Determination of interobserver variability for identifying inducible left ventricular wall motion abnormalities during dobutamine stress magnetic resonance imaging

Ingo Paetsch1*, Cosima Jahnke1, Victor A. Ferrari2, Frank E. Rademakers3, Patricia A. Pellikka4, W. Gregory Hundley5,6, Don Poldermans7, Jeroen J. Bax8, Karl Wegscheider9, Eckart Fleck1, and Eike Nagel1

1Department of Cardiology, German Heart Institute Berlin, Augustenburger Platz 1, 13353 Berlin, Germany; 2Department of Medicine, Cardiovascular Medicine Division, University of Pennsylvania Medical School, Pennsylvania, PA, USA; 3Department of Cardiology, Gasthuisberg University Hospital, Leuven, Belgium; 4Division of Cardiovascular Diseases, Mayo Clinic, Rochester, Minnesota, MN, USA; 5Department of Internal Medicine/Cardiology Section, Wake Forest University School of Medicine, Winston-Salem, North Carolina, NC, USA; 6Department of Radiology, Wake Forest University School of Medicine, Winston-Salem, North Carolina, NC, USA; 7Department of Cardiology, Thorax Center Rotterdam, The Netherlands; 8Department of Cardiology, Leiden University Medical Center, The Netherlands; and 9Institute for Statistics and Econometry, University of Hamburg, Hamburg, Germany

Received 17 October 2005; revised 28 February 2006; accepted 23 March 2006; online publish-ahead-of-print 13 April 2006

This paper was guest edited by Prof. Dr Markus Schwaiger, Klinik r.d. Isar der TU Muenchen, München, Germany

See page 1394 for the editorial comment on this article (doi:10.1093/eurheartj/ehl047)

Keywords
Cardiac magnetic resonance imaging; Dobutamine stress; Wall motion analysis; Reader variability

Introduction
Cardiovascular magnetic resonance (CMR) has gained widespread acceptance for assessment of global and regional left ventricular (LV) function at rest and exhibits low inter- and intraobserver variability for the assessment of LV volumes and regional wall motion. Recently, CMR has been implemented during pharmacological stress [dobutamine stress magnetic resonance (DSMR)] for the detection of inducible wall motion abnormalities due to flow-limiting coronary arterial luminal narrowings. Clinically, DSMR has proven to be safe, feasible, accurate, and reliable for identifying myocardial ischaemia and has been used successfully in limited patient populations to determine cardiac prognosis.2–7 Importantly, however, the interobserver variability for identifying inducible LV wall motion abnormalities during ischaemia has not been established. This issue is important because a low interobserver variability would allow application of a standardized DSMR test in serial patient studies. In particular, if DSMR were to be utilized in multicentre trials, unanimous reader decisions regarding the presence of inducible wall motion abnormalities would increase the consistency of the data.

Aims
To determine the interobserver variability for identifying inducible left ventricular (LV) wall motion abnormalities during high-dose dobutamine/atropine stress cardiovascular magnetic resonance (DSMR).

Methods and results
Four readers from various institutions were supplied with the image data from 150 consecutive DSMR examinations and asked to grade wall motion and image quality throughout graded doses of dobutamine infusion administered to achieve 85% of the maximum age-predicted heart rate. Inducible ischaemia was identified if more than one segment demonstrated a new or worsening LV wall motion abnormality, and significant stenosis was defined as >50% luminal diameter reduction by quantitative coronary angiography. Seventy-seven patients (51%) had luminal narrowings ≥50%. Diagnostic performance (sensitivity, specificity, diagnostic accuracy) of all readers was 78.2, 87.0 and 82.5%. Disagreement between two readers occurred in every seventh examination. Agreement on the presence or absence of inducible wall motion abnormalities was moderate (mean kappa value 0.59, range 0.52–0.76). Diagnostic performance and disagreement were independent of the presence of luminal narrowings ≥50% or the number of diseased coronary vessels. Image quality was regarded excellent in 89.3% of standard views.

Conclusion
In the setting of multiple observers from different institutions performing a diagnostic reading of DSMR examinations carried out at a single centre, the interobserver variability was low for identifying inducible LV wall motion abnormalities indicative of coronary arterial luminal narrowings ≥50%.

© The European Society of Cardiology 2006. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org
Accordingly, we assessed the interobserver variability of DSMR among four readers from various institutions. We excluded any information on patient history or stress testing.

Methods

Study population

The study was conducted in accordance with the standards of the Charité and Virchow-Klinikum Ethics Committee. Between January 2002 and July 2003, a total number of 162 consecutive patients with chest pain referred to the German Heart Institute for clinically indicated cardiac catheterization were screened for study inclusion and 150 patients willingly agreed to give written informed consent for prior DSMR testing. Patients were eligible if they had suspected or known coronary artery disease (CAD) (with or without prior percutaneous revascularization or a history of previous myocardial infarction). Patients with prior coronary surgery or typical contraindications for MR imaging and administration of dobutamine were not considered. All study participants were instructed to refrain from β-blockers and antianginal medication at least 24 h prior to the CMR study.

Magnetic resonance study

CMR was performed with the patient in the supine position using a 1.5 T MR scanner (Philips Intera CV, Best, The Netherlands) equipped with a PowerTrak6000 gradient system (23 mT/m; 219 μs rise time), a five-element phased array coil and software package release 9. Cardiac synchronization was performed using four electrodes placed on the left anterior hemithorax (Vector-ECG).

Magnetic resonance imaging technique

For cine imaging, a steady-state free precession sequence was used during an end-expiratory breathhold of 4–6 s in combination with parallel image acquisition (SENSe) and retrospective gating (25 phases/cardiac cycle, repetition time 2.7 ms, echo time 1.4 ms, flip angle 60°). Typical in plane spatial resolution was 1.8 × 1.8 mm² with a slice thickness of 8 mm.

Scan procedure

The following standardized scheme for image planning was rigorously applied by the technician performing the scan procedure: (i) rapid multislice survey, (ii) expiratory, single-angled, single-slice cine scan intersecting the apex and the coaptation point of the mitral valve on a transverse view, (iii) expiratory, double-angled long-axis cine scan intersecting the apex and the coaption point of the mitral valve planned on the previously acquired single-angulated axis cine scan intersecting the apex and the coaptation point of the mitral valve, (iv) expiratory cine imaging of three short-axis views acquired at 25, 50, and 75% of the distance between the epicardial border of the LV apex and the coaptation of the mitral valve (end-systolic phase), (v) expiratory cine imaging of four-, five-chamber views planned on the equatorial short-axis orientation.

Dobutamine stress MR protocol

The pharmacological stress protocol followed the previously described high-dose dobutamine/atropine regimen. Briefly, after acquisition of rest cine scans in the diagnostic standard views (apical-, mid-, and basal short-axis views, four-, two-, and three-chamber views), dobutamine was infused intravenously at 3 min stages at doses of 10, 20, 30, and 40 μg/kg/min and all standard views were acquired at each level. If target heart rate, defined as age-predicted submaximal heart rate [(220 − age) × 0.85], was not reached at the highest dobutamine level, atropine was applied in 0.25 mg fractions (maximal dose 2 mg). Termination criteria were as previously published.

Reader characteristics

Four readers (V.A.F.; F.E.R.; W.G.H.; E.N., with expertise in DSMR of 5–10 years involving 3–10 examinations per week) evaluated the examinations.

Image analysis

The DSMR examinations were anonymized and saved on the hard disk of four laptop computers being sent to the four readers. A PC-compatible dedicated viewing software (View Forum, Philips Medical Systems, The Netherlands) with a synchronized quadscreen image display was used for review of the DSMR examinations. The readers were supplied with the DSMR images only, thus being fully blinded to clinical, angiographic, and pharmacological stress-related data of the patients.

Each reader performed segmental analysis of LV wall motion by applying the standard visual scoring system to grade wall motion abnormalities (1, normokinesis, 2, hypokinesis, 3, akinesis, 4, dyskinesis) and documented the reading on a standardized evaluation sheet (17-segment model). Grading was done at rest and during dobutamine stress at 10 μg/kg/min, 20 μg/kg/min, and maximum stress level. Ischaemia was defined as more than one segment(s) showing an inducible wall motion abnormality (i.e. an increase in the segmental wall motion score of at least one) or a biphasic response. Other criteria for interpretation were not pre-arranged, no prior joint training session was organized.

Image quality was graded on a four-point scale as excellent, good, moderate, or non-diagnostic at each stress level and for each of the standard views.

Coronary arterial supply of myocardial segments was defined according to the guidelines of the American Heart Association, and for further analysis, the left anterior descending artery (LAD) and the combined right coronary artery (RCA)/left circumflex artery (LCX) territory were considered.

Quantitative coronary angiography

Coronary angiography was considered the standard of reference. The procedure was done according to the angiographic guidelines using a simultaneous biplane, multidirectional and isocentric X-ray system. At least two orthogonal views of every major coronary vessel and its side branches were acquired. Quantitative coronary angiography (Philips Inturis CardioView, QCA V3.3, Pie Medical Imaging) was performed off-line at the German Heart Institute Berlin by an independent observer (>10 years experience) being unaware of the results of CMR imaging. The contrast-filled catheter was used for image calibration. The severity of a coronary stenosis was derived from one single view showing the maximal reduction in absolute luminal diameter and a significant coronary stenosis was defined as ≥50% luminal diameter reduction in vessels with ≥2 mm diameter; a significant left main stenosis was considered double-vessel disease.

Statistical analysis

For all continuous parameters, mean ± standard deviation are given. The paired Student’s t-test was used to assess statistical significance of continuous variables. Group differences for categorical variables were tested with the χ² or Fisher’s exact test. All tests were two-tailed; P < 0.05 was considered significant.

Sensitivity, specificity, positive/negative predictive value, and accuracy were calculated according to the standard definitions. To correct for possible correlations due to the structure of the data (four readings in each of 150 patients), confidence limits were calculated from a population-averaged mixed logistic regression model with patients as random effect and readers as fixed effect. For positive and negative predictive values, where reference populations are different between readers, confidence limits were calculated for each reader separately based on a binomial model as described.
by Clopper and Pearson. Readers were compared with respect to their diagnostic performance using $F$-tests of two-way analysis of variance models with the factors patient and reader.

Two different quantitative measures of agreement were used to allow comparisons to other publications. Cohen’s kappa was applied to measure agreement between pairs of readers and averaged over all pairs to evaluate overall agreementusing the following grading: 0–0.2 (poor), 0.21–0.4 (fair), 0.41–0.6 (moderate), 0.61–0.8 (substantial), and 0.81–1.0 (nearly perfect). However, because the kappa coefficient does not allow linear model building to evaluate the influence of covariates on agreement, an individually determined measure of disagreement was introduced that permitted identification of conditions of rather good and rather moderate agreement. This disagreement measure for 0/1 scale was determined as follows. First, for each reader, the average absolute difference to all other readers was calculated. Then, the reader-specific disagreements were averaged to obtain an overall measure of disagreement ($D$). The reported values of $D$ represent the percentages of disagreement and are to be interpreted as follows. If disagreement is, for example, 10%, then two randomly selected readers will disagree in every 10th judgement.

General linear models were applied to evaluate the dependence of diagnostic performance and disagreement based on the presence or absence of significant CAD and the number of diseased coronary vessels. Data analysis was performed using SPSS for Windows 11.5 (SPSS Inc.) and STATA 9.0.

## Results

### Patient characteristics and dobutamine stress testing

Tables 1 and 2 summarize the clinical and haemodynamic data. One hundred and twenty-nine (86.0%) out of 150 patients reached target heart rate. The reasons for premature termination included chest pain and dyspnoea in six patients (4.2%), hypotension in one (0.7%), transient bradycardia in one (0.7%), arrhythmia in four (2.8%), and a new or worsening wall motion abnormality in nine (6.3%). In 57 patients (38.0%), there were wall motion abnormalities at rest in at least one segment (45 patients with hypokinesis and 12 with akinesis).

### Diagnostic performance

Table 3 shows the diagnostic performance for all readers and each single reader (MR1–MR4). The sensitivities of the readers with regard to single-, double-, and triple-vessel disease were 75.0, 81.7, and 80.4%. In addition, sensitivity for the detection of inducible wall motion abnormalities being indicative of coronary luminal narrowing ≥50% did not significantly differ between patients with or without wall motion abnormalities at rest (81.6 and 75.0%, $P = 0.161$).

### Disagreement in interpretation of DSMR

The values of the disagreement measure $D$ provide the percentage of disagreement and are to be interpreted as follows. If a patient is randomly selected from the study population, and two readers are randomly selected, then a disagreement $D$ of, for example, 10% indicates that two randomly selected readers will disagree in every 10th judgement.

In the present study, the readers disagreed in all patients in approximately every seventh case (14.9% of the cases) on test positivity or negativity. Disagreement of the DSMR readers was similar in patients with or without CAD (15.2 and 12.7%, respectively; $P = 0.130$) and was not significantly influenced by the number of diseased coronary vessels (18.2, 11.0, and 18.8% in single-, double-, and triple-vessel disease, respectively; $P = 0.063$).

### Agreement in interpretation of DSMR

The average kappa coefficient of agreement regarding the presence or absence of inducible wall motion was 0.59 for all readers; the kappa values ranged from 0.52 (MR1 and MR4) to 0.76 (MR3 and MR4).
For the assignment of inducible wall motion abnormalities to the LAD or the combined RCA/LCX territory, the average kappa value of agreement was 0.51 (range from 0.39 to 0.78).

Image quality of the single-centre image database

Image quality was considered to be good/excellent in 96.4% of standard views (Figure 1).

Representative examples of DSMR examinations are given in Figure 2A and B.

Discussion

Our data show that (i) multiple observers from different institutions performing a diagnostic reading of DSMR achieve a high diagnostic accuracy in detection of inducible LV wall motion abnormalities in patients with significant CAD, (ii) disagreement of experienced DSMR readers occurs in approximately every seventh case and reader agreement regarding test positivity or negativity results in a kappa value of 0.59, (iii) diagnostic performance and disagreement of experienced DSMR readers are independent of the presence of CAD and the number of diseased coronary vessels, and (iv) image quality of the DSMR approach as routinely used at the single centre is consistently high.

In previous studies using high-dose DSMR, high diagnostic accuracies have been reported ranging from 82 to 88%, and its superiority in comparison to dobutamine stress echocardiography has been shown, especially in patients with suboptimal echocardiographic image quality.

CMR imaging of LV function is technically easy to perform and has been well standardized with respect to image acquisition strategies. As the technique is now regarded as highly reproducible and reliable for the assessment of LV function and wall motion, the variability of visual wall motion assessment remains the most influential component and may limit its diagnostic value for detection of patients with significant CAD.

The diagnostic accuracy achieved in the current, fully blinded read lies within the range of previously published data on DSMR and thus further confirms its reliability in detection of significant CAD. Importantly, diagnostic accuracy and reader disagreement of DSMR were not influenced by the presence or absence of CAD or the number of diseased vessels. This is an advantage over stress echocardiography, for which a higher diagnostic accuracy and reader agreement have been shown in multiple-vessel disease.

Image quality is regarded as one of the most important factors for the accuracy of dobutamine stress echocardiography. Despite consecutive inclusion of patients, in the present DSMR study, readers from different institutions graded >96% of images as excellent/good and only 0.4% of the standard views were regarded as non-diagnostic.

The paradigm is that excellent endocardial border delineation alone ensures high diagnostic accuracy and excellent reader agreement. In our fully blinded read, the diagnostic accuracy achieved by DSMR experts was consistently high and did not significantly differ. However, as reported for stress echocardiography, the readers obviously applied heterogeneous interpretation criteria leading either to a markedly specific (MR1 and MR2) or to a more 'balanced' reading with similarly high values for sensitivity and specificity (MR3 and MR4).

This is as well reflected by the agreement of all readers: though DSMR reader agreement was found to be only moderate (kappa value 0.59), it was clearly higher than that reported for stress echocardiography using fundamental imaging (kappa value 0.37, fair agreement) or second harmonic imaging with prior definition of uniform interpretation criteria (kappa value 0.55, moderate agreement). Of note, the kappa value of two DSMR readers was 0.76 (MR3 and MR4), which represents substantial agreement.

It is noteworthy, though, that in face of only moderate overall reader agreement of DSMR experts, standardized reading criteria need to be set for DSMR, despite the consistently high diagnostic accuracy achieved in the present study. The impact of uniform reading and analysis criteria on image interpretation has been described for other imaging modalities also and generally, a higher agreement could be demonstrated when these interpretation criteria were followed (e.g. for myocardial perfusion scintigrams, the kappa value could be increased from 0.45 to 0.66).
It can reasonably be assumed that in clinical routine further improvements of DSMR accuracy will most likely rely on (i) implementation of standardized reading criteria and specific training for DSMR readers and (ii) potential methods to quantify regional wall motion. In preliminary studies, it has been shown that the application of myocardial tagging may help to better characterize the regional myocardial function and identify additional patients with significant CAD. Real-time analysis tools for direct visualization of regional myocardial strain (e.g. FastHARP) are becoming available and their value is being currently assessed.

Study limitations

Patients with contraindications to CMR imaging (non-compatible biometallic implants, claustrophobia) were not included and patients with claustrophobia can account for as much as 4-6% in a similar population. Thus, even though DSMR testing may not be applicable in all patients, the diagnostic value is high in all eligible patients.

Currently, only a few centres perform DSMR in daily clinical routine and thus, the number of experienced readers is limited. The present study did not assess how much and what kind of training is necessary to achieve the level of an experienced DSMR reader. Further studies are needed to address this issue in order to provide clear recommendations for a specific training and to ensure high-quality reading of DSMR.

For our study, we used the angiographically defined coronary diameter reduction as the standard of reference. Yet, it is known from invasive measurements of coronary flow reserve that the haemodynamical relevance of coronary luminal stenoses may vary considerably. The haemodynamic consequences of stenotic lesions, and thus the resultant ischaemic response of the myocardium, depend on several other parameters as well, e.g. stenosis morphology (excentricity), stenosis location (bifurcational lesions) or the combined flow limiting effects of serial stenoses. Our aim, though, was to predict, whether invasive therapy was required for a given patient in a routine clinical setting, and to determine the value of DSMR as a ‘gatekeeper’ for further invasive testing and therapy.

Commonly, non-invasive testing is conducted prior to invasive coronary angiography and, thus, the interpreter is unaware of the coronary dominance type; in addition, it has been recognized that the inferolateral region exhibits an extensive overlap of coronary arterial supply corresponding to the RCA and LCX territory. Thus, for our territorial analysis, we considered the combined RCA/LCX territory.

Conclusions

In the setting of multiple observers from different institutions performing a diagnostic reading of DSMR examinations conducted at a single centre, DSMR proved to be a reliable test for the detection of significant CAD with experienced DSMR readers achieving a high diagnostic accuracy. Disagreement regarding test positivity or negativity occurred in every seventh case. DSMR correctly identified patients with significant CAD independent of the number of diseased coronary vessels. Image quality of the DSMR technique as routinely performed at the single centre was highly rated. When compared with the diagnostic
performance of available techniques, the present DSMR approach is sufficient to be incorporated into clinical decision-making in patients with suspected or known CAD.

Acknowledgements

We are indebted to our CMR staff for expert assistance during the stress examinations, to Uwe Koehler, PhD, and Philip Krueger, PhD for their invaluable technical support during generation of the PC image database, and to Uwe Kokartis for performing the quantitative angiographic measurements. We thank Philips Medical Systems for supplying the PC-based CMR viewing software (Marc Kouwenhoven and Bert Verdonck).

Conflict of interest: none declared.

References