ESC European Society of Cardiology

Europace (2022) **24**, 234–244
doi:10.1093/europace/euab170

Combining home monitoring temporal trends from implanted defibrillators and baseline patient risk profile to predict heart failure hospitalizations: results from the SELENE HF study

Antonio D'Onofrio[1]*, Francesco Solimene[2], Leonardo Calò[3], Valeria Calvi[4], Miguel Viscusi[5], Donato Melissano[6], Vitantonio Russo[7], Antonio Rapacciuolo [8], Andrea Campana[9], Fabrizio Caravati[10], Paolo Bonfanti[11], Gabriele Zanotto[12], Edoardo Gronda[13], Antonello Vado[14], Vittorio Calzolari[15], Giovanni Luca Botto[11], Massimo Zecchin[16], Luca Bontempi[17], Daniele Giacopelli [18], Alessio Gargaro [18], and Luigi Padeletti[19]

[1]Cardiology Department - Electrophysiology and Cardiac Pacing Unit A.O.R.N. V. Monaldi, Via L. Bianchi, Naples, Italy; [2]Electrophysiology, Montevergine Clinic, Viale S. Modestino 8, 83013 Mercogliano, Italy; [3]Cardiology Division, Policlinico Casilino, Via Casilina 1049, 00169 Rome, Italy; [4]Cardiology Department, Policlinico G. Rodolico, AOU Policlinico V. Emanuele, Via S. Sofia 78, 95125 Catania, Italy; [5]Cardiology Division, Sant'Anna and San Sebastiano Hospital, Via F. Palasciano, 81100 Caserta, Italy; [6]Cardiology Division, F. Ferrari Hospital, Viale F. Ferrari 1, 73042 Casarano (LE), Italy; [7]Cardiology Division, SS. Annunziata Hospital, Via F. Bruno 1, 74121 Taranto, Italy; [8]Cardiology Department of Advanced Biomedical Sciences, Corso Umberto I 40, 80138 Federico II University of Naples, Italy; [9]Cardiology Division, A.O.U. San Giovanni di Dio e Ruggi D'Aragona, Via San Leonardo 1, 84131 Salerno, Italy; [10]Division of Cardiology, ASST Settelaghi, Di Circolo Hospital, Via L. Borri 57, 21100 Varese, Italy; [11]Cardiology Division, Rho Civil Hospital, Corso Europa 250, 20017 Rho (MI), Italy; [12]Cardiology Division, Mater Salutis Hospital, Via C. Gianella 1, 37045 Legnago, Italy; [13]Department of Medicine and Medical Specialties, I.R.C.C.S. Foundation Ca' Granda, Via M. Fanti 6, 20122 Milano, Italy; [14]Cardiology Division, S. Croce e Carle Hospital, Via M. Coppino 26, 12100 Cuneo, Italy; [15]Cardiology Division, Santa Maria di Ca' Foncello Hospital, Piazzale dell'Ospedale 1, 31100 Treviso, Italy; [16]Cardiology Department, Cattinara University Hospital, Strada di Fiume 447, 34149 Trieste, Italy; [17]Cardiology Division, Spedali Civili, Piazzale Spedali Civili 1, 25123 Brescia, Italy; [18]BIOTRONIK Italia, Via delle Industrie 11, 20090 Vimodrone (MI), Italy; and [19]Cardiology Department, I.R.C.C.S. MultiMedica, Via Milanese 300, 20099 Sesto San Giovanni, Milano, Italy

Received 10 November 2020; editorial decision 15 June 2021; accepted after revision 18 June 2021; online publish-ahead-of-print 15 August 2021

Aims We developed and validated an algorithm for prediction of heart failure (HF) hospitalizations using remote monitoring (RM) data transmitted by implanted defibrillators.

Methods and results The SELENE HF study enrolled 918 patients (median age 69 years, 81% men, median ejection fraction 30%) with cardiac resynchronization therapy (44%), dual-chamber (38%), or single-chamber defibrillators with atrial diagnostics (18%). To develop a predictive algorithm, temporal trends of diurnal and nocturnal heart rates, ventricular extrasystoles, atrial tachyarrhythmia burden, heart rate variability, physical activity, and thoracic impedance obtained by daily automatic RM were combined with a baseline risk-stratifier (Seattle HF Model) into one index. The primary endpoint was the first post-implant adjudicated HF hospitalization. After a median follow-up of 22.5 months since enrolment, patients were randomly allocated to the algorithm derivation group ($n = 457$; 31 endpoints) or algorithm validation group ($n = 461$; 29 endpoints). In the derivation group, the index showed a C-statistics of 0.89 [95% confidence interval (CI): 0.83–0.95] with 2.73 odds ratio (CI 1.98–3.78) for first HF hospitalization per unitary increase of index value ($P < 0.001$). In the validation group, sensitivity of predicting primary endpoint was 65.5% (CI 45.7–82.1%), median alerting time 42 days (interquartile range 21–89), and false

* Corresponding author. Tel: +39 081 706 2287; fax: +39 081 7062558. *E-mail address*: donofrioant1@gmail.com

(or unexplained) alert rate 0.69 (CI 0.64–0.74) [or 0.63 (CI 0.58–0.68)] per patient-year. Without the baseline risk-stratifier, the sensitivity remained 65.5% and the false/unexplained alert rates increased by ≈10% to 0.76/0.71 per patient-year.

Conclusion With the developed algorithm, two-thirds of first post-implant HF hospitalizations could be predicted timely with only 0.7 false alerts per patient-year.

Keywords Heart failure • Remote monitoring • Implantable cardioverter-defibrillator • Cardiac resynchronization therapy • Hospitalization • Predictors

What's new?

- In 918 patients with implantable defibrillators, we developed and validated an algorithm to predict heart failure (HF) hospitalizations using automatic, daily remote monitoring without patient interaction.
- Seven temporal trends contributed to the algorithm: diurnal and nocturnal heart rates, ventricular extrasystoles, atrial tachyarrhythmia burden, heart rate variability, physical activity, and thoracic impedance.
- To individualize predictive power of the algorithm, we tested the combination of the temporal trends with a baseline risk-stratifier (Seattle HF Model).
- The sensitivity of the algorithm (including baseline risk-stratifier) in predicting first HF hospitalizations was 65.5% with a median alerting time of 42 days.
- False alert rate (0.69 per patient-year) and unexplained alert rate (0.63 per patient-year) were remarkably lower than in other published algorithms, which may increase actionability of alerts and reduce workload for the attending physicians.

Introduction

Despite treatment improvements over the last 30 years, acute heart failure (HF) is associated with poor prognosis and high rehospitalization rates. In Europe, 44% of hospitalized HF patients are readmitted within the subsequent 12 months.[1] Therapies for acute HF do not halt disease progression and each HF hospitalization confers deteriorating prognosis.[2] Early prevention of decompensated HF events is therefore a key strategy to improve patient outcomes.

Several baseline and cross-sectional risk scores developed to stratify the risk of death in HF patients[3–4] are less effective in predicting HF hospitalizations early enough to allow timely intervention.[5] Implantable cardioverter-defibrillators (ICDs) and cardiac resynchronization therapy defibrillators (CRT-Ds) offer several HF-related diagnostics that are of interest as potential longitudinal predictors of HF events.[6–8] Combined with daily remote monitoring (RM) and automatic alerts, these diagnostic data may have remarkable implications on HF management and costs. However, for routine adoption of device-based predictive algorithms, it is essential to minimize the rate of inappropriate alerts.

In this study, we developed and validated a predictive algorithm based on temporal trends obtained daily by remote ICD or CRT-D monitoring. The algorithm performance was tested for acute HF events involving hospitalization, outpatient intravenous intervention (IVI), or death.

Methods

Study design

SELENE HF (Selection of potential predictors of worsening heart failure) was an observational, multicentre, event-driven study designed to prospectively collect follow-up and RM data trends from a population of patients with ICDs and CRT-Ds, to document HF hospitalizations and deaths, and to correlate these events with RM data. As described in the study design paper, the intent was to identify the combination with the greatest sensitivity and specificity in predicting HF events.[9]

The study included patients with an ICD capable of atrial sensing or a CRT-D, left ventricular ejection fraction (LVEF) ≤35%, and a New York Heart Association (NYHA) class II or III before the implantation. Patients were excluded if they had permanent atrial fibrillation, acute HF, previous stroke, planned cardiac surgery, short-life expectancy (<6 months), or insufficient mobile phone service coverage at home. Patients underwent in-hospital follow-up examinations biannually until the targeted number of adjudicated primary endpoints was reached.

The study was approved by the ethics committees in the 34 Italian and Spanish investigational sites listed in the Appendix. The study was conducted in compliance with the Declaration of Helsinki and the ISO14155:2013 Good Clinical Practice for medical device investigations (ClinicalTrials.gov Identifier: NCT01836510). All patients provided written informed consent before enrolment. All study data were source document verified.

Remote monitoring and blinding

All devices were manufactured by BIOTRONIK SE & Co. KG (Berlin, Germany) and used the Home Monitoring technology characterized by daily automatic transmissions of device data over the GSM (Global System for Mobile Communication) network.[10] In normal conditions, these data are available on the Home Monitoring webpage for hospital staff. However, during this study, patients were registered only in a restricted area not accessible by study team members, which ensured that HF-related medical decisions were not based on RM temporal trends. Only a limited number of safety-related alerts, such as low battery level, out-of-range lead impedance, or ineffective ICD shock were enabled.[9]

Table 1 Algorithm components

Component	Description
RM variables	Updated every day based on automatic, daily RM data transmission
24 h HR	Mean ventricular rate during 24 h. In the predicting algorithm, the variable is analysed during the last 90 days to detect periods with monotone increases
Night HR	Lowest 10-min average value during resting period (from 1 a.m. to 5 a.m.). In the algorithm, the variable is analysed to detect periods of instability within the last 45 days
HRV	Daily standard deviation of 5-min average atrial–atrial intervals recorded every 5 min. The algorithm searches for periods of monotone decrease of the relative 8-day moving average during the last 90 days
24 h activity	Trend of patient physical activity over the last 25 days assessed by an in-built accelerometer sensor and expressed in percent of 24 h (decreasing activity is indicative)
AHRE burden	Daily burden of atrial fibrillation and high rate atrial episodes over the last 7 days expressed in percent of 24 h
PVC/day	Trend of the number of premature ventricular complexes per hour. The slope of the relative 4-day moving average is analysed during the last 45 days
Thoracic impedance	Corresponding to the changes in thoracic fluid levels. Impedance trend is calculated from daily averages of 24 subthreshold impedance measurements between RV lead and device case. The variable is analysed within the last 90 days to detect periods with monotone decrease in the relative 8-day moving average
SHFM	The Seattle Heart Failure Model[3] score at baseline, before device implantation
Predicting index	Linear combination of the variables after numerical processing

The rationale for the inclusion of selected RM variables in the predicting model is provided in the study design paper.[9] The initial plan was to include nine RM variables: the seven variables listed above and two variables that were eventually excluded from the model: percentage of CRT pacing (daily rate of resynchronized ventricular beats) and intracardiac impedance measured between RV and LV lead, because data were not available in all patients and will be the objective of subsequent subanalyses. The large number of included variables reflects the previous experience that heart failure events manifested large variability of involved RM variables and trend changes. During the study, the set of longitudinal RM variables was expanded with the cross-sectional SHFM variable collected at baseline.
AHRE, atrial high rate episodes; CRT, cardiac resynchronization therapy; HR, heart rate; HRV, heart rate variability; LV, left ventricular; PVC, premature ventricular contractions (ventricular extrasystoles); RM, remote monitoring; RV, right ventricular; SHFM, Seattle Heart Failure Model.

Seattle Heart Failure Model score

The predictive algorithm was designed to combine temporal trends of RM parameters with a baseline risk-stratifier in order to individualize and potentially improve predictive power. The Seattle HF Model (SHFM)[3] risk-stratifier was chosen for this purpose, including demographics, NYHA class, LVEF, ischaemic aetiology, systolic blood pressure, medical therapy, and laboratory data (haemoglobin, lymphocytes, uric acid, cholesterol, and serum sodium) assessed at baseline, before device implantation.

Study endpoints

Study endpoints were independently adjudicated by a three-member board blinded to RM data. The primary endpoint was the first post-implant hospitalization for worsening HF, defined as a non-elective hospital admission with an overnight stay, triggered by symptoms, signs, or objective evidence of worsening HF (LVEF, electrocardiogram, other instrumental evidence) and requiring administration of intravenous therapy for HF (diuretics, vasodilators, or inotropic agents). Secondary endpoint was a composite of any (first or subsequent) hospitalization, outpatient IVI, or death related to worsening HF.

Study endpoints were classified as usable if occurring ≥30 days post-implantation (run-in period for algorithm stabilization) and associated with an RM transmission rate of at least 55%. This low cut-off value in comparison to a ≈90% median RM transmission rate in clinical practice was selected as it still ensured sufficient input information for the algorithm while allowing index evaluation also in conditions of moderate RM compliance. The RM transmission rate was defined as the proportion of days with data transmission among up to 90 days preceding an endpoint event.

Data analysis and statistics

The SELENE HF study was initially designed to collect HF-related events for the development of a predicting algorithm without its validation. Study closure was planned after adjudication of 50 primary endpoints, with a minimum 3-month follow-up of the last enrolled patient.[9] Between the occurrence and adjudication of the 50th potential primary endpoint, further HF-related hospitalizations occurred which underwent adjudication and were eventually included in the analysis. A higher than expected number of collected and adjudicated events allowed an expansion of the study to include algorithm validation.

Accordingly, before data analysis, patients were randomly allocated to a derivation and a validation group, stratified by usable primary endpoint events and device type. The aim was to obtain two independent cohorts of ∼460 patients with 30 primary endpoints and a balanced distribution of ICD and CRT-D devices. It was estimated that 30 primary endpoints in the validation cohort would allow testing an assumed sensitivity of 70% [95% confidence interval (CI): 50–85%, having >90% power to reject the null hypothesis of sensitivity ≤40% at one-sided binomial test ($\alpha = 0.025$)], along with a false alert rate of 1.0 per patient-year (CI 0.95–1.05, χ^2 distribution).

The predicting algorithm was developed with the primary endpoint events in the derivation cohort. A two-stage development process encompassed a cross-sectional analysis and a longitudinal analysis. In the cross-sectional analysis, the predicting index was derived from the Home Monitoring temporal trends of seven variables reported in *Table 1*. The trends were analysed up to 90 days before an event and compared with equivalent intervals randomly selected in patients without events. A number of numerical transformations for temporal trends was pre-defined. Each numerical transformation was tested with univariate logistic models. The transformation maximizing the associated C-statistics was found for

each variable. The best pull of processed variables was identified then by automatic numerical stepwise procedures and linearly combined with the baseline SHFM score to obtain the coefficients for index calculation. Next, projected sensitivity and specificity was evaluated by a receiver operating characteristic curve to identify optimal threshold value(s) for alerts.

In the longitudinal analysis, the predicting index was calculated daily in each patient of the derivation group. When it exceeded a nominal threshold (e.g. 4.5) for a number of consecutive days, an alert was triggered for a possible endpoint event. After an alert, the nominal threshold was replaced by a lower, recovery threshold (e.g. 3.0; the difference is called 'offset' of the recovery threshold, in this case −1.5) to filter out casual fluctuations. An alert was classified as true-positive if the index did not fall below the recovery threshold between alert and event. When the index falls below the recovery threshold, the alert state is cancelled, and nominal threshold is re-established as a criterion to trigger new alert. Both the number of consecutive days to trigger an alert and the offset of the recovery threshold were optimized to provide the best trade-off between sensitivity and false alert rate.

The predicting index and its nominal and recovery thresholds were finally validated in the validation cohort. Sensitivity, alerting time, specificity, and the rates of false alerts and unexplained alerts per patient-year were calculated for primary and secondary study endpoints. Sensitivity was defined as the percentage of endpoint events preceded by a true-positive alert. Alerting time is the number of days between a true-positive alert and the related event. Specificity was defined as the percentage of days with index values appropriately below nominal threshold (periods with adverse events occurring at the recovery threshold were excluded). False alert rate was the number of false-positive alerts (not followed by primary or secondary study endpoint) per patient-year. Some false-positive alerts were followed by adverse events that were mainly HF-related but not fulfilling the definition of study endpoint. False-positive alerts not followed by adverse events were unexplained alerts. Patients who were lost to follow-up or died contributed to the analyses up to the date of last available information.

Continuous variables are reported as median (interquartile range); binary or categorical variables as counts and percentages of non-missing data. Study groups were compared by the Mann–Whitney *U* test for continuous variables and Pearson's χ^2 or Fisher's exact test, as appropriate, for binary variables. All tests were significant with $P \leq 0.05$. Statistical tests were performed with the STATA 11 SE software (StataCorp LP, TX, USA).

Results

From May 2012 to March 2017, the study recruited 918 patients with dual-chamber ICDs (38%), single-chamber ICDs with atrial diagnostics (18%), or CRT-Ds (44%). Adjudication of endpoint events was completed in November 2018. Study population was thereafter randomly allocated to the derivation ($n = 457$) or validation ($n = 461$) group. No significant difference in main baseline variables was observed between study cohorts (*Table 2*). The patient flow is shown in *Figure 1*.

In the derivation cohort, 127 HF-related adverse events were reported (0.14 events per patient-year), including 75 events adjudicated as HF hospitalizations (9 terminals), 14 deaths with HF as primary cause, 4 outpatient visits with IVI, and 43 events that did not require hospitalization or IVI. Of the 75 HF hospitalizations, 34 were first post-implant hospital admissions, 31 of which met usability criteria to serve as primary endpoints for algorithm development.

In the validation cohort, of the 101 reported HF-related adverse events, 64 were HF hospitalizations (5 terminals), 9 were deaths for worsening HF, 6 were IVIs, and 27 did not require hospitalization or IVI. The composite of any (first or subsequent) worsening HF hospitalization, death, or IVI occurred in 73 cases (secondary endpoint), 11 of which did not fulfil usability criteria for the reasons listed in *Figure 1*. Eventually, 29 primary and 62 secondary usable endpoint events were available for algorithm validation tests.

The median remote monitoring rate was 91.3% of days (interquartile range, 83.5–95.8%) in the derivation cohort and 90.8% (83.1–95.5%) in the validation cohort. In 39 of 918 patients (4.2%) connection for Home Monitoring remote transmissions could not be established due to insufficient GSM coverage.

Results of the derivation analysis

In the derivation cohort, a unitary increase of the index value was associated with an odds ratio of 2.73 (CI 1.98–3.78; $P < 0.001$) for the first post-implant worsening HF-hospitalization. The area under the Receiver Operating characteristic curve was 0.89 (CI 0.83–0.95). Potential index nominal thresholds associated with highest prediction accuracy ranged from 3.5 to 4.5, yielding a projected sensitivity of 81.5% (CI 61.9–93.7%) to 63.0% (CI 42.4–80.6%) and a projected specificity of 82.6% (CI 78.2–86.5%) to 90.7% (CI 89.0–94.9%), respectively.

Left panel of *Figure 2* shows modified receiver operating characteristic curve of the algorithm sensitivity vs. false and unexplained alert rates, obtained during the longitudinal analysis in the derivation cohort. The curve was computed using 3 consecutive days with index above nominal threshold to raise an alert and an offset of −1.0 for the recovery threshold.

Validation tests

The performance of the predicting algorithm in the validation cohort is illustrated in the right panel of *Figure 2*. Detailed results for the nominal thresholds of interest are presented in *Table 3*. With the 4.5 nominal threshold, 65.5% of usable primary endpoint events could be predicted (CI 45.7–82.1%). Median alerting time was 42 days, false alert rate 0.69 alerts per patient-year, and unexplained alert rate 0.63 per patient-year. With a nominal threshold of 3.5, the sensitivity increased to 72.4%, albeit with higher false (1.07), and unexplained (0.99) alert rates per patient-year.

For the combined secondary endpoint of any HF hospitalization, outpatient IVI, or death related to worsening HF, at nominal thresholds 4.5–3.5, the sensitivity was 54.8–64.5%, median alerting time 43–60 days, false alert rate 0.67–1.05 and unexplained alert rate 0.63–0.98 alerts per patient-year.

Positive and negative predictive values (PPV and NPV) were calculated post-hoc for the secondary endpoint events. The PPV was defined as the fraction of appropriate recovery threshold periods (i.e. including one or more secondary endpoint events) over the total recovery threshold time. The NPV was defined as the fraction of appropriate nominal threshold periods (i.e. not including any secondary endpoint events) over the total nominal threshold time. For nominal thresholds of 3.5–4.5, PPV ranged from 5.3% to 7.7% and NPV ranged from 96.6% to 96.7%. No reasonable calculation of PPV and NPV is

Table 2 Patient population by derivation and validation cohorts

Variables	All	Derivation cohort	Validation cohort	*P*-value
Number of patients	918	457	461	
Follow-up (months)	22.5 (14.1–35.8)	21.9 (13.8–33.6)	23.4 (14.6–37.1)	
Age (years)	69.1 (60.7–75.9)	68.8 (60.7–75.7)	69.3 (60.8–76.1)	0.61
Gender (male)	744 (81.0%)	366 (80.1%)	378 (82.0%)	0.46
Body mass index (kg/m^2)	26.7 (24.2–29.4)	27.0 (24.5–29.4)	26.5 (24.2–29.4)	0.33
CRT-D devices	403 (43.9%)	202 (44.2%)	201 (43.6%)	0.85
QRS duration (ms)	120 (102–150)	121 (103–150)	120 (102–150)	0.69
LVEF (%)	30 (25–34)	30 (25–34)	30 (25–35)	0.25
Systolic blood pressure (mmHg)	120 (110–130)	120 (110–130)	120 (110–130)	0.13
NYHA Class II/III	446 (48.8%)/467 (51.2%)	225 (49.4%)/230 (50.6%)	221 (48.2%)/237 (51.8%)	0.72
SHFM-predicted 1-year mortality (%)	3.8 (2.3–6.6)	3.6 (2.2–3.6)	4.0 (2.4–6.6)	0.18
Primary aetiology				
Ischaemic cardiomyopathy	413 (45.0%)	206 (45.1%)	207 (44.9%)	0.95
Dilated cardiomiopathy	365 (39.8%)	185 (40.5%)	180 (39.1%)	0.66
Comorbidities				
History of hypertension	604 (65.8%)	295 (64.6%)	309 (67.0%)	0.43
Diabetes	323 (35.4%)	153 (33.6%)	170 (37.2%)	0.26
Chronic kidney disease	194 (21.1%)	107 (23.4%)	87 (18.9%)	0.09
Atrial fibrillation history	129 (14.1%)	68 (15.0%)	61 (13.3%)	0.46
Stroke/TIA	69 (7.5%)	33 (7.2%)	36 (7.8%)	0.73
Valvular surgery	68 (7.4%)	37 (8.1%)	31 (6.7%)	0.45
Blood, urine tests				
Sodium (mg/dL)	140 (138–142)	140 (138–142)	140 (138–142)	0.38
Blood urea nitrogen (mg/dL)	35.0 (22.4–52.0)	36.9 (23.0–52.0)	34.0 (22.4–50.5)	0.51
Haemoglobin (g/dL)	13.4 (12.2–14.6)	13.5 (12.3–14.7)	13.3 (12.1–14.5)	0.06
Lymphocytes (%)	25.5 (19.8–31.8)	25.6 (19.8–31.8)	25.3 (19.8–31.9)	0.98
Serum uric acid (mg/dL)	6.1 (4.8–7.6)	6.0 (4.8–7.7)	6.2 (4.8–7.5)	0.81
Cholesterol (mg/dL)	153 (127–188)	155 (129–187)	152 (125–190)	0.71
Baseline therapy				
Diuretics	797 (86.8%)	400 (87.5%)	397 (86.1%)	0.55
Beta-blockers	793 (86.4%)	395 (86.4%)	398 (86.3%)	0.96
ACE inhibitors	523 (57.0%)	259 (56.7%)	264 (57.3%)	0.86
Aldosterone antagonists	240 (26.1%)	133 (29.1%)	107 (23.2%)	0.04
Angiotensin receptor blockers	196 (21.3%)	100 (21.9%)	96 (20.8%)	0.70
Calcium-channel blockers	75 (8.2%)	36 (7.9%)	39 (8.5%)	0.75
Statins	553 (60.2%)	286 (62.6%)	267 (57.9%)	0.15
Antiplatelets	596 (64.9%)	298 (65.2%)	298 (64.6%)	0.86
Anticoagulants	228 (24.8%)	109 (23.9%)	119 (25.8%)	0.49
Amiodarone	169 (18.4%)	81 (17.7%)	88 (19.1%)	0.59

Data are shown as median (interquartile range) or as number (% of non-missing data).
ACE, angiotensin-converting enzyme; CRT-D, cardiac resynchronization therapy defibrillator; LVEF, left ventricle ejection fraction; NYHA, New York Heart Association; SHFM, Seattle Heart Failure Model; TIA, transient ischaemic attack.

possible for the primary endpoint because its occurrence as 'first post-implant' event automatically terminates follow-up regarding this endpoint.

Figure 3A illustrates the mean temporal trends of the index averaged over the group of patients with ($n = 60$) and without ($n = 858$) primary endpoint events. A progressive increase of the index average is visible weeks before the event. The relative contribution of each component to total index value varied during time, depending on the evolution of the relative parameter trend. An analysis at 1 week before events showed an average 20–25% contribution from monotone increase of 24-h heart rate and instability of nocturnal heart rate, with the remaining components contributing 13% or less (*Figure 3B*). The average contribution of the atrial high rate episode burden component was 4% as only 11 of the 60 primary endpoint events presented with ongoing atrial fibrillation (in these 11 cases the average contribution was 19%).

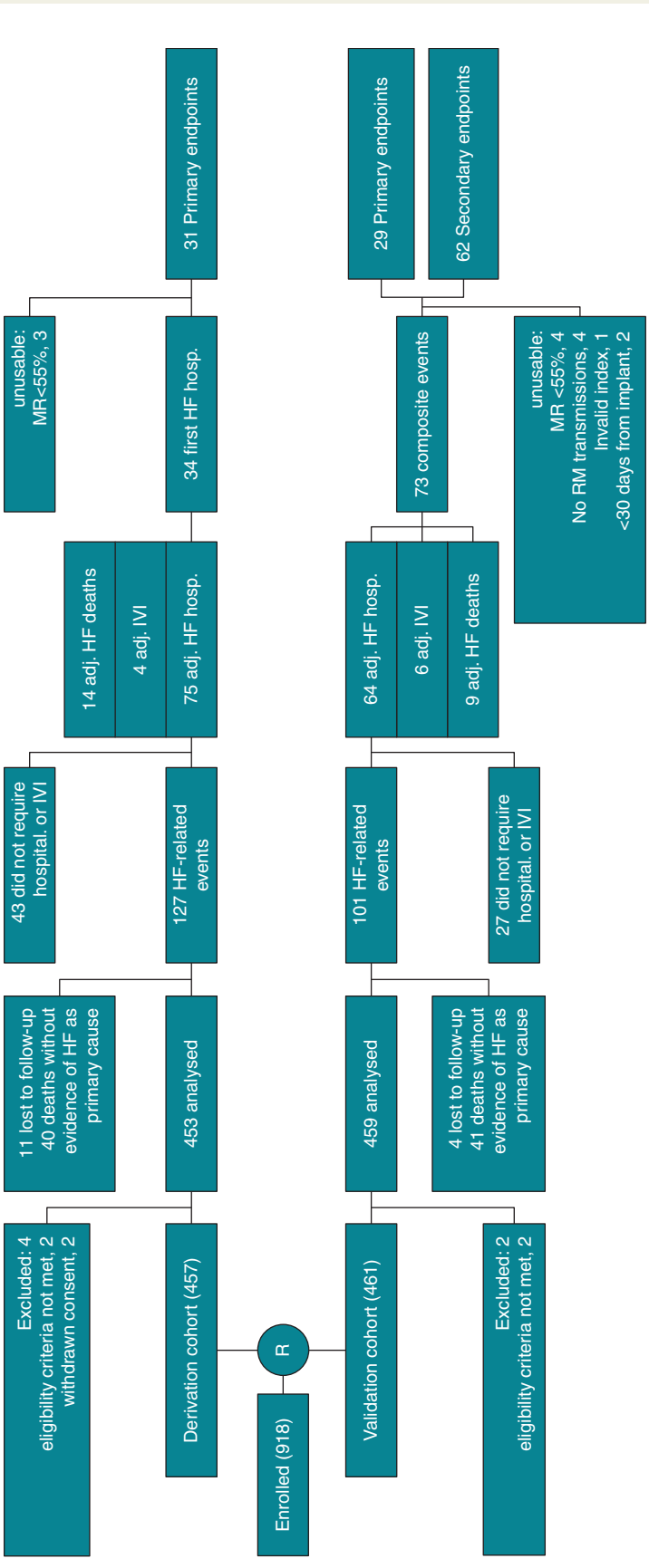

Figure 1 Consort diagram. Patients who were lost to follow-up or died contributed to the analyses up to the date of last available information. Adj., adjudicated as; HF, heart failure; hosp. or hospital., hospitalization; IVI, outpatient intravenous intervention; MR, minimum monitoring rate expressed as percentage of days with remote monitoring coverage in the last 90 days; R, random allocation stratified by primary endpoint and device type; RM, remote monitoring.

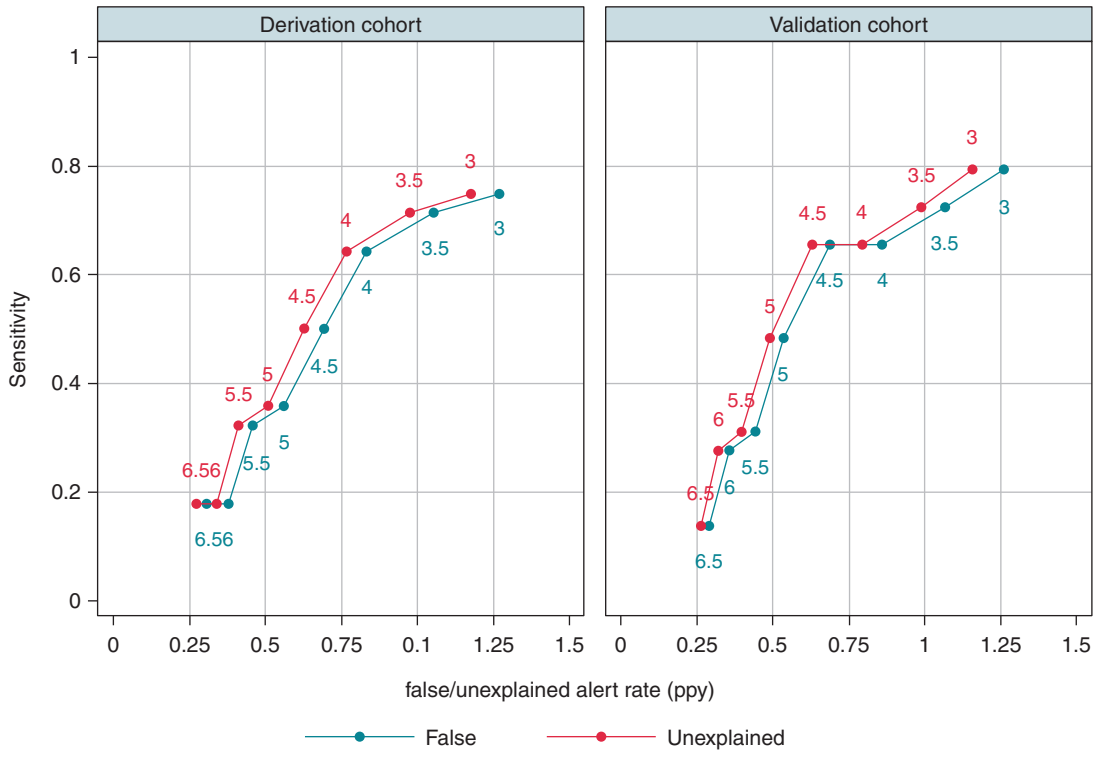

Figure 2 The modified receiver operating characteristic curve of algorithm sensitivity to primary endpoints vs. false and unexplained alert rates per patient-year (ppy). The curves were computed for the derivation cohort (left panel) and the validation cohort (right panel), using 3 consecutive days with index above nominal threshold to raise an alert, and an offset of −1.0 for the recovery threshold.

Table 3 **Validation of the predictive algorithm**

Endpoint	Nominal threshold	Predicted events/usablea events	Sensitivity (%)	Alerting time (days)	Specificity (%)	False alert rate (ppy)	Unexplained alert rate (ppy)
First post-implant HF hospitalization	3.5	21/29	72.4 (52.8–87.3)	61 (43–75)	75.8 (75.6–75.9)	1.07 (1.00–1.13)	0.99 (0.93–1.05)
	4.0	19/29	65.5 (45.7–82.1)	58 (22–87)	82.4 (82.3–82.5)	0.86 (0.80–0.92)	0.79 (0.74–0.85)
	4.5	19/29	65.5 (45.7–82.1)	42 (21–89)	86.7 (86.6–86.8)	0.69 (0.64–0.74)	0.63 (0.58–0.68)
Any HF hospitalization, outpatient IVI, or death related to worsening HF	3.5	40/62	64.5 (51.3–76.2)	60 (30–92)	75.3 (75.2–75.4)	1.05 (0.99–1.12)	0.98 (0.92–1.05)
	4.0	37/62	59.7 (46.4–71.9)	54 (24–92)	82.0 (81.9–82.2)	0.85 (0.79–0.91)	0.79 (0.73–0.85)
	4.5	34/62	54.8 (41.7–67.5)	43 (17–85)	86.5 (86.4–86.6)	0.67 (0.62–0.73)	0.63 (0.58–0.68)

Results are reported for different nominal thresholds of predicting index. Sensitivity, specificity, false alert rate, and unexplained alert rate are provided with the relative 95% confidence interval. Alerting time is reported as median (interquartile range).
[a]Endpoint events were usable if they occurred after a run-in period of 30 days and were associated with a minimum remote monitoring transmission rate of 55%.
HF, heart failure; IVI, intravenous intervention; ppy, per patient-year.

Results without SHFM score

We have validated the performance of the predictive algorithm also without the SHFM score. For the primary endpoint, at the nominal threshold of 4.5, the sensitivity remained 65.5%, whereas median alerting time increased from 42 to 62 days, false alert rate from 0.69 to 0.76/patient-year, and unexplained alert rate from 0.63 to 0.70/patient-year. At the nominal threshold of 3.5, the sensitivity and median alerting time remained 72.4% and 61 days, while the false and unexplained alert rates increased by ≈10% to 1.18/patient-year and 1.09/patient-year, respectively. The corresponding values for the secondary endpoint were 54.8%, 51 days, 0.75 and 0.70 per patient-year (4.5 threshold), and 62.9%, 61 days, 1.16 and 1.08 per patient-year,

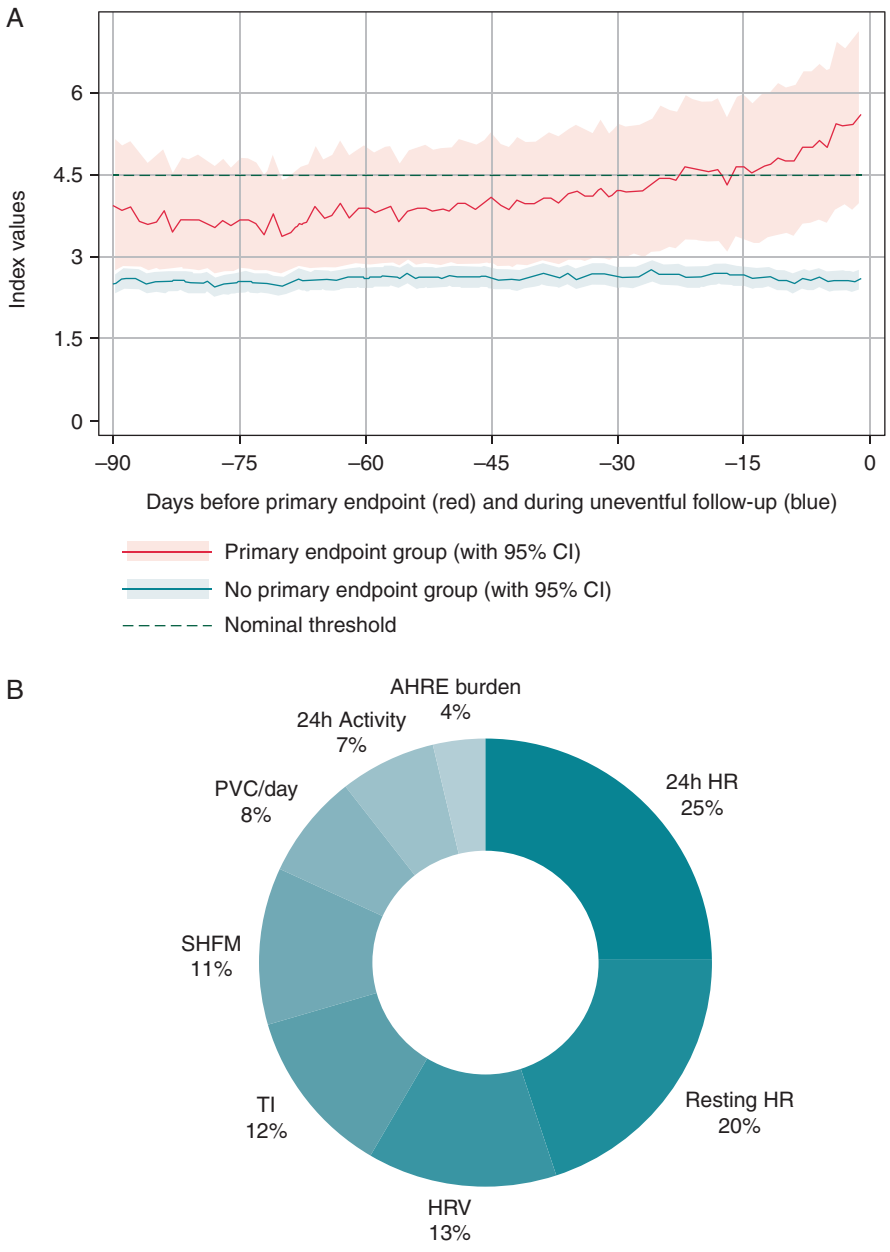

Figure 3 (*A*) Temporal trends of the predicting index. The daily average values of the predicting index are plotted in patients with primary endpoint events ($n = 60$, red line) vs. patients without primary endpoint events ($n = 858$, blue line). Data are aligned relative to the date of heart failure hospitalization (primary endpoint group) and up to 60 days before the end of follow-up (no primary endpoint group). Owing to the Seattle Heart Failure Model baseline component, index values are constantly numerically higher in the primary endpoint group. But statistical significance is not reached and baseline stratification is not sufficient for a reliable prediction unless the index value crosses certain nominal threshold. The apparent alerting time of about 20 days (shorter than the 42 days found in the validation analysis) is the result of averaging index values over all detected and undetected 60 events. (*B*) The relative contribution of all seven components to the index value, averaged for the last 7 days before 60 primary endpoint events. AHRE, atrial high rate episodes; HR, heart rate; HRV, heart rate variability; PVC, premature ventricular contractions; SHFM, Seattle Heart Failure Model; TI, thoracic impedance.

respectively (3.5 threshold). This again corresponds to a ≈ 10% increase in the false and unexplained alert rates compared to the algorithm with SHFM and nearly the same sensitivity, with an alerting time increase only for the 4.5 threshold (*Table 3*).

Discussion

We developed an algorithm for acute HF prediction in ICD and CRT-D patients based on the baseline SHFM score and seven RM temporal trends. Using the index nominal threshold of 4.5, the

algorithm predicted 65.5% of first post-implant HF hospitalizations with median alerting time of 42 days and one false alert every 17 months. All alerts generated by the algorithm were used for false alert rate calculation, including those preceding HF-related events which were less severe than required by endpoint definitions (possibly due to pre-emptive therapy unrelated to our study). When these cases were excluded and calculations were limited to totally unexplained alerts, there was one false alert every 19 months.

The specificity of 76–87% (depending on nominal threshold) shows that the index crossed the nominal threshold on up to one-quarter of days outside the alerting times. Therefore, a consistency criterion to raise an alert was needed. During algorithm development, we found that three consecutive days above nominal threshold is an optimal consistency criterion. Furthermore, a recovery threshold lower than the nominal threshold (the difference of −1.0) is introduced to avoid that a raised alert is inappropriately cancelled before an event due to small daily variations of the index value.

For the composite secondary endpoint combing HF-related deaths, hospitalizations, and IVIs, the algorithm showed similar performances as for the first post-implant HF hospitalization. However, the latter (primary endpoint) is of greater clinical relevance since current therapies for acute HF cannot stem disease progression.[11] Hence, early diagnosis is strategic in counteracting the disease, as confirmed by the latest recommendations on HF management.[12] Onset of HF symptoms is only a last stage of a decompensation process starting several weeks in advance.[13] Our algorithm-generated alerts at a median of 42–61 days before events, allowing sufficient time for patient contact, investigations, and preventive measures to reduce hospitalizations. *Figure 4* illustrates a case example collected during the study.

Six to eight weeks of alerting time are in line with previous observations: Zile *et al*.[14] reported a significant increase in diastolic pressure that can be observed on a 60-day time scale in systolic HF patients; in the MultiSENSE study (Multisensor chronic evaluation in ambulatory heart failure patients), 25% of all events could be detected earlier than 66 days before HF events.[8] Our results add evidence that the decompensation process starts several weeks and even months before HF hospitalizations. The combination of daily index updates with a 6–8 week prediction time appears as a fair compromise to mitigate the so-called 'test to event timing paradox' for which prolonged test-to-event timing as well as higher test frequency may increase prediction uncertainty especially in dynamic conditions.[15]

Several algorithms have been proposed to combine multiple sensors in an attempt to maximize prediction capacity. In the PARTNERS HF[6] study (Programme to access and review trending information and evaluate correlation to symptoms in patients with heart failure), an algorithm based on rolling evaluations of pairs of consecutive 30-day intervals was developed combining multiple diagnostic variables. The method was subsequently refined to identify low-, medium-, and high-risk levels for monthly evaluations. Sensitivity/specificity associated with different risk levels ranged from 83%/46% (low risk) to 46%/90% (high risk). However, 44% of evaluations were classified in the medium-risk category in which uncertainty about patient condition and medical reaction may persist.[7]

In the MultiSENSE study, the dynamic HeartLogic algorithm predicted 70% of impending HF-hospitalizations and IVIs with a median detection time of 34 days and 1.47 unexplained alerts per patient-year, as compared to 0.63 unexplained alerts per patient-year in our study. The HeartLogic is a multisensor algorithm combining first and third heart sounds and respiration rate, among other sensor data. In contrast, our predicting algorithm was based on temporal trends of parameters that are part of ordinary diagnostic equipment of dual-chamber defibrillators and cardiac resynchronization therapy pacemakers and defibrillators. In this regard, our algorithm has larger generalizability, showing that a comparable prediction accuracy may be achieved by appropriately utilizing information already available in most devices. Additionally, daily transmissions characterizing the RM system used in the study would allow a centralized implementation without the need of firmware modification or upgrade of implanted systems. This is of potential advantage for immediate applicability. Finally, unlike other multivariable predictors, our algorithm has been developed and validated in patient cohorts with both ICD and CRT-D devices in similar proportions, which can contribute further to a more general application.

It may be important to include information relative to individual patient risk stratification in multisensor algorithms. It has been recently reported that using the HeartLogic algorithm in combination with NT-proBNP (N-terminal pro-B-type natriuretic peptide) levels increased the risk-stratification ability dramatically,[16] but it would imply serial assessment of biomarker levels, which can raise some practical concerns.[17] We incorporated the baseline SHFM score in our model as a stationary risk-stratifier based on the demographic, aetiology, therapy, blood, and urine data that are commonly available at device implantation. Serving as an additive constant to tailor nominal thresholds to patient's individual risk profile, the SHFM reduced false and unexplained alert rates by ≈10%, with negligible impact on sensitivity. Knowing that false alerts are the main source of inefficient resource consumption, the lower false alert rate and reduced alerting time with SHFM are expected to increase actionability of alerts.

The anticipated workload for the attending physicians related to the predicting algorithm is low because implantable devices routinely send data required for automatic generation of alerts, and the false positive alert rate is minimized by algorithm design. Algorithm alerts incorporated into automatic daily RM system may therefore enhance efficiencies in contrasting disease progression by allowing early awareness of deteriorating HF, increasing therapy compliance, reducing hospitalizations, and ultimately improving outcomes. Early interventions post-alert in patients without symptoms may include reinforcing the need for diet and fluid restrictions and evaluating the use of current medication. If congestion is reported, an increase in diuretics may be required, with follow-up 1–2 weeks later. The ability to benefit is likely to differ according to the overall disease burden of the patient; those with comorbidities such as kidney/lung disease or diabetes are more susceptible to decompensation, more difficult to treat and more prone to recurrent and costly events. These hypotheses should be tested in subsequent controlled clinical investigations.

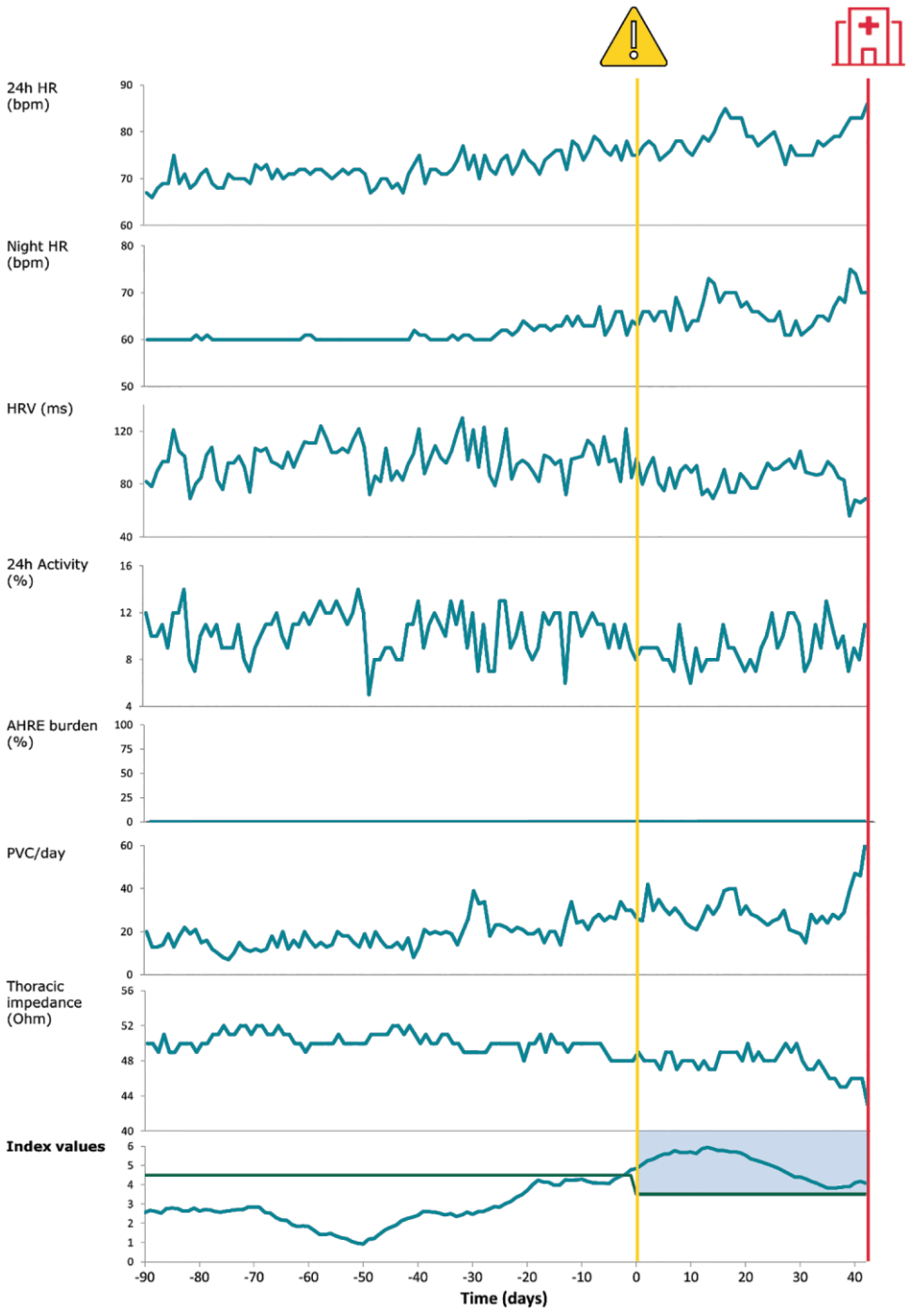

Figure 4 Trends of Home Monitoring variables and the predicting index in an 82-year-old man (implanted with a CRT-D, baseline SHFM = 0.23) before a worsening HF hospitalization (11-day hospital stay, treatment with loop and potassium-sparing diuretics). With the algorithm, an alert would have been raised 42 days before hospital admission, mainly driven by increasing 24-h HR and ventricular extrasystoles, instability of nocturnal HR, decreasing daily activity and thoracic impedance, visible 4–5 weeks before the alert and 8–9 weeks before the admission. The alert would have allowed proactive care and possibly prevent the exacerbation of HF and consequent hospitalization. Yellow line: day of alert, time = 0 (nominal threshold 4.5, recovery threshold offset −1.0); red line: day of HF hospitalization; light blue square: alerting state of index. AHRE, atrial high rate episodes; CRT-D, cardiac resynchronization therapy defibrillator; HR, heart rate; HRV, heart rate variability; PVC, premature ventricular contractions; SHFM, Seattle Heart Failure Model.

Limitations

Despite the low false and unexplained alert rates for our algorithm, the estimate of PPV for secondary endpoints is still below 8%, in line with alternative algorithms.[7–8] Owing to the complexity of the disease, patients may experience varying degrees of worsening HF,[18–20] while we analysed only the subset of adjudicated and usable events leading to IVI, hospitalization, or death. Therefore, we cannot exclude that some algorithm alerts classified as 'false' were actually related to decompensating conditions which did not ultimately lead to a study endpoint.

The predictor we describe was developed and tested in a population with an indication for a dual-chamber ICD or a CRT-D implantation. No data are currently available for the predictor performance in patients with permanent atrial fibrillation or in HF patients without cardiac implantable electronic devices.

Conclusions

A longitudinal worsening HF predictor combining seven RM temporal trends with the SHFM as baseline risk-stratifier was developed and validated. The predicting algorithm showed promising sensitivity and a remarkably low false alert rate. First post-implant HF hospitalizations, IVIs, subsequent and terminal HF hospitalizations could be predicted with similar accuracy. Randomized trials are needed to assess whether the application of the algorithm may be associated with improved outcomes.

Supplementary material

Supplementary material is available at *Europace* online.

Acknowledgements

The authors would like to sincerely thank Jochen Proff for his significant contribution to study design, enduring support to project realization, and interpretation of final results; Tobias Timmel for his accurate verification of analysis results; Dejan Danilovic and Stefan Paule for valuable inputs during manuscript editing; Nicola Rovai for study management; Xenia Antoniou for her precious support and assistance throughout the study course. Last but not least we are in debt with all investigators and personnel at BIOTRONIK Italia for their essential contribution to study realization. To the memory of L. Padeletti this work is gratefully dedicated.

Funding

The study was sponsored by BIOTRONIK SE & Co. KG.

Conflict of interest: E.G. has received consultancy fees and grants from Bayer, CVRx, and Medtronic. D.G. and A.G. are employees of BIOTRONIK Italia. All other authors reported no relationship relevant to the content of this article.

Data availability

The data underlying this article were provided by BIOTRONIK SE & Co. KG. Data will be shared on request to the corresponding author with permission of BIOTRONIK SE & Co. KG.

References

1. Maggioni AP, Dahlström U, Filippatos G, Chioncel O, Crespo Leiro M, Drozdz J *et al.*; Heart Failure Association of the European Society of Cardiology (HFA). EURObservational Research Programme: regional differences and 1-year follow-up results of the Heart Failure Pilot Survey (ESC-HF Pilot). *Eur J Heart Fail* 2013;**15**:808–17.
2. Gheorghiade M, Mebazaa A. Introduction to acute heart failure syndromes. *Am J Cardiol* 2005;**96**:1G–4G.
3. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB *et al.* The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation* 2006;**113**:1424–33.
4. Pocock SJ, Ariti CA, McMurray JJV, Maggioni A, Køber L, Squire IB *et al.*; Meta-Analysis Global Group in Chronic Heart Failure. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J* 2013;**34**:1404–13.
5. Rahimi K, Bennett D, Conrad N, Williams TM, Basu J, Dwight J *et al.* Risk prediction in patients with heart failure. *JACC Heart Fail* 2014;**2**:440–6.
6. Whellan DJ, Ousdigian KT, Al-Khatib SM, Pu W, Sarkar S, Porter CB *et al.* Combined heart failure device diagnostics identify patients at higher risk of subsequent heart failure hospitalizations: results from PARTNERS HF (Program to Access and Review Trending Information and Evaluate Correlation to Symptoms in Patients With Heart Failure) study. *J Am Coll Cardiol* 2010;**55**:1803–10.
7. Cowie MR, Sarkar S, Koehler J, Whellan DJ, Crossley GH, Tang WH *et al.* Development and validation of an integrated diagnostic algorithm derived from parameters monitored in implantable devices for identifying patients at risk for heart failure hospitalization in an ambulatory setting. *Eur Heart J* 2013;**34**:2472–80.
8. Boehmer JP, Hariharan R, Devecchi FG, Smith AL, Molon G, Capucci A *et al.* A multisensor algorithm predicts heart failure events in patients with implanted devices: results from the MultiSENSE study. *JACC Heart Fail* 2017;**5**:216–25.
9. Padeletti L, Botto GL, Curnis A, De Ruvo E, D'Onofrio A, Gronda E *et al.* Selection of potential predictors of worsening heart failure: rational and design of the SELENE HF study. *J Cardiovasc Med (Hagerstown)* 2015;**16**:782–9.
10. Varma N, Ricci RP. Telemedicine and cardiac implants: what is the benefit? *Eur Heart J* 2013;**34**:1885–95.
11. Setoguchi S, Stevenson LW, Schneeweiss S. Repeated hospitalizations predict mortality in the community population with heart failure. *Am Heart J* 2007;**154**:260–4.
12. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS *et al.*; ESC Scientific Document Group. 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**:2129–200.
13. Adamson PB. Pathophysiology of the transition from chronic compensated and acute decompensated heart failure: new insights from continuous monitoring devices. *Curr Heart Fail Rep* 2009;**6**:287–92.
14. Zile MR, Bennett TD, St John Sutton M, Cho YK, Adamson PB, Aaron MF *et al.* Transition from chronic compensated to acute decompensated heart failure: pathophysiological insights obtained from continuous monitoring of intracardiac pressures. *Circulation* 2008;**118**:1433–41.
15. Hawkins NM, Virani SA, Sperrin M, Buchan IE, McMurray JJ, Krahn AD. Predicting heart failure decompensation using cardiac implantable electronic devices: a review of practices and challenges. *Eur J Heart Fail* 2016;**18**:977–86.
16. Cleland JG, Antony R. It makes SENSE to take a safer road. *Eur Heart J* 2011;**32**:2225–7.
17. Gardner RS, Singh JP, Stancak B, Nair DG, Cao M, Schulze C *et al.* Heart Logic multisensor algorithm identifies patients during periods of significantly increased risk of heart failure events: results from the MultiSENSE study. *Circ Heart Fail* 2018;**11**:e004669.
18. Costanzo MR. The luck of having a cardiac implantable electronic device. *Circ Heart Fail* 2018;**11**:e004894.
19. Stephen JG, Robert JM, Felker GM. Outpatient worsening heart failure as a target for therapy. A review. *JAMA Cardiol* 2018;**3**:252–9.
20. Stephen JG, Felker GM, Butler J. Outpatient versus inpatient worsening heart failure: distinguishing biology and risk from location of care. *Eur J Heart Fail* 2019;**21**:121–4.