Published in final edited form as:

Magn Reson Imaging. 2011 May ; 29(4): 497–506. doi:10.1016/j.mri.2010.11.005.

A simulation-based comparison of two methods for determining relaxation rates from relaxometry images

Kenneth Marro^{1,2}, Randall Otto¹, Orpheus Kolokythas², Akiko Shimamura³, Jean E. Sanders⁴, George B. McDonald⁴, and Seth D. Friedman¹

¹Department of Radiology, Seattle Children's Hospital, Seattle WA

²Department of Radiology, University of Washington, Seattle WA

³Department of Hematology, Seattle Children's Hospital, Seattle, WA

⁴Fred Hutchinson Cancer Research Center, Seattle, WA

Abstract

When assessing liver iron content using relaxometry, an average relaxation rate (R_1 , R_2 or R_2^*) is usually determined from a region of interest, or the entire liver. This is commonly performed by fitting the signal decay in individual voxels to an appropriate relaxation function. The voxel-level parameters resulting from the fits are combined to determine the average relaxation rate and an empirically derived calibration curve is used to convert this single value to iron content. The goal of this study was to compare the precision and accuracy of this voxel-wise fitting to an alternative method that relies on first averaging the signals from all voxels within the region of interest and then determining the relaxation rate from a single fit. Systematic differences were observed when both methods were applied to clinical images. Mathematical simulations were employed to determine which method provided more robust estimates of the true relaxation rate. The mathematical simulations were then expanded to include a range of conditions expected in typical relaxometry images. The results show that voxel-wise fitting skews the relaxation rate estimates and increases variance, particularly when the true relaxation rate is moderate to fast, as it would be in liver with high iron content. The potential impact of these results on clinical decisions is discussed.

INTRODUCTION

MR relaxometry is now widely employed for measuring hepatic iron overload, a condition that poses a serious health problem for patients with genetic hemochromatosis, hyperabsorption of iron from the intestinal tract, and anemias requiring multiple blood transfusions. As the accuracy of MRI estimation of liver iron content approximates that of biochemical measurements from liver tissue, MRI is replacing invasive liver biopsy as the method of choice for liver iron determination. A recent report delineates MRI methods for assessing liver iron in multi-site clinical trials [1], building on initial reports describing calibration curves [2–6]. Assessing liver iron by MRI relies on the mechanism that

© 2010 Elsevier Inc. All rights reserved

Corresponding author: Kenneth Marro Department of Radiology Box 357115 University of Washington Seattle, WA 98195-7115 marro@u.washington.edu Phone: 206-543-3499 Fax: 206-543-3495.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

increasing iron content decreases transverse (T2 and T2*) and longitudinal (T1) relaxation times. The conventional way of expressing the relaxation times is in terms of rates: $R2=1/T2$, $R2^*=1/T2^*$, and $R1=1/T1$.

The typical liver relaxometry protocol is to acquire images at multiple echo or inversion times and then estimate amplitude versus time using an appropriate equation. In nearly all published protocols, the line fitting is performed on a voxel-by-voxel, or pixel-by-pixel, basis [7–14]. Voxel-wise fitting (VWF) generates R2, R2*, or R1 maps. Although there has been some suggestion that these maps contain clinically relevant information [4,15], the detailed distribution of relaxation rates is typically not used for diagnostic purposes. Instead, the average relaxation rate within a manually-defined region of interest (ROI) [1] or an entire cross-sectional [5,6] is computed and converted to average iron content using an empirically determined calibration curve.

An alternative to the VWF method is to first average the MR signal amplitude in multiple voxels within a defined ROI and then perform a single fit to the data [1], a method we will refer to as average-then-fit (ATF). This approach is much less computationally demanding since it requires only a single fit. While it sacrifices information about spatial distribution of iron, the ATF method has the advantage of fitting to data with higher signal to noise ratio (SNR) than the VWF method. In a recent paper describing the nuances of different modeling approaches, Beaumont et al. suggest that differences between the VWF and ATF methods should be evaluated for possible sources of error in estimates of relaxation rates [16]. However, a direct comparison between the VWF and ATF approaches has not been reported to date.

In our own center, while using both methods to analyze clinical data, we noted that the VWF method provided systematically higher estimates of liver iron than the ATF method in patients with high hepatic iron content (HIC). We undertook this study to better characterize this discrepancy and to provide some quantitative understanding of how accuracy and precision of the two methods change as a function of the true relaxation rate. We developed a mathematical model that closely simulated the noise conditions and relevant parameters in our relaxometry images. The mathematical simulations allowed us to assess the performance of the VWF and ATF methods when the true underlying relaxation function was known. We observed strong correlations between the modeling results and the in vivo results, providing some validation of the assumptions on which the model was based. We then expanded the model to include a broader range of conditions and parameters that might be relevant for acquisition protocols employed at other institutions.

The primary result we report is that the ATF method provides superior accuracy and precision than the VWF method under almost all the conditions we studied. The VWF method contains noise-dependent biases that increase the likelihood of incorrect clinical decisions due to variations in image quality, particularly in patients with iron content near the threshold at which therapy is recommended. The VWF method also displayed substantially higher variance than the ATF method when the true relaxation rate was high, suggesting that the ATF method would be the preferred method for monitoring response to therapy in patients with high HIC. These results raise questions about the validity of the VWF method as a basis for clinical decisions and suggest that existing protocols should be modified to implement the ATF method.

MATERIAL AND METHODS

In vivo relaxometry imaging

After institutional approval from the Seattle Children's Hospital Human Subjects Review Committee, we performed a retrospective study in 20 pediatric subjects (8 females, 12 males, age range 5 to 21 years) with HIC ranging from normal to high. Single breath-hold, gradient-recalled-echo (GRE) images were acquired. The echo time varied from 1.3 to 22.1 msec in 12 equal steps of 1.89 msec. The in-plane resolution varied from 2.3×2.3 to 3.2×3.2 mm, depending on the size of the subject. Other acquisition parameters included: TR = 200 msec, 10 mm slice thickness, 12 slices.

In each subject, both the VWF and ATF methods were used to obtain $R2^*$ estimates from eight relatively homogenous liver ROIs consisting of 36 voxels (6×6), as shown in Figure 1. This allowed comparison of descriptive parameters for ATF (e.g. standard deviation) that would not otherwise be available. The equation used for the fitting was a monoexponential decay function with a constant offset [6,17]:

$$S(\tau) = A \times \exp(-R2^* \times \tau) + C \quad (\text{Eq. 1}),$$

where τ is the echo time. The equation parameters, A , $R2^*$ and C , were restricted to be greater than zero and optimized with the standard nonlinear least squares fitting routine available in the Curve Fitting Toolbox of Matlab version 7.5.0 (The Mathworks Inc., Natick MA).

For the VWF method, equation 1 was fit to the signal versus echo time in each of the 36 voxels within the ROI. The results were then averaged to obtain the mean $R2^*$ within the ROI. This yielded eight $R2^*$ values, from which the mean and standard deviation were calculated for each subject.

For the ATF method, the signal was first averaged over the 36 voxels within each ROI and then a single fit to equation 1 was obtained. This also yielded an $R2^*$ estimate from each of the eight ROIs, from which the mean and standard deviation were calculated for each subject.

The image noise content was estimated by determining the coefficient of variation of parameter A in equation 1 as estimated from fits to all the voxels in the eight liver ROIs. The coefficients of variation ranged from 6% to 8% and this range of noise values appears to be typical for this type of image [18]. A noise content of 7% was used in the simulation analyses described below.

Mathematical simulation protocol

The intent of the modeling was to compare the relaxation estimates provided by the VWF and ATF methods when applied to functions that were analogous to in vivo data but where the true underlying relaxation rates were known. The modeling was designed to mimic the VWF and ATF methods, as applied to the in vivo images. MATLAB was used to generate ideal (noise free) free induction decays (FIDs) with relaxation rates representative of those that would be expected in liver tissue with normal to high HIC (A). The FIDs were sampled at discrete time points analogous to the echo train used for in vivo imaging (B). Random noise with a known standard deviation was added to the sampled FIDs (C). Estimates of the relaxation rates were then obtained using both the VWF and ATF fitting methods (D).

A. Create an ideal free induction decay, $FID_{true}(t)$ —The following equation was used to generate the true relaxation functions:

$$FID_{true}(t) = A1_{true} \times \exp(-R21_{true} \times t) + A2_{true} \times \exp(-R22_{true} \times t), \quad (\text{Eq. 2})$$

where $A1_{true}$ and $A2_{true}$ are the nominal amplitudes (arbitrary units) and $R21_{true}$ and $R22_{true}$ are the true relaxation rates (in sec^{-1}) of the assumed biexponential relaxation function for liver.[19] Analyses were run with $R21_{true}$ set to 50, 100, 150, 200, 250, 300, 400, 500, 600, 800, and 1000 sec^{-1} . In all analyses $R22_{true}$ was set to 33 sec^{-1} and the relative values of $A1_{true}$ and $A2_{true}$ (0.9 and 0.1, respectively) were selected to roughly approximate the curves generated by Ghugre et al [19].

B. Sample $FID_{true}(t)$ at discrete time points—The ideal FIDs were sampled at 12 discrete time points, $\tau = \{1.29, 3.14, 5.0, 6.94, 8.84, 10.74, 12.64, 14.54, 16.44, 18.34, 20.24, \text{ and } 22.14\}$ msec, analogous to the gradient echo train employed to acquire the in vivo images. The sampled, ideal FID was referred to as $FID_{true}(\tau)$.

C. Add noise to $FID_{true}(\tau)$ —Analysis of the in vivo relaxometry images indicated that noise content was about 7% of the nominal signal amplitude (see above). Equivalent noise was introduced to the ideal FIDs by adding random values, ϵ , to each sampled point, where ϵ is taken from a set of random numbers with a standard deviation of 7% of the nominal amplitude of FID_{true} (i.e., $A1_{true} + A2_{true}$). After adding ϵ , the noisy FIDs, referred to as $nFID(\tau)$, were converted to absolute values, as would be the case for in vivo relaxometry images where magnitude images are most commonly used.

D. Fit $nFID(\tau)$ using equation 1—The VWF estimates were obtained by repeating step C 36 times, fitting each resulting $nFID(\tau)$ separately and then averaging the 36 $R2$ estimates together. For the ATF method, step C was also repeated 36 times but the FIDs were then averaged together and the $R2$ estimate was obtained from a single fit to equation 1. In both cases the estimates were normalized by the $R2$ estimate obtained by fitting the corresponding, noise-free $FID_{true}(\tau)$. The justification for normalizing by this $R2$ value is that, while it is not a faithful representation of the true relaxation function (since the true relaxation was biexponential and the model was monoexponential), it represents the best estimate of the true relaxation function that could be generated by the fitting routine. The normalized estimates were referred to as $R2_{norm}$.

Step D was repeated eight times for each fitting method to obtain eight estimates of $R2$ (analogous to the eight ROIs placed in each of the in vivo images) for each of the values of $R21_{true}$. The eight $R2$ results were then used to determine the mean and standard deviation of the estimates as a function of $R21_{true}$.

The methods described above allowed comparison and validation of the modeling results against in vivo images acquired with the specific parameters described above. To allow more general interpretation to other acquisition protocols, the model was then expanded to include variable noise content and echo train timing. Ten different noise levels were explored by setting the standard deviation of ϵ to 1% through 12% in 10 equal increments. The echo train timing was varied by adding a variable time shift to τ , as defined in B, above. In this way the timing of the shortest echo in the train, τ_{min} , was varied from 0.5 to 3.5 msec in 12 equal steps.

Additional simulations were undertaken to assess how the number of voxels included in the ROI affects the performance of the VWF and ATF methods. For these simulations the noise

level was constant at 7%, $TE_{\min}$ was 2 msec and the nominal value of $R21_{\text{true}}$ was 200 sec^{-1} . The general procedures described above were repeated while allowing the number of voxels to vary from 1 to 128 in 32 equal steps. The parameters evaluated were the mean and coefficient of variation of $R2_{\text{norm}}$ from each of the two fitting methods. Finally, to explore a more realistic in vivo scenario, in which the true relaxation rate varies from voxel to voxel, this analysis was repeated using variable $R21_{\text{true}}$ values with a mean of 200 sec^{-1} and standard deviation of 45 sec^{-1} .

RESULTS

In vivo images

Figure 1 shows an example liver image, from a subject with high HIC, and the location of the eight ROIs in which $R2^*$ estimates were obtained. As shown in the table, the estimates obtained from the VWF method, $R2^*_{\text{VWF}}$, were consistently higher than those obtained with the ATF method, $R2^*_{\text{ATF}}$. In this particular subject, the ratio, $R2^*_{\text{VWF}}/R2^*_{\text{ATF}}$, ranged from 1.1 to 2.6, with a mean of 1.5. The standard deviation of the eight VWF estimates was also higher, by a factor of about 2.7, than the standard deviation of the ATF estimates.

Figure 2 shows the ratios of the means ($R2^*_{\text{VWF}}/R2^*_{\text{ATF}}$) and standard deviations ($SD_{\text{VWF}}/SD_{\text{ATF}}$) when both methods were applied to the cohort of 20 patients. The data are plotted against the nominal $R2^*$ estimates obtained using the ATF method. The VWF estimates were generally higher and exhibited more volatility (higher standard deviation) than the ATF estimates. The discrepancies between the estimates increased as the nominal relaxation rate increased.

Simulation results

The mathematical simulations revealed the same patterns observed in the in vivo results. The solid lines in Figure 2 show the ratios of the means and standard deviations of the VWF and ATF estimates obtained from the simulations. The ratio, $R2^*_{\text{VWF}}/R2^*_{\text{ATF}}$, was greater than one over the entire range of relaxation rates included in the model and tended to increase as the nominal $R2^*$ increased. At the highest relaxation rate studied (1000 sec^{-1}) the VWF estimates were about 1.35 times higher than the ATF estimates. The ratio, $SD_{\text{VWF}}/SD_{\text{ATF}}$, was also consistently greater than 1, except when the nominal relaxation rate was less than about 200 sec^{-1} , and increased to as high as 2.5 as the nominal relaxation rate increased.

Figure 3 shows surface plots of the model's VWF and ATF estimates of $R2$ when the noise content and the echo train timing were allowed to vary. To allow interpretation of the accuracy of the estimates, they have been normalized by the $R2^*$ estimate obtained by fitting the corresponding, noise-free FID (see Methods). The figure contains surface plots for three different values of $R21_{\text{true}}$, 42, 200, and 1000 sec^{-1} , corresponding to normal, intermediate and high HIC, respectively. For the normal and intermediate relaxation rates, the ATF estimates are all very close to 1, indicating a high degree of accuracy under these conditions. In contrast, the VWF estimates at the normal and intermediate relaxation rates reflected a noise-dependent bias. When the noise content was low the VWF estimates of $R2_{\text{norm}}$ were close to 1 but they increased as the noise content increased. These results indicate that the discrepancies in the $R2^*$ estimates shown in Figures 1 and 2 arise from a tendency for the VWF method to overestimate the relaxation rate rather than for the ATF method to underestimate it. At the highest value of $R21_{\text{true}}$ studied (1000 sec^{-1}), echo time and noise exerted competing effects on the $R2$ estimates. Increasing noise corresponded to increasing overestimation of $R2^*$, while increasing echo train timing tended to reduce the estimates of

$R2^*$. The biases in the estimates were observed for both fitting methods but were more pronounced with the VWF method.

Under most conditions the ATF method also resulted in higher precision when compared to the VWF method, as measured by the standard deviation of the $R2$ estimations (Figure 4). The VWF method generated more stable results at the lowest value of $R21_{\text{true}}$ analyzed, 42 sec^{-1} , but the standard deviations for both methods were relatively low under this condition. The volatility of the estimates at the intermediate and high values of $R21_{\text{true}}$ (200 and 1000 sec^{-1} , respectively) was consistently higher for the VWF method, by as much as a factor of 5 under the least optimal conditions ($TE_{\text{min}} = 3.5 \text{ msec}$, $\varepsilon = 12\%$). This pattern is consistent with the in vivo results shown in Figure 2.

The results in Figures 3 and 4 assumed an idealized situation in which the true relaxation rate was uniform in all the voxels involved in the analysis. The results depicted in Figure 5 may represent a more realistic in vivo scenario, in which the true relaxation rate varies from voxel to voxel. The solid lines in the graphs show the results when $R21_{\text{true}}$ had a constant value of 200 sec^{-1} and the dashed lines show the results when $R21_{\text{true}}$ was allowed to vary from this value with a standard deviation of 45 sec^{-1} . To simplify the presentation of these results, the noise level was held constant (7%) and a single value of TE_{min} (1.8 msec) was included in the analysis. For the ATF method, the solid lines and dashed lines nearly overlapped, indicating that variability in $R21_{\text{true}}$ had virtually no effect, when the number of voxels was 15 or higher. However, for the VWF method, variability in $R21_{\text{true}}$ amplified the bias and increased the volatility in the estimates of $R2_{\text{norm}}$.

DISCUSSION

Our intention in this project was to compare the accuracy and precision of two widely implemented methods of estimating average relaxation rates from liver images. There are many numerical models and optimization algorithms that can be implemented with relaxometry images [16,18,20]. Each has advantages and disadvantages and there is no clear consensus as to the best approach to use. There is also no clear consensus as to the nature of the true underlying transverse relaxation function in iron loaded liver, which has been suggested to be biexponential [19] or nonexponential [21,22] in form. As a result, calibration curves have been reported with monoexponential, monoexponential-plus-a-constant, or biexponential functions, with the monoexponential-plus-a-constant method being the most broadly implemented [6,17,23].

The general fitting approach that we analyzed, nonlinear least squares fitting to a monoexponential function with a constant baseline offset, is probably the most common approach described in the liver iron literature [15]. It is most frequently applied on a voxel-by-voxel basis to generate relaxation maps [7–14,17]. While there have been suggestions that the voxel-level variance contained in these maps may contain physiologically relevant information concerning the effects of chelation therapy [4,15], a single relaxation rate is typically used to calculate the iron content on which therapeutic decisions are based. This rate is obtained by averaging the voxel-level relaxation rates within a relatively large ROI. In this paper, we compare this method to a simple alternative that also provides a single relaxation rate within an ROI, which is to average the MR signal intensities within the ROI first, then performing a single fit.

While intuition might suggest that these two methods should lead to equivalent results, our analysis reveals systematic differences in the relaxation rate estimates they provide. For our in vivo images, the VWF method provided higher estimates of $R2^*$ over the entire range of nominal relaxation rates in the subject pool (Figure 2A). The difference between the

estimates was largest when the nominal relaxation rate was very high, with the ratio of the VWF estimate to the ATF estimates rising as high as 1.5.

The modeling results closely followed this pattern and the results in Figure 3 indicate that the discrepancy arises from a tendency for the VWF method to overestimate $R2^*$ rather than for the ATF method to underestimate it. Under virtually all of the conditions we analyzed with the modeling, the estimates provided by the ATF method were more accurate, with $R2_{\text{norm}}$ remaining close to 1 except when the true relaxation rate was high and the acquisition parameters were least optimal (i.e. long echo train timing and high noise content).

This systematic overestimation of relaxation rates by the VWF method does not necessarily lead to incorrect estimates of iron content. The conversion from relaxation rate to iron content is usually accomplished using empirically derived calibration curves. Most clinical centers do not create their own calibration curves but instead use one of several published relationships between relaxation rates and biopsy results. If the acquisition parameters and image quality were the same, these calibration curves would presumably contain the same biases as the clinical images. However, our results reveal two additional weaknesses in the VWF method that should be causes for concern to clinicians. First, the $R2$ estimates provided by the VWF method are dependent on the noise content in the images. The modeling results in Figure 6 show the range of $R2_{\text{norm}}$ values that might be expected from the two fitting methods when $R21_{\text{true}}$ is near the threshold at which decisions concerning chelation therapy might be made (200 sec^{-1}) and the noise content is allowed to vary. The VWF estimates ranged from 0.94 to 1.7. This raises the possibility that patients below the threshold may receive unnecessary treatment. This possibility appears to be reduced with the ATF method, for which the corresponding range of $R2_{\text{norm}}$ estimates was 1.0 to 1.17.

The second weakness in the VWF method that could have negative clinical consequences is that the standard deviation of the estimates is higher than that of the ATF method when the true relaxation rate is high. This tendency is clear in the graph of Figure 2B, where the standard deviation of the VWF estimates was roughly 2 to 3 times higher than that of the ATF methods when the relaxation rate was high. Here again, there was strong agreement between the in vivo results and the modeling results. The higher volatility of the VWF estimates is also apparent in the surface plots for the intermediate and high relaxation rates in Figure 4. With moderate noise content of 7%, the standard deviations of the VWF estimates was generally 2 to 3 times higher than those of the ATF method when $R21_{\text{true}}$ was 200 s^{-1} or higher. This suggests that the ATF method would be a better tool for use in monitoring the effects of therapy in patients with high HIC.

It may be appropriate to criticize the inherent assumption in our model that the relaxation rate is constant for all voxels within the ROI. In real images, the relaxation rate may vary within any reasonably sized ROI [19]. The results of Figure 5 are intended to address this issue. The figure shows the effect on the estimates from the two fitting methods when the true relaxation rate was allowed to vary from voxel to voxel. For the VWF method, the variability in the true relaxation rate resulted in increased overestimation of $R2$ and increased variability in the estimates. These effects were independent of the number of voxels included in the ROI. In contrast, variability in the true relaxation rate had only negligible effect on the ATF estimates as long as the number of voxels was greater than about 15. We repeated these simulations for other values of ϵ , TE_{min} , and $R21_{\text{true}}$ and the pattern was consistent (data not shown). These modeling results suggest that local relaxation rate variations that may be present in vivo would further diminish the accuracy and precision of the VWF method but would have negligible effect on the ATF method. Further, these results suggest that at least some of the local variations that are often seen in relaxation rate

maps are likely to arise from volatility in the VWF fitting method. That is, even if the in vivo relaxation rates were perfectly homogenous, the VWF method would erroneously create heterogeneous maps and the heterogeneity would grow worse as the relaxation rate increased.

One reason for the superior performance of the ATF method is that averaging the signal from multiple voxels increases the SNR for the fitted data. We averaged signals from 36 voxels prior to fitting, increasing the SNR by a factor of 6, and this should reasonably lead to improved precision in the ATF fits. However, averaging the 36 relaxation rates obtained from fits to individual voxels within an ROI, as is done with the VWF method, should also generate an SNR benefit in the $R2^*$ estimate and lead to similar levels of precision for the two methods. While this pattern tended to hold true when $R21_{\text{true}}$ was low, the results of Figure 3 suggest that the performance of the two methods is also complicated by other factors.

One feature to note in Figure 3 is that, when the noise content was very low, the relaxation rate estimates from the two methods were similar. As the noise content increased, there was a general tendency for the VWF estimates to be higher than the ATF estimates. The reason for this is the tendency for the distribution of VWF fits to transform from Gaussian to Rician as the noise level increases [24,25]. This phenomenon can be seen in Figure 7, which displays histograms of VWF estimates of $R2_{\text{norm}}$ resulting from 512 fits when $R21_{\text{true}}$ was set to 200 sec^{-1} , and TE_{min} to 1.8 msec. At low noise levels, $\epsilon = 2\%$, the $R2_{\text{norm}}$ values displayed a generally Gaussian distribution ranging from 0.8 to 1.2 with a mean of 1. As the noise content increased, the range of $R2_{\text{norm}}$ values also increased, as expected. At the intermediate noise level the $R2_{\text{norm}}$ values ranged from about 0.5 to 3. The shape of the distribution also began to change from Gaussian to Rician. The Rician shape is more obvious at the highest noise level depicted, $\epsilon = 11\%$. Some values of $R2_{\text{norm}}$ were over 9 and, since the model constrained $R2^*$ to be non-negative, there were no offsetting values below 0 and the distribution became more asymmetric. This resulted in a mean $R2_{\text{norm}}$ that was substantially greater than 1. This phenomenon is less apparent with the ATF method because prior averaging reduces the chance that fitting will produce extreme values of $R2^*$, narrows the spread in the distribution of $R2^*$ values, and raises the threshold at which the distribution becomes Rician.

The primary disadvantage to the ATF method is that any information about nonuniform spatial distribution of the iron content is lost. However, in almost all published protocols, information about iron content is condensed to a single number, which is assumed to represent the average HIC in the whole liver, and the clinical decision concerning treatment is based on this single number. If a sense for the distribution of iron within the liver is desired, the VWF method could be applied to generate an iron content image. Pre-processing using adaptive filtering could be employed to help reduce the effects of noise. Our results suggest that a high degree of caution should be used when interpreting these images because it may not be easy to distinguish true spatial variations in iron content from artifactual variations introduced by the VWF method.

The primary result we report is that the VWF method tends to introduce biases and increase the scatter of estimates of relaxation rates, particularly when the true underlying relaxation rate is moderate to high. These flaws could needlessly increase the chances of misdiagnosis, raising questions about the validity of the VWF method as a clinical tool for assessing liver relaxometry images. Both the precision and accuracy in the determination of relaxation rates from exponential fits to MR image data are substantially improved when the signal from multiple voxels is averaged first and then processed with a single fit. In general, the ATF method extracts more valuable information from MR relaxometry data and, since there was

nothing in our modeling approach that would restrict interpretation only to liver, this observation is likely to hold true for other organs.

Acknowledgments

This work was supported by NIH grants CA18029 (NCI), HL36444 (NHLBI), CA15704 (NCI), HL0795 (NHLBI), and EB009438 (NIBIB). We also wish to acknowledge the assistance of Gary McNeal of Siemens Medical Systems.

REFERENCES

1. Meloni A, Ramazzotti A, Positano V, Salvatori C, Mangione M, Marcheschi P, Favilli B, De Marchi D, Prato S, Pepe A, Sallustio G, Centra M, Santarelli MF, Lombardi M, Landini L. Evaluation of a web-based network for reproducible T2* MRI assessment of iron overload in thalassemia. *Int J Med Inform.* 2009; 78(8):503–512. [PubMed: 19345609]
2. Stark DD, Moseley ME, Bacon BR, Moss AA, Goldberg HI, Bass NM, James TL. Magnetic resonance imaging and spectroscopy of hepatic iron overload. *Radiology.* 1985; 154(1):137–142. [PubMed: 3964933]
3. Bulte JW, Miller GF, Vymazal J, Brooks RA, Frank JA. Hepatic hemosiderosis in non-human primates: quantification of liver iron using different field strengths. *Magn Reson Med.* 1997; 37(4): 530–536. [PubMed: 9094074]
4. Clark PR, St Pierre TG. Quantitative mapping of transverse relaxivity ($1/T(2)$) in hepatic iron overload: a single spin-echo imaging methodology. *Magn Reson Imaging.* 2000; 18(4):431–438. [PubMed: 10788721]
5. St Pierre TG, Clark PR, Chua-anusorn W, Fleming AJ, Jeffrey GP, Olynyk JK, Pootrakul P, Robins E, Lindeman R. Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. *Blood.* 2005; 105(2):855–861. [PubMed: 15256427]
6. Wood JC, Enriquez C, Ghugre N, Tyzka JM, Carson S, Nelson MD, Coates TD. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. *Blood.* 2005; 106(4):1460–1465. [PubMed: 15860670]
7. Argyropoulou MI, Astrakas L. MRI evaluation of tissue iron burden in patients with beta-thalassaemia major. *Pediatr Radiol.* 2007; 37(12):1191–1200. quiz 1308–1199. [PubMed: 17710390]
8. Argyropoulou MI, Kiortsis DN, Astrakas L, Metafratzi Z, Chalissos N, Efremidis SC. Liver, bone marrow, pancreas and pituitary gland iron overload in young and adult thalassemic patients: a T2 relaxometry study. *Eur Radiol.* 2007; 17(12):3025–3030. [PubMed: 17549485]
9. Kim D, Jensen JH, Wu EX, Sheth SS, Brittenham GM. Breathhold multiecho fast spin-echo pulse sequence for accurate R2 measurement in the heart and liver. *Magn Reson Med.* 2009; 62(2):300–306. [PubMed: 19526516]
10. Olynyk JK, Gan E, Tan T. Predicting iron overload in hyperferritinemia. *Clin Gastroenterol Hepatol.* 2009; 7(3):359–362. [PubMed: 19095082]
11. Papakonstantinou O, Alexopoulou E, Economopoulos N, Benekos O, Kattamis A, Kostaridou S, Ladis V, Efsthathopoulos E, Gouliamos A, Kelekis NL. Assessment of iron distribution between liver, spleen, pancreas, bone marrow, and myocardium by means of R2 relaxometry with MRI in patients with beta-thalassemia major. *J Magn Reson Imaging.* 2009; 29(4):853–859. [PubMed: 19306409]
12. Schein A, Enriquez C, Coates TD, Wood JC. Magnetic resonance detection of kidney iron deposition in sickle cell disease: a marker of chronic hemolysis. *J Magn Reson Imaging.* 2008; 28(3):698–704. [PubMed: 18777554]
13. Schwenzer NF, Machann J, Haap MM, Martirosian P, Schraml C, Liebig G, Stefan N, Haring HU, Claussen CD, Fritsche A, Schick F. T2* relaxometry in liver, pancreas, and spleen in a healthy cohort of one hundred twenty-nine subjects-correlation with age, gender, and serum ferritin. *Invest Radiol.* 2008; 43(12):854–860. [PubMed: 19002057]

14. Song R, Cohen AR, Song HK. Improved transverse relaxation rate measurement techniques for the assessment of hepatic and myocardial iron content. *J Magn Reson Imaging*. 2007; 26(1):208–214. [PubMed: 17659538]
15. Clark PR, Chua-anusorn W, St Pierre TG. Bi-exponential proton transverse relaxation rate (R2) image analysis using RF field intensity-weighted spin density projection: potential for R2 measurement of iron-loaded liver. *Magn Reson Imaging*. 2003; 21(5):519–530. [PubMed: 12878262]
16. Beaumont M, Odame I, Babyn PS, Vidarsson L, Kirby-Allen M, Cheng HL. Accurate liver T2 measurement of iron overload: a simulations investigation and in vivo study. *J Magn Reson Imaging*. 2009; 30(2):313–320. [PubMed: 19629985]
17. Hankins JS, McCarville MB, Loeffler RB, Smeltzer MP, Onciu M, Hoffer FA, Li CS, Wang WC, Ware RE, Hillenbrand CM. R2* magnetic resonance imaging of the liver in patients with iron overload. *Blood*. 2009; 113(20):4853–4855. [PubMed: 19264677]
18. Raya JG, Dietrich O, Horng A, Weber J, Reiser MF, Glaser C. T2 measurement in articular cartilage: impact of the fitting method on accuracy and precision at low SNR. *Magn Reson Med*. 2010; 63(1):181–193. [PubMed: 19859960]
19. Ghugre NR, Coates TD, Nelson MD, Wood JC. Mechanisms of tissue-iron relaxivity: nuclear magnetic resonance studies of human liver biopsy specimens. *Magn Reson Med*. 2005; 54(5): 1185–1193. [PubMed: 16215963]
20. He T, Gatehouse PD, Kirk P, Mohiaddin RH, Pennell DJ, Firmin DN. Myocardial T*2 measurement in iron-overloaded thalassemia: an ex vivo study to investigate optimal methods of quantification. *Magn Reson Med*. 2008; 60(2):350–356. [PubMed: 18666131]
21. Jensen JH, Chandra R. Strong field behavior of the NMR signal from magnetically heterogeneous tissues. *Magn Reson Med*. 2000; 43(2):226–236. [PubMed: 10680686]
22. Jensen JH, Chandra R. Theory of nonexponential NMR signal decay in liver with iron overload or superparamagnetic iron oxide particles. *Magn Reson Med*. 2002; 47(6):1131–1138. [PubMed: 12111959]
23. Alexopoulou E, Stripeli F, Baras P, Seimenis I, Kattamis A, Ladis V, Efstathopoulos E, Brountzos EN, Kelekis AD, Kelekis NL. R2 relaxometry with MRI for the quantification of tissue iron overload in beta-thalassemic patients. *J Magn Reson Imaging*. 2006; 23(2):163–170. [PubMed: 16374880]
24. van der Weerd L, Vergeldt FJ, Adrie de Jager P, Van As H. Evaluation of algorithms for analysis of NMR relaxation decay curves. *Magn Reson Imaging*. 2000; 18(9):1151–1158. [PubMed: 11118770]
25. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. *Magn Reson Med*. 1995; 34(6): 910–914. [PubMed: 8598820]

ROI	$R2^*_{VWF}$	$R2^*_{ATF}$	$R2^*_{VWF}/R2^*_{ATF}$
1	1132	1037	1.1
2	1511	1175	1.3
3	1590	1123	1.4
4	1582	1101	1.4
5	2359	910	2.6
6	1665	1122	1.5
7	1011	723	1.4
8	1396	962	1.5
mean	1531	1019	1.5
St.Dev.	406	149	

Figure 1. Example image

The image, from a patient with high HIC, shows the locations of eight ROIs in which fitting was performed. The table shows the results of the $R2^*$ values (in sec^{-1}) obtained with the VWF ($R2^*_{VWF}$) and ATF ($R2^*_{ATF}$) fitting methods. The right-most column shows the ratio, $R2^*_{VWF}/R2^*_{ATF}$. The VWF estimates are consistently higher than the ATF estimates, with the ratio ranging from 1.1 to 2.6 and a mean ratio of 1.5. The standard deviation of the VWF estimates was also substantially higher, by a factor of 2.7. These patterns are in complete agreement with the predictions of the modeling.

Figure 2. In vivo results

The analysis described in Figure 1 was repeated for 19 additional subjects with a wide range of HIC. Plot A shows the ratio, $R2^*_{VWF}/R2^*_{ATF}$, as a function of the nominal $R2^*$, as estimated by the ATF method. As the nominal $R2^*$ increased, the tendency for the VWF estimates to be higher than the ATF estimates increased. Plot B shows the ratio of the standard deviations of the $R2^*$ estimates for the two methods (SD_{VWF}/SD_{ATF}). The volatility of the $R2^*$ estimates increased for both methods as the nominal value of $R2^*$ increased (see Figure 4) but the increase was higher for the VWF method. The solid lines in both plots are the predicted ratios generated by the modeling program with a noise content of 7%.

Figure 3. Surface plots of the R2 estimates generated by the model

The left column shows the VWF estimates of R2 as a function of noise, ϵ , and minimum echo time, $TE_{\min}$. The right column shows the corresponding R2 estimates obtained from the ATF method. Three surface plots are displayed in each column for the three different true relaxation rates, $R21_{\text{true}} = 42, 200$ and 1000 sec^{-1} . To facilitate comparisons, the R2 values have been normalized by the R2 value obtained by fitting the corresponding noiseless FIDs sampled with the shortest echo train (see Methods). For convenience, the z-axis scales are different but the color shading is consistent for all plots. The VWF method tended to overestimate R2, particularly at intermediate and high relaxation rates, with the overestimation generally increasing as a function of noise. At the highest value of $R21_{\text{true}}$, increasing echo train timing exerted a competing effect, tending to decrease the estimates of R2. The ATF method provided more accurate estimates of R2 at all three relaxation rates.

Figure 4. Surface plots of the normalized standard deviations of the R2 estimates generated by the VWF and ATF methods

For convenience, the z-axis scales are different but the color shading is consistent for all plots. As expected, the standard deviation generally increased as the noise content increased. When $R21_{\text{true}}$ was low, 42 sec^{-1} , the standard deviation was greater for the ATF method. When $R21_{\text{true}}$ was intermediate or high, the standard deviation was greater for the VWF method.

Figure 5. Effects of voxel-to-voxel variations in R2

The plots show the means and standard deviations of the model's relaxation rate estimates as a function of the number of voxels included in the ROI. For these simulations the noise level was 7% and $TE_{\min}$ was 1.8 msec. (Similar results were obtained for other values of these parameters.) When the ROI included only one voxel the VWF and ATF methods were equivalent and yielded identical results. The accuracy and precision of the VWF estimates were generally poorer than the VWF estimates when more than one voxel was included. The solid lines show the results when $R21_{\text{true}}$ had a uniform value of 200 sec^{-1} for all voxels within the ROI. The dashed lines show the results when $R21_{\text{true}}$ was variable, with a mean of 200 sec^{-1} and a standard deviation of 45 sec^{-1} . Voxel to voxel variability of $R21_{\text{true}}$ increased the bias and volatility of the VWF estimates but had only negligible effect on the ATF estimates except when the number of voxels was less than about 15. These results suggest that, if the true relaxation rate varied from voxel to voxel, as might be expected in vivo, the accuracy and precision of the VWF estimates would likely decrease while the ATF estimates would be unaffected as long as the ROI contained more than about 15 voxels.

Figure 6. Potential impact on clinical decisions

The plots show the normalized $R2^*$ estimates from the VWF (solid line) and ATF (dashed line) methods as a function of noise content when $TE_{min} = 1.8$ msec and $R21_{true} = 200$ sec^{-1} , approximately the threshold at which a clinical decision about treatment would be made. For both methods, increasing noise content caused an increase in the tendency to overestimate $R2^*$, raising the possibility that poor image quality could lead to incorrect decisions to treat patients below the threshold. This possibility is higher for the VWF method where the range of estimates was from 0.94 at the lowest noise content to 1.7 at the highest noise content. The corresponding range of estimates for the ATF method was smaller, 0.99 to 1.17.

Figure 7. Histograms of $R2_{\text{norm}}$ fits resulting from the VWF method at three different noise levels

For these simulations, $R21_{\text{true}}$ was set to 200 sec^{-1} and TE_{min} was set to 1.8 msec. To create distributions with better definition, 512 fits were performed, as opposed to the 36 fits used for the other simulations. When the noise level was low, $\epsilon = 2\%$, the $R2_{\text{norm}}$ values fell between 0.8 and 1.2 with a mean of 1 and a generally Gaussian distribution. At the intermediate noise level, $\epsilon = 7\%$, the spread of $R2_{\text{norm}}$ values increased and the distribution began to assume an asymmetric, Rician shape. The Rician shape is more distinct at the high noise level, $\epsilon = 11\%$, where $R2_{\text{norm}}$ ranged nearly as high as 10. Since the model was constrained to $R2$ values that were greater than 1, these large values of $R2_{\text{norm}}$ were not offset by corresponding negative values so the mean of the values increased. The dashed lines in the graphs indicate the mean of the VWF fits. For reference, the corresponding ATF fits are indicated by the dotted lines. Note that the x-axis scales are different.