Automatic and visual reproducibility of perfusion and function measures for myocardial perfusion SPECT

Yuan Xu, PhD, Sean Hayes, MD, Iftikhar Ali, MD, Terrence D. Ruddy, MD, R. Glenn Wells, PhD, Daniel S. Berman, MD, Guido Germano, PhD, and Piotr J. Slomka, PhD

Background. We define the repeatability coefficients (RC) of key quantitative and visual perfusion and function parameters that can be derived by the QGS/QPS automated software and by expert visual observer from gated myocardial perfusion SPECT (MPS) scans.

Methods. Standard QGS/QPS algorithms have been applied to derive quantitative perfusion and function parameters in 200 99mTc-tetrofosmin rest/stress MPS scans, obtained in 100 consecutive patients who underwent 2 separate gated rest/stress scans on the same camera. Variables included stress, rest, and ischemic total perfusion deficit (TPD), ejection fraction, motion, and thickening. Visual perfusion/motion scores were derived by an expert reader using randomized scan order and normalized to % myocardium.

Results. Quantitative and visual parameters were highly reproducible with smaller RC for some quantitative measures as compared to visual measures ($P < .0001$). RC for quantitative measures were 3.3% for stress TPD, 1.8% for rest TPD, and 3.2% for ischemic TPD and for visual scoring 4.8% for stress, 3.8% for rest, and 4.3% for ischemic ($P \leq .002$). The results in each vessel territory showed that in the right coronary artery (RCA) territory the quantitative approach had improved reproducibility as compared to visual reading. Visual thickening scoring was more reproducible than motion scoring ($P < .0001$).

Conclusions. This study demonstrates that standard perfusion and function parameters derived from MPS by visual or quantitative analysis are highly reproducible with some advantages to the quantitative approach. (J Nucl Cardiol 2010;17:1050–7.)

Key Words: SPECT, myocardial perfusion imaging • SPECT, reproducibility

INTRODUCTION

Visual and quantitative perfusion and function assessments of myocardial perfusion SPECT (MPS) are important in detection and evaluation of coronary artery disease (CAD). Reproducibility analysis for these parameters is essential in understanding the significance of changes seen in serial studies. Most previous studies have reported the inter-/intra-observer agreement for one MPS test or the reproducibility of measures for retrospective studies. A recent study has reported the reproducibility of visual perfusion derived by current automated software for test–retest MPS studies obtained with the same radiopharmaceutical injection. The aim of this study is to comprehensively evaluate the reproducibility of both visual and automated quantifications of perfusion and function parameters from MPS stress and rest scans that were repeated in the same scanning session.

MATERIALS AND METHODS

Patients

One hundred studies with repeated stress and rest MPS acquisitions were obtained from July 2007 to March 2008 at Nuclear Cardiology Diagnostic Imaging Laboratory,
Reproducibility of SPECT measures

Protocol. For rest images, the injected dose of 99mTc-tetrofosmin was performed with a 99mTc-tetrofosmin rest/stress 1-day protocol. For direct comparison with a recently introduced measure—total perfusion deficit (TPD), all visual perfusion scores were also normalized to percent of the total myocardium (the summed scores divided by 68 \[4 \times 17\]) and multiplied by 100. Local summed and normalized scores were taken into account. The expert was also blinded to any computer-generated myocardial perfusion quantification results and paired information for the randomized patient identification codes. However, all available image data including raw projections gated stress and resting non-corrected scans were considered during scoring. The observer could also modify the default assignment of segments to each specific vascular territory. In addition, to eliminate visual bias, each image in first scanning session was presented to observer after 6 months for testing the intra-observer reproducibility. Subsequently, the visual summed scores for stress (VSSS) and summed scores for rest (VSRS) were calculated by summing of respective segmental perfusion scores. Visual summed difference score (VSDS) was also computed as the difference between VSSS and VSRS (VSSS – VSRS). Visual motion (VM) and thickening (VT) scores were obtained from stress image by summing of all corresponding segmental scores. Previous established abnormality threshold for VSSS (\(>3\)) was assumed for observer scoring. For direct comparison with a recently introduced measure—total perfusion deficit (TPD), visual perfusion scores were also normalized to percent of the total myocardium (the summed scores divided by 68 \[4 \times 17\]) and multiplied by 100. Local summed and normalized scores for each vascular territory were also obtained.

### Automated Quantification

QPS software computed TPD score by integrating the hypoperfusion severities below normal limits in polar map.
coordinates. Normal limits threshold was defined as 3.0 mean absolute deviations (approximately equivalent to 2.5 standard deviations, SD) for each polar map sample. TPD was measured independently at each stress and rest scan. Ischemic TPD (ITPD) was defined as the difference between stress TPD (STPD) and rest TPD (RTPD) (STPD – RTPD). Functional measures included ejection fraction (EF), ventricular wall motion (WM) (mm), and thickening (WT) (%), which were obtained from gated stress/rest scans using QGS software.12 TPD, WM, and WT values were also calculated for each vascular territory.

### Statistical Analysis

Continuous variables are denoted as mean ± SD. Bonferroni-corrected paired t-test was used to compare mean values of continuous variables of test–retest studies and χ² test was used to analyze categorical variables. To test reproducibility, several methods including linear regression, Pearson correlation analysis, and the Bland–Altman method were employed to determine the reproducibility between the same assessments of two repeated studies. To obtain visual assessment of intra-observer reproducibility, the reading scores in first scanning session were compared with visual scores in the second scanning session. Repeatability coefficient (RC), defined as 1.96 SD of the differences between pairs of repeated measures, was also determined.13 A P < .05 was considered statistically significant.

### RESULTS

160 of 800 contours generated from QPS/QGS (gated image: 21%; ungated image: 19%) needed adjustment, as determined by experienced technologist. Most of the adjustments were subtle, only rearranging the valve-plane positioning. Only 13 cases were adjusted by giving a corrected initial ellipsoid. Eleven cases with incorrect ellipsoid determination had high shape quality control flag value (≥4.11).9

### Perfusion Results

For the perfusion visual scores, 17–18% of patients had ≥5% myocardium abnormal at stress in both studies (test: 17%; retest: 18%), and 10% of the patients had ≥5% myocardium abnormal at rest or ischemic myocardium in both studies. TPD showed similar distribution of abnormality as compared to visual scores. The distribution differences between test and retest studies in either visual scores or quantitative scores were not significant. Table 2 shows the reproducibility of perfusion results (global visual perfusion score, TPD, and corresponding results in vessel territories). Table 2A shows the reproducibility of global variables and correlation between test and retest studies. The correlations of visual perfusion interpretation were excellent for VSSS and VSRS (both r = 0.95) and good for VSDS (r = 0.81). Similar correlation values were also obtained for stress, rest, and ischemic TPD with r = 0.98, r = 0.99, and r = 0.85, respectively. As compared to visual measures, the correlation between test and retest for stress and rest quantitative perfusion parameters was significantly higher (0.98 vs 0.95: P = .001; 0.99 vs 0.95: P < .0001).

The reproducibility of global perfusion quantification analyzed by Bland–Altman method is also shown in Table 2A and Figure 1. In general, the bias, 95% limits of agreement, and SD were smaller for quantitative perfusion measures (stress, rest, and ischemic) compared to visual findings. In addition, Table 2A shows that both quantitative and visual perfusion parameters were highly reproducible with small repeatability coefficients (RC) (RC = 4.8% for VSSS%; RC = 3.8% for VSRS%; RC = 4.3% for VSDS%; RC = 3.3% for STPD; RC = 1.8% for RTPD, and RC = 3.2% for ITPD), all P ≤ .002. Among abnormal cases, Figure 2 shows that reproducibility of stress or ischemic quantitative parameter was smaller than that of corresponding visual parameter for all abnormality thresholds.

Most of the correlations between test and retest studies for visual and automated perfusion measures in each vessel territory were either excellent or good (≥0.7) (Table 2B). In general, the correlations of both visual and automated perfusion measures in the LAD were high (>0.95), the correlations of both measure methods in the LCX were good, and the correlations of both methods in the RCA were also high (>0.80). In addition, the automated measurements in the RCA were correlated better than those of visual measures (P < .0001).

Similar to the correlation analyses, the automated perfusion results by Bland–Altman method (Table 2B) show similar results as compared with visual measures. Both quantitative and visual perfusion parameters in each vessel territory were also highly reproducible with small repeatability coefficients (all RC ≤ 4.1%). The RC for STPD and RTPD in RCA was significantly lower than that for normalized summed visual rest perfusion (P < .001).

### Functional Results

Global functional measures (Table 3A) showed high repeatability especially in thickening (VT: r = 0.94; WT: r = 0.91). Visual scoring was more reproducible for thickening score than for motion scores (P < .0001). In addition, stress EF had lower RC than rest EF (P = .0048). However, after removing 35 cases with operator adjusted contour, there are no significant differences between stress EF and rest EF (stress EF: r = 0.94, –1.88 ± 3.6%, RC = 6.4% and rest EF: r = 0.91, –0.66 ± 4.4%, RC = 8.7%).
Table 2. Global (A) and regional (B) reproducibility of perfusion parameters between test-retest studies in each vascular territory

|                     | Bias ± SD | 95% Limits of agreement | Correlation R value | RC |
|---------------------|-----------|--------------------------|---------------------|----|
| (A) Global          |           |                          |                     |    |
| VSSS                | 0.0 ± 1.7 | -3.3, 3.3                | 0.95                | 3.3|
| VSRS                | -0.1 ± 1.3| -2.7, 2.5                | 0.95                | 2.6|
| VSDS                | 0.08 ± 1.5| -2.8, 3.0                | 0.81                | 2.9|
| VSSS%               | 0.01 ± 2.5| -4.8, 4.8                | 0.95                | 4.8|
| VSRS%               | -0.10 ± 1.9| -3.9, 3.7              | 0.95                | 3.8|
| VSDS%               | 0.12 ± 2.2| -4.2, 4.4                | 0.81                | 4.3|
| STPD (%)            | 0.1 ± 1.7 | -3.2, 3.4                | 0.98*               | 3.3*|
| RTPD (%)            | -0.08 ± 0.9| -1.9, 1.8               | 0.99*               | 1.8*|
| ITPD (%)            | 0.19 ± 1.6| -3.0, 3.4                | 0.85                | 3.2*|
| (B) Regional        |           |                          |                     |    |
| LAD VSSS%           | 0.11 ± 1.66| -3.14, 3.36            | 0.96                | 3.3|
| LCX VSSS%           | -0.02 ± 1.48| -2.91, 2.87           | 0.78                | 2.9|
| RCA VSSS%           | -0.07 ± 2.10| -4.18, 4.04           | 0.81                | 4.1|
| LAD STPD (%)        | 0.2 ± 1.8 | -3.4, 3.8                | 0.96                | 3.5|
| LCX STPD (%)        | 0.08 ± 1.59| -3.04, 3.20            | 0.80                | 3.1|
| RCA STPD (%)        | -0.12 ± 0.53| -1.16, 0.91          | 0.98##              | 1.0##|
| LAD VSRS%           | 0.18 ± 1.03| -1.84, 2.20            | 0.98                | 2.0|
| LCX VSRS%           | 0.04 ± 0.73| -1.39, 1.48            | 0.72                | 1.4#|
| RCA VSRS%           | -0.32 ± 1.39| -3.05, 2.40          | 0.91                | 2.7|
| LAD RTPD (%)        | 0.0 ± 1.2 | -2.4, 2.4                | 0.98                | 2.4|
| LCX RTPD (%)        | -0.01 ± 1.04| -2.1, 2.0             | 0.71                | 2.0|
| RCA RTPD (%)        | -0.07 ± 0.47| -0.98, 0.85           | 0.98##              | 0.9##|

* P ≤ .002 vs corresponding value for normalized visual scores.
* P < .001 vs corresponding values for TPD.
** P < .0001 vs corresponding values for normalized visual scores.

Visual functional measures (VM and VT scores), automated raw motion, and thickening in each vessel territory were also analyzed (Table 3B). Table 3B shows that most of functional measures in each vessel territory have high or good correlation (≥0.72) except VM score in LCX and RCA area. Visual thickening score was more reproducible than visual motion score in each vessel territory (P ≤ .001). In addition, the RC for automated WM in each vessel territory was smaller than 2 mm and for automated WT in each vessel territory was smaller than 10.4%.

DISCUSSION

This is the first study, which comprehensively evaluates and compares intra-observer visual and quantitative reproducibility of wide range of function and perfusion parameters in MPS studies obtained with the same injection of radiopharmaceutical. We included most of visual and quantitative parameters, which have been used in clinical practice including TPD, WM, and thickening. It should be noted that in this study an experienced attending physician was reading blinded scans in two consecutive sessions within 6-month interval with the same display defaults and presets. Therefore, this represents the case for the best possible visual intra-observer reproducibility. We demonstrated that the software analysis showed equivalent or better reproducibility than this best possible (expert intra-observer) visual analysis.

In this study, the reproducibility of visual and quantitative MPS measurements has been evaluated from patients who underwent two separate gated stress/rest MPS studies. Most of correlations and reproducibility of perfusion and functional assessments of stress and rest MPS between these two consecutive studies were high by quantitative and by visual analyses. As compared with visual assessments, stress and rest perfusion automated measures had significantly higher correlations and smaller repeatability coefficients. Our
results also showed that VT scores were more reproducible than motion scores both globally and in local vessel territories.

Several previous serial studies have evaluated the variability of MPS assessments.\textsuperscript{2-6} Mahmarian et al\textsuperscript{3} showed a high correlation of myocardial perfusion defect size from 18 serial exercise thallium-201 SPECT images. Prigent et al\textsuperscript{6} also reported high reproducibility of the quantitative measure of percent hypoperfused myocardium by exercise thallium-201 SPECT images. Recently, our laboratory compared visual with automated assessments from 31 stable patients with abnormal stress MPI and suggested that automated quantification for abnormal perfusion was more

Figure 1. Differences of global perfusion defects. The left column shows the differences of visual perfusion results between test and retest; right column shows the differences of TPD results between test and retest. The first row shows the differences of normalized visual stress perfusion and STPD between test and retest; the second row shows the differences of normalized visual rest perfusion and RTPD between test and retest; the third row shows the differences of normalized visual ischemic perfusion and ITPD between test and retest.
reproducible than visual evaluation. In that study the RC of ITPD was found to be 5.7% for the studies repeated within 9-22 months and it is found to be 3.2% in our study with same-injection repeated studies. However in abnormal studies, the ITPD reproducibility is slightly higher at 4.3% (see Figure 2). In another recent study, the reproducibility of ischemic quantitative measurement was reported as 0.25 ± 3.81%. Therefore, the estimated RC at 95% confidence would be 7.5%. The values reported in our studies are lower, even in abnormal studies, which may be related to the differences in the software.

Compared to previous studies, this study utilized repeated stress and rest studies obtained in the same day with the same isotope injection to evaluate the reproducibility of perfusion and function parameters obtained from MPS. By using studies obtained on the same date with the same injection, we eliminated the effect of disease progression, changing patient medications, and other variations in the clinical state as well as the variability in the level of the exercise. Therefore, from the point of view of testing the visual or software analysis reproducibility, our data are not contaminated by the unrelated components and represent true evaluation of the analysis reproducibility.

This study also showed good correlation for visual and automated perfusion/functional assessments of SPECT MPI between these repeated studies, and higher reproducibility for some automated perfusion measures in stress and rest images. This study showed that reproducibility of stress or ischemic quantitative parameters in abnormal cases considered separately was significantly smaller than that of corresponding visual parameters. Besides evaluation of global parameters, quantifications in each vessel territory analyzed in our study showed that quantitative measures in RCA territory had some superiority as compared with visual measures. This might be related to slight patient movement, and artifacts variations in this location between these two repeated studies, which could affect expert’s evaluation during scoring.

There are several limitations in this study. The major limitation of this study is the small number of patients with abnormal perfusion and/or function included. In this study, we only compared software reproducibility to the intra-observer reproducibility. Presumably inter-observer reproducibility would be worse (with higher RC coefficients) than the intra-observer variability. However even in the case of intra-observer variability, we could demonstrate the superiority of the software supporting the overall conclusions of software analysis superiority as compared with visual analysis. In addition, although our data acquired during continuous sessions are not affected by the disease progress, patient medications, and other clinical state as well as the variability in the level of the exercise, the functional parameters such as EF, WM, and WT could be potentially affected by the timing of the acquisition. In particular, stress function could be affected by stunning and rest function could be affected by variable image quality (e.g., increased extra-cardiac uptake). Furthermore, in this study, not all of the standard variables generated from our software were evaluated for reproducibility. Some parameters such as transient ischemic dilation (TID) or diastolic function parameters were not evaluated due to nature of the imaging data obtained with single injection. TID would be associated with error in sequential imaging acquisitions and diastolic function parameters were only validated for the 16-gate data, whereas the data studied here were obtained with 8-gates.

**CONCLUSION**

This study demonstrates that standard perfusion and function parameters derived by expert visual or quantitative analysis are highly reproducible with significant advantages for the quantitative approach especially for the stress and ischemic perfusion variables.
Table 3. Global (A) and regional (B) reproducibility of functional parameters for each vascular territory

|                      | Bias ± SD | 95% Limits of agreement | Correlation R value | RC |
|----------------------|-----------|--------------------------|---------------------|----|
| (A) Global           |           |                          |                     |    |
| VM                   | 0.2 ± 5.5 | -10.4, 10.9              | 0.74                | 10.7 |
| VT                   | -0.2 ± 1.9| -3.9, 3.5                | 0.94                | 3.7  |
| WM (mm)              | -0.13 ± 0.63| -1.38, 1.11            | 0.89                | 1.3  |
| WT (%)               | -0.8 ± 4.3 | -9.2, 7.7                | 0.91                | 8.6  |
| Stress EF (%)        | -1.36 ± 3.6| -8.5, 5.8                | 0.94                | 7.1  |
| Rest EF (%)          | -0.07 ± 5.2| -10.2, 10.1              | 0.87                | 10.1 |
| Stress EF (%)*       | -1.88 ± 3.6| -8.9, 5.1                | 0.94                | 6.4  |
| Rest EF (%)*         | -0.66 ± 4.4| -9.2, 7.9                | 0.91                | 8.7  |
| (B) Regional         |           |                          |                     |    |
| LAD VM               | 0.2 ± 3.3 | -6.3, 6.6                | 0.84                | 6.5  |
| LCX VM               | 0.0 ± 0.8 | -1.6, 1.5                | 0.41                | 1.6  |
| RCA VM               | 0.1 ± 2.4 | -4.3, 4.7                | 0.56                | 4.7  |
| LAD WM (mm)          | -0.09 ± 0.75| -1.55, 1.37         | 0.90                | 1.5  |
| LCX WM (mm)          | 0.06 ± 0.94| -2.04, 1.65              | 0.87                | 1.9  |
| RCA WM (mm)          | -0.12 ± 0.86| -1.8, 1.57              | 0.84                | 1.7  |
| LAD VT               | 0.0 ± 0.9 | -1.6, 1.7                | 0.98                | 1.8  |
| LCX VT               | 0.0 ± 0.7 | -1.3, 1.3                | 0.72                | 1.4  |
| RCA VT               | -0.2 ± 1.3 | -2.8, 2.4                | 0.85                | 2.5  |
| LAD WT (%)           | 1.0 ± 5.2 | -11.2, 9.2               | 0.90                | 10.4 |
| LCX WT (%)           | -0.8 ± 5.2 | -10.9, 9.3              | 0.87                | 10.4 |
| RCA WT (%)           | -0.8 ± 4.5 | -9.6, 8.1                | 0.89                | 9.0  |

* The results without any observer’s contour adjustment on either scan (N = 65).

Acknowledgments

This research was supported in part by Grant RO1HL089765-01 from the National Heart, Lung, and Blood Institute/National Institutes of Health (NHLBI/NIH) (PI: Piotr Slomka). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NHLBI. We would like to thank James Gerlach and Mark Hyun for adjusting contours.

Open Access

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Min JK, Berman D. Anatomic and functional assessment of coronary artery disease: convergence of 2 aims in a single setting. Circ Cardiovasc Imaging 2009;2:163-5.

2. Danias PG, Ahlberg AW, Travin MI, et al. Visual assessment of left ventricular perfusion and function with electrocardiography-gated SPECT has high intraobserver and interobserver reproducibility among experienced nuclear cardiologists and cardiology trainees. J Nucl Cardiol 2002;9:263-70.

3. Mahmarian JJ, Moye LA, Verani MS, Bloom MF, Pratt CM. High reproducibility of myocardial perfusion defects in patients undergoing serial exercise thallium-201 tomography. Am J Cardiol 1995;75:1116-9.

4. Wackers FJ. Science, art, and artifacts: How important is quantification for the practicing physician interpreting myocardial perfusion studies? J Nucl Cardiol 1994;1:S109-17.

5. 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.

6. Prigent FM, Berman DS, Elashoff J, et al. Reproducibility of stress redistribution thallium-201 SPECT quantitative indexes of hypoperfused myocardium secondary to coronary artery disease. Am J Cardiol 1995;75:1116-9.

7. Ali I, Ruddy TD, Almagrahi A, Anstett FG, Wells RG. Half-time SPECT myocardial perfusion imaging with attenuation correction. J Nucl Cardiol 2009;50:554-62.

8. Hansen CL, Goldstein RA, Akinboboye OO, et al. Myocardial perfusion and function: Single photon emission computed tomography. J Nucl Cardiol 2007;14:e39-60.
9. Xu Y, Kavanagh P, Fish M, et al. Automated quality control for segmentation of myocardial perfusion SPECT. J Nucl Med 2009;50(9):1418-26.
10. Slomka PJ, Nishina H, Berman DS, et al. Automated quantification of myocardial perfusion SPECT using simplified normal limits. J Nucl Cardiol 2005;12:66-77.
11. Nishina H, Slomka PJ, Abidov A, et al. Combined supine and prone quantitative myocardial perfusion SPECT: method development and clinical validation in patients with no known coronary artery disease. J Nucl Med 2006;47:51-8.
12. Germano G, Kiat H, Kavanagh PB, et al. Automatic quantification of ejection fraction from gated myocardial perfusion SPECT. J Nucl Med 1995;36:2138-47.
13. Bland JM, Altman DG. Applying the right statistics: Analyses of measurement studies. Ultrasound Obstet Gynecol 2003;22:85-93.
14. Mahmarian JJ, Cerqueira MD, Iskandrian AE, et al. Regadenoson induces comparable left ventricular perfusion defects as adenosine: A quantitative analysis from the ADVANCE MPI 2 trial. JACC Cardiovasc Imaging 2009;2:959-68.