Multiple Linear Regression-Based Model for Predicting the Fibroscan Technique Outcome in Hepatic Diseases

Marina GORUNESCU[1], Monica LUPŞOR[2], Radu BADEA[2] and Horia ŞTEFĂNESCU[3]

[1]Faculty of Mathematics and Computer Science, University of Craiova,
[2]Department of Ultrasonography, 3rd Medical Clinic, University of Medicine and Pharmacy, Cluj-Napoca
[3]Department of Hepatology, 3rd Medical Clinic, University of Medicine and Pharmacy, Cluj-Napoca
mgorun@inf.ucv.ro
mmlupsor@yahoo.com
rbadea@umfcluj.ro
hstefanescu@umfcluj.ro

Abstract. The non-invasive medical diagnosis has become nowadays one of the most important applications of the applied Statistics boosted by the computers computation power. FibroScan is a rapid, non-invasive technology that uses low frequency vibration and ultrasound to assess the stiffness of the liver tissue. The aim of this study is to predict both the stiffness of the liver tissue and the fibrosis stages, depending on the most significant medical characteristics, based on a multiple linear regression model applied to a concrete database.
Keywords: fibroscan, forward stepwise regression, backward stepwise regression, stiffness prognosis, fibrosis stages
Math. Subject Classification 2000: 62J05

1 Introduction

The prognosis and management of chronic liver diseases largely depend on the amount and progression of liver fibrosis. In patients with chronic hepatitis C (CHC), the precise stage of liver fibrosis is the most important predictor of disease progression and determines the need for antiviral therapy [1]. An obvious trend in clinical practice observed in the latter years consists of finding a correct method for liver fibrosis evaluation in a non-invasive way, both by biochemical tests as well as imaging methods, as an alternative to liver biopsy. Virtually, all serological tests developed to date have not entirely met clinician expectations, as they insufficiently predict the stage of liver fibrosis on an individual level. The imaging diagnosis methods have the advantage of being non-invasive and, at the same time, they allow a complete evaluation of the entire organ, with a precise appreciation of the seriousness of the disease when

the fibrosis does not affect the liver uniformly [2]. The last technological discovery in the evaluation of liver fibrosis is the Fibroscan® (Echosens, Paris), a specially adapted ultrasound device using the principle of the one-dimension transient elastography (TE) for the assessment of liver stiffness. Basically, the practice of Fibroscan is based on establishing some cut-off values of the liver stiffness for each stage of fibrosis [3]. The aim of our study is to apply the multiple regression model to predict the stiffness values, provided by Fibroscan, depending on the main clinical, biochemical and histological parameters. This regression-based methodology will help physicians to automatically estimate both the stiffness and the fibrosis stages, thus saving up time and financial resources.

2 Multiple linear regression

The general purpose of multiple regression technique is to identify the relationship between several independent (predictor) variables and a dependent (criterion) variable. This relationship can be used in a multiple regression analysis to build the corresponding regression equation:

$$Y = b_0 + b_1 * X_1 + \ldots + b_k * X_k.$$

Once this regression equation has been determined, the analyst can easily compute the expected (predicted) value of the dependent variable, depending on the values of the known independent predictors.

There are several situations in which the multiple regression method is necessary:

1. We may wish to remove possible effects of other unimportant variables from the study concerning the relation between two variables only;
2. We wish to explore possible prognostic variables with no prior information of which variable are important;
3. We may wish to develop a prognostic methodology from several explanatory variables to predict the (unknown) value of the dependent variable (of interest).

Multiple regression is relatively straightforward when we previously know the influence of each predictor upon the dependent variable and the difficulties occur when we wish to identify from a large number of possible predictors the most important ones and to further identify the concrete mathematical relationship.

The statistical significance of each variable in the multiple regression model is obtained by using the P-level (corresponding to the t-statistics for the ratio of the regression coefficients to the standard error).

In this study we have used two different approaches concerning the multiple linear regression: (a) the *forward stepwise regression* and (b) the *backward stepwise regression*.

Forward stepwise regression. This method consists of three steps:

- Step1: find the unique variable that has the strongest connection with the dependent variable (that is the variable with the most significant correlation) and consider it in the model.
- Step 2: find the variable among those not included in the model yet, which the largest correlation with the residuals of the model so far.
- Step 3: repeat step 2 until the addition of an extra variable to the model is not statistically significant ($P < 0.05$).

Backward stepwise regression. Different from the above algorithm, the backward stepwise regression considers all the variables in the initial model and next removes the variables less important one at a time, until all the remaining variables in the model are statistically significant.

It is not unusual that using the two algorithms we obtain different models. However, neither approach is more correct than the other. Usually, both algorithms are used together and the final result is balanced by the two complementary approaches.

For more details concerning the multiple regression, see [4].

3 The dataset

The concrete dataset used in this study consists of 324 patients, with chronic HCV infection, examined at the 3rd Medical Clinic, University of Medicine and Pharmacy Cluj-Napoca, Romania, between May 2007 and March 2008. All of them had positive HCV-RNA in their serum and underwent percutaneous liver biopsy for grading and staging the diseases. All patients were referred to liver stiffness measurement. Besides the epidemiological, anthropometric clinical parameters and other important parameters, the following biological parameters were determined for all patients on the same day as liver stiffness: aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyl-transpeptidase (GGT), total bilirubin (TB), alkaline phosphatases (AP), platelet count, prothrombin time [international normalized ratio (INR)], sideremie, fasting blood glucose, fasting serum cholesterol, HDL- cholesterol, triglycerides, fatty load, prothrombin index (PI), abdominal perimeter, medium eritrocity volume (MEV) and body mass index (BMI).

For each patient, both the stiffness of the liver tissue (obtained by fibroscan) and the the fibrosis stages is recorded, and afterwards used to obtain the corresponding multi-linear regression equations. In this way, one can predict both the unknown value of the tissue stiffness and the fibrosis stages of a new patient, by using the known values of the histological, clinical and biochemical parameters.

The study was approved by the local Ethical Committee of the University of Medicine and Pharmacy Cluj-Napoca. The nature of the study was explained to the patients, each of whom provided written informed consent before the beginning of the study, in accordance with the principles of the Declaration of Helsinki (revision of Edinburgh, 2000).

4 Results

We have split the whole dataset into two subsets: (*a*) the *learning dataset* containing 300 cases, and (*b*) the *testing dataset* containing 24 cases.

We have performed several regression analyzes, varying the explanatory variables included in the model, in order to identify the most important connections between the medical parameters obtained by standard procedures and the stiffness obtained by the fibroscan technique.

Firstly, we analyzed the correlation between the two main research dependent variables: tissue stiffness and fibrosis stages (MetavirF), on the one hand, and all the explanatory variables (histological, clinical and biochemical parameters) on the other hand. The Pearson's r correlation coefficient shows the degree of connection between the two outcomes and the possible predictors, in this way explaining and predicting them at the same time. Table 1 shows the correlation analysis results (the significant correlations being marked, P-value < 0.05).

Table 1. Correlation analysis (significant correlations marked)

Variables	**Stiffness**	**MetavirF**	**Variables**	**Stiffness**	**MetavirF**
Stiffness	1	***0.65***	AST	***0.26***	***0.25***
Age	***0.17***	***0.3***	ALT	-0.01	0.03
IMC	***0.13***	0.08	GGT	***0.24***	***0.22***
Metavir A	***0.12***	***0.28***	TB	***0.51***	***0.2***
MetavirF	***0.65***	1	AP	***0.55***	***0.22***
IQR1	***0.52***	***0.43***	Prothrombin index (PI)	***-0.33***	***-0.24***
Success rate	-0.01	0.05	INR	***0.36***	***0.23***
Sex	-0.01	-0.06	VEM	***0.12***	0.08
Glycemia	0.09	0.1	Trombocites (Tr)	***-0.31***	***-0.29***
Triglycerides	0.02	0.03	Fatty load (FL)	-0.03	0.01
Cholesterol	-0.09	***-0.19***	Abdominal perimeter (AP)	0.02	0.05
HDL-CST	-0.06	-0.05	Sideremie	***0.14***	***0.15***

One of the main questions regarding the use of the tissue stiffness obtained the fibroscan technique is concerns the possible relation between stiffness and the fibrosis stages (MetavirF). There is indeed a strong relation between stiffness and MetavirF (highly positive correlated), given by the following linear regression equation:

Stiffness = -1.05 + 6.05 * MetavirF

The scatterplot of the two variables and the corresponding regression line (95% confidence interval included) are displayed in Fig. 1.

Thus, given the value of the fibrosis stages (MetavirF), it is easy to estimate the value of the stiffness, without directly using the fibroscan.

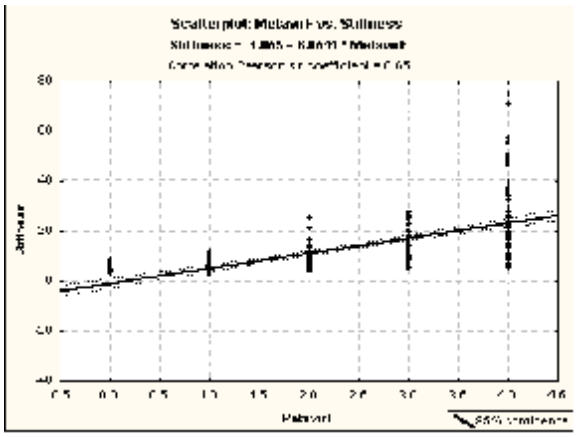

Fig. 1. Scatterplot: Stiffness vs. MetavirF and the regression equation

Next, we used the regression model to identify the possible relations between the medical analyzes (histological, clinical and biochemical parameters) and the tissue stiffness obtained by using the fibroscan technique. To accomplish this task, we have used both the forward stepwise regression (maximal model) and the backward stepwise regression (minimal model), the final decision being made by considering a balance between the two models.

The first step involved the investigation of the possible connections between stiffness and the histological possible predictors. The results of this analysis consist of the two regression equations (forward and backward models):

$$Stiffness = 0.47 + 6.49 * MetavirF - 1.15 * MetavirA$$

$$Stiffness = -1.41 + 6.30 * MetavirF$$

Remark. 1. In the first equation, the variable MetavirA is not significant ($P = 0.11$), but is worth to be included in the model, since it is a significant histological parameter.

2. Although the backward stepwise regression above has the same terms like the simple linear regression model, they have slight different coefficients. This means that the backward stepwise regression model significantly differs from the simple regression model, even if it provides a two terms equation.

The second step deals with the regression model applied to the stiffness and the clinical and biochemical parameters as independent variables. In this case, we obtained (forward and backward models):

$$Stiffness = -7.39 + 0.03 * AP + 0.11 * PI - 0.04 * Tr + 2.18 * TB + 11.71 * INR + 0.32 * IMC + 0.01 * GGT$$

$$Stiffness = -5.16 + 1.98 * TB + 0.04 * AP + 16.24 * INR - 0.05 * Tr$$

Let us notice that the variables: TB, AP, Tr, INR appear in both equations, which means that they are the most important predictors.

The third step consists in investigating the regression models involving the tissue stiffness and the histological, clinical and biochemical parameters. The two regression equations are displayed below:

$$Stiffness = -8.45 + 4.30 * MetavirF + 0.03 * AP - 0.07 * PI + 1.63 * TP - 0.02 * Tr + 8.19 * INR - 0.1 * AP$$

$$Stiffness = -11.37 + 4.37 * MetavirF + 1.59 * TB + 0.03 * AP + 10.39 * INR - 0.03 * Tr$$

In this case, the variables: MetavirF, AP, INR and Tr are the most important predictors.

Remark. In the important case when the patients have the same values for MetavirF (=2), MetavirA (=2) and FL (=0), only the forward regression equation can be computed:

$$Stiffness = -28.70 - 0.02 \times Tr + 0.38 \times IMC + 0.21 \times VEM$$

Finally, we have studied the relationship between the fibrosis stages (MetavirF) and the explanatory variables (clinical, biochemical and elastographical). Thus, we have obtained the following regression equations (forward and backward):

$$MetavirF = 1.03 + 0.08 \times Stiffness + 0.02 \times Age - 0.002 \times AP - 0.003 \times Cholesterol + 0.005 \times AST$$

$$MetavirF = 1.40 + 0.08 \times Stiffness + 0.02 \times Age - 0.003 \times Cholesterol - 0.002 \times AP$$

As we can see, in this case the two regression models provide almost the same equations, that is the fibrosis stages, given by the MetavirF variable are easier to be evaluated that the tissue stiffness.

5 Conclusions

This paper deals with the application of the multiple linear regression methodology (both forward stepwise and backward stepwise models) to a concrete dataset concerning the chronic hepatitis C. The regression model has been applied to forecast both the tissue stiffness obtained by using the Fibroscan, and the fibrosis stages obtained by biopsies. Moreover, using this methodology, we assessed the hierarchy of the most important clinical, biochemical and histological medical parameters, in order to be subsequently used in future studies for classification purposes.

References

1. ***, NIH Consensus Statement on Management of Hepatitis C (2002). In: NIH Consens State Sci Statements 2002; 19: pp. 1–46.
2. **Afdhal, N.H., Curry, M.**, Technology evaluation: a critical step in the clinical utilization of novel diagnostic tests for liver fibrosis. J. Hepatol 46, 543–545 (2007)
3. **Sandrin, L., Fourquet, B., Hasquenoph, J.M., et al.**, Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. Ultrasound Med Biol 29, 1705–1713 (2003)
4. **Pardoe, I.**, Applied Regression Modeling. A business Approach. Wiley: Hoboken, New Jersey, (2006)