The Influence of the Analysis Technique on Estimating Liver Iron Overload Using Magnetic Resonance Imaging T2* Quantification

El-Sayed H. Ibrahim, Ayman M. Khalifa, Ahmed K. Eldaly

Abstract— **Iron toxicity is the major cause of tissue damage in patients with iron overload. Iron deposits mainly in the liver, where its concentration closely correlates with whole body iron overload. Different techniques have been proposed for estimating iron content, with liver biopsy being the gold standard despite its invasiveness and influence by sampling error. Recently, magnetic resonance imaging (MRI) has been established as an effective technique for evaluating iron overload by measuring T2* in the liver. However, various factors associated with the adopted analysis technique, mainly the exponential fitting model and signal averaging method, affect the resulting measurements. In this study, we evaluate the influences of these factors on T2* measurement in numerical phantom, calibrated phantoms, and nine patients with different degrees of iron overload. The results show different performances among the fitting models and signal averaging methods, which are affected by SNR, image quality and signal homogeneity inside the selected ROI for analysis.**

I. INTRODUCTION

Iron toxicity is the main reason for tissue damage and organ failure in patients with iron overload. 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 [1].

Accurate evaluation of iron overload is critical for initialization of chelation therapy to minimize, and even reverse, its effects. 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 non-specific and is affected by inflammation and infection [2].

Recently, magnetic resonance imaging (MRI) has been considered for evaluating iron overload based on T2* measurement [3-6]. By acquiring a number of T2*-weighted images at different echo times (TE's), the signal intensity of the tissue of interest could be plotted against TE, from which T2* (measured in ms), and its inverse R2* = 1000/T2* (measured in 1/s), can be calculated using curve fitting. Iron overload has been shown to result in abnormally reduced T2* values that are not observed in other clinical conditions. At 1.5T, T2* of 20 is commonly used as the cutoff for identifying iron overload [7]. However, the reduced signal-to-noise ratio (SNR) at 1.5T is a limiting factor for reliable T2* quantification. Therefore, T2*-weighted imaging at 3.0T has been suggested for quantifying iron overload, especially when tissue iron levels are low or high image resolution is required [8]. A linear relationship has been shown between R2* and iron content over the entire clinical range of interest; however, there was no significant correlation between R2* and serum ferritin [8].

T2* is estimated by fitting the signal intensities acquired at multiple TE's to an exponentially decaying curve. Although the analysis process is conceptually straightforward, there are a number of 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.

The user's selection of the ROI for analysis affects T2* measurement due to heterogeneous iron distribution in the liver, susceptibility artifacts, and inclusion of vasculature, which results in partial volume effect that affects T2* measurement of the liver parenchyma [5,9]. T2* could be calculated pixel-wise (relaxivity map) or inside a defined ROI, with the latter method preferred to avoid heterogeneous results between adjacent pixels due the three previously mentioned reasons. The ROI-based techniques could be classified into three categories: pixel-wise (MAP), where exponential fitting is applied to each pixel inside the ROI, followed by obtaining the mean of the resulting T2* values; average (AVG), where the average signal intensity inside the ROI is first calculated, followed by exponential fitting to measure T2*; and median (MED), where the median signal intensity inside the ROI is first calculated, followed by exponential fitting to measure T2*.

The choice of the exponential fitting model is critical for T2* assessment. Although the bi-exponential signal decaying model (BEXP, equation 1) was adopted in human liver studies [5,10], the single exponential decaying model (EXP, equation 2) was successfully implemented [11]. The exponential-plus-constant model (CEXP, equation 3) [5] is an established criterion for approximating the BEXP model, where the slowly varying term in the BEXP model is replaced by a constant value, *C*. In the following equations, S and S_0 are the signal intensity and its initial value at TE = 0, respectively.

$$\underline{\text{BEXP}}\text{:}\ S = 0.9\, S_o e^{-\frac{TE}{T_2^*}} + 0.1\, S_o e^{-\frac{TE}{200}} \qquad [1]$$

$$\underline{\text{EXP}}\text{:}\quad S = S_o e^{-\frac{TE}{T_2^*}} \qquad [2]$$

$$\underline{\text{CEXP}}\text{:}\ S = S_o e^{-\frac{TE}{T_2^*}} + C \qquad [3]$$

Different exponential fitting models have been used in the literature with different degrees of success. Typically, a large ROI of homogeneous signal intensity is selected inside the liver away from vasculature, in which the mean signal intensity is measured and plotted against TE. T2* is then

E. H. Ibrahim (corresponding author) is with Mayo Clinic, Jacksonville, FL 32224, USA. Phone: 904-953-6037; Fax: 904-943-6581 (email: ibrahim.elsayed@mayo.edu).

A. M. Khalifa is with Helwan University, Cairo, Egypt (e-mail: aymankhalifa2002@yahoo.com).

A. K. Eldaly is with Helwan University, Cairo, Egypt (e-mail: ahmed.karam@h-eng.helwan.edu.eg).

measured by fitting the results to a EXP [4,11,12], BEXP [13], or CEXP model [6,14]. Other studies used more than one ROI for measuring T2* [5,9,15,16]. For example, Maris et al [15] used two ROI's extracted from two orthogonal slices in the liver, where the mean signal intensity inside each ROI was calculated at different TE's, and then the CEXP model was implemented on each dataset to measure T2*. The final T2* was calculated as the average of the resulting two T2* values. Melony et al [9] evaluated the effectiveness of using a single ROI for measuring hepatic T2* compared to T2* measurement inside the whole liver. The authors divided the liver into eight functionally independent segments and measured T2* in a ROI inside each segment, 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 ROI's are selected away from susceptibility artifacts and vasculature. Finally, Positano et al [5] used a method for measuring T2* inside the whole liver after removing vasculature using a semi-automated fuzzy clustering technique.

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

II. METHODS

A. Numerical Phantom

Numerical phantoms were generated with different T2* values. Different levels of noise, with variance ranging from 1 to 22, were added to the phantom, resulting in signal-to-noise ratio (SNR) ranging from 8 to 233 at first timeframe. The three exponential models (EXP, BEXP, and CEXP) and three averaging methods (AVG, MED, and MAP) were used to measure T2*. Root mean square errors were calculated between the actual and predicted T2* values and summed for all SNR levels to determine the analysis technique with minimum estimation error.

B. Phantom Scans

Calibrated phantoms with pre-determined amounts of iron were created (Figure 1) to compare the T2* values calculated by different analysis techniques to ground truth. Ten 50-mL vials were filled with agarose-based material doped with $MnCl_2$ to create a material with T1 and T2 values similar to those in the liver, as previously described [8]. Different amounts of iron sulphate were added to nine vials (the tenth vial was used as a reference), ranging from 0.5 to 4.5 g in equal increments, which resemble the range of iron content in the clinical setting. The phantoms were imaged on a 3.0T Siemens MRI scanner using a 12-echo GRE sequence with TE's ranging from 1 to 16.5 ms.

C. Human Objects Scans

Nine human subjects with different degrees of iron overload were imaged on a 3.0T Siemens MRI scanner, similar to the phantom experiments, where 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*.

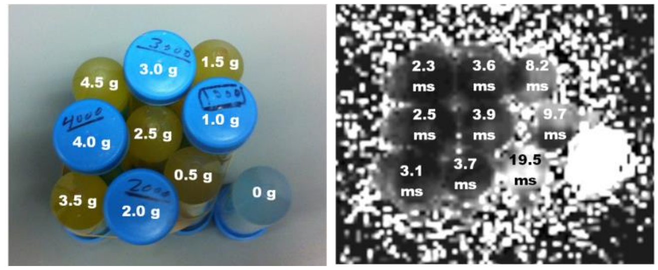


Figure 1. Calibrated agarose-based calibrated iron phantom (left) and T2* map (right). The amount of iron content (in gm) is marked on each tube. The T2* values measured on the scanner console are marked for all tube.

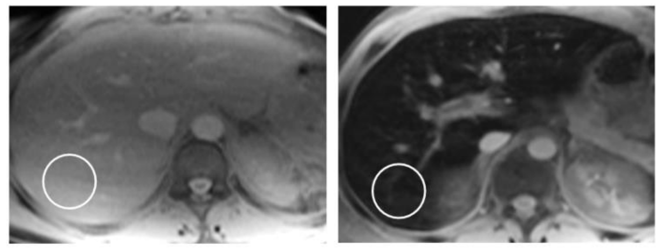


Figure 2. T2* weighted images for mild (left) and severe iron overload (right) patients. The white circles indicate the location and size of the ROI's used for analysis .

D. 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 averaging method (MAP, AVG, and MED); exponential fitting model (EXP, BEXP, and CEXP) using the Levenberg–Marquadt algorithm. A 2-cm^2 circular region of interest (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.

III. RESULTS

A. Numerical Phantom Results

Figure 3 shows the estimated T2* values for the numerical phantom 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 deviation in T2* estimation 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 and 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 averaging methods and exponential fitting models. The EXP model showed identical estimation errors for all signal averaging methods, although the error decreased for higher T2* 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.

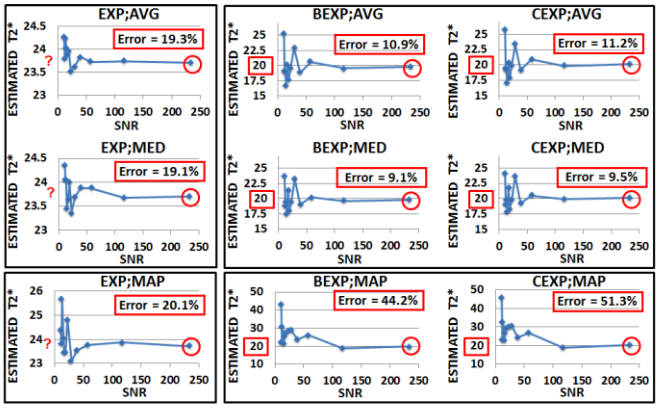


Figure 3. Numerical simulation of the SNR effects on T2* calculation based on the adopted exponential model and signal averaging method. T2* of 20 ms is chosen in this example as it considered the cutoff between normal and overloaded iron content. Curves with the same y-axis scaling are grouped in a black frame for easy visual comparison. The red circle shows the final T2* that each method converges to. The red mark on the y-axis shows the value of that convergence. The average percentage error across all SNR's is marked for each method.

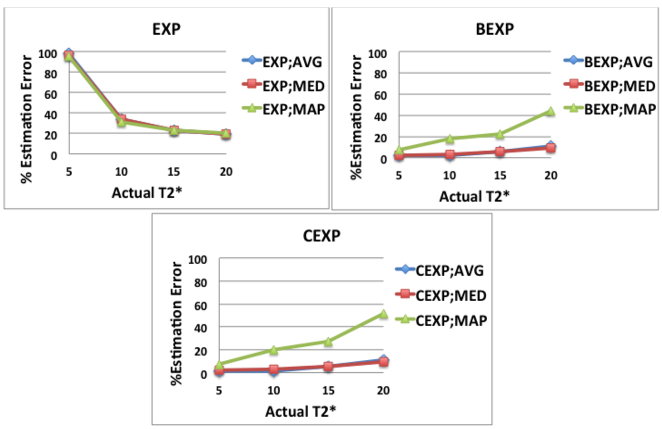


Figure 4. Numerical simulation of the effects of the exponential model and averaging method on T2* calculation for different T2* values (5, 10, 15, and 20 ms). The x-axis shows the actual T2* value and the y-axis shows the percentage error in the estimated value. Note that the red and blue curves show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

B. Phantom Scans Results

In figure 5, the estimated R2* values is compared with iron concentration in each phantom tube. The best linear relationships (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.

C. Human Subjects Results

Figure 6 shows the estimated T2* value for the 9 patients calculated using different exponential models and signal averaging methods. The results in the figure confirm the numerical simulations. The EXP model (methods # 1, 4, 7) resulted in the worst estimations. For example, methods #1 and #4 showed similar and very small T2* values (all values less than 10) 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 results between the BEXP and CEXP models (methods #2 and #3 provided very similar estimations; same for methods #5 and #6; and #8 and #9). The AVG and MED methods provided very similar results (methods #1 and #4; #2 and #5; and #3 and #6). 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.

IV. Discussion

T2* measurement with MRI has been established as an effective technique for non-invasive assessment of iron overload in the liver [5,6,9]. However, the image processing

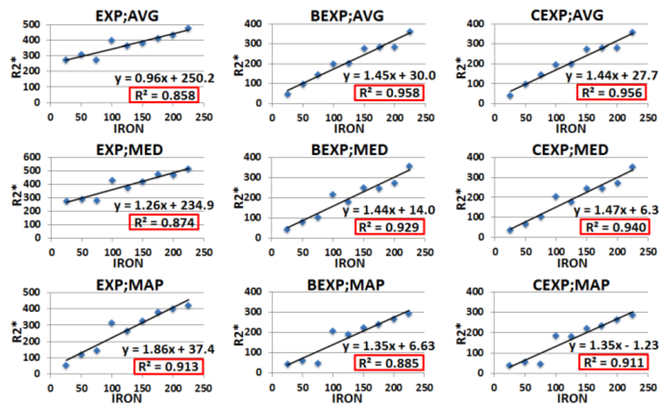
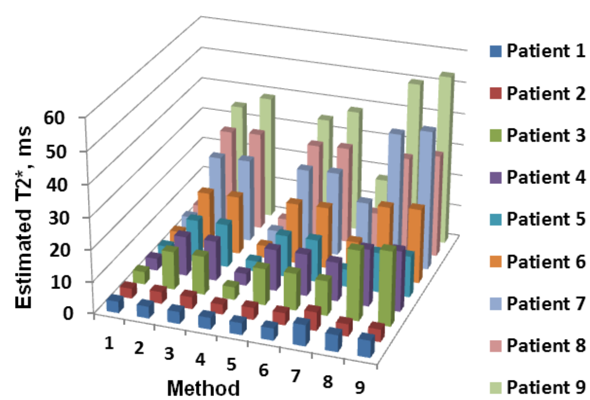


Figure 5. Relationship between iron concentration and R2* in the phantoms based on the adopted exponential model and signal averaging method. The regression equation and R^2 are shown for each method.



Method	EXP	BEXP	CEXP
AVG	1	2	3
MED	4	5	6
MAP	7	8	9

Figure 6. T2* calculated for different patients using different exponential models and signal averaging methods. The table in the bottom explains the fitting model and signal averaging method for each method number in the y-axis. The x-axis shows different patients represented by different color codes. The z-axis shows the calculated T2* value.

criterion adopted for T2* calculation plays a crucial role in determining the resulting value. Basically, T2* could be affected by the imaging sequence, parameter settings, the exponential fitting model, and the signal averaging method.

In this study, we examined the effects of the data fitting model and signal averaging method on T2* estimation. The results from human subjects were in agreement with the 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 BEXP and CEXP models were always 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 results from the AVG and MED methods were very similar, with slightly better performance of the AVG method. Basically, in the absence of signal heterogeneity in the selected ROI, both methods result in similar results. The results become slightly different when there's 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 averaging method or SNR.

One limitation of our study is the lack of 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. Further, 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.

The encouraging results in this study require conducting further research on a larger number of patients with wider ranges of iron overload and known pathology to confirm these results and to determine the optimal technique for evaluating iron overload in different diseases and patient groups. Further, advanced pulse sequences, e.g. ultra-short echo time (UTE) sequences, which allow for achieving TE on the order of micro seconds and thus measuring very short T2* accurately, should be investigated for quantifying iron content in patients with heavy iron overload, which is a challenging task using standard imaging sequences.

V. Conclusion

T2* measurement with MRI is a promising technique for evaluating iron overload in the liver. However, the adopted model fitting model and signal averaging method should be taken in consideration as they affect the results, and their performances vary based on a number of factors including the image quality, SNR, and selected ROI. 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. Future studies include scanning a larger number of patients with ground truth biopsy results, and studying the effects of image SNR and fat liver on iron quantification.

Acknowledgment

This work is fully supported by a research and development grant (PDP2012.R12.4) from ITIDA agency, Ministry of Communication and Information Tech., Egypt.

References

[1] J. C. Wood, “Magnetic resonance imaging measurement of iron overload,” Curr. Opin. Hematol., Vol. 14, no. 3, pp. 183-190, May 2007.

[2] A. Kolnagou , C. Economides , E. Eracleous , G. J. Kontoghiorghes, “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, Vol. 30, pp. 219-227, 2006.

[3] Bondestam S, Lamminen A, Anttila VJ, Ruutu T, Ruutu P. Magnetic resonance imaging of transfusional hepatic iron overload. Br. J. Radiol,Vol. 67, no. 769, pp. 339-341, 1994.

[4] 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. , Vol. 22, no. 23, pp. 2171-2179, 2001.

[5] V. Positano, B. Salani, A. Pepe, M. F. Santarelli, D. De Marchi, A. Ramazzotti, B. Favilli, E. Cracolici, M. Midiri, P. Cianciulli, M. Lombardi, L. Landini. “Improved T2* assessment in liver iron overload by magnetic resonance imaging,” Mag. Reson.Imag., Vol. 27, pp. 188–197 , 2009.

[6] J. C. Wood, C. Enriquez, N. Ghugre, J. M. Tyzka, S. Carson, M. D. Nelson, and T. D. Coates, “MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients,” Blood, Vol.106, no. 4, pp. 1460-1465, 2005 Aug 15.

[7] A. A. Di Tucci, G. Matta, S. Deplano, A. Gabbas, C. Depau, D. Derudas, G. Caocci, A. Agus, and E. Angelucci, “Myocardial iron overload assessment by T2* magnetic resonance imaging in adult transfusion dependent patients with acquired anemias,” haematologica, Vol. 93, no. 9, pp. 1385-1388, 2008

[8] E. H. Ibrahim, F. N. Rana, K. R. Johnson, and R. D. White, “Assessment of cardiac iron deposition in sickle cell disease using 3.0 Tesla cardiovascular magnetic resonance,” Hemoglobin, Vol. 36, pp.343-361, 20.

[9] A. Meloni, A. Luciani, V. Positano, D. De Marchi, G. Valeri, G. Restaino, E. Cracolici, V. Caruso, M. C. Dell’Amico, B. Favilli, M. Lombardi, and A. Pepe, “Single Region of Interest Versus Multislice T2* MRI Approach for the Quantification of Hepatic Iron Overload,” J. Mag. Reson. Imag., Vol. 33, pp. 348–355, 2011.

[10] V. Positano, B. Salani, B. Scattini, M. F. Santarelli, A. Ramazzotti, A. Pepe, M. Lombardi, and L. Landini, “A Robust Method for Assessment of Iron Overload in Liver by Magnetic Resonance Imaging,” Eng. in Med. and Biol. Conf. , France , pp. 2895-2898, 2007.

[11] A. Pepe, M. Lombardi, V Positano, et al., “Evaluation of the efficacy of oral deferiprone in beta-thalassemia major by multislice multiecho T2*,” *Eur J Haematol,* vol. 76, pp. 183-192, 2006.

[12] M. Westwood, L. J. Anderson, D. N. Firmin, P. D. Gatehouse, C. C. Charrier, B. Wonke, D. J. Pennell, “A single breath-hold multiecho T2* cardiovascular magnetic resonance technique for diagnosis of myocardial iron overload,” J. Magn. Reson. Imaging, Vol. 18, no. 1, pp. 33-39, 2003.

[13] P. R. Clark, W. Chua-anusorn , T. G. St Pierre, “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, Vol. 21, no. 5, pp. 519-530, Jun 2003.

[14] J. C. Wood, M. Otto-Duessel, M. Aguilar, H. Nick, M. D. Nelson, T. D. Coates, H. Pollack, R. Moats, “Cardiac iron determines cardiac T2*, T2, and T1 in the gerbil model of iron cardiomyopathy,” Circulation, Vol. 112, no. 4, pp. 535-543, Jul. 2005.

[15] T. G. Maris, O. Papakonstantinou, V. Chatzimanoli, A. Papadakis, K. Pagonidis, N. Papanikolaou, A. Karantanas, N. Gourtsoyiannis, “Myocardial and liver iron status using a fast T*2 quantitative MRI (T*2qMRI) technique,” Magn. Reson. Med., Vol. 57, pp. 742-753, 2007.

[16] E. H. Ibrahim, A. M. Khalifa, A. K. Eldaly, and A. W. Bowman “The effects of the analysis technique on hepatic iron evaluation using T2* mapping with magnetic resonance imaging,” MECBME, Qatar, Feb. 2014.