ORIGINAL RESEARCH

Detection of Left Ventricular Hypertrophy Using Bayesian Additive Regression Trees: The MESA

Rodney Sparapani, PhD; Noura M. Dabbouseh, MD, MS; David Gutterman, MD; Jun Zhang, PhD; Haiying Chen, MD, PhD; David A. Bluemke, MD, PhD; Joao A. C. Lima, MD; Gregory L. Burke, MD, MS; Elsayed Z. Soliman, MD, MSc, MS

Background—We developed a new left ventricular hypertrophy (LVH) criterion using a machine-learning technique called Bayesian Additive Regression Trees (BART).

Methods and Results—This analysis included 4714 participants from MESA (Multi-Ethnic Study of Atherosclerosis) free of clinically apparent cardiovascular disease at enrollment. We used BART to predict LV mass from ECG and participant characteristics using cardiac magnetic resonance imaging as the standard. Participants were randomly divided into a training set (n=3774) and a validation set (n=940). We compared the diagnostic/prognostic performance of our new BART-LVH criteria with traditional ECG-LVH criteria and cardiac magnetic resonance imaging–LVH. In the validation set, BART-LVH showed the highest sensitivity (29.0%; 95% CI, 18.3%–39.7%), followed by Sokolow-Lyon-LVH (21.7%; 95% CI, 12.0%–31.5%), Peguero–Lo Presti (14.5%; 95% CI, 6.2%–22.8%), Cornell voltage product (10.1%; 95% CI, 3.0%–17.3%), and Cornell voltage (5.8%; 95% CI, 0.3%–11.3%). The specificity was >93% for all criteria. During a median follow-up of 12.3 years, 591 deaths, 492 cardiovascular disease events, and 332 coronary heart disease events were observed. In adjusted Cox models, both BART-LVH and cardiac magnetic resonance imaging–LVH were associated with mortality (hazard ratio [95% CI], 1.88 [1.45–2.44] and 2.21 [1.74–2.81], respectively), cardiovascular disease events (hazard ratio [95% CI], 1.46 [1.08–1.98] and 1.91 [1.46–2.51], respectively), and coronary heart disease events (hazard ratio [95% CI], 1.72 [1.20–2.47] and 1.96 [1.41–2.73], respectively). These associations were stronger than associations observed with traditional ECG-LVH criteria.

Conclusions—Our new BART-LVH criteria have superior diagnostic/prognostic ability to traditional ECG-LVH criteria and similar performance to cardiac magnetic resonance imaging–LVH for predicting events. **(*J Am Heart Assoc*. 2019;8:e009959. DOI: 10.1161/JAHA.118.009959)**

Key Words: ECG • ensemble predictive modeling • left ventricular hypertrophy • nonparametric machine learning

Left ventricular hypertrophy (LVH) is a modifiable risk factor for cardiovascular disease (CVD) and mortality.[1,2] Early detection of LVH can have implications on patient outcomes. Although cardiac imaging provides a more accurate assessment of LVH than the ECG,[1,3,4] data acquisition feasibility and the low cost may make the ECG an ideal tool for routine screening and follow-up of patients at risk for LVH.

There are a large number of ECG-LVH criteria derived from visual inspection of the 12-lead ECG already in use by clinicians and investigators.[5] These ECG-LVH criteria vary in their diagnostic ability.[6–9] However, all of them tend to have low sensitivity and high specificity. Part of the inadequacy of traditional ECG-LVH criteria to detect true LVH may be because of the limited number of data elements used in these criteria. Subtle electrocardiographic

From the Institute for Health and Equity, Division of Biostatistics (R.S.), Cardiovascular Center (R.S., N.M.D., D.G.), Division of Cardiology, Department of Medicine (N.M.D., D.G.), Medical College of Wisconsin, Milwaukee, WI; Department of Electrical Engineering and Computer Science, University of Wisconsin–Milwaukee, Milwaukee, WI (J.Z.); Division of Public Health Sciences, Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston Salem, NC (H.C.); Department of Radiology, School of Medicine and Public Health, University of Wisconsin, Madison, WI (D.A.B.); Division of Cardiology and Department of Radiology, Department of Medicine, Johns Hopkins University, Baltimore, MD (J.A.C.L.); Division of Public Health Sciences, Wake Forest School of Medicine, Winston Salem, NC (G.L.B.); Epidemiological Cardiology Research Center, Department of Epidemiology and Prevention, Wake Forest School of Medicine, Winston Salem, NC (E.Z.S.); and Section on Cardiology, Department of Internal Medicine, Wake Forest School of Medicine, Winston Salem, NC (E.Z.S.).

Accompanying Tables S1 through S4 and Figures S1 through S5 are available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.118.009959

Correspondence to: Rodney Sparapani, PhD, Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Rd, Milwaukee, WI 53226. E-mail: rsparapa@mcw.edu

Received May 30, 2018; accepted November 14, 2018.

Clinical Perspective

What Is New?

- The ECG often serves as an initial screening tool to detect left ventricular hypertrophy despite its low sensitivity to do so.
- Capitalizing on modern digital acquisition of electrocardiographic signals, we present a new criterion derived by nonparametric machine learning that improves left ventricular hypertrophy detection: our new criterion had higher sensitivity than traditional electrocardiographic criteria in predicting left ventricular hypertrophy and stronger associations with clinical outcomes, like mortality, cardiovascular disease, and coronary heart disease.

What Are the Clinical Implications?

- Studies such as this herald a paradigm shift toward empowering clinicians with computer-aided diagnosis and prognosis tool kits: an orchestration of clinical insight, digital data acquisition, and machine learning to reliably detect potentially adverse conditions, like left ventricular hypertrophy.

signal characteristics alone, or combined with individual characteristics, might provide a better signature of LVH and LV mass (LVM) compared with traditional methods.

With the advent of high-frequency acquisition of digital electrocardiographic signals and processes for analyzing and comparing large data sets, it may be possible to develop more accurate LVH criteria using electrocardiographic signals not visible or typically considered before. Therefore, we hypothesized that substantial untapped information is contained in the digital electrocardiographic tracings, which could be combined with nonelectrocardiographic data, allowing for better detection of LVH. The MESA (Multi-Ethnic Study of Atherosclerosis), with its diverse population, high-quality digital ECG, cardiac magnetic resonance imaging (cMRI), and participant data, offers an ideal opportunity to test our hypothesis. Using the MESA data and a nonparametric machine-learning, ensemble predictive modeling technique known as Bayesian Additive Regression Trees (BART),[10] we developed a new LVH criterion and compared it with traditional ECG-LVH criteria. We also examined the prognostic significance of the newly derived LVH as a predictor of all-cause mortality, CVD events, and coronary heart disease (CHD) events.

Methods

The MESA is a population-based, prospective, longitudinal study initiated to elucidate the prevalence, risk factors, and progression of subclinical CVD. The details of MESA have been previously described.[11] From 2000 to 2002, 3214 men and 3600 women, aged 45 to 84 years, who were free of apparent clinical CVD were recruited from 6 sites in the United States: Baltimore, MD; Chicago, IL; Forsyth County, NC; Los Angeles, CA; New York, NY; and St Paul, MN. Institutional review board approval was obtained for each site. Written informed consent was completed by each participant at his or her enrollment. The data, analytic methods, and study materials will not be made available to other researchers for purposes of reproducing the results or replicating the procedure.

For the purpose of this analysis, we only included MESA participants with available 12-lead ECG and cMRI data. Participants with poor technical quality ECG data or with evidence of major intraventricular conduction delay were excluded. This yielded 4714 participants with ECG-cMRI pairings.

Three seated blood pressure measurements were taken 5 minutes apart using an automated device (Dinamap Pro 100; Critikon, Milwaukee, WI). The mean of the last 2 measurements was considered for analysis. Hypertension was defined as systolic blood pressure $\geq$130 mm Hg, diastolic blood pressure $\geq$80 mm Hg, or history of blood pressure–lowering drugs according to the new blood pressure guidelines. Trained technicians measured height, weight, and waist circumference following a standardized protocol. Obesity was defined as body mass index $\geq$30 kg/m^2. Diabetes mellitus was defined as current use of glucose-lowering medications, fasting glucose $\geq$126 mg/dL, or nonfasting glucose $\geq$200 mg/dL.

Electrocardiography and Determination of ECG-LVH

Standard 12-lead ECGs were digitally acquired at baseline using a Marquette MAC-PC electrocardiograph (Marquette Electronics, Milwaukee, WI) at 10 mm/mV calibration and speed of 25 mm/s. The same equipment was used at all sites. All ECGs were centrally read at the Epidemiological Cardiology Research Center located at Wake Forest School of Medicine (Winston Salem, NC). All ECGs were visually inspected for quality and identification of technical errors before being automatically processed with the GE Marquette 12-SL program 2001 version (GE Marquette, Milwaukee, WI). Numerical summaries of amplitudes and durations of the ECG waveforms were automatically measured. There were measurements for each of the 12 leads, yielding a total of 552 amplitude and duration measurements per ECG. This, in addition to 4 global ECG measurements (PR interval, P axis, QRS interval, and QRS axis), totaled 556 continuous ECG variables that we used in the analysis (Table S1). Using some of these amplitudes and durations, we derived the following traditional ECG-LVH criteria for comparison: Cornell voltage (SV3+RaVL >2.8 mV for men and >2.2 mV for women),[12] Cornell voltage product

([RaVL+SV3]×QRS duration ≥244 mV seconds; for both Cornell criteria, 0.6 mV was added to the voltage sum for women), Sokolow-Lyon (SV1+RV5/V6 ≥3.5 mV and/or RaVL ≥1.1 mV),[13] and Peguero–Lo Presti (deepest S wave in any single lead SD+SV4 >2.3 mV for women and >2.8 mV for men).[14] In the MESA, study-specific ECG-LVH criteria were developed (SV1+SV2+RV5 ≥4.2 mV).[6] However, these new criteria were not developed using the standard approach (ie, formulated via a training set and confirmed with a validation set either within MESA itself or in another cohort). Therefore, we opted not to use the MESA ECG-LVH criteria in the main analysis because of concerns of overfitting and lack of generalizability, but we commented on it as an additional analysis.

Prior efforts have been made to estimate LVM from simple electrocardiographic variables and participants characteristics, such as done by Rautaharju et al.[8] To compare estimated LVM by BART (an intermediate step to develop BART-LVH), we also calculated LVM using models by Rautaharju et al.[8]

cMRI and Determination of cMRI-LVH

The cMRI protocol in MESA has been previously described.[15] Briefly, MESA cMRI used fast-gradient echo to obtain cine images of the heart. The LVM was measured as the sum of the myocardial area (the difference between endocardial and epicardial contours) times slice thickness plus image gap in the end-diastolic phase multiplied by the specific gravity of the myocardium (1.05 g/mL).[15] Observed LVM was then determined in all MESA participants. Individual LVM was predicted using the following allometric height and weight equations previously derived from a separate reference MESA subpopulation of 822 men and women without LVH risk factors: predicted LVM=8.17×(height in meters)$^{0.561}$×(weight in kilograms)$^{0.608}$ for men and predicted LVM=6.82×(height in meters)$^{0.561}$×(weight in kilograms)$^{0.608}$ for women. The 95th percentile cutoff value of normalized LVM, defined as observed LVM/predicted LVM, was calculated as 1.31. This cutoff defined cMRI-LVH in our study (ie, participants with normalized LVM ratio ≥1.31 were considered to have LVH).[6]

Outcome Ascertainment

MESA participants were followed up from baseline up to 13 years later. Outcomes included in our analysis were all-cause mortality, fatal/nonfatal CVD events, and fatal/nonfatal CHD events. All events were adjudicated by an independent committee.

Statistical Analyses

To develop a refined LVH criterion, we relied on BART: a modern predictive modeling technique shown to have excellent properties when considering many covariates.[10,16] BART is a state-of-the-art, tree-based bayesian nonparametric machine-learning method. BART has been shown to be equal to, or better than, many competitors in out-of-sample, validation predictive performance while being relatively computationally efficient. For more details, the BART method is more thoroughly described elsewhere.[10] BART supports many types of outcomes: continuous, dichotomous, categorical, and time to event with right censoring.[10,17,18]

We used participant characteristics and the electrocardiographic variables for prediction of observed normalized LVM. We chose normalized LVM, rather than LVH, as our outcome because predictive modeling generally works best when predicting a continuous outcome and then dichotomizing as necessary. For this analysis, we used the BART R package.[18] The total list of variables considered included 565 ECG and participant characteristics: demographics, biometrics, and CVD risk factors, such as blood pressure and body mass index (Table S1). We used variable selection with BART[16] to narrow our focus to important covariates and refit based on the subset.

Using cMRI as our standard of reference for the diagnosis of LVH, we assessed the sensitivity, specificity, positive and negative predictive values, and F1 score for our BART method as well as several traditional ECG-LVH criteria applied to the MESA data set. The F1 score is the harmonic mean of the sensitivity, *a*, and the true positive rate, *b*, in which an F1 score of 0% is the worst possible accuracy and 100% is the best. Thus, the F1 score is a composite measure of diagnostic accuracy for a positive test: F1 score=$2(a^{-1}+b^{-1})^{-1}$ where a=TP/(TP+FN) and b=TP/(TP+FP), TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. The sensitivity, specificity, positive predictive value, negative predictive value, and F1 score for BART-LVH were also assessed in subgroups of the study participants stratified by younger versus older age (below and above the median), sex, race/ethnicity (black versus others), hypertension, obesity, and diabetes mellitus. We examined the correlation between LVM estimated by BART and other criteria to LVM measured by cMRI. We plotted the receiver operating characteristic curve and estimated the corresponding area under the curve for BART and other LVH criteria compared with LVH as determined by cMRI.

The associations between baseline BART-LVH and time-to-event outcomes (all-cause mortality, incident CVD, and incident CHD separately) were examined using Cox proportional hazards models adjusted as follows: model 1, unadjusted; model 2, sociodemographic variables, which included age, sex, race/ethnicity, and income; and model 3, which adjusted for the sociodemographic variables plus body mass index, systolic blood pressure, use of blood pressure–lowering medication, diabetes mellitus status, cigarette smoking pack-years, total cholesterol, and use of lipid-lowering medication.

Similar models were fitted for cMRI-LVH and traditional ECG-LVH criteria for comparison. We explore whether the risk associated with BART-LVH is because of its relationship with cMRI-LVH or because of another risk pathway. Therefore, we create a group variable based on the per-subject agreement between them: cMRI-LVH=BART-LVH=Yes, cMRI-LVH=Yes/BART-LVH=No, cMRI-LVH=No/BART-LVH=Yes, and cMRI-LVH=BART-LVH=No. Models 1 through 3, as described above, are fitted; and the risk profiles of these groups are summarized.

All analyses were performed with SAS 9.4 (SAS Institute Inc, Cary, NC) and R 3.4.0 (R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was determined as a 2-sided $P<0.05$.

Results

This analysis included 4714 participants (aged 61.3±10.1 years, 53.6% women, 38.4% whites, 13.2% American Chinese, 25.8% blacks, 22.6% Hispanic). Participants were randomly divided into 2 groups, a training set (n=3774) and a validation set (n=940). Table 1 shows the characteristics of the participants included in those 2 sets. As shown, the 2 groups were similar in the demographics and CVD risk factors.

We fitted our new BART model to the training data set. The R^2 of normalized LVM for the training (validation) set was 40.7% (26.2%) for all variables. With variable selection, we identified 26 contributing variables. We refitted the model that we call BART-LVH on the basis of these 26 variables, reaching an R^2 for the training (validation) set of 42.1% (26.0%). These variables were age, sex, height, systolic and diastolic blood pressures, plus the following electrocardiographic variables: heart rate, QRS duration, P duration (V4), R amplitude (aVF, aVL, V1, V3, V4, V5), R intrinsicoid deflection (V6), S amplitude (V1, V2, V3), secondary P (prime) duration (V2), STJ amplitude (V1, V2, V4), and T amplitude (I, V1, V4, V6). R amplitude, V1; R amplitude, V5; heart rate; sex; S amplitude, V1; and QRS interval were the top 6 variables with the highest partial R^2 contributing to the BART model (Table S2). In additional analysis, we fitted a submodel that excluded non-ECG participants' characteristics, which we call "BART-LVH ECG only." In Table S3 and Figure S1, we have a comparison of the performance of these 2 BART-LVH models along with a comparison of other LVM/normalized LVM/LVH criteria. As shown, the BART-LVH ECG-only model does not reach the performance of the BART-LVH model in either R^2 or area under the curve: R^2 is 21.5% versus 26.0% and area under the curve is 81.5% versus 82.9%, respectively. Therefore, the results of the BART-LVH model are presented from here on unless otherwise indicated.

Table 1. Baseline Participant Characteristics

Characteristics	Training Sample (n=3774)	Validation Sample (n=940)	*P* Value
Age, mean (SD), y	61.4 (10.1)	61.0 (9.9)	0.235
Women, n (%)	2025 (53.7)	502 (53.4)	0.890
Race/ethnicity, n (%)			
White	1453 (38.5)	355 (37.8)	0.981
Black	970 (25.7)	245 (26.1)	
Chinese	498 (13.2)	126 (13.4)	
Hispanic	853 (22.6)	214 (22.8)	
Diabetes mellitus, n (%)	419 (11.1)	115 (12.2)	0.327
Heart rate, mean (SD), bpm	62.9 (9.3)	63.0 (9.5)	0.608
Body mass index, mean (SD), kg/m^2	27.7 (4.9)	28.0 (5.1)	0.144
cMRI-LVM, mean (SD), g	143.9 (38.4)	144.3 (38.3)	0.757
LVH by MRI, n (%)	271 (7.2)	69 (7.3)	0.866
LVH by Cornell voltage, n (%)	128 (3.4)	28 (3.0)	0.527
LVH by Cornell voltage product, n (%)	242 (6.4)	49 (5.2)	0.171
LVH by Sokolow-Lyon, n (%)	348 (9.2)	66 (7.0)	0.033
LVH by Peguero–Lo Presti, n (%)	287 (7.6)	64 (6.8)	0.405
Systolic BP, mean (SD), mm Hg	125.3 (21.3)	125.3 (20.9)	0.750
Diastolic BP, mean (SD), mm Hg	71.8 (10.4)	71.9 (9.9)	0.574
Blood pressure medication, n (%)	1319 (35.0)	319 (34.0)	0.566
Cigarette pack-years, mean (SD)	10.7 (20.5)	10.0 (19.3)	0.416
Total cholesterol, mean (SD), mg/dL	194.8 (35.4)	193.2 (34.8)	0.165
Lipid-lowering medication, n (%)	597 (15.8)	152 (16.2)	0.767

Pearson's χ^2 test was used for qualitative variables, and Wilcoxon's rank sum test was used for quantitative variables. BP indicates blood pressure; bpm, beats per minute; cMRI, cardiac magnetic resonance imaging; LVH, left ventricular hypertrophy; LVM, left ventricular mass.

For comparison with other electrocardiographic criteria, we chose a cutoff with high specificity of ≈95%. We determined that those individuals with a BART estimated normalized LVM of 1.19 or higher represent LVH in both BART-LVH models; note that 1.19 is the 93rd percentile of estimated normalized LVM. Figure 1 shows the receiver operating characteristic curve for BART-LVH, defined by the estimated normalized LVM compared with LVH by cMRI (area under the curve=82.9%).

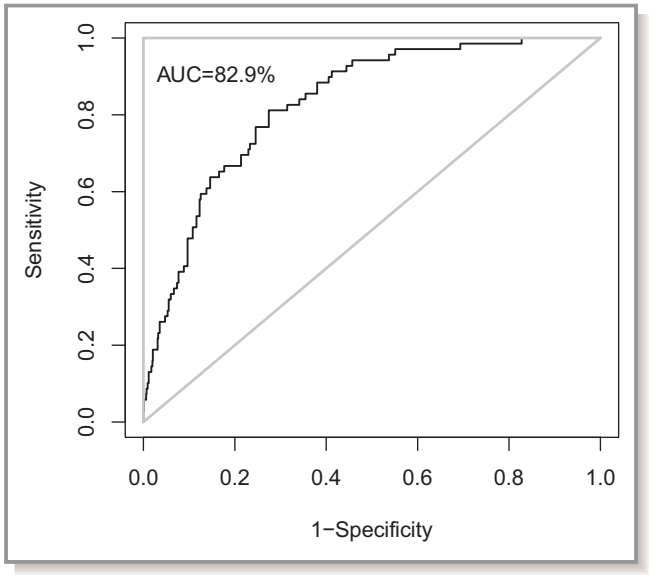


Figure 1. Receiver operating characteristic curve for Bayesian Additive Regression Trees–left ventricular hypertrophy (LVH) compared with LVH by cardiac magnetic resonance imaging in the validation sample (n=940). AUC indicates area under the curve.

Figure 2 and Figure S2 show the correlation between the estimated LVM by BART and the observed cMRI-LVM in the validation and training samples, with R^2 of 66.2% and 73.9%, respectively. As shown in Figures S3 through S5, the R^2 between the estimated LVM by the 3 models of Rautaharju et al[8] and the observed cMRI-LVM in the validation set were much lower: 29.2%, 49.2%, and 51.7%, respectively.

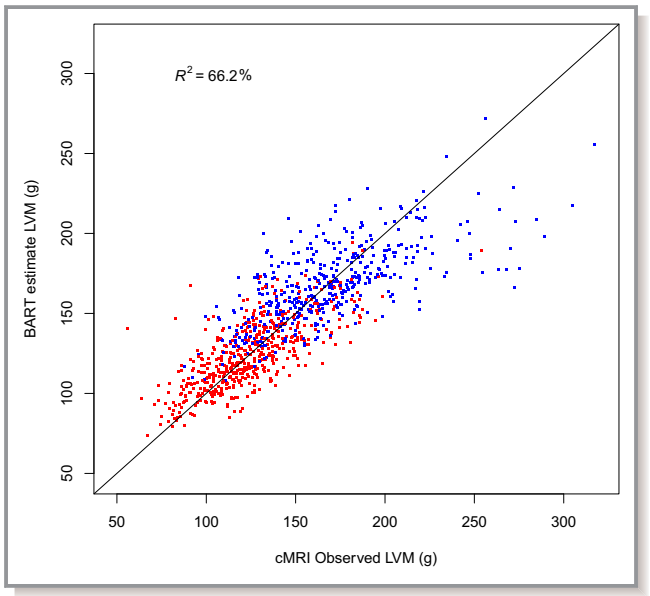


Figure 2. Correlation between left ventricular mass (LVM) estimated by Bayesian Additive Regression Trees (BART) and cardiac magnetic resonance imaging (cMRI) in men and women in the validation sample (n=940). Blue dots indicate men; red dots, women.

BART-LVH showed the highest positive and negative predictive values, followed by Sokolow-Lyon-LVH, Peguero–Lo Presti, Cornell voltage product, and Cornell voltage (Table 2). Similar patterns were observed for the sensitivity and overall F1 score, but Cornell voltage product and Cornell voltage showed a higher specificity than BART-LVH. Differences in sensitivity were statistically compared via McNemar's test among those determined to have LVH by cMRI: *P* values for BART versus Sokolow-Lyon, 0.251; versus Peguero–Lo Presti, 0.0184; versus Cornell voltage product, 0.002; and versus Cornell voltage, 0.0002. BART and Sokolow-Lyon are not statistically different via McNemar's test, but this test depends heavily on sample size. Therefore, we calculated the Cohen's κ measure of agreement with cMRI LVH for both BART and Sokolow-Lyon: BART κ=0.239 and Sokolow-Lyon κ=0.162. Compared with MESA-LVH, which is developed specifically for the MESA cohort, BART-LVH showed slightly inferior sensitivity, but better specificity: MESA-LVH versus BART-LVH sensitivity, 34.8% versus 29.0%, and specificity, 89.9% versus 94.6%, respectively.

In subgroup analysis, the diagnostic performance of BART-LVH was not different for sex, obesity present versus absent, or diabetes mellitus present versus absent. On the other hand, BART-LVH has better sensitivity in those older rather than younger, with hypertension present versus absent and blacks versus nonblacks (Table 3).

During a median follow-up of 12.3 years, 591 deaths, 492 CVD events, and 332 CHD events occurred in the analysis sample (training and validation samples combined). In separate multivariable Cox models using this overall sample (n=4710), both BART-LVH and cMRI-LVH were associated with greater risk of all-cause mortality, CVD events, and CHD events. These associations were stronger than the associations observed with the traditional ECG-LVH criteria. (Table 4). For all-cause mortality, CVD events, and CHD events with similar Cox models described above, we estimated the risk for subjects because of the agreement between BART-LVH and cMRI-LVH. In Table S4, we have the risk profiles of these agreement groups. Generally, cMRI-LVH=BART-LVH=Yes has the highest risk, cMRI-LVH=Yes/BART-LVH=No is second, and cMRI-LVH =No/BART-LVH=Yes is third. This suggests that BART-LVH approximates cMRI-LVH rather than being an alternative path to the outcome.

Discussion

In this report from the MESA, we tested the utility of BART in developing LVH criteria from electrocardiographic and non-electrocardiographic data. Our new BART-LVH criteria showed

Table 2. Diagnostic Performance of ECG LVH Criteria in the Validation Sample (n=940) Compared With cMRI-LVH

LVH Criteria	Sensitivity (95% CI), %	Specificity (95% CI), %	PPV, %	NPV, %	F1 Score, %
BART-LVH	29.0 (18.3–39.7)	94.6 (93.1–96.1)	29.9	94.4	29.4
Sokolow-Lyon	21.7 (12.0–31.5)	94.1 (92.6–95.7)	22.7	93.8	22.2
Peguero–Lo Presti	14.5 (6.2–22.8)	93.8 (92.2–95.4)	15.6	93.3	15.0
Cornell voltage product	10.1 (3.0–17.3)	95.2 (93.8–96.6)	14.3	93.0	11.9
Cornell voltage	5.8 (0.3–11.3)	97.2 (96.2–98.3)	14.3	92.9	8.2

BART indicates Bayesian Additive Regression Trees; cMRI, cardiac magnetic resonance imaging; LVH, left ventricular hypertrophy; NPV, negative predictive value; PPV, positive predictive value.

better diagnostic and prognostic performance than traditional ECG-LVH criteria, such as Cornell voltage, Cornell voltage product, and Sokolow-Lyon, as well as the more recently published Peguero–Lo Presti criteria. Our new LVH criteria also showed similar prognostic performance as a predictor of poor outcomes, similar to cMRI-LVH.

There are at least 38 criteria for diagnosis of LVH from ECG.[5,6,14] Most of these criteria rely on the QRS voltage to diagnose LVH. However, using computer simulations, it has been shown that the mass and shape of the left ventricle, on which most traditional ECG-LVH criteria rely, are not the only determinants of QRS voltage. LVH may manifest itself on ECG as diffuse or regional slowing in conduction velocity attributable to changes in the sequence of ventricular activation, even if the anatomical features of the left ventricle are unchanged.[19,20] These findings provide further support that the traditional electrocardiographic criteria for LVH do not necessarily mirror changes in LVM over time; deficiencies such as this explain the multitude of criteria proposed, none of which provide a high level of diagnostic accuracy.[7,21–24] In

Table 3. Sensitivity, Specificity, and Predictive Values of BART-LVH Compared With the Standard of Reference of LVH by MRI in Subgroups

Subgroups	Participants, n/BART-LVH, n (%)	Sensitivity, (95% CI), %	Specificity, (95% CI), %	PPV, %	NPV, %	F1 Score, %
Median age, y						
≥61	2431/217 (8.9)	51.4 (44.7–58.2)*	95.1 (94.2–96.0)*	49.3	95.4	50.4
<61	2283/123 (5.4)	34.1 (26.0–42.2)*	96.4 (95.6–97.2)*	36.6	96.0	35.3
Sex						
Men	2187/165 (7.5)	46.0 (38.4–53.7)	95.6 (94.7–96.5)	45.5	95.6	45.7
Women	2527/175 (6.9)	43.5 (36.2–50.8)	95.8 (95.0–96.6)	44.0	95.7	43.8
Race						
Blacks	1215/161 (13.3)	52.6 (44.1–61.1)*	91.6 (89.9–93.2)*	43.5	94.0	47.6
Nonblacks	3499/179 (5.1)	39.6 (33.0–46.3)*	97.1 (96.5–97.6)*	45.8	96.2	42.5
Hypertension						
Present	1978/271 (13.7)	52.7 (46.4–59.0)*	91.8 (90.5–93.1)*	47.2	93.3	49.8
Absent	2736/69 (2.5)	24.7 (16.2–33.3)*	98.3 (97.8–98.8)*	34.8	97.3	28.9
Obesity						
Present	1333/94 (7.1)	44.4 (34.7–54.2)	95.9 (94.8–97.0)	46.8	95.6	45.6
Absent	3381/246 (7.3)	44.8 (38.5–51.1)	95.6 (94.9–96.3)	43.9	95.8	44.4
Diabetes mellitus						
Present	534/62 (11.6)	49.3 (37.3–61.2)	93.8 (91.6–96.0)*	53.2	92.8	51.2
Absent	4180/278 (6.7)	43.6 (37.7–49.5)	95.9 (95.3–96.5)	42.8	96.1	43.2

BART indicates Bayesian Additive Regression Trees; LVH, left ventricular hypertrophy; MRI, magnetic resonance imaging; NPV, negative predictive value; PPV, positive predictive value.
*Represents statistically significant differences between subgroup levels.

Table 4. Associations of the BART-LVH and cMRI-LVH With Adverse Outcomes

	Participants, n/Events, n (%)		Model 1*	Model 2†	Model 3‡
LVH Criteria	No LVH	LVH	HR (95% CI)		
All-cause mortality					
cMRI-LVH	4372/496 (11.3)	338/95 (28.1)	2.72 (2.18–3.39)	2.25 (1.79–2.82)	2.21 (1.74–2.81)
BART-LVH	4372/507 (11.6)	338/84 (24.9)	2.33 (1.85–2.94)	1.81 (1.43–2.30)	1.88 (1.45–2.44)
Sokolow-Lyon	4295/518 (12.1)	414/73 (17.6)	1.49 (1.16–1.91)	1.15 (0.90–1.48)	1.12 (0.86–1.44)
Peguero–Lo Presti	4359/533 (12.2)	351/58 (16.5)	1.40 (1.07–1.84)	1.15 (0.88–1.51)	1.09 (0.82–1.44)
Cornell voltage	4554/560 (12.3)	156/31 (19.9)	1.72 (1.20–2.47)	1.33 (0.92–1.93)	1.25 (0.86–1.83)
Cornell voltage product	4419/544 (12.3)	291/47 (16.2)	1.37 (1.02–1.84)	1.08 (0.80–1.47)	1.06 (0.78–1.44)
Incident cardiovascular disease					
cMRI-LVH	4372/421 (9.6)	338/71 (21.0)	2.51 (1.94–3.24)	2.25 (1.73–2.92)	1.91 (1.46–2.51)
BART-LVH	4372/432 (9.9)	338/60 (17.8)	2.01 (1.54–2.64)	1.78 (1.35–2.35)	1.46 (1.08–1.98)
Sokolow-Lyon	4295/432 (10.1)	414/60 (14.5)	1.51 (1.15–1.98)	1.27 (0.96–1.68)	1.14 (0.86–1.52)
Peguero–Lo Presti	4359/441 (10.1)	351/51 (14.5)	1.50 (1.12–2.01)	1.35 (1.01–1.81)	1.16 (0.86–1.57)
Cornell voltage	4554/471 (10.3)	156/21 (13.5)	1.43 (0.92–2.21)	1.43 (0.92–2.24)	1.08 (0.68–1.70)
Cornell voltage product	4419/456 (10.3)	291/36 (12.4)	1.26 (0.90–1.77)	1.16 (0.82–1.64)	0.96 (0.68–1.36)
Incident coronary heart disease					
cMRI-LVH	4372/285 (6.5)	338/47 (13.9)	2.45 (1.80–3.34)	2.25 (1.65–3.09)	1.96 (1.41–2.73)
BART-LVH	4372/288 (6.6)	338/44 (13.0)	2.18 (1.58–2.99)	2.01 (1.45–2.78)	1.72 (1.20–2.47)
Sokolow-Lyon	4295/297 (6.9)	414/35 (8.5)	1.27 (0.89–1.80)	1.08 (0.75–1.53)	0.98 (0.68–1.41)
Peguero–Lo Presti	4359/299 (6.9)	351/33 (9.4)	1.41 (0.99–2.02)	1.29 (0.90–1.85)	1.11 (0.76–1.62)
Cornell voltage	4554/320 (7.0)	156/12 (7.7)	1.18 (0.66–2.10)	1.34 (0.75–2.42)	1.01 (0.56–1.84)
Cornell voltage product	4419/309 (7.0)	291/23 (7.9)	1.16 (0.76–1.78)	1.15 (0.75–1.77)	0.97 (0.62–1.50)

BART indicates Bayesian Additive Regression Trees; cMRI, cardiac magnetic resonance imaging; HR, hazard ratio; LVH, left ventricular hypertrophy.
*Unadjusted.
†Adjusted for age, sex, race/ethnicity, and income.
‡Adjusted for model 2 plus body mass index, diabetes mellitus, systolic blood pressure, use of blood pressure–lowering medications, smoking status, total cholesterol, and use of lipid-lowering medications.

early published data from the Framingham study, the prevalence of ECG-LVH was 3.2% compared with a 16% to 19% prevalence rate when LVH was assessed by echocardiography.[4] The low sensitivity of ECG to detect LVH and the subsequent concern of false-positive LVH has been a challenge in applying efficient screening for LVH using ECG.[21–24] Previous studies estimating LVM from ECG[8] have shown that LVM is dependent on individual subject factors, such as sex, body size, and race/ethnicity; therefore, sound predictive modeling dictates incorporating said information. These relationships were evident in our study, as demonstrated by the better performance of the BART-LVH model, which includes electrocardiographic and nonelectrocardiographic characteristics compared with the BART-LVH ECG-only model. Our proposed approach that uses more information from ECG (and non-ECG), enabled by the availability of digital electrocardiographic data and the modern predictive modeling approach we used, provides the best positive predictive value and a higher F1 score than any current electrocardiographic criteria for LVH. Furthermore, if the machine-learning process is automated, as we have demonstrated here, adding more variables that enhance the model would add more precision with little to no additional effort required.

We observed better sensitivity of BART-LVH with those older, those with hypertension, and blacks. It is known that the prevalence of the disease under investigation affects the predictive value of any test. This means that the same diagnostic test could have a different predictive accuracy according to the clinical setting in which it is applied. Because LVH is more prevalent in those who are older and those with hypertension, it is not unexpected for BART-LVH to show better diagnostic performance in these subgroups of the populations. Furthermore, far more blacks experienced hypertension in this study, 57.0%, than nonblacks, 36.8%; therefore, the higher sensitivity for blacks is likely related to their prevalence of hypertension.

Despite the low sensitivity of ECG to detect LVH, ECG-LVH has been shown to be associated with greater risk of poor CVD outcomes, and its regression reverses the risk.[25] Interestingly, LVH detected by ECG has been shown to be predictive of CVD outcomes in a similar manner to LVH detected by imaging.[23,26–28] These findings, along with wide availability and low cost, have made the ECG an ideal tool for initial evaluation of patients with hypertension to detect LVH.[5] Because of the known performance of ECG-LVH to predict poor outcomes better than their ability to detect anatomical features (ie, diagnose LVH), it has been suggested that risk stratification and prediction should be the primary use for ECG-LVH criteria.[21] Even the current electrocardiographic interpretation guidelines recommend developing new ECG-LVH criteria for prediction.[5] We showed that our new LVH is predictive of outcomes in a similar manner to LVH by cMRI and is better than the traditional ECG-LVH criteria.

It could be argued that machine learning is too complex to be used directly by clinicians to assess LVH, similar to traditional ECG-LVH. Nevertheless, with the wide use of digital electrocardiographic machines, it is feasible to incorporate machine-learning algorithms into contemporary digital electrocardiographic machines to produce automated interpretations of LVH using this approach. Also, it is understandable that digital electrocardiographic systems do not typically include all of the health information used in the BART LVH calculation, making its derivation by current digital electrocardiographic systems a challenge. However, in years to come, it may well be that automated electrocardiographic machines will import patient characteristics, such as blood pressure and body mass index, from the health information system to good effect. Furthermore, for example, suppose that systolic/diastolic blood pressure is not immediately available, then nominal values, such as 120/80 mm Hg, could be used in the interim. By developing automated models to assess LVH, it is hoped the diagnosis of LVH will be made in a less time-consuming/costly manner and will be picked up sooner in a patient's clinical course, thus leading to improved outcomes via earlier detection and treatment. The success of this approach in detecting LVH and predicting outcomes, as we have shown, opens the door for using the same approach in predicting several types of CVD outcomes.

Our study had some limitations. We compared our newly developed LVH with only a few of the traditional ECG-LVH criteria. However, we used the most common criteria, including those with sex-specific cutoffs (Cornell voltage), a mix of QRS duration and amplitudes (Cornell voltage product), and a mix of limb and chest leads (Sokolow-Lyon and Cornell voltage). All of these criteria have shown good diagnostic performance in multiethnic settings compared with other LVH criteria and high prognostic significance as a predictor for CVD events,[6] and some of them are invariant to obesity.[29] Furthermore, the current recommendations for the use of electrocardiographic criteria for detection of cardiac chamber enlargement[5] do not favor, or recommend, one set of LVH criteria over others. Therefore, using any criteria should serve the purpose in accord with these recommendations, and we used several, including a recently developed ECG-LVH.[14] Because BART-LVH includes electrocardiographic and nonelectrocardiographic data, it could be argued that the comparison between BART-LVH and traditional ECG-LVH would not be fair. Nevertheless, our comparison of BART-LVH with traditional ECG-LVH is in the context of comparing nonimaging LVH methods with each other. Also, Cornell voltage includes nonelectrocardiographic data, such as sex, in its definition, and we supplemented our results by using the LVM models of Rautaharju et al,[8] which include sex, race, and weight. Another limitation is that MESA participants were free of apparently clinical CVD at baseline. Hence, our results may not be generalizable to those with CVD.

Despite these limitations, we presented new LVH criteria via a novel approach extended to predict CVD outcomes. Our analysis builds on the strengths of the MESA, including its large sample size and community-based racially diverse population with well-ascertained variables and outcomes.

Conclusions

We developed a novel non–imaging-based model for LVM using BART. We showed that our method generates LVH criteria that have better diagnostic and prognostic performance than current traditional LVH criteria. Further study is needed to determine whether our criteria are broadly applicable to different populations and for detection and prediction of other outcomes.

Acknowledgments

The authors thank the other investigators, the staff, and the participants of the 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.

Sources of Funding

This research was supported by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, and N01-HC-95169 from the National Heart, Lung, and Blood Institute; and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences. Sparapani and Zhang were supported

by the Cardiovascular Center through the Research and Education Program Fund, a component of the Advancing a Healthier Wisconsin Endowment.

Disclosures

None.

References

1. Gradman AH, Alfayoumi F. From left ventricular hypertrophy to congestive heart failure: management of hypertensive heart disease. *Prog Cardiovasc Dis*. 2006;48:326–341.
2. Bauml MA, Underwood DA. Left ventricular hypertrophy: an overlooked cardiovascular risk factor. *Clevel Clin J Med*. 2010;77:381–387.
3. Bottini PB, Carr AA, Prisant LM, Flickinger FW, Allison JD, Gottdiener JS. Magnetic resonance imaging compared to echocardiography to assess left ventricular mass in the hypertensive patient. *Am J Hypertens*. 1995;8:221–228.
4. Levy D, Anderson KM, Savage DD, Kannel WB, Christiansen JC, Castelli WP. Echocardiographically detected left ventricular hypertrophy: prevalence and risk factors: the Framingham heart study. *Ann Intern Med*. 1988;108:7–13.
5. Hancock EW, Deal BJ, Mirvis DM, Okin P, Kligfield P, Gettes LS. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part V: electrocardiogram changes associated with cardiac chamber hypertrophy: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; The American College of Cardiology Foundation and the Heart Rhythm Society. *Circulation*. 2009;119:e251–e261.
6. Jain A, Tandri H, Dalal D, Chahal H, Soliman EZ, Prineas RJ, Folsom AR, Lima JA, Bluemke DA. Diagnostic and prognostic utility of electrocardiography for left ventricular hypertrophy defined by magnetic resonance imaging in relationship to ethnicity: the Multi-Ethnic Study of Atherosclerosis (MESA). *Am Heart J*. 2010;159:652–658.
7. Levy D, Garrison RJ, Savage DD, Kannel WB, Castelli WP. Prognostic implications of echocardiographically determined left ventricular mass in the Framingham Heart Study. *N Engl J Med*. 1990;322:1561–1566.
8. Rautaharju PM, Park LP, Gottdiener JS, Siscovick D, Boineau R, Smith V, Powe NR. Race- and sex-specific ECG models for left ventricular mass in older populations: factors influencing overestimation of left ventricular hypertrophy prevalence by ECG criteria in African-Americans. *J Electrocardiol*. 2000;33:205–218.
9. Wolf HK, Burggraf QW, Cuddy E, Milliken JA, Rautaharju P, Smith ER, Warren JW. Prediction of left ventricular mass from the electrocardiogram. *J Electrocardiol*. 1991;24:121–127.
10. Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. *Ann Appl Stat*. 2010;4:266–298.
11. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacobs DR Jr, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156:871–881.
12. Casale P, Devereux R, Kligfield P, Eisenberg RR, Miller DH, Chaudhary BS, Phillips MC. Electrocardiographic detection of left ventricular hypertrophy: development and prospective validation of improved criteria. *J Am Coll Cardiol*. 1985;6:572–580.
13. Sokolow M, Lyon TP. The ventricular complex in left ventricular hypertrophy as obtained by unipolar and limb leads. *Am Heart J*. 1949;37:161–186.
14. Peguero JG, Lo Presti S, Perez J, Issa O, Brenes JC, Tolentino A. Electrocardiographic criteria for the diagnosis of left ventricular hypertrophy. *J Am Coll Cardiol*. 2017;69:1694–1703.
15. Natori S, Lai S, Finn JP, Gomes AS, Hundley WG, Jerosch-Herold M, Pearson G, Sinha S, Arai A, Lima JA, Bluemke DA. Cardiovascular function in multi-ethnic study of atherosclerosis: normal values by age, sex, and ethnicity. *Am J Roentgenol*. 2006;186:S357–S365.
16. Linero A. Bayesian regression trees for high dimensional prediction and variable selection. *J Am Stat Assoc*. 2018;113:626–636.
17. Sparapani RA, Logan BR, McCulloch RE, Laud PW. Nonparametric survival analysis using Bayesian Additive Regression Trees (BART). *Stat Med*. 2016;35:2741–2753.
18. McCulloch RE, Sparapani R, Gramacy R, Spanbauer C, Pratola MT. BART: Bayesian Additive Regression Trees. The Comprehensive R Archive Network. https://cran.r-project.org/package=BART. Published November 18, 2018. Accessed November 19, 2018.
19. Bacharova L, Szathmary V, Kovalcik M, Mateasik A. Effect of changes in left ventricular anatomy and conduction velocity on the QRS voltage and morphology in left ventricular hypertrophy: a model study. *J Electrocardiol*. 2010;43:200–208.
20. Bacharova L, Szathmary V, Mateasik A. QRS complex and ST segment manifestations of ventricular ischemia: the effect of regional slowing of ventricular activation. *J Electrocardiol*. 2013;46:497–504.
21. Rautaharju PM, Soliman EZ. Electrocardiographic left ventricular hypertrophy and the risk of adverse cardiovascular events: a critical appraisal. *J Electrocardiol*. 2014;47:649–654.
22. Okin PM, Roman MJ, Devereux RB, Kligfield P. Electrocardiographic identification of left ventricular hypertrophy: test performance in relation to definition of hypertrophy and presence of obesity. *J Am Coll Cardiol*. 1996;27:124–131.
23. Almahmoud MF, O'Neal WT, Qureshi W, Soliman EZ. Electrocardiographic versus echocardiographic left ventricular hypertrophy in prediction of congestive heart failure in the elderly. *Clin Cardiol*. 2015;38:365–370.
24. Buchner S, Debl K, Haimerl J, Djavidani B, Poschenrieder F, Feuerbach S, Riegger GA, Luchner A. Electrocardiographic diagnosis of left ventricular hypertrophy in aortic valve disease: evaluation of ECG criteria by cardiovascular magnetic resonance. *J Cardiovasc Magn Reson*. 2009;11:1–11.
25. Bang CN, Devereux RB, Okin PM. Regression of electrocardiographic left ventricular hypertrophy or strain is associated with lower incidence of cardiovascular morbidity and mortality in hypertensive patients independent of blood pressure reduction—a LIFE review. *J Electrocardiol*. 2014;47:630–635.
26. O'Neal WT, Almahmoud MF, Qureshi WT, Soliman EZ. Electrocardiographic and echocardiographic left ventricular hypertrophy in the prediction of stroke in the elderly. *J Stroke Cerebrovasc Dis*. 2015;24:1991–1997.
27. Leigh JA, O'Neal WT, Soliman EZ. Electrocardiographic left ventricular hypertrophy as a predictor of cardiovascular disease independent of left ventricular anatomy in subjects aged ≥ 65 years. *Am J Cardiol*. 2016;117:1831–1835.
28. Oseni AO, Qureshi WT, Almahmoud MF, Bertoni AG, Bluemke DA, Hundley WG, Lima JAC, Herrington DM, Soliman EZ. Left ventricular hypertrophy by ECG versus cardiac MRI as a predictor for heart failure. *Heart*. 2016;103:49–54.
29. Abergel E, Tase M, Menard J, Chatellier G. Influence of obesity on the diagnostic value of electrocardiographic criteria for detecting left ventricular hypertrophy. *Am J Cardiol*. 1996;77:739–744.

SUPPLEMENTAL MATERIAL

Table S1. List of variables considered in development of BART-LVH.

Type	Names
Individual variables: 13	Sex, height, weight, race/ethnicity, age, BMI, SBP, DBP, heart rate, PR interval, P axis, QRS interval, QRS axis
Variables replicated in each of the 12 ECG leads: 46 variables X 12 leads = 552 variables	P duration, P amplitude, P area, P intrinsicoid, P' duration, P' amplitude, P' area, P' intrinsicoid, P total area, Q duration, Q amplitude, Q area, Q intrinsicoid, R duration, R amplitude, R area, R intrinsicoid, R' duration, R' amplitude, R' area, R' intrinsicoid, S duration, S amplitude, S area, S intrinsicoid, S' duration, S' amplitude, S' area, S' intrinsicoid, T amplitude, T area, T intrinsicoid, T' amplitude, T' area, T' intrinsicoid, T total area, special T amplitude, STM amplitude, STE amplitude, STJ amplitude, min of STM/STE/STJ amplitude, max of STM/STE/STJ amplitude, QRS area, QRS balance, QRS deflection balance, QRS intrinsicoid

LVH= left ventricular hypertrophy

Table S2. Relative Partial R^2 increase due to last variable added to BART model.

Variable	Relative Partial R^2 increase
R amplitude, V1	8.7%
R amplitude, V5	8.0%
Heart rate	7.1%
Sex	6.8%
S amplitude, V1	6.4%
QRS interval	5.5%
Height	5.1%
S amplitude, V3	5.0%
R amplitude, aVF	4.9%
Systolic blood pressure	4.8%
T amplitude, V1	4.5%
P interval, V4	4.5%
Age	3.8%
T amplitude, V6	3.8%
R intrinsicoid deflection, V6	3.7%
Diastolic blood pressure	3.3%
R amplitude, V3	3.0%
STJ amplitude, V1	2.3%
T amplitude, I	2.1%
S amplitude, V2	1.8%
STJ amplitude, V4	1.6%
R amplitude, aVL	1.3%
T amplitude, V4	1.3%
STJ amplitude, V2	1.1%

R amplitude in V4 and P′ duration in V2 contributed negatively and hence were removed from the list of increasers above

Table S3. Predictive and discriminatory comparison of criteria in the validation set.

Criteria	R^2 with Normalized LVM by cMRI	AUC for LVH by cMRI
BART-LVH	26.0%	82.9%
BART-LVH (ECG only)	21.5%	81.5%
Sokolow-Lyon	12.0%	71.4%
Peguero-Lo Presti	5.6%	66.3%
Cornell Voltage Product	6.5%	65.9%
Cornell Voltage	4.5%	67.9%
Rautaharju LVM model 3: Normalized LVM	5.3%	65.5%
Rautaharju LVM model 2: Normalized LVM	3.1%	62.0%
Rautaharju LVM model 1: Normalized LVM	2.5%	61.6%

Table S4. Associations of the BART-LVH and cMRI-LVH agreement with adverse outcomes.

LVH Criteria	Model1*	Model 2†	Model 3††
	All-cause Mortality		
	HR (95%CI)	**HR (95%CI)**	**HR (95%CI)**
cMRI-LVH=BART-LVH=Y	3.63 (2.73, 4.84)	2.45 (1.82, 3.29)	2.52 (1.84, 3.44)
cMRI-LVH=Y/BART-LVH=N	2.22 (1.62, 3.03)	2.15 (1.55, 2.96)	2.11 (1.52, 2.94)
cMRI-LVH=N/BART-LVH=Y	1.60 (1.12, 2.29)	1.40 (0.97, 2.02)	1.53 (1.05, 2.24)
	Incident Cardiovascular Disease		
	HR (95%CI)	**HR (95%CI)**	**HR (95%CI)**
cMRI-LVH=BART-LVH=Y	3.13 (2.23, 4.41)	2.49 (1.75, 3.53)	2.01 (1.38, 2.92)
cMRI-LVH=Y/BART-LVH=N	2.19 (1.55, 3.10)	2.11 (1.48, 3.02)	1.89 (1.32, 2.72)
cMRI-LVH=N/BART-LVH=Y	1.41 (0.94, 2.13)	1.36 (0.90, 2.07)	1.18 (0.76, 1.82)
	Incident Coronary Heart Disease		
	HR (95%CI)	**HR (95%CI)**	**HR (95%CI)**
cMRI-LVH=BART-LVH=Y	3.39 (2.28, 5.04)	2.88 (1.92, 4.33)	2.44 (1.57, 3.79)
cMRI-LVH=Y/BART-LVH=N	1.82 (1.15, 2.86)	1.83 (1.16, 2.90)	1.69 (1.06, 2.68)
cMRI-LVH=N/BART-LVH=Y	1.46 (0.90, 2.39)	1.45 (0.88, 2.38)	1.31 (0.78, 2.20)

HR (95%CI) = hazard ratio (95% Confidence interval), LVH= left ventricular hypertrophy; cMRI= cardiac magnetic resonance imaging

***** Unadjusted; **†**Adjusted for age, sex, race/ethnicity and income

†† Adjusted for model 2 plus body mass index, diabetes, systolic blood pressure, use of blood pressure lowering medications, smoking status, total cholesterol and use of lipid lowering medications

Figure S1. Receiver operating characteristic curves for LVH criteria compared to LVH by cMRI in the validation sample (n=940).

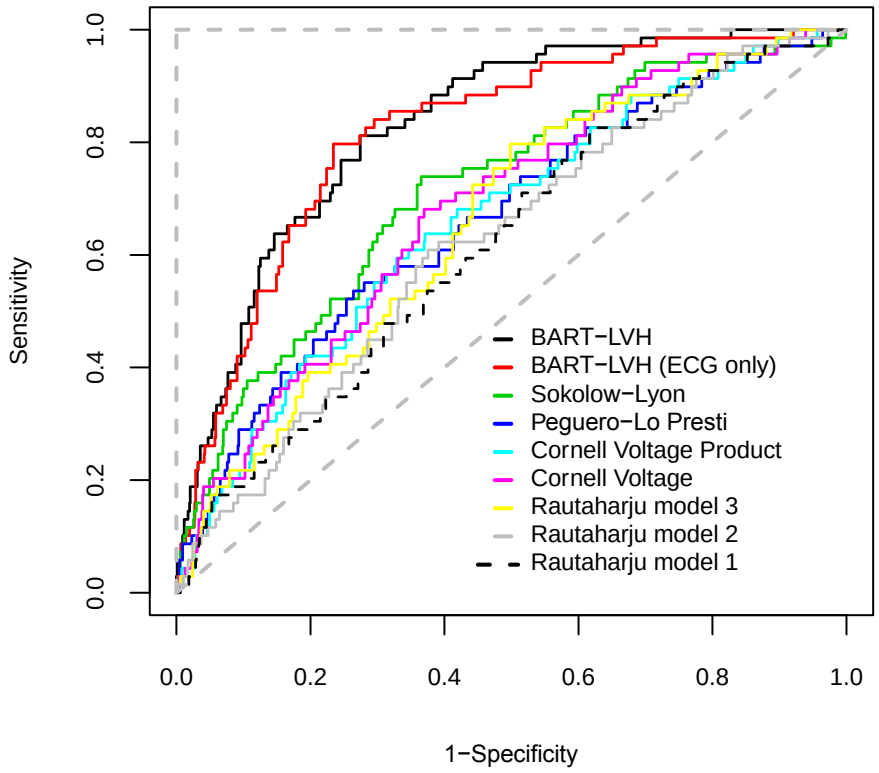
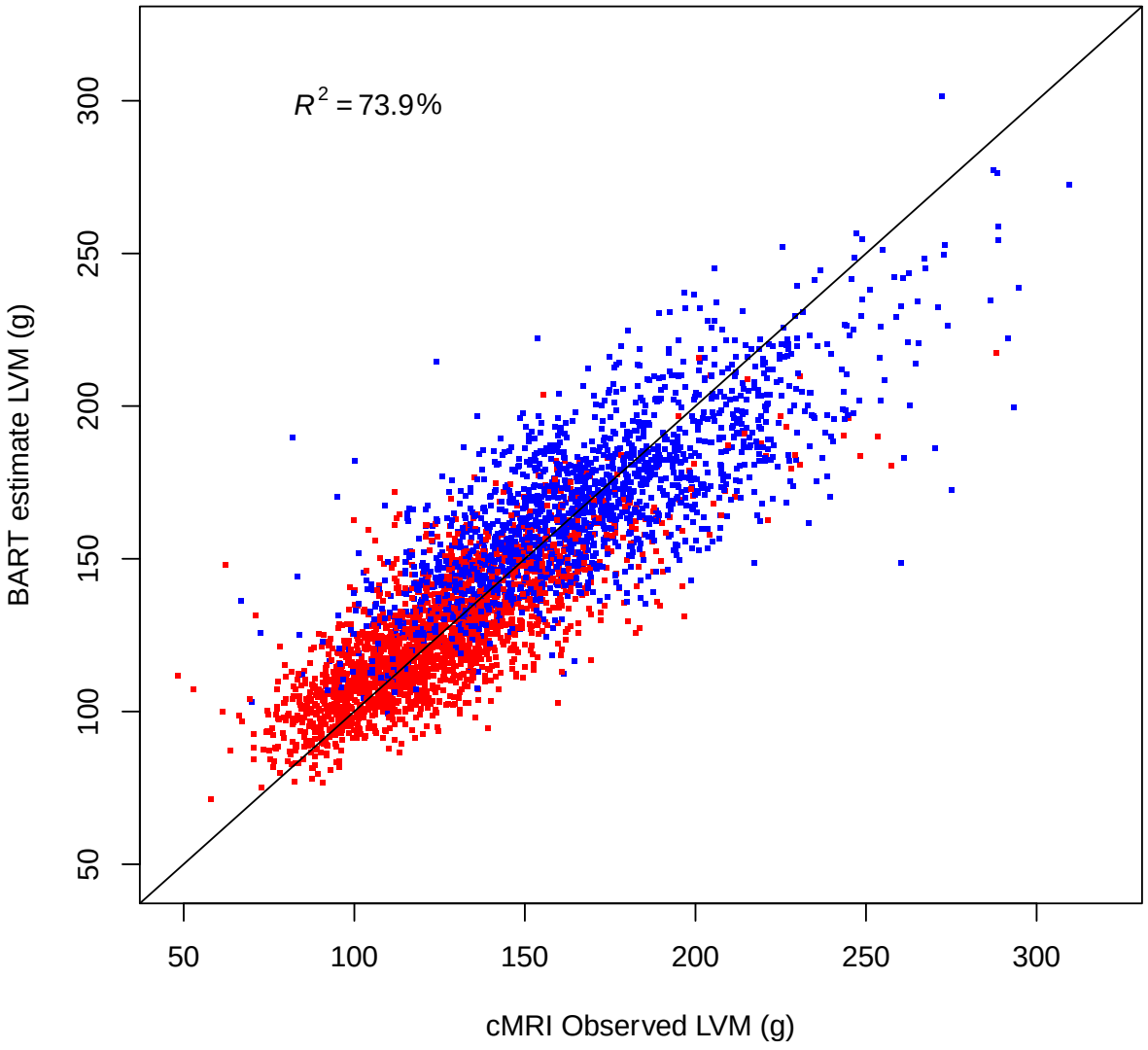
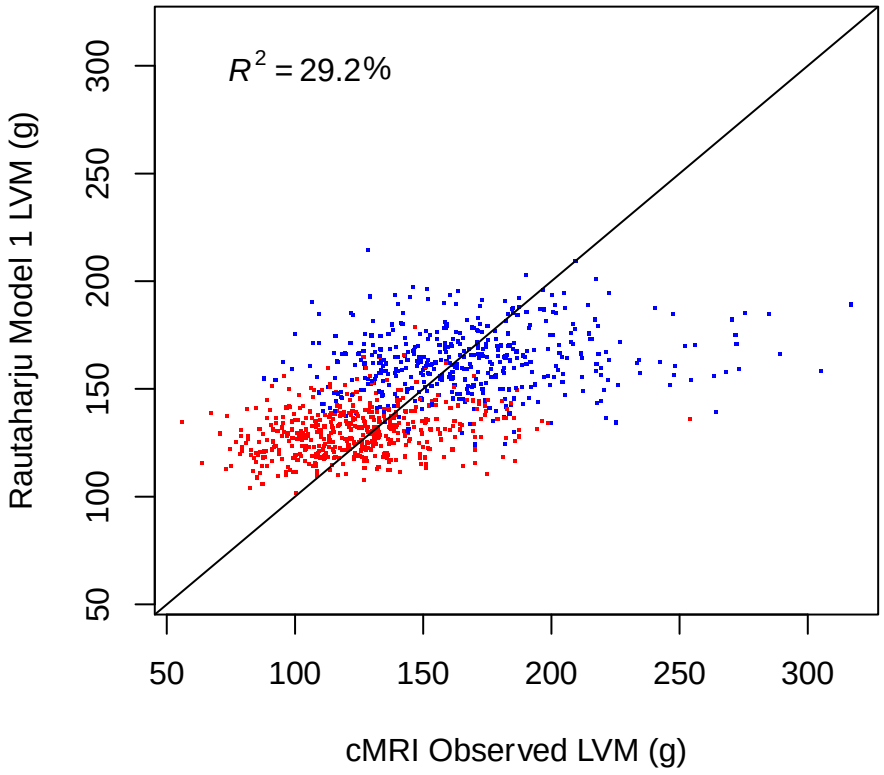


Figure S2. Correlation between left ventricular mass estimated by BART and cMRI in men and women in the training sample (n=3774).



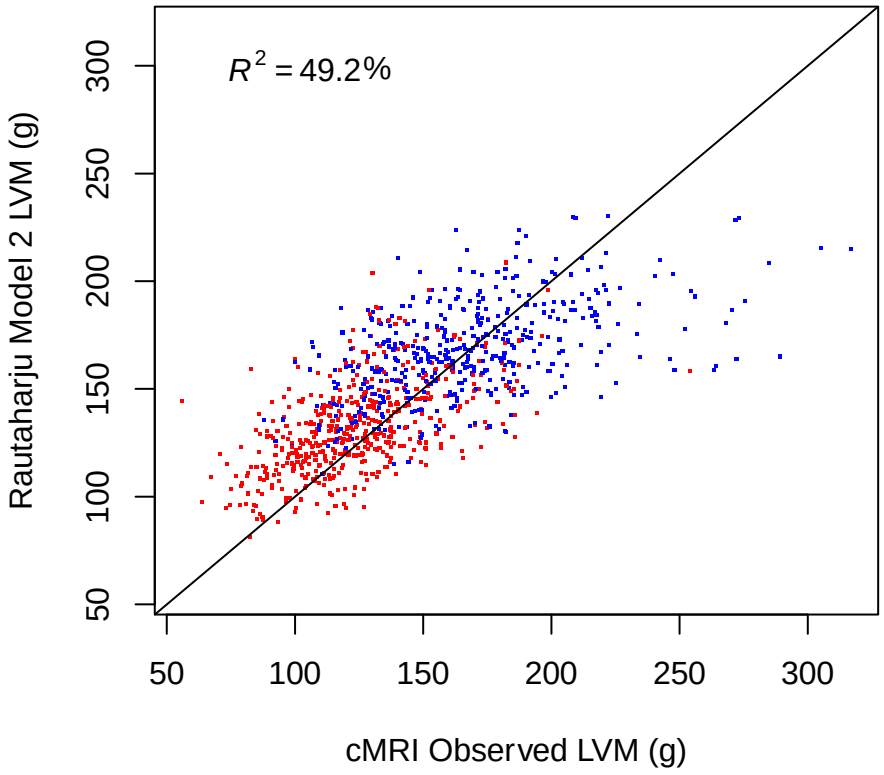
cMRI= cardiac magnetic resonance imaging; blue dots=men, red dots=women

Figure S3. Rautaharju Model 1; Model with Cornell Voltage only.



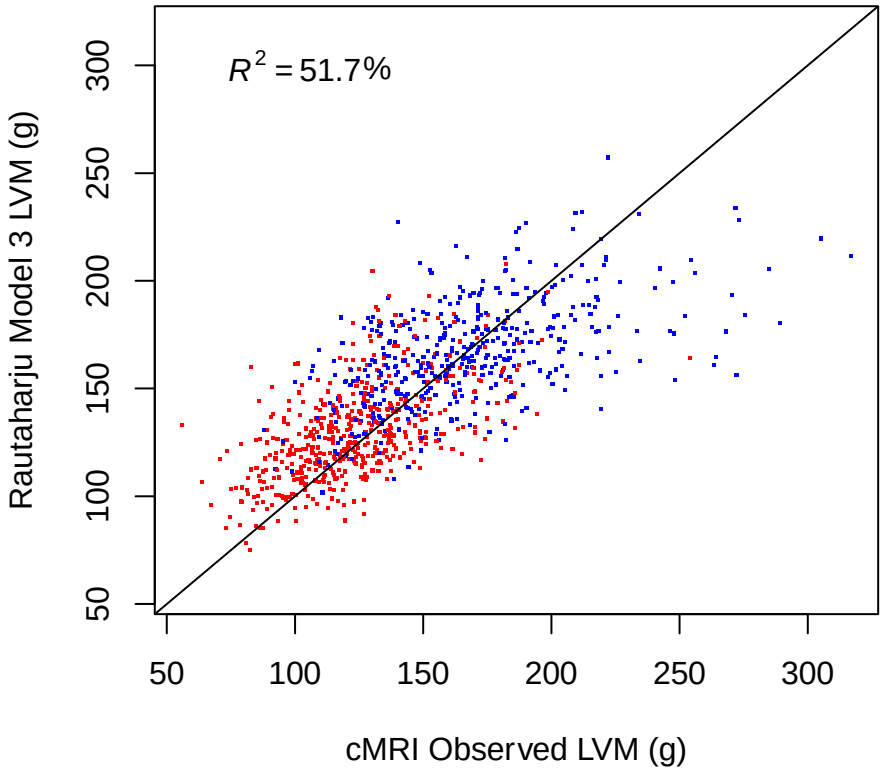
cMRI= cardiac magnetic resonance imaging; CV estimate LVM= Rautaharju Model 1; Model with Cornell Voltage only; blue dots=men, red dots=women

Figure S4. Rautaharju Model 2; Model with Cornell Voltage and weight.



cMRI= cardiac magnetic resonance imaging; CV and weight estimate LVM= Rautaharju Model 2; Model with Cornell Voltage and weight; blue dots=men, red dots=women

Figure S5. Rautaharju Model 3; Model with Cornell Voltage, weight and one ECG variable added.



cMRI= cardiac magnetic resonance imaging; CV, weight and 1 ECG variable estimate LVM= Rautaharju Model 3; Model and 1 ECG variable; blue dots=men, red dots=women