Multivariate Analysis of Follow-up Physiological Data Recorded by Cardiac Implantable Devices

E Roux[1], AI Hernández[1], L Graindorge[2], G Carrault[1], P Mabo[3]

[1] INSERM U642, LTSI, University of Rennes, Rennes, France
[2] ELA Medical, Inc., Le Plessis-Robinson, France
[3] Department of Cardiology, University Hospital, Rennes, France

Abstract

New cardiac implantable devices (IDs) allow the acquisition of an increasing amount of data relative to the patient's activity. A quantitative analysis of these data can be of particular interest to optimize the stimulation therapy and to improve patient follow-up. This work presents a method to: i) evaluate the information content of the ID memory data; ii) define synthetic indexes that would summarize these data and ease their interpretation and iii) characterize and compare different patient populations. The proposed approach is based on a coding stage allowing the representation of time-varying data, followed by a multivariate analysis stage, based on Principal Component Analysis (PCA). Results show that the initial 38 variables obtained from the IDs memory contain pertinent information that can be synthesized with a low number of indexes (the first four factorial axes of the PCA represent 70% of the total variance) and that these indexes can be useful to follow the evolution of a given patient's state.

1. Introduction

The development of new leads and efficient algorithms and the greater storage capacity of implantable devices (IDs) allow to acquire and record an increasing amount of data relative to pacemaker functions (*e.g.* logs of the pacemaker activity or lead status) and to cardiac activity dysfunctions, such as arrhythmia (*e.g.* electrograms of atrial fibrillation episodes) [1, 2].

A particularly interesting application of these new ID concerns patients suffering from refractory heart failure (RHF) that have an indication for a cardiac resynchronization therapy (CRT). RHF is one of the most important causes of morbidity and mortality on persons aged over 55 years [3, 4, 5]. Moreover, although the CRT is currently established as the most adapted treatment for patients suffering from RHF with delayed inter-ventricular conduction [6], 20 to 35% of the patients, called non-responders, do not benefit from CRT [7]. A study even shows that 10% of CRT recipients experienced worsening of symptoms and mechanical dys-synchrony [8]. Data acquired from these new IDs are potentially useful to extract relevant information about the state of the patient and to ease patient follow-up. However, these data are large, multivariate, time-dependent and heterogeneous, and experts (*i.e.* physicians, biomedical engineers, ...) have only poor background knowledge on their underlying physiological processes.

This paper presents a multivariate analysis method based on a coding stage that allows the representation of time-varying data, and a Principal Component Analysis (PCA) [9] that can facilitate the interpretation of these datasets. In fact, PCA seems to be particularly eligible and, to our knowledge, has not been applied yet in the given context. The present study shows how PCA can be used to i) identify the correlations between the available variables, ii) highlight the most informative variables, iii) study the relationships between synthetic indexes provided by PCA and the usually used clinical indexes and iv) compare two patient populations with different indications (heart rhythm and/or conduction defects and heart failure).

2. Methods

2.1. Clinical protocols and acquired data

Two clinical protocols have been defined for this study:

• Protocol 1 ($\mathscr{P}_1$) includes patients with rhythm and/or heart conduction defects and no concomitant cardiac diseases. The ID is assumed to totally compensate for their cardiac problem. These patients are supposed to have a relatively stable state of health after surgery. IDs are controlled at the end of the first and/or the third postoperative months. Daily data recorded in the ID are retrieved at the same time and concern a one-month length period;

• Protocol 2 ($\mathscr{P}_2$) includes patients suffering from RHF and having an indication for a CRT. The same variables as in $\mathscr{P}_1$ are recorded. Data are retrieved at the end of

Table 1. List of physiological variables

Description	Names	Units
Total duration within the activity level	$Duration_E$ [1]	s
	$Duration_R$	
Cumulative values of acceleration	$Acceleration_E$	$m \cdot s^{-2}$ (g)
	$Acceleration_R$	
Cumulative values of impedance	$Impedance_E$	millivolts (mV)
	$Impedance_R$	
Cumulative number of ventilation cycles	$NbBreaths_E$	$NbVC$
	$NbBreaths_R$	
Cumulative number of cardiac cycles	$NbCardCyc_E$	$NbCC$
	$NbCardCyc_R$	
"Mean" [2] activity intensity	$ActIntensity$	$g \cdot s^{-1}$
$Impedance_E$ over $Acceleration_E$	$ImpOverAcc$	$mV \cdot g^{-1}$
"Mean" heart rate	$HeartRate_E$	Beats per minute (bpm)
	$HeartRate_R$	
"Mean" Impedance minute ventilation	$ImpMinuteVent_E$	$mV \cdot min^{-1}$
	$ImpMinuteVent_R$	
"Mean" ventilation frequency	$VentilationFreq_E$	$NbVC \cdot min^{-1}$
	$VentilationFreq_R$	
$Impedance_E$ over $Impedance_R$	$ImpedanceRate$	none

[1] Subscripts E and R are for Exercise and Rest, respectively.
[2] The duration of each *Exercise* and *Rest* period that occur within 24 hours is unknown. Only the cumulative duration is known. Consequently, this "mean" is not the average of the variable values over 24 hours, excepted if all the periods are of the same duration.

the third, the sixth and the twelfth postoperative months and cover a one-month length period. At the same time, patients perform a six-minute walk test and their ejection fraction (*EF*) is evaluated. The six-minute walk test (*6'WT*) consists in measuring the total distance covered by a patient walking during at most six minutes. The *EF* is estimated by means of an echocardiography and corresponds to the ratio of the amount of blood pumped by the left ventricle of the heart on the amount of blood the ventricle contains. The *EF* characterizes the efficiency of the heart and especially of the left ventricle.

Data acquired from the IDs are provided by two sensors: an accelerometer and a transthoracic impedance sensor that reflect the physical activity and the ventilation of the patient [10], respectively. They enable to classify the activity level of the patients into two states: *exercise* and *rest*. These states are defined by the joint information given by the two sensors and by means of two thresholds. For each of these states, 24-hour cumulative values of all variables are computed and recorded in the ID memory over 30-day follow-up periods. The 19 physiological variables are listed in table 1.

2.2. Principal Component Analysis (PCA)

PCA [9] is a non-parametric method that aims at optimally representing individuals and variables in order to identify the pertinent information "hidden" in the data, by defining a new space with uncorrelated axes defined by linear combinations of the initial variables and called *factorial axes*. This representation is optimal in the sense that it maximizes the variance along the factorial axes given the constraint of orthogonality of the axes. The input data set of the PCA is a matrix in which each row corresponds to a statistical individual and each column to a variable. When previously centering and scaling the data, positions of variables along the factorial axes and mutual positions of variables can be directly interpreted in terms of correlations with the factorial axes and between variables, respectively. Moreover, supplementary variables and individuals can be jointly represented with the variables and individuals of analysis. They do not participate to the computation of the PCA and their projection on the PCA planes: i) facilitates the interpretation of the PCA factorial axes by relating them to meaningful variables (*e.g.* age, sex, etc.) and ii) enables the characterization of supplementary individuals according to their locations with respect to individuals of analysis. Consequently, PCA enables to compare a *reference* population (*i.e.* patients included in $\mathscr{P}_1$, considered as stable and that define individuals of analysis) with a *test* one (*i.e.* patients included in $\mathscr{P}_2$ considered as supplementary individuals).

However, the application of PCA on time series demands to code the information with a suitable format. In this work, each variable of a 30-day record is divided into two 15-day time series. A *statistical individual* is characterized by the mean and the standard deviation of a given patient's data during a 15-day time period. In this way, a given patient is represented by several statistical individuals. This method leads to 38 variables: 19 corresponding to averages and 19 to standard deviations, denoted with the prefixes "$M-$" and "$SD-$", respectively. Statistical individuals whose temporal support begin before the 10^{th} day after surgery were withdrawn from the analysis in order to minimize the effect of the transient state observed in many patients just after the ID implantation.

3. Results

Concerning $\mathscr{P}_1$, 86 records related to 59 patients have been obtained. 8 patients suffer from sinus dysfunction, 29 from AV blocks, 2 from AV blocks and sinus dysfunction. There are 25 women and 34 men. The mean age of the patients is 73 (standard deviation 10, minimum 46 and maximum 91). PCA is thus applied to 125 statistical individuals related to these 59 patients, *i.e.* to an array of 125 rows and 38 columns.

The $\mathscr{P}_2$ provided 16 records related to 13 patients: 10 men and 3 women. The mean age is 66 (minimum 38 and maximum 87). The mean *EF* of the patients is 39.9 (23-60) and the mean performance concerning the *6'WT* is 364.2 (108-470).

3.1. Factorial axes definition/interpretation

The first four factorial axes obtained by the PCA represent 35.1, 16.7, 9.9 and 6.9% of the total variance of the data, *i.e.* near 70% in total.

The first axis is mainly defined by the time spent in exercise and the intensity of the efforts, *i.e.* $M-ActIntensity$, $M-Impedance_E$ and $M-VentilationRate$. Consecuently, in figure 2, the more the individuals are located to the right of the plane, the longer is the time spent in exercise and the more important are the efforts they make. Correlations between mean and standard deviation for the variables contributing to the definition of the first factorial axis indicate that the capacity to modulate the daily activity duration and intensity is related to the baseline level of the activity duration and intensity.

The second axis is defined by the ventilation activity in terms of amplitude, frequency and flow rate, especially in rest. This axis can be interpreted as an "axis of cardiovascular efficiency". In figure 2, the more the individuals are located to the upper part of the plane, the more important is their ventilation in rest in terms of amplitude, frequency and flow rate (*i.e.* their cardiovascular system is less "efficient"), and this independently of the daily activity duration and intensity (*i.e.* of the position along the first axis).

The third and the fourth factorial axes cannot be intuitively related to a global functional or physiological state. Consequently, in the following, only the plane defined by the first two factorial axes, representing 51.8% of the overall variance, is studied.

Figure 2 presents patients from both populations on the first factorial plane. It can be observed that intra-patient variability is lower than inter-patients variability and that a higher intra patient variability is associated with higher activity levels. This is shown by the longer way covered along the first factorial axis by patients with high activity levels, showing a significant evolution with respect to the baseline (right side of figure 2).

3.2. Supplementary variables/individuals

Figure 1 represents the projection of the supplementary variable *Age* (for $\mathscr{P}_1$) on the first factorial plane. The linear correlation coefficient between the variable *Age* and the first factor is -0.54 with $p << 0.01$ (see Figure 1). This means that the older the patients are, the lower the total time spent in exercise and the intensity of the activity. This correlation remains significant by withdrawing patients under 60 years old (patients supposed to have a professional activity). Even if these results were expected, they show the suitability of the axes obtained with the PCA.

According to Figure 2, women and men in $\mathscr{P}_1$ seem to be discriminated by the second factorial axis. A Kruskal-Wallis (KW) test performed on the mean locations of each patient, showed that the difference between *Women* and *Men* is not statistically significant (result of the KW test = 0.09). However, women have significantly higher values than men ($p < 0.05$) for the following variables contributing to the second axis: $M-Impedance_R$, $M-ImpMinuteVent_E$ and $ImpOverVAcc$. These differences can be due to anatomic variations [10] possibly related to sex and not to different ventilation status in term of efficiency.

In the following, PCA axes are considered as axes of a reference frame in which patients of $\mathscr{P}_2$ can be projected to study their relative functional states. Given the relatively low number of patients included in $\mathscr{P}_2$ (13 patients), qualitative results observed with PCA are not corroborated by statistical tests.

Patients of $\mathscr{P}_2$ have higher values for variables positively correlated with the second axis whatever their level of activity (*i.e.* their positions along the first factorial axis) is (see Figure 2). This result highlights, from a general point of view, a lower cardiopulmonary efficiency for patients with HF. Figure 1 presents the projections, onto the first PCA plane, of the variables *EF* and *6'WT* related to patients of $\mathscr{P}_2$ that suffer from RHF. The coordinates of these clinical variables along a given axis are the linear correlation coefficients between these variables and the projections, on the factorial axis, of the individuals of $\mathscr{P}_2$.

The six-minute walk test values seems correlated with the coordinates of the individuals on the first factorial axis. Consequently, by referring to the interpretation of this axis and to Figure 1, this tends to show that the more active the patients with HF are, the better their performances during the *6'WT* are. On the other hand, the *EF* values are not correlated with the coordinates on the first or the second axes.

4. Discussion and conclusions

This study was designed to show the appropriateness of the PCA to analyze multivariate follow up data relative to patients suffering from different cardiac diseases. Two clinical protocols have been designed, providing data on patients suffering from rhythm and/or heart conduction defects and from RHF.

The present paper let us evaluate not only the pertinence of the PCA for analyzing follow-up data but also the informative potential of the physiological data recorded in IDs. The first two factorial axes of the PCA represent more than 50% of the total variance of the data and can be intuitively related to functional and physiological states of the patients. Supplementary variables and individuals have permitted to identify relationships between *sex*, *age* and the proposed synthetic indexes. Patients suffering from RHF have been characterized in a relative manner with respect

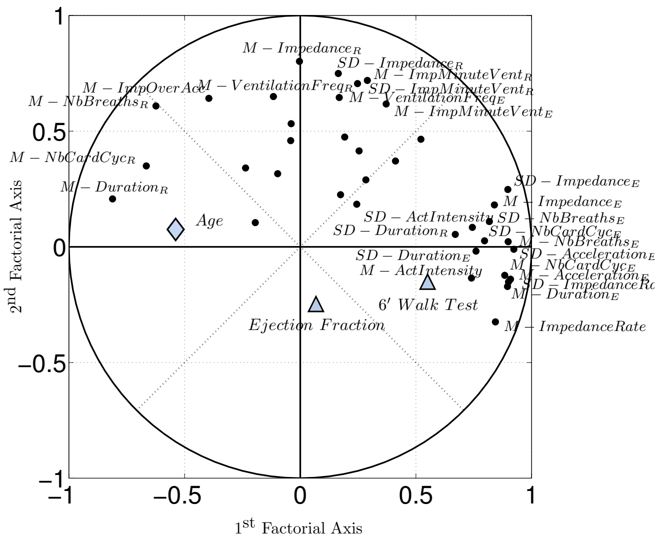

Figure 1. Analysis and illustrative variables represented in the first plane of the PCA. Only variables of analysis that are significantly correlated with the factors are labeled.

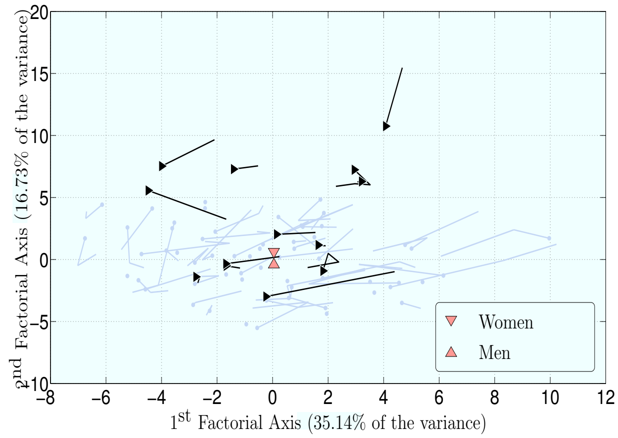

Figure 2. Analysis (in gray) and supplementary (in black) individuals represented in the first plane of the PCA. A patient is represented by a trajectory that links the statistical individuals related to him. The first point in time for each patient is marked up by a circle ($\mathscr{P}_1$) or a triangle ($\mathscr{P}_2$).

to reference patients' state and behavior. Moreover, correlation between the *EF*, the *6'WT* and the physiological variables recorded in the IDs have been studied. From a clinical standpoint, this preliminary study is important as it appears that the information recorded in the ID memory can be exploited and analyzed by a multidimensional data analysis method for the objective evaluation and prognosis of the patients' functional state. For example, data acquired from subsequent follow-ups can be represented as a trajectory on the PCA plane in order to study the evolution of the patients state with respect to a reference population. It would also be possible to identify typical trajectories related to health deterioration. Of course, data related to more patients would permit to perform a specific PCA with a population suffering from HF. New factorial axes definitions are expected due to the fundamental differences between these patients and the ones of $\mathscr{P}_1$. As patients of $\mathscr{P}_2$ are not stimulated in the atrium, the heart rate should become a more discriminant variable and contribute to the first two factorial axes in a greater proportion.

References

[1] Pollak WM, Simmons JD, *et al.* Pacemaker diagnostics: A critical appraisal of current technology. PACE 2003;26:76–98.

[2] Yang A, HochhŁusler M, *et al.* Advanced pacemaker diagnostic features in the characterization of atrial fibrillation: impact on preventive pacing algorithms. PACE Jan 2003; 26(1 Pt 2):310–313.

[3] Cowie MR, Wood DA, *et al.* Incidence and aetiology of heart failure - a population-based study. Eur Heart J 1999; 20:421–428.

[4] Hunt S, Baker D, *et al.* ACC/AHA Guidelines for the Evaluation and Management of Chronic Heart Failure in the Adult: Executive Summary. Circulation Dec 2001; 104(24):2996–3007.

[5] Mosterd A, Hoes AW, *et al.* Prevalence of heart failure and left ventricular dysfunction in the general population; The Rotterdam Study. Eur Heart J Mar 1999;20(6):447–455.

[6] Cleland JGF, Daubert JC, *et al.* The Effect of Cardiac Resynchronization on Morbidity and Mortality in Heart Failure. N Engl J Med Mar 2005;.

[7] Molhoek SG, Erven LV, *et al.* QRS duration and shortening to predict clinical response to cardiac resynchronization therapy in patients with end-stage heart failure. PACE Mar 2004;27(3):308–313.

[8] Arya A, Haghjoo M, *et al.* Who is the best candidate for cardiac resynchronization therapy: Implications of recent studies. Indian Heart Journal 2003;55(6).

[9] Le Roux B, Rouanet H. Geometric Data Analysis. From Correspondence Analysis to Structured Data Analysis. Kluwer Academic, 2004.

[10] Simon R, Ni Q, *et al.* Comparison of impedance minute ventilation and direct measured minute ventilation in a rate adaptive pacemaker. PACE Nov 2003;26(11):2127–2133.

Address for correspondence:

Alfredo I. Hernández
INSERM U642, LTSI, Université de Rennes I
Campus de Beaulieu, Bâtiment 22
35042 RENNES Cedex, FRANCE