

OP-D-20

MACHINE LEARNING-DERIVED LOW-DENSITY LIPOPROTEIN CHOLESTEROL (LDL-C) ESTIMATION AGREES BETTER WITH DIRECTLY MEASURED LDL-C THAN CONVENTIONAL EQUATIONS IN INDIVIDUALS WITH TYPE 2 DIABETES MELLITUS

<https://doi.org/10.15605/jafes.038.AFES.23>

Gerald Sng, Khoo You Liang, Tan Hong Chang, Bee Yong Mong
Singapore General Hospital, Singapore

INTRODUCTION

Elevated low-density lipoprotein cholesterol (LDL-C) is an important risk factor for atherosclerotic cardiovascular disease (ASCVD). Direct LDL-C measurement is not widely performed. LDL-C is typically estimated using the Friedewald (FLDL), Martin-Hopkins (MLDL), or Sampson (SLDL) equations, which may be inaccurate at high triglycerides (TG) or low LDL-C levels. We aimed to determine if machine learning (ML)-derived LDL-C levels agree better with direct LDL-C than conventional equations in patients with type 2 diabetes mellitus (T2DM).

METHODOLOGY

We performed a retrospective cohort study on patients with T2DM from a multi-institutional diabetes registry in Singapore from 2013 to 2020. Directly measured LDL-C values were compared against LDL-C values estimated by the FLDL, MLDL, and SLDL equations, and ML models using linear regression (LR), random forest (RF) and k-nearest neighbours (KNN) using measures of agreement and correlation. Values were considered discordant if the estimated LDL-C was 4.5 mmol/L.

RESULTS

There were 11,475 patients with 39,417 sets of unique lipid panel results included in the final analysis. In the training set, 31,533 sets of results were used and 7,884 sets of results were used in the test set. All three ML models demonstrated better goodness-of-fit with lower root-mean-square-error values than any of the conventional equations, as well as stronger correlation with higher R² and r values. Of the three ML models, LR performed the least well (rmse 0.231, R² 0.954 and r 0.977, $p < 0.001$) as compared to RF (rmse 0.209, R² 0.962 and r 0.981, $p < 0.001$) or KNN (rmse 0.212, R² 0.961 and r 0.98, $p < 0.001$). All three ML methods had much lower discordance rates (LR 2.17%, RF 2.18%, KNN 2.04%) than conventional equations (FLDL 23.14%, SLDL 17.90%, MLDL 14.22%). ML methods performed less well in the subset of patients with TG >4.5 mmol/L, although all three models still demonstrated better goodness of fit and correlation. Discordance rates were lower as well (LR 3.69%, RF 3.69%, KNN 2.30%), although the MLDL equation had the lowest discordance rate in this subgroup (1.84%).

CONCLUSION

Conventional LDL-C estimation equations have disadvantages and are reported to perform poorly at high TG levels. ML methods may offer an alternative to allow more accurate estimation of LDL-C and to reduce misclassification and undertreatment in T2DM patients at high ASCVD risk.

KEYWORDS

low-density lipoprotein cholesterol, type 2 diabetes, machine learning