Research article

DOI: 10.59715/pntjmp.4.3.14

Correlation, Agreement, and Treatment Misclassification of Sampson and Friedewald Equations for Estimated LDL-Cholesterol Across Triglyceride Levels

Trinh Thi Ngoc Ai[1], Nguyen Van Thanh[2], Le Huu Dung[3], Thai Huu Khoa[1]

[1] Department of Laboratory Medicine, Faculty of Nursing and Medical Technology, Pham Ngoc Thach University of Medicine, Ho Chi Minh City, Vietnam

[2] Faculty of Medical Laboratory, City Children's Hospital, Ho Chi Minh City - Vietnam

[3] Faculty of Medical Laboratory, Dai Phuoc Clinic, Ho Chi Minh City, Vietnam

Abstract

Objectives: Concordance between calculated and directly measured LDL-C levels, as assessed by various measures of test accuracy (correlation coefficient, Bland-Altman plot and percentage of patients misclassified at LDL-C treatment thresholds).

Methods: The study was conducted on data from 1277 subjects who measured all four lipid tests (triglycerides, total cholesterol, LDL-C, HDL-C). The LDL-C level measured by β-quantification derived method and the calculated LDL-C using formulas (Sampson, Friedewald) were analyzed for Pearson correlation with the correlation coefficient and regression equations. The agreement was assessed using Bland-Altman plots, and the percentage of patients misclassified at LDL-C treatment thresholds was evaluated using Cohen's Kappa statistic.

Results: The Sampson equation exhibited a stronger correlation compared to the Friedewald equation ($r = 0.892$, $p < 0.05$, vs. $r = 0.814$, $p < 0.05$). In the triglyceride (TG) < 200 mg/dL group, the Sampson and the Friedewald equation demonstrated high agreement, with %Mean LDLdiff values of 3.18% and 5.19%, respectively. In the $200 \leq TG < 400$ mg/dL group, only the Sampson formula maintained an acceptable level of agreement (%Mean LDLdiff = 11.98%), whereas the Friedewald formula showed a higher %Mean LDLdiff of 18.78%. In the highest TG group, both formulas exhibited significant discrepancies, with %Mean LDLdiff values of 27.93% and 43.05%, respectively.

Conclusions: LDL-C calculated using both the Sampson and Friedewald formulas exhibits a strong correlation with direct LDL-C. The Sampson equation maintains an acceptable level of agreement in the TG < 400 mg/dL group, whereas the Friedewald equation is only acceptable in the TG < 200 mg/dL group. The correlation and agreement of both formulas are inversely proportional to TG levels.

Keywords: LDL-C, Sampson, Friedewald

Received: 21/02/2025
Revised: 10/3/2025
Accepted: 20/7/2025

Author contact:
Trinh Thi Ngoc Ai
Email: aittn@pnt.edu.vn
Phone: +84767517443

1. INTRODUCTION

According to the 2018 guidelines from the American Heart Association/American College of Cardiology (AHA/ACC), to reduce the risk of cardiovascular disease and atherosclerosis through cholesterol management, LDL-C levels are considered a target parameter [1]. Therefore, accurately measuring LDL-C concentration is essential and crucial for the treatment and monitoring of cardiovascular diseases and lipid

disorders. The β-quantification method is the most commonly used approach for this measurement. However, this technique has drawbacks, such as being time-consuming and costly (equipment, reagents, calibration and control solution). As a result, several alternative formulas have been proposed to estimate LDL-C levels. The most classical and longstanding formula is Friedewald formula (1972), which addressed some of the limitations of the β-quantification method, but it still has several drawbacks: patients must fast, it cannot be applied when TG levels exceed 400 mg/dL, it is inaccurate when LDL-C levels are low and in type III hyperlipidemia [2]. Many recent studies have proposed new formulas to overcome these limitations, including the Sampson formula (2020), which has been shown to be reliable for TG levels ≤ 800 mg/dL [3]. However, the Sampson formula is not yet widely used in Vietnam, and there has been limited research on its application, particularly in terms of TG levels. Therefore, this study aims to assess the concordance between estimated and measured LDL-C levels.

2. MATERIALS AND METHODS

2.1. Study design and participants

This cross-sectional study was conducted on 1277 patients who visited the Dai Phuoc clinic from September 2022 to February 2023. The study selected patients who met the following criteria: (1) the laboratory test results included all four parameters (TG, total cholesterol (TC), LDL-C, HDL-C); (2) TG ≤ 800 mg/dL; (3) all lipid tests were performed using the same AU640 automatic biochemical analyzer system (Beckman Coulter, USA) by enzymatic method; (4) Blood samples collected during fasting. Test results that failed internal quality control or external quality assessment, as well as patients with a documented history of lipid-lowering treatment, were excluded.

The research team used the sample size calculation formula to compare the two correlation coefficients between the two groups:

$$n_1 = n_2 \geq 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\frac{1}{2} \times \log_e \frac{1+r_1}{1-r_1} - \frac{1}{2} \times \log_e \frac{1+r_2}{1-r_2}} \right)^2 + 3$$

The components include:

z: constant related to type I and II errors

α : type I error (Choose α = 0.05) ⇒ z = 1.96

β : type II error (Choose β = 0.2 ⇔ power = 0.8) ⇒ z = 0.84

n: required research sample size yes

r: correlation coefficient (taken from previous study)

Choose r_1 = 0.896 (according to previous study by Li J et al.)[4]

Choose r_2 = 0.943 (according to previous study by Li J et al.)[4]

Therefore, n ≥ 164/group. The study design includes 3 main groups, so it is necessary to select at least 492 patients who meet the sample inclusion criteria.

2.2. Procedure

Step 1: Collect the patient test results from the laboratory's Labconn software according to the selection and exclusion criteria.

Step 2: Calculate LDL-C using the Sampson equation (mg/dL).

$LDL\text{-}C_S = (TC/0.948) – (HDL\text{-}C/0.971) – ((TG/8.56)+((TG \times Non\text{-}HDL\text{-}C)/2.140) – (TG^2/16100)) – 9.44$

Step 3: Calculate LDL-C using the Friedewald equation (mg/dL).

$LDL\text{-}C_F = TC – HDL\text{-}C – TG/5$

Step 4: Data processing

Enter and store the data in MS Excel 2019. Process the data using SPSS 20.0 and MedCalc 22.009.

2.3. Statistical analysis

2.3.1. Assess the correlation between direct LDL-C (LDL_D) and LDL-C estimated by Sampson equation (LDL_S), LDL-C estimated by Friedewald equation (LDL_F) using Pearson correlation analysis.

2.3.2. Evaluate the agreement between LDL_D and LDL_S, LDL_F using a Bland-Altman plot based on: The mean difference of LDL_D and LDL_S, LDL_F (Mean LDLdiff); The standard deviation of the LDLdiff series (SD LDLdiff); The percentage of the mean difference of LDL-C (%Mean LDLdiff), the percentage of samples outside the limits of agreement (LoA) range. According to NCEP ATP III, the acceptable percentage of the %Mean LDLdiff for a test measuring LDL-C concentration is within ±12% [5].

2.3.3. Evaluate the percentage of patients misclassified at LDL-C treatment thresholds using the Cohen's Kappa statistic when classifying patients into treatment groups based on LDL-C levels.

2.4. Ethics

This study has been approved by the Ethics Committee in Medical Research of Pham Ngoc Thach University of Medicine under the Approval Certificate No.855/TĐHYKPNT-HĐĐĐ dated April 20, 2023.

3. RESULTS

3.1. Study population

Table 1. Characteristics of the study population

Characteristics		Sex		Age			Total n *(%)*
		Male n *(%)*	Female n *(%)*	< 40 n *(%)*	40 – 60 n *(%)*	> 60 n *(%)*	
Grouping by TG (mg/dL)	TG < 200	335 *(49.2)*	149 *(25.0)*	108 *(35.8)*	285 *(42.0)*	91 *(30.6)*	484 *(37.9)*
	200 ≤ TG < 400	203 *(29.8)*	253 *(42.4)*	91 *(30.1)*	233 *(34.4)*	132 *(44.4)*	456 *(35.7)*
	400 ≤ TG ≤ 800	143 *(21.0)*	194 *(32.6)*	103 *(34.1)*	160 *(23.6)*	74 *(24.9)*	337 *(26.4)*
Total population n (%)		681 *(53.3)*	596 *(46.7)*	302 *(23.6)*	678 *(53.1)*	297 *(23.3)*	1277

There are 1277 results of consecutive patients included in the study from September 2022 to February 2023, categorized into three TG concentration ranges. Among these, the group with TG < 200 mg/dL accounted for the highest proportion (37.9%). Males

had a higher rate than females (53.3% vs. 46.7%), and the predominant age group was 40-60 years (53.1%).

3.2. The correlation between LDL_D and LDL_S, LDL_F

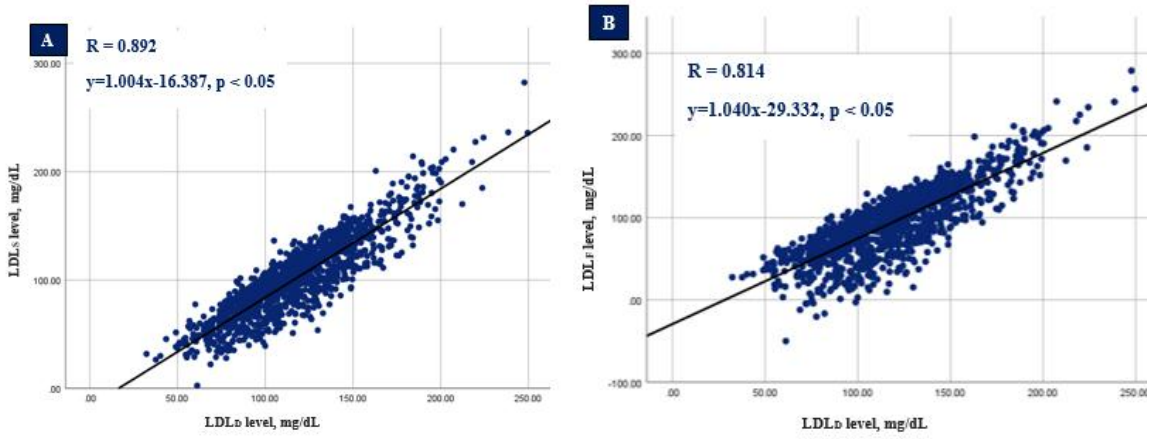

Figure 1. Regression lines between direct LDL-C and LDL-C values estimated with Sampson's and Friedewald's equation

A. Sampson equation B. Friedewald equation

The correlation between LDL_D and LDL_S, as well as LDL_F, was evaluated using the correlation coefficient (r) and regression equations. All the results, both the Sampson and the Friedewald equations demonstrated an acceptable correlation. However, the Sampson equation exhibited a stronger correlation compared to the Friedewald equation ($r = 0.892$, $p < 0.05$, compared to $r = 0.814$, $p < 0.05$).

Table 2. The correlation coefficient and regression equation between direct LDL-C and LDL-C values estimated by TG group

Group	Equation	Correlation (r)	Regression equation	p
TG < 200 mg/dL (n=484)	Sampson	0.976	y=1.134x-17.948	< 0.05
	Friedewald	0.971	y=1.117x-18.201	< 0.05
200 ≤ TG < 400 mg/dL (n=456)	Sampson	0.968	y=1.106x-27.758	< 0.05
	Friedewald	0.962	y=1.184x-44.984	< 0.05
400 ≤ TG ≤ 800 mg/dL (n=337)	Sampson	0.930	y=1.033x-39.417	< 0.05
	Friedewald	0.910	y=1.279x-89.649	< 0.05

The correlation coefficient was strongly positive, indicating a strong direct relationship, ranging from 0.910 to 0.976. The Sampson equation demonstrated a stronger correlation than the Friedewald equation within each TG group (r = 0.930 –

0.976, $p < 0.05$, compared to $r = 0.910 – 0.971$, $p < 0.05$). The highest correlation was observed in the TG < 200 mg/dL group ($r = 0.976$ and $r = 0.971$). The lowest correlation was found in the 400 ≤ TG ≤ 800 mg/dL group; however, it still indicated a strong correlation ($r = 0.930$ and $r = 0.910$).

3.3. The agreement between LDL_D and LDL_S, LDL_F

Table 3. Mean LDLdiff ± SD and %Mean LDLdiff values of equations

Group	Mean ± SD (mg/dL)			Mean LDL_{diff} ± SD (mg/dL)		%Mean LDL_{diff}	
	Sampson	Friedewald	Direct LDL-C	Sampson	Friedewald	Sampson	Friedewald
TG < 200 mg/dL (n=484)	104.64 ±35.80	102.47 ±35.45	108.08± 30.82	3.44 ±8.75	5.61 ±9.23	3.18	5.19
200 ≤ TG < 400 mg/dL (n=456)	105.98 ±35.31	98.25 ±38.06	120.97± 30.93	14.99 ±9.37	22.72 ±11.82	11.98	18.78
400 ≤ TG ≤ 800 mg/dL (n=337)	91.11 ±34.83	71.99 ±44.08	126.42± 31.36	35.31 ±12.86	54.43 ±20.26	27.93	43.05
Total population (n=1277)	101.55 ±35.90	92.92 ±40.80	117.52± 31.92	15.97 ±16.21	24.60 ±23.74	13.59	20.93

All the results and at each TG concentration level, showed that LDL_S exhibited greater agreement with LDL_D than LDL_F. The highest agreement was observed in the TG < 200 mg/dL group, with a %Mean LDLdiff of 3.18% (Sampson equation) and 5.19% (Friedewald equation). Conversely, the lowest agreement was found in the 400 ≤ TG ≤ 800 mg/dL group, with a %Mean LDLdiff of 27.93% (Sampson equation) and 43.05% (Friedewald equation).

3.3. The percentage of patients misclassified at LDL-C treatment thresholds

Table 4. Kappa coefficient of equations

Group	**Cohen's Kappa**	
	Sampson	Friedewald
TG < 200 mg/dL (n=484)	0.80	0.76
200 ≤ TG < 400 mg/dL (n=456)	0.50	0.35
400 ≤ TG ≤ 800 mg/dL (n=337)	0.02	0.00
Total population (n=1277)	0.47	0.38

Across the entire study as well as within each TG concentration subgroup, the kappa coefficient of the Sampson equation was consistently higher than that of the Friedewald equation. This indicates that the Sampson equation exhibited greater concordance in treatment classification based on LDL-C levels. However, for both equations, the level of agreement was inversely proportional to TG concentration, ranging from substantial agreement (κ-values of 0.8 and 0.76, respectively) in the TG < 200 mg/dL group to slight agreement (κ-values of 0.02 and 0.00, respectively) in the 400 ≤ TG ≤ 800 mg/dL group.

4. DISCUSSION

There is a lack of consistency in the classification of TG levels across previous studies, both domestically and internationally. A Consensus Statement from the European Atherosclerosis Society (EAS) and the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) has agreed that the Friedewald formula becomes increasingly inaccurate at TG concentrations ranging from 200 to 400 mg/dL (2.3–4.5 mmol/L) and is considered invalid when TG levels exceed 400 mg/dL (4.5 mmol/L) [6]. Additionally, Sampson et al. (2020) reported that the Sampson formula remains applicable for TG levels below 800 mg/dL [3]. Based on this foundation, the author categorized TG into three groups (TG <200 mg/dL, 201–400 mg/dL, and 401–800 mg/dL) to evaluate the correlation and agreement between the two aforementioned formulas.

4.1. The correlation between direct LDL-C and LDL-C values estimated with Sampson's and Friedewald's equation

Table 5. A Summary of the correlation and mean LDL-C deviation from selected similar studies conducted both domestically and internationally.

Author, year	Country	Equipment	Correlation (r)	
			Sampson	Friedewald
Maureen Sampson et al, 2020 [3]	US	-	0.965	0.881
Li et al, 2022 [4]	China	AU 5800 Beckman Coulter, USA	0.943	0.896
Alpdemir et al, 2024 [7]	Turkey	Roche Diagnostic, Indianapolis, IN, USA	0.905	0.897
Nga Le Hoang Bich et al, 2023 [8]	Vietnam	Cobas C702 - Roche	0.981	0.976
Thinh Tran Phuoc et al, 2024 [9]	Vietnam	AU680 Beckman Coulter, USA	0.890	0.850
This study	Vietnam	AU640 Beckman Coulter, USA	0.892	0.814

This study demonstrated strong positive correlations with LDL_D for both formulas, with the Sampson formula exhibiting a stronger correlation than the Friedewald formula across the entire study population as well as each TG subgroup (Figure 1, Table 2). This finding is consistent with the conclusions of other studies both domestically and internationally [3, 4, 7-9]. Notably, the correlation coefficient in this study was closest to that reported by Thinh Tran Phuoc et al. (2024), which may be attributed to the use of the same Beckman Coulter system (Table 5) [9].

When analyzed by TG subgroups, the correlation coefficient (r) exhibited a decreasing trend as TG concentrations increased for both formulas, ranging from 0.930 to 0.976 for LDL_S and from 0.910 to 0.971 for LDL_F. This trend is in agreement with the findings of Li et al. (2022), Alpdemir et al. (2024), and Thinh Tran Phuoc et al. (2024) [4, 7, 9].

4.2. The agreement between direct LDL-C and LDL-C values estimated with Sampson's and Friedewald's equation

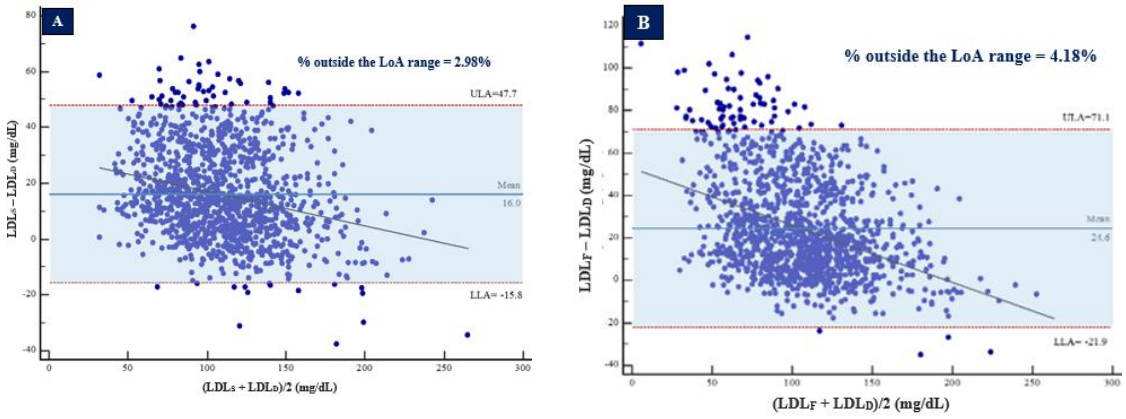

Figure 2. Bland-Altman plot between direct LDL-C and LDL-C values estimated with Sampson's and Friedewald's equation
(ULA= Upper limits of agreement; LLA= Lower limits of agreement)
A. Sampson equation B. Friedewald equation

Across the entire study population, the Sampson formula demonstrated greater agreement with direct LDL-C compared to the Friedewald formula across all metrics, including Mean LDLdiff, SD LDLdiff, and %Mean LDLdiff (Table 3). The Bland-Altman plot further supports this finding, as the percentage of samples falling outside the LoA for LDL_S was lower than that for LDL_F (2.98% vs. 4.18%). %). These results are consistent with the findings of Li et al. (2022), Alpdemir et al. (2024), and Nga Le Hoang Bich et al. (2023) [4, 7, 8].

When analyzed by TG groups, the discrepancy between direct LDL-C and estimated LDL-C values increased progressively across higher TG subgroups, consistent with findings from other studies. In the TG < 200 mg/dL group, both formulas met NCEP ATP III criteria, as %Mean LDLdiff <12%. In the 200 ≤ TG < 400 mg/dL group, while the Sampson formula met the NCEP ATP III criteria, the Friedewald formula did not (11.98% vs. 18.78%). This finding aligns with the results of previous global studies, which suggest that the Friedewald formula lacks accuracy at triglyceride (TG) concentrations exceeding 200 mg/dL [10-12]. In the highest TG groups, the Sampson formula exhibited better agreement than the Friedewald formula; however, neither met the NCEP ATP III criteria (27.93% and 43.05%). These findings are consistent with those reported by Li et al. (2022), Alpdemir et al. (2024), and Thinh Tran Phuoc et al. (2024) [4, 7, 9].

4.3. The percentage of patients misclassified at LDL-C treatment thresholds

Using Cohen's Kappa statistic, the agreement in treatment classification based on LDL-C decreased as TG levels increased. The Kappa coefficient for the Sampson formula was higher than that for the Friedewald formula. In the TG < 200

mg/dL and 400 ≤ TG < 800 mg/dL groups, both formulas demonstrated the same level of agreement, classified as substantial agreement and slight agreement, respectively. In the 200 ≤ TG < 400 mg/dL group, the Sampson formula showed moderate agreement, whereas the Friedewald formula exhibited fair agreement, one level lower (Table 4). In the TG < 200 mg/dL and 400 ≤ TG < 800 mg/dL groups, both formulas exhibited equivalent levels of agreement, categorized as substantial agreement and slight agreement, respectively. This finding aligns with the recommendation by Abdulrahman Naser et al. (2022) that when TG < 200 mg/dL, both equations demonstrate good concordance, whereas when TG > 400 mg/dL, neither of the calculated equations is accurate, and direct quantification should be employed instead [13]. In the 200 ≤ TG < 400 mg/dL group, the Sampson formula displayed moderate agreement, while the Friedewald formula showed fair agreement, one level lower. This observation is consistent with prior studies indicating that the Friedewald formula's accuracy diminishes when TG > 200 mg/dL [12], whereas the Sampson formula maintains an acceptable level of concordance [3].

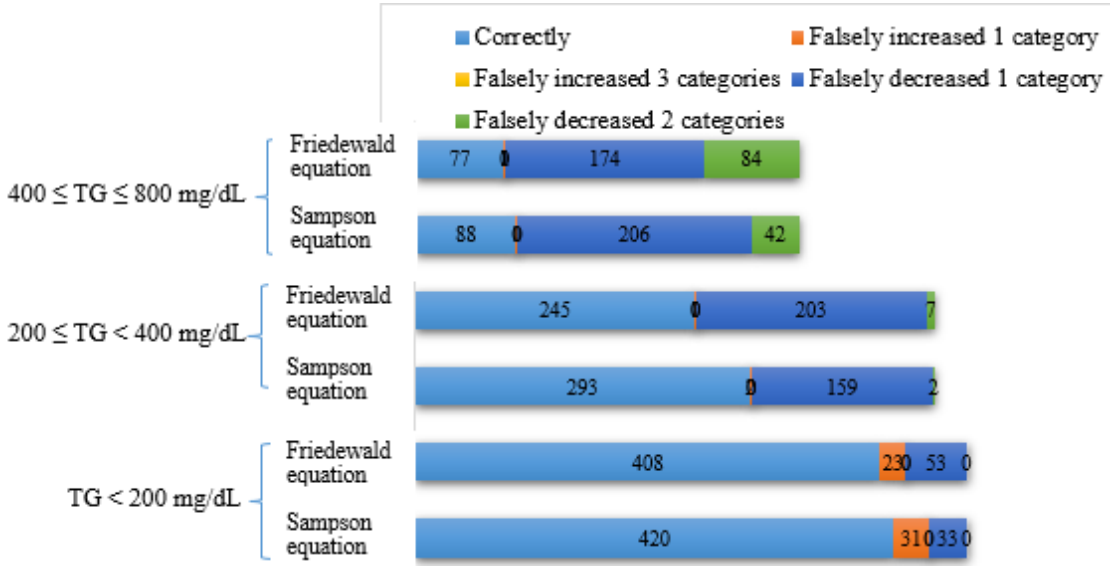

Chart 1. Misclassification for patients with TG levels (%)

When using the two formulas to classify treatment based on LDL-C, the Sampson formula provides more accurate classification across all TG groups. While misclassification in the TG < 200 mg/dL and 200 ≤ TG < 400 mg/dL groups primarily involves downward classification errors (mainly by one category), misclassification in the remaining group includes both downward and upward classification errors (Chart 1). This implies that when TG levels are below 400 mg/dL, the estimation formula may delay treatment decisions, whereas in the 400 ≤ TG ≤ 800 mg/dL group, it may lead to both overtreatment and delayed treatment.

5. CONCLUSION

LDL-C calculated using both the Sampson and Friedewald formulas exhibits a strong correlation with direct

LDL-C. In the TG < 200 mg/dL group, the two formulas demonstrate high agreement, resulting in minimal misclassification in treatment stratification. In the 200 ≤ TG < 400 mg/dL group, only the Sampson formula maintains an acceptable level of agreement with direct LDL-C and shows moderate agreement in treatment classification. In these two groups, misclassification errors tend to cause delays in treatment. In the highest TG group, both formulas show significant discrepancies, leading to frequent misclassification, which may result in both overtreatment and delayed treatment. The correlation and agreement of both formulas are inversely proportional to TG levels.

6. LIMITATIONS

This study was limited by time, manpower, and funding constraints, reducing its scope and depth. Its confined locations may have limited its representativeness across diverse populations. As a retrospective study using secondary data, controlling confounding factors was challenging. The authors addressed this by excluding non-fasting patients and those with prior lipid-lowering treatment to reduce bias. For more accurate and generalizable results, future studies should involve multiple facilities, diverse equipment, and a prospective design for comprehensive data collection.

REFERENCES

1. al., S.M.G.e., *2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines.* 2019. **73**(24): p. 3168-3209.
2. William T. Friedewald, R.I.L., and Donald S. Fredrickson, *Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge.* Clinical Chemistry, 1972. **18**(6): p. 499-502.
3. Sampson, M., et al., *A new equation for calculation of low-density lipoprotein cholesterol in patients with normolipidemia and/or hypertriglyceridemia.* 2020. **5**(5): p. 540-548.
4. Li, J., et al., *Evaluation of Sampson equation for LDL-C in acute coronary syndrome patients: a Chinese population-based cohort study.* 2022. **21**(1): p. 39.
5. Detection, N.C.E.P.E.P.o. and T.o.H.B.C.i. Adults, *Third report of the National Cholesterol Education Program (NCEP) Expert Panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III)*. 2002: The Program.
6. Langlois, M.R., et al., *Quantifying atherogenic lipoproteins: current and future challenges in the era of personalized medicine and very low concentrations of LDL cholesterol. A consensus statement from EAS and EFLM.* 2018. **64**(7): p. 1006-1033.
7. Alpdemir, M., M.F. Alpdemir, and M.J.J.o.M.B. Şeneş, *Comparison of Friedewald, Martin/Hopkins, and*

Sampson formulae with direct LDL measurement in hyperlipidaemic and normolipidaemic adults. 2024.

8. Le Hoang Bich Nga, T.T.T., Le Thu Hang, Do Thi Minh Anh, Nguyen Thi Ngoc Lan *Comparison of the LDL-C'S results by direct quantification and calculation using formulas.* Vietnam Medical Journal, 2023. **529**(8): p. 216-223.
9. Tran Phuoc Thinh, T.H.Q., Dang Thanh Ngan, Le Thi Ngoc Van, Huynh Quang Minh, Tran Do Hung, *Evaluation of correlation and agreement between directly measured and estimated LDL-Cholesterol using selected new formulas.* Can Tho Journal of Medicine and Pharmacy, 2024. **10**(75): p. 79-86.
10. Marniemi, J., et al., *Poor applicability of the Friedewald formula in the assessment of serum LDL cholesterol for clinical purposes.* 1995. **28**(3): p. 285-289.
11. Martin, S.S., et al., *Friedewald-estimated versus directly measured low-density lipoprotein cholesterol and treatment implications.* 2013. **62**(8): p. 732-739.
12. Anwar, M., D.A. Khan, and F.A.J.J.C.P.S.P. Khan, *Comparison of Friedewald formula and modified Friedewald formula with direct homogeneous assay for low density lipoprotein cholesterol estimation.* 2014. **24**(1): p. 8-12.
13. Naser, A., et al., *Comparison of novel Martin/Hopkins and Sampson equations for calculation of low-density lipoprotein cholesterol in diabetic patients.* 2022. **119**: p. 225-233.