

ORIGINAL ARTICLE

Variability of grade and stage in simultaneous paired liver biopsies in patients with hepatitis C

Silvia Skripenova, Thomas D Trainer, Edward L Krawitt, Hagen Blaszyk

J Clin Pathol 2007;**60**:321–324. doi: 10.1136/jcp.2005.036020

See end of article for authors' affiliations

Correspondence to: H Blaszyk, University of Vermont College of Medicine, Department of Pathology, 89 Beaumont Avenue, Burlington, VT 05405 USA; hagen.blaszyk@vtmednet.org

Accepted 1 March 2006
Published Online First
12 May 2006

Background: Grading and staging of liver biopsies in patients with chronic hepatitis remains an inexact "gold standard" that is influenced by variabilities in scoring systems, sampling, observer agreement and expertise. Spatial disease variability relative to markers of the adequacy of biopsy has not been studied previously.

Methods: Paired liver biopsy specimens were obtained from the right and left hepatic lobes of 60 patients with chronic hepatitis C. Histological grade and disease stage were assessed according to the Ludwig scoring system, and scores were evaluated in relation to differences in size and number of portal tracts in all paired samples.

Results: The relative difference (%) in aggregate biopsy size and number of portal tracts was similar between paired samples with and without a difference in grade. Paired samples with a difference in stage showed a larger relative difference in biopsy size ($p=0.09$) and in the number of portal tracts ($p=0.016$).

Conclusions: Our study shows a difference of one grade or one stage in 30% of paired liver biopsies, due to a combination of sampling variability and observer variability. Acknowledgment of "built-in" variability in grading and staging chronic hepatitis C by both clinicians and pathologists is essential for managing the individual patient with chronic hepatitis C.

Liver biopsy remains the gold standard for assessing disease severity and progression in chronic hepatitis C.^{1,2} Although virus-related factors seem to be better predictors of response to treatment than host-related factors, necroinflammatory activity (grade) and degree of fibrosis (stage) provide both a measure and a predictor of disease progression.¹ However, liver biopsy has several limitations, including interobserver and intraobserver variability in biopsy interpretation, sampling variation, and the morbidity and mortality associated with the procedure itself.³

Sampling variability is of considerable concern, as a standard liver biopsy specimen represents only about 0.0002% of the entire organ. This problem has been assessed in various ways through investigative studies.^{3–7} It is generally thought that size does matter,⁸ and that bigger is better,⁹ but universally accepted consensus criteria for the adequacy of biopsy regarding the minimum required biopsy size or number of sampled portal tracts do not exist.^{9,10} Moreover, several different grading/staging systems are used worldwide for the histological assessment of chronic hepatitis C.^{11–13} The different sources of variability in interpretation of liver biopsy influence each other at least in part, but the degree of observer experience is probably more important than characteristics of the specimen.³

Two studies have recently dealt with the problem of spatial disease variability through sampling more than one area of the liver in patients with chronic hepatitis C.^{9,14} Both studies found differing histological grades and stages in a large fraction (24–45%) of paired biopsy specimens, but the relationship between observed disease variability and biopsy size has not been studied.

PATIENTS AND METHODS

In total, 60 patients with chronic hepatitis C and no clinical or histological evidence for coexisting liver diseases were enrolled into the study (2001–3), which had been approved by the Institutional Review Board at the University of Vermont (Burlington, Vermont, USA). Paired liver biopsy specimens were obtained from the right and left hepatic lobes under ultrasound guidance using 18-gauge core biopsy needles. Specimens were fixed in formalin and routinely processed for histology including haematoxylin and eosin (H&E), iron and

trichrome stains. Each biopsy sample was scored after examination of at least three histological H&E-stained sections and corresponding trichrome stains. Biopsy samples from patients with coexisting steatohepatitis were excluded.

Histological grade (degree of necroinflammatory activity) and the disease stage (extent of fibrosis) were determined for each sample according to the four-tiered Ludwig scoring system,¹⁵ which is similar to the grading tool originally published by Scheuer.¹⁶ Biopsy samples in which the liver capsule was identified were graded and staged using all but a 1-mm subcapsular rim of tissue. The maximum aggregate length of each needle core biopsy specimen was measured and the number of portal tracts was recorded. All histological slides were coded to ensure blinded evaluation of the paired samples. The entire study set was then randomly evaluated by a hepatopathologist (HB) to obtain scores for both grade and stage, which were used for subsequent statistical analysis. Two weeks later, HB again blindly and randomly evaluated all paired biopsy specimens to assess intraobserver variability. A junior trainee (SS) with interest in liver pathology and experience with the Ludwig scoring system graded and staged the entire study set randomly to assess interobserver variability.

The relative difference (in %) in aggregate biopsy size and number of portal tracts between each pair of biopsies was calculated. For example, if the right lobe biopsy of a paired sample measured 7 mm, compared with a left lobe biopsy size of 10 mm, then the relative size difference was calculated as 30%.

Statistical analysis

Statistical analyses were performed using a two-tailed t test, assuming unequal variance, and the χ^2 test, assuming a significance level of $p<0.05$. Cohen κ statistics assessed intraobserver and interobserver agreement.

RESULTS

Paired bilateral needle biopsy samples from 60 patients (median age 44.5 years; range 20–55 years; 32 men, 28 women) with chronic hepatitis C and no other coexisting liver disease were evaluated. Biopsy characteristics were consistent with chronic

Table 1 Distribution of grade and stage in 60 paired right and left lobe liver biopsies

	Right	Left
Grade		
1	3 (5%)	8 (13%)
2	36 (60%)	27 (45%)
3	21 (35%)	25 (42%)
4	0	0
Stage		
1	3 (5%)	3 (5%)
2	36 (60%)	40 (67%)
3	15 (25%)	11 (18%)
4	6 (10%)	6 (10%)

hepatitis C in all cases. The percentage of hepatitis C-associated steatosis varied from 0% to 25%, with a patchy distribution pattern, but features of steatohepatitis were not present in this study set. The degree of steatosis did not vary significantly between paired samples. The median aggregate biopsy size of both right and left-sided biopsies was 14 mm, with a range of 4–30 mm for the right and 5–31 mm for the left. Similarly, the median number of portal tracts of all right-sided samples was 13.5 (range 4–28), compared with a median of 12 (range 5–27) in left-sided biopsies. The relative right versus left difference in number of portal tracts (mean (standard deviation) (SD)) for all paired samples was 27.4 (19.1), and that for aggregate biopsy size 21.4 (17). Table 1 shows the distribution of grade (degree of necroinflammatory activity) and stage (extent of fibrosis) according to the Ludwig scoring system for all samples.¹⁵

A difference of either one grade or one stage was seen in 18 (30%) paired samples. Of those, 10 showed a difference in both one grade and one stage. Paired biopsy samples with a difference in grade or stage were proportionally distributed (data not shown). The relative difference (in %) in aggregate biopsy size and number of portal tracts was similar between paired samples with and without a difference in grade (table 2).

The relative difference in aggregate biopsy size was larger in paired samples that showed a difference in stage, but this difference failed to reach statistical significance ($p = 0.09$). The relative difference in the number of portal tracts was significantly ($p = 0.016$) greater in paired samples that showed a difference in stage (table 2). The smaller specimen of those paired samples consistently scored lower.

The absolute size (in mm) and the absolute number of portal tracts did not differ significantly between right-sided or left-sided biopsy specimens, whether or not there was a difference in grade or stage (data not shown). The fraction of paired samples showing a difference in grade or stage did not change significantly when restricting analyses to paired samples of a certain minimum size or number of portal tracts—that is, size at least 10, 15 or 20 mm, or number of portal tracts at least 10, 15 or 20 (table 3).

Interobserver variability analysis (trainee *v* expert) indicated a moderate degree of agreement and showed differing grading

scores in 41 of 120 (34.2%) individual samples (κ coefficient 0.47). Similarly, differing staging scores were seen in 18 of 120 (15%) individual samples (κ coefficient 0.59). Samples with disagreement showed a consistent trend toward higher grading and staging by the trainee. Intraobserver variability analysis (expert *v* expert on different days) indicated a substantial degree of agreement and showed differing grading scores in 30 of 120 (25%) individual samples (κ coefficient 0.72). Similarly, differing staging scores were seen in 14 of 120 (11.7%) individual samples (κ coefficient 0.78). Individual biopsy samples that showed interobserver or intraobserver variability in scores of grade or stage did not differ in size or number of portal tracts from those samples that scored consistently (data not shown).

DISCUSSION

Liver biopsy continues to have a central role as the “gold standard” in evaluating patients with chronic hepatitis C,^{1 9 11 13} but limitations include cost, patient morbidity and mortality, observer variability, sampling variation and use of different histopathological scoring tools.

The Knodell histology activity index was the first system of its type to assess the severity of chronic hepatitis objectively, semiquantitatively and reproducibly. It was followed by the development of other grading/staging systems of variable complexity, such as the Ishak, Scheuer, Ludwig and METAVIR scoring systems.^{12 13} The resulting table with 552 possible numerals underscores the need to choose an appropriate system for the right setting. To be effective in everyday diagnostic practice, scoring systems must be simple to understand and to apply, be effectively communicated to the clinician, and be clinically relevant.¹³ Simpler systems have generally higher degrees of observer concordance,¹³ and agreement increases with the degree of observer experience.³ The Ludwig and Scheuer systems^{15 16} are less complex than the Knodell/Ishak or METAVIR systems, are associated with significant interobserver and intraobserver agreement,^{3 5} and have been adopted by many pathologists and hepatologists for the routine clinical evaluation of patients with chronic hepatitis C.⁵

Studies validating the most widely used systems have shown satisfactory degrees of intraobserver and interobserver variability.¹³ Our study was not designed to investigate observer variability of the Ludwig scoring system, but the data are consistent with previous studies. Agreement was generally greater for the assessment of fibrosis (stage), and the degree of expertise significantly influenced variability, using κ scores for assessment of intraobserver and interobserver agreement. A substantial degree of intraobserver agreement (κ scores of 72 for grade and 78 for stage in the current study) cannot minimise differing grading scores in 25% and differing staging scores in 12% of the 120 individual study samples. The generated numbers in hepatitis grading and staging do not represent measurements of a continuous variable, which presents a fundamental problem with histological scoring.¹¹ However, the relative simplicity of the Ludwig system used in this study would be expected to carry a low risk for substantial

Table 2 Relative differences (in %) in size and number of portal tracts in 60 paired right and left lobe liver biopsies

	n	Relative difference (%) in size	Relative difference (%) in number of portal tracts
Biopsy pairs with difference in grade	18	20.5 (16.5)	30.5 (21.2)
Biopsy pairs without difference in grade	42	21.8 (17.4)	26.2 (18.3)
Biopsy pairs with difference in stage	18	28.4 (21.9)*	37.7 (21.7)†
Biopsy pairs without difference in stage	42	18.5 (13.7)	22.9 (16.2)

* $p = 0.09$.

† $p = 0.016$.

Table 3 Number of paired liver biopsy specimens showing a difference in grade or stage relative to biopsy size and number of portal tracts in 60 biopsies of the right and left liver

Both specimens of the paired biopsy sample feature a minimum of:	n	Paired samples with a difference in grade	Paired samples with a difference in stage
10 mm in length	50 (83.3%)	16 (32.0%)	15 (30.0%)
15 mm in length	18 (30.0%)	6 (33.3%)	6 (33.3%)
20 mm in length	5 (8.3%)	2 (40.0%)	1 (20.0%)
10 portal tracts	50 (83.3%)	14 (28.0%)	12 (24.0%)
15 portal tracts	23 (38.3%)	7 (30.4%)	8 (34.8%)
20 portal tracts	6 (10.0%)	2 (33.3%)	1 (16.7%)

observer variation.^{3,5} We performed and compared the statistical analyses using both intraobserver and interobserver datasets without significant change in numbers, proportions and significance levels (data not shown). Thus, the observed intraobserver variation in grading and staging seems to be random, and did not change the interpretation of our data. It also indicates that grading and staging differences are more likely due to spatial disease variation.

Presently, there are no uniform criteria for assessing the adequacy of liver biopsies. On the basis of older reports, the recommended satisfactory length of liver biopsy in hepatitis C ranges from 10 to 40 mm, and a sample 15 mm long or containing 4–6 portal tracts was considered acceptable.^{6,9} The variability of the relative amount of fibrosis is negligible beyond a specimen length of 40 mm.⁴ When the criteria for minimum size (at least 10, 15 or 20 mm) or minimum number of portal tracts (at least 10, 15 or 20) were applied to our study set, the fraction of paired samples with a difference in grade or stage did not change. Earlier studies suggested that thin-needle biopsy may provide information similar to large-needle biopsy,⁷ but this notion has been generally refuted,⁸ and bigger seems to be better.⁹ Refined biopsy size criteria have been suggested to include a biopsy specimen at least 20 mm long and 1.4 mm wide, or the presence of 11 portal tracts.^{6,9} Schiano *et al*¹⁷ showed reliable grading and staging in specimens measuring at least 10 mm.

This study assessed the relationship between observed sampling variability and biopsy size. We showed that a size difference between paired samples did not influence the assigned inflammatory grade. By contrast, paired samples with a large relative size difference were more likely to show a difference in fibrosis stage. The smaller sample in each pair consistently scored lower, a finding supported by others.^{6,8} Two other groups found differing histological stages in a large fraction (24–45%) of paired biopsy specimens with hepatitis C obtained from different areas of the right lobe of the liver,¹⁴ or biopsies obtained from the right and left lobes of the liver chosen under direct visualisation during peritoneoscopy.⁵ This supports older studies showing confirmation of cirrhosis in only 50–80% when one biopsy specimen was examined, compared with the assessment of three spatially distinct samples.^{18,19} Interestingly, hepatic viral load does not seem to have the same degree of heterogeneity of sampling variability as does histology.²⁰ However, resolving the problem of disease heterogeneity by taking several biopsy specimens from the same patient must be balanced against the risk of increased morbidity and mortality.^{4,20} Moreover, the relationship between observer agreement and specimen size is debated,³ suggesting that as the specimen increases in size the probability of various types of lesions increases, creating a paradoxical increase in observer variability. Sampling variability can never be completely

Take-home messages

- Histological grading and staging in chronic hepatitis is an inexact gold standard.
- A “built-in” variability in histological grading/staging of chronic hepatitis has to be acknowledged by both the clinician and the pathologist.
- Variability in scoring systems, sampling, observer agreement and expertise necessitate a common ground approach in scoring biopsies.
- This common-ground approach is best facilitated by (a) choosing the simplest scoring system to serve the intended purpose, (b) pursuing minimum sample criteria, including biopsy size of >10 mm or presence of >10 portal tracts, and (c) histological biopsy evaluation by an experienced pathologist.

eliminated, but may be overcome in large cohorts of patients and biopsies, as variation is likely to be random and multi-directional.⁹ This should minimise the role of sampling variability in the interpretation of clinical trials. However, sampling variability may become an issue in the management of the individual patient, particularly in the assessment of disease activity or progression, which may then influence treatment decisions. The relatively small number of biopsy specimens with minimum and maximum grade and stage in this study set is consistent with a random sample of patients with chronic hepatitis C seen at our institution. Patients at the extremes of inflammatory grade or histological stage may be particularly important in clinical decision making. However, paired biopsy samples with a difference in grade or stage ($n = 18$) were proportionally distributed along the different grades and stages, and did not cluster at potentially important clinical decision points. All paired biopsy samples with a difference in grade or stage ($n = 18$) showed only a one-grade or one-stage difference. Understaging, and to a lesser degree undergrading, would have occurred if only the smaller of the paired sample were to be examined. This could have influenced clinical decision making in some patients. Other clinical data such as viral load may help to balance the “built-in” variability in grading and staging chronic hepatitis C, but the assessment of the adequacy of biopsy should accompany each pathology report.

In summary, grading and staging of liver biopsies in chronic hepatitis remains an inexact gold standard that is influenced by variabilities in scoring systems, sampling, observer agreement and expertise. Optimising one factor may confound the performance of other variables, necessitating a “common ground” approach. This would include (1) choice of the simplest scoring system that will serve the intended purpose; (2) pursuit of minimum sample criteria, including a size of at least 10 mm or presence of >10 portal tracts; and (3) assessment of biopsies by an experienced pathologist. Acknowledgment of a “built-in” variability in grading and staging chronic hepatitis C by both the clinician and the pathologist is essential for managing the individual patient with chronic hepatitis C.

Authors' affiliations

S Skripnova, T D Trainer, H Blaszyk, Department of Pathology, University of Vermont College of Medicine, Burlington, Vermont, USA

E L Krawitt, Department of Medicine, University of Vermont College of Medicine, Burlington, Vermont, USA

Competing interests: None.

Ethical approval: This study was approved by the Institutional Review Board at the University of Vermont.

REFERENCES

- 1 **Koukoulis GK.** Chronic hepatitis C: grading, staging, and searching for reliable predictors of outcome. *Hum Pathol* 2001;**32**:899–903.
- 2 **Saadeh S, Cammel G, Carey WD, et al.** The role of liver biopsy in chronic hepatitis C. *Hepatology* 2001;**33**:196–200.
- 3 **Rousselet MC, Michalak S, Dupre F, et al.** Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology*, 2005;**41**:257–64.
- 4 **Bedossa P, Dargere D, Paradis V.** Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003;**38**:1449–57.
- 5 **Regev A, Berho M, Jeffers LJ, et al.** Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol* 2002;**97**:2614–18.
- 6 **Colloredo G, Guido M, Sonzogni A, et al.** Impact of liver biopsy size on histological evaluation of chronic viral hepatitis: the smaller the sample, the milder the disease. *J Hepatol* 2003;**39**:239–44.
- 7 **Petz D, Klauk S, Roehl FW, et al.** Feasibility of histological grading and staging of chronic viral hepatitis using specimens obtained by thin-needle biopsy. *Virchows Arch* 2003;**442**:238–44.
- 8 **Demetris AJ, Ruppert K.** Pathologist's perspective on liver needle biopsy size? *J Hepatol* 2003;**39**:275–7.
- 9 **Scheuer PJ.** Liver biopsy size matters in chronic hepatitis: bigger is better. *Hepatology* 2003;**38**:1356–8.
- 10 **Guido M, Ruge M.** Liver biopsy sampling in chronic viral hepatitis. *Sem Liver Dis* 2004;**24**:89–97.
- 11 **Huebscher SG.** Histological grading and staging in chronic hepatitis: clinical applications and problems. *J Hepatol* 1998;**29**:1015–22.
- 12 **Scheuer PJ, Standish RA, Dhillon AP.** Scoring of chronic hepatitis. *Clinics Liver Dis* 2002;**6**:335–47.
- 13 **Brunt EM.** Grading and staging the histopathological lesions of chronic hepatitis: the Knodell histology activity index and beyond. *Hepatology* 2000;**31**:241–6.
- 14 **Siddique I, El-Naga HA, Madda JP, et al.** Sampling variability on percutaneous liver biopsy in patients with chronic hepatitis C virus infection. *Scand J Gastroenterol* 2003;**38**:427–32.
- 15 **Batts KP, Ludwig J.** Chronic hepatitis. An update on terminology and reporting. *Am J Surg Pathol* 1995;**19**:1409–17.
- 16 **Scheuer PJ.** Classification of chronic hepatitis: a need for reassessment. *J Hepatol* 1991;**13**:372–4.
- 17 **Schiano TD, Azeem S, Bodian CA, et al.** Importance of specimen size in accurate needle liver biopsy evaluation of patients with chronic hepatitis C. *Clin Gastroenterol Hepatol* 2005;**3**:930–5.
- 18 **Abdi W, Millan JC, Mezey E.** Sampling variability on percutaneous liver biopsy. *Arch Intern Med* 1979;**139**:667–9.
- 19 **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**:523–5.
- 20 **Fanning L, Loane J, Kenny-Walsh E, et al.** Tissue viral load variability in chronic hepatitis C. *Am J Gastroenterol* 2001;**96**:3384–9.

Clinical Evidence—Call for contributors

Clinical Evidence is a regularly updated evidence-based journal available worldwide both as a paper version and on the internet. *Clinical Evidence* needs to recruit a number of new contributors. Contributors are healthcare professionals or epidemiologists with experience in evidence-based medicine and the ability to write in a concise and structured way.

Areas for which we are currently seeking contributors:

- Pregnancy and childbirth
- Endocrine disorders
- Palliative care
- Tropical diseases

We are also looking for contributors for existing topics. For full details on what these topics are please visit www.clinicalevidence.com/ceweb/contribute/index.jsp

However, we are always looking for others, so do not let this list discourage you.

Being a contributor involves:

- Selecting from a validated, screened search (performed by in-house Information Specialists) epidemiologically sound studies for inclusion.
- Documenting your decisions about which studies to include on an inclusion and exclusion form, which we keep on file.
- Writing the text to a highly structured template (about 1500-3000 words), using evidence from the final studies chosen, within 8-10 weeks of receiving the literature search.
- Working with *Clinical Evidence* editors to ensure that the final text meets epidemiological and style standards.
- Updating the text every 12 months using any new, sound evidence that becomes available. The *Clinical Evidence* in-house team will conduct the searches for contributors; your task is simply to filter out high quality studies and incorporate them in the existing text.

If you would like to become a contributor for *Clinical Evidence* or require more information about what this involves please send your contact details and a copy of your CV, clearly stating the clinical area you are interested in, to CECommissioning@bmjgroup.com.

Call for peer reviewers

Clinical Evidence also needs to recruit a number of new peer reviewers specifically with an interest in the clinical areas stated above, and also others related to general practice. Peer reviewers are healthcare professionals or epidemiologists with experience in evidence-based medicine. As a peer reviewer you would be asked for your views on the clinical relevance, validity, and accessibility of specific topics within the journal, and their usefulness to the intended audience (international generalists and healthcare professionals, possibly with limited statistical knowledge). Topics are usually 1500-3000 words in length and we would ask you to review between 2-5 topics per year. The peer review process takes place throughout the year, and out turnaround time for each review is ideally 10-14 days.

If you are interested in becoming a peer reviewer for *Clinical Evidence*, please complete the peer review questionnaire at www.clinicalevidence.com/ceweb/contribute/peerreviewer.jsp