# Traveling Volunteers: A Multi-Vendor, Multi-Center Study on Reproducibility and Comparability of 4D Flow Derived Aortic Hemodynamics in Cardiovascular Magnetic Resonance

Aylin Demir, MD,[1] Stephanie Wiesemann, MD,[1,2] Jennifer Erley, MD,[3]

Sebastian Schmitter, PhD,[4] Ralf Felix Trauzeddel, MD,[1,2,5] Burkert Pieske, MD,[2,3,6]

Jochen Hansmann, MD,[7] Sebastian Kelle, MD,[2,3,6] and Jeanette Schulz-Menger, MD[1,2]* 🆔

**Background:** Implementation of four-dimensional flow magnetic resonance (4D Flow MR) in clinical routine requires awareness of confounders.
**Purpose:** To investigate inter-vendor comparability of 4D Flow MR derived aortic hemodynamic parameters, assess scan-rescan repeatability, and intra- and interobserver reproducibility.
**Study Type:** Prospective multicenter study.
**Population:** Fifteen healthy volunteers (age $24.5 \pm 5.3$ years, 8 females).
**Field Strength/Sequence:** 3 T, vendor-provided and clinically used 4D Flow MR sequences of each site.
**Assessment:** Forward flow volume, peak velocity, average, and maximum wall shear stress (WSS) were assessed via nine planes (P1–P9) throughout the thoracic aorta by a single observer (AD, 2 years of experience). Inter-vendor comparability as well as scan-rescan, intra- and interobserver reproducibility were examined.
**Statistical Tests:** Equivalence was tested setting the 95% confidence interval of intraobserver and scan-rescan difference as the limit of clinical acceptable disagreement. Intraclass correlation coefficient (ICC) and Bland–Altman plots were used for scan-rescan reproducibility and intra- and interobserver agreement. A $P$-value $<0.05$ was considered statistically significant. ICCs $\geq 0.75$ indicated strong correlation (>0.9: excellent, 0.75–0.9: good).
**Results:** Ten volunteers finished the complete study successfully. 4D flow derived hemodynamic parameters between scanners of three different vendors are not equivalent exceeding the equivalence range. P3–P9 differed significantly between all three scanners for forward flow ($59.1 \pm 13.1$ mL vs. $68.1 \pm 12.0$ mL vs. $55.4 \pm 13.1$ mL), maximum WSS ($1842.0 \pm 190.5$ mPa vs. $1969.5 \pm 398.7$ mPa vs. $1500.6 \pm 247.2$ mPa), average WSS ($1400.0 \pm 149.3$ mPa vs. $1322.6 \pm 211.8$ mPa vs. $1142.0 \pm 198.5$ mPa), and peak velocity between scanners I vs. III ($114.7 \pm 12.6$ cm/s vs. $101.3 \pm 15.6$ cm/s). Overall, the plane location at the sinotubular junction (P1) presented most inter-vendor stability (forward: $78.5 \pm 15.1$ mL vs. $80.3 \pm 15.4$ mL vs. $79.5 \pm 19.9$ mL [$P = 0.368$]; peak: $126.4 \pm 16.7$ cm/s vs. $119.7 \pm 13.6$ cm/s vs. $111.2 \pm 22.6$ cm/s [$P = 0.097$]). Scan-rescan reproducibility and intra- and interobserver variability were good to excellent (ICC $\geq 0.8$) with best agreement for forward flow (ICC $\geq 0.98$).

**Data Conclusion:** The clinical protocol used at three different sites led to differences in hemodynamic parameters assessed by 4D flow.
**Level of Evidence:** 2
**Technical Efficacy Stage:** 2

Three-dimensional (3D) cine (time-resolved) phase-contrast magnetic resonance with three-directional velocity-encoding (4D Flow MR) allows quantification and visualization of hemodynamics in the heart and great vessels.[1–3] It has been shown to be useful in different diseases for the evaluation of blood flow patterns and derived hemodynamic parameters, such as wall shear stress (WSS).[4] WSS is elevated in patients with severe aortic stenosis and/or bicuspid aortic valve, as well as in aortic valve replacements.[5–8] In systemic diseases with affection of the aorta, WSS is also altered. In patients with Marfan syndrome, these alterations included a local decrease in average WSS in the outer proximal ascending segment and an increase in the inner distal ascending aorta.[9] This additional information may help with clinical decision-making and therapy guiding in diseases.

The potential of integrating 4D Flow MR into daily clinical routine has been increased due to several image acceleration techniques, such as parallel imaging, yielding scan times for the assessment of aortic flow of under 2 minutes.[10] However, awareness of confounders is now of vital importance for successful implementation in the clinical setting.

4D Flow MR acquisitions have been applied using various protocols and different scanner types.[1–11] Although the acquisition of the whole heart is routinely recommended in congenital heart diseases, while the acquisition of the aorta is mainly used for aortic and aortic valve diseases, no consensus has been established on which technique to use in adults.[12,13] Additionally, 4D Flow MR sequences differ between sites with scanners from different vendors regarding their acquisition parameters, as each vendor has made different adjustments in the sequence and in the recommended protocol for optimal acquisition.

For other quantitative parameters like $T_1$ mapping or myocardial strain, variability of results in dependence on different manufacturers is known, thus indicating that variability may also exist in 4D Flow MR.[14,15] Without this information on inter-vendor agreement, comparison of 4D Flow MR examinations provided by scanners from different vendors may be difficult in the clinical environment.

Therefore, the aims of this study were to:

1. Examine the inter-vendor comparability and reproducibility of 4D Flow MR derived aortic hemodynamic parameters at scanners of three different vendors in order to evaluate if 4D Flow MR examinations acquired at different sites under diverging conditions are equivalent.

2. Test for scan-rescan reproducibility at one scanner for hemodynamic variability.

3. Determine the intra- and interobserver agreement for evaluation of hemodynamic parameters.

## Materials and Methods

### Study Population

Ethical approval was obtained by the local ethics committee (approval number EA2/208/17). The study complied with the Declaration of Helsinki. It was registered at the German Clinical Trials Register (registration number: 00013253) and the World Health Organization (universal trial number: U1111-1207-5874). Informed written consent was obtained from each participant prior to study enrollment. Healthy volunteers without a history of cardiovascular diseases were recruited and underwent three cardiovascular MR scans using different scanner types. Cardiovascular MR examinations verified normal left ventricular function and a tricuspid aortic valve without pathology.

### Image Acquisition

Volunteers were scanned at three different sites (scanners I–III). Each site was using a 3-T wide-bore scanner of different vendors (scanners sorted alphabetically not corresponding to sites): Ingenia (Philips, Best, The Netherlands), SIGNA Architect (GE Healthcare, Milwaukee, WI, USA), and MAGNETOM Verio (Siemens Healthineers, Erlangen, Germany). The same scan protocol was performed in all volunteers. Heart rate was recorded during each 4D Flow MR acquisition, while blood pressure was measured before each 4D Flow MR scan. At one site (scanner III) an additional 4D Flow MR acquisition for scan-rescan reproducibility was performed after a short break of 20–30 minutes, where the volunteers left the scanner requiring repositioning of the coil and replanning of image acquisition afterwards.

To validate the current level of agreement in 4D flow imaging of the aorta in a real-life setting, the clinical 4D Flow MR sequence of each site was used to image the volunteers. The study refrained of adapting sequence parameters to the published consensus paper as common ground to avoid potential bias favoring one vendor and to reflect the current clinical application of 4D Flow MR sequences.[2] Acquisition parameters between the scanners varied due to different vendor-provided protocols and each site's individual adjustments for optimal clinical use (Table 1).

The acquisition volume was defined according to the vendor's recommendation: on scanner III the scan covered the aorta, on scanners I and II the whole heart. For cardiac gating, an electrocardiogram (ECG) was used in all cases. On scanner III, prospective cardiac gating was applied along with a cross-paired respiratory navigator placed on the lung–liver interface allowing for free breathing. Retrospective triggered ECG gating without a respiratory navigator

**TABLE 1. Sequence Parameters of the Different Scanners**

| | Scanner I | Scanner II | Scanner III |
|---|---|---|---|
| Acquisition volume | Sagittal<br>Whole heart | Transversal<br>Whole heart | Sagittal<br>Full coverage aorta |
| ECG gating | Retrospective | Retrospective | Prospective |
| Respiratory navigator | No | No | Yes |
| Echo time (msec) | 2.2 | 2.0 | 2.6 |
| Repetition time (msec) | 3.5 | 4.2 | 5.1 |
| Temporal resolution (msec) | 28 | 66.8 | 40.8 |
| Number of cardiac phases | 25 ($\pm$0) | 25 ($\pm$0) | 19.2 ($\pm$3.4) |
| Temporal segmentation factor | 2 | 4 | 2 |
| Acquisition voxel size (mm$^3$) | 2.8 $\times$ 2.8 $\times$ 2.8 | 2.4 $\times$ 2.4 $\times$ 2.8 | 2.7 $\times$ 2.3 $\times$ 2.6 |
| Reconstructed voxel size (mm$^3$) | 1.3–1.9 $\times$ 1.3–1.9 $\times$ 2.8 | 1.4–1.5 $\times$ 1.4–1.5 $\times$ 1.4 | 2.3 $\times$ 2.3 $\times$ 2.6 |
| Field of view (mm$^3$) | 270 $\times$ 180 $\times$ 81.2 | 380 $\times$ 266 $\times$ 95.2 | 360 $\times$ 270 $\times$ 83.2 |
| Velocity encoding (cm/s) | 250 | 150 | 150 |
| Flip angle (degrees) | 5 | 8 | 7 |
| Radiofrequency coil | Multi-element receive coil array with flexible number of up to 32 elements | Multi-element receive coil array with flexible number of up to 32 elements | 32-channel body coil |
| Parallel imaging[a] | $R = 2$ | $R = 8$<br>kt-acceleration | $R = 5$<br>kt-acceleration |
| Encoding scheme | 4-point symmetric with Hadamard encoding | 4-point symmetric with Hadamard encoding | 4-point symmetric with MPS |
| Sequence | Product sequence | Product sequence | Prototype sequence |
| Acquisition duration (minutes) | 10.8 ($\pm$1.2) | 10.5 ($\pm$1.5) | 8.0 ($\pm$2.2) |

Number of cardiac phases and acquisition duration are presented as mean $\pm$ SD.
ECG = electrocardiogram; $R$ = acceleration factor; kt = k-adaptive-t; MPS = measurement/phase/slice.
[a]Used vendor-specific techniques in alphabetical order: Autocalibrating Reconstruction for Cartesian sampling (ARC), GeneRalized Autocalibrating Partially Parallel Acquisition (GRAPPA), and Sensitivity Encoding (SENSE).

was used at scanners I and II. Based on site routine, the velocity encoding (VENC) value was set to 150 cm/s at scanners II and III, and to 250 cm/s at scanner I. Acquisition time differed between the volunteers, depending on heart rate and breathing quality.

## Image Analysis

All 4D Flow MR images were analyzed by a single observer (AD, 2 years of 4D Flow MR experience) using CAAS MR Solutions 5.0 (Pie Medical Imaging BV, Maastricht, The Netherlands).[16] Maxwell fields were corrected in all MR systems automatically during image reconstruction.[17] Background phase offset correction using a linear fit and phase unwrapping were performed in all datasets.[2,18] In the applied software version, phase unwrapping was not possible for the data acquired at scanner II. Therefore, all images (including those from scanners I and III) were manually screened by the observer (AD) for potential phase aliasing, and in cases where phase aliasing occurred inside a plane region, the plane was excluded.

For segmentation of the aorta, the start point was placed in the upper region of the left ventricular outflow tract and the end point in the descending aorta below the cardiac apex (Fig. 1b). As previously described, a phase-specific segmented 3D aortic surface model was automatically detected for five cardiac phases and the peak systolic phase was selected for implementation of the succeeding analysis steps.[19,20] If necessary, this 3D volume model was manually corrected by the observer by editing the vessel boundary contours for each aortic 2D plane in the peak systolic phase (Fig. 1c) updating the centerline subsequently. In case of missing or truncated aortic segments, the scan was excluded from further evaluation.

Nine cross-sectional planes were positioned along the centerline (Figs. 1 and 2). Contours of the vessels were projected automatically on each plane based on the 3D segmentation of the selected peak systolic phase and, if needed, manually adapted for all phases along the cardiac cycle. Peak velocity and forward flow volume were evaluated. For WSS analysis the identical plane locations were
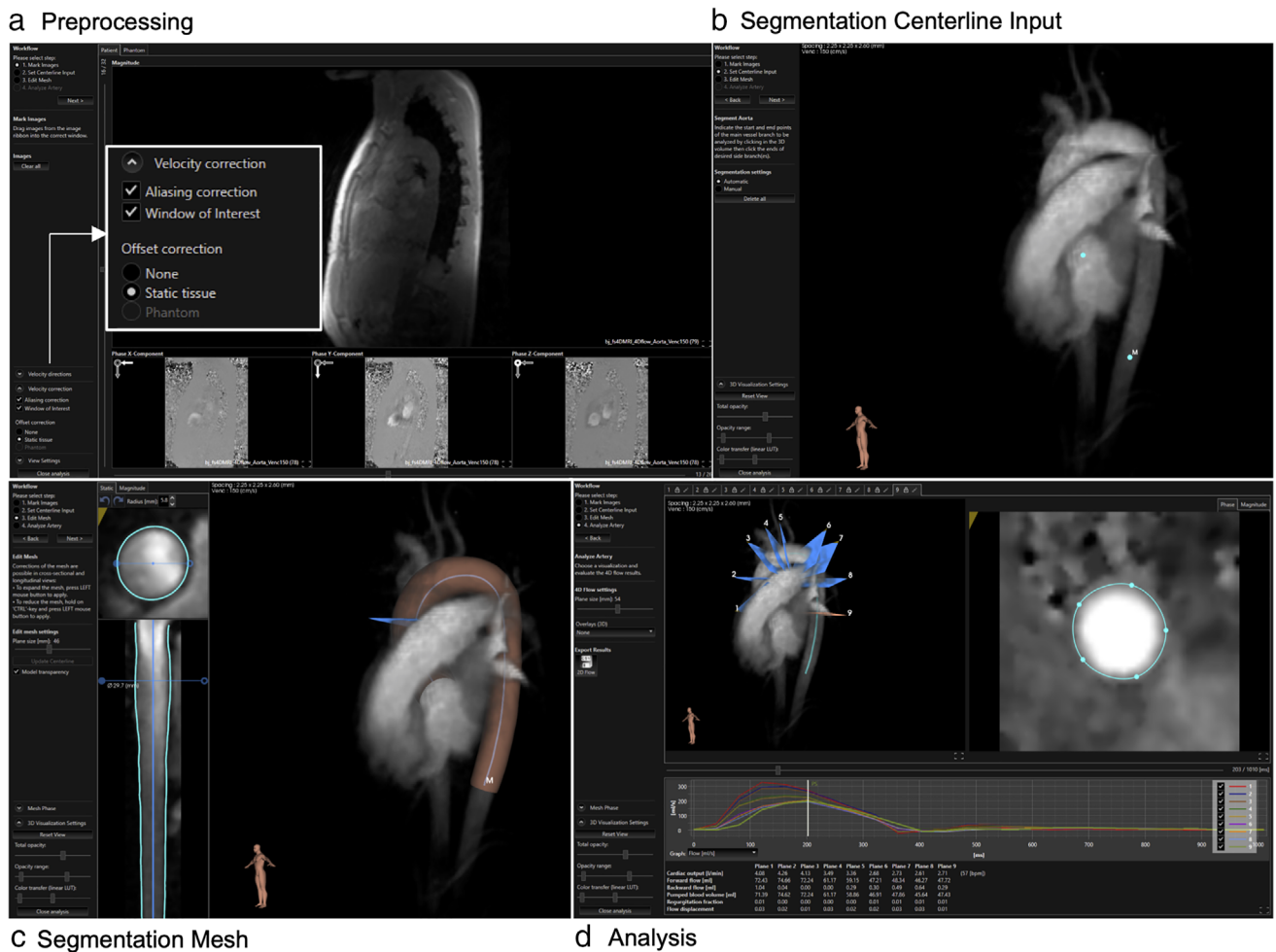


FIGURE 1: Process of 4D Flow MR analysis with CAAS. (a) Preprocessing with background phase offset and aliasing correction. (b) Segmentation of the aorta: determining the start and end point of the centerline using a 3D display of the heart and great vessels. (c) Segmentation of the aorta: editing the Mesh model based on the static tissue for each 2D plane along the aorta in the peak systolic phase. The Mesh model represents a 3D volume model of the aorta for one cardiac phase. It is automatically generated for five cardiac phases (peak systolic phase ± 2 cardiac phases). We selected the peak systolic phase for subsequent manual segmentation and analysis. (d) Localization of the nine planes in the thoracic aorta for analysis of flow parameters with exemplary drawn vessel contours for plane 9.
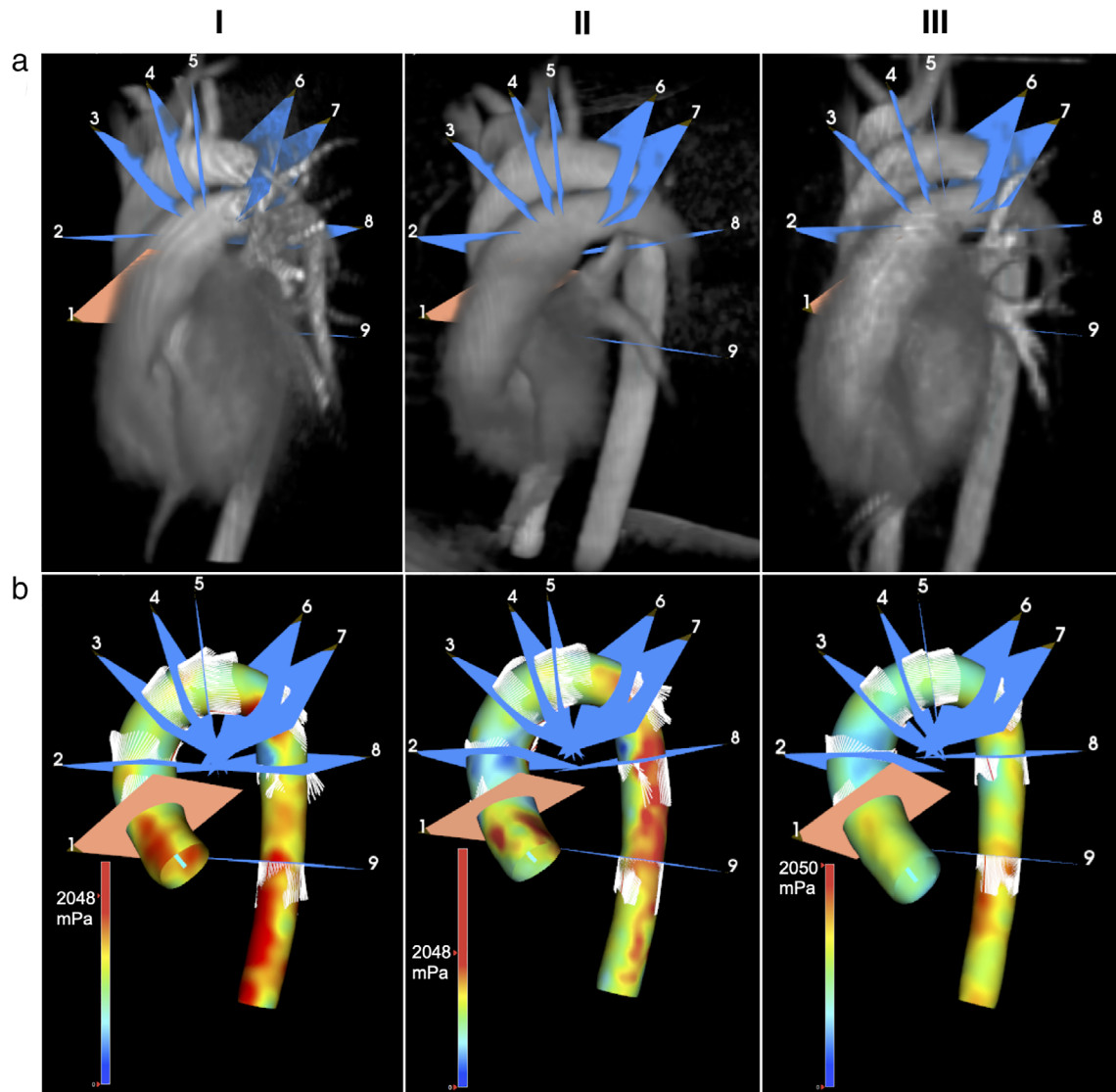
**FIGURE 2: Visualization of the 4D Flow MR data obtained by the different scanners for evaluation of flow parameters (a) and wall shear stress (WSS) (b) from one volunteer. (a) Nine cross-sectional planes were positioned along the centerline and perpendicularly to the longitudinal axis of the aortic wall as follows: at the level of the sinotubular junction (P1), in the mid-ascending aorta (P2), proximal to the brachiocephalic trunk (P3), between the brachiocephalic trunk and the left common carotid artery (P4), between the left common carotid artery and the left subclavian artery (P5), at the aortic isthmus (P6), in the descending aorta above the pulmonary artery (P7), in the descending aorta below the pulmonary artery (P8), in the descending aorta below the level of the aortic valve (P9). (b) Distribution of maximum WSS is visualized in a 3D segmented model of the aorta for the peak systolic phase. As shown in the color bar red-colored areas represent highest values for maximum WSS. The color bars were adjusted as far as possible between the three acquisitions for better visual comparison.**

automatically transferred. WSS was automatically calculated based on the segmented 3D volume model for the peak systolic phase and extrapolated to the point of zero velocity for defining the location of the aortic wall.[21] No manual adaption was necessary. For each plane, the WSS vector along the surface of the 3D model was calculated for 90 wall points. Maximum WSS was determined as the highest of all WSS values and average WSS as the average of all WSS values in the peak systolic phase.

For intraobserver analysis, the scans of all 10 volunteers at all three scanners (30 scans in total) were evaluated twice. All scans were analyzed by a blinded second observer (SW, 5 years of 4D Flow MR experience, >10 years of MR experience) for interobserver analysis as well. Both analyses only included plane positions in the ascending

aorta (P1–P3) as most clinically relevant pathologies occur in this region.

### Statistical Analysis

Statistical analyses were performed with SPSS (V25.0, IBM Corp., Armonk, NY, USA) and SAS (V9.4, SAS Institute Inc., Cary, NC, USA). Bland–Altman analyses were generated for agreement of the values acquired at the different scanners. As previously published, for equivalence testing the 95% confidence interval (CI) of the intraobserver difference for each parameter was set as the margin of acceptable disagreement regarding the comparison between vendors.[22] Results from two vendors were considered equivalent if the limits of their difference would lie within the 95% CI of

intraobserver difference. Additionally, equivalence was tested using the 95% CI of scan-rescan difference as the accepted range. Significance testing was performed for each plane separately using the Wilcoxon and Friedman tests. Numeric data of the combined planes (eg, P1–P9) were stated as the mean ± SD of their average value. Linear mixed models with repeated measures taking into account the correlation of multiple observations per patient were additionally applied to determine statistically significant differences jointly for the planes P1–P9 as well as in the ascending aorta (P1–P3), aortic arch (P4–P6), and descending aorta (P7–P9). A hierarchical test strategy was used. Only if significance was shown in the global test (scanner I vs. II vs. III) significance was assessed in the pairwise comparison between two scanners also at a two-sided level of 5%.

Depending on the distribution (normal vs. non-normal), significance of scan-rescan variability was tested with the paired $t$-test or Wilcoxon test. For agreement of scan-rescan reproducibility, Bland–Altman analyses and the intraclass correlation coefficient (ICC) were calculated. ICC was interpreted as follows: >0.9: excellent, 0.75–0.9: good, 0.5–0.75: moderate, <0.5: poor.[23] Intra- and interobserver reliability was determined using the ICC and Bland–Altman plots.

## Results

Fifteen healthy volunteers (mean age 24.5 ± 5.3 years, 8 females) without a history of cardiovascular diseases were enrolled in the study. Baseline characteristics are provided in Table 2. Out of these, 10 volunteers completed the study and could be included in further analysis. Following circumstances resulted in exclusion of the five volunteers. Due to unexpected technical problems, one volunteer could not be scanned at scanner II. Three datasets from scanner I were lost due to incomplete reconstruction after data acquisition. Another 4D Flow MR acquisition obtained at scanner I was excluded as inaccurate image planning led to missing aortic segments. In one volunteer, one plane (P2) had to be excluded as the aortic segment was truncated during image planning, while aliasing led to an exclusion of two additional planes (P1, P5) in two other volunteers. In total, we analyzed 90 segments and 267 planes in 10 volunteers. The time period between the first and second scan sessions (at scanners I and II) was 103 ± 4 days, while the second and third scans were performed within 18 ± 10 days.

### Comparison of Different Vendors

Average heart rate did not differ significantly between the scans (69.6 ± 9.4 bpm vs. 77.4 ± 6.8 bpm vs. 78.9 ± 13.5 bpm; $P = 0.067$). Systolic and diastolic blood pressure levels significantly varied (sys: 111.0 ± 10.0 mmHg vs. 121.4 ± 11.8 mmHg vs. 118.8 ± 10.1 mmHg; dia: 59.8 ± 5.4 mmHg

---

**TABLE 2. Baseline Characteristics of the Healthy Volunteers**

| Variables | Inter-Vendor ($N = 10$) | | Scan-Rescan ($N = 15$) | |
|---|---|---|---|---|
| Gender (female) | 6 | | 8 | |
| Age (years) | 24.1 (±5.4) | | 24.5 (±5.3) | |
| Height (cm) | 174.1 (±7.2) | | 173.9 (±8.8) | |
| Weight (kg) | 66.2 (±10.5) | | 65.6 (±10.6) | |
| BMI (kg/m$^2$) | 21.7 (±2.2) | | 21.6 (±2.2) | |
| BSA (m$^2$) | 1.8 (±0.2) | | 1.8 (±0.2) | |
| LVEDV (mL) | 145.4 (±31.5) | | 146.6 (±32.6) | |
| LVEF (%) | 63.8 (±4.5) | | 64.7 (±4.8) | |
| Heart rate at scanner I (bpm) | 69.6 (±9.4) | $P = 0.067$ | 76.1 (±11.5) | $P = 0.986$ |
| Heart rate at scanner II (bpm) | 77.4 (±6.8) | | 76.1 (±11.1) | |
| Heart rate at scanner III (bpm) | 78.9 (±13.5) | | | |
| Blood pressure at scanner I (mmHg) | 111.0 (±10.0)/59.8 (±5.4) | $P = 0.03$<br>$P = 0.005$ | 119.4 (±12.0)/61.5 (±9.6) | $P = 0.116$<br>$P = 0.599$ |
| Blood pressure at scanner II (mmHg) | 121.4 (±11.8)/69.0 (±8.0) | | 123.7 (±17.1)/60.6 (±10.5) | |
| Blood pressure at scanner III (mmHg) | 118.8 (±10.1)/60.4 (±6.1) | | | |

Values are presented as mean ± SD.
BMI = body mass index; BSA = body surface area; LVEDV = left ventricular end-diastolic volume; LVEF = left ventricular ejection fraction.

vs. 69.0 ± 8.0 mmHg vs. 60.4 ± 6.1 mmHg). Figure 2 exemplary illustrates a visual impression of the analyzed 4D Flow MR data from one volunteer at the three different scanners.

Forward flow volume showed significant differences between all scanners regarding P3–P9 (59.1 ± 13.1 mL vs. 68.1 ± 12.0 mL vs. 55.4 ± 13.1 mL), and for the descending aortic segment (52.2 ± 12.4 mL vs. 60.7 ± 13.3 mL vs. 50.1 ± 12.3 mL) (Tables S1 and S2 in the Supplemental Material). Only P1, located at the sinotubular junction, indicated no difference for pairwise comparison (I vs. II $P = 0.110$; I vs. III $P = 0.646$; II vs. III $P = 0.314$), as well as for comparison of all three vendors (78.5 ± 15.1 mL vs. 80.3 ± 15.4 mL vs. 79.5 ± 19.9 mL; $P = 0.368$). Forward flow volume between the values obtained from scanners I and III agreed best in Bland–Altman analyses (Table 3, Fig. S1 in the Supplemental Material), while the highest values were detected by scanner II. Peak velocity differed significantly between the three scanners within the descending aortic segment (125.6 ± 17.5 cm/s vs. 113.7 ± 13.2 cm/s vs. 111.1 ± 19.1 cm/s) and in a pairwise or three-way comparison within all planes except P1 (126.4 ± 16.7 cm/s vs. 119.7 ± 13.6 cm/s vs. 111.2 ± 22.6 cm/s; $P = 0.097$) (Tables S1 and S2 in the Supplemental Material). Highest values for peak velocity resulted from scanner I, lowest from scanner III. In Bland–Altman analyses the results of scanners II and III were most comparable (Table 3, Fig. S1 in the Supplemental Material). Equivalence could not be concluded for both flow parameters as the range of 95% CI of differences between the scanners exceeded the margin of intraobserver variability. When setting the 95% CI of scan-rescan variability as accepted range, only the difference between scanners II and III for peak velocity laid within the limits and were therefore found to be equivalent (Fig. 3, Table 3).

Maximum WSS varied significantly between all scanners for P1–P9, aortic arch, and the descending aorta as well as all planes separately except P1 ($P = 0.062$) and P2 ($P = 0.121$) (Tables S1 and S2 in the Supplemental Material). Average WSS results were similar, revealing a non-significant difference only for P2 ($P = 0.264$) and the ascending aortic segment ($P = 0.357$). Lowest values for maximum and average WSS resulted from data of scanner III (WSS max: I = 1790.5 ± 172.4 mPa vs. II = 1903.2 ± 430.5 mPa vs. III = 1458.7 ± 238.8 mPa; WSS avg: I = 1344.6 ± 136.9 mPa vs. II = 1251.9 ± 214.2 mPa vs. III = 1102.8 ± 187.6 mPa). Between scanners I and II, no significant difference for maximum WSS in all planes ($P = 0.406$) and segments (ascending: $P = 0.622$, arch: $P = 0.961$, descending: $P = 0.297$) and only significant differences in P5, P9, and the aortic arch for average WSS could be observed (ascending: $P = 0.204$, descending: $P = 0.555$) (Tables S1 and S2 in the Supplemental Material). Both WSS

TABLE 3. Mean Values and Confidence Intervals of Bland–Altman Analyses for Inter-Vendor, Scan-Rescan, Intra- and Interobserver Reproducibility

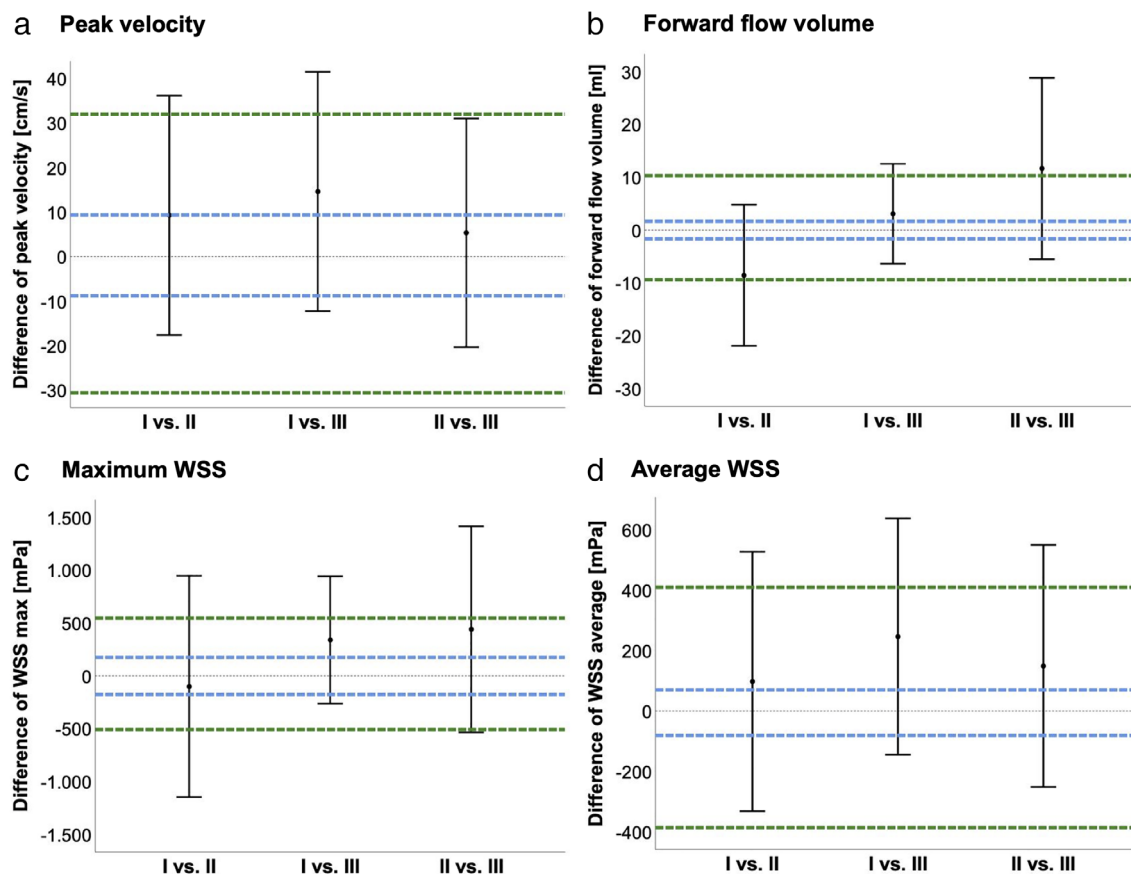| | Peak Velocity (cm/s) | Forward Flow Volume (mL) | Maximum WSS (mPa) | Average WSS (mPa) |
|---|---|---|---|---|
| Inter-vendor | | | | |
| I vs. II | 9.28 (36.17; −17.60) | −8.57 (4.80; −21.93) | −100.48 (945.79; −1146.75) | 97.71 (527.48; −332.06) |
| I vs. III | 14.19 (41.45; −13.07) | 3.01 (12.94; −6.92) | 332.12 (935.85; −271.61) | 242.53 (635.97; −150.90) |
| II vs. III | 5.23 (30.82; −20.35) | 11.54 (28.73; −5.65) | 433.78 (1409.89; −542.32) | 145.12 (549.81; −259.57) |
| Scan-rescan | 0.69 (31.99; −30.60) | 0.51 (10.31; −9.29) | 24.01 (548.70; −500.68) | 13.05 (412.80; −386.70) |
| Intraobserver | 0.27 (9.34; −8.81) | 0.00 (1.66; −1.66) | −0.48 (175.55; −176.51) | −5.54 (69.92; −81.00) |
| Interobserver | 1.92 (16.47; −12.64) | −2.41 (2.78; −7.61) | 29.31 (605.41; −546.79) | 71.19 (415.85; −273.47) |

WSS = wall shear stress.

**FIGURE 3:** Equivalence testing for inter-vendor comparison in relation to intraobserver and scan-rescan variability for peak velocity (a), forward flow volume (b), maximum wall shear stress (WSS) (c), and average WSS (d). Equivalence of two vendors is shown if the 95% confidence interval (CI) for the scanner comparison (indicated as gray vertical bars) lays within the equivalence limits of the intraobserver variability (blue dashed lines) or scan-rescan variability (green dashed lines). The dashed lines display the range of intraobserver/scan-rescan variability as 95% CI of the differences of the measured values.

parameters exceeded the margin of intraobserver and scan-rescan variability as accepted difference (Fig. 3, Table 3).

### Scan-Rescan Reliability

All 15 volunteers were scanned twice at scanner III and no acquisition or plane had to be excluded. Average heart rate (76.1 ± 11.5 bpm vs. 76.1 ± 11.1 bpm; $P = 0.986$) and blood pressure (sys: 119.4 ± 12.0 mmHg vs. 123.7 ± 17.1 mmHg [$P = 0.116$]; dia: 61.5 ± 9.6 mmHg vs. 60.6 ± 10.5 mmHg [$P = 0.599$]) did not vary significantly between scan and rescan. No significant difference was found for all parameters in all planes and segments (P1–P9: peak = 102.4 ± 14.6 cm/s vs. 101.7 ± 18.1 cm/s [$P = 0.859$]; forward flow = 60.0 ± 13.2 mL vs. 59.5 ± 13.7 mL [$P = 0.621$]; WSS max = 1470.4 ± 207.8 mPa vs. 1446.4 ± 284.5 mPa [$P = 0.955$]; WSS avg = 1118.3 ± 163.3 mPa vs. 1105.2 ± 211.7 mPa [$P = 0.795$] (Table S3 in the Supplemental Material). Agreement for forward flow volume was excellent (ICC = 0.98) and good for peak velocity (ICC = 0.83), average WSS (ICC = 0.84), and maximum WSS (ICC = 0.80). Bland–Altman analyses showed good agreement with narrow CI (Table 3, Fig. 4). Equivalence set by the intraobserver variability

could not be concluded as all parameters exceeded the margin (Fig. 3, Table 3).

### Inter- and Intraobserver Reliability

Bland–Altman plots and ICC indicated excellent intraobserver reproducibility (forward flow: ICC = 1.0; peak velocity: ICC = 0.99; maximum WSS: ICC = 0.99; average WSS: ICC = 1.0; Table 3, Fig. 4).

Excellent interobserver agreement was achieved for forward flow (ICC = 0.99), peak velocity (ICC = 0.96), and maximum WSS (ICC = 0.92), while average WSS displayed good agreement (ICC = 0.86) (Fig. 4).

### Discussion

This study revealed significant differences for 4D flow derived hemodynamic parameters when using vendor-provided and clinical used protocols at three different sites. All parameters exceeded the equivalence range set by the intraobserver analysis. For representation of the real clinical setting and the current level of standardization each site used their individual and vendor-provided standard 4D Flow MR sequence for acquisition. The study results do not provide information
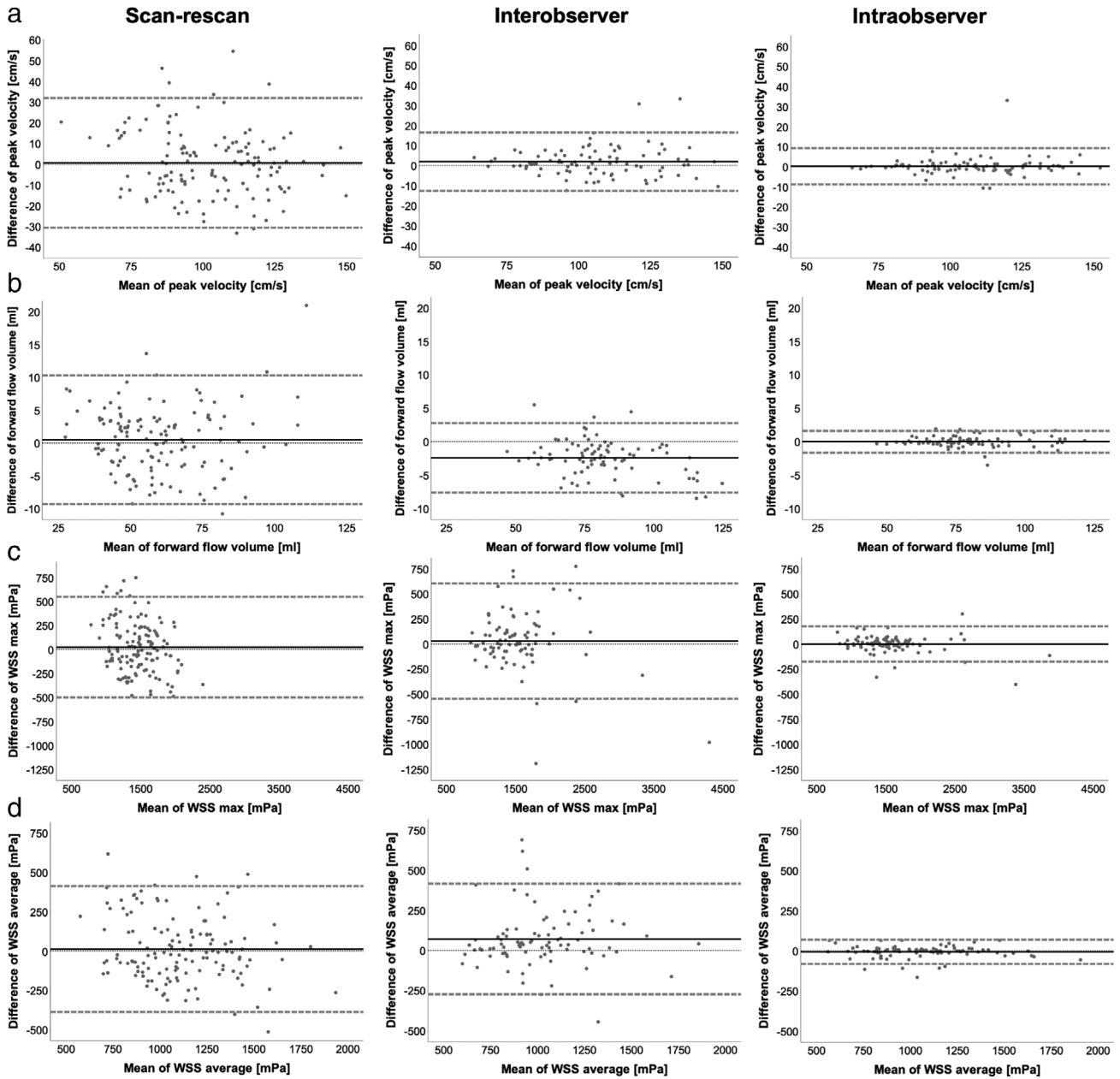
**FIGURE 4:** Bland–Altman plots for scan-rescan reproducibility and intra- and interobserver variability of peak velocity (**a**), forward flow volume (**b**), maximum wall shear stress (WSS) (**c**), and average WSS (**d**). Solid black lines indicate mean difference, dashed lines indicate limits of agreement (95% confidence interval of the differences of the measured values).

about which of the 4D flow sequences is the most accurate or precise. However, they show that 4D flow derived hemodynamic parameters cannot easily be compared for a follow-up when a subject is examined at different scanners under diverging conditions, exemplary different vendors, and/or sites. Also, the differences in protocols identify the lack of commercial 4D Flow MR sequences across platforms and centers in the clinical routine, yet reflecting the current status quo. Additionally, many data sets or planes needed to be discarded, highlighting the complexity of this method despite continuous technical development of 4D Flow MR sequences and postprocessing tools.

Bock et al validated 4D Flow MR sequences in a phantom and in vivo comparing flow volumes between a 1.5-T Siemens and Philips scanner, which showed a larger variation of kinetic energy between the two scanner types than between repeated measurements at one scanner.[24] This is in line with the findings of this study.

Watanabe et al investigated the impact of changing the vendor of the scanner on flow velocity using a straight-tube phantom. Their results showed satisfying accuracy of flow velocity profiles in all three 3 T scanners (GE, Philips, Siemens).[25] Additionally, they measured time-averaged flow velocity perpendicular to the tube central, which

corresponded well with the reference values obtained by a flowmeter for each vendor, but reference values slightly differed between the scanners. Absolute values for peak velocity were also similar between the scanners in the present study. Further, a statistical analysis including equivalence testing as a method for accepted difference range from clinical perspective was performed. Peak velocity was the most stable parameter as equivalence was achieved between two scanners within the limits of scan-rescan variability.

Wen et al evaluated the multicenter reproducibility of neurovascular 4D flow magnetic resonance imaging (MRI) scans on 10 healthy volunteers at three different sites using a 3-T GE scanner with identical coils and acquisition parameters. They observed great multisite reproducibility for measurement of blood flow and good reproducibility for peak velocity in intracranial vessels.[26] Their findings suggest the significant disagreement in 4D flow parameters detected in the present study to be rather induced by the vendor-associated difference of acquisition parameters than practical experience in the different sites.

Systolic and diastolic blood pressure levels varied significantly between the scans indicating hemodynamic changes in each volunteer. These physiological changes may have been induced by differences in fluid or food intake before 4D Flow MR acquisition, as the volunteers did not fast.[27] However, absolute values of blood pressure levels were similar and from a clinical perspective this discrepancy represents no clinical relevance as heart rate did not differ significantly. Additionally, all three scanners were wide-bore (70 cm) scanners leading to a similar level of rather common claustrophobia induced stress. This minimalizes the potential influence on hemodynamic parameters caused by different levels of claustrophobic stress. Also, in larger bores effects due to a more heterogeneous field are commonly found.

Since the plane at the level of the sinotubular junction (P1) showed the lowest number of significant differences in hemodynamic parameters, we considered this region as the most stable plane location. Additionally, when comparing the cardiac outputs resulting from the cine images and the 4D flow analysis (forward volume at P1 × heart rate during 4D flow) exemplarily for all 15 volunteers at scanner III, no significant difference could be found. This can serve as an external validation of the 4D flow derived flow volumes in this study.

The stability of P1 is especially interesting for clinical purposes as the position is identical to the localization applied in clinically used 2D flow measurements holding clinical importance for diagnosis of aortic flow measurements.[28] This stability could be explained by its central location independent of the defined acquisition volume (whole heart vs. aorta). Plane positions that are more distant from the isocenter of the magnet rather lead to inaccurate flow results mainly caused by eddy currents yielding phase offsets.[29]

Therefore, background correction has been applied by a linear fit to the static tissue. However, eddy current related phase contributions may vary between scanners and potentially fitting models including higher order terms may be needed for comparisons between different vendors, particularly for the more distant slices. Furthermore, a simple approach to verify the precision of the applied background compensation is missing for clinical routine and the correction may be insufficient.[30]

The different acquisition volumes applied in this study could represent a factor leading to disagreement in flow parameters due to various isocenter distances regarding P2–P9. Yet, comparison of both whole heart sequences showed significant differences in forward flow volume and peak velocity for many planes and segments, thus demonstrating that the definition of the acquisition volume may not be the only influential factor.

4D Flow MR sequences in this study were scanner-specific containing variations in acquisition parameters which may have contributed to the observed significant differences. Multiple research groups studied the impact of acquisition parameters on results of different flow parameters and WSS. Low spatial resolution had the most influence on WSS leading to underestimation of WSS values.[31–33] Stalder et al showed with a synthetic model that, eg, for a voxel length of 1 mm WSS is lowered to 60% of its actual value and decreases to above 30% for a voxel length of 10 mm, whereas total blood flow remained comparatively steady.[31] In this study, the highest acquired voxel size and thereby lowest spatial resolution was obtained at scanner I, but WSS values were lowest at scanner III. Also lower values for forward flow volume and peak velocity were found at scanner III, which should not be affected by the difference in spatial resolution, but which could be related to the different acceleration techniques applied at the scanners (Autocalibrating Reconstruction for Cartesian sampling, GeneRalized Autocalibrating Partially Parallel Acquisition, Sensitivity Encoding). Highest VENC was set at scanner I and may have induced higher WSS values as Zimmermann et al detected.[33] However, in scans acquired at scanner II, also higher WSS values were measured without significant difference to the values measured at scanner I, although the VENC was lower. Furthermore, varieties in temporal resolution and the amount of reconstructed cardiac phases could have led to determination of slightly different absolute time points of the peak systolic phase between the scans, elevating the variability in WSS results.

Montalba et al showed greater underestimation and variability for flow parameters when using 4D flow acquisitions with lower temporal resolutions.[34] In this study, the lowest temporal resolution was obtained at scanner II, but lowest flow parameters resulted from 4D flow images acquired at scanner III. The lowest forward flow volume

could be partly attributed to the difference in ECG gating, as the only prospectively gated sequence, which does not cover the entire cardiac cycle missing the late diastolic phases, was used at scanner III. Although aortic flow is primarily expected in the systolic phases in healthy volunteers, prospective gating could cause lower forward flow volumes missing the early systolic flow due to a delay in detecting the R-wave.

Considering knowledge on the influence of acquisition parameters on the results the differences in protocol settings represent potential confounders and sequence parameters should be matched as far as possible between different scanners. However, as each scanner type presents individual specifications, individual adjustments are needed. Furthermore, individual and site-dependent adjustments in protocol settings can be necessary for optimizing the vendor-provided protocols from a clinical perspective. Additional site-associated influence factors such as potential operator-dependent variabilities in levels of experience with 4D Flow MR planning at the different sites could have caused the differences as well.

Regarding scan-rescan reliability, the results showed no significant differences for all parameters and highlighted forward flow volume as most constant parameter with excellent scan-rescan and intra- and interobserver agreement. Markl et al also found excellent scan-rescan, intra- and interobserver agreement for calculation of total flow and peak velocity vs. greater variability for WSS.[35] Compared to flow parameters, evaluation of WSS seems to depend more on the definition of vessel wall contours, leading to more variability by differing segmentation contouring.[31,33,35] This might explain the wider CI for WSS parameters reached by the interobserver agreement. The broader CI becomes particularly clear outreaching the level of the scan-rescan CI for maximum WSS.

In relation to inter- and intraobserver reproducibility, scan-rescan comparison resulted in higher variability. This may likely be determined by differences in repositioning of the coil and image planning with, eg, various isocenter locations between both acquisitions. Further, length of still recumbency was longer before the first 4D Flow MR acquisition than the rescan and may have induced hemodynamic changes, although blood pressure and heart frequency were similar between both scans. In addition, the scan-rescan difference was statistically non-significant, and ICC reached good-excellent agreement.

### Limitations

The study consisted of only a smaller number of healthy volunteers with a high drop-out rate and did not include patients or elderly, but it was already challenging to organize and implement this quantity of scans at different sites throughout Germany. Reflecting the status quo, each site used their individual and vendor-provided standard 4D Flow

MR sequence, instead of agreeing on a common adapted sequence. Thus, acquisition parameters differed between the scans influencing the inter-vendor agreement but reflected the real-life setting and current level of standardization. Additional scans with matched sequence parameters could not be performed for logistical reasons and the study refrained of a comparison of 4D Flow MR results with a reference gold standard, eg, 2D flow acquisitions or invasive hemodynamic measurements. Therefore, the results do not allow conclusions about which of the 4D flow sequences is the most precise or if the differences were more vendor or site associated, but neither of those was actually this study's aim. Also, changes in hemodynamic parameters during the various time between 4D Flow MR examinations at the different sites may have elevated the inter-vendor variability, however as all blood pressure measurements lay within a narrow range and the volunteers were all healthy, the impact should be small.

No adjustment was applied for the significance level due to the explorative character of the study. However, hierarchical testing was used by applying first the global test on differences between scanners before interpreting pairwise comparisons of two scanners.

## Conclusion

4D flow derived aortic hemodynamic parameters assessed with various vendor-provided and clinical 4D Flow MR protocols at the three different sites are not equivalent. Overall, the plane positioned at the sinotubular junction showed most inter-vendor stability and agreement. To enable large multi-center studies and patient follow-up examinations at various scanners and sites further investigation of the differences of sequences across centers and vendors is needed. The identification of confounders will help to overcome current limitations. Due to the lack of existing commercial 4D Flow MR sequences for clinical routine across platforms and centers, longitudinal studies and patient's follow-up examinations should take place at the same scanner.

## References

1. Markl M, Chan FP, Alley MT, et al. Time-resolved three-dimensional phase-contrast MRI. J Magn Reson Imaging 2003;17(4):499-506.

2. Dyverfeldt P, Bissell M, Barker AJ, et al. 4D flow cardiovascular magnetic resonance consensus statement. J Cardiovasc Magn Reson 2015;17(1):72.

3. Markl M, Kilner PJ, Ebbers T. Comprehensive 4D velocity mapping of the heart and great vessels by cardiovascular magnetic resonance. J Cardiovasc Magn Reson 2011;13(1):7.

4. Frydrychowicz A, Stalder AF, Russe MF, et al. Three-dimensional analysis of segmental wall shear stress in the aorta by flow-sensitive four-dimensional-MRI. J Magn Reson Imaging 2009;30(1):77-84.

5. van Ooij P, Markl M, Collins JD, et al. Aortic valve stenosis alters expression of regional aortic wall shear stress: New insights from a 4-dimensional flow magnetic resonance imaging study of 571 subjects. J Am Heart Assoc 2017;6(9):e005959.

6. Barker AJ, Markl M, Burk J, et al. Bicuspid aortic valve is associated with altered wall shear stress in the ascending aorta. Circ Cardiovasc Imaging 2012;5(4):457-466.

7. von Knobelsdorff-Brenkenhoff F, Trauzeddel RF, Barker AJ, Gruettner H, Markl M, Schulz-Menger J. Blood flow characteristics in the ascending aorta after aortic valve replacement—A pilot study using 4D-flow MRI. Int J Cardiol 2014;170(3):426-433.

8. Trauzeddel RF, Löbe U, Barker AJ, et al. Blood flow characteristics in the ascending aorta after TAVI compared to surgical aortic valve replacement. Int J Cardiovasc Imaging 2016;32(3):461-467.

9. van der Palen RLF, Barker AJ, Bollache E, et al. Altered aortic 3D hemodynamics and geometry in pediatric Marfan syndrome patients. J Cardiovasc Magn Reson 2017;19(1):30.

10. Ma LE, Markl M, Chow K, et al. Aortic 4D flow MRI in 2 minutes using compressed sensing, respiratory controlled adaptive k-space reordering, and inline reconstruction. Magn Reson Med 2019;81(6):3675-3690.

11. Sieren MM, Berlin C, Oechtering TH, et al. Comparison of 4D Flow MRI to 2D Flow MRI in the pulmonary arteries in healthy volunteers and patients with pulmonary hypertension. PLoS One 2019;14(10):e0224121.

12. Stankovic Z, Allen BD, Garcia J, Jarvis KB, Markl M. 4D flow imaging with MRI. Cardiovasc Diagn Ther 2014;4(2):173-192.

13. Zhong L, Schrauben EM, Garcia J, et al. Intracardiac 4D flow MRI in congenital heart disease: Recommendations on behalf of the ISMRM Flow & Motion Study Group. J Magn Reson Imaging 2019;50(3):677-681.

14. Lee Y, Callaghan MF, Acosta-Cabronero J, Lutti A, Nagy Z. Establishing intra- and inter-vendor reproducibility of T1 relaxation time measurement with 3T MRI. Magn Reson Med 2019;81(1):454-465.

15. Erley J, Zieschang V, Lapinskas T, et al. A multi-vendor, multi-center study on reproducibility and comparability of fast strain-encoded cardiovascular magnetic resonance imaging. Int J Cardiovasc Imaging 2020;36(5):899-911.

16. Caas 4D Flow. 2018. Available from: https://www.piemedicalimaging.com/product/mr-solutions/4d-flow (accessed January 4, 2021).

17. Bernstein MA, Zhou XJ, Polzin JA, et al. Concomitant gradient terms in phase contrast MR: Analysis and correction. Magn Reson Med 1998;39(2):300-308.

18. Walker PG, Cranney GB, Scheidegger MB, Waseleski G, Pohost GM, Yoganathan AP. Semiautomated method for noise reduction and background phase error correction in MR phase velocity data. J Magn Reson Imaging 1993;3(3):521-530.

19. Delingette H. General object reconstruction based on simplex meshes. Int J Comput Vis 1999;32(2):111-142.

20. van der Palen RLF, Roest AAW, van den Boogaard PJ, de Roos A, Blom NA, Westenberg JJM. Scan-rescan reproducibility of segmental aortic wall shear stress as assessed by phase-specific segmentation with 4D flow MRI in healthy volunteers. MAGMA 2018;31(5):653-663.

21. Pie Medical Imaging. User Manual Caas MR Solutions 5.0. Chapter 6: 4D Artery Workflow 2018. Available on request via pmi@pie.nl (accessed January 4, 2021).

22. Zange L, Muehlberg F, Blaszczyk E, et al. 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(1):12.

23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155-163.

24. Bock J, Töger J, Bidhult S, et al. Validation and reproducibility of cardiovascular 4D-flow MRI from two vendors using 2 × 2 parallel imaging acceleration in pulsatile flow phantom and in vivo with and without respiratory gating. Acta Radiol 2019;60(3):327-337.

25. Watanabe T, Isoda H, Fukuyama A, et al. Accuracy of the flow velocity and three-directional velocity profile measured with three-dimensional cine phase-contrast MR imaging: Verification on scanners from different manufacturers. Magn Reson Med Sci 2019;18(4):265-271.

26. Wen B, Tian S, Cheng J, et al. Test-retest multisite reproducibility of neurovascular 4D flow MRI. J Magn Reson Imaging 2019;49(6):1543-1552.

27. Hauser JA, Muthurangu V, Steeden JA, Taylor AM, Jones A. Comprehensive assessment of the global and regional vascular responses to food ingestion in humans using novel rapid MRI. Am J Physiol Regul Integr Comp Physiol 2016;310(6):R541-R545.

28. Schulz-Menger J, Bluemke DA, Bremerich 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(1):19.

29. Gatehouse PD, Rolf MP, Graves MJ, et al. Flow measurement by cardiovascular magnetic resonance: A multi-Centre multi-vendor study of background phase offset errors that can compromise the accuracy of derived regurgitant or shunt flow measurements. J Cardiovasc Magn Reson 2010;12(1):5.

30. Lotz J, Meier C, Leppert A, Galanski M. Cardiovascular flow measurement with phase-contrast MR imaging: Basic facts and implementation. Radiographics 2002;22(3):651-671.

31. Stalder AF, Russe MF, Frydrychowicz A, Bock J, Hennig J, Markl M. Quantitative 2D and 3D phase contrast MRI: Optimized analysis of blood flow and vessel wall parameters. Magn Reson Med 2008;60(5):1218-1231.

32. Petersson S, Dyverfeldt P, Ebbers T. Assessment of the accuracy of MRI wall shear stress estimation using numerical simulations. J Magn Reson Imaging 2012;36(1):128-138.

33. Zimmermann J, Demedts D, Mirzaee H, et al. Wall shear stress estimation in the aorta: Impact of wall motion, spatiotemporal resolution, and phase noise. J Magn Reson Imaging 2018;48(3):718-728.

34. Montalba C, Urbina J, Sotelo J, et al. Variability of 4D flow parameters when subjected to changes in MRI acquisition parameters using a realistic thoracic aortic phantom. Magn Reson Med 2018;79(4):1882-1892.

35. Markl M, Wallis W, Harloff A. Reproducibility of flow and wall shear stress analysis using flow-sensitive four-dimensional MRI. J Magn Reson Imaging 2011;33(4):988-994.