Analysis of Big Data of HCV Patients

¹Sara Khaleel and ²Ahed J. Alkhatib
¹*Department of Pharmacy, School of Pharmacy, The University of Jordan, Amman, Jordan*
²*Department of Legal Medicine, Toxicology of Forensic Science and Toxicology, School of Medicine, Jordan University of Science and Technology, Jordan*

Key words: HCV, liver, jaundice, headache, nausea, big data

Corresponding Author:
Ahed J. Alkhatib
Department of Medicine and Critical Care, International Mariinskaya Academy

Abstract: The present study aimed to analyze big data posted on Kaggle about HCV infection and to find correlations between demographic variables and clinical variables related to HCV infection. The data posted on Kaggle is a large data consisting of 1385 patients. Data included some variables such as age, gender and Body Mass Index (BMI). Clinical manifestations were also included such as fever, jaundice, headache, nausea and vomiting. Variables including laboratory findings including white blood cells, red blood cells, platelets and hemoglobin were also included. Various statistical models were included such as descriptive statistics such as frequencies, percentages, means, and standard deviations. The correlations between study variable were assed using Pearson correlation. Significance was considered at $\alpha \leq 0.05$. Study findings showed that clinical manifestations were reported by about 50% of patients. The results reported some correlations between study variables including positively significant correlations between HB and BMI, nausea and vomiting. Also, there were some negatively significant correlation between jaundice and BMI and diarrhea and hemoglobin. Taken together, we recommend future studies to investigate the importance of such correlations.

INTRODUCTION

HCV is one of the most common human pathogens, contaminating in excess of 150 million individuals overall[1]. Roughly 3% of the overall populace is infected with the hepatitis C infection[2]. The commonness of HCV disease fluctuates all through the world, with the most noteworthy commonness revealed in Egypt[3]. HCV causes incessant diseases that can lead to liver fibrosis, cirrhosis and hepatocellular carcinoma[2]. Along these lines, creating proficient framework that can foresee the probability of patients getting HCV infection. It empowers critical information, for example designs, connections between clinical factors identified with HCV infection, to be set up. A significant test confronting healthcare industry is nature of administration. Quality of administration suggests diagnosing illness accurately and gives compelling medicines to patients. Poor conclusion

can prompt lamentable outcomes which are inadmissible. Information digging could be utilized for examining and finding shrouded designs inside patient's datasets. In this way, a smart framework for anticipating patients of HCV can be assembled and is viable. The primary issue, in mining the clinical databases is the modest number of patients comparative with the quantity of highlights[1].

HCV disease is the main source of cirrhosis (93%) which is a hazard factor for Hepatocellular Carcinoma (HCC)[4, 5]. It prompts both hepatic aggravation and fibrosis. Transformation what's more, threatening change of the contaminated cells are advanced by the HCV protein articulation[6-8]. HCV infection is described by its long time movement to cirrhosis-related HCC[9]. In view of the phylogenetic and arrangement investigations of HCV genomes, there are seven genotypes of HCV strains and 67 subtypes upon further grouping of every genotype[10]. HCV genotype 4 is considered the most overwhelming HCV genotype in Egypt (Ray *et al.*, 2000). Distinctive HCV commonness in Egypt were announced. The HCV commonness in the age gathering (15-59 years) was 14.7% in 2008 while it became 10% in 2015[5].

Kouyoumjian *et al.*[11] conducted a study about HCV in Egypt. Egypt is the most affected country by HCV and requirements an extensive portrayal of HCV the study of disease transmission to educate the scale-up of treatment and counteraction programs. Orderly audits, meta-examinations and meta-relapses were directed. An aggregate of 25 frequency, 259 commonness and 47 genotype contemplates were identified. Occurrence and predominance levels were high over all populaces. Genotype 4 represented 94.1% of diseases with a relative Shannon diversity file of just 14.4%. Pooled mean HCV commonness was assessed at 11.9% (95% CI = 11.1-12.6%) among everybody, 55.6% (95% CI = 49.4-61.7%) among populaces at high hazard, 14.3% (95% CI = 10.3-18.8%) among populaces at transitional hazard, 56.0% (95% CI = 50.4-61.6%) among populaces with liver-related conditions and 35.0% (95% CI = 27.3-43.1%) among uncommon clinical populaces. Mean HCV viremic rate was assessed at 66.7% (95% CI = 61.7-71.5%). Meta-regression showed 6% lower chances for HCV commonness for every one-year increase in production year (AOR = 0.94; 95% CI = 0.92-0.96). HCV predominance is high with proof for progressing transmission essentially through medicinal services. Genotype assorted variety is low with genotype 4 predominance. 66% of counter acting agent positive Egyptians are incessantly tainted and need treatment. Clinical populaces ought to be organized for screening. Regardless of the huge scope pestilence, commonness seems by all accounts, to be declining quickly steady with a contracting epidemic.

Study objectives: The main objectives of this study are to analyze large data of 1385 patients with HCV and to find correlations between study variables.

MATERIALS AND METHODS

This study is based on a dataset posted on Kaggle (https://www.kaggle.com/nourmibrahim/hepatitis-c-virus-hcv-for-egyptian-patients). This dataset is about Hepatitis C Virus (HCV) for Egyptian patients who were subjecting for treatment dosages for HCV about 18 months. Discretization should be applied based on expert recommendations. This data was obtained from UCI Machine Learning Repository

Citation: Dua and Graff, UCI Machine Learning Repository (http://archive.ics.uci.edu/ml). Irvine, CA: University of California, School of Information and Computer Science.

The dataset included description of 1385 patients with HCV. Dataset was well prepared without any missing data. The data was analyzed using SPSS Version 21. Descriptive statistics were applied. Frequencies and percentages were used to describe categorized variables such as gender. Means and standard deviations were used to describe continuous variables such as age and Body Mass Index (BMI). The relationships between variables were assessed using Pearson correlation. Significance were considered at $\alpha \leq 0.05$.

RESULTS

General characteristics of participants: As shown in Table 1, the mean age is 46.32±8.78 years. The mean BMI is 28.61±4.08. There is a close distribution of males (51%) and females (49%).

Clinical characteristics of participants: Table 2 summarizes clinical characteristics of participants. About 48% of participants had fever, nausea/vomiting was reported by about 50% of patients. Headache was reported by 50.4% of participants. Diarrhea was reported by 49.7% of patients. Fatigue was reported by 50.1% of patients. Jaundice was reported by 49.9% of patients. Epigastria pain was reported by 49.6% of patients.

Blood counts of study participants: Table 3 summarizes levels of blood counts among patients. The mean level of WBC is 7533.39±2668.22 mL^3, the mean level of RBC is 4422129.61±346357.71 m^3. The mean level of HB is 12.6±1.71 g dL^{-1}. The mean level of platelets is 158348.06±38794.79 mL^3.

Significant correlations between study variables: Significant correlations were found between HB and

Table 1: General characteristics of participants

Variables	Description
Age (M±SD) years	46.32±8.78
Body Mass Index (BMI)	28.61±4.08
Gender (N, %):	
Male	707 (51%)
Female	678 (49%)

Table 2: Clinical characteristics of participants

Variables	Frequencies (N)	Percentage
Fever:		
Yes	671	48.4
No	714	51.6
Nausea/vomiting:		
Yes	689	49.7
No	696	50.3
Headache:		
Yes	698	50.4
No	687	49.6
Diarrhea:		
Yes	689	49.7
No	696	50.3
Fatigue:		
Yes	694	50.1
No	691	49.9
Jaundice:		
Yes	691	49.9
No	694	50.1
Epigastric pain:		
Yes	687	49.6
No	698	50.4

Table 3: Blood counts of study participants

Variables	Mean	SD
White Blood Cells (WBC)	7533.39	2668.22
Red Blood Cells (RBC)	4422129.61	346357.71
Hemoglobin (HB) g dL^{-1}	12.6	1.71
Platelets	158348.06	38794.79

Table 4: Significant correlations between study variables

Variables	R-values	Significance
HB-BMI	0.06	0.031
Jaundice-BMI	-0.07	0.008
Nausea-Vomiting	0.86	0.001
Diarrhea-HB	-0.062	0.022

BMI (r = 0.06, p = 0.031). There was also another negative and significant correlation between jaundice and BMI (r = -0.07, p = 0.008). the results also showed a significantly positive correlation between nausea and vomiting (r =0.86, p = 0.001). Another significantly negative correlation was found between diarrhea and HB (r = -0.062, p = 0.022) (Table 4).

DISCUSSION

The present study was conducted to analyze large data about HCV for 1385 patients. The results showed that following treatment period of three months that clinical manifestations were in the range of 50%. However, this is still higher than reported in other studies that reported manifestations to be within the limit of 25%[12].

Regarding blood indices, the results showed that the mean levels of blood counts were within normal reference. Several studies reported that HCV patients are more likely to have higher levels of blood counts compared with control subjects[13-17].

Correlation studies showed significantly positive correlation between BMI and HB. This may due to the effect of disease and treatment options. In other words, the disease may affect weight and HB levels and treatment may reverse the situation. Overweight adversely affects the progression of chronic HCV liver disease and is associated with diminished response to antiviral therapy using weight-based dosing in a cohort with minimal comorbidities[12, 18, 19].

CONCLUSION

The present study showed that there is a significantly positive correlation between BMI and HB which could be attributed to the impacts of disease progression. There was also a positively significant correlation between nausea and vomiting. Other significantly negative correlations were found between BMI and jaundice, diarrhea and hemoglobin. These correlations may open the door for other scientific research works.

REFERENCES

01. Rarwan, A.A.A., T.E. Hafeez and H. Mamdouh, 2013. An analysis of hepatitis C virus prediction using different data mining techniques. Int. J. Comput. Sci. Eng. Inf. Technol. Res. (IJCSEITR.), 3: 209-220.

02. Chen, S.L. and T.R. Morgan, 2006. The natural history of hepatitis C virus (HCV) infection. Int. J. Med. Sci., 3: 47-52.

03. Frank, C., M.K. Mohamed, G.T. Strickland, D. Lavanchy, R.R. Arthur and L.S. Magder *et al.*, 2000. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. Lancet, 355: 887-891.

04. Yang, J.D., W.R. Kim, R. Coelho, T.A. Mettler and J.T. Benson *et al.*, 2010. Cirrhosis is present in most patients with hepatitis B and hepatocellular carcinoma. Clin. Gastroenterol. Hepatol., 9: 64-70.

05. Rashed, W.M., M.A.M. Kandeil, M.O. Mahmoud and S. Ezzat, 2020. Hepatocellular Carcinoma (HCC) in Egypt: A comprehensive overview. J. Egypt. National Cancer Inst., Vol. 32, 10.1186/s43046-020-0016-x

06. Tang, A., O. Hallouch, V. Chernyak, A. Kamaya and C.B. Sirlin, 2017. Epidemiology of hepatocellular carcinoma: Target population for surveillance and diagnosis. Abdom. Radiol., 43: 13-25.

07. El-Houseini, M.E., A. Ismail, A.A. Abdelaal, A.H. El-Habashy and Z.F. Abdallah *et al.*, 2019. Role of TGF-ß1 and C-Kit Mutations in the Development of Hepatocellular Carcinoma in Hepatitis C Virus-Infected Patients: in vitro Study Biochem. Moscow, 84: 941-953.

08. Neamatallah, M., M. El-Bendary, H. Elalfy, T. Besheer and M. Abd El-Maksoud *et al.*, 2019. Impact of toll-like receptors 2(TLR2) and TLR 4 gene variations on HCV susceptibility, response to treatment and development of hepatocellular carcinoma in cirrhotic HCV patients. Immunol. Invest., 49: 462-476.

09. Freeman, A., 2002. Estimating progression to cirrhosis in chronic hepatitis C virus infection. Hepatol., 34: 809-816.

10. Kumar, A., M.K. Rajput, D. Paliwal, A. Yadav, R. Chhabra and S. Singh, 2018. Genotyping & diagnostic methods for hepatitis C virus: A need of low-resource countries. Indian J. Med. Res., 147: 445-455.

11. Kouyoumjian, S.P., H. Chemaitelly and L.J. Abu-Raddad, 2018. Characterizing hepatitis C virus epidemiology in Egypt: Systematic reviews, meta-analyses, and meta-regressions. Sci. Rep., Vol. 8, 10.1038/s41598-017-17936-4

12. Hoofnagle, J.H., 1997. Hepatitis C: The clinical spectrum of disease. Hepatology, 26: 15S-20S.

13. Wang, C.S., W.J. Yao, S.T. Wang, T.T. Chang and P. Chou, 2004. Strong association of Hepatitis C Virus (HCV) infection and thrombocytopenia: Implications from a survey of a community with hyperendemic HCV infection. Clin. Infect. Dis., 39: 790-796.

14. Tsai, M.H., K.H. Lin, K.T. Lin, C.M. Hung and H.S. Cheng *et al.*, 2015. Predictors for early identification of hepatitis C virus infection. BioMed. Res. Int., 2015: 1-7.

15. El Barbary, M.A., A.E.M. Saad, F.M. Attia, M.I. Mandour, M.A. Haidara, M.M. Dallak and E.R. Isenovic, 2010. Thrombocytopenia in patients with chronic hepatitis C: A possible role of HCV on platelet progenitor cell maturation. Angiology, 61: 304-313.

16. Kauf, T.L., D.R. Nelson, J. Schelfhout, J.A. Delaney and P.F. Wang, 2012. Trends in the prevalence of thrombocytopenia among individuals iInfected with hepatitis C Virus in the United States, 1999-2008. BMC Res. Notes, Vol. 5, 10.1186/1756-0500-5-142.

17. Tsai, M.C., K.M. Kee, Y.D. Chen, L.C. Lin, L.S. Tsai, H.H. Chen and S.N. Lu, 2006. Excess mortality of hepatocellular carcinoma and morbidity of liver cirrhosis and hepatitis in HCV-endemic areas in an HBV-endemic country: Geographic variations among 502 villages in Southern Taiwan. J. Gastroenterol. Hepatol., 22: 92-98.

18. Delgado-Borrego, A., D. Healey, B. Negre, M. Christofi and S. Sabharwal *et al.*, 2010. The influence of body mass index on outcome of pediatric chronic hepatitis C virus infection. J. Pediatr. Gastroenterol. Nutr., 51: 191-197.

19. Ray, S.C., R.R. Arthur, A. Carella, J. Bukh and D.L Thomas, 2000. Genetic epidemiology of hepatitis C virus throughout Egypt. J. Infect. Dis., 182: 698-707.