

ポスター

ポスター8

医療データ分析4（データ・テキストマイニング）

2018年11月23日(金) 17:00 ～ 18:00 K会場(ポスター、HyperDemo) (2F 多目的ホール)

[2-K-4-4] 治療パターンと転倒インシデントの相関ルールデータマイニング

○大野 実^{1,2}, 島田 陽介^{1,2}, 佐藤 雅也^{1,3}, 川崎 志保理⁴, 堀 賢^{1,5} (1.順天堂大学大学院 医学研究科 電子医療情報管理学講座, 2.セコム株式会社 IS研究所, 3.セコム医療システム株式会社, 4.順天堂大学大学院 医学研究科 心臓血管外科学教室, 5.順天堂大学大学院 医学研究科 感染制御科学教室)

[目的]病院情報システム内には、入院患者の日々の治療行為や検査結果、転床歴などのデータが時系列に蓄積されている。こうした時系列データの出現パターンを解析することで、病棟での転倒患者に有意に多く見られる治療パターンを抽出し、転倒発生に関する知見を得ることを目的とする。[方法]順天堂医院において2012年4月～2016年3月の期間に転倒した316名の患者と、性別、年齢区分(5歳刻み)、資源病名、手術の有無等の条件で一致した非転倒患者316名の632名の患者群を対象とした。対象患者について、転倒日の直前10日間の治療行為、検査結果、転床歴をイベントアイテムとした出現パターンを時系列データとして作成した。非転倒患者については、対応する転倒患者を参照した適当な疑似転倒日を定義した。これらの時系列データに対して、統計ソフト「R」の「arulesSequences」パッケージを用いて、相関ルールのデータマイニングを実施した。[結果]支持度0.063、確信度0.62を下限とし、相関ルールの右辺に転倒イベントを含むルールは、30個のルールが抽出された。順天堂医院に2016年4月～6月の期間中入院した1695名の患者(23名の転倒患者を含む)のデータを検証用データとして、これら30個のルールをフィッシャーの正確性検定で検定したところ、精神神経用剤を含むパターンの出現頻度に有意差が見られた。[結論]精神神経用剤の使用が、転倒発生に関連性があることは示唆されたが、他の治療行為や検査結果との前後関係から、生物学的妥当性のある説明を導くことはできなかった。本手法は、治療行為の手順に起因するような要素の発見には適しているが、転倒インシデントはベッドやトイレ等の周辺環境状況に起因するような要素が支配的であったために、限定的な分析となってしまったと考える。

治療パターンと転倒インシデントの相関ルールデータマイニング

大野 実^{*1*2}、島田 陽介^{*1*2}、佐藤 雅也^{*1*3}、川崎 志保理^{*4}、堀 賢^{*1}

^{*1} 順天堂大学大学院 医学研究科 電子医療情報管理学講座、

^{*2} セコム株式会社 IS 研究所、^{*3} セコム医療システム株式会社

^{*4} 順天堂大学医学部附属順天堂医院 医療安全推進部 医療安全管理室

Data mining association rules between treatment patterns and falling incidents

Minoru Ohno^{*1*2}, Yosuke Shimada^{*1*2}, Masaya Satoh^{*1*3}, Shiori Kawasaki^{*4}, Satoshi Hori^{*1}

^{*1} Department of Medical Intelligence Management, Juntendo University, Graduate School of Medicine,

^{*2} Intelligent Systems Laboratory, SECOM Co., Ltd., ^{*3} SECOM Medical System Co., Ltd.,

^{*4} Patient Safety Unit, Department of Patient Safety Promotion, Juntendo University Hospital,

Aim: To extract significantly common treatment patterns that were observed in falling patients in wards, and to obtain knowledge about falling incidents by analyzing massive time-series data which had been accumulated in the hospital information system.

Methods: In total, 632 patients were recruited in this study. Patients were over 16 years old and had been hospitalized in Juntendo hospital from April 2012 to March 2016; 316 were falling patients, the others were non-falling patients and selected based on the combination of gender, age group, main disease, length of stay and so on. The association rules between their treatment patterns and falling incident were extracted from the time-series data of 10 days just before the day of the fall, by using statistical software “R” with “arulesSequences” package.

Findings: With a minimum support of 0.063 and a minimum confidence of 0.62, 30 rules which have falling incident as their right-hand-side were extracted. These 30 rules were tested by Fisher's exact test using the other patient's data-set, and significant differences were only found in the rules which include treatment pattern with psychoneurotic agents.

Conclusion: Although it was suggested that the use of psychoneurotic drugs was related to the occurrence of falls, it was impossible to derive a biological plausibility from the context of other treatment behaviors and examination results

Keywords: DPC data, data mining, association rule, time-series data

1. はじめに

我が国における電子カルテの導入率は年々増加しており、JAHIS(一般社団法人保健医療福祉情報システム工業会)の調査では、2017年時点で400床以上の大規模病院での電子カルテ導入率は75%を超えている。¹⁾ また、診療報酬包括支払制度であるDPC/PDPSの普及も進み、急性期一般入院基本料等に該当する病床の約83%を占めるまでになった。²⁾

このように大規模病院を中心として構造化・標準化された診療データが時間情報とともに日々蓄積される環境が整いつつある現在、データマイニングや機械学習の手法を用いて臨床現場での有意義な知見を見出していくことが期待されている。

一方で、入院患者の転倒インシデントは病院管理上の大きな課題であり、これまでも様々な転倒リスクアセスメントツールを用いてリスク分析が行われてきた。³⁾ しかしながら、転倒の時間帯や場所に着目した分析が主であり、患者に提供された治療行為の時系列データを用いた分析は限定的であった。

そこで本研究では、病院情報システム内に日々蓄積されている治療行為や検査結果、転床歴などの時系列データを解析し、病棟での転倒患者に高頻度でみられる治療パターンを抽出することで、転倒発生に関する知見を得ることを目的とした。

2. 方法

順天堂医院の入院患者のうち、転倒インシデントのあった患者(転倒事例)と患者属性の類似している転倒インシデントのなかった患者(非転倒事例)の二群を選定した。それぞれの

患者の転倒日前10日間の治療行為や検査値、転床歴を時系列データとして表現した。この際、非転倒事例の転倒日はDPCの入院期間IIを基準に疑似転倒日を設定することとした。これらの時系列データ(以下治療パターンと呼ぶ)を統計ソフト「R」の「arulesSequences」パッケージを用いて相関ルールをデータマイニングし、転倒事例に多く見られる頻出治療パターンを抽出した。抽出した各頻出治療パターンについて、ルール抽出に用いた患者セットとは別の患者セットを用いて出現率のノンパラメトリック検定を行った。

2.1 データ

順天堂医院に2012年4月～2016年3月の期間に入院していた患者を相関ルール抽出のための患者セットの対象とした。

このうち医療安全管理室で管理する転倒事案リストに含まれる1,653名の患者の中で、年齢が16歳以上かつ転倒が入院日より10日以上経過しており、DPCデータとリンク可能な660名を転倒事例として抽出した。なお、一入院期間に複数回転倒している患者は、初回転倒のみを対象とした。

患者への治療行為はDPCデータのEFファイルから、血液検査値は検査部門システムから、転床歴は電子カルテの移動歴からそれぞれ抽出した。

2.2 事例選定

上記660名の転倒患者に対して、以下の条件で一致する非転倒患者を1:1で割り当てた。

- 性別
- 年齢区分(5歳刻み)

- c) DPC6 桁分類(xxxxxx を含まない)
- d) 入院期間が各々の DPC 分類の入院期間 II より長いのか短い
- e) 手術の有無
- f) 入院時 ADL(歩行, 階段)
- g) 非転倒患者に疑似転倒日を設定可能(詳細は次節で述べる)

これにより, 転倒事例, 非転倒事例それぞれ 316 例ずつの患者ペアをマッチングした。

2.2.1 非転倒事例の疑似転倒日の設定方法

非転倒事例の治療パターンを抽出するにあたり, 以下の方法で疑似転倒日を設定した。

- 1) 非転倒事例と対になる転倒事例において, 入院期間 II を起点とした転倒日の経過日数を X 日と計算する。
- 2) 非転倒事例の入院期間 II を起点として X 日経過した日を非転倒事例の疑似転倒日と設定する。

※転倒日が入院期間 II 以前の場合には X は負の値となる。

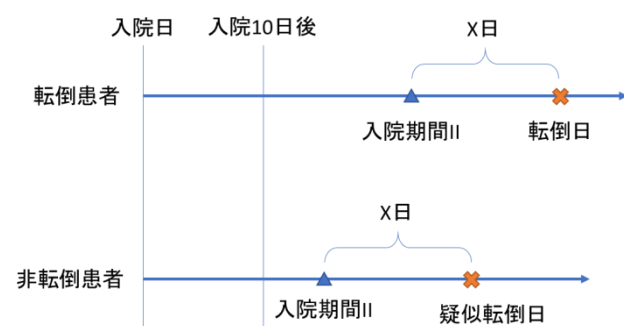


図1 非転倒事例の疑似転倒日の設定方法

ただし, 疑似転倒日が非転倒事例の退院日を超えてしまうケースや, 入院 10 日後以前になってしまうケースの場合は, マッチング候補から除外した。

2.3 治療パターンの時系列表現

両群の転倒日(疑似転倒日)前10日間の治療パターンを時系列表現する方法について説明する。

10 日間に行われた治療行為や血液検査結果, 転床歴及び転倒を1日毎のイベントアイテムとして表現する。

治療行為のイベントアイテムは, DPC データの EF ファイルから投薬・注射, 処置, 手術, 画像検査, リハビリイベントを抽出し, 以下の表に示す粒度でイベントアイテムとした。

表1 治療行為イベントアイテムの粒度

治療行為	粒度
投薬・注射	薬効分類4桁でまとめる(234種類) ただし, 1149(その他の解熱鎮痛消炎剤), 2329(その他の消化性潰瘍用剤), 3311(生理食塩液類)は除外した
処置	解釈番号3桁でまとめる(28種類)
手術	「手術」と「麻酔」としてまとめる(2種類)
画像検査	全て「画像検査」でまとめる(1種類)
リハビリ	全て「リハビリ」でまとめる(1種類)

血液検査結果イベントは, Batalらの例⁴⁾に倣い値を抽象化してイベントアイテムとした。例えば, BUN(血液尿素窒素)が

基準範囲と比較して高値であった場合は BUN-H(igh), 低値なら BUN-L(ow), 基準内なら BUN-N(ormal)というイベントアイテムとした。イベントアイテムとした血液検査項目は, BUN, Na, K, グルコース, HbA1c, BNP, CPK の7種類である。なお, イベント日は採血日とした。

転床等の病床移動に関するイベントは, 転科, 転棟, 転室・転床, 外泊・外出の4種類である。

これらのデータを, 患者IDをシーケンスID, 10日間の治療パターンの経過日数をイベントID, イベントアイテムをアイテムとしてデータベースに格納した。一例として図2に示すような治療パターンの患者のデータは, データベースのテーブルに図3のように格納される。

	1	2	3	4	5	6	7	8	9	10
注射A	○	○	○	○	○					
注射B				○	○	○				
処置C								○		
血液検査D	N		N	N			H	H	H	
:	:	:	:	:	:	:	:	:	:	:
転倒										○

図2 治療パターンの例

縦軸にイベントアイテムをとり, 横軸に治療パターン中の経過日数をとる。イベントアイテムを実施した日を○で表現する。血液検査の場合は, H,L,Nで表現する。

シーケンスID (患者ID)	イベントID (経過日数)	アイテム (イベントアイテム)
00123	1	注射A
00123	1	血液検査D-N
00123	2	注射A
00123	3	注射A
00123	3	血液検査D-N
:	:	:
00123	10	転倒

図3 データベースの格納テーブルの例

本研究では, 一連のイベントアイテムの時系列を, 次のように表現する。

まず, 一連の時系列は<>で括られて, 各イベントアイテムは, 発生日の古い順に左からカンマで区切られて並べられる。同一日に発生したイベントアイテムは{}で括られて, {}の内部では時間的順序の意味は持たない。例えば, 前述の患者(ID:00123)の治療パターン S_1 は次のように表現される。

$$S_1 = \langle \{A, D-N, \dots\}, \{A, \dots\}, \{A, D-N, \dots\}, \dots, \{\dots, \text{転倒}\} \rangle$$

2.4 時系列データからの相関ルールの抽出

時系列データから相関ルールを抽出するアルゴリズムは様々提案されているが, 本研究では Zaki の提案する SPADE 法⁵⁾の実装である統計ソフト「R」の「arulesequences」パッケージを利用した。

一般に, 時系列データから相関ルールを抽出する際には最小支持度と最小確信度を設定してから行う。ここで簡単に, 支持度と確信度について説明する。ある時系列データ α が別の時系列データ β の部分集合であるとき, 条件 α は時系

列データ β を満たすという。例えば、条件 $\alpha_1 = \langle \{A, B\}, C \rangle$ は、前述の S_1 を満たすが、条件 $\alpha_2 = \langle A, C, B \rangle$ は S_1 を満たさない。(処置 C の後に注射 B は発生していないので。) このとき、全時系列データの数に対する条件 α を満たす時系列データの数の割合を条件 α の支持度という。

また、条件 X を満たす時系列データで、同時に Y という条件も高い頻度で満たされるとき、相関ルール $R: \langle X \rangle \Rightarrow \langle Y \rangle$ が導かれる。このとき、条件 X を満たす時系列データの数に対する条件 X と Y の両方を満たす時系列データの数の割合を相関ルール R の確信度という。

言い換えれば、相関ルールの支持度は、そのルールが発現する頻度を表し、大きければ大きいほど、そのルールが登場しやすいことを意味する。また、相関ルールの確信度は、左辺(前提部)を満たす時系列データが右辺(帰結部)を満たす確率であり、大きければ大きいほど前提条件が満たされたときに結論部分が起きやすいことを意味する。

最小支持度を設定する際には、あまり小さい値だと、めったに発現しないルールまで抽出してしまうことになり、逆に大きすぎる値だとルールを全く抽出できないということになる。特に、本研究では右辺に転倒イベントを含むルールを抽出するのが目的であり、転倒事例は全体の 50% であるために、最小支持度は最大でも 0.5 となる。

一方、最小確信度を設定する際にも、あまり小さい値だと、前提部が起きた時に帰結部が起きる確率が低い、あまり意味のない相関ルールを多く抽出してしまうことになり、逆に大きすぎる値だとルールを全く抽出できないということになる。本研究では、転倒事例と非転倒事例がちょうど半数ずつであるために、偶然に頼っても 50% の確率で転倒事例となる。そこで、最小確信度は小さくとも 0.5 以上であることが望ましい。

本研究では、いくつかの試行の結果、最小支持度を 0.063、最小確信度を 0.62 と設定した。

2.5 抽出パターンの有意差検定

前節で抽出された相関ルールの妥当性を検証するために、ルール抽出に用いた患者セットとは異なる患者セットを用いて、前提条件となる治療パターンと転倒の関係性を検証した。

検証に用いたのは、順天堂医院に 2016 年 4 月～6 月の期間に入院期間が含まれる患者のうち、16 歳以上かつ 10 日以上入院、かつ DPC データとリンク可能な患者 1695 名である。このうち転倒インシデントのあった患者は 23 名である。

検証用患者の治療パターンを抽出するにあたり、転倒事例 23 例は、転倒日を基準に前 10 日間を時系列データ区間とした。残りの非転倒事例は、入院 10 日目以降、退院日までの期間でランダムに疑似転倒日を設定し、疑似転倒日前 10 日間を時系列データ区間とした。

各相関ルールに対して、検証用患者の治療パターンが前提部を満たすか否かと、転倒インシデントの有無を変数としたクロス集計表を作成し、フィッシャーの正確性検定を行った。

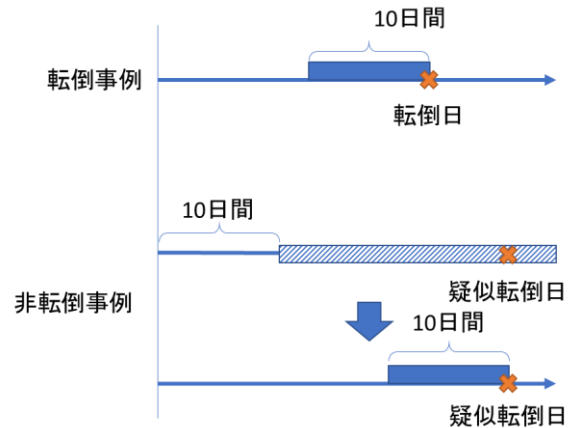


図4 検定用データの時系列データ区間の設定方法

3. 結果

3.1 相関ルールの抽出

最小支持度 0.063、最小確信度 0.62 で右辺に転倒イベントを含む相関ルールを抽出したところ、30 個のルールが抽出された。前提条件に含まれるアイテムは、投薬・注射では「その他の中枢神経系用薬」、「その他の催眠鎮静剤、抗不安剤」、「その他の精神神経用剤」、「抗アルドステロン製剤; トリアムテレン等」、「他に分類されない血液・体液用薬」、「その他の利尿剤」の 6 種類、血液検査では、K(カリウム)、Na(ナトリウム)、BUN(血液尿素窒素)、グルコースがいずれも基準値内というイベントアイテムの出現が認められた。それ以外には、転室/転床イベントとリハビリイベントが認められた。

30 個のルールは、巻末の表 2 に示す。

3.2 相関ルールの検証

前節で得られた 30 個の相関ルールに対して、前提部包含の有無と転倒インシデントの有無でクロス集計表を作成し、フィッシャーの正確性検定を行った。一例として相関ルール $\langle \{BS-N, K-N\}, \{精神\} \rangle \Rightarrow \langle \{転倒\} \rangle$ のクロス集計表を表 3 に示す。

表 3 相関ルールのクロス集計表の一例

	転倒有	転倒無	計
前提部満たす	6	125	131
前提部満たさない	17	1547	1564
計	23	1672	1695

フィッシャーの正確性検定を行うと p 値は 0.006402 (< 0.05) となり、相関ルールの前提部 $\langle \{BS-N, K-N\}, \{精神\} \rangle$ を満たすと有意に転倒しやすいという結果が得られた。

これ以外にも、

$\langle \{K-N\}, \{精神\}, \{精神\}, \{K-N\} \rangle$ ($p = 0.01327$)

$\langle \{K-N, Na-N\}, \{精神\}, \{BUN-N, K-N\} \rangle$ ($p = 0.01484$)

$\langle \{K-N\}, \{精神\}, \{精神, K-N\} \rangle$ ($p = 0.0123$)

$\langle \{K-N\}, \{精神, BS-N\} \rangle$ ($p = 0.01094$)

の 4 パターンに有意に差が認められた。

4. 考察

精神神経用剤の使用が、転倒発生に関連性があることは示唆されたが、他の治療行為や検査結果との前後関係から、生物学的妥当性のある説明を導くことはできなかった。本手

法は、因果関係が単純で治療行為の手順に起因するような要素の発見には適しているが、本研究のような転倒インシデントはベッドやトイレ等の周辺環境が大きく影響するので、この手法の適用には適さなかったと考えられる。

5. 結語

本発表では、DPC データや検査部門、看護部門の部門システムのデータを結合することで日単位の時系列データを作成し、入院中の転倒につながる治療パターンを探索的に発見する試みについて報告した。いくつか有意差のある相関ルールが見つかったが、臨床的な見地からの解釈は出来なかった。

今後は、転倒以外の有害事象についても同様の手法を試していきたい。

参考文献

- 1) JAHIS オーダエントリ・電子カルテシステム導入調査報告 ― 2017 年調査(平成 29 年)―
[https://www.jahis.jp/action/id=57?contents_type=23]
- 2) 厚生労働省 平成 30 年度新規 DPC 対象病院及び新規 DPC 準備病院説明会資料
[https://www.prrism.com/dpc/dpc_setumeikai_20180315.pdf]
- 3) 鈴木みずえ, 古橋玲子, 鶴見智子, 松下恵美, 岩田浩子, 内田敦子. 特定機能病院の内科・外科系病棟における転倒の実態と転倒リスクアセスメントツールの開発. 看護研究 2006 ; 39 : 139-151.
- 4) I. Batal, H. Valizadegan, G.F. Cooper, M. Hauskrecht. A Temporal Pattern Mining Approach for Classifying Electronic Health Record Data. ACM Trans Intell Syst Technol. 2013; Vol. 4: No. 4: pp. 63
- 5) M. J. Zaki. SPADE: An Efficient Algorithm for Mining Frequent Sequences. Machine Learning. 2001; 42, 31-60

表 2 抽出された相関ルール

No.	相関ルール	支持度	確信度	リフト
1	<{K-N, Na-N}, {転室/転床}> => <{転倒}>	0.0759	0.6760	1.3521
2	<{K-N}, {中枢}> => <{転倒}>	0.0680	0.6323	1.2647
3	<{K-N, 眠剤}> => <{転倒}>	0.0680	0.6231	1.2463
4	<{Na-N}, {転室/転床}, {K-N}> => <{転倒}>	0.0632	0.6451	1.2903
5	<{K-N}, {転室/転床}, {K-N}> => <{転倒}>	0.0680	0.6231	1.2463
6	<{BS-N, K-N}, {精神}, {K-N}> => <{転倒}>	0.0680	0.6417	1.2835
7	<{K-N}, {精神}, {精神}, {K-N}> => <{転倒}>	0.0664	0.6461	1.2923
8	<{精神, K-N, Na-N}, {BUN-N, K-N}> => <{転倒}>	0.0632	0.6349	1.2698
9	<{K-N, Na-N}, {精神}, {BUN-N, K-N}> => <{転倒}>	0.0712	0.625	1.25
10	<{抗ア}, {K-N, 抗ア}> => <{転倒}>	0.0632	0.625	1.25
11	<{K-N}, {精神}, {精神, K-N}> => <{転倒}>	0.0727	0.6216	1.2432
12	<{他血}, {他血}, {他血}, {リハ}, {リハ}> => <{転倒}>	0.0632	0.6451	1.2903
13	<{他血}, {他血, リハ}, {リハ}, {他血, リハ}> => <{転倒}>	0.0648	0.6406	1.2812
14	<{他血, リハ}, {他血}, {リハ}, {他血, リハ}> => <{転倒}>	0.0632	0.6451	1.2903
15	<{他血}, {リハ}, {他血, リハ}, {他血, リハ}> => <{転倒}>	0.0648	0.6406	1.2812
16	<{他血}, {他血}, {他血, リハ}, {他血, リハ}> => <{転倒}>	0.0632	0.625	1.25
17	<{K-N}, {精神, BS-N}> => <{転倒}>	0.0680	0.6417	1.2835
18	<{他血, K-N}, {他血}, {他血}, {K-N, Na-N}, {他血}> => <{転倒}>	0.0632	0.625	1.25
19	<{他血, K-N, Na-N}, {他血}, {他血}, {他血, Na-N}, {他血}> => <{転倒}>	0.0648	0.6406	1.2812
20	<{K-N}, {他血}, {他血}, {他血}, {他血, Na-N}, {他血}> => <{転倒}>	0.0632	0.6451	1.2903
21	<{他血}, {他血, K-N, Na-N}, {他血}, {他血, K-N}, {他血}> => <{転倒}>	0.0632	0.625	1.25
22	<{他血, K-N, Na-N}, {他血}, {他血}, {他血, K-N}, {他血}> => <{転倒}>	0.0632	0.6349	1.2698
23	<{他血}, {他血}, {他血}, {他血}, {他血, Na-N}, {他血}, {他血}> => <{転倒}>	0.0664	0.6268	1.2537
24	<{BUN-N, K-N, Na-N}, {K-N, 利尿}, {利尿}> => <{転倒}>	0.0664	0.6268	1.2537
25	<{Na-N}, {抗ア}, {抗ア}> => <{転倒}>	0.0632	0.6451	1.2903
26	<{K-N}, {K-N, 抗ア}, {抗ア}> => <{転倒}>	0.0632	0.6451	1.2903
27	<{抗ア, 利尿}, {抗ア}> => <{転倒}>	0.0632	0.625	1.25
28	<{抗ア}, {抗ア}, {抗ア}> => <{転倒}>	0.0680	0.6231	1.2463
29	<{他血, K-N, Na-N}, {他血}, {他血}, {他血, Na-N}, {他血}> => <{他血, 転倒}>	0.0648	0.6406	5.6232
30	<{他血}, {他血}, {他血}, {他血}, {他血, Na-N}, {他血}, {他血}> => <{他血, 転倒}>	0.0664	0.6268	5.5024

凡例:

中枢: その他の中枢神経系用薬(薬効分類 1190)…リリカパセル, ラジカット, メマリー, アリセプトなど
 眠剤: その他の催眠鎮静剤, 抗不安剤(薬効分類 1129)…マイスリー, アモバン, ゾルピデム, リスミーなど
 精神: その他の精神神経用剤(薬効分類 1179)…アタラックス, デパス, サインバルタ, セレネースなど
 抗ア: 抗アルドステロン製剤; トリアムテレン等(薬効分類 2133)…アルダクトン, ソルダクトン, トリテレンなど
 他血: 他に分類されない血液・体液用薬(薬効分類 3399)…バイアスピリン, プラビックス, エパデールなど
 利尿: その他の利尿剤(薬効分類 2139)…ラシックス, ダイアート, サムスカなど
 K-N: カリウム-基準値内 (5.0-3.5mmol/L)
 Na-N: ナトリウム-基準値内 (145-135mmol/L)
 BS-N: グルコース基準値内 (109-65mg/dl)
 BUN-N: 血液尿素窒素-基準値内 (21-9mg/dl)
 リハ: リハビリ(心大血管疾患リハ, 脳血管疾患等リハ, 運動器リハ, 呼吸器リハ等を含む)
 フィッシャーの正確性検定で有意差が認められたルールの番号を網掛けで示す。

