

University of Groningen

Balancing Speed and Accuracy in Cardiac Magnetic Resonance Function Post-Processing

Snel, Gert J H; Poort, Sharon; Velthuis, Birgitta K; van Deursen, Vincent M; Nguyen, Christopher T; Sosnovik, David; Dierckx, Rudi A J O; Slart, Riemer H J A; Borra, Ronald J H; Prakken, Niek H J

Published in:
Diagnostics

DOI:
[10.3390/diagnostics11101758](https://doi.org/10.3390/diagnostics11101758)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Snel, G. J. H., Poort, S., Velthuis, B. K., van Deursen, V. M., Nguyen, C. T., Sosnovik, D., Dierckx, R. A. J. O., Slart, R. H. J. A., Borra, R. J. H., & Prakken, N. H. J. (2021). Balancing Speed and Accuracy in Cardiac Magnetic Resonance Function Post-Processing: Comparing 2 Levels of Automation in 3 Vendors to Manual Assessment. *Diagnostics*, 11(10), [1758]. <https://doi.org/10.3390/diagnostics11101758>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.


Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Article

Balancing Speed and Accuracy in Cardiac Magnetic Resonance Function Post-Processing: Comparing 2 Levels of Automation in 3 Vendors to Manual Assessment

Gert J.H. Snel ^{1,*}, Sharon Poort ¹, Birgitta K. Velthuis ², Vincent M. van Deursen ³, Christopher T. Nguyen ^{4,5}, David Sosnovik ^{4,5,6}, Rudi A.J.O. Dierckx ⁷, Riemer H.J.A. Slart ^{7,8}, Ronald J.H. Borra ^{1,7} and Niek H.J. Prakken ¹

- ¹ Medical Imaging Center, Department of Radiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands; sharonpoort87@gmail.com (S.P.); r.j.h.borra@rug.nl (R.J.H.B.); n.h.j.prakken@umcg.nl (N.H.J.P.)
 - ² Department of Radiology, University Medical Center Utrecht, University of Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands; B.K.Velthuis@umcutrecht.nl
 - ³ Department of Cardiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands; v.m.van.deursen@umcg.nl
 - ⁴ Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, 149 13th Street, Charlestown, MA 02129, USA; Christopher.Nguyen@mgh.harvard.edu (C.T.N.); dsosnovik@mgh.harvard.edu (D.S.)
 - ⁵ Cardiovascular Research Center, Massachusetts General Hospital, Harvard Medical School, 149 13th Street, Charlestown, MA 02129, USA
 - ⁶ Division of Health Sciences and Technology, Harvard-MIT, 7 Massachusetts Avenue, Cambridge, MA 02139, USA
 - ⁷ Medical Imaging Center, Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands; r.a.dierckx@umcg.nl (R.A.J.O.D.); r.h.j.a.slart@umcg.nl (R.H.J.A.S.)
 - ⁸ Faculty of Science and Technology, Department of Biomedical Photonic Imaging, University of Twente, Dienstweg 1, 7522 ND Enschede, The Netherlands
- * Correspondence: g.j.h.snel@umcg.nl



Citation: Snel, G.J.H.; Poort, S.; Velthuis, B.K.; van Deursen, V.M.; Nguyen, C.T.; Sosnovik, D.; Dierckx, R.A.J.O.; Slart, R.H.J.A.; Borra, R.J.H.; Prakken, N.H.J. Balancing Speed and Accuracy in Cardiac Magnetic Resonance Function Post-Processing: Comparing 2 Levels of Automation in 3 Vendors to Manual Assessment. *Diagnostics* **2021**, *11*, 1758. <https://doi.org/10.3390/diagnostics11101758>

Academic Editor: Andrea D. Annoni

Received: 25 August 2021

Accepted: 22 September 2021

Published: 24 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Automating cardiac function assessment on cardiac magnetic resonance short-axis cines is faster and more reproducible than manual contour-tracing; however, accurately tracing basal contours remains challenging. Three automated post-processing software packages (Level 1) were compared to manual assessment. Subsequently, automated basal tracings were manually adjusted using a standardized protocol combined with software package-specific relative-to-manual standard error correction (Level 2). All post-processing was performed in 65 healthy subjects. Manual contour-tracing was performed separately from Level 1 and 2 automated analysis. Automated measurements were considered accurate when the difference was equal or less than the maximum manual inter-observer disagreement percentage. Level 1 (2.1 ± 1.0 min) and Level 2 automated (5.2 ± 1.3 min) were faster and more reproducible than manual (21.1 ± 2.9 min) post-processing, the maximum inter-observer disagreement was 6%. Compared to manual, Level 1 automation had wide limits of agreement. The most reliable software package obtained more accurate measurements in Level 2 compared to Level 1 automation: left ventricular end-diastolic volume, 98% and 53%; ejection fraction, 98% and 60%; mass, 70% and 3%; right ventricular end-diastolic volume, 98% and 28%; ejection fraction, 80% and 40%, respectively. Level 1 automated cardiac function post-processing is fast and highly reproducible with varying accuracy. Level 2 automation balances speed and accuracy.

Keywords: cardiac magnetic resonance; cardiac function; reproducibility; automation; standardization; tracing protocol

1. Introduction

Cardiovascular magnetic resonance (CMR) imaging is the reference standard for non-invasive assessment of ventricular volumes, function, and left ventricular (LV) mass [1]. Conventional post-processing by manually contour-tracing a stack of short-axis slices in end-diastolic and end-systolic phase is time-consuming [2,3]. As an alternative, commercially available artificial intelligence software has the ability to reduce post-processing time while increasing reproducibility [4], and several studies already showed promising results for Level 1 automated assessment [5–10]. The major challenge in measurement accuracy is basal tracing; dealing with cardiac through-plane motion combined with the complexity of basal anatomy which also varies between individuals [3,11]. These introduce disagreement with manual reference tracings, leading to different standards being used between sites, especially for the right ventricle (RV) [5,8,12]. Furthermore, Level 1 automated measurements can contain significant vendor-specific relative standard error in all slices [6,8,13].

To address these measurement inaccuracies in commercially available Level 1 automated post-processing solutions, the first phase of the Level 2 automated approach includes standardized manual adjustments of the Level 1 automated basal tracings. In the second phase of Level 2 automation, these results are corrected using the software package-specific relative-to-manual standard errors. Standardization of manual adjustments improves generalizability and reproducibility [14–16]. The Society for Cardiovascular Magnetic Resonance (SCMR) published general recommendations for standardized image interpretation of short-axis cines, including statements for identification of the basal slice and inclusion of specific anatomical structures [2]. Detailed operating procedures on handling basal slices with partially visualized LV or RV are nevertheless not available yet.

The first aim was to compare Level 1 automated post-processing of three commercially available software packages to manual contour-tracing on analysis time, reproducibility, and accuracy. The second aim was to assess the added value of the two phases of Level 2 automation; phase 1, manual adjustment of Level 1 generated basal tracings using a standardized contour-tracing protocol incorporating SCMR recommendations, and phase 2, correction of software package-specific relative-to-manual standard errors.

2. Materials and Methods

2.1. Study Population

This prospective single-center study was conducted in accordance with the Declaration of Helsinki, and the study protocol was approved by the Medical Ethical Committee of the University Medical Center Groningen (no. 2016/476; 19 December 2016). Sixty-five healthy volunteers (mean age, 30 ± 6 years; male sex, 57%) were recruited and signed informed consent prior to inclusion. The included subjects were divided among a training dataset ($n = 15$), validation dataset ($n = 10$) and segmentation dataset ($n = 40$) (study design in Figure 1).

2.2. CMR Imaging

All subjects underwent CMR on a 3.0 T scanner (MAGNETOM Prisma, Siemens Healthineers, Erlangen, Germany—software version VE11C) equipped with a 60-element phased-array coil. Electrocardiographically gated balanced steady-state free precession sequences were used to acquire long- and short-axis cines of 25 phases during breath-holds [17]. Long-axis cines were visualized in 4-chamber (4Ch), 2-chamber left ventricle (2ChLV), 2-chamber right ventricle (2ChRV) and outflow tract views. Short-axis cines were acquired from the atria to the ventricular apex with 6 mm slice thickness and 4 mm interslice gap [17]. All images were anonymized and exported directly from the scanner for offline analysis.

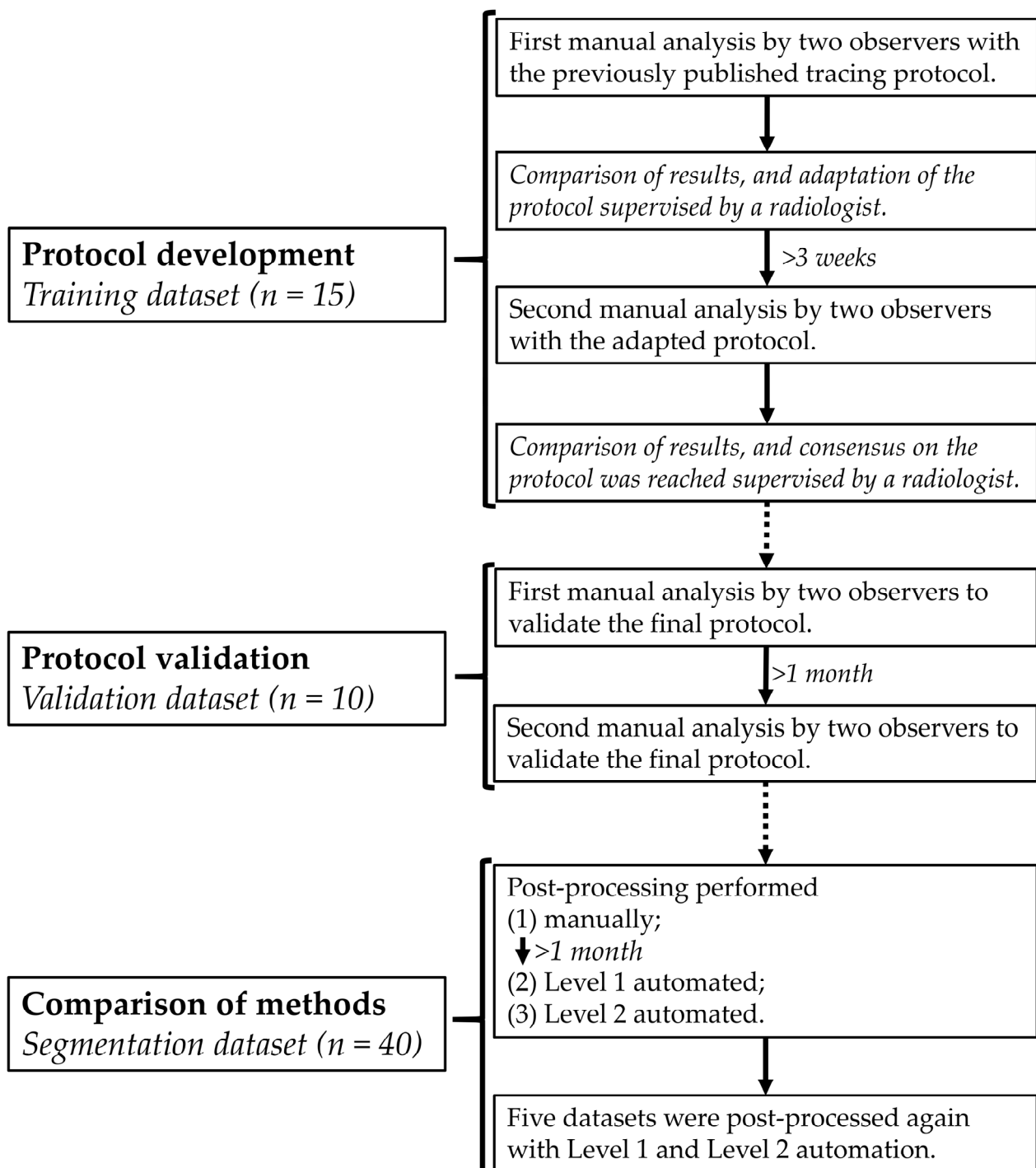


Figure 1. Flowchart of the study design.

2.3. Contour-Tracing Protocol

Two observers (a PhD student and a radiology resident, designated first and second), independently evaluated the previously published 1.5 T short-axis contour-tracing protocol using the training dataset (Figure 1) [18]. Manual contour-tracing was performed on a clinical workstation with contour-tracing software cvi42 (v.5.10.1, Circle Cardiovascular Imaging Inc., Calgary, AB, Canada), further referred to as vendor 1, enabling simultaneous comparison of long- and short-axis images. After both observers analyzed the training dataset once, the tracing protocol was adjusted to improve reproducibility in a consensus

meeting with an experienced cardiovascular radiologist. Adjustments included adaptation of rules how to in- or exclude the basal tracings, control mechanisms between LV and RV stroke volume and between end-diastolic and end-systolic LV mass, and CMR scan parameters, most importantly slice and interslice gap thickness, and number of phases [18]. After 3 weeks, both observers repeated the manual analysis of the training dataset using the adapted tracing protocol, after which consensus was reached on the final protocol between the above mentioned three observers.

The contour-tracing protocol with detailed textual and visual instructions is provided in Supplementary Material File S1. In brief, the cardiac phases with largest and smallest blood pool in the midventricular short-axis slice were defined as end-diastolic and end-systolic phase, respectively. In these two cardiac phases in all short-axis slices, endocardial contours were traced surrounding the LV and RV blood pool to assess volumes, and additionally epicardial contours surrounding the LV wall for LV mass calculation. Papillary muscles and trabecular tissue were included in the blood pool.

The two observers independently manually validated the contour-tracing protocol using the blinded validation dataset (Figure 1). After one month, both observers retraced the validation dataset to measure reproducibility.

2.4. Post-Processing Methods

The segmentation dataset was used to compare Level 1 and phase 1 of Level 2 automated results with gold standard manual contour-tracing results (Figure 1). Analysis time of all measurements was noted. Manual contour-tracing was performed by the first observer with vendor 1, the second observer checked all manual contours for adherence to the final tracing protocol.

Level 1 automated post-processing started one month after finishing manual contour-tracing. The first observer assessed three commonly used commercially available software packages; vendor 1, vendor 2 (Qmass MR, v.8.1, Medis Medical Imaging Systems, Leiden, The Netherlands) and vendor 3 (syngo.via, v.VB30A, Siemens Healthineers, Erlangen, Germany). The hardware specifications met the criteria as recommended by the vendors. All Level 1 automated contour-tracing was performed on short-axis cines exactly adhering to the instructions as provided in their respective user manuals. In all three software packages, the LV range was manually defined by selecting the mitral valve plane and apex on the 4Ch view. As vendor 1 allowed the use of multiple long-axis cines to define ventricular ranges, the LV range was also defined on the 2ChLV, and the RV range was defined on the 4Ch and 2ChRV. In vendor 2 it was not possible to define the RV range, and vendor 3 did not provide automated RV segmentation and was therefore excluded for RV evaluation.

Level 1 automated post-processing was performed with at least two weeks between vendors, and all subjects were post-processed in random order. None of the vendors used machine learning adapting to operator input, and therefore Level 1 automated contour-tracings were not affected by learning from user input. The first five datasets were automatically post-processed again for reproducibility assessment, and since this reproducibility proved to be nearly perfect, the number of datasets was not enlarged.

Phase 1 of Level 2 automated post-processing was performed directly after Level 1 automated assessment was completed. The analysis included manually adjusting the Level 1 automated contours of the most basal slice containing both LV and RV, and the even more basal slice, if only LV or RV remained visible, using the final tracing protocol. Subsequently, adherence to this protocol was checked by the second observer. Phase 1 of Level 2 automated procedure was repeated in the first five datasets, and this number proved to be sufficient to reliably assess post-processing reproducibility.

2.5. Statistical Analysis

The statistical analysis was performed using SPSS (v.24, Statistical Package for the Social Sciences, International Business Machines, Armonk, NY, USA). A p -value < 0.05 was

considered significant. Analysis time was noted as mean \pm standard deviation (SD) and analyzed using the analysis of variances with post-hoc tests. Reproducibility was analyzed using the intraclass correlation coefficient (ICC) and Bland-Altman statistics, noted as mean difference (± 1.96 SD) (95% limits of agreement) [19,20].

Agreement between methods was analyzed using Bland-Altman statistics and linear regression to investigate which part of the variation was removed by adjusting Level 1 automated basal tracings. Furthermore, the ratio between phase 1 of Level 2 automation and manual tracing results was calculated for each cardiac parameter. The median of this ratio was defined as the software package-specific relative-to-manual standard error used for phase 2 of Level 2 automation. Measurements differing equally or less than the maximum inter-observer disagreement percentage of manual tracings were considered accurate.

Based on Koo et al. [20], $ICC \geq 0.75$ was used to define good reliability. The ICC between Level 1 automated and manual tracing results was calculated for volumes and mass per vendor. If Level 1 automated results were unreliable, correcting basal slices alone was not expected to gain sufficiently accurate results. The statistical analysis of Level 2 automation was focused on vendors in which all parameters had an $ICC \geq 0.75$.

3. Results

3.1. Analysis Time

The analysis time of manual segmentation (21.1 ± 2.9 min) was significantly longer than Level 1 (2.1 ± 1.0 min) and Level 2 automated (5.2 ± 1.3 min) post-processing (Figure 2).

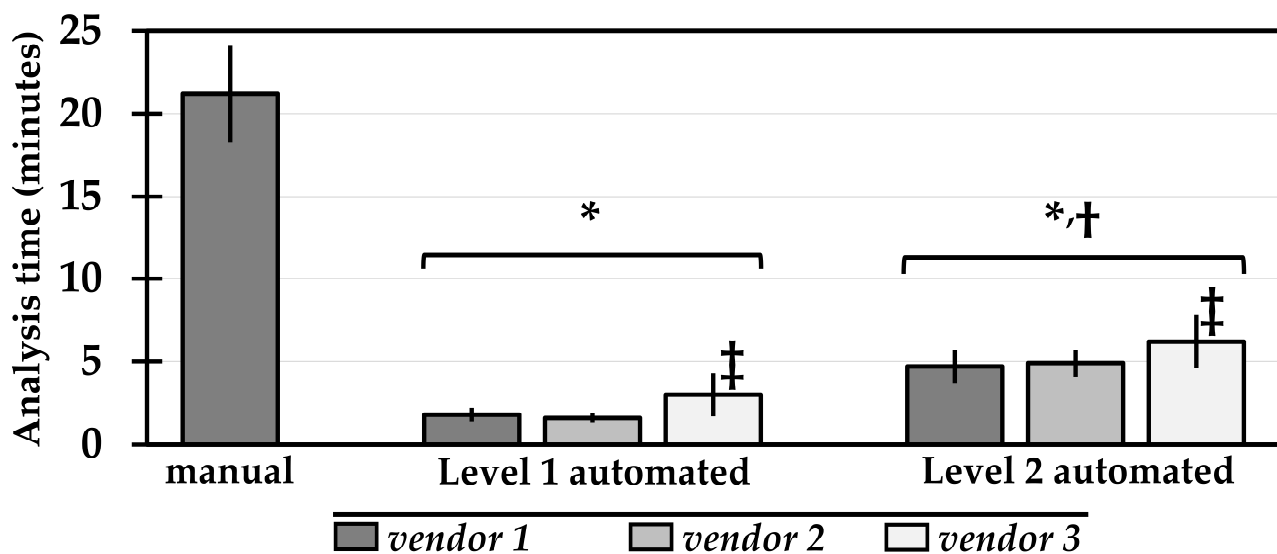


Figure 2. Analysis time (mean \pm standard deviation) of each post-processing method. $p < 0.05$ (*) versus manual, (†) versus Level 1 automated, (‡) versus other vendors within the same Level of automation.

3.2. Reproducibility

The reproducibility of the manually contour-traced validation dataset is shown in Table S1. The intra- and inter-observer agreement were excellent for all parameters ($ICC \geq 0.925$).

The intra-observer variability of the LVEF was -0.5% ($\pm 1.6\%$) and the inter-observer variability was -0.4% ($\pm 1.5\%$); for the RVEF, these were -0.2% ($\pm 1.2\%$) and -0.1% ($\pm 2.3\%$), respectively. The maximum inter-observer disagreement was 6%.

The reproducibility of the repeated Level 1 and phase 1 of Level 2 automated measurements is shown in Table S2. The reproducibility of all parameters was close to perfect in both Level 1 ($ICC = 0.974$ – 1.000) and phase 1 of Level 2 automated post-processing

(ICC = 0.949–1.000), higher than manual tracing. As expected, Level 1 post-processing showed higher reproducibility than phase 1 of Level 2 post-processing.

3.3. Accuracy

3.3.1. Level 1 Automation

LV parameters differed significantly from the gold standard manual measurements, and this variation was software package-specific (Figure 3, Table S3). Of the three software packages, vendor 1 performed best; the difference in LV end-diastolic volume (EDV) was -11 mL (± 21 mL), in LV ejection fraction (EF) 4% ($\pm 8\%$), and in LV mass 26 g (± 20 g). RV parameters showed larger differences than LV parameters compared to manual measurements. This variability in vendor 1 led to a small number of Level 1 measurements considered accurate (Figure 4, Table S4).

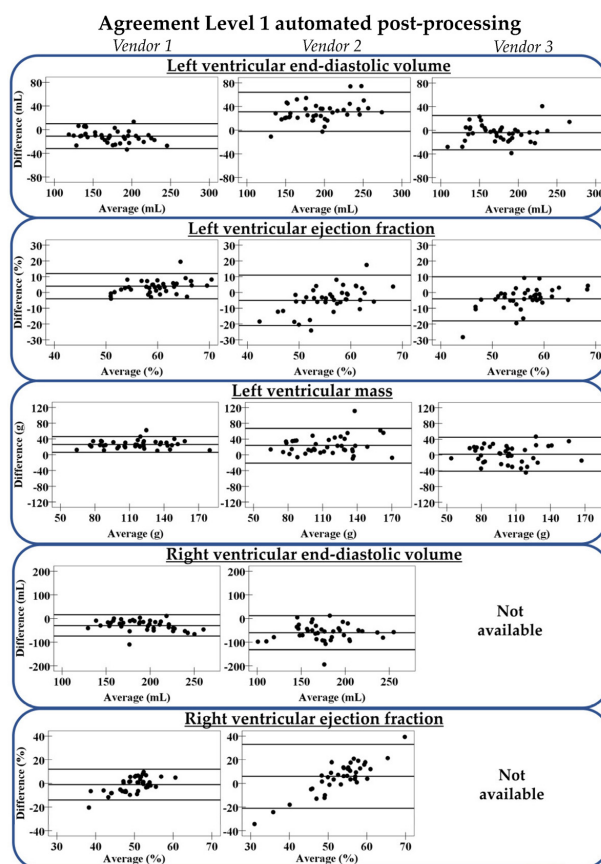


Figure 3. Agreement between Level 1 automation and gold standard manual contour-tracing. The three lines represent mean difference with limits of agreement. Vendor 3 does not support automated right ventricular post-processing.

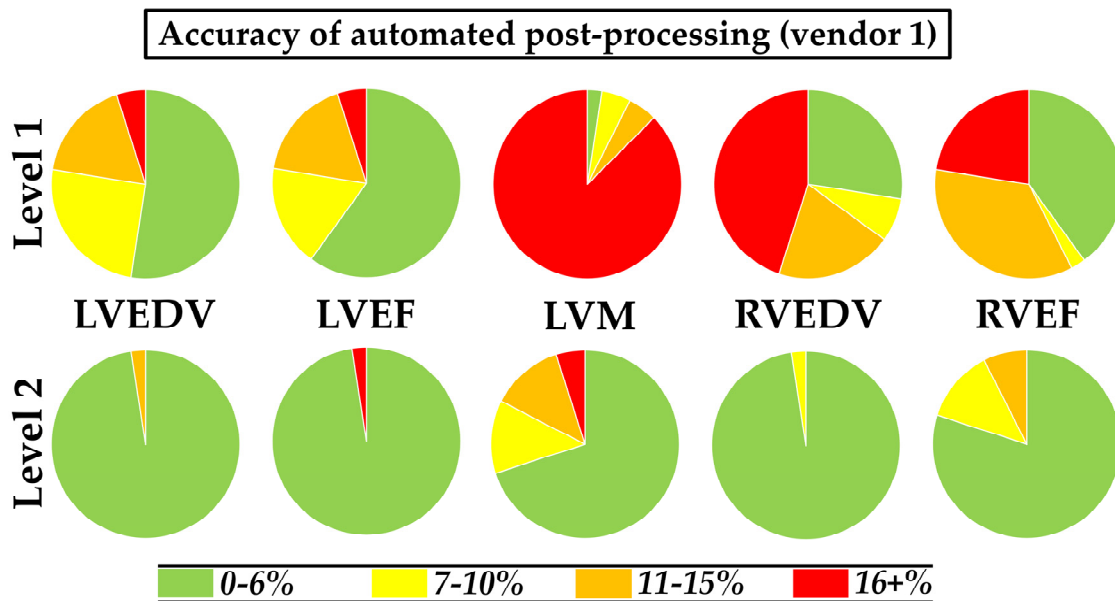


Figure 4. Accuracy of Level 1 and Level 2 automated post-processing results in vendor 1 compared to gold standard manual contour-tracing results. LV: left ventricle; EDV: end-diastolic volume; EF ejection fraction; M mass; RV: right ventricle.

Despite the differences between vendor 1 and the manual ground truth, the reliability was good for all cardiac parameters (lowest ICC was 0.771, highest ICC was 0.949) (Table S5). In vendor 2 and vendor 3, however, the reliability was moderate for all parameters ($ICC \leq 0.749$), except for the LVEDV. Therefore, further evaluation of Level 2 automation focused on vendor 1.

3.3.2. Level 2 Automation

The acquired Level 1 automated basal tracings in vendor 1 were manually adjusted using the standardized protocol (phase 1 of Level 2 automation), removing a large part of the variation between automated and manual measurements, especially for RV parameters (Table S3). The R^2 improved for the LVEDV from 0.91 to 0.98, LVEF from 0.53 to 0.72, LV mass from 0.86 to 0.93, RVEDV from 0.62 to 0.97, and RVEF from 0.17 to 0.63. Corresponding limits of agreement between phase 1 of Level 2 automated and manual results were smaller than those in Level 1 automation (Figure 5). Nevertheless, only a small number of phase 1 of Level 2 automated measurements was considered accurate (Table S4), as the software package-specific relative-to-manual standard error was still present.

The standard error in phase 1 of Level 2 automated measurements is provided in Table S6. The standard error for LVEDV was -7% (interquartile range (IQR); -8% , -6%) and for RVEDV this was -8% (IQR; -11% , -7%), suggesting that volumes were underestimated compared to manual contour-tracing. For the LV mass, this was the opposite as the standard error was positive (21% (IQR; 18% , 29%)). After correcting for these standard errors, most measurements in phase 2 of Level 2 automation were considered accurate: 98% of the LVEDV, 98% of the LVEF, 70% of the LV mass, 98% of the RVEDV, and 80% of the RVEF (Figure 4). Differences between phase 2 of Level 2 automated and manual results are shown in Figure 5.

Level 2 automated results of vendor 2 and vendor 3 are reported in Tables S3–S7. The differences with manual results were larger in vendor 2 and vendor 3 than in vendor 1. After correcting for the relative-to-manual standard errors, the number of accurate measurements in vendor 2 and vendor 3 remained low.

Agreement automated post-processing vendor 1

Level 1

Level 2 - phase 1

Level 2 - phase 2

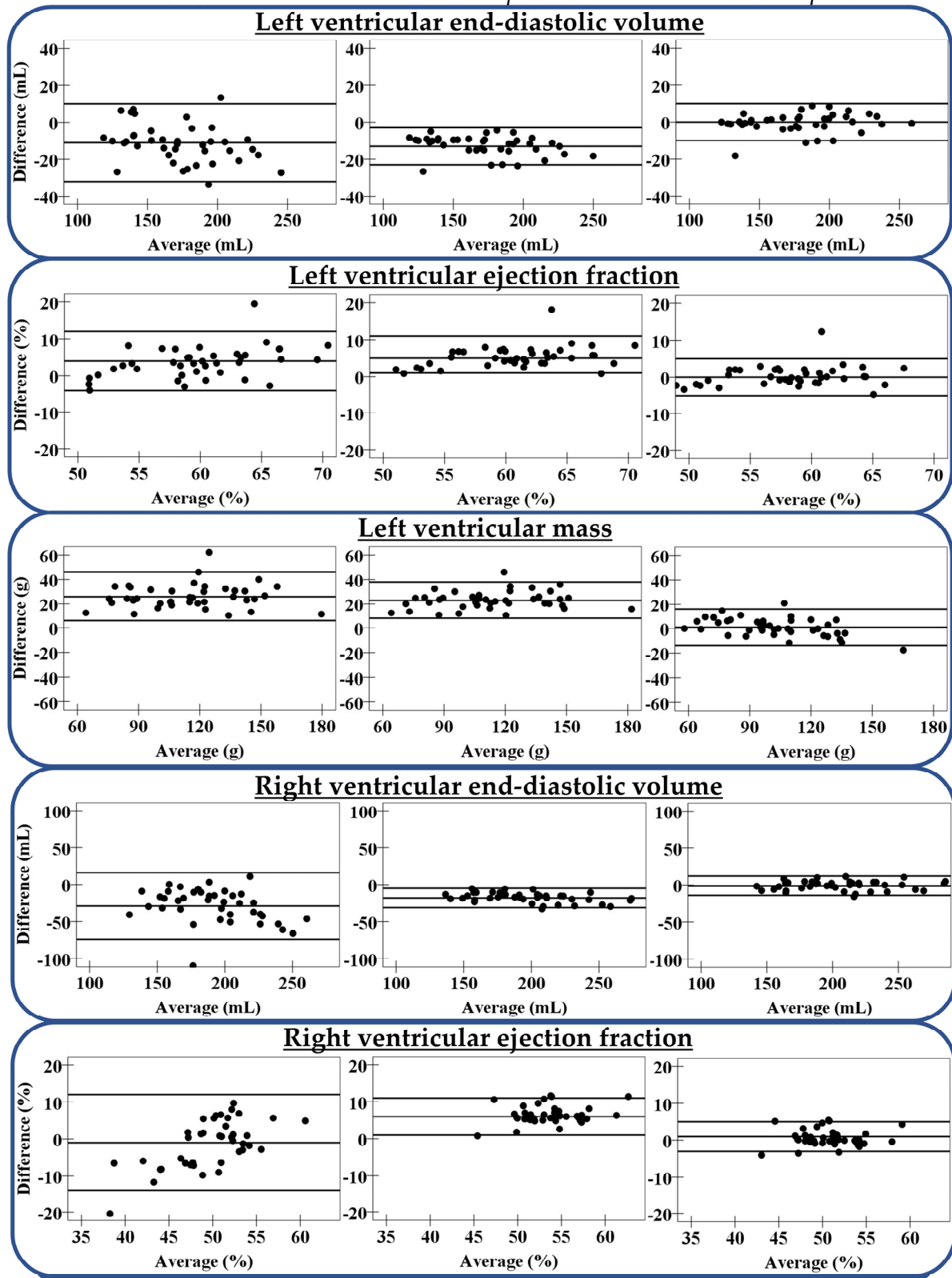


Figure 5. Agreement between different Levels of automated post-processing in vendor 1 and gold standard manual contour-tracing. From left to right: Level 1, Level 2 before correction of the relative-to-manual standard error (phase 1), and Level 2 after correction (phase 2). The three lines represent mean difference with limits of agreement.

4. Discussion

In this study we compared the speed and accuracy of Level 1 automated post-processing to manual contour-tracing. We additionally assessed the benefit of manually adjusting Level 1 automated basal tracings using a standardized protocol (phase 1 of Level 2 automation), combined with correction of the relative-to-manual standard error (phase 2 of Level 2 automation). Level 1 automated post-processing was ten times faster with higher reproducibility compared to manual contour-tracing, however, big differences resulted in few measurements that were considered accurate with large variation between software packages. Vendor 1 demonstrated the most reliable Level 1 automated results, and Level 2 automation showed high accuracy while preserving speed.

Level 1 automated post-processing in all three software packages was quick and highly reproducible, and this could improve clinical workflow as previously reported [8]. However, a large percentage of automated post-processing results was considered inaccurate, as the difference with manual results was higher than 6% (the maximum inter-observer disagreement). Therefore, automated post-processing seems inadequate for implementation in clinical workflow when contours remain unchecked. This supports the SCMR recommendation that the observer must check automated tracings for appropriateness [2].

Literature confirms that basal tracings introduce substantial variation between automated and manual post-processing [3,5,8,10,12], further validating manual adjustments. These adjustments should be standardized into easy to follow tracing instructions which incorporate SCMR recommendations (Supplementary Material File S1) [2,18]. Manual validation of these instructions showed higher intra- and inter-observer reproducibility than previous studies [6,16,21–25], proving it suitable for standardized manual correction of Level 1 automated basal tracings. The lower reproducibility in other studies could be caused by lacking standardization of unambiguous tracing instructions. In the image analysis section some studies referred to the SCMR guidelines [6,22], mentioned the definition of the most basal LV slice [21–25], or described structures that needed to be included in the RV [6,16,21,23,24]. Only Petersen et al. [23] showed similar reproducibility, possibly related to their control mechanisms which included checking equality of end-diastolic and end-systolic LV mass, and LV and RV stroke volumes in absence of significant valvular insufficiency, minimizing discrepancies. The time involved in manual contour-tracing both LV and RV using the standardized instructions (21 min) was representative for clinical routine [3,5], and comparable to previous studies that needed 13 to 14 min for only LV segmentation [26,27].

Vendor 1 demonstrated the most reliable automated post-processing with small variation compared to manual results after phase 1 of Level 2 automation, confirming differences in basal tracings are still the most challenging in automation. The other two vendors showed less reliable Level 1 automated post-processing, and substantial variation remained after adjusting the basal tracings, caused by inaccurate tracings in other slices. Consequently, phase 1 of Level 2 automation is only advantageous in software packages with relatively reliable Level 1 automation.

In all vendors, Level 1 automated contours are consistently traced too narrow or too wide resulting in relative-to-manual standard errors which, if known, can be used to improve accuracy. This correction step was performed during phase 2 of Level 2 automation. In the most reliable package (vendor 1), Level 1 automated volumes were systematically underestimated, and mass overestimated, caused by the endocardial contour being traced comparatively too close to the blood pool in all slices. After correction for this software package-specific relative-to-manual standard error, most measurements were considered accurate. In the other two packages, the calculated relative-to-manual standard errors were imprecise, evidenced by large IQRs, and this was caused by erroneous contour-tracing. In these cases, this final correction step is not beneficial.

In the best performing software package, completed Level 2 automation was still four times faster with higher reproducibility compared to manual, while the variability of volumes and mass remained comparable to literature [6,21–24]. Interestingly, the variability

of LVEF and especially RVEF in vendor 1 was lower than previously reported [6,21,22,24]. This demonstrates that the Level 2 automation approach can be an effective trade-off in software packages with reasonably well Level 1 automated post-processing.

There were several limitations in this study. First, we investigated three mainstream commercially available clinical software packages to perform automated post-processing where more software packages are available. However, we feel that the included software packages are representative of the field, as they are widely used in clinical assessment. Second, we performed gold standard manual contour-tracing on vendor 1 only. Nevertheless, this should not bias the results, as the included software packages ought to provide similar results as demonstrated previously [28]. Third, the segmentation dataset was post-processed multiple times which could induce some memory bias; however, this effect was minimized because of the time between measurements. For Level 1 automation, this bias was also irrelevant as user input was only required to define the ventricular range on long-axis cines, a step not applicable for manual contour-tracing. Fourth, we exclusively imaged healthy subjects in a relatively small group. Reproducibility of presented findings therefore need to be confirmed in a larger patient population with altered chamber geometry.

5. Conclusions

In the three studied vendors, Level 1 automated post-processing of cardiac volumes, function, and LV mass showed high reproducibility and speed 10-times faster than manual contour-tracing, however, when left unchecked differing degrees of variation made them inadequate for clinical use. Level 2 automation combines protocolized manual adjustments of Level 1 automated basal tracings with software package-specific relative-to-manual standard error correction while still four-times faster than manual assessment. In a reliable software package, Level 2 automation could be an effective approach to balance speed and accuracy in clinical workflow.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/diagnostics11101758/s1>, File S1: contour-tracing protocol. Table S1: Manual reproducibility, Table S2: Automated reproducibility, Table S3: Comparison between methods, Table S4: Accuracy of automated post-processing, Table S5: Intraclass correlation coefficient between automated and manual, Table S6: Software package-specific relative-to-manual standard error, Table S7: Differences between phase 2 of Level 2 automation and manual.

Author Contributions: Conceptualization, B.K.V., V.M.v.D., C.T.N., D.S. and N.H.J.P.; Formal analysis, G.J.H.S. and S.P.; Funding acquisition, R.A.J.O.D. and N.H.J.P.; Investigation, G.J.H.S. and S.P.; Methodology, G.J.H.S. and N.H.J.P.; Resources, R.A.J.O.D. and R.J.H.B.; Software, R.J.H.B.; Supervision, R.H.J.A.S. and N.H.J.P.; Writing—original draft, G.J.H.S. and N.H.J.P.; Writing—review and editing, S.P., B.K.V., V.M.v.D., C.T.N., D.S., R.A.J.O.D., R.H.J.A.S. and R.J.H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Dutch Heart Association (grant number 2016T042).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the University Medical Center Groningen (No. 2016/476; 19 December 2016).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ponikowski, P.; Voors, A.A.; Anker, S.D.; Bueno, H.; Cleland, J.G.F.; Coats, A.J.S.; Falk, V.; González-Juanatey, J.R.; Harjola, V.P.; Jankowska, E.A.; et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur. Heart J.* **2016**, *37*, 2129–2200. [[CrossRef](#)] [[PubMed](#)]
2. Schulz-Menger, J.; Bluemke, D.A.; Bremerich, J.; Flamm, S.D.; Fogel, M.A.; Friedrich, M.G.; Kim, R.J.; von Knobelsdorff-Brenkenhoff, F.; Kramer, C.M.; Pennell, D.J.; et al. Standardized image interpretation and post-processing in cardiovascular magnetic resonance—2020 update. Society for Cardiovascular Magnetic Resonance (SCMR): Board of Trustees Task Force on Standardized Post-Processing. *J. Cardiovasc. Magn. Reson.* **2020**, *22*, 19. [[CrossRef](#)]
3. Petitjean, C.; Dacher, J.N. A review of segmentation methods in short axis cardiac MR images. *Med. Image Anal.* **2011**, *15*, 169–184. [[CrossRef](#)]
4. Siegersma, K.R.; Leiner, T.; Chew, D.P.; Appelman, Y.; Hofstra, L.; Verjans, J.W. Artificial intelligence in cardiovascular imaging: State of the art and implications for the imaging cardiologist. *Neth. Hear. J.* **2019**, *27*, 403–413. [[CrossRef](#)] [[PubMed](#)]
5. Bai, W.; Sinclair, M.; Tarroni, G.; Oktay, O.; Rajchl, M.; Vaillant, G.; Lee, A.M.; Atung, N.; Lukaschuk, E.; Sanghvi, M.M.; et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J. Cardiovasc. Magn. Reson.* **2018**, *20*, 65. [[CrossRef](#)] [[PubMed](#)]
6. Backhaus, S.J.; Staab, W.; Steinmetz, M.; Ritter, C.O.; Lotz, J.; Hasenfuß, G.; Schuster, A.; Kowallick, J.T. Fully automated quantification of biventricular volumes and function in cardiovascular magnetic resonance: Applicability to clinical routine settings. *J. Cardiovasc. Magn. Reson.* **2019**, *21*, 24. [[CrossRef](#)] [[PubMed](#)]
7. Ruijsink, B.; Puyol-Antón, E.; Oksuz, I.; Sinclair, M.; Bai, W.; Schnabel, J.A.; Razavi, R.; King, A.P. Fully Automated, Quality-Controlled Cardiac Analysis From CMR: Validation and Large-Scale Application to Characterize Cardiac Function. *JACC Cardiovasc. Imaging* **2020**, *13*, 684–695. [[CrossRef](#)]
8. Bartolomé, P.; Caballeros, M.; Quílez-Larragan, A.; Núñez-Córdoba, J.M.; Fernández González, Ó.; Ezponda, A.; Bastarrika, G. Accuracy and Time-Efficiency of an Automated Software Tool to Assess Left Ventricular Parameters in Cardiac Magnetic Resonance Imaging. *J. Thorac. Imaging* **2020**, *35*, 64–70. [[CrossRef](#)] [[PubMed](#)]
9. Hautvast, G.L.T.F.; Salton, C.J.; Chuang, M.L.; Breeuwer, M.; O'Donnell, C.J.; Manning, W.J. Accurate Computer-Aided Quantification of Left Ventricular Parameters: Experience in 1555 CMR Studies from the Framingham Heart Study. *Magn. Reson. Med.* **2012**, *67*, 1478–1486. [[CrossRef](#)] [[PubMed](#)]
10. Böttcher, B.; Beller, E.; Busse, A.; Cantré, D.; Yücel, S.; Öner, A.; Ince, H.; Weber, M.A.; Meinel, F.G. Fully automated quantification of left ventricular volumes and function in cardiac MRI: Clinical evaluation of a deep learning-based algorithm. *Int. J. Cardiovasc. Imaging* **2020**, *36*, 2239–2247. [[CrossRef](#)] [[PubMed](#)]
11. Marcus, J.T.; Götte, M.J.W.; DeWaal, L.K.; Stam, M.R.; Van Der Geest, R.J.; Heethaar, R.M.; Van Rossum, A.C. The Influence of Through-Plane Motion on Left Ventricular Volumes Measured by Magnetic Resonance Imaging: Implications for Image Acquisition and Analysis. *J. Cardiovasc. Magn. Reson.* **1999**, *1*, 1–6. [[CrossRef](#)] [[PubMed](#)]
12. Paknezhad, M.; Marchesseau, S.; Brown, M.S. Automatic basal slice detection for cardiac analysis. *J. Med. Imaging* **2016**, *3*, 34004. [[CrossRef](#)] [[PubMed](#)]
13. Avendi, M.R.; Kheradvar, A.; Jafarkhani, H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med. Image Anal.* **2016**, *30*, 108–119. [[CrossRef](#)] [[PubMed](#)]
14. Suinesiaputra, A.; Bluemke, D.A.; Cowan, B.R.; Friedrich, M.G.; Kramer, C.M.; Kwong, R.; Plein, S.; Schulz-Menger, J.; Westenberg, J.J.M.; Young, A.A.; et al. Quantification of LV function and mass by cardiovascular magnetic resonance: Multi-center variability and consensus contours. *J. Cardiovasc. Magn. Reson.* **2015**, *17*, 63. [[CrossRef](#)]
15. Hedström, E.; Ishida, M.; Sepúlveda-Martínez, A.; Salehi, D.; Sperling, J.; Engblom, H.; Nagel, E. The effect of initial teaching on evaluation of left ventricular volumes by cardiovascular magnetic resonance imaging: Comparison between complete and intermediate beginners and experienced observers. *BMC Med. Imaging* **2017**, *17*, 33. [[CrossRef](#)]
16. Bonnemains, L.; Mandry, D.; Marie, P.Y.; Micard, E.; Chen, B.; Vuissoz, P.A. Assessment of right ventricle volumes and function by cardiac MRI: Quantification of the regional and global interobserver variability. *Magn. Reson. Med.* **2012**, *67*, 1740–1746. [[CrossRef](#)]
17. Kramer, C.M.; Barkhausen, J.; Bucciarelli-Ducci, C.; Flamm, S.D.; Kim, R.J.; Nagel, E. Standardized cardiovascular magnetic resonance (CMR) protocols 2020 update. *J. Cardiovasc. Magn. Reson.* **2020**, *22*, 17. [[CrossRef](#)]
18. Prakken, N.H.; Velthuis, B.K.; Vonken, E.-J.J.; Mali, W.P.; Cramer, M.-J.J. Cardiac MRI: Standardized Right and Left Ventricular Quantification by Briefly Coaching Inexperienced Personnel. *Open Magn. Reson. J.* **2008**, *1*, 104–111. [[CrossRef](#)]
19. Watson, P.F.; Petrie, A. Method agreement analysis: A review of correct methodology. *Theriogenology* **2010**, *73*, 1167–1179. [[CrossRef](#)]
20. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [[CrossRef](#)]
21. Hudsmith, L.E.; Petersen, S.E.; Francis, J.M.; Robson, M.D.; Neubauer, S. Normal human left and right ventricular and left atrial dimensions using steady state free precession magnetic resonance imaging. *J. Cardiovasc. Magn. Reson.* **2005**, *7*, 775–782. [[CrossRef](#)]

22. Suinesiaputra, A.; Sanghvi, M.M.; Aung, N.; Paiva, J.M.; Zemrak, F.; Fung, K.; Lukaschuk, E.; Lee, A.M.; Carapella, V.; Kim, Y.J.; et al. Fully-automated left ventricular mass and volume MRI analysis in the UK Biobank population cohort: Evaluation of initial results. *Int. J. Cardiovasc. Imaging* **2018**, *34*, 281–291. [[CrossRef](#)]
23. Petersen, S.E.; Aung, N.; Sanghvi, M.M.; Zemrak, F.; Fung, K.; Paiva, J.M.; Francis, J.M.; Khanji, M.Y.; Lukaschuk, E.; Lee, A.M.; et al. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. *J. Cardiovasc. Magn. Reson.* **2017**, *19*, 18. [[CrossRef](#)] [[PubMed](#)]
24. Alfakih, K.; Plein, S.; Thiele, H.; Jones, T.; Ridgway, J.P.; Sivananthan, M.U. Normal human left and right ventricular dimensions for MRI as assessed by turbo gradient echo and steady-state free precession imaging sequences. *J. Magn. Reson. Imaging* **2003**, *17*, 323–329. [[CrossRef](#)] [[PubMed](#)]
25. Csecs, I.; Czibalmos, C.; Suhai, F.I.; Mikle, R.; Mirzahosseini, A.; Dohy, Z.; Szűcs, A.; Kiss, A.R.; Simor, T.; Tóth, A.; et al. Left and right ventricular parameters corrected with threshold-based quantification method in a normal cohort analyzed by three independent observers with various training-degree. *Int. J. Cardiovasc. Imaging* **2018**, *34*, 1127–1133. [[CrossRef](#)] [[PubMed](#)]
26. Queirós, S.; Barbosa, D.; Engvall, J.; Ebberts, T.; Nagel, E.; Sarvari, S.I.; Claus, P.; Fonseca, J.C.; Vilaça, J.L.; D’Hooge, J. Multi-centre validation of an automatic algorithm for fast 4D myocardial segmentation in cine CMR datasets. *Eur. Heart J. Cardiovasc. Imaging* **2016**, *17*, 1118–1127. [[CrossRef](#)]
27. Bhuvra, A.N.; Bai, W.; Lau, C.; Davies, R.H.; Ye, Y.; Bulluck, H.; McAlindon, E.; Culotta, V.; Swoboda, P.P.; Captur, G.; et al. A Multicenter, Scan-Rescan, Human and Machine Learning CMR Study to Test Generalizability and Precision in Imaging Biomarker Analysis. *Circ. Cardiovasc. Imaging* **2019**, *12*, e009214. [[CrossRef](#)]
28. Zange, L.; Muehlberg, F.; Blaszczyk, E.; Schwenke, S.; Traber, J.; Funk, S.; Schulz-Menger, J. Quantification in cardiovascular magnetic resonance: Agreement of software from three different vendors on assessment of left ventricular function, 2D flow and parametric mapping. *J. Cardiovasc. Magn. Reson.* **2019**, *21*, 12. [[CrossRef](#)] [[PubMed](#)]