IVIM parameters have good scan-rescan reproducibility when evidential motion contaminated and poorly fitted image data are removed.

Olivier Chevallier[1,3§], Nan Zhou [2§,] Jian He[2*], Romaric Loffroy[3], Yì Xiáng J. Wáng[1*]

[1]Department of Imaging and Interventional Radiology, Faculty of Medicine, The Chinese University of Hong Kong, New Territories, Hong Kong SAR

[2]Department of Radiology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, China.

[3]Department of Vascular and Interventional Radiology, University of Bourgogne/Franche-Comté, François-Mitterrand Teaching Hospital, Dijon Cedex, France

[§] these two authors contributed equally to this study

Corresponding authors: Dr Yi-Xiang Wang, Department of Imaging and Interventional Radiology, Faculty of Medicine, The Chinese University of Hong Kong, New Territories, Hong Kong SAR, e-mail: yixiang_wang@cuhk.edu.hk; Dr Jian He, Department of Radiology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, China. e-mail: hjxueren@126.com

Acknowledgement: We thank Miss Yao Tina Li, former research student at the Chinese University of Hong Kong, for programming the image processing tool used in this study; Dr Jean-Pierre Cercueil at Department of Vascular and Interventional Radiology, François-Mitterrand Teaching Hospital, University of Bourgogne/Franche-Comté, Dijon, France, for discussions during the course of data analysis. Dr Olivier Chevallier was supported by a grant provided by the Société Française de Radiologie (SFR) together with the Collège des Enseignants de Radiologie de France (CERF).

IVIM parameters have good scan-rescan reproducibility when evidential motion contaminated and poorly fitted image data are removed.

Abstract

Background: Intravoxel Incoherent Motion (IVIM) diffusion MRI is a promising technique for liver pathology evaluation, but this technique's scan-rescan reproducibility has been reported to be unsatisfactory.

Objective: To understand whether IVIM MRI parameters for liver parenchyma can be good after removal of motion contaminated and/or poorly fitted image data.

Materials and Methods: Eighteen healthy volunteers had liver scanned twice at the same session to assess scan-rescan repeatability, and again in another session after an average interval of 13 days to assess reproducibility. Diffusion weighted image were acquired with a 3T scanner using respiratory-triggered echo-planar sequence and 16 *b*-values (0 to 800 s/mm^2). Measurement was performed on the right liver with segmented-unconstrained least square fitting. Image series with evidential anatomical mismatch, apparent artifacts, and poorly fitted signal intensity vs. b-value curve were excluded. A minimum of three slices was deemed necessary for IVIM parameter estimation.

Results: With total 54 examinations, 6 scans did not satisfy inclusion criteria, leading to a success rate of 89% and 14 volunteers were finally included. With each scan a mean of 5.3 slices (range: 3-10 slices) were utilized for analysis. Using threshold *b*-value=80s/mm^2, the coefficient of variation and within-subject coefficient of variation for repeatability and reproducibility were: 2.86% and 4.24% for Dslow, 3.81% and 4.24%, for PF, 18.16% and 24.88% for Dfast; and those for reproducibility were 2.48% and 3.24% for Dslow; 4.91% and 5.38% for PF; 21.18% and 30.89% for Dfast.

Conclusion*:* ~~T~~IVIM parameter scan-rescan reproducibility can be potentially good.

Keywords: Diffusion weighted imaging; intravoxel incoherent motion; liver; reproducibility

Introduction

Since the initial study of Yamada et al (1), there have been greater interests to explore Intravoxel Incoherent Motion (IVIM) technique to evaluate diffused liver diseases such as liver fibrosis and nonalcoholic fatty liver disease; to characterize liver tumor; and to evaluate treatment response (2, 3). A prerequisite to translating IVIM imaging into clinical applications is accurate measurement of IVIM parameters and acceptable reproducibility. Nevertheless, accurate liver IVIM quantification is challenging, partially due to the limited sampling and low signal-to-noise ratio (SNR) for fast diffusion data acquisition. A major obstacle for clinical application of IVIM technique for abdominal organs is its unsatisfactory scan-rescan reproducibility (2, 3). For example, in a short-term reproducibility study, Andreou et al (4) reported the 95% confidence intervals of percentage difference between paired measurements of liver parenchyma was (−24.3, 25.1) for PF (*f*), (−5.12, 8.09) for Dslow (*D*), and (−31.2, 59.1) for Dfast (*D*);* and the absolute limit was (0.140, 0.232) for PF, (0.951, 1.08) for Dslow, and 35.7, 82.5 for Dfast.

IVIM diffusing imaging typically involves long data acquisition time with images acquired at a series of *b*-values. It is usually acquired with respiratory gating, however, even respiratory gating is associated with substantial residual respiration induced motion (2). Respiratory motion can cause inter-*b*-value motion and intra-*b*-value motion. Inter-*b*-value motion causes mis-match of anatomical structures on images of difference *b*-values, and intra-*b*-value motion where motion occurs during the data acquisition for the slice causes image artifacts. Diffusion MRI is also influenced by artifacts related to magnet/sequence imperfections, such as B_0 inhomogeneity resulting from susceptibility variations; geometric distortions from residual motion probing gradients-induced eddy currents (5). In this study, we introduce a manual 'image data cleaning' process, with the aim to mitigate artifacts associated with respiration motion as well as with magnet/sequence imperfections. We hypothesize that if IVIM's scan-rescan reproducibility will be satisfactory after 'image data cleaning', then with further technical improvement such as acquisition of more *b*-values for curve fitting, advanced methods for motion correction, statistical

remove of ill fitted pixels, or accelerated data acquisition with single-breath hold, liver IVIM will eventually have clinical applicability.

Materials and Methods

This study was conducted with the approval of the institutional ethics committee and informed consent was obtained. Eighteen healthy volunteers underwent IVIM diffusion imaging with a 3T magnet and a 32 channels dStream Torso coil (Ingenia, Philips Healthcare, Best, The Netherlands). The IVIM diffusion imaging was based on a single-shot spin-echo-type echo-planar imaging sequence, with 16 b-values of 0, 3, 10, 25, 30, 40, 45, 50, 80, 200, 300, 400, 500, 600, 700 and 800 s/mm^2, NSA of 2 for b=700 s/mm2 and b=800 s/mm2, and NSA=1 for other *b*-values. Spectral presaturation with inversion recovery technique was used for fat suppression. Respiratory triggering was performed using an air-filled pressure sensor fixed on the upper abdomen, resulting in an average TR of 2149 ms. Other parameters included TE=55ms, slice thickness=6mm, matrix =100×116, field-of-view (FOV)=360×300 mm, EPI factor= 29, a sensitivity-encoding (SENSE) factor=4, number-of-slices =26. The scan subjects were trained so that they maintained shallow regular breathing during image acquisition. The average IVIM scan duration was 6 min. All volunteers were scanned twice in the same session to assess scan–rescan repeatability (scans 1.1 and scans 1.2), and additionally once again in another session (scan 2) with an interval of 5–21 days (mean 13 days) to assess scan–rescan reproducibility.

The IVIM signal attenuation was modeled according to Eq 1

$$SI(b) = SI0[(1-PF)\cdot exp(-b\cdot Dslow) + PF\cdot exp(-b\cdot Dfast)] \quad [1]$$

where SI(b) and SI0 denote the signal intensity acquired with the b-factor value of b and b =0 s/mm2, respectively (6). Perfusion fraction (PF) is the fraction of the pseudo-diffusion linked to microcirculation, Dslow is the diffusion coefficient representing the pure molecular diffusion, and Dfast is the pseudo-diffusion coefficient representing the incoherent microcirculation within the voxel (perfusion-related diffusion). With the segment-unconstrained approach, the estimation of

Dslow was obtained by a least-squares linear fitting of the logarithmized image intensity at *b* values greater than threshold, which was chosen of 50, 80 or 200 s/mm^2, to a linear equation (7). The fitted curve was then extrapolated to obtain an intercept at *b*-value= 0s/mm^2. The ratio between this intercept and SI_0 gave an estimate of PF. Finally, the obtained Dslow and PF were substituted into Eq. 1 and were nonlinear least squares fitted against all *b* values to estimate Dfast using the Levenberg–Marquardt algorithm.

All curve-fitting algorithms were implemented in a custom program developed on MatLab (Mathworks, Natick, MA, USA). Firstly, a pixel-wise analysis algorithm was performed in order to exclude pixels which presented a poor fitting for Dslow with a coefficient of determination R^2 value lower than 0.8 (8). Regions of interest (ROIs) were then positioned to cover a large portion of right liver parenchyma while avoiding large vessels on Dslow parametric map after the pixel exclusion process. For ROI analysis, the IVIM parameters were calculated based on the mean signal intensity of the whole ROI, which has been shown to offer better estimation than pixel-wise fitting when the signal-to-noise ratio (SNR) of the DW images is low (9, 10). The same as some other reports (12-19), only the right lobe of the liver was measured in the current study (Fig. 1).

A manual procedure was taken to 'clean the image data' for each examination. Firstly, slices which covered only the lowest part of segment V-VI (usually slices below the gallbladder) or the hepatic dome, near the visceral near or the diaphragmatic surfaces, were discarded. Then, each scan's image series were graded as 'good quality', 'fair quality', or 'insufficient quality'. Motion induced imaging data degrading was visually assessed between consecutives images at different *b*-values for each slice, noting the location of the following anatomic structures: kidneys, gall bladder, spleen, hepatic edges, main hepatic vessels (main portal vein, portal veins until second order, main hepatic veins). If no motion or artifact was noted, the slice series was graded 'good quality'. Image series of 'insufficient quality' were mainly due to motion leading to liver displacement between images of different *b*-values (inter-*b*-value motion), and sometimes apparent artifacts in the hepatic parenchyma which could be due to intra-*b*-value motion (Fig 2).

Slices presented only slight displacement or inconspicuous artifact were graded 'fair quality'. Image series of 'good quality' and 'fair quality' were included for the second step data cleaning.

Image series which generated poorly IVIM diffusion fitted curve were then excluded. Firstly, slices which presented parameters results with a coefficient of determination R^2 value lower than 0.95 for ROI-wise fitting were excluded (20). Then, the plots of signal intensity vs. *b*-values were individually evaluated. Slices which demonstrated evidential outliers with MRI signal vs. *b*-value relation and could not be properly fitted were discarded. The initial part of the curve associated with low b-values tend to be difficult to fit accurately. Very fast signal decay sometimes occurred below b=10 s/mm^2 (21, 22). This strong signal attenuation, which occurs at *b*-values less than 10 s/mm^2 and may reflects the velocity of flowing protons present in relative large vessels instead of the micro-vessels of IVIM theory, can lead to overestimation of Dfast. In addition, for an IVIM image series to be valid, we required that at least three slices can be included for final analysis after data-cleaning. The means of all included slices' measurements were then regarded the value of the examination.

Statistical analysis was performed using MedCalc Statistical Software (version 17.6, MedCalc Software bvba, Ostend, Belgium). Intra-scan repeatability between scan 1.1 and scan 1.2, and inter-scan reproducibility between scan 1.1 and scan 2 of PF, Dslow and Dfast were assessed by coefficient of variation (CoV), the within-subject coefficient of variation ($_w$CoV), and Bland-Altman mean difference and 95% limits of agreements (BA-LA). $_w$CoV is defined by Eq 2 and Eq 3

$$Within - subject\ SD = \sqrt{\frac{\sum(x_1 - x_2)^2}{2n}} \qquad \text{Eq 2}$$

$$Within - subject\ CoV(\%) = 100 * \frac{Within - subject\ SD}{Mean} \qquad \text{Eq 3}$$

With n being the number of subjects (=14 in this study) and x_1 and x_2 are the duplicate parameter measurements for each subject.

Results

With the total 54 examinations, data-cleaning removed 6 examinations without satisfying inclusion criteria; leading to a success rate of 89% (Fig. 2). 14 volunteers were finally included for measurement reproducibility analysis (5 males and 9 females; mean age: 25.7 years; range: 24-27 years). 68.11% of the scanned slices were finally included for the final analysis, with a mean of 5.3 slices (range: 3-10 slices) for each scan. The image series graded of 'good quality' presented a higher rate of acceptable fitted curve (81.2%) than slices graded 'Fair Quality' (62.1%).

The scan-rescan repeatability and reproducibility for PF, Dslow and Dfast using threshold *b*-value=80 s/mm^2 are summarized in tables 1, 2 and 3. Bland–Altman plots for each parameter are shown in Fig 5. The 95% Bland-Altman limits of agreements, mean of CoV and wCoV for repeatability and reproducibility are shown in tables 4 and 5. Typically, for threshold *b*-value=80 s/mm^2, Dslow had a CoV, $_w$CoV of 2.86% and 4.24% between scan 1.1 and scan 1.2, of 2.48% and 3.24% between scan 1.1 and scan 2; PF had a CoV, $_w$CoV of 3.81% and 4.24% between scan 1.1 and scan 1.2, and of 4.91% and 5.38% between scan 1.1 and scan 2; Dfast had a CoV, $_w$CoV of 18.16% and 24.88% between scan 1.1 and scan 1.2, of 21.18% and 30.89% between scan 1.1 and scan 2.

Discussion

In this study, we demonstrated that IVIM parameters can have good reproducibility when evidential motion contaminated and/or poorly fitted image data are removed. Quite a few papers have addressed IVIM parameter reproducibility (11-19, 21, 23-30). Those measured and reported reproducibility in comparable format to our results are listed in table 6, which also represented the majority of literature on this topic. Our results broadly represent the best results ever reported. The difficulty to accurately measure Dfast has been discussed previously (31). While at this stage we still do not fully understand 'how good of reproducibility is good enough', the recent report by Wáng *et al.* seems to suggest a combined use of IVIM DWI parameters may have both good sensitivity and good specificity for early stage liver fibrosis detection even when IVIM DWI data acquisition were not idealized (20). Therefore, though the reproducibility for Dfast may not be optimal, this level of reproducibility may be sufficient for liver fibrosis evaluation. Note automatic images registration and motion correction for IVIM image series have been

reported by other authors (11, 12, 21, 24). However, images registration does not remove artifacts associated with intra-*b*-value motion and scanner calibration/sequence imperfection. Padhani et al. (4) described approach for machine and sequence calibrations for diffusion weighted imaging. The TE is kept as short as possible as this results in better SNR and it minimizes susceptibility echo planar imaging-related artifacts. Eddy currents related to diffusion gradients and echo planar imaging techniques lead to geometric distortions and image shearing, which can be reduced by increasing readout bandwidths or reducing the echo spacing (4).

In addition to the 'data cleaning' process taken in this study, a few other steps may have additionally contributed to the good results in this study. The participants were trained to avoid irregular breathing or sudden deep breathing during the examination. Sixteen *b*-values were used, which is at the up end of *b*-value number compared with published results on IVIM reproducibility (table 6). We were able use a mean of 5.3 slices for reproducibility calculation, which is more than the number of slices used in most of the published papers on reproducibility (11-19, 21, 23-30). The signal measurement was ROI-based method, and the IVIM parameters were calculated based on the mean signal intensity of the whole ROI. ROI-based approach allows assessing the plots of signal measurements and fitted curves for each slices, while this is not possible for pixel-based method when IVIM parameters are generated on parametric maps.

Dslow for liver parenchyma has been shown to have best reproducibility, followed by PF, and then Dfast. Interestingly, we found a slightly better reproducibility using threshold *b*-value =50 or 80 s/mm^2 for Dslow and PF estimation than threshold *b*-value=200s/mm^2. On the contrary, a higher *b*-value threshold (such as 200 s/mm^2) can lead to better reproducibility for Dfast (tables 4 and 5). This may be partially due to more of *b*-values were included for Dslow calculation using lower *b*-value threshold. This highlights the importance of acquiring with sufficient b-value numbers. Note instead of the commonly used value of *b*=200 s/mm^2, recently Li et al proposed *b*-value=60 s/mm^2 as the threshold value for separating diffusion and perfusion compartments for liver fibrosis detection (7). It should be noted that Dslow is less sensitive to pathologies than perfusion compartment measurement. For example, the study of Luciani et al (32) reported that D_{slow} does not differ significantly between healthy and cirrhotic livers, and cirrhotic livers is mainly

associated D_{fast} reduction. Andreou et al. (4) reported no statistical difference between Dslow of liver parenchyma (=1.1 $\times 10^{-3}$ mm2/s) and liver metastasis (=1 $\times 10^{-3}$ mm2/s). Recent literature review also showed that among nine patient studies, only D_{fast}, despite being the least stable, consistently demonstrated liver fibrosis is progressively associated a reduced measurement (2). Therefore, it is critically important to measure the perfusion compartment which can be reflected by Dfast, PF, as well as Dslow computed with low threshold *b*-value to include perfusion compartment.

Our study has some limitations. The volunteer population in this study included only young healthy subjects. While our results may be applicable to diffused liver diseases such as hepatic fibrosis, how our approach can be applicable to focal liver lesions will require additional studies. Secondly, we didn't ask volunteers to fast before examination, while the hepatic flow may vary depending on the fasted/prandial status (33). The reproducibility of IVIM parameters could therefore may be better when the subjects are scanned in fasted status. The data cleaning criteria presented in this study remains subjective, not precisely defined and was not automatized. An objective assessment method, including machine-based recognizing anatomical landmarks and estimation of quantification of scattering of the MRI signal intensity vs *b*-value relationship, are being explored in our laboratory to automatically and consistently assess the data acquisition quality. Another limitation is that we only tested segment-unconstrained analysis of IVIM data, while segment-unconstrained analysis remains till now the most popular approach for IVIM analysis (7), it has been suggested that Bayesian probability may perform better in fitting consistency (34, 35). Finally, in this study, five out of 54 scanned could not be used for analysis, leading to a success rate of 89%. A better sequence design allowing over-sampling of the focused liver parenchyma regions is expected to minimize the failure rate, and potentially because of the increased number of 'sufficient quality slices' available for averaging, will further increase the measure reproducibility.

In conclusion, we demonstrated the proof-of-principle that the scan-rescan reproducibility of IVIM parameters can potentially be good. This understanding is important for further developing IVIM technique for diagnostic clinical application.

Fig legend

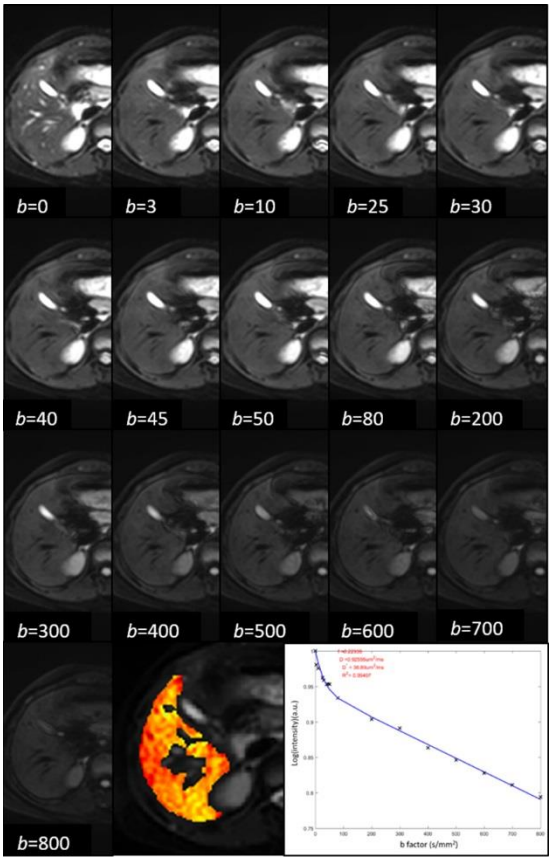

Figure 1, A IVIM image series of 'good quality', No evidential motion or artifacts could be seen in more than two images. ROI was drawn to cover as much as possible of liver parenchyma of right lobe, while avoiding signal vessels which are noticeable in the central parts of parenchyma and also in the area close to the gallbladder in this slice. The signal intensity and b-value curve has a R^2>0.96 for ROI-based fitting. The signal decay is consistent with b-value changes, and no obvious outlier value is noted.

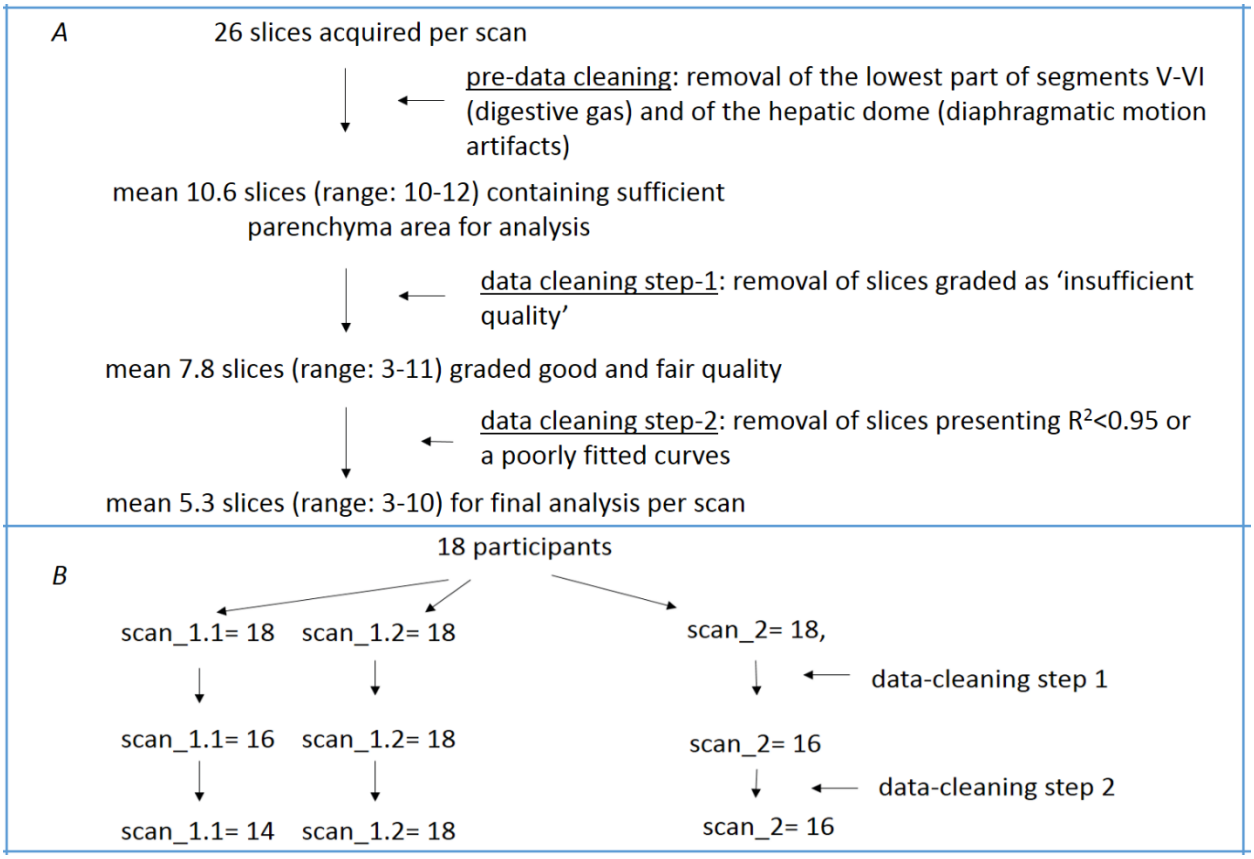

Figure 2, A: image-data cleaning process flow diagram. B: number of volunteer subjected included for analysis at each steps.

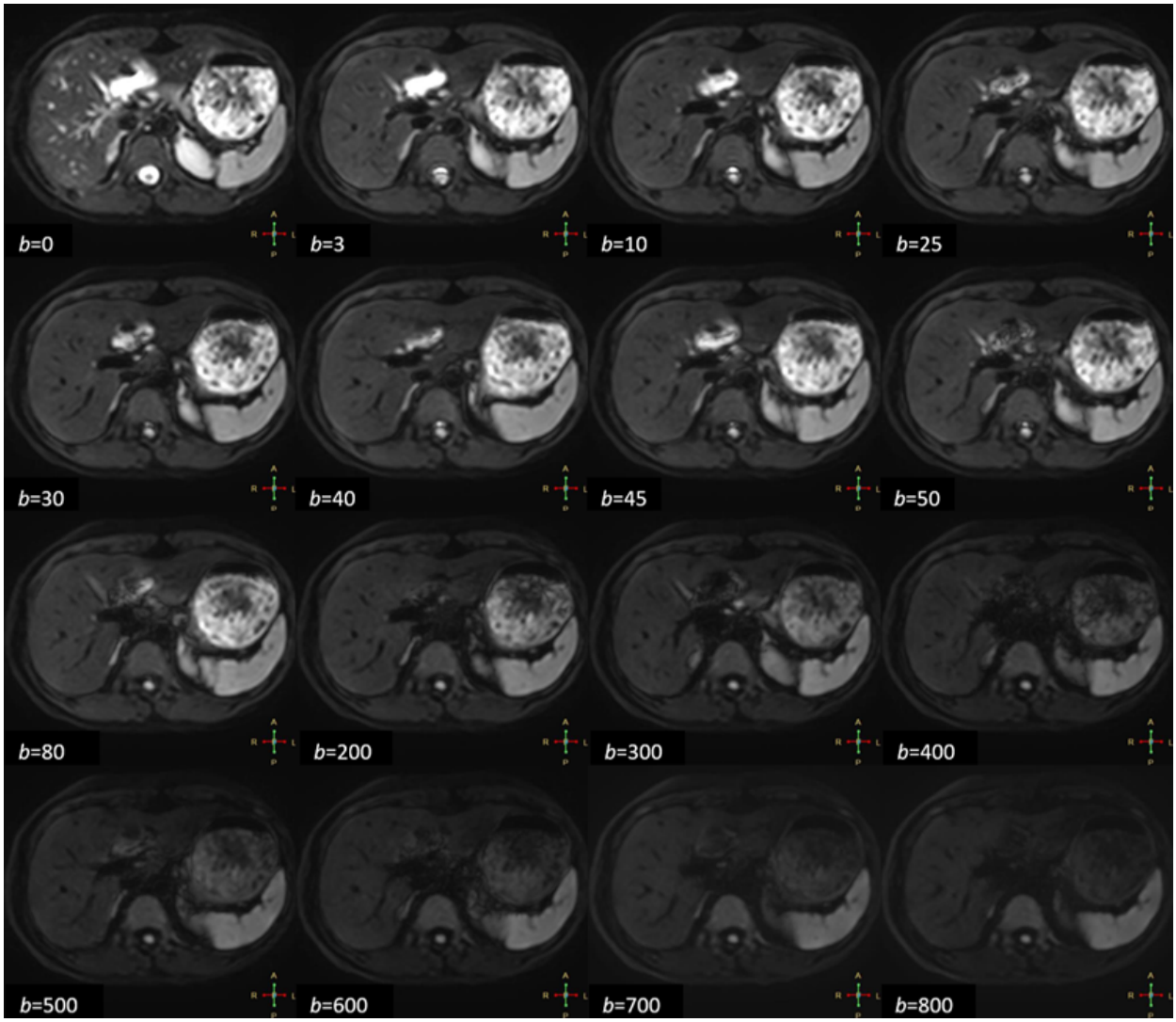

Figure 3, An example of excluded image series due to inter-image motion. Left kidney size varies among different images, and the posterior branch of the left portal branch vein can be seen in some images but not in other images.

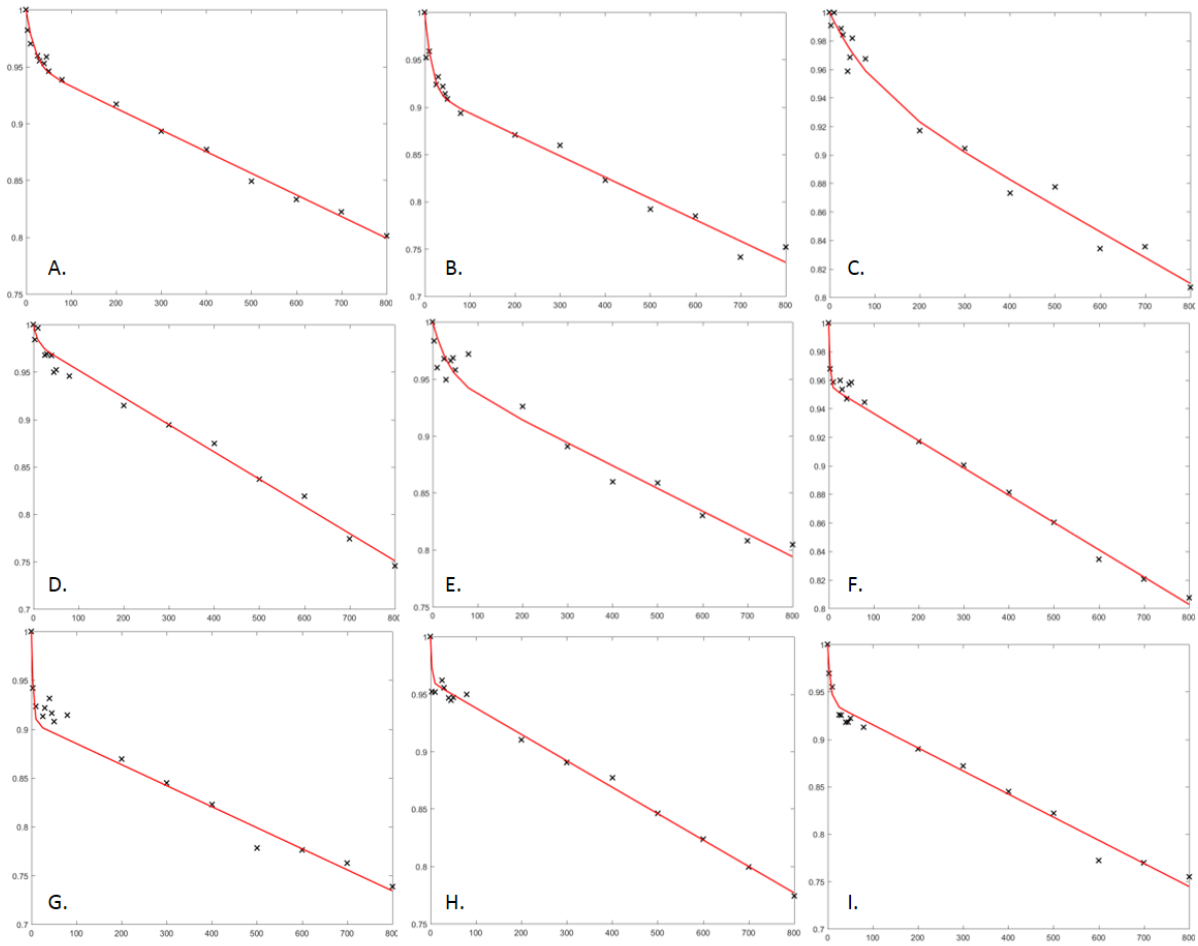

Fig 4, Examples of well fitted (A), acceptable (B) and unacceptable (C-I) signal intensity and b-value curves. Plots C, D, E, H show MRI signal increases and decreases erratically between *b*=0 s/mm^2 and *b*=80 s/mm^2. For plots E, F, I, H, at least 3 consecutive data points are outliers under or above the fitted curve between b=0 s/mm^2 and b=80 s/mm^2. Plots F, G, H present sharp signal drop at b-value=0 and b=3 s/mm^2, and therefore unreasonably high Dfast values.

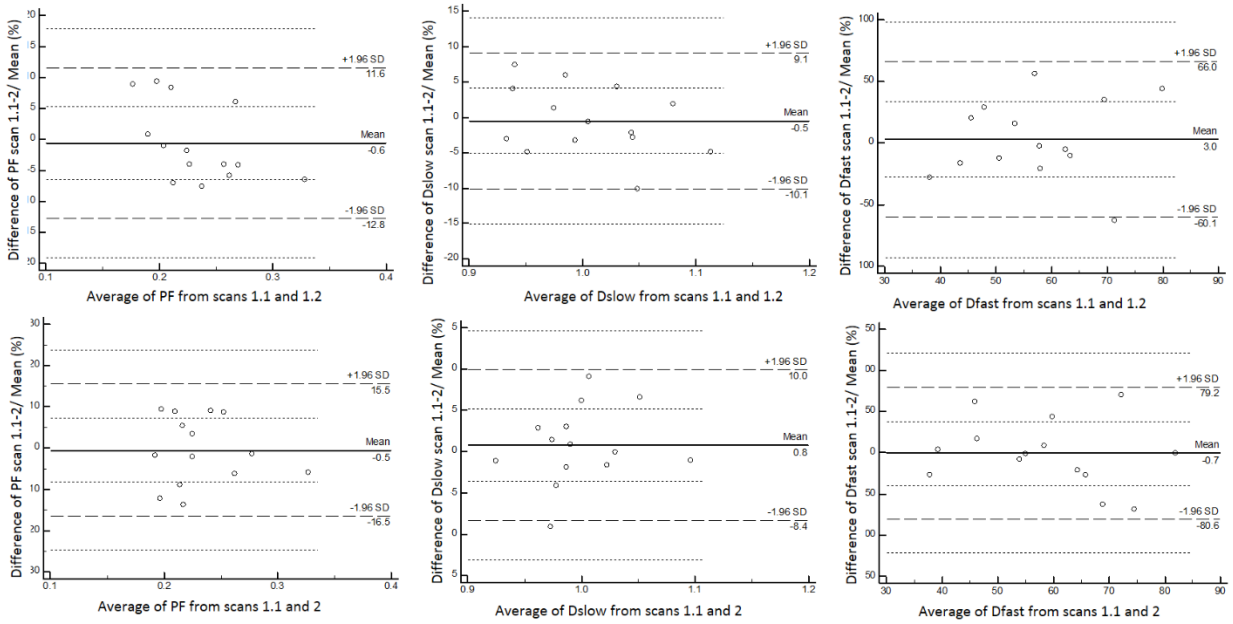

Fig 5, Bland–Altman plots of scan-rescan repeatability (scan 1.1 vs scan 1.2) and reproducibility (scan 1.1 vs scan 2).

Table 1. PF values in healthy volunteers with *b*-value=80 s/mm^2 threshold and the number of included slices per examination

	PF value (No. of included slices, CoV_{slices})		
Case No.	Scan 1.1	Scan 1.2	Scan 2
1	22.23 (5, 13.76)	23.14 (3, 12.98)	22.70 (4, 19.13)
2	25.40 (3, 28.63)	26.93 (4, 25.65)	27.03 (3, 9.08)
3	31.72 (3, 15.98)	33.86 (3, 17.64)	33.63 (4, 2.80)
4	25.20 (6, 14.85)	26.24 (3, 9.48)	23.00 (3, 7.82)
5	22.88 (3, 15.45)	24.69 (5, 6.84)	22.11 (7, 8.54)
6	27.53 (4, 12.44)	25.91 (3, 10.79)	27.91 (3, 15.76)
7	22.25 (4, 16.57)	22.65 (5, 16.08)	21.09 (4, 21.31)
8	20.25 (8, 13.07)	20.46 (9; 13.36)	23.23 (7, 17.88)
9	20.47 (6, 9.36)	21.97 (6; 15.99)	22.38 (5, 7.89)
10	26.38 (6, 13.86)	27.50 (5; 8.85)	24.19 (7, 12.21)
11	18.44 (8, 11.91)	16.87 (7; 32.05)	20.85 (5, 10.83)
12	20.73 (9, 11.61)	18.88 (10; 19.29)	18.86 (9, 11.20)
13	21.93 (6, 12.93)	20.18 (6; 11.37)	20.05 (7, 18.53)
14	19.09 (6, 8.12)	18.93 (4; 14.96)	19.44 (4, 10.15)

PF: in %; CoV_{slices} (in %): coefficient of variation of PF across the slices included after full data-cleaning

Table 2. Dslow values in healthy volunteers with *b*-value=80 s/mm^2 threshold

Case No.	Scan 1.1	Scan 1.2	Scan 2
1	0.96 (7.72)	0.92 (3.36)	1.00 (2.93)
2	1.09 (7.65)	1.07 (4.03)	1.10 (7.66)
3	0.98 (3.41)	0.91 (2.30)	0.95 (3.38)
4	0.98 (8.22)	0.97 (6.27)	0.97 (3.17)
5	1.05 (0.59)	1.01 (5.21)	0.96 (4.68)
6	1.03 (5.12)	1.06 (10.49)	1.03 (1.64)
7	1.03 (6.72)	1.05 (14.31)	0.97 (4.44)
8	0.98 (3.79)	1.01 (7.48)	1.00 (3.05)
9	1.00 (3.93)	1.01 (8.95)	0.97 (2.48)
10	1.01 (7.02)	0.96 (10.27)	1.03 (11.07)
11	0.93 (3.71)	0.97 (3.60)	1.02 (7.59)
12	1.09 (6.28)	1.14 (6.58)	1.02 (3.53)
13	1.00 (1.40)	1.10 (7.34)	0.99 (2.99)
14	0.92 (4.72)	0.95 (4.98)	0.93 (7.11)

Dslow: in (x10^{-3}mm^2/s); CoV_{slices} (in %): coefficient of variation of Dslow across the slices included after full data-cleaning.

Table 3. Dfast values in healthy volunteers with *b*-value=80 s/mm^2 threshold

Case No.	Scan 1.1	Scan 1.1	Scan 2
1	40.02 (19.61)	47.17 (17.58)	38.70 (20.91)
2	50.26 (26.07)	40.92 (8.17)	42.49 (22.08)
3	54.79 (24.04)	40.95 (25.25)	55.28 (12.61)
4	81.80 (46.55)	57.29 (67.71)	82.07 (119.20)
5	60.82 (44.72)	64.12 (57.65)	55.76 (21.60)
6	48.91 (28.25)	93.83 (74.85)	100.25 (90.14)
7	32.77 (32.79)	43.36 (37.18)	42.97 (11.26)
8	51.90 (32.92)	63.94 (38.19)	55.95 (31.65)
9	57.68 (24.53)	49.25 (28.48)	71.08 (27.25)
10	97.62 (46.29)	62.33 (34.16)	46.72 (25.86)
11	60.14 (64.20)	66.65 (47.33)	31.76 (29.84)
12	72.91 (42.31)	40.97 (24.16)	46.77 (34.68)
13	47.37 (18.40)	53.80 (31.23)	90.40 (54.72)
14	57.04 (41.15)	58.62 (24.75)	74.67 (56.94)

Dfast: in (x10^{-3}mm^2/s); CoV_{slices} (in %): coefficient of variation of Dfast across the slices included after full data-cleaning

Table 4. scan-rescan repeatability of IVIM parameters*

		Threshold 50	Threshold 80	Threshold 200
PF	Mean CoV (%, range)	2.40 (0.13 – 7.20)	3.81 (0.62 – 6.61)	4.89 (0.00 – 11.49)
	Within-subject CoV (%)	3.20	4.24	6.32
	Mean Diff % (BA – LA, %)	-0.9 (-10.1 – 8.3)	-0.6 (-12.8 – 11.6)	-3.0 (-19.1 – 13.0)
Dslow	Mean CoV (%, range)	2.07 (0.30 – 5.10)	2.86 (0.43 – 7.18)	3.65 (0.51 – 10.34)
	Within-subject CoV (%)	2.47	3.36	4.43
	Mean Diff % (BA – LA, %)	-0.7 (-7.6 – 6.3)	-0.5 (-10.1 – 9.1)	0.9 (-12.1 – 13.9)
Dfast	Mean CoV (%, range)	23.71 (5.01 – 49.02)	18.16 (1.93 – 44.50)	17.07 (0.47 – 50.39)
	Within-subject CoV (%)	28.28	24.88	26.50
	Mean Diff % (BA – LA, %)	0.9 (-77.2 – 79.0)	3.0 (-60.1 – 66.0)	7.6 (-54.6 – 69.9)

Diff: difference, BA-LA: Bland-Altman 95% limits of agreement

Table 5. Scan-rescan reproducibility of IVIM parameters*

		Threshold 50	Threshold 80	Threshold 200
PF	Mean CoV (%, range)	5.68 (0.75 – 10.94)	4.91 (0.97 – 9.71)	5.98 (0.60 – 13.87)
	Within-subject CV (%)	6.59	5.38	6.94
	Mean Diff % (BA – LA, %)	-0.7 (-19.9 – 18.5)	-0.5 (-16.5 – 15.5)	-1.8 (-22.2 – 18.6)
Dslow	Mean CoV (%, range)	2.99 (0.61 – 6.75)	2.48 (0.05 – 6.45)	3.10 (0.58 – 6.00)
	Within-subject CV (%)	3.72	3.24	3.53
	Mean Diff % (BA – LA, %)	0.8 (-9.7 – 11.3)	0.8 (-8.4 – 10.0)	1.3 (-8.4 – 11.0)
Dfast	Mean CoV (%, range)	21.60 (1.15 – 67.31)	21.18 (0.23 – 49.87)	19.33 (2.31 – 72.36)
	Within-subject CV (%)	34.33	30.89	35.93
	Mean Diff % (BA – LA, %)	3.2 (-84.0 – 90.3)	-0.7 (-80.6 – 79.2)	-4.3 (-80.1 – 71.5)

Diff: difference, BA-LA: Bland-Altman 95% limits of agreement

Table 6. Summary of liver IVIM parameters reproducibility in liver parenchyma.

Paper	Subjects	Repeatability or Reproducibility				*b*-value No.	Breathing and Post processing
		Statistics	PF	Dslow	Dfast		
Andreou *et al.*, (2013) (4)	14 Pts (tumor)	BA – LA	-24.3–25.1	-5.12–8.09	-31.2–59.1	8	RT V-by-V LSF
Dyvorne *et al.*, (2013) (13)	6 HV 4 Pts (CLD)	$_{w}$CoV (range)	11.5 (1.5 – 22)	4.7 (1.2 – 21.9)	30.1 (9.7 – 50.1)	16	FB ROI Bayesian
Kakite *et al.*, (2015) (15)	11 Pts (CLD)	CoV (range) BA-LA	25.3 (9.3 – 84.9) -74.6 – 56.4	13.2 (1.2 – 28.5) -32.8 – 28.4	59.0 (2.4 – 121.3) -151.2 – 132.7	16	FB ROI Bayesian
Park *et al.*, (2017) (19)	12 HV	$_{w}$CoV	9.65	4.05	29.52	9	RT V-by-V LSF (constrains)
Patel *et al.*, (2010) (23)	One Pts (benign lesion) 4 HV	CoV	7.7	3.8	14.6	9	RT ROI Seg-anal/ *b*=200
Jajamovich *et al.*, (2014) (14)	11 HV 19 Pts (CLD)	CoV BA – LA	20.4 -76.9 – 69.0	15.2 -63.5 – 69.7	51.6 -200.5 – 184.5	16	FB ROI Bayesian
Hectors *et al.*, (2016) (26)	8 Pts (CLD)	CoV (range) BA – LA	5.2 (0.0 – 11.8) -20.5 – 21.4	7.3 (0.4 – 19.8) -31.3 – 26.8	21.4 (4.6 – 42.8) -73.8 – 73.7	16	RT and FB ROI Bayesian
Lee *et al.*, (2015) (24)	12 HV	$_{w}$CoV (95% CI)	ET 9.86 (7.13 – 12.59)	RT 3.19 (1.83 – 4.54)	ET 71.87 (47.53 – 96.21)	9	ET and RT V-by-V LSF
Gurney-Champion *et al.*, (2016) (21)	intra-session: 14 HV inter-session: 16 HV	$_{w}$CoV	Intra-session: 32 Inter-session: 34	Intra-session: 9 Inter-session: 10		14	RT LSF with Dfast fixed

All statistics are in % and the results shown are for liver parenchyma except Andreou et al' results. When different acquisition (FB vs RT, etc.) or post-processing (full-fitting, segmented fitting, with or without constraint, etc.) schemes were compared, only the best results are presented in this table. Subject number corresponds to the number of subject included for reproducibility analysis. HV: Heathy Volunteers, Pts: Patients, CLD: chronic liver disease, CoV: Mean of the coefficient of variation, $_{w}$CoV: Within-subject coefficient of variation, BA – LA: Bland-Altman 95% limits of agreement; 95% CI: 95% confidence intervals; RT: Free Breathing, RT: respiratory triggering (including navigator triggering), ET: Electrocardiography triggering, V-by-V: Voxel by voxel analysis, ROI: ROI-analysis, Seg-anal/ *b*=150: Segment fitting method, with *b*-value threshold of 150 s/mm^2, LSF: least square fitting method, Bayesian: Bayesian fitting method.

REFERENCES

1. Yamada I, Aung W, Himeno Y, et al. Diffusion coefficients in abdominal organs and hepatic lesions: evaluation with intravoxel incoherent motion echo-planar MR imaging. Radiology 1999;210:617–623.

2. Li YT, Cercueil JP, Yuan J, et al. Liver intravoxel incoherent motion (IVIM) magnetic resonance imaging: a comprehensive review of published data on normal values and applications for fibrosis and tumor evaluation. Quant Imaging Med Surg 2017;7:59–78.

3. Koh DM, Collins DJ, Orton MR. Intravoxel incoherent motion in body diffusion-weighted MRI: reality and challenges. AJR Am J Roentgenol 2011;196:1351-1361

4. Andreou A, Koh DM, Collins DJ, et al. Measurement reproducibility of perfusion fraction and pseudodiffusion coefficient derived by intravoxel incoherent motion diffusion-weighted MR imaging in normal liver and metastases. Eur Radiol 2013;23:428–434.

5. Padhani AR, Liu G, Mu-Koh D, et al. Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations. Neoplasia 2009;11:102–125.

6. Turner R, Le Bihan D, Maier J, et al. Echo-planar imaging of intravoxel incoherent motion. Radiology 1990;177:407–414.

7. Li YT, Lu PX, Huang H, Leung JC, et al. Dependence of Intravoxel Incoherent Motion MR threshold b-value selection for separating perfusion and diffusion compartment and liver fibrosis diagnostic performance. BioRxiv doi: http://dx.doi.org/10.1101/164129.

8. Lu PX, Huang H, Yuan J, et al. Decreases in Molecular Diffusion, Perfusion Fraction and Perfusion-Related Diffusion in Fibrotic Livers: A Prospective Clinical Intravoxel Incoherent Motion MR Imaging Study. PLoS ONE 2014; 9: e113846.

9. Yuan J, Wong OL, Lo GG, et al. Statistical assessment of bi-exponential diffusion weighted imaging signal characteristics induced by intravoxel incoherent motion in malignant breast tumors. Quant Imaging Med Surg 2016;6:418–429.

10. Li YT, Huang H, Zhuo Z, et al. Bi-phase age-related brain gray matter magnetic resonance T1ρ relaxation time change in adults. Magn Reson Imaging 2017;39:200–205.

11. Moteki T, Horikoshi H. Evaluation of hepatic lesions and hepatic parenchyma using diffusion-weighted echo-planar MR with three values of gradient b-factor. J Magn Reson Imaging 2006;24:637–645.

12. Dyvorne HA, Galea N, Nevers T, et al. Diffusion-weighted imaging of the liver with multiple b values: effect of diffusion gradient polarity and breathing acquisition on image quality and intravoxel incoherent motion parameters--a pilot study. Radiology 2013;266:920–929.

13. Dyvorne H, Jajamovich G, Kakite S, et al. Intravoxel incoherent motion diffusion imaging of the liver: optimal b-value subsampling and impact on parameter precision and reproducibility. Eur J Radiol 2014 83:2109–2113.

14. Jajamovich GH, Dyvorne H, Donnerhack C, et al. Quantitative liver MRI combining phase contrast imaging, elastography, and DWI: assessment of reproducibility and postprandial effect at 3.0 T. PloS One 2014;9:e97355.

15. Kakite S, Dyvorne H, Besa C, et al. Hepatocellular carcinoma: short-term reproducibility of apparent diffusion coefficient and intravoxel incoherent motion parameters at 3.0T. J Magn Reson Imaging 2015;41:149–156.

16. Cohen AD, Schieke MC, Hohenwalter MD, et al. The effect of low b-values on the intravoxel incoherent motion derived pseudodiffusion parameter in liver. Magn Reson Med 2015;73:306–311.

17. Cui Y, Dyvorne H, Besa C, et al. IVIM Diffusion-weighted Imaging of the Liver at 3.0T: Comparison with 1.5T. Eur J Radiol Open 2015;2:123–128.

18. França M, Martí-Bonmatí L, Alberich-Bayarri Á, et al. Evaluation of fibrosis and inflammation in diffuse liver diseases using intravoxel incoherent motion diffusion-weighted MR imaging. Abdom Radiol 2017 ;42:468–77.

19. Park HJ, Sung YS, Lee SS, et al. Intravoxel incoherent motion diffusion-weighted MRI of the abdomen: The effect of fitting algorithms on the accuracy and reliability of the parameters. J Magn Reson Imaging 2017;45:1637–1647.

20. Wang YXJ, Deng M, Li YT, et al. A Combined Use of Intravoxel Incoherent Motion MRI Parameters Can Differentiate Early-Stage Hepatitis-b Fibrotic Livers from Healthy Livers. SLAS Technol 2017 doi: 10.1177/2472630317717049

21. Gurney-Champion OJ, Froeling M, Klaassen R, et al. Minimizing the Acquisition Time for Intravoxel Incoherent Motion Magnetic Resonance Imaging Acquisitions in the Liver and Pancreas. Invest Radiol. 2016;51:211–20.

22. Cercueil JP, Petit JM, Nougaret S, et al. Intravoxel incoherent motion diffusion-weighted imaging in the liver: comparison of mono-, bi- and tri-exponential modelling at 3.0-T. Eur Radiol. 2015;25:1541-50.

23. Patel J, Sigmund EE, Rusinek H, et al. Diagnosis of cirrhosis with intravoxel incoherent motion diffusion MRI and dynamic contrast-enhanced MRI alone and in combination: preliminary experience. J Magn Reson Imaging 2010;31:589–31600.

24. Lee Y, Lee SS, Kim N, et al. Intravoxel incoherent motion diffusion-weighted MR imaging of the liver: effect of triggering methods on regional variability and measurement repeatability of quantitative parameters. Radiology 2015;274:405–415.

25. Shan Y, Zeng M, Liu K, et al. Comparison of Free-Breathing With Navigator-Triggered Technique in Diffusion Weighted Imaging for Evaluation of Small Hepatocellular Carcinoma: Effect on Image Quality and Intravoxel Incoherent Motion Parameters. J Comput Assist Tomogr 2015;39:709–715.

26. Hectors SJ, Wagner M, Besa C, et al. Intravoxel incoherent motion diffusion-weighted imaging of hepatocellular carcinoma: Is there a correlation with flow and perfusion metrics obtained with dynamic contrast-enhanced MRI? J Magn Reson Imaging 2016;44:856–864.

27. Barbieri S, Donati OF, Froehlich JM, et al. Comparison of Intravoxel Incoherent Motion Parameters across MR Imagers and Field Strengths: Evaluation in Upper Abdominal Organs. Radiology. 2016;279:784–794.

28. Jerome NP, d'Arcy JA, Feiweier T, et al. Extended T2-IVIM model for correction of TE dependence of pseudo-diffusion volume fraction in clinical diffusion-weighted magnetic resonance imaging. Phys Med Biol. 2016;61:N667–N680.

29. Kurugol S, Freiman M, et al. Spatially-constrained probability distribution model of incoherent motion (SPIM) for abdominal diffusion-weighted MRI. Med Image Anal. 2016;32:173–183.

30. Park HJ, Sung YS, Lee SS, et al. Intravoxel incoherent motion diffusion-weighted MRI of the abdomen: The effect of fitting algorithms on the accuracy and reliability of the parameters. J Magn Reson Imaging 2017;45:1637–1647.

31. Zhang Q, Wang YX, Ma HT, et al. Cramér-Rao bound for Intravoxel Incoherent Motion Diffusion Weighted Imaging fitting. Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf. 2013;2013:511–514.

32. Luciani A, Vignaud A, Cavet M, et al. Liver cirrhosis: intravoxel incoherent motion MR imaging--pilot study. Radiology 2008;249:891–899.

33. Sabbá C, Ferraioli G, Genecin P, et al. Evaluation of postprandial hyperemia in superior mesenteric artery and portal vein in healthy and cirrhotic humans: an operator-blind echo-Doppler study. Hepatology. 1991;13:714–718.

34. Neil JJ, Bretthorst GL. On the use of Bayesian probability theory for analysis of exponential decay data: an example taken from intravoxel incoherent motion experiments. Magn Reson Med 1993;29:642–647.

35. Barbieri S, Donati OF, Froehlich JM, et al. Impact of the calculation algorithm on biexponential fitting of diffusion-weighted MRI in upper abdominal organs. Magn Reson Med 2016;75:2175–2184.