I, DOCTOR: The role of machine learning in phenotyping ARDS

*Andrea R. Levine and Carl B. Shanholtz**

Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of Maryland School of Medicine, 110 South Paca Street, Room 2-N-147, Baltimore, MD 21201, USA

EBioMedicine 2022;75: 103770
Published online xxx
https://doi.org/10.1016/j.ebiom.2021.103770

The diagnosis of Acute Respiratory Distress Syndrome (ARDS) encapsulates remarkable heterogeneity in etiology, presentation, severity, course, and treatment response. Undoubtedly, enrolling patients amalgamated under the loosely defined term of ARDS has contributed to innumerable "negative" clinical trials in ARDS.[1] This is likely due to heterogeneity of treatment effect (HTE), or the nonrandom variability in the direction or magnitude of a treatment effect. Understanding and detecting HTE is fundamental to providing personalized, or precision, medicine.[2]

Historically, investigators report subgroup analysis using forest plots. This is often statistically flawed and prone to identifying false-positives and overestimating the effect of true-positives.[2] Comparing individual characteristics between populations underestimates person-level heterogeneity.[2] Etiology, severity, radiographic presentation, protein biomarkers, gene expression, and intervention of ARDS are among the many ways patients have previously been sorted into more homogenous subgroups of ARDS.[3] Latent class analysis (LCA) is a probabilistic technique which allows for a multivariable approach to identifying unmeasured or unobserved subgroups based on selected predictor variables agnostic to the patient's outcome. LCA combining plasma biomarkers and clinical variables has identified inflammatory subphenotypes of ARDS which, *post-hoc*, have demonstrated a differential response to ARDS-directed therapies.[4] However, this 32-variable LCA model is clinically impractical and impossible to use to stratify patients in prospective clinical trials.

The use of complex machine learning (ML) algorithms to group patients far surpasses, and is unencumbered by, our clinical or methodological understanding of patients or disease. Prior work by Sinha et al. has demonstrated the merit of using recursive partitioning and regression ML algorithms to distill the cumbersome list of available clinical and biomarker variables into a parsimonious 3,4 variable model that can accurately identify ARDS subphenotypes and HTE.[5] More recent work by Dr. Sinha's group has established that gradient boosted tree ML algorithms can correctly classify patients into subphenotypes of ARDS, absent the need for protein biomarkers, using readily available clinical data and this clinical classifier model detected HTE which paralleled that of the prior LCA analysis.[6]

In contrast to decision tree algorithms, clustering algorithms for pattern recognition have gained popularity as a ML technique and shown promise in identifying HTE among subgroups in diseases related to ARDS. In this publication, Sinha and colleagues utilized nine unsupervised and supervised ML clustering algorithms to re-analyze three ARDS randomized clinical trials (RCTs) to identify subgroups of patients and HTE.[7] This study was novel in demonstrating the feasibility of applying commonly used ML techniques to ARDS phenotyping. However, these ML algorithms identified different numbers of clusters within the same trials and were inconsistent in establishing HTE. No single ML clustering algorithm emerged as the most robust or reliable. There was tremendous divergence in the contributing partitioning variables between models and the attempt to remove biomarkers from the models, presumably to improve the practicality of this strategy, significantly impeded the algorithms' ability to identify patient clusters or HTE.

ML algorithms are frequently criticized for being a "black box" where the intricacies of the algorithm are concealed and non-intuitive. In contrast, LCA is considered a more statistically robust technique with a lower rate of misclassifications. The variables included in LCA modeling must be deliberately chosen, driven by prior research, mechanistic plausibility, or biological feasibility.[8] The success of classifying patients into subgroups using LCA is unequivocally linked to the expertise of the investigator and the purposeful selection of each indicator variable, making this approach more hypothesis driven, intuitive, and transparent. Although clustering algorithms failed to consistently identify HTE in these three clinical trials, the complete anthology of work from the Sinha group reinforces the feasibility and inevitability of the ML strategy and illuminates the value of the black-box of machine learning compared to LCA alone in phenotyping ARDS.

Both LCA and ML algorithms have only been studied in *post-hoc* analysis. Prospective clinical trials are

*Corresponding author.
E-mail address: cshanhol@som.umaryland.edu (C.B. Shanholtz).

imperative to validate ARDS phenotypes and HTE. This hinges on point-of-care phenotyping of patients prior to clinical trial enrollment. Combining the herculean potential of ML in combination with the intuitive and accessible point-of care clinical indicators offers great promise in further phenotyping patients with ARDS and is the aim of the PHIND study (NCT04009330). Personalized medicine is here to stay, and with it the use of ML as a tool for phenotyping patients. We commend Dr. Sinha and his colleagues' ongoing efforts to exploit the black-box of ML to unwrap the black-box of ARDS heterogeneity.

Contributors

AL and CS contributing equally to the writing, review and editing of this commentary.

Declaration of interests

CS received grants from the department of defense, PR191560, from NIH NINR, R013664 and holds stocks from Teva Pharmaceuticals. AL received grants from NIH NIA, NIH NHLBI and University of Maryland.

References

1 Khan YA, Fan E, Ferguson ND. Precision medicine and heterogeneity of treatment effect in therapies for ARDS. *Chest* 2021;**160**(5):1729–38.
2 Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ* 2018;**363**:k4245.
3 Matthay MA, Arabi YM, Siegel ER, Ware LB, Bos LDJ, Sinha P, et al. Phenotypes and personalized medicine in the acute respiratory distress syndrome. *Intensive Care Med* 2020;**46**(12):2136–52.
4 Sinha P, Delucchi KL, Thompson BT, McAuley DF, Matthay MA, Calfee CS, et al. Latent class analysis of ARDS subphenotypes: a secondary analysis of the statins for acutely injured lungs from sepsis (SAILS) study. *Intensive Care Med* 2018;**44**(11):1859–69.
5 Sinha P, Delucchi KL, McAuley DF, O'Kane CM, Matthay MA, Calfee CS. Development and validation of parsimonious algorithms to classify acute respiratory distress syndrome phenotypes: a secondary analysis of randomised controlled trials. *Lancet Respir Med* 2020;**8**(3):247–57.
6 Sinha P, Churpek MM, Calfee CS. Machine learning classifier models can identify acute respiratory distress syndrome phenotypes using readily available clinical data. *Am J Respir Crit Care Med* 2020;**202**(7):996–1004.
7 Sinha P, Spicer A, Delucchi KL, McAuley DF, Calfee CS, Churpek MM. Comparison of machine learning clustering algorithms for detecting heterogeneity of treatment effect in acute respiratory distress syndrome: a secondary analysis of three randomised controlled trials. *EBioMedicine* 2021;**74**:103697.
8 Sinha P, Calfee CS, Delucchi KL. Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. *Crit Care Med* 2021;**49**(1):e63–79.