JACC Vol. 27, No. 2
February 1996:330–6

Analysis of Interinstitutional Observer Agreement in Interpretation of Dobutamine Stress Echocardiograms

RAINER HOFFMANN, MD, HARALD LETHEN, MD, THOMAS MARWICK, MD, FACC,*
MARIAROSARIA ARNESE, MD,† PAOLO FIORETTI, MD,† ALESSANDRO PINGITORE, MD,‡
EUGENIO PICANO, MD,‡ THOMAS BUCK, MD,§ RAIMUND ERBEL, MD, FACC,§
FRANK A. FLACHSKAMPF, MD, PETER HANRATH, MD, FACC

Aachen and Essen, Germany; Cleveland, Ohio; Rotterdam, The Netherlands; and Pisa, Italy

Objectives. **This study sought to determine the degree of interinstitutional agreement in the interpretation of dobutamine stress echocardiograms.**

Background. **Dobutamine stress echocardiography involves subjective interpretation. Consistent methods for acquisition and interpretation are of critical importance for obtaining high interobserver agreement and for facilitating communication of test results.**

Methods. **Five experienced centers were each asked to submit 30 dobutamine stress echocardiograms (dobutamine up to 40 μg/kg body weight per min and atropine up to 1 mg) obtained in patients undergoing coronary angiography. Thus, a total of 150 dobutamine stress echocardiograms were interpreted by each center without knowledge of any other patient data. Left ventricular wall motion was assessed using a 16-segment model but was otherwise not standardized. No patient was excluded because of poor image quality or inadequate stress level. Echocardiographic image quality was assessed using a five-point scale.**

Results. **Angiographically significant coronary artery disease (≥50% diameter stenosis) was present in 95 patients (63%). By a majority decision (three or more centers), the sensitivity, specificity and accuracy of dobutamine echocardiography were 76%, 87% and 80%, respectively. Abnormal or normal results of stress echocardiography were agreed on by four or all five of the centers in 73% of patients (mean kappa value 0.37, fair agreement only). Agreement on the left anterior descending artery territory (78%) was similar to that for the combined right coronary artery/left circumflex artery territory (74%), and for specific segments the agreement ranged from 84% to 97% and was highest for the basal anterior segment and lowest for the basal inferior segment. Agreement was higher in patients with no (82%) or three-vessel coronary artery disease (100%) and lower in patients with one- or two-vessel disease (61% and 68%, respectively). Agreement on positivity or negativity of stress test results was 100% for patients with the highest image quality but only 43% for those with the lowest image quality ($p = 0.003$).**

Conclusions. **The current heterogeneity in data acquisition and assessment criteria among different centers results in low interinstitutional agreement in interpretation of stress echocardiograms. Agreement is higher in patients with no or advanced coronary artery disease and substantially lower in those with limited echocardiographic image quality. To increase interinstitutional agreement, better standardization of image acquisition and reading criteria of stress echocardiography is recommended.**

(J Am Coll Cardiol 1996;27:330–6)

Dobutamine stress echocardiography is an accurate, noninvasive technique for the diagnosis of coronary artery disease (1,2). As the assessment of improvement or deterioration of regional wall motion during the test is subjective, not only the accuracy but also the agreement between interpreters are important considerations in the expansion of the test to the clinical arena. This is because interobserver agreement influences the ability of physicians to communicate with each other regarding the results and therapeutic implications.

From the Medical Clinic I, Rheinisch Wesfalische Techische Hochschule Aachen, Aachen, Germany; *Cleveland Clinic Foundation, Cleveland, Ohio; †Thoraxcenter Rotterdam, Rotterdam, The Netherlands; ‡University of Pisa, Pisa, Italy; and §University Gesamthochschule, Essen, Germany.

Manuscript received February 22, 1995; revised manuscript received September 21, 1995, accepted September 26, 1995.

Address for correspondence: Dr. Rainer Hoffmann, Medical Clinic I, Klinikum RWTH Aachen, Pauwelsstrasse 30, 52057 Aachen, Germany.

Most diagnostic methods have been examined for observer variability of assessment including clinical examination (3,4), electrocardiography (5,6), exercise electrocardiography (7), perfusion scintigraphy (8) and coronary angiography (9,10). Wall motion analysis by means of echocardiography is an example of the difficulty with interobserver agreement because interpretation of test results is inherently subjective. The interobserver variability of stress echocardiographic results has been examined in a small patient group (11) but was confined to readers from the same institution, and the results may have been influenced by local standards and conventions not explicitly stated, which may influence diagnostic decisions, especially in borderline situations. To circumvent these considerations, we designed a multicenter study to assess the interobserver agreement of experienced readers from different institutions in interpreting dobutamine stress echocardiographic images without any additional clinical data. The present report details the

0735-1097/96/$15.00
0735-1097(95)00483-1

analysis of agreement within a panel of five expert readers who independently read 150 dobutamine stress echocardiographic studies submitted in equal shares by all five institutions.

Methods

Patients. At each of the five institutions, 30 consecutive patients scheduled for angiography due to suspected coronary artery lesions underwent dobutamine stress echocardiography. The results of stress echocardiography did not influence the decision to perform angiography. No patient was excluded on the basis of poor echocardiographic image quality. Patients with a previous Q wave myocardial infarction were excluded from the study, as were those with congestive heart failure, severe congenital or acquired valvular heart disease or documented cardiomyopathy. A total of 146 patients (116 men, 30 women; mean [±SD] age 46 ± 12 years, range 29 to 78) were investigated, of whom 4 underwent a second examination after interventional therapy had been performed.

Previous myocardial revascularization had been performed in 26 patients (20 with coronary angioplasty, 6 with coronary artery bypass grafting), and 14 patients had a non-Q wave infarction. One hundred thirteen patients had angina at the time of examination. Antianginal therapy before the examination included nitrates in 78 patients, calcium channel antagonists in 48 and beta-adrenergic blocking agents in 45. In 39 patients antianginal medication was stopped before stress echocardiography.

Dobutamine stress protocol. Dobutamine was infused intravenously, starting at a dose of 5 μg/kg body weight per min for 3 min, increasing the dosage every 3 min to 10, 20, 30 and 40 μg/kg per min. In case 85% of age-predicted maximal heart rate was not reached and the test was negative, additional atropine was given intravenously in 0.25-mg steps every minute, up to a maximal dosage of 1.0 mg. The infusion was stopped before reaching maximal pharmacologic stress for any of the following reasons: development of new wall motion abnormalities; progressive and severe angina; >0.2 mV of downsloping ST segment depression; development of hypotension (decrease in systolic blood pressure ≥20 mm Hg); significant hypertension (systolic blood pressure >240 mm Hg); and dyspnea or severe ventricular arrhythmias.

Monitoring. Blood pressure and 12-lead electrocardiograms (ECGs) were recorded at baseline and at the end of each dobutamine stage or before the premature cessation of the test. The presence of horizontal or downsloping ST segment depression of at least 0.1 mV, 0.08 s after the J point versus baseline recordings was considered diagnostic for myocardial ischemia. The rate-pressure product was calculated by multiplying systolic blood pressure and heart rate.

Image acquisition. Image acquisition was performed with the patient in the left lateral decubitus position before, during and after dobutamine infusion. No patient was excluded because of low image quality. In each patient, imaging was attempted in the parasternal long-axis and short-axis as well as in the apical four- and two-chamber views; in 32 patients, one

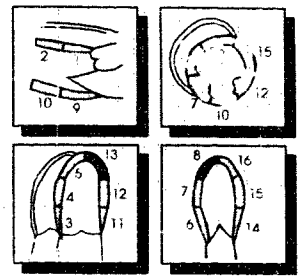


Figure 1. Sixteen-segment scheme according to regional wall motion evaluation with a two-region distribution of coronary circulation. **Leftward hatching** = left anterior descending distribution; **rightward hatching** = right coronary/left circumflex distribution; **numbers** are segment numbers.

or more views were not recorded because of insufficient acoustic windows. All views obtained at baseline were reacquired at the end of each dobutamine stress level and when new wall motion abnormalities or worsening of preexisting wall motion abnormalities occurred. However, only the images at rest and peak stress conditions were selected and stored for exchange. Individual centers maintained their standard practice for image acquisition, so that images from 82 patients (from three centers) were recorded on videotape only, and images from 68 patients (from three centers) were digitized on-line into quad-screen cine loop format. In all cases, recorded images were transferred to SVHS-videotapes to have a uniformly readable media and distributed to all centers for interpretation. For those studies primarily being digitized and transferred onto videotape, direct comparison of rest and stress images in the quad-screen cine loop format was still possible because long sessions of the digitized images were taped.

Stress echocardiographic interpretation. To base interpretation exclusively on imaging data, each of the five centers evaluated all 150 dobutamine stress investigations using a standardized report form without knowledge of any patient data apart from the echocardiographic images. Thus, no information on maximal dobutamine or atropine dosage and reason for termination of the dobutamine stress test was given to the interpreting physician. Clinical or angiographic data of these patients were also not made available.

For wall motion analysis, the left ventricle was divided according to a 16-segment model (Fig. 1). A two-region distribution of coronary circulation was used for further analysis of segmental wall motion abnormalities related to coronary circulation, with segments 3, 6, 7, 9, 10, 11 and 12 part of the combined right coronary or left circumflex territory, as opposed to the left anterior descending territory that included the remaining segments 1, 2, 4, 5 and 8. Segmental wall motion was evaluated and scored according to the American Society of Echocardiography: 1 = normal; 2 = hypokinetic; 3 = akinetic; 4 = dyskinetic. Results of dobutamine stress echocardiography were considered positive if the score increased by at least one

grade (e.g., hypokinesia to akinesia) in at least one segment with stress. Wall motion abnormalities at rest not deteriorating with stress were not considered positive test results. Other interpretive criteria were not stipulated.

Echocardiographic image quality was assessed by the coordinating center for each stress test on a five-point scale: *A* = complete endocardial definition and wall thickening with exactly similar image views at rest and during dobutamine; *B* = visualization of all segments but not adequate as A; *C* = inadequate visualization of one or two segments but adequate visualization of adjacent segments within the same territory; *D* = inadequate visualization of three or more segments but adequate visualization of adjacent segments of the same territory; *E* = inadequate visualization of segments comprising one or more whole territories.

Coronary angiography and angiographic analyses. The coronary anatomy of all patients was evaluated by coronary angiography, and the degree of coronary artery stenoses was determined by the caliper method in at least one projection. Significant coronary artery stenosis was considered present when ≥50% reduction of vessel diameter was observed in at least one major coronary artery.

Statistical analysis. Calculations of sensitivity, specificity and diagnostic accuracy of dobutamine stress echocardiography for the detection of coronary artery disease were performed on the basis of the majority opinion (three or more centers). In addition, calculations were also performed with the criterion of ≥70% lumen diameter narrowing.

Concordant interpretation was identified as the presence of identical readings from four or all five of the interpreting centers. In addition to analyzing the overall agreement of dobutamine stress echocardiographic interpretation between all institutions, the role of variations in acquisition and interpretation was investigated by focusing on echocardiograms submitted and interpreted at individual institutions. The influence of severity of coronary artery disease, echo image quality and the recording system used on the reproducibility of data were similarly recorded.

The kappa test was used to test the hypothesis that agreement was greater than chance alone (12). Average coefficients of agreement (kappa) were computed for the five readers of the different institutions. The coefficient of agreement (kappa) was graded as follows: 0 to 0.2 = poor to slight; 0.21 to 0.4 = fair; 0.41 to 0.6 = moderate; 0.61 to 0.8 = substantial; 0.81 to 1.0 = nearly perfect.

Results

Coronary angiography. Coronary angiography identified coronary artery disease in 95 patients. One-vessel disease was present in 59 patients (62% of those with significant coronary artery disease), two-vessel disease in 22 (23%) and three-vessel disease in 14 (15%). Distribution of affected vessels was left anterior descending coronary artery in 62 (41%) patients, circumflex artery in 33 (22%), and right coronary artery in 50 (33%). Bypass graft stenosis was counted as equivalent to stenosis of the grafted native vessel.

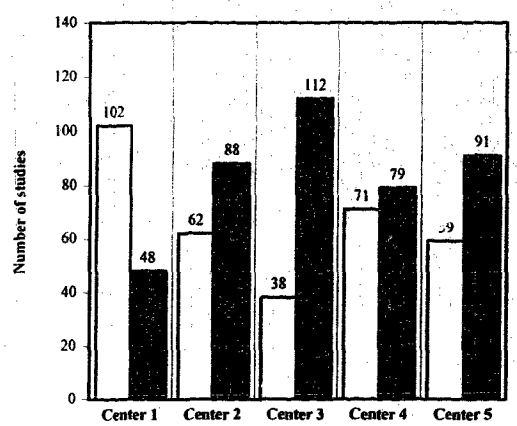


Figure 2. Evaluation of negativity (**solid bars**) versus positivity (**open bars**) for the 150 dobutamine stress echocardiograms by institution. **Numbers above bars** = number of patients

Dobutamine stress test. The average maximal dobutamine dosage was 35 ± 8 μg/kg per min, and additional atropine was given to 53 patients. This resulted in an increase in heart rate from 67 ± 12 (range beats/min 45 to 105) at baseline to 122 ± 27 beats/min (range 40 to 210) at peak stress. The rate-pressure product increased from 10,964 ± 2,623 mm Hg min^{-1} to 20,136 ± 5,245 mm Hg min^{-1} with maximal pharmacologic stress.

Fifty-seven patients developed angina during the stress test, and ECG changes with maximal stress developed in 41. Dobutamine stress was stopped prematurely in 37 patients.

Echocardiography. There was a significant difference between the centers in identifying dobutamine stress test results as positive. On average, the results of 67 dobutamine stress tests were evaluated as positive, but this result ranged from 38 to 102 between centers (Fig. 2).

Sensitivity, specificity and diagnostic accuracy. Using the majority opinion (three or more) of the five centers to define the presence of positive or negative results in all 150 patients, dobutamine stress echocardiography had a sensitivity of 76% for the detection of coronary artery disease. When studies submitted by different institutions were compared, the sensitivity ranged from 67% to 79%. The average specificity was 87% (range 73% to 100%), and accuracy for detection of coronary artery disease was 80% (range 77% to 87%). For those lesions having a ≥70% diameter stenosis, sensitivity, specificity and accuracy were 83%, 83% and 83%, respectively.

Agreement between institutions. Four or all five of the institutions agreed on dobutamine stress test result abnormality or normality in 110 (73%) of the 150 studies. The average coefficient of agreement among investigators from different institutions for normality or abnormality of the dobutamine echocardiography stress test result (kappa) was 0.37, which represents fair agreement. For those 31 studies in whom the

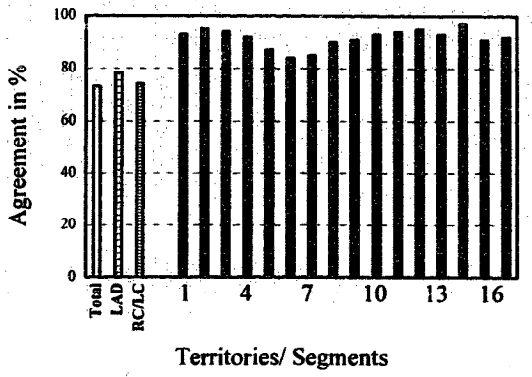


Figure 3. Majority interinstitutional agreement on a four of five or five of five basis for positivity or negativity of dobutamine stress test results by specific coronary territory and for each of 16 ventricular segments. LAD = left anterior descending coronary artery; RC/LC = right and left circumflex coronary arteries.

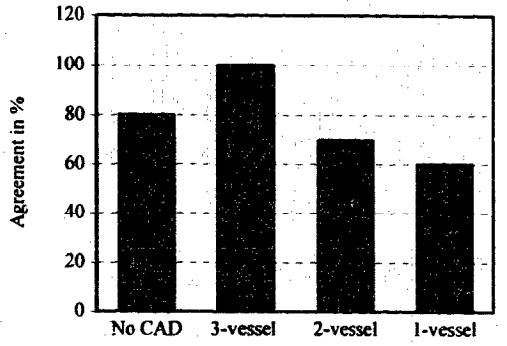


Figure 4. Majority interinstitutional agreement on positivity or negativity of dobutamine stress echocardiographic results by severity of coronary artery disease (CAD).

majority of the observers identified a baseline wall motion abnormality, majority agreement was slightly lower (71%) than for those 119 studies without baseline wall motion abnormality (74%). When only those left ventricular segments supplied by the left anterior descending artery are considered, agreement concerning the presence or absence of new wall motion abnormalities was 78% (mean kappa value 0.37), similar to the combined right coronary artery/left circumflex artery territory in whom a majority agreement of 74% (mean kappa value 0.33) was reached. Agreement in specific segments ranged from 84% to 97% and was highest for the basal anterior segment (segment 14 on the 16-segment model) and lowest for the basal inferior segment (segment 6 on the 16-segment model) (Fig. 3).

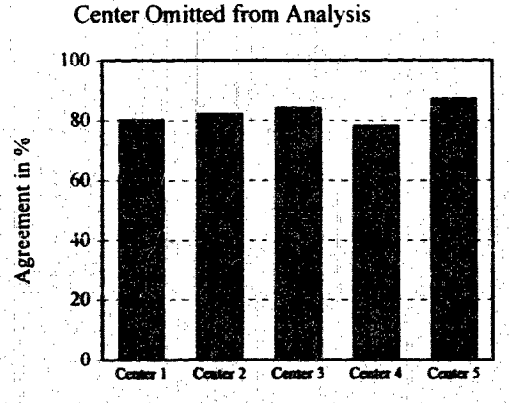
Agreement corresponding to disease severity. For patients with no coronary artery disease (n = 55), majority agreement on dobutamine stress echocardiographic result was 82% on a four of five or five of five basis. In patients with three-vessel disease (n = 14), the majority agreement was 100%, exceeding the concordance of results in patients with only one- or two-vessel disease (Fig. 4).

Agreement omitting one analyzing institution. As a result of the observation of different rates of positive stress echocardiographic evaluations among the centers, the agreement on dobutamine stress echocardiographic positivity or negativity was evaluated for the remaining four centers after omitting the interpretations of one center. Majority agreement (three of four or four of four centers) on overall positivity or negativity increased for the remaining institutions each time one of the centers was omitted. However, there were quite different increases in agreement related to the agreement of all five centers (73%). Agreement of three of four or four of four centers ranged between 77% and 87%, depending on which of the five analyzing centers was omitted (Fig. 5).

Agreement on dobutamine stress tests submitted by single centers. To analyze whether dobutamine echocardiographic stress tests results submitted by one institution were more clearly interpretable, and thus resulted in higher agreement than those of other institutions, the majority agreements for the 30 dobutamine echocardiographic tests submitted by each of the five institutions were analyzed separately. Majority agreement on the 30 dobutamine echocardiography stress tests submitted by single institutions ranged from 70% to 80%, indicating similar interpretability of stress echocardiograms (Fig. 6).

Effect of image quality on agreement. Image quality had a significant effect on the overall agreement on presence or absence of inducible wall motion abnormalities. In the 13 patients with dobutamine stress echocardiographic tests of the highest image quality (grade A, allowing good delineation of all left ventricular wall segments), agreement on the presence or absence of inducible wall motion abnormalities was 100%

Figure 5. Majority interinstitutional agreement on a three of four or four of four basis for positivity or negativity of dobutamine stress echocardiographic results when one institution is omitted from the analysis.



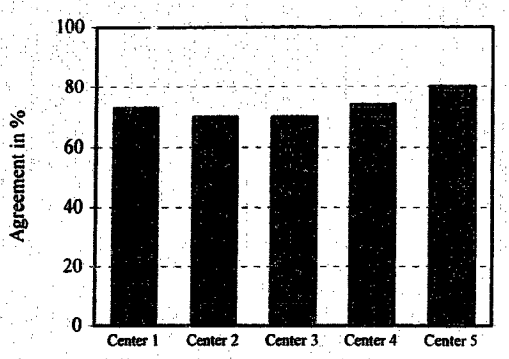


Figure 6. Majority interinstitutional agreement on dobutamine stress echocardiographic test results (n = 30) submitted by each institution.

on a four of five or five of five basis. However, for the 14 patients (9.5%) with the lowest image quality (grade E), agreement was only 43%, significantly lower than for those with the highest image quality ($p = 0.003$) (Fig. 7). Although no patient was excluded from the full analysis because of poor image quality, when patients with grade E images were omitted, the interinstitution agreement increased to 77%.

Agreement related to image recording system. Agreement on a four of five or five of five basis for those 82 dobutamine echocardiographic stress tests recorded in the videotape format was 71%. This was similar to the 76% agreement in 68 studies recorded on videotape after digital image processing.

Discussion

Variability in test interpretation. Interobserver variability is a well known problem in cardiology, and almost all diagnostic methods have been examined for interobserver and intraobserver agreement. A surprisingly low level of interobserver agreement has been reported for interpretation of the rest 12-lead ECG. In a study of 20 physicians who were asked to interpret 100 ECG responses as normal or showing a previous myocardial infarction or nonspecific abnormalities, Segall (5) found that ≥70% of the readers agreed in only 77% of the studies. A similar interobserver difference has been reported in the frequency of "abnormal" exercise ECG diagnosis. Blackburn (7) asked 14 readers from seven institutions to interpret 38 exercise ECG test results as normal, abnormal or borderline and found that abnormalities were identified in 5% to 58% of studies.

The results of imaging studies are also prone to interobserver variability. Atwood et al. (8) assessed the agreement of four experienced readers in the interpretation of 100 thallium perfusion images. The interobserver agreement for a majority of observers (three of four or four of four observers) from the same institution was found to be 75% for an abnormal and 68% for a normal interpretation. A striking interobserver variability has been reported in the interpretation of coronary stenosis severity and left ventricular function evaluation at coronary angiography. Zir et al. (9) found that in only 13 of 20 coronary angiograms (65%) did all four experienced coronary angiographers (from the same institution) concur regarding the significance of a stenosis (defined as ≥50% diameter lumen narrowing). De Rouen et al. (10) reported a 31% overall disagreement between the assessment of single readers and those of an expert panel on the classification of a coronary vessel as ≥70% stenotic. For left ventriculograms, the average percent disagreement in interpretation of wall motion between observers dividing the ventricle into five segments was 42%. The agreement on left ventricular contraction grade assessment was found to be only fair, with an average coefficient of agreement (kappa) of 0.34 among 11 observers (10).

Like most techniques in medicine, dobutamine stress echocardiography requires observer interpretation and is thus subjective to some extent. Several groups have reported their interobserver variabilities to both evaluate the degree of consistency in the interpretation of stress echocardiograms and to measure for interpretation validity. Sawada et al. (1) reported agreement between two observers from the same institution on the presence or absence of a stress-induced abnormality in 91% of cases, and Beleslin et al. (13) reported an interobserver agreement between two observers of 93%. A high degree of correlation between wall motion score indexes evaluated by two different observers has also been reported before as well as after dynamic stress during exercise echocardiography (11). However, the evaluation of wall motion score as a global measurement does not allow a statement on agreement of specific left ventricular segments, and has only limited value for assessment of specific regions. Moreover, these studies of interobserver variability have focused on observers from the same institution, with all observers participating frequently in joint stress echocardiographic reading sessions.

Sources of interobserver variability. The reasons for interobserver variability in stress echocardiography may be classified into four broad categories: 1) normal regional variability of left ventricular function in response to dobutamine stress; 2)

Figure 7. Majority interinstitutional agreement by assessment image quality of dobutamine stress echocardiograms.

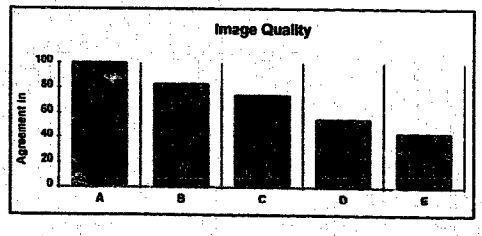


Image Quality	A	B	C	D	E
Agreement	100%	83%	74%	55%	43%
Number of patients	13	47	35	41	14

variability of image quality; 3) variability of stress echocardiographic hardware, in particular image storage on videotape or in digitized format; and 4) observer variability.

The normal human left ventricular functional response to dobutamine stress is characterized by substantial regional variability within each patient, as shown by studies with cine computed tomography (14), magnetic resonance imaging (15) and two-dimensional echocardiography (16). This heterogeneity of left ventricular wall thickening can be magnified by dobutamine infusion even in subjects without coronary artery disease, and there may be ambiguity with respect to differentiating this normal variation from the presence of disease.

To obtain a high degree of interobserver agreement, multiple views are necessary to visualize all segments of the left ventricle with high quality images. Competence of the observer in the interpretation of stress echocardiograms is essential to reach high accuracy and thus a high interobserver agreement. This has been stressed by Picano et al. (17), who found accuracy in the interpretation of stress echocardiograms to be significantly higher for echocardiographers with a high level of experience with stress echocardiography. However, for high interobserver agreement, there also needs to be homogeneity in the classification of a wall motion as being normal, hypokinetic, akinetic or dyskinetic. This applies especially to the classification of hypokinetic wall motion, which may range from slight to severe and where the cutoff from what is considered normal may vary between observers and institutions. Mild hypokinesia may be viewed as normal clinical variability at one institution and be considered pathologic at another.

Significance of the study. To our knowledge, the present study is the first to analyze interobserver agreement among experienced stress echocardiographers from different institutions. There was considerable discrepancy between centers in their evaluation of dobutamine stress echocardiograms. The number of studies evaluated as positive ranged from 38 to 102 between the five centers. This is reflected by an average coefficient of agreement (kappa) of 0.37, which is only fair. The fact that one center found significantly higher rates of dobutamine echocardiographic positivity than others indicates a systematic difference in reading criteria, implying different "thresholds" at different places to classify stress echocardiographic results as positive or negative. This stresses the necessity to further refine and unify reading criteria.

Interobserver agreement was significantly influenced by echocardiographic image quality, with higher agreement in patients with high image quality than in those with limited image quality. Although no patient was excluded due to low echocardiographic image quality, improvement in interobserver agreement could have been obtained by excluding the 10% of patients with the lowest image quality. These findings suggest that it is reasonable to exclude patients from stress echocardiography if image quality is insufficient to visualize all left ventricular segments.

Segmental agreement was found to be highest for the anterobasal and lowest for the inferobasal region. This might be due to the lower endocardial border definition in the inferobasal region in many patients and reduced wall thickening present at baseline compared with other left ventricular segments. Reduced accuracy in the detection of inducible wall motion abnormalities for this area has been reported previously (18).

Although we did not show better concordance on the basis of digitized image loops compared with videotaped recordings, the present study design may have placed the digital approach at a disadvantage, because recording digitized data on videotape is associated with a slight decrement in image resolution, and this approach limits the ability to use stop-frame images, which facilitate interpretation. Although digitized cine loop displays are desirable to simplify analysis, the place of digital image processing for dobutamine echocardiography is unresolved, with data supporting (19) as well as disproving (20) improvement in accuracy by application of this technique. Ideally, compatibility of different digitizing systems would permit digitized images to be readable by all systems independent of the system acquiring the images. Furthermore, it is desirable that the digitized segment of the original echocardiographic screen displays the left ventricle with maximal spatial resolution.

The problem of considerable interobserver variability has been reduced for other imaging techniques. Computer-assisted quantitative coronary angiography has diminished variability in interpretation of coronary angiograms. The agreement in interpretation of planar thallium-201 imaging as having ischemia or no ischemia could be increased from a kappa value of 0.36 without standardization on interpretation to 0.71 with standardization of display and quantification (21). Similar improvements should be the goal with stress echocardiography. In the future, improved image quality allowing better endocardial border definition, possibly obtainable by left heart contrast echocardiography or computer algorithms to evaluate wall motion abnormalities, may diminish the subjective nature of interpretation and reduce the interobserver variability.

Limitations of the study. Quantitative coronary angiography was not performed in the present study. However, this was not thought to be inadequate because the main purpose of the study was to determine agreement in interpretation of stress echocardiograms, and the issue of correlation between dobutamine stress echocardiography and angiography has been repeatedly analyzed in previous studies.

In the present study, data acquisition as well as equipment were heterogeneous, with videotaped recordings in 82 patients and digital image storage in 68. Videotapes were used as the final uniform media because different image acquisition and display systems were implemented at the different centers. Furthermore, in 32 patients only apical views were available for evaluation.

No standardization of stress echocardiographic reading criteria was defined. These criteria were clearly heterogeneous between centers, with aggressive, liberal "overreaders" as well as conservative "underreaders." We did nothing to prevent these disparate reading policies because the documentation of

these variations was part of the rationale of the study. The results describe the current heterogeneity of these criteria among expert groups.

Interinstitutional agreement was found to be dependent on the severity of the coronary artery disease. Thus, agreement might have been better in a patient group with a lower prevalence of one- or two-vessel coronary artery disease because this was the group with the greatest disagreement.

Conclusions and clinical implications. A high intercenter agreement on the interpretation of test results is important for clinical decisionmaking. The present study shows that the current heterogeneity in data acquisition and reading criteria between different centers result in relatively low interinstitutional agreement in the interpretation of stress echocardiograms. Thus, to increase agreement in the interpretation of stress echocardiograms between different institutions, additional efforts toward standardization and communication are needed.

We acknowledge the expert statistical analysis of Thomas Reinecke.

References

1. Sawada SG, Segar DS, Ryan T, et al. Echocardiographic detection of coronary artery disease during dobutamine infusion. Circulation 1991;83: 1605–14.
2. Marcovitz PA, Armstrong WF. Accuracy of dobutamine stress echocardiography in detecting coronary artery disease. Am J Cardiol 1992;69:1269–73.
3. Koran LM. The reliability of clinical methods, data and judgments. New Engl J Med 1975;293:642–95.
4. Raftery EB, Holland WW. Examination of the heart: an investigation into variation. Am J Epidemiol 1967;85:438–44.
5. Segall HN. The electrocardiogram and its interpretation: a study of reports by 20 physicians on a set of 100 electrocardiograms. Can Med Assoc 1960;82:?–6.
6. Simonson E, Tuna N, Okamoto N, Toshima H. Diagnostic accuracy of the vectorcardiogram and electrocardiogram. Am J Cardiol 1966;17:829–78.
7. Blackburn H. The exercise electrocardiogram: differences in interpretation. Am J Cardiol 1968;21:871–80.
8. Atwood JE, Jensen D, Froelicher V, et al. Agreement in human interpretation of analog thallium myocardial perfusion images. Circulation 1981;64: 601–9.
9. Zir LM, Miller SW, Dinsmore RE, Gilbert JP, Harthorne JW. Interobserver variability in coronary angiography. Circulation 1976;53:627–32.
10. De Rouen TA, Murray JA, Owen W. Variability in the analysis of coronary arteriograms. Circulation 1977;55:324–8.
11. Oberman A, Fan PH, Nanda N, et al. Reproducibility of two-dimensional exercise echocardiography. J Am Coll Cardiol 1989;14:923–8.
12. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Measurement 1960;20:37–46.
13. Beleslin BD, Ostojic M, Stepanovic J, et al. Stress echocardiography in the detection of myocardial ischemia. Circulation 1994;90:1168–76.
14. Lanzer P, Garrett J, Lipton MJ. Quantitation of regional myocardial function by cine computed tomography: pharmacologic changes in wall thickness. J Am Coll Cardiol 1986;8:682–92.
15. van Rugge FP, Holman ER, van der Wall EE, de Roos A, van der Laarse A, Bruschke AVG. Quantitation of global and regional left ventricular function by cine magnetic resonance imaging during dobutamine stress in normal human subjects. Eur Heart J 1993;14:456–63.
16. Hausnerova E, Gottdiener JS, Hecht GM, Hausner PF. Heterogeneity of regional left ventricular wall thickening during dobutamine stress echocardiography in normal subjects [abstract] Circulation 1994;90 Suppl I:I-391.
17. Picano E, Lattanzi F, Orlandini A, Marini C, L'Abbate A. Stress echocardiography and the human factor: the importance of being expert. J Am Coll Cardiol 1991;17:666–9.
18. Bach DS, Muller DWM, Gros BJ, Armstrong WF. False positive dobutamine stress echocardiograms: characterization of clinical, echocardiographic and angiographic findings. J Am Coll Cardiol 1994;24:928–33.
19. Marwick TH, D'Hondt AM, Mairesse GH, et al. Comparative ability of dobutamine and exercise stress in inducing myocardial ischaemia in active patients. Br Heart J 1994;72:31–8.
20. Attenhofer CH, Pellikka PA, Oh JK, Roger VL, McCully RB, Seward JB. Comparison of frame-grabbed cine-loop images and videotape record in stress echocardiography: a prospective study [abstract]. Circulation 1994;90 Suppl I:I-391.
21. Wackers FJT, Bodenheimer M, Fleiss JL, Brown M. Factors affecting uniformity in interpretation of planar thallium-201 imaging in a multicenter trial. J Am Coll Cardiol 1993;21:1064–74.