MÆDICA - a Journal of Clinical Medicine

ORIGINAL PAPERS: CLINICAL OR BASIC RESEARCH

Reproducibility in echocardiography: *clinical significance, assessment, and comparison with other imaging methods*

Andrei-Dumitru MARGULESCU, MD, Mircea CINTEZA, MD, PhD,
Dragos VINEREANU MD, PhD, FESC

University Hospital of Bucharest, Cardiology Department, and
"Carol Davila" University of Medicine and Pharmacy, Romania

ABSTRACT

In this paper we discuss the definition and application of reproducibility in echocardiography. Therefore, reproducibility of the most used echocardiographical parameters is discussed, and thereafter their value is compared with the reproducibility of the same parameters by other imaging methods. In order to improve reproducibility, echocardiographical measurements and interpretations have to be performed by standardized methods, as the ones stated in the current guidelines.

Keywords: reproducibility, echocardiography

1. CLINICAL SIGNIFICANCE

The clinical significance of reproducibility in medicine, and particularly in echocardiography, can be outlined by the following example. The PREAMI trial (1) proved that perindopril 8 mg, given after an acute myocardial infarction with preserved ejection fraction (>40%), significantly reduced (by 22%) the combined endpoint of death, hospitalization for heart failure, and ventricular remodeling at 1 year. In this study the ventricular remodeling process was assessed by 2D echocardiographical estimation of the left ventricular (LV) volumes. Perindopril was shown to reduce the end-diastolic volume of the left ventricle (LVEDV) by 8% at 1 year, a statistically significant result. However, the reproducibility of the LV volume estimation in 2D echo is about 15% for both intraobserver and interobserver evaluation (see below), that is the double of the result in the PREAMI trial. In order to interpret the clinical significance of these "statistically significant" reductions of the LVEDV, the value of the reproducibility of the method is essential to be discussed and taken into account. ❑

2. DEFINITIONS AND ASSESSMENT

Reproducibility is defined as the closeness of agreement between independent results, obtained with the same method, on identical test materials, but under different condition (2). These conditions might be: different human operators, or different time intervals of measurements.

From a practical and also clinical point of view, reproducibility can be divided in three types. In medicine, and particularly in echocardiography, these types are:

1. Inter-observer reproducibility. In this type of reproducibility, each of two blinded echocardiographers will do the same echocardiographic measurements in one patient, under identical conditions. These measurements are then compared.
2. Intra-observer reproducibility. In this type, one experienced investigator will

perform echocardiographical measurements in one patient, and then repeat them several minutes apart, under identical examination conditions. The two measurements are then compared.

3. Test-retest reproducibility. In this type, one experienced investigator will perform echocardiographical measurements in one patient, but several days apart. This type of reproducibility assumes that no important clinical changes have emerged in the time interval between the two examinations. This might not be true, and thus, this type of reproducibility is seldom used in practice. The test-retest reproductibility is also called "repeatability".

In order to measure accurately the reproducibility of a method, a minimum number of subjects should be included. The sample size depends on the number of observers (usually two), and can be calculated using the following formula (3):

$$SS = [\text{number of observers} - 1] \times [\text{number of patients} - 1] > 60$$

Therefore, for two observers, the minimum number of subjects needed for the correct calculation of reproducibility is 61. Unfortunately, this requirement is rather rare fulfilled by the reproducibility studies. Real reproducibility means acquiring and measuring all the data by one investigator during a particular echocardiographic study. However, there are many papers reporting only the variability of the measurements of the already acquired echocardiographical studies on different recording media. This can not be labeled as reproductibility, and should be referred as "measurement variability".

Data analysis of the reproducibility may be assessed by three statistical tests, each with its own advantages and drawbacks, and differing with the number of the observers:

1. The Pearson correlation coefficient (r), which is the most simple to use, but the less accurate;
2. The Bland Altman analysis (when two observers exist);
3. The coefficient of variations (when more than 2 observers exist).

In order to use reproducibility data in clinical practice, three levels of acceptance of the measurements have been defined. A satisfactory limit of reproducibility is defined below 10%. A value of the reproducibility between 10% and 30% may be satisfactory only in some conditions, depending on the magnitude of the use in practice, the cost of the method, and the availability of alternative methods. A variation of the reproducibility level over 30% is unsatisfactory; the method of measurement with such a reproducibility needs corrective actions (Figure 1) (4). ❑

<10%	Satisfactory
10-30%	May be satisfactory
>30%	Unsatisfactory. Needs corrective actions

FIGURE 1. The levels of acceptance of reproducibility. Adapted from reference 4.

3. REPRODUCIBILITY IN ECHOCARDIOGRAPHY

We will discuss briefly the reproducibility of some of the echocardiographic techniques used frequently in routine clinical practice: 2D echocardiography, Doppler parameters, valvular assessment, 3D echocardiography, and tissue Doppler imaging.

3.1. 2D echocardiography

The measurements of LV dimensions by M-mode or 2D echocardiography have an excellent intra- and inter- observer reproducibility (5% and 10%, respectively) (5). The test-retest reproducibility is lower (at 15%), and it may represent the variations that are occurring naturally within the heart (preload, afterload, etc), during different examination conditions between the two measurements (5).

By contrast, LV volume assessment by 2D echocardiography has only moderate intra-, inter, and test-retest reproducibility (15%, 15%, and 18%, respectively) (Figure 2) (6-8). As a consequence, the estimation of LV ejection fraction by 2D echocardiography, which may eliminate bias by dividing the values of the volumes, has a better reproducibility than LV volumes (intraobserver 10%, interobserver 14%) (8-11).

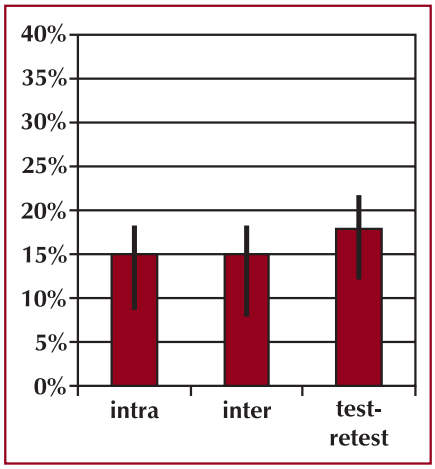

FIGURE 2. The reproducibility of the assessment of left ventricular volumes by 2D echocardiography. Data from references 6-8

It is important to stress meanwhile that 2D echocardiography underestimates true LV volumes (as compared with the gold standard, magnetic resonance imaging or contrast ventriculography). The LV end-diastolic volume is underestimated by 11 to 98 ml, and the LV end-systolic volume is underestimated by 10 to 59 ml. However, because both volumes are underestimated, the LV ejection fraction (which is a ratio of the two) is not much affected. Indeed, LVEF estimated by 2D echocardiography has a good correlation with LVEF assessed by MRI or ventriculography (r = 0.80 – 0.90) (6,8,9,12).

Currently, there are three ways of improving estimation of LV volumes and its reproducibility by echocardiography:

1. Using the harmonic acoustic quantification technique (r increased to 0.91 vs. ventriculography for the estimation of the LV volumes, and the intra- and interobserver reproducibility decreased to less than 10%) (7);
2. Using contrast echocardiography (r increased to 0.87 vs. angiographic computed tomography) (8);
3. Using 3D echocardiography (r increased up to 0.99 vs. ventriculography, with a very high reproducibility, of less than 5%) (13).

Diastolic function of the LV, assessed from the transmitral flow pattern, has good reproducibility when the velocities of the E and A waves and their ratio are measured (intraobserver 10%, interobserver 12%). However, assessment of either the deceleration time of the E wave (TdE) or the pressure half time (PHT) has a poor reproducibility (up to 40%) (14,15).

The assessment of myocardial viability by dobutamine stress echocardiography has moderate reproducibility. For viability, dobutamine stress echocardiography has an intraobserver reproducibility of 8%, and an interobserver reproducibility of 16%. For ischemia, the corresponding values are 15% and 18%, respectively (16).

3.2. VALVULAR ASSESSMENT

Overall, echocardiography has only moderate reproducibility for the assessment of the severity of valvular disease.

Thus, for the assessment of the severity of mitral regurgitation, PISA method has a reproducibility of only 12% (17). Meanwhile, the area of the regurgitant jet method has an overall reproducibility of 15% (18,19). The vena contracta method for assessing mitral regurgitation by color Doppler echocardiography overestimates true mitral regurgitant orifice, it is markedly influenced by flow rate and the ultrasound system that is used. However, a diameter of a vena contracta over 8 mm has good sensitivity and specificity for discriminating severe from non-severe mitral regurgitation (20). The estimation of the diameter of the vena contracta has a reproducibility of 10-15% (21).

For mitral stenosis, both the PHT method of estimation of the mitral valve effective area and the planimetric estimation of the mitral

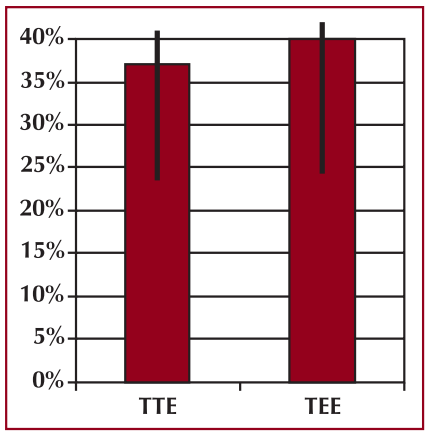

FIGURE 3. Overall reproducibility of the assessment of the severity of aortic stenosis by Doppler ecocardiography. Data from reference 23.

Notes: TTE: transthoracic echocardiography; TEE: transesophageal echocardiography

valve area have only moderate to low reproducibility (intraobserver 20%, interobserver 25%) (22).

The reproducibility of the assessment of the severity of aortic stenosis by Doppler echocardiography is very low, both in transthoracic and transoesophagial echocardiography (28% to 41% for transthoracic echocardiography, 25% to 43% for transoesophagial echocardiography) (Figure 3) (23). This is because the best indicator of the severity of the aortic stenosis is the Doppler flow acceleration trough the valve. In order to obtain the maximal velocity, careful alignment of the ultrasonic Doppler signal with the direction of flow is essential. This is often very difficult in clinical practice.

The reproducibility of the assessment of the severity of aortic regurgitation is better than for aortic stenosis if the severity is assessed by measuring the jet height or the vena contracta of the regurgitant jet. A vena contracta over 6 mm accurately differentiate between severe and non-severe aortic regurgitation. This measurement has a good interobserver and intraobserver reproducibility (10-15%) (24).

The measurement of the systolic pulmonary arterial pressure by the maximal velocity of the tricuspid regurgitant jet has excellent reproducibility (in the range of 5%). Moreover, the correlation of the echocardiographical estimate of the pulmonary arterial systolic pressure almost completely overlaps the invasive measurements ($r = 0.99$) (25).

These values of reproducibility are very important when important decisions are based on individual echocardiographic examinations. This is especially true in cardiac surgery, where the indication for valve replacement is based on the echocardiographical assessment of the severity of the valvular lesion and of the function of the

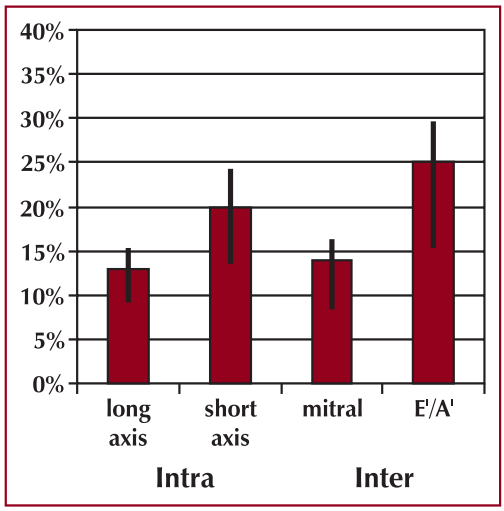

FIGURE 4. The reproducibility of different assessments of myocardial velocities by tissue Doppler echocardiography. Data from references 26,27,29.

left ventricle. Thus, it is important to assess the severity of the valvular lesions and the function of the left ventricle by using only the echocardiographic measurements that have the best reproducibility together with the best sensitivity and specificity for that assessment.

3.3. Tissue Doppler Imaging

Measurement of myocardial velocities by tissue Doppler imaging has only moderate reproducibility. For the LV, the intra- and interobserver reproducibility for the assessment of the longitudinal and short axis systolic functions oscillate between 10% and 20% (26-29). For the right ventricle, the values are about the same. The assessment of the diastolic velocities (E', A', and their ratio) has good to moderate reproducibility (10-28%), depending on the wall segment that was analyzed (Figure 4) (29). Measuring of the diastolic velocities of the posterior basal segments, which are the more difficult to visualize, has the best reproducibility (15%), while the basal anterior segments have the worst reproducibility (30%) (29). ❑

4. COMPARISON BETWEEN ECHOCARDIOGRAPHY AND OTHER IMAGISTING TECHNIQUES

As we have shown before, 2D echocardiography has usually a moderate reproducibility. By comparison, the reproducibility of the other imaging methods (MRI, and angiography) for the assessment of valvular lesions, LV volumes, and LV systolic function is very good, in the range of 5% (6,9,12,13,23). However, these methods are either invasive (angiography) or much more complex (MRI), and they also have limited availability. Thus, echocardiography still serves as the method of first choice for these assessments. ❑

5. CLINICAL IMPLICATIONS

The mathematical formula that is used to calculate the sample size of a study that is needed in order to affirm that the observed differences are not due to the reproducibility of the method, is the following one (3):

$$N = f(a,P) \times SD^2 \times 2/d^2$$

Where:

f(a,P) = 10.5 (for a power P of 90% and an a error of 0.05)
SD = interobserver standard deviation (interobserver reproducibility)
d = difference to be detected

Coming back to the example that was provided at the beginning of this article from the PREAMI study (1), we can now calculate which is the sample size required in order to affirm that the reported difference in the LV volumes are not due to the reproducibility of the measurements. Thus, 2D echocardiography

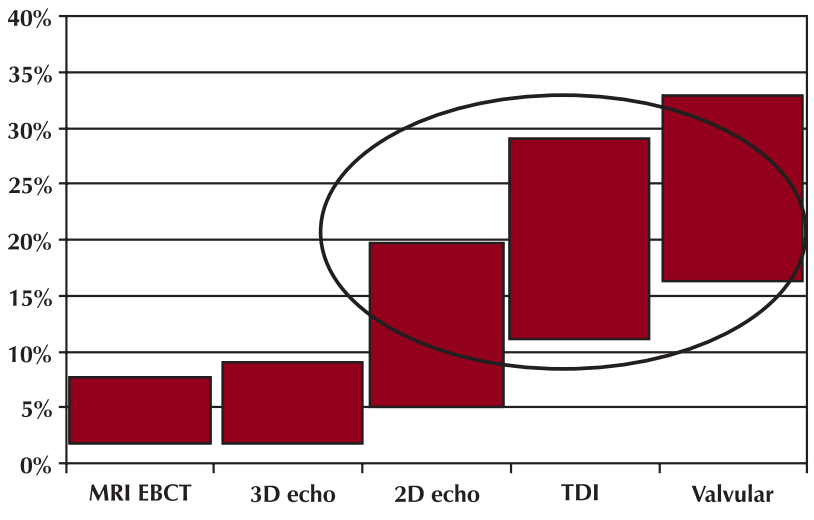

FIGURE 5. Reproducibility in echocardiography. The majority of the echocardiographical techniques have only moderate reproducibility

Notes: MRI, magnetic resonance imaging; EBCT, electron beam computed tomography; TDI, tissue Doppler imaging.

has a reproducibility of around 15% (15% intra-observer, 15% interobserver, and 18% test-retest reproducibility, respectively) (6-8). By introducing in the above formula the reproductibility of the method, at a difference detected of 8% of the LV volumes, the calculated sample size required for the PREAMI study is 74 patients in each subgroup analyzed. This was clearly accomplished, since PREAMI included 1252 patients (randomized 1:1 in each treatment group). Thus, we can conclude that the echocardiographical results of the PREAMI study can be attributed indeed to the treatment itself. ❑

CONCLUSION

Reproducibility has a major impact in clinical studies and in the real life evaluation of echocardiographical parameters.

Overall, the echocardiographical evaluation has only moderate (acceptable) reproducibility (10-20%). Major exceptions from this moderate reproducibility are: 1) on the good side, high reproducibility (<10%) for 3D echo (only when good images are available); 2) on the bad side, low reproducibility (over 30%) for valvular assessment, particularly for aortic stenosis. TDI assessment, particularly for diastolic function, has only moderate reproducibility.

The clinical implication is that any echocardiographical study has to be performed using only the recommendations mentioned in the available guidelines (30). The measurements outlined in these guidelines might not be perfect, but they are the only one used in clinical studies, and they are the only ones validated against gold standards. Any other non-validated method of measurement in echocardiography has a lower reproducibility than the validated equivalents. ❑

REFERENCES

1. **Ferrari R** – Perindopril and Remodeling in Elderly with Acute Myocardial Infarction Investigators. Effects of Angiotensin-Converting Enzyme Inhibition with Perindopril on left Ventricular Remodeling and Clinical Outcome: Results of the Randomized Perindopril and Remodeling in Elderly with Acute Myocardial Infarction (PREAMI) Study. *Arch Intern Med.* 2006; 166:659-666
2. **IUPAC (International Union of Pure and Applied Chemistry)** – Compendium, 1997. Available at www.iupac.org/goldbook/R05305.pdf
3. **Bland M** – An introduction to medical statistics. 3rd ed. Oxford (UK): Oxford University Press, 2000
4. http://www.sixsigmaspc.com/dictionary/RandR-repeatability-reproducibility.html
5. **Dai S, Ayres NA, Harrist RB, Bricker JT, Labarthe DR** – Validity of Echocardiographic Measurement in an Epidemiological Study. Project HeartBeat! *Hypertension.* 1999; 34: 236-241
6. **Grothues F, Smith GC, Moon JC, Bellenger NG, Collins P, Klein HU, Pennell DJ** – Comparison of Interstudy Reproducibility of Cardiovascular Magnetic Resonance with two-dimensional Echocardiography in Normal Subjects and in Patients with Heart Failure or left Ventricular Hypertrophy. *Am J Cardiol.* 2002; 90:29-34
7. **Tsujita-Kuroda Y, Zhang G, Sumita Y, Hirooka K, Hanatani A, Nakatani S, Yasumura Y, Miyatake K, Yamagishi M** – Validity and Reproducibility of Echocardiographic Measurement of left Ventricular Ejection Fraction by Acoustic Quantification with Tissue Harmonic Imaging Technique. *J Am Soc Echocardiogr.* 2000; 13:300-305
8. **Thomson HL, Basmadjian AJ, Rainbird AJ, Razavi M, Avierinos JF, Pellikka PA, Bailey KR, Breen JF, Enriquez-Sarano M** – Contrast Echocardiography Improves the Accuracy and Reproducibility of left Ventricular Remodeling Measurements: a Prospective, Randomly Assigned, Blinded Study. *J Am Coll Cardiol.* 2001; 38:867-875
9. **Malm S, Frigstad S, Sagberg E, Larsson H, Skjaerpe T** – Accurate and Reproducible Measurement of left Ventricular Volume and Ejection Fraction by Contrast Echocardiography: a Comparison with Magnetic Resonance Imaging. *J Am Coll Cardiol.* 2004; 44:1030-1035
10. **Mondelli JA, Di Luzio S, Nagaraj A, Kane BJ, Smulevitz B, Nagaraj AV, Greene R, McPherson DD, Rigolin VH –** The Validation of Volumetric Real-time 3-Dimensional Echocardiography for the Determination of left Ventricular Function. *J Am Soc Echocardiogr.* 2001; 14:994-1000
11. **Gottdiener JS, Livengood SV, Meyer PS, Chase GA** – Should Echocardiography be Performed to Assess Effects of Antihypertensive Therapy? Test-retest Reliability of Echocardiography for Measurement of left Ventricular Mass and Function. *J Am Coll Cardiol.* 1995; 25:424-430
12. **Kuroda T, Seward JB, Rumberger JA, Yanagi H, Tajik AJ** – Left Ventricular Volume and Mass: Comparative Study of Two-dimensional Echocardiography and Ultrafast Computed Tomography. *Echocardiography.* 1994; 11:1-9
13. **Nosir YFM, Fioretti PM, Vletter WB, Boersma E, Salustri A, Postma JT, Reijs AEM, Cate FJT, Roelandt JRTC** – Accurate Measurement of Left Ventricular Ejection Fraction by Three-dimensional Echocardiography. A Comparison With Radionuclide Angiography. *Circulation.* 1996; 94: 460-466
14. **Galderisi M, Benjamin EJ, Evans JC, D'Agostino RB, Fuller DL, Lehman B, Wolf PA, Levy D** – Intra- and Interobserver Reproducibility of Doppler-assessed Indexes of left Ventricular Diastolic Function in a Population-based Study (the Framingham Heart Study). *Am J Cardiol.* 1992; 70:1341-1346
15. **Palmieri V, Arezzi E, Sabatella M, Celentano A** – Interstudy Reproducibility of Parameters of left Ventricular Diastolic Function: a Doppler Echocardiography Study. *J Am Soc Echocardiogr.* 2003; 16:1128-1135
16. **Bjornstad K, al Amri M, Lingamanaicker J, Oqaili I, Hatle L** – Interobserver and intraobserver variation for analysis of left ventricular wall motion at baseline and during low- and high-dose dobutamine stress echocardiography in patients with high prevalence of wall motion abnormalities at rest. *J Am Soc Echocardiogr.* 1996; 9:320-328
17. **Grossmann G, Hoffmeister A, Imhof A, Giesler M, Hombach V, Spiess J** – Reproducibility of the Proximal Flow Convergence Method in Mitral and Tricuspid Regurgitation. *Am Heart J.* 2004; 147:721-728
18. **Castello R, Lenzen P, Aguirre F, Labovitz A** – Variability in the Quantitation of Mitral Regurgitation by Doppler Color Flow Mapping: Comparison of Transthoracic and Transesophageal Studies. *J Am Coll Cardiol.* 1992; 20:433-438
19. **Smith MD, Grayburn PA, Spain MG, DeMaria AN** – Observer Variability in the Quantitation of Doppler Color Flow jet Areas for Mitral and Aortic Regurgitation. *J Am Coll Cardiol.* 1988; 11:579-584
20. **Mascherbauer J, Rosenhek R, Bittner B, Binder J, Simon P, Maurer G, Schima H, Baumgartner H** – Doppler echocardiographic assessment of valvular regurgitation severity by measurement of the vena contracta: an in vitro validation study. *J Am Soc Echocardiogr.* 2005; 18:999-1006
21. **Hall SA, Brickner ME, Willett DL, Irani WN, Afridi I, Grayburn PA** – Assessment of Mitral Regurgitation Severity by Doppler Color Flow Mapping of the Vena Contracta. *Circulation.* 1997; 95:636-642
22. **Chen Q, Nosir YF, Vletter WB, Kint PP, Salustri A, Roelandt JR** – Accurate Assessment of Mitral Valve Area in Patients with Mitral Stenosis by Three-dimensional Echocardiography. *J Am Soc Echocardiogr.* 1997; 10:133-140.
23. **Kupfahl C,Honold M,Meinhardt G,Vogelsberg H,Wagner A, Mahrholdt H, Sechtem U** – Evaluation of Aortic Stenosis by Cardiovascular Magnetic Resonance Imaging: Comparison with Established Routine Clinical Techniques. *Heart.* 2004; 90:893-901
24. **Tribouilloy CM, Enriquez-Sarano M, Bailey KR, Seward JB, Tajik AJ** – Assessment of Severity of Aortic Regurgitation Using the Width of the Vena Contracta. A Clinical Color Doppler Imaging Study. *Circulation.* 2000; 102:558-564
25. **Lanzarini L, Fontana A, Lucca E, Campana C, Klersy C** – Noninvasive Estimation of both Systolic and Diastolic Pulmonary Artery Pressure from Doppler Analysis of Tricuspid Regurgitant Velocity Spectrum in

Patients with Chronic Heart Failure. *Am Heart J.* 2002; 144:1087-1094

26. **Vinereanu D, Khokhar A, Fraser AG** – Reproducibility of Pulsed Wave Tissue Doppler Echocardiography. *J Am Soc Echocardiogr.* 1999; 12:492-499
27. **Nikitin NP, Witte KK, Thackray SD, de Silva R, Clark AL, Cleland JG** – Longitudinal Ventricular Function: Normal Values of Atrioventricular Annular and Myocardial Velocities Measured with Quantitative Two-dimensional Color Doppler Tissue Imaging. *J Am Soc Echocardiogr.* 2003; 16:906-921
28. **Belohlavek M, Tanabe K, Jakrapanichakul D, Breen JF, Seward JB** – Rapid Three-dimensional Echocardiography: Clinically Feasible Alternative for Precise and Accurate Measurement of left Ventricular Volumes. *Circulation.* 2001; 103:2882-2884
29. **Fraser AG, Payne N, Madler CF, Janerot-Sjoberg B, Lind B, Grocott-Mason RM, Ionescu AA, Florescu N, Wilkenshoff U, Lancellotti P, Wutte M, Brodin LA** – MYDISE Investigators. Feasibility and Reproducibility of off-line Tissue Doppler Measurement of Regional Myocardial Function during Dobutamine Stress Echocardiography. *Eur J Echocardiogr.* 2003; 4:43-53
30. **Gottdiener JS, Bednarz J, Devereux R, Gardin J, Klein A, Manning WJ, Morehead A, Kitzman D, Oh J, MD, Quinones M, Schiller NB, Stein JH, Weissman NJ** – American Society of Echocardiography Recommendations for Use of Echocardiography in Clinical Trials A Report from the American Society of Echocardiography's Guidelines and Standards Committee and The Task Force on Echocardiography in Clinical Trials. *J Am Soc Echocardiogr.* 2004; 17:1086-1119

Address for correspondence:
Andrei-Dumitru Margulescu, Emergency University Hospital, Splaiul Independentei Blvd., No. 169, SE 5, Bucharest, Romania
email address: andrei_marg@yahoo.com