JACC: CARDIOVASCULAR IMAGING
© 2020 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION
PUBLISHED BY ELSEVIER

VOL. 13, NO. 5, 2020

ORIGINAL RESEARCH

Interpatient Similarities in Cardiac Function

A Platform for Personalized Cardiovascular Medicine

Márton Tokodi, MD,[a,b] Sirish Shrestha, MSc,[a] Christopher Bianco, MD,[a] Nobuyuki Kagiyama, MD, PhD,[a] Grace Casaclang-Verzosa, MD,[a] Jagat Narula, MD, PhD,[c] Partho P. Sengupta, MD, DM[a]

ABSTRACT

OBJECTIVES The authors applied unsupervised machine-learning techniques for integrating echocardiographic features of left ventricular (LV) structure and function into a patient similarity network that predicted major adverse cardiac event(s) (MACE) in an individual patient.

BACKGROUND Patient similarity analysis is an evolving paradigm for precision medicine in which patients are clustered or classified based on their similarities in several clinical features.

METHODS A retrospective cohort of 866 patients was used to develop a network architecture using 9 echocardiographic features of LV structure and function. The data for 468 patients from 2 prospective cohort registries were then added to test the model's generalizability.

RESULTS The map of cross-sectional data in the retrospective cohort resulted in a looped patient network that persisted even after the addition of data from the prospective cohort registries. After subdividing the loop into 4 regions, patients in each region showed unique differences in LV function, with Kaplan-Meier curves demonstrating significant differences in MACE-related rehospitalization and death (both $p < 0.001$). Addition of network information to clinical risk predictors resulted in significant improvements in net reclassification, integrated discrimination, and median risk scores for predicting MACE ($p < 0.05$ for all). Furthermore, the network predicted the cardiac disease cycle in each of the 96 patients who had second echocardiographic evaluations. An improvement or remaining in low-risk regions was associated with lower MACE-related rehospitalization rates than worsening or remaining in high-risk regions (3% vs. 37%; $p < 0.001$).

CONCLUSIONS Patient similarity analysis integrates multiple features of cardiac function to develop a phenotypic network in which patients can be mapped to specific locations associated with specific disease stage and clinical outcomes. The use of patient similarity analysis may have relevance for automated staging of cardiac disease severity, personalized prediction of prognosis, and monitoring progression or response to therapies.
(J Am Coll Cardiol Img 2020;13:1119–32) © 2020 by the American College of Cardiology Foundation.

Cardiovascular disease continues to be a leading cause of death worldwide (1). One of the major research priorities is to prevent adverse clinical events and hospitalization by risk factor management and by earlier detection of subclinical cardiac dysfunction. This research has led to a plethora of noninvasive approaches, with diverse technical underpinnings, to assess various

From the [a]Division of Cardiology, West Virginia University Heart & Vascular Institute, Morgantown, West Virginia; [b]Heart and Vascular Center, Semmelweis University, Budapest, Hungary; and the [c]Division of Cardiology, Icahn School of Medicine at Mount Sinai, New York, New York. Dr. Sengupta has been a consultant for Heartsciences, Ultromics Ltd., and Kencor Health. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the *JACC: Cardiovascular Imaging* author instructions page.

Manuscript received August 2, 2019; revised manuscript received October 31, 2019, accepted December 19, 2019.

ISSN 1936-878X/$36.00
https://doi.org/10.1016/j.jcmg.2019.12.018

ABBREVIATIONS AND ACRONYMS

A = late diastolic transmitral flow velocity

ACC/AHA = American College of Cardiology/American Heart Association

CI = confidence interval

E = early diastolic transmitral flow velocity

EF = ejection fraction

HR = hazard ratio

LV = left ventricular

MACE = major adverse cardiac event(s)

MAGGIC = Meta-Analysis Global Group in Chronic Heart Failure

NYHA = New York Heart Association

TDA = topological data analysis

and often overlapping aspects of cardiac function. For example, assessment of left ventricular (LV) systolic and diastolic function is an integral part of evaluating patients with subclinical or overt cardiac disease (2). However, a diagnostic imaging protocol can produce numerous parameters, each with its strengths and limitations (3). Several of these parameters are routinely used in clinical cardiology in conjunction with the 4 functional class stages of the New York Heart Association (NYHA) and the American College of Cardiology/American Heart Association (ACC/AHA) classification of heart failure. However, due to the lack of unanimity on the combination and use of these parameters to depict a single patient or a group of patients with similar characteristics, there is a paramount need to develop a staging method that can integrate multiple tests and diagnostic variables at the point of care.

In the present study, we explore topological data analysis (TDA) (4), a state-of-the-art data analytical approach that provides well-founded mathematical, statistical, and algorithmic methods to present the underlying geometric structures in data. It can be used in an unsupervised machine learning pipeline to compare multiple variables and clusters similar patients into nodes. A node connects with other nodes with edges if the same patients are clustered into >1 nodes. This allows us to summarize the complex data to a simple connected patient similarity network that can be visualized to attain novel insights into disease mechanisms (5,6). TDA has been successfully used to derive mechanistic insights from biological datasets in disciplines such as oncology, genomics, immunology, diabetes, and pre-clinical spinal cord injury (7-13). We applied the TDA-based network analysis to detect patient similarity patterns using a cross-sectional multiparametric echocardiographic dataset. Accordingly, we pooled echocardiographic data from both ambulatory and hospitalized settings to develop a broad cross-sectional representation of patients across different stages of cardiac disease that can be identified readily on a patient–patient similarity network. We subsequently investigated the prognostic value of the topological network and explored whether the longitudinal course of disease observed could be tracked along the topological map to assess the risks of cardiac events in an index patient.

METHODS

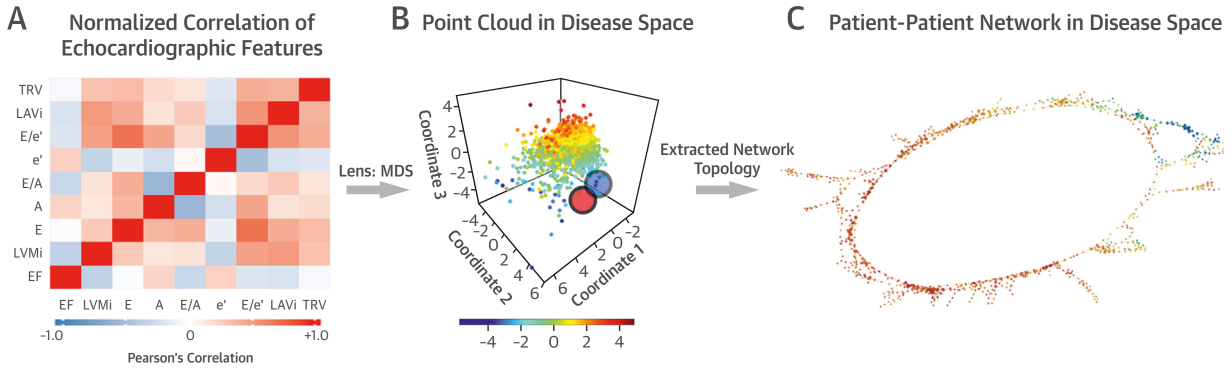
DATA. There were 2 parts to our study. First, we developed a broad cross-sectional representation of patients across different stages of cardiac disease as a primary cohort by pooling patients from a retrospective study and 2 prospective registries. The network topology was first developed from the retrospective study. The addition of prospective data was used as a step for validating the persistence and stability of network topology.

Second, after the stability of the network shape was confirmed, we tested the personalized prediction for a new patient to be represented on this network structure by including a second cohort of patients who had 2 echocardiographic evaluations (secondary cohort) for clinically indicated reasons. As patients underwent 2 echocardiographic examinations, the change in the location of these patients over the network was monitored to understand whether the network also represented a change in cardiac disease staging.

STUDY POPULATION. A flowchart of the study population, inclusion and exclusion criteria are provided in Supplemental Figure 1. The retrospective group included a convenience sample of 866 outpatients (age 65 ± 17 years; 387 men and 479 women) in sinus rhythm who were referred for echocardiographic assessment of cardiac function between March 2013 and December 2015 at the Icahn School of Medicine at Mount Sinai (New York, New York). The prospective groups included 468 patients (age 55 ± 15 years; 195 men and 273 women) enrolled between July 2017 to February 2018 in 2 ongoing patient registries at West Virginia University (Morgantown, West Virginia) that followed 2 prospective trials (Analysis of Surface EKG Signals to Identify Myocardial Dysfunction in Patients at Risk for Coronary Artery Disease, NCT02560168; and Evaluation of Cardiopulmonary Diseases by Ultrasound, NCT02248831). The pooled patients from the retrospective study and the 2 prospective registries formed the primary cohort of patients used for developing the patient–patient similarity network.

For personalized patient predictions, we further tested the model in a secondary cohort of 96 patients (age 58 ± 15 years; 49 men and 47 women) from the prospective registries who had 2 consecutive echocardiographic examinations. Follow-up data for this cohort were collected after the second echocardiographic assessment. The additional details of comprehensive echocardiographic evaluations performed according to published guidelines (14) is

FIGURE 1 **Steps of TDA**



Topological data analysis (TDA) pipeline. **(A)** Dataset containing echocardiographic features was analyzed using TDA from a bivariate correlation matrix. **(B)** Data were processed with 2 multidimensional scaling (MDS) lenses to generate the disease space. TDA resamples the disease space multiple times to identify similar patients and links them to nodes (**red and blue circles**). **(C)** The patient–patient network was created to provide a simple visual representation of the data. A = late diastolic transmitral flow velocity; é = early diastolic relaxation velocity at septal mitral annular position; E = early diastolic transmitral flow velocity; E/A = early to late diastolic transmitral flow velocity ratio; E/é = early diastolic transmitral flow to annular velocity ratio; EF = ejection fraction; LAVi = left atrial volume index; LVMi = left ventricular mass index; TRV = tricuspid regurgitation peak volume.

listed in the Supplemental Appendix. The institutional review board approved the study protocol, and all study participants in the prospective studies provided written informed consent.

TDA. We applied TDA to detect patterns in multidimensional echocardiographic parameters by studying the geometrical structure obtained from the network that signified a compressed representation of high-dimensional data (5) for patient similarity analysis. The notion of expressing the shape of data using TDA was extensively validated and successfully applied to different areas of health sciences (7,10-12,15). In the created topological network, nodes represented a cluster of patients, whereas edges connected nodes that contained patients who existed in both nodes. Nodes were color-coded based on the average value of the parameter of interest (e.g., ejection fraction [EF] or LV mass) of the clustered patients in the node. The nodes were colored red for the most extreme abnormal values, whereas they were colored blue for the maximum normal value. There were a gamut of colors for the average values in between.

TDA was performed using the cloud based Ayasdi Workbench v7.4 (Ayasdi Inc., Menlo Park, California). Nine echocardiographic parameters were used to create the topological network, namely, EF, LV mass index, early diastolic transmitral flow velocity (E), late diastolic transmitral flow velocity (A), E/A ratio, early diastolic relaxation velocity (é), E/é ratio, left atrial volume index, and tricuspid regurgitation peak velocity. The steps in generating the TDA network is provided in **Figure 1**. After the network was created in the primary cohort (containing retrospective and prospective cohorts), we trained a random forest-based classifier (16) using the echocardiographic data of the primary cohort to predict the region each patient from the secondary cohort might belong to. This allowed us to predict the characteristics of the patients and the outcomes the patient might have experienced. Additional details for the TDA can be found in the Supplemental Appendix.

CLINICAL OUTCOMES, ENDPOINTS, AND STAGING. Patient electronic medical records were reviewed for post-echocardiographic follow-up. Hospitalizations were classified based on the International Classification of Diseases-10th Revision coding. Endpoints were defined as death from a major adverse cardiac event (MACE) (defined as myocardial infarction, acute coronary syndrome, acute decompensated heart failure, cardiac arrest, or arrhythmia) and first MACE-related rehospitalization. The time to each endpoint was measured from the date of the echocardiographic examination used in the study. Clinical cardiac disease staging was performed using NYHA functional class assessment, ACC/AHA heart failure staging, and assessment of the MAGGIC (Meta-Analysis Global Group in Chronic Heart Failure) risk score. The ability of the MAGGIC score to predict death and cardiovascular hospitalization events related to MACE was well validated (17,18).

TABLE 1 Clinical and Echocardiographic Characteristics of the Retrospective Cohort

	Retrospective Cohort (n = 866)	Region I (n = 177)	Region II (n = 138)	Region III (n = 286)	Region IV (n = 212)	Overall p Value
Demographics						
Men	387 (45)	73 (41)	39 (28)*	102 (36)*	142 (67)*	<0.001
Age, yrs	66 (54-79)	49 (37-59)*	64 (58-72)*	78 (67-86)*	66 (55-80)	<0.001
Body mass index, kg/m^2	26.8 (23.6–30.4)	25.9 (23.0–29.8)	25.7 (22.5–29.4)	27.0 (23.8–30.3)	27.5 (24.3–31.4)	0.013
Hypertension	441 (51)	48 (27)*	54 (39)†	189 (66)*	126 (59)†	<0.001
Hyperlipidemia	358 (41)	37 (21)*	45 (33)*	153 (53)*	99 (47)	<0.001
Diabetes mellitus	178 (21)	16 (9)*	14 (10)†	79 (28)*	60 (28)†	<0.001
COPD	61 (7)	4 (2)†	14 (10)	25 (9)	15 (7)	0.012
Tobacco abuse	377 (44)	76 (43)	50 (36)	122 (43)	105 (50)	0.282
History of CKD	214 (25)	18 (10)*	18 (13)*	96 (34)*	62 (29)‡	<0.001
Clinical outcomes						
MACE rehospitalization	147 (17)	8 (5)*	10 (7)*	77 (27)*	45 (21)	<0.001
MACE death	10 (1)	0 (0)	0 (0)	3 (1)	5 (2)‡	0.083
Echocardiography						
LVEF, %	62 (57–67)	63 (59–66)†	65 (61–68)*	64 (60–68)*	56 (45–74)*	<0.001
LV mass index, g/m^2	85 (67–106)	66 (58–77)*	64 (58–76)*	93 (78–115)*	105 (87–129)*	<0.001
E, m/s	0.79 (0.60–0.90)	0.73 (0.63–0.88)†	0.70 (0.60–0.80)*	0.80 (0.70–1.00)*	0.80 (0.60–1.00)†	<0.001
A, m/s	0.76 (0.60–1.00)	0.58 (0.50–0.67)*	0.86 (0.72–0.93)*	1.11 (0.91–1.25)*	0.61 (0.49–0.75)*	<0.001
E/A	0.90 (0.80–1.20)	1.30 (1.10–1.50)*	0.80 (0.70–0.90)*	0.80 (0.70–0.90)*	1.25 (1.00–1.90)*	<0.001
é, cm/s	6.0 (4.4–7.6)	9.0 (8.0–11.0)*	6.7 (6.0–7.0)*	4.3 (4.0–5.0)*	5.7 (4.5–7.0)‡	<0.001
E/é	12.5 (9.2–17.9)	8.3 (7.1–9.9)*	10.4 (8.5–12.0)*	18.2 (14.5–24.5)*	13.8 (10.4–19.5)†	<0.001
LA volume index, ml/m^2	34 (27–43)	28 (22–34)*	27 (22–33)*	38 (31–48)*	42 (33–55)*	<0.001
TR peak velocity, m/s	2.30 (2.00–2.70)	2.10 (1.90–2.30)*	2.20 (2.00–2.50)‡	2.50 (2.20–2.80)*	2.40 (2.10–2.84)†	<0.001

Values are n (%) or median (interquartile range). *$p < 0.001$. †$p < 0.01$. ‡$p < 0.05$ between the region and the remaining regions, Kolmogorov-Smirnov test. Overall p values are calculated using analysis of variance or Kruskal-Wallis test.

A = late diastolic transmitral flow velocity; COPD = chronic obstructive pulmonary disease; CKD = chronic kidney disease; é = early diastolic relaxation velocity at septal mitral annular position; E = early diastolic transmitral flow velocity; EF = ejection fraction; LA = left atrial; LV = left ventricular; MACE = major adverse cardiovascular event(s); TR = tricuspid regurgitation.

STATISTICAL ANALYSIS. Between-group comparisons were performed using Pearson's chi-square test or Fisher exact test (for categorical variables), and analysis of variance, Kruskal-Wallis test or Kolmogorov-Smirnov test (for continuous variables); use of the preceding statistical tests to compare patient groups within the topological network was previously performed (10,19). Correlations between categorical variables were computed using Goodman and Kruskal's γ coefficient. The rates of hospitalization and survival were analyzed with Cox proportional hazard models, Kaplan-Meier curves, and the log-rank test. Cox proportional hazard models were constructed to elucidate independent prognostic values of region information after adjustment with ACC/AHA heart failure stage, NYHA functional class, and the MAGGIC risk score. Furthermore, to evaluate improvement of Cox proportional hazard models by adding region information to these risk predictors, integrated discrimination improvement, net reclassification improvement, and median improvement in risk score were calculated using R package survIDINRI version 1.1-1 (R Foundation, Vienna, Austria) (20,21). A p value of <0.05 was considered statistically significant. We used R version 3.4.0 (R Foundation for Statistical Computing, Vienna, Austria) for all statistical analyses.

RESULTS

Clinical characteristics of the study population are shown in Tables 1 and 2.

CONTINUUM OF CARDIAC FUNCTION. The use of TDA to create a patient–patient similarity network in the retrospective dataset resulted in the formation of a looped structure. After the addition of cases from the prospective data, the loop was persistent, which validated that the phenotypic network loop structure of the network model was intrinsic to the data and not an artifact. The combined network was used for discovering feature distributions and developing associations with clinical and outcome data. We noted that echocardiographic variables followed a gradually changing pattern throughout the loop (Figure 2). Moving counterclockwise, starting from the top of the loop, we observed gradually decreasing EF and é values and increasing LV mass index, E/é ratio, left atrial volume index, and tricuspid regurgitation peak velocity values. In addition to the echocardiographic features, cardiovascular risk factors, MACE-related

TABLE 2 Clinical and Echocardiographic Characteristics of the Prospective Cohort

	Prospective Cohort (n = 468)	Region I (n = 165)	Region II (n = 113)	Region III (n = 59)	Region IV (n = 112)	Overall p Value
Demographics						
Male	195 (42)	72 (44)	33 (29)*	12 (20)†	68 (61)†	<0.001
Age, yrs	57 (47–66)	46 (36–56)†	60 (53–66)†	66 (58–75)†	63 (53–71)*	<0.001
Body mass index, kg/m^2	30.7 (25.7–36.5)	29.5 (25.1–37.1)	31.5 (25.1–37.8)	28.6 (24.4–36.3)	31.0 (27.3–34.9)	0.509
SBP, mm Hg	126 (114–140)	122 (110–135)‡	124 (115–140)	135 (122–147)*	127 (111–147)	<0.001
DBP, mm Hg	75 (68–82)	75 (69–83)	76 (68–81)	72 (68–80)	73 (66–79)	0.252
Hypertension	318 (68)	88 (53)†	79 (70)	45 (76)	91 (81)†	<0.001
Hyperlipidemia	337 (72)	101 (61)†	89 (79)	47 (80)	84 (75)	0.003
Diabetes mellitus	115 (25)	22 (13)†	26 (23)	16 (27)	43 (38)†	<0.001
COPD§	53 (19)	13 (8)	7 (6)‡	4 (7)	28 (25)†	<0.001
Tobacco abuse	206 (44)	63 (38)‡	53 (47)	26 (44)	58 (52)	0.131
History of CAD	114 (24)	12 (7)†	22 (19)	19 (32)	55 (49)†	<0.001
History of CVA§	31 (11)	7 (7)	4 (6)	3 (12)	15 (21)*	0.026
History of CKD	81 (17)	5 (3)†	12 (11)	16 (27)‡	41 (37)†	<0.001
HF diagnosed ≥18 months	68 (14)	1 (1)†	2 (2)†	10 (17)	53 (47)†	<0.001
Medications						
ACEI/ARB	158 (34)	38 (23)†	43 (38)	25 (42)	45 (40)	0.003
Beta-blocker	160 (34)	34 (21)†	29 (26)‡	23 (39)	66 (58)†	<0.001
Calcium channel blocker#	71 (17)	20 (12)	23 (20)	8 (16)	13 (21)	0.208
Statin#	166 (41)	47 (28)†	53 (47)	26 (53)‡	28 (44)	0.001
Clinical staging and outcomes						
NYHA functional class, I/II/III/IV	268/123/62/15	123/32/9/1†	63/44/6/0†	25/25/9/0*	44/19/35/14†	<0.001
ACC/AHA stage, A/B/C/D	122/168/171/7	80/43/42/0†	26/37/50/0	4/28/26/1*	8/48/50/6†	<0.001
MAGGIC score	11 (7–16)	9 (5–12)‡	12 (9–15)†	14 (11–20)†	18 (9–26)†	<0.001
MACE rehospitalization	38 (8)	3 (2)†	4 (4)‡	5 (8)	26 (23)†	<0.001
MACE death	9 (2)	0 (0)‡	1 (1)	0 (0)	8 (7)†	<0.001
Echocardiography						
LVEF, %	62 (57–66)	62 (59–65)*	65 (62–68)†	63 (59–67)	54 (39–60)†	<0.001
LV mass index, g/m^2	74 (60–95)	64 (55–75)†	64 (56–76)†	83 (71–102)*	108 (89–136)†	<0.001
E, m/s	0.79 (0.65–0.94)	0.81 (0.69–0.92)‡	0.69 (0.60–0.82)†	0.81 (0.63–0.94)	0.90 (0.69–1.11)†	<0.001
A, m/s	0.71 (0.57–0.87)	0.62 (0.52–0.72)†	0.86 (0.75–0.94)†	0.99 (0.85–1.16)†	0.58 (0.42–0.70)†	<0.001
E/A	1.10 (0.82–1.40)	1.28 (1.10–1.50)†	0.82 (0.73–0.91)†	0.79 (0.70–0.93)†	1.43 (1.10–2.10)†	<0.001
é, cm/s	7.0 (5.3–9.0)	9.1 (8.0–11.0)†	7.0 [6.0–8.0)†	5.0 (4.0–5.6)†	5.0 (4.0–6.5)‡	<0.001
E/é	10.5 (8.4–14.4)	8.4 (6.9–10.1)†	10.2 (9.0–11.9)†	15.5 (13.4–19.4)†	16.5 (11.2–25.0)†	<0.001
LA volume index, ml/m^2	29 (22–37)	25 (21–31)†	24 (19–29)†	36 (28–43)†	42 (34–56)†	<0.001
TR peak velocity, m/s	2.19 (1.90–2.50)	2.01 (1.70–2.30)†	2.10 (1.81–2.35)	2.40 (2.10–2.63)‡	2.42 (2.11–2.90)†	<0.001

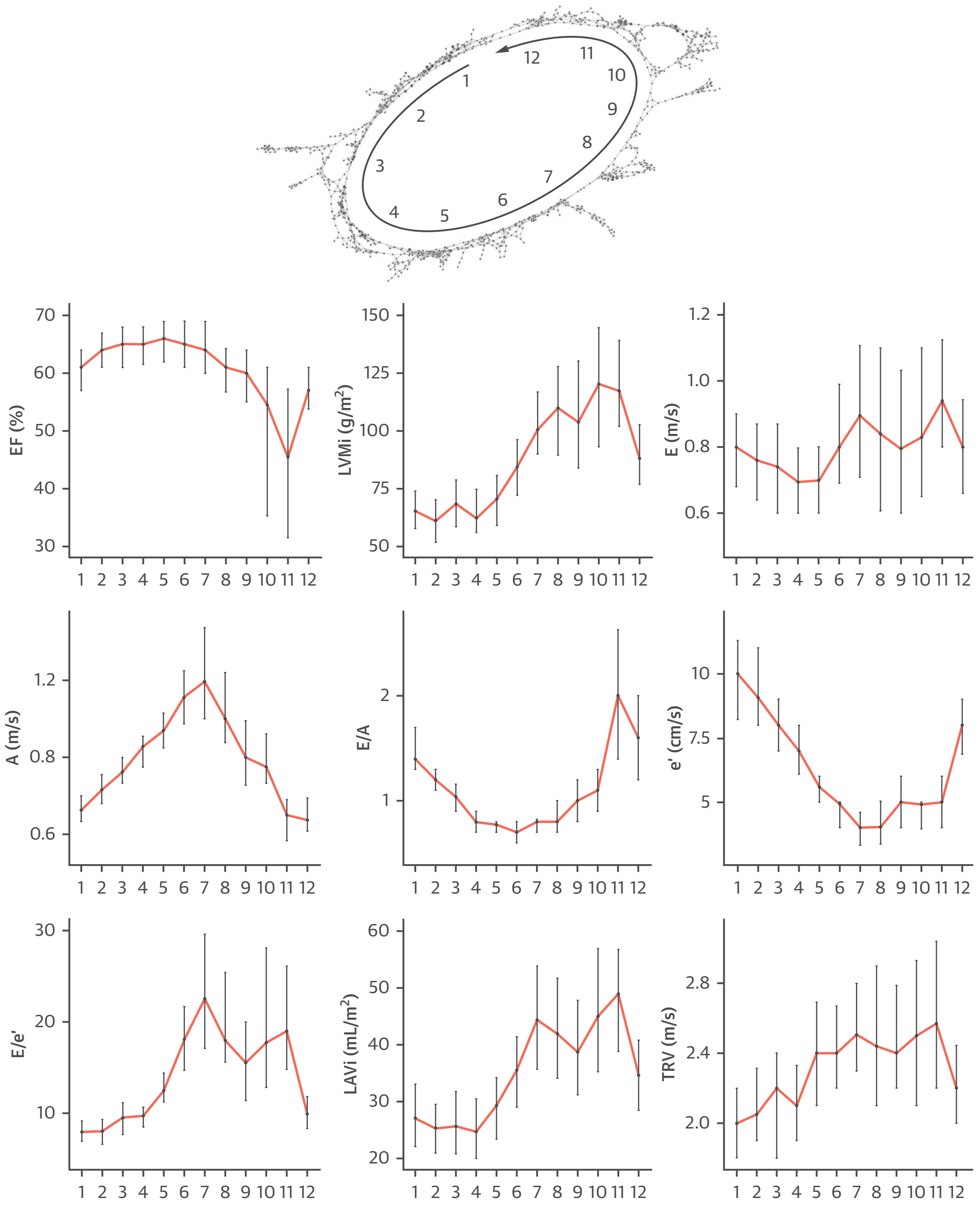
Values are n (%), median (interquartile range), or n. *p < 0.01. †p < 0.001. ‡p < 0.05; between the region and the remaining regions, Kolmogorov-Smirnov test. Overall p values are calculated using analysis of variance or Kruskal-Wallis test. §Data available only for 274 patients. #Data available only for 407 patients.

ACC/AHA = American College of Cardiology/American Heart Association; ACEI = angiotensin-converting-enzyme inhibitor; ARB = angiotensin II receptor blocker; CAD = coronary artery disease; CVA = cerebrovascular accident; DBP = diastolic blood pressure; HF = heart failure; HDL = high-density lipoprotein; LDL = low-density lipoprotein; MAGGIC = Meta-Analysis Global Group in Chronic Heart Failure; NYHA = New York Heart Association; SBP = systolic blood pressure; other abbreviations as in Table 1.

rehospitalization, and death also showed accumulation in distinct sections of the loop (Table 1).

Upon seeing the gradations of echocardiographic variables, and to create clinically useful categories, we measured multidimensional Euclidean distance of nodes using 9 echocardiographic parameters that were used to create the network (Supplemental Appendix). The distance was used to create the groups—which was correlated with the gradation of the variables—on the network and subsequently collated based on clinical characteristics into 4 regions (Central Illustration). The 4 groups were chosen for reaching equivalence to 4 empiric categories of symptoms or disease class used in clinical cardiology (NYHA functional classes I to IV and ACC/AHA heart failure class A to D). Patients in each region showed unique differences in clinical characteristics (Central Illustration, Tables 1 and 2, Supplemental Tables 1 and 2). Notably, progressing from the first to the fourth region, an increasing trend was seen in age and prevalence of underlying risk factors and comorbidities. LVEF remained preserved in the first to third regions; however, it was significantly reduced in the fourth region (p < 0.001). LV mass

FIGURE 2 Gradation of Echocardiographic Features on TDA Network

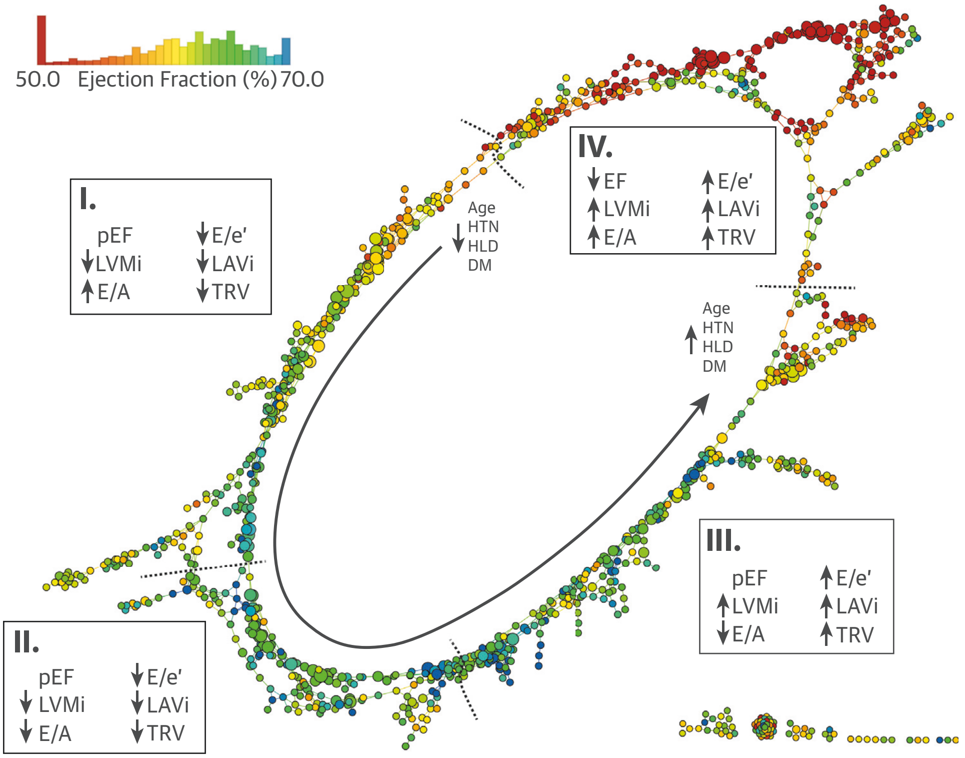


The gradation of echocardiographic features while moving counterclockwise on the loop. Abbreviations as in Figure 1.

progressively increased, and diastolic function parameters progressively worsened from region I to IV. We found a correlation between the regions and both NYHA functional classes ($\gamma = 0.47$; $p < 0.001$) and ACC/AHA stages ($\gamma = 0.52$; $p < 0.001$) in the prospective cohort, which showed more symptomatic patients in the third and fourth regions than in the first region (Figure 3A).

ASSOCIATION OF REGIONS WITH MACE. The median follow-up time in the primary cohort was 309 days (quartile 1 to 3: 100 to 531 days). A total of 207 (16%) patients were observed to have MACE-

CENTRAL ILLUSTRATION The Looped Network of Cardiac Dysfunction



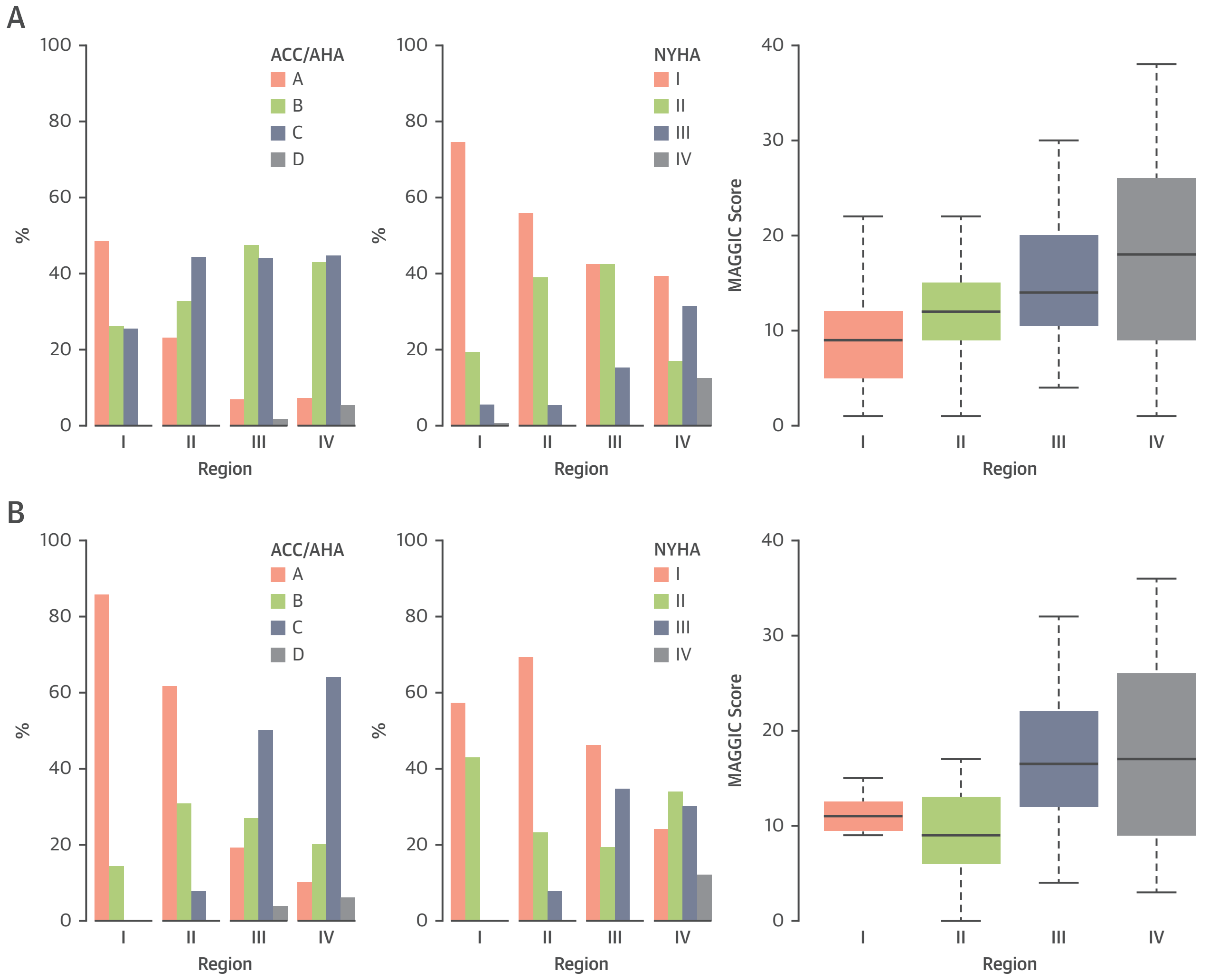
Tokodi, M. et al. J Am Coll Cardiol Img. 2020;13(5):1119–32.

Multiparametric echocardiographic datasets were used to develop a patient–patient similarity network using topological data analysis. Nodes indicate ≥1 patient who have similar echocardiographic characteristics, and nodes including similar patients are connected by edges. The color of nodes has been colorized with the mean ejection fraction (EF) of the node. Moving counterclockwise along with the **gray arrow** starting from the top of the loop, 4 regions were identified with different clinical presentations and outcomes. Region I consisted mostly of patients with risk factors but no obvious symptoms or disease. Patients in region II had more cardiac risk factors (especially hypertension [HTN]) with impaired left ventricular LV relaxation. Region III showed presence of patients with advanced diastolic dysfunction and heart failure with preserved EF (pEF) and pulmonary HTN, but region IV included patients with heart failure with reduced EF with increased LV mass, left atrial volume and pulmonary pressures. Although the map is developed using cross-sectional data, distinct regions of the networks correspond to distinct parts of the disease, along which patients can move on the map, signaling progression, treatment, and recovery of the disease. DM = diabetes mellitus; E/A = early to late diastolic transmitral flow velocity ratio; E/é = early diastolic transmitral flow to annular velocity ratio; HLD = hyperlipidemia; LAVi = left atrial volume indexed to body surface area; LVMi = LV mass indexed to body surface area; TRV = tricuspid regurgitation velocity.

related hospitalizations, and 19 (1%) patients died due to MACE during follow-up. The number of MACE-related hospitalizations increased progressively from regions I to IV ($p < 0.001$), with MACE-related deaths seen only in the third and fourth regions ($p < 0.001$). The Kaplan-Meier curves for MACE-related rehospitalization in the regions differed significantly ($p < 0.001$) (**Figure 4A**). Patients in the fourth region had a >5-fold increased risk of re-hospitalization (hazard ratio [HR]: 5.89; 95% confidence interval [CI]: 3.39 to 10.24; $p < 0.001$), whereas patients in the third region had a >6-fold increased risk of re-hospitalization (HR: 6.88; 95% CI: 3.98 to 11.90; $p < 0.001$) compared with the first region. Subjects in the second region did not have a significantly increased risk of hospital admission due to MACE (HR: 1.45; 95% CI: 0.72 to 2.93; $p = 0.301$) compared with those in the first region. The number of MACE-related deaths was low in the first and second regions, whereas the probability of death was significantly higher in the third and fourth regions than in the first region ($p < 0.001$) (**Figure 4B**).

INDIVIDUALIZED PATIENT PREDICTIONS. Individualized patient predictions for clinical stages, severity, and future adverse events were tested in a secondary cohort analysis. Detailed demographics

FIGURE 3 Heart Failure Stages and Classification by Four Regions on TDA Network

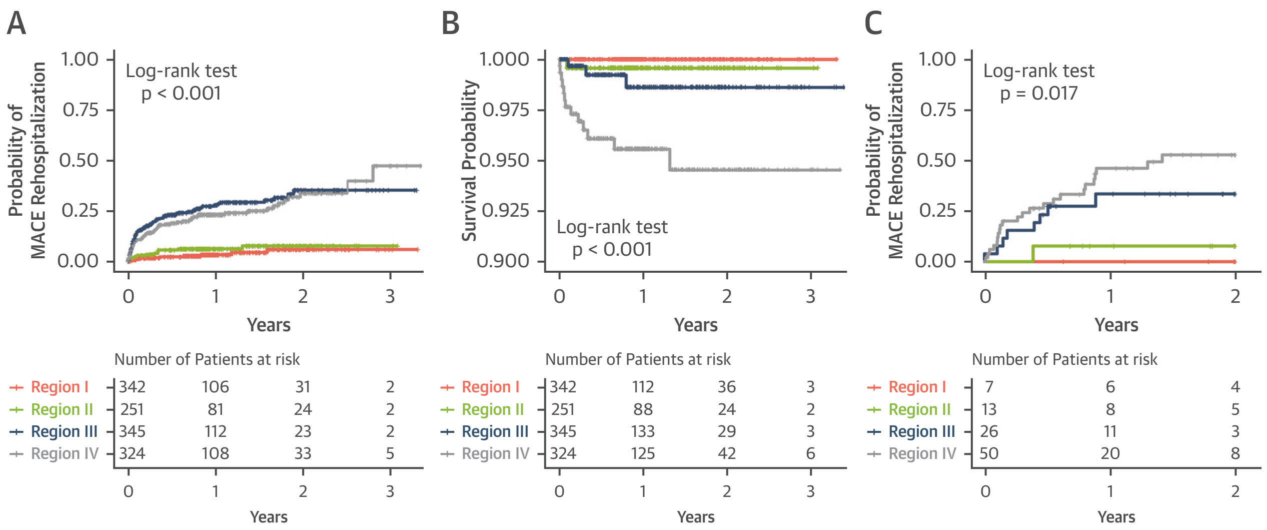


New York Heart Association (NYHA) functional classes and American College of Cardiology/American Heart Association (ACC/AHA) heart failure stages in the 4 regions of the **(A)** primary and **(B)** secondary cohort analysis. MAGGIC = Meta-Analysis Global Group in Chronic Heart Failure.

and the region comparison in this cohort are shown in Table 3 and Supplemental Table 3. After predicting the region membership of patients from the secondary cohort using the random forest classifier, the same tendency was observed for the probability of MACE-related rehospitalizations in the regions as was those in the primary cohort (Figure 4C, Supplemental Table 3). Subjects predicted to be in the fourth region had a >2-fold increased risk of MACE-related rehospitalization (HR: 2.75; 95% CI: 1.27 to 45.95; p = 0.010), whereas those belonging to the third and second regions were not associated with a significantly increased risk compared with those in the first region (HR: 0.95; 95% CI: 0.42 to 2.11; p = 0.890; and HR: 0.17 95% CI: 0.02 to 1.22; p = 0.078, respectively). Patients in the first region were free of any MACE. A correlation of NYHA functional classes and ACC/AHA stages with regions was also observed ($\gamma = 0.56$ and $\gamma = 0.67$; both $p < 0.001$, respectively), which indicated that more symptomatic patients were found in the fourth region than in other regions (Figure 3B).

We also wanted to demonstrate whether changing the location of a patient on the loop was associated with worsening or improvement of cardiac function. To illustrate the motion of patients on the loop, the predicted regions of the first and second echocardiograms were compared (Figure 5). Both

FIGURE 4 **Kaplan-Meier Curve for MACE-Related Outcomes by TDA Regions**



Number of Patients at risk

	0	1	2	3
Region I	342	106	31	2
Region II	251	81	24	2
Region III	345	112	23	2
Region IV	324	108	33	5

Years

Number of Patients at risk

	0	1	2	3
Region I	342	112	36	3
Region II	251	88	24	2
Region III	345	133	29	3
Region IV	324	125	42	6

Years

Number of Patients at risk

	0	1	2
Region I	7	6	4
Region II	13	8	5
Region III	26	11	3
Region IV	50	20	8

Years

Kaplan-Meier curves of the 4 regions: **(A)** major adverse cardiovascular event(s) (MACE)-related rehospitalization in the primary cohort, **(B)** MACE-related death in the primary cohort, and **(C)** MACE-related rehospitalization in the secondary cohort. Abbreviation as in **Figure 1**.

echocardiograms in 13 patients were in low-risk regions (region I or II), whereas those in 63 patients were in the high-risk region (region III or IV). Fifteen patients showed improvement (moved from region III and/or IV to region I and/or II) in echocardiographic results, and 5 patients showed worsening (moved from region I and/or II to region II and/or IV) echocardiographic results. Improvement or staying in the low-risk regions was associated with lower MACE-related rehospitalization rates after the second echocardiogram was performed than worsening or staying in high-risk regions (3% vs. 37%; p < 0.001).

DISCRIMINATION AND RECLASSIFICATION. The incremental value of the topological regions was assessed in the prospective cohort. Even after adjustment with NYHA functional class, ACC/AHA heart failure stages, and MAGGIC scores, the predictive value of being in region IV was consistently significant with a 8- to 10-fold risk (**Table 4**). Net reclassification improvement, integrated discrimination improvement, and median improvement in risk score consistently showed that adding region information to NYHA functional class, ACC/AHA stage, and MAGGIC score significantly improved the prediction of the MACE-related events (**Table 4**). A combination of NYHA functional class (symptoms) and region information (cardiac function) performed better than that of the ACC/AHA stage, which also accounted for symptoms and cardiac function (C-index: 0.819 vs. 0.720).

DISCUSSION

The notion of patient similarity is a growing idea in personalized predictive analytics to support clinical assessment (10,22-25). Patient similarity is a method that can empower precision medicine to stratify patients into clinically relevant subgroups (22). Such subgroup identification generally involves the use of unsupervised machine learning methods for clustering patients (26). However, most clustering techniques discretize the continuous patient data to develop discrete groups, using arbitrary thresholds. In contrast, TDA refers to a collection of powerful geometric approaches that integrate complex high-dimensional data to develop a patient–patient similarity network. The network involves partial clustering, which allows cluster overlaps that illustrate the entire study population as a continuous network of similar patients. This method can allow us to capture the notion of connectivity and continuum to describe the different stages of a disease.

Using retrospectively and prospectively collected echocardiography data from 1,334 patients, we illustrated, for the first time, the potential role of a patient–patient similarity network for mapping cardiac dysfunction without the constraint of any a priori diagnostic system in varying degrees of LV structural and functional remodeling. Specifically, the TDA model in our analysis clustered the multiparametric data without using a hierarchical structure

TABLE 3 Clinical and Echocardiographic Characteristics of the Longitudinal Cohort

	Longitudinal Cohort (n = 96)	Region I (n = 7)	Region II (n = 13)	Region III (n = 26)	Region IV (n = 50)	Overall p Value
Demographics						
Male	47 (49)	3 (43)	4 (31)	7 (27)*	33 (66)†	0.004
Age, yrs	59 (50–67)	53 (30–59)*	55 (46–59)	65 (57–72)*	57 (49–66)	0.005
Body mass index, kg/m^2	31.0 (27.4–34.7)	24.2 (22.2–27.9)‡	29.3 (28.3–34.7)	31.4 (27.6–34.6)	31.6 (28.2–35.1)	0.020
SBP, mm Hg	131 (117–147)	117 (111–134)	126 (120–142)	145 (123–151)	130 (112–149)	0.100
DBP, mm Hg	79 (72–84)	77 (71–84)	82 (72–86)	77 (73–80)	80 (72–84)	0.684
Hypertension	83 (86)	3 (43)‡	12 (92)	26 (100)*	42 (84)	0.002
Hyperlipidemia	75 (78)	4 (57)	9 (69)	23 (88)	39 (78)	0.219
Diabetes mellitus	38 (40)	0 (0)*	4 (31)	15 (58)*	19 (38)	0.029
COPD	27 (28)	3 (43)	2 (15)	7 (27)	15 (30)	0.620
Tobacco abuse	49 (51)	4 (57)	6 (46)	11 (42)	28 (56)	0.681
History of CAD	57 (59)	2 (29)	6 (46)	16 (62)	33 (66)	0.203
History of CVA	21 (22)	1 (14)	5 (38)	4 (15)	11 (22)	0.430
History of CKD	23 (24)	0 (0)	0 (0)*	8 (31)	15 (30)	0.034
HF ≥18 months	37 (39)	0 (0)*	0 (0)‡	10 (38)	27 (54)‡	<0.001
Medications						
ACEI/ARB	45 (47)	2 (29)	7 (54)	13 (50)	23 (46)	0.772
Beta-blocker	52 (54)	2 (29)	5 (38)	15 (58)	30 (60)	0.277
Calcium channel blocker	16 (17)	2 (29)	3 (23)	7 (27)	4 (8)	0.249
Statin	40 (42)	3 (43)	6 (46)	15 (58)	16 (32)	0.091
Clinical risks and outcomes						
NYHA functional class, I/II/III/IV	37/28/25/6	4/3/0/0	9/3/1/0	12/5/9/0	12/17/15/6‡	0.026
ACC/AHA Stage, A/B/C/D	24/22/46/4	6/1/0/0‡	8/4/1/0‡	5/7/13/1	5/10/32/3‡	<0.001
MAGGIC score	15 (9–22)	11 (10–13)	9 (6–13)*	17 (12–22)	17 (9–26)	0.009
MACE rehospitalization	32 (33)	0 (0)	1 (8)	8 (30)	23 (46)‡	0.009
MACE death	4 (4)	0 (0)	0 (0)	0 (0)	4 (8)	0.525
Echocardiography (first echocardiogram)						
LVEF, %	56 (47–62)	55 (54–59)	63 (60–69)‡	59 (54–64)	50 (33–58)†	<0.001
LV mass index, g/m^2	96 (75–117)	54 (48–62)†	60 (40–73)†‡	94 (85–112)	103 (94–126)†	<0.001
E, m/s	0.79 (0.67–1.01)	0.67 (0.62–0.87)	0.72 (0.67–0.79)	0.78 (0.69–1.08)	0.83 (0.69–1.08)	0.221
A, m/s	0.75 (0.54–0.94)	0.64 (0.48–0.65)*	0.89 (0.79–0.91)	1.02 (0.91–1.19)†	0.60 (0.42–0.75)†	<0.001
E/A	1.06 (0.80–1.67)	1.30 (1.26–1.52)	0.88 (0.79–0.92)*	0.77 (0.68–0.91)†	1.64 (1.07–2.18)†	<0.001
é, cm/s	6.0 (5.0–8.0)	10.0 (8.5–10.5)*	8.0 (6.0–8.0)	6.0 (4.0–7.0)	6.0 (4.0–8.0)	0.002
E/e′	12.6 (9.5–17.1)	8.2 (6.2–10.7)*	9.8 (8.6–13.2)*	15.0 (11.6–18.0)	13.1 (9.9–21.2)	0.001
LA volume index, ml/m^2	31 (22–41)	20 (16–22)‡	22 (19–25)‡	32 (28–36)	39 (28–48)†	<0.001
TR peak velocity, m/s	2.33 (1.88–2.69)	1.86 (1.35–2.20)	2.10 (1.95–2.21)*	2.11 (1.76–2.69)	2.55 (2.06–2.94)‡	0.016

Values are n (%), median (interquartile range), or n. *$p < 0.05$. †$p < 0.001$. ‡$p < 0.01$, between the region and the remaining regions, Kolmogorov-Smirnov test. Overall p values are calculated using analysis of variance or Kruskal-Wallis test.

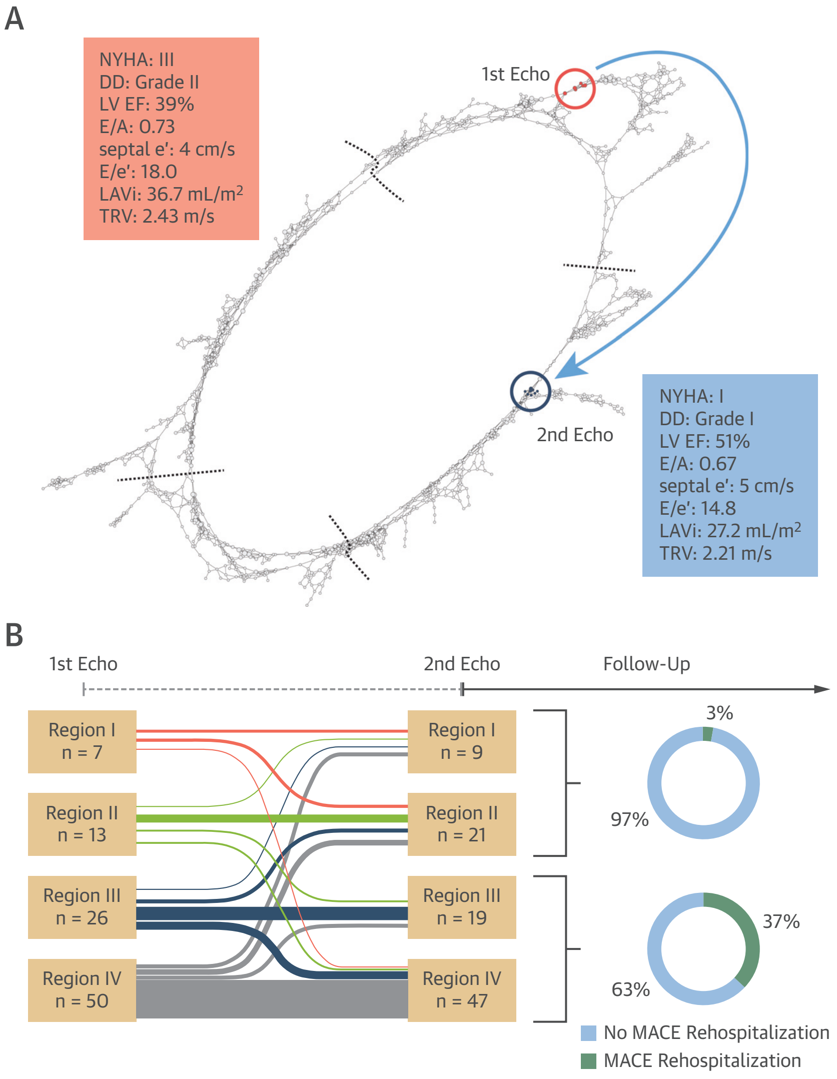
Abbreviations as in Tables 1 and 2.

or branching tree (4) but rather meaningfully represented the geometry of the data based on the similarity of the patients. Remarkably, the nodes clustered to produce a network in the form of a loop. Moreover, this loop demonstrated the relationships with the outcome of interest, which suggested a valid method of risk stratification for patients. We further illustrated the potential value of this loop for individualized prediction. Using a group of patients with longitudinally collected echocardiographic studies in which the patients were sampled in different stages of cardiac dysfunction, our analysis suggested that this looped space might also represent the periodic or recurrent behavior of the disease, thereby tracing the path that patients traveled through cycles of worsening cardiac function and recovery.

PATIENT SIMILARITY VERSUS AN AVERAGE PATIENT. Echocardiography remains the most versatile tool in clinical practice, offering an ever-increasing array of measurements. Although novel multivariate data-driven analytic approaches to stratifying cardiac dysfunction have been recently made available (27), the integration of clinical and echocardiographic data for precision phenotyping has been arduous using traditional techniques.

Classically, studies and results are based on meticulous experimental designs and statistical analyses to produce exhaustive results for an average

FIGURE 5 **Inter-Region Motion of Secondary Cohort Patients on TDA Network**



(A) Pre- and post-operative position of a representative patient (74 years of age, female) who underwent coronary artery bypass grafting on the loop. **(B)** Inter-region motion of patients in the secondary cohort. DD = diastolic dysfunction; other abbreviations as in Figures 1, 3, and 4.

patient. However, no two patients are alike, making it difficult to generalize study results for average patients to actual patients with cardiac dysfunction. To this end, novel bioinformatics and machine-learning approaches have been suggested to help support integration of high-dimensional data for rapid medical decision-making (28-31).

The concept of the patient similarity network using TDA has been shown in well-known studies, such as those aiming to subgroup patients with diabetes (10), to identify individuals resistant to malaria infections (15), and patients with pre-clinical traumatic brain and spinal cord injury (11). TDA has specifically enabled real-time exploration of the concept of disease space. For example, Torres et al. (15) demonstrated a similar loop in analyzing disease tolerance to malaria in mice and humans and stratified the resilience of patients based on the size of the loop through the disease space. Similarly, we demonstrated a gradual change in echocardiographic variables throughout the disease cycle that outlined similarity among patients and that described different phenotypes of cardiac functions in the disease space. Despite the abundance and complexity of echocardiographic features, distinct paths emerged for the patients with cardiac dysfunction in clockwise or counterclockwise directions on

TABLE 4 Independency and Incremental Value of Regions Upon Clinical Risk Predictors

	Adjusted HR*				Model Improvement†		
	HR	95% CI	p Value		Estimate	95% CI	p Value
Model with NYHA functional class + regions: C-index 0.819 vs. 0.749 for model without regions							
Region II	2.20	0.53–9.21	0.280	IDI	0.079	0.017–0.166	0.007
Region III	3.90	0.92–16.46	0.064	NRI	0.775	0.077–0.899	0.027
Region IV	8.87	2.53–31.05	<0.001	MIRS	0.058	0.002–0.245	0.020
Model with ACC/AHA stage + regions: C-index 0.815 vs. 0.720 for model without regions							
Region II	1.91	0.45–8.00	0.380	IDI	0.187	0.105–0.288	0.007
Region III	3.72	0.88–15.77	0.074	NRI	0.737	0.268–0.903	0.007
Region IV	12.39	3.71–41.35	<0.001	MIRS	0.275	0.118–0.386	0.013
Model with MAGGIC score + regions: C-index 0.815 vs. 0.775 for model without regions							
Region II	2.01	0.48–8.43	0.340	IDI	0.098	0.033–0.192	0.007
Region III	2.92	0.67–12.76	0.160	NRI	0.437	0.071–0.821	0.020
Region IV	8.16	2.25–29.58	0.001	MIRS	0.146	0.025–0.250	0.013

*Summarizes hazard ratio (HR) for each region adjusted by NYHA functional class, ACC/AHA stage, and Meta-Analysis Global Group in Chronic heart failure (MAGGIC) score, respectively. †Summarize model improvement by adding region information upon each risk predictors.

CI = confidence interval; IDI = integrated discrimination improvement; MIRS = median improvement in risk score; NRI = net reclassification improvement; other abbreviations as in Table 2.

the loop based on the progression, treatment, and recovery of the disease.

CLINICAL IMPLICATIONS. There are several pathophysiological and clinical implications for our study. First, the continuity of our patient similarity network suggested the pathophysiological classification of cardiac dysfunction should be viewed as a continuum rather than as arbitrary divisions of the patient population into discrete subgroups as heart failure with reduced, mid-range, or preserved EF. Measures of LV systolic and diastolic function did not exhibit abrupt changes at any level of cardiac function, but they covered a gradual and continuous spectrum, creating an overlapping and interconnected spectrum of disease phenotypes that was previously suggested but not shown (32). Second, the 4 regions of the loop showed incremental value over NYHA functional class, ACC/AHA stages, and commonly used risk scores (e.g., MAGGIC risk score), which suggested the clinical usefulness of this approach in patient risk stratification. Finally, unlike the consensus-driven algorithms (e.g., NYHA functional class and ACC/AHA stage) that first use expert knowledge and then develop the stages and decision pathways, the computational technique described in this study learns automatically and requires no a priori knowledge or training to develop meaningful disease representation. This ability to integrate multiple parameters pragmatically to define patient phenotypes and to reproduce known clinical knowledge provides a strong foundation for why a provider could rely on this simplified staging scheme. Moreover, the TDA characterization system allows identifying patients on a disease map much like the Global Positioning System; such illustrated steps can help with automated classification, risk stratification, and monitoring progression or response to therapies. Such decision support systems are critically needed not only for clinical care but also for clinical trials in which heterogeneity of disease presentation affects patient matching and discovery of novel therapies.

STUDY LIMITATIONS. The follow-up duration and patient sample size for patients with reduced EFs was modest and potentially averted us from capturing a greater number of cardiac events to test the applicability, disease trajectory over time, and ability of our model to predict measurable isolated endpoints. Furthermore, specific therapeutic interventions that targeted any specific region on the loop or types of therapies that could change the patients' prognoses were not included in the study and would be a logical next step that would be required to address in future studies. Addition of biomarkers and novel echocardiographic parameters, such as strain and strain rate, could provide an added benefit to the model and should also be investigated in the future.

CONCLUSIONS

TDA may have broad implications on developing clinical risk stratification schemes using patient–patient similarity networks. TDA can be used for an integrated assessment of multiple echocardiographic parameters that measure the extent of cardiac structural and functional remodeling. Moreover, such topological networks may be a viable data analytical approach to trace the progression of cardiac dysfunction in patients as they travel through cycles of compensation and decompensation within a looped disease space. Overall, identifying diverse cardiac phenotypes brings us one step closer to precision medicine.

ADDRESS FOR CORRESPONDENCE: Dr. Partho P. Sengupta, Heart & Vascular Institute, West Virginia University, 1 Medical Center Drive, Morgantown, West Virginia 26506-8059. E-mail: partho.sengupta@wvumedicine.org.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: A patient similarity network integrated echocardiographic parameters of cardiac structure and function to develop a looped network in which patients could be mapped precisely to specific disease stage and clinical outcomes. This network representation allowed automated classification of cardiac function and personalized prediction of MACE in an individual patient.

TRANSLATIONAL OUTLOOK: Patient similarity analysis can be combined with machine learning approaches to offer a practical solution for personalized risk stratification and may enable identification of patient populations with similar risk and those who are likely to respond to targeted therapies.

REFERENCES

1. Benjamin EJ, Blaha MJ, Chiuve SE, et al. Heart disease and stroke statistics' 2017 update: a report from the American Heart Association. Circulation 2017;135:e146–603.

2. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2016; 37:2129–200m.

3. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging 2015;16:1–39.e14.

4. Lum PY, Singh G, Lehman A, et al. Extracting insights from the shape of complex data using topology. Sci Rep 2013;3:1236.

5. Carlsson G. Topology and data. Bulletin of the American Mathematical Society 2009;46: 255–308.

6. Singh G, Memoli F, Carlsson G. Topological methods for the analysis of high dimensional data sets and 3D object recognition. Proc Eurographics Symp Point-Based Graphics 2007:91–100.

7. Nicolau M, Levine AJ, Carlsson G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proc Natl Acad Sci U S A 2011; 108:7265–70.

8. Camara PG, Rosenbloom DIS, Emmett KJ, Levine AJ, Rabadan R. Topological data analysis generates high-resolution, genome-wide maps of human recombination. Cell Syst 2016;3: 83–94.

9. Lakshmikanth T, Olin A, Chen Y, et al. Mass cytometry and topological data analysis reveal immune parameters associated with complications after allogeneic stem cell transplantation. Cell Rep 2017;20:2238–50.

10. Li L, Cheng W-Y, Glicksberg BS, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med 2015;7:311ra174.

11. Nielson JL, Paquette J, Liu AW, et al. Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. Nat Commun 2015;6:8581.

12. Hinks T, Zhou X, Staples K, et al. Multidimensional endotypes of asthma: topological data analysis of cross-sectional clinical, pathological, and immunological data. Lancet 2015;385:S42.

13. Hinks TSC, Zhou X, Staples KJ, et al. Innate and adaptive T cells in asthmatic patients: relationship to severity and disease mechanisms. J Allergy Clin. Immunol 2015;136:323–33.

14. Nagueh SF, Smiseth OA, Appleton CP, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2016;29: 277–314.

15. Torres BY, Oliveira JHM, Thomas Tate A, Rath P, Cumnock K, Schneider DS. Tracking resilience to infections by mapping disease space. PLoS Biol 2016;14:e1002436.

16. WO2019006213A1 - Systems and methods for topological data analysis using nearest neighbors - Google Patents. Available at: https://patents.google.com/patent/WO2019006213A1/en. Accessed October 25, 2019.

17. Rich JD, Burns J, Freed BH, Maurer MS, Burkhoff D, Shah SJ. Meta-Analysis Global Group in Chronic (MAGGIC) heart failure risk score: validation of a simple tool for the prediction of morbidity and mortality in heart failure with preserved ejection fraction. J Am Heart Assoc 2018;7: e009594.

18. Simpson J, Jhund PS, Silva Cardoso J, et al. Comparing LCZ696 with enalapril according to baseline risk using the MAGGIC and EMPHASIS-HF risk scores: an analysis of mortality and morbidity in PARADIGM-HF. J Am Coll Cardiol 2015;66: 2059–71.

19. Casaclang-Verzosa G, Shrestha S, Khalil M, et al. Network tomography for understanding phenotypic presentations in aortic stenosis. J Am Coll Cardiol Img 2018;12:236–48.

20. Long Q, Xu J, Osunkoya AO, et al. Global transcriptome analysis of formalin-fixed prostate cancer specimens identifies biomarkers of disease recurrence. Cancer Res 2014;74: 3228–37.

21. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. Stat Med 2013;32: 2430–42.

22. Parimbelli E, Marini S, Sacchi L, Bellazzi R. Patient similarity for precision medicine: a systematic review. J Biomed Inform 2018;83: 87–96.

23. Zhang P, Wang F, Hu J, Sorrentino R. Towards personalized medicine: leveraging patient similarity and drug similarity analytics. AMIA Jt Summits Transl Sci Proc 2014:32–6.

24. Sharafoddini A, Dubin JA, Lee J. Patient similarity in prediction models based on health data: a scoping review. JMIR Med Informatics 2017;5:e7.

25. Ng K, Sun J, Hu J, Wang F. Personalized predictive modeling and risk factor identification using patient similarity. AMIA Summits Transl Sci Proc 2015:132–6.

26. Kagiyama N, Sirish S, Peter DF, Sengupta PP. Artificial intelligence: practical primer for clinical research in cardiovascular disease. J Am Heart Assoc 2019;8:e012788.

27. Omar AMS, Bansal M, Sengupta PP. Advances in echocardiographic imaging in heart failure with reduced and preserved ejection fraction. Circ Res 2016;119:357–74.

28. Antman EM, Loscalzo J. Precision medicine in cardiology. Nat Rev Cardiol 2016;13:591–602.

29. Mirnezami R, Nicholson J, Darzi A. Preparing for precision medicine. N Engl J Med 2012;366:489–91.

30. Krittanawong C, Zhang HJ, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. J Am Coll Cardiol 2017;69:2657–64.

31. Samad MD, Ulloa A, Wehner GJ, et al. Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning. J Am Coll Cardiol Img 2019;12:681–9.

32. De Keulenaer GW, Brutsaert DL. Systolic and diastolic heart failure are overlapping phenotypes within the heart failure spectrum. Circulation 2011;123:1996–2004.

KEY WORDS echocardiography, patient similarity, topological data analysis

APPENDIX For an expanded Methods section as well as a supplemental figure and tables, please see the online version of this paper.