


Multivendor Evaluation of Automated MRI Postprocessing of Biventricular Size and Function for Children With and Without Congenital Heart Defects

Jelle P. G. van der Ven, MD,^{1,2} Wouter van Genuchten, MD,^{1,2} Zaheda Sadighy, MD,¹ Emanuela R. Valsangiacomo Buechel, PhD,³ Samir Sarikouch, PhD,⁴ Eric Boersma, PhD,⁵ and Willem A. Helbing, PhD^{1,6*} 

Background: Manually segmenting cardiac structures is time-consuming and produces variability in MRI assessments. Automated segmentation could solve this. However, current software is developed for adults without congenital heart defects (CHD).

Purpose: To evaluate automated segmentation of left ventricle (LV) and right ventricle (RV) for pediatric MRI studies.

Study Type: Retrospective comparative study.

Population: Twenty children per group of: healthy children, LV-CHD, tetralogy of Fallot (ToF), and univentricular CHD, aged 11.7 [8.9–16.0], 14.2 [10.6–15.7], 14.6 [11.6–16.4], and 12.2 [10.2–14.9] years, respectively.

Sequence/Field Strength: Balanced steady-state free precession at 1.5 T.

Assessment: Biventricular volumes and masses were calculated from a short-axis stack of images, which were segmented manually and using two fully automated software suites (Medis Suite 3.2, Medis, Leiden, the Netherlands and SuiteHeart 5.0, Neosoft LLC, Pewaukee, USA). Fully automated segmentations were manually adjusted to provide two further sets of segmentations. Fully automated and adjusted automated segmentation were compared to manual segmentation. Segmentation times and reproducibility for each method were assessed.

Statistical Tests: Bland Altman analysis and intraclass correlation coefficients (ICC) were used to compare volumes and masses between methods. Postprocessing times were compared by paired t-tests.

Results: Fully automated methods provided good segmentation (ICC > 0.90 compared to manual segmentation) for the LV in the healthy and left-sided CHD groups (eg LV-EDV difference for healthy children 1.4 ± 11.5 mL, ICC: 0.97, for Medis and 3.0 ± 12.2 mL, ICC: 0.96 for SuiteHeart). Both automated methods gave larger errors (ICC: 0.62–0.94) for the RV in these populations, and for all structures in the ToF and univentricular CHD groups. Adjusted automated segmentation agreed well with manual segmentation (ICC: 0.71–1.00), improved reproducibility and reduced segmentation time in all patient groups, compared to manual segmentation.

Data Conclusion: Fully automated segmentation eliminates observer variability but may produce large errors compared to manual segmentation. Manual adjustments reduce these errors, improve reproducibility, and reduce postprocessing times compared to manual segmentation. Adjusted automated segmentation is reasonable in children with and without CHD.

Evidence Level: 3.

Technical Efficacy: Stage 2.

J. MAGN. RESON. IMAGING 2022.

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.28568

Received Aug 4, 2022, Accepted for publication Dec 5, 2022.

*Address reprint requests to: W.A.H., Department of Pediatrics, Division of Cardiology, Erasmus University Medical Center-Sophia Children's Hospital, Room SP-2430, Wytemaweg 80, 3015 CN, Rotterdam, the Netherlands. E-mail: w.a.helbing@erasmusmc.nl

Grant support: This study was supported by the Magnetic Resonance Imaging Project of the Competence Network for Congenital Heart Defects funded by the German Federal Ministry of Education and Research (BMBF, FKZ 01G10210, 01G10601) J.P.G. van der Ven is supported by a research grant from the Dutch Heart Foundation (grant 2013T091).

From the ¹Department of Pediatrics, Division of Cardiology, Erasmus University Medical Center-Sophia Children's Hospital, Rotterdam, the Netherlands;

²Netherlands Heart Institute, Utrecht, the Netherlands; ³Pediatric Heart Centre, University Children's Hospital, Zurich, Switzerland; ⁴Department of Heart, Thoracic, Transplantation and Vascular Surgery, Hannover Medical School, Hannover, Germany; ⁵Department of Cardiology, Erasmus University Medical Center, Rotterdam, the Netherlands; and ⁶Department of Radiology, Erasmus University Medical Center, Rotterdam, the Netherlands

Additional supporting information may be found in the online version of this article

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

MRI is considered the reference standard for the assessment of volumetric parameters of the heart.¹ It can provide excellent spatial resolution, can provide excellent contrast between blood and tissue, and does not expose the patient to ionizing radiation.² However, several limitations of MRI limit its widespread use. First, image acquisition is relatively time-consuming and thus patient movement, including breathing, can reduce image quality. This is especially limiting in young children, who may not be able to follow instructions for breath-holds. Second, postprocessing of the images is needed to measure parameters of cardiac function. This is a time-consuming activity, which needs to be performed by trained observers.^{3,4} Even then, important interobserver variability exists among volumetric MRI measurements.^{3,5}

Automated postprocessing by software algorithms might help to reduce postprocessing time and reduce interobserver variability. Improvements in computing power, artificial intelligence, and the increased use of MRI have led to the development of several commercially available automated postprocessing software packages.⁶ These algorithms have shown promising results in the adult population.^{7–9} The algorithms are generally trained using large datasets of structurally normal adult hearts, such as the UK Biobank cohort,¹⁰ but may not include pediatric subjects. Assessment of ventricular volumes and function for the follow-up of congenital heart defects (CHD) is an important indication for MRI in children.² However, children and patients with CHD are mostly under-represented in—or excluded entirely from—the training datasets of automated segmentation algorithms.^{11,12} Structurally abnormal hearts may confound algorithms, which are trained exclusively on structurally normal hearts.¹³ Furthermore, there are large differences in body proportions between adults and children.¹⁴ These differences could affect the performance of postprocessing algorithms even in children with structurally normal hearts. For legal, ethical, and training reasons, human supervision and corrections are probably warranted when using such algorithms in practice for both children and adults.

We hypothesized that fully automated segmentation of ventricular volumes, although it eliminates observer variability, has poor agreement with manual segmentation and great vessel flow derived stroke volume (SV) in children. Automated segmentation followed by manual adjustment might increase agreement with manual segmentation, increase reproducibility compared to manual segmentation, and reduce postprocessing time compared to manual segmentation. Thus, the aims of this study were 1) to compare left and right ventricular (LV, RV) volumes and masses obtained from two automated cardiovascular MRI segmentation software packages (Medis Medical Imaging, Leiden, the Netherlands and SuiteHeart, Neosoft LLC, Pewaukee, USA), both fully automated and with manual adjustments, with those derived from manual segmentation in healthy children and children with different CHD diagnoses and 2) to compare postprocessing times, reproducibility, and

internal consistency (as determined by the comparison of great vessel flow derived stroke volume [SV] and segmentation derived SV) between techniques.

Materials and Methods

This retrospective study was approved by the institutional review board of the Erasmus Medical Center, Rotterdam, the Netherlands. All patients, or their legal guardians, gave informed consent for either participation in clinical research or reuse of clinical MRI studies for research purposes in accordance with national legislature.

Study Population

Twenty healthy children, as well as children with aortic stenosis or regurgitation ($n = 20$), tetralogy of Fallot (ToF) ($n = 20$), and univentricular CHD following Fontan palliation ($n = 20$) were included in this study. Healthy children were screened by medical history and physical examination. Cardiopulmonary conditions were exclusion criteria for healthy children.¹⁵ The CHD groups represent CHD affecting primarily the left ventricle (LV-CHD), CHD affecting the right ventricle, and complex CHD, respectively. Healthy children, patients with ToF, and patients with univentricular CHD were randomly selected—stratified by age groups—from previously published cohorts.^{15–20} Children with LV-CHD were selected from studies performed in clinical practice. The sample size of 20 subjects per group is adequately powered (i.e. statistical power 80%, $\alpha = 0.05$) to assesses volume differences of 5.5 mL between measurements, based on a measurement standard deviation of 8 mL.

Image Acquisition

Balanced steady-state free precession MR images were obtained using either a Signa 1.5 T whole-body MR imaging system for subjects from the University Children's Hospital Zurich and Erasmus MC Sophia Children's Hospital Rotterdam (General Electric, Milwaukee, USA) or an Intera R11 1.5-T whole-body MR scanner for subjects from the Bad Oeynhausen Hospital center (Philips Medical Systems, Best, the Netherlands). MRI protocols have previously been described extensively.^{21–24} Scan parameters are summarized in the online supplement S1. Images were obtained as a stack of 8–19 images in the short axis orientation from the apex up to the atrioventricular and ventriculo-arterial valves. Subjects were instructed to hold their breath during image acquisition. Young children unable to be instructed for effective breath-holding were sedated and images were obtained while freely breathing. No subjects received general anesthesia with intubation for MRI. Flow measurements were obtained on the level above the aortic and pulmonary valves, perpendicular to the vessel plane with a retrospectively gated 2D velocity-sensitized gradient echo sequence. Typical scan parameters were repetition time 4.5–15 msec, echo time 2.4–6.5 msec, flip angle 18°–30°, slice thickness 6–7 mm, field of view 290 × 220 mm, matrix 144–256 × 128, 1–6 phase-encoded lines per cardiac phase per cardiac cycle, velocity encoding 150 cm/sec, 24–40 reconstructed phases.

MRI Postprocessing

Images were analyzed on a Windows 10 workstation using commercially available software packages. Manual contouring of cardiac

structures was performed according to current postprocessing guidelines using Medis Suite version 3.2.60 (Medis Medical Imaging Systems, Leiden, The Netherlands).^{25,26} Endocardial contours were delineated both tracing trabeculae—and including papillary muscles in the ventricular wall mass—(trabeculated) and with smooth endocardial borders—including papillary muscles in the blood pool (smoothed). For univentricular CHD patients, all ventricular structures that unload into the aorta were considered as a single ventricle.²⁷ For analysis of great vessel flow the vessel were manually identified, automatically propagated across all phases, and manually reviewed and revised where needed. Forward flow, rather than net flow, was used to ensure fair comparisons to volumetrically derived stroke volumes, accounting for possible valve regurgitation.²⁵ Observers were blinded to the great vessel flow derived SV and the results of segmentation for intraobserver and interobserver variability during manual segmentation. Manual segmentation of the same study for intraobserver variability was performed with an interval of at least 1 month.

Two software packages for automated contour detection (Medis Suite version 3.2.60, Medis Medical Imaging, Leiden, the Netherlands and SuiteHeart 5.0.0, Neosoft LLC, Pewaukee, USA) automatically segmented the studies in the same manner as the manual segmentation protocol to the extent possible. SuiteHeart automated segmentation was performed tracing trabeculated contours. Medis automated segmentation was performed tracing smoothed contours, as no options for tracing trabeculated contours are available. For comparisons between manual segmentation and automated segmentation—that is, both fully automated and automated adjusted by human input—trabeculated automated contours were compared to trabeculated manual contours for SuiteHeart and smoothed automated contours were compared to smoothed manual contours for Medis. When not otherwise specified, we refer to trabeculated contours throughout this article. End-diastolic and end-systolic phases for the RV were manually adjusted to match these phases of the LV, in accordance with the manual segmentation protocol. For fully automated segmentation, the apical and basal limits of the ventricle were determined without human input or supervision. Examples of fully automated segmentation are provided as an online supplement S1.

We also assessed the effects of manual corrections on the results of automated segmentation. Three observers involved in manual segmentation (JvdV with 8 years of experience, ZS with 1 year, and WvG with 6 years) were instructed to adjust the automatically segmented contours, where necessary, in accordance with the manual segmentation protocol. Each study was segmented by JvdV twice and once by either ZS or WvG for analysis of interobserver variability. If the automated segmentation did not (correctly) identify ventricular structures in ≥ 3 slices, the automated segmentation was considered failed. If automated segmentation failed, measurements from manual segmentation for intraobserver or interobserver variability were used instead. Excluding these cases from analysis entirely may introduce bias as potentially only less complex cases are included in the statistical analysis. Furthermore, this approach more closely simulates the application of automation in practice, where studies need to be manually segmented if automated segmentation fails.

In this study, we assessed 1) differences between biventricular volumes, ejection fraction and mass derived between fully automated segmentation and automated segmentation followed by manual

adjustment vs. fully manual segmentation, 2) agreement of SV derived by different segmentation strategies with great vessel flow-derived SV, 3) Intraobserver and interobserver variability for different segmentation strategies, and 4) postprocessing times.

The postprocessing times of manual segmentation and manual corrections on automated segmentation were clocked by a stopwatch. Time recordings for each observer were started when observers started postprocessing (i.e. excluding study loading times) and were stopped when segmentation was complete. Time spent reviewing and adjusting contours were included in the postprocessing time.

Statistical Analysis

All continuous data are presented as mean \pm standard deviation (SD) or median and interquartile range (IQR), depending on the distribution. Volumetric data were not indexed for body surface area. Assessment of differences in volumes and mass between automated segmentation strategies versus manual segmentation, and for reproducibility analyses, were performed by paired *t* tests, Bland Altman analysis, and intraclass correlation coefficient (ICC). ICC is presented as ICC [95% confidence interval].²⁸ An ICC ≥ 0.90 was considered good agreement, and ICC ≥ 0.80 was considered moderate agreement.²⁹ ICC values are presented in color according to this classification in tables to allow for an easy overview. All statistical analyses were performed using R version 3.6.1 (R Core team, Vienna, Austria).

Results

Manual Segmentation

Subject characteristics, including ventricular volumes and function derived by manual segmentation, are summarized in Table 1. One (5%) healthy subject, 3 (15%) patients with LV-CHD, 1 (5%) patient with ToF, and 20 (100%) univentricular CHD patients were imaged while free breathing.

For healthy subjects, LV and RV SV derived from manual segmentation agreed excellently with great vessel flow derived LV and RV SV (difference 5.5 ± 14.5 mL, ICC: 0.90 [0.73–0.97] for the LV and 2.6 ± 11.6 mL, ICC: 0.92 [0.81–0.97] for the RV, see online supplement S1). This was also true for the LV SV in the LV-CHD group (difference -11.9 ± 11.6 mL, ICC: 0.97 [0.93–0.99]), but not for the RV SV (difference 1.8 ± 26.8 mL, ICC: 0.73 [0.29–0.91]). For ToF and univentricular CHD, the agreement between manual segmentation SV and great vessel flow-derived SV ranged from poor for single ventricular SV (difference -6.3 ± 18.2 mL, ICC: 0.68 [0.39–0.84]) to moderate for RV SV of ToF patients (difference -9.6 ± 16.5 mL, ICC: 0.87 [0.69–0.95]).

Intraobserver and interobserver variability of manual segmentation is presented in Table 2. Intraobserver and interobserver variability was excellent for healthy children and patients with LV-CHD (ICC > 0.91 for all parameters). For patients with ToF and univentricular CHD, reproducibility ranged from poor (eg interobserver variability of RV ESV for ToF: 12.9 ± 21.8 mL, ICC: 0.76 [0.54–0.88]) to excellent

TABLE 1. Patient Characteristics

	Healthy children (<i>n</i> = 20)	LV-CHD (<i>n</i> = 20)	ToF (<i>n</i> = 20)	Univentricular CHD (<i>n</i> = 20)
Patient characteristics				
Age	11.7 [8.9–16.0]	14.2 [10.6–15.7]	14.6 [11.6–16.4]	12.2 [10.2–14.9]
Male sex	9 (45%)	19 (95%)	15 (75%)	16 (80%)
BSA	1.39 [1.04–1.70]	1.58 [1.12–1.77]	1.58 [1.36–1.73]	1.29 [1.13–1.54]
Subjective image quality				
No diagnostic quality	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Reduced diagnostic quality	2 (10%)	3 (15%)	1 (5%)	2 (10%)
Many artifacts	2 (10%)	1 (5%)	2 (10%)	5 (25%)
Some artifacts	5 (25%)	7 (35%)	4 (20%)	5 (25%)
Optimal diagnostic quality	11 (55%)	9 (45%)	13 (65%)	8 (40%)
Ventricular size and function				
LV EDV (mL)	98 [73–136]	121 [103–220]	128 [105–145]	SiV EDV 115 [100–135]
LV ESV (mL)	33 [25–46]	39 [23–92]	47 [35–57]	SiV ESV 51 [47–66]
LV SV (mL)	63 [46–88]	88 [69–107]	77 [65–94]	SiV SV 61 [49–71]
LV EF (%)	66 [64–68]	65 [59–71]	61 [57–65]	SiV EF 51 [49–57]
LV Mass (g)	75 [49–92]	109 [84–206]	80 [59–102]	SiV Mass 72 [67–92]
RV EDV (mL)	110 [79–139]	142 [96–173]	186 [146–220]	-
RV ESV (mL)	41 [32–57]	60 [34–82]	84 [60–112]	-
RV SV (mL)	63 [46–87]	73 [56–88]	98 [87–120]	-
RV EF (%)	60 [59–63]	60 [50–63]	53 [50–57]	-

Reported volumes and masses are manually segmented by a single observer (JvdV).
 LV = left ventricular; ToF = tetralogy of Fallot; CHD = congenital heart disease; BSA = body surface area; EDV = end diastolic volume; ESV = end systolic volume; SV = stroke volume; EF = ejection fraction; RV = right ventricular; SiV = single ventricular.

(eg intraobserver variability of LV EDV for ToF: 1.9 ± 9.4 mL, ICC 0.96 [0.91–0.98]). In general, interobserver variability was poorer than intraobserver variability and variability was poorer in the end-systolic phase than in the end-diastolic phase.

Fully Automated Segmentation

Automated segmentation was successful for all studies using SuiteHeart. Six studies were unable to be segmented automatically using Medis (one healthy subject, two with LV-CHD, one with ToF and two with univentricular CHD). Failed automated segmentation was more common in younger patients (9.0 [3.4–10.1] vs. 13.2 [10.4–15.8], $P = 0.002$) and for free breathing studies (3/9 vs. 3/68, $P = 0.010$). Failed automated segmentation was not

statistically significantly related to any diagnosis group or manual segmentation time.

Differences in measurements between manual and fully automated segmentation are shown in Table 3. For healthy children, agreement between fully automated and manual segmentation is good for left ventricular parameters for both software suites. Agreement for right ventricular parameters ranged from poor-to-good Medis and moderate-to-good for SuiteHeart (eg RV EDV -3.0 ± 20.8 mL, ICC: 0.92 [0.83–0.96] for Medis and 4.6 ± 16.9 mL, ICC: 0.92 [0.82–0.96] for SuiteHeart). For patients with LV-CHD, the agreement between fully automated and manual segmentation of both software suites was similar to that of healthy children, that is, agreement is good for LV parameters and ranges from bad to good for RV parameters. However, this agreement was poor for several parameters in patients with ToF (eg RV SV

TABLE 2. Reproducibility Manual Segmentation

	Healthy children				Left-sided CHD				
	Intraobserver variability		Interobserver variability		Intraobserver variability		Interobserver variability		
	Error	ICC	Error	ICC	Error	ICC	Error	ICC	
LV EDV	0.8 ± 9.1	0.98 [0.95–0.99]	7.9 ± 9.5	0.98 [0.95–0.99]	LV EDV	1.7 ± 9.2	0.99 [0.99–1]	1.5 ± 9.2	0.99 [0.99–1]
LV ESV	0.9 ± 5.9	0.93 [0.86–0.97]	8.0 ± 5.0	0.96 [0.91–0.98]	LV ESV	2.4 ± 8.2	0.98 [0.96–0.99]	2.4 ± 7.0	0.99 [0.97–0.99]
LV SV	1.8 ± 10.1	0.94 [0.88–0.97]	0.4 ± 9.1	0.95 [0.90–0.98]	LV SV	4.1 ± 13.9	0.96 [0.92–0.98]	3.9 ± 6.5	0.99 [0.98–1]
LV Mass	2.5 ± 12.9	0.96 [0.91–0.98]	9.0 ± 13.0	0.95 [0.89–0.98]	LV Mass	4.1 ± 12.4	0.99 [0.97–0.99]	34.8 ± 24.5	0.94 [0.88–0.97]
RV EDV	4.0 ± 14.0	0.96 [0.92–0.98]	3.7 ± 16.1	0.95 [0.89–0.97]	RV EDV	2.1 ± 10.5	0.98 [0.97–0.99]	5.3 ± 19.1	0.95 [0.89–0.98]
RV ESV	2.6 ± 7.9	0.94 [0.87–0.97]	4.2 ± 7.2	0.95 [0.89–0.97]	RV ESV	3.0 ± 7.8	0.97 [0.94–0.99]	2.6 ± 11.2	0.95 [0.89–0.97]
RV SV	1.4 ± 11.6	0.93 [0.85–0.97]	1.1 ± 12.3	0.91 [0.82–0.96]	RV SV	0.9 ± 7.6	0.97 [0.93–0.99]	5.2 ± 12.2	0.92 [0.84–0.96]
Fallot									
	Intraobserver variability		Interobserver variability		Intraobserver variability		Interobserver variability		
	Error	ICC	Error	ICC	Error	ICC	Error	ICC	
	LV EDV	1.9 ± 9.4	0.96 [0.91–0.98]	7.9 ± 13.1	0.91 [0.82–0.96]	SiV EDV	0.2 ± 9.1	0.98 [0.95–0.99]	0.3 ± 15.2
LV ESV	2.0 ± 9.8	0.76 [0.54–0.88]	7.1 ± 13.4	0.54 [0.22–0.76]	SiV ESV	6.4 ± 13.0	0.83 [0.67–0.92]	0.4 ± 12.6	0.86 [0.72–0.93]
LV SV	0.1 ± 11.9	0.88 [0.76–0.94]	0.8 ± 16.1	0.75 [0.52–0.87]	SiV SV	6.6 ± 12.3	0.88 [0.76–0.94]	0.7 ± 12.5	0.85 [0.70–0.93]
LV Mass	2.7 ± 8.7	0.94 [0.88–0.97]	12.4 ± 13.7	0.85 [0.71–0.93]	SiV Mass	8.8 ± 19.6	0.75 [0.52–0.87]	14.5 ± 20.9	0.65 [0.38–0.82]
RV EDV	6.6 ± 15.9	0.96 [0.91–0.98]	7.9 ± 14.5	0.97 [0.93–0.99]					
RV ESV	4.1 ± 13.1	0.92 [0.83–0.96]	12.9 ± 21.8	0.76 [0.54–0.88]					
RV SV	2.5 ± 17.4	0.83 [0.67–0.92]	4.9 ± 18.0	0.82 [0.64–0.91]					

CHD = congenital heart defects, ICC = intraclass correlation; LV = left ventricular; RV = right ventricular; EDV = end-diastolic volume; ESV = end-systolic volume; SV = stroke volume; SiV = single ventricular.

Errors are presented in mL. Intraclass correlation coefficients are color coded for clarity: <0.80 is colored red; <−0.90 is colored yellow; >0.90 is colored green. Volumes and masses are derived from trabeculated contours.

TABLE 3. Fully automated vs. manual segmentation

	SuiteHeart															
	Healthy children				Left-sided CHD				Fallot				Univentricular CHD			
	Error	ICC	Error	ICC	Error	ICC	Error	ICC	Error	ICC	Error	ICC	Error	ICC		
LV EDV	3.0 ± 12.2	0.96 [0.91–0.98]	8.0 ± 26.6	0.96 [0.92–0.98]	-1.2 ± 14.6	0.90 [0.79–0.95]	SIV EDV	-8.2 ± 19.3	0.89 [0.77–0.95]							
LV ESV	-2.1 ± 4.0	0.97 [0.93–0.99]	-3.6 ± 15.5	0.95 [0.89–0.98]	-4.9 ± 10.9	0.73 [0.50–0.87]	SIV ESV	-1.4 ± 18.8	0.74 [0.50–0.87]							
LV SV	5.0 ± 10.5	0.93 [0.85–0.97]	11.6 ± 16.8	0.94 [0.88–0.97]	3.7 ± 14.0	0.81 [0.63–0.91]	SIV SV	-6.9 ± 16.0	0.78 [0.58–0.89]							
LV Mass	1.9 ± 12.4	0.96 [0.91–0.98]	23.6 ± 23.3	0.95 [0.90–0.98]	9.5 ± 8.6	0.94 [0.88–0.97]										
RV EDV	3.7 ± 16.5	0.94 [0.87–0.97]	13.6 ± 30.1	0.86 [0.73–0.94]	25.6 ± 27.9	0.87 [0.74–0.94]										
RV ESV	-7.3 ± 7.8	0.93 [0.86–0.97]	2.7 ± 12.2	0.93 [0.85–0.97]	-4.7 ± 15.4	0.88 [0.77–0.95]										
RV SV	11.0 ± 15.7	0.83 [0.67–0.92]	12.4 ± 25.0	0.66 [0.38–0.82]	30.1 ± 25.5	0.62 [0.32–0.80]										
	Medis															
	Healthy children				Left-sided CHD				Fallot				Univentricular CHD			
	Error	ICC	Error	ICC	Error	ICC	Error	ICC	Error	ICC	Error	ICC	Error	ICC		
	LV EDV	1.4 ± 11.5	0.97 [0.94–0.99]	-6.8 ± 15.6	0.99 [0.97–0.99]	-4.7 ± 14.2	0.92 [0.84–0.96]	SIV EDV	-8.4 ± 52.5	0.39 [0.00–0.67]						
LV ESV	0.6 ± 7.7	0.93 [0.85–0.97]	-1.8 ± 10.0	0.98 [0.96–0.99]	-4.6 ± 15.4	0.68 [0.41–0.84]	SIV ESV	-12.9 ± 27.4	0.53 [0.17–0.76]							
LV SV	0.8 ± 8.4	0.96 [0.91–0.98]	-5.0 ± 10.9	0.98 [0.95–0.99]	-0.1 ± 9.2	0.93 [0.85–0.97]	SIV SV	-0.5 ± 31.5	0.37 [0.00–0.66]							
LV Mass	-4.7 ± 14.4	0.93 [0.85–0.97]	-26.2 ± 20.1	0.96 [0.91–0.98]	-14.4 ± 13.8	0.82 [0.65–0.91]										
RV EDV	-3.0 ± 20.8	0.92 [0.83–0.96]	-18.9 ± 25.5	0.92 [0.83–0.96]	-20.5 ± 28.2	0.88 [0.76–0.94]										
RV ESV	8.8 ± 12.8	0.87 [0.74–0.94]	-0.9 ± 17.8	0.90 [0.79–0.95]	-6.1 ± 32.4	0.65 [0.37–0.82]										
RV SV	-8.8 ± 23.3	0.62 [0.32–0.81]	-18.0 ± 16.2	0.85 [0.70–0.93]	-14.4 ± 19.8	0.81 [0.63–0.91]										

CHD = congenital heart defects, ICC = intraclass correlation; LV = left ventricular; RV = right ventricular; EDV = end-diastolic volume; ESV = end-systolic volume; SV = stroke volume; SIV = single ventricular.
 Intraclass correlation coefficients are color coded for clarity: <0.80 is colored red; >0.90 is colored yellow; >0.90 is colored green. Manual contours are trabeculated for SuiteHeart and smoothed for Medis.

TABLE 4. Reproducibility automated segmentation followed by manual adjustment

		SuiteHeart									
(A)	Healthy children			Left-sided CHD							
	Intraobserver variability		Interobserver variability		Intraobserver variability		Interobserver variability				
	Error	ICC	Error	ICC	Error	ICC	Error	ICC			
LV EDV	0.6 ± 2.2	1 [1-1]	0.9 ± 2.6	1 [1-1]	LV EDV	1.5 ± 7.7	1 [0.99-1]	4.4 ± 12.3	0.99 [0.98-1]		
LV ESV	0.6 ± 2.1	0.99 [0.98-1]	0.6 ± 2.2	0.99 [0.98-0.99]	LV ESV	0.1 ± 6.2	0.99 [0.99-1]	6.6 ± 15.3	0.95 [0.90-0.98]		
LV SV	1.0 ± 2.8	0.99 [0.99-1]	1.3 ± 3.3	0.99 [0.98-1]	LV SV	1.7 ± 9.3	0.98 [0.96-0.99]	2.3 ± 12.3	0.97 [0.93-0.99]		
LV Mass	0.3 ± 4.1	1 [0.99-1]	0.8 ± 2.3	1 [1-1]	LV Mass	1.6 ± 3.8	1 [1-1]	2.7 ± 11.3	0.99 [0.97-0.99]		
RV EDV	2.9 ± 5.9	0.99 [0.98-1]	4.9 ± 6.7	0.99 [0.98-1]	RV EDV	1.4 ± 7.4	0.99 [0.98-1]	8.3 ± 16.0	0.97 [0.93-0.98]		
RV ESV	1.9 ± 4.9	0.98 [0.95-0.99]	0.7 ± 3.3	0.99 [0.97-0.99]	RV ESV	0.7 ± 5.8	0.99 [0.98-0.99]	2.4 ± 7.4	0.98 [0.96-0.99]		
RV SV	1.0 ± 5.0	0.98 [0.96-0.99]	4.2 ± 5.6	0.97 [0.95-0.99]	RV SV	0.8 ± 7.2	0.96 [0.92-0.98]	5.8 ± 14.0	0.87 [0.74-0.94]		
Univentricular CHD											
		Fallot		Interobserver variability		Intraobserver variability		Interobserver variability		ICC	
		Error	ICC	Error	ICC	Error	ICC	Error	ICC	Error	ICC
LV EDV	1.7 ± 4.2	0.99 [0.98-1]	0.7 ± 4.5	0.99 [0.98-1]	SV EDV	2.8 ± 12.4	0.94 [0.88-0.97]	1.9 ± 7.2	0.98 [0.95-0.99]		
LV ESV	1.1 ± 5.9	0.93 [0.86-0.97]	1.1 ± 6.0	0.93 [0.86-0.97]	SV ESV	4.6 ± 10.5	0.91 [0.81-0.96]	0.5 ± 7.1	0.95 [0.90-0.98]		
LV SV	3.7 ± 7.2	0.94 [0.87-0.97]	1.9 ± 5.1	0.97 [0.93-0.98]	SV SV	1.8 ± 6.6	0.92 [0.84-0.96]	1.4 ± 8.0	0.89 [0.77-0.95]		
LV Mass	2.1 ± 5.8	0.97 [0.95-0.99]	2.0 ± 5.3	0.98 [0.95-0.99]	SV Mass	2.8 ± 8.0	0.93 [0.86-0.97]	9.3 ± 11.1	0.90 [0.79-0.95]		
RV EDV	2.0 ± 6.6	0.99 [0.98-1]	7.2 ± 14.9	0.96 [0.92-0.98]							
RV ESV	1.0 ± 4.0	0.99 [0.98-1]	3.7 ± 10.5	0.95 [0.90-0.98]							
RV SV	3.3 ± 6.1	0.97 [0.93-0.99]	1.5 ± 14.2	0.85 [0.71-0.93]							
Medis											
		Healthy children			Left-sided CHD						
		Intraobserver variability		Interobserver variability		Intraobserver variability		Interobserver variability			
		Error	ICC	Error	ICC	Error	ICC	Error	ICC		
LV EDV	0.4 ± 2.2	1 [1-1]	0.2 ± 2.8	1 [1-1]	SV EDV	2.7 ± 7.0	1 [0.99-1]	3.2 ± 8.7	1 [0.99-1]		
LV ESV	1.9 ± 4.3	0.98 [0.96-0.99]	0.5 ± 7.0	0.94 [0.88-0.97]	SV ESV	0.5 ± 9.0	0.99 [0.97-0.99]	5.4 ± 11.5	0.98 [0.95-0.99]		
LV SV	2.3 ± 4.6	0.99 [0.97-0.99]	0.7 ± 5.8	0.98 [0.96-0.99]	SV SV	1.8 ± 12.0	0.97 [0.94-0.99]	0.5 ± 13.9	0.96 [0.92-0.98]		
LV Mass	0.6 ± 2.4	1 [1-1]	2.5 ± 3.3	1 [0.99-1]	SV Mass	2.7 ± 6.3	1 [0.99-1]	6.6 ± 8.7	0.99 [0.98-1]		
RV EDV	4.3 ± 8.6	0.99 [0.97-0.99]	4.1 ± 8.7	0.99 [0.97-0.99]		0.6 ± 7.4	0.99 [0.99-1]	3.3 ± 18.1	0.96 [0.92-0.98]		
RV ESV	4.5 ± 7.2	0.96 [0.92-0.98]	3.1 ± 11.9	0.90 [0.79-0.95]		2.2 ± 10.0	0.97 [0.94-0.99]	7.5 ± 15.8	0.93 [0.85-0.97]		
RV SV	0.2 ± 5.1	0.98 [0.97-0.99]	0.9 ± 11.4	0.94 [0.87-0.97]		2.8 ± 7.2	0.96 [0.92-0.98]	10.6 ± 14.5	0.88 [0.76-0.94]		

TABLE 4. Continued

	Fallot				Univentricular CHD				
	Intraobserver variability		Interobserver variability		Intraobserver variability		Interobserver variability		
	Error	ICC	Error	ICC	Error	ICC	Error	ICC	
LV EDV	0.4 ± 8.2	0.97 [0.94–0.99]	0.4 ± 9.7	0.96 [0.92–0.98]	SiV EDV	2.0 ± 7.3	0.98 [0.96–0.99]	4.9 ± 17.8	0.88 [0.75–0.94]
LV ESV	4.0 ± 10.1	0.85 [0.71–0.93]	2.0 ± 10.5	0.83 [0.67–0.92]	SiV ESV	1.1 ± 6.9	0.97 [0.94–0.99]	3.8 ± 10.4	0.93 [0.86–0.97]
LV SV	3.6 ± 10.7	0.89 [0.78–0.95]	1.6 ± 6.7	0.96 [0.91–0.98]	SiV SV	0.9 ± 9.0	0.84 [0.68–0.92]	0.5 ± 13.8	0.56 [0.24–0.77]
LV Mass	1.9 ± 7.2	0.93 [0.86–0.97]	2.8 ± 7.2	0.93 [0.86–0.97]	SiV Mass	2.3 ± 7.7	0.95 [0.89–0.98]	12.2 ± 13.9	0.84 [0.66–0.92]
RV EDV	0.0 ± 11.1	0.98 [0.96–0.99]	1.1 ± 14.0	0.97 [0.93–0.99]					
RV ESV	1.8 ± 13.3	0.94 [0.87–0.97]	3.2 ± 11.1	0.96 [0.91–0.98]					
RV SV	1.7 ± 11.0	0.91 [0.82–0.96]	5.3 ± 16.9	0.79 [0.60–0.90]					

CHD = congenital heart defects, ICC = intraclass correlation; LV = left ventricular; RV = right ventricular; EDV = end-diastolic volume; ESV = end-systolic volume; SV = stroke volume; SiV = single ventricular.
 Intraclass correlation coefficients are color coded for clarity: <0.80 is colored red; <0.90 is colored yellow; >0.90 is colored green. Contours are trabeculated for SuiteHeart and smoothed for Medis.

difference 30.1 ± 25.5 mL, ICC: 0.62 [0.32–0.80] for SuiteHeart), and for most parameters in patients with univentricular CHD (eg single ventricular SV difference -6.9 ± 16.0 , ICC: 0.78 [0.56–0.89] for SuiteHeart).

Fully automated segmentation had agreement with great vessel flow-derived LV SV of a magnitude similar to that of manual segmentation across study groups (online supplement S1). For the RV and single ventricle, fully automated segmentation had worse agreement with great vessel flow-derived SV than manual segmentation (eg single ventricular SV in Medis: difference -2.4 ± 27.0 mL, ICC: 0.38 [0.0–0.69] vs. single ventricular SV manual: difference -6.3 ± 18.2 mL, ICC: 0.68 [0.39–0.84]).

Fully automated segmentation has, by definition, perfect intraobserver and interobserver reproducibility for all parameters (difference 0 ± 0 mL, ICC: 1).

Automated Segmentation Followed by Manual Adjustment

Large errors in automated segmentation may be easily recognized and corrected by experienced observers. Automated segmentation followed by manual adjustment resulted in moderate-to-good agreement with manual segmentation for healthy children and LV-CHD (ICC: 0.88–0.99) and mostly moderate-to-good agreement for patients with ToF and univentricular CHD (ICC: 0.70–0.97, presented in the online supplement S1). Note that the agreement between automated segmentation followed by manual adjustment and manual segmentation is of a similar magnitude as the interobserver agreement of manual segmentation.

Agreement between great vessel flow-derived LV and RV SV and automated segmentation followed by manual adjusted-derived LV and RV SV was similar to—or, for some analyses, more favorable than—that of manual segmentation for healthy children, patients with LV-CHD, and patients with ToF (presented in the online supplement S1: ICC: 0.74–0.95 for automated segmentation followed by manual adjustment vs. 0.72–0.97 for manual segmentation). For patients with univentricular CHD agreement between great vessel flow-derived single and automated segmentation followed by manual adjustment derived single ventricular SV was poor for both Medis and SuiteHeart (ICC: 0.33 [0–0.64] for Medis vs. 0.62 [0.31–0.81] for SuiteHeart).

Automated segmentation followed by manual adjustment resulted in excellent reproducibility for both software packages and exceeded that of manual segmentation, even for univentricular CHD (see Tables 2 and 4: segmentation was considered good for 35 of the 50 parameters for manual segmentation vs. 47 of the 50 (SuiteHeart) and 41 of the 50 (Medis) for automated segmentation followed by manual adjustment). For healthy children, ICC exceeded 0.96 for intraobserver variability for all parameters and exceeded 0.90 for interobserver variability. For CHD, intraobserver variability

was good for most parameters for Medis, and for all parameters for SuiteHeart (Table 4). Interobserver variability ranged from generally good for LV-CHD (ICC: 0.88–1) to poor to good for univentricular CHD (ICC: 0.56–0.93) in Medis. For SuiteHeart, interobserver variability was moderate to good for all patient groups (Table 4).

Postprocessing Time of Segmentation Strategies

An overview of postprocessing times is presented in the online supplement S1. Manual segmentation times were 24 (IQR 16–28) minutes for healthy children, 23 (IQR 23–24) minutes for LV-CHD, 19 (IQR 17–23) minutes for ToF, and 13 (IQR 12–14) minutes for univentricular CHD. Automated segmentation followed by manual adjustment postprocessing times did not differ significantly between Medis and SuiteHeart ($P = 0.429$). Postprocessing times for automated segmentation followed by manual adjustment were importantly lower than those for manual segmentation (5 [IQR 4–6] minutes for healthy children, 6 [IQR 4–8] minutes for LV-CHD, 5 [IQR 4–6] minutes for ToF, and 7 [IQR 6–9] minutes for univentricular CHD) (combined results for Medis and SuiteHeart).

Discussion

Using current commercially available software suites, we found fully automated segmentation of the ventricles in pediatric subjects was good for the left ventricle in healthy children and patients with LV-CHD. However, errors were larger for the right ventricle in these patients, as well as for patients with ToF and univentricular CHD (for the left, right, and single ventricle). These errors were generally easily recognized and may be adjusted by human observers. Automated segmentation followed by manual adjustment: 1) agreed well with manual segmentation, 2) improved reproducibility compared to manual segmentation, and 3) reduced segmentation time.

Several differences between adults and children warrant researching the efficacy of postprocessing methods in children. MRI studies are often obtained during breath-holds to minimize motion and increase image quality, but young children are unable to hold their breath and are often imaged while free-breathing which may affect image quality. CHD is an important indication for MRI in pediatrics and as such, structural defects are more common than in adults.² Differences in body proportions may also confound algorithms as landmarks may not be recognized.³⁰ Lastly, children have smaller hearts, higher heart rates, a lower mass-to-volume ratio and slightly more spherical left ventricles than adults.³¹ These factors could affect the performance of postprocessing algorithms. It is important to consider that most algorithms have been developed for, trained on, and validated in populations of adults with structurally normal hearts.^{11,12} The performance of these algorithms in children requires

further study. Hammon et al evaluated a commercially available fully automated segmentation algorithm for the LV in 45 children and found the algorithms failed to detect the LV in the three patients under 6 years old.³⁰ Similarly, in our present study, failure to automatically segment studies in Medis was more common for younger patients and studies obtained while breathing freely. It should be noted that studies obtained while free breathing were performed exclusively in younger patients. Whether free breathing or younger age leads to failure to automatically segment studies remains uncertain. Fully automated segmentation may result in large errors at any slice, but it is generally more common in the basal or apical slice (data not shown). The current study demonstrates that fully automated segmentation may be feasible only in limited cases (the LV for healthy children and LV-CHD).

Automated segmentation for CHD may be difficult to achieve reliably as training of automated segmentation algorithms generally requires large training datasets.¹³ Backhaus et al. evaluated the performance of the SuiteHeart automated segmentation algorithm in adults undergoing clinically indicated MRI (including patients with ToF and aortic valve replacements) and found reduced agreement with manual segmentation for LV measurements in subjects with ToF and in studies with reduced image quality.⁹ This is in agreement with the results of the current study. We demonstrated that manual adjustments to automated segmentation provide high agreement with conventional manual segmentation methods, even for complex CHD. Agreement with great vessel flow generally increased with manual adjustments to automated segmentation, compared to fully automated segmentation. However, this was not true for univentricular CHD. This may relate to inaccuracy of great vessel flow-derived stroke volume in this population, or reflect variability in segmenting these studies (despite the use of automated segmentation). Furthermore, we found automated segmentation followed by manual adjustment reduces segmentation time compared to manual segmentation and improves reproducibility. The improved reproducibility of automated segmentation followed by manual adjustment compared to fully manual segmentation may relate to large segments of the automated segmentation being satisfactory and remaining unaltered by both observers.

Several factors in our study complicate direct comparisons between software suites. The automated segmentation in Medis failed in six patients, compared to none in SuiteHeart. These cases were excluded from analysis for Medis. Automated segmentation may be more likely to fail in complicated cases (i.e. smaller children, reduced image quality, more complex cardiac defects).¹³ The automated segmentation algorithm of Medis provided smoothed contours only and does not provide an option to trace trabeculae of the endocardial border. Smoothing endocardial contours results in higher

volumes and lower wall mass and may improve interobserver variability.^{32,33} To account for these differences, smoothed manual contours were used for all comparisons with the Medis automated segmentation algorithm (both fully automated and adjusted by human input). Overall, the post-processing time required for automated segmentation followed by manual adjustment was similar between the software suites. It should be noted that the manual and automated followed by manual adjustment segmentation times for univentricular CHD were lower than for other groups, as only a single ventricle needs to be segmented. The overall results for automated segmentation followed by manual adjustment can—subjectively—be considered excellent for both software suites.

Limitations

The age distribution of included subjects differs across diagnosis groups, which probably introduced some bias. This was primarily due to limited availability of studies across all age ranges for each diagnosis group. Studies from patients under 8 years old were scarce for complex CHD and Fallot patient groups, as these studies are rarely clinically indicated. Echocardiography is generally a suitable imaging modality for these indications at this age range. To assure we included a broad range of ages for each diagnosis available studies were stratified by age quartiles and randomly selected from these quartiles. We included 20 subjects per group, which was an adequate sample size according to our power calculation. However, this sample size did not allow for reliable subset analyses. Due to the nature of this study, not true reference standard reference was available. We compared volumetrically derived stroke volumes to those derived from great vessel flow. However, these measurements are also subject to variability. We did not time postprocessing times for automated postprocessing as this is heavily influenced by computer hardware. Furthermore, it does not require user attention as both software suites allow for queuing of the preprocessing of multiple studies as a background process. Differences between segmentations were reported only as differences in resulting volume measurements, rather than more descriptive differences of 2D contours, such as dice indices or Hausdorff distances.^{34,35} We expressed error size in volumes and mass as these are more readily applicable for clinicians.^{4,5,36} Variability (across observers and segmentation methods) is usually expressed by systematic and random error. We used ICC values for general statements regarding the agreement between measurements as this parameter accounts for both systematic and random error in a single dimension. As ICC values are influenced by the total variability within the dataset (which may vary across our study groups), ICCs cannot be compared directly across study groups. Absolute differences between measurements should always be considered.

Conclusions

Fully automated segmentation provided good results only for the left ventricle in healthy children or those with LV-CHD. Compared to manual segmentation, automated segmentation followed by manual adjustment provided similar volumes and mass, improved reproducibility, and importantly reduced segmentation time, even for patients with complex CHD.

Acknowledgments

The authors acknowledge Joep Schoenmaeckers, BSc. for his technical support in realizing this study.

REFERENCES

1. Leiner T, Bogaert J, Friedrich MG, et al. SCMR Position Paper (2020) on clinical indications for cardiovascular magnetic resonance. *J Cardiovasc Magn Reson*. 2020;22(1):76.
2. Valsangiacomo Buechel ER, Grosse-Wortmann L, Fratz S, et al. Indications for cardiovascular magnetic resonance in children with congenital and acquired heart disease: An expert consensus paper of the Imaging Working Group of the AEPIC and the Cardiovascular Magnetic Resonance Section of the EACVI. *Eur Heart J Cardiovasc Imaging*. 2015; 16(3):281-297.
3. Hedstrom E, Ishida M, Sepulveda-Martinez A, et al. 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(1):33.
4. Beerbaum P, Barth P, Kropf S, et al. Cardiac function by MRI in congenital heart disease: Impact of consensus training on interinstitutional variance. *J Magn Reson Imaging*. 2009;30(5):956-966.
5. Luijnenburg SE, Robbers-Visser D, Moelker A, Vliegen HW, Mulder BJ, Helbing WA. Intra-observer and interobserver variability of biventricular function, volumes and mass in patients with congenital heart disease measured by CMR imaging. *Int J Cardiovasc Imaging*. 2010;26(1): 57-64.
6. Shu L-Q, Sun Y-K, Tan L-H, Shu Q, Chang AC. Application of artificial intelligence in pediatrics: Past, present and future. *World J Pediatr* 2019;15(2):105-108.
7. Queirós S, Barbosa D, Engvall J, et al. Multi-centre validation of an automatic algorithm for fast 4D myocardial segmentation in cine CMR datasets. *Eur Heart J* 2016;17(10):1118-1127.
8. Hautvast GL, Salton CJ, Chuang ML, Breeuwer M, O'Donnell CJ, Manning WJ. Accurate computer-aided quantification of left ventricular parameters: Experience in 1555 cardiac magnetic resonance studies from the Framingham Heart Study. *Magn Reson Med*. 2012;67(5):1478-1486.
9. Backhaus SJ, Staab W, Steinmetz M, et al. Fully automated quantification of biventricular volumes and function in cardiovascular magnetic resonance: Applicability to clinical routine settings. *J Cardiovasc Magn Reson*. 2019;21(1):24.
10. Petersen SE, Aung N, Sanghvi MM, 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(1):18.
11. Tan LK, Liew YM, Lim E, McLaughlin RA. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. *Med Image Anal* 2017;39:78-86.
12. Bai W, Sinclair M, Tarroni G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson* 2018;20(1):65.

13. Arafati A, Hu P, Finn JP, et al. Artificial intelligence in pediatric and adult congenital cardiac MRI: An unmet clinical need. *Cardiovasc Diagn Ther* 2019;9(S2):S310-S325.
14. Dewey FE, Rosenthal D, Murphy DJ Jr, Froelicher VF, Ashley EA. Does size matter? Clinical applications of scaling cardiac size and function for body size. *Circulation*. 2008;117(17):2279-2287.
15. van der Ven JPG, Sadighy Z, Valsangiacomo Buechel ER, et al. Multi-centre reference values for cardiac magnetic resonance imaging derived ventricular size and function for children aged 0–18 years. *Eur Heart J Cardiovasc Imaging*. 2020;21(1):102-113.
16. van den Berg J, Hop WC, Strengers JL, et al. Clinical condition at mid-to-late follow-up after transatrial-transpulmonary repair of tetralogy of Fallot. *J Thorac Cardiovasc Surg*. 2007;133(2):470-477.
17. Luijnenburg SE, Helbing WA, Moelker A, et al. 5-year serial follow-up of clinical condition and ventricular function in patients after repair of tetralogy of Fallot. *Int J Cardiol*. 2013;169(6):439-444.
18. van den Bosch E, Cuypers J, Luijnenburg SE, et al. Ventricular response to dobutamine stress cardiac magnetic resonance imaging is associated with adverse outcome during 8-year follow-up in patients with repaired Tetralogy of Fallot. *Eur Heart J Cardiovasc Imaging*. 2020;21(9):1039-1046.
19. van den Bosch E, Bossers SSM, Kamphuis VP, et al. Associations between blood biomarkers, cardiac function, and adverse outcome in a Young Fontan cohort. *J Am Heart Assoc*. 2021;10(5):e015022.
20. Bossers SS, Kapusta L, Kuipers IM, et al. Ventricular function and cardiac reserve in contemporary Fontan patients. *Int J Cardiol*. 2015;196:73-80.
21. Robbers-Visser D, Boersma E, Helbing WA. Normal biventricular function, volumes, and mass in children aged 8 to 17 years. *J Magn Reson Imaging*. 2009;29(3):552-559.
22. Sarikouch S, Peters B, Gutberlet M, et al. Sex-specific pediatric percentiles for ventricular size and mass as reference values for cardiac MRI: Assessment by steady-state free-precession and phase-contrast MRI flow. *Circ Cardiovasc Imaging*. 2010;3(1):65-76.
23. Buechel EV, Kaiser T, Jackson C, Schmitz A, Kellenberger CJ. Normal right- and left ventricular volumes and myocardial mass in children measured by steady state free precession cardiovascular magnetic resonance. *J Cardiovasc Magn Reson*. 2009;11:19.
24. Robbers-Visser D, Jan Ten Harkel D, Kapusta L, et al. Usefulness of cardiac magnetic resonance imaging combined with low-dose dobutamine stress to detect an abnormal ventricular stress response in children and young adults after Fontan operation at young age. *Am J Cardiol*. 2008;101(11):1657-1662.
25. Schulz-Menger J, Bluemke DA, Bremerich J, et al. Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic Resonance (SCMR) board of trustees task force on standardized post processing. *J Cardiovasc Magn Reson*. 2013;15:35.
26. Van Der Ven JPG, Sadighy Z, Valsangiacomo Buechel ER, et al. Multi-centre reference values for cardiac magnetic resonance imaging derived ventricular size and function for children aged 0–18 years. *Eur Heart J* 2020;21(1):102-113.
27. van der Ven JPG, Alsaied T, Juggan S, et al. Atrial function in Fontan patients assessed by CMR: Relation with exercise capacity and long-term outcomes. *Int J Cardiol*. 2020;312:56-61.
28. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.
29. Liljequist D, Elfving B, Skavberg RK. Intraclass correlation – A discussion and demonstration of basic features. *PLOS one*. 2019;14(7):e0219854.
30. Hammon M, Janka R, Dankerl P, et al. Pediatric cardiac MRI: Automated left-ventricular volumes and function analysis and effects of manual adjustments. *Pediatr Radiol*. 2015;45(5):651-657.
31. Kaku K, Takeuchi M, Otani K, et al. Age- and gender-dependency of left ventricular geometry assessed with real-time three-dimensional transthoracic echocardiography. *J Am Soc Echocardiogr*. 2011;24(5):541-547.
32. Winter MM, Bernink FJ, Groenink M, et al. Evaluating the systemic right ventricle by CMR: The importance of consistent and reproducible delineation of the cavity. *J Cardiovasc Magn Reson*. 2008;10:40.
33. Kawel-Boehm N, Maceira A, Valsangiacomo-Buechel ER, et al. Normal values for cardiovascular magnetic resonance in adults and children. *J Cardiovasc Magn Reson*. 2015;17:29.
34. Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1945;26(3):297-302.
35. Rockafellar RT, Wets RJ-B. *Variational analysis*. Berlin Heidelberg: Springer-Verlag; 1998.
36. Gnanappa GK, Rashid I, Celermajer D, Ayer J, Puranik R. Reproducibility of cardiac magnetic resonance imaging (CMRI)-derived right ventricular parameters in repaired Tetralogy of Fallot (ToF). *Heart, Lung Circul* 2018;27(3):381-385.