Madness and method in stress echo reading

A. Varga, E. Picano, C. Dodi*, A. Barbieri, L. Pratali and O. Gaddi†

Institute of Clinical Physiology, Pisa, Italy; *Cardiovascular Unit, Guastalla Hospital, Reggio Emilia, Italy; †Cardiology Department, S.M.N. Hospital, Reggio Emilia, Italy

Aim To assess whether ‘eye education’ through short-term, high-intensity joint reading sessions may improve diagnostic accuracy and inter-observer agreement among beginners.

Methods and Results Seventeen cardiologists with absent to minimal (<100 studies performed) previous stress echo experience independently and blindly read 18 stress echo studies, nine at the beginning (‘pre-training’ set) and nine at the end (‘post-training’ set) of a 2 day stress echo school which included a joint reading session of 50 tapes. The two sets were balanced as far as type of stress and image quality. The 17 observers had an average accuracy score of 51±16.4 before and 64.3±8.7% after the training (P<0.005). Concordant (i.e. >14 readers giving the same response) interpretation occurred in three out of nine studies before and in eight out of nine studies after the training (33% vs 88%, P<0.01). Kappa values went from 0.14 (poor) before to 0.39 (fair, close to moderate) after the training.

Conclusion Short-term, high-intensity dedicated training in stress echo, with joint reading sessions and consensus development of reading criteria significantly increased accuracy and markedly reduced the inter-observer variability in the reading of stress echos by beginners. If there is a Shakespearean madness in stress echo reading, ‘yet there is a method in’t’ (Hamlet, II, II, 205–206).

(Eur Heart J 1999; 20: 1271–1275)

Key Words: Agreement, stress echo, variability.

See page 1216 for the Editorial comment on this article

Introduction

Stress echocardiography is being increasingly promulgated as a user-friendly and cost effective diagnostic tool for coronary artery disease[1]. However, easy access to ultrasound technology can paradoxically become a problem in everyday clinical practice. In the absence of quality control, with simple 2D echocardiography and a drug costing a few Euros, every physician can become a stress echocardiographer. The combination of unrestricted access and the subjective, qualitative nature of reading, creates enormous potential for ‘stress echocardiographic disease’, and sets the stage for the transition of the technique from ‘established technology’ to a phase of discredit and backlash[2].

It was previously shown that the interpretation of stress echocardiographic tests by an echocardiographer without specific training severely underestimates the diagnostic potential of this technique. To build a learning curve and reach a plateau of diagnostic accuracy examination of 100 stress echocardiographic studies were found to be more than adequate to offset this deficit[3]. However, it has recently been shown that, even among experienced observers working in an institution with an undisputed reputation in stress echo, inter-institutional variability can be substantial[4]. This inter-institutional variability often appears to be a clear demonstration of the clinical unreliability, perhaps close to madness, of the method[5]. The aim of this study was to evaluate whether there is a method in this stress echo madness. Accordingly, we evaluated individual diagnostic accuracy and class inter-observer agreement before and after ‘eye’ education through short-term, high-intensity, joint reading sessions during a 2 day ‘Stress Echo School’ organized for cardiologist beginners in stress echo.

Methods

Twenty-two Italian cardiologists attending a 2 day closed-number Stress Echo School were initially considered. By selection criteria, all had completed a full training programme in echocardiography (M-mode,
B-mode and colour Doppler techniques). All 21 cardiologists were considered as stress-echo beginners, with absent to minimal (<100 studies performed) previous stress echo exposure. Five participants had to be excluded from the final analysis because they were either unable to complete the second test at the end of the school (n=2), or to attend the full duration of the course (n=3). The remaining 17 cardiologists (five females, 12 males, age 37 ± 4 years) completed both reading sessions and were included in the data analysis. Each of them read independently and blindly 18 stress echo studies, nine before (‘pre-training’ set) and nine after (‘post-training’ set) a 2 day joint reading session of 50 tapes.

Image display

Reading was tested using two sets of stress echos, each consisting of nine studies. The two sets were similar in terms of type of stress (dipyridamole: n=4; dobutamine: n=3; ergonovine n=2 for each group); image quality (excellent in 3; good in 4; sufficient in two studies for each group); left ventricular dysfunction at rest (absent in 7 and present in two studies for each group). Only the images at rest and peak stress were selected and shown. All images were digitally acquired with a Vingmed CFM 800 machine (General Electric-Vingmed, Trondheim, Norway) and shown in quad-screen cine-loop format with Echo-Pac software.

Stress echocardiographic interpretation

Each reader evaluated the stress echo studies using a standardized report form, without knowing any patient data apart from the echocardiographic images. The patients’ clinical and angiographic data were also withheld from the reader. The left ventricle was divided into 16 segments and segmental wall motion was scored according to the American Society of Echocardiography (from 1=normal up to 4=dyskinetic)⁶. A test was considered positive when the regional score index increased by one grade or more in at least one segment.

Data analysis

Individual accuracy analysis

For each case, the gold standard against which the individual reader’s accuracy was assessed was the expert reader, in agreement with angiographic verification: negative dipyridamole or dobutamine studies with angiographically normal coronary arteries, or positive studies with significant (≥ 70% visually assessed stenosis) coronary artery disease. All cases with a positive ergonovine test had the coronary vasospasm documented during coronary angiography. In all positive cases, the wall motion dysfunction during stress was consistent with the site of stenosis and/or spasm during coronary angiography. Each reading for each case was scored as ‘concordant’ (=1, agreement regarding positivity vs negativity, and in ‘positive’ cases, this also included the region involved); ‘partially concordant’ (=0.5, agreement regarding positivity, but with discrepancies on >2 segments and/or, for each segment, with discrepancies >2 grades of regional wall motion score); ‘discordant’ (=0, disagreement regarding positivity vs negativity or, in positive cases, complete disagreement regarding the involved region).

Inter-individual agreement analysis

Independent of the accuracy, concordant interpretation was identified as the presence of fundamentally identical readings (regarding positivity vs negativity and, in positive cases, regarding the main site of dysfunction) in >14 out of 17 observers.

Statistical analysis

Continuous data were expressed as means and standard deviation. Pre- vs post-training differences were tested for significance by means of the Student’s t-test for paired values. The chi-square test and the Fisher’s exact test were used for comparing discrete variables (number of concordant interpretations, pre- vs post-training). The required level of significance was $P$ less than 0.05. The kappa test was used to test the hypothesis that agreement was greater than chance alone⁷,⁸ and was measured on two sets of stress echo studies, before and after the training. Average coefficients of agreement (kappa) were computed for the 17 readers. The coefficient of agreement 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=nearly perfect.

Results

Expert (‘teacher’) reading

In the ‘teacher’ reading of the 18 tests, 14 studies were positive and four negative for ischaemia. For each of the two sets of nine studies, the expert (‘teacher’) reading gave positive results in seven and negative results in two studies. In all cases, the reading was consistent with subsequent angiographic verification. In all positive studies the site of the wall motion abnormality was consistent with the anatomical location of the coronary artery stenosis or spasm.

Beginner (‘student’) reading: individual accuracy data

A comparison of the student and teacher readings (the latter was assumed to be the diagnostic gold standard)
was 4·58 ± 1·48 out of nine at the start, and 5·79 ± 0·79 out of nine at the end of the sessions. As a percentage, the average agreement went from 51 ± 16·4% to 64·3 ± 8·7% (P<0·005): Fig. 1. Nine readers (‘responders’) had a >20% improvement in reading performance, whereas the remaining five (‘non responders’) had only a marginal improvement (<20%) or an actual decrease in accuracy.

**Beginner (‘student’) reading:**

**Inter-individual agreement data**

For each case, concordant interpretations (with 15 or more readers in agreement) occurred in 3/9 studies before and in 8/9 studies after training: 33% vs 88%; P<0·01 (Fig. 2). Kappa values went from 0·14 (poor) before to 0·39 (fair, close to moderate) after the training.

**Discussion**

To acquire intra-reader diagnostic power and to deflate inter-reader variability, there is no substitute for a joint reading with an expert echocardiographer. Obviously, not all echocardiographers are created equal and not all respond in the same way to training. Phenomena such as fatigue, saturation and demotivation may explain why some students do not respond to training. In some, accuracy and reproducibility even worsen. However, on average, training increases accuracy and decreases variability.

**How to minimize variability**

Variability was significantly reduced when agreement before and after training were compared. This improved agreement could not be attributed to known factors modulating agreement, which can be technology-related, stress-related, or patient-related. It has been suggested that the use of digital images instead of videotapes may improve agreement[9,10]. However, in our study only digital images were used.

Not all stresses deteriorate image quality to the same extent. Image pollution is greater with exercise than pharmacological stresses[11], and among the pharmacological stresses, image pollution is greater with dobutamine than with dipyridamole[12] or ergonovine[13]. However, in our study only pharmacological stresses were used, and the proportion of stresses (dobutamine, dipyridamole, and ergonovine), was similar in the sets before and after training. It is also known that reproducibility is poor in patients with resting images of borderline quality[4]. However, the quality of resting images was similar in our two sets of studies and cannot have affected the improvement in agreement. The last, and certainly not least factor affecting variability is the ‘human factor’[3], i.e. eye education obtained through...
Previously it has been shown that a learning curve of >100 stress echo studies read with expert supervision is necessary to optimize the diagnostic yield of the stress echo technique[3]. The present study reinforces this message, demonstrating that 50 studies with expert supervision determine a significant, albeit clearly suboptimal, improvement in diagnostic performance. Another important aspect of stress echo performance is variability, which may be largely independent of accuracy. Diagnostic accuracy is a function of experience; for a given diagnostic accuracy every observer has his/her own sensitivity/specificity curve: there are ‘overreaders’ (high sensitivity–low specificity) and ‘underreaders’ (low sensitivity–high specificity), depending on whether images are aggressively or conservatively interpreted as abnormal. It appears likely that these statements exceed the framework of the stress echo reading and can be universally applied in cardiological diagnosis. Such variability, striking even between experienced observers[4], represents the recognized limitation of stress echo[17], and may be so substantial as to be close to madness[5]. Our study, however, confirms what clinical stress echocardiographers have long suspected[18], i.e. that ‘though this be madness, yet there is method in’t’[19].

The organization of Stress Echo School was made possible by unrestricted educational grant of Sonotron Vingmed-Italia. We are grateful to Andrea Ripoli for his expert statistical review. We are also grateful to the attendees of the School, who volunteered to take part in the reading sessions: Ilaria Moneta, Mirka Bacchi, Paolo Vasini, Sergio Callegari, Virginia Agnoletto, Alessandro Navazio, Giovanni Tortorella, Gigliola Bedetti, Isabella Abelli, Francesco Marzani, Biagio Sassone, Stefano Cappelli, Fabrizio Corghi, Fabio Sgura, Massimo Panteleoni, Maria Cristina Turina, Fogliero Pignatti, Carlo Piscicelli.

References


