

# Clinical Interpretation Standards and Quality Assurance for the Multicenter PET/CT Trial Rubidium-ARMI

Jennifer M. Renaud<sup>1</sup>, Ilias Mylonas<sup>1</sup>, Brian McArdle<sup>1</sup>, Taylor Dowsley<sup>1</sup>, Kathy Yip<sup>2</sup>, Eric Turcotte<sup>3</sup>, Jean Guimond<sup>4</sup>, Mikael Trottier<sup>4</sup>, Philippe Pibarot<sup>4</sup>, Conor Maguire<sup>5</sup>, Lucille Lalonde<sup>5</sup>, Karen Gulenchyn<sup>6</sup>, Gerald Wisenberg<sup>7</sup>, R. Glenn Wells<sup>1</sup>, Terrence Ruddy<sup>1</sup>, Benjamin Chow<sup>1</sup>, Rob S.B. Beanlands<sup>1</sup>, and Robert A. deKemp<sup>1</sup>

<sup>1</sup>University of Ottawa Heart Institute, Ottawa, Ontario, Canada; <sup>2</sup>KMH Cardiology and Diagnostic Centres, Mississauga, Ontario, Canada; <sup>3</sup>Centre Hospitalier Universitaire de Sherbrooke, Quebec, Canada; <sup>4</sup>Institut Universitaire de Cardiologie et de Pneumologie de Québec, Quebec, Canada; <sup>5</sup>University of Alberta Hospital, Edmonton, Alberta, Canada; <sup>6</sup>St. Joseph's Healthcare, Hamilton, Ontario, Canada; and <sup>7</sup>St. Joseph's Health Care, London, Ontario, Canada

Rubidium-ARMI (<sup>82</sup>Rb as an Alternative Radiopharmaceutical for Myocardial Imaging) is a multicenter trial to evaluate the accuracy, outcomes, and cost-effectiveness of low-dose <sup>82</sup>Rb perfusion imaging using 3-dimensional (3D) PET/CT technology. Standardized imaging protocols are essential to ensure consistent interpretation.

**Methods:** Cardiac phantom qualifying scans were obtained at 7 recruiting centers. Low-dose (10 MBq/kg) rest and pharmacologic stress <sup>82</sup>Rb PET scans were obtained in 25 patients at each site. Summed stress scores, summed rest scores, and summed difference scores (SSS, SRS, and SDS [respectively] = SSS–SRS) were evaluated using 17-segment visual interpretation with a discretized color map. All scans were coread at the core lab (University of Ottawa Heart Institute) to assess agreement of scoring, clinical diagnosis, and image quality. Scoring differences greater than 3 underwent a third review to improve consensus. Scoring agreement was evaluated with intraclass correlation coefficient (ICC-r), concordance of clinical interpretation, and image quality using  $\kappa$  coefficient and percentage agreement. Patient <sup>99m</sup>Tc and <sup>201</sup>Tl SPECT scans ( $n = 25$ ) from 2 centers were analyzed similarly for comparison to <sup>82</sup>Rb. **Results:** Qualifying scores of SSS = 2, SDS = 2, were achieved uniformly at all imaging sites on 9 different 3D PET/CT scanners. Patient scores showed good agreement between core and recruiting sites: ICC-r = 0.92, 0.77 for SSS, SDS. Eighty-five and eighty-seven percent of SSS and SDS scores, respectively, had site–core differences of 3 or less. After consensus review, scoring agreement improved to ICC-r = 0.97, 0.96 for SSS, SDS ( $P < 0.05$ ). The agreement of normal versus abnormal (SSS  $\geq 4$ ) and nonischemic versus ischemic (SDS  $\geq 2$ ) studies was excellent: ICC-r = 0.90 and 0.88. Overall interpretation showed excellent agreement, with a  $\kappa = 0.94$ . Image quality was perceived differently by the site versus core reviewers (90% vs. 76% good or better;  $P < 0.05$ ). By comparison, scoring agreement of the SPECT scans was ICC-r = 0.82, 0.72 for SSS, SDS. Seventy-six and eighty-eight percent of SSS and SDS scores, respectively, had site–core differences of 3 or less. Consensus review again improved scoring agreement to ICC-r = 0.97, 0.90 for SSS, SDS ( $P < 0.05$ ). **Conclusion:** <sup>82</sup>Rb myocardial perfusion imaging protocols were implemented with highly repeatable interpretation in centers using 3D PET/CT technology, through an effective standardization and quality

assurance program. Site scoring of <sup>82</sup>Rb PET myocardial perfusion imaging scans was found to be in good agreement with core lab standards, suggesting that the data from these centers may be combined for analysis of the rubidium-ARMI endpoints.

**Key Words:** positron emission tomography; myocardial perfusion imaging; <sup>82</sup>Rb

**J Nucl Med 2014; 55:58–64**

DOI: 10.2967/jnumed.112.117515

Myocardial perfusion imaging (MPI) using <sup>201</sup>Tl- and <sup>99m</sup>Tc-labeled tracers is an accepted indication for detection of obstructive coronary artery disease and stratification of patients at risk for adverse cardiovascular events (1,2). <sup>82</sup>Rb is an alternative isotope with the lowest radiation dose among perfusion imaging tracers (0.7–1.3 mSv/GBq) (3,4) and is considered to have superior accuracy and incremental prognostic value (5–10).

Rubidium-ARMI (<sup>82</sup>Rb as an Alternative Radiopharmaceutical for Myocardial Imaging) is a multicenter trial to evaluate the accuracy, clinical outcomes, and cost-effectiveness of low-dose <sup>82</sup>Rb MPI using 3D PET/CT, compared with conventional <sup>99m</sup>Tc and <sup>201</sup>Tl SPECT imaging (10). Standardization of acquisition and interpretation methods is essential to allow consistent analysis of pooled data from multiple participating centers. A single core site (University of Ottawa Heart Institute) had previous experience performing <sup>82</sup>Rb PET in Canada because the tracer was not yet approved for sale. Therefore, an initial process of knowledge transfer was proposed to establish the training and standardized procedures for high-quality <sup>82</sup>Rb perfusion imaging and interpretation at nuclear imaging centers across Canada. We hypothesized that a comprehensive quality assurance (QA) program would achieve low interobserver variability between the new <sup>82</sup>Rb imaging sites and the QA core site.

## MATERIALS AND METHODS

### Qualifying Scans

**Image Acquisition.** Standards were established using the Discovery 690 PET/CT system (GE Healthcare) at the core site (University of Ottawa Heart Institute) (8,9). Qualifying scans simulating normal rest and abnormal stress <sup>82</sup>Rb MPI were then obtained at all sites using an anthropomorphic torso phantom (Data Spectrum) to standardize

Received Dec. 10, 2012; revision accepted Jul. 31, 2013.

For correspondence or reprints contact: Robert deKemp, University of Ottawa Heart Institute, 40 Ruskin St., Ottawa, ON, Canada K1Y 4W7.

E-mail: radekemp@ottawaheart.ca

Published online Nov. 18, 2013.

COPYRIGHT © 2014 by the Society of Nuclear Medicine and Molecular Imaging, Inc.

reconstructed image resolution, perfusion defect contrast, and CT attenuation correction (CTAC) image alignment. The phantom contained heart and liver inserts allowing scatter from abdominal organs to be simulated (Fig. 1A). With the phantom placed in a prone position, a CT scout scan was obtained to center the heart in the field of view. The phantom was removed from the bed, and 1,000–1,500 MBq of  $^{82}\text{Rb}$  activity were infused rapidly into the water-filled liver cavity. After vigorous mixing, 60 mL were withdrawn from the liver and injected directly into the heart wall chamber and the remaining volume filled with water. The resulting 2:1 activity concentration in liver:myocardium simulated image contrast observed at rest, when the liver, spleen, pancreas, or stomach wall is often visualized. The phantom was repositioned on the scanner bed, and 2 min after infusion an 8-min scan was acquired, followed by a CTAC scan. To simulate stress-induced ischemia, a 1-cm<sup>3</sup> transmural plastic defect was placed in the inferior wall of the myocardium chamber. The rubidium imaging procedure was repeated, except 120 mL were withdrawn from the liver and injected into the heart wall, resulting in liver:myocardium stress contrast of 1:1.

**Image Reconstruction.** Tracer uptake images were reconstructed using the vendor default iterative reconstruction settings with 12-mm postfiltering and all corrections enabled. Images were corrected explicitly for the 776.5-keV cascade/prompt  $\gamma$  emissions from  $^{82}\text{Rb}$  decay (~15% abundance) on some PET systems (Table 1). Prompt  $\gamma$  emissions recorded in coincidence with annihilation photons produce a background signal distinct from scatter and randoms, reported to produce septal artifacts on some 3D PET systems (11). On other

systems without explicit correction, a 50-cm CTAC field of view was used to minimize the potential prompt  $\gamma$  effects. Vendor-specific fusion display of the PET and CTAC images was used to correct or verify alignment of the CT images for proper attenuation correction. Sites were instructed to verify that the 50% PET activity contour fell within the CT soft-tissue contour on the fusion display (Figs. 1B and 1C).

**Semiquantitative Analysis.** The uptake defect summed rest score (SRS), summed stress score (SSS), and summed difference score (SDS) was computed automatically with Corridor4DM (INVIA), comparing scans against a simulated database with a uniform mean of 75% and SD of 15%. Defect scores were assigned on a 0–4 scale (0, normal tracer uptake; 4, absent tracer) of defect severity in 17 myocardial segments according to the thresholds shown in Figure 2.

#### Patient Scans

**Population.**  $^{82}\text{Rb}$  PET images acquired consecutively from the trial start date at each site (May 2010 to February 2012) were evaluated from 25 patients enrolled at 6 of the sites

and 24 patients from the seventh site ( $n = 174$ ); 1 patient could not complete the stress scan because of claustrophobia. The  $^{99\text{m}}\text{Tc}$  SPECT images of 25 patients acquired at a single recruiting site from October to November 2011 and  $^{201}\text{Tl}$  SPECT images of 25 patients acquired at another site during the  $^{99\text{m}}\text{Tc}$  shortage (June to September 2009) were assessed. All patients were referred for clinically indicated myocardial perfusion scans for diagnosis or risk stratification of coronary artery disease. Summary demographic data are presented for PET and SPECT patients in Table 2 and by site in Supplemental Table 1 (supplemental materials are available at <http://jnm.snmjournals.org>). The study was approved by the research ethics boards at all participating centers. All patients signed a written informed consent form before enrollment.

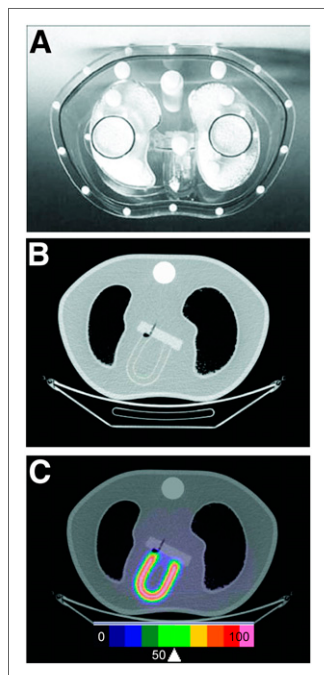
Patients were instructed to fast overnight and abstain from caffeine and theophylline-containing medications for 12 h before the test as per guidelines of the American Society of Nuclear Cardiology (12). Anti-anginal medications ( $\beta$ -blockers, calcium antagonists, and nitrates) were withheld on the morning of the study.

**PET Imaging.** Patients underwent low-dose rest and dipyridamole stress  $^{82}\text{Rb}$  PET MPI; 10 MBq/kg of body weight were infused over 30 s using a custom infusion system (13). A 6- to 8-min static scan was started 2 min after injection, when the randoms rate was less than two thirds the total coincidence counting rate. The acquisition sequence was rest CTAC, rest PET, dipyridamole (0.140 mg/kg/min  $\times$  4–5 min), stress PET, aminophylline (optional), and stress CTAC (Fig. 3). The CTAC scans were fast helical (<5 s), low-dose scans acquired after breath-hold or normal end expiration. Static images were reconstructed using the same methods as described above.

**SPECT Imaging.** Patients imaged with  $^{201}\text{Tl}$  SPECT underwent a 1-d stress-redistribution study with a single injection of 130 MBq at peak dipyridamole (0.140 mg/kg/min  $\times$  5 min) stress. Stress scans were acquired on an Infinia Hawkeye 4 dual-head SPECT/CT camera (GE Healthcare) at 10–15 min after injection and after 4 h of redistribution. Thirty projections were obtained at 45 s/projection over 180° rotation. Static images were reconstructed using filtered back projection with a 10th-order Butterworth filter, 0.35 cycles/cm cutoff frequency, and no attenuation correction as per standard clinical practice at the site.

$^{99\text{m}}\text{Tc}$ -sestamibi SPECT patients underwent a 2-d stress–rest protocol. Eight minutes after dipyridamole injection (0.140 mg/kg/min  $\times$  5 min), patients received 555–1,110 MBq of  $^{99\text{m}}\text{Tc}$ -sestamibi, and 4 min later 100–200 mg of aminophylline were administered. Stress imaging was performed 45–90 min after dipyridamole using a Vertex or Forte dual-head SPECT camera (ADAC Laboratories). Counts were collected over 180° rotation with 64 projections of 28 s each. Static images were reconstructed using ordered-subset expectation maximization (10 subsets, 2 iterations), a fifth-order Butterworth filter, 0.52 cycles/cm cutoff frequency, and no attenuation correction as per standard clinical practice at the site. Rest imaging was performed the day before or after the stress study using the same protocol.

**Semiquantitative Analysis.** PET and SPECT scans were assessed visually for image quality as good, fair, or poor. Semiquantitative segmental scoring of SSS, SRS, and SDS was performed with Corridor4DM as described above. Default scores were set automatically using the simulated database with a uniform 75% normal cutoff to establish a consistent starting point for clinical interpretation, as scanner-specific normal databases were not available for 3D PET/CT. A discretized, 10-step color map was used for visualization, with the scores corresponding to the colors and database percentages shown in Figure 2. Default scores were modified by the interpreting physician according to their expert visual assessment and the following guide-



**FIGURE 1.** (A) Anthropomorphic torso phantom with heart and liver inserts, used to simulate rest and stress perfusion scans. Corridor4DM display of CTAC (B) and fused PET CTAC images (C) used to verify alignment for proper attenuation correction. Fifty percent PET activity (green-yellow) should fall within CT soft-tissue region for proper alignment.

**TABLE 1**  
PET/CT Systems Used in Rubidium-ARMI Study

Scanner	Reconstruction algorithm	Filter	Cutoff	Prompt $\gamma$ correction	SSS	SRS	SDS
D690 PET/VCT-64	VUE Point HD (4 iterations, 24 subsets)	Hann	12 mm	Yes	2	0	2
D600 PET/CT-16	VUE Point HD (4 iterations, 32 subsets)	Hann	12 mm	Yes	2	0	2
DRX PET/CT-16	3D FORE OSEM (5 iterations, 21 subsets)	Hann	12 mm	No	2	0	2
TF PET/CT-16	3D RAMLA	NA	Normal	No	2	0	2
GXL PET/CT-16	3D RAMLA	NA	Normal	No	2	0	2
DSTE/VCT-16	3D FORE OSEM (8 iterations, 20 subsets)	Hann	12 mm	No	2	0	2
TF PET/CT-16	3D RAMLA (2 iterations, 4 subsets)	NA	Normal	No	2	0	2
Biograph PET/CT-16	3D FORE OSEM (2 iterations, 24 subsets)	Gauss	12 mm	Yes	2	0	2
DSTE/VCT-16	3D FORE OSEM (8 iterations, 20 subsets)	Hann	12 mm	No	2	0	2

FORE = Fourier rebinning; OSEM = ordered-subset expectation maximization; RAMLA = row action maximum likelihood algorithm.

lines: stress defect scores represent infarct plus ischemia, rest defect scores represent infarct only, and stress–rest difference scores reflect ischemia. Physicians reviewed 5 example cases before the start of the trial to familiarize themselves with these guidelines. The example scans were selected by the core lab reviewer and included 2 normal and 3 abnormal cases, representing a combination of straight-forward versus challenging interpretations. The site reviewers scored these cases independently based on the guidelines above and then compared their scores with those assigned by the core lab reviewer; any discrepancies were discussed with the core lab to increase interpretation consistency and experience.

**Site Versus Core Interpretation.** Scans were analyzed by a single physician per scan, per site, with several sites having two or more reporting physicians. The 25 cases from the recruiting sites were then coread by a single physician at the core lab to assess the variability in image quality, SSS, SDS, and clinical diagnosis between the sites and core. All physicians were senior nuclear medicine reporting staff experienced with SPECT perfusion imaging. With the exception of the core lab reviewer, all were new to the reporting of PET perfusion scans because  $^{82}\text{Rb}$  was not yet approved for sale in Canada.  $^{82}\text{Rb}$  PET MPI was initiated at the recruiting sites for the rubidium-ARMI study, hence the need for the common interpretation software and specific interpretation guidelines above. Scores with differences greater than 3 underwent a third review via discussion between the core lab reader and the recruiting site physician to improve scoring consensus. A difference greater than 3 was chosen as consensus review cutoff because it is equivalent to the range of SSS diagnostic categories (e.g., 0–3, 4–7, and 8–11) used commonly in prognosis studies of MPI (9).

After the quantitative perfusion scoring, clinical diagnosis was classified as normal, abnormal, equivocal, or uninterpretable. The core reviewer diagnosis was based only on the perfusion scores when electrocardiogram-gated and coronary calcium results were not available at the core lab. Cases in which the image quality was not indicated by one or both of the interpreters ( $n = 21$ ) were excluded from the analysis of that metric. The core lab reader was masked to clinical history.

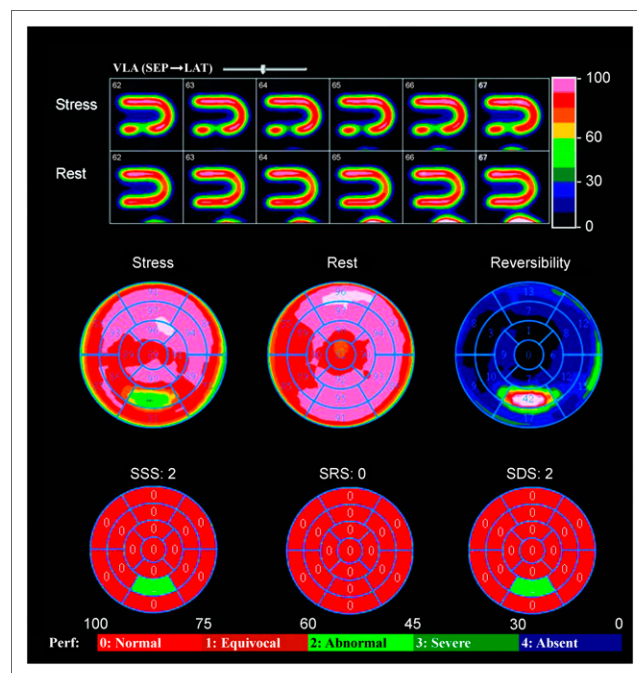
**Statistical Analysis.** PET and SPECT demographics were compared via nonparametric Wilcoxon rank-sum tests for continuous variables and Fisher exact tests for categorical variables. The interclass correlation coefficient (ICC-r), determined with a 2-way random model with single measures, was used to assess agreement of  $^{82}\text{Rb}$  and combined  $^{201}\text{Tl}$  and  $^{99\text{m}}\text{Tc}$  SSS and SDS values, before and after consensus review. ICC-r was also used to assess the scoring agreement of normal versus abnormal ( $\text{SSS} \geq 4$ ) and ischemic ( $\text{SDS} \geq 2$ ) versus nonischemic cases. Bland–Altman analysis was used to evaluate the mean  $\pm$  SD and reproducibility coefficient (RPC) between the site and core scores. Means were compared via the Student  $t$  test. The Fisher exact

test was used to determine the significance of the change in correlation before versus after consensus review and between PET versus SPECT. The F test was used to compare the RPC values before versus after consensus review and between PET versus SPECT.  $P$  values of less than 0.05 were considered significant. Agreement between image quality was assessed with percentage agreement and the Fisher exact test. The agreement of clinical interpretation was assessed with the  $\kappa$  coefficient. Site-by-site variability of image quality assessment was evaluated using ANOVA.

## RESULTS

### Qualifying Scans

Phantom scans consistently resulted in polar map scores of  $\text{SSS} = 2$ ,  $\text{SRS} = 0$ , and  $\text{SDS} = 2$  at all imaging sites using 9



**FIGURE 2.** Qualifying phantom scan results showing normal perfusion at rest and inferior wall defect at stress, resulting in  $\text{SRS} = 0$ ,  $\text{SSS} = \text{SDS} = 2$ . Vertical-long-axis (VLA) views of myocardium at stress and rest are shown. Normalized stress, rest, and reversibility 17-segment polar maps show assigned database percentages and spectrum 10-step colors. SSS, SRS, and SDS polar maps show expected defect in inferior wall.

**TABLE 2**  
Patient Demographics

Description	PET	SPECT
Male sex	56%	58%
Age (y)	65 ± 11	65 ± 11
Body mass index (kg/m <sup>2</sup> )	30 ± 7	29 ± 5
Diabetes	36%	30%
Smoker*	63%	66%
Hypertension	68%	68%
Dyslipidemia	79%	76%
Family history	46%	50%
Typical angina	21%	12%
Previous myocardial infarction	28%	26%
Previous percutaneous coronary intervention	26%	28%

\*Past or current.

different 3D PET/CT scanners (Table 1). Normal resting scans with an SRS = 0 were expected by design, using the phantom with uniform myocardial activity. Similarly, all simulated stress scans with the defect in the inferior wall resulted in an SSS = 2, with an SDS = 2 accurately reflecting the pattern of reversible ischemia under these idealized conditions with a small 1-cm<sup>3</sup> defect centered in the midinferior segment (Fig. 2).

#### Patient Scans

PET and SPECT patients had a mean age of 65 y (34–93 and 39–89 y, respectively). None of the demographics were significantly different between the groups (Table 2).

#### <sup>82</sup>Rb PET Site Versus Core Interpretation

Recruiting site SSS and SDS scores were  $5.6 \pm 7.9$  and  $3.3 \pm 5.2$ , respectively. Comparison between core and site scores resulted in good overall agreement: ICC-r = 0.92 for SSS and 0.77 for SDS. Eighty-five percent (148/174) of SSS and 87% (151/174) of SDS scores had absolute differences  $|\text{site} - \text{core}| \Delta \leq 3$ . The difference was 0 in 40% of the cases for SSS and 51% of cases for SDS, with  $\Delta \leq 6$  in 95% of the cases for SSS (range, -15 to +15) and 98% of cases for SDS (range, -15 to +13) (Figs. 4A and 4B). Mean differences were SSS  $\Delta = -0.52 \pm 3.3$  and SDS  $\Delta = 0.35 \pm 3.5$  ( $P = \text{not significant vs. zero}$ ) (Figs. 4C and 4D). The reproducibility coefficient for SSS and SDS was 6.4 and 5.4, respectively. Scoring agreement was most highly correlated in the normal to moderate range (SSS and SDS < 15); this is important because small scoring differences in this range can shift a patient from one diagnostic category to another.

After consensus review, overall agreement was significantly improved to ICC-r = 0.97 for SSS and 0.96 for SDS ( $P < 0.05$  for both). Considering the sites individually, 5 of 7 had significant improvement in SSS or SDS agreement after consensus review. In cases for which a nonsignificant improvement was observed, the

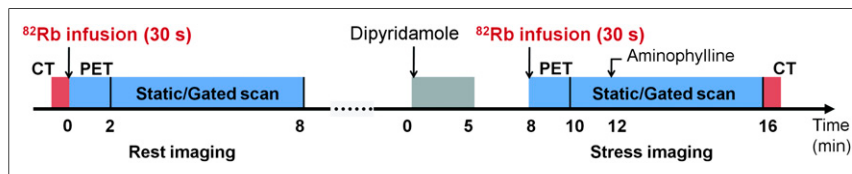
initial agreement rate was already high (ICC-r > 0.8) so there was less room for improvement. The number of cases with an overall scoring  $\Delta$  of 0 increased to 44% for SSS and 55% for SDS. The range of SSS and SDS differences was also reduced (Figs. 4A and 4B). Mean differences were SSS  $\Delta = -0.36 \pm 1.8$  and SDS  $\Delta = 0.26 \pm 1.4$  ( $P = \text{not significant vs. zero}$ ) (Figs. 4C and 4D). RPC for SSS and SDS improved significantly to 3.5 and 2.8, respectively ( $P < 0.05$ ). A smaller range of scoring differences was observed after consensus review for both SSS and SDS. There was little change in the low-score range; however, the distribution was narrower in the highly abnormal cases (i.e., high SSS and SDS). Ninety-three percent of the SSS data and 95% of the SDS were within a difference of 3 after consensus review. The largest discrepancies occurred in cases with large defects spanning multiple segments, as in the example in Figure 5. Despite the scoring differences, these cases were all correctly identified as abnormal at both the recruiting site and the core lab. The scoring agreement of normal versus abnormal (SSS  $\geq 4$ ) scans and ischemic (SDS  $\geq 2$ ) versus nonischemic was found to be excellent, with an ICC-r = 0.90 and 0.88, respectively.

For overall diagnostic interpretation, before consensus review the site and core interpretations were in 86% agreement ( $\kappa = 0.74$ ); after review agreement improved to 94% ( $\kappa = 0.89$ ). The same 82 of 82 cases were interpreted as abnormal by both core and site reviewers, and 81 of 85 were also considered normal by both core and site. In the other 4 of 85 cases interpreted as normal by the core, 3 were considered abnormal and 1 was equivocal by the site. In 5 cases, the site reported the scans as uninterpretable because of technical difficulties including CTAC misregistration ( $n = 4$ ) or truncation of the heart within the field of view ( $n = 1$ ), whereas the core reported these cases as abnormal. Lastly, in 2 cases considered equivocal by the core, the site reported 1 as normal and 1 as abnormal.

Good diagnostic image quality was obtained for most scans acquired (Supplemental Fig. 1). Recruiting sites ranked 90% of their combined images as good quality, whereas the core reviewer indicated that 76% of the same images fell into that category. Site versus core percentages in the other categories were 7% versus 16% fair and 3% versus 9% poor. These represent a significant difference in the perception of image quality between the core and site reviewers ( $P < 0.05$ ). ANOVA of site rankings of image quality showed no significant difference between sites, but each site consistently ranked its own image quality higher than the core reviewer's ranking ( $P < 0.001$ ).

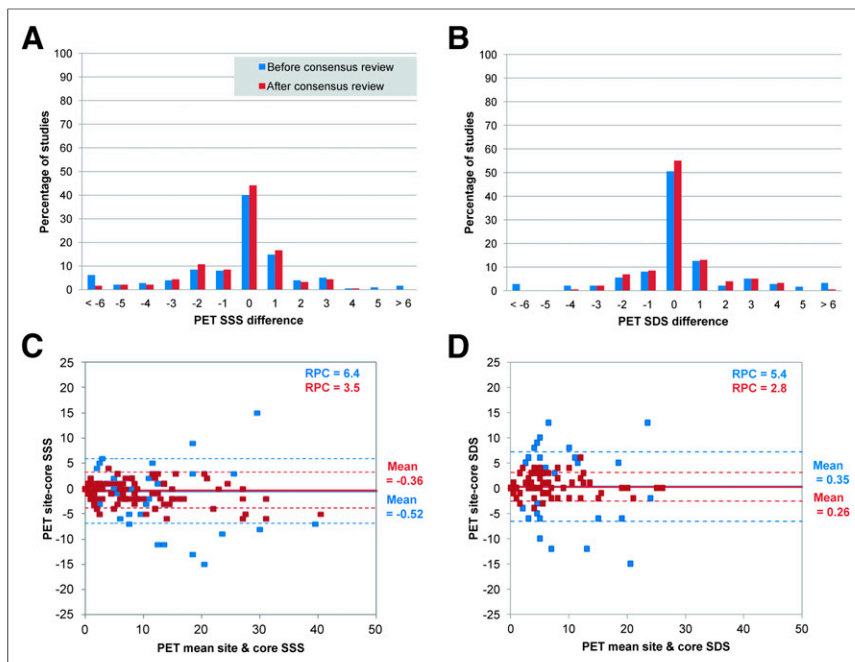
#### <sup>201</sup>Tl and <sup>99m</sup>Tc SPECT Site Versus Core Interpretation

SPECT scans were merged into a single cohort for analysis. Site SSS and SDS scores were  $5.7 \pm 8.3$  and  $3.3 \pm 5.7$ , respectively. Moderate to good agreement between site and core scores was observed: ICC-r = 0.82 for SSS and 0.72 for SDS. Seventy-six percent (38/50) of SSS scores and 88% (44/50) of SDS scores had differences  $|\text{site} - \text{core}| \Delta \leq 3$ . Thirty-six percent of the cases for SSS and 54% for SDS had  $\Delta = 0$ ; the scoring difference was  $\leq 6$  in 86% of the cases for SSS (range, -18 to 14) and 98% for SDS (range, -6 to +13) (Figs. 6A and 6B). There was a moderate correlation between site and core scores:  $r = 0.81$  for SSS and  $r = 0.77$  for SDS, with a mean SSS  $\Delta = -0.40 \pm 4.8$  and SDS  $\Delta = 0.54 \pm 2.7$  ( $P = \text{not significant vs. zero}$ ) (Figs. 6C and 6D). RPC values for SSS



**FIGURE 3.** <sup>82</sup>Rb PET rest and stress MPI protocol.





**FIGURE 4.** (A and B) Difference between core lab and recruiting site PET SSS (A) and SDS (B) before and after consensus review. (C and D) Bland-Altman analysis of site vs. core PET SSS (C) and SDS (D) before and after consensus review.

and SDS were 9.4 and 5.3, respectively. Overall diagnostic interpretation between the core and sites showed 84% agreement ( $\kappa = 0.74$ ).

After consensus review, scoring agreement improved significantly to ICC-r = 0.97 for SSS and ICC-r = 0.90 for SDS ( $P < 0.05$  for both). SDS correlation was significantly lower versus PET ( $P < 0.001$ ), whereas SSS correlation was not significantly

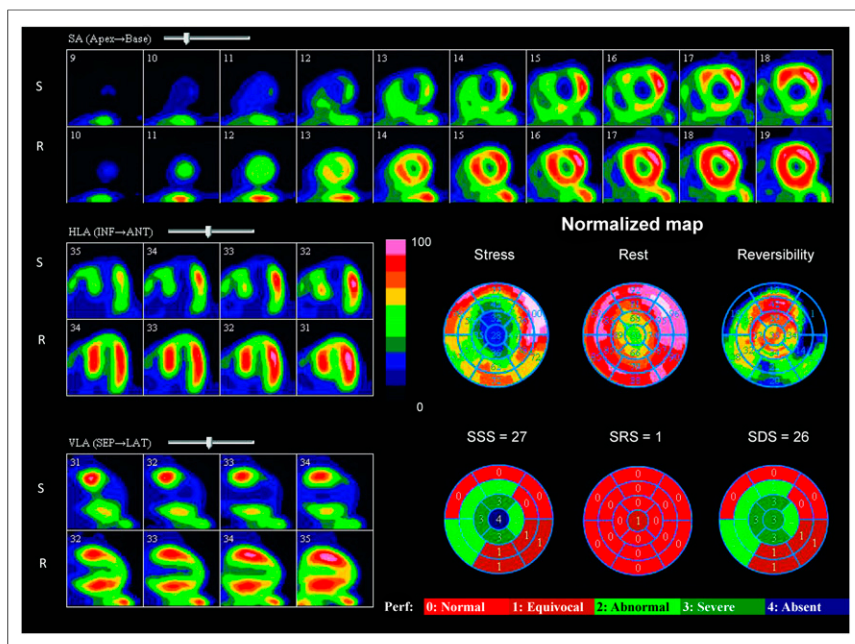
different. Cases with an overall SSS difference of 0 increased to 48%, with no change observed for SDS. The consensus ranges were reduced accordingly (Figs. 6A and 6B). Mean differences were SSS  $\Delta = 0.16 \pm 1.9$  and SDS  $\Delta = -0.14 \pm 1.4$  ( $P =$  not significant vs. zero and vs. PET) (Figs. 6C and 6D). RPC for SSS and SDS improved to 3.8 and 2.8, respectively ( $P < 0.05$ ) but were not significantly different from PET. Scoring agreement of normal versus abnormal (SSS  $\geq 4$ ) scans and ischemic (SDS  $\geq 2$ ) versus nonischemic was good: ICC-r = 0.84 and 0.73, respectively. The latter is significantly worse than the PET scoring agreement for ischemic (SDS  $\geq 2$ ) versus nonischemic cases, for which ICC-r = 0.90 and 0.88, respectively. Overall agreement in diagnostic classification between the core and sites improved to 98% ( $\kappa = 0.96$ ), similar to the PET results. The same 24 of 24 cases were interpreted as normal by both core and site reviewers, and 22 of 23 were considered abnormal by both core and site. In 1 of 23 cases interpreted as normal by the core, the site considered the case equivocal because of a perceived artifact.

Image quality was rated as fair in most cases (54%) by the core reviewer, whereas the sites ranked the majority as good (54%). Overall rankings of the site versus core were 54% versus 46% good, 44% versus 54% fair, and 2% versus 0% poor, representing lower overall image quality as compared with PET ( $P < 0.05$ ). This reflects a 64% overall agreement in SPECT image quality rating, with no significant difference in the perception of quality between the site and core reviewers ( $P =$  not significant).

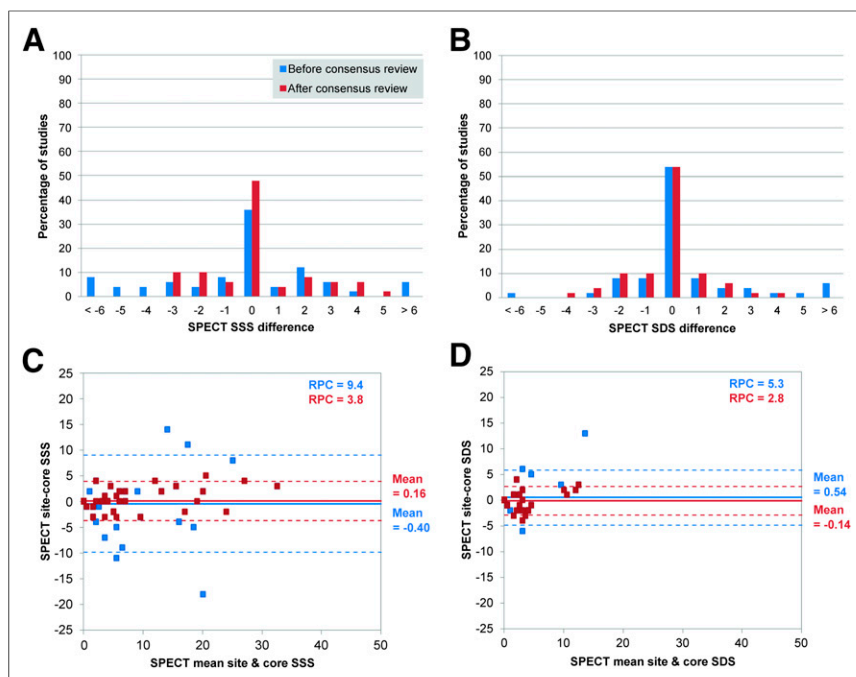
## DISCUSSION

This study successfully standardized  $^{82}\text{Rb}$  imaging protocols at several centers using 3D PET/CT scanners. After initial qualifying phantom scans, clinical  $^{82}\text{Rb}$  scans were coread to assess the agreement of perfusion scores (SSS, SDS), image quality, and overall interpretation between the recruiting site and core lab reviewers. Polar map scoring consensus improved after cases with differences of more than 3 were reviewed. Comparison with the combined results of standard  $^{201}\text{Tl}$ - and  $^{99\text{m}}\text{Tc}$ -based SPECT imaging was performed. SPECT data were merged into a single cohort because separate analysis of the  $^{201}\text{Tl}$ - and  $^{99\text{m}}\text{Tc}$ -based data did not show a significant difference between the 2 tracers.

All 3D PET/CT scanners in this study achieved consistent qualifying phantom scan results despite differences in technology, such as scintillation detectors, number of CT slices, and prompt  $\gamma$  correction. This



**FIGURE 5.**  $^{82}\text{Rb}$  perfusion images and 17-segment consensus scores for patient where core-site scoring difference was greater than 3 (core: SSS, SDS = 28; site: SSS, SDS = 13). Discrepancy is from large defect spanning multiple segments. Regardless of scoring differences, the case was correctly identified as definitely abnormal by both interpreters. S = stress; R = rest.



**FIGURE 6.** (A and B) Difference between core lab and recruiting site SPECT SSS (A) and SDS (B) before and after consensus review. (C and D) Bland-Altman analysis of site vs. core SPECT SSS (C) and SDS (D) before and after consensus review.

latter effect is highly dependent on the particular vendor implementation of scatter correction, which has not been systematically investigated on the GE Healthcare and Philips systems. The phantom studies suggest that any potential bias is small but should be confirmed in future studies comparing perfusion results against an accepted gold standard such as invasive coronary angiography.

The standardized imaging and scoring methods allowed excellent agreement of SSS and SDS between site and core lab interpretations. The improvement in scoring agreement after consensus review demonstrates the added value of this process in improving consistency and training experience. Significant improvement in half of the sites demonstrates that they required the consensus review process to improve their technique and experience to accurately score more difficult cases, whereas the other sites already had adequate understanding of the scoring methods after initial training. Greater differences were observed in the site-core scoring in cases with large SSS and SDS. These cases were still recognized as highly abnormal by both readers; thus, these scoring differences were not clinically significant. However, in the cases of mild-to-moderate disease (SSS, SDS < 15), for which a small scoring difference can change the disease classification, a high correlation between the site and core scores was observed. The equivocal range of 60%–75% used in the present study is consistent with the established abnormal threshold of PET perfusion, less than 60% of maximum, as reported previously (14). Without an independent gold standard, the present study results may rely on the core lab expertise and experience relative to the newly recruited centers. Consensus reviewing was shown to improve agreement but did not introduce a bias toward the core lab interpretation scores because an equal proportion was shifted either toward the site scores or toward the core scores (31% each), and the remainder shifted to the average of the 2 scores (39%). Overall agreement might be further improved using scanner-specific  $^{82}\text{Rb}$  normal databases to assign default segmental scores,

reducing the need for user modification and decreasing interoperator variability. We recently reported the accuracy of low-dose  $^{82}\text{Rb}$  MPI using a 3D PET/CT normal database (8), showing results similar to standard-dose imaging (15) and traditional 2D PET methods. Consistent scan interpretation between recruiting centers is essential to permit pooling of the data for further evaluation of the rubidium-ARMI endpoints of clinical outcomes and cost-effectiveness. Agreement between site and core scoring was assessed only after the initial training period. It is not known whether these results may diverge or converge over the course of the study, but it seems likely that the agreement should be maintained over time, assuming that the site reviewers continue to follow the simple interpretation methods as described.

The 94% agreement of PET interpretation after consensus review is similar to, or better than, previous studies comparing SPECT perfusion scores between readers (16). In those studies, agreement was in the range of 68%–97% ( $\kappa = 0.56$ –0.89). In the present study, SPECT interpretation agreement was 98% ( $\kappa = 0.96$ ), similar to PET. The PET interpretation agreement might have been even further improved if the core reviewer had access to the CTAC images for all studies. This is a limitation of this study and highlights the importance of reviewing the CTAC and PET alignment in all cases. Attenuation-corrected SPECT images were not evaluated in the present study. While this may be viewed as a limitation, the procedure is not widely used in clinical practice.

Good PET image quality was observed in most cases when the site and core reviewers used a 3-point scale of good, fair, or poor. The higher percentage of site rankings of good PET image quality may reflect the reviewers' relative inexperience at the start of the trial. The core interpreter had observed a larger number of cases across the sites and thus had a wider frame of reference than the site reviewers, who reviewed only images from their own particular site. Image quality may vary from one site to another because of the specific equipment and patient populations—for example, one site had a significant population of bariatric surgery patients with high body mass index, resulting in lower-quality images. Comparison of image quality in patients scanned on both PET and SPECT systems may further elucidate the relative quality of  $^{82}\text{Rb}$  PET versus SPECT images.

$^{82}\text{Rb}$  is produced from a  $^{82}\text{Sr}/\text{Rb}$  generator, allowing widespread distribution. The short 76-s half-life allows rapid scanning times and lower radiation exposure to the patient. The short scan time permits rapid rest and stress imaging to be completed in 30–45 min, which is convenient for the patient and permits high-throughput imaging and efficient use of the technology. The low-dose (10 MBq/kg) protocol may also allow simultaneous quantification of absolute flow on 3D PET systems with adequate dynamic range to permit accurate measurement of the bolus first-pass activity (9,17,18) and measurement of ventricular function close to peak hyperemia with pharmacologic stress. Ischemia-induced perfusion and wall-motion changes with  $^{82}\text{Rb}$  PET may enhance

sensitivity for detecting clinically significant disease (19). The reported sensitivity and specificity for  $^{82}\text{Rb}$  PET imaging was approximately 90%–93% and 81%–88%, respectively, in systematic reviews by our group and others (2,7,20). These reviews also demonstrated that  $^{99\text{m}}\text{Tc}$ -based SPECT sensitivity and specificity were 80%–85% and 76%–85%, respectively (7,20).  $^{82}\text{Rb}$  PET was shown to be superior to electrocardiogram-gated  $^{99\text{m}}\text{Tc}$ -SPECT MPI in terms of overall accuracy, normalcy rates, and improved image quality, with increased confidence in interpretation (fewer equivocal studies) (6). In comparison, mean accuracies were reported to be 85% versus 79% for PET and  $^{201}\text{Tl}$  SPECT (5).

Our results show the benefit of a QA and standardization program for the dissemination and use of a new PET tracer in the clinical routine. The present study supports the ability to combine data from several centers for evaluation of the full multicenter trial results. This model may also be helpful in the investigation and implementation of other novel tracers, such as  $^{18}\text{F}$ -labeled perfusion agents currently in phase III development (21).

## CONCLUSION

Reproducible imaging standards are essential for clinical trial results to be pooled across participating centers. With effective training and standardization of  $^{82}\text{Rb}$  PET MPI, good interpretation agreement was found between the recruiting sites and the core lab. Agreement was further improved with consensus reading of the most discrepant cases, suggesting that improved scoring consistency is achieved with increased training and experience. The results indicate that repeatable interpretation is achievable across multiple imaging centers using different 3D PET/CT scanners, following these imaging standards and quality assurance methods.

## DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734. This study was funded by research grants from CIHR MIS-100935 (Rubidium-ARMI) and ML1-112246 (MITNEC), HSFO PRG-6242 (Molecular Function and Imaging program), and ORF RE-02038 (ICT) in partnership with Jubilant DraxImage (JDI) and INVIA in-kind. Rob S.B. Beanlands was supported in part by a career investigator award from HSFO and Tier 1 Research Chair at the University of Ottawa. Ilias Mylonas, Brian McArdle, and Taylor Dowsley were supported in part by the HSFO program grant, the Whit & Heather Tucker Endowed Fellowship, the Vered-Beanlands Endowed Research Fellowship in Cardiology, and the UOHI Associates in Cardiology. Robert A. deKemp receives royalties from rubidium PET technology licenses. Rob S.B. Beanlands and Robert A. deKemp are consultants with JDI. No other potential conflict of interest relevant to this article was reported.

## ACKNOWLEDGMENTS

We thank the research coordinators, technologists, and nurses at the participating sites for their efforts acquiring the PET and SPECT scans. We acknowledge Jubilant DraxImage for providing the  $^{82}\text{Rb}$

generators, elution system, training, and support and INVIA Medical Solutions for providing Corridor4DM software and support.

## REFERENCES

- Hendel RC, Berman DS, Di Carli MF, et al. ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/SNM 2009 appropriate use criteria for cardiac radionuclide imaging. *J Am Coll Cardiol*. 2009;53:2201–2229.
- Beanlands RS, Chow BJ, Dick A, et al. CCS/CAR/CANM/CNCS/CanSCMR joint position statement on advanced noninvasive cardiac imaging using positron emission tomography, magnetic resonance imaging and multidetector computed tomographic angiography in the diagnosis and evaluation of ischemic heart disease: executive summary. *Can J Cardiol*. 2007;23:107–119.
- Ruby-Fill™ product monograph, rubidium Rb 82 generator [package insert]. Kirkland, Quebec, Canada: Jubilant DraxImage; 2011.
- Senthazhchelvan S, Bravo PE, Lodge MA, et al. Radiation dosimetry of  $^{82}\text{Rb}$  in humans under pharmacologic stress. *J Nucl Med*. 2011;52:485–491.
- Stewart RE, Schwaiger M, Molina M, et al. Comparison of rubidium-82 positron emission tomography and thallium-201 SPECT imaging for detection of coronary artery disease. *Am J Cardiol*. 1991;67:1303–1310.
- Bateman TM, Heller GV, McGhie A, et al. Diagnostic accuracy of rest/stress ECG-gated  $^{82}\text{Rb}$  myocardial perfusion PET: comparison with ECG-gated Tc-99m sestamibi SPECT. *J Nucl Cardiol*. 2006;13:24–33.
- McArdle BA, Dowsley TF, Wells G, et al. Systematic review of the accuracy of  $^{82}\text{Rb}$  PET for the diagnosis of CAD in comparison to SPECT. *J Am Coll Cardiol*. 2012;60:1828–1837.
- Kaster T, Mylonas I, Renaud JM, et al. Accuracy of low-dose rubidium-82 myocardial perfusion imaging for detection of coronary artery disease using 3D PET and normal database interpretation. *J Nucl Cardiol*. 2012;19:1135–1145.
- Yoshinaga K, Katoh C, Manabe O, et al. Incremental diagnostic value of regional myocardial blood flow quantification over relative perfusion imaging with generator-produced rubidium-82 PET. *Circ J*. 2011;75:2628–2634.
- Rubidium-82: an alternative radiopharmaceutical for myocardial imaging (ARMI). Clinicaltrials.gov. <http://clinicaltrials.gov/ct2/show/NCT01128023>. Accessed October 21, 2013.
- Esteves FP, Nye JA, Khan A, et al. Prompt-gamma compensation in  $^{82}\text{Rb}$  myocardial perfusion 3D PET/CT. *J Nucl Cardiol*. 2010;17:247–253.
- Dilsizian V, Bacharach SL, Beanlands R, et al. ASNC imaging guidelines for nuclear cardiology procedures: PET myocardial perfusion and metabolism imaging. *J Nucl Cardiol* [serial online]. <http://www.asnc.org/imageuploads/ImagingGuidelinesPETJuly2009.pdf>. Accessed November 8, 2013.
- Klein R, Adler A, Beanlands RS, deKemp RA. Precision-controlled elution of a  $^{82}\text{Sr}/^{82}\text{Rb}$  generator for cardiac perfusion imaging with positron emission tomography. *Phys Med Biol*. 2007;52:659–673.
- Sdringola S, Nakagawa K, Nakagawa Y, et al. Combined intense lifestyle and pharmacologic lipid treatment further reduce coronary events and myocardial perfusion abnormalities compared with usual-care cholesterol-lowering drugs in coronary artery disease. *J Am Coll Cardiol*. 2003;41:263–272.
- Nakazato R, Berman DS, Dey D, et al. Automated quantitative  $^{82}\text{Rb}$  3D PET/CT myocardial perfusion imaging: normal limits and correlation with invasive coronary angiography. *J Nucl Cardiol*. 2012;19:265–276.
- Johansen A, Høiland-Carlson 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;11:710–718.
- Klein R, Renaud JM, Ziadi MC, et al. Intra- and inter-operator repeatability of myocardial blood flow and myocardial flow reserve measurements using rubidium-82 pet and a highly automated analysis program. *J Nucl Cardiol*. 2010;17:600–616.
- Lortie M, Beanlands RS, Yoshinaga K, et al. Quantification of myocardial blood flow with  $^{82}\text{Rb}$  dynamic PET imaging. *Eur J Nucl Med Mol Imaging*. 2007;34:1765–1774.
- Dorbala S, Vangala D, Sampson U, et al. Value of vasodilator left ventricular ejection fraction reserve in evaluating the magnitude of myocardium at risk and the extent of angiographic coronary artery disease: a  $^{82}\text{Rb}$  PET/CT study. *J Nucl Med*. 2007;48:349–358.
- Parker MW, Iskandar A, Limone B, et al. Diagnostic accuracy of cardiac positron emission tomography versus single photon emission computed tomography for coronary artery disease: a bivariate meta-analysis. *Circ Cardiovasc Imaging*. 2012;5:700–707.
- Rischpler C, Park MJ, Fung GS, Javadi M, Tsui BM, Higuchi T. Advances in PET myocardial perfusion imaging: F-18 labeled tracers. *Ann Nucl Med*. 2012;26:1–6.