10.1148/radiol.222239

Appendix S1

Machine Learning Algorithms

Ground Truth Production

Ground truth data were produced by i) delineation of the endocardial and epicardial contours; ii) definition of two remote reference regions (refROIs) in the normal myocardium (one in the lateral wall, one in the interventricular septum, thereby avoiding regions with scar and the right ventricular insertion points); and iii) elimination of ghost artifacts or wrap-around artifacts projecting onto normal myocardium by manually setting such hyperintense areas to normal (GTvolume, version 2.2.7, GyroTools, Zurich, Switzerland). Difficult cases (eg, with major artifacts) were analyzed by consensus (first and last author). These 4 contours yielded total LV mass, as well as dense scar (= $\text{denseScar}_{\text{hum}}$) defined as scar signal: > 5 standard deviations (SD) above the mean of refROIs, total scar (= $\text{totalScar}_{\text{hum}}$) defined as scar signal: > 2 SD above the mean of the refROIs, and nondense scar (= $\text{nondenseScar}_{\text{hum}}$) defined as scar signal: 2 SD ≤ scar signal ≤ 5 SD of the mean of the refROIs. The thresholds of > 5SD and > 2SD were chosen as they were already successfully applied by others (5,7,19).

Machine Learning Algorithm

In the fully convolutional neural network (20), ie, a Ternaus-Net (21,22) the down-sampling path is initialized with the weights from VGG16 (36), see also Figure 2. Such weight initialization has previously been demonstrated as effective in (medical) segmentation tasks (37). As VGG16 expects three-channel (ie, RGB) input images we prepend a single convolutional layer of three 1 × 1 kernels to the network, so that our single channel images can be given as input. Our network ends with a final sigmoid activation (Fig 2).

Data Preprocessing and Network Training

As part of preprocessing we normalized the intensity values of the LGE input images to the range [0,1] (by subtracting the minimum, dividing by the 99.5th percentile, and then clipping to [0,1]). All LGE images were also resampled to 1 mm × 1 mm resolutions and cropped to 128 × 128 pixels, centered around the LV. Only the short-axis LGE images were used as algorithm input.

During training, we used three channel binary images as GT target segmentation, where the first channel contains the segmentation mask of the myocardium and the second and third channels contain total scar and dense scar (= myocardium with signal > 2SD and > 5SD above the mean of the refROIs), respectively. During training we minimized the sum of the dice loss and binary cross entropy loss for each of the three channels. We trained the network using Adam (38) with a learning rate of 0.0001 and a batch-size of 32.

During training, the learning group was split into training (n = 413) and validation (n = 102) sets to allow for network design decisions and hyperparameter tuning, such as selecting the learning rate and deciding when to stop training (to avoid overfitting).

Test Time Augmentation

In cardiac segmentation in particular it is common to perform data augmentation, for example, by image rotations and/or reflections to increase the variability of images shown to the network during training (39). To improve the final predictions of the trained network, we made use of a test-time augmentation approach. During training we augmented our training images on the fly with random rotations, reflections, zooms and shears to improve the robustness of the final model. Specifically, when predicting the segmentation masks for an image x, we actually predicted the segmentation masks for all 8 views of x resulting from mirroring and/or rotations of multiples of 90 degrees. For each of these predictions the inverse mirroring/rotation was applied to the resulting mask, such that the prediction aligned with the original ×. This resulted in 8 separate predictions for x, which we averaged to produce a single final prediction yielding TTA (test-time augmentation) prediction.

Prediction Outputs

Four prediction outputs were generated: 1). the original prediction output, 2). the TTA prediction output, 3). the threshold-based (TB) output (from the network's predictions, refROIs were estimated and these were used to produce the two thresholds to define the total 2SD-scar and dense 5SD-scar, emulating the approach used in the human segmentation), and 4). a combination of the TTA and TB output (TTA+TB).

References

36. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://arxiv.org/abs/1409.1556. Posted 2014. Accessed February 11, 2019.

37. Shvets A, Rakhlin A, Kalinin A, Iglovikov V. Automatic instrument segmentation in robot-assisted surgery using deep learning. arXiv preprint arXiv:1803.01207. https://arxiv.org/abs/1803.01207. Posted 2018. Accessed February 11, 2019.

38. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. https://arxiv.org/abs/1412.6980. Posted 2014. Accessed February 11, 2019.

39. Krizhevsky A, Sutskever I, Hinton G. Image net classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012;25:1097–1105. https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

Appendix S2

Statistics

Calculation of Kaplan-Meier Curves

Considering time from study entry to major adverse cardiac events (MACE), end of follow-up or censoring event-free survival was estimated using the Kaplan–Meier method and survival curves were compared by means of the Log-Rank test.

Univariable and Multivariable Cox Proportional Hazard Models

Univariable and multivariable Cox proportional hazard models were used to assess the association of demographic and imaging data with the risk of MACE. As current guidelines assign a class 1 indication for ICD implantation in CAD patients with LVEF ≤ 35% AND dyspnea NYHA II or III (11–13), this combination called guidelines-criterion was also tested. In the testing group (n = 246), the subgroup of 3D breath-hold LGE images (n = 30) was of limited quality (good or acceptable quality in 61% versus 80% for 2D acquisitions, P < .001). Therefore, the univariable and multivariable Cox proportional hazard models were performed on the 2D data sets of the final testing group (n = 216, Fig 1).

Cox proportional hazard modeling was performed by including candidate predictors-identified as those with P < .10 at univariable analysis–and by excluding predictors causing collinearity on the basis of the variance inflation factor (VIF). Missing data for covariates were handled with the use of multiple imputation. Different multivariable models were adapted to evaluate whether human-based or machine-based scar measures improve the currently established guidelines-criterion (= model 0), ie, we included to this model 0 the $\text{denseScar}_{\text{hum}}$ (yielding model 1), the $\text{denseScar}_{\text{mach}}$ (yielding model 2), and both, the $\text{denseScar}_{\text{hum}}$ and $\text{denseScar}_{\text{mach}}$ (yielding model 3). Thus, the multivariable models 1–3 are adjusted for the guidelines-criterion. The same procedure (ie, adjusting for guidelines-criterion) was applied by including $\text{totalScar}_{\text{hum}}$, $\text{totalScar}_{\text{mach}}$, or both yielding models 1a, 2a, and 3a, respectively.

To assess the performance of the multivariable model in terms of discriminatory ability, estimated areas under the curve (AUCs) for the time-dependent receiver-operator-characteristics (ROC) curves were derived.

To account for noncardiovascular death as competing event, the analyses were also repeated considering Fine and Gray competing risk regression. To assess the performance of the competing risk multivariable models estimated areas under the curve (AUCs) for the time-dependent receiver-operator-characteristics (ROC) curves were derived. This competing risk modeling was performed as a sensitivity analysis.

In the multivariable analyses scar mass was considered as a dichotomous variable using as thresholds the values for totalScar and denseScar that best predicted MACE in the training data set (n = 515) as evaluated by time-dependent ROC analyses considering with 10,000 bootstrap replicates. The areas under the time-dependent ROC curves in the testing group ranged from 0.52 ($\text{TotalScar}_{\text{hum}}$) to 0.56 ($\text{DenseScar}_{\text{mach}}$) and were statistically not significantly different from those of the learning group.

Proportional hazards assumption was evaluated using visual inspection of the log-log survival curves and the Schoenfeld residuals test.

Table S1

Inclusion and Exclusion Criteria

Inclusion criteria
• HF (according to the ACC/AHA classification, C.W. Yancy et al, J. Am. Coll. Cardiol, 62,2013) secondary to ICM in adult patients (age ≥ 18 years)
• Reduced LVEF of < 50% (as the ACC/AHA definition of HF with preserved ejection fraction (HFpEF) establishes reference values of EF ≥ 50%, S.A. Hunt et al, J. Am. Coll. Cardiol, 53,2009)
• Stage B and C (according to the ACCF/AHA classification)
• The following variables available at baseline: clinical variables standard TTE and CMR acquisition
• Ischemic origin of LV dysfunction was defined as the presence of at least one of the following criteria:
-i) history of previous percutaneous coronary intervention or coronary artery bypass grafting
-ii) angiographic evidence of coronary artery disease with ≥ 70% stenosis in ≥ 1 epicardial vessel or ≥ 50% stenosis of the left main coronary artery
-iii) evidence of ischemic scar at LGE imaging in a specific coronary perfusion territory involving at least 2 contiguous segments according to the AHA LV segmentation
Exclusion criteria
• Pregnancy
• Current alcohol or drug abuse
• Unstable angina
• Decompensated HF (NYHA class IV) in the previous 3 months
• Acute myocarditis in the previous 3 months
• Recent MI (< 40 days)
• Severe valvular disease
• Cardiac amyloidosis
• Hypertrophic cardiomyopathy
• Arrhythmogenic right ventricular cardiomyopathy
• Takotsubo cardiomyopathy
• Congenital heart disease
• Constrictive pericarditis
• Non CMR compatible device
• Estimated GFR ≤ 30 mL/min/1.73 m2 (KDOQI ≥ 4)
• Other contraindication to gadolinium contrast agent
• Severe claustrophobia
• Cancer
• Cardiotoxicity of cancer therapy
• Postradiation
• Participating in other trials with an active treatment arm (not to exclude patients who are in trials of diagnostic techniques or approved therapies)
• Unwilling or unable to provide informed consent

HF = heart failure, ACC/AHA = American College of Cardiology/American Heart Association, CAD = coronary artery disease, ICM = ischemic cardiomyopathy, LGE = late gadolinium enhancement, LVEF = left ventricle ejection fraction, BNP = brain natriuretic peptide, NYHA = New York Heart Association, TTE = transthoracic echocardiography, CMR = cardiac magnetic resonance, MI = myocardial infarction, GFR = glomerular filtration rate, KDOQI = Kidney Disease Outcomes Quality Initiative.

Table S2

Cardiac MRI Parameters

	All Patients	Learn Group	Test Group	*P* Value = Learn versus Test
2D/3D (n)	676/85	460/55	216/30	
Breath-hold	all	all	all	—
Slice thickness				
−2D (mm)	8.2 ± 0.8 (7–12)	8.3 ± 0.8 (7–12)	8.2 ± 0.7 (7–12)	0.13
−3D (mm)	5.0 ± 0.5 (4.5–8.0)	5.1 ± 0.6 (4.5–8.0)	5.0 ± 0.2 (4.5–6.0)	0.50
Gap between slices				
−2D (mm)	0.5 ± 1.0 (0–4.5)	0.5 ± 1.0 (0–4.5)	0.5 ± 0.9 (0–2.6)	0.64
−3D (mm)	—	—	—	—
Resolution in-plane	1 × 1 mm^2	1 × 1 mm^2	1 × 1 mm^2	—
Type of read-out				
-FGRE (n)	706	478	229	0.90
-SSFP (n)	54	37	17	0.90
Output image				
-Mag (n)	602	408	195	0.99
PSIR (n)	158	107	51	0.99
CM dose (mMol/kg)	0.158 ± 0.032	0.158 ± 0.030	0.156 ± 0.031	0.60
range	(0.10–0.20)	(0.10–0.20)	(0.10–0.20)	—
Gadobenate dim. (%)	331 (43.5)	221 (42.9)	110 (44.7)	0.64
Gadoterate meg. (%)	192 (25.2)	128 (24.9)	64 (26.0)	0.73
Gadobutrol (%)	172 (22.6)	122 (23.7)	50 (20.3)	0.30
Gadoteridol (%)	42 (5.5)	28 (5.4)	14 (5.7)	0.89
Gadodiamide (%)	24 (3.2)	16 (3.1)	8 (3.3)	0.92

Note.—In-plane resolution: all late gadolinium enhanced (LGE) input images were preprocessed to yield 1 × 1 mm^2 spatial resolution for all cases. To this end, input images were spline-interpolated to 1 × 1 mm^2 resolution using the zoom function of the scipy.ndimage package, centered at the left ventricle, and subsequently cropped to a 128 × 128 matrix as required for the Ternaus-Net. The process was automatized and required a few seconds of compute time per patient. Further preparatory analysis and transformation were not needed. Note that test-time augmentation was used ie, predictions on a number of variants of the input images were made, and then combined to yield a single refined prediction. Per study protocol, after acquisitions of localizers and cine steady-state free precession images, a conventional gadolinium-chelate contrast medium (CM) was administered intravenously as a bolus of 0.1–0–2 mmol/kg according to each center's customary protocol for LGE imaging. According the CM dose used, 10–15 min after CM injection, breath-hold inversion recovery gradient-echo images were acquired with the same image orientations as used for cine acquisitions to detect LGE. CM = contrast medium, FGRE = fast gradient echo, Mag = magnitude image, PSIR = phase-sensitive inversion recovery image, SSFP = steady-state free-precession, Gadobenate dim = Gadobenate dimeglumine, Gadoterate meg. = Gadoterate meglumine.

Table S3

Criteria for LGE Image Quality Per Patient

Quality Grade	Definition: % of Slices with Artifacts
Good	no slice with artifacts
Acceptable	<30%
Sufficient	30% to < 40%
Borderline	40% to < 50% or one slice not analyzable
Poor	≥50% or 2 slices not analyzable
Patient excluded	≥ half of slices not analyzable

Note.—Image quality is graded according the following criteria: 1. A minor degree of blurring of apical slices is to be expected (due to partial volume artifacts) and is therefore accepted, unless the apical slices are crucial in decision making (eg, when apical scars are present). 2. Artifact: if the artifact interferes with scar assessment (eg, when a ghost or a wrap-around artifact crosses the scar region). Artifacts not interfering with scar assessment are note noted as artifacts. 3. Excluding criteria per slice (ie, not analyzable): more than 50% of myocardium not analyzable due to artifact. For examples of quality gradings, see Figure E1.

Table S4

Late Gadolinium Enhanced Image Quality Scoring by Two Observers

	Observer 1	Observer 2	*P* Value	Kendall τ
Mean ± SD	4.1 ± 1.1	4.2 ± 1.1	0.73*	
Median (IQR)	4 (1)	4 (1)		
Quality Scores n (%)			<0.001†	0.497
-Good (score 5)	18 (45)	19 (47.5)		
-Acceptable (score 4)	14 (35)	15 (37.5)		
-Sufficient (score 3)	3 (7.5)	2 (5)		
-Borderline (score 2)	3 (7.5)	2 (5)		
-Poor (score 1)	2 (5)	2 (5)		

Note.—IQR = interquartile range.

* Mann Whitney test.

† Kendall rank correlation.

Table S5

Scar Quantification Results by Human and the Machine Algorithm in the Testing Group (*n* = 216)

CMR Scar Machine (TTA)	Human	Machine	*P* Value
LV mass (g)†	108.9 ± 33.2	116.9 ± 30.1	<0.001
total Scar% (%LV)†	39.3 ± 14.2	36.5 ± 12.5	<0.001
dense Scar% (%LV)†	21.3 ± 13.0	18.6 ± 11.6	<0.001
nondense Scar% (%LV)	18.0 ± 5.8	17.9 ± 6.1	0.75
total Scar (g)	43.4 ± 21.6	42.6 ± 17.5	0.30
dense Scar (g)†	23.6 ± 16.6	21.7 ± 14.2	0.002
nondense Scar (g) †	19.8 ± 9.4	20.8 ± 8.8	0.02

Note.—LV = left ventricular, TTA = test time augmentation.

† Denotes a significant *P* value < 0.05.

Table S6

Univariable Cox Model with Competitive Risk Assessment for MACE (Testing Group, 2D LGE Images, *n* = 216)

Univariable Cox Model with Competitive Risk Assessment			
	SHR	*P* Value	(95%-CI)
Demographics			
Age (y)	1.01	0.39	(0.99–1.04)
Gender	1.56	0.25	(0.73–3.33)
BMI (kg/m^2)	1.02	0.80	(0.89–1.16)
BSA (m^2)	1.06	0.96	(0.22–5.08)

Family History	1.39	0.21	(0.83–2.34)
Smoking	0.69	0.19	(0.40–1.20)
Hyperlipidemia†	0.61	0.049	(0.37–0.99)
Diabetes	1.26	0.39	(0.74–2.15)
Hypertension	0.80	0.40	(0.49–1.33)
NYHA Class†	1.35	0.02	(1.05–1.74)
Left bundle branch block	0.87	0.62	(0.50–1.52)
Echocardiography			
LVEDVI (mL/m^2)†	1.01	0.02	(1.00–1.01)
LVESVI (mL/m^2)†	1.01	0.001	(1.06–1.02)
LVEF echo (%)†	0.94	<0.001	(0.92–0.97)
LVEF echo (threshold ≤ 35%)†	4.07	<0.001	(2.03–8.16)
LVEF echo (≤ 35%) and NYHA II/III*†	3.57	<0.001	(2.13–5.97)
CMR Function/volumes			
LVEDVI (mL/m^2)†	1.01	<0.001	(1.00–1.01)
LVESVI (mL/m^2)†	1.01	<0.001	(1.00–1.01)
LVSVI (mL/m^2)	0.99	0.09	(0.98–1.00)
LVEF (%)†	0.95	<0.001	(0.92–0.97)
RVEDVI (mL/m^2)	1.01	0.14	(1.00–1.02)
RVESVI (mL/m^2)†	1.01	0.045	(1.00–1.02)
RVSVI (mL/m^2)	0.99	0.30	(0.98–1.01)
RVEF (%)†	0.98	0.03	(0.97–1.00)
CMR LGE-Quality			
-acceptable versus good	0.59	0.11	(0.31–1.13)
-sufficient versus good	0.54	0.22	(0.20–1.46)
-borderline versus good	0.54	0.44	(0.11–2.59)
-poor versus good	0.58	0.42	(0.15–2.19)
CMR scar human (GT)			
$LVmass_{hum}$ (g)	1.01	0.06	(1.00–1.01)
$totalScar\%_{hum}$ (%LV)	5.07	0.09	(0.77–33.53)
$denseScar\%_{hum}$ (%LV)†	7.27	0.03	(1.17–45.23)
$nondenseScar\%_{hum}$ (%LV)	0.20	0.41	(0.00–9.48)
$totalScar_{hum}$ (g)†	1.01	0.02	(1.00–1.02)
$denseScar_{hum}$ (g)†	1.02	0.009	(1.00–1.03)
$nondenseScar_{hum}$ (g)	1.01	0.45	(0.99–1.03)
Threshold			
-ROC: $totalScar_{hum}$ (> 30.7 g)†	1.01	0.02	(1.00–1.02)
-ROC: $denseScar_{hum}$ (> 20.3 g)	1.66	0.06	(0.98–2.80)
CMR scar machine (TTA)			
$LVmass_{mach}$ (g)†	1.01	0.02	(1.00–1.02)
$totalScar\%_{mach}$ (%LV)†	7.95	0.02	(1.37–46.25)
$denseScar\%_{mach}$ (%LV)†	16.98	0.003	(2.60–111.08)
$nondenseScar\%_{mach}$ (%LV)	0.37	0.65	(0.01–27.48)
$totalScar_{mach}$ (g)†	1.02	0.001	(1.01–1.03)
$denseScar_{mach}$ (g)†	1.03	<0.001	(1.01–1.04)
$nondenseScar_{mach}$ (g)	1.01	0.45	(0.99–1.03)
Threshold			
-ROC: $totalScar_{mach}$ (> 30.7 g)†	1.02	0.001	(1.01–1.03)
-ROC: $denseScar_{mach}$ (> 20.3 g)†	2.31	0.002	(1.35–3.95)

Note.—BMI = body mass index, BSA = body surface area, 95%-CI = 95% confidence interval, LVEDVI = left ventricular end-diastolic volume index, LVESVI = left ventricular end-systolic volume index, LVSVI = left

ventricular stroke volume index, LVEF = left ventricular ejection fraction, ROC = receiver-operator-characteristics curve, RVEDVI = right ventricular end-diastolic volume index, RVESVI = right ventricular end-systolic volume index, RVSVI = right ventricular stroke volume index, RVEF = right ventricular ejection fraction, SHR = subhazard ratio, TTA = test time augmentation.

† Denotes a significant *P* value < 0.05.

* Criterion for ICD implantation according international Guidelines (11–13).

Table S7

Univariable Cox Model for All-cause Mortality (Testing Group, 2D LGE Images, *n* = 216)

Univariable Cox Model for All-cause Mortality	HR	*P* Value	(95%-CI)
Demographics			
Age (y)†	1.03	0.02	(1.01–1.01)
Gender	2.64	0.06	(0.99–7.17)
BMI (kg/m^2)	0.98	0.79	(0.87–1.11)
BSA (m^2)	0.26	0.29	(0.03–2.93)
Family History	1.77	0.18	(0.77–4.05)
Smoking	0.48	0.15	(0.18–1.30)
Hyperlipidemia	0.75	0.49	(0.33–1.71)
Diabetes	1.94	0.13	(0.83–4.54)
Hypertension	1.24	0.62	(0.52–2.96)
NYHA Class	1.48	0.08	(0.96–2.29)
Left bundle branch block	0.78	0.65	(0.26–2.30)
Echocardiography			
LVEDV (mL/m^2)	1.01	0.09	(1.00–1.02)
LVESV (mL/m^2)	1.01	0.07	(1.00–1.03)
LVEF echo (%)†	0.94	0.01	(0.90–0.99)
LVEF echo (threshold ≤ 35%)	2.34	0.08	(0.91–5.97)
LVEF echo (≤ 35%) and NYHA II/III*	1.87	0.16	(0.82–4.26)
CMR Function/volumes			
LVEDVI (mL/m^2)	1.00	0.36	(1.00–1.01)
LVESVI (mL/m^2)	1.00	0.21	(1.00–1.02)
LVSV (mL/m^2)	1.00	0.38	(1.00–1.01)
LVEF (%)\dag\	0.95	0.02	(0.90–1.00)
RVEDVI (mL/m^2)	1.01	0.52	(1.00–1.02)
RVESVI (mL/m^2)	1.01	0.08	(1.00–1.03)
RVSV (mL/m^2)	0.99	0.35	(0.97–1.01)
RVEF (%)†	0.96	0.003	(0.93–0.99)
CMR scar human (GT)			
LVmass$_{hum}$ (g)	1.01	0.11	(1.00–1.02)
totalScar%$_{hum}$ (%LV)	4.63	0.36	(0.21–104.22)
denseScar%$_{hum}$ (%LV)	1.40	0.84	(0.06–34.75)
nondenseScar%$_{hum}$ (%LV)	1.04	0.99	(0.01–109.72)
totalScar$_{hum}$ (g)	1.02	0.08	(1.00–1.04)
denseScar$_{hum}$ (g)	1.01	0.36	(1.00–1.04)
nondenseScar$_{hum}$ (g)	1.02	0.21	(0.99–1.05)
Threshold			
-ROC: totalScar$_{hum}$ (> 30.7 g)†	8.12	0.04	(1.09–60.39)
-ROC: denseScar$_{hum}$ (> 20.3 g)	1.21	0.66	(0.52–2.79)
CMR scar machine (TTA)			

$LVmass_{mach}$ (g)	1.01	0.09	(1.00–1.03)
$totalScar\%_{mach}$ (%LV)	4.73	0.35	(0.19–118.95)
$denseScar\%_{mach}$ (%LV)	1.07	0.97	(0.03–35.16)
$nondenseScar\%_{mach}$ (%LV)	0.69	0.88	(0.00–68.18)
$totalScar_{mach}$ (g)	1.02	0.09	(1.00–1.04)
$denseScar_{mach}$ (g)	1.01	0.60	(0.98–1.03)
$nondenseScar_{mach}$ (g)	1.02	0.13	(0.99–1.06)
Threshold			
-ROC: $totalScar_{mach}$ (> 30.7 g)†	4.57	0.04	(1.07–19.53)
-ROC: $denseScar_{mach}$ (> 20.3 g)	0.97	0.95	(0.43–2.21)

Note.—BMI = body mass index, BSA = body surface area, 95%-CI = 95% confidence interval, GT = ground truth, LVEDV = left ventricular end-diastolic volume, LVEDVI = left ventricular end-diastolic volume index, LVEF = left ventricular ejection fraction, LVESV = left ventricular end-systolic volume, LVESVI = left ventricular end-systolic volume index, ROC = receiver-operator-characteristics curve, RVEDVI = right ventricular end-diastolic volume index, RVEF = right ventricular ejection fraction, RVESVI = right ventricular end-systolic volume index, RVSV = right ventricular stroke volume, SHR = subhazard ratio, TTA = test time augmentation.

† Denotes a significant *P* value < 0.05.

* Criterion for implantable cardioverter defibrillator (ICD) implantation according international Guidelines (11–13).

Table S8

Multivariable Competing Risk Models for MACE (Testing Group, 2D LGE Images, *n* = 216)

Dense Scar	SHR	*P* Value	(95%-CI)	$AUC_{Time\text{-}dependent\ ROC}$
Model 0				
LVEF echo (≤ 35%) and NYHA II/III*	3.57	<0.001	(2.13–5.97)	0.61 (0.49–0.72)
Model M1-Dense Scar				
LVEF echo (≤ 35%) and NYHA II/III*	3.39	<0.001	(2.00–5.76)	0.63 (0.51–0.77)
threshold-$denseScar_{hum}$ (> 20.3 g)	1.28	0.36	(0.76–2.15)	*P* = .18 versus M0
Model M2-Dense Scar				
LVEF echo (≤ 35%) and NYHA II/III*	3.16	<0.001	(1.84–5.43)	0.64 (0.53–0.79)
threshold-$denseScar_{mach}$ (> 20.3 g)	1.79	0.050	(1.00–3.19)	*P* = .01 versus M0
				P = .20 versus M1
Model M3-Dense Scar				
LVEF echo (≤ 35%) and NYHA II/III*	3.25	<0.001	(1.90–5.55)	0.68 (0.57–0.79)
threshold-$denseScar_{hum}$ (> 20.3 g)	0.61	0.22	(0.28–1.35)	*P* = .01 versus M0
threshold-$denseScar_{mach}$ (> 20.3 g)	2.62	0.02	(1.13–6.05)	*P* = .27 versus M1
Total Scar	SHR	*P* value	(95%-CI)	$AUC_{Time\text{-}dependent\ ROC}$
Model 0				
LVEF echo (≤ 35%) and NYHA II/III*	3.57	<0.001	(2.13–5.97)	0.61 (0.49–0.72)
Model M1b-Total Scar				
LVEF echo (≤ 35%) and NYHA II/III*	3.43	<0.001	(2.03–5.79)	0.67 (0.56–0.77)
threshold-$totalScar_{hum}$ (> 30.7 g)	1.29	0.43	(0.69–2.42)	*P* = .001 versus M0
Model M2b-Total Scar				
LVEF echo (≤ 35%) and NYHA II/III*	3.03	<0.001	(1.74–5.28)	0.66 (0.56–0.77)
threshold-$totalScar_{mach}$ (> 30.7 g)	1.90	0.07	(0.94–3.83)	*P* = .005 versus M0
				P = .55 versus M1b
Model M3b-Total Scar				
LVEF echo (≤ 35%) and NYHA II/III*	3.02	<0.001	(1.73–5.26)	0.66 (0.54–0.75)

threshold-totalScar$_{hum}$ (> 30.7 g)	0.84	0.62	(0.42–1.68)	*P* = .045 versus M0
threshold-totalScar$_{mach}$ (> 30.7 g)	2.11	0.06	(0.97–4.59)	*P* = .09 versus M1b

Note.—AUC = area under the receiver-operator-characteristics curve, 95%-CI = 95% confidence interval, LVEF = left ventricular ejection fraction, ROC = receiver-operator-characteristics curve, SHR = subhazard ratio.

* Criterion for implantable cardioverter defibrillator (ICD) implantation according international Guidelines (11–13).