Author’s response to reviews

Title: Accurate assessment of LV function using the first automated 2D-border detection algorithm for small animals - Evaluation and application to models of LV dysfunction

Authors:

Jana Grune (jana.grune@charite.de)
Daniel Ritter (daniel.ritter@charite.de)
Kristin Kräker (kristin.kraeker@charite.de)
Kathleen Pappritz (kathleen.pappritz@charite.de)
Niklas Beyhoff (niklas.beyhoff@charite.de)
Till Schütte (till.schuette@charite.de)
Christiane Ott (Christiane.Ott@dife.de)
Cathleen John (Cathleen.John@dife.de)
Sophie van Linthout (sophie.van-linthout@charite.de)
Carsten Tschöpe (Carsten.Tschoepe@charite.de)
Ralf Dechend (ralf.dechend@charite.de)
Dominik Muller (dominik.mueller@mdc-berlin.de)
Nadine Haase (nadine.haase@mdc-berlin.de)
Tilman Grune (scientific.director@dife.de)
Ulrich Kintscher (ulrich.kintscher@charite.de)
Wolfgang Kuebler (wolfgang.kuebler@charite.de)

Version: 1 Date: 25 Mar 2019

Author’s response to reviews:

Manuscript CARU-D-19-00014
Accurate assessment of LV function using the first automated 2D-border detection algorithm for small animals - Evaluation and application to models of LV dysfunction

Jana Grune; Daniel Ritter; Kristin Kräker; Kathleen Pappritz; Niklas Beyhoff; Till Schütte; Christiane Ott; Cathleen John; Sophie van Linthout; Carsten Tschöpe; Ralf Dechend; Dominik N. Muller; Nadine Haase; Tilman Grune; Ulrich Kintscher; Wolfgang M. Kuebler

We sincerely thank the editor and the reviewers for the thorough evaluation of our manuscript and the pertinent and helpful comments. We have gladly followed the reviewers’ suggestions and revised the manuscript accordingly. We have also responded to all concerns in a detailed point-by-point response. We feel confident that we have adequately addressed all concerns and hope that the editor and the reviewers will consider our work suitable for publication in Cardiovascular Ultrasound.

Point by Point Response

Reviewer reports:

Reviewer #1: Grune et al. have performed an interesting methodological study in order to evaluate cardiac function by novel automated two-dimensional software algorithm (Auto2DE) for small animals (mice and rats) and to compare it to the standard use of manual 2D-echocardiographic assessment (2DE). They have hypothesized that novel Auto2DE will provide rapid and robust data sets in accord with manually assessed data of small animals. The study is well conducted yet some issues should be clarified and additional data are mandatory.

We thank the reviewer for the supportive comments.

1) Table 1: the authors should add data on heart rate, LV end-systolic and end-diastolic thickness.

We thank the reviewer for this helpful suggestion. In response to this comment, we have now included data on heart rate in Table 1. Unfortunately, the LV end-systolic and end-diastolic thicknesses are not calculated by conventional 2DE or Auto2DE-tracings of B-mode images, since both methods are based on the monoplane Simpson’s method of discs. Although we share the reviewer’s opinion that this would be an interesting data set, these analyses could therefore not be performed and included in the present validation study.

2) Table 2: the authors should add data on heart weight of mice db/db+, db+/db+ and rats. Moreover, the authors should calculate the HW/BW ratio for each experimental condition.

We are grateful for this comment. In the revised manuscript, we have now included data on heart weight and heart weight/body weight-ratio for mice cohorts in Table 2.
The rat cohorts are part of ongoing studies and were used for evaluation purposes only in the current manuscript. Therefore, we provide the data on body weight, heart weight and heart weight/body weight-ratio in the point-by-point-response, but would prefer to omit them from the revised manuscript in order to allow future publication of these data in the context of their mechanistic background.

Table 1 Physiological data on heart weights of rat cohorts.

<table>
<thead>
<tr>
<th></th>
<th>HW</th>
<th>BW</th>
<th>HW/BW-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ctrl (n=5)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TetO (n=5)</td>
<td>660.0±22.2</td>
<td>398.6±7.8</td>
<td>1.65±0.1</td>
</tr>
<tr>
<td>mRen (n=5)</td>
<td>1869.0±70.4</td>
<td>447.8±15.4</td>
<td>4.18±0.1</td>
</tr>
<tr>
<td>TetO/mRen (n=6)</td>
<td>1012.7±85.2</td>
<td>345.2±12.9</td>
<td>2.93±0.2</td>
</tr>
<tr>
<td>Ctrl (n=5)</td>
<td>558.3±12.0</td>
<td>170.4±2.4</td>
<td>3.29±0.1</td>
</tr>
<tr>
<td>dTGR (n=8)</td>
<td>964.4±32.0</td>
<td>181.3±8.1</td>
<td>5.36±0.2</td>
</tr>
</tbody>
</table>


3) Table 2: animals recruited in the study have different age. The authors should discuss the impact of age on cardiac function assessed by echocardiography in order to justify the methodological approach in animals of different age.

We agree with the reviewer’s comment that aging impacts relevantly on cardiac performance. In general, the animals used in our study can be considered as adolescents (2-4.5 month old mice and 2-6 month old rats), when taking into account their total life span. Cumulative evidence from the literature reports on echocardiographic-assessed cardiac aging in mice upon 15 month or older [1–4] and rats upon 22 month or older [5–8]. Therefore, it seems unlikely that aging as a pathophysiological process has influenced the echocardiographic assessment in mice and rat cohorts described in our manuscript, since none of them would be considered as “aged”. That notwithstanding, we concur with the reviewer that for generalized applications of Auto2DE in the future it would be of interest to determine whether Auto2DE is sensitive enough to detect age-associated systolic function decline in aged animals. Since we presently do not have data on
aged cohorts, we added this fact as a limitation in the discussion section of our revised manuscript.

“Our study has some limitations, which should be taken into account when interpreting the presented data set. First, mouse and rat models analyzed with 2DE and novel Auto2DE were of different age. The possibility that aging as a pathophysiological process itself could have had impact on cardiac performance was not investigated and cannot be ruled out in our methodological approach to compare both aforementioned imaging modalities. Future studies may address whether Auto2DE is suitable for the analysis of age-associated cardiac function decline.”

(Discussion, p. 20, line 460-466, revised manuscript)

4) The authors should add in the text, figure legends and tables the sample size for each experimental group.

We thank the reviewer for this important advice and accordingly moved the paragraphs on Validation cohorts in mice and rats and Cardiovascular Disease Models from the Online Supplement to the Method section of the revised manuscript. These paragraphs contain all information on n-numbers/sample size for the individual study groups and animal models, and can now be found at:

(Methods, p. 6-8, line 133-170, revised manuscript)

To further address the reviewer’s concern, we included the sentence “Numbers in brackets indicate n-numbers” in the figure legends of Figure 1 and 2, where this information was missing. We apologize for this mistake and hope that the presentation of figures and figure legends is now more clear.

5) The authors should add informations about the type and dose of anesthesia used to perform the echocardiographic assessment in each experimental group. How much was the body temperature of mice and rats during the echocardiographic assessment?

The reviewer is of course correct, and we apologize for the missing information on body temperature and anesthesia, which was only provided through our referenced prior publications, but not directly stated in the method sections. In the method section of our revised manuscript, we have now included the following paragraph:

“Animals were anesthetized with 3% isoflurane (Baxter International, Deerfield, Illinois, USA) and fixed in supine position on a heatpad at 37°C (FUJIFILM VisualSonics, Toronto, Ontario, Canada). Isoflurane concentrations were further reduced to a minimum of 1-2% to achieve constant and comparable heart rates during image acquisition.”

(Methods, p. 8, line 177-181, revised manuscript)
We did not directly measure body temperature of mice and rats during echocardiographic examinations. We agree with the reviewer’s comment that body temperature may impact cardiac function but consider that placement on the 37°C heatpad should have avoided episodes of relevant hypothermia and corresponding decline in cardiac performance. However, since we did not directly monitor body temperature during image acquisition we added this fact as a potential limitation to the Discussion section of the revised manuscript. Further, we mentioned the possibility of potential confounding effects on assessment of cardiac function.

“Second, while animals were positioned on a 37°C heatpad during image acquisition we did not monitor body temperature directly. Therefore, we cannot fully exclude that variations in body temperature had potential confounding effects on the assessment of cardiac performance in the present study.”

(Discussion, p. 20, line 466-469, revised manuscript)

Reviewer #2: This study aimed to compare an automatic cardiac 2D echographic analysis software against manual approach.

For this purpose, authors compared ultrasound images acquired from both healthy and pathologic mice and rats.

Main conclusions of the authors were that the new automatic software can be used for automatic LV analysis in most of the enrolled animals but it still suffers of issues when dealing with specific pathologies.

The paper is well written but some issues should be discussed before considering it for publication.

1) Line 94: information about manufacturer are missing. Please correct.

We apologize for the missing information which is now included as follows:

“Advancements in ultrasound technology recently also paved the way for the first automated 2D-border detection algorithm (Auto2DE, FUJIFILM VisualSonics, Toronto, Ontario, Canada) for the assessment of LV systolic function in small animals.”

(Introduction, p. 4, line 92-95, revised manuscript)

2) Lines 124-125: MX550 or MS550 should be used for cardiac analysis in mice in order to obtain optimal results in terms of spatial resolution. Using this probe, conclusion of the study could be different. Please comment and add this point as a limitation.
The reviewer’s point is well taken, since the MX550D linear array transducer (32-55 MHz, Centre Transmit: 40 MHz, axial resolution: 40 µm) offers a higher spatial resolution, when compared to the MX400 linear array transducer (22-55 MHz, Centre Transmit: 30 MHz, axial resolution: 50 µm) used for echocardiographic image acquisition of the mice cohorts in our present study. Both transducers are recommended by the manufacturer for the application in cardiovascular disease models (https://www.visualsonics.com/resource/shared-stories/using-echocardiography-evaluate-development-heart-failure-aging-mice). The reason for using the MX400 linear array transducer in the present study was that higher resolution automatically limits penetration depth of ultrasound. Since we aimed to investigate type II DM (db+/db+) mice with higher body weights in comparison to control groups or other cardiovascular disease models, we chose the MX400 linear array transducer due to its higher penetration depth in order to keep the assessed data sets comparable (all mice cohorts measured with the same resolution/transducer). To address the reviewer’s concern we added the lower resolution of the MX400 vs the MX550 linear array transducer as a potential limitation in the Discussion section.

“In the present study, we opted to use the MX400 linear array transducer due to its superior performance in tissue penetration for the echocardiographic examination of mice cohorts, which however comes at the cost of a slightly lower spatial resolution as compared to the MX550D linear array transducer (Visualsonics).”

(Discussion, p. 20, line 469-473, revised manuscript)

3) Line 198: details about definition of Q1-Q4 quality levels should be added.

We thank the reviewer for this important suggestion. In response to this comment, we have now moved the paragraphs on Assessment of image quality form the Online Supplement to the Method section of the revised manuscript. These section can now be found at:

(Methods, p. 10, line 221-230, revised manuscript)

Additionally, we included a brief summary of image qualities in the respective text passage:

“To investigate the degree to which image quality affects the performance of Auto2DE analysis, cine loops of the murine validation cohort were graded into four distinct quality levels (Q1 - good, Q2 - fair, Q3 - poor, or Q4 - insufficient) by an expert in small animal echocardiography (Figure 2A).”

(Results, p. 12, line 282-285, revised manuscript)

4) Lines 190-193: These results should be commented in Discussion section.

We thank the reviewer for the opportunity to comment on these results. Accordingly, we have included a new paragraph into the revised manuscript to discuss why the difference between
absolute values of Auto2DE-assessed data and results obtained by standard of use 2DE is larger in rats as compared to mice:

“However, there seem to be differences regarding the tracing with Auto2DE between mice and rats, since absolute values of Auto2DE-assessed LV function parameters differed more in rats compared to mice. Due to the rat’s larger size, physiological noise originating from cardiac and respiratory motion is larger as compared to mice. As cardiorespiratory noise is well known to cause severe artifacts, such an effect may hamper the proper analysis of rat cine loops by Auto2DE [22–24]. Another explanation for this finding could be an underrepresentation of rats among the 200 expertly curated LV analysis traces used to train the Auto2DE algorithm (which is unknown as the original tracing library is not public domain).”

(Results, p. 17, line 387-395, revised manuscript)

5) Line 187: the number of healthy mice and rats is quite different (52 vs 14). What is the correct number in terms of power of the study? A post hoc analysis on study power should be added.

The reviewer’s point is well taken. Indeed, the lack of a significant difference between both imaging modalities may either indicate that both techniques are indeed comparable, or that the sample size taken was too small to detect such a difference. The latter concern is particularly relevant for the smaller rat validation cohort (n=14). Therefore, we did a post hoc power calculation based on the given sample size, effect size, and type I error α (tested for both 0.05 and 0.10) for the results of both validation cohorts and all cardiac function metrics to determine whether the tested sample size was large enough to allow for the detection of differences between 2DE and Auto2DE. The results are shown as a table in the Online Supplement S2. Corresponding sentences mentioning the post-hoc power analysis were included in the method and results sections of our revised manuscript. In brief, the analysis shows that the sample size of the mouse cohort (n=52) is sufficient to identify differences between 2DE and Auto2DE, since the calculated post-hoc power is >.99 for all cardiac function metrics. In contrast, the computed post-hoc power for EF and FS in the rat cohort is rather low due to the small sample size of only 14 animals. Additionally, we calculated how much rats we would have needed in the rat cohort to detect whether or not there was a difference between both imaging techniques. For a type I error α of 0.05 we would have required n=31 rats and n=44 for EF and FS, respectively. As these animal numbers were presently not available for analysis, we added this as a limitation in the discussion section of the revised manuscript as follows:

“Furthermore, post-hoc power analysis demonstrated that the sample size in the rat validation cohort was not sufficient to detect differences between 2DE and Auto2DE for the relative metrics EF and FS. Ongoing studies using larger sample sizes will thus be required to verify that the tested echocardiographic modalities yield similar results in rats.”

(Results, p. 20, line 473-476, revised manuscript)
6) Line 217: are these bias significative or not? Limits of agreements of biases should be added.

We thank the reviewer for the opportunity to add these important analyses. In the present study, the bias indicates the systematic measurement error by Auto2DE (method under evaluation) when compared to the considered true value assessed by conventional 2DE (standard of use method).

In the revised manuscript, we now provide 95% confidence intervals for bias in addition to the limits of agreement for all individual cardiac function parameters stratified by image quality in Table 1 and Figure 2D. The calculated Bias was found to be not statistical significant for all cardiac function parameters stratified by image quality, which is indicated by 95% Limits of Agreement levels, which were never statistically significant from zero. Therefore, we included a column with the information on statistical significance of Bias in Table 1.

7) Lines 211-214, 253, 263-264, 269-273 and 284-286: these sentences are comments and they should be deleted or moved to Discussion section.

We gladly followed the reviewer’s advice and deleted the sentences in lines 211-214, 253, 263-264, and 284-286 from the revised version of our manuscript. The sentences in line 269-273 (“In contrast to our original hypothesis that pronounced cardiac phenotypes would in general be less suitable for automated analysis tools, we found Auto2DE analysis to be specifically hampered in three diabetic animal models. Based on the impact of image quality on Auto2DE performance shown previously in this study, we speculated that diabetic cardiomyopathy may result in poor image quality per se“) provide the rationale for the next set of experiments (Figure 5) in our experimental approach. In order to allow the readership of Cardiovascular Ultrasound to follow our train of thought, we would thus prefer to keep this sentence as a logical introduction for the next paragraph.

8) Lines 298-307: inter/intra-observed variability was not evaluated with the present work. This part should be removed or strongly limited.

The reviewer is correct that inter- or intra-observer variability were not evaluated in the present work. That notwithstanding, we consider that this point should at least be briefly discussed. In line with the reviewer’s suggestion, we have thus condensed the respective paragraph by 50%, focusing mainly on prior experience with clinical automated software tools.

9) Line 352: numerical results should be moved to Results section.

Thank you – we have accordingly removed all numerical results from the Discussion section.
10) Lines 231-236: this analysis try to reflect the suitability of LV automatic analysis respect to manual one in different animal models. However, correlation only reflects a relationship between the two approaches while an evaluation of the agreement would useful here. Please comment and add the analysis or the limitation.

We thank the reviewer for the opportunity to comment on this point. We concur with the reviewer that calculation of Bias and Limits of Agreements according to Bland & Altman is the state-of-the-art statistical approach for the comparison of two techniques. In the original manuscript, we attempted to acknowledge for this fact by presenting exemplary Bland & Altman plots for EF and CO in one mouse and one rat model, respectively, in Figure 4. We opted for this approach as a comprehensive Bland & Altman analysis for all analyzed small animal models and cardiac function parameters would yield a table of 224 individual values (Bias and LOA for 8 cardiovascular disease models with 2 groups each for 7 different cardiac function parameters) which we feel would dilute rather than clarify content and focus. In the revised manuscript, we now state the fact that we provide Bland & Altman analyses only for a few exemplary parameters as a potential limitation in the discussion section.

“Lastly, it should be emphasized that correlation analyses reflect relationships rather than agreement between two imaging modalities. The latter was exemplarily addressed in detail by Bland & Altman analyses for two cardiovascular disease models and two cardiac function parameters.”

(Discussion, p. 20-21, line 477-480, revised manuscript)

References


