

## ORIGINAL RESEARCH ARTICLE



# Heterogeneity in the Association Between the Presence of Coronary Artery Calcium and Cardiovascular Events: A Machine-Learning Approach in the MESA Study

Kosuke Inoue<sup>1</sup>, MD, PhD, Teresa E. Seeman, PhD, Tamara Horwich<sup>2</sup>, MD, MS, Matthew J. Budoff<sup>3</sup>, MD, Karol E. Watson<sup>4</sup>, MD, PhD

**BACKGROUND:** Coronary artery calcium (CAC) has been widely recognized as an important predictor of cardiovascular disease (CVD). Given the finite resources, it is important to identify individuals who would receive the most benefit from detecting positive CAC by screening. However, the evidence is limited as to whether the burden of positive CAC on CVD differs by multidimensional individual characteristics. We sought to investigate the heterogeneity in the association between positive CAC and incident CVD.

**METHODS:** This cohort study included adults from MESA (Multi-Ethnic Study of Atherosclerosis) ages  $\geq 45$  years and free of cardiovascular disease. After propensity score matching in a 1:1 ratio, we applied a machine learning causal forest model to (1) evaluate the heterogeneity in the association between positive CAC and incident CVD, and (2) predict the increase in CVD risk at 10-years when  $CAC > 0$  (versus  $CAC = 0$ ) at the individual level. We then compared the estimated increase in CVD risk when  $CAC > 0$  to the absolute 10-year atherosclerotic CVD (ASCVD) risk calculated by the 2013 American College of Cardiology/American Heart Association pooled cohort equations.

**RESULTS:** Across 3328 adults in our propensity score–matched analysis, our causal forest model showed the heterogeneity in the association between  $CAC > 0$  and incident CVD. We found a dose–response relationship of the estimated increase in CVD risk when  $CAC > 0$  with higher 10-year ASCVD risk. Almost all individuals (2293 of 2428 [94.4%]) with borderline risk of ASCVD or higher showed  $\geq 2.5\%$  increase in CVD risk when  $CAC > 0$ . Even among 900 adults with low ASCVD risk, 689 (69.2%) showed  $\geq 2.5\%$  increase in CVD risk when  $CAC > 0$ ; these individuals were more likely to be male, Hispanic, and have unfavorable CVD risk factors than others.

**CONCLUSIONS:** The expected increases in CVD risk when  $CAC > 0$  were heterogeneous across individuals. Moreover, nearly 70% of people with low ASCVD risk showed a large increase in CVD risk when  $CAC > 0$ , highlighting the need for CAC screening among such low-risk individuals. Future studies are needed to assess whether targeting individuals for CAC measurements based on not only the absolute ASCVD risk but also the expected increase in CVD risk when  $CAC > 0$  improves cardiovascular outcomes.

**Key Words:** atherosclerosis ■ cardiovascular disease ■ machine learning ■ precision medicine ■ propensity score

Coronary artery calcium (CAC) is a key marker of subclinical atherosclerotic cardiovascular disease (ASCVD).<sup>1</sup> Ample evidence has shown the relationship between CAC and incident cardiovascular dis-

ease (CVD) and suggested that CAC can aid in enhancing traditional CVD risk prediction models.<sup>2–5</sup> In addition, a recent study reported that statin therapy was associated with a reduced risk of incident CVD across all CAC burden

Correspondence to: Kosuke Inoue, MD, PhD, Department of Social Epidemiology, Graduate School of Medicine, Kyoto University, Floor 2, Science Frontier Laboratory, Yoshida-konoe-cho, Sakyo-ku, Kyoto, Kyoto 604-8146, Japan. Email inoue.kosuke.2j@kyoto-u.ac.jp

This manuscript was sent to Craig Glastonbury, Guest Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at <https://www.ahajournals.org/doi/suppl/10.1161/CIRCULATIONAHA.122.062626>.

For Sources of Funding and Disclosures, see page 140.

© 2022 The Authors. *Circulation* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the [Creative Commons Attribution Non-Commercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited, the use is noncommercial, and no modifications or adaptations are made.

*Circulation* is available at [www.ahajournals.org/journal/circ](http://www.ahajournals.org/journal/circ)

## Clinical Perspective

### What Is New?

- Using the longitudinal cohort study of the multiethnic population, along with propensity score matching and machine learning causal forest algorithm, we found heterogeneity in the association between coronary artery calcium (CAC) >0 and incident cardiovascular disease (CVD) risk (ie, the expected increase in CVD risk when CAC>0 varied across individuals).
- The 10-year atherosclerotic CVD risk was associated in a dose–response manner with the expected increase in CVD risk when CAC>0.
- Even among people with low absolute atherosclerotic CVD risk, nearly 70% had a large, expected increase in CVD risk when CAC>0, and they were more likely to be male, Hispanic, and have unfavorable CVD risk factors.

### What Are the Clinical Implications?

- These findings provide empirical evidence for the potential utility of targeted CAC measurements for such subpopulations with a large increase in CVD risk related to a positive CAC even when their atherosclerotic CVD risk is low.
- It is important to consider not only outcome risk but also the exposure–outcome association (or the treatment effect) at individual levels to achieve the most effective distribution of finite resources such as CAC screening with advancing cardiovascular health equity.
- Additional studies are needed to evaluate whether targeting people for CAC measurements based on the potential burden of positive CAC on CVD (ie, CAC–CVD association) at individual levels improves cardiovascular outcomes as a future precision medicine approach.

### Nonstandard Abbreviations and Acronyms

<b>ACC/AHA</b>	American College of Cardiology/American Heart Association
<b>ASCVD</b>	atherosclerotic cardiovascular disease
<b>CAC</b>	coronary artery calcium
<b>CVD</b>	cardiovascular disease
<b>MESA</b>	Multi-Ethnic Study of Atherosclerosis

strata except the absence of CAC (ie, CAC=0), highlighting the role of CAC=0 in decision-making regarding statin initiation.<sup>6</sup> The 2018 American Heart Association (AHA)/American College of Cardiology (ACC)/Multi-Society Cholesterol Guidelines and the 2019 ACC/AHA Primary Prevention Clinical Practice Guidelines upgraded CAC measurement as a Class IIa recommendation and

endorsed the use of CAC as a supporting tool for personalized CVD risk management in primary prevention.<sup>7,8</sup> Although the optimal target population for CAC measurements has often been considered according to the estimated CVD risks under an implicit assumption of the correlation between risk and benefit,<sup>9</sup> it is unclear whether individuals at high risk of CVD would receive the most benefit from the information gained with CAC measurement.

In recent years, heterogeneity in the association between risk factors and incident CVD has received substantial attention. Previous studies showed that the strength of the association between CAC and incident CVD increased by the number of traditional CVD risk factors<sup>10</sup> although it did not statistically differ by age, sex, and race/ethnicity.<sup>11</sup> However, these studies were based on a traditional statistical approach—subgroup analysis—that limits the identification of individuals with the strongest or weakest associations because it reduces the dimension of characteristics into a binary or simple category. To overcome this limitation, there has been rapid advancement in the machine learning–based approach which allows us to investigate how associations vary by multidimensional individual characteristics.<sup>12,13</sup> The causal forest model is one such method that uses an ensemble of trees optimized to detect heterogeneity in the association across the individual levels.<sup>14,15</sup> Although it uses a similar algorithm to the random forest—one of the common machine learning algorithms for outcome prediction, causal forest allows us to evaluate the outcome risk associated with the exposure at individual levels rather than the outcome itself. Given the finite resources and increasing concerns about overtesting and overtreatment,<sup>1,16</sup> identifying individuals with a large change in CVD risk related to positive CAC by applying the causal forest model would provide new insight into the precision-medicine approach of CAC screening to effectively reduce the CVD burden.

Therefore, using data from MESA (Multi-Ethnic Study of Atherosclerosis), along with the causal forest model and propensity score matching, we examined the heterogeneity in the association between positive CAC and incident CVD among US adults (ie, whether the estimated increase in CVD risk when CAC>0 varied across individuals). We then examined the correlation between the estimated increase in CVD risk when CAC>0 and the 10-year ASCVD risk calculated by the 2013 ACC/AHA pooled cohort equations.<sup>17</sup> Lastly, we compared characteristics between individuals with a high burden of positive CAC on CVD and those with a low burden of positive CAC on CVD.

## METHODS

### Data Sources and Study Population

The data and materials are available and can be requested at <http://www.mesa-nhlbi.org>. The MESA is a prospective observational cohort of 6814 men and women aged 45 to 84 years

without known CVD at enrollment.<sup>18</sup> The participants, who identified themselves as White, Black, Hispanic, or Asian (mostly of Chinese origin), were enrolled from 6 communities in the United States (Baltimore, Maryland; Chicago, Illinois; Forsyth County, North Carolina; Los Angeles, California; New York, NY; St. Paul, Minnesota) between 2000 and 2002. Of the 6814 participants, we excluded participants with missing data on covariates listed below (n=157). We then excluded participants who did not have follow-up information on outcomes during 10 years since the enrollment (n=1063), resulting in a sample of 5594 participants before propensity score matching (Figure 1). MESA was approved by the Institutional Review Boards of the participating institutions, and the study was conducted following the Declaration of Helsinki. All participants gave written informed consent for participation in the studies. Full details of the MESA study design can be found elsewhere.<sup>18</sup> This study followed the PRIME (Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation) checklist (Table S1).<sup>19</sup>

### Exposure Ascertainment

At baseline, the CAC score was measured twice for each individual using a cardiac-gated electron-beam computed tomography scanner (Chicago, Los Angeles, New York) or a 64-slice multidetector CT (Baltimore, Forsyth County, St. Paul).<sup>20</sup> The intra- and interobserver agreements were  $\kappa = 0.93$  and  $0.90$ , respectively. Participants were told either they had no CAC or that the amount was less than average, average, or greater than average and that they should discuss the results with their physicians.

### Outcome Ascertainment

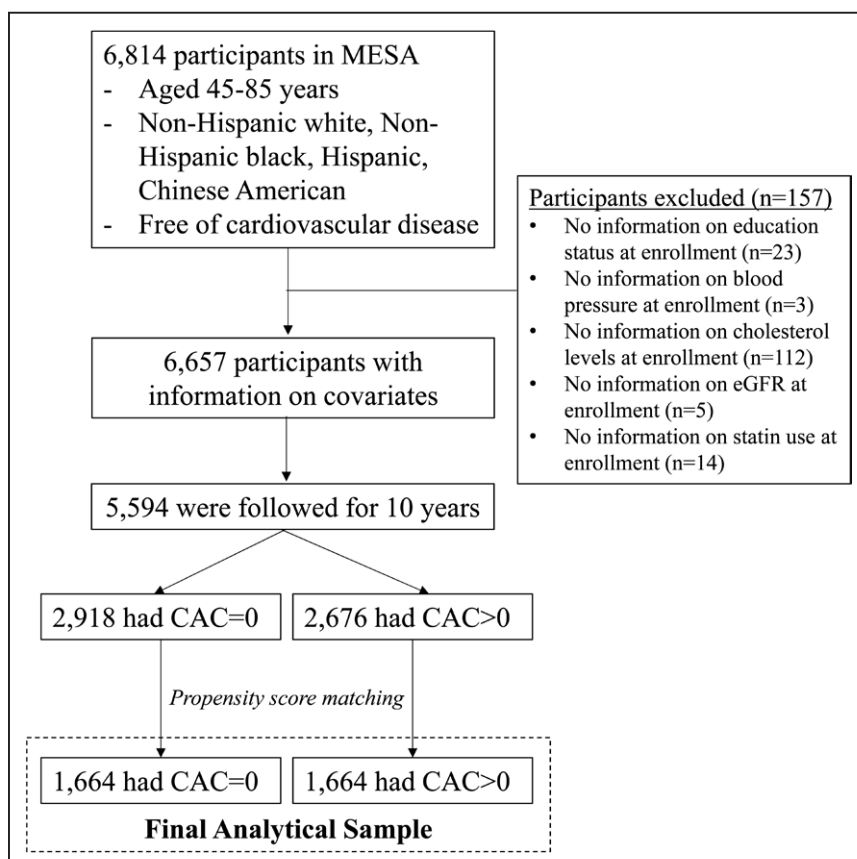
For the purpose of this study, we used incident hard CVD events at 10 years of follow-up, which included myocardial infarction, resuscitated cardiac arrest, stroke, coronary heart disease death, and stroke death. During follow-up examinations conducted every 9 to 12 months, a telephone interviewer inquired about all interim hospital admissions, cardiovascular diagnoses, and deaths. Two physicians from the MESA study events committee independently reviewed all the medical records for end-point classification and assignment of incidence dates. A more detailed description of the follow-up methods and event adjudication is available at [www.mesa-nhlbi.org](http://www.mesa-nhlbi.org).

### Other Covariates

Participants completed self-administered questionnaires including age, sex (male, female), race/ethnicity (White, Black, Hispanic, Asian), education status (less than college, college or above), health insurance status (public, private, uninsured), smoking status (never, former, current), history of diabetes, and medication use. Weight, height (which were used to calculate body mass index), and blood pressure were measured by trained staff. Creatinine and cholesterol levels (total, high-density lipoprotein [HDL], and low-density lipoprotein [LDL]) were measured from peripheral blood samples. The estimated glomerular filtration rate (eGFR) was calculated by the Chronic Kidney Disease Epidemiology Collaboration equation.<sup>21</sup>

### Propensity Score Matching

We employed 1-to-1 propensity score matching without replacement to match participants who had CAC>0 with those



**Figure 1. Flow of study sample selection.**

Sensitivity analysis was conducted using all 5594 samples before propensity score matching. CAC indicates coronary artery calcium; and MESA, Multi-Ethnic Study of Atherosclerosis.

who had CAC=0 at baseline, to adjust for potential confounders listed in Table 1. We used a logistic regression model to derive the propensity scores for positive CAC, and a caliper of 0.1 standard deviations of the logit of the propensity score. The absolute standardized mean difference <0.1 indicated successful balance.<sup>22</sup>

## Statistical Analyses

After describing baseline characteristics before and after propensity score matching, we applied a machine learning causal forest algorithm (using *grf* package in R) among the matched sample to build a model to predict the increase in hard CVD risk when CAC>0 (compared to CAC=0) at the individual level. In this model, an ensemble of 2000 causal trees was built using all of the covariates included in propensity score matching. To minimize the overfitting of the model, we employed an honest splitting approach along with 10-fold cross-fitting.<sup>14,15,23</sup> In this approach, each tree algorithm was built using a randomly chosen 50% subsample from the original data without replacement. The fractional subsample was then divided into 2 halves, with the first half used to construct the tree structure and the second half used to make predictions. The parameters of our causal forest model (eg, the minimum number of samples a node should contain, the number of variables considered during each split) were tuned using 10-fold cross-validation. Model calibration was evaluated by fitting the best linear fit of the regression of the observed association on the predicted association.<sup>24</sup> The c-for-benefit was also calculated to evaluate the discrimination performance of our model.<sup>25</sup> The variable importance was assessed based on a simple weighted sum of how many times each variable was split in the causal forest model without consideration of the stage of split.<sup>24</sup> More details on causal forest analysis are shown in the Expanded Methods, Figures S1 to S4, and elsewhere.<sup>14,15,24</sup>

Then, we applied our causal forest model to MESA participants to predict the increase in CVD risk when CAC>0 at the individual level (because the model was built using 10-fold cross-fitting, estimates for individuals were calculated based on trees fitted without their own observations and thus had low risk of overfitting<sup>23</sup>). We computed Spearman's correlation coefficient and Pearson's correlation coefficient to investigate the correlation between the estimated increase in CVD risk when CAC>0 and the 10-year ASCVD risk calculated by the 2013 ACC/AHA pooled cohort equations<sup>7</sup> as well as the 4 major risk factors of CVD (age, systolic blood pressure, body mass index, and LDL-cholesterol levels).

Lastly, we identified subpopulations with low 10-year ASCVD risk (<5% based on the 2013 AHA/ACC pooled cohort equations) and a large increase in CVD risk when CAC>0 (≥2.5% based on our causal forest model), and compared their characteristics with those with low 10-year ASCVD risk and a small increase in CVD risk when CAC>0 (<2.5% based on our causal forest model). We set 2.5% increase as the threshold given that borderline 10-year ASCVD risk ranges from 5 to 7.5%.<sup>7,8</sup> We also compared the characteristics according to the increase in CVD risk when CAC>0 among individuals with borderline or higher 10-year ASCVD risk (≥5% based on the 2013 AHA/ACC pooled cohort equations). As sensitivity analyses, we reset the threshold using the median for each subpopulation instead

of 2.5% increase (3.2% increase for adults with low 10-year ASCVD risk; and 5.5% increase for adults with borderline or higher 10-year ASCVD risk). *P* values were adjusted for multiple comparisons using the Benjamini–Hochberg method.<sup>26</sup> All analyses were conducted using R version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria).

## Additional Analyses

We conducted 2 additional analyses. First, we applied the causal forest model to the entire 5594 participants before propensity score matching rather than 3328 propensity score-matched sample. Second, given the usefulness of CAC measurement for shared decision-making of initiating statin therapy for untreated patients in clinical practice,<sup>8</sup> we restricted individuals to those without statin use at baseline, in order to assess whether our findings can be applied to adults who have not started statin therapy yet.

## RESULTS

### Participant Characteristics

Across 5594 participants in the MESA before propensity score matching, the mean±SD age was 61.3 (10.0) years, and 53% were female. Compared to participants with CAC=0 (n=2918 [52.6%]), those with positive CAC (n=2676 [47.8%]) were more likely to be older, male, White (as opposed to Black, Hispanic, or Asian), less educated, publicly insured, and former smokers (Table 1). They also showed unfavorable CVD risk profiles: higher systolic blood pressure, higher diastolic blood pressure, lower eGFR, higher prevalence of diabetes, higher prevalence of antihypertensive use, and higher prevalence of statin use. The 2 groups (CAC>0 and CAC=0) were well balanced on all baseline covariates after propensity score matching; standardized mean differences were 0.1 for all covariates (Table 1; Figure S5).

### Causal Forest Model Predicts Increased CVD Risk When CAC>0

During 10 years of follow-up among 3328 propensity score-matched samples, 231 (6.9%) participants experienced composite CVD events (CAC=0; 69 of 1674 [4.1%]; CAC>0, 162 of 1674 [9.7%]). Our causal forest model suggested the presence of heterogeneity in the association between CAC>0 and incident CVD events (ie, the CAC–CVD association varied across individuals; Figure S6). In the best linear fit model for the observed association, the coefficient of the mean forest prediction was 1.00 (*P* value <0.001), indicating that the mean forest prediction was well calibrated. Also, the coefficient of the out-of-bag predicted association was 0.93 (*P* value <0.001), indicating that the forest captured heterogeneity. The c-for-benefit of our causal forest model was 0.87 (95% CI, 0.74 to 1.00), indicating the high discrimination

**Table 1. Baseline Characteristics in the MESA Cohorts Before and After Propensity Score Matching**

Variable*	Cohort before propensity score matching†			Cohort after propensity score matching†		
	CAC=0	CAC>0	ASMD	CAC=0	CAC>0	ASMD
Participants, n	2918	2676	–	1664	1664	–
Age	57.6 ± 9.0	65.4 ± 9.4	0.854	61.8 ± 8.7	62.1 ± 9.1	0.036
Sex, %						
Male	36.3	57.6	0.437	46.8	49.0	0.043
Female	63.7	42.4	−0.437	53.2	51.0	−0.043
Race/ethnicity, %						
White	35.1	45.2	0.207	39.6	40.8	0.025
Black	29.9	23.6	−0.142	27.3	26.4	−0.02
Hispanic	23.1	19.2	−0.096	20.9	20.2	−0.018
Asian	11.9	12.0	0.003	12.2	12.6	0.013
Education status, %						
Less than college	32.5	35.9	0.071	35.6	34.1	−0.032
College or above	67.5	64.1	−0.071	64.4	65.9	0.032
Health insurance, %						
Public	14.9	24.9	0.251	20.4	20.8	0.009
Private	75.0	68.8	−0.137	70.9	71.5	0.012
Uninsured	10.1	6.3	−0.139	8.7	7.8	−0.033
Smoking status, %						
Never	56.5	45.5	−0.222	50.7	49.9	−0.016
Former	31.0	42.6	0.243	36.7	37.5	0.017
Current	12.4	11.8	−0.018	12.7	12.6	−0.002
BMI, kg/m <sup>2</sup>	28.2 ± 5.7	28.4 ± 5.2	0.03	28.4 ± 5.5	28.5 ± 5.5	0.019
Systolic blood pressure, mmHg	121.8 ± 20.3	130.2 ± 21.3	0.403	126.4 ± 21.0	126.6 ± 20.4	0.008
Diastolic blood pressure, mmHg	71.1 ± 10.3	72.7 ± 10.0	0.159	72.1 ± 10.4	72.4 ± 9.9	0.023
eGFR, mL/(min·1.73 m <sup>2</sup> )	76.2 ± 14.7	72.7 ± 17.5	−0.214	74.5 ± 14.5	74.5 ± 18.5	−0.002
Total cholesterol, mg/dL	193.6 ± 34.0	194.4 ± 34.7	0.024	195.6 ± 34.8	195.3 ± 34.8	−0.009
HDL cholesterol, mg/dL	52.8 ± 15.1	49.6 ± 14.5	−0.219	51.3 ± 14.9	51.1 ± 15.3	−0.018
LDL cholesterol, mg/dL	116.4 ± 30.6	118.7 ± 31.8	0.073	118.7 ± 30.9	118.6 ± 32.1	−0.004
Diabetes, %	8.3	15.0	0.21	11.7	12.2	0.015
Antihypertensive use, %	27.6	44.7	0.363	35.7	36.5	0.018
Statin use, %	9.8	20.7	0.307	14.4	15.1	0.019

ASMD indicates absolute standardized mean difference; BMI, body mass index; CAC, coronary artery calcium score; CVD, cardiovascular diseases; eGFR, estimated glomerular filtration rate; HDL, high-density lipoprotein; LDL, low-density lipoprotein; and MESA, Multi-Ethnic Study of Atherosclerosis.

\*Mean ± SD was described for continuous variables, otherwise indicated.

†1-to-1 pair matching was performed by nearest-neighbor matching without replacement, with the use of a caliper width equal to 0.1 of the SD of the logit of the propensity score. ASMD <0.1 was considered to indicate a negligible imbalance between the 2 groups (CAC>0 vs CAC=0).

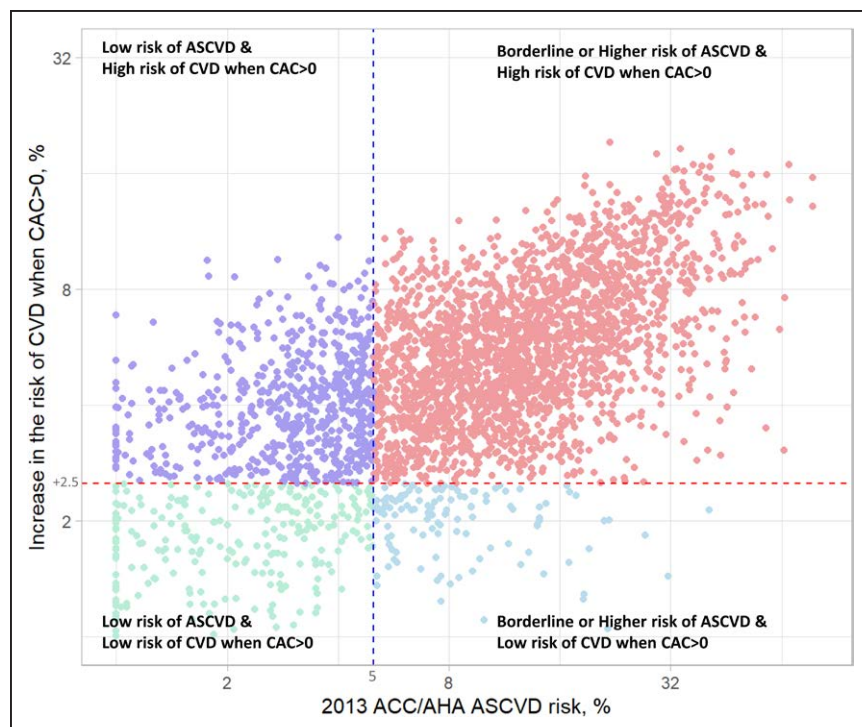
performance of the model. The variable importance calculation showed that age, eGFR, systolic and diastolic blood pressure, HDL, LDL, body mass index, and sex were frequently split when building the causal forest model (Figure S7).

### Estimated CVD Risk Increase When CAC>0 and Its Correlation With 10-Year ASCVD Risk

The increase in CVD risk when CAC>0 based on the causal forest model was associated in a dose–response manner with the 10-year ASCVD risk based on the

ACC/AHA pooled cohort equations (Spearman correlation coefficient = 0.60, *P* value <0.001; Pearson correlation coefficient = 0.61, *P* value <0.001; Figure 2). A similar pattern was also observed when we compared the increase in CVD risk when CAC>0 with age, body mass index, systolic blood pressure, and LDL-cholesterol levels (Figure 3).

Among 2428 individuals with borderline ASCVD risk or greater, 2293 (94.4%) showed a large increase in CVD risk when CAC>0, supporting the need to evaluate CAC for such a high-risk population (Table S2). Even among 900 individuals with low ASCVD risk, 623 (69.2%)



**Figure 2. Association between the 10-year ASCVD risk and the estimated increase in the risk of cardiovascular events when CAC > 0 compared to CAC = 0.**

X-axis shows the 10-year ASCVD risk calculated by the 2013 ACC/AHA pooled cohort equations. Y-axis shows the estimated increase in the risk of cardiovascular events when CAC > 0 (calculated by the causal forest model). Spearman correlation coefficient and Pearson correlation coefficient between the 10-year ASCVD risk and the estimated increase in the risk of cardiovascular events when CAC > 0 were 0.60 ( $P$  value < 0.001) and 0.61 ( $P$  value < 0.001), respectively. ASCVD indicates atherosclerotic cardiovascular disease; CAC, calcium coronary artery calcium; and CVD, cardiovascular disease.

showed a large increase in CVD risk when CAC > 0; they were more likely to be male and Hispanic (as opposed to White, Black, and Asian), and had more unfavorable CVD profiles than those with < 2.5% increase in CVD risk when CAC > 0 (Table 2). We found similar patterns when we used the median instead of 2.5% as the threshold of increase in CVD risk when CAC > 0 (ie, 3.2% increase for adults with low 10-year ASCVD risk; and 5.5% increase for adults with borderline or higher 10-year ASCVD risk; Tables S3 and S4).

### Additional Analyses

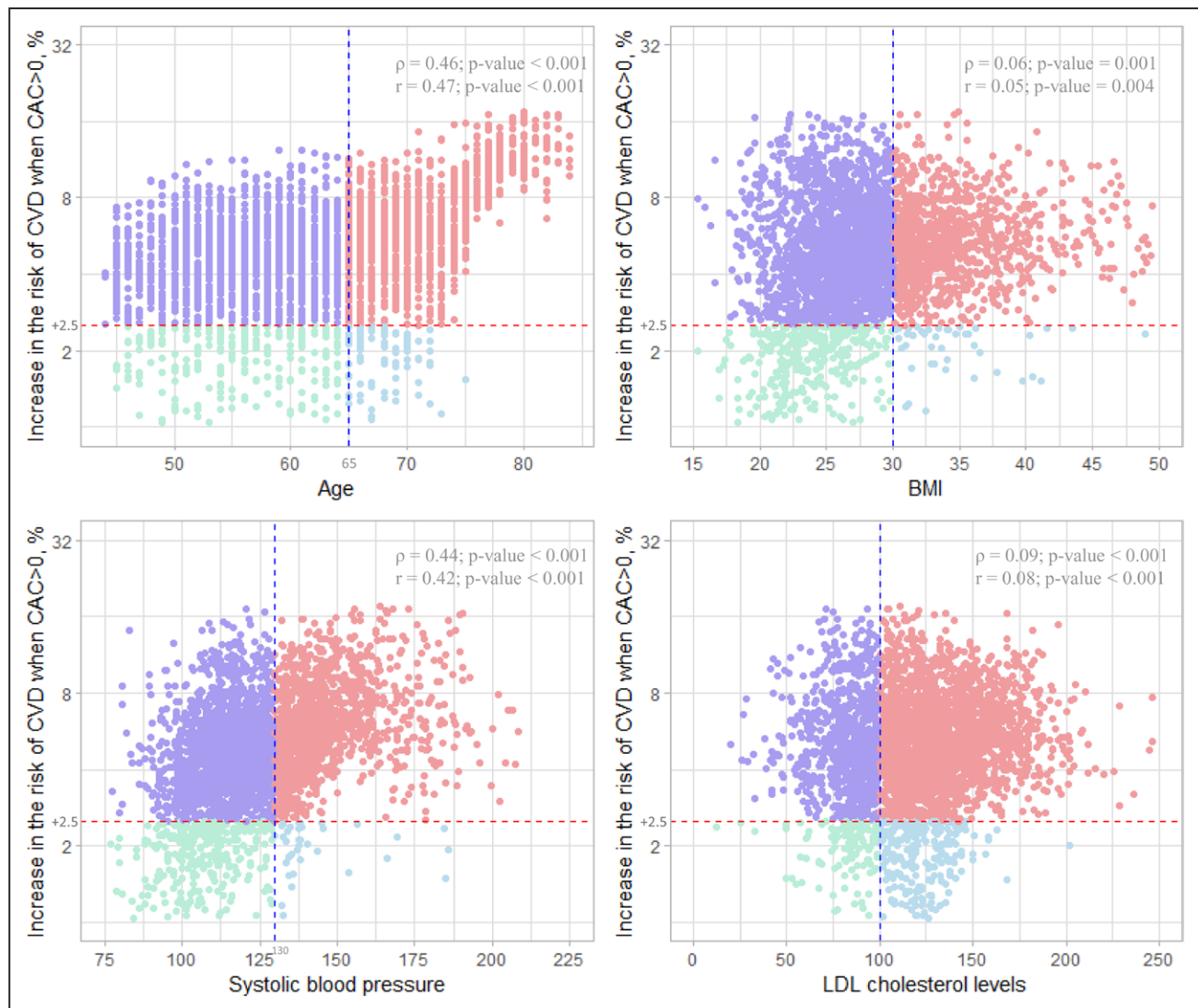
The results did not change when we used the entire cohort before propensity score matching (Figure S8), and restricted individuals to those without statin use (Figure S9).

## DISCUSSION

In this longitudinal cohort study of the multiethnic population, we found that the expected increase in CVD risk when CAC > 0 was heterogeneous across individuals. The 10-year ASCVD risk was associated in a dose-response manner with the expected increase in hard CVD risk when CAC > 0. Even among individuals with low ASCVD risk, nearly 70% of them showed a  $\geq 2.5\%$  increase in CVD risk when CAC > 0. They were more likely to be male, Hispanic, and had unfavorable CVD profiles. These results provide empirical evidence for the potential utility of targeted CAC measurements for such subpopulations with a large increase in CVD risk related to positive CAC even when their CVD risk is low. They also highlight

the importance to consider the individual-level burden of positive CAC on CVD as well as the absolute CVD risk as a precision medicine approach.

Our findings would advance the ongoing discussion about who should be screened for CAC to prevent future CVD events. An increasing body of literature has shown the potential of CAC to reduce CVD risk in precision medicine through facilitating shared decision-making between physicians and patients and promoting long-term adherence to diet, exercise, statin, and aspirin therapy.<sup>27,28</sup> Some studies have reported that CAC measurements would improve risk assessment by traditional approaches which can underdetect individuals who could potentially benefit from early preventive interventions.<sup>29,30</sup> The 2019 CVD Primary Prevention Clinical Practice Guidelines<sup>9</sup> assigned CAC a class IIa recommendation as a useful tool when the treatment decision is uncertain only by traditional CVD risk estimation (eg, CAC > 0 could be used to assist with shared decision-making about statin use<sup>31</sup>). This was, at least partially, based on the fact that the benefit from the CAC screening is proportional to the absolute risk of CVD, which was supported by our findings of the dose-response relationship between the increase in CVD risk when CAC > 0 and the absolute 10-year ASCVD risk. Meanwhile, it is also important to note that a substantial number of individuals at low risk of 10-year ASCVD showed a strong relationship between CAC > 0 and incident CVD. These findings indicate that our machine learning-based approach would help us to identify such individuals, and thus facilitate our effort to achieve precision medicine<sup>32</sup> with appropriate and cost-effective resource allocation of CAC measurements and preventive approaches such as statin therapy.<sup>8</sup>



**Figure 3. Association between the traditional cardiovascular risk factors and the estimated increase in the risk of cardiovascular events when CAC > 0 compared to CAC = 0.**

X-axis shows 4 major traditional cardiovascular risk factors (age, BMI, systolic blood pressure, and LDL cholesterol levels). The vertical dashed line (blue) for each variable corresponds to age (65 years), BMI (30 kg/m<sup>2</sup>), systolic blood pressure (130 mmHg), and LDL cholesterol levels (100 mg/dL). Y-axis shows the estimated increase in the risk of cardiovascular events when CAC > 0 compared to CAC = 0 (calculated by the causal forest model). Spearman correlation coefficients ( $\rho$ ) and Pearson correlation coefficient ( $r$ ) between each traditional cardiovascular risk factor and the estimated increase in the risk of cardiovascular events when CAC > 0 (vs CAC = 0) and their  $P$  values are shown at the upper right of each plot. BMI indicates body mass index; CAC, coronary artery calcium; CVD, cardiovascular disease; and LDL, low-density lipoprotein.

Of note, among adults with low ASCVD risk, those with a large increase in CVD risk associated with CAC > 0 were more likely to be Hispanic and have unfavorable CVD risk profiles (ie, higher body mass index, higher blood pressure, higher LDL cholesterol levels, and greater prevalence of diabetes and antihypertensive use) than those with a small increase in CVD risk associated with CAC > 0. This observed pattern might reflect the influence of social determinants of health on the relationship between CAC > 0 and CVD, which are not captured in the ASCVD risk pooled cohorts equation. Thus, the correct identification of such subpopulations would potentially help us to minimize the social disparity in cardiovascular health. In other words, targeting only individuals at a

high absolute risk of CVD for CAC measurements may not be sufficient to achieve the AHA's 2024 Impact Goal of removing barriers to health equity.<sup>33</sup> Further investigations are warranted to (1) disentangle the underlying mechanisms of the observed heterogeneity across social determinants of cardiovascular health and (2) evaluate whether the tailored approach of CAC screening using the individual-level prediction of the increase in CVD risk associated with CAC > 0 as well as their absolute CVD risk would improve cardiovascular health equity.

Our study demonstrated for the first time that the association between CAC > 0 and incident CVD varied across multidimensional characteristics of individuals. Previous studies have investigated the association

**Table 2. Baseline Characteristics of Adults With Low 10-Year ASCVD Risk (<5%) According to Estimated Risk Increase of Hard CVD Events When CAC>0 vs CAC=0**

Variable*	Estimated increase in the risk of hard CVD events when CAC>0		P value†
	Low (<2.5%)	High (≥2.5%)	
Participants, n	277	623	–
Age	54.7 ± 5.7	53.9 ± 5.9	0.08
Sex, %			0.05
Male	32.5	40.0	
Female	67.5	60.0	
Race/ethnicity, %			0.01
White	63.9	54.9	
Black	9.4	11.2	
Hispanic	11.2	20.1	
Asian	15.5	13.8	
Education status, %			0.08
Less than college	19.9	25.8	
College or more	80.1	74.2	
Health insurance, %			0.35
Public	5.4	7.9	
Private	83.8	83.0	
Uninsured	10.8	9.2	
Smoking status, %			0.14
Never	49.1	55.9	
Former	42.6	38.4	
Current	8.3	5.8	
BMI, kg/m <sup>2</sup>	26.3 ± 5.5	28.9 ± 5.9	<0.01
Systolic blood pressure, mmHg	106.7 ± 12.0	117.5 ± 15.5	<0.01
Diastolic blood pressure, mmHg	64.9 ± 7.4	71.2 ± 9.7	<0.01
eGFR, mL/(min·1.73 m <sup>2</sup> )	74.5 ± 10.9	76.5 ± 21.0	0.15
Total cholesterol, mg/dL	190.9 ± 26.7	199.2 ± 35.0	<0.01
HDL cholesterol, mg/dL	59.6 ± 16.9	50.8 ± 14.7	<0.01
LDL cholesterol, mg/dL	111.1 ± 22.8	121.8 ± 31.7	<0.01
Diabetes, %	1.1	3.1	0.11
Antihypertensive use, %	8.3	22.0	<0.01
Statin use, %	13.4	13.4	0.99

BMI indicates body mass index; CAC, coronary artery calcium score; CVD, cardiovascular diseases; eGFR, estimated glomerular filtration rate; HDL, high-density lipoprotein; and LDL, low-density lipoprotein.

\*Mean ± SD was described for continuous variables, otherwise indicated.

†P values for between-group differences. We used the Wilcoxon rank-sum test for continuous variables and the Fisher exact test for categorical variables. P value was adjusted for multiple comparisons using the Benjamini–Hochberg method.

within traditional risk factor strata. Using a registry of 10,377 asymptomatic individuals who were referred by their primary care physicians for CAC screening, Shaw et al<sup>34</sup> reported the association between elevated CAC and 5-year mortality rate among low-, intermediate-, and

high-risk individuals based on Framingham risk score, respectively. Using a prospective cohort study (the South Bay Heart Watch) including 1461 individuals with ≥1 coronary risk factor, Greenland et al<sup>2</sup> also showed the increased risk of composite CVD outcomes related to elevated CAC, even among people with low Framingham risk scores. More recently, based on data from a large cohort of 44 052 asymptomatic individuals in the United States, Nasir et al<sup>35</sup> showed that individuals with no risk factors and elevated CAC had higher all-cause mortality rates than those with ≥3 risk factors and CAC=0 during a median follow-up of 5 years. Subsequently, using MESA, Silverman, et al<sup>10</sup> reported that CAC=0 was associated with a low CVD event rate even among individuals with multiple risk factors, suggesting the clinical usefulness of identifying CAC=0 to avoid unnecessary or costly interventions. However, subgroup analyses in these previous studies only allowed us to investigate the association by the category of single or a few characteristics. In this context, our study provided more detailed and flexible information than the traditional approach by capturing the heterogeneous patterns complicated by their multidimensional combinations including continuous information such as summarized CVD risk score, age, body mass index, systolic blood pressure, and LDL cholesterol levels.

## Limitations

Despite the strength of using MESA—a longitudinal study design with a large sample size of an ethnically diverse population—our study has several limitations. First, although we have applied propensity score matching to balance the distribution of measured covariates including cardiovascular risk factors between participants with CAC>0 and those with CAC=0 and also adjusted for these covariates through causal forest model, our results might have suffered from confounding bias related to unmeasured confounders. Second, our results likely suffered from bias because of differential loss to follow-up for 16% of MESA participants, such that the likelihood of being lost to follow-up was related to positive CAC and incident CVD. Although we set a binary outcome (incident CVD at 10 years) to build a causal forest model, future studies are warranted to develop and apply the causal forest model in the time-to-event data taking account of competing risks. Third, because some of the baseline characteristics were self-reported in MESA, we cannot rule out the possibility of information bias of covariates. Fourth, we used baseline covariates to assure temporality, and therefore, information on the change in therapy after the results of CAC screening was not included. Lastly, because the MESA enrolled people aged ≥45 years at enrollment, our findings are not generalizable or transportable to the younger population.



## Conclusions

Our machine learning–based approach along with propensity score matching showed that the increases in CVD risk associated with positive CAC were heterogeneous across individuals. There was a substantial number of individuals with a larger increase in CVD risk when CAC>0 despite having a low 10-year ASCVD risk. Identifying such subpopulations with severe CVD burden related to positive CAC would facilitate our current effort to maximize the utility and effectiveness of CAC screening to prevent CVD. Future prospective studies are needed to evaluate whether this new optimal targeting strategy for CAC measurements with information not only on the absolute CVD risk but also the expected increase in CVD risk due to positive CAC improves cardiovascular outcomes in the precision medicine era.

## ARTICLE INFORMATION

Received September 24, 2022; accepted October 24, 2022.

### Affiliations

Department of Social Epidemiology, Graduate School of Medicine, Kyoto University, Japan (K.I.); Department of Epidemiology, University of California Los Angeles (UCLA) Fielding School of Public Health (T.E.S.); Department of Medicine, Divisions of Geriatrics (T.E.S.); and Cardiology (T.H., K.E.W.), UCLA David Geffen School of Medicine; and Department of Medicine, Lundquist Institute at Harbor UCLA Medical Center, Torrance (M.J.B.).

### Acknowledgments

The authors thank the other investigators, the staff, and the participants of MESA (Multi-Ethnic Study of Atherosclerosis) for their valuable contributions. A full list of participating MESA investigators and institutions can be found at <http://www.mesa-nhlbi.org>. This paper has been reviewed and approved by the MESA Publications and Presentations Committee.

Drs Inoue and Watson had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Inoue and Watson were responsible for the concept and design. Drs Inoue, Seeman, Horwich, Budoff, and Watson were responsible for the acquisition, analysis, or interpretation of data. Drs Inoue, Seeman, Horwich, Budoff, and Watson drafted the manuscript. Drs Inoue, Seeman, Horwich, Budoff, and Watson were responsible for critical revision of the manuscript for important intellectual content. Drs Inoue and Seeman did the statistical analysis.

### Sources of Funding

This research was supported by the National Heart, Lung, and Blood Institute (contract Nos. 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, and N01-HC-95169) and the National Center for Advancing Translational Sciences (grant nos. UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420). This work was also supported by a fund from the Barbra Streisand University of California Los Angeles Women's Heart Health Program (to T.H., K.W.). Dr Inoue is supported by the Japan Society for the Promotion of Science (21K20900; 22K17392), the Japan Endocrine Society, Meiji Yasuda Life Foundation of Health and Welfare, and the Program for the Development of Next-Generation Leading Scientists with Global Insight sponsored by the Ministry of Education, Culture, Sports, Science and Technology (Japan). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

### Disclosures

None.

## Supplemental Material

Supplemental Methods

Figures S1–S9

Tables S1–S4

## REFERENCES

- Nasir K, Cainzos-Achirica M. Role of coronary artery calcium score in the primary prevention of cardiovascular disease. *BMJ*. 2021;373:n776. doi: 10.1136/bmj.n776
- Greenland P, LaBree L, Azen SP, Doherty TM, Detrano RC. Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals. *JAMA*. 2004;291:210–215. doi: 10.1001/jama.291.2.210
- Taylor AJ, Bindeman J, Feuerstein I, Cao F, Brazaitis M, O'Malley PG. Coronary calcium independently predicts incident premature coronary heart disease over measured cardiovascular risk factors: mean three-year outcomes in the Prospective Army Coronary Calcium (PACC) project. *J Am Coll Cardiol*. 2005;46:807–814. doi: 10.1016/j.jacc.2005.05.049
- Budoff MJ, Shaw LJ, Liu ST, Weinstein SR, Mosler TP, Tseng PH, Flores FR, Callister TQ, Raggi P, Berman DS. Long-term prognosis associated with coronary calcification: observations from a registry of 25,253 patients. *J Am Coll Cardiol*. 2007;49:1860–1870. doi: 10.1016/j.jacc.2006.10.079
- Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, Liu K, Shea S, Szklo M, Bluemke DA, et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N Engl J Med*. 2008;358:1336–1345. doi: 10.1056/NEJMoa072100
- Mitchell JD, Fergestrom N, Gage BF, Paisley R, Moon P, Novak E, Cheezum M, Shaw LJ, Villines TC. Impact of statins on cardiovascular outcomes following coronary artery calcium scoring. *J Am Coll Cardiol*. 2018;72:3233–3242. doi: 10.1016/j.jacc.2018.09.051
- Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, Braun LT, de Ferranti S, Faiella-Tommasino J, Forman DE, et al. 2018 ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2019;73:e285–e350. doi: 10.1016/j.jacc.2018.11.003
- Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, Himmelfarb CD, Khera A, Lloyd-Jones D, McEvoy JW, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2019;74:1376–1414. doi: 10.1016/j.jacc.2019.03.009
- Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 1985;14:32–38. doi: 10.1093/ije/14.1.32
- Silverman MG, Blaha MJ, Krumholz HM, Budoff MJ, Blankstein R, Sibley CT, Agatston A, Blumenthal RS, Nasir K. Impact of coronary artery calcium on coronary heart disease events in individuals at the extremes of traditional risk factor burden: the Multi-Ethnic Study of Atherosclerosis. *Eur Heart J*. 2014;35:2232–2241. doi: 10.1093/eurheartj/ehf508
- Budoff MJ, Young R, Burke G, Jeffrey Carr J, Detrano RC, Folsom AR, Kronmal R, Lima JAC, Liu KJ, McClelland RL, et al. Ten-year association of coronary artery calcium with atherosclerotic cardiovascular disease (ASCVD) events: the Multi-Ethnic Study of Atherosclerosis (MESA). *Eur Heart J*. 2018;39:2401–2408. doi: 10.1093/eurheartj/ehy217
- Powers S, Qian J, Jung K, Schuler A, Shah NH, Hastie T, Tibshirani R. Some methods for heterogeneous treatment effect estimation in high dimensions. *Stat Med*. 2018;37:1767–1787. doi: 10.1002/sim.7623
- Luedtke AR, van der Laan MJ. Optimal individualized treatments in resource-limited settings. *Int J Biostat*. 2016;12:283–303. doi: 10.1515/ijb-2015-0007
- Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc*. 2018;113:1228–1242. doi: 10.1080/01621459.2017.1319839
- Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. *PNAS*. 2016;113:7353–7360. doi: 10.1073/pnas.1510489113
- Nasir K. Overhauling cardiovascular risk prediction in primary prevention: difficult journey worth the destination. *Circ Cardiovasc Qual Outcomes*. 2015;8:466–468. doi: 10.1161/CIRCOUTCOMES.115.002207
- Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, et al; American College of Cardiology/American Heart Association Task Force on

- Practice Guidelines. 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults. *Circulation*. 2014;129(25\_suppl\_2):S1–45. doi: 10.1161/01.cir.0000437738.63853.7a
18. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR, Kronmal R, Liu K, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156:871–881. doi: 10.1093/aje/kwf113
  19. Sengupta PP, Shrestha S, Berthon B, Messas E, Donal E, Tison GH, Min JK, D'hooge J, Voigt J-U, Dudley J, et al. Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): a checklist. *JACC: Cardiovascular Imaging*. 2020;13:2017–2035. doi: 10.1016/j.jcmg.2020.07.015
  20. Carr JJ, Nelson JC, Wong ND, McNitt-Gray M, Arad Y, Jacobs DR, Sidney S, Bild DE, Williams OD, Detrano RC. Calcified coronary artery plaque measurement with cardiac CT in population-based studies: standardized protocol of Multi-Ethnic Study of Atherosclerosis (MESA) and Coronary Artery Risk Development in Young Adults (CARDIA) study. *Radiology*. 2005;234:35–43. doi: 10.1148/radiol.2341040439
  21. Jha HC, Divya A, Prasad J, Mittal A. Plasma circulatory markers in male and female patients with coronary artery disease. *Heart Lung*. 2010;39:296–303. doi: 10.1016/j.hrtlng.2009.10.005
  22. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*. 2009;28:3083–3107. doi: 10.1002/sim.3697
  23. Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W. Double/Debiased/Neyman machine learning of treatment effects. *Amer Econ Rev*. 2017;107:261–265. doi: 10.1257/aer.p20171038
  24. Athey S, Wager S. Estimating treatment effects with causal forests: an application. arXiv. Preprint posted online February 20, 2019. 10.48550/arXiv.1902.07409
  25. van Klaveren D, Steyerberg EW, Serruys PW, Kent DM. The proposed "concordance-statistic for benefit" provided a useful metric when modeling heterogeneous treatment effects. *J Clin Epidemiol*. 2018;94:59–68. doi: 10.1016/j.jclinepi.2017.10.021
  26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995;57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
  27. Nasir K, McClelland RL, Blumenthal RS, Goff DC, Hoffmann U, Psaty BM, Greenland P, Kronmal RA, Budoff MJ. Coronary artery calcium in relation to initiation and continuation of cardiovascular preventive medications: the Multi-Ethnic Study of Atherosclerosis (MESA). *Circ Cardiovasc Qual Outcomes*. 2010;3:228–235. doi: 10.1161/CIRCOUTCOMES.109.893396
  28. Kambhampati S, Ashvetiya T, Stone NJ, Blumenthal RS, Martin SS. Shared decision-making and patient empowerment in preventive cardiology. *Curr Cardiol Rep*. 2016;18:49. doi: 10.1007/s11886-016-0729-6
  29. Michos ED, Nasir K, Braunstein JB, Rumberger JA, Budoff MJ, Post WS, Blumenthal RS. Framingham risk equation underestimates sub-clinical atherosclerosis risk in asymptomatic women. *Atherosclerosis*. 2006;184:201–206. doi: 10.1016/j.atherosclerosis.2005.04.004
  30. Nasir K, Michos ED, Blumenthal RS, Raggi P. Detection of high-risk young adults and women by coronary calcium and National Cholesterol Education Program Panel III guidelines. *J Am Coll Cardiol*. 2005;46:1931–1936. doi: 10.1016/j.jacc.2005.07.052
  31. Barrett B, Ricco J, Wallace M, Kiefer D, Rakel D. Communicating statin evidence to support shared decision-making. *BMC Fam Pract*. 2016;17:41. doi: 10.1186/s12875-016-0436-9
  32. Hodson R. Precision medicine. *Nature*. 2016;537:S49–S49. doi: 10.1038/537S49a
  33. Lloyd-Jones DM, Elkind M, Albert MA. American Heart Association's 2024 impact goal: every person deserves the opportunity for a full, healthy life. *Circulation*. 2021;144:e277–e279. doi: 10.1161/CIRCULATIONAHA.121.057617
  34. Shaw LJ, Raggi P, Schisterman E, Berman DS, Callister TO. Prognostic value of cardiac risk factors and coronary artery calcium screening for all-cause mortality. *Radiology*. 2003;228:826–833. doi: 10.1148/radiol.2283021006
  35. Nasir K, Rubin J, Blaha MJ, Shaw LJ, Blankstein R, Rivera JJ, Khan AN, Berman D, Raggi P, Callister T, et al. Interplay of coronary artery calcification and traditional risk factors for the prediction of all-cause mortality in asymptomatic individuals. *Circ Cardiovasc Imaging*. 2012;5:467–473. doi: 10.1161/CIRCIMAGING.111.964528