

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 8 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2011～2014

課題番号：23790561

研究課題名(和文)因果連鎖オントロジーに基づく糖尿病の病態遷移予測に関する研究

研究課題名(英文)A research on modeling Diabetes Disease Progression based on Causal Chain Ontology

研究代表者

今井 健 (Imai, Takeshi)

東京大学・医学(系)研究科(研究院)・講師

研究者番号：90401075

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：糖尿病患者の将来起こりうる合併症リスクを予測するため、糖尿病病態遷移モデルの構築を目的とした。そのため、(1)糖尿病の病態遷移オントロジーを開発し、(2)実際の患者診療情報のオントロジーへのマッピング結果を時系列で追うことで、(3)大血管障害や細小血管障害を発症するまでの期間を、HbA1c、喫煙歴、家族歴、高血圧、脂質異常などの各種患者状態パラメータで予測する重回帰モデルを得た。

研究成果の概要(英文)：This study aimed to model diabetes disease progression for predicting future risks of diabetes complications. We first built a diabetes disease progression ontology which represent causal chains of abnormal states, and mapped patient's clinical status to the ontology in chronological order. Then we constructed the diabetes disease progression model to predict the length of transition to diabetic macroangiopathy and microangiopathy based on patient's status (e.g. HbA1c, smoking status, family history, hypertension, hyperlipidemia) and multiple regression model.

研究分野：医療情報学

キーワード：オントロジー 糖尿病 病態遷移予測

1. 研究開始当初の背景

近年、世界の糖尿病患者数は著しく増加しており、我が国でも平成 19 年の国民健康・栄養調査では、糖尿病が疑われる人の割合が予備軍含め約 2,210 万人(40 歳以上では 29.0%)に達している。また 2 型糖尿病などの生活習慣病にかかる医療費は 10.4 兆円(2004 年)と約 3 割に達しており、医療費が財政上大きな負担となりつつある我が国にとって糖尿病患者の増加は深刻な問題である。そのため糖尿病の早期発見が必要であると共に、その後の病態遷移、すなわち患者の血糖コントロール状況・各種病態の発現および進展・様々なリスク因子を正確に把握し、将来起こりうる合併症とその可能性、程度を予測するための「病態遷移モデル」の開発が重要な課題となっている。これは将来のリスク情報を患者に提供することで、治療に対するコンプライアンスを高めるためにも非常に重要である。

これまで、海外では糖尿病の診断や合併症のリスクを予測する数理モデルの研究が行われてきている。例えば米国糖尿病学会では糖尿病を含む人体に起こる様々な病態、関連する治療法、医療費等の関係をモデル化し、個々の糖尿病患者における合併症の発症リスクを予測する Archimedes Model などの研究が行われてきた。しかしこれは糖尿病の病態遷移知識を記述したものではなく、網羅性の欠如(記述されている知識の範囲が限定的で他疾患との関係の記述が不十分なこと)、知識記述方式の改善(概念の定義と概念間の意味関係が記述されておらず、電子カルテや他のデータベースとの連携がとれないこと)が課題となっていた。

この問題を解決するための 1 つの方法として、概念の定義を概念間の意味関係を用いて計算機処理可能な形式で記述したデータベース(以下オントロジー)を用いる方式が近年盛んになっている。申請者はこれまで厚生労働省受託研究開発事業「医療情報システムのための医療情報知識基盤研究開発事業(H19-21,22-24)」に参画し、我が国における臨床医学オントロジーの開発研究を行ってきた。これは網羅的な疾患・部位・症状所見概念について意味関係、因果連鎖関係を記述したデータベースであり、電子カルテや各種外部データベースとの連携を目指した記述フレームワークに基づいている。この経験から先行研究の欠点を克服するためには、本課題においてもオントロジー工学に基づいた知識記述フレームワークを活用し、上記のオントロジーを発展させ糖尿病病態遷移モデルの構築に役立てることが必要であるとの着想に至った。

2. 研究の目的

以上の背景のもと、本研究は以下の 2 つの柱を元に、病態の進展や将来リスクを定量的に予測するための、糖尿病患者の病態遷移モ

デルの開発を目的とする。

(1) 糖尿病関連領域の病態遷移オントロジーの開発

既存の臨床医学オントロジーを元に、患者診療情報を記述するのに必要な異常/正常状態、検査項目等の用語概念セットを追加する。また診療ガイドラインや既存の知見から BMI、血圧、LDL、HDL、中性脂肪値等、糖尿病の病態因果連鎖の促進要因を抽出し上記に追加し、糖尿病病態遷移オントロジーを開発する。

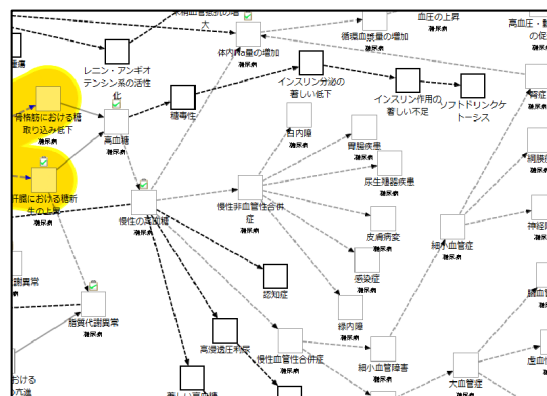
(2) 糖尿病病態遷移モデルの構築

患者診療情報を病態遷移オントロジーにマッピングした結果を経時的に追うことにより、病態進行の促進要因から遷移のしやすさ・推定速度を算出し、将来の病態進展・合併症発生リスクを予測する病態遷移モデルを構築する。

3. 研究の方法

(1) 糖尿病病態遷移オントロジーの開発

既存の臨床医学オントロジーの糖尿病領域の成果を借用し、患者診療情報を記述するための異常状態や検査項目等、本研究の目的に必要な概念を追加した。また病態遷移は下記の図 1 のように、ある疾患において人体に発生する異常状態が次々と次の異常状態を誘発していく因果連鎖の形で、専用のエディタで記述した。構築にあたっては、研究協力者である糖尿病専門医の助言を得て行った。



(図 1. 糖尿病の病態遷移の記述例)

(2) 患者状態のオントロジーへのマッピング

上で開発したオントロジー中の各異常状態ノードに患者状態をマッピングした時系列情報を得るため、東京大学医学部附属病院の糖尿病代謝内科に入院歴を持つ患者の退院サマリーを収集した。検索条件は、電子カルテ入力が始まり病院情報システム内のデータベースに記録が存在する 2008 年 10 月から 2012 年 3 月 31 日までの間に同科に入退院歴を持ち、また病名に糖尿病という文字を含む患者である。収集した退院サマリーデータは連結可能匿名化を施し、以降の処理を行った。

次に、得られた退院サマリーから、病態遷

移オントロジー中の異常状態ノードにマッピングするための情報を抽出するルールを専門家の協力のもと策定した。退院サマリーから抽出した項目は、「入力年月日、患者基本情報、糖尿病発症年、診断名、家族歴、喫煙歴、インスリン投与開始時、糖尿病およびその他内服薬、LDL、HDL、TG(中性脂肪)、空腹時血糖値、HbA1c、糖尿病性腎症の有無・重症度、網膜症の有無・重症度、末梢神経障害の有無・部位、大血管障害の有無と部位(脳血管/冠動脈/その他)、蛋白尿、AB/C比、sCPR、sCr、GFR、各種既往症と年齢、各種合併症(悪性腫瘍/心疾患/肝障害/高血圧/脂質異常症等)」などである。これらの情報を策定されたルールに従って、退院サマリーの文章から人手で抽出した。時系列変化の単位を月とするため、糖尿病の診断時点(年齢)はその年の中間時点(7月)の月とした。

(3) 糖尿病病態遷移モデルの開発

上で得られた結果から、各患者が病態遷移オントロジー中のノード(異常状態)にいつ到達したかを知ることができる。糖尿病発症年が得られている患者については、その時点で糖尿病の定義となる異常状態ノード(例えば2型糖尿病患者であれば「4.研究成果」図3の領域Aのノード)が起きていることになる。また心筋梗塞の発症であれば、領域Fの「虚血性心疾患」ノードに対応づけられる。患者状態の時系列変遷を追い、ある入院時に前回入院時より増悪していた場合、そこまでの経過時間(単位:月)とその時の各種患者状態パラメータを取得した。また、各異常状態(腎症、網膜症、神経障害、冠動脈疾患)が発症に至るまでの経過時間(月)が家族歴、喫煙歴、BMI、HbA1c、HT(高血圧)、HL(脂質異常)などの患者状態パラメータで説明できると仮定し、重回帰分析により次の回帰式を得た。

発症に至るまでの経過時間(月)

$$= (A_i \cdot P_i) + C$$

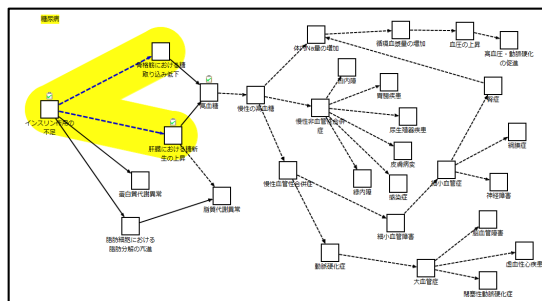
(A は重み係数, P は各患者状態パラメータ, C は定数項)

4. 研究成果

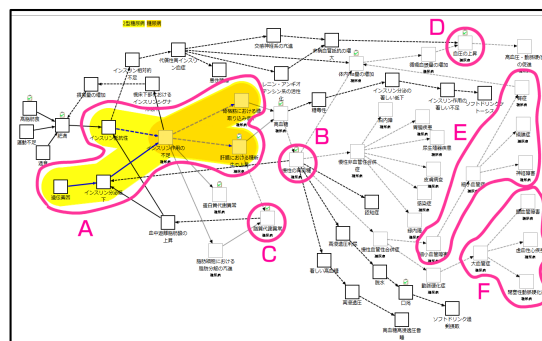
(1) 糖尿病病態遷移オントロジー

構築した糖尿病病態遷移オントロジーは1型糖尿病、2型糖尿病、その他各種サブタイプを含む42の疾患からなり、それらは“is-a関係”により階層的に整理されている。最上位の「糖尿病」の病態遷移オントロジーを図2に示す。黄色で囲まれている部分は「インスリン作用の不足」「骨格筋における糖取り込み低下」「肝臓における糖新生の上昇」の3つのノードからなる部分因果連鎖であり、これはこの疾患の定義を表す。つまり全患者に共通して起こるものである。他の部分は患者によっては見られたり見られなかったりする異常状態の連鎖である。糖尿病

の各サブタイプは、この病態連鎖を継承し、下位疾患においては、「上位疾患の病態連鎖に新しいノードを追加する」、あるいは「特定のノードの意味をより詳細化する」という2つの操作の組み合わせにて作成される。図3に、下位疾患である「2型糖尿病」の病態遷移オントロジーを示す。上位疾患である糖尿病の連鎖に比べ、定義領域(黄色の部分)に「インスリン抵抗性」「インスリン分泌低下」などのノードが追加されており(A)、2型糖尿病と診断された時点でこれらの異常ノードが患者に起こっていることを意味する。またその後次々と起こる異常状態の連鎖が矢印で表現されており、例えば「慢性の高血糖(B)」「脂質代謝異常(C)」「血圧の上昇(D)」「腎症・網膜症・神経障害といった各種細小血管障害(E)」「脳血管障害・虚血性心疾患・閉塞性動脈硬化症といった各種大血管障害(F)」などが存在する。これらオントロジー中のノードは、次の段階で患者の時系列診療情報のマッピング対象となる。



(図2. 糖尿病の病態遷移オントロジー)



(図3. 2型糖尿病の病態遷移オントロジー)

(2) 患者診療データのマッピング

本研究で取得した患者診療情報は全体で1,183名であったが、電子カルテ稼働後の2008年10月~2012年3月と収集期間が短かったため、糖尿病代謝内科で複数回入院している患者は566名で、退院サマリーから抽出された診療データ取得点のはのべ1128回であった。そのうち大血管障害(図3のF領域)、もしくは細小血管障害(図3のE領域)まで起こっている患者は267名、またこのうち2型糖尿病が234名を占めていた。また糖尿病発症時点が分かっている患者は150名存在した。

(3) 糖尿病病態遷移モデル

次に、遷移モデルの計算のため、複数回の

観測点において糖尿病の病態が進行しているケースのみを抜粋した。病態遷移中の細かな概念にはマッピングできないことも多く、十分なデータ数が確保できないため、2型糖尿病の患者だけに絞り、糖尿病発症、各細小血管障害の発症（腎症、網膜症、神経障害）、各大血管障害の発症（脳血管障害、冠動脈障害）に大別し、糖尿病発症から5個の合併症のどれかが発症する度に「増悪した」と見なし、そこに至る経過時間(月)と発症時の各種パラメータを取得した。結果、病態遷移としては216名の2型糖尿病患者について586パスが得られた。

全てのパスについて説明変数となるパラメータが揃っている訳ではないため、共通して入力されている頻度が高く、また臨床的にもリスク要因として挙げられるBMI、HbA1c、糖尿病家族歴、喫煙歴、HT（高血圧症）、HL（脂質異常症）を説明変数とし、重回帰分析により5カテゴリの合併症を発症するまでの期間についての各々の回帰式を得た。

例として、大血管障害の1つである「冠動脈疾患合併状態」への移行(月数)についての重回帰分析結果を表1に示す。変数選択の過程において、 $p > 0.2$ のものは回帰式から除去されている。また偏回帰係数が負のものは冠動脈疾患発症までの月数を減らすため、より早く冠動脈疾患の合併状態に移行してしまうことを意味する。喫煙歴、脂質異常症については有意に移行を早めてしまう要因であるという結果が得られた。

変数	偏回帰係数	P値
喫煙歴	-92.46	* 0.044
BMI	5.18	0.170
脂質異常症	-178.64	* 0.005
定数項	220.73	0.051

(表1: 冠動脈疾患への移行速度に関する重回帰分析結果。*は $p < 0.05$ で有意)

本研究の限界として、電子的に収集可能な期間のデータに限定したため、退院サマリーに糖尿病発症年齢が記載されている場合でも発症以降電子カルテ稼働までの期間の詳細については不明である。そこで、2008年10月～2012年3月末の期間内で遷移したケース(35パス)のみを集め分析を行った結果が表2である。データ数が少ないため、3つの細小血管症をまとめ、「何らかの細小血管障害の発症」への移行速度とした。有意なものとしてHbA1cが高いと移行期間を短くするという結果が得られた。

変数	偏回帰係数	P値
BMI	0.93	0.067
HbA1c	-2.42	* 0.006
脂質異常症	9.19	0.059
定数項	1.02	0.940

(表2: 細小血管障害への移行速度に関する重回帰分析結果。*は $p < 0.05$ で有意)

(4) 成果の位置づけ・今後の展望

現在電子カルテの導入が進み、膨大な診療情報が電子的に蓄積されつつある。また日本医療情報学会・日本糖尿病学会の合同委員会で、糖尿病患者について必ず記載することを目標とした糖尿病ミニマム項目セットも策定され、基本的な情報が全て入力された糖尿病患者の診療情報が今後全国的に蓄積されていくことが予想される。またSS-MIX2などの標準ストレージ規格の推進などにより多施設間でのデータ共有に向けての環境も整いつつある。本研究で得られた遷移モデルはデータ件数の問題から限定的な結果となったが、同じ手法を多施設の巨大な臨床データベースに対して適用することでより精緻なモデルが得られるだろうことは明らかである。本研究は国内で共通に参照される医学知識基盤である臨床医学オントロジーを基にし、糖尿病の病態遷移を記述するためのフレームワークを示したという点で、貴重なパイロット研究であると位置づけられる。特に視覚的にも分かりやすい病態遷移マップ上に患者の現在の状態をプロットすることで、現在の状態・過去の状態・そして将来さらに進行する危険がある状態を一覧して把握することができ、医師にとって有用であるのみならず、患者にとっても治療に対するコンプライアンスを高める意味で有益な情報提供をすることが可能であると考えられ、これも本研究の大きな特徴である。

今後の展望としては、まず、より大規模・長期間のデータに対してこの手法を適用することが挙げられる。今回は電子データが存在する2008年途中以降を対象としたため、時系列データ取得点の数が十分では無かったが、しかし本研究終了時点2015年において既に本研究で利用したデータの2倍の期間のデータが蓄積されており、今後ますます長期のデータの利用が容易になっていくと考えられる。次に、病院情報システム内のデータから自動的に異常状態へマッピングするための機械学習の手法の開発も重要である。今回は人手でオントロジー内のノードへのマッピングを行ったが、今後対象とするデータ量が多くなった場合、全て人手でマッピングを行うのは困難であると予想される。そこで各種検査結果やカルテ内の記載から「ある特定の異常状態が起こっている」と自動で高精度に判断する手法の開発は重要な課題である。さらに今後、糖尿病のみならず、慢性腎臓病など他の慢性疾患についても同様の手法を検討する予定である。

5. 主な発表論文等

〔雑誌論文〕(計1件)

Mizocughi R, Kozaki K, Kou H, Yamagata Y, Imai T, Waki K, Ohe K. River Flow Model of Diseases. Proc. Of International Conference on Biomedical Ontology (ICBO), 査読有, 2011:pp.63-70.

〔学会発表〕(計 5 件)

Yamagata Y, Kou H, Kozaki K, Mizoguchi R, Imai T, Ohe K. Ontological Model of Abnormal States and its Application in the Medical Domain. 4th International Conference on Biomedical Ontology (ICBO2013), Jul.8-9,2013, Montreal, Qc, Canada.査読有.

Kozaki K, Kou H, Yamagata Y, Imai T, Ohe K, Mizoguchi R. Browsing Causal Chain in a Disease Ontology. 11th International Semantic Web Conference (ISWC2012), Nov.11-15,2012, Boston, USA. 査読有.

今井 健, 古崎晃司, 国府裕子, 山縣友紀, 溝口理一郎, 大江和彦. 我が国における臨床医学オントロジーの現状と今後の展望. 第61回 SLUD 研究会・医療情報コミュニケーションシンポジウム(人工知能学会 SIG-SLUD-B003), 2011/10/2, 東京大学大学院情報学環(東京都文京区).

古崎晃司, 国府裕子, 山縣友紀, 今井 健, 大江和彦, 溝口理一郎. 臨床医学オントロジーにおける疾患連鎖モデルの考察. 第23回セマンティックウェブとオントロジー研究会(人工知能学会 SIG-SWO), 2011/6/22, 浜名湖かんざんじ荘(静岡県浜松市).

山縣友紀, 国府裕子, 古崎晃司, 今井 健, 大江和彦, 溝口理一郎. 疾患オントロジーにおけるバランス複合異常と汎用異常連鎖に関する考察. 人工知能学会第25回全国大会, 2011/6/3, いわて県民情報交流センター(岩手県盛岡市).

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

6. 研究組織

(1)研究代表者

今井 健 (IMAI, Takeshi)

東京大学・大学院医学系研究科・講師

研究者番号：90401075

(2)研究協力者

大江和彦 (OHE, Kazuhiko)

東京大学・大学院医学系研究科・教授

脇 嘉代 (WAKI, Kayo)

東京大学・大学院医学系研究科・

特任准教授