Phenotyping heart failure using model-based analysis and physiology-informed machine learning

Edith Jones, E Benjamin Randall, Scott L Hummel, David M Cameron, Daniel A Beard, and Brian E. Carlson
DOI: 10.1113/JP281845

Corresponding author(s): Brian Carlson (bcarl@umich.edu)

The following individual(s) involved in review of this submission have agreed to reveal their identity: Pablo Lamata (Referee #2)

Review Timeline:

Submission Date:	28-Apr-2021
Editorial Decision:	19-May-2021
Revision Received:	20-Aug-2021
Accepted:	07-Sep-2021

Senior Editor: Don Bers

Reviewing Editor: Eleonora Grandi

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

1st Editorial Decision 19-May-2021

Dear Professor Carlson,

Re: JP-RP-2021-281845 "Phenotyping heart failure using model-based analysis and physiology-informed machine learning" by Edith Jones, E Benjamin Randall, Scott L Hummel, David M Cameron, Daniel A Beard, and Brian E. Carlson

Thank you for submitting your manuscript to The Journal of Physiology. It has been assessed by a Reviewing Editor and by 2 expert Referees and I am pleased to tell you that it is considered to be acceptable for publication following satisfactory revision.

Please advise your co-authors of this decision as soon as possible.

The reports are copied at the end of this email. Please address all of the points and incorporate all requested revisions, or explain in your Response to Referees why a change has not been made.

NEW POLICY: In order to improve the transparency of its peer review process The Journal of Physiology publishes online as supporting information the peer review history of all articles accepted for publication. Readers will have access to decision letters, including all Editors' comments and referee reports, for each version of the manuscript and any author responses to peer review comments. Referees can decide whether or not they wish to be named on the peer review history document.

I hope you will find the comments helpful and have no difficulty returning your revisions within 4 weeks.

Your revised manuscript should be submitted online using the links in Author Tasks Link Not Available.

Any image files uploaded with the previous version are retained on the system. Please ensure you replace or remove all files that have been revised.

REVISION CHECKLIST:

- Article file, including any tables and figure legends, must be in an editable format (eg Word)

- Upload each figure as a separate high quality file

- Upload a full Response to Referees, including a response to any Senior and Reviewing Editor Comments;

- Upload a copy of the manuscript with the changes highlighted.

You may also upload:

- A potential 'Cover Art' file for consideration as the Issue's cover image;

- Appropriate Supporting Information (Video, audio or data set https://jp.msubmit.net/cgi-bin/main.plex?form_type=display_requirements#supp).

To create your 'Response to Referees' copy all the reports, including any comments from the Senior and Reviewing Editors, into a Word, or similar, file and respond to each point in colour or CAPITALS and upload this when you submit your revision.

I look forward to receiving your revised submission.

If you have any queries please reply to this email and staff will be happy to assist.

Yours sincerely,

Professor Don M. Bers
Senior Editor
The Journal of Physiology
https://jp.msubmit.net
http://jp.physoc.org
The Physiological Society
Hodgkin Huxley House
30 Farringdon Lane
London, EC1R 3AW
UK
http://www.physoc.org
http://journals.physoc.org

REQUIRED ITEMS:

-Author photo and profile. First (or joint first) authors are asked to provide a short biography (no more than 100 words for one author or 150 words in total for joint first authors) and a portrait photograph. These should be uploaded and clearly labelled with the revised version of the manuscript. See Information for Authors for further details.

-You must start the Methods section with a paragraph headed Ethical Approval. If experiments were conducted on humans confirmation that informed consent was obtained, preferably in writing, that the studies conformed to the standards set by the latest revision of the Declaration of Helsinki, and that the procedures were approved by a properly constituted ethics committee, which should be named, must be included in the article file. If the research study was registered (clause 35 of the Declaration of Helsinki) the registration database should be indicated, otherwise the lack of registration should be noted as an exception (e.g. The study conformed to the standards set by the Declaration of Helsinki, except for registration in a database.). For further information see: https://physoc.onlinelibrary.wiley.com/hub/human-experiments

-Your manuscript must include a complete Additional Information section

-The Journal of Physiology funds authors of provisionally accepted papers to use the premium BioRender site to create high resolution schematic figures. Follow this link and enter your details

and the manuscript number to create and download figures. Upload these as the figure files for your revised submission. If you choose not to take up this offer we require figures to be of similar quality and resolution. If you are opting out of this service to authors, state this in the Comments section on the Detailed Information page of the submission form.

-Please upload separate high-quality figure files via the submission form.

-Please ensure that any tables are in Word format and are, wherever possible, embedded in the article file itself.

-Please ensure that the Article File you upload is a Word file.

-Your paper contains Supporting Information of a type that we no longer publish. Any information essential to an understanding of the paper must be included as part of the main manuscript and figures. The only Supporting Information that we publish are video and audio, 3D structures, program codes and large data files. Your revised paper will be returned to you if it does not adhere to our Supporting Information Guidelines

-A Data Availability Statement is required for all papers reporting original data. This must be in the Additional Information section of the manuscript itself. It must have the paragraph heading "Data Availability Statement". All data supporting the results in the paper must be either: in the paper itself; uploaded as Supporting Information for Online Publication; or archived in an appropriate public repository. The statement needs to describe the availability or the absence of shared data. Authors must include in their Statement: a link to the repository they have used, or a statement that it is available as Supporting Information; reference the data in the appropriate sections(s) of their manuscript; and cite the data they have shared in the References section. Whenever possible the scripts and other artefacts used to generate the analyses presented in the paper should also be publicly archived. If sharing data compromises ethical standards or legal requirements then authors are not expected to share it, but must note this in their Statement. For more information, see our Statistics Policy.

EDITOR COMMENTS

Reviewing Editor:

See rev 1 - modifications of previous model(s) and parameters used in machine learning algorithms should be clarified to ensure reproducibility.

Patient data are from a biobank (retrospective) and no data are collected for this study - was

ethics approval obtained?

Both reviewers commented positively on an interesting new approach that utilizes mechanistic modeling and statistical learning to cluster HFpEF patients into different phenotypes. This is seen as a potentially powerful approach to both deepen the fundamental understanding of this complex disease, with various etiologies, and to guide personalized approaches to therapy. A major weakness is the lack of a streamlined a clear presentation of the methodology, in particular with respect to the rationale for some aspects of the study design and methodological choices. The reviewers raise a number of questions that are helpful in making the manuscript more accessible to the broad readership, while also clarifying important details.

REFEREE COMMENTS

Referee #1:

HFpEF presents a wide range of clinical heterogeneities and the treatment of HFpEF requires deep phenotyping to determine the correct cohort for appropriate therapies. In this MS, Jones and colleagues present an interesting study combining mathematical modelling and machine learning techniques to deep phenotype HFrEF and HFpEF. They applied genetic algorithms to obtain patient-specific closed-loop models of the cardiovascular system based on the patient-specific transthoracic echocardiography (TTE) and right heart catheterization (RHC) measures. They show that analysis of the optimized model parameters using machine learning techniques (PCA, clustering) reveals mechanistic differences between HFpEF groups that are not seen when analyzing clinical measures alone. Interestingly, these analyses identify three subgroups of HFpEF, including an "HFrEF-like HFpEF".

* Strengths: This MS represents a powerful combination of mathematical modelling and machine learning and adds novel insights into HF classification. The originality is high, the study design is appropriate, and the conclusions are supported by robust data.

* Weakness: 1) Part of the MS can be hard to understand, considering the general readership of this journal. Some descriptions of the methods are unclear. 2) There are no examples of outputs for the mathematical models (time traces of outputs). How well the patient-specific model describes the clinical data? 3) Cohort size is small. 4) Potential minor error in editing the MS/supplementary materials.

* The authors are encouraged to add a figure to illustrate exemplary outcomes from the mathematical model (e.g., time courses of volume and pressure of the cardiovascular system components), and present data to show how well the patient-specific models describe the

clinical data (e.g., side-by-side comparison of model output vs clinical data). A comparison between a patient-specific model vs the normal (non-HF) model would be a plus.

* The closed-loop cardiovascular model is a reduced version of the Smith et al. model and is "similar" to the previous study from the same group. However, the description here is not clear enough to understand what exactly is reduced from the Smith model and what is different from the previous study (Colunga et al.). The schematic of the closed-loop cardiovascular model presented in Colunga et al., 2020 by the same group seems more informative than the one in Fig 1 and better link to the model descriptions in the supplementary materials. I would suggest adding a similar schematic to illustrate the model used in the present study (at least included in the supplementary material).

* Fig 4 shows machine learning-based analysis of the clinical data listed in Table 1. Were patient biometrics and EF also included in the clinical data for PCA and clustering determination? Since there were discrepancies in determining EF and CO, would these discrepancies affect the outcome of the analyses in Fig 4? Repeating these analyses with reported EF, for example, may help answer this question.

* PCA analyses were focused on two principal components which account for ~ 50% of the variance. What is the rationale for selecting the first 2 PCA components? Would adding components affect the PCA hull and thus the interpretation?

* It appears that the PCA and cluster analyses consider the clinical data or model parameters equally meaningful (without weighting) regardless of the underlying clinical and physiological significance of each biomarker/property. Could the authors comment on the data selection criterion for such analysis?

* It would be interesting to see where the normal model (non-HF) outputs (those describing the clinical data) would sit in the PCA space (Fig 4.) following the PCA transformation. Would it be distant from both PCA hulls (HFrEF and HFpEF)? Likewise in Fig 6?

* Would combine the clinical data and model parameters for the machine learning analyses give any meaningful results?

* Cohort size is small (e.g., 4 for HFpEF1). This should be discussed.

* Model parameters were optimized through a genetic algorithm implemented with Matlab. How

was the optimization convergence defined (criteria for reaching the best model parameters, methods for avoiding local minimum)? How were the machine learning algorithms implemented (e.g., using a Matlab built-in toolbox, or a python package)? What were the hyperparameters applied in the optimization? Adding such information would aid the reproducibility of the optimization procedure.

* Some of the figures and tables in the supplementary material are not referenced/described in the main text. Also, I could not locate Supplemental Table 1 (Table S1 in the Supplementary material shows parameters for the mathematical model instead).

Minor points:

* It would be helpful to indicate/describe the criteria applied to diagnose HFpEF and HFrEF for the original clinical data, to gain an understanding of how the clinical measurements listed in Table 1 are used in clinical settings.

* Definition of SV_LVOT in Eqs. 1&2?

* Colour-coding for the k-means clusters A&B in Fig. 4B appears inconsistent with the caption description. Likewise, Fig. 6C hierarchical cluster colours seem also confusing with the panel caption. Fig 6, it appears that patient 18 falls just outside as opposed to inside the HFrEF PCA hull.

* Under "Global sensitivity analysis", "E_sv was calculated nominally as described above", however, I could not locate the relevant description.

* The authors may consider adding the corresponding symbols of the parameters to the y-axis labelling in Figs. 7&8. This would be helpful for a general readership.

* Figs. S1 & 3 appear identical. Please verify.

* Fig captions: "Fig. S1. Supplemental Figure 3.", "Fig. S2. Supplemental Figure 7." seem confusing.

Referee #2:

Authors present an interesting approach to improve the phenotyping of HFpEF, a proof of concept with a small sample of subjects where they show the ability of computational models to integrate sparse sources of information and derive hidden useful diagnostic biomarkers. Two main HFpEF aetiologies are presented, one as an early HFrEF (cluster 1) and another as one with impaired relaxation (cluster 2).

In general, I found the reading of the manuscript cumbersome. The contributions are there, but the storyline is a bit bumpy into too many methodological details, and still left with the core methodological doubt: were parameters uniquely constrained? How ill-posed vs. well-posed was the problem of parameter identifiability? Despite this main limitation, it is a sound approach, and having found a sensible and rounded interpretation of results further increases the chances of plausibility (it does not matter if some patients will change clusters, the core group phenotype is the main finding here).

Comments:

- On the state of the art, authors miss an important work in the literature that tackles the issue of HFpEF, they are not the first to tackle this question by fitting computational models. They must discuss their findings with respect to this work: https://link.springer.com/article/10.1007/s12265-018-9816-y, specially related to their findings and general understanding of stiffness in HFrEF and HFpEF.

- In the broader context, authors are invited to frame their work within the context of the synergies between mechanistic and statistical models, i.e. the digital twin in cardiovascular medicine: https://academic.oup.com/eurheartj/article/41/48/4556/5775673

- On the methods

o The model framework would benefit from a rationale of what are the key parameters that need to be personalised, and any other relevant justification of the 16 parameters left (why not more, why not less), and a much smaller level of detail in other aspects.

o A specific example of model components/parameters that is quite confusing is the concept of "stressed volume of blood". This is a quite confusing concept to this reviewer, and a better clarity is encouraged, together with a rationale of why this model component is important in the research question posed.

o A cleaner presentation of the step-wise search of parameters would also be appreciated. What is the confidence that the method finds the unique set of parameters for each subject? This is the core question that remains unanswered.

- On results and interpretations:

o Although I quite like the idea that the model parameters are the ones revealing the true clusters, this claim needs to be put in context of methodological limitations, such as the very small data explored, and the serendipity of methodological choices (the specific cluster techniques and thus associated group belonging criteria, the uncertainty in data and model parameters found, the clinical measures might simply need extra PCA dimensions to reveal the same, etc).

END OF COMMENTS

Confidential Review **28-Apr-2021**

Response to Reviewers

Phenotyping heart failure using model-basedanalysis and physiology-informed machine learning

E. Jones, E. B. Randall, S. L. Hummel, D. M. Cameron, D. A. Beard, B. E. Carlson.

1 General Response

We appreciate the reviewers' thoughtful comments and critiques regarding this manuscript. We hope we have addressed all concerns to your satisfaction. In particular, we would like to make note of the following substantial changes:

- We have added two new Figures to the manuscript: new Figure 3, which gives a schematic of the model, and new Figure 4, which shows time-series for representative HFrEF and HFpEF patients.
- We have included figures for the normal subject and all the patients optimizations in the Supplemental Material (Figures S5-S34).
- After rerunning all of the optimizations, some changes occurred in the placement of some patients in the clustering analysis. Further detail are given below.

2 Reviewing Editor

See rev 1 - modifications of previous model(s) and parameters used in machine learning algorithms should be clarified to ensure reproducibility.
Response: We have addressed all of Reviewer 1's concerns below.

Patient data are from a biobank (retrospective) and no data are collected for this study - was ethics approval obtained?
Response: Ethics approval was obtained for the research use of all data in the Cardiovascular Health Improvement Project (CHIP) database. We added the following to the manuscript (Section Methods, subsection Clinical data, page 3):
"This retrospective data capture was approved by the Institutional Review Board at the University of Michigan, and informed consent was obtained for all subjects in the database."

Both reviewers commented positively on an interesting new approach that utilizes mechanistic modeling and statistical learning to cluster HFpEF patients into different phenotypes. This is seen as a potentially powerful approach to both deepen the fundamental understanding of this complex disease, with various etiologies, and to guide personalized approaches to therapy. A major weakness

is the lack of a streamlined a clear presentation of the methodology, in particular with respect to the rationale for some aspects of the study design and methodological choices. The reviewers raise a number of questions that are helpful in making the manuscript more accessible to the broad readership, while also clarifying important details.

3 Reviewer #1

HFpEF presents a wide range of clinical heterogeneities and the treatment of HFpEF requires deep phenotyping to determine the correct cohort for appropriate therapies. In this MS, Jones and colleagues present an interesting study combining mathematical modelling and machine learning techniques to deep phenotype HFrEF and HFpEF. They applied genetic algorithms to obtain patient-specific closed-loop models of the cardiovascular system based on the patient-specific transthoracic echocardiography (TTE) and right heart catheterization (RHC) measures. They show that analysis of the optimized model parameters using machine learning techniques (PCA, clustering) reveals mechanistic differences between HFpEF groups that are not seen when analyzing clinical measures alone. Interestingly, these analyses identify three subgroups of HFpEF, including an "HFrEF-like HFpEF".

* Strengths: This MS represents a powerful combination of mathematical modelling and machine learning and adds novel insights into HF classification. The originality is high, the study design is appropriate, and the conclusions are supported by robust data.

* Weaknesses:

1. Part of the MS can be hard to understand, considering the general readership of this journal. Some descriptions of the methods are unclear.
   Response: We have revised several parts of the Methods section for clarity. We sincerely hope this version is more comprehensible.

2. There are no examples of outputs for the mathematical models (time traces of outputs). How well the patient-specific model describes the clinical data?
   Response: We have now included model fits for representative HFpEF and HFrEF patients in new Figure 4. We have also included fits to all of the patients in the Supplemental Material.

   In the process of plotting out simulations for all patients, it became apparent that eleven patient fits had RV diastolic volumes that were much larger than the LV diastolic volumes. This was a result of the fact that no clinical measures of RV volumes are available to constrain this part of the model. Even though little is known about the relative sizes of RV and LV volumes in HFpEF and HFrEF, we decided to constrain the model to have a RV diastolic volume no greater than 1.5 times the LV diastolic volume. This changed some of the parameters slightly, and therefore, some shifting of patients occurred in the PCA and clustering analysis. Specifically, patient 30 moved from NCC to HFpEF1, and further shuffling occurred between NCC and HFpEF2 patients. These shifts do not change the overall conclusions made in this study.

3. Cohort size is small.
   Response: With a larger patient cohort, we would likely have more HFrEF-like HFpEFs. Our cohort size yielded a HFpEF1 subgroup with 5 patients, which is 25% of the HFpEF patients studied here. We have added the following to the manuscript (Section Discussion, subsection

Limitations, page 13):
"This methodology determined five HFpEF1 (HFrEF-like HFpEF) patients. Though this is a small cohort of subjects, this accounts for 25% of the total HFpEF patients in our study. It is of interest to see if this percentage holds with a larger patient cohort in the future."

4. Potential minor error in editing the MS/supplementary materials.
Response: We have proofread the manuscript and the Supplemental Material in detail.

3.1 Major comments

1. The authors are encouraged to add a figure to illustrate exemplary outcomes from the mathematical model (e.g., time courses of volume and pressure of the cardiovascular system components), and present data to show how well the patient-specific models describe the clinical data (e.g., side-by-side comparison of model output vs clinical data). A comparison between a patient-specific model vs the normal (non-HF) model would be a plus.
Response: We have now included model fits for representative HFpEF and HFrEF patients in new Figure 4. We have also included the fits to all patients for transparency in the Supplemental Material. Details of the model parameterization to normal cardiovascular function are included in the Supplemental material as well as the clinical measures that represent normal CV function that the model was optimized to.

2. The closed-loop cardiovascular model is a reduced version of the Smith et al. model and is "similar" to the previous study from the same group. However, the description here is not clear enough to understand what exactly is reduced from the Smith model and what is different from the previous study (Colunga et al.). The schematic of the closed-loop cardiovascular model presented in Colunga et al., 2020 by the same group seems more informative than the one in Fig 1 and better link to the model descriptions in the supplementary materials. I would suggest adding a similar schematic to illustrate the model used in the present study (at least included in the supplementary material).
Response: We have revised our methodology overview figure (Figure 1) to include a more informative schematic of the model. We have also added a new figure that contains the detailed schematic of the model in new Figure 3 similar to that of the Colunga et al. 2020 study.

3. Fig 4 shows machine learning-based analysis of the clinical data listed in Table 1. Were patient biometrics and EF also included in the clinical data for PCA and clustering determination?
Response: EF, height, and weight were not included in the PCA of the clinical data, but rather EF was used to determine the convex hulls as part of the clinical diagnosis. Since EF is calculated from LV volumes, we did not want to bias our results by effectively represent the LV volumes twice in the analysis by including EF in the PCA and clustering analysis of the clinical data. Our intent was to use only the clinical measures used for our parameter optimization. To clarify, we have added the following to the manuscript (Section Results, subsection HF subgroups determined from clinical data, page 8):
"All RHC and TTE patient data to which the model was optimized (Table 1) except EF, height, and weight were included in the PCA. Since EF was a major factor used to determine clinical diagnosis and LV diastolic and systolic volumes are already included in the PCA analysis, EF was excluded."

4. Since there were discrepancies in determining EF and CO, would these discrepancies affect the outcome of the analyses in Fig 4? Repeating these analyses with reported EF, for example, may help answer this question.
Response: Using the reported EF would not affect the outcome from the clinical data PCA since the EF was not used in the PCA itself so as not to bias the results towards the LV volumes and since the convex hulls were prescribed based on the patient's diagnosis (not the actual EF value). We have added the following to the manuscript (Section Results, subsection HF subgroups determined from clinical data, page 8)
"All RHC and TTE patient data to which the model was optimized (Table 1) except EF, height, and weight were included in the PCA. Since EF was a major factor used to determine clinical diagnosis and LV diastolic and systolic volumes are already included in the PCA analysis, EF was excluded."

5. PCA analyses were focused on two principal components which account for ∼50% of the variance. What is the rationale for selecting the first 2 PCA components? Would adding components affect the PCA hull and thus the interpretation?
Response: We selected the first 2 principal components for two reasons. (1) Each subsequent principal component accounts for <15% of the total variance, and they are not plotted for clarity. This is the case for both the clinical data PCA and the optimized parameter PCA. (2) Analysis and visualization of more than two principal components is more difficult in a multidimensional space, and it is questionable whether a clearer picture of our clusters would emerge with this small cohort. For these reasons, we chose to analyze the two-dimensional PCA. We have added the following to the manuscript (Section Methods, subsection Machine learning, subsubsection Principal component analysis, page 7):
"Figure 6A plots the two-dimensional space of the first two principal components describing more than 50% of the total variance. Subsequent principal components each accounted for less than 15% of the total variance and are not plotted for clarity."

6. It appears that the PCA and cluster analyses consider the clinical data or model parameters equally meaningful (without weighting) regardless of the underlying clinical and physiological significance of each biomarker/property. Could the authors comment on the data selection criterion for such analysis?
Response: The total variance for the clinical data and the optimized parameters are two separate entities, and it is a coincidence that the first two principal components describe similar percentages of their respective total variances. Hence, the two are not comparable. Also, selection criteria included the clustering as well as the PCA. We have added the following to the manuscript (Section Results, subsection HF subgroups determined from clinical data, page 8):
"The first two principal components of our clinical data PCA describe 52% of the total variance."
and (Section Results, subsection HF subgroups determined from optimized parameter values, page 9):
"The first two principal components of our optimized parameter PCA describe 59% of the total variance. Since the clinical data and parameter space are two entirely different representations of the patient population no conclusion should be drawn from the fact that both PCA analyses represent an equivalent total variance for the first two principal components."

Also, the reviewer is correct that based on knowledge of cardiovascular physiology and what is known about HFpEF and HFrEF we could weigh the clinical measures or the parameters we deem most critical. However, the intent of our approach is not to do this and see what measures or parameters exhibit emergent significance and then evaluate if this aligns with our expectations or not. All of the PCA and clustering analysis is performed on clinical measures or optimized parameters that are centered and then normalized by the standard deviation of the measure or parameter. The selection of the data used in this study is based on measures from RHC and TTE procedures that are present for all patients in the study. The selection of parameters is based on those that show sensitivity and minimal correlation with other parameters therefore giving us confidence in their optimized values. We have added the following to the manuscript (Section Methods, subsection Machine learning, page 7):
"To mitigate any bias in these analyses, no additional weighting is placed on any of the clinical measurements or optimized parameters."

7. It would be interesting to see where the normal model (non-HF) outputs (those describing the clinical data) would sit in the PCA space (Fig 4.) following the PCA transformation. Would it be distant from both PCA hulls (HFrEF and HFpEF)? Likewise in Fig 6?
Response: The reviewer makes a good point, and the normal outputs lie above the HFpEF hull. However, we do not think this adds to the depiction of the clustering analysis and have chosen not to show this in our plots.

8. Would combining the clinical data and model parameters for the machine learning analyses give any meaningful results?
Response: The reviewer raises another good point. However, we looked at the clinical data and optimized parameters separately as a first pass in this study. With a larger cohort, we might be able to discern subgroups more clearly by combining clinical measurements and parameters.

9. Cohort size is small (e.g., 4 for HFpEF1). This should be discussed.
Response: Please see above. Note that in the reanalysis performed here the HFpEF1 group has expanded to 5; however, as noted before a larger number of patients is needed to validate this HFpEF1 phenotype and possible find other underlying HFpEF subgroups.

10. Model parameters were optimized through a genetic algorithm implemented with Matlab. How was the optimization convergence defined (criteria for reaching the best model parameters, methods for avoiding local minimum)? What were the hyperparameters applied in the optimization? Adding such information would aid the reproducibility of the optimization procedure.
Response: The specifications for the genetic algorithm were a population size of 500 and a stall generation limit of 10 generations. All other hyperparameters are set to the MATLAB GA defaults. To test for convergence, we repeated the optimization for each patient 10 times and selected the run with the lowest cost. We added the following to the manuscript (Section Methods, subsection Mathematical modeling framework, subsubsection Optimization, page 7):
"Estimates for the adjustable parameters are obtained using the genetic algorithm with a population size of 500 and a stall generation limit of 10 generations implemented in MATLAB (MathWorks Natick, Ma). All other specifications were set to their default MATLAB value. To check to see if the parameter space was explored adequately, we ran the optimization for

each patient 10 times and observed a consistent residual across the best few runs. The run with the lowest cost was chosen for our final results. More details about MATLAB's implementation of the genetic algorithm can be found at mathworks.com."

11. How were the machine learning algorithms implemented (e.g., using a Matlab built-in toolbox, or a python package)?
Response: We used the built-in MATLAB functions for both the k-means and hierarchical clustering methods. For the k-means, we chose to group the points into 2 clusters using the L_1-norm. We mention this in the manuscript, which is highlighted in Section Methods, subsection Machine learning, subsubsection k-means clustering, page 7. For the hierarchical clustering, using the built-in linkage and cluster functions in MATLAB, we measure distance using the Ward metric and group into two clusters. This also is highlighted in the manuscript in Section Methods, subsection Machine learning, subsubsection hierarchical clustering, page 7. We have added the following to the manuscript (Section Methods, subsection Machine learning, page 7):
"We utilized one classification and two different clustering techniques using the built-in MATLAB k-means and hierarchical clustering functions to group individuals within a population based on similar characteristics. "

12. Some of the figures and tables in the supplementary material are not referenced/described in the main text. Also, I could not locate Supplemental Table 1 (Table S1 in the Supplementary material shows parameters for the mathematical model instead).
Response: We have updated the manuscript to reference all of the tables and figures in the Supplemental Material throughout the document.

3.2 Minor comments

1. It would be helpful to indicate/describe the criteria applied to diagnose HFpEF and HFrEF for the original clinical data, to gain an understanding of how the clinical measurements listed in Table 1 are used in clinical settings.
Response: The criteria for determining HFpEF and HFrEF is a history of heart failure symptoms and an ejection fraction above 50% and below 50%, respectively. We have added the following to the manuscript (Section Methods, subsection Clinical data, page 3):
"The criteria for determining whether a patient has HFpEF or HFrEF is a history of HF symptoms and an EF above 50% or below 50%, respectively."

2. Definition of SV_LVOT in Eqs. 1&2?
Response: We have defined this variable after equation (1) (Section Methods, subsection Data discrepancy/inconsistency, subsubsection Ejection fraction, page 4):
" ... where SV_{LVOT} is the stroke volume (SV) determined by LVOT VTI, and $V_{LV,diast}$ is the diastolic LV volume determined by MOD or Teichholz."

3. Colour-coding for the k-means clusters A&B in Fig. 4B appears inconsistent with the caption description. Likewise, Fig. 6C hierarchical cluster colours seem also confusing with the panel caption. Fig 6, it appears that patient 18 falls just outside as opposed to inside the HFrEF PCA hull.
Response: We corrected the coloring for the k-means clustering in Fig. 4B (now Fig. 6B in the updated manuscript) and hierarchical clustering in Fig. 6C (now Fig. 8C).

4. Under "Global sensitivity analysis", "E_{SV} was calculated nominally as described above", however, I could not locate the relevant description.
   Response: We have added the description for E_{SV} in the Supplemental Material, Equation S47. We have updated the manuscript (Section Methods, subsection Mathematical modeling framework, subsubsection Global sensitivity analysis, page 6):
   "... and E_{sv} was calculated using Equation (S47) in the Supplemental Material."

5. The authors may consider adding the corresponding symbols of the parameters to the y-axis labelling in Figs. 7&8. This would be helpful for a general readership.
   Response: We have added the symbols for the parameters in the figures as requested.

6. Figs. S1 & 3 appear identical. Please verify.
   Response: We have updated the supplemental figures to ensure that they are all distinct.

7. Fig captions: "Fig. S1. Supplemental Figure 3.", "Fig. S2. Supplemental Figure 7." seem confusing.
   Response: We have fixed all of the relevant supplemental figure captions.

4 Reviewer #2

Authors present an interesting approach to improve the phenotyping of HFpEF, a proof of concept with a small sample of subjects where they show the ability of computational models to integrate sparse sources of information and derive hidden useful diagnostic biomarkers. Two main HFpEF aetiologies are presented, one as an early HFrEF (cluster 1) and another as one with impaired relaxation (cluster 2).

In general, I found the reading of the manuscript cumbersome. The contributions are there, but the storyline is a bit bumpy into too many methodological details, and still left with the core methodological doubt: were parameters uniquely constrained?
Response: Yes, the model parameters were uniquely constrained. To clarify this, we have included the upper and lower parameter bounds in Table S1 of the Supplemental Material. We have also added the following to the manuscript (Section Methods, subsection Mathematical modeling framework, subsubsection Global sensitivity analysis, page 6):
"All parameters were varied within their physiological bounds, listed in the Table S1 of the Supplemental Material."

How ill-posed vs. well-posed was the problem of parameter identifiability?
Response: The problem is fairly ill-posed, as are most inverse problems in mathematical biology. However, we have used sensitivity analysis to address the practical identifiability of the parameters estimated here, that is, whether the parameters can be uniquely identified given the data available. We have added the following to the manuscript (Section Methods, subsection Mathematical modeling framework, subsubsection Global sensitivity analysis, page 6):
"Since the inverse problem investigated here is ill posed, a sensitivity analysis is performed to assess the practical identifiability of the parameters, i.e., determine which of the parameters can be identified with the given clinical patient data. "

Despite this main limitation, it is a sound approach, and having found a sensible and rounded interpretation of results further increases the chances of plausibility (it does not matter if some patients will change clusters, the core group phenotype is the main finding here).

4.1 Major comments

1. On the state of the art, authors miss an important work in the literature that tackles the issue of HFpEF, they are not the first to tackle this question by fitting computational models. They must discuss their findings with respect to this work: `https://link.springer.com/article/10.1007/s12265-018-9816-y`, specially related to their findings and general understanding of stiffness in HFrEF and HFpEF.
   Response: We have expanded our discussion of stiffness in HFrEF and HFpEF. In Wang et al, they observed increased LV stiffness in HFrEF, which we also observe. However, we also observe increased LV stiffness in the HFpEF population as well, so this parameter did not necessarily serve as an indicator of diastolic dysfunction in the HFrEF patients in our work. We have added a discussion of the Wang et al paper in the manuscript (Section Introduction, page 3):
   " Others have attempted to classify HF patients using clinical data to inform cardiovascular modeling (Wang et al., 2018). To our knowledge, ours is the first study that uses model-based analysis of clinical data and physiology-informed machine learning to determine subclassifications of HFpEF. "
   And (Section Discussion, subsection HFpEF1 as HFrEF-like HFpEF, page 11):
   "In the HFrEF population, we observe elevated λ_{LV}, an observation in accordance with the increased diastolic myocardial stiffness reported in HFrEF patients (Wang et al., 2018)."

2. In the broader context, authors are invited to frame their work within the context of the synergies between mechanistic and statistical models, i.e. the digital twin in cardiovascular medicine: `https://academic.oup.com/eurheartj/article/41/48/4556/5775673`
   Response: The reviewer makes a good point that this approach is somewhat similar to ours in the fact that they combine mathematical and statistical models to predict physiological function. We have added the following to the manuscript (Section Introduction, page 3):
   "This synergistic approach is in line with similar studies that combine mathematical and statistical techniques to predict physiological function at the patient-specific level (e.g., the "digital twin" (Corral-Acero et al., 2020)). "

3. The model framework would benefit from a rationale of what are the key parameters that need to be personalised, and any other relevant justification of the 16 parameters left (why not more, why not less), and a much smaller level of detail in other aspects.
   Response: We agree that a description of the particular parameters that are optimized would be useful for the reader. We have added the following to the manuscript (Section Methods, subsection Mathematical modeling framework, page 5):
   "Overall, the model used here has 6 states (compartmental blood volumes listed in Equations (S18)-(S23) of the Supplemental Material) and 16 parameters each with a specific physiological interpretation (Table 2). Equations for the reduced cardiovascular system model used in this study are given in Section S1 of the Supplemental Material and model code without parameter optimization can be found at (DOI: 10.5281/zenodo.5215892). Figure S5 in the Supplemental Material displays the model predictions for normal cardiovascular function corresponding to the parameters listed in Table S1 of the Supplemental Material. Figure 4 shows the model predictions for representative HFrEF (panels A-D) and HFpEF (panels E-H) patients showing the LV and systemic pressures (panels A and E), RV and pulmonary pressures (panels B and F), LV and RV volumes (panels C and G), and pressure-volume loops (panel D and H). Figures

for all model predictions for each patient can be found in Figures S6-S34 of the Supplemental Material."
And (Section Methods, subsection Mathematical modeling framework, subsubsection Global sensitivity analysis, page 6):
"This subset consists of parameters λ_{lv}, λ_{rv}, E_{lv}, and E_{rv}, which are used to describe cardiac function. All others are hemodynamic parameters that define cardiovascular function as a whole, which may be important for distinguishing particular subgroups of HFpEF."

4. A specific example of model components/parameters that is quite confusing is the concept of "stressed volume of blood". This is a quite confusing concept to this reviewer, and a better clarity is encouraged, together with a rationale of why this model component is important in the research question posed.
Response: We have added the following definitions for stressed and unstressed volumes (Section Methods, subsection Mathematical modeling framework, subsubsection Nominal parameters and initial conditions, page 6):
"This total blood volume is comprised of stressed and unstressed volumes. The unstressed blood volume is the volume in each compartment at which the pressure is zero. The stressed volume is the difference between the total and unstressed volumes."
Modeling stressed versus unstressed volume is important for our future work. In the HF field, regulation of stressed volume has recently been considered. We have included further rationale of our choice to model the stressed volume in the manuscript (Section Methods, subsection Mathematical modeling framework, subsubsection Nominal parameters and initial conditions, page 6):
"In this study, the percent of stressed volume remains the same across all patients. However, regulation of stressed and unstressed volume is a current topic of discussion in the field of HF (Fallick et al., 2011; Fudim et al., 2017), and the ability to change the ratio of stressed and unstressed volume can be explored in future studies."

5. A cleaner presentation of the step-wise search of parameters would also be appreciated.
Response: If the reviewer is asking for more details about the determination of the estimated subset of parameters, we agree that this is important. We have added the following to the manuscript (Section Methods, subsection Mathematical modeling framework, subsubsection Global sensitivity analysis, page 6):
" The Sobol' indices were calculated using Monte Carlo integration by computing $10^3(16+2) = 1.8e4$ *model evaluations similar to the procedure described in Randall et al. (Randall E. B. et al., 2021)."*

   If the reviewer is asking about the search for parameters using the genetic algorithm, a full description of the optimization method is slightly more comprehensive. The genetic algorithm in short randomly draws a set of 500 parameter sets from the bounded parameter space and first evaluates each one (generation 0). The parameter sets in this initial population then generate offspring using selection, crossover and mutation, mimicking evolutionary biology to generate a new fitter population (generation 1) and the process is repeated until a stopping criterion is reached. We have added the following to the manuscript (Section Methods, subsection Mathematical modeling framework, subsubsection Optimization, page 7):
   "More details about MATLAB's implementation of the genetic algorithm can be found at mathworks.com."

6. What is the confidence that the method finds the unique set of parameters for each subject? This is the core question that remains unanswered.
   Response: Technically, this method of optimization is a heuristic method that cannot guarantee a parameter set that produces a global minimum. Though this method may not necessarily find a unique identifiable subset for each patient, the subset found in this study is optimized for each subject and none of the parameters reach their physiological bounds. Furthermore, we have repeated each optimization 10 times to assess convergence. Hence, we are confident that this particular subset can help answer the physiological questions here. We have added the following to the manuscript (Section Methods, subsection Mathematical modeling framework, subsubsection Global sensitivity analysis, page 6):
   "This methodology produced a subset of uncorrelated parameters that can be estimated for each patient. In particular, none of the parameters reached their physiological bounds when estimated, giving confidence that the subset in Equation (3) is well prescribed to investigate the HF questions discussed here."

7. Although I quite like the idea that the model parameters are the ones revealing the true clusters, this claim needs to be put in context of methodological limitations, such as the very small data explored, and the serendipity of methodological choices (the specific cluster techniques and thus associated group belonging criteria, the uncertainty in data and model parameters found, the clinical measures might simply need extra PCA dimensions to reveal the same, etc).
   Response: We agree that the nature of the methods chosen may influence the way the clusters are chosen in some way, but we have used common, unsupervised methods and have made strides not to bias our results, even with such a small cohort. However, upon further analysis, we realized that a potential biomarker for the HFrEF-like HFpEF patients could be the elevated LV systolic and diastolic volumes. To address these concerns, we have added the following to the manuscript (Section Results, subsection Analysis of the clinical data from the 4 HF subgroups, page 11):
   "Overall, analysis of the clinical data with 4 HF subgroups reveals that all patients have higher pressures at rest, with HFpEF2 showing significantly higher pressures when compared to HFrEF. The main distinguishing factor between groups are systolic and diastolic LV volumes where HFrEF and HFpEF1 both have ventricular volume overload, signifying that greater LV volumes could be used as a biomarker for HFrEF-like HFpEF patients. "
   And (Section Discussion, page 11):
   "These groups could not be determined from clinical data alone but reveal that large LV volumes could be used as a biomarker to indicate HFrEF-like HFpEF patients."
   And (Section Discussion, subsection HFpEF1 as HFrEF-like HFpEF, page 11):
   "These results suggest a possible biomarker in high LV volumes for HFpEF patients, identifying patients belonging to HFpEF1."
   And (Section Discussion, subsection Limitations, page 13):
   "Here, two clustering methods were selected that used different approaches, but we could have used other common methods (e.g., mean-shift).The selection of k-means and hierarchical clustering in this study was made since these are robust and complementary approaches that can be applied to a wide variety of data sets. "
   And (Section Conclusions):
   "Moreover, our methodology reveals that potential biomarkers for identifying HFpEF-like HFrEF patients are elevated left ventricular systolic and diastolic volumes. However, these

biomarker differences necessary to determine HFpEF subgroups could not be distinguished based on the clinical data alone."

Dear Dr Carlson,

Re: JP-RP-2021-281845R1 "Phenotyping heart failure using model-based analysis and physiology-informed machine learning" by Edith Jones, E Benjamin Randall, Scott L Hummel, David M Cameron, Daniel A Beard, and Brian E. Carlson

I am pleased to tell you that your paper has been accepted for publication in The Journal of Physiology.

NEW POLICY: In order to improve the transparency of its peer review process The Journal of Physiology publishes online as supporting information the peer review history of all articles accepted for publication. Readers will have access to decision letters, including all Editors' comments and referee reports, for each version of the manuscript and any author responses to peer review comments. Referees can decide whether or not they wish to be named on the peer review history document.

Are you on Twitter? Once your paper is online, why not share your achievement with your followers. Please tag The Journal (@jphysiol) in any tweets and we will share your accepted paper with our 23,000+ followers!

The last Word version of the paper submitted will be used by the Production Editors to prepare your proof. When this is ready you will receive an email containing a link to Wiley's Online Proofing System. The proof should be checked and corrected as quickly as possible.

Authors should note that it is too late at this point to offer corrections prior to proofing. The accepted version will be published online, ahead of the copy edited and typeset version being made available. Major corrections at proof stage, such as changes to figures, will be referred to the Reviewing Editor for approval before they can be incorporated. Only minor changes, such as to style and consistency, should be made a proof stage. Changes that need to be made after proof stage will usually require a formal correction notice.

All queries at proof stage should be sent to TJP@wiley.com

Yours sincerely,

Professor Don M. Bers
Senior Editor
The Journal of Physiology
https://jp.msubmit.net
http://jp.physoc.org
The Physiological Society
Hodgkin Huxley House
30 Farringdon Lane
London, EC1R 3AW
UK
http://www.physoc.org
http://journals.physoc.org

professional video, design, and writing services to create shareable video abstracts, infographics, conference posters, lay summaries, and research news stories for your research at www.wileyauthors.com/eeo/promotion.

* IMPORTANT NOTICE ABOUT OPEN ACCESS *

Information about Open Access policies can be found here
https://physoc.onlinelibrary.wiley.com/hub/access-policies

To assist authors whose funding agencies mandate public access to published research findings sooner than 12 months after publication The Journal of Physiology allows authors to pay an open access (OA) fee to have their papers made freely available immediately on publication.

You will receive an email from Wiley with details on how to register or log-in to Wiley Authors Services where you will be able to place an OnlineOpen order.

You can check if you funder or institution has a Wiley Open Access Account here
https://authorservices.wiley.com/author-resources/Journal-Authors/licensing-and-open-access/open-access/author-compliance-tool.html

Your article will be made Open Access upon publication, or as soon as payment is received.

If you wish to put your paper on an OA website such as PMC or UKPMC or your institutional repository within 12 months of publication you must pay the open access fee, which covers the cost of publication.

OnlineOpen articles are deposited in PubMed Central (PMC) and PMC mirror sites. Authors of OnlineOpen articles are permitted to post the final, published PDF of their article on a website, institutional repository, or other free public server, immediately on publication.

Note to NIH-funded authors: The Journal of Physiology is published on PMC 12 months after publication, NIH-funded authors DO NOT NEED to pay to publish and DO NOT NEED to post their accepted papers on PMC.

EDITOR COMMENTS

Reviewing Editor:

Both reviewers are enthusiastic about the first paper to define the HFrEF and HFpEF phenotypic landscape combining clinical data and mechanistic model.

Note that Reviewer 2 suggests making all patient data available, which I think the authors should consider doing.

REFEREE COMMENTS

Referee #1:

I would like to thank the authors for thoroughly considering and addressing my previous comments. I find the MS has been greatly improved with the revision and do not have further concerns/comments.

Referee #2:

Authors have addressed all concerns raised, congratulations.

One final remark: there is still data to be made available that would enable other researchers to reproduce their results. Patient data used in this study is much more than Table 1! For example, the temporal transients of pressures and volumes are also needed and precious for the research community. Can all data used be made available?

1st Confidential Review 20-Aug-2021