Liver biopsy has long been the gold standard for the grading of hepatic inflammation and the staging of hepatic fibrosis, and has been used as the reference standard method in evaluations of plasma markers of liver disease. In the modern era with the use of ultrasound guidance, biopsy is an essentially safe procedure with a 0.3% rate of serious complications, such as postprocedure pain or bleeding. However, a liver biopsy samples only 1/50,000th of the liver parenchyma and is prone to sampling error, especially when there is a focal component to the liver disease. In addition, the quality of the liver biopsy and the experience of the pathologist are essential for correct grading and staging of liver disease.

More recently, clinical investigators have been searching for noninvasive serum markers of fibrosis (1). These can be individual markers or a series of markers from which a fibrosis index can be derived. In either case, these marker tests must have the following characteristics: They must be reliable, accurate, reproducible, and easy to perform. In addition, they must reflect total mass of liver collagen and extracellular matrix (ECM) and be able to reflect both fibrogenesis and fibrosis regression. The ideal marker test would be able to accurately stage disease and also be sensitive to changes in fibrosis induced by therapy or the natural history of disease progression.

In this issue of Clinical Chemistry, Poynard et al. (2) address one of the most important issues in this area: Is there a true gold standard for liver fibrosis? The traditional approach to the development of serologic tests for fibrosis has been to compare them with liver biopsy, which is taken as the gold reference standard, and to attribute discordance to the failure of the fibrosis markers. Poynard et al. (2) have taken a more creative approach and examined all discordant results between the FibroTest/ActiTest and liver biopsy, assigning a rigid set of criteria as to what institutes a failure of each test. Discordance was seen in 154 of 537 patients (28.7%). Discordances were evenly distributed between discordances for both fibrosis and inflammation. Discordance was attributable to fibrosis-marker failure in 13 patients (2.4%) and to biopsy failure in 97 (18%). No obvious reason for discordance was seen in 44 patients (8.2%). Fibrosis-marker failure was most often attributable to inflammation or hemolysis, whereas biopsy failure was secondary to small biopsy size or steatosis.

Overall these findings are not particularly surprising, but nevertheless, they are very important. In recent years, the inability of liver biopsy to accurately stage fibrosis has received renewed attention. In a recent seminal report, Bedossa et al. (3) examined the sampling variability of liver biopsy in chronic hepatitis C. Measurement of fibrosis by image analysis of liver biopsies showed a CV of 55% with 15-mm biopsies and 45% with 25-mm biopsies. When they used the Metavir scoring system, the CVs improved to 35% and 25% for the respective biopsy sizes.

Only 13.8% of the 537 biopsies in the report by Poynard et al. (2) were >25 mm, and all were performed at an experienced medical center. We can deduce from these studies that bigger is clearly better for liver biopsy but that this is seldom achieved, even in experienced centers.

In addition to the issues with biopsy size, there is also variability in sampling, which can lead to incorrect staging of disease. A recent study (4) compared percutaneous biopsy with laparoscopic biopsy and demonstrated that cirrhosis was missed in almost 30% of cases by the percutaneous biopsy. We can deduce from these studies that the potential for error in staging disease can be as high as 35% and that even cirrhosis can be missed in 30% of patients. Clearly the liver-biopsy standard is not gold but a burnished bronze.

Understanding the limitations of biopsy has important clinical implications. The majority of liver biopsy studies have been done in patients with hepatitis C virus (HCV) infections. Liver biopsy is clearly not necessary for the diagnosis of HCV but is used by many clinicians for treatment decisions. Patients with fibrotic disease, stage 2 or greater by Metavir, are usually encouraged to undergo treatment with interferon and ribavirin. The inability to correctly stage disease has profound clinical implications, and the biopsy should always be interpreted in the clinical context of disease.

As treatments for liver disease improve, biopsy may be less relevant for treatment decisions. Already in patients with HCV genotypes 2 and 3, where treatment achieves sustained viral eradication in 80% of cases, biopsy is often not performed. Although most available serum markers do not attempt to exactly stage disease as biopsies do, but rather give an indication of whether there is mild fibrosis (stage 0–1) or significant fibrosis, the information from markers is equivalent to that from biopsy in deciding to treat or not treat HCV infection as described in the clinical scenario above.

The finding that discordance is more often explainable by failure of the liver biopsy than by errors in the markers might be considered as evidence that fibrosis markers are more accurate than biopsy. Unfortunately, this may not be the case because it is impossible to discern what is truth. To really compare biopsy with biomarkers, the gold standard would have to be laparoscopic examination and biopsy of the liver. The cost, risks, and logistics make such a study very unlikely. The FibroTest/ActiTest is based on a five-marker algorithm for fibrosis and the addition of alanine aminotransferase as a sixth marker for inflammation. Although this index is not measuring direct markers of ECM production or removal, studies by Poynard's group (5–7) have shown that FibroTest provides a linear biochemical evaluation of liver fibrosis. FibroTest has been validated in cross-sectional studies for fibrosis staging in many liver diseases and has an overall area under the ROC curve (ROC area) for fibrosis diagnosis of ~0.8
This is certainly comparable to the current status of liver biopsy for diagnosing cirrhosis and fibrosis.

Many different serum fibrosis markers have been studied in liver disease, including single markers of ECM, such as hyaluronic acid, and combinations of ECM markers and markers of hepatic function, and these have been reviewed recently. Overall, nearly all markers have ROC areas for fibrosis staging of 0.8 or better, and the clinical question now is not whether we need more markers but how to best use the current established markers. In this study, Poynard et al. have suggested a clinical algorithm for the incorporation of serum tests into the evaluation and treatment of patients with HCV. These types of clinical paradigms require validation in prospective studies, and doing so should be a priority of clinical research studies. Studies should incorporate how to use the serum markers to follow clinical disease over time and whether serum markers can predict patients likely to have disease progression.

Although serum markers clearly have an emerging clinical role in cross-sectional staging of disease, many questions must be answered before we can abandon liver biopsy. The major pitfall of all of the current serum fibrosis markers is their lack of ability to differentiate small changes in the state of the ECM. The ability to quantify changes in the ECM will be critical as novel therapies for liver fibrosis are being developed. We need to focus on novel techniques for measuring serum fibrosis, and there have been some promising advances in the areas of proteomics and proteoglycomics. Translating the promise of new laboratory advances in diagnosis of fibrosis into clinical patient care will be required before we retire liver biopsy from our diagnostic armamentarium in the evaluation of liver disease.

References

Nezam H. Afdhal
Beth Israel Deaconess Medical Center
Harvard Medical School
110 Francis St.
Liver Center, Boston, MA 02125
E-mail nafdhal@bidmc.harvard.edu

DOI: 10.1373/clinchem.2004.035899