Parkinsonism and Related Disorders 62 (2019) 251–252

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis

Correspondence

Studying reproducibility of data-driven Parkinson's disease subtypes

With great interest we read the article of Mestre, Eberly, Tanner, Grimes, Lang, Oakes, and Marras on the reproducibility of data-driven Parkinson's disease (PD) subtypes [1]. We fully subscribe the aim of their study assessing the reproducibility of the results of currently performed studies on data-driven subtypes, since a variety of classifications has been published in the last decades. Assessing the reproducibility of data-driven subtypes is one approach to value the relevance of PD subtypes.

Replication of previous studies on data-driven PD subtypes is necessary to perform further research based on these classifications. It is crucial that meaningful subtypes can be detected in other study populations than the original population and by other research groups. However, to determine whether studies can be replicated, it is essential that the methodology resembles the original methodology.

Mestre, Eberly, Tanner, Grimes, Lang, Oakes, and Marras [1] could not confirm the classification of our data-driven subtypes in the Dutch PROPARK cohort [2], which has been validated on the Spanish ELEP cohort [3]. Based on the description in their publication, it appears that there are differences in the methodology, which may have prevented reproducibility of the original classification. An important difference is that the variables included in the original cluster analysis do not match the variables reported in the supplemental panel 8, table 1. In this replication study, autonomic symptoms were not included, whereas the patient characteristics gender, disease duration, age-at-onset, and disease severity were included. In our study, we deliberately classified subtypes solely on clinical signs and symptoms, and subsequently evaluated differences in patient characteristics like gender, and age-at-onset, to study differences between patients with different phenotypes. Clustering techniques may force patients in the same cluster based on their patient characteristics if included in the analyses, although they are not necessarily related to different phenotypes. Therefore, selection of variables in the cluster analyses will have an effect on the classification, especially when variables have a larger variability within the study population.

Second, another clustering technique was applied, namely K-means instead of model based clustering. The K-means method has certain drawbacks, as we previously pointed out [4]. One of the risks is to end up in a local minimum, leading to a suboptimal solution. Therefore it is recommended to repeat the cluster analyses with a large number of starting points. Whether this was applied in the study of Mestre and colleagues is unclear. Differences in clustering techniques might result in a different classification of subtypes, although we do agree with the authors that a classification should be robust and not highly dependent on the method that was used. However, clustering techniques need to be optimally controlled for their limitations.

The same holds for the measurement instruments. Different scales will have a different sensitivity to measure clinical signs or symptoms. Hence, replication of analyses based on different measurement instruments may end up in different results. Again, to be able to implement defined subtypes on a larger scale in further research, results should not be highly dependent on the measurement instrument that was used. Although differences in sensitivity, and applicability specifically for measuring PD signs and symptoms, should be taken into account.

Furthermore, the authors discuss that a study population with a wide range of disease duration is of limited value, since patients may develop a different clinical presentation during the disease, and may be classified to different PD subtypes throughout the course of the disease. We are of the opinion that analyses performed on a population with a larger variety in disease duration, with severity scores of clinical signs and symptoms converted relative to disease duration, may reveal a spectrum of subtypes for the complete course of the disease, rather than for only early or later stages of the disease. As scores are relative to the disease duration, differences between patients may become overt already during the first stages of the disease and could have been missed in the Longitudinal and Biomarker Studies in Parkinson's Disease (LABS-PD) cohort. To enhance further research on biomarkers (or other fields of interest) related to subtypes, PD subtypes should be able to be replicated. We therefore support the idea of guidance of research on data-driven PD subtypes, including guidance on reproducibility studies of PD subtypes, to move forward.

References

[1] T.A. Mestre, S. Eberly, C. Tanner, D. Grimes, A.E. Lang, D. Oakes, C. Marras, Reproducibility of data-driven Parkinson's disease subtypes for clinical research, Park. Relat. Disord. (2018), https://doi.org/10.1016/j.parkreldis.2018.07.009.

[2] S.M. Rooden, F. Colas, P. Martínez-Martin, M. Visser, D. Verbaan, J. Marinus, R.K. Chaurdhuri, J.N. Kok, J.J. van Hilten, Clinical subtypes of Parkinson's disease, Mov. Disord. 26 (2011) 51–58, https://doi.org/10.1002/mds.23346.

[3] ELEP group, A longitudinal study of Patients with Parkinson's disease (ELEP): aims and methodology, Rev. Neurol. 42 (2006) 360–365.

[4] S.M. van Rooden, W.J. Heiser, J.N. Kok, D. Verbaan, J.J. van Hilten, J. Marinus, The identification of Parkinson's disease subtypes using cluster analysis: a systematic review, Mov. Disord. 15 (2010) 969–978, https://doi.org/10.1002/mds.23116.

https://doi.org/10.1016/j.parkreldis.2018.12.027
Received 9 November 2018; Accepted 26 December 2018
1353-8020/ © 2019 Elsevier Ltd. All rights reserved.

Stephanie M. van Rooden*
Julius Centre for Health Sciences and Primary Care, Huispost nr. STR 6. 131, P.O. Box 85500, 3508, GA, Utrecht, the Netherlands
E-mail address: S.M.vanRooden@umcutrecht.nl.

Dagmar Verbaan
Amsterdam UMC Location AMC, Department of Neurosurgery, PO Box 22660, 1100DD, Amsterdam Zuidoost, the Netherlands
Amsterdam UMC Location AMC, Department of Rehabilitation, PO Box 22660, 1100DD, Amsterdam Zuidoost, the Netherlands
E-mail address: d.verbaan@amc.uva.nl.

Martine Jeukens-Visser
Amsterdam UMC Location AMC, Department of Rehabilitation, PO Box 22660, 1100DD, Amsterdam Zuidoost, the Netherlands
E-mail address: m.jeukens-visser@amc.uva.nl.

Jacobus J. van Hilten
Dept. of Neurology, Leiden University Medical Centre, PO Box 9600, 2300, RC, Leiden, the Netherlands
E-mail address: j.j.van_hilten@lumc.nl.

* Corresponding author.