MUGA processing: intra and interoperator variability impact using manual and automated methods

Rita Belo¹, Cláudia Alves¹, Cristiana Carvalhal¹, Sérgio Figueiredo²³, Elisabete Carolino²⁴, Lina Vieira²³⁵

¹. Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa. Lisboa, Portugal. ritabelo96@gmail.com
⁵. CIMOSM, ISEL – Centro de Investigação em Modelação e Optimização de Sistemas Multifuncionais. Lisboa, Portugal.

ABSTRACT: Introduction – Multigated acquisition (MUGA) scan is mainly used for the assessment of left ventricular ejection fraction (LVEF) in patients who undergo cardiotoxic chemotherapy drugs. When applying automatic (A) or manual (M) processing methods, some biases in the quantitative metrics can be obtained. The aim of this study is to evaluate the influence of A and M methods, specifically, the inter and intraoperative variability in accordance with the professional experience. Methods – A retrospective study was performed with 14 MUGA exams available in ESTeSL's Xeleris™ Functional Imaging Workstation v. 1.0628 database. Three operators (OP) with no professional experience and two with more than 10 years of experience, processed every study five times for each method, using the EF Analysis™ and the Peak Filling Rate™. To perform the multiple comparisons, the Repeated Measures ANOVA, Friedman, t-test and Wilcoxon tests were used, considering $\alpha$=0.05. Results – Four of the OP presented statistically significant differences between methods in one or more parameters; similar values between experienced OP and between the non-experienced were observed when the A method was applied, and higher discrepancies were present for all parameters obtained by the M mode; higher LVEF, peak filling rate, and peak emptying rate values were observed for the M method. Conclusion – Variability was found when comparing M and A processing methods, as well as interoperator variability associated with their level of experience. Despite that, there was a trend of less variability between the two experienced OP and in the A method.

Keywords: Equilibrium radionuclide angiography; Cardiac function; Segmentation; Left ventricular ejection fraction; Diastolic parameters

Processamento de estudos de angiografia de radionuclídeos em equilíbrio: impacto da variabilidade intra e interoperator por métodos manuais e automáticos

RESUMO: Introdução – A angiografia de radionuclídeos em equilíbrio (ARNe) é principalmente realizada para determinar a fração de ejeção do ventrículo esquerdo (FEVE) em doentes submetidos a quimioterápicos cardiotoxícos. Quando aplicados métodos de processamento automáticos (A) ou manuais (M) podem ser obtidas distorções métricas. Este estudo teve como objetivo aferir a influência dos métodos A e M e avaliar a variabilidade inter e intraoperator associada a diferentes experiências profissionais. Métodos – Estudo retrospetivo com 14 exames ARNe existentes na base de dados da estação de processamento Xeleris™ Functional Imaging Workstation v. 1.0628 da ESTeSL. Três operadores (OP) sem experiência profissional e dois com mais de dez anos de experiência processaram cada estudo cinco vezes por cada método, recorrendo ao EF Analysis™ e ao Peak Filling Rate™. As múltiplas comparações foram realizadas com os testes
ANOVAs of repeated measures, Friedman, teste-t and Wilcoxon, considering α=0.05. **Resultados** – Quatro dos OP apresentaram diferenças estatisticamente significativas entre métodos para um ou mais parâmetros; foram obtidos valores semelhantes entre os OP experientes e entre os não experientes quando se aplicou o método A e observaram-se maiores discrepâncias para todos os parâmetros obtidos pelo método M; obtiveram-se valores superiores de FEVE, taxas de esvazamento e preenchimento máximos com o método M. **Conclusão** – Verificou-se variabilidade dos resultados obtidos a partir da comparação dos métodos de processamento M e A, bem como variabilidade do interoperator associada ao seu nível de experiência profissional. Contudo os dois OP experientes apresentaram menor discrepância de valores entre si e para o método A.

**Palavras-chave**: Angiografia de radionuclídeos em equilíbrio; Função cardíaca; Segmentação; Fração de ejeção do ventrículo esquerdo; Parâmetros diastólicos

**Introduction**

Multigated acquisition (MUGA), also known as equilibrium radionuclide angiography, is a nuclear medicine procedure that uses $^{99m}$Technetium labelled erythrocytes to acquire cardiac images in synchronism with an electrocardiogram. It is a well-established and noninvasive method used to assess left ventricular (LV) ejection fraction (EF) and is mainly used for serial assessment of LVEF in patients who undergo cardiotoxic chemotherapy drugs since the LV dysfunction is the most common manifestation of cardiotoxicity.

The cardiac alterations identified as late injuries and since the deterioration of the diastolic function precedes the systolic one, it is important to have tools available to detect early cardiac damages, which can be achievable by measuring the peak filling rate (PFR) and peak emptying rate (PER) using the same acquisition used to determine LVEF.

In order to obtain these quantitative metrics, image segmentation is an important processing step for the detection of the LV cavity to determine LVEF and consequently other physiological parameters such as PFR and PER. Specifically, it is essential to apply different regions of interest (ROI), based on automatic (A) and/or manual (M) segmentation.

To calculate the EF, three ROI need to be drawn, such as a periventricular one to cover only background (BKG) structures and two around the LV at the end of the diastole (ED) and at the end of systole (ES). Therefore, EF can be determined using the equation:

$$\%\text{EF} = \left( \frac{\text{BKG corrected counts} - \text{BKG corrected ES counts}}{\text{BKG corrected ED counts}} \right) \times 100$$

**Equation 1. %LVEF calculation.**

Generally, a LVEF value greater than 55% is considered normal and a drop greater than 10% is consistent with early cardiotoxicity.

In the clinical field, there are different commercial software applications, with both A and M processing approaches available. The EF Analysis” program is one of them and it’s used for the quantification of the LV function. The A and M processing methods differ in the way that ventricular ROIs are obtained, such that they can be drawn manually by the operator (OP) or automatically defined by the program after the OP centers an elliptical ROI in the LV. In the A mode, an edge detection algorithm is used to determine the ED and ES ROI. ED and ES ROI are drawn in frames with the highest and lowest counts in the LV, respectively. The background ROI is automatically created in the segment with the lowest average count rate, regardless of the processing mode chosen. The LVEF is calculated and a time-activity curve is created.

The Peak Filling Rate application is used as a diagnosis and prognosis tool for early detection of deterioration of LV diastolic function, which is an early manifestation of developing coronary disease that if not treated or if the cardiotoxic drugs are not discontinued, it can evolve to systolic dysfunction. This tool uses the resulting series created by the EF Analysis as input and smooths the LV time-activity curve in order to create a derivative curve. The commonly used diastolic parameters obtained from this curve are the PFR and the PER. The PFR represents the maximum rate of filling and the PER is obtained from the systolic phase of the time-activity curve and determines the maximum rate of LV emptying. These parameters are more sensitive than the EF index, which only decreases when the LV function has adulterated quite significantly, which justifies their importance.

All these quantitative parameters obtained by MUGA processing applications require an accurate segmentation of the LV. However, when applied for LVEF estimation, some inter and intraoperator biases were obtained as well as between centers and commercial applications systems. These facts create certain inconsistencies in some of the quantitative metrics obtained, which need experimental validation. This is especially important regarding the diastolic parameters like PFR and PER, the first ones to change and remit to early deterioration of ventricular function.

The main goal of this study was to evaluate the influence of A and M processing methods on MUGA studies. Specifically, we intend to evaluate the influence of inter and intraoperative variability in the determination of LVEF, PFR and PER parameters obtained by the A and M methods, in accordance with the professional experience.

**Methods**

A retrospective study was performed considering MUGA studies integrated into the database of the Xeleris Functional
Imaging Workstation v. 1.0628 at ESTeSL. A non-probabilistic sample of 14 patients with a clinical indication for MUGA was used. The patients with a valid left anterior oblique MUGA dataset for processing were included, all the others were excluded.

In order to calculate the LVEF, data were processed using the EF Analysis™ software and the Peak Filling Rate™ was used to obtain the PFR and PER values. Five OP were selected to process the MUGA exams and categorized by their degree of professional experience, respectively: OP1 and OP2 with more than 10 years of experience, and OP3, OP4 and OP5 with basic knowledge of nuclear medicine and without clinical experience. Each OP processed each study five times per method, as exemplified for an experienced OP in Figure 1. The data was processed non-consecutively, and it was guaranteed that the OPs did not know the previously obtained values, as well as the information related to the patient.

The LVEF, PFR and PER values were analyzed using the International Business Machine Statistical Package for the Social Sciences™ v. 23.0 for macOS. The results were considered significant at the significance level of 5%. The Shapiro-Wilk test was performed to test the normality of the data. When comparing the multiple parameters between OP, the Repeated Measures ANOVA test was used when the normality assumption was verified ($p \geq 0.05$) and when not verified ($p < 0.05$), the Friedman’s test was performed. When comparing the two methods for the same OP, the t-test was used for two paired samples when the normality assumption was verified ($p \geq 0.05$) and when not verified ($p < 0.05$), the Wilcoxon test was used.

Results

When comparing the obtained LVEF values with the A method, statistically significant differences were detected between OP [Greenhouse-Geisser statistic (2.804)=5.897, $p=0.003$]. Of the paired multiple comparisons, the differences were between the OP1 and the OP 3 ($p=0.013$), 4 ($p=0.020$) and 5 ($p=0.006$). Analogously, there were differences between the OP2 and the OP 3 ($p=0.024$), 4 ($p=0.018$) and 5 ($p=0.004$). In this case, not only the experienced OP obtained values were close to each other but they both obtained LVEF values significantly different than the ones obtained by the non-experienced OP. The experienced OP also obtained higher LVEF values as seen in Figure 2. Regarding the LVEF values obtained with the M method, there were statistically significant differences identified between OP [Greenhouse-Geisser statistic (2.460)=12.02, $p=0.000$]. Of the paired multiple comparisons, the differences found were between the OP1 and the observers OP 2 ($p=0.001$), 3 ($p=0.048$) and 4 ($p=0.000$); between the OP2 and the OP 4 ($p=0.018$) and 5 ($p=0.033$); between OP3 and OP4 ($p=0.005$); and between OP4 and OP5 ($p=0.001$). It is verified that there is a discrepancy between the entire OP in the values of the LVEF, with higher dispersion in LVEF values for this method (Figure 2).

Comparing both methods, we found statistically significant differences in the LVEF value of OP1 ($t(13)=-5.538$, $p=0.000$), OP3 ($t(13)=-3.12$, $p=0.008$) and OP5 ($t(13)=-6.505$, $p=0.000$), with higher values of LVEF obtained with the M method for all 5 OP.

Concerning the PFR values obtained by the A processing, statistically significant differences were detected between OP [X$^2(4)=16.16$]. From Friedman’s multiple comparisons, the differences obtained were between OP1 and the OP 4 ($p=0.041$) and 5 ($p=0.04$); and between OP4 and OP5 ($p=0.028$). It is verified that there are differences between experienced and non-experienced OP, with proximal PFR values between each group. Experienced OP with higher PFR values and with proximal values between non-experienced OP (Figure 3). For the PFR values obtained with the M processing, statistically significant differences were also detected between OP [$f^2(4)=29.943$]. From Friedman’s multiple comparisons the differences obtained were between OP4 and OP3 ($p=0.034$); and between OP2 ($p=0.023$) and OP1 ($p=0.001$). There are variations in the PFR values between all OP regardless of the processing method used; however, the M one presented higher discrepancies (Figure 3).
Concerning the PFR value, differences between methods were found in OP1 (z = 3.107, p = 0.008), OP3 (z = 2.103, p = 0.008) and OP5 (z = 3.171, p = 0.013). Values obtained by M processing were also higher than the A ones.

For the A PER values, statistically significant differences were detected between OP [F(4) = 13.200]. From Friedman’s multiple comparisons the differences obtained were between OP5 and the two experienced OP (p = 0.008, p = 0.005); and between the OP4 and the two experienced ones (p = 0.028, p = 0.019). As we can verify, there is a difference between experienced and non-experienced OP, besides, experienced OP obtained closer values and the non-experienced obtained values closer to each other. For the M PER values, statistically significant differences were detected between OP [F(4) = 29.943]. From the Friedman multiple comparisons, the differences obtained were between OP4 and OP5 (p = 0.028) and OP1 (p = 0.000); and between OP3 and OP1 (p = 0.008). There is a discrepancy between the entire OP in the value of the PER, with more impact in OP4 (Figure 4).

Concerning the LVEF parameter, we observed similar values between the OP involved. This can be related to the different ROI dimensions and geometry used, which tend to overestimate the LV edges, resulting in higher LVEF values. This should be taken into consideration since the sum of the counts in a ROI is assumed to be a proportional measure of a clinically relevant factor.

Additionally, the difficulty to keep the BKG ROI in exactly the same position may be a limitation, regardless of the method applied and the OP involved. This is important because it can be correlated with the variations obtained between OP for the M method, higher values of LVEF for the M method and consequent changes in the diastolic parameters. We realized, particularly for the operators with basic knowledge of nuclear medicine, in both processing methods, that occasionally the BKG ROI overlaid the diastolic ROI, which is not supposed to happen since it may result in overestimated values. Therefore, it is important to evaluate the final results of the processed data and check if BKG ROI is in an overlay position. Nevertheless, it is important to analyze its size as well, since the smaller the BKG area, the higher the %LVEF value and consequently change in the diastolic parameters can introduced and possibly generate biases results.

The determination of the diastolic function parameters is more vulnerable and sensitive than the systolic ones, which may justify the fact that most of the OP presented values significantly different between both methods for these parameters. Considering they are determined by the geometric slope of the final LVEF curve, if the LVEF value is affected, there is a chance that the PFR and PER values are artefactual, which leads us to reinforce the idea that the segmentation process is reflected in the determination of important physiological parameters.

Between methods, the OP1, OP3, and OP5 showed significant differences for all ventricular function parameters, and most discrepancies were produced by the less experienced OP. However, we did not expect to see these differences in one experienced OP. While analyzing the results from the exams processed by OP1, we saw that polygonal ROI were used to draw the limits manually so these variations can be correlated with the ROIs’ shape, dimensions, and area.

More, with the exception of the OP4, which did not present any discrepancies, all other OP presented statistically significant differences between methods in one or more parameters. However, the application of different processing methods should result in different physiological measurements since the M method should only be used when the values obtained by the other are not consensual. The analyzed data results from exams processed by OP4 showed an agreement between A and M ROI, which justifies similar results between the two processing methods.

**Figure 3.** Comparison of the values of A and M PFR of each OP.

**Figure 4.** Comparison of the values of A and M PER of each OP.
Regarding the processing methods available for MUGA exams and as documented by Boudraa et al., the A method should be used preferably if available since it is more reproducible, in comparison with the M method that presents greater interoperator variability, which is consistent with our results. However, in clinical practice, there are some studies where the bounding boxes created by the A mode do not correspond to the real LV edges and in those cases reprocessing by M mode should be performed. For example, in patients with heterogeneous ventricular contraction, it is difficult to accurately segment the ventricular cavity from other cardiac structures, thus sometimes there is the need to delineate manually the structure of interest.

In cases of cardiotoxicity studies, patients usually perform multiple MUGA scans and there is the chance the LV function values differ significantly from the previous values. So, in those cases, there is also the need to reprocess using the M mode to analyze their reliability.

Concerning this intra- and inter-variability and compared with other imaging techniques, echocardiography became a routine to assess LVEF decrease. Although the equipment required being smaller, cheaper, more mobile and more available, MUGA was shown to be more reproducible than echo-cardiography. Additionally, with the emerge of myocardial perfusion imaging, SPECT cardiac imaging became the state of the art in nuclear cardiology. But will MUGA have a comeback? Chen et al. compared LVEF, ED and ES volumes derived from conventional 24-frame gated planar MUGA with the same LV function parameters derived from 24-frame gated CZT SPECT MUGA and 24-frame gated reprojected at 45°CZT planar MUGA. Good overall correlations between each data were found but planar MUGA LVEF and CZT reprojected LVEF values were lower than the CZT SPECT LVEF, suggesting that reprojection of 3D CZT gated data indeed may substitute planar MUGA.

In addition, the results of our study may be influences by some factors. The small and random sample may be a limitation, despite the fact that we tried to overcome this aspect by processing each exam five times per mode. Besides, there are some influencing variables that weren’t taken into consideration such as the gender and weight of the patients that can induce some attenuation in female and obese patients. Also, the best septal angle and detector tilt can influence the final results, since these factors when optimized, may enhance the accuracy of this technique.

**Conclusion**

MUGA scans play an important role in the assessment of cardiotoxicity. On the other hand, PFR and PER values predict early cardiac damages. The evaluation of measurement errors is irrefutable in the search for a better diagnostic quality in clinical practice.

Varying levels of OP professional experience in clinical practice along with the application of different processing methods can lead to discrepancies in the values obtained by the MUGA technique. Variability was found when comparing M and A processing methods, as well as variability inter-operator associated with their level of experience. Despite the overall interoperator oscillation, there was a trend of less variability between the two experienced OP for each processing method. However, the observations of any inequalities existing between OP or within operators may need further studies in the clinical field, in order to obtain a reliable impact on diagnosis and patient management, especially concerning the diastolic parameters.

**References**


Conflito de interesses
Os autores declaram não ter quaisquer conflitos de interesse. Artigo submetido em 05.06.2019 e aprovado em 07.02.2020