Morphometric and Visual Evaluation of Fibrosis in Renal Biopsies

Alton B. Farris,*† Catherine D. Adams,* Nicole Brousaides,* Patricia A. Della Pelle,* A. Bernard Collins,* Ellie Moradi,‡ R. Neal Smith,* Paul C. Grimm,‡ and Robert B. Colvin*

*Pathology Service, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts; ‡Division of Pediatric Nephrology, Stanford University School of Medicine, Stanford, California; and †Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia

ABSTRACT

Interstitial fibrosis is an outcome measure of increasing importance in clinical trials of both renal transplantation and native disease, but data on the comparative advantages of fibrosis measurement methods are limited. We compared four morphometric techniques and contrasted these with two visual fibrosis-scoring methods on trichrome-stained slides. Two morphometric methods included whole-slide digital images: collagen III immunohistochemistry and a new technique using trichrome and periodic acid–Schiff subtraction morphometry; the other two methods included Sirius Red with and without polarization on multiple digital fields. We evaluated 10 serial sections from 15 renal biopsies with a range of fibrosis extent and diagnoses on duplicate sections with each method on separate days. Three pathologists performed visual scoring on whole-slide images. Visual and morphometric techniques had good to excellent interassay reproducibility ($R^2 = 0.62$ to $0.96$) and interobserver reproducibility ($R^2 = 0.75$ to $0.99$, all $P < 0.001$). Morphometry showed less variation between observers than visual assessment (mean of 1% to 5% versus 11% to 13%). Collagen III, Sirius Red unpolarized, and visual scores had the strongest correlations ($R^2 = 0.78$ to $0.89$), the greatest dynamic range, and the best correlation with estimated GFR ($R^2 = 0.38$ to $0.50$, $P < 0.01$ to $0.001$). Considering efficiency, reproducibility, and functional correlation, two current techniques stand out as potentially the best for clinical trials: collagen III morphometry and visual assessment of trichrome-stained slides.


All chronic renal diseases result in the accumulation of extracellular matrix in the interstitium, referred to as fibrosis, which can eventually contribute to renal failure.1–7 For this reason, assessment of interstitial fibrosis is widely used in the evaluation of kidney biopsies.8–11 Several studies have shown that interstitial fibrosis quantification is predictive of renal allograft outcome and may be considered a surrogate marker.12–16 Current research focuses on the therapeutic inhibition of fibrosis13–19 and requires accurate measures of fibrosis to investigate and validate these therapies. Unfortunately, however, there is little comparative information on the relative advantages of the various techniques to quantify fibrosis.

Visual assessment using trichrome stained slides is the standard practice,20 but has been reported to be poorly reproducible.21,22 Several different morphometry techniques have been employed to assess fibrosis, including Sirius Red, specific for collagen types I and III under polarized light,23–25 trichrome,26,27 and immunohistochemistry for type III collagen.28–30 A number of studies have demonstrated the usefulness of computer-assisted morphometry of Sirius Red,
some of which have shown measurements that correlate with GFR. Studies have demonstrated that collagen III staining methods can be predictive of decreased GFR. Given the limitations inherent in the currently used methods for fibrosis measurement, we developed and evaluated a new morphometric method that could be applied in routine stains, using just trichrome and periodic acid–Schiff (PAS) stained slides, taking advantage of whole slide scanners. We then compared this method with the most commonly used alternative methods, Sirius Red, collagen III, and visual assessment, to identify the most robust and efficient technique (Figure 1).

RESULTS

Trichrome-PAS Fibrosis Measurement
Because trichrome stains basement membranes and sometimes brush borders, which should not be included in measurements of interstitial fibrosis, a new method was developed to exclude basement membranes and brush borders. Because basement membranes are stained strongly by PAS, we reasoned that subtraction of the area stained by PAS from that stained blue by trichrome might be a robust measure of interstitial collagen.

Qualitatively, trichrome stained both interstitial fibrosis and basement membranes, whereas PAS stained basement membranes (Figure 2). In addition, both trichrome and PAS stained casts and neither trichrome nor PAS stained brush borders strongly. Therefore, subtracting PAS from trichrome allowed for subtraction of basement membranes and casts in a quantitative manner. Measurements obtained for trichrome yielded a relatively broad range of values (49% to 92%); however, PAS values tended to be in a relatively narrower range (2% to 21%). Therefore, the T-P values obtained through subtraction tended to be related to trichrome values by a relatively fixed amount (Figure 3).

Collagen III Immunohistochemistry (Figure 3)
As reported, collagen III staining was consistently detected in the interstitium as fine lines between normal tubules, in broader areas with fibrosis, and intimal fibrosis of larger arteries. Glomeruli did not stain. One antibody tested gave variable and patchy staining of the normal fibrous tissue between tubules and was not used for analysis (data not shown).

Sirius Red (Figure 4)
As described, under polarized light Sirius Red stains fibrils in the interstitium and in areas of intimal fibrosis. Glomeruli were negative. With unpolarized light, arteries and glomeruli stained.
Comparison of Morphometric Techniques

The three techniques (Sirius Red, collagen III, and trichrome-PAS) each had good interassay reproducibility, as judged by the mean fibrosis percentage and correlation coefficient between values obtained on duplicate sections stained and analyzed on different days (Table 1). The best reproducibility was Sirius Red, unpolarized ($R^2 = 0.96, P < 0.0001$), followed by Sirius Red, polarized ($R^2 = 0.94, P < 0.0001$), collagen III ($R^2 = 0.83, P < 0.0001$), and trichrome-PAS ($R^2 = 0.69, P < 0.0001$). In a comparison of “observers” selecting analysis areas and executing the analysis algorithm, interobserver reproducibility was excellent for collagen III ($R^2 = 0.99, P < 0.0001$) and trichrome-PAS ($R^2 = 0.93, P < 0.0001$), with average absolute differences between observers of 0.9% ± 1.2% (collagen III) and 4.8% ± 4.7% (trichrome-PAS). Sirius Red staining had excellent interobserver reproducibility in prior studies and was not reassessed.12,24,25

Differences in the techniques were most apparent in the absolute level of fibrosis percentage and the dynamic range of fibrosis percentage values. The mean calculated fibrosis percentage was highest for trichrome-PAS (59% to 64%), followed by collagen III (39% to 48%), Sirius Red unpolarized (31%), and Sirius Red polarized (13%). The dynamic range was the greatest for collagen III (24% to 74%), followed by Sirius Red unpolarized (12% to 53%), trichrome-PAS (38% to 82%), and Sirius Red polarized (4% to 22%).

Visual Assessment

Two definitions of the area of fibrosis were used: The area percentage of the cortex occupied by fibrous tissue and the percentage of the cortex that was fibrotic (Figure 5). Three pathologists scored duplicate trichrome stained coded slides.

Table 1. Comparison of morphometric techniques: Extent of fibrosis and interassay reproducibility in duplicate sections

<table>
<thead>
<tr>
<th>Technique</th>
<th>Assay Day</th>
<th>Mean (%)</th>
<th>SD (%)</th>
<th>Minimum (%)</th>
<th>Maximum (%)</th>
<th>Average Change (%)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sirius Red polarized</td>
<td>1</td>
<td>13</td>
<td>5</td>
<td>4</td>
<td>22</td>
<td>0.9 ± 0.8</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13</td>
<td>5</td>
<td>4</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sirius Red unpolarized</td>
<td>1</td>
<td>31</td>
<td>12</td>
<td>12</td>
<td>51</td>
<td>1.8 ± 1.8</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>31</td>
<td>12</td>
<td>15</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collagen III</td>
<td>1</td>
<td>39</td>
<td>11</td>
<td>24</td>
<td>60</td>
<td>9.0 ± 5.0</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>48</td>
<td>12</td>
<td>32</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trichrome-PAS</td>
<td>1</td>
<td>64</td>
<td>13</td>
<td>38</td>
<td>82</td>
<td>7.8 ± 3.6</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>59</td>
<td>12</td>
<td>43</td>
<td>78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aSerial sections stained and evaluated on separate days from 15 biopsies.

bMean ± SD of absolute difference between day 1 and day 2 values, 15 sections in each group.

$R^2$ Coefficient of determination between day 1 and day 2 values, all $P \leq 0.0001$. 

Figure 3. Quantitation of fibrosis using collagen III immunohistochemistry. (A) Collagen III immunohistochemistry stain. (B) The corresponding “mark-up” image generated by the quantitation algorithm applied to the collagen III immunohistochemistry stain. Tissue considered “positive” is “marked up” either yellow, orange, or red, in that order, with increasing positivity of match to the algorithm parameters.

Figure 4. Quantitation of fibrosis using Sirius Red staining. (A) Under polarized light. (B) Screen shot of the quantitative analysis performed on the Sirius Red stain.

Table 1. Comparison of morphometric techniques: Extent of fibrosis and interassay reproducibility in duplicate sections
on different days (Table 2). The intraobserver reproducibility was good to excellent, with $R^2$ values of 0.62 to 0.90 ($P < 0.0001$). The absolute differences between slides evaluated by one observer varied from 7.3% to 13.1%. As expected, the mean percentage of fibrotic cortex had somewhat higher levels than the fibrous area percentage (47% versus 36%) and the maximum was also somewhat higher (100% versus 90%). Correlations between the values using the two definitions for individual observers was high ($R^2 = 0.82, 0.90, and 0.81$). The interobserver reproducibility was calculated using the average of the measurements obtained on day 1 and day 2 for all 15 samples (Figure 6). With use of different pairs of observers, the $R^2$ values for fibrous tissue percentage were 0.90, 0.92, and 0.99 and the $R^2$ values for fibrotic cortex percentage were 0.84, 0.93, and 0.93 ($P < 0.0001$).

**Correlations between Techniques**

The correlation of the visual assessment on duplicate slides was nearly as good as with the morphometric techniques ($R^2 = 0.62$ to 0.90 for visual versus $R^2 = 0.69$ to 0.96 for morphometry); however, the absolute difference in measurements was greater (0.9% to 9.0% for morphometry, 7.3% to 13.1% for visual). Interobserver reproducibility was typically higher for collagen III and trichrome-PAS morphometry ($R^2 = 0.96$ to 0.99) than for visual assessment ($R^2 = 0.75$ to 0.99).

One group of measurements (collagen III, Sirius Red unpolarized, and visual) was highly correlated ($R^2 = 0.78$ to 0.89, $P = 0.001$) (Table 3, Figure 7). The remaining measurement methods (polarized Sirius Red, trichrome-PAS) showed little to no correlation with the other groups ($R^2 = 0.01$ to 0.37). Collagen III correlated with the unpolarized Sirius Red ($R^2 = 0.89$) and with the visual assessments (0.78, 0.88) but not well with the trichrome-PAS (0.24). Unpolarized Sirius Red correlated better with visual assessments ($R^2 = 0.80, 0.86$) than did the polarized Sirius Red ($R^2 = 0.24, 0.25$). Polarized Sirius Red fibrosis percentage did not correlate well with any of the methods ($R^2 = 0.33$ for collagen III method, 0.24 and 0.25 for the visual methods, and 0.14 for the trichrome-PAS method). Correlations were primarily compared using the average of the measurements obtained in 2 days for each of the 15 samples. This average value did not differ substantially from the correlation assessment obtained from the set of day 1 or day 2 values except for the trichrome-PAS method, which showed better correlations with other measurements on day 2 (Supplemental Data). Bland Altman plots for the morphometric methods of measurement showed that the difference between most measurements fell within the limits of agreement (mean differ-

### Table 2. Intra-assay, intraobserver reproducibility of visual assessment of fibrosis

<table>
<thead>
<tr>
<th>Definition of Fibrosis</th>
<th>Observer</th>
<th>Assay Day</th>
<th>Mean (%)</th>
<th>SD (%)</th>
<th>Minimum (%)</th>
<th>Maximum (%)</th>
<th>Average Difference$^a$</th>
<th>$R^2_{bc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>% Fibrous area</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>46</td>
<td>29</td>
<td>15</td>
<td>90</td>
<td></td>
<td>7.3 ± 7.2</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>27</td>
<td>23</td>
<td>2</td>
<td>70</td>
<td></td>
<td>11.0 ± 8.9</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>38</td>
<td>28</td>
<td>4</td>
<td>85</td>
<td></td>
<td>7.7 ± 6.4</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>% Fibrotic cortex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>52</td>
<td>32</td>
<td>6</td>
<td>100</td>
<td></td>
<td>13.1 ± 14.5</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>47</td>
<td>29</td>
<td>5</td>
<td>88</td>
<td></td>
<td>11.7 ± 16.4</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>44</td>
<td>38</td>
<td>3</td>
<td>95</td>
<td></td>
<td>9.3 ± 9.2</td>
<td>0.88</td>
</tr>
</tbody>
</table>

*Three observers scored 15 biopsies stained and evaluated on two different days, according to two definitions of % fibrosis (see Concise Methods).*

*Mean absolute difference ± SD (as in Table 1).*

*Coefficient of determination, as in Table 1, all $P < 0.0001$.*
A principal component analysis showed that measurement methods clustered into three groups: the trichrome-PAS method alone; a group with the Sirius Red unpolarized method, visual methods, and collagen III; and the Sirius Red polarized alone (data not shown).

Correlation of Fibrosis with Renal Function

The patients had a range of serum creatinine (0.4 to 11.9 mg/dl, mean 2.8). The estimated GFR (eGFR) was calculated by three standard formulae: Modification of Diet in Renal Disease (MDRD) capped at 60, Cockcroft-Gault, and CKD-EPI (Table 4).

Table 3. Correlation of percentage of fibrosis measurements between techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Sirius Red Polarized</th>
<th>Sirius Red Unpolarized</th>
<th>Collagen III</th>
<th>Trichrome-PAS</th>
<th>Visual % Fibrous Area</th>
<th>Visual % Fibrotic Cortex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sirius Red polarized</td>
<td>0.30&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>0.33</td>
<td>0.01</td>
<td>0.24</td>
<td>0.26</td>
</tr>
<tr>
<td>Sirius Red unpolarized</td>
<td></td>
<td>0.89&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.14</td>
<td>0.86&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.80&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Collagen III</td>
<td>0.33&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.89&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.24</td>
<td>0.88&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.78&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Trichrome-PAS</td>
<td>0.01</td>
<td>0.14</td>
<td>0.24</td>
<td>0.38&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.36&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Visual % fibrous area</td>
<td>0.24</td>
<td>0.86&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.88&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.38&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.89&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Visual % fibrotic cortex</td>
<td>0.26</td>
<td>0.80&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.78&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.36&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.89&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

Coefficient of determination ($R^2$) between methods with the results obtained from an average of measurements obtained by each method in 2 days for 15 samples.

<sup>a</sup>0.05 > P > 0.001.

<sup>b</sup>P ≤ 0.001.
The various measures of renal function correlated with each other, with $R^2$ values from 0.29 (Cr versus Cockcroft-Gault) to 0.67 (Cr versus MDRD) and $r$ values from 0.81 (MDRD versus CKD-EPI) to −0.82 (Cr versus MDRD), with average absolute differences between the methods ranging from 27–29 ml/min (Cockcroft-Gault versus CKD-EPI) to 41–87 ml/min (MDRD versus CKD-EPI) (other data not shown). The extent of fibrosis correlated with functional estimates with varied strength and breadth, according to the fibrosis measurement and the eGFR formulae. The best performance was with the collagen III, unpolarized Sirius Red, and the two visual assessments ($R^2 = 0.35$ to 0.50). Trichrome-PAS showed a strong correlation with Cockcroft-Gault and no other measure, and the polarized Sirius Red did not show a correlation with any of the functional parameters. The slope of the eGFR versus the fibrosis percentage can be used to calculate the expected

Table 4. Correlation of renal function with percentage of fibrosis measured by each technique

<table>
<thead>
<tr>
<th>Technique</th>
<th>Sirius Red Polarized</th>
<th>Sirius Red Unpolarized</th>
<th>Collagen III</th>
<th>Trichrome-PAS</th>
<th>Visual % Fibrous Area</th>
<th>Visual % Fibrotic Cortex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum Cr (mg/dl)</td>
<td>0.004</td>
<td>0.21*</td>
<td>0.26</td>
<td>0.12</td>
<td>0.25</td>
<td>0.32*</td>
</tr>
<tr>
<td>eGFR CKD-EPI (ml/min per 1.73 m²)</td>
<td>0.07</td>
<td>0.45*</td>
<td>0.50*</td>
<td>0.07</td>
<td>0.38*</td>
<td>0.42*</td>
</tr>
<tr>
<td>eGFR MDRD (ml/min per 1.73 m²)</td>
<td>0.02</td>
<td>0.43*</td>
<td>0.42*</td>
<td>0.05</td>
<td>0.35*</td>
<td>0.41*</td>
</tr>
<tr>
<td>eGFR Cockcroft-Gault (ml/min)</td>
<td>0.04</td>
<td>0.22</td>
<td>0.29*</td>
<td>0.36*</td>
<td>0.25</td>
<td>0.28*</td>
</tr>
</tbody>
</table>

Coefficient of determination ($R^2$) between the creatinine, the mean eGFR using three methods, and the average of % fibrosis measurements by each method in 2 days for 15 samples.

$^{*} 0.05 > P > 0.001$. 

**Figure 7.** Correlation of computer-based and visual measurements. Regression lines show the correlations of computer-based and visual measurements with measurements of day 1 and day 2 sections of each sample averaged. Regression $R^2$ values are given for the corresponding combined day 1 and day 2 measurements. Curved lines bound a density ellipse containing 95% of the measurements obtained. Corresponding $P$ values are given in Table 3. SR, Sirius Red; T-P, trichrome-PAS.
change in eGFR per unit change in fibrosis percentage (Table 5, Figure 8). Correlations and slopes were primarily compared using the average of the measurements obtained in 2 days for each of the 15 samples. This average value did not differ substantially from the correlation assessment obtained from the set day 1 or day 2 values (Supplemental Data).

DISCUSSION

In this study, we compared the standard published methods for measuring fibrosis by morphometry and by visual inspection, to determine their ease of use, reproducibility, and correlation with function, including a new method based on trichrome and PAS stained slides (Table 6).

We found that the trichrome-PAS method had excellent interobserver reproducibility and moderate interassay reproducibility, as well as a moderate dynamic range. The assay correlated well with GFR measured by the Cockcroft-Gault method and with visual assessment of fibrosis and has the advantage of less observer variability. Although we used slides scanned with the Aperio slide scanner and analyzed with Aperio software, similar techniques could be applied with slides scanned with other instruments and analyzed with other morphometry software packages. The assay relies on routine stains prepared on renal biopsies, which is regarded as an advantage. Disadvantages are potential variation in trichrome stains, the need for a digital slide scanner, and the extra labor involved. Alternative methods to select interstitial areas have been used with image recognition software, which have the advantage of requiring a single slide.

We evaluated two methods of assessing fibrosis by visual inspection. In one, percentage of fibrosis area is defined as a morphometric algorithm, estimating the percentage of the cortex that is fibrous tissue. The second definition measures the percentage of cortex affected by fibrosis. We found that either method gives moderate to high interassay and interobserver reproducibility. This is not always the case for visual methods for fibrosis. We propose four reasons for the success: (1) we defined what was to be measured as percentage of fibrosis, (2) the area to be measured was specified (cortex bounded by glomeruli), (3) all slides were scanned and evaluated on a computer screen, and (4) the stains were done in the same laboratory (although on different days). An additional effect,
namely, that the pathologists work together, may also have been a factor. We found no objective reason to choose between the two definitions: both had similar reproducibility and correlation with function. It is notable that no standard definition is used by pathologists, which adds to the confusion in the literature. It will be important to indicate which definition is used in clinical studies, however, because they have different ranges. Theoretically, the percentage of fibrotic cortex could be 100%, but the percentage of fibrous tissue would necessarily be less because some cortical area is always occupied by tubules, glomeruli, and vessels. An international multicenter project is underway to evaluate the reproducibility of this system with a larger group of pathologists, initiated by the Banff Fibrosis Scoring Working Group.33

We compared the different techniques for reproducibility, dynamic range, and correlation with function (Table 6) because we were unable to find a study in the literature with a side by side comparison on the same samples of all these techniques. We conclude from our analysis that each method is satisfactorily reproducible and applicable to clinical studies of fibrosis. The methods sort into collagen-based measurements (collagen III and Sirius Red) and trichrome-based assays. This is probably due to measurement of different components of the matrix. Previous studies showed that polarized Sirius Red correlates poorly with pathologists score of fibrosis.24, 25 Which method is more relevant to long-term outcome remains to be established. The highest dynamic ranges were with collagen III and visual assessment. There is wide variability in how easy assays are to perform ranging from the most difficult (Sirius Red with polarization) to the easiest (visual). Sirius Red requires individual fields to be photographed in polarized light and analyzed, a process that will be considerably more efficient with whole slide scanners capable of polarized light. Sirius Red stains are robust and probably less dependent on the fixation than collagen III immunohistochemistry. Collagen III stains also require validation of the sensitivity of the commercial antibody, which can show variability with different antibodies.

Although there is no perfect measure of renal function based on limited clinical parameters,34,35 many of the computer-assisted and pathologist assessments of fibrosis in this study correlated with clinical measures of renal function as determined by the serum creatinine level and by eGFR according to the MDRD, Cockcroft-Gault, and CKD-EPI formulae. Interestingly, the correlation between the various eGFR calculations were no better than those between pathologists’ visual assessments of fibrosis. As to concurrent renal function, the closest correlations were with the visual, collagen III, and unpolarized Sirius Red. The discovery that the unpolarized Sirius Red measurements correlate better than the polarized Sirius Red with GFR has been shown in prior studies.36 Furthermore, in at least one study, the visual assessment of fibrosis had greater predictive value of later renal function than Sirius Red morphometry.25

There remain intrinsic limitations of the measurement of fibrosis, primarily because of sampling. Other studies have estimated that repeat biopsies show a decrease of fibrosis, presumably because of sampling, in 12% of cases.27 We addressed this question in our study by comparing different cores from the same patient and found that the absolute difference averaged 2.5% and 4.9% for trichrome-PAS and collagen III, respectively (data not shown). In addition to extent, the nature and pattern of the fibrosis is also probably important. For example, fibrosis that is “active” or “young” may have greater potential for remodeling. Diffuse fine fibrosis may have different consequences than broad scars. Inflammation in areas of fibrosis has also been noted in several studies to be an adverse risk factor for progression of renal disease.16,37,38

At the present time, for the purpose of research studies and clinical trials, particularly studies of renal fibrosis, this study implies that studies could benefit from using more than one fibrosis measurement method. Studies might benefit from using one of the methods that correlated with trichrome and one of the methods based on collagen measurement. In addition, measurements by pathologists can also be useful because pathologists can give reproducible results that correlate statistically in a relative manner between pathologists, although the absolute measurements given by pathologists may vary somewhat. We conclude that considering the efficiency, reproducibility, and functional correlation, two current techniques are best for the interpretation of renal fibrosis: collagen III morphometry and visual assessment of trichrome stained slides. These tests are optimal for clinical trials, particularly when a laboratory has established a well-controlled collagen III immunohistochemical staining method. Improvements in scanning, spectral analysis, and software will provide better measurements of fibrosis for evaluation of renal biopsies and outcome analysis of therapeutic trials for renal fibrosis.

**CONCISE METHODS**

**Biopsy Samples**

Fifteen renal biopsies were selected from our database from 2008 that had at least seven glomeruli in the block available for sectioning, regardless of diagnosis, and that had a wide range of fibrosis from normal to extensive. Fourteen were native kidneys and one was an allograft. The diagnoses were lupus nephritis (four), IgA nephropathy (three, including one in an allograft), focal segmental glomerulosclerosis (three, one with thrombotic microangiopathy), diabetic glomerulosclerosis (two), Henoch-Schoenlein purpura (one), membranous glomerulonephritis (one), and membranoproliferative glomerulonephritis (one). The average age was 32.6 ± 19.9 years (3 to 64 years); nine patients were Caucasian, three Hispanic, two African American, and one Asian.

**Stains**

Ten 6-μm-thick serial sections were prepared by the same experienced histotechnologist on the same microtome and duplicate sections stained with trichrome, Sirius Red, periodic acid-Schiff (PAS), and antibodies to collagen III on separate days as shown in Figure 1. Trichrome stains were done in a standard manner with the Toulouse-
Latrec 1 Step Trichrome Stain Kit (Biocare Medical, Concord Park, CA). For collagen III immunohistochemistry (IHC) antigen retrieval was performed using the BORG Decloaker (Biocare Medical). Polyclonal rabbit anti-human collagen III (LS-B693; Lifespan Biosciences, Seattle, WA) was used at a dilution of 1:400. This antibody proved the more sensitive and consistent of the two we tested (data not shown). Antibody binding was detected with Envision+ Dual Link System with horseradish peroxidase (Dako, Carpineria, CA) and developed with diaminobenzidine using the Autostainer Plus automated immunohistochemical stainer (Dako). Sirius Red staining was done as described, by treated rehydrated sections stained 1 hour with saturated picric acid containing 0.1% Sirius Red F3BA (Sigma-Aldrich-Fluka Chemicals, St. Louis). The slides were washed in two changes of acidified water, then dehydrated in three changes of 100% ethanol, and then washed in xylene.24,25

To minimize the contribution of subcapsular regions and normal fibrous tissue around larger arteries, we measured fibrosis only bounded by glomeruli, for both morphometry and visual techniques.

**Morphometry**

For collagen III, stained sections were scanned with an Aperio ScanScope CS (Aperio Technologies, Inc., Vista, CA) and analyzed using the ImageScope Positive Pixel Count algorithm.39,40 The Aperio ScanScope allowed scanning and quantitation of the whole slide using a ×20 objective lens with a numerical aperture of 0.75 coupled with a doubler objective to achieve a scan of whole slides at ×40 magnification. The default parameters of the Positive Pixel Count (hue of 0.1 and width of 0.5) detected collagen III IHC adequately.

Sirius Red stained slides were analyzed using methods previously published.24 In brief, images were examined with a Nikon E600 microscope, and a Spot Digital Firewire camera was used to capture 24 bit RGB color images that were stored as TIFF files. A background image of a blank area of the slide was initially obtained and background correction was performed in real time to adjust for subtle irregularities in the illumination of the microscope field. The images were acquired using the ×20 objective with a numerical aperture of 0.5. Images of the cortex of the biopsy from one end to the other were obtained in a serpentine fashion starting at one end of the tissue and working toward the other. Image analysis was performed on the stored images using an automated macro specially written for the software package NIH ImageJ.41,42 Automated analysis of the images was performed with operator supervision. After the software was set to differentiate the positively stained from negatively stained areas on the first image, the software sequentially opened each image, did the analysis, stored the data, closed the image, and moved on to the next image until the entire biopsy was analyzed. The operator’s only function during the analysis phase was to watch the screen and stop the process if an error was detected.

For trichrome-PAS measurements, stained sections were scanned with an Aperio ScanScope CS and analyzed using the ImageScope Positive Pixel Count algorithm.39,40 For trichrome and PAS, hue values for blue and pink were measured in all cases, and an average hue for the trichrome (0.64) and PAS (0.854) were used when evaluating these stains. The default hue width was used for the trichrome stained slides, and a narrower hue width (0.035) was selected for the PAS stained slides to set an adequate threshold for detecting only PAS-positive basement membranes. In this manner, the percentage of cortical interstitial fibrosis (T-P fibrosis) was calculated (% positive trichrome pixels − % positive PAS pixels).

**Visual Assessment**

Three pathologists viewed whole scanned trichrome stained slides from the first and second days of trichrome staining on a computer monitor and assigned visual assessment of fibrosis. What area of fibrosis is to be measured is often not defined in the literature, whether the area percentage of the cortex occupied by fibrous tissue (as used in morphometry) or the percentage of the cortex that is fibrotic (as commonly used by pathologists). In this study we compared these two contrasting definitions. As in the morphometric analysis, the cortex was assessed in that part bounded by glomeruli, which eliminates the immediate subcapsular area and the larger vessels at the corticomedullary junction.

**Renal Functional Data**

Creatinine values and patient parameters were obtained from the clinical database and eGFR was calculated by both the Cockcroft-Gault,63 MDRD,44 and Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI)65 equations. The MDRD eGFR was capped at 60, which is the highest value typically accepted with the MDRD equation. Regression with fibrosis analysis measurements and renal function measurements was performed.

**Statistical Analysis**

Microsoft Excel (Microsoft Corporation, Redmond, WA) was used to record data and perform preliminary statistical analysis. Further statistical analysis, including multivariate regression analysis and principal component analysis, was performed in SAS JMP version 7.0 (SAS Institute, Cary, NC).

**ACKNOWLEDGMENTS**

This work has been supported by grants from the National Institutes of Health (U01-AI-63623, U01-AI-077816, U01-AI-070107).

**DISCLOSURES**

None.

**REFERENCES**


Supplemental information for this article is available online at http://www.jasn.org/.