RESEARCH ARTICLE

NMR IN BIOMEDICINE WILEY

Validation of cardiac diffusion tensor imaging sequences: A multicentre test–retest phantom study

Irvin Teh[1] | William A. Romero R.[2] | Jordan Boyle[3] | Jaume Coll-Font[4] | Erica Dall'Armellina[1] | Daniel B. Ennis[5,6] | Pedro F. Ferreira[7,8] | Prateek Kalra[9] | Arunark Kolipaka[9] | Sebastian Kozerke[10] | David Lohr[11] | François-Pierre Mongeon[12] | Kévin Moulin[6] | Christopher Nguyen[4] | Sonia Nielles-Vallespin[7,8] | Brian Raterman[9] | Laura M. Schreiber[11] | Andrew D. Scott[7,8] | David E. Sosnovik[4] | Christian T. Stoeck[10] | Cyril Tous[13] | Elizabeth M. Tunnicliffe[14,15] | Andreas M. Weng[16] | Pierre Croisille[2] | Magalie Viallon[2] | Jürgen E. Schneider[1]

[1]Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, UK

[2]Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294, F-42023, Saint Etienne, France

[3]School of Mechanical Engineering, University of Leeds, Leeds, UK

[4]Cardiovascular Research Center and A. A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

[5]Division of Radiology, VA Palo Alto Health Care System, Palo Alto, California, USA

[6]Department of Radiology, Stanford University, Stanford, California, USA

[7]Cardiovascular Magnetic Resonance Unit, The Royal Brompton and Harefield NHS Foundation Trust, London, UK

[8]National Heart and Lung Institute, Imperial College London, London, UK

[9]Department of Radiology, The Ohio State University Wexner Medical Center, Columbus, Ohio, USA

[10]Institute for Biomedical Engineering, University and ETH Zurich, Zurich, Switzerland

[11]Department of Cardiovascular Imaging, Comprehensive Heart Failure Center, Würzburg, Germany

[12]Division of Non-invasive Cardiology, Montreal Heart Institute, Montreal, Canada

[13]Department of Radiology, Radiation-Oncology and Nuclear Medicine and Institute of Biomedical Engineering, Université de Montréal, Montréal, Canada

[14]Radcliffe Department of Medicine, University of Oxford, Oxford, UK

[15]Oxford NIHR Biomedical Research Centre, Oxford, UK

[16]Department of Diagnostic and Interventional Radiology, University Hospital Würzburg, Würzburg, Germany

Correspondence
Irvin Teh, University of Leeds, LICAMM, Worsley Building, Room 8.49j, Clarendon Way, Leeds, LS2 9NL, UK.
Email: i.teh@leeds.ac.uk

Cardiac diffusion tensor imaging (DTI) is an emerging technique for the in vivo characterisation of myocardial microstructure, and there is a growing need for its validation and standardisation. We sought to establish the accuracy, precision, repeatability and reproducibility of state-of-the-art pulse sequences for cardiac DTI among

Abbreviations used: ADC, apparent diffusion coefficient; CODE, convex optimised diffusion encoding; CV, coefficient of variation; DTI, diffusion tensor imaging; DW, diffusion-weighted; DWI, diffusion-weighted imaging; EPI, echo planar imaging; FA, fractional anisotropy; HCM, hypertrophic cardiomyopathy; M2B1resist, M2SE with improved robustness to B_1 inhomogeneities; M2SE, second-order motion-compensated spin echo; MD, mean diffusivity; MODE, motion-compensated optimised diffusion encoding; PGSE, pulsed gradient spin echo; PVP, polyvinylpyrrolidone; RMSD, root mean squared difference; ROI, region of interest; SD, standard deviation; SE, spin echo; SNR, signal-to-noise ratio; STEAM, stimulated echo acquisition mode.

10 centres internationally. Phantoms comprising 0%–20% polyvinylpyrrolidone (PVP) were scanned with DTI using a product pulsed gradient spin echo (PGSE; N = 10 sites) sequence, and a custom motion-compensated spin echo (SE; N = 5) or stimulated echo acquisition mode (STEAM; N = 5) sequence suitable for cardiac DTI in vivo. A second identical scan was performed 1–9 days later, and the data were analysed centrally. The average mean diffusivities (MDs) in 0% PVP were (1.124, 1.130, 1.113) x 10^{-3} mm^2/s for PGSE, SE and STEAM, respectively, and accurate to within 1.5% of reference data from the literature. The coefficients of variation in MDs across sites were 2.6%, 3.1% and 2.1% for PGSE, SE and STEAM, respectively, and were similar to previous studies using only PGSE. Reproducibility in MD was excellent, with mean differences in PGSE, SE and STEAM of (0.3 ± 2.3, 0.24 ± 0.95, 0.52 ± 0.58) x 10^{-5} mm^2/s (mean ± 1.96 SD). We show that custom sequences for cardiac DTI provide accurate, precise, repeatable and reproducible measurements. Further work in anisotropic and/or deforming phantoms is warranted.

KEYWORDS
cardiac DTI, isotropic phantom, multicentre, polyvinylpyrrolidone, pulse sequence validation, reproducibility

1 | INTRODUCTION

Diffusion tensor imaging (DTI) is an emerging noninvasive and contrast agent-free method for the characterisation of cardiac microstructure. It provides measurements, such as mean diffusivity (MD) and fractional anisotropy (FA), that are sensitive to the diffusion of water molecules, and therefore local tissue structure. Increased MD, for example, corresponded well to late gadolinium-enhanced segments in myocardial infarction,[1,2] while decreased FA can reflect cardiomyocyte disarray and increased extracellular volume in hypertrophic cardiomyopathy (HCM).[3] Changes in MD and FA have also been associated with a range of conditions including myocardial infarction,[4,5] hypertrophy,[6] athlete's heart,[7] fibrosis,[8] amyloidosis[9] and dilated cardiomyopathy.[10,11]

Despite recent advances, several major challenges need to be addressed to facilitate integration of DTI in clinical routine. The first is the uncertainty in choice of pulse sequence. As a result of cardiac motion, methods for motion compensation have had to be developed. These can be classed into two broad approaches, both employing cardiac triggering for synchronising to the cardiac cycle, and a single-shot echo planar imaging (EPI) readout for reduced motion sensitivity. The first method is based on stimulated echo acquisition mode (STEAM)[12] with monopolar[13] diffusion gradient waveforms. The second method is based on spin echo (SE) with motion-compensated diffusion gradient waveforms.[14] The most widely used SE implementation involves up to second-order motion compensation, rendering the sequence insensitive to constant velocity and acceleration. Second-order motion-compensated spin echo (M2SE) can be achieved with a range of methods, including asymmetric bipolar waveforms,[15] symmetric tripolar waveforms with improved robustness to B$_1$ inhomogeneities (M2B1resist)[16] and numerically optimised waveforms convex optimised diffusion encoding (CODE)[17] and motion-compensated optimised diffusion encoding (MODE)[18] with reduced echo times. Two studies compared the performance of STEAM and M2SE. One reported more than twofold higher signal-to-noise ratio (SNR) efficiency in M2SE compared with STEAM,[19] while the other, performed on a system with standard clinical gradients, observed that STEAM was more robust over a range of cardiac phases.[20]

The second challenge is the known sensitivity of diffusion MRI, not just to tissue properties, but also to acquisition parameters. Fitting of a tensor to diffusion-weighted (DW) data ignores non-Gaussian diffusion effects[21] stemming from the multiple compartments and barriers present within the complex cell microenvironment, leading to measures of apparent diffusivity that are sensitive to parameters such as pulse sequences,[19,20] diffusion times[22] and b-value.[23] Furthermore, DTI relies on low SNR DW images, leading to errors in measured parameters,[24,25] and sensitivity to image resolution and SNR.[26] The low SNR necessitates larger voxel sizes that enhance partial volume and residual motion artefacts,[27] while the rapid switching of diffusion gradient waveforms may enhance eddy current effects.[27] Additionally, early STEAM studies erroneously ascribed b-values of 0 s/mm^2 to the non-DW data during the reconstruction, leading to bias in MD, even within one sequence type.[28,29] Consequently, studies in the myocardium of healthy volunteers have reported a wide range of MD (0.87 x 10^{-3} to 1.72 x 10^{-3} mm^2/s) and FA (0.29 to 0.61).[19,20,30,31] This hampers comparison of data between sites and studies, and a standardised protocol is needed for comparison of absolute values.

The third challenge is the need to establish reproducibility of DTI parameters within and across sites. One intrasite study in healthy volunteers using a STEAM-EPI sequence showed no significant differences in myocardial MD and FA acquired at two time points.[28] Another single-site study in patients with HCM[32] reported coefficients of variation (CVs) of myocardial MD and FA across two time points of 19% and 7.2%, respectively. Intersite reproducibility is typically measured by scanning standardised subjects and substrates, such as travelling volunteers and phantoms, respectively, at multiple sites. Such studies have been performed most widely in the context of brain imaging, and have reported good reproducibility in MD and FA across up to 11 sites, with intersite CV_{MD} in brain ranging from 1.6% to 5.4% and CV_{FA} ranging from 2.0% to 4.5%.[33–35] One cardiac DTI study involving healthy volunteers at two sites reported myocardial CV_{MD} and CV_{FA} of up to 7% and 6%, respectively.[29]

While volunteers are key to assessing the real-world performance of DTI, phantoms are more cost-effective for larger scale studies, offer longer term stability and allow for customisation of features for isolation of sources of variation in the data. Isotropic phantoms, for example, ice water[33] or aqueous solutions of polyvinylpyrrolidone (PVP)[34,36,37] or nickel chloride and sodium chloride,[35] provide a homogeneous substrate with known MD, while biomimetic phantoms simulating cardiac microstructure[38] additionally permit reproducible measurements of FA. While such anisotropic phantoms could facilitate more realistic evaluation of anisotropy, these remain generally unavailable. Moreover, isotropic phantoms have a unique advantage, in that there exist gold standard reference measurements of diffusivity in the case of water, and corroborating MD values in the literature in the case of PVP.

In this multicentre study, we sought to establish the accuracy, precision, repeatability and reproducibility of state-of-the-art pulse sequences for cardiac DTI among 10 centres internationally. We employed a standardized, custom-built, temperature-controlled PVP phantom to permit baseline evaluation of MD and FA, as well as intrasite and intersite reproducibility in a homogeneous, isotropic substrate with known diffusivity, and acquired data using a product noncardiac DTI pulse sequence to serve as a reference.

2 | MATERIALS AND METHODS

Ten sites participated in the study (one Philips 1.5-T scanner, nine Siemens 3-T scanners). Phantoms were produced at a single site and shipped to participating sites. Data were acquired using a standardised protocol prescribing both product and custom DTI sequences. A second scan was performed from 1–9 days following the first scan. Data were uploaded to a central server, and analysed by a single site using a standardised processing pipeline.

2.1 | Phantom preparation

The phantom comprised 7 x 50-ml Falcon tubes filled with PVP (Sigma Aldrich, Dorset, UK) in distilled water in concentrations ranging from 0% to 20%. The tubes were positioned vertically in an outer container using a laser-cut plastic holder. Prior to imaging, the tubes were immersed in an ice-water bath and given sufficient time to equilibrate to 0°C (Figure 1). The phantoms were placed at the isocentre in identical orientation.

2.2 | Data acquisition

To verify consistent temperature, a fast diffusion-weighted imaging (DWI) scout scan was performed using the product sequence (i.e. pulsed gradient spin echo [PGSE] with single-shot EPI readout). The mean apparent diffusion coefficient (ADC) was measured in a region of interest (ROI) in
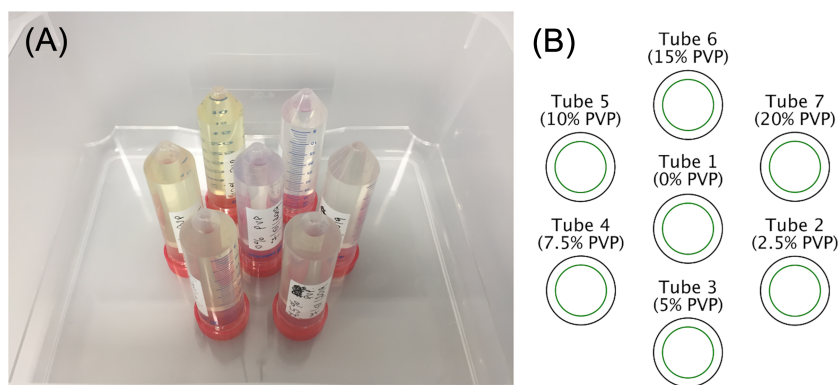

FIGURE 1 Phantom design. (A) Photograph of phantom prior to filling with ice-water bath. (B) Key for identifying tubes with different concentrations of polyvinylpyrrolidone (PVP)

the central tube. The scout scan was repeated at 10-min intervals, until the mean ADCs across two consecutive scans were within 3% agreement. Sites performed experiments using the product PGSE sequence[13] and a custom DTI sequence of their choice optimised for cardiac-specific applications. The scan parameters for the product sequence were TR/TE = 3000/85 ms, coronal view, field of view (FOV) = 300 × 230 mm, in-plane resolution = 2.5 mm, slice thickness = 8 mm, gap = 8 mm, slices = 3, bandwidth ∼ 3000 Hz, parallel imaging acceleration = 2x, ECG-triggered with simulated heart rate = 60 beats per min. Specified parameters such as TE and resolution were fairly conservative, to ensure that all sites could meet the sequence specifications in the presence of different hardware capabilities. To ensure consistency in diffusion schemes across different scanners, the following diffusion schemes were specified: (i) DWI scout scan: non-DW images = 1, diffusion vectors (3) = [1 0 0; 0 1 0; 0 0 1], b = 450 s/mm^2, repetitions = 3, acquisition time ∼36 s; (ii) DTI scan: non-DW images = 1, diffusion vectors[39] (6) = [0.5257 0.8507 0; 0.5257–0.8507 0; 0 0.5257 0.8507; 0 0.5257–0.8507; 0.8507 0 0.5257; −0.8507 0 0.5257], b = 100, 300, 450 s/mm^2, repetitions = 30, acquisition time ∼29 min. Each site performed imaging with one custom sequence of their preference, including spin echo-based sequences with up to second-order motion compensation (M2SE, CODE, MODE, M2B1resist) and STEAM (Figure 2). The relevant acquisition parameters can be found in Table 1 and the supporting information.

2.3 | Data analysis

Image reconstruction was performed on the scanners using standard vendor reconstruction software. Data storage and management were performed using the Human Heart Project, an online platform for heart imaging research.[40,41] Data were analysed centrally using open source Matlab
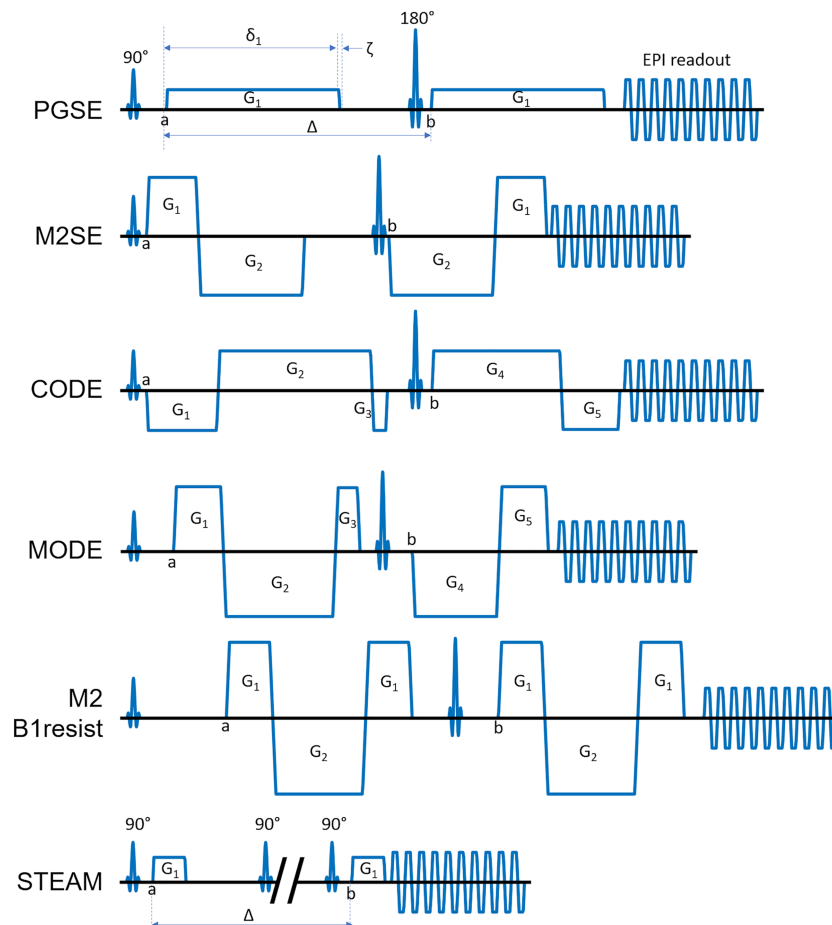
FIGURE 2 Schematic pulse sequence diagrams describing the diffusion waveforms used. These include the pulsed gradient spin echo (PGSE) product sequence and custom sequences as follows: motion-compensated spin echo (M2SE), convex optimised diffusion encoding (CODE), motion-compensated optimised diffusion encoding (MODE), motion-compensated symmetric spin echo (M2B1resist) and stimulated echo acquisition mode (STEAM). All custom spin echo sequences were motion-compensated up to the second order, and all sequences used a single-shot echo planar imaging readout. For clarity, preparation pulses, navigators, crushers and spoilers are not shown. Gradient waveform parameters, including diffusion gradient duration (δ_n) for each unique gradient lobe, time (b −a), slew time (ζ) and diffusion time (Δ) where applicable, are given in Table 1. δ_n was measured from the start of each unique gradient lobe (G_n) to the end of its plateau. Pulse sequences are not shown to scale

TABLE 1 Relevant hardware and custom sequence parameters

	Hardware						Pulse sequence (custom)								
Site	B_0 (T)	Scanner	G_{max} (mT/m)	Max slew rate (T/m/s)	Gradient coil	RF coil	Pulse sequence	TR (ms)	TE (ms)	Diffusion duration,$\delta_1 \dots \delta_n$ (ms)	Time (b – a) (ms)	Slew time, ζ (ms)	Grad Amp, $G_1 \dots G_n$ (mT/m)	b-value for non-DW (b_{low}; s/mm^2)	Crushers during DW scans
A	3	Prisma	80	200	XR	32Ch body + 18Ch spine	M2SE	3000	74	7.8/16	36	1.6	59	5	N
B	3	Prisma	80	200	XR	32Ch body + 18Ch spine	M2 B1resist	3000	97	7.0/14	41	1.0	76	2	Y
C	3	Prisma	80	200	XR	32Ch body + 18Ch spine	M2 B1resist	3000	97	7.0/14	41	1.0	76	2	Y
D	3	Prisma	80	200	XR	32Ch body + 18Ch spine	MODE	3000	75	7.3/17/3.7/13/7.2	36	1.4	65/ −65/66/ −65/65	0	Y
E	3	Skyra	45	200	XQ	32Ch body + 18Ch spine	CODE	3000	85	11/23/2.3/22/11	42	0.7	−40/40/ −40/40/ −39	5	N
F	3	Vida	60	200	XT	18Ch body + 18Ch spine	STEAM	2000	35	5.4	1000	0.9	15	42	N
G	3	Skyra	45	200	XQ	18Ch body + 32Ch spine	STEAM	2000	35	4.9	1003	1	16	76	N
H	1.5	Achieva	80/40	100/200	QD	32Ch cardiac (16/16 anterior/ posterior)	STEAM	6000	37	4.0	1000	0.21	20	39	N
I	3	Prisma	80	200	XR	18Ch body + 32Ch spine	STEAM	2000	34	4.0	1000	1.5	20	20	N
J	3	Prisma Fit	80	200	XR	32Ch body + 18Ch spine	STEAM	2000	24	2.8	998	0.4	28	16	N

Note: The timings and amplitudes of diffusion gradient waveforms (b = 450 s/mm^2) are stated with reference to the schematic pulse sequence diagrams given in Figure 2. In the case of PGSE and STEAM, time (b – a) corresponds to the diffusion time, Δ.

Abbreviations: CODE, convex optimised diffusion encoding; M2B1resist, M2SE with improved robustness to B_1 inhomogeneities; M2SE, second-order motion-compensated spin echo; MODE, motion-compensated optimised diffusion encoding; PGSE, pulsed gradient spin echo; STEAM, stimulated echo acquisition mode.

code (https://github.com/vigente/gerardus). The first step was semiautomatic tube segmentation, where tube centres were defined manually, and ROIs generated up to a fixed radius from the tube centres. Tensors were fit directly to the data using linear least squares, using all image repetitions without prior averaging. MD and FA were calculated as follows (Equations 1 and 2, respectively).

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \tag{1}$$

$$FA = \frac{1}{\sqrt{2}} \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}{\lambda_1{}^2 + \lambda_2{}^2 + \lambda_3{}^2}}, \tag{2}$$

where λ_1, λ_2 and λ_3 are the principal eigenvalues of the diffusion tensor. To obtain the reference diffusivity of water (D_{ref}), the self-diffusion of water, as measured using an independent non-MR technique, was extrapolated to 0°C using a second-order polynomial fit.[42] To assess the effect of b-values and rationalise an appropriate choice of b-values, average values of MD and FA across an ROI in tube 1 using data acquired with different b-value combinations are reported. Root mean squared differences (RMSDs) between scans 1 and 2 were assessed. MD and FA across ROIs in all tubes were reported using data from all repetitions, and also expressed as a time course reconstructed from subsampled single repetitions. The reference diffusivity for water was defined as D_{ref} (H_2O) = 1.113 x 10^{-3} mm^2/s at 0°C extrapolated from Mills.[42]

To calculate drift, the MD and FA across repetitions were first smoothed with a sliding window of five repetitions. Drift was determined by the difference between final and initial values. The stability of MD and FA measurements across repetitions were expressed in terms of the standard deviation (SD) across repetitions. As the non-DW data had variable b-values, image SNR was calculated based on the b = 100 s/mm^2 images averaged across six DW directions: SNR = mean/SD over repetitions.[43]

Accuracy in MD in each tube and site was expressed as[44]:

$$\text{Accuracy}, \partial_{j,k} = \frac{\overline{MD}_{j,k} - \overline{MD}_{ref,j}}{\overline{MD}_{ref,j}} \times 100\%, \tag{3}$$

where $\overline{MD}$ is averaged across each ROI and both scans. The reference $\overline{MD}_{ref}$ is the MD measured by PGSE averaged across each ROI, and all sites and both scans. j and k are the tube and site indexes.

Precision in MD in each tube and site was expressed in terms of[45]:

$$\text{Coefficient of variation, } CV(ROI)_{j,k} = \frac{[\sigma(MD_i)]_{j,k}}{[\mu(MD_i)]_{j,k}} \times 100\%, \tag{4}$$

where y σ is the SD of MD and μ is the mean MD across each ROI in scan 1. i, j and k are the voxel, tube and site indexes, respectively.

Intrasite repeatability was assessed by Bland–Altman plots, with mean differences and 95% limits of agreement reported.

Intersite reproducibility was assessed in terms of:

$$\text{Coefficient of variation, } CV_{k,l,m} = \frac{\sigma(\overline{MD}_{k,l,m})}{\mu(\overline{MD}_{k,l,m})} \times 100\%, \tag{5}$$

where σ is the SD of MD and μ is the mean MD across each variable separately, $\overline{MD}$ is averaged across each ROI and both scans. k is the site index (1, 2, ... 10), l is the sequence index (1:PGSE, 2:SE, 3:STEAM), m is the scan index (1, 2). We report the interscan, intersite (PGSE, SE, STEAM), intersequence (PGSE vs. SE, PGSE vs. STEAM) and intersite/sequence CV (SE vs. STEAM).

2.4 | Statistical analysis

Differences in MD with respect to D_{ref} were assessed by one-sample t-tests. Shapiro–Wilk tests were performed to test for normality in the accuracy of MD. Wilcoxon rank-sum tests were performed in each tube to determine the significance of differences between the medians of each sequence with respect to other sequences. The Bland–Altman results were analysed by pairwise comparisons of mean differences between PGSE, SE and STEAM, using two-sample t-tests with unequal variances. A significance level of $p = 0.05$ was used. To compare statistical distributions of the reproducibility of MD and FA, bootstrapped histograms were calculated by sampling the differences in MD and FA between scans 1000 times with replacement. Medians and 95% confidence intervals are reported. Histograms with nonoverlapping 95% CIs were deemed to be significantly different from one another.

3 | RESULTS

3.1 | Data quality

All sites acquired DTI data at two time points. The custom sequence acquisitions were split evenly into two groups, SE (N = 5) and STEAM (N = 5). While the product sequence acquisitions were standardised apart from small scanner variations in diffusion gradient duration (25.6 ± 2.9 ms) and diffusion time (39.9 ± 3.5 ms; mean ± SD across sites), there was greater variation across the custom sequence acquisitions. The main difference was that of all those sites that acquired STEAM data, only one site (site H) was able to acquire three slices of data as stipulated in the protocol, by slice interleaving and using TR = 6000 ms. To maintain consistent scan time, 15 repetitions were used in the analysis. The other four sites were unable to acquire multislice data in an interleaved fashion because of implementation limitations, and therefore only acquired single-slice data. As TR = 2000 ms was used, 45 repetitions were acquired to match scan times (sites F, G and I), although one site acquired 30 repetitions (site J). In the custom SE data, 30 repetitions were acquired with TR = 3000 ms, matching the product sequence. There were, however, instances in the custom SE data where not all 30 repetitions were available for all DW directions: 27 (site A, scan 1), 29 (site A, scan 2) and 28 (site E, scan 2) repetitions. Where a given repetition did not include all required DW directions, that repetition was excluded from analysis. In one site (site H), the data were exported at a reconstructed in-plane resolution of 0.94 x 0.94 mm, and had to be resampled to the nominal 2.5 x 2.5 mm resolution. Variations in parameters such as TR, TE and diffusion gradient timings are given in Table 1.

Figure 3 shows representative MD and FA maps in a single slice acquired at scan 1, using both product and custom sequences. Qualitatively, the maps were of good quality within the tubes. By contrast, the data in the surrounding ice-water bath were highly variable, and some voxels were masked out by the scanner reconstruction (site F). Gibb's ringing (all sites), geometric distortions (site G) and SENSE unfolding artefacts (site H) were observed.

3.2 | Mean diffusivity

The dependence of MD on b-values in the DTI reconstruction is illustrated in Figure 4. The results show that MD calculated from the product sequence data was relatively stable at $(1.120 \pm 0.022) \times 10^{-3}$ mm²/s across all b-value combinations. By contrast, greater sensitivity to b-value combinations was observed in MODE and STEAM custom sequences with MD = $(1.61 \pm 0.73, 1.06 \pm 0.13) \times 10^{-3}$ mm²/s across b-value combinations, respectively. The RMSDs of MD between scans 1 and 2 in the custom sequence data were $(0.214, 0.078, 0.048, 0.010, 0.005, 0.010, 0.012) \times 10^{-3}$ mm²/s across the b-value combinations b_{low}, 100; b_{low}, 300; b_{low}, 450; 100, 300; 100, 450; 300, 450; and b_{low},100, 300, 450, respectively. The lowest RMSD for MD was found using the combination of b = 100 and 450 s/mm², and this was therefore used for subsequent analysis. As phantom was isotropic, we focus on MD; corresponding FA measurements can be found in the supporting information.

Figure 5 depicts the MD averaged over ROIs within each tube. In tube 1 (0% PVP), the average MDs across all sites and scans 1 and 2 were $(1.124, 1.130, 1.113) \times 10^{-3}$ mm²/s for PGSE, SE and STEAM, respectively. Relative to the reference diffusivity of H₂O at 0°C of
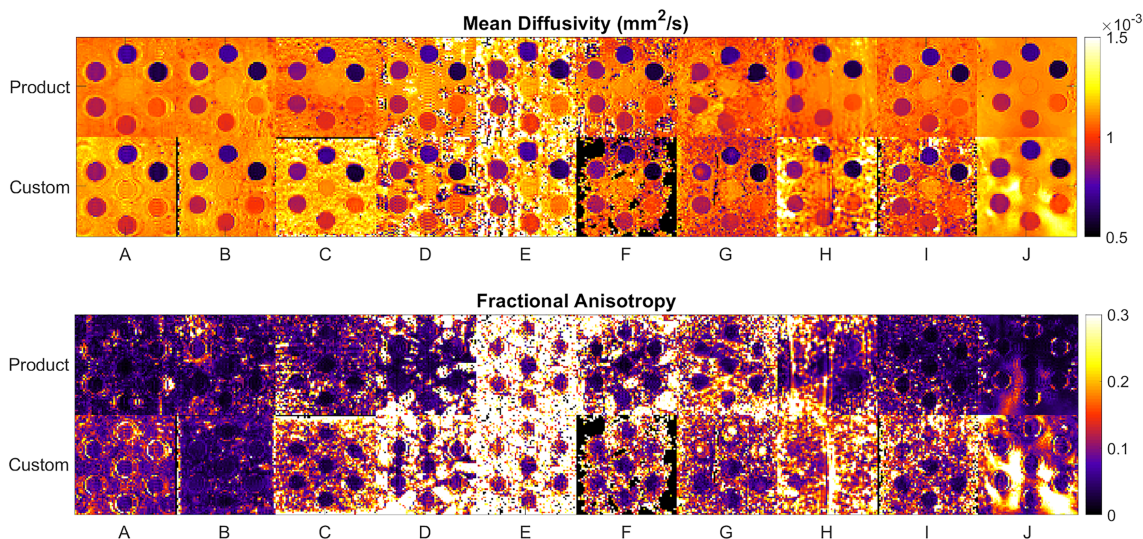

FIGURE 3 Mosaic of mean diffusivity (MD; top) and fractional anisotropy (FA; bottom) maps. Single-slice data are shown and were acquired using product (top) and custom sequences (bottom) at scan 1
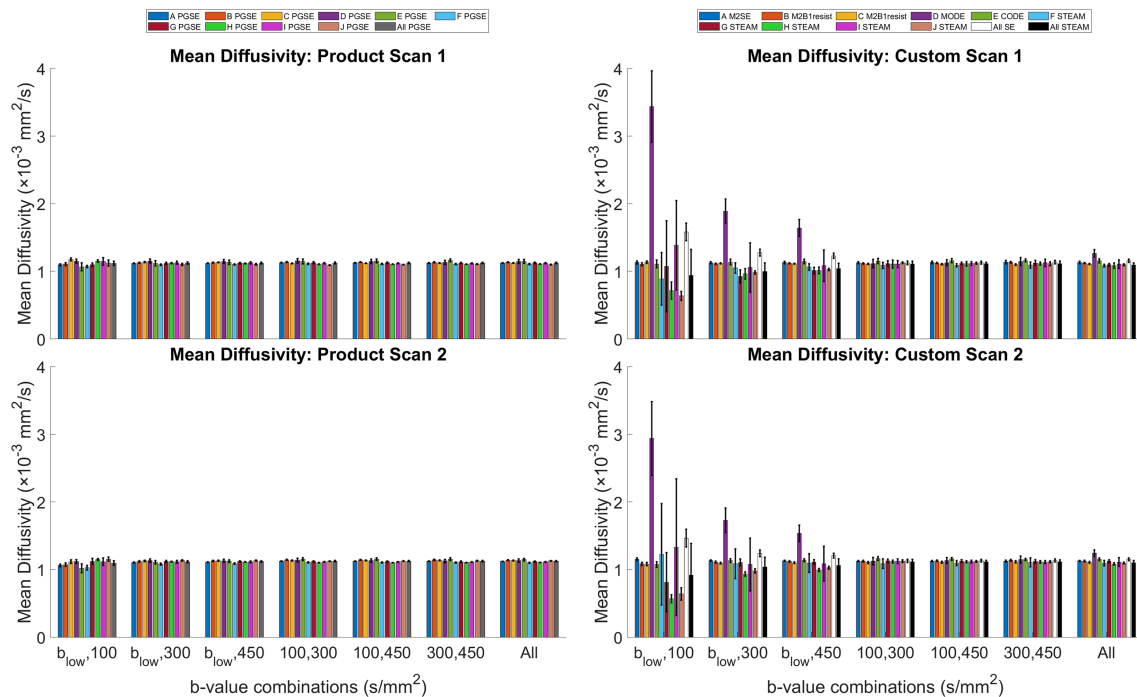
FIGURE 4 Sensitivity of mean diffusivity (MD) to b-values used in the diffusion tensor imaging (DTI) reconstruction. Average MD in tube 1 (0% polyvinylpyrrolidone) across sites, two time points, product (left) and custom sequence data (right) and b-value combinations. Mean ± SD across voxels in the region of interest. Average values for pulsed gradient spin echo (PGSE), spin echo (SE) and stimulated echo acquisition mode (STEAM) are given by grey, white and black bars, respectively. The b-values of nondiffusion-weighted data are denoted by b_{low}, and ranged from 0 to 76 s/mm^2 across sites (Table 1). MD and FA reconstructed using pairs of b-values that include b_{low} are generally further from expected values and less reproducible across scans. The combination of b = (100, 450) s/mm^2 was considered in subsequent analysis

1.113 x 10^{-3} mm^2/s, the average MD of tube 1, as measured by PGSE and SE, was overestimated by 1.0% ($p = 0.01$) and 1.5% ($p = 0.01$), respectively, while STEAM provided a result that was accurate to within 0.04% ($p = 0.9$). The SD across ROIs in tubes 1–7 increased with PVP concentration, particularly in the STEAM data. Summary values of MD and FA grouped by pulse sequence are presented in Table 2, alongside literature values where available.

3.3 | Accuracy

The accuracy of MD relative to the average MD measured across sites is presented in Figure 6. The null hypothesis that samples were normally distributed was rejected in three cases (STEAM tube 1, SE tube 3 and PGSE tube 7). Median accuracies between sequences were not significantly different, except for PGSE tube 2 versus STEAM tube 2 ($p < 0.05$). Within pulse sequences and between tubes, there were no significant differences in median accuracies.

3.4 | Precision

The precision of MD across pairs of sequences and tubes show that SE has significantly higher median CV than PGSE in three tubes, whereas STEAM has significantly higher median CV than PGSE in six tubes (Figure 7). There were clear differences as well when comparing medians between pairs of tubes, with precision worsening with increasing tube number (i.e. PVP concentration).

3.5 | Intrascan stability

Figure 8 shows the SNR of the b = 100 s/mm^2 data in a single repetition. In tube 1, the average SNRs across scans 1 and 2 were 23.3, 19.9 and 20.8 in PGSE, SE and STEAM, respectively, whereas in tube 7, the corresponding average SNRs were 44.1, 41.8 and 31.2. The stability of measurements across repetitions is reported in the supporting information.
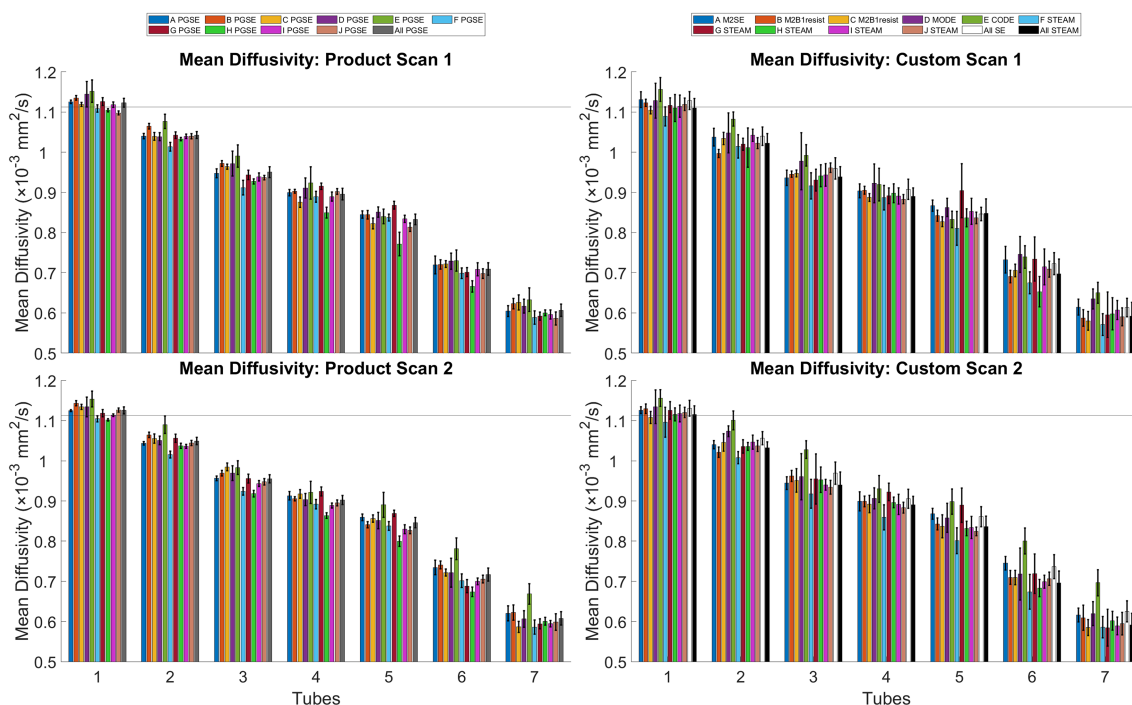
FIGURE 5 Average mean diffusivity (MD) across regions of interest (ROIs) as a function of polyvinylpyrrolidone (PVP) concentration. Tensors were reconstructed using b = (100, 450) s/mm^2 data. Tubes 1–7 corresponded to 0%, 2.5%, 5%, 7.5%, 10%, 15% and 20% PVP, respectively. Mean ± SD across voxels in the ROI. The horizontal black line indicates the reference diffusivity (H$_2$O at 0°C) = 1.113 x 10^{-3} mm^2/s

3.6 | Intrasite repeatability

Bland–Altman plots of MD and FA across two scans are presented in Figure 9. In tube 1, the mean differences of MD in PGSE, SE and STEAM were (0.3 ± 2.3, 0.24 ± 0.95, 0.52 ± 0.58) x 10^{-5} mm^2/s, respectively (mean ± 1.96 SD). The corresponding mean differences in FA were 0.0006 ± 0.0099, 0.006 ± 0.018 and −0.012 ± 0.039. The differences between pulse sequences, as observed in the mean differences of MD and FA, were not significant at $p = 0.05$. Linear regression of difference by average values of MD and FA showed low proportional bias, with R^2 of 0.16 or less in all cases. Figure 10 summarises the bootstrapped scan-wise mean differences in MD and FA, grouped by pulse sequence. On average and across tubes, the greatest median scan-wise difference in MD of 0.96 x 10^{-5} mm^2/s was seen in the SE data, whereas STEAM exhibited the greatest median scan-wise difference in FA of −0.012. Table S1 compares the mean difference averaged across all tubes across sites. The lowest absolute mean differences in MD and FA in the custom sequence data were seen in sites A and D, respectively.

Overall, STEAM yielded more accurate MD than SE at the isocentre ($p = 0.02$), while MD and FA across repetitions were more stable in SE compared with STEAM ($p < 0.001$). Although SE tended to higher reproducibility in MD and FA than STEAM, these were not significant at $p = 0.05$.

3.7 | Intersite reproducibility

The average CV$_{MD}$ was evaluated across scans, sites (PGSE, SE and STEAM separately) and sequence pairs (PGSE vs. SE, PGSE vs. STEAM and SE vs. STEAM). The results were: CV$_{MD(scans)}$ = 1.0%, CV$_{MD(sites, PGSE)}$ = 2.6%, CV$_{MD(sites, SE)}$ = 3.1%, CV$_{MD(sites, STEAM)}$ = 2.1%, CV$_{MD(sequences, PGSE vs SE)}$ = 1.2%, CV$_{MD(sequences, PGSE vs. STEAM)}$ = 1.6% and CV$_{MD(sequences, SE vs. STEAM)}$ = 2.6% (Figure 11). Within the intersite grouping, there were no significant differences in median CV between different sequences. Within the intersequence grouping, median CV in SE versus STEAM was significantly higher than both PGSE versus SE and PGSE versus STEAM.

4 | DISCUSSION

In this study, we compared several state-of-the-art methods suitable for in vivo cardiac DTI alongside conventional PGSE data. The results at the isocentre (i.e. tube 1) showed that, averaged across sites, PGSE, SE and STEAM yielded accurate MD, agreeing with the reference diffusivity of

TABLE 2 Mean diffusivity (MD) and fractional anisotropy (FA) across regions of interest in tubes 1–7 by sequence

Tube	[PVP] (%)	MD (x 10^{-3} mm²/s)				FA			
		PGSE	SE	STEAM	Literature	PGSE	SE	STEAM	Literature
1	0	1.124 ± 0.017	1.13 ± 0.017[†]	1.113 ± 0.011*	1.115 ± 0.033,[29] 1.123 ± 0.023,[30] 1.12 ± 0.01[34]	0.028 ± 0.016	0.052 ± 0.016*	0.059 ± 0.030*	0.037 ± 0.010[29]
2	2.5	1.046 ± 0.018	1.048 ± 0.030	1.027 ± 0.014*	1.05 ± 0.02[34]	0.036 ± 0.023	0.059 ± 0.023*	0.064 ± 0.031*	
3	5	0.953 ± 0.023	0.964 ± 0.028[†]	0.939 ± 0.015	0.98 ± 0.02[34]	0.049 ± 0.034	0.065 ± 0.032	0.069 ± 0.038	
4	7.5	0.899 ± 0.020	0.907 ± 0.014[†]	0.890 ± 0.016		0.042 ± 0.024	0.060 ± 0.021*	0.066 ± 0.023*	
5	10	0.839 ± 0.026	0.854 ± 0.022	0.842 ± 0.032	0.850 ± 0.024,[30] 0.85 ± 0.01[34]	0.049 ± 0.036	0.062 ± 0.034	0.090 ± 0.047*	
6	15	0.713 ± 0.025	0.729 ± 0.031[†]	0.697 ± 0.015	0.71 ± 0.01[34]	0.052 ± 0.032	0.078 ± 0.042	0.113 ± 0.066*	
7	20	0.607 ± 0.021	0.619 ± 0.035[†]	0.592 ± 0.010*	0.607 ± 0.019,[30] 0.59 ± 0.02[34]	0.048 ± 0.028	0.078 ± 0.032*	0.123 ± 0.067*	

Note: MD and FA given as mean ± SD across sites and time points. Values from the literature using PGSE at a single time point are given for matching PVP concentrations and temperature. Significant differences at $p < 0.05$ (i) in SE and STEAM with respect to PGSE are denoted by *, and (ii) in SE with respect to STEAM by [†].

Abbreviations: PGSE, pulsed gradient spin echo; PVP, polyvinylpyrrolidone; SE, spin echo; STEAM, stimulated echo acquisition mode.

water at 0°C to within 1.5% or less. Moreover, the MDs in the PGSE, SE and STEAM data were within ±1% of averaged literature values,[33,34,37] suggesting good conformance with prescribed experimental conditions. The diffusivity of PVP depends on the concentration, temperature and length of polymer chains, and an independently measured ground truth is unavailable. We therefore used the MD acquired using PGSE averaged across sites and scans as a reference for assessing accuracy in PVP with different concentrations: these values were in good agreement with the literature (Table 2).

The expected value of FA in an isotropic phantom is zero. However, factors including noise, imperfect gradient calibration, eddy currents, and convection and vibration within the phantom could in principle contribute to nonzero FA. Both SE and STEAM had lower SNRs than PGSE, which could have resulted in the higher FA in the custom sequences. We chose not to report CV_{FA} because the expected FA in isotropic media is zero, which leads to unstable CV_{FA} that is biased by the measured FA. For instance, poor gradient calibration would lead to a higher FA, and therefore artificially lower CV_{FA}. Vibration and convection were not deemed to be major issues due to the good accuracy of MD.

CV (ROI) in MD and FA across sites was lowest in PGSE and highest in STEAM. In an isotropic liquid phantom, the underlying substrate is expected to be homogeneous. Reasons that could increase CV (ROI) include noise and image artefacts. From Figure 3, Gibb's ringing can be seen in all datasets due to the hard edges of the phantom. Additional artefacts are visible in the STEAM data (sites G and H), potentially increasing the CV (ROI). Furthermore, as the tube number and PVP concentration increased, we observed that the mean FA and CV (ROI) in MD and FA were increasingly elevated, particularly in STEAM. This may be associated with (i) effects of concomitant fields away from the isocentre (site E), (ii) localised image artefacts in tubes 5, 6 and 7 possibly caused by vibration (site G), and (iii) the increased effect of Gibb's ringing in the tubes with higher PVP concentrations due to the increased differential in DW image intensities relative to the surrounding ice-water bath.

In cardiac DTI, DW data with finite, low b-values are often used instead of non-DW data to suppress the effects of microvascular perfusion.[23] There was no perfusion in the phantom, and therefore MD and FA would be expected to be relatively insensitive to the b-value combinations used, insofar as perfusion is concerned. This was generally the case for PGSE data. Greater deviations were seen when custom DTI data were reconstructed from b-value pairs that included non-DW data. See b_{low} in Table 1 for corresponding b-values. Deviations from the expected values of MD and FA were particularly striking in MODE and STEAM data. With MODE, high MD was seen whenever the b_{low} data were used. This was due to a shading artefact seen in the DW data, but not in the non-DW data, therefore resulting in a positive bias in MD when the non-DW data were included in the reconstruction. The shading artefact was potentially due to residual concomitant gradient field effects, and this was reflected in higher regional heterogeneity compared with other SE data. In STEAM, the value of b_{low} can be substantial due to the effect of crushers and long diffusion times. The diffusion contrast between b_{low} and $b_{high} = 100$ s/mm² can therefore be insufficient for reliable estimation of MD and FA. This is a general result of the increased effects of noise as the range of b-values used for reconstruction are reduced, for instance, resulting in a positive bias in FA. As b_{high} increased to 450 s/mm², the errors were reduced. Where combinations of b = (100, 300), (100, 450) and (300, 450)
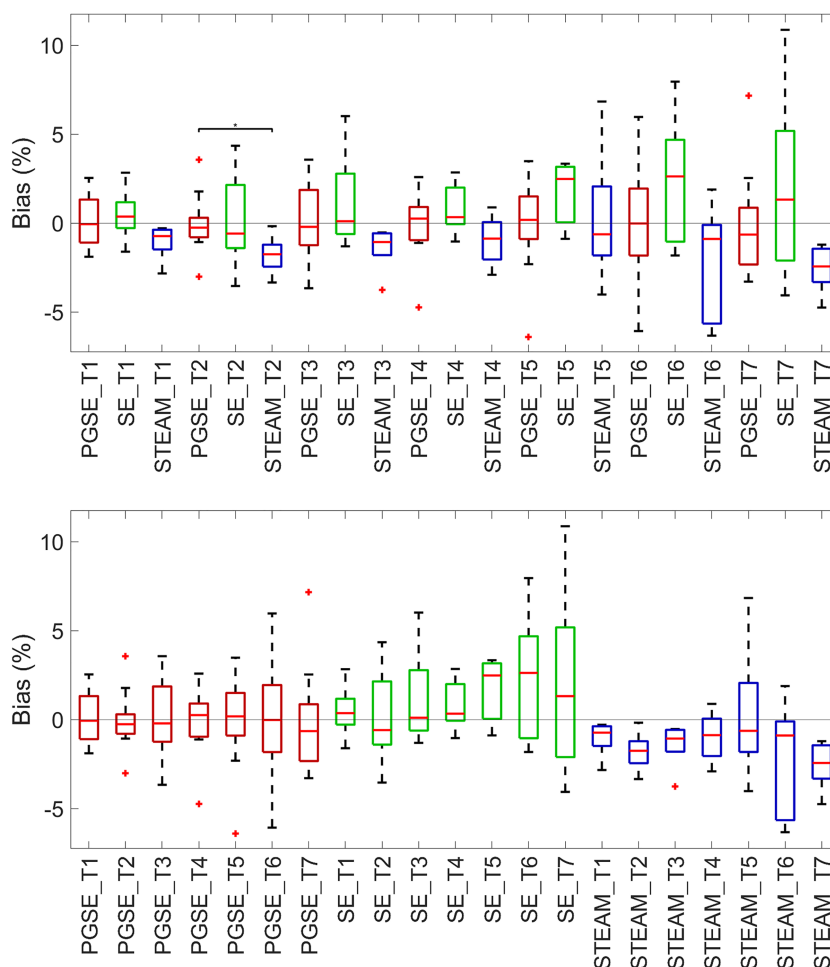
FIGURE 6 Accuracy in mean diffusivity (MD) measurements. Accuracy of MD relative to the reference MD was reported across pulse sequences and tubes (T). The reference MD in each tube was determined by the mean MD across sites and scans. Data from scans 1 and 2 were first averaged. Wilcoxon rank-sum tests were performed between pairs of sequences (pulsed gradient spin echo [PGSE] vs. spin echo [SE]; SE vs. stimulated echo acquisition mode [STEAM]; PGSE vs. STEAM) for each separate tube, and between pairs of tubes for each separate sequence; *$p < 0.05$). Differences in median accuracy between tubes within the same sequence were not significant at $p < 0.05$. Data were sorted by sequences-tubes (top) and tubes-sequences (bottom) for clarity

s/mm^2 were used, the MD and FA were qualitatively indistinguishable. However, we found that the combination of b = (100, 450) s/mm^2 yielded the best reproducibility in MD between scans 1 and 2, and this was consistent with typical b-values used in the literature.[19,22]

The SNRs in the b = 100 s/mm^2 images at the isocentre were comparable between SE and STEAM. The SNR generally increased with PVP concentration as the mean signal intensity was less attenuated by diffusion-weighting. Away from the isocentre in tube 7, the image SNRs in PGSE and SE were 41% and 34% higher than that in STEAM, which may have arisen from image artefacts as described (sites G and H). Furthermore, in four instances of STEAM, only a single slice was acquired due to limitations in implementation. In these cases, reduced FOV was used, which involved excitation slice-selection gradients that were orthogonal to subsequent slice-selection gradients. As a result, acquisition of multiple slices while maintaining constant acquisition time would necessitate a reduction in the number of repetitions proportional to the number of slices (site H). By implementing a slab-tilted reduced FOV imaging method that is compatible with multislice imaging, this enabled site H to extend the TR to 6 s, rather than the 2 s used by the other STEAM sites, which recovers some of the SNR efficiency lost by using the STEAM technique. This motivates a shift from standard slab-perpendicular methods towards better methods of reduced FOV imaging, such as slab-tilted methods. We note that image SNR does not fully reflect the stability of measurements, as it does not take into account higher b-value data and associated potential issues. For instance, among the STEAM acquisitions, lower SNR was found alongside lower CV in MD and FA (site H).

The mean differences in MD in tube 1 across scans were relatively small at 0.22%, 0.21% and 0.47% in PGSE, SE and STEAM, respectively, expressed as a percentage of the respective average MD. The 95% limits of agreement were higher at ±0.84% in SE compared with ±0.52% in the
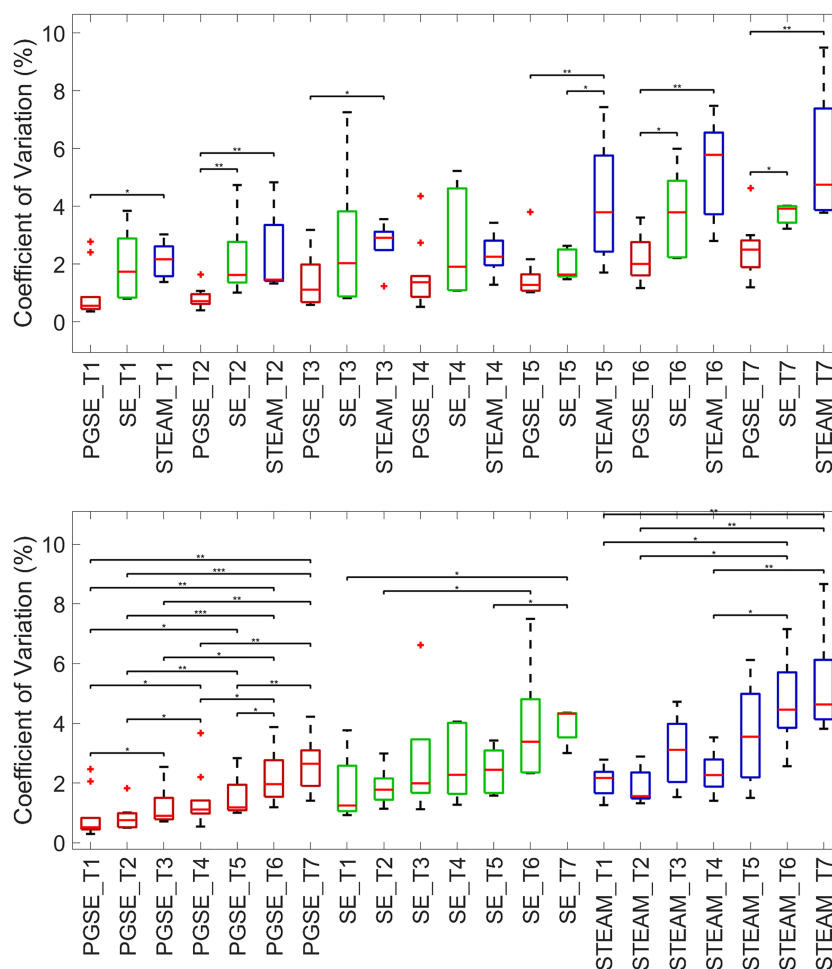
FIGURE 7 Precision in mean diffusivity (MD) measurements. Precision of MD was reported across pulse sequences and tubes (T). Wilcoxon rank-sum tests were performed between pairs of sequences and tubes; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. In several tubes, precision was poorer in spin echo (SE) and stimulated echo acquisition mode (STEAM) compared with pulsed gradient spin echo (PGSE). There was a clear dependence of precision on tube number. Data were sorted by sequences-tubes (top) and tubes-sequences (bottom) for clarity

STEAM data. For FA, both the absolute mean difference and 95% limits of agreement were lower in SE (0.006 ± 0.018) compared with STEAM (-0.012 ± 0.039). The MD data suggest that there was good reproducibility across the different classes of sequences. That the reproducibility with respect to FA was poorer could be due to the intrinsically low FA and lack of underlying microstructure.

Intrasite repeatability was superior to intersite and intersequence reproducibility, as measured in terms of CV_{MD}. This is consistent with the literature. Intersite reproducibility ranging from 2.1% (STEAM) to 3.1% (SE) was similar to other studies using PGSE alone,[33,34] where values ranged from 2.1% to 3%. We also observed that intersite reproducibility was marginally poorer than intersequence reproducibility. The CV in all cases was low compared with CV in vivo,[29] as the phantoms were static, isotropic and consistently fabricated. We would anticipate greater variation between sites and sequences, with added phantom complexity simulating anisotropy and/or motion.

While we consider the various SE sequences collectively, it is worth noting that each of the four SE sequences used have their specific characteristics. Despite having the longest TE, the M2B1resist sequence had the lowest FA, regional heterogeneity in MD and FA, and SD_{MD} and SD_{FA} across repetitions, and highest SNR among custom sequences. Owing to the symmetric design of the diffusion gradients, the sequence, along with M2SE, is inherently more robust to concomitant fields. Furthermore, as the zeroth and first gradient moments are nulled prior to the refocusing pulse, its insensitivity to motion with constant velocity is less dependent on the refocusing pulse being close to 180° compared with other custom SE sequences.[46] However, the additional gradient lobes relative to M2SE, and the need for crushers due to the nulling of zeroth gradient moments before the refocusing pulse, extend the minimum TE. M2SE and MODE yielded the best interscan reproducibility in MD and FA, respectively. While shorter TE are feasible with CODE and MODE compared with M2SE and M2B1resist, the former two sequences are sensitive to spatially varying concomitant fields due to the asymmetry of diffusion gradient waveforms about the refocusing pulse, and corrections are
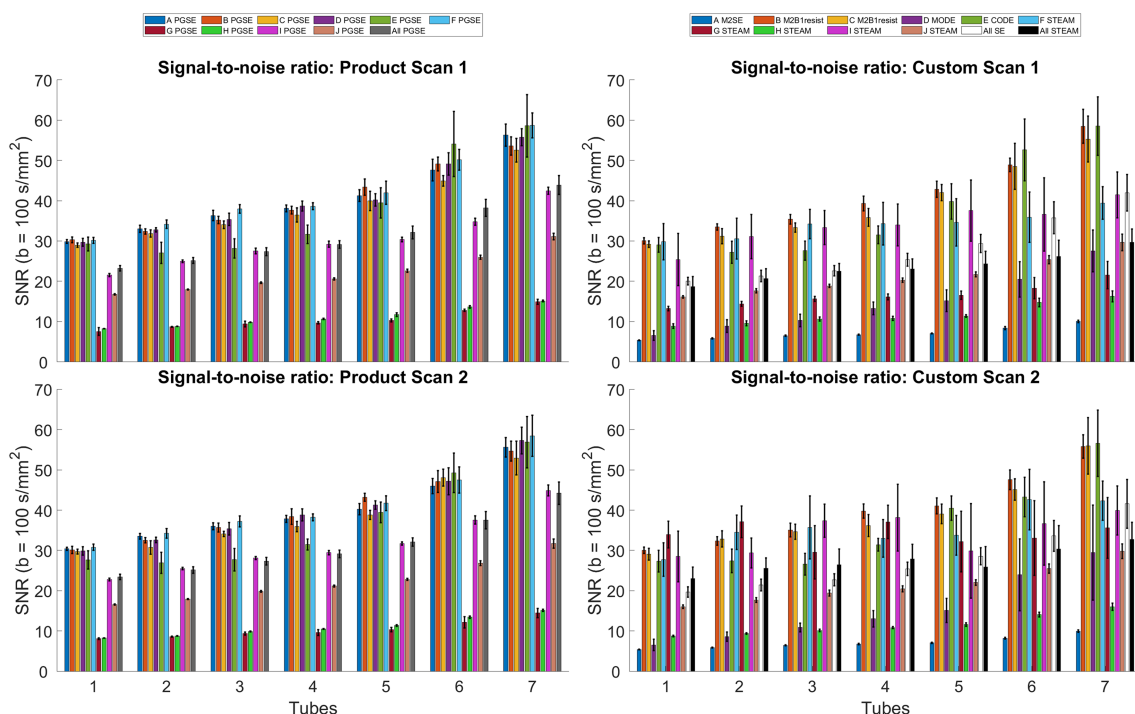
FIGURE 8 Signal-to-noise ratio (SNR) of $b = 100$ s/mm^2 images. SNR was highest in the pulsed gradient spin echo (PGSE) data: 23.3 and 44.1 in tubes 1 and 7, respectively. Relative to PGSE, SNRs (spin echo [SE]) in tubes 1 and 7 were 85.1% and 94.8%, respectively, while SNRs (stimulated echo acquisition mode [STEAM]) were 89.4% and 70.8%, respectively. Mean ± SD across voxels in the region of interest

employed to mitigate image artefacts arising from concomitant fields. As the custom SE group data include data acquired with four different SE sequences, the within-group variation would be expected to decrease were a single sequence used.

The current work provides a promising basis for future clinical studies. While other quantitative CMR techniques have arguably been hampered by methods that differ even in phantoms and are very sensitive to readout parameters,[47] here we demonstrate that a wide range of sequences from different centres have excellent agreement in a simple phantom. Clinically, the availability of multiple sequences for cardiac DTI offers sites the flexibility to choose the sequence to suit the pathology of interest and hardware configuration available. SE, for instance, yields higher SNR efficiency but requires higher performance gradient systems. STEAM, on the other hand, is able to acquire data over a wider range of cardiac phases, but requires breath-holding, which may be difficult in some patient cohorts. While the comparison of cardiac DTI sequences remains an active area of research, there may be a clinical case for both classes of sequences.

There are several limitations of using a static isotropic phantom for quantifying the performance of cardiac DTI sequences. First, the phantom lacks motion. This prevents the assessment of the quality of motion compensation in the custom sequences, where inadequate motion compensation is a primary reason for rejected images, elevated MD and failed scans. Incorrect triggering could be difficult to identify from the static phantom images. Second, the phantom lacks anisotropic microstructure, leading to an expected FA of zero. This results in non-Gaussian distribution of errors in FA, and enhancement in the errors when expressed as a percentage. Third, the phantom substrate is liquid and therefore could be prone to scanner vibration and thermal convection, which in turn could increase MD, although this was not observed to be a major issue. This effect may be more pronounced in STEAM, where diffusion times are an order of magnitude longer. Fourth, the T_1 and T_2 relaxation times of the liquid substrate are different compared with that found in the heart,[37,48] with $T_{1,H2O@0°C} = 1525$ ms, $T_{2,H2O@0°C} = 1472$ ms, $T_{1,20\%PVP@0°C} = 753$ ms, $T_{2,20\%PVP@0°C} = 623$ ms, $T_{1,Heart\ in\ vivo} = 1184$ ms and $T_{2,Heart\ in\ vivo} = 52$ ms, as measured at 3 T. This translates to higher SNR in the phantom relative to heart, and sequences with longer TE such as custom SE, would suffer greater SNR penalties in vivo. Adjusting for T_1 and T_2 in the in vivo setting relative to iced water, the signal in SE would be expected to decrease by 73% to 82%, while the signal in the STEAM data would decrease by 29% to 49% based on TR and TE in the custom sequences. This would lead to a theoretical 2.7x increase in signal in STEAM relative to SE in tube 1. In 20% PVP, the relative theoretical improvement in signal in STEAM over SE would be 2.4x.

Despite its shortcomings, the use of a static, isotropic phantom is an important first step for assessing custom sequence performance in a multicentre study, as it allows a baseline assessment of parameters in the absence of additional variables associated with tissue microstructure and motion. The phantom was robust and consistent, having been manufactured centrally, and stable over time, permitting repeated scanning with negligible change in substrate. The use of the ice-water bath permitted accurate temperature control, which is important, as
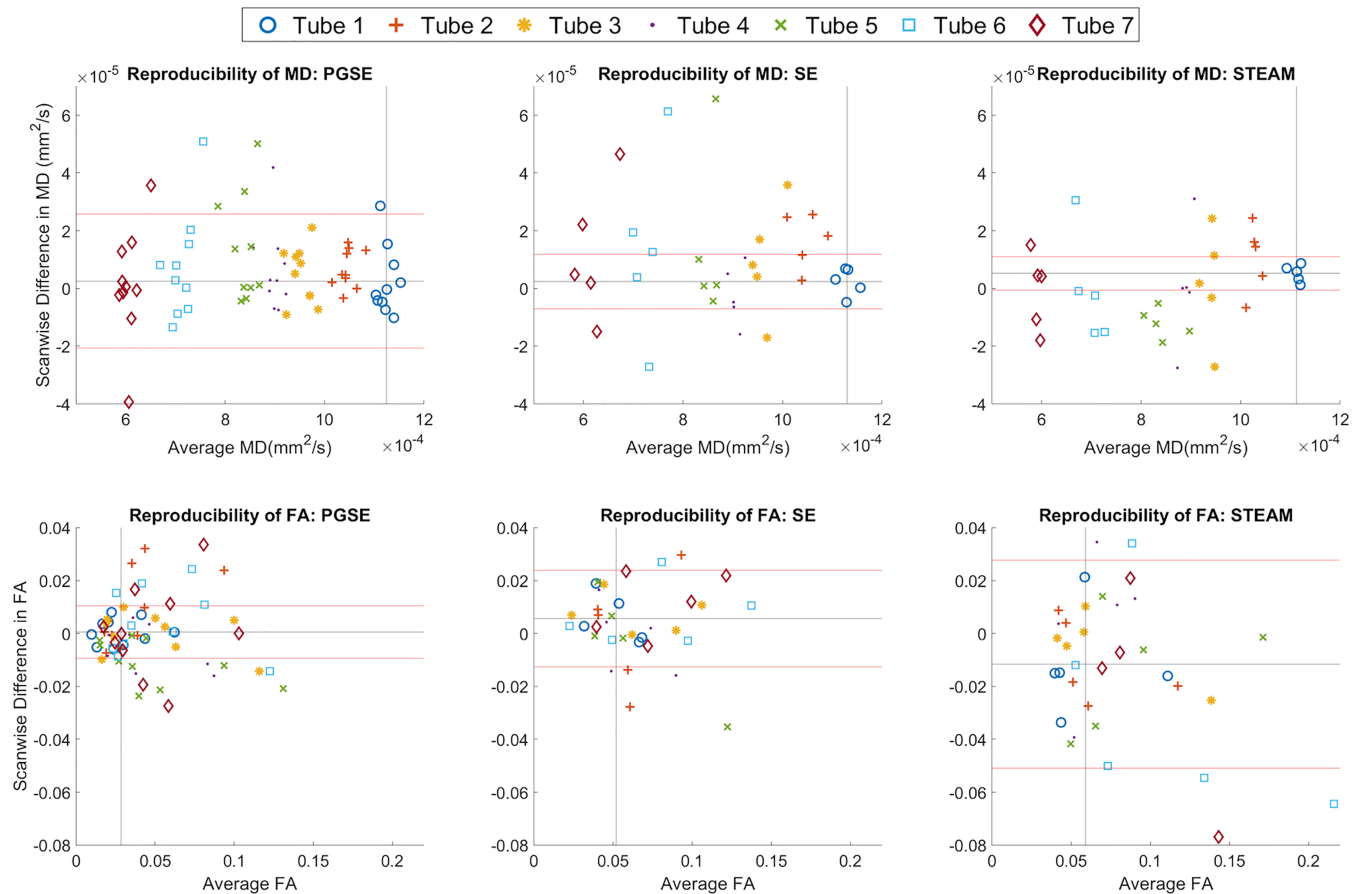
FIGURE 9 Bland–Altman plots of mean diffusivity (MD) and fractional anisotropy (FA) across two scans (scan 2 – scan 1). Data acquired using pulsed gradient spin echo (PGSE; left), spin echo (SE; middle) and stimulated echo acquisition mode (STEAM; right) sequences in tubes 1 to 7 (0%–20% polyvinylpyrrolidone [PVP]) are colour-coded by increasing PVP concentration. The black and red horizontal lines indicate the mean and ±1.96 SD values for tube 1, respectively. The black vertical line indicates the average values for tube 1 across sites
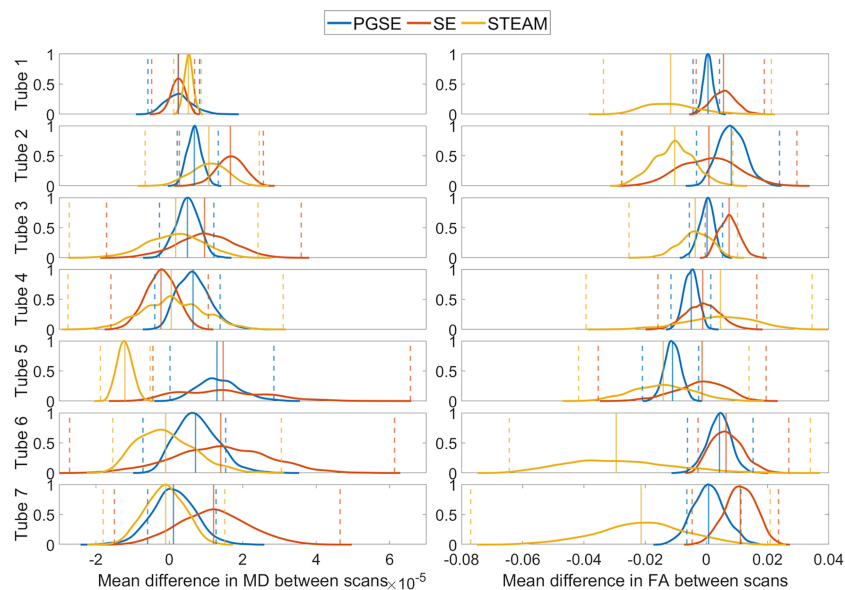
FIGURE 10 Bootstrapped normalised histograms of mean differences in mean diffusivity (MD) and fractional anisotropy (FA) between scans. Median and 95% confidence intervals (CIs) are given by vertical solid and dashed lines. The MD difference in stimulated echo acquisition mode (STEAM) was significantly lower than that of pulsed gradient spin echo (PGSE) and spin echo (SE) in tube 5. All other differences in reproducibility between sequences were not significant
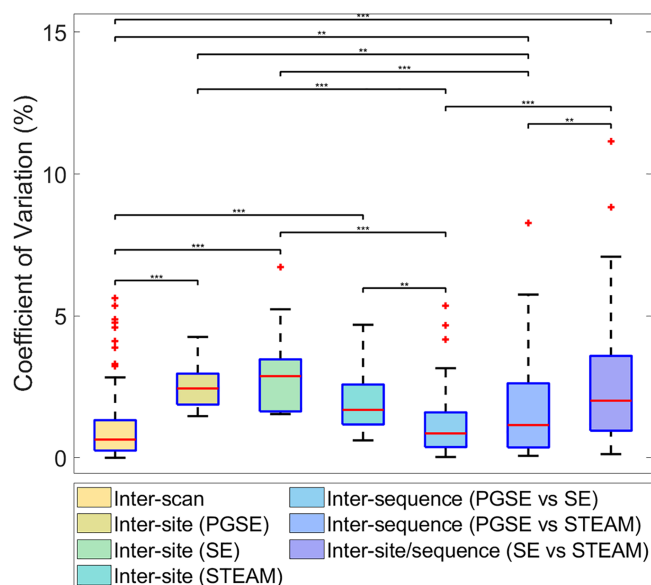
FIGURE 11 Reproducibility of mean diffusivity (MD) across scans, sites and sequences. Coefficients of variation in MD across scans, sites and sequences. Wilcoxon rank-sum tests were performed pairwise between groups; $**p < 0.01$, $***p < 0.001$. PGSE, pulsed gradient spin echo; SE, spin echo; STEAM, stimulated echo acquisition mode

every $1°C$ increase in temperature at $0°C$ leads to an $\sim5\%$ increase in diffusivity. More elaborate systems for monitoring and maintaining accurate temperature are feasible, but would be costly to implement. In this ice-water phantom, the maximum MD achievable was $\sim1.1 \times 10^{-3}$ mm^2/s in water. This was in the middle of the range of MDs measured in the heart in vivo with SE and STEAM (0.87×10^{-3} to 1.72×10^{-3} mm^2/s), and attributable to the low temperature. To simulate higher MD, a higher temperature would be needed. At $35°C$, water would have too high an MD (3.0×10^{-3} mm^2/s). Instead, PVP solutions with concentrations (20% to 40%) would cover an appropriate MD range (0.9×10^{-3} to 1.8×10^{-3} mm^2/s), based on a previous study.[49] Crucially, there exist gold standard, independent reference measurements of MD in water over a range of temperatures, and corroborating values of MD in temperature-controlled PVP phantoms in the literature.

5 | CONCLUSIONS

In summary, we have validated the accuracy, precision, repeatability and reproducibility of state-of-the-art custom sequences for in vivo cardiac DTI. This study benchmarks the performance of custom SE and STEAM against product PGSE sequences, and identifies baseline variation across sites, scanners and sequences. Some areas of inconsistency in the measurements have been highlighted, which warrant further methodological refinement. Future work includes the development of more sophisticated phantoms with more physiological relaxation times, diffusion anisotropy and motion characteristics. This would add further insight into the robustness and behaviour of cardiac DTI sequences, and the availability of appropriate phantoms will be key to facilitating quality assurance protocols. Quantitative validation of pulse sequences in phantoms represents an important step towards rationalising pulse sequences, and will contribute to protocol harmonisation and the establishment of cardiac DTI in the clinical setting.

CONFLICT OF INTEREST

Royal Brompton Hospital has research collaboration agreements with Siemens AG Medical Solutions.

AUTHOR CONTRIBUTIONS

DATA AVAILABILITY STATEMENT

ORCID

Irvin Teh https://orcid.org/0000-0002-6705-3129

William A. Romero R. https://orcid.org/0000-0001-7312-6745

Jaume Coll-Font https://orcid.org/0000-0001-9341-6838

Erica Dall'Armellina https://orcid.org/0000-0002-2165-7154

Daniel B. Ennis https://orcid.org/0000-0001-7435-1311

Pedro F. Ferreira https://orcid.org/0000-0002-0436-3496

Sebastian Kozerke https://orcid.org/0000-0003-3725-8884

David Lohr https://orcid.org/0000-0002-6509-3776

Kévin Moulin https://orcid.org/0000-0002-9188-6403

Christopher Nguyen https://orcid.org/0000-0003-1475-2329

Laura M. Schreiber https://orcid.org/0000-0002-8827-1838

Andrew D. Scott https://orcid.org/0000-0001-7656-3123

David E. Sosnovik https://orcid.org/0000-0002-6486-7434

Christian T. Stoeck https://orcid.org/0000-0001-8670-0929

Cyril Tous https://orcid.org/0000-0003-2664-5272

Elizabeth M. Tunnicliffe https://orcid.org/0000-0002-6945-5201

Andreas M. Weng https://orcid.org/0000-0001-8943-3539

Pierre Croisille https://orcid.org/0000-0003-4019-3460

Magalie Viallon https://orcid.org/0000-0001-9118-0438

Jürgen E. Schneider https://orcid.org/0000-0003-0999-5684

REFERENCES

1. Nguyen C, Fan Z, Xie Y, et al. In vivo contrast free chronic myocardial infarction characterization using diffusion-weighted cardiovascular magnetic resonance. *J Cardiovasc Magn Reson.* 2014;16:68.
2. Mekkaoui C, Jackowski MP, Kostis WJ, et al. Myocardial scar delineation using diffusion tensor magnetic resonance tractography. *J Am Heart Assoc.* 2018;7(3):e007834.
3. Ariga R, Tunnicliffe EM, Manohar SG, et al. Identification of myocardial disarray in patients with hypertrophic cardiomyopathy and ventricular arrhythmias. *J Am Coll Cardiol.* 2019;73(20):2493-2502.
4. Wu MT, Tseng WY, Su MY, et al. Diffusion tensor magnetic resonance imaging mapping the fiber architecture remodeling in human myocardium after infarction: correlation with viability and wall motion. *Circulation.* 2006;114(10):1036-1045.
5. Moulin K, Viallon M, Romero W, et al. MRI of reperfused acute myocardial infarction edema: ADC quantification versus T1 and T2 mapping. *Radiology.* 2020;295(3):542-549.
6. Tseng WYI, Dou JG, Reese TG, Wedeen VJ. Imaging myocardial fiber disarray and intramural strain hypokinesis in hypertrophic cardiomyopathy with MRI. *J Magn Reson Imaging.* 2006;23(1):1-8.
7. Das A, Chowdhary A, Kelly C, et al. Insight into myocardial microstructure of athletes and hypertrophic cardiomyopathy patients using diffusion tensor imaging. *J Magn Reson Imaging.* 2020;53(1):73-82.
8. Nguyen C, Lu MJ, Fan ZY, et al. Contrast-free detection of myocardial fibrosis in hypertrophic cardiomyopathy patients with diffusion-weighted cardiovascular magnetic resonance. *J Cardiovasc Magn Reson.* 2015;17(1):107.
9. Gotschy A, von Deuster C, van Gorkum RJH, et al. Characterizing cardiac involvement in amyloidosis using cardiovascular magnetic resonance diffusion tensor imaging. *J Cardiovasc Magn Reson.* 2019;21(1):56.
10. Abdullah OM, Drakos SG, Diakos NA, et al. Characterization of diffuse fibrosis in the failing human heart via diffusion tensor imaging and quantitative histological validation. *NMR Biomed.* 2014;27(11):1378-1386.
11. von Deuster C, Sammut E, Asner L, et al. Studying dynamic myofiber aggregate reorientation in dilated cardiomyopathy using in vivo magnetic resonance diffusion tensor imaging. *Circ Cardiovasc Imaging.* 2016;9(10):e005018.
12. Edelman RR, Gaa J, Wedeen VJ, et al. In vivo measurement of water diffusion in the human heart. *Magn Reson Med.* 1994;32(3):423-428.

13. Stejskal EO, Tanner JE. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J Chem Phys.* 1965;42(1):288-292.

14. Gamper U, Boesiger P, Kozerke S. Diffusion imaging of the in vivo heart using spin echoes--considerations on bulk motion sensitivity. *Magn Reson Med.* 2007;57(2):331-337.

15. Stoeck CT, von Deuster C, Genet M, Atkinson D, Kozerke S. Second-order motion-compensated spin echo diffusion tensor imaging of the human heart. *Magn Reson Med.* 2016;75(4):1669-1676.

16. Ma S, Nguyen CT, Christodoulou AG, et al. Accelerated cardiac diffusion tensor imaging using joint low-rank and sparsity constraints. *IEEE Trans Biomed Eng.* 2018;65(10):2219-2230.

17. Aliotta E, Wu HH, Ennis DB. Convex optimized diffusion encoding (CODE) gradient waveforms for minimum echo time and bulk motion-compensated diffusion-weighted MRI. *Magn Reson Med.* 2016;77(2):717-729.

18. Majeed W, Kalra P, Kolipaka A. Motion Compensated, Optimized Diffusion Encoding (MODE) Gradient Waveforms. In: *Paper presented at: Proceedings of the 26th Annual Meeting of ISMRM, Paris, France*; 2018.

19. von Deuster C, Stoeck CT, Genet M, Atkinson D, Kozerke S. Spin echo versus stimulated echo diffusion tensor imaging of the in vivo human heart. *Magn Reson Med.* 2016;76(3):862-872.

20. Scott AD, Nielles-Vallespin S, Ferreira PF, et al. An in-vivo comparison of stimulated-echo and motion compensated spin-echo sequences for 3 T diffusion tensor cardiovascular magnetic resonance at multiple cardiac phases. *J Cardiovasc Magn Reson.* 2018;20(1):1.

21. McClymont D, Teh I, Carruth E, et al. Evaluation of non-Gaussian diffusion in cardiac MRI. *Magn Reson Med.* 2017;78(3):1174-1186.

22. Lasic S, Szczepankiewicz F, Dall'Armellina E, et al. Motion-compensated b-tensor encoding for in vivo cardiac diffusion-weighted imaging. *NMR Biomed.* 2020;33(2):e4213.

23. Scott AD, Ferreira PF, Nielles-Vallespin S, et al. Optimal diffusion weighting for in vivo cardiac diffusion tensor imaging. *Magn Reson Med.* 2015;74(2):420-430.

24. Jones DK, Basser PJ. "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. *Magn Reson Med.* 2004;52(5):979-993.

25. Gahm JK, Kindlmann G, Ennis DB. The effects of noise over the complete space of diffusion tensor shape. *Med Image Anal.* 2014;18(1):197-210.

26. McClymont D, Teh I, Schneider JE. The impact of signal-to-noise ratio, diffusion-weighted directions and image resolution in cardiac diffusion tensor imaging - insights from the ex-vivo rat heart. *J Cardiovasc Magn Reson.* 2017;19(1):90.

27. Stoeck CT, von Deuster C, van Gorkum RJH, Kozerke S. Motion and eddy current-induced signal dephasing in in vivo cardiac DTI. *Magn Reson Med.* 2020;84(1):277-288.

28. Nielles-Vallespin S, Mekkaoui C, Gatehouse P, et al. In vivo diffusion tensor MRI of the human heart: reproducibility of breath-hold and navigator-based approaches. *Magn Reson Med.* 2013;70(2):454-465.

29. Tunnicliffe EM, Scott AD, Ferreira P, et al. Intercentre reproducibility of cardiac apparent diffusion coefficient and fractional anisotropy in healthy volunteers. *J Cardiovasc Magn Reson.* 2014;16(1):31.

30. Moulin K, Croisille P, Feiweier T, et al. In vivo free-breathing DTI and IVIM of the whole human heart using a real-time slice-followed SE-EPI navigator-based sequence: A reproducibility study in healthy volunteers. *Magn Reson Med.* 2016;76(1):70-82.

31. Stoeck CT, Kalinowska A, von Deuster C, et al. Dual-phase cardiac diffusion tensor imaging with strain correction. *PLoS ONE.* 2014;9(9):e107159.

32. McGill LA, Ismail TF, Nielles-Vallespin S, et al. Reproducibility of in-vivo diffusion tensor cardiovascular magnetic resonance in hypertrophic cardiomyopathy. *J Cardiovasc Magn Reson.* 2012;14:86.

33. Grech-Sollars M, Hales PW, Miyazaki K, et al. Multi-centre reproducibility of diffusion MRI parameters for clinical sequences in the brain. *NMR Biomed.* 2015;28(4):468-485.

34. Palacios EM, Martin AJ, Boss MA, et al. Toward precision and reproducibility of diffusion tensor imaging: a multicenter diffusion phantom and traveling volunteer study. *Am J Neuroradiol.* 2017;38(3):537-545.

35. Prohl AK, Scherrer B, Tomas-Fernandez X, et al. Reproducibility of structural and diffusion tensor imaging in the TACERN multi-center study. *Front Integr Neurosci.* 2019;13:24

36. Pierpaoli C, Sarlls J, Nevo U, Basser PJ, Horkay F. Polyvinylpyrrolidone (PVP) water solutions as isotropic phantoms for diffusion MRI studies. In: Proceedings of the 17th Annual Meeting of ISMRM, Honolulu, USA; 2009.

37. Jerome NP, Papoutsaki MV, Orton MR, et al. Development of a temperature-controlled phantom for magnetic resonance quality assurance of diffusion, dynamic, and relaxometry measurements. *Med Phys.* 2016;43(6):2998-3007.

38. Teh I, Zhou FL, Cristinacce PLH, Parker GJM, Schneider JE. Biomimetic phantom for cardiac diffusion MRI. *J Magn Reson Imaging.* 2016;43(3):594-600.

39. Szczepankiewicz F, Hoge S, Westin CF. Linear, planar and spherical tensor-valued diffusion MRI data by free waveform encoding in healthy brain, water, oil and liquid crystals. *Data Brief.* 2019;25:104208.

40. Romero W, Viallon M, Froeling M, et al. CMRDiffTools: A Processing and Analysis Tool for Cardiac Diffusion MR images. In: *Paper presented at: Proc Int Soc Magn Reson Med.* Paris, France; 2018.

41. CREATIS The Human Heart Project, 2020. Accessed January 17, 2022. https://humanheart-project.creatis.insa-lyon.fr/

42. Mills R. Self-diffusion in normal and heavy-water in range 1-45 degrees. *J Phys Chem.* 1973;77(5):685-688.

43. Dietrich O, Raya JG, Reeder SB, Reiser MF, Schoenberg SO. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. *J Magn Reson Imaging.* 2007;26(2):375-385.

44. Zhu T, Hu R, Qiu X, et al. Quantification of accuracy and precision of multi-center DTI measurements: a diffusion phantom and human brain study. *NeuroImage.* 2011;56(3):1398-1411.

45. Farrell JA, Landman BA, Jones CK, et al. Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. *J Magn Reson Imaging.* 2007;26(3):756-767.

46. Nguyen C, Fan Z, Xie Y, et al. In vivo diffusion-tensor MRI of the human heart on a 3 tesla clinical scanner: An optimized second order (M2) motion compensated diffusion-preparation approach. *Magn Reson Med.* 2016;76(5):1354-1363.

47. Kellman P, Hansen MS. T1-mapping in the heart: accuracy and precision. *J Cardiovasc Magn Reson.* 2014;16:2.

48. Granitz M, Motloch LJ, Granitz C, et al. Comparison of native myocardial T1 and T2 mapping at 1.5T and 3T in healthy volunteers: reference values and clinical implications. *Wien Klin Wochenschr.* 2019;131(7–8):143-155.

49. Wagner F, Laun FB, Kuder TA, et al. Temperature and concentration calibration of aqueous polyvinylpyrrolidone (PVP) solutions for isotropic diffusion MRI phantoms. *PLoS ONE.* 2017;12(6):e0179276.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.