Accurate liver T*2 measurement of iron overload: A simulations investigation and in vivo study

Marine Beaumont, Isaac Odame, Paul S. Babyn, Logi Vidarsson, Melanie Kirby-Allen, and Hai-Ling Margaret Cheng

Version Post-print/accepted manuscript


Publisher's Statement This is the peer reviewed version of the following article: [Beaumont M, Odame I, Babyn PS, Vidarsson L, Kirby-Allen M, Cheng HL. Accurate liver T measurement of iron overload: A simulations investigation and in vivo study. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2009 Aug;30(2):313-20.], which has been published in final form at [10.1002/jmri.21835]. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

How to cite TSpace items

Always cite the published version, so the author(s) will receive recognition through services that track citation counts, e.g. Scopus. If you need to cite the page number of the author manuscript from TSpace because you cannot access the published version, then cite the TSpace version in addition to the published version using the permanent URI (handle) found on the record page.

This article was made openly accessible by U of T Faculty. Please tell us how this access benefits you. Your story matters.
Accurate Liver T$_2^*$ Measurement of Iron Overload: a Simulations Investigation and In-vivo Study

Marine Beaumont, PhD$^{1,2}$, Isaac Odame, MD$^3$, Paul S Babyn, MD$^2$, Logi Vidarsson, PhD$^2$, Melanie Kirby-Allen, MD$^3$, Hai-Ling Margaret Cheng, PhD$^{1,2,4}$

$^1$The Research Institute, The Hospital for Sick Children
$^2$Diagnostic Imaging, The Hospital for Sick Children
$^3$Haematology/Oncology, The Hospital for Sick Children
$^4$Department of Medical Biophysics, University of Toronto
Toronto, Canada

Corresponding author:
Hai-Ling Margaret Cheng, Ph.D.
The Hospital for Sick Children
555 University Avenue, Toronto, Ontario M5G 1X8 Canada
Phone: 416-813-5415
Fax: 416-813-7362
hai-ling.cheng@sickkids.ca

Word count: 4547

Running title: Accuracy of T$_2^*$ Measurement of Liver Iron
ABSTRACT

Purpose: To investigate the accuracy of T$_2^*$ liver iron quantification using different curve-fitting models under varying acquisition conditions, and to compare in iron-overloaded patients the reliability of rapid T$_2^*$ measurements against approved and slower T$_2$ protocols.

Materials and Methods: Simulations were conducted to assess the influence of various factors on the accuracy of T$_2^*$ measurement: curve-fitting model, signal-to-noise ratio (SNR), and echo time (TE) spacing. Fifty-four iron-overloaded pediatric patients were assessed using a standard T$_2$ and two variations of T$_2^*$ acquisitions. In both simulations and in-vivo data, three analysis models were evaluated: monoexponential, constant offset, and truncated.

Results: Simulations show the truncated model provides the best accuracy but is susceptible to underestimating high iron species under low SNR or high minimum TE. In contrast, the offset model tends to overestimate but maintains the most reliable measurements across the relevant range of iron levels. Furthermore, a much lower SNR can be tolerated if the acquisition employs a low minimum TE. In-vivo results confirm theoretical findings and show that T$_2^*$ measurements can be as reliable as those from approved and slower T$_2$ protocols.

Conclusion: Guidelines are provided on choosing an appropriate model under specific noise conditions and acquisition schemes to ensure accurate and rapid T$_2^*$ liver iron quantification.

Key words: T$_2^*$ relaxation; liver iron overload; noise; curve-fitting models
INTRODUCTION

Pathologic iron overload is a common occurrence in children with certain genetic disorders, including hereditary hemochromatosis and anemias. The majority of cases are associated with frequent blood transfusions to treat anemic patients, specifically those with thalassaemia major and sickle cell disease (1,2). Without aggressive iron chelation therapy, these patients die from endocrine and cardiac dysfunction in the second decade of life (3). Therefore, iron burden must be monitored accurately and closely for early diagnosis and chelation titration (4,5). The gold standard measurement has been liver biopsy, as the liver is the dominant iron store and liver iron reflects total body iron (6). However, biopsy is invasive and subject to sampling errors, and is particularly ill-suited to young children (7,8). In recent years, magnetic resonance imaging (MRI) has emerged as the most widely used non-invasive technique to quantify liver iron content, and correlation with biopsy results has been demonstrated (9-11). To date, there have been many applications of MRI to study liver iron (12-17), and a number of methods exist for acquiring and analyzing the MRI data. However, our understanding of how to optimize these procedures for accurate iron measurement needs to be improved.

Either T2 or T2* techniques can be used for MRI iron quantification. Most experience to date has been with T2 methods, which demonstrate correlation with histological findings over clinically relevant iron levels (9,10) and have received FDA-approval (Ferriscan™) as a non-invasive imaging surrogate to liver biopsy. Potentially more attractive are very rapid T2*-based methods that would avoid the need for sedation (e.g. in pediatric patients). To capture the rapid T2* signal decay in heavily iron-loaded tissue, one requires a very short minimum echo time (18). Analysis of T2* decay curves can be performed using a single exponential decay model (19) or a non-monoexponential model such as an exponential plus constant offset (10,18) to
correct for noise and other artifacts in later echo images (20). However, recent simulations studies have shown that iron levels are generally overestimated using the constant offset model (21,22). More accurate measurement can be achieved using a truncated model by discarding points below the noise level from data-fitting (23), but this was proposed for measuring lower iron levels in the heart (22). In these few investigations on technical considerations, the influence of noise (22), analysis models (18,21,22), and acquisition parameters (18) on measurement accuracy were considered separately. Moreover, the range of T₂* values investigated were either not applicable to higher iron content found in the liver (22) or inadequately resolved in the high liver iron (i.e. very low T₂*) regime (21). There currently remains a need for an integrated understanding of these collective influences in order to optimize the T₂* approach for rapid and accurate liver iron measurement.

In this study, we investigate the ability of T₂* quantification to measure liver iron levels encountered in children with iron overload. Three analysis models are considered: mono-exponential, constant offset, and truncated. Performance of these models is evaluated through simulations under varying conditions of noise and different choices of minimum echo time and echo spacing. In-vivo validation is performed in a prospective study involving pediatric patients with iron overload.

MATERIALS AND METHODS

Curve-Fitting Models

Three models were evaluated to fit the T₂* signal versus echo time (TE) decay curve. The simplest is a mono-exponential function:

\[ S = S_0 \exp\left(-TE \cdot R_2^*\right) \]  \[1\]
where $S$ is the measured signal, $S_0$ is the initial signal amplitude, and $R_2^* = 1/T_2^*$ is the effective transverse relaxation rate. The main disadvantage of this model is its sensitivity to noise. To minimize the influence of noise on late echo signal, two alternative models have been applied. The first is a constant offset model (10,18), where the constant $C$ accounts for noise and artifacts:

$$S = S_0 \exp\left(-TE \cdot R_2^*\right) + C$$  \[2\]

The second is a truncated model, where only data well above the noise level are fitted to a mono-exponential function (Eq. [1]) (23). In this study, the criterion $S < 2\sigma_{\text{noise}}$, where $\sigma_{\text{noise}}$ is the noise standard deviation, was used to identify the first echo (and all subsequent echoes) to discard. Curve-fitting employed the Levenberg-Marquardt algorithm and assumed physically meaningful (i.e. positive) bounds on the parameters $S_0$ and $R_2^*$.

**Simulations**

The range $R_2^* = 20–1000$ Hz typically seen in liver iron patients was investigated. Signal intensity, $S$, was generated using the signal equation for a gradient-echo sequence. A longitudinal relaxation time of $T_1 = 586$ ms for liver (24) was assumed, and MR parameters were chosen to correspond to two $T_2^*$ multi-echo gradient-echo sequences adopted in our institution for clinical assessment of liver iron: breath-held ($TE = [1-16]$ ms, repetition time (TR) = 200 ms, flip angle (FA) = 20°, 12 echoes) and non-breath-held ($TE = [2.3-30]$ ms, TR = 500 ms, FA = 60°, 11 echoes).

Gaussian-distributed complex noise, with standard deviation $\sigma_{\text{noise}}$, was added to the generated signal. Different signal-to-noise ratios (SNR = $M_0/\sigma_{\text{noise}}$, where $M_0$ is the fully relaxed magnetization) in the range 20 to 300 were evaluated. For each SNR level and $R_2^*$ value, 10000 trials were run. The mean $R_2^*$ and standard deviation were then computed to determine
measurement accuracy and precision for all three models (*cf. Models*) and both T$_2^*$ acquisition protocols. All simulations were implemented in Matlab® (v.7.0, The Mathworks, Natick, NA).

**Patient Study**

Fifty-four patients (13 ± 5 years of age) with iron overload were enrolled in this Institutional Review Board approved prospective study. T$_2$ and T$_2^*$ data were acquired on a clinical 1.5T Siemens (Avanto). The T$_2$ protocol employed a multi-slice spin-echo sequence (total scan time = 30 min 20 s): 11 axial slices, slice thickness (TH) = 5 mm, matrix size $= 192 \times 256$, FA = 90°, TR = 2500 ms, number of excitations (NEX) = 1; repeated for 5 different TE = 6, 9, 12, 15, 18 ms. The T$_2^*$ protocols employed multi-slice multi-echo gradient-echo sequences. One was a non-breath-held sequence (total scan time = 3 min 22 s): 11 axial slices, TH = 6 mm, matrix size $= 101 \times 192$, variable field-of-view (FOV) (default = 350 mm), FA = 60°, TR = 500 ms, NEX = 4; 11 equally spaced TE $= [2.3-30]$ ms. The second was a breath-held sequence (scan time = 13 s per breath-hold): 11 axial slices, TH = 5.5 mm, matrix size $= 64 \times 128$, variable FOV (default = 400 mm), FA = 20°, TR = 200 ms, NEX = 1; 12 equally spaced TE $= [1-16]$ ms. The T$_2$ and non-breath-held T$_2^*$ protocols were applied to all patients, of which twenty-five also underwent the breath-held T$_2^*$ sequence.

Reference iron concentration measurements based on T$_2$ data were provided by FerriScan™. Data from the T$_2^*$ protocols were analyzed using in-house developed software (Matlab v.7.0) on a pixel-wise basis, using the single slice location corresponding to that in the FerriScan analysis. The first step was removal of background pixels by excluding those with signal intensity below 5% of the image maximum (on the first TE image). R$_2^*$ maps were then computed on the remaining pixels using the constant offset and the truncated models. In applying the truncated model, the threshold $\sigma_{\text{noise}}$ was determined from the signal standard deviation in a
homogeneous region of the liver on the first TE image. A single ROI-derived mean $R_2^*$ value was then computed from the $R_2^*$ map, where the ROI was manually drawn to encompass the entire liver and exclude blood vessels and ducts. Finally, the mean $R_2^*$ value was converted to liver iron concentration $[Fe]$ (mg/g) using the liver calibration curve described in (10):

$$[Fe] = 0.0254R_2^* + 0.202 \quad [3]$$

RESULTS

Simulations

Figure 1 shows how the constant offset and truncated models fit the $T_2^*$ signal decay curve under different noise conditions (SNR = 50 and 200), using the sequence $TE = [2.3-30]$ ms for illustration. The constant in the offset model is seen to track the noise level, and the number of points for fitting the truncated model is greatly reduced at low SNR. Thus, although the two model fits are theoretically identical under noiseless conditions, their divergence is apparent as SNR is reduced.

Figure 2 compares the performance of all three curve-fitting models using two different acquisition schemes ($TE = [1-16]$ ms and $[2.3-30]$ ms). Under optimal SNR conditions (SNR = 200, Fig.2a) and echo spacing (min TE = 1 ms), all models provide accurate $R_2^*$ estimates. So long as a high SNR is achieved, reasonable accuracy is maintained even with a higher minimum TE (min $TE = 2.3$ ms), although now model frailties become apparent, namely, overestimation associated with the constant offset model and underestimation of high $R_2^*$ species associated with the truncated model. In contrast, under poor SNR conditions (SNR = 50, Fig.2b), using a short minimum TE is critical to maintaining accuracy of $R_2^*$ estimates. As before, either the constant offset or the truncated model may be applied. In the extreme case of poor SNR and echo
spacing (min TE = 2.3 ms), the strengths of both models are most evident: the constant offset model is less accurate (overestimation) than the truncated model but provides the best robustness across the entire R₂* range.

Figure 3 illustrates the precision of R₂* estimates. No major difference in precision was observed between the constant offset and truncated models. Also, precision is seen to be determined mainly by SNR, with the influence of echo sampling being evident only at low noise levels (Fig.3a).

Figure 4 illustrates the influence of SNR on R₂* measurement accuracy. Low to moderate R₂* species (200 and 500 Hz) can be accurately measured down to low SNR levels irrespective of the acquisition scheme, with the truncated model providing the best accuracy and the constant offset model slightly overestimating R₂*. In contrast, high R₂* species (900 Hz) are more susceptible to underestimation from poor conditions (i.e. low SNR and/or high minimum TE), and accuracy is best ensured using the constant offset model. Furthermore, much lower SNR levels can be tolerated if a short minimum TE is used.

**Patient Study**

Figure 5 illustrates an anatomical image of the liver (Fig.5a) and corresponding R₂* maps obtained from the constant offset (Fig.5b) and truncated model (Fig.5c). Lower R₂* estimates observed in the truncated model map are consistent with simulations predictions. A typical example of a ROI encompassing the liver and excluding blood vessels is also shown (yellow outline). The SNR in the in-vivo data was approximately 200 and 50 for the non-breath-held (TE = [2.3-30] ms) and breath-held (TE = [1-16] ms) protocols, respectively.

Figure 6 shows T₂* versus reference (i.e. FerriScan) measurements of iron levels for all 54 patients. T₂* data are derived from the non-breath-held protocol. Both the constant offset and
Accuracy of $T_2^*$ Measurement of Liver Iron

truncated models provide accurate measurements, with differences emerging for iron levels higher than 10 mg/g. Consistent with theoretical predictions, the truncated model incurs an underestimation, while the constant offset model slightly overestimates but maintains reasonable accuracy throughout the entire range of iron levels.

Figure 7 compares the breath-held (low min TE, low SNR) versus non-breath-held (high min TE, high SNR) $T_2^*$ sequence for measuring iron concentrations. Low to moderate iron levels (up to 10 mg/g) are in close agreement, independent of acquisition protocol or analysis model. This trend confirms the simulations predictions of Figure 4. High iron levels, on the other hand, are lower using the breath-held protocol, most likely due to an SNR level below that required for achieving comparable accuracy with the non-breath-held sequence (see Figure 4).

**DISCUSSION**

The ability of $T_2^*$ methods to measure liver iron content rapidly and non-invasively is extremely attractive for the pediatric patient. Not only does it address the invasiveness and sampling error to which traditional liver biopsy is prone, but it can eliminate the need for sedation that lengthy $T_2$ acquisitions often require. However, the reliability of $T_2^*$ measurements need to be better understood to gain similar acceptance as FDA-approved $T_2$ protocols. In this study, simulations were conducted to quantify the necessary conditions on both acquisition and analysis to achieve accurate $T_2^*$ measurement of iron concentration. A cohort of pediatric patients with elevated iron levels was assessed using $T_2^*$ and standard $T_2$ protocols. The simulations and in-vivo results are consistent, and they demonstrate that $T_2^*$ can provide accurate measurement of iron levels. Analysis using the truncated model is most accurate, but
under conditions of low SNR and/or suboptimal echo spacing, the constant offset model maintains the best robustness even for high iron species.

Two T$_2^*$ protocols were investigated in this study (breath-held and non-breath-held), because they represent different solutions that are uniquely suited to some children. To understand the clinical motivation for these two approaches and the choice of parameter settings, consider first the ideal scenario. Ideally, the acquisition sequence should employ a short minimum TE so that high iron species can be accurately quantified. Accuracy also benefits from rapid scanning to eliminate motion artifacts. Both traits are achieved with the breath-held sequence (TE = [1-16] ms). However, the SNR associated with this sequence is generally low due to a short minimum TE and other parameter settings (e.g. TR, NEX) to maintain a short acquisition time. Therefore, in choosing this acquisition approach, one must consider the tradeoff between SNR and minimum TE in determining the T$_2^*$ measurement accuracy achievable (see Fig.4). Consideration must also be given to the patient status. Although breath-holding effectively removes respiratory motion, many children cannot breath-hold for 10 to 20 s. To address this issue, one solution is to blur out the motion artifacts by averaging (i.e. NEX > 1). Since time constraints are no longer a limitation, the maximum TE can be extended for improved measurement of low iron levels (without sacrificing slice coverage). More importantly, the freedom to scan longer allows manipulation of various MR parameters (e.g. increasing TR and minimum TE) with the aim of substantially increasing SNR to improve T$_2^*$ accuracy. These were all factored into the design of the non-breath-held sequence (TE = [2.3-30] ms).

The major results of this study can now be interpreted within the clinical framework described above. Foremost, T$_2^*$ assessment of iron levels correlates well with reference T$_2$ measurements in the patient population. Deviation of T$_2^*$ estimates from true iron levels is seen
mainly for high iron species, because of a lower than desirable SNR level for the particular TE spacing scheme, and can be predicted from simulations given the SNRs achieved in-vivo (see Fig. 4). To maintain robust measurements across the entire range of iron concentrations under suboptimal conditions, the constant offset model is a better choice than the truncated model, as demonstrated in both simulations and in-vivo results (Fig. 2 and 6). Another important observation is that the breath-held and non-breath-held sequences offer comparable accuracy (Fig. 7), and as such, children who cannot breath-hold are at an equal advantage. The superior accuracy expected from a breath-held sequence was not obtained, despite a much lower SNR requirement, most likely because noise remained problematic. Accuracy was more easily attained through substantial SNR gains possible from a non-breath-held sequence to outweigh the negative impact of a higher minimum TE. Future work will focus on optimizing the two protocols, using our simulations results for guidance, to achieve the best SNR/TE combination for accurate T2* measurement.

Because the goal of this study was to optimize T2* measurement accuracy, consistency with similar studies was ensured before conducting an integrated, detailed investigation. For example, our finding that over- and underestimation of iron is associated with the constant offset and truncated models, respectively, have also been reported by others (21,22). However, in those studies, either a single or a limited range of T2* values was shown, neither of which is adequate to characterize the full behavior over the T2* range relevant for elevated liver iron. Ghugre et al (18) examined a range of T2* values as in our study, but lower 1/T2* species (up to 500 Hz instead of 1000 Hz in our study) found in myocardial iron overload were examined. A second, much less well addressed topic in the literature is the influence of noise. Most reports mention noise in reference to appropriate modeling and do not explicitly consider the noise level. The
only detailed study to consider noise effects is He et al (22), but they reported noise as the number of signal averages and not the absolute SNR. Thus, their conclusions cannot easily be generalized to other acquisition sequences. Our study uniquely investigates $T_2^*$ measurement accuracy as a function of SNR and can, therefore, be used to guide the design of parameter settings for any sequence. There also exist a couple of discrepancies that need to be noted. First, the reported accuracy of a truncated model (22) may seem at odds with the more widely used and well recognized constant offset model (18,21). This discrepancy is consistent with our findings, which show that the truncated model is more accurate under ideal sampling and SNR, but below a certain SNR threshold the constant offset model offers better accuracy. The second note is that in-vivo iron measurement curves (Fig.6) appear to underestimate true iron levels more than predicted by simulations (Fig.2). This phenomenon is expected, since the simulations $T_2^*$ represents the mean estimate under noisy conditions, whereas the in-vivo mean $T_2^*$ is additionally influenced by the $T_2^*$ distribution in the liver. Therefore, in-vivo, iron species higher than the mean $1/T_2^*$ that are more severely underestimated will contribute to an overall lower $1/T_2^*$ estimate, and this effect will be larger for the truncated model.

There exist additional factors on liver iron measurement, which were not considered in this study. One component is the effect of slowly relaxing species, such as blood, for which a biexponential model has been applied and validated (25). Another is alteration in the signal intensity on a gradient-echo sequence from a high fat content, known as the “beating effect” (18). This arises from different phase accruals between fat and water protons, which results in signal cancellation and signal reinforcement at TE of 2.3 ms and 4.6 ms, respectively. An interval of 2.3 ms can be used to minimize this effect (and then fitting every other datapoint). However, these conditions on echo times put restrictions on optimizing TE for accurate $T2^*$
measurement. Further investigation is required on the relative merits of optimizing TE for either T2* accuracy or balancing fat/water signal. A third issue is possible differences between pixel-wise and ROI-based analysis of signal decay. The simpler ROI-based approach was not assessed, since significant differences were not expected in pediatric patients, who generally have more homogeneous iron deposition patterns. In fact, distribution homogeneity was evidenced by similar average mean and median iron values calculated from pixel-maps. The more complex pixel-wise approach was adopted because of its ability to discern heterogeneous iron deposition patterns, which is especially valuable in adult patients, such as those with cirrhosis. In general, a single average iron level measurement, however derived (e.g. from an ROI-based analysis), should be supplemented with an iron distribution map for a complete picture of iron overload.

In conclusion, T2* MRI is a rapid method for quantifying liver iron and offers similar accuracy as approved T2 approaches. Regardless whether a breath-held or non-breath-held acquisition is most suited to a particular patient, accuracy can be ensured so long as the minimum SNR requirement, as quantified in this study, is met. The choice of model also depends on the acquisition conditions and range of iron levels considered. Under ideal conditions of short minimum TE and high SNR, the truncated model is preferred. However, the constant offset model offers more robust measurements, especially for high iron species, when the trade-off between SNR and TE scheme is insufficient to maintain measurement accuracy.
ACKNOWLEDGMENTS

The authors thank Dr. John Wood for helpful discussions.
REFERENCES


FIGURE LEGENDS

**Figure 1**: Fitting T₂* signal decays curves with constant offset (dashed line) and truncated model (solid line). Simulated noise levels are (a) SNR = 200 and (b) SNR = 50.

**Figure 2**: Accuracy of R₂* measurements using three fitting models under different acquisition conditions. The two data sampling schemes (TE = [1-16] ms and TE = [2.3-30] ms) are presented in the left and right columns, respectively; with (a) SNR = 200 and (b) SNR = 50.

**Figure 3**: Precision of R₂* measurements using the constant offset and truncated models under different acquisition conditions. Error bars represent standard deviation. The two data sampling schemes (TE = [1-16] ms and TE = [2.3-30] ms) are presented in the left and right columns, respectively; with (a) SNR = 200 and (b) SNR = 50.

**Figure 4**: Influence of SNR on the quantification of low, medium, and high R₂* (200, 500, and 900 Hz, respectively) using three different models on data acquired with (a) TE = [1-16] ms and (b) TE = [2.3-30] ms.

**Figure 5**: (a) Anatomical image acquired with the non-breath-held T₂* protocol; the image shown corresponds to TE = 2.3 ms. Corresponding R₂* maps derived from (b) the constant offset model and (c) the truncated model. A typical ROI encompassing the whole liver and excluding blood vessels is shown in yellow.

**Figure 6**: Comparison of T₂* assessment of iron concentration [Fe] using the constant offset model (gray triangles) and the truncated model (black circles) versus the reference T₂ [Fe] by FerriScan in 54 patients. T₂* data was acquired using the non-breath-held sequence. Line of identity is shown.
Figure 7: Comparison of $T_2^*$ assessment of iron concentration [Fe] from data acquired with the breath-held (TE = [1-16] ms) versus non-breath-held sequence (TE = [2.3-30] ms) in 25 patients. Iron concentrations derived from the constant offset model (gray triangles) and the truncated model (black circles) are compared. Line of identity is shown.
a) Real data: SNR=200
- Constant offset model
- Truncated model

b) Real data: SNR=50
- Constant offset model
- Truncated model
Estimated $R^*_e$ (Hz) - TE: [2.3-30] ms

Estimated $R^*_e$ (Hz) - TE: [1-16] ms

Mono-exponential  ○ Constant offset  ● Truncated
$T_2^* \{\text{Fe}\} \text{ (mg/g) - TE: [2.3-30]} \text{ ms}$

- Constant offset
- Truncated
- Identity