

Artificial intelligence should always be guided by human intelligence. Response to 'Augmented vs. artificial intelligence for stratification of patients with myositis' by Mahler *et al*

We thank Mahler *et al* for their comment¹ on our work in which we used hierarchical clustering on principal components to define clinically meaningful subgroups of patients with anti-Ku antibodies.²

Mahler *et al* argue for the use of machine learning alongside expert decision, thus relying on augmented judgement in making the final decision on patient stratification. We share this view.

In this regard, we disagree with the statement according to which the hierarchical clustering on principal components applied to 1000 observations with a multivariate normal distribution proposed by Pinal-Fernandez and Mammen³ and the hierarchical clustering on principal components applied to our observations from 42 anti-Ku patients yielded similar results.

As Mahler *et al* stated, clustering “will always give an optimal solution for the number of clusters present in a dataset”.

But here are three major differences between the two results that should not be overlooked, in order to avoid false discovery:

1. Mahler *et al* stated that “it is to the user’s discretion to determine whether those clusters exist or not”. At first (human) glance on the factorial maps, one can deduce that the results yielded by the clustering applied in Pinal-Fernandez and Mammen’s dataset and ours are radically different. Examination with the naked eye of the projection in the factorial map of the 1000 observations proposed by Pinal-Fernandez and Mammen reveals a unique dot that should lead to the human intelligence–driven conclusion that no relevant cluster exists in the dataset. In contrast, projection in a factorial map of our observations in 42 anti-Ku patients yielded three dots when observed with the naked eye that should lead to the human intelligence–driven conclusion that partitioning these patients on their principal components is relevant.
2. To further appreciate the relevance of the clusters, human intelligence can be assisted (‘augmented intelligence’) by the use of statistical tools such as Bartlett’s test of sphericity and K-fold cross-validation. We ran these two tests on the example proposed Pinal-Fernandez and Mammen and showed that the irrelevance, already obvious at the stage of ‘human intelligence’, was confirmed at the stage of ‘augmented intelligence’.⁴ On the contrary, suitability of our data for hierarchical clustering on principal components and the presence of the three clusters observed on the naked-eye examination was confirmed by these tests.²
3. In essence, the example provided by Pinal-Fernandez and Mammen is a smart statistical sophistry in which artificial intelligence is voluntarily misguided by human intelligence since one should not use hierarchical clustering on principal components on a dataset that is not suspected to contain relevant clusters. On the contrary, we used this technique

because prior data from independent teams have highlighted that patients with anti-Ku are heterogeneous and thus well fitted for this analysis.

In conclusion, we share the view of Mahler *et al* according to which artificial intelligence should always be guided by human intelligence in order to avoid the risk of false discovery, elegantly illustrated by the example provided by Pinal-Fernandez and Mammen. As briefly pointed above, certain good practice rules do exist to prevent both authors and readers from this danger. Ultimately, hypotheses generated by these techniques must be validated in independent cohorts.

Alain Meyer,^{1,2} Lionel Spielmann,³ François Séverac^{4,5}

¹Exploration Fonctionnelle Musculaire, Service de physiologie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

²Centre National de Référence des Maladies Auto-Immunes Systémiques Rares de l’Est et du Sud-Ouest, Service de rhumatologie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

³Service de Rhumatologie, Hôpitaux civils de Colmar, Colmar, France

⁴Service de Santé Publique, GMRC, CHU de Strasbourg, Strasbourg, France

⁵Cube, UMR 7357, équipe IMAGeS, Université de Strasbourg, Strasbourg, France

Correspondence to Dr Lionel Spielmann, Service de Rhumatologie, Hospices civils de Colmar, Colmar, Alsace 68024, France; lionel.spielmann@ch-colmar.fr

Handling editor Josef S Smolen

Contributors AM and LS wrote the manuscript with support from FS. All authors contributed to the final version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.



To cite Meyer A, Spielmann L, Séverac F. *Ann Rheum Dis* Epub ahead of print: [please include Day Month Year]. doi:10.1136/annrheumdis-2019-216014

Received 18 July 2019

Accepted 19 July 2019



► <http://dx.doi.org/10.1136/annrheumdis-2019-216000>

Ann Rheum Dis 2019;0:1. doi:10.1136/annrheumdis-2019-216014

REFERENCES

- 1 Mahler M, Rossin B, Kubassova O. Augmented vs. artificial intelligence for stratification of patients with myositis. *Ann Rheum Dis*.
- 2 Spielmann L, Nespolo B, Séverac F, *et al*. Anti-Ku syndrome with elevated CK and anti-Ku syndrome with anti-dsDNA are two distinct entities with different outcomes. *Ann Rheum Dis* 2019;78:1101–6.
- 3 Pinal-Fernandez I, Mammen AL. On using machine learning algorithms to define clinically meaningful patient subgroups. *Ann Rheum Dis*. doi:10.1136/annrheumdis-2019-215852. [Epub ahead of print: 21 Jun 2019].
- 4 Meyer A, Spielmann L, Severac F. On how to not misuse hierarchical clustering on principal components to define clinically meaningful patient subgroups. *Ann Rheum Dis*. In Press