borders; fissural nodularity; ground-glass opacities; mediastinal adenopathy; and pleural effusions (4).

Second, pulmonary involvement by KS in immunocompetent patients is extremely rare, as stated by Chapman and colleagues (1), and the finding of cavitated lesions is even rarer. Although the authors noted that 10% of patients with KS present with cavitations, based on the literature (3), this percentage cannot be correlated with the case in question (an immunocompetent patient with endemic KS), because it pertains to HIV-positive patients.

Third, and most important, the authors did not indicate whether a histopathological analysis of the pulmonary nodules was performed. The differential diagnosis of multiple cavitated nodules is broad and includes neoplastic diseases (e.g., metastases and lymphomas), infectious diseases (e.g., septic embolism and granulomatous diseases), and less common etiologies (e.g., nodular sarcoidosis, rheumatoid nodules, granulomatosis with polyangiitis, and nodular amyloidosis).

Thus, until further information is provided, we urge caution when including KS in the differential diagnosis of the numerous conditions that may present with the CT pattern of multiple cavitary pulmonary nodules. ■

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

Correspondence and requests for reprints should be addressed to Edson Marchiori, M.D., Ph.D., Department of Radiology, Federal University of Rio de Janeiro, Rua Thomaz Cameron, 438. Valparaiso, CEP 25685.120. Petrópolis, Rio de Janeiro, Brazil. Email: edmarchiori@gmail.com.

References

1. Chapman R, Hussain N, Tinson G, Moonsie I. Pneumothorax secondary to cavitating pulmonary Kaposi sarcoma in an immunocompetent adult. *Am J Respir Crit Care Med* 2024;210:e18–e19.
2. Zeinaty PE, Lebbé C, Delyon J. Endemic Kaposi's sarcoma. *Cancers (Basel)* 2023;15:872.
3. Dupin N, Jary A, Boussouar S, Syrykh C, Gandjbakhche A, Bergeret S, *et al.* Current and future tools for diagnosis of Kaposi's sarcoma. *Cancers (Basel)* 2021;13:5927.
4. Gasparetto TD, Marchiori E, Lourenço S, Zanetti G, Vianna AD, Santos A, *et al.* Pulmonary involvement in Kaposi sarcoma: correlation between imaging and pathology. *Orphanet J Rare Dis* 2009;4:18.

Limitations of Linear Dimensional Reduction Methods in Chronic Obstructive Pulmonary Disease Phenotyping

Yoshiyasu Takefuji

Faculty of Data Science, Musashino University, Tokyo, Japan

ORCID ID: 0000-0002-1826-742X (Y.T.).

To the Editor:

Bell and colleagues conducted a comprehensive investigation into the temporal exploration of chronic obstructive pulmonary disease (COPD) phenotypes, drawing insights from the COPDGene and SPIROMICS cohorts (1). To address the issue of missing data, they used singular value decomposition (SVD) for imputation, thereby creating a complete data matrix suitable for further analysis. Based on Python code, principal component analysis (PCA) was performed on the completed data matrix. They selected the first six principal components for analysis, basing their choice on the extended elbow criterion observed in the corresponding scree plot (1). This methodological framework enabled a nuanced understanding of COPD phenotypes over time, contributing valuable insights to the research field.

This paper raises significant concerns regarding the use of SVD and PCA because of their nature of linearity and neglecting true associations between the targets and features, leading to erroneous interpretations and conclusions against the nonlinear nonparametric nature of biological data analysis (2–8). When linear methods such as SVD and PCA are applied to nonlinear biological data, the outcomes are potentially distorted by eliminating important nonlinear features, leading to erroneous interpretations. For example, complex interactions between genetic factors and environmental exposures in COPD may follow nonlinear patterns that cannot be adequately captured by linear transformations.

Both SVD and PCA focus solely on variance maximization and orthogonal projections, neglecting true causal or correlative associations between the target disease outcomes and features. This is particularly problematic in COPD research, where certain biomarkers may have nonlinear relationships with disease progression that are not reflected in their variance contribution. Dimensional reduction or feature selection should be based on true associations with clinical outcomes rather than mathematical variance; otherwise, important features with strong biological significance but modest variance may be eliminated, potentially missing crucial disease mechanisms specific to COPD subpopulations.

This paper advocates two types for feature selection. One type contains Spearman's correlation with *P* values and Kendall's tau, both

Artificial Intelligence Disclaimer: No artificial intelligence tools were used in writing this manuscript.

Originally Published in Press as DOI: 10.1164/rccm.202504-0857LE on June 23, 2025

accompanied by *P* values for pairwise relationships with monotonic patterns (9). These nonparametric correlation methods can detect monotonically increasing or decreasing relationships between variables without assuming linearity, making them particularly valuable for analyzing biomarkers that may have consistent directional relationships with COPD progression but not necessarily linear ones. Another includes mutual information analysis (10) and effective transfer entropy (11) for complex interactions among multiple variables with nonmonotonic patterns.

Mutual information analysis quantifies the amount of information shared between variables regardless of relationship shape, capturing complex dependencies in COPD pathophysiology that may be missed by correlation-based approaches. Effective transfer entropy extends this by determining causal relationships and information flow direction between variables, potentially revealing mechanistic insights about how certain genetic or environmental factors influence disease progression over time. These information-theoretic approaches are especially suited for analyzing the complex, multifactorial nature of COPD, where multiple interacting systems contribute to disease heterogeneity. ■

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

Correspondence and requests for reprints should be addressed to Yoshiyasu Takefuji, Ph.D., Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan. Email: takefuji@keio.jp.

References

1. Bell AJ, Ram S, Labaki WW, Murray S, Kazerooni EA, Galban S, *et al.* Temporal exploration of chronic obstructive pulmonary disease phenotypes: insights from the COPDGene and SPIROMICS cohorts. *Am J Respir Crit Care Med* 2025;211:569–576.
2. Leblond F, Tichauer KM, Pogue BW. Singular value decomposition metrics show limitations of detector design in diffuse fluorescence tomography. *Biomed Opt Express* 2010;1:1514–1531.
3. Vaidya PG, S Anand PS, Nagaraj N. Erratum to: a nonlinear generalization of singular value decomposition and its applications to mathematical modeling and chaotic cryptanalysis. *Acta Appl Math* 2010;112:223.
4. Dyer EL, Kording K. Why the simplest explanation isn't always the best. *Proc Natl Acad Sci U S A* 2023;120:e2319169120.
5. Cristian P-M, Aarón V-J, Armando E-HD, Estrella M-LY, Daniel N-R, David G-V, *et al.* Diffusion on PCA-UMAP manifold: the impact of data structure preservation to denoise high-dimensional single-cell RNA sequencing data. *Biology (Basel)* 2024;13:512.
6. Yao Y, Ochoa A. Limitations of principal components in quantitative genetic association models for human studies. *eLife* 2023;12:e79238.
7. Elhaik E. Principal component analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Sci Rep* 2022;12:14683.
8. Mohseni N, Elhaik E. Biases of principal component analysis (PCA) in physical anthropology studies require a reevaluation of evolutionary insights. *eLife* 2024;13:RP94685.
9. Okoye K, Hosseini S. Correlation tests in R: Pearson Cor, Kendall's tau, and Spearman's rho. In: R programming. Singapore: Springer; 2024.
10. Erman B. Mutual information analysis of mutation, nonlinearity, and triple interactions in proteins. *Proteins* 2023;91:121–133.
11. Li WX, Lin QH, Zhang CY, Han Y, Calhoun VD. A new transfer entropy method for measuring directed connectivity from complex-valued fMRI data. *Front Neurosci* 2024;18:1423014.

Reply to Takefuji: Limitations of Linear Dimensional Reduction Methods in Chronic Obstructive Pulmonary Disease Phenotyping

Alexander J. Bell[1], Andrei Zinovyev[2], and Alexander N. Gorban[3,4]

[1]Department of Radiology, University of Michigan, Ann Arbor, Michigan; [2]In Silico R&D, Evotec, Toulouse, France; [3]School of Computing and Mathematical Sciences, University of Leicester, Leicester, United Kingdom; and [4]Department of Mathematics, King's College London, London, United Kingdom

ORCID ID: 0000-0002-6151-0915 (A.J.B.).

From the Authors:

This paper responds to the letter "Limitations of Linear Dimensional Reduction Methods in Chronic Obstructive Pulmonary Disease Phenotyping" by Takefuji, examining recently published work by Bell and colleagues (1). In the evaluated study, clinical trajectory analysis was performed using an algorithm published by Golovenkin and colleagues (2), in which key steps included imputation of missing data using singular value decomposition (SVD), dimensionality reduction using principal component analysis (PCA), and consequent structural analysis using elastic principal graphs (3, 4). The purpose of the study was to map chronic obstructive pulmonary disease (COPD) phenotypes, including their interconnectedness, to infer potential developmental pathways.

Takefuji criticized the use of the linear methods PCA and SVD, as complex biological data is expected to contain nonlinear relationships. Specifically, Takefuji suggests such linear analysis likely neglects true associations within the data, leading to erroneous interpretation. References on the limitations of SVD and PCA are provided. Takefuji suggests alternative methods, including Spearman's correlation, mutual information analysis, and effective transfer entropy (ETE), to better account for nonlinearity in the complex biological data in COPD.

The evaluation by Takefuji is highly appreciated, and it is true that capturing nonlinear dependencies between variables is certainly important for analyzing clinical and complex biological data in COPD. The methods highlighted—Spearman correlation, mutual information analysis, and ETE—are indeed valuable tools for detecting nonlinear relationships between pairs of variables and are widely used in biological data analysis. However, Takefuji appears to have overlooked some crucial points regarding the precise roles of SVD and PCA in the examined work and does not account for the key role of elastic principal graphs, an essentially nonlinear method designed specifically to uncover complex nonlinear structures within data.

Artificial Intelligence Disclaimer: No artificial intelligence tools were used in writing this manuscript.

Originally Published in Press as DOI: 10.1164/rccm.202504-1019LE on June 23, 2025