Heart Failure Prediction in Patients with Remotely Monitored Implanted Cardiac Devices: a Multiparametric Model

Eleonora Malloni[1], Alberto Bandini[2], Stefano Severi[1], Cristiana Corsi[1]

[1]DEI, University of Bologna, Cesena, Italy
[2]Morgagni-Pierantoni Hospital, AUSL della Romagna, Forlì, Italy

Abstract

Heart failure (HF) and atrial fibrillation (AF) are widely spread among the global population, especially in elderly patients. Usually, they co-exist, and they are characterized by a complex cause-effect mechanism, responsible for patient's clinical status worsening, which can lead to recurrent hospitalisations. The main goal of this retrospective study was to reduce the clinical and economic impact of HF hospitalisations on the national healthcare system. This was done by implementing a predicting model to timely identify HF exacerbations, before they require a hospital set-up management. The predicting algorithm combines the daily data trends that are remotely collected by ICD and CRT devices to build-up day-by-day a risk index. The index is then compared to an upper nominal threshold to activate an alarm, which indicates a risk to experience a hospital admission in a short time. In this version of the model, we added the AF burden as a new parameter employed in the final risk-score computation, with daily and nightly heart rates and physical activity. Results show a good algorithm performance in the validation group, with sensitivity = 57%, median alerting time = 47.5 days, specificity = 75.5%, false-positive rate per patient-year = 0.83, positive predictive value = 10% and negative predictive value = 99%. The obtained metrics are in line with both the version not including the AF burden, and the alert algorithms designed by the main biomedical device manufacturers.

1. Introduction

Nowadays heart failure (HF) is widely spread out among the global population, and it represents a significant clinical and economic burden. The course of the disease is frequently characterized by clinical exacerbations and phases of instability, defined as acute decompensation episodes, where the rapid onset of severe symptoms and signs leads to recurrent hospital admissions. In fact, HF-related hospitalisations represent the 1-2% of all hospital admissions [1] and HF is also associated with the highest re-admission rate (around 20-25%) after the discharge [2].

Atrial fibrillation (AF) is the most common sustained arrhythmia encountered in HF. Indeed, AF and HF share a common risk profile in elderly patients with cardiovascular risk factors. They are related by a complex pathophysiological inter-relationship, which can justify how they strongly affect each other's outcome with high hospitalisation rates [3].

By exploiting the recent remote monitoring technologies of last generation cardiac electronic implantable devices (CIEDs), significant changes in physiological parameters, which may reflect HF progression and worsening, can be timely detected. In this way, the adequate therapy can be provided to avoid unnecessary hospitalisations [4].

Therefore, the main goal of this study is to implement and validate a risk-predicting model, which, in addition to parameters based on heart rate and physical activity, includes the AF burden, to verify if we can promptly predict HF-related hospitalisations.

2. Materials and methods

The Selene HF study [5] proposed by Biotronik is the main source of inspiration for this study because it proves that by exploiting simple parameters collected by common sensors (available in the majority of devices currently on the market) it is possible to derive an index to reliably predict HF decompensations. At the cardiology department of the Morgagni-Pierantoni hospital in Forlì, a model was previously developed to generate a score for assessing the risk of HF hospitalisations in patients with remotely monitored devices, regardless of the brand of their device.

2.1. Data collection

The study included patients with remotely monitored ICDs, both single (ICD VR) and dual chamber (ICD DR), and CRTs, followed by the four HF units of the AUSL Romagna (Cesena, Forlì, Ravenna and Rimini). Three hundred and fifty-four patients implanted with devices manufactured by Abbott, Biotronik, Boston Scientific and

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.164

Medtronic met the inclusion criteria and were included in the trial. Quantification of AF burden requires the implantation of a device with a working atrial catheter that is able to detect the arrhythmia. Therefore, a preliminary skim was performed to exclude single-chamber devices, devices with a plugged atrial catheter, and those ones that have an atrial catheter but are programmed in VVI mode. After this step, only 221 patients were selected. They were predominantly men (83%) with a mean age of around 71 years, 35% of average left ventricular ejection fraction (LVEF), belonging to NYHA Class I (21%), NYHA Class II (72%) and NYHA Class III (7%). They are implanted with ICDs DR (75), ICDs VDR (19), CRT-Ds (118), CRT-Ps (9) and CRT-D VDR (1).

All remote transmissions were evaluated up to October/November 2023. Then, hospital admissions due to HF exacerbations were collected by accessing the digital records on the cardiac information system Log-80, used by the four units. Overall, 47 hospitalisations over a mean follow-up period of 3.43 years were recorded (0.062 events per patient-year). Data belonging to patients who died or stopped remote monitoring for any reason were used until the last date of available information.

Finally, eligible patients were grouped by centres in order to be assigned to a training and a validation group. Patients from Cesena, Ravenna and Rimini, having together two-thirds of the total number of events, were allocated to the training group, while those from Forlì were assigned to the validation set.

Table 1. Baseline parameters collected in the study population (ICD: Implantable Cardioverter Defibrillator; CRT: Cardiac Resynchronization Therapy; LVEF: Left Ventricular Ejection Fraction; NYHA: New York Heart Association).

Variables	All	Derivation	Validation
Patient number (n)	221	168	53
Follow-up (years)	3.43	3.42	3.46
Age (years)	71.2	70.8	72.6
Gender (male)	184	138	46
Device (%)			
ICD DR	75	57	18
ICD VDR	19	12	7
CRT-D	118	93	24
CRT-D VDR	1	0	1
CRT-P	9	6	3
CRT-P VDR	0	0	0
LVEF (%)	35.4	35.2	35.7
NYHA Class			
I	46	34	12
II	160	124	36
III	15	10	5
Aetiology			
Ischaemic	97	72	25
Dilated idiopathic	98	76	22
Others	26	20	6

This partition ensured larger generalization, since device types, NYHA classes, aetiology and events number are distributed in similar proportions between the two sets (Table 1).

2.2. Risk score computation and alerts

The final risk score is a weighted linear combination of five parameters collected by the devices, whose values were normalized in the range [0,1] to standardize data onto a common scale without distorting the differences in value ranges or losing valuable information. These five parameters are the following: the point value of the daily average frequency without any numerical transformation, the transforms of the daily and nightly average heart rates, physical activity, and AF burden. In particular, in the transformations adopted we evaluated: the monotone increase in the daily average heart rate over periods of 90 days; the dispersion of the nightly average heart rate analysed in a 45-day window; the decreasing trend in the daily patient activity over 25 days; and the daily minutes of AF in a 7-day timeframe.

Each parameter employed in the linear combination is weighted by a factor obtained through the Latin Hypercube Sampling statistical technique which provides a simultaneous randomization of a given set of parameters by defining a priori the number of samples.

For what concerns the automatic alerts, the score is compared day-by-day with an upper nominal threshold, varying theoretically in the range [0:max(score)] with a step 0.01, and with a lower recovery threshold, obtained by multiplying the nominal threshold by a further weight factor, derived with the Latin Hypercube Sampling technique. If the score exceeds the upper threshold for at least three consecutive days, the alarm is switched on. Then, the alarm remains active until the score drops below the lower recovery threshold (Figure 1).

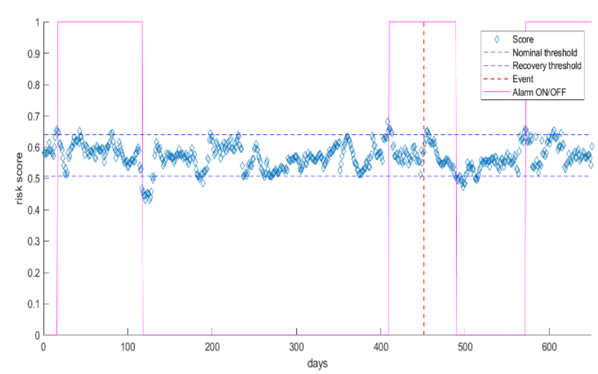

Figure 1. Alarm working principle: the alarm thresholds are represented by dashed blue lines and the alarm on/off is depicted as a pink wave: when the score (blue points) overcomes the upper threshold for at least three days the alarm is on, otherwise the alarm is off.

3. Results

Two hundred and twenty-one patients, who met all the inclusion criteria, were allocated to a training and to a validation group according to the criterion of division by centres. One hundred and sixty-eight patients, belonging to Cesena, Ravenna and Rimini, with 33 events (0.057 events per patient-year) were included in the training cohort, whereas 53 patients from Forlì with 14 events (0.076 events per patient-year) were assigned to the validation one.

By varying the threshold in a pre-identified range, a hundred simulations were run on the training cohort in order to test a hundred possible combinations of the six weights derived with the Latin Hypercube Sampling, whose values are forced to be in the range [0,1]. The best combination of weights responsible to produce the greatest area under the ROC curve (AUC = 0.757), is extracted and applied to the validation cohort to derive the ROC curve (AUC = 0.684 with 95% C.I. 0.678-0.689) (Figure 2).

By analysing the point corresponding to the nominal threshold of 0.64, the performance metrics were derived. At this threshold, the sensitivity was 57%, meaning that 8 events over 14 had been properly detected with a median alerting time of 47.5 days (interquartile range: 30 to 80.5 days). The false-positive rate per patient-year was 0.83 and the specificity was 75.5%. Moreover, the time during which the alarm was active corresponds to the 24.6% of the entire observation period of the follow-up. Finally, the positive (PPV) and negative (NPV) predictive values have been computed, resulting 10% and 99.91% respectively.

3.1 Results obtained without AF burden

The obtained results have been compared with the performance of the same algorithm but without including the AF burden, tested on the original 353 patients implanted also with single-chamber ICDs and CRTs.

For this dataset, the overall mean follow-up time was 2.93 years. Patients were allocated to the training and validation groups according to their 64 usable events. The training cohort included 268 patients from Cesena, Ravenna and Rimini with 39 events (0.05 events per patient-year), while 85 patients from Forlì with 25 events (0.1 events per patient-year) were assigned to the validation cohort. In this case, the optimal weights combination identified in the training cohort, after running one hundred simulations, was used to derive the ROC curve in the validation cohort (AUC = 0.70 with 95% C.I. 0.694-0.705) (Figure 3).

At the analysed threshold of 0.97, the sensitivity of the predictive model was 56%, meaning that 14 events over 25 had been properly identified with a median alerting time of 43 days (interquartile range: 25 to 59 days). The false-positive rate per patient-year was 1.16 and the specificity was 82%. Moreover, the percentage of days in which the alarm was active is 18% of the entire observation period, whereas the positive and negative predictive values were 6.7% and 99.88%, respectively.

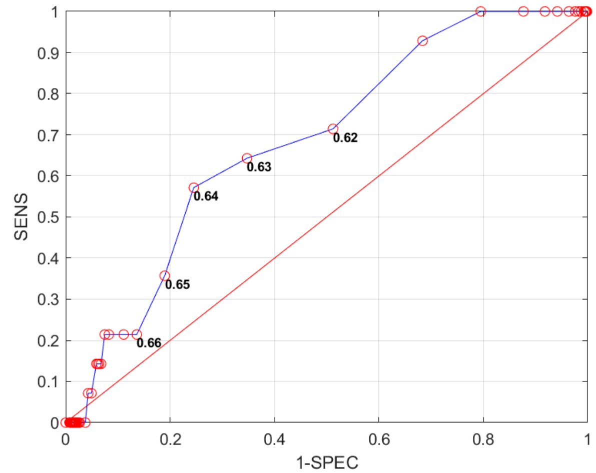

Figure 2. ROC curve of the validation cohort including the AF burden in the score computation, evaluated by plotting the sensitivity versus 1- specificity.

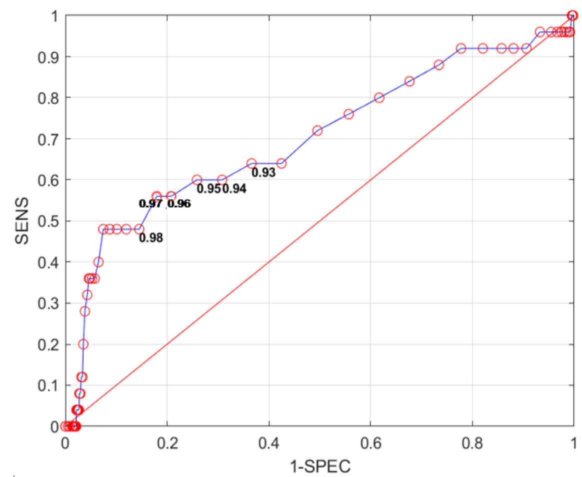

Figure 3. ROC curve of the validation cohort not including the AF burden in the score computation, evaluated by plotting the sensitivity versus 1- specificity.

4. Discussion

Our model for HF prediction showed comparable results and performances compared to the multi-parametric risk-score indices for the prediction of HF-related hospitalisations implemented by the main biomedical devices companies that are currently on the market (Table 2). On note, our approach is device-independent and it can be applied to remote monitoring data acquired with devices

Table 2: Sensitivity (SENS), specificity (SPEC), false-positive rate, positive predictive value (PPV), negative predictive value (NPV) and median alerting time of the multi-parameter risk indices available in the literature and of those implemented in the present study.

Multi-parameter index	SENS (%)	SPEC (%)	FP rate (ppy)	PPV (%)	NPV (%)	Median alerting time (days)
TriageHF (high risk)	46	90.2	/	/	/	/
HeartLogic	70	85.7	1.47	5.6	99.98	34
HeartInsight (for thresholds 3.5-4.5)	64.5-54.8	75.3-86.5	1.05-0.67	5.3-7.7	96.6-96.7	60-43
Risk-score with AF burden	57	75.5	0.83	10	99.91	47.5
Risk-score without AF burden	56	82	1.16	6.7	99.88	43

from all manufacturers.

An interesting comparison between the two algorithm versions, with and without AF burden, concerns the metrics derived in the validation groups. In fact, both models have comparable performances in terms of sensitivity, median alerting time and NPV. On the contrary, specificity and false-positive rate per patient-year are slightly different. The predicting model including the AF burden has both lower specificity and false-positive rate, meaning that the probability of correctly identifying the absence of an event and the number of improper alarm activations (false positives: 140 with AF burden, 256 without) are both lower compared to the model not including AF burden. On the contrary, the PPV is higher because the number of days in which the alarm is improperly active is lower, compared to the case without AF burden.

The higher percentage of alert activations depends on the scale factor between the nominal and recovery thresholds. Since the recovery threshold is 79% of the nominal one, it takes more days after the event for the risk score to drop below the lower threshold, in comparison to the algorithm without the AF burden, where the scale factor between the two thresholds is 0.9.

5. Conclusions

A model for HF-related hospitalisations prediction was designed and validated in patients implanted with a remotely monitored cardiac device with a working atrial catheter from AUSL Romagna in Italy.

. The developed risk-score index included standard information available from all devices: daily and nightly average heart rates, physical activity and AF burden (in one version of the model).

In general, both versions of the predicting algorithms, including the AF burden in the score computation or not, achieved good performance. Hence, they represent the right path to follow in order to ensure the well-being of patients and reduce both the clinical and economic burden associated to recurrent HF hospitalisations for the healthcare system.

References

[1] Ambrosy, Andrew P., et al. «The Global Health and Economic Burden of Hospitalizations for Heart Failure». *JACC*, vol. 63(12), pp. 1123–33, 2014.

[2] Groenewegen, Amy, et al. «Epidemiology of Heart Failure». *European Journal of Heart Failure*, vol. 22(8), pp. 1342–56, 2020.

[3] Verhaert, Dominique V. M., et al. «The Bidirectional Interaction between Atrial Fibrillation and Heart Failure: Consequences for the Management of Both Diseases». *EP Europace*, vol. 23, Supplement_2, pp. ii40–45, 2021.

[4] Theuns, Dominic A. M. J., et al. «Remote Monitoring of Heart Failure in Patients with Implantable Cardioverter-Defibrillators: Current Status and Future Needs». *Sensors*, vol. 21(11), pp. 3763, 2021.

[5] D'Onofrio, Antonio, et al. «Combining Home Monitoring Temporal Trends from Implanted Defibrillators and Baseline Patient Risk Profile to Predict Heart Failure Hospitalizations: Results from the SELENE HF Study». *EP Europace*, vol. 24(2), pp. 234–44, 2022.

Address for correspondence:

Eleonora Malloni
DEI, University of Bologna, Campus di Cesena
Via dell'Università 50, 47521 Cesena, Italy
E-mail address: eleonora.malloni@studio.unibo.it