

令和 2 年 5 月 29 日現在

機関番号：20101

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K09062

研究課題名(和文) データマイニング手法による健診受診集団の保健事業立案のための効果的階層化の試み

研究課題名(英文) Establishment of effective strategy for classification of individuals who need to receive health guidance using data mining technique

研究代表者

大西 浩文(Ohnishi, Hirofumi)

札幌医科大学・医学部・教授

研究者番号：20359996

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：地域住民健診受診者のデータを用い、決定木分析とクラスター分析を応用することで、健診結果に基づく生活習慣病罹患リスクと、生活習慣や保健サービスへの期待等の対象者の特徴に基づいて、生活習慣病予防対策のための健診受診者集団の効果的な階層化を試みた。将来の高血圧や糖尿病の罹患に対する危険因子の保有状況と特定健診・特定保健指導の流れを考慮し、かつ食塩摂取量や保健指導の希望の有無など対象者の特徴が近い集団に分類して優先順位をつけることが可能となり、市町村の保健従事者が少ない資源投入で効果的な生活習慣病予防対策を実施できる可能性が示唆された。

研究成果の学術的意義や社会的意義

市町村国保がデータヘルス計画の中の保健事業を立案する上で、被保険者集団を疾病罹患リスクや、年代や生活スタイル、生活習慣改善意欲などに基づいてターゲットを絞ることは重要であるが、どのような方法で効果的に対象を分類・階層化するかの方法論はまだ確立していない。本研究によって、過去に蓄積された健診データをデータマイニングの手法を用いて分析することで、受診者の健診結果に基づき将来の疾病罹患リスクや対象特性の近い集団に効果的に分類・階層化し、市町村の保健従事者の限られた人的・経済的資源においても効果的な保健事業が運営できる可能性が示唆された。

研究成果の概要(英文)：We attempted to effectively stratify groups of health checkup examinees for lifestyle-related disease prevention by applying a decision tree analysis and a cluster analysis in our community-based cohort study. We were able to classify health checkup examinee into several groups with similar characteristics of the subject such as risk for future occurrence of hypertension and type 2 diabetes from the results of longitudinal analyses and one's current lifestyle including diet and demand for health service. It was suggested that health workers in local governments could implement effective lifestyle-related disease prevention with less human resource when using these classification methods.

研究分野：疫学・公衆衛生学

キーワード：特定健診・特定保健指導 保健事業 生活習慣病罹患ハイリスク者階層化 データマイニング 地域一般住民コホート

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

全ての健康保険組合に対し、レセプト等のデータの分析、それに基づく加入者の健康保持増進のための『データヘルス計画』の策定・公表、事業実施、評価等の取り組みが求められており、市町村は、国保保険者としてレセプトデータ、健診データ、介護認定データが一元管理されている国保データベース (KDB) 等を活用してデータの分析を行い、被保険者の健康課題を抽出し、特定健診・特定保健指導を軸として効果的な保健事業を Plan・Do・Act・Check の PDCA サイクルで実施していくことになる。特定健診・特定保健指導では、肥満に該当しないものは、メタボリックシンドロームやその予備群のような介入は行われず情報提供レベルにとどまっているが、データヘルス計画においてはこれらの非該当者を対象とした保健事業も組み込んでいく必要がある。限られた人的・経済的資源を効果的に活用するためには、保健事業の対象となる集団を疾病罹患リスクや効果的な介入を行うことができる特徴を持った幾つかの集団に分けることが必要になる。健康教室のような保健事業は意識の高い比較的高齢者の出席が多く、またメールを活用した介入などは IT 機器に慣れ親しんでいる若い年代に効果的であること、保健事業の必要性を感じていない者については応募による参加型事業では介入できないことなど、効果的な保健事業を行うためには、疾病罹患リスクの高低のみならず、年代や改善意識、どのような生活スタイルかなどの要因も考慮しなければいけない。しかし、これまでにこのような要因まで考慮して、効果的に保健事業を行うために対象集団の分類・階層化を行う方法についての検討はほとんどないのが現状である。

データマイニングの手法の一つである決定木分析は成長予測モデルであり、予測因子候補の中から適切な予測因子を選択し、分割値で分岐することで決定木を構築し、事象の予測を行う。予測因子の選択及び分割値の算出に際しては、集団のバラツキの指標である GINI インデックス等を使用し、分岐毎にその集団のバラツキが最小になるようにモデルを構築する。健診受診者のデータベースに決定木分析の手法を用いることで将来の高血圧、糖尿病、動脈硬化性疾患の罹患リスクの高い者を効果的に分類できると考えられる。また、クラスター分析は、異なる性質のものが混ざりあっている集団の中から互いに似たものを集めて集落 (クラスター) を作り、対象を複数のクラスターに分類する方法である。クラスター分析を用いると客観的な基準に従って科学的に分類ができるため、同じような生活習慣病罹患危険因子を持ち合わせる者や、同じような年代、食生活や運動習慣、生活習慣改善意欲や保健指導利用意思を持つ複数の集団に分類することが可能となる。よってこれらのデータマイニング手法を活用することで、保健事業を効果的に行うための集団の階層化を検討することが可能になると考えられる。

2. 研究の目的

本研究は、継続中の地域一般住民コホート研究のデータを用い、横断的・縦断的検討により、健診結果をもとに決定木分析やクラスター分析といったデータマイニング手法を応用することで積極的な介入が必要と考えられる高血圧や糖尿病などの生活習慣病の罹患リスクで健診受診者集団を分類すること、また実際に保健事業を計画する上で必要となる年齢や生活習慣、保健指導の希望の有無などの要因も考慮して、保健事業を効果的に行うための健診受診集団の分類・階層化が可能となるかの検討を行うことを目的とする。

3. 研究の方法

(1) 将来の高血圧罹患を判別する決定木作成とクラスター分析

1994 年の端野・壮瞥町住民健診受診者中、初年度で既に高血圧に該当する者 (収縮期血圧 (SBP) 140mmHg かつ/または拡張期血圧 (DBP) 90mmHg かつ/または降圧薬内服中の者) と分析に必要なデータの欠損者を除外した 1,000 名を 2007 年までの 13 年間追跡したデータに基づいて将来の高血圧罹患の危険因子を特定した上で、それらの項目を用いて決定木の作成を行った。決定木の成長アルゴリズムには CHAID (Chi-squared, Automatic Interactive Detector) を用いた。作成された決定木における各ノードを一つのカテゴリとし、決定木分岐に使用されていないが保健指導を行う上で検討が必要な要因 (性別, body mass index (BMI), 中性脂肪値 (TG), HDL コレステロール値 (HDL-C), LDL コレステロール値 (LDL-C)) を考慮した階層クラスター分析を行った。

(2) 特定保健指導の流れを考慮したクラスター分析

2017 年の健診受診者 605 名を対象に、(1) において作成した高血圧罹患ハイリスク者判別のための決定木モデルの分岐要因と、メタボリックシンドローム該当や予備群該当など特定保健指導の流れを考慮して対象を層別化し、これらの層をもとに、その他の保健指導を行う上で必要な要因 (性別, BMI, TG, HDL-C, LDL-C, 推定食塩摂取量, 飲酒状況, 保健指導利用希望の有無) を考慮した階層クラスター分析を行った。

(3) 2000 年代コホートの利用による高血圧罹患新規要因の探索と特定保健指導の流れを考慮したクラスター分析

2002 年の健診受診者 1,514 名を対象に、2007 年まで 5 年間の追跡による高血圧罹患を判別する危険因子について空腹時インスリン値や高感度 CRP、血清アディポネクチンなどの新たな要因も候補として決定木分析を行った。決定木の成長アルゴリズムとしては分岐値を算出させる CART (Classification and Regression Tree) 法を用いた。分岐値として選択された要因に基づき、また特定保健指導の流れを考慮した層に分け、それら複数の層

を対象とした階層クラスター分析を行った。

(4) 糖尿病罹患ハイリスク者のクラスター分析

2002年の健診受診者1,514名を対象に、2007年まで5年間の追跡による新規糖尿病罹患(空腹時血糖値126mg/dLまたはHbA1c 6.5%または新規糖尿病治療開始)を判別する決定木モデルを作成し、各ノードを一つのカテゴリとして、特定保健指導の流れを考慮した階層クラスター分析を行った。

4. 研究成果

(1) 将来の高血圧罹患を判別する決定木作成とクラスター分析

1994年の男女合わせた対象では、年齢50歳以上、正常高値血圧、 γ -GTP中央値以上、空腹時血糖値100mg/dL以上を分岐とした決定木により感度68.2%、特異度71.1%で13年後までの高血圧罹患ハイリスク者の選別が可能となり、年齢<50歳のノード1、年齢50歳+正常高値血圧ありのノード2、年齢50歳+正常高値血圧なし+ γ -GTP中央値以上のノード3、年齢50歳+正常高値血圧なし+ γ -GTP中央値未満+空腹時血糖値100mg/dLのノード4、年齢50歳+正常高値血圧なし+ γ -GTP中央値未満+空腹時血糖値<100mg/dLのノード5の5群に分けて、分岐に用いられていない危険因子であるbody mass index, 中性脂肪, HDLコレステロール, LDLコレステロールと男性比率をもとに階層クラスター分析を行うと、ノード1, 5の高血圧罹患リスクの比較的低い群が一つのクラスターにまとまり、ノード2, 3, 4の3群がクラスター化された。ノード1+5からは503名中168名で33%が高血圧に、ノード2+3+4からは497名中361名で72.6%が高血圧に罹患していた。

(2) 特定保健指導の流れを考慮したクラスター分析

特定健診・特定保健指導の流れを念頭に置いた高血圧対策としては、上記(1)の決定木分析の枝分かれ要因と合わせて健診受診者は以下の層に分かれると考えられる: 1) 高血圧治療中、2) メタボリックシンドローム(MetS)該当、3) MetS予備群該当、4) MetS非該当で未治療高血圧、5) 非該当で正常高値血圧、6) 非該当で γ -GTP中央値以上、7) 非該当で空腹時血糖100mg/dL以上、8) 非該当で50歳以上、9) 非該当で50歳未満である。この9つの層に分けた上で、BMI, TG, HDL-C, LDL-C, 推定塩分摂取量、男性頻度、保健指導利用希望者頻度、毎日飲酒者頻度を要因として階層クラスター分析を行うと、a) 非該当で50歳未満、b) メタボ該当+予備群該当、c) 非該当で50歳以上、d) その他の4つのクラスターに分けられた。

(3) 2000年代コホートの利用による高血圧罹患新規要因の探索と特定保健指導の流れを考慮したクラスター分析

2002年を初年度とする決定木モデルで選択された将来の高血圧罹患を判別する要因としては、空腹時インスリン値と空腹時血糖値が新たに選定され、それぞれの分岐値が4.5た。これらを分岐要因に加えた場合のクラスター分析を行った。保健事業の流れに基づいて、対象を1) 高血圧治療中、2) MetS該当、3) MetS予備群該当、4) MetS非該当で未治療高血圧、5) 非該当で正常高値血圧、6) 非該当でIRI4.5 μ U/mL以上、7) 非該当で空腹時血糖100mg/dL以上、8) 非該当で50歳以上、9) 非該当で50歳未満の9つの層に分け、BMI, TG, HDL-C, LDL-C, 推定塩分摂取量、男性頻度、保健指導利用希望者頻度、毎日飲酒者頻度を要因としてクラスター分析を行ったところ、a) MetS該当、b) 50歳未満、c) 非該当で50歳以上、d) その他の群の4つのクラスターに分けられた。

(4) 糖尿病罹患ハイリスク者のクラスター分析

糖尿病罹患ハイリスク者を選別する決定木分析およびクラスター分析を行った。2002年を初年度とした糖尿病罹患に対する決定木分析で選択された空腹時血糖値とHbA1c、および中性脂肪値を考慮し、保健指導の流れに沿って1) 糖尿病該当、2) メタボリックシンドローム(MetS)該当、3) MetS予備群該当、4) 非該当で血糖 ≥ 122 または非該当でHbA1c 5.9%以上、5) 非該当でTG 99mg/dL以上、6) それ以外の6つの層に分け、年齢, BMI, HDL-C, LDL-C, 推定塩分摂取量、男性頻度、保健指導利用希望者頻度、現在飲酒者頻度、現在喫煙者を要因としたクラスター分析の結果、a) MetS該当+糖尿病該当、b) 予備群該当+TG高値、c) 血糖高値+その他の3つのクラスターに分けられた。

(5) コホートの繰り返し健診およびデータベース構築

本研究においては、長期間での健診データの変化を考慮した縦断的検討を行うための繰り返し健診と測定データの蓄積および整備というデータベース構築も目的としている。研究期間中の毎年の健診において、特定健診では測定されない空腹時インスリン値、高感度CRP、尿アルブミン等を測定することができ、これまでのデータベースに加えて研究期間中の4年間のデータの蓄積と整備ができた。

(6) 研究成果のまとめと考察

将来の高血圧罹患を判別する決定木作成とクラスター分析を行うことで、高血圧罹患リスクの高低に加えて、肥満度や脂質異常などの危険因子の保有状況も近い集団にまとめることができたことから、保健指導を行う上でも効率的に介入を行うことが可能になると考えられた。具体的には、c) 非該当で50歳以上、d) その他 MetS 非該当で50歳未満には生活習慣改善の重要性についての情報提供を、メタボ該当 + 予備群該当には内臓脂肪減少に向けた保健指導を、それに加えて高血圧予防の重点対策を行うのであればまず血糖値や β -GTP 高値等の高血圧罹患リスクを保有する対象に、次いで MetS 非該当で50歳以上の集団に、という介入優先度を決定することにつながると考えられた。また、特定健診には含まれない新たな検査項目についても検討を行ったところ、MetS や MetS 予備群非該当者であっても空腹時インスリン高値該当者は高血圧罹患リスクの高いグループとして判別され、未治療高血圧や正常高値血圧者と同様の特徴を持つクラスターに分類されることから、減塩を始めとする高血圧予防・管理のための生活習慣への介入を考慮する必要性が示唆された。高血圧罹患以外にも、糖尿病罹患ハイリスク者の判別とクラスター分類を行ったところ、同様の結果が得られ、高血圧や糖尿病といった生活習慣病予防対策を効果的に行う対象の層別化の可能性が示唆された。しかし、こうした決定木モデルやクラスター分析の結果の妥当性に関しては、同じコホート内で検討しているだけでは有効性・妥当性の評価としては不十分であり、他の集団への適応効果等の検証も必要である。また、本研究においては心血管疾患罹患の判別に関してはイベント発生数が少ないため検討できなかったため、より大きな集団で同様の検討を行う必要があると考えられた。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Toshiaki Seko, Hiroshi Akasaka, Masayuki Koyama, Nobuaki Himuro, Shigeyuki Saitoh, Tetsuji Miura, Mitsuru Mori, Hirofumi Ohnishi	4. 巻 19
2. 論文標題 Lower limb muscle mass is associated with insulin resistance more than lower limb muscle strength in non-diabetic older adults	5. 発行年 2019年
3. 雑誌名 Geriatrics Gerontology International	6. 最初と最後の頁 1254-1259
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/ggi.13805	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 大西 浩文
2. 発表標題 地域一般住民における尿中Na/K比の変化と血圧変化との関連-端野・壮瞥町研究-
3. 学会等名 第39回日本高血圧学会
4. 発表年 2016年

1. 発表者名 世古 俊明, 樋室 伸顕, 赤坂 憲, 齋藤 重幸, 三浦 哲嗣, 大西 浩文
2. 発表標題 経口ブドウ糖負荷試験におけるブドウ糖曲線下面積と糖尿病発症との関連
3. 学会等名 第54回日本循環器病予防学会学術集会
4. 発表年 2018年

1. 発表者名 世古 俊明, 赤坂 憲, 樋室 伸顕, 齋藤 重幸, 三浦 哲嗣, 大西 浩文
2. 発表標題 非糖尿病高齢者の下肢筋量および筋力とインスリン抵抗性の関連
3. 学会等名 第33回日本糖尿病合併症学会
4. 発表年 2018年

1. 発表者名 世古 俊明, 赤坂 憲, 樋室 伸顕, 齋藤 重幸, 三浦 哲嗣, 大西 浩文
2. 発表標題 地域在住高齢者の身体的フレイルと骨格筋量及び筋力との関連：端野・壮瞥町研究
3. 学会等名 第30回日本老年医学会北海道地方会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	齋藤 重幸 (Saitoh Shigeyuki) (60253994)	札幌医科大学・保健医療学部・教授 (20101)	
研究分担者	赤坂 憲 (Akasaka Hiroshi) (70468081)	大阪大学・医学部附属病院・講師 (14401)	
研究協力者	世古 俊明 (Seko Toshiaki)	札幌医科大学・公衆衛生学講座・訪問研究員	
研究協力者	樋室 伸顕 (Himuro Nobuaki)	札幌医科大学・公衆衛生学講座・講師	
研究協力者	小山 雅之 (Koyama Masayuki)	公衆衛生学講座・助教	