Influence of the Analysis Technique on Estimating Hepatic Iron Content Using MRI

El-Sayed H. Ibrahim, PhD,[1]* Ayman M. Khalifa, PhD,[2] and Ahmed K. Eldaly, MSc[2]

Purpose: To investigate the effect of the analysis technique on estimating hepatic iron content using MRI.

Materials and Methods: We evaluated the influences of single-exponential (EXP), bi-exponential (BEXP), and exponential-plus-constant (CEXP) models; and pixel-wise (MAP), average (AVG), and median (MED) signal calculation methods on T2* measurement using numerical simulations, calibrated phantoms, and nine patients scanned on 3 Tesla MRI, based on regression, correlation, and t-test statistical analysis.

Results: The T2* measurement error varied from 9 to 51% in the numerical simulations (T2*: 5–20 ms), depending on signal-to-noise ratio (SNR; range: 8–233) with significant ($P < 0.05$) difference between actual and predicted values. The MAP method performed well (error $< 10\%$) at high SNR (>100), but resulted in severe estimation errors at low SNR (<50). The EXP model resulted in significant measurement differences ($P < 0.05$) compared with all other methods, irrespective of SNR. In vivo T2* values ranged from 3.1 to 53.6 ms, depending on the amount of iron overload and implemented analysis method. The BEXP (range: 3.7–50 ms) and CEXP (range: 3.8–53.6 ms) models, and the AVG (range: 3.2–38.8 ms) and MED (range: 3.1–38.5 ms) methods provided more accurate measurements than the EXP model (range: 3.1–18.3 ms) and MAP (range: 3.8–53.6 ms) method, respectively ($P < 0.05$). The BEXP and CEXP models provided very similar measurements ($P > 0.87$). Similarly, the AVG and MED methods provided very similar results ($P > 0.97$), with slightly better performance of the AVG method.

Conclusion: Different analysis techniques show different performances based on the fitting model and signal calculation method. Based on this study, the CEXP model and AVG method are recommended due to simpler implementation and less influence by the selected analysis region.

Iron toxicity IS the main reason for tissue damage and organ failure in patients with iron overload.[1] Excess iron content accumulates in different body organs, although most iron overload is deposited in the liver, with liver iron concentration closely correlating with the degree of iron overload.[2] Accurate evaluation of iron overload is critical for initialization of chelation therapy to minimize, and even reverse, its effects.[1,3] Liver biopsy is the gold standard for quantifying hepatic iron contents. However, the technique's invasiveness, high cost, and influence by the sampling error preclude its widespread use. Serum ferritin could be used for monitoring iron overload; however, this technique is nonspecific and is affected by inflammation and infection.[4]

Recently, MRI has been considered for evaluating iron overload based on T2* measurement.[5–8] At 1.5 Tesla (T), T2* of 20 ms is commonly used as the cutoff for identifying iron overload.[9] Although similar cutoff values have not yet been determined for 3.0T, T2*-weighted imaging at 3.0T has been suggested for quantifying iron overload with high sensitivity, especially when tissue iron levels are low or high image resolution is required.[10] Furthermore, a linear relationship has been shown between R2* ($= 1000/T2^*$) and iron content over the entire clinical range of interest; however, there was no significant correlation between R2* and serum ferritin.[10]

T2* is estimated by fitting the signal intensities acquired at multiple echo times (TEs) to an exponentially decaying curve. Although the analysis process is conceptually straightforward, there are several factors related to the analysis technique that could affect the resulting T2* value,

including the choice between pixel-wise and region-of-interest (ROI) analysis methods and the selected data fitting model. For example, T2* could be calculated pixel-wise (relaxivity map) or using the average or median value inside the ROI. Furthermore, different exponential fitting models could be used in the analysis, including single-exponential decaying model (EXP), bi-exponential signal decaying model (BEXP), and exponential-plus-constant model (CEXP). The expected effects of these factors on the resulting measurements and compare these different analysis methods on T2* estimation, which is the purpose of this study.

Although T2* measurement in the heart is usually performed on the septal wall (because it suffers less susceptibility artifacts than the lateral wall) mostly using single-exponential model, the analysis criteria in the liver are less clear due to the existence of vascular structures and signal-to-noise ratio (SNR) variation that could affect the resulting measurements based on the selected ROI and analysis technique.[6,11] Therefore, we focused in this study on T2* measurements in the liver. It should be noted that besides T2* measurement, other MRI techniques, e.g., based on T2 measurements, have been developed for estimating iron overload in the liver, which are largely used in clinical practice.[12]

In this study, we evaluated the influences of the adopted exponential fitting model and signal calculation method at different SNR on T2* calculation based on numerical simulations, calibrated phantoms, and patients with different degrees of iron overload.

Materials and Methods
Factors Affecting the Analysis Method

MEASUREMENTS INSIDE THE SELECTED ROI. Calculations inside the ROI were conducted using three approaches: (1) pixel-wise (MAP), where exponential fitting is applied to each pixel inside the ROI, followed by obtaining the mean of the resulting T2* values; (2) average (AVG), where the average signal intensity inside the ROI is first calculated, followed by exponential fitting of the resulting values at different TEs; and (3) median (MED), where the median signal intensity inside the ROI is first calculated, followed by exponential fitting of the resulting values at different TEs.

EXPONENTIAL FITTING MODELS. Three different exponential fitting models were used in the analysis, as shown in the following equations (S and S_0 are the signal intensity and its initial value at TE = 0, respectively, and is C a constant)[13,14]:

$$\mathbf{BEXP}: S = 0.9\ S_o e^{(-TE/T_2^*)} + 0.1\ S_o e^{(-TE/200)} \quad (1)$$

$$\mathbf{EXP}: S = S_o e^{(-TE/T_2^*)} \quad (2)$$

$$\mathbf{CEXP}: S = S_o e^{(-TE/T_2^*)} + C \quad (3)$$

Numerical Simulations
Numerical simulations were conducted to test the behavior of different analysis techniques under different SNR levels. Basically, 5 cm² circular regions were generated with different signal intensity based on different T2* ranging from 5 to 20 ms in increments of 5 ms. Twelve "timeframes" were created for each image to simulate signal relaxation for image acquisition at different echo times ranging from 1 to 16 ms in equal increments. Different levels of Gaussian noise, with variance ranging from 1 to 22, were added to the calculated images, resulting in SNRs ranging from 8 to 233 in the first timeframe. The three exponential fitting models (EXP, BEXP, and CEXP) and three signal calculation methods (AVG, MED, and MAP) were used to measure T2* using the Levenberg-Marquardt algorithm. Root mean square errors were calculated between the actual and predicted T2* values, and absolute error values were summed for all SNR levels to determine the analysis technique with minimum estimation error.

Phantom Scans
Calibrated phantoms with known iron concentrations were created (Fig. 1) to compare the T2* values calculated by different analysis techniques with ground truth. Ten 50-mL tubes were filled with agarose-based material, which was created by dissolving 0.5% agarose in distilled water and doping with 0.085 milli-molar of manganese chloride to create gel-like material with T1 and T2 values similar to those in the liver, as previously described.[10] Different amounts of iron-sulphate were added to nine tubes (the tenth tube was used as a reference with no iron content), ranging from 0.5 to 4.5 g in equal increments, which resemble the range of iron overload in the clinical setting. The phantoms were imaged on a 3.0T Siemens Skyra MRI scanner (Siemens Healthcare, Erlangen, Germany) using a body surface coil and a 12-echo gradient echo (GRE) sequence with TEs ranging from 1 to 16.5 ms in equal increments. The imaging parameters were as follows: repetition time (TR) = 200 ms, matrix = 256 × 192; bandwidth = 1776 Hz/pixel, field of view (FOV) = 380 × 285 mm², flip angle = 20°, and slice thickness = 10 mm.

In Vivo Scans
Nine human subjects with sickle cell disease (five males, four females; age = 38 ± 12 years) were imaged on the 3.0T Siemens MRI scanner using imaging sequence and parameters similar to those used in the phantom experiment (except for FOV that depended on the patient size, and matrix that was adjusted accordingly to maintain resolution closer to that in the phantom experiments) after giving informed consent in this Institutional Review Board-approved study. A mid-liver axial slice was acquired in a single end-expiration breath-hold. Figure 2 shows T2* weighted Liver images and the location of ROI used to estimate the T2*.

Data Analysis
An in-house software was created in Matlab to analyze the images while modifying the factors that could affect the measured T2*, specifically the signal calculation method (MAP, AVG, and MED);
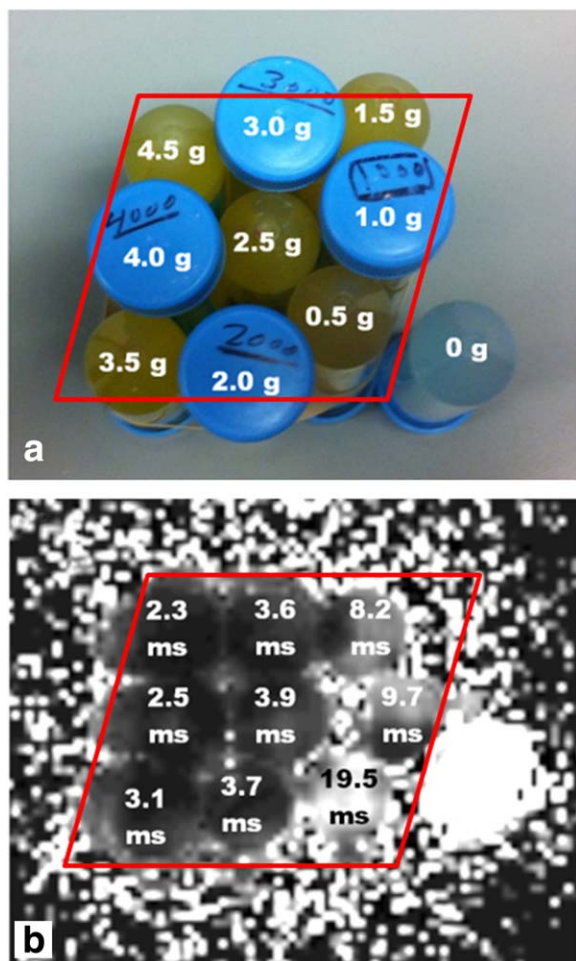
FIGURE 1: Calibrated agarose-based calibrated iron phantom (a) and T2* map (b). The amount of iron content (in grams) is marked on each tube. The T2* values measured on the scanner console are marked for all tube.

exponential fitting model (EXP, BEXP, and CEXP) using the Levenberg-Marquardt algorithm. A 2-cm^2 circular ROI in the center of each phantom tube is analyzed, and a 4-cm^2 circular ROI in the right lobe of the liver away from vasculature is analyzed for each human subject. The location of the selected ROI was not exactly the same for all patients; however, it was selected away from the boundary and vasculature in the region of most homogeneous signal intensity in the right lobe. Regression and correlation analysis was conducted using 2010 Microsoft Excel software to determine the relationship between estimated T2* values in the phantoms using different analysis methods. Statistical t-test was conducted to examine the significance of the in vivo measurement differences by different analysis methods, where $P < 0.05$ was considered statistically significant.

Results

Numerical Simulation Results

The T2* measurement error varied from 9 to 51% in the numerical simulations (T2* range: 5–20 ms), depending on SNR (range: 8–233) with significant ($P < 0.05$) difference between actual and predicted values. The MAP method per-

formed well (error $< 10\%$) at high SNR (>100), but resulted in severe estimation errors at low SNR (<50). The EXP model results in significant measurement differences ($P < 0.05$) compared with all other methods, irrespective of SNR (range: 8–233).

Figure 3 shows the estimated T2* values for the numerical simulation with T2* = 20 ms and SNR ranges from 8 to 233. The figure shows the average percentage error in T2* estimation for each method. All methods showed reducing estimation error with increasing SNR. The EXP model resulted in overestimation of T2* regardless of SNR. The MAP methods results in severe errors at low SNR and then approached the correct estimation with improved SNR. The BEXP and CEXP models as well as the AVG and MED methods showed very similar behaviors. Any combination of these two models and two methods resulted in small estimation errors at low SNR and accurate estimates with improved SNR.

Figure 4 shows the change in percentage estimation error with actual T2* value for all different signal calculation methods and exponential fitting models. The EXP model showed identical estimation errors for all signal calculation methods, although the error decreased for higher T2*
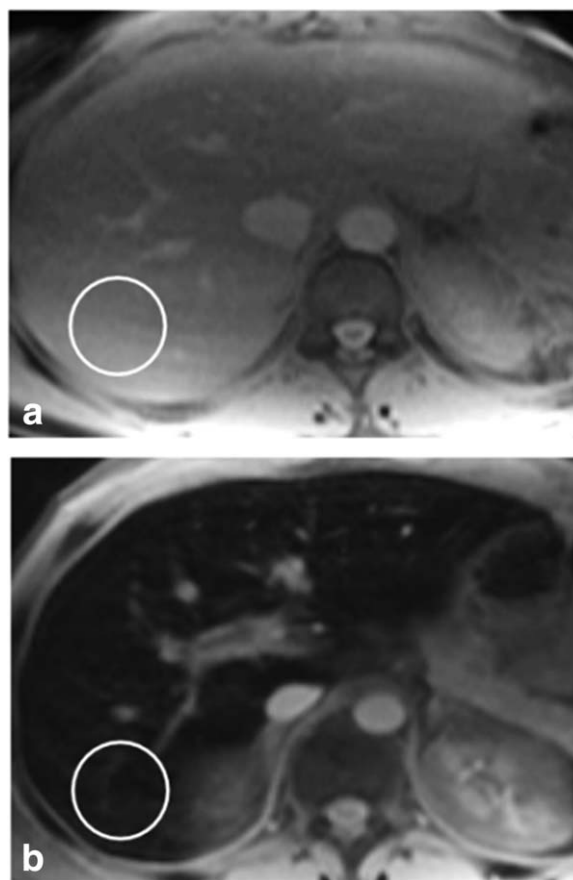

FIGURE 2: T2* weighted images for mild (a) and severe (b) iron overload patients. The white circles indicate the location and size of the ROIs used for analysis.
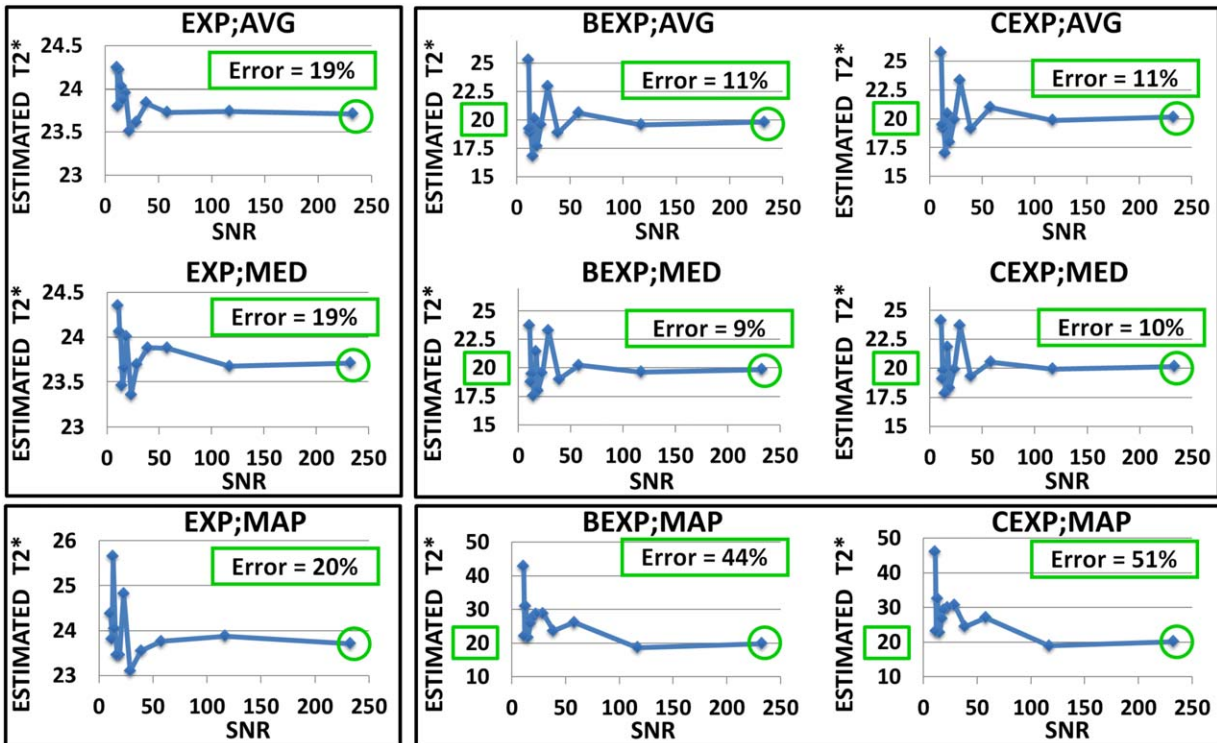
FIGURE 3: Numerical simulation of the SNR effects on T2* calculation based on the adopted exponential model and signal calculation method. T2* of 20 ms is chosen in this example as it considered the cutoff between normal and overloaded iron content. Curves with the same y-axis scaling are grouped in a black frame for easy visual comparison. The green circle shows the final T2* that each method converges to. The green/red mark on the y-axis shows the value of that convergence. The average percentage error (with respect to the true T2* value) across different SNRs is marked for each method.

values. In all cases, the AVG and MED methods showed identical errors at each T2*, which was always less than that for the EXP model. The errors from the BEXP and CEXP models were very similar at each T2*. The MAP method always resulted in larger error than the AVG or MED methods.

Phantom Results
In Figure 5, the estimated R2* values is compared with iron concentration in each phantom tube. The best linear relationship (max R^2) was provided by the BEXP and CEXP models, which showed very similar behaviors. The AVG and MED methods provided the best linear relationships with slightly higher R^2 with the AVG method.

Human Subjects Results
Figure 6 and Table 1 show the estimated T2* value for the nine patients calculated using different exponential models and signal calculation methods. The T2* values ranged from 3.1 to 53.6 ms, depending on the amount of iron overload and implemented analysis method. The BEXP (range: 3.7–50 ms) and CEXP (range: 3.8–53.6 ms) models, and the AVG (range: 3.2–38.8 ms) and MED (range: 3.1–38.5 ms) methods provided more accurate measurements than the EXP model (range: 3.1–18.3 ms) and MAP (range: 3.8–53.6 ms) method, respectively ($P < 0.05$).

The results in Figure 6 confirm the numerical simulations. The EXP model (methods #1,4,7) resulted in significant measurement differences ($P < 0.05$) compared with other methods. For example, methods #1 and #4 showed similar and very small T2* values (all values < 10 ms) for all patients, which contradicts the fact that these patients have different disease stages as confirmed by the varying signal intensity of their T2*-weighted images as shown in Figure 2.

Figure 6 shows close measurements between the BEXP and CEXP models (methods #2 and #3 provided very similar estimations [$P = 0.91$]; same for methods #5 and #6 [$P = 0.91$]; and #8 and #9 [$P = 0.87$]). The AVG and MED methods provided very similar results (methods #1 and #4 [$P = 0.97$]; #2 and #5 [$P = 0.99$]; #3 and #6 [$P = 0.99$]). The MAP method (methods #7,8,9) always showed different results than those obtained with either the AVG or MED methods (MAP showed a trend for overestimating the results). The results from the BEXP and CEXP models combined with either AVG or MED methods showed varying T2* for different patients, which is consistent with the varying liver signal intensity in the T2*-weighted images of these images.

Discussion
The analysis of the MRI images for measuring T2*, as an estimate of iron overload, is affected by several factors. First,

EXP

a

BEXP
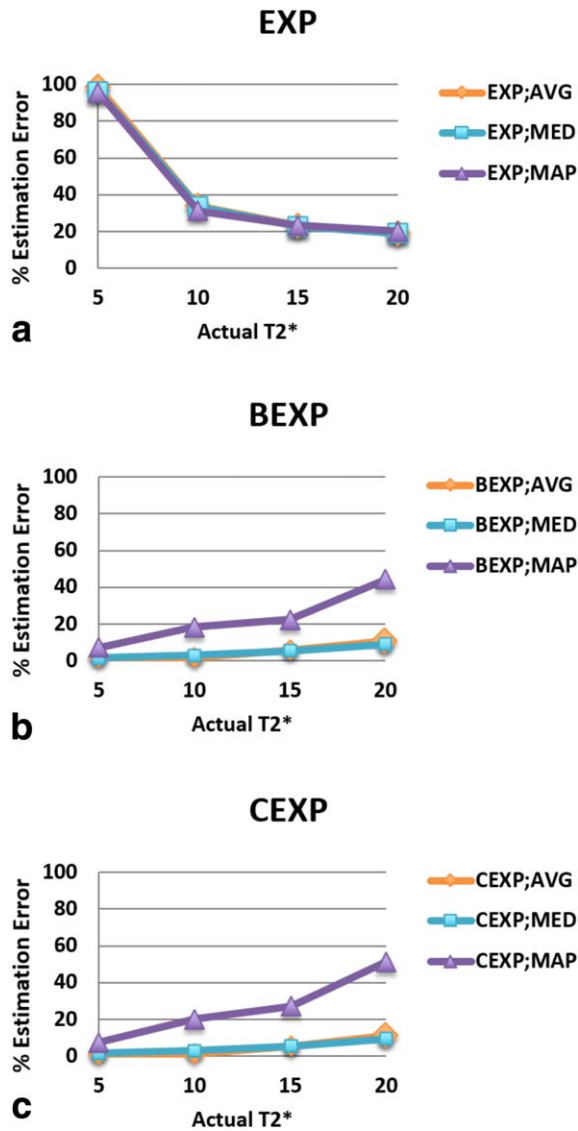

b

CEXP

c

FIGURE 4: a–c: Numerical simulation of the effects of the exponential model and signal calculation method on T2* calculation for different T2* values (5, 10, 15, and 20 ms). The x-axis shows the actual T2* value and the y-axis shows the percentage error in the estimated value. Note that the red and blue curves show almost identical behaviors (appear on top of each other).

the user's selection of the ROI affects T2* measurement due to heterogeneous iron distribution in the liver, susceptibility artifacts, and/or inclusion of vasculature.[7,15] Furthermore, T2* could be calculated pixel-wise or inside a defined ROI, with the latter method preferred to avoid heterogeneous results between adjacent pixels due the three previously mentioned reasons. Finally, different exponential fitting models have been used in the literature with different degrees of success[6,7,16]. Although the BEXP model has been adopted in human liver studies[7,16,17] and the EXP model has been successfully implemented in several studies,[6,11,18] the CEXP model[7,8,19] has been regarded as a good approximation of the BEXP model, where the slowly varying term in the BEXP model is replaced by a constant value.

Although mono-exponential fitting with echo truncation is a viable data analysis approach, especially in cases where SNR is low and later echoes harm, rather than benefit, the exponential fitting, we did not include this approach in our analysis, as measurements from all echoes showed sufficient SNR, secondary to the use of 3T imaging in this study.

In this study, we examined the effects of the data fitting model and signal calculation method on T2* estimation. The results from human subjects were in agreement with the numerical simulation and phantom results. Basically, the BEXP and CEXP models and the AVG and MED methods provided more accurate results than the EXP model and MAP method, respectively. The results from the BEXP and CEXP models were very similar. The close results between these two models could be explained by noting that the CEXP model can be considered as an approximation of the BEXP model, where the tissue with large T2* (usually representing vasculature) is replaced by a constant. The superiority of the BEXP model to the EXP model could be explained by the addition of a constant term in BEXP, which reflects more realistic approximation of signal relaxation in the presence of background noise or the existence of a second tissue in the analyzed voxel with large T2* value (partial volume effect, especially in low-resolution images).

The results from the AVG and MED methods were very similar, with slightly better performance of the AVG method. This may be attributed in part to the fact that the AVG method uses the average value within the ROI to fit the exponential curve, which results in improved SNR. Basically, in the absence of signal heterogeneity in the selected ROI, both methods result in similar results. The results become slightly different when there is signal heterogeneity inside the ROI, e.g., vasculature, artifact, or boundary effects. The MAP method performed well at high SNR, but resulted in severe estimation errors at low SNR (as in the case of severe iron overload with very low signal close to the noise level). The EXP model resulted in the worst estimations irrespective of the implemented signal calculation method or SNR.

In this study, a single ROI inside the liver has been used for analysis; rather than measuring T2* map inside the whole liver. Meloni et al[15] have evaluated the effectiveness of using a single ROI for measuring hepatic T2* compared with T2* measurement inside the whole liver, where the results showed that the single ROI approach provided reliable values for evaluating hepatic iron overload in six out of the eight segments, provided that the ROIs are selected away from susceptibility artifacts and vasculature. This was the approach we followed in our analysis.

This study showed good quality images based on T2* imaging at 3T magnetic field strength. While T2* imaging
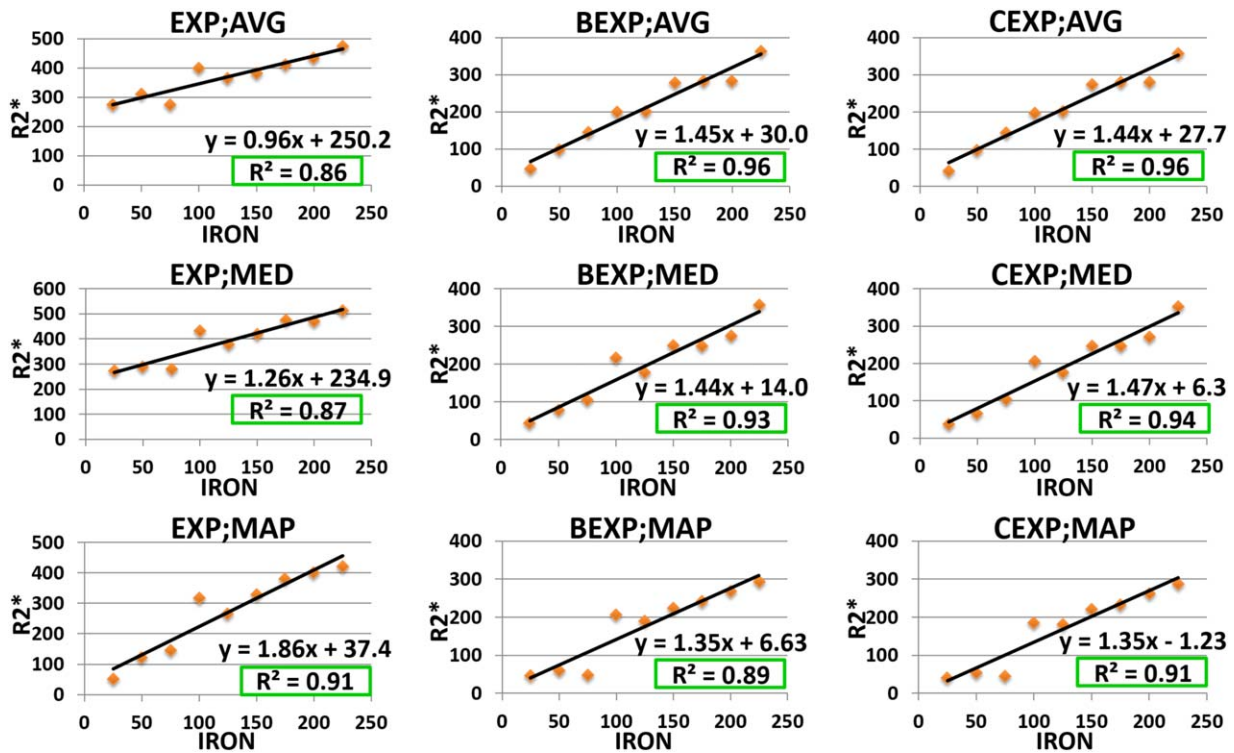
FIGURE 5: Relationship between iron concentration and R2* in the phantoms based on the adopted exponential model and signal calculation method. The regression equation and R^2 are shown for each method.

has been established at 1.5T with cutoff T2* values in the literature for determining iron overload and severe iron overload,[20] the results from the current study suggest that iron overload assessment at 3T has a potential for future applications. Despite the increased technical challenges for T2* imaging at 3T, e.g., the need for excellent field shimming to avoid severe off-resonance artifacts, high field strength results in improved SNR, which may lead to even earlier detection of small iron concentrations compared with imaging at 1.5T. Nevertheless, our study cannot affirm the superiority of iron imaging at 3T compared with 1.5T; a point that needs to be further investigated in a larger study.

One limitation of our study is the lack of reference-standard biopsy samples from the patients, as this was not part of standard-of-care procedures. However, the performances of the different analysis techniques on the data from the human subjects were very similar with those in the numerical simulations. Furthermore, we used the calibrated iron phantoms as our ground truth for comparison against the level of iron concentration, which provides an excellent means for comparing the different analysis techniques, as the amount of iron content is known precisely, which may not be the case in biopsy due to sampling errors, small number of subjects, and dependency on disease model.

It should be noted here that although lower signal intensity of the liver is associated with increased T2* effect due to increased iron content, the lack of biopsy data precludes us from affirming the amount of iron or its variability among the studied subjects. Another limitation of the current study is the lack of reproducibility measurements due to the small number of studied subjects. The encouraging results in this study require conducting further research on a larger number of patients with wider ranges of iron overload with biopsy results to confirm these results, assess
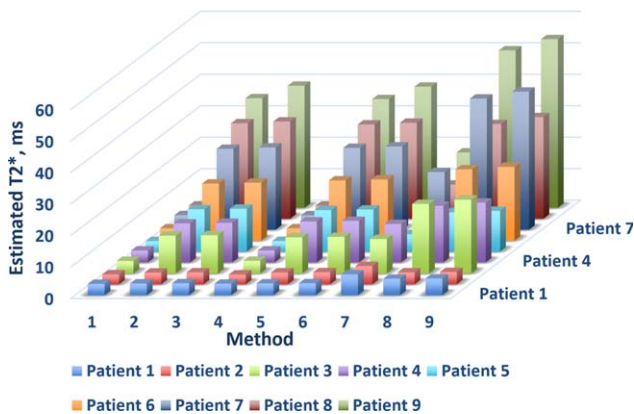

FIGURE 6: T2* calculated for different patients using different exponential models and signal calculation methods. The fitting model and signal calculation method for each method number in the x-axis are:1, AVG-EXP; 2, AVG-BEXP; 3, AVG-CEXP; 4, MED-EXP; 5, MED-BEXP; 6, MED-CEXP; 7, MAP-EXP; 8, MAP-BEXP; 9, MAP-CEXP. The y-axis shows different patients represented by different color codes. The z-axis shows the calculated T2* value.

TABLE 1. T2* Measurements (ms) in the Patients Based on Signal Calculation Method and Fitting Model

	AVG			MED			MAP		
	EXP	BEXP	CEXP	EXP	BEXP	CEXP	EXP	BEXP	CEXP
Method	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**
Patient 1	3.6	3.7	3.8	3.7	3.8	3.8	6.6	5.2	5.3
Patient 2	3.2	3.7	3.8	3.1	3.8	3.8	5.8	3.8	3.9
Patient 3	4.0	12.0	12.1	4.0	11.4	11.6	10.9	22.1	23.4
Patient 4	3.8	12.4	12.6	3.8	13.0	13.2	12.2	18.0	19.0
Patient 5	3.4	13.4	13.6	3.4	13.2	13.3	5.5	12.5	12.9
Patient 6	3.9	18.1	18.4	3.9	19.1	19.4	9.9	22.6	23.4
Patient 7	4.6	25.7	26.1	4.6	26.0	26.4	18.3	41.7	43.8
Patient 8	4.2	30.3	30.9	4.2	30.0	30.5	10.9	30.1	32.3
Patient 9	4.3	34.8	38.8	4.4	34.5	38.5	17.6	50.0	53.6

AVG = average; BEXP = bi-exponential; CEXP = exponential-plus-constant; EXP = single-exponential; MAP = pixel-wise mapping; MED = median.

measurements reproducibility of each technique, and determine the optimal technique for evaluating iron overload in different diseases and patient groups. Another point that is worth mentioning is the possible effect of hepatic steatosis on T2* measurements. Although we did not study this effect in the current study, the inclusion of fat quantification techniques, e.g., mDixon imaging,[12] is planned for future work to fully investigate this effect.

In conclusion, T2* measurement with MRI is a promising technique for evaluating iron overload in the liver. However, the adopted model fitting model and signal calculation method should be taken in consideration as they affect the results. The BEXP and CEXP models provide very close results that are more accurate than the EXP model. The AVG and MED methods provide similar results, and their estimation errors are much less than the MAP method, especially at low SNR. Therefore, we recommend the CEXP model due to its simpler implementation than the BEXP model. The AVG method is also recommended as it is less affected by the ROI selection.

References

1. Porter JB. Concepts and goals in the management of transfusional iron overload. Am J Hematol 2007;82:1136–1139.

2. Wood JC. Magnetic resonance imaging measurement of iron overload. Curr Opin Hematol 2007;14:183–190.

3. Wacker P, Halperin DS, Balmer-Ruedin D, Oberhansli I, Wyss M. Regression of cardiac insufficiency after ambulatory intravenous deferoxamine in thalassemia major. Chest 1993;103:1276–1278.

4. Kolnagou A, Economides C, Eracleous E, Kontoghiorghes GJ. Low serum ferritin levels are misleading for detecting cardiac iron overload and increase the risk of cardiomyopathy in thalassemia patients. The importance of cardiac iron overload monitoring using magnetic resonance imaging T2 and T2*. Hemoglobin 2006;30:219–227.

5. Bondestam S, Lamminen A, Anttila VJ, Ruutu T, Ruutu P. Magnetic resonance imaging of transfusional hepatic iron overload. Br J Radiol 1994;67:339–341.

6. Anderson LJ, Holden S, Davis B, et al. Cardiovascular T2-star (T2*) magnetic resonance for the early diagnosis of myocardial iron overload. Eur Heart J 2001;22:2171–2179.

7. Positano V, Salani B, Pepe A, et al. Improved T2* assessment in liver iron overload by magnetic resonance imaging. Magn Reson Imaging 2009;27:188–197.

8. Wood JC, Enriquez C, Ghugre N, et al. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. Blood 2005; 106:1460–1465.

9. Di Tucci AA, Matta G, Deplano S, et al. Myocardial iron overload assessment by T2* magnetic resonance imaging in adult transfusion dependent patients with acquired anemias. Haematologica 2008;93: 1385–1388.

10. el Ibrahim SH, Rana FN, Johnson KR, White RD. Assessment of cardiac iron deposition in sickle cell disease using 3.0 Tesla cardiovascular magnetic resonance. Hemoglobin 2012;36:343–361.

11. Westwood M, Anderson LJ, Firmin DN, et al. A single breath-hold multiecho T2* cardiovascular magnetic resonance technique for diagnosis of myocardial iron overload. J Magn Reson Imaging 2003;18: 33–39.

12. Gandon Y, Olivie D, Guyader D, et al. Non-invasive assessment of hepatic iron stores by MRI. Lancet 2004;363:357–362.

13. Wood JC, Ghugre N. Magnetic resonance imaging assessment of excess iron in thalassemia, sickle cell disease and other iron overload diseases. Hemoglobin 2008;32:85–96.

14. Ghugre NR, Enriquez CM, Coates TD, Nelson MD Jr, Wood JC. Improved R2* measurements in myocardial iron overload. J Magn Reson Imaging 2006;23:9–16.

15. Meloni A, Luciani A, Positano V, et al. Single region of interest versus multislice T2* MRI approach for the quantification of hepatic iron overload. J Magn Reson Imaging 2011;33:348–355.

16. 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;21:519–530.

17. Positano V, Salani B, Scattini B, et al. A robust method for assessment of iron overload in liver by magnetic resonance imaging. Conf Proc IEEE Eng Med Biol Soc 2007;2007:2895–2898.

18. Pepe A, Lombardi M, Positano V, et al. Evaluation of the efficacy of oral deferiprone in beta-thalassemia major by multislice multiecho T2*. Eur J Haematol 2006;76:183–192.

19. Wood JC, Otto-Duessel M, Aguilar M, et al. Cardiac iron determines cardiac T2*, T2, and T1 in the gerbil model of iron cardiomyopathy. Circulation 2005;112:535–543.

20. Pennell DJ. T2* magnetic resonance and myocardial iron in thalassemia. Ann N Y Acad Sci 2005;1054:373–378.