A comparative assessment of intra- and inter-observer repeatability of three widely used software packages forthe quantification of defect sizein stress myocardial perfusion scintigraphy

Christos Sachpekidis[1]MD,
Annette Kopp-Schneider[2] PhD,
Vasileios Sachpekidis[3] MD,
Efstratios Moralidis[1]MD

1. Department of Nuclear Medicine, Aristotle University, Papageorgiou Hospital, Thessaloniki, Greece
2. Department of Biostatistics, German Cancer Research Center, Heidelberg, Germany
3. Department of Cardiology, Papageorgiou Hospital, Thessaloniki, Greece

Keywords: Myocardial perfusion scintigraphy
- Ischaemia quantification
- Software packages
- Repeatability

Corresponding author:
Christos Sachpekidis MD,
Department of Nuclear Medicine
Aristotle University Medical School, Papageorgiou Hospital
Ring Road, Nea Efkarpia
Thessaloniki 56429, Greece
christos_saxpe@yahoo.gr

Received:
16 August 2022
Accepted revised:
6 December 2022

Abstract

Objective: To assess the intra- and inter-observer repeatability of popular software packages for the quantitative determination of abnormality size in stress myocardial perfusion scintigraphy. **Subjects and Methods:** A total of 182 tomographic stress myocardial perfusion scans were processed in duplicate by an experienced and trainee observer to assess SSSext (summed stress score multiplied by 100/68) and total defect extent (TDE), as % of the left ventricle, with 4 dimension-myocardial (4DM), emory cardiac toolbox (ECTb) and quantitative perfusion SPECT (QPS) packages. The Bland-Altman (B-A) analysis and Lin's concordance correlation coefficient (CCC) were used to assess agreement. **Results:** In SSSext's intra-observer repeatability, CCC showed substantial agreement for 4DM and QPS, and moderate agreement for ECTb for both observers. In inter-observer repeatability, CCC revealed substantial agreement for 4DM and QPS, and poor agreement for ECTb. Regarding TDE, CCC showed substantial intra-observer repeatability for both operators using all packages, while the inter-observer repeatability was substantial for 4DM and QPS, and moderate for ECTb. In SSSext's intra-observer repeatability for 4DM, ECTb and QPS, the B-A analysis provided (mean±1.96SD of paired measurements) 0.0±4.3, 0.2±7.8, -0.6±7.6 for the experienced physician and 0.2±5.9, 0.0±7.5, -0.5±5.4 for the trainee, respectively; in inter-observer repeatability it provided 0.2±5.4, 0.1±9.6, 0.2±8.1, respectively. Regarding TDE, the B-A values for intra-observer repeatability were 0.1±5.2, 0.1±7.9, 0.1±2.8 for the experienced reader and 0.3±6.6, -0.1±6.4, -0.1±2.4 for the trainee, respectively; in inter-observer agreement the B-A provided 0.6±6.4, -0.2±10.3, -0.1±4.3, respectively. **Conclusion:** Considerable differences in intra- and inter-observer agreement were noted for the quantitative determination of defect size using widely employed software packages, suggesting limitations in the clinical use of these measurements. Quantitative perfusion SPECT appears preferable, but with no significant advantage over 4DM. There were no significant differences between the observers.

Hell J Nucl Med 2022; 25(3): 274-284 *Epub ahead of print: 14 December 2022* *Published online: 30 December 2022*

Introduction

Myocardial perfusion scintigraphy (MPS) is a well-established modality offering important diagnostic and prognostic information for patients with known or suspected coronary artery disease (CAD) [1,2]. It has been shown that patients with extensive ischaemia are more likely to realize a survival benefit with revascularization, as opposed to patients with limited ischaemia which fare better with medical therapy [3,4]. For management decisions a cut-off point of the extent of ischaemia of >10% of the left ventricle(LV)is recommended [5]. Therefore, the determination of the amount of perfusion abnormality is essential.

Visual assessment has been the backbone of MPS interpretation for several years, but this approach may suffer from intra- and inter-observer variability and may depend on the user's experience [6-9]. Software packages have been introduced in clinical practice to improve the standardization of MPS interpretations [10-12]. The three most widely used software packages for automated quantification of myocardial perfusion are the quantitative perfusion SPECT (QPS, Cedars-Sinai Medical Center, Los Angeles, CA, USA) [13], emorycardiac toolbox (ECTb, Emory University, Atlanta, GA, USA) [12], and 4 dimension-myocardial SPECT (4DM, University of Michigan, Ann Arbor, MI, USA) [14].

From a clinical standpoint, a tight agreement between repeated measurements may be more important than the accuracy [15]. In addition, the accuracy may not be quantified in the absence of an available gold standard, as in our case. Repeatability (agreement between different analyses of a single acquisition of MPS data) may discern the smallest amount by which two measurements must differ to determine whether the difference is significant. Hence, the control of intra- and inter-observer variability by these programs may improve the consistency of image interpretation and set serial imaging from a more comprehensive perspective. However, the repeatability of software packages in quantitatively assessing perfusion abnormalities in a comparative setting has not been previously evaluated [16].

The present study aimed to assess the intra- and interobserver repeatability of the QPS, ECTb and 4DM software packages in the quantification of stress MPS abnormalities by experienced and trainee physicians. These software packages were selected because of their popularity, automated quantitative features, potential to process the same reconstructed data, and the ability to provide in a consistent manner both the summed stress score (SSS, which can be used for the assessment of defect size) and the total defect extent (TDE, the extent of the left ventricle being hypoperfused). The indices SSS and TDE provide powerful outcome data that have stood the test of time [3, 4, 17].

Subjects and Methods

Study population

One hundred and eighty-two consecutive patients referred for routine 1-daytechnetium-99m (^{99m}Tc)-sestamibi MPS in our department were retrospectively enrolled from our data base. Ninety-two patients underwent a stress/rest protocol and 90 patients underwent a rest/stress protocol. Patients were assigned randomly to either protocol, according to their appointment in the morning (stress/rest) or afternoon (rest/stress) sessions. Hence, as adults had been submitted to a routine examination, with no excess radioactivity or other interventions, and their scans were recruited in a retrospective fashion, involvement in an ethical committee process was unnecessary according to local ethical standards. Nevertheless, as part of departmental routine, all patients were asked to provide signed consent before undergoing the scan, for use of their imaging and clinical data for academic and research purposes. Clearance for this procedure has been obtained from our Institutional Scientific (and Ethics) Committee (No 330/568).This study was performed in accordance with the principles of the Declaration of Helsinki. The characteristics of the participants are listed in Table 1. There were no essential differences between the patients who underwent the different protocols.

Patient preparation, stress protocol and image acquisition

Medication withdrawal was left at the discretion of the referring physician, but in general patients were instructed to abstain from caffeine for 24h, nitrates in the morning of the study, and the ophylline derivatives for 48h. Patients were stressed using either coronary vasodilators or exercise testing using the standard symptom-limited Bruce protocol. For patients weighing approximately 70-80kg 250-300MBq of ^{99m}Tc-sestamibi were injected intravenously for the initial study, followed by the administration of threefold this activity for delayed imaging. For heavier patients doses were upscaled according to published recommendations [18]. For all patients stress images were acquired 15-60 minutes after ^{99m}Tc-sestamibi injection, and rest images were acquired 45-60 minutes after radiotracer injection. Acquisition was performed with a dual-head gamma-camera (GE Healthcare, Discovery NM 630, Haifa, Israel), equipped with low-energy, high-resolution collimators, in the supine position, in step and shoot mode using a circular orbit and a 64×64 matrix, a zoom factor of 1.4 and a pixel size of 6.3mm, with 60 projections over 180° from the 45° right anterior oblique to the 45° left posterior oblique projection and 25s and 20s per projection for the initial and delayed images, respectively.

Table 1. *Patient characteristics (N = 182).*

Patient characteristics	Values
Male gender	142 (78.0%)
Age (yrs, mean ±SD)	64.4±10.4
BMI (kg/m^2, mean ±SD)	30.1±4.9
Hypertension [N (%)]	127 (69.8%)
Diabetes [N (%)]	65 (35.7%)
Dyslipidemia [N (%)]	123 (67.6%)
Smoking [N (%)]	42 (23.1%)
Previous history (MI/PCI/CABG) [N (%)]	78 (42.9%)/ 73 (40.1%)/ 46 (25.3%)
Indication (Diagnosis/Risk stratification) [N (%)]	66 (36.3%)/ 116 (63.7%)
Stress (Exercise/Adenosine/ Dipyridamole)	22 (12.1%)/ 48 (26.4%)/ 112 (61.5%)

BMI, body mass index; MI, myocardial infarction; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting

Image processing

Singe photon emission tomography computed tomography data were processed on a Xeleris3 workstation using the Myovation software package (GE Healthcare, Haifa, Israel). Images were reconstructed using OSEM/MLEM (2 iterations, 10 maximum number of subsets) with a post-processing 3D Butterworth filter of an order 10 and a cut-off frequency 0.35 cycles/cm for the lower activity and an order 5 and a critical frequency 0.40 cycles/cm for the images with the higher activity administered. Automated quantitative analysis was performed using QPS (version 4.0), ECTb (version 3.0), and 4DM (version 4.0) software packages. Minimal manual interaction was allowed by the program to ensure correct placement of a region of interest enclosing the heart and separating it from surrounding areas of extracardiac activity, such as the liver or bowel, and to adjust the orientation of the heart to the standard coordinates before proceeding with automated myocardial perfusion boundary detection. The latter step allows the correct co-registration of stress and rest images; it is an essential part of processing and it is invariably included in software programs. All analyses were carried out as suggested by the manufacturers of the software packages, and suitable vendor databases were used for all analyses. Both the SSS and TDE, expressed as a percentage of the LV, were automatically calculated for every patient using each software tool.

The assessment of SSS was based on a17-segment model of the LV and a 5-point scale for grading segmental tracer uptake, as suggested by guidelines: 0, normal (<10%); 1, mild reduction (10%-<25%); 2, moderate reduction (25%-50%); 3, severe reduction (≥50%); and 4, absent (background level of counts) [19]. Subsequently, SSS was used to express the percentage of LV hypoperfusion by dividing it by 68 (the maximum potential score in the 17-segment model) and multiplying by 100, thus providing an index of the extent of the perfusion abnormality at stress (SSSext). This approach has been coined previously in a seminal and widely quoted work on the prognostic significance of the extent of ischaemia that proposed the 10% threshold of ischaemia for the management of patients [3]. In that study a 20-segment semiquantitative visual interpretation was performed. The TDE was based on the number of pixels with perfusion lower than the threshold set by the normal database of each software tool. Hence, TDE is different from SSSext which incorporates the size and severity of the defects.

Stress acquisition data were analyzed independently by an experienced nuclear cardiologist with more than 25 years of clinical experience, and a trainee who had completed two years of training in nuclear cardiology. Images were processed in duplicate (baseline-repeat) by each operator to assess intra-observer repeatability, allowing at least two weeks between sessions. Both operators were unaware of the patient data. The baseline analysis of the experienced physician was compared with the baseline analysis of the trainee to determine inter-observer repeatability. In an attempt to better delineate the observer's agreement in determining the critical 10% of the LV threshold of the extent of ischaemia, patients with 5%-15% area of perfusion abnormality were evaluated separately. In this subgroup analysis, pair wise comparisons comprised measurements of the extent of perfusion abnormality between ≥5% and ≤15% of the LV, as measured by either the first or second processing of data (for intra-observer assessment) by either operator (for inter-observer assessment) or when the mean value of paired measurements was within that range (for both intra- and inter-observer assessment).

In addition, for every software package the estimation of the extent of the perfusion abnormality as assessed with SSSext was compared with that of TDE. For this purpose, for every software package the baseline measurement of SSSext by each operator was compared with the baseline assessment of the TDE, and the repeat measurement of SSSext was compared with the repeat determination of the TDE.

Statistical analysis

Continuous variables are presented as mean±1 standard deviation (SD). Categorical data are presented as numbers or proportions. Welch tests were used to compare independent means. The bias of the pairs of measurements was assessed using the mean of their differences. The t-test was used to investigate non-zero bias. The agreement between variables was assessed with Bland-Altman analysis, including the 95% limits of agreement (LoA), which are defined as the mean±1.96 SD of the differences of paired measurements. In addition to the LoA derived for paired measurements in the entire population, the analysis was also performed for the subgroup of the two measurements where either one of the two values or the mean of the two values was between 5% and 15%. The Pitman-Morgan test was used to assess whether the LoA's were different between the different software packages, that isto compare the LoA. In addition, Lin's concordance correlation coefficient (CCC) was calculated and interpreted as follows: CCC<0.90 was considered to represent poor agreement, CCC=0.90-0.95 moderate agreement, CCC=0.95-0.99 substantial agreement and CCC> 0.99 almost perfect agreement [20]. Concordance correlation coefficient was calculated using the R package epiR. To compare the intra- and inter-observer agreement in patients with a body mass index (BMI) ≥30 versus those with a BMI<30, the variance of the differences of paired measurements was compared between the two patient groups by the F-test. The statistical significance level was set to 0.05, and P-values were adjusted for multiplicity with the Bonferroni-Holm correctionas required. The calculations were performed using R (version 3.4.3) or SAS (version 9.4).

Results

The participants demonstrated a wide range of SSSext and TDE values. Descriptive statistics of the measured parameters by both examiners using all three software packages are presented in Table 2. Interestingly, with the 4DM software package, TDE values invariably significantly over estimated the extent of ischaemia in comparison to SSSext in both experienced and trainee observers, but the opposite occurred with the QPS software. Regarding the ECTb program, there was a trend for overestimation of the area of ischaemia with TDE in comparison to SSSext, although statistical significance was occasionally observed.

The results of the Bland-Altman analysis for SSSext and TDE are illustrated in Figures 1 and 2 for intra-observer agreement, and Figure 3 for inter-observer agreement and summarized in Table 3.

Regarding the SSSext assessment, the 4DM software program attained the best intra-observer agreement (i.e. the narrowest LoA) for the experienced observer and the best inter-observer agreement, whereas the ECTb offered the worst intra-observer repeatability for the trainee and the worst inter-observer repeatability. With regard to the TDE, the QPS program attained the best intra-observer agreement for both operators and the best inter-observer agre-

Table 2. Descriptive statistics (mean±SD) of the studied parameters obtained with the three different software packages by the two observers.

Software	SSSext		TDE	
	Baseline	Repeat	Baseline	Repeat
Experienced physician				
4DM	13.0±11.8**	13.1±12.1##	17.6±15.1**	17.6±15.2##
ECTb	14.9±10.8	14.7±11.1#	17.3±12.9	17.2±13.2#
QPS	16.3±14.5*	17.0±15.0##	13.2±11.1*	13.1±11.1##
Trainee				
4DM	13.2±11.8**	13.1±11.8##	18.2±15.3**	18.0±15.2##
ECTb	15.0±10.5	15.0±10.2	17.1±12.7	17.2±12.3
QPS	16.5±14.7*	17.0±15.0##	13.2±11.3*	13.3±11.3##

SSSext, summed stress score extent; TDE, total defect extent.
**,#P<0.05, for the comparison between SSSext and TDE of the same vendor in the baseline (*) and the repeat (#) assessments, for both operators.*
***, ##P<0.01, for the comparison between SSSext and TDE of the same vendor in the baseline (*,**) and the repeat (#,##) assessments, for both operators.*

ement for both operators and the best inter-observer agreement, whereas the ECTb offered the wider LoA in the experienced physician's intra-observer assessment and the inter-observer evaluation. The imperfections of intra- and inter-observer agreement were also addressed in patients with "5%≤area of ischaemia ≤15%" (impeded in Figures 1-3). Notably, in that subgroup the 95% LoA closely followed the patterns of the entire population, as described previously (Table 3). It is also worth mentioning that despite the disparities among the software packages, the assessment of TDE with the QPS software consistently provided the narrowest limits in both intra- and inter-observer agreement.

Concordance correlation coefficient values are presented in Table 4. These results to a certain extent parallel those in Table 3 and show that the QPS software package somehow outperforms the remaining packages in the measurement of TDE, in terms of intra- and inter-observer agreement. Conversely, ECTb in no case is superior to any other software package in assessing both SSSext and TDE.

In separate analyses of the subpopulations studied regarding the sequence of image acquisition (stress/rest versus rest/stress protocol), no essential differences were observed between the two protocols concerning intra- and inter-observer repeatability for all software packages under consideration. Similarly, no differences in repeatability were found according to the sex.

In obese patients (BMI≥30), the experienced operator had a significantly narrower LoA for intra-observer agreement in the assessment of SSSext with the QPS software than in patients with a BMI<30 (-0.8±5.5 vs 0.4±9.0, respectively, P<0.001). In addition, obesity improvedinter-observer agreement, compared to non-obese patients, albeit statistical significance was attained in the measurement of SSSext with the 4DM (0.3±3.7 vs 0±6.4, P<0.001) and the QPS programs (0.1±5.7 vs 0.3±9.5, P<0.001) and also in the assessment of TDE with the ECTb (-0.1±7.0 vs -0.3±12.1, P<0.001) and the QPS programs (-0.1±3.3 vs 0.1±4.7, P<0.05). Similar trends were observed in patients with "5%≤ area of ischaemia ≤15%", as defined previously.

Discussion

This study evaluated intra- and inter-observer repeatability and highlighted the difficulties ofthe commercially available and widely used 4DM, QPS and ECTb software packages, as they are commonly applied in clinical practice, to assess defect size with stress myocardial perfusion SPECT. Two different approaches were tested; one based on SSS, similar to that originally used to establish the cut-off point of 10% of the LVfor the extent of ischaemia [3], and the direct measurement ofthe area of hypoperfusion based on thresholds from normal databases, from an experienced and a trainee observer. The overall impression was that in terms of intra- and inter-observer variability the QPS software appears preferable, but with no substantial advantage over the 4DM, whereas the ECTb package is comparatively less satisfactory, although all three programs were suboptimal. This was particularly evident in the range of ischaemia of 5%-15% of the LV, in which a clinically satisfactory degree of agreement could not be consistently attained.

Regarding SSSext, the best results in intra-observer agreement were provided by 4DM and secondarily by the QPS software package, although the LoA were wide in both cases. This variability was even more pronounced with the use

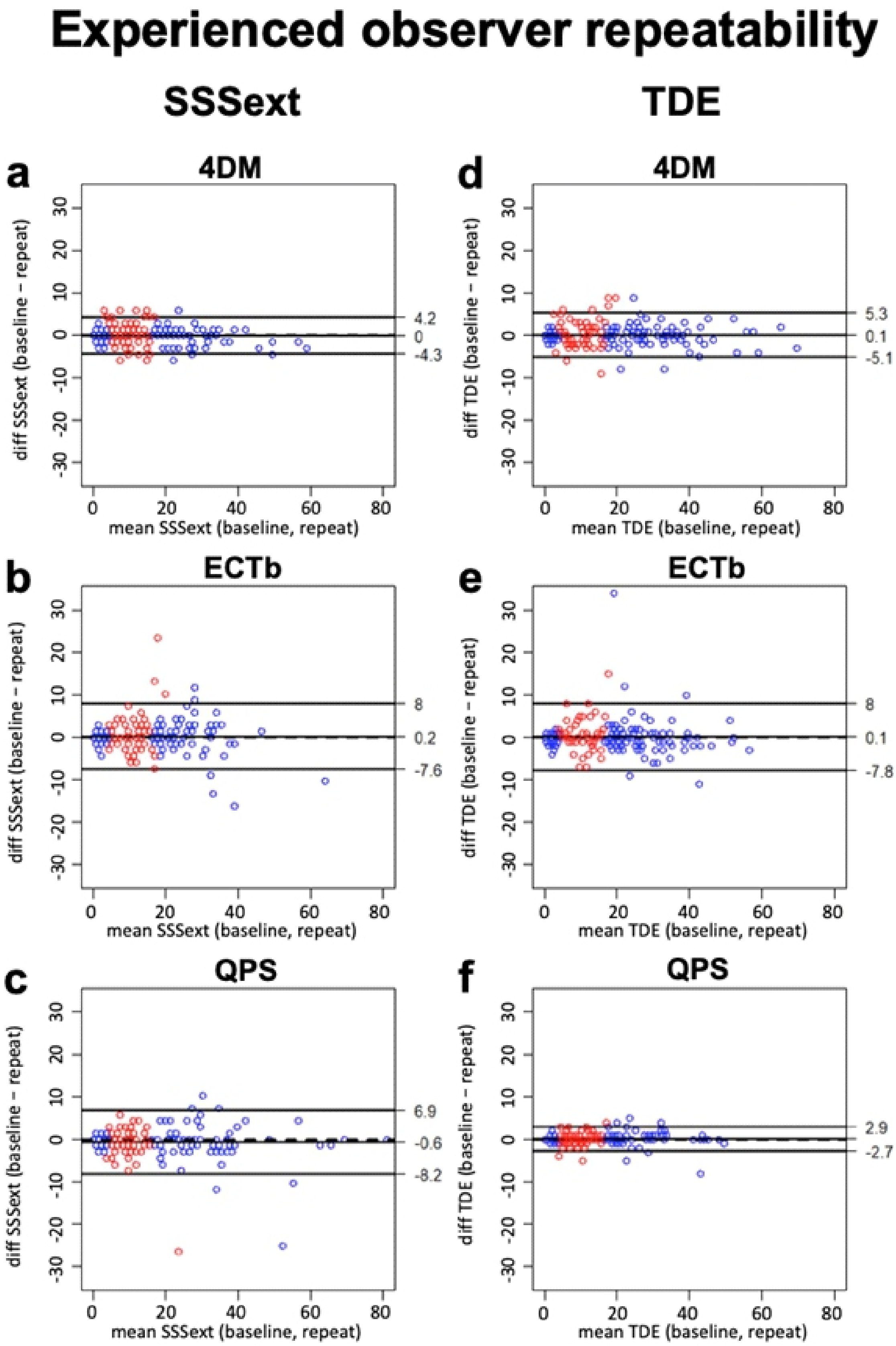

Figure 1. Bland-Altman plots of intra-observer repeatability of the experienced observer for SSSext (left column) and TDE (right column) calculations based on the 4DM (a, d), ECTb (b, e), and QPS (c, f) software packages. Cases with 5%≤"area of ischaemia"≤15% (as calculated with the first or second processing of data or the mean of paired measurements) are presented in red circles. The middle line illustrates the mean of paired measurements and the lower and upper lines illustrate the 95% limits of agreement.

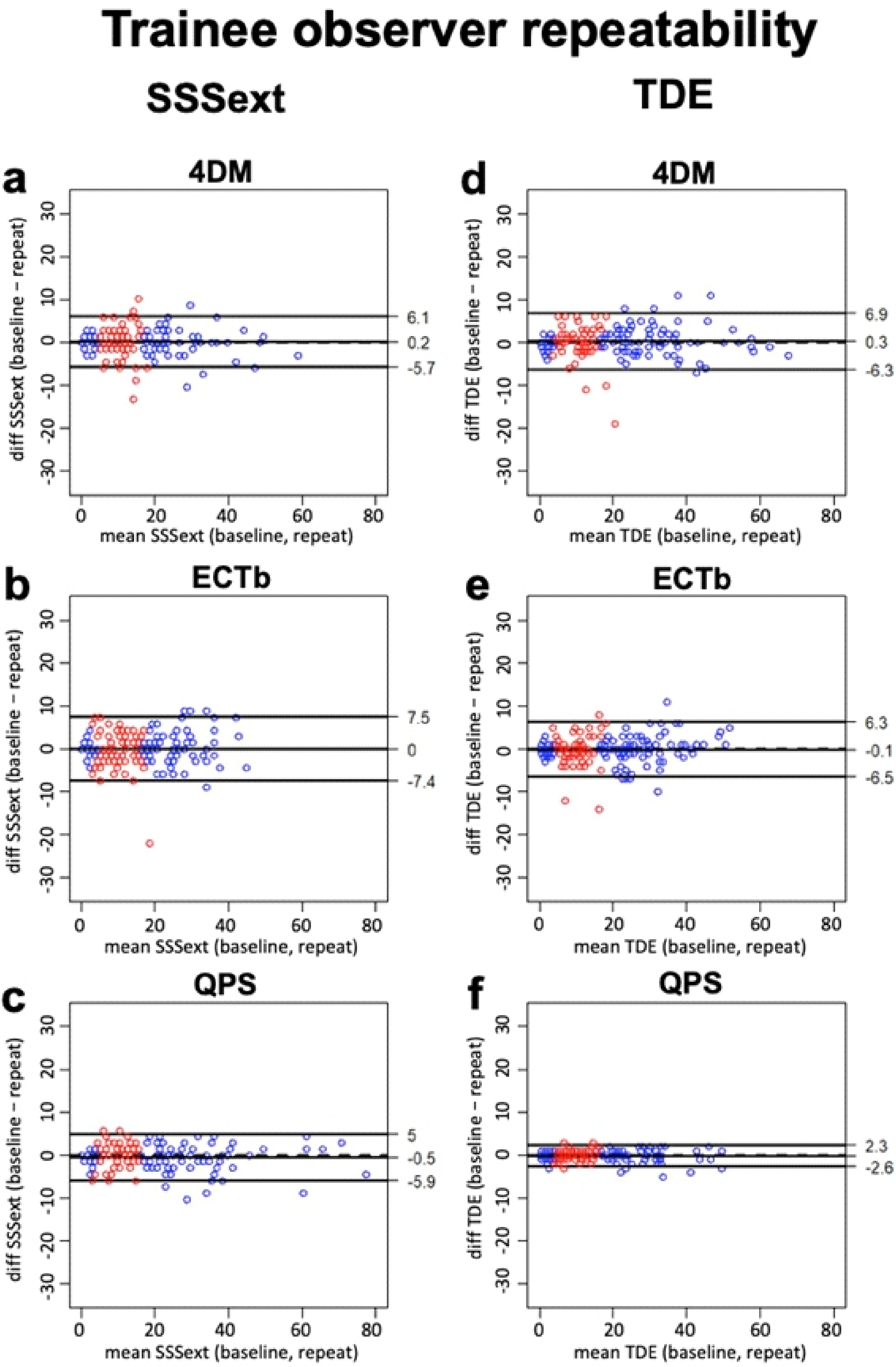

Figure 2. Bland-Altman plots of intra-observer repeatability of the trainee observer for SSSext (left column) and TDE (right column) calculations based on the 4DM (a, d), ECTb (b, e), and QPS (c, f) software packages. Cases with 5%≤area of ischaemia ≤15%" (as calculated with the first or second processing of data or the mean of paired me-asurements) are presented in red circles. The middle line illustrates the mean of paired measurements and the lower and upper lines illustrate the 95% limits of agre-ement.

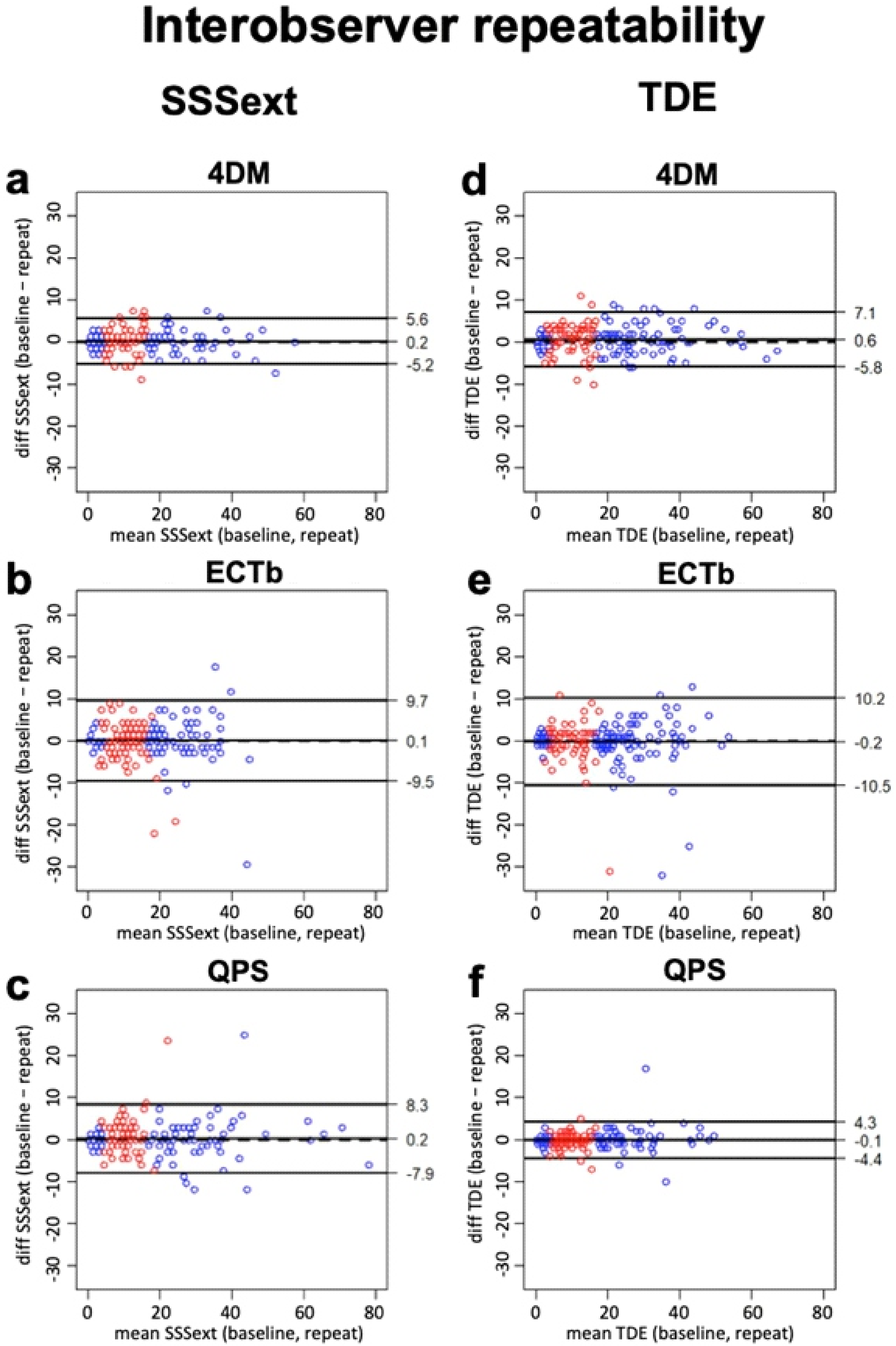

Figure 3. Bland-Altman plots of inter-observer repeatability for SSSext (left column) and TDE (right column) calculations based on the 4DM (a, d), ECTb (b, e), and QPS (c, f) software packages. Cases with "5% ≤ area of ischaemia ≤ 15%" (as calculated by either operator or the mean of paired measurements) are presented in red circles. The middle line illustrates the mean of paired measurements and the lower and upper lines illustrate the 95% limits of agreement.

Table 3. *Intra- and inter-observer repeatability data for SSSext and TDE with the three different software packages, presented as mean difference (bias) ±1.96 SD (LoA) according to the Bland-Altman analysis.*

	SSSext (%LV)	TDE (%LV)
Intra-observer repeatability		
Experienced physician		
4DM	0.0±4.3*,#	0.1±5.2*,#
ECTb	0.2±7.8#	0.1±7.9#,§
QPS	-0.6±7.6*	0.1±2.8*,§
Trainee		
4DM	0.2±5.9*	0.3±6.6*
ECTb	0.0±7.5*,#	-0.1±6.4#
QPS	-0.5±5.4#	-0.1±2.4*,#
Inter-observer repeatability		
4DM	0.2±5.4*,#	0.6±6.4*,#
ECTb	0.1±9.6#,§	-0.2±10.3#,§
QPS	0.2±8.1*,§	-0.1±4.3*,§

Abbreviations as in Table 2
**,#,§P<0.05 for the paired comparisons of the 95% limits of agreement (LoA) for each parameter, e.g. for the TDE assessments by the experienced physician (intra-observer repeatability) the QPS package shows significantly narrower LoA than both 4DM and ECTb, while 4DM shows significantly narrower LoA than ECTb. The same approach holds for the other comparisons.*

Table 4. *Intra- and inter-observer agreement with Lin's concordance correlation coefficient (CCC) for SSSext and TDE for both observers with the three different software packages.*

	SSSext (%LV)	TDE (%LV)
Intra- observer repeatability		
Experienced physician		
4DM	Substantial (0.983)	Substantial (0.984)
ECTb	Moderate (0.934)	Substantial (0.952)
QPS	Substantial (0.965)	Almostperfect (0.991)
Trainee		
4DM	Substantial (0.967)	Substantial (0.975)
ECTb	Moderate (0.932)	Substantial (0.965)
QPS	Substantial (0.982)	Almostperfect (0.994)
Inter- observer repeatability		
4DM	Substantial (0.973)	Substantial (0.976)
ECTb	Poor (0.894)	Moderate (0.915)
QPS	Substantial (0.960)	Substantial (0.980)

of the ECTb (Tables 3,4). In terms of inter-observer repeatability, 4DM provided the best results. Concerning the TDE, the QPS program offered the best degree of intra- and inter-observer agreement (Tables 3, 4). Largely, no remarkable differences were noted between the experienced and trainee observers, similarto a certain series [21], but dis similar to other previously published repeatability data[9] based on visual semi-quantitative analysis.

A large body of evidence attests to a favourable intra- and inter-observer interpretative reproducibility of myocardial perfusion SPECT using thallium-201, ^{99m}Tc-sestamibi,or dual-isotope imaging [22-30]. Kappa values ranged 0.70-0.81 and 0.62-0.89 and absolute agreement was attained in 85%-91% and 81%-92% of cases for intra- and inter-observer agreement, respectively. However, images in those studies were assessed visually, and interpretation concurrence was judged in a binary fashion (normal-abnormal, or ischaemia-scar). As expected, the inter-observer agreement for the presence or absence of reversibility was higher than the agreement for more detailed exact segmental scoring (kappa values 0.89 versus 0.71 and absolute agreement in 95% vs. 84% of cases, respectively) [24]. Moreover, in a different setting, comprising two sequential exercise imaging tests, in which the extent of ischaemia was ranked by visual analysis, the inter-observer agreement rates were lower than those provided when the ischaemia categories were amalgamated into the presence or absence of reversibility [31].

The software programs differ in many ways including the algorithms used to define the left ventricular valve plane, analyze circumferential profiles, generate polar maps, normalize patient data, determine the threshold of abnormality, assign segmental scores and quantify ischaemia based on normal databases used for comparison. An additional source of uncertainty may be introduced by the manual adjustment of software processes, which is not infrequent [16]. Most previous publications have reported considerable differences in the quantification of myocardial perfusion defects with 4DM, ECTb, and QPS, discouraging inter change ability for serial imaging purposes [32-35].

Regarding the repeatability of popular software packages in quantitatively assessing myocardial perfusion, only QPS has been systematically validated. High reproducibility of standard perfusion variables has been reported with the visual or quantitative assessment of repeat MPS scans with some supremacy of the quantitative approach [16, 36, 37]. Our data parallel published findings for the quantitative determination of TDE, but reveal certain variations in the quantitative assessment of SSSext (Table 3). Unfortunately, detailed comparisons cannot be made since older publications assessed SSS visually and also tested the automated derived total perfusion deficit, which is a different variable representing the extent and severity of ischaemia, in a test-retest setting.

For clinical purposes, a high degree of confidence in the repeatability of quantitative measurements is of paramount importance [15]. In this respect, the estimated thresholds for appreciating differences in stress defects, using either SSSext or TDE, are set by the 95% LoA (Table3), which span from ±2.4% to ±7.8% and from ±4.3% to ±10.3% in repeat processing of a myocardial SPECT acquisition by the same or different operators, respectively. This corresponds to a substantial proportion or surpasses the cut-off value of the defect size recommended for management decisions (10% of the LV) [3]. The difficulty in reliably determining the extent of ischaemia specifically in cases in the critical range of 5%-15% of the LV, is illustrated in Figures 1-3. In this subgroup the results of the Bland-Altman analysis practically replicated those of the entire population and denoted the level of uncertainty in determining the extent of abnormality with software packages in patients with small to moderate perfusion defects, which may be the vast majority [38].Notably, in a test-retest setting, the 95% LoA of the paired measurements are expected to be wider. Although the available quantitative techniques may be able to establish small differences between groups, they may be less sensitive in discerning small changes in serial imaging of individual patient [39].

It is worth adding that TDE measurements consistently provided significantly larger values of the extent of perfusion abnormality than SSSext with the 4DM and ECTb packages, although statistical significance was attained invariably only with 4DM. Conversely, QPS software provided a significantly smaller defect size with TDE than with SSSext. This discrepancy may be partly explained by the fact that SSSext is an index of the extent of abnormality, incorporating the size and severity of defects, unlike TDE, which only delineates the area of abnormality. Nevertheless, such observations have not been reported before and they may be worrisome and merit further elucidation because the threshold of 10% of the LV to select treatment management has been derived from a series using scoring values and not defect extents. Hence, liberally using SSSext or TDE with any software package for the quantitative assessment of a perfusion defect may potentially lead to misjudgment and mistreatment [3].Quantitative perfusion SPECT software, in particular, despiteoffering the best interpretive repeatability with TDE, may also significantly underestimate the extent of hypo perfusion with that variable (Table 2).

The results were not influenced by gender. However, there was a trend for improved inter-observer repeatability in obese patients compared with their non-obese counterparts, which reached statistical significance in a few paired measurements. The reason for this finding is not apparent. To a certain extent this may be an effect of the normal databases used by the software packages which would be more suitable for patients with characteristics similar to those of the population in whichthey were developed [40]. In this respect, the prevalence of obesity is remarkably higher in the US than in the country where this work was conducted and this is particularly accentuated in males, which comprised the majority of patients enrolled [41]. Thus, provided that the population composition is reflected in normal databases, inter-observer agreement may have been favoured in obese study participants. This, in turn, implies that the application of normal institutional database may improve the performance of the software, unlike earlier reports [8,33]. Nonetheless, even more favourable results in obese patients show a considerable degree of uncertainty in determining the amount of perfusion abnormality.

The use of vendor databases may be considered as a limitation of our study. Ideally, MPS should be assessed with a normal database usinga similar protocol, imaging system and population characteristics, as discussed previously. However, to our understanding, most physicians use a normal database provided by the vendor rather than their own institutional version [34]. Nevertheless, in an effort to account for potential constraints inflicted by the particular imaging protocols used to build up these databases, patients submitted to both stress/rest and rest/stress imaging protocols were enrolled and provided similar intra-/inter-observer agreement. Likewise, our work compares the quantification of perfusion defects by software packages in the real world, not the diagnostic or prognostic performance of these packages, by using normal databases that are available to all users of the software programs. We also tested only two variables, SSSext and TDE, because they both address the same clinicalentity, as a percentage of the LV, and thus they can be compared directly to each other. In addition, SSS and the extent of hypoperfusion at stress are invariably calculated using all packages, and both have been used in clinical studies [3,17].

In conclusion, these data provide evidence of the advantages and limitations of the three most widely used software packages in the quantitative assessment of defect size in stress myocardial perfusion imaging, in terms of intra- and inter-observer repeatability. In this respect, QPS appears to be preferable for the quantitation of the extent of a perfusion abnormality, although a significant advantage over 4DM cannot be appreciated. No substantial differences were noted between the experienced and trainee observer. Regarding repeatability, the selection of a particular software package for the quantitative assessment of MPS requires careful validation and judicious clinical consideration. Nevertheless, physicians should be aware of the degree of uncertainty in the quantitative determination of the amount of hypo perfused myocardium, particularly in the critical range of clinical decision-making, because of the limitations imposed by intra- and inter-observer agreement.

The authors declare that they have no conflicts of interest.

Bibliography

1. Bourque JM, Beller GA. Stress Myocardial Perfusion Imaging for Assessing Prognosis: An Update. *JACC Cardiovasc Imaging* 2011; 4: 1305-19.
2. Miller TD, Askew JW, Herrmann J. Assessing clinical impact of myocardial perfusion studies: ischemia or other prognostic indicators? *Curr Cardiol Rep* 2014; 16: 465.
3. Hachamovitch R, Berman DS, Kiat H et al. Comparison of the short-term survival benefit associated with revascularization compared with medical therapy in patients with no prior coronary artery disease undergoing stress myocardial perfusion single photon emission computed tomography. *Circulation* 2003; 107: 2900-7.
4. Hachamovitch R, Rozanski A, Shaw LJ et al. Impact of ischaemia and scar on the therapeutic benefit derived from myocardial revascularization vs. medical therapy among patients undergoing stress-rest myocardial perfusion scintigraphy. *Eur Heart J* 2011; 32: 1012-24.
5. Neumann FJ, Sousa-Uva M, Ahlsson A et al. ESC Scientific Document Group. 2018 ESC/EACTS Guidelines on myocardial revascularization. *Eur Heart J* 2019; 40: 87-165.
6. Candell-Riera J, Santana-Boado C, Bermejo B et al. Interhospital observer agreement in interpretation of exercise myocardial Tc-99m tetrofosmin SPECT studies. *J Nucl Cardiol* 2001; 8: 49-57.
7. Akesson L, Svensson A, Edenbrandt L. Operator dependent variability in quantitative analysis of myocardial perfusion images. *Clin Physiol Funct Imaging* 2004; 24: 374-9.
8. Wolak A, Slomka PJ, Fish MBet al. Quantitative myocardial-perfusion SPECT: Comparison of three state-of-the-art software packages. *J Nucl Cardiol* 2008; 15: 27-34.
9. Dondi M, Rodella C, Giubbini Ret al. I-MAP investigators. Inter-reader variability of SPECT MPI readings in low- and middle-income countries: Results from the IAEA-MPI Audit Project (I-MAP). *J Nucl Cardiol* 2020; 27: 465-78.
10. Lindahl D, Lanke J, Lundin A et al. Improved classifications of myocardial bull's-eye scintigrams with computer-based decision support system. *J Nucl Med* 1999; 40: 96-101.
11. Germano G, Kavanagh PB, Waechter P et al. A new algorithm for the quantitation of myocardial perfusion SPECT. I: Technical principles and reproducibility. *J Nucl Med* 2000; 41: 712-9.
12. Garcia EV, Faber TL, Cooke CD et al. The increasing role of quantification in clinical nuclear cardiology: the Emory approach. *J Nucl Cardiol* 2007; 14: 420-32.
13. Germano G, Kavanagh PB, Slomka PJ et al. Quantitation in gated perfusion SPECT imaging: The Cedars-Sinai approach. *J Nucl Cardiol* 2007; 14: 433-54.
14. Ficaro EP, Lee BC, Kritzman JN, Corbett JR. Corridor4 DM: The Michigan method for quantitative nuclear cardiology. *J Nucl Cardiol* 2007; 14: 455-65.
15. Germano G. Quantitative measurements of myocardial perfusion and function from SPECT (and PET) studies depend on the method used to perform those measurements. *J Nucl Cardiol* 2018; 25: 925-8.
16. Germano G, Kavanagh PB, Fish MB et al. "Same-patient processing" for multiple cardiac SPECT studies. 2. Improving quantification repeatability. *J Nucl Cardiol* 2016; 23: 1435-41.
17. Mahmarian JJ, Shaw LJ, Filipchuk NGal. INSPIRE Investigators. A multinational study to establish the value of early adenosine technetium-99m sestamibi myocardial perfusion imaging in identifying a low-risk group for early hospital discharge after acute myocardial infarction. *J Am Coll Cardiol* 2006; 48: 2448-57.
18. Henzlova MJ, Duvall WL, Einstein AJ et al. ASNC imaging guidelines for SPECT nuclear cardiology procedures: Stress, protocols, and tracers. *J Nucl Cardiol* 2016; 23(3): 606-39.
19. Dorbala S, Ananthasubramaniam K, Armstrong IS et al. Single Photon Emission Computed Tomography (SPECT) Myocardial Perfusion Imaging Guidelines: Instrumentation, Acquisition, Processing, and Interpretation. *J Nucl Cardiol* 2018; 25: 1784-846.
20. McBride GB. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. Available at: https://www.medcalc.org/download/pdf/McBride2005.pdf. Accessed at 15 Dec 2021.
21. Ejlersen JA, May O, Mortensen J et al. Stress-only myocardial perfusion scintigraphy: a prospective study on the accuracy and observer agreement with quantitative coronary angiography as the gold standard. *Nucl Med Commun* 2017; 38: 904-11.
22. Alazraki NP, Krawczynska EG, DePuey EG et al. Reproducibility of thallium-201 exercise SPECT studies. *J Nucl Med* 1994; 35: 1237-44.
23. Mahmarian JJ, Moye LA, Verani MS et al. High reproducibility of myocardial perfusion defects in patients undergoing serial exercise thallium-201 tomography. *Am J Cardiol* 1995; 75: 1116-9.
24. Berman DS, Kiat H, Friedman JD et al. Separate acquisition rest thallium-201/stress technetium-99m sestamibi dual-isotope myocardial perfusion single-photon emission computed tomography: a clinical validation study. *J Am Coll Cardiol* 1993; 22: 1455-64.
25. Weinmann P, Foult JM, Le Guludec Det al. Dual-isotope myocardial imaging: feasibility, advantages and limitations. Preliminary report on 231 consecutive patients. *Eur J Nucl Med* 1994; 21: 212-5.

26. Johansen A, Høilund-Carlsen PF, Christensen HW et al. Observer variability in the evaluation of dual-isotope Tl-201/Tc-99m sestamibi rest/stress myocardial perfusion SPECT in men and women with known or suspected stable angina pectoris. *J Nucl Cardiol* 2004; 11: 710-8.
27. Suleiman K, Desai D, Raval U et al. The effect of training on the interpretation of ^{99m}Tc-sestamibi myocardial perfusion SPET in patients with suspected coronary artery disease. *Nucl Med Commun* 1997; 18: 922-6.
28. Golub RJ, Ahlberg AW, McClellan JR et al. Interpretive reproducibility of stress Tc-99m sestamibi tomographic myocardial perfusion imaging. *J Nucl Cardiol* 1999; 6: 257-69.
29. Johansen A, Gaster AL, VejeA et al. Interpretive intra- and inter-observer reproducibility of rest/stress ^{99m}Tc-sestamibi myocardial perfusion SPECT in a consecutive group of male patients with stable angina pectoris before and after percutaneous transluminal angioplasty. *Nucl Med Commun* 2001; 22: 531-7.
30. Danias PG, Ahlberg AW, Travin MI et al. Visual assessment of left ventricular perfusion and function with electrocardiography-gated SPECT has high intra-observer and inter-observer reproducibility among experienced nuclear cardiologists and cardiology trainees. *J Nucl Cardiol* 2002; 9: 263-70.
31. Iskandrian AE, Bateman TM, Belardinelli L et al. Adenosine versus regadenoson comparative evaluation in myocardial perfusion imaging: results of the ADVANCE Phase 3 Multi-Center International Trial. *J Nucl Cardiol* 2007; 14: 645-58.
32. Svensson A, Akesson L, Edenbrandt L. Quantification of myocardial perfusion defects using three different software packages. *Eur J Nucl Med Mol Imaging* 2004; 31: 229-32.
33. Guner LA, Karabacak NI, Cakir T et al. Comparison of diagnostic performances of three different software packages in detecting coronary artery disease. *Eur J Nucl Med Mol Imaging* 2010; 37: 2070-8.
34. Johansson L, Lomsky M, Marving J et al. Diagnostic evaluation of three cardiac software packages using a consecutive group of patients. *EJNMMI Res* 2011; 1: 22.
35. Ather S, Iqbal F, Gulotta J et al. Comparison of three commercially available softwares for measuring left ventricular perfusion and function by gated SPECT myocardial perfusion imaging. *J Nucl Cardiol* 2014; 21: 673-81.
36. Berman DS, Kang X, Gransar H et al. Quantitative assessment of myocardial perfusion abnormality on SPECT myocardial perfusion imaging is more reproducible than expert visual analysis. *J Nucl Cardiol* 2009; 16: 45-53.
37. Xu Y, Hayes S, Ali I et al. Automatic and visual reproducibility of perfusion and function measures for myocardial perfusion SPECT. *J Nucl Cardiol* 2010; 17: 1050-7.
38. Shaw LJ, Berman DS, Maron DJ et al. Optimal medical therapy with or without percutaneous coronary intervention to reduce ischemic burden: Results from the clinical outcomes utilizing revascularization and aggressive drug evaluation (COURAGE) trial nuclear substudy. *Circulation* 2008; 117: 1283-91.
39. Iskandrian AE, Garcia EV, Faber T. Analysis of serial images: A challenge and an opportunity. *J Nucl Cardiol* 2008; 15: 23-6.
40. Rubeaux M, Xu Y, Germano G et al. Normal Databases for the Relative Quantification of Myocardial Perfusion. *Curr Cardiovasc Imaging Rep* 2016; 9: 22.
41. https://data.worldobesity.org/tables/prevalence-of-adult-overweight-obesity-2/.Accessed at 19 March 2022.