Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

ApoB–Sampson LDL-C Discordance: A Cross-Sectional Analysis of Baseline Data from a Prospective Cohort Using Robust Statistical Modeling and an Evidence-Traceable Knowledge Graph

Ashok Soota
SKAN Research Trust
Yogesh Shouche
SKAN Research Trust
Pradeep B S
National Institute of Mental Health and Neurosciences
Srikanth K V
Narayana Health
Komal Prasad Chandrachari
Narayana Health
Prasan Shankar
University of Trans-Disciplinary Health Sciences and Technology
Sanketh V Sharma
University of Trans-Disciplinary Health Sciences and Technology
Chetankumar Naik
chetankumar.naik@happiestminds.com
Happiest Minds Technologies Limited
Bebi Negi
Happiest Minds Technologies Limited
Shashikumar P
Happiest Minds Technologies Limited
Geetha S
Happiest Minds Technologies Limited
Amit Debnath
Happiest Minds Technologies Limited
Devyani Azad
SKAN Research Trust

Ahalya P Gopi
SKAN Research Trust
Amritha S
SKAN Research Trust
Hemanth R
SKAN Research Trust
Lavanya Garady
SKAN Research Trust

Article

Keywords: Apolipoprotein B, LDL cholesterol, biomarker discordance, Knowledge Graph, cardiometabolic phenotype, robust statistical inference

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11508978/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

ApoB–Sampson LDL-C Discordance: A Cross-Sectional Analysis of Baseline Data from a Prospective Cohort Using Robust Statistical Modeling and an Evidence-Traceable Knowledge Graph

Authors: Ashok Soota[1], Yogesh Shouche[2], Pradeep B S[3], Srikanth K V[4], Komal Prasad Chandrachari[5], Prasan Shankar[6], Sanketh V Sharma[7], Chetankumar Naik[8], Bebi Negi[9], P Shashikumar[10], Geetha S[11], Amit Debnath[12] Azad Devyani[13], Ahalya P Gopi[14], Amritha S[15], Hemanth R[16], Lavanya Garady[17]

Authors' affiliations:
[1]Chairman & Managing Trustee, Scientific Knowledge for Ageing and Neurological Ailments Research Trust (SKAN-RT), Bengaluru, Karnataka, India
Email Id: ashok.soota@ask-mr.com

[2]Director, Scientific Knowledge for Ageing and Neurological Ailments Research Trust (SKAN-RT), Bengaluru, Karnataka, India
Email Id: yogesh.shouche@skanrt.in

[3]Professor and Former Head, Department of Epidemiology, Deputy Medical Superintendent, National Institute of Mental Health and Neuro Sciences, Bengaluru, Karnataka, India
Email Id: doctorpradeepbs@gmail.com

[4]Senior Consultant Cardiologist, Narayana Institute of Cardiac Sciences, Bengaluru, Karnataka, India
Email Id: srikanthkv02@gmail.com

[5]Senior Consultant Neurosurgeon and Spinal Surgeon, Narayana Health, Bengaluru, Karnataka, India
Email Id: komal.prasad.dr@narayanahealth.org

[6]Medical Director, The University of Trans-disciplinary, Health Sciences and Technology, Bengaluru, Karnataka, India
Email Id: prasan.shankar@iaimhealthcare.org

[7]Assistant Professor, The University of Trans-disciplinary Health Sciences and Technology, Bengaluru, Karnataka, India
Email Id: sanketsharma@tdu.edu.in

[8]Senior Data Scientist, Happiest Minds Technologies Limited, Bengaluru, Karnataka, India
Email id: chetankumar.naik@happiestminds.com

[9]Associate Principal Data Scientist, Happiest Minds Technologies Limited, Bengaluru, Karnataka, India
Email id: bebi.negi@happiestminds.com

[10]Data Science Architect, Happiest Minds Technologies Limited, Bengaluru, Karnataka, India
Email id: p.shashikumar@happiestminds.com

[11]Senior Data Scientist, Happiest Minds Technologies Limited, Bengaluru, Karnataka, India
Email id: geetha.s3@happiestminds.com

[12]Senior Data Scientist, Happiest Minds Technologies Limited, Bengaluru, Karnataka, India
Email id: amit.debnath@happiestminds.com

[13]Research Assistant, Scientific Knowledge for Ageing and Neurological Ailments Research Trust (SKAN-RT), Bengaluru, Karnataka, India
Email Id: devyani.azad@skanrt.in

[14]Research Assistant, Scientific Knowledge for Ageing and Neurological Ailments Research Trust (SKAN-RT), Bengaluru, Karnataka, India
Email Id: ahalya.gopi@skanrt.in

[15]Senior Executive, Scientific Knowledge for Ageing and Neurological Ailments Research Trust (SKAN-RT), Bengaluru, Karnataka, India
Email Id: amritha.s@skanrt.in

[16] Research Assistant, Scientific Knowledge for Ageing and Neurological Ailments Research Trust (SKAN-RT), Bengaluru, Karnataka, India
Email Id: hemanth.r@skanrt.in

[17]Group Head of Longitudinal Study, Scientific Knowledge for Ageing and Neurological Ailments Research Trust (SKAN-RT), Bengaluru, Karnataka, India
Email Id: lavanya.garady@skanrt.in

Corresponding Author details:
Chetankumar Naik
Senior Data Scientist
Happiest Minds Technologies Limited
Bengaluru, Karnataka, India
Email id: chetankumar.naik@happiestminds.com

Abstract

Apolipoprotein B (ApoB) and low-density lipoprotein cholesterol (LDL-C) reflect related but distinct aspects of atherogenic lipoprotein burden: ApoB approximates atherogenic particle number, whereas LDL-C reflects cholesterol mass within LDL particles. We investigated participant characteristics associated with ApoB variation beyond Sampson-calculated LDL-C in a cross-sectional analysis of 2,802 participants from the Heart Attack and Stroke Predictability (HASP) cohort study.

ApoB and LDL-C were strongly correlated (Pearson r = 0.854), indicating substantial concordance with residual variation suitable for discordance analysis. The analysis proceeded sequentially across adiposity-related, socioeconomic, demographic, and clinical domains. BMI showed a small positive association with ApoB beyond LDL-C (standardized $\beta = 0.047$, $p < 0.001$) with 100% bootstrap directional stability. Standard-of-living achievement showed no significant overall association beyond LDL-C ($F = 2.36$, $p = 0.095$; $\Delta R^2 = 0.00045$), and category contrasts did not remain significant after false-discovery-rate correction.

In a multivariable model, each decade of age was associated with +0.71 mg/dL ApoB, male gender with +1.92 mg/dL, and hypertension history with +3.52 mg/dL after covariate adjustment. All three associations passed FDR correction and showed 100% directional stability across 500 bootstrap resamples. Their inclusion increased R^2 from 0.7298 to 0.7365 ($\Delta R^2 = 0.0066$; robust joint $p = 8.6 \times 10^{-}15$). Evidence-qualified associations were represented in an evidence-traceable Knowledge Graph.

LDL-C explained most ApoB variation, while selected adiposity-related, demographic, and clinical characteristics captured small but reproducible discordance-related heterogeneity.

Keywords: Apolipoprotein B; LDL cholesterol; biomarker discordance; Knowledge Graph; cardiometabolic phenotype; robust statistical inference

1 Introduction

Low-density lipoprotein cholesterol (LDL-C) remains one of the most widely used lipid measures for cardiovascular risk assessment and treatment decisions. However, LDL-C and apolipoprotein B (ApoB) do not describe exactly the same biological property. LDL-C represents the cholesterol mass carried predominantly within LDL particles, whereas ApoB more directly reflects the number of circulating atherogenic lipoprotein particles, because each ApoB-containing atherogenic particle carries a single ApoB molecule (Puri et al. 2024; Soffer et al. 2024). Two individuals may therefore have similar LDL-C concentrations while differing in the number of atherogenic particles circulating in the blood.

This difference becomes important when the two biomarkers do not move proportionately. Previous prospective studies have shown that strong overall agreement between LDL-C and particle-related measures can coexist with clinically relevant discordance at the individual level. In the prospective cohort studied by Mora et al. (2014), LDL-C was strongly related to ApoB and other LDL-related measures, yet cardiovascular risk estimates differed among individuals in whom the measures were discordant. Similar findings from the INTERHEART study demonstrated that when cholesterol concentration and ApoB-related particle burden were discordant, cardiovascular risk was more closely aligned with the particle-number phenotype than with cholesterol concentration alone (Sniderman et al. 2012). These studies established the broader clinical relevance of discordance, but they primarily addressed its relationship with cardiovascular outcomes.

Within the Heart Attack and Stroke Predictability (HASP) cohort, the key question was whether individuals with similar LDL-C concentrations could still differ meaningfully in ApoB, and whether those differences were associated with recognizable participant characteristics. This is clinically relevant because LDL-C reflects the amount of cholesterol carried within atherogenic particles, whereas ApoB more closely reflects the number of those particles. In Indian populations, where dyslipidemia, adiposity, hypertension, and other cardiometabolic abnormalities commonly occur together, these factors may help explain why cholesterol concentration and atherogenic particle burden do not always align closely (Berry et al. 2024; Guptha et al. 2014; Puri et al. 2024; Ramakrishnan et al. 2019).

Indian studies provide important context for this question. Berry et al. (2024), in a community-based study involving participants from urban Delhi and rural Ballabhgarh, reported a strong relationship between ApoB and LDL-C while also showing that ApoB concentrations corresponding to LDL-C targets varied according to triglyceride status. In urban Asian Indians, Guptha et al. (2014) demonstrated substantial variation in lipid characteristics according to BMI, waist circumference, dietary factors, and socioeconomic characteristics. Hypertension, age, and gender also remain central components of cardiovascular risk assessment in Indian populations, with the Great India Blood Pressure Survey demonstrating a substantial burden of hypertension and marked variation across demographic groups (Ramakrishnan et al. 2019). Together, these findings suggest that population-level concordance between lipid measures may conceal smaller participant-level differences associated with anthropometric, socioeconomic, demographic, or clinical characteristics.

The resulting framework therefore moves from a simple biomarker question—how closely do ApoB and LDL-C agree? —to a more clinically interpretable one: which participant characteristics are associated with the residual ApoB variation that remains once LDL-C has already been taken into account? By combining continuous discordance modelling, robust statistical inference, and an evidence-traceable Knowledge Graph, the study aimed

to characterize this heterogeneity within the HASP cohort while maintaining a clear distinction between cross-sectional association, biological interpretation, and future cardiovascular prediction.

2 Methodology

The Heart Attack and Stroke Predictability (HASP) study is a large-scale, population-based retrospective–prospective cohort study being conducted across multiple locations in India, with the overarching objective of developing an Indian-specific risk-prediction framework for myocardial infarction and stroke based on clinically relevant risk factors (Soota et al. 2026). Participants enrolled in the cohort are planned to be followed at regular intervals for tentatively five years. The present study represents a cross-sectional analysis of baseline data from the initial HASP cohort established in Bengaluru, Karnataka, India.

Adults aged 30–80 years were recruited from urban and rural areas of Bengaluru. Eligible participants were expected to remain in Bengaluru during the follow-up period and provided written informed consent. Individuals with a prior self-reported or documented history of heart disease or stroke, as well as those who were chronically debilitated, bedridden, unable to communicate, or otherwise unable to participate in study procedures, were excluded.

For the present analysis, participants were additionally required to have complete ApoB data and the lipid measurements required for calculation of Sampson LDL-C. Following application of the study eligibility and analytical inclusion criteria, 2,802 participants included in the cross-sectional analysis.

2.1 Sampson LDL-C Calculation and Definition of Discordance

Sampson LDL-C was calculated from total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TG) using the Sampson equation (Sampson et al. 2020):

$$LDL-C = \frac{TC}{0.948} - \frac{HDL-C}{0.971} - \left(\frac{TG}{8.56} + \frac{TG \times (non-HDL-C)}{2140} - \frac{TG^2}{16100}\right) - 9.44$$

where non-HDL-C was calculated as:

$$non-HDL-C = TC - HDL-C$$

Where, TC = total cholesterol, HDL-C = high-density lipoprotein cholesterol, and TG = triglycerides

All lipid concentrations were expressed in mg/dL.

ApoB–LDL-C discordance was conceptualized as continuous variation in ApoB relative to that expected for a given Sampson LDL-C concentration. Relatively higher ApoB at a given LDL-C concentration represented positive discordance, whereas relatively lower ApoB represented negative discordance. No binary or percentile-based threshold was imposed to classify participants as concordant or discordant.

For inferential analyses, this concept was operationalized by modelling ApoB directly with Sampson LDL-C included as a conditioning covariate, rather than by using a separately calculated residual as the primary outcome. Participant-characteristic coefficients therefore represented associations with ApoB conditional on LDL-C, thereby quantifying variation in ApoB not accounted for by the corresponding Sampson LDL-C concentration.

2.2 Analytical Framework and Rationale

Figure 1 summarizes the staged, domain-informed analytical framework used to investigate factors associated with ApoB–Sampson LDL-C discordance. The analysis first characterized the overall relationship between ApoB and Sampson LDL-C to establish the degree of biomarker concordance within the cohort. Participant characteristics were then examined sequentially across prespecified domains: BMI as an adiposity-related characteristic, standard-of-living achievement as a socioeconomic characteristic, and age, gender, and hypertension history as demographic and clinical characteristics.

The staged design distinguished the overall ApoB–LDL-C relationship from associations observed after conditioning on LDL-C and from those estimated in a multivariable setting. Statistical evidence was evaluated using robust inference, control for multiple testing, incremental model contribution, and bootstrap assessment of directional stability. Evidence-supported associations were subsequently incorporated into an evidence-traceable Knowledge Graph together with their statistical provenance. The Knowledge Graph was used to organize and trace the analytical evidence and was not interpreted as representing causal relationships.

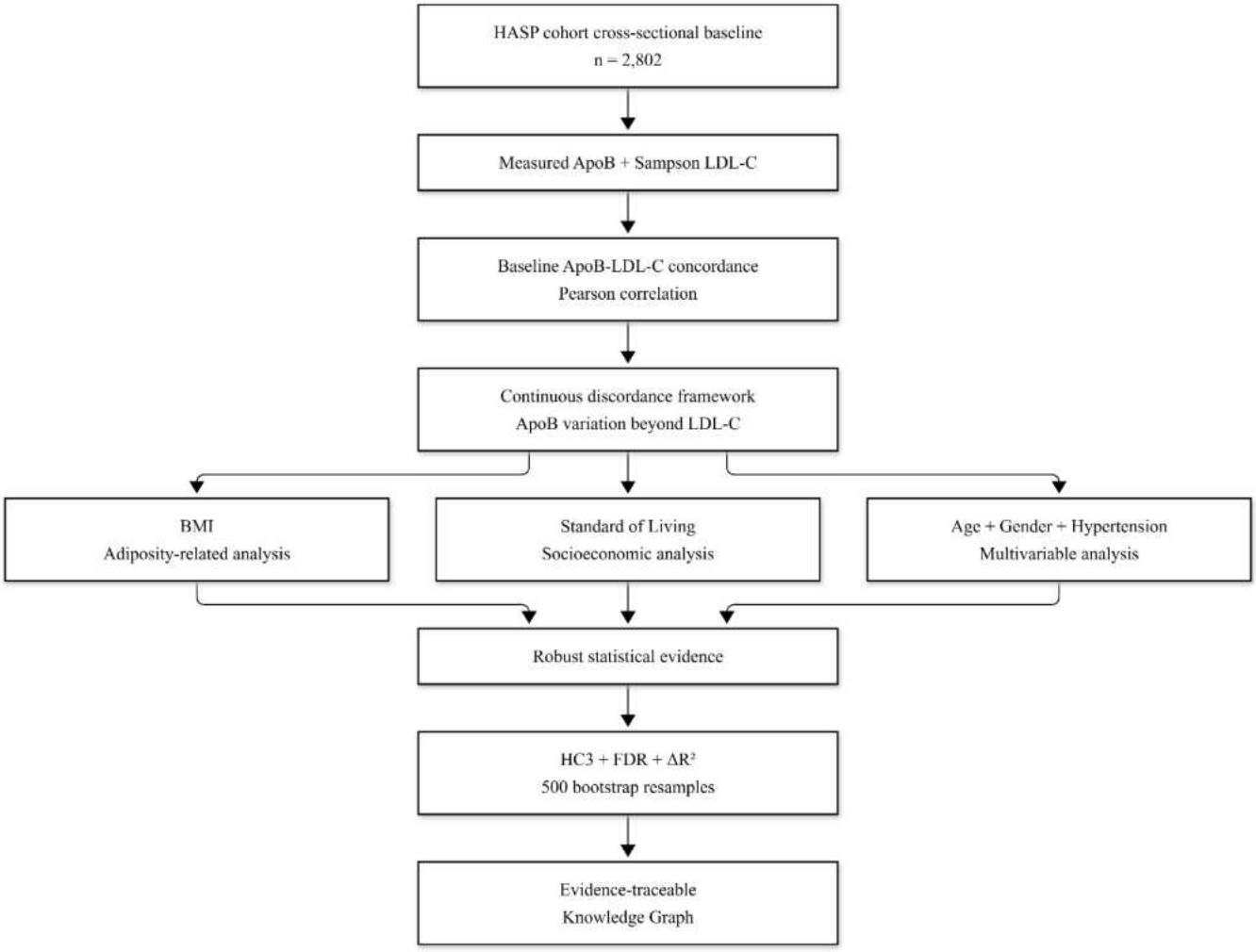

Fig. 1. Staged domain-informed analytical framework for investigating ApoB–Sampson LDL-C discordance.

2.4 Domain-Specific Analytical Strategy

BMI was expressed in kg/m² and represented the adiposity-related domain. Standard-of-living achievement was examined as a socioeconomic characteristic using the three categories recorded at the HASP baseline assessment: "not so much," "to some extent," and "very much." Its overall association with ApoB beyond that explained by Sampson LDL-C was assessed jointly across categories, with category-specific estimates subsequently derived using "not so much" as the reference group.

Age, gender, and hypertension history were evaluated together in a multivariable model adjusted for Sampson LDL-C. Age was modelled as a continuous variable and scaled per 10-year increase, whereas gender and hypertension history were included as binary participant-level variables as recorded in the baseline dataset. This specification

allowed the independent association of each demographic or clinical characteristic with ApoB to be estimated while accounting for LDL-C and the other covariates included in the model.

2.5 Statistical Analysis and Robustness Assessment

Pearson correlation was used to describe crude linear relationships between continuous variables. Linear regression models were then used to evaluate ApoB variation conditional on Sampson LDL-C and the participant characteristics of interest. Effect estimates were reported with 95% confidence intervals where applicable.

2.5.1 Adiposity-Related Discordance: BMI

For the BMI analysis, the principal conditional model was of the form:

$$ApoB_i = \beta_0 + \beta_1(LDL - C_i) + \beta_2(BMI_i) + \varepsilon_i$$

It represents:

- = ApoB concentration for participant i
- = intercept
- = association of LDL-C with ApoB
- = association of BMI with ApoB after accounting for LDL-C
- = residual variation not explained by the model

The BMI association was also reported as a standardized coefficient to show the relative strength of its association with ApoB after accounting for LDL-C.

2.5.2 Socioeconomic Discordance: Standard of Living

For standard-of-living achievement, the LDL-C-only model was compared with a model containing Sampson LDL-C and indicator terms for the standard-of-living categories. The two category coefficients were evaluated relative to the “not so much” reference group, allowing assessment of whether ApoB concentrations differed between standard-of-living groups after accounting for LDL-C. The category terms were also evaluated jointly to determine whether standard-of-living achievement contributed additional explanatory information beyond LDL-C.

$$ApoB_i = \beta_0 + \beta_1(LDL - C_i) + \beta_2 I(SOL_i = \text{to some extent}) + \beta_3 I(SOL_i = \text{very much}) + \varepsilon_i$$

It represents:

= ApoB concentration for participant (i)

= intercept, corresponding to the "not so much" reference category

= association of LDL-C with ApoB

= adjusted ApoB difference for participants reporting "to some extent" compared with "not so much," after accounting for LDL-C

= adjusted ApoB difference for participants reporting "very much" compared with "not so much," after accounting for LDL-C

= indicator variable taking the value 1 when the participant belongs to the specified category and 0 otherwise

= residual variation not explained by the model

The two standard-of-living coefficients were evaluated relative to the "not so much" reference group and were also tested jointly to determine whether standard-of-living achievement contributed additional explanatory information about ApoB beyond Sampson LDL-C.

2.5.3 Demographic and Clinical Discordance: Multivariable Model

The multivariable demographic and clinical model were specified as:

$$ApoBi = \beta 0 + \beta 1(LDL - Ci) + \beta 2(Age10i) + \beta 3(Genderi) + \beta 4(Hypertensioni) + \varepsilon i$$

Where:

- = ApoB concentration for participant
- = intercept
- = association of LDL-C with ApoB
- β_2 = association of age with ApoB after accounting for the other variables
- β_3= association of gender with ApoB after accounting for the other variables
- β_4 = association of hypertension with ApoB after accounting for the other variables
- ε_i = residual variation not explained by the model

2.5.4 Robustness, Multiple-Testing Correction and Model Diagnostics

HC3 heteroskedasticity-consistent covariance estimation was used for regression inference to provide robust standard errors in the presence of potential heteroskedasticity and leverage. Incremental model contribution was

assessed using changes in R^2 together with appropriate effect-size measures. Where the collective contribution of multiple parameters was of interest, robust joint hypothesis tests were performed. For the demographic and clinical model, the joint contribution of age, gender, and hypertension history beyond Sampson LDL-C was assessed using an HC3-robust joint F test.

Where multiple related hypotheses were evaluated, p values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) procedure. FDR correction was applied within related analytical families, including the BMI–biomarker comparisons, standard-of-living category contrasts, and age, gender, and hypertension coefficients in the multivariable demographic and clinical model.

Association stability was additionally assessed using 500 participant-level bootstrap samples generated with replacement. The corresponding model was refitted in each bootstrap sample, and directional stability was defined as the proportion of bootstrap estimates retaining the same coefficient direction as that observed in the original analytical sample. Regression diagnostics included assessment of residual patterns, linearity, leverage, influential observations, and multicollinearity. Potential nonlinear relationships for continuous variables were examined in sensitivity analyses where appropriate.

2.6 Evidence-Traceable Knowledge Graph

Statistically evaluated associations were subsequently organized in an evidence-traceable Knowledge Graph. The graph was used as an evidence-integration and visualization framework rather than as a separate statistical or causal model. Participant characteristics, biomarkers, and the ApoB–LDL-C discordance phenotype were represented as semantically defined nodes. Statistical relationships were represented as edges connecting the corresponding nodes. Each evidence-bearing edge retained information concerning the underlying statistical analysis, including effect direction and magnitude, confidence interval where applicable, statistical support, FDR status, adjustment context, and bootstrap directional stability. The Knowledge Graph therefore did not generate new statistical associations independently of the regression analyses. Instead, it provided a structured representation in which graphical relationships remained traceable to the statistical evidence from which they were derived. The graph was consequently interpreted as an evidence-integration framework rather than as evidence of causal biological pathways.

2.7 Statistical Software and Reproducibility

Data processing and statistical analyses were performed in Python using pandas and NumPy for data management and numerical operations, SciPy for statistical calculations and statsmodels.api for regression modelling, HC3 robust inference, model-fit assessment, joint statistical testing, and multiple-testing procedures.

Python libraries, matplotlib and seaborn were used for statistical figures, while NetworkX was used to construct and represent the evidence-traceable Knowledge Graph. Bootstrap resampling was implemented within the Python analytical workflow using fixed random seeds to support reproducibility. Analytical code, variable definitions, model specifications, statistical testing procedures, and Knowledge Graph construction rules were maintained within the same reproducible analytical workflow.

3 Results

3.1 Relationship Between ApoB and Sampson LDL-C

ApoB and Sampson LDL-C were strongly positively correlated across the 2,802-participant analytical cohort (Pearson r = 0.854), indicating substantial overall concordance between the two biomarkers (Fig. 2). Despite this strong correlation, variation in ApoB remained among participants with similar Sampson LDL-C concentrations, providing the basis for evaluating ApoB–LDL-C discordance as a continuous phenotype.

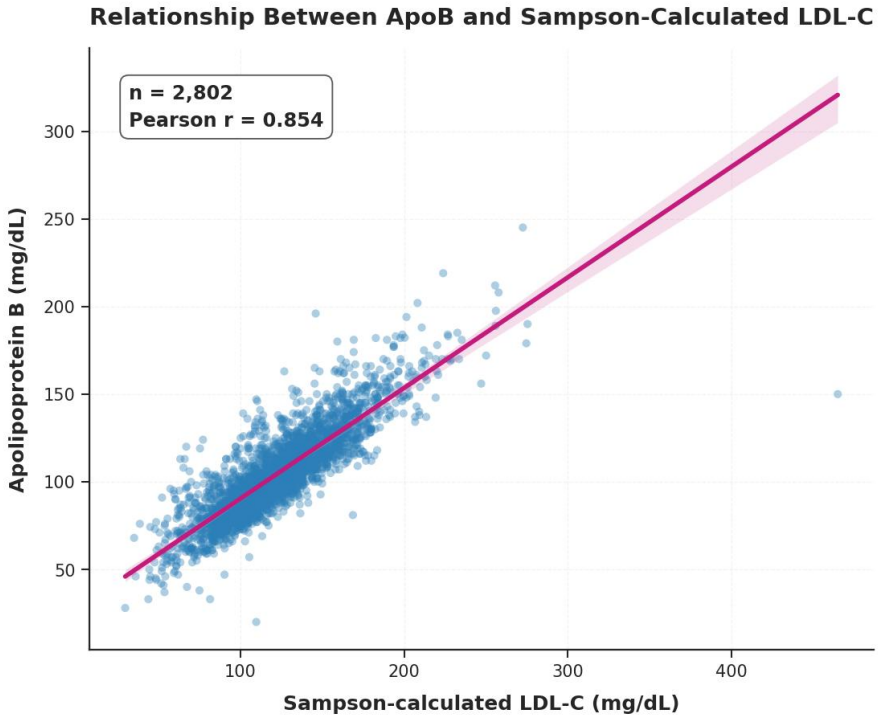

Fig. 2. Relationship between ApoB and Sampson LDL-C in the analytical cohort.

3.2 Association of BMI With ApoB–Sampson LDL-C Discordance

BMI showed little association with Sampson LDL-C (r = 0.017, p = 0.370), whereas a small positive association was observed with ApoB (r = 0.061, FDR-adjusted q = 0.0016).

After accounting for Sampson LDL-C, BMI remained positively associated with ApoB variation beyond LDL-C (standardized $\beta = 0.047$, $p < 0.001$). Although the estimated effect was small, its direction remained positive in 100% of the 500 bootstrap resamples.

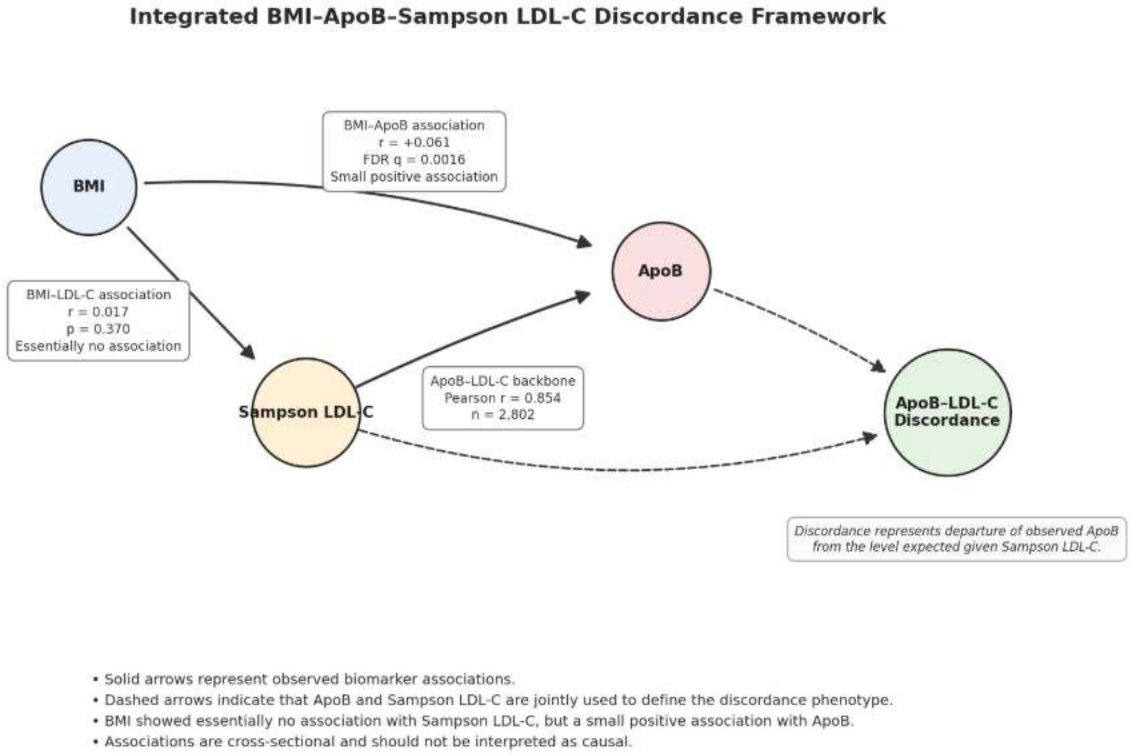

Fig. 3. Knowledge Graph representation of the relationships among BMI, ApoB, Sampson LDL-C, and ApoB–LDL-C discordance.

3.3 Association of Standard-of-Living Achievement with ApoB–Sampson LDL-C Discordance

Standard-of-living achievement showed only a weak association with ApoB after accounting for Sampson LDL-C. Mean ApoB concentrations were slightly higher across the "to some extent" and "very much" groups than in the "not so much" group, but the overall association was not statistically significant ($F = 2.36$, $p = 0.095$) and contributed negligibly to model fit ($\Delta R^2 = 0.00045$; partial $\eta^2 = 0.0017$).

Although the "to some extent" category showed a small positive adjusted ApoB difference (+1.86 mg/dL), this did not remain supported after FDR correction ($q = 0.088$), while the "very much" category was also non-significant ($q = 0.230$). Overall, these findings indicate limited evidence that standard-of-living achievement materially contributed to ApoB–LDL-C discordance in this cohort.

3.4 Multivariable Associations with ApoB Beyond LDL-C

Descriptive mean ApoB and Sampson LDL-C concentrations across hypertension history, gender, and age groups are shown in Fig. 4 A, Fig. 4 B and Fig. 4 C respectively.

Participants with a history of hypertension had higher mean ApoB than those without hypertension (108.5 vs 104.6 mg/dL), whereas mean Sampson LDL-C concentrations were similar between the two groups (123.3 vs 123.7 mg/dL).

Males had a slightly higher mean ApoB concentration than females (106.1 vs 104.8 mg/dL), while mean Sampson LDL-C was marginally lower in males (122.9 vs 124.0 mg/dL).

Across age groups, both biomarkers showed a broadly similar nonlinear pattern. Mean ApoB increased from 99.8 mg/dL at ages 30–39 years to 108.9 mg/dL at ages 50–59 years and subsequently decreased to 105.3 mg/dL at ages 70–80 years. Mean Sampson LDL-C increased from 118.9 mg/dL to 127.6 mg/dL and subsequently declined to 121.4 mg/dL.

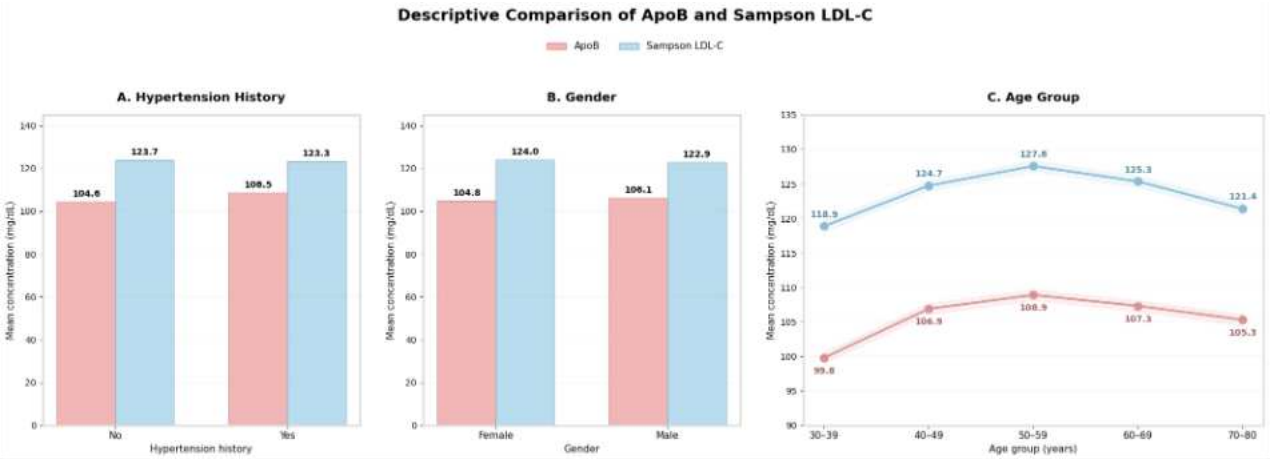

Fig. 4. Descriptive mean ApoB and Sampson LDL-C concentrations across hypertension history, gender, and age groups.

Figure 5 represents the adjusted associations of age, gender, and hypertension history with ApoB after accounting for Sampson LDL-C and the remaining covariates. Edge annotations summarize the corresponding effect estimates, FDR-adjusted statistical support, and bootstrap directional stability.

In the multivariable model containing Sampson LDL-C, age, gender, and hypertension history, each 10-year increase in age was associated with 0.71 mg/dL higher ApoB, male gender with 1.92 mg/dL higher ApoB relative to female gender, and hypertension history with 3.52 mg/dL higher ApoB relative to no hypertension history.

These associations remained statistically supported after FDR correction ($q = 0.0034$, $q = 0.0006$, and $q \approx 0.000002$, respectively) and retained a positive direction in 100% of the 500 bootstrap resamples.

These estimates indicate that participants with otherwise comparable Sampson LDL-C concentrations could nevertheless have systematically different ApoB concentrations according to age, gender, and hypertension history, although the absolute magnitude of these differences was modest.

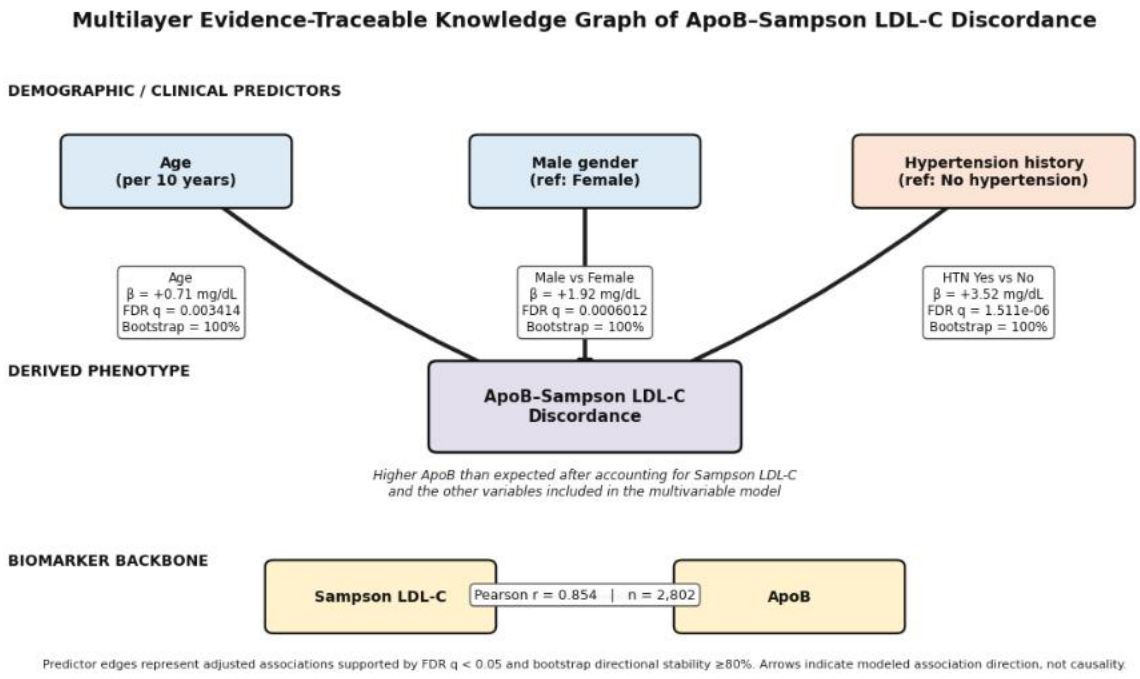

Fig. 5. Evidence-traceable Knowledge Graph representation of multivariable associations with ApoB beyond Sampson LDL-C.

3.5 Incremental Contribution of Demographic and Clinical Predictors Beyond Sampson LDL-C

In the 2,802-participant analytical cohort, Sampson LDL-C alone explained 72.98% of the variation in ApoB ($R^2 = 0.7298$). Addition of age, gender, and hypertension history increased the model R^2 to 0.7365, corresponding to an incremental R^2 of 0.0066, or approximately **0.66% additional explained variance. The** three predictors were jointly associated with ApoB beyond Sampson LDL-C in the HC3-robust global test ($F(3, 2797) = 23.14$, $p = 8.60 \times 10^{-15}$). Thus, age, gender, and hypertension history collectively accounted for a statistically detectable but quantitatively small component of ApoB variation beyond that already explained by Sampson LDL-C.

4. Discussion

This study characterized ApoB–Sampson LDL-C discordance as a continuous phenotype. ApoB and Sampson LDL-C were strongly correlated ($r = 0.854$), indicating substantial overall concordance, while still leaving measurable variation in ApoB among participants with comparable LDL-C concentrations.

Previous Indian studies have linked adiposity with adverse lipid patterns (Guptha et al. 2014), while studies comparing ApoB with cholesterol-based measures have shown that particle burden does not necessarily vary proportionately with cholesterol concentration (Berry et al. 2024). Consistent with this distinction, BMI in the present study showed little association with Sampson LDL-C but a small positive association with ApoB after adjustment for LDL-C (standardized $\beta = 0.047$). Although modest in magnitude, the association retained the same direction across all bootstrapped resamples, suggesting a small but reproducible adiposity-related component of ApoB–LDL-C heterogeneity.

Socioeconomic characteristics have also been associated with lipid patterns in Indian populations (Guptha et al. 2014). In contrast, perceived achievement of expected standard of living contributed almost no additional explanatory information in the present analysis. The overall association was non-significant, and the nominal association observed for the "to some extent" category did not remain supported after FDR correction($Q=0.08$). This finding is specific to the single socioeconomic measure examined and does not exclude associations with broader dimensions of socioeconomic circumstances.

Age, gender, and hypertension history were examined because of their established relevance to cardiovascular health in Indian populations (Chandra et al. 2014; Guptha et al. 2014; Ramakrishnan et al. 2019; Sawhney et al. 2024). After adjustment for Sampson LDL-C and the remaining covariates, ApoB was 0.71 mg/dL higher per decade of age, 1.92 mg/dL higher in males, and 3.52 mg/dL higher among participants with a history of hypertension. All three associations remained FDR-supported and directionally stable across 500 bootstrap resamples. However, their combined incremental contribution to explained ApoB variation was only approximately 0.66%, indicating statistically robust but quantitatively modest associations.

These findings are biologically plausible because LDL-C reflects the cholesterol content carried predominantly within LDL particles, whereas ApoB more closely reflects the number of circulating atherogenic particles. Consequently, individuals with similar LDL-C concentrations may have different particle burdens, providing a basis for ApoB–LDL-C discordance (Mora et al. 2014; Puri et al. 2024; Soffer et al. 2024). Previous discordance studies have demonstrated similar heterogeneity between cholesterol-based and particle-related measures (Mora et al. 2014; Sniderman et al. 2012).

The present study extends this literature by focusing not on future cardiovascular events, but on participant characteristics associated with ApoB variation remaining after LDL-C has been accounted for. The observed

associations should therefore be interpreted as cross-sectional relationships rather than causal effects or indicators of future cardiovascular risk.

4.2 Knowledge Graph and Methodological Contribution

The Knowledge Graph provided a structured way to connect statistical evidence across biomarker, anthropometric, socioeconomic, demographic, and clinical domains while preserving traceability to the underlying analyses. By linking each relationship to its effect estimate, adjustment context, statistical support, and stability measures, the graph made it possible to distinguish stronger evidence from weaker or uncertain associations without collapsing them into a single conclusion.

This approach complements conventional statistical modelling by transforming separate analytical results into an interconnected, evidence-traceable representation that can be extended as additional variables, analyses, or longitudinal outcomes become available. The graph should therefore be viewed as an evidence-integration framework rather than a causal network.

5 Strengths, Limitations and Future Work

5.1 Strengths

A key strength was the use of an interpretable statistical framework that quantified ApoB variation at comparable LDL-C levels without relying on opaque or unnecessarily complex modelling. Continuous discordance preserved the full biomarker signal, while HC3 inference, FDR correction, ΔR^2, and bootstrap resampling tested robustness, multiplicity, explanatory contribution, and stability from complementary perspectives. This allowed statistical significance to be separated from effect magnitude and reproducibility. The Knowledge Graph further strengthened interpretability by linking these evidence dimensions across domains while retaining their analytical provenance, creating a transparent alternative to complex models that may offer greater predictive flexibility but less explanatory clarity.

5.2 Limitations

The cross-sectional design limits conclusions about temporality, causality, or future cardiovascular risk. Only a restricted set of participant characteristics were examined, and the additional variance explained beyond LDL-C was modest. Potential contributors such as triglyceride-related metabolism, glycemic status, central adiposity, diet, physical activity, medication use, genetics, and broader socioeconomic factors were not comprehensively assessed. Bootstrap resampling supports internal stability but does not replace independent external validation.

5.3 Future Work

Future analyses should incorporate broader metabolic, behavioral, pharmacological, socioeconomic, and genetic variables and examine nonlinear relationships and biologically plausible interactions. Alternative interpretable statistical or machine-learning approaches could be compared with the present regression framework to identify additional structure in discordance. External validation in independent populations is needed, while longitudinal follow-up would allow assessment of whether ApoB–LDL-C discordance is associated with subsequent cardiovascular outcomes. Knowledge Graph could then evolve to integrate cross-sectional and longitudinal evidence while preserving traceability.

6. Concluding Perspective

Although ApoB and Sampson LDL-C showed strong overall concordance at the population level, correlation alone does not capture the heterogeneity that may exist among individuals with similar LDL-C concentrations. This study therefore shifted the focus from overall agreement between the biomarkers to the residual ApoB variation present at comparable LDL-C levels and examined whether this discordance was systematically associated with anthropometric, socioeconomic, demographic, and clinical characteristics. The findings suggest that high population-level concordance can coexist with small but reproducible individual-level differences, supporting the value of studying discordance as a distinct and interpretable phenotype.

Declarations

Ethics approval and consent to participate:

Ethical approval for the study has been obtained from multiple institutional ethics committees, including SKAN-RT (SKAN/IEC/2025/05), NIMHANS (NIMHANS/35th IEC [BS & NS DIV.]/2022), Narayana Health (ECR/772/Inst/KA/2016/RR-22), and TDU (TDU/IEC/2024/3). The study is being conducted in accordance with the Indian Council of Medical Research (ICMR) National Ethical Guidelines for Biomedical and Health Research involving Human Participants (2017) and the principles of the Declaration of Helsinki. Written informed consent will be obtained from all participants prior to data collection.

Consent for publication:

All the authors have consented to publication the manuscript.

Competing interests:

The authors don't have any competing interests.

Availability of data and materials:

The study is ongoing, and data might be available on request.

Funding:

This study is funded by Scientific Knowledge for Ageing and Neurological Ailments Research Trust (SKAN-RT).

Authors' contributions:

AS contributed to the conceptualization and study of design. LG and DA contributed to the screening of studies and coordination of the review process. CN contributed to data curation, statistical analysis, methodological development, Knowledge Graph construction, visualization, interpretation of findings, and drafting of the manuscript. PBS, KPC, SKV, PrS, SVS, BN, PS, GS, AD, AG, AS and HR contributed to the critical review and revision of the manuscript. All authors reviewed and approved the final version of the manuscript.

Acknowledgements:

The authors express their gratitude to SKAN-Research trust by granting permission to publish this paper. We sincerely thank all SKAN Research team members for their support.

References

Berry P, Amarchand R, Ramakrishnan L, Gupta R, Kondal D, Bharadiya V, Krishnan A, Tandon N, Prabhakaran D, Roy A (2024) Establishing Apolipoprotein-B and non-high-density-lipoprotein-C goals in Indian population: A cross-sectional study. *Indian Heart J* 76:154–158. doi:10.1016/j.ihj.2024.06.003

Chandra KS, Bansal M, Nair T et al (2014) Consensus statement on management of dyslipidemia in Indian subjects. *Indian Heart J* 66(Suppl 3):S1–S51. doi:10.1016/j.ihj.2014.12.001

Guptha S, Gupta R, Deedwania P et al (2014) Cholesterol lipoproteins and prevalence of dyslipidemias in urban Asian Indians: A cross-sectional study. *Indian Heart J* 66:280–288. doi:10.1016/j.ihj.2014.03.005

Mora S, Buring JE, Ridker PM (2014) Discordance of low-density lipoprotein cholesterol with alternative LDL-related measures and future coronary events. *Circulation* 129:553–561. doi:10.1161/CIRCULATIONAHA.113.005873

Puri R, Bansal M, Mehta V et al (2024) Lipid Association of India 2023 update on cardiovascular risk assessment and lipid management in Indian patients: Consensus statement IV. *J Clin Lipidol* 18:e351–e373. doi:10.1016/j.jacl.2024.01.006

Ramakrishnan S, Zachariah G, Gupta K et al (2019) Prevalence of hypertension among Indian adults: Results from the Great India Blood Pressure Survey. *Indian Heart J* 71:309–313. doi:10.1016/j.ihj.2019.09.012

Sampson M, Ling C, Sun Q et al (2020) A new equation for calculation of low-density lipoprotein cholesterol in patients with normolipidemia and/or hypertriglyceridemia. *JAMA Cardiol* 5:540–548. doi:10.1001/jamacardio.2020.0013

Sawhney JPS, Ramakrishnan S, Madan K et al (2024) CSI clinical practice guidelines for dyslipidemia management: Executive summary. *Indian Heart J* 76(Suppl 1):S6–S19. doi:10.1016/j.ihj.2023.11.271

Sniderman AD, Islam S, Yusuf S, McQueen MJ (2012) Discordance analysis of apolipoprotein B and non-high-density lipoprotein cholesterol as markers of cardiovascular risk in the INTERHEART study. *Atherosclerosis* 225:444–449. doi:10.1016/j.atherosclerosis.2012.08.039

Soffer DE, Marston NA, Maki KC et al (2024) Role of apolipoprotein B in the clinical management of cardiovascular risk in adults: An Expert Clinical Consensus from the National Lipid Association. *J Clin Lipidol* 18:e647–e663. doi:10.1016/j.jacl.2024.08.013

Soota A, Shouche Y, Pradeep BS et al (2026) Assessment of hypertension prevalence and care cascade performance toward the 80-80-80 targets in the HASP cohort study, India. *Sci Rep*. doi:10.1038/s41598-026-65588-0