American Journal of Preventive Cardiology 22 (2025) 100964

Contents lists available at ScienceDirect

American Journal of Preventive Cardiology

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

Original Research

Composite socio-environmental risk score for cardiovascular assessment: An explainable machine learning approach

Zhuo Chen [a], Jean-Eudes Dazard [a], Pedro Rafael Vieira de Oliveira Salerno [a], Santosh Kumar Sirasapalli [a], Mohamed HE Makhlouf [a], Sanjay Rajagopalan [a,1,*], Sadeer Al-Kindi [b,1,*]

[a] Herman K Hellerstein Professor of Cardiovascular Research, Director, Cardiovascular Research Institute, Harrington Heart and Vascular Institute, University Hospitals, and School of Medicine, Case Western Reserve University, 11100 Euclid Ave, Cleveland, OH 44106, United States

[b] Associate Professor of Cardiology, Associate Director, Cardiovascular Prevention & Wellness, Center for CV Computational & Precision Health, Jerold B Katz Investigator, Academy of Translational Research, Houston Methodist DeBakey Heart & Vascular Center, Medical Director, Center for Health and Nature and Department of Cardiology, Houston Methodist, Houston, TX, United States

ARTICLE INFO

Keywords:
Socio-environmental risk factors
Major adverse cardiovascular events
Machine learning
Exposome

ABSTRACT

Background: Cardiovascular disease (CVD) is the leading global cause of death, with socio-environmental factors significantly influencing morbidity and mortality. Understanding these factors is essential for improving risk assessments and interventions.

Objective: To develop and evaluate the predictive power of a composite socio-environmental (SE) cardiovascular risk score in forecasting major adverse cardiovascular events (MACE) among patients, considering both traditional and novel socio-environmental risk factors.

Methods: A Survival Random Forest (RSF) model was used to create a composite socio-environmental (SE) cardiovascular risk score using 22 census-tract level variables from 62,438 patients in the CLARIFY registry undergoing coronary artery calcium (CAC) scoring. A Cox Proportional Hazard (CPH) model was then applied to assess the association between the SE-MACE risk score and MACE in a hold-out test set. SHapley Additive exPlanations (SHAP) values were used to identify variable importance.

Results: The study included 62,438 individuals (mean age 59.6 years, 53.2 % female, 87.7 % White). Hypertension (55.4 %), diabetes (15.7 %), and dyslipidemia (72.3 %) were common, with a median CAC score of 168. The RSF model showed a concordance index of 0.58, with significant factors including smoking prevalence, insurance status, and median household income impacting cardiovascular risk. The SE-MACE risk score was robustly associated with MACE (HR, 1.21 [95 % CI, 1.11-1.32]), independent of clinical variables and the CAC score. Kaplan-Meier analysis highlighted clear risk stratification across SE-MACE score quartiles.

Conclusion: The SE-MACE risk score effectively incorporates socio-environmental factors into cardiovascular risk assessment, identifying individuals at higher risk for MACE and supporting the need for holistic assessment frameworks. Further validation in diverse settings is recommended to confirm these findings.

1. Introduction

Social Determinants of Health (SDOH), encompassing economic, social, environmental, and psychosocial factors, are now recognized as critical drivers of cardiovascular disease (CVD) risk, morbidity, and mortality [1,2]. These conditions in which individuals are born, grow, live, work, and age, significantly impact cardiovascular health, often acting as 'upstream factors' that influence both health behaviors and traditional CVD risk factors [3]. Recent reviews and studies have consistently demonstrated the significant role of SDOH in CVD [4,5]. For instance, an umbrella review of 70 systematic reviews highlighted that factors related to economic circumstances and early childhood development are consistently associated with increased CVD risk and mortality [4]. This review also found evidence linking social/community

* Corresponding authors.
E-mail addresses: sanjay.rajagopalan@uhhospitals.org (S. Rajagopalan), sal-kindi@houstonmethodist.org (S. Al-Kindi).
[1] Authors contributed equally as co-senior authors.

https://doi.org/10.1016/j.ajpc.2025.100964
Received 1 October 2024; Received in revised form 9 March 2025; Accepted 15 March 2025
Available online 17 March 2025

context and neighborhood/built environment themes to CVD. These findings underscore the urgent need to address SDOH to reduce the global burden of CVD and achieve health equity.

While epidemiologic and mechanistic investigation have provided compelling links between environmental factors and SDOH and cardiovascular risk, most studies have incorporated small number of environmental risk factors often exclusively [6,7]. We and others have shown significant intersection and interaction between social, natural and built environmental factors and health risk [8–10]. For example, at the county level, social deprivation increases the risk of air pollution cardiovascular mortality [9]. Similarly, multiple aspects in the built environment exaggerate exposure to pollutants, the health risks of which are disproportionately felt by marginalized communities [11]. There remains a significant gap in our understanding of the complex interactions of various risk factors at the neighborhood level within the broader context of how socio-environmental risk factors exert their effects. As such, there is a compelling need for detailed understanding of the socio-environmental exposome (a comprehensive measure of all the socio-environmental exposures of an individual in a lifetime) and its effects on cardiovascular events, particularly in at-risk individuals. Coronary calcium score (CAC) is widely believed to be one of the best predictors of future cardiovascular events [12]. The presence of CAC helps stratify asymptomatic individuals into groups at higher risk for major adverse cardiovascular events. There have been very few studies that have examined the impact of baseline CAC and the impact of additional social and environmental factors on adverse cardiovascular events.

We, therefore, sought to investigate the neighborhood's impact on cardiovascular risks, and develop a composite socio-environmental cardiovascular risk score from diverse variables, and evaluated its association with major adverse cardiovascular events.

2. Methods

We employed a Survival Random Forest (RSF) model to build a composite socio-environmental (SE) cardiovascular risk score using 22 census-tract level environmental and socioeconomic features (obtained from various sources) with 62,438 patients undergoing coronary artery calcium scoring in CLARIFY registry (clinicaltrials.gov identifier NCT04075162) [13]. Subsequently, we used a Cox Proportional Hazard (CPH) model to identify the association between the risk score and MACE in a hold-out test set.

Cohort Description: The study cohort was drawn from University Hospitals Health System (UHHS) in northeast Ohio, one of the region's largest healthcare networks, including 18 large hospitals and over 31 Health Centers. Recognizing the lack of Medicare reimbursement for CAC testing in Ohio, UHHS initiated a low-cost CAC screening program in 2014, reducing the test's cost to $99 [6,13]. The no-cost phase of this program was fully implemented in January 2017 after a successful pilot phase. In this program, CAC screening is offered to men aged 45 or older and women 55 or older, without a history of cardiovascular disease but with one or more heart disease risk factors, such as dyslipidemia, hypertension, smoking, diabetes, and family history of coronary artery disease. It was also available to men and women over 40 with chronic inflammatory conditions. The CAC testing was conducted at 21 locations across northeast Ohio. For our study, we included all patients who participated in the UHHS CAC program from January 2014 to November 2022. Data were extracted from electronic medical records and maintained in the CLARIFY registry (ClinicalTrials.gov NCT04075162). Outcomes, including cardiovascular events, were identified via ICD-10 codes and linkage with the Ohio death index. Patients lost to follow-up were censored at their last recorded visit. Patient consent waived for registry inclusion by the UHHS institutional review board.

CAC scores were determined using Multi-Detector CT (MDCT) scanners, adhering to standardized protocols recommended by the Society for Cardiovascular Computed Tomography (SCCT) [14]. Scans from various locations were centralized for quantification using specialized software (Heartbeat-CS, EBW, Philips Medical Systems, Cleveland, OH). The CAC results were shared with patients and referring providers through electronic medical records and an online patient portal.

Outcomes: The primary outcomes of this study encompassed a composite of major adverse cardiovascular events (MACE), defined as the first occurrence of myocardial infarction, stroke, coronary revascularizations, heart failure admissions, and all-cause mortality following CAC screening. These events were identified using international classification of diseases codes and data linkage with the Ohio death index [15]. Additional data on cardiovascular comorbidities and self-reported race/ethnicity were extracted from the electronic medical records. Patients lost to follow-up were censored at their last recorded hospital encounter date in the system.

Social and Environmental Factors Data: A total 22 social and environmental exposures at the census tract level were compiled from a variety of sources (Supplemental Table 1 and Supplemental Fig. 1). From the CDC's PLACES Dataset [16] we retrieved four variables: the proportion of individuals without health insurance, smoking prevalence, binge drinking prevalence, and the proportion lacking sufficient sleep. The Environmental Protection Agency's Smart Location Database (SLD) contributed eight variables, [17] including the percentage of working-age population, percentage of households without a car, and various ranks and indices related to employment, walkability, and transit access. Food access was assessed using two metrics from the U.S. Department of Agriculture's Food Access Research Atlas Data [18] the percentage of low-income individuals living more than half a mile from a supermarket and the percentage of households without a vehicle in similar conditions. Additionally, three socio-economic and demographic factors were sourced from the American Community Survey 5-year estimates (2019): [19] median household income, the percentage of individuals with less than a high school education, and unemployment rates. The Social Vulnerability Index from the CDC/ATSDR [20], which assesses community resilience to disasters, was also included. Finally, we incorporated four types of raster environmental data (detailed in Supplemental Table 1), including satellite-derived $PM_{2.5}$ levels, [21] night light pollution, [22] noise exposure, [23] and the Normalized Difference Vegetation Index (NDVI), for greenness [24]. These data were integrated with census tract boundaries from the Census Bureau, and exposures were estimated for each tract using zonal statistics in QGIS V 3.16.

2.1. Analytic approach

Our analytical approach comprises two main steps (Central Illustration):

(1) Generation of the SE-MACE Risk Score: We employ a Random Survival Forest (RSF) algorithm [25] to understand the relationship between socio-environmental features and MACE outcomes. Initially, the data is split into an 80 % training set and a 20 % test set. To optimize the RSF model parameters, we use a 5-fold cross-validation method with a grid search approach, targeting the lowest prediction error, which is measured by 1 minus Harrell's concordance index [26] (C-index). The C-index, ranging from 0 to 1, evaluates the model's ability to correctly rank pairs of individuals in terms of survival, accounting for censoring. The chosen model is then applied to the unseen test data to generate the SE-MACE risk scores, which are standardized z-scores with a mean of 0 and a standard deviation of 1. For interpretability, we augmented the RSF model with SHapley Additive exPlanations (SHAP), [27] which provide both local and global interpretations of the model's predictions.
(2) Assessment of the SE-MACE Risk Score Effectiveness: We evaluate the hazard ratio of the SE-MACE risk scores using a CPH model.

This assessment is performed on the test dataset to avoid biases from the training data. We compare three models: (a) one using only the SE-MACE risk score, (b) another adding individual clinical variables (age, sex, race/ethnicity, hypertension, diabetes, and dyslipidemia), and (c) a third incorporating all aforementioned variables plus the Coronary Artery Calcium (CAC) score. Two-sided $p < 0.05$ was considered statistically significant. Additionally, we divided the SE-MACE risk scores into quartiles to categorize patients into four distinct groups. To visualize the differences in MACE risks over time among these groups, we plotted cumulative MACE risks for each quartile. The significance of the differences in MACE risks between these groups was assessed using the log-rank test.

For these analyses, R version 4.2.3 is used for the CPH model, and Python 3.10.3 for the RSF model and SHAP analysis.

3. Results

A total of 62,438 individuals were identified and geocoded. Detailed baseline characteristics of the cohort are presented in Table 1. During the median follow-up period of 26.86 months, there were 2604 incidents of MACE (4.2 % of the cohort). The participants had an average age of 59.6 years, and the majority (87.7 %) identified as White. Females represented 53.2 % of the cohort. Additionally, 55.4 % of patients were diagnosed with hypertension, 15.7 % with diabetes, and 72.3 % had dyslipidemia. Mean value of CAC score was 168 (0-4,993).

3.1. Feature importance from SHAP

The RSF model, cross-validated on the training set and incorporating 22 socio-environmental variables, attained a concordance index (C-index) of 0.58. This model configuration was optimized with a forest of 1000 trees, using approximately the square root of the number of variables (~5) for potential splits at each node, and a minimum terminal node size of 760. We ranked the 22 SE variables in predicting the MACE in test set, using mean absolute SHAP values. Fig. 1 provides an analysis of the impact of SE features on the SE-MACE risk score, using mean absolute SHAP values within the RSF model. Correlation analysis (Pearson's r) between the SHAP values and corresponding feature values are performed to examine the directionality of the features [28]. Red bars indicate a positive correlation with the SE-MACE risk score, signifying that as these features increase, so does the MACE risk score. In contrast, blue bars represent a negative correlation, indicating an inverse relationship with the risk score. The intensity of the color saturation along with the bar length allows for the visualization of both the direction and magnitude of each SE feature's impact on the risk score, providing insights into the diverse factors that contribute to cardiovascular risk in the population studied.

Among the features, 'Percentage of Smokers' shows the most considerable positive correlation with the risk score, depicted by the length and red color of its bar. This is followed by 'Percentage of No Insurance', which also positively correlates with an increase in the risk score. 'Median Household Income' displays a significant negative correlation, as evidenced by its blue bar, suggesting that higher income levels are associated with a lower SE-MACE risk score. Other notable features include 'Percentage of Sleep Less Than 7 h' and 'Percentage Less than High School Education', both positively correlated with increased risk, alongside 'Social Vulnerability Index' (SVI), particulate matter (PM2.5) and 'Light Pollution', although to a lesser degree than smoking prevalence and insurance status. Negative correlations are also observed with built environmental factors such as 'National Walkability Index', Employment Mix' and 'Proximity to Transit Stops', all of which are indicated by blue bars, suggesting these factors are associated with a decrease in the SE-MACE risk score. Interestingly, while the 'Percentage of Binge Drinking' is associated with lower risk scores, its feature importance was low as indicated by its low mean absolute SHAP value. Additionally, certain variables, including 'NDVI', 'Low Income and One-Half Mile from Supermarket', and 'Noise', were found to have no significant association with the risk score.

Table 1
Descriptive statistics of the patients from the CLARIFY cohort.

Characteristic	No. (%) ($N = 62,438$)
Age, mean (SD), y	59.62 (9.71)
Gender	
Female	33226 (53.2)
Male	29205 (46.8)
Unknown	7 (0.0)
Race/ethnicity	
White	54777 (87.7)
Black	5450 (8.7)
Other/Unknown	2211 (3.5)
CAC Score (SD)	168 (420)
Hypertension	34578 (55.4)
Diabetes	9814 (15.7)
Dyslipidemia	45146 (72.3)
MACE_DAYS (SD)	949 (689)
MACE	2604 (4.2)

*Characteristic with SD is a continuous variable, and reported as median (standard deviation)

3.2. Association between SE-MACE risk score and outcomes

Table 2 details the outcomes of the CPH models assessing the composite SE-MACE risk score's impact on MACE. In Model 1 (unadjusted), the SE-MACE risk score was associated with a hazard ratio (HR) of 1.28 (95 % CI 1.19-1.39, $p < 0.001$). Upon adjustment for age, race, sex, hypertension, diabetes, and dyslipidemia in Model 2, the risk score maintained significantly associated with MACE (HR 1.22, 95 % CI 1.12-1.33, $p < 0.001$). Further adjustment in Model 3, which included the CAC score, yielded continued strong association (HR 1.21, 95 % CI 1.11-1.32, $p < 0.001$). Notably, in both Model 2 and Model 3, race did not exhibit a significant effect on MACE outcomes (Model 3: HR for Black 0.86, 95 % CI 0.50-1.50, $p = 0.59$). Additionally, the significance of sex as a variable diminished in Model 3 after including the CAC score, transitioning from a significant factor in Model 2 (HR for male 1.46, 95 % CI 1.23-1.75, $p < 0.001$) to a non-significant one in Model 3 (HR for male 1.15, 95 % CI 0.96-1.39, $p = 0.13$).

The Kaplan-Meier curves displayed in the Fig. 2, illustrates the cumulative risk of MACE over a 2500-day period among individuals within four quartiles of the SE-MACE risk score in the hold-out test set. The curves clearly stratify risk, with the highest quartile (Q4) showing a more rapid increase in cumulative risk, distinctly separating from the lowest risk quartile (Q1). Notably, the dashed lines highlight the time at which each quartile reaches a 5 % cumulative risk of MACE, demonstrating a significant separation between quartiles. At the 1500-day follow-up mark, the incidence of MACE in the highest risk quartile (Q4) was nearly twice that observed in the lowest quartile (Q1), with 160 events compared to 82 events. The statistical significance of the differences in MACE risk across the quartiles is confirmed by a log rank test with a p-value of less than 0.001.

4. Discussion

The main finding of this study is the derivation of a novel socio-environmental (SE) composite risk score from 22 census-tract level variables which exhibits a consistent and significant association with major adverse cardiovascular events (MACE) after adjusting for individual level baseline calcium score. The SE-MACE risk score generated using a machine learning approach was found to be significantly associated with MACE risk, after adjusting for common clinical variables and CAC score in our hold-out test data with the ability to further delineate

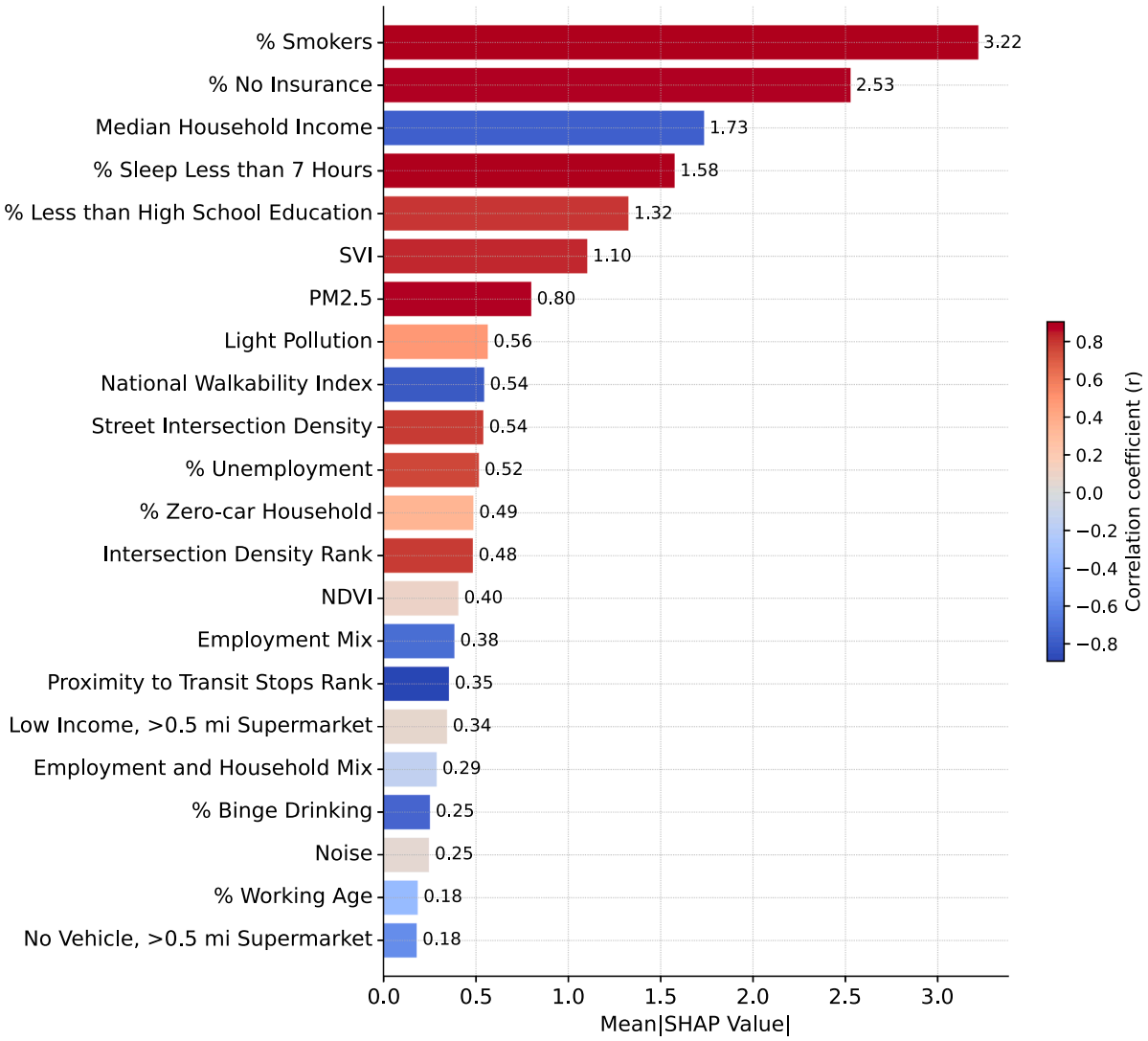


Fig. 1. Relative importance of the 22 features from the Random Survival Forest (RSF) model using SHAP values, illustrating: (i) the direction of correlation between each feature and its SHAP values, indicated by color (red for positive, blue for negative); (ii) the intensity of this correlation, denoted by the depth of color saturation; and (iii) the average absolute impact of each feature, represented by bar length.

and stratify risk across quartiles of the SE-MACE score. A particularly compelling finding was the nearly twofold increase in MACE incidence in the highest risk quartile (Q4) compared to the lowest (Q1) over the follow-up period, providing a robust ability of SE risk to identify at-risk individuals. The lack of significant association between race and MACE outcomes in the adjusted models, and the diminishing significance of sex after including CAC scores, are intriguing findings that highlight the fact that socio-environmental disparities and racism, rather than race, are the drivers of MACE risk. Our data thus raise the possibility that underlying socio-environmental factors may mediate MACE risk above and beyond those of traditional risk factors and surrogates of MACE and indicate the potential utility of the SE-MACE risk score in clinical and public health settings for risk stratification and community wide identification of vulnerable populations respectively.

Prior research has established links between socio-environmental factors and cardiovascular risk [1,3]. However, many studies have examined these factors in isolation or as limited indices [6,29], For instance, while the impact of individual factors like economic hardship [30], or air pollution [29] on CVD is recognized, fewer studies have comprehensively assessed the combined influence of diverse socio-environmental variables on individual cardiovascular risk [31,32] Our work builds upon neighborhood-level analyses, including our own prior mapping of socio-environmental factors to geographic CHD prevalence [33]. This study advances the field by integrating a broad spectrum of 22 socio-environmental factors to evaluate individual cardiovascular risk. To our knowledge, this is a novel effort to synthesize such diverse variables into a unified, individual-level risk assessment model.

In this study, we strategically utilized the RSF model followed by the CPH model to capitalize on their distinct strengths. The RSF model's ability to handle complex, high-dimensional data without prior assumptions made it ideal for developing a comprehensive SE risk score. With to its non-parametric nature, RSF allowing for the capture of complex, non-linear relationships and interactions between the numerous SE variables without the need for a priori assumptions about the data. This approach effectively manages the multi-collinearity and interaction effects inherent in multiple SE variables. The explainability of ML models becomes progressively crucial in studying the impact of a range of factors. In this work, the use of SHAP features provides additional context in terms of directionality of predicted features. Transitioning to the CPH model with the composite SE risk score allowed for a more streamlined and interpretable analysis. The CPH model's strength lies in its clear quantification of hazard ratios and the facility to control for additional covariates. This two-pronged approach combines

Table 2
Cox proportional hazards models with composite SE-MACE risk score.

	Model 1[a]		Model 2[b]		Model 3[c]	
Variable	HR (95 CI)	p value	HR (95 CI)	p value	HR (95 CI)	p value
SE-MACE Risk Score	1.28 (1.19, 1.39)	<0.001	1.22 (1.12, 1.33)	<0.001	1.21 (1.11, 1.32)	<0.001
Age			1.04 (1.03, 1.05)	<0.001	1.03 (1.01, 1.04)	<0.001
Sex						
Female (ref)			–	–	–	–
Male			1.46 (1.23, 1.75)	<0.001	1.15 (0.96, 1.39)	0.13
Race						
White (ref)			–	–	–	–
Black			0.83 (0.62, 1.11)	0.22	0.94 (0.70, 1.26)	0.66
Other/ Unknown			0.82 (0.47, 1.43)	0.48	0.86 (0.50, 1.50)	0.59
Hypertension			3.97 (2.96, 5.32)	<0.001	3.63 (2.70, 4.87)	<0.001
Diabetes			1.92 (1.60, 2.32)	<0.001	1.75 (1.45, 2.11)	<0.001
Dyslipidemia			1.37 (1.03, 1.83)	0.03	1.23 (0.92, 1.65)	0.16
Log2(CAC)					1.12 (1.09, 1.15)	<0.001

[a] Model 1: unadjusted.
[b] Model 2: adjusted for age, race, sex, hypertension, diabetes, and dyslipidemia.
[c] Model 3: adjusted for age, race, sex, hypertension, diabetes, dyslipidemia, and CAC score.
All the models were evaluated in the hold-out test set.
HR, hazard ratio; CI, confidence interval; SE, social environmental; MACE, major adverse cardiovascular events; CAC, coronary artery calcium.

the robust feature selection capability of machine learning with the interpretability of traditional survival analysis, offering a more nuanced understanding of SE factors on cardiovascular risk.

The absence of a significant association for race in the CPH model suggests that race may not be a valid covariate in widely used risk predictors, likely due to its erroneous treatment as a biological construct. We hypothesize that previous association typically attributed to race is perhaps captured by variation in socio-environmental exposures, as indicated by the moderate correlation between race and our SE-MACE risk score (Fig. 3). This perspective aligns with the American Heart Association's (AHA) recent shift in their risk prediction model, where race has been replaced by social determinants of health (SDOH) to provide a more accurate and equitable assessment of cardiovascular risk [34,35]. Our study aligns with the PREVENT model's use of a place-based Social Determinants Index (SDI), which prioritizes place-based social determinants over race in CVD risk assessment. However, our approach extends this framework by using census-tract-level data and an unbiased, data-driven model that incorporates a broader set of socio-environmental and behavioral risk factors, allowing for a more detailed assessment of cardiovascular risk.

Our study's SHAP analysis, identifying key socio-environmental variables, largely reinforces existing literature while also highlighting the composite impact of these factors. Consistent with established research [36,37] smoking prevalence showed a strong positive correlation with CVD risk. Similarly, the importance of insurance status and the negative correlation of median household income align with the recognized roles of healthcare access and socioeconomic factors in cardiovascular health [1,38,39]. Our findings also extend to environmental factors like PM2.5 and walkability, supporting the growing evidence base on living conditions and health [40], as well as the relevance of social vulnerability [41,42]. While our SE-MACE score incorporates 22 variables for optimal prediction, these top predictors – smoking, insurance, income, sleep, and education – suggest potential for a simplified, yet effective, risk assessment tool for future applications.

4.1. Limitations

Our study's limitations include the potential limited generalizability of findings from the CLARIFY registry, which may not represent diverse healthcare systems and socio-environmental contexts. The observational nature of the study restricts causal inference and acknowledges the possibility of unmeasured confounders. Additionally, selection bias might arise from the criteria for CAC screening and voluntary participation, potentially affecting cohort representativeness. A temporal mismatch between socio-environmental data collection and health outcomes recording could impact the accuracy of associations. Finally, the lack of external validation also leaves questions about the model's wider applicability. Despite these constraints, our study underscores the significant impact of socio-environmental factors on cardiovascular disorders, supported by data from a large, racially diverse cohort.

5. Conclusion

In summary, our study provides a comprehensive analysis of the impact of socio-environmental factors on major adverse cardiovascular events (MACE) through the development and validation of a composite SE-MACE risk score. The SE-MACE score was found to has robust association with MACE, after adjusting for common clinical risk factors. We have identified key socio-environmental determinants such as smoking prevalence, insurance status, median household income, lack of sleep, education and $PM_{2.5}$ that significantly influence cardiovascular risk. The implications of our findings are substantial for public health and clinical practice. They highlight the necessity of integrating socio-environmental factors into cardiovascular risk assessment, suggesting that a more holistic approach could lead to improved prevention and management strategies for cardiovascular diseases. This research contributes to a deeper understanding of MACE risk and underscores the importance of addressing socio-environmental determinants to effectively tackle the challenges of cardiovascular health in diverse populations.

Disclosures

None of the authors have conflicts of interest relevant to the contents of this manuscript.

Funding

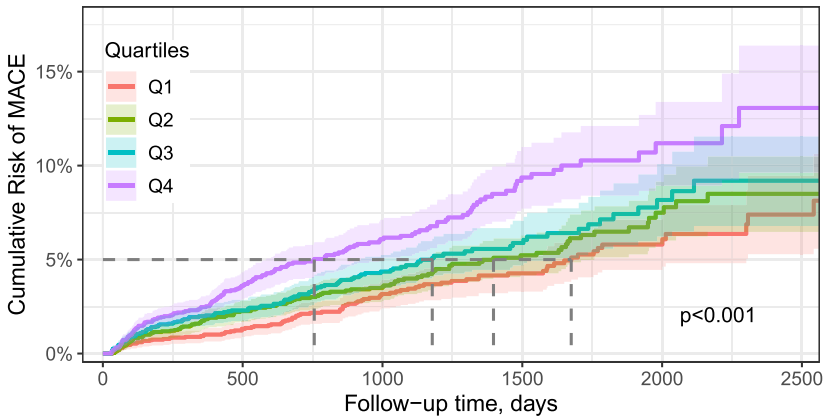
This work was funded by the National Institute on Minority Health and Health Disparities Award # P50MD017351 and 1R35ES031702-01 awarded to Dr. Rajagopalan.

Ethical review statement

Data were collected from the CLARIFY registry (NCT04075162), with patient consent for inclusion waived by the University Hospitals Health System (UHHS) Institutional Review Board.

CRediT authorship contribution statement

Zhuo Chen: Writing – review & editing, Writing – original draft,



	At Risk					
Q1	3083	2186	1436	791	340	134
Q2	3236	2179	1384	751	323	133
Q3	3057	2035	1209	565	220	76
Q4	3107	1970	1108	502	170	57
	Events					
Q1	0	36	69	82	92	96
Q2	0	63	87	106	118	121
Q3	0	63	97	111	120	122
Q4	0	95	135	160	166	168

Fig. 2. Cumulative risk of MACE for individuals in the holdout test set within four SE-MACE risk score quartiles. The shading straddling each line represents the confidence interval. Dashed lines indicate the days for each group to reach 5 % cumulative risk. *p* value was calculated in a log rank test.
Abbreviations: *MACE* major adverse cardiovascular events; SE socio-environmental.

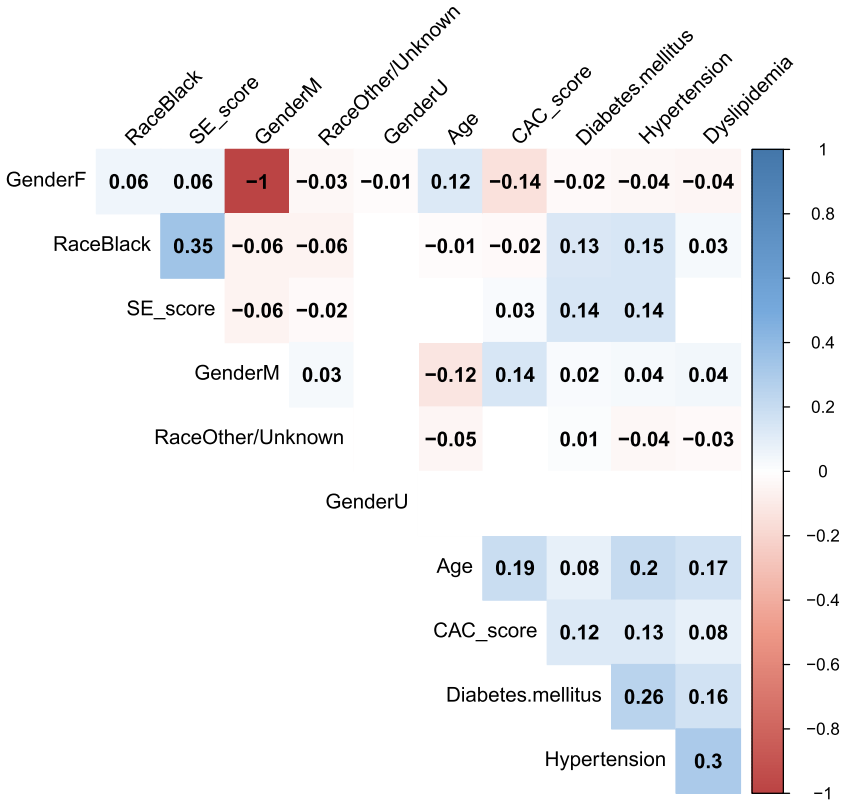
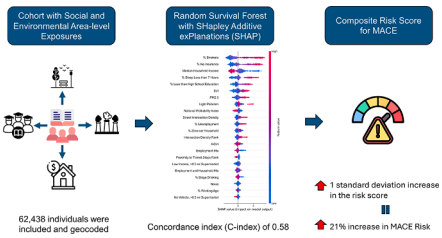


Fig. 3. Correlation matrix of the variables used in the CPH model. Numbers represent the Pearson correlation coefficients at the 0.01 significance level. Blank square indicates the pair is not significantly correlated.

Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jean-Eudes Dazard:** Writing – review & editing, Visualization, Methodology, Investigation. **Pedro Rafael Vieira de Oliveira Salerno:** Writing – review & editing, Methodology, Investigation. **Santosh Kumar Sirasapalli:** Software, Investigation, Data curation. **Mohamed HE Makhlouf:** Software, Data curation. **Sanjay Rajagopalan:** Writing – review & editing, Validation, Supervision,



Central illustration. The development of a composite cardiovascular risk score using socio-environmental data from over 62,000 individuals, processed through a Random Survival Forest with SHAP analysis to predict MACE. Abbreviations: MACE major adverse cardiovascular events, C-index: the concordance index.

Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Sadeer Al-Kindi:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Sanjay Rajagopalan reports financial support was provided by National Institute on Minority Health and Health Disparities. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ajpc.2025.100964.

References

[1] Havranek EP, Mujahid MS, Barr DA, et al. Social determinants of risk and outcomes for cardiovascular disease. Circulation 2015;132(9):873–98. https://doi.org/10.1161/CIR.0000000000000228.

[2] Powell-Wiley TM, Baumer Y, Baah FO, et al. Social determinants of cardiovascular disease. Circ Res 2022;130(5):782–99. https://doi.org/10.1161/CIRCRESAHA.121.319811.

[3] Brandt EJ, Tobb K, Cambron JC, et al. Assessing and addressing social determinants of cardiovascular health. JACC 2023;81(14):1368–85. https://doi.org/10.1016/j.jacc.2023.01.042.

[4] Teshale AB, Htun HL, Owen A, et al. The role of social determinants of health in cardiovascular diseases: an umbrella review. J Am Heart Assoc 2023;12(13):e029765. https://doi.org/10.1161/JAHA.123.029765.

[5] McNeill E, Lindenfeld Z, Mostafa L, et al. Uses of social determinants of health data to address cardiovascular disease and health equity: a scoping review. J Am Heart Assoc 2023;12(21):e030571. https://doi.org/10.1161/JAHA.123.030571.

[6] Rajagopalan S, Al-Kindi SG, Brook RD. Air pollution and cardiovascular disease: JACC state-of-the-art review. J Am Coll Cardiol 2018;72(17):2054–70.

[7] Münzel T, Sørensen M, Lelieveld J, et al. Heart healthy cities: genetics loads the gun but the environment pulls the trigger. Eur Heart J 2021;42(25):2422–38. https://doi.org/10.1093/eurheartj/ehab235.

[8] Bevan GH, Al-Kindi SG, Brook RD, Münzel T, Rajagopalan S. Ambient air pollution and atherosclerosis: insights into dose, time, and mechanisms. Arterioscler Thromb Vasc Biol 2021;41(2):628–37.

[9] Bevan GH, Freedman DA, Lee EK, Rajagopalan S, Al-Kindi SG. Association between ambient air pollution and county-level cardiovascular mortality in the United States by social deprivation index. Am Heart J 2021;235:125–31.

[10] Motairek I, Sharara J, Makhlouf MH, et al. Association between particulate matter pollution and CKD mortality by social deprivation. Am J Kidney Dis 2022. Published online.

[11] Rajagopalan S, Vergara-Martel A, Zhong J, et al. The urban environment and cardiometabolic health. Circulation 2024;149(16):1298–314. https://doi.org/10.1161/CIRCULATIONAHA.123.067461.

[12] Greenland P, Blaha MJ, Budoff MJ, Erbel R, Watson KE. Coronary calcium score and cardiovascular risk. J Am Coll Cardiol 2018;72(4):434–47. https://doi.org/10.1016/j.jacc.2018.05.027.

[13] Al-Kindi S, Tashtish N, Rashid I, et al. Effect of no-charge coronary artery calcium scoring on cardiovascular prevention. Am J Cardiol 2022;174:40–7. https://doi.org/10.1016/j.amjcard.2022.03.019.

[14] Gupta A, Lau E, Varshney R, et al. The identification of calcified coronary plaque is associated with initiation and continuation of pharmacological and lifestyle preventive therapies. JACC Cardiovasc Imaging 2017;10(8):833–42. https://doi.org/10.1016/j.jcmg.2017.01.030.

[15] Choi BG, Rha SW, Yoon SG, Choi CU, Lee MW, Kim SW. Association of major adverse cardiac events up to 5 years in patients with chest pain without significant coronary artery disease in the Korean population. J Am Heart Assoc 2019;8(12):e010541. https://doi.org/10.1161/JAHA.118.010541.

[16] CDC. PLACES: Local Data for Better Health. Centers for disease control and prevention. April 12, 2023. Accessed June 20, 2023. https://www.cdc.gov/places.

[17] Chapman J, Fox EH, Bachman W, Frank LD, Thomas J, Rourk Reyes A. *Smart Location Database Technical Documentation and User Guide Version 3.0.*; 2021. Accessed December 11, 2023. https://www.epa.gov/system/files/documents/2023-10/epa_sld_3.0_technicaldocumentationuserguide_may2021_0.pdf.

[18] Economic Research Service (ERS), U.S. Department of Agriculture (USDA). Food Access Research Atlas. Accessed December 11, 2023. https://www.ers.usda.gov/data-products/food-access-research-atlas/.

[19] United States Census Bureau. American Community Survey 5-Year Data (2009-2022). Census.gov. Accessed December 11, 2023. https://www.census.gov/data/developers/data-sets/acs-5year.html.

[20] Flanagan BE, Gregory EW, Hallisey EJ, Heitgerd JL, Lewis B. A social vulnerability index for disaster management. J Homel Secur Emerg Manag 2011;8(1).

[21] van Donkelaar A, Hammer MS, Bindle L, et al. Monthly global estimates of fine particulate matter and their uncertainty. Environ Sci Technol 2021;55(22):15287–300. https://doi.org/10.1021/acs.est.1c05309.

[22] Falchi F, Cinzano P, Duriscoe D, et al. The new world atlas of artificial night sky brightness. Sci Adv 2016;2(6):e1600377. https://doi.org/10.1126/sciadv.1600377.

[23] Bureau of Transportation StatisticsNational Transportation Noise Map . Accessed December 11, 2023. https://www.bts.gov/geospatial/national-transportation-noise-map.

[24] U.S. Geological Survey (USGS) Earth resources observation and science (EROS) center. eVIIRS Global NDVI. U.S. Geological Survey; 2012. doi:10.5066/P9QOEFNP.

[25] Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat 2008;2(3):841–60. https://doi.org/10.1214/08-AOAS169.

[26] Harrell Jr FE, RM C, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247(18):2543–6. https://doi.org/10.1001/jama.1982.03320430047030.

[27] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30. Accessed December 11, 2023, https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

[28] Taye AD, Borga LG, Greiff S, Vögele C, D'Ambrosio C. A machine learning approach to predict self-protecting behaviors during the early wave of the COVID-19 pandemic. Sci Rep 2023;13(1):6121. https://doi.org/10.1038/s41598-023-33033-1.

[29] Al-Kindi SG, Brook RD, Biswal S, Rajagopalan S. Environmental determinants of cardiovascular disease: lessons learned from air pollution. Nat Rev Cardiol 2020;17(10):656–72.

[30] Bevan GH, Josephson R, Al-Kindi SG. Socioeconomic deprivation and heart failure mortality in the United States. J Card Fail 2020;26(12):1106–7.

[31] Khadke S, Kumar A, Al-Kindi S, et al. Association of environmental injustice and cardiovascular diseases and risk factors in the United States. J Am Heart Assoc 2024;13(7):e033428. https://doi.org/10.1161/JAHA.123.033428.

[32] Salerno PR, Motairek I, Dong W, et al. County-level socio-environmental factors associated with stroke mortality in the United States: a cross-sectional study. Angiology 2024. Published online00033197241244814.

[33] Al-Kindi S, Bevan G, Madabhushi A, et al. Socioeconomic and environmental exposures (Socio-enviromics) and cardiometabolic health in the United States. Circulation 2021;144(Suppl_1):A13960. -A13960.

[34] Ghosh AK, Venkatraman S, Nanna MG, et al. Risk prediction for atherosclerotic cardiovascular disease with and without race stratification. JAMA Cardiol 2023. https://doi.org/10.1001/jamacardio.2023.4520. Published online December 6.

[35] Khan SS, Coresh J, Pencina MJ, et al. Novel prediction equations for absolute risk assessment of total cardiovascular disease incorporating cardiovascular-kidney-metabolic health: a scientific statement from the American Heart Association. Circulation 2025;0(0). https://doi.org/10.1161/CIR.0000000000001191.

[36] Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease. J Am Coll Cardiol 2004;43(10):1731–7. https://doi.org/10.1016/j.jacc.2003.12.047.

[37] Duncan MS, Freiberg MS, Greevy Jr RA, Kundu S, Vasan RS, Tindle HA. Association of smoking cessation with subsequent risk of cardiovascular disease. JAMA 2019;322(7):642–50. https://doi.org/10.1001/jama.2019.10298.

[38] McClurkin MA, Yingling LR, Ayers C, et al. Health insurance status as a barrier to ideal cardiovascular health for U.S. adults: data from the national health and nutrition examination survey (NHANES). PLoS One 2015;10(11):e0141534. https://doi.org/10.1371/journal.pone.0141534.

[39] Azap RA, Nolan TS, Gray DM, et al. Association of socioeconomic status with ideal cardiovascular health in black men. J Am Heart Assoc 2021;10(23):e020184. https://doi.org/10.1161/JAHA.120.020184.

[40] Rajagopalan S, Landrigan PJ. Pollution and the heart. N Engl J Med 2021;385(20):1881–92. https://doi.org/10.1056/NEJMra2030281.
[41] Khan SU, Javed Z, Lone AN, et al. Social vulnerability and premature cardiovascular mortality among US counties, 2014 to 2018. Circulation 2021;144(16):1272–9. https://doi.org/10.1161/CIRCULATIONAHA.121.054516.
[42] Bevan G, Pandey A, Griggs S, et al. Neighborhood-level social vulnerability and prevalence of cardiovascular risk factors and coronary heart disease. Curr Probl Cardiol 2022:101182. https://doi.org/10.1016/j.cpcardiol.2022.101182. Published online March 27.