーダ依頼科名称	○
28	オーダ依頼者コード	○
29	オーダ依頼者名	○
30	実施科コード	○
31	実施科名称	○
32	実施者コード	○
33	実施者名称	○
34	実施端末名	○
35	主治医コード	○
36	主治医名	○
37	患者名	○
38	患者名(カナ)	○
39	患者名(ミドルネーム)	○
40	生年月日	○
41	性別	○
42	緊急病名コード	○
43	緊急病名名称	○
44	ICD10コード	○
45	定数配置使用区分	○
46	患者別薬物対象区分	○
47	計算対象区分	○
48	単価換算率	○
49	数量計算方法	○
50	単価換算率1	○
51	単価換算率2	○
52	単価換算率3	○
53	数量1	○
54	数量2	○
55	数量3	○
56	数量4	○
57	数量5	○
58	時間1	○
59	時間2	○
60	時間3	○
61	時間4	○
62	フェイス情報	×

② 管理会計システム：患者属性データ

番号	内容	必要：○ 不要：×
1	病院ID	○
2	患者ID	○
3	患者名(性)	○
4	患者名(名)	○
5	患者名(姓・カナ)	○
6	患者名(名・カナ)	○
7	患者ミドルネーム	○
8	患者ミドルネーム(カナ)	○
9	性別	○
10	生年月日	○
11	入院管理番号	○
12	入院診療科コード	○
13	入院診療科略称	○
14	入院診療コード	○
15	入院診療略称	○
16	入院病室コード	○
17	入院病室名称	○
18	入院ベッド有無	○
19	入院ベッドコード	○
20	入院病室コメント	○
21	入院チームコード	○
22	主治医コード	○
23	主治医名	○
24	入院予定日	○
25	入院日	○
26	入院時刻	○
27	移動先診療科コード	○
28	移動先診療科略称	○
29	移動先診療コード	○
30	移動先診療室コード	○
31	移動先診療室名称	○
32	移動先ベッドコード	○
33	確保ベッド有無	○
34	看護床コード	○
35	看護床名称	○
36	看護区分コード	○
37	看護区分名称	○
38	看護区分コード	○
39	看護区分名称	○
40	搬送区分コード	○
41	搬送区分名称	○
42	フェイス情報	×

③ 医事会計システム：患者病名データおよび入院日データ

番号	内容	必要：○ 不要：×
1	患者番号	○
2	氏名	○
3	姓	○
4	名	○
5	入外区分	○
6	主病名	○
7	開始日	○
8	終了日	○
9	病名コード	○
10	病名	○
11	入院日	○
12	退院日	○
13	科コード	○

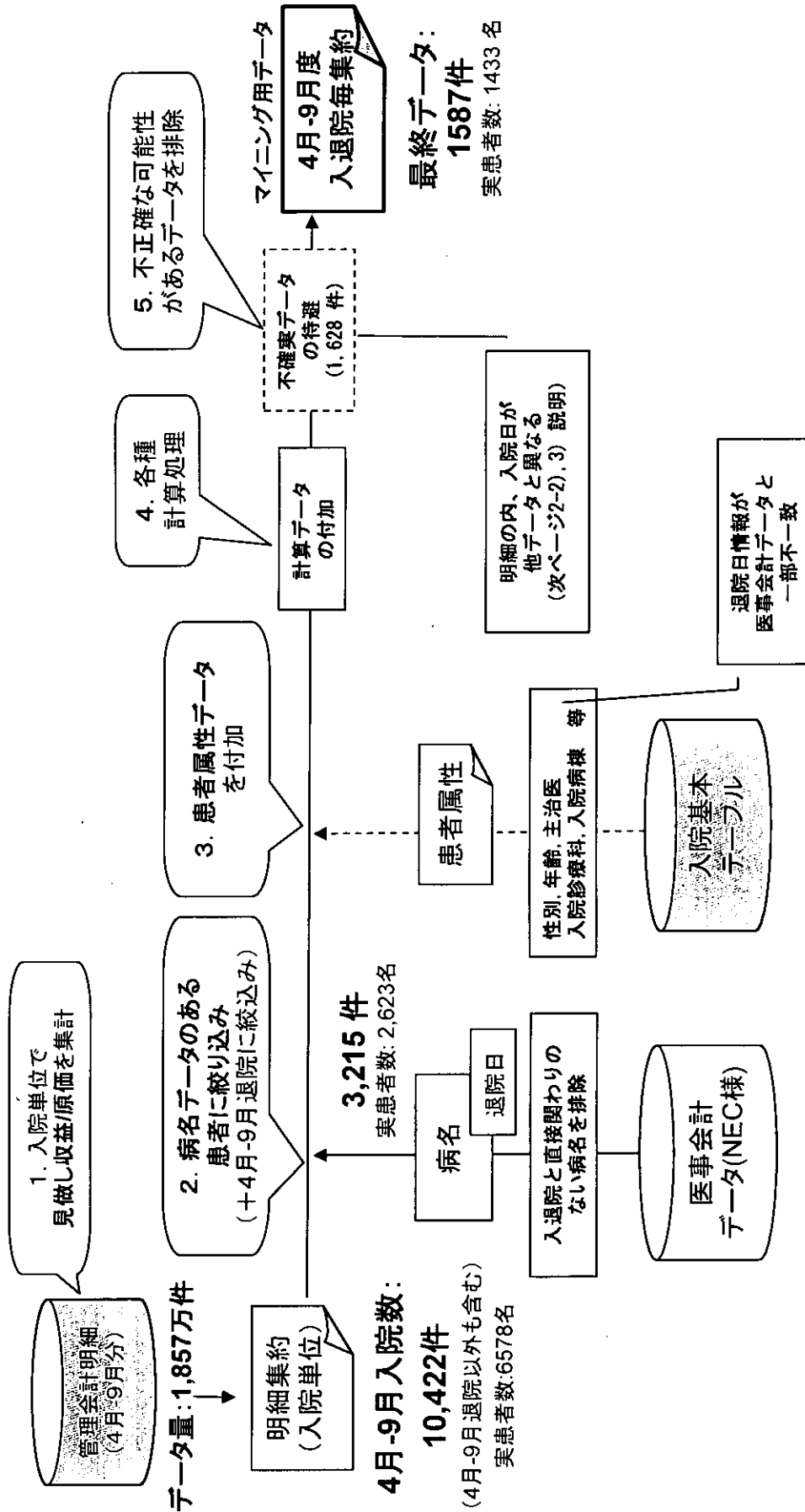
④ (手動作成) 医師属性データ

1	医師コード	○
2	医師氏名	○
3	医師氏名(カナ)	○
4	医師診療科コード	○
5	医師単価	○
6	医師ワークグループ	○

※ ×印は主に個人情報のためデータより削除しております。

マイニングデータ作成手順 (2003年4月-9月分)

2003年4月～9月のデータを抽出した際の、マイニングデータ作成手順は以下のとおりです。



4月-9月入院: 7,586件 (実患者数: 6257名)

4月-9月入退院: 3,579件 (実患者数: 2785名)

データ前処理の内訳

解析データの正確性を期すため、マイニングの前処理として以下のデータ洗浄作業を実施しています。

1. データの抽出

管理会計システムからは以下の条件に基づきデータの抽出を行っております。

- ・入院データ
 - 入院毎の集約とするため、外来、および入院中他科外来は除きます。
 - 本試行では4月-9月の入院/退院患者のデータとなります。
- ・在院日数が計算可能なデータ
- ・病名が判別可能なデータ
- ・患者が判別可能なデータ
 - 患者IDが入力されていないデータ(物流請求データ等)は除きます。

2. 不確実なデータの排除

以下のケースではデータが正確ではない可能性があるため、取り除きます。

1) 病名データ

- 入院日以前に終了日が登録されている病名、および退院日後に開始日が登録されている病名のデータは、当該入院とは直接関連しない病名として、除きます。(364入院分)

2) 入院基本テーブルと入院日が異なるデータ

- 管理会計データの内、医事会計システムより取得しているデータ(手術/処置/麻酔)の中に、入院日が異なっていると思われるデータが見られます。該当患者に関しては他のデータも不確実と考えられるため、除きます。(940入院分)

3) 入院日と原価計上日に矛盾があるデータ

- 管理会計データの内、入院日以前に原価計上されているようなデータが見られます。一部はシステムトラブルの影響で入院日が間違っており入力されたものと判明していますが、残りは各元データによって状況が異なり、正否の判断、また正しいデータに置き換えるのが困難と考えられます。2)と同じく、該当患者のデータを除きます。(1,157入院分)

マイニングデータ

※データの詳細は補足資料をご参照ください。

入退院毎の見做し収益/原価をオーダー種毎に集計します。

Key 患者属性

患者ID+入院日	在院日数、性別、年齢、等	病名(5件まで)	主治医	入院診療科、入院病棟
----------	--------------	----------	-----	------------

金額集計 (入退院あたりの各原価/見做し収益 構成要素)

・・・ 各原価 構成要素の合計(処方-薬品、処方-固定費、・・・計22項目) 原価合計

・・・ 各見做し収益 構成要素の合計(処方、注射、検体検査、・・・計11項目) 見做し収益合計

見做し利益

原価回収率

赤字/黒字区分

各原価/見做し収益 構成要素 の1日あたりの額

各原価/見做し収益 構成要素 の全体に占める割合 (%)

各原価/見做し収益 構成要素 の原価回収率 (%)

4. 分析アプローチ

分析の着眼点

「黒字予備軍」の発見

- ・赤字を黒字転換させるために、「少しの努力で黒字になりやすい赤字グループ」を探します。
- 「黒字グループと似た特徴を持つ」赤字グループに着目
- ・そのグループと黒字グループを比較し、違いを明確にすることで、改善に結びつく施策を実施し、黒字転換をめざします。

