Interpretation of transbronchial lung biopsies from lung transplant recipients: Inter- and intraobserver agreement

Anne Stephenson MD, Julia Flint MD, John English MD, Sverre Vedal MD, Guy Fradet MD, Dean Chittock MD, Robert D Levy MD

BACKGROUND: Transbronchial lung biopsy results are crucial for the management of lung transplant recipients. Little information is available regarding the reliability and reproducibility of the interpretation of transbronchial lung biopsies.

OBJECTIVE: To examine the inter-reader variability between two lung pathologists with expertise in lung transplantation.

METHODS: Fifty-nine transbronchial lung biopsy specimens were randomly selected. Active infection had been excluded in all cases. The original interpretations (as per the Lung Rejection Study Group) for acute rejection grade included 19 biopsies scored as A0 (none), 14 scored as A1 (minimal), 12 as A2 (mild), 11 as A3 (moderate) and three as A4 (severe). The pathologists worked independently without clinical information or knowledge of the original interpretation. The specimens were graded using the Lung Rejection Study Group criteria for acute rejection (grades A0 to A4), airway inflammation (grades B0 to B4) and bronchiolitis obliterans (C0 absent and C1 present). Between-reader agreement for each category was analyzed using a Kappa statistic.

RESULTS: Because many transplant specialists initiate augmented immunosuppression with biopsy grades of A2 or higher, results for each reader were dichotomized as A0/A1 versus A2/A3/A4. Using this dichotomy, there was only moderate agreement (kappa 0.166, P = 0.108). The intrareader agreement for acute rejection was substantial (kappa 0.795, P = 0.0001; kappa 0.676, P = 0.0001). The intrareader agreement for acute rejection (grades A0 to A4) and airway inflammation (grades B0 to B4) and bronchiolitis obliterans (C0 absent and C1 present) was substantial (kappa 0.795, P = 0.0001; kappa 0.676, P = 0.0001). Between-reader agreement for each category was analyzed using a Kappa statistic.

CONCLUSIONS: Because the agreement between expert pathologists is only modest, optimum clinical decision-making requires that transbronchial lung biopsy results be used in an integrated clinical context.

Key Words: Acute rejection; Lung transplantation; Transbronchial biopsy

The use of surveillance transbronchial biopsies for the early detection of complications following lung transplant is standard practice in most transplant centres (1,2). In addition, biopsies are performed in settings of clinical concern such as increasing symptoms or declining pulmonary function. Biopsy results are essential when making clinical decisions regarding patient management. Often, biopsy results are a major factor in the decision-making regarding augmentation or a change in...
immunosuppressive medication in the setting of suspected allograft rejection (3). Because the development of repeated or severe episodes of acute rejection have been shown to be risk factors for the subsequent development of bronchiolitis obliterans (BO), biopsy diagnosis of this condition has both treatment and prognostic implications for patients (4). Conversely, transbronchial biopsy is rarely relied on for the identification of chronic allograft dysfunction because of its well-recognized low sensitivity for the diagnosis of BO.

In 1990, the Lung Rejection Study Group (LRSG) developed a classification system for pulmonary allograft histopathology (5), which was subsequently revised in 1995 (6). The purpose of this grading system was to establish a uniform, reproducible standard for diagnosing and qualifying acute rejection, airway inflammation, BO and chronic vascular rejection. Using this classification, data can be compared among centers to develop a consistent approach toward patient management. Although using an objective standard approach to biopsy interpretation is crucial, the authors also emphasized the need for pathologists to interpret the biopsy data in an integrated clinical context to improve the accuracy of diagnosis. Furthermore, because infection and rejection can be confused histologically, the presence of infection must always be ruled out.

Although this detailed classification system is used at many institutions that perform lung transplantation, little attention has been paid to the validity and reproducibility of this approach, specifically with respect to the inter- and intrareader variability of biopsy interpretation. This is an important area to explore in view of the fact that the biopsy result is often a major factor taken into consideration when making decisions regarding therapy. The purpose of the current study was to investigate the inter- and intrareader variability of two lung pathologists with expertise in lung transplantation at Vancouver General Hospital in Vancouver, British Columbia.

MATERIALS AND METHODS

Fifty-nine transbronchial lung biopsy specimens were randomly selected for the study. All grades of acute rejection (6) (based on the original interpretations) were represented. There were 19 biopsies scored as A0 (none), 14 scored as A1 (minimal), 12 scored as A2 (mild), 11 scored as A3 (moderate) and three as A4 (severe). All cases selected were thought to contain adequate tissue for diagnosis at the time of the original interpretation. The original interpretations were primarily performed by a third pathologist not involved with the current study.

The slides were independently graded by two experienced lung pathologists without clinical information or previous biopsy results using the LRSG classification system. The pathologists completed an extensive detailed questionnaire developed for the study, outlining multiple factors including specimen adequacy, number of tissue fragments, visible airways, vessels and alveoli, evidence of infection and histological grade. Although specimen adequacy is typically defined as at least five pieces of alveolated lung parenchyma, the pathologist was allowed to subjectively evaluate the tissue specimen; if there were fewer than five pieces of alveolated tissue but they felt confident of the diagnosis, the specimen was considered to be adequate.

In many centers, depending on the clinical circumstances, the threshold for a decision to augment or alter immunosuppression in acute rejection is initiated with a biopsy grade of A2 or higher. For this reason, acute rejection results were dichotomized as A0/A1 versus A2/A3/A4. For airway inflammation, the results were dichotomized as B0 versus B1/B2/B3/B4. BO is considered to be normal, B1 is minimal airway inflammation, B2 is mild, B3 is moderate and B4 is severe. BO was dichotomized as C0 (no evidence of BO) versus C1 (presence of BO).

Inter-reader agreement for each category was analyzed using a kappa statistic with values less than 0.21 indicating poor agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement and 0.81 to 1.00 almost perfect agreement (7).

In the second part of the study, intrareader agreement was examined. Three months after the first interpretation, the same set of slides was given to the same two pathologists at which time they graded them a second time using the LRSG criteria. The level of agreement between the first and second reading for each pathologist was analyzed using a kappa statistic.

RESULTS

In terms of acute rejection, reader 1 rated the pathology specimens as follows: 17 biopsies were rated as A0, 14 rated as A1, 12 as A2, 10 as A3 and one as A4. The breakdown for acute rejection for reader 2 was 18 biopsies were rated as A0, 15 as A1, 11 as A2, 13 as A3 and one as A4. A total of five slides were excluded from the final kappa analysis for the following reasons: one slide was deemed undiagnosable by both pathologists and was not graded for acute rejection, and four slides were not graded for acute rejection by reader 2; therefore, they could not be used in the statistical analysis. A total of 54 slides were included in the final analysis for acute rejection. The inter-reader agreement for acute rejection was moderate (kappa 0.470, P<0.001) between the two lung pathologists (Table 1). There was disagreement for 26% of the slides in this category. The level of agreement between the readers was fair for airway inflammation (kappa 0.333, P=0.014) (Table 2) and poor for BO (kappa 0.166, P=0.108). The two pathologists disagreed on whether airway inflammation or BO was present for 33% of the biopsies.

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Inter-reader agreement for acute rejection*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader 1</td>
<td>Reader 2</td>
</tr>
<tr>
<td>Grade A2-A4</td>
<td>16</td>
</tr>
<tr>
<td>Grade A0-A1</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Inter-reader agreement for airway inflammation*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader 1</td>
<td>Reader 2</td>
</tr>
<tr>
<td>Grade B1-B4</td>
<td>12</td>
</tr>
<tr>
<td>Grade B0</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
</tr>
</tbody>
</table>

*A0 Normal; A1 Minimal; A2 Mild; A3 Moderate; A4 Severe

*B0 Normal; B1 Minimal; B2 Mild; B3 Moderate; B4 Severe
The intrareader agreement for both reader 1 and reader 2 for acute rejection was substantial with kappas of 0.795, \(P=0.0001\) and 0.676 \(P=0.0001\), respectively.

DISCUSSION
The present study shows that agreement between expert lung pathologists is moderate when interpreting acute rejection for transbronchial biopsy specimens. There was less agreement when airway inflammation and BO were interpreted. Furthermore, although intrareader agreement was high, it was not perfect. This latter observation reflects the element of subjectivity inherent in the interpretation of pathological specimens. The LRSG classification is a helpful guide but it is important to recognize its limitations. Ideal conditions are needed to maximize the accuracy of interpretation. Adequate tissue sampling, clinical information and good communication between clinicians and lung pathologists are crucial to ensure accurate diagnosis and improved patient management.

One possible limitation of the present study is the generalizability of the results. The setting in which the pathologists read the biopsy specimens was artificial because they did not have access to clinical information or previous biopsy results. We felt this was necessary to try to get an objective assessment of the accuracy of biopsy interpretation using the LRSG criteria alone. In this regard, the intent of the present study was not to assess the accuracy of biopsy interpretation itself, but rather the ability of two pathologists to agree on the classification of the specimens. Furthermore, the pathologists read the biopsies in the context of a study that may have altered their interpretations; however, because they were both under the same influence, this should not have affected their ability to agree on the final diagnosis.

The current study is representative of 'real-world' clinical medicine in which decisions are made based on biopsy interpretations, recognizing that the interpretation may differ depending on a number of factors, including, most importantly, the experience of the pathologist. Even with specialized pulmonary pathologists, there is a subjective component to biopsy interpretation and agreement on diagnosis is not always achieved. Many transplant centres do not have access to dedicated experienced pulmonary pathologists to interpret the transbronchial biopsy specimens, which may limit interpretation even further. From a clinician's standpoint, optimum decision-making clearly requires that transbronchial biopsy results be used in an integrated context with additional clinical information and previous biopsy results to increase the reliability of the interpretation of the biopsy.

CONCLUSIONS
Pathological specimens play a crucial role in the diagnosis of acute allograft rejection. The interpretation of these specimens affects patient care and directs medical therapy. The present study showed that the diagnosis of acute rejection based on transbronchial lung biopsy is moderate between two lung pathologists with expertise in pulmonary transplantation. There is less agreement on the presence or absence of airway inflammation and BO. Our findings support the original recommendations made by the authors of the LRSG (5), who suggested that the biopsy specimens need to be interpreted in an integrated clinical context to allow for optimum patient management and clinical decisions. In an effort to improve the accuracy of diagnosis, all clinically relevant information must be made available to the pathologists because biopsy interpretation has an important bearing on therapy and potential significant consequences for the patients.

REFERENCES