The effect of measurement area size on the reliability of myocardial iron load measurement in cardiac magnetic resonance imaging examinations of thalassemia patients

Abstract

Background/aim: The aim of this study was to evaluate the intraobserver and interobserver reliability of cardiac T2* MRI measurements in the different region of interest (ROI) sizes.

Materials and methods: Cardiac T2* MRIs of 24 thalassemia major patients were evaluated. Two different (ROI) sizes were used for measurement. In the first measurement, an ROI with approximately 5 mm in diameter was used in the interventricular septal myocardium. In the other method, all interventricular septal myocardium was used to measurement. The intra-observer and interobserver variabilities were assessed with intraclass correlation coefficient (ICC).

Results: Measurements of the first observer, the ICC of the small-sized ROI (ssROI) were 0.869, and for the second observer, the ICC of the ssROI was 0.659. The ICC of the whole-septal ROI (wsROI) were 0.991 for the first observer and 0.980 for the second observer. Interobserver variability, for the mean measurement, the ssROI’s ICC was 0.442, and the wsROI’s ICC was 0.883

Conclusion: In the evaluation of myocardial iron load with T2* MRI, we suggest making measurements with ROI including all of the interventricular septum in consequence of high intraobserver and interobserver consistency.

Keywords

Thalassemia major, myocardial iron load, T2* MRI, ROI size
1. Introduction

In refractory anemias as thalassemia major, there is iron overload in internal organs especially in the liver, heart, and pancreas due to recurrent transfusions. The heart is the most important organ because of effect to mortality and morbidity in thalassemia major patients [1]. There is a strong correlation between the serum ferritin level and iron load in internal organs [2]. Moreover, the serum ferritin level can be affected by inflammatory conditions [3]. Nonetheless, there isn’t a strong correlation between the serum ferritin level and cardiac iron load [4]. It is very important that myocardial iron load be accurately diagnosed, so that effective treatment can be administered.

Today, non-invasive measurement of myocardial iron load is performed using T2* MRI and the software supplied by equipment manufacturers. There is a lot of software for this purpose other than MR suppliers. Despite MRI technological improvements, due to heterogeneity of myocardial iron distribution, MRI sequence standardization problems, and software problems, many clinicians are concerned about the evaluation of myocardial iron load [5]. Of particular note is the problem of the lack of consistency of repeated measurements (reliability), which makes accurate and confident diagnosis difficult. In addition, the literature lacks sufficient data concerning how region of interest (ROI) size affects measurement of myocardial iron load. As such, the present study aimed to determine the intraobserver and interobserver reliability of several ROI sizes used for measurement of myocardial iron load via T2* MRI in thalassemia major patients.

2. Material and Methods

2.1. Patients

T2* MRI images of 24 thalassemia major patients (14 female and 10 male) that were obtained between June and September 2016 were retrospectively evaluated. Exclusion
criteria included motion artifacts and artifacts due to the MRI technique. The study protocol was approved by the our University Ethics Committee.

2.2. MRI

MRI was performed using a 1.5 T MR scanner (GE, Signa, USA). T2*-weighted MRI was used to measure myocardial iron load, using a fast gradient-echo multi-echo sequence (8 echoes) and ECG triggering. Short-axis views (midventricular) of the left ventricle were obtained. Multi-echo sequence parameters were as follows; matrix: 192 × 256 mm; flip angle: 25°; TR: 120 ms; TE: 1.5-13.31 ms; FOV: 41 × 41 mm; slice thickness: 5 mm.

2.3. Image analysis

Measurement was performed using CMRtools software (CMRtools Thalassemia-Tools, Cardiovascular Imaging Solutions, United Kingdom). The DICOM images of thalassemia patients in our hospital radiology archive were copied to local memory and measurements were done in the local memory. All T2* measurements were performed by 2 observers using mid-ventricular short axis slices. An ROI was drawn in the septum in the first slice and the ROI in the other 7 slices was automatically selected by CMRtools; however, it was possible to manually select the ROI in the latter 7 slices if necessary due to motion artifacts.

For measurement of the first slice, an ROI approximately 5 mm in diameter was used and it was referred to as a small-sized ROI (ssROI). The lower limit for interventricular septum thickness is 6 mm; therefore, a 5-mm ROI was used to ensure that it included the entire septum, but none of the lumen. In fact, there are some centers that measure cardiac iron load by this type of method. The ssROI was drawn so that it did not include anything beyond the myocardium (Figure 1). The ssROI was selected manually by inputting 3 points in CMRtools. In the second measurement method, an inner and outer contour was
drawn manually. Next, a reference ROI—including the septum—was defined automatically by CMRtools. The operator could make changes and arrange the area of ROI. This measurement was referred to as whole-septal ROI (wsROI) (Figure 2). The drawing did not extend beyond the myocardium because the blood’s signal can affect measurement. ssROI and wsROI measurements were performed by 2 observers, with a 1-week interval between each measurement and blinding of the previous values and each other. First observer had 1 year of experience with cardiovascular imaging and had been using CMRtools for 6 months. Second observer was performing T2* MRI measurement and using CMRtools for the first time.

2.4. Statistical analysis

Intraobserver and interobserver reliability were assessed using with intraclass correlation coefficient (ICC). The ICC of ssROI and wsROI were calculated for both observers. For the interobserver reliability, ICC estimates and their 95% confident intervals were calculated using SPSS statistical package version 20 (SPSS Inc, Chicago, IL) based on a mean-rating (k = 2), consistency-agreement, 2-way random-effects model. For the intraobserver reliability, ICC estimates and their 95% confident intervals were calculated using SPSS statistical package version 20 (SPSS Inc, Chicago, IL) based on a mean-measurement (k = 2), absolute-agreement, 2-way mixed-effects model. We accepted that ICC values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability [6].

3. Results

Among the 24 patients, 14 (58.3%) were female and 10 (41.7%) were male. Mean age of the entire patient population was 26.08 years (range: 17–44 years), versus 26.27 years
Observers’ measurements are summarized in Table 1. For the first observer, the ICC of ssROI was found 0.869 (0.697-0.943 in the %95 confidence interval), and the ICC of wsROI was found 0.991 (0.978-0.996 in the 95% confidence interval). Among the measurements of observer 1, the reliability of the ssROI method was “good” and the wsROI method was “excellent” (Table 2). For the second observer, the ICC of ssROI was found 0.659 (0.211-0.852 in the %95 confidence interval), and the ICC of wsROI was found 0.980 (0.954-0.991 in the 95% confidence interval). Among the measurements of observer 1, the reliability of the ssROI method was “moderate” and the wsROI method was “excellent” (Table 3).

In terms of interobserver reliability, the ICC of ssROI was found 0.442 (0.056-0.713 in the %95 confidence interval), and the ICC of wsROI was found 0.883 (0.749-0.948 in the 95% confidence interval). In terms of observers’ measurements, the reliability of the ssROI method was “poor” and the wsROI method was “good-excellent” (Table 4).

Figure 3a and figure 3b shows that interobserver measurement consistency using the wsROI method ($R^2 = 0.842$) was slightly higher than that based on the ssROI method ($R^2 = 0.696$). Due to values are mean of both measurement method, calculated $R^2$ values are close to each other. When we evaluate single measurements beside mean values of $T2^*$ measurement, $R^2$ values for each measurement method (wsROI and ssROI) were significantly different (figure 4a and 4b). The first wsROI measurement value for each observer, $R^2$ value 0.790 results almost to the mean value; however, the first ssROI measurement for each observer was slightly low ($R^2 = 0.248$). While single measurements using the wsROI method are as confidential as the mean of multiple measurements with
the ssROI method, mean values obtained with the ssROI method in different parts of the septum results relatively confidential and reproducible results.

4. Discussion

In the present study wsROI’s ICC was significantly higher than ssROI’s ICC. The wide variation in ssROI measurements might have been due to heterogeneity of myocardium iron load or MRI artifacts. In a study that included 14 idiopathic hemochromatosis patients, Olson et al. [7] emphasized heterogeneity of myocardial iron load in cardiac autopsies. They also suggested that iron load might be focal and recommended repeating endomyocardial biopsies. Presently, endomyocardial biopsies are not routinely used in the management of iron overload, as they are invasive and such non-invasive modalities as T2* MRI are readily available.

A study on the heterogeneity of iron load in the myocardium reported heterogenous iron load in 50% of the study’s thalassemia patients [8]. The researchers made multiple measurements in the myocardium using the multislice technique and obtained global T2*-weighted images of the myocardium. They showed that global myocardial T2* measurements correlated well with one measurement of the midventricular septum. Of note, the researchers used a whole septal ROI, which is important because an ROI of that size has the advantage of providing the mean of different parts of the myocardium.

Roghi et al. observed significantly different myocardial T2-weighted values based on different tools [5]. They used CMRtools and the pixel-wise (PW) inline myocardial T2*-weighted mapping system for characterizing myocardial iron load in thalassemia major patients. They concluded that different methods can categorize iron accumulation differently, and that consecutive measurements of a patient can result in misinterpretation,
especially of borderline values, which can negatively affect iron chelation treatment.

Earlier studies have shown that there is lack of standardization across iron load measurement methods, including the unit of measure [2,9]. In addition, as different measurement programs might categorize a patient differently, it is of paramount importance to standardize software, so as to minimize intraobserver and interobserver variability.

The present study aimed to determine the effect of ROI size on the reliability of myocardial iron load measurements. Both intraobserver and interobserver ICC were higher for measurements obtained via wsROI than via ssROI, indicating that as ROI size decreased measurement inconsistency increased. Figure 5 shows both observers’ measurements in a patient; interobserver evaluation showed that the difference in measurements with the ssROI method was 19.28%, versus 1.12% with the wsROI method. Additionally, there was some software measured with ssROI. Whatever why using ssROI, aroused suspicion in repeatability of measurements. And also the main dilemma in this situation was which measurements should establish on management. In terms of chelation therapy, clinicians often must decide if they should use the highest iron load measurement or the mean load measurement of the septum; however, based on the results of long-term studies mean overload measurement of the septum is the best choice.

The present study has some limitations, including its retrospective design and the exclusion of a large number of patients due to MRI artifacts (primarily because of technician inexperience). To the best of our knowledge the present study is the first to examine the effect of ROI size on the reliability of T2* MRI measurement of myocardial iron load in thalassemia patients. In cardiac T2*-weighted measurements, some of the update software was user-dependent, while some of them contained template about ROI
size and shape and measurements. Based on the present findings, we strongly advocate the standardization of T2* MRI measurement of myocardial iron load.

5. Conclusion

Different measurements of consecutive T2* MRIs using ssROI in the same patient makes challenging the monitoring, management, and treatment of thalassemia patients. Based on the present findings, for the evaluation of myocardial iron load based on T2*-weighted MRI we suggest using an ROI that includes the entire septum, as it is associated with high intraobserver and interobserver reliability. Lastly, the present findings clearly show that as ROI size increases the reliability of T2* MRI measurement of myocardial iron load increases.

Acknowledgement/Disclaimers/Conflict of interest disclosure

Authors we have no conflicts of interest to disclose.

References


Table 1: Both observers made two measurements by two different methods. As a result, each observer made four measurements. All measurements of the observers are given in the table.

<table>
<thead>
<tr>
<th></th>
<th>Observer 1 measurement</th>
<th>Observer 2 measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wsROI-1</td>
<td>wsROI-2</td>
</tr>
<tr>
<td>N</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Mean</td>
<td>39.6129</td>
<td>39.8896</td>
</tr>
<tr>
<td>Minimum</td>
<td>11.91</td>
<td>11.79</td>
</tr>
<tr>
<td>Maximum</td>
<td>117.15</td>
<td>106.00</td>
</tr>
</tbody>
</table>

(ws: whole-septal, ss: small-sized, ROI: region of interest, N: Number of cases).

Table 2: Comparison of ssROI and wsROI measurements for observer 1

<table>
<thead>
<tr>
<th></th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC Coefficient</td>
</tr>
<tr>
<td>ssROI Avarage measure</td>
<td>0.869</td>
</tr>
<tr>
<td>wsROI Avarage measure</td>
<td>0.991</td>
</tr>
</tbody>
</table>

**Table 3:** Comparison of ssROI and wsROI measurements for observer 2.

<table>
<thead>
<tr>
<th></th>
<th>95% confidence interval</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC Coefficient</td>
<td>Lower bound</td>
</tr>
<tr>
<td>ssROI</td>
<td>Average measure</td>
<td>0.659</td>
</tr>
<tr>
<td>wsROI</td>
<td>Average measure</td>
<td>0.980</td>
</tr>
</tbody>
</table>

(ws: whole-septal, ss: small-sized, ROI: region of interest.)

**Table 4:** Interobserver reliability of ssROI and wsROI measurements, based on both observers’ means.

<table>
<thead>
<tr>
<th></th>
<th>95% confidence interval</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC Coefficient</td>
<td>Lower bound</td>
</tr>
<tr>
<td>ssROI</td>
<td>Average measure</td>
<td>0.442</td>
</tr>
<tr>
<td>wsROI</td>
<td>Average measure</td>
<td>0.883</td>
</tr>
</tbody>
</table>

(ws: whole-septal, ss: small-sized, ROI: region of interest.)
Figures and figures legend

Figure 1: An ROI approximately 5 mm in diameter (ssROI) was used.

Figure 2: ROI including the entire septum (wsROI).
Figure 3: Interobserver reliability of measurements using the wsROI method (on the left) is slightly higher than that of measurements using the ssROI method (on the right). Comparisons were made using mean measurements.

Comparisons were made using mean measurements.

Figure 4: Interobserver reliability of measurements using the wsROI method (on the left) is slightly higher than that of measurements using the ssROI method (on the right). Comparisons were made using the first measurements of each observer.
Figure 5: Measurements of both observers with the wsROI method were shown in the same patient. The top one is the first observer and the bottom one is the measurement of the second observer. T2* values are shown in the upper left corner of each measurement.