Data Article

Korean public and hospital data for estimating LDL-cholesterol

CrossMark

Taesic Lee [a], Juwon Kim [b], Young Uh [b,*], Hyunju Lee [a,c,**]

[a] Department of Biomedical Science and Engineering, Gwangju Institute of Science and Technology, Gwangju, South Korea
[b] Department of Laboratory Medicine, Yonsei University Wonju College of Medicine, Wonju, South Korea
[c] School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea

A R T I C L E I N F O

A B S T R A C T

The data is composed of 14,812 subjects, and was obtained from the Korean National Health and Nutritional Examination Survey (KNHANES) from 2009 to 2015. The KNHANES data is publicly available at https://knhanes.cdc.go.kr/knhanes/eng/index.do.

The dataset consisting of 4520 participants was obtained from Wonju Severance Christian Hospital (WSCH) in South Korea. To protect the patient's personal information, we removed ID and gender of patients. Detail information of the data presented in the present article is available in the research article "Deep neural network for estimating low density lipoprotein cholesterol" (Lee et al., 2018).

© 2018 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications table

Subject area	*Biology*
More specific subject area	*Medical informatics, Lipids, Lipoproteins*
Type of data	*Online data*
How data was acquired	*KNHANES (Publicly available)* *– Total cholesterol, triglyceride, HDL-C, and LDL-C levels were measured by the Hitachi analyzer 7600 (Hitachi, Tokyo, Japan)* *WSCH (extracting from hospital database)* *– The total cholesterol, triglyceride, HDL-C, and LDL-C for the testing dataset were analyzed using the modular DPE system (Roche Diagnostics, Basel, Switzerland). The lipoprotein subfraction test was performed using the Lipoprint LDL System (Quantimetrix, Redondo Beach, CA, USA).*
Data format	*Raw*
Experimental factors	*Public data including 14,812 subjects with information about age, gender, and levels of lipid profiles.* *Korea hospital data including 4520 patients with information about levels of lipid profiles.*
Experimental features	*Medians of total cholesterol, HDL cholesterol, and triglyceride in KNHANES were 186, 47, 120, respectively. Medians of total cholesterol, HDL cholesterol, and* *triglyceride in Korea hospital data were 165, 37, 103, respectively.*
Data source location	*KNHANES (Seoul, South Korea)* *WSCH (Wonju, South Korea)*
Data accessibility	The KNHANES data is publicly available at https://knhanes.cdc.go.kr/knhanes/eng/index.do.
Related research article	Taesic Lee, Juwon Kim, Young Uh, Hyunju Lee, Deep neural network for estimating low density lipoprotein cholesterol (Clin Chim Acta; in press) [1].

Value of the data

- Representative for Korea: The KNHANES study was carried out by Korean Centers for Disease Control and Prevention (K-CDC), which is nation-wide and population-based data. Participants are selected by using cluster-based random sampling according to demographic features to make the KNHANES representative of Korea.
- Clinical application: After constructing the final model using the KNHANES, the model was applied to the local hospital dataset. The present framework suggests that it is immediately applicable to a clinical setting.
- Potential to construct more accurate model: The data has potential to be used for constructing more accurate and feasible LDL cholesterol estimating model

1. Data

Subjects with lipid profiles in 2009–15 of KNHANES and in 2008–13 of a Korean hospital were included in the healthy person-based and the patient-based data, respectively. Note that LDL cholesterol of both data was directly measured. The KNHANES data is publicly available at https://knhanes.cdc.go.kr/knhanes/eng/index.do.

2. Experimental design, materials and methods

We obtained a dataset of 14,812 anonymized laboratory test records from 2009–15 KNHANES. Subjects with any missing results of total cholesterol, triglyceride, HDL cholesterol, and LDL cholesterol were excluded. Participants with dyslipidemia medication or higher triglyceride concentration of more than 400 mg/dL [2] were not excluded. All lipid profiles were tested on fasting state more than 12 h.

For the Korean hospital data, lipoprotein subfraction test results were obtained from September 2008 to March 2013 recorded at WSCH. All subjects fasted overnight for 12 h before blood collection, and the levels of four lipid profiles were measured on the day of blood collection. If a patient record contained multiple test results of the collection during the study period, only the first result was chosen in the dataset. The total cholesterol, triglyceride, HDL-C, and LDL-C for the Korea hospital dataset were analyzed using the modular DPE system (Roche Diagnostics, Basel, Switzerland) [1]. The lipoprotein subfraction test was performed by using the Lipoprint LDL System (Quantimetrix, Redondo Beach, CA, USA) [1]. The Lipoprint system utilizes polyacrylamide gel electrophoresis to separate the various lipoprotein subfractions on the basis of particle size as VLDL band, three mid-bands [MID-C = VLDL remnants, MID-B = large intermediate density lipoprotein (IDL), and MID-A = small IDL], and seven LDL bands and an HDL band. LDL-1 and -2 are defined as large LDL subfractions whereas LDL-3, 4, 5, 6, and -7 are defined as small LDL subfractions [1,3,4].

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.12.009.

References

[1] T.S. Lee, J.W. Kim, Y. Uh, H.J. Lee, Deep neural network for estimating low density lipoprotein cholesterol, Clin. Chim. Acta (2018), https://doi.org/10.1016/j.cca.2018.11.022.

[2] S.S. Martin, M.J. Blaha, M.B. Elshazly, P.P. Toth, P.O. Kwiterovich, R.S. Blumenthal, et al., Comparison of a novel method vs the Friedewald equation for estimating low-density lipoprotein cholesterol levels from the standard lipid profile, JAMA 310 (2013) 2061–2068. https://doi.org/10.1001/jama.2013.280532.

[3] H. Arai, Y. Kokubo, M. Watanabe, T. Sawamura, Y. Ito, A. Minagawa, et al., Small dense low-density lipoproteins cholesterol can predict incident cardiovascular disease in an urban Japanese cohort: the Suita study, J. Atheroscler. Thromb. 20 (2013) 195–203. https://doi.org/10.5551/jat.14936.

[4] R.C. Hoogeveen, J.W. Gaubatz, W. Sun, R.C. Dodge, J.R. Crosby, J. Jiang, et al., Small dense low-density lipoprotein-cholesterol concentrations predict risk for coronary heart disease: the Atherosclerosis Risk In Communities (ARIC) study, Arterioscler. Thromb. Vasc. Biol. 34 (2014) 1069–1077. https://doi.org/10.1161/ATVBAHA.114.303284.