

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 17 日現在

機関番号：16401

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24590615

研究課題名(和文)医療データに基づく動的病態力学の構築

研究課題名(英文)Construction of Dynamic Patho-Mechanics Based on Medical Informataion Database

研究代表者

奥原 義保 (OKUHARA, Yoshiyasu)

高知大学・教育研究部医療学系・教授

研究者番号：40233473

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：理論的な考察と、病院情報システムに蓄積されたデータから生活習慣病等の病態推移を記述・予測するモデルを構築し、病院情報システムのデータを活用した動的病態力学の創生を目指した。

実際の診療データにおける欠損値の存在による予測誤差の増大を、粒子型フィルタを用いることで解決し、長期予測処理を可能とした。また、データのみから推定するのではなく、適切な数理モデルを組み込んだ階層的ベイズモデルの適用が病態の推移予測に有望であることを示した。さらに、迅速な予測が必要とされる場合においても、限られたデータから有用な予測が可能であることを示し、病院情報システムのデータを活用した動的病態力学の可能性を示せた。

研究成果の概要(英文)：Combining theoretical considerations and the data stored in the hospital information system, we tried to build a model to describe and predict the disease state transition such as lifestyle-related diseases. With this approach, we aimed at creation of dynamic patho-mechanics utilizing the data of the hospital information system.

The increase of the prediction error due to the presence of missing values in the actual clinical data, has been solved by using a particle filter, and enables long-term prediction process. Also, instead of estimating from the data alone, the application of hierarchical Bayesian model that incorporates the appropriate mathematical model has shown to be promising for transition prediction of the condition. Furthermore, even when the rapid prediction is required, we have shown examples of useful predictions from limited data. With these results, we have shown the possibility of a dynamic patho-mechanics utilizing the data of the hospital information system.

研究分野：医療情報学

キーワード：病態推移予測 病院情報システム 予防医学

1. 研究開始当初の背景

急速な高齢化が進む我が国においては、加齢と食生活などの生活習慣が関係する生活習慣病の増加が深刻な問題になっている。生活習慣病は脳血管系の疾患や心血管系の疾患につながる深刻な疾病である反面、生活習慣にかかわる主要な危険因子を除く努力をすれば予防可能な疾患であるとも考えられる。

このため、従来のように発症後治療するという後追いの医療ではなく、発症を予測し対策を講ずることによって未然に防ぐ予見性の医療が重要であると考えられる。こうした予見性の医療のためには、対象となる疾患の病態推移について理解し、発症前から発症に至るまでを記述・予測できる方法を開発して、医療提供者が将来の推移を予測できるようにすることが重要であると考えられる。

経時的に病態の推移を記述するためには、検査値等観測量のシステムティックな推移をデータから推測し、病態の推移を記述するモデルを構築するという手順も考えられるが、観測データだけから、システムティックな時間変動部分を記述する決定論的メカニズムを推測するためには、詳細な粒度を有する膨大な量のデータが必要である。

ところが、外来もしくは定期健診での受診者を対象としたデータを用いてモデル構築を行う場合、年単位における変動予測が必要となってくるが、通常の診療では詳細な検査が行われていないため、データが存在せず適用できない、定期的に頻繁なデータ取得が行われているとは限らないため、欠損値だけのデータを適用せざるを得ず、実装が非常に困難となる。そのため、通常の診療で実施される検査の中から疾病初期において変動する項目を見つけ出して用い、長期間解析を行う際に発生する欠損値及びノイズに対して頑健なモデルの構築を行うことが必要となる。

そのためには、観測データ以外の情報からシステムティックな時間的変動部分を決定する決定論的メカニズムの部分の構築することが必要だと考えられる。そこでは、理論的アプローチと実際の医療データによる検証との双方向のアプローチの融合が必要である。

2. 研究の目的

本研究の目的は、理論的な考察と、病院情報システムに蓄積された医療データから生活習慣病などの病態推移を記述・予測するモデルを作成し、過去から現在にかけての一般的な臨床検査データから患者の将来の病態推移の予測を可能にする方法の構築、言わば動的病態力学を構築することである。このことにより、発症後治療するという後追いの医療ではなく、発症を予測し対策を講ずることによって未然に防ぐ予見性の医療が実現すると期待される。このため、本研究では、

(1) 糖尿病初期に至る状態を予測するために血糖及びインスリン以外に糖尿病初期を把握できる検査項目を用い、長期間解析を行う際に発生する欠損値及びノイズに頑健な**状態空間モデルによる糖尿病の病態推移予測モデルの構築**、

(2) 糖尿病を対象に、検査データが十分に揃っていない患者の病態推移予測を可能とする**階層ベイズモデルによる病態推移予測モデルの構築**、

(3) **糖尿病以外の疾患についての病態推移予測モデルの構築**、
を行った。

これらにより、病院情報システムのデータを活用した動的病態力学の創生を目指した。

3. 研究の方法

本研究では、高知大学医学部附属病院において蓄積しているデータを解析対象データとした。附属病院には、1981年からのデータが蓄積されており、総患者数28万人、10年以上の来院データを持つ患者数3万2千人の長期間かつ大規模なデータである。これらのデータは、連結不可能な匿名化処理を施された上で、研究・教育用データウェアハウスRYOMA(Retrieval sYstem for Open Medical Analysis)に蓄積されている。

(1) 状態空間モデルによる糖尿病の病態推移予測

糖尿病を対象とした状態空間モデルによる病態推移予測では、第一段階として、RYOMAのデータに基づき、糖代謝機能の観測データであるHbA1cを他の検査項目と連動させた予測モデルの構築を行った。そのため、糖尿病発症前の一般的な検査項目データの変動を抽出し、どの検査項目をモデルに組み込むことが適切かについて解析を行った。具体的には、まず、蓄積データ中の75gブドウ糖負荷試験(OGTT)データから、対象群を血糖値、インスリン値、インスリン抵抗性指数(HOMA-IR)値によって糖尿病群(DM)、耐糖能異常群(IGT)、及び正常群(Normal)の3群に分類した。これら3群につき、糖尿病発症より以前の影響を評価するため、OGTTを実施した日を基準として、OGTT実施日から前10年以内のデータの集計を行った。集計の対象とした検査項目は、TG(中性脂肪)、HDL(高比重リポたんぱく)、LDL(低比重リポたんぱく)、PLT(血小板)、ALT(アラニンアミノトランスフェラーゼ)、AST(アスパラギン酸アミノトランスフェラーゼ)、 γ -GT(ガンマ・グルタミール・トランスペプチターゼ)、ChE(コリンエステラーゼ)、RBC(赤血球)、WBC(白血球)である。これらの各検査項目の集計値につき、群間の有意差等を調べた。

その結果、長期データによる群間の統計的に有意な差が見られたこと、分散が小さいことを基準に、モデルに組み込むパラメータとしてはHDLが適切であると判断した。このた

め、本研究では糖尿病を発症する前に HDL の低値への変動が存在すること、低 HDL 値の患者において HbA1c 値が上昇する可能性があることを仮定した。その仮定に基づき上述の予測確率モデルを構築した。

本研究では、病院情報システムに蓄積されているデータに基づき糖代謝異常に関しての長期予測モデルを脂質代謝異常と連携させて構築することを目指したが、これらのモデルのパラメータを診療データから取得することは困難であり、HbA1c と HDL の関係を直接定義することは困難であるため、各検査値を推定するための糖代謝及び脂質代謝機能を、直接観測はできないが、状態の時間変化を決定する動的システム(内部モデル)として捉えた。これらの代謝能力の変動に伴って、直接観測可能な検査の結果が変化すると考え、その検査結果に基づいて内部モデルの同定を行った。これらの観測可能なモデルを観測モデルと定義した。この2つのモデル対を「状態空間モデル」として構築した。この状態空間モデルを予測モデルとして扱い、「観測モデル」における各検査値は予測データとなる。第一段階の解析結果とそれから導かれた仮説に基づき、リスク因子となる HDL の変動を定義する内部モデルである脂質代謝異常が内部モデルの糖代謝異常よりも先に発生していること及び低 HDL を生成する脂質代謝異常の状態では HbA1c が上昇しやすいように糖代謝異常が変動することを、脂質代謝異常モデルから糖代謝異常モデルに影響を与える形で記述し、それぞれの内部モデルが観測モデルである HDL 及び HbA1c に対応している状態空間モデルとした。

内部モデルの脂質代謝異常、糖代謝異常は高い値ほど悪化を示すよう定義し、内部モデル間の関係は、HDL を定義する内部モデルの脂質代謝異常が HbA1c を決定する内部モデルである糖代謝異常に影響する形でモデルを構築した。観測モデル及び内部モデル間の関係は、これまでに RYOMA に蓄積してきたデータに基づき、ファジィ推論によって、その変動量モデルを定義した。糖代謝異常から計算される HbA1c を表す観測モデルの式は、RYOMA のデータから得られた時系列分布に基づく指数関数モデルとして定義した。

内部モデル及び観測モデルを確率分布として表現するが、ノイズの影響を考慮する必要があるため、この確率分布をパラメトリックモデルとして仮定できず、各モデル分布を定式化するのが困難である。そのため、粒子型フィルタ(パーティクルフィルタ)^[1]によって、構築モデルの実装を行った。モデル評価は 116 人の実診療データに対し適用し、実際の HbA1c 及び観測モデルにて予測された HbA1c の二乗平均平方根誤差(RMSE)にて行った。

なお、最終年度にはこのモデルでは明確に考慮していないインスリン分泌能の要素を取り入れるため、インスリンを測定していな

い場合でも血糖値の時間変化データのみからインスリン値を推定できる手法の開発も試みた。そこでは、OGTT における血糖値とインスリン値を記述する数理モデル^[2]を用いて、対象とする患者以外の OGTT 実施患者集団のデータにおいて個人ごとの最適パラメータを求め、そのパラメータを用いて計算した個々の血糖値時間変化が、対象とする患者の血糖値の時間変化に近い患者を複数選択し、それら患者のパラメータセットに対して、正規ノイズを加えてそれぞれ複数のパラメータセットを生成、さらに、生成したパラメータセットに対して複数のインスリン初期値を与え、これらの全組み合わせに対して血糖値の時間変化を計算し、対象患者の血糖値実測値との二乗誤差が小さい上位のインスリン計算値をインスリン予測値として評価した。

(2) 階層的ベイズモデルによる糖尿病の病態推移予測

背景でも触れたように、日常的な診療においては糖尿病やその疑いのある患者に対してインスリンの値は測定されない。このように検査データが十分に揃っていない患者の病態推移予測を可能とするため、(1)とは別の方法として、糖代謝を記述する最低限の要素を取り入れた連立微分方程式からなる数理モデルを構築して決定論的に血糖値およびインスリン濃度の時間変化を記述するモデルとし、階層ベイズモデルに取り入れることにより、検査データが十分に揃っていない患者の糖尿病の病態進展段階を推定するモデルを構築した。数理モデルとしては、よく知られた minimal model^[3]に基づいたモデルを用いた。

この階層ベイズモデルを、leave-one-out 交差確認法によって評価した。評価のための比較は、数理モデルを含まず、データのみから単純な Bayes 推定によって推定した分類方法的な中程度の間で行った。

さらに、このモデルを RYOMA に蓄積されているデータに適用し、糖尿病の病態進展段階を推定する手法を構築した。

(3) 糖尿病以外の疾患に対する病態推移予測

妊婦の CTG データから胎児のリスクを予測するモデルの構築

ノイズを多く含む CTG (CardioTocoGram : 胎児心拍陣痛図) データに基づいて胎児のリスクレベル評価を行うアルゴリズムを構築した。まず欠損値の多い胎児心拍数データについて粒子フィルタ法によって補完を行った。次に、補完したデータを用いて、一過性除脈のパターンを抽出した。さらに、一過性除脈のパターンと判断され抽出されたデータにつき、「日本産婦人科学会診療ガイドライン」^[4]に従ってレベル評価を行った。パターン抽出とレベル評価においてはファジィ

推論の手法を用いた。このアルゴリズムを高知大学医学部附属病院に蓄積されている CTG データを用いてレベル分けをし、分娩時の臍帯血の pH 値との関係で評価した。

入院時の情報に基づく急性腎障害発症の予測モデル

ロジスティック回帰モデルを用い、RYOMA に蓄積されたデータのうち 1981 年～2012 年における、18 歳以上の入院患者のデータから、入院時の患者基本情報、検査情報、投薬や手術の予定など入院時に入力可能な情報を説明変数として入院中の急性腎障害(AKI)の発症予測モデルを構築した。なお、AKI 発症は国際的な診断基準である KDIGO2012 ガイドライン^[5]に基づき、クレアチニン値によって判断した。

さらに、モデルの適応可能性を検討するため、モデル作成には用いていない 2013 年 1 月から 2015 年 1 月の入院患者のデータを新規データとしてモデルを適応し予測シミュレーションを行い、感度、特異度を求めた。

4. 研究成果

(1) 状態空間モデルによる糖尿病の病態推移予測

116 人のデータに対し予測を行い、検査初回日以外の予測データ (285 件) について、予測 HbA1c 値と観測 HbA1c 値の比較を行い、RMSE が 0.25 となった。誤差の絶対値の最大及び最小値は 1.21 及び 0.00 となった。図 1 に予測 HbA1c と観測 HbA1c との関係図を示す。

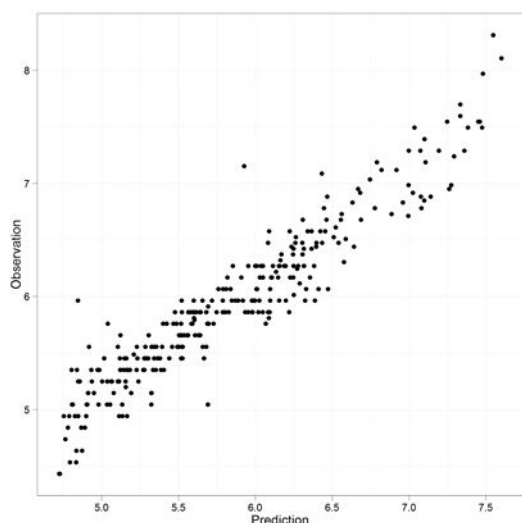


図 1

横軸が予測値であり、縦軸が観測値となっている。ピアソンの積率相関係数の有意性検定において有意性が示された ($r=0.95$)。入力データは欠損値が存在している時系列データであり、また様々な外乱ノイズ等の影響を考えると、得られた RMSE 値は妥当性を示す十分な値であると考えられる。また、予測値及び観測値の相関も極めて高く構築モデルによる予測が適切に行われていることを示している。

実際の診療データにおいて欠損値が存在し、そのことによって予測誤差が増大することが長期間予測処理を困難としている理由の一つであるが、本研究では粒子型フィルタを用いることで処理継続が可能となっている。さらに、予測データを任意の確率分布として表現することにより、日内変動などのノイズに柔軟に対応することが可能となり、結果として欠損値が多く含まれる 10 年以上のデータに対しても適用可能であることを示している。これらの結果から、健診等において使われている検査項目だけでも適切に関係を記述し適用することにより、診療上妥当な予測が行えると考えられる。

(2) 階層的ベイズモデルによる糖尿病の病態推移予測

RYOMA に蓄積されているデータからは、Glucose disposal と Insulin secretory の機能が互いを補い合うように働いている様子を観測することが出来た。

データのみから患者の病態進展度を推定する単純な Bayes 推定モデルでは、75gOGTT 検査における 2 時間後の血糖値を推定する分布関数が正常状態、境界領域の患者群では低血糖状態である 50mg/dL 以下の領域まで尾を引くという、実際には考えられない分布になった(図 2A)のに対して、数理モデルを取り入れたモデルでは 50mg/dL 以下の領域が充分抑制されている現実的な分布を得た(図 2B)。

図 2A 単純なモデル
横軸は 2 時間後の血糖値

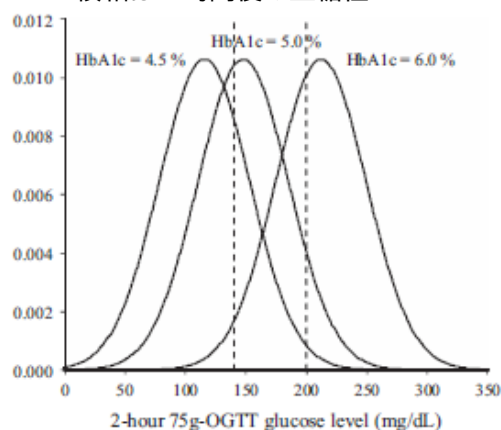
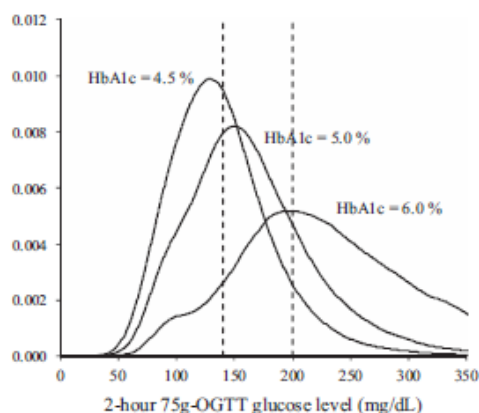


図 2B 数理モデルを取り入れたモデル
横軸は 2 時間後の血糖値



また、leave-one-out 交差確認法による比較の結果、単純な Bayes 推定モデルでは、正常状態の正解率が 56%であったのに対して、本モデルでは 84%という高い正解率であった。一方、空腹時血糖 $\geq 126\text{mg/dL}$ と OGTT120 分の血糖値 $\geq 120\text{mg/dL}$ で判定された糖尿病状態の正解率は、単純モデル 100%、本モデル 83%という結果であった。

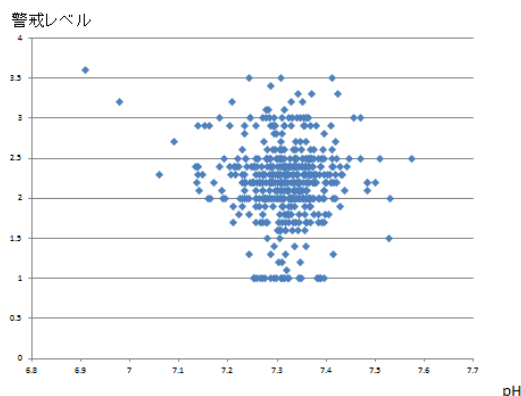
これらの結果は、データのみから患者の病態進展度を推定するだけでなく、適切な定量的数理モデルを組み込んだ階層的ベイズモデルの適用が病態の推移予測に有望であることを示している。

(3) 糖尿病以外の疾患に対する病態推移予測

妊婦の CTG データから胎児のリスクを予測するモデルの構築

図 3 に示されるように、胎児アシドーシスに分類される臍帯血 $\text{pH} < 7.2$ を示すほとんどのケースに対してレベル 2 (監視の強化か保存的措置の施行) 以上の評価となり、 pH 値が低くなるにつれて自動判定の警戒レベルが高くなる結果となった。実際には pH 値が 7.2 以上であっても警戒レベルが高く評価された例も多かったが、CTG による判定はもともと胎児機能不全の診断特異度は低いと言われており、その限界が反映しているものと考えられる。こうした限界はあるものの、本モデルは、CTG データからのスクリーニング的な胎児リスク自動判定や、大量の CTG データから医学生の教育に有用なパターンを自動抽出し、医学教育の教材作成に用いるなどの応用が考えられる。

図 3



入院時の情報に基づく急性腎障害発症の予測モデル

最終的に得られたモデルの判別能は、 $\text{AUC} = 0.867$ であり、 $\text{cut off} = 0.141$ において、感度 79.1%、特異度 78.9%であった。

新規データ (2013 年 1 月から 2015 年 1 月までの患者データ) に対してモデルによる予測シミュレーションを行った結果は、感度 80.3%、特異度 77.1%であった。

本研究で得られたモデルでは、尿酸、尿素窒素など、先行研究^[6]では、解析対象とされ

ていなかった臨床検査項目や手術による影響が考慮されている。入院時に実施頻度の高い検査項目を説明変数に組み込むことで、先行研究で示されている AKI 発症予測モデルの $\text{AUC} 0.78$ よりも予測能の高い AKI の発症予測モデルが構築できた。

本研究で構築したモデルに組み込まれた説明変数は、いずれも入院時に入力可能なものである。こうした予測モデルを病院情報システムに組み込めば、入院初期の情報からリスクの高い患者に関するアラートを上げることで、早期に対応できるようになると考えられる。

本研究では、理論的な考察と、病院情報システムに蓄積された医療データから生活習慣病などの病態推移を記述・予測するモデルを構築し、病院情報システムのデータを活用した動的病態力学の創生を目指した。

「状態空間モデルによる糖尿病の病態推移予測」では、実際の診療データにおける欠損値の存在とそれによる予測誤差の増大という問題を、状態空間モデルに対して粒子型フィルタを用いることで解決し、長期間予測処理を可能とした。また、予測データを任意の確率分布として表現することにより、日内変動などのノイズに柔軟に対応することが可能となった。「階層的ベイズモデルによる糖尿病の病態推移予測」では、データのみから患者の病態進展度を推定するのではなく、適切な定量的数理モデルを組み込んだ階層的ベイズモデルの適用が病態の推移予測に有望であることを示した。

これらにより、病院情報システムのデータを活用した動的病態力学の創生において課題となる、限られた観測値と不可避なノイズの存在による長期予測の困難さを解決する方策を見出したと考える。

また、「糖尿病以外の疾患に対する病態推移予測」では、

妊婦の CTG データから胎児のリスクを予測するモデルの構築

入院時の情報に基づく急性腎障害発症の予測モデル

という、短期間に迅速な予測が必要とされる場合においても、粒子フィルタ法、ファジイ推論、ロジスティック回帰モデルなどを適宜用いて、限られたデータから有用な予測が行える例を示し、病院情報システムのデータを活用した動的病態力学の可能性を示せたと考える。

我々は、今後も病院情報システムのデータを活用した動的病態力学の確立を目指した研究を継続する予定である。

<参考文献>

- [1] Kitagawa G., Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models, Journal of Computational and Graphical Statistics,

1996, 5(1), 1-25

- [2] De Gaetano A, Panunzi S, Matone A, et al: Routine OGTT: a robust model including incretin effect for precise identification of insulin sensitivity and secretion in a single individual, PLoS One, 8(8): e70875, 2013.
- [3] Bergman RN, Ider YZ, Bowden CR, Cobelli C. Quantitative estimation of insulin sensitivity. Am J Physiol. 1979 Jun;236(6):E667-77.
- [4] 日本産婦人科学会「婦人科診療ガイドライン 産科編 2014」: http://www.jsog.or.jp/activity/pdf/gl_sanka_2014.pdf
- [5] KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney International, VOL2, SUPPLEMENT 1, 2012(http://www.kdigo.org/clinical_practice_guidelines/pdf/KDIGO%20AKI%20Guideline.pdf)
- [6] Matheny ME, Miller RA, Ikizler TA, Waitman LR, Denny JC, Schildcrout JS, Dittus RS, Peterson JF. Development of Inpatient Risk Stratification Models of Acute Kidney Injury for Use in Electronic Health Records. Medical Decision Making. 2010; 30:639-650.

5. 主な発表論文等

[雑誌論文](計2件)

Y. Hatakeyama, H. Kataoka, N. Nakajima, T. Watabe, S. Fujimoto, Y. Okuhara, Prediction Model for Glucose Metabolism Based on Lipid Metabolism. Methods of Information in Medicine.(査読有) Vol. 53,357-363,2014.DOI:10.3414/ME14-01-034

Teruaki Watabe, Yoshiyasu Okuhara, Yusuke Sagara, A hierarchical Bayesian framework to infer the progression level to diabetes based on deficient clinical data. Computers in Biology and Medicine. (査読有) Vol. 50,107-115, 2014.DOI:10.1016/j.compbiomed.2014.04.017

[学会発表](計5件)

皇山豊、片岡浩巳、中島典昭、渡部輝明、奥原義保、モンテカルロ手法による数理モデルを利用したインスリン値の変動推定、第19回日本医療情報学会春季学術大会、2015年6月12日、仙台国際センター(宮城県・仙台市)

Yutaka Hatakeyama, Hiromi Kataoka, Noriaki Nakajima, Teruaki Watabe, Yoshiyasu Okuhara, Level evaluation

system for Cardiotocography, Joint 7th International Conference on Soft Computing and Intelligent Systems and 15th International Symposium on Advanced Intelligent Systems, TP4-2-3-(1),2014 Dec,3-6 Kita-Kyushu(Japan).kitakyushu International Conference Center.

三木 健生、片岡 浩巳、皇山 豊、堀野 太郎、寺田 典生、奥原 義保、入院患者における急性腎障害(AKI)発症の事前予測モデルの構築、第57回日本腎臓学会総会、2014年7月5日、パシフィコ横浜(神奈川県・横浜市)

渡部輝明、中島典昭、奥原義保、相良祐輔、欠損検査データに基づいた病態推移の推定とその階層ベイズによる数学的枠組み、日本医療情報学会中国四国支部第30回中国四国医療情報学研究会、2013年3月13日、ウエルピア伊予(愛媛県・伊予市)

皇山豊、片岡浩巳、中島典昭、渡部輝明、奥原義保、脂質代謝を考慮した糖尿病における病態予測モデルの構築、第16回日本医療情報学会春季学術大会、2012年6月1日、函館国際ホテル(北海道・函館市)

6. 研究組織

(1)研究代表者

奥原 義保(OKUHARA, Yoshiyasu)
高知大学・教育研究部医療学系・教授
研究者番号: 40233473

(2)研究分担者

皇山 豊(HATAKEYAMA, Yutaka)
高知大学・教育研究部医療学系・準教授
研究者番号: 00376956

渡部 輝明(WATABE, Teruaki)
高知大学・教育研究部医療学系・講師
研究者番号: 90325415

片岡 浩巳(KATAOKA, Hiromi)
高知大学・教育研究部医療学系・助教
研究者番号: 80398049

中島 典昭(NAKAJIMA, Noriaki)
高知大学・教育研究部医療学系・助教
研究者番号: 00335928