Journal of the American Heart Association

ORIGINAL RESEARCH

Vocal Biomarker Is Associated With Hospitalization and Mortality Among Heart Failure Patients

Elad Maor, MD, PhD; Daniella Perry, PhD; Dana Mevorach, MA; Nimrod Taiblum, PhD; Yotam Luz, PhD; Israel Mazin, MD; Amir Lerman, MD; Gideon Koren, MD; Varda Shalev, MD, MPH

BACKGROUND: The purpose of this article is to evaluate the association of voice signal analysis with adverse outcome among patients with congestive heart failure (CHF).

METHODS AND RESULTS: The study cohort included 10 583 patients who were registered to a call center of patients who had chronic conditions including CHF in Israel between 2013 and 2018. A total of 223 acoustic features were extracted from 20 s of speech for each patient. A biomarker was developed based on a training cohort of non-CHF patients (N=8316). The biomarker was tested on a mutually exclusive CHF study cohort (N=2267) and was evaluated as a continuous and ordinal (4 quartiles) variable. Median age of the CHF study population was 77 (interquartile range 68–83) and 63% were men. During a median follow-up of 20 months (interquartile range 9–34), 824 (36%) patients died. Kaplan–Meier survival analysis showed higher cumulative probability of death with increasing quartiles (23%, 29%, 38%, and 54%; $P<0.001$). Survival analysis with adjustment to known predictors of poor survival demonstrated that each SD increase in the biomarker was associated with a significant 32% increased risk of death during follow-up (95% CI, 1.24–1.41, $P<0.001$) and that compared with the lowest quartile, patients in the highest quartile were 96% more likely to die (95% CI, 1.59–2.42, $P<0.001$). The model consistently demonstrated an independent association of the biomarker with hospitalizations during follow-up ($P<0.001$).

CONCLUSIONS: Noninvasive vocal biomarker is associated with adverse outcome among CHF patients, suggesting a possible role for voice analysis in telemedicine and CHF patient care.

Key Words: congestive heart failure ■ telemedicine ■ vocal biomarkers ■ voice

Telemedicine allows monitoring of patients remotely and holds the potential to reduce congestive heart failure (CHF) related hospitalizations, improve quality of life, and optimize the use of the limited resources in this field.[1] Data from clinical trials investigating telemedicine-based intervention are conflicting, with some trials suggesting no improvement in patient outcome while other studies suggesting benefits from telemonitoring that include improvement in quality of life, reduction in the risk of all-cause mortality, and CHF-related hospitalizations.[2–6] Wireless implantable hemodynamic monitoring systems of pulmonary artery pressure, for example, allow for improved heart failure management and reduced heart failure hospitalizations.[7,8] Noninvasive biomarker-based approaches for telemedicine use incremental data from home measurements, following acute decompensated CHF events to identify high-risk patients.[9] Voice signal analysis is an emerging noninvasive biomarker. We recently reported a possible relationship between specific vocal biomarkers and coronary artery disease underscoring the potential use of this simple biomarker to identify patients at

Correspondence to: Elad Maor, MD, PhD, The Olga & Lev Leviev Heart Center, Sheba Medical Center, Sheba Medical Center Hospital-Tel Hashomer, 52621 Ramat Gan, Israel. E-mail: elad.maor@sheba.health.gov.il

Supplementary material for this article is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.119.013359

For Sources of Funding and Disclosures, see page 10.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Wearable medical devices and smartphones hold the potential to assist in telemedicine. In the specific case of noninvasive voice analysis, preliminary data suggest a potential relationship between voice and cardiovascular disease.
- Using a large cohort of patients with congestive heart failure who were registered to a call center in Israel, this study describes for the first time how voice analysis can assist in risk stratification of such patients with respect to overall survival and/or hospitalizations.

What Are the Clinical Implications?

- Clinically, this study supports the hypothesis that vocal biomarkers might assist in identification of high-risk subjects in areas where access to services such as physical examination, blood tests, and cardiovascular imaging are limited.

Nonstandard Abbreviations and Acronyms

HR	hazard ratio
ICD-9	*International Classification of Diseases, Ninth Revision*

risk.[10] The objective of the present study was to test the hypothesis that vocal biomarkers are associated with increased risk of hospitalizations and mortality among CHF patients.

METHODS

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Study Population

MOMA is a multidisciplinary healthcare center founded at 2012 by Maccabi HealthCare Services. The Center provides telemedical services to complex patients with various chronic conditions including CHF, lung disease, diabetes mellitus, fragility, homecare, and oncological patients under treatment. It is staffed by multidisciplinary health practitioners including nurses, consulting physicians, clinical pharmacists, physical therapists, social workers, and nutritionists.[11] In this study, we analyzed anonymized medical records and audio recordings of patients enrolled in the center between June 2013 and October 2018 (N=10 583). A periodic phone conversation between nurse and patient is part of the center's protocol, and these sessions were recorded.

The population was divided into 2 mutually exclusive patients' groups: a training cohort of patients with chronic conditions excluding CHF (N=8316) and a CHF study cohort (N=2267). Baseline characteristic of the training population are summarized in Table S1. Inclusion criteria for the current analysis were age >18, at least 20 s of recorded voice, and either Hebrew or Russian language. Patients were excluded from the study if there were technical difficulties with identifying patients' voice (eg, when more than 1 person shares the same phone line) or if there was significant noise according to the subjective impression of the labeler. Voice analysis and data processing were blinded to patients' personal identifiers and clinical data. In addition, in order to avoid hearing the identification stage of the conversation, the first 10 s of each recording were removed from analysis. The Institutional Review Board of Maccabi Health Services approved this study on the basis of strict maintenance of participants' anonymity during database analyses. Data provided for this study are nonidentifiable and deidentified using strict pseudonymization methods. Only specific voice elements, rather than verbal recordings, are being analyzed. The patient can only be re-identified with the aid of a key code, which is not available for the researchers. In agreement with Recital 26 in European Union general data protection regulation, no opting in is required for this type of study. No individual consent was obtained.

Voice Characteristics and Vocal Biomarker Definition

Trained personnel listened to the calls and labeled the time segments in which only the patient's voice is heard. When more than 1 call was available for a patient, the most recent call was used in this study. These manually labeled segments were then concatenated to form 20 s length of voice sample per patient. The voice analysis process was robust and was only marginally influenced by disturbances. Overall, analysis was possible for >99% of the recordings screened.

Low-level acoustic features were extracted from the samples, using "Vocalis Health" voice processing techniques. These feature sets included: Mel Cepstrum representation, Pitch and Formant Measures, Jitter, Shimmer and Loudness.[12] These features were extracted in a temporal resolution of 100 points per second, forming a matrix of 2000 columns, by the number of extracted temporal features (as rows). Figure 1 demonstrates an example of the matrix representation. Next, 600 high-level features are extracted from these images (matrices). These features are the result of an eighth-order moments analysis applied on the time series of low-level

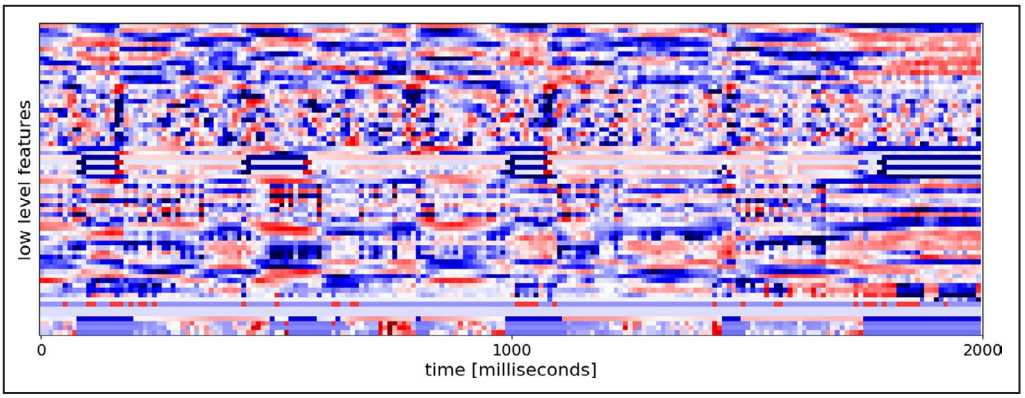

Figure 1. Matrix representation of the acoustic signal. The image demonstrates a 2-s snapshot of such matrix representation of the acoustic signal.
The horizontal is the "time" axis with 100 points per s resolution, and the vertical axis is categorical, where each line represents a specific time-series of low-level features from the abovementioned feature sets.

features mentioned above.[13] Next, features were excluded from the study using a previously described feature selection process,[14] finally giving 223 valid features from each voice sample. The exclusion criterion was a Pearson correlation lower than 0.65. The correlations were applied upon each of the high-level features when calculated twice, on the left and on the right halves of all the images in the training cohort.

A machine learning linear model[15] was then constructed based on the training cohort, while being blinded to the CHF study cohort. For the purpose of the current analysis, this model is described as the vocal biomarker, a unitless unbounded scalar, which is a linear combination of the 223 high-level acoustic features mentioned above. The vocal biomarker was optimized for the training cohort using cross-validation techniques.[16] With cross-validation, various linear models were evaluated (logistic regression, support vector machine, linear discriminant analysis and elastic net) at different regularization levels. The biomarker was calculated based on the training cohort only, and its prediction capabilities were estimated based on the biomarker's hazards ratio (HR) *P* value, with respect to overall survival. In order to estimate the biomarker variability, the biomarker was calculated for 2 separate 20-s segments of the same call for 290 CHF subjects. This analysis demonstrated strong statistical correlation (Pearson's correlation coefficient 0.79, $P=10^{-63}$). In the training cohort, the final model yielded hazards ratio (HR) of 1.66 for all-cause mortality with the biomarker as a continuous variable (95% CI, 1.58–1.74, $P<0.001$) and HR of 2.32 (95% CI, 2.09–2.57, $P<0.001$) with the biomarker dichotomized to highest versus lower quartiles.

CHF Definition and Study End Point

Using advanced networking information technology, the Maccabi Health Services Cardiovascular Registry identified CHF patients as previously described.[17] Diagnosis is based on personal medical records (*International Classification of Diseases, Ninth Revision* [*ICD-9*] codes 404.X and 428.x), hospitalization records, laboratory tests, medications, physiological signals (eg, ECGs), radiological images (eg, echocardiograms, angiograms), and reports from investigations and procedures. CHF diagnosis was then validated by comparing the registry identified *ICD-9* codes with diagnoses and other relevant clinical and personal data derived from a comprehensive review of the patient's medical records and by performing logic tests on aberrant results.

The primary outcome of the current study was all-cause mortality. Survival data were available for all subjects from the Israeli Population Register up to October 2018. Secondary outcome included any hospitalization during follow-up. Hospitalization data were available from Maccabi Health Services medical records.

Statistical Analysis

Study population was categorized to 4 quartiles based on the vocal biomarker with Q1 being the lowest quartile. Variables were expressed as median with interquartile range and as frequency (%) for categorical variables. The 4 biomarker groups were compared using an ANOVA test for continuous variables, and χ^2 test for categorical variables. The probability of death according to the vocal biomarker groups was graphically displayed according to the method of Kaplan–Meier, with comparison of cumulative survival across strata by the log-rank test. Univariate and multivariate Cox proportional hazards regression modeling was used to determine the HR for all-cause mortality of the biomarker as a continuous as well as dichotomous variable with the lower quartile as

a reference. Covariates that were significant in the univariate model were incorporated to the multivariate Cox model. Binary logistic regression with adjustment for age and sex was used to determine the odds ratio for 30-, 90-, and 180 days. The diagnostic accuracy of the vocal biomarker in predicting overall survival in comparison to other predictors of adverse outcome was evaluated using a time-dependent receiver-operating curve analysis. Similar methods were used for the secondary end point of hospitalizations during follow-up. The vocal biomarker was further assessed using the NRI approach in a multivariate logistic regression model for 1-year all-cause mortality, with and without the biomarker. Similar analysis was used for 30-day hospitalizations. Finally, interaction analysis was performed to assess the biomarker with respect to important subgroups of the study. Statistical significance was accepted for a 2-sided $P<0.05$. The statistical analyses were performed with Python programming language (Python Software Foundation, https://www.python.org). All supporting data are available within the article and its online supplementary file.

RESULTS

The final study population comprised 2267 CHF patients. Median age of the study population was 77 (interquartile range 68–83) years and 1434 (63%) were men. Baseline characteristic of the study population are summarized in Table 1. Overall rate of comorbidities in the study cohort was relatively high: there were 1964 (87%) hypertensive patients, 1729 (76%) patients with chronic kidney disease (estimated glomerular filtration rate <60), 624 (28%) patients with active cancer, and 573 (25%) patients with lung disease. During the year before their voice recording, 1420 (63%)

Table 1. Baseline Characteristics of the Study Cohort by Biomarker Quartiles

	Q1 (n=570)	Q2 (n=568)	Q3 (n=563)	Q4 (n=566)	All (n=2267)	*P* Value
Age, y	70 [64, 78]	76 [67, 82]	78 [70, 84]	80 [74, 86]	77 [68, 83]	<0.001
Male (%)	368 (65%)	370 (65%)	364 (65%)	332 (59%)	1434 (63%)	NS
Hebrew language (%)	453 (79%)	361 (63%)	344 (61%)	381 (67%)	1539 (68%)	<0.001
CHF diagnosis (duration, y)	4 [1, 7]	3 [1, 7]	4 [1, 7]	4 [1, 7]	4 [1, 7]	NS
Mean BP, mm Hg	90 [84, 97]	90 [83, 98]	90 [82, 98]	89 [80, 97]	90 [84, 97]	<0.05
Hypertension	469 (82%)	492 (86%)	494 (88%)	509 (90%)	1946 (87%)	NS
DM	343 (60%)	346 (61%)	337 (60%)	328 (58%)	1354 (60%)	NS
CAD	391 (69%)	385 (68%)	395 (70%)	384 (68%)	1555 (69%)	NS
PVD	98 (17%)	96 (17%)	123 (22%)	108 (19%)	425 (19%)	NS
CKD	438 (77%)	477 (84%)	473 (84%)	506 (90%)	1894 (84%)	NS
Lung disease	168 (29%)	125 (22%)	129 (23%)	151 (27%)	573 (25%)	<0.05
Active cancer	134 (24%)	148 (26%)	167 (30%)	175 (31%)	624 (28%)	NS
Hospitalizations*	321 (56%)	360 (63%)	349 (62%)	390 (69%)	1420 (63%)	NS
Obesity (BMI >30)	249 (44%)	250 (44%)	238 (42%)	206 (36%)	943 (42%)	NS
Fasting glucose	111 [99, 145]	114 [100, 140]	113 [97, 144]	110 [95, 138]	112 [98, 141]	NS
eGFR <60	393 (69%)	414 (73%)	446 (79%)	476 (84%)	1729 (76%)	<0.05
Triglycerides	132 [96, 178]	121 [90, 170]	118 [86, 168]	119 [90, 156]	122 [90, 169]	<0.05
TC, mg/dL	155 [132, 182]	150 [129, 182]	148 [128, 177]	153 [127, 181]	152 [129, 180]	NS
β-Blockers	554 (97%)	559 (98%)	550 (98%)	549 (97%)	2212 (98%)	NS
ACE inhibitors	559 (98%)	560 (98%)	556 (99%)	553 (98%)	2228 (98%)	NS
Diuretic	553 (97%)	558 (98%)	552 (98%)	557 (99%)	2220 (98%)	NS
Spironolactone	415 (73%)	408 (72%)	396 (70%)	413 (73%)	1632 (72%)	NS
Lipid-lowering drugs	547 (96%)	536 (94%)	535 (95%)	534 (95%)	2152 (95%)	NS
Vocal biomarker	−1 [−1.3, −0.8]	−0.3 [−0.4, −0.2]	0.3 [0.1, −0.4]	1.2 [0.8, −1.6]	0.00 [−0.6, 0.06]	<0.001
Follow-up, mo	22 [17, 35]	21 [12, 34]	20 [8, 34]	18 [5, 30]	20 [9, 34]	NS
NYHA ≥3	79 (14%)	82 (14%)	86 (15%)	115 (20%)	362 (16%)	<0.05
EF <40	149 (26%)	118 (21%)	130 (23%)	115 (20%)	512 (23%)	NS

ACE indicates angiotensin-converting enzyme; BMI, body mass index; BP, blood pressure; CAD, coronary artery disease; CHF, congestive heart failure; CKD, chronic kidney disease; DM, diabetes mellitus; EF, ejection fraction; eGFR, estimated glomerular filtration rate mL/min per 1.73 m^2; NYHA, New York Heart Association Classification Guidelines; NS, not significant; PVD, pulmonary vascular disease; and TC, total cholesterol.

*Any hospitalization in the previous year.

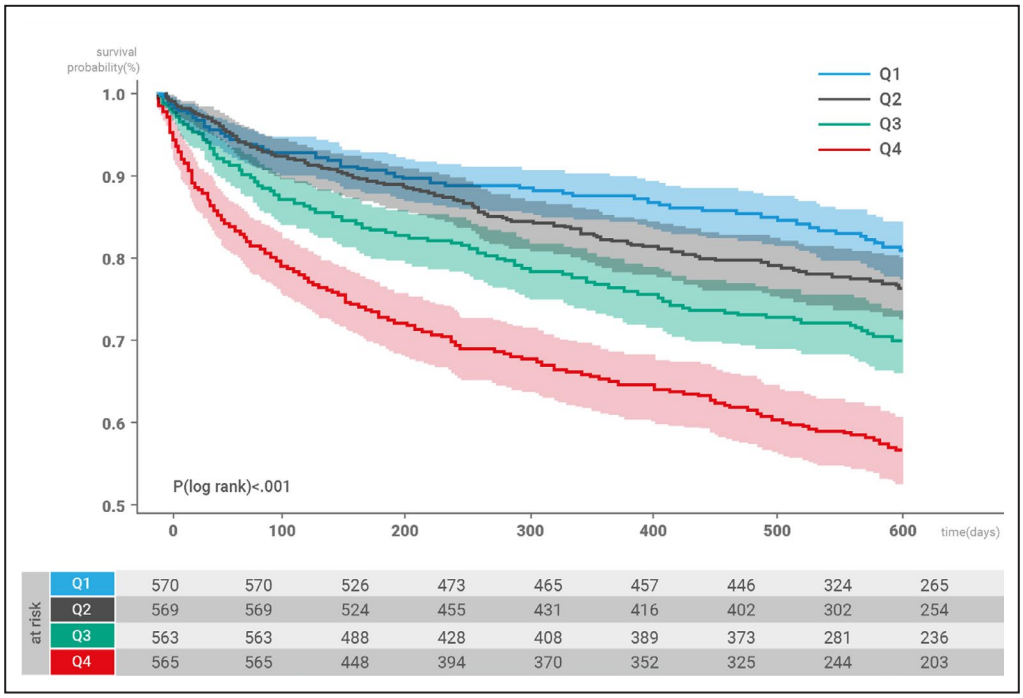

Figure 2. Kaplan–Meier survival analysis.
Kaplan–Meier curve showing overall cumulative survival probability for the 4 biomarker quartile groups.

patients had at least 1 hospitalization. The majority of patients were treated with β-blockers (2212 [98%]), angiotensin-converting enzyme inhibitors (2228 [98%]) and lipid-lowering drugs (2152 [95%]). Aldosterone antagonists were used by 1632 (72%) patients. Subjects with higher vocal biomarker values were older, were more likely to be men, and had higher rates of chronic kidney disease. Comparison of baseline characteristics by the prespecified biomarker groups are summarized in Table 1.

Association of the Vocal Biomarker With Mortality During Follow-Up

There were 824 (36%) deaths during a median follow-up of 20 months (interquartile range 9–34). There were 135 (23%), 168 (29%), 215 (38%), and 306 (54%) deaths in the 4 biomarker quartile groups, respectively. Kaplan–Meier survival analysis showed that the cumulative probability of death in the lowest (Q1) and highest (Q4) biomarker groups at 20 months of follow-up were 53%±11% and 64%±13%, respectively (Figure 2; $P<0.001$ for the overall difference during follow-up). When the vocal biomarker was categorized according to the primary analysis to 4 equal quartiles, univariate Cox regression model showed 30%, 70%, and 270% increased risk of deaths in Q2, Q3, and Q4 compared with the lower quartile as reference ($P<0.05$ for Q2, and $P<0.001$ for Q3 and Q4; Table 2). The same model with the biomarker as a continuous variable consistently showed that each standard deviation increase of the biomarker was associated with a significant 48% elevated risk of death (95% CI, 1.39–1.58; $P<0.001$). Additional clinical and laboratory predictors of increased risk of death in the univariate model are summarized in Table 2. Compared with other strong predictors of 1-year mortality, the area under the receiver operating characteristic curve of the highest vocal biomarker was higher at 67% versus 54% and 61% for hypertension, peripheral artery disease, and prior hospitalization, respectively. To further assess how the biomarker improves the prediction of 1-year all-cause mortality, NRI approach was calculated based on a multivariate logistic regression model with and without the biomarker. Categorical NRI showed a 5% improvement with 0.5% improvement because of reclassification of high risk subjects and 4% improvement because of reclassification of low-risk subjects.

Association of the Vocal Biomarker With Hospitalization During Follow-Up

During the same follow-up period, 1403 (62%) patients were admitted to the hospital for at least 24 hours. Median time to first hospital admission was 6 (interquartile range 2–19) months. There were 313 (54%), 335 (58%), 357 (63%), and 398 (70%) hospitalizations in the four biomarker quartile groups. Kaplan–Meier survival analysis showed that patients in the higher biomarker group were more likely to

Table 2. Univariate Predictors of Mortality

Predictor	Events		HR	95% CI	Wald Chi	*P* Value
	(–) Group	(+) Group				
Q4 vs Q1	135 (23%)	306 (54%)	2.7	2.20–3.30	9.60	<0.001
Q3 vs Q1	135 (23%)	215 (38%)	1.7	1.40–2.15	5.00	<0.001
Q2 vs Q1	135 (23%)	168 (29%)	1.3	1.02–1.61	2.17	<0.05
Q4 vs Q1–Q3	518 (30%)	306 (54%)	2.04	1.77–2.35	9.88	<0.001
Age >77 y	337 (27%)	487 (46%)	1.05	1.04–1.05	12.34	<0.001
Male	312 (37%)	512 (35%)	0.9	0.80–1.06	–1.10	NS
Hebrew language (%)	273 (37%)	551 (35%)	0.85	0.75–1.01	–1.85	NS
CHF duration >4 y	415 (36%)	409 (36%)	1.00	1.00–1.00	2.38	<0.05
BP >90 mm Hg	452 (38%)	361 (33%)	0.99	0.98–0.99	–4.36	<0.001
Hypertension	63 (20%)	761 (38%)	2.12	1.64–2.74	5.71	<0.001
DM	326 (35%)	498 (6%)	1.01	0.88–1.16	0.18	NS
PVD	623 (33%)	201 (47%)	1.55	1.32–1.82	5.42	<0.001
CAD	245 (34%)	579 (37%)	1.15	0.99–1.34	1.85	NS
CKD	82 (21%)	742 (39%)	2.16	1.72–2.71	6.60	<0.001
Lung disease	583 (34%)	241 (42%)	1.34	1.16–1.56	3.84	<0.001
Active cancer	554 (33%)	270 (43%)	1.46	1.26–1.69	5.11	<0.001
Hospitalization†	223 (26%)	601 (42%)	1.88	1.61–2.19	8.02	<0.001
Obesity (BMI >30)	438 (40%)	286 (30%)	0.68	0.59–0.79	–5.02	<0.001
Fasting glucose	354 (36%)	321 (32%)	0.99	1.00–1.00	–0.80	NS
eGFR <60	67 (13%)	739 (42%)	2.69	2.09–3.46	7.70	<0.001
Anemia*	574 (43%)	234 (26%)	0.77	0.74–0.80	–12.17	<0.001
Triglycerides >122 mg/dL	469 (40%)	354 (31%)	1.00	1.00–1.00	–4.11	<0.001
TC	410 (36%)	413 (36%)	0.99	1:00–1:00	–1.44	NS
β-blockers	21 (38%)	803 (36%)	0.95	0.62–1.47	–0.02	NS
ACE inhibitors	15 (38%)	809 (36%)	0.95	0.6–1.66	–0.02	NS
Diuretic	12 (25%)	812 (36%)	1.7	0.94–2.9	1.75	0.08
Spironolactone	230 (36%)	594 (36%)	1.05	0.91–1.23	0.69	NS
Lipid-lowering drugs	50 (43%)	774 (35%)	0.84	0.63–1.12	–1.17	NS
NYHA ≥3	257 (28%)	181 (50%)	1.62	1.00–2.00	8.03	<0.001
EF <40	202 (35%)	188 (36%)	1.08	1.00–1.00	0.26	NS

ACE indicates angiotensin-converting enzyme; BMI, body mass index; BP, blood pressure; CAD, coronary artery disease; CHF, congestive heart failure; CKD, chronic kidney disease; DM, diabetes mellitus; EF, ejection fraction; eGFR, estimated glomerular filtration rate mL/min per 1.73 m^2; HR, hazard ratio; NS, not significant; NYHA, New York Heart Association Classification Guidelines; PVD, peripheral vascular disease; and TC, total cholesterol.

*Hemoglobin >13.5 for males, 12 for females.

†Any hospitalization in the previous year.

hospitalize (Figure 3). Univariate Cox regression model showed that compared with the lowest quartile, patients in Q2, Q3, and Q4 were 18%, 35%, and 69% more likely to be hospitalized during follow-up (*P*<0.05 for Q2; *P*<0.001 for Q3 and Q4). The same model with the biomarker as a continuous variable consistently showed that each SD increase of the biomarker was associated with a significant 25% elevated risk of hospitalization during follow-up (95% CI, 1.19–1.31; *P*<0.001). Binary logistic regression with adjustment for age and sex yielded consistent results (Table 3) such that compared with the lowest biomarker group (Q1), patients in the highest quartile (Q4) were more than 3 times more likely to hospitalize within 30 days (*P*<0.001). Compared with other strong predictors of 30-day hospitalization, the time-dependent area under the receiver operating characteristic curve of the highest vocal biomarker was high at 62% versus 56% and 52% for chronic kidney disease and peripheral vascular disease, respectively (Figure 4). Categorical NRI showed a 3% improvement with a negative –1% reclassification of

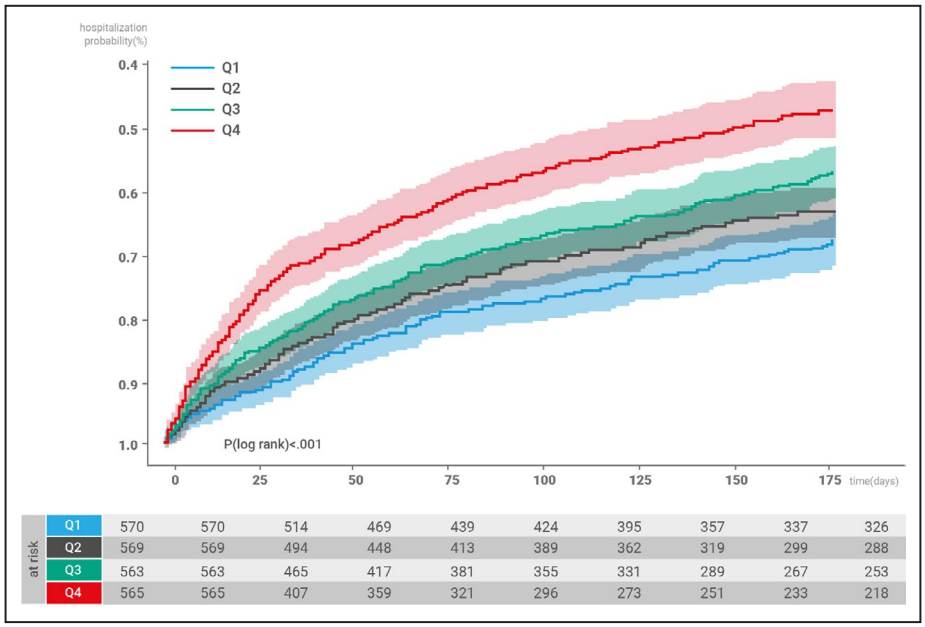

Figure 3. Kaplan–Meier hospitalization probability.
Kaplan–Meier curve showing overall cumulative hospitalization probability for the 4 biomarker quartile groups.

high-risk subjects and 4% reclassification of low-risk subjects.

Multivariate and Subgroup Analysis

Multivariate Cox proportional hazards regression modeling with adjustment to age, sex, and additional clinical predictors of death showed that patients in Q3 and Q4 were 37% and 96% more likely to die during follow-up (*P* NS for Q2, *P*<0.01 for Q3, *P*<0.001 for Q4; model 2, Table 4). Similar results were obtained for the secondary end point such that patients in Q3 and Q4 groups were 23% and 56% more likely to hospitalize during follow-up (*P*<0.01 for Q3; *P*<0.001 for Q4; model 2, Table 4). Subgroup analysis of 913 (40%) patients with available estimated glomerular filtration rate, baseline body mass index, mean blood pressure, hemoglobin, triglycerides, prior hospitalizations, New York Heart Association functional class, and echocardiographic ejection fraction yielded similar results such that compared with Q1, the highest vocal biomarker quartile was associated with a significant and independent 97% increased risk of death (*P*<0.001, model 3, Table 4) and with a significant 50% increased risk for hospitalizations during follow up (*P*<0.01; model 3, Table 4).

The consistency of the association of the biomarker with overall survival was assessed with the use of statistical tests of interaction between biomarker categories and multiple important predictors of death within the Cox model. The analysis showed that while the independent increased risk of death associated with higher vocal biomarker was significant in all subgroups, there was a significant interaction with respect to 2 subgroups: compared with patients with estimated glomerular filtration rate ≥60, patients with low estimated glomerular filtration rate showed attenuated association of the high vocal biomarker with increased risk of death (HR 1.59 versus 1.79 respectively; *P* for interaction 0.001). In addition, while the association of the vocal biomarker with increased risk of death was significant for both obese and nonobese patients (*P*<0.001 for both), there was a significant interaction such that the risk was significantly more pronounced in nonobese patients (HR 1.64 versus 1.58, *P* for interaction 0.01).

DISCUSSION

The main novel finding of the current study is that noninvasive voice signal characteristics are associated with adverse clinical outcome among patients with symptomatic heart failure. The association of the vocal biomarker with poor survival persisted after adjustment for relevant confounders and was consistent in each risk subset analyzed, suggesting an independent association. Thus, the current study supports the use of voice signal analysis as a noninvasive diagnostic biomarker for identifying high-risk CHF patients.

Voice signal analysis and voice recognition are being used extensively for commercial purposes: Companies such as Amazon, Google, Apple, and

Table 3. Binary Logistic Regression for Hospitalizations*

	Events				Continuous Model		Q2 vs Q1		Q3 vs Q1		Q4 vs Q1	
	Q1	Q2	Q3	Q4	OR†	95% CI	OR†	95% CI	OR†	95% CI	OR†	95% CI
30-d	60 10%)	83 (14%)	97 (17%)	151 (26%)	1.64	1.47–1.82	1.25	0.89–1.76	1.63	1.17–2.28	3.14	2.27–4.34
90-d	130 (22%)	163 (28%)	178 (31%)	234 (41%)	1.50	1.37–1.65	1.19	0.92–1.56	1.44	1.11–1.88	2.34	1.79–3.05
180-d	185 (32%)	212 (37%)	239 (42%)	294 (52%)	1.41	1.29–1.55	1.12	0.87–1.45	1.43	1.11–1.84	2.00	1.55–2.59
365-d	254 (44%)	271 (47%)	300 (53%)	338 (59%)	1.34	1.22–1.47	1.11	0.87–1.43	1.41	1.10–1.82	1.76	1.36–2.29

NS indicates not significant; and OR, odds ratio.
*P<0.001 for all comparisons except for Q2 vs Q1 (NS).
†The model is adjusted for age and sex.

Netflix use this technology in order to allow customers to talk to products directly with a wake word ("Alexa," "SIRI," "OK Google," etc) and receive voice responses and content instantly. The same technology is used for home thermostats and other smart home applications. However, robust data on the association of voice with disease states and clinical outcomes are lacking. Voice signal characteristics have been suggested to be associated with a number of different pathological entities including dyslexia, attention deficit hyperactive disorder, Parkinson's disease, and other neurological disorders.[18–20] However, only 2 studies investigated voice in the context of cardiovascular disease. We have recently studied subjects who underwent clinically indicated coronary angiogram and had their voice recorded to their personal smartphone devices using "Vocalis Health" application.[10] In that study we were able to identify 2 voice features that were associated with the presence of coronary artery disease. In another pilot study, voice of 10 patients with decompensated heart failure was analyzed during treatment. The study suggested a possible correlation between several voice measures and improvement in heart failure symptoms.[21] While the current study did not investigate a specific voice measure, it adds additional evidence that associates voice with the cardiovascular system. Thus, the current study extends these previous studies and supports a potential role for voice biomarker in heart disease.

This study does not offer a direct mechanism. However, with a sample size of more than 2200 patients and strong independent association using several different multivariate models, the novel association reported in this study might have clinical significance. Voice production is the results of 3 components: the lungs, the larynx, and the articulators (tongue, palate, and mouth muscles).[22] The vagus nerve, which is participating in voice production together with several other cranial nerves, is critical for autonomic control of the heart through its superior inferior and thoracic branches. It is also associated with heart rate control and variability and was suggested by Sugrue and colleagues as a possible participant in the biological basis underlying the finding of this study.

Numerous prognostic markers of death and/or hospitalization have been identified in patients with heart failure, but their clinical applicability is limited and risk stratification remains a challenge, particularly with respect to heart failure hospitalizations.[23] Notable risk predictors include age, renal function, blood pressure, blood sodium level, left ventricular ejection fraction, sex, brain natriuretic peptide level, New York Heart Association functional class, diabetes mellitus, body mass index, and exercise

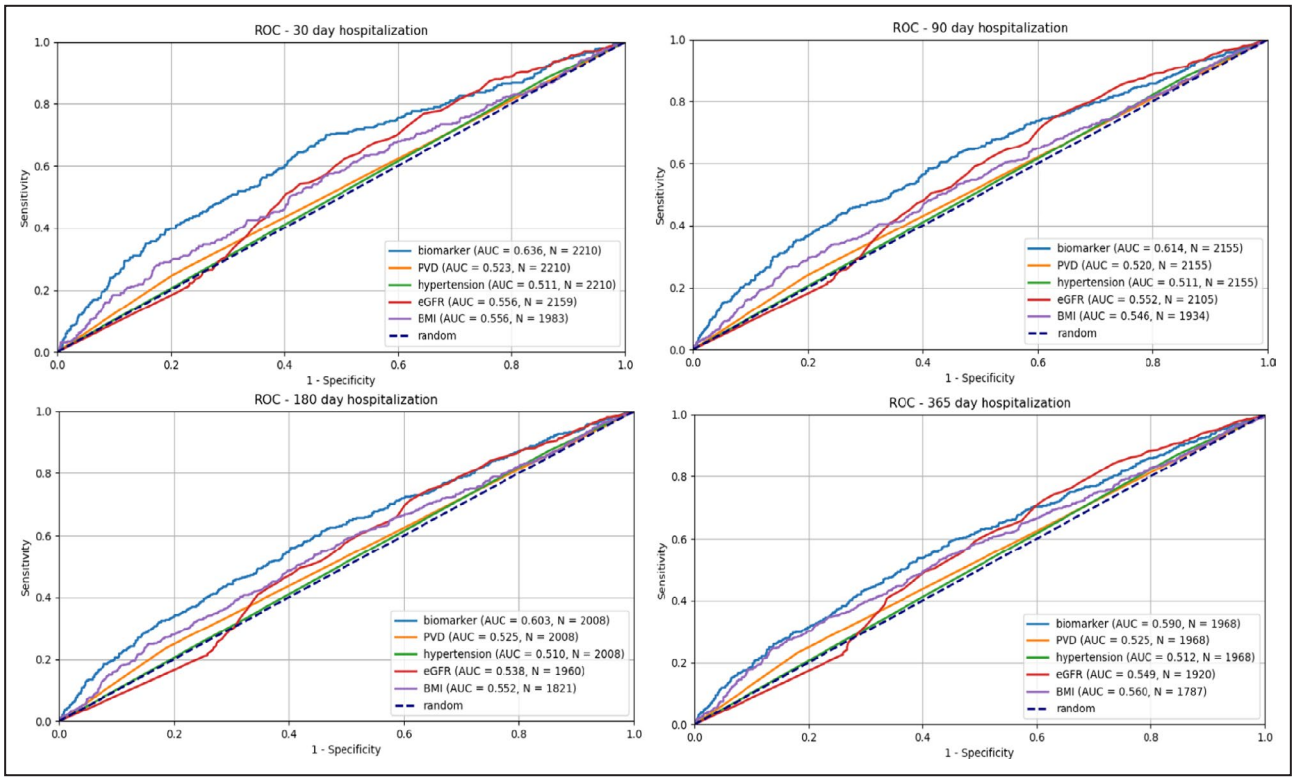

Figure 4. Time-dependent receiver operator curve for 30-, 90-, 180-, and 365-day hospitalization.
Receiver-operating curves for selected univariate predictors of 30-, 90-, 180-, and 365-day hospitalization along with the vocal biomarker. All receiver-operating curves and all calculated areas under the curve are for the outcome of hospitalization, and are calculated at 30 days (top left), 90 days (top right), 180 days (bottom left), and 365 days (bottom right). Please see text for categorical NRI analysis.

capacity.[24] A recent analysis showed how prediction of mortality and in particular heart failure hospitalizations remains only moderately successful, with contemporary prediction models having a mean C-statistic of 0.66±0.0005 for predicting mortality, which is similar to the C-statistic of 0.67 for the vocal biomarker in estimating 1-year mortality reported in this study.[25] Our study therefore supports the use of vocal biomarkers, which are noninvasive and can be incorporated to any smartphone or even landline phone, for risk assessment of heart failure patients in telemedicine settings, when clinical health care is provided at a distance.

There are multiple approaches for telemedicine, and each approach needs to be assessed on its individual merit. Examples include nurse-coordinated disease management programs, monitoring of body weight, blood pressure or ECG, and telephone-based interactive voice-response system.[3,22–24] The current analysis identifies a new opportunity for home telemedical monitoring and management that has not been explored before. The vocal biomarker is a noninvasive tool that, together with other advances in telecommunication technologies, can be used as adjunct to medical management of patients with heart failure. While data on the effectiveness and the exact role of telemedicine in heart failure are conflicting, the vocal biomarker holds the potential to assist in identification of high-risk subjects in areas where access to services such as physical examination, blood tests, and cardiovascular imaging are limited.

Limitations of the Analysis

This study has several limitations. First, our training cohort included non-CHF patients and therefore we could not use our training cohort for optimizing the biomarker cutoff. This is why data are presented in 4 equal quartiles, with no suggested cutoff for clinical use. Second, this is a retrospective analysis of selected population of Hebrew/Russian speaking patients with symptomatic heart failure. It is unclear whether our findings could be generalized to other populations. Third, laboratory data were available for 90% of the study population, New York Heart Association functional class was available for 56% and echocardiographic ejection fraction data were available for 48% of the study population. Data on other important predictors of poor outcome such as brain natriuretic peptide level and exercise capacity

Table 4. Multivariate Cox Regression Models

	Hospitalization				Mortality			
	HR	Wald	95% CI	*P* Value	HR	Wald	95% CI	*P* Value
Model 1								
Continues*	1.21	6.77	1.15–1.28	<0.001	1.34	8.60	1.25–1.43	<0.001
Q4 vs Q1	1.54	5.45	1.32–1.8	<0.001	1.98	6.34	1.6–2.44	<0.001
Q3 vs Q1	1.26	2.90	1.08–1.47	0.003	1.39	2.92	1.11–1.73	<0.01
Q2 vs Q1	1.1	1.35	0.95–1.3	NS	1.08	0.73	0.87–1.37	NS
Model 2								
Continues*	1.21	6.71	1.15–1.28	<0.001	1.32	8.30	1.24–1.41	<0.001
Q4 vs Q1	1.56	5.57	1.33–1.82	<0.001	1.96	6.24	1.59–2.42	<0.001
Q3 vs Q1	1.23	3.08	1.09–1.49	0.002	1.37	2.82	1.10–1.71	<0.01
Q2 vs Q1	1.14	1.69	0.98–1.34	0.09	1.10	0.83	0.88–1.39	NS
Model 3								
Continues*	1.18	3.34	1.07–1.30	<0.001	1.31	4.41	1.17–1.49	<0.001
Q4 vs Q1	1.5	2.76	1.12–1.86	<0.01	1.97	3.55	1.36–2.87	<0.001
Q3 vs Q1	1.23	1.57	0.95–1.58	NS	1.7	2.73	1.16–2.49	<0.01
Q2 vs Q1	1.01	0.06	0.78–1.31	NS	1.17	0.79	0.79–1.74	NS

HRs indicates hazard ratios; and NS, not significant.

*Vocal biomarker as a continuous variable. HRs for the continuous vocal biomarker are for each 1 unit increase in the SD. Wald—Wald χ^2 statistics; Model 1—adjusted for age (continuous) and sex; Model 2—adjusted for age (continuous), sex, lung disease, chronic kidney disease, active cancer, hypertension, peripheral artery disease, language, and congestive heart failure duration (continuous, days). Model 3—subgroup analysis of 913 (40%) patients with available laboratory data. Model 3 is further adjusted to the following continuous covariates: estimated glomerular filtration rate, mean blood pressure, triglycerides level, body mass index and hemoglobin, as well as the following categorical variables: New York Heart Association functional class (4-level), echocardiographic ejection fraction (EF <40), and prior hospitalizations in the previous year (Y/N).

were not available. Fourth, this analysis does not explain the observed association between voice and adverse outcome. Lastly, this is an observational study and therefore the consistency of our findings including the validation and incremental prognostic value of the vocal biomarker need to be addressed in future controlled studies. The data presented in this study do not support the use of vocal biomarker for a diagnostic "stand-alone" tool.

CONCLUSIONS AND CLINICAL IMPLICATIONS

This is the first study to document a relationship between a vocal biomarker and adverse outcome among CHF patients including mortality and future hospitalizations. Vocal signal analysis is a noninvasive biomarker that can assist healthcare providers with individual patient risk stratification. Together with other digital health tools such as virtual visits and home monitoring, it holds the potential to assist in providing quality care in rural, remote communities. There is a need for further studies to assess how the use of vocal biomarkers can be used to improve patient outcome.

ARTICLE INFORMATION

Received June 21, 2019; accepted January 14, 2020.

Affiliations

From the Chaim Sheba Medical Center, Tel Hashomer, Israel (E.M., I.M.); Vocalis Health, Tel Aviv, Israel (D.P., D.M., N.T., Y.L.); Department of Cardiovascular Disease, Mayo Clinic, Rochester, MN (A.L.); Kahn-Maccabi Institute of Research and Innovation, Tel Aviv, Israel (G.K., V.S.); Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel (E.M., I.M., G.K., V.S.).

Acknowledgments

The authors would like to thank Yonathan Sason, Liran Azaria, and Yoram Levanon for their contribution in this project.

Disclosures

Perry, Mevorach, Luz, and Taiblum are employed by the Company. Dr Maor serves as a consultant for Vocalis Health. The remaining authors have no disclosures to report.

Supplementary Material

Table S1

REFERENCES

1. Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. *Lancet*. 2011;378:731–739.
2. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, Phillips CO, Hodshon BV, Cooper LS, Krumholz HM. Telemonitoring in patients with heart failure. *N Engl J Med*. 2010;363:2301–2309.
3. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JG. Structured telephone support or non-invasive telemonitoring for patients with heart failure. *Cochrane Database Syst Rev*. 2015;10:CD007228. DOI: 10.1002/14651858.CD007228.pub3.
4. van Veldhuisen DJ, Braunschweig F, Conraads V, Ford I, Cowie MR, Jondeau G, Kautzner J, Aguilera RM, Lunati M, Yu CM, et al. Intrathoracic impedance monitoring, audible patient alerts, and outcome in patients with heart failure: clinical perspective. *Circulation*. 2011;124:1719–1726.
5. Olivari Z, Giacomelli S, Gubian L, Mancin S, Visentin E, Francesco VD, Iliceto S, Penzo M, Zanocco A, Marcon C, et al. The effectiveness of

remote monitoring of elderly patients after hospitalisation for heart failure: the renewing health European project. *Int J Cardiol*. 2018;257:137–142.
6. Hindricks G, Taborsky M, Glikson M, Heinrich U, Schumacher B, Katz A, Brachmann J, Lewalter T, Goette A, Block M, et al. Implant-based multiparameter telemonitoring of patients with heart failure (IN-TIME): a randomised controlled trial. *Lancet*. 2014;384:583–590.
7. Abraham WT, Adamson PB, Bourge RC, Aaron MF, Costanzo MR, Stevenson LW, Strickland W, Neelagaru S, Raval N, Krueger S, et al. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. *Lancet*. 2011;377:658–666.
8. Desai AS, Bhimaraj A, Bharmi R, Jermyn R, Bhatt K, Shavelle D, Redfield MM, Hull R, Pelzel J, Davis K, et al. Ambulatory hemodynamic monitoring reduces heart failure hospitalizations in "real-world" clinical practice. *J Am Coll Cardiol*. 2017;69:2357–2365.
9. Maisel A, Barnard D, Jaski B, Frivold G, Marais J, Azer M, Miyamoto MI, Lombardo D, Kelsay D, Borden K, et al. Primary results of the HABIT Trial (heart failure assessment with BNP in the home). *J Am Coll Cardiol*. 2013;61:1726–1735.
10. Maor E, Sara JD, Orbelo DM, Lerman LO, Levanon Y, Lerman A. Voice signal characteristics are independently associated with coronary artery disease. *Mayo Clin Proc*. 2018;93:840–847.
11. Irony A, Avi P, Segal A, Iris M, Nachman A. Managing chronic care & tele-monitoring model for enhancing integrated care—the MOMA suggestion. *Int J Integr Care*. 2017;17:A19.
12. Rabiner L, Schafer R. Theory and Applications of Digital Speech Processing. 1st ed. Upper Saddle River, NJ: Pearson; 2010.
13. Hamilton JD. Time Series Analysis. New Jersey, United States: Princeton University Press; 1994.
14. Liu H, Yu L. Toward integrating feature selection algorithms for classification and clustering. *IEEE Trans Knowl Data Eng*. 2005;17:491–502.
15. Alpaydin E. Introduction to Machine Learning. 2nd ed. Cambridge, MA: MIT Press; 2009.
16. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. 1137–1143. San Francisco CA: Morgan Kaufmann; 1995.
17. Shalev V, Chodick G, Goren I, Silber H, Kokia E, Heymann AD. The use of an automated patient registry to manage and monitor cardiovascular conditions and related outcomes in a large health organization. *Int J Cardiol*. 2011;152:345–349.
18. Bonneh YS, Levanon Y, Dean-Pardo O, Lossos L, Adini Y. Abnormal speech spectrum and increased pitch variability in young autistic children. *Front Hum Neurosci*. 2011;4:237.
19. Levanon Y, Lossos-Shifrin L. Method and system for diagnosing pathological phenomenon using a voice signal [Internet]. 2008. Available at: http://www.google.com/patents/US7398213. Accessed September 18, 2016.
20. Ur K, Holi MS. Automatic detection of neurological disordered voices using mel cepstral coefficients and neural networks. In: 2013 IEEE Point-of-Care Healthcare Technologies (PHT). 2013:76–79.
21. Murton OM, Hillman RE, Mehta DD, Semigran M, Daher M, Cunningham T, Verkouw K, Tabtabai S, Steiner J, Dec GW, et al. Acoustic speech analysis of patients with decompensated heart failure: a pilot study. *J Acoust Soc Am*. 2017;142:EL401–EL407.
22. Sugrue A, Mahowald J, Asirvatham SJ. Hey goglexiri, do I have coronary artery disease? *Mayo Clin Proc*. 2018;93:818–820.
23. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, Falk V, González-Juanatey JR, Harjola V-P, Jankowska EA, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J*. 2016;37:2129–2200.
24. Rahimi K, Bennett D, Conrad N, Williams TM, Basu J, Dwight J, Woodward M, Patel A, McMurray J, MacMahon S. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC Heart Fail*. 2014;2:440–446.
25. Ouwerkerk W, Voors AA, Zwinderman AH. Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. *JACC Heart Fail*. 2014;2:429–436.