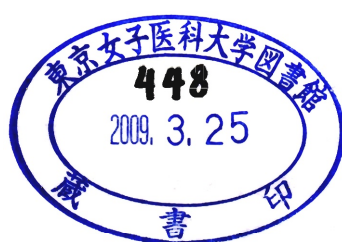

電子カルテのデータマート化による糖尿病医療情報解析

18590499

平成 18 年度～平成 19 年度科学研究費補助金
(基盤研究 (C)) 研究成果報告書



平成 21 年 3 月

研究代表者 佐倉 宏

東京女子医科大学医学部准教授



<はしがき>

平成 18～19 年度科学研究費補助金（基盤研究（C））により、「電子カルテのデータマート化による糖尿病医療情報解析」の研究を行い、ここにその成果を報告する。なお、東京女子医科大学大学院生 春木武徳および石澤香野の両氏には本研究に多大に協力していただいた。また、検査技師の田中康富氏にはデータ入力・解析に協力していただいた。

研究組織

研究代表者：佐倉 宏（東京女子医科大学医学部准教授）
研究分担者：菅野宙子（東京女子医科大学医学部助教）
研究分担者：丸山聡子（東京女子医科大学医学部助教）
研究分担者：岩本安彦（東京女子医科大学医学部教授）
（研究協力者：春木武徳（東京女子医科大学大学院生））
（研究協力者：石澤香野（東京女子医科大学大学院生））

交付決定額（配分額）

（金額単位：円）

	直接経費	間接経費	合計
平成 18 年度	2,400,000	0	2,400,000
平成 19 年度	1,100,000	3,300,000	1,430,000
総計	3,500,000	3,300,000	3,830,000

研究発表

(1) 雑誌論文

1. 佐倉 宏：電子カルテシステムおよび情報システムを利用した糖尿病臨床データ解析（糖尿病診療の診療の新時代と IT）.肥満と糖尿病（査読有）5（別冊 5）: 2006, 90-92
2. 佐倉 宏, 岩本安彦, 菅野宙子, 宇都祐子, 丸山聡子, 藤川径子, 春木武徳, 田中香野: 患者臨床データベースを用いた経口血糖降下薬の効果の解析. 臨床薬理の進歩（査読無）27: 2006, 69-75
3. 佐倉 宏, 岩本安彦: 経口血糖降下薬の大規模臨床試験の特殊性と提言. 糖尿病（査読有）50: 2007, 635
4. 佐倉 宏: 電子カルテを用いたデータ解析とは? 肥満と糖尿病（査読有）6: 2007, 471-472
5. Haruki T, Sakura H, Iwamoto Y: Identification of factors predicting glycemic control in new patients with type 2 diabetes. J Tokyo Wom Med Univ（査読有）79: 2009 「印刷中」
6. Maruyama T, Sakura H, Kanno H, Iwamoto Y: Factors associated with glycemic control after an inpatient program. Metabolism. 「発表予定」

7. Sakura H, Tanaka Y, Iwamoto Y: Seasonal fluctuations of glycosylated hemoglobin levels in Japanese diabetic patients. 「発表予定」

(2) 学会発表

1. Sakura H, Haruki T, Iwamoto Y: The analysis of the factors which predict the glycemic control of the newly-visited type 2 diabetic patients using a hospital-based datamart. 66th Annual Meeting of American Diabetes Association, 2006.6, Washington, USA
2. 佐倉 宏, 春木武徳, 菅野宙子, 丸山聡子, 藤川径子, 田中香野, 岩本安彦: 電子カルテを利用した効率的な糖尿病データマートの構築. 第 49 回日本糖尿病学会年次学術集会, 2006.5, 東京
3. 春木武徳, 佐倉 宏, 岩本安彦: 2 型糖尿病患者の初診時現病歴の HbA1c に影響する因子の検討. 第 49 回日本糖尿病学会年次学術集会, 2006.5, 東京
4. 田中香野, 熊野宏昭, 佐藤 徳, 佐倉 宏, 岩本安彦: 肥満に伴う前頭葉機能変化の解析. 第 49 回日本糖尿病学会年次学術集会, 2006.5, 東京
5. 田中康富, 佐倉 宏, 岩本安彦: HbA1c の季節変動の解析. 第 49 回日本糖尿病学会年次学術集会, 2006.5, 東京
6. Sakura H, Tanaka Y, Iwamoto Y: Seasonal fluctuation of HbA1c in Japanese diabetic patients. 19th World Diabetes Congress, 2006.12, Cape Town, South Africa
7. 佐倉 宏, 菅野宙子, 岩本安彦: 経口糖尿病薬による二次無効の検討. 第 50 回日本糖尿病学会年次学術集会, 2007.5, 仙台
8. 佐倉 宏, 春木武徳, 菅野宙子, 丸山聡子, 石澤香野, 田中健富, 岩本安彦: 解析ツールを用いた臨床新知見の分析 (シンポジウム: 検診義務化などの新しい動きをふまえて電子情報を糖尿病診療にどう生かすか). 第 50 回日本糖尿病学会年次学術集会, 2007.5, 仙台
9. 菅野宙子, 佐倉 宏, 岩本安彦: スルホニル尿素薬単独療法 5 年後の臨床経過の変遷について. 第 50 回日本糖尿病学会年次学術集会, 2007.5, 仙台
10. 丸山聡子, 佐倉 宏, 菅野宙子, 田中康富, 岩本安彦: 血糖コントロール目的で初診した 2 型糖尿病患者における入院と外来での治療の比較. 第 50 回日本糖尿病学会年次学術集会, 2007.5, 仙台
11. 春木武徳, 佐倉 宏, 岩本安彦: 初診 2 型糖尿病患者における尿酸値と動脈硬化危険因子の関係の検討. 第 50 回日本糖尿病学会年次学術集会, 2007.5, 仙台
12. 田中康富, 佐倉 宏, 岩本安彦: 糖尿病患者における HbA1c の季節変動パターンの解析. 第 50 回日本糖尿病学会年次学術集会, 2007.5, 仙台
13. Sakura H, Tanaka Y, Iwamoto Y: Seasonal fluctuation of glycosylated hemoglobin in Japanese diabetic patients. 14th Japan-Korea Symposium on Diabetes Mellitus, 2007.10, Kyoto
14. 佐倉 宏: 肥満 2 型糖尿病にはどのような経口血糖降下薬が有効か? 第 22 回日本糖尿病合併症学会, 2007.10, つくば

(3) 図書

1. 佐倉 宏: 西村書店, 疫学とエビデンス 糖尿病と医療情報—データベースの作成
(門脇 孝, 石橋 俊, 佐倉 宏, 戸辺一之, 野田光彦 編) .糖尿病学 基礎と臨床,
2007, 820-823

研究成果による産業財産権の出願・取得状況
なし

研究目的

- ① 電子カルテは公正・可視・保存性など多くの点で有用であり、急速に普及してきている。しかし、まだ診療と医事会計に利用されているにすぎず、臨床研究に応用する試みはほとんど行われていない。電子カルテは端末から入力された情報を瞬時に正確に処理・保存を行う必要がある基幹系システムと位置づけられる。一方、保存された膨大な医療情報を分析するためには、基幹系システムとは独立したデータマート（データウェアハウス）などの情報系システムを構築する必要である。本研究の目的は、第一に、電子カルテシステムに保存される医療情報を容易に解析できる形式に変換してデータマート化する方法を確立することである。第二に、上記の方法でデータマート化された糖尿病に関する診療情報を用いて、多変量解析、データマイニング、最適化解析を行い、糖尿病の病因・病態について有用な知識を発見し、治療に応用することである。
- ② 今回の研究の対象疾患である糖尿病は単一疾患にもかかわらず病態が複雑であり、発症や予後を規定する因子は多数存在し、治療も他の疾患に比べて困難である。また有病率が非常に高く、さまざまな重症度の患者が混在していることも大きな特徴である。このような複雑かつ患者数の多い疾患を解析しより良い治療法を確立するためには、病態に関与する多数の因子をデータマート化して解析するのが有力な方法である。医事・検査・薬物情報といったすでに構造化されている医療情報については比較的容易にデータマート化することができるが、自由記載形式で入力された診療録情報についてのデータマート化は困難であった。紙カルテの場合は人手を使ってひとつひとつデータを収集していく他に方法はないが、電子カルテを利用した場合でも構造化されていないSOAPなどの医療情報はそのままの形で解析を行うことは困難である。電子カルテには効率的に入力するためテンプレートが用意されているが、それをうまく活用すればデータの構造化も可能と考えられる。そこで本研究においては、テンプレートを利用してできるだけ多くの糖尿病に関する医療情報をデータマート化することを試みる。そしてデータマート化した医療情報を多変量解析・データマイニング・最適化解析などさまざまな方法で解析することにより、糖尿病の病因・病態について有用な知識を発見したり、治療に応用したりすることが可能と考えられる。このように大量のデータを電子カルテから収集して解析する手法は疫学研究の新しい手法としても有用であると考えられる。電子カルテはすでに医療の現場に広く普及してきており、システムとしてはかなり満足すべきレベルまで到達している。今後はより便利なシステムへの追求と、データの二次利用が大きなテーマになると考えられる。本研究は電子カルテの二次利用についての先駆的な研究であり、意義が大きいものと考えられる。
- ③ 申請者の所属する施設では2003年7月に外来診療が全面的に電子カルテ化されたが、最も早くから電子カルテが稼働している大規模病院のひとつである。また、申請者が所属する糖尿病センターは常時通院している糖尿病患者数が約15,000名と世界最大規模である。申請者は2000年よりデータマート化に取り組んでおり、データマートに蓄積した医療情報もすでに600万レコードを超えている。そして、データマートの構築とその解析についてすでに多くの学会・論文発表を行ってきた（研究業績1-11）。またデータマートの構築と電子カルテ情報の利用についてはすでに施設内の倫理委員会に承

認されている。

テンプレートを用いた診療情報の入力についても電子カルテ導入当初から行っている。今回の申請研究で提案しているテンプレートから情報を引き出す試みについてパイロット的に検討したところ、3,000人以上の初診患者の間診・身体所見を一括化してデータマート化できることが確認できた(第25回日本医療情報学会にて発表予定)。したがって、本研究の遂行は十分可能であるし、最適な環境に恵まれていると考える。国内外を見ても類似した研究はほとんど見当たらない。その第一の理由として、データマートの構築はかなり専門的な知識が必要なことが挙げられる。申請者は経済産業省の情報処理技術者の資格を有しており、データベースの構築・運用に関する知識を十分に有している。第二の理由として、データマートを構築する能力のあっても電子カルテを導入している施設はまだ限定されていることが挙げられる。このような観点からも、本研究は独創的であり、今後の医療情報技術発展全体に対しても意義が大きいものとする。

従来の研究経過・研究成果

- 1) 基盤研究(C)、平成 16・17 年度、データマイニング手法を用いた糖尿病臨床情報の解析、370 万円
 1. データ収集とデータマート化
II で得た研究費を用いて申請者が所属する部署に入院・外来患者の糖尿病臨床情報のデータマート構築を開始していたが、収集レコード数を約 600 万件に充実させた。
 2. 統計学的解析とデータマイニング解析
経口血糖降下薬の効果、入院患者の血糖コントロール推移に及ぼす因子について多変量解析を行い(研究業績 4・11)、さらにデータマイニング手法(因子分析、アソシエーション分析、決定木)を用いて上記の研究結果を発展させた。これらの研究を通じて、糖尿病発症後から治療開始までの期間が入院・薬物いずれの効果にも非常に大きな影響を与えることが明らかとなった。また、EBM の観点から推奨される治療が行われているにもかかわらず、良好な血糖コントロールレベルを達成していない患者がまだ非常に多い実態が明らかになった。
- 2) 武田科学振興財団報彰基金、2001 年度、糖尿病患者情報のデータベース化と薬物療法の効果の解析、佐倉 宏(研究代表者)、150 万円
- 3) ダノン学術研究助成金、2002 年度、糖尿病臨床情報のデータベース化とその解析―食事・運動・薬物療法の効果の解析とビタミン E の研究への応用―、岩本安彦(研究代表者)、佐倉 宏(研究分担者)、300 万円
- 4) 臨床薬理振興財団研究奨励交付金、平成 14 年度、患者臨床情報データベースを用いた経口血糖降下薬のテーラーメイド処方、佐倉 宏(研究代表者)、300 万円

上記はいずれも、申請者が所属する部署における糖尿病患者臨床情報をデータマート化し、その解析を行った研究である。通院歴のある約 30,000 人の糖尿病患者について、①医事システムからの患者基本情報、②薬剤部システムからの薬物情報と処方情報、③検査部

システムからの検査情報、④所属部署システムからの退院サマリー情報を中心に収集した。

入力した情報は、リレーショナルデータベースの形で保存し、そこから必要な情報をデータベース言語 SQL (Structured Query Language)を用いて選択・抽出し、統計学的な解析を行なった。

申請者の施設に通院している患者の平均的臨床像の解析、経口血糖降下薬の投与状況の解析、薬物効果の解析、薬物に対する responder、non-responder の背景因子の相違点についての解析、糖負荷試験・食事負荷試験の解析の結果などを学会・論文にて発表した。

研究準備状況

I.

1. データマートの構築に関しては、施設内倫理委員会から平成16年8月12日に承認された。その承認に基づき、病院内関係部署に電子カルテ情報の利用許可を得て、医療情報を収集している。
2. セキュリティーの観点から、データマートは孤立した1台のPC内に構築してあり、すでに600万件を超える医療情報が蓄積されている。
3. いったん入力されたら更新されることのない初診患者の問診・身体所見について、テンプレートから表形式に移行してデータマートに取り込むパイロットスタディーを試み、動作することを確認した。
4. データマートからのSQL(Structured Query Language)を用いた医療情報の抽出はすでに日常的に行っており、多変量解析を中心とした解析実績も十分に有している。
5. データマイニングについてはまだ中規模のデータ解析しか行っていないが、多変量解析と矛盾がなくさらに発展した結果が得られている。

II.

1. サーバ・クライアントシステムについては、申請者の所属する部署にすでに設置済みであり、データマートを移行する準備が整っている。
2. クライアント用のPCについてはすでに2台有しているが、平成18年度にさらに購入予定である。
3. OCR機器と関連ソフトウェアはすでに導入済みである。
4. 基本ソフトウェアやデータマイニングソフトウェアの多くをすでに所有しているが、upgradeやソフトウェアの更なる充実は平成18-19年度の予算で行う予定である。

本研究には新たな物理的スペースは必要でないので、現在の研究環境で十分遂行が可能である。

研究計画・方法 (平成18年度)

I.

1. テンプレートの見直しと新規作成

すでに、テンプレートを用いて入力された初診患者の問診・身体所見はデータマートへ移行できることが確認できたが、項目名の不備などのため一部の医療情報の移行が困難であることが明らかになった。この点を修正したテンプレートを作成する。糖尿病の

解析に必要な再診患者の身体所見、合併症関連情報、心理的因子についてもテンプレートを用いて電子カルテに入力できるようにする（研究代表者：佐倉 宏）。

2. OCRを用いた電子カルテ記載情報の変換

テンプレート内に入力された情報はSOAP欄に書き込まれる。テンプレートは更新されると古い情報は削除されるが、電子カルテ内のSOAP欄に情報は残る。ただし、SOAP欄の情報は構造化されていないので、そのままの形ではデータマートに取り込めない。そこで、OCRを用いて、SOAP欄の情報を表形式に変換することを試みる（研究代表者：佐倉 宏）。

3. 電子カルテ情報のデータマートへの移行

テンプレート内の医療情報について、順次データマートに移行する。更新されないうちに取り込む必要があるため、毎月1回データマートへ取り込む。平成18年度は医療情報部にテンプレート情報の表形式への変換を依頼してデータマートへ移行する（研究代表者：佐倉 宏）。

テンプレートを用いることができないが糖尿病解析に重要な医療情報については、電子カルテから研究者自身でデータを抽出し、データマートへ手入力する（分担研究者：菅野宙子、丸山聡子）。

4. データマートのサーバ・クライアントシステムへの移行

複数の研究者でデータマートを効率よく利用できるようにするために、データマートをサーバ・クライアントシステムに移行する。クライアント用のPCを2台購入する。セキュリティの確保については、専門業者に依頼する。

5. データ解析

データマート化した医療情報を使って、①血糖コントロールに及ぼす初診時間診・身体所見の背景因子の解析（研究協力者：春木武徳—大学院生）、②薬物効果に関与する指標に関する研究（分担研究者：菅野宙子）、③血糖コントロールに関与する心理因子の解析（研究協力者：田中香野—大学院生）、④気温・体温が血糖コントロールに及ぼす影響についての解析（分担研究者：丸山聡子）を行う。

II. 生命倫理・安全対策に関する留意事項

本研究に関して、すでに「疫学研究に関する倫理指針」（平成14年6月17日 文部科学省・厚生労働省告示第2号）及び「疫学研究に関する倫理指針の施行等について」（平成14年6月17日 文部科学省研究振興局長・厚生労働省大臣官房厚生科学課長連名通知）に基づいた研究計画書を施設内倫理委員会に申請して、承認が得られている（2004年8月12日、承認番号567）。さらに、医事・薬剤・検査情報に関しては、担当部署に研究計画書を提出して、許可を得てからデータを収集した。

電子カルテから膨大な情報をデータマートに移行して解析を行うため、個人情報の保護とセキュリティには最大限の留意を払う必要がある。現段階では、孤立したPCにのみデータを蓄積しているが、より強固なデータマートを構築するために、本格的なデータベース管理システムとサーバ・クライアントシステムを構築する。同時に、外部に患者情報が漏れないように、データ管理者（研究分担者：岩本安彦）およびデータベース管理者（研

究代表者：佐倉 宏)をおく。また、システム構築の一部を専門業者に依頼し、セキュリティ一面を万全なものとする。

研究計画・方法 (平成19年度)

I.

1. テンプレートからデータマートへの移行の自動化

医事情報・薬剤情報については、1ヶ月に1回データが自動的にデータマートへ送信されるシステムをすでに確立してある。検査情報、テンプレート内の医療情報についても自動的にデータマートへ送信されるシステムを確立する。これにより、より効率的にデータマートが構築できる(研究代表者：佐倉 宏)。

2. データマートの充実

データマート内に蓄積された医療情報の整合性の検討、データの洗浄(はずれ値の検討、入力ミスの排除など)、重要な導出関数(身長と体重からBMI、糖尿病発症年と初診日から糖尿病経過年、など)をデータマートに組み込み、データマートの充実化をはかる(研究代表者：佐倉)。順調に行けば、蓄積情報数は1,000万件を超えると予想される。

3. データ解析

前年度に開始したデータ解析を継続し、結果をまとめる。

4. データマイニング手法による解析

前年度に開始したテーマについて、データマイニング手法(因子分析、決定木、アソシエーション分析など)を使ってさらに解析を進める。

また、データマートに蓄積された情報全体をデータマイニング手法を用いて解析し、未知の因子間の関連性の発見、血糖コントロールや合併症を低下させるにはどの因子が重要で治療によってどの程度改善しうるかについて最適化解析を用いて検討する。

II. 生命倫理・安全対策に関する留意事項

前年度と同じである。

研究成果

1. テンプレートの見直し、新規作成、データマートへの移行の自動化

病棟カルテの電子カルテ化に伴い、退院サマリーをテンプレートを用いて入力できるようにし、データマートへ簡単に移行できるようにした。

2. OCRを用いた電子カルテ記載情報の変換

従来の退院サマリーに関しては、ほとんど同一の形式でワープロを用いて作成されていた。病名に関してはOCRを用いてデータマートへ取り込むことができるようになった。

3. 電子カルテ情報のデータマートへの移行とその充実

従来から取得していた医療情報に加えて退院サマリーもデータマート化することができ、全部で1,000万件以上の医療情報を収集し、効率よく解析することが可能となった。

4. データマートのサーバ・クライアントシステムへの移行

データマートの一部をサーバ・クライアントシステムに移行し、本研究に参与した研究者がデータにアクセスすることが可能になった。

5. データ解析

データマート化された医療情報を用いて、HbA1cの季節変動に関する検討、血糖コントロールを予知しうる初診時因子をはじめとする解析を行い、学会・論文を通じて発表した。雑誌論文として発表予定の3編（文献5-7）を添付する。

6. データマイニング手法による解析

【目的】データマイニングは大規模なデータベースの中から有用な情報を抽出するための新しい手法であり、多変量解析の発展型として位置づけられる。種々の解析アルゴリズムが開発・利用されているが、医学分野における意義はまだ明らかではない。今回、初診患者の背景因子から6か月後のHbA1c値を予測する方法として、通常の変量解析とデータマイニング手法を用いた解析のどちらが優れているかについて比較検討した。

【方法】2003年7月-2005年12月に血糖コントロール目的で初診し、初診時HbA1cが6.5%以上かつ6か月以上通院した2型糖尿病788例の背景因子を電子カルテから一括して抽出し、6か月後のHbA1c値および血糖コントロール指標（HbA1c 6.5%未満：良好、HbA1c 6.5%以上：不良）との関連性を、さまざまな手法を用いて予測した。通常の変量解析としてロジスティック回帰分析および重回帰分析、データマイニング手法として決定木（C5.1、C&R tree、CHAID、QUEST）、Neural network、その他（Bayes、Decision list、SVM）の手法を、SPSS社のClementine 12.0を用いて解析した。

【結果】対象患者の初診時および6か月後のHbA1cはそれぞれ $9.0 \pm 1.9\%$ 、 $7.3 \pm 1.4\%$ であり、6か月HbA1c 6.5%の良好な血糖コントロール達成率は38.2%だった。いずれの解析法によっても、罹病期間、初診時の薬物療法の有無、BMI、T-Chol、HDL-C、UA、Crが6か月後のHbA1c値への寄与度が大きいという結果が得られた。6か月後のHbA1c値と各解析法を用いた予測値との相関係数が最も良かったのは、Neural network手法（ $r=0.536$ ）であり、重回帰分析（ $r=0.424$ ）よりも優れていた。また、6か月後の血糖コントロール指標が良、不良のいずれかであるかの2値予測の精度が最も良かったのは決定木（C5.1）（79.8%）であり、二項ロジスティック回帰分析（72.5%）よりも優れていた。

【総括】通常の変量分析は、背景因子とエンドポイントとの間に線形性の関係があると

ということが前提となっている。ところが、糖尿病は多数の因子が複雑に関与している疾患であり、背景因子とエンドポイントは非線形的な関係があるため、そのような関連性の分析により適しているデータマイニング手法がより優れていた結果であったと考えられた。今回はセカンダリーエンドポイントである HbA1c 値について解析したが、糖尿病の複雑性を考えると、死亡や合併症のようなプライマリーエンドポイントの解析についても、データマイニング手法が有用である可能性が高いと考えられる。