Authors:
Yosuke Nabeshima (y.nabeshima1016@gmail.com)
Hidehiro Namisaki (namisaki30@yahoo.co.jp)
Toshihiro Teshima (t.teshima.6.21@gmail.com)
Yasuhiko Kurashige (yasuhikokurashige@ybb.ne.jp)
Akiko Kakio (ta-kakio@jcom.home.ne.jp)
Azusa Fukumitsu (azusa_fukumitsu@yahoo.co.jp)
Yutaka Otsuji (otsujiy@med.uoeh-u.ac.jp)
Masaaki Takeuchi (takeuchi@med.uoeh-u.ac.jp)

Version: 1 Date: 03 Oct 2019

Author’s response to reviews:

We thank both reviewers for their thoughtful and constructive comments and suggestions. The following are our responses to each comment raised by the reviewers.

Responses to the Reviewer #1:

Nabeshima et al. present the improvement of the accuracy of two-dimensional echocardiographic (2DE) left ventricular (LV) measurements after a specific training program using cardiac magnetic resonance (CMR). They also argue that since the degree of improvement is individual, we would need to individualize training programs for each institution to reduce inter-institutional variability.

The use of coverage probability (CP) as a quality control parameter in this paper is original and potentially innovative. The paper is very well-written and statistical analyses are appropriate and well-conducted, giving a comprehensive good quality to the manuscript. However, there are some issues that must be addressed before the paper can be accepted.
Major:

1. The lack of standardization of practice period after training intervention and the lack of objective evaluation of that period make unfair conclusions about that issue.

RESPONSE: Thank you for this comment. We addressed these points in the study limitation as well as in the conclusion of this revised manuscript.

Page 17, Study limitations, lines 5-7:

Seventh, since the training was arbitrary, the lack of standardization of the practice period after the training intervention and the lack of objective evaluation were other limitations of this study.

Page 2, Abstract, Conclusions, lines 2-4:

Since the degree of improvements differed among hospitals, individualization of training programs and periodical objective evaluation may be required to reduce inter-institutional variability.

Page 17 Conclusion, lines 4-7:

To improve accuracy and reproducibility, individualization of the training program and a periodical objective evaluation in each subject for each hospital with a standardization of the practice period is required.

2. Assuming the known LV volumes underestimation with 2DE in comparison with CMR, how do the authors explain that 2DE-LV end-systolic volume (LVESV) was higher underestimated than 2DE-LV end-diastolic volume (LVEDV), resulting in LV ejection fraction (LVEF) overestimation?

RESPONSE: Thank you for this question. The following table shows percentage of underestimation of EDV and ESV in all subjects and each hospital at protocol 1.
<table>
<thead>
<tr>
<th></th>
<th>all</th>
<th>Site A</th>
<th>Site B</th>
<th>Site C</th>
<th>Site D</th>
<th>Site E</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDV</td>
<td>-14.1±22.4%</td>
<td>-6.6±23.8%</td>
<td>-3.8±25.5%</td>
<td>-9.8±20.5%</td>
<td>-29.4±14.9%</td>
<td>-28.8±14.1%</td>
</tr>
<tr>
<td>ESV</td>
<td>-28.1±21.1%</td>
<td>-11.3±21.8%</td>
<td>-15.1±22.7%</td>
<td>-28.3±16.9%</td>
<td>-46.4±10.9%</td>
<td>-41.4±13.3%</td>
</tr>
</tbody>
</table>

This table shows there was a systematic higher underestimation of ESV compared with EDV. We think that there are at least two reasons for this. First, tracing the endocardial border (between compacted layer and non-compacted layer) is more difficult when we trace the border at end-systole because the contraction of non-compacted layer makes the layer similar to the compacted layer. Thus, there is a risk of tracing a more inward position at end-systole. Second, the denominator (volumes assessed by CMR) is smaller at end-systole than at end-diastole. Thus, a similar degree of underestimation makes a higher degree of percentage underestimation at end-systole.

Minor:

1. The possibility to choose one from three consecutive cardiac cycles for 2DE examination could introduce a possible confounder, please detailed how you can minimize possible differences between consecutive cardiac cycle (i.e. requiring patient breath-holding).

RESPONSE: Thank you. This point has been added in the Methods section.

Methods, Acquisition and analysis of 2DE images, Page 5, lines 2 – Page 6, line2:

Echocardiographic image acquisition was systematically performed according to the American Society of Echocardiography guidelines(8) with breath-holding in all patients, from which we used apical LV-focused 4- and 2-chamber views for the analysis.

2. Since the acquisition of 2DE images impacts on the reliability of 2DE measurements, it is important to know if 2DE datasets were acquired by one or more different sonographers.

RESPONSE: Thank you. This point has been incorporated in the Methods section.

Page 5, Methods, Acquisition and analysis of 2DE images, lines 1-2:
All 2DE images were acquired by one expert sonographer using a commercially available ultrasound machine and equipment (iE33, Philips Medical Systems, Andover, MA, USA).

3. It was not specified if CMR manual LV endocardial border traces was made by the same operator as the expert that visited each hospital for training intervention.

RESPONSE: Thank you for pointing this out. This point has been incorporated in the Methods section.

Page 6, Methods, Acquisition and analysis of CMR images, second paragraph, lines 1-3:

CMR LV volumes and LVEF were measured by the same operator (the expert that visited each hospital for the training intervention) via feature tracking analysis (2D CPA MR; TomTec Imaging Systems, Unterschleissheim, Germany).

4. It seems that lines list does not perfectly fit with actual paper line-spacing, please correct and consider it for right positioning of the following issues.

RESPONSE: Thank you. We corrected this error as suggested by this reviewer.

5. Page 4 Line 44: please correct to "LV volumes" in line with the other parts of the article.

RESPONSE: Thank you. We corrected this error as suggested by this reviewer.

Page 4, Background, Second paragraph, line 2:

LV volumes and LVEF.(9)

6. Page 12 Lines 13 and 15: it is the first time that appears "EDV" and "ESV" acronyms in the paper, so please provide spelling or change to LVEDV and LVESV, as in the other parts of the paper.

RESPONSE: Thank you for pointing this out. We changed “EDV” and “ESV” to LVEDV and LVESV as suggested.

Page 12, Results, Second paragraph, lines 2-4:
Although CP values before and after the intervention were not different between the expert group and the novice group, the expert group had significantly lower CV values of LVEDV and LVESV than those of the novice group after the intervention.

7. Page 14 Lines 15 to 20: the sentence "Another possible explanation is that the number of echocardiographic examinations performed per one sonographer per year could negatively affect the impact of improvement" is not very clear and not supported by data, please try to better explain or rephrase it.

RESPONSE: Thank you for this comment. We deleted this sentence in the revised manuscript.

Responses to the Reviewer #2: This interesting paper, aimed to evaluate the efficacy of a novel training intervention based on CMR as a reference in improving inter-observer and inter-institutional variability of left ventricular (LV) volumes and LV ejection fraction (EF) measurements by two-dimensional echocardiography (2DE). The authors concluded that the training program, conducted by 54 sonographers from 5 hospitals, was useful in improving the accuracy of measuring LV volumes and LVEF in most of participants. Since it was observed that the degree of improvements differed among hospitals, authors proposed that an individualization of training programs may be required. I suggest the following major and minor revisions.

MAJOR ISSUES

Page 9 line 29 Results: Please, delete any comment from the "Results" section and move it to the "Discussion" (i.e. "As expected…").

RESPONSE: Thank you for this suggestion. Because the sentence “LV volumes measured by 2DE were significantly underestimated and LVEF was significantly overestimated compared to the corresponding values of CMR measurements in each hospital and for the whole cohort.” were the results observed in this study, we simply deleted “As expected,” in the revised manuscript.

Page 9 line 34-37 Results: About the paragraph "As expected, LV volumes measured by 2DE were significantly underestimated and LVEF was significantly overestimated compared to the corresponding values of CMR measurements in each hospital and for the whole cohort. However, the degree of this discordance differed substantially among hospitals.". Are differences in measurement accuracy among hospitals statistically significant? Please, report the P value.
RESPONSE: Thank you for the thoughtful comments. We created a new table that reports p values between hospitals (Supplemental Table 1).

Page 11 lines 41-44 Results: Please, add in text the p value of the comparison about the prevalence of expert sonographers among the hospitals.

RESPONSE: Thank you for this comment. We added p values for the comparison about the years of experience for echocardiography and the prevalence of expert sonographers among hospitals in the revised manuscript.

Page 11, Results, Sonographer characteristics and reproducibility, lines 1-4:

Years of experience for echocardiography was not different among five hospitals (p=0.801). If we defined an expert sonographer as > 10 years’ experience for echocardiography and a novice sonographer as ≤ 10 years’ experience,(13) the prevalence of the expert sonographer was not different (p=0.409) (Table 1).

Page 15 lines 7-41 Discussion: I suggest citing more briefly the studies by Daubert et al. and Thavendiranathan et al., without specifying data and results in deep details.

RESPONSE: Thank you for this suggestion. We revised the sentence according to the comment from this reviewer.

Page 15, Discussion, lines 1-10:

Daubert et al.(11) first reported the usefulness of CP for quality control of echocardiography. They reported that the CP for LVEDV was significantly improved after training; however, the CP for LVEF was excellent without training. These results suggest that intra-institutional variability of visual EF is very low under the supervision of an expert in each hospital, because the EF calculation is usually affected by visual EF. Thus, determination of visual EF is crucial to calculate the accurate EF. However, the authors did not report the measurement accuracy because of a lack of a reference value for LVEF. Thavendiranathan et al.(15) reported that using CMR in a training program is effective for improving inter-observer variability and accuracy of visual EF assessment.

Table 4: Most of CV and CP values in Table 4 are different from the corresponding values in Table 3. Please, align tables.
RESPONSE: Thank you for this important comment. In Table 3, we calculated CV and CP in 15 cases of protocol 1. However, we did not find one propensity score matched subject in protocol 2. Table 4 showed CV and CP values in 14 propensity score matched pairs in both protocols 1 and 2. This is the reason why the numbers of CV and CP in protocol 1 were a little bit different between Tables 3 and 4. This point was described in the Methods section.

Page 5, Patients:

We selected a consecutive series of 15 patients who underwent clinically indicated CMR examinations and also agreed to undergo 2DE examinations on the same day among the CMR database of our laboratory. We did not exclude any patients who had poor 2DE image quality. We then aimed to select an additional 15 patients with propensity score matching of LV end-diastolic volume (LVEDV) and LVEF on the CMR and 2DE image quality from the same CMR database. However, one of the patients among the first 15 cases did not have an appropriate propensity score match in the database. Thus, we did not use this patient for the comparative analysis before and after the intervention.

We also added number of subjects in protocols 1 and 2 in Tables 3 and 4 in the revised manuscript for clarity.

MINOR ISSUES

Page 11 lines 36-41 Results: The following sentence "If we defined an expert sonographer as > 10 years' experience for echocardiography and a novice sonographer as ≤ 10 years' experience." is not a result; please, move it to "Methods" section.

RESPONSE: Thank you for pointing this out. We moved the sentence to the Methods section.

Page 7, Methods, Examiners, lines 2-3:

We defined an expert sonographer as one with > 10 years of experience in echocardiography and a novice sonographer one having ≤ 10 years’ experience.(10)

Page 11 lines 51-54 Results: The following sentence "If we defined an active sonographer as the sonographer performing echocardiography examinations > 1000 cases/year" is not a result; please, move it to "Methods" section.

RESPONSE: Thank you for pointing this out. We moved the sentence to the Methods section.
Moreover, we defined an active sonographer as the sonographer who performs > 1000 echocardiography examinations/year.

Page 11 lines 54-56 Results: The information about the percentage of active sonographers that ranged from 0 to 100 is also in Table 1. Please, do not repeat in text the results that are cited in Tables.

RESPONSE: We deleted the sentence in the revised manuscript.

Page 12 lines 56-59 Discussion: Please, restate the sentence "Owing to its high spatial resolution, CMR SSFP images provide clearer visualization of the LV endocardial border, and the LV volumes and LVEF measured by CMR are associated with reduced inter-observer variability" as follows "Owing to its high spatial resolution and tissue contrast, CMR SSFP images provide clearer visualization of the LV endocardial border, and the LV volumes and LVEF measured by CMR are associated with reduced inter-observer variability".

RESPONSE: Thank you for this suggestion. We changed the sentence according to this reviewer’s suggestion.

Discussion, page 12 second paragraph, line 6 – page 13, line 2:

Owing to its high spatial resolution and tissue contrast, CMR SSFP images provide clearer visualization of the LV endocardial border, and the LV volumes and LVEF measured by CMR are associated with reduced inter-observer variability.(9)

Page 13 line 1 Discussion: I suggest citing more recent papers about the low inter-center variability of LV volumes and LVEF measurements by CMR (i.e. Marsella M. et al., Haematologica, 2011 and Aquaro G. et al JMRI, 2017) using the current sequences. The technique reported by Semelka RC in 1990 is obsolete: please, delete it.

RESPONSE: Thank you. We quoted 2 suggested references and deleted one reference (Semelka RC et al.) in the revised manuscript.
Page 13 lines 12-15 Discussion: I suggest citing recent papers (i.e. Aquaro G. et al JMRI, 2017) about reference values of LV volumes by CMR.

RESPONSE: Thank you. We quoted the suggested reference in the revised manuscript.

Page 13, Discussion, lines 6-9:

Since CMR provides reference values of LVEDV and LVESV for each patient (19), this offers an opportunity to examine precisely where to trace so as to obtain similar values for LV volumes on 2DE images.

Page 17 line 2: Please, delete the following sentence "Contrast echocardiography could replace CMR". There are no evidence supporting the sentence.

RESPONSE: Thank you for this thoughtful comment. We deleted the following sentence in the revised manuscript.

Contrast echocardiography could replace CMR, but contrast agent is not commercially available in Japan.

Table 3: I suggest specifying in table legend that p values are referred to the comparison between 2DE measurements in each hospital and CMR values.

RESPONSE: Thank you for this useful suggestion. We updated the Table 3 footnote in the revised manuscript.

Table 3: In Protocol 2 section, EF values in Site E and in Site A (41+-16) are like the global EF value (All sites), but there is not the symbol of p value <0,05. Maybe there is a typing error.

RESPONSE: Thank you. We double checked the numbers, and the values and statistical analysis were correct.