

CORRESPONDENCE

Open Access



Lost in a number: concealed heterogeneity within the sequential organ failure assessment (SOFA) score

Neville Dusaj¹, Eleni Papoutsis², Katherine L. Hoffman³, Ilias I. Siempos^{2,4} and Edward James Schenck^{4*}

Keywords Predictive enrichment, Prognostic enrichment, Heterogeneity of treatment effect, Multiple organ dysfunction scores, SOFA

To the editor:

Organ dysfunction scores [1] are used in critical care research to benchmark the risk of death in ICU populations and to explore potential heterogeneity of treatment effects in clinical trials. The SOFA score, an updatable organ dysfunction score made of six individual subscores, is used to define sepsis [2] and has been used in randomized clinical trials of sepsis and ARDS to define quantiles of risk to explore heterogeneity of the average treatment effect.

Implicit in the use of multiple organ dysfunction as a stratification method is the expectation that the approach will result in sub-populations that will be more homogeneous and share a similar prognosis. Unfortunately, this approach may not account for potential clinical and biologic heterogeneity. Such heterogeneity may dilute the predictive effect of grouping by a similar prognosis.

Recent work has identified ICU subphenotypes using SOFA scores together with other biologic variables [3]. More simply, a single SOFA score number contains multiple combinations of disparate organ dysfunctions. For example, a score of 6 has 426 subscore combinations, and 12 has 1751. This heterogeneity may conceal varied pathobiology leading to a similar prognosis in critical illness. To illustrate this potential, we explored the heterogeneity within groups of patients sharing a single SOFA score.

We did a retrospective study using two data sources: a single-center ICU cohort, see supplemental methods for details, and the PETAL-ROSE multicenter randomized clinical trial of neuromuscular blockers for patients with ARDS [4]. We identified patients by Sepsis-3 criteria [2] in the ICU cohort, and then, we explored the heterogeneity within patients sharing a day 1 SOFA score of 6, 9, and 12. To validate this heterogeneity in a more specific disease we explored a population with a non-neurologic SOFA score of 9 in the ARDS clinical trial.

Within each strata of patients sharing the same total SOFA score, we performed a clustering analysis to identify subphenotypes. We compared SOFA subscores components, demographics and other baseline factors across clusters in each strata to identify underlying biologic differences. We then compared 28-day mortality and markers accounting for duration of organ failure support.

Within the ICU cohort population, there were 760, 469, 206 patients with a SOFA score of 6, 9 and 12,

*Correspondence:

Edward James Schenck
ejs9005@med.cornell.edu

¹ Tri-Institutional MD-PhD Program, Weill Cornell Medicine, Memorial Sloan Kettering Cancer Center, Rockefeller University, New York, NY, USA

² First Department of Critical Care Medicine and Pulmonary Services, Evangelismos Hospital, National and Kapodistrian University of Athens Medical School, Athens, Greece

³ Division of Biostatistics, Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA

⁴ Division of Pulmonary and Critical Care Medicine, Department of Medicine, New York-Presbyterian Hospital-Weill Cornell Medical Center, Weill Cornell Medicine, 1300 York Avenue, Box 96, New York, NY 10065, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

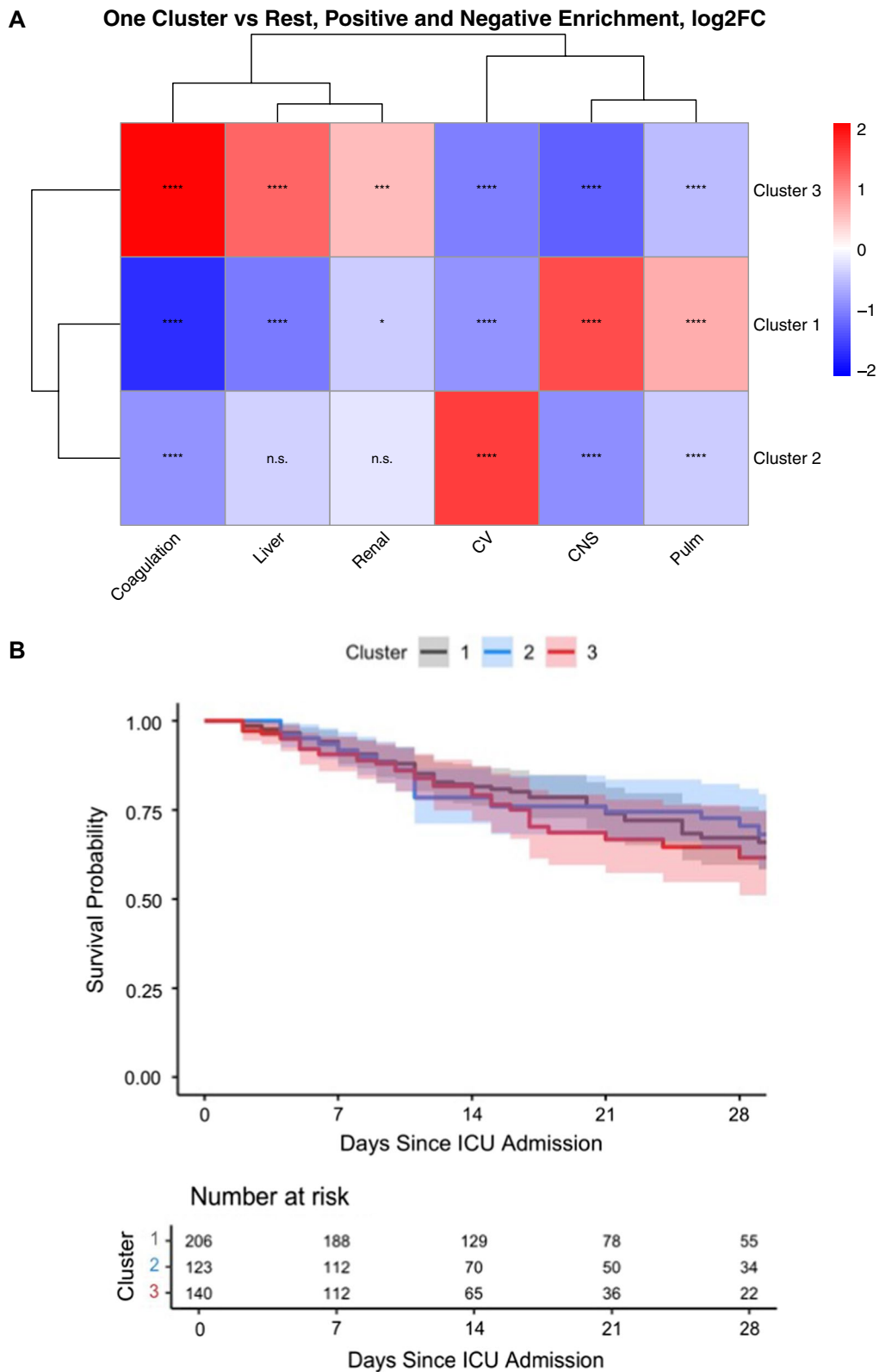


Fig. 1 **A** Heat map of differing clusters with log₂ fold differences in SOFA subscores. Color intensity corresponds to log₂ fold changes, and number of * correspond to statistical significance. Abbreviations: CV Cardiovascular, CNS Central Nervous System, **B** Kaplan–Meier plot comparing survival time between clusters

respectively. Three distinct subscore defined subphenotypes were seen in each group. For example, in the group with a SOFA of 9, higher cardiovascular failure scores, higher respiratory failure scores and higher mixed organ failure were seen as distinct clusters. Panel A of Fig. 1 displays the log₂ fold change of each SOFA subscore in each subphenotype. Similar findings were seen in the SOFA 6 and SOFA 12 strata with different subscore distributions. Consistently, three distinct clusters were seen in the clinical trial population. Details of the total populations and each subphenotype can be found in the supplement, Additional file 1: Tables S1–S4, and Figures S1–S4.

In the SOFA 9 strata in the ICU cohort, patients in the cardiovascular failure cluster were older, more likely to be women, to have a blood stream infection, and have septic shock compared to the two other cohorts. Patients in the respiratory failure cohort had more comorbidities and were more likely to have pneumonia. Patients in the mixed group were younger and were more likely to have immunosuppression compared to patients in other subphenotypes. Differential clusters were seen in the other SOFA score strata and in the ARDS clinical trial, Additional file 1: Table S4 and Figures S5, S6. All SOFA score strata in each case shared a similar prognosis, Additional file 1: Tables S1–S4; however, individual organ dysfunction durations and clinical characteristics were different.

In two independent cohorts, we identified distinct clusters of patients within different SOFA scores each with a similar prognosis but with markedly different clinical characteristics. This analysis displays the hidden heterogeneity within multiple organ dysfunction scores despite accuracy in identifying similar outcomes. This study compliments work that established that an organ dysfunction scores' validity is a function of the scores' uniformity of fit to the population under study [5] and highlights that predictive enrichment may not be achieved with methods that are prognostically valid.

Strengths include the simple design and inclusion of a broad range of patients from two distinct data sources reflecting a range of ICU patients. Limitations include using the relatively inclusive definition of sepsis from a single academic center with a high severity of disease. Moreover, the use of electronic health records may lead to missingness and confounding regarding neurologic injury. However, we confirmed similar findings in a more restrictive ARDS clinical trial population with manually extracted data. We chose to explore the hidden heterogeneity in the simple SOFA score, a more complicated risk prediction scoring system would by definition conceal more heterogeneity. This analysis

supports explicit hypothesis-driven predictive enrichment in the design of clinical trials. A number is not a surrogate for clinical homogeneity.

Abbreviations

APACHE	Acute physiology and chronic health valuation
ICU	Intensive care unit
SOFA	Sequential organ failure assessment

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13054-023-04782-2>.

Additional file 1. Supplemental Methods, Tables and Figures.

Acknowledgements

This analysis was in part prepared by using ROSE research material obtained from the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) of the National Heart, Lung, and Blood Institute (NHLBI). The article does not necessarily reflect the opinions or views of the researchers who performed this trial or the NHLBI. The authors acknowledge the incredible work by the PETAL Network researchers, without which part of this study would not have been possible.

Author contributions

Conceptualization, formal analysis, methodology, and writing and original draft preparation were done by all authors; and supervision was done by EJS. All authors read and approved the final manuscript.

Funding

EJS is supported by the NHLBI of the National Institutes of Health through grant K23 HL151876. N.D. is supported by a F30 Predoctoral Fellowship from the NHLBI of the National Institutes of Health (F30HL156496) and a Medical Scientist Training Program grant from the National Institute of General Medical Sciences of the National Institutes of Health to the Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program (T32GM007739). IS is supported by a grant from the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "2nd Call for H.F.R.I. Research Projects to support Post-Doctoral Researchers" (Project 80- 1/15.10.2020).

Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The institutional review board at Weill Cornell Medicine approved of this study (#181101976) with a waiver of informed consent. With regard to the usage of data from the ROSE clinical trial, because data would be received in de-identified form from the NHLBI BioLINCC, the Institutional Review Board of Evangelismos Hospital waived the need for informed consent and approved the study (protocol #210/2023-05-10).

Consent for publication

Not applicable.

Competing interests

EJS reports fees from Axle informatics outside of the current work all other authors report no competing interests.

Received: 14 December 2023 Accepted: 18 December 2023
Published online: 02 January 2024

References

1. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 1996;22(7):707–10.
2. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA.* 2016;315(8):801–10. <https://doi.org/10.1001/jama.2016.0287>.
3. Knox DB, Lanspa MJ, Kuttler KG, Brewer SC, Brown SM. Phenotypic clusters within sepsis-associated multiple organ dysfunction syndrome. *Intensive Care Med.* 2015;41(5):814–22. <https://doi.org/10.1007/s00134-015-3764-7>.
4. National Heart L, Blood Institute PCTN, Moss M, et al. Early neuromuscular blockade in the acute respiratory distress syndrome. *N Engl J Med.* 2019;380(21):1997–2008. <https://doi.org/10.1056/NEJMoa1901686>.
5. Moreno R, Apolone G, Miranda DR. Evaluation of the uniformity of fit of general outcome prediction models. *Intensive Care Med.* 1998;24(1):40–7. <https://doi.org/10.1007/s001340050513>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.