

健診での糖尿病指摘後に医療機関受診をしない集団を機械学習により予測

1. 発表者：

岡田 啓（東京大学大学院医学系研究科 糖尿病・生活習慣病予防講座 特任助教）
橋本 洋平（東京大学大学院医学系研究科 公共健康医学専攻 臨床疫学・経済学 特任研究員）
後藤 匡啓（東京大学大学院医学系研究科 公共健康医学専攻 臨床疫学・経済学 客員研究員）
山口 聡子（東京大学大学院医学系研究科 糖尿病・生活習慣病予防講座 特任准教授）
山内 敏正（東京大学大学院医学系研究科/東京大学医学部附属病院 糖尿病・代謝内科 教授）
南学 正臣（東京大学大学院医学系研究科/東京大学医学部附属病院 腎臓・内分泌内科 教授）
康永 秀生（東京大学大学院医学系研究科 公共健康医学専攻 臨床疫学・経済学 教授）
門脇 孝（虎の門病院 院長/東京大学 名誉教授）

2. 発表のポイント：

- ◆特定健診などの生活習慣病の健診を受診した成人のうち、糖尿病（注1）の診断基準を満たし受診勧奨を受けるにも関わらず医療機関を受診しない（=未受診の）人を予測するモデルを、機械学習（注2）を用いて構築しました。
- ◆13個の因子を用いた従来のモデル（注3）よりも、39個の候補から機械学習を用いて選んだ4個の因子のみを用いたモデルの方が、予測性能が高く、特に健診前の12ヶ月の医療機関利用月数が少ないことが、未受診の重要な予測因子であることを発見しました。
- ◆少ない因子で効率よく未受診を予測出来ることにより、早期にハイリスクの集団を同定して介入することが可能になり、糖尿病の合併症を抑制するための医療政策立案に役立つ可能性があります。

3. 発表概要：

国内に約1000万人の患者が存在すると考えられている糖尿病は、未治療のまま放置すると心筋梗塞、脳卒中や慢性腎不全などの合併症のリスクが高まりますが、適切な治療の継続によりそれらの発症が抑えられることが知られています。しかしながら、特定健診などの生活習慣病の健診で糖尿病を指摘され医療機関の受診を勧奨されても、半数以上の患者が受診しないことが分かっており、受診率の向上が求められていました。

この度、東京大学の岡田啓特任助教、山口聡子特任准教授、山内敏正教授、南学正臣教授、康永秀生教授、門脇孝名誉教授（虎の門病院院長）らの研究グループは、機械学習を用いて、健診で糖尿病を指摘された後、未受診となる集団を予測するモデルの構築を試みました。診療報酬請求明細書・健診のデータを含む大規模疫学データベースを用いて、従来の13個の因子で予測するモデルと比較して予測能が高いモデルを、機械学習を用いることにより、4つの因子のみで構築できることを明らかにしました。本研究の成果は、糖尿病の合併症の予防を目的とした政策立案に大きく貢献するエビデンスとなることが期待されます。

本研究は、日本時間2022年4月18日に米国科学誌 *Diabetes Care* オンライン版 **Published Ahead of Print** に掲載されます。

なお本研究は、厚生労働省および文部科学省科学研究費補助金（厚生労働行政推進調査事業費補助金・政策科学総合研究事業（政策科学推進研究事業）「診療現場の実態に即した医療ビッグデータを利活用できる人材育成促進に資するための研究」（課題番号 21AA2007 研究代

表 康永秀生)、文部科学省科学研究費補助金「糖尿病受診中断の予測因子探索と政策提言」(課題番号 20K18957 研究代表 岡田啓)の支援により行われました。

東京大学大学院医学系研究科 糖尿病・生活習慣病予防講座は、朝日生命保険相互会社との社会連携講座です。

4. 発表内容:

【研究の背景】

糖尿病の治療を受けずに放置することにより、深刻な合併症になるリスクが高まることが知られており、糖尿病と診断された場合には、受診して治療を受けることが非常に重要です。しかしながら、糖尿病は一般に無症状であることも多いため、健診で糖尿病を指摘された後も、医療機関の受診率は低いのが現状です。実際、過去の報告によると、糖代謝異常で受診勧奨を受けた患者のうち、35%しか受診していませんでした。このような健診後に受診しない集団についての研究は、世界でも数が少なく、どのような集団が受診しにくいのか、ということはいまだに明らかではありませんでした。日本では、13個の因子を用いた予測モデルの報告がありますが、機械学習を用いることにより、より「良く」「効率的に」糖尿病受診勧奨後に未受診となる集団を予測できれば、このようなリスクの高い集団に早期に介入することにより、受診率を上げて合併症を予防する医療政策立案に役立つ可能性があると考えられました。本研究では、機械学習を用いて、糖尿病の受診勧奨後の未受診を予測するモデルの構築を試みました。

【研究内容】

診療報酬請求明細書・健診のデータを含むJMDCデータベースには2019年10月までの健診情報が登録されています。この中から、健診受診から半年以上追跡が可能な成人のうち、健診で糖尿病の診断基準(HbA1c値が6.5%以上かつ空腹時血糖値が126 mg/dL以上)を満たした10,645人を解析対象としました。このうち5,450人(51.2%)が受診勧奨にも関わらず、糖尿病に関して医療機関を受診しませんでした。対象を4:1に分け、前者をモデル構築用、後者をモデル検証用の集団とし、モデル構築用の集団でLasso回帰を行い、未受診を予測する重要な因子を同定しました。より少ない因子で予測能を保つことが出来ると言われるLasso回帰の1標準誤差ルールを用いて、39個の因子の中から予測因子を選択しました。対照となる予測モデルとしては、2014年に科学誌に掲載された、13個の因子を含む予測モデルを用いました。Lasso回帰の1標準誤差ルールで選ばれた予測因子は、(1)過去12ヶ月の受診頻度、(2)HbA1c値、(3)脂質異常症薬処方、(4)降圧薬処方であり、この4つだけで予測したモデルの方が対照モデルの予測性能よりもDelong検定で有意に高い予測能を示しました(c統計量:Lassoモデル0.71(95%信頼区間0.69-0.73)、対照モデル0.67(95%信頼区間0.65-0.69)、P値<0.001)。

本研究で機械学習によって得られた予測モデルは、健診で糖尿病を指摘後に未受診となる集団を、既存モデルよりも高い予測能で予測できることを明らかにしました。

【社会的意義・今後の予定】

糖尿病の合併症予防は、日本のみならず、世界中で大きな公衆衛生上の問題であり、本研究の結果が活用されれば、日本全体における糖尿病患者の受診率が向上する可能性があります。受診率が向上すれば、心筋梗塞、脳卒中や慢性腎不全などの合併症を予防できる可能性が増加し、心筋梗塞や脳梗塞といった急性期疾患の予防に繋がり、ひいては健康寿命の延伸にも繋がることが期待されます。

5. 発表雑誌：

雑誌名：「*Diabetes Care*」 (オンライン版：4月18日)

論文タイトル：A machine learning-based predictive model to identify patients who failed to attend a follow-up visit for diabetes care after recommendations from a national screening program

著者：Akira Okada, Yohei Hashimoto, Tadahiro Goto, Satoko Yamaguchi, Sachiko Ono, Kayo Ikeda Kurakawa, Masaomi Nangaku, Toshimasa Yamauchi, Hideo Yasunaga, Takashi Kadowaki*

DOI 番号：10.2337/dc21-1841

6. 注意事項：

日本時間 4月18日 (月) 午後 11 時 (米国東部夏時間：18日 (月) 午前 10 時) 以前の公表は禁じられています。

7. 問い合わせ先：

<研究内容に関するお問い合わせ先>

東京大学医学部附属病院糖尿病・生活習慣病予防講座

特任助教 岡田 啓 (おかだ あきら)

特任准教授 山口 聡子 (やまぐち さとこ)

電話：03-3815-5411 (内線：34347)

E-mail：aokada@m.u-tokyo.ac.jp (岡田)

syamaguc@m.u-tokyo.ac.jp (山口)

<広報担当者連絡先>

東京大学医学部附属病院 パブリック・リレーションセンター

担当：渡部、小岩井

電話：03-5800-9188 (直通) E-mail：pr@adm.h.u-tokyo.ac.jp

国家公務員共済組合連合会 虎の門病院 総務課

担当：萩

電話：03-3560-7752 (直通) E-mail：mihoko-h@toranomom.gr.jp

8. 用語解説：

(注1) 糖尿病：

糖尿病は、血糖を下げるホルモンであるインスリンの膵臓からの分泌不足または、筋肉や肝臓、脂肪組織でのインスリンの効きが悪くなっているインスリン抵抗性のどちらか、または両方が合わさったインスリン作用不足による慢性的な高血糖状態のことです。糖尿病を無治療で放置しておく、心筋梗塞や脳梗塞のような心血管病を引き起こしたり、透析に至る腎不全になったりするリスクが高まることがわかっています。日本では糖尿病がある人が約1,000万人いるとされています。血糖マネジメントの指標として、空腹時血糖値や過去1~2ヶ月の平均血糖値を反映するHbA1c値が用いられます。糖尿病と診断される場合の一つに、空腹時血糖値126 mg/dL以上かつHbA1c値6.5%以上があり、今回の対象集団を選ぶのに用いられています。

(注 2) 機械学習：

機械学習とは、人間が経験に基づいてものごとを判断するように、コンピュータを使ってある一定のアルゴリズムに従って判断や分別を予測することです。Lasso 回帰は、数多くある機械学習のなかでも古典的な手法です。

(注 3) 13 個の因子を用いた従来モデル：

2014 年に京都大学の研究グループによって発表されたモデルです。13 因子は、年齢、性別、体格指数 BMI、HbA1c 値、尿糖陽性の有無、蛋白尿陽性の有無、降圧剤処方の有無、脂質異常症薬処方の有無、抗うつ薬処方の有無、喫煙状態、飲酒頻度、ヘモグロビン値、被保険者本人かどうか、を指します。

<参考論文>

Tsujimura Y, Takahashi Y, Ishizaki T, Kuriyama A, Miyazaki K, Satoh T, et al. Predictors of hyperglycaemic individuals who do not follow up with physicians after screening in Japan: a cohort study. *Diabetes Res Clin Pract* 2014;105:176–184.