



HHS Public Access

Author manuscript

Clin Gastroenterol Hepatol. Author manuscript; available in PMC 2015 October 13.

Published in final edited form as:

Clin Gastroenterol Hepatol. 2009 March ; 7(3): 335–338. doi:10.1016/j.cgh.2008.11.029.

A Comparison of Hepatopathologists' and Community Pathologists' Review of Liver Biopsy Specimens From Patients With Hepatitis C

MAIRE ROBERT^{*}, ANDRE N. SOFAIR^{*}, ANN THOMAS[‡], BETH BELL[§], STEPHANIE BIALEK[§], CHRISTOPHER CORLESS^{||}, GRACE VAN NESS[‡], SHARON HUIE-WHITE^{*}, NICOLE STABACH^{*}, and ATIF ZAMAN^{||}

^{*}Yale University School of Medicine, New Haven, Connecticut

[‡]Oregon Public Health Division, Portland, Oregon

[§]Centers for Disease Control and Prevention, Atlanta, Georgia

^{||}Division of Gastroenterology and Hepatology, Oregon Health & Science University, Portland, Oregon

Abstract

Background & Aims—Assessment of liver histology has an important role in the management of chronic liver disease. It is not clear whether there are interobserver variabilities between hepatopathologists and general community pathologists. We evaluated the effect of type of pathologist and biopsy specimen size on interobserver agreement for hepatic fibrosis.

Methods—Subjects were identified from a population-based sample of adults from a chronic liver disease surveillance network. Biopsy slides from 391 hepatitis C patients who had undergone liver biopsy were obtained and read by 2 study hepatopathologists blinded to the patients' diagnoses (the gold standard). The interobserver agreement of the fibrosis stage between the hepatopathologists and the general pathologists' report were evaluated by kappa index.

Results—There was complete agreement between the study pathologist and community pathologist in 49.9% of biopsy specimens. The overall kappa index across all stages of fibrosis was 0.409, with the best agreement occurring at higher stages of fibrosis (kappa: 0.482 for stage 3, 0.776 for stage 4). Overall agreement was good (kappa, 0.465) when biopsy samples were greater than 1.5 cm in size. The community pathologist under-staged fibrosis in 73% of biopsy specimens with disagreement. A total of 26% of patients with stages 2 to 4 fibrosis were understaged by the community pathologist.

Conclusions—Results from this population-based study show good overall interobserver agreement between hepatopathologists and general pathologists when determining fibrosis stage in liver biopsy specimens from hepatitis C patients when liver biopsy sizes are adequate. However,

Address requests for reprints to: Atif Zaman, MD, MPH, Division of Gastroenterology and Hepatology, Oregon Health & Science University, Mail code L461, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201. zamana@ohsu.edu; fax: (503) 494-7556.

Conflicts of interest

The authors disclose no conflicts.

community pathologists tended to understage fibrosis, which could keep patients from receiving proper treatment.

Chronic hepatitis C is one of the most common causes of chronic liver disease worldwide.¹ Progressive hepatic fibrosis with the development of cirrhosis is a feature of chronic hepatitis C, as well as most other forms of chronic liver disease. The goal of current hepatitis C therapy is to achieve sustained viral clearance to halt or even improve ongoing histologic damage, which can lead to cirrhosis. However, current therapy is effective in only 50% to 60% of patients and is associated with significant side effects.²⁻⁴ Therefore, current guidelines suggest that treatment should be considered only in hepatitis C patients who have evidence of histologic damage, specifically significant hepatic fibrosis defined as stage 2 fibrosis or higher.¹

The current standard of practice is to consider liver biopsy in hepatitis C patients who appear to be good candidates for interferon-based treatment, and if the biopsy specimen shows significant fibrosis, to proceed with treatment. However, there are limitations to the use of liver biopsy for the histologic assessment of fibrosis. First, a liver biopsy removes only about 1/50,000 of the liver and thus carries a substantial risk of sampling error. Studies suggest that the minimum size of an adequate liver biopsy specimen is at least 1.5 cm.^{5,6} Second, assessment of fibrosis is at best a semiquantitative analysis. Studies have shown that there is significant interobserver variability in assessing histologic necroinflammation and fibrosis. For example, a study by Regev et al,⁷ using the Scheuer staging system, reported that there was a difference in the Scheuer stage of at least one grade in 33% of patients. Because of the critical role histologic assessment plays in the management of liver disease, numeric staging systems such as the Histological Activity Index of Knodell and the Metavir scoring system (specifically used for evaluating fibrosis and inflammatory activity in hepatitis C) have been developed to minimize the subjective aspect of histologic assessment.^{8,9}

Although these scoring systems have been shown in previous studies to improve precision and reproducibility in histologic assessment, most of these studies were small and performance of these scoring systems was based on readings by specialized pathologists working in academic settings, and therefore may not reflect their performance in a clinical setting.^{10,11} Furthermore, the impact of the type of pathologist, hepatopathologist, or a general pathologist (typically seen in community settings) on interobserver variability has been poorly studied. The aims of this study were as follows: (1) to evaluate the interobserver agreement across all stages of fibrosis between nonacademic (community-based) pathologists versus hepatopathologists practicing in an academic setting, focusing on interobserver agreement regarding significant fibrosis, and (2) to determine the impact of biopsy specimen size on interobserver agreement.

Methods

The study population was selected from participants in the Chronic Liver Disease Surveillance Network, a population-based sample of adults with newly diagnosed hepatitis C seen in gastroenterology practices in 3 US counties.¹² Subjects were identified during the

calendar years 1998 to 2001 by active surveillance at all community and academic gastroenterology practices in the surveillance areas. An eligible case was defined as a patient newly diagnosed with chronic liver disease who had visited a participating gastroenterologist during the study period; was at least 18 years old; was a resident of the following counties: New Haven, CT, Alameda, CA, or Multnomah, OR; did not have recognized human immunodeficiency virus infection; and, in the case of Alameda County residents, was enrolled in the Northern California Kaiser Permanente Medical Care Program. Because of the limited number of patients who underwent a biopsy, participants from the Kaiser Permanente Medical Care Program were excluded from the current study.

Newly diagnosed chronic liver disease was defined as follows: (1) abnormal liver tests of at least 6 months' duration; or (2) pathologic findings on liver biopsy including cirrhosis, fibrosis, or chronic hepatitis; or (3) abnormal findings on imaging studies, including nodularity or evidence of portal hypertension; or (4) occurrence of a diagnostic clinical event such as variceal bleeding, evidence of portal hypertension found on endoscopy, encephalopathy, or portal hypertensive ascites. The study protocols for each surveillance site were approved by the Institutional Review Boards of all collaborating institutions.

Trained interviewers administered a survey to eligible subjects after obtaining informed consent. Interview data gathered included demographic characteristics, medical history, history of risk factors for hepatitis, and a lifetime drinking history. Relevant clinical data, including the biopsy report, if available, were obtained from a review of the patients' medical records in the gastroenterologists' offices. Subjects also underwent serologic testing for markers of hepatitis C virus infection, including antibody to hepatitis C virus (HCV) (anti-HCV) (ORTHO-HCV enzyme-linked immunosorbent assay v. 3; Ortho-Clinical Diagnostics, Raritan NJ); positive specimens were tested for HCV RNA (COBAS AMPLICOR Hepatitis C Virus Test, v2.0; Roche Diagnostics, Basel, Switzerland) or, if HCV RNA negative, with a supplemental recombinant immunoblot assay (RIBA, Chiron RIBA HCV Strip Immunoblot Assay, version 3.0; Chiron Corp., Emeryville, CA). Specimens were not retested for markers when the result was available in the chart.

We obtained biopsy slides for patients who previously had undergone liver biopsy as part of their clinical evaluation. A total of 53 community pathologists originally were involved in reading the liver biopsy slides. These slides were reread by 1 of the 2 study hepatopathologists blinded to the patients' diagnoses, using a standardized pathology score sheet to stage fibrosis and grade inflammation.¹³ The study hepatopathologists practice in academic settings in the 2 surveillance areas. At the beginning of the study and before biopsy scoring, the study hepatopathologists evaluated the interobserver variability between themselves by examining the same set of 10 slides from an external sample of hepatitis C patients blindly and independently. Agreement was 100% for no fibrosis and cirrhosis. Disagreement was noted in the intermediate fibrosis stages by no more than one stage in 4 cases. Differences in scoring and scoring criteria were discussed in detail to achieve a uniform approach to scoring based on published criteria and to minimize interobserver disagreement during the study. On the pathology form, fibrosis was defined as follows: 0, none; 1, portal fibrosis; 2, periportal fibrosis; 3, bridging fibrosis; or 4, cirrhosis. Consistent with the hepatology literature, for purposes of analysis, any fibrosis score of 2 or higher was

considered to be significant fibrosis. Although portal inflammation also was scored, we did not study interobserver agreement for inflammation because hepatic fibrosis has a greater impact on treatment decisions in hepatitis C patients.

We limited our analysis to hepatitis C patients whose liver biopsy slides were reread by one of the study hepatopathologists, the full pathology report from the general pathologist practicing in the community who originally read the biopsy was available, and the study hepatopathologist was not also the community pathologist (ie, read the initial liver biopsy as well). Therefore, a total of 391 (78%) of 499 hepatitis C patients from the 2 surveillance sites who had undergone a liver biopsy comprised the study cohort. The variables studied included stage of fibrosis reported by the study hepatopathologist and the general pathologist, and whether the biopsy specimen was less than 1.5 cm in length.

Appropriate descriptive statistics were computed for all variables. By using the study pathologist's reading as the reference, the interobserver agreement of the fibrosis scores was evaluated using the kappa index, which accounts for expected agreement by chance. The interpretation of kappa values was as follows: > 0.75 , excellent; 0.40 kappa 0.75 , fair to good; $0 < \text{kappa} < 0.40$, poor.¹⁴ Interobserver agreement was assessed for all fibrosis stages, as well as for significant (stage ≥ 2) versus nonsignificant (stage < 2) fibrosis. Finally, each of these analyses was stratified by length of biopsy specimen (< 1.5 cm or ≥ 1.5 cm) to evaluate its effect on the kappa index. SPSS software version 15.0 (SPSS Inc, Chicago, IL, 2004) was used for database management and statistical analysis.

Results

Of the study cohort of 391 cases, 245 (63%) were from Connecticut and the remainder were from Oregon. Adequate liver biopsy specimen, defined as a length of 1.5 cm or greater, was observed in 67% of biopsy specimens. The percentage of adequate biopsy samples was 74% at the Oregon site and 63% at the Connecticut site. Considering the hepatopathologists' readings as the reference, the distribution of the fibrosis stage was 28 patients (7.2%) with stage 0 fibrosis, 90 patients (23%) with stage 1 fibrosis, 139 patients (35.5%) with stage 2 fibrosis, 63 patients (16.1%) with stage 3 fibrosis, and 71 patients (18.2%) with stage 4 fibrosis. In 49.9% of biopsy specimens, the study hepatopathologist's and general community pathologist's readings completely agreed. There was one fibrosis stage of disagreement in 37.3% of cases, 2 fibrosis stages in 12%, and 3 or 4 fibrosis stages in 0.8% of cases.

Table 1 describes the degree of interobserver agreement based on the kappa index. Without taking liver biopsy size into account, agreement across all stages of fibrosis was fair. Interobserver agreement was poor for stages 0 through 2, agreement was good for stage 3, and agreement was excellent for stage 4. Interobserver agreement was good when the samples were stratified as stages 0 to 1 versus stages 2 to 4. When taking liver biopsy size into account, overall as well as for each fibrosis stage, the kappa index was higher for biopsy sizes of 1.5 cm or greater.

Table 2 describes the percentage of disagreement between the hepatopathologist and community pathologist across each stage of fibrosis. The least disagreement was seen at the extreme stages of fibrosis, stage 0 and stage 4. In general, a biopsy size of 1.5 cm or greater reduced the degree of disagreement. Readings disagreed in a higher proportion of biopsy specimens showing stage 1 or stage 2 fibrosis compared with biopsy specimens read as other stages. There was disagreement by one or more stages in half of the biopsy specimens with a reference reading of stage 1 or stage 2. General pathologists tended to understage fibrosis compared with the study hepatopathologists. The general pathologist understaged fibrosis in 73% of biopsy specimens with disagreement. There was complete agreement in only 41.7% of biopsy specimens showing stage 2 fibrosis; 50.3% of these biopsy specimens were understaged by the general pathologists. Similar findings were noted at stages 0 to 1 versus stages 2 to 4, which is typically the cut-off level where the decision for hepatitis C treatment is made. At this cut-off level, 273 patients were considered to have stages 2 to 4 fibrosis by the hepatopathologist. A discrepancy was seen in 91 of 391 patients between the study hepatopathologist and community pathologist. If the study hepatopathologist's reading was considered the gold standard, then 71 of the 91 discrepant patients were understaged by the community pathologist and would not have been offered hepatitis C treatment. Therefore, of the 273 stage 2 to 4 patients, 26% would not have been offered hepatitis C treatment if the community pathologist's reading was used. This percentage decreased to 20% if only patients with a biopsy size of 1.5 cm or greater was used.

Discussion

This large study compared interobserver variability between hepatopathologists practicing in academic institutions and general pathologists practicing in community settings in staging of hepatic fibrosis in liver biopsy specimens of hepatitis C patients. By using statistical criteria, agreement both overall and in distinguishing clinically significant fibrosis generally was fair to good, and was best among biopsy specimens with the highest stages of fibrosis. However, there was poor agreement between pathologists for biopsy specimens showing only portal or periportal fibrosis. More importantly, this interobserver variability would lead to about a quarter of the patients not being offered hepatitis C treatment because the community pathologists' fibrosis reading would have understaged the fibrosis below the threshold of stage 2 fibrosis, at which point treatment typically is offered.

Despite the fact that liver biopsy remains the primary tool for the assessment of liver fibrosis, it is far from an ideal test. Because a standard biopsy needle only samples a small portion of the liver, it is prone to sampling error. Typically, sampling error understages the degree of fibrosis; several studies have shown that cirrhosis is missed on a single blind liver biopsy in 10% to 30% of cases.¹⁵⁻¹⁷ In a study by Abdi et al,¹⁸ diagnosing cirrhosis correctly increased from 80% to 100% when comparing a single biopsy specimen with 3 specimens. Previous studies have suggested that accuracy also improves when the specimen size is at least 1.5 cm in length and/or greater than 5 portal tracts are present.^{5,6} In the present study, when the biopsy specimen size was 1.5 cm or greater, interobserver agreement improved both overall and at each fibrosis stage.

Over the years, standardized staging systems have been developed, such as the Knodell and METAVIR classifications, to improve the reproducibility of histologic findings. Indeed, many studies have shown good to excellent interobserver agreement in staging fibrosis using these staging systems.^{9,10,19} However, even among academic hepatopathologists, interobserver disagreement has been reported to be 10%.⁷ Furthermore, a recent study by Rousselet et al²⁰ found that interobserver agreement was poor between academic and nonacademic pathologists, with an overall kappa value of 0.22. In contrast, we found fair to good agreement, especially when the biopsy size was 1.5 cm or greater. It is unclear why the agreement was higher in the present study. A possible reason is that at the time of the current study hepatitis C was well known in the medical community and nonacademic pathologists may have received formal training using standardized histologic scoring systems for hepatitis C. Also, other studies, such as the Rousselet et al²⁰ study, were non-US studies and there may be geographic factors playing a role in the differences.

The decision to treat patients with hepatitis C is based on many factors including underlying comorbid conditions, the ability of the patient to tolerate the regimen, and patient compliance. However, the major determining factor is the degree of histologic damage seen on liver biopsy. Typically, patients with stage 2 fibrosis or higher are considered appropriate candidates for treatment from a histologic standpoint. In this study, hepatopathologists and general pathologists disagreed most often when reading biopsy specimens with fibrosis stages 1 and 2, and the majority of the time the general pathologist understaged the degree of fibrosis compared with the hepatopathologist. These findings may have clinical implications. One can speculate that some hepatitis C patients being evaluated in the community found to have stage 1 fibrosis would not be offered therapy because of evidence of only mild histologic damage, whereas these same patients evaluated by a hepatopathologist could have been candidates for treatment based on an indication of stage 2 fibrosis on the same biopsy specimen. This may be even more likely if the biopsy specimen is less than 1.5 cm in size.

This study had several strengths. The study included a large group of hepatitis C patients, ascertained in a systematic manner from the population in 2 different geographic locations. The reference histologic staging of fibrosis was performed by 2 hepatopathologists who used an a priori agreed-on histologic classification. Also, the histologic findings and biopsy specimen sizes reflect what one actually would expect in the community because the study cohort was community-based. However, the study had some limitations. Because both academic hepatopathologists did not read each biopsy specimen, the interobserver agreement for every biopsy specimen could not be determined. Further, we considered the study pathologists' readings to be the reference, that is, the correct reading, which might not have been accurate in all cases. Interobserver differences have occurred even among academic pathologists and even when a pathologist rereads their own slides at a later date. Therefore, it is impossible to have a true gold standard. However, in the current study, the study hepatopathologists reviewed a sample of liver biopsy specimens and discussed any disagreements to minimize interobserver disagreements during the study. Finally, the histologic readings by the hepatopathologists and the general pathologists were not performed close in time to each other; in many cases the readings were years apart.

In conclusion, the data from this population-based study show that there is fair to good overall interobserver agreement between hepatopathologists and general pathologists in determining the degree of fibrosis in liver biopsy specimens from hepatitis C patients. This agreement improves when liver biopsy specimens are greater than 1.5 cm in size. Furthermore, the agreement is the best when advanced fibrosis is present. However, there was poor agreement in the lower to intermediate stages of fibrosis, especially at stages 1 and 2, with general pathologists tending to understage fibrosis. Hence, the study's findings have potential implications for clinical practice in the community when hepatic fibrosis stage 2 is used as a criterion to identify candidates for interferon-based therapy. Clinicians should keep the following issues in mind when basing treatment decisions on liver biopsy findings: (1) adequacy of liver biopsy specimen size, (2) their pathologists should use standardized scoring methods, and (3) if stage 1 fibrosis is reported, then consideration should be given to a second reading by another pathologist (ideally by a hepatopathologist).

Abbreviation used in this paper

HCV hepatitis C virus

References

1. Seeff LB, Hoofnagle JH. National Institutes of Health consensus development conference statement: management of hepatitis C: 2002—June 10–12, 2002. *Hepatology*. 2002; 36:S1–S20. [PubMed: 12407571]
2. Manns MP, McHutchison J, Gordon SC, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C; a randomised controlled trial. *Lancet*. 2001; 358:958–965. [PubMed: 11583749]
3. Fried MW, Shiffman ML, Reddy KR, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med*. 2002; 347:975–982. [PubMed: 12324553]
4. Hadziyannis S, Sette H, Morgan TR, et al. Peginterferon alfa-2a and ribavirin combination therapy in chronic hepatitis C. *Ann Intern Med*. 2004; 140:346–355. [PubMed: 14996676]
5. Holund B, Poulsen H, Schlichting P. Reproducibility of liver biopsy diagnosis in relation to the size of the specimen. *Scand J Gastroenterol*. 1980; 15:329–335. [PubMed: 7433892]
6. Schlichting P, Holund B, Poulsen H. Liver biopsy in chronic aggressive hepatitis. Diagnostic reproducibility in relation to size of specimen. *Scand J Gastroenterol*. 1983; 18:27–32. [PubMed: 6675176]
7. Regev A, Berho M, Jeffers LJ, et al. Sampling error and interobserver variation in liver biopsy in patients with chronic hepatitis C. *Am J Gastroenterol*. 2002; 97:2614–2618. [PubMed: 12385448]
8. Knodell RG, Ishak KG, Black WC, et al. Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. *Hepatology*. 1981; 1:431–435. [PubMed: 7308988]
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. [PubMed: 8020885]
10. Goldin RD, Goldin JG, Burt AD, et al. Intraobserver and interobserver variation in the histopathological assessment of chronic viral hepatitis. *J Hepatol*. 1996; 25:649–654. [PubMed: 8938541]
11. Gronbaek K, Christensen PB, Hamilton-Dutoit S, et al. Interobserver variation in interpretation of serial liver biopsies from patients with chronic hepatitis C. *J Viral Hepat*. 2002; 9:443–449. [PubMed: 12431207]

12. Bell BP, Navarro VJ, Manos MM, et al. The epidemiology of newly-diagnosed chronic liver disease in the United States: findings of population-based sentinel surveillance (abstr). *Hepatology*. 2001; 34:468A.
13. Ludwig J. The nomenclature of chronic active hepatitis: an obituary. *Gastroenterology*. 1993; 105:274–278. [PubMed: 8514045]
14. Fleiss, JL. *Statistical methods for rates and proportions*. New York: Wiley; 1981. The measurement of interrater agreement; p. 212-236.
15. Bruguera M, Bordas JM, Mas P, et al. A comparison of the accuracy of peritoneoscopy and liver biopsy in the diagnosis of cirrhosis. *Gut*. 1974; 15:799–800. [PubMed: 4279817]
16. Pagliaro L, Rinaldi F, Craxi A, et al. Percutaneous blind biopsy versus laparoscopy with guided biopsy in diagnosis of cirrhosis. A prospective, randomized trial. *Dig Dis Sci*. 1983; 28:39–43. [PubMed: 6217961]
17. Maharaj B, Maharaj RJ, Leary WP, et al. Sampling variability and its influence on the diagnostic yield of percutaneous needle biopsy of the liver. *Lancet*. 1986; 1:523–525. [PubMed: 2869260]
18. Abdi W, Millan JC, Mezey E. Sampling variability on percutaneous liver biopsy. *Arch Intern Med*. 1979; 139:667–669. [PubMed: 443970]
19. Westin J, Lagging LM, Wejstal R, et al. Interobserver study of liver histopathology using the Ishak score in patients with chronic hepatitis C virus infection. *Liver*. 1999; 19:183–187. [PubMed: 10395036]
20. Rousselet M, Michalak S, Dupré F, et al. Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology*. 2005; 41:257–264. [PubMed: 15660389]

Table 1

Interobserver Agreement Between Study Pathologists and Community Pathologists in Assessing Stage of Fibrosis in Patients With Hepatitis C

Fibrosis stage (study pathologist)	Kappa index	Kappa index stratified by liver biopsy specimen size	
		<1.5 cm	1.5 cm
Overall (across all stages)	0.409	0.379	0.465
0	0.274	0.142	0.345
1	0.304	0.282	0.330
2	0.282	0.223	0.401
3	0.482	0.454	0.549
4	0.776	0.703	0.857
0–1 vs 2–4	0.508	0.447	0.629

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Percentage Agreement Between Study Pathologists and Community Pathologists in Assessing Stage of Fibrosis in Patients With Hepatitis C

Fibrosis stage (study pathologist)	Overall percentage disagreement	Percentage disagreement stratified by liver biopsy specimen size	
		<1.5 cm	1.5 cm
0	35.7	60	22.2
1	52.2	54.5	49.1
2	58.3	59.2	55.6
3	44.4	47.8	35.3
4	22.5	29.7	14.7
0-1 vs 2-4	23.3	26.1	17.6

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript