Journal of Cardiovascular Magnetic Resonance (2005) **7,** 763–768

ISSN: 1097-6647 print / 1532-429X online
DOI: 10.1080/10976640500287414

Taylor & Francis
Taylor & Francis Group

VENTRICULAR FUNCTION

Reproducibility and inter-observer variability of dobutamine stress CMR in patients with severe coronary disease: implications for clinical research

MUSHABBAR A. SYED,[1] D. IAN PATERSON,[1] W. PATRICIA INGKANISORN,[1] KENNETH L. RHOADS,[1] JONATHAN HILL,[2] RICHARD O. CANNON III,[2] and ANDREW E. ARAI, M.D.[1,*]

[1]*Laboratory of Cardiac Energetics, National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA*
[2]*Cardiovascular Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA*

Purpose. The purpose of this study was to analyze reproducibility and inter-observer variability of dobutamine stress cardiac magnetic resonance imaging (dobutamine CMR) and its implications on serial studies. *Methods.* Nineteen consecutive patients underwent two dobutamine CMR each (median 12 days apart), as part of eligibility criteria for phase I/II stem cell therapy trial. These patients had Canadian Cardiovascular Society Class III/IV angina despite maximal therapy. The two studies were compared for reproducibility of stress response. To assess inter-observer variability, 29 randomly selected dobutamine CMR studies were analyzed by three experienced observers and Kappa values were computed to measure the agreement. *Results.* Dobutamine CMR studies were completed without any major complications. The left ventricular function, dobutamine and atropine dose, hemodynamic response, symptomatic response and the results of wall motion and perfusion abnormalities were highly reproducible between the two studies ($p = .91$). Sample size calculations suggested that a clinical trial using dobutamine CMR to detect an endpoint of resolution of two ischemic segments would require a sample size of 20 subjects and to detect an improvement in perfusion of two segments would require a sample size of 8 subjects. Inter-observer variability between individual and consensus interpretation of dobutamine CMR was good to very good (kappa = 0.81 for wall motion and 0.70 for perfusion). *Conclusion.* Dobutamine CMR is a highly reproducible technique with very good inter-observer variability and could be used as a specific endpoint in a relatively small clinical trial.

Key Words: Dobutamine stress test; Magnetic resonance imaging; Reproducibility; Inter-observer variability

1. Introduction

Cardiac magnetic resonance imaging (CMR) is increasingly being used during pharmacological stress testing. The use of dobutamine stress magnetic resonance imaging (dobutamine CMR) was reported in the early 1990's but feasibility was limited (1, 2). Technical advances in image resolution and spatial coverage have translated into higher clinical accuracy and diagnostic value of dobutamine CMR (3, 4). Dobutamine CMR was also shown to have an independent prognostic value in patients suspected of coronary artery disease (5) and in patients undergoing non-cardiac surgery (6).

Reproducibility of a test is an important measure of its diagnostic value and is defined as the ability of a test to give the same results when conducted under similar conditions. Reproducibility is dependent upon a variety of technical and biological conditions as well as observer variability. Although reproducibility of a test is not a marker of accuracy, accurate tests tend to be reproducible under similar testing conditions. Despite common use of pharmacological stress testing, there are relatively few studies reporting the reproducibility of this modality (7–9). The reproducibility and inter-observer variability of dobutamine CMR has not been reported.

We investigated the reproducibility and inter-observer variability of dobutamine CMR in patients referred for ischemic heart disease evaluation. We hypothesized that dobutamine CMR is a reproducible technique with low inter-observer variability based on its proven diagnostic and prognostic value. These characteristics will have important

Received 10 September 2004; accepted 22 May 2005.
This project was funded by the intramural program of the National Heart, Lung, and Blood Institute.
*Address correspondence to Andrew E. Arai, M.D., National Heart, Lung and Blood Institute, National Institutes of Health, Bldg. 10, Room B1D 416, MSC 1061, 10 Center Drive, Bethesda, MD 20892, USA; Fax: (301) 402-2389; E-mail: araia@nih.gov

implications for sample size in clinical trials using dobutamine CMR for serial studies as well as patient management in clinical practice.

2. Methods

2.1. *Patients*

2.1.1. *Reproducibility*

We studied 19 consecutive patients (> 21 years old) who underwent two high-dose symptom limited dobutamine CMR studies (1st study = baseline, 2nd study = reproducibility) as part of eligibility criteria for a phase I/II stem cell therapy protocol. All patients had known coronary artery disease as demonstrated by coronary angiography within the previous six months and a history of Canadian Cardiovascular Society class III or IV angina despite maximal medical therapy and previous revascularization. Patients were excluded from the stem cell therapy protocol if they had a myocardial infarction within the previous 2 months, unstable angina, left ventricular ejection fraction (EF) less than 30%, decompensated heart failure, or other significant structural heart disease. We also excluded pregnant or lactating women, and patients with hemoglobin < 12.5 g/dL or abnormal liver or renal function tests.

2.1.2. *Inter-observer variability*

We randomly selected 29 dobutamine CMR studies to assess for inter-observer variability. Out of these, 14 studies were from the reproducibility cohort (stem cell therapy protocol) while the rest were from 15 patients who had undergone dobutamine CMR for clinical indications. Overall, the dobutamine CMR was positive by wall motion in 17 patients and negative in 12 patients. Perfusion was abnormal in 19 patients and normal in 10 patients.

2.2. *Dobutamine stress protocol*

Patients had their medications withheld only on the morning of the test because of the concern for severe angina. Dobutamine was infused in a standard fashion, starting at 5 mcg/kg/minute (stage 1) and increased to 10 mcg/kg/minute (stage 2), 20 mcg/kg/minute (stage 3), 30 mcg/kg/minute (stage 4), and 40 mcg/kg/minute (stage 5) in three minute intervals (Fig. 1). Atropine was administered intravenously in 0.25 mg increments (total dose 1 mg) starting in stage 3 if heart rate response to dobutamine remained sub-optimal. Dobutamine infusion was discontinued for severe angina, on patient's request, significant drop in systolic blood pressure (> 20 mmHg), severe hypertension (systolic blood pressure of > 250 and/or diastolic blood pressure of > 115 mmHg), severe arrhythmias or other serious adverse effects, new significant wall motion abnormality or if target heart rate was achieved (≥ 85% age related predicted maximum heart rate [PMHR]). During the stress test, patients were monitored for heart rate and rhythm, blood pressure, oxygen saturation, symptoms and wall motion abnormalities.

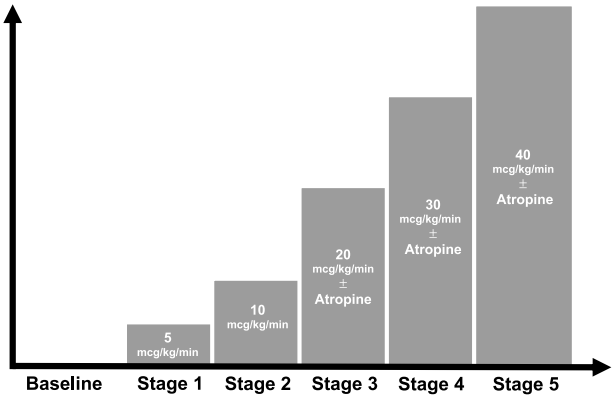

Figure 1. Dobutamine CMR protocol.

2.3. *Magnetic resonance imaging*

CMR was performed on a 1.5 T clinical scanner (General Electric Medical Systems, Waukesha, WI, USA) using a 4-element cardiac phased array coil. Five imaging planes, 3 short axis (basal, mid, apical) and 2 long axis (2 chamber and 4 chamber) views were acquired at rest and at each stress stage using either a fast gradient-echo (FGRE) or steady state free precession (SSFP) technique. We paid particular attention to acquiring all imaging planes at the peak, symptom limited dose. The imaging parameters for FGRE sequence (25 studies) were TR 6.2 ms, TE 2.4 ms, flip 15°, 28 to 36 cm field of view, slice thickness 5 mm, matrix 192 × 160. The imaging parameters for the SSFP sequence (13 studies) were TR 3.6 ms, TE 1.6 ms, flip 45°, 28 to 36 cm field of view, slice thickness 8 mm, matrix 192 × 160. At peak stress, a myocardial perfusion scan was also performed using Gadolinium 0.1 mmol/kg intravenously at a rate of 5 cc/s followed by 20 cc saline flush at the same rate. The perfusion scan utilized a saturation recovery method for T1 weighting and echo planar sequence for image acquisition (TR 6.4 ms, TE 1.5 ms, notch saturation prep flip 90°, read out flip 20°, echotrain length of 4, field of view 32–36, matrix 128 × 96, slice thickness 8 mm) and acquired 5–6 short axis slices of the left ventricle.

2.4. *Image analysis*

Dobutamine CMR images at baseline and each stage were displayed side-by-side in continuous cine loops on a computer screen. An example of the display is shown in Fig. 2. The image analysis software provided options for displaying and viewing each slice location at different stress stages on one screen. The software could display the whole cardiac cycle or only the systolic frames in a cine loop.

All studies were read jointly by four cardiologists experienced in performing and interpreting dobutamine CMR studies and the report was finalized by consensus. Segmental wall motion analysis was reported using the American Society of Echocardiography 16-segment model of the left ventricle (10). Wall motion abnormalities were classified as normal, hypokinetic, akinetic, or dyskinetic. The

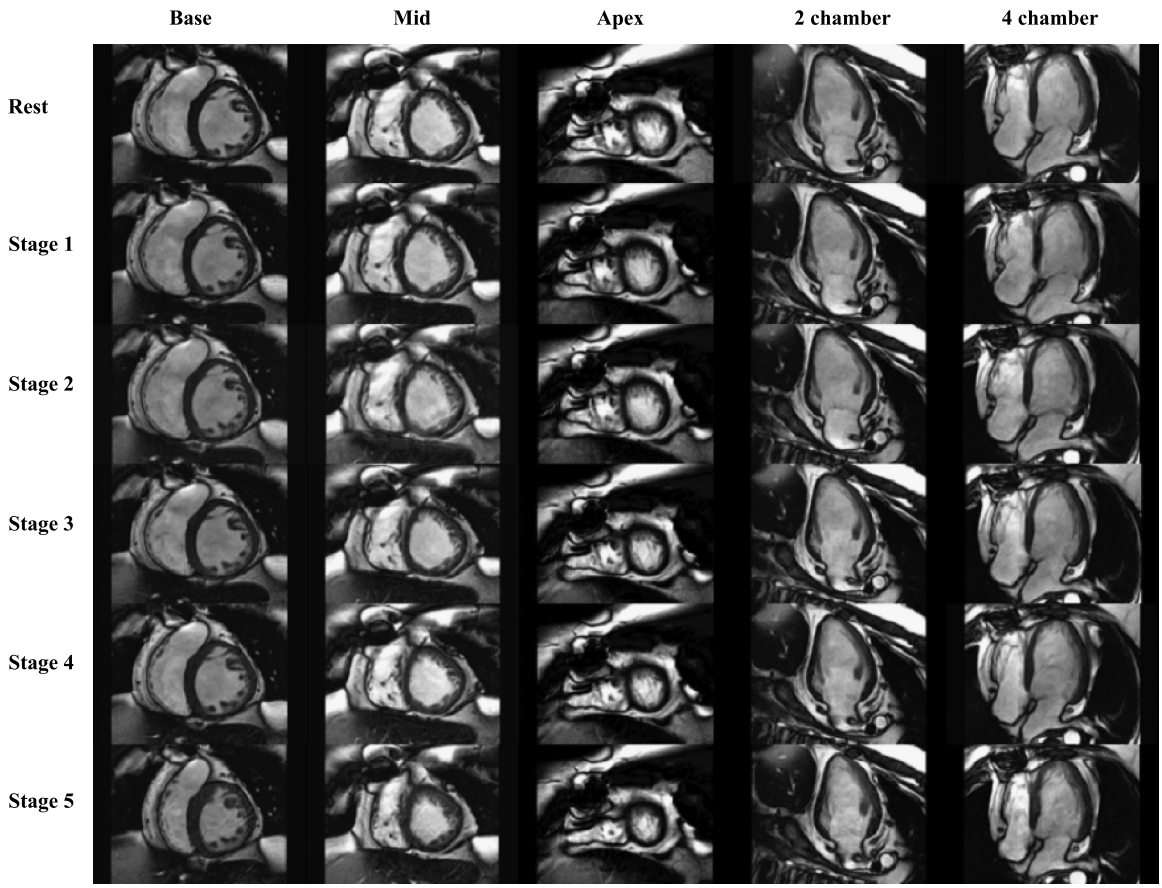

Figure 2. Dobutamine CMR (example of the display).

ischemic response was defined as a new or worsening of wall motion abnormality or a biphasic response (improvement of baseline wall motion abnormality at low dose with worsening at higher dose dobutamine) in one or more segments.

The wall motion score was calculated by summing up the score of each of 16 segments as: normal = 1, hypokinetic = 2, akinetic = 3, dyskinetic = 4. Wall motion score index was the total score divided by 16 (number of total LV segments).

Perfusion scans were analyzed using qualitative and semi-quantitative analysis on the peak enhancement image. Perfusion scans were scored as normal = 0, mild abnormality = 1, moderate abnormality = 2, and severe abnormality = 3, the score for each segment was summed to give a total perfusion score for each study using the 16 segment model.

2.4.1. *Reproducibility*

The dobutamine CMR studies for each patient were read in a random order without the knowledge of the paired study results. The studies were analyzed for the reproducibility of dobutamine and atropine dose, hemodynamic response, symptoms, wall motion ischemic response and perfusion abnormalities.

2.4.2. *Inter-observer variability*

To assess inter-observer variability in reading dobutamine CMR images, three cardiologists blinded to the clinical data analyzed a total of 29 randomly selected studies independently. Their interpretation of the wall motion and perfusion studies was then compared with the consensus reading to assess agreement.

2.5. *Statistical analysis*

Continuous variables were expressed as mean ± SD; a p value of $< .05$ was considered significant. Baseline and

Table 1. Baseline characteristics

Age (years)	55 ± 8
Sex	13 men, 6 women
Body surface area	2.0 ± 0.2 m^2
Risk factors	
Hypertension n (%)	14 (74%)
Diabetes mellitus n (%)	7 (37%)
Smoking n (%)	9 (47%)
Hyperlipidemia n (%)	18 (95%)
Prior history	
Myocardial infarction n (%)	19 (100%)
PTCA n (%)	15 (79%)
CABG n (%)	17 (89%)

PTCA = percutaneous coronary angioplasty; CABG = coronary artery bypass graft surgery.

Table 2. Hemodynamic data

Variables	Baseline study	Reproducibility study	p value
Ejection fraction (EF)	52 ± 9%	53 ± 10%	.58
Resting heart rate (bpm)	65 ± 11	61 ± 9	.18
Peak heart rate (bpm)	125 ± 24	126 ± 17	.74
Resting SBP (mmHg)	141 ± 20	134 ± 18	.19
Peak SBP (mmHg)	158 ± 27	155 ± 20	.74
Resting DBP (mmHg)	79 ± 9	70 ± 10	.002
Peak DBP (mmHg)	78 ± 16	68 ± 20	.03
Resting rate-pressure product	9339 ± 2465	8215 ± 1732	.06
Peak rate-pressure product	19440 ± 4346	19669 ± 3947	.83
Maximum dobutamine dose (mcg/kg/min)	32 ± 9	32 ± 9	.82
Maximum atropine dose (mg)	0.25 ± 0.35	0.30 ± 0.35	.49

bpm = beats per minute; SBP = systolic blood pressure; DBP = diastolic blood pressure.

reproducibility study variables were compared using a paired two-tailed *t* test. Kappa statistics were used to measure the agreement between individual and consensus interpretations. The coefficient of agreement (kappa) was graded as follows: 0.90–1.00 = excellent agreement, 0.80–0.89 = very good agreement, 0.60–0.79 = good agreement, 0.40–0.59 = fair agreement, 0.20–0.39 = poor agreement, and < 0.20 = no agreement. Sample size calculations for a clinical trial using dobutamine CMR as an endpoint were calculated for a paired *t*-test aiming to detect resolution of 2 ischemic wall motion responses or 2 abnormal perfusion segments with an alpha of 0.05 and a power of 0.80.

3. Results

Dobutamine CMR was successfully completed in all 19 patients (38 total studies) without any significant adverse events. Patients averaged 55 ± 8 years of age (13 males) with a body surface area of 2.0 ± 0.2 m^2 (Table 1). In keeping with stem cell therapy protocol eligibility criteria, all patients had refractory angina associated with inoperable coronary artery disease. Additionally, all patients had a prior myocardial infarction and the majority had either coronary angioplasty or coronary bypass surgery (68% had both angioplasty and bypass surgery). Most of the patients (89%) had two or more risk factors. The sensitivity and specificity of dobutamine CMR for the detection of ischemia, defined as > 75% angiographic stenosis, on a per patient basis was 89% and 100% respectively and on a per coronary artery territory basis was 79% and 78% respectively.

3.1. *Reproducibility of dobutamine CMR*

Out of the 38 total studies, gating problems made two wall motion and one perfusion study non-diagnostic; another perfusion study had an error in contrast injection. Table 2 shows the hemodynamic data during the baseline and reproducibility dobutamine CMR studies. The median duration between the baseline and reproducibility study was 12 days. Overall, the left ventricular systolic function was mildly reduced (EF = 53 ± 9%). The average dose of dobutamine and atropine used was 31.8 ± 8.6 micrograms/kg/min and 0.28 ± 0.34 mg intravenously respectively, achieving a peak heart rate of 126 ± 21 beats per minute and a rate pressure product (peak systolic BP × peak heart rate) of 19554 ± 4097. The resting left ventricular function, dobutamine and atropine dosages, resting and peak heart rate, systolic blood pressure,

Table 3. Dobutamine CMR results

Variables	Baseline study	Reproducibility study	p value
PMHR achieved (%)	76 ± 14	77 ± 11	.73
Chest pain	19 (100%)	18 (94%)	.33
Heart rate at the onset of chest pain	101 ± 25	89 ± 33	.09
Resting wall motion score	23.7 ± 6.8	23.7 ± 6.9	1.00
Resting wall motion score index	1.5 ± 0.4	1.5 ± 0.4	.97
Number of ischemic segments	4 ± 3	3 ± 3	.71
Peak wall motion score	25 ± 9	26 ± 8	.96
Peak wall motion score index	1.6 ± 0.5	1.6 ± 0.5	.76
Number of abnormal perfusion segments	5 ± 4	5 ± 4	.45
Perfusion score	9 ± 9	9 ± 8	.92

PMHR = predicted maximum heart rate.

and rate pressure product was not significantly different between the two studies; the average diastolic blood pressure was lower in reproducibility study than the baseline study.

On average, the PMHR achieved was 76 ± 12%. Almost all patients had chest pain during the dobutamine CMR (37/38 studies) at a mean heart rate of 95 ± 30 beats per minute. The baseline study was positive by wall motion in 15 patients and by perfusion in 14 patients (p = .58); the reproducibility study was positive in 15 patients by both wall motion and perfusion (p = 1.0). On comparing the two dobutamine CMR studies (Table 3), no significant differences were found in the PMHR achieved, symptoms of chest pain in relation to the stress stage and the results of the wall motion and perfusion analysis (p value is not significant for all these comparisons).

We also compared the reproducibility of wall motion ischemic response by FGRE and SSFP techniques and did not find any significant differences. On the baseline studies, the mean number of ischemic segments by FGRE were 3.91 ± 3.65 and by SSFP were 3.83 ± 3.54 (p = .97). On the reproducibility studies the mean number of ischemic segments by FGRE compared to SSFP were 3.50 ± 4.09 vs. 3.67 ± 2.25 (p = .93). Additionally, no significant differences were found when wall motion ischemic response by either technique was compared with the entire group by combining the two imaging techniques.

3.2. *Sample size calculation*

Based on the consensus reading, the absolute difference in the number of ischemic segments between baseline and reproducibility studies was 2.4 and abnormal perfusion segments was 1.9. The standard deviation (SD) of the absolute difference in number of ischemic segments was 3.0 and SD of the absolute difference in abnormal perfusion segments was 1.6. Sample size calculations based on the SD of the absolute differences suggested that for a clinical trial using dobutamine CMR to detect an endpoint defined as resolution of two ischemic segments, a sample size of 20 subjects would be needed. Similarly to detect an improvement in perfusion of two segments, a sample size of 8 subjects would be needed.

3.3. *Inter-observer variability*

The results of correlation statistics and kappa values are displayed in Table 4. In comparisons with the consensus readings, kappa values for wall motion assessment averaged 0.81, while kappa values for perfusion assessment averaged 0.70. These kappa values show good to very good correlation between individual and consensus interpretations of dobutamine CMR.

Table 4. Inter-observer variability

	Reader 1	Reader 2	Reader 3
Wall motion			
R-value	0.86	0.82	0.86
Kappa	0.86	0.93	0.65
Perfusion			
R-value	0.86	0.79	0.85
Kappa	0.75	0.77	0.58

4. Discussion

To our knowledge, this is the first report of the reproducibility and interobserver variability of dobutamine CMR. In symptomatic patients with advanced coronary artery disease, reproducibility of paired high dose symptom limited dobutamine CMR (median 12 days apart) was excellent. On comparing the two dobutamine CMR studies, there were no significant differences in symptomatic or hemodynamic response as well as the outcome results. Sample size calculations based on the high reproducibility of the dobutamine CMR suggested that this method could be used as an end point defined as resolution of ischemic segments or perfusion abnormalities in a relatively small clinical trial. The correlation between three independent observers and a consensus reading was very good with a mean Kappa value of 0.81 for wall motion and 0.70 for perfusion assessment.

Despite routine clinical use, the reproducibility and interobserver variability of pharmacological stress testing in general and dobutamine stress testing in particular have rarely been studied. A Medline search of the English literature over the last ten years revealed only one report of interstudy reproducibility of dobutamine stress testing using echocardiography (8). In this report, Takeuchi et al. studied 15 patients with prior myocardial infarction who underwent two dobutamine stress echocardiograms at a mean interval of 19 days. Overall, paired dobutamine stress echocardiography had good reproducibility. Two observers analyzed the studies independently without comparison to a consensus interpretation; the interobserver agreement was 87% with kappa value of 0.71. Our study differed from this report not only by inclusion of a population with more symptomatic coronary artery disease, but we also analyzed a variety of stress function and perfusion variables and compared individual interpretations with a consensus interpretation. Our results indicate a comparable if not better reproducibility and inter-observer variability for dobutamine CMR than described by Takeuchi et al. for dobutamine echocardiography.

The reproducibility and interobserver variability of exercise Thallium-201 and Tc-99 SPECT studies were found to be excellent in selected patients when the exercise test performance was reproducible (9) or the image quality was good to excellent (10). In earlier reports, the interobserver agreement of dobutamine echocardiography was found to be between 89–93% (11, 12). In a multicenter study of 150 patients who underwent dobutamine stress echocardiography, Hoffmann et al. (13) found only fair inter-institutional agreement (kappa 0.37). A similar level of agreement was

found when single center interpretation was analyzed separately. The image quality had a significant effect on the overall agreement, as only 13 out of 150 studies were rated as good delineation of all left ventricular segments.

We believe that the high reproducibility and interobserver variability in our study relates largely to the ability of CMR to provide high resolution imaging in multiple planes at each stress stage. There was clear visualization of the endocardium and assessment of wall thickness in all 16 left ventricular segments.

4.1. *Limitations*

Our study analyzed patients with advanced symptomatic coronary artery disease who have failed mechanical revascularization interventions. Confirmation of these results in a larger and more diverse patient population is needed. Reproducibility studies of stress testing are difficult to perform outside of a research protocol due to clinical limitations. In our study, dobutamine stress CMR was used as eligibility criteria for phase I/II trial of stem cell therapy. Out patient population, therefore, was limited by the enrollment criteria of this clinical trial that only enrolled patients with advanced coronary artery disease who failed maximum medical therapy. Because of concerns for potential adverse toxicity of this treatment, it was decided to only include patients who had objective and reproducible evidence of ischemia. Cardiologists trained in interpretation and performance of CMR and stress testing interpreted these studies and therefore interobserver variability in other clinical settings may be less optimal. Similarly, the reproducibility and interobserver variability in a multicenter setting may be lower than our single center experience.

5. Conclusions

Dobutamine CMR is a highly reproducible technique with very good interobserver variability. Dobutamine CMR could be used as a specific endpoint in a relatively small clinical trial, with an endpoint defined as resolution of ischemic or perfusion abnormalities. The reproducibility of the method could reduce sample size, study costs, and study duration compared with less reproducible methods.

6. Abbreviations

CMR	cardiac magnetic resonance imaging
mcg/kg/min	microgram/kilogram/minute
PMHR	predicted maximum heart rate
FGRE	fast gradient echo
SSFP	steady state free precession
SD	standard deviation
EF	ejection fraction

References

1. Pennell DJ, Underwood SR, Manzara CC, Swanton RH, Walker JM, Ell PJ, Longmore DB. Magnetic resonance imaging during dobutamine stress in coronary artery disease. Am J Cardiol 1992; 70:34–40.
2. Baer FM, Voth E, Theissen P, Schicha H, Sechtem U. Gradient-echo magnetic resonance imaging during incremental dobutamine infusion for the localization of coronary artery stenoses. Eur Heart J 1994; 15:218–225.
3. Hundley WG, Hamilton CA, Thomas MS, Harrington DM, Salido TB, Kitzman DW, Little WC, Link KM. Utility of fast cine magnetic resonance imaging and display for the detection of myocardial ischemia in patients not well suited for second harmonic stress echocardiography. Circulation 1999; 100:1697–1702.
4. Nagel E, Lehmkuhl HB, Bocksch W, Klein C, Vogel U, Frantz E, Ellmer A, Dreysse S, Fleck E. Noninvasive diagnosis of ischemia induced wall motion abnormalities with the use of high dose dobutamine stress MRI. Circulation 1999; 99:763–770.
5. Hundley WG, Morgan TM, Neagle CM, Hamilton CA, Rerkpattanapipat P, Link KM. Magnetic resonance imaging determination of cardiac prognosis. Circulation 2002; 106:2328–2333.
6. Rerkpattanapipat P, Morgan TM, Neagle CM, Link KM, Hamilton CA, Hundley WG. Assessment of preoperative cardiac risk with magnetic resonance imaging. Am J Cardiol 2002; 90:416–419.
7. Takeuchi M, Sonoda S, Miura Y, Kuroiwa A. Reproducibility of dobutamine digital stress echocardiography. J Am Soc Echocardiogr 1997; 10:344–351.
8. Alazraki NP, Krawczynska EG, DePuey G, Ziffer JA, Vansant JP, Pettigrew RI, Taylor A, King SB III, Garcia EV. Reproducibility of thallium-201 exercise SPECT studies. J Nucl Med 1994; 35:1237–1244.
9. Golub RJ, Ahlberg AW, McClellan JR, Herman SD, Travin MI, Mather JF, Aitken PW, Baron JI, Heller GV. Interpretive reproducibility of stress Tc-99m Sestamibi tomographic myocardial perfusion imaging. J Nucl Cardiol 1999; 6:257–269.
10. American Society of Echocardiography Committee on Standards, Subcommittee on Quantitation of Two Dimensional Echocardiograms. Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. J Am Soc Echocardiogr 1989; 2:358–367.
11. Sawada SG, Segar DS, Ryan T, Brown SE, Dohan AM, Williams R, Fineberg NS, Armstrong WF, Feigenbaum H. Echocardiographic detection of coronary artery disease during dobutamine infusion. Circulation 1991; 83:1605–1614.
12. Beleslin BD, Ostojic M, Stepanovic J, Djordjevic-Dikic A, Stojkovic S, Nedeljkovic M, Stankovic G, Petrasinovic Z, Gojkovic L, Vasiljevic-Pokrajcic Z, Nedeljkovic S. Stress echocardiography in the detection of myocardial ischemia: head to head comparison of exercise, dobutamine and dipyridamole stress tests. Circulation 1994; 90:1168–1176.
13. Hoffmann R, Lethen H, Marwick T, Arnese M, Fioretti P, Pingitore A, Picano E, Buck T, Erbel R, Flachskamp FA, Hanrath P. Analysis of inter-institutional observer agreement in interpretation of dobutamine stress echocardiography. J Am Coll Cardiol 1996; 27:330–336.