Hospital Emergency Department: An Insight by Means of Quantitative Methods

Paolo Cremonesi, Marcello Montefiori* and Marina Resta

Diem – Via *Vivaldi 5, 16126 Genova, Italy*

Abstract: In this work we will examine the activity of the Emergency Department (ED) of an Italian primary Hospital by way of real data. Data will be analyzed both *via* econometrics and data mining (namely: dimensions reduction) models. Our findings demonstrate that using a quantitative exploratory approach to the study of ED data makes it possible to gain suitable information for both the hospital's management and the policymaker, hence contributing to a better understanding of EDs activity and to address its accurate programming. The new approach we suggest is intended to put at decision-maker disposal a set of tools that surfing on the available data make it possible to skim the very relevant information (and hence to reject negligible elements) extracting from the whole set of determinants only those of effective relevance. This, in our opinion, could be a key issue to both verifying the actual performance, and to put forth new policies to improve efficiency and quality as well.

Jel Code: I10; C60.

Keywords: Data mining, econometrics, efficiency, emergency room, principal component analysis.

1. INTRODUCTION

With reference to almost all the developed countries public health care expenditure growth represents one of the most relevant policy problem to cope with. The challenge is to effectively overcome the quality/cost trade off [1-3], *i.e.,* to assure citizens a constant (or even increasing) quality for health services, meeting at once the cost containment goal.

We are firmly persuaded that the solution to this issue has to pass through the efficient management of the Emergency Departments (EDs) activity: EDs, in fact, play a prominent role both in terms of economic resources consumption and in terms of programming; it is a matter of fact that EDs are responsible for a large component of the total amount of patients hospitalization and hospital diagnostic activity [4-6]. Besides, it is evident that the concern among policymakers about both the cost and the activity of ED is increasing, since it could seriously compromise not only the goal of cost containment, but also the citizens' health.

The main issues concerning the Emergency Department relate to the possibility that ED may be either overused or used in an inappropriate way; for example, this situation can occur when a large share of visits are devoted either to non-urgent patients, or to patients with minor medical problems, thus seriously compromising the quality and efficiency of provided services. The challenge for the ED is then to gain in efficiency, while guaranteeing or even improving the quality of the services; i.e., by empowering ED appropriateness in care.

*Address correspondence to this author at the Diem – Via Vivaldi 5, 16126 Genova, Italy; Tel: + 39 010 209 5221; Fax: +39 010 209 5223;
E-mail: montefiori@unige.it, marcello.montefiori@gmail.com

However, despite of the importance of EDs with respect to both the national health care system and the single hospital budget, very little is known and studied about their costs and activity. This is probably due to the peculiarities of hospital's ED, that render very tricky its understanding and analysis, as it is somewhat confirmed by the existing sparse literature.

Most relevant contributions focus on the re-organization of the ED taking into account either the high marginal cost for non-urgent patients [7] or the definition of different criteria to measure and forecast the ED overcrowding [8-11]. Besides, from those research it emerges that overcrowding is viewed as the main responsible for deficiencies in terms both of quality and effectiveness of the treatments provided at the ED [12, 13]. Finally, some studies observed that an effective optimization of the ED activity cannot be implemented disregarding the peculiarities of elderly, i.e., people over 75 that are admitted to ED [14].

Thus our primary aim is to offer a unified approach gathering the instances emerged from the cited contributions and going deepest in the analysis of ED data. To such extent, we will examine the activity of the ED of an Italian primary Regional Hospital by way of observable data. The collected data will be studied using both econometrics and dimensions reduction techniques, in order to provide new insights on the Emergency Department activity, hence grasping useful information for both the hospital's management and the policymaker, and providing a better knowledge of EDs activity to address its accurate programming.

With a look to such aims, what remains of the paper is organized as follows. Section 2 will be devoted to the description of data, while in Section 3 we will use an econometric approach to analyze the dataset we have gathered. Section 4 will discuss an application of Principal

1876-8245/12

Components Analysis (PCA) tailored to provide a better representation of the examined data. Finally, Section 5 will conclude.

2. DATA DESCRIPTION

We investigate a sample of data drawn from the population of patients admitted to the ED of *Ospedali Galliera* of Genova, Italy, throughout the whole 2010[1]. In particular, the sample is represented by the data collected in the ED during the week from Thursday 9th December 2010 to Thursday 16th December 2010; we gathered information for 729 patients out of 1045 that were triaged at the ED: the lack of data for the remaining patients is due to the fact that on the one hand 34 patients abandoned the ED after triaging, but before the first examination, while on the other hand for 282 patients records referring to the variable "visiting time" (a really crucial variable for our analysis, as we are going to illustrate in short) were missing.

The information retrieved for each patient is a bunch of records pertaining both the patient himself and his clinical pathway. In particular the ED computer system provided a huge number of information for each patient and gathered them into a file. The variable of interest in our analysis are given below.

(i) *Date and time of arrival;*

(ii) *Medical attendant* (that is the identification code of the accepting medical staff);

(iii) *Triage entrance code;*

(iv) *Patient'spersonal information* (in particular: gender and date of birth);

(v) *Date and hour of first visit;*

(vi) *Number of Laboratory and non-laboratory prescriptions;*

(vii) *Patient outcome;*

(viii) *Attending Physician;*

(ix) *Date and hour of discharging* (it refers to the patient report closing time).

In addition, a self-reporting data collection has been implemented in order to assess the time each physician devoted to each patient for his care: we will discuss it in deeper detail in next Section 3. The motivation inside this task may be found on the crucial role of such information to attribute the medical cost component to each patient in a proper way.

In our analysis we match up patient's clinical pathway data with those of accountancy type belonging to the ED balance sheet, arranged by the cost-centered criterion.

The data set we have built combines economic and clinic information, since on the one hand it includes details that refer to the patient's pathway within the ED, and on the other hand it points to costs, i.e. to scores that pertains the ED balance sheet. As a result, we got a quite complex database, spanning over several aspects of the patients' cost and path within the hospital. This complies to our declared aim to provide the decision maker with new and useful information, thus making it possible for him to verify (in a positive perspective) the actual level of performance and put forth suitable policies (in a normative perspective) intended to improve both efficiency and quality of the services provided.

One could straightforward argue that one week of observation is a too short period to draw robust conclusions about the costs of the emergency department. However, basic statistics computed on our sample suggested that in spite of the short time horizon, the results are consistent with respect to the population it is extracted from (i.e., all patients treated by the ED in 2010)[2].

3. THE ECONOMETRIC MODEL

Our study combines a mix of consolidated econometric techniques with data mining tools. Such a combination has its rationale in the goals we are pursuing.

As anticipated in the previous section, during the recording sample week, we collected data of 729 patients. One of the most important records we collected is the *visiting time*. By this variable we mean the time spent by each physician in treating the patient. In essence, it measures the total amount of time devoted to each patient by the physician providing him cares, which includes: the time spent for visiting, reading reports, speaking with the patients or his relatives, etc.

As first result we provided evidence of a neat variability among physicians, as shown in Fig. (**1**).

Fig. (**1**) presents a box-plot representation of the *visiting time* variable. The x-axis is labeled with the ID identifying single physicians (numbered from 1 to 19), while on y-axis the "visiting time" distribution per physician is reported (expressed in minutes). Following the standard box-plot representation, grey colored boxes include the "central" 50% of physicians sample distribution (i.e., patients encompassed between the 25^{th} and the 75^{th} percentile). The dots outside the linking line represent outlier patients for the sample distribution.

To make the explanation clearer, let us consider for instance, Physician 1 (P01): the reader can note that the outlier patient is on the twenty minutes visiting time line; conversely, if we move to Physician 12 (P12) the outlier patient corresponds to a visiting time of 60 minutes. Besides, if we take a look to Physician 6 (P06) we observe that the outlier patient is associated to a very short visiting time (below five minutes). A glimpse into Physician 7 (P07) informs us that the whole set of patients examined by P07, always take him an amount of time very close to his 50% percentile. What clearly emerges from the graphical representation is then a marked variability *among physician's behaviors*. This could be a devious information for policy maker. We needed to test whether such variability was somewhat endogenous or rather it was induced by the individual bias, namely by physicians subjective differences in registering the time span. To do this data have been

[1]E.O. OspedaliGalliera is an Italian primary Regional Hospital and its ED treated in 2010 more than 54000 patients. Looking at the number of visits provided by the Ed of E.O. OspedaliGalliera in 2010, it turns out that it is the third most important ED in the Liguria district.

[2]The significance of the sample clearly emerges by comparing the distribution of gender, triage code, age frequencies and outcome with that referring to the year 2010 as a whole. See Appendix 1 for further details.

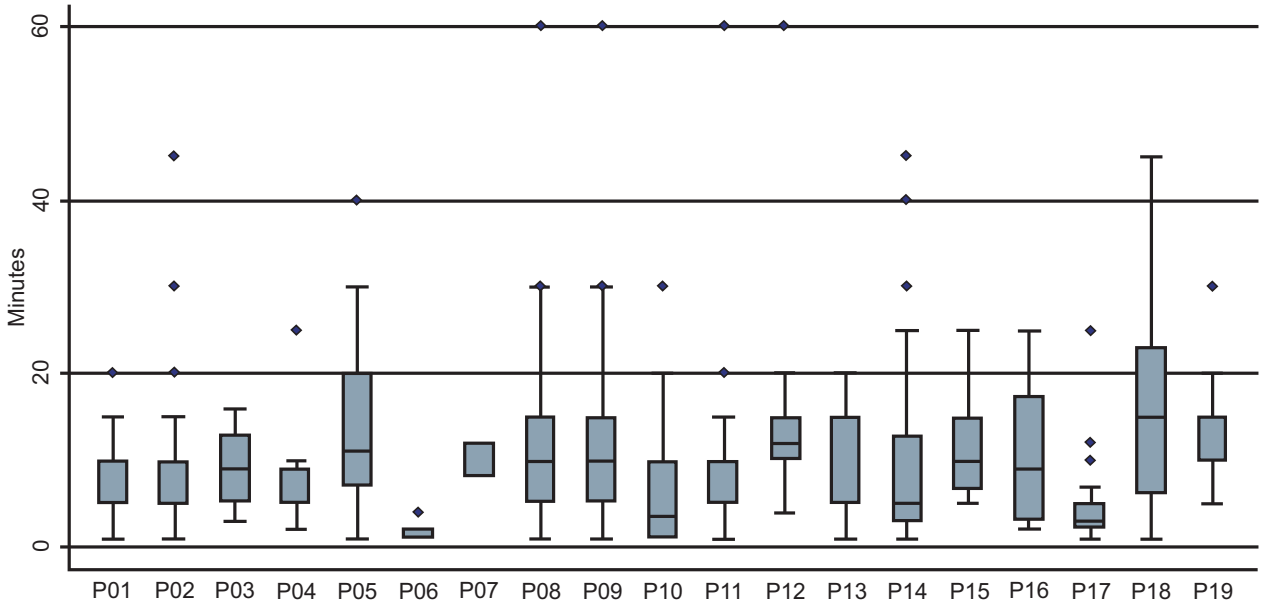


Fig. (1). Box plot representing the "visiting time" per physician.

rescaled per each physician in the range [0, 1] according to the well known formula:

$$rv_{k,i} = \frac{v_{k,i} - t_k}{T_k - t_k}$$

where:

rv_k, is the i-th rescaled value for the k-th physician;

v_k is the original i-th visiting time value for the k-th physician;

t_k is the minimum value of visiting time for the k-th physician;

T_k is the maximum value of visiting time for the k-th physician.

This procedure was necessary to avoid the risk of a constant subjective over/under-estimation of the values. The rescaled data allowed to compensate for such kind of bias. By this procedure we were able to map the time devoted by all physicians to patients within the same range, making them more suitable for comparisons. The box plot of the "rescaled visiting time" is provided in Fig. (**2**).

The rescaled data now exhibit a reduced variability among physicians; this is a direct consequence of the fact that the rescaled values have curbed the "subjective" differences in data collection that might have occurred. In this spirit, the "anomalous" behaviours of P06 and P07 can now be easily intepreted: P06 patients sample distribution, for instance, is fully contained below the 40% line: probably P06 spent very similar amounts of time to visit his patients; whereas it now clearly emerges that P07 devoted a very different amount of time to the patients he examined.

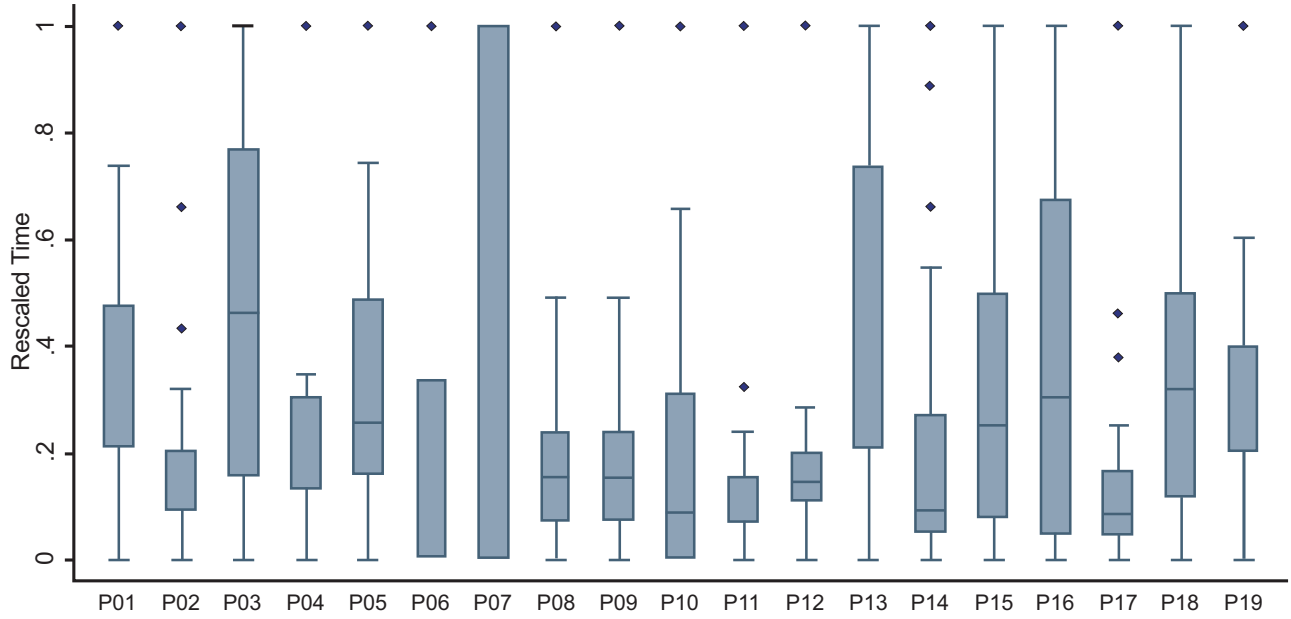


Fig. (2). Box plot representing the "rescaled visiting time" per physician.

As second stage, we moved to develop an econometric model to investigate the main drivers of the amount of time physicians spent for their patients. Although we believe that the model we are going to introduce and explain can be useful both in a positive perspective (allowing to highlight the determinants of the visiting time, as well as the way they affect such dependent variable) and in a predictive wise (providing accurate estimations for the visiting time values that were missed), however we will limit our attention to the former aspect, which is the one of main interest to what this study concerns.

The understanding of the determinants of the time physicians spend with their patients is fundamental for at least two reasons related to efficiency and cost. Proper management of the ED personnel assures human resources to be efficiently arranged, and hence to improve quality (for example by a waiting time containment with reference to urgency and emergency cases.) The latter reason moves from the observation that the personnel cost (medical doctors, nurses and others) represents almost the 70% of total cost as reported in the ED balance sheet. Any reorganization aiming at cost containment has to take this element in duly consideration.

For each patient we find out the following model to hold:

$$RV_T = \alpha + \beta_G G_r + \beta^e_{Age} Age + \beta_{D75} D_{75} + \beta_{OC} Oc + \beta_{DG} D_G + \\ + \beta_{DY} D_Y + \beta_{DR} D_R + \beta_H H_r + \beta_{FT} FT_r + \beta_{Pnl} Pnl + \beta_{Pl} Pl$$

where:

RV_T is the rescaled visiting time, i.e. the dependent variable.

α is a constant term.

Age is the age of each patient at the time of his admission to ED.

β_{75} represents an elderly dummy which is intended to capture the role (if any) played by elderly in affecting the time spent by physicians to visit elders. This variable assumes value equal to 1 if the patient's age is greater (or equal) than 75, 0 otherwise.

D_G; D_Y; D_R are, respectively, dummies variable for green, yellow and red triage codes. By these variables we intend to catch differences depending on the urgency of the treatment, assuming the urgency to be a good proxy of the patient severity.

Pnl and *Pl* are, respectively, the number of non-laboratory and the number of laboratory prescriptions

G_r is the patient gender dummy (r=1 male patient, r=0, female patient).

Oc is the number of patient waiting contemporaneously at ED.

H_r is the *hospitalization* dummy (r=1 for Hospitalized patient; r=0, for Not Hospitalized patient).

FT_k is a *fast track* dummy (*k*=1 for fast tracked patients; *k=0* otherwise). The *fast track* is a particular service that aims to decrease the waiting time in the ED. In case of well defined acute diseases, the nurse at the acceptance desk of the ED will immediately address the patient to the proper hospital's ward, so that the patient will quickly receive the examinations required by his condition. After the visit, the patient will come back to the ED carrying on the examinations ready to be analyzed by an ED's physician.

Table 1. Ols and Tobit Econometric Analysis

Dependent Variable: *Rescaled Reported Time*				
Coefficients	**A-OLS**		**B-TOBIT**	
G	-.0017234		-.0048197	
	(.0136296)		(.0152568)	
Age	.0002644		.0004832	
	(.0005731)		(.000641)	
D_{75}	.0049009		-.0108496	
	(.0281026)		(.0312954)	
OC	-.0010873	*	-.0014501	**
	(.0006424)		(.0007268)	
D_G	.0487668	*	.1072558	***
	(.0264416)		(.0329632)	
D_Y	.1140928	***	.1751758	***
	.0343155		(.0406823)	
D_R	.1819895	***	.2381251	***
	(.055602)		(.0628189)	
H	.0026254		.0036416	
	(.0219041)		(.0245007)	
FT	-.0765695	***	-.1765068	***
	(.0282432)		(.0366663)	
P_{nl}	.0313545	***	.0336321	***
	(.0038439)		(.0042259)	
P_l	.0100285	***	.0106718	***
	(.0013329)		(.0014683)	
constant	.0284809		-.0483844	
	(.0369169)		(.0435827)	
	F(11, 717) = 64.31		LR chi2(11) = 412.26	
	Prob> F = 0.0000		Prob> chi2 = 0.0000	
	R-squared = 0.4097		Pseudo R2 = 1.3669	
Number of obs = 729				
Tobit regression obs. Summary:				
			100 left-cens. obs. at rescaled_all<=0	
			629 uncensored observations	
			0 right-censored observations	
Significance levels: 1%***, 5%**, 10%*				

The coefficients model have been estimated in two different ways, using both Ordinary Least Squares (OLS) and Tobit estimators. The use of Tobit estimator is therein motivated to incorporate into the model a censored dependent variable. In our dataset, in fact, the dependent

variable (the self reported visiting time) is censored (i.e. lies within the range 0-1) by construction.

The econometric analysis pointed on very interesting and not trivial information: conversely to what one might expect, the elderly condition does not seem to affect the time required for treatment at ED and, as a consequence, the cost is also not driven by age. Table **1**, in fact, shows that the coefficients of the variables *Age* and D_{75} (dummy for elderly) are not statistically significant in both analyses. This result suggests that they do not affect the time the physicians devote to the treatment of patients. The same reasoning applies to the *G* (gender) variable, coherently, this time, with what one may expect.

The variable *OC (overcrowding)* was intended to capture a sort of *pressure* that physicians might be subject to when the number of people in the waiting room of the ED increases. This variable is relevant at 10% and 5% level of significance respectively in the OLS and Tobit estimation. In both cases it shows a negative sign, and its interpretation is trivial: the larger the number of people awaiting for care at the same time inside the ED, the lower is the time spent for patients by physicians.

Besides, in our opinion, very interesting information can be grasped from the analysis of the dummy variables referring to the triage color codes. All those variables seem to affect the dependent variable. In other words, the time devoted by physicians to patients with different color codes exhibits a statistically relevant difference: the positive sign associated to the coefficients of triage code dummies means that on average, ceteris paribus, the visiting time increases with severity (which is correlated to the triage code) of the patient.

Another interesting issue is related to hospitalized patients (*H* variable). When patients are hospitalized their ER "time for care" is shortened and their clinical pathway moves to other hospital's wards. We can then claim that hospitalization allows somehow diminishing the pressure on the ED. Similarly to the *H* variable, the *fast track*(*FT*) variable also allows for a very prominent time saving. The explanation for this result is straightforward and rooted in the fast track procedure itself.

To conclude our analysis an attentive consideration is deserved by the variables representing the number of laboratory and non laboratory prescriptions: P_{nl} and P_l. P_{nl} and P_l variables are strongly correlated with the dependent variable RV_T; this result seems to suggest that the larger the number of (lab/ non lab) prescriptions is, the more the clinical condition of the patient (and as a consequence, the time required for diagnosis and treatment) becomes "tricky". We might expect that when a large number of prescriptions is required, the patient cost jack up, since personnel and examinations cost are the two most relevant items of cost.

4. SOME INSIGHTS WITH DATA MINING AND DIMENSION REDUCTION TECHNIQUES

As already said in the introduction ED represents a relatively unexplored field, despite of its crucial role for the assessment of both hospital's efficiency and cost management. We then used an exploratory approach helping us to provide further insights on how ED variables work together.

This holding, the role of data mining [15, 16] is to let (in a broad sense) the data to speak for themselves. Data mining, however, is a general and wider branch of data analysis, and its goal may be achieved in many different ways. Our way will consist in studying data *via* dimension reduction techniques.

Dimension reduction is nothing but a technique of mapping data to a lower dimensional space, to both discard uninformative variance in the data, and detects the subspace in which the data are effectively embedded into.

Methods assessing dimensionality reduction can be divided into two broad categories: the ones that rely on projections (like Principal Components Analysis [17, 18]) and those attempting to model the manifold on which the data lies (such as Self Organizing Maps [19]). Due to the declared aim of this paper, i.e. an explorative approach, we will focus only on the first group, giving particular attention to Principal Components Analysis whose main features will be described in Appendix 2.

Joining together both the clinical path variables as discussed in Section 3, and the cost components that can be read from the balance sheet[3], we were able to manage a 729*x*27 matrix M, each row reporting the complete patient characterization. The complete list of examined 27 variables is reported in Appendix 3.

We then performed Principal Components Analysis (PCA) on M, in order to find what types of drivers has the major impact on costs resulting from the hospital's financial statement.

The outcomes of the analysis are given in Fig. (**3**).

Fig. (**3**) may be interpreted in the following way: each point represents a patient or, better, his projection from the original 27-dimensions space into a 5-dimensions space: the axes appearing in the figure shows values for the five main directions towards which patients are driven, i.e. in decreasing order of importance: the Triage Code (CTr), the Number of non lab (NrNLP) and laboratory prescriptions (NrLP), the related cost of this latter (CLP), and the Operative Cost (OPC), intended as the cost of medical and nursing activity.

This analysis served as the starting point for a deeper investigation into the relations existing among the aforementioned drivers.

In particular, some interesting information was provided by the scatter diagrams obtained by coupling the five examined determinants, whose results are provided in Figs. (**4-6**).

As it is known, scatter plots allow the user to display points using as two variables coordinate, one on the x-axis, and the second on the y-axis. In our case the scattering procedure has been somewhat enforced, since each point (i.e. the patient) has been maintained with the corresponding triage code and colored accordingly.

[3]Data used in our analysis have been estimated by activity based costing methodology [20]

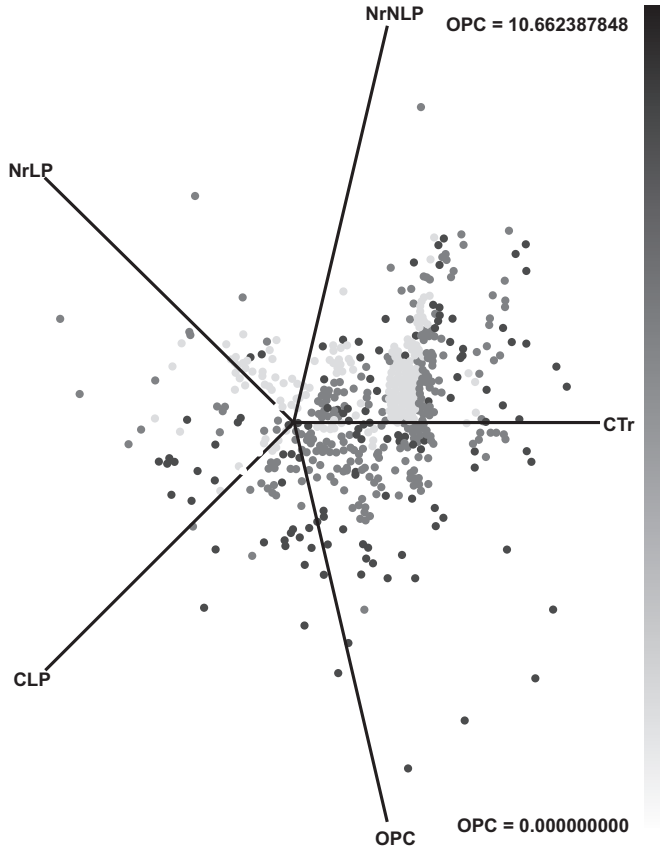


Fig. (3). Patients PCA projection. Each point represents a patient, while the main drivers of the patients status are those appearing along the axis.

Looking to Fig. (**4**), the scatter diagrams oppose the triage codes *vs* the remaining determinants of patient's status highlighted by PCA. With a view to the way points tend to aggregate, one could note that, in general, both green (1) and yellow (2) Triage Codes impact stronger that the remaining triage codes, i.e., white (0) and red (3). In other words, green and yellow codes patients generally are more requesting in terms of both non lab and lab prescriptions, as well as for what concerns medical and nursing efforts, and therefore they result costly with respect to the other triage codes individuals.

We then searched for evidences in such sense by opposing the other drivers one to each other, still using the same scattering procedure, as one can see by looking at Figs. (**5**, **6**). In particular, Fig. (**5**) shows the scatter diagrams that oppose the Number of Non Laboratory Prescriptions (NrNLP) to the Number of Laboratory Prescriptions (NrLP) and to Operative Costs (OP) respectively, while in Fig. (**6**) the Number of Laboratory Prescriptions (NrLP) are plotted *vs* Operative Costs (OP).

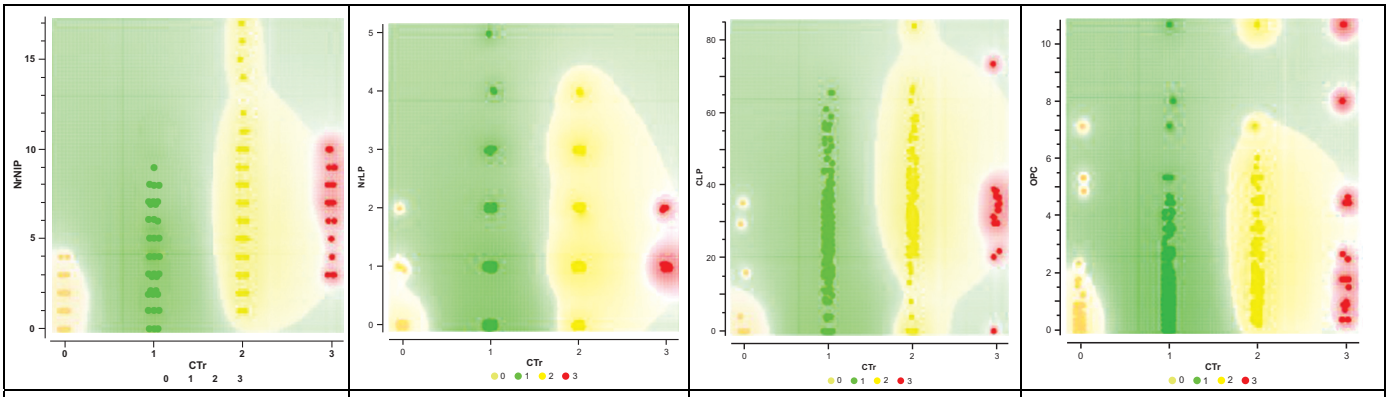
If we look at the way points aggregate in both Figs. (**5**, **6**), our first impression is once again confirmed, because we discover that the impact (in terms of cost) of red codes is overtaken by that of both green and yellow codes.

A striking example of what we claim is offered by Fig. (**5b**): opposing the number of non Lab prescriptions to the Operative Costs yields in a number of very costly yellow and green codes patients whose number exceeds that of red code patients. Similar remarks hold also if we refer to Fig. (**6**), where the Nr of Lab Prescription is again opposed to Operative Costs.

CONCLUDING REMARKS

It is a matter of fact that managing the efficiency of Emergency Departments (ED) is a crucial issue that concerns not only the single hospital, but also the national health system as a whole. The present scenario of reduced resources forces both the social decision maker and the hospital management to closely scrutinize all the available information, in order to apply cost containment policy and to improve the quality of the services provided. The information available for each patient is quite complex, and spans over several aspects (from triaging time to either laboratory or non laboratory prescriptions, just to cite some of them).

Within the aforementioned scenario we believe that new approaches can allow for a better understanding of the ED activity, both from a clinical and economical perspective. With this in mind, we moved towards two distinct but related directions. First of all we suggested an econometric procedure that works on the visiting time (VT), one of the most crucial variables to assess both the efficiency and costs containment of EDs. In this way we were able to stress the



(**a**) (**b**) (**c**) (**d**)

Fig. (4). Scatter diagrams of the Triage Code (CTr) opposed to Nr of Non Lab prescriptions (**a**), Nr of Lab Prescription (**b**), Cost of Lab Prescription (**c**), and Operative Costs (**d**), respectively. Numbers on the x-axis stand for proper triage codes, as also explained by the legend.

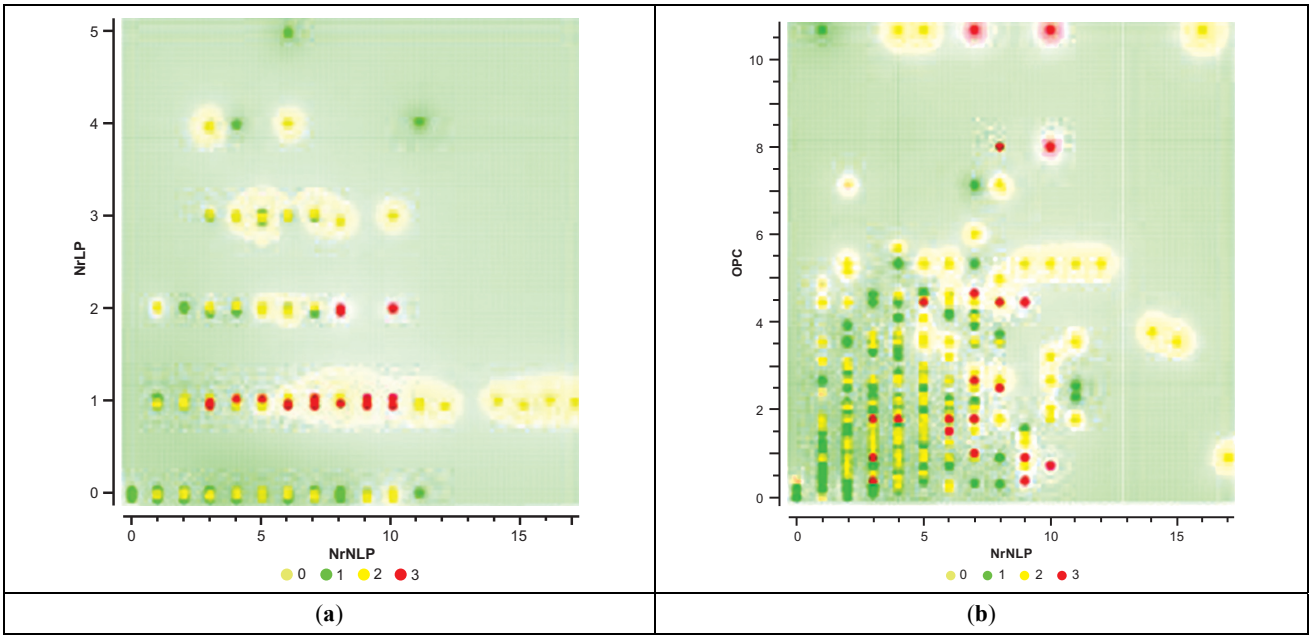


Fig. (5). Scatter diagrams of the Nr of Non Lab prescriptions opposed to Nr of Lab Prescription (**a**), and Operative Costs (**b**).

main determinants of VT. In addition, we provided an exploratory data analysis approach to ER data that have been examined by means of dimensions reduction techniques, namely: by means of Principal Components Analysis (PCA).

By the econometric analysis and the data mining approach it has been possible to grasp information not available otherwise, providing the decision maker with new and useful information that make him possible to verify (in a positive perspective) the actual level of performance, and to put forth suitable policies (in a normative perspective), in order to improve both efficiency and quality for health services.

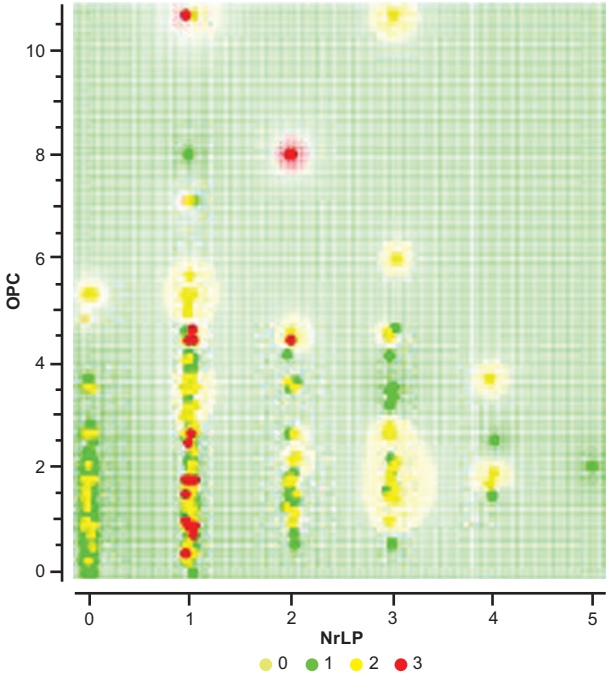


Fig. (6). Scatter diagrams of the Nr of Lab prescriptions opposed to Operative Costs.

Besides, the econometric analysis has shown the impact of a number of variables in affecting the time devoted to patients by physicians. This analysis is then also relevant to assess the ED activity and cost composition. In fact, looking at the ED balance sheet it emerges that the cost for physicians and other medical staff prevails on every other items. In particular, medical staff costs represent approximately 70% of the overall cost ascribed to ED in its balance sheet. To this extent the patient visiting time is extremely relevant information to be investigated.

For what it concerns PCA, we have shown how a relatively simple technique may be of help to extract knowledge from data. In particular, PCA made us possible to observe that despite the huge number of determinants that can be recorded for each patient, only a reduced subset of them is really important to determine the total cost of the patient itself. Using this methodology it has been possible to verify the variability "between and within" the different types of patients, as well as the main determinants of their costs. Besides, the possibility to manage a procedure whose results may be easily reported *via* an intuitive graph representation gave us both an additional tool for a better understanding of the obtained results, and a good starting point for future analysis now in a confirmatory sense.

ACKNOWLEDGEMENTS

The authors would like to thank E.O. Ospedali Galliera for the precious cooperation in providing the data.

This work has been funded by the MIUR within the FIRB project N. RBFR08IKSB.

CONFLICT OF INTEREST

Declared none.

APPENDIX 1: SAMPLE WEEK SIGNIFICANCE ANALYSIS

In this section we provide the reader with some statistics referring to the data of the sample week used in our

investigation, in order to confirm its representativeness with respect to the population of data the sample has been drawn from.

To this extent, data belonging to the sample week are hereafter compared with those referring to the whole population of patients which have been triaged by the ED during the year 2010.

As first step (Table **2**) we compared the percentage distribution of triage composition of the two data sets (the sample week opposed to the whole 2010). The sample fits quite well the general trend.

Table 2. Patients Composition by Triage Color

	WHITE	GREEN	YELLOW	RED	TOT
Sample week	8,01%	68,15%	21,66%	2,18%	100,00%
Whole 2010	8,68%	72,16%	17,21%	1,96%	100,00%

Looking at gender composition, Table **3** highlights homogeneity in distribution between the two cases with reference to all triage codes except the white. The data of the sample week, in fact, show for white code patients a neat prevalence of males (64%) with respect to females (36%). On the opposite, the data referring to the whole population of white codes present a distribution aligned with that of the other colors. However, since white codes represent the 8% of the total number of ED accessed, which does not affect the results obtained by our study.

Table 3. Gender Composition by Triage Code

Gender Composition		Triage Classification			
		WHITE	GREEN	YELLOW	RED
Sample week	**M**	64,20%	49,64%	52,05%	45,45%
	F	35,80%	50,36%	47,95%	54,55%
	TOT	100,00%	100,00%	100,00%	100,00%
Whole 2010	**M**	48,52%	52,05%	49,81%	44,71%
	F	51,48%	47,95%	50,19%	55,29%
	TOT	100,00%	100,00%	100,00%	100,00%

Looking at Table **4** it clearly emerges that also with reference to the per age distribution the two dataset present strong similarity. The elderly patients (i.e., patients older than 65) are detectable in almost exactly the same percentage both in the sample week and in the whole 2010.

Table 4. Age Composition by Triage Code

	Age by Class	WHITE	GREEN	YELLOW	RED	TOT
Sample Week	**0-14**	1,23%	1,74%	0,46%	0,00%	1,38%
	14-64	87,65%	77,21%	43,84%	18,18%	69,54%
	65+	11,11%	21,04%	55,71%	81,82%	29,08%
	TOT	100,00%	100,00%	100,00%	100,00%	100,00%
Whole 2010	**0-14**	1,14%	1,21%	0,33%	0,00%	1,03%
	14-64	82,28%	76,33%	42,74%	24,05%	70,04%
	65+	16,58%	22,46%	56,93%	75,95%	28,93%
	TOT	100,00%	100,00%	100,00%	100,00%	100,00%

Finally, Table **5** provides information on the outcome associated to each patient. The patients outcome can be: i) the discharge from hospital; ii) the hospitalization in some ward of the same hospital; iii) the transfer to other medical structure; iv) death; v) hospital abandoning (patients might opt to abandon the hospital, because of their own decision, generally due to a very low severity illness); and vi) other outcomes (very infrequent), as expulsion or hospitalization refusal.

By this comparison between the sample week and the population from which it has been drawn, we are empirically confident that the data used in our analysis well represent the phenomenon we investigated and therefore that the results we have found are reliable and they can be applied to the entire population.

APPENDIX 2: PRINCIPAL COMPONENTS ANALYSIS: A MATHEMATICAL SKETCH.

Principal component analysis (PCA) is a technique which uses linear algebra to make data more readable, and to retrieve any available information that is not directly exploitable from them.

Table 5. Outcome Composition (in %) by Triage Code

OUTCOME	%									
	WHITE		GREEN		YELLOW		RED		TOT	
	Week	2010	Week	2010	Week	2010	Week	2010	Week	2010
DISCHARGED	74,07	85,67	84,18	83,03	50,23	48,75	4,55	5,55	**74,28**	**75,85**
HOSPITALIZED	7,41	2,78	10,16	10,87	48,86	47,89	90,91	91,72	**20,08**	**18,11**
TRANSFERRED	1,23	0,11	0,44	0,30	0,46	1,41	4,55	1,95	**0,59**	**0,50**
DECEASED	0,00	0,07	0,00	0,01	0,00	0,01	0,00	0,49	**0,00**	**0,02**
ABANDON	17,28	11,05	4,21	4,84	0,46	1,46	0,00	0,29	**4,35**	**4,71**
AL.	0,00	0,33	1,02	0,95	0,00	0,47	0,00	0,00	**0,69**	**0,80**
TOT	100	100	100	100	100	100	100	100	100	100

Let X is the original input *m x n* matrix. Then:

$$P'X = Y \quad (1)$$

represents a change of basis, where Y is another *m x n* matrix linked to X by a linear transformation made possible through the square (*m x m*) matrix P; in another words, P allows a change of basis thanks to which the row vectors in P will become the principal components of X.

The matrix P may be computed in various ways: basically the available methods focus either on the Eigen values decomposition of the covariance matrix, or in the singular value decomposition of the original data matrix; however, since the paper does not intend to be an essay on PCA, rather focusing on an application of it, here we will use (and briefly discuss) the former procedure, i.e. the covariance method.

This method consists of a few numbers of steps starting from the matrix X performed as follows.

1. Derive from X the matrix B, subtracting X by its mean E[X]: B=X−E,X.
2. Compute the covariance matrixΣ:
$$\Sigma=B,B\text{-}T.$$
3. Compute the matrix V of eigenvectors which diagonalizes the covariance matrix Σ:
$$V^{-1}\Sigma V = D.$$

where:

— D is the diagonal matrix of Eigen values of Σ, with $d_{i,j} = \lambda_i$ for i=j, and $d_{i,j} = 0$, otherwise;

— V, also of dimension $m \times m$, is the Eigen vectors matrix of Σ.

Once steps 1-3 have been performed, the procedure ends by sorting the columns of the eigenvector matrix V and Eigen value matrix D in order of *decreasing* Eigen value. The principal components of X will be then in number k<m, chosen as the earlier k elements on the diagonal of D achieving, on percentage, a reasonably high value of variance explained.

In order to make clear those concepts, we provide an explicative toy example. Let us assume to have the following situation for 4 ER patients.

Table 6. The Matrix of Data for the Toy Example

Triage Code	Patients Age	Nr of Prescriptions	Time Spent for Diagnosis (Minutes)
3	53	2	10
2	53	2	11
3	52	1	9
2	52	1	11

Then, let us imagine we must select only two of the above patients balancing their clinical condition with the potential impact they can have on the ER costs. Following the indication provided for the PCA method, our first step should consist in forming the matrix:

$$X = \begin{bmatrix} 3 & 53 & 2 & 10 \\ 2 & 53 & 2 & 11 \\ 3 & 52 & 1 & 9 \\ 2 & 52 & 1 & 11 \end{bmatrix}$$

The reader could note that each column of X corresponds to a column in Table **6**. Moving to the covariance matrix we should get:

$$\Sigma = \begin{bmatrix} 0.3333 & 0 & 0 & -0.5 \\ 0 & 0.3333 & 0.3333 & 0.1667 \\ 0 & 0.3333 & 0.3333 & 0.1667 \\ -0.5 & 0.1667 & 0.1667 & 0.9167 \end{bmatrix}$$

And hence by applying Step3 formulas:

$$V = \begin{bmatrix} 0 & 0.8346 & -0.3260 & -0.4440 \\ -0.7071 & -0.1340 & -0.6553 & 0.2293 \\ -0.7071 & -0.1340 & -0.6553 & 0.2293 \\ 0 & 0.5172 & 0.1864 & 0.8353 \end{bmatrix}$$

and:

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.0235 & 0 & 0 \\ 0 & 0 & 0.6193 & 0 \\ 0 & 0 & 0 & 1.2739 \end{bmatrix}$$

Now looking at the longest diagonal in matrix D, its components should be taken in the following order: d_{44}, d_{33}, d_{22}, d_{11}. The columns ov V should be ordered accordingly: v_4, v_3, v_2, v_1, way to move along the direction of the maximum explained variance. This in turn, suggests the way to find the direction of the maximum explained variance. Such value can be easily computed as:

$$\frac{100 * d_{ii}}{\sum_{i=1}^{4} d_{ii}}$$

Looking back to our example, the calculation provides the following results: 66.4663; 32.3090; 1.2247; 0. this means that the first principal component aids to explain approximately 66% of the total variance of data; the second component is able to explain an additional 32.30% of variance, and so on. As consequence, in order to project X into a lower dimensional space we can build the matrix P whose columns will be the fourth and third column of V respectively, i.e. those eigenvectors that are associated to the greater percentage of explained variance:

$$P = \begin{bmatrix} -0.444 & -0.3260 \\ 0.2293 & -0.6553 \\ 0.2293 & -0.6553 \\ 0.8353 & 0.1864 \end{bmatrix}$$

and then we can write:

$$P'X = Y$$

thus obtaining:

Table 7. Recorded Variables for Each Patient

NR.	Description	Abbreviation
1	Gender (Sex)	SX
2	Age	A
3	Time between triage and 1st visit	TTD
4	Time between 1st visit and exit	TVD
5	Triage Code	CTr
6	Days of Prognosis	GgPrognosys
7	Diagnosis	Diagnosi
8	Short Observation	OBI
9	Number of Non laboratory Prescriptions	NrNLP
10	Number of laboratory Prescriptions	NrLP
11	Nr of Laboratory Exams	NrEL
12	X-Rays cost	CRX
13	Non laboratory prescriptions Cost	CONLP
14	Laboratory prescriptions Cost	CLP
15	Other Costs	OC
16	Drugs Cost	PC
17	Hospital specific Cost	SPC
18	Medical services Cost	SSC
19	Cleaning Cost	CC
20	Kitchen and Laundry Cost	KLC
21	AdministrativeCost	APC
22	Common Cost	CCQ
23	Physicians Cost	OPC
24	Nursing Costs	OPC
25	Auxiliary Personnel Cost	APC
26	Other graduate personnel Cost	OGC
27	Total Cost	TC

$$\begin{bmatrix} -0.444 & 0.2293 & 0.2293 & 0.8353 \\ -0.3260 & -0.6553 & -0.6553 & 0.1864 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 53 & 2 & 10 \\ 2 & 53 & 2 & 11 \\ 3 & 52 & 1 & 9 \\ 2 & 52 & 1 & 11 \end{bmatrix}$$

$$= \begin{bmatrix} 1.485 & 49.974 & 0.635 & 9.333 \\ -3.882 & -76.396 & -2.432 & -14.316 \end{bmatrix}$$

The matrix on the right hand side of the equality is nothing but the original matrix X projected into a reduced dimension space by the change of basis operated through P.

APPENDIX 3. RECORDED VARIABLES FOR EACH PATIENT

Table **7** shows the variables we have recorded for each patient. Note that Physician personnel cost and Nursing personnel cost (records number 23 and 24 in our list) have been gathered into the variable that we labeled by OPC.

REFERENCES

[1] Chalkley M, Malcomson JM. Contracting for health services with unmonitored quality. Econ J 1998a; 108: 1093-110.

[2] Chalkley M, Malcomson JM. Contracting for health services when patient demand does notreflect quality. J Health Econ 1998b; 17: 1-19.

[3] Ma CA. Health care payment systems: cost and quality incentives. J Econ Manag Strategy 1994; 3: 93-112.

[4] Williams RM. The costs of visits to emergency departments. N Engl J Med 1996; 334: 642-6.

[5] Sartini M, Cremonesi P, Tamagno R, Cristina ML, Orlando P, Simeu Group. Quality in emergency departments: a study on 3,285,440 admissions. J Prev Med Hyg 2007; 48: 17-23.

[6] Cremonesi P, di Bella E, Montefiori M. Cost analysis of emergency department. J Prev Med Hyg 2010; 51(4): 157-63.

[7] Bamezai A, Melnick G, Nawathe A. The cost of an emergency department visit and its relationship to emergency department volume. Ann Emerg Med 2005; 45(5): 483-90.

[8] Hoot N, Aronsky D. An early system for overcrowding in the emergency department. AMIA Annu Symp Proc 2006; p: 339.

[9] Hoot N, Zhou C, Jones I, Aronsky D. Measuring and forecasting emergency department crowding in real time. Ann Emerg Med 2007; 49(6): 747-55.

[10] McCarthy ML, Aronsky D, Jones ID, *et al*. The emergency department occupancy rate: a simple measure of emergency department crowding? Ann Emerg Med 2008; 51(1): 15-24.

[11] McCarthy ML, Zeger SL, Ding R, *et al*. Crowding delays treatment and lengthens emergency. Ann Emerg Med 2009; 54(4): 492-503.

[12] Pines JM, Yealy DM. Advancing the science of emergency department crowding. Measurement and Solutions. Ann Emerg Med 2009; 54(4): 511-3.

[13] Horwitz LI, Green J, Bradley EH. US emergency department performance on wait time and length of visit. Ann Emerg Med 2010; 55(2): 133-41.

[14] Rossille D, Cuggia M, Arnault A, Bouget J, Le Beux P. Managing an emergency department by analysis HIS medical data: a focus on elderly patient clinical pathways. Health Care Manag Sci 2008; 11: 139-46.

[15] Friedman J, Hastie T, Tibshirani R. Elements of statistical learning: prediction, inference and data mining. New York: Springer 2001.

[16] Kantardzic M. Data mining: concepts, models, methods, and algorithms. New York: John Wiley & Sons 2003.

[17] Jackson JE. A user's guide to principal components. New York: John Wiley and Sons 1991.

[18] Jolliffe IT. Principal component analysis. New York: Springer-Verlag 1986.

[19] Kohonen T. Self-organizing maps. New York: Springer 2001.

[20] Cooper R, Kaplan R, Maisel L, Morrissey E, Oehm R. Implementing activity-based cost management: moving from analysis to action. New Jersey: Institute of Management Accountants, Montvale 1992.

Received: January 20, 2012 Revised: February 23, 2012 Accepted: March 3, 2012