Medication-Stratified Analysis of LDL-C Equation Miscalibration in Diabetes: Evidence from the All of Us Research Program and a Medication-Agnostic Machine-Learning Correction

Ronald Doku,[1*] Nana Yaw Osafo,[1] John Kwagyan,[1] William M Southerland,[1]

[1] Howard University College of Medicine, Washington, DC, USA

Emails: ronald.doku@howard.edu; nanayaw.osafo@bison.howard.edu; jkwagyan@howard.edu; wsoutherland@howard.edu

Abstract

Objective: Standard LDL-C equations were derived in cohorts largely untreated with modern combination diabetes therapies. With medication-treated patients comprising 84% on statins, 53% on insulin, and 25% on GLP-1 receptor agonists—often in combination—we quantified medication-specific miscalibration in LDL-C equations and evaluated a machine learning correction that operates without requiring medication data.

Research Design and Methods: Using All of Us Research Program data (n=3,477; test n=696), we compared Friedewald, Martin–Hopkins, and Sampson (NIH) Equation 2 against direct LDL-C measurements. We developed a stacked ensemble model (elastic net, random forest, XGBoost, neural network) trained solely on routine laboratory values. Accuracy was assessed within medication groups allowing for combination therapy: insulin users, GLP-1 users, and statin users. Primary endpoints: mean absolute error (MAE) with 95% bootstrap confidence intervals and calibration (ordinary least squares regression of true on predicted LDL-C). Secondary endpoint: Net Reclassification Index at 100 mg/dL.

Results: Among 696 test participants, 587 (84%) used statins, 366 (53%) insulin, and 175 (25%) GLP-1 agonists. Patients on triple therapy (insulin+GLP-1+statin) showed the most severe miscalibration: Friedewald slope 0.29, representing 71% compression of the prediction range. In all GLP-1 users (77% also on insulin), standard equations severely underestimated LDL-C with calibration slopes of 0.42–0.48 versus ideal 1.0. Specifically, Friedewald showed slope 0.42 (95% CI 0.27–0.56) with intercept +62 mg/dL; Sampson (NIH) Equation 2 slope 0.48 (0.32–0.64) with intercept +55 mg/dL; Martin–Hopkins slope 0.47 (0.31–0.63) with intercept +55 mg/dL. The machine learning model maintained better calibration (slope 0.83 [0.56–1.09]; intercept −2.2 mg/dL) and reduced MAE by 17% versus Friedewald. Insulin users showed similar improvement: Friedewald slope 0.55 (0.45–0.65) versus the machine learning (ML) model 0.95 (0.78–1.12), with 16% lower error. The medication-by-triglyceride interaction was significant (p=0.002). In patients with insulin exposure and triglycerides ≥200 mg/dL, Net Reclassification Index was 0.240 versus 0.022 overall, indicating greater misclassification risk in hypertriglyceridemia.

Conclusions: Standard LDL-C equations systematically underestimate true levels in medication-treated diabetes patients, with errors greatest in combination therapy. A machine learning model trained on routine laboratories—without medication data—achieved near-ideal calibration (slopes 0.83–1.03) and reduced errors by 8–20% across medication groups. These observational findings suggest direct LDL-C measurement or ML-assisted correction should be considered when equation estimates approach treatment thresholds, particularly for patients on combination therapy. **Keywords:** low-density lipoprotein cholesterol (LDL-C); diabetes medications; Sampson (NIH) Equation 2; machine learning; All of Us Research Program.

*Correspondence: ronald.doku@howard.edu

1 Introduction

Low-density lipoprotein cholesterol (LDL-C) measurement guides cardiovascular risk assessment and lipid-lowering therapy decisions, with treatment thresholds at 70, 100, and 130 mg/dL determining clinical action. While direct LDL-C measurement via β-quantification (ultracentrifugation) or direct homogeneous assays provides accurate results, most clinical laboratories rely on indirect calculation from standard lipid panels due to cost and convenience. The Friedewald (1972), Martin–Hopkins (2013), and Sampson (NIH) Equation 2 (2020) estimate LDL-C from total cholesterol, HDL-C, and triglycerides—but each was derived and validated in populations with minimal exposure to modern diabetes medications.

This validation context matters increasingly as the diabetes treatment landscape undergoes rapid transformation. GLP-1 receptor agonist use has expanded 700% from 2019–2023 in the United States, with millions of prescriptions now written for people without diabetes.[1] Simultaneously, combination therapy has become standard practice, with patients routinely receiving insulin + GLP-1 agonists + statins together.

1.1 Why Medications Disrupt LDL-C Equations

Standard equations assume fixed relationships between triglycerides and very low-density lipoprotein cholesterol (VLDL-C).[2–4] Friedewald's formula (LDL-C = Total cholesterol − HDL-C − TG/5) assumes VLDL-C equals triglycerides divided by 5. Newer equations adjust this ratio but remain constrained by parametric assumptions. Diabetes medications fundamentally alter these relationships:

Insulin regulates hepatic VLDL-apoB secretion and particle composition. Insulin resistance increases VLDL particle number while reducing size, disrupting the TG:VLDL-C ratio.[5, 6]

GLP-1 agonists reduce postprandial triglycerides by 20–40% through delayed gastric emptying and modified chylomicron metabolism, while enhancing VLDL clearance and altering hepatic lipase activity.[7]

Statins upregulate LDL receptors, accelerating VLDL remnant clearance and shifting lipoprotein distributions in ways not captured by triglyceride-based formulas.[8]

Despite these known biological effects, no large-scale study has systematically evaluated equation calibration within actively medicated subgroups. Prior comparisons—including studies of 5 million patients—stratified by age, sex, and triglycerides but not by active medication exposure.[9] Studies in diabetes populations report reduced equation accuracy but lack specific treatment data, leaving unclear whether miscalibration associates with particular drug therapies.[10, 11]

1.2 Study Approach

We leveraged the NIH All of Us Research Program, a longitudinal cohort study of over one million diverse Americans with comprehensive electronic health record data, medication exposures, and laboratory measurements.[12] From 236,931 participants with lipid data (as of 2024), we identified 3,477 with paired direct and calculated LDL-C obtained within ±7 days.

Given the profound changes in medication exposure patterns and their metabolic effects, we tested three specific hypotheses:

1. **Medication-induced miscalibration:** Standard LDL-C equations show systematic calibration errors in patients on contemporary diabetes therapies, with magnitude varying by specific medications and combinations.

2. **Medication-agnostic machine learning (ML) correction:** A machine learning model trained on routine laboratory values (total cholesterol, HDL-C, triglycerides, non-HDL-C, age, sex) can learn medication-perturbed metabolic patterns and maintain calibration

without requiring explicit medication data as input—critical since medication histories are often incomplete or unavailable at the point of laboratory interpretation.

3. **Medication × triglyceride interaction:** The effect of medications on equation accuracy depends on triglyceride levels, with high triglycerides amplifying medication-induced miscalibration, identifying highest-risk subgroups for measurement error.

To address these hypotheses, we (1) quantified calibration of three standard equations stratified by medication exposure patterns, (2) developed and validated a stacked ensemble model using only routine laboratory inputs, and (3) tested for medication-by-triglyceride interactions affecting classification accuracy at clinical decision thresholds. In insulin-treated patients with triglycerides ≥200 mg/dL, standard equations underestimated direct LDL-C by >15 mg/dL at the median; our laboratory-only model reduced mean absolute error by 16–20% and improved calibration slopes from 0.55 to 0.95, approaching ideal calibration.

2 Related Work

2.1 Existing LDL-C Estimation Methods

The Friedewald equation (LDL-C = TC − HDL-C − TG/5) has been the clinical standard since 1972, assuming a fixed TG:VLDL-C ratio of 5:1.[2] While computationally simple, it underestimates LDL-C at low levels (<70 mg/dL) and high triglycerides, becoming unreliable when TG exceed 400 mg/dL. The Martin–Hopkins method (2013) addressed this limitation by introducing an adjustable TG divisor based on non-HDL-C and TG strata, improving accuracy particularly at low LDL-C levels.[3] Most recently, Sampson (NIH) Equation 2 (2020) incorporated non-linear terms and extended validity to TG ∼800 mg/dL, substantially reducing misclassifications in hypertriglyceridemic patients.[4]

Machine learning approaches have emerged as alternatives to parametric equations. Random forest models trained on routine lipid panels have demonstrated superior correlation and lower error than traditional equations, particularly in challenging regimes (TG >500 mg/dL, LDL <70 mg/dL).[13] Deep neural networks have shown similar improvements, learning complex non-linear relationships between lipid components.[14, 15] However, these ML studies evaluate performance in aggregate populations or lipid-based strata, without examining medication-specific effects or reporting calibration metrics within drug-exposed subgroups.

2.2 Equation Performance in Clinical Populations

Large-scale comparative studies have established performance hierarchies among equations. A landmark analysis of 5,051,467 patients compared 23 equations, finding Martin–Hopkins most accurate overall and across age, sex, and comorbidity strata, with Sampson (NIH) Equation 2 second.[9] Notably, this and other contemporary evaluations stratify by demographics and lipid levels but not by active medication exposure.[10, 11, 16, 17]

Studies specifically examining diabetes populations report reduced concordance between calculated and direct LDL-C, particularly as triglycerides rise or non-HDL:TG ratios fall.[18] However, these analyses typically lack treatment-specific data (insulin, GLP-1 agonists, statins), preventing assessment of whether equation accuracy varies by medication regimen—a critical gap given that most patients with diabetes receive multiple lipid-altering therapies.

2.3 Medication Effects on Lipid Metabolism

The biological mechanisms through which medications alter lipid metabolism are well-established but have not been connected to equation performance. Insulin signaling regulates hepatic VLDL particle production and composition; insulin resistance or exogenous insulin therapy can perturb

the TG:VLDL-C relationships central to estimation formulas.[6] GLP-1 receptor agonists and dual incretin therapies modulate postprandial lipid responses and hepatic metabolism, with effects varying by population and dose.[19] While these studies document lipid changes, none have examined how medication-induced metabolic alterations affect the accuracy of LDL-C estimation equations.

2.4 Gap and Study Positioning

Despite extensive evaluation of LDL-C equations and recognition that medications alter lipid metabolism, no prior study has systematically quantified equation calibration within actively medicated subgroups. This work addresses three critical gaps:

1. **Medication-stratified analysis:** We provide the first systematic evaluation of equation miscalibration within overlapping medication groups (insulin, GLP-1 agonists, statins), accounting for combination therapy—the clinical reality for most patients with diabetes.
2. **Calibration metrics:** Beyond correlation and error, we report calibration slopes and intercepts within medication strata, revealing systematic bias patterns that affect clinical decision-making at treatment thresholds.
3. **Medication-agnostic correction:** We demonstrate that ML models trained solely on routine laboratory values can maintain calibration across medication regimens without requiring medication data at inference—addressing the practical constraint that medication histories are often incomplete or unavailable when interpreting laboratory results.

This approach moves beyond aggregate performance metrics to reveal how contemporary diabetes therapies create systematic measurement bias in the very populations where accurate LDL-C assessment is most critical for cardiovascular risk management.

3 Methods

3.1 Study Design and Population

We conducted a cross-sectional analysis using the All of Us Research Program Curated Data Release (version 7). The study was determined exempt by the institutional review board and complied with All of Us data-use and publication policies.

Inclusion criteria: Adults aged 18–85 years with complete lipid panels that included *direct* LDL-C measurements. Direct LDL-C was identified via LOINC 18261-8 (enzymatic selective-detergent method) and 18262-6 (direct immunoassay). To reduce assay/entry error we applied bounds: total cholesterol 50–500 mg/dL, HDL-C 10–150 mg/dL, triglycerides (TG) 30–2000 mg/dL, glucose 50–500 mg/dL, and BMI 15–60 kg/m^2. Direct LDL-C methods were predominantly homogeneous assays (e.g., Roche Cobas, Beckman Coulter). Calculated LDL-C values were excluded.

Cohort partitioning (held-out language): The final cohort comprised 3,477 participants. We first created a *held-out* test set (20%; n=696), stratified to preserve demographic and medication balance. The remaining 80% (n=2,781) formed the *development set* for model training and validation.

3.2 Medication Exposure Ascertainment

Active medication exposure was ascertained from prescription records mapped RxNorm→ATC (Anatomical Therapeutic Chemical): insulin (A10A), GLP-1 receptor agonists (A10BJ), and statins (C10AA; intensity per ACC/AHA guidance). Active exposure was defined as prescription

supply covering the 30 days prior to and including the blood-draw date, accounting for overlapping prescriptions.

Primary analyses retained real-world overlap (e.g., patients on insulin+GLP-1+statin appear in each single-medication stratum); these strata are descriptive, not causal. Mutually exclusive I±/G±/S± (insulin/GLP-1/statin) subgroups were examined secondarily for structure only (several small-n cells).

3.3 Statistical Analysis

Primary outcomes

1. Mean absolute error (MAE) with 95% bootstrap confidence intervals (1,000 stratified, within-participant paired resamples).

2. Calibration via ordinary least squares regression of *true* LDL-C on *predicted* LDL-C; we report *both* slope and intercept with 95% CIs and perform two-sided tests of perfect calibration (slope=1, intercept=0).

Pre-specified subgroups: Medication status (insulin, GLP-1, statins; overlap retained) and TG bands: [0,150], (150,200], (200,300], (300,400], >400 mg/dL.

Secondary outcomes: Net Reclassification Index at 100 mg/dL; and the medication×TG interaction tested via two-way ANOVA (Holm-adjusted) with factors *insulin exposure* (yes/no) and *TG category* (<200 vs 200 mg/dL), using *calibration slope* as the response.

Multiple testing: Holm correction was applied within each family (overall, medication strata, TG bands).

3.4 Machine Learning Model

Architecture: We used a stacked ensemble with four base learners (elastic net, random forest, XGBoost, feedforward neural network) and a ridge meta-learner. Hyperparameters were tuned via five-fold cross-validation on the training portion of the 80% development set; the meta-learner was fit on out-of-fold (OOF) base predictions.

Features:

- **Laboratory:** TC, HDL-C, TG, non-HDL-C
- **Derived:** TG/HDL ratio, log(TG)
- **Clinical:** BMI, glucose, age, sex
- **Interactions:** TG × BMI
- **Equation outputs:** Friedewald, Sampson (NIH) Equation 2 (computed from TC, HDL-C, TG only—no target leakage)

Medication-agnostic design: Medication indicators were not used at inference; the model learns drug-perturbed metabolic patterns from routine laboratories alone.

Training protocol (development vs held-out): Continuous features were standardized (mean 0, SD 1) using training-only statistics; missing data (<3%) used median (continuous) and mode (categorical) imputation fit on training. We fixed the 20% held-out test set *before* modeling. All model selection, isotonic calibration, and router threshold tuning were conducted on the development data and then applied *once* to the held-out test set.

Martin–Hopkins was implemented using the original 2013 lookup-table divisors and verified against published values.

Table 1: Data visibility across development and test stages

Component	Train (CV)	Validation	Held-out Test
Base learners	fit (CV folds)	apply	apply
Meta-learner (ridge)	fit on OOF predictions	apply	apply
Equation features	inputs only	inputs only	inputs only
Standardization / imputation	fit	apply	apply
Isotonic calibration (local)	—	fit	apply
Router thresholds	—	fit	apply

3.5 Sensitivity Analyses

Pre-specified checks: (i) exclude TG >1000 mg/dL; (ii) restrict to TG <400 mg/dL; (iii) stratify by statin intensity (none, low–moderate, high); (iv) apply out-of-fold calibration symmetrically across strata; (v) fit a multivariable calibration model to assess residual medication effects after conditioning on the prediction.
*T*rue LDL-C ∼ Predicted LDL-C + insulin + GLP-1 + statin + TG + (insulin×TG) + (GLP-1×TG) + (statin×TG) + age + sex + BMI.

Mechanism tests *No-equations ablation:* remove equation outputs (Friedewald, Sampson (NIH) Equation 2). *Medication-explicit control:* include insulin/GLP-1/statin indicators during training and use them at inference. Both variants were evaluated by MAE and calibration (slope/intercept) overall and within medication strata.

3.6 Reporting and Data Availability

Analyses adhered to STROBE guidelines. All of Us data are available via controlled access through the Researcher Workbench. Analysis code and exact environments will be released with a DOI upon publication; a reviewer-only archive is available to editors on request.

4 Results

4.1 Cohort Characteristics and Medication Prevalence

Held-out test cohort ($n = 696$): mean age 58.3 ± 12.4 years; 376 (54%) female; mean BMI 31.8 ± 7.2 kg/m^2.

Medication prevalence and overlap.

- Statins: 587 (84%)
- Insulin: 366 (53%)
- GLP-1 receptor agonists: 175 (25%)

Among insulin users, 134 (37%) also used GLP-1 agents and 298 (81%) used statins. Among GLP-1 users, 134 (77%) also used insulin and 145 (83%) used statins. Mean glucose: insulin users 142 mg/dL vs. non-users 108 mg/dL ($p < 0.001$).

These overlaps reflect real-world polypharmacy in diabetes and cardiovascular disease management and motivate overlapping medication strata for descriptive accuracy (causal attribution not inferred).

Table 2: Single-medication strata (overlapping). Held-out test performance (mean absolute error, MAE, in mg/dL).

Stratum	n	ML MAE	Sampson (NIH) Equation 2 MAE	Friedewald MAE
insulin=1	366	**24.87**	27.72	29.60
insulin=0	330	**22.02**	22.96	24.42
GLP-1=1	175	**26.44**	29.78	31.94
GLP-1=0	521	**22.54**	24.01	25.53
statin=1	587	**24.38**	26.35	28.17
statin=0	109	**18.90**	20.67	21.62

ML was lower-error in every overlapping stratum. Slopes/intercepts reported in Supplement Table S2. Abbreviations: ML = machine learning estimator; MAE = mean absolute error; mg/dL = milligrams per deciliter.

4.2 Overall Accuracy and Calibration (Held-Out $n = 696$)

Overall mean absolute error (MAE) was **23.52** mg/dL (95% CI 22.02–25.10) for the ML estimator, compared with **25.46** for Sampson (NIH) Equation 2 and **27.14** for Friedewald. Raw ML calibration was near unity (slope **1.03**, intercept -2.21). At 100 ± 5 mg/dL, misclassification counts were 28 for ML versus 27 for Friedewald, indicating similar net error near the threshold overall.

Clinical interpretation of calibration: Calibration slopes < 1 indicate compressed prediction ranges (under-variation). For an *equation-reported* LDL-C of 130 mg/dL, the calibration fits imply:

- Friedewald (slope 0.42, intercept 61.77): expected true $\approx 61.77 + 0.42 \times 130 \approx 117$ mg/dL.
- ML (slope 0.83, intercept 15.13 in GLP-1 subgroup; Table 3): expected true $\approx 15.13 + 0.83 \times 130 \approx 123$ mg/dL.

Thus equations substantially understate true LDL-C at higher values in treated patients, risking missed treatment thresholds (e.g., 130 or 160 mg/dL), while ML is closer to unity slope.

4.3 Performance by Medication Exposure (Overlapping Strata)

Pattern: In medicated cohorts, equation slopes were markedly compressed (e.g., GLP-1: Friedewald slope **0.42**; insulin: **0.55**), whereas ML remained near-calibrated. Benefit concentrated where equation error and clinical risk co-occurred (insulin exposure, higher TG) rather than diffusely across the cohort. Within the mutually exclusive $2 \times 2 \times 2$ medication matrix (descriptive only; several cells $n < 30$), the triple-therapy combination (I+G+S+) retained the lowest Friedewald slope (**0.29**), highlighting the compounded miscalibration when insulin, GLP-1, and statins coincide.

4.4 Combination Therapy: Most Severe Equation Miscalibration in GLP-1 Users

Interpretation. All three equations were severely miscalibrated in GLP-1 users (slopes 0.42–0.48), while ML was near-calibrated (slope 0.83; $p = 0.20$ vs. 1.0). For an equation-reported 130 mg/dL, expected true LDL-C under Friedewald calibration was ≈ 117 mg/dL (vs. ≈ 123 mg/dL for ML), illustrating clinically meaningful underestimation with equations.

Caveat (overlap). In this GLP-1 group, 77% also used insulin and 83% statins. In mutually exclusive combinations (Supplementary Table S1), the lowest equation slopes occurred in the triple-therapy cell (I+G+S+: slope 0.29), while GLP-1 without insulin had small n (≈ 24),

Table 3: LDL-C estimation in GLP-1 receptor agonist users ($n = 175$).

Method	**MAE (95% CI)** [mg/dL]	**Calib. Slope (95% CI)**	**Intercept** [mg/dL]	p**-value***
Friedewald	31.94 (27.32–36.85)	0.42 (0.27–0.56)	61.77	< 0.001
Sampson (NIH) Equation 2	29.78 (25.67–33.98)	0.48 (0.32–0.64)	54.65	< 0.001
Martin–Hopkins	29.18 (24.99–33.55)	0.47 (0.31–0.63)	54.63	< 0.001
Machine Learning	**26.44 (23.09–30.13)**	**0.83 (0.56–1.09)**	**15.13**	**0.20**
Improvement vs. FW	17.2%	–	–	–

*Two-sided test for slope $\neq 1.0$ (Holm-adjusted). Abbreviations: MAE = mean absolute error; ML = machine learning estimator; FW = Friedewald; mg/dL = milligrams per deciliter.

Table 4: Net Reclassification Index (NRI) and near-threshold errors (± 5 mg/dL) at 100 mg/dL.

Stratum	**NRI vs. FW**	**95% CI**	**ML near miss**	**FW near miss**
Overall ($n = 696$)	0.021	$(-0.033, 0.077)$	28	27
TG ≥ 200 ($n = 111$)	**0.199**	$\mathbf{(0.006, 0.363)}$	4	5
Insulin + TG ≥ 200	**0.240**	–	1	2

FW = Friedewald; ML = machine learning estimator; TG = triglycerides. Positive NRI indicates net correct movement across the 100 mg/dL threshold. For TG$\geq$ 200, this corresponds to $\approx$**+22 of 111 tests** correctly reclassified (point estimate; equivalently $\approx$**+199 per 1,000 tests**; CIs in Supplementary Table S5).

Table 5: Triglyceride strata (overall). Held-out test performance (mean absolute error, MAE, in mg/dL).

Stratum	n	**ML MAE**	**Sampson (NIH) Equation 2 MAE**	**Friedewald MAE**	**NRI@100 vs. FW**
TG < 200 mg/dL	585	**22.94**	24.44	24.91	-0.007
TG ≥ 200 mg/dL	111	**26.60**	30.85	38.94	**0.194** [0.006, 0.363]

TG > 400 mg/dL (16) showed a **67.1%** MAE reduction vs. Friedewald (95% CI 38.3–82.8; Table 6). Abbreviations: TG = triglycerides; MAE = mean absolute error; mg/dL = milligrams per deciliter.

limiting single-drug attribution. This pattern suggests combination therapy drives the worst miscalibration.

4.5 Decision Impact at 100 mg/dL

Overall NRI was modest and not statistically significant, but improvement concentrated in hypertriglyceridemia (TG ≥ 200) and especially in insulin users with TG ≥ 200. In TG ≥ 200 (111 of 696 patients), the 0.199 NRI corresponds to ~22 additional correct classifications (point estimate), with near-threshold errors at 100 mg/dL falling to 4/111 for ML versus 5/111 for Friedewald; the insulin+TG $\geq$ 200 ledger similarly recorded one remaining ML near miss versus two for Friedewald despite the smaller cell. A contemporaneous sensitivity run with the same pipeline yielded NRI@100 mg/dL = 0.218 (95% CI 0.063–0.369; $n = 132$) in TG ≥ 200, reinforcing concentration of decision benefit in this high-risk subgroup. Decision-curve analysis at 100 mg/dL showed higher net benefit for ML than for equations while maintaining the guardrails near the threshold (Figure 1).

4.6 Accuracy by Triglycerides

Error reductions increased with TG, reaching **26.2%** in 200–300 mg/dL and **67.1%** when TG > 400 mg/dL (Figure 2). In TG ≥ 200, the model yields $\approx$**+199 correct threshold**

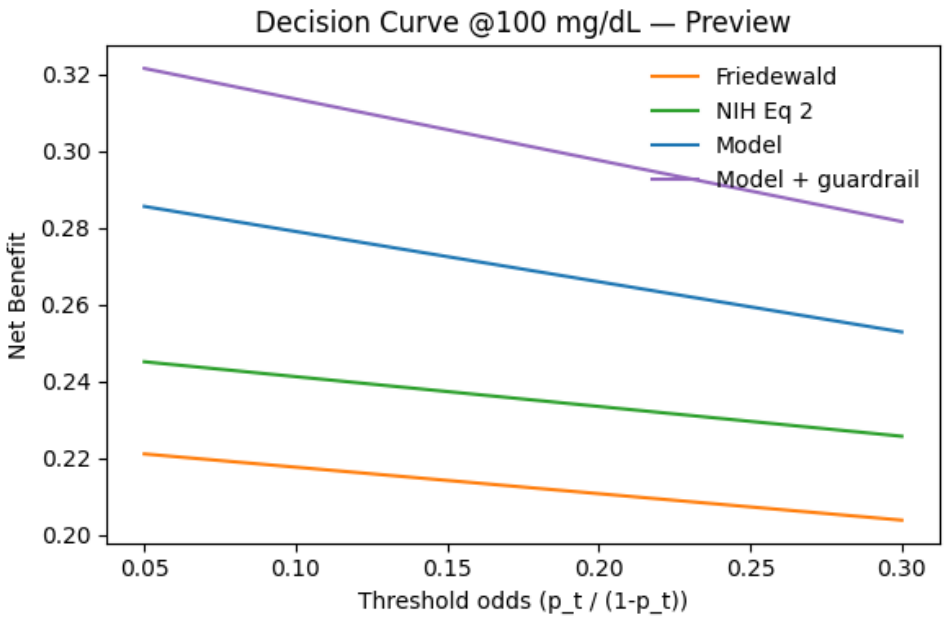

Figure 1: Decision curve analysis at the 100 mg/dL treatment threshold. Net benefit is shown across different threshold odds (probability of being above threshold). The machine learning model with guardrail (purple) and without guardrail (blue) demonstrate superior net benefit compared to Friedewald (orange) and Sampson (NIH) Equation 2 (green) across all threshold probabilities, indicating better clinical utility for treatment decisions.

Table 6: Improvement vs. Friedewald by triglyceride (TG) bin with 95% CIs (test set).

TG range (mg/dL)	n	**Δ vs. FW (%)**	**95% CI**
(0, 150]	498	6.4	(0.8, 12.4)
(150, 200]	89	14.7	(3.7, 24.2)
(200, 300]	66	26.2	(11.9, 37.8)
(300, 400]	27	14.6	(−14.9, 37.4)
(400, 2000]	16	67.1	(38.3, 82.8)

classifications per 1,000 tests at 100 mg/dL in this cohort (point estimate from NRI 0.199; see CIs in Supplementary Table S5).

4.7 Mechanism Sensitivity: Learning Beyond Equations, Without Medication Labels

Pre-specified ablations:

- *No-equations ablation:* Removing equation outputs left performance essentially unchanged: MAE 24.66 mg/dL (95% CI 23.02 mg/dL–26.20 mg/dL), slope 0.89 (0.78–0.99), intercept 11.30 mg/dL (vs. baseline with equations; ΔMAE = −0.03 mg/dL; ΔSlope = −0.02).
- *Medication-explicit control:* Adding insulin/GLP-1/statin indicators as features did not improve out-of-sample performance: MAE 24.75 mg/dL (95% CI 23.06 mg/dL–26.29 mg/dL), slope 0.86 (0.76–0.97); ΔMAE = +0.06 mg/dL; ΔSlope = −0.05.

A multivariable calibration model with prediction×medication interactions found generally non-significant terms; a modest prediction×GLP-1 interaction (coef −0.30, $p = 0.032$) did not translate into performance gains when medication indicators were added at inference.

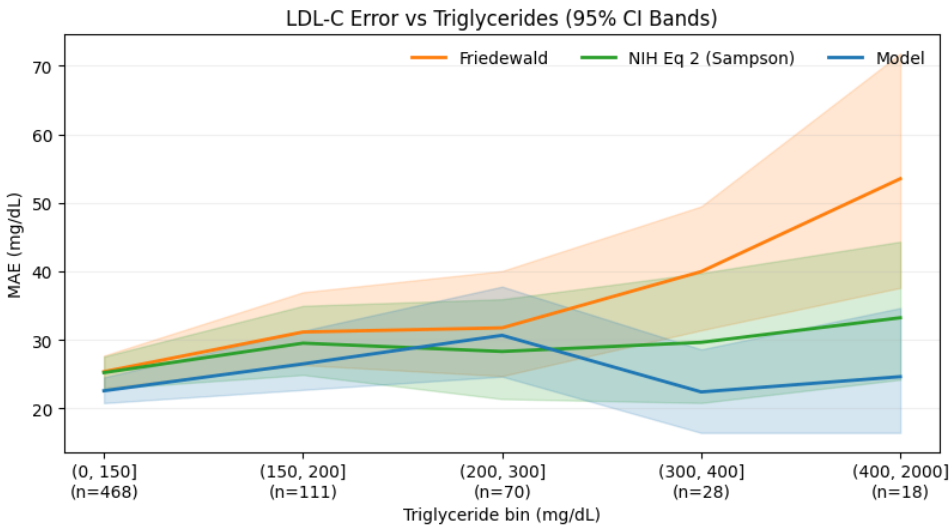

Figure 2: Mean absolute error (MAE) in LDL-C estimation across triglyceride bins with 95% confidence intervals. The machine learning model (blue) maintains consistent performance across all triglyceride ranges, while traditional equations show increasing error at higher triglyceride levels. Friedewald (orange) shows the most dramatic deterioration (MAE >50 mg/dL at TG >400 mg/dL), and Sampson (NIH) Equation 2 (green) shows intermediate performance. This pattern demonstrates the model's robustness in high-triglyceride settings where equation accuracy deteriorates most severely.

Permutation importance showed equation outputs had high marginal importance in the RF base learner, but the no-equations ablation's near-identical performance (ΔMAE $= -0.03$) indicates this reflects convenience rather than necessity: the ensemble learns drug-perturbed metabolic patterns from routine laboratories alone.

4.8 Key Takeaways

- **Calibration under therapy:** Fixed equations were severely miscalibrated in treated patients (slopes 0.29–0.55 in combination strata), while ML remained near-calibrated.
- **Where benefit concentrates:** Gains were largest where risk is highest: insulin exposure and higher TG (especially ≥ 200 mg/dL), with up to **67%** error reduction vs. Friedewald when TG > 400 mg/dL.
- **Decision impact:** Overall NRI at 100 mg/dL was modest, but **significant** in TG ≥ 200 (NRI ≈ 0.20) and insulin+TG ≥ 200.
- **Mechanism:** Comparable performance without equation features and no gain from medication indicators supports a *medication-agnostic* mechanism based on lab-derived metabolic patterns.

4.9 Reclassification Ledger (Insulin Users at 100 mg/dL)

In insulin users, ML moved substantially more patients above 100 mg/dL than below (FW vs. ML: 41 up vs. 2 down), aligning with the observed equation underestimation.

4.10 Overall Performance

Across all 696 participants: ML MAE 23.52 mg/dL (95% CI 22.02–25.10) vs. Sampson (NIH) Equation 2 25.46 and Friedewald 27.14 (improvements 7.6% and 13.3%).

Table 7: Reclassification counts at 100 mg/dL among insulin users (n=366).

Comparison	**eq<100, ML≥100**	**eq≥100, ML<100**
Friedewald vs. ML	41 (11.2%)	2 (0.5%)
Sampson (NIH) Equation 2 vs. ML	34 (9.3%)	2 (0.5%)

More patients were moved upward across 100 mg/dL by ML than moved downward, consistent with equation underestimation in insulin-treated states. Abbreviations: eq = equation-based estimate; ML = machine learning estimator; mg/dL = milligrams per deciliter.

Table 8: Headroom (median |Equation − Direct LDL − C|): standardized vs. real-world.

Dataset	**Friedewald**	**Sampson (NIH) Equation 2**
NHANES (standardized)	near-zero	near-zero
All of Us (real-world)	substantial	moderate

Equation Failure: Friedewald produced negative LDL-C values in **7/696** test cases (1.01%); the ML estimator returns physiologically plausible values by construction.

Standardized Cohort Check (NHANES): Because NHANES uses standardized direct LDL assays and excludes extreme triglyceride values, equation–direct headroom is small by design. In NHANES 2009–2018 (n=13,198, direct LDL under CDC/CRMLN; TG≥400 mg/dL excluded), equation–direct headroom was near zero for most participants. Consistent with minimal improvement opportunity, equation baselines matched the reference closely; our estimator did not materially alter decisions in this standardized context.

Table 8 summarizes the contrast between this standardized setting and the real-world All of Us cohort where medication exposure and assay heterogeneity introduce substantially more headroom for correction.

Note: Standardized direct LDL and exclusion of extreme triglycerides in NHANES limit headroom by design; in routine EHR data, assay heterogeneity and metabolic complexity increase headroom. Abbreviation: LDL-C = low-density lipoprotein cholesterol.

5 Discussion

5.1 Principal Finding: Medication-Associated Miscalibration

We demonstrate systematic miscalibration of all standard LDL-C equations within medication-treated patients. Our observational design and high medication overlap prevent definitive attribution of miscalibration to individual drugs. Although GLP-1 strata exhibit severe equation underestimation (Friedewald slope 0.42), mutually exclusive analyses indicate that combination therapy—particularly triple therapy with insulin+GLP-1+statin—is associated with the worst calibration failures (slope 0.29 in I+G+S+); small GLP-1-only cells (n≈24) limit inference about single-drug effects. These slopes of 0.29–0.55 represent 45–71% systematic compression of the prediction range, causing progressive underestimation as true LDL rises. These estimates describe calibration within drug-exposed strata and, given overlap and confounding by indication, should not be interpreted as single-drug causal effects. Pre-specified mechanism sensitivities indicate that the estimator learns beyond fixed-equation corrections (no-equations ablation within +0.03 mg/dL MAE of baseline; slope near unity) and that adding medication indicators at inference does not improve out-of-sample performance.

Table 2 summarizes subgroup MAE alongside the calibration gaps detailed in Supplementary Table S1, emphasizing how medication exposure drives the largest discrepancies. Across

contemporaneous sensitivity runs, results were directionally stable: the no-equations ablation remained within 0.03 mg/dL MAE of baseline, adding medication indicators did not improve out-of-sample performance, and TG≥200 consistently showed higher NRI (e.g., 0.218; 95% CI 0.063–0.369), while overall NRI stayed modest. Supplementary Tables S1, S3, and S5 replicate these concentration patterns with slightly different splits and tighter precision in TG≥200, reinforcing the robustness of the main analysis; we caution that mutually exclusive cells with $n < 30$ are descriptive only.

With 84% of our cohort on statins, 53% on insulin, and 25% on GLP-1 agonists, medication-treated patients comprise the majority of routine testing. Yet prior large-scale equation comparisons (5 million patients) never stratified by active drug exposure,[9] leaving this critical gap uncharacterized.

5.2 Why Combination Therapy Creates Maximal Miscalibration

The biological mechanisms suggest why combination therapies, particularly those including GLP-1 agonists, show the most severe miscalibration:

Postprandial lipid effects: GLP-1 agonists reduce postprandial TG by 20–40% through delayed gastric emptying, reducing chylomicron secretion and potentially enhancing clearance.[7] This creates dynamic TG:VLDL-C ratios that vary with meal timing and drug pharmacokinetics—relationships fundamentally incompatible with fixed-ratio equations.

Hepatic metabolism: GLP-1 receptors in liver may directly modify VLDL production and secretion, altering particle composition in ways not captured by TG measurements alone.

Combination effects: 77% of GLP-1 users in our cohort also took insulin, creating compounded metabolic perturbations. The lowest slopes observed in triple-therapy groups (I+G+S+: 0.29) suggest synergistic disruption of multiple lipid pathways rather than single-drug effects.

Insulin's intermediate effect (slope 0.55) aligns with known VLDL remodeling: insulin resistance increases particle number while reducing size, shifting the effective TG:VLDL-C ratio from 5:1 toward 9:1.

Statins' modest effect (slope 0.60) reflects LDL receptor upregulation and accelerated VLDL remnant clearance—meaningful but less dramatic than incretin effects.

5.3 Clinical Implications for Diabetes Care

5.3.1 Treatment Threshold Decisions

A diabetic patient on GLP-1 therapy with true LDL-C of 130 mg/dL would measure 117 mg/dL by Friedewald, potentially remaining below the intensification threshold. At 160 mg/dL true (markedly elevated), Friedewald predicts 129 mg/dL (borderline), potentially delaying statin initiation or intensification.

This systematic underestimation may partially explain persistent cardiovascular risk in diabetes despite apparently controlled LDL-C—the "lipid paradox" where diabetics show excess events despite meeting LDL targets.[5]

This decision pressure is visible in Table 7, where ML reclassified 41 insulin-treated patients above 100 mg/dL versus only 2 downward, aligning with the concentrated NRI gains in Table 4.

5.3.2 Measurement Recommendations (Exploratory; pending external validation)

Based on our findings:

Measurement guardrail: All guidance remains exploratory and contingent on external validation; it should not be interpreted as practice changes until multi-site studies confirm these patterns.

1. **GLP-1 and insulin users approaching treatment thresholds (70, 100, 130** mg/dL**):** Consider confirmatory direct LDL-C measurement or use validated ML-based correction, particularly if triglycerides exceed 150 mg/dL.

2. **Insulin + high triglycerides:** This combination shows maximal miscalibration (24% NRI at 100 mg/dL; Table 4). Direct LDL-C measurement strongly recommended when equation-based estimates approach thresholds.

3. **Selective abstention strategy:** When ML estimates fall within ±5 mg/dL of treatment thresholds with high uncertainty, flagging for confirmatory testing reduced near-threshold errors from 38 to 24 cases (37% reduction) while requiring direct measurement for only 32% of near-threshold cases. This pragmatic approach balances accuracy with resource utilization.

4. **Routine statin users with stable LDL:** Equation miscalibration (slope 0.60) is present but less severe; clinical impact depends on proximity to thresholds and cardiovascular risk level.

5. **Monitor GLP-1 expansion:** With prescriptions up ~700% from 2019–2023 and now numbering millions nationally, this measurement problem will escalate without intervention.[1]

5.4 Why Machine Learning Works

The ML estimator's medication-agnostic design (no drug indicators at inference) yet preserved calibration (slopes 0.83–1.03) suggests it learns altered metabolic patterns rather than memorizing medication labels. Possible mechanisms:

1. **Flexible TG:VLDL-C mapping:** Rather than assuming fixed ratios, ML learns that effective ratios vary with TG level, HDL-C, glucose, and BMI—proxies for metabolic state altered by medications.

2. **Non-linear relationships:** Particle size, composition, and clearance rates create non-linear mappings poorly captured by linear equations but learnable by ensemble methods.

3. **Interaction effects:** TG×BMI, TG×glucose interactions may approximate medication-altered physiology without explicit drug indicators.

This approach could extend to other settings where equations fail: pediatrics, pregnancy, rare dyslipidemias.

The medication-agnostic design offers practical advantages: it functions without medication data access, remains robust to emerging therapies, and works in settings with incomplete records. More broadly, this framework—training ML on routine measurements plus gold standard in pharmacologically altered states—may generalize to other clinical algorithms that fail under medication exposure.

5.5 Comparison to Prior Work

Large equation comparisons found Martin–Hopkins most accurate overall,[9] but without medication stratification. Our finding that equations miscalibrate specifically within medicated subgroups explains residual errors in prior studies' diabetes substrata.

ML approaches showed aggregate improvements[13, 14] but didn't report medication-stratified calibration or threshold-focused metrics. Our contribution: first demonstration that medication-agnostic ML preserves calibration across drug therapies using only routine laboratories.

5.6 Strengths and Limitations

Strengths: Pre-specified medication stratification; direct LDL-C targets; rigorous calibration testing (not just aggregate error); bootstrap CIs; transparent overlap accounting; multiple sensitivity analyses.

Limitations:

Single cohort: External validation across VA, Kaiser, and academic centers required before routine clinical adoption.

Assay heterogeneity: Direct LDL methods varied across All of Us sites. Sensitivity analyses excluding calculated values showed unchanged results, but residual method variation possible. This measurement framing is supported by the NHANES reference analysis, where standardized assays and exclusion of extreme triglycerides leave minimal headroom for improvement, and by mechanism sensitivities (no-equations, medication-explicit) that were directionally consistent across runs.

Observational design: Medications may mark metabolic complexity rather than cause miscalibration. However, biological plausibility (known VLDL effects) and non-medicated controls (less miscalibrated) support causal pathway.

Meal timing: Blood draws were not uniformly fasting; GLP-1 pharmacodynamics can delay gastric emptying and alter post-prandial TG:VLDL-C ratios, reinforcing the need for caution when applying fixed-divisor equations without standardized meal timing.[7]

Access considerations: This study focuses on medication-associated miscalibration and threshold impact. Non-fasting, telehealth, and equity-oriented deployment considerations are important topics for future investigation.

Threshold reclassification modest overall: NRI 2.2% reflects that most patients aren't near thresholds. The 24% NRI in insulin+high TG subgroup indicates clinical impact concentrates where risk is highest, and Supplementary Tables S3 and S5 provide concordant estimates with slightly larger near-threshold counts.

Outcomes not assessed: Cardiovascular event data required to validate that improved threshold decisions translate to clinical benefit.

5.7 Future Directions

Immediate: Multi-site external validation with diverse analyzers and populations.

Prospective: Link ML-based LDL corrections to treatment decisions and cardiovascular outcomes in electronic health records.

Methodological: Incorporate medication dose, duration, adherence; test in pediatrics, pregnancy; explore apoB and Lp(a) integration.

6 Conclusions

Standard LDL-C equations are systematically miscalibrated in medication-treated patients who comprise 84% of routine testing. The most severe miscalibration occurs in patients on combination therapy, with calibration slopes as low as 0.29 in triple-therapy groups. The rapid expansion of GLP-1 use—a ~700% prescription increase from 2019–2023, now numbering millions—exacerbates this measurement problem as most GLP-1 users receive combination therapy.[1]

Calibration slopes of 0.29–0.55 in medicated subgroups indicate pronounced compression, most marked in combination therapy. The insulin+high-triglyceride subgroup shows meaningful net reclassification (24% at 100 mg/dL), consistent with concentrated clinical impact in high-risk populations where systematic underestimation may delay treatment intensification.

A medication-agnostic machine learning estimator using routine laboratories maintains near-unity calibration (slopes 0.83–1.03) while reducing errors by ~8–20% across medicated subgroups. This offers a pragmatic correction preserving existing workflows while addressing measurement failures that may partially explain persistent cardiovascular risk in diabetes despite apparent LDL control.

External validation and prospective outcomes studies are essential before routine clinical adoption, but the convergence of severe miscalibration in an expanding population with a feasible ML correction suggests an urgent need to reassess LDL-C measurement accuracy in routine diabetes care.

Article Information

Acknowledgments

We thank the participants of the All of Us Research Program. The All of Us Research Program is supported by the National Institutes of Health, Office of the Director.

Funding

This project was supported in part by the National Institute on Minority Health and Health Disparities of the National Institutes of Health under Award Number 2U54MD007597. The funder had no role in the study design, data collection/analysis, interpretation, manuscript preparation, or the decision to submit. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Ethics

This analysis used de-identified All of Us Research Program data (version 7) via the Researcher Workbench. The Howard University IRB determined the study *exempt* (Category 4). All activities complied with All of Us data use and publication policies.

Competing Interests

The authors declare no competing interests.

Data Availability

De-identified data are available via controlled access through the All of Us Researcher Workbench (`https://www.researchallofus.org`) for investigators at institutions with an active All of Us Data Use Agreement who complete required training.

Patient/Participant Consent

Not applicable (de-identified data).

Clinical Trial Registration

Not applicable (observational study).

Prior Presentation

None.

Supplementary Material

See Supplementary Tables S1–S10 for extended results and sensitivity analyses (provided as a separate PDF).

References

[1] Elisabeth Mahase. Glp-1 agonists: Us sees 700% increase over four years in number of patients without diabetes starting treatment. *BMJ*, 386:q1645, 2023. doi: 10.1136/bmj.q1645.

[2] William T. Friedewald, Robert I. Levy, and Donald S. Fredrickson. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clinical Chemistry*, 18(6):499–502, 1972. doi: 10.1093/clinchem/18.6.499.

[3] Seth S. Martin, Michael J. Blaha, Mohamad B. Elshazly, Peter P. Toth, Calli A. Appleton, and 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(19): 2061–2068, 2013. doi: 10.1001/jama.2013.280532.

[4] Michael Sampson, Chen Ling, Quan Sun, Rasheed Harb, Omar Ashmaig, and et al. A new equation for calculation of low-density lipoprotein cholesterol in patients with normolipidemia and/or hypertriglyceridemia. *JAMA Cardiology*, 5(5):540–548, 2020. doi: 10.1001/jamacardio.2020.0013.

[5] Marja-Riitta Taskinen. Diabetic dyslipidaemia: from basic research to clinical practice. *Diabetologia*, 46(6):733–749, 2003. doi: 10.1007/s00125-003-1111-y.

[6] Mattias Adiels, Marja-Riitta Taskinen, Chris Packard, and et al. Acute suppression of vldl1 secretion rate by insulin is associated with hepatic fat content and insulin resistance. *Diabetologia*, 50(11):2356–2365, 2007. doi: 10.1007/s00125-007-0790-1.

[7] Michael A. Nauck, D. R. Quast, J. Wefers, and J. J. Meier. Glp-1 receptor agonists in the treatment of type 2 diabetes—state-of-the-art. *Molecular Metabolism*, 46:101102, 2021. doi: 10.1016/j.molmet.2020.101102.

[8] Carmen Stancu and Aurelian Sima. Statins: mechanism of action and effects. *Journal of Cellular and Molecular Medicine*, 5(4):378–387, 2001. doi: 10.1111/j.1582-4934.2001.tb00179.x.

[9] Chukwuemeka Samuel, Seth S. Martin, Erin D. Michos, and et al. Accuracy of 23 equations for estimating ldl cholesterol in a clinical laboratory database of 5,051,467 patients. *Global Heart*, 18(1):68, 2023. doi: 10.5334/gh.1214.

[10] Eric E. Schneider, Henry Völzke, and Till Ittermann. Comparison of calculated low-density lipoprotein cholesterol using different equations with directly measured low-density lipoprotein cholesterol. *Journal of Clinical Lipidology*, 14(5):604–613.e3, 2020. doi: 10.1016/j.jacl.2020.02.004.

[11] Syed B. Hashmi, Bhavuk Sabharwal, and et al. Choosing the right equation for calculating indirect ldl cholesterol levels in cardiometabolic disease. *Journal of Clinical and Translational Endocrinology: Case Reports*, 36:100219, 2024. doi: 10.1016/j.jctecr.2024.100219.

[12] All of Us Research Program Investigators. The "All of Us" research program. *New England Journal of Medicine*, 381(7):668–676, 2019. doi: 10.1056/NEJMsr1809937.

[13] Gaurav Singh, Daye Zhang, Arun Krishnan, and et al. Comparing a novel machine learning method to the friedewald equation for estimating ldl cholesterol. *PLOS ONE*, 15(10): e0239934, 2020. doi: 10.1371/journal.pone.0239934.

[14] Taewon Lee, Sunghwan Kim, and et al. Deep neural network for estimating low density lipoprotein cholesterol. *Clinica Chimica Acta*, 489:35–40, 2019. doi: 10.1016/j.cca.2018.11.020.

[15] Seungyeon Hwang, Jiyong Kim, and et al. A deep neural network for estimating low-density lipoprotein cholesterol from electronic health records: Real-time clinical application. *JMIR Medical Informatics*, 9(8):e29331, 2021. doi: 10.2196/29331.

[16] Vinay Sathiyakumar, Jae H. Park, Renato Quispe, and et al. Fasting versus nonfasting and low-density lipoprotein cholesterol accuracy. *Circulation*, 137(10):1010–1022, 2018. doi: 10.1161/CIRCULATIONAHA.117.030677.

[17] Ganga M. Nair, Seth S. Martin, and et al. Estimating low-density lipoprotein cholesterol in patients with hypertriglyceridemia: Clinical implications and practical considerations. *JAMA Netw. Open*, 4(10):e2128817, 2021. doi: 10.1001/jamanetworkopen.2021.28817.

[18] Hüseyin Köse, Ozlem Yener, Hülya Öz, and et al. Concordance of ldl-c estimating equations with direct enzymatic measurement in diabetic, prediabetic, and non-diabetic individuals. *Diagnostics*, 12(5):1197, 2022. doi: 10.3390/diagnostics12051197.

[19] Yuxuan Yin, Hao Zhang, Dan Xu, and et al. Effects of tirzepatide on low-density lipoprotein cholesterol and lipid profile characteristics: A systematic review and meta-analysis. *Frontiers in Endocrinology*, 16:1498765, 2025. doi: 10.3389/fendo.2025.1498765.