PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Identification of Acute Respiratory Distress Syndrome subphenotypes denovo using routine clinical data: a retrospective analysis of ARDS clinical trials
AUTHORS	Duggal, Abhijit; Kast, Rachel; Van Ark, Emily; Bulgarelli, Lucas; Siuba, Matthew T.; Osborn, Jeff; Rey, Diego; Zampieri, Fernando; Cavalcanti, Alexandre; Maia, Israel; Paisani, Denise M; Laranjeira, Ligia N; Serpa Neto, Ary; Deliberato, Rodrigo Octávio

VERSION 1 – REVIEW

REVIEWER	Jacky Suen Prince Charles Hospital, Critical Care Research Group
REVIEW RETURNED	19-Jul-2021

GENERAL COMMENTS	Thank you very much for the opportunity to review this manuscript. I truly enjoy reading the paper and sincerely hope that my comments can help to improve it further. In this paper, Duggal et al worked on the highly important area of ARDS subphenotype, with an attempt to identify patient subphenotype using clinical data that's routinely available. The authors correctly pointed out that in order for ARDS subphenotype to be implemented in clinical setting, it is important to consider feasibility and how common some of the parameters are. As such, the authors developed 10 different models based on combination of various clinical parameters, and using mortality difference as the primary endpoint, with the hope that more commonly used parameter can help broader-application of ARDS subphenotype. While the approach taken by the authors are admirable, it is unclear where is the significance of the finding. Namely, based on the presented data, it is unclear if 2 distinct subphenotypes were identified, or are the pateints simply categorised into "average" and "more severe" ARDS patients. The latter could equally explained the increased mortality, as patients in these group seems to be more severe. This of course is in line with previous attempts in identification of ARDS subphenotype, however, a major difference is that treatment efficacy were also examined among the 2 subphenotypes, which highlight potential differences in underlying mechanisms and biology, leading to different response to treatment. Unfortunately this key part is missing in the current manuscript, consequently, it is unclear of the potential benefits of the potential way of subphenotyping patients. Unless this information can provide better assessment and evaluation of treatment options, it is hard to know if it's better than the current patient stratification based on P/F ratio.

	Therefore, I would disagree with the authors on line 309-311, the a minimum of preliminary analysis of treatment outcome based on the proposed clustering be included in this manuscript. This will significantly improve the impact of the manuscript by demonstration application of the finding. Some minor points: - suggest reviewing and proof-reading of the manuscripts. There seems to be some errors. E.g. line 48 "For feature selections, sets." Line 41-42 seems incomplete. - I would suggest to have the set of clinical parameters chosen to be included in the main text at least once, as I would consider that to be one of the key founding. Likewise, I have spent significant time on eTable 5 and sugges to include it in the main text. - unclear how "frequently observed in the routine care of ARDS patients" is defined. - unclear the meaning of "patients scored" in table 2.

REVIEWER	Tina Chen Montefiore Health System, Critical Care Med, Department of Medicine
REVIEW RETURNED	14-Oct-2021

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript by Abbhijit et al. The study used unsupervised learning to sub phenotypes of ARDS using clinically available (EHR) based variables from large ARDS clinical trials. I thought the authors did a good job in explaining the algorithm chosen to perform their initial training cohort (from Eden and FACTT). The subtypes were tested on multiple ARDS studies and consistently showed separations between the 2 subtypes. The Clinical variables and biomarkers separated nicely between the 2 sub-phenotypes. Advantage of the algorithm: 1) using minimal clinically available variable improves applicability. 2) Validated across different studies and continue to retained separation. 3) the clinical outcomes correlated well with the subtype classifications. 4) biomarkers also demonstrated plausibility 5) low rate of missing-ness 6) clinical studies included spanned across a long time span which taken into account in the change ARDS management (Affecting outcome) Limitation: Retrospective of clinical studies decreases generalizability. Biomarkers were not as available to validate the phenotypes (but in some way this is a strength of the model able to classifiy only using limited number of variables) What are the features that really contribute to the model in sub typing? The model chosen for the sub typing has the least number of variables (pH, PaO2, Cr , Bili, HCO3, MAP, HR, RR, and FiO2). The variables are not all significantly different between the 2 subtypes. I am curious why that is. (ie in ALVEOLI, ARMA, SALS and ARD: PaO2 is not significantly different between the 2 groups) Does that mean that these variables are not important? Does this method of analysis be able to tell us what variables are important in determining subgroups?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Jacky Suen, Prince Charles Hospital
Comments to the Author:

Thank you very much for the opportunity to review this manuscript. I truly enjoy reading the paper and sincerely hope that my comments can help to improve it further.

Thank you for your comments, we appreciate it.

In this paper, Duggal et al worked on the highly important area of ARDS subphenotype, with an attempt to identify patient subphenotype using clinical data that's routinely available. The authors correctly pointed out that in order for ARDS subphenotype to be implemented in clinical setting, it is important to consider feasibility and how common some of the parameters are. As such, the authors developed 10 different models based on combination of various clinical parameters, and using mortality difference as the primary endpoint, with the hope that more commonly used parameter can help broader-application of ARDS subphenotype.
While the approach taken by the authors are admirable

Thank you.

it is unclear where is the significance of the finding. Namely, based on the presented data, it is unclear if 2 distinct subphenotypes were identified, or are the pateints simply categorised into "average" and "more severe" ARDS patients. The latter could equally explained the increased mortality, as patients in these group seems to be more severe. This of course is in line with previous attempts in identification of ARDS subphenotype, however, a major difference is that treatment efficacy were also examined among the 2 subphenotypes, which highlight potential differences in underlying mechanisms and biology, leading to different response to treatment.

Unfortunately this key part is missing in the current manuscript, consequently, it is unclear of the potential benefits of the potential way of subphenotyping patients. Unless this information can provide better assessment and evaluation of treatment options, it is hard to know if it's better than the current patient stratification based on P/F ratio.

Therefore, I would disagree with the authors on line 309-311, the a minimum of preliminary analysis of treatment outcome based on the proposed clustering be included in this manuscript. This will significantly improve the impact of the manuscript by demonstration application of the finding.

Excellent point, thank you for your comment and allowing us to provide some additional clarification.

We acknowledge the importance of the differential treatment response. Since we pool data from different trials with different outcomes and sample sizes, the analysis required to confidently provide this information may demand more detailed explanation that would not fit all in this single paper. In fact, we have found heterogeneity in the treatment effect after analyzing several past trials and will be submitting the work for peer-review subsequently to this work. However, as mentioned above, the amount of information generated to assess the differential response along with the current content of this article explaining the development of our unsupervised learning model with the appropriate data to back up our findings for a clinical audience, would make a single paper inconveniently convoluted. For that reason, we chose first to focus on explaining in detail our approach for model development and provide an extensive validation using several and diverse data sets. In any case, your observation is on point, and to acknowledge so we have added the following sentence at the end of our discussion/limitation, to give it the appropriate relevance:

"Lastly, future work should analyze previous trials to identify possible differential treatment responses for the subphenotypes of ARDS patients identified in this study."

Some minor points:
- suggest reviewing and proof-reading of the manuscripts. There seems to be some errors. E.g. line 48 "For feature selections, sets." Line 41-42 seems incomplete.

Thank you for noticing, we have further proof-read the manuscript and made the necessary corrections.

- I would suggest to have the set of clinical parameters chosen to be included in the main text at least once, as I would consider that to be one of the key founding. Likewise, I have spent significant time on eTable 5 and sugges to include it in the main text.
We have included the list of parameters chosen under the subsection *"Clinical characteristics of each cluster"* of the results. We have also moved eTable 5 to the manuscript, and due to the limit of figures/tables, moved table 2 to the supplement.

- unclear how "frequently observed in the routine care of ARDS patients" is defined.

From the list of variables available across all datasets, we had all ICU physician (co-authors) familiar with ARDS care establish a list of measurements that they would commonly use and/or register in the routine care of these patients. We have altered a sentence in the manuscript to clarify that.

"The list of potential candidates was then further refined to include only those that are frequently observed in the routine care of ARDS patients at the time of its diagnosis according to judgement provided by ICU physicians who participated in this study"

- unclear the meaning of "patients scored" in table 2.

We agree that the meaning of that label is not clear, we have added a description for such in the table legend:

" Number of patients without any missing data, allowing their assignment to one of the clusters."*

Reviewer: 2
Dr. Tina Chen, Montefiore Health System
Comments to the Author:

Thank you for the opportunity to review this manuscript by Abbhijit et al.

We appreciate you taking the time to review - thank you!

The study used unsupervised learning to sub phenotypes of ARDS using clinically available (EHR) based variables from large ARDS clinical trials. I thought the authors did a good job in explaining the algorithm chosen to perform their initial training cohort (from Eden and FACTT). The subtypes were tested on multiple ARDS studies and consistently showed separations between the 2 subtypes. The Clinical variables and biomarkers separated nicely between the 2 sub-phenotypes.

Advantage of the algorithm: 1) using minimal clinically available variable improves applicability. 2) Validated across different studies and continue to retained separation. 3) the clinical outcomes correlated well with the subtype classifications. 4) biomarkers also demonstrated plausibility 5) low rate of missing-ness 6) clinical studies included spanned across a long time span which taken into account in the change ARDS management (Affecting outcome)

We agree that this algorithm model appears to provide consistent identification of two subphenotypes using a small amount of data that shows applicability over time with the available biomarker data supporting our hypothesis.

Limitation: Retrospective of clinical studies decreases generalizability. Biomarkers were not as available to validate the phenotypes (but in some way this is a strength of the model able to classifiy only using limited number of variables)

Thank you for the diligent observations. We agree that the use of retrospective randomized clinical trials does not yet show generalizability to the overall ARDS population. Furthermore, we agree that biomarker analysis was limited to the data available in ALVEOLI and ARMA. We hope to have

acknowledged all of the limitations in our manuscript and hope to be able to address them in future work.

What are the features that really contribute to the model in sub typing? The model chosen for the sub typing has the least number of variables (pH, PaO2, Cr , Bili, HCO3, MAP, HR, RR, and FiO2). The variables are not all significantly different between the 2 subtypes. I am curious why that is. (ie in ALVEOLI, ARMA, SALS and ARD: PaO2 is not significantly different between the 2 groups) Does that mean that these variables are not important? Does this method of analysis be able to tell us what variables are important in determining subgroups?

That is indeed a great question.

Unfortunately, unlike supervised algorithms (e.g., regression analyses), unsupervised algorithms such as K-means clustering do not provide one straightforward and established metric to describe feature importance. In that sense, our approach of testing multiple sets of variables was also meant to select features that were most likely to be relevant, serving as surrogate for the feature selection step normally employed in supervised algorithms.

While each individual variable by itself may not be significantly different across sub-phenotypes, their interaction in the 9-dimensional space of our model may be relevant. T-tests, however, do not account for this interaction.

VERSION 2 – REVIEW

REVIEWER	Jacky Suen Prince Charles Hospital, Critical Care Research Group
REVIEW RETURNED	03-Nov-2021

GENERAL COMMENTS	Dear Authors, Thank you very much for the opportunity to review this work. Overall, I am satisfied with the revision and comments from the authors. It is a shame that the authors cannot provide further information about the consequence of patient subphenotyping and consequent efficacy/response to treatment. If the authors can even provide a little information on this, it will be a nice way to conclude this manuscript.

REVIEWER	Tina Chen Montefiore Health System, Critical Care Med, Department of Medicine
REVIEW RETURNED	18-Nov-2021

GENERAL COMMENTS	Thank you for addressing my comments. Regarding my concern about features contribution to the sub-typing using your un-supervised model. I suggest adding to the discussion section as you have explained in your response to comments.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Jacky Suen, Prince Charles Hospital

Comments to the Author:

Dear Authors,

Thank you very much for the opportunity to review this work. Overall, I am satisfied with the revision and comments from the authors. It is a shame that the authors cannot provide further information about the consequence of patient subphenotyping and consequent efficacy/response to treatment. If the authors can even provide a little information on this, it will be a nice way to conclude this manuscript.

Thank you for your insightful review, we acknowledge the importance of differential treatment responses to treatment in our sub-phenotypes. We tried to analyze the heterogeneity in treatment response using these datasets, but in order to answer this question appropriately we had to undertake a complex Bayesian analysis to confidently answer the question at hand. Given this complexity we are submitting the detailed heterogeneity to treatment effect for peer-review as a separate manuscript. We have acknowledged the need for this analysis in our limitations section and have also mentioned that this is a key step that needs to be studied in relation to our subphenotypes

Reviewer: 2

Dr. Tina Chen, Montefiore Health System

Comments to the Author:

Thank you for addressing my comments.

Thank you for your comments

Regarding my concern about features contribution to the sub-typing using your unsupervised model. I suggest adding to the discussion section as you have explained in your response to comments

We have added the following sentence to the discussion section:

"*Unfortunately, unlike supervised algorithms (e.g., regression analyses), unsupervised algorithms such as K-means clustering do not provide one straightforward and established metric to describe feature importance. In that sense, our approach of testing multiple sets of variables was also meant to select features that were most likely to be relevant, serving as surrogate for the feature selection step normally employed in supervised algorithms.*

While each individual variable by itself may not be significantly different across sub-phenotypes, their interaction in the 9-dimensional space of our model may be relevant."