Intermediate fibrosis staging in hepatitis C: a problem not overcome by optimal samples or pathologists’ expertise

Maria C. Chindamo,* Vera L. Nunes-Pannain,** João M. Araújo-Neto,* Henrique S. Moraes-Coelho,* Ronir R. Luiz,*** Cristiane A. Villela-Nogueira,* Renata M. Perez*

* Department of Internal Medicine. ** Department of Pathology. *** Institute of Public Health Studies. Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

ABSTRACT

Background and aims. The prediction of intermediate stage of fibrosis in chronic hepatitis C represents a prognostic factor for disease progression. Studies evaluating biopsy performance in intermediate stage considering current patterns of liver samples and pathologists’ variability are scarce. We aimed to evaluate the effect of optimal liver specimens (≥ 20 mm and/or ≥ 11 portal tracts) and pathologists’ expertise on agreement for intermediate stage of fibrosis in chronic hepatitis C. Material and methods. Guided biopsies with large TruCut needle were initially scored by four pathologists with different expertise in liver disease and posteriorly reviewed by a reference hepatopathologist to evaluate fibrosis agreement. Results. Of the 255 biopsies initially selected, 240 met the criteria of an optimal fragment (mean length 24 ± 5 mm; 16 ± 6 portal tracts) and were considered for analysis. The overall agreement among all fibrosis stages was 77% (κ = 0.66); intraobserver and interobserver agreement was, respectively, 97% (k = 0.96) and 73% (κ = 0.60). Excluded samples (< 20 mm and < 11 portal tracts) presented a lower agreement (40%; κ = 0.24). Stratifying fibrosis stages, an interobserver agreement of 42% was found in intermediate stage (F2), ranging from 0 to 56% according to pathologists’ expertise, compared to 97% in mild (F0-F1) and 72% in advanced fibrosis (≥ F3) (p < 0.001). Of the 23% misclassified cases, fibrosis understaging occurred in 82% of specimens, predominantly in F2, even when evaluated by a hepatopathologist. Conclusions. Liver biopsy presents intrinsic limitations to assess intermediate stage of fibrosis not overcome by optimal samples and experienced pathologists’ analysis, and should not be considered the gold standard method to evaluate intermediate fibrosis in chronic hepatitis C.

Key words. Chronic hepatitis C. Liver biopsy. Intermediate stage. Interobserver agreement.

INTRODUCTION

Despite the well-recognized limitations of liver biopsy1-6 and the widespread use of non-invasive methods as substitutes for liver fibrosis assessment,7-10 biopsy is still considered the best and the only procedure available for grading and staging chronic hepatitis C (CHC) in many countries. Studies have shown differences in biopsy performance when assessing intermediate stage of fibrosis in comparison to the extremes stages, and a consequent low performance of non-invasive methods which are validated according to histological staging.11,12 Even considering the tendency to reduce the importance of precise fibrosis assessment with the recent improvement in CHC therapy,13,14 the identification of the intermediate stage remains important to predict disease progression and necessary to prioritize the effective use of new direct-acting antiviral agents, which are costly and not available in many parts of the world.

Considering that smaller samples are associated to fibrosis understaging,15-17 attempts were made in the last decade to define adequate liver specimens aiming to reduce misclassification related to chronic viral hepatitis analysis. Specimens of 20 mm and/or containing at least 11 complete portal tracts (CPT),18 as well as larger samples up to 25 mm,11 are currently considered as the optimal standards.

Correspondence and reprint request: Maria Chiara Chindamo, M.D., Ph.D., Avenida Lúcio Costa 3602, apto 1304 bloco 2, Barra da Tijuca-Rio de Janeiro-Brazil-CEP 22630-010. E-mail: mchiara@terra.com.br

opposed to previous recommendation of 15 mm of length with 6 to 8 portal tracts.19 Nevertheless, it remains unclear whether optimal samples can prevent understaging of intermediate stage of fibrosis. Studies concerning this subject are scarce in the literature and were performed using special technologies, such as slide digitalization, which are not usually applied in daily practice.11,12 Thus, we conducted a study to evaluate the impact of optimal liver samples and the analysis from pathologists with different skills in liver disease, on the agreement for intermediate stage of fibrosis in patients with CHC, considering liver biopsy in the routine daily practice.

MATERIAL AND METHODS

We conducted a cross-sectional study at the Federal University of Rio de Janeiro with prospective inclusion of patients with diagnosis of CHC submitted to percutaneous liver biopsies, as a protocol to evaluate antiviral treatment. Patients with concomitant CHC and human immunodeficiency virus infection, hepatitis B virus, alcohol abuse, metabolic, autoimmune, biliary diseases or liver transplantation were excluded. All consecutive biopsies were guided by ultrasonography using a 14 or 16 G disposable Tru Cut needle (Surecutw, TSK Laboratory, Akasaka, Japan) obtaining specimens with a maximum of 20 mm length on each pass and a core of 1.6 or 1.4 mm in diameter, respectively. Liver biopsies were performed by a senior hepatologist or by a resident supervised by a senior. The liver specimen obtained was visually inspected and if it presented a length of less than 20 mm, an additional fragment was obtained in the same procedure. Liver biopsies were formalin-fixed and embedded in paraffin. Serial sections, 5 μm thick were cut from each paraffin block and routinely stained with hematoxylin and eosin, periodic acid-Schiff diastase, reticulin, Masson Trichrome and Picrosirius red. The length of fragment was verified before and after paraffin inclusion considering the sum of all fragments obtained.

Biopsies were classified according to METAVIR score.20 The initial evaluation of liver samples in the University Hospital is usually assessed by general pathologists on training and further revised by a pathologist with expertise in liver disease. Considering this approach, the slides were initially evaluated by four random pathologists, as follows: a hepatopathologist (H1) with more than 30 years’ experience in liver disease, another hepatopathologist (H2) with 20 years’ experience, a general pathologist, and a post-graduate student. All biopsies were then reviewed a second time by the most experienced hepatopathologist (H1) of the initial group, referred to as the gold standard, who was unaware of patients’ diagnosis or results previously reported, including her own. This procedure was performed in order to analyze intra and interobserver agreement considering the ideal current patterns of length and number of portal tracts reported in literature.18 For a practical approach, fibrosis was categorized as mild (F0-F1), intermediate (F2) and advanced (F3-F4). We considered as major complication of liver biopsy any need for further hospitalization related to procedure within seven days, surgery, or death. The study was approved by the Ethics Committee of the Institution. All patients signed the informed consent form.

Statistical analysis

All data were presented as mean values or proportions. Chi-square ($\chi^2$) test was used to compare categorical variables, applying the Fisher exact test when necessary. For the comparative analysis of non-parametric continuous measures, Mann-Whitney test was applied and, for variables with normal distribution, Student’s $T$ test or ANOVA were performed. The agreement on fibrosis stage between the hepatopathologist H1 and the other initial pathologists’ report was evaluated by Kappa index. Kappa index agreement was interpreted according to following scale: less than 0.20 = poor agreement; 0.20 to 0.40 = fair agreement; 0.40 to 0.60 = moderate agreement; 0.60 to 0.80 = good agreement; 0.80 to 1.00 = very good agreement. Data were analyzed using the statistical package SPSS version 20 for Windows. A $p$ value of $\leq 0.05$ was considered statistically significant.

RESULTS

Liver biopsy procedure and quality of fragments

The initial series of liver specimens consisted of 255 biopsies, of which 240 (94%) were considered adequate according to current patterns of liver samples (length $\geq 20$ mm and/or containing $\geq 11$ CPT) and thus, were included for analysis. We obtained a median length of $31 \pm 8$ mm before and $24 \pm 5$ mm after fixation of the specimen, determining an average reduction of 23% from the original size ($7 \pm 6$
The mean number of portal tracts in the sample was 16±6. Fragment size after fixation presented the following distribution: 98% ≥ 15 mm, 88% ≥ 20 mm and 58% ≥ 25 mm. A number of portal tracts ≥ 11 was found in 86% of the total liver sample.

Concerning the liver biopsy procedure, Tru Cut needle 14 was used in 80% of cases with no difference in fragment size according to the type of needle 14 or 16 both before (31 ± 9 vs. 33 ± 8 mm; p = 0.90) or after fixation (25 ± 5 vs. 24 ± 5 mm; p = 0.728) and no differences regarding the number of CPT (17 ± 6 vs. 15 ± 5; p = 0.63). An average of two passes was performed in each procedure. A greater number of CPT (16 ± 6 vs. 11 ± 3; p = 0.038) was found when two or more passes were made in comparison to one. There was no difference between the number of passes done according to the needle size 14 or 16 G used (p = 0.247). Regarding liver biopsy complications, only one patient presented a symptomatic hepatic hematoma which improved with a conservative approach (0.4%). Distribution of fibrosis stages according to reviser evaluation, scored by META-VIR, was as follows: 2% of patients staged F0, 40% staged F1, 26% staged F2, 24% staged F3 and 8% staged F4.

**Analysis on fibrosis agreement between pathologists**

Considering the 240 patients included presenting optimal liver samples, the overall Kappa index for fibrosis agreement between the reviser hepatopathologist (H1) and the initial report of all four groups was κ = 0.66, representing a complete agreement of 77% across all stages of fibrosis. Conversely, the inadequate fragments which were excluded (< 20 mm long and containing < 11 CPT) showed an agreement in all stages of fibrosis of 40% (κ = 0.24).

The hepatopathologist (H1) intraobserver agreement was κ = 0.96 (n = 35), considering 100% of agreement in mild fibrosis and cirrhosis and one case of F3 stage rescored as F2, demonstrating a high reproducibility of results by this senior hepatopathologist.

The interobserver agreement between the hepatopathologist (H1) analysis and the group comprising three pathologists with different expertise in liver disease is shown in table 1. The interobserver agreement was good between the senior hepatopathologist (H1) and the pathologist experienced in liver disease (H2), moderate when compared to the general pathologist and poor with the post-graduate student. There was no difference in the overall concordant and discordant reports regarding the length (24 ± 5 vs. 24 ± 6; p = 1.0) or number of CPT (16 ± 6 vs. 16 ± 6; p = 0.89).

In respect to the different fibrosis stage, we observed 97% of general interobserver agreement for mild fibrosis (F0-F1), and 72% for advanced fibrosis (F3-F4); however, the agreement for intermediate fibrosis (F2) was only 42% (p < 0.001). The three groups did not differ regarding mean length (p = 0.107). Both presented mean number of CPT higher than 11 nevertheless, with a lower number of CPT in mild fibrosis in comparison to intermediate and advanced fibrosis (14.7 vs. 17.7 vs. 17.8; p = 0.001). Remarkably, the interobserver agreement was progressively lower for intermediate stages of fibrosis according to the level of experience and specialty of pathologists: 56% for the pathologist with experience in liver disease (H2), 18% for the general pathologist and 0% for the post-graduate student (p = 0.004).

Discordant results were found in 56 of the 240 patients (23%) who had the fibrosis stage modified after revision by the experienced hepatopathologist (H1), including one case of her own previous results. The less experienced pathologists underestimated liver fibrosis in 46 of 56 (82%) cases, predominantly in intermediate stage of fibrosis. Of the 46 patients who had the stage of fibrosis underscored, 24 (52%) were initially classified as F1 and then rescored by hepatopathologist (H1) as F2 (Table 2).

**Table 1.** Overall Interobserver agreement in mild, intermediate and advanced fibrosis stages between the experienced hepatopathologist reviser (H1) and the group of pathologists of the first report.

<table>
<thead>
<tr>
<th>Pathologist group of the first report (n = 205)</th>
<th>Kappa index for fibrosis stage</th>
<th>Overall percentage agreement*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathologist with experience in liver disease (H2) (n = 134)</td>
<td>0.68</td>
<td>78%</td>
</tr>
<tr>
<td>General pathologist (n = 50)</td>
<td>0.55</td>
<td>72%</td>
</tr>
<tr>
<td>Post-graduate student (n = 21)</td>
<td>0.18</td>
<td>48%</td>
</tr>
</tbody>
</table>

* p < 0.002

**DISCUSSION**

Identification of intermediate stage of fibrosis in CHC is still an important tool to select patients for antiviral treatment and to optimize the use of resources in clinical management. Nevertheless, studies conducted in laboratory scenario showed a low performance of liver biopsy in the assessment of intermediate stage. The present study aimed to analyze the impact of optimal liver samples and the contribution of pathologists’ variability on agreement for intermediate stage of fibrosis in CHC, considering a clinical setting.

A good quality liver sample is the first requirement to ensure reduction of understaging and interobserver variability. Our study contributed by demonstrating the feasibility of obtaining adequate liver fragments, since performing liver biopsies guided by ultrasound with large needles and an average of two passes yielded mean lengths of 24 ± 5 mm and mean number of CPT of 16 ± 6, with no major complications. These results were more expressive than data presented in a systematic review addressing the quality of liver specimens, where only 32 of 162 studies described quality of samples and reported a mean length of 17.7 ± 5.8 mm and mean number of CPT of 7.5 ± 3.4.

Considering the good characteristics of the samples in our study, an overall agreement between the reviser and the group of pathologists with different skills was 97% for mild fibrosis and 72% for advanced fibrosis. Intermediate fibrosis stage had the worst agreement (42%) ranging from 0 to 56% according to pathologists’ experience and expertise. This suggests that misclassification of the intermediate stage of fibrosis may not be avoided despite optimal liver fragments.

A low level of diagnostic performance in liver fibrosis stages F2 vs. F1 in comparison to performance of F1 vs. F0 or F4 vs. F3 has been recently described by Poynard, et al. They evaluated large surgical samples collected from 20 consecutive patients with chronic liver disease and analyzed digitalized images of 27,869 virtual biopsies of increasing length, demonstrating a raise of overall performance according to biopsy length, except when comparing performance for the diagnosis of F2 vs. F1. However, in that study, the characteristics of the 20 patients included was artificially exacerbated and CHC was represented by only 3 cases. Similarly, Bedossa, et al. using virtual liver specimens of CHC observed that performance of biopsy was lower for the intermediate adjacent stages F2 vs. F1 and better for the extreme stages of F0 and F4. Our study conducted in a clinical scenario of a routine pre-treatment liver biopsy confirmed these observations. This finding represents a great limitation of liver biopsy when considering patients with fibrosis stage METAVIR F2 as candidates for treatment from a histological standpoint.

A condition that greatly influences the misclassification of the intermediate stage of fibrosis is the subjective interpretation of fibrosis considering the

<table>
<thead>
<tr>
<th>Table 2. Analysis of fibrosis staging misclassification considering the experienced hepatopathologist (H1) report as the reference.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrepancy in fibrosis staging according to less experienced pathologists’ analysis (n = 55)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Understaging fibrosis (n = 46)</td>
</tr>
<tr>
<td>F1 to F0</td>
</tr>
<tr>
<td>F2 to F1</td>
</tr>
<tr>
<td>F3 to F2</td>
</tr>
<tr>
<td>F3 to F1</td>
</tr>
<tr>
<td>F4 to F3</td>
</tr>
<tr>
<td>Total, n (%)</td>
</tr>
<tr>
<td>Overstaging fibrosis (n = 9)</td>
</tr>
<tr>
<td>F0 to F1</td>
</tr>
<tr>
<td>F1 to F2</td>
</tr>
<tr>
<td>F2 to F3</td>
</tr>
<tr>
<td>Total, n (%)</td>
</tr>
</tbody>
</table>

H2: hepatopathologist with 20 years’ experience. GP: general pathologist. PGS: post graduate student.
current scoring systems which are prone to considerable intraobserver and interobserver disagreement. This includes difficulties when differentiating true bridging fibrosis from a normal large portal tract extension and an incompletely represented septum located in sample periphery which, in turn, may lead to errors of classification between F1 and F2 METAVIR score.

The impact of the pathologist’s specialty, i.e.: hepatopathologist or a general pathologist, on interobserver variability as well as the level of experience has rarely been evaluated. Rousselet, et al., when analyzing 254 liver biopsies from patients with chronic viral hepatitis, found that the level of experience (duration and location of practice) besides specialization had more influence on agreement than the characteristics of the specimen alone. Our study also demonstrated that pathologists with less experience tend to underscore fibrosis regardless of adequate length and number of CPT, even when considering those pathologists with expertise in liver diseases. The less experienced pathologists understaged fibrosis in 82% of biopsy specimens with disagreement predominantly in the intermediate stage of fibrosis (52%).

A limitation to be considered in this study was the absence of a real gold standard to compare performance between different pathologists’ analyses even after choosing a highly qualified hepatopathologist as the reference for fibrosis evaluation. Consensus readings between two or more pathologists may be an alternative approach to improve the final results. An additional limitation, also shared by most clinical trials, is that the histopathological scoring of liver fibrosis is actually a categorical assessment of architectural changes and not a true measurement of the amount of fibrosis in the liver sample.

Although the applicability of liver biopsy in CHC has been questioned with the development of non-invasive markers and highly efficient antiviral drugs, the adoption of these methods differs from country to country as well as the indication of costly proteases inhibitors-based triple therapy and other new potent direct-acting agents, which still require an allocation based on medical priorities. Thus, liver biopsy still plays an important role for stage scoring in the research setting, which has to be improved to represent the best benchmark for validation of serum surrogate markers of liver fibrosis and other non-invasive techniques for liver disease staging. In this context, one possible explanation for the apparent failure of non-invasive markers to distinguish intermediate stage of fibrosis could be the resulting misclassifications of the biopsy itself, which is an imperfect gold standard as demonstrated in our study.

In order to restore liver biopsy as the true reference when evaluating liver fibrosis, some improvements should be considered. A precise assessment of the amount of liver fibrosis is highly required. For this purpose Picrosirius red staining is the preferred histochemical method to quantify fibrosis, superior to the usually employed Masson trichrome or reticulin stain, due to the affinity for most hepatic collagens. Moreover, computer assisted image analysis associated to properly stained liver sections is the recommended method for measuring fibrosis morphologically using digital image segmentation to assess the area of collagen and the area of tissue. Image analysis do not enable evaluation of the architectural changes included in stage scoring systems (nodularity, fibrous portal linking and portal-central fibrous bridging), thus this technique must be applied as a complementary tool together with histopathological analysis to quantify and evaluate fibrosis progression. To date, image analyses have been used infrequently in clinical settings. Further studies are required to better understand these techniques and to incorporate them in daily practice and future clinical studies regarding fibrosis in CHC.

In conclusion, the main contribution of this study was to demonstrate how the diagnosis of intermediate fibrosis is missed in the daily practice even when a good quality fragment is available and evaluated by a skilled pathologist. Clinicians have to take into account that the risk of biopsy error is greater between stages F2 and F1 than for the extreme stages F1 vs. F0 and F4 vs. F3. In countries where liver biopsy is still the principal method used to evaluate fibrosis in CHC, an understanding of the limitations on the components of the fibrosis assessment process, including the biopsy procedure, quality of sample and pathologists’ reading is of utmost importance. Besides, the application of digitalized images analysis to adequately quantify fibrosis can optimize clinical decision making and ensure reliable information concerning the development of antiviral therapy trials as well as non-invasive techniques for liver fibrosis assessment.

**ABBREVIATIONS**

- **CHC**: chronic hepatitis C.
- **CPT**: complete portal tracts.

DISCLOSURE/ CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES