Accurate, Race-Free LDL-C Estimation in Non-Fasting Settings: A Machine-Learning Study in 3,477 Adults

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

Importance: Traditional LDL-C testing barriers—mandatory 9–12 hour fasting and in-person visits—disproportionately limit access for rural populations (60% of US counties lack cardiology services), shift workers, patients with diabetes, and caregivers. These barriers persist despite 2016 European guidelines endorsing non-fasting lipid panels and explosive post-COVID telehealth growth.

Objective: To determine whether machine learning can maintain and enhance LDL-C accuracy in real-world access-expanding scenarios: non-fasting collection, remote assessment without vital signs, and race-free inputs with equity audit.

Design, Setting, and Participants: Cross-sectional analysis of All of Us Research Program (n=3,477 adults; 40.1% tested outside traditional fasting windows). We evaluated accuracy across fasting proxies, labs-only configurations, and racial/ethnic groups using 80/20 train/test split with bootstrap confidence intervals.

Main Outcomes and Measures: Mean absolute error (MAE) and calibration in likely non-fasting states (blood drawn after 10:00 local time, triglycerides (TG) $\geq 200\ \mathrm{mg\,dL^{-1}}$, glucose $> 110\ \mathrm{mg\,dL^{-1}}$); labs-only configuration (no blood pressure); racial equity (Black-/African American vs. White MAE gap); economic impact of single-visit workflows.

Results: Among 696 test participants, 279 (40.1%) were tested in likely non-fasting states. Standard equations degraded substantially in non-fasting conditions (Friedewald MAE 29.05 $\mathrm{mg\,dL^{-1}}$vs. 25.87 $\mathrm{mg\,dL^{-1}}$fasting; calibration slope remained far from ideal at 0.58–0.61), while the ML system maintained accuracy (MAE 24.04 $\mathrm{mg\,dL^{-1}}$non-fasting vs. 23.18 $\mathrm{mg\,dL^{-1}}$fasting; slope 0.99–1.07)—yielding 17.2% improvement over Friedewald in non-fasting states vs. 10.4% when fasting.

Removing blood pressure measurements changed MAE by only -0.12 $\mathrm{mg\,dL^{-1}}$(90% CI -0.33 to 0.08), meeting non-inferiority within $\pm 0.5\ \mathrm{mg\,dL^{-1}}$margin and enabling telehealth workflows. The system achieved racial equity without race input (Black/African American–White gap -0.19 $\mathrm{mg\,dL^{-1}}$, 95% CI -4.12 to 3.74, indicating no meaningful difference).

Economic modeling showed eliminating fasting requirements prevents 4,000 repeat visits annually per 10,000 tests. Base-case net savings of $185,000 per 10,000 tests (40% repeat-visit rate at $50/visit minus $15,000 implementation/QA costs). Break-even occurs at 750 tests in year one, then 250 tests/year thereafter.

Conclusions and Relevance: Machine learning enables accurate LDL-C assessment without fasting or in-person visits, addressing critical access barriers for rural, diabetic, and underserved populations. With roughly 40% of real-world testing already occurring outside fasting windows, single-visit workflows can substantially reduce repeat visits while maintaining measurement quality and achieving racial equity.

*Correspondence: ronald.doku@howard.edu

Keywords: low-density lipoprotein cholesterol; machine learning; non-fasting lipid panel; telehealth; health equity; cardiovascular risk

1 Introduction

1.1 The Access Crisis in Lipid Testing

Cardiovascular disease remains the leading cause of death in the United States, with lipid management central to prevention.[1] Yet access to accurate LDL-C testing is profoundly inequitable. Traditional testing requires 9–12 hour overnight fasting and in-person blood draws, creating cascading barriers:

Geographic barriers: 60% of US counties lack cardiology services; rural patients travel median 40 miles for specialty lipid clinics.[2] Fasting requirements necessitate early morning appointments, often requiring overnight stays or 3–4 hour pre-dawn drives.

Socioeconomic barriers: Shift workers cannot easily accommodate fasting windows or morning appointments. Caregivers (predominantly women) struggle with childcare during testing. Hourly workers lose wages for clinic visits.[3]

Medical barriers: Patients with diabetes face hypoglycemia risk during prolonged fasting. Those on insulin cannot safely skip meals. Multiple comorbidities require more frequent monitoring, multiplying access challenges.[4]

The scale of the problem: In our cohort, 40.1% were tested outside traditional fasting windows—representing either protocol violations or practical accommodations. Nationally, this suggests 20 million annual non-fasting lipid panels in suboptimal conditions.

1.2 The Telehealth Opportunity

The COVID-19 pandemic catalyzed explosive telehealth growth: virtual visits increased 154-fold (0.1% to 15.4% of encounters).[5] Home phlebotomy services expanded. Retail pharmacies added point-of-care testing. Yet lipid management hasn't kept pace because:

- Standard equations assume fasting and often lose accuracy in non-fasting states [4].
- Blood pressure is not routinely available in retail or home testing settings.
- Race-based corrections in clinical algorithms raise equity concerns [6].

The 2016 European Atherosclerosis Society endorsed non-fasting lipids,[4] but US adoption lags due to accuracy concerns with calculated LDL-C in non-fasting states.

1.3 The Technical Challenge

Standard LDL-C equations (Friedewald, Martin-Hopkins, NIH Equation 2) were validated in fasting populations and degrade in non-fasting states—particularly at elevated triglycerides where postprandial lipemia compounds errors.[7] This creates a cruel paradox: the populations needing convenient testing most (diabetics, rural patients, shift workers) are precisely those where equation failures are most severe.

Recent machine learning approaches show promise[8] but haven't demonstrated: (1) maintained accuracy across fasting states, (2) feasibility with labs-only data (no vitals), or (3) racial equity without race-based adjustments.

1.4 Study Objectives

We hypothesized that machine learning could maintain LDL-C accuracy in three access-enabling scenarios:

- **Non-fasting states:** Patients can be tested at any time, eliminating scheduling barriers.
- **Labs-only configuration:** No blood pressure required, enabling telehealth and retail models.
- **Race-free inputs with equity audit:** Equitable performance without race-based corrections.

We evaluated real-world performance, economic impact, and implementation pathways to inform access-expanding deployment.

2 Related Work

2.1 Non-fasting Lipid Testing and Guideline Evolution

For decades, routine lipid testing required 9–12 hours of fasting, in part because the Friedewald equation assumes fasting triglycerides (TG). Accumulating evidence shows fasting is unnecessary for most patients. In a 2016 European consensus statement, Nordestgaard *et al.* reviewed large cohorts and recommended *non-fasting* lipid profiles as the default, noting only small average postprandial shifts (e.g., modest TG increases) and advocating fasting primarily for special circumstances (e.g., TG $>400\,\mathrm{mg\,dL^{-1}}$, suspected familial hypertriglyceridemia, pancreatitis risk) [4]. Subsequent work has reinforced that population-level lipid distributions and cardiovascular risk prediction are largely comparable between fasting and non-fasting panels [3, 9]. Despite these data, U.S. uptake has been uneven, with lingering concerns about calculated LDL-C accuracy in non-fasting states—a gap directly addressed by modern formulae and ML-based estimators (below).

2.2 Formula-based LDL-C Estimation: Beyond Friedewald

The Friedewald equation [10] has been the mainstay of LDL-C estimation, but it underperforms at low LDL-C and higher TG due to its fixed TG:VLDL-C ratio. Martin and colleagues derived a lookup-table method (the Martin–Hopkins equation) that adapts the TG:VLDL-C ratio across TG and non-HDL strata, improving risk-category concordance versus Friedewald, especially for LDL-C $< 70\,\mathrm{mg\,dL^{-1}}$and in non-fasting samples [11]. More recently, Sampson *et al.* introduced "NIH Equation 2", a bivariate quadratic model calibrated to β-quantification that extends usable accuracy up to TG $\sim 800\,\mathrm{mg\,dL^{-1}}$and reduces misclassification at guideline thresholds [7]. Independent evaluations generally confirm that both Martin–Hopkins and NIH Equation 2 outperform Friedewald at low LDL-C and elevated TG—precisely where clinical decisions hinge on small errors [12]. Still, even refined formulas remain fixed functional forms that can misestimate LDL-C in atypical metabolic states (e.g., medication-perturbed lipemia) or in non-fasting conditions.

2.3 Machine Learning for LDL-C and Lipid Analytics

Leveraging large laboratory datasets, several groups have shown that machine learning (ML) can surpass traditional formulas by learning flexible relations among routine variables. In two large Korean cohorts, a deep neural network (DNN) for LDL-C prediction achieved lower error across risk strata than Friedewald, Martin–Hopkins, and NIH equations, including at low LDL-C and high TG [13]. Other evaluations similarly report that tree-based or ensemble models reduce bias and RMSE versus equation baselines, including in challenging subgroups (non-fasting, LDL-C

$< 70\,\mathrm{mg\,dL^{-1}}$, TG $> 400\,\mathrm{mg\,dL^{-1}}$) [12]. Collectively, this literature demonstrates that ML can (i) accommodate postprandial variability without explicit fasting indicators and (ii) maintain accuracy using only routine lab inputs—features essential for access-expanding, telehealth-ready workflows. Our study extends this line of work by (1) explicitly stress-testing performance in non-fasting states, (2) validating a *labs-only* configuration (no vital signs) for remote settings, and (3) deploying race-free inputs with equity audit.

2.4 Telehealth, Remote Testing, and Access Disparities

Telehealth scaled dramatically during COVID-19 and remains integral to outpatient care [5]. Remote blood-draws, home phlebotomy, and retail-lab collection now make same-day lipid testing feasible without clinic visits. Access gains are most salient for rural communities and underserved populations who face long travel distances and appointment frictions; enabling accurate *non-fasting* LDL-C in remote workflows removes repeat-visit burdens from fasting protocols. Prior work highlights how geographical maldistribution of cardiology services and socioeconomic barriers depress testing completion [5]. By demonstrating non-fasting, labs-only accuracy, our approach operationalizes lipid management for "anytime, anywhere" testing.

2.5 Equity and Race-Free Modeling

A growing body of scholarship cautions against race-based adjustments in clinical algorithms, which can entrench disparities [6]. Although standard LDL-C equations do not explicitly include race, subgroup accuracy gaps can arise from data or model misspecification. Recent lipid-ML reviews stress validating models across populations and reporting subgroup performance [12]. In this context, we design and evaluate an estimator with *race-free inputs with equity audit*, showing no measurable performance gap across racial groups while improving overall accuracy. Our results align with broader aims to deliver equitable preventive cardiology without race corrections, via more faithful modeling of underlying biology and real-world testing conditions.

3 Methods

3.1 Study Population

We analyzed 3,477 adults from All of Us Research Program (v7) with complete lipid panels, glucose, and BMI. Direct LDL-C (LOINC 18261-8, 18262-6) served as reference standard. The cohort was split 80/20 stratified by race/ethnicity (test n=696: 54% female, 23% Black/African American, 8% Hispanic, mean age 58.3 years, mean BMI 31.8 kg/m^2). The present study confines endpoints to fasting status, labs-only workflows, equity, and economic deployment.

3.2 Fasting Status Determination

Because fasting labels were limited, we used validated proxies documented in prior non-fasting lipid studies:[4]

- **Collection time proxy:** Blood drawn after 10:00 local time (outside typical morning fasting window; indicates likely non-fasting)
- **Metabolic proxy:** Triglycerides 200 $\mathrm{mg\,dL^{-1}}$(postprandial elevation)
- **Glycemic proxy:** Random glucose >110 $\mathrm{mg\,dL^{-1}}$(non-fasting)
- **Combined non-fasting index:** Any of the above three criteria

Sensitivity analyses repeated all comparisons with 08:00 and 12:00 time-of-day cutoffs. Because the combined non-fasting index aggregates metabolic and glycemic signals in addition to time-of-day, its prevalence (78.6%) exceeds the time-of-day proxy alone (40.1%).

3.3 Model Development

We trained an ensemble (random forest, XGBoost, neural network, elastic net) with ridge meta-learner using routine laboratory values:

Base configuration: TC, HDL-C, TG, glucose, BMI, age, sex + engineered features (non-HDL, TG/HDL, log(TG))

Labs-only configuration: Excluded blood pressure to assess telehealth feasibility

Critical design: Race/ethnicity excluded from training; used only for post-hoc equity auditing

Missing data (<3%) handled via median/mode imputation. Five-fold cross-validation selected hyperparameters. Equations compared: Friedewald, NIH Equation 2 (Sampson), Martin-Hopkins.

3.4 Primary Outcomes

1. **Non-fasting accuracy:** MAE with 95% bootstrap CIs stratified by fasting proxies
2. **Non-fasting calibration:** OLS regression slope/intercept (true predicted)
3. **Labs-only feasibility:** Two one-sided tests (TOST) for non-inferiority (± 0.5 mg dL^{-1}margin)
4. **Racial equity:** Black/African American vs. White MAE gap with 95% CI

3.5 Economic Analysis

We modeled costs on a per–10,000–test basis, considering:

- **Eliminated repeat visits for fasting:** assumed 40% of tests require a repeat visit at \$50/visit $\Rightarrow$ savings $= 0.40 \times 10{,}000 \times \$50$.
- **Implementation costs:** one-time integration \$10,000; annual QA/monitoring \$5,000.
- **Prevented misclassification events:** downstream cost offsets from improved threshold decisions.
- **Break-even volume:** solve for test volume where total savings minus implementation/QA equals zero (year 1 includes \$10,000 one-time integration cost).

A simple base-case expression is:

$$\text{Net Savings} \;=\; (0.40 \times 10{,}000 \times \$50) \;-\; (\$10{,}000 + \$5{,}000) \;+\; \text{downstream offsets.}$$

3.6 Statistical Analysis

We performed bootstrap resampling with 1,000 iterations, using patient-level pairing across methods to preserve within-subject comparisons. Unless otherwise specified, confidence intervals (CIs) are reported at the 95% level. For multiple comparisons, p-values were adjusted using the Holm procedure. Non-inferiority for the labs-only workflow was evaluated with two one-sided tests (TOST) using 90% confidence intervals; all other intervals are 95%. Calibration quality was summarized using the Integrated Calibration Index (ICI) for regression, computed with 10 equal-frequency bins on predicted LDL-C, weighting by bin prevalence and reporting the mean absolute calibration error (observed minus predicted) aggregated to a single scalar (mg dL^{-1}).

4 Results

4.1 The Real-World Non-Fasting Population

Among the 696 test participants, 279 (40.1 %) were tested outside traditional fasting windows. By non-fasting proxy:

- **Collection time after 10:00:** 279 (40.1 %)
- **Triglycerides (TG)** $\geq 200\,\mathrm{mg\,dL^{-1}}$**:** 111 (15.9 %)
- **Glucose** $> 110\,\mathrm{mg\,dL^{-1}}$**:** 428 (61.5 %)
- **Any non-fasting criterion (combined index):** 547 (78.6 %)

This sizable non-fasting subset reflects real-world testing patterns rather than protocol violations.

4.2 Equations Degrade in Non-Fasting States; ML Does Not

Table 1 shows the critical finding: standard equations lose substantial accuracy when patients aren't fasting, while ML maintains performance.

Table 1: Performance in Fasting vs. Non-Fasting States

State	N (%)	Method	MAE (95% CI)	Slope	Improv.
Fasting	417 (60)	Friedewald	25.87 (23.52–28.15)	0.58	–
		NIH Equation 2	24.31 (22.19–26.51)	0.63	–
		ML	**23.18 (21.26–25.06)**	**0.99**	**10.4%**
Non-Fasting	279 (40)	Friedewald	29.05 (26.31–31.82)	0.61	–
		NIH Equation 2	27.18 (24.67–29.75)	0.65	–
		ML	**24.04 (21.43–26.84)**	**1.07**	**17.2%**

Equation degradation: +3.18 $\mathrm{mg\,dL^{-1}}$ *Friedewald, +2.87* $\mathrm{mg\,dL^{-1}}$ *NIH Equation 2*
ML stability: +0.86 $\mathrm{mg\,dL^{-1}}$ *(minimal change)*

All MAE values in $\mathrm{mg\,dL^{-1}}$. Improvement = % reduction vs. Friedewald.

Key finding: ML improvement over equations *increases* in non-fasting states (17.2% vs. 10.4%), exactly when access-expanding workflows are deployed. Across the full test set, ICI was 7.6 $\mathrm{mg\,dL^{-1}}$for Friedewald versus 2.3 $\mathrm{mg\,dL^{-1}}$for ML; in non-fasting strata ICI was 8.1 $\mathrm{mg\,dL^{-1}}$versus 2.5 $\mathrm{mg\,dL^{-1}}$, respectively, confirming superior calibration quality.

4.3 All Fasting Proxies Show Consistent Pattern

Table 2 demonstrates robustness across multiple non-fasting indicators.

Table 2: ML Performance Across Multiple Non-Fasting Proxies (time, triglycerides, glucose)

Proxy	N	ML MAE ($\mathrm{mg\,dL^{-1}}$)	FW MAE ($\mathrm{mg\,dL^{-1}}$)	Improvement
Time after 10:00	279	24.04	29.05	17.2%
TG 200 $\mathrm{mg\,dL^{-1}}$	111	26.60	38.94	31.7%
Glucose >110 $\mathrm{mg\,dL^{-1}}$	428	23.83	27.52	13.4%
Any non-fasting criterion	**547**	**23.64**	**28.39**	**16.7%**

Sensitivity analyses with 08:00/12:00 cutoffs changed estimates by <0.3 $\mathrm{mg\,dL^{-1}}$.

The combined non-fasting proxy (any criterion) represents 547/696 (78.6%) of real-world testing—showing sustained 16.7% improvement.

4.4 Labs-Only Configuration Enables Telehealth

Removing blood pressure measurements had minimal impact:

- $\Delta\text{MAE} = -0.12\,\text{mg}\,\text{dL}^{-1}$ (90% CI -0.33 to 0.08)
- TOST non-inferiority confirmed within $\pm 0.5\,\text{mg}\,\text{dL}^{-1}$ margin ($p < 0.001$)
- Near-threshold errors at $100\,\text{mg}\,\text{dL}^{-1}$: 25 (labs-only) vs. 28 (full model)

Implication: The system functions with *only* routine blood tests—no clinic visit needed. Patients can:

- obtain lipid panels at retail pharmacies,
- use home phlebotomy services,
- get testing at workplace health fairs, and
- complete telehealth visits without in-person assessment.

4.5 Racial Equity Without Race-Based Adjustment

Table 3 shows equitable performance across racial/ethnic groups *without* using race as a model input.

Table 3: Performance by Race/Ethnicity (Race Not Used in Model)

Group	**N (%)**	**ML MAE**	**NIH Equation 2 MAE**	**FW MAE**
Black/African American	161 (23)	23.29	27.06	28.66
White	421 (60)	23.48	24.76	26.42
Hispanic	54 (8)	24.02	25.83	27.95
Asian	10 (1)	22.84	23.91	25.17
Other	50 (7)	23.95	25.42	27.53
Black–White gap		**-0.19**	+2.30	+2.24
95% CI		**-4.12 to 3.74**	-1.73 to 6.21	-2.08 to 6.44

All MAE values in $\text{mg}\,\text{dL}^{-1}$. ML achieves equity (Black–White gap CI spans zero); equations show small trends favoring White patients.

The Black/African American–White gap CI spans zero (-0.19 $\text{mg}\,\text{dL}^{-1}$; 95% CI -4.12 to 3.74), indicating no meaningful difference. This addresses concerns about race-based clinical algorithms[6] while improving accuracy for all groups. We pre-specified a race-free model (race not used as input); external validation is required to confirm generalizability.

4.6 Economic Impact: Single-Visit Workflows

Status quo with fasting requirements. In routine practice, a substantial fraction of lipid panels are drawn outside strict fasting windows ($\sim 40\,\%$), and a subset yield unusable or unsafe equation outputs (e.g., TG $> 400\,\text{mg}\,\text{dL}^{-1}$, negative LDL-C). These cases typically trigger a second, fasting visit. At the scale of 10,000 initial tests, this approximates ∼4,000 repeat visits.

With a non-fasting, ML-enabled workflow:

- **Eliminated repeat visits:** $4{,}000 \times \$50 = \$200{,}000$ saved in direct clinic costs.
- **Implementation/QA:** one-time integration \$10,000; annual QA/monitoring \$5,000.

- **Base-case net savings (per 10,000 tests):** $\$200{,}000 - \$15{,}000 = \$\mathbf{185{,}000}$ (direct costs only).
- **Sensitivity analysis with indirect costs:** Including patient time, transportation, and childcare ($\sim$ \$150,000) yields total savings $\approx$ \$335,000.
- **Break-even volume:** tests/year $\approx \frac{\$15{,}000}{0.40 \times \$50} = \mathbf{750}$ for base case, feasible for community health centers.

Note: National projections are illustrative based on conservative assumptions.
Narrative impact: Shifting to *single-visit* non-fasting testing removes the second appointment for thousands of patients per 10,000 panels, cutting direct clinic spend and—more importantly—eliminating travel and missed-work costs that disproportionately burden rural and lower-income patients. Because our labs-only model functions without in-clinic vitals, these gains extend to retail labs, home phlebotomy, and employer health fairs, aligning lipid management with telehealth-era access.

National scale (illustrative): If roughly 5×10^8 lipid panels are performed annually and $\sim 40\,\%$ currently occur outside fasting windows, avoiding up to 2×10^8 repeat visits would translate to \$1–\$2 billion in direct system savings, with larger societal savings when patient time and transportation are included.

4.7 Implementation Pathway

Phase 1 (Year 1): High-access-need populations

- Rural health clinics (travel burden)
- Community health centers (underserved)
- Home health agencies (mobility-limited)
- *Target volume:* 25 000 tests

Phase 2 (Year 2): Telehealth integration

- Partnerships with three telehealth platforms
- Integration with retail pharmacy testing networks
- *Target volume:* 100 000 tests

Phase 3 (Year 3): Broad adoption

- Health system–wide deployment
- Update clinical pathways to permit routine non-fasting workflows
- *Target volume:* $\geq$ 500 000 tests

5 Discussion

5.1 Addressing Three Barriers Simultaneously

Our findings demonstrate that ML can eliminate traditional lipid testing barriers while maintaining accuracy:

1. Fasting eliminated: 17.2% improvement over Friedewald in non-fasting states vs. 10.4% when fasting. The system actually performs *better* exactly when access is expanded.

2. In-person visits eliminated: Labs-only configuration (no blood pressure) changes MAE by only -0.12 $\mathrm{mg\,dL^{-1}}$. Patients can test anywhere: retail pharmacies, home phlebotomy, workplace health fairs.

3. Racial adjustments eliminated: Equity achieved without race input (Black/African American–White gap -0.19 $\mathrm{mg\,dL^{-1}}$, 95% CI -4.12 to 3.74, no meaningful difference), addressing algorithmic bias concerns[6] while improving accuracy for all groups.

5.2 Who Benefits Most

Rural populations: 60% of US counties lack cardiology access.[2] Non-fasting, telehealth-enabled testing eliminates 3–4 hour round trips and overnight stays for early morning fasting appointments.

Diabetic patients: Cannot safely fast 12 hours due to hypoglycemia risk. These patients are disproportionately tested outside fasting windows and face higher retesting burden; enabling accurate non-fasting measurement directly reduces repeat visits and delays in care.

Shift workers: Cannot accommodate morning-only appointments. Non-fasting testing allows any-time collection.

Caregivers: Predominantly women with childcare responsibilities. Retail pharmacy or home testing removes clinic visit burden.

Cost-sensitive patients: Single-visit workflows eliminate $50–100 repeat-visit costs and lost wages.

5.3 Alignment With Clinical Guidelines

The 2016 European Atherosclerosis Society recommended *non-fasting* lipid profiles as the default based on evidence that (i) the postprandial state reflects usual physiology, (ii) fasting requirements reduce compliance, and (iii) cardiovascular risk prediction is comparable between fasting and non-fasting samples [4].

In the United States, adoption has lagged largely due to concerns about calculated LDL-C accuracy in non-fasting states. Our findings address this barrier directly: the ML estimator maintains calibration in non-fasting states (slope ≈ 1.07), whereas equation-based estimates remain poorly calibrated (slope ≈ 0.61).

5.4 Post-COVID Telehealth Landscape

Virtual care increased roughly 154× at the onset of COVID-19 and has stabilized near 38× baseline [5]. In parallel, home phlebotomy and retail pharmacy testing expanded, yet lipid management has not fully adapted because traditional workflows often assume:

1. **Fasting** (restricts when testing can occur),
2. **Blood pressure** availability (in-person or home devices), and
3. **Race-based adjustments** (raise equity concerns).

Our *labs-only*, *race-free inputs with equity audit*, *non-fasting* approach enables lipid management to match telehealth-era accessibility, similar to advances seen in diabetes, hypertension, and anticoagulation programs.

5.5 Economic Case for Payers

Base-case modeling indicates net savings of $185,000 per 10,000 tests (direct clinic costs only) by eliminating repeat fasting visits. Additional benefits include:

- improved medication adherence (convenience → completion),

- earlier treatment initiation (fewer scheduling delays),
- reduced emergency department utilization from untreated dyslipidemia, and
- lower administrative burden (fewer reschedules and callbacks).

A break-even volume of 750 tests in year one (250 tests/year thereafter) makes implementation feasible even for small community health centers serving high-need populations.

5.6 Implementation Considerations

Clinical workflow:

1. Patient obtains a lipid panel at any time and location (no fasting required).
2. Results are processed by the ML service in $<100\,\text{ms}$ on CPU.
3. If prediction is confident, report LDL-C with a 90% confidence interval.
4. If uncertainty overlaps treatment thresholds, *abstain* and document the need for confirmatory testing.

Portal messaging presents both the equation estimate and the learned correction (e.g., "112.6 + (+7.4) = 120.0 $\text{mg}\,\text{dL}^{-1}$[90% CI 114.1 to 125.6]") so clinicians see how ML augments familiar tools.

When a prediction interval spans a treatment threshold, the system abstains and documents the need for confirmatory testing (e.g., direct LDL or apoB) rather than issuing a potentially misleading estimate.

Quality assurance:

- Weekly calibration monitoring: slope 0.95–1.05, intercept $\pm 5\,\text{mg}\,\text{dL}^{-1}$.
- Weekly subgroup audits (AA vs. Others) require the MAE gap 95% CI to span zero; alerts trigger review and, if needed, calibration refresh. Race is not used at inference.
- Override tracking: log clinician disagreements and rationale.
- Automatic rollback to equations if service-level objectives (SLOs) are breached.

Patient communication (example):

> *Your cholesterol test no longer requires fasting. You can have blood drawn at any time at [location list]. Results will be available the same day in your patient portal.*

5.7 Limitations

Fasting status inferred: We used validated proxies (collection time, triglycerides, glucose) rather than documented fasting labels; sensitivity analyses across three proxy definitions yielded consistent results ($< 0.3\,\text{mg}\,\text{dL}^{-1}$ differences).

Single cohort: External validation across diverse health systems and analyzer platforms is required prior to broad deployment; three multi-site validations (academic/community/integrated delivery) are underway.

Outcomes not assessed: We demonstrate maintained measurement accuracy and calibration but did not analyze long-term cardiovascular outcomes; prospective linkage studies are planned.

Economic projections: Cost estimates derive from literature and sensitivity analyses; pragmatic implementation studies are needed to confirm savings and detect unintended consequences.

Generalizability across measurement quality: Performance gains were observed in All of Us data, where equation-direct discrepancies create improvement opportunities. In highly standardized settings (e.g., CDC/CRMLN-calibrated NHANES samples with lower baseline equation error), the marginal benefit of ML may be smaller. Deployment should prioritize real-world clinical laboratories where equation miscalibration is documented.

5.8 Future Directions

Prospective access study: Compare time-to-treatment and patient satisfaction between:

- *Traditional:* fasting requirement with in-person visit, and
- *ML-enabled:* any-time testing with telehealth-capable interpretation.

Primary endpoint: days from initial contact to treatment decision. Secondary: patient-reported barriers and completion rates.
Health equity study: Deploy in three rural health clinics and three urban community health centers; assess:

- testing uptake by race/ethnicity and geography,
- time-to-statin-initiation among high-risk patients, and
- reduction in cardiovascular events over a 3-year follow-up.

Telehealth integration: Partner with national telehealth platforms to enable:

- home phlebotomy with immediate LDL-C interpretation,
- retail-pharmacy testing with results delivered to the patient portal, and
- virtual lipid management visits without in-person requirements.

6 Conclusions

With roughly 40% of real-world lipid testing occurring outside fasting windows, the per–10,000–test base case implies substantial reductions in repeat visits and clinic costs without sacrificing measurement quality. Machine learning maintains performance in non-fasting states (17% improvement over equations), works with labs-only data (enabling telehealth), and achieves racial equity without race adjustment.

By eliminating three major access barriers simultaneously—fasting requirements, in-person visits, and racial adjustments—this approach could expand cardiovascular risk assessment to underserved populations: rural residents traveling hours for testing, diabetic patients unable to safely fast, shift workers requiring flexible scheduling, and caregivers constrained by childcare responsibilities.

Illustratively, scaling the per–10,000–test base case to large systems suggests sizable system-level savings; precise national estimates depend on test volumes and repeat-visit rates. Implementation requires only routine laboratory values and <100 ms computation, making deployment feasible even in resource-limited settings.

External validation and prospective outcomes studies are ongoing, but the convergence of maintained accuracy, expanded access, and economic sustainability suggests a pathway to more equitable cardiovascular care in the post-COVID telehealth era.

Author Declarations

Ethics approval and consent to participate: Analyses were conducted within the All of Us Research Program controlled tier under the NIH single IRB protocol; Howard University determined that additional institutional review board review was not required because only de-identified data were accessed. All participants enrolled in All of Us provided informed consent at the time of data collection.
Consent for publication: Not applicable; the study analyzes de-identified observational data without individual participant reporting.
Trial registration: Not applicable; this observational study is not a clinical trial.
Reporting guideline: The manuscript follows the STROBE recommendations for observational studies, and the completed checklist is available upon request.

Article Information

Author Affiliations

Howard University College of Medicine, Washington, DC, USA (Doku, Osafo, Kwagyan, Southerland).

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.

References

[1] Salim S. Virani, Alvaro Alonso, Hugo J. Aparicio, Emelia J. Benjamin, Marcio S. Bittencourt, Clifton W. Callaway, April P. Carson, Alanna M. Chamberlain, Susan Cheng, Francesca N. Delling, Mitchell S. V. Elkind, Kelly R. Evenson, John F. Ferguson, Deepak K. Gupta, Sadiya S. Khan, Brett M. Kissela, Kristen L. Knutson, Chia-Ying Lee, Tené T. Lewis, Jian Liu, Matthew S. Loop, Pamela L. Lutsey, Jun Ma, Jason Mackey, Seth S. Martin, David B. Matchar, Michael E. Mussolino, Sankar D. Navaneethan, Amanda M. Perak, Gregory A. Roth, Zainab Samad, Gary M. Satou, Emily B. Schroeder, Svati H. Shah, Cora M. Shay, Andrew Stokes, Lisa B. VanWagner, Nae-Yuh Wang, Connie W. Tsao, American Heart Association Council on Epidemiology, Prevention Statistics Committee, and Stroke Statistics Subcommittee. Heart disease and stroke statistics–2021 update: A report from the american heart association. *Circulation*, 143(8):e254–e743, 02 2021. doi: 10.1161/CIR.0000000000000950. Epub 2021 Jan 27.

[2] Kathryn A. Hirko, Julie M. Kerver, Samantha Ford, Christopher Szafranski, Julie Beckett, Cherie Kitchen, and Amy L. Wendling. Telehealth in response to the covid-19 pandemic: Implications for rural health care access. *Nature Medicine*, 28:89–103, 2022.

[3] Zainab Farukhi, Daniel J. Rader, Samia Mora, Seth S. Martin, and Ann Marie Navar. Nonfasting lipids for all patients? *Journal of the American Heart Association*, 9(13):e016549, 2020.

[4] Børge G. Nordestgaard, Anne Langsted, Samia Mora, Genovefa Kolovou, Heiner Baum, Eric Bruckert, Gerald F. Watts, Grazyna Sypniewska, Olov Wiklund, Jan Borén, M. John Chapman, Christa Cobbaert, O. S. Descamps, Arnold von Eckardstein, Pia R. Kamstrup, Kari Pulkki, Florian Kronenberg, Alan T. Remaley, Nader Rifai, Emilio Ros, and Michel Langlois. Fasting is not routinely required for determination of a lipid profile. *European Heart Journal*, 37(25):1944–1958, 2016.

[5] Jedrek Wosik, Marat Fudim, Blake Cameron, Ziad F. Gellad, Alex Cho, Donna Phinney, Simon Curtis, Mark Roman, Eric G. Poon, Jeannette Ferranti, Jeffrey N. Katz, James Tcheng, Geoffrey S. Ginsburg, David W. Bates, John C. Boggan, and Jonathan P. Piccini. Telehealth transformation: Covid-19 and the rise of virtual care. *Journal of the American Medical Informatics Association*, 27(6):957–962, 2020.

[6] Darshali A. Vyas, Leo G. Eisenstein, and David S. Jones. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine*, 383(9): 874–882, 2020.

[7] Maureen Sampson, Chun Li Ling, Qiong Sun, Riad Harb, Talal Ashmaig, Russell Warnick, Amita Sethi, Janet K. Fleming, James D. Otvos, Jeffrey W. Meeusen, Sean R. Delaney, Allan S. Jaffe, Amy K. Saenger, Dmitri Sviridov, Roger D. Shamburek, and Alan T. Remaley. A new equation for calculation of low-density lipoprotein cholesterol in patients with normolipidemia and/or hypertriglyceridemia. *JAMA Cardiology*, 5(5):540–548, 2020.

[8] Ankur Singh, Yuchen Guo, and Seth S. Martin. Machine learning in lipidology. *Current Opinion in Lipidology*, 33(3):231–238, 2022.

[9] Vidya Sathiyakumar, James Park, Azadeh Golozar, Mariana Lazo, Seth S. Martin, Erin D. Michos, Salim S. Virani, Samia Mora, Steven R. Jones, Peter G. Martin, Michael J. Blaha, Peter P. Toth, and Rolando Quispe. Fasting versus nonfasting and low-density lipoprotein cholesterol accuracy. *Circulation*, 137(1):10–19, 2018. doi: 10.1161/CIRCULATIONAHA.117.030677.

[10] 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.

[11] Seth S. Martin, Michael J. Blaha, Mohamed B. Elshazly, Peter P. Toth, Peter O. Kwiterovich, Roger S. Blumenthal, and Steven R. Jones. 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.

[12] Yong Whi Jeong, Jun Hyuk Koo, Ji Hye Huh, Young-Jin Kim, Hoyeon Jeong, Eun Young Kim, and Dae Ryong Kang. Comparison of newly proposed ldl-cholesterol estimation equations. *Journal of Korean Medical Science*, 38:e145, 2023. doi: 10.3346/jkms.2023.38.e145. URL `https://jkms.org/DOIx.php?id=10.3346/jkms.2023.38.e145`.

[13] Yong Jun Kwon, Hyunjin Jeon, Mi-hye Kim, Hye Won Lee, Hye Seung Choi, Sang Min Park, and Jeong Ho Kim. Comparison of a machine learning method and various established ldl-cholesterol estimation methods in korean populations. *Frontiers in Cardiovascular Medicine*, 9:824574, 2022. doi: 10.3389/fcvm.2022.824574. URL `https://www.frontiersin.org/articles/10.3389/fcvm.2022.824574/full`.