

Objective Scoring of Extent of Steatosis and Fibrosis Using Computer Assisted Image Analysis System in Liver Biopsy of Non-alcoholic steatohepatitis Patients

Prateek Kinra^{1*}, Prasenjit Das¹, Rajni Yadav¹, Saumyaranjan Mallik², Shalimar³, SK Acharya⁴,
Guresh Kumar², Maneesh Vijay K² and Siddhartha Datta Gupta²

¹Dept of Pathology, AFMC Pune, India

²Dept of Path, AIIMS, New Delhi, India

³Dept of Gastroenterology, AIIMS, New Delhi, India

⁴Dept of Biostatistics, AIIMS, New Delhi, India

ABSTRACT

Background: The histological grading/staging system for Non-alcoholic steatohepatitis (NASH) encompasses the following criteria: extent of steatosis, lobular-inflammation, ballooning hepatocytes and extent of fibrosis. The existing scoring system has a drawback of high subjectivity. This study was designed with aims of studying (i) Inter & intra-observer variation; (ii) Correlation of the visual grading/staging with computer assisted image-analysis technique (IA); ultrasonography grade of fatty liver and fibroscan for extent of fibrosis.

Methods: 48 known cases of NASH were studied. The Sirius-red stained slides were assessed for % area of steatosis, fibrosis and parenchyma using Image Pro Plus® 6.1 image-analysis software.

Results: The Kappa coefficient value for inter-observer variation was 0.338 (SE: 0.111-fair agreement); intra-observer variation was -0.083 (SE: 0.143-poor agreement). There was no agreement between the steatosis scoring between visual grading and IA score (K= -0.16). IA score unlike visual score correlated well with the USG grading system (K= 0.65) and biochemical variables [P values TG (0.04), BMI (0.039)]. For the first time in literature image analysis fibrosis score has been assessed in cases of NASH. The correlation between fibroscan score and IA fibrosis showed good correlation ($r=0.766$). The paper also lays down the cut off values of %fibrosis score on image analysis. A linear progression curve has been plotted to estimate extent of actual fibrosis in liver as against the existing fibroscan test.

Conclusion: Visual assessment over-estimates the area of steatosis, resulting in false higher NASH histological grade, unlike the image-analysis system. The image-analysis adds objectivity to score steatosis and fibrosis.

Keywords: Image Analysis, NASH, Scoring

Introduction

Non-alcoholic fatty liver disease (NAFLD), described by Ludwig et al. in 1980,^[1] is characterized histologically by fat accumulation in >5% of hepatocytes [steatosis] with associated inflammation and hepatocyte injury [non-alcoholic steatohepatitis (NASH)] or their complications so forth. While simple steatosis seen in NAFLD does not increase the short-term morbidity or mortality; NASH considerably increases risks of cirrhosis, liver failure and hepatocellular carcinoma. It is estimated that NAFLD/NASH will increase the 5-year direct and indirect medical costs by 26% by World Gastroenterology Organisation Global Guidelines.^[2] When NAFLD becomes chronic, the degenerated hepatocytes are replaced gradually by fibrosis. It is beyond doubt that fibrosis score is extremely important in predicting treatment outcome. Visual fibrosis scoring especially in liver core biopsies suffers from (i) sampling error that can be overcome by two core biopsies and (ii)

lack of agreement in assigning the correct stage due to high inter-observer variation. Though many histopathological scoring systems have been proposed as by Matteoni et al for NAFLD and Brunt *et al* for NASH,^[3,4] the histological scoring scheme of the NASH- Clinical Research Network (NASH CRN) is most popular and accepted by both clinical and research communities⁵. In principle, this scoring system comprises of NAFLD Activity Score (NAS) and fibrosis staging. The NAS grading can range from 0 to 8 and is calculated by the sum of scores of steatosis (0-3), lobular inflammation (0-3) and hepatocyte ballooning (0-2). In patients with NAFLD, NAS score of ≥ 5 strongly correlate with a diagnosis of “definite NASH” whereas NAS score ≤ 3 correlates with a diagnosis of “not NASH”.^[5] However these scoring systems are primarily dependent on visual approximation of the histopathologists and thus are prone to subjective variations. General pathologists

lack the exposure to numerous liver biopsies and tend to make errors in assessing the pattern and extent of fibrosis in smaller biopsy samples.^[6] Hence there is requirement of following a widely accepted staging method across the geographical barriers, so that appropriate information can be gained from previous medical records and appropriate management can be instituted when the patient migrates to another country. Similar subjective variation is also noted while assessing steatosis, lobular inflammation and ballooning degeneration. In this regard non-invasive biochemical markers as: body mass index (BMI), AST/ALT ratio, presence of diabetes into a weighted sum, AST-to-platelet ratio index (APRI), gamma glutamyl transferase (GGT), hyaluronic acid (HA), tissue inhibitor of metalloproteinase 1 (TIMP1), amino-terminal peptide of pro-collagen III (P3NP), haptoglobin, α 2-macroglobulin, apolipoprotein A1 and total bilirubin have been studied in different studies;^[7,8] however none of them have been validated till date or found not to be useful in isolation. The non-invasive transient elastography (Fibroscan, Echosens, Paris, France) however have gained popularity amongst the hepatologists to assess liver stiffness score (LSM), which employs ultrasound-based technology to measure liver stiffness and has been validated for use in chronic hepatitis C, HIV/HCV co-infection and cholestatic liver diseases. However elastography score can be influenced by other parameters as degree of inflammation as well as hepatic fat content. Hence whether it can be used blindly in assessment of chronicity in NAFLD/NASH needs to be investigated.^[8]

Moreover, with the recent advent of whole slide digital scanners, tissue histopathology slides can now be scanned and stored in a digital form. As a result digitized tissue histopathology has now become amenable to the application of computerized image analysis. Analogous to its role in radiology, the computer-assisted diagnosis algorithms have begun to be developed for disease detection, diagnosis, and prognosis prediction to complement to the opinion of the pathologist.^[9] The use of image analysis methods has questioned the accuracy of the traditional approach to assessing the severity of steatosis according to the estimated percentage of hepatocytes containing fat droplets while assessing the liver biopsies of patients with NASH.^[10] This study was designed with the aim to analyse the- (i) inter& intra-observer variations for assessing the degree of steatosis& fibrosis using conventional scoring systems; (ii) correlation of the extent of steatosis& fibrosis among visual grading and staging system (NASH CRN) with computer assisted image analysis technique (IA) and ultrasonographic (USG) assessment score - of liver steatosis and fibrosis.

Material and methods

HE stained slides of 48 known cases of NASH were retrieved from archives and NASH CRN grading/staging was done by two pathologists independently. The principal investigator reassessed all the slides after a gap of 4 weeks of initial screening to calculate intra-observer variation. Masson trichrome (MT) and reticulin silver (Retic) stained slides were used to assess the stage of NASH. Sirius red (SR) special stain was carried out in each case to impart a distinct red colour to the areas of fibrosis. Multiple digital images were acquired from the whole SR stained liver core at 4x magnification without significant overlap and were then stitched digitally and processed further by using the Irfan view software (Fig No.1), version 4.36 (free public domain software by IrfanSkiljan, Austria, Europe). Any artefact (such as cracks) or sinusoidal and vascular spaces were assigned white background prior to image analysis. These processed stitched images were then analyzed by the Image analysis software®, version 6.1, Media Cybernetics, USA) for % area of steatosis, fibrosis and total area of liver parenchyma. Each variable was assigned different colours (digital segmentation) and the % area and total area were calculated. The software was calibrated before these measurements. The morphometric measurement was repeated twice for each stitched liver core and the mean values were calculated. The various steps for image analysis are appended below in Fig No. 2. Following additional data for each patient were collected: (i) body mass index (BMI) (ii) lipid profile (iii) waist hip ratio (iv) fibroscan score (v) ultrasound grade of fatty liver. All these parameters were correlated with the image analysis grade and visual NASG CRN steatosis grade.

Exclusion Criteria: (i) patients with known history of chronic alcoholic intake or significant high mean corpuscular volume (MCV)/ Gamma glutamyltransferase (GGT) (ii) serological positivity indicating towards HBV/HCV chronic hepatitis or autoimmune hepatitis (iii) histological evidences of haemochromatosis, alpha 1 antitrypsin deficiency or Wilson's disease.

Statistical analysis: The inter-observer & intra-observer variations were calculated using Kappa coefficient. Kappa statistics was also used to find the agreement levels between the scoring systems. The statistical comparison was performed using paired 't' test and Spearman correlation. ROC curve was made to find out cut off for the % fibrosis.

Results

The mean age of the indexed patients was 34.8 years (range 22-54 years), with a male predominance (M: F-10:1). Two independent observers graded the extent of steatosis, lobular inflammation, ballooning degeneration and extent

of fibrosis in this group of patients using the NASH CRN system and the same was repeated after 4 weeks using the same criteria. When compared between two observers the agreement pattern was as follows: steatosis ($K=-0.083$, No agreement); ballooning degeneration ($k= 0.267$, fair agreement); lobular inflammation ($k=0.06$, slight agreement) and fibrosis ($k= 0.317$, moderate agreement) [Table 1]. Similarly, the intra-observer variation was also prominent as only in 45.83% of the observations agreement could be achieved [Table 1].

Steatosis Grading: When the image analysis (IA) steatosis grade was compared and correlated to the conventional NASH CRN scoring system, there was no agreement between these two grading systems [Table 2]. Then the steatosis grades of NASH CRN system and IA steatosis grade were compared with the USG fatty score. Though there was a good agreement between IA steatosis grade and USG steatosis grades ($K= 0.65$; $P=0.02$), the visual estimation of steatosis did not show any agreement [Table 3]. The biochemical parameters as described in methodology were also correlated with steatosis grading by NASH-CRN system and IA steatosis scores. Though no agreement was noted between the biochemical changes

and visual grading of steatosis in liver biopsies; positive correlation was noted amongst steatosis grading by image analysis system and blood triglyceride level ($P 0.04$) and body mass index ($P 0.03$) [Table 3].

Fibrosis Staging: As already has been said that inter-observer agreement in staging of liver fibrosis on liver biopsies achieved moderate agreement [Table 1]. Fibroscan values was available in 24/48 patients of the study group. The visual fibrosis staging by NASH CRN staging system showed very low correlation with either image analysis fibrosis scores or fibroscan liver stiffness values [Table 4]. While the fibrosis scores as measured by IA technique and fibroscan values were in agreement ($r=0.76$; $P \text{ value}=0.01$) [Fig 3; Table 4]. Based on the receiver operating characteristic (ROC) curve findings, the fibrosis score by IA technique was classified as follows: No fibrosis/ grade 1 fibrosis-IA fibrosis score $<0.5\%$; Minimal fibrosis/ grade 2 fibrosis-IA fibrosis score $0.5-1\%$; Moderate fibrosis/ grade 3 fibrosis-IA fibrosis score $1-5\%$; severe fibrosis/ grade 4 fibrosis-IA fibrosis score $>5\%$ [Fig No.4; Table 4]. As the image analysis scores were repeated twice, there was minimal variations between two values in each case (average $CV=4.2\%$).



Fig. 1: A stitched digital image of a core biopsy of liver showing maintained lobular architecture (Portal tracts-pink; central vein-blue) [SR stain x40].

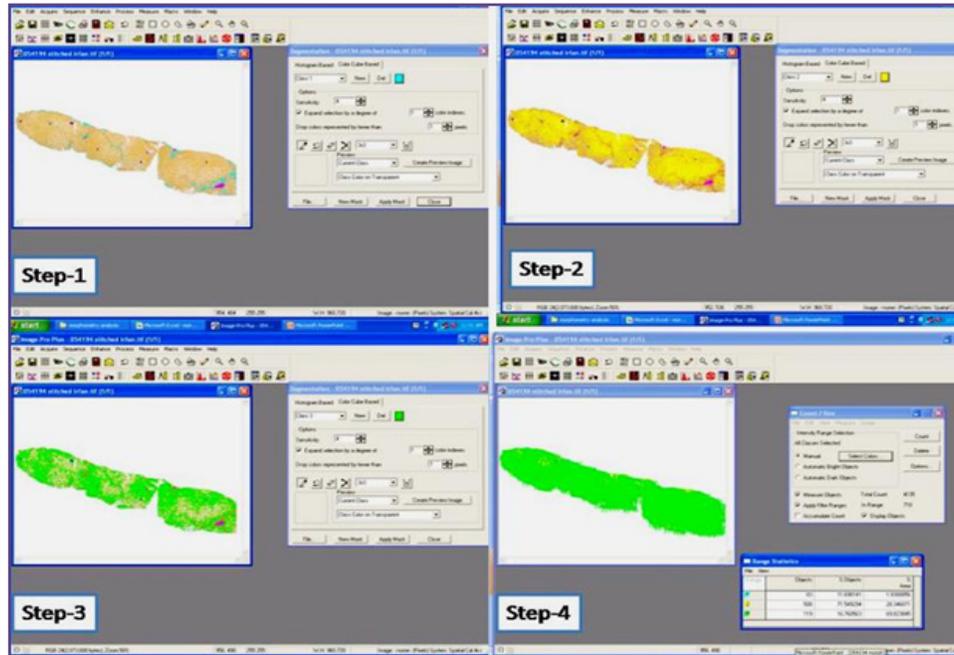


Fig. 2: Image processing: The Sirius red stained stitched core of liver was processed to exclude any background colour and the central hepatic and the portal veins were imparted specific colour to prevent any colour overlap during image analysis at a later stage. Further processing:

Step 1: The areas depicting fibrosis are assigned light blue colour; Step 2: The steatosis area was imparted yellow colour; Step 3: Hepatic parenchyma was imparted light green colour [Segmentation]. Step 4: The computer automatically generates the percentage area of each components by Image Proplus software version 6.1.

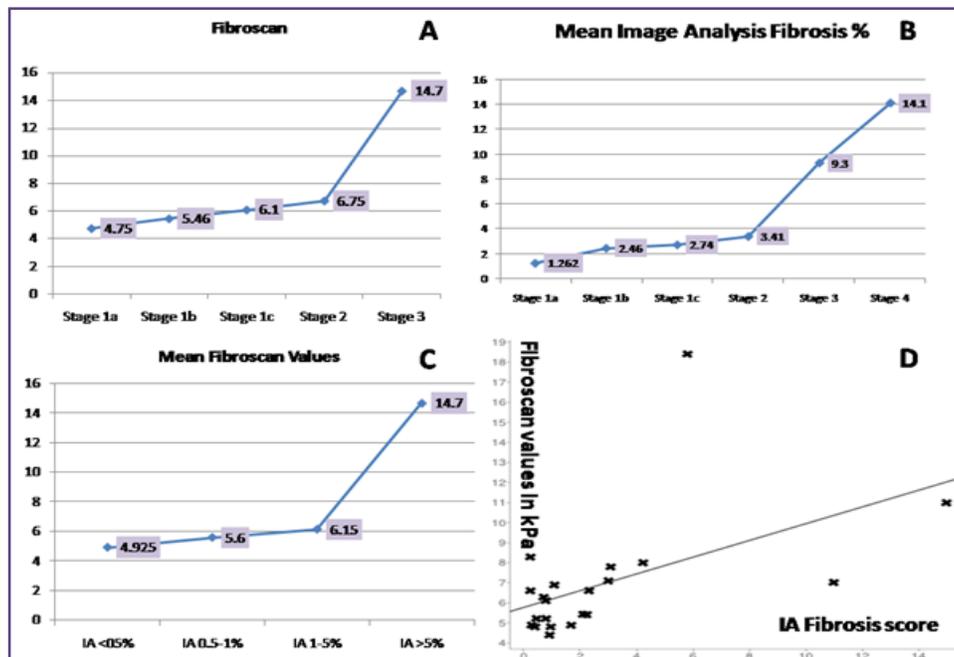


Fig. 3: Graph shows mean Fibroscan values in KPa (Y axis) observed in various stages of NASH (X axis) [Fig 3A]. Mean fibrosis % assessed by image analysis (Y axis) observed in various stages of NASH (X axis) [Fig 3B]. Mean Fibroscan values in KPa (Y axis) observed in various categories of IA analysis fibrosis % (X axis) [Fig 3C]. Linear regression plot (X axis: Fibrosis % score as assessed by image analysis; Y axis fibroscan values in kPa). Mean x:2.30, Mean y:7.25; intercept (a) 5.97; Slope (b): 0.551 [Fig 3D].

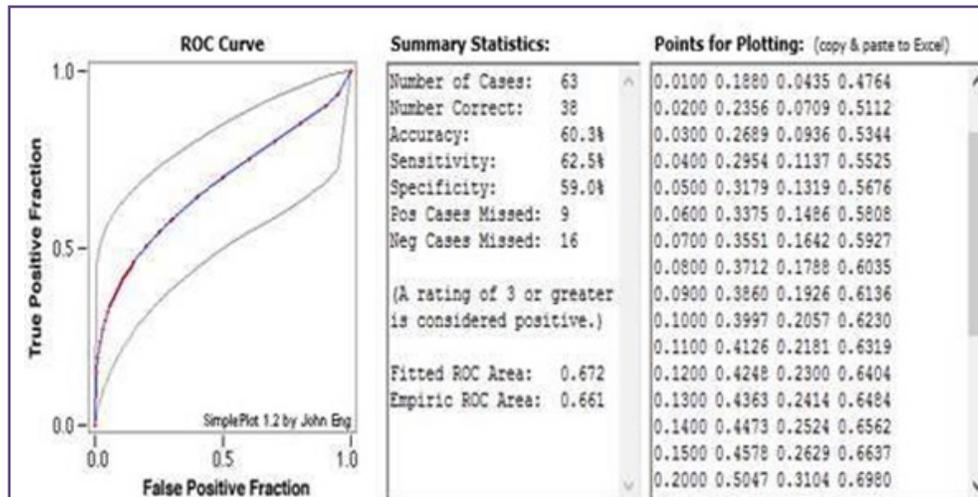


Fig. 4: ROC curve depicting the sensitivity, specificity and Area under Curve at the best cut offs. The cut offs achieved for fibrosis % are: (Grade 1: IA fibrosis % score <0.5%- normal); Grade 2 (IA fibrosis score 0.5-1%- minimal fibrosis); Grade 3 (IA fibrosis score 1-5%- moderate fibrosis) and Grade 4 (IA fibrosis score >5%- severe fibrosis).

Discussion

Grading and staging of NAFLD/NASH is the most important input what is expected from the histopathologists to comment on to understand the natural history of the disease, to determine possible outcome and to understand if the histological changes are only due to NASH or there is any associated disease. A widely acceptable semi quantitative grading and staging scheme was introduced by Brunt et al In 1999^[4] and was followed by the scoring system proposed by the NASH Clinical Research Network scoring system^[5]. This classification was however was designed to correlate the histological changes before and after therapeutic intervention trials. As has been demonstrated that there is wide variability of interpretation of disease activity grade and fibrosis stage amongst different histopathologists including intra-observer variability, a more objective scoring system is needed to be introduced. Amongst the grading parameters, observer agreement is generally good for steatosis, moderate for hepatocyte ballooning and slight for lobular inflammation or identification of Mallory's hyaline^[10]. In our study the agreement levels were: steatosis (K=-0.083, No agreement), ballooning degeneration (k=0.267, fair agreement), lobular inflammation (k=0.06, slight agreement) and fibrosis (k=0.317, moderate agreement). WHO characterized kappa values as follows: values < 0 - no agreement; 0-0.20- slight agreement; 0.21-0.40- fair agreement; 0.41-0.60- moderate agreement; 0.61-0.80- substantial and 0.81-1 as almost perfect agreement.^[11] The intra-observer variation was also rated as 'slight' using Kappa statistics with agreement in only 45.83% observations. Moreover, with the advent of radiological techniques non-invasive elastogram

was found gradually easier to perform in clinical practise, though the "gold standard" liver biopsy was continued to be taken. Serum biomarkers as described previously have also failed to gain user confidence due to heterogeneity in results.^[2] Julluri et al carried out a study to compare the scoring patterns in slides of NASH cases between a general pathologist and expert hepatopathologists in a blinded fashion. As expected there was very low level of agreement: - inter-observer agreement (kappa) between two pathologists was 0.62 for steatosis, 0.44 for lobular inflammation, 0.25 for ballooning, 0.40 for NAS and 0.35 for fibrosis. The two pathologists diagnosed "definite NASH" in a similar proportion of patients (56% vs. 57%), but their inter-observer agreement was only 0.46 (0.24-0.67) as they both diagnosed different levels of NASH (borderline vs. definite) in different subjects.^[6] Hence it was clearly apparent that we need to think beyond the light microscopic grading and staging system in NASH.

In recent years, computer-aided image processing and analysis systems have played a significant role in quantitative pathology as it adds subjectivity to the histological observations. After adapting the IA technique, the variation between two morphometric measurement scores was minimal in our study (average CV=4.2%). There was also very low level of agreement between scores of NASH CRN system and IA scores (k=-0.16). We also found that image analysis score of liver steatosis significantly matches with the ultrasonic fat estimation scores (k=0.65), in contrast to its variability with light microscopic steatosis area%. On applying paired 't' test between the estimated biochemical parameters and light microscopic and IA systems, the image analysis score

fares better [P values-TG (0.04), BMI (0.039)]. This signifies that the image analysis system is more accurate in grading of hepatic steatosis. During analysis it was realised that there is a tendency to overestimate the level of hepatic fat content on visual scoring. Visual grading (performed twice) in our study erroneously over-read the area of steatosis in 3/48 (6.25%) and 5/48 (10.4%) patients respectively, resulting in their inclusions in 'definite NASH category' and 'probable NASH category' respectively. Mei et al also compared visual score, image analysis score and total fat values in twenty-six patients. The results showed that individually, each method produced highly precise and reproducible measurements. Compared with each other, they showed very strong correlations (correlation coefficient $r=0.81-0.95$) unlike our study where the correlations were not appreciable ($P=0.09$). The authors found that visual estimation appeared to have a systematic bias, giving results nearly 4-fold higher than other methods as ranges of fat content in all 26 specimens were as follows: 2.2–15% by biochemical, 0.8–82.5% by visual, and 0.3–19.6% by morphometry method. This may be because visual estimation denotes the fraction of hepatocytes containing fat droplets, instead of the true fraction of fat.^[12]

Patients of NASH with advanced stage have a higher risk of developing complications such as portal hypertension or hepatocellular carcinoma, and therefore require close follow up. Transient elastography which was found to be useful for accurate estimation of hepatic fibrosis in patients with chronic hepatitis C has also been used to stage fibrosis in patients of NASH.^[13] However fibroscan has a limitation: liver tissue containing fat deposition is presumably softer than healthy liver parenchyma, and sound velocity has been reported to decrease as fatty liver progressed in an animal experiment.^[13] In our present study the IA fibrosis score was compared to the light microscopic NAS-CRN stage and fibroscan values. The study population was divided into 5 groups based on established NAS-CRN staging system. The Spearman correlation coefficient test showed that IA fibrosis score associated fairly well while light microscopic scoring associated moderately against the fibroscan values ($r=0.76, 0.55$ respectively). No association was noted between light microscopic staging and IA fibrosis scores ($r=0.49, P=0.06$). The fibrosis score as assessed by image analysis was plotted against the fibroscan values. A linear regression curve has been plotted for the first time between these variables. The same can be utilised for calculating the actual amount of fibrosis percentage in the liver based on non-invasive fibroscan values. Since image analysis fibrosis percentage was being calculated for the first time, a ROC curve was made to determine the cut offs and find out

the value of this new diagnostic modality. The area under curve was 0.672 signifying a moderately acceptable new diagnostic test. Sensitivity and specificity of IA technique calculated were 62.5%, 59% respectively [Fig No.4]. The cut offs derived for this highest sensitivity and specificity were as follows:- Grade 1 (IA fibrosis score $<0.5\%$ - normal); Grade 2 (IA fibrosis score $0.5-1\%$ - minimal fibrosis); Grade 3 (IA fibrosis score $1-5\%$ - moderate fibrosis) and Grade 4 (IA fibrosis score $>5\%$ - severe fibrosis).

Conclusion

We conclude that although the NASH CRN scoring system is robust, it suffers from high subjectivity both in histological grading and staging of fibrosis. Light microscopic assessment tends to overestimate hepatic fat content in liver biopsies. Computer assisted image analysis technique is an acceptable alternative to visual grading and staging with moderate sensitivity and specificity and shows better association with hepatic biochemical derangements and transient elastography findings. The entire morphometric procedure including digital image acquisition, image processing, segmentation and final analysis take 10 minutes a case and both free and commercially available software can be used. This technique hence seems a striking alternative to light microscopic grading and staging of liver biopsies in NAFLD and will drastically reduce the inter and intra- observer variations.

Acknowledgments

All laboratory technicians of Dept of Pathology, AIIMS, New Delhi

References

1. Ludwig J, Viggiano TR, McGill DB, Oh BJ. Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. *Mayo Clin Proc.* 1980 Jul;55(7):434-8.
2. LaBrecque DR, Abbas Z, Anania F, Ferenci P, Khan AG, Goh KL et al; World Gastroenterology Organisation. World Gastroenterology Organisation global guidelines: Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *J Clin Gastroenterol.* 2014 Jul;48(6):467-73.
3. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. *Gastroenterology.* 1999 Jun;116(6):1413-9.
4. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. *Am J Gastroenterol.* 1999 Sep;94(9):2467-74.
5. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al; Nonalcoholic Steatohepatitis Clinical Research Network. Design and validation of a

- histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. 2005 Jun;41(6):1313-21.
6. Juluri R, Vuppalanchi R, Olson J, Unalp A, Van Natta ML, Cummings OW et al. Generalizability of the nonalcoholic steatohepatitis Clinical Research Network histologic scoring system for nonalcoholic fatty liver disease. *J Clin Gastroenterol*. 2011 Jan;45(1):55-8.
 7. Obika M, Noguchi H. Diagnosis and evaluation of nonalcoholic fatty liver disease. *Exp Diabetes Res*. 2012;2012:145754.
 8. Dowman JK, Tomlinson JW, Newsome PN. Systematic review: the diagnosis and staging of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis *Aliment Pharmacol Ther*. Mar 2011; 33(5): 525–40.
 9. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: a review. *IEEE Rev Biomed Eng*. 2009; 2:147-71.
 10. Hübscher SG. Histological assessment of non-alcoholic fatty liver disease. *Histopathology*. 2006 Nov;49(5):450-65.
 11. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol*. 2008 May;61(Pt 1):29-48.
 12. Li M, Song J, Liu W. Comparing Morphometric, Biochemical and Visual Measurements of Macrovesicular Steatosis of Liver. *Hum Pathol*. Mar 2011; 42(3): 356–60.
 13. Yoneda M, Fujita K, Hiraishi H. Transient elastography in patients with non-alcoholic fatty liver disease (NAFLD). *Gut*. Sep 2007; 56(9): 1330–1.

***Corresponding author:**

Gp Capt Prateek Kinra, Dept Of Pathology, AFMC, Pune-411040, India

Phone: +91 9945277110

Email: pkinra_in@yahoo.com

Financial or other Competing Interests: None.