

# Potential Role of Artificial Intelligence for the Previous Study Using Traditional Analysis

Kei Nakajima<sup>a, b, c</sup> , Manami Igata<sup>a</sup>, Ryoko Higuchi<sup>a</sup>

## To the Editor

In the past decade, it has become evident that artificial intelligence (AI) contributes to the analysis of various social and scientific areas including medical studies, particularly large epidemiological studies consisting of big data [1, 2]. However, the analysis of medical big data using AI is still generally unfamiliar and likely to be unfeasible for most clinical investigators [3, 4]. The exception is investigators who work in special fields such as radiology because of the implicated procedures in the algorithm and required computer technologies [4, 5].

We recently published a study using the healthcare data of 387,642 people in the general population (age range: 40 - 68 years) showing that an extremely high concentration of high-density lipoprotein cholesterol (HDL-C) (> 110 mg/dL) was associated with the incidence of diabetes after 6 years compared with the tentative reference of an HDL-C concentration of 80 to 90 mg/dL [6].

The relative risk of an extremely high HDL-C concentration of > 110 mg/dL for diabetes was 1.46 (95% confidence interval: 1.18 - 1.81), 2.45 (1.70 - 3.53), and 1.53 (1.02 - 2.29), for all study subjects, men, and women, respectively, after adjustment for relevant confounding factors including age, treatments, alcohol consumption, and smoking. Therefore, the results showed the possibility that an extremely high HDL-C concentration may not be beneficial for health and that the optimal HDL-C concentration should not be the highest to reduce the risk of diabetes.

In our previous analysis, we used the SAS Enterprise Guide (SAS-EG 7.1) in SAS software, version 9.4 (SAS Institute, Cary, NC, USA), which has been used for numerous medical studies worldwide for several decades.

In the present study, we challenged the findings of our

above-mentioned previous study using an AI analysis system (Prediction One; Sony Network Communications Inc., Tokyo, Japan) [7]. Such an analysis is easy to perform with Prediction One once the data sheet has been properly prepared.

The outcome was the incidence of diabetes determined by the hemoglobin A1c (HbA1c) concentration ( $\geq 6.5\%$ ), fasting plasma glucose concentration ( $\geq 126$  mg/dL), and treatment for diabetes, and 10 variables were selected as contributing factors. Data learning, evaluation, and neural network analysis were automatically performed, and it took about 20 min to complete.

Table 1 shows the results of the AI analysis. In all study subjects, the incidence of diabetes was predicted by the body mass index, age, systolic blood pressure, triglyceride concentration, HDL-C concentration, smoking, sex, history of cardiovascular disease, habitual exercise, and alcohol consumption in order by weight. The weights may be lower because of the large sample size and low incidence of diabetes (3.8% in total). The area under the curve, accuracy, precision, recall, and F value of the overall model were 74.5%, 91.4%, 13.8%, 23.8%, and 17.5%, respectively.

Figure 1 shows the details of the contributions in the HDL-C categories. Among nine categories of HDL-C ( $\leq 39$ , 40 - 49, 50 - 59, 60 - 69, 70 - 79, 80 - 89, 90 - 99, 100 - 109, and  $\geq 110$  mg/dL), the first and second degrees of HDL-C categories for the prediction of a positive incidence of diabetes were:  $\leq 49$  mg/dL (a) and 50 - 59 mg/dL (b), whereas the third degree was  $\geq 110$  mg/dL (c). The first degree of HDL-C for the negative incidence of diabetes was 80 - 89 mg/dL (a), but not an extremely high HDL-C of > 110 mg/dL. The second and third degrees were 90 - 109 g/dL and 70 - 79 mg/dL.

For comparison with other representative variables contributing to the incidence of diabetes, the results of the body mass index and age are also shown in Figure 1. These results for the incidence of diabetes appear to be reasonable, and some may bring new insights.

Taking these results obtained by AI into consideration, it may be concluded that an extremely high HDL-C concentration of > 110 mg/dL is not beneficial in terms of the development of diabetes and that the optimal HDL-C concentration for the prevention of diabetes may be 80 - 89 mg/dL.

Although the methods and expression of the results differ between traditional analysis and AI, we confirmed that these results obtained by AI were almost the same as those in our previous study involving logistic regression analysis using SAS software.

The currently used AI analysis system (Prediction One) does not require the user to possess specific AI skills, which

Manuscript submitted July 21, 2021, accepted July 26, 2021  
Published online July 28, 2021

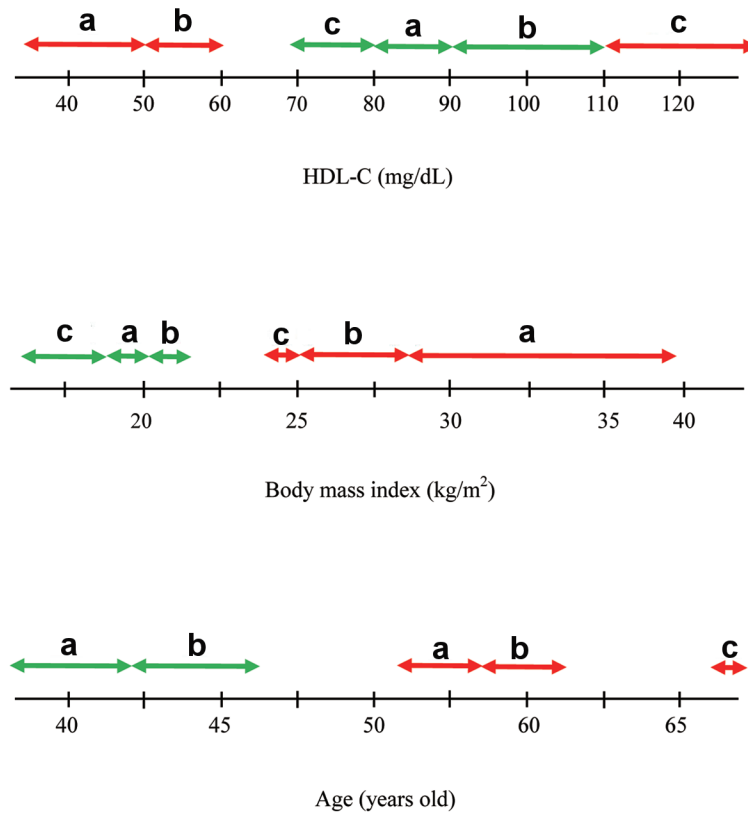
<sup>a</sup>School of Nutrition and Dietetics, Faculty of Health and Social Services, Kanagawa University of Human Services, 1-10-1 Heisei-cho, Yokosuka, Kanagawa 238-8522, Japan

<sup>b</sup>Department of Endocrinology and Diabetes, Saitama Medical Center, Saitama Medical University, 1981 Kamoda, Kawagoe, Saitama 350-8550, Japan

<sup>c</sup>Corresponding Author: Kei Nakajima, School of Nutrition and Dietetics, Faculty of Health and Social Services, Kanagawa University of Human Services, 1-10-1 Heisei-cho, Yokosuka, Kanagawa 238-8522, Japan.

Email: nakajima-rsh@kuhs.ac.jp

doi: <https://doi.org/10.14740/jocmr4568>



**Figure 1.** Detailed contributions of variables to the incidence of diabetes. Red and green arrows indicate prediction of the positive and negative incidence of diabetes. a: First degree. b: Second degree. c: Third degree. The space indicates the categories of lower degrees (≥ fourth degree). HDL-C: high-density lipoprotein cholesterol.

allowed us to utilize the high-spec functions of AI.

In this context, we believe that some potential roles of analysis using this AI system include confirmation and support of the data analysis using traditional statistical software [1, 8, 9]. In addition, AI may prevent us from overlooking important factors [9].

At present, however, it is unknown whether AI analysis will become the primary method for medical study and deci-

sion-making in the healthcare system beyond the traditional analytical methods in the near future. Much more research using traditional analysis and AI is needed to elucidate this issue.

### Acknowledgments

We thank Angela Morben, DVM, ELS, from Edanz (<https://>

**Table 1.** Variables Contributing to Prediction of the Incidence of Diabetes

Degree of contribution (order)	Variables	Weight
1	Body mass index (kg/m <sup>2</sup> )	0.033
2	Age (years old)	0.020
3	Systolic blood pressure (mm Hg)	0.020
4	Serum triglyceride (mg/dL)	0.016
5	Serum HDL-C (mg/dL, nine categories)	0.013
6	Smoking (yes or no)	0.011
7	Sex (men, women)	0.011
8	History of cardiovascular disease (yes or no)	0.005
9	Habitual exercise (yes or no)	0.004
10	Alcohol consumption (daily, sometimes, or almost none)	0.004

All listed variables are those measured at baseline. HDL-C: high-density lipoprotein cholesterol.

jp.edanz.com/ac), for editing a draft of this manuscript.

### Financial Disclosure

None to declare.

### Conflict of Interest

None of the authors have any potential conflict of interest.

### Informed Consent

Not applicable.

### Author Contributions

KN contributed to the study design and the interpretation of the initial analysis. MI and RH conducted the discussion of the literature. KN prepared the first draft of the manuscript. All authors read and edited the manuscript.

### Data Availability

Any inquiries regarding supporting data availability of this study should be directed to the corresponding author.

### References

1. Musacchio N, Giancaterini A, Guaita G, Ozzello A, Pellegri MA, Ponzani P, Russo GT, et al. Artificial intelligence and big data in diabetes care: a position statement of the Italian Association of Medical Diabetologists. *J Med Internet Res*. 2020;22(6):e16922.
2. Khan ZF, Alotaibi SR. Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective. *J Healthc Eng*. 2020;2020:8894694.
3. Mehta N, Pandit A, Shukla S. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. *J Biomed Inform*. 2019;100:103311.
4. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. *BMC Med Inform Decis Mak*. 2021;21(1):125.
5. Ahmad Z, Rahim S, Zubair M, Abdul-Ghaffar J. Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. *Diagn Pathol*. 2021;16(1):24.
6. Nakajima K, Higuchi R, Iwane T, Shibata M, Takada K, Sugiyama M, Matsuda M, et al. High incidence of diabetes in people with extremely high high-density lipoprotein cholesterol: results of the Kanagawa investigation of total checkup data from the National Database-1 (KITCHEN-1). *J Clin Med*. 2019;8(3):381.
7. Sony Network Communications, Prediction One; 2020. Available from: <https://www.predictionone.sony.biz>. Last accessed on July 29, 2021.
8. Choi DJ, Park JJ, Ali T, Lee S. Artificial intelligence for the diagnosis of heart failure. *NPJ Digit Med*. 2020;3:54.
9. Romiti S, Vinciguerra M, Saade W, Anso Cortajarena I, Greco E. Artificial intelligence (AI) and cardiovascular diseases: an unexpected alliance. *Cardiol Res Pract*. 2020;2020:4972346.