Abstract—This research is aiming to set higher standards in healthcare research in Bosnia and Herzegovina from the aspect of mathematical and information science methodology. The data on Hepatitis C virus treatment published in previous work was reinvestigated using the Bayesian network methodology. The results are still in the line with the results published worldwide, but this research provides an additional quality, through the conditional probabilities for treatment outcome. Further development of this model may result in an acceptable prediction model for HCV treatment outcome.

I. INTRODUCTION

The goal of this paper is to present a Bayesian network model for early prediction of therapy response of patients infected with Hepatitis C virus (HCV). The data used in this research was collected over a five year period (2005 – 2009) with nearly 300 HCV patients admitted at the Clinic for Gastro-entero-hepatology, Sarajevo, Bosnia and Herzegovina. Since the above Clinic is the main clinic for the region of Federation of Bosnia and Herzegovina (FBiH), the predictive model should be valid for the whole FBiH.

Background

The authors of this paper have worked together for a long time on HCV treatment data analysis [1]. It seemed natural to upgrade the cooperation into a new level – that is, to start using more advanced prediction techniques. One of them is Bayesian network, and here is presented a model for prediction of response to HCV therapy based on typical data for patients from FBiH.

Data Collection and Preparation

The data for this research were collected in the University of Sarajevo Clinical Centre over the five year period (2005-2009). The patients, total 231 with the complete data, were from the whole region of FBiH, and were admitted to the Clinic for Gastro-Entero-Hepatology. They were treated with the Pegylated Interferon and Ribavirin therapy.

The criterion for therapy continuation was Early Virologic Response (EVR), which is HCV RNA test after 12 weeks of treatment (EVR12). The patients that failed that test were not treated any more (were listed as having negative response to the therapy). The criterion for the therapy success was the SVR, the HCV RNA test performed six months after the end of therapy.

It is very important to note that all further probability calculations were performed only for the patients that completed the whole therapy, so all of the probabilities in the next sections should be interpreted that way. That is, if we say that probability of therapy success is 0.5, we assume that 50% of patients who completed the therapy will have a successful therapy outcome.

For the purpose of this research, we also selected HCV Genotype (1a, 1b, 3 or 4), age of patients at the beginning of therapy (Age), and two Liver Inflammation indicators obtained after the liver biopsy, Necroinflammatory Score (NiSc) and the Stage of Fibrosis (FibS). We selected these variables following the suggestion of medical part of our team; for example, Hepatic Steatosis was not chosen as predictor following the results published in [2].

II. METHODS

In order to develop a Bayesian network model, several steps had to be taken: variables had to be categorized; dependences among the selected variables tested, and conditional probabilities calculated.

Since the most of the variables in this research are continuous, authors decided to categorize them in the following way:

For EVR12 and SVR were first calculated as the logarithmic values, and then categorized into the following categories: 0, 1-4, and 5+

The variable Age was grouped into 5 categories: <25, 25-34, 35-44, 45-54, 55+

For NiSc (ranging from 0 to 18) the authors decided on the following categories: 0, 1-4, 5-9, 10+, while for FibS (ranging from 0-3) on 0, 1, 2, 3+

The dependences among variables served to establish dependency arcs in the Bayesian network model, while the
direction of the arc was decided following the suggestion of the medical part of the team.

Conditional probabilities served to measure the strength of the dependency. The authors decided on the naive Bayesian Network (BN) model, for it is proven to assure good results compared to other BN models [3].

Statistical analysis was performed in open source software R [4], while the Bayesian network modeling was performed in open-source software MSBNX [5].

III. RESULTS AND DISCUSSION

The relation among EVR12 and SVR was tested with paired t-test (for both original and logarithmic values), and the obtained p-value was <0.01. Therefore one can conclude that SVR significantly depends on EVR.

The first dependency arc is set from Genotype directly to SVR, according the results presented in [6].

From the statistical analysis it turned out that Age and EVR12 are not independent; for example, it is known that younger patients show better therapy response [7]. Since age is immanent to a person, the dependency arc is directed from Age to EVR12, that is, we say that early viral response depends on the age of patients at the beginning of therapy.

The only other set of dependencies in this model is between NiSc and FibS. Since Necroinflammatory Score is the first predictor of the Liver Inflammation, and Fibrosis is developing after the Inflammation processes of the Liver Cells, we set the dependency arc from NiSc to FibS, that is, the Fibrosis Stage depends on the Necroinflammatory Score.

From Table 2 it is visible that for the same predictor values of Age, Genotype, NiSc, and FibS, it turns out that the probability of a successful therapy outcome is somewhat smaller, and it has value of 0.8226. Therefore, one can conclude that of such patients who completed the therapy 82.26% will be cured.

It seems that the later model is more appropriate for prediction, so EVR12 will be unobserved for the next examples as well.

The authors also wanted to check the validity of the model in relation to HCV Genotype. It is well known that older patients with Genotype 3 react worse to therapy than other Genotypes [8]. Nevertheless, from this model it follows that patient with HCV Genotype 3 reaction to the therapy is more dependent on the stage Fibrosis.

Table 3 presents the model evaluation for the same Age group (35-44), NiSc (1-4), FibS (1) but for Genotype 3. EVR12 is unobserved.

From Table 3 it is visible that patients infected with HCV Genotype 3 have 70% chance of being cured.

On the other hand, if such patients have no Fibrosis, (Table 4) they have a slightly greater chance of being cured (77.6%).

From Table 5 it is visible that patients with other HCV Genotypes have slightly less chances of being cured (75.38%) under the same assumptions as in Table 4.

When it comes to the age of patients at the beginning of therapy (variable Age), it can be interpreted in two ways: as age of patients at the moment of infection, or, as the therapy should be cured assuming the above determined values for predictors.

In other words, 81.8% of all patients who completed the therapy 82.26% will be cured. It seems that the later model is more appropriate for prediction, so EVR12 will be unobserved for the next examples as well.

The authors also wanted to check the validity of the model in relation to HCV Genotype. It is well known that older patients with Genotype 3 react worse to therapy than

<table>
<thead>
<tr>
<th>Node Name</th>
<th>State 0</th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
<th>State 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &lt;25</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>24-34</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>35-44</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>45-54</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>55+</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

From Table 2 it is visible that for the same predictor values of Age, Genotype, NiSc, and FibS, it turns out that the probability of a successful therapy outcome is somewhat smaller, and it has value of 0.8226. Therefore, one can conclude that of such patients who completed the therapy 82.26% will be cured.

From Table 3 it is visible that patients infected with HCV Genotype 3 have 70% chance of being cured.

The predictive model is presented in Figure 1.

The example of model evaluation for the assigned value of EVR12 is presented in Table 1. In that table is shown an example for fixed values of all predictor. It turns out that for the patients in the age group 35-44 with HCV Genotype 1b, with primary stage of Liver Inflammation (NiSC in 1-4), and primary stage of Fibrosis (FibS=1), the probability of successful treatment outcome is 0.8180.

In other words, 81.8% of all patients who completed the therapy should be cured assuming the above determined values for predictors.

Since there exists a significant dependence of SVR from EVR12, Table 2 presents the same model but with EVR12 unobserved.
Nevertheless, there are patients who repeatedly start planning to investigate a model in which they will observe the therapy treatment outcome, in the next research the authors plan to explore gender instead. Predictive model with multiple variables, so in our next research we plan to explore gender instead.

The situation is better for younger age groups, and the worst for the oldest age group. In order to compare the results, the other predictor variables fixed are presented in Table 5. Now it turns out that patients of an older age group (45-54) have smaller chances of being cured (58.76% compared to the previously obtained 75.38%).

The situation is similar for the other HCV Genotypes, but not in Bosnia and Herzegovina. The results obtained in this model do not differ in general neither from the results published in international papers, nor from the known facts about HCV.

It is shown that the early virology response (EVR12) is the best predictor for the final therapy outcome (SVR), and that the age of patients at the beginning of therapy (Age) influences SVR in the way that younger patients have greater chance for a successful therapy outcome than older ones. Furthermore, the presented model gives different probabilities of a successful therapy outcome for different HCV Genotypes, as it should be.

When it comes to the Liver Inflammation stage (NiSc), it turns out that there is little difference for patients with no Inflammation and the patients with the primary Inflammation stage (NiSc values 0, and 1-4). On the other hand, later stages of Liver Inflammation, and especially stage of Fibrosis (FibS) affect the therapy outcome, in particular when reflected to HCV Genotype.

Nevertheless, there is place for model improvements, and this team will explore other predictors in future.

IV. CONCLUSIONS

The most important result of this research, as the authors see it, is the variety of obtained results. This way, the healthcare workers can gain a more complete picture of any treatment outcome when having to cope with multiple predictors. On the other hand, professionals with mathematical or IT background are gaining a chance to work with real-life data and obtain multidisciplinary specialization.

The results obtained in this model do not differ in general neither from the results published in international papers, nor from the known facts about HCV.

It is shown that the early virology response (EVR12) is the best predictor for the final therapy outcome (SVR), and that the age of patients at the beginning of therapy (Age) influences SVR in the way that younger patients have greater chance for a successful therapy outcome than older ones. Furthermore, the presented model gives different probabilities of a successful therapy outcome for different HCV Genotypes, as it should be.

When it comes to the Liver Inflammation stage (NiSc), it turns out that there is little difference for patients with no Inflammation and the patients with the primary Inflammation stage (NiSc values 0, and 1-4). On the other hand, later stages of Liver Inflammation, and especially stage of Fibrosis (FibS) affect the therapy outcome, in particular when reflected to HCV Genotype.

Nevertheless, there is place for model improvements, and this team will explore other predictors in future.

ACKNOWLEDGMENT

The authors want to thank all staff of University of Sarajevo Clinical Centre who contributed to data collection and analysis.

REFERENCES


as a Predictor of Early Response to Pegylated Interferon and Ribavirin Therapy in Patients with Chronic Hepatitis C. Journal of US-China Medical Science, ISSN 1548-6648, USA, Sep. 2010, Volume 7, No. 9 (Serial No. 70).


