Liver biopsy: The best, not the gold standard

Pierre Bedossa¹,*, Fabrice Carrat²

¹Department of Pathology, Beaujon Medical Center, Assistance Publique-Hôpitaux de Paris, INSERM, U 773, Paris-Diderot University, Clichy, France
²Epidemiology of Infectious Diseases, UMR-S 707, UPMC & INSERM, Public Health Unit, Assistance Publique-Hôpitaux de Paris, Hôpital Saint-Antoine, Paris, France

Fibrosis, the hallmark of chronic liver diseases, is one of the major deleterious processes associated with chronic hepatitis C. Staging of fibrosis relies on an evaluation of several histological features including assessment of extent of the extracellular matrix deposit, the localization of the deposits within the liver lobule and changes in lobular architecture. These features are then integrated into a semiquantitative scoring system. Histological staging of fibrosis has gained acceptance as a major element in evaluation of liver damage in hepatitis C. Indeed, staging mirrors the natural evolution of chronic hepatitis, predicts evolution toward development of cirrhosis and end-stage liver complications, contributes to predicting a sustained response to antiviral treatment. This is crucial as cirrhosis, the end-point of fibrosis, is the main cause of morbidity and mortality in chronic liver diseases [1–4].

Because fibrosis implies morphological damage, liver biopsy has come to be the natural gold standard for staging the disease. However, the high prevalence of chronic hepatitis C in addition to the cost and constraints generated by this procedure has triggered an intensive search for alternative methods for staging the disease. How to evaluate the performance of these surrogates and how the inherent limits of the biopsy influence the evaluation of accuracy of surrogates are discussed in this issue of the Journal by Mehta and colleagues [5]. This is a relevant question since liver biopsy carries potential limitations including sampling errors and interobserver variations [6,7]. Although several means exist for minimizing these risks such as procurement of biopsies of sufficient length [8] and interpretation of biopsies by experienced liver pathologists [9], staging of fibrosis with biopsy will always carry a risk, albeit low, of misclassification thus making the term “best” standard more appropriate than “gold” standard for liver biopsy.

The performance of any surrogates is classically evaluated by calculation of the area under the receiver operating characteristic curve (AUROC) using liver biopsy as the reference. In this setting, the AUROC represents the probability that a surrogate will correctly rank two randomly chosen patients, one with a liver biopsy considered “normal” and the other “diseased”. Because liver biopsy is not the gold standard but is the best available standard, a perfect surrogate will never reach maximal value (i.e. 1). Taking into account a range of accuracies of the biopsy and a range of prevalences of significant disease (that influence the AUROC), Metha et al. demonstrate that in the most favorable scenario, an AUROC > 0.90 cannot be achieved when assessing the so-called “significant fibrosis” even for a perfect marker [5]. This is important for several reasons. First, studies have already shown that these maximal AUROC values have been reached for surrogates, especially when assessing cirrhosis versus non-cirrhosis, suggesting that these surrogates may be at least as good as liver biopsy in the diagnosis of cirrhosis [10].
suggest that a definitive method for assessing the performance of surrogate markers would employ a clinical end-point rather than biopsy as gold standard. These conclusions should be discussed in further detail before accepting them definitively.

The main alternatives to liver biopsy that have been developed in the past 10 years are based on two very different concepts: serum markers and liver stiffness [11]. They differ substantially both in their rationale and in their conception.

Stiffness, as assessed by ultrasound (Fibroscan) and more recently by MRI, evaluates the velocity of propagation of a shock wave within the liver tissue. This method examines a physical parameter of liver tissue which is related to its elasticity. Thus, liver biopsy is used to choose the best discriminative thresholds to predict histological stage. The main drawback is that additional space-occupying lesions often encountered in hepatitis C such as steatosis, edema and inflammation will develop within an organ wrapped in a distensible non-elastic envelope (Glisson’s capsule), contribute to modifying liver texture and may act as a confounding factors when stiffness is concerned. Nevertheless, there exist strong arguments supporting the hypothesis that elasticity parallels staging at precirrhotic or cirrhotic stages. A recent meta-analysis showed that the AUROC of liver elasticity parallels staging at precirrhotic or cirrhotic exists strong arguments supporting the hypothesis that factors when stiffness is concerned. Nevertheless, there exist strong arguments supporting the hypothesis that elasticity parallels staging at precirrhotic or cirrhotic stages. A recent meta-analysis showed that the AUROC reaches the “holy grail” of 0.90 for diagnosis of cirrhosis with Fibroscan [8]. However, it is noteworthy that changing the definition of “diseased” liver from F4 to F3F4 or F2F3F4 is associated with a progressive decrease in the AUROC, suggesting that this approach is valid for diagnosis of cirrhosis but less adequate when assessing transition from one stage to the upper one, a crucial goal for treatment decision or patient follow-up. In this setting, the proposal of a clinical reference (liver-related death, end-stage liver complications) for comparing the performances of Fibroscan and biopsy for diagnosis of cirrhosis is meaningful and seems feasible. In the mean time, assessing the prognostic value of the wide range of stiffnesses observed within cirrhotic livers should be useful since this would overcome one major limitation of the biopsy (i.e. one histological stage for all type of cirrhosis).

Validation of surrogates compared to a reference other than biopsy is completely different when addressing serum markers. Serum markers are combinations of several blood parameters that are optimized to mirror the stage of liver fibrosis. Despite the wide number of proposed combinations, they are all designed in the same way: they are meant to optimize the choice of blood parameters and to maximize the algorithm to match histological stages as assessed using liver biopsy. This is a fundamental difference compared to Fibroscan. While Fibroscan assesses one genuine characteristic of liver tissue, serum marker algorithm is built to mimic biopsy irrespective of the biopsy accuracy. In that case, the findings presented by Mehta et al. will hold only if biopsy and serum marker misclassifications are not correlated at any given stage of fibrosis – a challenging hypothesis. Otherwise, since biopsy was used for choosing the optimal combination of serum markers, a perfect serum marker could theoretically reach an AUROC of 1.0 and a lower AUROC value is related to serum marker own limitations rather to limitation of biopsy for assessing fibrosis.

One major limitation of any of these surrogates lies in their conception and/or their validation using a dichotomized approach (significant versus non-significant fibrosis). In addition to the question of what is considered to be “significant” fibrosis, a definition which is variable according to the study and aims pursued, staging fibrosis cannot be summed up by such a binary approach. Histological staging systems comprise 5 (METAVIR) or even 7 (Ishak score) different stages [7,12]. This level of complexity has been shown to be relevant not only for individual assessment and follow-up of disease evolution, but also for defining the rate of fibrosis progression and the right moment for using antiviral therapy or starting prevention of complications from cirrhosis. The dichotomized approach used for surrogates is imposed by the use of AUROC that tests a binary hypothesis. Using this approach there is a significant loss of information and a dependency on the proportion of each stage of fibrosis in the study sample. Other accuracy measures designed for ordinal gold standard have recently been published and should overcome these limitations [13]. However in most works these limitations have been bypassed by considering the different histological stages as linear variables and extrapolating intermediate values for each of the stages. However, this is an erroneous supposition since scores are categorical not continuous variables. When considering the extent of fibrosis, a variable that can be easily quantified by image analysis, studies have shown the absence of linearity between extent of fibrosis and histological stage [8,14]. Such an approximation explains why, when considering only adjacent stages (F1vsF2 or F2vsF3...) AUROC values are unacceptably low, prompting us to consider the surrogate as an inadequate tool for individual follow-up [15].

There is an urgent need to pursue the development of a surrogate for staging fibrosis. Because of the conditional relationship with biopsy, the serum marker might represent a dead-end. Hopefully, physical imaging will eventually be refined to an acceptable level of accuracy, especially for evaluation of early stages of fibrosis. Indeed, promising results have recently been shown using elastography with MRI.

Although much effort has been made in evaluation of fibrosis as a major decision criterion for hepatolo-
gists, it is only one among the many elementary histopathologic features present at the same time on liver biopsy performed for hepatitis C. Fibrosis is not an autonomous feature, but rather a tissue lesion resulting from other pathologic mechanisms such as inflammatory, degenerative or dystrophic processes leading to other pathologic mechanisms such as hepatocellular carcinoma and portal hypertension. In order to provide relevant information, fibrosis should be viewed in light of its full histopathologic context. Simultaneous evaluation of necroinflammation allows to assess whether fibrosis is the result of a past event that has stabilized or even regressed or is an ongoing process that may continue to worsen. Frequently, biopsy also detects associated lesions such as steatosis or steatohepatitis which provide information useful for management and prognosis of patients with chronic hepatitis C [16]. Finally, it is noteworthy that, in diseases with a high prevalence, like hepatitis C, liver biopsy may also reveal that abnormal liver function tests are related to an unexpected liver disease in addition to hepatitis C. Clearly, all this information may influence patient management. Therefore, equating chronic liver disease with the extent of fibrosis alone is an oversimplification that could be useful for physicians but it could also prove misleading.

After more than 10 years of active investigations, alternatives to liver biopsy for staging chronic liver diseases have revealed both their strength and weakness. As emphasized by Mehta et al. “Novel strategies are needed to move the field forward”. This implies not only long-term prospective studies using clinical end-points to validate surrogate markers that might be difficult to perform especially when addressing validation of markers for the diagnosis of early stages of fibrosis but also development of new innovative tools. Whether these tools will reach a satisfactory level of accuracy prior to the discovery of highly efficient and innocuous antiviral treatments remains an open question.

To date, liver biopsy remains the gold/best standard for accurate staging and grading in chronic hepatitis C and the major question that remains concerns the moment at which such an accurate evaluation is needed in chronic hepatitis C [17].

References

[3] Poynard T, Ratziu V, Benhamou Y, Opolon P, Cacoub P, Bedossa P. Natural history of its full histopathologic context. Simultaneous evaluation of necroinflammation allows to assess whether fibrosis is the result of a past event that has stabilized or even regressed or is an ongoing process that may continue to worsen. Frequently, biopsy also detects associated lesions such as steatosis or steatohepatitis which provide information useful for management and prognosis of patients with chronic hepatitis C [16]. Finally, it is noteworthy that, in diseases with a high prevalence, like hepatitis C, liver biopsy may also reveal that abnormal liver function tests are related to an unexpected liver disease in addition to hepatitis C. Clearly, all this information may influence patient management. Therefore, equating chronic liver disease with the extent of fibrosis alone is an oversimplification that could be useful for physicians but it could also prove misleading.

After more than 10 years of active investigations, alternatives to liver biopsy for staging chronic liver diseases have revealed both their strength and weakness. As emphasized by Mehta et al. “Novel strategies are needed to move the field forward”. This implies not only long-term prospective studies using clinical end-points to validate surrogate markers that might be difficult to perform especially when addressing validation of markers for the diagnosis of early stages of fibrosis but also development of new innovative tools. Whether these tools will reach a satisfactory level of accuracy prior to the discovery of highly efficient and innocuous antiviral treatments remains an open question.

To date, liver biopsy remains the gold/best standard for accurate staging and grading in chronic hepatitis C and the major question that remains concerns the moment at which such an accurate evaluation is needed in chronic hepatitis C [17].

References

[3] Poynard T, Ratziu V, Benhamou Y, Opolon P, Cacoub P, Bedossa P. Natural history of its full histopathologic context. Simultaneous evaluation of necroinflammation allows to assess whether fibrosis is the result of a past event that has stabilized or even regressed or is an ongoing process that may continue to worsen. Frequently, biopsy also detects associated lesions such as steatosis or steatohepatitis which provide information useful for management and prognosis of patients with chronic hepatitis C [16]. Finally, it is noteworthy that, in diseases with a high prevalence, like hepatitis C, liver biopsy may also reveal that abnormal liver function tests are related to an unexpected liver disease in addition to hepatitis C. Clearly, all this information may influence patient management. Therefore, equating chronic liver disease with the extent of fibrosis alone is an oversimplification that could be useful for physicians but it could also prove misleading.

After more than 10 years of active investigations, alternatives to liver biopsy for staging chronic liver diseases have revealed both their strength and weakness. As emphasized by Mehta et al. “Novel strategies are needed to move the field forward”. This implies not only long-term prospective studies using clinical end-points to validate surrogate markers that might be difficult to perform especially when addressing validation of markers for the diagnosis of early stages of fibrosis but also development of new innovative tools. Whether these tools will reach a satisfactory level of accuracy prior to the discovery of highly efficient and innocuous antiviral treatments remains an open question.

To date, liver biopsy remains the gold/best standard for accurate staging and grading in chronic hepatitis C and the major question that remains concerns the moment at which such an accurate evaluation is needed in chronic hepatitis C [17].

References