Understanding diffusion-weighted MRI analysis: Repeatability and performance of diffusion models in a benign breast lesion cohort

Neil Peter Jerome1,2 | Igor Vidi1,2 | Liv Egnell2,3 | Torill E. Sjøbakk1,2 | Agnes Østlie4 | Hans E. Fjøsne4,5 | Pål Erik Goa3 | Tone F. Bathen1,2

1Department of Circulation and Medical Imaging, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
2Clinic of Radiology and Nuclear Medicine, St. Olavs Hospital, Trondheim, Norway
3Department of Physics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
4Department of Radiology, St. Olavs Hospital, Trondheim, Norway
5Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Correspondence
Neil Peter Jerome, Department of Circulation and Medical Imaging, Norwegian University of Science and Technology (NTNU), 7030 Trondheim, Norway.
Email: neil.p.jerome@ntnu.no

Funding information
Norwegian University of Science and Technology, Grant/Award Number: 90065000; Central Norway Regional Health Authority

Diffusion-weighted MRI (DWI) is an important tool for oncology research, with great clinical potential for the classification and monitoring of breast lesions. The utility of parameters derived from DWI, however, is influenced by specific analysis choices. The purpose of this study was to critically evaluate repeatability and curve-fitting performance of common DWI signal representations, for a prospective cohort of patients with benign breast lesions. Twenty informed, consented patients with confirmed benign breast lesions underwent repeated DWI (3 T) using: sagittal single-shot spin-echo echo planar imaging, bipolar encoding, TR/TE: 11,600/86 ms, FOV: 180 x 180 mm, matrix: 90 x 90, slices: 60 x 2.5 mm, iPAT: GRAPPA 2, fat suppression, and 13 b-values: 0–700 s/mm². A phase-reversed scan (b = 0 s/mm²) was acquired for distortion correction. Voxel-wise repeat-measures coefficients of variation (CoVs) were derived for monoexponential (apparent diffusion coefficient [ADC]), biexponential (intravoxel incoherent motion: f, D, D*) and stretched exponential (α, DDC) across the parameter histograms for lesion regions of interest (ROIs).

Goodness-of-fit for each representation was assessed by Bayesian information criterion. The volume of interest (VOI) definition was repeatable (CoV 13.9%). Within lesions, and across both visits and the cohort, there was no dominant best-fit model, with all representations giving the best fit for a fraction of the voxels. Diffusivity measures from the signal representations (ADC, D, DDC) all showed good repeatability (CoV < 10%), whereas parameters associated with pseudodiffusion (f, D*) performed poorly (CoV > 50%). The stretching exponent α was repeatable (CoV < 12%). This pattern of repeatability was consistent over the central part of the parameter percentiles. Assumptions often made in diffusion studies about analysis choices will influence the detectability of changes, potentially obscuring useful information. No single signal representation prevails within or across lesions, or across repeated visits;
parameter robustness is therefore a critical consideration. Our results suggest that stretched exponential representation is more repeatable than biexponential, with pseudodiffusion parameters unlikely to provide clinically useful biomarkers.

**KEYWORDS**
breast cancer, breast neoplasms, diffusion magnetic resonance imaging, biomarkers, magnetic resonance imaging, repeatability

1 | INTRODUCTION

Diffusion-weighted imaging (DWI) is an established component of magnetic resonance imaging (MRI) protocols in oncology.\(^1\)–\(^3\) Deriving meaningful diffusion parameters from signal curves, however, requires an appreciation of the limitations of the diffusion models being used, and what relationship they have to tissue microstructure. For clinical evaluation of breast cancer, DWI studies have shown an ability to characterise lesions\(^4,5\) and have also displayed sensitivity for measuring and predicting response to treatment.\(^6\)–\(^9\) Commonly, DWI is quantified using the apparent diffusion coefficient (\(ADC\)), and is used as an indication of tissue cellularity. It is important to remember that \(ADC\) is an empirical parameter based on an assumption of Gaussian (random) diffusion and is an oversimplification in many tissues.

More complex representations of DWI signal attempt to quantify non-Gaussian behaviour as either perfusion components to the signal (the intravoxel incoherent motion, or IVIM model),\(^10\) or allow for a distribution of diffusion coefficients within each voxel (the stretched exponential representation).\(^11\) These representations, summarised in Table 1, have the potential to offer more sensitive imaging biomarkers for treatment response,\(^12,13\) specific to changes in vascularity, compartmentalised and distributed diffusion processes, and contributions from the extracellular matrix and collagen.

Using more advanced DWI makes acquisition and analysis more challenging, requiring the acquisition of a larger number of \(b\)-values, and currently additional time for analysis performed outside of the normal clinical data workflow. The representations themselves also have limitations that are often overlooked. Assumptions in models may be violated in certain tissues, and DWI signal is known to be sensitive to acquisition parameters, such as diffusion time (\(\Delta\)) and echo time (TE), which are often not standardised.\(^14,15\) It is thus important that the cost of performing extended DWI protocols is critically examined in the context of practical utility. While the sensitivity and specificity of DWI-derived biomarkers are critical for clinical decision-making, the repeatability of these parameters is similarly crucial for translation into clinical use.\(^16\) Unfortunately, incorporating repeatability data into clinical studies is challenging and places an extra demand on patients.\(^17\) This is especially true for patients with malignant lesions that may be changing faster and where there is less time for additional MR examinations before treatment.

In this study, we report the findings from fitting the monoexponential (ADC), biexponential (IVIM), and stretched exponential signal representations to diffusion signal acquired with multiple \(b\)-values in a cohort of patients with benign breast lesions, and use repeated measurements to assess repeatability. In this setting, the benign breast lesions present a useful and accessible model system, which may be considered simpler and better behaved than malignant lesions, with which to highlight the complexity of diffusion analysis in a wider context.

2 | EXPERIMENTAL

2.1 | Patient cohort and MR protocol

This prospective study was approved by the Regional committee for Medical and Health Research Ethics (REC) of central Norway (identifier 2011/568). Twenty-one informed and consenting patients (median age 26, range 19–50 years) with benign breast lesions (fibroadenomas, confirmed by histopathological assessment with core needle biopsy, or clinical follow-up) were recruited from September 2016 to October 2017. Alongside a standard clinical protocol, including anatomical (\(T_2\)-weighted) and dynamic contrast-enhanced imaging (DCE-MRI, with one baseline and seven postcontrast images at a time resolution of 1 min), the clinical MRI examination (‘visit 1’) included a multiple \(b\)-value DWI protocol, which is described below. The diffusion protocol was repeated exactly during an additional examination (‘visit 2’) after 7 (median; range 2–18) days with no intervening treatment. The hospital where this study was performed does not schedule examinations with regard to the menstrual cycle, and so these data were not recorded. All imaging was performed on a clinical 3-T scanner (Skyra, Siemens Healthineers) equipped with a 16-channel breast coil. The repeatability thus explicitly represents elements from underlying physiology and from patient set-up and scanner performance, to most accurately reflect clinical use.
2.2 Diffusion analysis

Diffusion images were corrected for distortion using the method reported by Teruel et al. Volumes of interest (VOIs) were drawn on a diffusion-weighted image for all slices of the tumour in both examinations, avoiding cystic regions and using DCE and T2 images as a reference. VOIs were drawn by a scientist with 12 years’ experience of cancer imaging (NPJ) and validated by an experienced breast radiologist (AØ). The repeatability of the delineated volume was assessed by a log-normalised repeat-measures coefficient of variation (CoV) suitable for constrained positive values.

Multiple representations of the DWI signal decay (see Table 1 for details of signal representations) were fitted to the DWI data on a voxel-wise basis, and also for the mean VOI signal in each tumour. The fitted representations included an exponential decay (giving ADC), a biexponential decay (as per the IVIM model, giving the volume fraction of pseudodiffusion compartment with coefficient $D^*$ in addition to the true diffusion coefficient $D$), and stretched exponential decay (giving the stretching exponent $\alpha$ and distributed diffusion coefficient DDC). For these representations, all $b$-value data were used in the fitting. All signal fitting was performed using a Levenberg–Marquardt algorithm using MATLAB (Mathworks, Natick, MA, USA); for IVIM, initial parameter estimates were provided by a segmented approach; these results are also reported. No parameter restraints were used, to avoid unintentional bias from voxels returning limiting values, and to allow for critical assessment of the information content of the acquired data.

| Signal representation | Parameters | Assumptions | Interpretation |
|-----------------------|------------|-------------|---------------|
| Monoexponential (ADC) | $S_b = S_0 \cdot \exp(-b \cdot ADC)$ | ADC apparent diffusion coefficient | Diffusion is isotropic and Gaussian in nature, giving monoexponential signal decay with $b$-value. ADC is the diffusion coefficient in an unimpeded single compartment system that would lead to the monoexponential DWI signal decay that best matches the observed signal. | ADC is sensitive to impendence of water, arising from tissue structure (cellularity, tortuosity, extracellular matrix, etc.). ADC is non-specific and contains influence from all factors (see above). |
| Biexponential (IVIM) | $S_b = S_0 \cdot [f \cdot \exp(-b \cdot D^*) + (1-f) \cdot \exp(-b \cdot D)]$ | $f$ pseudodiffusion volume fraction, $D^*$ pseudodiffusion coefficient, $D$ true diffusion coefficient | Diffusion is isotropic. Two compartments: (1) true (Gaussian) compartment with volume fraction $1-f$ and empirical diffusion coefficient $D$; and (2) pseudodiffusion compartment of volume fraction $f$ with coefficient $D^*$. Perfusion in microcapillaries is equal in all directions (incoherent thus can be modelled as a pseudodiffusion (i.e. monoexponential)). | $D$ is diffusion-insensitive, thus is true tissue diffusion. $f$ reflects (but is not a direct measure of) tissue perfusion. $D^*$ reflects non-diffusion process, including microcapillary flow. |
| Stretched exponential | $S_b = S_0 \cdot \exp(-b \cdot DDC^*)$ | $\alpha$ diffusion exponent, DDC distributed diffusion coefficient | Diffusion is isotropic. Non-monoeponential signal decay can be modelled by a summation of a distribution of curves with a continuum of diffusion coefficients. | Gaussian diffusion is represented by $\alpha = 1$ (i.e. DDC = ADC). When $\alpha < 1$, diffusion is interpreted as non-Gaussian, with underlying distribution of coefficients summarised by DDC. Exponent $\alpha$ reflects spread of coefficients, i.e. intravoxel diffusion character spread. |

Note: for all representations of diffusion-weighted MRI (DWI) signal, $S_0$ is the signal intensity in the absence of diffusion weighting (i.e. $b = 0$ s.mm$^{-2}$), but contains a term allowing for T2 decay over the imaging echo time $TE$, which is assumed constant across all DW images. $b$ is the applied diffusion weighting.
2.3 | Statistical analyses

Following fitting, the optimal representation was determined by Bayesian information criterion (BIC), essentially comparing ‘goodness-of-fit’ using signal residuals but allowing comparison across representations that contain different numbers of variables (to avoid favouring overfitting). Repeat-measure CoV and 95% limits of agreement (LoA) were calculated for all DWI parameters for the values at every percentile across the VOI histogram for voxel-wise results, and for the mean and median VOI values.\textsuperscript{17,20} Median diffusion parameters were compared using a student’s t-test, with a significance threshold of 0.05. All statistical analyses were performed in MATLAB using standard functions.

3 | RESULTS

3.1 | Cohort characteristics and repeatability of lesion volumes

One patient withdrew before completing both scans, and data from one patient were excluded owing to incorrect slice positioning; eight patients had additional lesions within the imaging FOV that were analysed independently, and five lesions were excluded owing to being small or not clearly delineable on DWI, giving a total of 26 lesions with repeated datasets. The volumes of the benign lesions ranged from 0.29 to 25.39 cm\textsuperscript{3}, with CoV of 13.88%, indicating good VOI repeatability; the Bland–Altman plot is given in Figure 1, along with illustrative lesion delineations in Figure 2. Consistently good DWI SNR was observed by eye at maximum diffusion-weighting; two lesions randomly sampled from the cohort both gave SNR values above 35 for the highest \(b\)-value image (SNR at the maximum \(b\)-value for the lesion in Figure 2 was 37.57). SNR values this high are suitable for modelling with an assumed Gaussian noise distribution, and so do not require an explicit term in the model for (e.g. Rician) noise contribution. For summarised DWI results see Table 2, with full details per lesion in Table S1; note that the size of this cohort defies any useful inference from the values themselves, but that a range of values for each parameter is clearly seen across and within lesions, indicating both inter- and intratumour heterogeneity.

3.2 | Different signal representations in benign breast lesions

Figure 3 shows a summary of the DWI representations preferred by BIC for voxel-wise fitting of each tumour in each scan, as well as for VOI mean signal. Within each tumour, diffusion decay curves for different voxels are best described by different representations. Proportions of voxels preferring each representation are present in all lesions (as illustrated in Figure 4), with a substantial proportion of voxels best described by the monoexponential ADC model (range 19.2\%–59.8\%; the mode in 33 of 52 cases), with the remainder of non-Gaussian voxels showing that cohort data do not appear to dramatically favour either IVIM or stretched exponential. Averaging the VOI signal before modelling gives a lower BIC for non-Gaussian models in a majority (46/52 cases for mean VOI signal) of the DWI acquisitions in the study. Interestingly, there was only moderate concordance of the dominant model across repeated visits, for either voxel-wise or from VOI-averaged DWI fitting (14 and 12 out of 26, respectively).

3.3 | Repeatability of DWI biomarkers

CoVs for the medians of all DWI parameters for the cohort are given in Table 2; for the measures ADC, \(D\) and DDC, the returned CoV is low (<10\%). For the pseudodiffusion parameters (\(f\) and \(D^*\)), CoV is far higher (>50\%), whereas the CoV of the stretching exponent \(\alpha\) is also low. Using

![FIGURE 1 Bland–Altman plot showing lesion volume of interest (VOI) volume and repeatability details for the study cohort (solid and dashed lines are mean ± 1.96 std), with an acceptable coefficient of variation (CoV) of 13.88%](image-url)
the segmented IVIM method, rather than a full least-squares estimation initialised by these values, shows a reduction in the CoV for each parameter, although the CoV for \( f \) and \( D^* \) remain considerably high compared with \( D \). Figure 5 displays the CoVs for each parameter across each histogram percentile, illustrating the wider stability of diffusivity-related parameters (\( \text{ADC}, D \) and \( \text{DDC} \)) as low as the fifth percentile (expected to correspond to the regions with the highest cellular density). CoV values at the minima and maxima of all parameters are large, because they relate directly to the most extreme fitted values in each scan, and as such display the expected ‘bathtub’ profile with a stable centre and raised CoV at these extreme values. It is important to note that imposing boundaries on the fitting would show a misleadingly low CoV of zero should any

**FIGURE 2** Typical lesion regions of interest (ROIs) for repeated visits, outlined in yellow for three successive diffusion-weighted MRI (DWI) slices (\( b = 10 \text{ s.mm}^{-2} \)) alongside the corresponding \( T_2 \)-weighted images (top row; from visit 1). The representative signal-to-noise ratio (SNR) at the highest \( b \)-value was measured at 37.57 in this lesion (ROI mean signal over std of noise in corresponding air ROI).

**TABLE 2** Tumour cohort diffusion-weighted MRI (DWI) characteristics (mean ± std of volume of interest [VOI] median values), with coefficients of variation (CoV), 95% limits of agreement (LoA), and paired \( t \)-tests (all \( p > 0.05 \)). The units in this table are: volume (cm\(^3\)); apparent diffusion coefficient (\( \text{ADC} \)), \( D \), distributed diffusion coefficient (\( \text{DDC} \) \( (\times 10^{-3} \text{ mm}^2\text{s}^{-1}) \)); \( f \) and \( \alpha \) (no units); CoV and LoA (%)

| Lesion | Visit | Volume | Monoexponential | Segmented IVIM | Least squares IVIM | Stretched exponential |
|--------|-------|--------|----------------|---------------|-------------------|---------------------|
|        |       |        | \( \text{ADC} \) | \( f \) | \( D^* \) | \( D \) | \( f \) | \( D^* \) | \( D \) | \( \alpha \) | \( \text{DDC} \) |
| mean ± std | 1 | 6.19 ± 6.56 | 1.85 ± 0.36 | 0.09 ± 0.07 | 12.66 ± 4.33 | 1.69 ± 0.27 | 0.07 ± 0.20 | 11.23 ± 8.18 | 1.65 ± 0.23 | 0.85 ± 0.10 | 1.84 ± 0.36 |
| | 2 | 6.24 ± 6.87 | 1.84 ± 0.29 | 0.10 ± 0.04 | 14.88 ± 5.06 | 1.61 ± 0.26 | 0.11 ± 0.17 | 14.18 ± 8.17 | 1.60 ± 0.39 | 0.82 ± 0.08 | 1.83 ± 0.30 |
| \( p \) | | | | | | | | | | | |
| | | | | | | | | | | | |
| CoV | 13.88 | 9.41 | 97.35 | 29.32 | 4.74 | 122.33 | 51.06 | 7.33 | 11.81 | 9.45 |
| 95% LoA | (−31.8, | (−22.9, | (−89.7, | (−54.9, | (−12.3, | (−92.2, | (−73.7, | (−18.4, | (−27.8, | (−23.0, |
| | 46.6) | 29.7) | 861.4) | 121.7) | 14.0) | 131.7) | 279.7) | 22.5) | 38.6) | 29.9) |

Abbreviation: IVIM, intravoxel incoherent motion.
FIGURE 3  (A) Example images showing preferred signal representation choice (lowest Bayesian information criterion value) for each voxel for illustrative tumour slices in repeated visits (the same lesions and slices as Figure 2). There is no overall concordance of the best performing diffusion-weighted MRI (DWI) model within or across slices, or between visits. All cohort lesions contain voxels that are best described by each representation. (B) Real data and fitted curve for apparent diffusion coefficient (ADC), intravoxel incoherent motion (IVIM), and stretched representations for one voxel giving the lowest Bayesian information criterion (BIC) for each (left-to-right); the difference between residuals is not necessarily large.

FIGURE 4  Cumulative percentage graphs for diffusion-weighted MRI (DWI) modelling in each benign lesion in the cohort, showing the proportion of voxels with the lowest Bayesian information criterion (BIC) value for monoexponential apparent diffusion coefficient (ADC), biexponential intravoxel incoherent motion (IVIM), and stretched exponential across all $b$-values.
FIGURE 5  Repeat-measures coefficient of variation (CoV) at each volume of interest (VOI) percentiles, in each of the signal representations for the cohort (from 0th percentile, meaning the minimum values, up to the 100th, meaning the maximum values, with the median at the 50th). Good repeatability for apparent diffusion coefficient (ADC), D, and both stretched exponential parameters are observed in the flat profiles around the median, with the expected increase in CoV at the extreme values (outliers by definition) within each VOI. Pseudodiffusion parameters from intravoxel incoherent motion (IVIM) perform substantially less well, even for median values

voxels hit the limit during fitting. Figure 6 is potentially the most interesting, and shows an example of repeated parameter values from a patient scan. Here, the results indicate nonphysiological behaviour (i.e. impossible) of IVIM parameters across the VOI (negative f values, f ranging up to 1, negative D, extremely high D*) that indicate that fitting of the IVIM model is neither especially robust, nor necessarily appropriate in this patient (or cohort).

4 | DISCUSSION

The results from this study illustrate the complexities of interpreting clinical DWI, and the danger of neglecting repeatability assessments. The high CoVs of pseudodiffusion parameters from the IVIM model indicate that only large differences or changes in these values can be treated as reliably detectable, and thus will limit their utility for classification and assessment of treatment response. The median IVIM D, together with ADC, was shown to be most repeatable, as well as the central portion of the VOI histogram. It is worth noting that the stretched exponential model parameters DDC and α are more repeatable than IVIM for this b-value range, likely arising from the fact that both parameters are fitted using the full data range. The stretched exponential representation is not a true biophysical model, because α does not have a specific interpretation describing tissue structure, but nevertheless may provide useful biomarkers that are sensitive to microstructural changes. The exponent α acknowledges and attempts to capture tissue complexity by providing an empirical assessment of the apparent range of diffusion coefficient distributions present in a voxel, and it is true that tissue microstructure is on a smaller scale than the typical DWI voxel size. Also, it should not be forgotten that ADC itself is an empirical parameter, as are IVIM parameters, and that an explicit biophysical interpretation is not a requirement for good repeatability or reliable detection of clinically relevant changes.

Alongside the variable repeatability of the diffusion parameters, these data show that there is no objectively best-performing signal representation for benign breast lesions. This is true for voxels within individual lesions, for lesions across the cohort, and even for the same lesion between repeated scans, illustrating the difficulty of attempting to compare across scans at a voxel level. This is an important result, especially because these lesions are expected to be well behaved (i.e. slow growing), and more homogeneous than a cohort of malignant lesions. Analysis using a single representation thus implicitly makes assumptions about what information the data contains, and may only show a part of the whole picture. In particular, an implicit assumption of the model being appropriate, and the data being sufficient to support the model, are able to result in nonphysiological parameters (for IVIM), as demonstrated here. Identification of these results may be obscured by summary statistics or the imposition of boundaries when fitting, and may have implications for the conclusions drawn. In particular, voxels where fitted parameters have hit
limits may be mistaken for converged values if not critically examined, and would give biased summary measures and artificially low variability. Both have the effect of obscuring the degree of information contained in the data, allowing erroneous inferences (and ultimately conclusions) from the examination. Such assumptions are often built into study design before any data are acquired, and can easily be lost if not attended to carefully. We ourselves recognise that this study implicitly assumes that results should derive from the use of a single representation across all voxels and scans, which, while standard practice and an intuitive choice, is not a given and should be a topic for deeper consideration.

The consensus in the extracranial MR imaging community that DWI is able to provide biomarkers more sophisticated than ADC is not yet matched with a similar agreement on how best to perform such analysis. Optimal \( b \)-value selection depends on both the nature of the tissue being studied and the intended analysis, and fitting of DWI signal curves can be performed using many different algorithms. These discussions essentially arise because of the inherent difficulty of obtaining high-quality data at low \( b \)-values. Pulsatile motion in larger vessels, as well as cardiac and respiratory motion, decreases data quality and is difficult to fully remove using flow-compensated and triggered techniques. Beyond simple summary statistics, histogram and spatial analyses within lesions further increase the demand for high-quality data at a voxel level, or the need for utilising prior assumptions of DWI parameter behaviour.

Some research articles report that the IVIM model provides clinically useful biomarkers in breast, but a closer reading of some studies shows them supporting the use of only, and not the pseudodiffusion parameters and \( D^* \), leaving the issue of the value of the full IVIM model unclear. Methods that split the difference between complexity and brevity, such as the three \( b \)-value minimal acquisition that does not attempt to fit \( D^* \), and relative enhanced diffusivity, may represent a pragmatic partial solution, at the expense of precluding retrospective alternate analyses.

The underlying assumptions associated with the DWI signal representation and algorithm choices are not always questioned, and may not be valid for the chosen tissue. Nonetheless, DWI studies commonly ‘lock in’ to a certain interpretation before the data are acquired, or any longitudinal changes observed, for example, the formation of necrotic regions following treatment; here, any non-Gaussian model should be robust to observed changes and still give well-behaved parameters that can, for example, capture the collapse of a specific diffusion component. The great potential of DWI in oncology is tempered by a need for critical appraisal of the model applicability in the context of the application, and it is
for this reason that repeatability and reproducibility studies are vital. It has been demonstrated that repeatability of the ADC in extracranial oncological applications can be generalised, but similar work for non-Gaussian diffusion analyses remains to be fully explored. The main limitation of this study is that it reflects only a small cohort of benign breast lesions, some of which were in themselves small, where nonmonoexponential signal behaviour has been shown to be less prominent than for malignant lesions that are of greater clinical interest. Nevertheless, the signal curves are clearly shown to be nonmonoexponential, and to dismiss the utility of a benign lesion cohort on the basis that the pseudodiffusion component is smaller illustrates precisely the problem with beginning from an unstated assumption of the correctness of one particular model (in this case IVIM). Fitting the mean VOI signal in these data shows the stretched exponential as the best fit, but this may indicate the heterogeneity among voxels, not necessarily non-Gaussian diffusion itself. This is the premise of the stretched exponential model, which assumes heterogeneity within voxels. Benign lesions are more suited for repeatability studies because they are less likely to change between repeated scans, and thus better report on the repeatability of the elements that comprise the clinical scanning process itself. In reality, it is difficult to acquire repeated scans for malignant lesions given the need to begin treatment promptly, when weighed against the (perception that the) additional MR examination provides no additional benefit to the patient. Studies that follow and build on this work, ideally in larger cohorts including malignant lesions, will add more data to these observations and examine additional repeatability limits arising from the lesion type. Additionally, this study only examined short-term repeatability as per the Quantitative Imaging Biomarkers Alliance (QIBA) definition, although the design aimed to reflect the timescale of expected changes in DWI from successful treatment. It is also important to note that data on menstrual cycle and menopause were not collected, and that it can be expected that this may vary between visits. In general, the cohort is taken to be premenopausal from population statistics. Longer term repeatability will also reflect the instability of scanner performance, as well as changes arising from hardware or software. In this study, the effect of specific acquisition parameters on repeatability was not considered, but the chosen protocol is a realistic reflection of clinical DWI protocols and so presents a fair indication of expected CoVs. Reproducibility of DWI parameters will also be important in the context of multiple sites, where a higher variation of patient positioning, scanner hardware and acquisition parameters can be expected and are challenging to standardise. Where gradient nonlinearity is known to affect the accuracy of DWI parameters away from the isocentre, consistency of patient positioning is an additional factor to consider, and these data reflect the true expected repeatability for routine clinical examination.

In conclusion, this study unambiguously demonstrates that, for a cohort of patients with benign breast lesions, there is no clearly favoured diffusion signal representation. This applies across the cohort, and within the same lesion scanned repeatedly. Appreciation of DWI measure repeatability in context will be essential in fully realising the potential, and limitations, of more complex DWI biomarkers in the clinical setting. More complex fitting algorithms that leverage additional prior information are able to generate better parameter maps, which may ameliorate poor repeatability to some degree. These techniques, though, do not absolve the experimenter of the responsibility to critically examine precisely what information the data contain, and not merely what they purport to show. In particular, the choice of representation used for analysis is not arbitrary because the resulting parameters have different repeatability characteristics, and will influence the ability to confidently infer physiological changes by DWI. This is illustrated by the different coefficients of variations derived for the investigated parameters, all drawn from the same data. The incorporation of repeated measurements for assessment of parameter repeatability in situ is critical in clinical imaging trials. The same applies to assessment of the type and details of analysis, not just the resulting parameters, and indicates the need for a deliberate effort to fully report acquisition and analysis protocols, which will also allow for more informed retrospective and alternative analyses.

ACKNOWLEDGEMENTS

We acknowledge support from the liaison Committee between the Central Norway Regional Health Authority and the Norwegian University of Science and Technology (project ID 90065000).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Neil Peter Jerome https://orcid.org/0000-0002-9556-5930

REFERENCES

1. Koh D-M, Collins DJ. Diffusion-weighted MRI in the body: applications and challenges in oncology. Am J Roentgenol. 2007;188(6):1622-1635.
2. Dietzel M, Ellmann S, Schulz-Wendtland R, et al. Breast MRI in the era of diffusion weighted imaging: do we still need signal-intensity time curves? Eur Radiol. 2019;30(1):47-56.
3. Clauser P, Mann R, Athanasiou A, et al. A survey by the European Society of Breast Imaging on the utilisation of breast MRI in clinical practice. Eur Radiol. 2018;28(5):1909-1918.
4. Vidić I, Eg nell L, Jerome NP, et al. Support vector machine for breast cancer classification using diffusion-weighted MRI histogram features: Preliminary study. J Magn Reson Imaging. 2018;47(5):1205-1216.
5. Sigmund EE, Cho GY, Kim S, et al. Intravoxel incoherent motion imaging of tumor microenvironment in locally advanced breast cancer. Magn Reson Med. 2011;65(5):1437-1447.

6. Pickles MD, Gibbs P, Lowry M, Turnbull LW. Diffusion changes precede size reduction in neoadjuvant treatment of breast cancer. Magn Reson Imaging. 2006;24(7):843-847.

7. Sharma U, Danishad KKA, Seenu V, Jagannathan NR. Longitudinal study of the assessment by MRI and diffusion-weighted imaging of tumor response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. NMR Biomed. 2009;22(1):104-113.

8. Thoeny HC, Ross BD. Predicting and monitoring cancer treatment response with diffusion-weighted MRI. J Magn Reson Imaging. 2010;32(1):2-16.

9. Iacconi C, Giannelli M, Marini C, et al. The role of mean diffusivity (MD) as a predictive index of the response to chemotherapy in locally advanced breast cancer: A preliminary study. Eur Radiol. 2010;20(2):303-308.

10. Le Bihan D, Breton E, Lallemand D, Aubin M-L, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology. 1988;168(2):497-505.

11. Bennett KM, Schmaida KM, Bennett RT, Rowe DB, Lu H, Hyde JS. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. Magn Reson Med. 2003;50(4):727-734.

12. Cho GY, Moy L, Kim SG, et al. Evaluation of breast cancer using intravoxel incoherent motion (IVIM) histogram analysis: comparison with malignant status, histological subtype, and molecular prognostic factors. Eur Radiol. 2016;26(8):2547-2558.

13. Etz RA, Onaygil C, Akin Y, Kaya H, Aribal E. Quantitative differentiation of breast lesions at 3T diffusion-weighted imaging (DWI) using the ratio of distributed diffusion coefficient (DDC). J Magn Reson Imaging. 2016;44(6):1633-1641.

14. 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.

15. Cheng L, Blackledge MD, Collins DJ, et al. T2-adjusted computed diffusion-weighted imaging: A novel method to enhance tumour visualisation. Comput Biol Med. 2016;79(August):92-98.

16. Winfield JM, Tunaru N, Rata M, et al. Extracranial soft-tissue tumors: repeatability of apparent diffusion coefficient estimates from diffusion-weighted MR imaging. Radiology. 2017;284(1):88-99.

17. Jerome NP, Miyazaki K, Collins DJ, et al. Repeatability of derived parameters from histograms following non-Gaussian diffusion modelling of diffusion-weighted imaging in a paediatric oncological cohort. Eur Radiol. 2016;26:345-353.

18. Tenel J, Fjesne HE, Østlie A, et al. Inhomogeneous static magnetic field-induced distortion correction applied to diffusion weighted MRI of the breast at 3T. Magn Reson Med. 2015;74(4):1138-1144.

19. Limpert E, Stahel WA, Abbt M. Log-normal distributions across the sciences: Keys and clues. Bioscience. 2001;51(5):341-352.

20. Orton MR, Jerome NP, Rata M, Koh D-M. IVIM in the Body: A General Overview. In: Le Bihan D, Iima M, Federau C, Sigmund EE, eds. Intravoxel Incoherent Motion (IVIM) MRI: Principles and Applications. Pan Stanford Publishing Pte. Ltd; 2018:145-174.

21. Novikov DS, Kiselev VG, Jespersen SN. On modeling. Magn Reson Med. 2018;79(6):3172-3193.

22. Kim S, Decarlo L, Cho GY, et al. Interstitial fluid pressure correlates with intravoxel incoherent motion imaging metrics in a mouse mammary carcinoma model. NMR Biomed. 2012;25:787-794.

23. Taouli B, Beer AJ, Chenevert T, et al. Diffusion-weighted imaging outside the brain: Consensus statement from an ISMRM-sponsored workshop. J Magn Reson Imaging. 2016;44(3):521-540.

24. Lemke A, Laun FB, Simon D, Stieltjes B, Schad LR. An in vivo verification of the intravoxel incoherent motion effect in diffusion-weighted imaging of the abdomen. Magn Reson Med. 2010;64(6):1580-1585.

25. Meeus EM, Novak J, Dehghani H, Peet AC. Rapid measurement of intravoxel incoherent motion (IVIM) derived perfusion fraction for clinical magnetic resonance imaging. Magn Reson Mater Phys, Biol Med. 2018;31(2):269-283.

26. Peters NHGM, Vincken KL, Van Den Bosch MAAJ, Luijten P, Mali WPTM, Bartels LW. Quantitative diffusion weighted imaging for differentiation of benign and malignant breast lesions: The influence of the choice of b-values. J Magn Reson Imaging. 2010;31(5):1100-1105.

27. Jansenfied O, Montelius M, Starck G, Ujbangberg M. Optimization of b-value schemes for estimation of the diffusion coefficient and the perfusion fraction with segmented intravoxel incoherent motion model fitting. Magn Reson Med. 2019;82(4):1541-1552.

28. Orton MR, Collins DJ, Koh DM, Leach MO. Improved intravoxel incoherent motion analysis of diffusion weighted imaging by data driven Bayesian model selection. Magn Reson Med. 2014;71(1):411-420.

29. Freiman M, Perez-Rossello JM, Callahan MJ, et al. Reliable estimation of incoherent motion parametric maps from diffusion-weighted MRI using fusion bootstrap moves. Med Image Anal. 2016;44(3):325-336.

30. Wurzig MC, Kenkel D, Filli L, Boss A. A standardized parameter of diffusion for combined intravoxel incoherent motion and diffusion Kurtosis analysis of diffusion imaging data. Invest Radiol. 2016;51(3):203-210.

31. Gurney-Champion OJ, Klaassen R, Froeling M, et al. Comparison of six fit algorithms for the intravoxel incoherent motion model of diffusion-weighted magnetic resonance imaging data of pancreatic cancer patients. PLoS One. 2018;13(4):1-18.

32. Meeus EM, Novak J, Withey SB, Zarinabad N, Dehghani H, Peet AC. Evaluation of intravoxel incoherent motion fitting methods in low-perfused tissue. J Magn Reson Imaging. 2017;45(5):1325-1334.

33. Wetscherek A, Stieltjes B, Laun FB. Flow-compensated intravoxel incoherent motion diffusion imaging. Magn Reson Med. 2015;74(2):410-419.

34. Jerome NP, Orton MR, D'Arcy JA, Collins DJ, Koh D-MM, Leach MO. Comparison of free-breathing with navigator-controlled acquisition regimes in abdominal diffusion-weighted magnetic resonance images: Effect on ADC and IVIM statistics. J Magn Reson Imaging. 2014;39(1):235-240.

35. While PT. A comparative simulation study of bayesian fitting approaches to intravoxel incoherent motion modeling in diffusion-weighted MRI. Magn Reson Med. 2017;76(6):2373-2387.

36. Chen F, Chen P, Muhammed HH, Zhang J. Intravoxel incoherent motion diffusion for identification of breast malignant and benign tumors using chemometrics. Biomed Res Int. 2017;2017:3845409.

37. Zhu Y, Li X, Wang F, et al. Intravoxel incoherent motion diffusion-weighted magnetic resonance imaging in characterization of axillary lymph nodes: Preliminary animal experience. Magn Reson Imaging. 2018;52(May):46-52.

38. Wang Q, Guo Y, Zhang J, Wang Z, Huang M, Zhang Y. Contribution of IVIM to conventional dynamic contrast-enhanced and diffusion-weighted MRI in differentiating benign from malignant breast masses. Breast Care. 2016;11(4):254-258.
39. Iima M, Kataoka M, Kanao S, et al. Intravoxel incoherent motion and quantitative non-Gaussian diffusion MR imaging: evaluation of the diagnostic and prognostic value of several markers of malignant and benign breast lesions. Radiology. 2018;287(2):432-441.
40. Kim Y, Kim SH, Lee HW, et al. Intravoxel incoherent motion diffusion-weighted MRI for predicting response to neoadjuvant chemotherapy in breast cancer. Magn Reson Imaging. 2018;48:27-33.
41. Bokacheva L, Kaplan JB, Giri DD, et al. Intravoxel incoherent motion diffusion-weighted MRI at 3.0 T differentiates malignant breast lesions from benign lesions and breast parenchyma. J Magn Reson Imaging. 2014;40(4):813-823.
42. Pieper CC, Willinek WA, Meyer C, et al. Intravoxel incoherent motion diffusion-weighted MR imaging for prediction of early arterial blood flow stasis in radioembolization of breast cancer liver metastases. J Vasc Interv Radiol. 2016;27(9):1320-1328.
43. Teruel JR, Goa PE, Sjøbakk TE, Østlie A, Fjøsne HE, Bathen TF. A simplified approach to measure the effect of the microvasculature in diffusion-weighted MR imaging applied to breast tumors: preliminary results. Radiology. 2016;281(2):373-381.
44. Vidić I, Jerome NP, Bathen TF, Goa PE, While PT. Accuracy of breast cancer lesion classification using intravoxel incoherent motion diffusion-weighted imaging is improved by the inclusion of global or local prior knowledge with Bayesian methods. J Magn Reson Imaging. 2019;50(5):1478-1488.
45. Iima M, Yano K, Kataoka M, et al. Quantitative non-Gaussian diffusion and intravoxel incoherent motion magnetic resonance imaging. Invest Radiol. 2015;50(4):205-211.
46. 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(1):9-26.
47. Gottschalk MS, Eskild A, Hofvind S, Gran JM, Bjelland EK. Temporal trends in age at menarche and age at menopause: A population study of 312 656 women in Norway. Hum Reprod. 2020;35(2):464-471.
48. Newitt DC, Tan ET, Wilmes LJ, et al. Gradient nonlinearity correction to improve apparent diffusion coefficient accuracy and standardization in the American college of radiology imaging network 6698 breast cancer trial. J Magn Reson Imaging. 2015;42(4):908-919.
49. Tan ET, Manielli L, Slavens ZW, King KF, Hardy CJ. Improved correction for gradient nonlinearity effects in diffusion-weighted imaging. J Magn Reson Imaging. 2013;38(2):448-453.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Jerome NP, Vidić I, Egnell L, et al. Understanding diffusion-weighted MRI analysis: Repeatability and performance of diffusion models in a benign breast lesion cohort. NMR in Biomedicine. 2021;34:e4508. [https://doi.org/10.1002/nbm.4508](https://doi.org/10.1002/nbm.4508)