A Systematic Review of the Quality of Liver Biopsy Specimens

Evangelos Cholongitas, MD,1 Marco Senzolo, MD,1 Richard Standish, MD,2 Laura Marelli, MD,1 Alberto Quaglia, MD,2 David Patch, MD,1 Amar P. Dhillon, MD2 and Andrew Kenneth Burroughs, MD1

Key Words: Liver biopsy; Percutaneous liver biopsy; Transjugular liver biopsy; Length; Portal tract

DOI: 10.1309/W3XCNT4HKFBN2G0B

Abstract

Characteristics for an optimal liver biopsy specimen were recently defined as 20 to 25 mm long and/or containing more than 11 complete portal tracts (CPTs). A systematic review of percutaneous liver biopsy (PLB) and transjugular liver biopsy (TJLB) series yielded only 32 PLB studies in which these characteristics were evaluated: mean ± SD length, 17.7 ± 5.8 mm and number of CPTs, 7.5 ± 3.4; and 15 TJLB studies: mean ± SD length, 13.5 ± 4.5 mm and number of CPTs, 6.8 ± 2.3. Studies of sampling heterogeneity and intraobserver and interobserver variability also used inadequate specimens by present standards. Only 11 (5.3%) of 207 therapeutic studies for chronic hepatitis B and C documented length and/or number of CPTs. Of the current 12 studies evaluating noninvasive fibrosis tests, only 8 documented length or number of CPTs, and only 1 documented length and number of CPTs. New studies are needed based on adequate liver biopsy samples to provide reliable estimation of grading and staging in chronic liver disease.

Liver biopsy (LB) is an important diagnostic tool and helps make therapeutic decisions in acute and chronic liver disease.1 Histopathologic examination is the “gold standard” in chronic hepatitis C (CHC) for assessing changes after antiviral therapy2 and is considered mandatory for grading (necroinflammatory activity) and staging (fibrosis) in most patients,3,4 including patients with persistently normal aminotransferase values,5-8 and also for evaluating steatosis, all histologic features that affect the natural history and therapeutic outcome.9-13 In chronic hepatitis B (CHB), the same applies.14

An LB specimen consists of approximately 1/50,000 of the hepatic mass, but it is considered reasonably representative of the whole liver.15 Several studies have evaluated the following: (1) the optimal size of the LB specimen necessary for accurate evaluation of diffuse liver disease, (2) whether heterogeneity of liver disease represents a real problem in clinical practice, and (3) whether intraobserver and interobserver variation significantly affect LB interpretation and which liver disease severity scoring system is the most reliable.

The status of LB is being challenged by noninvasive tests for the evaluation of fibrosis and its value questioned owing to variable specimen quality. Recent studies have evaluated optimal length16,17 and number of portal tracts17 for accurate grading and staging in chronic viral hepatitis. Thus, there has been further emphasis on the quality of LB leading to an accurate interpretation.

Percutaneous LB (PLB) and transjugular LB (TJLB) are the 2 main techniques. Laparoscopic biopsies and biopsies during laparotomy are more invasive18; endoscopic ultrasound-guided biopsy19 has not been established as an alternative method. The needles are considered large when the external diameter is 1.0 mm or more (14-19 gauge) and thin when
it is less than 1.0 mm (≥20 gauge). Suction (Menghini) and cutting (Tru-Cut) needles are used most often.

PLB is the most common procedure and lasts just a few seconds; it is performed under local anesthesia with lido-caine while the patient holds or her breath after expiration.20,21 Although compared with Menghini needles, Tru-Cut needles usually produce a less fragmented sample,22,23 there are conflicting results about their relative safety.18,23 Major and minor complications occur in up to 6% with PLB, and 0.04% to 0.11% can be life threatening15,24,25 related to the following: (1) technical factors, including experience of operators,26-28 larger needles,21,22 more than 1 pass,23,28-32 and, possibly, not using ultrasonography before or during LB,1,28,30,33-38 and (2) impaired coagulation beyond current safe limits.24 The most important complication is bleeding.

TJLB has been used since 1970 and is an alternative and safe method in high-risk patients, ie, massive obesity, gross ascites, severe coagulopathy, or previous failure of PLB. TJLB has the advantage that the Glisson capsule is not breached except as a procedural complication from within the liver. Bleeding, therefore, is extremely rare.1,15,18,27 Hepatic venous pressure gradient measurements18,27 and carbon dioxide portography can be performed concomitantly.1 Initially, TJLB was an aspiration technique, resulting in excessive fragmentation and small specimens, making diagnosis difficult.39,40 However, the Tru-Cut TJLB needle has improved the technique without increasing complications.41-43 Few series report deaths: mortality is 0.5% or less,44 and complications range from 0.1% to 20% and include abdominal pain, cardiac arrhythmias, capsular perforation, and, rarely, intraperitoneal hemorrhage.1,41 Multiple passes with the TJLB needle have not led to increased complication rates.43,45

Specimen Size and Histologic Evaluation of Grading and Staging

Holund et al46 first studied diagnostic reproducibility relative to specimen size in 100 selected LB specimens that were 25 mm or longer and 1 mm or wider in patients with acute or chronic hepatitis or cirrhosis. Part of the histologic slide was covered with opaque paper to evaluate the influence of various “artificial lengths.” The conclusion was that an LB specimen 5 mm long or longer was adequate for diagnosing acute hepatitis but inadequate for chronic hepatitis or cirrhosis. Subsequently, the same group47 focused on chronic hepatitis, using the same selection criteria (16-gauge Menghini needles with an external diameter of 1.65 mm). Specimens of 15 mm or more were necessary for an accurate diagnosis of chronic aggressive hepatitis.

Another study using the same methods in patients with CHC included 100 PLB specimens that were 20 mm or longer; the needle used was not described.48 The METAVIR scoring system was assessed in different lengths: 5, 10, 15, and 20 mm or longer, with the latter as the reference standard for concordance.49 A 10-mm length was adequate for reliable assessment of necroinflammatory activity and fibrosis (weighted κ, 0.81 and 0.85, respectively).

Colloredo et al17 evaluated 161 LB specimens from patients with CHC and CHB using the Ishak scoring system, excluding biopsy specimens less than 3 cm long but none based on width. Initially, each sample was scored 3 times by evaluating the entire specimen (≥3 cm), the first 1.5 cm, and the first 1.0 cm. Then, the width was reduced to 1.0 mm by using an optical device, and each biopsy specimen was rescored, blindly, twice by evaluating the entire length (≥3 cm) and the first 1.5 cm. The reduction in length led to a significant decrease in number of CPTs and underestimation of necroinflammatory activity. Thus, severe grade was diagnosed in 11.8% of LB specimens that were 3 cm or longer but only 0.6% of LB specimens that were 1.5 cm long (both were 1.4 mm wide), and severe stage was diagnosed in 11.2% of LB specimens 3 cm or longer but in only 3.1% of LB specimens 3 cm or longer but 1 mm wide. A specimen 20 mm or longer and/or containing 11 or more CPTs was necessary for reliable assessment of grading and staging in chronic viral hepatitis. These criteria have been adopted rapidly as optimal standards. However, it is clear that more than 50% were a priori inadequate because 194 were excluded from 355 LB specimens. Second, whether a 16-gauge needle (external diameter, 1.65 mm) results in a constant width of 1.4 mm needs to be questioned because other studies using larger needles (14 gauge; external diameter, 2.1 mm) describe a mean ± SD width of 0.9 ± 0.3 mm.50 In practice, biopsy width is not uniform because of variable tissue shrinkage and because the plane of section cannot always be through the maximum diameter of the biopsy cylinder. Third, the method of changing the width by covering with a straight-edged mask cannot be accurate; most histologic sections do not lie in straight lines.

Bedossa et al16 evaluated the adequacy of LB samples obtained at least 3 cm from the tumor by using image analysis of 17 surgical specimens following resection for hepatocellular carcinoma in patients with CHC. They derived 10,659 virtual liver samples varying from 2.5 to 200 mm in length and a constant 1.2 mm width. The image analysis of fibrosis was converted to the METAVIR scoring system (the reference METAVIR stage was based on the whole sample, at least 2 × 3 cm). Accurate evaluation of fibrosis was achieved in only 65% of 15-mm-long and 75% of 25-mm virtual samples, with no significant improvement with longer samples. The conclusion was that a specimen 25 mm long was the minimum length for reliable staging.

We aimed to evaluate the literature in terms of size and quality of PLB and TJLB specimens in relation to the recently proposed minimum requirements for the assessment of
chronic viral hepatitis and heterogeneity and intraobserver and interobserver variation, particularly with thin-needle LB, using the current histologic scoring systems for grading and staging in CHC and CHB and to examine these variables in clinical trials of antiviral therapy and in studies assessing noninvasive markers of fibrosis.

Materials and Methods

We performed a systematic review of the length and number of CPTs in published series of PLBs using a MEDLINE search (English/non-English) and the following key words: “percutaneous liver biopsy,” “needle,” “Menghini,” “Tru-Cut,” “sample size,” and “length.” Published abstracts from European and American gastroenterology and hepatology conferences during the previous 10 years also were reviewed. Reference lists from these studies were hand searched to identify further relevant articles. A total of 162 studies were evaluated, but only 32 (27 full articles and 5 abstracts) had information about length and/or number of CPTs.22,24,33,35,36,44,50-75 The following variables were extracted: number of patients, number of PLBs performed, type of needle (Menghini or Tru-Cut), size of needle (diameter), and whether the procedure was ultrasound-guided or “blind.”

All data were analyzed by using the statistical package SPSS (version 10.0, SPSS, Chicago, IL). The $\chi^2$ test was used to compare qualitative variables and the $t$ test and Mann-Whitney test to compare quantitative variables, as appropriate. Quantitative variables with normal distribution were expressed as the mean ± 1 SD and with skewed distribution as the median (range). The significance level was set at a $P$ value of .05 or less (2-sided).

Results

Systematic Review of Length and Number of CPTs Obtained With PLB

All 32 studies reported length but only 12 reported the number of CPTs. Table II, 5 did not report mean and range,22,32,53,55,61 but just categories, eg, longer or longer than a certain length. Fragmentation was described in 8 studies, but in only 4 was the mean number of fragments given.35,50,54,59 There were 8,746 patients from whom 10,027 PLB specimens were obtained. The needle size was from 14 to 19 gauge (median, 16 gauge) (Table 1). There were 4,481 Menghini needle biopsies, 4,134 Tru-Cut needle biopsies, and 1,412 of unknown type. The mean ± SD length and number of CPTs were 17.7 ± 5.8 mm and 7.5 ± 3.4 mm, respectively. The correlation between length and CPTs was poor (Spearman $r = 0.45; P = .04$). PLB specimens obtained during the 1996-2005 period compared with those obtained before 1996 were significantly longer (19.8 vs 15.7 mm; $P = .033$) and were obtained more frequently with ultrasound guidance (9 vs 2 studies; $P = .001$) using smaller needles (18 or 19 gauge; 6 vs 2 studies; $P = .023$). The Menghini needle yielded significantly longer samples (19.9 ± 6.6 mm) compared with the Tru-Cut needle (14.3 ± 3.2 mm; $P = .016$), but without a significant difference in the number of CPTs (7.3 vs 6.9; $P = .8$).

Table 2. Only 1 study using the Tru-Cut needle70 documented the range in length, so it was not possible to assess whether there was less variability in length with the Tru-Cut than with the Menghini needle.

Because the Tru-Cut needle provides a maximum length of sample determined by the notch in the needle shaft (usually 20-25 mm; compared with the Menghini in which the length depends on the force of aspiration and operator experience), this could explain the longer samples obtained with the Menghini. Another reason could be that more passes were performed with the Menghini. However, 16 of 32 studies that gave such information showed that more than 1 pass was performed in 108 (3.1%) of 3,535 biopsies using the Menghini compared with 199 (12.1%) of 1,646 biopsies using the Tru-Cut.

Ultrasound Guidance

A total of 5,392 specimens from ultrasound-guided procedures and 1,369 specimens from blind biopsy procedures were analyzed. Specimens from ultrasound-guided biopsies were longer than specimens from blind biopsy procedures (20.5 vs 14.4 mm; $P = .021$), possibly because ultrasound guidance gives rise to greater confidence in performing a biopsy. However, ultrasound-guided biopsy specimens did not contain significantly more CPTs than specimens from blind biopsy procedures (8.3 vs 5.3; $P = .13$). Specimens obtained using the Menghini needle and ultrasound guidance were significantly longer than specimens obtained with the Tru-Cut and ultrasound guidance (24.4 vs 13.6 mm; $P = .017$) and specimens obtained “blindly” using the Menghini (15.8 mm; $P = .017$) (Table 2). There was no significant difference in the length of specimens obtained with the Tru-Cut needle with ultrasound guidance vs blindly.

Center Experience

We assumed that larger studies would be published by more experienced operators. LB samples were longer in studies with 100 or more PLBs than in those with fewer than 100 PLBs (20.4 mm vs 16 mm; $P = .026$). However, this difference was not significant for the number of CPTs (8 vs 7.3; $P = .7$). Specimens obtained with Menghini needles were significantly longer in studies with 100 or more PLBs than in those with fewer than 100 PLBs (24 vs 16.1 mm; $P = .005$), in contrast with studies in which Tru-Cut needles were used.
and there were more or fewer than 100 PLBs (12.3 vs 14.8 mm, respectively; \( P = .27 \); Table 2). In studies with fewer than 100 PLBs, ultrasound guidance did not help to obtain longer specimens (ultrasound-guided vs non–ultrasound-guided, 17.9 vs 13.6 mm; \( P = .19 \)).

### Needle Size

There was no significant difference in length (range, 16.3-20.7 mm) or number of CPTs (range, 4.6-9.7) according to needle diameter. Longer biopsy specimens (mean, 20.7 mm) containing a larger number of CPTs (mean, 9.7) were obtained by using 17-gauge needles, but these results are derived from only 3 studies (2 from the same center\(^57,59\) and 1 on cadavers\(^59\); Table 1). PLB specimens obtained by 18- or 19-gauge needles compared with smaller ones had similar mean length (18.4 vs 18.6 mm) but contained more CPTs (8.0 vs 6.0); however, this difference was not significant. The Menghini and Tru-Cut needles were compared only in studies using 14-, 15-, and 18-gauge needles. LB specimens were, on average, longer when 14-gauge Menghini (23 vs 15.5 mm; \( P = .18 \)) or 15-gauge Menghini (21 vs 14 mm; \( P = .47 \)) needles were used, whereas specimens obtained using 18-gauge Menghini needles were significantly longer than those obtained with 18-gauge Tru-Cut needles (26 vs 12.8 mm; \( P = .012 \)).

### Quality of LB Specimens in Trials of Antiviral Therapy

We evaluated clinical trials from 1996 to 2004 using interferon, ribavirin, lamivudine, or adefovir in 147 trials for antiviral therapy. LB specimens were, on average, longer when 14-gauge Menghini (23 vs 15.5 mm; \( P = .18 \)) or 15-gauge Menghini (21 vs 14 mm; \( P = .47 \)) needles were used, whereas specimens obtained using 18-gauge Menghini needles were significantly longer than those obtained with 18-gauge Tru-Cut needles (26 vs 12.8 mm; \( P = .012 \)).

### Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of Patients</th>
<th>Tru-Cut</th>
<th>Menghini</th>
<th>B</th>
<th>US</th>
<th>Mean Length (mm)</th>
<th>Mean No. of Portal Tracts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gilmore et al(^{22})</td>
<td>1,500</td>
<td>990</td>
<td>510</td>
<td>930</td>
<td>570</td>
<td>&lt;10 mm: M, 4.9%; T, 4.6%</td>
<td>NR</td>
</tr>
<tr>
<td>Gunnesson et al(^{24})</td>
<td>708</td>
<td>NR</td>
<td>1,086 (15 G)</td>
<td>NR</td>
<td>1,086</td>
<td>32</td>
<td>NR</td>
</tr>
<tr>
<td>Lindor et al(^{25})</td>
<td>836</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>B, 16; US, 17</td>
<td>NR</td>
</tr>
<tr>
<td>Papini et al(^{36})</td>
<td>200</td>
<td>200</td>
<td>100 (16 G)</td>
<td>100 (14 G)</td>
<td>100</td>
<td>B, 22; US, 28</td>
<td>NR</td>
</tr>
<tr>
<td>Crawford et al(^{50})</td>
<td>16</td>
<td>NR</td>
<td>16 (14 G)</td>
<td>NR</td>
<td>NR</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>Farrell et al(^{55})</td>
<td>166</td>
<td>201 (15/18 G)</td>
<td>NR</td>
<td>91</td>
<td>110</td>
<td>B, 16.2; US, 15</td>
<td>B, 78; US, 6.3</td>
</tr>
<tr>
<td>Galani et al(^{37})</td>
<td>212</td>
<td>NR</td>
<td>212 (18 G)</td>
<td>NR</td>
<td>212</td>
<td>29</td>
<td>15(^3)</td>
</tr>
<tr>
<td>ter Borg et al(^{52})</td>
<td>184</td>
<td>184 (14 G)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>70% &lt;15</td>
<td>NR</td>
</tr>
<tr>
<td>Angelucci et al(^{53})</td>
<td>501</td>
<td>1,184 (14 G)</td>
<td>NR</td>
<td>NR</td>
<td>1,184</td>
<td>M, 8.4-9.9 mm(^{2}); T, 7.85 mm(^{2})</td>
<td>M, 3.9-4.5; T, 3.96</td>
</tr>
<tr>
<td>Hopper et al(^{54})</td>
<td>10</td>
<td>28 (14 G)</td>
<td>52 (16-18 G)</td>
<td>NR</td>
<td>NR</td>
<td>M, 8.4-9.9 mm(^{2}); T, 7.85 mm(^{2})</td>
<td>M, 3.9-4.5; T, 3.96</td>
</tr>
<tr>
<td>Colombo et al(^{55})</td>
<td>1,179</td>
<td>569 (14 G)</td>
<td>610 (16 G)</td>
<td>NR</td>
<td>NR</td>
<td>&lt;10 mm: M, 12.4%; T, 3.6%</td>
<td>NR</td>
</tr>
<tr>
<td>Rocken et al(^{56})</td>
<td>79</td>
<td>NR</td>
<td>79 (17 G)</td>
<td>NR</td>
<td>79</td>
<td>25.3</td>
<td>9.7</td>
</tr>
<tr>
<td>Brunetti et al(^{57})</td>
<td>149</td>
<td>NR</td>
<td>149 (18 G)</td>
<td>NR</td>
<td>149</td>
<td>21.2</td>
<td>NR</td>
</tr>
<tr>
<td>Petz et al(^{58})</td>
<td>100</td>
<td>NR</td>
<td>41 (17 G)</td>
<td>NR</td>
<td>100</td>
<td>25.5</td>
<td>NR</td>
</tr>
<tr>
<td>Goldner(^{59})</td>
<td>3</td>
<td>15 (14 G)</td>
<td>30 (16/17 G)</td>
<td>NR</td>
<td>NR</td>
<td>M, 11.6; T, 16.4</td>
<td>NR</td>
</tr>
<tr>
<td>Chau et al(^{60})</td>
<td>50</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>McAfee et al(^{64})</td>
<td>50</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>22</td>
<td>NR</td>
</tr>
<tr>
<td>Vargas-Tank et al(^{61})</td>
<td>66</td>
<td>NR</td>
<td>132</td>
<td>NR</td>
<td>NR</td>
<td>≥5 mm: M, 46%; T, 94%</td>
<td>NR</td>
</tr>
<tr>
<td>Torp-Pedersen et al(^{62})</td>
<td>77</td>
<td>NR</td>
<td>77 (19 G)</td>
<td>NR</td>
<td>77</td>
<td>17</td>
<td>NR</td>
</tr>
<tr>
<td>Catulloli et al(^{63})</td>
<td>753</td>
<td>NR</td>
<td>753 (18 G)</td>
<td>NR</td>
<td>753</td>
<td>278</td>
<td>NR</td>
</tr>
<tr>
<td>Chevallier et al(^{64})</td>
<td>600</td>
<td>600 (18 G)</td>
<td>NR</td>
<td>NR</td>
<td>600</td>
<td>9.9</td>
<td>5.7</td>
</tr>
<tr>
<td>Fliam et al(^{65})</td>
<td>74</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>12.3</td>
<td>NR</td>
</tr>
<tr>
<td>Siddique et al(^{66})</td>
<td>30</td>
<td>NR</td>
<td>30 (15 G)</td>
<td>NR</td>
<td>NR</td>
<td>16.5</td>
<td>4.5</td>
</tr>
<tr>
<td>Kim et al(^{67})</td>
<td>304</td>
<td>NR</td>
<td>171</td>
<td>304</td>
<td>B, 11; US, 16.2</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Bateson et al(^{68})</td>
<td>77</td>
<td>41 (14 G)</td>
<td>36 (15 G)</td>
<td>NR</td>
<td>NR</td>
<td>M, 20.7; T, 15</td>
<td>NR</td>
</tr>
<tr>
<td>Meng et al(^{69})</td>
<td>277</td>
<td>NR</td>
<td>277 (15 G)</td>
<td>NR</td>
<td>NR</td>
<td>15</td>
<td>NR</td>
</tr>
<tr>
<td>Maharaj et al(^{70})</td>
<td>40</td>
<td>40</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>16.3</td>
<td>NR</td>
</tr>
<tr>
<td>Gurakar et al(^{71})</td>
<td>76</td>
<td>76 (14/15 G)</td>
<td>NR</td>
<td>NR</td>
<td>21 (14 G); 14 (15 G)</td>
<td>5.2 (14 G); 6 (15 G)</td>
<td></td>
</tr>
<tr>
<td>Spirecho et al(^{72})</td>
<td>145</td>
<td>78 (18 G)</td>
<td>67 (18 G)</td>
<td>NR</td>
<td>145</td>
<td>M, 12.6; T, 16</td>
<td>M, 7.2; T, 8.1</td>
</tr>
<tr>
<td>Regan et al(^{73})</td>
<td>98</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>19.1 mm(^{2})</td>
<td>5.2</td>
</tr>
<tr>
<td>Judmaier et al(^{74})</td>
<td>136</td>
<td>62</td>
<td>74</td>
<td>NR</td>
<td>NR</td>
<td>M, 8; T, 12</td>
<td>M, 6; T, 16</td>
</tr>
<tr>
<td>Steadman et al(^{75})</td>
<td>50</td>
<td>NR</td>
<td>50</td>
<td>NR</td>
<td>NR</td>
<td>15.9 mm(^{2})</td>
<td>NR</td>
</tr>
</tbody>
</table>

B, blind; NR, not reported; US, ultrasound.

\(^1\) Data are given as number of liver biopsy specimens; when reported, the needle gauge (G) is given in parentheses. Translation of gauge to external diameter of needle (mm) is as follows: 14 G = 2.1; 15 G = 1.83; 16 G = 1.65; 17 G = 1.47; 18 G = 1.24; and 19 G = 1.06.

\(^2\) When reported the type of needle (M, Menghini; T, Tru-Cut) and relevant data are given.

\(^3\) Expression of length and/or number of portal tracts as a percentage more than a specific level.

\(^4\) Inaccurate measurement (by assumption of study authors).

\(^5\) Cadavers.

\(^6\) Expression of size as surface area (mm\(^{2}\)).
CHC and 60 for CHB in which the histologic grade and stage were evaluated formally. Only 8 studies documented the type and only 9 the size of the needle. Only 11 studies for CHC provided information on LB specimen quality: length in 3,76-78 number of CPTs in 6,79-84 and both in 2.85,86 Thus, surprisingly only 2 of 147 studies had the relevant background information to assess whether histologic assessment was based on an adequate or optimal biopsy sample. There was even less information on interobserver and intraobserver variation, which was evaluated in only 379,87,88 and 279,82 studies, respectively. In CHB studies, none provided information on the quality of liver biopsy specimens, and only 3 studies assessed intraobserver and interobserver variation.89-91

**Evaluation of Potential Heterogeneity of Liver Disease With PLB**

We found 5 studies Table 3.65,66,92-94 Only 1 study had biopsy specimens of adequate length, and in the 50 patients with CHC studied, ultrasound-guided PLB of the right lobe (28 ± 11 mm) and left lobe (25 ± 9 mm) showed no difference between grading and staging in the paired biopsy specimens.94 In the other studies, all had significant variability: 1 did not document length,92 1 had a mean length of only 12.3 mm,65 and 2 evaluated biopsy specimens selected as 15 mm or longer, one laparoscopic93 and the other PLB.66

### Intraobserver and Interobserver Variation and Scoring Systems in PLB

We found 6 studies95-100 Table 4. and only 196 used samples of adequate length (≥40 mm). The Scheuer system had excellent results for intraobserver and interobserver agreement, as did the Knodell system for fibrosis but not for inflammatory score.

In the other studies, 1 did not document length,97 3 used samples 10 mm or longer,95,98,100 and 1 used samples 15 mm or longer.99 The study that included histopathologists with different levels of expertise, duration, and location of practice100 and had an excellent design only used specimens 10 mm or longer, and, thus, its results may not be applicable to optimal LB specimens. In fact, agreement increased in relation to length and number of portal tracts.

### Thin-Needle vs Large-Needle PLB for Assessment of Diffuse Liver Disease

Rocken et al56 compared the Menghini thin needle, 20 and 21 gauge, with the conventional Menghini large needle, 17 gauge, in cases with no differences in indications for biopsy or histologic diagnoses. LB specimens that were Ishak stage 5 and 6 were excluded; 343 biopsy specimens were obtained from 258 patients: 17-gauge needle used by surgeons using several passes for 28 biopsies (17Gs); single-pass percutaneous for 79 biopsies using a 17-gauge needle (17Gp); and ultrasound guidance with a 20-gauge needle in 88 biopsies (20Gp) and a 21-gauge needle in 80 biopsies (21Gp). The authors found that specimens in the 20Gp group, compared with specimens in the 17Gp group, were longer (29.8 vs 25.3 mm; P < .05) but contained fewer

---

**Table 2**

Liver Biopsy Specimen Length and Number of Portal Tracts in 32 Studies Categorized by Use of Tru-Cut and Menghini Needles in PLBs and by Guidance (Ultrasound vs Blind) and Experience

<table>
<thead>
<tr>
<th>PLB Tru-Cut</th>
<th>Menghini</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (n = 8,615)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length, mm</td>
<td>14.3 ± 3.2</td>
<td>19.9 ± 6.6</td>
</tr>
<tr>
<td>No. of CPTs</td>
<td>6.9 ± 3.6</td>
<td>7.3 ± 3.6</td>
</tr>
<tr>
<td>Ultrasound (n = 5,392)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length, mm</td>
<td>13.6 ± 3.2</td>
<td>24.4 ± 5.9</td>
</tr>
<tr>
<td>No. of CPTs</td>
<td>6.9 ± 1.6</td>
<td>8.4 ± 3.9</td>
</tr>
<tr>
<td>Blind (n = 1,369)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length, mm</td>
<td>12 ± 2.8</td>
<td>15.8 ± 4.1</td>
</tr>
<tr>
<td>No. of CPTs</td>
<td>5.6 ± 0.7</td>
<td>4.2 ± 0.4</td>
</tr>
<tr>
<td>Experience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length, mm</td>
<td>12.3 ± 2.9</td>
<td>24 ± 5.7</td>
</tr>
<tr>
<td>No. of CPTs</td>
<td>5.6 ± 1.6</td>
<td>9.7 (4.5-15)</td>
</tr>
<tr>
<td>No experience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length, mm</td>
<td>14.8 ± 3.2</td>
<td>16.1 ± 5.1</td>
</tr>
<tr>
<td>No. of CPTs</td>
<td>5.7 (4.6-16)</td>
<td>6.5 ± 2.1</td>
</tr>
</tbody>
</table>

CPT, complete portal tract; PLB, percutaneous liver biopsy.

*Experience was defined as studies with more than 100 PLBs and no experience as studies with fewer than 100 PLBs. Variables with normal distribution are expressed as mean ± 1 SD and those with a nonnormal distribution as median (range).

---

**Table 3**

Studies Evaluating the Heterogeneity in Grading and Staging of Chronic Hepatitis C

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of Cases</th>
<th>Needle Size</th>
<th>Specimen Length (mm)</th>
<th>No. of Portal Tracts</th>
<th>Scoring System</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flamm et al95</td>
<td>74</td>
<td>NR</td>
<td>12.3 (mean)</td>
<td>NR</td>
<td>Knodell</td>
<td>66%</td>
</tr>
<tr>
<td>Fanning et al92</td>
<td>12</td>
<td>NR</td>
<td>16</td>
<td>All ≥5</td>
<td>Ishak</td>
<td>66%; grade, 75%; stage, 75%</td>
</tr>
<tr>
<td>Regev et al93</td>
<td>124</td>
<td>16 G</td>
<td>All ≥15</td>
<td>All ≥5</td>
<td>Scheuer</td>
<td>98.4%; stage, 92.7%</td>
</tr>
<tr>
<td>Persico et al94</td>
<td>50</td>
<td>18 G</td>
<td>All ≥15</td>
<td>NR</td>
<td>Ishak</td>
<td>Right vs left lobe: grade, 8.13 vs 8.06; stage, 2.16 vs 2.13</td>
</tr>
<tr>
<td>Siddique et al86</td>
<td>29</td>
<td>1.7 mm</td>
<td>All ≥15</td>
<td>All ≥5</td>
<td>Knodell</td>
<td>31%; grade, 79.3%</td>
</tr>
</tbody>
</table>

NR, not reported.

* Translation of gauge (G) to external diameter of needle (mm) is as follows: 16 G = 1.65; and 18 G = 1.24.

† Defined as <2-point difference in score.
portal tracts (6.7 vs 9.7). An insufficient sample was obtained in 4 cases in the 20Gp group, and in only 1 in the 17Gp group. The authors concluded that 20Gp could be a reliable alternative for patients with diffuse liver disease and contraindications for large-needle (eg, 17Gp) percutaneous biopsy.

Petz et al.58 examined the feasibility of thin-needle biopsy for grading and staging in chronic viral hepatitis: 59 patients underwent thin-needle biopsy (20-gauge, 0.9-mm needle) and 41 underwent large-needle biopsy (17-gauge, 1.4-mm needle). All samples were read first separately and then together by 2 independent pathologists using the Ishak scoring system. The sample was considered adequate in all but 4 thin-needle biopsies. No significant difference was found for grading and staging between thin-needle and large-needle specimens. However, in thin-needle specimens, severe fibrosis (stage 5) and cirrhosis (stage 6) tended to be underestimated. The limitations of the study were that thin-needle and large-needle samples were not paired, the biopsy procedure was neither randomized nor standardized, and there may have been a bias because there was a significantly lower platelet count in patients who had undergone thin-needle biopsy that likely represented more advanced liver disease and/or cirrhosis.

These limitations were overcome in a study in which paired thin-needle (0.8 mm) and large-needle (1.2 mm) biopsy specimens were obtained through the same puncture site from 149 consecutive patients with CHC.57 LB samples were considered adequate if they were 10 mm or longer, contained 4 or more portal tracts, and were not too fragmented. Two hepatopathologists made a joint evaluation using the Ishak scoring system. Large-needle specimens were significantly longer than thin-needle specimens (21.2 vs 12.2 mm; \( P < .001 \)) and less fragmented (11% vs 42%; \( P < .001 \)) and considered adequate more frequently (94% vs 55.7%; \( P < .001 \)). Comparison of the 83 paired and adequate specimens showed that in thin-needle specimens, fibrosis and all 4 categories of necroinflammatory activity were underscored systematically. Finally, thin-needle biopsy resulted in underestimation of cirrhosis (2 of 3 biopsy specimens with stage 5/6).

Similar results were obtained when paired samples of similar length were compared and when the METAVIR and Scheuer scoring systems were used. The authors concluded that thin-needle biopsy should be avoided for grading and staging in patients with CHC.

**Transjugular Liver Biopsy**

Heterogeneity of liver disease and interobserver or intraobserver variation have not been evaluated in TJLB. By using the same search criteria for TJLB as for PLB, we found only 15 studies that documented the length and number of CPTs, the needle size, or number of passes. Table 5 using a Tru-Cut and/or Menghini-type needle,41,42,44-45,60,73,101-109 and most described only small series. Only our large series of TJLB (n = 326) detailed the number of passes (n = 3), needle size (Tru-Cut 19 gauge), length (mean, 22.5 mm), number of CPTs (mean, 8.7) and fragmentation (median, 5).109 Only 1 study compared Tru-Cut (18 gauge) and Menghini-type (16 gauge) needles and found that using the Tru-Cut resulted in significantly longer specimens (12 vs 7 mm; \( P < .05 \)).102 Overall, 1,389 TJLB specimens were evaluated (mean, 2.5 passes per patient); the mean ± SD length was 13.5 ± 4.5 mm (13 of 15 studies), and the mean ± SD number of CPTs was 6.8 ± 2.3 (6 studies) (Table 5). Quality of TJLB requires study because this method allows multiple passes (to obtain adequate samples) with far less likelihood of increasing complication rates.33,45
Histologic Assessment Without LB

LB should be performed only “if the expected benefit exceeds the small risk associated with this procedure.” There has been renewed interest in the noninvasive evaluation of diffuse liver disease. Aspartate aminotransferase (AST)–alanine aminotransferase (ALT) and AST–platelet count ratios have been shown to have significant correlation with the degree of liver fibrosis in patients with chronic viral hepatitis, nonalcoholic fibrotic liver disease, or alcoholic liver disease. A combination of age, γ-glutamyltransferase and cholesterol levels, and platelet count had a very good correlation (area under the receiver operating characteristic curve, 0.86) with liver fibrosis in patients with CHC and a score of less than 4.2 identified patients with a META VIR stage of fibrosis of 0 or 1 with 96% accuracy. Recently, another index using the platelet count, the AST/ALT ratio, and the international normalized ratio was compared with Ishak fibrosis scores of 5 and 6 and was found to have an area under the receiver operating characteristic curve of 0.776 (training set) and 0.808 (validation set), but only 15% of biopsy specimens were ≥25 mm or longer (there was no evaluation of portal tracts because cirrhosis was being evaluated, not chronic hepatitis).

More complex tests, the FibroTest (Biopredictive, Paris, France; FibroSURE LabCorp, Burlington, NC) and ActiTest (Biopredictive; FibroSURE LabCorp), use less common serologic markers and are at least as sensitive as the Forns scale for excluding fibrosis and discriminating significant fibrosis. Transient elastography (FibroScan, Biopredictive; FibroSURE LabCorp) is a novel noninvasive method for the assessment of liver fibrosis. Combined with FibroTest, FibroScan was better for discriminating severe fibrosis and cirrhosis than FibroScan or FibroTest alone. Procollagen III aminopeptide has been considered sufficient to monitor methotrexate-induced fibrosis. None of these noninvasive tests is able to distinguish different stages of fibrosis, and they are considered less reliable than LB; few studies have been done outside of CHC.

Table 5
Systematic Review of 15 TJLB Series Reporting at Least One of the Characteristics of Length, Portal Tracts, Number of Passes, and Fragments of Liver Biopsy Specimens

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of TJLBs</th>
<th>Needle Type*</th>
<th>No. of Passes</th>
<th>No. of Fragments</th>
<th>Mean Specimen Length (mm)</th>
<th>Mean No. of Portal Tracts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chau et al[60]</td>
<td>18</td>
<td>Tr (18 G)</td>
<td>1-3</td>
<td>NR</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>McAfee et al[44]</td>
<td>146</td>
<td>M and Tr</td>
<td>NR</td>
<td>NR</td>
<td>9</td>
<td>NR</td>
</tr>
<tr>
<td>Regan et al[74]</td>
<td>123</td>
<td>Tr</td>
<td>NR</td>
<td>NR</td>
<td>18</td>
<td>NR</td>
</tr>
<tr>
<td>Bull et al[81]</td>
<td>193</td>
<td>Tr</td>
<td>NR</td>
<td>NR</td>
<td>5.6</td>
<td>NR</td>
</tr>
<tr>
<td>Buzzi et al[42]</td>
<td>50</td>
<td>Tr (18 G)</td>
<td>2.2</td>
<td>NR</td>
<td>1-20†</td>
<td>10.4</td>
</tr>
<tr>
<td>Papatheodoridis et al[16]</td>
<td>157</td>
<td>Tr</td>
<td>1.8</td>
<td>NR</td>
<td>14.8</td>
<td>NR</td>
</tr>
<tr>
<td>Sawyer et al[101]</td>
<td>44</td>
<td>M (16 G); Tr (NR)</td>
<td>≤3</td>
<td>NR</td>
<td>6</td>
<td>NR</td>
</tr>
<tr>
<td>Choo et al[102]</td>
<td>711</td>
<td>M (16 G); Tr (18 G)</td>
<td>M, 2.3; Tr, 2.9</td>
<td>NR</td>
<td>M, 7; Tr, 12</td>
<td>NR</td>
</tr>
<tr>
<td>DiMichele et al[103]</td>
<td>13</td>
<td>Tr (19 G)</td>
<td>&gt;3-5</td>
<td>NR</td>
<td>13.6</td>
<td>6</td>
</tr>
<tr>
<td>Kardache et al[74]</td>
<td>29</td>
<td>Tr (18 G)</td>
<td>1</td>
<td>NR</td>
<td>12</td>
<td>≥8‡</td>
</tr>
<tr>
<td>De Hoyos et al[105]</td>
<td>52</td>
<td>Tr (18 G)</td>
<td>2.5</td>
<td>NR</td>
<td>17</td>
<td>6.2</td>
</tr>
<tr>
<td>Elsharkawy et al[106]</td>
<td>100</td>
<td>Tr³</td>
<td>NR</td>
<td>NR</td>
<td>16</td>
<td>NR</td>
</tr>
<tr>
<td>Gorriz et al[107]</td>
<td>77</td>
<td>Tr (18 G)</td>
<td>5.2</td>
<td>NR</td>
<td>15.2</td>
<td>NR</td>
</tr>
<tr>
<td>Little et al[108]</td>
<td>43</td>
<td>Tr (18, 19, 20 G)</td>
<td>2.7</td>
<td>NR</td>
<td>11 (18 G); 15 (19 G)</td>
<td>NR</td>
</tr>
<tr>
<td>Cholongitas et al[109]</td>
<td>326</td>
<td>Tr (19 G)</td>
<td>3</td>
<td>5</td>
<td>22.5</td>
<td>8.7</td>
</tr>
</tbody>
</table>

M. Menghini needle; NR, not reported; Tr, Tru-Cut needle; TJLB, transjugular liver biopsy.
* When the needle gauge (G) was reported, it is given in parentheses. Translation of gauge to external diameter of needle (mm) is as follows: 16 G = 1.65; 18 G = 1.24; and 19 G = 1.06.
† Length per core.
‡ In 14 patients with cirrhosis.
§ 2.2-mm needle diameter.

Discussion

The minimum standards for an optimal PLB for assessing chronic viral hepatitis require longer specimens than before (≥20-25 mm long and/or containing ≥11 CPTs), or the quality of the biopsy was not mentioned and neither was the needle size with 2 exceptions, both of which used 16-gauge needles. The current FibroScan may have limited value for assessing fibrosis in overweight or obese patients.
Table 6

Studies Evaluating Liver Fibrosis With Noninvasive Tests

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of Patients</th>
<th>Cause of Liver Disease</th>
<th>Scoring System</th>
<th>Liver Biopsy Specimen</th>
<th>Needle Size*</th>
<th>Noninvasive Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wai et al,112 2003</td>
<td>192</td>
<td>CHC</td>
<td>Ishak</td>
<td>NR</td>
<td>NR</td>
<td>APRI</td>
</tr>
<tr>
<td>Forns et al,113 2002</td>
<td>476</td>
<td>CHC</td>
<td>META VIR</td>
<td>CPT, ≥6</td>
<td>NR</td>
<td>Age, PLT, GGT, cholesterol</td>
</tr>
<tr>
<td>Imbert-Bismut et al,116 2001</td>
<td>339</td>
<td>CHC</td>
<td>META VIR</td>
<td>L, &gt;10 mm</td>
<td>NR</td>
<td>FibroTest</td>
</tr>
<tr>
<td>Castera et al,119 2005</td>
<td>183</td>
<td>CHC</td>
<td>META VIR</td>
<td>L, 17 mm (median)</td>
<td>NR</td>
<td>FibroScan, FibroTest, APRI</td>
</tr>
<tr>
<td>Poynard et al,117 2003</td>
<td>352</td>
<td>CHC</td>
<td>META VIR</td>
<td>NR</td>
<td>NR</td>
<td>FibroTest, ActiTest</td>
</tr>
<tr>
<td>Hui et al,121 2005</td>
<td>235</td>
<td>CHB</td>
<td>Ishak and Knodell</td>
<td>L, ≥15 mm; CPT, ≥5</td>
<td>16 gauge</td>
<td>BMI, PLT, albumin, bilirubin</td>
</tr>
<tr>
<td>Rosenberg et al,122 2004</td>
<td>211</td>
<td>CHC</td>
<td>META VIR</td>
<td>L, &gt;10 mm</td>
<td>NR</td>
<td>FibroTest</td>
</tr>
<tr>
<td>Rossi et al,123 2003</td>
<td>125</td>
<td>CHC</td>
<td>META VIR</td>
<td>NR</td>
<td>NR</td>
<td>FibroTest</td>
</tr>
<tr>
<td>Ziol et al,124 2005</td>
<td>327</td>
<td>CHC</td>
<td>META VIR</td>
<td>CPT, ≥10</td>
<td>NR</td>
<td>FibroScan</td>
</tr>
<tr>
<td>Lok et al,125 2005</td>
<td>1,141</td>
<td>CHC</td>
<td>Ishak</td>
<td>15% ≥25 mm</td>
<td>NR</td>
<td>PLT AST/ALT INR</td>
</tr>
<tr>
<td>Colletta et al,126 2005</td>
<td>40</td>
<td>CHC</td>
<td>META VIR</td>
<td>Mean, 20 mm; range, 14-25 mm</td>
<td>16 gauge</td>
<td>FibroTest, FibroScan</td>
</tr>
<tr>
<td>Poynard et al,126 2005</td>
<td>283</td>
<td>CHC</td>
<td>META VIR</td>
<td>NR</td>
<td>NR</td>
<td>FibroTest, ActiTest</td>
</tr>
</tbody>
</table>

ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; CHB, chronic hepatitis B; CHC, chronic hepatitis C; CPT, complete portal tracts; GGT, γ-glutamyltransferase; INR, international normalized ratio; L, length of specimen; NR, not reported; PLT, platelet count.

* A 16-gauge needle has an external diameter of 1.65 mm.

However, more than 1 pass is likely to be needed to obtain a PLB specimen of adequate size, which has the potential to increase the complication rate, which increases with needle size and number of passes.28-30,32 For this reason, the clinical applicability of the histopathologic requirement for larger liver biopsy specimens has to be explored critically. A minimum requirement for a routine LB specimen to always have 11 or more CPTs could be unrealistic and dangerous for the patient on one hand; on the other hand, the realization that inadequate samples are unreliable would make LB histopathologic examination irrelevant at best and dangerous at worst.

This said, it is surprising that only 2 of 147 studies of antiviral therapy for CHC and none for CHB had details on both the length of the LB specimen and the number of CPTs. In addition, inadequate LB specimens have been used (whenever data about the quality of LB specimens was given) in studies of noninvasive markers of fibrosis, so that no study to date is sufficiently reliable to establish the validity of nonhistologic markers. Thus, the interpretation of results in all of these studies will have flaws.

In this systematic review, comprising all documented series of PLB in the literature, the LB specimens had an average length and number of portal tracts well below the published minimum sample size requirements16,17 in more than half the cases. How can adequate biopsy samples be obtained for reliable grading and staging of chronic liver disease?16 Rocken et al56 showed that all methods of LB resulted in an insufficient sample size in a significant proportion of patients: 42% of PLBs with a large 17-gauge needle contained 10 or more portal tracts. Only the surgically obtained LB specimens with multiple passes provided adequate liver samples in a very high proportion of cases. Although using a thin needle allows multiple passes without increasing complications, this advantage is overcome by its low diagnostic performance.57 Although specimens obtained with Menghini needles are significantly longer than those obtained with Tru-Cut needles (19.5 vs 14.3 mm; P = .01) the number of CPTs was no different. A new Tru-Cut needle with a larger notch (at least 30 mm) may overcome this but could result in more complications.

The number of CPTs emerged as the key factor for considering the adequacy of LB specimens.15,17,50 However, we could not completely assess the data on CPTs because the definition of completeness was rarely stated in the relevant studies. Rocken et al50 for PLBs, similar to our study for TJLBs,109 used the definition of Crawford et al50 for CPTs: complete circumference with at least 2 portal structures within them. Colloredo et al17 considered CPTs as only the portal triads with complete circumference in the normal liver. This may explain in part why poor correlation was documented between length and CPTs (although still statistically significant, r = 0.45; P = .04). In addition, fragmentation will reduce number of CPTs if the break occurs through them.

In contrast with the risks of PLB with multiple passes, TJLB offers the possibility of using multiple passes without increasing complications.28-30,127 TJLB has been considered a second-class biopsy owing to the small specimens and increased fragmentation compared with PLB. However, our review has shown that with a mean of 2.5 passes, the biopsy specimens are on average only 4.2 mm shorter compared with...
PLB (13.5 mm vs 17.7 mm, respectively), and, it is important to note, contain almost the same number of CPTs (6.8 vs 7.5, respectively), which is similar to the difference between Tru-Cut and Menghini needles. Even though the Menghini needle under ultrasound guidance gave the best average length, 24.4 mm, the mean number of CPTs was only 8.4. In our center, TJLBs are always performed with 3 passes providing LB specimens with a mean length of 22.5 mm and a mean number of CPTs of 8.7 (Table 5), similar to the “best” PLB technique. Fragmentation was not excessive using the Tru-Cut technique (median fragment number, 5) and in only 5 (1.5%) of 326 were the biopsy specimens too small or fragmented to provide a diagnosis. We are now evaluating the use of 4 passes to see whether an “ideal” specimen can be obtained consistently and in most patients. Therefore, TJLB could be an alternative and safe approach to obtain samples of adequate size and a reliable assessment of liver histologic features, particularly in clinical trials.

Despite the current enthusiasm for using noninvasive tests to diagnose the degree of fibrosis, further prospective studies are needed to validate diagnostic accuracy and usefulness. However, these studies must use optimally interpreted and adequate LB specimens. The question is whether LB can be regarded as the gold standard for the staging and grading of diffuse liver diseases when risks of biopsy, inadequate sampling, and intraobserver and interobserver error are taken into account. If the currently proposed minimal criteria for an LB specimen (≥20-25 mm long and ≥11 CPTs) are to be used as a gold standard, more than 1 pass using a standard PLB will be required, with more risk of complications. Our review suggests that recent improvements in TJLB techniques offer the possibility of safely obtaining ideal LB samples. These issues assume additional importance when changes in LB histopathologic features and noninvasive tests are used as endpoints in clinical trials. Studies that have been performed using inadequate biopsy specimens by present standards must be considered insufficiently reliable to guide clinicians. New studies are needed based on adequate LB samples.

References


