CHEST

Machine Learning Algorithms to Differentiate Among Pulmonary Complications After Hematopoietic Cell Transplant

Husham Sharifi, MD; Yu Kuang Lai, MD; Henry Guo, MD, PhD; Mita Hoppenfeld, MD; Zachary D. Guenther, MD; Laura Johnston, MD; Theresa Brondstetter, MS; Laveena Chhatwani, MD; Mark R. Nicolls, MD; and Joe L. Hsu, MD

BACKGROUND: Pulmonary complications, including infections, are highly prevalent in patients after hematopoietic cell transplantation with chronic graft-vs-host disease. These comorbid diseases can make the diagnosis of early lung graft-vs-host disease (bronchiolitis obliterans syndrome) challenging. A quantitative method to differentiate among these pulmonary diseases can address diagnostic challenges and facilitate earlier and more targeted therapy.

STUDY DESIGN AND METHODS: We conducted a single-center study of 66 patients with CT chest scans analyzed with a quantitative imaging tool known as parametric response mapping. Parametric response mapping results were correlated with pulmonary function tests and clinical characteristics. Five parametric response mapping metrics were applied to K-means clustering and support vector machine models to distinguish among post-transplantation lung complications solely from quantitative output.

RESULTS: Compared with parametric response mapping, spirometry showed a moderate correlation with radiographic air trapping, and total lung capacity and residual volume showed a strong correlation with radiographic lung volumes. K-means clustering analysis distinguished four unique clusters. Clusters 2 and 3 represented obstructive physiology (encompassing 81% of patients with bronchiolitis obliterans syndrome) in increasing severity (percentage air trapping 15.6% and 43.0%, respectively). Cluster 1 was dominated by normal lung, and cluster 4 was characterized by patients with parenchymal opacities. A support vector machine algorithm differentiated bronchiolitis obliterans syndrome with a specificity of 88%, sensitivity of 83%, accuracy of 86%, and an area under the receiver operating characteristic curve of 0.85.

INTERPRETATION: Our machine learning models offer a quantitative approach for the identification of bronchiolitis obliterans syndrome vs other lung diseases, including late pulmonary complications after hematopoietic cell transplantation. CHEST 2020; 158(3):1090-1103

KEY WORDS: bone marrow; bronchiolitis obliterans; graft vs host; medical informatics; organizing pneumonia; radiology—thoracic

FOR EDITORIAL COMMENT, SEE PAGE 852

ABBREVIATIONS: ATS = American Thoracic Society; BOS = bronchiolitis obliterans syndrome; cGVHD = chronic graft-vs-host disease; ERS = European Respiratory Society; HCT = hematopoietic cell transplantation; NIH = National Institutes of Health; OP = organizing pneumonia; PCA = principal component analysis; PFT = pulmonary function test; PRM = parametric response mapping; RV = residual volume; TLC = total lung capacity; TS = truncal sclerosis

AFFILIATIONS: From the Department of Medicine, the Division of Pulmonary, Allergy, and Critical Care Medicine (Drs Sharifi, Lai, Chhatwani, Nicolls, and Hsu and Ms Brondstetter), the Departments

Bronchiolitis obliterans syndrome (BOS) is a pulmonary manifestation of chronic graft-vs-host disease (cGVHD) that typically occurs within 2 years after hematopoietic cell transplantation (HCT).[1] BOS is characterized by small airways inflammation and fibrinous obliteration of the bronchiolar lumen.[2,3] Clinical presentation includes insidious onset of dyspnea, cough, and fixed obstructive defect on spirometry. Rapid irreversible decline in lung function is the hallmark of the disease.[3]

The National Institutes of Health (NIH) diagnostic criteria for BOS comprise clinical history, air trapping, and hyperinflation on pulmonary function testing (PFT) and CT.[2] However, its sensitivity for diagnosing BOS is poor, because only half of patients with biopsy-proven BOS meet NIH diagnostic criteria.[4,5] Comorbid conditions, such as pulmonary infections, inflammatory pneumonia (eg, organizing pneumonia [OP]), and extraparenchymal restriction resulting from cGVHD-related sclerodermatous disease of the thorax (truncal sclerosis [TS]) can directly affect static and dynamic pulmonary assessment and thus make the diagnosis difficult.[6-8] Steroid-induced myopathy is also highly prevalent in this population and can further confound identification of disease. Because of these etiologic uncertainties, patients are often misdiagnosed or diagnosed after a fixed obstructive defect has become apparent late in the disease, increasing morbidity and mortality.[3]

Pulmonary parametric response mapping (PRM), a quantitative imaging tool, has been studied in patients with BOS, with promising results.[9] In addition to assessing inspiratory and expiratory volumes, PRM quantifies percentage of lung involved in air trapping, a marker of BOS,[9] and percentage of lung that is normal, emphysema-like, or has parenchymal opacities.[10]

PRM has been used to assess progression of small airways damage in patients with COPD.[11,12] In lung transplant patients with BOS, PRM showed a moderate correlation between air trapping and decline in FEV_1 and provided prognostic information on survival.[13,14] However, whether these findings can be extended to BOS after HCT is unknown.

Therefore, we evaluated the utility of PRM in a well-characterized cohort of patients at our academic medical center. We analyzed the association of PFTs and PRM metrics and applied an unsupervised machine learning algorithm to explore whether PRM metrics alone can represent patient physiology. From the output of this algorithm, we created a machine learning model to diagnose BOS and compared the model's performance with radiologists' interpretations. Our results provide new insight into the diagnosis and management of lung GVHD and other pulmonary complications after HCT.

Materials and Methods

Eligible patients were those who received HCT and were seen in our Lung GVHD clinic from June 22, 2015 to June 22, 2018, with the following inclusion criteria: (1) a documented complaint of dyspnea on exertion; (2) PFTs in the 3-month period before HCT; (3) at least one PRM CT chest; and (4) PFTs in the 3-month period after PRM CT chest (all but one patient received PFTs within 1 month of PRM). Demographic features, medical history, medication use, specific organ manifestation and staging of cGVHD, and mortality were abstracted from the medical record as well as our center's Blood and Marrow Transplantation database. Human subject's approval was obtained from our institutional review board (IRB number 43215). Data were collected and analyzed in compliance with the Health Insurance Portability and Accountability Act.

A panel of four clinicians (three pulmonary physicians [J. L. H., L. C., and H. S.) and one blood and marrow transplant physician [L. J.]) adjudicated the clinical diagnosis for all eligible patients. Two thoracic radiologists (HG, ZG) reached a consensus diagnosis for BOS based on established criteria in the radiology literature.[15]

Imaging Methods

Noncontrast thoracic CT scanning at full inspiration (total lung capacity [TLC]) and end expiration (residual volume [RV]) was performed using helical CTs (Siemens Force, Siemens Medical Systems; or GE Discovery CT750 HD, GE Healthcare). Refer to e-Appendix 1 for the protocolized patient coaching in acquisition of volume measurements.

Scans performed with a Siemens scanner were with 515 × 512 reconstruction matrix, 192 × 0.6 mm collimation, 120 kV, rotation time 0.25 seconds, pitch 1, CareDose4D at QRM125. Images were

of Radiology (Drs Guo and Guenther) and Medicine (Dr Hoppenfeld), and the Department of Medicine, Division of Blood and Marrow Transplantation (Dr Johnston), Stanford University School of Medicine, Stanford, CA.

Drs Sharifi and Lai contributed equally to this manuscript.

Part of this article has been presented as a poster at the 2019 CHEST Annual Meeting, October 6-10, 2018, San Antonio, TX.

FUNDING/SUPPORT: This work was supported by a grant from the NIH, the National Heart, Lung, and Blood Institute [Grant 1K08HL122528-01A1 to J.L.H.]. The Blood and Marrow Transplant database used in this study was funded by P01 CA049605 and P01 HL075462 (NIH, Bethesda, MD).

CORRESPONDENCE TO: Joe Hsu, MD, MPH, the Division of Pulmonary, Allergy, and Critical Care Medicine, Stanford University School of Medicine, Grant Bldg S069, 300 Pasteur Dr, Stanford, CA 94305-5236; e-mail: joehsu@stanford.edu

DOI: https://doi.org/10.1016/j.chest.2020.02.076

reconstructed at 1-mm axial slice thickness with sharp reconstruction algorithm (filter Br54) for visual assessment and neutral reconstruction algorithm (filter Bf32) at 0.7 mm increment for PRM quantitative analysis. Scans performed with a GE scanner were with 515 × 512 reconstruction matrix, 64 × 0.625 mm collimation, 120 kV, rotation time 0.5 seconds, pitch 1.375, SmartMA at NI52. Images were reconstructed at 1.25-mm axial slice thickness with bone kernel for visual assessment and standard kernel at 0.8-mm increment for PRM quantitative analysis.

In addition to measurement of inspiratory and expiratory volumes, PRM processing of CT images produced a percentage for each of four metrics—normal lung, air trapping, emphysema-like lung disease, and parenchymal opacities. Values greater than or equal to 950 Hounsfield units (HU) and less than −810 HU at inspiration and greater than or equal to −856 HU at expiration were classified as normal (PRM^{Normal}, green voxels). Greater than or equal to 950 HU and less than −810 HU at inspiration and less than −856 at expiration were classified as air trapping (PRM^{AT}, yellow voxels), also known as functional small airways disease. Greater than or equal to −1,000 HU and less than −950 HU at inspiration and less than −856 HU at expiration were classified as emphysema-like lung (PRM^{Emp}, red voxels). Voxels greater than or equal to −810 HU and less than −500 HU at inspiration and less than −500 HU at expiration were classified as parenchymal disease (PRM^{PD}, purple voxels).

PRM metrics were compared with PFTs to assess physiological relevance for obstructive and restrictive lung disease. Mean PRM metrics were compared across the four clinical disease states. CT-based lung volumes were adjusted by American Thoracic Society (ATS)/European Respiratory Society (ERS) guidelines and are presented in detail in e-Appendix 1.[16] Additional details of GVHD diagnostic criteria and of radiology diagnosis are presented in the Methods section e-Appendix 1 and e-Table 1.[2]

Statistics

Metrics for obstructive and restrictive disease were analyzed with Pearson Correlation. Statistical significance of differences was assessed with the two-sample *t*-test for pairs, the Kruskal Wallis test for group comparisons, and the Wilcoxon rank-sum test for pairwise comparisons of multiple groups. The Benjamini-Hochberg procedure was applied to control for multiple comparisons. $P < .05$ was considered significant. K-means clustering was used for clustering analysis. Optimal cluster number K was chosen by the elbow method with a Scree plot, as described in the online supplemental material and seen in e-Figure 1. A support vector machine algorithm was trained with 10-fold cross-validation on a 75% training set, and the model's performance was assessed on a held-out 25% test set. Further detail on the machine learning analysis is in e-Appendix 1. All statistical analyses were performed in R version 3.6.1, with the stats package for K-means clustering and the caret package for classification analytics.

Results

Seventy-nine patients were screened for inclusion in the study. Thirteen patients were excluded based on loss to follow-up (n = 1), lack of follow-up PFTs (n = 2), inability to locate pre-HCT PFTs (n = 1), or technically suboptimal PRM studies (n = 9). No significant differences in demographic or disease characteristics were seen between included and excluded subjects (e-Table 2) Sixty-six patients with PRM scans were included in the study. The cohort comprised 37 men and 29 women, with a mean age of 52 years. Most patients had no lung disease before HCT (Table 1).[17] Most received myeloablative transplant regimens (56.1%), and a similar distribution of patients underwent transplants from matched related donors vs matched/mismatched unrelated donors (45.5% vs 51.6%, respectively).

Ninety-one percent of patients were diagnosed with cGVHD (Table 2). Mean time from transplantation to last follow-up was 25 months (median = 19 months; interquartile range = 12-30 months). Mean time from transplantation to diagnosis of lung GVHD was 35 months (median = 26 months, interquartile range = 10-46 months). During study follow-up, all-cause mortality was 16.4%. BOS was the most common posttransplantation lung complication (35%), and 17% of patients had OP either in isolation or in combination with a second posttransplantation lung complication. TS was observed in 14% of patients.

TABLE 1] Demographic Data Before Hematopoietic Cell Transplantation

No.	66
Age, mean (SD)	52.2 (12.5)
Male, No. (%)	37 (56.7)
Race/ethnicity, No. (%)	
Asian	8 (11.9)
White	51 (76.1)
Hispanic	8 (11.9)
HCT-CI score[a]	
<3 Comorbidities	50
≥3 Comorbidities	11
Underlying reason for HCT, No. (%)	
AML	20 (30.3)
ALL	14 (21.2)
MDS	12 (18.2)
CML	6 (9.1)
CLL	3 (4.5)
NHL	6 (9.1)
HL	2 (3.0)
Other	3 (4.5)
Type of donor, No. (%)	
Match-related	30 (45.5)
Unrelated-identical	24 (36.4)
Unrelated-mismatched	10 (15.2)
Haploidentical/cord	2 (3)

(Continued)

TABLE 1] *(Continued)*

Regimen, No. (%)	
Myeloablative	37 (56.1)
Non-myeloablative	19 (28.8)
Reduced intensity	10 (15.2)
PFT characteristics before HCT, mean (SD)	
FEV_1, L	3.20 (0.76)
FEV_1, % predicted	100.5 (16.0)
FVC, L	4.13 (0.98)
FVC, % predicted	99.6 (13.8)
FEV_1/FVC, %	78.14 (7.64)
TLC, L	6.10 (1.68)
TLC, % predicted	104.7 (15.8)
RV, L	1.89 (0.88)
RV, % predicted	99.1 (32.9)
RV/TLC, %	29.42 (8.37)
D_{LCO}, % predicted	89.51 (20.10)
History of prior lung disease, No. (%)	
Asthma	12 (17.9)
COPD	1 (1.5)
No prior lung disease	54 (80.6)
History of smoking, No. (%)	21 (31.3)

ALL = acute lymphoblastic leukemia; AML = acute myeloid leukemia; CLL = chronic lymphocytic leukemia; CML = chronic myelogenous leukemia; D_{LCO} = diffusing capacity for carbon monoxide; HL = Hodgkin's lymphoma; MDS = myelodysplastic syndrome; NHL = non-Hodgkin's lymphoma; RV = residual volume; TLC = total lung capacity.
[a]Hematopoietic cell transplantation comorbidity index score (HCT-CI) was calculated as described by Sorror.[17]

TABLE 2] Demographic Data After Hematopoietic Cell Transplantation

No.	66
Median length of clinical follow-up, mo (IQR)	19 (12-30)
Median time from HCT to lung cGVHD diagnosis, mo (IQR)	26 (10-46)
Acute GVHD > grade II, No. (%)	30 (68.2)
cGVHD, No. (%)	60 (90.9)
cGVHD grade, No. (%)[a]	
Mild	4 (6.7)
Moderate	18 (30)
Severe	38 (63.3)
Other sites of cGVHD involvement, No. (%)	
Eyes	36 (60)
Mouth	41 (68.3)
GI tract	13 (21.6)
Liver	25 (41.7)
Skin	47 (78.3)
Musculoskeletal	14 (23.3)
Fascia	14 (23.3)
Other	9 (15)
Pulmonary complications after HCT, No. (%)	
BOS	26 (38.8)
OP	12 (17.9)
TS	10 (14.9)
Infection	3 (4.5)
Fibrosis	3 (4.5)
Criteria for BOS diagnosis, No. (%)[b]	
Clinical criteria	2 (6.5)
Biopsy	2 (6.5)
NIH	27 (87)
Treated with FAM, %[c]	77
Mortality, No. (%)	11 (16.4)
No. of PRM CT scans analyzed	66

BOS = bronchiolitis obliterans syndrome; cGVHD = chronic graft-vs-host disease; FAM = inhaled fluticasone, azithromycin, montelukast; NIH = National Institutes of Health; OP = organizing pneumonia; PRM = parametric response mapping; TS = truncal sclerosis.
[a]cGVHD was staged based on the National Institute of Health consensus guidelines.[2]
[b]Of the patients diagnosed with BOS by NIH criteria, four had co-existing OP and therefore were assigned OP as their diagnosis in the analysis. Of the two patients diagnosed with BOS by clinical criteria, one patient was assigned as having BOS in the analysis because of a strong response to BOS therapy by functional activity, respiratory status, and PFTs. With inclusion of the two patients diagnosed by biopsy, 26 patients were classified as having BOS for the analysis.
[c]All patients with a diagnosis of BOS were treated with the FAM regimen.

Posttransplantation PFTs indicated obstructive and restrictive physiology (Table 3). BOS patients demonstrated moderate obstruction ($P < .001$) and a higher TLC in absolute value and percent predicted ($P < .001$). RV was significantly higher in absolute value ($P = .04$) and percent predicted ($P < .001$) in patients with BOS compared with patients with OP or TS, as was RV/TLC ($P < .001$). In TS, extraparenchymal restriction was evident from a reduced absolute and percent predicted TLC ($P < .001$).

PFTs, PRM, and Disease States

Figure 1 depicts associations between PFT and PRM parameters. Moderate negative correlation was found between FEV_1 and air trapping (-0.42) (Fig 1A), and RV and air trapping had a moderate positive correlation (0.62) (Fig 1C). The strongest correlations were observed for CT-based absolute volume measures (Fig 1D, 1E): 0.70 between RV and expiratory volume, and 0.88 between TLC and inspiratory volume. In patients with BOS, TLC and inspiratory volume had a correlation coefficient of 0.94; RV and expiratory volume had a

TABLE 3] Pulmonary Function Tests After HCT for Patients With BOS, OP, or TS

	Bronchiolitis Obliterans Syndrome	Organizing Pneumonia	Truncal Sclerosis	*P*[a]
No.	26	12	10	NA
PFTs, mean (SD)				
FEV_1 (L)	1.72 (0.79)	2.41 (0.80)	2.18 (0.74)	< .001
FEV_1, % predicted	59.6 (35.7)	75.2 (31.3)	68.9 (21.7)	< .001
FVC, L	3.13 (1.09)	3.94 (0.99)	2.76 (0.99)	.003
FVC, % predicted	80.1 (16.9)	75.6 (29.5)	66.4 (22.8)	< .001
FEV_1/FVC, %	54.8 (11.9)	74.2 (10.2)	80.1 (8.4)	< .001
TLC, L	5.64 (1.50)	4.86 (1.46)	4.42 (1.05)	< .001
TLC, % predicted	98.6 (16.4)	74.2 (21.0)	77.4 (18.3)	< .001
RV, L	3.16 (7.25)	1.44 (0.70)	1.60 (0.59)	.041
RV, % predicted	129.1 (36.1)	69.0 (26.6)	79.2 (22.4)	< .001
RV/TLC, %	44.6 (10.1)	35.0 (12.9)	35.8 (10.6)	< .001
D_{LCO}, % predicted	78.2 (20.1)	66.1 (19.6)	65.7 (14.1)	< .001

All patients were assigned a diagnosis for post-transplantation complication based on adjudicated criteria (see Methods). PFT = pulmonary function test. See Table 1 legend for expansion of other abbreviations.
[a]Kruskall-Wallis test used for r statistical significance.

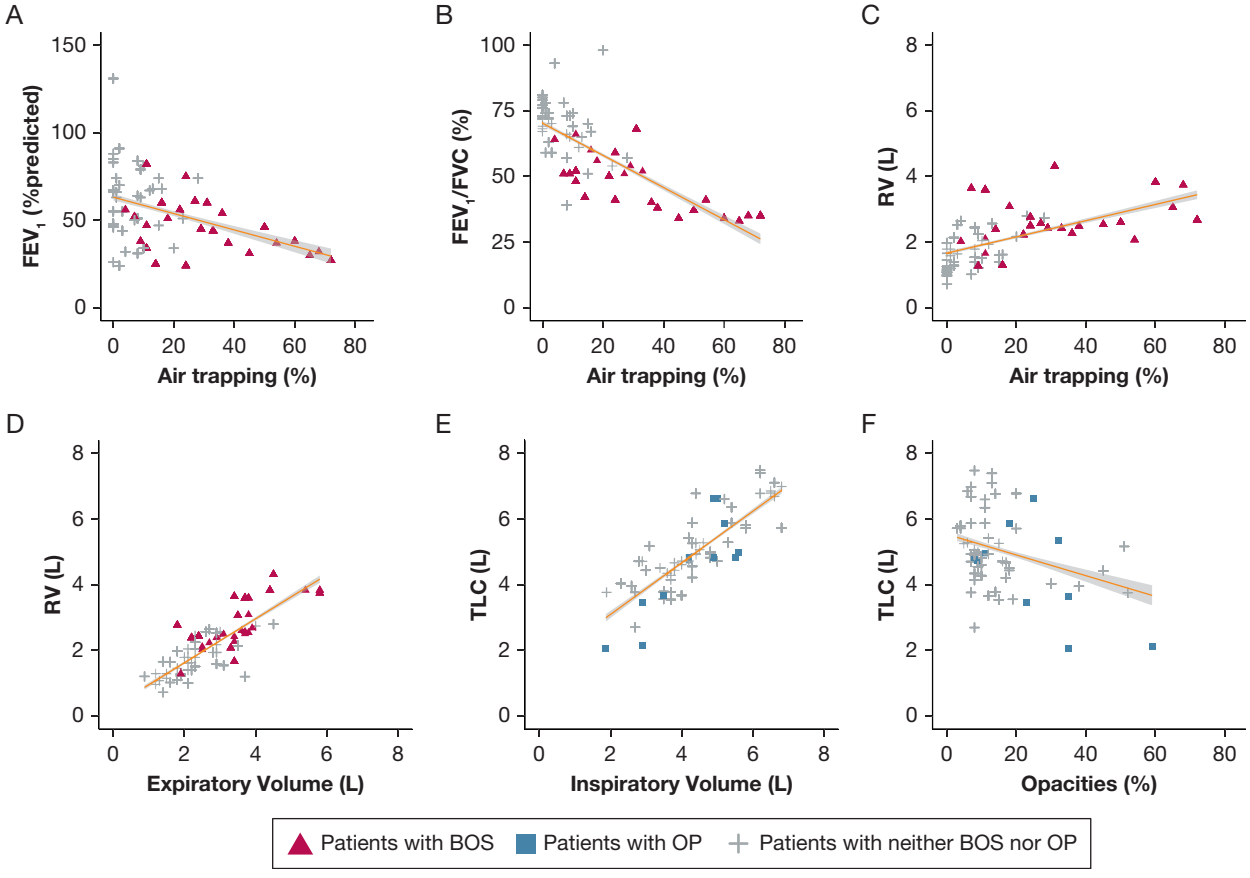

Figure 1 – *Fitted regression lines for pulmonary function test metrics vs parametric response mapping (PRM) metrics. A, FEV_1 percent predicted vs percentage air trapping by PRM. B, FEV_1/FVC vs percentage air trapping by PRM. C, RV vs percentage air trapping by PRM. D, RV vs expiratory volume by PRM. E, TLC vs inspiratory volume by PRM. F, TLC vs percentage parenchymal opacities by PRM. The gray region represents the 95% CI. OP = organizing pneumonia; RV = residual volume; TLC = total lung capacity.*

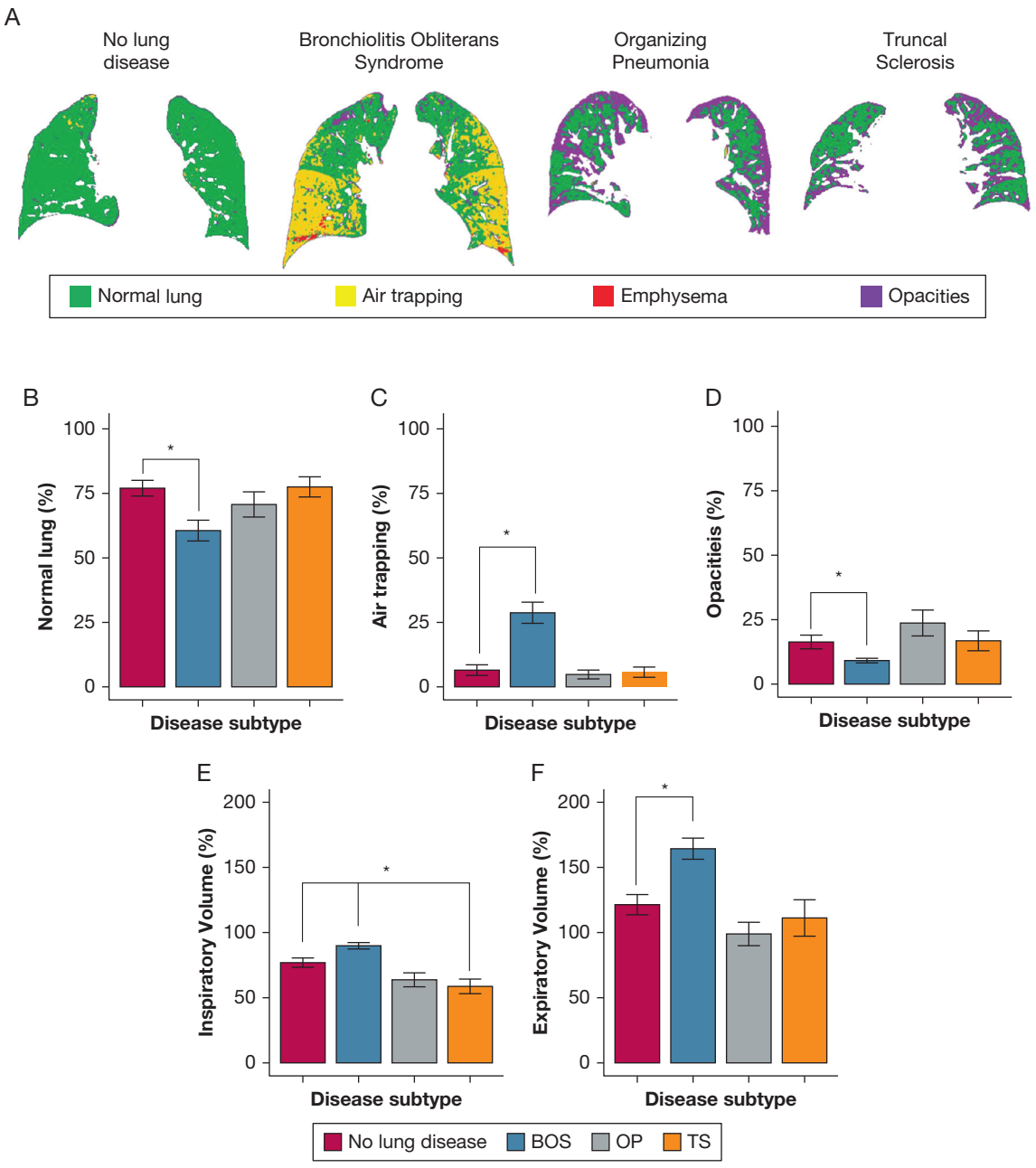

Figure 2 – *PRM metrics by posttransplantation lung complication. Representative coronal cuts by disease state classified by PRM colors. B, Percentage normal lung by posttransplantation lung complication. C, Percentage air trapping by posttransplantation lung complication; D, Percentage parenchymal opacities by posttransplantation lung complication. E, Percentage predicted inspiratory volume (adjusted for height and sex) by posttransplantation lung complication. F, Percentage predicted expiratory volume (adjusted for age, height, and sex) by posttransplantation lung complication.*[19] *Each percentage is a mean. Error bars show the SEM. By Wilcoxon rank sum test of each disease relative to people with no lung disease, statistical significance to P < .05 is labeled with a star. Benjamini-Hochberg procedure is used to control for the false discovery rate. Diagnostic categories after hematopoietic cell transplantation contain 18 patients with no lung disease, 26 patients with BOS, 12 patients with OP, and 10 patients with TS. TS = truncal sclerosis. See Figure 1 legend for expansion of other abbreviations.*

correlation coefficient of 0.66 ($P < .0001$). These data suggest that PRM has a role in assessing the degree of obstructive and restrictive lung physiology, especially in the form of volume measurements. Given that no significant differences were seen using absolute or percent predicted CT-based volumes, the volumes in what follows are presented as the predicted values based on established ATS/ERS criteria.[16]

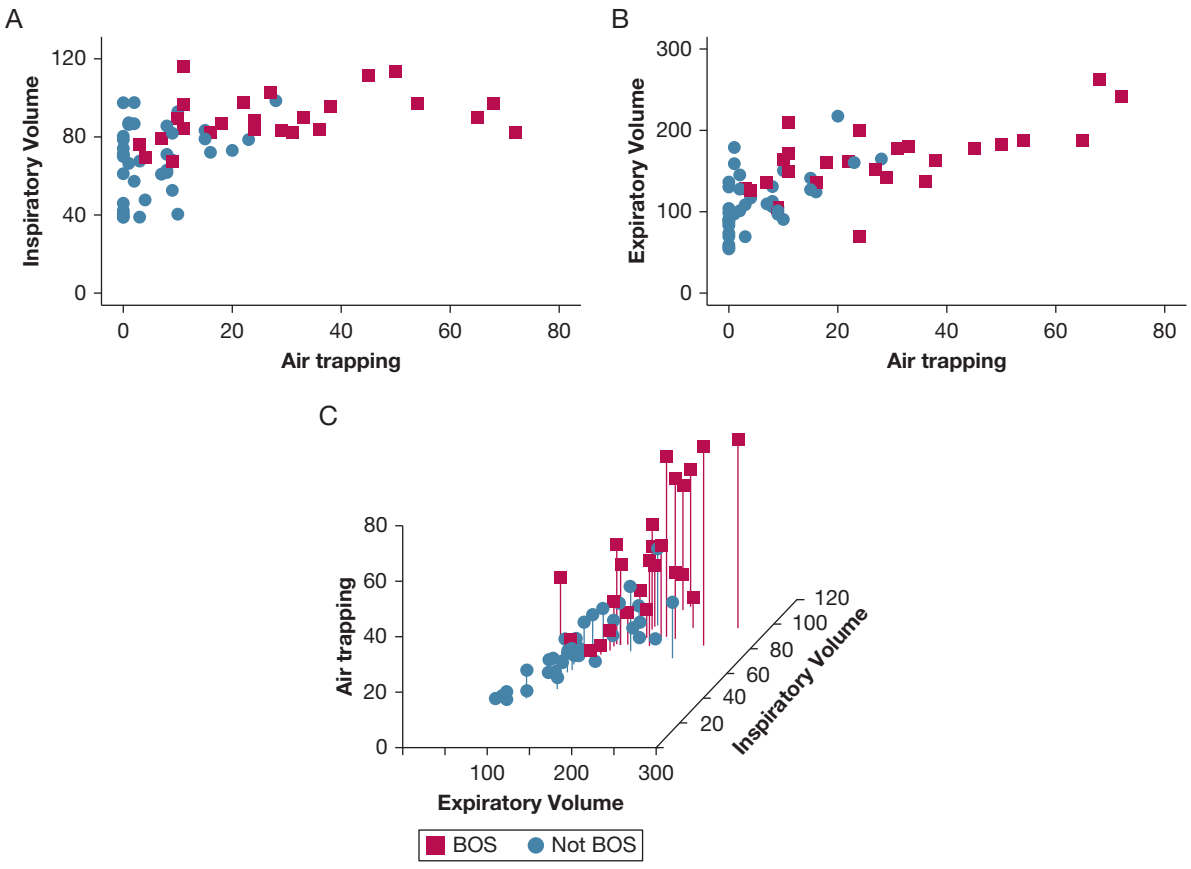

Figure 3 – *Scatterplots for BOS by relevant PRM metrics. A, Inspiratory volume vs air trapping for patients with BOS and without BOS. B, Expiratory volume vs air trapping for patients with BOS and without BOS. C, Three-dimensional scatterplot for BOS by PRM metrics of air trapping percentage vs percentage predicted expiratory volume vs percentage predicted inspiratory volume. An interactive, three-dimensional scatterplot is also found online in e-Figure 2.*[18] *See Figure 1 and 2 legends for expansion of abbreviations.*

Figure 2A depicts representative coronal CT cuts processed by PRM. In comparing BOS patients with patients without lung disease (Fig 2C), there was significantly higher air trapping (28.7% vs 6.5%, respectively; $P < .0001$) and inspiratory volume (Fig 2E, 89.9% vs 76.9%, respectively; $P = .005$). In contrast, for patients with TS, inspiratory volume was significantly lower than in patients with no lung disease (58.7% vs 76.9%, respectively; $P = .01$; Fig 2E). Expiratory volume also revealed hyperinflation in patients with BOS vs those without lung disease ($P = .0004$; Fig 2F). These associations and ground truth diagnosis are illustrated in two- and three-dimensional scatterplots (Fig 3).[18] Together, these data suggest that the PRM metrics of air trapping and inspiratory and expiratory volumes may be useful in differentiating pulmonary complications after HCT.

Machine Learning for Pulmonary Complications After HCT

K-means clustering by PRM metrics captured 76.4% of the cohort's variance and identified four patient clusters (Table 4; Fig 4). Cluster 1 was characterized by normal lung physiology (82.3% normal lung by PRM). Mean values for

TABLE 4] Parametric Response Mapping Metrics by Mean Percentage per Cluster

Cluster	No.	Normal Lung, %	Parenchymal Opacities, %	Air Trapping, %	Inspiratory Volume, %	Expiratory Volume, %
1	24	82.3	12.1	5.3	68.0	104.3
2	21	71.5	12.1	15.6	89.9	150.8
3	11	45.4	9.2	43.0	92.9	202.2
4	7	57.9	40.1	1.7	44.8	71.6

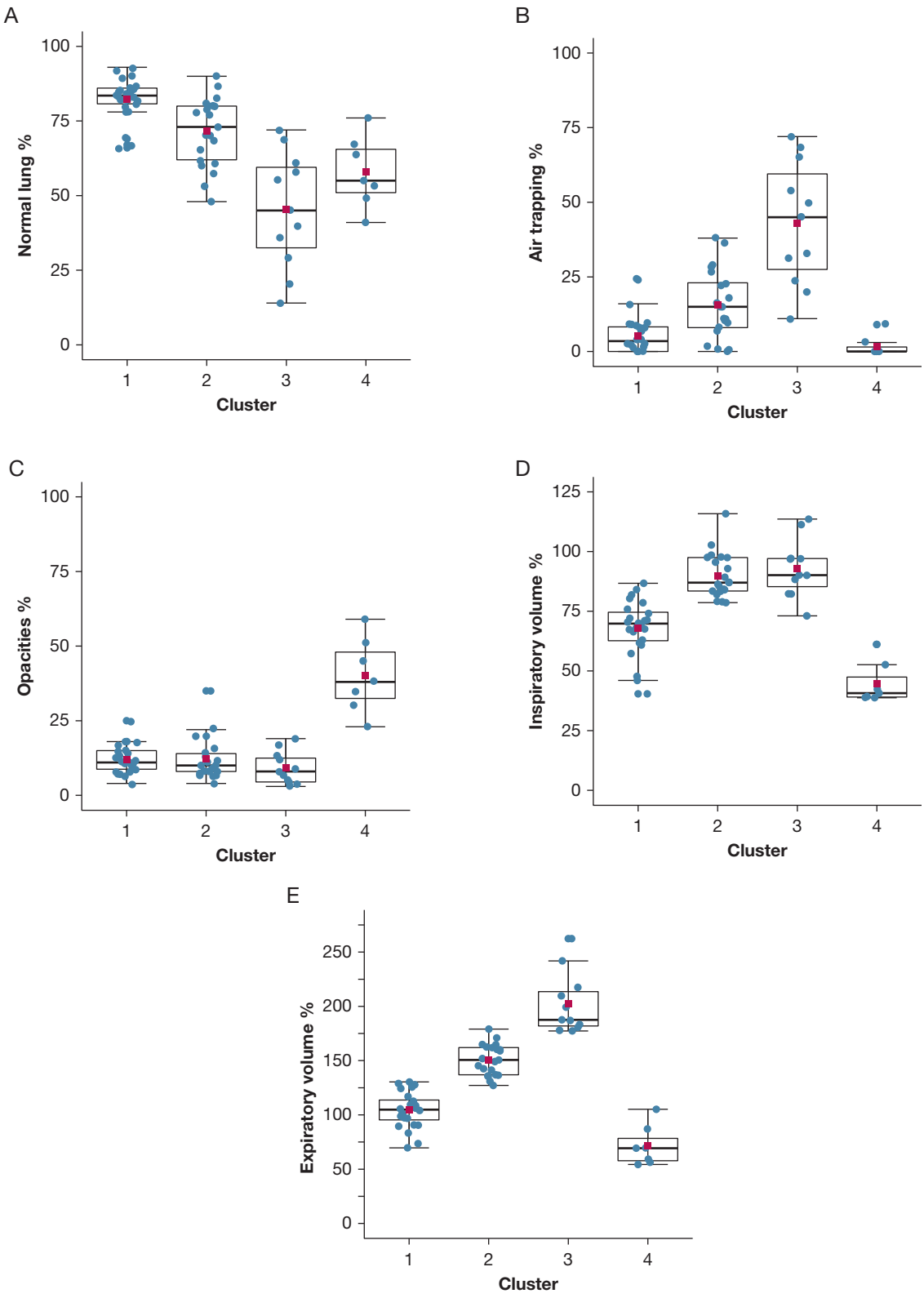

Figure 4 – *PRM metrics by clusters. A, Difference in mean normal lung between clusters 3 and 4 does not achieve statistical significance. All other means of cluster pairs differ with P < .02. B, Mean air trapping of clusters 1 and 4 differ with P = .08, and means of all other cluster pairs differ with P ≤ .002. C, Mean opacities in cluster 4 differs from all other clusters with P < .001. D, Mean inspiratory volumes for clusters 2 and 3 do not differ with statistical significance. All other means of cluster pairs have P < .0001. E, All mean expiratory volumes for all cluster pairs differ with P < .001. The box represents the interquartile range (IQR), and the thick line in the box is the median. Whiskers are 1.5 of the IQR. The red square is the mean. All points for the PRM scans are shown, including outliers. Wilcoxon rank-sum was used to calculate significance, and Benjamini-Hochberg was used to control for multiple comparisons.*

TABLE 5] Pretransplantation PFTs by Cluster[a]

	Cluster 2	Cluster 3	*P*[a]
No.	21	11	NA
Age, mean (SD)	54.81 (9.96)	49.73 (10.71)	.191
Female, No. (%)	9 (43)	7 (64)	.457
PFTs, mean (SD)			
FEV_1, L	3.16 (0.82)	3.22 (0.69)	.846
FEV_1, % predicted	103.8 (14.3)	102.6 (16.0)	.834
FVC, L	4.20 (1.07)	4.12 (0.90)	.843
FVC, % predicted	104.57 (12.66)	102.64 (13.89)	.694
FEV_1/FVC, %	75.62 (6.52)	78.18 (5.69)	.279
TLC, L	6.57 (1.52)	6.35 (2.56)	.831
TLC, % predicted	117.00 (15.33)	104.00 (16.92)	.128
RV, L	2.01 (0.88)	1.88 (1.29)	.813
RV, % predicted	112.10 (38.77)	86.67 (41.38)	.235
RV/TLC, %	32.00 (9.21)	26.00 (9.63)	.225
DLCO, % predicted	98.81 (17.27)	89.36 (18.50)	.162

See Tables 1 and 3 legends for expansion of abbreviations.
[a]Two-sample t test used to calculate P value.

clusters 2 and 3 suggested increased severity of BOS by air trapping (15.6% vs 43.0%, respectively, $P < .0001$) and expiratory volume (151% vs 202%, respectively; $P < .0001$). Moreover, clusters 2 and 3 encompassed 81% of patients with the clinical diagnosis of BOS.

Pretransplantation PFTs for patients in clusters 2 and 3 revealed no statistically significant differences (Table 5). In contrast, posttransplantation PFTs for patients in clusters 2 and 3 indicated differences that were clinically and statistically significant in absolute FEV_1, percentage predicted FEV_1, percent predicted FVC, percent predicted RV, and RV/TLC (Table 6). Together, this suggests that patients in cluster 2 and cluster 3 had marked worsening of obstruction longitudinally after transplantation.

We conducted a principal component analysis (PCA) to understand the contribution of each PRM metric to the ground truth diagnosis of pulmonary complications

TABLE 6] Post-transplantation PFTs by Cluster

	Cluster 2	Cluster 3	*P*[a]
No.	21	11	NA
All-cause mortality, No. (%)[b]	2 (9.5)	2 (18.2)	.888
PFTs, mean (SD)			
FEV_1, L	1.96 (0.88)	1.24 (0.45)	.017
FEV_1, % predicted	64.95 (21.67)	41.82 (14.74)	.004
FVC, L	3.35 (1.23)	2.59 (0.81)	.080
FVC, % predicted	82.95 (18.55)	66.64 (18.47)	.028
FEV_1/FVC, %	57.63 (11.46)	50.73 (20.31)	.241
TLC, L	5.80 (1.49)	5.62 (1.35)	.739
TLC, % predicted	97.21 (14.15)	97.55 (13.13)	.949
RV, L	2.37 (0.77)	2.90 (0.86)	.093
RV, % predicted	109.84 (31.08)	148.45 (23.37)	.001
RV/TLC, %	41.37 (11.53)	49.64 (7.38)	.042
DLCO, % predicted	74.63 (30.81)	69.82 (19.39)	.645

See Tables 1 and 3 legends for expansion of abbreviations.
[a]Two-sample t-test used to calculate P value.
[b]The two deaths in cluster 2 were secondary to relapse-related mortality. The two deaths in cluster 3 were secondary to BOS-related mortality.

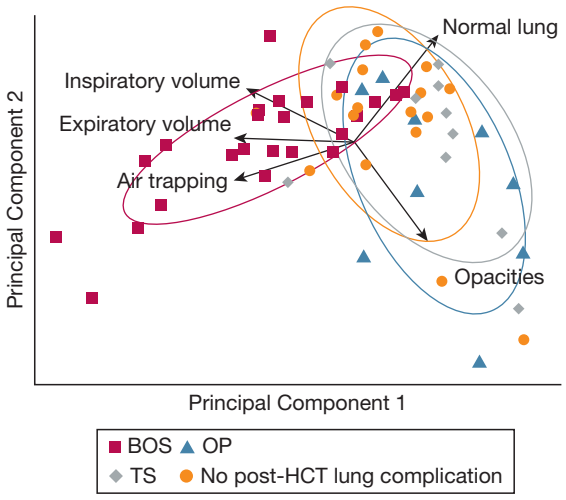

Figure 5 – *Principal components analysis of PRM metrics. The first two principal components (PCs) for five PRM metrics are shown. PC 1 explains 49.4% of cohort variance, whereas PC 2 explains 29.2% of cohort variance (78.6% of total variance). Each arrow represents a PRM metric, with the arrow direction representing its contribution to a PC and the arrow length representing the magnitude of its contribution to variance. Each ellipse is the normal probability for the group. The units of each axis are a composite of linear combinations of the original PRM variables. See Figure 1 legend for expansion of abbreviation.*

after HCT (Fig 5). The PRM metrics of air trapping and expiratory and inspiratory volume pointed in the direction of BOS (red squares), indicating that these metrics contribute to the variance that makes BOS patients differ from others. Percentage normal lung pointed in the direction of patients with no lung disease (orange circles), and volume measurements pointed in the opposite direction of TS (gray diamonds). In sum, these results imply that PRM analysis can create a taxonomy for pulmonary complications after HCT, forming the foundation of a diagnostic model for BOS based on air trapping and inspiratory and expiratory volumes.

Given the association and exploratory analyses, we applied a support vector machine algorithm to classify patients with BOS. Figure 6 shows the decision boundary for identifying BOS from air trapping and expiratory volume (Fig 6A, 6B) and from air trapping and inspiratory volume (Fig 6C, 6D). For example, in a patient with air trapping of 15%, the algorithm diagnosed BOS when the expiratory volume was greater than or equal to approximately 150% of predicted. When air trapping was the sole variable, the model had an accuracy of 79% ($P = .05$), specificity of 88%, and sensitivity of 67% ($P = .05$). Adding expiratory volume did not alter diagnostic performance, but adding inspiratory volume to air trapping increased accuracy to 86% ($P = .02$) and sensitivity to 83% ($P = .03$). Table 7 shows model performance in comparison with the consensus diagnosis of two board-certified thoracic radiologists (H. G. and Z. G.), with a specificity (60%) and sensitivity (70%) ($P = .02$) comparable to those of previous studies assessing CT-based diagnosis of BOS.[19,20] These results suggest that the thresholds created by our machine learning models can help refine and enhance interpretation of volumetric chest CTs in HCT patients with posttransplantation lung complications.

Discussion

In this study we used machine learning algorithms to identify pulmonary complications after HCT, to diagnose BOS, and to assess BOS severity from PRM processing of chest CT scans. By PRM metrics alone, these models offer an approach to assisting in the differentiation and clinical care of HCT patients with cGVHD.

Lung GVHD after HCT is increasingly appreciated to be accompanied by heterogeneous comorbidities. Although BOS is well-described by NIH guidelines, conditions such as OP and TS in this population have only been studied in descriptive cohorts.[21,22] Our well-characterized and uniformly observed and treated cohort comprised people who manifested BOS, OP, and TS, with a prevalence consistent with the literature.[5,7] Following-up with the patients in our clinic for years allowed us to assess response to therapy and thus improved our diagnostic accuracy.

Our findings of a moderate negative and positive correlation, respectively, of PRM air trapping with FEV_1 and RV/TLC recapitulated those of studies conducted in patients with BOS after lung transplantation, which

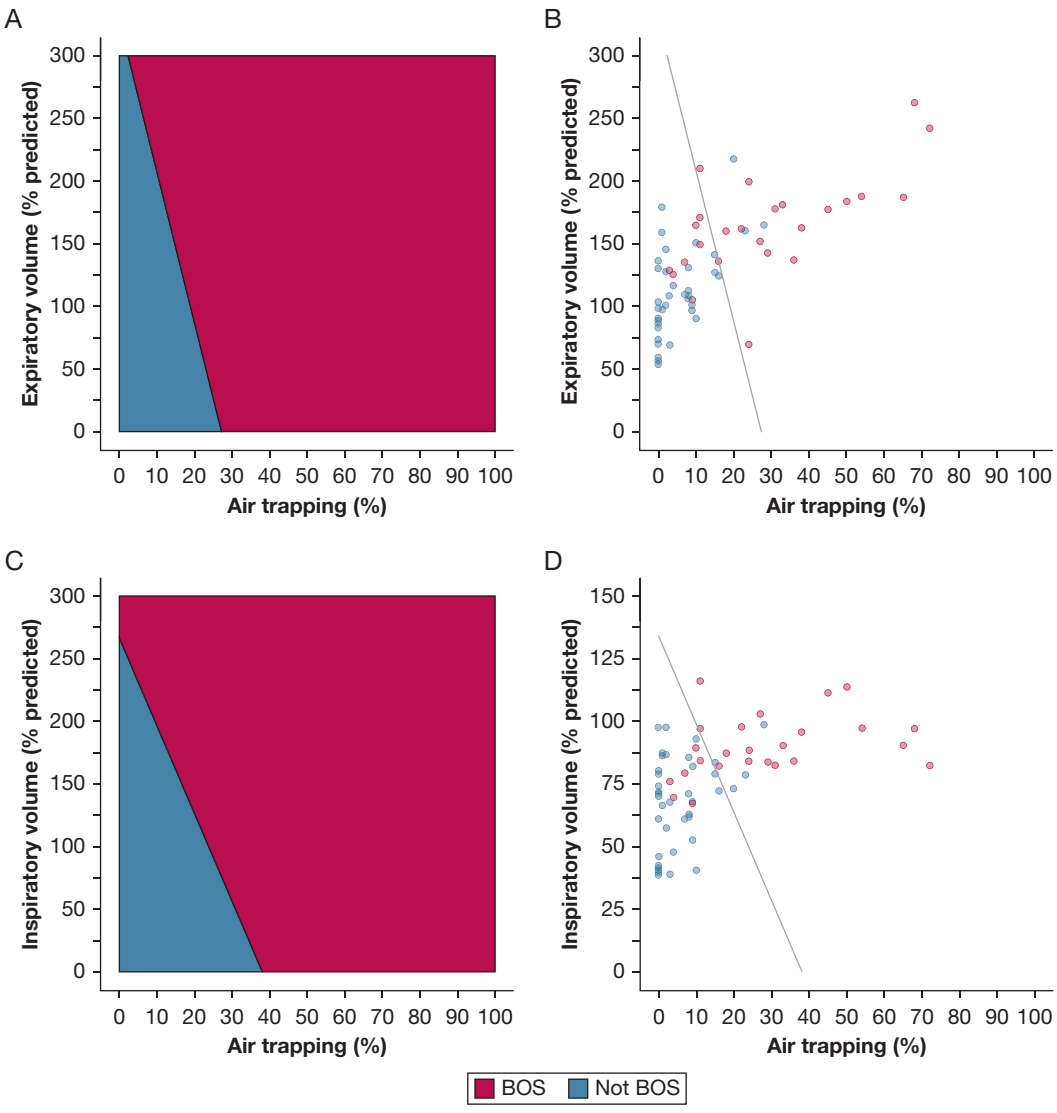

Figure 6 – *Machine learning decision boundaries for diagnosing BOS. A, Machine learning decision boundaries for diagnosing BOS from air trapping and expiratory volume. The algorithm diagnoses that everyone to the right of the line has BOS and that everyone to the left of the line does not have BOS. B, Figure depicts the clinical diagnosis assigned to each patient relative to the diagnosis by the machine learning algorithm comparing air trapping and expiratory volumes. C, Machine learning decision boundaries for diagnosing BOS from air trapping and inspiratory volume. D, Figure depicts the clinical diagnosis assigned to each patient relative to the diagnosis by the machine learning algorithm comparing air trapping and inspiratory volumes. The boundary was created from the 75% training set, and its performance is from the 25% test set. It is superimposed on all PRM scans to illustrate clinical application. Sixty-three of the 66 scans were included, with three scans excluded because of missing at least one volume measurement. See Figures 1 and 2 legends for expansion of abbreviations.*

report moderate negative correlation ($R = -0.59$) between air trapping and decline in FEV_1[14] and moderate correlation ($R = 0.663$) between air trapping and increased RV/TLC.[23] Our classification model also builds on prior work from the lung transplant literature, in which machine learning has been applied to the analysis of functional respiratory imaging to predict early BOS with an accuracy of 85% (combining volumes, airway resistance, and surface area of the right middle lobe, right upper lobe, and central airway, respectively).[24] Ultimately, comparison of BOS after HCT with BOS after lung transplantation can constitute a future study, because the latter represents a more progressive disease and is the most common long-term cause of mortality.[25]

We found that PRM measures were compelling when used to differentiate clinical diagnoses. Mean metrics differed significantly for BOS (increased air trapping, inspiratory and expiratory volumes) and for TS (decreased inspiratory volume). Machine learning algorithms applied without prior knowledge of patient characteristics identified similar distinctions in

TABLE 7] BOS Diagnosis by Machine Learning Models Compared With Thoracic Radiologists

	Diagnosis of BOS With Air Trapping and Expiratory Volume	Diagnosis of BOS With Air Trapping and Inspiratory Volume	Radiologist Diagnosis of BOS
Sensitivity	0.67	0.83	0.70
Specificity	0.88	0.88	0.60
PPV	0.80	0.83	0.53
NPV	0.78	0.88	0.76
Accuracy	0.79	0.86	0.64
F1 score	0.73	0.83	0.67
AUC (95% CI)	0.77 (0.53-1.0)	0.85 (0.65-1.0)	0.65 (0.54-0.76)

Radiologist diagnosis of BOS is shown for expiratory phase. AUC = area under the receiver operating characteristic curve; NPV = negative predictive value; PPV = positive predictive value. See Table 1 legend for expansion of other abbreviations.

physiology. Clustering analysis found two clusters representing BOS in its early or mild stages (cluster 2, 17% of all patients) and in its moderate to severe stages (cluster 3, 15% of all patients). In total, our models offer an approach to address current gaps in clinical care, such as diagnosis of early BOS, which is commonly delayed for many months by clinicians.[26]

Compared with this study, a prior study by Galbán et al[9] identified BOS from air trapping alone and did so with a similar accuracy (79% vs 75%, respectively), a lower specificity (88% vs 72%, respectively) and a higher sensitivity (67% vs 76%, respectively). Notable differences are present between these studies: (1) differences in comparison groups, because Galbán et al[9] evaluated BOS vs infection; the current study evaluated BOS vs a broader comparison (no lung disease, infection, fibrotic lung disease and other posttransplantation pulmonary complications, including TS and OP), which increases the difficulty of screening for BOS by air trapping alone; and (2) differences in statistical techniques, because Galbán et al[9] assessed a cross-validated model for classification on the entire cohort; the current study assessed a cross-validated model for classifier performance on an unexposed test set, a standard approach in machine learning to maximize external generalizability.[27] For our model, the addition of inspiratory volume improved accuracy and sensitivity with increasing statistical significance, suggesting that normal or low lung volumes likely represent a disease other than BOS.

Our study was subject to some limitations. First, our reliance on the experience of a single center may limit external generalizability. We addressed this by comparing PRM metrics with standardized measures, such as PFTs, by minimizing unexplained variance in clustering, by validating the machine learning model on an unexposed test set, and by comparing performance with radiologists. Future studies will validate our machine learning model in cGVHD patients at other academic medical centers. Second, we converted lung volumes by PRM to percentage predicted values. To our knowledge, such plethysmographic standards of converting volumes have not been validated specifically with the inhalation and exhalation maneuvers of PRM. In studies that employ PRM with COPD, computational output of the software is correlated with clinical disease state,[10,28] with the intention of PRM metrics acting as biomarkers. The approach of adjusting inspiratory volume by sex and height and expiratory volume by age, sex, and height is a unique feature of our study. It normalized the inclusion of volumes in statistical analysis and model analysis, given the dramatic differences in absolute values that can occur between, for example, shorter individuals and taller individuals. In this adjustment, we used the standard conversion equations as established by ATS/ERS guidelines.[16] Third, our small sample size introduces the possibility of random effects yielding a statistical signal. To mitigate the risk, we used multiple approaches for statistical inference in the relationship of PRM metrics to ground truth—correlation, comparison of means, PCA, clustering, and classification. A consistent clinical and statistical significance was seen for segregation of BOS patients from non-BOS patients in physiologically relevant ways. Moreover, to assess the robustness of the clustering model, we employed a method described by Hastie and colleagues[29] for the iterative removal of a known number of observations at random and subsequent determination of clustering results on the remaining subset of data. We found similar results as the final analysis. A mean total variance of 76.7% of the data was captured from the

subsets, as compared with 76.4% total variance captured in the composite analysis. Mean PRM metrics by cluster were largely unchanged (see e-Table 3 and Clustering Analysis section in e-Appendix 1 for details). To ensure that the training and test set for the classification model contained well balanced distributions of disease with similar physiological characteristics, we verified that disease states and physiological parameters across the sets did not differ (e-Table 4 in e-Appendix 1). To assess that the clustering model was detecting a new signal in the post-HCT context, we verified that pre-HCT demographics were comparable across clusters (e-Table 5 in e-Appendix 1). In contrast PRM metrics alone phenotyped patients in a manner that was validated by relevant posttransplant physiology, with disease states and all PFTs after HCT revealing differences that achieved group-wise statistical significance (e-Table 6 in e-Appendix 1). Fourth, selection bias may exist, because patients came to our attention only if they were referred to our lung GVHD clinic. This referral pattern may impact positive predictive value and consequent applicability in low prevalence settings.[30] Fifth, our study patients received multiple PFTs, with the PFT most proximal to their PRM being used for analysis. No clear guidelines are available for repeating PFTs in these patients, although studies show that repeat PFTs allow for earlier diagnosis of BOS[31,32] and are actionable.[31,33] Finally, our study is retrospective, with patients receiving HCT before study inclusion. However, data central to our analysis were collected on a rolling basis according to predefined methodology (eg, PFTs proximal to PRM scan), and all diagnoses of posttransplantation lung complications occurred during the study time span. Importantly, most of our BOS patients (94%) met the NIH clinical definition of BOS or were diagnosed by biopsy, and independent of any clinical information, clustering analysis identified 81% of patients with the diagnosis of BOS. Insofar as PFTs—and by extension PRM metrics—represent true physiology, this supports our clinical definitions.

Interpretation

Our study suggests that a machine learning-based taxonomy for pulmonary complications after HCT may be possible, allowing for more precise medical management for lung GVHD. By driving the process of discovery as an outgrowth of patient care, our models can be enhanced to assist in clinician decision-making across a spectrum of posttransplantation lung complications.

Acknowledgments

Author contributions: H. S. had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis, including and especially any adverse effects. H. S., Y. K. L., and J. H. contributed substantially to the study design, data analysis and interpretation, and the writing of the manuscript. H. G., M. H., Z. G., L. J., T. B., L. C., and M. N. contributed substantially to the study design, interpretation, and writing of the manuscript.

Financial/nonfinancial disclosures: None declared.

Role of sponsors: The sponsor had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript.

Other contributions: The authors thank and acknowledge the technicians in our center's 3-D lab, specifically Derrick Laurel, RT, who facilitated PRM processing. We also thank Dr Sally Glaser, PhD, for her critical review of the manuscript. Above all, we thank the patients in our lung GVHD clinic.

Additional information: The e-Appendix, e-Tables, and e-Figures can be found in the Supplemental Materials section of the online article.

References

1. Williams KM. How I treat bronchiolitis obliterans syndrome after hematopoietic stem cell transplantation. *Blood*. 2017;129(4):448-455.
2. Jagasia MH, Greinix HT, Arora M, et al. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: I. The 2014 Diagnosis and Staging Working Group Report. *Biol Blood Marrow Transplant*. 2015;21(3):389-401.
3. Abedin S, Yanik GA, Braun T, et al. Predictive value of bronchiolitis obliterans syndrome stage 0p in chronic graft-versus-host disease of the lung. *Biol Blood Marrow Transplant*. 2015;21(6):1127-1131.
4. Uhlving HH, Andersen CB, Christensen IJ, et al. Biopsy-verified bronchiolitis obliterans and other noninfectious lung pathologies after allogeneic hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2015;21(3):531-538.
5. Holbro A, Lehmann T, Girsberger S, et al. Lung histology predicts outcome of bronchiolitis obliterans syndrome after hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2013;19(6):973-980.
6. Gratwhol AA. Sjögren-type syndrome after allogeneic bone-marrow transplantation. *Ann Intern Med*. 1977;87(6):703.
7. Hildebrandt GC, Fazekas T, Lawitschka A, et al. Diagnosis and treatment of pulmonary chronic GVHD: report from the consensus conference on clinical practice in chronic GVHD. *Bone Marrow Transplant*. 2011;46(10):1283.
8. Williams KM, Chien JW, Gladwin MT, Pavletic SZ. Bronchiolitis obliterans after allogeneic hematopoietic stem cell transplantation. *JAMA*. 2009;302(3):306-314.
9. Galbán CJ, Boes JL, Bule M, et al. Parametric response mapping as an indicator of bronchiolitis obliterans syndrome after hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2014;20(10):1592-1598.
10. Galbán CJ, Han MK, Boes JL, et al. Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med*. 2012;18(11):1711-1715.
11. Boes JL, Hoff BA, Bule M, et al. Parametric response mapping monitors temporal changes on lung CT scans in the subpopulations and intermediate outcome

measures in COPD study (SPIROMICS). *Acad Radiol*. 2015;22(2):186-194.

12. Vasilescu DM, Martinez FJ, Marchetti N, et al. Non-invasive imaging biomarker identifies small airway damage in severe COPD. *Am J Respir Crit Care Med*. 2019;200(5):575-581.
13. Belloli EA, Degtiar I, Wang X, et al. Parametric response mapping as an imaging biomarker in lung transplant recipients. *Am J Respir Crit Care Med*. 2017;195(7):942-952.
14. Verleden SE, Vos R, Vandermeulen E, et al. Parametric response mapping of bronchiolitis obliterans syndrome progression after lung transplantation. *Am J Transplant*. 2016;16(11):3262-3269.
15. Gunn MLD, Godwin JD, Kanne JP, Flowers ME, Chien JW. High-resolution CT findings of bronchiolitis obliterans syndrome after hematopoietic stem cell transplantation. *J Thorac Imaging*. 2008;23(4):244-250.
16. Stocks J, Quanjer PhH. Reference values for residual volume, functional residual capacity and total lung capacity. *Eur Respir J*. 1995;8(3):492-506.
17. Sorror ML. Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT. *Blood*. 2005;106(8):2912-2919.
18. Plotly Chart Studio. https://plot.ly/~lungGVH/7/#/. Accessed June 5, 2020.
19. Lee E-S, Gotway MB, Reddy GP, Golden JA, Keith FM, Webb WR. Early bronchiolitis obliterans following lung transplantation: accuracy of expiratory thin-section CT for diagnosis. *Radiology*. 2000;216(2):472-477.
20. Siegel MJ, Bhalla S, Gutierrez FR, Hildebolt C, Sweet S. Post-lung transplantation bronchiolitis obliterans syndrome: usefulness of expiratory thin-section CT for diagnosis. *Radiology*. 2001;220(2):455-462.
21. Nakasone H, Onizuka M, Suzuki N, et al. Pre-transplant risk factors for cryptogenic organizing pneumonia/bronchiolitis obliterans organizing pneumonia after hematopoietic cell transplantation. *Bone Marrow Transplant*. 2013;48(10):1317-1323.
22. Apisarnthanarax N, Donato M, Körbling M, et al. Extracorporeal photopheresis therapy in the management of steroid-refractory or steroid-dependent cutaneous chronic graft-versus-host disease after allogeneic stem cell transplantation: feasibility and results. *Bone Marrow Transplant*. 2003;31(6):459-465.
23. Solyanik O, Hollmann P, Dettmer S, et al. Quantification of pathologic air trapping in lung transplant patients using CT density mapping: comparison with other CT air trapping measures. *PLoS One*. 2015;10(10):e0139102.
24. Barbosa EJM, Lanclus M, Vos W, et al. Machine learning algorithms utilizing quantitative CT features may predict eventual onset of bronchiolitis obliterans syndrome after lung transplantation. *Acad Radiol*. 2018;25(9):1201-1212.
25. Yusen RD, Edwards LB, Dipchand AI, et al. The Registry of the International Society for Heart and Lung Transplantation: Thirty-third Adult Lung and Heart-Lung Transplant Report—2016; Focus theme: primary diagnostic indications for transplant. *J Heart Lung Transplant*. 2016;35(10):1170-1184.
26. Cheng G-S, Storer B, Chien JW, et al. Lung function trajectory in bronchiolitis obliterans syndrome after allogeneic hematopoietic cell transplant. *Ann Am Thorac Soc*. 2016;13(11):1932-1939.
27. Bzdok D, Krzywinski M, Altman N. Machine learning: a primer. *Nat Methods*. 2017;14(12):1119-1120.
28. Pompe E, Galbán CJ, Ross BD, et al. Parametric response mapping on chest computed tomography associates with clinical and functional parameters in chronic obstructive pulmonary disease. *Respir Med*. 2017;123:48-55.
29. Hastie T, Tibshirani R, Friedman J. *Unsupervised learning. In: The Elements of Statistical Learning*. New York, NY: Springer; 2009.
30. Leeflang MMG, Rutjes AWS, Reitsma JB, Hooft L, Bossuyt PMM. Variation of a test's sensitivity and specificity with disease prevalence. *Can Med Assoc J*. 2013;185(11):E537-544.
31. Bergeron A, Chevret S, Chagnon K, et al. Budesonide/formoterol for bronchiolitis obliterans after hematopoietic stem cell transplantation. *Am J Respir Crit Care Med*. 2015;191(11):1242-1249.
32. Kuzmina Z, Krenn K, Petkov V, et al. CD19+CD21low B cells and patients at risk for NIH-defined chronic graft-versus-host disease with bronchiolitis obliterans syndrome. *Blood*. 2013;121(10):1886-1895.
33. Williams KM, Cheng G-S, Pusic I, et al. Fluticasone, azithromycin, and montelukast treatment for new-onset bronchiolitis obliterans syndrome after hematopoietic cell transplantation. *Biol Blood Marrow Transplant*. 2016;22(4):710-716.