

# Agreement Between Centers on the Interpretation of Exercise Echocardiography

Jesús Peteiro,<sup>a</sup> Ángel M. Alonso,<sup>b</sup> Rafael Florenciano,<sup>c</sup> Carlos González Juanatey,<sup>d</sup> Gonzalo de la Morena,<sup>c</sup> Ignacio Iglesias,<sup>f</sup> Mar Moreno,<sup>e</sup> and Miguel A. Rodríguez<sup>f</sup>

<sup>a</sup>Unidad de Ecocardiografía, Hospital Juan Canalejo, A Coruña, Spain.

<sup>b</sup>Unidad de Ecocardiografía, Hospital Txagorritxu, Vitoria, Álava, Spain.

<sup>c</sup>Unidad de Ecocardiografía, Virgen de la Arrixaca, Murcia, Spain.

<sup>d</sup>Unidad de Ecocardiografía, Hospital Xeral, Lugo, Spain.

<sup>e</sup>Unidad de Ecocardiografía, Hospital Gregorio Marañón, Madrid, Spain.

<sup>f</sup>Unidad de Ecocardiografía, Hospital de León, León, Spain.

**Introduction and objectives.** A low-to-moderate level of agreement on the interpretation of dobutamine echocardiography has been reported, but there are no similar findings on exercise echocardiography. The objectives of this study were to assess the level of agreement between centers on the use of exercise echocardiography and to evaluate the accuracy of the technique when used in a blinded manner.

**Patients and methods.** Six institutions with experience in exercise echocardiography each sent 25 study results to the other centers. Of these, 15 were positive or negative studies on consecutive patients undergoing coronary angiography, and 10 were on non-diabetic patients who had non-coronary chest pain or were asymptomatic and whose pretest probability of coronary artery disease was <10%. Each institution evaluated 150 studies: 125 blinded and 25 of their own with knowledge of clinical data.

**Results.** For 116 patients (78%), four or more of the 5 centers blindly evaluating each study agreed with the positive or negative result. The average kappa coefficient was 0.48 (intercenter range, 0.45-0.52). The percentage agreement was higher with three-vessel disease (93%, range, 85%-95%), with left anterior descending coronary artery disease (83%; range, 80%-86%), and when the referring institution reported baseline dyssynergy (86%; range, 82%-90%), dyssynergy in left anterior descending coronary artery territory (81%; range, 76%-84%), or a peak wall motion score index >1.50 (88%; range, 85%-90%). When the technique was used blinded to detect ≥50% coronary narrowing in ≥1 vessel, its sensitivity, specificity and accuracy were 68%, 66%, and 67%, respectively, with wide variability between centers.

**Conclusions.** There was moderate agreement between centers on the interpretation of exercise echocardiography.

When used blinded, the technique's accuracy was lower than that reported when clinical data is known.

**Key words:** Exercise echocardiography. Intercenter agreement. Accuracy.

## Concordancia intercentros en la interpretación de la ecocardiografía de ejercicio

**Introducción y objetivos.** Se ha demostrado un acuerdo bajo-moderado en la interpretación de la ecocardiografía con dobutamina, pero no se ha realizado un estudio similar con ejercicio. El objetivo fue evaluar la concordancia intercentros de la ecocardiografía de ejercicio y la precisión de la técnica realizada de forma «ciega».

**Pacientes y método.** Cada uno de los 6 centros remitió 25 casos a los demás centros: 15 eran pacientes consecutivos en los que se había realizado una coronariografía y 10 eran no diabéticos, asintomáticos o con dolor precordial atípico y con una probabilidad pretest < 10%. Cada centro corrigió 150 casos: 125 de forma «ciega», y 25 de su propio centro conociendo los datos clínicos.

**Resultados.** Cuatro o más de los 5 centros que evaluaban de forma ciega cada caso coincidieron en un diagnóstico positivo o negativo en 116 pacientes (78%). La media de los valores kappa obtenidos era 0,48 (rango intercentros, 0,45-0,52). El grado de acuerdo era mayor en la enfermedad de 3 vasos (93%; rango, 85-95%) y de la arteria descendente anterior (83%; rango, 80-86%), y cuando el remitente informaba disineria basal (86%; rango, 82-90%), disineria en el territorio de la descendente anterior (81%; rango, 76-84%) o índice de motilidad segmentaria pico > 1,50 (88%; rango, 85-90%). La sensibilidad, la especificidad y la precisión diagnóstica media de la técnica «a ciegas» para estenosis coronaria ≥ 50% en ≥ 1 vaso fue del 68, el 66 y el 67%, respectivamente, pero con una amplia variabilidad intercentros.

**Conclusiones.** La concordancia intercentros de la ecocardiografía de ejercicio es moderada. La precisión diagnóstica «a ciegas» es inferior a la comunicada cuando se conocen los datos clínicos.

**Palabras clave:** Ecocardiografía de ejercicio. Acuerdo intercentros. Precisión diagnóstica.

SEE EDITORIAL ON PAGES 9-11

Study financed by the RECAVA Cardiovascular Network.

Correspondence: Dr. J.C. Peteiro.  
P.º Ronda, 5, 4.º izqda. 15011 A Coruña. España.  
E-mail: pete@canalejo.org

Received June 9, 2005.

Accepted for publication October 18, 2005.

## ABBREVIATIONS

CAD: coronary artery disease.  
ECG: electrocardiogram.  
EE: exercise echocardiography.

## INTRODUCTION

One of the main limitations of stress echocardiography is its variability. Hoffmann's first study, although carried out with fundamental imaging and without uniform reading criteria, found only low agreement in the interpretation of dobutamine stress echocardiography.<sup>1</sup> This improved in a subsequent study by the same author when using harmonic imaging and uniform reading criteria.<sup>2</sup>

However, and surprisingly, although exercise echocardiography (EE) is the oldest,<sup>3</sup> most sensitive and safest<sup>4,5</sup> method of administering stress, as well as being the most widely used,<sup>6</sup> no study has been done to investigate intercenter agreement using this technique. Thus, the purpose of this study was to evaluate: *a*) intercenter agreement on EE, and *b*) the sensitivity, specificity, and diagnostic accuracy of the technique under blinded conditions.

## PATIENTS AND METHODS

Six centers participated in the study, each having broad experience with stress echocardiography and, in particular, with EE (having carried out between 1000 and 7000 EE). Each of the 6 centers sent 25 study results. Of these, 15 were positive or negative EE studies on consecutive patients undergoing coronary angiography within 3 months of EE; and the other 10 studies were on non-diabetic patients, also consecutive, asymptomatic or with non-coronary chest pain and with a <10% pretest probability of coronary artery disease (CAD) according to sex, age, and risk factors.<sup>7</sup> Thus, each center evaluated 150 cases: 125 under blinded conditions (data from other centers) and 25 from their own center with knowledge of the clinical data.

State-of-the-art equipment was used with second harmonic imaging and stress digitalization packs (Sonos-5500, Philips, used by 4 centers and Vivid-5, GE, used by 2 centers). Each study was sent to the coordinating center on optical disk, which then re-distributed them to the other centers either in the same format or on video tape, depending on each center's capabilities. Apical 4- and 2-chamber and parasternal long-axis and short-axis views were compared, at rest and under stress in quad-screen format.

## Reading Criteria

Uniform reading criteria<sup>8</sup> were used. A positive EE was defined when there was at least 1 abnormal segment at rest or under stress, or tardokinesia in the event that there were no alterations in conduction, and negative EE when no segment was abnormal at rest or under stress, or there was hypokinesia isolated from the posterobasal and/or septobasal segment, unless accompanied by dyssynergy in one adjacent segment.

Each center categorized every positive result as necrosis (regional alteration in wall motion that persisted or improved with stress), ischemia (alteration in wall motion with stress), ischemia plus necrosis in the same territory (alteration in baseline wall motion that worsened in the same territory with stress), or ischemia at a distance (alteration in wall motion in 1 or more territories at baseline, with the appearance of new alterations in wall motion in a different territory with stress). Wall motion score index at rest and under stress was calculated in each reading by dividing the left ventricle into 16 segments.<sup>9</sup> The territories affected in each study were determined according to whether they were dependent on the left anterior descending coronary artery (LAD), circumflex artery (Cx), right coronary artery (RC), or a combination of them.

In addition, each center objectively and subjectively assessed the quality of each study. A segment quality score was used for the objective assessment where a score of 3 was assigned to each segment with good visibility (thickness and displacement), 2 to those with fair visibility, 1 to those with poor visibility, and 0 to the non-visible. For the subjective assessment, each study was qualified as good, fair, poor, or non-interpretable.

## Statistical Analysis

The SPSS 12.0 statistical package was used. Continuous variables are presented as mean $\pm$ SD. Discrete variables are presented as percentages. Comparisons between patients with and without CAD were done via X<sup>2</sup> test for discrete variables and Student *t* test for continuous variables. Agreement between 2 centers was estimated by the percentage agreement (negative or positive EE) found after analyzing studies from other centers without including the cases of the centers themselves (150-50 cases=100 cases). The percentage agreement and kappa coefficients ( $\kappa$ ) (proportion of agreement higher than that due to chance) were as follows: a  $\kappa$  coefficient between 0 and 0.20 was considered very low; between 0.21 and 0.40, low; between 0.41 and 0.60, moderate; between 0.61 and 0.80, good; and between 0.80 and 1.0, excellent.<sup>10</sup> The sensitivity, specificity, and diagnostic accuracy for each center were calculated by the centers assessing their own cases, as well as by blinded assessment of the other

**TABLE 1. Clinical Characteristics, Medication, and Baseline Electrocardiogram\***

|                            | CAD<br>(n=58) | Without CAD<br>(n=91) | P      |
|----------------------------|---------------|-----------------------|--------|
| Age, years, mean±SD        | 63±10         | 57±13                 | <.01   |
| Male sex, n (%)            | 47 (81)       | 48 (53)               | <.0001 |
| Diabetes mellitus, n (%)   | 20 (34)       | 6 (7)                 | <.0001 |
| Hyperetension, n (%)       | 30 (52)       | 38 (42)               | NS     |
| Cholesterol, n (%)         | 33 (58)       | 20 (22)               | <.0001 |
| Smoking, n (%)             | 28 (48)       | 11 (12)               | <.0001 |
| Medical history, n (%)     | 2 (3)         | 7 (8)                 | NS     |
| Previous AMI, n (%)        | 24 (41)       | 12 (13)               | <.0001 |
| PTCA/surgery, n (%)        | 15 (26)       | 14 (15)               | NS     |
| Nitrates, n (%)            | 21 (36)       | 10 (11)               | <.0001 |
| Calcium antagonists, n (%) | 8 (14)        | 10 (11)               | NS     |
| Beta-blockers, n (%)       | 25 (43)       | 13 (14)               | <.0001 |
| ACEI, n (%)                | 21 (36)       | 21 (23)               | NS     |
| Digoxin, n (%)             | 1 (2)         | 0 (0)                 | NS     |
| Atrial fibrillation, n (%) | 1 (2)         | 2 (2)                 | NS     |
| LBBB, n (%)                | 2 (3)         | 5 (5)                 | NS     |

\*PTCA indicates percutaneous transluminal coronary angioplasty; LBBB, complete left bundle branch block; SD, standard deviation; CAD, coronary artery disease; AMI, acute myocardial infarction; ACEI, angiotensin-converting enzyme inhibitors.

centers' cases. Sensitivity was defined as the percentage of cases with positive EE among patients with significant coronary stenosis in at least 1 vessel. Specificity was defined as the percentage of cases with negative EE among patients without angiographically demonstrated coronary lesions or with a low pretest probability. Diagnostic accuracy was defined as the percentage of successes (cases with positive EE and CAD, plus cases with negative EE and absence of CAD) from total patients.

## RESULTS

One hundred and forty-nine studies were available for analysis (1 study was excluded due to poor images). Contrast agents were used for left ventricular opacification in 9 studies (6%) and the stress study was done with peak stress imaging in 124 cases (83%).

### Baseline Clinical Characteristics and Response to Stress

Significant CAD was found in 58 patients (39%) as defined by stenosis  $\geq 50\%$  in  $\geq 1$  coronary artery, main branch, or coronary artery bypass graft, whereas 91 patients (61%) had angiographically demonstrated non-significant CAD (n=37), or low pretest probability according to the previous definition (n=54). There was 1-vessel disease in 24 patients with CAD, 2-vessel disease in 18, and 3-vessel disease in 16. The LAD was stenosed in 40 patients, the RC in 39 and the Cx

**TABLE 2. Response to Stress\***

|                                                   | CAD      | Without CAD | P      |
|---------------------------------------------------|----------|-------------|--------|
| ME, mean±SD                                       | 8±2      | 10±3        | <.0001 |
| Maximum cardiac frequency, beats/min, mean±SD     | 127±24   | 144±20      | <.0001 |
| Clinical data (+), n (%)                          | 14 (24%) | 7 (8%)      | <.01   |
| ECG (+), n (%)                                    | 20 (34%) | 14 (15%)    | <.001  |
| Non-diagnostic ECG, n (%)                         | 18 (31%) | 16 (18%)    | <.001  |
| <85% theoretical maximum cardiac frequency, n (%) | 32 (55%) | 28 (31%)    | <.01   |

\*SD indicates standard deviation; ME, metabolic equivalents; CAD, coronary artery disease; ECG, electrocardiogram.

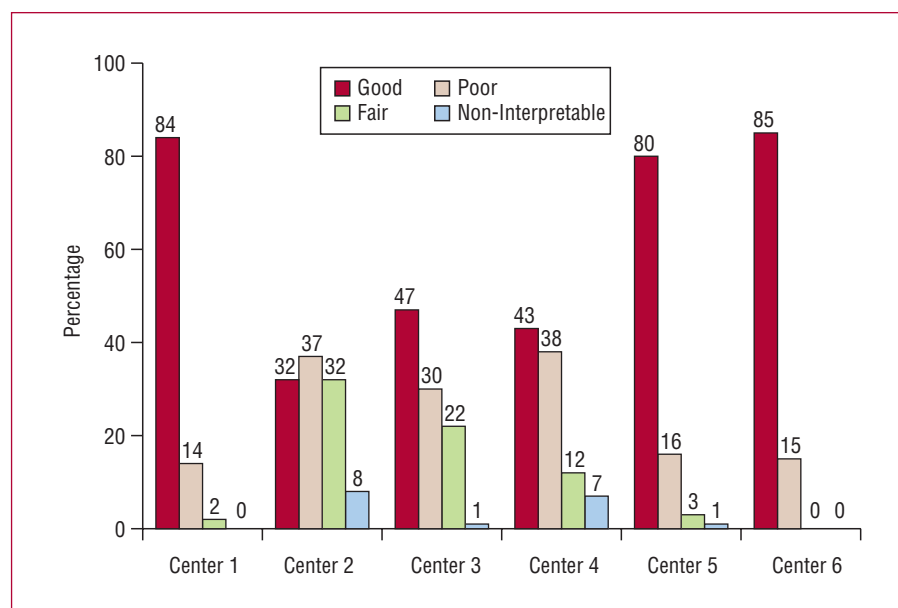
in 29. Table 1 shows baseline clinical characteristics, medication, and baseline electrocardiogram (ECG) data in patients with and without CAD. Table 2 shows data on response to stress in patients with and without CAD.

### Image Quality

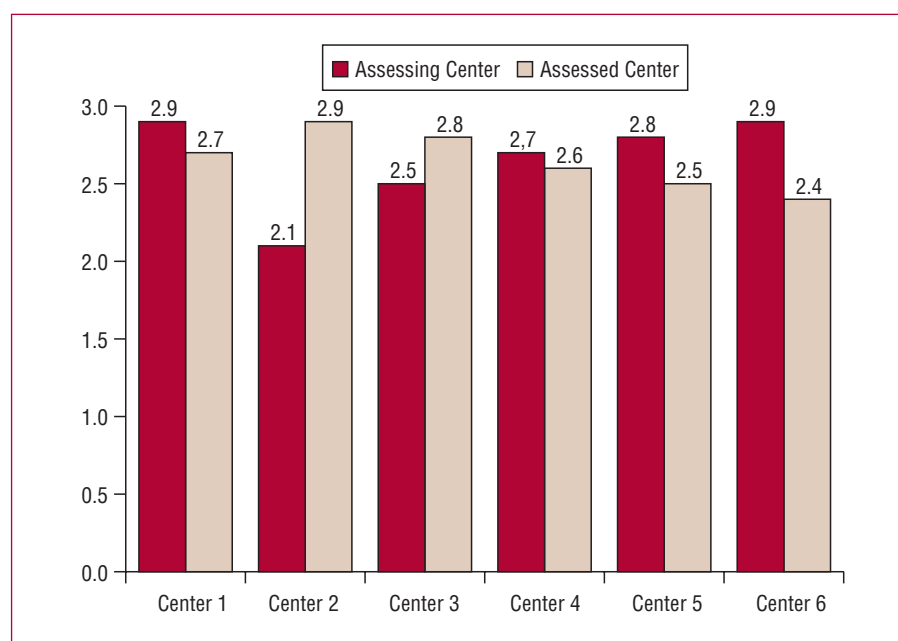
The subjective assessment of the quality of the studies differed significantly between the different centers. Some centers described a high percentage of studies as good ( $\geq 80\%$  of studies), whereas others only considered less than half the cases as good and between 0 and 8% as non-interpretable (Figure 1). The same differences were found when the different centers calculated the quality of the segment wall motion score (Figure 2). In general, the centers that qualified the others as worse tended to have better quality images according to the other centers.

### Agreement

Four or more of the 5 centers that assessed each case under blinded conditions agreed on a positive diagnosis of CAD in 51 patients and on a negative diagnosis in 65 patients, which means that there was agreement on a total of 116 of the 149 patients (78%). There was agreement regarding a positive or negative diagnosis of CAD in  $4.1 \pm 0.9$  centers out of the 5 centers. There was a mean  $\kappa$  coefficient of 0.48 between the different centers, with mean intercenter  $\kappa$  coefficients ranging from 0.45 to 0.52. The percentage agreement and the  $\kappa$  coefficients in different scenarios are shown in the Table 3. The percentage agreement and the  $\kappa$  coefficient differed according to the diagnosis of regional contractility anomalies by the referring center, and the percentage agreement was greater when the referring center had detected baseline anomalies in regional contractility in a given territory, contractility anomalies at rest and/or with stress in the LAD territory, or when a worse wall motion score index with stress were reported (Table 4).



**Figure 1.** Percentage of studies qualified as good, fair, poor, and non-interpretable according to the different centers.



**Figure 2.** Scoring of quality of studies from other centers according to the referring center (light columns) and scoring of quality of studies from each center according to the other centers (dark columns).

## Sensitivity, Specificity, and Diagnostic Accuracy

The percentage of positive and negative readings, as well as the sensitivity, specificity, and diagnostic accuracy differed between the different centers when assessed under blinded conditions (Figure 3). There were 2 centers with high sensitivity but low specificity and 1 where the opposite occurred.

The mean sensitivity, specificity, and diagnostic accuracy of the 6 centers regarding stenosis  $\geq 50\%$  in at least 1 vessel (according to visual estimation) was 68%, 66%, and 67%, respectively. The mean sensitivity and

specificity of the different centers was similar in tests which were higher or lower than submaximal (68% vs 64% and 66% vs 65%, respectively). These data contrast with the mean sensitivity, specificity and diagnostic accuracy of each center when they assessed their own cases (Figure 4). If we consider that the positive cases of coronary artery disease were those with stenosis  $\geq 50\%$  in at least 1 vessel or with a history of acute myocardial infarction (AMI) and baseline dyssynergy according to the diagnosis of the referring center, then the sensitivity, specificity, and diagnostic accuracy in the blinded reading were similar to those obtained with stenosis  $\geq 50\%$  as the only criterion: 69% (intercenter range, 53%-82%), 70%

**TABLE 3. Percentage Agreement and Kappa Coefficients According to the Results of the Coronary Angiography and Echocardiographic Image Quality\***

|                                        | Percentage Agreement<br>(Mean Intercenter<br>Range) | κ (Mean Intercenter<br>Range) |
|----------------------------------------|-----------------------------------------------------|-------------------------------|
| 3 vessels (n=16)                       | 93 (85-95)                                          | †                             |
| 2 vessels (n=18)                       | 63 (48-71)                                          | 0.32 (0.20-0.46)              |
| 1 vessels (n=24)                       | 76 (67-82)                                          | 0.52 (0.37-0.62)              |
| 0 vessels (n=37)†                      | 71 (69-78)                                          | 0.42 (0.32-0.57)              |
| LAD disease (±other vessels)<br>(n=40) | 83 (80-86)                                          | 0.57 (0.46-0.63)              |
| RC/Cx disease (±LAD)<br>(n=46)         | 74 (72-76)                                          | 0.39 (0.31-0.45)              |
| Without CAD (n=91)                     | 71 (68-75)                                          | 0.37 (0.32-0.40)              |
| Good study quality (n=68)              | 73 (69-76)                                          | 0.46 (0.41-0.53)              |
| Suboptimal quality (n=81)              | 75 (74-77)                                          | 0.51 (0.48-0.53)              |

\*RC indicates right coronary artery; Cx, circumflex artery; LAD, left anterior descending coronary artery; CAD, coronary artery disease.

†Demonstrated by coronary angiography.

‡The κ coefficient is not stated, because it cannot be calculated for some of the comparisons between centers, as occurs when one or both centers only give 1 of the 2 possible diagnoses (CAD or normal) for all the cases analyzed.

(range, 49%-89%) and 69% (range, 64%-78%). The sensitivity, specificity, and diagnostic accuracy according to the decisions of the majority (4 or more centers when 5 centers were assessing data; 3 or more when 4 centers were assessing data) were similar regarding  $\geq 50\%$  stenosis in one vessel and for a criterion of  $\geq 50\%$  stenosis or a history of AMI and baseline dyssynergy depending on the referring center (72% vs 73%; 74% vs 80%; and 73% vs 77%, respectively;  $P=NS$ ):

### False Positive Readings

Out of 403 readings corresponding to 83 patients without coronary stenosis, previous AMI, or baseline

dyssynergy according to the referring center, there were 124 false (+) readings corresponding to 31% of the assessments without CAD, with a wide intercenter range (11%-51%). These false (+) readings were mainly due to ascribing contractile alterations to the RC territory (36% of the readings) or to the LAD (35%), and less often to the Cx territory (10%) or to several territories (19%). The segment wall motion score index (WMSI) measured by the assessing centers in these cases was  $1.1 \pm 0.2$  at rest and  $1.3 \pm 0.2$  with stress.

### False Negative Readings

There were 319 readings corresponding to 66 patients. Of those who had angiographically demonstrated coronary stenosis some had a history of AMI. In those who did not undergo coronary angiography, or where this was negative, all had a medical history of AMI and dyssynergy according to the referring center. There were 102 false (–) readings which corresponded to 32% of the assessments with CAD (intercenter range, 18%-47%). In most of these cases there was only 1-vessel disease (45%; LAD disease in 23 of them) or 2-vessel disease (34%), and on fewer occasions 3-vessel disease (9%) or disease in no vessels (13%). The referring center reported dyssynergy in 32 of these 66 patients (48%), which was severe (WMI, 1.50) in 12 of them (18%).

## DISCUSSION

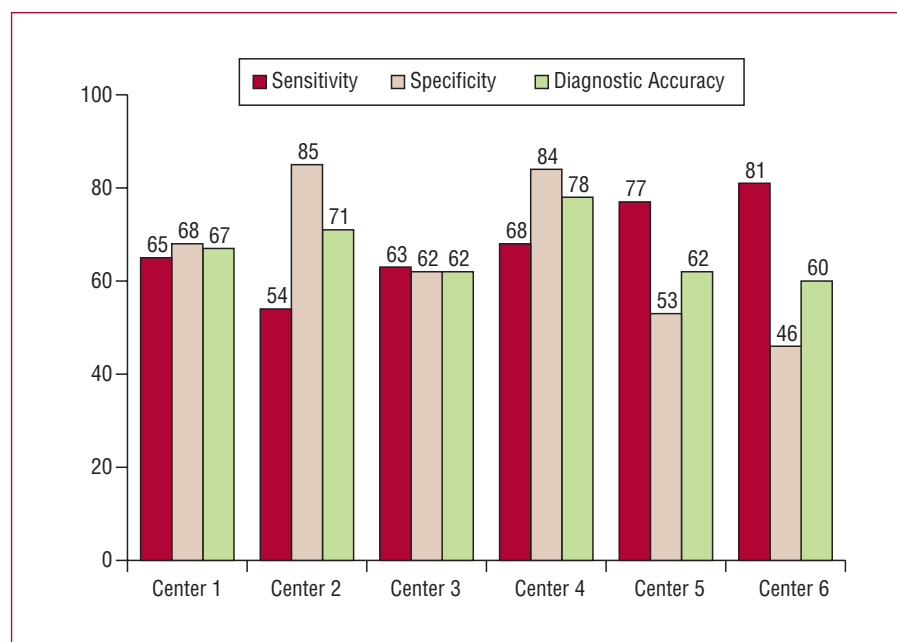
The main interest of this study lies in it being the first in which intercenter agreement on exercise echocardiography has been assessed. The main findings were as follows: a) the intercenter agreement on exercise echocardiography was moderate, and b) the sensitivity, specificity and diagnostic accuracy of the technique when carried out under blinded conditions were lower than those commonly reported

**TABLE 4. Percentage Agreement and Kappa Coefficient According to the Diagnosis of the Referring Center\***

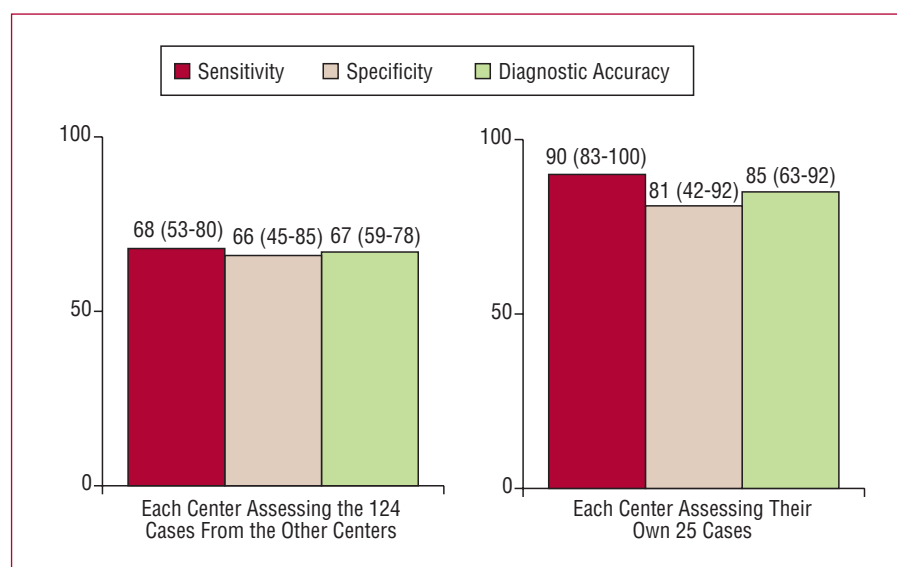
|                                               | Percentage Agreement<br>(Mean Intercenter<br>Range) | κ (Mean Intercenter<br>Range) |
|-----------------------------------------------|-----------------------------------------------------|-------------------------------|
| Necrosis at baseline echocardiography (n=33)  | 86 (82-90)                                          | †                             |
| Normal baseline echocardiography (n=116)      | 71 (66-74)                                          | 0.36 (0.28-0.40)              |
| EE: LAD territory (±other territory) (n=45)   | 81 (76-84)                                          | 0.45 (0.37-0.57)              |
| EE: RC/Cx territory (±other territory) (n=45) | 73 (65-77)                                          | 0.39 (0.29-0.47)              |
| EE: RC/Cx territory only (n=24)               | 70 (65-74)                                          | 0.35 (0.19-0.43)              |
| WMI with stress =1.00 (n=80)                  | 72 (64-79)                                          | †                             |
| WMI with stress =1.06-1.25 (n=19)             | 64 (52-68)                                          | †                             |
| WMI with stress =1.26-1.50 (n=22)             | 75 (62-85)                                          | †                             |
| WMI with stress >1.50 (n=28)                  | 88 (85-90)                                          | †                             |

\*RC indicates right coronary artery; Cx, circumflex artery; LAD, left anterior descending coronary artery; CAD, coronary artery disease; ECG, exercise echocardiography; WMI, wall motion score index.

†The κ coefficient is not stated, because it cannot be calculated for some of the comparisons between centers, as occurs when one or both centers only give one of the 2 possible diagnoses (CAD or normal) for all the cases analyzed.



**Figure 3.** Sensitivity, specificity and diagnostic accuracy of each center that assessed, under blinded conditions, the cases referred by the other centers.



**Figure 4.** Sensitivity, specificity and diagnostic accuracy of exercise echocardiography in detecting coronary lesions with  $\geq 50\%$  stenosis assessed under blinded conditions or with knowledge of clinical data and response to stress. The means of the centers and intercenter ranges are shown.

when baseline characteristics and patient response stress are known.

### Intercenter Agreement on Exercise Echocardiography

Although Hoffmann et al studied intercenter agreement on dobutamine stress echocardiography,<sup>1,2</sup> there are no similar studies on exercise echocardiography despite being more frequently used, sensitive and safe.<sup>4,6</sup> Low agreement was observed ( $\kappa=0.37$ ) in Hoffmann's first study<sup>1</sup> which was carried out with fundamental imaging and without uniform reading criteria, whereas in the second study, carried out with harmonic imaging and

uniform reading criteria, agreement was moderate ( $\kappa=0.55$ ).<sup>2</sup> The improvement in agreement seemed to be due both to using harmonic imaging and the standardization of the reading criteria, since the degree of agreement on the same patients studied with fundamental imaging was greater than in Hoffmann's first study. We used the same reading criteria as in Hoffmann's second study,<sup>2</sup> which, in general, did not involve any change in the normal clinical practice followed in each center. It could be expected that the degree of agreement on exercise echocardiography would be less than that carried out with dobutamine, since there should be better quality images with the latter technique. However, by means of uniform reading criteria and harmonic imaging the

percentage agreement was moderate, with a mean  $\kappa$  coefficient of 0.48, which is better than that of Hoffmann's first study and similar to the author's second study. The percentage agreement was greater in 3-vessel disease, in left anterior descending coronary artery disease, when there were baseline alterations in regional contractility and when the referring center reported dyssynergy in the LAD territory or serious dyssynergy: the fact of higher percentage agreement in these circumstances has clinical diagnostic and prognostic relevance, since the patients with these characteristics have a worse prognosis.<sup>11-13</sup>

### Intercenter Agreement on Other Diagnostic Techniques

Concern over the degree of agreement is not exclusive to stress echocardiography. Different degrees of variability in interpretation have been observed with other techniques. Thus, very low levels of agreement have been reported regarding the interpretation of ST-segment elevation ( $\kappa=0.05$ ) or ST-segment depression ( $\kappa=0.38$ ) between 2 centers in patients with acute coronary syndrome.<sup>14</sup> Studies on myocardial perfusion with nuclear medicine procedures also present difficulties in interpretation, since these techniques are subjective and, as in exercise echocardiography, the experience of the observer and image quality can influence the interpretation. A moderate-high agreement has been reported with thallium imaging, with  $\kappa$  coefficients ranging between 0.56 and 0.74 in 2 studies.<sup>15,16</sup> However, in a multicenter study with 25 participating hospitals, agreement between different centers without uniform reading criteria was low ( $\kappa=0.27$ ).<sup>17</sup> In a study by Candell-Riera et al.,<sup>18</sup> good agreement was found with exercise technetium-99m tetrofosmin myocardial perfusion single-photon emission computed tomography, with  $\kappa$  coefficients between 0.62 and 0.70 depending on whether topographical images or polar mapping were evaluated.<sup>18</sup> This study also found that the sensitivity of the report under blinded conditions was significantly lower than that reported when the clinical data of the patient were known.

### Sensitivity, Specificity, and Diagnostic Accuracy

Although we present mean sensitivity and specificity scores for the different centers, the variability among them regarding interpretation under blinded conditions is of more interest. However, the sensitivity, specificity, and diagnostic accuracy of the technique carried out under blinded conditions were lower for each center than when they assessed their own cases and the baseline characteristics of the patients and their response to stress were known. This finding is not surprising, but it gives

us an idea of the limitations of the technique when the pretest probability, hemodynamic, clinical, and ECG response to stress are not known. It is clear that EE can and should be used in clinical practice, but not under blinded conditions.

### Limitations

The reading format of the studies was the same for all the centers (apical 4- and 2-chamber and parasternal long-axis and short-axis views at baseline and with stress), although the quality differed since, depending on whether the centers could read optical disks, the study was recorded on video or was sent via optical disk. However, the percentage agreement and the  $\kappa$  coefficients were similar for studies of optimal and suboptimal quality. The complete study recorded on video was not sent out as was done in the study by Hoffmann et al.<sup>1</sup> This fact could lead to overestimating agreement, since the operator tends to acquire and store the images he/she considers more representative taking into account other test characteristics different from those of the image. Twenty-five percent of the patients were receiving treatment with beta-blockers and up to 40% of the tests were lower than submaximal. This fact could have led to underestimating sensitivity, although we have not found higher sensitivity in the tests that were maximal in comparison to those that were submaximal.

### REFERENCES

- Hoffmann R, Lethen H, Marwick T, Arnesen M, Fioretti P, Pingitore A, et al. Analysis of interinstitutional observer agreement in interpretation of dobutamine stress echocardiograms. *J Am Coll Cardiol.* 1996;27:330-6.
- Hoffmann R, Marwick TH, Poldermans D, Lethen H, Ciani R, van der Meer P, et al. Refinements in stress echocardiographic techniques improve interinstitutional agreement in interpretation of dobutamine stress echocardiograms. *Eur Heart J.* 2002;23:821-9.
- Wann LS, Faris JV, Childress RH, Weyman AE, Feigenbaum H. Exercise cross-sectional echocardiography in ischemic heart disease. *Circulation.* 1979;60:1300-8.
- Beleslin BD, Ostojic M, Stepanovic J, Djordjevic-Dikic A, Stojkovic S, Nedeljkovic M, et al. Stress echocardiography in the detection of myocardial ischemia. Head-to-head comparison of exercise, dobutamine, and dipyridamole tests. *Circulation.* 1994;90:1168-76.
- Acquatella H. Ecocardiografía de estrés en Latinoamérica. Revisión de 5 años (1997-2002). *Rev Esp Cardiol.* 2003;56:21-8.
- Rodríguez García MA, Iglesias-Garriz I, Corral Fernández F, Garrate Coloma C, Alonso-Orcajo N, Branco L, et al. Evaluación de la seguridad de la ecocardiografía de estrés en España y Portugal. *Rev Esp Cardiol.* 2001;54:941-8.
- Prevention of coronary heart disease in clinical practice. Recommendations of the Second Joint Task Force of European and other Societies on Coronary Prevention. *Eur Heart J.* 1998;19:1434-503.
- Hoffmann R, Lethen H, Marwick T, Rambaldi R, Fioretti P, Pin-

- gitore A, et al. Standardized guidelines for the interpretation of dobutamine echocardiography reduce inter-institutional variance in interpretation. *Am J Cardiol*. 1998;82:1520-4.
9. Bourdillon PD, Broderick TM, Sawada SG, Armstrong WF, Ryan T, Dillon JC, et al. Regional wall motion index for infarct and non-infarct regions after reperfusion in acute myocardial infarction: comparison with global wall motion index. *J Am Soc Echocardiogr*. 1989;9:398-407.
10. Fleiss JL. The measurement of interrater agreement. In: Fleiss JL, editor. *Statistical methods for rates and proportions*. New York: John & Sons; 1981. p. 212-36.
11. Arruda AM, Das MK, Roger VL, Klarich KW, Mahoney DW, Pellikka PA. Prognostic value of exercise echocardiography in 2,632 patients  $\geq 65$  years of age. *J Am Coll Cardiol*. 2001;37: 1036-41.
12. Elhendy A, Mahoney DW, Khandheria BK, Paterick TE, Burger KN, Pellikka PA. Prognostic significance of the location of wall motion abnormalities during exercise echocardiography. *J Am Coll Cardiol*. 2002;40:1623-9.
13. Peteiro J, Monserrat L, Mariñas J, Garrido I, Bouzas M, Muñoz J, et al. Valor pronóstico de la ecocardiografía de ejercicio en cinta rodante. *Rev Esp Cardiol*. 2005;58:924-33.
14. Holmvang L, Hasbak P, Clemmensen P, Wagner G, Grande P. Differences between local investigator and core laboratory interpretation of the admission electrocardiogram in patients with unstable angina pectoris or non-Q-wave myocardial infarction (a Thrombin Inhibition in Myocardial Ischemia [TRIM] substudy). *Am J Cardiol*. 1998;82:54-60.
15. Okada RD, Boucher CA, Kirshenbaum HK, Kushner FG, Strauss HW, Block PC, et al. Improved diagnostic accuracy of thallium-201 stress test using multiple observers and criteria derived from interobserver analysis of variance. *Am J Cardiol*. 1980; 46:619-24.
16. Atwood JE, Jensen D, Froelicher V, Witztum K, Gerber K, Gilpin E, et al. Agreement in human interpretation of analog thallium myocardial perfusion images. *Circulation*. 1981;64:601-9.
17. Wackers FJ, Bodenheimer M, Fleiss JL, Brown M. Factors affecting uniformity in interpretation of planar thallium-201 imaging in a multicenter trial. The Multicenter Study on Silent Myocardial Ischemia (MSSMI) Thallium-201 Investigators. *J Am Coll Cardiol*. 1993;21:1064-74.
18. Candell-Riera J, Santana-Boado C, Bermejo B, Armadans L, Castell J, Casás I, et al. Impacto de los datos clínicos y concordancia interhospitalaria en la interpretación de la tomografía miocárdica de perfusión. *Rev Esp Cardiol*. 1999;52:892-7.