Concordance Among Pathologists in the Second Cardiac Allograft Rejection Gene Expression Observational Study (CARGO II)

Maria G. Crespo-Leiro,[1] Andreas Zuckermann,[2] Christoph Bara,[3] Paul Mohacsi,[4] Uwe Schulz,[5] Andrew Boyle,[6] Heather J. Ross,[7] Jayan Parameshwar,[8] Michael Zakliczyński,[9] Roberto Fiocchi,[10] Joerg Stypmann,[11] Daniel Hoefer,[12] Hans Lehmkuhl,[13] Mario C. Deng,[14] Pascal Leprince,[15] Gerald Berry,[16] Charles C. Marboe,[14] Susan Stewart,[8] Henry D. Tazelaar,[17] Helen M. Baron,[18] Ian-Charles Coleman,[1] and Johan Vanhaecke[19]

Background. There has been no large evaluation of the ISHLT 2004 acute cellular rejection grading scheme for heart graft endomyocardial biopsy specimens (EMBs).

Methods. We evaluated agreement within the CARGO II pathology panel and between the panel (acting by majority) and the collaborating centers (treated as a single entity), regarding the ISHLT grades of 937 EMBs (with all grades ≥2R merged because of small numbers).

Results. Overall all-grade agreement was almost 71% both within the panel and between the panel and the collaborating centers but, in both cases, was largely because of agreement on grade 0: for the average pair of pathologists, fewer than a third of the EMBs assigned grade ≥2R by at least one were assigned this grade by both.

Conclusion. The 2004 revision has done little to improve agreement on the higher ISHLT grades. An EMB grade ≥2R is not by itself sufficient as a basis for clinical decisions or as a research criterion. Steps should be taken toward greater uniformity in EMB grading, and efforts should be made to replace the ISHLT classification with diagnostic criteria—EMB based or otherwise—that correspond better with the pathophysiology of the transplanted heart.

Keywords: Biopsy, Gene expression, Heart transplantation, Pathology, Rejection.

Immunosuppression after organ transplantation should be continually adapted to the individual and varying needs of a particular organ recipient. Current methods for monitoring immunosuppression—measuring serum levels of immunosuppressive drugs, noting the occurrence of side effects, looking for functional or morphologic evidence of rejection—seem to afford only a crude approximation that fails to allow adequate discrimination between underimmunosuppression and overimmunosuppression. After heart transplantation, the main criterion of underimmunosuppression is currently the detection of acute cellular rejection (ACR) by evaluation of the extent and distribution of lymphocytic infiltrates and associated alterations in tissue samples taken by endomyocardial biopsy (EMB) (1). However, a major weakness of this criterion lies in the large subjective component of the evaluation procedure. Improvement of concordance among pathologists was one of the aims of the International Society for Heart and Lung Transplantation (ISHLT) in its 2004 simplification of its 1990 classification (2, 3), but discrepant evaluations are still common, and there has been no large-scale evaluation of concordance among pathologists using the 2004 scheme. That EMB grading cannot by itself serve as the sole measure of immunosuppressive adequacy has been pointed out by Baran et al. (4), among others.

A noninvasive technique that is more objective than EMB grading is gene expression profiling of peripheral blood mononuclear cells (5–7). Further evidence for the clinical performance of a test using this technique is expected to emerge from the recently completed second Cardiac Allograft Rejection Gene Expression Observational Study (CARGO II) (http://clinicaltrials.gov/ct2/show/NCT00761787). However, like other studies in which rejection is a criterion variable, studies evaluating novel tests of rejection face a major problem in interrater disagreement on EMB grades, the reference method ("gold standard") for evaluation of ACR. Such disagreement has, on occasions, led to crucial ambiguity in study outcomes (e.g., to discrepancy in the significance of differences between two study arms (8)), and in the CARGO study (the predecessor of CARGO II), the three central pathologists achieved unanimous agreement on only 23% of EMB specimens assigned ISHLT 1990 grades 3A or 3B by the pathologists of the centers in which they were obtained (9). In evaluating a new technique, this difficulty must be dealt with by basing evaluation on consensus or majority EMB grading, which is labor intensive and operationally complex.

In CARGO II, hematoxylin-eosin–stained EMB slides provided by the collaborating centers were independently reread by members of a central panel of pathologists. Here, we report concordance in EMB grading among the members of the panel and between the panel as a whole (considered as acting by majority) and the collaborating centers (considered as a single entity).

RESULTS

Table S1 (see **SDC,** http://links.lww.com/TP/A716) summarizes the characteristics of the 462 patients who provided EMBs included in the present study. Among the 937 EMBs they provided, 425 (45.4%) were locally assigned grade 0, 439 (46.9%) grade 1R, and 73 (7.8%) a grade ≥2R.

Agreement Among Panel Members

In what follows, the four panel members are anonymously denoted P1, P2, P3 and P4. **Table S2** (see **SDC,** http://links.lww.com/TP/A716),which lists the numbers of EMBs assigned each grade by each pathologist, shows that P1 tended to award higher grades than the others. The discrepancy between P1 and the others is also reflected in the agreement indices listed in Table 1 (except that P2, who assigned fewer EMBs to grade ≥2R than any other panel member, had even lower average agreement than P1 for this grade). The estimated 95% confidence intervals of the pairwise agreement indices, generally ±4–6%, show the precision of these data to be sufficient for the purposes of this study, even in the case of positive agreement on grade ≥2R, for which confidence intervals up to 36% wide do little to

TABLE 1. Positive grade-specific agreement, negative grade-specific agreement, and overall (all-grade) agreement between panel pathologists on the ISHLT 2004 grade of EMB samples

Pathologists	Grade 0R		Grade 1R		Grade ≥2R		Overall
	PA	NA	PA	NA	PA	NA	
P1–P2	60.8 (55.3–66.1)	58.3 (52.7–63.8)	40.7 (35.1–46.4)	56.4 (51.5–61.3)	17.6 (9.5–26.6)	88.1 (85.3–90.8)	64.8 (60.6–69.0)
P1–P3	57.4 (52.1–62.8)	52.7 (47.2–58.4)	40.6 (34.8–46.6)	58.3 (53.4–63.4)	27.4 (16.2–38.9)	90.2 (87.4–92.9)	64.6 (60.5–69.0)
P1–P4	57.3 (51.9–62.7)	53.8 (47.9–59.4)	43.7 (37.7–49.5)	57.6 (52.5–62.7)	28.0 (15.7–40.5)	92.2 (89.6–94.5)	66.0 (61.7–70.2)
P2–P3	71.3 (66.6–75.7)	60.6 (54.8–66.3)	53.7 (47.7–59.6)	70.1 (65.6–74.6)	21.2 (8.0–35.5)	95.0 (93.1–96.8)	76.5 (72.9–79.9)
P2–P4	68.6 (64.1–73.3)	55.3 (49.4–61.5)	50.4 (44.2–56.8)	69.0 (64.6–73.6)	37.0 (19.2–55.6)	96.7 (95.1–98.1)	75.3 (71.6–79.1)
P3–P4	71.6 (66.9–76.1)	57.9 (51.5–64.1)	53.0 (46.3–59.6)	72.4 (67.9–76.8)	34.5 (17.4–52.4)	96.0 (94.3–97.7)	77.3 (73.6–81.0)
P1	58.5	55.0	41.6	57.4	23.7	90.1	65.1
P2	67.1	58.1	47.9	65.1	22.4	93.3	72.2
P3	67.1	56.9	48.8	67.0	27.4	93.8	73.0
P4	66.2	55.5	48.7	66.5	32.1	95.0	73.0
Whole panel	64.8	56.4	46.6	64.0	25.8	93.1	70.8

Top panel: pairwise agreement for each pair of pathologists. Middle panel: average agreement of each pathologist with the other three. Bottom panel: average over all pathologist pairs. Agreement values are expressed as percentages; in parentheses, estimated 95% confidence intervals.

PA, positive agreement; NA, negative agreement.

hide the poor performance on this grade. The salient feature of the data of Table 1 is the increasing discrepancy between positive and negative agreement values as grades increase.

Agreement Between Local Center and Panel Gradings

Table 2 shows, for each set of EMBs assigned a given grade by the local center, its distribution by panel-assigned grade, together with the number for which the panel failed to reach a majority decision ("ungraded"). It may be noted that the panel reached a majority agreement on all but 0.9% of EMBs that were locally assigned grade 0, but failed to reach agreement on 4.1% of local grade 1Rs and 19.2% of local grade ≥2Rs. Also, there was a general trend for the panel to assign lower grades than local centers (except, of course, for grade 0): 35% of local grade 1Rs and 52% of local grade ≥2Rs were given lower grades by the panel, while only 14% of local grade 0s and 3% of local grade 1Rs were given higher grades by the panel. Table 3 lists the corresponding formal agreement results. That the estimated 95% confidence intervals are somewhat narrower than for the within-panel data of Table 1 is attributable to the larger number of cases, each pair of panel members having only seen about 500 EMBs. Note that the discrepancy between positive and negative agreement values increases with grade in the same way as between panel members, as does negative agreement itself.

DISCUSSION

The above results confirm that EMB grading is far from being a totally objective procedure, if "objective" is understood to mean that a given EMB sample will be given the same grade by any competent, appropriately trained pathologist. Overall pairwise agreement between panel members regarding ISHLT 2004 grades ranged from 65% to 77%, while agreement in this respect between the panel and the local centers, 70.7%, was virtually identical to overall within-panel agreement, 70.8%. These figures are nevertheless within the range reported for a number of other areas of medicine, such as ankle fracture classification (*10*), emergency radiography (*11*), or the evaluation of patients for antibody treatment of lung cancer (*12*).

In the CARGO study (*9*), overall within-panel concordance was 66.4% when calculated as in the present study. The lower value is expected, because the pathology substudy of that study considered four grades (ISHLT 1990 grades 0, 1A plus 1B, 2, and ≥3A) as against the three employed here. That the difference with respect to the 70.8% of the present

TABLE 3. Agreement (%) between panel and local centers on ISHLT 2004 grades

Grade	PA	NA
0R	62.1 (58.1–66.1)	58.7 (54.5–62.9)
1R	50.0 (45.5–54.5)	61.0 (57.2–64.8)
≥2R	28.4 (18.4–38.8)	94.0 (92.3–95.5)
Overall	70.7 (67.7–73.7)	

In parentheses, estimated 95% confidence intervals.
PA, positive agreement; NA, negative agreement.

study is so small indicates that the 2004 revision of the ISHLT grading scheme has done little if anything to improve overall concordance.

Comparison of the bottom row of Table 1 with Table 3 shows that the within-panel and panel-local agreement results are similar not only for overall agreement but also for each grade, in spite of the clear tendency in Table 2 for the panel to assign lower grades than the local centers. This suggests that the tendency to assign lower grades was due to a difference in effective criteria rather than to better or worse application of criteria. Such a difference in criteria may possibly have resulted from the panel members being unable to use additional information on the patient's history, and this relative lack of information may be seen as a flaw in the study design. It is nevertheless interesting, if this possible lack of information was indeed responsible for lower rating by the panel, that the additional information available to local centers was associated with a general tendency for these centers to assign *higher* grades than the panel. This would suggest that pathologists may effectively tend to use the ISHLT histological grading criteria to establish a lower limit for their grade assignation, and use additional information to raise the histologically assigned grade but not, or only quite rarely, to lower it. Another possible explanation for generally lower grading by the panel is the panel's lack of clinical responsibility for the study patients, which may have freed them from the temptation to prefer a false positive to a false negative.

Both within the panel, and between panel and local centers, positive agreement decreased with increasing grade from 62–65% to 26–28%, while negative agreement increased from 56–59% to 93–94%. The pathologists showed moderate agreement both as to which EMBs were grade 0 and which were not, but whereas EMBs classified as not of grade ≥2R by one pathologist were highly likely to be similarly classified by others, an EMB graded ≥2R by one pathologist was quite unlikely to receive this grade from a given second pathologist. Thus, discouragingly, ISHLT grading is least objective when trying to fulfill one of its main purposes (detection of significant ACR), and most objective when trying to prove a negative result (absence of significant ACR), a goal for which it is unsuited because of the possibility of damaged tissue being missed at biopsy.

In principle, the discrepancies noted above should not have been influenced by any antibody-mediated rejection (AMR). AMR and ACR are sufficiently clearly distinguished in the 2004 ISHLT recommendations, and the present analysis concerned only ACR. Nevertheless, it cannot absolutely be

TABLE 2. Agreement between panel and local centers on ISHLT 2004 grades (numbers of EMBs)

Panel grade	Local grade		
	0	1R	≥2R
0	363	153	11
1R	58	253	27
≥2R	0	15	21
Ungraded	4	18	14

ruled out that there may have been cases in which the presence of AMR may have hampered ACR grading, and that these may have influenced the results.

The fact that on average any pathologist assigns grades <2R to over half the EMB samples that another pathologist would classify as grade ≥2R must contribute to reported apparent anomalies in patient outcome. For example, it has been observed that among patients with rejection and severely compromised hemodynamics, an initial rejection grade <3A (2R) is associated with a worse 2-year survival rate than a grade ≥3A (2R) (13). Such contradictions highlight the absence of a solid pathophysiological basis for the ISHLT EMB classification (14). Unless and until EMB samples can be classified with regard to ACR in a physiologically more meaningful way, doubt attends both their validity for evaluation of new therapies - in this field their exclusive use should probably be replaced by their joint consideration together with other factors determining long-term outcome (4) - and their role in long-term follow-up (15). An EMB classification with greater physiological correlation will no doubt require the use of immunohistochemical, genomic and/or proteomic methods (16).

One of the weaknesses of this study, asymmetry of information between panel and local centers, has already been discussed above. A related limitation was the inability of the panel pathologists to seek further information by requesting additional procedures such as the cutting of additional sections or the use of additional stains; had panel members been able to obtain this information they might have achieved a greater degree of agreement (although this does not invalidate the agreement results in regard to the EMB material actually seen).

A further weakness of the study, perhaps its greatest, was the restriction of the number of EMBs that were locally assigned ISHLT 1990 grades of 0 or 1A. Although the set of EMB specimens provided to the panel was appropriate for the main goals of CARGO II, the sample of 937 EMBs used in this pathology substudy was not quantitatively representative of the general population of HT EMB specimens, and the observed concordance may accordingly be slightly biased.

In conclusion, this study has shown that the pathologists constituting the central pathology panel of the CARGO II study tended to assign lower ISHLT ACR grades to hematoxylin-eosin–stained EMB samples than did local pathologists, but within-panel agreement was no better than agreement between their majority decision and the grade assigned at the centers of origin of the samples. Overall (all-grade) agreement using the 2004 ISHLT classification showed little if any improvement with respect to a previous evaluation of the 1990 classification (9), although it was similar to those reported in many other medical fields. Moreover, the fair degree of overall agreement was due very largely to agreement on grade 0 specimens: for the average pair of pathologists, fewer than a third of EMBs assigned grade ≥2R by at least one were assigned this grade by both. In keeping with two recent small studies (17, 18), these findings suggest that, although the 2004 revision simplified the ISHLT classification, it has done little to improve agreement on the upper grades. Greater uniformity in EMB interpretation should be sought in the short term by means of workshops, "clinics," or other such events and in the long term by research to identify EMB characteristics that more accurately reflect the physiology of the transplanted heart. Because, in practice, EMB findings are not the sole basis for clinical decisions but are used together with other information (especially hemodynamic parameters), it is also possible that benefit for clinical practice and/or research methodology might accrue from the formal integration of these various factors in a single ACR grading system or, more generally, in a classification encompassing both ACR and AMR.

MATERIALS AND METHODS

Patients, Samples, and Sample Evaluation

Between January 2005 and November 2008, 754 heart transplant recipients were enrolled in the CARGO II study at 17 centers in Austria (2), Belgium, Canada, France, Germany (4), Italy, Poland, Spain, the United Kingdom, the USA (3) and Switzerland (hereinafter the "local centers") after giving informed consent to the performance of EMB as part of routine care, the collection of blood samples for research purposes, and the use of clinical data. The study protocol was approved by the institutional review board or ethics committee of each center. Patients could be enrolled at any time after transplantation, and received the standard of care at the institution treating them, the study requiring no additional intervention. No patients were excluded on medical grounds.

Within the study period, 4,767 EMB samples were obtained from 631 of the enrolled patients. These samples were included in the main CARGO II analyses only if they fulfilled the following conditions: 1) a blood sample had been properly collected in accordance with the CARGO II protocol during the clinical visit in which the EMB was performed; 2) the patient had not experienced an episode of rejection (defined by a locally assigned grade ≥3A (2R) or record of rejection treatment and/or augmentation of the immunosuppressive regimen in the case report form) within the 60 days before EMB and had not been receiving 20 mg or more of oral corticosteroids in the 21 days before EMB; and 3) EMB had been performed at least 55 days after transplantation, at least 21 days after any therapy for ACR, and at least 30 days after any transfusion of red blood cells. EMBs fulfilling these conditions were included in the present study of concordance among pathologists only if they also fulfilled the following additional conditions: 4) an ISHLT 1990 grade assigned to the EMB by the pathologist at the local center of origin was available; 5) they were not samples of locally assigned grades 0 or 1A from patients from whom a sample of the same grade had already been included in the study (this criterion was applied to prevent the study from being overloaded with the highly prevalent grade 0 and grade 1A samples); 6) all hematoxylin-eosin–stained slides used for evaluation by the local center pathologist were sent to the study organizers for blind grading by an independent central panel (hereinafter "the panel") composed of the pathologists G. Berry, C. Marboe, S. Stewart, and H. Tazelaar; and 7) grades were obtained from at least two of the three panel members to whom each EMB was sent for independent evaluation. These seven conditions were satisfied by 937 EMBs from 462 patients.

Evaluation of Concordance

In evaluating concordance between the central panel and local centers (these latter considered as a single entity), ISHLT 1990 grades were mapped onto ISHLT 2004 grades in the usual way, and grades 2R and 3R were merged as "grade ≥2R" because of small numbers of cases. Each EMB was considered to have been awarded an ISHLT grade by the panel if at least two panel members awarded it the same grade (hereinafter "the panel grade"); otherwise it was considered as ungraded. The concordance between panel and local center, and pairwise concordances between panel members, were assessed in terms of specific positive and negative agreement for each ISHLT grade on a per EMB basis, that is, positive agreement on grade x was calculated as the proportion of the EMBs seen by both raters and assigned

grade *x* by at least one, that were actually assigned that grade by both; and negative agreement as the proportion of those assigned some other grade by at least one that were actually assigned grade *x* by neither (*19*). Both positive and negative agreements were calculated to avoid the possibly misleading effects of large differences in grade prevalences. In addition, pairwise overall agreement within each pair of panel pathologists, or between panel and local center assignments, was calculated as the proportion of the EMBs seen by both to which they both assigned the same grade; the average grade-specific and overall concordance of a given panel pathologist P*i* with the other three was calculated; and the grade-specific and overall internal agreement of the whole panel was determined as the weighted average of those of the four pathologists.

Statistical Analyses

Estimated 95% confidence intervals for positive and negative agreement were obtained using a percentile bootstrap method. All statistical calculations were performed using R.

ACKNOWLEDGMENTS

1. CARGO II Principal Investigators and Study Sites

Maria G. Crespo-Leiro, Hospital Universitario A Coruña, La Coruña, Spain

Johan Vanhaecke, Cardiovascular Diseases, University Hospitals Leuven, and Department of Cardiovascular Sciences, KU Leuven, Belgium

Andreas Zuckermann, Medical University of Vienna, Vienna, Austria

Christoph Bara, Hannover Medical School, Hannover, Germany

Paul Mohacsi, University Hospital Bern, Bern, Switzerland

Uwe Schulz, Ruhr University of Bochum, Bad Oeynhausen, Germany

Roberta C. Bogaev, Texas Heart Institute at St. Luke's Episcopal Hospital, Houston, TX

Andrew Boyle, University of Minnesota, Minneapolis, MN

Heather Ross, Toronto General Hospital, Toronto, Canada

Jayan Parameshwar, Papworth Hospital, Cambridge, United Kingdom

Michal Zakliczyński, Silesian Center for Heart Disease, Zabrze, Poland

Roberto Fiocchi, Ospedali Riuniti di Bergamo, Bergamo, Italy

Joerg Stypmann, Department of Cardiovascular Medicine, Division of Cardiology, University Hospital Münster, Münster, Germany

Daniel Hoefer, Innsbruck Medical University, Innsbruck, Austria

Hans Lehmkuhl, Deutsches Herzzentrum, Berlin, Germany

Mario C. Deng, Columbia University Medical Center, New York, NY

Pascal Leprince, Groupe Hospitalier Pitié-Salpêtrière, Paris, France

2. CARGO II Study Pathologists and Institutions

Gerald Berry, Stanford University Medical Center, Stanford, CA

Charles C. Marboe, College of Physicians and Surgeons of Columbia University, New York, NY

Susan Stewart Papworth Hospital Pathology Department, Papworth Everard, Cambridge, UK

Henry D Tazelaar, Department of Laboratory Medicine and Pathology, Mayo Clinic, Scottsdale, AZ

3. XDx, study sponsor

Helen M. Baron, Michael Brown, Kenneth C. Fang, Tod M. Klingler, Hans J. Wolters, Upen Patil, Debbie Pierretti, Joyce Alejo-Stone, Weiyao Liu, Nitzan Sternheim

REFERENCES

1. Billingham M. Diagnosis of cardiac rejection by endomyocardial biopsy. *J Heart Transplant* 1982; 1: 25.
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: Heart Rejection Study Group. The International Society for Heart Transplantation. *J Heart Transplant* 1990; 9: 587.
3. 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.
4. Baran DA, Taylor DO, Kobashigawa JA. Immunosuppression and heart transplantation: how do we define success? *Am J Transplant* 2010; 10: 205.
5. Horwitz P, Tsai E, Putt M, et al. Detection of cardiac allograft rejection and response to immunosuppressive therapy with peripheral blood gene expression. *Circulation* 2004; 110: 3815.
6. Deng MC, Eisen HJ, Mehra MR, et al. Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *Am J Transplant* 2006; 6: 150.
7. Pham MX, Teuteberg JJ, Kfoury AG, et al. Gene-expression profiling for rejection surveillance after cardiac transplantation. *N Engl J Med* 2010; 362: 1890.
8. Grimm M, Rinaldi M, Yonan N, et al. Superior prevention of acute rejection by tacrolimus vs. cyclosporine in heart transplant recipients—a large European trial. *Am J Transplant* 2006; 6: 1387.
9. Marboe C, 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: S219.
10. Malek I, Machani B, Mevcha A, et al. Inter-observer reliability and intra-observer reproducibility of the Weber classification of ankle fractures. *J Bone Joint Surg* 2006; 88-B: 1204.
11. Robinson P, Wilson D, Coral A, et al. Variation between experienced observers in the interpretation of accident and emergency radiographs. *Br J Radiol* 1999; 72: 323.
12. Barlési F, Balleyguier C, Besse B, et al. Inter- and intraobserver consistency in assessing eligibility for bevacizumab (BVZ) in non-small-cell lung cancer (NSCLC) patients with centrally located tumors. *Ann Oncol* 2010; 21: 1682.
13. Mills R, Naftel D, Kirklin J, et al. Heart transplant rejection with hemodynamic compromise: a multiinstitutional study of the role of endomyocardial cellular infiltrate. Cardiac Transplant Research Database. *J Heart Lung Transplant* 1997; 16: 813.

14. Marboe CC. Editorial: the molecular phenotype of heart transplant biopsies. *Am J Transplant* 2010; 10: 1957.

15. Mehra M, Parameshwar J. Gene expression profiling and cardiac allograft rejection monitoring: Is IMAGE just a mirage? *J Heart Lung Transplant* 2010; 29: 599.

16. Holweg CT, Potena L, Luikart H, et al. Identification and classification of acute cardiac rejection by intragraft transcriptional profiling. *Circulation* 2011; 123: 2236.

17. Yang H-M, Lai CK, Gjertson 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? *Cardiovascular Pathology* 2009; 18: 198.

18. Angelini A, Andersen CB, Bartoloni G, et al. 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. *J Heart Lung Transplant* 2011; 30: 1214.

19. Grant J. The fetal heart rate trace is normal, isn't it? *The Lancet* 1991; 337: 215.