Analysis of variability and reproducibility of echocardiography measurements in valvular aortic valve stenosis [3]

LUÍS M. MOURA, SANDRA F. RAMOS, FAUSTO J. PINTO, ISABEL M. BARROS, F. ROCHA-GONÇALVES

1Oporto University School of Medicine, Oporto, Portugal, 2Pedro Hispano Hospital, Matosinhos, Portugal, 3Lisbon University School of Medicine, Lisbon, Portugal

ABSTRACT

Background: Doppler echocardiography is the most frequent method for detecting and evaluating the severity of valvular aortic stenosis. The aim of this study was to assess the variability and reproducibility of echocardiographic parameters including aortic valve area (AVA), peak aortic jet velocity (Vmax), velocity ratio (V1307/Vmax), peak gradient (Gmax) and mean gradient (Gmean) in aortic stenosis (AS) patients.

Methods: Doppler echocardiograms were obtained from 150 randomly selected patients (56.7% male; mean age 73±9 years) with asymptomatic moderate aortic valve stenosis. The echocardiographic measurements were performed by two independent level III (expert) blinded observers. To assess intra-observer variability, we evaluated parameters of AS progression at two different times (mean of two weeks after the first examination).

Results: For intra-observer variability (observer 1), the variation and reproducibility coefficients were, respectively, 1.88% and 0.16 m/s for Vmax, 2.08% and 0.14 for V1307/Vmax, 2.05% and 0.18 cm² for AVA,

Variabilidade e reprodutibilidade da análise das medições ecocardiográficas na estenose valvular aórtica

RESUMO

Introdução: A ecocardiografia Doppler é o método mais frequente de detecção e avaliação da gravidade da estenose valvular aórtica. O objectivo deste estudo foi avaliar a variabilidade e reprodutibilidade dos parâmetros ecocardiográficos como a área valvular aórtica (AVA), velocidade pico (Vmax), relação de velocidades (V1307/Vmax), gradiente pico (Gmax) e gradiente médio (Gmédio) nos doentes com estenose aórtica.

Métodos: Um ecocardiograma Doppler foi realizado em 60 doentes consecutivos selecionados aleatoriamente (da população do estudo RAAVE) com estenose aórtica moderada a grave assintomática (56.7% sexo masculino; idade média 73 ± 9 anos). As medidas ecocardiográficas foram efectuadas por dois ecocardiografistas de nível III numa estratégia de dupla ocultação. Para o estudo da variabilidade intra-observador, avaliamos os parâmetros de progressão da estenose aórtica em dois
3.89% and 5.18 mmHg for $G_{\text{max}}$ and 7.87% and 6.30 mmHg for $G_{\text{mean}}$. For inter-observer variability, the variation and reproducibility coefficients were, respectively, 2.00% and 0.14 m/s for $V_{\text{max}}$, 2.91% and 0.14 for $V_{\text{LVOT}}/V_{\text{max}}$, 7.67% and 0.16 cm² for AVA, 8.53% and 7.06 mmHg for $G_{\text{mean}}$ and 3.90% and 5.58 mmHg for $G_{\text{max}}$. Both intra- and inter-observer studies showed excellent intraclass correlation coefficients (ICC) for all echocardiographic parameters (ICC ranged from 0.943 to 0.990 for intra-observer variability and from 0.955 to 0.992 for inter-observer variability).

**Conclusion:** Doppler echocardiographic measurements of AVA, $V_{\text{max}}$, $G_{\text{max}}$ and $G_{\text{mean}}$ are highly reproducible when performed by expert observers. Of all echocardiographic parameters, $V_{\text{max}}$ and $V_{\text{LVOT}}/V_{\text{max}}$ showed the best variability and reproducibility, and thus constitute reliable tools for clinical and research purposes in aortic stenosis diagnosis and follow-up.

**Key words**
Aortic stenosis; Echocardiography; Variability and reproducibility; Bland-Altman; RAAVE; Follow-up

Resultados: No que respeita à variabilidade intra-observador (observador 1), os coeficientes de variação e reprodutibilidade foram, respectivamente, 1.88% e 0.16 m.s⁻¹ para $V_{\text{max}}$, 2.08% e 0.14 cm² para a relação $V_{\text{LVOT}}/V_{\text{max}}$, 2.05% e 0.18 m.s⁻¹ para a AVA, 3.89% e 5.18 mmHg para $G_{\text{max}}$ and 7.87% e 6.30 mmHg para o $G_{\text{média}}$. No que respeita à variabilidade inter-observador, os coeficientes de variação e reprodutibilidade foram, respectivamente, 2.00% e 0.14 m.s⁻¹ para $V_{\text{max}}$, 2.91% e 0.14 m.s⁻¹ para a relação $V_{\text{LVOT}}/V_{\text{max}}$, 7.67% e 0.16 cm² para a AVA, 8.53% e 7.06 mmHg para a $G_{\text{média}}$ e 3.90% e 5.58 mmHg para o $G_{\text{max}}$. Os estudos intra-observador e inter-observador mostraram ter excelentes coeficientes de correlação intra-classe (CCI), para todos os parâmetros ecocardiográficos (CCI varia de 0.943 até 0.990 para a variabilidade intra-observador e de 0.955 até 0.992 para a variabilidade inter-observador).

**Conclusões:** As medições ecocardiográficas da AVA, $V_{\text{max}}$, $G_{\text{max}}$ e $G_{\text{média}}$ são altamente reprodutíveis quando realizadas por ecocardiografistas experientes. De todos os parâmetros ecocardiográficos a $V_{\text{max}}$ e a relação $V_{\text{LVOT}}/V_{\text{max}}$ apresentam os melhores valores de variabilidade e reprodutibilidade e assim constituem provavelmente a melhor ferramenta no diagnóstico e seguimento criterioso dos doentes com estenose valvular aórtica.

**Palavras Chave:**
Estenose aórtica; Ecocardiografia; Variabilidade e reprodutibilidade; Bland-Altman; Estudo RAAVE; Seguimento e prognóstico
INTRODUCTION

Calcific aortic stenosis (AS) is common and hemodynamically significant valvular stenosis affects approximately 5% of people aged 75 to 86 years\(^1\). Severe symptomatic aortic stenosis has a poor prognosis and the only definitive treatment is aortic valve replacement\(^2\). The rate of disease progression is not well known and patients with mild to moderate aortic stenosis require regular clinical and echocardiographic surveillance\(^3,4\).

Doppler echocardiography has become the accepted mode of surveillance for patients with aortic stenosis, avoiding the need for repeated cardiac catheterization. Results have been shown to correlate well with invasive hemodynamic data\(^5-7\).

Calculation of aortic valve area (AVA) by the continuity equation requires precise measurements of the left ventricular (LV) outflow tract (LVOT) diameter, and peak velocities of blood flow in the LVOT and across the aortic valve\(^8,9\).

Failure to acquire satisfactory images of the LVOT can be a major limitation in a significant proportion of patients because of body habitus, sigmoid interventricular septum, or aortic valve calcification\(^10,11\).

The continuity equation further amplifies error by using the square of the LVOT diameter to calculate cross-sectional area, contributing to significant intra- and inter-observer variability.

Reproducibility studies are designed to measure the level of concordance between observations made under the same circumstances by the same observer (intra-observer concordance) or by different observers (inter-observer concordance). Intra-observer variability is predominantly random, while inter-observer variability can be random or systematic.

A measurement is said to be reproducible if it comes from a reproducible procedure, that is, the same measurements will be obtained if it is repeated under the same conditions. Lack of reproducibility in measurements can lead to major scientific, clinical and medical-legal consequences. Sometimes, the lack of reproducibility obtained from several measurements can only be improved by basing decisions on multiple, independent, arbitrated or consensual opinions.

In a biomedical study the only source of variability in parameters should be the intrinsic biological variability of the subjects under study. However, there is also very often variability which depends on measurements made by the observer or by the tool used for the measurements.

Accurate measurements are sometimes needed in situations in which they are difficult to obtain in clinical practice, with wide variability in observations that can limit their validity.

Previous studies on reproducibility analyzing this problem have mainly evaluated the correlation more than the concordance of measurements, which is the main reason we have used Bland-Altman analysis.

The aim of this study was to assess the parameters of reproducibility of echocardiographic measurements in aortic stenosis patients.

METHODS

Patients and echocardiographic measurements

One hundred and fifty patients with asymptomatic moderate calcific aortic stenosis and AVA ≥1.0 cm\(^2\) (56.7% male; mean age 73±9 years) were included in this Rosuvastatin Affecting Aortic Valve Endothelium to Slow the Progression of Aortic Stenosis (RAAVE) sub-study. We decided to include moderate AS because within this level of severity there is a wide range of variability.

The local research ethics committee approved the study and written informed consent was obtained from all participants. The patients were randomly selected from the total study population of the RAAVE study.

Comprehensive transthoracic echocardiograms were performed in a single echocardiographic laboratory. Immediate physician
review (level III) enabled re-imaging for quality control. Standard Doppler measurements of the LVOT and aortic valve were recorded from multiple windows to obtain the maximum velocity, and the mean gradient, peak velocity, and aortic valve area were measured and calculated as defined by the American Heart Association/American College of Cardiology guidelines for the clinical application of echocardiography.[12, 13]

To assess intra-observer variability, we evaluated parameters of aortic stenosis progression (peak velocity, aortic valve area, peak gradient, mean gradient and \( V_{LVOT}/V_{max} \) ratio) at two different times (two weeks after the first examination).

For inter-observer variability, off-line measurements were made from the records by a second observer who was unaware of the results of the first studies.

The echocardiograms were performed by two echocardiographers with over five years of experience (IB and LM).

The images were recorded digitally and analyzed off-line. A total of five measurements of each parameter were made for patients in sinus rhythm and seven for patients with atrial fibrillation, and the mean of these measurements was used.

### Statistical analysis

Data are presented as mean ± standard deviation. The variability of the measurements was evaluated based on the variability coefficient. Reproducibility was assessed by the interclass correlation coefficient (ICC).

In addition, the Bland-Altman method[14] (the mean of the differences ± twice the standard deviation of the differences) was used to calculate the reproducibility coefficient. This method is an excellent tool for examination of models with discordance between measurements, as its graphical nature makes it easier to interpret.

The Student’s t test for paired samples was used to verify whether the mean of the differences between the measurements (different observers or different times) was statistically equal to zero. Analyses were performed using SPSS software, version 14.0 (SPSS Inc., Chicago, Illinois). A two-tailed \( p \) value less than 0.05 was considered to indicate statistical significance.

### RESULTS

#### Intra-observer variability and reproducibility

Comparisons of variables for both observer 1 and observer 2 are shown as coefficients of variation (Table I), as reproducibility coefficients and as ICCs (Table II), and as a Bland-Altman plot in Figure 1.

The variation coefficients (Table I) were proportionately better, corresponding to a small variation, for \( V_{max} \) (observer 1 = 1.88%; observer 2 = 5.86%), \( V_{LVOT}/V_{max} \) (observer 1 = 2.08%; observer 2 = 2.15%) and for \( G_{max} \) (observer 1 = 3.89%; observer 2 = 4.12%) than for A VA (observer 1 = 2.05%; observer 2 = 7.89%) and \( G_{mean} \) (observer 1 = 7.87%; observer 2 = 6.67%).

The intra-observer reproducibility coefficients (Table II) were 0.16 m/s for \( V_{max} \), 0.14 for \( V_{LVOT}/V_{max} \) and 0.18 cm² for A VA. For \( G_{max} \) and \( G_{mean} \), reproducibility coefficients were 5.18 mmHg and 6.30 mmHg, respectively. When the null hypothesis was verified the mean difference of measurements (obtained at two different times) was equal to zero, and this cannot be rejected for the variables analyzed (\( p \geq 0.251 \)).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient of variation</th>
<th>Observer 1</th>
<th>Observer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_{max} ) (m/s)</td>
<td>1.88%</td>
<td>5.86%</td>
<td></td>
</tr>
<tr>
<td>A VA (cm²)</td>
<td>2.05%</td>
<td>7.89%</td>
<td></td>
</tr>
<tr>
<td>( G_{max} ) (mmHg)</td>
<td>3.89%</td>
<td>4.12%</td>
<td></td>
</tr>
<tr>
<td>( G_{mean} ) (mmHg)</td>
<td>7.87%</td>
<td>6.67%</td>
<td></td>
</tr>
<tr>
<td>( V_{LVOT}/V_{max} )</td>
<td>2.08%</td>
<td>2.15%</td>
<td></td>
</tr>
</tbody>
</table>

Table I. Intra-observer variability of echocardiography measurements of AS
Table II. Intra-observer reproducibility of echocardiographic measurements of AS

<table>
<thead>
<tr>
<th></th>
<th>Mean ± SD</th>
<th>Mean ± SD of the differences</th>
<th>RC</th>
<th>ICC</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{max}$ (m/s)</td>
<td>3.51±0.55</td>
<td>0.01±0.08</td>
<td>0.16</td>
<td>0.989</td>
<td>0.479</td>
</tr>
<tr>
<td>AVA (cm²)</td>
<td>1.01±0.32</td>
<td>0.00±0.09</td>
<td>0.18</td>
<td>0.969</td>
<td>0.966</td>
</tr>
<tr>
<td>$G_{max}$ (mmHg)</td>
<td>50.56±16.50</td>
<td>0.29±2.59</td>
<td>5.18</td>
<td>0.957</td>
<td>0.550</td>
</tr>
<tr>
<td>$G_{mean}$ (mmHg)</td>
<td>33.44±11.47</td>
<td>0.74±3.15</td>
<td>6.30</td>
<td>0.943</td>
<td>0.251</td>
</tr>
<tr>
<td>$V_{LVOT}/V_{max}$</td>
<td>0.31±0.05</td>
<td>0.01±0.07</td>
<td>0.14</td>
<td>0.990</td>
<td>0.729</td>
</tr>
</tbody>
</table>

* The hypotheses of the test are H0: $d_{mean}$ (mean difference) = 0 vs. H1: $d_{mean} \neq 0$. RC: reproducibility coefficient; twice the standard deviation of the differences, ICC: intraclass correlation coefficient.

Figure 1. Intra-observer reproducibility (A-E) of echocardiographic measurements of aortic valve stenosis. The central horizontal line corresponds to the mean of the differences of the two measurements and the two exterior lines correspond to 2 x SD of the differences.

Table III. Inter-observer variability of echocardiographic measurements of AS

<table>
<thead>
<tr>
<th></th>
<th>Coefficient of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{max}$ (m/s)</td>
<td>2.00%</td>
</tr>
<tr>
<td>AVA (cm²)</td>
<td>7.67%</td>
</tr>
<tr>
<td>$G_{max}$ (mmHg)</td>
<td>3.90%</td>
</tr>
<tr>
<td>$G_{mean}$ (mmHg)</td>
<td>8.35%</td>
</tr>
<tr>
<td>$V_{LVOT}/V_{max}$</td>
<td>2.91%</td>
</tr>
</tbody>
</table>

* The hypotheses of the test are H0: $d_{mean}$ (mean difference) = 0 vs. H1: $d_{mean} \neq 0$. RC: reproducibility coefficient; twice the standard deviation of the differences, ICC: intraclass correlation coefficient.
Whichever variables were analyzed, excellent intraclass correlation coefficients were obtained (Table II), reflecting excellent accuracy in measurements of all echocardiographic variables.

The distribution of the points in Figure 1 (A-E) shows that there were no systematic errors in the readings made by the observer.

Among the echocardiographic variables studied, \( V_{\text{max}} \) and \( V_{\text{LVOT}}/V_{\text{max}} \) provide better performance for the evaluation of aortic stenosis (lower variation coefficient and higher ICC).

### Inter-observer variability and reproducibility

Comparisons between observers 1 and 2 are presented as coefficients of variation (Table III), as reproducibility coefficients and as ICCs (Table IV), and as a Bland-Altman plot in Figure 2 (A-E).

In analysis of inter-observer variability, Table III shows that coefficients of variation were better for \( V_{\text{max}} \) (2.00%) and \( V_{\text{LVOT}}/V_{\text{max}} \) (2.91%) than for \( G_{\text{max}} \) (3.90%), \( A_{\text{VA}} \) (7.67%) and \( G_{\text{mean}} \) (8.53%).

As can be seen in Table IV, the coefficients of inter-observer reproducibility were 0.14 m/s for \( V_{\text{max}} \), 0.14 for \( V_{\text{LVOT}}/V_{\text{max}} \) and 0.16 cm\(^2\) for \( A_{\text{VA}} \). For \( G_{\text{max}} \) and \( G_{\text{mean}} \), the reproducibility coefficients were 5.58 mmHg and 7.06 mmHg, respectively. When the hypothesis was proven, the mean differences in the measurements between the two observers were equal to zero, so this hypothesis cannot be rejected for the variables analyzed (\( p \geq 0.495 \)).

The distribution of the points in Figure 2 (A-E) suggests that there is no divergence in the observations produced by the two observers. In addition, excellent intraclass correlation coefficients were recorded (Table IV) (\( r=0.992 \) for \( V_{\text{max}} \), \( r=0.981 \) for \( V_{\text{LVOT}}/V_{\text{max}} \), \( r=0.967 \) for \( A_{\text{VA}} \), \( r=0.961 \) for \( G_{\text{max}} \) and \( r=0.995 \) for \( G_{\text{mean}} \)).

We obtained a correlation coefficient \( r \) for \( A_{\text{VA}} \) for 2D echocardiography vs. catheterization of 0.71 in 12 patients who were referred for aortic valve replacement surgery.

### DISCUSSION

What is in fact the clinical utility of these data in an era when multimodality imaging is becoming more and more relevant and echo has been shown to be a reproducible imaging technique? What does it add to the body of current knowledge that will make clinicians rely even further on it? Will these results in any way alter clinical practice?

Additionally, we believe there is some confusion among scientists and statisticians about the right test to assess inter- and intra-observer variability. What does this really add over and above analysis of variance, standard deviation and simple averages?

To start to answer to these questions we already know that assessment of aortic stenosis requires evaluation of patient symptoms and quantification of the severity of valve obstruction. Methods to assess severity include cardiac catheterization, echocardiography, intravascular ultrasound, computed tomography, and magnetic resonance imaging. The ideal method should be simple, noninvasive, and widely available while remaining both accurate and reproducible; echocardiography would seem to best fulfill these requirements.
High standard deviations were observed in the hemodynamic progression of AS, highlighting the considerable individual variability in disease progression.

Calculation of aortic valve area according to the continuity equation is a reliable and reproducible method for evaluating the severity of aortic stenosis. However, it requires accurate measurements, because a calculation error could be amplified proportionally to its square (for example, in determination of the diameter of the LVOT), which contributes to greater intra- and inter-observer variability, thus limiting its validity in some cases.

Other works studying these problems mainly evaluated correlation rather than concordance of data\(^{16}\). We calculated the variability of observations based on the variation coefficient and difference between two measurements as the percentage of the mean\(^{10,15-18}\).
However, as these coefficients do not show potential systematic errors, we calculated the reproducibility coefficient, which does not vary with the value of the mean. Prior findings on inter-observer variability for AVA indicate that this can reach up to 24%, although studies assessing reproducibility that use regression analysis should be interpreted with caution (18).

In this study, peak aortic jet velocity and the \( \frac{V_{\text{LVOT}}}{V_{\text{max}}} \) ratio (interrelated by the Bernoulli equation) demonstrated lower values of intra-and inter-observer variation. Similarly, these two parameters already had the best correlation coefficients. This shows that an increase in peak transvalvular velocity on AS follow-up constitutes the earliest sign of disease progression and is the best parameter revealing AS severity because it evaluates the vena contracta (anatomical area).

The coefficients of variation of AVA are similar to those in the literature, which confirms the clinical limitations of this parameter in monitoring AS progression over time (19). Recall we obtained a correlation coefficient (r) for AVA for 2D echocardiography vs. catheterization of 0.71. New methods such as real-time 3D echocardiography (RT3D) will probably be more accurate than 2D echocardiography and two-dimensional volumetric methods to calculate the area and to grade the severity of AS (20). In the near future evaluating AVA by RT3D will probably demonstrate the best agreement among all techniques. Planimetry of the aortic valve by transesophageal echocardiography shows good correlation with the Gorlin equation and with the continuity equation by RT3D, but this parameter is by itself rarely sufficient for making clinical decisions.

As in other published papers, our study shows through serial measurements that peak transvalvular velocity and aortic valve area are the parameters with least variability and best reproducibility for the evaluation and follow-up of patients with AS.

Using the same rationale we additionally propose that the \( \frac{V_{\text{LVOT}}}{V_{\text{max}}} \) ratio, i.e. peak transvalvular velocity (CW) divided by LLVOT peak velocity (PW), is a better parameter to evaluate AS progression. This ratio is another approach to reducing error related to LVOT diameter measurements by removing cross-sectional area (CSA) from the simplified continuity equation. This dimensionless velocity ratio expresses the size of the effective valve area as a proportion of the CSA of the LVOT. A normal velocity ratio is slightly less than 1, with smaller ratios indicating more severe aortic stenosis. For example, a velocity of 0.25 (21) means that the valve opening is reduced to one-fourth (25%) its normal size. The velocity ratio has the advantage of being already “indexed” to body size. Normal intra-cardiac velocities are similar in people of all ages and sizes. By looking at the velocities alone, the velocity ratio assumes that the proximal cross-sectional area is “normal” for the patient and thus the resulting descriptor of stenosis severity is indexed for body size (22).

Therefore, we suggest their routine use in quantification of AS, especially when technical doubts arise (deficient acoustic windows) or when other measurements are close to the limits of different degrees of severity.

The major clinical implications and utility of this study are that it has demonstrated, using appropriate methodology (Bland-Altman plots), that velocity ratio is the most useful parameter to follow aortic stenosis patients in the long term. We came to this conclusion on the basis of findings in previous studies on this topic and after correlation with angiographic data.


12. Lang RM, Bierig M, Devereux RB, Flachskampf FA, Foster E, Pellikka PA, et al. Recommendations for chamber quantification: a report from the American Society of Echocardiography’s Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. J Am Soc Echocardiogr. 2005 Dec;18(12):1440-63.


