Open Peer Review of Fleischer et al. (2024) Cytokines (IL1β, IL6, TNFα) and serum cortisol levels may not constitute reliable biomarkers to identify individuals with post-acute sequelae of COVID-19

Prof. Dr. Matthias Kohl
Institute of Precision Medicine
Furtwangen University

In the paper by Fleischer et al. (2024), the authors report on their investigation of four selected biomarkers (IL1β, IL6, TNFα and serum cortisol) for the identification of post-acute sequelae of COVID-19 (PASC). Four groups of people were examined including people *"who had never contracted SARS-CoV-2 (n = 13), infected but had no PASC (n = 34), infected with former PASC that resolved (n = 40) and patients with ongoing PASC after infection (n = 91)"*.

The authors state that this is a prospective cohort study, which is obviously observational in nature. Based on the reference numbers given for the local ethics committee, the study appears to be closely related to the study by Fleischer et al. (2022), where the same reference numbers are given and which was also described by the authors as a prospective observational cohort study. Unfortunately, neither of the two studies was deposited in a registry for clinical studies, even though this should be the standard for observational studies today (Hemingway et al. 2009, Williams et al. 2010, Ioannidis 2012) and the current registries also include many observational studies. This makes it impossible to determine the relationship between these two studies. It is also not possible to verify whether the reported results also match the original study objectives of the studies. Particularly in the case of observational studies, great importance should be attached to this (Thomas and Peterson (2012)) and this must be considered as a major methodological weakness of the study. It is also unclear whether sample size calculations were performed and what sample sizes were originally planned for the study groups.

Furthermore, it is not explained in more detail in which sense the study was planned prospectively. For the four biomarkers investigated and their significance in the context of PASC, reference is made to studies whose results were published in 2021 and 2022. In particular, reference is also made to the WHO Delphi consensus criteria of October 2021 for the definition of PASC, although the connection between the consensus criteria and the publication cited in the article (reference 17 by Buonsenso et al. (2021)), which deals with *"a clinically based classification of disease severity for paediatric COVID-19"*, is unclear. Since the inclusion of patients began in January 2021, at a time when neither the definition of the PASC group was possible nor the significance of the selected biomarkers in this context was known, it must be assumed that the study objectives were originally different. Again it is not possible to verify this as the study was not registered.

Another important methodological weakness lies in the insufficient consideration of the STROBE guideline (van Elm et al. (2007), Vandenbroucke et al. (2007)), which describes how cohort studies should be reported. Important points of the STROBE checklist, which were insufficiently addressed in the article, are for instance points 9 (Bias), 12 (Statistical methods) and 13 (Participants). It is not clear, for example, how an attempt was made to avoid selection bias, whether and how potential confounders were examined and how exactly the 178 participants examined in the study were came about (e.g. numbers examined for eligibility, confirmed, included in the study, and analyzed).

In addition, the description of the statistical analysis also reveals methodological weaknesses. The important results of the Kruskal-Wallis test were not reported. Suitable post-hoc tests in this case would not be Dunnett's multiple comparison tests (Dunnett (1964)), but Dunn's multiple comparisons tests (Dunn (1964)). Strictly speaking, however, these post-hoc tests would not be necessary if the Kruskal-Wallis tests were not significant, as post-hoc tests are usually only applied in the case of a significant result. They are used to uncover the specific differences between three or more groups.
Since a non-parametric test was used to compare the groups, it is also unclear why the Shapiro-Wilk test (Shapiro and Wilk (1965)) was also applied, which is a test of normality (not parametric distribution!). Based on the data distributions shown in Figure 1 of Fleischer et al. (2024), it can be assumed that there was clearly no normal distribution, at least for the cytokines. Accordingly, it is also highly doubtful that Pearson's correlation is applicable in this case, as the necessary assumption of a linear relationship between the parameters investigated is extremely questionable. In view of the (right-)skewed data distributions for the cytokines and the sometimes visible extreme data points (outliers), which are also the causes of the very large standard deviations of the "No prior COVID-19" group in Figure 1, it must be assumed that the results given for the correlation are clearly distorted. It is known that Pearson's correlation is very sensitive to outliers (Devlin et al. (1975)).
Due to the skewed data distributions and since the Kruskal-Wallis test (Kruskal and Wallis (1952)) is not a test for the mean value, but generally examines a shift in the location of the distribution (more precisely, stochastic dominance), the specification of mean +/- SD in Table 2 and Figure 1 is also questionable from a statistical point of view. The specification of median and IQR and the use of box-and-whisker plots is usual in this case and also statistically more meaningful.

In observational studies, as in the present study, confounders must always be taken into account, which can have a decisive influence on the result (D'Onofrio et al. (2020)). This point was apparently given very little to no attention in the present study. Nothing is mentioned in the Statistics section and only a few demographic parameters are listed in Table 1, the relevance of these parameters for the results does not appear to have been examined in detail.

Due to the serious methodological and statistical weaknesses described above, it must be concluded that the reported results are highly likely to be biased and that the existence of false negative results does not seem unlikely. The final conclusion, that *"the above-mentioned cytokines and cortisol are not appropriate biomarkers. The results of this study are consistent with our previous findings and those of others who did not find any laboratory changes and have suggested a non-organic/psychosomatic genesis of PASC.",* which is already implied in the title of the paper, therefore appears completely unjustified. It also ignores the long-established fact that absence of evidence is not evidence of absence (Altman and Bland (1995)). Further studies are certainly needed to be able to exclude IL1β, IL6, TNFα and serum cortisol as biomarkers associated with PASC. In particular, it seems completely illogical why one can conclude a "non-organic/ psychosomatic genesis of PASC" on the basis of only four negative biomarkers when there is contradictory evidence in the literature, as also stated in the introduction by Fleischer et al. (2024), and when there are tens of thousands of possible biomarkers that could prove an organic genesis of PASC.

References

Fleischer M., Szepanowski F., Mausberg A.K., Asan L., Uslar E., Zwanziger D., Volbracht L., Stettner M., Kleinschnitz C. (2024). Cytokines (IL1β, IL6, TNFα) and serum cortisol levels may not

constitute reliable biomarkers to identify individuals with post-acute sequelae of COVID-19. Therapeutic Advances in Neurological Disorders. 17. https://doi.org/10.1177/17562864241229567.

Fleischer M., Szepanowski F., Tovar M., Herchert K., Dinse H., Schweda A., Mausberg A.K., Holle-Lee D., Köhrmann M., Stögbauer J., Jokisch D., Jokisch M., Deuschl C., Skoda E.-M., Taufel M., Stettner M., Kleinschnitz C. (2022). Post-COVID-19 Syndrome is Rarely Associated with Damage of the Nervous System: Findings from a Prospective Observational Cohort Study in 171 Patients. *Neurol Ther* **11**, 1637–1657 (2022). https://doi.org/10.1007/s40120-022-00395-z

Hemingway H., Riley R.D., Altman D.G. (2009). Ten steps towards improving prognosis research. BMJ 339:b4184 https://doi.org/10.1136/bmj.b4184.

Williams R.J., Tse T., Harlan W.R., Zarin D.A. (2010). Registration of observational studies: is it time? CMAJ. 182(15):1638-42. https://doi.org/10.1503/cmaj.092252.

Ioannidis J.P.A. (2012). The Importance of Potential Studies That Have Not Existed and Registration of Observational Data Sets. *JAMA* 308(6):575-576. https://doi.org/10.1001/jama.2012.8144.

Thomas L., Peterson E.D. (2012). The Value of Statistical Analysis Plans in Observational Research: Defining High-Quality Research From the Start. *JAMA* 308(8):773-774. https://doi.org/10.1001/jama.2012.9502.

Buonsenso D., Parri N., De Rose C., Valentini P.; Gemelli-pediatric COVID-19 team (2021). Toward a clinically based classification of disease severity for paediatric COVID-19. Lancet Infect Dis. 21(1):22. https://doi.org/10.1016/S1473-3099(20)30396-0.

von Elm E., Altman D.G., Egger M., Pocock S.J., Gøtzsche P.C., Vandenbroucke J.P.; STROBE Initiative (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med. 147(8):573-7. https://doi.org/10.7326/0003-4819-147-8-200710160-00010.

Vandenbroucke J.P., von Elm E., Altman D.G., Gøtzsche P.C., Mulrow C.D., Pocock S.J., Poole C., Schlesselman J.J., Egger M.; STROBE Initiative (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Ann Intern Med. 147(8):W163-94. https://doi.org/10.7326/0003-4819-147-8-200710160-00010-w1.

Dunnett C.W. (1964). New tables for multiple comparisons with a control. *Biometrics*, **20**:482-491. https://doi.org/10.2307/2528490.

Dunn O.J. (1964). Multiple comparisons using rank sums. Technometrics. 6 (3): 241-252. https://doi.org/10.2307/1266041.

Shapiro S.S., Wilk M.B. (1965). An analysis of variance test for normality (complete samples). Biometrika. 52 (3-4): 591-611. https://doi.org/10.1093/biomet/52.3-4.591.

Devlin S.J., Gnanadesikan R., Kettenring J.R. (1975). Robust estimation and outlier detection with correlation coefficients. Biometrika. 62 (3): 531-545. https://doi.org/10.2307/2335508.

Kruskal W.H., Wallis A. (1952). Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association. 47 (260): 583-621. https://doi.org/10.1080/01621459.1952.10483441.

D'Onofrio B.M., Sjölander A., Lahey B.B., Lichtenstein P., Öberg A.S. (2020). Accounting for Confounding in Observational Studies. Annu Rev Clin Psychol. 7(16):25-48. https://doi.org/10.1146/annurev-clinpsy-032816-045030.

Altman D.G., Bland J.M. (1995). Absence of evidence is not evidence of absence. BMJ. 19;311(7003):485. https://doi.org/10.1136/bmj.311.7003.485.