



# Display colour scale effects on diagnostic performance and reader agreement in cardiac CT and prostate apparent diffusion coefficient assessment

S. Zabala-Travers<sup>a,\*</sup>, B.D. Gallas<sup>a</sup>, S. Busoni<sup>b</sup>, M.C. Williams<sup>c</sup>, L. Noferini<sup>b</sup>, L. Fedeli<sup>b</sup>, S. Lucarini<sup>b</sup>, L. Galastri<sup>b</sup>, S. Mirsadraee<sup>c</sup>, A. Badano<sup>a</sup>

<sup>a</sup> Division of Imaging, Diagnostics and Software Reliability, OSEL/CDRH/FDA, 10903 New Hampshire Ave, Silver Spring, MD 20993-0002, USA

<sup>b</sup> Radiology and Medical Physics Department, Firenze University Hospital, Largo Brambilla 3, 50134 Firenze, Italy

<sup>c</sup> Clinical Research Imaging Centre, University of Edinburgh, Chancellor's Building SU315, 49 Little France Crescent, Edinburgh EH16 5U4, UK

## ARTICLE INFORMATION

### Article history:

Received 16 May 2018

Accepted 30 August 2018

**AIM:** To investigate the effect of colour scale choice on diagnostic performance in the interpretation of medical images.

**MATERIALS AND METHODS:** Twelve clinicians interpreted 210 myocardial computed tomography (CT) perfusion (CTP) examinations, and nine clinicians interpreted 165 magnetic resonance imaging (MRI) apparent diffusion coefficient (ADC) prostate images. In three separate sessions, each participant read the same image set using greyscale, hot-iron, and rainbow scales, respectively. Participants scored their level of confidence for tumour presence in the ADC study, and for ischaemia in the CTP study, from 0 to 100. The area under the receiver operating characteristic (ROC) curve (AUC) was used as the performance metric. For cases that scored >50, CTP readers' agreement on the ischaemic transmural extent was analysed, and ADC map readers' selected values and coordinates for the lowest ADC within the detected tumour were compared across different colour scales.

**RESULTS:** For CTP detection, the AUC was up to 0.10 higher with greyscale,  $0.67 \pm 0.02$  (standard error), compared to rainbow,  $0.56 \pm 0.02$ , and detection with hot-iron was in between ( $0.61 \pm 0.03$ ). For ischaemic transmural lesion categorisation, observed inter-reader agreement was highest with greyscale for category 25–50%. There is a small tendency for rainbow and greyscale to outperform hot-iron in the detection of prostate tumours. The selected lowest ADC value and pixel localisation was similar with all colour scales.

**CONCLUSIONS:** The present findings suggest that colour visualisation has a measurable effect on CTP and ADC performance. Further investigation is necessary to determine the magnitude of the effect in diagnostic tasks.

© 2018 The Royal College of Radiologists. Published by Elsevier Ltd. All rights reserved.

\* Guarantor and correspondent: S. Zabala-Travers, Present address: Department of Pediatric Radiology, Children Hospital Pereira Rossell, Bulevar Artigas 1550, Montevideo, 11300, Uruguay.

E-mail address: [s.zabala.travers@gmail.com](mailto:s.zabala.travers@gmail.com) (S. Zabala-Travers).

## Introduction

Colour-coded maps are the preferred visualisation approach in most recently developed imaging techniques; however, no standards or guidelines are available for radiologists to use in deciding on a colour scale that optimises image interpretation.<sup>1</sup> Magnetic resonance imaging (MRI) apparent diffusion coefficient (ADC) maps, and myocardial perfusion computed tomography (CTP) are two of many imaging methods using colour visualisation, with questionable reproducibility.<sup>2–6</sup> Could colour visualisation be playing a role in this variability?

Colour scales with a linear increase in luminance (e.g., greyscale and hot-iron scales) have been described as suitable for such tasks as lesion detection. On the other hand, rainbow-like scales, often called spiral scales for their hue variations, have been considered useful for categorisation and localisation.<sup>7,8</sup> A recent study<sup>9</sup> suggested that rainbow scales outperform the hot-iron scale and greyscale on intensity quantification tasks using functional MRI-like synthetic imagery.<sup>7,9,10</sup> Imaging biomarker assessment often requires a multitude of tasks, including detection, localisation, and quantification. A knowledge gap exists as to how the most prevalent colour scales in medical imaging — the rainbow-like, hot iron, and standard greyscale — affect clinical tasks.<sup>10</sup>

In prostate ADC maps and myocardial CTP, colour plays a significant role in qualitative and quantitative assessment.<sup>5,11–13</sup> ADC maps measure water movement in tissue. Prostate tumours depict water movement restrictions and show lower ADC than normal tissue. The Prostate Imaging Reporting and Data System (PI-RADS) recommends the use of ADC maps for improved detection, localisation, and characterisation of lesion aggressiveness.<sup>3</sup> Significant correlation between ADC map assessment and pathology has been reported and ADC values below 750–900  $\mu\text{m}^2/\text{s}$  have been correlated with clinically significant cancers.<sup>3,6</sup> Accurate ADC quantification can thus reduce unnecessary biopsies. Conversely, correct localisation is essential for biopsy guidance, lesion follow-up, surgery, and radiation therapy.<sup>14</sup> In addition, higher lesion localisation concordance among radiologists can provide more accurate risk stratification and patient surveillance.

Colour maps are also used for evaluating myocardial CTP. Because of reduced tissue iodine concentration, ischaemic and infarcted myocardium appear as hypodense compared to healthy tissue in contrast-enhanced CTP. Accurate localisation and quantification of the transmural extent of the ischaemic territory is an important determinant of complications, prognosis, and mortality.<sup>15</sup> The sensitivity and specificity of CTP range from 0.79 to 0.97 and 0.72 to 0.98, respectively.<sup>5</sup> To date, no report has established whether colour selection plays a role in detection, quantification, or reader variability.

The present study was undertaken to investigate whether the choice of colour scale affects malignant prostate tumour detection and characterisation in ADC maps, and myocardial ischaemic lesion detection and transmural extension assessment in CTP.

## Materials and methods

### Readers and cases

Twelve readers (radiologists, cardiologists, and one biomedical scientist; ages 27–36 years), four of them with high (>3 years) and eight with low (<1 year) experience, read CTP images at a cardiac imaging centre. Nine radiologists from a different radiology centre (ages 31–44 years) three of them with high, three with medium (1–3 years) and three with low experience, interpreted prostate ADC maps. Participants from each centre were randomly divided into three groups, with each group randomly assigned a different case subset using a “split-plot” study design.<sup>16</sup> Within each group, all readers read all cases with all colour scales (Table 1). Readers performed the Farnsworth–Munsell 100 hue colour vision test. One participant with red–green colour blindness only read images using the greyscale.

Cases were images corresponding to a single CT or MRI section. They were selected retrospectively by local experienced clinicians from studies locally acquired between 2014 and 2015. For the CTP study, 210 short-axis oriented CT sections representative of the left base, mid, and apex ventricle regions were manually selected from

**Table 1**  
Study design.

	Readers		
	Group 1	Group 2	Group 3
Myocardial perfusion CT study			
Case set 1: cases 1–70	F, High M, High M, Low F, Low M, Low M, Low		
Case set 2: cases 71–140		M, High M, Low M, Low	
Case set 3: cases 141–210			M, High F, Low F, Low
Prostate ADC MAP study			
Case set 1: cases 1–55	M, High F, Medium F, Low		
Case set 2: cases 56–110		F, High F, Medium F, Low	
Case set 3: cases 111–165			F, High F, Medium F, Low

*Split-plot design.* Readers in each centre were split into three groups. Cases were split into three subgroups. Each group of readers was assigned a subset of cases; 35 positive and 35 negative cases for CTP readers, and 22 positive and 33 negative cases for the ADC group. All clinicians in a subgroup read the same subset of cases three times, using a different colour scale each time. For each reader, gender (M: male F: female), age, level of experience (high >3 years, medium 1–3 years, low <1 year) is indicated.

44 stress and 31 rest CT examinations performed on 48 patients. One hundred and five of these cases were positive and 105 were negative for ischaemic lesions in both invasive coronary angiography and fractional flow reserve (reference standard). CTP was performed as previously described.<sup>17</sup> Sample cases from the myocardial study are displayed in Fig 1.

One hundred and sixty-five MRI ADC map sections representative of all prostate levels were used as cases for the prostate study. Sixty-six of these cases were positive for tumour and 99 were negative. They were manually selected from 56 (26 tumour-negative and 30 tumour-positive) histopathology-confirmed patient examinations. Pathology reports were considered the reference standard. Section A in the [Electronic Supplemental Material](#) summarises acquisition parameters for both studies.

As the reader tasks were not focused on patient outcomes and lesion as well as normal organ appearance was variable at different levels, up to three non-consecutive sections from each patient study, separated by at least 20 mm in CTP and 10 mm in ADC maps, were selected to increase case availability. Sections from the same patient examination were assigned to different case sets so that no reader assessed two cases from the same examination. Images were cropped to display only the prostate or heart region of interest, and were anonymised before selection. Both studies were locally approved by the institutions review board.

#### Reader tasks

Clinicians' performance with greyscale, hot-iron, and rainbow scales was tested for detection, categorisation, and quantification. Each group was presented with a set of 70 CTP cases (35 positive and 35 negative) or 55 prostate ADC map cases (22 positive and 33 negative). Readers interpreted the same set of cases three times, in three separate sessions, using a different colour scale each time. Cases were presented in random order in each session, and the sequence of colour scales for each participant was also randomised. Due to very limited reader availability, a minimum washout period of 30

minutes was set to minimise eye fatigue. Prostate readers used an inverted rainbow (IR) scale, which they requested. Readers received training on software, task demonstration, and use of the confidence-scoring scheme with a 14-case training set. The prevalence of lesions in the training set was similar to that in the study set. Readers were unaware of lesion prevalence during all sessions.

#### Task 1: detection

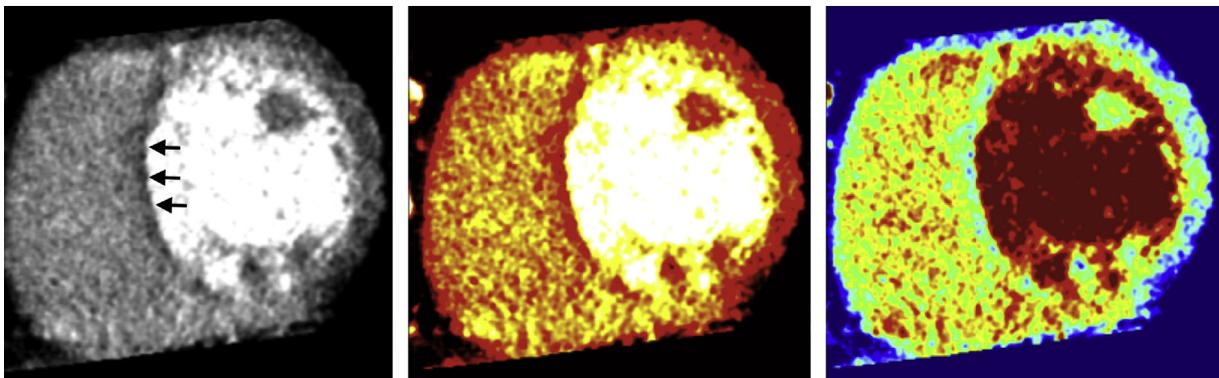
A horizontal slider from 0–100 was used for rating confidence in the presence or absence of a hypoperfused myocardial area (CTP) or a malignant prostate tumour (ADC). Readers were trained to use a range of 0–50 for cases they considered negative, using zero when the case was undoubtedly negative, 50 to 100 for cases considered positive, 100 for definitely positive and 50 when unable to decide on the presence or absence of a lesion. A plain language guide was displayed below the slider for easy reference. Invasive coronary angiography with fractional flow reserve and prostate pathology reports were used as truth for this task. Readers were blind to patient records and other imaging. For cases that scored positive for lesion (50 or above), they were prompted to perform a second task.

#### Task 2 for CTP: ischaemic lesion categorisation

In this CTP task, readers were asked to classify the transmural extent of the detected perfusion defect as a percentage of the myocardial wall: <25%, 26–50%, 51–75%, or >75%.

#### Task 2 for ADC maps: intensity quantification and localisation

The prostate group was asked to localise the lowest ADC value within the tumour using a mouse to position a circular 5-pixel region of interest (ROI). The median ROI value and the pixel coordinates were recorded. For cases where a second area was considered suspicious, readers were instructed to use the index lesion defined by PIRADS as the most aggressive for analysis. The software recorded performance time.



**Figure 1** Sample positive cases from the myocardial CTP study. Short axis-oriented sections representative of the left ventricle base displayed with the three evaluated colour scales. Left to right: greyscale, hot-iron, and rainbow scales. Arrows depict an ischaemic lesion.

## Workstation

Surface tablets (Windows 8.1, Microsoft, Redmond, WA, USA) were used to run custom-designed software (MATLAB R2015b, MathWorks, Natick, MA, USA) for presenting cases and recording answers. Cases were assessed on a medical display (Eonis MDRC-2224 BL, Barco, Kortrijk, Belgium), sized 512×512 pixels on a full-screen black background. A screenshot of the user interface is displayed in Fig 2. No image manipulation tools were provided. A fixed window width (WW) of 300 HU and window level (WL) of 100 HU was used for the myocardial study following institutional practice. For the prostate study, a fixed customised WW/WL of 1100/820 HU was selected based on the intensity value range of all cases. On-screen display modes were used to select colour settings. The standard RGB display mode was used for rainbow, IR and hot-iron sessions, and the Digital Imaging and Communications in Medicine (DICOM) mode, incorporating a greyscale standard display function (GSDF) look-up table, was selected for greyscale. Ambient light was kept within the American Association of Physicists in Medicine (AAPM) CT/MRI/NM reading room recommendations (50 lx).<sup>18</sup>

## Sizing and data analysis

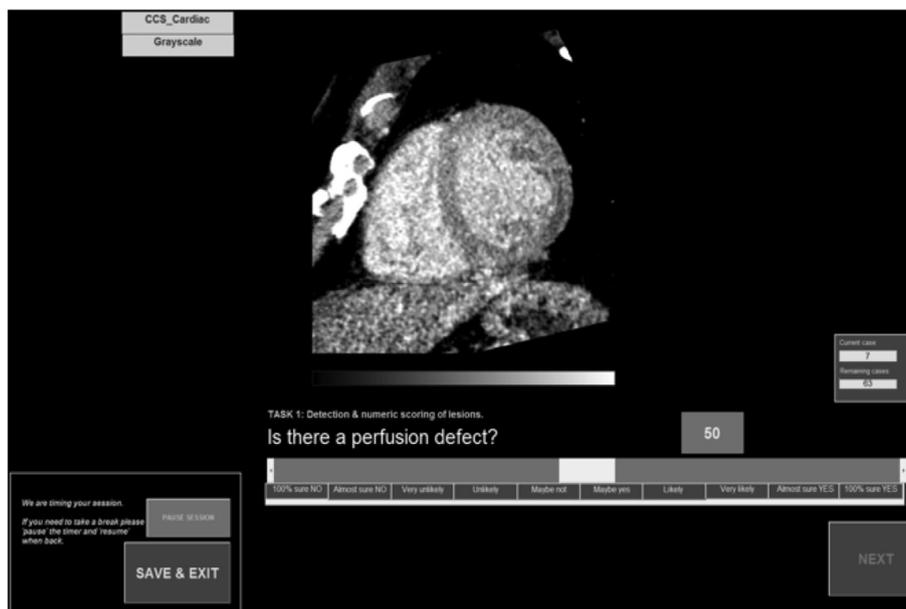
iMRMC software was used,<sup>19</sup> case variance estimates from a previous study,<sup>20</sup> an effect size of 0.05, and a 95% significance for sizing. The “split-plot” study design described above corresponds to power estimates of 98% for the CTP and 99% for the ADC study, in a fixed reader scenario. iMRMC software was also used for the statistical analysis of the colour scale comparison. The primary

endpoint was the area under the receiver operating characteristic curve (AUC).

No multiple-hypothesis adjustment was made for the three null hypotheses:  $AUC_{\text{greyscale}}=AUC_{\text{hot-iron}}$ ,  $AUC_{\text{greyscale}}=AUC_{\text{rainbow}}$ ,  $AUC_{\text{hot-iron}}=AUC_{\text{rainbow}}$ . Alternative hypotheses were  $AUC_{\text{greyscale}} \neq AUC_{\text{hot-iron}}$ ,  $AUC_{\text{greyscale}} \neq AUC_{\text{rainbow}}$ ,  $AUC_{\text{hot-iron}} \neq AUC_{\text{rainbow}}$ . To describe these hypothesis tests, we used 95% confidence intervals (CI) and *p*-values. Results were presented following an MRMC analysis approach,<sup>19</sup> which accounts for reader and case variability and is optimised for small sample sizes. This analysis is based on the *t*-distribution, *U*-statistic estimates of variance, and an estimate of the degrees of freedom.<sup>21</sup> A table including results with the more optimistic normal approximation is available in the [Electronic Supplemental Material, Section B](#). This table also offers detailed analysis by level of experience.

Additional analysis was performed to compare colour scale performance. Considering a confidence score of 50, readers' answers were binarised as positive (50 or above) or negative (below 50) for each lesion, in accordance with training received for this study. Answers were then compared to the ground truth, and the percentage of correct answers per case was estimated. The binarised analysis showed that all readers had incorrectly answered 24 cases for all colour scales (see Discussion for the hypothesis on this issue). A post-hoc analysis was then performed without these problem cases.

For the second task, no ground truth was available. When two readers scored a CTP case >50, their agreement on the lesion transmural extent was analysed. The observed percent agreement normalised by extension category was calculated for each colour scale (see [Electronic Supplemental Material, Section C](#), for details on the analysis methods). For the



**Figure 2** Screen capture of the software used for the CTP, task 1 (lesion detection). Below the 0–100 confidence score slider, a reference guide displayed a plain language description for each portion of the slider, reading from left to right: 100% sure no, almost sure no, very unlikely, unlikely, maybe not, maybe yes, likely, very likely, almost sure yes, 100% sure yes.

prostate study, the average difference in the lowest selected ADC value ( $\Delta\text{ADC}$ ,  $\mu\text{m}^2/\text{s}$ ) and the average Euclidean distance ( $d$ , mm) between all pairs of readers' pixel coordinates were analysed using the two-sample  $t$ -test to compare readers' consistency. Lacking a more rigorous method for agreement than  $\Delta\text{ADC}$  and  $d$ , all observations were pooled and assumed to be independent and identically distributed (see [Electronic Supplemental Material, Section D](#)).

## Results

### Task 1: detection

Detection results are presented in [Fig 3](#).

#### CTP planned analysis

For ischaemic lesion detection (CTP), the  $\text{AUC} \pm \text{standard error (SE)}$  was  $0.59 \pm 0.03$  with greyscale,  $0.57 \pm 0.03$  with hot-iron and  $0.52 \pm 0.03$  with rainbow. The biggest difference in these results was between greyscale and rainbow ( $\Delta\text{AUC} = 0.07$ , 95% CI =  $-0.01$  to  $0.16$ ).

#### CTP post-hoc analysis

The follow-up analysis without considering the 24 problem cases revealed performance was 0.1 higher when using greyscale,  $0.67 \pm 0.02$ , compared to rainbow,  $0.56 \pm 0.02$  (95% CI =  $0.004$  to  $0.2$ ,  $p = 0.04$ ). There was significantly improved detection with hot-iron,  $0.61 \pm 0.03$ , compared to that with rainbow when considering only the highly experienced readers (95% CI =  $0.06$  to  $0.09$ ). [Fig 4](#) depicts detection results by level of experience. Sensitivity with greyscale was higher than with hot-iron (95% CI =  $0.034$

to  $0.31$ ) and with rainbow (95% CI =  $0.02$  to  $0.29$ ), considering readers with all levels of experience. Sensitivity and specificity for detection are depicted in [Table 2](#).

For prostate, the AUC was  $0.79 \pm 0.03$ ,  $0.77 \pm 0.03$ ,  $0.80 \pm 0.03$  with greyscale, hot-iron, and IR, respectively ([Fig 2](#)). No significant differences were observed, although slightly lower AUCs were measured when using hot-iron (see [Table 2](#)). No significant differences were detected for experience.

Observed inter-reader percent agreement  $\pm \text{SE}$  for both lesion-present (LP) and lesion-absent (LA) cases was approximately 0.60 (see [Fig 5](#)). Observed intra-reader agreement for the same case across colour scales was  $0.68 \pm 0.11$  and  $0.72 \pm 0.14$  for CTP and ADC map, respectively.

There was no significant difference in AUC between hot-iron and IR using GSDF settings, although average sensitivity was lower for GSDF ( $0.69 \pm 0.08$  for hot-iron and  $0.68 \pm 0.15$  for IR) compared to the same session performed with RGB ( $0.73$  and  $0.73$ , respectively).

### Task 2 for CTP: ischaemic lesion categorisation

[Table 3](#) depicts observed inter-reader agreement for greyscale, hot-iron, and rainbow on transmural ischaemic lesion categorisation ([Fig 4](#)). Intra-reader observed agreement on categorisation was  $0.22 \pm 0.27$ ,  $0.33 \pm 0.14$ ,  $0.34 \pm 0.26$ , and  $0.27 \pm 0.23$  on categories,  $<25\%$ ,  $25\text{--}50\%$ ,  $50\text{--}75\%$ , and  $75\text{--}100\%$ , respectively (not displayed). Use of category distribution across readers is available in [Electronic Supplemental Material, Section E](#).

### Task 2 for ADC map: lesion quantification

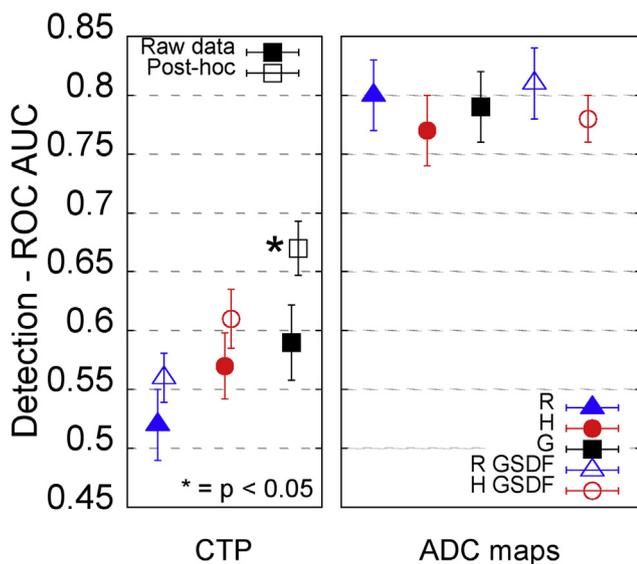
The average  $\Delta\text{ADC}$  value between readers was  $119.4 \pm 43.2$ ,  $101 \pm 33.3$ , and  $118.3 \pm 42.7$   $\mu\text{m}^2/\text{s}$  when reading with greyscale, hot-iron and IR, respectively (not significant). The average  $d$  between readers' coordinates of the lowest ADC value was  $4 \pm 5.8$ ,  $3.8 \pm 5.1$ , and  $3.7 \pm 4.3$  mm, when reading with greyscale, hot-iron, and IR, respectively ([Fig 6](#)). Sample ADC map cases and reader answers are displayed in [Fig 7](#).

No significant difference in time to decision was observed. Experienced readers completed sessions faster (see [Electronic Supplemental Material, Section F](#)). Raw data from this study can be found at <https://github.com/DIDSR/iMRMC/wiki/zabala2018-colorScaleEffects>.

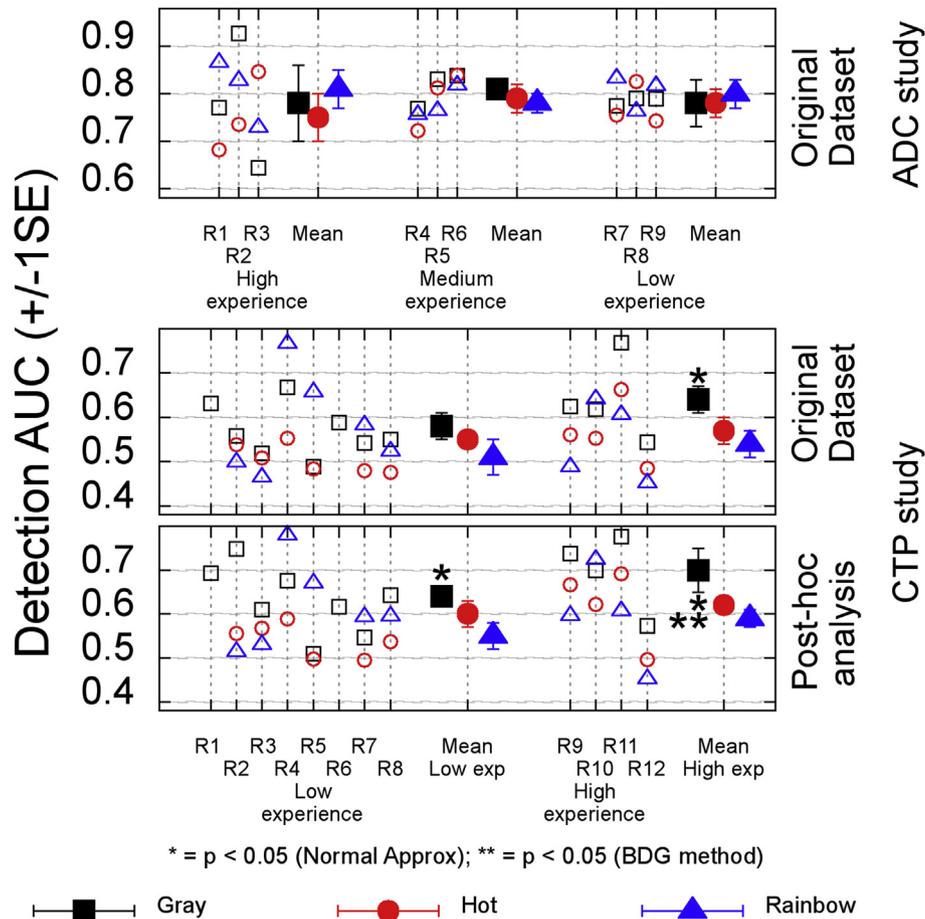
## Discussion

An effect of colour was observed on ischaemic lesion evaluation. Both greyscale and hot-iron scales improved detection performance compared to rainbow, consistent with previous publications on linear scales. The lower AUC registered with rainbow is consistent with a 20% lower detectability reported with spiral scales, attributed to the lack of perceptual order.<sup>7,22–24</sup>

Balancing case difficulty beforehand is challenging, particularly when using subjects of varied experience levels,



**Figure 3** Readers' detection performance with the different colour scales with myocardial CTP (left) and prostate ADC maps (right). For CTP, AUC are presented for raw data analysis and data without difficult cases, labelled as *Post-hoc*. Error bars correspond to 1SE. R: Rainbow, H: Hot-Iron, G: Grey, R-GSDF and H-GSDF: Rainbow and hot-iron sessions performed with greyscale display function; \*  $p$ -value for AUC comparison between greyscale and rainbow was  $<0.05$ .



**Figure 4** Detection AUCs classified by readers' level of experience in the ADC study (upper plot), the CTP study original dataset (middle plot), and CTP study dataset without difficult cases (lower plot). Each column represents a reader and each colour represents a session with a different colour scale. \* $p < 0.05$  with the normal approximation; \*\* $p < 0.05$  using Gallas' methods.

as is usual in diagnostic centres. The binarised response analysis was motivated by generally low AUCs for the CTP study, which triggered concerns that some cases were excessively difficult for the task requested, particularly as the study task differed in some ways from the regular clinical task (see below). Given this observation, 24 cases (approximately the same number of positives and

negatives) answered incorrectly by all readers in all colour scales were removed for post-hoc analysis. It was hypothesised that the registered improvement in AUCs, which followed the same trend found in the originally planned analysis, together with a reduction in variability (possibly associated with readers operating further away from the "guessing line"), highlighted differences across colour scales (Fig 2). Even though adjusting case difficulty *a posteriori* is a limitation of the study, the removal of cases from all colour scales and readers ensured that no visualisation method was favoured over others.

Surprisingly, the observed results for the ADC task showed that IR had a slight advantage over greyscale and hot-iron on tumour detection, consistent with Borland's statement regarding the absence of a universal scale.<sup>7</sup> ADC map-readers used the IR scale. Lower human contrast sensitivity between hues, with small changes in luminance in the green–cyan transition has been described.<sup>23</sup> Coincidentally, normal myocardium is shown in the green–cyan range with most of the ischaemic lesions being cyan and light blue. On the other hand, prostate tumours displayed with the IR scale are in yellows and reds, contrasting with the green and cyan representing normal prostate ADC values (see Section G in the [Electronic Supplemental Material](#)).

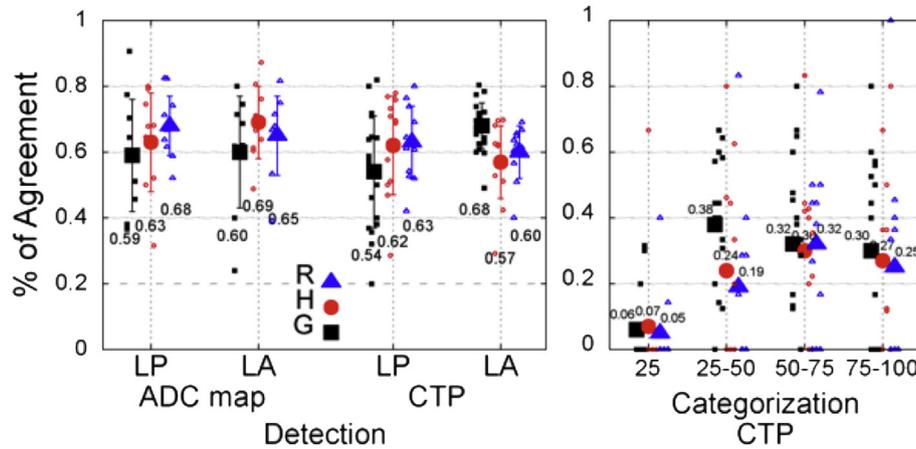
**Table 2**  
Sensitivity and Specificity for lesion detection.

	Myocardial CTP		Prostate ADC maps	
	Sensitivity (SE)	Specificity (SE)	Sensitivity (SE)	Specificity (SE)
Greyscale	0.69 (0.04) <sup>a</sup>	0.57 (0.06)	0.72 (0.08)	0.66 (0.09)
Hot-Iron	0.52 (0.05)	0.66 (0.06)	0.73 (0.06)	0.60 (0.07)
Rainbow	0.54 (0.04)	0.58 (0.05)	0.73 (0.06)	0.69 (0.07)
H-GSDF			0.69 (0.08)	0.70 (0.10)
R-GSDF			0.68 (0.15)	0.69 (0.13)

Sensitivity and specificity for lesion detection in the myocardial CTP and prostate ADC studies. SE: 1 standard error.

H-GSDF and G-GSDF: hot-iron and greyscale session performed with greyscale standard display function settings.

<sup>a</sup> Sensitivity with greyscale is higher than with hot-iron (95% CI=0.034,0.31) and with rainbow (95% CI=0.02,0.29).



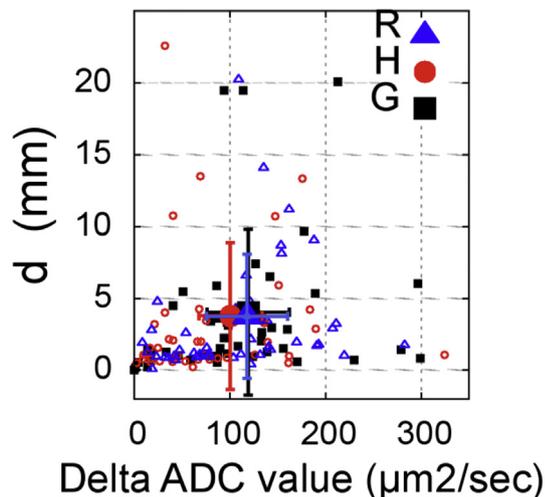
**Figure 5** Percent agreement (SD) on lesion detection (left) and categorisation of ischaemic lesion transmural extension (right). Small marks represent percent agreement for all possible reader combinations; larger marks represent the average percent agreement for each visualisation method: greyscale (G), hot-iron (H) and rainbow (R), for lesion present (LP) and lesion absent (LA) cases. Error bars correspond to 1 SE of pooled data.

**Table 3**  
Inter-reader agreement for ischaemic lesion categorisation.

	0–25%	25–50%	50–75%	75–100%
Grayscale	0.06 (0.11)	0.38 (0.22)	0.32 (0.24)	0.30 (0.27)
Hot-iron	0.07 (0.21)	0.24 (0.28)	0.30 (0.23)	0.27 (0.28)
Rainbow	0.05 (0.12)	0.19 (0.23)	0.32 (0.23)	0.25 (0.30)

Inter-reader agreement for transmural extent of the detected ischaemic lesion across colour scales. Agreement is noticeably higher for grayscale in the 25–50% category.

Readers assessed each case once with each colour scale; information on intra-reader agreement within scales is not available. Intra-reader agreement across colour scales was approximately 70%, while inter-reader agreement within colour scales was 60–70% for all methods, consistent with



**Figure 6** Correlation between average differences in distance (*d*) and ADC value across readers using the different scales. Each small mark corresponds to one case; the larger mark corresponds to the reader- and case-average for each method (G: grayscale, H: hot-iron, R: rainbow). Error bars correspond to 1 SE.

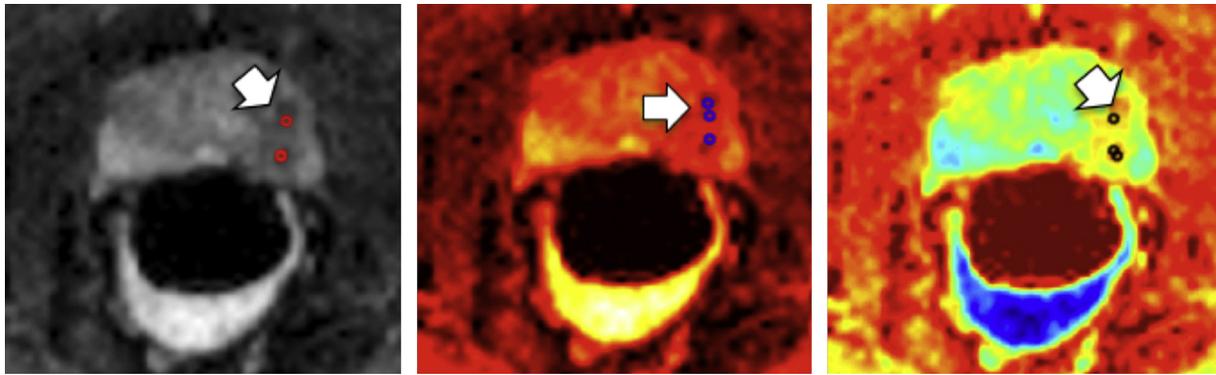
previous reports.<sup>2–4,6</sup> No significant difference by reader level of experience was detected, except for time to decision, which was lower for experienced readers.

Agreement was higher with grayscale for less conspicuous ischaemic lesions (26–50% wall thickness) compared to hot-iron and rainbow, but the data are quite noisy. Finally, the use of higher categories (>50%) was generally more prevalent with hot-iron and rainbow, suggesting that readers might tend to overestimate size with these scales.

The average difference for the lowest ADC value (101–119  $\mu\text{m}^2/\text{s}$ ) is higher than 0.1 of the 750–900  $\mu\text{m}^2/\text{s}$  threshold for significant cancer.<sup>3</sup> Although similar across methods, this difference should be carefully considered when using threshold ADC in support of clinical decisions. Automated image quantification products are not yet available, and readers are still required to select areas for ADC calculation. The present results suggest that measured values of the order of 650  $\mu\text{m}^2/\text{s}$  could relate to unmeasured cancer regions.

Localisation is related to quantification. A higher *d* could be interpreted as readers considering different areas suspicious for malignancy, or as a mismatch in readers’ intended and recorded localisations, including marks with very low intensity values (such as rectum or urethra), resulting in large ADC differences.

The following differences were observed between the proposed and clinical tasks. First, readers were blind to clinical records. Second, cases were read as single sections, making the differentiation of lesions from artefacts more difficult. Third, no windowing tool was available, and image contrast could have been suboptimal for some cases. Fourth, no other MRI sequence was available to assist in differentiating malignant from benign findings, which exhibit minimal ADC values; including hyperplasia nodules, calculi and other calcifications, areas of fibrosis, dense fibromuscular stroma, or blood products.<sup>3</sup> Finally, the 30-minute session spacing is a limitation in the study; however, the large number of cases, the random case display, as well as



**Figure 7** Sample prostate positive cases from the ADC map study. Readers' selected location for the lowest ADC value in the tumour is represented with a mark ('o'), see arrows. Left to right: greyscale, hot-iron, rainbow scale.

the colour scale session randomisation among readers were used to minimise any effect of case recall on the study results.

The results of this research do not support recommending a colour scale for myocardial CTP or for prostate ADC map evaluation. In addition, findings apply only to the diagnostic methods evaluated, and should not be unjustifiably extrapolated. The present findings constitute a first investigation of a previously unaddressed issue. It is possible that some colour scales are more suitable for detection, localisation, quantification, or intra-lesion localisation tasks; however, as changing scales for each task might not be practical in a clinical scenario, efforts must focus on finding the best-performing colour scale for a mix of tasks.

In conclusion, myocardial perfusion CT readers show improved detection performance (up to 0.10 AUC), and are more consistent on lesion transmural extension categorisation when using greyscale visualisation. A small tendency was found for the IR scale and greyscale to outperform the hot-iron scale for prostate cancer detection. An average difference in the detected lowest ADC value of 101–119  $\mu\text{m}^2/\text{s}$  and an average distance of 4 mm for lesion localisation were measured and were similar across scales. The present findings indicated a measurable effect of colour on image interpretation that merits further investigation, with study designs focusing on reducing variability in order to provide optimal presentation in imaging techniques relying on colour visualisation.

## Conflict of interest

There are no conflicts of interest to report.

## Disclaimer

The mention of any commercial products herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services. This is a contribution of the U.S Food and Drug Administration and is not subject to copyright.

## Acknowledgements

The authors acknowledge Dr Adam Wunderlich for his valuable assistance in data analysis, Tom Kimpe (BARCO) for useful discussion in the study design and for providing the medical monitors for the study. We also acknowledge readers at Careggi Hospital (F Mungai, G Grazzini, L Bartolini, L Galastri, S Pradella, S Lucarini, B Matteuzzi, E Mazzoni, Marta B) and the Center of Cardiovascular Science, UoE (S McLenachan, J Davis, T Gordon, G Papanastasiou, A Martin, A Moss, M Doris, S Alam and C Serrat) for their invaluable participation in this study. S.Z.-T. acknowledges funding by appointment to the Research Participation Program at the Center for Devices and Radiological Health (CDRH) administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the US Food and Drug Administration. This work was partially supported by a CRADA between CDRH and BARCO, and it was partly funded by the British Heart Foundation (FS/11/014).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.crad.2018.08.016>.

## References

1. Badano A, Revie C, Casertano A, et al. Consistency and standardization of color in medical imaging: a consensus report. *J Digit Imaging* 2014;1–12.
2. Saba L, Argiolas GM, Raz E, et al. Carotid artery dissection on non-contrast CT: does color improve the diagnostic confidence? *Eur J Radiol* 2014;83(12):2288–93. <https://doi.org/10.1016/j.ejrad.2014.09.001>.
3. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS prostate imaging - reporting and data system: 2015, version 2. *Eur Urol* 2016;69(1):16–40. <https://doi.org/10.1016/j.eururo.2015.08.052>.
4. Rana AK, Wardlaw JM, Armitage PA, et al. Apparent diffusion coefficient (ADC) measurements may be more reliable and reproducible than lesion volume on diffusion-weighted images from patients with acute ischaemic stroke—implications for study design. *Magn Reson Imaging* 2003;21(6):617–24. [https://doi.org/10.1016/S0730-725X\(03\)00087-0](https://doi.org/10.1016/S0730-725X(03)00087-0).
5. Williams MC, Reid JH, McKillop G, et al. Cardiac and coronary CT comprehensive imaging approach in the assessment of coronary heart disease. *Heart* 2011;97(15):1198–205.

6. Muller BG, Shih JH, Sankineni S, et al. Prostate cancer: interobserver agreement and accuracy with the revised prostate imaging reporting and data system at multiparametric MR imaging. *Radiology* 2015;**277**(May 2016):142818. <https://doi.org/10.1148/radiol.2015142818>.
7. Borland D, Taylor II RM. Rainbow color map (still) considered harmful. *IEEE Comput Graph Appl* 2007;**27**(2):14–7.
8. Pizer SM, Zimmerman JB. Color display in ultrasonography. *Ultrasound Med Biol* 1983;**9**(4):331–45.
9. Zabala-Travers S, Choi M, Cheng W-C, et al. Effect of color visualization and display hardware on the visual assessment of pseudocolor medical images. *Med Phys* 2015;**42**(6), <https://doi.org/10.1118/1.4921125>.
10. Griggs B. *The end of rainbow? An exploration of color in scientific visualization*. BSc thesis. University of Oregon; 2014.
11. Tan CH, Hobbs BP, Wei W, et al. Dynamic contrast-enhanced MRI for the detection of prostate cancer: meta-analysis. *AJR Am J Roentgenol* 2015 Apr;**204**(4):W439–48. <https://doi.org/10.2214/AJR.14.13373>.
12. Chang Z, Wang C. Treatment assessment of radiotherapy using MR functional quantitative imaging. *World J Radiol* 2015;**7**(1):1.
13. Bauer RW, Kerl JM, Fischer N, et al. Dual-energy CT for the assessment of chronic myocardial infarction in patients with chronic coronary artery disease: comparison with 3-T MRI. *AJR Am J Roentgenol* 2010;**195**(3):639–46. <https://doi.org/10.2214/AJR.09.3849>.
14. Costa DN, Pedrosa I, Donato F, et al. MR imaging—transrectal US fusion for targeted prostate biopsies: implications for diagnosis and clinical management. *RadioGraphics* 2015;**35**(3):696–708. <https://doi.org/10.1148/rg.2015140058>.
15. Geltman E, Ehsani A, Campbell M, et al. The influence of location and extent of myocardial infarction on long-term ventricular dysrhythmia and mortality. *Circulation* 1979;**60**(4):805–14.
16. Obuchowski NA. Reducing the number of reader interpretations in MRMC studies. *Acad Radiol* 2009;**16**(2):209–17.
17. Williams MC, Mirsadraee S, Dweck MR, et al. Computed tomography myocardial perfusion vs 150-water positron emission tomography and fractional flow reserve. *Eur Radiol*. 2016;1–11.
18. Samei E, Badano A, Chakraborty D, et al. Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. *Med Phys* 2005;**32**(4):1205–25.
19. Gallas BD. One-shot estimate of MRMC variance: AUC. *Acad Radiol* 2006;**13**(3):353–62. <https://doi.org/10.1016/j.acra.2005.11.030>.
20. Van Dyke CW, White RD, Obuchowski NA, et al. Cine MRI in the diagnosis of thoracic aortic dissection. In: *79th annual meeting of the radiological society of north America*. Chicago, Illinois: Radiological Society of North America; 1993.
21. Obuchowski NA, Gallas BD, Hillis SL. Multi-reader ROC studies with split-plot designs. A comparison of statistical methods. *Acad Radiol* 2012;**19**(12):1508–17. <https://doi.org/10.1016/j.acra.2012.09.012>.
22. Rogowitz B, Treinish LA. Data visualization: the end of the rainbow. *Spectr IEEE*. 1998;**35**(12):52–9.
23. Marchak FM, Cleveland WS, Rogowitz BE, et al. The psychology of visualization. In: *Proceedings of the 4th conference on visualization'93*; 1993. p. 351–4.
24. Li H, Burgess AE. Evaluation of signal detection performance with pseudocolor display and lumpy backgrounds. In: Kundel HL, editor. *SPIE proceedings*. International Society for Optics and Photonics; 1997. p. 143–9. <https://doi.org/10.1117/12.271286>.