Reproducibility of the Banff schema in reporting protocol biopsies of stable renal allografts

James Gough1, David Rush2, John Jeffery2, Peter Nickerson2, Rachel McKenna2, Kim Solez3 and Kiril Trpkov3

1Department of Pathology and 2Department of Nephrology, University of Manitoba Health Sciences Centre, Winnipeg, Manitoba, Canada and 3Department of Laboratory Medicine and Pathology, University of Alberta, 5B4.02 WC MacKenzie Health Sciences Centre, Edmonton, Alberta, Canada

Keywords: Banff; inter-observer agreement; protocol renal biopsy; rejection

Abstract

Background. There is evidence that biopsy of stable renal allografts may be of value in predicting chronic allograft nephropathy, the main cause of graft loss. However, the reproducibility of such histological evaluation has not been tested in this setting. We tested the reproducibility of the Banff schema for this purpose.

Methods. We rated acute and chronic changes in 184 protocol biopsies. Individual pathologists at two different Canadian transplant centres reported independently.

Results. There was agreement in 73.53, 42.86, and 77.08% of cases in assigning a diagnosis of acute rejection, borderline changes (as defined in the schema), and no acute rejection, respectively. Applying kappa statistics, there was very good agreement in making the diagnosis of acute rejection vs no acute rejection (kappa 0.77). There was good inter-observer agreement in scoring glomerulitis, intimal arteritis, interstitial infiltrates, tubulitis, and arteriolar hyalinosis. Rating chronic changes also gave good inter-observer agreement (kappa 0.53, 0.65, and 0.62, respectively, for mild, moderate, and severe chronic allograft nephropathy). Agreement on transplant glomerulopathy was, however, poor.

Conclusions. We conclude that the Banff classification provides a reproducible method for the histological assessment of protocol renal allograft biopsies in stable grafts. Such biopsies may be valuable in detecting subclinical rejection and early chronic allograft nephropathy and may also be used as surrogate end-points in the evaluation of therapy to prevent the latter.

Introduction

Protocol renal allograft biopsy is a potentially valuable diagnostic and research tool. Important applications of the technique include uncovering ‘subclinical rejection’ [1,2], the prediction of chronic allograft nephropathy [3], and the establishment of a surrogate end-point for therapeutic trials for preventing the latter [4]. The reproducibility of the histological interpretation of protocol biopsies is thus of great importance, but, to our knowledge, no published study has specifically examined this issue. In the present study we applied the Banff schema to a large number of protocol biopsies interpreted by pathologists at two large Canadian transplant centres.

Subjects and methods

All biopsies performed between April 1992 and April 1995 on 64 consecutive patients at one centre who had consented to take part in a clinical trial on the effects of treating ‘subclinical rejection’ were examined. The outcome of this study is published elsewhere [5]. Biopsies were done according to a protocol at 1, 2, 3, 6, and 12 months after transplantation. Slides were prepared according to standard techniques. One hundred and eighty-four biopsies contained at least seven glomeruli and an artery, which is defined as adequate for interpretation [6]. Biopsies were scored according to the Banff schema [6] for evidence of acute and chronic changes and for arteriolar hyalinosis. The schema recognizes three types of acute cell-mediated rejection; tubulointerstitial, vascular, and severe vascular rejection based on the extent of inflammation present in the interstitium, tubules (‘tubulitis’), and vessels. A ‘borderline’ category is also used. Acute antibody-mediated rejection is also categorized in the schema. Chronic changes are graded...
using a semi-quantitative score (score 0–3 (normal to severe)) for interstitial fibrosis and tubular atrophy. Sclerosing transplant arteriopathy and transplant glomerulopathy are graded separately as specific indicators of chronic rejection. Finally, there is a category for ‘other diagnoses’ recognized in the allograft.

All biopsies were interpreted by pathologists at two transplant centres: by either one of two pathologists at centre A and by one pathologist at centre B. All of these individuals were blinded to each other’s reports and to clinical information.

We applied kappa statistics to our results. This statistical technique has been widely used to study the reproducibility of histological diagnoses, especially those using a grading or scoring system [7].

Results

Table 1 shows the level of agreement between the two centres in assigning one of three histological diagnoses to each biopsy, namely ‘acute rejection’, ‘no acute rejection’, and ‘borderline changes’.

The kappa value for the diagnosis of acute rejection vs no acute rejection (the latter including the ‘borderline’ category) was 0.77.

Taking these three diagnoses (no acute rejection, borderline, and acute rejection) as a scale from 0 to 2, the value of kappa in assigning them to a biopsy was 0.69 with 95% confidence limits of 0.60–0.79 (see Table 2). As can be deduced from Table 2, the prevalence of the three categories (agreed upon by both centres) ‘no acute rejection’, ‘borderline’, and ‘acute rejection’ in the 184 biopsies was 27, 13 and 40%, respectively.

Table 3 shows the kappa values for inter-observer agreement in scoring glomerulitis, intimal arteritis, interstitial infiltrates, and tubulitis from 0 to 3. Glomerulitis and intimal arteritis were uncommon (10 and seven biopsies, respectively).

Although the histological diagnosis alone defined rejection in these cases, in 17 of the 186 biopsies (9.3%), a clinical diagnosis of rejection was made at the time of protocol biopsy. All of these were reported as acute rejection in the biopsy.

Unlike acute rejection, chronic graft injury is not well defined either clinically or histologically. In the Banff schema the histological lesions of chronic graft damage are rated according to the severity of interstitial fibrosis and tubular atrophy. Transplant glomerulopathy and sclerosing transplant vasculopathy are taken as indicators of chronic rejection. Table 4 shows kappa values for these four items as rated by the two centres.

The kappa values for mild, moderate, and severe chronic allograft nephropathy were 0.53, 0.65 and 0.62, respectively (with 95% confidence intervals of 0.50–0.69). Arteriolar hyalinosis is scored from 0 (no arteriolar hyalinosis) to 3 (severe arteriolar hyalinosis) in the Banff system. Inter-observer agreement for hyalinosis yielded a kappa value of 0.63 (with 95% confidence intervals of 0.58–0.73).

Discussion

Our study shows that the Banff system offers a reproducible histological diagnosis of acute rejection in stable renal allograft patients and we have shown elsewhere that this finding is clinically significant for graft function at 2 years [5]. There is a greater discrepancy in assigning the ‘borderline’ category but the practical consequences of this were minimal insofar as centre B tended to call ‘borderline’ what centre A called ‘no acute rejection’. In only one case was the ‘borderline’ diagnosis of centre B called ‘acute rejection’ by centre A. While Saad et al. [8] have shown that the ‘borderline’ category may be important in diagnostic graft biopsies, they have also pointed out that this conclusion cannot be extrapolated to protocol biopsies. Nevertheless, the significance of borderline changes in protocol biopsies must be further evaluated.

Table 1. Inter-observer agreement between two centres in the assignment of three histological diagnoses to protocol renal transplant biopsies

<table>
<thead>
<tr>
<th>Histological diagnosis</th>
<th>Inter-observer agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>No acute rejection</td>
<td>77.08%</td>
</tr>
<tr>
<td>Borderline changes</td>
<td>42.86%</td>
</tr>
<tr>
<td>Acute rejection</td>
<td>73.53%</td>
</tr>
</tbody>
</table>

Table 2. Inter-observer agreement between two centres in assigning three histological diagnoses to 184 protocol renal allograft biopsies

<table>
<thead>
<tr>
<th></th>
<th>Centre A</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No acute rejection</td>
<td>Borderline</td>
<td>Acute rejection</td>
</tr>
<tr>
<td>Centre B</td>
<td></td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>No acute rejection</td>
<td>74</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Borderline</td>
<td>14</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>Acute rejection</td>
<td>4</td>
<td>13</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>92</td>
<td>41</td>
<td>51</td>
</tr>
</tbody>
</table>

The diagnoses of centre A are read in columns while the corresponding diagnoses of B are read in rows. Thus, 74 cases were assigned a diagnosis of ‘no acute rejection’ by centre A whereas centre B assigned 78 cases to this category, the four additional cases having been diagnosed as ‘borderline’ by centre A.
A kappa value greater than 0.75 is believed to signal excellent agreement in diagnosis [7]. Concordance for the biopsy diagnosis of invasive carcinoma of the cervix, for example, is reported by one group as giving a kappa value of 0.832 [9].

There have been few studies of inter-observer agreement in the biopsy diagnosis of renal allograft rejection. One such report showed that knowledge of the clinical findings greatly influenced biopsy interpretation [10]. Marcussen et al. [11] examined the reproducibility of the Banff schema in 77 biopsies, mostly clinically indicated, from 46 patients. They found a kappa value of 0.56 (compared with our figure of 0.77) for inter-observer agreement in the diagnosis of acute rejection vs no acute rejection (the latter, including the borderline category).

Recently, Colvin et al. [12] reported a kappa value of 0.80 in a study of the reproducibility of histological criteria for acute rejection devised by the Co-operative Clinical Trials in Transplantation (CCTT) group. More than 90% of these biopsies were performed for diagnostic rather than protocol purposes. The CCTT schema does not evaluate chronic changes.

In evaluating chronic allograft nephropathy, our kappa values were 0.53, 0.65, and 0.62 for mild, moderate, and chronic allograft nephropathy, respectively. While not as good as those for acute rejection, these figures are comparable with those achieved in assessing inter-observer agreement, for example, in the grading of breast carcinoma [13].

Arteriolar hyalinosis is an important finding in graft biopsies. It may signify donor disease, chronic rejection, or cyclosporin toxicity. We found very good inter-observer agreement in grading this parameter (kappa = 0.63).

Our kappa value (0.14) for ‘cg’ (transplant glomerulopathy) is poor. The presence and extent of splitting of the glomerular basement membrane (GBM) is the principal marker for cg in the Banff schema [6]. Centre B tended to over-diagnose GBM splitting especially in the presence of glomerulosclerosis. Both centres agreed that no transplant glomerulopathy was present in 86.4% of cases. In a further 11.4% of cases the scores for cg differed by only 1. The inappropriate rating of cg was acknowledged by centre B.

Although glomerulosclerosis is not a specific indicator of transplant glomerulopathy, it may be the end-result of that condition. Whatever its aetiology, it may be an important prognostic finding in a transplant biopsy and for this reason we have incorporated glomerulosclerosis as a measure of chronic injury in other studies [5] and the Banff schema recommends that it be independently recorded in biopsy reports [6].

The overall results of this exercise in inter-observer comparison in rating protocol renal allograft biopsies are encouraging and show good agreement in assigning a diagnosis of acute rejection (kappa = 0.77). While the kappa values for rating the individual parameters evaluated in arriving at this diagnosis are lower than this (ranging from 0.42 to 0.50), there is no ‘right’ kappa established for such parameters and our kappa values must await comparison with those of other studies.

In summary, we believe that the Banff schema provides a reproducible method of assessing protocol renal transplant biopsies and facilitates reliable studies of the significance and long-term consequences of ‘subclinical rejection’ and the establishing of a surrogate endpoint for chronic rejection.

References


*Received for publication: 16.1.01*  
*Accepted in revised form: 8.1.02*