BJA

British Journal of Anaesthesia, xxx (xxx): xxx (xxxx)

CORRESPONDENCE

Changes in circulating extracellular vesicle cargo are associated with cognitive decline after major surgery. Comment on *Br J Anaesth* 2025; 134: 1683–95

Yoshiyasu Takefuji

Faculty of Data Science, Musashino University, Tokyo, Japan

E-mail: takefuji@keio.jp

Keywords: cognitive decline; extracellular vesicles; feature agglomeration; nonlinear data analysis; principal component analysis

Editor—Mkrtchian and colleagues[1] conducted an observational case–control study examining the relationship between alterations in the cargo of circulating extracellular vesicles (EVs) and cognitive decline after major surgery. They used PCA to analyse the EV proteome, which comprised normalised abundances of 214 proteins consistently expressed across all samples. The PCA results revealed two distinct clusters of EV proteome signatures at 24 and 48 h after surgery, which were independent of the postoperative cognitive outcome groups.[1]

However, this study raises important theoretical concerns regarding the application of PCA, particularly because of its linear nature. Although PCA is a widely used and powerful tool, this characteristic can lead to erroneous interpretations and conclusions when analysing nonlinear, nonparametric biological data.[2–14] For accurate data analysis, researchers must thoroughly understand the assumptions underlying the analytical tools they use. Violating these assumptions can result in distorted outcomes including all metric scores, ultimately leading to misleading interpretations.

There are three primary types of misapplications in data analysis: (1) violating fundamental assumptions, (2) challenges related to ground truth in model interpretation, and (3) critical misapplications such as improper data normalisation and transformation. Here I highlight the first and third categories of misapplications identified in Mkrtchian and colleagues' study.

PCA is predicated on several key assumptions: the existence of linear relationships among variables, the use of continuous and standardised data, adequate sample sizes, homoscedasticity, minimal outliers, and the orthogonality of principal components. When linear models such as PCA are applied to nonlinear datasets, or when parametric methods are used inappropriately with nonparametric data, the results can become significantly distorted. This distortion can result from issues such as loading magnitudes, which can lead to local optimisation errors, or from feature selection based on cumulative contribution, which can incorrectly aggregate contributions across orthogonal components. Therefore, strict adherence to these underlying assumptions is essential for drawing valid conclusions from PCA and similar analytical techniques. Violating assumptions cannot guarantee calculated metric scores.

In light of these limitations, I advocate for using multifaceted approaches that integrate unsupervised machine learning models, such as feature agglomeration[15] and highly variable gene selection.[16] Feature agglomeration helps in reducing dimensionality while preserving important structural information within the dataset. By grouping similar features together, researchers can capture complex relationships in the data without making linearity assumptions. Meanwhile, highly variable gene selection prioritises the selection of genes or proteins that exhibit the most significant variability across samples, ensuring that the most informative features are analysed for their biological relevance. After these steps, researchers can then apply nonlinear, nonparametric statistical methods, such as Spearman's correlation, to draw more reliable and interpretable insights from the data. These approaches collectively enhance the robustness of the analysis and facilitate a more accurate understanding of the biological mechanisms underlying cognitive decline after surgery.

DOI of original article: 10.1016/j.bja.2024.07.040.

Declaration of interest

The author declares that they have no conflict of interest.

References

1. Mkrtchian S, Eldh M, Ebberyd A, et al. Changes in circulating extracellular vesicle cargo are associated with cognitive decline after major surgery: an observational case–control study. *Br J Anaesth* 2025; **134**: 1683–95
2. Dyer EL, Kording K. Why the simplest explanation isn't always the best. *Proc Natl Acad Sci USA* 2023; **120**: e2319169120
3. Cristian PM, Aarón VJ, Armando ED, et al. Diffusion on PCA-UMAP manifold: the impact of data structure preservation to denoise high-dimensional single-cell RNA sequencing data. *Biology* 2024; **13**: 512
4. Yao Y, Ochoa A. Limitations of principal components in quantitative genetic association models for human studies. *eLife* 2023; **12**: e79238
5. Elhaik E. Principal component analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Sci Rep* 2022; **12**: 14683
6. Mohseni N, Elhaik E. Biases of principal component analysis (PCA) in physical anthropology studies require a reevaluation of evolutionary insights. *eLife* 2024; **13**: RP94685
7. Lenz M, Müller FJ, Zenke M, Schuppert A. Principal components analysis and the reported low intrinsic dimensionality of gene expression microarray data. *Sci Rep* 2016; **6**: 25696
8. Dey R, Lee S. Asymptotic properties of principal component analysis and shrinkage-bias adjustment under the generalized spiked population model. *J Multivar Anal* 2019; **173**: 145–64
9. Mehta P, Wang CH, Day AGR, et al. A high-bias, low-variance introduction to machine learning for physicists. *Phys Rep* 2019; **810**: 1–124
10. Prasad S, Bruce LM. Limitations of principal components analysis for hyperspectral target recognition. *IEEE Geosci Remote Sens Lett* 2008; **5**: 625–9
11. Alanis-Lobato G, Cannistraci CV, Eriksson A, Manica A, Ravasi T. Highlighting nonlinear patterns in population genetics datasets. *Sci Rep* 2015; **5**: 8140
12. Black D, Byrne D, Walke A, et al. Towards machine learning-based quantitative hyperspectral image guidance for brain tumor resection. *Commun Med (Lond)* 2024; **4**: 131
13. Caprihan A, Pearlson GD, Calhoun VD. Application of principal component analysis to distinguish patients with schizophrenia from healthy controls based on fractional anisotropy measurements. *Neuroimage* 2008; **42**: 675–82
14. Nyamundanda G, Brennan L, Gormley IC. Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics* 2010; **11**: 571
15. Ward JH. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 1963; **58**: 236–44
16. Brennecke P, Anders S, Kim J, et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods* 2013; **10**: 1093–5

doi: 10.1016/j.bja.2025.09.022