MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press

Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method

Franck Le Duff [a], Cristian Muntean[b], Marc Cuggia[a], Philippe Mabo[b]

[a] *Medical Informatics Laboratory, Medical School, Av du Pr Léon Bernard, F-35043 Rennes Cedex France*
[b] *Cardiology and Vascular Disease Department, Regional Hospital, 2 rue Henri le Guilloux 35033 Rennes Cedex France*

Abstract

BACKGROUND: The prognosis of life for patients with heart failure remains poor. By using data mining methods, the purpose of this study was to evaluate the most important criteria for predicting patient survival and to profile patients to estimate their survival chances together with the most appropriate technique for health care. METHODS: Five hundred and thirty three patients who had suffered from cardiac arrest were included in the analysis. We performed classical statistical analysis and data mining analysis using mainly Bayesian networks. RESULTS: The mean age of the 533 patients was 63 (± 17) and the sample was composed of 390 (73 %) men and 143 (27 %) women. Cardiac arrest was observed at home for 411 (77 %) patients, in a public place for 62 (12 %) patients and on a public highway for 60 (11 %) patients. The belief network of the variables showed that the probability of remaining alive after heart failure is directly associated to five variables: age, sex, the initial cardiac rhythm, the origin of the heart failure and specialized resuscitation techniques employed. CONCLUSIONS: Data mining methods could help clinicians to predict the survival of patients and then adapt their practices accordingly. This work could be carried out for each medical procedure or medical problem and it would become possible to build a decision tree rapidly with the data of a service or a physician. The comparison between classic analysis and data mining analysis showed us the contribution of the data mining method for sorting variables and quickly conclude on the importance or the impact of the data and variables on the criterion of the study. The main limit of the method is knowledge acquisition and the necessity to gather sufficient data to produce a relevant model.

Keywords

Knowledge extraction; Data Mining; Heart Failure; Congestive /mortality /physiopathology/ therapy; Retrospective Studies; Survival Rate; Bayes Theorem, Artificial Intelligence, Clinical Trials/statistics & numerical data, Data Interpretation, Statistical; Heart Diseases.

Introduction

Cardiac arrest is defined as a spontaneous irreversible arrest of the general circulation by cardiac inefficacity. It is recognized with the absence of the femoral pulse for more than 5 seconds.

Without resuscitation, cardiac arrest leads to sudden cardiac death. The public health impact of sudden cardiac death is heavy since the survival rate is estimated at between 1 and 20 % for cardiac arrest patients. This represents 300,000 to 400,000 deaths a year in the United States and 20,000 to 30,000 deaths in France. The profile of the patient is now well known since it generally concerns men from about 45 to 75 years. Hospitalization must be optimal and fast. According to the type of cardiac attack, the procedure of assumption of responsibility can vary and some studies [1] show the interest of various techniques over others, according to the cause of the cardiac arrest.

Heart disease is the number one cause of death in the U.S. According to the American Heart Association[1], an estimated 1,100,000 people in the U.S. will have a coronary attack each year. Ninety five percent of sudden cardiac arrest victims die before reaching the hospital and heart disease claims more lives each year than the following six leading causes of death combined (cancer, chronic lower respiratory diseases, accidents, diabetes mellitus, influenza and pneumonia). Almost 150,000 people in the U.S. who die from heart disease each year are under the age of 65.

These data show the interest for predicting the risk of death after heart failure and the need to analyze the events that occurred during care to provide prognostic information. Classic statistical analyses have already been done and provide some information about epidemiology of the heart failure and causes of the failure. This paper presents the use of a probability in a statistical approach in order to profile heart failure in a sample of patients and predict the impact of some events in the care process.

Goals

The main objective of our work was to represent the relationship between variables to determine which variables were the most important for patient survival, with data mining methods. Other goals were firstly to build a decision tree to evaluate which was main factor and its impact on the patient activity in the service or physician activity. Secondly, we wanted to produce a real patient profile to determine the best practice for these patients. One interest of the Bayesian network is the possibility of monitoring the propagation of the impact of an event in the network on other variables or events.

1. http://www.americanheart.org/

Material and Methods

Data collection

Patients suffering from cardiac arrest were managed by the Emergency Medical Service (EMS) of Rennes Medical School in partnership with the Emergency Medical Services of the General Hospitals of the Brittany Region. The resuscitation on the spot conformed with the recommendations of the European Resuscitation Council [2]. This study is a retrospective analysis of data collected from April 1998 to April 2002. During this period, six hundred and seven (607) adults patients who suffered from an out of hospital cardiac arrest were included in our study. Twenty eight (28) patients were excluded because we found a traumatic etiology for the cardiac arrest. Forty six (46) patients were also excluded because the cardiac arrest occurred after the arrival of the emergency service or because some information was missing. For each patient, we collected the following data: age, sex, place of the heart failure, cause, presence of a witness, first aid carried out by the witness, initial cardiac rhythm, time delay between emergency call and arrival of the medical staff, realization of specialized resuscitation, death in the first 24 hours, survival for the first year and Pittsburgh Glasgow score (Cerebral performance categories) during the first year (from 1 for absence of neurological sign to 5 for death).

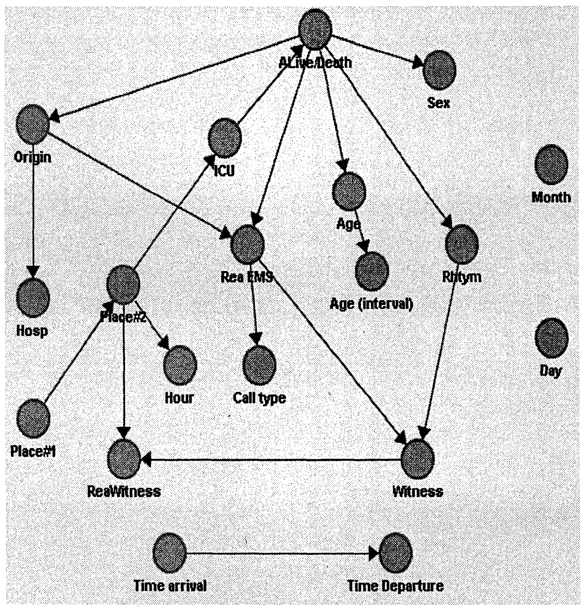


Figure 1 - Bayesian Network (Taboo Order learning method)

Analysis

A classic statistical analysis was done using descriptive method and statistical tests. According to the variables, distribution and the type of the data we used either Anova analysis and Chi^2 tests or Mann and Whitney tests. A logistic regression was performed. The significant threshold was fixed at 5 per cent. A Kaplan Meier method was used for survival analysis.

To complete and confirm the previous results we performed an analysis using the data mining method and we use mainly Bayesian network to model and analyze the data [2, 3, 4].

The Bayes' rules can be described with the following formula:

$$P(R=r \mid e) = \frac{P(e \mid R=r)\, P(R=r)}{P(e)}$$

where P(R = r | e) denotes the probability that random variable R has value r given evidence e. In this probability graph model, the nodes represent the variables and the arcs represent dependency between variables. In these models, a simple representation of a complex problem is possible. We also used the hidden Markov model to analyse data [6]. This model allowed us to deduce what the variables directly related with the target node were (this model provides a simplified model with the target node and the relationship with the nodes which represent parents and sons).

Results

Description

The study used 533 patients. The mean age was 63 (± 17) and the data sample comprised 390 (73 %) men and 143 (27 %) women. Heart failure was observed at home for 411 (77 %) of the 533 patients, in a public place for 62 (12 %) patients and on a public highway for 60 (11 %) patients. A witness was present for (58 %) 308 patients and only 47 (9 %) patients had first aid provided by a witness. The average time delay between emergency call and arrival was 16 minutes (2-70).

At the arrival of the emergency services, 335 (63%) patients were in asystole. The main cause supposed to explain the cardiac arrest was a cardiac etiology in 334 (62 %) cases. Finally 424 (80 %) patients received specialized cardiac resuscitation by emergency services.

Prediction and Bayesian networks

We performed the analysis in three steps:

- Step 1: Learning step
- Step 2: Analysis of associations
- Step 3: Inference and prediction

For the learning step, we use the Taboo Order method to build the network (Fig. 1)The graph of the variables showed that the probability of being alive after heart failure is directly associated with five variables: age sex, the initial cardiac rhythm, the origin of heart failure, and the type of specialized resuscitation employed.

By monitoring the main node (Alive/death), we can infer conclusions about the patient profile. Figure 2 shows the patient profile observed after cardiac arrest. Forty nine point one six persent of the patients died, 6.04 % left the hospital and the 44.80 % represent non cardiac arrest or cardiac arrest where resuscitation was not possible.

For each variable directly related with this node, we can see that the rhythm was equal to 1 asystole in 62.32 % of the cases, equal to 2 ventricular tachy arrhythmias FV/TV in 31.03 % and equal to 3 pulseless electrical activity in 6.65 % of the cases. Nineteen point six four percent of the patients were younger than 46 years and 87.37 % of the patients were not hospitalised in intensive care.

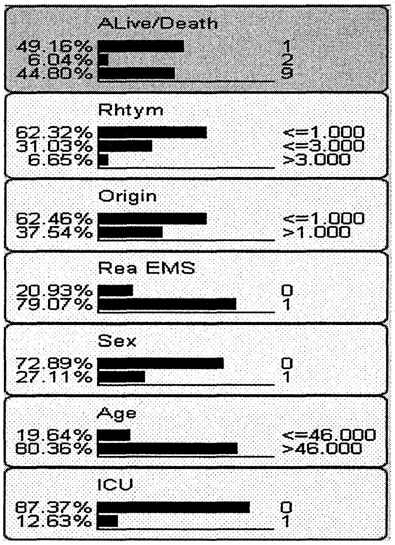


Figure 2 - Data observed

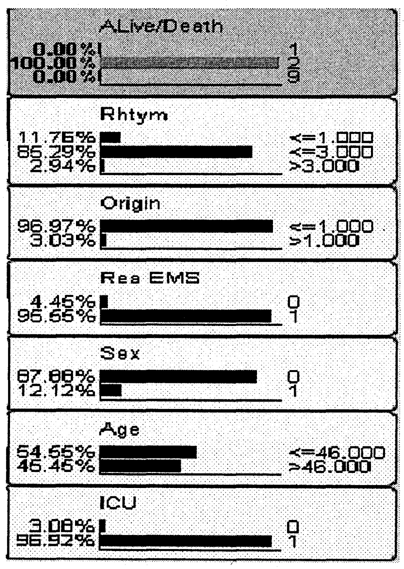


Figure 3 - Alive patient profile

In figure 3, we selected the live status and the propagation of this choice on the other variables can be seen. In this profile, the patients who left the hospital alive suffered from electromechanical dissociation in 85.29 % of the cases and were younger (54.55 % were 46 or less), and 96.92 % profited from hospitalisation in intensive care.

The hidden Markov layer (Fig. 4) demonstrated that one node appears to be highly related to the main node. This node is the 'ReaWitness' node. It appears that where a witness was present and active i.e. the witness helped the patient by realizing cardiac resuscitation, the patient had a better chance of remaining alive.

Forty eight point five percent of the patients who left hospital alive had immediate cardiac resuscitation versus 9.44 % for the entire sample.

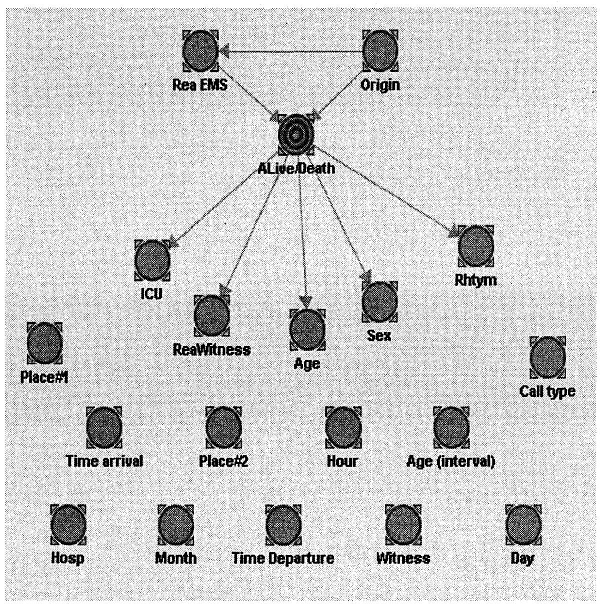
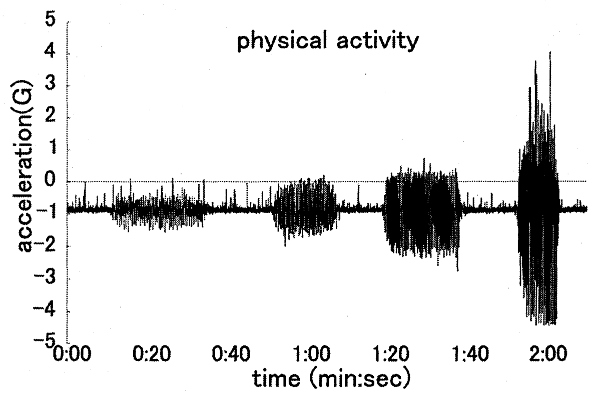


Figure 4 - Hidden Markov Layer

Discussion

The results of the statistical analyses and the probability analyses are equivalent. It appears that heart failure is mainly related with the patient environment, his/her history and the cause that is responsible for the failure. According to others studies in the same field of research, it is highly probable that in-hospital factors are associated with survival from out-of-hospital cardiac arrest [8][9] and it is possible to build a valid cardiac arrest score to evaluate a risk for cardiac arrest victims [10] [11]. Some others articles show the factors that are important to explain long-term survival and quality of life [12] Whereas the classical analysis published by now propose various models to explain what is the best score to evaluate the expected survival or to find at the end of the care process which variables are the most important to explain the survival after cardiac failure. Few of these articles try to evaluate the impact of some factors to others. [11] develops a graphic model that describes survival from sudden out-of-hospital cardiac arrest as a function of time intervals to critical prehospital interventions and it appears that time can become a diagnostic element and a predicting value that can modulate the care. The impact of environment and intangible elements as time, chronology, pathways of patients and relations between these elements can be more easily explored with data mining methods and this point is one of the most important input of these methods.

Since statistical analysis enable the clinician's hypothesis to be tested and this can produce results with significance notions, the mining method allows a complex universe to be untangled. The work required to represent relationships between variables and data was instantaneous with the mining method whereas it was necessary to spend one day of analysis to obtain the same conclusion with the statistical analysis. More interesting was the possibility to see what the relationship was between variables that provided the collected data. Without a preliminary hypothesis or mental construction, the sample of data 'speaks for itself', and illustrates alone which variables influence the others. The inference power offered by this belief network is well accepted by physicians and is comprehensive and allows the expert to test or imagine other relationships or impact of events for a diagnosis or a disease..



In conclusion, it seems that the use of the Bayesian network in medical analysis could be useful to explore data and to find hidden relationships between events or characteristics of the sample. It is a first approach for discussing hypotheses between clinicians and statistical experts. The main limit of these tools is the necessity to have enough data to find regularity in the relationships.

References

[1] A comparison of antiarrhythmic-drug therapy with implantable defibrillators in patients resuscitated from near-fatal ventricular arrhythmias. The Antiarrhythmics versus Implantable Defibrillators (AVID) Investigators. *N Engl J Med.* 1997 Nov 27;337(22):1576-83.

[2] Guidelines 2000 Cardiopulmonary Resuscitation and Emergency Cardiovascular Care: International Consensus on Science. *Circulation* 2000; 102 Supplement I.

[3] Chamberlain D, Cummins RO. Recommended guidelines for uniform reporting of data from out-of-hospital cardiac arrest: the 'Utstein style'. European Resuscitation Council, American Heart Association, Heart and Stroke Foundation of Canada and Australian Resuscitation Council. *Eur J Anaesthesiol.* 1992 May;9(3):245-56.

[4] Cheeseman P, Stutz J. Bayesian classification (autoclass): Theory and results. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R.Uthurusamy, editors, *Advances in Knowledge Discovery and DataMining,* pages 153–180. AAAI/ MIT Press, Cambridge, MA, 1996.

[5] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Machine Learning*, 29:131–163, 1997.

[6] Zoubin Ghahramani, Michael I. Jordan Factorial Hidden Markov Models (1997) *Proc. Conf. Advances in Neural Information Processing Systems*, NIPS.

[7] Le Duff F, Happe A, Burgun A, Levionnois S, Bremond M, Le Beux P. Sharing medical data for patient path analysis with data mining method. *Medinfo. 2001*;10(Pt 2):1364-8.

[8] Skrifvars MB, Pettila V, Rosenberg PH, Castren M. A multiple logistic regression analysis of in-hospital factors related to survival at six months in patients resuscitated from out-of-hospital ventricular fibrillation. *Resuscitation.* 2003 Dec; 59(3): 319-28.

[9] Ghuran A, Reid F, La Rovere MT, Schmidt G, Bigger JT Jr, Camm AJ, Schwartz PJ, Malik M; ATRAMI Investigators. Heart rate turbulence-based predictors of fatal and nonfatal cardiac arrest (The Autonomic Tone and Reflexes After Myocardial Infarction substudy). *Am J Cardiol.* 2002 Jan 15; 89(2): 184-90.

[10]McCullough PA, Thompson RJ, Tobin KJ, Kahn JK, O'Neill WW. Validation of a decision support tool for the evaluation of cardiac arrest victims. *Clin Cardiol.* 1998 Mar; 21(3): 195-200.

[11]Larsen MP, Eisenberg MS, Cummins RO, Hallstrom AP. Predicting survival from out-of-hospital cardiac arrest: a graphic model. *Ann Emerg Med.* 1993 Nov; 22(11): 1652-8.

[12]Bunch TJ, White RD, Gersh BJ, Meverden RA, Hodge DO, Ballman KV, Hammill SC, Shen WK, Packer DL. Long-term outcomes of out-of-hospital cardiac arrest after successful early defibrillation. *N Engl J Med.* 2003 Jun 26; 348(26): 2626-33.

Address for correspondence

Dr Franck Le Duff

Laboratoire d'Informatique Médicale, Faculté de Médecine, Av du Pr Léon Bernard, F-35043 Rennes Cedex.

Franck.leduff@univ-rennes1.fr