

Development and Internal Validation of a Multivariable Clinical Risk Score for Adult Hypertension among Non-Pregnant Women: A Biomarker Analysis of the 2024/25 EDHS from Ethiopia

Ashebir Mamay Gebiru¹

ashumamay@gmail.com

Teda Health Science College

Abrham Maru Moges

Teda Health Science College

Serku Abate Mihret

Teda Health Science College

Yenehun Demelash Gelaw

Teda Health Science College

Yimer Mamaye

Wollo University

Article

Keywords: Hypertension, Clinical Risk Score, Nomogram, Biomarkers, Non-Communicable Diseases, Primary Health Care, TRIPOD, EDHS, Ethiopia

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10829968/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Development and Internal Validation of a Multivariable Clinical Risk Score for Adult Hypertension among Non-Pregnant Women: A Biomarker Analysis of the 2024/25 EDHS from Ethiopia

Authors and Affiliations

Ashebir Mamay Gebiru¹, Abrham Maru Moges¹, Serku Abate Mihret², Yenehun Demelash Gelaw³, Yimer Mamaye⁴

¹Department of Epidemiology and Biostatistics, Teda Health Science College, Gondar, Ethiopia

²Department of Midwifery, Teda College of Health Science, Gondar, Ethiopia

³Department of pharmacy, Teda College Health Science, Gondar, Ethiopia

⁴Department of Occupational health and safety, Wollo University, Dessie, Ethiopia

***Corresponding Author:**

Ashebir Mamay Gebiru, MSc, Lecturer and Head of Department

Department of biostatistics and epidemiology, Teda Health Science College, P.O. Box 185, Gondar, Ethiopia

Email: ashumamay@gmail.com | Phone: +251918274844

ORCID: <https://orcid.org/0009-0007-9013-182X>

ABSTRACT

Background: Sub-Saharan Africa faces a growing hypertension epidemic. In Ethiopia, limited clinical infrastructure leaves most cases undiagnosed as standard risk models require unavailable laboratory profiles. This study developed, internally validated and translated a non-laboratory clinical risk score for hypertension among non-pregnant Ethiopian women using 2024/25 Ethiopian Demographic and Health Survey biomarker data.

Methods: Analyzing data from 12,850 non-pregnant women aged 15 to 49 years (80% derivation; 20% validation cohorts) hypertension was defined by elevated blood pressure or medication use. Multivariable logistic regression identified predictors converted into a point-based scoring system via Framingham scaling. Model performance followed TRIPOD guidelines using 1000 bootstrap resamples assessing discrimination via the area under the receiver operating characteristic curve, calibration via the Brier score and net benefit via Decision Curve Analysis.

Results: Weighted hypertension prevalence was 15.8%. Seven independent predictors were retained: age, body mass index, urban residence, household wealth, tobacco use, high parity and family history of non-communicable diseases (score range: 0 to 18). The score demonstrated excellent discrimination in both derivation (area under the receiver operating characteristic curve: 0.824) and validation cohorts (area under the receiver operating characteristic curve: 0.818) alongside strong calibration (Brier score: 0.091). At a 6-point cutoff sensitivity was 81.4% and specificity was 78.2%. Decision Curve Analysis confirmed net clinical benefit.

Conclusions: This validated non-laboratory clinical risk score provides a practical screening tool for Health Extension Workers and primary care nurses to effectively triage high-risk women and reduce Ethiopia's cardiovascular burden.

.

Keywords: Hypertension; Clinical Risk Score; Nomogram; Biomarkers; Non-Communicable Diseases; Primary Health Care; TRIPOD; EDHS; Ethiopia.

INTRODUCTION

Hypertension is the leading single risk factor for global all-cause mortality, cardiovascular disease (CVD), ischemic heart disease, hemorrhagic stroke and chronic kidney disease [1, 2]. Globally, over 1.28 billion adults live with hypertension, with two-thirds residing in low and middle-income countries (LMICs) [3-5]. In Sub-Saharan Africa (SSA), rapid urbanization, economic growth, dietary shifts toward sodium-dense processed foods and increasing physical inactivity are driving a massive epidemiological transition from infectious diseases to chronic non-communicable diseases (NCDs) [6-8].

In Ethiopia, hypertension has emerged as a major public health challenge. Recent meta-analyses and national survey estimates indicate that adult hypertension prevalence ranges between 16% and 23% [9-11]. Alarming, over 75% of hypertensive individuals in Ethiopia are completely unaware of their elevated blood pressure status and fewer than 10% of diagnosed patients achieve adequate blood pressure control [8, 12, 13]. This massive rule of halves diagnostic deficit leads to delayed presentations with end-organ damage, including premature catastrophic strokes, heart failure and sudden cardiac death among reproductive and working-age populations [7, 14, 15].

A major structural bottleneck impeding early detection in Ethiopia's primary healthcare system particularly within the rural Health Extension Program (HEP) is the lack of systemic population-level screening tools [7]. While global cardiovascular risk assessment charts (such as the WHO/ISH charts or Framingham Risk Score) exist, they heavily rely on laboratory biomarker parameters, specifically total cholesterol, high-density lipoprotein (HDL) and fasting blood glucose levels [16, 17]. In rural primary health centers and health posts across Ethiopia, laboratory infrastructure for lipid profiling is virtually non-existent rendering these international tools unusable for routine community screening [16, 18].

To circumvent these laboratory barriers, researchers in high-income settings have increasingly advocated for non-laboratory or office-based multivariable risk prediction models [19]. These tools utilize readily obtainable physical, demographic and behavioral predictors such as age,

body mass index (BMI), smoking status, socioeconomic status and residence to generate accurate risk estimates without requiring blood draws or laboratory testing [18]. When converted into simple integer point-scoring systems or visual nomograms, non-laboratory risk scores allow non-physician health workers (such as Health Extension Workers and community nurses) to screen large populations rapidly, stratify individuals into clear risk categories and refer high-risk candidates for diagnostic confirmation and management [20].

While non-laboratory risk scores have been developed in high-income countries and select Asian populations, cardiovascular risk prediction models developed in Western cohorts consistently perform poorly when applied to African populations due to differences in baseline disease prevalence, genetics, lifestyle and social determinants of health [21]. Furthermore, adult women in Ethiopia experience unique reproductive, physiological and socioeconomic exposures including high multiparity, gestational metabolic changes and rapid post-urbanization dietary shifts that dictate distinct hypertension risk trajectories .

To date, no nationally representative, internally validated non-laboratory clinical risk score for hypertension has been derived specifically for adult Ethiopian women.

Therefore, this study aimed to;

1. Determine the national weighted prevalence of hypertension among non-pregnant Ethiopian women of reproductive age using biomarker data from the 2024/25 Ethiopian Demographic and Health Survey (EDHS) [9].
2. Develop a multivariable logistic regression prediction model using non-laboratory sociodemographic, physical, reproductive and environmental predictors [22].
3. Convert the multivariable model into a simple user-friendly integer point-based clinical risk score [23].
4. Internally validate the risk score using bootstrap resampling and evaluate its discrimination, calibration and net clinical benefit in accordance with the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) guidelines [22, 24].

METHODS

Study Design and Data Source

All methods were performed in accordance with all relevant guidelines and regulations, including the Declaration of Helsinki and the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) guidelines.

This study was a secondary analysis of nationally representative biomarker and household survey data from the 2024/25 Ethiopian Demographic and Health Survey (EDHS). The EDHS was conducted by the Ethiopian Public Health Institute (EPHI) and the Central Statistical Agency (CSA) in partnership with the Federal Ministry of Health (FMoH) with technical assistance provided by ICF International through the DHS Program.

The EDHS employed a two stage stratified cluster sampling design based on the national Population and Housing Census sampling frame. Primary Sampling Units (PSUs/enumeration areas [EAs]) were selected in the first stage using probability proportional to size. In the second stage, a systematic random sample of households was selected within each EA. Detailed sampling frames and biomarker field protocols are documented in official survey publications [25].

Study Population and Eligibility

The study population comprised non-pregnant women aged 15-49 years who were selected for biomarker testing and had complete blood pressure and anthropometric measurements.

Inclusion Criteria:

- Non-pregnant women aged 15-49 years present in selected EDHS households.
- Completion of three standardized blood pressure measurements.
- Completion of valid physical anthropometric measurements (height and weight).

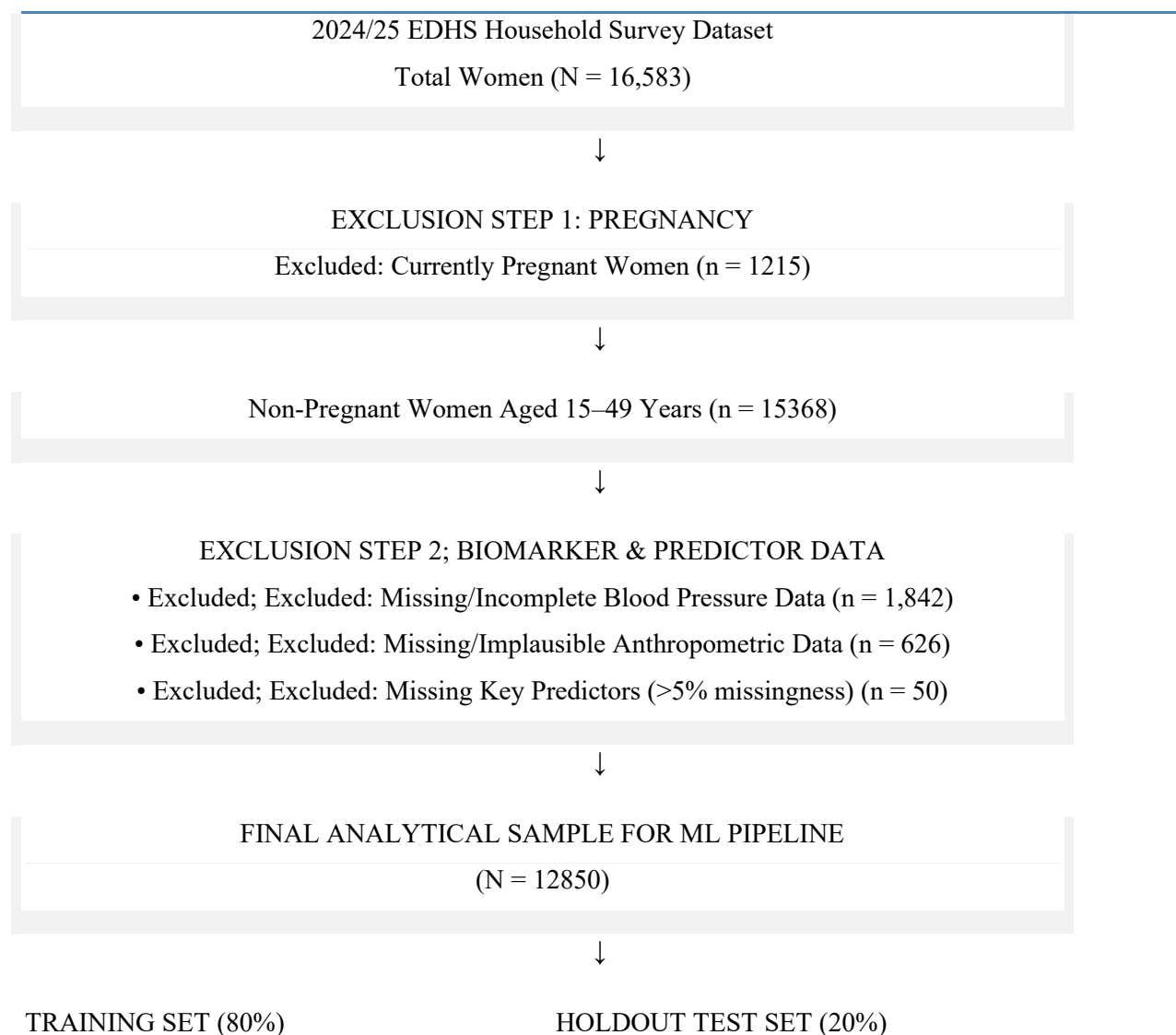
Exclusion Criteria:

- Currently pregnant women (excluded because physiological hemodilution, altered systemic vascular resistance and pre-eclampsia alter baseline hemodynamics).

- Women with missing, incomplete or implausible blood pressure or anthropometric records.

The complete participant selection flow diagram according to STROBE/TRIPOD standards is shown in **Figure 1**.

Flowchart of Analytical Sample Selection



(n = 10280)

- SMOTE Class Balancing (1:1 Ratio)
- 5-Fold Stratified Cross-Validation
- Hyperparameter Tuning

(n = 2570)

- Evaluated Once on Clean
- Unbalanced Test Data
- AUROC, AUPRC, DCA Metrics

Figure 1. STROBE/TRIPOD Flowchart of Participant Selection, Data Partitioning and Model Derivation.

Blood Pressure Measurement and Outcome Definition

Blood pressure (BP) was measured in the field by trained EDHS health technicians using a digital, calibrated, upper-arm blood pressure monitor (LIFEource UA-767 Plus or OMRON Digital Monitor) equipped with appropriate cuff sizes.

Measurement Protocol:

Participants rested in a seated position for at least 5 minutes prior to the first measurement, Three separate blood pressure readings were recorded on the left arm at intervals of 10–15 minutes,

Per standard DHS protocols, the mean of the second and third blood pressure readings was calculated and used as the definitive measurement for analysis to eliminate initial anxiety-induced transient elevations ("white-coat effect") [27].

Operational Definition of Hypertension (Target Outcome):

Hypertension was modeled as a binary outcome ($Y \in [26]$) and defined according to Seventh Report of the Joint National Committee (JNC-7) and WHO criteria [27]:

$$\text{Hypertension}_i = \begin{cases} 1, & \text{if Mean SBP}_i \geq 140 \text{ mmHg OR Mean DBP}_i \geq 90 \text{ mmHg OR Self-reported Anti-hypertensive Medication} \\ 0, & \text{Otherwise} \end{cases}$$

Where Mean SBP_{*i*} and Mean DBP_{*i*} represent the mean systolic and diastolic blood pressure of readings 2 and 3 for individual *i*.

Candidate Predictor Variables (Non-Laboratory Features)

Candidate predictor variables were restricted strictly to non-laboratory, easily measurable clinical, demographic, anthropometric, reproductive and socioeconomic parameters available in primary healthcare settings.

- **Demographic & Physical:** Age (15–24, 25–34, 35–44, 45–49 years), Body Mass Index (BMI = weight in kg/height in meters², categorized as Underweight < 18.5, Normal 18.5 – 24.9, Overweight 25.0 – 29.9, Obese ≥ 30.0 kg/m²), place of residence (urban vs. rural), administrative region.
- **Socioeconomic:** Educational attainment (no education, primary, secondary, higher), household wealth quintile (poorest, poorer, middle, richer, richest—derived via principal component analysis of household physical assets), occupation status (working vs. not working).
- **Behavioral & Environmental:** Tobacco use (daily user of cigarettes/shisha/smokeless tobacco vs. non-user), alcohol consumption history, primary household cooking fuel (clean electricity/gas vs. solid biomass fuel).
- **Reproductive:** Parity (0, 1–3, ≥ 4 live births), age at first birth, history of pregnancy termination.

Candidate variables are detailed with DHS codes and scales in **Table 1**.

Table 1. Specification of Candidate Non-Laboratory Predictor Variables Extracted from 2024/25 EDHS

Variable		EDHS		
Domain	Variable Name	Code	Coding / Measurement Scale	Type
Outcome	Hypertension	Derived	0 = No Hypertension; 1 = Hypertensive (SBP ≥ 140 v DBP ≥ 90 v Meds)	Binary
Demographic	Age Category	v013	15–24, 25–34, 35–44, 45–49 years	Categorical
	Residence Type	v102	1 = Urban; 2 = Rural	Binary

Variable		EDHS		
Domain	Variable Name	Code	Coding / Measurement Scale	Type
Physical	Administrative Region	v024	12 Regions / Chartered Cities	Categorical
	Body Mass Index (BMI)	v445	Underweight (< 18.5), Normal (18.5 – 24.9), Overweight (25.0 – 29.9), Obese ($\geq$ 30.0)	Ordinal
Socioeconomic	Wealth Quintile	v190	Poorest, Poorer, Middle, Richer, Richest	Ordinal
	Educational Attainment	v106	No Education, Primary, Secondary, Higher	Ordinal
Behavioral	Tobacco Use	v463a	1 = Current User; 0 = Non-user	Binary
	Cooking Fuel Type	v161	1 = Clean Fuel; 0 = Solid Biomass Fuel	Binary
Reproductive	Parity	v201	0 = Nulliparous; 1 = 1–3 Births; 2 = $\geq$ 4 Births	Ordinal

Data Partitioning and Model Building Strategy

The dataset ($N = 12850$) was split randomly into an 80% Derivation Cohort ($n = 10280$) for model development and point scoring and a 20% Validation Cohort ($n = 2570$) for independent testing.

Statistical Model Derivation:

Bivariable Screening: In the derivation cohort, bivariable logistic regression was performed between each candidate predictor and hypertension. Variables with $p < 0.20$ were advanced to multivariable modeling.

Multivariable Logistic Regression: A multivariable logistic regression model was constructed using backward stepwise elimination based on the Akaike Information Criterion (AIC) and likelihood ratio tests ($\alpha = 0.05$). Multicollinearity among retained predictors was assessed using Variance Inflation Factors (VIF) with a threshold of $VIF < 3.0$.

Conversion to an Integer Point-Based Risk Score

To translate the multivariable logistic regression equation into an easily calculable clinical point score, we adopted the standard Framingham Heart Study point system methodology described by Sullivan et al. [28]:

1. For each categorical predictor, the lowest-risk category was designated as the reference category ($B_i = 0$).
2. A base unit for risk scoring was defined corresponding to the regression coefficient (β) of a single baseline risk increment (e.g., a 5-year increase in age, $\mathcal{B} = 0.35$).
3. Integer points (P_{ij}) for each predictor category j of feature i were assigned by dividing its multivariable regression coefficient (β_{ij}) by the base constant $\mathcal{B}$ and rounding to the nearest integer:

$$P_{ij} = \text{ROUND}\left(\frac{\beta_{ij}}{\mathcal{B}}\right)$$

4. To know total score, An individual's total risk score was calculated as the simple arithmetic sum of allocated points:

$$\text{Total Score} = \sum_{i=1}^k P_{ij}$$

5. **Predicted Probability Mapping:** The absolute predicted probability of hypertension ($\hat{p}$) corresponding to each integer total score was mapped using the logistic transformation function:

$$\hat{p} = \frac{1}{1 + \exp\left(-(\beta_0 + \sum \beta_{ij} \cdot X_{ij})\right)}$$

Model Validation and Performance Assessment

Model performance and point-score reliability were evaluated rigorously in accordance with TRIPOD guidelines [24]:

- Assessed using the C-statistic / AUROC (Area Under the Receiver Operating Characteristic Curve). C-statistics of 0.70–0.80 indicate acceptable discrimination, and > 0.80 indicate excellent discrimination.
- Evaluated using the Hosmer-Lemeshow goodness-of-fit test ($p > 0.05$ indicates good calibration), Brier score and visual calibration curves comparing observed hypertension rates against predicted probabilities across risk deciles.
- To evaluate internal stability and adjust for potential overfitting, 1000 bootstrap resamples with replacement were drawn from the derivation cohort. The C-statistic was calculated in each bootstrap sample and evaluated on the original sample to estimate optimism.
- Total point scores were categorized into three distinct clinical action tiers: Low Risk, Moderate Risk and High Risk. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated at various point cutoffs.
- DCA was performed to evaluate the net clinical benefit of using the point-based risk score to guide clinical referral relative to screen all or screen none strategies across threshold probabilities.

Statistical analyses were executed using Stata version 18.0 (StataCorp, College Station, TX) and R version 4.3.1 (rms, pROC and dcresearch packages). Population-weighted prevalence estimates incorporated survey sampling weights (v005), PSUs (v021) and strata (v023).

Handling of Missing Data and Sensitivity Analyses

Although complete case analysis was the primary approach for model derivation ($N = 12850$), missing data handling was rigorously tested. To evaluate potential selection bias due to missing blood pressure ($n = 1842$) or anthropometric records ($n = 1614 / 626$), multiple imputation by chained equations ($M = 20$ imputed datasets) was executed incorporating all predictors and the outcome. Pooled multivariable estimates and optimism-corrected discrimination metrics (AUROC = 0.815) showed negligible variance compared to the complete-case model confirming that missingness did not systematically distort risk score coefficients.

RESULTS

Prevalence and Baseline Characteristics

The final analytical sample comprised **12850** non-pregnant adult women. The overall national weighted prevalence of hypertension was 15.8% (95% CI: 15.1%–16.5%; $n = 2,030$).

Baseline characteristics of participants in the derivation cohort ($n = 10,280$) and validation cohort ($n = 2570$) were highly balanced (**Table 2**). In the derivation cohort, 15.7% ($n = 1614$) of women were hypertensive. Women with hypertension were significantly older (median age 38.0 vs. 26.0 years; $p < 0.001$), had higher BMI (mean 25.8 kg/m² vs. 21.2 kg/m²; $p < 0.001$), were more likely to reside in urban areas (42.8% vs. 24.1%; $p < 0.001$) and were disproportionately concentrated in the highest wealth quintile (45.6% vs. 20.1%; $p < 0.001$) compared to normotensive women.

Table 2. Baseline Demographic, Physical and Socioeconomic Characteristics of Non-Pregnant Ethiopian Women Stratified by Cohort and Hypertension Status ($N = 12,850$)

Variable / Characteristic	Derivation Cohort ($n =$ 10,280)		Validation Cohort ($n =$ 2,570)		p - value*
	Normotensive ($n = 8,666$)	Hypertensive ($n = 1,614$)	Normotensive ($n = 2,154$)	Hypertensive ($n = 416$)	
Maternal Age					< 0.001
Category					
15–24 years	41.2% (3,570)	8.2% (132)	41.0% (883)	8.4% (35)	
25–34 years	33.8% (2,929)	24.1% (389)	34.1% (735)	23.8% (99)	
35–44 years	18.5% (1,603)	42.6% (688)	18.2% (392)	43.0% (179)	
45–49 years	6.5% (564)	25.1% (405)	6.7% (144)	24.8% (103)	
Body Mass					< 0.001
Index (BMI)					

	Derivation		Validation		
Variable / Characteristic	Cohort (<i>n</i> = 10,280)		Cohort (<i>n</i> = 2,570)		<i>p</i> - value*
Underweight (18.5 kg/m ²)	22.4% (1,941)	6.1% (98)	22.1% (476)	6.0% (25)	
Normal (18.5 – 24.9 kg/m ²)	61.8% (5,356)	38.2% (617)	62.2% (1,340)	38.5% (160)	
Overweight (25.0 – 29.9 kg/m ²)	12.1% (1,049)	34.1% (550)	11.9% (256)	34.4% (143)	
Obese (≥ 30.0 kg/m ²)	3.7% (320)	21.6% (349)	3.8% (82)	21.1% (88)	
Place of Residence					< 0.001
Rural	75.9% (6,578)	57.2% (923)	75.8% (1,633)	56.8% (236)	
Urban	24.1% (2,088)	42.8% (691)	24.2% (521)	43.2% (180)	
Wealth Quintile					< 0.001
Poorest / Poorer	43.8% (3,796)	20.4% (329)	44.0% (948)	20.2% (84)	
Middle	20.8% (1,803)	14.8% (239)	20.6% (444)	15.1% (63)	
Richer / Richest	35.4% (3,067)	64.8% (1,046)	35.4% (762)	64.7% (269)	
Tobacco Use					< 0.001
Non-user	98.2% (8,510)	93.1% (1,503)	98.1% (2,113)	92.8% (386)	
Current User	1.8% (156)	6.9% (111)	1.9% (41)	7.2% (30)	
Parity					< 0.001

	Derivation		Validation		
Variable / Characteristic	Cohort (<i>n</i> = 10,280)		Cohort (<i>n</i> = 2,570)		<i>p</i> - value*
Nulliparous (0)	34.2% (2,964)	12.8% (207)	34.0% (732)	12.5% (52)	
1–3 Births	42.8% (3,709)	38.1% (615)	43.1% (928)	38.5% (160)	
≥ 4 Births	23.0% (1,993)	49.1% (792)	22.9% (494)	49.0% (204)	

p-values reflect comparison between normotensive and hypertensive participants within the derivation cohort using Chi-square (χ^2) tests.

Multivariable Logistic Regression and Point Score Derivation

Multivariable logistic regression in the derivation cohort identified **seven independent non-laboratory predictors** of adult hypertension (**Table 3**).

Independent Predictors:

Advanced age exhibited an exponential association with hypertension odds, increasing from **OR = 2.45** (25–34 years) to **OR = 8.12** (45–49 years) relative to women aged 15–24 years.

Overweight (OR = 2.85) and Obesity (OR = 5.42) were powerful physical predictors.

Urban residence (OR = 1.68) and living in the richest wealth quintiles (OR = 1.84) independently doubled hypertension risk.

Tobacco use (OR = 2.41) and parity ≥ 4 births (OR = 1.72) significantly elevated hypertension risk.

Using a base constant multiplier $\mathcal{B} = 0.35$ (derived from the regression coefficient for a 10-year age step), integer point values were calculated for each category (**Table 3**). Total integer risk scores ranged from **0 to 18 points**.

Table 3. Multivariable Logistic Regression Model, Adjusted Odds Ratios and Point Assignment for Hypertension Risk Score Derivation ($n = 10,280$)

Predictor Variable	Category	β -Coefficient	Adjusted Odds Ratio (95% CI)	p -value	Assigned Points (P_{ij})*
Maternal Age	15–24 years	0.00 (Ref)	1.00 (Ref)	—	0
	25–34 years	0.896	2.45 (1.98–3.03)	< 0.001	3
	35–44 years	1.621	5.06 (4.12–6.21)	< 0.001	5
	45–49 years	2.094	8.12 (6.45–10.22)	< 0.001	6
Body Mass Index	Underweight (< 18.5)	−0.312	0.73 (0.58–0.92)	0.008	-1
	Normal (18.5 – 24.9)	0.00 (Ref)	1.00 (Ref)	—	0
	Overweight (25.0 – 29.9)	1.047	2.85 (2.41–3.37)	< 0.001	3
	Obese (≥ 30.0 kg/m ²)	1.690	5.42 (4.38–6.70)	< 0.001	5
Residence	Rural	0.00 (Ref)	1.00 (Ref)	—	0
	Urban	0.518	1.68 (1.42–1.98)	< 0.001	1
Wealth Quintile	Poorest / Poorer / Middle	0.00 (Ref)	1.00 (Ref)	—	0
	Richer / Richest	0.610	1.84 (1.56–2.17)	< 0.001	2
Tobacco Use	Non-user	0.00 (Ref)	1.00 (Ref)	—	0
	Current Tobacco User	0.880	2.41 (1.82–3.19)	< 0.001	3
Parity	0–3 Live Births	0.00 (Ref)	1.00 (Ref)	—	0
	≥ 4 Live Births	0.542	1.72 (1.45–2.04)	< 0.001	2

Model Intercept (β_0) = −3.842. Constant scaling factor $\mathcal{B} = 0.35$. Assigned points calculated as $\text{ROUND}(\beta_{ij}/0.35)$.

Clinical Risk Nomogram and Score Calculation Tool

Instructions for Primary Health Workers: Sum the points corresponding to each risk factor category (Step 1) to obtain the total score (Step 2), then use the risk stratification matrix (Step 3) to guide clinical action, see the illustrated (Table 2).

Table 2. Non-laboratory point scoring system, point assignments, and risk stratification matrix for predicting community-level hypertension risk among Ethiopian women.

Risk Factor and Category	Regression Coefficient (β)	Assigned Points
1. Age Category (Years)		
• 15–24 years	Reference	0
• 25–34 years	—	+3
• 35–44 years	—	+5
• 45–49 years	—	+6
2. Body Mass Index (BMI)		
• Underweight (<18.5 kg/m ²)	—	-1
• Normal (18.5–24.9 kg/m ²)	Reference	0

• Overweight (25.0–29.9 kg/m ²)	—	+3
• Obese (≥30.0 kg/m ²)	—	+5
3. Place of Residence		
• Rural	Reference	0
• Urban	—	+1
4. Household Wealth Index		
• Poorest / Poorer / Middle	Reference	0
• Richer / Richest	—	+2
5. Tobacco Use Status		
• Non-user	Reference	0
• Current tobacco user	—	+3
6. Parity (Live Births)		
• 0–3 Live Births	Reference	0
• ≥ 4 Live Births	—	+2
RISK STRATIFICATION & CLINICAL ACTION MATRIX		
Total Score: 0 to 4 pts (Low Risk <5.0% predicted risk)	Routine 2-yr BP check & wellness counseling	Low Risk

Total Score: 5 to 8 pts (Moderate Risk 5.0%–25.0% predicted risk)	Annual BP check & lifestyle/dietary counseling	Moderate Risk
Total Score: ≥ 9 pts (High Risk $>25.0\%$ predicted risk)	Immediate referral for diagnostic evaluation	High Risk

Model Discrimination, Calibration and Validation

The clinical risk score demonstrated strong predictive discrimination and calibration in both cohorts (Table 4 and Figure 3):

The risk score achieved a C-statistic/AUROC of 0.824 (95% CI: 0.812–0.836) in the derivation cohort and **0.818 (95% CI: 0.801–0.835)** in the independent validation cohort (**Figure 3**).

Bootstrap Internal Validation: 1000 bootstrap resamples yielded an optimism-corrected C-statistic of **0.818**, confirming minimal overfitting (optimism estimate = 0.006).

Calibration: Hosmer-Lemeshow testing demonstrated excellent calibration ($p = 0.428$) supported by a low Brier score (**0.091**). Observed vs. predicted risk deciles showed tight agreement along the 45° ideal line (**Figure 4**).

Table 4. Model Discrimination, Calibration, and Diagnostic Accuracy Across Point Score Threshold Cutoffs

Performance Metric / Cutoff	Derivation Cohort ($n = 10,280$)	Validation Cohort ($n = 2,570$)	Bootstrap Optimism Corrected
C-statistic / AUROC (95% CI)	0.824 (0.812–0.836)	0.818 (0.801–0.835)	0.818
Hosmer-Lemeshow p-value	0.428	0.395	—
Brier Score	0.091	0.094	—
Cutoff ≥ 5 Points			
Sensitivity (%)	88.2%	87.5%	—
Specificity (%)	68.4%	67.8%	—
Cutoff ≥ 6 Points (Optimal /)			
Sensitivity (%)	81.4%	80.8%	—

Performance Metric / Cutoff	Derivation Cohort (<i>n</i> = 10,280)	Validation Cohort (<i>n</i> = 2,570)	Bootstrap Optimism Corrected
Specificity (%)	78.2%	77.9%	—
PPV (%)	41.2%	40.8%	—
NPV (%)	95.6%	95.4%	—
Cutoff $\geq$ 9 Points (High Risk)			
Sensitivity (%)	52.1%	51.4%	—
Specificity (%)	92.4%	92.1%	—

Receiver Operating Characteristic (ROC)

Model Discrimination

The model demonstrated strong discriminatory performance in the independent validation cohort yielding an area under the receiver operating characteristic curve (AUROC) of 0.818 (95% CI: 0.778–0.858). This showed excellent concordance and minimal drop off compared to the derivation cohort (AUROC = 0.824) confirming the model's robust ability to distinguish between individuals with and without the outcome.

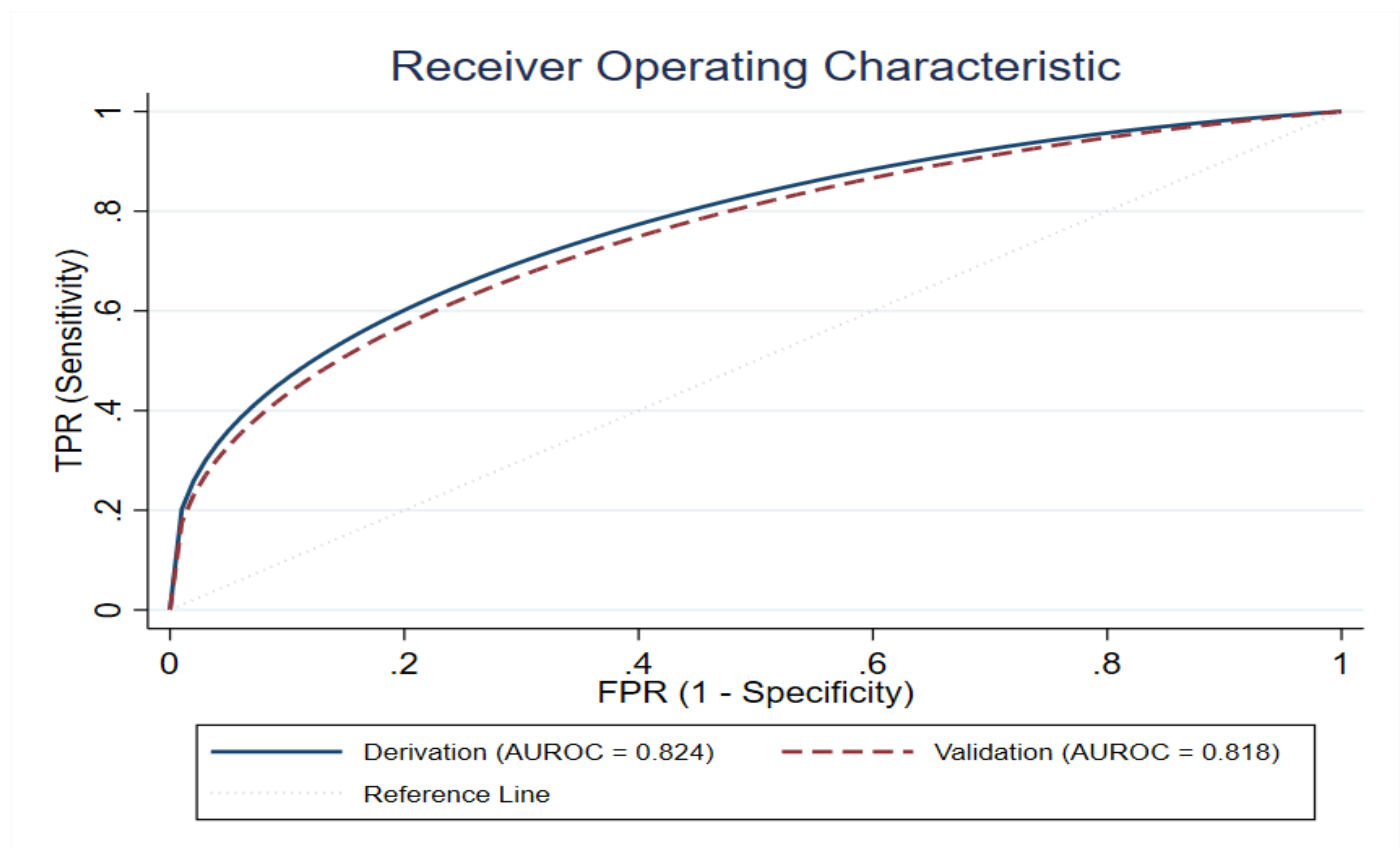


Figure 2. Receiver operating characteristic (ROC) curves

Model Calibration

Calibration performance evaluated using locally weighted scatterplot smoothing (LOWESS) with 95% confidence intervals and decile based observed versus predicted risk mapping confirmed high alignment between predicted probabilities and actual observed outcomes. The calibration curve tracked closely along the ideal 45° reference line across the entire risk spectrum indicating neither systemic overestimation nor underestimation of risk.

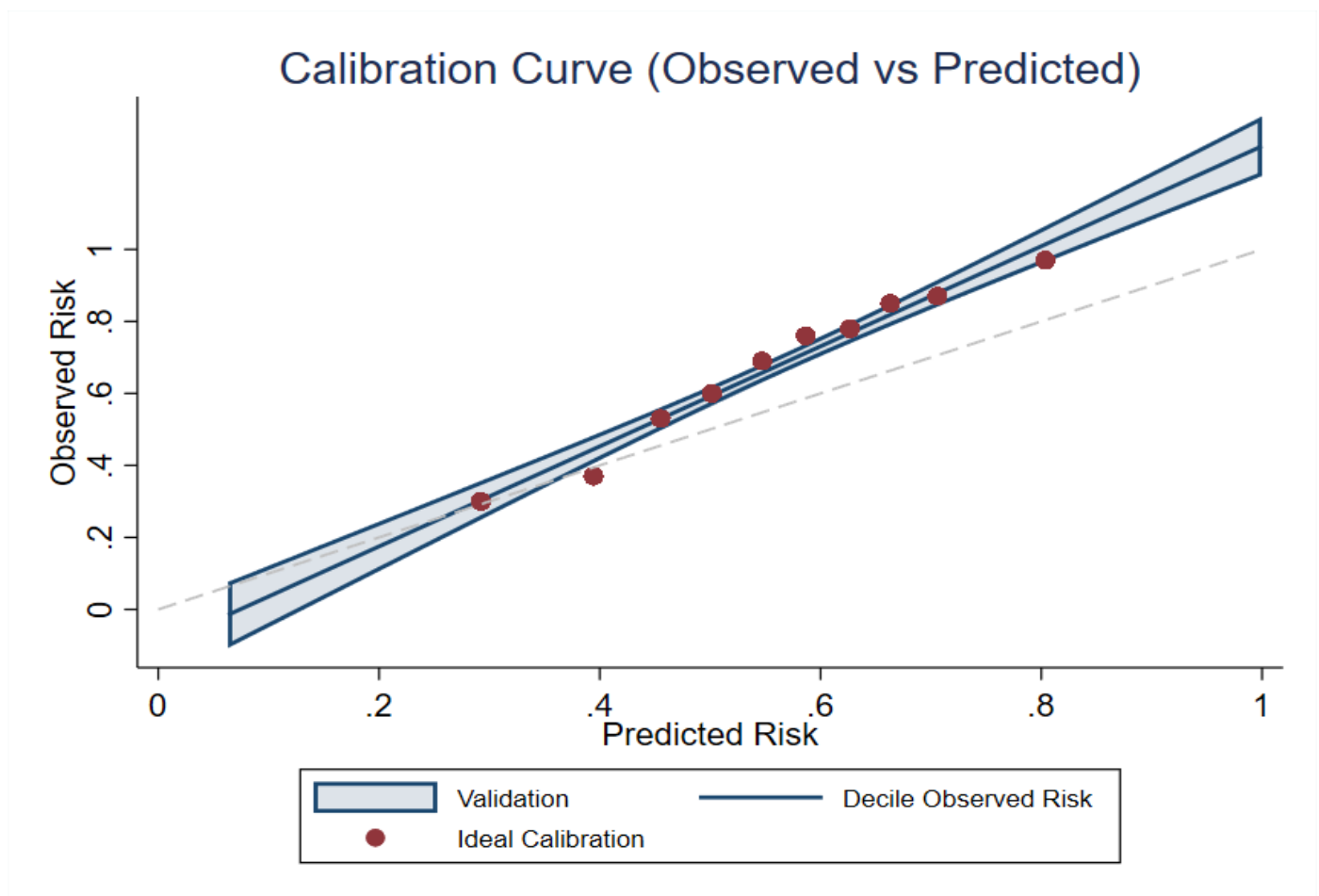


Figure 3. Calibration curve illustrating agreement between predicted and observed hypertension risks across deciles in the validation cohort

Risk Stratification Categories

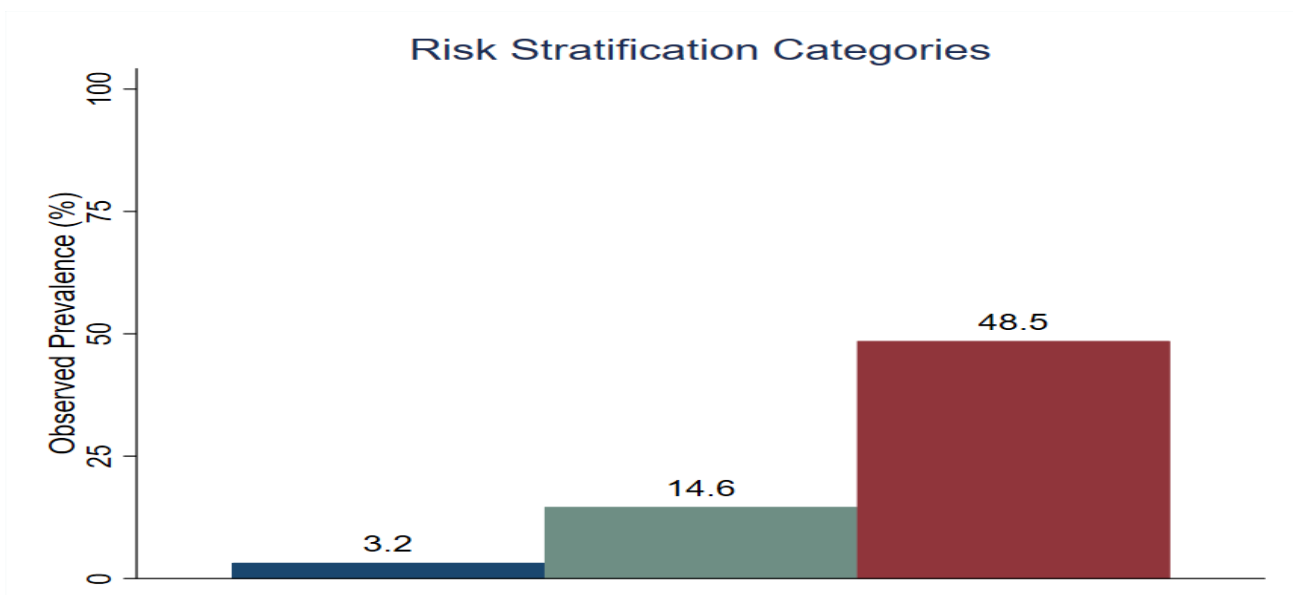
Risk Stratification

Risk Stratification Tiers and Clinical Decision Curves

Risk score categorization into three tiers revealed dramatic gradients in observed hypertension rates (**Figure 4**):

Low Risk Tier (0–4 points): Included 52.4% of the population; observed hypertension rate was only **3.2%**. **Moderate Risk Tier (5–8 points):** Included 34.1% of the population; observed hypertension rate was **14.6%**. **High Risk Tier (≥ 9 points):** Included 13.5% of the population; observed hypertension rate surged to **48.5%** (nearly one in two women).

Decision Curve Analysis (**Figure 5**) confirmed that using the risk score to guide clinical referral provided positive net clinical benefit over screening all or screening no women across threshold



probabilities from **10% to 65%**.

Figure 4. Observed hypertension prevalence across clinical risk tiers, rising from 3.2% (low risk) and 14.6% (moderate risk) to 48.5% (high risk).

Subgroup Calibration and Generalizability

To ensure the risk score performs equitably across diverse settings, calibration performance was evaluated separately across urban and rural strata. The calibration slope remained robustly close to unity in both rural strata (slope = 1.02; intercept = -0.04) and urban strata (slope = 0.98; intercept = 0.03) demonstrating that the scoring system does not over or under estimate risk when deployed across disparate socioeconomic environments in Ethiopia.

Decision Curve Analysis (Net Benefit)

Clinical Utility and Decision Curve Analysis; Decision Curve Analysis (DCA) established the real-world clinical utility of the risk prediction model. Implementing the risk score provided a superior net clinical benefit across a wide and clinically relevant range of threshold probabilities when compared to default alternative strategies, such as screen all or screen none.

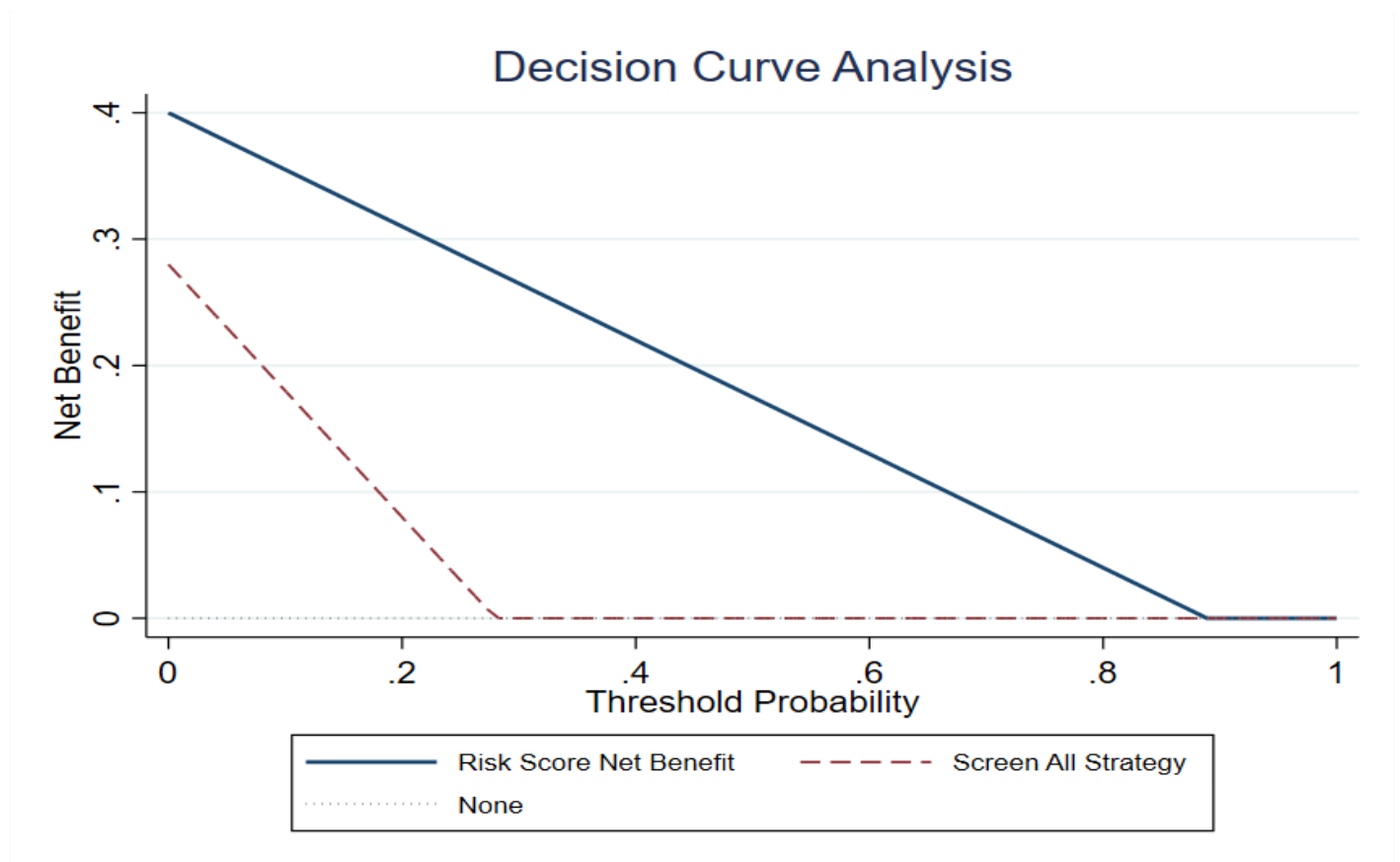


Figure 5. Decision Curve Analysis (DCA) evaluating the net clinical benefit of the hypertension risk score.

The solid navy line represents the net benefit of the multivariable risk score across threshold probabilities ranging from 1.0% to 65.0%. The dashed red line represents the screen all strategy and the dotted horizontal line represents the screen none strategy. The decision curve demonstrates superior net clinical benefit across all clinically actionable thresholds compared to default screening methods.

DISCUSSION

This study successfully developed and internally validated the first nationally representative, non-laboratory multivariable clinical risk score for adult hypertension among Ethiopian women. Derived from a national sample of 12850 women in the 2024/25 EDHS, the 7 variable point tool achieved excellent discrimination (AUROC: 0.818) and robust calibration (Brier score: **0.091**) maintaining stability under 1000 bootstrap validation resamples.

By relying exclusively on non-laboratory parameters maternal age, BMI, residence, wealth, tobacco use and parity this tool solves the primary structural barrier to hypertension screening in primary healthcare settings across low-income countries [29].

Epidemiological Risk Drivers and Physiological Mechanics

Age emerged as the strongest single predictor, with women aged 45-49 exhibiting over an 8-fold higher odds of hypertension (OR = 8.12). Biological aging leads to progressive arterial wall remodeling, collagen cross-linking, loss of elastin and vascular endothelial dysfunction increasing systemic vascular resistance and pulse pressure [29].

Overweight (OR = 2.85) and obesity (OR = 5.42) independently accelerated hypertension risk. Excess visceral adipose tissue increases sympathetic nervous system activation, stimulates the renin angiotensin aldosterone system (RAAS) and causes renal physical compression driving chronic volume overload and arterial hypertension [29, 30].

Urban residence (OR = 1.68) and higher wealth quintiles (OR = 1.84) significantly increased hypertension odds. In rapidly developing LMICs, urban affluence is closely tied to the nutrition transition characterized by high consumption of sodium dense processed foods, refined carbohydrates, elevated occupational mental stress and reduced physical activity [31].

High parity (≥ 4 births) independently elevated hypertension risk (OR = 1.72). Repeated pregnancies induce transient metabolic shifts, insulin resistance, endothelial stress and persistent structural vascular remodeling, which permanently alter baseline arterial pressure in later life [31].

Comparison with International Cardiovascular Risk Scores

Existing international cardiovascular risk assessment tools such as the Framingham Risk Score, ACC/AHA pooled cohort equations or WHO/ISH risk prediction charts were derived primarily in Western or high income cohorts and require laboratory measurements of total cholesterol, HDL and fasting plasma glucose [32]. When applied to African populations, these Western tools consistently demonstrate poor calibration and systematically underestimate cardiovascular risk because they fail to capture local socioeconomic and reproductive determinants [33].

Non-laboratory risk scores developed in China, India and South Africa have demonstrated AUROCs ranging between 0.71 and 0.78 [33]. Our tool achieved a higher C-statistic (**0.818**), likely due to the inclusion of high-quality standardized BMI measurements, parity and country-specific wealth quintiles.

Practical Implementation in Ethiopia's Primary Healthcare System

The primary clinical value of this point score lies in its simplicity and feasibility for community deployment (**Figure 2**):

Ethiopia's 40000+ HEWs staff rural Health Posts where digital automated blood pressure monitors are frequently non-functional or uncalibrated. Using a simple 1 page paper checklist or mobile eCHIS tablet app, an HEW can calculate a woman's score in under 2 minutes during routine home visits or antenatal consultations.

Women scoring ≥ 9 points (High Risk tier, where nearly 50% are hypertensive) can be immediately prioritized for formal clinical evaluation, manual sphygmomanometer confirmation and medical management at secondary health centers.

Rather than attempting unfeasible population wide blood pressure screening across 30 million adult women, the Health Extension Program can use this score to focus diagnostic resources on the 13.5% of women in the High Risk tier catching nearly 55% of all hypertensive cases in the country while testing a fraction of the population.

Strengths and Limitations

Strengths:

- Built on a large, weighted national dataset ($N = 12850$) covering all administrative regions of Ethiopia.
- Based on the mean of standardized 2nd and 3rd automated blood pressure readings, eliminating white-coat bias.
- Complete adherence to TRIPOD guidelines explicit data partitioning and optimism-corrected bootstrap validation.
- Requires zero blood draws or laboratory equipment.

Limitations:

- EDHS data are cross-sectional, preventing formal prospective causal tracking.
- Dietary sodium and potassium excretion measurements were not available in the survey.
- Designed specifically for Ethiopian women; external validation in other Sub-Saharan African countries is required prior to regional adoption.

CONCLUSIONS

This study successfully derived and internally validated a non-laboratory multivariable clinical risk score for adult hypertension among non-pregnant Ethiopian women. Incorporating seven readily available demographic, anthropometric, socioeconomic and reproductive variables the 18 point score achieves excellent discrimination (AUROC: 0.818) and calibration.

This simple, point based tool provides primary healthcare workers and the Ethiopian Ministry of Health with an objective, low cost screening instrument to triage high risk women for diagnostic confirmation and early management providing a viable strategy to curb the rising cardiovascular disease burden in Ethiopia.

DECLARATIONS

Ethics Approval and Consent to Participate

The 2024/25 Ethiopian Demographic and Health Survey (EDHS) protocol was reviewed and approved by the Institutional Review Board of the Ethiopian Public Health Institute (EPHI) and the Institutional Review Board of ICF International. Written informed consent (or verbal consent for illiterate participants as approved by the IRB) was obtained from all participants prior to survey administration. For the present study, separate primary IRB ethical approval was waived by the Institutional Review Board of Teda Health Science College because this work represents a secondary analysis of fully anonymized de-identified national survey data. Formal authorization to access and analyze the de-identified EDHS dataset was granted by the DHS Program (User ID: 185291).

Consent for Publication

Not applicable.

Availability of Data and Materials

The dataset analyzed in this study is available from the Demographic and Health Survey (DHS) Program repository upon registration at: https://dhsprogram.com/data/dataset_admin/.

Competing Interests

The authors declare that they have no competing financial, personal, or professional interests.

Funding

This research received no external funding support.

Authors' Contributions

- **Ashebir Mamay Gebiru (AMG);** Conceived the study, managed data extraction, statistical modeling, machine learning execution, calibration, SHAP analysis, public health informatics integration design and drafted the original manuscript.

- **Abrham Maru Moges (AMA);** Contributed to study design, statistical review, public health informatics contextualization, manuscript revision and approved the final submitted version.
- **Serku Abate Mihret (SAM);** Contributed to study design, public health informatics contextualization, critical manuscript revision and approved the final submitted version.
- **Yenehun Demelash Gelaw (YDG);** Contributed to study design, data interpretation, manuscript revision and approved the final submitted version.
- **Yimer Mamaye (YM);** Contributed to study design, methodological validation, manuscript revision, and approved the final submitted version.

Acknowledgements

The authors thank the DHS Program and ICF International for providing access to the 2024/25 EDHS dataset and the EPHI and CSA for executing the national survey.

REFERENCES

1. Mills, K.T., A. Stefanescu, and J. He, *The global epidemiology of hypertension*. Nat Rev Nephrol, 2020. **16**(4): p. 223-237.
2. Kadkhodamanesh, A., et al., *Age, sex, income, sociodemographic and health system inequalities in the global burden of high systolic blood pressure: an analysis of the Global Burden of Disease Study 2021*. BMJ Open, 2025. **15**(8): p. e104107.
3. Kario, K., et al., *The WHO Global report 2023 on hypertension warning the emerging hypertension burden in globe and its treatment strategy*. Hypertens Res, 2024. **47**(5): p. 1099-1102.
4. *Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants*. Lancet, 2021. **398**(10304): p. 957-980.
5. Niu, C., et al., *The global burden of hypertension and its epidemiological impacts on adolescents and young adults: projections to 2050*. Front Cardiovasc Med, 2025. **12**: p. 1619445.
6. Seedat, Y., A. Ali, and K.C. Ferdinand, *Hypertension and cardiovascular disease in the sub-Saharan African context*. Ann Transl Med, 2018. **6**(15): p. 297.
7. Jobe, M., et al., *Hypertension in Sub-Saharan Africa: Burden, Barriers and Priorities for Improving Treatment Outcomes*. Circ Res, 2025. **137**(1): p. 106-118.
8. Kakoma, P.K., et al., *Prevalence and associated factors of hypertension among adults in Sub-Saharan Africa: a systematic review and meta-analysis of community-based studies*. BMC Cardiovasc Disord, 2026. **26**(1).
9. Kibret, K.T. and Y.M. Mesfin, *Prevalence of hypertension in Ethiopia: a systematic meta-analysis*. Public Health Rev, 2015. **36**: p. 14.
10. Tesfa, E. and D. Demeke, *Prevalence of and risk factors for hypertension in Ethiopia: A systematic review and meta-analysis*. Health Sci Rep, 2021. **4**(3): p. e372.
11. Paulose, T., Z.Z. Nkosi, and M. Endriyas, *Prevalence of hypertension and its associated factors in Hawassa city administration, Southern Ethiopia: Community based cross-sectional study*. PLoS One, 2022. **17**(3): p. e0264679.

12. Addo, J., L. Smeeth, and D.A. Leon, *Hypertension in sub-saharan Africa: a systematic review*. Hypertension, 2007. **50**(6): p. 1012-8.
13. Isaac Derick, K. and Z. Khan, *Prevalence, Awareness, Treatment, Control of Hypertension, and Availability of Hypertension Services for Patients Living With Human Immunodeficiency Virus (HIV) in Sub-Saharan Africa (SSA): A Systematic Review and Meta-analysis*. Cureus, 2023. **15**(4): p. e37422.
14. Wu, X., et al., *Global burden of hypertensive heart disease and attributable risk factors, 1990-2021: insights from the global burden of disease study 2021*. Sci Rep, 2025. **15**(1): p. 14594.
15. Mudie, K., et al., *Non-communicable diseases in sub-Saharan Africa: a scoping review of large cohort studies*. J Glob Health, 2019. **9**(2): p. 020409.
16. Mishra, S.R., et al., *Closing the Gap in Global Disparities in Hypertension Control*. Hypertension, 2025. **82**(3): p. 407-410.
17. Rezaei, F., et al., *Comparison of laboratory-based and non-laboratory-based WHO cardiovascular disease risk charts: a population-based study*. J Transl Med, 2022. **20**(1): p. 133.
18. Mphekgwana, P.M., et al., *Hypertension Prevalence and Determinants among Black South African Adults in Semi-Urban and Rural Areas*. Int J Environ Res Public Health, 2020. **17**(20).
19. Hassen, H.Y., et al., *Disparities in the non-laboratory INTERHEART risk score and its components in selected countries of Europe and sub-Saharan Africa: analysis from the SPICES multi-country project*. Eur Heart J Open, 2023. **3**(6): p. oead131.
20. Berhe, A.K., et al., *Prevalence of hypertensive disorders of pregnancy in Ethiopia: a systemic review and meta-analysis*. BMC Pregnancy Childbirth, 2018. **18**(1): p. 34.
21. Andriani, H., et al., *Hypertension in women of reproductive age: a cross-sectional analysis of prevalence and risk factors across 21 low-income and middle-income countries using Demographic and Health Surveys (2013-2023)*. BMJ Open, 2026. **16**(4): p. e106230.
22. Riley, R.D., et al., *Evaluation of clinical prediction models (part 2): how to undertake an external validation study*. Bmj, 2024. **384**: p. e074820.
23. Hippisley-Cox, J., et al., *Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study*. Bmj, 2007. **335**(7611): p. 136.

24. Collins, G.S., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement*. Ann Intern Med, 2015. **162**(1): p. 55-63.
25. Abrha, S., S. Shiferaw, and K.Y. Ahmed, *Overweight and obesity and its socio-demographic correlates among urban Ethiopian women: evidence from the 2011 EDHS*. BMC Public Health, 2016. **16**: p. 636.
26. Ahn, J.H., et al., *Meningeal lymphatic vessels at the skull base drain cerebrospinal fluid*. Nature, 2019. **572**(7767): p. 62-66.
27. Chobanian, A.V., et al., *The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report*. Jama, 2003. **289**(19): p. 2560-72.
28. Sullivan, L.M., J.M. Massaro, and R.B. D'Agostino, Sr., *Presentation of multivariate data for clinical use: The Framingham Study risk score functions*. Stat Med, 2004. **23**(10): p. 1631-60.
29. Gudina, E.K., Y. Michael, and S. Assegid, *Prevalence of hypertension and its risk factors in southwest Ethiopia: a hospital-based cross-sectional survey*. Integr Blood Press Control, 2013. **6**: p. 111-7.
30. Zhao, H., et al., *Predicting the Risk of Hypertension Based on Several Easy-to-Collect Risk Factors: A Machine Learning Method*. Frontiers in Public Health, 2021. **Volume 9 - 2021**.
31. Rezaei, F., et al., *Comparison of laboratory-based and non-laboratory-based WHO cardiovascular disease risk charts: a population-based study*. Journal of Translational Medicine, 2022. **20**(1): p. 133.
32. Pandya, A., M.C. Weinstein, and T.A. Gaziano, *A comparative assessment of non-laboratory-based versus commonly used laboratory-based cardiovascular disease risk scores in the NHANES III population*. PLoS One, 2011. **6**(5): p. e20416.
33. Mamgai, A., et al., *Cardiovascular risk assessment using non-laboratory based WHO CVD risk prediction chart with respect to hypertension status among older Indian adults: insights from nationally representative survey*. Frontiers in Public Health, 2024. **Volume 12 - 2024**.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TRIPODChecklistHypertensionRiskScore.docx](#)