Circulation
Cardiovascular Quality and Outcomes
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Use of Hundreds of Electrocardiographic Biomarkers for Prediction of Mortality in Postmenopausal Women : The Women's Health Initiative

Eiran Z. Gorodeski, Hemant Ishwaran, Udaya B. Kogalur, Eugene H. Blackstone, Eileen Hsich, Zhu-ming Zhang, Mara Z. Vitolins, JoAnn E. Manson, J. David Curb, Lisa W. Martin, Ronald J. Prineas and Michael S. Lauer

Circ Cardiovasc Qual Outcomes 2011;4;521-532; originally published online August 23, 2011;
DOI: 10.1161/CIRCOUTCOMES.110.959023
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514
 Print ISSN: 1941-7705. Online ISSN: 1941-7713

Use of Hundreds of Electrocardiographic Biomarkers for Prediction of Mortality in Postmenopausal Women

The Women's Health Initiative

Eiran Z. Gorodeski, MD, MPH*; Hemant Ishwaran, PhD*; Udaya B. Kogalur, PhD; Eugene H. Blackstone, MD; Eileen Hsich, MD; Zhu-ming Zhang, PhD; Mara Z. Vitolins, DrPH, RD; JoAnn E. Manson, MD, DrPH; J. David Curb, MD; Lisa W. Martin, MD; Ronald J. Prineas, MD, PhD; Michael S. Lauer, MD

Background—Simultaneous contribution of hundreds of electrocardiographic (ECG) biomarkers to prediction of long-term mortality in postmenopausal women with clinically normal resting ECGs is unknown.

Methods and Results—We analyzed ECGs and all-cause mortality in 33 144 women enrolled in the Women's Health Initiative trials who were without baseline cardiovascular disease or cancer and had normal ECGs by Minnesota and Novacode criteria. Four hundred and seventy-seven ECG biomarkers, encompassing global and individual ECG findings, were measured with computer algorithms. During a median follow-up of 8.1 years (range for survivors, 0.5 to 11.2 years), 1229 women died. For analyses, the cohort was randomly split into derivation (n=22 096; deaths, 819) and validation (n=11 048; deaths, 410) subsets. ECG biomarkers and demographic and clinical characteristics were simultaneously analyzed using both traditional Cox regression and random survival forest, a novel algorithmic machine-learning approach. Regression modeling failed to converge. Random survival forest variable selection yielded 20 variables that were independently predictive of long-term mortality, 14 of which were ECG biomarkers related to autonomic tone, atrial conduction, and ventricular depolarization and repolarization.

Conclusions—We identified 14 ECG biomarkers from among hundreds that were associated with long-term prognosis using a novel random forest variable selection methodology. These biomarkers were related to autonomic tone, atrial conduction, ventricular depolarization, and ventricular repolarization. Quantitative ECG biomarkers have prognostic importance and may be markers of subclinical disease in apparently healthy postmenopausal women. **(*Circ Cardiovasc Qual Outcomes*. 2011;4:521-532.)**

Key Words: electrocardiography ■ epidemiology ■ women ■ prognosis

Among postmenopausal women, quantitative electrocardiographic (ECG) biomarkers have a prognostic value.[1–4] Prior studies focused on single ECG measures such as QRS width,[5] small groups of measures such as ventricular repolarization abnormalities,[1,2] and categories of findings such as minor and major ECG abnormalities.[3] Modern digital ECG software is able to abstract hundreds of quantitative measures from a standard 12-lead ECG. To date, there have been no studies exploring the prognostic value of such a large number of ECG measures in a nonparsimonious manner.

Risk stratification based on use of hundreds of quantitative ECG biomarkers presents several unique challenges, which make use of traditional regression methods difficult. First, ECG measures are highly correlated, making their simultaneous use in a regression model problematic. Second, ECG measures may have nonlinear effects that require complex transformations. Third, manual identification of 2-way and 3-way interactions among hundreds of variables is challenging. Fourth, regression models with hundreds of variables may be overfit, consequently performing poorly in testing scenarios. Random forest methodology, a nonparametric decision tree-based approach, has been proposed as a cutting-edge analytic method to address these issues.[6–8] Recently, random forest methodology has been extended to deal with

Received August 26, 2010; accepted June 23, 2011.

From the Heart and Vascular Institute (E.Z.G., E.H.B., E.H.) and Department of Quantitative Health Sciences (H.I., U.B.K.), Cleveland Clinic, Cleveland, OH; Department of Epidemiology, Wake Forest University School of Medicine, Winston-Salem, NC (Z.M.Z., M.V., R.J.P.); Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA (J.E.M.); John A. Burns School of Medicine, Division of Cardiovascular Medicine, University of Hawaii at Manoa, Honolulu, HI (J.D.C.); Division of Cardiology, George Washington University, Washington, DC (L.W.M.); and Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, Bethesda, MD (M.S.L.).

*Drs Gorodeski and Ishwaran are joint first authors.

The online-only Data Supplement is available at http://circoutcomes.ahajournals.org/cgi/content/full/CIRCOUTCOMES.110.959023/DC1.

Correspondence to Michael S. Lauer, MD, Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, National Institutes of Health, Rockledge Center II, Rm 10122, 6701 Rockledge Dr, Bethesda, MD 20892. E-mail lauerm@nhlbi.nih.gov

Circ Cardiovasc Qual Outcomes **is available at http://circoutcomes.ahajournals.org** **DOI: 10.1161/CIRCOUTCOMES.110.959023**

time-to-event data, an approach termed *random survival forests* (RSF).[8]

The objective of the present study was to evaluate the prognostic importance of quantitative ECG biomarkers in postmenopausal women without known cardiovascular disease or cancer who had normal baseline resting ECGs, using a data-rich model. We studied women with normal ECGs because they have been shown to have a lesser risk of mortality than those with major or minor ECG abnormalities.[3] We used RSF methodology to classify women into subgroups of risk and to identify clinical and ECG predictors of mortality. With this approach, numerous decision trees were developed and used to (1) identify the most important predictors (ie, variable selection) and (2) construct risk stratification models.

WHAT IS KNOWN

- Prior studies demonstrated that among postmenopausal women, single ECG measures, or small groups of ECG measures, are prognostic of long-term mortality.
- Simultaneous contribution of hundreds of ECG measures to prediction of mortality in this population has not been studied.

WHAT THE STUDY ADDS

- We used RSFs, a novel machine-learning statistical approach, to demonstrate that among apparently healthy postmenopausal women with clinically normal ECGs, ECG biomarkers related to autonomic tone, atrial conduction, and ventricular depolarization and repolarization have long-term prognostic significance.

Methods

Study Population

The Women's Health Initiative Clinical Trial (www.whiscience.org/about) enrolled 68 132 postmenopausal women (online-only Data Supplement Figure 1) aged 50 to 79 years into randomized trials testing 3 prevention strategies (hormone therapy, dietary modification, or calcium/vitamin D). Eligible women had a choice of enrolling into 1, 2, or all 3 components. At baseline, demographic and clinical characteristics, physical measures, and a standard 12-lead ECG were collected. Exclusion criteria were component specific and were related to competing risks, safety reasons, and adherence or retention reasons.[9]

We focused only on those women who had available a baseline ECG of good quality and without arm lead reversal. We excluded women who had any minor or major ECG abnormalities[3] according to Minnesota[10,11] or Novacode[12] criteria. The remaining 35 774 women had ECGs with sinus rhythm, normal AV conduction, no evidence of old myocardial infarction as suggested by Q waves, normal QRS duration, normal ventricular repolarization, no left atrial enlargement, no right ventricular hypertrophy, no right atrial enlargement, and no fascicular block.

We further excluded 2510 women who had suspected or known cardiovascular disease (history of angina, prior percutaneous coronary intervention, prior coronary artery bypass graft, peripheral arterial disease, prior carotid endarterectomy, aortic aneurysm, or stroke) or a history of cancer (breast, ovarian, colon, cervical, liver, lung, brain, bone, or stomach cancer or leukemia, lymphoma, or Hodgkin disease). Finally, 120 women had missing outcome values and were excluded. The final sample included 33 144 women without known cardiovascular disease or cancer with normal baseline 12-lead ECGs.

ECG Analysis

Standard 12-lead ECGs were recorded at baseline using standardized procedures.[1,3,13] These ECGs were processed at a central laboratory (EPICORE Center; University of Alberta; Edmonton, Alberta, Canada [and later at EPICARE; Wake Forest University; Winston-Salem, NC]) and classified by Minnesota code and Novacode criteria with the use of the Marquette 12-SL program, 2001 version (General Electric; Menomonee Falls, WI).[1,2] Software also abstracted continuous duration and voltage measures by lead for the median beats in each lead, all of which were recorded simultaneously for 10 seconds.

Four hundred and seventy-seven ECG measures abstracted by the Marquette program were studied, encompassing both global and individual ECG measures. Global measures included ventricular rate, median PR duration, median QT duration, median QTc interval, median P-wave axis, median QRS axis, and median T-wave axis. Two measures of ultrashort heart rate variability were studied: SD of the mean value of RR intervals over a 10-second recording (SDNN) and the square root of the mean value of the squares of the differences among all adjacent RR intervals (RMS-SD).

The Marquette program assigned to biphasic (ie, first inflection above or below baseline and second inflection in opposite polarity) P waves and T waves 2 sets of variables, where the second set was termed *prime*, which is different from and not to be confused with the term *prime* used in clinical ECG interpretation, which refers to wave notching.

Individual ECG measures were as follows:

- P-wave measures included P-wave and P′-wave amplitudes, intrinsicoid times (ie, time from onset to peak), durations, and areas in all 12 leads.
- Q-wave measures included Q-wave amplitudes, intrinsicoid times, durations, and areas in all 12 leads.
- R-wave measures included R-wave and R′-wave amplitudes, intrinsicoid times, durations, and areas in all 12 leads.
- S-wave measures included S-wave and S′-wave amplitudes, intrinsicoid times, durations, and areas in all 12 leads.
- QRS complex measures included QRS intrinsicoid times (time from onset of QRS complex to middle of QRS complex) in all 12 leads.
- ST-segment measures included beginning of ST-segment amplitudes (at J point), middle of ST-segment amplitudes (at J+1/16 average RR interval), end of ST-segment amplitudes (at J point+⅛ average RR interval), and ST-segment amplitudes at J point+60 ms in all 12 leads.
- T-wave measures included T-wave and T′-wave amplitudes, intrinsicoid times, and areas in all 12 leads.

Amplitudes were recorded to the nearest 100th of a millivolt and times recorded to the nearest millisecond.

Outcome

All-cause mortality, a clinically relevant and unbiased end point,[14] was recorded centrally by the Women's Health Initiative clinical coordinating center.[15]

Statistical Analysis

Random Survival Forests

RSF analysis[8] used all-cause mortality for the outcome. Candidate predictor variables included all 477 ECG measures described previously in addition to 22 baseline demographic and clinical predictors (Table 1).

Table 1. Baseline Characteristics

Characteristic	Derivation (n=22 096)	Validation (n=11 048)
Age, y	61 (50–79)	61 (50–79)
Ethnicity		
White	18 395 (83)	9172 (83)
Black	1792 (8)	925 (8)
Hispanic	975 (4)	511 (5)
American Indian	94 (0)	31 (0)
Asian/Pacific Islander	541 (2)	270 (2)
Unknown	299 (1)	139 (1)
Smoking		
Never smoked	11 436 (52)	5738 (52)
Past smoker	9018 (41)	4463 (40)
Current smoker	1642 (7)	847 (8)
Hypertension	5715 (26)	2839 (26)
Treated diabetes	628 (3)	344 (3)
Systolic blood pressure, mm Hg	124 (113–135)	124 (113–135)
Diastolic blood pressure, mm Hg	75 (70–81)	75 (70–81)
Body mass index, kg/m^2	27.5 (24.3–31.3)	27.4 (24.4–31.5)
Statin use	1116 (5)	538 (5)
Other antihyperlipidemic medication use	1304 (6)	634 (6)
Aspirin use	4013 (18)	1987 (18)
Bilateral oophorectomy	3370 (17)	1936 (18)
Hysterectomy	8430 (38)	4308 (39)
Waist-to-hip ratio	0.80 (0.76–0.85)	0.80 (0.76–0.85)
Pregnancy		
Never pregnant	1864 (8)	929 (8)
1	1534 (7)	751 (7)
2–4	13 129 (59)	6550 (59)
5+	5569 (25)	2818 (26)
HRT Usage status		
Never used	10 210 (46)	5089 (46)
Past user	3763 (17)	1810 (16)
Current user	8123 (37)	4149 (38)
Income		
<$10,000	807 (4)	373 (3)
$10 000-$19 999	2285 (10)	1206 (11)
$20 000-$34 999	4997 (23)	2446 (22)
$35 000-$49 999	5198 (24)	2575 (23)
$50 000-$74 999	4410 (20)	2233 (20)
$75 000-$99 999	2023 (9)	1012 (9)
$100 000-$149 999	1288 (6)	639 (6)
≥$150 000	623 (3)	285 (3)
Unknown	465 (2)	279 (3)
Alcoholic drinks per week	0.4 (0–2.7)	0.4 (0–2.7)
Marital status		
Never married	908 (4)	463 (4)
Divorced/separated	3490 (16)	1813 (16)
Widowed	3270 (15)	1685 (15)
Presently married/living as married	14 428 (65)	7087 (64)
Medical insurance	20 716 (94)	10 372 (94)
Education		
0–8 y	293 (1)	158 (1)
Some high school	715 (3)	342 (3)
High school diploma/GED	3958 (18)	1871 (17)
School after high school	8714 (39)	4283 (39)
College degree or higher	8416 (38)	4394 (40)

Data are presented as n (%) or median (25th to 75th percentile), except for age, which is presented as median (range). GED indicates general educational development.

Derivation and Validation Subsets

Two thirds of the women were randomly selected for primary analysis (derivation cohort, 22 096; deaths, 819), and the remainder were selected for external validation (validation cohort, 11 048; deaths, 410). When randomly selecting the derivation and validation cohorts, we stratified according to event type (death or censoring) to ensure a similar event rate in both cohorts. The mortality rates in these cohorts were similar (online-only Data Supplement Figure 2).

Forest Analysis

Using the derivation cohort, an RSF of 1000 trees was constructed, with each tree from an independent and unique bootstrap sample of the data (Figure 1A). At each node of the tree, we randomly selected a subset of candidate variables (Figure 1B). For example, the variable occupying the level 0 branch/node was chosen through a "competition" of 22 randomly selected variables; the number of variables randomly selected is the square root of the number of total candidate variables (in this case the square root of 499, which is ≈22). For each of the 22 variables, we split the bootstrap sample into 2 groups, constructed Kaplan-Meier survival curves, and calculated a log-rank statistic. The variable whose split yielded the highest log-rank value "won the competition" and was thus chosen to occupy the node. We split categorical variables according to their natural categories and continuous variables at 10 randomly selected cut points.

For each subsequent node of the tree, we repeated the same process: random selection of candidate variables, splitting of each variable with construction of survival plots and calculation of log-rank statistic, and selection of the best splitting variable. The process continued down each branch of the tree until we reached a unique subset that contained no fewer than 3 deaths,[8] (ie, a terminal node). This approach yielded extensively grown trees having, on average, 143 terminal nodes, where each terminal node included a group of women having similar characteristics and survival outcomes.

Maximal Subtrees for Identification of Predictive Variables

As we have described elsewhere,[16] the most important variables for prediction were identified as those that most frequently split nodes nearest to the trunks of the trees (ie, the root node). Figure 2 demonstrates a random tree with color coding of maximal subtrees. A maximal subtree for a variable v is the largest subtree whose lowest branch is split using v (ie, no other parent branches of the subtree are split using v). There may be no maximal subtree, or there may be several. The shortest distance from the tree trunk to the root of a maximal subtree of v is the minimal depth of v. For example in Figure 2, income splits the tree trunk and has a minimal depth of 0, whereas age occupies the root of 2 yellow subtrees with minimal

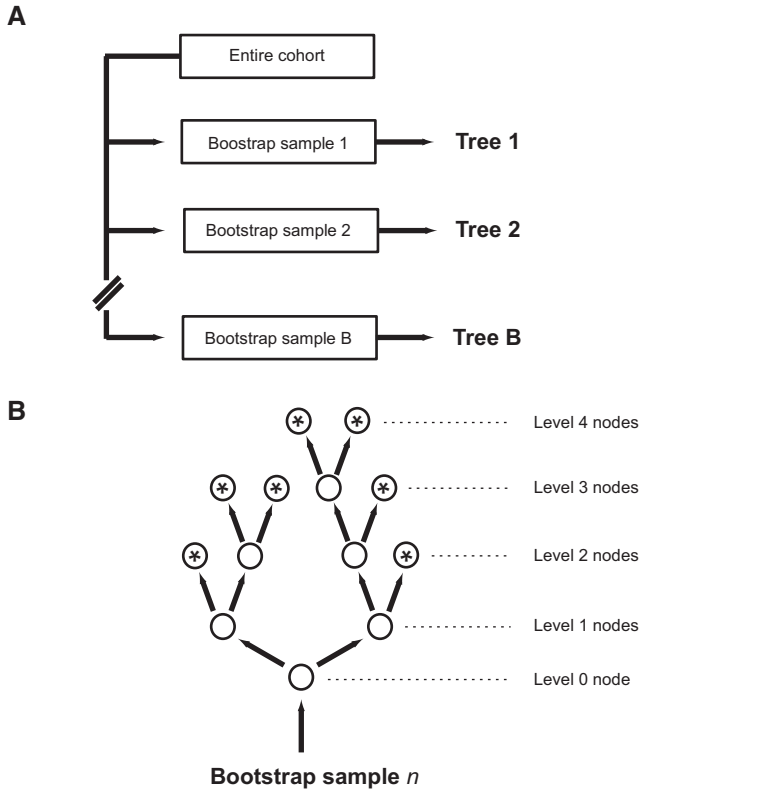


Figure 1. Approach to constructing a random survival forest. **A**, One thousand bootstrap samples of women were derived from the full cohort. **B**, Each sample was then used to construct a unique and independent decision tree.

depths of 3 and 6, respectively. The most predictive variables are those whose minimal depth (averaged over the forest) is smaller than a threshold value determined under the null hypothesis that a variable is unrelated to the survival distribution.[16] For variables like age in which there are >1 maximal subtrees, we used only the lowest value of minimal depth for calculating average minimal depth across the forest. We have previously shown that this variable approach successfully identifies the strongest predictors, with no loss of overall model accuracy because of excessive parsimony.[16]

Construction of Prediction Models

We constructed 8 different prediction models using the derivation cohort: model 1, RSF using all 499 demographic, clinical, and ECG variables; model 2, Cox regression using all 499 variables; model 3, L1-penalized Cox regression using all 499 variables; model 4, Akaike information criterion-penalized Cox regression; model 5, RSF using the 20 variables identified by the maximal subtree algorithm; model 6, Cox regression using the 20 variables identified by the maximal subtree algorithm; model 7, L1-penalized Cox regression using the top-100 RSF variables with lasso parameter selected by 10-fold cross-validation; and model 8, Akaike information criterion-penalized Cox regression using the top-50 RSF variables. The choices of 100 variables for model 7, and 50 variables for model 8, were arbitrary but necessary for these penalized Cox regression methods to converge.

Validation of Prediction Models

Predictive accuracy for all models was assessed using Harrell concordance index both internally (using out-of-bag cross-validation in the derivation cohort) and externally (using the validation cohort). We assessed the individual predictiveness of the top variables identified by the maximal subtrees algorithm by constructing a sequence of nested models and then calculating measures of discrimination (Harrell concordance index) and calibration (continuous ranked probability score,[17] defined as the area under the prediction error curve using the Brier score) for each. Values were calculated using out-of-bag cross-validation. We investigated interactions among our top-20 variables using linkage hierarchical clustering analysis. Specifics regarding methods and results can be found in the online-only Data Supplement.

Missing Data Imputation

Data were missing on 32 of the 499 variables, although very few of these data were missing (maximum amount missing for a variable, 14.3%; average missed per variable, 1.5%). Missing data were imputed using the forest method[8] such that imputed data were not guided by outcomes (ie, survival behavior of patients did not bias imputation).

Computational Methods

Data assembly was performed with SAS version 9.1.3 (SAS Institute Inc; Cary, NC) software. Analyses were performed using R version 2.7.2 (www.r-project.org), using the publically available RSF library[18,19] written by 2 of the authors (H.I., U.B.K.). I1 penalization was performed using the coxnet function in the glmnet library (http://cran.r-project.org/web/packages/glmnet), and Akaike information criterion penalization and fitting was performed using stepAIC from the MASS library (http://cran.r-project.org/web/packages/MASS).

Results

Characteristics and Outcomes

Table 1 shows the baseline characteristics of the derivation and validation cohorts. Global ECG measures are shown in

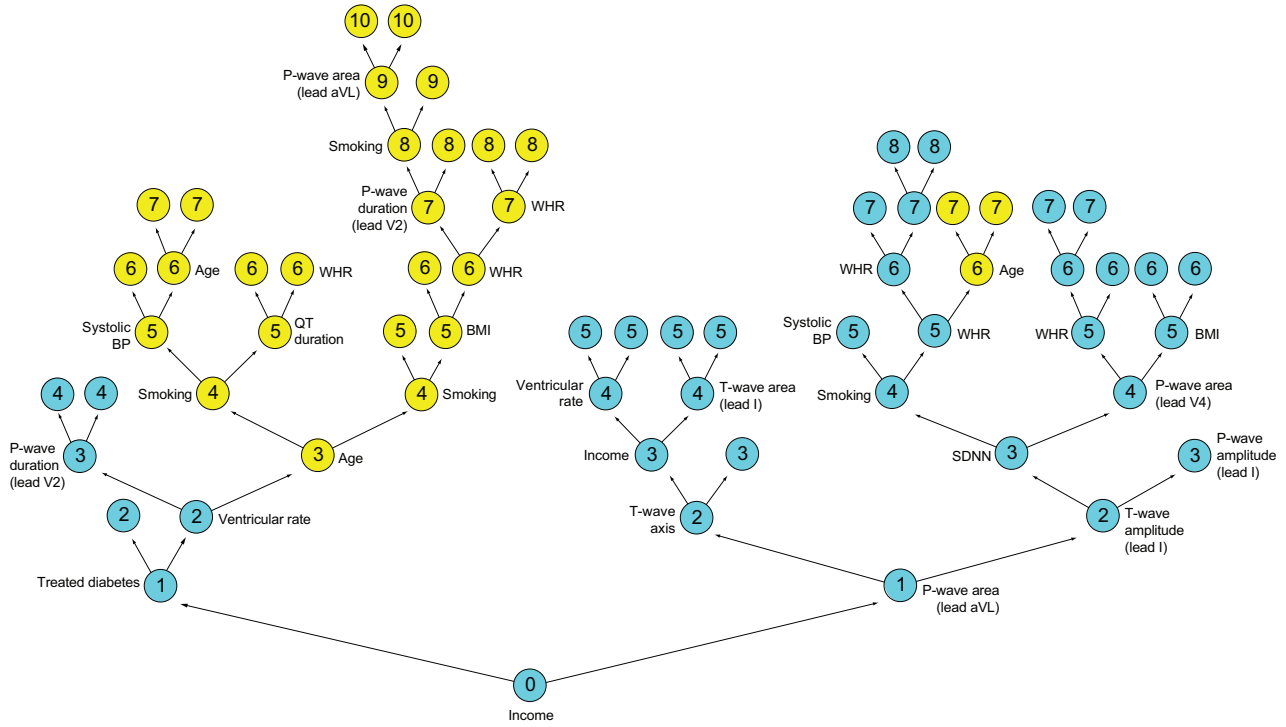


Figure 2. Example of 1 decision tree from the forest. Depth of a branch (node) is indicated by numbers 0 to 10. Highlighted are maximal subtrees (ie, largest subtree whose lowest branch is split using the variable of interest) for the variables income (blue) and age (yellow). Income has 1 maximal subtree at minimal depth 0. Age has 2 maximal subtrees at minimal depths 3 and 6.

Table 2, and all other individual ECG measures are shown in Table 3.

During a median follow-up of 8.1 years (range for survivors, 0.5 to 11.2 years), 1229 (3.7%) women died. Causes of death included cardiovascular diseases (251, 20%), cancer (664, 54%), homicide/suicide (13, 1%), accident/injury (42, 3%), and other/unknown (259, 21%).

Identification of Predictors

In the derivation cohort using all demographic, clinical, and ECG predictors, the 20 variables identified by RSF that were most predictive of long-term all-cause mortality (Figure 3) were the following:

Table 2. Global ECG Measures

	Derivation (n=22 096)	Validation (n=11 048)
Ventricular rate, beats/min	65 (59–71)	65 (59–71)
Median PR duration, ms	158 (144–172)	156 (144–172)
Median QT duration, ms	400 (382–418)	400 (382–418)
Median QTc interval, ms	413 (406–423)	413 (406–423)
Median P-wave axis, °	54 (42–65)	55 (42–65)
Median QRS axis, °	27 (8–48)	27 (8–48)
Median T-wave axis, °	40 (28–51)	40 (28–51)
SDNN, ms	16 (11–25)	17 (11–25)
RMS-SD, ms	17 (11–26)	17 (11–26)

Data are presented as median (25th-75th percentile). ECG indicates electrocardiographic; RMS-SD indicates square root of the mean value of the squares of the differences among all adjacent RR intervals; SDNN, SD of the mean value of RR intervals over a 10-second recording.

- ECG variables representing autonomic tone, including ventricular variability (SDNN, RMS-SD) and ventricular rate.
- ECG variables representing atrial conduction, including P-wave durations (P-wave intrinsicoid duration in leads V3 and V4, P-wave duration in lead V2), P-wave areas (P-wave area in lead V2), P-wave amplitude (P-wave amplitude in lead I), and P-wave axis (median of all leads).
- ECG variables representing ventricular depolarization and repolarization, including QT duration (median of all leads).
- ECG variables representing ventricular repolarization, including T-wave areas (T-wave area in lead I, T-wave area in lead aVL), T-wave amplitude (T-wave amplitude in lead I), and T-wave axis (median in all leads).
- Traditional variables, including age, waist-to-hip ratio, smoking, income, systolic blood pressure, and body mass index.

External Validation

We used the validation subset (n=11 048) to externally validate 8 RSF and Cox prediction models (Table 4). The Cox regression models (models 2 to 4) using all 499 variables did not converge. The RSF and Cox regression models constructed with covariates selected by various variable selection methods demonstrated similar discriminative accuracy in the derivation and validation data sets. Hazard ratios and 95% CIs derived from Cox model (model 6) are shown in online-only Data Supplement Table 1.

We assessed the individual contribution of 20 variables (6 demographic/clinical variables and 14 ECG variables) selected by RSF variable selection method to discrimination (C

Table 3. Lead-Specific ECG Quantitative Measures

	I	II	III	aVL	aVR	aVF	V1	V2	V3	V4	V5	V6
P-wave amplitude, μV												
Q1	63	92	39	−24	−112	63	24	39	48	53	53	48
Q2	78	117	58	−14	−97	87	34	53	63	63	63	58
Q3	92	141	83	53	−83	112	48	73	78	78	73	73
P-wave duration, ms												
Q1	98	98	67	52	98	96	39	80	98	98	98	98
Q2	106	106	98	90	106	104	46	102	106	106	106	106
Q3	114	114	110	106	114	112	55	110	114	114	114	114
P-wave area, μV×ms												
Q1	156	254	60	−34	−317	151	27	80	135	148	148	140
Q2	200	330	132	−10	−271	229	50	122	172	183	181	170
Q3	247	404	212	102	−227	305	76	163	210	221	216	203
P-wave intrinsicoid duration, ms												
Q1	50	44	28	26	46	36	20	26	34	38	44	46
Q2	60	50	40	44	54	46	26	34	42	46	52	54
Q3	66	58	50	64	62	54	32	40	52	58	66	66
P′-wave amplitude, μV												
Q1	0	0	−24	0	0	0	−48	0	0	0	0	0
Q2	0	0	0	0	0	0	−34	0	0	0	0	0
Q3	0	0	0	34	0	0	0	0	0	0	0	0
P′-wave duration, ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	59	0	0	0	0	0
Q3	0	0	27	48	0	0	68	0	0	0	0	0
P′-wave area, μV×ms												
Q1	0	0	−16	0	0	0	−81	0	0	0	0	0
Q2	0	0	0	0	0	0	−51	0	0	0	0	0
Q3	0	0	0	31	0	0	0	0	0	0	0	0
P′-wave intrinsicoid duration, ms												
Q1												
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	64	68	0	0	64	0	0	0	0	0
Q-wave amplitude, μV												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	24	0	0	24	0	0	0	0	0	0	0	34
Q3	53	43	68	63	688	39	0	0	0	0	48	63
Q-wave duration, ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	13	0	0	13	0	0	0	0	0	0	0	15
Q3	18	16	21	19	51	16	0	0	0	0	16	18
Q-wave area, μV×ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	10	0	0	10	483	0	0	0	0	0	0	15
Q3	27	20	44	32	871	18	0	0	0	0	23	33
Q-wave intrinsicoid duration, ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	6	0	0	8	32	0	0	0	0	0	0	8
Q3	10	10	14	12	36	10	0	0	0	0	10	12

(*Continued*)

Table 3. Continued

	I	II	III	aVL	aVR	aVF	V1	V2	V3	V4	V5	V6
R-wave amplitude, μV												
Q1	600	590	73	249	14	219	73	273	551	937	1005	800
Q2	781	771	156	439	34	410	126	424	815	1201	1240	996
Q3	991	976	375	664	63	629	195	629	1123	1503	1503	1215
R-wave duration, ms												
Q1	48	47	20	40	6	39	20	28	40	42	42	48
Q2	63	60	29	55	15	52	24	34	45	47	49	63
Q3	74	75	51	68	20	70	28	40	50	52	59	72
R-wave area, μV×ms												
Q1	673	637	38	264	0	215	44	215	583	974	1040	907
Q2	913	895	116	514	16	458	88	374	882	1271	1327	1161
Q3	1209	1208	411	820	38	766	152	603	1217	1624	1678	1457
R-wave intrinsicoid duration, ms												
Q1	26	28	12	24	8	24	12	18	26	28	28	28
Q2	34	34	23	32	12	32	14	22	30	32	34	36
Q3	38	40	40	40	42	40	18	28	34	36	38	40
R′-wave amplitude, μV												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0
R′-wave duration, ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0
R′-wave area, μV×ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0
R′-wave intrinsicoid duration, ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0
S-wave amplitude, μV												
Q1	0	0	0	0	0	0	527	644	405	190	24	0
Q2	0	19	141	0	590	53	712	874	605	346	131	0
Q3	73	131	415	126	844	175	917	1132	825	527	263	63
S-wave duration, ms												
Q1	0	0	0	0	0	0	52	40	30	23	7	0
Q2	0	7	26	0	40	15	59	48	38	33	27	0
Q3	27	28	51	34	66	34	65	56	45	40	36	25
S-wave area, μV×ms												
Q1	0	0	0	0	0	0	693	676	300	115	9	0
Q2	0	8	92	0	0	25	977	1035	537	267	87	0
Q3	47	96	466	97	943	150	1289	1455	834	472	218	42
S-wave intrinsicoid duration, ms	0	0	0	0	0	0	40	46	52	52	46	0
Q1												
Q2	0	30	38	0	0	44	42	50	54	56	56	0
Q3	58	60	50	54	40	58	46	54	58	60	60	58

(*Continued*)

Table 3. Continued

	I	II	III	aVL	aVR	aVF	V1	V2	V3	V4	V5	V6
S′-wave amplitude, μV												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0
S′-wave duration, ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0
S′-wave area, μV×ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0
S′-wave intrinsicoid duration, ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0
QRS intrinsicoid duration, ms												
Q1	34	36	38	34	36	38	40	42	34	34	34	36
Q2	38	38	44	40	38	42	42	48	38	36	38	38
Q3	40	42	48	44	40	46	46	52	50	40	40	42
ST-segment at J-point amplitude, μV												
Q1	4	4	−15	−5	−35	−5	−20	−10	−15	−10	−5	4
Q2	14	19	4	4	−20	14	−5	14	9	9	9	19
Q3	29	39	24	24	−10	29	9	39	29	29	29	34
Middle ST-segment amplitude, μV												
Q1	4	9	−5	−5	−35	4	14	43	29	14	9	4
Q2	14	24	9	4	−20	14	24	63	48	34	19	14
Q3	29	39	19	14	−10	29	39	92	78	53	39	24
End ST-segment amplitude, μV												
Q1	19	24	−10	4	−59	9	9	73	63	39	29	14
Q2	34	43	9	14	−44	24	29	112	97	68	48	29
Q3	53	63	24	29	−25	43	48	161	141	102	78	48
ST 60 ms after J-point amplitude, μV												
Q1	7	12	−4	−3	−32	4	12	43	31	17	9	4
Q2	17	24	7	4	−21	16	25	67	53	35	23	14
Q3	28	39	19	14	−11	27	40	96	79	56	39	26
T-wave amplitude, μV												
Q1	166	209	−29	48	−297	112	−92	219	273	263	234	180
Q2	219	263	53	92	−244	156	−34	332	380	366	317	239
Q3	278	327	112	141	−200	209	63	458	507	483	415	312
T-wave area, μV×ms												
Q1	930	1198	−68	221	−1654	609	−392	1392	1662	1531	1305	985
Q2	1204	1506	236	465	−1377	872	−101	2036	2268	2065	1734	1294
Q3	1518	1836	607	738	−1127	1185	351	2755	2966	2697	2255	1666
T-wave intrinsicoid duration, ms												
Q1	102	106	72	88	104	104	62	82	94	98	102	104
Q2	114	116	106	108	116	118	100	96	106	112	114	116
Q3	126	128	124	124	128	130	120	110	118	124	126	128

(*Continued*)

Table 3. Continued

	I	II	III	aVL	aVR	aVF	V1	V2	V3	V4	V5	V6
T′-wave amplitude, μV												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0
T′-wave area, μV×ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0
T′-wave intrinsicoid duration, ms												
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	0	0	0
Q3	0	0	0	0	0	0	0	0	0	0	0	0

P-wave intrinsicoid duration indicates time from P onset to peak of P; P′-wave intrinsicoid duration, time from P′ onset to peak of P′, where P′ is a second deflection of the P wave that is opposite in polarity to the original P wave; Q1, 25th percentile; Q2, 50th percentile or median; Q3, 75th percentile; Q-wave intrinsicoid duration, time from Q onset to peak of Q; QRS intrinsicoid duration, time from onset of QRS complex to middle of QRS complex; R-wave intrinsicoid duration, time from Q onset to peak of R; R′-wave intrinsicoid duration, time from Q onset to peak of R′; S-wave intrinsicoid duration, time from Q onset to peak of S; S′-wave intrinsicoid duration, time from Q onset to peak of S′; T-wave intrinsicoid duration, time from end of ST segment to peak of T; T′-wave intrinsicoid duration, time from end of ST segment to peak of T′, where T′ is a second deflection of the T wave that is opposite in polarity to the original T wave.

index) and calibration (continuous ranked probability score) in sequential nested RSF models, where the first model used only age; the second, age and waist-to-hip ratio; the third, age, waist-to-hip ratio, and smoking; and so forth. Figure 4 shows that these performance measures stabilized in the range of 15 to 20 variables, near the size of the model identified by the primary analysis (Figure 3, Table 4).

Discussion

Among 33 144 postmenopausal women without known cardiovascular disease or cancer and with normal resting ECGs by Minnesota and Novacode criteria, we found that 20 variables were independently predictive of long-term mortality, 14 of which were ECG biomarkers representing autonomic tone (ventricular rate and variability), atrial conduction (P-wave durations and areas), ventricular depolarization (QT duration), and ventricular repolarization (T-wave axis, amplitude, and areas). Selected plots demonstrating adjusted predicted survival for an ECG biomarker from each 1 of these 4 categories are shown in Figure 5 (all others are shown in online-only Data Supplement Figure 3). Further, we found that parsimonious prediction models incorporating these ECG measures along with demographic and clinical characteristics selected by an RSF variable selection procedure yielded better predictive accuracy than the nonparsimonious RSF model using all variables (Table 4). Finally, the parsimonious

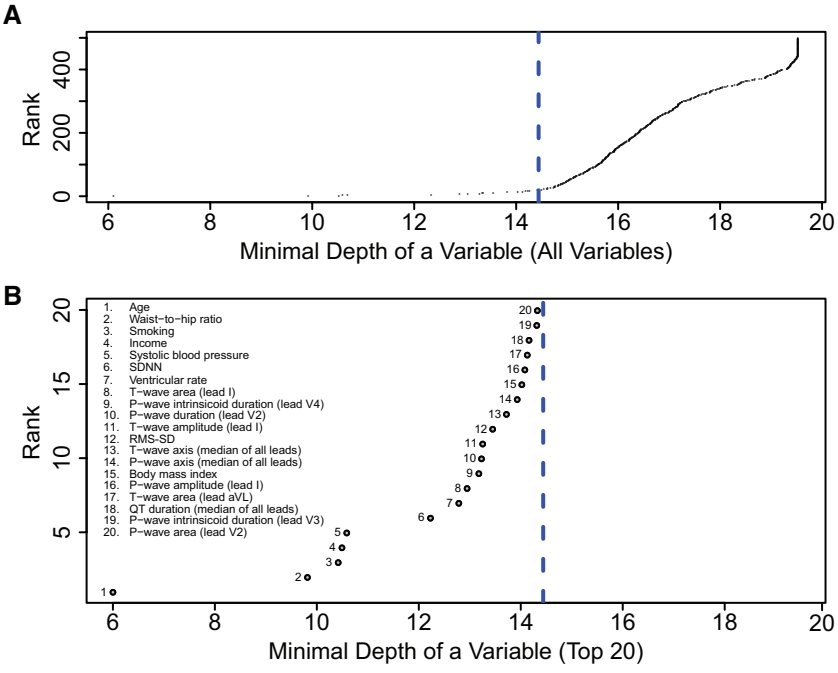


Figure 3. Minimal depth (variable importance). **A**, All variables averaged out from all trees in the forest (1000 trees). **B**, Zoom-in of the top-20 variables. Dashed line is the threshold for filtering variables; variables to left of the line are predictive. On the *y* axis is the ranking of variables, where age is most predictive, then waist-to-hip ratio, and so forth. RMS-SD indicates square root of the mean value of the squares of the differences among all adjacent RR intervals; SDNN, SD of the mean value of RR intervals over a 10-second recording.

Table 4. C Index Values

	No. Variables in Model	Derivation Cohort	Validation Cohort
n	…	22 097	11 048
Deaths, n (%)	…	819 (3.7)	410 (3.7)
Prediction models using all covariates			
Model 1 RSF	499	0.6815	0.6710
Model 2 Cox	499	Did not converge	…
Model 3 L1-penalized Cox	499	Did not converge	…
Model 4 AIC-penalized Cox	499	Did not converge	…
Prediction models using covariates selected by variable selection methods			
Model 5 RSF	20	0.6992	0.6934
Model 6 Cox	20	0.6954	0.6975
Model 7 L1-penalized Cox	59	0.7003	0.6978
Model 8 AIC-penalized Cox	22	0.7005	0.6980

Models 1–4 used all 499 demographic, clinical, and ECG variables available. Models 5 and 6 used 20 variables selected by RSF variable selection method (demographic/clinical: age, waist-to-hip ratio, smoking, income, systolic blood pressure, body mass index; ECG: SDNN, ventricular rate, T-wave area [lead I], P-wave intrinsicoid duration [leads V3, V4], P-wave duration [lead V2], T-wave amplitude [lead I], RMS-SD, T-wave axis, P-wave axis, P-wave amplitude [lead I], T-wave area [lead aVL], QT duration, P-wave area [lead V2]). Model 7 used 59 variables selected by lasso approach from the top-100 RSF variables (demographic/clinical: age, waist-to-hip ratio, smoking, systolic blood pressure, income, body mass index, hypertension, education, diastolic blood pressure, marital status, alcoholic drinks per week, treated diabetes; ECG: SDNN, P-wave intrinsicoid duration [leads I, aVL, V2, V4, V5, V6], ventricular rate, P-wave duration [leads I, aVL, V2, V3, V6], RMS-SD, T-wave axis, P-wave axis, R-wave duration [leads aVF, V1, V4], P-wave area [leads I, V1], QRS intrinsicoid duration [lead I], T-wave intrinsicoid duration [leads I, III, aVL], P-wave amplitude [leads I, aVL, V5], T-wave area [leads aVR, V3], R-wave amplitude [leads II, V1, V5, V6], R-wave intrinsicoid duration [leads II, aVF, V1, V3, V4], P′-wave area [lead V1], R-wave area [leads III, aVL, V3, V6], T-wave amplitude [lead V1], QTc duration, P′-wave amplitude [lead V2]). Model 8 used 22 variables selected by AIC stepwise approach from the top-50 RSF variables (demographic/clinical: age, waist-to-hip ratio, smoking, systolic blood pressure, income, body mass index, hypertension, education, marital status; ECG: ventricular rate, P-wave duration [lead V2], T-wave axis, P-wave axis, R-wave duration [lead V4], P-wave area [lead I], P′-wave intrinsicoid duration [lead aVL], QRS intrinsicoid duration [lead I], T-wave area [leads aVR, aVL], P-wave amplitude [lead I], R-wave intrinsicoid duration [lead aVF], R-wave area [lead V5]). AIC indicates Akaike information criterion; RSF, random survival forest. Other abbreviations as in Table 2.

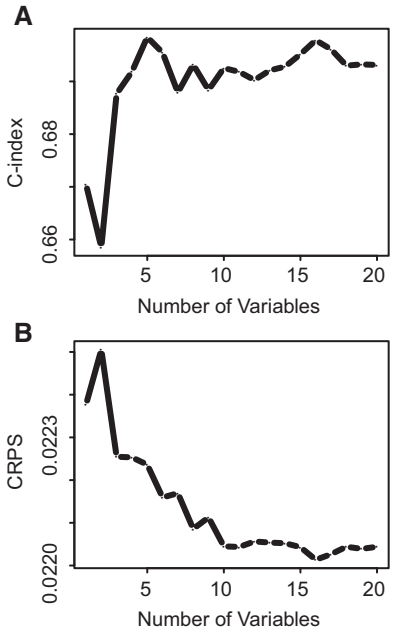


Figure 4. Measures of (**A**) discrimination and (**B**) calibration using validation cohort for nested models, with variables ordered by increasing minimal depth for the top-20 variables. First model included the top variable (age), the second model included top-2 variables (age and waist-to-hip ratio), the third model included top-3 variables (age, waist-to-hip ratio, and smoking), and so forth. CRPS indicates continuous ranked probability score.

RSF model populated by the RSF variable selection procedure was sparser (ie, containing fewer covariates) than parsimonious regression models populated by various other variable selection approaches but performed similarly well in terms of prediction (Table 4). Although other investigators have reported on the predictive utility of ECG findings in women,[1–3] we are, to our knowledge, the first to use an algorithmic approach to simultaneously assess hundreds of digitally measured ECG variables without the bias of preselection.

Use of hundreds of ECG measures for prediction modeling presents a unique challenge. Many of these variables are highly correlated, may have complex interactions that are difficult to detect, and may have nonlinear associations with outcome. Traditional regression and variable selection methods perform poorly under these types of conditions and tend to produce biased results.[6] Our findings confirm these challenges. When we attempted to use standard Cox modeling, we were unable to generate models that converged (Table 4). Additionally, for the penalized Cox regression modeling (model 7 and model 8), it was necessary to restrict the selection of the model variables in an arbitrary manner in order for these methods to converge. To address these challenges, we used RSF methodology both for risk modeling and for variable selection.

Machine learning, the scientific discipline from which RSF methodology is derived, is a field concerned with the design and development of algorithms that allow computers to change behavior based on data.[20] This approach assumes that "nature produces data in a black box whose insides are complex, mysterious, and at least, partly unknowable."[6] As such, instead of attempting to model data from the black box (ie, traditional regression), machine learning is concerned with iterative algorithms, such as RSF, that are intensely focused on prediction.

Unlike classification and regression trees where only a single tree is constructed, RSF uses a large number of survival trees for prediction and variable selection.[8] Growing extensive trees with hundreds of decision branches is a general principle of random forest methodology.[7] Doing so yields trees with low bias (ie, prediction models that better estimate the predictor being estimated). To ensure low variance (ie, amount of variation within predicted results),

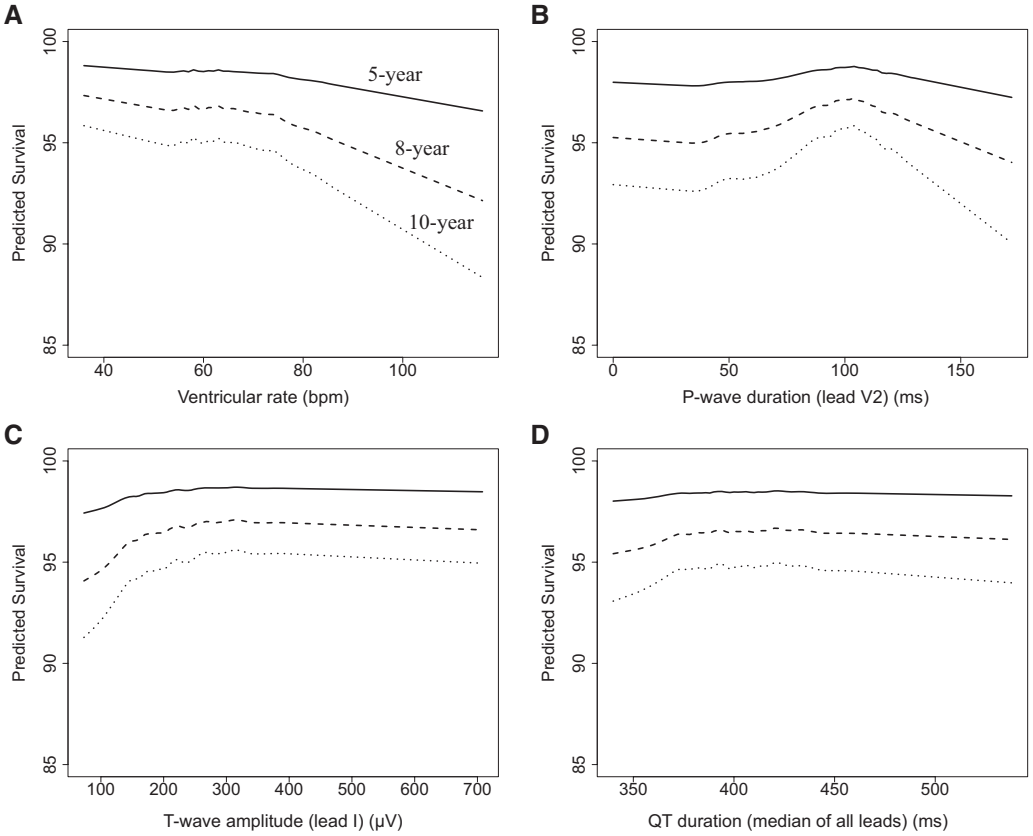


Figure 5. Adjusted predicted survival (%) at 5, 8, and 10 years for ventricular rate (**A**), P-wave duration (lead V2) (**B**), T-wave amplitude (lead I) (**C**), and QT duration (median of all leads) (**D**).

trees must be decorrelated, which is accomplished by introducing 2 forms of randomization when growing a tree (Figure 1). First, trees are grown using independent bootstrap samples of data. Second, each tree is grown by randomly selecting a subset of candidate variables for splitting at each node. Using this 2-stage randomization yields a stable and accurate inference and resolves the instability of classification and regression trees.[21] Random forest has been shown to be accurate and comparable to state-of-the-art predictors, such as bagging,[22] boosting,[23] and support vector machines.[24] Further, random forest has been shown to be highly effective in problems involving large numbers of correlated variables.[7,16,25–27] Examples in the literature include genetics,[28,29] environmental science,[30] and rheumatology.[31]

We believe that RSF analysis may have potential future applications in clinical practice. The RSF prediction model can be stored as an object in the statistical software to be used at a later time on external data sets. This is possible because the random seed chain used to generate the original model is stored. Thus, once a model is generated, it can be used repeatedly on test data sets and will yield identical results if repeated on the same data set. Further, if the original data are used on the restored model, the results will be identical to that of the original analysis. Moreover, this applies even when the training and test data have missing values because we also store the seed chain used to impute missing data values. Thus, when the model is restored, the seed chain used to impute data is reinitialized, and the original forest and its imputation mechanism are reproduced exactly as before. These properties may allow RSF to be used as a prediction tool in clinical settings. It is technologically feasible to create Web-based or even hand-held RSF calculators for use in practice.

The present study has several important limitations. First, the Women's Health Initiative clinical trials enrolled mostly white, highly educated women and, therefore, may have limited generalizability. Second, many of the clinical variables were by self-report, and data regarding standard blood biomarkers were lacking. Lastly, we did not have an external data set (replication cohort) with which to validate our prediction models, although we attempted to do so by setting a portion of our data aside for validation. We are not aware of a similar cohort of postmenopausal women with detailed ECG data to allow such replication and validation. It is possible that several other National Heart, Lung, and Blood Institute cohorts, including the Framingham Heart Study, may soon digitize ECG data and make it available to investigators.

In summary, we found that ECG biomarkers representing autonomic tone, atrial conduction, and ventricular depolarization and repolarization were independently predictive of long-term mortality in postmenopausal women who had no known cardiovascular disease or cancer and had normal ECGs by standard clinical criteria. These findings suggest that further research will be necessary to identify underlying pathophysiological mechanisms and potential therapeutic implications. Additionally, we introduced RSF, a machine-

learning approach to data analysis, which may be of utility in other complex data problems in cardiovascular medicine.

Acknowledgments

For a list of the Women's Health Initiative investigators, go to www.whi.org/about/investigators.php.

Sources of Funding

Supported by National Heart, Lung, and Blood Institute CAN #8324207 (to Drs Gorodeski and Lauer).

Disclosures

None.

References

1. Rautaharju PM, Kooperberg C, Larson JC, LaCroix A. Electrocardiographic abnormalities that predict coronary heart disease events and mortality in postmenopausal women: the Women's Health Initiative. *Circulation*. 2006;113:473–480.
2. Rautaharju PM, Kooperberg C, Larson JC, LaCroix A. Electrocardiographic predictors of incident congestive heart failure and all-cause mortality in postmenopausal women: the Women's Health Initiative. *Circulation.* 2006; 113:481–489.
3. Denes P, Larson JC, Lloyd-Jones DM, Prineas RJ, Greenland P. Major and minor ECG abnormalities in asymptomatic women and risk of cardiovascular events and mortality. *JAMA*. 2007;297:978–985.
4. Zhang ZM, Prineas RJ, Eaton CB. Evaluation and comparison of the Minnesota Code and Novacode for electrocardiographic Q-ST wave abnormalities for the independent prediction of incident coronary heart disease and total mortality (from the Women's Health Initiative). *Am J Cardiol.* 2010;106:18–25.
5. Iuliano S, Fisher SG, Karasik PE, Fletcher RD, Singh SN. QRS duration and mortality in patients with congestive heart failure. *Am Heart J*. 2002;143:1085–1091.
6. Breiman L. Statistical modeling: the two cultures. *Stat Sci*. 2001;16: 199–231.
7. Breiman L. Random forests. *Machine Learning*. 2001;45:5–32.
8. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2:841–860.
9. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, Rossouw JE. The Women's Health Initiative recruitment methods and results. *Ann Epidemiol*. 2003;13:S18–S77.
10. Prineas RJ, Crow RS, Blackburn H. *The Minnesota Code Manual of Electrocardiographic Findings*. Boston, MA: John Wright PSB; 1982:203.
11. Prineas RJ, Crow RS, Zhang ZM. *The Minnesota Code Manual of Electrocardiographic Findings.* 2nd ed. London, UK: Springer; 2009: 277–324.
12. Rautaharju PM, Park LP, Chaitman BR, Rautaharju F, Zhang ZM. The Novacode criteria for classification of ECG abnormalities and their clinically significant progression and regression. *J Electrocardiol*. 1998;31: 157–187.
13. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials.* 1998;19:61–109.
14. Lauer MS, Blackstone EH, Young JB, Topol EJ. Cause of death in clinical research: time for a reassessment? *J Am Coll Cardiol*. 1999;34: 618–620.
15. Curb JD, McTiernan A, Heckbert SR, Kooperberg C, Stanford J, Nevitt M, Johnson KC, Proulx-Burns L, Pastore L, Criqui M, Daugherty S. Outcomes ascertainment and adjudication methods in the Women's Health Initiative. *Ann Epidemiol*. 2003;13:S122–S128.
16. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-dimensional variable selection for survival data. *J Am Stat Assoc*. 2010; 105:205–217.
17. Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biom J*. 2008;50:457–479.
18. Ishwaran H, Kogalur UB. RandomSurvivalForest 3.5.1 R Package. The Comprehensive R Archive Network Web site. http://cran.r-project.org. Accessed on May 5, 2009.
19. Ishwaran H, Kogalur UB. Random Survival Forests for R. *Rnews*. 2007; 7:25–31.
20. Mitchell TM. *Machine Learning*. New York: McGraw-Hill; 1997.
21. Breiman L. Heuristics of instability and stabilization in model selection. *Ann Statist.* 1996;24:2350–2383.
22. Breiman L. Bagging predictors. *Machine Learning*. 1996;1996:123–140.
23. Freund Y, Shapire RE. Experiments with a new boosting algorithm. In: *Proceedings of the 13th International Conference on Machine Learning.* Burlington, MA: Morgan Kaufmann; 1996:148–156.
24. Cortes C, Vapnik VN. Support-vector networks. *Machine Learning*. 1995;20:273–297.
25. Bureau A, Dupuis J, Falls K, Lunetta KL, Hayward B, Keith TP, Van Eerdewegh P. Identifying SNPs predictive of phenotype using random forests. *Genet Epidemiol*. 2005;28:171–182.
26. Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*. 2006;7:3.
27. Lunetta KL, Hayward LB, Segal J, Van Eerdewegh P. Screening large-scale association study data: exploiting interactions using random forests. *BMC Genet*. 2004;5:32.
28. Pang H, Lin A, Holford M, Enerson BE, Lu B, Lawton MP, Floyd E, Zhao H. Pathway analysis using random forests classification and regression. *Bioinformatics*. 2006;22:2028–2036.
29. Minn AJ, Gupta GP, Padua D, Bos P, Nguyen DX, Nuyten D, Kreike B, Zhang Y, Wang Y, Ishwaran H, Foekens JA, van de Vijver M, Massague J. Lung metastasis genes couple breast tumor size and metastatic spread. *Proc Natl Acad Sci U S A*. 2007;104:6740–6745.
30. Parkhurst DF, Brenner KP, Dufour AP, Wymer LJ. Indicator bacteria at five swimming beaches-analysis using random forests. *Water Res*. 2005; 39:1354–1360.
31. Ward MM, Pajevic S, Dreyfuss J, Malley JD. Short-term prediction of mortality in patients with systemic lupus erythematosus: classification of outcomes using random forests. *Arthritis Rheum*. 2006;55:74–80.

SUPPLEMENTAL MATERIAL

Supplemental Methods

Interactions

We investigated interactions between our top 20 variables by using a modified version of minimal depth. For each variable ***v***, we determined its minimal depth. We then calculated the "relative minimal depth" for each of the other 19 variables to ***v*** by counting the number of splits needed until that variable split for the first time under ***v***. A variable with a small minimal depth relative to ***v*** is highly associated with ***v*** because it indicates a variable with a tendency to split whenever ***v*** does.[1] The relative minimal depth for each variable was determined by averaging over the forest. This resulted in 19 values for each variable. These values were converted to a distance matrix and complete linkage hierarchical clustering was applied to this matrix.

Supplemental Figure 4 displays the dendrogram from the hierarchical clustering analysis (for convenience, the bottom part of the figure displays the minimal depth for each variable rounded to the nearest integer). This figure identifies three to four distinct clusters. The "top" cluster identified by the clustering algorithm is age (green cluster on the extreme left). Its height in the dendrogram and the fact that its minimal depth is by far the smallest of all variables suggests that all variables must in some way interact with it. On the right-hand side systolic blood pressure (blue cluster) and smoking, waist-to-hip ratio, and SDNN (orange cluster) comprise two near similar clusters that appear primarily to be a clinical effect. These variables must be interrelated to one another. Finally, the large cluster in the middle (red cluster) is predominantly ECG-based. These variables must also be interrelated.

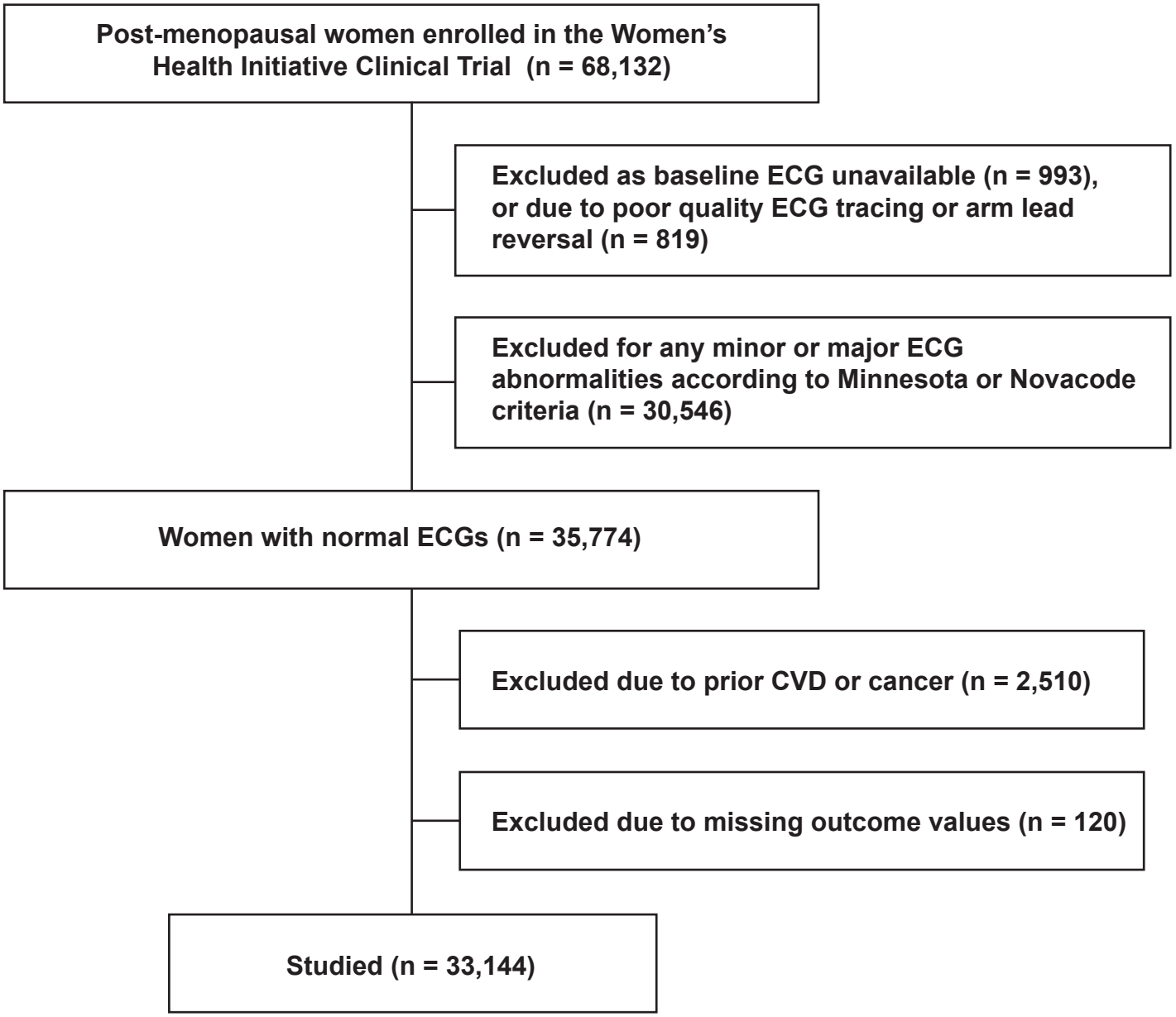
As an example, **Supplemental Figure 5** illustrates how survival depends upon age and the three variables in the orange cluster. Plotted on the left vertical axis is 10-year predicted survival against age on the bottom horizontal axis. Each panel is conditioned by waist-to-hip ratio (top horizontal axis) and SDNN (right vertical axis). Red and blue curves are survival stratified by smoking behavior, with blue indicating smoking. In all

panels, survival decreases with increasing age and with smoking. One can see an interaction involving the remaining two variables, with survival being generally worse for patients with large waist-to-hip ratios and with low SDNN values.

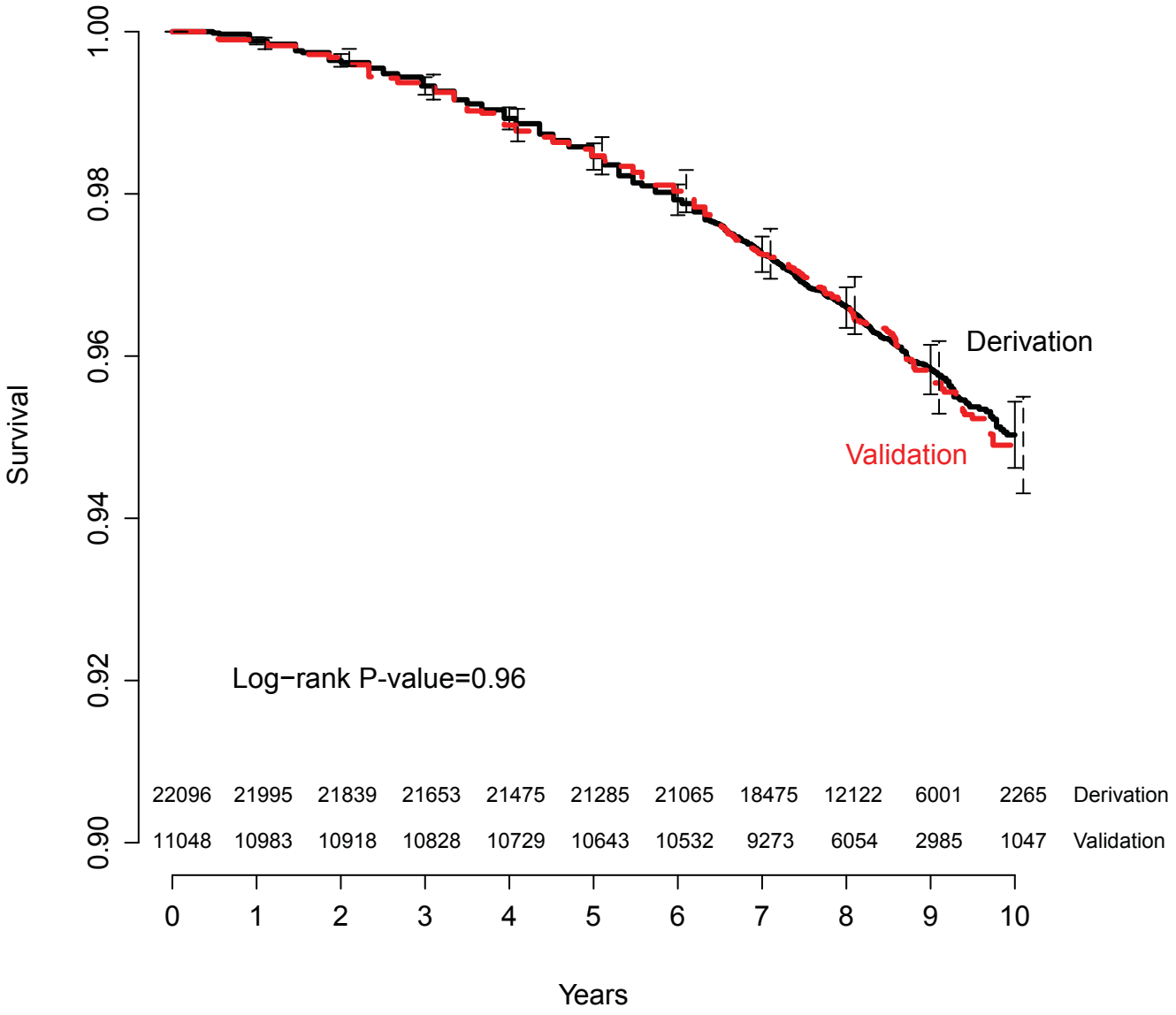
Supplemental Table 1. Hazard ratios and 95% confidence intervals (for 1 standard deviation of difference in continuous variables) amongst 20 variables identified by random survival forest

	Hazard Ratio	Lower 95% CI	Upper 95% CI
Age	1.70	1.58	1.83
Waist-to-hip ratio	1.06	1.00	1.11
Smoking	1.58	1.42	1.76
Income	0.83	0.77	0.90
Systolic blood pressure	1.13	1.06	1.21
SDNN	0.98	0.86	1.12
Ventricular rate	1.24	1.09	1.40
T-wave area (lead I)	0.72	0.55	0.94
P-wave intrinsicoid time (lead V4)	1.02	0.94	1.10
P-wave duration (lead V2)	0.92	0.85	0.99
T-wave amplitude (lead I)	1.07	0.85	1.34
RMS-SD	0.98	0.87	1.11
T-wave axis (median of all leads)	1.30	1.13	1.49
P-wave axis (median of all leads)	1.04	0.99	1.09
Body mass index	1.11	1.04	1.19
P-wave amplitude (lead I)	0.97	0.89	1.05
T-wave area (lead avL)	1.43	1.15	1.79
QT duration (median of all leads)	1.14	1.00	1.29
P-wave intrinsicoid duration (lead V3)	0.98	0.90	1.06
P-wave area (lead V2)	1.01	0.92	1.10

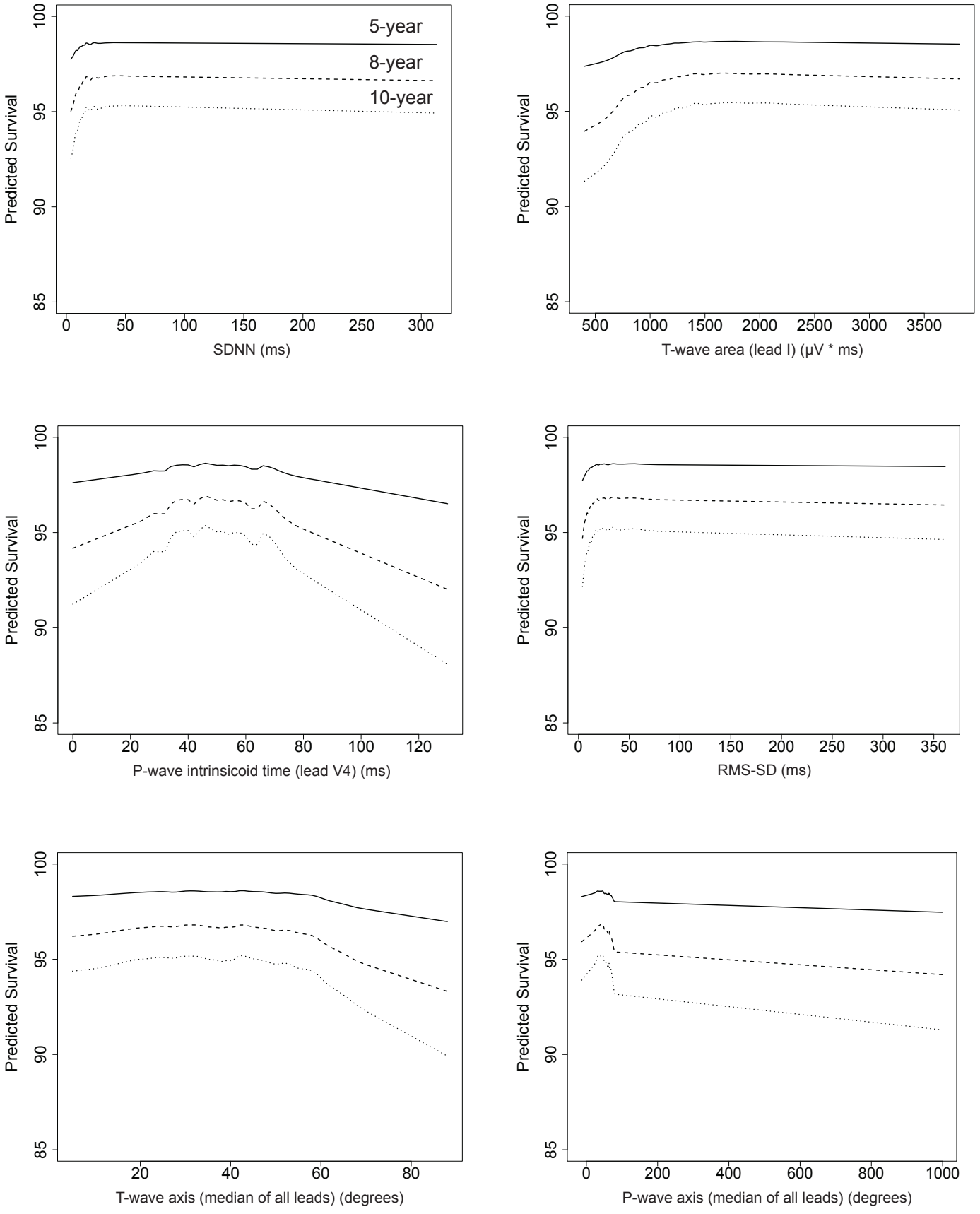
Supplemental Figure 1.



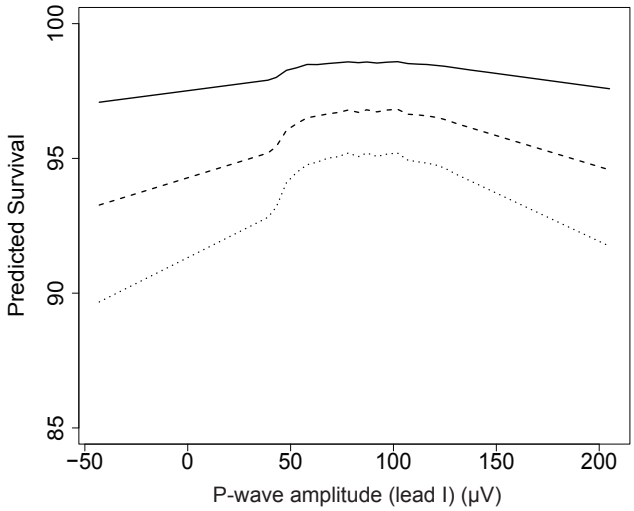
Supplemental Figure 2.

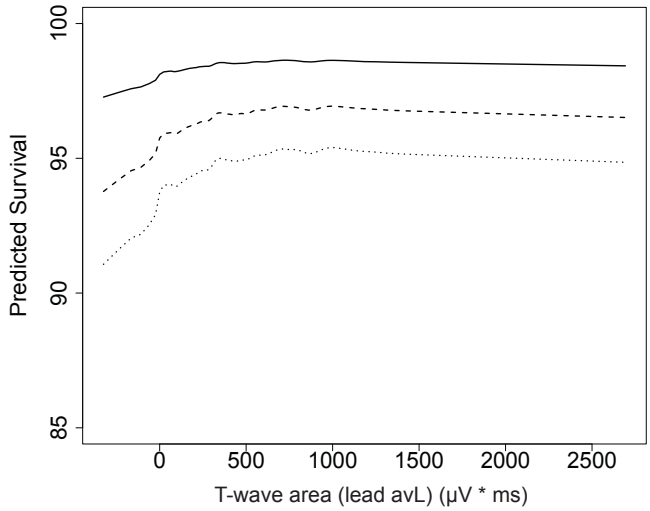


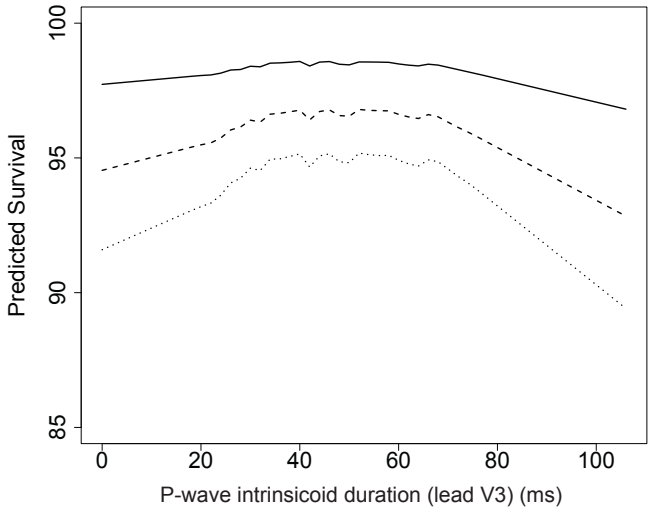
Supplemental Figure 3.

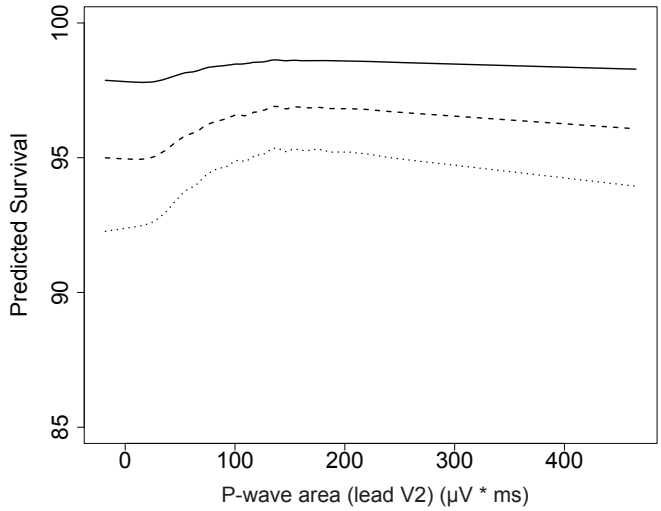


Supplemental Figure 3. (continued)

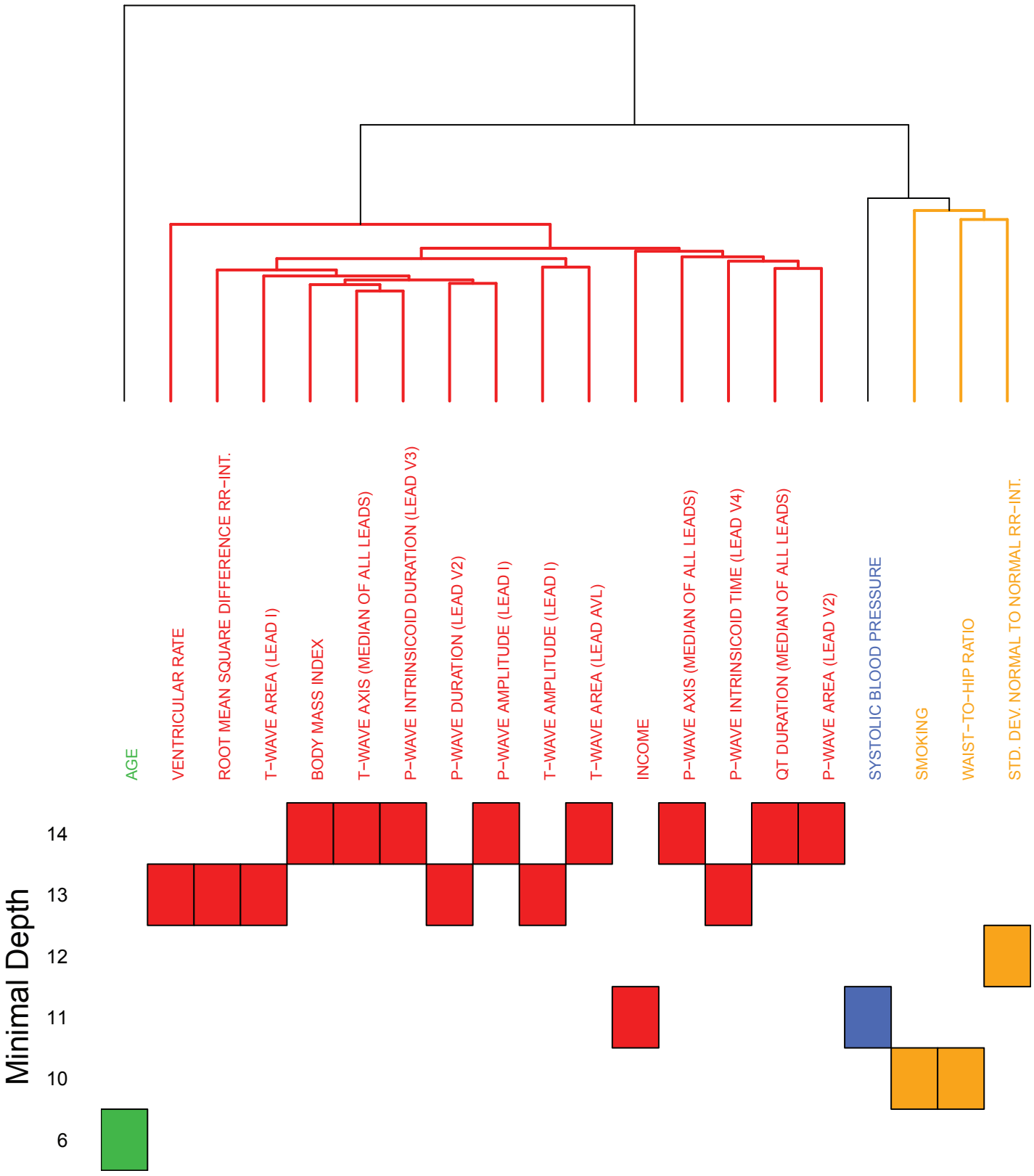




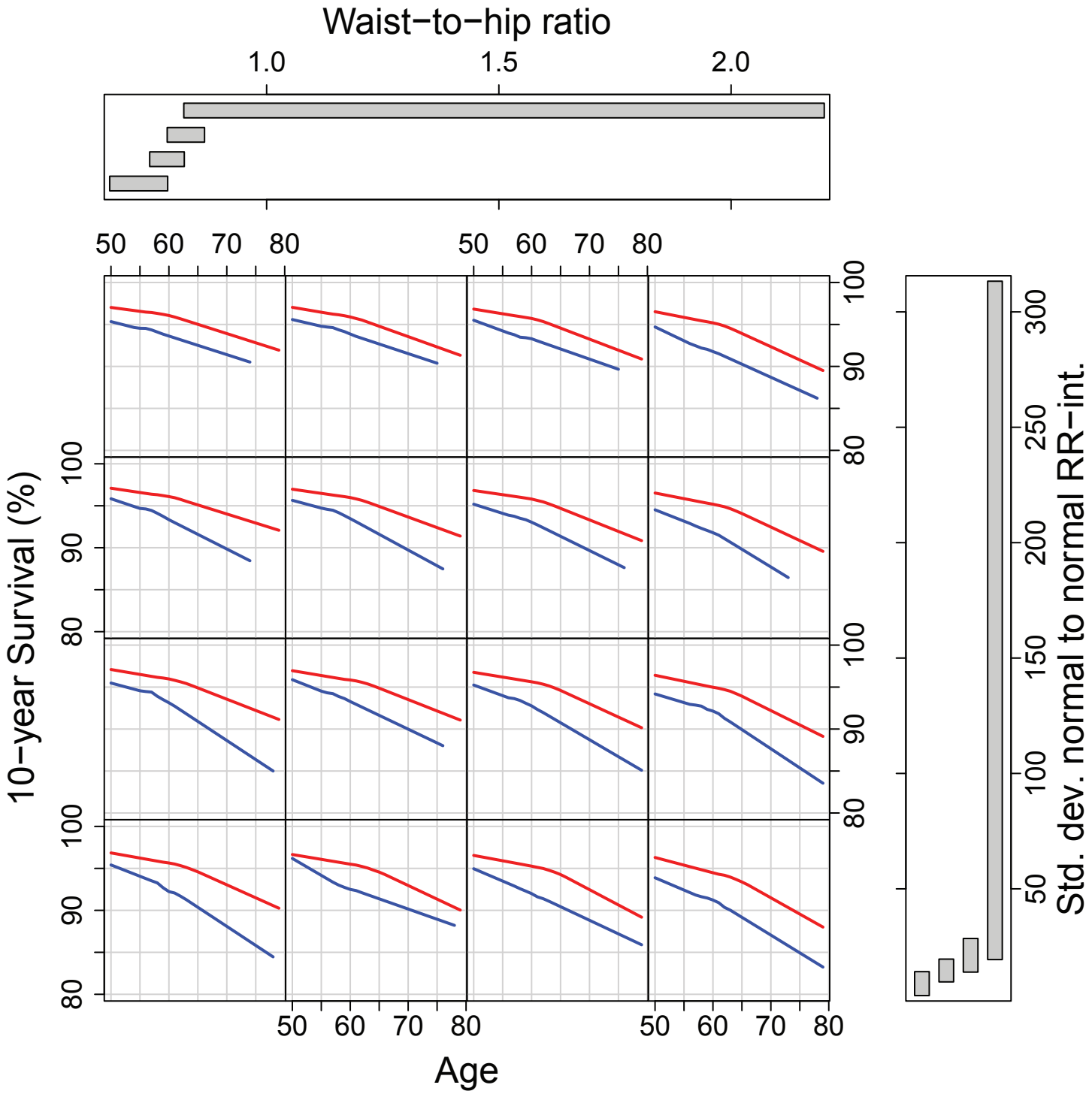




Supplemental Figure 4.



Supplemental Figure 5.



Supplemental Figure Legends

Supplemental Figure 1. CONSORT diagram.

Supplemental Figure 2. Kaplan-Meier plot comparing outcomes between derivation and validation cohorts.

Supplemental Figure 3. Adjusted-predicted survival (%) at 5, 8, and 10 years.

Supplemental Figure 4. Dendrogram presenting results of hierarchical clustering analysis.

Supplemental Figure 5. RSF-estimated 10-year survival as a function of age, SDNN, and WHR for smokers (blue) and non-smokers (red). Smoothed curves are loess curves of the estimated survival for each individual.

Supplemental References

1. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-Dimensional Variable Selection for Survival Data. *J Am Stat Assoc.* 2010;105:205-217.