Table 2. Pulmonary vascular pressure values before and after esophageal pressure correction

	Mean (SD)
Pulmonary artery pressure—whole breath	
Mean pulmonary artery pressure	40.4 (15.8)
Esophageal pressure	1.0 (2.8)
Corrected pulmonary artery pressure	39.4 (15.7)
PAWP—end expiratory	
PAWP	19.1 (7.5)
Esophageal pressure	4.7 (3.8)
Corrected PAWP	14.4 (7.7)
PAWP—whole breath	
PAWP	13.1 (7.2)
Esophageal pressures	0.95 (2.9)
Corrected PAWP	12.1 (7.0)

Definition of abbreviation: PAWP = pulmonary artery wedge pressure.

Author disclosures are available with the text of this article at www.atsjournals.org.

Ahmed Jawad, M.D.
 Adriano R. Tonelli, M.D., M.Sc.
 Robert L. Chatburn, M.H.H.S., R.R.T.-N.P.S.
 Xiaofeng Wang, Ph.D.
 Umur Hatipoğlu, M.D.
 Cleveland Clinic
 Cleveland, Ohio

ORCID ID: 0000-0002-2321-9545 (A.R.T.).

References

- Behazin N, Jones SB, Cohen RI, Loring SH. Respiratory restriction and elevated pleural and esophageal pressures in morbid obesity. *J Appl Physiol* (1985) 2010;108:212–218.
- Owens RL, Campana LM, Hess L, Eckert DJ, Loring SH, Malhotra A. Sitting and supine esophageal pressures in overweight and obese subjects. *Obesity (Silver Spring)* 2012;20:2354–2360.
- Akoumianaki E, Maggiore SM, Valenza F, Bellani G, Jubran A, Loring SH, Pelosi P, Talmor D, Grasso S, Chiumello D, et al.; PLUG Working Group (Acute Respiratory Failure Section of the European Society of Intensive Care Medicine). The application of esophageal pressure measurement in patients with respiratory failure. *Am J Respir Crit Care Med* 2014;189:520–531.
- Baydur A, Behrakis PK, Zin WA, Jaeger M, Milic-Emili J. A simple method for assessing the validity of the esophageal balloon technique. *Am Rev Respir Dis* 1982;126:788–791.
- Hoepfer MM, Bogaard HJ, Condliffe R, Frantz R, Khanna D, Kurzyna M, Langleben D, Manes A, Satoh T, Torres F, et al. Definitions and diagnosis of pulmonary hypertension. *J Am Coll Cardiol* 2013;62(25, suppl):D42–D50.
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–381.
- Parameswaran K, Todd DC, Soth M. Altered respiratory physiology in obesity. *Can Respir J* 2006;13:203–210.
- Sharp JT, Henry JP, Sweany SK, Meadows WR, Pietras RJ. Effects of mass loading the respiratory system in man. *J Appl Physiol* 1964;19:959–966.
- Pankow W, Podszus T, Gutheil T, Penzel T, Peter J, Von Wichert P. Expiratory flow limitation and intrinsic positive end-expiratory pressure in obesity. *J Appl Physiol* (1985) 1998;85:1236–1243.
- Ferretti A, Giampiccolo P, Cavalli A, Milic-Emili J, Tantucci C. Expiratory flow limitation and orthopnea in massively obese subjects. *Chest* 2001;119:1401–1408.
- Rochester DF, Enson Y. Current concepts in the pathogenesis of the obesity-hypoventilation syndrome. Mechanical and circulatory factors. *Am J Med* 1974;57:402–420.
- Friedman SE, Andrus BW. Obesity and pulmonary hypertension: a review of pathophysiologic mechanisms. *J Obes* 2012;2012:505274.
- Weyman AE, Davidoff R, Gardin J, Ryan T, St John Sutton M, Weissman NJ. Echocardiographic evaluation of pulmonary artery pressure with clinical correlates in predominantly obese adults. *J Am Soc Echocardiogr* 2002;15:454–462.
- Kovacs G, Avian A, Pienn M, Naeije R, Olschewski H. Reading pulmonary vascular pressure tracings: how to handle the problems of zero leveling and respiratory swings. *Am J Respir Crit Care Med* 2014;190:252–257.
- Rice DL, Awe RJ, Gaasch WH, Alexander JK, Jenkins DE. Wedge pressure measurement in obstructive pulmonary disease. *Chest* 1974;66:628–632.
- Boerrigter BG, Waxman AB, Westerhof N, Vonk-Noordegraaf A, Systrom DM. Measuring central pulmonary pressures during exercise in COPD: how to cope with respiratory effects. *Eur Respir J* 2014;43:1316–1325.
- LeVarge BL, Pomerantsev E, Channick RN. Reliance on end-expiratory wedge pressure leads to misclassification of pulmonary hypertension. *Eur Respir J* 2014;44:425–434.
- Galiè N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, Simonneau G, Peacock A, Vonk Noordegraaf A, Beghetti M, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS); endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Respir J* 2015;46:903–975.
- Pecchiari M, Loring SH, D'Angelo E. Esophageal pressure as an estimate of average pleural pressure with lung or chest distortion in rats. *Respir Physiol Neurobiol* 2013;186:229–235.
- Washko GR, O'Donnell CR, Loring SH. Volume-related and volume-independent effects of posture on esophageal and transpulmonary pressures in healthy subjects. *J Appl Physiol* (1985) 2006;100:753–758.

Copyright © 2017 by the American Thoracic Society

Improving Care for Patients with Interstitial Lung Disease Using Machine Learning Requires Transparency and Reproducibility

To the Editor:

I read with interest the recent article by Pankratz and colleagues (1) that tested machine-learning methods to identify usual interstitial pneumonia using tissue obtained by transbronchial biopsy (TBBx). The authors used RNA sequence data from

TBBx specimens to enrich the performance of a classification model that, in the future, might obviate a more invasive surgical lung biopsy.

The authors report that RNA sequence data was “input to machine learning algorithms” and “feature selection and parameter estimation were performed by logistic regression with elastic net penalty.” Was elastic net regression the only statistical model type used to build a classifier? If other machine-learning algorithms were used, as is implied, which ones? The authors report using the R language for statistical

computing, which currently supports interfaces to more than 100 such machine-learning algorithms (2). Parameter tuning was performed using cross-validation, but which parameters were tuned, and for which models? This lack of reported detail is not in accordance with best practices for reproducible research using computational methods or gene expression data (3, 4).

In the results section, the authors report the sensitivity and specificity of various models with wide 95% confidence intervals. The area under the curve is reported without any 95% confidence intervals. A new pooling analysis is also introduced in the results section. Although the point estimates of the area under the curve for this analysis are higher than in the primary analysis, they are also presented without confidence intervals. Furthermore, the online supplement indicates that the pooling analysis used all available subjects without split samples for training and validation, thus differing from the primary analysis. Although the sensitivity from these analyses appears quite optimistic when presented as box plots, there are no checks for overfitting, for which this analysis is at high risk, given the small sample size and large number of predictor variables. Perhaps most importantly, no statistical tests are performed on these performance measures to support a conclusion of improved performance.

Although the authors rightly describe their model as a “classifier” in the text, the title of the article indicates that the model is used for detection. Fibrotic lung disease was likely already detected by imaging studies to be eligible for enrollment. The remaining diagnostic uncertainty that the reported model seeks to overcome would require nuanced classification for a particular type of interstitial lung disease, rather than outright detection. Computational approaches to the tasks of detection (5) and classification (6) are distinct, each with different clinical applications and analytic methods.

Finally, only 72% of the analytic sample had a gold standard histopathologic diagnosis by surgical lung biopsy. Including labeled observations in the training set from TBBx and cryobiopsy specimens does not provide useful information to answer the clinical question under investigation. Furthermore, the analysis does not provide a comparison of the incremental classification performance gained through use of RNA sequence data compared with expert histopathologic diagnosis of the TBBx specimens alone. Nor is there a comparison of how the machine-learning approach, albeit not specifying as to which

one(s), compared with traditional logistic regression. Especially in the case of an industry-funded study, such transparency around incremental performance relative to currently available methods is necessary when carefully weighing trade-offs in costly diagnostic tests, invasive procedures, and treatments with significant side effects.

In summary, the authors are tackling an important question by leveraging RNA technology and machine-learning approaches to improve the care of patients with interstitial lung disease. However, clinicians and patients making difficult clinical decisions under uncertainty are best served by scientific results that can be widely replicated and validated.

Author disclosures are available with the text of this letter at www.atsjournals.org.

Gary E. Weissman, M.D., M.S.H.P.
University of Pennsylvania
Philadelphia, Pennsylvania

ORCID ID: 0000-0001-9588-3819 (G.E.W.).

References

- 1 Pankratz DG, Choi Y, Imtiaz U, Fedorowicz GM, Anderson JD, Colby TV, Myers JL, Lynch DA, Brown KK, Flaherty KR, *et al*. Usual interstitial pneumonia can be detected in transbronchial biopsies using machine learning. *Ann Am Thorac Soc* 2017; 14:1646–1654.
- 2 Hothorn T. CRAN task view: machine learning & statistical learning. 2017 Jun 30 [accessed 2017 Jul 10]. Available from: <https://CRAN.R-project.org/view=MachineLearning>
- 3 Peng RD. Reproducible research in computational science. *Science* 2011;334:1226–1227.
- 4 Ioannidis JPA, Allison DB, Ball CA, Coulibaly I, Cui X, Culhane AC, Falchi M, Furlanello C, Game L, Jurman G, *et al*. Repeatability of published microarray gene expression analyses. *Nat Genet* 2009;41: 149–155.
- 5 Park SC, Tan J, Wang X, Lederman D, Leader JK, Kim SH, Zheng B. Computer-aided detection of early interstitial lung diseases using low-dose CT images. *Phys Med Biol* 2011;56:1139–1153.
- 6 Xu Y, van Beek EJ, Hwanjo Y, Guo J, McLennan G, Hoffman EA. Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM). *Acad Radiol* 2006; 13:969–978.

Copyright © 2017 by the American Thoracic Society

Reply: Improving Care for Patients with Interstitial Lung Disease, Using Machine Learning, Requires Transparency and Reproducibility

From the Authors:

We are encouraged by Dr. Weissman’s interest in our study (1). To enhance readability, we de-emphasized the analytical background in the manuscript (2). Here, we briefly outline the robust analytical strategy utilized. A detailed description of algorithm development will be published separately (3).

We evaluated random forest, support vector machine, gradient boosting, neural network, and penalized logistic regression models, as well as ensembles of these models (4). We chose a penalized logistic regression model based on cross-validation performance

and tuned the L1 and L2 penalty parameters within each cross-validation fold. We used a LASSO (least absolute shrinkage and selection operator) penalty to select the genes in our model, as it is not feasible to perform linear regression on more than 20,000 genes when training is limited to hundreds of samples; there would likely be overfitting. We internally validated all code according to best practices, and multiple internal analysts reviewed the results. We computed areas under the curve (AUCs) and associated 95% confidence intervals of 0.86 (0.79–0.93) and 0.92 (0.81–1.00) for sample-level and subject-level evaluation, respectively, on the independent test set.

The *in silico* mixing simulation of two to five transbronchial biopsies (TBBs) within subjects was performed by averaging normalized counts. Experimental noise was estimated using samples run in replicates across several sequencing plates and added at the