Published in final edited form as:

J Heart Lung Transplant. 2019 March ; 38(3): 327–328. doi:10.1016/j.healun.2018.12.002.

Inter-rater agreement in the diagnosis of chronic lung allograft dysfunction following lung transplantation

Cynthia E. Fisher¹, Siddhartha G. Kapnadak², Erika D. Lease², Jeffrey D. Edelman², and Ajit P. Limaye¹

¹Division of Allergy and Infectious Diseases, University of Washington, Seattle, WA

²Division of Pulmonary, Critical Care, and Sleep Medicine, University of Washington, Seattle, WA

Chronic lung allograft dysfunction (CLAD) affects over 50% of lung transplant recipients (LTR) by five years and remains the primary reason for graft failure¹. It has been increasingly recognized that CLAD is a heterogeneous diagnosis with variable prognosis and different phenotypes: bronchiolitis obliterans syndrome (BOS) and restrictive allograft syndrome (RAS). Precise diagnosis of CLAD remains difficult: There are no clinically-validated biomarkers, and diagnosis can be subjective, hinging on the correct interpretation of pulmonary function tests (PFTs) coupled with exclusion of clinical confounders. Given the inherent subjectivity, there may be significant variability in both the designation of CLAD subtype and the time of onset. This has important implications both clinically and in research, where the accuracy and consistency of CLAD endpoints are crucial.

To our knowledge there is only a single study evaluating inter-rater agreement of CLAD diagnosis². In that study, Kapila, *et al.* retrospectively analyzed data collected separately for the Assessment of Immunosuppressive Regimen in Suppressing Acute and Chronic Rejection (AIRSAC) trial, for which three reviewers evaluated 181 LTR for the presence of BOS^{2,3}. The authors reported an 81% initial concordance between at least two of the three reviewers. They next randomly selected 5 out of the 35 discordant cases, and distributed serial forced expiratory volume in one second (FEV1) values and brief clinical scenarios to members of the International Society for Heart & Lung Transplantation (ISHLT) via a web-based platform. Among the 87 respondents (response rate 47%) they found a 70% inter-observer agreement (average of 5 cases, range:60%–83%) of presence or absence of BOS, but only a 41% agreement regarding the time of onset (range:21%–70%).

There were several limitations to this important study, including the use of older 2001 guidelines, assessment for only BOS (and not RAS), evaluation of BOS onset date agreement in only five patients, and the calculation of percent agreements rather than a more robust kappa statistic that accounts for expected chance agreement. Additionally, respondents in the web-based portion of the study were only given access to serial FEV1

Reprints and Correspondence: Ajit P. Limaye, MD, University of Washington, Box 356174, 1959 NE Pacific St, Seattle, WA 98195-6714, Tel: 206 598 1041 FAX: 206 598 6662, limaye@u.washington.edu. **Alternate Corresponding Author:** Cynthia Fisher, MD MPH, University of Washington, Box 356175, 1959 NE Pacific St, Seattle, WA 98195-6714, Tel: 206 598 9149 FAX: 206 598 6662, celaine@uw.edu.

C.E.F., S.G.K., J.D.E., and A.P.L. have no conflicts of interest to disclose.

values and basic descriptions of potential confounders that were already abstracted by the previous reviewers. Finally, by only evaluating five of the more difficult, discordant, cases, the authors acknowledged that the survey was not designed to reflect the true overall concordance rate of BOS diagnosis by physicians; instead, this allowed them to focus on identifying specific issues that reviewers found most difficult.

To address some of these limitations, we retrospectively reviewed serial spirometry, clinical, laboratory, and radiographic data for 250 consecutive LTR at a single US institution transplanted between 2007–2012. Each patient was evaluated independently by two lung transplant pulmonologists. Using an updated classification system⁴, for a given patient CLAD was designated if there was a sustained (≥ 3 weeks), $\geq 20\%$ decline in FEV1 as compared with the baseline value, in the absence of clinical confounders. As previously described, subtype was determined at time of CLAD onset, when patients were considered to have RAS if the forced vital capacity (FVC) was $\leq 80\%$ of FVC baseline, and BOS if the FVC was $\geq 80\%$ of FVC baseline⁴. Subtypes were thus defined only at CLAD onset, and this study was not designed to capture patients who evolved from one phenotype to a mixed pattern later. After initial independent review, discordant results were resolved through joint data review and discussion between the two reviewers. Cohen's kappa statistic was used to determine the initial inter-rater agreement for both presence or absence of CLAD and for CLAD subtype.

Of the 250 LTR, 51 (20.4%) were designated as having CLAD by Reviewer 1, and 50 (20%) by Reviewer 2. Table 1 shows the initial reviewer agreement between presence or absence of CLAD (regardless of subtype or date of onset), which demonstrates excellent agreement (kappa 0.938, 95% confidence interval [CI]:0.884–0.992). Of 48 cases initially concordantly designated as CLAD, the reviewers also agreed on the time of onset in 44 (91.7%). Two cases each of BOS and RAS were initially assigned different dates of onset (difference of 7 and 103 days for BOS, and 19 and 180 days for RAS). Of the 48 cases initially concordant CLAD diagnoses, there was more discrepancy between the classification of CLAD subtype (Table 2): In 13 of the 48 cases (27%), the reviewers initially assigned different subtypes. When restricting the inter-rater agreement analysis to the initial BOS versus RAS classifications in Table 2, the concordance dropped significantly (kappa 0.291, 95% CI: 0–0.62). The final diagnoses after discussion of discordant cases were: 34 BOS, 16 RAS, and 200 No CLAD.

Using a large cohort of LTR and incorporating newer diagnostic guidelines, we found high inter-rater agreement between two transplant pulmonologists in diagnosing CLAD (any subtype), but lower concordance with the subtypes BOS and RAS. To our knowledge, this is the first study that evaluates the diagnostic agreement of the major CLAD subtypes. Interestingly, there was much higher initial agreement on date of onset than that reported by Kalipa. This may be due in part to having two reviewers rather than 87. All discrepancies in our study were resolved after the two reviewers jointly discussed the cases. However, the results of this study highlight the potential for variability in CLAD diagnosis, and especially its subtypes and time of onset. One limitation of our study was that the reviewers were not blinded to the patient identity in order to facilitate the assessment of non-CLAD causes for spirometry decline. While it is possible that a reviewer's prior knowledge of a patient

affected the assessment, the use of two independent reviewers would help mitigate this effect. Given the importance of well-defined clinical endpoints for research studies, use of multiple reviewers should be considered in future clinical studies of CLAD. Additionally, due to there being no universally accepted criteria for the diagnosis of RAS, we recognize that the definition used in this paper may differ from some other studies. The publication of consensus guidelines, such as that planned by the ISHLT, will help with this diagnostic uncertainty. As the definition of CLAD and its subtypes evolve and diversify, and until more objective and precise diagnosis of CLAD becomes feasible, further research on inter-rater agreement should be conducted.

Acknowledgments

Disclosure Statement:

E.D.L. reports a grant from the Cystic Fibrosis Foundation. This work was supported by the National Institute of Health (K23HL143050 to CEF) and the Cystic Fibrosis Foundation (EDL).

References:

1. Chambers DC, Yusen RD, Cherikh WS, et al. The Registry of the International Society for Heart and Lung Transplantation: Thirty-fourth Adult Lung And Heart-Lung Transplantation Report-2017; Focus Theme: Allograft ischemic time. *J Heart Lung Transplant* 2017;36(10):1047–1059. [PubMed: 28784324]
2. Kapila A, Baz MA, Valentine VG, Bhorade SM, investigators A. Reliability of diagnostic criteria for bronchiolitis obliterans syndrome after lung transplantation: a survey. *J Heart Lung Transplant* 2015;34(1):65–74. [PubMed: 25447585]
3. Bhorade S, Ahya VN, Baz MA, et al. Comparison of sirolimus with azathioprine in a tacrolimus-based immunosuppressive regimen in lung transplantation. *Am J Respir Crit Care Med* 2011;183(3):379–387. [PubMed: 20833822]
4. Todd JL, Jain R, Pavlisko EN, et al. Impact of forced vital capacity loss on survival after the onset of chronic lung allograft dysfunction. *Am J Respir Crit Care Med* 2014;189(2):159–166. [PubMed: 24325429]

Table 1.

Initial agreement for diagnosis of CLAD vs. No CLAD (regardless of subtype or date of onset) between two transplant pulmonologists^a

Reviewer 1	Reviewer 2		Total
	CLAD	No CLAD	
CLAD	48	3	51
No CLAD	2	197	199
Total	50	200	250

CLAD: Chronic lung allograft dysfunction

^aFour cases where reviewers agreed on the diagnosis of CLAD but not the date of onset were included

Table 2.

Initial agreement of BOS and RAS diagnoses (among cases initially designated as CLAD, regardless of date of onset) between two transplant pulmonologists^a

Reviewer 1	Reviewer 2		
	BOS	RAS	Total
BOS	30	2	32
RAS	11	5	16
Total	41	7	48

BOS: Bronchiolitis obliterans syndrome; CLAD: Chronic lung allograft dysfunction

RAS: Restrictive allograft syndrome

^aFour cases where reviewers agreed on the diagnosis but not the date of onset were included (2 BOS, 2 RAS)