



Reference range determination for imaging biomarkers: Myocardial T1

Journal:	<i>Journal of Magnetic Resonance Imaging</i>
Manuscript ID	JMRI-18-1394.R1
Wiley - Manuscript type:	Original Research
Classification:	Clinical Science, Cardiovascular and interventional imaging < Clinical Science, Heart < Cardiovascular and interventional imaging < Clinical Science
Manuscript Keywords:	Imaging biomarker, T1 mapping, reference range, reference interval, normal values



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Reference range determination for imaging biomarkers: Myocardial

T1

FOR PEER REVIEW ONLY

Abstract

BACKGROUND: Imaging biomarkers, such as the T1 relaxation time of the myocardium using MRI, can be valuable in cardiac medicine if they are properly validated. Consensus statements recommend that for myocardial T1, each investigator should establish a reference range.

PURPOSE: To describe a statistically valid method for determining and reporting the reference range in each center, which simultaneously minimizes the twin risks of undersampling, leading to a uselessly uncertain range, and oversampling, which exposes volunteers to unnecessary scanning and wastes resources.

STUDY TYPE: Cohort

POPULATION: 278 normal human subjects without cardiac disease from two cardiac MR centers.

FIELD STRENGTH/SEQUENCE: 1.5T and 3T; Modified Look-Locker Inversion recovery sequence

ASSESSMENT: n/a The T1 relaxation time was estimated from multiple samples of tissue magnetization after inversion. A valid method for calculating a reference range is used.

STATISTICAL TESTS: Shapiro-Wilk test for Normality; Tukey robust approach for identification of outliers; reference range calculation with confidence intervals.

RESULTS: Reference ranges for measurement of myocardial T1 were calculated, with confidence intervals, enabling comparison with clinically important differences. At 3.0T: 1129 to 1301 ms at site 1 (n=21) and 1160 to 1309 ms at site 2 (n=59), and at 1.5T at site 2: 933 to 1020 ms (male, n=130) and 965 to 1054 ms (female, n=68). The 3T reference range from site 1 was successfully benchmarked against the 3T reference range at site 2.

1
2
3 DATA CONCLUSION: Myocardial T1 reference ranges can be properly characterized,
4
5 enabling clinical comparison to a valid reference range with known confidence intervals,
6
7
8 using methodology similar to that described in this report.
9
10
11
12
13
14
15
16

17 **Keywords**
18
19

20 Imaging biomarker, T1 mapping, reference range, reference interval, normal values
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FOR PEER REVIEW ONLY

INTRODUCTION

Imaging biomarkers play an increasing role in cardiac medicine. The BEST Resource (1) from the FDA–NIH Biomarker Working Group defines a biomarker as “a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions”. It continues “...molecular, histologic, radiographic, or physiologic characteristics are types of biomarkers”, making imaging biomarkers such as the MR T1 relaxation time of the myocardium, if validated, formally equivalent to blood-borne molecular biomarkers such as troponin T and LDL-cholesterol, or to physiologic biomarkers such as systolic blood pressure.

A biomarker measured in similar patients with different equipment in different hospitals should yield similar values (2). This is an essential prerequisite before embarking on large multicenter outcome studies to create an evidence base for use. Unfortunately, for myocardial T1, this basic technical validation step is rather challenging. The design and operation of different manufacturers’ MR equipment may affect accuracy, and in the case of myocardial fibrosis, even quite small differences in T1 may be biologically important. Recognising this risk of between-center irreproducibility, consensus statements recommend that each investigator should establish a reference range for myocardial T1 for his or her own equipment (3, 4).

A study to fully define the reference range involves a considerable number of subjects and significant costs – particularly if it has to be replicated by every investigator and for every new scanner installation or upgrade. More importantly it poses an ethical challenge: How best to define an adequate reference range while minimizing the exposure of volunteers to the burden of MR scanning from which they gain no benefit.

1
2
3 The purpose of this work is to provide statistically valid methodology for investigators to
4 calculate and interpret a reference range in their centre, while minimizing the twin risks of
5 undersampling, leading to a uselessly uncertain range, and oversampling, which exposes
6 volunteers to unnecessary scanning and wastes resources. Related tasks are also addressed,
7 such as predicting the number of subjects needed, dealing with outliers, and benchmarking
8 one reference range against another. Myocardial T1 reference ranges are reported. With the
9 recent availability of mapping-based myocardial tissue characterization on clinical magnetic
10 resonance systems, mapping of the T1 relaxation time is an exemplar imaging biomarker, but
11 our approach is general.
12
13
14
15
16
17
18
19
20
21
22
23
24

25 **METHODS**

26
27
28 Data acquired for this work was acquired at site 1 using a healthy volunteer database (each
29 subject provided informed consent and the study was approved by the local Ethics Review
30 Board), and from site 2 using data from the local T1 quality control collective (aggregate data
31 used for service evaluation does not require further ethical review (decision of the local
32 Ethical Review Board)).
33
34
35
36
37
38
39

40 The calculation of four myocardial T1 reference ranges of subjects without cardiac disease is
41 reported here: a 3T reference range which includes both male and female subjects from site 1
42 (n=21, 5 male and 16 female, with age mean/sd 57/12y); a 3T reference range which includes
43 both male and female subjects from site 2 (n=59, 38 male and 21 female, with age mean/sd
44 52/9y); 1.5T reference range for males at site 2 (n=132 with age mean/sd 50/8 y); and a 1.5T
45 reference range for females at site 2 (n=68 with age mean/sd 48/10 y). Site 1 used a single
46 Achieva TX MR system (Philips, Best, The Netherlands) using patient-adaptive RF
47 shimming and a 6 channel receiver array. Site 2 used Ingenia MR systems (Philips, Best, The
48 Netherlands) at both 1.5T and 3.0T, both using a 28 channel receiver array. All acquisitions
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 employed a MOLLI 5s(3s)3s product sequence (breath hold duration 11 seconds not
4 dependent on heart rate). At 3.0T, sequence parameters include: acquired voxel size
5 2.0x2.0mm; TR/TE/flip angle 2.2ms/0.92ms/20°; SENSE acceleration factor 2, partial
6 Fourier factors kx/ky 0.63/0.85; receiver bandwidth 1086 Hz/px; acquisition in end-systole
7 (site 1) and end-diastole (site 2). At 1.5T, sequence parameters include: acquired voxel size
8 2.0x2.0mm; TR/TE/flip angle 2.4ms/1.1ms/35°; SENSE acceleration factor 2, partial Fourier
9 factors kx/ky 0.63/0.85; receiver bandwidth 1082 Hz/px; acquisition in end-diastole. No
10 motion correction was applied and MOLLI source images were checked to ensure a steady
11 breath hold position and myocardium alignment between source images. In cases of poor
12 breath holding, the slice/map was reacquired. A blood correction was not applied. ROIs over
13 the whole left ventricle myocardium were used at site 1 and over a left ventricle myocardium
14 of a mid-slice at site 2.

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The Shapiro-Wilk test was employed to provide evidence that sample data are likely drawn from a Normally distributed population. Outliers are identified using the Tukey robust approach with box-and-whisker plots displaying interquartile ranges (IQR: Q3-Q1; Q1: lower quartile, Q3: upper quartile); at levels of $< Q1 - 1.5 \text{ IQR}$ and / or $> Q3 + 1.5 \text{ IQR}$, the outliers are discarded (5, 6).

An ordered list of calculations which can be used to create a reference range is presented in Table 1. To facilitate the reader, various commands are identified for Microsoft Excel (2010 or later) and Google Sheets. Confidence intervals (CIs) are used to indicate a probability that values calculated from a random sample will include the population value of interest.

Calculated values are labelled in Table 1 and correspond to labels in Figure 1.

Thus, referring to Figure 1 and Table 1:

- the 95% reference range is from J to H

- the 90% confidence intervals around the reference range limits are R to Q, and P to N
- the mean value is A with a 90% confidence interval of U to T.

RESULTS

Site 1 at 3.0T Males and Females

Whole-heart ROI T1 data at 3 tesla for 21 normal subjects are listed here to allow repetition of calculations by the reader and were: 1245, 1243, 1299, 1261, 1243, 1245, 1144, 1224, 1133, 1186, 1179, 1213, 1202, 1244, 1220, 1174, 1212, 1211, 1157, 1233, 1245 ms. Visual inspection of the data using a box-and-whisker plot confirmed no outliers are present (see Figure 2).

Using the Shapiro-Wilk test, the calculated W statistic was 0.96, higher than the critical value of W of 0.91 (significance level 0.05), therefore the null hypothesis that the data are Normally distributed was not rejected. The data are likely consistent with the Normal distribution.

The 95% reference range is 1215 ± 86 ms. I.e. the 95% reference range is 1129 to 1301 ms.

The 90% CIs around the reference range limits are 1102 to 1156 ms and 1274 to 1328 ms.

These 90% CIs around the reference range limits each are 32% of the reference range size.

The confidence intervals around the reference range limits allow us to state with increased confidence that a patient's T1 value is outside the reference range for the population, if it is outside the range from the lower CI boundary of the lower end of the reference range to the upper CI boundary of the upper end of the reference range (i.e. R to N in Figure 1). For this data set this corresponds to values less than 1102 and more than 1328 ms.

The mean T1 in the normal subjects was 1215 ms with a 90% confidence interval of 1199 to 1230 ms (note that this is not the reference range).

Site 2 at 3.0T Males and Females

The box-and-whisker plot confirmed no outliers are present (see Figure 2).

Using the Shapiro-Wilk test, the calculated W statistic was 0.99, higher than the critical value of W of 0.96 (significance level 0.05), therefore the null hypothesis that the data are Normally distributed was not rejected. The data are likely consistent with the Normal distribution.

The 95% reference range is 1234 ± 75 ms. I.e., the 95% reference range is 1160 to 1309 ms.

The 90% CIs around the reference range limits are 1145 to 1174 ms and 1295 to 1323 ms.

These 90% CIs around the reference range limits each are 19% of the reference range size.

The range outside of which data may be more confidently placed outside the reference range for the population (i.e. R to N in Figure 1) for this data set this corresponds to 1145 to 1323 ms.

The mean T1 in the normal subjects was 1234 ms with a 90% confidence interval of 1226 to 1242 ms (note that this is not the reference range).

Site 2 at 1.5T Males

The box-and-whisker plots showed two outliers present (see Figure 3). Upon clinical review of the outlier subjects, the high-T1 outlier was excluded due to clinical confounders. Using the Tukey robust approach described above it was confirmed that both outliers were outside the range for excluding unexplained outliers for this data (920-1034ms), and therefore the low-T1 outlier was also omitted from the following results (i.e. updated n = 130).

Using the Shapiro-Wilk test, the calculated W statistic for the male T1 data is 0.99, higher than the critical value of W of 0.98 (significance level 0.05). Therefore, the null hypothesis

1
2
3 that the data are Normally distributed was not rejected. The data are likely consistent with the
4
5 Normal distribution.
6

7
8 The 95% reference range is 976 ± 44 ms. I.e., the 95% reference range is 933 to 1020 ms.
9

10
11 The 90% CIs around the reference range limits are 927 to 938 ms and 1014 to 1025 ms.
12

13
14 These 90% CIs around the reference range limits each are 13% of the reference range size.
15

16
17 The range outside of which data may be more confidently placed outside the reference range
18
19 for the population (i.e. R to N in Figure 1) for this data set this corresponds to 927 to
20
21 1025 ms.
22

23
24 The mean T1 in the normal subjects was 976 ms with a 90% confidence interval of 973 to
25
26 979 ms (note this is not the reference range).
27
28

29 30 ***Site 2 at 1.5T Females***

31
32 The box-and-whisker plot confirmed no outliers are present (see Figure 3).
33

34
35 Using the Shapiro-Wilk test, the calculated W statistic for female T1 data is 0.97, higher than
36
37 the critical value of W of 0.96 (significance level 0.05). Therefore the null hypothesis that the
38
39 data are Normally distributed was not rejected. The data are likely consistent with the Normal
40
41 distribution.
42
43

44
45 The 95% reference range is 1009 ± 44 ms. I.e., the 95% reference range is 965 to 1054 ms.
46
47

48
49 The 90% CIs around the reference range limits are 957 to 973 ms and 1046 to 1061 ms.
50

51
52 These 90% CIs around the reference range limits each are 18% of the reference range size.
53

54
55 The range outside of which data may be more confidently placed outside the reference range
56
57 for the population (i.e. R to N in Figure 1) for this data set this corresponds to 957 to
58
59 1061 ms.
60

1
2
3 The mean T1 in the normal subjects was 1009 ms with a 90% confidence interval of 1005 to
4
5 1014 ms (note this is not the reference range).
6
7

8 Comparison of the reference ranges can be made with Table 2.
9

10 11 **DISCUSSION**

12
13
14
15 Thorough technical validation of an imaging biomarker includes assessment of whether the
16
17 imaging biomarker measurement can be utilised in any geographical location and give
18
19 comparable data (2). This requires an understanding of its repeatability (over measurements
20
21 performed multiple times in the same subject using the same equipment, software and
22
23 operators over a short timeframe), reproducibility (over measurements performed using
24
25 different equipment, different software or operators, or at different sites and times, either in
26
27 the same or in different subjects), and bias (systematic difference between measurements of a
28
29 parameter and its real value). Even after minimizing all confounding technical and
30
31 physiological pitfalls that influence the myocardial T1 (7), high quality reference ranges are
32
33 still required to inform studies establishing clinical utility. The calculations and discussion
34
35 provided in this work will enable clinical researchers using myocardial T1 measurement to
36
37 create a robust local reference range and apply it carefully. It is not the intention of this work
38
39 to review published data of myocardial T1 values; such reviews may be found in the
40
41 literature (8). Nor is it proposed that the reference ranges reported in this work are ideal.
42
43 Variations in workflow for any range may change utility, e.g., variations in pulse sequence
44
45 used, whether motion correction is applied, how ROIs are placed including choice of
46
47 segmentation, imaging in systole or diastole, subgroup analysis by gender or application of
48
49 blood correction, *et cetera*. Rather, they illustrate that, whatever the end-to-end workflow of
50
51 data acquisition and analysis, any reference range can be properly characterised and its utility
52
53 more clearly understood, enabling appropriate use of a reference range in the clinical setting.
54
55
56
57
58
59
60

1
2
3 The increased precision of the reference range limits can be appreciated with a larger number
4
5 of subjects in our 3.0T data from two sites.
6
7

8 We use the term “reference range”, outside of which T1 would be suspected to be abnormal,
9
10 possibly affecting patient treatment. This term is equivalent to “reference interval”; we avoid
11
12 the alternative term “*normal range*” as it may be confused with the Normal distribution. The
13
14 reference population should be as similar as possible to the population on which the reference
15
16 range will be applied. In some settings it may be necessary for a subgroup, such as
17
18 older/younger age or male/female gender, to be considered with its own reference range;
19
20 statistical criteria have been described for biochemical biomarkers which can identify such
21
22 groups by focusing on the proportions of the subgroups outside the reference-range limits of
23
24 the entire population (5). Whether subgroups are required will also depend on the effect size
25
26 of the T1 change in disease. Subgroup reference ranges should not, however, be calculated
27
28 unless there is a strong statistical rationale, as this increases the total required number of
29
30 healthy participants, which must be counted separately in each subgroup.
31
32
33
34
35

36 For myocardial T1, data acquisition methods require attention. For example, in the case of
37
38 Modified Look-Locker Imaging (MOLLI) T1 mapping methods as used in this work,
39
40 particular attention should be paid to the MOLLI scheme; there are significant improvements
41
42 in the performance of newer schemes as used in cardiac T1 mapping product software such as
43
44 5s(3s)3s, versus the original MOLLI scheme 3b(3b)3b(3b)5b (9). However, the discussions
45
46 made in this work are not specific to MOLLI or to any specific MR system manufacturer;
47
48 whatever T1 mapping method is used, the scanning procedures used to acquire the reference
49
50 data should be performed in the same way when obtaining subsequent clinical data. Details
51
52 such as the heart phase used for data acquisition may be noted; in normal myocardium,
53
54 diastolic and systolic myocardial T1 values differ but correlate strongly, and blood correction
55
56 reduces variability and gender differences (10, 11).
57
58
59
60

1
2
3 Guidelines for recording region of interest (ROI) values from T1 maps are provided in the
4
5 2017 consensus statement (4). Where options are available, decisions should be recorded, so
6
7 that the same approach can be mandated in subsequent diagnostic work, for example whether
8
9 to draw an ROI on the septum only or over the whole myocardium. The mean pixel value of
10
11 each ROI is recorded, which is used in reference range calculation. The standard deviation of
12
13 individual ROIs would not generally be used in the calculation, but may also be recorded; if
14
15 the standard deviations are very large or variable, it may be necessary review certain
16
17 subjects' data.
18
19
20
21

22
23 The calculations which are described assume that the sample of healthy T1 values measured
24
25 are drawn from a population which is Normally distributed. Though a Normal distribution of
26
27 T1 values are expected, it is essential to check that the shape of the distribution of the sample
28
29 data is consistent with a Normal distribution. It is incorrect to use the calculations below
30
31 without preliminary verification or else an erroneous reference range might be created (5).
32
33 Various parametric tests can be used to provide evidence that sample data are likely drawn
34
35 from a Normally distributed population. The Shapiro-Wilk test is the most accurate test for
36
37 small sample sizes and is recommended, although a range of methods are available (12). We
38
39 are not aware of any previous findings of non-Normal distributions of healthy myocardial T1.
40
41 In the unexpected scenario that sample data are not Normally distributed (or cannot be made
42
43 Normal after transformation e.g. log, power, or some other function of the original scale),
44
45 non-parametric statistics would be required, which are not covered here. Such a finding
46
47 would have significant implications because although it simplifies calculation of the
48
49 reference range, many more samples (healthy subjects) would be necessary.
50
51
52
53
54

55
56 Outliers may be discarded for a variety of reasons (conditions of the subject, measurement
57
58 problems etc.). Nevertheless, sometimes outliers may occur with no such obvious reason for
59
60 their occurrence. Whilst the Normal distribution allows for extreme values, such values may

1
2
3 be removed if they are clinically implausible. Outliers can be identified from a visual
4
5 inspection of the data, for example using the straightforward box-and-whisker plot using the
6
7 ROI mean values (readily available in many spreadsheet applications and statistics
8
9 packages/languages as well as online). In ambiguous cases a statistical method is useful to
10
11 justify or rule out exclusion such as that described in this work.
12
13

14
15 A common reference range principle is to arbitrarily consider the central 95% of samples as
16
17 being within the reference range, though adoption of a different interval may be acceptable
18
19 with clinical justification. Here the standard 95% value is used.
20
21

22
23 The 2017 consensus statement (4) recommends calculation of the limits of the reference
24
25 range “with the upper and lower range of normal defined by the mean plus and minus 2
26
27 standard deviations of the normal data”. To contain 95% of all values in the population, the
28
29 advised standard deviation multiplier of 2 is approximate, and valid only for large sample
30
31 counts. The calculations described here provide a more accurate multiplier value using the
32
33 Student’s t-distribution which is appropriate when the sample size is small and the whole
34
35 population (not sample) standard deviation is unknown, as is the case in this context. Careful
36
37 reading of normative values reported in the literature is advised; the mean plus and minus
38
39 only one standard deviation may be reported.
40
41
42

43
44 A reference range is intended to characterise a population. The correct interpretation of the
45
46 reference range can have important consequences for patient care and should be done in
47
48 consultation with a professional statistician. A distinction should be made between such
49
50 descriptive statistics and clinical decision thresholds. A value within a reference range may
51
52 be “typical” but it does not indicate “healthy”. Similarly a value outside a reference range
53
54 may be atypical but may not necessarily indicate disease. It is useful to remember that it is
55
56 only an estimate of the true reference range which is, in general, unknowable, and hard
57
58
59
60

1
2
3 thresholds for “normal” and “abnormal” reflect our choices in calculations and not nature.

4
5 Clinical decision thresholds are based on diagnostic questions, are defined by consensus, and
6
7 are thresholds above or below which a specific medical decision is proposed (8). They are
8
9 often derived from Receiver Operating Characteristic (ROC) curves and predictive values
10
11 (13, 14). Where a reference range has wide confidence intervals around its reference range
12
13 limits due to limited sample size, it may be prudent to use three classes of values instead of a
14
15 standard normal/abnormal dichotomy, to manage the imprecision in the reference range limits
16
17 (15). For example, the three classes of values may be: normal, indicating values within the
18
19 reference range but more conservatively not including any values within the confidence
20
21 intervals of the reference range limits (i.e. Q to P, in Figure 1); indeterminate, indicating
22
23 values contained within the confidence intervals of the reference range limits (R to Q, and P
24
25 to N, in Figure 1); or abnormal, indicating values outside both the reference range and the
26
27 confidence intervals of the reference range limits ($<R$ and $>N$ in Figure 1, see also extremes of
28
29 90% CIs of reference range limits in Table 2). From a health-economic perspective, an
30
31 “indeterminate” finding represents a futile use of MRI: if the proportion of “indeterminate”
32
33 findings is unacceptably high this may suggest that the number of subjects used to determine
34
35 the reference range was too low.
36
37
38
39
40
41
42

43 The number of healthy subjects, or normal individuals (referred to MR without any abnormal
44
45 findings) required to create a reference range is not straightforward (16); it depends on the
46
47 precision desired, which in turn depends on the proposed clinical application and diseases
48
49 which may be encountered. Moreover, the number required also depends on the precision of
50
51 the measurement method which varies from one method to the next (which for myocardial T1
52
53 mapping includes variations such as patient setup, system type, sequence used, analysis steps,
54
55 etc). However some help can be given by asking a different question: How many samples are
56
57 needed such that the confidence intervals for the reference range limits are at most a
58
59
60

1
2
3 particular proportion (p , e.g. one-third) of the reference range size? In other words, a
4
5 judgement can be made about the precision of the limits of the reference range, compared to
6
7 the width of the reference range itself; this can be visualised with the help of Figure 1 looking
8
9 at R to Q (or P to N) versus the reference range. We may estimate the number of samples
10
11 needed by observing the following (cf 18E.1. in ref. 16). The standard error of the limit of the
12
13 reference range (SE_{lim} , result K in Table 1) may be approximated to $\sqrt{3s^2/n}$ where n is the
14
15 number of samples and s is the standard deviation. The confidence interval around the
16
17 reference range limit is $2t$ times SE_{lim} where t is the critical t value (result L in Table 1), and
18
19 the width of the reference range is $2ts$. Therefore, the proportion p which we wish to
20
21 calculate is $2tSE_{lim}/2ts$, which reduces to $p = \sqrt{(3/n)}$. Thus, the number of samples
22
23 required to yield an acceptable proportion of the reference range size for the confidence
24
25 intervals of the reference range limits is $n = 3/p^2$. Then, for example: if the acceptable
26
27 proportion of the reference range size for the confidence intervals of the reference range
28
29 limits is one quarter, the approximate number of samples required can be calculated using
30
31 $3/(0.25^2) = 48$.

32
33
34
35
36
37
38
39 The number of subjects required for the reference range can be related to the expected T1 of
40
41 myocardium for specific diseases, if the width of the reference interval is known or can be
42
43 approximately established. Similarly, the utility of an existing reference range can be
44
45 checked. For example: if a local myocardial T1 reference range at 1.5T happens to be 940 to
46
47 1060 ms, and in disease a hypothetical T1 of 920 ms is expected, in order to limit the
48
49 confidence intervals of the reference interval limits to 40 ms ($\pm(940-920)$) so that a value of
50
51 920 ms may be confidently interpreted as abnormal, the number of samples required can be
52
53 calculated using $3/((40/(1060-940))^2) = 27$.

54
55
56
57
58
59
60

1
2
3 For diseases in which the change in myocardial T1 is small (e.g. diseases causing diffuse
4 myocardial fibrosis), smaller CIs around the reference range limits are required, necessitating
5 a larger sample size when creating the reference range to be able to constructively classify
6 patients' T1 values.
7
8
9
10

11
12
13 There are challenges involved in harmonising reference ranges from different institutions or
14 from different equipment manufacturers. However, benchmarking local data against
15 published reference ranges is possible and recommended (4). The following method is
16 standardised practice in biochemistry (17). Twenty subjects' data (after removing outliers)
17 are collected representing the local apparently healthy population satisfying the population
18 criteria. Then, if no more than two of the 20 subjects fall outside the reference range, the
19 reference range can be adopted, at least provisionally (5). If three or more are outside the
20 reference range, the experiment should be repeated with another 20 subjects (after removing
21 outliers). Again, if no more than two of the 20 subjects fall outside the reference range, the
22 reference range can be adopted. If three or more are outside the reference range, the reference
23 range is probably not valid and a local reference range should be created. This benchmarking
24 verification method is valid where the population from which the reference values are
25 measured has a Normal distribution. It is statistically sound; the probability of rejecting the
26 reference range when in fact at least 95% of the whole local population do fall within the
27 reference range being tested is 5 to 7% (17). In our 3.0T data from site 1, it can be observed
28 only two data points fall outside of the 3.0T reference range from site 2, despite differences
29 in: specific scanner type; heart phase of data acquisition (end systole vs end diastole,
30 respectively); and ROI inclusion criteria (whole heart vs only mid-slice, respectively). Thus,
31 the benchmarking of the 3.0T reference ranges from site 1 against site 2 is successful.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1
2
3 reproducibility in hemodialysis patients (19); at different field strengths (20); evaluation of
4
5 different T1 mapping methods (21). It is advised that regularly repeated phantom-based
6
7 quality control is performed to periodically confirm that the status and stability of the MR
8
9 system have not changed significantly during the time between establishing a reference range
10
11 and clinical scanning (4). Detailed instruction on quality control in MRI including of
12
13 relaxometry, as well as discussion on appropriate phantom design, is available (22, 23).
14
15

16
17 A limitation of this study is the arbitrary numbers of normal subjects included from sites 1
18
19 and 2, according to the data available at the time. These numbers are not designed to produce
20
21 a reference range with a specific utility according to a particular clinical question.
22
23

24
25 Nevertheless, the reference ranges created show the steps to follow and considerations to
26
27 make, and may be used for benchmarking.
28
29

30
31 In order for imaging biomarkers to cross clinical translation gaps, validation of the imaging
32
33 biomarker must be robust. Myocardial T1 mapping in routine clinical practice still faces
34
35 challenges including the creation of a local reference range of values, or the validation of a
36
37 previously established and/or external reference range at the local site. The calculations and
38
39 discussions provided in this work will equip the reader to perform these tasks.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 *Acknowledgements:* We thank the Cardiovascular Medicine Research team at the University
4
5 of Aberdeen for acquisition of healthy volunteer data.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FOR PEER REVIEW ONLY

References

1. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Silver Spring (MD): Food and Drug Administration (US); 2016-. Glossary. 2016 Jan 28 [Updated 2018 May 2].
<https://www.ncbi.nlm.nih.gov/books/NBK338448/> accessed 28th September 2018
2. O'Connor JP, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol*. 2017;14:169-186.
3. Moon JC, Messroghli DR, Kellman P, et al. Myocardial T1 mapping and extracellular volume quantification: a Society for Cardiovascular Magnetic Resonance (SCMR) and CMR Working Group of the European Society of Cardiology consensus statement. *J Cardiovasc Magn Reson*. 2013;15:92.
4. Messroghli DR, Moon JC, Ferreira VM, et al. Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). *J Cardiovasc Magn Reson*. 2017;19:75.
5. Ceriotti F, Hinzmann R, Panteghini M. Reference intervals: the way forward. *Ann Clin Biochem*. 2009;46:8-17.
6. Ozarda Y. Reference intervals: Current status, recent developments and future considerations. *Biochimica Medica*. 2016;26:5-11.
7. Cameron D, Vassiliou VS, Higgins DM, Gatehouse PD. Towards accurate and precise T1 and extracellular volume mapping in the myocardium: a guide to current pitfalls and their solutions. *Magn Reson Mater Phy* 2018;31:143-163.

- 1
2
3 8. Puntmann VO, Peker E, Chandrashekar Y, Nagel E. T1 Mapping in Characterizing
4
5 Myocardial Disease: A Comprehensive Review. *Circ Res.* 2016;119:277-99.
6
7
- 8 9. Higgins DM and Moon JC. Review of T1 Mapping Methods: Comparative Effectiveness
9
10 Including Reproducibility Issues. *Curr Cardiovasc Imaging Rep* 2014;7:9252.
11
12
- 13 10. Reiter U, Reiter G, Dorr K, Greiser A, Maderthaner R, Fuchsjäger M. Normal diastolic
14
15 and systolic myocardial T1 values at 1.5-T MR imaging: correlations and blood
16
17 normalization. *Radiology.* 2014;271:365-72.
18
19
- 20 21 11. Nickander J, Lundin M, Abdula G, et al. Blood correction reduces variability and gender
22
23 differences in native myocardial T1 values at 1.5 T cardiovascular magnetic resonance - a
24
25 derivation/validation approach. *J Cardiovasc Magn Reson.* 2017;19:41.
26
27
- 28 29 12. Ghasemi A, Zahediasl S. Normality Tests for Statistical Analysis: A Guide for Non-
30
31 Statisticians. *Int J Endocrinol Metab* 2012;10:486-489.
32
33
- 34 35 13. Sikaris KA. Physiology and its Importance for Reference Intervals. *Clin Biochem Rev.*
36
37 2014;35:3–14.
38
- 39 40 14. Sikaris K. Application of the Stockholm hierarchy to defining the quality of reference
41
42 intervals and clinical decision limits. *Clin Biochem Rev.* 2012;33:141–8
43
44
- 45 46 15. Leslie WD, Greenberg ID. Reference range determination: the problem of small sample
47
48 sizes. *J Nucl Med.* 1991;32:2306-10.
49
- 50 51 16. Bland, M. *An Introduction to Medical Statistics.* 3rd Ed. New York: Oxford University
52
53 Press; 2001, 405 p.
54
- 55 56 17. CLSI document C28-A3c, 2008. CLSI. *Defining, Establishing, and Verifying Reference*
57
58 *Intervals in the Clinical Laboratory; Approved Guideline – Third Edition.* Wayne, PA:
59
60 Clinical Laboratory and Standards Institute.

- 1
2
3 18. Pica S, Sado DM, Maestrini V, et al. Reproducibility of native myocardial T1 mapping in
4 the assessment of Fabry disease and its role in early detection of cardiac involvement by
5 cardiovascular magnetic resonance. *J Cardiovasc Magn Reson*. 2014;16:99.
6
7
- 8
9
10 19. Graham-Brown MP, Rutherford E, Levelt E, et al. Native T1 mapping: inter-study, inter-
11 observer and inter-center reproducibility in hemodialysis patients. *J Cardiovasc Magn*
12 *Reson*. 2017;19:21.
13
14
- 15 20. Raman FS, Kawel-Boehm N, Gai N, et al. Modified look-locker inversion recovery T1
16 mapping indices: assessment of accuracy and reproducibility between magnetic resonance
17 scanners. *J Cardiovasc Magn Reson*. 2013;15:64.
18
19
- 20 21. Roujol S, Weingärtner S, Foppa M, et al. Accuracy, precision, and reproducibility of four
21 T1 mapping sequences: a head-to-head comparison of MOLLI, ShMOLLI, SASHA, and
22 SAPPHERE. *Radiology*. 2014;272:683-689.
23
24
- 25 22. Quality Control and Artefacts in Magnetic Resonance Imaging. Report 112, Institute of
26 Physics and Engineering in Medicine (IPEM). 1st Ed. York: IPEM; 2017, 222 p.
27
28
- 29 23. Captur G, Gatehouse P, Keenan KE, et al. A medical device-grade T1 and ECV phantom
30 for global T1 mapping quality assurance-the T1 Mapping and ECV Standardization in
31 cardiovascular magnetic resonance (TIMES) program. *J Cardiovasc Magn Reson*.
32 2016;18:58.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Tables

Value	Spreadsheet	Comment	Result label
Mean	AVERAGE(<range>)		A
Standard deviation	STDEV.S(<range>)	Appropriate for a sample	B
Sample size	COUNT(<range>)		C
Degrees of freedom	C-1		D
Significance level	1-0.95	One, minus proportion of values to be considered “normal”, usually 95%	E
Critical t value for reference interval	T.INV.2T(E,D)	Two-tailed	F
Reference interval tolerance	B*F	This result is one side of the reference range	G
Upper limit of 95% reference range	A+G		H
Lower limit of 95% reference range	A-G		J
Standard error of the tolerance	SQRT(B^2*((1/C)+(2/(D))))	Will be used for confidence interval around reference interval limits	K
Critical t value for reference interval limits	T.INV.2T(1-0.9,D)	Typically 90% CI is used	L
Reference range limits tolerance	L*K	This result is one side of the range-limit CI	M
Upper 90% CI limit for upper limit of reference range	H+M		N
Lower 90% CI limit for upper limit of reference range	H-M		P
Upper 90% CI limit for lower limit of reference range	J+M		Q
Lower 90% CI limit for lower limit of reference range	J-M		R
90% CI of the mean value	CONFIDENCE.T(1-0.9,B,C)	Using t distribution. Typically 90% CI is used	S
Mean upper 90% CI limit	A+S		T
Mean lower 90% CI limit	A-S		U

Table 1. Series of spreadsheet calculations required for creating a reference interval and describing it. Commands are for Microsoft Excel 2010 or later but other tools have similar commands. For example if using Google Sheets change only “STDEV.S” to “STDEV”, and “CONFIDENCE.T(1-0.9,B,C)” to “T.INV.2T(1-0.9,D)*B/SQRT(C)”.

	Reference range (ms)	Extremes of 90% CIs of reference range limits (ms)	90% CIs of reference range limits as proportion of the reference range	n (excl. outliers)
Site 1, 3.0T	1129 to 1301	1102 to 1328	32%	21
Site 2, 3.0T	1160 to 1309	1145 to 1323	19%	59
Site 2, 1.5T male	933 to 1020	927 to 1025	13%	130
Site 2, 1.5T female	965 to 1054	957 to 1061	18%	68

Table 2. Comparison of reference ranges.

FOR PEER REVIEW ONLY

Figure/Table legends

Figure 1. Illustrative relationship between various calculated values.

Figure 2. 3.0T data from both site 1 and site 2. Both data sets include males and females. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; no outliers are present; cross represents sample mean; data points are plotted as open circles. n = 21 (site 1) and n = 59 (site 2) sample points.

Figure 3. 1.5T data from site 2 only. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; cross represents sample mean; data points are plotted as open circles. Male n = 132 (two outliers are seen) and female n = 68 sample points.

Table 1. Series of spreadsheet calculations required for creating a reference interval and describing it. Commands are for Microsoft Excel 2010 or later but other tools have similar commands. For example if using Google Sheets change only “STDEV.S” to “STDEV”, and “CONFIDENCE.T(1-0.9,B,C)” to “T.INV.2T(1-0.9,D)*B/SQRT(C)”.

Table 2. Comparison of reference ranges.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

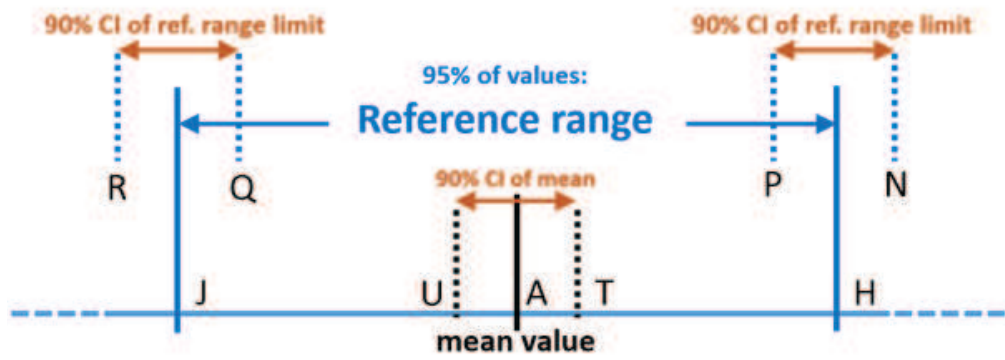


Figure 1: Illustrative relationship between various calculated values.

51x19mm (300 x 300 DPI)

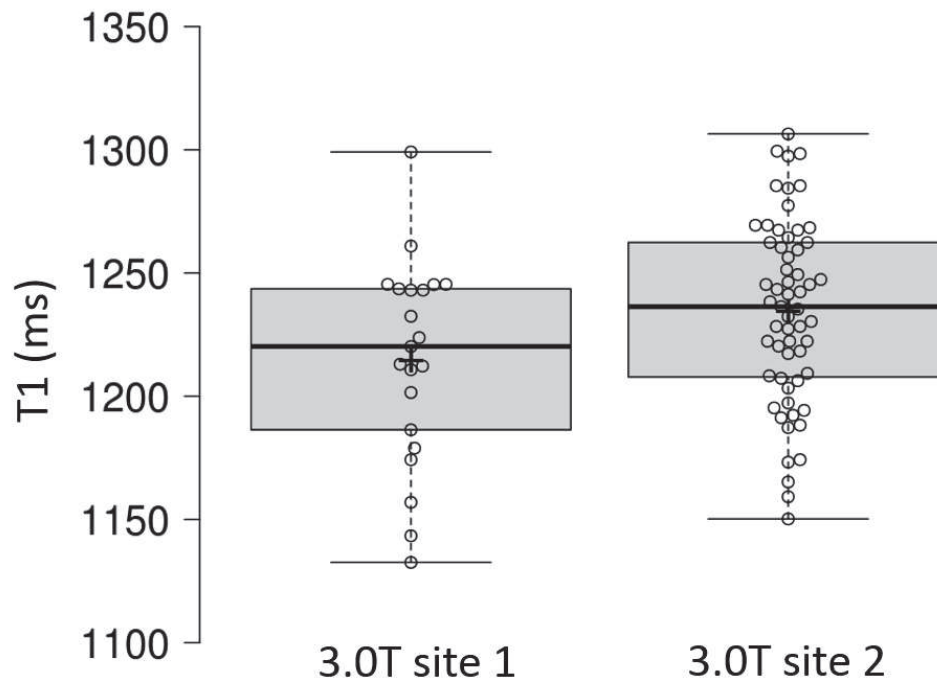


Figure 2. 3.0T data from both site 1 and site 2. Both data sets include males and females. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; no outliers are present; cross represents sample mean; data points are plotted as open circles. $n = 21$ (site 1) and $n = 59$ (site 2) sample points.

81x57mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

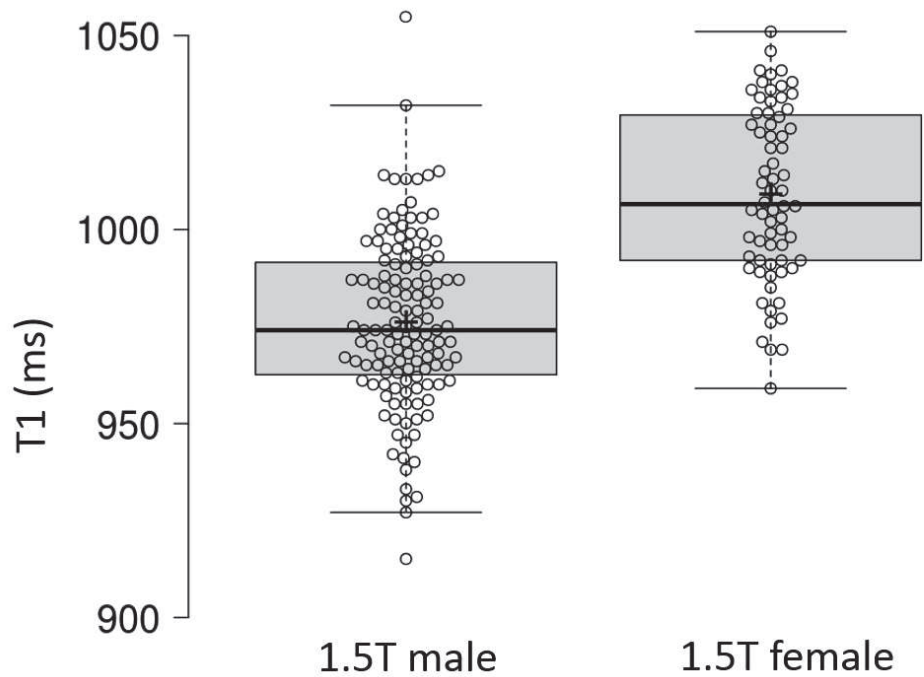


Figure 3. 1.5T data from site 2 only. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; cross represents sample mean; data points are plotted as open circles. Male n = 132 (two outliers are seen) and female n = 68 sample points.

85x59mm (300 x 300 DPI)