ELSEVIER

The Journal of Heart and Lung Transplantation

http://www.jhltonline.org

A web-based pilot study of inter-pathologist reproducibility using the ISHLT 2004 working formulation for biopsy diagnosis of cardiac allograft rejection: The European experience

Annalisa Angelini, MD,[a] Claus Boegelund Andersen, MD, DMSc,[b] Giovanni Bartoloni, MD,[c] Fiona Black, MD, FRCPath,[d] Paul Bishop,[e] Helen Doran, MB, FRCPath,[e] Marny Fedrigo, MD,[a] Jochen W.U. Fries, MD, PhD,[f] Martin Goddard, MD, FRCPath,[g] Heike Goebel, MD,[f] Desley Neil, BMedSc, FRCPath,[h] Ornella Leone, MD,[i] Andrea Marzullo, MD,[j] Monika Ortmann, MD,[f] Francois Paraf, MD,[k] Samuel Rotman, MD,[l] Nesrin Turhan, MD,[m] Patrick Bruneval, MD,[n] Anna Chiara Frigo, MSc,[o] Francesco Grigoletto, MD,[o] Alessio Gasparetto, MSc,[p] Roberto Mencarelli, MD,[p] Gaetano Thiene, MD,[a] and Margaret Burke, MD, FRCPath[q]

From the [a]Department of Medical Diagnostic Sciences and Special Therapies, University of Padua, Padua, Italy; [b]Rigshospitalet, Copenhagen, Denmark; [c]Department of Pathology, University of Catania, Catania, Italy; [d]Freeman Cardiothoracic Centre, Newcastle upon Tyne, UK; [e]Department of Pathology, Wythenshawe Hospital, Manchester, UK; [f]Department of Pathology, Uniklinik Koeln, Cologne, Germany; [g]Department of Pathology, Papworth Hospital, Cambridge, UK; [h]Department of Cellular Pathology, University Hospitals Birmingham and Queen Elizabeth Hospital, Birmingham, UK; [i]Department of Pathology, S. Orsola-Malpighi Hospital, Bologna, Italy; [j]Department of Pathology, University of Bari, Bari, Italy; [k]Service d'anatomie pathologique, Centre Hospitalier Universitaire Dupuytren, Limoges, France; [l]Insitut Universitaire de Pathologie de Lausanne, Lausanne, Switzerland; [m]Department of Pathology, Turkiye Yuksek Ihtisas Hospital, Ankara, Turkey; [n]Department of Pathology, Hopital Europeen Georges Pompidou, Paris, France; [o]Department of Environmental Medicine and Public Health, University of Padua, Padua, Italy; [p]Clinical Pathology Department and Information Technology Department, Azienda ULSS 18 Rovigo – Local Healthcare Authority, Rovigo, Italy; and [q]Department of Pathology , Royal Brompton and Harefield NHS Trust, London, UK.

BACKGROUND: The aim of this study was to assess, at the European level and using digital technology, the inter-pathologist reproducibility of the ISHLT 2004 system and to compare it with the 1990 system We also assessed the reproducibility of the morphologic criteria for diagnosis of antibody-mediated rejection detailed in the 2004 grading system.

METHODS: The hematoxylin–eosin-stained sections of 20 sets of endomyocardial biopsies were pre-selected and graded by two pathologists (A.A. and M.B.) and digitized using a telepathology digital pathology system (Aperio ImageScope System; for details refer to http://aperio.com/). Their diagnoses were considered the index diagnoses, which covered all grades of acute cellular rejection (ACR), early ischemic lesions, Quilty lesions, late ischemic lesions and (in the 2005 system) antibody-mediated rejection (AMR). Eighteen pathologists from 16 heart transplant centers in 7 European countries participated in the study. Inter-observer reproducibility was assessed using Fleiss's kappa and Krippendorff's alpha statistics.

RESULTS: The combined kappa value of all grades diagnosed by all 18 pathologists was 0.31 for the 1990 grading system and 0.39 for the 2005 grading system, with alpha statistics at 0.57 and 0.55, respectively. Kappa values by grade for 1990/2005, respectively, were: 0 = 0.52/0.51; 1A/1R = 0.24/0.36; 1B = 0.15; 2 = 0.13; 3A/2R = 0.29/0.29; 3B/3R = 0.13/0.23; and 4 = 0.18. For the 2 cases of AMR, 6 of 18 pathologists correctly suspected AMR on the hematoxylin–eosin slides, whereas, in each of 17 of the 18 AMR-negative cases a small percentage of pathologists (range 5% to 33%) overinterpreted the findings as suggestive for AMR.

1053-2498/$ -see front matter
doi:10.1016/j.healun.2011.05.011

KEYWORDS:
heart transplant;
endomyocardial biopsy;
acute cellular rejection;
antibody-mediated rejection;
classification;
scoring system

CONCLUSIONS: Reproducibility studies of cardiac biopsies by pathologists in different centers at the international level were feasible using digitized slides rather than conventional histology glass slides. There was a small improvement in interobserver agreement between pathologists of different European centers when moving from the 1990 ISHLT classification to the "new" 2005 ISHLT classification. Morphologic suspicion of AMR in the 2004 system on hematoxylin–eosin-stained slides only was poor, highlighting the need for better standardization of morphologic criteria for AMR. Ongoing educational programs are needed to ensure standardization of diagnosis of both acute cellular and antibody-mediated rejection.

J Heart Lung Transplant 2011;30:1214–20

Reprint requests: Annalisa Angelini, MD, Department of Medical-Diagnostic Sciences and Special Therapies, University of Padua Medical School, Via Gabelli 61, 35121 Padua, Italy. Telephone: 0039-049-827-2260. Fax: 0039-049-827-2205.
E-mail address: annalisa.angelini@unipd.it

In late 2004 an expert group of the International Society for Heart and Lung Transplantation (ISHLT), led by Dr Susan Stewart, published its long-awaited revision of the 1990 working formulation for biopsy diagnosis of acute cellular rejection.[1,2] A recommendation was that the revised system should be audited to assess reproducibility, which was shown to have been a problem with the more complex 1990 system.[3–7] Under the auspices of the Association for European Cardiovascular Pathology (AECVP), a transplant working group was established in January 2009 to facilitate collaboration between heart transplant pathologists across Europe. The aims of the working group are to facilitate contact between experts in this field, to guarantee educational and training activities using web-based technology where possible, and to promote research projects. The resulting standardization of diagnostic and therapeutic approaches would contribute to improving standards of healthcare offered to European citizens undergoing heart transplantation.

One obvious target is to ensure standardization in the use of the ISHLT 2004 revision of the 1990 working formulation for biopsy diagnosis of cardiac allograft rejection, to assess its reproducibility, and to define its strengths and weaknesses. In the 15 years since publication of the 1990 working formulation it has become apparent that there were widespread inconsistencies in the use of the grading system,[3–7] as highlighted by multicenter therapeutic trials of which central review of endomyocardial biopsies (EMBs) by an expert pathology panel was a part.[3] We therefore sought to assess, at the European level, interobserver pathologist reproducibility of the ISHLT 2004 classification system for biopsy diagnosis of cardiac allograft rejection, and to compare it with the 1990 grading scheme and thus highlight its strengths and weaknesses. We also tested reproducibility for morphologic diagnosis of antibody-mediated rejection (AMR), the criteria for which were detailed only in the 2004 grading scheme. The study was designed to be web-based using digital technology to facilitate participation, and to reduce the costs and risks of mailing slides throughout Europe and the time needed to carry out the study (the Aperio ImageScope System; for details refer to http://aperio.com/).

As we anticipated that many pathologists lacked sufficient experience with fully electronic studies such as this, and were often skeptical about reliability and quality of digital imaging systems for biopsy evaluation, a pilot study of cases selected by the study organizers was done to enable participants to familiarize themselves with the electronic skills required to simulate microscope manipulation.

Methods

A call for voluntary participation in this study was sent out through the membership circulation list of the AECVP and through personal contacts already existing between pathologists involved in European heart transplant programs. Eighteen pathologists from 16 different heart transplant centers in 7 European countries volunteered for the study. The inclusion criteria for individual pathologists were that they were:

- Working in or associated with a center with an "active" heart transplant program (i.e., their centers should still perform heart transplant and they should assess EMBs on a regular basis: 4 pathologists assessed >500 EMBs; 7 pathologists between 100 and 300 EMBs; and 7 pathologists between 30 and 80 EMBs).
- Regularly evaluating endomyocardial biopsies for allograft rejection either as the lead cardiac pathologist for their center or on a transplant biopsy shift.
- Using the new 2004 ISHLT classification.

Twenty cases pre-selected by two pathologists (M.B. and A.A.) from their respective centers were digitized using the Aperio ImageScope system. The 20 cases covered all grades of ACR, early ischemic lesions, Quilty lesions, late ischemic lesions and AMR (index diagnoses). They were all adequate with regard to number of pieces and representative myocardium and sufficient technical quality. For each case, three hematoxylin–eosin slides were used for digitalization. A response form was designed by two pathologists (M.B. and A.A.) and a link inserted in the webpage for each case in the study to collect data electronically (see response form, Figure 1) for statistical analysis. Each pathologist was given a unique code and blinded as to the source and diagnosis of each of the 20 cases.

Interobserver reliability was assessed using Fleiss's kappa statistics, a category-specific measure considered standard for assessment of interrater reliability.[8] Because Fleiss's kappa statistics were developed for nominal scales (categories), we also calculated Krippendorff's alpha statistics, which are appropriate for ordinal data (rank order) but are not category-specific.[9] For both methods, we calculated the bootstrap 95% confidence interval with the percentile method and performed 5,000 simulations.[10]

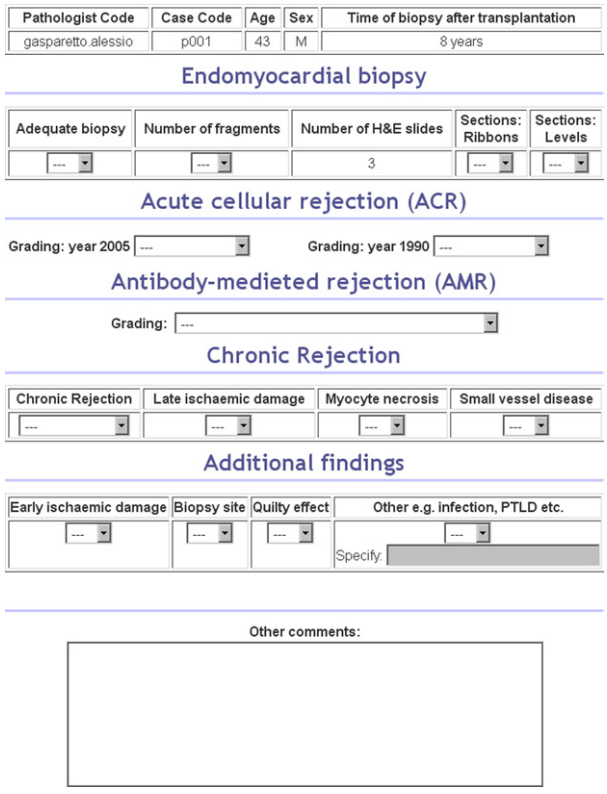

Figure 1 A template for the electronic response form.

To determine whether reproducibility in the diagnosis of ACR improves with experience we evaluated the percentage of agreement, overestimation and underestimation, according to the time on a transplant program and the volume of EMBs evaluated. The kappa score based on degree of inter-pathologist agreement was defined as: poor, <0.00; slight, 0.00 to 0.20; fair, 0.21 to 0.40; moderate, 0.41 to 0.60; substantial, 0.61 to 0.80; or almost perfect, 0.81 to 1.00.[11,12]

All analyses were conducted with SPSS (version 17.0) software.

Results

The combined kappa value of all grades of ACR diagnosed by all 18 pathologists was 0.3178 for the 1990 grading system and 0.3923 for the 2004 grading system. Kappa values for each grade of ACR according to the 1990 and 2004 classifications, respectively, are presented in Table 1. If we grouped the grades as 0+1R and 2R+3R, the kappa values would be 0.48. Taking into account the ordinal nature of the classification systems, Krippendorff's alpha gave a result of 0.5668 (bootstrap 95% confidence interval 0.5370 to 0.5957) for the 1990 system and 0.5511 (bootstrap 95% confidence interval 0.5214 to 0.5796) for the 2004 system.

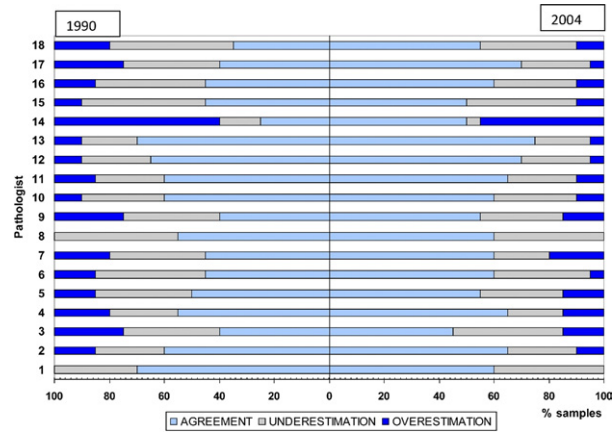

Figure 2 Percentage distribution of agreement, underestimation and overestimation in diagnoses of EMBs made by pathologists compared with the index diagnoses by referent pathologists according to the 1990 and 2004 classifications.

The distribution of agreement, overestimation or underestimation in applying the two classifications of ACR, as compared with the index diagnosis by the referent pathologists, is reported in Figure 2 for each of the 18 pathologists and in Figure 3 as mean agreement. As expected, agreement increased with the 2004 classification and showed a reduction in overestimation compared with the 1990 classification.

Agreement, overestimation or underestimation data for the two classifications of ACR according to the experience of the pathologist, measured as the number of years on a transplant program, are reported in Figure 4. Three of the pathologists were with the transplant program for between 2 and 8 years, 9 pathologists for 12 to 18 years and 6 for >25

Table 1 Fleiss's Kappa Statistics for 1990 and 2004 Classifications

	1990	2004		1990 to 2004
Grade	Kappa (95% bootstrap confidence interval)	Grade	Kappa (95% bootstrap confidence interval)	Kappa (95% bootstrap confidence interval)
0	0.5156 (0.3590 ÷ 0.7682)	0	0.5064 (0.3512 ÷ 0.7603)	0.4847 (0.5417 ÷ 1.0000)
1A	0.2353 (0.1659 ÷ 0.3624)	1R	0.3569 (0.2697 ÷ 0.5266)	
1B	0.1469 (0.0142 ÷ 0.2620)			
2	0.1277 (0.0435 ÷ 0.2033)			
3A	0.2923 (0.1451 ÷ 0.5090)	2R	0.2945 (0.1552 ÷ 0.5090)	0.4847 (0.2724 ÷ 0.6813)
3B	0.1257 (−0.0025 ÷ 0.1961)	3R	0.2319 (−0.0025 ÷ 0.3652)	
4	0.1824 (−0.0588 ÷ 0.2402)			
Combined	0.3178 (0.2081 ÷ 0.4059)	Combined	0.3923 (0.2642 ÷ 0.4899)	0.4847 (0.2477 ÷ 0.6322)

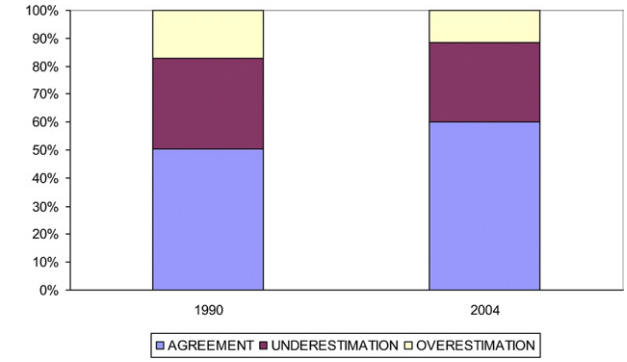

Figure 3 Percentage distribution of agreement, underestimation and overestimation by the 18 pathologists compared with the index diagnoses by referent pathologists when applying the two classifications.

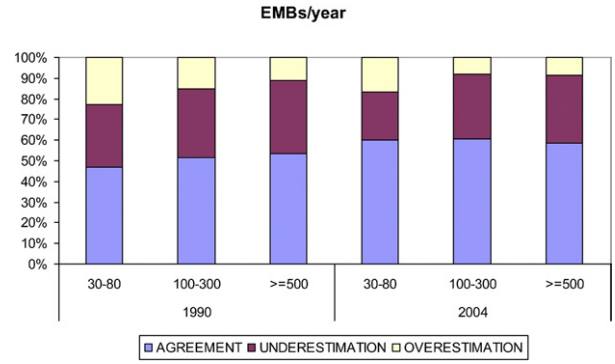

Figure 5 Percentage distribution of agreement, underestimation and overestimation of the 18 pathologists compared with the index diagnoses by referent pathologists according to number of EMBs evaluated per year when applying the two classifications.

years. A small percentage of improvement in agreement between pathologists could be observed among the more experienced pathologists. When we took into account the volume of EMBs evaluated by pathologists, similar results of increased concordance with index diagnosis by the referent pathologists were obtained (Figure 5). Seven pathologists reported between 30 and 80 EMBs/year, 7 pathologists between 100 and 300 EMBs/year and 4 pathologists >500 EMBs/year. In most transplant programs biopsy-diagnosed grade of ACR is one parameter used by clinicians to determine whether amplification of immunosuppression is needed. Therefore, we divided our cases into two groups: (a) ISHLT 2004 Grades 2R and 3R (normally needing increased immunosuppression); and (b) ISHLT 2004 Grades 0 and 1R (normally needing no modification of immunosuppression).

Re-examining our results by grouping the grades of rejection in this way showed a much higher level of agreement than for each grade of rejection, regardless of whether the 2004 and 1990 system was used. There was agreement within each group of 89.4%, underestimation (false negatives, those who did not receive enhanced immunosuppression) of 5.6% and overestimation (false positives, those who received an unnecessary increase in immunosuppression) of 5.0% (Figure 6). The cohort of cases included 2 AMRs. A total of 33% and 39% of the pathologists, respectively, correctly identified features suspicious of AMR, whereas, in 17 of the 18 cases negative for AMR, a small percentage of pathologists, ranging from 6% to 39%, suggested AMR in each case (Table 2).

Figure 7 shows the percentage distribution of correctly identified, false-negative and false-positive diagnoses expressed by pathologists compared with the index diagnosis by the referent pathologists for the identification of additional findings such as chronic ischemic damage, myocyte necrosis, small-vessel disease and early ischemic damage.

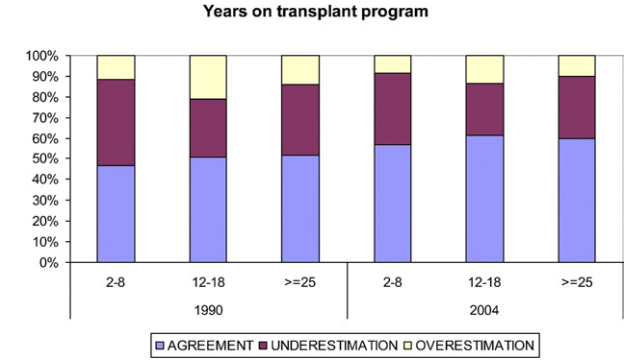

Figure 4 Percentage distribution of agreement, underestimation and overestimation of the 18 pathologists compared with the index diagnoses by referent pathologists according to years with transplant program when applying the two classifications.

Discussion

The results of our pilot study show that introduction of the 2004 classification produced only a small improvement in interobserver reproducibility of the ACR grading system among different centers in Europe. The kappa value increased from 0.31 to 0.39 for the 1990 and 2004 classifica-

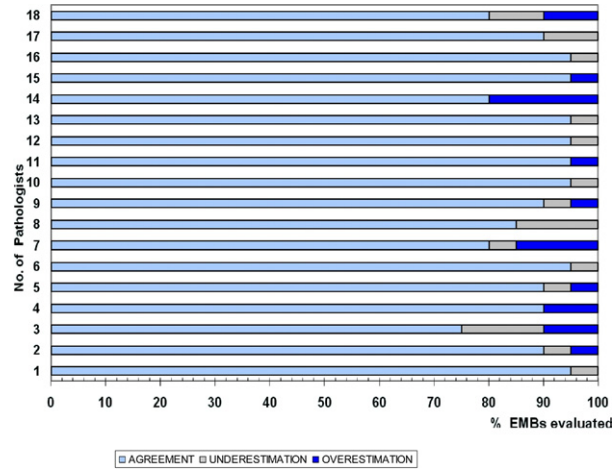

Figure 6 Percentage distribution of agreement, underestimation and overestimation in diagnoses made by pathologists compared with the index diagnoses by referent pathologists taking into account the need of immunosuppression modification.

Table 2 Distribution of Agreement Between Referent Pathologists (ACR/AMR Index Diagnosis) and 18 Pathologists for Diagnosis of AMR on Hematoxylin Eosin Staining for Each of the 20 Cases

Case code	Referent pathologists: ACR/AMR diagnosis	Pathologists diagnosis for AMR		
		Negative for AMR	Suggestive for AMR	Total
p001	0 (AMR)	11 (61%)	7 (39%)	18
p002	0/neg	17 (94%)	1 (6%)	18
p003	0/neg	15 (83%)	3 (17%)	18
p004	1A/neg	17 (94%)	1 (6%)	18
p005	3A/neg	16 (89%)	2 (11%)	18
p006	0/neg	15 (83%)	3 (17%)	18
p007	1A/neg	15 (83%)	3 (17%)	18
p008	1A/neg	14 (78%)	4 (22%)	18
p009	1B (AMR)	12 (67%)	6 (33%)	18
p010	4/neg	13 (72%)	5 (28%)	18
p011	1A/neg	13 (72%)	5 (28%)	18
p012	0/neg	17 (94%)	1 (6%)	18
p013	1A/neg	12 (67%)	6 (33%)	18
p014	1A/neg	14 (78%)	4 (22%)	18
p015	3B/neg	15 (83%)	3 (17%)	18
p016	3B/neg	11 (61%)	7 (39%)	18
p017	1A/neg	16 (89%)	2 (11%)	18
p018	2/neg	17 (94%)	1 (6%)	18
p019	1A/neg	16 (89%)	2 (11%)	18
p020	0/neg	18 (100%)	0	18
TOTAL		294 (82)	66 (18)	360 (100)

tion and, according to the Landis and Koch[11,12] guidelines for interpretation of kappa values, the strength of agreement remained within the range of fair. When taking into account the ordinal nature of the two classifications with the alpha statistic, the agreement for the two classifications was similar (1990: 0.57; 2004: 0.55) and greater when compared with the nominal version. Interestingly, the kappa values were in the range of moderate if Grades 0 and 1R (no modification of therapy required) and 2R and 3R (need for modification/adjustment of therapy) were considered together.

Although slightly disappointing these results are in keeping with the results obtained by the UCLA group in their reproducibility study in which they evaluated inter- and intraobserver reproducibility in a group of pathologists from the same center using glass slides.[13] Reproducibility of grading systems is usually tested within small groups of pathologists who have previously worked or trained together[13] and thus should be more similar than individuals who have never worked together. Our study represents the first attempt, at the European level, to assess reproducibility of the ISHLT 2004 grading scheme among pathologists at different transplant centers and with varied experience. We have confirmed the importance of performing these types of surveys at an international level to assess the worldwide variability in the application of scoring systems, and highlighted the need to improve the accuracy of diagnosis. We have shown that manipulation of digitized images of the histology slides on a standard desktop computer can be done similarly to using a microscope to assess the glass slides, thus allowing international surveys.

International studies such as ours can be performed in a reasonable time-frame and at low cost, and can foster contact and discussion on diagnostic criteria to ultimately improve reproducibility and, through this closer collaboration between units, can even improve the technical aspects of biopsy handling and staining. Telepathology is increasingly being applied worldwide, as many centers are adopting digital slide imaging acquisition systems, which are becoming cost-accessible to many pathology laboratories.[14,15] These systems can allow storage of large archives of cases and ease of retrieval. Moreover, they may find a wider application in consultations among task force members or experts for classification schemes. The virtual slides could be useful for teaching purposes as well as for external quality assurance programs to maintain high standards, improve diagnostic skills, and serve as an instrument for continuing medical education and revalidation.[16,17] The long-term economic impact of such a solution, compared with the more traditional approach with glass slides, is intuitive in terms of time, human resources and costs. Presently, the major limitation resides in the resistance of traditional pathologists to adopt such technological solutions.

As expected, the experience of pathologists, as judged by time with the transplant program and volume of cases evaluated, improves the percentage of cases correctly evaluated. This is consistent with the experience of the UCLA group[13] and with other classification systems in different areas of medicine.[18–20] Low kappa values have been reported for other reproducibility studies of renal allograft rejection[21,22] and in a hepatitis survey,[23] thus showing that the problem is not organ-specific but related to semi-quantitative classification systems. Difficulties in defining abnormal substrates in simple numeric terms are inevitable with diseases that are a continuum of observed abnormalities of patterns rather than a collection of discrete (yes/no) abnormalities.[24,25]

Our data could also be assessed according to the clinical implications of the new classification and its impact on management of patients in the follow-up period, even though we recognize that this could be pure speculation

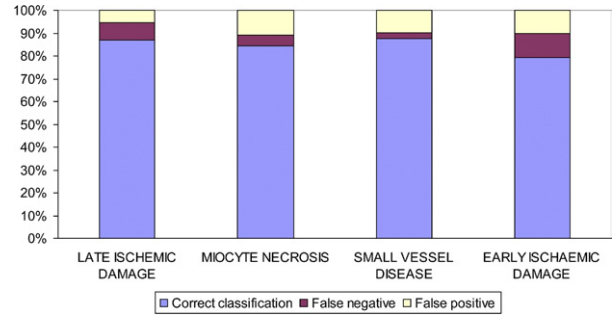

Figure 7 Percentage distribution of correct diagnoses and false-negative and false-positive diagnoses expressed by pathologists compared with the index diagnoses by referent pathologists for the identification of additional findings.

because we assumed that treatment is standardized among clinicians and this is not always the case. Based on the treatment threshold for modification of immunosuppressive therapy being Grade 2R and evaluating the agreement or disagreement between pathologists according to this cut-off, the data show that only 5% of the diagnoses were improperly graded above the threshold and would have resulted in inappropriate overtreatment of the patients. In another 6% of the diagnoses, the grading was below the expected treatment threshold, resulting in undertreatment of patients. Our results seem better than those previously reported in a UK survey[26] and by the UCLA group, which showed a tendency to overestimate diagnosis in 20% of cases. In only a small percentage of cases (3%) after further evaluation by immunohistochemistry and discussion were the diagnoses upgraded, and this could be considered the percentage of underestimation.[13]

There has been considerable debate on the use of histologic criteria as a trigger to further histochemical and immunologic evaluation for the diagnosis of AMR.[27–32] Our results confirm that morphology on hematoxylin–eosin is not specific for a diagnosis of AMR,[32] and routine use at set time intervals of immunohistochemical investigation of biopsies and assessment of serum for donor-specific antibodies, as concluded by the recent ISHLT review of AMR, is important.[33]

Our results relating to chronic rejection, detectable as late ischemic damage, myocyte necrosis and small-vessel disease, and to non-rejection findings, such as early ischemic damage, showed agreement in >80% of the cases. Use of quantification tools such as morphometry or immunohistochemistry as an aid for detection of early structural cytoplasmic and nuclear abnormalities could improve our ability to diagnose inflammation and myocyte injury, thus further reducing interobserver variability.

We emphasize that our results, although preliminary and requiring further and larger studies, indicate that "virtual" microscopy is feasible as a tool for international collaboration and standardization in this challenging area of biopsy diagnosis.

In conclusion, our preliminary study has shown that reproducibility studies among different centers at the international level are feasible using digital technology and the data are comparable with microscopy using glass slides. There was a small improvement, although not statistically significant, in interobserver agreement between pathologists of the different European centers when moving from 1990 ISHLT classification to the 2004 ISHLT "new" classification and with increasing experience, as evidenced by the numbers of biopsies evaluated annually and the years with a transplant program. Morphologic identification of features suspicious for AMR, as defined in the 2004 system, was poor. Better definition of the morphologic criteria for diagnosis of AMR on EMBs is urgently required. There is a need for continuing educational programs, both locally and internationally, to standardize the diagnosis of acute cellular rejection and antibody-mediated rejection.

Disclosure statement

The authors express sincere gratitude to the General Director of the Azienda Sanitaria of Rovigo, Dr Adriano Marcolongo, who supported our study and offered use of the telepathology system. We also acknowledge the patronage of the Association for European Cardiovascular Pathology, which facilitates the European network and collaboration between heart transplant pathologists across Europe. The authors have no conflicts of interests to disclose.

References

1. Stewart S, Winters GL, Fishbein MC, et al. Revision of the 1990 working formulation for the standardization of nomenclature in the diagnosis of heart rejection. J Heart Lung Transplant 2005;24:1710-20.
2. Billingham ME, Cary NR, Hammond ME, et al. A working formulation for the standardization of nomenclature in the diagnosis of heart and lung rejection. J Heart Transplant 1990;9:587-93.
3. Winters GL, McManus BM. Consistencies and controversies in the application of the International Society for Heart and Lung Transplantation working formulation for heart transplant biopsy specimens. J Heart Lung Transplant 1996;15:728-35.
4. Winters GL. The challenge of endomyocardial biopsy interpretation in assessing cardiac allograft rejection. Curr Opin Cardiol 1997;12:146-52.
5. Fishbein MC, Bell G, Lones MA, et al. Grade 2 cellular heart rejection: does it exist? J Heart Lung Transplant 1994;13:1051-7.
6. Marboe CC, Billingham M, Eisen H, et al. Nodular endocardial infiltrates (Quilty lesions) cause significant variability in diagnosis of ISHLT Grade 2 and 3A rejection in cardiac allograft recipients. J Heart Lung Transplant 2005;24(suppl):S219-26.
7. Milano A, Caforio AL, Livi U, et al. Evolution of focal moderate (International Society for Heart and Lung Transplantation grade 2) rejection of the cardiac allograft. J Heart Lung Transplant 1996;15:456-60.
8. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378-82.
9. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Commun Methods Meas 2007;1:77-89.
10. Efron B, Tibshirani RJ. Bootstrap measures for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1986;1:54-77.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
12. Koch GG, Landis JR, Freeman JL, et al. A general methodology for the analysis of experiments with repeated measurement of categorical data. Biometrics 1977;33:133-58.
13. Yang H-M, Laia CK, Gjertsona DW, et al. Has the 2004 revision of the International Society of Heart and Lung Transplantation grading system improved the reproducibility of the diagnosis and grading of cardiac transplant rejection? Cardiovasc Pathol 2009;18:198-204.
14. LeRouge C, Tulu B, Forducey P. The business of telemedicine: strategy primer. Telemed J 2010;16:898-909.
15. Krupinski EA, Tillack AA, Richter L, et al. Eye-movement study and human performance using telepathology virtual slides: implications for medical education and differences with experience. Hum Pathol 2006;37:1543-56.
16. Kumar RK, Velan GM, Korell SO, et al. Virtual microscopy for learning and assessment in pathology. J Pathol 2004;204:613-8.
17. Sergi C, Mikuz G. External quality assurance as a revalidation method for pathologists in pediatric histopathology: comparison of four international programs. BMC Clin Pathol 2008;8:11-9.
18. El-Husseiny M, Coleman N. Inter- and intra-observer variation in classification systems for impending fractures of bone metastases. Skeletal Radiol 2010;39:155-60.

19. Cross SS. Grading and scoring in histopathology. Histopathology 1998;33:99-106.
20. Silcocks PB. Some issues in observer error studies in pathology. J Pathol 1992;168:255-6.
21. Marcussen N, Olsen TS, Benediktsson H, et al. Reproducibility of the Banff classification of renal allograft pathology. Inter- and intraobserver variation. Transplantation 1995;60:1083-9.
22. Gibson IW, Gwinner W, Brocker V, et al. Peritubular capillaritis in renal allografts: prevalence, scoring system, reproducibility and clinicopathological correlates. Am J Transplant 2008;8:819-25.
23. Netto GJ, Watkins DL, Williams JW, et al. Interobserver agreement in hepatitis C grading and staging and in the Banff grading schema for acute cellular rejection: the "Hepatitis C 3" Multi-institutional Trial experience. Arch Pathol Lab Med 2006;130:1157-62.
24. Furness P, Taub N, Assmann K, et al. International variation in histologic grading is large, and persistent feedback does not improve reproducibility. Am J Surg Pathol 2003;27:805-10.
25. Reed EF, Demetris AJ, Hammond E, et al. Acute antibody mediated rejection of cardiac transplantations. J Heart Lung Transplant 2006;25:153.
26. Lindop GBM, Burke MM, Ogston S, et al. Interobserver variability in grading acute rejection in endomyocardial biopsies. J Heart Lung Transplant 2003;22(suppl):S97.
27. Burke M, Andersen C, Ashworth M, et al. C4d methodology and interpretation in biopsy diagnosis of cardiac antibody-mediated rejection: a European survey from the transplant working group of the Association for European Cardiovascular Pathology (AECVP). J Heart Lung Transplant 2010;29(suppl):37-8.
28. Wu GW, Kobashigawa JA, Fishbein MC, et al. Asymptomatic antibody-mediated rejection after heart transplantation predicts poor outcomes. J Heart Lung Transplant 2009;28:417-22.
29. Rodriguez ER, Skojec DV, Tan CD, et al. Antibody-mediated rejection in human cardiac allografts: evaluation of immunoglobulins and complement activation products C4d and C3d as markers. Am J Transplant 2005;5:2778-85.
30. Tan CD, Sokos GG, Pidwell DJ, et al. Correlation of donor-specific antibodies, complement and its regulators with graft dysfunction in cardiac antibody-mediated rejection. Am J Transplant 2009;9: 2075.
31. Fedrigo M, Gambino A, Tona F, et al. Can C4d immunostaining on endomyocardial biopsies be considered a prognostic biomarker in heart transplant recipients? Transplantation 2010;90:791-8.
32. Smith RN, Brousaides N, Grazette L, et al. C4d deposition in cardiac allografts correlates with alloantibody. J Heart Lung Transplant 2005; 24:1202-10.
33. Kobashigawa J, Crespo-Leiro MG, Ensminger SM, et al. Report from a consensus conference on antibody-mediated rejection in heart transplantation J Heart Lung Transplant 2011;30:252-69.