Research Article

Comparison of LDL-C Estimation Using Ridge Regression and Four Established Equations Against Direct Determination of LDL-C in a Northeastern Population in Thailand

Sirawich Sonsok[1], Pongdech Sarakarn[2,3*]

[1]Department of Medical Technology, Kosumphisai Hospital, Mahasarakham, Thailand
[2]ASEAN Cancer Epidemiology and Prevention Research Group (ACEP), Faculty of Public Health, Khon Kaen University, Khon Kaen, Thailand
[3]Department of Epidemiology and Biostatistics, Faculty of Public Health, Khon Kaen University, Khon Kaen, Thailand

Article Info

***Corresponding Author:**
Assoc. Prof. Pongdech Sarakarn
Faculty of Public Health, Khon Kaen University
123 Mittraphap Road, Muang District, Khon Kaen 40002, Thailand
Phone: +66-43-202222
E-mail: spongd@kku.ac.th

Keywords

Low-density lipoprotein cholesterol (LDL-C), Ridge regression, Machine learning, LDL-C calculation equations, Lipid profile Classification Description: Clinical chemistry, Public health

Abstract

Background: Equations traditionally used for estimating low-density lipoprotein cholesterol (LDL-C) have limitations in accuracy and reliability. This study aimed to compare the performance of established equations with a machine learning approach to determine the most appropriate method for LDL-C estimation.

Methods: A retrospective cross-sectional study was conducted using 14,109 lipid profile records from inpatients and outpatients at Kosumphisai Hospital, Northeastern Thailand (2017–2021). LDL-C was estimated using the Friedewald, Puavilai, National Institutes of Health (NIH), and Martin equations, as well as a Ridge regression model. Direct LDL-C measurement served as the reference standard. Model performance was evaluated using mean absolute error (MAE), the proportion of estimates within ±12% of the direct measurement, and Bland–Altman analysis.

Results: The calculation of LDL-C using Ridge regression provided the highest proportion of estimates within the ±12% error margin (75.37%), the lowest MAE (10.05 mg/dL), and the narrowest 95% limits of agreement (−31.19 to 31.57 mg/dL) in Bland–Altman analysis.

Conclusions: Ridge regression provided greater accuracy and reliability for LDL-C estimation compared with the four established equations. Future research should consider incorporating additional predictors and alternative penalized regression techniques, such as Lasso or Elastic Net, to enhance model robustness.

Introduction

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, accounting for approximately 31% of all global deaths, with ischemic heart disease being a significant contributor [1]. In Thailand, CVDs have been identified as a primary public health concern, contributing to a growing number of hospital admissions and healthcare costs [2]. The accurate assessment of lipid profiles, particularly low-density lipoprotein cholesterol (LDL-C), is essential for the early diagnosis, risk stratification, and management of CVDs.

LDL-C is widely recognized as a key biomarker in atherosclerosis progression and is used to guide lipid-lowering therapies, such as statins, to reduce cardiovascular risk [3]. While direct LDL-C measurement using ultracentrifugation or homogeneous assays provides accurate results, these methods are costly, require specialized equipment, and are not routinely available in resource-limited settings, particularly in community hospitals in Thailand.

Due to the limited availability of direct LDL-C measurement, several formulas have been developed to estimate LDL-C using commonly measured lipid parameters. The most widely used equation is the Friedewald equation (LDL-C = TC – HDL-C – TG/5), which assumes a fixed triglyceride-to-VLDL-C (very low-density lipoprotein cholesterol) ratio of 5 [4]. However, this formula is known to be inaccurate in individuals with high triglyceride levels (greater than 400 mg/dL), potentially leading to a misclassification of LDL-C levels.

Alternative equations, such as the Puavilai equation [5], the Martin-Hopkins equation, which utilizes an adaptive triglyceride-to-VLDL-C ratio [6], and the National Institutes of Health (NIH), USA, as the NIH equation [7], have been developed to enhance LDL-C estimation. However, discrepancies between these formulas have been observed, particularly in patients with metabolic syndrome, diabetes, or hypertriglyceridemia, making it challenging to determine the most suitable equation for different populations.

Despite advances in LDL-C estimation methods, no single equation works best for all patient groups, emphasizing the need for more flexible and precise approaches. Machine learning (ML) has emerged as a promising tool in healthcare, capable of detecting complex patterns in medical data [8]. Various ML models, such as random forests, support vector machines (SVMs), and deep learning, have shown better performance in LDL-C estimation compared to traditional equations in certain populations [9-13].

However, most existing studies have been conducted in Western populations, where genetic, dietary, and lifestyle factors differ significantly from those in Southeast Asia, particularly in Thailand. Moreover, ML-based LDL-C estimation models often require large datasets, advanced computational infrastructure, and model interpretability, which can limit their applicability in clinical practice. Additionally, few studies have compared ML models with multiple LDL-C equations within the same setting, creating a gap in understanding their relative performance in real-world scenarios.

An important but often overlooked aspect of LDL-C estimation is its potential effect on health disparities. Traditional LDL-C formulas can introduce bias in certain demographic groups, especially in patients with metabolic disorders, high triglycerides, or unique dietary habits. Likewise, ML models trained on datasets that lack diversity may perform differently when applied to various ethnic or socioeconomic populations. Addressing these disparities is essential to ensure that ML-based LDL-C estimation models are fair and applicable to all groups.

Considering the limitations of traditional LDL-C equations, the potential of ML models, and the lack of research in a Thai population, this study aims to develop and evaluate ML-based LDL-C estimation models using lipid profile data from a community hospital in Northeastern Thailand. Additionally, we compare the performance of the selected ML model with four traditional LDL-C equations (Friedewald, Puavilai, Martin-Hopkins, and NIH) using direct LDL-C measurements to assess their feasibility in clinical settings.

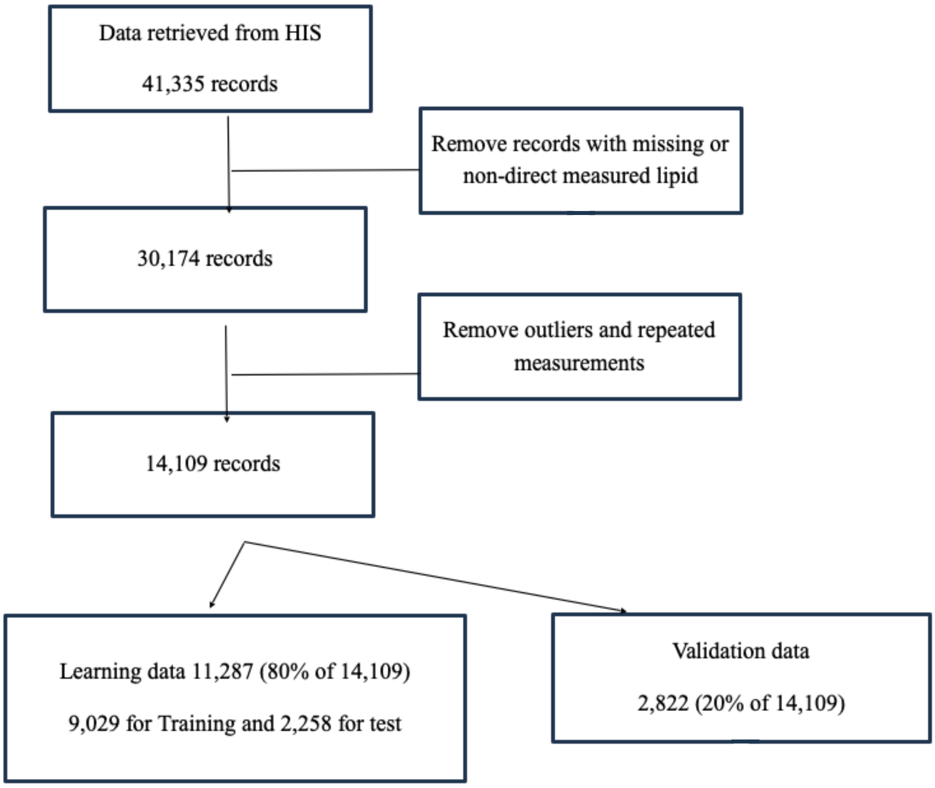
Material and Method

Study Design and Population

This study employed a retrospective cross-sectional design, using lipid profile datasets obtained from both inpatient and outpatient sections of Kosumphisai Hospital, Mahasarakham, a district hospital in the northeastern region of Thailand. Data were extracted from the Hospital Information System (HIS) and included lipid profile measurements from 14,109 individuals gathered between 2017 and 2021. The study population had a mean age of approximately 62 years, with age ranging from about 10–98 years in the training set, 10–93 years in the testing set, and 12–96 years in the validation set. Men accounted for roughly 38% of participants across all three datasets (38.5% in the training set, 37.7% in the testing set, and 38.4% in the validation set). The dataset comprised total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), and low-density lipoprotein cholesterol (LDL-C). Inclusion criteria included individuals with data on TC, TG, HDL-C, and LDL-C. In contrast, exclusion criteria involved individuals with missing lipid profile results and measurements exceeding the linearity limits of the specific assays used.

The entire data selection process, including record retrieval, exclusion steps, and dataset partitioning, is summarized in Figure 1.

Figure 1: Flowchart showing the data selection process, including inclusion/exclusion criteria and how the dataset is split for training and validation.



All lipid measurements were obtained from 12-hour fasting plasma samples, which were analyzed using an automated clinical chemistry analyzer (Abbott ARCHITECT, Illinois, USA) according to the manufacturer's instructions. TC and TG levels were measured using enzymatic colorimetric assays, while HDL-C and LDL-C were determined using homogeneous enzymatic assays. All the included patient data was retrieved from the direct LDL-C measurement using, two-step enzymatic colorimetric method. In the first step a polyanion reagent and detergent 1 selectively complex and mask non-LDL lipoproteins (chylomicrons, HDL, VLDL), and cholesterol released from these fractions is consumed by the enzymatic system in a non-chromogenic reaction to eliminate background signal. In the second step detergent 2 liberates cholesterol from the LDL fraction; the liberated LDL-cholesterol is hydrolyzed by cholesterol esterase and oxidized by cholesterol oxidase, and the resulting hydrogen peroxide is detected via peroxidase-catalyzed coupling with 4-aminoantipyrine/TOOS to generate a colored product. The absorbance was read photometrically (primary wavelength ~604 nm) at 37 °C and converted to concentration using HDL/LDL calibrators. The assay is linear from 5 to 600 mg/dL with a limit of detection near 4.5 mg/dL; fresh fasting serum was used, and reagents were handled according to manufacturer recommendations to ensure stability and accuracy [14]. Routine systematic quality control procedures were applied to ensure precision and accuracy.

LDL-C Calculation Equations

LDL-C was estimated using four mathematical equations:

1. Friedewald equation:

LDL-C (mg/dL) = TC – HDL-C – TG/5

2. Puavilai equation:

LDL-C (mg/dL) = TC – HDL-C – TG/6

3. NIH equation:

LDL-C (mg/dL) = (TC/0.948) – (HDL-C/0.971) − (TG/8.56) – [(TG*NonHDL-C)/2140] + (TG2 /16100) − 9.44

where NonHDL-C = TC – HDL-C

4. Martin-Hopkins equation:

LDL-C was determined using the Martin LDL calculator (LDL-Calculator.com).

The accuracy of these formulas was evaluated using direct LDL-C measurement as the reference method.

Principle of Ridge regression

Ridge regression is a regularized linear regression technique that minimizes the sum of squared errors while incorporating a penalty term (L2 regularization) to shrink the regression coefficients [15]. The objective function for ridge regression is defined as:

$$L(\beta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

where

y_i = observed LDL-C value
y_i = predicted LDL-C value
j = regression coefficients
λ = regularization parameter controlling shrinkage

The penalty term $(\lambda \sum_{j=1}^{p} \beta_j^2)$ prevents overfitting by reducing the magnitude of regression coefficients, making the model more stable when dealing with multicollinearity or small sample sizes [15].

Ridge regression was chosen as the primary machine learning method in this study for several reasons. First, it prevents overfitting through regularization, which reduces model complexity and enhances generalizability to unseen data [16]. This is particularly advantageous when predictor variables such as total cholesterol (TC), triglycerides (TG), and HDL-C tend to be correlated, as ridge regression effectively addresses multicollinearity. Second, the method reduces variance, yielding more stable coefficient estimates and decreasing sensitivity to noise in lipid measurements [17]. Finally, ridge regression is both interpretable and computationally efficient. Unlike more complex non-linear models (e.g., XGBoost or deep learning), it preserves transparency regarding variable contributions and is straightforward to implement in routine clinical practice [18].

Model Training and Evaluation

The remaining data was randomly split into 11,287 (80%) as learning data and 2,822 (20%) as validation data. Afterward, the 80% of 11,287 (n=9,029) was assigned as model training data and 20% of 11,287 (n=2,258) was assigned as the model testing data. Finally, the validation data was used to evaluate all LDL-C methods compared to direct measured LDL-C. The independent variables were TC, TG, and HDL-C, while direct LDL-C measurement was the target variable. A 10-fold cross-validation approach was used to tune the hyperparameter λ by partitioning the dataset into 10 subsets, training the model on 9 subsets while validating on the remaining subset in each iteration. Ridge regression was implemented using Python 3.9 and Scikit-learn 0.23.2.

Statistical Analysis

To assess the accuracy of LDL-C estimation methods, three statistical measures were used:

1. Mean Absolute Error (MAE)

MAE quantifies the absolute difference between estimated and directly measured LDL-C values [19]:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$

where:
n = number of data points
y_i = directly measured LDL-C
y_i = estimated LDL-C

A lower MAE indicates a more accurate calculation method.

2. Bland-Altman Analysis

Bland-Altman analysis was used to assess the agreement between estimated and directly measured LDL-C values base on how much the differences between the calculated LDL-C and direct measured LDL-C (20). The mean difference (bias) and limits of agreement (LOA = mean difference ± 1.96 SD) were reported to determine systematic bias and variability between methods.

3. Percentage Within 12% Error of Direct LDL-C

To assess the clinical acceptability of the LDL-C estimation models, the amount of data of estimates within 12% of direct LDL-C measurements was calculated as:

$$\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100 \leq 12$$

where:

yi = directly measured LDL-C
yi = estimated LDL-C

A higher number and percentage of calculated LDL-C data within 12% differed from directly measured LDL-C, indicating better clinical agreement. This metric aligns with clinical practice guidelines that suggest LDL-C estimations should be within 12% of direct measurements for reliable cardiovascular risk assessment [21, 22].

Results

The baseline characteristics of the study population are outlined in Table 1. The learning dataset (n = 11,287) was split into a training set (n = 9,029) and a testing set (n = 2,258), while the validation dataset included 2,822 participants. The average age was 61.5 ± 11.7 years in the training set, 61.6 ± 11.6 years in the testing set, and 61.9 ± 11.8 years in the validation set, with similar age ranges across datasets (10–98, 10–93, and 12–96 years, respectively). The percentage of male participants was comparable across groups, representing 38.5% (n = 3,480) in the training set, 37.7% (n = 851) in the testing set, and 38.4% (n = 1,083) in the validation set.

Table 1: Baseline characteristics of the study population.

Variables	**Learning data (n = 11,287)**				**Validation data (n = 2,822)**	
	Training (n = 9,029)		**Testing (n = 2,258)**			
	Mean ± SD	**Range**	**Mean ± SD**	**Range**	**Mean ± SD**	**Range**
Age (years)	61.5 ± 11.7	10–98	61.6 ± 11.6	10–93	61.9 ± 11.8	12–96
Male (n, %)	3480 (38.5%)	-	851 (37.7%)	-	1083 (38.4%)	-
TC (mmol/L)	4.87 ± 1.23	1.37–17.48	4.82 ± 1.18	1.37–16.64	4.83 ± 1.18	1.34–20.67
TG (mmol/L)	2.06 ± 1.57	0.27–21.32	2.05 ± 1.50	0.46–15.76	2.08 ± 1.63	0.49–21.11
HDL-C (mmol/L)	1.16 ± 0.31	0.28–3.72	1.15 ± 0.32	0.28–3.10	1.16 ± 0.32	0.28–4.13
Direct LDL-C (mmol/L)	2.96 ± 1.06	0.28–13.33	2.94 ± 1.00	0.49–8.98	2.94 ± 1.00	0.31–9.12

TC: Total Cholesterol; LDL-C: Low-Density Lipoprotein Cholesterol; TG: Triglycerides; HDL-C: High-Density Lipoprotein Cholesterol; Direct LDL-C: Directly measured LDL-C.

Lipid profile distributions were similar across datasets, with minor variations in total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), and low-density lipoprotein cholesterol (LDL-C) levels. Table 1 presents the detailed distributions. The mean TC levels were 4.85 ± 1.23 mmol/L in the training set, 4.82 ± 1.18 mmol/L in the testing set, and 4.83 ± 1.18 mmol/L in the validation set, with respective ranges of 1.37–17.48 mmol/L, 1.37–16.64 mmol/L, and 1.34–20.67 mmol/L. The mean TG levels were 2.06 ± 1.57 mmol/L in the training set, 2.05 ± 1.50 mmol/L in the testing set, and 2.08 ± 1.63 mmol/L in the validation set, with respective ranges of 0.27–21.32 mmol/L, 0.46–15.76 mmol/L, and 0.49–21.11 mmol/L. The mean HDL-C levels were 1.16 ± 0.31 mmol/L in the training set, 1.15 ± 0.32 mmol/L in the testing set, and 1.16 ± 0.32 mmol/L in the validation set, with corresponding ranges of 0.28–3.72 mmol/L, 0.28–3.10 mmol/L, and 0.28–4.13 mmol/L. Lastly, direct LDL-C levels remained consistent across datasets, with means of 2.96 ± 1.06 mmol/L in the training set, 2.94 ± 1.00 mmol/L in the testing set, and 2.94 ± 1.00 mmol/L in the validation set, with respective ranges of 0.28–13.33 mmol/L, 0.49–8.98 mmol/L, and 0.31–9.12 mmol/L. The consistency of lipid profile values across datasets supports the reliability of model development, minimizing dataset-related bias.

The performance of the LDL-C estimation models was assessed using Mean ± Standard Deviation (SD) and Mean Absolute Error (MAE), as summarized in Table 2.

Table 2: Comparison of LDL-C estimation methods (Validation data, n = 2,822).

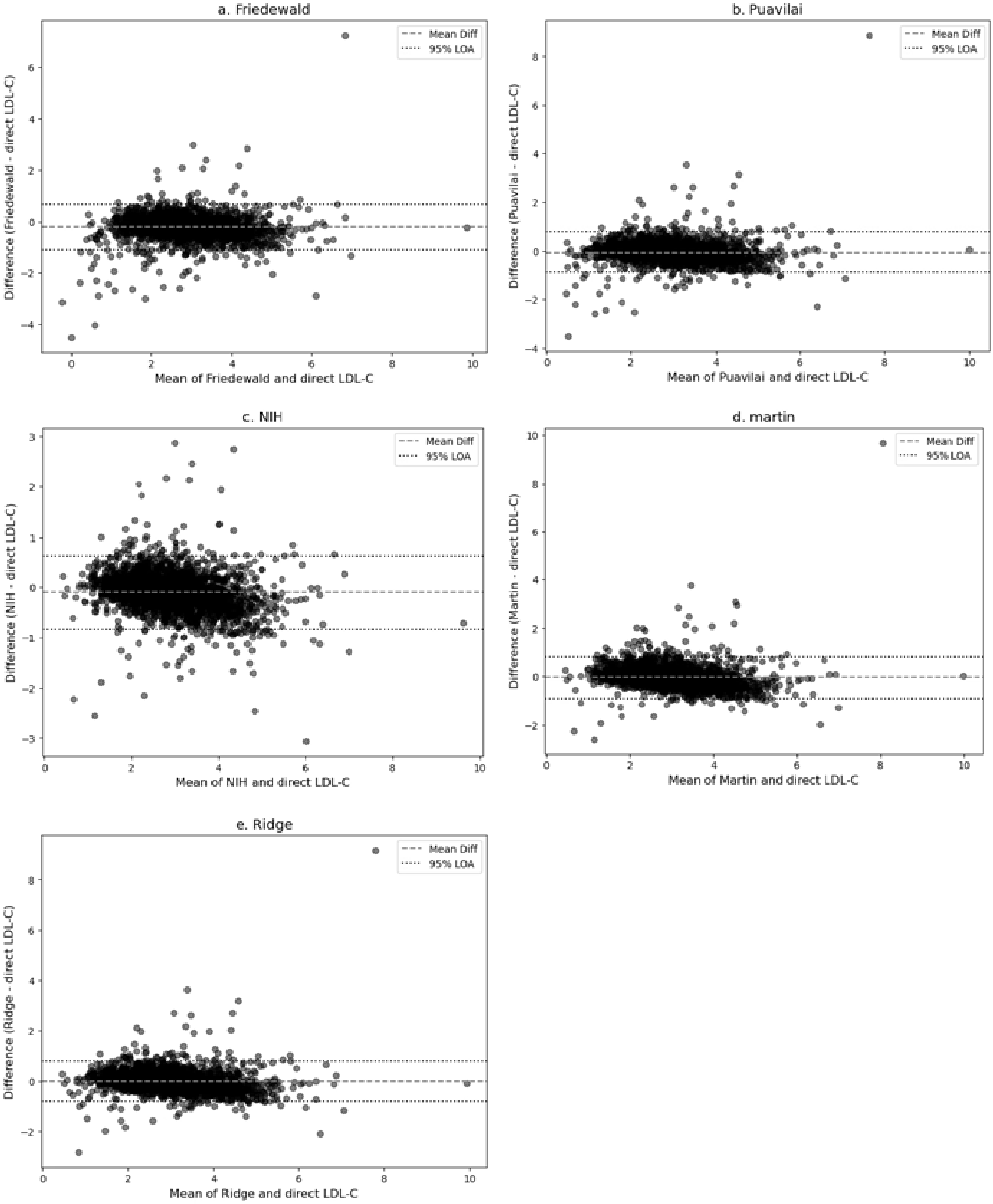
Method	**Mean ± SD (mmol/L)**	**MAE (mmol/L)**
Friedewald	2.72 ± 0.96	0,34
Puavilai	2.88 ± 0.95	0,27
NIH	2.83 ± 0.89	0,27
Martin	2.90 ± 0.90	0,28
Ridge regression	2.93 ± 0.92	0,26

MAE: Mean Absolute Error, calculated as the absolute difference between estimated and direct LDL-C values.

Among traditional estimation methods, the Friedewald equation showed the lowest accuracy, with the highest MAE (0.34 mmol/L). In contrast, the Ridge Regression model, developed through machine learning, outperformed all traditional equations, achieving the lowest MAE (0.26 mmol/L), which indicates improved estimation accuracy.

To assess the agreement between estimated LDL-C values and direct measurements, Bland-Altman analysis was performed for each estimation method (Table 3, Figure 2a–e). This analysis evaluated mean bias (the difference between estimated and direct LDL-C values) and 95% limits of agreement (LOA). The Ridge Regression model showed the smallest mean bias (-0.005 mmol/L), indicating its LDL-C estimates closely matched the direct measurements. Conversely, the Friedewald equation exhibited the largest deviation, with a mean bias of -0.213 mmol/L, implying consistent underestimation of LDL-C levels. The 95% LOA indicates the range within which 95% of the differences between estimated and direct LDL-C values fall. Ridge Regression had a fairly narrow LOA (-0.807 to 0.816 mmol/L), suggesting lower variability and better precision, while the Friedewald equation had the widest range (-1.096 to 0.671 mmol/L), reflecting greater inconsistency.

Figure 2 a–e: Bland-Altman plots comparing LDL-C estimation methods with direct LDL-C measurement (Validation data, n=2,822).



a. Friedewald

b. Puavilai

c. NIH

d. martin

e. Ridge

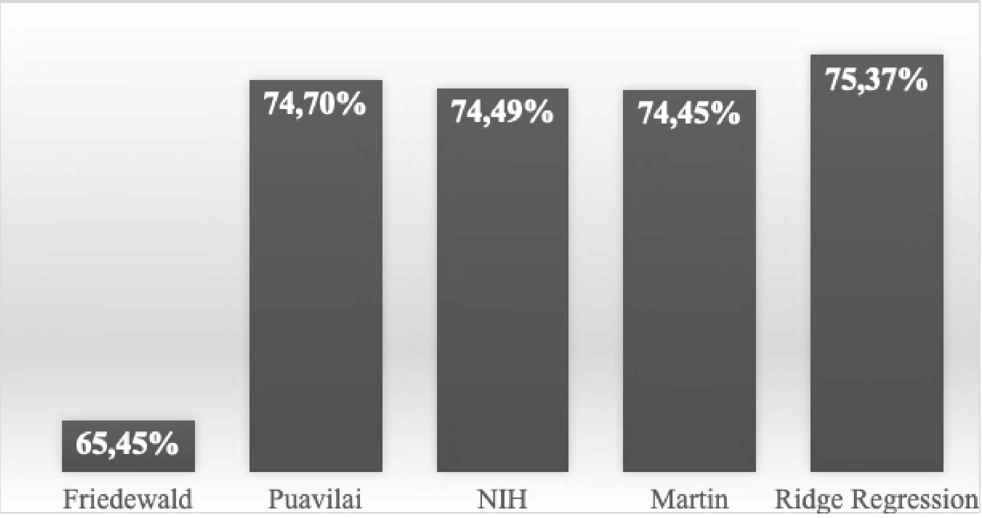
Table 3: Bland–Altman analysis (Validation data, n = 2,822).

Method	Mean Diff (mmol/L)	Lower LOA (mmol/L)	Upper LOA (mmol/L)
Friedewald	-0,213	-1,096	0,671
Puavilai	-0,054	-0,885	0,778
NIH	-0,099	-0,833	0,634
Martin	-0,035	-0,897	0,827
Ridge regression	-0,005	-0,807	0,816

LOA (Limits of Agreement): The range within which 95% of the differences between estimated and directly measured LDL-C values fall. Mean Diff: The average difference between estimated LDL-C values and direct LDL-C measurements. A lower bias indicates closer agreement with the reference measurement.

The percentage of LDL-C estimates within ±12% error was analyzed as a measure of clinically acceptable agreement with direct LDL-C measurement.

Figure 3: Percentage of calculation results within 12% deviated from direct LDL-C.



In Figure 3, the Friedewald equation (65.45%) had the lowest accuracy, showing greater variability. In contrast, the Puavilai (74.70%), NIH (74.49%), and Martin (74.45%) equations demonstrated moderate improvements. Ridge Regression (75.37%) had the highest accuracy; thus, machine learning-based approaches may improve the accuracy of LDL-C estimation. These findings highlight the benefits of alternative LDL-C estimation methods over the Friedewald equation, particularly in improving cardiovascular risk assessment.

Discussion

This study assessed the performance of five low-density lipoprotein cholesterol (LDL-C) estimation methods, including four traditional equations and a machine learning-based Ridge regression model, in a Northeastern Thai population. Our results show that Ridge regression outperforms traditional formulas, offering more accurate and dependable LDL-C estimates.

Consistent with previous studies, our results highlight the limitations of traditional LDL-C estimation methods, particularly the Friedewald equation. Despite its widespread use due to its simplicity, the Friedewald equation is known to underestimate LDL-C in individuals with high triglyceride (TG) levels or low LDL-C concentrations [11, 13, 23-25]. This underestimation is clinically concerning, as it may lead to non-detection of cardiovascular risk and suboptimal treatment decisions. Alternative equations, such as the Martin-Hopkins and Sampson-NIH equations, have attempted to address these limitations by incorporating more sophisticated TG: VLDL-C ratio adjustments or refined algorithms for very-low-density lipoprotein cholesterol (VLDL-C) estimation [6,7]. However, despite these improvements, our study demonstrates that Ridge regression provides superior accuracy and reliability in LDL-C estimation.

The effectiveness of machine learning (ML) models for LDL-C estimation has been corroborated by recent studies. For instance, research in Turkish pediatric populations found that ML models produced more concordant LDL-C estimates than traditional formulas [10]. Similarly, a study in Eastern India found that ML approaches, including XGBoost and Random Forests, outperformed conventional formulas in predicting LDL-C levels [9]. These findings align with our results, supporting the potential of ML-driven LDL-C estimation across diverse populations. Accurate LDL-C estimation is essential for cardiovascular risk assessment and optimizing lipid-lowering therapies [24]. Errors in LDL-C estimation can lead to misclassification of risk categories, increasing the likelihood of overtreatment or undertreatment. Our findings suggest that integrating Ridge regression into clinical workflows could improve LDL-C assessment precision, enabling more precise treatment strategies.

Machine learning methods often face challenges such as overfitting, multicollinearity, and interpretability. Ridge regression was chosen in this study because it can address these issues while keeping clinical relevance. Unlike deep learning models, Ridge regression provides transparency in how variables contribute, making it easier for clinicians to interpret. It also effectively handles collinear predictors, which is especially important when working with lipid profile data where some variables like total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), and LDL-C tend to be correlated. The regularization techniques in Ridge regression help ensure the model remains stable, reducing the risk of excessive prediction variance and improving its ability to generalize [16-18].

While our study offers valuable insights into the clinical use of ML-based LDL-C estimation, several limitations need to be acknowledged. First, our findings are based on a Northeastern Thai population, which may limit their applicability to other ethnicities or regions. Second, seasonal variations in lipid levels, influenced by diet and lifestyle factors, may have affected LDL-C estimates [24]. Third, all LDL-C measurements were obtained from a single laboratory, which could affect the external validity of our results in multi-center settings. Additionally, the timing of lipid testing is often driven by institutional practices, such as annual health checkups or screening programs, which can differ between countries. In Thailand, for example, these services often occur at the start of the fiscal year or during specific seasons, leading to seasonal overrepresentation in the dataset. This clustering may introduce temporal bias into the model's training and evaluation. If other countries follow different testing schedules, the performance of the same LDL-C estimation methods might vary due to differences in seasonal data structures. Therefore, context-specific validation is essential when applying machine learning-based LDL-C models across different healthcare systems.

To further establish the clinical usefulness of regression models, future research should focus on external validation using independent datasets from diverse healthcare settings. This strategy would help confirm the model's ability to work across different patient groups. Improving LDL-C prediction accuracy might be possible by adding more biomarkers, genetic data, or patient-specific clinical parameters. Additionally, examining the robustness of other penalized regression techniques such as Lasso and ElasticNet could enhance model performance across various subgroups. Prospective studies assessing how machine learning–based LDL-C estimation influences clinical decision-making and long-term cardiovascular outcomes would provide valuable insights into its practical application.

Conclusion

Our study shows that a machine learning-based Ridge regression model offers better accuracy in LDL-C estimation compared to traditional equations. This increased precision has important implications for cardiovascular risk assessment and lipid-lowering therapy management. By incorporating machine learning methods into clinical workflows, healthcare professionals can improve LDL-C evaluation, leading to more personalized and effective treatment strategies.

Declaration of Generative AI and AI-assisted technologies in the writing process

We would like to declare that generative AI (ChatGPT-5) was used solely for assistance in checking and refining the English language in this manuscript, including minor translations. The authors entirely generated the content, ideas, and findings presented in the manuscript without AI assistance. After language editing, the authors reviewed and validated the final version to ensure its accuracy and integrity.

Conflict of Interest

The authors declare no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical approval

The study was approved by the Khon Kaen University Ethics Committee for Human Research (Approval No: HE651263). Informed consent was not required as no human samples were collected directly for this study.

References

1. World Health Organization. Cardiovascular diseases (CVDs). 2021 [cited 2024 Oct 31]. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases.
2. World Health Organization. WHO Data Repository. 2023 [cited 2024 Oct 31]. Available from: https://data.who.int/countries/.
3. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 Cholesterol Clinical Practice Guidelines: Synopsis of the 2018 AHA/ACC/Multi-Society Cholesterol Guideline. Ann Intern Med. 2019;170(11):779-783.
4. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin

Chem. 1972;18(6):499-502.
5. Puavilai W, Laorugpongse D, Deerochanawong C, Muthapongthavorn N, Srilert P. The accuracy of using a modified Friedewald equation to calculate LDL from non-fasting triglycerides: a pilot study. Med Assoc Thai. 2009;92(2):182-188.
6. Martin SS, Blaha MJ, Elshazly MB, Toth PP, Kwiterovich PO, Blumenthal RS, et al. Comparison of a novel method vs. the Friedewald equation for estimating LDL-C levels from the standard lipid profile. JAMA. 2013;310(19):2061-2068.
7. Sampson M, Ling C, Sun Q, Harb R, Ashmaig M, Warnick R, et al. A new equation for LDL-C calculation in patients with normolipidemia and hypertriglyceridemia. JAMA Cardiol. 2020;5(5):540-548.
8. Krittanawong C, Virk HUH, Bangalore S, Wang Z, Johnson KW, Pinotti R, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. Sci Rep. 2020;10(1):16057.
9. P P A, Kumari S, Rajasimman AS, Nayak S, Priyadarsini P. Machine learning predictive models of LDL-C in an Eastern Indian population and comparison with directly measured LDL-C. Ann Clin Biochem. 2021;59(1):76-86.
10. 10. Koçhan N. Estimation of LDL-C using machine learning models and its comparison with directly measured and calculated LDL-C in Turkish pediatric population. Abant Med J. 2023;12(1):63-75.
11. Oh GC, Ko T, Kim JH, Lee MH, Choi SW, Bae YS, et al. Estimation of low-density lipoprotein cholesterol levels using machine learning. Int J Cardiol. 2022;352:144-149.
12. Ghayad JP, Barakett-Hamadé V, Sleilaty G. Prospective validation of a machine learning model for LDL-C estimation. Lab Med. 2022;53(6):629-635.
13. Paydaş Hataysal E, Körez MK, Yeşildal F, İşman FK. A comparative evaluation of LDL-C estimation: Machine learning algorithms vs. traditional equations. Clin Chim Acta. 2024; 557:117853.
14. Archem Sağlık Sanayi ve Tic. A.Ş. LDL Direct Cholesterol. Instructions for use. REF 02R05-31/02R05-21. Rev V3.5, 08.2021. İstanbul (Turkey): Archem Sağlık Sanayi ve Tic. A.Ş.; 2021. p.1–4.
15. Safi S, Alsheryani M, Alrashdi M, Suleiman R, Awwad D, Abdalla Z. Optimizing linear regression models with Lasso and Ridge Regression: A study on UAE financial behavior during COVID-19. Migr Lett. 2023;20(6):139-153.
16. Tsigler A, Bartlett PL. Benign overfitting in ridge regression. Journal of Machine Learning Research. 2023;24:1–76.
17. Akhtar N, Alharthi MF. A comparative study of the performance of new ridge estimators for multicollinearity: Insights from simulation and real data application. AIP Adv. 2024;14(11):115311. doi:10.1063/5.0236631
18. Ridge Regression. Columbia University Mailman School of Public Health. Population Health Methods. [Internet]. [cited 2025 Dec 1]. Available from: https://www.publichealth.columbia.edu/research/population-health-methods/ridge-regression
19. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing model performance. Clim Res. 2005;30(1):79-82.
20. Bland JM, Altman D. Statistical methods for assessing agreement between two clinical measurement methods. Lancet. 1986;327(8476):307-310.
21. Dintshi M, Kone N, Khoza S. Comparison of measured LDL-C with calculated LDL-C using the Friedewald and Martin-Hopkins formulae in diabetic adults. PLoS One. 2022;17(12):e0277981.
22. National Cholesterol Education Program. Third report of the NCEP Expert Panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). Bethesda, MD: National Heart, Lung, and Blood Institute; 2002.
23. Chen L, Rong C, Ma P, Li Y, Deng X, Hua M. A new equation for estimating LDL-C concentration based on machine learning. Medicine (Baltimore). 2024;103(15):e37766.
24. Packard C, Chapman MJ, Sibartie M, Laufs U, Masana L. Intensive LDL-C lowering in cardiovascular disease prevention: Opportunities and challenges. Heart. 2021;107(17):1369-1375.
25. Ma X, Yan H, Zhang H, Wang M, Zhang Q, Zhou X. Seasonal variations of blood lipids: a mini-review. Lipids Health Dis. 2020;19(1):1-8.