Using Machine-Learning for Prediction of the Response to Cardiac Resynchronization Therapy: the SMART-AV Study

Stacey J. Howell, MD,[1] Tim Stivland, BSME, MBA,[2] Kenneth Stein, MD, FHRS,[2] Kenneth A. Ellenbogen, MD, FHRS,[3] Larisa G. Tereshchenko, MD, PhD, FHRS, CCDS.[1,4]

[1]Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR; [2]Boston Scientific; [3]Medical College of Virginia/VCU School of Medicine; [4]Cardiovascular Division, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD.

Brief Title: SMART-AV CRT Response prediction

Correspondence: Larisa Tereshchenko, 3181 SW Sam Jackson Park Rd; UHN62; Portland, OR, 97239. E-mail:tereshch@ohsu.edu.

Words:0

Abstract

Introduction—We aimed to apply machine learning (ML) to develop a prediction model for cardiac resynchronization therapy (CRT) response.

Methods and Results—Participants from the SmartDelay Determined AV Optimization (SMART-AV) trial (n=741; age, 66 ±11 yrs; 33% female; 100% NYHA III-IV; 100% EF≤35%) were randomly split into training & testing (80%; n=593), and validation (20%; n=148) samples. Baseline clinical, ECG, echocardiographic and biomarker characteristics, and left ventricular (LV) lead position (43 variables) were included in 6 ML models (random forests, convolutional neural network, lasso, adaptive lasso, plugin lasso, elastic net, ridge, and logistic regression). A composite of freedom from death and heart failure hospitalization and a >15% reduction in LV end-systolic volume index at 6-months post-CRT was the endpoint. The primary endpoint was met by 337 patients (45.5%). The adaptive lasso model was more accurate than class I ACC/AHA guidelines criteria (AUC 0.759; 95%CI 0.678-0.840 versus 0.639; 95%CI 0.554-0.722; $P<0.0001$), well-calibrated, and parsimonious (19 predictors; nearly half are potentially modifiable). The model predicted CRT response with 70% accuracy, 70% sensitivity, and 70% specificity, and should be further validated in prospective studies.

Conclusions—ML predicts short-term CRT response and thus may help with CRT procedure planning.

Clinical trial registration—ClinicalTrials.gov Identifier: NCT00677014

Keywords: cardiac resynchronization therapy, machine learning

Introduction

Cardiac resynchronization therapy (CRT) is an established treatment for patients with systolic heart failure (HF) and ventricular dyssynchrony.[1] However, despite proven benefit, nearly a third of CRT recipients are considered to be "non-responders".[2]

Guided left ventricular (LV) lead placement considering the timing of LV activation and electrical delay[3], together with dynamic atrioventricular (AV) optimization[4], can potentially reduce the CRT non-response rate. Previous analysis of the SMART-AV (SmartDelay Determined AV Optimization: A Comparison to Other AV Delay Methods Used in Cardiac Resynchronization Therapy) study suggested a strategy for using measures of LV electrical delay at implantation to guide LV lead placement.[5] However, a complex interaction between cardiac veins anatomy and cardiomyopathy substrate can make guided LV lead placement procedure technically difficult. Prediction of the probability of a CRT response can possibly help with the allocation of resources and CRT procedure planning.

Machine learning (ML) has taken hold in a number of fields to improve risk prediction as compared to traditional methods.[6,7] Several studies have applied ML to address the clinical challenge of CRT patient selection and showed that ML algorithms perform better than guidelines-recommended QRS duration and bundle branch block (BBB) morphology.[8-11] However, all previous ML-prediction models targeted the long-term (≥ 1 year) CRT outcomes. At present, there is no short-term (6-month) CRT response prediction tool that can be used to plan CRT implantation and delivery.

We conducted the current study with the goal to use ML to predict short-term (6-month) response to CRT.

Methods

The authors used the deidentified SMART-AV study dataset, provided by the executive study committee. The Oregon Health & Science University Institutional Review Board determined the deidentified nature of the dataset. Open-source code for statistical data analysis is provided at https://github.com/Tereshchenkolab/statistics.

Study population

The SMART-AV was a randomized, multicenter, single-blinded clinical trial[12,13] that sought to determine whether AV delay optimization would improve CRT response at six months post-implant. The trial enrolled New York Heart Association (NYHA) class III-IV HF patients with left ventricular ejection fraction (LVEF)≤35% despite optimal medical therapy, and QRS duration ≥120 ms, in sinus rhythm. HF patients who were in complete heart block, could not tolerate pacing at VVI-40-RV for up to two weeks, or previously received CRT were excluded. Enrollment was completed from May 2008 through December 2009. In the current study, we excluded participants with missing candidate predictor variables and lost to follow-up. Of the 980 randomized SMART-AV participants, 741 CRT recipients were included in this study.

Candidate predictor variables

At the enrollment visit, baseline clinical characteristics data were collected, which included medical history, current cardiovascular evaluation (NYHA class) and medications list, the 6-minute walk test, quality of life (Minnesota Living with Heart Failure Questionnaire), and blood draw for biomarkers.[12,13] We calculated estimated glomerular filtration rate (eGFR) using the chronic kidney disease (CKD) Epidemiology Collaboration equation (CKD-EPI).[14] LV lead location was selected at the discretion of the implanting physician. Baseline ECG and

echocardiogram were recorded post-implant (no biventricular pacing).[12,13] We normalized LV volumes and dimensions by body surface area (BSA).

The study endpoint

We defined the primary endpoint as a composite of freedom from death and HF hospitalization and a >15% reduction[4,5,15,16] in LV end-systolic volume index (LVESVI) at six months of follow-up. LVESV was the primary endpoint in the SMART-AV trial.[12,13] A single core laboratory performed all echocardiographic measurements in a blinded fashion.

Statistical machine learning analysis

We randomly split the study population into two non-overlapping samples: training & testing (80%; n=593), and validation (20%; n=148). Considering future clinical implementation, we included routinely available predictor variables that describe baseline clinical, ECG, echocardiographic and biomarker characteristics, and LV lead position (43 variables, Table 1).

We fitted eight different models (random forests[17], convolutional neural network[18], lasso, adaptive lasso, plugin lasso, elastic net, ridge, and logistic regression).

To train the random forests algorithm, we arranged the data in a randomly sorted order and tuned the number of subtrees and number of variables to randomly investigate at each split. We calculated both out-of-bag error (tested against training data subsets that are not included in subtree construction) and a validation error (tested against the validation data) to find the model with the highest testing accuracy.

We trained the convolutional neural network with 20 hidden layers, using 500 iterations with a training factor 2 and 4 normalization parameters. The network was comprised of 3 layers, 64 neurons per layer, and 901 synapse weights.

The family of lasso (least absolute shrinkage and selection operator) models employed ten-fold cross-validation in the training & testing sample. In lasso model, cross-validation selected the tuning parameter λ that minimized the out-of-sample deviance. The adaptive lasso performed multistep cross-validation, performing the second cross-validation step among the covariates selected in the first cross-validation step. The plugin lasso used partialing-out estimators to determine which covariates belong in the model, achieving an optimal bound on the number of covariates it included.[19] The elastic net permitted retention of correlated covariates.[20] In the ridge model, the penalty parameter used squared terms and kept all predictors in the model.

We validated the predictive accuracy of the models by comparing the area under the receiver operator curve (ROC AUC) in the validation sample. To assess calibration, we compared the observed and predicted proportions within the groups formed by the Hosmer-Lemeshow test[21], and used the calibration belt[22] to examine the relationship between out-of-sample estimated probabilities and observed CRT response rates. For the lasso family of models, we also calculated the out-of-sample deviance and deviance ratio.

We compared the performance of the selected model to the current 2013 American College of Cardiology Foundation/American Heart Association class I guideline criteria (QRS>150 ms and the presence of LBBB).[23]

Statistical analysis was performed using STATA MP 16.1 (StataCorp LP, College Station, TX). *P*-value < 0.05 was considered statistically significant.

Results

The SMART-AV study population characteristics are shown in Table 1 and have been previously reported elsewhere.[16] The primary endpoint was met by 337 patients (45.5%). Out of

404 participants who failed to respond, 13 died, 75 participants were hospitalized because of HF, and 316 participants failed to achieve a volumetric response.

In tuning the random forests algorithm, we observed that both out-of-bag error and validation error stabilized after 300 iterations at 30-35% (Figure 1), and we conservatively chose 500 subtrees. The minimum validation error was observed for 7 variables, and we chose 7 variables to investigate at each split randomly. The final random forests model reported 26% error in validation sample; it accurately predicted freedom from composite CRT response endpoint in 71 out of 83 participants (specificity 85.5%), and correctly predicted CRT response in 38 out of 65 individuals (sensitivity 58.5%), having a positive predictive value of 76% and negative predictive value of 72.4%. The single most important predictor was diabetes (Figure 2), which, together with demographic characteristics (age, sex, race) and other comorbidities (hypertension, smoking) comprised six the most important predictors.

A comparison of the prediction models' performance is shown in Table 2. The convolutional neural network demonstrated the highest predictive accuracy in the training & testing sample, with a final error of only 6%. However, the calibration of the convolutional neural network model was unsatisfactory (Hosmer-Lemeshow test $P<0.0001$; Figure 3), and predictive accuracy in the validation sample did not differ from the lasso family of models.

Several models (lasso, adaptive lasso, elastic net, ridge, and logistic regression) demonstrated similar fit and predictive accuracy both in training & testing, and validation samples (Table 2), which was significantly higher than current class I clinical guidelines (AUC 0.639; 95%CI 0.554-0.722), $P<0.0001$. Figure 4 shows the cross-validation function and selected λ for each model. Only a few models (logistic regression, adaptive lasso, and plugin lasso) showed

satisfactory out-of-sample calibration (Figure 5). Ultimately, we selected the adaptive lasso model as the most accurate, well-calibrated, and parsimonious (19 predictors listed in Table 3).

In the adaptive lasso model, the most important predictors (Figure 6) characterized dyssynchrony (ventricular conduction type, QRS duration), underlying disease substrate (cardiomyopathy type, primary prevention indication), and modifiable characteristics (NT-proBNP, systolic blood pressure), including PR interval. Nonischemic cardiomyopathy, female sex, primary prevention indication, history of valvular heart disease and cancer, and higher QRS duration, systolic blood pressure, LVEDVI, 6-min walk distance, $eGFR_{CKD\text{-}EPI}$, and age were associated with CRT response. Non-LBBB, AV block, and higher NT-proBNP, CRP, PR interval, LVEF, LVESDI, and weight were associated with non-response. Participants in the 5th quantile as compared to those in the 1st quantile had 14-fold higher odds of composite CRT response (Figure 7). The online calculator is freely available at http://www.ecgpredictscd.org/crt.

Discussion

In this study, using the ML approach, we developed a parsimonious model for the prediction of CRT response that is comprised of routinely available baseline clinical, ECG, and echocardiographic characteristics - measures of the disease substrate, dyssynchrony, and comorbidities. Several included predictors could be potentially modifiable. Developed in this study, the CRT response prediction model opens an avenue for a future randomized controlled trial, testing CRT implantation planning strategy, incorporating targeted lead placement and dynamic AV optimization programming.[4,5]

It has been previously shown that increasing degrees of interventricular (rather than intraventricular) dyssynchrony is expected to result in improved rates of clinical CRT response.[24]

Previous analysis of the SMART-AV study showed that optimally timed AV delay provides an incremental benefit to the substantial interventricular conduction delay[4,5], suggesting that both LV lead and right ventricular (RV) lead placement should target maximizing RV-LV delay. Pre-procedural planning may involve expensive and time-consuming cardiac imaging. Our risk score can predict the probability of the short-term composite CRT response and, therefore, can help to preserve resources while improving clinical outcomes. Careful pre-procedural planning would be particularly critical for CRT candidates with a moderate or low probability of CRT response, especially if they have modifiable factors. Notably, both the baseline PR interval and the presence of AV block were selected by the adaptive lasso model as essential predictors in the model, indicating the likely benefit of dynamic AV optimization.

Consistently with prior studies[5,8-11], we confirmed that ML could improve patient selection for CRT therapy beyond current guidelines. The strength of ML algorithms is the ability to capture complex interactions.[25] Several prior studies have used ML to predict CRT response. Kalscheur et al analyzed 595 COMPANION NYHA III/IV patients,[8] Cikes et al studied 1106 MADIT-CRT NYHA class ≤ II patients,[11] Feeny et al evaluated 470 NYHA I-IV patients from an observational cohort, and Hu et al retrospectively analyzed 990 predominately NYHA II-III patients from a single-center cohort.[26] Of note, all previous studies considered long-term CRT benefits, answering a question of CRT candidate selection. In contrast, our prediction model is focusing on a short-term CRT response and can help planning the CRT delivery strategy, in addition to selecting the most appropriate CRT candidate. Distinguishing those at high risk of non-response could alert cardiologists to a specific group that requires special attention within the first six months after CRT implantation.

Presently response and outcomes following CRT implantation vary significantly[2], making it crucial to improve patient selection for CRT. Improved identification of CRT responders could help to avoid CRT implantation in patients unlikely to benefit and to disproportionately incur undue harm and risk. Better prediction of CRT non-responders could be used to identify patients that may be better served with earlier consideration of advanced HF therapies, including mechanical circulatory support and transplantation rather than CRT, which would carry a lower yield of clinical improvement.

In this study, an absence of sustained ventricular tachyarrhythmia (primary prevention indication) was an important predictor of CRT response. This finding is consistent with previous studies that showed the antiarrhythmic effect of CRT and reversed electrical remodeling[27], which can be facilitated by the autonomic nervous system response.[28]

A comparison of ML models and selection of the "best" model also deserves discussion. We observed similar accuracy in all but one (plugin lasso) models, leaving seven models for consideration. However, only two of them (logistic regression and adaptive lasso) demonstrated satisfactory calibration. The parsimonious model (adaptive lasso) won because of (1) convenience (19 versus 43 predictors), and (2) approach to feature importance ranking. The most important predictors in the random forests model describe comorbidities and demographic characteristics, which unlikely to be modified (age, sex, race, diabetes, hypertension, smoking). In contrast, the most important predictors in the adaptive lasso model provide a meaningful characterization of the disease substrate and its electrophysiology (a type of cardiomyopathy and conduction abnormality, QRS duration, history of sustained ventricular tachyarrhythmia or cardiac arrest, NT-proBNP and systolic blood pressure), which can guide CRT delivery.

Strengths and Limitations

SMART-AV is a large multicenter randomized control trial with careful phenotyping that included blinded analysis of echocardiograms and biomarkers in core laboratories, and appropriate follow-up, providing an opportunity to study composite CRT response. A strength of the present study was the use of a composite endpoint of clinical outcomes (death, HF hospitalization) and volumetric remodeling. However, limitations of the study have to be taken into account. The study population was predominantly men, although this is characteristic and similar to other CRT trials. We limited candidate predictor variables by currently widely available and did not include novel ECG measures of dyssynchrony that can potentially further improve prediction.[15,29]

Conclusion

In summary, in this study, using ML, we developed and validated a parsimonious model that is comprised of routinely available baseline clinical, ECG, and echocardiographic characteristics. The model outperforms the current guidelines and predicts CRT response with 70% accuracy, 70% sensitivity, and 70% specificity, and should be further validated in prospective studies. The calculator is available at http://www.ecgpredictscd.org/crt.

Funding

This work was supported in part by HL118277 (LGT).

Disclosures

SMART AV trial was sponsored by Boston Scientific.

References

1. Tracy CM, Epstein AE, Darbar D, et al. 2012 ACCF/AHA/HRS Focused Update of the 2008 Guidelines for Device-Based Therapy of Cardiac Rhythm Abnormalities: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Heart Rhythm.* 2012;9(10):1737-1753.
2. Chatterjee NA, Singh JP. Cardiac resynchronization therapy: past, present, and future. *Heart Fail Clin.* 2015;11(2):287-303.
3. Singh JP, Fan D, Heist EK, et al. Left ventricular lead electrical delay predicts response to cardiac resynchronization therapy. *Heart Rhythm.* 2006;3(11):1285-1292.
4. Gold MR, Yu Y, Singh JP, et al. Effect of Interventricular Electrical Delay on Atrioventricular Optimization for Cardiac Resynchronization Therapy. *Circ Arrhythm Electrophysiol.* 2018;11(8):e006055.
5. Field ME, Yu N, Wold N, Gold MR. Comparison of measures of ventricular delay on cardiac resynchronization therapy response. *Heart Rhythm.* 2020;17(4):615-620.
6. Deo RC. Machine Learning in Medicine. *Circulation.* 2015;132(20):1920-1930.
7. Haq KT, Howell SJ, Tereshchenko LG. Applying Artificial Intelligence to ECG Analysis: Promise of a Better Future. *Circ Arrhythm Electrophysiol.* 2020;13(8):e009111.
8. Kalscheur MM, Kipp RT, Tattersall MC, et al. Machine Learning Algorithm Predicts Cardiac Resynchronization Therapy Outcomes: Lessons From the COMPANION Trial. *Circ Arrhythm Electrophysiol.* 2018;11(1):e005499.
9. Tokodi M, Schwertner WR, Kovács A, et al. Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: the SEMMELWEIS-CRT score. *European Heart Journal.* 2020;41(18):1747-1756.

10. Feeny AK, Rickard J, Patel D, et al. Machine Learning Prediction of Response to Cardiac Resynchronization Therapy: Improvement Versus Current Guidelines. *Circ Arrhythm Electrophysiol.* 2019;12(7):e007316.
11. Cikes M, Sanchez-Martinez S, Claggett B, et al. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *European journal of heart failure.* 2019;21(1):74-85.
12. Stein KM, Ellenbogen KA, Gold MR, et al. SmartDelay determined AV optimization: a comparison of AV delay methods used in cardiac resynchronization therapy (SMART-AV): rationale and design. *Pacing Clin Electrophysiol.* 2010;33(1):54-63.
13. Ellenbogen KA, Gold MR, Meyer TE, et al. Primary results from the SmartDelay determined AV optimization: a comparison to other AV delay methods used in cardiac resynchronization therapy (SMART-AV) trial: a randomized trial comparing empirical, echocardiography-guided, and algorithmic atrioventricular delay programming in cardiac resynchronization therapy. *Circulation.* 2010;122(25):2660-2668.
14. Matsushita K, Selvin E, Bash LD, Astor BC, Coresh J. Risk implications of the new CKD Epidemiology Collaboration (CKD-EPI) equation compared with the MDRD Study equation for estimated GFR: the Atherosclerosis Risk in Communities (ARIC) Study. *Am J Kidney Dis.* 2010;55(4):648-659.
15. Tereshchenko LG, Cheng A, Park J, et al. Novel measure of electrical dyssynchrony predicts response in cardiac resynchronization therapy: Results from the SMART-AV Trial. *Heart Rhythm.* 2015;12(12):2402-2410.

16. Cheng A, Gold MR, Waggoner AD, et al. Potential mechanisms underlying the effect of gender on response to cardiac resynchronization therapy: insights from the SMART-AV multicenter trial. *Heart Rhythm.* 2012;9(5):736-741.
17. Schonlau M, Zou RY. The random forest algorithm for statistical learning. *The Stata Journal.* 2020;20(1):3-29.
18. *BRAIN: Stata module to provide neural network* [computer program]. Version 1. Boston: Boston College Department of Economics; 2018.
19. Belloni A, Chen D, Chernozhukov V, Hansen C. Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica.* 2012;80(6):2369-2429.
20. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2005;67(2):301-320.
21. Lemeshow S, Hosmer DW, Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol.* 1982;115(1):92-106.
22. Nattino G, Lemeshow S, Phillips G, Finazzi S, Bertolini G. Assessing the calibration of dichotomous outcome models with the calibration belt. *Stata Journal.* 2017;17(4):1003-1014.
23. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation.* 2013;128(16):1810-1852.
24. Waks JW, Perez-Alday EA, Tereshchenko LG. Understanding Mechanisms of Cardiac Resynchronization Therapy Response to Improve Patient Selection and Outcomes. *Circ Arrhythm Electrophysiol.* 2018;11(4):e006290.

25. Rahman QA, Tereshchenko LG, Kongkatong M, Abraham T, Abraham MR, Shatkay H. Utilizing ECG-Based Heartbeat Classification for Hypertrophic Cardiomyopathy Identification. *IEEE Trans Nanobioscience.* 2015;14(5):505-512.
26. Hu SY, Santus E, Forsyth AW, et al. Can machine learning improve patient selection for cardiac resynchronization therapy? *PloS one.* 2019;14(10):e0222397.
27. Tereshchenko LG, Henrikson CA, Stempniewicz P, Han L, Berger RD. Antiarrhythmic effect of reverse electrical remodeling associated with cardiac resynchronization therapy. *Pacing Clin Electrophysiol.* 2011;34(3):357-364.
28. Tereshchenko LG, Henrikson CA, Berger RD. Strong coherence between heart rate variability and intracardiac repolarization lability during biventricular pacing is associated with reverse electrical remodeling of the native conduction and improved outcome. *J Electrocardiol.* 2011;44(6):713-717.
29. Jacobsson J, Borgquist R, Reitan C, et al. Usefulness of the Sum Absolute QRST Integral to Predict Outcomes in Patients Receiving Cardiac Resynchronization Therapy. *Am J Cardiol.* 2016;118(3):389-395.

Table 1. Baseline Clinical Characteristics in study participants

Characteristics		All (n=741)
Age(SD), y		66.0(11.0)
Female, n(%)		241(32.5)
White, n(%)		575(77.6)
LVEF(SD), %		27.5(8.7)
Weight(SD), kg		87.4(20.8)
Height(SD), cm		171.6(10.3)
Body mass index (SD), kg/m^2		29.6(6.2)
BP systolic(SD), mmHg		124.5(20.9)
BP diastolic(SD), mmHg		71.4(12.7)
Ischemic cardiomyopathy Hx, n(%)		426(57.5)
Primary prevention, n(%)		589(79.5)
Smoking Hx(current or former), n(%)		461(62.2)
Hypertension Hx, n(%)		528(71.3)
Diabetes Hx, n(%)		289(39.0)
Revascularization Hx, n(%)		380(51.3)
Autoimmune disease Hx, n(%)		19.0(2.6)
Sleep apnea Hx, n(%)		89(12.0)
Cancer Hx, n(%)		67(9.0)
Renal disease Hx, n(%)		119(16.1)
COPD Hx, n(%)		109(14.7)
Valve disease Hx, n(%)		40(5.4)
Pacemaker implant Hx, n(%)		15(2.0)
AV block, n(%)		138(18.6)
PR interval(SD), ms		198.2(50.4)
Heart rate(SD), bpm		71.3(12.5)
QRS duration(SD), ms		151.8(19.9)
Conduction disease:LBBB, n(%)		552(74.5)
	RBBB	81(10.9)
	IVCD	86(11.6)
	RBBB+left hemiblock	22(3.0)
NYHA class	II, n (%)	21(2.8)
	III	698(94.2)
	IV	22 (3.0)
6 minute walk(SD), m		268.2(124.7)
Quality of life(SD), points		47.2(25.0)
Potassium(SD), mmol/L		4.3(0.5)
Sodium(SD), mmol/L		138.7(3.1)

C-reactive protein(SD), ng/mL	6,438(4,425)
NT-proBNP median(IQR), pmol/L	1691(863-3952)
$eGFR_{CKD-EPI}$ (SD), mL/min/1.73 m²	63.6(22.8)
Use of ACEI/ARB, n (%)	485(65.5)
Use of beta blocker, n(%)	681(91.9)
Use of aldosterone antagonist, n(%)	262(35.4)
LV end systolic volume index (SD), mL/m^2	64.7(29.8)
LV end diastolic volume index (SD), mL/m^2	87.0(32.0)
LV end systolic diameter index (SD), cm/m^2	2.8(0.5)
LV end diastolic diameter index (SD), cm/m^2	3.2(0.5)
Lead location Apical n(%)	98(13.2)
Basal	47(6.3)
Mid	596(80.4)

Table 2. Development and validation of composite CRT response prediction tool

		Training & testing sample (N=593)					Validation sample (N=148)				
Group	Model	Deviance	Deviance ratio	Number of predictors	ROC AUC (95%CI)	P-value	Deviance	Deviance ratio	N predictors	ROC AUC (95%CI)	P-value
All participants	Ridge	1.201	0.129	43	0.753(0.714-0.792)		1.164	0.151	43	0.778(0.699-0.857)	
	Elastic net	1.196	0.133	30	0.751(0.711-0.790)		1.163	0.152	30	0.769(0.688-0.849)	
	Lasso	1.187	0.140	29	0.752(0.713-0.792)	0.277	1.155	0.158	29	0.770(0.690-0.850)	
	Adaptive lasso	1.184	0.142	19	0.751(0.712-0.790)		1.169	0.148	19	0.759(0.678-0.840)	0.368
	Logistic regress	1.147	0.168	43	0.768(0.730-0.805)		1.135	0.172	43	0.774(0.697-0.851)	
	CNN	-	-	43	0.979(0.966-0.993)	<0.0001	-	-	43	0.759(0.682-0.837)	
	Random forest	-	-	43	0.642(0.600-0.683)	<0.0001	-	-	43	0.720(0.649-0.791)	
	Plugin lasso	1.295	0.061	2	0.655(0.613-0.696)	<0.0001	1.296	0.055	2	0.667(0.582-0.751)	0.028

All coefficients are penalized except plugin lasso (postselection)

Table 3. Lassoknots, or predictors in the adaptive lasso model listed in the order of their importance

Predictor	Importance Rank	Beta-coefficient
Conduction type (non-LBBB)	1	-0.339
Nonischemic CM	2	0.267
QRS duration, ms	3	0.217
NT-proBNP	4	-0.180
Systolic BP, mmHg	5	0.186
Primary prevention	5	0.167
PR interval, ms	6	-0.155
Female	7	0.171
LVEF	8	-0.210
CRP	9	-0.100
LVEDVI	9	0.260
LVESDI	9	-0.368
6-min walk	10	0.114
$eGFR_{CKD\text{-}EPI}$	11	0.110
Age, y	12	0.155
Valve disease	13	0.069
Any AV block	13	-0.056
Weight, kg	13	-0.135
Cancer	14	0.011
constant		-0.202

Conduction type categories include:
1=LBBB
2=RBBB+left hemiblock
3=IVCD
4=RBBB

Figure Legends

Figure 1. Out-of-bag error and validation error plotted versus (A) number of iterations or subtrees, and (B) number of variables randomly investigated at each split in a random forests model.

Figure 2. Importance scores of predictor variables in the random forests model.

Figure 3. The calibration plot shows the observed and predicted CRT response proportions in convolutional neural network model for all participants. The size of the circles is proportional to the amount of data.

Figure 4: Cross-validation (CV) function (the mean deviance in the CV samples) is plotted over the search grid for the lasso penalty parameter λ on a reverse logarithmic scale for (**A**) lasso, (**B**) adaptive lasso, (**C**) elastic net, (**D**) ridge models. The first λ tried is on the left, and the last λ tried is on the right.

Figure 5. The calibration belt with 80% and 95% confidence intervals on the external sample shows the observed and predicted CRT response proportions in (**A**) logistic regression, (**B**) lasso, (**C**) adaptive lasso, (**D**) plugin lasso, (**E**) elastic net, and (**F**) ridge models for all participants.

Figure 6. Importance of the selected predictors in the adaptive lasso model. The most important predictors were added to the model early.

Figure 7. Probabilities of composite CRT response by quantiles of the adaptive lasso model.

Figure 1.

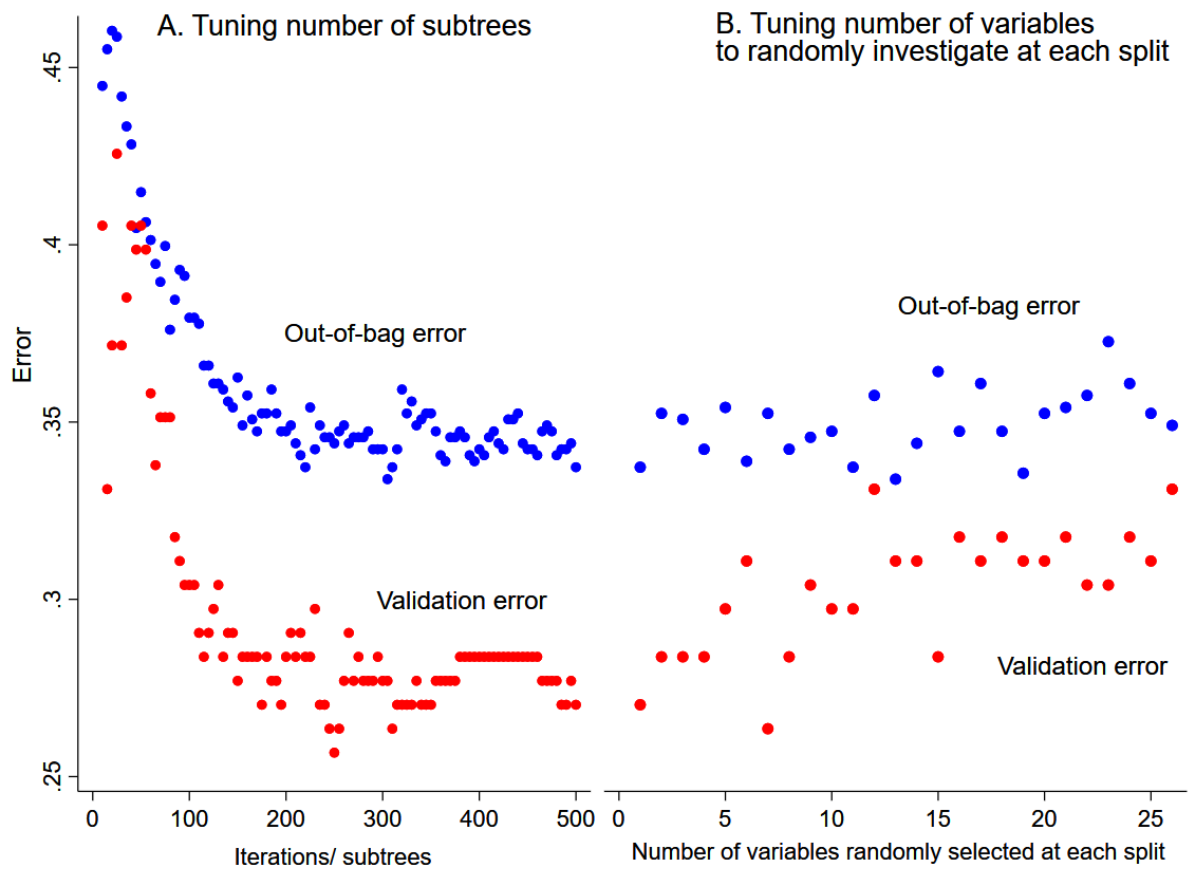

A. Tuning number of subtrees
B. Tuning number of variables
to randomly investigate at each split
Error
.25
.3
.35
.4
.45
Out-of-bag error
Validation error
0
100
200
300
400
500
Iterations/ subtrees
Out-of-bag error
Validation error
0
5
10
15
20
25
Number of variables randomly selected at each split

Figure 2:

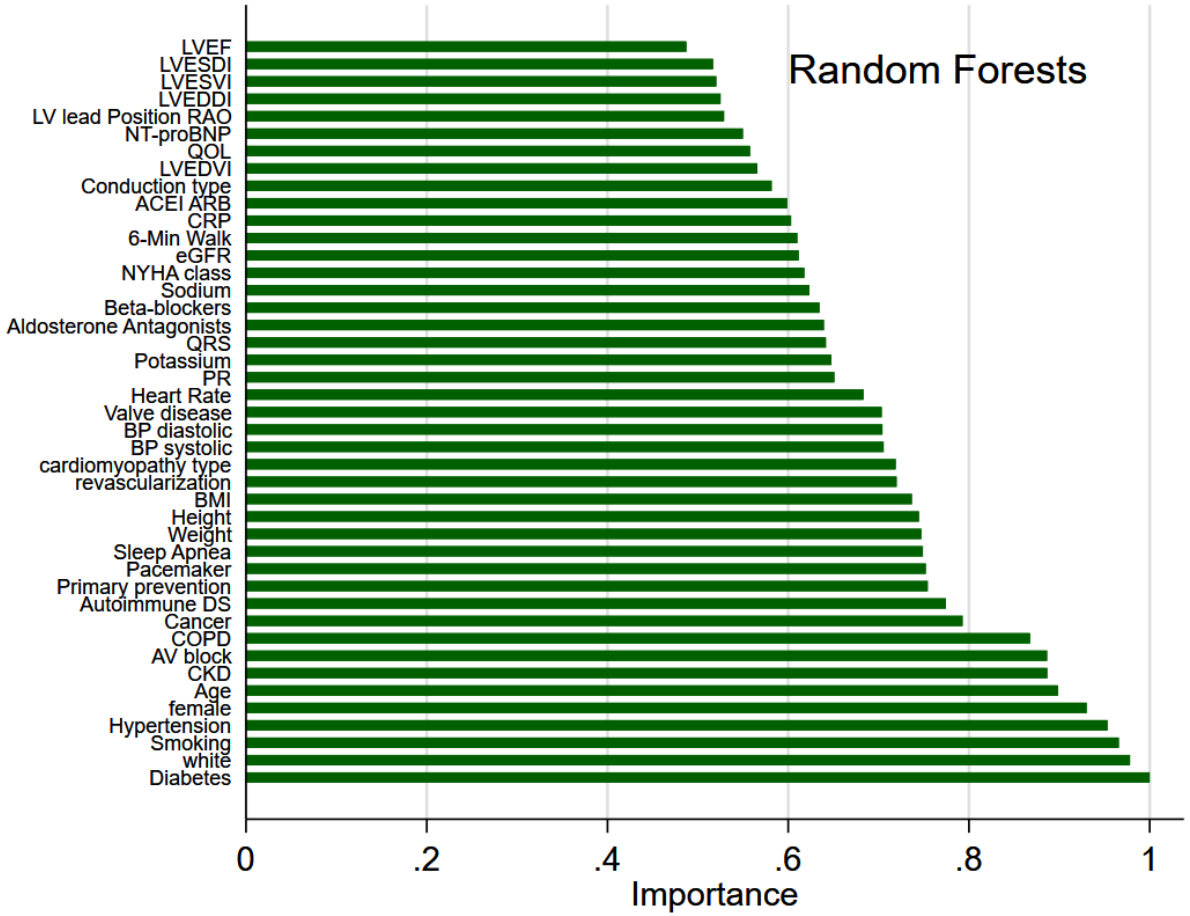
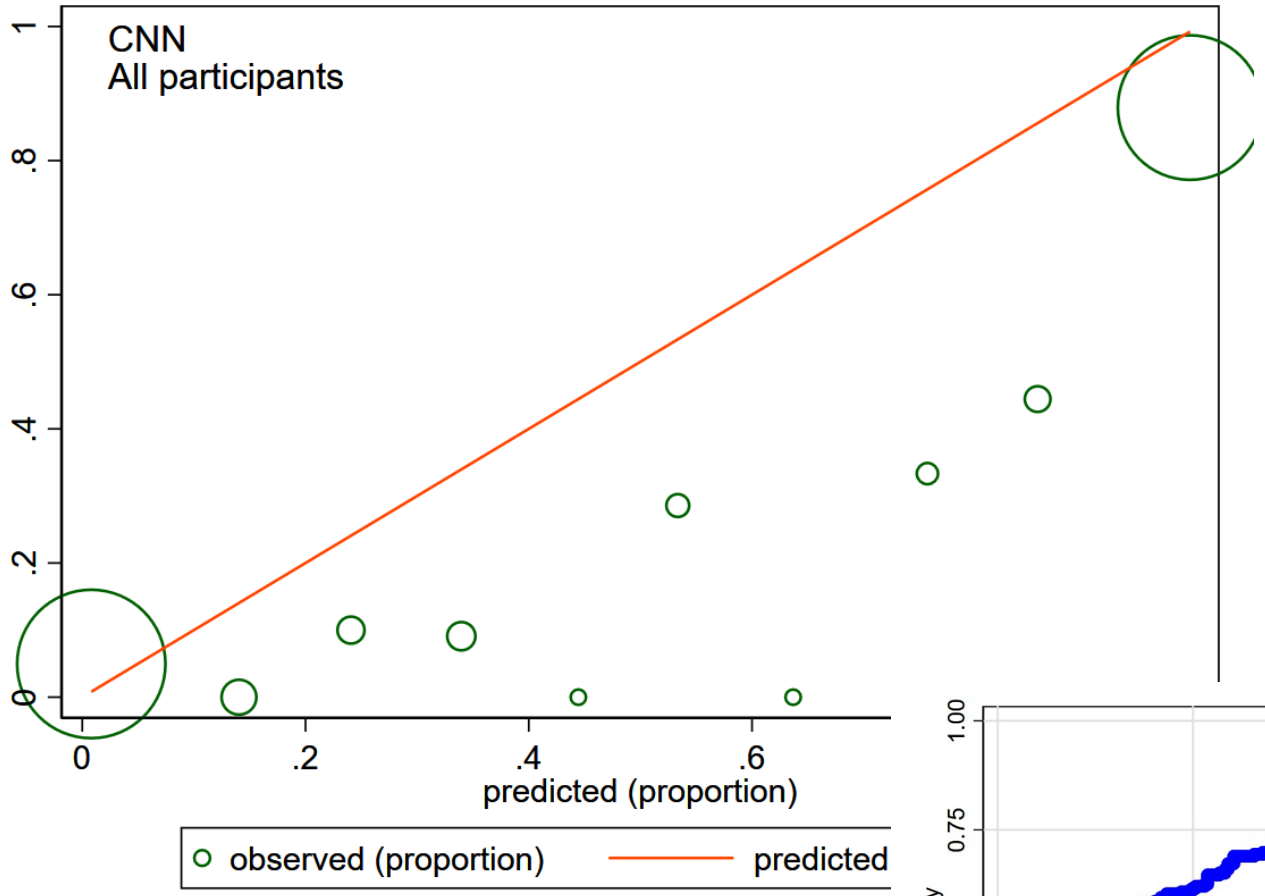

Random Forests
LVEF
LVESDI
LVESVI
LVEDDI
LV lead Position RAO
NT-proBNP
QOL
LVEDVI
Conduction type
ACEI ARB
CRP
6-Min Walk
eGFR
NYHA class
Sodium
Beta-blockers
Aldosterone Antagonists
QRS
Potassium
PR
Heart Rate
Valve disease
BP diastolic
BP systolic
cardiomyopathy type
revascularization
BMI
Height
Weight
Sleep Apnea
Pacemaker
Primary prevention
Autoimmune DS
Cancer
COPD
AV block
CKD
Age
female
Hypertension
Smoking
white
Diabetes
0
.2
.4
.6
.8
1
Importance

Figure 3:


CNN
All participants
0
.2
.4
.6
.8
1
predicted (proportion)
observed (proportion)
predicted

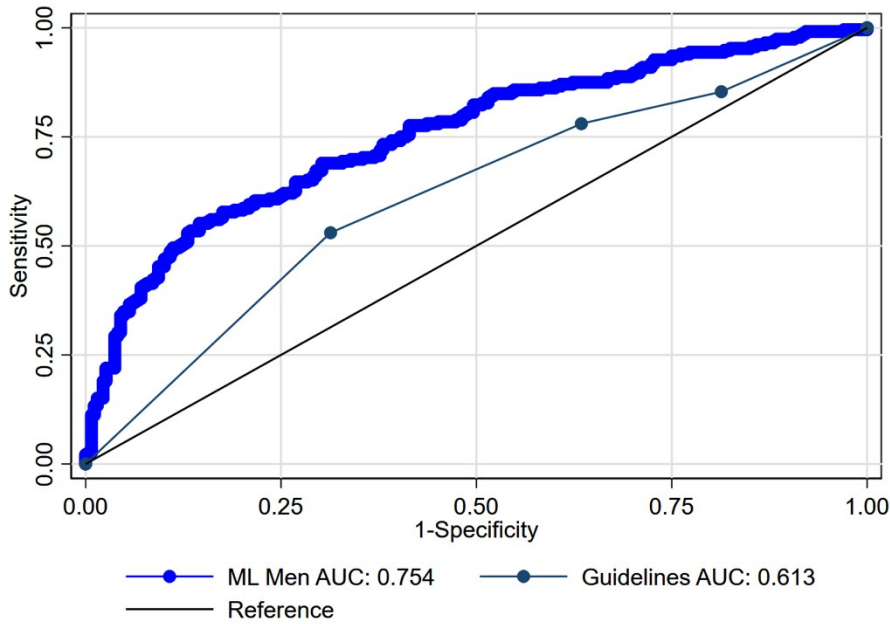

Sensitivity
0.00
0.25
0.50
0.75
1.00
1-Specificity
ML Men AUC: 0.754
Guidelines AUC: 0.613
Reference

Figure 4:

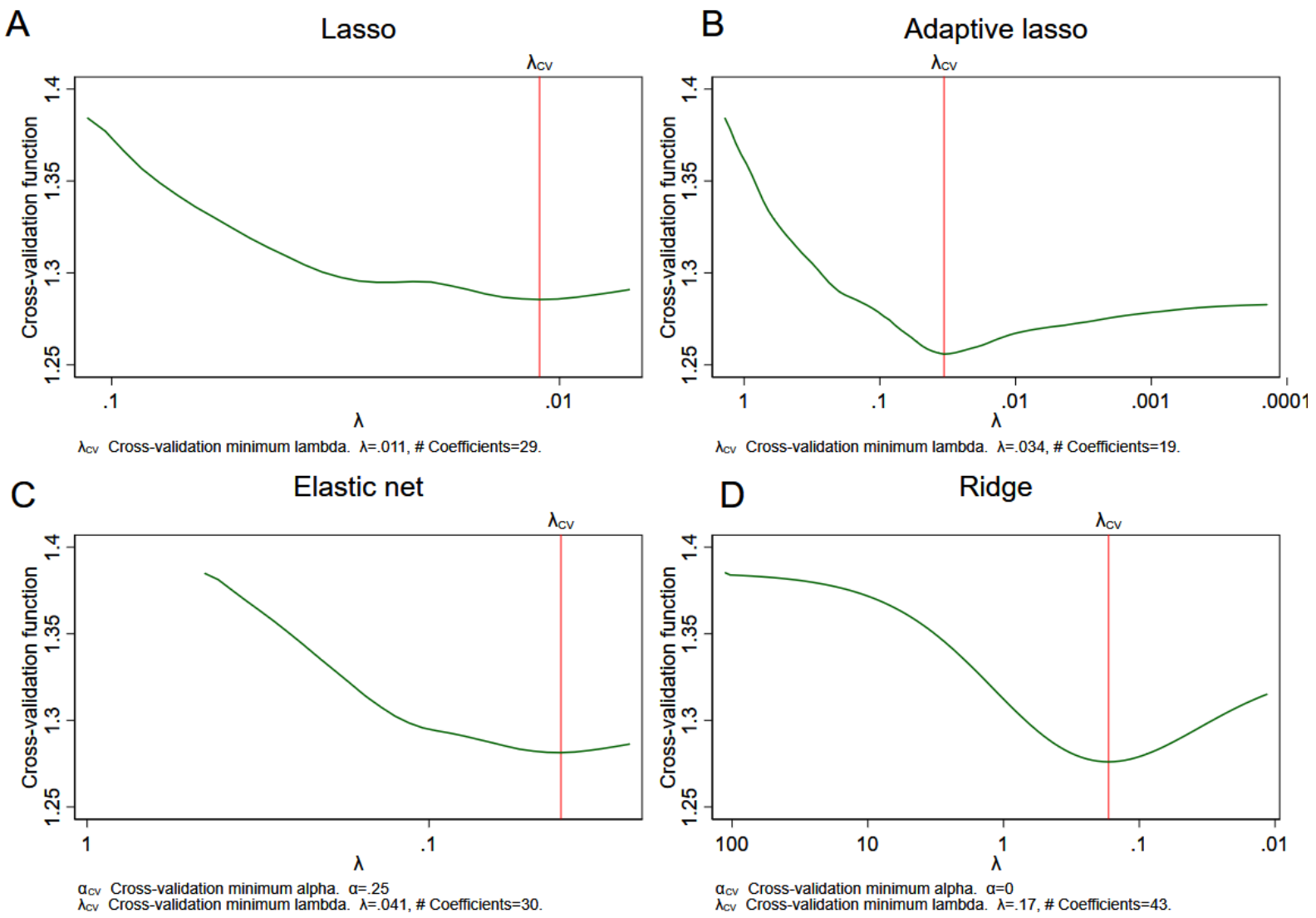

A
Lasso
B
Adaptive lasso
C
Elastic net
D
Ridge
λcv
Cross-validation function
λ
λcv Cross-validation minimum lambda. λ=.011, # Coefficients=29.
λcv Cross-validation minimum lambda. λ=.034, # Coefficients=19.
αcv Cross-validation minimum alpha. α=.25
λcv Cross-validation minimum lambda. λ=.041, # Coefficients=30.
αcv Cross-validation minimum alpha. α=0
λcv Cross-validation minimum lambda. λ=.17, # Coefficients=43.

Figure 5:

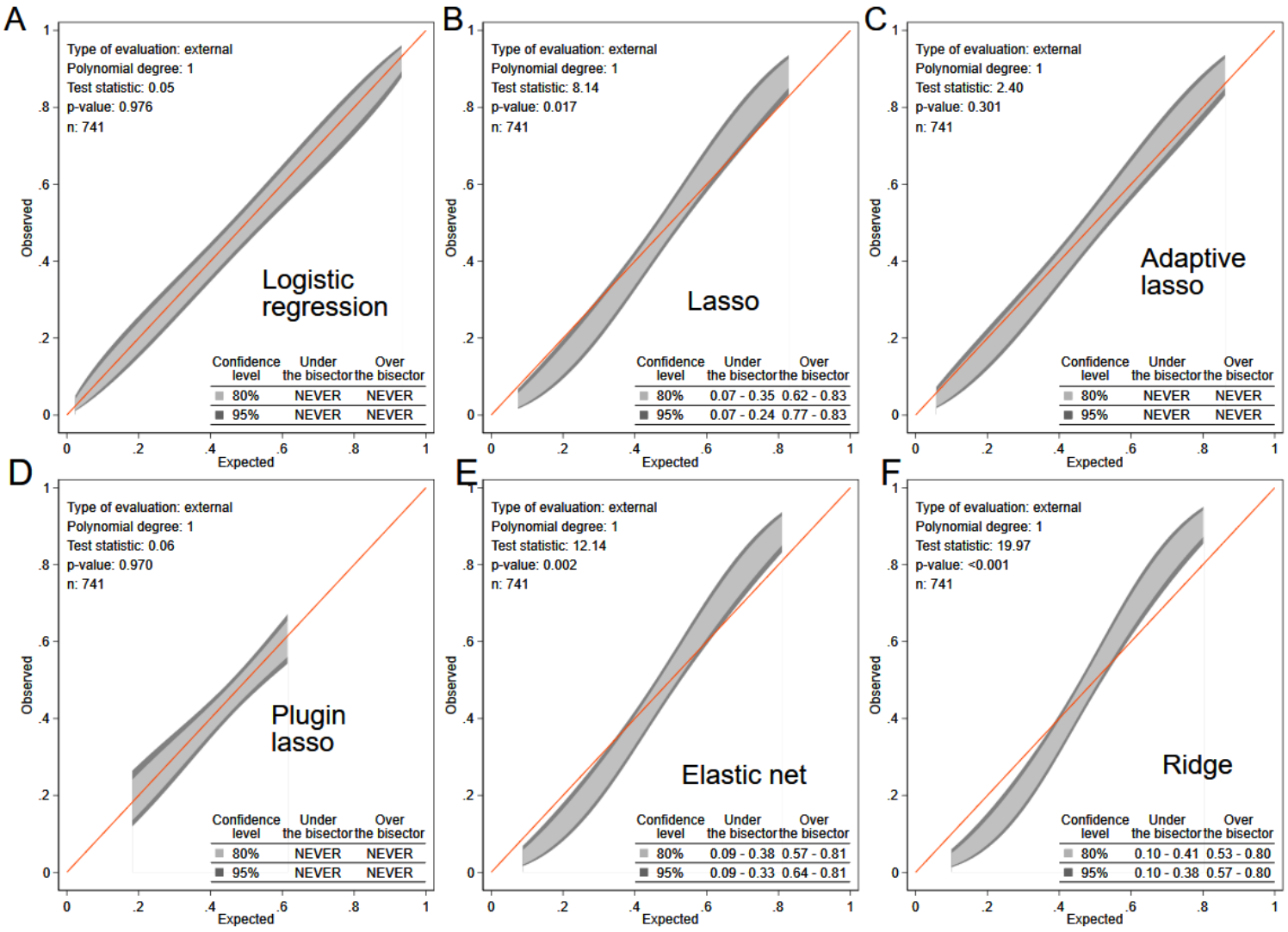

A
Type of evaluation: external
Polynomial degree: 1
Test statistic: 0.05
p-value: 0.976
n: 741
Logistic regression
Confidence level | Under the bisector | Over the bisector
80% | NEVER | NEVER
95% | NEVER | NEVER
B
Type of evaluation: external
Polynomial degree: 1
Test statistic: 8.14
p-value: 0.017
n: 741
Lasso
Confidence level | Under the bisector | Over the bisector
80% | 0.07 - 0.35 | 0.62 - 0.83
95% | 0.07 - 0.24 | 0.77 - 0.83
C
Type of evaluation: external
Polynomial degree: 1
Test statistic: 2.40
p-value: 0.301
n: 741
Adaptive lasso
Confidence level | Under the bisector | Over the bisector
80% | NEVER | NEVER
95% | NEVER | NEVER
D
Type of evaluation: external
Polynomial degree: 1
Test statistic: 0.06
p-value: 0.970
n: 741
Plugin lasso
Confidence level | Under the bisector | Over the bisector
80% | NEVER | NEVER
95% | NEVER | NEVER
E
Type of evaluation: external
Polynomial degree: 1
Test statistic: 12.14
p-value: 0.002
n: 741
Elastic net
Confidence level | Under the bisector | Over the bisector
80% | 0.09 - 0.38 | 0.57 - 0.81
95% | 0.09 - 0.33 | 0.64 - 0.81
F
Type of evaluation: external
Polynomial degree: 1
Test statistic: 19.97
p-value: <0.001
n: 741
Ridge
Confidence level | Under the bisector | Over the bisector
80% | 0.10 - 0.41 | 0.53 - 0.80
95% | 0.10 - 0.38 | 0.57 - 0.80
Observed
Expected
0
.2
.4
.6
.8
1

Figure 6:

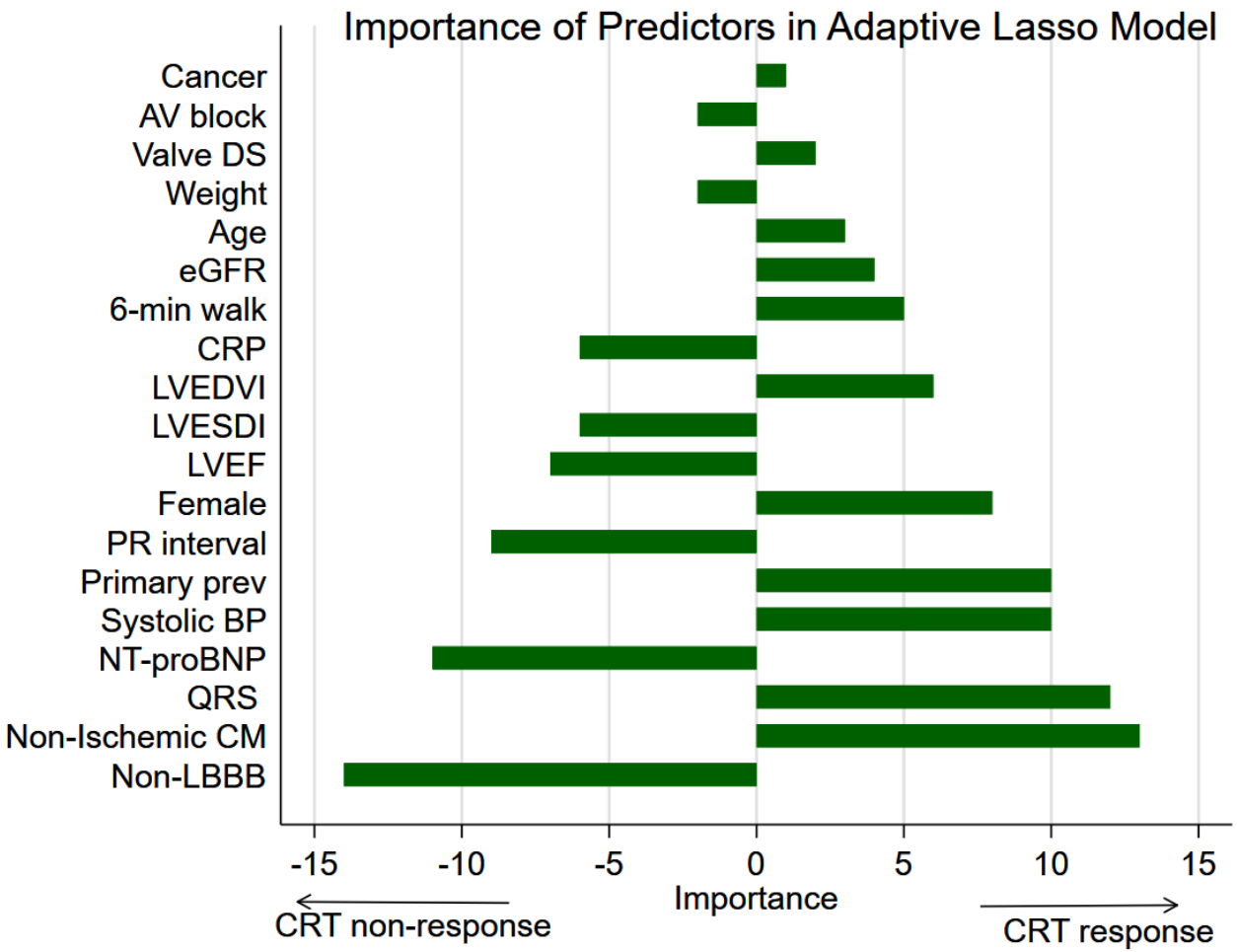
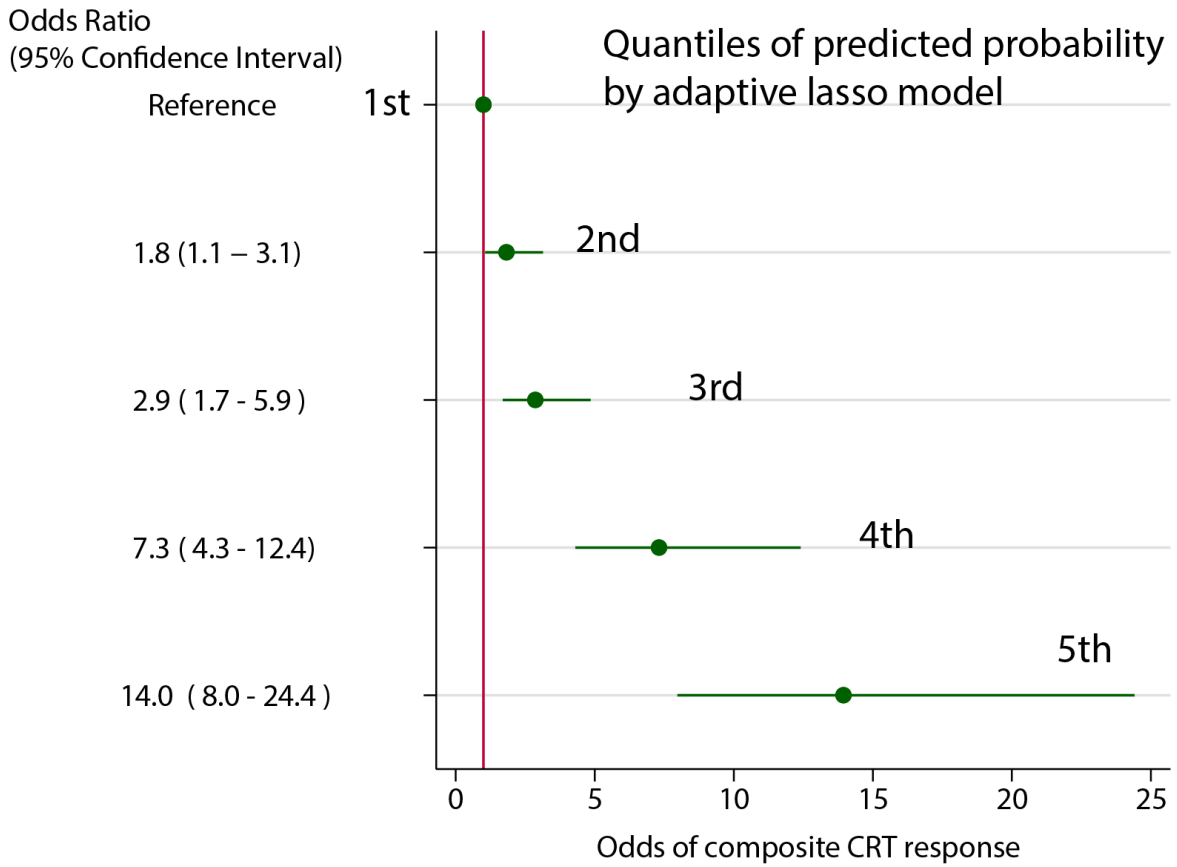

Importance of Predictors in Adaptive Lasso Model
Cancer
AV block
Valve DS
Weight
Age
eGFR
6-min walk
CRP
LVEDVI
LVESDI
LVEF
Female
PR interval
Primary prev
Systolic BP
NT-proBNP
QRS
Non-Ischemic CM
Non-LBBB
-15
-10
-5
0
5
10
15
Importance
CRT non-response
CRT response

Figure 7:


Odds Ratio
(95% Confidence Interval)
Reference
1.8 (1.1 – 3.1)
2.9 (1.7 - 5.9)
7.3 (4.3 - 12.4)
14.0 (8.0 - 24.4)
Quantiles of predicted probability
by adaptive lasso model
1st
2nd
3rd
4th
5th
0
5
10
15
20
25
Odds of composite CRT response