Standardized image post-processing of cardiovascular magnetic resonance T1-mapping reduces variability and improves accuracy and consistency in myocardial tissue characterization

Valentina Carapella*^{1,2}, Henrike Puchta¹, Elena Lukaschuk¹, Claudia Marini¹, Konrad Werys¹,
Stefan Neubauer¹, Vanessa M. Ferreira¹, Stefan K. Piechnik¹

1 Oxford Centre for Clinical Magnetic Resonance Research (OCMR), Radcliffe Department of Medicine, University of Oxford

2 Biomedical Engineering Department, King's College London, 5th Floor Becket House, London, United Kingdom

All authors take responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation

List of authors' email addresses:

Valentina Carapella (corresponding author): vcarapella@gmail.com

Henrike Puchta: henrike.puchta@cardiov.ox.ac.uk

Elena Lukaschuk: elena.lukaschuk@cardiov.ox.ac.uk

Claudia Marini: claudia.marini@outlook.com

Konrad Werys: konrad.werys@cardiov.ox.ac.uk

Stefan Neubauer: stefan.neubauer@cardiov.ox.ac.uk

Vanessa M. Ferreira: vanessa.ferreira@cardiov.ox.ac.uk

Stefan K. Piechnik: stefan.piechnik@cardiov.ox.ac.uk

Keywords: T1 mapping; cardiovascular magnetic resonance imaging; manual contouring; quality assurance; standardisation; image post-processing

Abstract¹

Background: Myocardial T1-mapping is increasingly used in multicentre studies and trials. Inconsistent image analysis introduces variability, hinders differentiation of diseases, and results in larger sample sizes. We present a systematic approach to standardize T1-map analysis by human operators to improve accuracy and consistency.

Methods: We developed a multi-step training program for T1-map post-processing. The training dataset contained 42 left ventricular (LV) short-axis T1-maps (normal and diseases; 1.5 and 3 Tesla). Contours drawn by two experienced human operators served as reference for myocardial T1 and wall thickness (WT). Trainees (n=26) underwent training and were evaluated by: (a) qualitative review of contours; (b) quantitative comparison with reference T1 and WT.

Results: The mean absolute difference between reference operators was 8.4 ± 6.3 ms (T1) and 1.2 ± 0.7 pixels (WT). Trainees' mean discrepancy from reference in T1 improved significantly post-training (from 8.1 ± 2.4 to 6.7 ± 1.4 ms; $p < 0.001$), with a 43% reduction in standard deviation (SD) ($p = 0.035$). WT also improved significantly post-training (from 0.9 ± 0.4 to 0.7 ± 0.2 pixels, $p = 0.036$), with 47% reduction in SD ($p = 0.04$). These experimentally-derived thresholds served to guide the training process: T1 (± 8 ms) and WT (± 1 pixel) from reference.

¹ **List of Abbreviations:** **CMR:** Cardiovascular Magnetic Resonance. **LGE:** Late Gadolinium Enhancement. **LV:** Left Ventricle. **OCMR:** Oxford Centre for Clinical Magnetic Resonance Research. **ROI:** Region of Interest. **RV:** Right Ventricle. **SD:** Standard Deviation. **Sh-MOLLI:** Shortened Look-Locker Inversion Recovery. **WT:** Wall Thickening.

1 **Conclusion:** A standardized approach to CMR T1-map image post-processing leads to significant
2 improvements in the accuracy and consistency of LV myocardial T1 values and wall thickness.
3 Improving consistency between operators can translate into 33-72% reduction in clinical trial
4 sample-sizes. This work may: (a) serve as a basis for re-certification for core-lab operators; (b)
5 translate to sample-size reductions for clinical studies; (c) produce better-quality training datasets
6 for machine learning.

7

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

1. Introduction

Cardiovascular magnetic resonance (CMR) T1-mapping is an emerging technique for advanced non-invasive myocardial tissue characterization. T1-mapping has diagnostic power at least comparable to late gadolinium enhancement (LGE) imaging, with potential to detect subtle pathology not readily discernible on LGE, such as diffuse myocardial fibrosis (1). There are a variety of T1-mapping techniques and approaches to utilize T1-maps for clinical applications. Efforts to standardize T1-mapping methods are ongoing, as seen in two mapping consensus papers published by the Society for Cardiovascular Magnetic Resonance (SCMR) (1, 2).

Over the last decade, robustness and reproducibility of T1 mapping results has been an active subject of research. Different aspects which could help reduce myocardial T1 variability have been investigated. Firstly, the influence of multiple centres, MR system vendors, magnetic field strength, observers and acquisition set-ups has been examined (3). Secondly, phantom-based quality assurance of T1 mapping data has been introduced (4, 5). Thirdly, the implementation of more consistent image analysis protocols has been explored but was limited to T1-map analysis of the septum only (6). Against the demonstrated benefit, this may limit the potential of T1-maps and the visual-spatial information that could aid clinical diagnosis and differentiation of disease entities beyond an average septal T1 value. In the aforementioned CMR mapping consensus papers (1, 2), the technical recommendations provided contain only a few general suggestions on image analysis (7). Thus far, there has been little effort on standardizing T1-map image post-processing, which is integral to fully utilizing T1-maps for clinical diagnosis.

The Oxford Centre for Clinical Magnetic Resonance Research (OCMR) has contributed a large body of quality original research in cardiac T1-mapping (8) as well as rigorous work towards the

1 standardisation of the analysis of CMR scans (9). Importantly, quality assurance of the image
2 quality of T1-maps has always been paramount in our T1-mapping methods (10-20). In a previous
3 study towards the definition of a CMR image analysis protocol, we have found a strong
4 dependency of average T1 values on partial volume, and the impact of contouring methods on
5 inter and intra-study variability (21). Thus, standardization of cardiac T1-map analysis is key to
6 control the heterogeneity in inter-operator differences and other biases introduced by untrained
7 human operators, which, in turn, may impact on the variability of T1-maps and their ability to
8 diagnose disease. Reliable, good quality annotations data, achieved by thorough standardisation
9 of all stages from acquisition to post-processing, are critical for artificial intelligence approaches.
10 The aim of this work is to introduce and describe the impact of a standardized training program for
11 human operators on the variability, consistency and accuracy of cardiac T1 parameters derived
12 from T1-maps.

13

14 **2. Methods**

15

16 **2.1 Training dataset.** The training dataset consists of fully anonymized T1-maps from research
17 participants previously enrolled in an ethically-approved study (10). The dataset contains 42 single
18 mid-ventricular short-axis native T1-maps acquired at OCMR using the established Shortened
19 Look-Locker Inversion Recovery (ShMOLLI) T1-mapping sequence (22) at 1.5T and 3T. Data
20 included 21 images from healthy volunteers and 21 patients with acute myocardial injury. The
21 images chosen were largely free of artefacts, as their detection and appropriate processing
22 requires more advanced training. During the training process, the trainee is exposed to the same
23 dataset (D) multiple times, referred to as D0, D1, D2, etc.

24

25 **2.2 Post-processing software.** T1-maps were analysed using dedicated software MC-ROI
26 (MyoCardial Regions Of Interest; programmed by SKP in Interactive Data Language, version 6.1,
27 Exelis Visual Information Solutions, Boulder, Colorado, USA), previously validated and employed

1 in a number of published studies (8, 10-22). The software can be made available to collaborators
2 upon request for research purposes, and training can be provided by experienced OCMR Core Lab
3 members by a separate arrangement. The core of the contouring feature is a semi-automatic
4 algorithm that initializes the myocardial outlines (i.e. both endocardial and epicardial contours of
5 the left ventricle), as shown in the left panel of Figure 1. It requires the manual identification of the
6 anterior RV-LV insertion point and an extent of the ventricle, to which the myocardial ROI is fitted
7 based on internally developed heuristic approaches. The contours placed by the software are
8 intended to provide a guideline of an arbitrary cut through a partial volume between the
9 myocardium and neighbouring tissues. The outlines are subjected to human assessment and
10 manual correction, often required in the presence of artefacts, pathology or in low quality scans.

11

12 The trainees were taught to manually correct significant deviations of the automatic contours;
13 otherwise, they were advised to retain the automatic contours as much as possible to reduce
14 subjective bias. During this process of manual correction, operators may either use a free-hand
15 drawing tool or a nudge-ball to optimise the contour consistency in view of several possible
16 presentation methods. A representative T1-map with several display options used for validation is
17 shown in the right panel of Figure 1.

18

19

20 **2.3 Reference dataset.** Two operators (Exp1 and Exp2) familiar with the MC-ROI software and its
21 use for analysis of cardiac T1-maps contoured the dataset to provide reference values of T1 and
22 wall thickness (WT). Exp1 is a senior clinical researcher in cardiology, while Exp2 is a senior
23 researcher with a physical sciences background.

24

25 **2.4 Trainees.** Training of operators for analysing T1-maps is carried out on a regular basis at
26 OCMR. Therefore, a cut-off of 26 consecutive observers was chosen to summarise the results and
27 finalise the training protocol. All trainees worked in OCMR at the time of training and were

competent PC users with different educational backgrounds, including clinical, biomedical, engineering and physical sciences specialties. All of them were involved in cardiovascular research and underwent similar training steps according to the developed training protocol.

2.5 Training program. The training is subdivided into three main sessions with an iterative approach, as detailed in Figure 2. The novice operator was permitted to take their time to feel comfortable with the dataset, without imposed time limits or potentially perceived added pressure by formal timing of their image analysis duration. For reference, an experienced analyst can typically process the whole dataset in a single session in approximately two hours.

Phase I - "Introductory training". The trainee is supervised by an experienced operator and receives a verbal introduction as well as a reference manual on the use of the software. The full training dataset of 42 scans is presented for the first time to the trainee, who would attempt the analysis of up to 5 scans under direct supervision. The main instruction at this stage is to make minimal corrections to the automatic LV myocardial contours, unless the software has visibly included extra-myocardial structures or tissue, or excluded significant parts of the LV myocardium. The aim is to maximise consistency between trainees by minimising the impact of individual contouring habits.

Phase II - "Contouring practice". This is a session with minimal supervision in which the trainee analyses the same training dataset for at least three times and up to four. The first two repeats of the dataset analysis (D1 and D2) are performed on consecutive days. Then, D3 is analysed a few days later to limit the risk of the trainee recognising the image and re-contouring to memory. Further repeats (e.g. D4 and so on) are optional, depending on the trainee's performance.

Phase III - "Assessment and feedback". Assessment of the results took place at the OCMR Core Lab, with at least one experienced operator for critical review. The overall performance of the

1 trainee was evaluated both qualitatively by examining the contours, and quantitatively by
2 comparing the resulting average myocardial T1 and WT with the reference dataset. Intra- and inter-
3 dataset variability were quantified and discussed. If the contours at the final stage of the training
4 were found to be consistently incorrect (e.g. including large portions of extra-cardiac tissue, such
5 as blood pool or pericardial fat), then the trainee was required to perform additional contouring on
6 at least one more dataset (D4) until significant trends and deviations from the reference values
7 were eliminated (see results below for examples).

8
9 **2.6 Statistical methods.** We analysed the average LV myocardial T1 values (ms) and the
10 average WT (pixels) for each T1 map, as previously published (22). The choice of pixels, rather
11 than millimetres (mm), to present WT values was due to slight differences in the image resolution
12 across the dataset (median=0.885, range 0.885-1.042 mm/pixel). Internal validity of the reference
13 dataset was assessed by the Bland-Altman analysis for T1 and WT, separately. The overall effect
14 of training was measured by comparing the average departure from the reference values for T1
15 and WT at the first and final training session, and the statistical significance of this difference was
16 evaluated by the Wilcoxon signed rank test. The statistical significance of the reduction of
17 variability after training among the group of trainees was tested with Levene's test for the
18 homogeneity of variance (23).

20 **3. Results**

21
22 **3.1 Consistency between the reference operators.** Prior to evaluating the training outcome for
23 all trainees, the two reference operators were tested for internal consistency. A Bland-Altman
24 analysis for LV myocardial T1 and WT showed the presence of bias (-5.4 ms in T1, and -1.2 pixels
25 in WT, respectively). This bias was deliberately not corrected to allow for reasonable inter-
26 individual variations in contouring approaches. Between the two reference operators, the mean
27 absolute difference over the whole dataset for LV myocardial T1 was 8.4 ± 6.3 ms (95% CI -23.3 to

1 12.6 ms) and for WT was 1.2 ± 0.7 pixels (95% CI -2.6 to 0.13 pixels). Intra-operator variability
2 analysis on this training dataset had been previously reported for the same reference operators
3 (10).

4
5 **3.2 Evaluation of training.** The group of 26 trainees who underwent Phase I training showed a
6 range of approaches to contouring, or contouring habits. Figure 3 (top row) provides an example
7 for Trainee A, with an initial tendency to consistently underestimate the LV wall thickness of the
8 myocardial ring. In contrast, Trainee B (Figure 3 bottom row) tended to include some partial
9 volume both at the endocardial and epicardial borders. The right panels of Figure 3 show a
10 summary plot for trainees A, B of the average difference compared to the reference dataset over
11 the whole dataset.

12
13 As shown in the right top panel, Trainee A had a consistent contouring pattern through the repeats,
14 underestimating the average T1 value compared to the reference values. Feedback and additional
15 contouring of two more datasets (D5 and D6) were sufficient to correct the significant
16 disagreement, although a tendency to deviate from training is suggested by the result in the latest
17 repeat (D6). On the contrary, trainee B corrected their own habit such that, upon Phase III – “Visual
18 Assessment”, no further training was required to achieve an accepted level of agreement with the
19 reference values.

20
21 It should be noted that the evaluation of the overall training performance is guided by a
22 combination of qualitative visual inspection of the myocardial contours, and quantitative
23 comparison of the average discrepancy of T1 values and wall thickness from the reference values,
24 especially in the developmental phase of the training protocol.

25
26 The variability in mean LV myocardial T1 values and wall thickness derived from contoured T1-
27 maps before and after training is summarized below:

1

2 1. **Reference operators** (n=2): the mean absolute difference (SD) for T1 was 8.4 (6.3) ms, for

3 WT was 1.2 (0.7) pixels.

4 2. **Trainees before training** (n=26): the mean absolute difference (SD) from reference for T1

5 was 8.1 (2.4) ms, for WT was 0.9 (0.4) pixels.

6 3. **Trainees after training** (n=26): the mean absolute difference (SD) from reference for T1

7 was 6.7 (1.4) ms, for WT was 0.7 (0.2) pixels.

8 4. **The percentage reduction in mean** absolute difference after training was 18% for T1 and

9 24% for WT ($p < 0.05$ by the Wilcoxon test);

10 5. **The percentage reduction in variance** after training was 43% for T1 and 47% for WT

11 ($p < 0.05$ by the Levene's test).

12

13 Overall, training significantly improved the estimation of mean LV myocardial T1 values and wall

14 thickness, with significant reductions in standard deviations in these measurements, and less

15 discrepancy when compared to the reference values.

16

17 The overall impact of training between the first and last dataset analysed is also summarized in the

18 top panel of Figure 4. The red dashed lines mark the range of variation achieved after training,

19 rounded to the nearest integer. Training reduced the average discrepancy from the reference

20 dataset in the trainees for both T1 and WT. At the end of the training, the ranges of variation

21 achieved were ± 8 ms (T1) and ± 1 pixel (WT), consistent with the level of observed difference

22 between expert operators. The lower-right panel of Figure 4 is a Bland-Altman plot comparing the

23 deviation from reference before and after training. There was a bias of -2.1 ms between the two

24 sets of measurements, indicating that after training, the discrepancy with reference was

25 consistently lower. Additionally, the coefficient of variation (as SD/Mean %) was 29.6% before

26 training, improving to 21.0 % after training (lower-left panel, Figure 4).

27

4. Discussion

To our knowledge, this is the first report to demonstrate the impact of standardized training of human operators on the manual post-processing of CMR T1-maps and consistency in quantification of T1 biomarkers. We have: (1) developed and tested a multi-step iterative training program alternating supervised and unsupervised training, aiming towards minimising bias and a pro-active learning approach in the trainees; (2) quantified the impact of training on the variability of average T1 and wall thickness outcomes based on a large sample size of trainees ($n=26$) which provided a rich dataset (9, 24); (3) obtained experimentally-derived numerical thresholds to assess acceptable degrees of inter-operator variability in average T1 and WT that can be used to evaluate the training outcome of Core Lab image analysts for certification and re-validation procedures.

Currently, manual analysis and human input are a critical part of CMR T1-map image post-processing, and training of operators adds significant value in improving the accuracy and consistency of T1 biomarkers. When determining sample size required for clinical studies or trials, reproducibility of a marker has a direct relation to sample size by a square function, with supra-proportionally larger sample sizes required to compensate for any mechanism that increases intra- or inter-operator variation (25). The impact of training on reduction of variability in T1 biomarkers was demonstrated in this study to range from 18-47%. Sample sizes are proportional to the square of the underlying data variation; hence these numbers translate into potential reductions in the material required for detecting the same statistical signal by as much as 33-72% in clinical trials. A similar conclusion may be derived from the improved consistency in myocardial thickness from 0.8 mm pre-training, to 0.6 mm post-training, i.e. with an average reduction of 0.2 mm (given the average image resolution being 0.908 mm per pixel across the studied data).

It is possible that standardized manual labelling of imaging datasets will help the quality of training in machine learning applications to automate mundane image analysis tasks. While further improvements are achievable with longer training periods for MR image post-processing (9), we

1 feel that further optimization of the operator alignment may not be time- or cost-effective to achieve
2 beyond the inter-expert differences observed in this study. It is important to note that the reported
3 overall scan-rescan reproducibility of LV myocardial T1 estimates is around 2% due to various
4 other factors (21), and, thus, improvements in the consistency of T1-map post-processing are likely
5 not resource-effective beyond this level.

6
7 Other than reducing the inter-rater variability in the average LV myocardial T1 and wall thickness,
8 the training program also provides a systematically tested, standardized approach to T1-map post-
9 processing, with clear advantages for the clinical researchers. The multi-step iterative training
10 program provides the trainee with initial instructions and human supervision (Phase I), and hinges
11 on the inherent human ability to distinguish anatomical structures to follow and contour their
12 outlines. Repeats of unsupervised image analysis (Phase II) permit examination and quantification
13 of any operator trends developed during the training process, as well as the opportunity for
14 correction when necessary. Feedback to operators of their contouring habits is fundamental for the
15 learning process of consistent and accurate contouring. Through discussion of the training results,
16 the trainee receives feedback, develops a critical approach to contouring, gains experience, and is
17 then better prepared for more advanced image analysis intended to ultimately aid clinical diagnosis
18 and decision-making.

19 For instance, Trainee A (Figure 3, top row) initially placed very conservative LV myocardial
20 contours on the T1-map; while this may provide “core” myocardial T1 values within the myocardial
21 ring, a T1-map contoured this way may miss potential sub-endocardial myocardial infarctions, sub-
22 epicardial lesions in myocarditis, or fail to fully characterize a heart with hypertrophic
23 cardiomyopathy. Accurately determining the wall thickness and delineating myocardial borders
24 from extra-myocardial tissues are important for fully utilizing myocardial T1-maps for various
25 disease presentations that require consistent characterization.

26

1 An important outcome of our work is the experimentally-derived numerical thresholds to assess
2 acceptable degrees of inter-operator variability in average T1 and WT that can be used to evaluate
3 the training outcomes of new trainees. We believe that the ranges of variability obtained are robust
4 and reliable, providing a solid platform to further develop a revalidation or re-certification process
5 for core lab image analysts at regular intervals. Drifts detected on regular interval revalidation may
6 prompt the need for refresher training procedures. This is highly relevant when CMR core labs
7 require appropriate regulatory approvals (e.g. FDA in the USA, CE Mark in Europe) and
8 certifications (e.g. ISO) for quantitative image analyses, especially for participation in clinical trials.

9

10 **4.1 Limitations**

11 Firstly, the current training process requires a large qualitative component of visual inspection of
12 myocardial contours by an experienced image analyst. While the authors believe that it is a
13 cardinal part of any training, the use of quantitative approaches (e.g. analysis of training on a per-
14 myocardial segment basis, pixel-wise comparison of contours) would increase the efficiency and
15 standardisation of the training; this is based on our findings that contours that were largely
16 incorrect can, through training and revision, achieve numerical measures for myocardial T1 and
17 WT that are closer to the reference values. Secondly, the training dataset was based on mid-
18 ventricular slices of native T1-maps acquired with a single T1-mapping technique in a single-
19 centre, analysed using a single software. While the basic principle of the proposed training
20 program should be translatable, the improvements in variability, consistency and accuracy may
21 need testing for other imaging planes (e.g. basal, apical, long-axis), and other T1-mapping
22 variants, including post-contrast T1-maps. The training dataset was carefully selected to be largely
23 free from artefacts to limit their impact on the initial learning experience, which focussed on
24 contouring; we are developing an advanced training programme dedicated to labelling and
25 handling of T1-map artefacts. The impact of training on speed of image analysis was not assessed
26 formally, although we observed a reduction of image processing time from some 6-8 hours initially,
27 down to 1-2 hours at completion of training. Comparison of different software tools with varied

1 approaches to streamline contour drawing, including both commercial and academic sources,
2 would help demonstrate feasibility in standardisation of T1-map image processing across software
3 platforms. Finally, this is a single-centre study, whereas a multi-centre approach would increase
4 the validity of our results; extension of the programme to other collaborating centres is an integral
5 part of our future work plan.

6

7 **4.2 Implications and future plans**

8 To generalise our findings, the image post-processing of T1 mapping would greatly benefit from a
9 more integrated effort across centres to address the standardisation of software tools, image
10 analysis protocols and training. Training and harmonisation of T1 mapping analysis will optimise
11 research outputs and comparability of clinical studies. Importantly, sample sizes may be
12 significantly reduced to detect treatment effects or other clinically-relevant endpoints, since there
13 would be less variability in the imaging biomarker (26). Moreover, highly trained and consistent
14 image analysts are needed more than ever in an era of increasing automation of post-processing,
15 since they lay the foundation of training material for traditional machine learning and deep learning
16 methods (27). Even though the first applications of automatic contouring on large CMR cine
17 imaging databases has had limited success when compared to human processing (28), automatic
18 contouring still holds promise to significantly accelerate the analysis of CMR images, with better
19 consistency and accuracy, especially with evolving deep learning methods. Well-designed and
20 validated training methods will allow for the development of uniform, semi-automated and
21 automated contouring approaches to address big data challenges presented by, for example, the
22 HCMR registry (>20,000 T1-maps from 2750 patients) (5) and the UK Biobank imaging component
23 (100,000 CMR datasets including native T1-maps) (29). Hence, the harmonisation of human-driven
24 post-processing in image analysis is critical for the future of CMR T1-mapping applications for
25 clinical use, before the machines can take full responsibility for this process.

26

27

1 **5. Conclusion**

2 A standardized approach to CMR T1-map image post-processing leads to significant
3 improvements in the accuracy and consistency of LV myocardial T1 values and wall thickness.
4 This work may serve as the basis for re-certification of core lab image analysts to ensure
5 adherence to standard operating procedures, quality and consistency of image analysis outputs.
6 Standardized T1-map analyses may translate to significant sample-size reductions for clinical
7 studies and trials, better-quality reference datasets for machine learning, and improved clinical
8 applications.

10 **Declarations**

11 **Ethics approval and consent to participate**

12 The training dataset used in this study consists of fully anonymized T1-maps from research
13 participants previously enrolled in research studies conducted at the University of Oxford Centre
14 for Clinical Magnetic Resonance Research (OCMR) approved by local ethics committees. All
15 subjects gave written informed consent to the use of their de-identified data.

17 **Consent for publication**

18 Not applicable.

20 **Availability of data and material**

21 The software can be made available to collaborators upon request for research purposes, and
22 training can be provided by experienced OCMR Core Lab members by a separate arrangement.

24 **Funding**

25 This study was supported in part by the National Institute for Health Research (NIHR) Oxford
26 Biomedical Research Centre at The Oxford University Hospitals NHS Foundations Trust,
27 University of Oxford, UK. SKP, VMF and SN acknowledge support from the British Heart

1 Foundation Centre of Research Excellence, Oxford. VMF is supported by the British Heart
2 Foundation.

3

4 **Author's contributions**

5 All authors made appropriate contributions according to the ICMJE guidance, and as such have
6 read and approved the final manuscript. All authors take public responsibility for appropriate
7 portions of the manuscript content and agree to be accountable in ensuring that questions related
8 to the accuracy or integrity of the work are appropriately investigated and resolved. VC, HP, EL,
9 CM, KW, VMF and SKP provided their training data for the analysis. Additional author-specific
10 contributions are as follow: VC, data collation and analysis, preparation of the manuscript. HP, data
11 collation and analysis. EL, data analysis & supervision of training. CM, initial concept and testing.
12 KW, contribution of trainee subjects. VMF, contributed the scan data used for training and original
13 training approaches, and edited the draft manuscripts. SN, contributed to final stages of manuscript
14 revisions. SKP, study conception and design of the training methods, data analysis and
15 interpretation, study supervision and manuscript revisions.

16

17 **Acknowledgements**

18 The authors would like to acknowledge all the trainees whose training results have been analysed
19 for the purpose of this paper.

20

21 **Competing interests**

22 SKP has patent authorship rights for U.S. patent 9,285,446 B2. Systems and methods Sh-MOLLI
23 cardiac gated mapping of T1. Granted March 15, 2016. IP is managed by Oxford University
24 Innovations; the licence exclusively transferred to Siemens Healthcare. All other authors have no
25 relationships relevant to the contents of this paper to disclose.

26

27 **References**

- 1 1. Moon JC, Messroghli DR, Kellman P et al. Myocardial T1 mapping and extracellular volume
2 quantification: a Society for Cardiovascular Magnetic Resonance (SCMR) and CMR Working
3 Group of the European Society of Cardiology consensus statement. *J Cardiovasc Magn Reson.*
4 2013;15:92.
- 5 2. Schulz-Menger J, Bluemke DA, Bremerich J et al. Standardized image interpretation and
6 post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic
7 Resonance (SCMR) board of trustees task force on standardized post processing. *J Cardiovasc*
8 *Magn Reson.* 2013;15:35.
- 9 3. Graham-Brown MP, Rutherford E, Levelt E et al. Native T1 mapping: inter-study, inter-
10 observer and inter-center reproducibility in hemodialysis patients. *J Cardiovasc Magn Reson.*
11 2017;19(1):21.
- 12 4. Captur G, Gatehouse P, Keenan KE et al. A medical device-grade T1 and ECV phantom
13 for global T1 mapping quality assurance—the T1 Mapping and ECV Standardization in
14 cardiovascular magnetic resonance (T1MES) program. *Journal of Cardiovascular Magnetic*
15 *Resonance.* 2016;18(1):58.
- 16 5. Kramer CM, Appelbaum E, Desai MY et al. Hypertrophic Cardiomyopathy Registry: The
17 rationale and design of an international, observational study of hypertrophic cardiomyopathy. *Am*
18 *Heart J.* 2015;170(2):223-30.
- 19 6. Rogers T, Dabir D, Mahmoud I et al. Standardization of T1 measurements with MOLLI in
20 differentiation between health and disease--the ConSept study. *J Cardiovasc Magn Reson.*
21 2013;15:78.
- 22 7. Messroghli DR, Moon JC, Ferreira VM et al. Clinical recommendations for cardiovascular
23 magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by
24 the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European
25 Association for Cardiovascular Imaging (EACVI). *J Cardiovasc Magn Reson.* 2017;19(1):75.
- 26 8. Liu JM, Liu A, Leal J et al. Measurement of myocardial native T1 in cardiovascular diseases
27 and norm in 1291 subjects. *J Cardiovasc Magn Reson.* 2017;19(1):74.
- 28 9. Karamitsos TD, Hudsmith LE, Selvanayagam JB, Neubauer S, Francis JM. Operator
29 induced variability in left ventricular measurements with cardiovascular magnetic resonance is
30 improved after training. *J Cardiovasc Magn Reson.* 2007;9(5):777-83.
- 31 10. Ferreira VM, Piechnik SK, Dall'Armellina E et al. Non-contrast T1-mapping detects acute
32 myocardial edema with high diagnostic accuracy: a comparison to T2-weighted cardiovascular
33 magnetic resonance. *J Cardiovasc Magn Reson.* 2012;14:42.
- 34 11. Bull S, White SK, Piechnik SK, Flett AS et al. Human non-contrast T1 values and
35 correlation with histology in diffuse fibrosis. *Heart.* 2013;99(13):932-7.
- 36 12. Dall'Armellina E, Piechnik SK, Ferreira VM et al. Cardiovascular magnetic resonance by
37 non contrast T1-mapping allows assessment of severity of injury in acute myocardial infarction. *J*
38 *Cardiovasc Magn Reson.* 2012;14:15.
- 39 13. Ferreira VM, Marcelino M, Piechnik SK et al. Pheochromocytoma Is Characterized by
40 Catecholamine-Mediated Myocarditis, Focal and Diffuse Myocardial Fibrosis, and Myocardial
41 Dysfunction. *J Am Coll Cardiol.* 2016;67(20):2364-74.
- 42 14. Ferreira VM, Piechnik SK, Dall'Armellina E et al. T(1) mapping for the diagnosis of acute
43 myocarditis using CMR: comparison to T2-weighted and late gadolinium enhanced imaging. *JACC*
44 *Cardiovasc Imaging.* 2013;6(10):1048-58.
- 45 15. Ferreira VM, Piechnik SK, Dall'Armellina E et al. Native T1-mapping detects the location,
46 extent and patterns of acute myocarditis without the need for gadolinium contrast agents. *J*
47 *Cardiovasc Magn Reson.* 2014;16:36.
- 48 16. Levelt E, Piechnik SK, Liu A et al. Adenosine stress CMR T1-mapping detects early
49 microvascular dysfunction in patients with type 2 diabetes mellitus without obstructive coronary
50 artery disease. *J Cardiovasc Magn Reson.* 2017;19(1):81.
- 51 17. Liu D, Borlotti A, Viliani D et al. CMR Native T1 Mapping Allows Differentiation of
52 Reversible Versus Irreversible Myocardial Damage in ST-Segment-Elevation Myocardial Infarction:
53 An OxAMI Study (Oxford Acute Myocardial Infarction). *Circ Cardiovasc Imaging.* 2017;10(8).

- 1 18. Mahmod M, Piechnik SK, Levelt E et al. Adenosine stress native T1 mapping in severe
2 aortic stenosis: evidence for a role of the intravascular compartment on myocardial T1 values. J
3 Cardiovasc Magn Reson. 2014;16:92.
- 4 19. Ntusi NA, Piechnik SK, Francis JM et al. Subclinical myocardial inflammation and diffuse
5 fibrosis are common in systemic sclerosis--a clinical study using myocardial T1-mapping and
6 extracellular volume quantification. J Cardiovasc Magn Reson. 2014;16:21.
- 7 20. Ntusi NAB, Piechnik SK, Francis JM et al. Diffuse Myocardial Fibrosis and Inflammation in
8 Rheumatoid Arthritis: Insights From CMR T1 Mapping. JACC Cardiovasc Imaging. 2015;8(5):526-
9 36.
- 10 21. Piechnik SK, Ferreira VM, Lewandowski AJ et al. Normal variation of magnetic resonance
11 T1 relaxation times in the human population at 1.5 T using ShMOLLI. J Cardiovasc Magn Reson.
12 2013;15:13.
- 13 22. Piechnik SK, Ferreira VM, Dall'Armellina E et al. Shortened Modified Look-Locker Inversion
14 recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat
15 breathhold. J Cardiovasc Magn Reson. 2010;12:69.
- 16 23. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed.
17 Malden, MA: Blackwell Science; 2001. xi, 817 p.
- 18 24. Suinesiaputra A, Cowan BR, Al-Agamy AO et al. A collaborative resource to build
19 consensus for automated left ventricular segmentation of cardiac MR images. Med Image Anal.
20 2014;18(1):50-62.
- 21 25. Liu S, Han J, Nacif MS, Jones J et al. Diffuse myocardial fibrosis evaluation using cardiac
22 magnetic resonance T1 mapping: sample size considerations for clinical trials. J Cardiovasc Magn
23 Reson. 2012;14:90.
- 24 26. Bellenger NG, Davies LC, Francis JM, Coats AJ, Pennell DJ. Reduction in sample size for
25 studies of remodeling in heart failure by the use of cardiovascular magnetic resonance. J
26 Cardiovasc Magn Reson. 2000;2(4):271-8.
- 27 27. Valindria VV, Lavdas I, Bai W et al. Reverse Classification Accuracy: Predicting
28 Segmentation Performance in the Absence of Ground Truth. IEEE Trans Med Imaging.
29 2017;36(8):1597-606.
- 30 28. Suinesiaputra A, Sanghvi MM, Aung N et al. Fully-automated left ventricular mass and
31 volume MRI analysis in the UK Biobank population cohort: evaluation of initial results. Int J
32 Cardiovasc Imaging. 2018;34(2):281-91.
- 33 29. Petersen SE, Matthews PM, Francis JM et al. UK Biobank's cardiovascular magnetic
34 resonance protocol. J Cardiovasc Magn Reson. 2016;18:8.
- 35 30. Carapella V, Jiménez-Ruiz E, Lukaschuk E et al. Towards the semantic enrichment of free-
36 text annotation of image quality assessment for UK Biobank cardiac Cine MRI scans. Deep
37 Learning and Data Labeling for Medical Applications: Springer, Cham; 2016. p. 238-48.
- 38 31. Hann E, Ferreira VM, Neubauer S, Piechnik SK. Deep Learning for Fully Automatic
39 Contouring of the Left Ventricle in Cardiac T1 Mapping CMR; 2018; Barcelona, Spain.

List of Figures

Figure 1 Example of use of MC-ROI interface. On the left panel (I): MC-ROI contour initialisation by means of semi-automatic myocardial outlines (blue and magenta dotted lines). The white dotted lines show the 6-segment subdivision of the left ventricular ring, computed from the manual positioning of the anterior RV-LV. Added red arrows signify major deviations from the desired myocardial contour just on the edge of the myocardium (green arrows). Analysts were tasked with smoothing the major shape departures and noise and match the prevalent shape and contrast of the outlines. **On the right panel (II),** additional pop-up panel in MC-ROI used to validate the consistency of the myocardial outlines against several chosen display modes. Note that contours have been corrected by the operator with respect to the initial outlines shown in the left panel. (a) Enlarged neighbourhood of the analysed T1 map in the conventional ShMOLLI colour scheme. (b) Outlier map to indicate areas with T1 pixels outside 3SD range (white). (c-e) in reference to green range established at 1-3 SD of normal T1 range.

1

2

3 **Figure 2. Training program for standardization of cardiac T1-map image post-processing.**
4 The iterative training process consists of three phases. Additional training may be required upon
5 assessment to change contouring habits.
6

7

8

9

Figure 3 Example of trainees' performance. Left and middle panels show for Trainee A (top row) and Trainee B (bottom row) an example of different contouring tendencies on T1-maps. For Trainee A: (A) At first attempt, Trainee A had heavily eroded the automatically-placed LV contours for endo- and epicardium; (B) however, by the end of the training, this habit was corrected. For Trainee B: (C) At first attempt, Trainee B tended to include more partial volume within the myocardial ring (black arrows). (D) Overall contouring approach has remained consistent but some incorrect inclusions seen at first attempt have been eliminated by the end of the training. **Right panel shows** the effect of training on the average difference from the reference mean LV myocardial T1 values. Red vertical lines indicate when feedback was given to the trainee (Phase III: Assessment and Feedback).

Figure 4 Summary of the effect of training on average deviation from reference. Top panels: Impact of training on departures of left ventricular myocardial T1 (left panel) and wall thickness WT (right panel) estimates from reference values. The red dashed lines show the range of variation achieved at the end of the training, ± 8 ms (T1) and ± 1 pixel (WT), respectively. **Bottom panels:** Summary table (left panel) for the case of T1, showing the mean absolute deviation before and after training as well as the 95% lower and upper limits of agreement and bias for the Bland-Altman plot (right panel). In this plot the comparison is between deviation from reference before training (*dev_before*) versus deviation from reference after training (*dev_after*) in the case of T1.

Commented [VC1]: Q2 and Q7. Figure 4 has a new bottom panel to address the request for a summary table and a bland-altman plot of deviation before and after training