

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 10 日現在

機関番号：32653

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22590610

研究課題名（和文） 糖尿病のアウトカム研究に適したデータマイニングアルゴリズムの開発

研究課題名（英文） Development of data-mining algorithm appropriate for outcome research of diabetes mellitus

研究代表者

佐倉 宏（SAKURA HIROSHI）

東京女子医科大学・医学部・教授

研究者番号：70240710

研究成果の概要（和文）：

糖尿病のアウトカム研究に適したデータマイニングアルゴリズムの開発を試みた。現在は患者背景因子から患者のアウトカムを予測する統計学手法として重回帰分析やロジスティック回帰分析が主に用いられているが、データマイニング手法のニューラルネットワーク手法や決定木手法を用いるとより優れた予測が行えることが明らかとなった。今後、条件設定や前向き試験による検証が必要である。

研究成果の概要（英文）：

We tried to develop the data-mining algorithm appropriate for the outcome study of diabetes mellitus. Data-mining methods such as neural network or decision tree were superior to multiple regression analysis or logistic regression analysis, respectively.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010 年度	1,800,000	540,000	2,340,000
2011 年度	800,000	240,000	1,040,000
2012 年度	800,000	240,000	1,040,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：医歯薬学

科研費の分科・細目：社会医学・公衆衛生学・健康科学

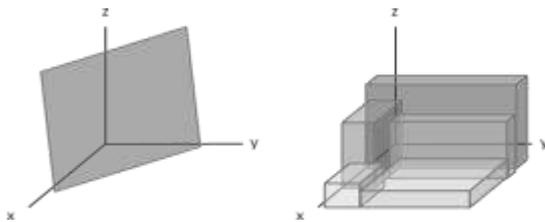
キーワード：糖尿病、データベース、データマイニング、医療情報学、多変量解析

1. 研究開始当初の背景

糖尿病は慢性高血糖を主徴とする疾患と定義される。発症数年以降にさまざま合併症を生じて QOL の低下が起こり、有病率が高いために社会的な損失も大きい疾患である。しかし、合併症の程度やそれに関与する血糖

コントロール（HbA1c）は患者ごとに大きな差があるので、患者背景因子（説明変数）からこれらのアウトカム（反応変数）を適切に予知することができれば診療の質は大いに向上する。現在、疾患のアウトカム研究では、主として重回帰分析、ロジスティック回帰分

析、Cox 比例ハザード分析が用いられている。これらの多変量解析では、説明変数と反応変数が線形関係にあることが前提条件となっているが、糖尿病をはじめとする多くの疾患の説明変数と反応変数の関係はより複雑であり、従来の解析では集団におけるアウトカムに關与する因子の同定はできても、個々の患者のアウトカムを予知するには限界がある。多変量解析の発展型とも言えるデータマイニング手法はアウトカム予知の点ではより優れている可能性が高い。すなわち、模式的に示せば、従来の多変量解析は左の図で示されるように変数間の最適な関係を平面的に求めるのに対し、データマイニング手法は右の図で示されるように変数間の複雑な関係をより細分化して最適な関係を求めることができるからである。



現在、臨床疫学において、最高のエビデンス得る方法は大規模臨床試験とされるが、結果を得るためには多大な費用と膨大な時間がかかる。また、患者の選択基準が限定されており、糖尿病のような病態が多様な疾患においては、選択基準に合致する患者は少数に過ぎず、診療の現場で大規模臨床試験の結果に基づく EBM の実践にも限界がある。

一方、電子カルテの普及に伴い、膨大な量のデータが蓄積されつつあり、そこから有用な医療情報を導き出すことができれば大規模臨床試験より優れたエビデンスとなる可能性が高い。このようにデータベース化された膨大な情報の解析においてもデータマイニング手法は優れた方法である。しかし、マーケティングや金融の分野ではよく利用

されるデータマイニング手法ではあるものの、医学研究における応用は国内・国外を問わずほとんど行われていない。

研究申請者が属する施設では 2003 年から電子カルテが導入されており、糖尿病の定期通院患者も約 12,000 名と世界最大規模の入院・外来診療を行っている。このようにデータベースの活用には有利な環境を利用して、2000 年から糖尿病関連情報のデータマート構築を行っている。そして、すでに 1,000 万件を超える医療情報を収集し、研究業績に示すようにさまざまな解析を行っている。データマイニング手法に関する研究にも着手しており、初診患者の 6 ヶ月後の HbA1c 値という数値予測においては、重回帰分析に比べてニューラルネットワークアルゴリズム、初診患者の 6 ヶ月後の血糖コントロールの良否という二値予測においては、ロジスティック回帰分析に比べて決定木アルゴリズム (C5.1) がより優れていることを見出している (研究業績 4,5)。以上のように、本研究の実施にあたっては他施設に比べて有利な環境と研究実績を有している。

2. 研究の目的

血糖コントロール指標 (HbA1c) および合併症 (細小血管障害および大血管障害) をアウトカムとし、特定の診療時点の患者背景因子 (説明変数) から、個々の患者の将来のアウトカム指標 (反応変数) を最もよく予測するためのデータマイニングアルゴリズムを開発する。また、開発したデータマイニングアルゴリズムが Web を通して広く利用できるようにシステムの構築を目指す。

3. 研究の方法

(1) データマイニング手法を行う基盤となるデータマートのさらなる充実をはかる。(研究代表者: 佐倉 宏、研究協力

者：飛田杏子、落合華奈)

- (2) データマイニングの有用性を解析する
課題として以下のテーマを行う。

入院患者の6ヶ月後の血糖コントロールや5
年後の合併症の予知(研究代表者：佐倉 宏、
研究協力者：落合華奈)

- 75gOGTT および食事負荷試験の詳細な
解析(分担研究者：菅野宙子、研究協力
者：高井孝典)

- (3) 上記の課題について、重回帰分析、ロジ
スティック回帰分析(ニューラルネット
ワーク、決定木手法他)の比較を行い、
最適なデータマイニング手法を確立す
る。

4. 研究成果

- (1) 電子カルテテンプレートからの医療情
報収集、電子カルテからのデータ収集の
半自動化を通じて、2013年3月の時点
で、約1,500万件の糖尿病関連データ
を含んだデータマートとなっている。
- (2) 各課題について、論文ならびに学会発表
で報告した(詳細は5.主な発表論文等に
列挙)
- (3) 初診患者の6ヶ月後の血糖コントロール
(HbA1c)の予知に関して検討を行った。
主なデータマイニング手法と従来の多
変量解析の比較を表に示す。
二値予測に関しては、決定木手法(C5.1)
の予測精度が79.8%、変数をあらかじめ
層別化しておく88.8%と最も高く、ロ
ジスティック回帰分析の72.6%、69.0%
よりも優れていた。
数値予測に関しては、Neural network
重回帰分析での予測値と実際の値の相
関係数がニューラルネットワークで
0.54と重回帰分析の0.42よりも優れて
いた。
他の課題についての従来手法とデータ

マイニング手法の比較は現在解析中
である。

	二値 予測 (%)	変数層別 二値予測 (%)	数値 予測 (R)
C5.1	79.8	88.8	
C&R tree	76.8	77.7	
CHAID	74.7	75.3	
Neural network	74.4	74.9	0.54
QUEST	73.0	74.6	
ロジスティッ ク回帰分析	72.6	69.0	
Bayes	69.9	72.7	
Decision list	69.8	71.8	
SVM	69.0	78.0	0.46
判別 一般化線形	60.9	56.7	0.43
重回帰分析			0.42

％：精度、R：相関係数

今後、データマイニング手法の条件の設定や
前向きな検討での検証作業を行う必要があ
る。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者に
は下線)

[雑誌論文](計4件)

- ① Takai T, Sakura H, Uchigata Y,
Iwamoto Y, Insulin sensitivity and
secretion indices in Japanese subjects
with normal and impaired glucose
tolerance, Diabetology International, 3,
80-85, 2012, 査読有

- ② Sakura H, Tanaka Y, Iwamoto Y, Seasonal fluctuations of glycated hemoglobin levels in Japanese diabetic patients. *Diabetes Research & Clinical Metabolism*, 88, 65-70, 2010, 査読有
- ③ 清村杏子, 佐倉 宏, 落合華奈, 菅野宙子, 丸山聡子, 岩本安彦, 外来および入院患者医療情報の電子カルテからの抽出とそのデータマート化, 肥満と糖尿病9(別冊) 日本糖尿病情報学会誌, 9, 26-32, 2010, 査読有
- ④ 落合華奈, 佐倉 宏, 清村杏子, 菅野宙子, 丸山聡子, 岩本安彦, 電子カルテから抽出した医療情報を用いた1型糖尿病のHbA1cとインスリン用法の解析, 肥満と糖尿病9(別冊) 日本糖尿病情報学会誌, 9, 26-32, 2010, 査読有

[学会発表] (計 10 件)

- ① 佐倉 宏, 菅野宙子, 石澤香野, 飛田杏子, 高井孝典, 落合華奈, 田中康富, 内潟安子, 長期通院患者の2005-2010年の血糖コントロール改善および改善に寄与する因子の解析, 第 55 回日本糖尿病学会年次学術集会, 2011.5.17, 横浜
- ② 佐倉 宏, インスリンと経口薬の併用ーDPP-4 阻害薬も含めてー, 第 55 回日本糖尿病学会年次学術集会, 2011.5.18, 横浜
- ③ 高井孝典, 佐倉 宏, 内潟安子, 75g ブドウ糖負荷試験におけるインスリン分泌・感受性指標の性および年齢の影響, 第 55 回日本糖尿病学会年次学術集会, 2011.5.18, 横浜
- ④ 落合華奈, 佐倉 宏, 内潟安子, 教育入院した糖尿病患者の5年後のイベント発症に関与する因子の解析, 第 55 回日本糖尿病学会年次学術集会, 2011.5.19, 横浜

- ⑤ Takai T, Sakura H, Uchigata Y, Iwamoto Y, Insulin sensitivity and insulin secretion indices in Japanese subjects with normal and impaired glucose tolerance, 20th World Diabetes Congress, 2011.12.5, Dubai, UAE
- ⑥ 高井孝典, 佐倉 宏, 岩本安彦, 正常および耐糖能異常者におけるインスリン抵抗性・分泌能と臨床背景の重回帰分析, 第 54 回日本糖尿病学会年次学術集会, 2011.5.20, 札幌
- ⑦ 菅野宙子, 佐倉 宏, 岩本安彦, 2型糖尿病のインスリン抵抗性・分泌能の推移と背景因子の解析, 第 54 回日本糖尿病学会年次学術集会, 2011.5.19, 札幌
- ⑧ 飛田杏子, 佐倉 宏, 岩本安彦, シタグリプチン投与による HbA1c 改善群と非改善群の背景因子の比較および副作用, 第 54 回日本糖尿病学会年次学術集会, 2011.5.19, 札幌
- ⑨ 高井孝典, 佐倉 宏, 岩本安彦, 正常および耐糖能異常者のインスリン分泌・抵抗性の年齢別解析と経年的な横断的解析, 第 53 回日本糖尿病学会年次学術集会, 2010.5.27, 岡山
- ⑩ 清村杏子, 佐倉 宏, 岩本安彦, データマートを用いた1型および2型糖尿病患者の血糖コントロールと糖尿病治療法の経年変化の解析, 第 53 回日本糖尿病学会年次学術集会, 2010.5.27, 岡山

6. 研究組織

(1) 研究代表者

佐倉 宏 (SAKURA HIROSHI)
東京女子医科大学・医学部・教授
研究者番号：70240710

(2) 研究分担者

菅野 宙子 (KANNO HIROKO)
東京女子医科大学・医学部・助教
研究者番号：90338971

岩本 安彦 (IWAMOTO YASUHIKO)
東京女子医科大学・医学部・教授
研究者番号：60143434
(平成 22 年度のみ)

(3) 連携研究者
なし

(4) 研究協力者
高井 孝典 (TAKAI TAKANORI)
東京女子医科大学・医学部・助教
研究者番号：なし

落合 華奈 (OCHIAI KANA)
東京女子医科大学・医学部・研究生
研究者番号：なし

飛田 杏子 (OTOBITA KYOKO)
東京女子医科大学・医学部・研究生
研究者番号：なし