Turning subtypes into disease axes to improve prediction of COPD progression

Junxiang Chen,[1] Michael Cho,[2,3] Edwin K Silverman,[2,3] John E Hokanson,[4] Greg L Kinney,[4] James D Crapo,[5] Stephen Rennard,[6,7] Jennifer Dy,[1] Peter Castaldi[2,8]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2018-213005).

[1]Department of Electrical and Computer Engineering, Northeastern University, Boston, United States
[2]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, United States
[3]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, United States
[4]Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado, USA
[5]Department of Medicine, National Jewish Health, Denver, Colorado, USA
[6]Division of Pulmonary, Critical Care, Sleep and Allergy, University of Nebraska Medical Center, Omaha, United States
[7]IMED Biotech Unit, AstraZeneca, Cambridge, United Kingdom
[8]Division of Primary Care and Internal Medicone, Brigham and Women's Hospital, Boston, United States

Correspondence to
Dr Peter Castaldi, Brigham and Women's Hospital Department of Medicine, Boston, MA 02115, USA;
peter.castaldi@channing.harvard.edu

Received 18 December 2018
Revised 6 May 2019
Accepted 13 May 2019
Published Online First 12 June 2019

To cite: Chen J, Cho M, Silverman EK, *et al*. *Thorax* 2019;**74**:906–909.

ABSTRACT

Chronic obstructive pulmonary disease (COPD) is an umbrella definition encompassing multiple disease processes. COPD heterogeneity has been described as distinct subgroups of individuals (subtypes) or as continuous measures of COPD variability (disease axes). There is little consensus on whether subtypes or disease axes are preferred, and the relative value of disease axes and subtypes for predicting COPD progression is unknown. Using a propensity score approach to learn disease axes from pairs of subtypes, we demonstrate that these disease axes predict prospective forced expiratory volume in 1 s decline and emphysema progression more accurately than the subtype pairs from which they were derived.

INTRODUCTION

The heterogeneity of chronic obstructive pulmonary disease (COPD) obscures our understanding of its natural history and molecular mechanisms. COPD heterogeneity is often represented as distinct subgroups of subjects (subtypes), but it can also be represented as continuous axes of variability, that is, disease axes.[1] A multicohort study demonstrated that subtypes identified by clustering were not reproducible across cohorts, whereas disease axes from the same cohorts were more consistent.[2] There is currently no consensus on the best approach to characterise COPD heterogeneity.

We define a subtype as a single subgroup of subjects and a COPD disease axis as any continuous representation of COPD heterogeneity. We describe a method, similar in concept to propensity scores,[3] where a pair of COPD subtypes can be used to define a single disease axis by using the subtype pair as the response in a logistic regression model that predicts the likelihood of subtype membership. These predictions constitute a subtype-defined disease axis. For example, in the case of chronic bronchitis (CB), the CB subtype is a binary yes/no classification based on patient symptoms. Conversely, the CB disease axis is a continuous measure derived from a predictive model that describes the propensity of each subject to have CB. Using longitudinal data from the Genetic Epidemiology of COPD (COPDGene) Study, we demonstrate that subtype-defined disease axes provide better prediction of prospective COPD progression than the original subtype pairs from which they were derived.

METHODS

Subjects in COPDGene with complete 5-year follow-up data were analysed (n=4726). Four general subtype classes were selected for study: CB per the American Thoracic Society for the Division of Lung Diseases (ATS-DLD) definition,[4] the pink puffer (PP)/blue bloater (BB) subtype,[5] frequent exacerbators (≥2 COPD exacerbations over the previous 12 months)[6] and upper/lower lobe emphysema predominant subjects with a log U/L ratio >1.5 for upper lobe or <-1.5 for lower lobe predominance. We refer to a subtype pair as two subtypes that are conceptually related and therefore used to construct a disease axis. For example, the CB subtype class yields a single subtype pair (CB present vs absent), whereas the PP/BB subtype class yields two pairs (PP/neither and BB/neither).

For each subtype pair, we used weighted logistic regression to identify a linear combination of predictors that provide optimal classification for that pair. The beta coefficients of this regression were used to calculate the disease axis value for each analysed subject. This software is available at https://github.com/Chen-Jxiang/SODA. We selected the baseline values of 27 variables to serve as the predictors in the regression models (see online supplementary materials for variables used). Disease axes were generated only from visit 1 data.

For the analyses of COPD progression, separate regression models were used to relate subtypes or disease axis scores to either 5-year change in forced expiratory volume in 1 s (FEV_1%) of predicted or change in emphysema. To formally test for whether disease axes provide incremental improvement in prediction beyond that provided by a subtype pair, we constructed nested regression models in which a disease axis was added to a base model containing the original subtype pair. Additional information is included in the online supplementary file 1.

RESULTS

A conceptual overview of our approach is shown in figure 1. Subtype definitions and characteristics of the subjects are shown in online supplementary tables 1 and 2. One disease axis was identified for each subtype pair, resulting in a total of six disease axes (one each for frequent/non-frequent exacerbators and presence/absence of CB, and two each for the PP/BB and upper/lower emphysema subtypes). To determine how well the disease axes could correctly classify their original subtypes, we examined the discrimination performance which was excellent for the PP/BB and upper/lower emphysema subgroups (Area under the receiver operating characteristic curve (AUC) >0.98), and reasonable for the frequent exacerbator (AUC=0.79) and CB subgroups (AUC=0.67). The predictors and beta coefficients from these models are shown in online supplementary tables 3 and 4.

We then studied how well each subtype pair and

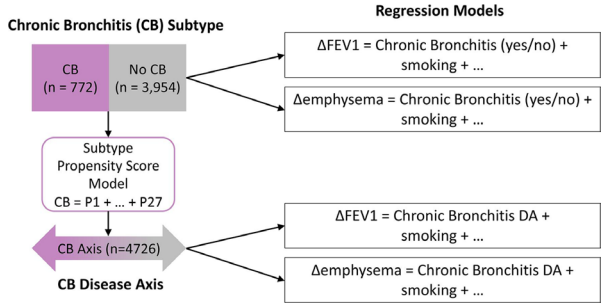

Figure 1 Overview of subtype-oriented disease axis approach. Chronic bronchitis (CB) subtypes are used as the response variable for a predictive model that uses 27 predictors (P1–P27) to classify subjects into the proper CB subtype group. The resulting predicted values constitute a continuous CB disease axis. Both subtype assignment and disease axis values are used as predictors in separate regression models in which 5-year change in FEV_1 or emphysema serve as the response variables. FEV_1, forced expiratory volume in 1 s.

its respective disease axis could predict two measures of COPD progression, change in FEV_1 and quantitative CT emphysema progression. For both outcomes, we observed that regression models containing disease axes as predictors explained a greater proportion of the variance of COPD progression than similar models containing the subtype pair, with particularly marked improvement noted for emphysema progression (table 1). To formally test for significant improvement in prediction from disease axes, for each subtype-containing model, we added the corresponding disease axes and compared the two models (table 2).

We also examined how well baseline disease axis values predicted the consistency of subtype assignment over time, which is an important issue for the CB and frequent exacerbator subtypes. We classified subjects as persistent or intermittent members of these subtypes according to their status at both COPDGene Study visits, and we observed that persistent subjects had higher disease axis values than intermittent subjects (online supplementary figures 1 and 2, $p<0.001$ for CB and $p=0.007$ for frequent exacerbators).

DISCUSSION

Previous work has shown that COPD variability typically occurs along a continuum.[2] Thus, while subtypes may have intuitive appeal, disease axes are more accurate. The method presented here turns subtypes into disease axes, providing representations of COPD heterogeneity that represent a continuum defined by two COPD subtypes. These disease axes were more predictive of COPD progression than the subtypes from which they were derived; because (1) disease axes 'expand' subtype information to all subjects in a dataset and (2) disease axes extract subtype-related information from a large number of input variables and thus contain more COPD-related information than subtypes alone.

Since this method uses predefined subtypes to guide data-driven analysis, the strengths of this approach are the interpretability of the disease axes and the improved prediction of disease progression. However, when the sole goal is prediction, purely data-driven methods may yield superior performance. These disease axes were generated in a single cohort, so independent

Table 1 Regression models using either subtypes or disease axes to predict change in emphysema or change in FEV_1% of predicted

			Subtype models			Disease axis models		
Progression measure	Subtype class	Subtype pair	Beta (SE)	P value	% variance explained	Beta (SE)	P value	% variance explained
Δ emphysema (Perc15)	Chronic bronchitis	Chronic bronchitis (no vs yes)	−1.1 (0.5)	0.05	7.5	−8.0 (0.5)	<0.001	12.8
	Frequent exacerbator	Frequent exacerbators (no vs yes)	0.3 (0.8)	0.68	7.5	−2.1 (0.4)	<0.001	8.0
	Pink puffer (PP)/blue bloater (BB)	PP/BB (neither vs PP)	−2.7 (2.0)	0.17	7.6	−1.2 (0.2)	<0.001	8.5
		PP/BB (neither vs BB)	−7.8 (3.2)	0.01		−0.3 (0.2)	0.07	
	Upper/lower emphysema	Upper/lower emphysema (neither vs LLE)	−6.9 (2.1)	0.001	8.3	6.7 (0.4)	<0.001	12.8
		Upper/lower Emphysema (neither vs ULE)	−4.7 (0.8)	<0.001		−5.1 (0.3)	<0.001	
Δ FEV1 % of predicted	Chronic bronchitis	Chronic bronchitis (no vs yes)	−1.9 (0.4)	<0.001	6.0	−2.5 (0.4)	<0.001	6.4
	Frequent exacerbator	Frequent exacerbators (no vs yes)	−1.9 (0.6)	0.002	5.8	−2.2 (0.3)	<0.001	6.5
	PP/BB	PP/BB (neither vs PP)	−3.5 (1.5)	0.03	5.7	−0.7 (0.1)	<0.001	6.4
		PP/BB (neither vs BB)	−3.2 (2.5)	0.19		−0.6 (0.1)	<0.001	
	Upper/lower emphysema	Upper/lower emphysema (neither vs LLE)	−0.7 (1.7)	0.68	5.7	−0.6 (0.3)	0.09	5.8
		Upper/lower emphysema (neither vs ULE)	−2.3 (0.6)	<0.001		0.3 (0.2)	0.19	

For each outcome and subtype class, two regression models were constructed including either the subtype(s) or disease axes as predictors. The COPD progression outcomes were either change in FEV1 % of predicted or change in emphysema between visit 1 and 2. Values in the subtype model columns are from models that include categorical subtype assignment as a predictor, and values in the disease axis columns are from models that include the corresponding disease axes as a predictor. All models also include baseline FEV1% predicted, baseline emphysema, and current smoking status at visit 1 and 2. PP/BB and upper/lower emphysema subtypes have three categories and thus two contrasts are included in the same model for each subtype group.
Perc15, 15th percentile of the lung density histogram.
COPD, chronic obstructive pulmonary disease; FEV1, forced expiratory volume in 1 s; LLE, lower lobe emphysema predominant subtype; ULE, upper lobe emphysema predominant subtype.

Table 2 Regression models using both subtypes and disease axes to predict change in emphysema or change in FEV1

			Subtypes		Disease axes		% variance explained
Progression measure	**Subtype class**	**Subtype pair**	**Beta (SE)**	**P value**	**Beta (SE)**	**P value**	
Δ emphysema (Perc15)	Chronic bronchitis	Chronic bronchitis (no vs yes)	−0.1 (0.5)	0.78	−8.0 (0.5)	<0.001	12.8
	Frequent exacerbator	Frequent exacerbators (no vs yes)	0.9 (0.8)	0.27	−2.2 (0.4)	<0.001	8.0
	Pink puffer (PP)/blue bloater(BB)	PP/BB (neither vs PP)	−0.2 (2.0)	0.92	−1.2 (0.2)	<0.001	8.6
		PP/BB (neither vs BB)	−8.2 (3.2)	0.01	−0.2 (0.2)	0.21	
	Upper/lower emphysema	Upper/lower emphysema (neither vs LLE)	−1.1 (2.2)	0.60	−3.0 (0.2)	<0.001	13.9
		Upper/lower Emphysema (neither vs ULE)	−0.20 (1.0)	0.84	−2.6 (0.2)	<0.001	
Δ FEV1% of predicted	Chronic bronchitis	Chronic bronchitis (no vs yes)	−1.7 (0.4)	<0.001	−2.3 (0.4)	<0.001	6.8
	Frequent exacerbator	Frequent exacerbators (no vs yes)	−1.3 (0.6)	0.02	−2.1 (0.3)	<0.001	6.6
	rPP/BB	PP/BB (neither vs PP)	−2.3 (1.6)	0.20	−0.7 (0.1)	<0.001	6.4
		PP/BB (neither vs BB)	−1.4 (2.5)	0.59	−0.6 (0.1)	<0.001	
	Upper/lower emphysema	Upper/lower emphysema (neither vs LLE)	−0.8 (1.7)	0.63	0.2 (0.1)	0.20	5.9
		Upper/lower emphysema (neither vs ULE)	−2.8 (0.8)	<0.001	0.2 (0.1)	0.13	

For each outcome and subtype class, one regression model was constructed for each subtype class. This model contained the relevant categorical subtypes as well as the corresponding disease axes. The COPD progression outcomes were either change in FEV1 % of predicted or change in emphysema between visit 1 and 2. All models also include baseline FEV1% predicted, baseline emphysema, and current smoking status at visit 1 and 2. PP/BB and upper/lower emphysema subtypes have three categories and thus two contrasts are included in the same model for each subtype group, respectively.
Perc15, 15th percentile of the lung density histogram.
COPD, chronic obstructive pulmonary disease; FEV1, forced expiratory volume in 1 s; LLE, lower lobe emphysema predominant subtype; ULE, upper lobe emphysema predominant subtype.

assessment of their generalisability is needed. These results provide proof of concept that subtype-defined disease axes provide more powerful prediction of COPD progression. In the future, it would be useful to define disease axes that can be produced from readily available variables, which would allow disease axes to be generated in a larger set of COPD studies.

In summary, relative to subtypes, disease axes provide more accurate clinical predictions, and in the future, disease axes may improve our clinical characterisation of COPD and enable more powerful biological discovery.

Collaborators Administrative Center: James D. Crapo; Edwin K. Silverman; Barry J. Make; Elizabeth A. Regan; Genetic Analysis Center: Terri Beaty; Ferdouse Begum; Peter J. Castaldi; Michael Cho; Dawn L. DeMeo; Adel R. Boueiz; Marilyn G. Foreman; Eitan Halper-Stromberg; Lystra P. Hayden; Craig P. Hersh; Jacqueline Hetmanski; Brian D. Hobbs; John E. Hokanson; Nan Laird; Christoph Lange; Sharon M. Lutz; Merry-Lynn McDonald; Margaret M. Parker; Dandi Qiao; Elizabeth A. Regan; Edwin K. Silverman; Emily S. Wan; Sungho Won; Phuwanat Sakornsakolpat; Dmitry Prokopenko; Imaging Center: Mustafa Al Qaisi; Harvey O. Coxson; Teresa Gray; MeiLan K. Han; Eric A. Hoffman; Stephen Humphries; Francine L. Jacobson; Philip F. Judy; Ella A. Kazerooni; Alex Kluiber; David A. Lynch; John D. Newell; Elizabeth A. Regan; James C. Ross; Raul San Jose Estepar; Joyce Schroeder; Jered Sieren; Douglas Stinson; Berend C. Stoel; Juerg Tschirren; Edwin Van Beek; Bram van Ginneken; Eva van Rikxoort; George Washko; Carla G. Wilson; PFT QA Center, Salt Lake City, UT: Robert Jensen; Data Coordinating Center and Biostatistics, National Jewish Health, Denver, CO: Douglas Everett; Jim Crooks; Camille Moore; Matt Strand; Carla G. Wilson; Epidemiology Core, University of Colorado Anschutz Medical Campus, Aurora, CO: John E. Hokanson; John Hughes; Gregory Kinney; Sharon M. Lutz; Katherine Pratte; Kendra A. Young; Mortality Adjudication Core: Surya Bhatt; Jessica Bon; MeiLan K. Han; Barry Make; Carlos Martinez; Susan Murray; Elizabeth Regan; Xavier Soler; Carla G. Wilson; Biomarker Core: Russell P. Bowler; Katerina Kechris; Farnoush Banaei-Kashani; COPDGene® Investigators – Clinical Centers: Ann Arbor VA: Jeffrey L. Curtis; Carlos H. Martinez; Perry G. Pernicano; Baylor College of Medicine, Houston, TX: Nicola Hanania; Philip Alapat; Mustafa Atik; Venkata Bandi; Aladin Boriek; Kalpatha Guntupalli; Elizabeth Guy; Arun Nachiappan; Amit Parulekar; Brigham and Women's Hospital, Boston, MA: Dawn L. DeMeo; Craig Hersh; Francine L. Jacobson; George Washko; Columbia University, New York, NY: R. Graham Barr; John Austin; Belinda D'Souza; Gregory D.N. Pearson; Anna Rozenshtein; Byron Thomashow; Duke University Medical Center, Durham, NC: Neil MacIntyre; H. Page McAdams; Lacey Washington; Health Partners Research Institute, Minneapolis: Charlene McEvoy; Joseph Tashjian; Johns Hopkins University, Baltimore: Robert Wise; Robert Brown; Nadia N. Hansel; Karen Horton; Allison Lambert; Nirupama Putcha; Los Angeles Biomedical Research Institute at Harbor UCLA Medical Center, Torrance, CA: Richard Casaburi; Alessandra Adami; Matthew Budoff; Hans Fischer; Janos Porszasz; Harry Rossiter; William Stringer; Michael E. DeBakey VAMC, Houston, TX: Amir Sharafkhaneh; Charlie Lan; Minneapolis VA: Christine Wendt; Brian Bell; Morehouse School of Medicine, Atlanta, GA: Marilyn G. Foreman; Eugene Berkowitz; Gloria Westney; National Jewish Health, Denver, CO: Russell Bowler; David A. Lynch; Reliant Medical Group, Worcester, MA: Richard Rosiello; David Pace; Temple University, Philadelphia, PA: Gerard Criner; David Ciccolella; Francis Cordova; Chandra Dass; Gilbert D'Alonzo; Parag Desai; Michael Jacobs; Steven Kelsen; Victor Kim; A. James Mamary; Nathaniel Marchetti; Aditi Satti; Kartik Shenoy; Robert M. Steiner; Alex Swift; Irene Swift; Maria Elena Vega-Sanchez; University of Alabama, Birmingham, AL: Mark Dransfield; William Bailey; Surya Bhatt; Anand Iyer; Hrudaya Nath; J. Michael Wells; University of California, San Diego, CA: Joe Ramsdell; Paul Friedman; Xavier Soler; Andrew Yen; University of Iowa, Iowa City, IA: Alejandro P. Comellas; Karin F. Hoth; John Newell; Brad Thompson; University of Michigan, Ann Arbor, MI: MeiLan K. Han; Ella Kazerooni; Carlos H. Martinez; University of Minnesota, Minneapolis, MN: Joanne Billings; Abbie Begnaud; Tadashi Allen; University of Pittsburgh, Pittsburgh, PA: Frank Sciurba; Jessica Bon; Divay Chandra; Carl Fuhrman; Joel Weissfeld; University of Texas Health Science Center at San Antonio, San Antonio, TX: Antonio Anzueto; Sandra Adams; Diego Maselli-Caceres; Mario E. Ruiz.

Contributors Study concept and design: JC, GLK, JEH and PC. Acquisition, analysis or interpretation of data: JC, MC, EKS, JDC, SR, JD and PC. Drafting of the manuscript: JC and PC. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: JC, JD and PC. Obtained funding: EKS, JDC and PC.

Funding The COPDGene project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens and Sunovion. The project described was supported by award number U01 HL089897 and award number U01 HL089856 from the National Heart, Lung and Blood Institute.

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health.

Competing interests JC reports consulting fees and grant support from GSK and Novartis outside the submitted work. MC has received grant support from GSK.

In the past 3 years, EKS received honoraria from Novartis for Continuing Medical Education Seminars and grant and travel support from GlaxoSmithKline.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Kinney GL, Santorico SA, Young KA, *et al*. Identification of chronic obstructive pulmonary disease axes that predict all-cause mortality: the COPDGene study. *Am J Epidemiol* 2018;187:2109–16.
2 Castaldi PJ, Benet M, Petersen H, *et al*. Do COPD subtypes really exist? COPD heterogeneity and clustering in 10 independent cohorts. *Thorax* 2017;72:998–1006.
3 ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
4 Ferris BG, Project ES. American Thoracic Society). *Am Rev Respir Dis* 1978;118:1–120.
5 Hersh CP, Make BJ, Lynch DA, *et al*. Non-emphysematous chronic obstructive pulmonary disease is associated with diabetes mellitus. *BMC Pulm Med* 2014;14.
6 Hurst JR, Vestbo J, Anzueto A, *et al*. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med* 2010;363:1128–38.