An appraisal of the histopathological assessment of liver fibrosis

R A Standish, E Cholongitas, A Dhillon, A K Burroughs and A P Dhillon

Updated information and services can be found at:

http://gut.bmj.com/cgi/content/full/55/4/569

These include:

References

This article cites 65 articles, 5 of which can be accessed free at:
http://gut.bmj.com/cgi/content/full/55/4/569#BIBL

1 online articles that cite this article can be accessed at:
http://gut.bmj.com/cgi/content/full/55/4/569#otherarticles

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Notes

To order reprints of this article go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to *Gut* go to:
http://journals.bmj.com/subscriptions/

AN APPRAISAL OF THE HISTOPATHOLOGICAL ASSESSMENT OF LIVER FIBROSIS

R A Standish, E Cholongitas, A Dhillon, A K Burroughs, A P Dhillon

SUMMARY

One of the most important aspects of the histopathological assessment of liver biopsies in the setting of chronic liver disease is determination of the degree of fibrosis and architectural change. Most of the work in this regard has been concerned with chronic viral hepatitis. This article attempts to assess critically our current and historical biopsy practice, from subjective fibrosis scoring systems to biopsy sample size; and the appropriate use of the data that scoring systems generate in the research and clinical setting. An understanding of the limitations of each of the components of the fibrosis assessment process can help to devise appropriate protocols to ensure that the information obtained is optimised, and its degree of reliability appreciated. It is only from this starting point that recently promulgated antifibrotic medications and ''non-invasive'' liver fibrosis assessment techniques can be evaluated properly.

INTRODUCTION

The degree of liver fibrosis is one of the most important diagnostic and prognostic assessments in chronic liver disease. Histological assessment of fibrosis is regarded as the ''gold standard'' in this respect. Clinical manifestations of liver disease and liver dysfunction accompany architectural changes of the liver parenchyma that are a result of advanced stages of liver fibrosis. Previously, it was thought that liver fibrosis and end stage liver disease (cirrhosis) were irreversible, and therefore crude determinations of liver fibrosis were acceptable because the therapeutic impact of this assessment was relatively minor. Recent work suggests that liver fibrosis may be modified by treatment[1–4] and so critical re-evaluation of the histopathological methods used to assess liver fibrosis is necessary. Liver biopsy assessment of fibrosis is not only the end point for the development of antifibrotic treatments but also forms the benchmark for validation of serum ''surrogate markers'' of liver fibrosis and other non-invasive techniques that purport to measure liver fibrosis and stage of liver disease.

There are several important issues which make interpretation of the current literature difficult.

► Firstly, the bulk of the literature published recently regarding the adequacy of liver biopsies is based on studies of chronic viral hepatitis. This literature suggests that the majority of liver biopsy samples are too small to be representative and therefore are inadequate for a reliable assessment of histological stage or grade.

► Secondly, that the histopathological scoring of the stage of liver disease is largely an assessment of architectural change and not a measurement of amount or degree of fibrosis.

► Thirdly, that stage scores are often mistaken for measurements, and inappropriate statistical techniques are then applied.

► Lastly, although recognised but often forgotten, that the scores generated are prone to considerable intraobserver and interobserver error.

Thus histopathological liver fibrosis assessment has become a tarnished ''gold standard'' and it needs to be improved. Progress will depend on recognition of the problem, a better understanding of the need for an adequate sample, and a much better understanding of the nature of ''staging scores''—including the use of more appropriate statistical tools and strategies to minimise observer variation. For proper quantification of liver collagen, image analysis (IA) of appropriately stained liver sections or biochemical assay might be necessary.

SCORING SYSTEMS AND VARIABILITY

Before going on to address the important issue of biopsy size and adequacy, we would like first to consider analysis of the biopsy. Currently, most centres carry out descriptive reporting when assessing diagnostic biopsies from an individual patient. However, in the trial and research setting, evaluation of biopsies is carried out using subjective scoring systems that produce shorthand values for various categories of inflammation (grade), and fibrosis and architectural

See end of article for authors' affiliations

Correspondence to:
Professor A P Dhillon, Academic Department of Histopathology, Royal Free Campus, Royal Free and University College Medical School, Rowland Hill St, London NW3 2PF, UK; a.dhillon@medsch.ucl.ac.uk
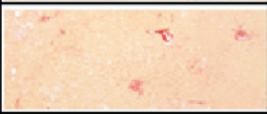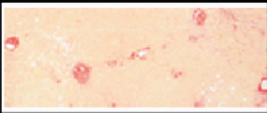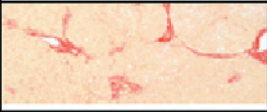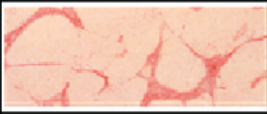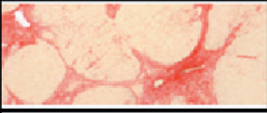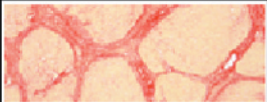
Appearance	Ishak stage: Categorical description	Ishak stage: Categorical assignment	Fibrosis measurement*
	No fibrosis (normal)	0	1.9%
	Fibrous expansion of some portal areas ± short fibrous septa	1	3.0%
	Fibrous expansion of most portal areas ± short fibrous septa	2	3.6%
	Fibrous expansion of most portal areas with occasional portal to portal (P-P) bridging	3	6.5%
	Fibrous expansion of portal areas with marked bridging (portal to portal (P-P) as well as portal to central (P-C))	4	13.7%
	Marked bridging (P-P and/or P-C), with occasional nodules (incomplete cirrhosis)	5	24.3%
	Cirrhosis, probable or definite	6	27.8%

Figure 1 Stage component of the Ishak system.[6] *Proportion (%) of area of illustrated section showing Sirius red staining for collagen (collagen proportionate area).

disruption (stage). There are many such systems[2 5 6] but essentially they "score" (that is, categorise) similar features.

The history of these scoring systems dates back to 1981 when the histological features of chronic hepatitis were evaluated for potential importance in determining the prognosis according to the understanding at the time of the pathophysiology of chronic hepatitis B virus (HBV) infection, and organised into a scoring system by Knodell and colleagues.[5] Evaluations were based on review of 14 biopsies from five patients: one with chronic HBV and four with presumed chronic non-A-non-B infection—at that time, hepatitis C virus (HCV) had not been identified. In the Knodell "histological activity index" system (HAI), each of four histopathological features (periportal ± bridging necrosis, intralobular degeneration/focal necrosis, portal inflammation, and fibrosis) are assessed separately and assigned a score. In order to exaggerate the difference between mild and serious disease, Knodell and colleagues[5] eliminated the number 2 from their scoring system for each of these features. Some consider this a drawback of this system, as it makes the scores discontinuous, but by omitting the 2, the Knodell system conveniently and simply allows the four key histopathological aspects of chronic liver disease (noted above) to be assigned to categories of: normal or minor change and two degrees of major change for each of the axes of assessment apart from periportal ± bridging necrosis (which is weighted further in the presence of bridging necrosis because of its perceived pathophysiological importance).
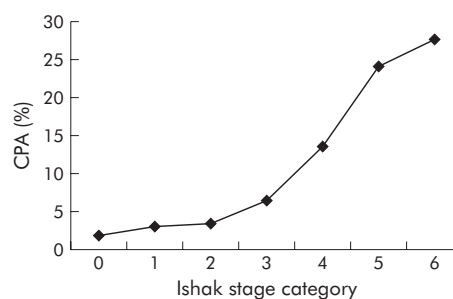

Figure 2 Relationship between Ishak stage scores and measured amount of fibrosis. Examples of the Ishak categories illustrated in fig 1 have been plotted against their individually measured collagen proportionate area (CPA). Clearly, the measured amount of fibrosis, and the corresponding Ishak category, are different evaluations. The increase in fibrosis with Ishak disease stage category is not a straight line.

The currently most widely used system is the Ishak, or "revised Knodell", system[6] which attempts to correct the criticism of numerical discontinuity by reintroducing the number 2. The criticism is itself erroneous when it is understood that the "numbers" associated with any histopathological scoring system represent a numerical shorthand for a descriptive categorical assignment, and are neither integers nor numerical measurements along a continuum in a mathematical sense (see below).

The first three axes of the Knodell HAI relate to the necroinflammatory grade of the disease. The fourth feature

("fibrosis": which incorporates architectural changes) assesses the stage of disease by evaluating the degree of fibrous portal tract expansion, fibrous portal-portal linking, portal-central fibrous bridges, and the formation of fibrous septa and parenchymal nodules.

The Ishak system[6] assesses fibrosis in seven categories, ranging from normal to cirrhosis (see fig 1) and so has potentially more discriminant descriptive power. All scoring systems basically use the same principles to record liver disease stage. It is obvious that the "stage" or "fibrosis" score is composed of a mixture of features, none of which specifically depends on the amount of fibrous tissue in a liver biopsy sample. The higher Ishak scores particularly depend more on architectural changes and degree of nodularity rather than amount of fibrous tissue. Even the lower scores (for example, "portal tract expansion") are only partly dependent on the amount of portal tract collagen because portal tracts can be considerably expanded by inflammatory infiltrates as well (fig 2).

Although in the original Knodell publication HAI was stated to be "numerical, objective, and reproducible",[5] it has, strictly speaking, never been any of these things. The considerable observer variation in the subjective interpretation of the categorical assignments has been often underestimated or ignored as a source of variability, despite the fact that both inter- and intraobserver variability have been well documented with respect to the various scoring systems.[7–9] Indeed, Knodell and colleagues recommended that a change of four points in the aggregate score was required before there was a high probability of a real or significant change, rather than merely representing a shift attributable to reinterpretation or mistaken interpretation of the biopsy features.[5] Not surprisingly, studies have shown that this variability is reduced if there are fewer categories to choose from within each axis of assessment, but this approach reduces the descriptive power of the more abbreviated scoring systems that have been proposed.

The METAVIR scoring system was designed for the assessment of HCV chronic hepatitis specifically and reproducibly. In its current formulation,[10] HCV hepatitis activity is based on evaluation of piecemeal necrosis and lobular necrosis. Portal inflammation is excluded from the METAVIR algorithm "because this feature is a prerequisite for the definition of chronic hepatitis even without activity." Nevertheless, the recent study by Rousselet and colleagues,[11] using the METAVIR system to examine 254 liver biopsies from patients with chronic viral hepatitis, found that: "the level of experience (specialisation, duration, and location of practice) had more influence on agreement than the characteristics of the specimen (length, fibrosis class number, miscellaneous factors). Agreement can be improved by experienced pathologist or consensus reading".[11] The experience of the histopathologist is an important variable that must be considered and optimised.

Procedures to minimise observer variation are required when any scoring system is used in studies of liver fibrosis: pre-trial consensus meetings should be held by the histopathologists involved who should have sufficient experience of the disease to be studied to debate and decide the precise fibrosis assignment categories and the borders between them. Then the entire study series should be assessed by at least two hepatopathologists independently, followed by consensus sessions to discuss significant disagreements; and the

assessments should be conducted within as short a period as is practical (that is, the assessment should not be of single biopsies weeks or months apart).[12] An additional component that can be included in study protocols to improve detection of subtle changes is the comparative review of each subject's paired biopsies, from the beginning and end of a trial, after the initial blinded histological scoring has been performed. This should be performed with limited unblinding only (that is, comparison of biopsy pairs without knowledge of intervention versus control, or time point of biopsy). However, the subtler the change that is detected in this way, the greater the risk of over-interpretation if the samples are unrepresentative.

In reality, few studies are structured in such a way. Routine scoring of liver biopsies is not appropriate for sporadic assessments in daily diagnostic practice because of the likelihood of inter- and intraobserver errors, and promotion of the spurious idea that something quantitative has been achieved. Routine scoring is done with increasing frequency under the impression that the UK National Institute for Health and Clinical Excellence (formerly known as the National Institute for Clinical Excellence) HCV treatment guidelines require this: the guidelines state that "Interferon alpha and ribavirin as combination therapy is recommended for the treatment of moderate to severe hepatitis C (defined as histological evidence of significant scarring (fibrosis) and/or significant necrotic inflammation), at standard doses for patients over the age of 18 years".[13] The original guidelines (Appraisal 14, 2000) suggested in a more detailed discussion (than is present in the 2003 revision) that "Liver biopsy is undertaken, if there are no increased risks, in order to assess liver scarring and necro-inflammation according to an accepted severity scale such as the Knodell". A later document (Appraisal 75, 2004) eliminates reference to the Knodell system. Wisely, there is neither reference to any particular scoring system nor is any score specified that mandates treatment in the 2004 guidelines. The decision of how much histological fibrosis, necrosis, and/or inflammation is "significant" should be a matter for discussion between the patient and their hepatologist, taking into account the potential risks and benefits of treatment versus no treatment in any individual case. Furthermore, recent studies suggest that in non-genotype 1 HCV, the benefit/risk

Strategies to minimise interobserver and intraobserver scoring variability.

► At least two pathologists with appropriate experience should be involved in studies involving histopathological scoring of chronic liver disease stage.
► Pre-trial consensus meetings should define precisely the interpretation of the categorical definitions of the scoring system to be used, and clarify the boundaries between the categories.
► Independent blinded assessment of the complete study series should be performed by each of the pathologists.
► Further consensus meetings to discuss scoring discrepancies should take place.
► Assessments should be carried out in as short a time as possible.
► Partially unblinded (without knowledge of time sequence or treatment group) qualitative comparison of biopsies from individual subjects after the initial scoring may be informative.
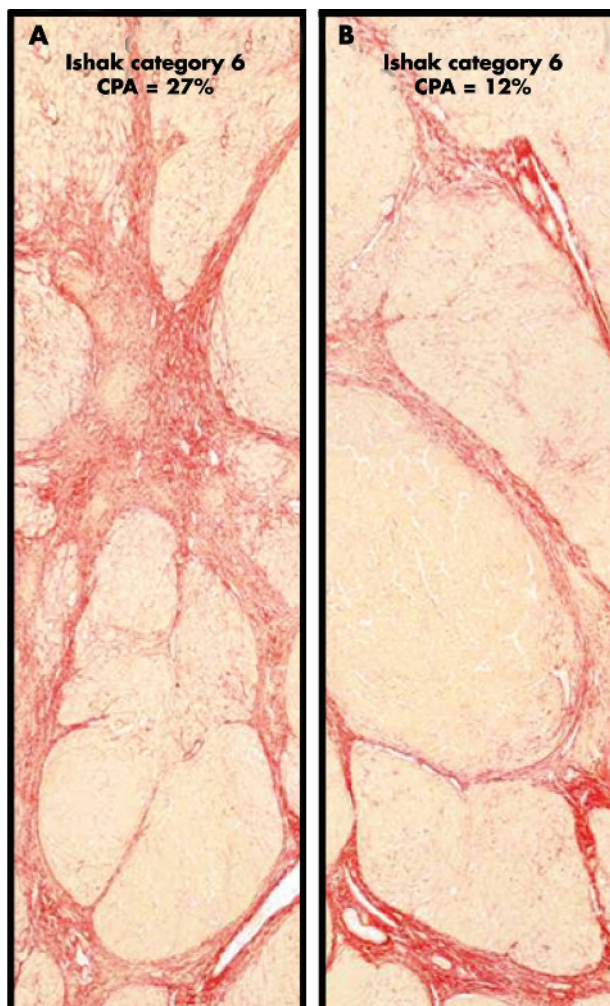
Figure 3 Ishak stage scores are neither numbers nor measurements. Two liver biopsies, each of which is stained with Sirius red for collagen. Both biopsies show parenchymal nodules surrounded by fibrous tissue, fulfilling the histological diagnostic criteria of cirrhosis. Therefore, each biopsy can also be described as ''Ishak stage 6'' (which is merely a symbol for the histological definition of cirrhosis). The overall area of one of the biopsies consists of 27% collagen, and 12% of the other biopsy is collagenous (collagen proportionate area (CPA)). Obviously, the histopathological diagnosis of cirrhosis (and assignment of ''Ishak stage 6''), and the measured amount of fibrosis are entirely different assessments.

balance currently favours treatment in even histologically mild disease, so that there may no longer be the need for therapeutic decisions to be biopsy driven at all.[14]

HISTOPATHOLOGICAL SCORES AND NUMBERS

The central dictum of Pythagoras' followers was ''all is number''. Unfortunately, where histological scoring systems are concerned, all is not number (fig 3). Failure to appreciate this essential point has led to misinterpretation of the histological scores and application of inappropriate statistical techniques. Consequently, the conclusions of most of the relevant literature are questionable.

Histological scores are at best ordered categorical data and not numbers. Any statistical analysis must be carried out with this in mind. One would no more consider these scores to be treatable numerically than one would add, subtract, and divide different categories under the heading of "colour", or to

mix and match apples and oranges. Appropriate statistical treatment of categorical data includes the use of contingency tables to convert categorical data into frequency data, which can then be numerically manipulated.[8] This important aspect of the nature of scoring has eluded many investigators, and little of the available literature on the subject has converted categorical data into frequency data.

Clinicians, including those who work with drug regulatory authorities (in contrast with hepatopathologists), are less familiar perhaps with the nature of the scores and the underlying definitions which lead to their assignment. For example, the French METAVIR group analysed scores of stage in biopsies from patients with chronic hepatitis C virus infection to produce an ''annual fibrosis progression rate''.[15] Not only does this assume a linear progression of fibrosis in HCV infection (which has not been demonstrated), but the work produced annual fibrosis progression rates with bewildering score fractions. One accepts that any histopathological fibrosis scoring scale consists of rather arbitrary points on a continuum between normality and cirrhosis, but when those points have been strictly defined at the start of a study, because the in-between points lack any histological descriptive definition, the sense of the subsequently invented fractions also lacks meaning. Despite this, the ''annual fibrosis progression rate'' has become popular in some similar studies, and this ''measure'' is in danger of becoming as entrenched as the aggregated HAI score.

The same issues are pertinent when considering other histopathological scoring systems. A scoring system suitable for a particular disease process is unlikely to be suitable for a different process. For example, a scheme designed for the assessment of steatohepatitis such as that devised by Brunt and colleagues[16] would be inappropriate for the assessment of chronic viral hepatitis, for which we should use scoring systems such as the Knodell or Ishak, because fibrosis develops and progresses in different ways in these diseases. Such misunderstandings can be confusing until one realises that all scoring systems are based on our understanding of the pathophysiology of the disease in question and the histopathologist is only trying to describe what is seen. The

---

Histopathological scoring of chronic liver disease stage guidelines.

► Scoring is not recommended for routine daily diagnostic practice.
► Diagnostic evaluation of longitudinal changes of disease stage in individual patients should be performed by direct comparison and review of all biopsies, and not by comparison of recorded scores.
► Stage scoring is not a measurement, but a categorical assignment (based largely on architectural changes), and any statistical analysis should take this into account.
► Numerical manipulation of scores (for example, calculation of average and aggregate numbers) is statistically invalid, and may lead to erroneous conclusions.
► There remains an important role for stage scoring in the research setting, perhaps alongside other methods of fibrosis assessment.
► The scoring system used should be appropriate to the disease process that is being studied according to an understanding of the pathophysiology of the condition.
► Strategies should be used to minimise interobserver and intraobserver scoring variability.

misuse of histopathological scoring systems confounds efforts to understand the natural history of disease which should be based on long term longitudinal follow up and outcome. The danger of misleading others with inappropriate manipulations of histopathological scores is clear and present, and the ease with which one can fool colleagues, and oneself, is alarming.

BIOPSY SIZE

Several years ago it was stated rather arbitrarily, and probably dictated by clinical pragmatism and a compromise with reality, that "a liver biopsy containing six portal tracts satisfies most hepatopathologists".[17] Now it is timely to move forward from the laudable desire to achieve histopathological happiness to a more scientific approach with regard to liver biopsy adequacy. In the case of liver fibrosis assessment, to determine the size of an adequate (representative) liver biopsy we must know the variability of distribution of liver fibrosis within the liver for the particular disease under consideration, which is likely to change with the stage of disease. An assessment when there is a less than adequate biopsy sample needs to be considered in the light of the degree to which that analysis is reliable or unreliable (that is, the confidence interval).

Sampling error usually results in underestimation of the feature being assessed but, for example, inclusion of capsular/subcapsular tissue overestimates both histological activity and fibrosis, and inclusion of the connective tissue normally seen in large (septal) portal tracts will overestimate chronic viral hepatitis related fibrosis (see fig 4). Colloredo and colleagues[18] showed that reduction of the size of the biopsy available for histological review (both the length and diameter of the needle core) influences histological interpretation, with shorter cores more likely to be scored with lower grade and stage—interpreted as "underestimation" of overall disease activity. Following this study they recommended a biopsy length of 20 mm or greater—with more than 11 complete portal tracts—in order to reliably grade and stage a biopsy for chronic viral hepatitis.[18] When these liver biopsy characteristics are compared with the average biopsy (even in specialised centres), then efforts to improve sample adequacy are clearly necessary.

Biopsy size is also one of the variables that impact on the reliability of scoring (both grade and stage). Two similar studies of pairs of synchronous liver biopsies from patients with HCV infection have been performed in efforts to assess the reliability of disease stage and grade assessments in chronic HCV infection. In the first study, by Siddique and

Adequacy of liver biopsy

- ▶ Adequate samples must be obtained which are representative of the disease process that is being evaluated.
- ▶ The current recommendation for chronic viral hepatitis is that the biopsy should be at least 20 mm length (1.4 mm diameter) and should contain at least 11 complete portal tracts.
- ▶ Smaller biopsies are likely to be unrepresentative.
- ▶ Small inadequate biopsies tend to underestimate both disease grade and stage.
- ▶ Variability increases with smaller samples.
- ▶ Recognition of the limitations of interpretation of small biopsies is essential in the management of patients and in the therapeutic trial setting.

colleagues,[19] pairs of right lobe biopsies from 29 patients were evaluated. Each biopsy contained 4–5 portal tracts (around half the currently recommended number). Using the Knodell HAI, Siddique et al showed that in approximately 20% of cases the stage score differed between the synchronous biopsy pairs by two categories or more.[19] Persico et al compared Ishak stage scores of synchronous left and right lobe biopsies.[20] The left and right lobe biopsies had mean lengths of 2.5 ($\pm$0.9) cm and 2.8 ($\pm$1.1) cm, respectively. There was minimal variation of stage score between the paired biopsies.[20] These two studies together show the need for adequate biopsies to minimise variation in stage assessment due to sampling error. The studies demonstrate that inadequate samples give unreliable results and adequate samples give reproducible results.

Bedossa and colleagues[21] studied large liver tissue sections of resections from patients with chronic HCV infection (including both liver resections and livers removed at the time of transplantation), which were given a METAVIR stage score. Internal variation within each tissue section was not directly addressed in this part of the study, and a single stage score was ascribed to each large section. Picro Sirius Red F3BA (Sirius red) staining and IA—measuring the area of Sirius red staining per unit area of liver tissue (see discussion of IA below)—were used to determine the "virtual" biopsy size that adequately reflected the overall stage of liver disease. "Virtual" biopsies of 25 mm length (1 mm diameter, which approximately corresponds to a needle of internal diameter 1.2 mm) "correctly" staged 75% of cases. Considering that the same data set was used both to construct the IA range for each stage score and to test it, this is perhaps a little disappointing. This could be because this attempt to use IA to validate the stage score ignores the disparate nature of the techniques—scoring being to a large extent a subjective architectural assessment and IA being a measurement of fibrosis. Variation attributable to sample size decreased progressively up to a (1 mm diameter) biopsy length of 40 mm but intrinsic variability of liver fibrosis persisted in virtual biopsies of greater than 40 mm.[21]

Given these recent revelations with regard to liver biopsy adequacy, there has been concern regarding the feasibility of safely obtaining adequate samples for reliable grading and staging of chronic liver disease. Sample size is probably related to operator experience and efforts should be made to improve operator performance. Percutaneous (PLB) and transjugular (TJLB) liver biopsies are the two main techniques used to obtain liver specimens. Reviewing the literature, 11 studies were identified that have information about both the length and number of complete portal tracts (CP) in PLB [19 22–31]. In order to adjust for the different cohort sizes, mean length and number of CP were summated relative to the number of biopsies in each study (n, that is, overall mean length $= \Sigma$ [n $\times$ mean length]/Σn). The mean number of CP in all of the studies was evaluated similarly (that is, overall mean CP $= \Sigma$ [n $\times$ mean CP]/Σn). The raw CP data were not available from the studies to enable calculation of an overall median, which would have been more appropriate for assessment of complete portal tracts (table 1). Overall mean length was 13.5 mm, and overall mean number of CP was 6.58. Some studies did not specify the type of needle used. Where specified, Menghini-type PLB were significantly longer compared with Tru-cut type PLB (16.2 v 12.1 mm), but with almost the same mean number of CP (6.09 v 6.35).

Table 1 Systematic review of studies reporting the quality of percutaneous liver biopsies in which both biopsy length and number of complete portal tracts were reported

No of studies	No of biopsies	Total mean length (mm)	Menghini mean length (mm)	Trucut mean length (mm)	Total mean CP	Menghini mean CP	Trucut mean CP
11[19 22–31]	1675	13.5	16.2	12.1	6.58	6.09	6.35

CP, complete portal tracts.

Insufficient information was included in these papers regarding the number of needle passes, which is a significant omission, as it has been shown that the complication rate correlates with the number of passes.[32] Fewer studies had information for both the length and number of CP obtained with TJLB using Tru-cut and/or Menghini-type needles.[25 33–35] Based on the data given by these four studies and following the same methodology as above, these show that with an average of 2.9 passes, the total mean length and CP were 20.9 mm and 8.09, respectively (table 2).

The implication of these findings is that, in well documented series of PLB and TJLB, the average length and mean number of portal tracts for over half the patients are well below the current recommendations for reliable grading and staging of biopsies taken for the assessment of chronic viral hepatitis.[36] Thus in order to obtain an adequate sample, a standard percutaneous approach will require two passes on average, increasing the risk of complications.[32] Via a transjugular route, several passes can be made without increasing complications. In a series of 326 TJLB from our centre, using three passes as standard, a median length of 22 mm and a median number of 8 CP was achieved without complications.[35] It seems reasonable to suppose that an adequate number of CP ($\geqslant$11) could be achieved safely by performing four or more passes.

MEASUREMENT

As the clinical importance of liver fibrosis assessment grows, the limitations of biopsy scoring systems are becoming more apparent. The changes described in the scoring categories are largely architectural, with little reference to the actual amount of collagen (fibrosis) in the liver sample. Routine (diagnostic) fibrosis assessment is usually carried out on a trichrome or reticulin stain. These are not specific collagen stains, and no measurement is made. If, after due consideration, it is felt to be desirable to determine precisely how much liver fibrosis there is (if liver collagen is what is meant), to validate studies of surrogate markers of fibrosis for example, or changes with antifibrotic treatment, then proper measurement of liver collagen is unavoidable. This will require time, effort, adequate sampling, and appropriate methodology. We cannot continue to pretend that subjective scoring has measured something when it has not.

Table 2 Systematic review of studies reporting the quality of transjugular liver biopsies in which both biopsy length and number of complete portal tracts were reported

No of studies	No of biopsies	Total mean length (mm)	Total mean CP
4[25 33–35]	409	20.9	8.09

CP, complete portal tracts.

Lord Kelvin said "In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of *Science*, whatever the matter may be".[37]

His reference to suitable "principles of numerical reckoning" alludes to the use of appropriate statistical analysis, which we have already touched on. As the value of routine histopathology demonstrates, clinically relevant data do not have to be numerical to be useful, but to pretend that they are numerical (when they are not) is wrong. It may be informative to measure what is seen, but to imagine it has been measured without doing so is self-delusion.

This brings us to the finding of a "practicable method for measuring" liver fibrosis. It is possible to use biochemical methods to measure the amount of collagen in tissue homogenates, but this requires the destruction of the tissue, so no further information can be gathered, which is unsatisfactory in the assessment of a diagnostic liver biopsy (in which context most human study material is gathered). In a recent study, Lee *et al* took multiple biopsies from the same livers (each of 12 livers sampled after transplantation, resection, or autopsy), and showed that hydroxyproline measurements of the biopsies from each liver had a coefficient of variance between 13% and 36% (mean 23%), and that this variation was present at all stages of liver disease.[38] In this study, 16 gauge Tru-cut needle biopsies with biopsy lengths of 10–15 mm were used and "obviously identifiable large vessels on the biopsy samples were excised and excluded from hydroxyproline measurements". Because of the nature of the biochemical assay, we cannot know the number of complete portal tracts in the samples but this study illustrates that even biochemical approaches to the quantification of liver biopsy fibrosis will be subject to appreciable sampling variability, and dependent on sample size.

Computer assisted IA of histochemically stained sections is a method for measuring fibrosis morphologically, it does not hinder the other necessary diagnostic evaluations, and collagenous structures irrelevant to the disease process (and which contribute to the variability between samples) can be excluded precisely (fig 4). Broadly speaking, IA uses segmentation of digital images to measure the area of collagen and the area of tissue, producing a "fibrosis ratio" or collagen proportionate area (CPA).

To quantify hepatic collagen histologically, it is necessary first to stain it specifically. Sirius red has an affinity for most

hepatic collagens, including types I and III (the major components of liver collagen)[39] [40] and this binding correlates with chemical hydroxyproline assay under standardised laboratory conditions.[41] Although the nature of the binding of the dye is not fully understood, staining of liver sections appears to be reliable and reproducible. Thus Sirius red staining is the preferred histochemical method when quantifying liver fibrosis, even though the staining may not be stoichiometric.[42] Some groups interested in IA have used reticulin or trichrome stains, and others have suggested the use of immunohistochemistry, but due to variability in staining and difficulty in accurate thresholding, these techniques can suffer from poor reproducibility.

Normal human liver is estimated to contain approximately 5.5 mg/g of collagen, and cirrhotic liver contains of the order of 30 mg/g.[39] Other estimates of normal liver collagen concentration suggest 2–8 mg/g wet weight of liver.[43] We must also consider that different disease processes probably lead to different amounts of collagen deposition at different stages of disease. Even the same disease process at the same stage is likely to produce different amounts of collagen in different individual patients. This aspect of collagen deposition has not yet been fully evaluated.[44]
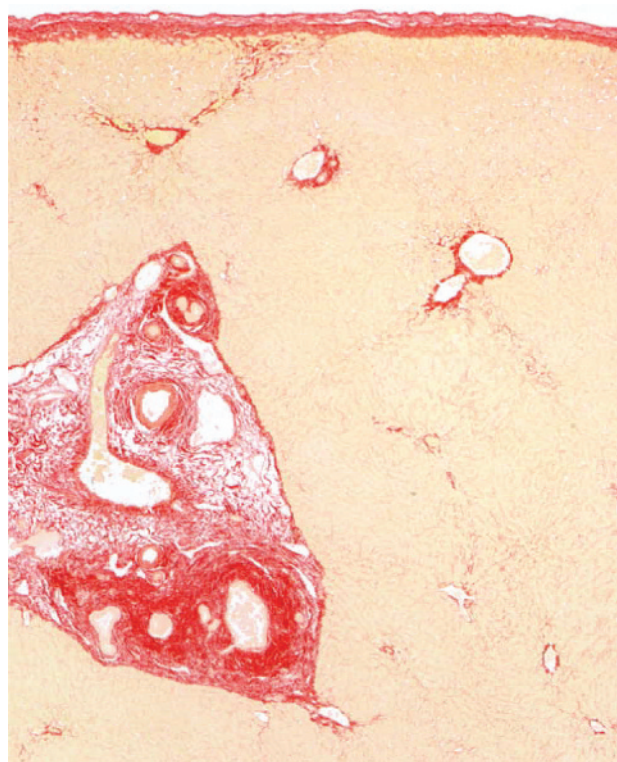

Figure 4 Normal liver contains more collagen in normal structures than in the small portal tracts which are usually affected by chronic viral hepatitis. This is a section of normal liver stained with Sirius red for collagen. The section includes normal liver capsule (top) and a large portal tract (middle left). In studies of chronic viral hepatitis, these areas should be excluded from measurements of liver fibrosis because chronic viral hepatitis affects small portal tracts primarily. Otherwise, subtle changes due to disease or treatment will be obscured. Uncritical inclusion of structural collagen components confounds attempts to measure disease related collagen accurately, so that editing is necessary of structures (for example, large portal tracts, capsular collagen), and technical staining artefacts that are irrelevant to the pathological process being studied.
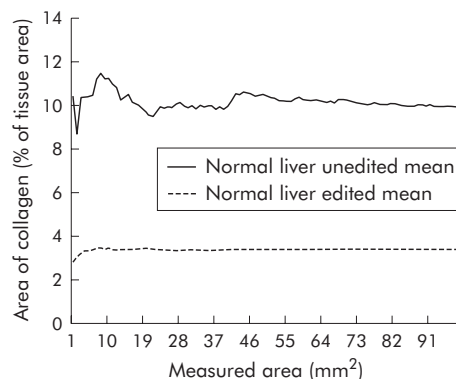
Figure 5 Cumulative mean of collagen proportionate area of normal liver stained with Sirius red (unedited versus editing of structural collagen). A smaller sample is adequate and representative if collagen that is not relevant to the disease process being studied is removed from the measurement. A traditional morphometric method of determining representative sample size is by measuring the feature of interest (in this case, liver fibrosis) in consecutive microscopic fields, and calculating the ''cumulative mean''. When a stable cumulative mean is achieved, the overall area measured can be regarded as an adequate (representative) sample. Editing (exclusion) of large portal tracts (>0.5 mm diameter), capsule, and technical artefacts in this experiment (measurement of collagen proportionate area) achieves a stable cumulative mean with a much smaller sample size. Without editing, fluctuations of the cumulative mean persist at sample sizes equivalent to quite large liver biopsies. The graph also illustrates the dominant proportion of collagen that resides in normal structures in normal liver, and the relatively small amount of fibrous tissue that is normally present in small (<0.5 mm diameter) portal tracts.

IA studies have looked at fibrosis in many conditions, producing overall similar results ranging from 1–4% fibrous tissue in normal liver to 15–35% fibrous tissue in cirrhosis.[21] [45–50] However, the differences in the technology and methodology used in these studies means that the results are not directly comparable.

Many of these studies have also attempted to correlate the image analysis measurements of fibrosis with the categorical stage scores. Manabe et al found that CPA did not correlate well with the components of the Knodell scoring system.[50] Pilette et al used the Knodell and a modified METAVIR scoring system and found that these correlated with liver biopsy percentage area of fibrosis.[48] Kage et al found that CPA correlated with their modified staging system based on the Desmet and Scheuer staging systems.[46] This collection of inconsistent results can be explained by the notion that IA and histopathological stage scores assess quite different things, and in different ways. In assigning histological stage score categories, much depends on architectural changes. It may be in some (or many) cases that the amount of fibrous tissue increases with these architectural derangements, but that is not specifically addressed in any of the scoring systems. No heed is paid to the width of a fibrous septum or the size of a nodule, for example, so two cirrhotic livers scoring 6 on Ishak staging might contain vastly different amounts of collagen, and thus have different IA measurements (fig 3). Despite this argument, O'Brien et al showed that IA measurements correlated well with Ishak stage scores, but only in the higher stages.[45] Results such as these may be chance occurrences, or only applicable to the limited framework of the individual study.

576

We are in the early days of elucidating the best way of using IA reliably, and the clinical utility or otherwise of IA has yet to be established. At present, a range of computer and camera hardware, software, and experimental procedures (and liver sample sizes) are being used, and so it is not surprising that some of the results are contradictory and confusing. When some of the methodological aspects have been resolved, IA may find a useful place in the armamentarium available to assess liver biopsy fibrosis.[51][52] Just as in biopsy scoring, there are elements of study design that can be put in place to minimise variability (see above), so with IA it is important to pay attention to various aspects of the methodology (including section thickness and staining, light source stability, camera characteristics, and accurate thresholding[53] of the image) to ensure stability and reproducibility of measurements. In addition, just as a mental note is made of capsular collagen, large septal portal tracts, and blood vessels when assessing fibrosis by scoring (so that these normal collagenous structures can be excluded from the disease stage scoring because they do not represent disease related collagen), so too in IA these structures need to be edited out of the measurement, particularly in the analysis of biopsy specimens, so as not to overshadow small disease or treatment related changes in the measured amount of fibrosis (figs 4, 5).

It is important to remember that IA in its simplest form is a measurement of area, and is unable to evaluate architectural changes such as nodularity, fibrous portal linking, and portal-central fibrous bridging, which are the histological features of architectural change included in stage scoring systems. Thus IA and stage scoring are different assessments (although they may in some circumstances be linked) and should be understood as complementary aspects of liver biopsy assessment (figs 2, 3).

NON-INVASIVE ASSESSMENT OF FIBROSIS

The procedure of obtaining a liver biopsy is not without its attendant risk of morbidity and even mortality. The risk of complications is of the order of 1%, and the risk of mortality has been put at between 0.1% and 0.01%. The advent of the transjugular approach may reduce these further, particularly in high risk patients, and may allow multiple samples to be taken without the increased risk that is seen with multiple percutaneous passes in standard liver biopsies.[54][55] With this small but significant danger in mind and taking into account intra/interobserver error as well as sample variability, non-invasive methods of liver fibrosis assessment have been developed. These include using various imaging techniques and serological markers—''surrogate'' markers of fibrosis.

Imaging techniques such as ultrasound, computed tomography, and magnetic resonance scanning as yet cannot detect small changes in fibrosis in an individual patient although some studies have shown good correlation with stage scoring. Newer techniques may be more sensitive but are still experimental. Sandrin *et al* have shown good correlation between a non-invasive ''liver stiffness'' measurement (based on transient elastography of the liver) with the METAVIR stage score, a correlation which was shown by Ziol *et al* to improve with larger biopsy specimens.[56–58]

Traditional serological markers of liver function—''liver function tests''—give little indication of the various underlying pathological processes, including fibrosis. More complex panels of markers have been evaluated to examine the possibility of non-invasive fibrosis assessment.[59][60] The panel of markers

developed by Rosenberg *et al* has limited application because it was only reliable for estimating the absence of fibrosis and not its extent.[61] Poynard *et al* have developed a patented algorithm (the ''Fibrotest'') which they have shown can predict significant fibrosis in chronic HCV infected patients according to the METAVIR stage score, with an ''area under the receiver operator curve'' of 0.73–0.87 and a negative predictive value (excluding significant fibrosis) of 91%.[60] However, we must consider what these calculations mean biologically, and how they have been validated. We have already mentioned that the use of stage scoring systems—particularly those with few categories, such as the METAVIR system—has well recognised limitations, and we must also consider the issue of sample adequacy. In the study by Rosenberg *et al*, a liver biopsy length of ⩾12 mm length or content of ⩾5 portal tracts (it is not specified if the portal tracts were complete or incomplete) was necessary for inclusion of the specimen in the study.[59] Generally, the performance of non-invasive tests of liver fibrosis has been evaluated using reference histological samples which were biopsies of suboptimal size by current standards (<20–25 mm length and/or containing <11 complete portal tracts) or the quality of which was not mentioned.[58][62–65] In a recent editorial, Thuluvath and Krok remind us that the validity of non-invasive serum marker assessments is yet to be established in longitudinal (as opposed to cross sectional) studies, and the results of already published studies need to be replicated in different laboratories before these assessments can play an accepted part in the clinical management of patients.[66]

The utility of each of the new ''non-invasive'' analytical approaches must be validated against long term patient follow up and precisely measured end points. If one of the end points is liver fibrosis, this must be measured specifically, reliably, reproducibly, accurately, and using adequate samples. Until this has been done, surrogate markers of liver fibrosis must be regarded as qualitative approximations, and not quantitative assessments.

CONCLUSION

With the emergence of potentially efficacious antiviral and antifibrotic agents, much depends on making sure that assessment of liver fibrosis is reliable. The present method of subjective assessment of liver fibrosis and architecture by a single histopathologist is reasonable in the daily diagnostic situation, but application of grading and staging scoring systems is inappropriate routinely. Adequate samples according to current standards for the disease in question must be obtained to avoid error, and if these are not available, the limitations of suboptimal samples must be recognised, acknowledged, and stated clearly.

Histopathological stage scoring is sufficient for many clinical trials, and is the correct approach for observational studies. However, better design, improved organisation, and more appropriate statistical evaluation than has generally been the case up to now is needed in studies involving liver biopsy analysis. The use of fibrosis stage scores as numerical data is an unacceptable statistical technique in studies using subjective histopathological categories as an end point. Reviewers, researchers, and readers require education with regard to this matter, not only with respect to the design and interpretation of current and future studies, but also in the interpretation of studies already published. Formal quantification of liver collagen may be necessary in some studies.

There are great dangers attendant upon a failure to recognise the pitfalls in our current liver fibrosis assessment

practice and if on recognition of such we do not make the required improvements. There are dangers from the point of view of future patients, who may be inappropriately treated or denied treatment that has been mistakenly discarded because of ignorance of the histopathological process. The ''golden standard'' of liver biopsy assessment of fibrosis has become tarnished with age, but with effort it can be polished to regain its rightful place as a more worthy gold standard.[67]

Summary

▶ Adequate representative liver biopsy samples are necessary, and the limitations of inadequate samples must be understood properly.

▶ Histological scoring of liver disease stage is not the same as measurement of liver fibrosis (with which it is often confused).

▶ Scoring variability can be minimised using careful methodology.

▶ Histological scoring of liver disease stage is a categorical assignment and not a numerical measurement, so the use of appropriate statistical techniques is essential.

▶ If measurement of liver collagen (fibrosis) is needed (for example, to detect the effects of antifibrotic therapy), then an appropriate technique such as image analysis of Sirius red stained sections, or biochemical assay, must be used (not histological stage scoring alone).

▶ Neither quantitative assessment of liver tissue collagen nor ''non-invasive'' tests give the same information as the histological stage score because, of these assessments, only histopathological examination is able to evaluate liver architecture.

▶ Quantitative assessment of liver tissue collagen and ''non-invasive'' tests must be validated against appropriate clinical outcomes (not histological stage scores alone), and any clinical utility they may or may not have will be determined thereby.

· · · · · · · · · · · · · · · · · ·

Authors' affiliations

R A Standish, A Dhillon, A P Dhillon, Academic Department of Histopathology, Royal Free and University College Medical School, London, UK

E Cholongitas, A K Burroughs, Academic Department of Liver Transplantation and Hepatobiliary Medicine, Royal Free and University College Medical School, London, UK

REFERENCES

1 **Suzuki Y**, Kumada H, Ikeda K, et al. Histological changes in liver biopsies after one year of lamivudine treatment in patients with chronic hepatitis B infection. J Hepatol 1999;**30**:743–8.

2 **Dufour JF**, DeLellis R, Kaplan MM. Regression of hepatic fibrosis in hepatitis C with long-term interferon treatment. Dig Dis Sci 1998;**43**:2573–6.

3 **Dufour JF**, DeLellis R, Kaplan MM. Reversibility of hepatic fibrosis in autoimmune hepatitis. Ann Intern Med 1997;**127**:981–5.

4 **Dienstag JL**, Goldin RD, Heathcote EJ, et al. Histological outcome during long-term lamivudine therapy. Gastroenterology 2003;**124**:105–17.

5 **Knodell RG**, Ishak KG, Black WC, et al. Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. Hepatology 1981;**1**:431–5.

6 **Ishak K**, Baptista A, Bianchi L, et al. Histological grading and staging of chronic hepatitis. J Hepatol 1995;**22**:696–9.

7 **The French METAVIR Cooperative Study Group**. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. Hepatology 1994;**20**:15–20.

8 **Westin J**, Lagging LM, Wejstal R, et al. Interobserver study of liver histopathology using the Ishak score in patients with chronic hepatitis C virus infection. Liver 1999;**19**:183–7.

9 **Goldin RD**, Goldin JG, Burt AD, et al. Intra-observer and inter-observer variation in the histopathological assessment of chronic viral hepatitis. J Hepatol 1996;**25**:649–54.

10 **Bedossa P**, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. Hepatology 1996;**24**:289–93.

11 **Rousselet MC**, Michalak S, Dupre F, et al. Sources of variability in histological scoring of chronic viral hepatitis. Hepatology 2005;**41**:257–64.

12 **Scheuer PJ**, Standish RA, Dhillon AP. Scoring of chronic hepatitis. Clin Liver Dis. 2002;6: 335–47, v–vi).

13 **National Institute for Clinical Excellence**. Summary of guidance issued to the NHS in England and Wales. Compilation 2003;**7**:187–8.

14 **Wright M**, Forton D, Main J, et al. Treatment of histologically mild hepatitis C virus infection with interferon and ribavirin: a multicentre randomized controlled trial. J Viral Hepat 2005;**12**:58–66.

15 **Poynard T**, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet 1997;**349**:825–32.

16 **Brunt EM**, Janney CG, Di Bisceglie AM, et al. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gastroenterol 1999;**94**:2467–74.

17 **Bravo AA**, Sheth SG, Chopra S. Liver biopsy. N Engl J Med 2001;**344**:495–500.

18 **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.

19 **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.

20 **Persico M**, Palmentieri B, Vecchione R, et al. Diagnosis of chronic liver disease: reproducibility and validation of liver biopsy. Am J Gastroenterol 2002;**97**:491–2.

21 **Bedossa P**, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. Hepatology 2003;**38**:1449–57.

22 **Farrell RJ**, Smiddy PF, Pilkington RM, et al. Guided versus blind liver biopsy for chronic hepatitis C: clinical benefits and costs. J Hepatol 1999;**30**:580–7.

23 **Hopper KD**, Abendroth CS, Sturtz KW, et al. Blinded comparison of biopsy needles and automated devices in vitro: 1. Biopsy of diffuse hepatic disease. AJR Am J Roentgenol 1993;**161**:1293–7.

24 **Rocken C**, Meier H, Klauck S, et al. Large-needle biopsy versus thin-needle biopsy in diagnostic pathology of liver diseases. Liver 2001;**21**:391–7.

25 **Chau TN**, Tong SW, Li TM, et al. Transjugular liver biopsy with an automated trucut-type needle: comparative study with percutaneous liver biopsy. Eur J Gastroenterol Hepatol 2002;**14**:19–24.

26 **Chevallier P**, Ruitort F, Denys A, et al. Influence of operator experience on performance of ultrasound-guided percutaneous liver biopsy. Eur Radiol 2004;**14**:2086–91.

27 **Crawford AR**, Lin XZ, Crawford JM. The normal adult human liver biopsy: a quantitative reference standard. Hepatology 1998;**28**:323–31.

28 **Gurakar A**, Balkan M, Fagiuoli S, et al. Experience with percutaneous liver biopsies using two different needles: a 14 G Tru cut needle or a 15 G needle gun. Hepatology 1993;**18**(suppl 1):P911.

29 **Spirchez Z**, Fayad E, Albu S, et al. Comparison of two needles, the surecut (18G) versus the biopty gun (18G) in percutaneous liver biopsy. Hepatology 1999;**30**(suppl 1):P1279.

30 **Regan J**, Mihalov M, Limjoko A, et al. Transjugular liver biopsy: evaluation and comparison with percutaneous biopsy. Hepatology 1997;**26**(suppl 1):P281.

31 **Judmaier G**, Prior C, Klimpfinger M, et al. Is percutaneous liver biopsy using the Trucut (Travenol) needle superior to Menghini puncture? Z Gastroenterol 1989;**27**:657–61.

32 **Tobkes AI**, Nord HJ. Liver biopsy: review of methodology and complications. Dig Dis 1995;**13**:267–74.

33 **DiMichele DM**, Mirani G, Wilfredo CP, et al. Transjugular liver biopsy is safe and diagnostic for patients with congenital bleeding disorders and hepatitis C infection. Haemophilia 2003;**9**:613–18.

34 **De Hoyos A**, Loredo ML, Martinez-Rios MA, et al. Transjugular liver biopsy in 52 patients with an automated Trucut-type needle. Dig Dis Sci 1999;**44**:177–80.

35 **Cholongitas E**, Quaglia A, Samonakis D, et al. Transjugular liver biopsy: is it effective for accurate histological interpretation? Gut 2005;**54**:P26.

36 **Guido M**, Rugge M. Liver fibrosis: natural history may be affected by the biopsy sample. Gut 2004;**53**:1878.

37 **Thomson W (Baron Kelvin)**. Popular lectures and addresses, vol 1. Constitution of matter. London: Macmillan and Co, 1889.

38 **Lee HS**, Shun CT, Chiou LL, et al. Hydroxyproline content of needle biopsies as an objective measure of liver fibrosis: Emphasis on sampling variability. J Gastroenterol Hepatol 2005;**20**:1109–14.

39 **Rojkind M**, Ponce-Noyola P. The extracellular matrix of the liver. Coll Relat Res 1982;**2**:151–75.

40 **Feldman G**. Critical analysis of the methods used to morphologically quantify hepatic fibrosis. J Hepatol 1995;**22**(suppl 2):49–52.

41 **James J**, Bosch KS, Aronson DC, et al. Sirius red histophotometry and spectrophotometry of sections in the assessment of the collagen content of liver tissue and its application in growing rat liver. Liver 1990;**10**:1–5.

42 **Puchtler H**, Meloan SN, Waldrop FS. Are picro-dye reactions for collagens quantitative? Chemical and histochemical considerations. Histochemistry 1988;**88**:243–56.

577

578

43 **Biagini G**, Ballardini G. Liver fibrosis and extracellular matrix. *J Hepatol* 1989;**8**:115–24.

44 **Poynard T**, Mathurin P, Lai CL, *et al.* A comparison of fibrosis progression in chronic liver diseases. *J Hepatol* 2003;**38**:257–65.

45 **O'Brien MJ**, Keating NM, Elderiny S, *et al.* An assessment of digital image analysis to measure fibrosis in liver biopsy specimens of patients with chronic hepatitis C. *Am J Clin Pathol* 2000;**114**:712–18.

46 **Kage M**, Shimamatu K, Nakashima E, *et al.* Long-term evolution of fibrosis from chronic hepatitis to cirrhosis in patients with hepatitis C: morphometric analysis of repeated biopsies. *Hepatology* 1997;**25**:1028–31.

47 **Duchatelle V**, Marcellin P, Giostra E, *et al.* Changes in liver fibrosis at the end of alpha interferon therapy and 6 to 18 months later in patients with chronic hepatitis C: quantitative assessment by a morphometric method. *J Hepatol* 1998;**29**:20–8.

48 **Pilette C**, Rousselet MC, Bedossa P, *et al.* Histopathological evaluation of liver fibrosis: quantitative image analysis vs semi-quantitative scores. Comparison with serum markers. *J Hepatol* 1998;**28**:439–46.

49 **Masseroli M**, Caballero T, O'Valle F, *et al.* Automatic quantification of liver fibrosis: design and validation of a new image analysis method: comparison with semi-quantitative indexes of fibrosis. *J Hepatol* 2000;**32**:453–64.

50 **Manabe N**, Chevallier M, Chossegros P, *et al.* Interferon-alpha 2b therapy reduces liver fibrosis in chronic non-A, non-B hepatitis: a quantitative histological evaluation. *Hepatology* 1993;**18**:1344–9.

51 **Lazzarini A**, Levine R, Ploutz-Snyder R, *et al.* Advances in digital quantification technique enhance discrimination between mild and advanced liver fibrosis in chronic hepatitis C. *Liver Int* 2005;**25**:1142–9.

52 **Hui AY**, Liew CT, Go MY, *et al.* Quantitative assessment of fibrosis in liver biopsies from patients with chronic hepatitis B. *Liver Int* 2004;**24**:611–18.

53 **Wright M**, Thursz M, Pullen R, *et al.* Quantitative versus morphological assessment of liver fibrosis: semi-quantitative scores are more robust than digital image fibrosis area estimation. *Liver Int* 2003;**23**:28–34.

54 **Papatheodoridis GV**, Patch D, Watkinson A, *et al.* Transjugular liver biopsy in the 1990s: a 2-year audit. *Aliment Pharmacol Ther* 1999;**13**:603–8.

55 **Corr P**, Beningfield SJ, Davey N. Transjugular liver biopsy: a review of 200 biopsies. *Clin Radiol* 1992;**45**:238–9.

56 **Sandrin L**, Fourquet B, Hasquenoph JM, *et al.* Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. *Ultrasound Med Biol* 2003;**29**:1705–13.

57 **Ziol M**, Handra-Luca A, Kettaneh A, *et al.* Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C. *Hepatology* 2005;**41**:48–54.

58 **Castera L**, Vergniol J, Foucher J, *et al.* Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005;**128**:343–50.

59 **Rosenberg WM**, Voelker M, Thiel R, *et al.* Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* 2004;**127**:1704–13.

60 **Poynard T**, Imbert-Bismut F, Munteanu M, *et al.* Overview of the diagnostic value of biochemical markers of liver fibrosis (FibroTest, HCV FibroSure) and necrosis (ActiTest) in patients with chronic hepatitis C. *Comp Hepatol* 2004;**3**:8.

61 **Bissell DM**. Assessing fibrosis without a liver biopsy: are we there yet? *Gastroenterology* 2004;**127**:1847–9.

62 **Imbert-Bismut F**, Ratziu V, Pieroni L, *et al.* Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. *Lancet* 2001;**357**:1069–75.

63 **Forns X**, Ampurdanes S, Llovet JM, *et al.* Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. *Hepatology* 2002;**36**:986–92.

64 **Hui AY**, Chan HL, Wong VW, *et al.* Identification of chronic hepatitis B patients without significant liver fibrosis by a simple noninvasive predictive model. *Am J Gastroenterol* 2005;**100**:616–23.

65 **Wai CT**, Greenson JK, Fontana RJ, *et al.* A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* 2003;**38**:518–26.

66 **Thuluvath PJ**, Krok KL. Noninvasive markers of fibrosis for longitudinal assessment of fibrosis in chronic liver disease: are they ready for prime time? *Am J Gastroenterol* 2005;**100**:1981–3.

67 **Claassen J**. The gold standard: not a golden standard. *BMJ* 2005;**330**:1121.