



**HAL**  
open science

## Efficiency and Accuracy Evaluation of Multiple Diffusion-Weighted MRI Techniques Across Different Scanners

Frederik Crop, Clémence Robert, Romain Viard, Julien Dumont, Marine Kawalko, Pauline Makala, Xavier Liem, Imen El Aoud, Aicha Ben Miled, Victor Chaton, et al.

► **To cite this version:**

Frederik Crop, Clémence Robert, Romain Viard, Julien Dumont, Marine Kawalko, et al.. Efficiency and Accuracy Evaluation of Multiple Diffusion-Weighted MRI Techniques Across Different Scanners. *Journal of Magnetic Resonance Imaging*, 2024, 59 (1), pp.311-322. 10.1002/jmri.28869 . hal-04155928

**HAL Id: hal-04155928**

**<https://hal.science/hal-04155928v1>**

Submitted on 7 Jul 2023



**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Efficiency and Accuracy Evaluation of Multiple Diffusion-Weighted MRI Techniques Across Different Scanners

Frederik Crop, PhD,<sup>1,2\*</sup>  Clémence Robert, MS,<sup>1</sup> Romain Viard, PhD,<sup>3,4</sup> Julien Dumont, RT,<sup>3</sup> Marine Kawalko, RT,<sup>5</sup> Pauline Makala, RT,<sup>6</sup> Xavier Liem, MD,<sup>6</sup> Imen El Aoud, MD,<sup>5</sup> Aicha Ben Miled, MD,<sup>5</sup> Victor Chaton, MD,<sup>5</sup> Lucas Patin, MD,<sup>5</sup> David Pasquier, MD,<sup>6,7</sup>  Ophélie Guillaud, MD,<sup>5</sup> Benjamin Vandendorpe, MD,<sup>6</sup> Xavier Mirabel, MD,<sup>6</sup> Luc Ceugnart, MD,<sup>5</sup> Camille Decoene, PhD,<sup>1</sup> and Thomas Lacornerie, MS<sup>1</sup>

**Background:** The choice between different diffusion-weighted imaging (DWI) techniques is difficult as each comes with tradeoffs for efficient clinical routine imaging and apparent diffusion coefficient (ADC) accuracy.

**Purpose:** To quantify signal-to-noise-ratio (SNR) efficiency, ADC accuracy, artifacts, and distortions for different DWI acquisition techniques, coils, and scanners.

**Study Type:** Phantom, in vivo intraindividual biomarker accuracy between DWI techniques and independent ratings.

**Population/Phantoms:** NIST diffusion phantom.

51 Patients: 40 with prostate cancer and 11 with head-and-neck cancer at 1.5 T

**Field Strength/Sequence:** Echo planar imaging (EPI): 1.5 T and 3 T Siemens; 3 T Philips.

Distortion-reducing: RESOLVE (1.5 and 3 T Siemens); Turbo Spin Echo (TSE)-SPLICE (3 T Philips).

Small field-of-view (FOV): ZoomitPro (1.5 T Siemens); IRIS (3 T Philips).

Head-and-neck and flexible coils.

**Assessment:** SNR efficiency, geometrical distortions, and susceptibility artifacts were quantified for different b-values in a phantom. ADC accuracy/agreement was quantified in phantom and for 51 patients. In vivo image quality was independently rated by four experts.

**Statistical Tests:** QIBA methodology for accuracy: trueness, repeatability, reproducibility, Bland-Altman 95% Limits-of-Agreement (LOA) for ADC. Wilcoxon Signed-Rank and student tests on  $P < 0.05$  level.

**Results:** The ZoomitPro small FOV sequence improved b-image efficiency by 8%–14%, reduced artifacts and observer scoring for most raters at the cost of smaller FOV compared to EPI. The TSE-SPLICE technique reduced artifacts almost completely at a 24% efficiency cost compared to EPI for b-values  $\leq 500$  sec/mm<sup>2</sup>. Phantom ADC 95% LOA trueness were within  $\pm 0.03 \times 10^{-3}$  mm<sup>2</sup>/sec except for small FOV IRIS. The in vivo ADC agreement between techniques, however, resulted in 95% LOAs in the order of  $\pm 0.3 \times 10^{-3}$  mm<sup>2</sup>/sec with up to  $0.2 \times 10^{-3}$  mm<sup>2</sup>/sec of bias.

**Data Conclusion:** ZoomitPro for Siemens and TSE SPLICE for Philips resulted in a trade-off between efficiency and artifacts. Phantom ADC quality control largely underestimated in vivo accuracy: significant ADC bias and variability was found between techniques in vivo.

**Level of Evidence:** 3

**Technical Efficacy Stage:** 2

J. MAGN. RESON. IMAGING 2023.

View this article online at [wileyonlinelibrary.com](http://wileyonlinelibrary.com). DOI: 10.1002/jmri.28869

Received Sep 25, 2022, Accepted for publication May 23, 2023.

\*Address reprint requests to: F.C., Medical Physics, Centre Oscar Lambret, 3, Rue Frédéric Combemale, 59000 Lille, France. E-mail: [f-crop@o-lambret.fr](mailto:f-crop@o-lambret.fr)

Contract grant sponsor: Cancéropôle Nord-Ouest (CNO).

From the <sup>1</sup>Department of Medical Physics, Centre Oscar Lambret, Lille, France; <sup>2</sup>University of Lille, IEMN, Lille, France; <sup>3</sup>University of Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, PLBS UAR 2014-US 41, Lille, France; <sup>4</sup>University of Lille, Inserm, CHU Lille, U1172-LiNCog-Lille Neuroscience & Cognition, Lille, France; <sup>5</sup>Department of Radiology, Centre Oscar Lambret, Lille, France; <sup>6</sup>Academic Department of Radiotherapy, Centre Oscar Lambret, Lille, France; and <sup>7</sup>University of Lille, Centre de recherche en informatique, Signal et automatique de Lille, Lille, France

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

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

Diffusion-weighted imaging (DWI) images the water diffusivity, which can characterize tissue cellularity. DWI is part of the standard prostate MRI protocol at diagnosis or recurrence after local treatment such as radiotherapy or high intensity focal ultrasound.<sup>1–3</sup> It has been used for macroscopic target volume definition for prostate radiotherapy or focal prostate treatment, but also more recently as a biomarker.<sup>4–6</sup> DWI has also shown promising results for therapy outcome prediction for head and neck cancer.<sup>7</sup> In the clinical setting, a compromise has to be found between image resolution, signal to noise ratio (SNR), acquisition time and distortions for diagnostic b-value images. However, this tradeoff can also have an effect on the accuracy of the quantitative apparent diffusion coefficient (ADC) imaging biomarker.<sup>8–12</sup> ADC calculation depends on multiple b-value images and thus depends on the SNR of each b-value image. An example of this optimization problem is the use of the number of averages/excitations for a low number of b-value images versus additional b-values with lower SNR for better ADC calculation.<sup>13</sup>

Current research for quantitative MRI biomarkers focuses on quantifying uncertainties through metrology, following, for example, Quantitative Imaging Biomarker Association (QIBA) and European Imaging Biomarkers Alliance (EIBALL) consortia and their recommendations.<sup>9,12,14</sup>

Different DWI sequences exist. The first, most employed, is single-shot echo planar imaging (EPI), which is efficient but can lead to distortions and artifacts. The second class of DWI sequences aims to reduce artifacts and distortions through segmentation of k-space (readout segmentation of long variable echo trains, RESOLVE)<sup>15</sup> or through split-echo acquisition of turbo spin echo (TSE-SPLICE).<sup>16</sup> A third, small field-of-view (FOV), type applies a combination of a tilt in the excitation plane with variable gradient amplitudes to reduce EPI artifacts and distortions at the cost of smaller FOV.<sup>17</sup> Liney et al.<sup>18</sup> investigated geometric accuracy for prostate radiotherapy planning in a radiotherapy setting. They found that the RESOLVE sequence provided the best in vivo agreement. Other authors also investigated in vivo differences between EPI and small FOV for thyroid<sup>19</sup> or prostate,<sup>20</sup> which were in favor of small FOV DWI. But combining results for both diagnostics and ADC accuracy proves to be difficult.

The aim of this work was to investigate the quantification of SNR efficiency, ADC accuracy, artifacts, and distortions between sequence types, coils, and MRI scanners on a phantom. This work also evaluated in vivo ADC agreement and expert-ratings between DWI techniques in order to try to validate phantom results.

## Materials and Methods

### Ethics Approval

Patients provided written informed consent for the anonymous use of data. This study was approved by the local ethics committee. Forty-two consecutive prostate cancer patients and eleven head-and-neck

cancer patients were originally considered. The prostate patients were randomly assigned an EPI-RESOLVE or EPI-small FOV comparison. One prostate cancer patient was excluded due to important prostate deformations and different levels of rectal gas, between the subsequent DWI acquisitions. Another prostate cancer patient was excluded as no DWI imaging technique was possible to interpret due to the presence of two severely artifacting prosthesis. This led to a final 40 prostate cancer patients population.

### MRI Data Acquisition

An overview of the different MRI scanners, DWI techniques, and coils used in this study is provided in Table 1 with full sequence details in Table S1 in the Supplemental Material. This study investigated three different types of sequences for DWI. The first was standard EPI with GRAPPA or SENSE acceleration. The second was oriented towards geometric fidelity: segmented EPI RESOLVE (Siemens)<sup>15</sup> and split-echo Turbo Spin Echo (TSE SPLICE, Philips).<sup>16</sup> The third was a small FOV DWI (Siemens: ZoomitPro, Philips: Iris).<sup>17</sup> Finally, these methods were also compared between two different coil types: flexible and head-and-neck coils.

All DWI sequences were clinically optimized in-house in collaboration with Siemens and Philips engineers to obtain the minimum echo time and interecho time. These sequences were used on phantom and in vivo with the same acquired and reconstructed voxel size (respectively,  $3 \times 3 \times 3 \text{ mm}^3$  and  $1.5 \times 1.5 \times 3 \text{ mm}^3$ ) and clinical appropriate FOV, except the small FOV sequences, which imply smaller FOVs (see Table S1 in the Supplemental Material for sequence details). Acquisitions were for all patients limited at equal acquisition time: 6 minutes for prostate cancer and 3 minutes 30 for head and neck cancer. The in vivo small FOV acquisitions used, relative to EPI, 14% smaller voxel sizes and an extra average/excitation for b1000, following the SNR efficiency gain to EPI as explained in the results. Fat saturation was applied equally for all sequences.

3D distortion correction was applied for all sequences and MRI scanners. Four encoding directions were applied for diffusion sensitizing gradients, as this study investigated isotropic DWI. The k-space-based GRAPPA parallel imaging acceleration was used for EPI and RESOLVE (Siemens), which has a beneficial influence on echo time reduction.<sup>21</sup> Image domain-based parallel imaging acceleration SENSE was applied for Philips sequences.

The NIST standard diffusion phantom (serial n° 128-A-03-097, CaliberMRI, Boulder, Colorado, US) was used and prepared in an ice bath at 0°C. This phantom contained 13 vials with poly vinyl propylide solutions of 0%, 10%, 20%, 30%, 40%, and 50% corresponding to NIST calibrated ADC values of 1.127, 0.843, 0.607, 0.403, 0.248, and  $0.128 \cdot 10^{-3} \text{ mm}^2/\text{sec}$  with a listed uncertainty of  $0.001\text{--}0.008 \cdot 10^{-3} \text{ mm}^2/\text{sec}$ .<sup>22</sup> All phantom measurements were based on the central 3D part of the spheres, which were originally automatically contoured on a high-resolution computed tomography (CT) scan. The central 3D part was defined by a 3D 5 mm retraction of the outer contour, resulting in a  $\pm 10 \text{ cm}^3$  volume. For each DWI phantom acquisition series, a high-resolution T1 3D Turbo Spin Echo was acquired in order to fuse with the original CT scan. The data were analyzed using custom python (3.9) code and the R project (V4.3.0, R Core Team, Vienna, Austria).<sup>23</sup>

Analysis was performed according to the QIBA workgroup methodology.<sup>14,24,25</sup> Using definitions in ISO 5725-1, a “true

**TABLE 1. Overview of Different Scanners and DWI Techniques**

Scanner number		1	2	3
Field strength (T)		1.5	3	3
Type		Sola	Vida	Achieva dstream
Manufacturer		Siemens	Siemens	Philips
Software version		XA11B	XA11B	5.7.1.1
SNR accuracy		X		
SNR efficiency		X	X	X
Techniques investigated	EPI	EPI	EPI	EPI overplus
	Distortion-reducing	Resolve	Resolve	TSE SPLICE
	Small FOV	ZoomitPro	–	IRIS
Coils investigated	H&N coils	20 channel	20 channel	32 channel
	Body + spine coils (channels)	12 + spine	18 + spine	–
Max gradient amplitude (mT/m)		33	45	80
Slew rate (T/m/sec)		125	200	200

EPI: echo planar imaging; H&N: head and neck; TSE SPLICE: split echo turbo spin echo acquisition.

value” corresponds with an exactly known value. For phantom measurements, the term “accuracy” was applied as the NIST ADC value of the vials was close to a true value. The term “agreement” should be used instead of “accuracy” in the case a true value was not known, which was the case for in vivo comparisons. Accuracy can then be divided into:

1. Precision ( $\sigma$ )
  - a. Repeatability ( $\sigma_{\text{repeat}}$ ): variability when repeating the measure or sequence under the same unchanged setup conditions
  - b. Reproducibility ( $\sigma_{\text{repro}}$ ): variability when reproducing the measure or sequence under a newly installed setup
2. Trueness
  - a. Systematic errors or fixed bias from measure to true value
  - b. Proportional bias (fixed, proportional or non-fixed)

The coefficient of variation (CV) was calculated as the relative standard deviation, where  $\sigma$  is the precision or variability of the measure and  $\mu$  is the mean of the measure:

$$CV = \frac{\sigma}{\mu} * 100\%$$

### **$SNR_{\text{NEMA1}}$ Accuracy for DWI**

SNR was evaluated using the NEMA1 difference method.<sup>26,27</sup> This method uses two identical acquisitions for SNR evaluation. These are summed and subtracted to obtain, respectively, a mean signal

intensity and difference image.  $\bar{S}_I$  corresponds to the mean signal intensity in a 3D region of interest of the mean image, and  $SD_{\text{diff}}$  corresponds to the standard deviation in the region of interest of the difference image. From this, SNR can be calculated as:

$$SNR_{\text{NEMA1}} = \sqrt{2} \frac{\bar{S}_I}{SD_{\text{diff}}}$$

The calculated SNR values correspond to the observed SNR and integrates the acceleration g-factor.<sup>28</sup> The accuracy of this method was validated with a voxelwise noise evaluation of a series of measurements, as detailed Data S1 in the Supplemental Material.<sup>13</sup> SNR was evaluated on phantom for all techniques and coils for DWI at multiple  $b$  values ( $b = 0, 10, 20, 30, 50, 100, 200, 500, 1000, 1500, 2000 \text{ mm}^2/\text{sec}$ ).

### **DWI Efficiency: $SNR/\sqrt{\text{Second}}$**

The efficiency of the sequence in clinical routine between different MRI scanners, techniques and coils can be measured by evaluating the SNR gain per square root second, or in other words, the amount of SNR provided per second for each  $b$  value.

SNR for DWI can be improved through several methods. A first is echo time shortening, which leads to higher SNR due to lower T2 decay. However, echo time shortening is dependent on resolution, bandwidth and is limited to hardware and does not shorten acquisition time. A second method commonly used in clinical practice is acquiring multiple averages (also called excitations or means), directly implemented in the console. But then the SNR is only

raising as the square root of the number of averages. For example, for “one” average, if the first sequence obtains SNR1 in 1 sec, the second SNR2 in 4 sec, and the third SNR3 in 24 sec, the first one will obtain  $\text{SNR1} \times \sqrt{24/1}$ , and the second will obtain  $\text{SNR2} \times \sqrt{24/4}$  within the same 24 s when applying multiple averages. Inversely, the third sequence will only obtain  $\text{SNR3} \times \sqrt{1/24}$  theoretically within the time SNR1 was obtained. A logical normalization to remove time influence and time taken per average is thus by dividing by the square root of time, which allows direct absolute comparison in efficiency:  $\text{SNR}/\sqrt{\text{second}}$ .

The time in seconds for each  $b$ -value was obtained by adding 10 averages/excitations in the console for that specific  $b$ -value and dividing the additional required time by 10 to reduce the rounding error for very fast acquisitions. Relative efficiency percentages between coils and DWI techniques were calculated as the mean of relative efficiency over all measured  $b$ -values according to the uncertainty budget. SNR was not corrected for bandwidth and the number of phase encoding steps, as these were intricate to the optimized sequence, such as reducing echo time.

### ADC Accuracy: Diffusion Phantom

The quantitative imaging biomarker for DWI is the ADC value. In this section, the accuracy of the obtained ADC values was measured for clinical acquisitions using the NIST standard diffusion phantom (details in “MRI data acquisition” section). These sequences used clinically relevant  $b$ -values and number of averages/excitations per  $b$ -value (“clinical levels”). For prostate, this was based on  $b = 50 \text{ sec/mm}^2$ ,  $b = 400 \text{ sec/mm}^2$ , and  $b = 1000 \text{ sec/mm}^2$  with 6 minutes acquisition time. For head and neck, this was based on  $b = 0 \text{ sec/mm}^2$  and  $b = 1000 \text{ sec/mm}^2$  acquisitions with 3 minutes and 30 sec acquisition time (see Table S1 in the Supplemental Material for all sequence details).

ADC reproducibility was evaluated using measurements performed on eight different days using the clinical  $b$ -values and number of averages/excitations (details in Table S1 in the Supplemental Material). ADC trueness for the different scanners, available techniques, and coils was evaluated by comparing to the NIST standard diffusion phantom values using Bland–Altman plots.

### Artifacts and Distortions

A  $\text{CuSO}_4$ -water-doped geometrical phantom was equipped with additional inserts for artifact assessment: a titanium needle and air insert simulating the tumor/air interfaces for prostate and head and neck cancers, but also gold fiducial, silicone oil, maize oil/water, and grid inserts, which were not used in this study. “Many to one mapping” artifacts due to nonlinear susceptibility induced distortions were evaluated for  $b = 1000 \text{ sec/mm}^2$  images by contouring the hypersignal region for the titanium needle and air insert.<sup>29</sup> The artifact surface, as well as its relative intensity to a homogeneous region, was evaluated. The same user, CR, contoured manually the artifacts on all images, verified by FC. Reproducibility of the artifact size contouring was evaluated by reproducing artifact delineations on five different days by the same user.

Geometrical distortions were assessed on the diffusion phantom using a mutual information-based deformable registration (Raystation, Raysearch Laboratory, Stockholm, Sweden) with a

1-mm grid between a high-resolution 3D-corrected T1 3D TSE acquisitions and the b1000 diffusion image both acquired with the same setup for each evaluation. Distortions on a grid phantom can generate artifacts and increase distortions due to very large susceptibility differences between the thick plastic slices and doped water. Therefore, this was evaluated on the NIST phantom directly, in an attempt to simulate clinical practice more closely in more homogeneous regions excluding regions with artifacts.

### Clinical Evaluation: Image Quality and ADC Agreement

The three different techniques were clinically evaluated on the 1.5 T MRI scanner. For patients with prostate cancer, a standard EPI was applied and combined randomly with a small FOV ( $n = 18$  patients) or RESOLVE acquisition ( $n = 22$  patients). All acquisitions used  $b$ -values equal to 50, 400, and 1000  $\text{sec/mm}^2$ , equal fat saturation, and limited at the same 6 minutes acquisition time. For patients with head and neck cancer, small FOV and RESOLVE were acquired ( $n = 11$  patients) with  $b = 0$  and 1000  $\text{sec/mm}^2$  acquisitions with 3 minutes and 30 sec acquisition time. No EPI was acquired for patients with head and neck cancer, as it was considered to result in too many artifacts.

All patients were positioned using MRI compatible radiotherapy immobilization devices: omniboard and thermoplastic masks (Macromedics, NL). These immobilization devices reduced movement issues compared with standard acquisitions but could reduce SNR due to distance to coils. As such, an optimized flexible coil setup used for head and neck: a posterior 6-channel coil and an anterior 12-channel coil.

Qualitative image rating was performed by OG (5 years experience), XL (7 years), VC (4 years), and LP (4 years) for head and neck DWI whilst DP (20 years), BV (6 years), EAI (3 years), and BMA (6 years) rated prostate DWI.

Next, a quantitative evaluation was performed for in vivo ADC agreement between techniques. 3D volumes were manually contoured in RayStation (V11b, RaySearch laboratories, Stockholm, Sweden). For patients with prostate cancer, the prostate/clinical target volume, gross target volume if present, internal obturator muscle, and an artefact-free bladder region were contoured. For patients with head and neck cancer, this concerned parotid glands, eyes, pterygoid muscles, masseters, grey matter, submandibular glands, and gross target volumes, where applicable depending on the localization. All volumes were contoured by radiation oncologists (XL: head and neck, DP & BV: prostate) with assistance of radiologists when required. Image contouring was also supported by fusion of multi parametric MRI with CT, dual energy CT for head and neck and PET/CT when available, as shown in Figures S4–S6 in the Supplemental Material. Image fusion was effective due to the use of immobilization devices reducing movements and improving position reproducibility, gold fiducial implants for prostate fusion but also bladder and dietary preparation. Contours were equal size and location on each pairwise DWI acquisition and were verified for coherence by FC. All MRI sequences were corrected for 3D distortions and under geometrical quality control.

### Statistical Analysis

Phantom and in vivo ADC accuracy were assessed by Bland–Altman plots and 95% limits of agreement (LOAs). Phantom ADC values were compared with the NIST provided values whilst in vivo the

DWI techniques were compared pairwise between DWI techniques. The 95% confidence intervals of the bias and LOA results were calculated and plotted in the Bland–Altman plots as additional dotted lines using the *BlandAltmanLeh* method in the R project.<sup>23,30</sup> Bland–Altman plots indicate bias if the 0 difference value is outside of the confidence intervals (remark: not the 95% LOAs). Next to the Bland–Altman plots, all ADC evaluations were subjected to a paired two sided *P*-test with prior verification of normal distribution through a Shapiro–Wilk test both at *P* < 0.05 level in the R project.

Qualitative pairwise evaluations were performed by four raters for each discipline. For each patient, a rater was given b1000 and ADC images of two techniques, combined with medical history and all other imaging. They then attributed a 0–10 score for both DWI techniques to b1000 and ADC images. The subtraction of these scores resulted in a difference score, indicating the preference for a specific technique even if the absolute score was different between raters. This resulted also in a scale by how much they preferred each technique over the other. ADC and b1000 could also be scored different by radiologists.

**TABLE 2. Summary of All Results Obtained in This Study**

**Result (uncertainties: see main text)**

Phantom	SNR <sub>NEMA1</sub> accuracy for DWI	Precision	Repeatability	CV = 5%		
			Reproducibility	CV = 12%		
		Trueness		+3.4% bias		
	SNR/ $\sqrt{s}$ efficiency		Body to H&N coil		27% loss	
			ZoomitPro to EPI		14% gain	
			RESOLVE to EPI		42% loss	
			TSE SPLICE to EPI		13% loss	
	Distortions/artifacts				24% loss for $b < 500 \text{ sec/mm}^2$	
		Artifacts			SPLICE TSE > RESOLVE > ZoomitPro > EPI	
	ADC accuracy		Distortions		Least: RESOLVE 1.5 T body coil	
Reproducibility			H&N coil	CV = 1.3%		
			Body coil	CV = 2.2%		
Trueness			All types/coils except IRIS	95% LOA = $\pm 0.03 \times 10^{-3} \text{ mm}^2/\text{sec}$ ; bias $< 0.01 \times 10^{-3} \text{ mm}^2/\text{sec}$		
			IRIS	95% LOA $[-0.04; -0.013] \times 10^{-3} \text{ mm}^2/\text{sec}$ ; bias $-0.03 \times 10^{-3} \text{ mm}^2/\text{sec}$		
In vivo	Image quality-raters			ZoomitPro = EPI > RESOLVE		
	ADC accuracy	EPI-ZoomitPro	Prostate-target	95% LOAs ( $10^{-3} \text{ mm}^2/\text{sec}$ )	Bias ( $10^{-3} \text{ mm}^2/\text{sec}$ )	
						$[-0.13; +0.30]$
		EPI-RESOLVE	Pelvis-all		$[-0.16; +0.63]$	0.23
			Prostate-target		$[-0.47; +0.02]$	-0.23
		RESOLVE-ZoomitPro	Pelvis-all		$[-0.53; +0.09]$	-0.22
			H&N-target		$[-0.33; +0.54]$	0.11
			H&N-all		$[-0.21; +0.41]$	0.1

The first part lists all phantom results ranging from SNR<sub>NEMA1</sub> accuracy, SNR efficiency per technique and coil use, distortion/artefact evaluation, and ADC accuracy. The second part gives an overview of the in vivo validation.

CV: coefficient of variation; DWI: diffusion weighted imaging; EPI: echo planar imaging; H&N: head and neck; SNR: signal-to-noise-ratio; TSE SPLICE: split echo turbo spin echo acquisition.

The null hypothesis  $H_0$  was there is no difference between the methods: each pairwise comparison could indicate that a specific technique was evaluated better than the other. Therefore, a two-sided pairwise Wilcoxon Signed Rank test evaluated the order of techniques. A one-sided pairwise student test then evaluated by how much each technique was evaluated as better. Tests were evaluated at a  $P < 0.05$  level. Ratings were performed independently and raters were not aware of any of the results of this study.

## Results

A summary of the results detailed in the following sections is given in Table 2. The SNR accuracy assessment results for DWI are detailed in the Supplemental Material Data (S2) with finally a +3.4% bias, 5% repeatability CV and 12% reproducibility CV.

### DWI Efficiency: Phantom $\text{SNR}/\sqrt{s}$

Figure 1 shows the  $\text{SNR}/\sqrt{s}$  efficiency results of the central insert of the NIST diffusion phantom, with ADC value of  $1.127 \times 10^{-3} \text{ mm}^2/\text{sec}$ , as the noise levels vary due to geometrical and acceleration considerations (see Data S2.4 in the Supplemental Material “SNR trueness: dependencies”). Uncertainty intervals for the  $\text{SNR}/\sqrt{s}$  shown in the Fig. 1 represent the reproducibility uncertainty, modified with a 3.4% bias.

The use of body coils resulted in a mean 27% efficiency loss compared to dedicated head and neck coils for both scanner 1 and 2 (Sola and Vida), ranging between 36% and 10% loss ( $\pm 11\%$ ) depending on the sequence type.

The ZoomitPro Siemens small FOV technique was 8% (head & neck coil) to 14% (Body coils)  $\pm 5\%$  more efficient than the standard EPI technique, whereas the IRIS technique had low efficiency compared to EPI: 68% loss. The segmented EPI RESOLVE technique for Siemens scanners was associated with efficiency reduction (36%–65% relative to EPI), whereas for Philips, the TSE SPLICE technique approximated the efficiency of the EPI technique for b-values of  $>500 \text{ sec}/\text{mm}^2$ . Details of all efficiency comparisons can be found in the Table S2 in the Supplemental Material.

### ADC Accuracy: Diffusion Phantom

Phantom reproducibility for ADC obtained values, using clinical sequences for the Sola 1.5 T (scanner 1) MRI scanner over all inserts, resulted in mean absolute  $\text{SD}_{\text{repro}}$  values of  $0.0057 \times 10^{-3} \text{ mm}^2/\text{sec}$  and  $0.0062 \times 10^{-3} \text{ mm}^2/\text{sec}$  or mean  $\text{CV}_{\text{repro}}$  values of 1.3% and 2.2% for head and neck coil and body coil, respectively. The CVs were ranging between 0.5%–3% and 0.3%–10%, respectively, for the head and neck and body coil, depending on the ADC values (highest ADC to lowest). These values were of the same order of magnitude as the listed NIST uncertainty (trueness) of the inserts.

Figure 2 shows that all the Siemens sequences considered in this study using clinical b levels and means for both 1.5 T and 3 T obtained 95% LOA (ADC) in the order of  $\pm 0.03 \times 10^{-3} \text{ mm}^2/\text{sec}$  with (if present) statistically significant bias  $< 0.01 \times 10^{-3} \text{ mm}^2/\text{sec}$  from the NIST values. For the Philips Achieva (scanner 3), a statistically significant bias of  $-0.028 \times 10^{-3} \text{ mm}^2/\text{sec}$  for the small FOV EPI (IRIS) sequence was measured, although with LOAs lower than that for EPI and the segmented EPI. The Philips Achieva EPI and segmented TSE presented a small positive and negative linear bias, which lead subsequently to a slightly larger 95% LOA compared with Siemens sequences.

### Artifacts and Distortions

Figure 3 shows that the size of artifacts was greatly reduced by the segmented techniques, but not the relative intensity for the Vida 3 T (scanner 2). The ZoomitPro technique resulted in a compromise between EPI and RESOLVE, whereas the IRIS technique resulted in equivalent artifacts as EPI for the Achieva 3 T (scanner 3).

Distortions in more homogeneous regions, using the NIST diffusion phantom showed subtle differences between techniques and coils, with the 1.5 T RESOLVE technique using body coils showing the least distortions (see Figure S3 in the Supplemental Material for more details).

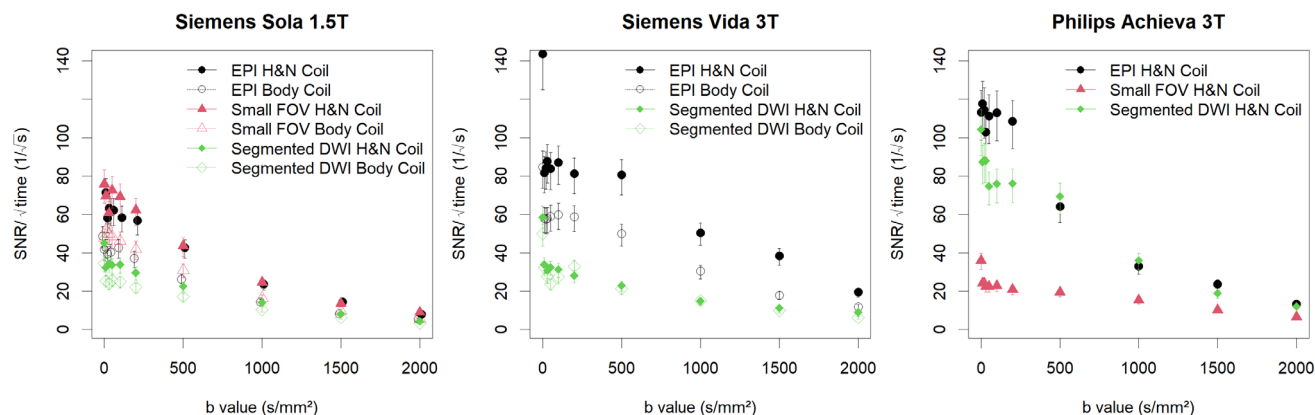


FIGURE 1: Efficiency ( $\text{SNR}/\sqrt{s}$  for 1 average, in vitro) at different b-values for the central insert with an ADC value of  $1.127 \times 10^{-3} \text{ mm}^2/\text{sec}$ .

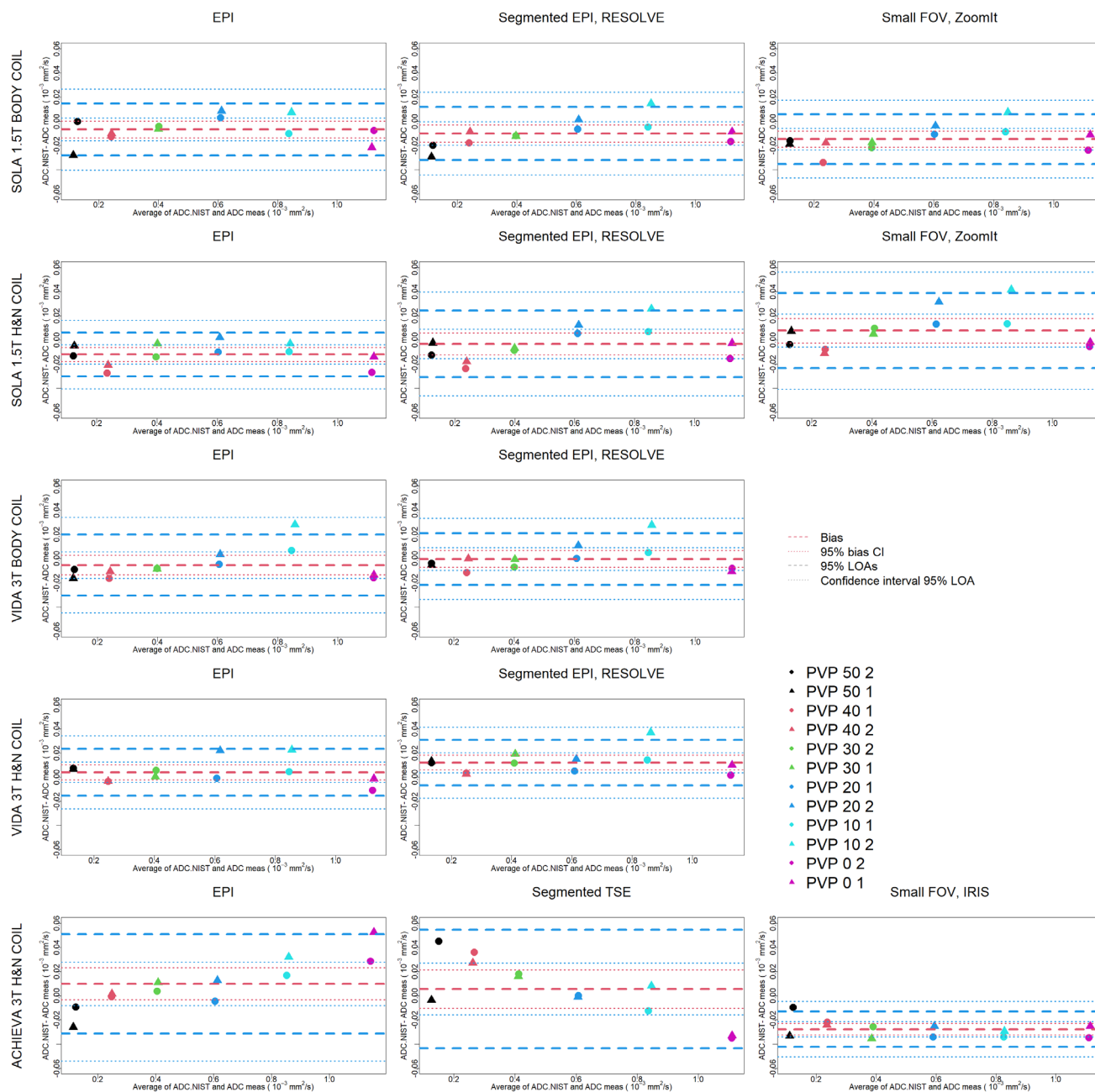


FIGURE 2: Phantom ADC trueness for all MRI scanners, sequences, and coils. The red dashed line represents the mean bias, with the red dotted lines representing the statistical uncertainty 95% confidence interval. Blue dashed lines represent the 95% LOA, with respective uncertainty intervals in blue dotted lines.

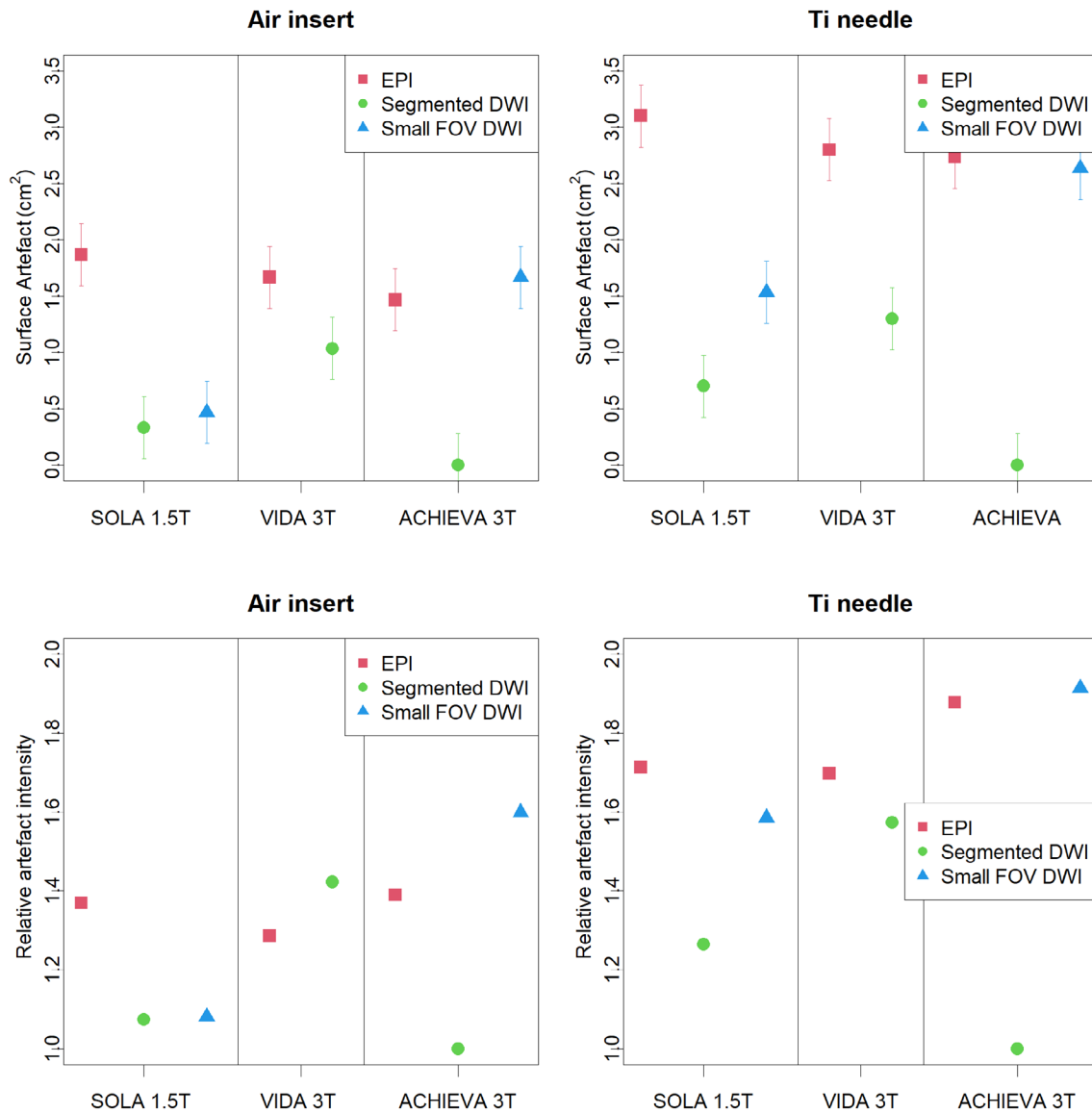
### Clinical Evaluation: Image Quality and ADC Agreement

**IN VIVO IMAGE QUALITY.** Clinical cases are shown in Figures S4–S6 in the Supplemental Material. The results indicated a statistically significant preference of the small FOV technique over RESOLVE technique for head and neck cancer patients for both b1000 and ADC with respective mean difference scores of 2.5 and 1.7 (Fig. 4). The small FOV technique offered in most cases better b1000 evaluation but sometimes ADC was visually better for the resolve

technique. In specific high susceptibility gradient regions like the sinus or nose, however, artifacts were still present for the small FOV technique, and the RESOLVE technique would be required.

The combined ratings of all four raters for prostate DWI resulted in a statistically non-significant difference between EPI and small FOV technique for prostate for both b1000 ( $P = 0.06$ ) and ADC ( $P = 0.68$ ) but statistically significantly favoring EPI over RESOLVE. When going into rating detail, the small FOV technique was statistically significant favored by three raters for both b1000 and ADC,





**FIGURE 3:** Artifact assessment: top graphs represent the surface area of the artifact and bottom graphs represent the relative intensity of the artifact to a homogeneous artifact-free region.

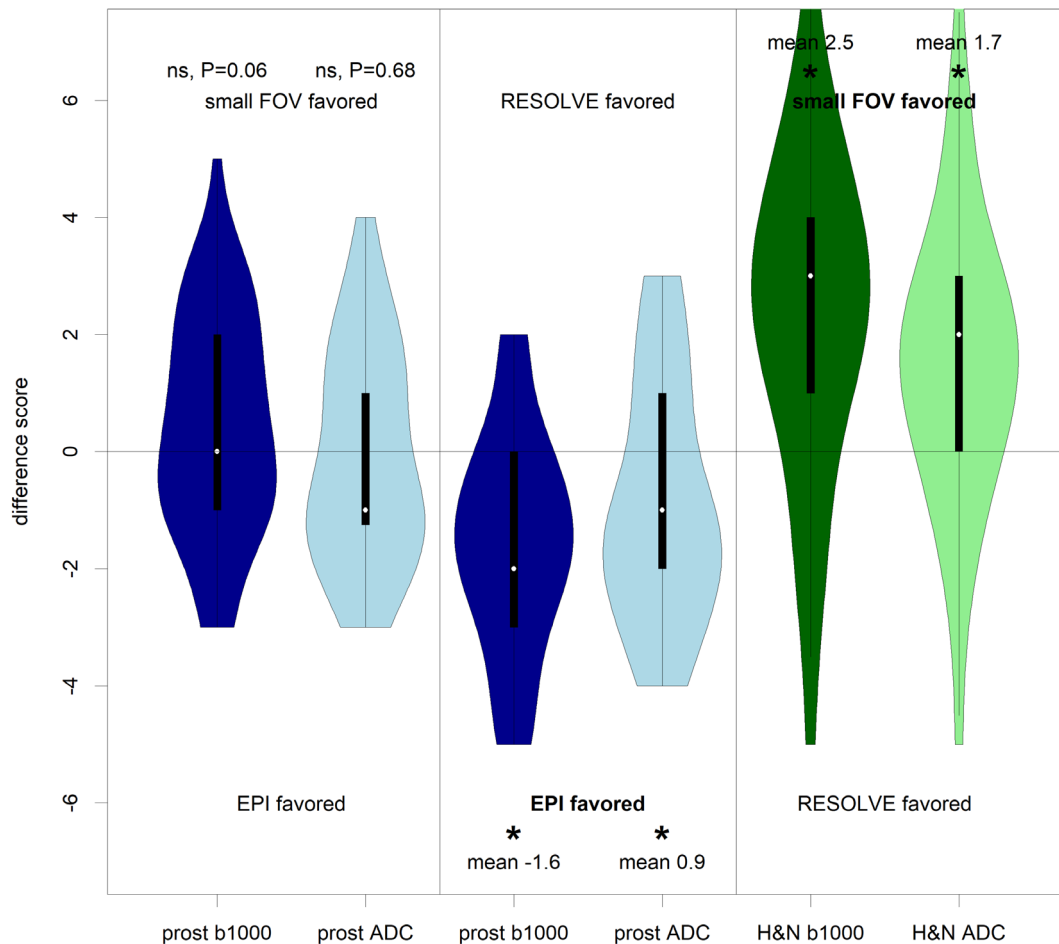
but not systematically for all patients. The ratings of these three raters resulted in respectively scores favoring small FOV for b1000 and ADC at +1 and +1.22. To note, these three raters sometimes preferred EPI over small FOV. One rater favored larger FOV EPI over the small FOV technique more systematically with a statistically significant difference score for, respectively, b1000 and ADC of  $-1.4$  and  $-1.44$ .

**IN VIVO ADC AGREEMENT.** Figure 5 shows that ADC values were affected in vivo by statistically significant biases in the order of  $0.1$  to  $0.2 \times 10^{-3} \text{ mm}^2/\text{sec}$  and variabilities with 95% LOA in order of  $\pm 0.2$  to  $0.3 \times 10^{-3} \text{ mm}^2/\text{sec}$  between techniques. Even for homogeneous volumes such as the bladder, large differences between techniques were present, in the order of 10%–20% bias and  $\pm 10\%$ –30% variability between techniques.

## Discussion

This study provided a quantitative evaluation of tradeoffs in SNR efficiency, artifacts and ADC accuracy between DWI MRI sequences and coils, but showed also ADC differences in vivo, not present on phantom, between techniques.

The obtained phantom DWI  $\text{SNR}_{\text{NEMA1}}$  accuracies were slightly greater than those reported by Dietrich et al.<sup>31</sup> for regular images (bias of 2.7% with  $\pm 1.6\%$  repeatability) but consistent with results for diffusion tensor imaging<sup>27</sup> (4% bias). Most likely, this bias with the more accurate voxelwise series of measurements approach<sup>14</sup> comes from the slightly non-gaussian distribution of the difference noise image. However, the accuracy was sufficient for the intended efficiency study especially in the light of the otherwise required number of acquisition for a voxelwise series of measurements approach.



**FIGURE 4:** Violin plots of difference of pairwise visual DWI scoring of b1000 and ADC images for prostate (blue, 4 raters, EPI vs. small FOV and EPI vs. RESOLVE) and Head and Neck (green, 4 raters, small FOV vs. RESOLVE). Significance was indicated as:  $P < 0.05$  (\*). For prostate, there was no statistically significant difference between the small FOV and EPI technique for all raters combined for both b1000 ( $P = 0.06$ ) and ADC ( $P = 0.68$ ). However, the small FOV technique was slightly, but statistically significant, favored to EPI by three raters, whereas the fourth rater favored EPI to small FOV systematically. The small FOV technique was in general favored for head and neck cancer, except in specific cases.

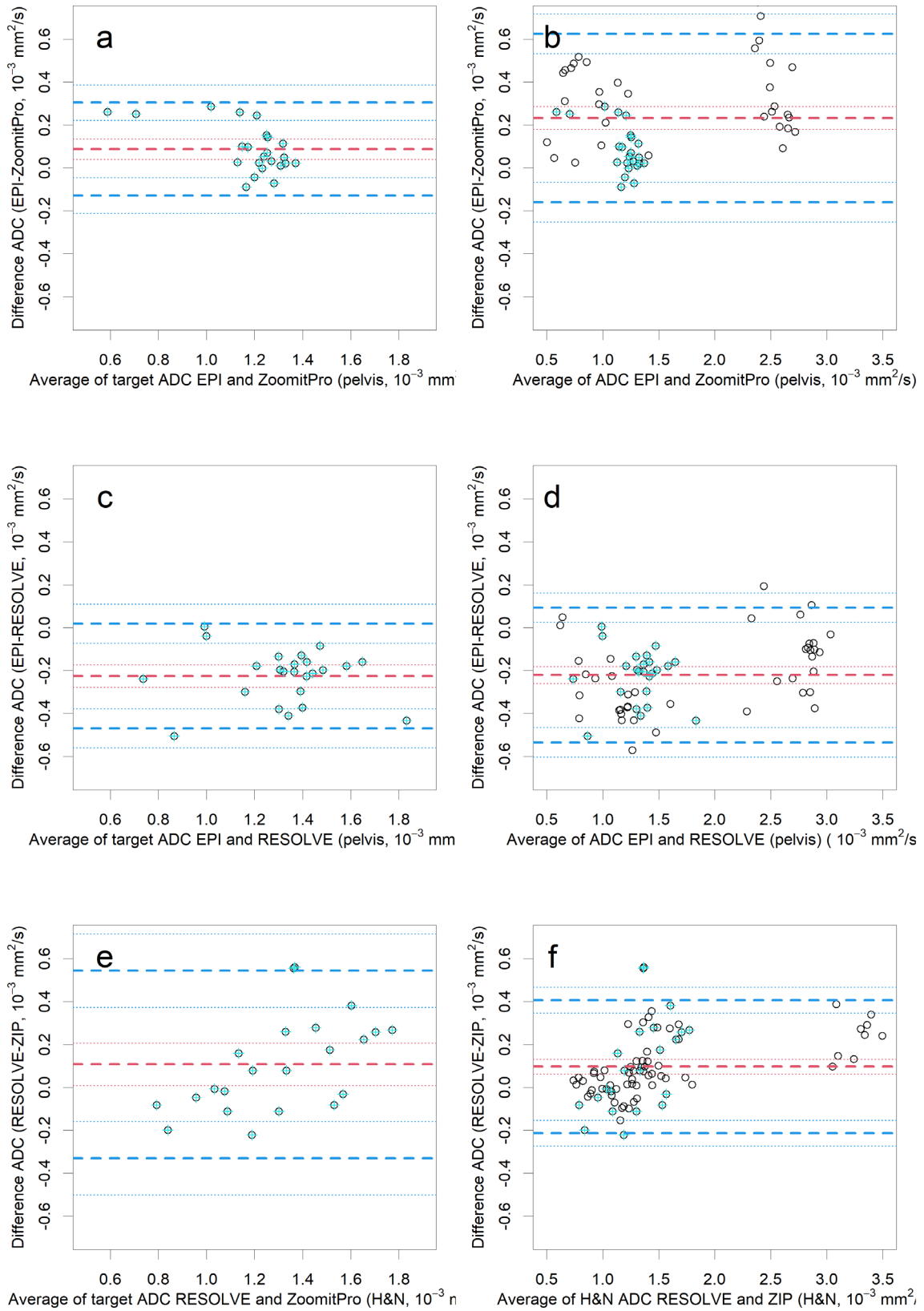
The  $\text{SNR}/\sqrt{s}$  efficiency can be an objective metric to compare techniques for b-images supporting practical decisions on technique and  $b$ -values. However, an interaction between voxel sizes, repetition time, bandwidth, acceleration, and echo time parameters is still possible. For  $b = 0 \text{ sec}/\text{mm}^2$  measurements, different gradients (and crusher gradients) were used by the scanners, resulting in considerably lower acquisition time, higher noise levels, but more efficient  $\text{SNR}/\sqrt{s}$ , with the Vida 3 T (scanner 3) EPI being the most efficient. For body applications, a  $b = 50\text{--}100 \text{ sec}/\text{mm}^2$  value was used to reduce the influence of slow-moving blood with high T2 values. Vascularity appears as a high signal due to the slow-moving protons in the gradient field introducing different phase information. However, the accuracy of tumor ADC calculations could benefit these efficient  $b = 0 \text{ sec}/\text{mm}^2$  acquisitions.

Artifacts are important for both diagnostic and radiotherapy MRI. Distortions can complicate diagnostics but are more problematic for applications requiring high geometric

accuracy such as radiotherapy contouring. Overall, geometrical distortions in homogeneous regions were not important due to the short echo time optimization. The small FOV ZoomitPro technique on the Sola 1.5 T (scanner 1) resulted in a compromise between EPI and RESOLVE, which was however not noticed for the IRIS Philips technique. On the other hand, the RESOLVE Siemens technique resulted in lower artifacts, but not as low as the TSE SPLICE Philips technique.

The results of this study indicated very good ADC accuracy on phantom for all, except the IRIS technique. This was not completely in line with the results of Liney et al. for the ZOOMit technique.<sup>18</sup> However, they used the predecessor of the ZOOMitPro technique on a 3 T scanner, the RESOLVE sequence used took approximately double the time of the other techniques and agreement was not evaluated extensively in vivo.

In vivo results followed the phantom results for head and neck cancer favoring the small FOV technique, except in



**FIGURE 5:** Bland–Altman plots of ADC value comparison in vivo. The left column shows only tumor or clinical target volumes (blue). The right column shows all results including nontarget volumes. The first two rows show Bland–Altman plots for EPI, RESOLVE, and ZoomitPro techniques in the pelvis region. The final row shows Bland–Altman plots for RESOLVE and ZoomitPro techniques in the head and neck region. The bias is shown in dashed red lines and its corresponding 95% confidence interval in dotted red lines. The 95% LOA are shown in dashed blue, with corresponding 95% confidence interval in dotted blue lines.

high susceptibility gradient regions requiring the RESOLVE technique. For prostate, this was more subtle with three raters favoring small FOV, however not systematically. One rater systematically favored EPI, due to the larger FOV offering better localization possibilities.

In vivo ADC results however, indicated large bias and large 95% LOA between techniques, despite the extensive phantom quality control indicating accurate results. This result is consistent with the in vivo bias noticed by Lu et al for thyroid glands between EPI and small FOV for a general electric (GE) 3 T scanner, but not with that found by Korn et al. for prostate on a GE 3 T with endorectal coil.<sup>19,20</sup> The agreement obtained in this study was not as good as the repeatability obtained in a breast MRI study, which was in the order of  $\pm 0.15 \times 10^{-3} \text{ mm}^2/\text{sec}$  LOAs.<sup>32</sup> The 95% LOAs obtained in the current study were however consistent with those of reproducibility studies of the prostate: showing <35% ADC differences and median ADC values of  $\pm 0.28 \times 10^{-3} \text{ mm}^2/\text{sec}$ .<sup>33,34</sup> The in vivo ADC results obtained in the head and neck region were in line with the results of Habrich et al. where they obtained between 13% and 31%.<sup>35</sup> Paudyal et al. obtained 95% LOA's for ADC repeatability in the order of  $\pm 0.1 \times 10^{-3} \text{ mm}^2/\text{sec}$  for HNSCC with EPI on a Philips Ingenia 3 T and in the order of  $\pm 0.5 \times 10^{-3} \text{ mm}^2/\text{sec}$  for papillary thyroid cancer with small FOV technique on GE scanners.<sup>36</sup>

### Limitations

The SNR quantification study focused on phantom acquisitions and not clinical imaging of the human body. However, clinical comparison of the SNR efficiency would be challenging due to reproducibility as one could expect movement and flow issues for the subtracted image. In this study, isotropic DWI were acquired using the four gradient approach, but simultaneous multislice acceleration or BLADE methods were not included. Tumor conspicuity and high or low contrast resolution were not assessed in detail but only globally qualitatively by raters.

This work was limited to two vendors, three scanners and three techniques. However, the obtained results, such as the SNR/ $\sqrt{s}$  efficiency and ADC accuracy based on the NIST standard diffusion phantom, should provide an absolute point of comparison when applied to other scanners and DWI techniques.

The high ADC bladder values were partly impacted by different artifacts in the homogeneous bladder region. Care was taken to not incorporate these regions but could not be avoided totally. An important factor in our study could be the in vivo noise floor<sup>37</sup> or the non-Gaussian diffusion at larger  $b$ -values not being comparable in vitro to in vivo, even though care was taken to obtain sufficient SNR. T2 dependence was verified on the NIST 106–0030 standard phantom

with calibrated T2 values but showed no difference in ADC values between sequence types.

### Conclusion

The SNR<sub>NEMA1</sub> method for DWI was found to provide reasonable accuracy. The combination of the proposed quantifiable SNR/ $\sqrt{s}$  metric with the assessment of ADC accuracy, artifact and distortion measurements can provide an objective supporting method for the choice of technique in practice.

Although ADC accuracy on phantom was excellent, important differences were observed in vivo for patients with both head and neck cancer and prostate cancer between the three techniques. Therefore, ADC values should be used with caution for quantitative imaging biomarker evaluations especially when changing techniques.

### Acknowledgments

We wish to thank Editage for English language editing.

### References

- Luzurier A, Jouve De Guibert PH, Allera A, et al. Dynamic contrast-enhanced imaging in localizing local recurrence of prostate cancer after radiotherapy: Limited added value for readers of varying level of experience. *J Magn Reson Imaging* 2018;48:1012-1023.
- Vargas HA, Hötter AM, Goldman DA, et al. Updated prostate imaging reporting and data system (PIRADS v2) recommendations for the detection of clinically significant prostate cancer using multiparametric MRI: Critical evaluation using whole-mount pathology as standard of reference. *Eur Radiol* 2016;26:1606-1612.
- Lotte R, Lafourcade A, Mozer P, et al. Multiparametric MRI for suspected recurrent prostate cancer after HIFU: Is DCE still needed? *Eur Radiol* 2018;28:3760-3769.
- Onal C, Erbay G, Guler OC, Oymak E. The prognostic value of mean apparent diffusion coefficient measured with diffusion-weighted magnetic resonance image in patients with prostate cancer treated with definitive radiotherapy. *Radiother Oncol* 2022;173:285-291.
- Alexander EJ, Murray JR, Morgan VA, et al. Validation of T2- and diffusion-weighted magnetic resonance imaging for mapping intraprostatic tumour prior to focal boost dose-escalation using intensity-modulated radiotherapy (IMRT). *Radiother Oncol* 2019;141:181-187.
- Foltz WD, Porter DA, Simeonov A, et al. Readout-segmented echo-planar diffusion-weighted imaging improves geometric performance for image-guided radiation therapy of pelvic tumors. *Radiother Oncol* 2015;117:525-531.
- Martens RM, Noij DP, Ali M, et al. Functional imaging early during (chemo)radiotherapy for response prediction in head and neck squamous cell carcinoma; a systematic review. *Oral Oncol* 2019;88:75-83.
- Gurney-Champion OJ, Mahmood F, van Schie M, et al. Quantitative imaging for radiotherapy purposes. *Radiother Oncol* 2020;146:66-75.
- Shukla-Dave A, Obuchowski NA, Chenevert TL, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J Magn Reson Imaging* 2019;49:e101-e121.
- Keenan KE, Gimbutas Z, Dienstfrey A, et al. Multi-site, multi-platform comparison of MRI T1 measurement using the system phantom. *PLoS One* 2021;16:1-19.
- Glide-Hurst CK, Paulson ES, McGee K, et al. Task group 284 report: Magnetic resonance imaging simulation in radiotherapy: Considerations for

- clinical implementation, optimization, and quality assurance. *Med Phys* 2021;48:e636-e670.
12. McGee KP, Hwang KP, Sullivan DC, et al. Magnetic resonance biomarkers in radiation oncology: The report of AAPM task group 294. *Med Phys* 2021;48:e697-e732.
  13. Peña-Nogales Ó, Hernando D, Aja-Fernández S, de Luis-García R. Determination of optimized set of b-values for apparent diffusion coefficient mapping in liver diffusion-weighted MRI. *J Magn Reson* 2020; 310:1-14.
  14. Sullivan DC, Obuchowski NA, Kessler LG, et al. For the RSNA-QIBA Metrology Working Group: Metrology standards for quantitative imaging biomarkers. *Radiology* 2015;277:813-825.
  15. Porter DA, Heidemann RM. High resolution diffusion-weighted imaging using readout-segmented echo-planar imaging, parallel imaging and a two-dimensional navigator-based reacquisition. *Magn Reson Med* 2009;62:468-475.
  16. Schick F. SPLICE: Sub-second diffusion-sensitive MR imaging using a modified fast spin-echo acquisition mode. *Magn Reson Med* 1997;38: 638-644.
  17. Finsterbusch J. Improving the performance of diffusion-weighted inner field-of-view echo-planar imaging based on 2D-selective radiofrequency excitations by tilting the excitation plane. *J Magn Reson Imaging* 2012;35:984-992.
  18. Liney GP, Holloway L, Al Harthi TM, et al. Quantitative evaluation of diffusion-weighted imaging techniques for the purposes of radiotherapy planning in the prostate. *Br J Radiol* 2015;88:20150034.
  19. Lu Y, Hatzoglou V, Banerjee S, et al. Repeatability investigation of reduced field-of-view diffusion weighted magnetic resonance imaging on thyroid glands. *J Comput Assist Tomogr* 2015;39:334.
  20. Korn N, Kurhanewicz J, Banerjee S, Starobinets O, Saritas E, Noworolski S. Reduced-FOV excitation decreases susceptibility artifact in diffusion-weighted MRI with Endorectal coil for prostate cancer detection. *Magn Reson Imaging* 2015;33:56-62.
  21. Larkman DJ, Nunes RG. Parallel magnetic resonance imaging. *Phys Med Biol* 2007;52:R15-R55.
  22. Keenan KE, Carmicka S, Gottlieb SC, Stupic KF. *Assessing changes in MRI measurands incurred in a scanner upgrade: Is my study comprised*. CA, USA: ISMRM; 2017.
  23. Ripley BD. The R project in statistical computing. *MSOR Connect* 2001; 1:23-25.
  24. Kessler LG, Barnhart HX, Buckler AJ, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Stat Methods Med Res* 2015;24:9-26.
  25. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. *Stat Methods Med Res* 2015;24:27-67.
  26. American College of Radiology. *Magnetic resonance imaging quality control manual*. Virginia, USA: American College of Radiology; 2015.
  27. Griffanti L, Baglio F, Preti MG, et al. Signal-to-noise ratio of diffusion weighted magnetic resonance imaging: Estimation methods and in vivo application to spinal cord. *Biomed Signal Process Control* 2012; 7:285-294.
  28. Reeder SB, Wintersperger BJ, Dietrich O, et al. Practical approaches to the evaluation of signal-to-noise ratio performance with parallel imaging: Application with cardiac imaging and a 32-channel cardiac coil. *Magn Reson Med* 2005;54:748-754.
  29. Jones DK, Cercignani M. Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR Biomed* 2010;23:803-820.
  30. Martin Bland J, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;327: 307-310.
  31. 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;26:375-385.
  32. Newitt DC, Zhang Z, Gibbs JE, et al. Test-retest repeatability and reproducibility of ADC measures by breast DWI: Results from the ACRIN 6698 trial. *J Magn Reson Imaging* 2019;49:1617-1628.
  33. Gibbs P, Pickles MD, Turnbull LW. Repeatability of echo-planar-based diffusion measurements of the human prostate at 3 T. *Magn Reson Imaging* 2007;25:1423-1429.
  34. Barrett T, Lawrence EM, Priest AN, et al. Repeatability of diffusion-weighted MRI of the prostate using whole lesion ADC values, skew and histogram analysis. *Eur J Radiol* 2019;110:22-29.
  35. Habrich J, Boeke S, Nachbar M, et al. Repeatability of diffusion-weighted magnetic resonance imaging in head and neck cancer at a 1.5 T MR-Linac. *Radiother Oncol* 2022;174:141-148.
  36. Paudyal R, Konar AS, Obuchowski NA, et al. Repeatability of quantitative diffusion-weighted imaging metrics in phantoms, head-and-neck and thyroid cancers: Preliminary findings. *Tomography* 2019;5:15-25.
  37. Baltzer P, Mann RM, Iima M, et al. Diffusion-weighted imaging of the breast—a consensus and mission statement from the EUSOBI international breast diffusion-weighted imaging working group. *Eur Radiol* 2020;30:1436-1450.