Semiquantitative evaluation overestimates the degree of steatosis in liver biopsies: a comparison to stereological point counting

Lennart E Franzén[1], Mattias Ekstedt[2], Stergios Kechagias[3] and Lennart Bodin[4]

[1]*Department of Pathology, Clinical Research Centre, University Hospital, Örebro, Sweden;* [2]*Department of Molecular and Clinical Medicine, Division of Gastroenterology and Hepatology, Linköping University, University Hospital, Linköping, Sweden;* [3]*Department of Medicine and Care, Division of Internal Medicine, Linköping University, University Hospital, Linköping, Sweden and* [4]*Unit of Biostatistics and Epidemiology, Clinical Research Centre, University Hospital, Örebro, Sweden*

The degree of steatosis in liver biopsies is usually assessed by a morphological semiquantitative approach in which the histopathologist uses a four-graded scale: 0–3 or none, slight, moderate and severe. Scores 1–3 are considered to correspond to fat deposition in <33, 33–66 and $>66\%$ of the hepatocytes. There is a considerable inter- and intra-individual variation in such scoring methods and a more standardized and quantitative approach is preferable. In the present study, we compare the semiquantitative technique with the stereological point counting method in the assessment of hepatic steatosis. A total of 75 archived liver needle biopsies were used. They were selected according to the original routine diagnosis of slight, moderate or severe steatosis. In all, 10 randomly selected images from each biopsy were digitized into a computer, a point grid lattice was superimposed and the number of hits on fat globules was counted. A pathologist scored the specimens in a four-graded scale as described above. The mean liver biopsy area (volume) with fat in hepatocytes was 2.2% for grade 1, 9.2% for grade 2 and 23.1% for grade 3. The kappa value for the semiquantitative estimates was 0.71 for the unweigthed kappa and 0.87 for weighted kappa. The intraclass correlation coefficient (ICC) was 0.99 for images counted twice and 0.95 when two sets of images were captured from the same biopsy. These ICCs indicate excellent agreement and above that of the semiquantitative estimates. In conclusion, the area/volume of fat content of the hepatocytes is greatly overemphasized in semiquantitative estimation. Furthermore, the point counting technique has a better reproducibility than visual evaluation and should be preferred in estimates of liver steatosis in scientific studies and in clinical contexts when the amount of steatosis is important for treatment and prognosis, such as liver transplantation.
Modern Pathology (2005) **18,** 912−916. doi:10.1038/modpathol.3800370; Published online 27 May 2005

Keywords: liver; quantification; steatosis

Steatosis is a common finding in liver biopsies both clinically and in experimental studies and the extent of fat accumulation is often asked for. The degree of steatosis is commonly assessed by visually and semiquantitatively estimating the area of the section of a needle biopsy that is occupied by fat vacuoles.[1] The result is often expressed as 'the percent of hepatocytes in the biopsy involved' and can be 0–3 where 0 is none, 1 is up to 33%, 2 is 33–66% and 3 is $>66\%$[2] and such grading is also recommended by the American Gastroenterological Association when assessing nonalcoholic fatty liver disease.[3] However, it should be pointed out that actual counting of hepatocytes with fat vacuoles is not performed and that the basis of the grading is the estimated area of fat vacuoles (Brunt EM, 2003, personal communication).

In several studies, objective computer-assisted methods[4,5] and stereological point counting[6,7] have been used to quantify the degree of steatosis. These studies have shown that usually not more than 10–20% of the sectional area or volume density is occupied by fat vacuoles.[4,5] The conception of the amount of fat in the liver therefore appears to be greatly overemphasized when graded semiquantitatively. Furthermore, there is a substantial inter- and intra-individual variation regarding the degree of steatosis using semiquantitative scoring.[4,8–11]

The aim of the present study was to assess in further detail the discrepancies between semiquantitative scoring of liver steatosis and morphometric/stereological assessment of the area fraction of steatosis, which equals the volume fraction of steatosis. The reproducibility of the two techniques was also determined using kappa analysis and the intraclass correlation.

Materials and methods

A total of 75 liver biopsies from archived slides stained with hematoxylin–eosin were used. They were selected according to the original grade of steatosis diagnosed, twenty-five of each slight, moderate and severe steatosis (grades 1, 2 and 3). The local ethics committee in Örebro, Sweden approved the study.

The specimens were blinded and evaluated twice regarding grade of steatosis (both macro- and microvesicular) by the pathologist (LF). The interval between the evaluations was 2 months. The degree of steatosis was graded 0–3 based on area of the section of a needle biopsy that was occupied by fat vacuoles.[1] Biopsies without steatosis were not included in this study and therefore no specimens have score 0.

A Leica DMRXA 2 microscope with a Leica DC 200 digital camera was used for image capturing. In all, 10 images from each biopsy were captured and stored in a computer using the software Adobe Photoshop 6.0. The first field of view was chosen in the end of the biopsy closest to the end of the microscopic slide. After the first image had been grabbed, the next field of view was chosen by moving along the length axis of the biopsy 1.25 fields of view in order not get overlapping images for evaluation. This procedure was continued until 10 images had been grabbed. A point grid consisting of 100 crosses, 35 μm apart, was superimposed on each image. The final magnification on the computer screen when counting was $\times 400$. The number of hits on fat vacuoles in hepatocytes (including both macro- and microvesicular) and normal hepatocytes was counted. Ballooned hepatocytes were omitted. They can almost always be separated from macrovesicular steatotic cells due to centrally placed nuclei and fragmented membranes or cell organelles in the cytoplasm. Hits on damaged tissue and larger areas with connective tissue were excluded. The results are given as the percentage of biopsy area

with fat deposition. Images from 20 randomly chosen specimens were recounted to assess the reproducibility of the point counting in the same images.

In all, 20 specimens were then selected randomly and a new set of 10 images was captured from each of these specimens. These new images were counted as above and the results were used to assess the reproducibility of the point counting technique when new images were resampled.

Statistics

Agreement for the scoring results was analysed by the kappa coefficient, both in the unweigthed and the weigthed form, the weights chosen as quadratic weights.[12] The percentage of absolute agreement was also calculated. The estimates of agreement were supplemented with 95% confidence intervals (95% CI) to account for sampling variability.

For the point counting technique, which is a purely quantitative variable, we calculated the intraclass correlation coefficient (ICC), to estimate agreement,[13] supplemented with 95% CI. The formulas for and the interpretation of the weighted kappa with quadratic weights and the ICC are similar as stated by Fleiss,[12] and values close to or above 0.75 signify very well to excellent agreement.
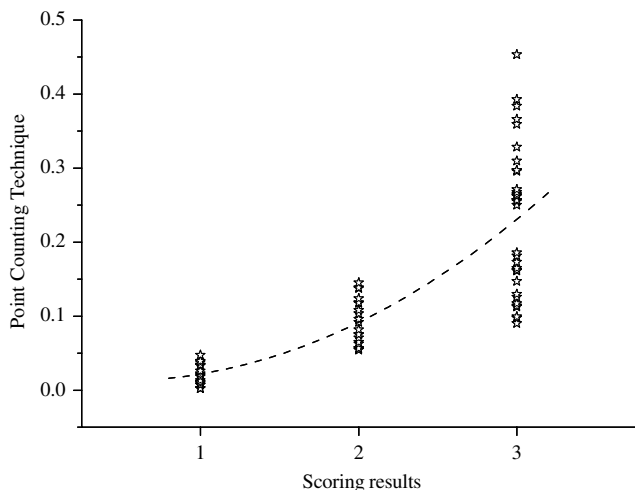
Results

The evaluation of the degree of steatosis by the semiquantitative approach resulted in the following scoring: 21 specimens obtained score 1, 20 specimens score 2 and 34 specimens score 3. This is to compare with the original scoring with 25 specimens in each score group.

The mean values of the point counting as well as the minimum and maximum values for each scoring group are shown in Table 1. The coefficients of variation regarding the percentage of the fat vacuole density between images taken from the same biopsy was calculated and found to be high (Table 1). This indicates an uneven distribution of steatosis in individual liver specimens. Since these coefficients of variation decreased with higher steatosis score, the uneven distribution of steatosis was especially seen with low-grade steatosis. When the scoring results were correlated to those obtained with the

Table 1 Basic characteristics of the point counting technique, stratified for three grades of the scoring results

	Grade 1	Grade 2	Grade 3
% Area fat globules Mean (minimum–maximum)	2.2 (0.2–4.7)	9.2 (5.4–14.5)	23.1 (9.0–45.3)
Coefficient of variation (%)	134	69	47

Mean values, range and coefficient of variation for percentage area of fat globules.

914

Figure 1 Scatterplot for point counting technique *vs* scoring results. Second degree polynomial fitted to the observations, – – – – – – –.

point counting technique (Figure 1), a polynomial of the second degree gave a good fit to the data.

A substantial overlap ($n = 18$) was found regarding score groups 2 and 3 (Figure 1). The overlapping specimens were re-examined and compared in order to find some morphological characteristic that made them scored to either group irrespective of the amount of fat globules measured with the point counting. No such features could be found. The degree of fibrosis did not differ between the groups 2 and 3 or between the specimens that were overlapping (data not shown). Nor were there any differences regarding the coefficients of variation, that is, uneven distribution of fat, in the specimens that were overlapping. No specimens were overlapping between degrees 1 and 2 (Figure 1).

The semiquantitative scoring was performed twice, 2 months apart and, the agreement was 81% (95% CI 72–90%) and the unweighted kappa was 0.71 (95% CI 0.58–0.85). Weighted kappa with quadratic weights was 0.87 (95% CI 0.81–0.94). When the images of 20 randomly chosen specimens were reassessed by point counting and compared to the initial counting, the ICC value was 0.99 (95% CI 0.98–1.00). In all, 10 images from each of 20 randomly chosen specimens were captured a second time and the concordance calculated. The ICC value was found to be 0.95 (95% CI 0.87–1.00).

Discussion

Steatosis is a frequent finding in liver biopsies and the extent of involvement is often asked for. Traditionally the degree of steatosis is assessed through a visual and semiquantitative estimation of the area of the section occupied by fat globules in hepatocytes[1] or the 'percent of hepatocytes' in the biopsy involved.[2] An alternate approach is to assess

the area of the section involved with steatosis, either by stereological point counting or by image analysis with thresholding of areas with fat globules. In the present study, we scrutinized the discrepancies between semiquantitative grading and point counting and the reproducibility of the methods.

A few earlier studies have considered and compared the semiquantitative and the quantitative approaches to assess the degree of steatosis in liver biopsies. Auger *et al*[4] used 2 μm thick sections from plastic-embedded tissue stained with hematoxyline–eosin. The automated analysis was based on thresholding of unstained areas, automatic omission of sinusoidal empty spaces with a 'form factor' and manual exclusion of vessels with red blood cells. The semiquantitative grading was performed by two pathologists estimating the percentage of fatty hepatocytes using a 10-graded scale. They found that the pathologists scoring varied between 0 and 80% whereas the automatic calculated densities were much lower and varied between 0 and 15%. Kumar *et al*[5] studied biopsies from patients with hepatitis C and their highest semiquantitative score was 2. They found that the maximum value of score 2 was approximately 11% area of the biopsy occupied by fat. The correlation between per cent steatosis and histological grade was high ($r = 0.87$). In this study, we found a mean fat density value of 23.1% and a maximum value of 45.3% in the score 3 group. These studies thus show that the semiquantitative estimates greatly overemphasize the true density of fat accumulation in the liver tissue.

The quantitative methods used for the estimation of fat accumulation in liver have been image analysis with thresholding and stereological point counting. The thresholding technique demarcates the unstained areas in the sections and these are mostly fat globules. One drawback with this approach is that sinusoids and, vessel and bile duct lumina also remain unstained. The sinusoids can be excluded automatically since their 'form factor' obtained in image analysis differ significantly from fat globules but vessels and bile duct lumina have to be manually excluded from the measuring area.[4,5] One way to circumvent this problem would be to use osmium tetroxide as a fat-specific stain and embed the tissue in resin. With this procedure fat is stained black.[14] However, this requires that the biopsies are handled separately from routine biopsy material and one also has to consider that osmium tetroxide is hazardous to handle.

The point counting technique has also been used to quantify liver steatosis.[6,7,15] The technique is simple and can be performed either by feeding images into the computer and use an overlay lattice as in this study or supplying one eye-piece of the microscope with a graticule with a point lattice and do the counting directly in the microscope. Different structures in the liver are identified immediately and can be included or excluded from counting. Another advantage with the point

counting technique is that the area of fibrosis can be assessed at the same time by counting points hitting the connective tissue. We therefore prefer point counting technique since it is simpler, no camera for capturing images and computer is needed, it measures the area/volume fraction of fat and do not require intervention such as manual exclusion of structures.

Reproducibility is of great importance in all quantification. Semiquantitative scoring methods mostly show moderate or low reproducibility and high inter- and intraobserver variability. A total of 41 morphological characteristics were evaluated by two pathologists in liver biopsies from 362 alcoholic patients.[8] The kappa values ranged from 0.13 (size of liver cell) to 1 (presence of hepatocellular carcinoma). Degree of steatosis was agreed upon in 48% of the cases and the kappa value was 0.35. The interobserver variation for steatosis was $k = 0.63$ in a study by the French METAVIR study group[9] and others have presented interindividual kappa values for steatosis of 0.508[10] and 0.64.[11] The intraindividual variation also shows low kappa values for the degree of steatosis, kappa $= 0.42$[9] and 0.63–0.65.[11] In the present study, the agreement between the two semiquantitative evaluations was 81% and the unweighed kappa value 0.71. This value is similar or slightly higher than previously reported. The ICC value, which is equivalent to the kappa value, was 0.99 when 25 cases were re-evaluated with point counting. When sets of new images were captured randomly from the 25 specimens and the point counting results were compared to those of the first evaluation, the ICC value was still very high, 0.95. These results thus show that the point counting technique has a very high reproducibility and is superior to the semiquantitative scoring methods.

A very precise measurement of the degree of liver steatosis is mostly of interest in experimental studies and usually not needed in clinical practice. However, in orthoptic liver transplantation, the degree of steatosis of the donor transplant significantly influences the outcome[16,17] possibly due to impairment of hepatic microcirculation.[18] A retrospective study of steatosis in cadaveric liver transplants was recently performed by Marsman et al[19] in which they compared automated measurement of steatosis with the semiquantitative grading of a pathologist. As has been shown earlier and discussed above, the pathologist was found to constantly overestimate the degree of steatosis. They conclude that an automated analysis system can be used to determine the fat content in liver biopsies but that further studies are needed to determine the role of such a technique in the evaluation of donor livers for transplantation. They used specimens stained with hematoxyline–eosin as in earlier studies with automated analysis[4,5] but nothing is mentioned how hollow structures with lumina such as sinusoids, vessels and bile ducts were omitted from the measurements.

In conclusion, the present study highlights the difference in reproducibility of scoring methods and point counting methods when assessing the degree of steatosis in liver biopsies. The point counting technique is simple and superior to scoring techniques and is also preferable to the image analysis thresholding method. Point counting can be used when accurate measurements of liver steatosis are required such as in liver biopsies from prospective transplant livers and in studies involving histological evaluation of the liver pathology.

Acknowledgements

References

1 Mathiesen UL, Franzén LE, Åselius H, et al. Increased liver echogenicity at ultrasound examination reflects degree of steatosis but not of fibrosis in asymptomatic patients with mild/moderate abnormalities of liver transaminases. Digest Liver Dis 2002;34: 516–522.

2 Brunt EM, Janney CG, Di Bisceglie AM, et al. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gastroenterol 1999;94:2468–2474.

3 AGA technical review on non-alcoholic fatty liver disease. Gastroenterology 2002;123:1705–1725.

4 Auger J, Schoevaert D, Martin ED. Comparative study of automated morphometric and semiquantitative estimations of alcoholic liver steatosis. Analyt Quant Cytol Histol 1986;8:56–62.

5 Kumar D, Farrell GC, Fung C, et al. Hepatitis C virus genotype 3 is cytopathic to hepatocytes: reversal of hepatic steatosis after sustained therapeutic response. Hepatology 2002;36:1266–1272.

6 Koch MM, Freddara U, Lorenzini I, et al. A stereological and biochemical study of the human liver in uncomplicated cholelithiasis. Digestion 1978;18: 162–177.

7 Zaitoun AM, Al Mardini M, Record CO. Stereology and morphometry of steatosis in human alcoholic (ALD) and non-alcoholic liver disease (NALD). Acta Stereol 1998;17:49–56.

8 Bedossa P, Poynard T, Naveau S, et al. Observer variation in assessment of liver biopsies of alcoholic patients. Alcoholism: Clin Exp Res 1988;12:173–178.

9 The French METAVIR Cooperative Study Group. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. Hepatology 1994;20:15–20.

10 Theodossi A, Skene AM, Portmann B, et al. Observer variation in assessment of liver biopsies including analysis by kappa statistics. Gastroenterology 1980;79: 232–241.

11 Younossi ZM, Gramlich T, Chang Liu Y, et al. Nonalcoholic fatty liver disease: assessment of varia-

bility in pathologic interpretations. Mod Pathol 1998;
11:560–565.

12 Fleiss JL, Levin B, Cho Paik M. Statistical Methods for Rates and Proportions, 3rd edn. New Jersey: Wiley Interscience, 2003.

13 Dunn G. Design and Analysis of Reliability Studies. The Statistical Evaluation of Measurement Errors. New York: Oxford University Press, 1989.

14 Hall P, Gormley BM, Jarvis LR, *et al.* A staining method for the detection and measurement of fat droplets in hepatic tissue. Pathology 1980;12:605–608.

15 Zaitoun AM, Al Mardini H, Awad S, *et al.* Quantitative assessment of fibrosis and steatosis in liver biopsies from patients with chronic hepatitis C. J Clin Pathol 2001;54:461–465.

16 Fernandez-Merino FJ, Nuno-Garza J, Lopez-Hervas P, *et al.* Impact of donor, recipient, and graft features on the development of primary dysfunction in liver transplants. Transplant Proc 2003;35:1793–1794.

17 Verran D, Kusyk T, Painter D, *et al.* Clinical experience gained from the use of 120 steatotic donor livers for orthoptic liver transplantation. Liver Transpl 2003;9:500–505.

18 Ijaz S, Yang W, Winslet MC, *et al.* Impairment of hepatic microcirculation in fatty liver. Microcirculation 2003;10:447–456.

19 Marsman H, Matsushita T, Dierkhishing R, *et al.* Assessment of donor liver steatosis: pathologist or automated software? Hum Pathol 2004;35:430–435.