Published in final edited form as:

NMR Biomed. 2018 January ; 31(1): . doi:10.1002/nbm.3843.

Sources of Systematic Error in Proton Density Fat Fraction (PDFF) Quantification in the Liver Evaluated From Magnitude Images With Different Numbers of Echoes

Mark Bydder, PhD¹, Gavin Hamilton, PhD², Ludovic de Rochefort, PhD¹, Ajinkya Desai, MD², Elhamy R Heba, MD², Rohit Loomba, MD, MHSc^{3,4}, Jeffrey B Schwimmer, MD^{5,6}, Nikolaus M. Szeverenyi, PhD², and Claude B Sirlin, MD²

¹Aix-Marseille Université, Centre de Résonance Magnétique Biologique et Médicale, Marseille, France

²Liver Imaging Group, Department of Radiology, University of California, San Diego, CA

³Division of Gastroenterology, Department of Medicine, University of California, San Diego, CA

⁴Division of Epidemiology, Department of Family Medicine and Preventive Medicine, University of California, San Diego, CA

⁵Department of Pediatrics, University of California, San Diego, CA

⁶Department of Gastroenterology, Rady Children's Hospital, San Diego, CA

Abstract

The purpose of this work was to investigate sources of bias in MRI liver fat quantification that leads to a dependence of the proton density fat fraction (PDFF) on number of echoes. This was a retrospective analysis of liver MRI data from 463 subjects. The magnitude signal variation with echo time from spoiled gradient echo images was curve-fitted to estimate the PDFF using a model that included monoexponential $R2^*$ decay and a multi-peak fat spectrum. Additional corrections for non-exponential decay (Gaussian), bi-exponential decay, degree of fat saturation, water frequency shift and noise bias were introduced. The fitting error was minimized with respect to $463 \times 3 = 1389$ subject-specific parameters and 7 additional parameters associated with these corrections. The effect on PDFF was analyzed, notably the dependence on number of echoes. Effects on $R2^*$ were also analyzed. Results show that the inclusion of bias corrections resulted in an increase in the quality of fit (r-squared) in 427 out of 463 subjects (i.e. 92.2 %) and a reduction in the total fitting error (residual norm) of 43.6 %. This was largely due to the Gaussian decay (57.8 % of the reduction), fat spectrum (31.0 %) and biexponential decay (8.8 %) terms. The inclusion of corrections was also accompanied by a decrease in the dependence of PDFF on number of echoes. Similar analysis of the $R2^*$ showed a decrease in the dependence on number of echoes. Comparison of PDFF with spectroscopy indicated excellent agreement before and after correction, but the latter exhibited lower bias on a Bland Altman plot (1.35 % versus 0.41 %). In conclusion, correcting for known and expected biases in PDFF quantification in liver reduces the fitting error, reduces the dependence on number of echoes and increases accuracy.

Keywords

PDFF; NAFLD; Fat quantification; Bias; Liver

Introduction

Quantitative measures of the fat content in liver are increasingly used as biological markers of non-alcoholic fatty liver disease (NAFLD) and other metabolic disorders (1,2,3,4). Qualitative evaluation of the fat content using magnetic resonance imaging (MRI) has been routinely used for a long time by assessing the difference between the in-phase (IP) and opposed-phase (OP) images (5,6).

However qualitative and semi-quantitative methods are difficult to standardize and can be implementation dependent, which can complicate the use of MRI for monitoring patients over time, especially if they are scanned on different scanners or with different techniques. Thus there are advantages of developing objective, quantitative techniques and adopting stringent tests for robustness and accuracy. Quantitative proton density fat fraction (PDFF) techniques model the proton densities of water and fat but must also include additional physical/chemical processes such as $R2^*$ decay and the degree of unsaturation of triglyceride molecule quantified in terms of the number of double bonds (ndb).

There are many considerations potentially relevant to PDFF quantification. Firstly, there are parameters that are expected to vary pixel by pixel: a scaling term due to receiver sensitivity, PDFF and $R2^*$ (7,8). Secondly there are parameters that are global over all pixels but, to a first approximation, can be expected to be constant within one data set from one acquisition: the chemical shift frequency difference between H_2O and CH_2 , fat spectrum, B_0 homogeneity and noise bias (9,10). Thirdly there are functional forms due to the physics of MRI (e.g. Bloch equations) that determine whether the decay is exponential, biexponential or non-exponential. In qualitative applications, all but the scaling term and PDFF are neglected, leading to a biased estimate of PDFF with this oversimplified model.

Certain features of the signal are readily apparent in the data (e.g. $R2^*$ decay) while subtle manifestations are more difficult to detect, however they may still introduce systematic errors in quantification. A pragmatic way to identify systematic errors is to look for spurious dependences of the PDFF on parameters that should not affect PDFF; for example, by measuring PDFF before/after administration of a contrast agent (11,12) or measuring correlations between PDFF and imaging parameters (13). One recent study in 84 subjects with 16 echoes observed that the PDFF in liver decreased significantly and systematically with number of echoes (from 3 to 16) in a zig-zag pattern despite correction for the major confounds ($R1$, $R2^*$ and fat spectrum) (14). The authors noted there are several areas for refinement in their model to account for these effects, such as spectral peak correction, $T2^*$ estimation, Gaussian or sinc decay and noise bias.

The purpose of the present study is to characterize and correct for sources of bias in PDFF estimation. Starting with the model of Ref (14), the effect of varying the parameters and functional forms described above was explored. The retrospective study design restricted the

analysis to magnitude images whereas there also exist biases in the phase part of the images that can contribute to inaccuracies (15,16); these are beyond the scope of the present study.

Methods

Study design and patients

This was a HIPAA-compliant and IRB-approved retrospective analysis of a cohort of research subjects with known or suspected NAFLD scanned between 2009 and 2012. A total of 463 subjects were included: 379 scanned with 6 echoes (199 male, 180 female, age range 8–81 years, mean $\pm$ standard deviation 28.5 ± 18.5) and 84 scanned with 16 echoes (45 male, 39 female, age range 10–74 years, mean $\pm$ standard deviation 35.9 ± 22.2). There was no overlap between the cohorts. Inclusion criteria were liver MRI using low-flip-angle, multiple-echo (6 or 16 echoes) spoiled-gradient-recalled-echo (SPGR) sequence and single voxel liver spectroscopy. Exclusion criteria were data acquired post-gadolinium or with parallel imaging or surface coil correction, as these invalidate the method of measuring background noise. One subject who had very rapid $R2^*$ decay (2 times larger than the next largest $R2^*$) was also excluded. Few subjects with NAFLD have iron overload and if they do the iron overload tends to be mild; without biopsy it is not possible to say definitively. Adults age 18 and older provided written informed consent. For children ages 8 to 17 written consent was obtained from a parent and written assent from the child.

Acquisition

Imaging was performed at 3.0T on a GE Healthcare (Milwaukee, WI) HDx scanner using zoom gradients, 8-element torso or cardiac phased array coil, multi-slice 2D SPGR, flip angle (FA) 10 degrees, repetition time (TR) between 100–240 ms to accommodate the subject's breath-hold capacity, field of view (FOV) 40–44 cm with rectangular adjustment to body habitus, base matrix 192×192 , bandwidth ± 142 kHz and slice thickness 8 mm. A fractional echo factor of 0.8 was necessary to attain minimum TEs of 1.1–1.3 ms and TE spacing of 1.1–1.2 ms in a single bipolar echo-train. Images were created by the vendor reconstruction software (homodyne correction and sum of squares) and magnitude images were exported to an Osirix DICOM viewer for analysis (17).

Magnetic resonance spectroscopy (MRS) was performed using the STEAM sequence at a single $2 \times 2 \times 2$ cm³ voxel in the right lobe of the liver: spectra were acquired in a single breath-hold using TE 10, 15, 20, 25 and 30 ms, TR 3500 and 10 ms mixing time. Analysis was performed in the MRUI software package as described in Ref (9). The spectroscopy-determined PDFF served as the reference standard for evaluating accuracy.

Data Modeling

Circular regions of interest (ROI) of 20 mm diameter were drawn in the right lobe of the liver (co-localized to the spectroscopy voxel) and in the background noise (the top left corner of the image). Care was taken to avoid signal from plastic in the coils. The means over the ROIs had a very high signal to noise ratio (SNR) so bias was the dominant source of error in these signals. The signal versus echo time was modeled to obtain estimates of 3 parameters for each subject: PDFF, $R2^*$ and a scale term

$$\text{signal} = \text{scale} \cdot |(1 - \text{PDFF}) + \text{PDFF} \cdot \text{fat basis}| \cdot \exp(-R2 * \text{TE}) \quad [1]$$

where the fat basis is a weighted sum of complex exponentials representing the fat spectrum. Around this, 7 bias parameters were optimized over all subjects, corresponding to sources of bias described below.

The overall procedure minimizes the sum of squares residual over 3618 data points ($379 \times 6 + 84 \times 16$ ROIs) with respect to 1396 free parameters (463×3 subject-specific + 7 bias-related). Computation took 2 hours on an Intel Xeon workstation, MATLAB 2016a (Natick MA), Levenberg Marquardt algorithm (*lsqnonlin* with tolerance 10^{-10}).

Noise bias—Magnitude images have Rician-like noise distribution that can be approximated as Gaussian with a bias term added in quadrature, i.e. $\text{signal} \rightarrow (\text{signal}^2 + \text{noise bias}^2)^{1/2}$ (10). The noise bias is equal to the noise standard deviation scaled by factors associated with the receiver circuitry and reconstruction method (18). A common way to estimate noise bias is to measure the signal in an ROI where there is only air (e.g. in the corner of the image). When performing such measurements, it was found that the noise bias was sometimes greater than the signal in the liver, which should not be possible. Phantom experiments revealed a substantial variation in the background noise (Figure 1), the cause of which was determined to be the gradient distortion correction. Since “zoom” gradients were used for all *in vivo* scans, the background noise measurements needed to be corrected by a factor of ~2 to reflect the values at the center.

The ROI data from each subject were divided by the background noise for that subject in order to standardize the noise variance and remove scaling differences between subjects. This meant the noise bias term could be modeled as a single parameter for all subjects and its value should be close to 1.

Gaussian decay—Magnetic field variations on the scale of the voxel (notably in the slice direction) result in a distribution of frequencies that evolve over time as the Fourier transform of the slice profile. The slice profile is often idealized as rect or Gaussian which leads to the decay resembling a sinc or Gaussian, but in practice the shape is not easily characterized (19). To a low-order Taylor approximation these functions have the same form, a constant plus a quadratic term, and so, for simplicity, the present study adopted the Gaussian. This means the signal decay term was modified to be of the form

$$\exp(-a \cdot \text{TE} - b \cdot \text{TE}^2) \quad [2]$$

where a represents the exponential decay and b represents the Gaussian decay.

Biexponential decay—There are plausible reasons to consider that the exponential decay rate is different for water and fat, since their compartmentalization and local environment are different. Furthermore fat and water have different magnetic susceptibilities, which can lead

to signal dephasing and an associated reduction in $T2^*$ (20). Despite these considerations, monoexponential fitting has generally been found adequate for estimating PDFF accurately (e.g. see Ref 8 among others). In order to accommodate a contribution from a biexponential term while retaining the numerical stability of monoexponential fitting, the $R2^*$ difference ($\Delta R2^*$) was modeled over all subjects as:

$$\Delta R2^* \equiv R2^* (\text{water}) - R2^* (\text{fat}) = c_0 + c_1 \cdot \text{PDFF} \quad [3]$$

This $R2^*$ is composed of a constant plus a linear term in PDFF, which, respectively, model systematic differences in $R2^*$ (e.g. due to $T2$ or other fundamental differences between water and fat) and the effects of susceptibility-induced dephasing caused by the difference in magnetic susceptibility (or any other differences that depend on PDFF).

Fat spectrum—The fat spectrum of triglyceride can be characterized in a number of ways, e.g. by the number of saturated, singly unsaturated and doubly unsaturated fatty acid chains (21) or by considerations about the molecular structure (9,22). This allows 3 parameters to completely characterize the fat spectrum rather than the 9 or more individual peaks; the use of 2 empirical constraints described in Ref (23) reduces this to a single parameter, ndb , which is the approach adopted in the present study.

In addition to optimizing the average ndb , there may be compositional changes in the fat compartment that cause ndb to systematically vary with PDFF. Changes in free fatty acid concentration and triglyceride composition with fat content have been reported in chemical analysis of biopsy samples (24,25) and there have been suggestive findings in MRS of correlation (both positive and negative) between ndb and the volume of fat in various depots, although not specifically in liver (26,27). In order to accommodate a systematic variation of this type, the ndb was modeled as a constant plus linear term in PDFF:

$$ndb = ndb_0 + ndb_1 \cdot \text{PDFF} \quad [4]$$

Water frequency correction—The chemical shift of water protons varies with temperature and pH, e.g. 4.7 ppm (at 39 °C) to 4.9 ppm (at 20 °C). Fat protons vary to a lesser extent that is essentially negligible. Since PDFF is a sensitive function of the water frequency, it is useful to consider the water frequency as a free parameter and estimate its value directly. In terms of Eq 1 this amounts to a frequency shift of the fat basis.

Data Analysis

Analysis 1—Using all subjects and all echoes, nonlinear optimization was performed over the 7 bias parameters and 1389 subject-specific parameters (PDFF, $R2^*$ and scale) to minimize the sum of squares of the residual norm ($\|r\|$) between 3618 data points and the model. Starting from the original model of Ref 14, the optimized parameter values were obtained.

Analysis 2—Following Ref (14), the PDFF was calculated using different numbers of echoes from 3 to 16. The variation in PDFF with number of echoes was plotted using the same model as Ref (14) and using the best-fit parameters from Analysis 1. Corrections were applied in order of their contribution to the change in residual. The effect of including each parameter was evaluated in terms of the correlation between PDFF and number of echoes. Since PDFF and number of echoes are independent parameters, any correlation can be considered evidence of a systematic modeling error.

Analysis 3—The accuracy of the estimated PDFF was compared using spectroscopy as a reference. The agreement between the original and optimized models versus spectroscopy was evaluated. Note that the spectroscopy PDFF was corrected to match the *ndb* used in the imaging PDFF estimate, since a proportion of the fat signal is hidden under the water peak and must be inferred from *ndb*. For example, the fat spectrum assumed in Ref (14) assumes an *ndb* of ≈ 1.92 , which implies 8.7 % of the fat signal is obscured by water whereas an *ndb* of 2.436 (26) would mean a fraction of 9.7 % is obscured. Corrections were also applied to remove systematic T1 bias that arises from the use of finite TRs and non-zero FAs. These were small: assuming T1 = 790 ms (liver) and T1 = 260 ms (fat), the maximum errors in PDFF were 0.3 % for spectroscopy and 1.6 % for imaging.

Results

Analysis 1—Results for the optimization are given in Table 1. As seen from the shaded row, the global optimum corresponds to an unphysical value for *ndb*, which should be in the range 2 – 3 for human and animal fats. Stepping through discrete values for *ndb*₁ shows this physical constraint necessitates *ndb*₁ to be in the range ± 1 in order that *ndb* stays in a valid range; it is not possible to determine a precise value for *ndb*₁ from the data in this study. A propensity for a negative value is supported by Refs (24,25) which indicate liver fat is more saturated at higher PDFF levels and, on this basis, the present study hypothesizes a value at the low end of the valid range of *ndb*₁ = -1.

A physically meaningful range for *c*₁ can be obtained from the magnetic susceptibility difference between water and fat. Under the static dephasing regime, a spherical dipole (fat droplet) embedded in a substrate (water) would give rise to a change in R2* that depends on the volume fraction η of the fat droplet (28):

$$\Delta R2^* \sim \frac{2\pi}{9\sqrt{3}} \cdot \gamma B_0 \cdot \Delta\chi \cdot \eta \sim c_1 \cdot \text{PDFF}$$

The substitution of η by PDFF is a reasonable approximation since the densities and proton to mass ratios for fat and water are similar (2). Taking the susceptibility difference $\chi = 0.65\text{--}0.75$ ppm from *in vivo* measurements in muscle and marrow (29) yields an estimate of $210\text{--}243 \text{ s}^{-1}$ for *c*₁. The fitted values for *c*₁ in Table 1 are consistent with this estimate for valid ranges of *ndb*.

The contribution to the change in residual norm ($\|\mathbf{r}\|$) between the original and the optimal parameter choices is shown in the final column. The most important effects (in terms of the

reduction in $\|r\|$) are Gaussian decay (57.8 %), followed by ndb (31.0 %) and $R2^*$ (8.8 %). The remaining effects account for less than 2.5 %.

Example fits from 4 subjects are shown in Figure 2. The contribution of the Gaussian decay can be seen in the low PDFF plots, which exhibit some of the characteristic bell shape. With increasing PDFF, another noticeable feature is the improved fitting of the first echo, which is frequently overestimated by the original model. The quality of fit (R^2) was increased in 427 out of 463 data sets (92.2 %) with a mean value of 0.9899 (range 0.9414–0.9998) before and 0.9963 (range 0.9525–1.000) after bias corrections.

Analysis 2—Figure 3 shows the variation of the PDFF with number of echoes when successively adding the different bias corrections. The original model is that described in Ref (14) and different lines represent the cumulative inclusion of the bias parameters in order of their contribution to $\|r\|$.

Despite its large effect on $\|r\|$ (see Analysis 1), the Gaussian did not alter the downward trend with number of echoes, however it did attenuate the zig-zag pattern. Introducing ndb elevated the mean PDFF and also reduced the downward trend. The $R2^*$ had a similar effect, although smaller. The water frequency term only had a minor effect (presumably because its value changed very little between models) and the noise bias term elevated PDFF slightly. Numerical tests for the strength of the association between PDFF and number of echoes are given in Table 2. These are in accord with the qualitative observations in Figure 3.

Figure 3 also shows the variation of the $R2^*$ with number of echoes when adding different bias corrections. The original model showed an upward trend in $R2^*$ with number of echoes. This trend was greatly reduced by introducing the Gaussian correction. The other terms have little effect on $R2^*$ apart from $R2^*$, which reduced the $R2^*$ by a constant amount approximately equal to $R2^*$ calculated at the mean PDFF of the cohort of ~ 0.12 .

Analysis 3—Figure 4 shows Bland Altman plots to illustrate the accuracy of the PDFF with respect to spectroscopy. The PDFF in these plots was calculated using the maximum number of echoes (either 6 or 16) available for each subject. Both models are in excellent agreement with spectroscopy. The mean bias and 95 % confidence intervals are 1.35 ± 2.63 for (original) and 0.41 ± 2.67 (optimized), indicating reduced bias with the optimized parameters.

Discussion

The purpose of this study was to identify biases that can affect PDFF accuracy. The main sources of bias ($T1$ and $T2^*$ relaxation) are well-known and easily recognized (6,7,8) however smaller effects are harder to detect. They can manifest as consistent mis-fitting of certain data points over many datasets or spurious dependences of the fitted parameters (PDFF and $R2^*$) on independent imaging parameters; e.g. Ref (14) showed an example in which PDFF exhibited a significant dependence on the number of echoes.

The following biases were considered: Gaussian decay, biexponential decay, number of double bonds (ndb), noise bias and water frequency shift. Results showed a 43.6 % reduction in fitting error (residual norm) when including all of the corrections. Gaussian decay and ndb accounted for almost 90 % of the reduction but a dependence of PDFF on number of echoes was still present until corrections for an $R2^*$ difference between water and fat were included. Corrections for water frequency shift and noise bias had a very small effect, the former because it was already close to its optimal value. In regard to the practical computation of PDFF with bias corrections, it should be noted that the complexity of including these terms in the model is hidden from the user and that the demands of the fitting process are no different than existing approaches for estimating PDFF; the only change is the use of pre-calibrated constants to correct for the biases.

The Gaussian decay arises from inhomogeneity of the main field (B_0) across the voxel and is thus a function of voxel size (approximately $8 \times 2 \times 2 \text{ mm}^3$ in the present study). It was modeled as an empirical factor although a non-exponential contribution to the decay can be inferred from the spatial derivative of B_0 (19). Incorporating the spatial derivative of B_0 thus provides an elegant way to model non-exponential decay and also serves as a natural smoothness constraint on B_0 , since discrete changes (related to 2π ambiguities) must be accompanied by a rapid decay of the signal. Complex images (i.e. magnitude and phase) are necessary to model B_0 , which also obviates the need for noise bias correction, albeit at a cost of requiring other bias corrections specific to the phase (15,16).

In regard to number of double bonds, the values of ndb_0 and ndb_1 from the present study yield an average ndb over the range PDFF 0 – 0.5 that is close to a recent literature value in liver of 2.436 (0.812 per fatty acid chain) (26). Finding agreement on the average fat properties is satisfying although an open question remains about the variation in ndb with PDFF. It was not possible to measure ndb_1 with precision in the present study but physical considerations limit the feasible values to ± 1 . Differentiating the effects of ndb and $R2^*$ changes using low spectral resolution data is challenging, since small changes in the fat spectrum typically mimic $R2^*$ decay (30). Ultimately the ndb parameter could be considered a biomarker in its own right and be measured on an individual basis in each subject, which would avoid the need to make assumptions about ndb.

An $R2^*$ difference between water and fat was modeled as a constant plus a linear term in PDFF. These were found to have opposite signs and therefore tended to cancel out at low to moderate PDFFs, which is supportive of the use of monoexponential $R2^*$ in PDFF estimation. The negative constant term ($c_0 = -43.2 \text{ s}^{-1}$) indicates that $R2^*(\text{fat}) > R2^*(\text{water})$ in liver, i.e. fat peaks have a broader linewidth than water by approximately 15 Hz. Previous work has reported the opposite relationship for the $R2$ values in liver, i.e. $R2(\text{fat}) = 12\text{--}20 \text{ s}^{-1} < R2(\text{water}) = 43 \text{ s}^{-1}$ (9), which would predict a narrower linewidth for fat. Broader fat peaks may reflect the presence of non-triglyceride organic compounds in the fat compartment.

The linear term c_1 is compatible with the effects of susceptibility-induced dephasing due to the difference in magnetic susceptibility between water and fat. Previous studies have also reported a positive correlation between PDFF and $R2^*$ with slopes of between 25 and ≈ 100

s^{-1} (31,32,33). The $R2^*$ in these studies is a monoexponential approximation of the decay and therefore not directly comparable to c_1 . In order to make a direct comparison, Figure 5 plots $R2^*$ versus PDFF using the same methodology as Refs 31–33 (i.e. the original model), which gives a slope of $79.3 s^{-1}$ in good agreement with the previous studies. Using the bias-corrected model, however, gives a slope of $-131.9 s^{-1}$. This would imply that, above a certain PDFF, the $R2^*$ is negative; clearly this is not physically meaningful but indicates that the decay rates built into the model (i.e. c_0 and c_1) overestimate the rate of decay at high PDFF. Reducing c_0 and c_1 by a factor of 3 removes the correlation between PDFF and $R2^*$ (slope of 0) with the implication that χ is 3 times smaller than Ref (29). As indicated by Table 1, this would put ndb in an unphysical range, which in turn could be resolved by holding ndb_1 at -1 ; the residual norm for such a constrained fit is $\|r\| = 5.5206$. It is appealing in the sense of Ref (19) to correct $R2^*$ for the confounding effects of susceptibility and ndb although it is not clear whether correlation between PDFF and $R2^*$ should be considered artefactual. Independent studies of ndb versus PDFF, $R2^*$ versus PDFF and χ for fat and liver would be helpful to put physical constraints on the parameters.

There are several interesting aspects relevant to $R2^*$ relaxometry in the liver that go beyond the scope of the present study. The assumption of static dephasing to predict c_1 may be questionable given the size of hepatocytes relative to the mean free path of water. Also the model of a spherical dipole embedded in a substrate is unrealistic at high PDFF when fat occupies most of the hepatocyte. Another relevant issue is the presence of a small amount of iron, which would increase $R2^*(\text{water})$ but also decrease the magnetic susceptibility difference χ ; this means that, for a certain range of iron and PDFF, iron may have no effect on (or may even decrease) the rate of decay. At higher iron concentrations, iron and susceptibility-dephasing work in the same direction to reduce $R2^*$ and ultrashort TE imaging methods may be necessary. Extremely short T2 components in the liver (ostensibly macromolecules from fibrosis) were not considered in the present study but may comprise 11.6 % of the MRI signal at TE = 0 (34).

Parameter values in the present study were optimized over a large dataset with different echo times but may not be optimal for other acquisitions and field strengths. Further study would help to establish the ranges and relative importance of each parameter. Some useful extensions would be high flip angle and/or short TR scans with averaging, which would increase the SNR and allow smaller voxels to be used. This would help model the lineshapes (i.e. ndb and $R2^*$) without such a strong Gaussian component; the introduced T1 bias is relatively well understood. Ultra high field imaging would improve SNR and also emphasize components of $R2^*$ that scale with field strength, such as susceptibility dephasing.

In conclusion, the present study has examined the influence of several minor confounding factors in quantification of PDFF in the liver using magnitude data. The inclusion of tuned constants in the signal model can reduce bias in the PDFF estimate leading to measurements that are less dependent on the methods used to acquire the images. The effect of PDFF as a confounder for $R2^*$ was also considered. In the spirit of reproducible research, the data and code used in the present study are made available online at <https://github.com/marcsous/bias>.

Acknowledgments

This work was also supported in part by GE Healthcare (Milwaukee, WI) and from grants from the National Institutes of Health including R01 DK075128, R01 DK088925 and K23 DK090303 (RL).

The research leading to these results received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n. PCOFUND-GA-2013-609102, through the PRESTIGE programme coordinated by Campus France.

References

1. Waterton JC. What imaging biomarkers are and how they are used. Proc ISMRM 2009: Educational Lecture.
2. Reeder SB, Hu HH, Sirlin CB. Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration. J Magn Reson Imaging. 2012; 36:1011. [PubMed: 22777847]
3. Le T-A, Chen J, Changchien C, Peterson MR, Kono Y, Patton H, Cohen BL, Brenner D, Sirlin C, Loomba R. the San Diego Integrated NAFLD Research Consortium (SINC). Effect of colessevelam on liver fat quantified by magnetic resonance in nonalcoholic steatohepatitis: a randomized controlled trial. Hepatology. 2012; 56:922. [PubMed: 22431131]
4. Nouredin M, Lam J, Peterson MR, Middleton M, Hamilton G, Le TA, Bettencourt R, Changchien C, Brenner DA, Sirlin C, Loomba R. Longitudinal comparison between MRI, MRS and histology-determined steatosis in NAFLD patients at two-time points in a randomized trial. Hepatology. 2013; 58(6):1930. [PubMed: 23696515]
5. Dixon WT. Simple Proton Spectroscopic Imaging. Radiology. 1984; 153:189. [PubMed: 6089263]
6. Ma J. Dixon Techniques for Water and Fat Imaging. J Magn Reson Imag. 2008; 28:543.
7. Westphalen ACA, Qayyum A, Yeh BM, Merriman RB, Lee JA, Lamba A, Lu Y, Coakley FV. Liver Fat: Effect of Hepatic Iron Deposition on Evaluation with Opposed-Phase MR Imaging. Radiology. 2007; 242:450. [PubMed: 17255416]
8. Bydder M, Yokoo T, Hamilton G, Middleton MS, Chavez AD, Schwimmer JB, Lavine JE, Sirlin CB. Relaxation effects in the quantification of fat using gradient echo imaging. Magn Reson Imaging. 2008; 26:347–59. [PubMed: 18093781]
9. Hamilton G, Yokoo T, Bydder M, Cruite I, Schroeder ME, Sirlin CB, Middleton MS. In vivo characterization of the liver fat 1H MR spectrum. NMR Biomed. 2011; 24:784. [PubMed: 21834002]
10. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. Magn Reson Med. 1995; 34:910. [PubMed: 8598820]
11. Takeshi, Yokoo, Collins, Julie M., Hanna, Robert F., Bydder, Mark, Middleton, Michael S., Sirlin, Claude B. Effects of intravenous gadolinium administration and flip angle on the assessment of liver fat signal fraction with opposed-phase and in-phase imaging. J Magn Reson Med. 2008; 28:246.
12. Liao J, Shiehorteza M, Girard OM, Sirlin CB, Bydder M. Evaluation of MRI fat fraction in the liver and spine pre and post SPIO infusion. Magnetic Resonance Imaging. 2013; 31:1012. [PubMed: 23602721]
13. Hansen KH, Schroeder ME, Hamilton G, Sirlin CB, Bydder M. Robustness of fat quantification using chemical shift imaging. Magn Reson Imaging. 2012; 30:151. [PubMed: 22055856]
14. Levin YS, Yokoo T, Wolfson T, Gamst AC, Collins J, Achmad EA, Hamilton G, Middleton MS, Loomba R, Sirlin CB. The effect of echo-sampling strategy on the accuracy of out-of-phase and in-phase multi-echo gradient-echo magnetic resonance imaging hepatic fat fraction estimation. 2014; 39:567.
15. Hernando D, Hines CDG, Yu H, Reeder SB. Addressing phase errors in fat-water imaging using a mixed magnitude/complex fitting method. Magn Reson Med. 2012; 67:638–644. [PubMed: 21713978]
16. Ruschke S, Eggers H, Kooijman H, Diefenbach MN, Baum T, Haase A, Rummeny EJ, Hu HH, Karampinos DC. Correction of Phase Errors in Quantitative Water–Fat Imaging Using a

- Monopolar Time-Interleaved Multi-Echo Gradient Echo Sequence. *Magn Reson Med.* (Early View 2016).
17. Rosset A, Spadola L, Ratib O. OsiriX: An Open-Source Software for Navigating in Multidimensional DICOM Images. *Journal of Digital Imaging.* 2004; 17:205. [PubMed: 15534753]
 18. Dietrich O, Raya JG, Reeder SB, Ingrisch M, Reiser MF, Schoenberg MO. Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics. *Magn Reson Imaging.* 2008; 26:754. [PubMed: 18440746]
 19. Hernando D, Vigen K, Shimakawa A, Reeder S. R2* mapping in the presence of macroscopic B0 field variations. *Magn Reson Med.* 2012; 68:830. [PubMed: 22161866]
 20. Hopkins JA, Wehrli FW. Magnetic Susceptibility Measurement of Insoluble Solids by NMR: Magnetic Susceptibility of Bone. *Magn Reson Med.* 1997; 37:494. [PubMed: 9094070]
 21. Ren J, Dimitrov I, Sherry AD, Malloy CR. Composition of adipose tissue and marrow fat in humans by 1H NMR at 7 Tesla. *Journal of Lipid Research.* 2008; 49:2055. [PubMed: 18509197]
 22. Berglund J, Ahlström H, Kullberg J. Model-Based Mapping of Fat Unsaturation and Chain Length by Chemical Shift Imaging—Phantom Validation and In Vivo Feasibility. *Magn Reson Med.* 2012; 68:1815. [PubMed: 22334300]
 23. Bydder M, Girard O, Hamilton H. Mapping the double bonds in triglycerides. *Magn Reson Imaging.* 2011; 29:1041. [PubMed: 21868182]
 24. Singer P, Gnauck G, Honigsmann G, Stolz P, Schliack V, Kettler LH, Buntrock P, Thoeke H. The Fatty Acid Pattern of Adipose Tissue and Liver Triglycerides According to Fat Droplet Size in Liver Parenchymal Cells of Diabetic Subjects. *Diabetologia.* 1974; 10:455. [PubMed: 4375642]
 25. Mavrelis PG, Ammon HV, Gleysteen JJ, Komorowski RA, Charaf UK. Hepatic free fatty acids in alcoholic liver disease and morbid obesity. *Hepatology.* 1983; 3:226. [PubMed: 6832713]
 26. Lundbom J, Hakkarainen A, Soderlund S, Westerbacka J, Lundbom N, Taskinen MR. Long-TE 1H MRS suggests that liver fat is more saturated than subcutaneous and visceral fat. *NMR Biomed.* 2011; 24:238. [PubMed: 20821410]
 27. Lundbom J, Hakkarainen A, Söderlund S, Westerbacka J, Lundbom N, Taskinen MR. Visceral fat saturation is positively correlated with liver fat content. *Proc ISMRM.* 2011:850.
 28. Haacke, EM., Brown, RW., Thompson, MR., Venkatesan, R. *Magnetic Resonance Imaging: Physical Principles and Sequence Design.* John Wiley & Sons; 1999. p. 614(Eq. 20.85)
 29. de Rochefort L, Brown R, Prince MR, Wang Y. Quantitative MR Susceptibility Mapping Using Piece-Wise Constant Regularized Inversion of the Magnetic Field. *Magn Reson Med.* 2008; 60:1003. [PubMed: 18816834]
 30. Kühn J-P, Hernando D, Muñoz del Rio A, Evert M, Kannengiesser S, Völzke H, Mensel B, Puls R, Hosten N, Reeder SB. Effect of Multipeak Spectral Modeling of Fat for Liver Iron and Fat Quantification: Correlation of Biopsy with MR Imaging Results. *Radiology.* 2012; 265:133. [PubMed: 22923718]
 31. Lee JT, Liau J, Murphy P, Schroeder ME, Sirlin CB, Bydder M. Cross-sectional investigation of correlation between hepatic steatosis and IVIM perfusion on MR imaging. *Magn Reson Imag.* 2012; 30:572.
 32. Hernando D, Haufe WM, Hooker CA, Schlein A, Wolfson T, Artz NS, Reeder SB, Sirlin CB. Relationship between liver proton density fat fraction and R2* in the absence of iron overload. *Proc ISMRM.* 2015:4118.
 33. Mamidipalli A, Hamilton G, Manning P, Hong CW, Park CC, Wolfson T, Hooker J, Heba E, Schlein A, Gamst AM, Durelle J, Paiz M, Middleton MS, Schwimmer JB, Sirlin CB. Cross-Sectional Correlation Between Hepatic R2* and Proton Density Fat Fraction (PDFF) in Children With Hepatic Steatosis. *J Magn Reson Imaging.* (Early View 2017)
 34. Zhu A, Hernando D, Johnson K, Reeder SB. Quantification of Short-T2* Signal Components in the Liver Using Radial 3D UTE Chemical Shift-Encoded MRI. *Proc ISMRM.* 2017:120.

Figure 1.

Images acquired with the “zoom” (left) and “whole” (right) gradients (40 mT/m and 23 mT/m, respectively) with otherwise identical imaging parameters (based on typical parameters used for a liver examination). The window / level settings were adjusted to emphasize the background noise. Note in the zoom image the noise bias in the top left corner is approximately two times higher than at the center of the image (10.8 ± 2.7 versus 5.6 ± 1.4). In the whole image the noise bias is uniform (6.5 ± 1.6 versus 5.8 ± 1.4).

Figure 2.

Example plots showing the fits of the original (left) and optimized (right) models for 16 acquired echoes. The signal is in arbitrary units but has been normalized by the background noise. Four subjects with varying fat fractions are illustrated. Each row corresponds to a different subject.

Figure 3.

Variation in mean PDFF with number of echoes (note only the 16 echo datasets are shown). The model with no correction for biases (circle markers) exhibits a decreasing zig-zag dependence of the estimated PDFF versus number of echoes. The different lines show the cumulative effect of applying bias corrections in order of their contribution to the change in residual norm (Table 2). The dependence of PDFF on number of echoes is greatly reduced after applying the corrections. Slopes and correlation coefficients are given in Table 3.

Figure 4. Bland Altman plots of MRI versus MRS for the original and optimized models. Both indicate high agreement between MRI and MRS; the mean bias and 95 % confidence intervals are 1.35 ± 2.63 for (original) and 0.31 ± 2.67 (optimized), indicating reduced bias with the optimized parameters.

Figure 5.

Plot of $R2^*$ versus PDFF for the original and optimized models. Without bias corrections there is a positive correlation whereas after correction there is a negative correlation. As noted in the text, these effects are within the range that can be explained by the magnetic susceptibility difference between water and fat. The $R2^*$ correction parameters (c_0 and c_1) have a strong influence on the slope obtained for the regression and the slope can be made zero for certain values.

Table 1

Best-fit parameters and residual norms for the bias parameters are shown for several discrete values of ndb_1 (the last 2 shaded columns). The global minimum of $\|r\|$ occurs at $\text{ndb} = -5$ (indicated by the shaded row) but corresponds to non-physical values of the other parameters. Physically reasonable values for ndb and c_1 are obtained only in the range $\text{ndb}_1 \pm 1$ (indicated by the *italicized* text).

Gaussian (s^{-2})	$R2^*$ (s^{-1}) c_0	$R2^*$ (s^{-1}) c_1	Water freq. (ppm)	Noise bias	Dbl. bonds ndb_0	Dbl. bonds ndb_1	Residual norm $\ r\ $
1005.6	-47.1	218.1	4.68784	0.707	2.271	1	5.3850
1007.0	-45.1	209.7	4.68781	0.704	2.504	0	5.3675
1008.3	-43.2	201.5	4.68780	0.701	2.737	-1	5.3534
1009.4	-41.4	193.6	4.68779	0.697	2.970	-2	5.3428
1010.3	-39.6	185.9	4.68782	0.692	3.202	-3	5.3355
1011.1	-38.0	178.5	4.68782	0.688	3.434	-4	5.3315
1011.8	-36.4	171.3	4.68784	0.683	3.666	-5	5.3307
1012.3	-34.9	164.5	4.68788	0.678	3.898	-6	5.3331

Table 2

Parameter values for the original and optimized models. Uncertainties represent 95% confidence intervals. Note that a precise value for ndb_1 could not be obtained by optimization but was fixed at -1 , as described in the text. The final column is the change in residual norm ($\|r\|$) following incremental addition of each parameter into the model (ordered from greatest change to least).

Parameter		Original $\ r\ = 9.4975$	Optimized $\ r\ = 5.3534$	$\ r\ _{\text{total}} 4.1441$
Gaussian (s^{-2})		0	1008 ± 50	57.8 %
No. double bonds	ndb_0	≈ 1.92	2.737 ± 0.132	31.0 %
	ndb_1	0	-1 (fixed)	
$R2^*$ (s^{-1})	c_0	0	-43.2 ± 6.0	8.8 %
	c_1	0	201.5 ± 23.1	
Water freq. (ppm)		4.70	4.6878 ± 0.003	1.6 %
Noise bias		0	0.701 ± 0.173	0.8 %

Table 3

Regression analysis for the dependence of PDFF on number of echoes (c.f. Figure 3). The correlation coefficient (r) measures the strength of the linear association. The decreasing correlation puts a numeric value on the qualitative observations in Figure 3 that the dependence of PDFF on number of echoes becomes progressively weaker as bias corrections are incrementally added to the model.

Parameter	Slope (10^{-2})	r	p-value
Original	-0.099	0.044	0.024
Gaussian	-0.111	0.049	0.012
ndb	-0.059	0.025	0.200
R2* offset	-0.038	0.017	0.392
Water freq.	-0.036	0.016	0.410
Noise bias	-0.031	0.013	0.494