Logistic Regression to Enhance Risk Assessment by Left Ventricular Ejection Fraction and f99

Corrado Giuliani1, Cees A Swenne2, Sumche Man2, Sandro Fioretti1, Francesco Di Nardo1, Laura Burattini1

1Department of Information Engineering, Polytechnic University of Marche, Ancona, Italy
2Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands

Abstract

Sudden cardiac death remains one of the leading causes of death in developed countries. Left ventricular ejection fraction (LVEF) and f99 are two noninvasive indexes of cardiovascular risk (traditional the former and innovative the latter) which, taken singularly, have not shown sufficiently high SCD predictive power to justify preventive actions. Thus, the aim of the present study was to investigate if their combination improves predictability of ventricular arrhythmias. To this aim, ECG recordings from 266 ICD patients, of which 76 developed ventricular tachycardia or fibrillation during the 4-year follow-up (ICD_Cases), and 190 did not (ICD_Controls). The ECGs of each patient was used to compute the f99, a repolarization index defined as the frequency at which the cumulative power energy reaches 99%. Eventually, a logistic regression between LVEF and f99 was performed in order to derive a combined predictor (CP) of ventricular arrhythmia. Goodness of each index was evaluated in terms of the area under the receiver operator curve (AUC). When used singularly, LVEF and f99 respectively provided an AUC of 0.67 and 0.64. When combined to get CP=-0.15-0.05·LVEF+0.03·f99, this provided an AUC of 0.71. In conclusion, use of logistic regression improves LVEF and f99 predictability of ventricular arrhythmias.

1. Introduction

Notwithstanding recent advances in the treatment of life-threatening ventricular arrhythmias, sudden cardiac death (SCD) remains one of the leading causes of death in developed countries [1]. Nowadays, patients are selected for clinical evaluation and treatment of ventricular arrhythmias only after they have experienced and survived a major cardiac event. Thus, preventive identification and treatment of these high-risk subjects is expected to have a great impact on the problem of SCD. There are several non-invasive indexes of risk for the occurrence of SCD that can be used for prevention. Among these, left ventricular ejection fraction (LVEF), which is defined as the quantity of blood ejected at every heartbeat, is considered as the standard SCD predictor [2]. However, it is measured only via echocardiography, which is not performed on routine clinical tests. Instead, electrocardiography (ECG) is a more simple, common, and cheap non-invasive test to evaluate cardiac electrical functionality, so that ECG-derived indexes would be preferable. Out of the vast amount of ECG indexes produced in the last decades, such as the ones describing the heart rate variability [3,4], many authors have focused on repolarization-based indexes, such as the QT-interval [5,6], the T-wave alternans [7,8] and others [9-14]. Indeed, defects of cardiac repolarization, noninvasively identifiable in the ECG ST segment and T wave, are among the major causes of SCD [5-7,15]. Nevertheless, neither LVEF [16] nor the aforementioned indexes have proven sufficiently reliable for SCD prevention [4,5,16-18]. Thus, we recently introduced a new frequency-based repolarization index termed f99, which is defined as the frequency at which the repolarization normalized cumulative energy reaches or overcomes 99% [19-21]. Notwithstanding, although f99 was found able to discriminate ventricular arrhythmias [21], still its predictive power is not sufficient to justify preventive diagnostic and therapeutic procedures in previously asymptomatic patients [21]. To overcome this issue, some authors have reported an increase in the predictive power when combining different indexes through various techniques [22,23], such as multiple linear regression or logistic regression (LR) [24]. Among these, the latter is the best fit in a case-control study since it uses a binary logistic model to predict a binary or dichotomous response variable from on one or more independent predictor variables [25], opposite to the former, which would be more suitable when the outcome variable is continuous [26]. Thus, being LVEF and f99 uncorrelated [21], the aim of the present study was to evaluate if combination of both LVEF and f99 by LR could enhance their predictive value. To this aim, the Leiden University Medical Center database of exercise ECGs in heart failure patients with implanted cardiac defibrillator (ICD) was used.
2. Methods

2.1. Study population and clinical data

The study population consisted of 266 patients with an ICD for primary prevention because of a depressed LVEF (<35%). The clinical data of the ICD patients, including age, gender, LVEF, body mass index (BMI), heart rate (HR), New York Heart Association (NYHA) functional class, therapies and medications are collected in the so-called Leiden University Medical Center database of exercise ECGs in heart failure patients with ICDs. All patients underwent a bicycle ergometer test, which consisted of an approximately 10-minute bicycle test during which the workload was incremented from zero to the patient’s maximal capacity by applying load-increments of 10% of the expected maximal capacity every minute. During the bicycle ergometer test, 8 leads (I, II, V1 to V6) ECG recordings were obtained using a CASE 8000 stress test recorder (GE Healthcare, sampling rate: 500 Hz; resolution: 4.88 μV/LSB). After the exercise test, all patients underwent a 4-year follow-up at the end of which they were classified as either ICD_Cases (76 patients) if they developed ventricular tachycardia (VT) or ventricular fibrillation (VF) during the follow-up, or ICD_Controls (190 patients) if they did not.

2.2. f99 computation

The f99 is a recently published [21] frequency-content-based index of repolarization. Briefly, first, from the 8-leads ECG signals available, leads III, aV1, aVr, aVf, X, Y and Z were obtained through a well-known transformation [27]. Second, the first 30 seconds of each 15-lead ECG tracing were selected since showing conditions close to rest, and each lead was resampled at 200 Hz and pre-processed for noise removal and baseline subtraction (0.5-35 Hz band-pass filter). Only leads characterized by stable HR (RR-interval standard deviation < 10% of mean RR) and clean and sinus rhythm (no more than 2 ectopic beats and artifacts) were used for computation of f99. Third, for each of the previously identified 30-s 15-lead ECG windows, the median beat is computed and repolarization is identified through empirical formulae. Fourth, frequency content is evaluated through Fourier transform to get signal’s energy, which is then normalized by the total energy. Fifth, f99 is defined as the frequency at which the repolarization normalized cumulative (ERPS%) energy reaches or overcomes 99% (Fig. 1 [20]). For further details see [21].

To overcome issues due to repolarization dispersion, we used a lead-system analysis. We initially computed a f99 value for each one of the 15 available ECG leads. Then, the single-lead f99 values were maximized over the Frank’s orthogonal XYZ leads because previously found to be able discriminate the two ICD groups of this population [21].

2.3. Logistic regression

LR analysis was first introduced in 1970s [24] as an alternative technique to analyze the relationship of a binary or dichotomous outcome variable, such as the presence/absence of a disease, with one or more quantitative predictor variable, thus expanding the technique of multiple regression analysis [25]. In many medical applications [22-23, 28-32], such as the one in this study, there may be more predictor variables which concur in determining the outcome variable Y. Thus, given N predictors, the logistic model predicts the “logit” (Eq. 1), that is the natural logarithm of odds, of Y from the N predictors. Consequently, the model has the following form:

\[ \ln \left( \frac{\pi}{1-\pi} \right) = \log(odds) = \text{logit} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_N X_N \]  

(1)

Hence,

\[ \pi = \text{Probability} (Y = \text{outcome of interest}) = \frac{e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_N X_N}}{1 + e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_N X_N}} \]  

(2)

where \( \pi \) is the probability of the event, \( \alpha \) is the intercept of Y, \( \beta_N \) are slope parameters and \( X_N \) are the predictors. \( \alpha \) and \( \beta_N \) are estimated by the maximum likelihood method [25].

Consequently, in this study, LR is used to combine LVEF and f99 indexes to get a combined predictor CP. Thus, the resulting equation (Eq. 3) has the following form:

\[ CP = \alpha + \beta_{\text{LVEF}} \text{LVEF} + \beta_{\text{f99}} \text{f99} \]  

(3)
2.5. Statistics

The values of f99 were computed for both ICD_Controls and ICD_Cases populations in Matlab. Then, to evaluate a parameter predictive power for the occurrence of ventricular arrhythmias the area under the receiving operating characteristic curve (AUC) was computed using the pROC package [33] in R, as well as the linear regression analysis.

3. Results

When used singularly, LVEF and f99 were respectively significantly lower and significantly higher in the ICD_Cases than in the ICD_Controls (median LVEF: 32% vs. 36%, P<10^-4; median f99: 37 Hz vs. 30 Hz, P<10^-5) and respectively provided an AUC of 0.67 and 0.64. LR was then performed and the CP was obtained as follows:

\[
CP = -0.15 - 0.05 \cdot \text{LVEF} + 0.03 \cdot f99
\]

thus being the CP intercept \(\alpha = -0.15\) and the two slope parameters \(\beta_{\text{LVEF}}\) and \(\beta_{f99}\) -0.05 and 0.03, respectively. CP was then evaluated for risk assessment, resulting significantly lower in the ICD_Cases than in the ICD_Controls (median CP: -1.1 vs. -0.7, P<10^-6) and providing an AUC of 0.71.

4. Discussion

This study aimed to evaluate if combination of LVEF and f99 by LR could be used to enhance their predictive value, that would have a great impact in the problem of SCD, which is one of the leading causes of death with more than 350 000 death per year [1,34]. Indeed, a multitude of indexes describing various SCD aspects have been published in the last decades [2-14,21]. Among these, LVEF is commonly considered as the standard SCD predictor [2,16], while f99 [19-21] is the most recently published repolarization index. However, neither LVEF nor f99 have provided a sufficient predictive power. Indeed, when used singularly, LVEF and f99 were respectively significantly lower and significantly higher in the ICD_Cases than in the ICD_Controls and respectively provided an AUC of 0.67 and 0.64. Moreover, as for LVEF and f99, none of all the published indexes has yet stood out with an acceptable predictive power [16-18]. In addition, this large number of variables may confuse physicians in the decision-making process, thus making them rely more on subjective impressions than on a strong, evidence-based background. Nevertheless, this is a common problem in other medical field, such as breast cancer [28], HIV detection [29] and others [30-32]. Thus, to avoid these problems and besides willing to improve the predictive power, some authors [28-32] have introduced the use of LR analysis against a more traditional regression technique for two or more predictors, such as multiple linear regression, which has proven to be unsatisfactory when the outcome is dichotomous or binary [26,35]. Consequently, LR was chosen in this study to evaluate if this technique could be used to obtain a CP combining LVEF and f99 indexes, thus reducing possible confusion in the index choice and increasing the predictive power. To this aim, the Leiden University Medical Center database of exercise ECGs in 266 heart failure patients with ICDs was employed, classified as either ICD_Cases (76 patients) if they developed ventricular tachycardia (VT) or ventricular fibrillation (VF) during the follow-up, or ICD_Controls (190 patients) if they did not. A value of LVEF and f99 was obtained for each of these subjects. Then, LR was applied and CP was obtained, resulting significantly lower in the ICD Controls than in the ICD_Cases and providing an AUC of 0.71.

5. Conclusion

LR analysis could be used to combine LVEF and f99 in order to improve their predictability of ventricular arrhythmias.

References


