

健康診断・診療情報ビッグデータを用いた疾病発症予測

近年、様々な分野でビッグデータを用いた研究が発展し成果を上げつつあるが、医療分野におけるビッグデータを用いた解析への取り組みはまだ始まったばかりであり、研究方法もシステム化されていないのが、現状である。

現在、高血圧や糖尿病などの生活習慣病の発症リスクについては未だ不明な点が多く、個々の患者において疾患表現型や併存疾患が異なるにも関わらず、医療従事者から患者に提供するものは平均的医療に留まっている。これを解明し、発症リスクを抽出し数理モデル化することで、個々の疾病発症を予測することが可能となり、個々の状態に応じた治療を提供する精密医療(Precision Medicine)が実現につながると考えられる。

我々はまず ABC Study, PPAR Study (耐糖能異常や軽度糖尿病を有し心筋梗塞の既往のある患者群を対象とした心筋梗塞再発予防に関する研究) に登録した臨床データを対象とし、心血管イベントの出現を、データマイニング法の一つである LAMP (Limitless-Arity Multiple-testing Procedure) を用いて解析し、有害事象に関与するリスクファクター、およびその組み合わせを求めた。LAMP は、超高速アルゴリズムを用いて無為な出現頻度の低い組み合わせを特定し取り除くことによって、補正係数を大幅に削減した状態での多重検定を可能とする方法である。この研究により高い血漿 BNP レベル (200pg/ml) は、心筋梗塞の再発を予測することが判明した。次に、グローバル企業の健保組合から提供を受けた 2011 年から 2015 年の約 5 万人分の情報を対象に LAMP を用いて、生活習慣病や心血管病の発症に関与するリスクファクター、およびその組み合わせを求めた。高血圧症、脂質異常症、糖尿病については、すべてにおいて年齢、BMI、体重増加、その他の生活習慣病の既往が疾病の発症に関与しており、また男性であることも強い予測因子であった。これらの結果は新規の発見ではないものの、LAMP の解析の正確性、汎用性の高さを示すものと考えられる。

次のステップとして、別の集団において疾病発症のリスクファクター、およびその組み合わせを多く持つ個人が実際に疾病を発症するかを検証し、各疾病ごとの発症リスクの抽出法の確立とその応用を目指す。

※ 榎本 竜郎 ※※ 宮下 洋平 ※※ 鷺尾 隆 ※※※ 北風 政史

研究テーマ

健康診断や多施設の診療情報、およびレセプトなどの臨床データ (ビックデータ) を対象として、高血圧や糖尿病といった生活習慣病が発症するプロセ

スを数理的に法則化、モデル化することで、疾病の発症リスクの抽出法の確立とその応用を目指す。

緒言

近年、様々な分野でビッグデータを用いた研究が発展し成果を上げつつあるが、医療分野におけるビッグデータを用いた解析への取り組みはまだ始まったばかりであり、研究方法もシステム化されていないのが、

※国立循環器病研究センター
 ※※大阪大学
 ※※※国立循環器病研究センター・
 阪和第二泉北病院

現状である。しかしながら、そのような中でも我々はこれまでに、以下の2つの点についてパイロット的な研究を行い、実証してきた。すなわち、1) 当院、国立循環器病研究センターにおける診療情報のビッグデータを用いた、心不全死死亡または心不全増悪による再入院までの期間を予測する数理モデルの作成¹⁾、および2) データマイニング法の一つである LAMP (Limitless-Arity Multiple-testing Procedure) を用いることにより、心血管病の病態悪化や予後に関与する未知のリスクファクターおよびその組み合わせを求めるなどの研究²⁾、である。

現在、高血圧や糖尿病などの生活習慣病の発症リスクについては未だ不明な点が多く、個々の患者において疾患表現型や併存疾患が異なるにも関わらず、医療従事者から患者に提供するものは平均的医療に留まっている。これを解明し、発症リスクを抽出し数理モデル化することで、個々の疾病発症を予測することが可能となり、個々の状態に応じた治療を提供する精密医療 (Precision Medicine) が実現につながると考えられる。

目 的

高血圧や糖尿病などの生活習慣病や心血管病等の発症リスクについてビッグデータを用いた解析することを、本研究の目的とする。企業の健康保険組合より入手した一般健康診断のデータ、および大規模臨床研究で得られたデータを対象として LAMP を用いた心血管

疾患発症リスクの抽出法の確立とその応用を目指す。

方 法

我々は今回以下の2種類のビッグデータを対象に解析を行った。

1. ABC Study, PPAR Study (耐糖能異常や軽度糖尿病を有し心筋梗塞の既往のある患者群を対象とした心筋梗塞再発予防に関する研究) に登録した臨床データ
2. 特殊企業従業員 5 万人からの健康診断およびレセプトのデータ (2011 年 -2015 年の期間)

これらを対象に LAMP を用いてデータマイニングを行い、生活習慣病や心血管病の発症に関与するリスクファクター、およびその組み合わせを求めた。LAMP は、超高速アルゴリズムを用いて無為な出現頻度の低い組み合わせを特定し取り除くことによって、補正係数を大幅に削減した状態での多重検定を可能とする方法である。

結 果

1.ABC Study, PPAR Study (耐糖能異常や軽度糖尿病を有し心筋梗塞の既往のある患者を対象とした心筋梗塞再発予防に関する研究) に登録した臨床データを対象とし、心血管イベント (死亡, 心筋梗塞, 狭心症, 心不全, 脳卒中) の出現を、LAMP を用いて解析し、有害事象に関与するリスクファクター、およびその組み合わせを求めた。この研究により高い

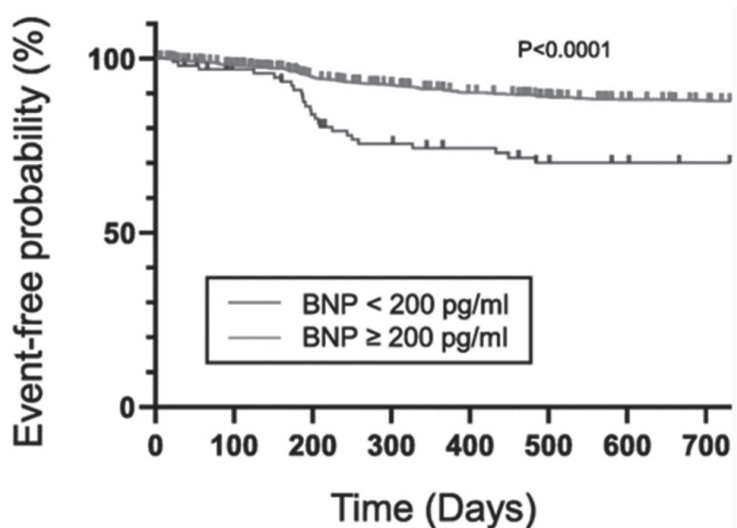


図 1 BMP 値による心筋梗塞再発の Kaplan-Meier 曲線

血漿 BNP レベル (200pg/ml) は、心筋梗塞の再発を予測する (図 1) ことが判明し、論文化を行った (Cardiovascular Drugs and Therapy volume 34, pages535-545 (2020)).

2. グローバル企業の健保組合から提供を受けた 2011 年から 2015 年の約 5 万人分の情 (経年健診データ, 受診履歴・病名データ, 処方データ等) を対象に LAMP を用いて、生活習慣病の発症に関与するリスクファクター、およびその組み合わせを求めた。ある疾病を対象に 2011 年の時点での非発症者のうち、その後 5 年間における疾病の新規発症を正例、それ以外を負例として、正例となる発症リスク因子の組み合わせを、非発症期の 2011 年のデータを利用して同定するという手法を用いた。因子の組み合わせは最大を 3 として解析を行い、高血圧症は 26,313 人のうち新規発症が 3,498 人、脂質異常症は 20681 人のうち新規発症が 5,997 人、糖尿病は 30,200 人のうち新規発症が 1,962 人であった。代表的な結果を下記 (図 2) に示す。3 疾患すべてにおいて年齢、BMI、体重増加、その他の生活習慣病の既往が疾病の発症に関与しており、また男性であることも強い予測因子であった。これらの

結果は新規の発見ではないものの、LAMP の解析の正確性、汎用性の高さを示すものと考えられる。

結 論

データマイニング法の一つである LAMP を用いて、耐糖能異常や軽度糖尿病を有し心筋梗塞の既往のある患者群を対象として解析した結果、高い血漿 BNP レベル (200pg/ml) は、心筋梗塞再発の予測因子であることが判明した。また、一般人を対象に高血圧、脂質異常症、糖尿病などの生活習慣病の発症予測因子を検討したところ、年齢、BMI、体重増加、その他の生活習慣病の既往が関係することが明らかとなった。後者については、LAMP 解析の正確性、汎用性の高さが証明されたものと考えられた。次のステップとして、別の集団において疾病発症のリスクファクター、およびその組み合わせを多く持つ個人が実際に疾病を発症するかを検証し、今後は各疾病ごとの発症リスクの抽出法の確立とその応用を目指す。

文 献

- 1) Hiroki Fukuda, Masafumi Kitakaze, et al. Non-

図 2 LAMP を用いた生活習慣病の発症に関与するリスクファクター、及びその組み合わせ

高血圧症

The combinatiin of clinical parameters		Adjusted p value
【性別】 男		1.7282E-189
【腹囲】 85cm(male) or 90cm(female)cm ≤		1.4793E-130
【BMI】 25~30kg/m ²		2.1126E-124
【体重増加】 はい		1.8564E-116
【酒】 毎日		3.425E-55
【尿酸】 7.1mg/dl ≤		2.7703E-52
【GPT】 41~100IU/L		1.4065E-47
【中性脂肪】 150mg/dl ≤		5.7308E-44
【LDL】 140mg/dl		8.9498E-31
【タバコ】 はい		2.5687E-08
【年齢(年度末)】 60 ≤		8.7244E-07
【HbA1C(NGSP)】 6.2% ≤		0.0013327
【収縮張期血圧】 80~140mmHg		0.0054505
【HDL】 <40mg/dl		0.014587
【性別】 男	【収縮張期血圧】 80~140mmHg	6.21E-175
【性別】 男	【腹囲】 85cm(male) or 90cm(female)cm ≤	9.33E-134
【腹囲】 85cm(male) or 90cm(female)cm ≤	【収縮張期血圧】 80~140mmHg	6.07E-128
【体重増加】 はい	【年齢(年度末)】 30~60	3.10E-109
【運動30分以上】 いいえ	【BMI】 25~30kg/m ²	4.79E-93
【性別】 男	【BMI】 25~30kg/m ²	【収縮張期血圧】 80~140mmHg
【体重増加】 はい	【年齢(年度末)】 30~60	【性別】 男
【運動30分以上】 いいえ	【体重増加】 はい	【収縮張期血圧】 80~140mmHg

脂質異常症

The combinatiin of clinical parameters			Adjusted p value
【体重増加】 はい			2.86E-158
【腹囲】 85cm(male) or 90cm(female)cm ≦			3.49E-126
【BMI】 25~30kg/m ²			2.17E-98
【GPT】 41~100IU/L			3.21E-40
【年齢(年度末)】 60 ≦			1.28E-37
【尿酸】 7.1~9mg/dl			1.33E-37
【収縮張期血圧】 140~180mmHg			6.64E-31
【拡張期血圧】 90~110mmHg			8.21E-31
【タバコ】 はい			1.21E-16
【HbA1C(NGSP)】 6.5~7%			4.24E-11
【体重増加】 はい	【性別】 男		2.1148E-167
【体重増加】 はい	【年齢(年度末)】 30~60		1.8051E-152
【運動30分以上】 いいえ	【体重増加】 はい		1.9837E-116
【性別】 男	【BMI】 25~30kg/m ²		6.2291E-106
【体重増加】 はい	【収縮張期血圧】 80~140mmHg		2.1846E-104
【運動30分以上】 いいえ	【体重増加】 はい	【年齢(年度末)】 30~60	3.23E-121
【性別】 男	【腹囲】 85cm(male) or 90cm(female)cm ≦	【年齢(年度末)】 30~60	2.61E-114
【体重増加】 はい	【年齢(年度末)】 30~60	【収縮張期血圧】 80~140mmHg	3.17E-106

糖尿病

The combinatiin of clinical parameters			Adjusted p value
【HbA1C(NGSP)】 6.2~6.5%			4.4236E-92
【年齢(年度末)】 60 ≦			5.6602E-65
【BMI】 25kg/m ² ~			2.4449E-35
【腹囲】 85cm(male) or 90cm(female)cm ≦			7.4574E-33
【空腹時血糖】 110~126mg/dl			4.5273E-32
【体重増加】 はい			1.0711E-29
【LDL】 140~200mg/dl			1.1652E-23
【γGTP】 71IU/L ≦			3.3349E-23
【中性脂肪】 150mg/dl ≦			9.3937E-22
【拡張期血圧】 90~110mmHg			2.6011E-18
【収縮張期血圧】 140~180mmHg			1.3636E-17
【心電図判定】 H(専門医要受診)			2.7214E-06
【性別】 男			0.00041704
【HbA1C(NGSP)】 6.2~6.5%	【性別】 男		6.7497E-86
【HbA1C(NGSP)】 6.2~6.5%	【年齢(年度末)】 30~60		2.3015E-58
【体重増加】 はい	【BMI】 25~30kg/m ²		1.1233E-27
【運動30分以上】 いいえ	【BMI】 25~30kg/m ²		3.6268E-23
【酒】 毎日	【γGTP】 71 ≦		0.000011003
【HbA1C(NGSP)】 6.2~6.5%	【収縮張期血圧】 80~140mmHg	【拡張期血圧】 <90mmHg	2.33E-59
【HbA1C(NGSP)】 6.2~6.5%	【性別】 男	【年齢(年度末)】 30~60	3.65E-56
【HbA1C(NGSP)】 6.2~6.5%	【体重増加】 はい	【年齢(年度末)】 30~60	3.90E-35
【運動30分以上】 はい	【空腹時血糖】 110~126mg/dl	【年齢(年度末)】 60 ≦	3.09E-08

linear Equation using Plasma Brain Natriuretic Peptide Levels to Predict Cardiovascular Outcomes in Patients with Heart Failure. Sci Rep. 2016 Nov 15;6:37073. (2016)

- 2) Hiroki Fukuda, Masafumi Kitakaze, et al. Elucidation of the Strongest Predictors of Cardiovascular Events in Patients with Heart Failure. EBioMedicine. Jul;33:185-195(2018)

研究成果

Kazuhiro Shindo, Tatsuro Hitsumto, Masafumi Kitakaze, et al. Cardiovascular Drugs and Therapy volume 34, pages535–545 (2020)

PREDICTION OF THE ONSET OF VARIOUS DISEASES USING THE DATA-MINING METHOD OF CLINICAL BIG DATA

※ Tatsuro HITSUMOTO, ※※ Yohei MIYASHITA, ※※ Takashi WASHIO, ※※※ Masafumi KITAKAZE

※ National Cerebral and Cardiovascular Center

※※ Osaka University

※※※ National Cerebral and Cardiovascular Center • hanwa

In recent years, research using big data has been developing and producing results in various fields. However the introduction of analysis using big data in the medical field has only just begun, and research methods have not been systematized yet.

Currently, in the treatment of lifestyle diseases such as hypertension and diabetes, medical workers provide only averaged medical care to patients, even though individual patients have different disease phenotypes and comorbidities. By elucidating these, extracting the risk of onset and utilizing mathematical modeling, it will be possible to predict the onset of individual diseases, lead to the realization of precision medicine that provides treatment according to each condition.

We first targeted clinical data enrolled in the ABC Study, PPAR Study (a study on the prevention of myocardial infarction recurrence in a group of patients with impaired glucose tolerance or mild diabetes and a history of myocardial infarction). The appearance of cardiovascular events was analyzed, using one of the novel data mining methods of limitlessness multiple-testing procedure (LAMP: Limitless-Arity Multiple-testing Procedure). LAMP allowed us to analyse all significant combinations of clinical parameters without any limits, via the limitlessness multiple testing procedure. This study found that high plasma BNP levels (200 pg/ml) predicted recurrence of myocardial infarction.

Next, we analyzed the data of health examinations for about 50,000 people from 2011 to 2015 using LAMP, and determined the risk factors involved in the development of lifestyle diseases, cardiovascular diseases and their combinations. For hypertension, dyslipidemia and diabetes, age, BMI, weight gain and a history of other lifestyle diseases were all involved in the onset of the disease. In addition, being male was also a strong predictor. Although these results are not new discoveries, they are considered to indicate the accuracy and versatility of LAMP analysis.

As the next step, we will examine whether people who have many risk factors, as well as their combinations for the onset of disease, actually develop the disease in another population; and then establish and apply a method for determining the risk of developing the disease for each case.