Machine learning for multidimensional response and survival after cardiac resynchronization therapy using features from cardiac magnetic resonance

Derek J. Bivona, PhD,*† Srikar Tallavajhala, BA,* Mohamad Abdi, MS,†
Pim J.A. Oomen, PhD,‡ Xu Gao, MD,§ Rohit Malhotra, MD,* Andrew E. Darby, MD,*
Oliver J. Monfredi, MD, PhD, FHRS,* J. Michael Mangrum, MD,*
Pamela K. Mason, MD, FHRS,* Sula Mazimba, MD,* Michael Salerno, MD, PhD,‖
Christopher M. Kramer, MD,*¶ Frederick H. Epstein, PhD,†¶
Jeffrey W. Holmes, MD, PhD,** Kenneth C. Bilchick, MD, MS, FHRS*

*From the *Department of Medicine, University of Virginia Health System, Charlottesville, Virginia, †Department of Biomedical Engineering, University of Virginia Health System, Charlottesville, Virginia, ‡Department of Biomedical Engineering, University of California, Irvine, California, §Department of Medicine, Northwestern University, Chicago, Illinois, ‖Departments of Medicine and Radiology, Stanford University, Palo Alto, California, ¶Department of Radiology and Medical Imaging, University of Virginia Health System, Charlottesville, Virginia, and **Departments of Medicine, Surgery, and Biomedical Engineering, University of Alabama at Birmingham, Birmingham, Alabama.*

BACKGROUND Cardiac resynchronization therapy (CRT) response is complex, and better approaches are required to predict survival and need for advanced therapies.

OBJECTIVE The objective was to use machine learning to characterize multidimensional CRT response and its relationship with long-term survival.

METHODS Associations of 39 baseline features (including cardiac magnetic resonance [CMR] findings and clinical parameters such as glomerular filtration rate [GFR]) with a multidimensional CRT response vector (consisting of post-CRT left ventricular end-systolic volume index [LVESVI] fractional change, post-CRT B-type natriuretic peptide, and change in peak VO_2) were evaluated. Machine learning generated response clusters, and cross-validation assessed associations of clusters with 4-year survival.

RESULTS Among 200 patients (median age 67.4 years, 27.0% women) with CRT and CMR, associations with more than 1 response parameter were noted for the CMR CURE-SVD dyssynchrony parameter (associated with post-CRT brain natriuretic peptide [BNP] and LVESVI fractional change) and GFR (associated with peak VO_2 and post-CRT BNP). Machine learning defined 3 response clusters: cluster 1 (n = 123, 90.2% survival [best]), cluster 2 (n = 45, 60.0% survival [intermediate]), and cluster 3 (n = 32, 34.4% survival [worst]). Adding the 6-month response cluster to baseline features improved the area under the receiver operating characteristic curve for 4-year survival from 0.78 to 0.86 ($P = .02$). A web-based application was developed for cluster determination in future patients.

CONCLUSION Machine learning characterizes distinct CRT response clusters influenced by CMR features, kidney function, and other factors. These clusters have a strong and additive influence on long-term survival relative to baseline features.

KEYWORDS Machine learning; Magnetic resonance imaging; Heart failure; Cardiac resynchronization therapy; Implantable cardioverter-defibrillator

(Heart Rhythm O^2 2022;3:542–552)

Introduction

Although there have been many informative clinical trials of cardiac resynchronization therapy (CRT) for chronic systolic heart failure over the past 2 decades[1–5] and development of alternative approaches based on conduction system pacing,[6,7] a major gap in our understanding of the clinical course post CRT in its different forms is how to weight and integrate different short-term post-CRT response parameters for the purpose of predicting long-term survival. In this regard, clinicians need to know how much diagnostic testing obtained 6–12 months after CRT adds to baseline characteristics and immediate procedural parameters of success, such as the timing of the QRS to the left ventricular (LV) electrogram at the LV pacing site (QLV). This is an issue of critical

Address reprint requests and correspondence: Dr Kenneth Bilchick, UVA Health System, Cardiovascular Division, P.O. Box 800158, Charlottesville, VA 22908. E-mail address: bilchick@virginia.edu.

2666-5018/

https://doi.org/10.1016/j.hroo.2022.06.005

KEY FINDINGS

- Evaluation of a multidimensional cardiac resynchronization therapy (CRT) response outcome based on left ventricular function, peak VO_2, and the neurohormonal axis is feasible using multivariate multiple linear regression and provides a more complete measure of short-term CRT response 6 months after device implantation than any individual single parameter.
- The baseline features (including cardiac magnetic resonance findings) predictive of each of the 3 components of this CRT response assessment have limited overlap, which helps explain the uncoupling frequently observed among response parameters in these 3 domains.
- Machine learning methods including the Gaussian mixture model applied to the multidimensional response assessment can be used to group patients into 3 clusters that identify patients with excellent, intermediate, and very poor long-term survival, respectively.
- The response clusters have additive predictive value relative to baseline features and offer excellent prediction of long-term survival (area under the curve = 0.86), demonstrating that combining response clusters and baseline features provides the most complete prognostic assessment.
- A web-based calculator to identify the expected response cluster and long-term survival of future patients is provided.

importance, as an algorithm to integrate pre-CRT data, procedural data, and short-term response measures could help heart failure specialists determine the appropriate timing for implantation of ventricular assist devices and/or listing for heart transplantation.[8] Short-term response measures of particular interest include the reduction in left ventricular end-systolic volume index (LVESVI),[9,10] improvement in peak VO_2,[11,12] and the serum B-type natriuretic peptide (BNP) level.[13]

Advanced mathematical methods based on machine learning and multivariate statistics are well suited to address this clinical issue. Instead of an analysis based on just 1 scalar (single-value) CRT response parameter as outcome or predictor in a linear regression model, a CRT response vector (consisting of multiple traditional scalar parameter values) from each patient can be used to determine the relative contributions of a multivariate short-term response measure and a patient's baseline features to long-term survival, as well as how much of the short-term response can be explained by baseline features.

High-quality inputs are important for such an analysis, and the dataset used in this study includes cardiac magnetic resonance (CMR) data for LV and right ventricular (RV) volumetric function[14] and strain from Displacement Encoding with Stimulated Echoes (DENSE).[10,15] The focus of this analysis is a cohort undergoing traditional CRT with LV pacing leads, although this machine learning approach can be easily applied to future analyses of cohorts undergoing conduction system pacing. Ultimately the goal is to apply these advanced analytic methods in implementing a patient-centered approach to device therapy in heart failure.

Methods

Study design, cohort selection, informed consent

All patients had LV ejection fraction (EF) 35% or less, New York Heart Association (NYHA) functional class II–IV, and QRS >120 ms, and had a class I or II indication for CRT based on AHA/ACC/HRS guidelines.[16] All patients had CRT defibrillators, with the exception of 1 patient who received a CRT pacemaker. Before receiving CRT implants at the University of Virginia Health System, patients completed intake forms for demographic characteristics, comorbid conditions, and medications, and these data were verified in the electronic health record. They underwent laboratory studies and vital sign measurements along with exercise testing, and received CMR imaging, echocardiography, and electrocardiograms. Optimization of pacing parameters for CRT was left to the operator's discretion; however, the following approach was recommended for all patients. Selection of the quadripolar LV pacing vector for CRT should be based on evaluation of the QLV, the capture threshold, and the presence of phrenic nerve stimulation with candidate pacing vectors. In patients with devices offering synchronized LV pacing (adaptive CRT),[17] this option was recommended if it resulted in an equivalent or shorter QRS duration compared with biventricular pacing. Use of device-based atrioventricular optimization algorithms was encouraged; alternatively, the A-V interval resulting in the shortest paced QRS duration could be selected. Using an offset between left and right ventricular pacing was encouraged if it shortened the paced QRS duration.

At 6 months after CRT implantation, the patients once again underwent echocardiography, laboratory studies, and exercise testing while also being followed for survival with routine interrogations. The main objectives of the study were (1) to identify pre-CRT clinical variables that were strongly associated with post-CRT response measures using multivariate statistical models and (2) to predict long-term survival of CRT patients with short-term response parameters using cluster analysis and logistic regression. All patients provided informed consent for this study, which was approved by the Institutional Review Board for Human Subjects Research at the University of Virginia.

Patient characteristics, laboratory measurements, exercise testing, and imaging

Demographic characteristics (age, sex, and race), comorbid conditions in addition to heart failure (hypertension, atrial fibrillation, chronic kidney disease, diabetes mellitus, prior

coronary artery bypass grafting surgery, and ischemic cardiomyopathy, which was defined as cardiomyopathy associated with prior myocardial infarction [assessed also with late gadolinium enhancement on CMR] and a significant contribution of ischemic heart disease to LV dysfunction), and medications (beta-blockers, angiotensin-converting enzyme inhibitor / angiotensin receptor blockers, loop diuretic usage and dosage, digoxin, and statins) at the time of CRT implantation were documented from intake data during enrollment in the study and information accessible in the electronic medical record. Patients had laboratory studies (including BNP, creatinine, sodium, and hemoglobin), blood pressure assessments, and exercise testing before the CRT procedure. Twelve-lead electrocardiograms were used to calculate the baseline QRS duration and type of conduction delay. Standard 2-dimensional echocardiographic images with Doppler were obtained for all patients at baseline and 6 months after CRT. Volumetric measurements indexed for body surface area were determined using standard methodology based on Simpson's rule. CMR examinations were performed for all patients before CRT and for 38% of patients after CRT. The CMR protocol included steady-state free precession cine imaging, cine DENSE, and late gadolinium enhancement. Cine DENSE was performed in 4 short-axis planes at basal, 2 midventricular, and midapical levels.[18,19] Circumferential strain from 2-dimensional cine DENSE was calculated semiautomatically to determine the CURE-SVD (range, 0–1, 1 = greatest synchrony).[20] In patients with CMR performed 6 months after CRT, CMR cine imaging was used to calculate the change in LV function, while echocardiographic measurements before and after CRT was used for this purpose in other patients.

Post-CRT response measures

The 3 6-month response measures were as follows: (1) fractional change (FC) in the LVESVI (LVESVI-FC = [post-CRT LVESVI – baseline LVESVI] / baseline LVESVI; negative number = favorable response); (2) BNP post-CRT; and (3) the change in peak oxygen output (Δ peak VO_2 = VO_2 post-CRT − VO_2 pre-CRT). The fractional change was used for the LVESVI and the net change was used for the peak VO_2 because these are the standard parameters in the literature. Pre-CRT and post-CRT magnetic resonance images (MRIs) were used to determine the LVESV response in the 38% of patients with post-CRT MRIs, while pre-CRT and post-CRT echocardiograms were used to determine the LVESV response in the remaining 62% of patients. The post-CRT BNP rather than the change in BNP was used because it was considered to be a more meaningful parameter than the change in BNP. These 3 measures were chosen because they reflect distinct aspects of heart failure response to CRT that can be measured objectively, as opposed to more subjective measures of heart failure response such as heart failure questionnaires or symptom scores. In addition, these measures were found to have stronger associations with survival than the Minnesota Living with Heart Failure Score, which was also acquired in these patients but was not included in this analysis for that reason.

Statistical analysis

Missing data

Only 2% of imaging-based parameters (CURE-SVD and ventricular volumes) before CRT implantation were missing and were imputed using the respective median values. The change in peak VO_2 was missing in 20% of patients because some patients had difficulty exercising both before and after CRT. These missing response measures were imputed by iteratively optimizing the covariance matrix and mean vector to get the conditional expected values using the expectation-maximization algorithm[21] for matrix completion.

Multivariate multiple linear regression

Thirty-nine features (including categorical and continuous variables) were identified and used as input for the multivariate linear regression models for the 3 response parameters. Stepwise linear regression models for the 3 response measures were implemented using the *statsmodels* package in Python (Python Software Foundation). The parameters associated with at least 1 of the 3 response measures (based on $P < .1$) were then included in a multivariate multiple linear regression in which the dependent variable was the CRT response vector (post-CRT LVESVI fractional change, post-CRT BNP, and change in peak VO_2). Pillai trace values and F-statistics were determined for each of the input variables. The Supplemental Material (Statistical Methods) describes these statistical methods and others in more detail.

Machine learning

Clustering algorithm, Gaussian mixture model, and survival analysis

Two types of machine learning algorithms for clustering, the k-means method and Gaussian mixture model (GMM), were considered for stratifying patients based on CRT response measures. Ultimately, the GMM was implemented as it yielded superior predictive prognostic information as discussed in the Results. The GMM was optimized to cluster patients based on only the 3 post-CRT response parameters as inputs using the *sklearn* package in Python. The Bayesian information criterion (BIC), a measure that calculates the likelihood of fit with a penalty for model complexity, was used to identify the optimal number of clusters (2–6) and covariance structure (tied, diagonal, or full). χ^2 tests were used for comparisons of categorical variables, and analysis of variance was used for comparisons of continuous variables between the cluster groups. Kaplan-Meier analysis was used to construct the survival curves based on clusters identified with the GMM, while the log-rank test was used to determine the P values for the differences in survival among the clusters.

Table 1 Baseline characteristics and cardiac resynchronization therapy response measures of patient cohort and cluster groups

	Cohort (N = 200)	Group 1 (N = 123)	Group 2 (N = 45)	Group 3 (N = 32)	*P* value
Demographics					
Age, years	67.4 (58.0-74.0)	66.6 (57.0-72.7)	68.0 (62.0-75.0)	68.5 (60.4-75.2)	.8
BMI	28.9 (25.4-33.7)	30.4 (26.6-34.9)	27.6 (23.8-32.4)	25.9 (22.6-30.6)	.007
Weight (kg)	89.4 (75.1-103.0)	92.5 (79.4-104.8)	84.6 (73.0-99.8)	80.3 (69.1-93.1)	.02
Female	54 (27.0)	36 (29.3)	12 (26.7)	6 (18.8)	.5
NYHA heart failure class					.0003
II	73 (36.5)	57 (46.3)	13 (28.9)	3 (9.4)	
III	126 (63.0)	66 (53.7)	32 (71.1)	28 (87.5)	
IV	1 (0.50)	0 (0.0)	0 (0.0)	1 (3.1)	
Race					.1
Black	27 (13.5)	15 (12.2)	4 (8.9)	8 (25.0)	
White/other	173(86.5)	108 (87.8)	41 (91.1)	24 (75.0)	
Comorbid conditions					
Ischemic cardiomyopathy	87 (43.5)	47 (38.2)	27 (60.0)	13 (40.6)	.04
Hypertension	115 (57.5)	79 (64.2)	19 (42.2)	17 (53.1)	.03
Atrial fibrillation	52 (26.0)	36 (29.2)	7 (15.6)	9 (28.1)	.5
Chronic kidney disease	62 (31.0)	35 (28.5)	12 (26.7)	15 (46.9)	.1
Diabetes mellitus	73 (36.5)	44 (35.8)	15 (33.3)	14 (43.8)	.6
Prior CABG	35 (17.5)	21 (17.0)	10 (22.2)	4 (12.5)	.5
Medications					
Beta-blocker	191 (95.5)	117 (95.1)	43 (95.6)	31 (96.9)	.9
ACE inhibitor or ARB	175 (87.5)	111 (90.2)	39 (86.7)	25 (78.1)	.2
Loop diuretic dose, mg					.07
0	58 (29.0)	41 (33.3)	10 (22.2)	7 (21.9)	
20–40	90 (45.0)	57 (46.3)	23 (51.1)	10 (31.2)	
60–80	34 (17.0)	16 (13.0)	7 (15.6)	11(34.4)	
>100	18 (9.0)	9 (7.3)	5 (11.1)	4 (12.5)	
Digoxin	17 (8.5)	7 (5.7)	6 (13.3)	4 (12.5)	.2
Statin	120 (60.0)	76 (61.8)	26 (57.8)	18 (56.3)	.8
Laboratory studies, vital signs, & exercise testing					
Systolic BP, mm Hg	118.0 (104.0-130.0)	119.0 (104.0-128.5)	122.0 (108.0-134.0)	110.0 (103.5-130.0)	.6
Sodium, mEq/L	138.0 (137.0 -140.0)	139.0 (137.0-140.0)	138.0 (137.0-141.0)	137.5 (136.0-140.0)	.2
Creatinine, mg/dL	1.1 (0.9-1.3)	1.0 (0.9-1.2)	1.2 (1.0-1.4)	1.3 (1.1-1.5)	.0003
Hemoglobin, g/dL	13.3 (12.3-14.7)	13.6 (12.5-14.8)	13.5 (12.3-14.9)	12.8 (12.1-14.1)	.2
GFR, mL/min/1.72 m^2	67.2 (54.1-84.1)	74.0 (60.1-88.1)	62.0 (51.0-70.7)	59.2 (45.5-74.4)	.0003
BNP, pg/mL	272.0 (130.0-632.3)	190.0 (92.5-298.5)	351.0 (220.0-654.0)	1147.5 (820.8-2550.0)	<.0001
Peak VO_2, mL/kg/min	14.4 (12.5-15.7)	14.4 (12.9-16.5)	14.3 (11.2-15.5)	14.0 (12.0-14.4)	.02
CMR & echocardiography assessment parameters					
LVEF, %	24.0 (17.7-30.5)	25.9 (19.0-31.0)	22.9 (17.2-28.6)	19.7 (14.4-24.8)	.01
LVEDVI, mL/m^2	126.3 (102.5-157.0)	117.0 (98.7-137.0)	134.7 (115.3-158.0)	167.9 (122.9-195.5)	<.0001
LVESVI, mL/m^2	93.7 (73.7-123.6)	88.4 (68.7-108.9)	100.7 (85.6-127.0)	134.2 (91.4-167.0)	<.0001
RVEF, %	37.5 (25.8-45.6)	37.9 (31.3-47.4)	38.3 (20.7-45.2)	27.3 (18.7-39.7)	.02
RVEDVI, mL/m^2	65.8 (52.9-83.1)	60.0 (48.8-77.0)	67.2 (59.6-81.0)	99.6 (76.7-121.5)	<.0001
RVESVI, mL/m^2	38.8 (29.9-55.5)	36.3 (26.5-46.5)	38.8 (33.3-56.0)	70.7 (48.8-91.2)	<.0001
LGE presence	95 (47.5)	52 (42.2)	25 (55.6)	18 (56.3)	.2
CURE-SVD	0.59 (0.45-0.76)	0.55 (0.41-0.74)	0.62 (0.45-0.72)	0.76 (0.58-0.85)	.001
ECG parameters					
QRS, ms	158 (142-175)	160.0 (147.0-175.5)	152.0 (138.0-160.0)	158.0 (139.0-180.0)	.3
QLV, ms	120.0 (87.0-149.3)	125.0 (92.5-150.0)	105.0 (80.0-130.0)	102.5 (73.8-150.0)	.04
LBBB	151 (75.5)	92 (74.8)	38 (84.4)	21 (65.6)	.2
RBBB	22 (11.0)	12 (9.8)	2 (4.4)	8 (25.0)	.01
Paced rhythm	28 (14.0)	19 (15.4)	4 (8.9)	5 (15.6)	.5
Upgrade or new device					.4
De novo device	153 (76.5)	90 (73.2)	37 (82.2)	26 (81.2)	
Upgrade device	47 (23.5)	33 (26.8)	8 (17.8)	6 (18.8)	
Response measures at 6 months post CRT					
Fractional change in LVESVI	-0.18 (-0.33 - -0.01)	-0.24 (-0.42 - -0.1)	-0.1 (-0.21-0.05)	0.025 (-0.13-0.13)	<.0001
BNP, pg/mL	177.0 (59.5-592.0)	77.0 (30.0-145.0)	524.0 (367.0-741.0)	2088.0 (1550.0-2845.8)	<.0001

(*Continued*)

Table 1 (*Continued*)

	Cohort (N = 200)	Group 1 (N = 123)	Group 2 (N = 45)	Group 3 (N = 32)	*P* value
Change in peak VO_2, mL/kg/min	0.0 (-1.0-1.2)	0.30 (-0.8-1.7)	-0.20 (-1.5-0.5)	-0.37 (-2.5 - -0.12)	.005
Survival status at 4 years					<.0001
Alive	149 (74.5)	111 (90.2)	27 (60.0)	11 (34.4)	
Dead	51 (25.5)	12 (9.8)	18 (40.0)	21 (65.6)	

Values are median (interquartile range) or n (%).

ACE = angiotensin-converting enzyme; ARB = angiotensin receptor blocker; BMI = body mass index; BNP = B-type natriuretic peptide; BP = blood pressure; CABG = coronary artery bypass graft; CURE-SVD = circumferential uniformity ratio estimate with singular value decomposition; GFR = glomerular filtration rate; LBBB = left bundle branch block; LGE = late gadolinium enhancement; LVEDVI = left ventricular end-diastolic volume index; LVEF = left ventricular ejection fraction; LVESVI = left ventricular end-systolic volume index; NYHA = New York Heart Association; QLV = QRS-LV electrogram time; RBBB = right bundle branch block; RVEDVI = right ventricular end-diastolic volume index; RVEF = right ventricular ejection fraction; RVESVI = right ventricular end-systolic volume index.

Cross-validation methodology

Using a 5-fold cross-validation framework, multivariable logistic regression models were used to predict survival at 4 years after CRT. In the first model, the input data consisted of the pre-CRT parameters with the strongest associations with survival (pre-CRT CURE-SVD from CMR, pre-CRT BNP levels, and pre-CRT peak VO_2) based on a separate regression model. In the second model, the response cluster was added to these 3 baseline variables. Receiver operating characteristic (ROC) curves for each fold within the cross-validation for both models were generated, and areas under the ROC curves (AUC) were calculated to evaluate and compare model performances.

Results

Baseline characteristics

Demographic characteristics, comorbid conditions, medications, laboratory findings, vital signs, exercise testing results, and imaging findings for the cohort of 200 patients (aged 66.1 ± 11.4 years; 27.0% female; 13.5% Black) are shown in Table 1. The median change in the LVESVI-FC following CRT was -0.18 (interquartile range [IQR] -0.33 to -0.01). Of note, the median and IQR for the LVESVI-FC were nearly identical for the 38% of patients with calculation based on MRI before and after CRT (-0.18; -0.34 to -0.03) and the 62% with the calculation based on echocardiography before and after CRT (-0.18; -0.33 to 0.02; $P = .5$). A total of 56.0% of patients had 15% or greater reduction in the LVESVI post CRT (LVESVI-FC ≤-0.15). The median BNP level was 177.0 pg/mL (IQR 59.5–592.0 pg/mL), and the median change in the peak VO_2 was 0.0 mL/kg/min (IQR -1.0 to 1.2). During a median follow-up of 4 years, 51 (25.5%) patients died. The distribution of LV pacing site was typical of other CRT studies and is shown in Supplemental Figure 1.

Identification of baseline features associated with short-term response vector

The stepwise linear regression analyses showed that different pre-CRT parameters were associated with each of the 3 response variables of interest, which describe cardiac function (LVESV-FC), neurohormonal activity (post-CRT BNP), and oxygen utilization (Δ peak VO_2). The pre-CRT parameters most associated with each response feature are shown in Table 2.

The pre-CRT variables most significantly associated with the LVESVI-FC (Table 2, section A) were the CURE-SVD ($P < .0001$), QLV ($P = .0003$), RVEF ($P = .0008$), ischemic cardiomyopathy ($P = .02$), and diabetes mellitus ($P = .05$). A lower CURE-SVD value (more dyssynchrony) was associated with greater response based on the positive coefficient. The CURE-SVD was also confirmed to be uncorrelated with the RVEF (r = -0.014, $P = .85$).

The pre-CRT parameters most associated with post-CRT BNP levels (Table 2, section B) were the baseline BNP ($P < .0001$), CURE-SVD ($P < .0001$), LVEDVI ($P < .0001$), glomerular filtration rate (GFR) ($P = .0005$), NYHA classification ($P = .02$), and diabetes mellitus ($P = .03$). As in the model for LVESVI-FC, the lower CURE-SVD score was associated with a lower post-CRT BNP.

The pre-CRT variables most associated with the change in peak VO_2 (Table 2, section C) were peak VO_2 at baseline ($P < .0001$), GFR ($P = .0002$), and systolic blood pressure ($P = .03$). Based on the positive coefficient, better kidney function at baseline had a strong association with favorable change in peak VO_2. As a sensitivity analysis, we also show that this model based on the matrix completion algorithm for missing peak VO_2 data points chooses the same covariates, gives nearly identical *P* values, and has standardized regression coefficients agreeing within a margin of 10%–15%, compared with the model in Table 2, section D, which is based on only patients with measured VO_2 before and after CRT.

The baseline features most associated with the overall response vector were the pre-CRT BNP ($P < .0001$), the pre-CRT CURE-SVD ($P < .0001$), GFR ($P < .0001$), the pre-CRT peak VO_2 ($P = .0001$), LVEDVI ($P = .0007$), NYHA classification ($P = .003$), QLV ($P = .006$), RVEF ($P = .03$), the presence of ischemic cardiomyopathy ($P = .04$), systolic blood pressure ($P = .09$), and the presence of diabetes mellitus ($P = .1$) (Table 3).

Response clusters and associations with survival

The GMM that resulted in the lowest BIC score included 3 clusters with a diagonal covariance structure as demonstrated in Figure 1. Table 4 shows the mean vector and variance vector (composed of diagonal values of the covariance matrix) for the mixture clusters, and the 3 clusters of patients are displayed in Figure 2. The summary of the response measures (along with the baseline characteristics not used for

Table 2 Stepwise linear regression models for the 3 response measures

(A) Model variable	Multivariable linear regression model for Δ LVESVI			
	Model coefficient	Standard error	*P* value	Standardized coefficient
Intercept	-0.0920	0.0827	.3	0
CURE-SVD	0.354	0.0749	<.0001	0.301
QLV	-0.00149	0.000402	.0003	-0.238
RVEF	-0.00346	0.00102	.0008	-0.210
Ischemic cardiomyopathy	0.0700	0.0308	.02	0.141
Diabetes mellitus	-0.0628	0.0315	.05	-0.123
$R^2 = 0.29$; adjusted $R^2 = 0.27$				

(B) Model variable	Multivariable linear regression model for BNP			
	Model coefficient	Standard error	*P* value	Standardized coefficient
Intercept	-792	292	.007	0
BNP at baseline	0.659	0.0545	<.0001	0.600
CURE-SVD	864	183	<.0001	0.211
LVEDVI	3.68	0.854	<.0001	0.205
GFR	-7.09	1.99	.0005	-0.166
NYHA classification	195	79.7	.02	0.111
Diabetes mellitus	-186	82.9	.03	-0.105
$R^2 = 0.62$; adjusted $R^2 = 0.61$				

(C) Model variable	Multivariable linear regression model for Δ peak VO_2			
	Model coefficient	Standard error	*P* value	Standardized coefficient
Intercept	2.82	1.31	.03	0
Peak VO_2 at baseline	-0.176	0.0419	<.0001	-0.287
GFR	0.0298	0.00796	.0002	0.254
SBP	-0.0187	0.00859	.03	-0.146
$R^2 = 0.13$; adjusted $R^2 = 0.12$				

(D) Model variable	Multivariable linear regression model for Δ peak VO_2: complete case analysis (153/200) vs original analysis (imputed data, 200/200)			
	Standardized coefficient (complete case analysis)	Standardized coefficient (original analysis)	*P* value (complete case analysis)	*P* value (original analysis)
Peak VO_2 at baseline	-0.337	-0.287	<.0001	<.0001
GFR	0.300	0.254	.0001	.0002
SBP	-0.163	-0.146	.03	.03

BNP = B-type natriuretic peptide; GFR = glomerular filtration rate; LVEDVI = left ventricular end-diastolic volume index; LVESVI = left ventricular end-systolic volume index; NYHA = New York Heart Association; QLV = QRS-LV electrogram time; RVEF = right ventricular ejection fraction; SBP = systolic blood pressure.

clustering) for the patients within each cluster is shown in Table 1.

The GMM clustered 123 patients into group 1, and this group had a median LVESVI-FC of -0.24 (IQR -0.42 to -0.1). A total of 83 patients (67.5%) within this group met echocardiographic criteria for favorable CRT response. The median post-CRT BNP was 77.0 pg/mL (IQR 30.0–145.0 pg/mL), and the median change in peak VO_2 was 0.30 mL/kg/min (IQR -0.8 to 1.7 mL/kg/min) in this group.

The GMM assigned 45 patients into group 2, which had a median LVESVI-FC of -0.1 (IQR -0.21 to 0.05). Twenty-one patients (46.7%) within this group had at least a 15% reduction in the LVESVI, the median BNP level was 524.0 pg/mL (IQR 367.0–741.0 pg/mL), and the median change in peak VO_2 was -0.2 mL/kg/min (IQR -1.5 to 0.5 mL/kg/min) in this group.

The GMM assigned 32 patients into group 3, and this group had a median percent reduction in the LVESVI of 0.025 (IQR -0.13 to 0.13). Only 8 patients (25.0%) within this group had a 15% or greater reduction in the LVESVI, the median post-CRT BNP level was 2088.0 pg/mL (IQR 1550.0–2845.8 pg/mL), and the median change in peak VO_2 was -0.37 mL/kg/min (IQR -2.5 to -0.12 mL/kg/min).

While the baseline characteristics of the cohort were not used to generate the response clusters, we did note significant differences among groups ($P < .05$) for the following baseline variables: BMI, NYHA classification, ischemic cardiomyopathy, hypertension, GFR, BNP at baseline, peak VO_2 at baseline, LV and RV volumetric parameters, CURE-SVD, and ECG parameters.

The Kaplan-Meier survival analysis is displayed in Figure 3 and demonstrates that patients in group 1 had

Table 3 Multivariate multiple linear regression for response vector

Model variable	Pillai trace value	F statistic	*P* value
BNP at baseline	0.451	50.7	<.0001
CURE-SVD	0.148	10.7	<.0001
GFR	0.109	7.51	<.0001
Peak VO_2 at baseline	0.107	7.37	.0001
NYHA classification	0.102	3.35	.003
LVEDVI	0.087	5.98	.0007
QLV	0.064	4.23	.006
RVEF	0.046	3.03	.03
Ischemic cardiomyopathy	0.045	2.89	.04
SBP	0.034	2.18	.09
Diabetes mellitus	0.033	2.10	.1

BNP = B-type natriuretic peptide; GFR = glomerular filtration rate; LVEDVI = left ventricular end-diastolic volume index; NYHA = New York Heart Association; QLV = QRS-LV electrogram time; RVEF = right ventricular ejection fraction; SBP = systolic blood pressure.

excellent survival, patients in group 2 had intermediate survival, and patients in group 3 had the markedly poor survival. The hazard ratio (HR) for cluster 2 (intermediate survival) vs cluster 1 (best survival) was 1.84 [95% confidence interval (CI): 1.35–2.50; $P < .001$]; the HR for cluster 3 (worst survival) vs cluster 1 (best survival) was 2.23 [95% CI: 1.71–2.90; $P < .001$]; and the HR for cluster 3 (worst survival) vs cluster 2 (intermediate survival) was 1.31 [95% CI: 1.04–1.65; $P < .05$]. Of note, the survival curves based on the clusters generated with the GMM showed more separation than those associated with the k-means clusters, providing justification for use of the GMM over k-means to generate the response clusters.

Effect of adding 6-month response cluster to baseline features for prediction of long-term survival

The ROC curves of the logistic regression model with the CURE-SVD score, pre-CRT BNP levels, and pre-CRT peak VO_2 levels as inputs (baseline features most strongly

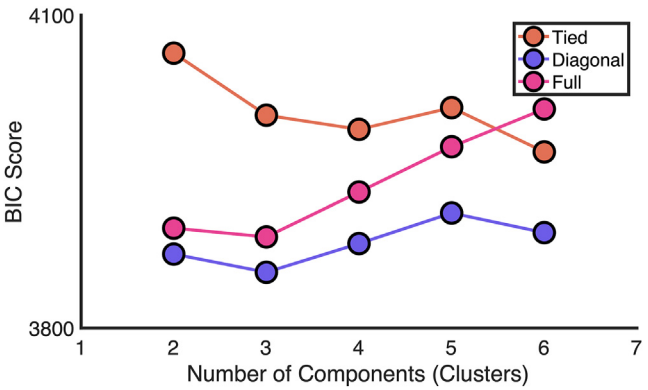

Figure 1 Gaussian mixture model (GMM) selection. A GMM with each covariance type and number of clusters ranging from 2 to 6 was generated. The Bayesian information criterion (BIC) score of each model was calculated and is shown above. For any number of clusters, the GMM with a diagonal covariance structure exhibited favorable BIC scores. The model with 3 clusters and a diagonal covariance structure had the lowest (most favorable) BIC score.

Table 4 Mean vector and variance vector for Gaussian mixture model clusters

	Mean vector			Variance vector		
	LVESVI-FC	BNP (pg/mL)	Δ Peak VO_2 (mL/kg/min)	LVESVI-FC	BNP	Δ Peak VO_2
Group 1	-0.265	94.8	0.47	0.0516	5333.0	5.24
Group 2	-0.0844	499.6	0.0354	0.0316	7.32E+04	5.15
Group 3	-0.00181	2101.6	-0.914	0.0618	9.78E+05	5.60

BNP = B-type natriuretic peptide; LVESVI-FC = left ventricular end-systolic volume index fractional change.

associated with the short-term response vector) are shown in Figure 4A. The AUC for each fold within the cross-validation is displayed along with the area under the average ROC curve (0.78 ± 0.04). The effect of adding the 6-month response cluster to baseline features is shown in Figure 4B. For each fold within the cross-validation, the AUC increased, and the area under the average ROC curve for this model (0.86 ± 0.021) was also higher compared with the model based on baseline features alone ($P = .02$).

Discussion

A major finding of this study was that machine learning could be used to discern 3 short-term response clusters based on 3 6-month parameters of CRT response (LVESVI-FC, post-CRT BNP, Δ peak VO_2), and these response clusters, in turn, were highly associated with survival times after CRT. Furthermore, the 61.5% of patients within this cohort that made up response cluster 1 had excellent 4-year survival (>90%), while the 22.5% in cluster 2 had intermediate survival, and the 16% in cluster 3 had very poor 4-year survival (<35%). In the future, this type of analysis could be incorporated into clinical management by increasing follow-up frequency and lowering the threshold for referral of cluster 3 patients for LV assist device or transplant consideration. An interesting observation with respect to the designation of the response clusters was that the logistic regression model with the response clusters performed at least as well as or better than the model with each of the 3 short-term response parameters used as separate covariates, highlighting the utility of using machine learning to identify response clusters.

We show in this study that not only are the short-term response measures predictive of long-term survival, but also the response cluster covariate provides substantial improvement in performance when added to baseline and procedural findings. Although the response vector had additional predictive value for survival apart from baseline findings, the specific baseline findings that influenced the overall response vector are also of particular interest. In particular, the following 9 baseline/procedural findings were associated with the overall multivariate response vector, here listed in the order of strength of association:

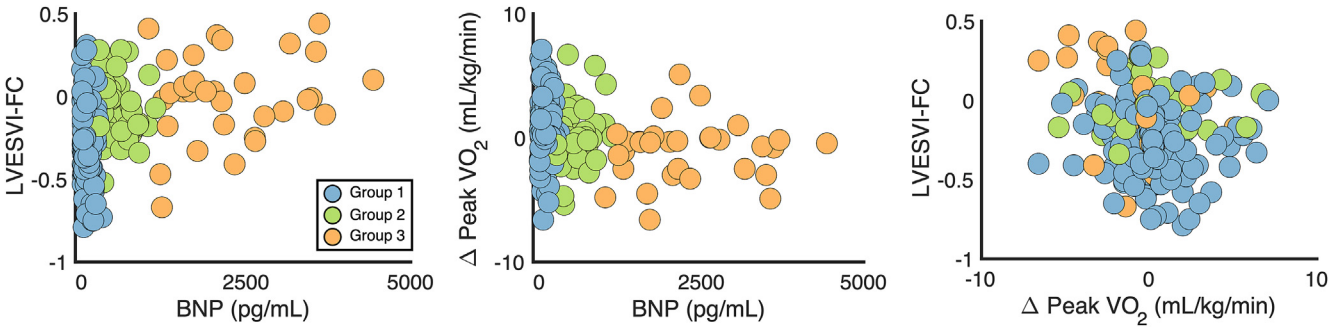

Figure 2 Gaussian mixture model (GMM) cluster analysis. The GMM stratified the patients (based on their cardiac resynchronization therapy [CRT] response measures) into 3 distinct groups. The clusters are plotted as 2-dimensional sets of the 3 response measures. Each point represents 1 patient with the respective left ventricular end-systolic volume index fractional change (LVESVI-FC), post-CRT B-type natriuretic peptide (BNP), and Δ peak VO_2 values and is colored based on the assigned cluster.

BNP at baseline, CURE-SVD, peak VO_2 at baseline, GFR, LVEDVI, QLV, NYHA classification, ischemic cardiomyopathy, and RVEF. Of particular interest, no baseline characteristic or procedural finding was a statistically significant covariate in the 3 separate linear regression models for the 3 scalar CRT response parameters, which highlights the complexity of CRT response and the potential value of machine learning clustering methods and multiple multivariate regression in this area. Of these 9 baseline/procedural parameters, 2 were found to have associations with 2 of the 3 response measures: renal function and the CMR CURE-SVD parameter.

The GFR was associated with both the post-CRT BNP response measure and the peak VO_2, as patients with a lower GFR tended to have higher post-CRT BNPs even after adjustment for the baseline BNP and a lower improvement in the peak VO_2. The former finding can be explained by the effect of chronic kidney disease on the neurohormonal axis in heart failure (that involves both secretion and clearance of BNP).[22] Although chronic kidney disease affects N-terminal proBNP more than BNP, BNP has still been shown to be influenced by chronic kidney disease.[23] With respect to the association of GFR and the change in VO_2, increasing stages of chronic kidney disease have been associated with lower peak VO_2,[24] likely related to systemic oxygen delivery factors. Right ventricular dysfunction, which often coexists with chronic kidney disease, has also been implicated as a mechanism to explain the association of chronic kidney disease with lower peak VO_2.[25] In this sense, it is interesting that RV function by CMR had an independent association with the reduction in LV end-systolic volume in the present study. The association of GFR with the peak VO_2 and RVEF with LVESVI-FC may be explained by a stronger association of GFR vs RVEF with peak VO_2 and the association of RV dysfunction with a lower GFR.

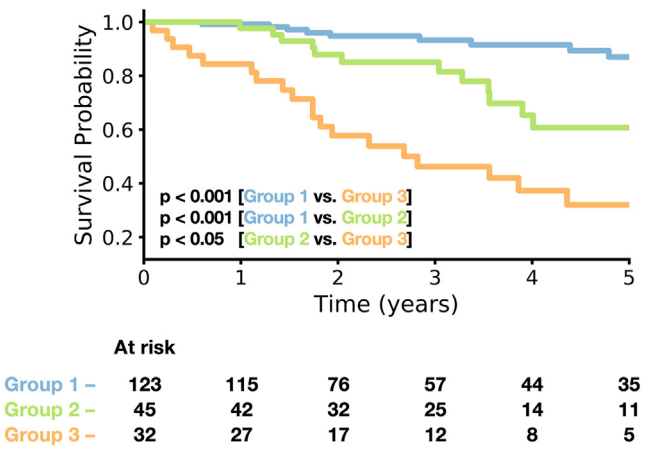

Figure 3 Kaplan-Meier analysis for each cluster. Kaplan-Meier curves demonstrating the probability of survival following cardiac resynchronization therapy implantation are presented for each of the 3 cluster groups. Patients in group 1 have a greater survival probability than patients in group 2 ($P < .001$) and in group 3 ($P < .001$). Patients in group 2 have a greater survival probability than patients in group 3 ($P < .05$).

The CMR CURE-SVD dyssynchrony parameter was also associated with 2 of the 3 response measures, but not the same ones as the GFR. A lower CURE-SVD was associated with both the favorable LV functional response (LVESVI-FC) and the post-CRT BNP, but not the peak VO_2. The CURE-SVD measures the extent of simultaneous contraction (negative circumferential strain) and stretch (positive circumferential strain) in opposing LV segments with CMR using a parameter with range 0–1, such that values trending toward 0 indicate greater dyssynchrony. Our group has demonstrated the robustness of the CURE-SVD for predicting LV functional improvement in CRT in prior cohorts[10,14,15,26]; however, the association with the neurohormonal axis is a new finding. This association with the neurohormonal axis provides an important mechanistic insight for how baseline CMR-based mechanical dyssynchrony and strain findings influence the neurohormonal axis in heart failure and long-term survival. As noted, the CURE-SVD is a pre-CRT patient characteristic, while the QLV is a procedural parameter indicating late activation of LV pacing site, which is a desired result in traditional CRT. In this regard, it is interesting that the CURE-SVD at baseline modified both the LV functional and neurohormonal response, while the QLV was primarily associated with the LV functional response in multivariate analysis.

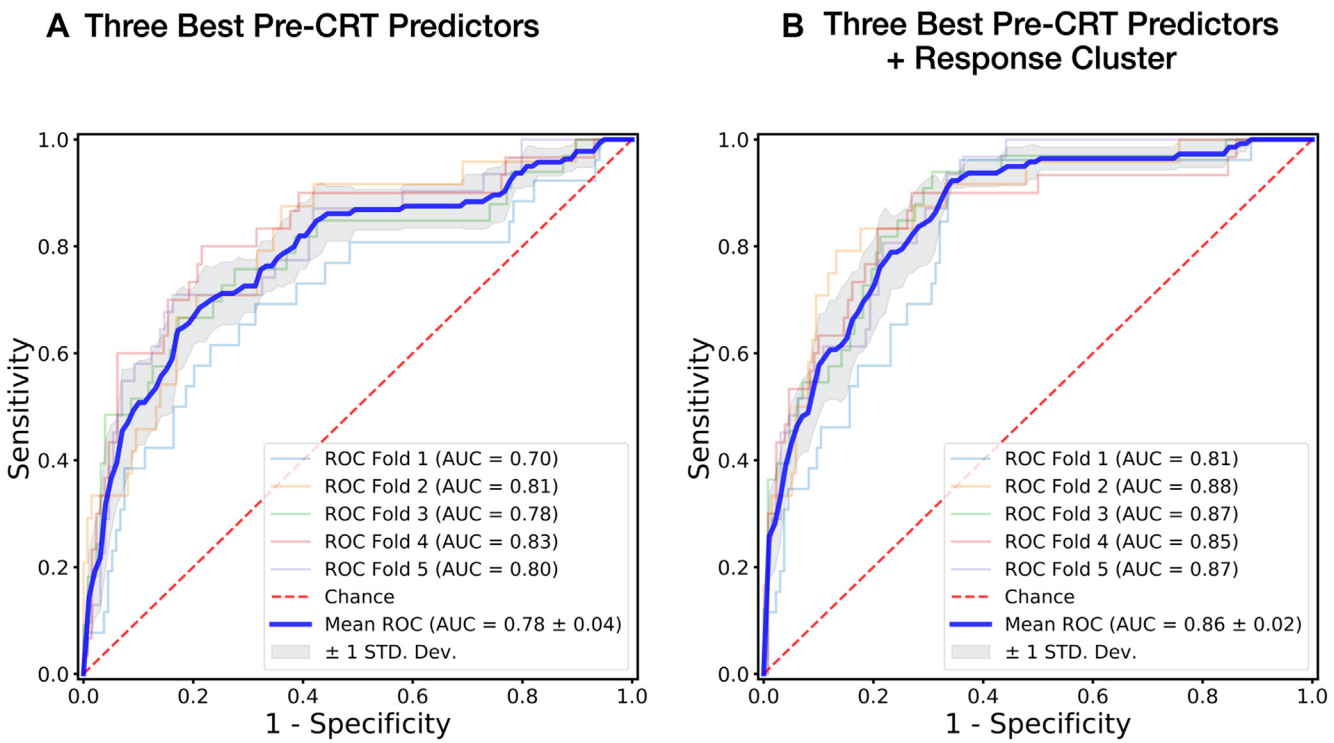

Figure 4 Receiver operating characteristic (ROC) curves for 4-year survival prediction. ROC curves from the logistic regression models are shown. **A:** Three best pre–cardiac resynchronization therapy (CRT) predictors. The CURE-SVD, pre-CRT B-type natriuretic peptide, and pre-CRT peak VO_2 are used as inputs. **B:** Three best pre-CRT predictors plus response cluster. The response clusters were added to the model in panel A. The addition of the short-term response clusters to these baseline features improved prediction of long-term response (area under the curve [AUC] 0.86 vs 0.78, $P = .02$).

Our study presents novel findings relative to 2 other recent papers on applications of machine learning for CRT[27,28] and is substantially different from these other studies in the following ways. First, the analysis based on the SMART-AV study uses a single binary response parameter based on a composite outcome,[27] and the other analysis uses a single binary response parameter based on whether LVEF improved by 10%.[28] In contrast, we used a multidimensional response outcome with 3 continuous variables representing different aspects of CRT response, which is then incorporated into a prediction model for long-term survival after CRT. Second, our study includes baseline CMR volumetric, strain, and scar data for all patients, whereas the other studies do not incorporate CMR data. Third, both of the other studies use baseline characteristics only, while our study demonstrates that addition of short-term response clusters to the baseline predictors resulted in an increased AUC of 0.86. Fourth, we provide an online calculator based on our model for estimation in future patients of long-term survival after CRT without an LV assist device or heart transplantation.

The results of this study could be applied in clinical practice by determining the statistical likelihood that a patient with a given short-term response vector belongs to each of the 3 clusters defined in this study, and then assigning the patient to the cluster for which the response vector had the highest likelihood. This can be done by using the trained/optimized GMM with a given response vector as input. We developed a web-based application that allows the user to input patient metrics, calculates the response vector, and predicts the cluster to which a new patient belongs using the trained GMM. This program can be accessed at http://gmmxcrt.pythonanywhere.com with username *tester* and password *BilchickCRT* (the Supplemental Material walks through the web-based app). This information is expected to be important for both patients and their physicians, as it provides approximately 62% of patients with reassurance that their 4-year survival should be quite good and identifies 16% of patients who are much more likely to merit evaluation for advanced heart failure therapies after CRT. In addition, as the pre-CRT parameters were also associated with long-term survival, these baseline features could also be used in the shared decision-making process for patients being evaluated for CRT.

Limitations

We did consider additional short-term response parameters such as a heart failure symptom score for inclusion in the response vector; however, having 3 components in the response vector worked well from a machine learning and statistical standpoint, and the symptom score was not as robust as the others. With respect to missing data, approximately 20% of patients could not exercise both before and after the study to facilitate determination of the change in peak VO_2. This was addressed in the analysis by using a matrix completion algorithm based on the expectation-maximization algorithm using a multivariate normal distribution. We also note that medical therapy for heart

failure has evolved in recent years, and more patients are being prescribed sodium-glucose cotransporter-2 inhibitors and angiotensin-neprilysin inhibitors. Longitudinal studies of outcomes for devices in heart failure with long-term follow-up by necessity to some extent lag behind recent trends in medical therapy, and certainly there will be interest in using these methods in future cohorts of patients to examine if the use of these heart failure therapies modifies the model. For example, angiotensin-neprilysin inhibitors have been shown to increase BNP expression.[29] As device-detected arrhythmias would occur mostly after the assessment of the 6-month response endpoint, which was used in the models along with baseline characteristics to predict subsequent clinical outcomes, incorporation of arrhythmia events was not possible with this study design. While not including arrhythmic death vs all-cause mortality as an outcome is a limitation, we do note a degree of uniformity with respect to protection from arrhythmic death, as nearly all patients had CRT defibrillators. Lastly, as this cohort with long-term follow-up had traditional CRT, we did not include patients with conduction system pacing in the present analysis; however, future analyses of a cohort of patients with conduction system pacing using these methods is planned.

Conclusion

Machine learning methods provide an effective way to understand the complex nature of CRT response, and the classifications generated by the methods promise to be useful for patients and providers with respect to clinical management strategies.

Funding Sources: This work was supported by grant R01 HL159945 from the NHLBI (PI: Bilchick).

Disclosures: Dr Bilchick has research grant support from Medtronic and Siemens Healthineers. Dr Malhotra has research grant support from Biosense Webster. Dr Darby has research grant support from Medtronic and Biosense Webster. Dr Mangrum has research grant support from Boston Scientific, CardioFocus, and St. Jude Medical. Dr Kramer and Dr Epstein have research grant support from Siemens Healthineers.

Authorship: All authors attest they meet the current ICMJE criteria for authorship.

Patient Consent: All patients provided written informed consent.

Ethics Statement: The research reported in this study was conducted according to the principles of the Declaration of Helsinki. The study was approved by the Institutional Review Board for Human Subjects Research at the University of Virginia.

Appendix

Supplementary data

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.hroo.2022.06.005.

References

1. Bristow MR, Saxon LA, Boehmer J, et al. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. N Engl J Med 2004;350:2140–2150.
2. Abraham WT, Fisher WG, Smith AL, et al. Cardiac resynchronization in chronic heart failure. N Engl J Med 2002;346:1845–1853.
3. Chung ES, Leon AR, Tavazzi L, et al. Results of the predictors of response to CRT (PROSPECT) trial. Circulation 2008;117:2608–2616.
4. Linde C, Abraham WT, Gold MR, Sutton MSJ, Ghio S, Daubert C. Randomized trial of cardiac resynchronization in mildly symptomatic heart failure patients and in asymptomatic patients with left ventricular dysfunction and previous heart failure symptoms. J Am Coll Cardiol 2008;52:1834–1843.
5. Arshad A, Moss AJ, Foster E, et al. Cardiac resynchronization therapy is more effective in women than in men: the MADIT-CRT (Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy) trial. J Am Coll Cardiol 2011;57:813–820.
6. Vinther M, Risum N, Svendsen JH, Mogelvang R, Philbert BT. A randomized trial of His pacing versus biventricular pacing in symptomatic HF patients with left bundle branch block (His-alternative). JACC Clin Electrophysiol 2021;7:1422–1432.
7. Vijayaraman P, Ponnusamy S, Cano O, et al. Left bundle branch area pacing for cardiac resynchronization therapy: results from the International LBBAP Collaborative Study Group. JACC Clin Electrophysiol 2021;7:135–147.
8. Ebong I, Mazimba S, Breathett K. Cardiac biomarkers in advanced heart failure: how can they impact our pre-transplant or pre-LVAD decision-making. Curr Heart Fail Rep 2019;16:274–284.
9. Bilchick KC, Kuruvilla S, Hamirani YS, et al. Impact of mechanical activation, scar, and electrical timing on cardiac resynchronization therapy response and clinical outcomes. J Am Coll Cardiol 2014;63:1657–1666.
10. Ramachandran R, Chen X, Kramer CM, Epstein FH, Bilchick KC. Singular value decomposition applied to cardiac strain from MR imaging for selection of optimal cardiac resynchronization therapy candidates. Radiology 2015;275:413–420.
11. Arora S, Aarones M, Aakhus S, et al. Peak oxygen uptake during cardiopulmonary exercise testing determines response to cardiac resynchronization therapy. J Cardiol 2012;60:228–235.
12. De Marco T, Wolfel E, Feldman AM, et al. Impact of cardiac resynchronization therapy on exercise performance, functional capacity, and quality of life in systolic heart failure with QRS prolongation: COMPANION trial sub-study. J Card Fail 2008;14:9–18.
13. Bilchick KC, Stafford P, Laja O, et al. Relationship of ejection fraction and natriuretic peptide trajectories in heart failure with baseline reduced and mid-range ejection fraction. Am Heart J 2022;243:1–10.
14. Gao X, Abdi M, Auger DA, et al. Cardiac magnetic resonance assessment of response to cardiac resynchronization therapy and programming strategies. JACC Cardiovasc Imaging 2021;14:2369–2383.
15. Bilchick KC, Auger DA, Abdishektaei M, et al. CMR DENSE and the Seattle heart failure model inform survival and arrhythmia risk after CRT. JACC Cardiovasc Imaging 2020;13:924–936.
16. 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;9:1737–1753.
17. Birnie D, Lemke B, Aonuma K, et al. Clinical outcomes with synchronized left ventricular pacing: analysis of the adaptive CRT trial. Heart Rhythm 2013;10:1368–1374.
18. Kim D, Gilson WD, Kramer CM, Epstein FH. Myocardial tissue tracking with two-dimensional cine displacement-encoded MR imaging: development and initial evaluation. Radiology 2004;230:862–871.
19. Zhong X, Spottiswoode BS, Meyer CH, Kramer CM, Epstein FH. Imaging three-dimensional myocardial mechanics using navigator-gated volumetric spiral cine DENSE MRI. Magn Reson Med 2010;64:1089–1097.
20. Spottiswoode BS, Zhong X, Hess A, et al. Tracking myocardial motion from cine DENSE images using spatiotemporal phase unwrapping and temporal fitting. IEEE Trans Med Imaging 2007;26:15–30.
21. Moon TK. The expectation-maximization algorithm. IEEE Signal Processing Magazine 1996;13:47–60.
22. Wiley CL, Switzer SP, Berg RL, Glurich I, Dart RA. Association of B-type natriuretic peptide levels with estimated glomerular filtration rate and congestive heart failure. Clin Med Res 2010;8:7–12.
23. Takase H, Dohi Y. Kidney function crucially affects B-type natriuretic peptide (BNP), N-terminal proBNP and their relationship. Eur J Clin Invest 2014;44:303–308.
24. Wallin H, Asp AM, Wallquist C, et al. Gradual reduction in exercise capacity in chronic kidney disease is associated with systemic oxygen delivery factors. PLoS One 2018;13:e0209325.
25. Dini FL, Demmer RT, Simioniuc A, et al. Right ventricular dysfunction is associated with chronic kidney disease and predicts survival in patients with chronic systolic heart failure. Eur J Heart Fail 2012;14:287–294.
26. Auger DA, Bilchick KC, Gonzalez JA, et al. Imaging left-ventricular mechanical activation in heart failure patients using cine DENSE MRI: validation and

implications for cardiac resynchronization therapy. J Magn Reson Imaging 2017; 46:887–896.
27. Howell SJ, Stivland T, Stein K, Ellenbogen KA, Tereshchenko LG. Using machine-learning for prediction of the response to cardiac resynchronization therapy: the SMART-AV study. JACC Clin Electrophysiol 2021; 7:1505–1515.
28. 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:e007316.
29. Myhre PL, Vaduganathan M, Claggett B, et al. B-type natriuretic peptide during treatment with sacubitril/valsartan: the PARADIGM-HF trial. J Am Coll Cardiol 2019;73:1264–1272.