American Journal of Preventive Cardiology 26 (2026) 101511

Contents lists available at ScienceDirect

American Journal of Preventive Cardiology

journal homepage: www.journals.elsevier.com/american-journal-of-preventive-cardiology

Short Report

A machine learning-derived cardiovascular risk score in people with HIV: the ML-ICONA score

Ovidio De Filippo [a,1], Francesco Codicè [b,1], Alessandro Tavelli [c,d,1], Tiziana Sanavia [b], Flavio Sartori [b], Andrea Bosio [e], Martina Ranzenigo [f,g], Alessandra Guida [h], Andrea Gori [i,j], Lucia Taramasso [k], Nicola Squillace [l], Pierre Meynet [a,m], Benedetta De Guidi [a,m], Alessandra Vergori [n], Giovanni Guaraldi [o,p], Piero Fariselli [b], Gaetano Maria De Ferrari [a,m], Fabrizio D'Ascenzo [a,m,2,*], Andrea Calcagno [e,q,2]

[a] Division of Cardiology, Cardiovascular and Thoracic Department, "Citta della Salute e della Scienza" Hospital, Turin, Italy
[b] Computational Biomedicine Unit, Department of Medical Sciences, University of Torino, Turin, Italy
[c] Icona Foundation, Milan, Italy
[d] National PhD Programme in One Health approaches to infectious diseases and life science research, Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Pavia, 27100, Italy
[e] Unit of Infectious Diseases, Department of Medical Sciences, University of Turin, Turin, Italy
[f] San Raffaele Vita-Salute University, Milan, Italy
[g] IRCCS San Raffaele Scientific Institute, Department of Infectious Diseases, Milan, Italy
[h] Infectious Diseases and Gender Medicine Unit, Cotugno Hospital, AORN dei Colli, Naples, Italy
[i] Infectious Diseases and Immunopathology, Department of Clinical Sciences, University of Milan, Luigi Sacco Hospital, Milan, Italy
[j] III Division of Infectious Diseases, ASST Fatebenefratelli Sacco, Luigi Sacco Hospital, Milan, Italy
[k] Infectious Disease Clinic, IRCCS Ospedale Policlinico San Martino, Genova, Italy
[l] Clinic of Infectious Disease, Fondazione IRCCS San Gerardo dei Tintori, Monza, Italy
[m] Department of Medical Sciences, University of Turin, Turin, Italy
[n] Viral Immunodeficiencies Unit, National Institute for Infectious Diseases Lazzaro Spallanzani IRCCS, Rome, Italy
[o] Department of Infectious Diseases, Azienda Ospedaliero-Universitaria di Modena, Policlinico di Modena, Modena, Italy
[p] Department of Surgical, Medical, Dental, and Morphological Sciences University of Modena and Reggio Emilia, Modena, Italy
[q] Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy

ARTICLE INFO

Keywords:
Human immunodeficiency virus
Cardiovascular prevention
Risk score
Machinelearning
Antiretrovirals

1. Introduction

By the end of 2022, approximately 39 million people worldwide were living with Human Immunodeficiency Virus (HIV). Widespread adoption of combination active antiretroviral therapy (cART) has substantially prolonged the lifespan of people living with HIV (PWH). Nevertheless, morbidity and mortality remain higher than in the general population, largely driven by cardiovascular (CV) disease [1].

Epidemiological studies report a 1.5– to 2–fold higher CV risk in PWH, even among those with full viral suppression, reflecting a complex interplay of poor control of CV risk factors, immune dysregulation, inflammation, metabolic derangements, and exposure to antiretroviral

---

* Corresponding author at: Division of Cardiology, University of Turin, Corso Bramante n 88.
E-mail address: fabrizio.dascenzo@gmail.com (F. D'Ascenzo).

[1] These authors equally contributed to this work and share first authorship

[2] joint last authorship

https://doi.org/10.1016/j.ajpc.2026.101511
Received 24 July 2025; Received in revised form 24 February 2026; Accepted 25 February 2026

drugs. Nevertheless, the association between integrase strand transfer inhibitors (INSTIs) and CV events remains a matter of ongoing and substantial debate [1,2]. European Society of Cardiology (ESC) guidelines recommend the use of SCORE2 [3] for risk estimation in PWH; however, this tends to underestimate their CV risk. HIV-specific models (e.g. D:A:D score) partially address this gap but still show limited discrimination and calibration across cohorts [4].

Machine learning (ML) approaches, allowing for integration of multiple variables, may improve risk prediction, particularly in populations with low event rates. In this context, we developed and validated a ML-based CV risk prediction model for PWH, named ICONA-ML, derived from the Italian Cohort Naïve Antiretrovirals (ICONA) Foundation Study [5]. Model performance in predicting the risk of long-term major cardiovascular events was compared with ESC SCORE-2.

2. Methods

The ICONA Foundation Study is a multicentre, prospective observational cohort enrolling PWH across Italy since 2008. Epidemiological, clinical, laboratory, and treatment-related data were systematically collected in an electronic case report form (www.icona.org). All participants provided written informed consent, in accordance with the ethical standards of the committee on human experimentation and the Helsinki Declaration (last amendment October 2013).

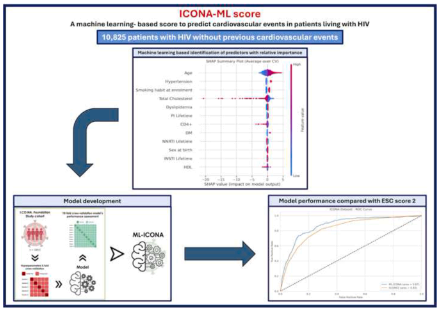
The derivation cohort included all participants without previous CV events and with at least 10 years of follow-up. Baseline features included age, sex at birth, smoking status, hypertension, diabetes mellitus, dyslipidaemia, blood pressure, lipid profile, baseline CD4+ cell count at ART initiation, HIV RNA levels at ART initiation, and cumulative exposure to cART drugs. The primary endpoint was a composite of mutually exclusive events, including CV death, myocardial infarction, unplanned coronary revascularization, or stroke, assessed over 10 years.

Categorical variables are reported as counts and percentages, and continuous variables as mean ± standard deviation or median and interquartile range (IQR). Differences between continuous variables were assessed using parametric tests (T-test) or non-parametric tests (Mann-Whitney U test), depending on data distribution. The chi-square test was employed to evaluate differences between categorical variables.

The ICONA-ML model was developed using a regularized logistic regression algorithm implemented using Python 3.10 programming language and the Scikit-learn library. L1–L2 regularization was applied to mitigate overfitting and improve model generalizability. Associated hyperparameters were fine-tuned through a 5-fold cross-validation within the training set. The a priori candidate feature set (see Supplementary Table 3) included traditional CV predictors (e.g. age, sex, smoking, diabetes, BMI, systolic blood pressure/antihypertensive treatment, lipid measures/lipid-lowering treatment, family history), routine laboratory measures (e.g. renal function), and HIV-related factors (e.g. current/nadir CD4, HIV RNA and cumulative exposure to major cART drug classes); all candidates were entered into the modelling pipeline and retained only if they provided incremental predictive value, with non-contributory variables excluded when they did not improve discrimination and overall performance during cross-validation. Overall performance was evaluated using ten-fold cross-validation across the full ICONA dataset. In each iteration, nine folds were used for training and one for testing, with performance metrics averaged across folds (**see** Central Illustration). Model performance was compared with ESC SCORE2 applied to the same cohort. Sensitivity, specificity, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), and Area Under the Precision-Recall Curve (AUCPRC) were calculated for both models.

3. Results

In the derivation cohort of 10,825 participants, 193 (1.7 %) experienced the primary endpoint at 10 years follow-up (**Supplementary Table 1**). Participants with events were older (mean age 52 ± 11 vs. 37 ±12 years, $p < 0.001$), more frequently smokers (55 % vs. 42 %, $p < 0.001$), and had a higher prevalence of hypertension, diabetes mellitus, dyslipidaemia, and family history of CVD. They also showed lower baseline CD4+ cell counts (292±259 vs. 354±268.7 cells/mm^3, $p = 0.002$). Exposure to protease inhibitors (PIs) and INSTIs was more common among participants with events (54 % and 75.6 %, vs. 40.5 % and 64 %, respectively, both $p < 0.001$), although the duration of exposure did not differ significantly (**Supplementary Table 2**).



Central Illustration. ICONA-ML score. *Top panel*: SHapley Additive exPlanations (SHAP) Summary Plot: Feature Contributions to Cardiovascular Risk Prediction. *Bottom left panel*: process of development of ICONA-ML model. The model was trained using L1-L2 regularization to minimize overfitting and enhance generalizability. Associated hyperparameters were fine-tuned through a 5-fold cross-validation within the training set. The model's performance was assessed using 10-fold cross-validation within the main ICONA dataset. In each iteration, nine folds were used for training and one for testing, with metrics calculated by averaging the results across all folds. *Bottom right panel*: Receiver Operating Characteristic (ROC) curve Comparison: ICONA-ML Score vs European Society of Cardiology (ESC) Score 2. *Legend*. PI: Protease Inhibitors; DM: diabetes mellitus; NNRTI: Non-nucleoside reverse transcriptase inhibitors; INSTI: integrase Strand Transfer Inhibitor; HDL: high density lipoprotein. ROC: Receiver Operating Characteristic.

In the ML analysis, age, sex at birth, smoking status at enrolment, hypertension, dyslipidaemia (along with total cholesterol and low-density lipoprotein), and diabetes mellitus showed the strongest association with the primary endpoint. By contrast, CD4+ count and cART drugs were not among the most relevant features (**see** Central Illustration). Mean Absolute SHapley Additive exPlanations (SHAP) values of relevant variables are reported in the Central Illustration**.**

In the derivation cohort, the ICONA-ML model achieved an AUC of 0.87 ± 0.04, with high sensitivity (0.94 ± 0.04) and moderate specificity (0.62 ± 0.03). In comparison, SCORE2 yielded an AUC of 0.83 ± 0.02, with a slightly lower sensitivity (0.92 ± 0.05) and a lower specificity (0.53 ± 0.03) than ICONA-ML. ICONA-ML also demonstrated higher precision (0.04 ± 0.00 vs. 0.03 ± 0.00) and a superior AUCPRC (0.13 ± 0.04 vs. 0.11 ± 0.04) (**Supplementary Table 4**).

Participants were ranked by predicted 10-year cardiovascular risk and stratified into deciles to examine risk gradients and calibration across the full prediction spectrum. Observed 10-year event rates increased monotonically across deciles, with the greatest separation in the highest two deciles (observed 10-year event rate 9.6 %). For clinical interpretability, adjacent deciles were collapsed into three strata: low (deciles 1–4), moderate (deciles 5–8), and high (deciles 9–10), and observed and predicted event rates were compared across strata (**Supplementary Figures 1–2**).

4. Discussion

To the best of our knowledge, this is the first study to apply a ML–based model to long-term cardiovascular risk prediction for primary prevention in PWH. ICONA-ML demonstrated high discrimination and sensitivity, with improved precision and AUCPRC compared with ESC SCORE2, underscoring the superiority of ML-based approaches in handling class imbalances [6].

Although overall accuracy was comparable between ICONA-ML and SCORE2 [3], the latter may underestimate CV risk in PWH due to the absence of HIV-related variables. Notably, even if both traditional CV and HIV-specific risk factors were included in our model, conventional CV risk factors contributed most strongly to risk prediction, while CD4+ cell count and antiretroviral drug exposure had a more limited impact. This suggests that chronic metabolic and lifestyle-related conditions might outweigh immunological and virological parameters in long-term cardiovascular risk estimation. Despite its strong association with CV risk [7], the relatively limited contribution of CD4+ cell count may reflect its ability to capture inflammatory status only within a limited temporal window. With respect to cART, it should be acknowledged that certain drugs may be prescribed preferentially to individuals with specific characteristics, such as renal dysfunction or resistance-associated mutations, which could indirectly influence CV risk [2].

From a clinical standpoint, ICONA-ML high sensitivity could be particularly useful for early identification of high-risk participants who could benefit from intensive primary prevention strategies. Furthermore, ICONA-ML can guide clinicians in categorizing participants into risk levels (low, moderate, high), tailoring management plans.

5. Limitations

This study has several limitations. The evolving nature of antiretroviral therapy and management may not be fully captured, especially considering the retrospective nature of the study. The relatively low number of CV events may have led to some degree of overfitting, despite cross-validation. Some immune-history metrics (e.g. CD4/CD8 ratio, CD4 nadir and cumulative viremia) were unavailable, not reliably ascertainable, or not retained after cross-validated feature selection; thus, their potential incremental contribution to prediction could not be fully evaluated. The decile-based stratification was used to describe calibration and clinically interpretable risk gradients; the resulting categories should not be interpreted as treatment thresholds and would require external validation and clinical-utility assessment before being used for decision-making. Furthermore, the study cohort consisted predominantly of Caucasian individuals of Italian origin, potentially limiting generalizability. Finally, emerging biomarkers and imaging markers of subclinical atherosclerosis were not included and may further enhance risk prediction in future iterations of the model.

6. Conclusion

ICONA-ML is a novel ML-based tool for CV risk prediction in PWH for primary prevention. By providing high sensitivity and clinically actionable risk stratification, it may support earlier and more personalized preventive interventions. External validation in diverse populations is warranted before widespread clinical implementation.

Disclosures

Conflict of interest: none declared.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The Icona Foundation is supported by unrestricted grants from Gilead Sciences, ViiV Healthcare, Merck Sharp & Dohme, and Janssen-Cilag. The funders of the ICONA Foundation had no role in the study design, data collection, analysis, decision to publish, or preparation of this study.

Data availability statement

The datasets generated during the current study are not publicly available because they contain sensitive data to be treated under data protection laws and regulations. Appropriate agreement of data sharing can be arranged after a reasonable request to the corresponding author.

CRediT authorship contribution statement

Ovidio De Filippo: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Data curation, Conceptualization. **Francesco Codicè:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation. **Alessandro Tavelli:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Data curation. **Tiziana Sanavia:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Formal analysis, Data curation. **Flavio Sartori:** Writing – review & editing, Supervision. **Andrea Bosio:** Writing – review & editing, Investigation, Data curation. **Martina Ranzenigo:** Writing – review & editing, Investigation, Data curation. **Alessandra Guida:** Writing – review & editing, Investigation, Data curation. **Andrea Gori:** Writing – review & editing, Validation, Supervision. **Lucia Taramasso:** Writing – review & editing, Validation, Supervision. **Nicola Squillace:** Writing – review & editing, Validation, Supervision. **Pierre Meynet:** Writing – review & editing, Writing – original draft, Investigation, Data curation. **Benedetta De Guidi:** Writing – review & editing, Writing – original draft, Investigation, Data curation. **Alessandra Vergori:** Writing – review & editing. **Giovanni Guaraldi:** Writing – review & editing, Validation, Supervision. **Piero Fariselli:** Writing – review & editing, Software, Formal analysis. **Gaetano Maria De Ferrari:** Writing – review & editing, Validation, Supervision. **Fabrizio D'Ascenzo:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Andrea Calcagno:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Icona Foundation Study Group

BOARD OF DIRECTORS: A d'Arminio Monforte (President), A Antinori (Vice-President), S Antinori, A Castagna, R Cauda, G Di Perri, E Girardi, R Iardino, A Lazzarin, GC Marchetti, C Mussini, E Quiros-Roldan, L Sarmati, B Suligoi, F von Schloesser, P Viale.

SCIENTIFIC SECRETARY: A d'Arminio Monforte, A Antinori, A Castagna, F Ceccherini-Silberstein, A Cingolani, A Cozzi-Lepri, A Di Biagio, E Girardi, A Gori, S Lo Caputo, G Marchetti, F Maggiolo, C Mussini, M Puoti, CF Perno, C Torti.

STEERING COMMITTEE: A Antinori, A Bandera, S Bonora, A Calcagno, D Canetti, A Castagna, F Ceccherini-Silberstein, A Cervo, A Cingolani, P Cinque, A Cozzi-Lepri, A d'Arminio Monforte, A Di Biagio, R Gagliardini, A Giacomelli, E Girardi, N Gianotti, A Gori, G Guaraldi, S Lanini, G Lapadula, M Lichtner, A Lai, S Lo Caputo, G Madeddu, F Maggiolo, V Malagnino, G Marchetti, A Mondi, V Mazzotta, C Mussini, S Nozza, CF Perno, S Piconi, C Pinnetti, M Puoti, E Quiros Roldan, R Rossotti, S Rusconi, MM Santoro, A Saracino, L Sarmati, V Spagnuolo, N Squillace, V Svicher, L Taramasso, C Torti, A Vergori.

STATISTICAL AND MONITORING TEAM: A Cozzi-Lepri, S De Benedittis, I Fanti, M Giotta, C Marelli, A Rodano', A Tavelli.

COMMUNITY ADVISORY BOARD: M Cernuschi, L Cosmaro, A Perziano, V Calvino, D Russo, M Farinella, N Policek, VL Del Negro.

BIOLOGICAL BANK INMI AND SAN PAOLO: M Augello, S Carrara, S

Graziano, G Prota, S Truffa, D Vincenti, R Rovito, M Sgarlata.

PARTICIPATING PHYSICIANS AND CENTERS: Italy A Giacometti, A Costantini, V Barocci (Ancona); A Saracino, C Santoro, E Milano (Bari); L Comi, C Suardi (Bergamo); P Viale, L Badia, S Cretella (Bologna); EM Erne, A Pieri (Bolzano); E Quiros Roldan, E Focà (Brescia); B Menzaghi, C Abeli (Busto Arsizio); L Chessa, F Pes (Cagliari); P Maggi, L Alessio (Caserta); G Nunnari, BM Celesia (Catania); J Vecchiet, K Falasca (Chieti); A Pan, S Dal Zoppo (Cremona); D Segala (Ferrara); F Bartalesi, A Bartoloni, B Borchi, C Costa (Firenze); S Lo Caputo, S Ferrara (Foggia); M Bassetti, E Pontali, S Blanchi, N Bobbio (Genova); C. Del Borgo, R. Marocco, G. Mancarella (Latina); S Piconi, C Molteni (Lecco); S Rusconi, G Canavesi (Legnano); G Pellicanò, Russotto Y (Messina); G Marchetti, S Antinori, A Gori, M Puoti, A Castagna, A Bandera, V Bono, MV Cossu, A Giacomelli, R Lolatto, MC Moioli, L Pezzati, S Diotallevi, C Tincati (Milano); C Mussini, M Menozzi (Modena); P Bonfanti, G Lapadula (Monza); V Sangiovanni, I Gentile, V Esposito, N Coppola, FM Fusco, G Di Filippo, V Rizzo, N Sangiovanni, S Martini (Napoli); AM Cattelan, D Leoni (Padova); A Cascio, M Trizzino (Palermo); D Francisci, E Schiaroli (Perugia); G Parruti, F Sozio (Pescara); D Messeri, SI Bonelli (Pistoia); C Lazzaretti, R Corsini (Reggio Emilia); A Antinori, R Cauda, C Mastroianni, L Sarmati, A Latini, A Cingolani, I Mastrorosa, S Lamonica, M Capozzi, M Camici, I Mezzaroma, M Rivano Capparuccia, G Iaiani, C Stingone, L Gianserra, J Paulicelli, MM Plazzi, G d'Ettore, M Fusto (Roma); I Coledan (Rovigo); G Madeddu, A De Vito (Sassari); M Fabbiani, F Montagnani (Siena); A Franco, R Fontana Del Vecchio (Siracusa); D Francisci, C Di Giuli (Terni); GC Orofino, G Calleri, G Di Perri, S Bonora, G Accardo (Torino); C Tascini, A Londero (Udine); G Battagin, S Nicolè (Vicenza); G Starnini, S Dell'Isola (Viterbo).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ajpc.2026.101511.

References

[1] Shah ASV, Stelzle D, Lee KK, et al. Global burden of atherosclerotic cardiovascular disease in people living with HIV. Circulation 2018;138(11):1100–12. https://doi.org/10.1161/CIRCULATIONAHA.117.033369.

[2] Neesgaard B, Greenberg L, Miró JM, et al. Associations between integrase strand-transfer inhibitors and cardiovascular disease in people living with HIV: a multicentre prospective study from the RESPOND cohort consortium. Lancet HIV 2022;9(7):e474–85. https://doi.org/10.1016/S2352-3018(22)00094-7.

[3] Hageman S, Pennells L, Ojeda F, et al. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J 2021;42(25):2439–54. https://doi.org/10.1093/eurheartj/ehab309.

[4] Yu J, Liu X, Zhu Z, et al. Prediction models for cardiovascular disease risk among people living with HIV: a systematic review and meta-analysis. Front Cardiovasc Med 2023;10. https://doi.org/10.3389/fcvm.2023.1138234.

[5] d'Arminio Monforte A, AC Lepri, Rezza G, et al. Insights into the reasons for discontinuation of the first highly active antiretroviral therapy ART regimen in a cohort of antiretroviral naïve participants. AIDS 2000;14(5):499–507. https://doi.org/10.1097/00002030-200003310-00005.

[6] Lüscher TF, Wenzl FA, D'Ascenzo F, Friedman PA, Antoniades C. Artificial intelligence in cardiovascular medicine: clinical applications. Eur Heart J 2024;45 (40):4291–304. https://doi.org/10.1093/eurheartj/ehae465.

[7] Lichtenstein KA, Armon C, Buchacz K, et al. Low CD4+ T cell count is a risk factor for cardiovascular disease events in the HIV outpatient study. Clin Infect Dis. 2010; 51(4):435–47. https://doi.org/10.1086/655144.