STUDIA UBB CHEMIA, LX, 2, Tom I, 2015 (p. 101-108)
(RECOMMENDED CITATION)

PROGNOSTIC FACTORS IN LIVER FAILURE IN CHILDREN BY DISCRIMINANT ANALYSIS OF CLINICAL DATA. A CHEMOMETRIC APPROACH

HORIA F. POP[a], COSTEL SÂRBU[b,*], ANA ŞTEFANESCU[c], AUREL BIZO[c], TUDOR LUCIAN POP[c]

ABSTRACT. Discriminant analysis was applied as an efficient method to identify an objective score concerning liver failure in children using clinical data. Discriminant analysis was not only used for classifying the patients according to the survival status, but also for detecting the most important factors that discriminate between surviving and deceased patients. Based on the considered factors, we were able to compute a complete separation between surviving and deceased patients. The factors responsible for the separation were age, K and total bilirubin (3rd sampling day). The smallest contribution was obtained for aspartate aminotransferase (3rd sampling day), hemoglobin, thrombocytes, albumin. The obtained results confirm that clinical analysis combined with the multidimensional analysis of data gives an interesting and very useful way for correlations, interpretations, problem solving and cost effectiveness.

Keywords: *Chemometrics, clinical data, discriminant analysis, liver failure prognosis*

INTRODUCTION

Liver failure in children could have acute or chronic evolution. Acute Liver Failure (ALF) is a relatively rare but often fatal event in children [1]. Definition of ALF is hepatic necrosis resulting in loss of liver function within few months of the onset of clinical liver disease. Acute liver failure accounts

[a] *Babeş-Bolyai University, Faculty of Mathematics and Computer Science, 1 M. Kogălniceanu St., Cluj-Napoca, RO-400084, Romania*
[b] *Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, 11 Arany Janos St., Cluj-Napoca, RO-400028, Romania*
[c] *Iuliu Haţieganu University of Medicine and Pharmacy, 8 Victor Babeş St., Cluj-Napoca, RO-400012, Romania*
* *Corresponding author: csarbu@chem.ubbcluj.ro*

for 10-15% of liver transplantation in USA [2]. The mortality without treatment, including liver transplantation, in ALF patients is over 70%. There are few patients with spontaneous regeneration of the liver with excellent long-term evolution [1].

It is very important to have an accurate evaluation of the prognosis of the patients with ALF in order to select the patients that will need liver transplantation in order to survive.

There are few reliable criteria for determination of the prognosis in ALF in children. Many attempts have been made to correlate clinical and laboratory data in order to establish the prognosis. Pediatric Acute Liver Failure (PALF) Study Group is a multicenter and multinational consortium created for this kind of studies with the most important results in children with ALF. Overall prognosis in children with ALF is variable with the etiology of the disease, between 68% survival in acute hepatitis A and 12% in drug toxicities [1]. Other parameters analyzed were age, degree of hepatic encephalopathy, severity of coagulopathy, bilirubin level [3]. In a British study of ALF due to acetaminophen intoxication prothrombin time, hypoglycemia, serum creatinine, acidosis and grade III encephalopathy were factors with poor prognosis [4].

Liver transplantation is the ultimate solution for patients with liver failure, acute or acute on chronic liver disease. King's College Criteria (using biochemical and clinical parameters available on admission: age, etiology, pH, prothrombin time, serum bilirubin, serum creatinine, encephalopathy) are widely used for selection of the patient with ALF for liver transplantation but they have not been validated in a large pediatric cohort [5].

The current prognostic score in use for organ allocation and the stratification of the need for liver transplantation is the MELD score (Model for End-Stage Disease), with its pediatric variant, the PELD score. The scores were developed in the early years of this century to improve organ allocation, discarding the “waiting time” on the transplant list as not being a good indicator of medical urgency [6-8]. The MELD score included creatinine, bilirubin and International Normalized Ratio measuring blood coagulability (INR) as parameters in the calculation formula, while the team designing PELD chose albumin, INR, total bilirubin, age (with more points attributed to children under the age of one) and evidence of failure to thrive. It proved to be a more efficient system, with fewer deaths while on the waiting list, despite the still high rate of mortality in the under two years of age group [9]; in this particular study, the change in PELD score proved to be an important predictor of outcome for children on the list. However, after the tempered enthusiasm of the first success, with higher rates of survival [10], criticisms ensued: the scoring system underestimated the near-term risk of death and subjected children to several serious complications while “waiting”, which forced health-care professionals and patients to resort to the “exception” mechanism that enabled the access to graft in spite of lower score. Further studies raised suspicions on the objectivity of the system, claiming inter-laboratory variability in the determination of INR [11], and in 2010 a modified PELD was proposed as a means of correcting and improving its efficacy [12].

Although the access to liver transplantation in children is rather limited in our country, the importance of a better risk stratification for children with liver failure remains unquestionable. Our goal is to analyze the PELD parameters as well as other factors of possible influence on morbidity and survival (sodium, ammonia, creatinine, and lactate) of patients with acute or acute-on-chronic liver failure and propose an alternate scoring system, based on their survival rates. In this order, the discriminant analysis (DA) has successfully been applied.

RESULTS AND DISCUSSION

The computed data set included 49 patients and the following 31 factors (variables or characteristics): age, leucocytes, hemoglobin, thrombocytes, C reactive protein, Aspartate aminotransferase (1st sampling day, AST_1), Alanine aminotransferase (1st sampling day, ALT_1), Aspartate aminotransferase (2nd sampling day, AST_2), Alanine aminotransferase (2nd sampling day, ALT_2), Aspartate aminotransferase (3rd sampling day, AST_3), Alanine aminotransferase (3rd sampling day, ALT_3), total bilirubin (1st sampling day, TB_1), Direct bilirubin (1st sampling day, DB_1), Total bilirubin (2nd sampling day, TB_2), Total bilirubin (3rd sampling day, TB_3), protein, albumin, pediatric end-stage liver disease (PELD), International Normalized Ratio measuring blood coagulability, value on the first day (INR_1), Na, K, glycaemia, urea, creatinine, worst creatinine level (V52_A), QT level on hospitalization day 1 (TQ_1), QT level on hospitalization day 2 (TQ_2), QT level on hospitalization day 3 (TQ_3), International Normalized Ratio measuring blood coagulability (INR), worst prothrombin index level (IP). We have to mention that the majority of measured variables (age, leucocytes, hemoglobin, thrombocytes, protein, albumin, Na, K) have a normal distribution according to the Kolmogorov-Smirnov statistical test. The chemometric analysis has been performed by using Statistica 7.1 software (StatSoft, Inc., Tulsa, USA).

After application of the stepwise DA to the matrix data (49 x 31) the factors (variables) presented in Table 1 were retained in the model. The statistics from this table illustrates the contribution to the patients discrimination of the considered factors (clinical data) according to different parameters.

Knowing that Wilks' lambda (λ^*) describes the unique contribution of each variable to the discriminatory power of the model (the smaller the value of λ^*, the more the model is discriminating and the larger the lambda λ^*, the more likely it is significant) and large values for F and close to 0 for λ^* shows that the variable has a significant contribution, the following statements may be retained.

It is easy to observe that the greatest contribution is given by age ($\lambda^* = 0.831$; $F = 6.524$). The next highest are K ($\lambda^* = 0.834$; $F = 6.352$) and BT_3 ($\lambda^* = 0.0843$; $F = 5.958$). The smallest contribution was obtained for AST_3 ($\lambda^* = 1.000$; $F = 0.002$), hemoglobin ($\lambda^* = 0.993$; $F = 0.223$), thrombocytes ($\lambda^* = 0.984$; $F = 0.509$), albumin ($\lambda^* = 0.981$; $F = 6.33$). Also a small contribution brings the protein ($\lambda^* = 0.958$; $F = 1.387$).

We performed also a canonical correlation analysis that determined the successive functions and canonical roots (the term root refers to the eigenvalues that are associated with the respective canonical function). The maximum number of functions will be equal to the number of groups minus one, or the number of variables in the analysis, whichever is smaller.

The corresponding standardized canonical discriminant function coefficients (c) corresponding to the single eigenvalue (4.126) are also showed in Table 1.

Table 1. Statistic results concerning discriminant analysis of clinical data

Factor in the model	Wilks' λ	λ*	F	p-level	c	r
INR	0.208	0.936	2.195	0.148	0.366	0.294
Albumin	0.199	0.981	0.633	0.432	-0.227	-0.257
AST_3	0.195	1.000	0.002	0.963	0.018	0.271
K	0.234	0.834	6.352	0.017	0.561	0.285
Hemoglobin	0.196	0.993	0.223	0.640	-0.183	-0.132
Age	0.235	0.831	6.524	0.016	-0.697	-0.174
Thrombocytes	0.198	0.984	0.509	0.481	-0.207	-0.107
V52_A	0.222	0.878	4.453	0.043	0.612	0.141
BT_3	0.231	0.843	5.958	0.020	0.903	0.160
Leukocytes	0.221	0.882	4.298	0.046	-0.737	0.004
ALT_3	0.205	0.950	1.682	0.204	0.539	0.205
TQ_3	0.208	0.937	2.145	0.153	0.328	0.224
Urea	0.215	0.908	3.224	0.082	-0.585	0.084
AST_1	0.222	0.878	4.457	0.043	-0.685	0.096
AST_2	0.214	0.911	3.127	0.087	0.569	0.251
Protein	0.204	0.958	1.387	0.248	-0.289	-0.216

As we know the higher the discriminant coefficient (absolute value) and the closer the correlation coefficient (r) is to 1 respectively, the more the variable importance for the separation of patients in defined groups**.**

The highest standardized discriminant coefficients correspond to BT_3 (0.903), leukocyte (0.737), age (0.697), AST_1 (0.685), K (0.561).

A common result that one looks at in order to determine how well the current classification functions predict group membership of cases is the classification matrix.

The classification matrix shows the number of cases that were correctly classified (on the diagonal of the matrix) and those that were misclassified.

The classification matrix presented in Table 2 indicates a complete separation of patients in a good agreement to their behavior.

Table 2. Matrix classification of patients

Group	Percent Correct	Yes p=.5102	No p=.4898
Yes	100.00	25	0
No	100.00	0	24
Total	100.00	25	24

We can also visualize how the functions discriminate between groups by plotting the individual scores for the discriminant function. The Figure 1 supports the excellent separation of the patients and their (dis)similarities according to scores obtained as linear combinations of the variables (factors) retained in the model.

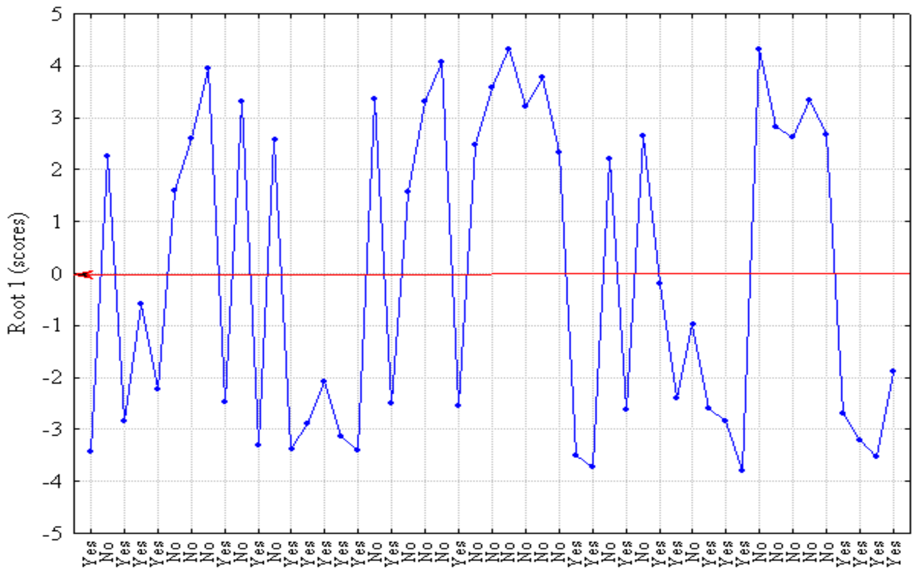

Figure 1. Graphic representation of scores corresponding to the 49 analyzed patients

CONCLUSIONS

Generally, Discriminant Analysis is a very useful tool for (1) detecting the variables that allow the researcher to discriminate between different (naturally occurring) groups, and (2) classifying cases into different groups with a better than chance accuracy**.**

Discriminant analysis was used not only for classifying the patients according to their disease but also for detecting the most important factors (variables) that discriminate between the groups.

This study illustrates two features of DA: the ability to determine the factor making the most important contribution to the difference between the two groups, and the ability to make an important contribution in the clinical setting. When support for a positive or negative diagnosis is required, DA may be able to provide such a binary decision, based on the multiple factors already available.

Clinical analysis combined with the multidimensional interpretation of data gives an interesting and very useful way of disease correlations, interpretations, problem solving and cost effectiveness.

DISCRIMINANT ANALYSIS

Discriminant Analysis (DA) was introduced and discussed in 1936 by R.A. Fisher [13], as a supervised classification method with large applications even today.

The method's main purpose is to predict class membership from a set of predictor variables by creating a function to produce the maximum between-group variance and the minimum intra-group variance. The predictor variables are related to these classes and the constructed memberships are then compared to the groups memberships indicated *a priori* by the user. This enables the user to test the grouping validity based on actual data, to test the created groups, or to assign groups membership to objects.

DA assumes the calculation of linear discriminant functions of independent variables starting from a qualitative dependent variable and two or more quantitative independent variables [14-17]. This is a parametric method, which means that it is based on certain statistical assumptions. The equality between the variance-covariance matrices of the groups to be separated and normal distribution of data are two of requirements for optimal application of DA. However, the difficulties due to the unfavourable statistical characteristics only influence the boundaries and therefore the classification by DA, but not the determination of the discriminant functions [18].

The contribution of the independent variables to the discrimination of groups can be appreciated either by the assay of the classes homogeneity using statistics F, like in the case of analysis of variance method, or by using Wilks' lambda for each variable. Wilks' lambda is the standard statistics used to express the significance of the overall discriminatory power of the variables in the model, where the value 1.0 indicates no discrimination power, while the value 0 indicates a perfect discrimination power. The partial Wilks' lambda describes the unique contribution of each variable to the model's discrimination power. The closer the partial lambda is to 0, the better the discrimination force of the variable is. In addition, the tolerance value illustrates the redundancy of the variable in the model. It is defined as the proportion of the variance contributed by respective variable, and is computed as 1 minus R-square of the respective variable, with all other variables included in the model. If the variable is completely redundant, the squared tolerance value approaches zero.

This information can also be obtained from the discriminant coefficients associated to the descriptive variables, and from the correlation coefficients between the descriptive variables and the scores. The higher the discriminant coefficient is in absolute value, and the closer the correlation coefficient is to one, the higher the variable importance for the cases separation into groups is. As well, the standardized discriminant coefficients, like, for example, the beta weights in regression methods, are used to asses the relative classification importance of the independent variables.

Multivariate analysis methods, including DA, have successfully been applied for the prognosis of liver failure and liver diseases using different clinical data [19, 20].

ACKNOWLEDGMENTS

This work was supported by a grant of the Romanian National Authority for Scientific Research, CNDIUEFISCDI, project number PN-II-PT-PCCA-2011-3.2-0917.

REFERENCES

1. P.F. Whitington, E.M. Alonso, R.H. Squires, *Acute liver failure.* In: Kelly D. (editor). *Diseases of the liver and biliary system in children*, Third Ed., Wiley-Blackwell, Oxford, **2008**, pp.169-188.
2. E.M. Alonso, R.H. Squires, P.F. Whitington, *Acute liver failure in children*. In: F.J. Suchy, R.J. Sokol, W.F. Balistreri WF (editors). *Liver disease in children*. Trd. Ed., Cambridge University Press, New York, **2007**, pp. 71-96.
3. R.H. Jr. Squires, B.L. Shneider, J. Bucuvalas, *J. Pediatr.*, **2006**, *148*, 652.

4. S.B.K. Mahadevan, P.J. McKiernan, P. Davies, D.A. Kelly, *Arch. Dis. Child*., **2006**, *91*, 598.
5. J.G. O'Grady, G.J.M. Alexander, K.M. Hayllar, *Gastroenterol*., **1989**, *97*, 439.
6. R.H. Wiesner, S.V. McDiarmid, P.S. Kamath, *Liver Transpl*., **2001**, *7*, 567.
7. R.B. Jr. Freeman, R.H. Wiesner, A.M. Harper, *Liver Transpl*., **2002**, *8*, 851.
8. R.B. Jr. Freeman, *Transplantation Proceedings*, **2003**, *35*, 2425.
9. R.B. Jr. Freeman, R.H. Wiesner, J.P. Roberts, *Am. J. Transplant*., **2004**, *9,* 114.
10. S.V. McDiarmid, R.M. Merion, D.M. Dykstra, A.M. Harper, *Liver Transpl*., **2004**, *10*, S23.
11. R. Arjal, J.F. Trotter, *Clin. Liver Dis*., **2009**, *13*, 67.
12. J. Seda Neto, E. Carone, *Liver Transpl*., **2010**, *16*, 426.
13. R.A. Fisher, *Ann. Eugenic*., **1936**, *7*, 179.
14. W.R. Klecka, *Discriminant Analysis*, Sage Publications, Iowa, **1980**.
15. G.H. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, **2004**.
16. C. Sârbu, H.F. Pop, R.S. Elekes, G. Covaci, *Rev. Chim*.(Bucharest), **2008**, *59*, 1237.
17. H.F. Pop, C. Sârbu, *MATCH*, **2013**, *69*, 391.
18. D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, *Chemometrics, a Textbook*, Elsevier Science Publishers, Amsterdam, **1988**.
19. M. Hamilton, *J. Clin. Path.*, **1977**, *30*, 454.
20. R.T. Russell, I.D. Feurer, P. Wisawatapnimit, E.S. Lillie, E.T. Castaldo, C. Wright Pinson, *HPB*, **2008**, *10*, 30.