

Original Article

Sampling Error in Histopathology Findings of Nonalcoholic Fatty Liver Disease: A Post Mortem Liver Histology Study

Shahin Merat MD¹, Rasoul Sotoudehmanesh MD¹, Mehdi Nourai MD^{1,2}, Masoumeh Peikan-Heirati MD¹, Sadaf G.Sepanlou MD¹, Reza Malekzadeh MD¹, Masoud Sotoudeh MD¹

Abstract

Background: Many clinical trials and natural history studies on nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH) rely heavily on liver histology to define their endpoints. There are many indications that the liver is not uniformly involved in NAFLD thus sampling error is a major concern. This study aims to evaluate the uniformity of various histologic features in livers affected with NAFLD.

Methods: Samples from a forensic autopsy series were studied and subjects with NAFLD identified. We took specimens from three different parts of each liver and recorded the degrees of steatosis, hepatocyte ballooning, lobular inflammation, portal inflammation, and fibrosis. A NASH activity index (NAI) which is the sum of scores of histologic features was also calculated. The agreement between the 3 samples from each liver was studied.

Results: There were 945 autopsies performed; 896 were suitable for histologic evaluation and 283 had NAFLD. Of these, 146 livers were available to our study from which 438 samples were taken. Fibrosis[intra-class correlation (ICC) = 0.87], lobular inflammation (kappa = 0.83), and portal inflammation (kappa = 0.83) were fairly uniformly distributed in the damaged liver. Steatosis was less uniform (kappa = 0.64), and hepatocyte ballooning was least uniformly distributed (kappa = 0.57). The ICC for NAI was 0.86, which indicated good agreement.

Conclusions: The individual histologic features of NAFLD and NASH are not uniformly distributed in the liver. Hepatocyte ballooning is especially non-uniform. Such non-uniformity should be taken into account when interpreting results of studies that rely on paired biopsies. A summary score such as NAI is less affected by sampling error.

Keywords: Autopsy, biopsy, fatty liver, pathology

Cite this article as: Merat S, Sotoudehmanesh R, Nourai M, Peikan-Heirati M, Sepanlou SG., Malekzadeh R, et al. Sampling error in histopathology findings of nonalcoholic fatty liver disease: a post mortem liver histology study. *Arch Iran Med.* 2012; **15**(7): 418 – 421.

Introduction

Nonalcoholic steatohepatitis (NASH) is currently the most common chronic liver disease in the United States and many other countries, including Iran.^{1,2} NASH may progress to chronic end stage liver disease and it is anticipated that this disease would become the most common etiology for liver-related mortality in the near future.³

Although NASH was first described clinically by Ludwig et al. in 1980,⁴ its diagnosis and staging are still mainly histological. Many studies on the natural history of NASH⁵ and numerous clinical trials^{6,7} rely on liver histology to prove their exactness. Noninvasive diagnostic tests such as imaging techniques⁸ or serum and genetic markers⁹ are not sensitive and specific enough to detect and stage NASH.¹⁰ Thus, liver biopsy is not only critical in defining the diagnosis and prognosis of NASH, but also remains pivotal in the evaluation of therapeutic measures.

Despite the elaborate systems devised for grading and staging of liver biopsies,¹¹⁻¹³ there is still no precise histological definition for nonalcoholic fatty liver disease (NAFLD) or NASH.⁵ Errors

in the assessment of liver histology might originate from interobserver variability, small size of biopsy samples,^{14,15} and the fact that NASH does not uniformly affect the liver mass. In one study, biopsies from both lobes of the liver obtained during bariatric surgery revealed a 50% disagreement of fibrosis stage.¹⁶ Sampling error in paired samples and interobserver discordance has been reported in several other studies, both for NAFLD and other liver diseases.^{15,17-23}

In studies evaluating paired biopsy samples, such as studies on treatment^{6,24,25} or natural history,^{26,27} the uncertainty in evaluating liver histology is a serious problem that potentially undermines the validity of the results.

In order to clarify the degree of sampling error in NAFLD subjects, we have compared histological features in large samples taken from different parts of the same post-mortem liver in a relatively large number of subjects with NAFLD.

Materials and Methods

We analyzed samples collected during a forensic autopsy series performed on 945 subjects in Tehran.²⁸ Autopsy was performed within 24 hours of death unless the death occurred on an official holiday, where in the autopsy was performed in 48 hours. Samples with autolysis were excluded.

Family members of subjects were contacted and information on alcohol intake and possible known liver disease was sought. Sub-

Authors' Affiliations:¹Digestive Disease Research Center, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran, ²Howard University, Department of Internal Medicine and Center for Sickle Cell Disease, Washington, District of Columbia, United States.

•**Corresponding author and reprints:** Masoud Sotoudeh MD, Sharitai Hospital, North Kargar Street, Tehran 14117, Iran. Tel: +98 21 8241 5300, Fax: +98 21 8241 8400, E-mail: msotoudeh@tums.ac.ir
Accepted for publication: 23 November 2011

Table 1. Scoring of histology findings in NAFLD.*

Variable	Score	Description
Steatosis	0	None
	1	Up to 33% of acini, mainly macrovesicular
	2	34%–66% of acini, commonly mixed steatosis
	3	Over 66% of acini (panacinar), commonly mixed steatosis
Hepatocyte ballooning	0	None
	1	Occasional in zone III
	2	Obvious in zone III
	3	Marked, predominantly in zone III
Lobular inflammation	0	None
	1	Scattered neutrophils, occasional mononuclear cells, 1 or 2 foci per 20x objective.
	2	Neutrophils associated with ballooned hepatocytes, mild chronic inflammation, 3 or 4 foci per 20x objective
	3	Acute and chronic inflammation, neutrophils may concentrate in zone III, over 4 foci per 20x objective
Portal inflammation	0	None
	1	Mild, some portal areas
	2	Mild to moderate, most portal areas
	3	Moderate to severe, most portal areas
Stage	0	No fibrosis
	1	Zone III perivenular, perisinusoidal (pericellular) fibrosis
	2	Stage 1 changes + periportal fibrosis
	3	Bridging fibrosis
	4	Cirrhosis

* From Merat et al.

jects with known liver disease and those with a history of alcohol intake greater than 40 grams per week were excluded.

In the original study, 49 liver samples were unusable due to autolysis and 283 (31.6%) had steatosis. From this number, only 146 livers were available for our study.

Large liver specimens, 2×2×2 cm each, were taken from 3 different parts of each liver; including left, right and caudate lobes, avoiding the capsule by at least 1 cm. Specimens were fixed in 10% neutral buffered formalin and processed, sectioned, and stained with hematoxylin and eosin following standard procedures. A subject was considered as having NAFLD and included in our study only if macrovesicular steatosis affected at least 5% of the hepatocytes in any of the 3 specimens. Further staining with Masson's trichrome, and Sweet's reticulin was performed on specimens from subjects with NAFLD.

Histological sections were examined by a single pathologist who was unaware of which samples belonged to the same liver. Four major histological features of steatohepatitis were recorded: steatosis, hepatocyte ballooning, lobular inflammation, and portal inflammation. Each feature was scored from 0 to 3 following the criteria adopted from Brunt et al. and according to the modification introduced by Merat et al.^{11,12} A NASH activity index (NAI) was calculated by summing the scores for these features, which yielded a number between 0 and 12 as described previously.¹¹ Fibrosis was separately scored from 0 to 4 according to Brunt et al.¹² The scoring of histological features is detailed in Table 1. We also calculated the NAFLD activity score (NAS) as proposed by Kleiner et al.¹³

Statistical methods

In order to evaluate agreement between the 3 samples from each liver, multi-rater kappa was used for each histological variable;

intra-class correlation (ICC) coefficient was calculated for NAI, NAS, and fibrosis. Values of 1 indicated perfect agreement; greater than 0.8 were considered as excellent; between 0.6 and 0.8 good; between 0.4 and 0.6 moderate; between 0.2 and 0.4 fair; and lower than 0.2 indicated slight agreement. Statistical analyses were performed using SPSS 18.0 (SPSS Inc., Chicago, IL) and STATA 10.1 (Stata Corp, College Station, TX).

Results

There were 146 livers with NAFLD suitable for histologic evaluation. Subjects included 128 males and 18 females. Mean age was 47 years (range: 11 – 94 years). According to data provided by family members, 46 subjects were smokers and 26 were drug abusers (mainly inhalational opium). None were alcohol abusers. The causes of death were myocardial infarction in 49, trauma and accidents in 36, and opium overdose in 14. Other causes of death included cerebrovascular accidents, infection, and homicide. None of the subjects had known liver disease.

Samples were scored as described in Methods and results compared. We calculated the multi-rater kappa for each of the histologic variables (Table 2). The results indicated that fibrosis, lobular inflammation, and portal inflammation were somewhat uniformly distributed in the damaged liver whereas steatosis and, particularly, hepatocyte ballooning was not.

The ICC for NAI from the 2-way random model was 0.86 (95% CI = 0.82 – 0.89, $P < 0.0001$) and for fibrosis was 0.87 (95% CI = 0.83 – 0.90, $P < 0.0001$), both indicating good agreement.

Agreement was generally better for higher scores indicating that more severe damage is more uniformly distributed, e.g. severe steatosis was more uniform than mild steatosis (Table 2). The promi-

Table 2. Multi-rater kappa for correlation of various histology features between 3 samples taken from the same liver.*

	Score 0	Score 1	Score 2	Score 3	Overall
Steatosis	NA	0.66	0.52	0.81	0.64
Hepatocyte ballooning	0.68	0.44	0.58	0.50	0.57
Lobular inflammation	0.50	0.83	0.83	0.90	0.83
Portal inflammation	0.68	0.84	0.80	0.89	0.83

*NA: Not applicable.

ment exception was hepatocyte ballooning where the best agreement was observed for score 0 (absence of ballooning).

As hepatocyte ballooning had the worst agreement, we calculated the ICC for the sum of portal inflammation, lobular inflammation, and steatosis, excluding hepatocyte ballooning. The resulting ICC was 0.87 (95% CI = 0.84 – 0.90, $P < 0.0001$), not significantly better than NAI.

Despite the poor agreement for hepatocyte ballooning, this feature is an important characteristic of liver damage in NASH. As seen in Table 2, the absence of hepatocyte ballooning (score 0) was fairly uniformly distributed, thus we recoded this variable as 0 for absence and 2 for any degree of hepatocyte ballooning. The multi-rater kappa for the recoded variable was 0.68 and when substituted into NAI the resultant ICC was 0.84 (95% CI = 0.80 – 0.88, $P < 0.0001$).

Staging hepatocyte ballooning from 0 to 2 as proposed by Kleiner et al.¹³ (instead of 0 to 3 as used in NAI) did not change the multi-rater kappa for this variable which remained at 0.57. The ICC for NAS was 0.83 (95% CI = 0.79 – 0.87, $P < 0.0001$), not significantly worse than NAI. NAS was originally developed to make the diagnosis of “NASH” vs. “not NASH” based on histologic scoring.¹³ Using this criteria, the diagnosis of NASH (NAS > 4) had good agreement when various sections of the same liver were evaluated (multi-rater kappa = 0.74).

Discussion

Liver biopsy is an invasive procedure that carries a small risk of complications not considered acceptable by many patients and physicians. It is also subject to errors in interpretation, and more importantly, sampling. Sampling error in liver disease is well recognized and may lead to a misdiagnosis of cirrhosis in up to 30% of subjects.²⁹

As seen in Table 2, the worst agreement was noted for hepatocyte ballooning. Interestingly, during the validation of NAI, the worst inter- and intra-observer agreements also belonged to this histologic feature.¹¹ This might indicate that the poor agreement we observed for hepatocyte ballooning in different samples of the same liver might be due to rater variability rather than true non-uniform involvement of the liver. In either case, hepatocyte ballooning, although an important diagnostic criterion for NASH, might not be a suitable variable for scoring liver histology.

It would be expected that excluding hepatocyte ballooning from the total score of NAI should result in a much better agreement. However as previously mentioned in the results, we did not observe any improvement in ICC. The number of zero scores (no hepatocyte ballooning) in our data was high (78% vs. 21% for other variables) and the absence of ballooning (score 0) had the best kappa score among various severities of hepatocyte ballooning (Table 2). Thus by excluding this variable we also lost the good agreement observed in “no hepatocyte ballooning”; this probably

explains why we did not observe a better ICC by excluding hepatocyte ballooning from NAI. The large number of zero scores in hepatocyte ballooning probably also explains why we did not observe better ICC when recoding this variable from 4 levels (0 to 4) to 2 levels (0 and 2).

We observed that despite the variability of the individual histologic features, there was a better agreement in NAI between different samples from the same liver. Apparently the variability in individual features was diluted in the summary score that has defined NAI. Thus, NAI might be a more suitable indicator of general liver damage than each of the individual features.

It should be noted that we have analyzed large tissue samples (> 2×2×2cm). In clinical settings such tissue size will not be available. Therefore in real-life the agreement could be even less. This might explain the inconsistencies seen in studies that depend on paired liver biopsies to define disease progression.

Other researchers have also studied sampling variation in NASH. Arun et al. have described a series of 31 obese patients in each of which 2 biopsies were taken from the left liver lobe. They observed that portal fibrosis had the greatest sampling discordance followed by ballooning degeneration.²² Larson et al., who have studied biopsies from the right and left liver lobes of 41 morbidly obese patients reported the greatest degree of sampling error in lobular inflammation followed by ballooning necrosis.²³ Both studies concluded that the agreement between various features was good. It should be noted that they only included morbidly obese patients which generally have more advanced liver lesions. Our data have also confirmed that advanced lesions are more uniformly distributed throughout the liver (Table 2). Most studies used core needle biopsies which yielded much smaller samples than our study and have smaller numbers of subjects. Those studies only compared 2 sections of each liver, not 3, as we did. The large sample sizes used in our study allowed us to conclude that the variations we observed were real and not a side effect of small sample sizes of needle biopsies.

We did not have detailed clinical data on our subjects. But as those with known liver disease have been excluded, it is relatively safe to assume that the prevalence of various liver diseases among our subjects is comparable to the general population. Recent studies on the prevalence of liver disease in the general population of Iran indicate a seroprevalence of 2.6% for hepatitis B followed by 0.5% for hepatitis C.^{30, 31} The prevalence of other liver diseases is much lower. Thus it is logical to assume that over 95% of our subjects had pure NAFLD or NASH. Alcohol abuse might have been underreported by family members. But considering the low rate of alcohol abuse in a Muslim community, we feel confident that a very small minority of our subjects, if any, had true alcoholic steatohepatitis. Furthermore, the 4 major histological criteria used for determination of the severity of the necroinflammatory process in NAFLD have diagnostic value irrespective of the cause or background circumstances which led to fat deposition and subsequent

changes in the liver.

The results of our study will add to previous evidence emphasizing the limitation of liver biopsy in diagnosis and monitoring of NASH. This should urge hepatologists to seek a better alternative that should ideally be a noninvasive indicator of general histologic damage.

Unfortunately, all currently available alternatives fail to reliably identify the early disease stage which is more common. For this reason, in the absence of better alternatives to identify early disease and despite the numerous drawbacks of liver biopsy, histologic evaluation is still the best option.³² There might be ways to improve the usefulness of histologic evaluation. One might think of seeking histological features or scoring systems which are more consistent across the liver.

Acknowledgment

The study was funded by a grant from the Digestive Disease Research Center of Tehran University of Medical Sciences.

References

- Pourshams A, Malekzadeh R, Monavvari A, Akbari MR, Mohamdkhani A, Yarahmadi S, et al. Prevalence and etiology of persistently elevated alanine aminotransferase levels in healthy Iranian blood donors. *J Gastroenterol Hepatol*. 2005; **20**(2): 229 – 233.
- Sohrabpour AA, Rezvan H, Amini-Kafiabad S, Dayhim MR, Merat S, Pourshams A. Prevalence of nonalcoholic steatohepatitis in Iran: A population based study. *Mid East J Dig Dis*. 2010; **2**: 14 – 19.
- Ekstedt M, Franzen LE, Mathiesen UL, Thorelius L, Holmqvist M, Bodemar G, et al. Long-term follow-up of patients with NAFLD and elevated liver enzymes. *Hepatology*. 2006; **44**(4): 865 – 873.
- Ludwig J, Viggiano TR, McGill DB, Oh BJ. Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. *Mayo Clin Proc*. 1980; **55**(7): 434 – 438.
- Ong JP, Younossi ZM. Epidemiology and natural history of NAFLD and NASH. *Clin Liver Dis*. 2007; **11**(1): 1 – 16.
- Merat S, Aduli M, Kazemi R, Sotoudeh M, Sedighi N, Sohrabi M, et al. Liver histology changes in nonalcoholic steatohepatitis after one year of treatment with probucol. *Dig Dis Sci*. 2008; **53**(8): 2246 – 2250.
- Neuschwander-Tetri BA, Brunt EM, Wehmeier KR, Oliver D, Bacon BR. Improved nonalcoholic steatohepatitis after 48 weeks of treatment with the PPAR-gamma ligand rosiglitazone. *Hepatology*. 2003; **38**(4): 1008 – 1017.
- Saadeh S, Younossi ZM, Remer EM, Gramlich T, Ong JP, Hurley M, et al. The utility of radiological imaging in nonalcoholic fatty liver disease. *Gastroenterology*. 2002; **123**(3): 745 – 750.
- Suzuki A, Lymp J, Sauver JS, Angulo P, Lindor K. Values and limitations of serum aminotransferases in clinical trials of nonalcoholic steatohepatitis. *Liver Int*. 2006; **26**(10): 1209 – 1216.
- Mazhar SM, Shiehmozteza M, Sirlin CB. Noninvasive assessment of hepatic steatosis. *Clin Gastroenterol Hepatol*. 2009; **7**(2): 135 – 140.
- Merat S, Khadem-Sameni F, Nouraei M, Derakhshan MH, Tavangar SM, Mossaffa S, et al. A modification of the Brunt system for scoring liver histology of patients with non-alcoholic fatty liver disease. *Arch Iran Med*. 2010; **13**(1): 38 – 44.
- 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; **94**(9): 2467 – 2474.
- Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. 2005; **41**(6): 1313 – 1321.
- Bravo AA, Sheth SG, Chopra S. Liver biopsy. *N Engl J Med*. 2001; **344**(7): 495 – 500.
- Vuppalanchi R, Unalp A, Van Natta ML, Cummings OW, Sandrasegaran KE, Hameed T, et al. Effects of liver biopsy sample length and number of readings on sampling variability in nonalcoholic Fatty liver disease. *Clin Gastroenterol Hepatol*. 2009; **7**(4): 481 – 486.
- Merriman RB, Ferrell LD, Patti MG, Weston SR, Pabst MS, Aouizerat BE, et al. Correlation of paired liver biopsies in morbidly obese patients with suspected nonalcoholic fatty liver disease. *Hepatology*. 2006; **44**(4): 874 – 880.
- Maharaj B, Maharaj RJ, Leary WP, Cooppan RM, Naran AD, Pirie D, et al. Sampling variability and its influence on the diagnostic yield of percutaneous needle biopsy of the liver. *Lancet*. 1986; **327**(8480): 523 – 525.
- Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology*. 2005; **128**(7): 1898 – 1906.
- Younossi ZM, Gramlich T, Liu YC, Matteoni C, Petrelli M, Goldblum J, et al. Nonalcoholic fatty liver disease: assessment of variability in pathologic interpretations. *Mod Pathol*. 1998; **11**(6): 560 – 565.
- Ratziu V, Bugianesi E, Dixon J, Fassio E, Ekstedt M, Charlotte F, et al. Histological progression of non-alcoholic fatty liver disease: a critical reassessment based on liver sampling variability. *Aliment Pharmacol Ther*. 2007; **26**(6): 821 – 830.
- Siddique I, El-Naga HA, Mada JP, Memon A, Hasan F. Sampling variability on percutaneous liver biopsy in patients with chronic hepatitis C virus infection. *Scand J Gastroenterol*. 2003; **38**(4): 427 – 432.
- Arun J, Jhala N, Lazenby AJ, Clements R, Abrams GA. Influence of liver biopsy heterogeneity and diagnosis of nonalcoholic steatohepatitis in subjects undergoing gastric bypass. *Obes Surg*. 2007; **17**(2): 155 – 161.
- Larson SP, Bowers SP, Palekar NA, Ward JA, Pulcini JP, Harrison SA. Histopathologic variability between the right and left lobes of the liver in morbidly obese patients undergoing Roux-en-Y bypass. *Clin Gastroenterol Hepatol*. 2007; **5**(11): 1329 – 1332.
- Sanyal AJ, Chalasani N, Kowdley KV, McCullough A, Diehl AM, Bass NM, et al. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. *N Engl J Med*. 2010; **362**(18): 1675 – 1685.
- Ratziu V, Charlotte F, Bernhardt C, Giral P, Halbron M, Lenaour G, et al. Long-term efficacy of rosiglitazone in nonalcoholic steatohepatitis: results of the fatty liver improvement by rosiglitazone therapy (FLIRT 2) extension trial. *Hepatology*. 2010; **51**(2): 445 – 453.
- Adams LA, Sanderson S, Lindor KD, Angulo P. The histological course of nonalcoholic fatty liver disease: a longitudinal study of 103 patients with sequential liver biopsies. *J Hepatol*. 2005; **42**(1): 132 – 138.
- Harrison SA, Torgerson S, Hayashi PH. The natural history of non-alcoholic fatty liver disease: a clinical histopathological study. *Am J Gastroenterol*. 2003; **98**(9): 2042 – 2047.
- Sotoudehmanesh R, Sotoudeh M, Ali-Asgari A, Abedi-Ardakani B, Tavangar SM, Khakinejad A, et al. Silent liver diseases in autopsies from forensic medicine of Tehran. *Arch Iran Med*. 2006; **9**(4): 324 – 328.
- Manning DS, Afdhal NH. Diagnosis and quantitation of fibrosis. *Gastroenterology*. 2008; **134**(6): 1670 – 1681.
- Merat S, Rezvan H, Nouraei M, Jafari E, Abolghasemi H, Radmard AR, et al. Seroprevalence of hepatitis C virus: the first population-based study from Iran. *Int J Infect Dis*. 2010; **14**(suppl 3): 113 – 116.
- Merat S, Rezvan H, Nouraei M, Jamali A, Assari S, Abolghasemi H, et al. The prevalence of hepatitis B surface antigen and anti-hepatitis B core antibody in Iran: a population-based study. *Arch Iran Med*. 2009; **12**(3): 225 – 231.
- Bedossa P, Carrat F. Liver biopsy: the best, not the gold standard. *J Hepatol*. 2009; **50**(1): 1 – 3.