

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional



**DOPPLER FLOW PATTERN IN PATIENTS WITH
AORTIC COARCTATION**

Susana Cristina Oliveira Cordeiro

Dissertação orientada por
Prof. João Gomes e Dr. Rui Anjos

Mestrado em Bioestatística

2015

Acknowledgements

To my mentor Rui Anjos, my greatest teacher in Pediatric Cardiology, I owe the idea and motivation for this dissertation.

To my mentor João Gomes, I am grateful for all the patience and time spent in answering all my questions. And for making his aim for all his students to use statistical theory as practical knowledge.

To all the Biostatistic course teachers, thank you for always being available to support students.

To my family, and particularly to my husband, a big appreciation for keeping me on the right track.

Abstract

Patients with coarctation of the aorta (CoAo) often show a Doppler flow pattern with diastolic flow in the descending aorta. The effect of arterial stiffness on CoAo flow pattern was described *in-vitro* and with computer models. Study of Doppler flow patterns may provide helpful data to support the decision of CoAo treatment. Fifty studies were obtained in 31 patients (14 women, 21.5±15.5 years of age). In 19 patients, studies were performed before and after intervention. Systolic invasive gradients were measured (Sgrad). Doppler parameters measured at the time of invasive evaluation, included Doppler corrected gradient (Dgrad), diastolic velocity at end of T wave (DVT), end diastolic velocity (DVQ), systolic and diastolic half pressure times (SHPTc and DHPTc) and velocity runoff (VRc). Arterial stiffness was assessed by measuring pulsed wave velocity (PWV) between right carotid and radial arteries.

With simple regression models, Sgrad showed correlation with Dgrad, DVT, DVQ, SHPTc, DHPTc and VRc ($p < 0.01$). The best multiple linear regression model provided the formula $Sgrad = -4.61 + 0.75 \times Dgrad + 0.06 \times DHPTc$ ($R^2 = 0.77$). A binary variable named Sign were Sign=0 if $Sgrad < 20$ mmHg and Sign=1 if $Sgrad \geq 20$ mmHg was created. The best multiple logistic regression model provided the formula $Ln\left(\frac{p_{Sign}}{1-p_{Sign}}\right) = -4.70 + 0.12 \times Dgrad + 0.06 \times DHPTc - 0.0008 \times Dgrad \times DHPTc$. A cutoff value of 0.34 for $Ln\left(\frac{p_{Sign}}{1-p_{Sign}}\right)$ resulted in a sensibility of 96% and specificity of 74% for this model.

A variable named DTail was obtained with DTail=0 if DHPTc=0 and DTail=1 if DHPTc>0. In the group with Sgrad below 30mmHg, a negative correlation was found between DTail and PWV ($p = 0.05$) suggesting that low aortic stiffness may contribute to persistent diastolic flow in the descending aorta.

Doppler systolic and diastolic parameters correlated well with severity of CoAo. In mild to moderate CoAo, Doppler diastolic flow in the descending aorta was more likely in patients with lower arterial stiffness.

KEYWORDS: Coarctation; Invasive gradients; Arterial Stiffness

Resumo

A coartação da aorta (CoAo) é uma cardiopatia congênita caracterizada pelo estreitamento de um segmento da aorta torácica ou abdominal, mais frequentemente localizada no istmo aórtico. O Doppler codificado a cor e o Doppler espectral são ferramentas utilizadas no ecocardiograma transtorácico de rotina que permitem a avaliação do fluxo sanguíneo na aorta descendente. Em pessoas sem cardiopatia, o padrão de fluxo Doppler na aorta descendente apresenta uma velocidade máxima inferior a 2 metros por segundo e o fluxo ocorre apenas em sístole. Em doentes com CoAo, o padrão de fluxo Doppler apresenta um aumento da velocidade de fluxo sanguínea na aorta descendente e, em alguns casos, exibe uma persistência de fluxo em diástole, denominada extensão diastólica.

A gravidade da doença tem sido avaliada de forma invasiva, semi-invasiva ou com técnicas mais complexas, estudando a relação entre o diâmetro da CoAo e o diâmetro da aorta ao nível do diafragma (CoAo/DAo). No estudo de Carvalho et al. (1990), que utilizou a variável CoAo/DAo calculada por angiografia como referência padrão de gravidade da doença, o estudo por Doppler demonstrou ser mais eficaz em avaliar a gravidade da CoAo quando as quantificações Doppler sistólicas e diastólicas foram consideradas em conjunto. No estudo de Tan et al. (2005), utilizando a variável CoAo/DAo obtida por ressonância magnética e estudos Doppler antes e após implantação de stent, a velocidade de fluxo Doppler diastólica na onda T permitiu prever a gravidade da CoAo. Apesar da relação entre quantificações diastólicas do padrão de fluxo Doppler e a gravidade da CoAo ter sido descrita anteriormente, estudos prévios não utilizaram os gradientes invasivos.

Inicialmente pensou-se que a persistência de fluxo em diástole na aorta descendente de doentes com CoAo dependia apenas da gravidade da CoAo. Estudos prévios *in-vitro* (Tacy et al., 1999) e com modelos computacionais (DeGroff et al., 2003) sugeriram que a rigidez arterial deve ser considerada na avaliação de doentes com CoAo, visto que a extensão diastólica aumenta quando a rigidez arterial diminui.

Assim, é objetivo desta dissertação descrever a relação entre o padrão de fluxo Doppler da CoAo, os gradientes invasivos da CoAo e a rigidez arterial, num grupo selecionado de doentes, utilizando modelos de regressão.

Cinquenta estudos foram obtidos em 31 doentes (14 mulheres, 21.5 ± 15.5 anos de idade). Em 19 doentes, foram realizadas avaliações antes e após intervenção percutânea (dilatação com balão e/ou implantação de stent). Nos 50 estudos obtidos, 12 foram apenas de diagnóstico, 19 foram prévios à intervenção percutânea e 19 foram obtidos após a intervenção percutânea. Foram medidos os gradientes sistólicos invasivos (Sgrad) no cateterismo cardíaco. Por ecocardiograma transtorácico foram avaliados vários parâmetros de Doppler obtidos na altura do procedimento invasivo, que incluíram o gradiente de Doppler corrigido (Dgrad), a velocidade diastólica no final da onda T (DVT), a velocidade em telediástole (DVQ), os tempos de hemipressão sistólicos e diastólicos (SHPTc e DHPTc) e a velocidade runoff (VRc - tempo para a velocidade decrescer do seu valor máximo até 33%). VRc, SHPTc e DHPTc foram corrigidos com a fórmula de Bazett para normalizar as medições de tempos para diferentes valores de frequência cardíaca. A rigidez arterial foi estimada através da medição da velocidade da onda de pulso (PWV - pulse wave velocity) entre as artérias carótida e radial direita, por tonometria.

Através de regressão linear simples, Sgrad apresentou relação com Dgrad, DVT, DVQ, SHPTc, DHPTc e VRc ($p < 0.01$). O modelo de regressão linear múltipla foi obtido, resultando na fórmula $Sgrad = -4.61 + 0.75 \times Dgrad + 0.06 \times DHPTc$. Apesar de Dgrad ter como objetivo estimar o valor que Sgrad por ecocardiografia, o modelo de regressão linear múltipla demonstrou que as variáveis que devem ser utilizadas para prever Sgrad são Dgrad e DHPTc. Este modelo apresentou um melhor ajustamento aos dados comparando com o modelo que inclui apenas a variável Dgrad ($R^2 = 0.77$ incluindo Dgrad e DHPTc versus $R^2 = 0.74$ incluindo apenas Dgrad), demonstrando que DHPTc resolve parte da imprecisão de Dgrad em prever Sgrad.

Foi criada uma variável binária Sign onde Sign = 0 se Sgrad apresentasse valores inferiores a 20 mmHg e Sign = 1 se Sgrad apresentasse valores iguais ou superiores a 20 mmHg. O modelo de regressão logística múltipla foi obtido, resultando na

formula $Ln\left(\frac{p_{Sign}}{1-p_{Sign}}\right) = -4.70 + 0.12 \times Dgrad + 0.06 \times DHPTc - 0.0008 \times$

$Dgrad \times DHTPc$. Porque a consequência de uma CoAo não tratada tem um impacto significativo na saúde do doente, foi escolhido o *cutoff* de 0.34 para $Ln\left(\frac{p_{Sign}}{1-p_{Sign}}\right)$ que resultou em 96% de sensibilidade e 74% de especificidade para este modelo, ou seja apenas 4% de falsos negativos e 26% de falsos positivos. Para os doentes identificados como falsos positivos no modelo de regressão logística proposto, a consequência é apenas a realização de um cateterismo de diagnóstico que tem baixa probabilidade de complicações.

Os estudos prévios que demonstraram o efeito da rigidez arterial no padrão de fluxo Doppler simularam esse efeito para CoAo com o mesmo grau de severidade. Posto isto, foi estudada a relação de PWV com as restantes variáveis na amostra inteira e em grupos selecionados de doentes. Como esperado, não foi encontrada relação significativa entre PWV e as restantes variáveis nos 28 doentes em que foi obtida medição de PWV (em 22 doentes não foi obtida medição de PWV). Através de análise gráfica, o grupo com Sgrad inferior a 30 mmHg foi selecionado (19 doentes). Para avaliar a relação entre PWV e DHPTc neste grupo, foi obtida uma nova variável DTail com DTail = 0 se DHPTc fosse também igual a zero e DTail = 1 se DHPTc apresentasse valores superiores a zero. Neste grupo com Sgrad inferior a 30 mmHg, verificou-se uma correlação negativa entre DTail e PWV através do modelo de regressão logística com a formula $Ln\left(\frac{p_{DTail}}{1-p_{DTail}}\right) = 6.69 - 0.94 \times PWV$, com $p=0.05$ para o teste da razão de verosimilhanças. Este modelo prevê que para cada redução de 1 metro por segundo da PWV, existe uma probabilidade 2.57 vezes maior para DTail = 1, no grupo de doentes com Sgrad inferior a 30 mmHg. Este achado é concordante com os estudos prévios *in-vitro* e com modelos computacionais. Uma artéria com menor rigidez aumenta de diâmetro quando a onda de pulso a atravessa e depois regressa ao diâmetro basal. Na presença de CoAo, este retorno ao diâmetro basal no início da diástole pode ser o responsável pelo fluxo sanguíneo tardio que resulta na extensão diastólica. Por outro lado, uma artéria com maior rigidez não sofre o aumento de diâmetro fisiológico quando a onda de pulso a atravessa. Assim, a ausência de extensão diastólica em doentes com CoAo significativa pode ser devido a uma rigidez arterial aumentada. A

relação significativa entre PWV e DTail sugere que a rigidez arterial reduzida pode contribuir para persistência de fluxo em diástole, sobretudo em doentes com CoAo ligeira ou moderada. Em doentes com CoAo grave, o obstáculo significativo parece induzir *per se* a presença de extensão diastólica, seja o PWV baixo ou elevado.

Em suma, Dgrad e DHPTc são variáveis obtidas por Doppler espectral no ecocardiograma que podem ser utilizadas para prever o gradiente invasivo da CoAo, Sgrad. Em doentes com CoAo ligeira a moderada, a rigidez arterial parece influenciar o valor de DHPTc visto valores mais elevados de DHPTc serem mais prováveis em doentes com menor rigidez arterial. Estes resultados foram obtidos pela aplicação de modelos de regressão que permitiram construir fórmulas matemáticas que podem ser utilizadas ao estudar os doentes com CoAo.

PALAVRAS-CHAVE: Coartação; Gradientes invasivos; Rigidez arterial

Contents

List of Abbreviations.....	xi
List of Figures.....	xiii
List of Tables.....	xv
Preface	xvii
1 Biological Background	1
1.1 Coarctation of the aorta	1
1.1.1 Morphology.....	1
1.1.2 Pathophysiology.....	3
1.1.3 Diagnosis.....	3
1.1.4 Treatment options.....	5
1.2 Arterial Stiffness	6
2 Statistical background.....	9
2.1 Linear Regression	9
2.1.1 Simple linear regression.....	9
2.1.2 Multiple linear regression	13
2.2 Logistic Regression	17
2.2.1 Simple logistic regression.....	17
2.2.2 Multiple logistic regression.....	19
3 Methods and Results.....	21
3.1 Invasive Gradients.....	25
3.2 Arterial Stiffness.....	34
4 Discussion	39
References.....	45
A R Scripts	47

List of Abbreviations

AIC	Akaike's Information Criterion
ANOVA	Analysis of Variance
BLUE	Best linear unbiased estimator
CoAo	Coarctation of the aorta
CHD	Congenital heart disease
Dgrad	Doppler corrected gradient
DHPTc	Diastolic half pressure time, corrected with Bazett's formula
DVQ	Diastolic velocity at Q wave
DVT	Diastolic velocity at the end of T wave
DTail	Diastolic tail
OR	Odds ratio
PWV	Pulsed wave velocity
ROC	Receiver Operating Characteristic
Sgrad	Systolic invasive gradients
Sign	Significant invasive gradient
SHPTc	Systolic half pressure time, corrected with Bazett's formula
TTE	Transthoracic echocardiogram
VIF	Variance inflation factor
VRc	Velocity runoff, corrected with Bazett's formula
2D	Two-dimensional

List of Figures

Figure 1	Normal aortic arch (left) and aortic arch with CoAo (right).
Figure 2	Color Doppler (left) and spectral Doppler (right) in CoAo.
Figure 3	Treatment options for CoAo: surgery (left), balloon dilatation (middle) and stenting (right).
Figure 4	Difference of blood flow velocity in compliant (left) and stiff (right) arteries.
Figure 5	Possible plots of the residuals. The horizontal line marks the zero. The top left plot shows Normal distribution of the residuals. The top 3 plots are all of homoscedastic residuals, but the middle and right are biased residuals. The bottom 3 plots are all of heteroscedastic residuals, but only the left is unbiased.
Figure 6	μ_x has a logistic function (left) and $\text{Ln} \left(\frac{p_x}{1-p_x} \right)$ is linear with x (right).
Figure 7	Study sample.
Figure 8	Method used for measuring the PWV with tonometry.
Figure 9	Scatter plot analysis between Sgrad and the different TTE variables. The black lines correspond to the regression line for the simple linear regression models. The black dots in the VRc plot represent the simple regression model for Sgrad above and below 30 mmHg.
Figure 10	Best multiple regression model for the dependent variable Sgrad, obtained with R software.
Figure 11	Multicollinearity studied by correlation matrix and VIF.

Figure 12	Scatter plot (left) and Normal Q-Q plot (right) of the residuals.
Figure 13	Scatter plot analysis between Sign and the different TTE variables. Notice that all patients with DVQ superior to zero had Sign = 1.
Figure 14	Best logistic multiple regression model for odds ratio of the dependent variable Sign, obtained with R software.
Figure 15	Sensibility and specificity curves of the multiple logistic model.
Figure 16	ROC curve of the multiple logistic model.
Figure 17	Sensibility and specificity values for choosing the cutoff of the model.
Figure 18	Scatter plots between Sgrad and the TTE variables, Dgrad and DHPTc. In the group of patients with Sgrad bellow 30 mmHg (horizontal line in the left plot), patients with higher PWV seem to have lower values of DHPTc. The black dots are from patients without measurement of PWV (missing data).
Figure 19	Scatter plot between Sgrad DHPTc, in the group of patients with Sgrad bellow 30 mmHg. Patients with higher PWV seem to have lower values of DHPTc. The black dots are from patients without measurement of PWV (missing data).
Figure 20	Boxplot between PWV and DTail, in the group with Sgrad bellow 30 mmHg.
Figure 21	Simple logistic regression model with DTail as the independent variable and PWV as the independent variable (top). OR of DTail for each unit of PWV (bottom).

List of Tables

Table 1	ANOVA table with n the total of data, m groups being compared, SS the sum of squares where $SS_{Tot} = SS_{Reg} + SS_E$, Y the dependent variable, \hat{Y} the estimator for Y , \bar{Y} the mean of Y , ε the statistical error and $i=1, \dots, n$.
Table 2	Variables included in the study.

Preface

Coarctation of the aorta (CoAo) is a congenital heart disease (CHD) characterized by the narrowing of a segment of the aorta. Color and spectral Doppler, tools used in standard routine echocardiography studies, allow for evaluation of blood flow in the aorta. In normal subjects, Doppler flow pattern in the descending aorta shows a peak velocity lower than 2 meters per second and flow occurs only during systole. In patients with CoAo, Doppler flow pattern shows an increased velocity in the descending aorta and, in some cases, a persistence of flow in diastole often referred as a diastolic tail. Color Doppler provides information of the narrowing site whereas spectral Doppler allows for quantitative analysis.

Previous studies have described a relation between CoAo Doppler flow pattern and CoAo severity. Using *in-vitro* and computer models, a relation between CoAo Doppler flow pattern and arterial compliance was demonstrated.

Carvalho et al. (1990) conducted a study where indices of severity of CoAo derived from non-invasive Doppler echocardiography were compared with measurements derived angiography. Using 24 studies from 17 patients, Doppler variables, including peak systolic and diastolic gradients and time to half peak systolic and diastolic velocities, were compared to a single angiography variable, ratio of CoAo diameter to the diameter of descending aorta at the level of diaphragm. A peak systolic gradient above 40 mmHg or time to half peak diastolic velocity above 100 ms were found to be highly specific for detecting an angiographic ratio below 0.5, with diastolic measurements more sensitive for diagnosis of severe coarctation than systolic measurements. They concluded that Doppler echocardiography was an effective non-invasive method of assessing severity of CoAo, particularly when systolic and diastolic events were considered together.

Another study conducted by Tan et al. (2005) focused on evaluating the effect of successful stenting on the CoAo Doppler flow pattern and identifying TTE indexes that could be used for follow-up. A TTE analysis before and after stenting included the variables peak systolic pressure gradient, diastolic velocity, end-diastolic tail velocity, systolic and diastolic velocity half-time index, systolic and diastolic

pressure half-time index. All TTE variables were compared to a single value for CoAo severity evaluation, the CoAo index (defined by the ratio of the narrowest coarctation cross-sectional area to the area of the abdominal aorta at diaphragm level) measured by magnetic resonance imaging. Diastolic velocity above 193 centimeters per second and a diastolic/systolic velocity ratio above 0.53 were highly predictive for a CoAo index below 0.25, and thus severe CoAo.

Using an *in-vitro* pulsatile flow model with four levels of CoA severity, Tacy et al. (1999) studied the relation between Doppler flow patterns at CoAo site and aortic compliance (calculated from local arterial stiffness). Diastolic runoff, defined as time between Vmax and 33% Vmax, had a positive linear relation with aortic compliance. They concluded that absence of a longer diastolic runoff in Doppler flow pattern should not exclude the diagnosis of significant CoAo.

In a computer model study, DeGroff et al. (2003) sought to investigate fluid and wall mechanics present in CoAo. They studied the relationship between diastolic runoff in Doppler flow pattern and aortic compliance (calculated from local arterial stiffness), using 3 computational numeric models of CoAo. In these simulations, the degree of diastolic runoff increased with arterial compliance. It was concluded that an increased aortic compliance induced greater dilatation and stored energy upstream the CoAo site in systole, with downstream release of the stored energy in diastole as the aortic wall recoils.

Although the relation between diastolic quantifications from the CoAo Doppler flow pattern and CoAo severity was already described, previous studies used variables obtained with the diameter of CoAo and not invasive gradients. But persistence of flow in diastole in the descending aorta of CoAo patients was thought to be solely dependent on lesion severity. Previous *in-vitro* and computer model studies suggested that arterial stiffness and compliance should be considered when evaluating a patient with CoAo, since the diastolic tail increases with decreasing arterial stiffness.

Therefore, this dissertation aims to describe the relation between CoAo Doppler flow pattern, CoAo invasive gradients and arterial stiffness, in a group of selected patients, using regression models.

1 Biological Background

This section is meant to describe CoAo and the concept of arterial stiffness. It provides an overview of CoAo definition, morphology and pathophysiology, diagnosis and treatment options. The concept of arterial stiffness is also described in this chapter.

1.1 Coarctation of the aorta

CoAo, or aortic coarctation, refers to narrowing of a segment of the thoracic or abdominal aorta, but it is most commonly located to the aortic isthmus (juxtaductal CoAo). CoAo is the seventh most common form of CHD, occurring in approximately 5 to 7% of CHD patients (approximately 36 per 100 000 live births), more likely in males. In some patients, CoAo manifests soon after birth, whereas in others it manifests at an older age. Whether it is genetic, environmental or hemodynamic, no single cause for CoAo has been proven. Abnormal flow distribution during fetal life with decreased aortic flow has long been suspected and histologic studies have demonstrated ductal tissue circumferentially surrounding the juxtaductal portion of the aorta. The role of genetic factors is increasingly recognized, for example CoAo occurs in approximately 12% of patients with Turner syndrome. CoAo is commonly associated with bicuspid aortic valve, ventricular septal defect and almost every type of CHD, particularly left heart obstructive lesions. A high incidence of associated cardiac abnormalities suggests that CoAo is a more complex defect than isolated narrowing of the aorta. (Lai et al., 2009)

1.1.1 Morphology

A juxtaductal CoAo usually results from narrowing of the aortic isthmus, in the proximal descending aorta at the arterial duct insertion. Different morphologic CoAo patterns can be distinguished based on age at diagnosis. In fetus and infant CoAo, the distal transverse aortic arch between the left common carotid and left

subclavian arteries is often hypoplastic, the angle between the ascending aorta and transverse arch is acute, the aortic isthmus is diffusely hypoplastic, and the arterial duct is usually patent. In older children and adult CoAo, aortic arch hypoplasia is less common, the coarctation segment is usually discrete and collateral arteries bypassing the CoAo are common. Unrelated to the age at diagnosis, the CoAo segment is characterized by luminal narrowing due to thickening of intima and media layers, hypoplasia and/or tortuosity. The length of the stenotic segment varies and the proximal descending aorta immediately past the CoAo often exhibits post-stenotic dilation. (Lai et al., 2009)

The morphology of abdominal CoAo is different from juxtaductal CoAo in that the involved segment is often long, the involvement of the renal and mesenteric vessels is common and the aortic media is thickened. (Lai et al., 2009)

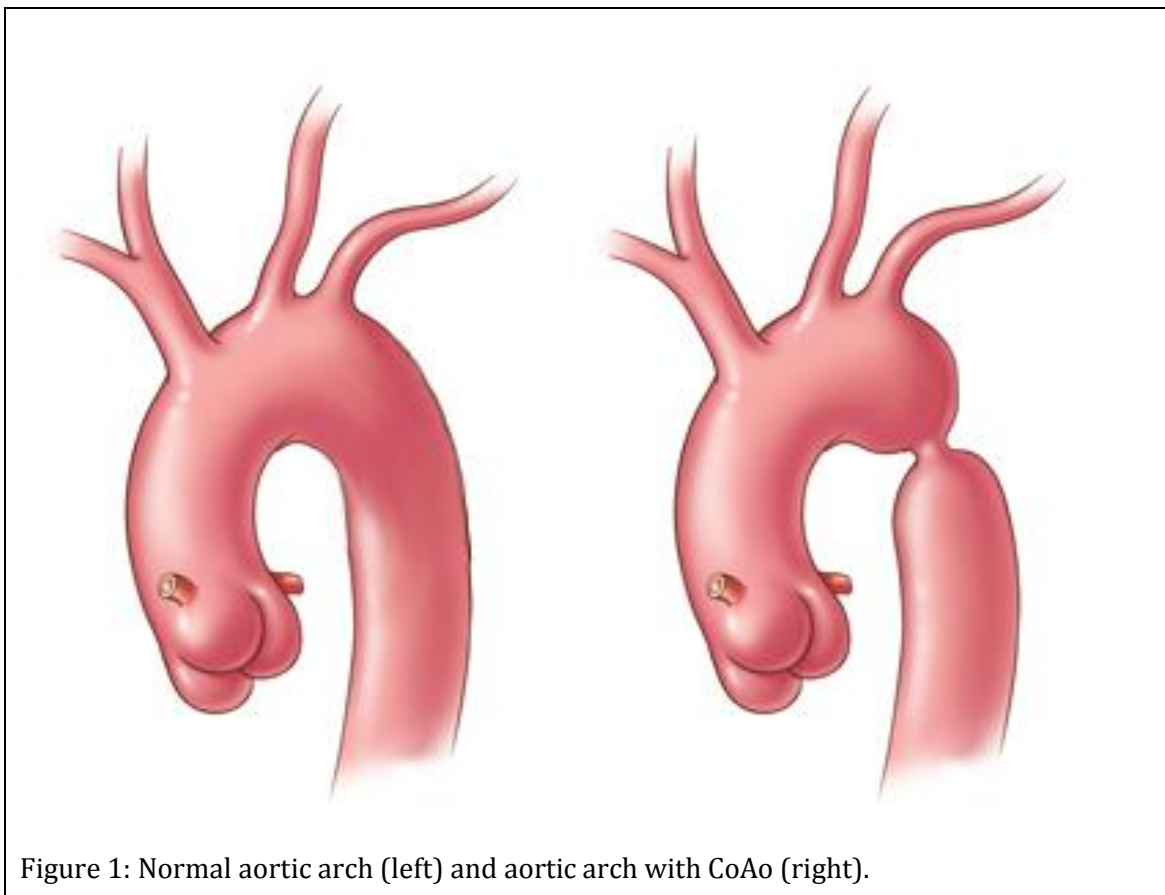


Figure 1: Normal aortic arch (left) and aortic arch with CoAo (right).

1.1.2 Pathophysiology

CoAo is usually well tolerated in the fetus because blood flow to the lower body and the placenta is supplied predominantly through the arterial duct. But narrowing of the aortic isthmus may result in diversion of blood flow from the aorta to the arterial duct and pulmonary artery. As a result, the right ventricle dilates whereas the left ventricle is pressure loaded, resulting in the ventricular size disproportion in the fetus. The hemodynamic burden imposed by severe CoAo manifests after birth as the foramen ovale closes and the arterial duct constricts and ultimately closes. As a result, the cardiac output and flow to the lower body must cross the narrow aortic segment. Under these conditions, systolic blood pressure in the upper body is increased. With no patent arterial duct, systolic blood pressure is low distal to the CoAo, clinically manifesting as reduced pulse amplitude in the femoral arteries. In the neonatal period, a patent arterial duct may be present, allowing right-to-left systolic flow from the main pulmonary artery to the descending aorta, providing adequate perfusion to the lower body and normal volume of the femoral pulse. In the absence of an associated cardiac lesion, oxygen saturation in the upper extremities is greater than that in the lower extremities, a phenomenon called differential cyanosis. Constriction and closure of the arterial duct in neonates with severe CoAo may also lead to left ventricular dysfunction. (Lai et al., 2009)

Later in life, during childhood or in adults, CoAo is usually diagnosed either due to a heart murmur, low pulse amplitude in the lower extremities or systemic hypertension. Multiple collateral vessels tend to develop between the high-pressure aortic branches proximal to the CoAo and the low-pressure distal to the CoAo. (Lai et al., 2009)

1.1.3 Diagnosis

Clinical diagnosis of CoAo usually rests on identifying a blood pressure difference between the upper and lower extremities, information that can be obtained by palpation of both radial and femoral arteries. If a substantial difference between the two is found, coarctation of the aorta should be suspected. In a patient without

cardiac disease, blood pressure should be the same in upper and lower extremities. If the systolic blood pressure is at least 20 mmHg higher in the arms when compared to the legs, it is very likely that the patient has CoAo. In neonates or infants, the signs of congestive cardiac failure may be also present. (Johnson & Moller, 2014)

Transthoracic Echocardiography (TTE) provides adequate diagnostic information in most CoAo cases. However, some patients may require other diagnostic modalities, such as magnetic resonance imaging, computed tomography or angiography (Lai et al., 2009).

The basic elements of a standard TTE study are two-dimensional (2D) images enhanced by Doppler, color Doppler and M-mode information in multiple imaging planes. The TTE is organized by acoustic “windows” from which the heart is examined. In pediatric echocardiography, a segmental approach should always be used to describe all of the major cardiovascular structures in sequence, allowing the imaging of any structural or functional CHD. Any laboratory performing a pediatric echocardiogram should have a written examination protocol that outlines the views to be obtained, the imaging modalities and the methods for recording. (Lai et al., 2009)

The main goals of echocardiographic examination in the setting of CoAo are: evaluation of heart situs and segmental cardiac anatomy; evaluation of aortic arch anatomy; color Doppler evaluation of the flow profile in the descending aorta; assessment of spectral Doppler flow pattern at the CoAo site; evaluation of flow in the arterial duct; evaluation of morphology, size and function of the ventricles, inflow and outflow; identification of associated anomalies (Lai et al., 2009). Cross-sectional images for evaluation of aortic arch anatomy, that reveal the narrowing site, are usually best obtained with the transducer positioned near the suprasternal notch. Color Doppler shows a turbulent signal at the CoAo site and spectral Doppler shows high-velocity flow to the descending aorta, sometimes with persistence of blood flow in diastole. In neonates, the diagnosis may be difficult due to the presence of the ductus arteriosus. (Johnson & Moller, 2014)

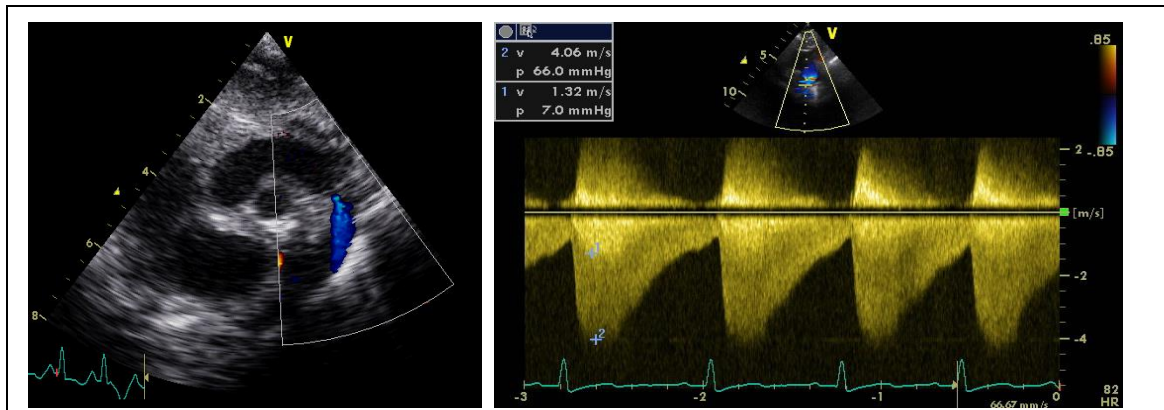
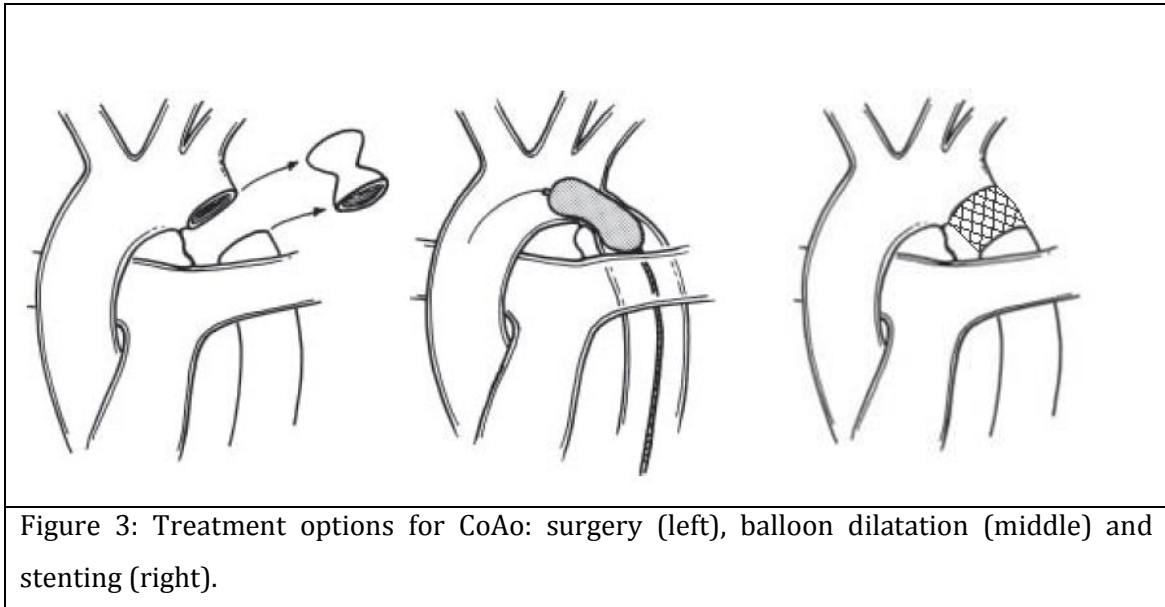


Figure 2: Color Doppler (left) and spectral Doppler (right) in CoAo.

Obtained by cardiac catheterization, an invasive diagnosis method, angiography is the gold standard imaging method for demonstrating the CoAo morphology. Invasive pressure measurements demonstrate systolic hypertension proximal to the coarctation and a gradient at the site of the coarctation, shown by pullback of the catheter across the lesion during pressure recording. (Johnson & Moller, 2014)

1.1.4 Treatment options

Treatment options for correction of CoAo include surgery, percutaneous balloon angioplasty and endovascular stent implantation. Surgery is the preferred treatment option for CoAo in neonates or for severe/complex CoAo anatomy in all patients. Balloon angioplasty and/or stent implantation are commonly used for treatment of CoAo in older children and adults. Balloon dilation of native coarctation avoids some surgery disadvantages but it is often less effective. Implantation of a metallic stent at the time of balloon dilation may provide better results but in small patients the stents do not allow for growth. Evaluation after surgery or transcatheter treatment of CoAo is similar to that of a native CoAo. Potential complications after treatment of CoAo that should be addressed include residual CoAo (immediately after treatment), recoarctation (development of narrowing after successful treatment), aneurysm formation, aortic dissection and persistent arterial hypertension. (Johnson & Moller, 2014)



1.2 Arterial Stiffness

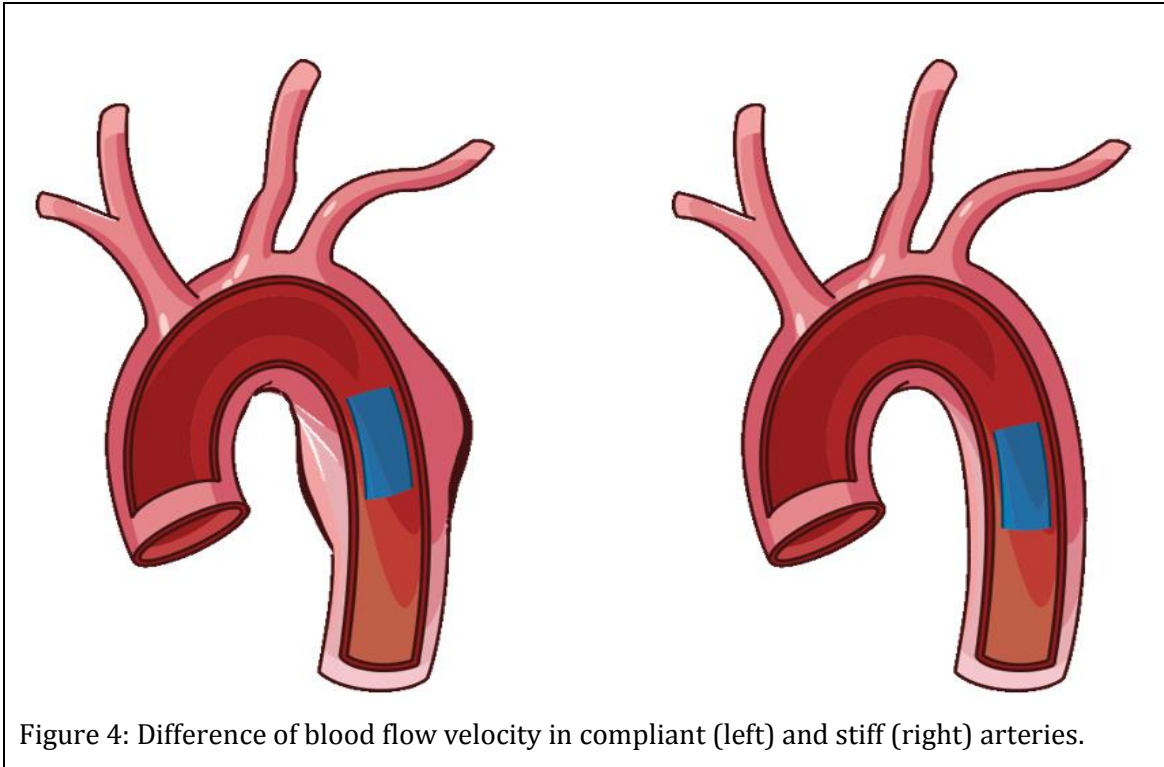
The mechanical behavior of the circulatory system is extremely complex. Arteries are known to have viscoelastic properties and powerful adaptive mechanisms. Arterial stiffness, and the opposite concept of arterial compliance, is related mainly to the viscoelastic properties of arteries. Local and regional arterial stiffness can be measured at various sites along the arterial tree. The downside is that elastic properties of arteries vary along the arterial tree. (Laurent et al., 2006)

Local arterial stiffness can be measured through high quality imaging methods, such as magnetic resonance imaging or computed tomography. Usually it is obtained by measuring the difference between the area or diameter of the artery in systole and in diastole. (Laurent et al., 2006)

Pulse wave velocity (PWV) is generally accepted as a simple, non-invasive and reproducible measurement to determine regional arterial stiffness. Carotid-femoral PWV has been used in epidemiological studies demonstrating the predictive value of aortic stiffness for cardiovascular events. (Laurent et al., 2006)

Because compliant arteries suffer a small increase in diameter when the pulse wave passes through, this arterial wall motion absorbs some of the energy produced by the left ventricle in systole. This small and brief dilatation of the artery slows down the velocity of the pulse wave, but the propagation of the pulse

wave is more efficient due to arterial wall elasticity. Therefore, PWV is lower in more compliant arteries and higher in stiffer arteries (Figure 4).



2 Statistical background

This section is meant to describe theoretical knowledge for simple and multiple linear and logistic regression, explaining how to build regression models and how to evaluate them.

2.1 Linear Regression

In regression, when a variable Y is influenced by a variable X , Y is called the dependent variable and X is the independent variable. If Y is a continuous variable, linear regression aims to represent the influence of one or more independent variables on Y through a straight line, called the regression line. (Oliveira, 2009)

2.1.1 Simple linear regression

In the particular case of simple linear regression, it is assumed that there is only the relationship between X and Y . Represented by Y_x , each element of the population is described by the values of X and Y . A linear regression model allows for the equation:

$$Y_x = \mu_x + \varepsilon_x = \beta_0 + \beta_1 x + \varepsilon_x .$$

In this equation: Y_x is the dependent variable; μ_x is the mean value of the dependent variable Y_x when the value for the independent variable X is x ; ε_x is the statistical error corresponding to possible unknown effects to Y_x beside the effect of x ; β_0 is the mean value of the dependent variable when the independent variable is equal to zero; β_1 is the change in the mean value of Y_x for each unit of X . Hence, Y answers to X through a straight line (regression line) with origin in β_0 and slope equal to β_1 . The values for ε for the different elements in the population are not related and should be independent by having a Gaussian probability distribution. (Gomes, 2011)

Assuming a random sample of size equal to n of a population Y with the characteristic X , $((x_1, Y_1), \dots, (x_n, Y_n))$, such as $Y_i = \mu_i + \varepsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, with $i=1, \dots, n$, $\varepsilon_i \cap N(0, \sigma)$ and $Y_i \cap N(\mu_i = \beta_0 + \beta_1 x_i, \sigma)$, we can estimate the parameters

of the regression model, β_0 and β_1 , by the method of ordinary least squares or the method of maximum-likelihood estimation. (Gomes, 2011)

The **method of least squares** is a standard approach to estimate β_0 and β_1 , where the regression line is obtained by defining the point where the variance of the distance of that point to all others is minimal (Oliveira, 2009). So, this method consists on finding the values that produce the minimum value of $SQ = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$ using the partial derivatives of SQ with respect to β_0 and β_1 .

Another method for estimation of the parameters of the regression model is the **maximum-likelihood estimation**. Through the assumptions defined earlier for the population Y and Y_i , we can conclude that (Y_1, \dots, Y_n) is a vector of variables Y_i with Gaussian distribution with $\mu_x = \beta_0 + \beta_1 x$ and σ , the log-likelihood function is: $\ln L(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$. The estimators for β_0 and β_1 can be found by using the partial derivatives of $\ln L$ with respect to β_0 and β_1 . Both of these methods provide the **estimator for β_1 and β_0** through the formulas:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

and

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}.$$

The properties of these estimators stand that $\hat{\beta}_1$ is an unbiased estimator for β_1 , as $\hat{\beta}_0$ is an unbiased estimator for β_0 . Both methods also provide the estimator for σ^2 through the formula: $\hat{\sigma}^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 / n$. (Gomes, 2011)

For testing the significance of the independent variable X , the test *null hypothesis* (H_0) states that there is no significant alteration in the mean value of the dependent variable Y when X changes ($H_0 : \beta_1 = 0$), were the *alternative hypothesis* (H_1) states that there is a significant alteration in the mean value of Y when X changes ($H_1 : \beta_1 \neq 0$ or $\beta_1 > 0$ or $\beta_1 < 0$). Analysis of Variance (ANOVA) table is used to analyze the simple regression model. In an **ANOVA table** (Table 1), the column headings are usually "Source", "DF", "SS", "MS", "F" and "P", where "Source" is the source of variation in the data (Regression, Error and Total), "DF" is the degrees of freedom in the source, "SS" is the sum of squares due to the source, "MS" is the

mean sum of squares, “F” is the test statistic F and “P” is the p-value for the F -test. (Pennsylvania State University, 2015)

Source	DF	SS	MS	F	P
Regression	$m-1$	$SS_{Reg} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	$MS_{Reg} = \frac{SS_{Reg}}{(m-1)}$	$\frac{MS_{Reg}}{MS_E}$	$p\text{-value}$
Error	$n-m$	$SS_E = \sum_{i=1}^n \varepsilon_i^2$	$MS_E = \frac{SS_E}{(n-m)}$		
Total	$n-1$	$SS_{Tot} = \sum_{i=1}^n (Y_i - \bar{Y})^2$			

Table 1: ANOVA table with n the total of data, m groups being compared, SS the sum of squares where $SS_{Tot} = SS_{Reg} + SS_E$, Y the dependent variable, \hat{Y} the estimator for Y , \bar{Y} the mean of Y , ε the statistical error and $i=1, \dots, n$.

Assuming the null hypothesis and with $n-2$ degrees of freedom in all simple regression models, the test statistic F has F distribution with parameters $(1, n-2)$. With the level of significance of $100(1-\alpha)\%$, it is also possible to obtain a **confidence interval** for β_1 , through $[\hat{\beta}_1 - t_{(n-2, \frac{\alpha}{2})} S_{\hat{\beta}_1}, \hat{\beta}_1 + t_{(n-2, \frac{\alpha}{2})} S_{\hat{\beta}_1}]$, with $S_{\hat{\beta}_1} = \sqrt{\frac{MS_E}{\sum_{i=1}^n (x_i - \bar{x})^2}}$. It is also possible to estimate the strength of the linear relation between X and Y through the square of the sample correlation coefficient (square of r_{XY}), called the **determination coefficient** (R^2):

$$R^2 = \frac{SS_{Reg}}{SS_{Tot}} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{[\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2} = \left[\frac{s_{XY}}{s_X s_Y} \right]^2$$

that assumes values within the interval $[0,1]$, where $R^2 \approx 1$ means that there is a perfect relation between X and Y while $R^2 \approx 0$ means that there is no relation. (Gomes, 2011)

The residuals of the model should obey the Gauss Markov assumptions: the expected value is equal to zero for all observations (unbiased); conditional variance is a constant (homoscedasticity); have Normal distribution; the residuals are all uncorrelated. Verifying these assumptions can be performed by the analysis of the residuals, where a visual examination of the residuals plot alone provides indications of possible violation of the Gauss Markov assumptions (Figure 5).

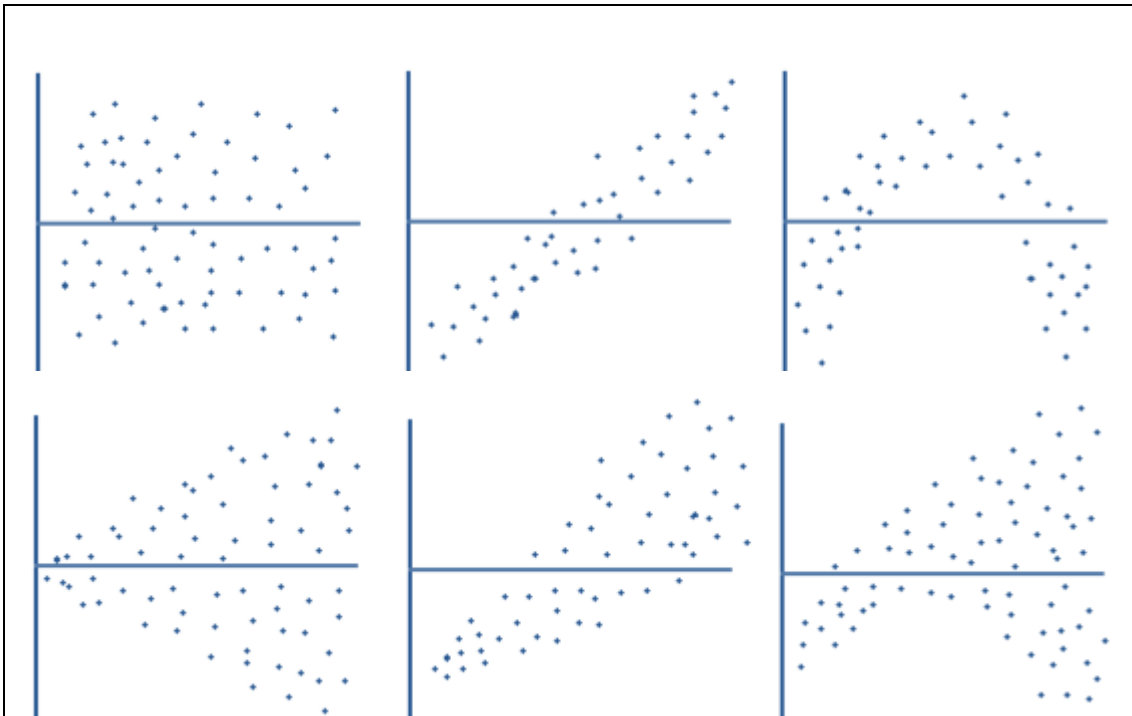


Figure 5. Possible plots of the residuals. The horizontal line marks the zero. The top left plot shows Normal distribution of the residuals. The top 3 plots are all of homoscedastic residuals, but the middle and right are biased residuals. The bottom 3 plots are all of heteroscedastic residuals, but only the left is unbiased.

If the points in a residual plot are randomly dispersed around the horizontal axis with no particular pattern, a linear regression model is appropriate for the data because the residuals are homoscedastic (homogeneity of variance) and unbiased, otherwise a non-linear model is more appropriate. If the point in a residual plot are dominant in one side of the horizontal line marking the zero, this means that the distribution of the residuals is not normal. If the dispersion of the residuals is not constant and increase for greater values of the predicted values, this is called heteroscedasticity and may be corrected by changing the dependent variable (for example, using logarithm). When the residuals plot show a particular pattern, the residuals are biased, even if there is homoscedasticity. After a visual examination of the residuals plot, it is possible to test if the residuals have a normal distribution using, for example, the Kolmogorov-Smirnov test. (Oliveira, 2009)

2.1.2 Multiple linear regression

The aim for multiple linear regression is to find a linear equation that can predict the mean value of the dependent variable Y as a function of p independent variables X_j , with $j=1, \dots, p$. Considering an individual of the population with the characteristics $\mathbf{x} = (x_1, \dots, x_p)$, if there is a linear relation with the value associated to the characteristic $Y (Y_x)$, that value can be described by the equation:

$$Y_x = \mu_x + \varepsilon_x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon_x .$$

In this model: Y_x is the dependent variable; μ_x is the mean value of the dependent variable; β_0 is the mean value of the dependent variable when the value of all independent variable is equal to zero; ε_x is the statistical error corresponding to possible unknown effects to Y_x beside the effect of \mathbf{x} . The coefficients β_j indicate the changes in the mean value μ_x for each unit of x_j , when all other variables x_j are constant. This model allows for accessing the marginal effect for each variable x_j in μ_x . (Gomes, 2011)

Assuming a sample of size equal to n of a population Y with the characteristic \mathbf{X} where $\varepsilon_i \cap N(0, \sigma)$, $Y_i \cap N(\mu_i, \sigma)$ and $E(\varepsilon_i \varepsilon_j) = 0$ ($i \neq j$), then:

$$Y_i = \mu_i + \varepsilon_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \text{ with } i=1, \dots, n$$

or in the matrix form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ where } \mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & x_{21} & \vdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \text{ and } E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} .$$

Similar to the simple linear regression, both methods of ordinary least squares and of maximum-likelihood estimation provide the same estimators for the vector $\boldsymbol{\beta}$. When estimating the parameters of the model, vector $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)$, the method of least squares consists on finding the values that produce the minimum value of $SQ(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2$, or in the matrix form $SQ(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$, by using the partial derivatives of SQ with respect to $\boldsymbol{\beta}$. In the alternative maximum-likelihood estimation method, the aim is to maximize the log-likelihood function, $L(\beta_0, \dots, \beta_p, \sigma^2) = \prod_{i=1}^n \phi\left(\frac{Y_i - \mu_i}{\sigma}\right) / \sigma$, where ϕ is the probability density function of the reduced Gaussian distribution, by finding the values for $\boldsymbol{\beta}$ that maximize the logarithm of

the log-likelihood function using the partial derivatives of $l(\boldsymbol{\beta}, \sigma^2)$ with respect to $\boldsymbol{\beta}$. Both methods provide the formula:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

$\widehat{\boldsymbol{\beta}}$ is a vector with $(p+1)$ components. The mean value of $\widehat{\boldsymbol{\beta}}$ is $\boldsymbol{\beta}$ and the matrix of covariance of $\widehat{\boldsymbol{\beta}}$, $Cov(\widehat{\boldsymbol{\beta}})$, is $\sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$. (Gomes, 2011)

The ANOVA table can also be obtained for statistical inference of the multiple simple regression model (Table 1). $MS_E = \frac{SS_E}{n-p-1}$ is a unbiased estimator for σ^2 . If

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{MS_E [(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}}} \cap T_{n-p-1}, \text{ with } j=1, \dots, p \text{ and } i=1, \dots, n, \text{ then the T-test of the coefficients}$$

states that there is significant change in the mean value of \mathbf{Y} when \mathbf{X} changes if

$$\frac{|\widehat{\beta}_j|}{\sqrt{MS_E [(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}}} > t_{n-p-1; 1-\alpha/2}. \text{ (Gomes, 2011)}$$

If $\boldsymbol{\theta} = \mathbf{C}^T \boldsymbol{\beta}$ is a linear combination of the elements of $\boldsymbol{\beta}$, then $\widehat{\boldsymbol{\theta}} = \mathbf{C}^T \widehat{\boldsymbol{\beta}}$ is a best linear unbiased estimator (BLUE) according to the Gauss-Markov theorem. $\widehat{\boldsymbol{\theta}}$ is a Gaussian vector with $E(\widehat{\boldsymbol{\theta}}) = \mathbf{C} \boldsymbol{\beta}$ and $Cov(\widehat{\boldsymbol{\theta}}) = \sigma^2 \mathbf{C} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T$. (Faraway, 2002)

If the aim is to test $H_0: \widehat{\boldsymbol{\theta}} = \mathbf{h}$ versus $H_1: \widehat{\boldsymbol{\theta}} \neq \mathbf{h}$, one option is to use the test statistic:

$$F = \frac{(\widehat{\boldsymbol{\theta}} - \mathbf{h})^T (\mathbf{C} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T)^{-1} (\widehat{\boldsymbol{\theta}} - \mathbf{h})}{q MS_E}$$

Assuming H_0 , this test statistic has F distribution with parameters $(q, n-p-1)$ and $\mathbf{C} = [\mathbf{c}_1^T, \dots, \mathbf{c}_q^T]$ is the matrix with the linear combinations interesting to test. For example, if the aim is to test if $\beta_1=0$ and $\beta_2 + \beta_3=0$, then we have $q=2$, $\mathbf{c}_1^T = [0 \ 1 \ 0 \ 0 \ \dots \ 0]$, $\mathbf{c}_2^T = [0 \ 0 \ 1 \ 1 \ \dots \ 0]$ and $\mathbf{h} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. (Gomes, 2011)

As seen in the simple linear regression, the **determination coefficient** (R^2) is defined by $R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{SS_{Reg}}{SS_{Tot}} = 1 - \frac{SS_E}{SS_{Tot}}$. But R^2 raises with the number of parameters, not allowing a reliable information of the number of variables to include in the model. Therefore, R^2 is often replaced by the adjusted- R^2 , or R_a^2 , defined by:

$$R_a^2 = 1 - \frac{MS_E}{MS_{Tot}} = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / (n - p - 1)}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / (n - 1)}$$

and whose name comes from the fact that it can be calculated using the same values for R^2 adjusted with the respective degrees of freedom. (Gomes, 2011)

As a consequence of $\hat{\beta}$, it is possible to find the predicted values \hat{Y} , corresponding to the values in the linear regression line, and find the residuals through $e = Y - \hat{Y}$. The analysis of the residuals should be performed the same way as it was described in the simple linear regression model.

A multiple regression model should not have unnecessary independent variables and should have all significant independent variables to explain the variability of the dependent variable. Including unnecessary independent variables in the regression model may lead to imprecise estimators for the coefficients due to excessive fitting of the data, a problem called overfitting. On the other hand, by not including all significant independent variables, its effect on the dependent variable will be distributed on the other independent variables, biasing the coefficients. (Oliveira, 2009)

Selecting the variables according to the higher value of R^2 always leads to the choice of the complete model. Even when using the adjusted- R^2 , this correction is still not enough and tends to select too many variables. If the model has less than the necessary independent variables, the estimator for σ^2 tends to be higher than the complete model, therefore it is important to seek a model that has a small value for MS_E . Backward elimination is a method of selecting the best subset of the "complete model" (model with all independent variables) through eliminating one by one the independent variables with the higher p -values in the T-test until all variables included in the model have p -value lower than the chosen level of significance. Another method is to select the independent variables sequentially, evaluating the importance of each dependent variable by identified the model which produces the lowest value of MS_E , called Forward selection. Bidirectional elimination is a combination of the two methods. (Gomes, 2011)

In multiple regression, one independent variable should not be completely determined by one or several other independent variables, but it is also possible to violate this condition through the presence of **multicollinearity**. It is important to identify and also solve this problem without compromising the quality of the model. Sometimes, a simple analysis of the correlation matrix is enough to realize the presence of multicollinearity and which independent variables are correlated. (Oliveira, 2009) Another method for evaluating the presence of multicollinearity is

the term $\frac{1}{1-R^2}$, referred to as a variance inflation factor (VIF). The square root of VIF explains how much larger the standard error is, compared with what it would be if that variables were uncorrelated. (Hsieh, 1998)

Another situation that can occur in multiple regression is **interaction** and **confounding**. Interaction occurs when the effect of one independent variable on the dependent variable is not the same for every values of another independent variable. For example, when studying the relation between different hospitals (dependent variable) and time form diagnosis to treatment, the CoAo severity may motivate different times. Confounding happens when adding an independent variable provides an alternative explanation for the relation between the dependent variable and another independent variable. For example, because in CoAo patients different ages require different treatments (Chapter 1.1.4), the patient's age changes the relation between arterial hypertension (dependent variable) and type of treatment. These problems are difficult to recognize and the solution might be to remove the variable causing the problem or to build separate regression models (Oliveira, 2009).

2.2 Logistic Regression

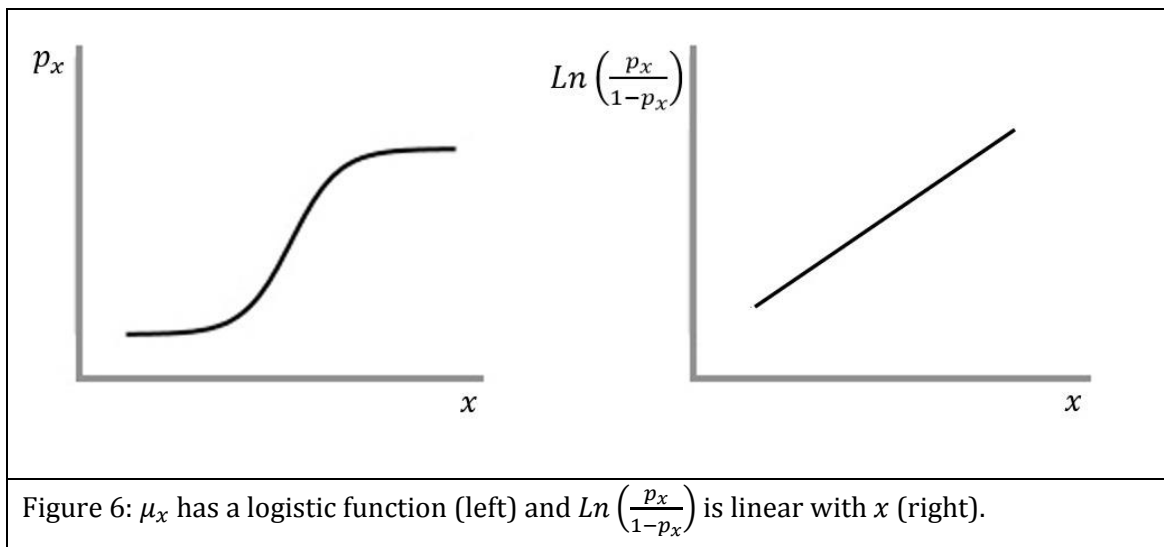
Binary regression models are a consequence of a categorized dependent variable. Logistic regression model are a particular case of binary regression models where the dependent variable assumes only the values 0 and 1.

2.2.1 Simple logistic regression

Considering a regression model where the dependent variable Y_x has a Bernoulli distribution, or $Y_x \cap B(p_x)$, the probability mass function is $P(Y_x = y) = p_x^y(1 - p_x)^{1-y}$, or $Y_x \begin{cases} 0 & 1 - p_x \\ 1 & p_x \end{cases}$. For each element of this population, the simple logistic model aims to find the constants β_0 and β_1 that resolve the formulas:

$$\mu_x = p_x = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad \text{and} \quad \text{Ln} \left(\frac{p_x}{1 - p_x} \right) = \beta_0 + \beta_1 x \quad (\text{canonic link}).$$

This means that p_x has a logistic function and the logarithm of the odds, $\text{Ln} \left(\frac{p_x}{1 - p_x} \right)$, is linear with x (Figure 6). (Gomes, 2011)



The maximum-likelihood estimation method, used in linear regression aims to find the estimators for β_0 and β_1 that maximize the log-likelihood function defined by the formula: $\text{Ln} L(\mathbf{y}, \mathbf{x}) = \sum l_i = \sum [y_i(\beta_0 + \beta_1 x_i) - \text{Ln}(1 + \exp(\beta_0 + \beta_1 x_i))]$. The estimators for β_0 and β_1 can be found by using the partial derivatives of the log-

likelihood function. Both of these methods provide the equation:

$$\begin{cases} \sum_{i=1}^n \frac{\partial l_i}{\partial \beta_0} = \sum_{i=1}^n (y_i - p_i) = 0 \\ \sum_{i=1}^n \frac{\partial l_i}{\partial \beta_1} = \sum_{i=1}^n x_i (y_i - p_i) = 0 \end{cases} . \text{ Or in the matrix form: } \mathbf{X}^T (\mathbf{Y} - \mathbf{p}) = 0 \text{ with } \mathbf{Y} \text{ and}$$

\mathbf{p} size $n \times 1$ matrixes and $\mathbf{X}^T = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix}$. These equations do not have an

analytical solution, therefore in is necessary to use an iterative method.

Considering a U function such as $U(\beta_0, \beta_1) = (\sum_{i=1}^n \frac{\partial l_i}{\partial \beta_0}, \sum_{i=1}^n \frac{\partial l_i}{\partial \beta_1})^T$, the **Newton-**

Raphson method can be applied. $[-\frac{\partial U}{\partial \beta}]$ is the partial derivative matrix of U and it

can be replaced by $\mathbf{X}^T \mathbf{W} \mathbf{X}$ with $\mathbf{W} = \text{Diag}(\mathbf{p}(1-\mathbf{p}))$. This iterative method aims to

find the values of β_0 and β_1 where the U function is close to zero. With $\boldsymbol{\beta} = (\beta_0, \beta_1)$,

this can be obtained through the formula:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [\mathbf{X}^T \mathbf{W} \mathbf{X}]_{\boldsymbol{\beta}^{(m)}}^{-1} [\mathbf{X}^T (\mathbf{Y} - \mathbf{p})]_{\boldsymbol{\beta}^{(m)}}^{-1}.$$

After finding the estimators for $\boldsymbol{\beta}$, the analysis the logistic regression model should

be performed. In logistic regression, the **Deviance** of a model is calculated by

comparing the simple regression model (M) to the perfectly adjusted model (M_s)

or to the null model (M_0) (Oliveira, 2009). Assuming L as the likelihood function,

the likelihood-ratio is calculated by $\lambda_1 = \frac{L(M)}{L(M_s)}$ and the measurement of the

Deviance of M to M_s is calculated by $G_1 = -2Ln(\lambda_1)$. In the same way, assuming

$\lambda_0 = \frac{L(M_0)}{L(M_s)}$, the measurement of the Deviance of M_0 to M_s is $G_0 = -2Ln(\lambda_0)$. A

higher Deviance translates into a larger difference between the two model that are

being compared. Because $\lambda_0 \cap \chi_{n-1}^2$ and $\lambda_1 \cap \chi_{n-2}^2$, the value for $G_0 - G_1$ can be

compared to the χ_1^2 quantil of chosen level of significance. This test is called the

likelihood-ratio test. (Gomes, 2011)

Other statistical tests can be used in evaluating the logistic regression model such

as the Wald test, the Score test, the Hosmer-Lemeshow test. (Gomes, 2011)

In the case of a continuous independent variable and because the canonic link is

the logarithm of the odds, it is possible to estimate the **Odds Ratio (OR)** for each

unit of the independent variable through $OR = \exp(\beta_1)$. (Oliveira, 2009)

The **Receiver Operating Characteristic (ROC)** curve is used to evaluate the capacity of the model to discriminate between “ $Y=1$ ” and “ $Y=0$ ” and it is obtained using the sensibility and specificity curves. By means of choosing a cutoff value c , $\hat{Y} = 1$ if the result for the odds is equal or higher than c . The sensibility of a model is the probability of correctly predicting “ $Y=1$ ”, calculated by $(\hat{Y} = 1|Y = 1) = \frac{P(\hat{Y}=1 \cap Y=1)}{P(Y=1)}$. The specificity of a model is the probability of correctly predicting “ $Y=0$ ”, calculated by $P(\hat{Y} = 0|Y = 0) = \frac{P(\hat{Y}=0 \cap Y=0)}{P(Y=0)}$. The probability of false negatives can be calculated by $1 - P(\hat{Y} = 1|Y = 1)$ and the probability of false positives can be calculated by $1 - P(\hat{Y} = 0|Y = 0)$. Different cutoff values for the model provide different values of sensibility and specificity. (Gomes, 2011)

The area under the ROC curve measures the percentage of the data that is correctly categorized, therefore reflecting the performance of the model. Authors suggest that a logistic regression model should have an area under the ROC curve above 80% for the data to have clinical interest. (Oliveira, 2009)

2.2.2 Multiple logistic regression

A regression model with the dependent variable Y where $Y \in B(p)$ and $Y \begin{cases} 0 & 1-p \\ 1 & p \end{cases}$, but instead of one independent variable there are p independent variables, $\mathbf{x} = (x_1, \dots, x_p)$, is a multiple logistic regression model. The mean value of Y conditional to \mathbf{x} is $E[Y_i|\mathbf{x}_i = (x_i^1, \dots, x_i^p)] = \mu_i = p_{x_i}$. Then, the formulas for the multiple logistic regression are:

$$p_x = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad \text{and} \quad \text{Ln} \left(\frac{p_x}{1-p_x} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

Similar to the simple logistic regression, the aim is to find the estimators for $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$ that maximize the log-likelihood function, in this case defined by: $\text{Ln} L(\mathbf{y}, \mathbf{x}) = \sum [y_i(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) - \text{Ln}(1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p))]$. This method also results in $\mathbf{X}^T(\mathbf{Y} - \mathbf{p}) = 0$ with \mathbf{Y} and \mathbf{p} size $n \times 1$ matrixes and \mathbf{X}^T a size $(p+1) \times n$ matrix, which means no analytical solution. Using the iterative Newton-Raphson method, the estimators for the coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$ can be

obtained through the same formula explained in the simple logistic regression:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [\mathbf{X}^T \mathbf{W} \mathbf{X}]_{\boldsymbol{\beta}^{(m)}}^{-1} [\mathbf{X}^T (\mathbf{Y} - \mathbf{p})]_{\boldsymbol{\beta}^{(m)}}^{-1}. \text{ (Gomes, 2011)}$$

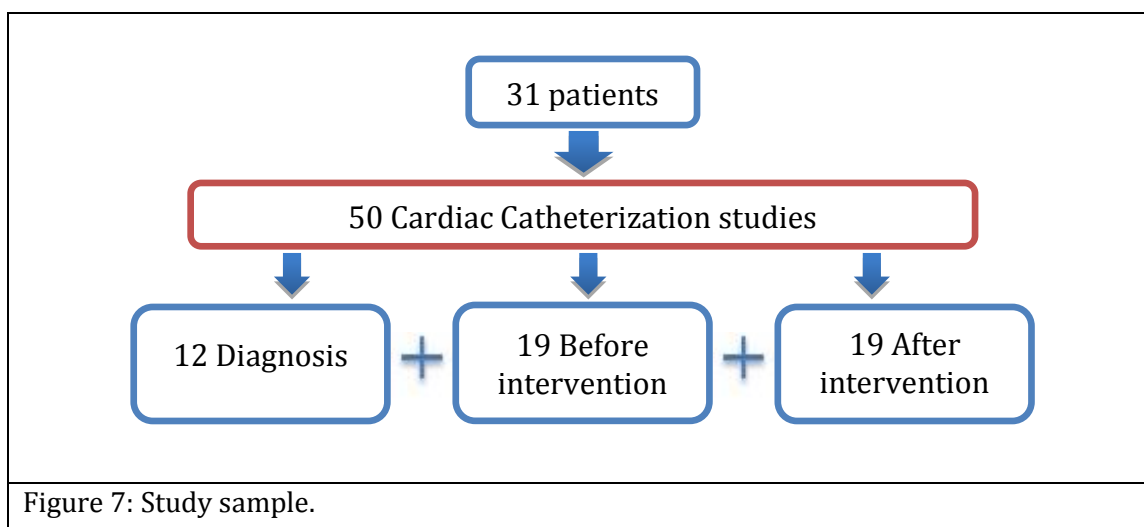
After finding the estimators for $\boldsymbol{\beta}$, evaluation of the model can be made through a value that represents a relative measurement of the information lost in a particular model, called **Akaike's Information Criterion (AIC)**. With k parameters and L the likelihood of the model, this value can be calculated by $AIC = -2[\ln(L) - k]$. The best model should have low AIC. (Gomes, 2011)

To select the best set of variables for the multiple logistic regression model, it is necessary to first understand how each independent variable behaves in relation to the dependent variable (for example, using plot analysis). Then, an analysis of the importance of each independent variable should be performed, using a quantitative values such as the Deviance and likelihood-ratio test (see Chapter 2.2.1) and the value of AIC. Similar to the multiple linear regression, the selection of the best set of variables can be performed through backward elimination (best subset of the complete model), forward selection (selecting independent variables sequentially) or bidirectional elimination (combination of backward elimination and forward selection). (Gomes, 2011)

In multiple logistic regression, ROC curve analysis can also be used to evaluate the performance of the selected model model. (Oliveira, 2009)

3 Methods and Results

Our sample included 31 patients with CoAo that underwent cardiac catheterization in Santa Cruz Hospital. Native CoAo was found in 23 patients and 8 patients underwent previous intervention. Overall mean age was 21.5 years, from patients between 4 and 62 years old, and 14 were women. In 19 patients, data was obtained from before and after intervention (balloon dilatation and/or stenting). Therefore, 50 cardiac catheterization studies were included, 12 with only diagnostic data, 19 with data before percutaneous intervention and 19 with data after percutaneous intervention (Figure 7).



Because there was no significant modification to the patient's aortic arch before or after cardiac catheterization, TTE and arterial stiffness studies performed within 45 days before or after were not excluded from the data. In the TTE studies associated to the diagnostic only cases and before intervention, 28 studies were performed the same day or the day before intervention and only 3 studies performed within 45 days before intervention. In the TTE studies associated to the data after the cardiac catheterization, 17 studies were performed 1 day after intervention and only 2 studies performed within 45 days after intervention. From the arterial stiffness studies associated to the diagnostic only cases and before intervention, 14 studies were performed the same day or the day before

intervention and only 2 studies were performed within 45 days before intervention. In the arterial stiffness studies associated to the data after cardiac catheterization, 11 studies were performed 1 day after intervention and only 1 study was performed within 45 days after intervention. In total there were 28 arterial stiffness studies performed and, because there are 50 cardiac catheterization studies, 22 patients had missing data for the arterial stiffness study.

The variables included in the data were obtained by cardiac catheterization, TTE and arterial stiffness studies (Table 2).

Method	Variables
Cardiac Catheterization	Sgrad - Systolic invasive gradient
TTE	Dgrad - Doppler corrected gradient DVT - Diastolic velocity at the end of T wave DVQ - Diastolic velocity at Q wave SHPTc - Systolic half pressure time DHPTc - Diastolic half pressure time VRc - Velocity runoff
Arterial Stiffness	PWV - Pulse wave velocity between right carotid and radial arteries

Table 2. Variables included in the study.

In the cardiac catheterization, systolic invasive gradients were measured (Sgrad) with an invasive catheter and a Transpac® disposable pressure transducer.

TTE studies were all performed with the Vivid7® ultrasound machine from General Electric®. TTE variables included were obtained by measuring the Doppler corrected gradient (Dgrad), diastolic velocity at end of T wave (DVT), diastolic velocity at Q wave (DVQ), systolic and diastolic half pressure times (SHPTc and DHPTc) and velocity runoff (VRc). Doppler gradients are obtained through the Bernoulli's formula where the gradient is equal to 4 times the squared

Doppler velocity, which reads $\Delta p = p_1 - p_2 = 4v^2$. Dgrad was obtained by subtracting the Doppler gradient before the CoAo site to the Doppler gradient after the CoAo site, measured in millimeters of mercury (mmHg). Both in meters per second (m/s), DVT as measured at the end of the electrocardiographic T wave, marking the beginning of the diastole, while DVQ was measured at the electrocardiographic Q wave, marking the end of the diastole. SHPTc was defined as the time for the maximum value of the Doppler gradient to decrease to half its value, while DHPTc was defined as the time for the Doppler gradient starting at the electrocardiographic T wave to decrease to half its value, both in milliseconds (ms). VRc was defined as the time for velocity to decrease from maximum value (V_{max}) to 33% V_{max} , in m/s. VRc, SHPTc and DHPTc were corrected with Bazett's formula, where the time is divided by the squared root of the RR interval, which is calculated through the heart rate, to resolve the problem that different heart rates produce different time measurements in each cardiac cycle.

Arterial stiffness studies were assessed by tonometry, with the Sphygmocor® equipment, and the variable included in the data was obtained by measuring pulsed wave velocity (PWV). PWV was measured between right carotid and radial arteries, in order to avoid any influence that the CoAo obstruction may have in the left radial artery and in the femoral arteries. The method used for measuring the PWV (Figure 8) consists on subtracting the time between the electrocardiographic Q wave and the start of the carotid pulse wave (t_1) to the time between the electrocardiographic Q wave and the start of the radial pulse wave (t_2), where $\Delta t = t_1 - t_2$. It is also necessary to subtract the distance between the sternal furcula and the location of the radial pulse (d_1) to the distance between the sternal furcula and the location of the carotid pulse (d_2), where $\Delta d = d_1 - d_2$. Then, PWV is a velocity calculated by the simple formula $\Delta t / \Delta d$, in m/s.

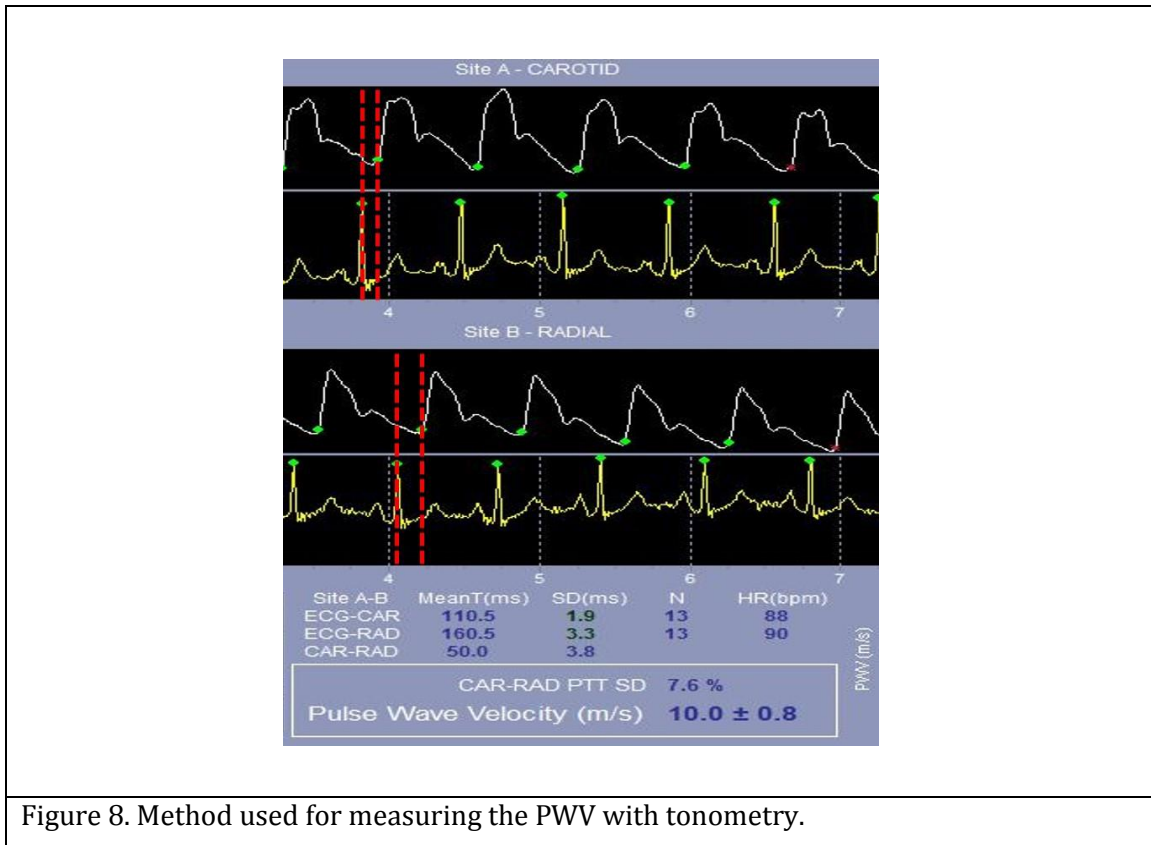


Figure 8. Method used for measuring the PWV with tonometry.

Overall, mean Sgrad was 23.5 mmHg, mean Dgrad was 32.2 mmHg, mean DVT was 1.0 m/s, mean DVQ was 0.3 m/s, mean SHPTc 101.5 ms, mean DHPTc 70.1 ms, mean VR 368.7 ms and mean PWV was 7.18 m/s.

To describe the relation between CoAo Doppler flow pattern, CoAo invasive gradients and arterial stiffness, the results were divided into results for invasive gradients and results for arterial stiffness. In the results for invasive gradients, only the relation between CoAo Doppler flow pattern and CoAo invasive gradients was studied. In the results for arterial stiffness, the relation between CoAo Doppler flow pattern and arterial stiffness was studied. All statistical analysis was performed using R version 3.1.2 software.

3.1 Invasive Gradients

Scatter plot analysis between Sgrad and the different TTE variables was performed (Figure 9). All plots suggest a positive relation with Sgrad. The variables Dgrad and DHPTc seemed to have the best linear relation with Sgrad. Dgrad seems to slightly overestimate Sgrad. The variables SHPTc, DVT and DVQ also showed a positive relation with Sgrad. The variable VRc showed a better linear relation for Sgrad below 30 mmHg.

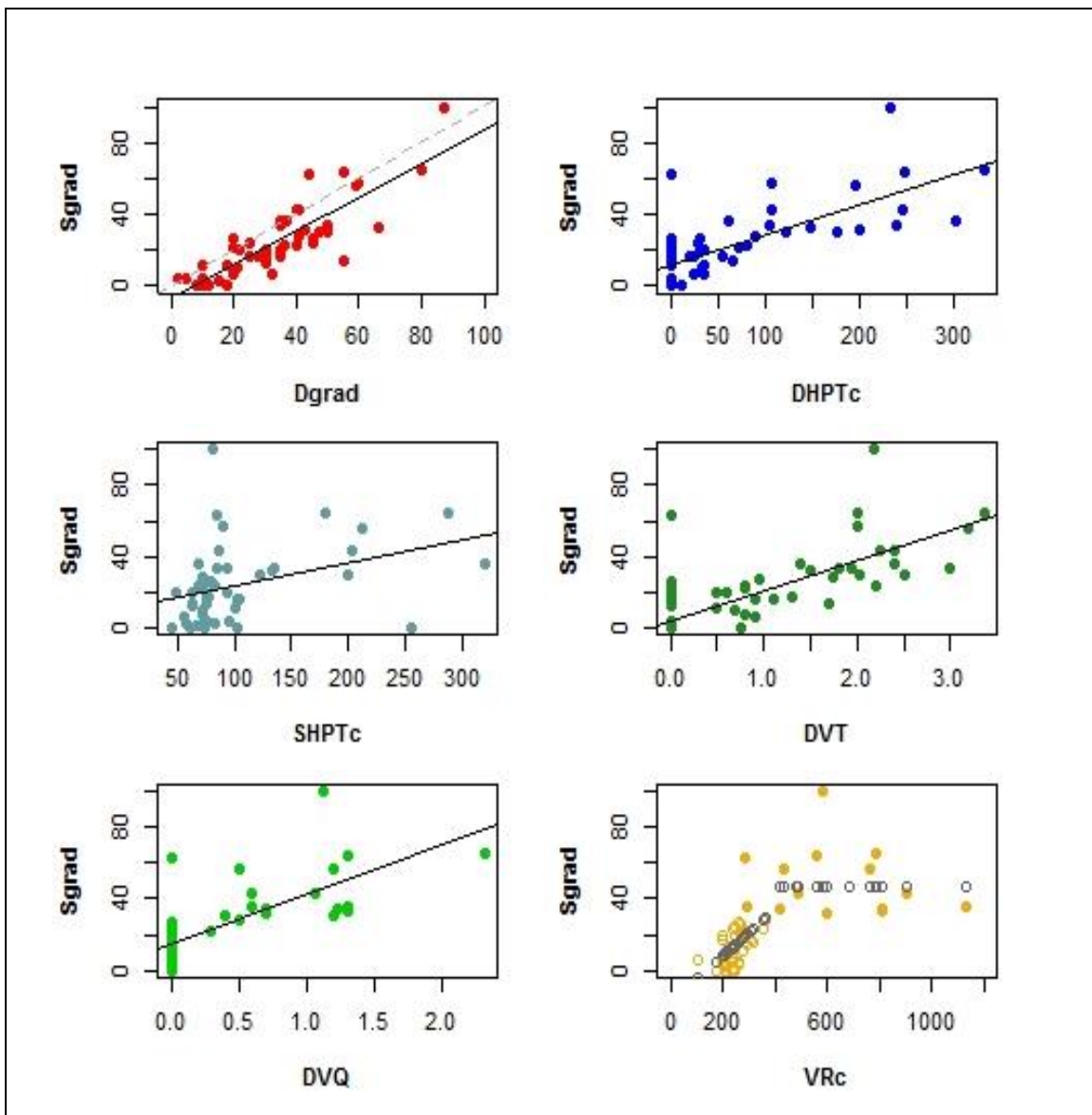


Figure 9: Scatter plot analysis between Sgrad and the different TTE variables. The black lines correspond to the regression line for the simple linear regression models. The black dots in the VRc plot represent the simple regression model for Sgrad above and below 30 mmHg.

Linear regression models were obtained with Sgrad as the dependent variable. Using the level of significance of 5%, a significant relation between variables was defined as a *p-value* for the statistical tests inferior to 0.05. Using simple linear regression models, all TTE variables showed a significant relation with Sgrad and a positive β_1 coefficient: the model using Dgrad had $\beta_1 = 0.93$ with *p-value* = 1.59×10^{-15} ; the model using DHPTc had $\beta_1 = 0.17$ with *p-value* = 2.62×10^{-9} ; the model using SHPTc had $\beta_1 = 0.13$ with *p-value* = 0.007; the model using DVT had $\beta_1 = 13.94$ with *p-value* = 7.49×10^{-8} ; the model using DVQ had $\beta_1 = 27.76$ with *p-value* = 6.39×10^{-9} ; and the model using VRc had $\beta_1 = 0.06$ with *p-value* = 1.65×10^{-6} .

Because all TTE variables showed a significant relation with Sgrad, the complete multiple linear regression model included all 6 variables. Backward elimination, forward selection and bidirectional elimination methods for selecting the best multiple regression model were used. All three methods provided the same 4 variables Dgrad, DHPTc, SHPTc and DVT with the formula $Sgrad = -0.10 + 0.82 \times Dgrad + 0.13 \times DHPTc - 0.05 \times SHPTc - 6.19 \times DVT$. Because SHPTc had a *p-value* of 0.18 for the *T*-test and also had a negative coefficient, contrary to what was observed in the simple linear regression model, SHPTc was removed from the multiple linear regression model. The multiple linear regression model that included the 3 variables (Dgrad, DHPTc and DVT had negative coefficient for the variable DVT, opposing to what was observed in the simple linear regression model. Therefore, the variable DVQ was also removed from the model. The best multiple regression model (Figure 10) included the 2 independent variables Dgrad and DHPTc and provided the formula:

$$\mathbf{Sgrad = -4.61 + 0.75 \times Dgrad + 0.06 \times DHPTc.}$$

This model has a test statistic *F* equal to 76.9 and with a *p-value* of 1.51×10^{-15} . Both coefficients had a *p-value* inferior to 0.05 for the *T*-test, were the coefficient for the variable Dgrad had *p-value* = 9.87×10^{-9} and the coefficient for the variable DHPTc had *p-value* = 0.02. This model's determination coefficient (R^2) was equal to 0.766 and the adjusted R^2 was equal to 0.756, meaning that the model provides a good fit for the data.

```

Call:
lm(formula = Sgrad ~ Dgrad + DHPTc, data = coao)

Residuals:
    Min       1Q   Median       3Q      Max
-26.271  -4.764  -1.625   5.348  34.567

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.60526     3.00362  -1.533   0.1319
Dgrad         0.75087     0.10810   6.946 9.87e-09 ***
DHPTc         0.05505     0.02284   2.410  0.0199 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.4 on 47 degrees of freedom
Multiple R-squared:  0.7659,    Adjusted R-squared:  0.756
F-statistic:  76.9 on 2 and 47 DF,  p-value: 1.513e-15

```

Figure 10: Best multiple regression model for the dependent variable Sgrad, obtained with R software.

Multicollinearity of the model was studied through the analysis of the correlation matrix and calculation of the VIF of the model (Figure 11). The square root of VIF explains that the standard error is 1.4 times larger than what it would be if the variables were uncorrelated, thus the multicollinearity of the model was considered not to be significant and no variable was removed from the model.

```

> cor(XX[,-1])
           Dgrad    DHPTc    Y
Dgrad 1.0000000 0.7037733 0.8584835
DHPTc 0.7037733 1.0000000 0.7250148
Y      0.8584835 0.7250148 1.0000000
> vif(model)
           Dgrad    DHPTc
1.981362 1.981362
> sqrt(1.98)
[1] 1.407125

```

Figure 11: Multicollinearity studied by correlation matrix and VIF.

By scatter plot analysis, the residuals of the model seem to be unbiased but heteroscedastic (Figure 12). The Normal Q-Q plot seems to demonstrate heavy

tailed residuals, but the Kolmogorov-Smirnov normality test for the residuals had a $p\text{-value} = 0.46$ which means that, for a level of significance of 5%, the residuals of the model were considered to have a Normal distribution.

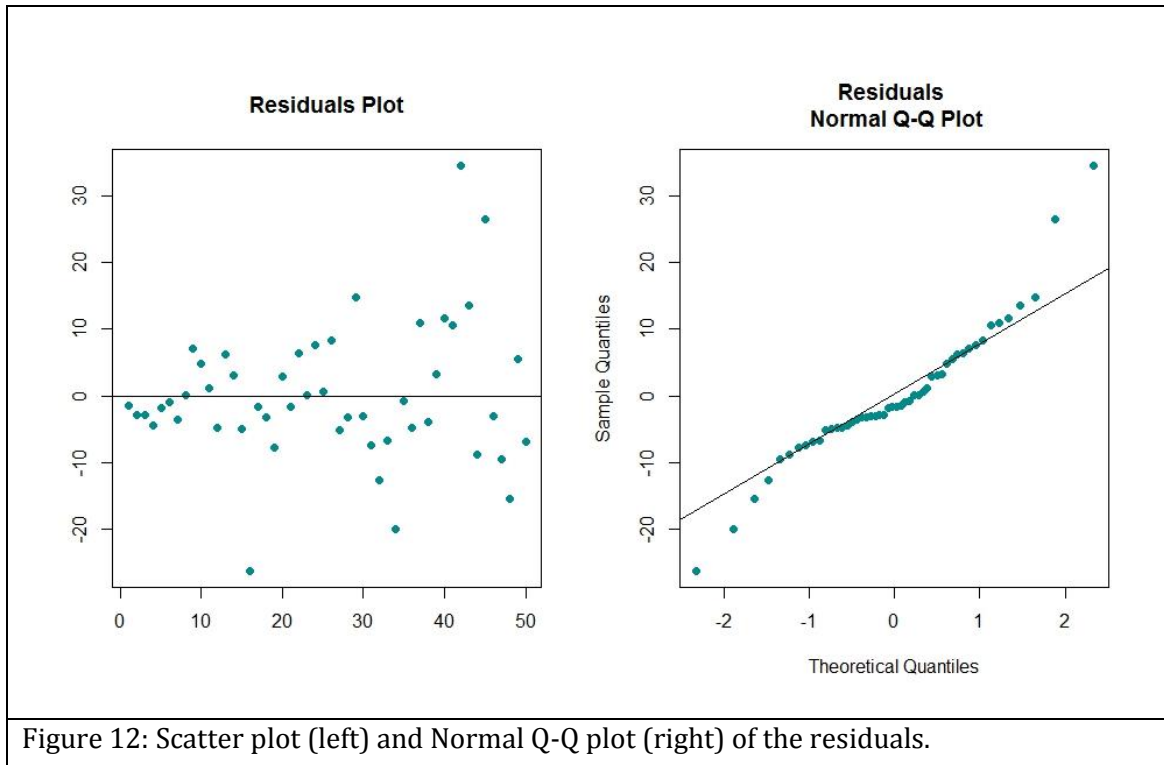


Figure 12: Scatter plot (left) and Normal Q-Q plot (right) of the residuals.

Because of the apparent heteroscedastic residuals and the heavy tailed Normal Q-Q plot of the residuals, multiple generalized regression models were obtained but with no improvement of the model's R^2 or of the residual's distribution.

Logistic regression models were obtained by transforming the dependent variable Sgrad into a binary variable named Sign. This new variable represents the presence or absence of a significant invasive gradient, where $\text{Sign} = 0$ if Sgrad is inferior to 20 mmHg and $\text{Sign} = 1$ if Sgrad is superior or equal to 20 mmHg, such as:

$$\text{Sign} \begin{cases} 0 & 1 - p_{\text{Sign}} \\ 1 & p_{\text{Sign}} \end{cases}$$

From the whole sample, 23 patients had $\text{Sign} = 0$ and 27 patients had $\text{Sign} = 1$. Scatter plot analysis between Sign and the different TTE variables was performed (Figure 13). All plots suggest that higher values for all TTE variables are more

likely to be associated with Sign = 1. A significant result was found by noticing that all patients with DVQ superior to zero had Sign = 1.

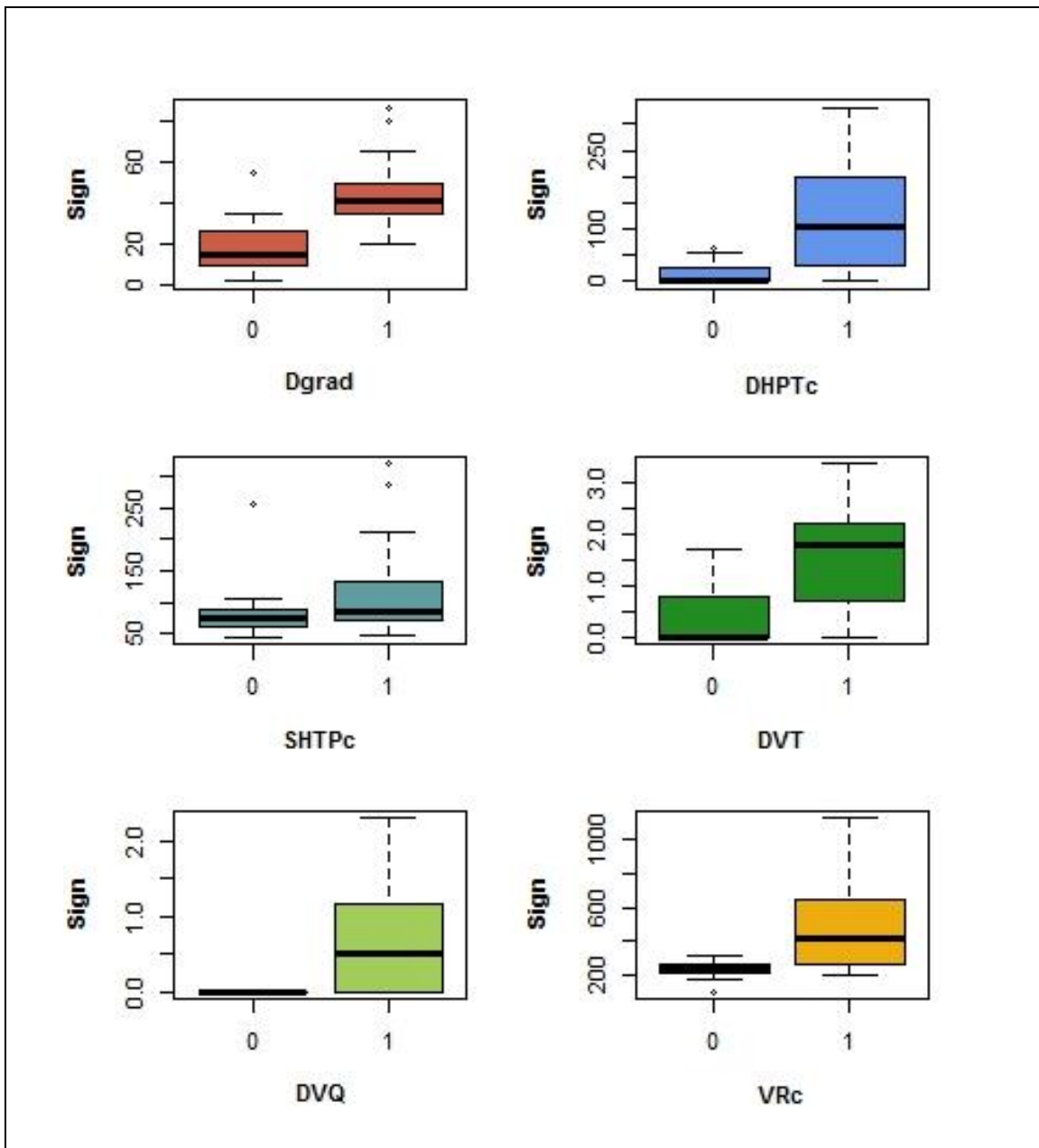


Figure 13: Scatter plot analysis between Sign and the different TTE variables. Notice that all patients with DVQ superior to zero had Sign = 1.

Simple logistic models were obtained with each TTE variable as the independent variable. Using the level of significance of 5%, a significant relation between variables was defined as a *p-value* for the likelihood-ratio test inferior to 0.05. With Sign as the dependent variable of the simple logistic model: the model using Dgrad

had $\beta_1 = 0.14$ with $p\text{-value} = 0.0001$; the model using DHPTc had $\beta_1 = 0.003$ with $p\text{-value} = 1.14 \times 10^{-5}$; the model using SHPTc had $\beta_1 = 0.002$ with $p\text{-value} = 0.04$; the model using DVT had $\beta_1 = 0.29$ with $p\text{-value} = 0.003$; the model using DVQ had $\beta_1 = 0.52$ with $p\text{-value} = 2.06 \times 10^{-5}$; and the model using VRc had $\beta_1 = 0.001$ with $p\text{-value} = 3.23 \times 10^{-5}$. All TTE variables and a positive β_1 coefficient and all $p\text{-values}$ for the likelihood-ratio test were inferior to 0.05 which means that higher values of the TTE variables Dgrad, DHPTc, SHPTc, DVT, DVQ and VRc are more likely to be associated with Sign = 1.

Due to the significant relation with Sign in the simple logistic regression, the complete multiple logistic regression model included the variables Dgrad, DHPTc, SHPTc, DVT and VRc. The variable DVQ had to be excluded from the multiple regression model because fitted probabilities numerically 0 or 1 occurred due to the fact that all patients with DVQ>0 had Sign=1. Backward elimination, forward selection and bidirectional elimination methods for selecting the best multiple regression model were also used. All three methods provided the same 3 variables Dgrad, DHPTc and DVT with the formula for the logarithm of the odds ratio: $Ln\left(\frac{p_{Sign}}{1-p_{Sign}}\right) = -3.88 + 0.12 \times \text{Dgrad} + 0.11 \times \text{DHPTc} - 4.24 \times \text{DVT}$. Because the coefficient for the variable DVT was negative and opposing to what was observed in the simple linear regression model, this variable was removed from the model. After removing DVT the $p\text{-value}$ of the likelihood-ratio test for the coefficient DHPTc became not significant. Instead of removing the variable DHPTc, AIC of 3 logistic regression models was compared (Figure 14). The model with only Dgrad as the independent variable had an AIC of 42.32. The model with Dgrad and DHPTc had an AIC of 40.45. The model that includes the independent variables Dgrad, DHPTc and their interaction has an AIC of 40.33, which means that this model provides more information. The model that included the interaction between Dgrad and DHPTc was considered the best model (Figure 14) and provided the formula:

$$Ln\left(\frac{p_{Sign}}{1-p_{Sign}}\right) = -4.70 + 0.12 \times \text{Dgrad} + 0.06 \times \text{DHPTc} - 0.0008 \times \text{Dgrad} \times \text{DHPTc} .$$

The coefficients had a *p-value* inferior to 0.05 for the likelihood-ratio test, were the coefficient for the variable Dgrad had *p-value* = 0.006, the coefficient for the variable DHPTc had *p-value* =0.046 and the coefficient for the interaction between Dgrad and DHPTc had *p-value* =0.059.

```

> m1=glm(Sign~Dgrad,binomial,coao)
> m2=glm(Sign~Dgrad+DHPTc,binomial,coao)
> m3=glm(Sign~Dgrad*DHPTc,binomial,coao)
> AIC(m1,m2,m3)
  df    AIC
m1  2 42.32566
m2  3 40.45131
m3  4 40.33317
> # The model with inferior AIC includes the interaction
> summary(m3)

Call:
glm(formula = Sign ~ Dgrad * DHPTc, family = binomial, data = coao)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.53834  -0.33706   0.06113   0.36832   1.54475

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.6981514   1.5225430  -3.086  0.00203 **
Dgrad         0.1299758   0.0469618   2.768  0.00565 **
DHPTc         0.0557919   0.0280051   1.992  0.04635 *
Dgrad:DHPTc  -0.0008101   0.0004290  -1.888  0.05900 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 68.994  on 49  degrees of freedom
Residual deviance: 32.333  on 46  degrees of freedom
AIC: 40.333

```

Figure 14: Best logistic multiple regression model for odds ratio of the dependent variable Sign, obtained with R software.

For further evaluation of the logistic regression model, the ROC curve (Figure 16) was obtained through the sensibility and specificity curves (Figure 15). The area under the ROC curve was 93.4%, reflecting the percentage of the data that was correctly predicted by this model.

Because it is more important no to have false negatives for a significant invasive gradient (Sign = 1) the best cutoff value choice for $Ln\left(\frac{p_{Sign}}{1-p_{Sign}}\right)$ of this logistic model is 0.34, resulting in a sensibility of 96% and specificity of 74% (Figure 17), which means that this model has only 4% of false negatives and 26% of false positives. When comparing the ROC curve of the selected model to the model that included only the independent variable Dgrad, the area under the ROC curve for the simple regression model was lower (91.5%), with a best cutoff value choice of 0.26 resulting in a lower sensibility (93%) and lower specificity (65%).

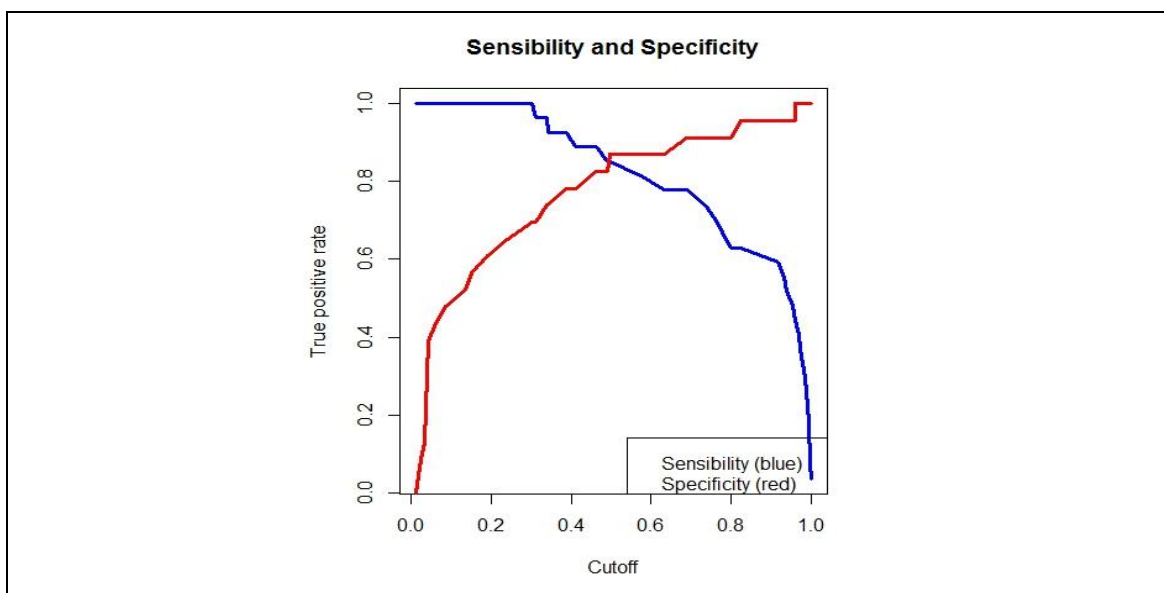


Figure 15: Sensibility and specificity curves of the multiple logistic model.

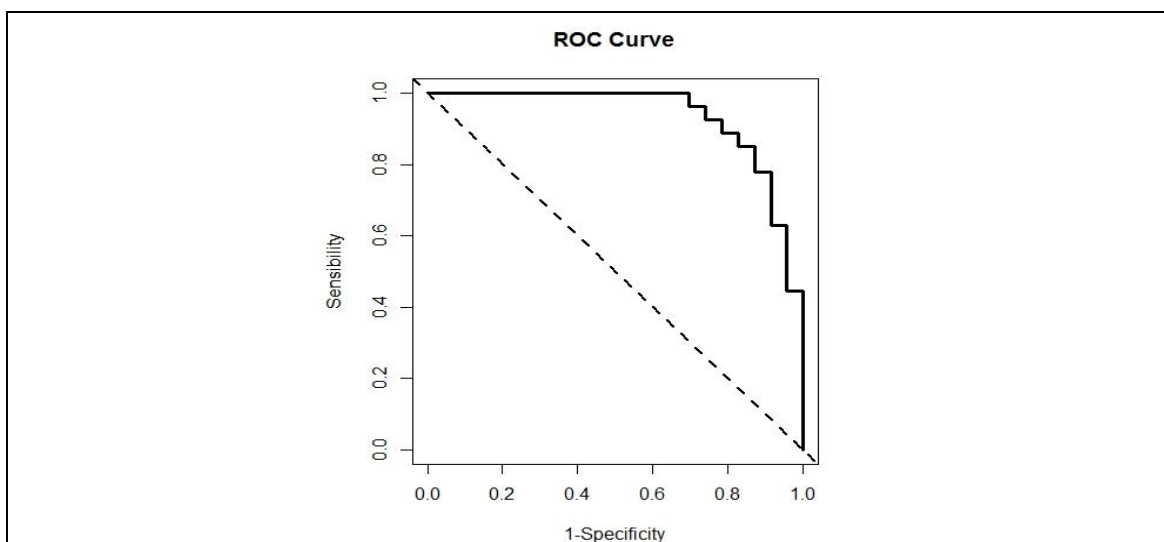


Figure 16: ROC curve of the multiple logistic model.

	Inf	0.00000000	1.00000000
0.99964171	0.03703704	1.00000000	
0.99835560	0.07407407	1.00000000	
0.99789717	0.11111111	1.00000000	
0.99471233	0.14814815	1.00000000	
0.99383256	0.18518519	1.00000000	
0.98941398	0.22222222	1.00000000	
0.98891580	0.25925926	1.00000000	
0.98557620	0.29629630	1.00000000	
0.97969634	0.33333333	1.00000000	
0.97314021	0.37037037	1.00000000	
0.96835516	0.40740741	1.00000000	
0.96036557	0.44444444	1.00000000	
0.96010786	0.44444444	0.95652174	
0.95376818	0.48148148	0.95652174	
0.93740510	0.51851852	0.95652174	
0.93340829	0.55555556	0.95652174	
0.91646702	0.59259259	0.95652174	
0.82522720	0.62962963	0.95652174	
0.80019965	0.62962963	0.91304348	
0.75964981	0.70370370	0.91304348	
0.73512543	0.74074074	0.91304348	
0.68813619	0.77777778	0.91304348	
0.63096110	0.77777778	0.86956522	
0.57523176	0.81481481	0.86956522	
0.49524469	0.85185185	0.86956522	
0.48921496	0.85185185	0.82608696	
0.46281935	0.88888889	0.82608696	
0.41010282	0.88888889	0.78260870	
0.38849181	0.92592593	0.78260870	
0.34295582	0.92592593	0.73913043	
0.34037442	0.96296296	0.73913043	
0.31026585	0.96296296	0.69565217	
0.30327275	1.00000000	0.69565217	
0.24076686	1.00000000	0.65217391	

Figure 17: Sensibility and specificity values for choosing the cutoff of the model.

3.2 Arterial Stiffness

Previous studies with *in-vitro* (Tacy et al., 1999) and computed models (DeGroff et al., 2003) have suggested that the diastolic tail increases with decreasing arterial stiffness, in CoAo with the same degree of stenosis. The aim is to find out if and in what way arterial stiffness is influencing the diagnosis of CoAo severity. Therefore, the TTE variables included for this subchapter are the ones that were included in the multiple linear and logistic regression models seen in the subchapter 3.1, which are Dgrad and DHPTc. The variable Dgrad suggests CoAo severity because it estimates the invasive gradient, but it does not measure the diastolic tail. The variable DHPTc can be used for demonstrating diastolic tail due to the fact that patients with DHPTc equal to zero have no diastolic tail and, for the patients that have diastolic tail, higher DHPTc means more significant diastolic tail. Because the *in-vitro* and computed model studies used only CoAo with the same degree of stenosis, the variable Sgrad was used in this subchapter for selecting groups of patients with similar CoAo severity.

The variable PWV determines the patients that have stiffer arteries (lower PWV) and the patients that have more compliant arteries (higher PWV).

Scatter plots between Sgrad and the TTE variables, Dgrad and DHPTc, were obtained (Figure 18). Patients with lower PWV were represented by green dots and patients with higher PWV were represented by red dots. Because the mean value for PWV in our sample was 7.18 m/s, patients with lower PWV were defined as the patients with PWV below or equal to 7.18 m/s, and patients with higher PWV were defined as the patients with PWV above 7.18 m/s. In the scatter plot between Sgrad and Dgrad, no difference was found between patients with higher and lower PWV. In the scatter plot between Sgrad and DHPTc, patients with higher PWV seem to have lower values of DHPTc, but only in the group of patients with Sgrad below 30 mmHg.

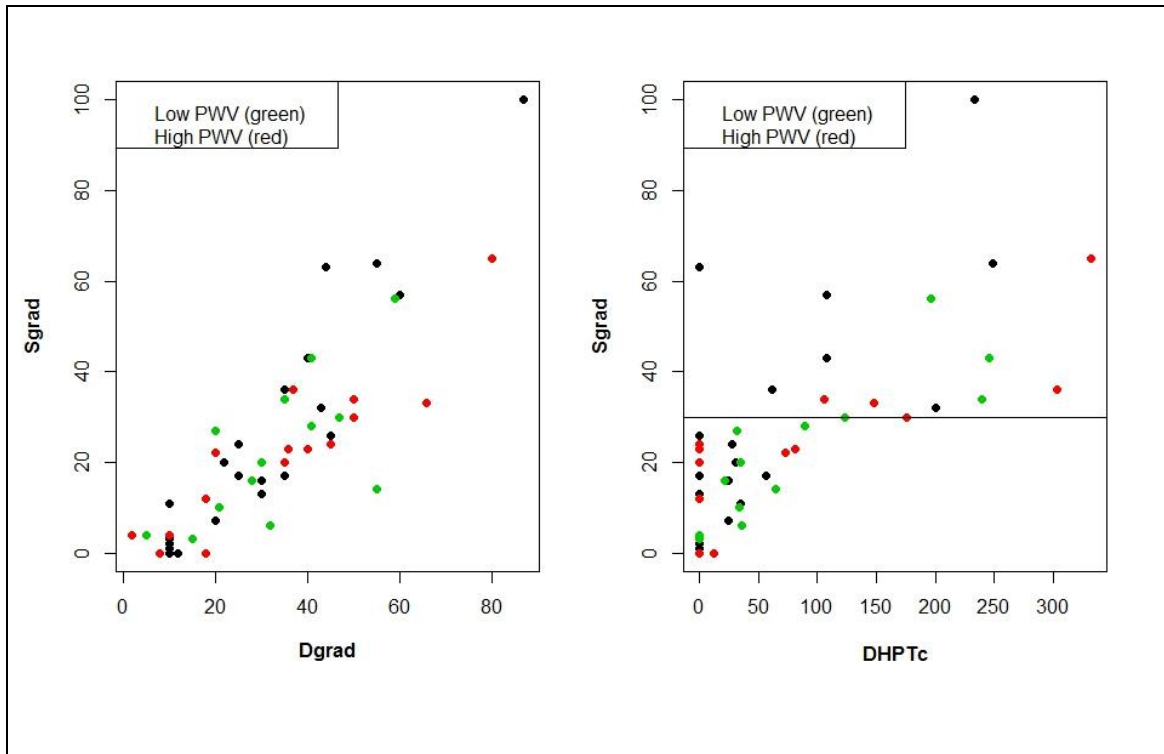


Figure 18: Scatter plots between Sgrad and the TTE variables, Dgrad and DHPTc. In the group of patients with Sgrad below 30 mmHg (horizontal line in the left plot), patients with higher PWV seem to have lower values of DHPTc. The black dots are from patients without measurement of PWV (missing data).

Using simple linear regression, no significant relation was found between PWV and the variables Sgrad, Dgrad or DHPTc in the 28 patients with PWV measurement (22 patients had missing data for PWV). Using the level of significance of 5%, a significant relation between variables was defined as a p -value for the F -test inferior to 0.05. None of these simple linear regression models had a significant relation between the two variables: the model between Sgrad and PWV had p -value = 0.80; the model between Dgrad and PWV had p -value = 0.44; and the model between DHPTc and PWV had p -value = 0.27.

Since the *in-vitro* and computed model studies used only CoAo with the same degree of stenosis, PWV was studied in groups with similar Sgrad. Because the scatter plot between Sgrad and DHPTc showed that patients with higher PWV seem to have lower values of DHPTc when Sgrad is below 30 mmHg, the group of patients with Sgrad inferior to 30 mmHg was studied (total of 19 patients). The scatter plot between Sgrad and DHPTc in this group was obtained (Figure 19), with

the same green and red codification for patients with lower and higher PWV used in Figure 17.

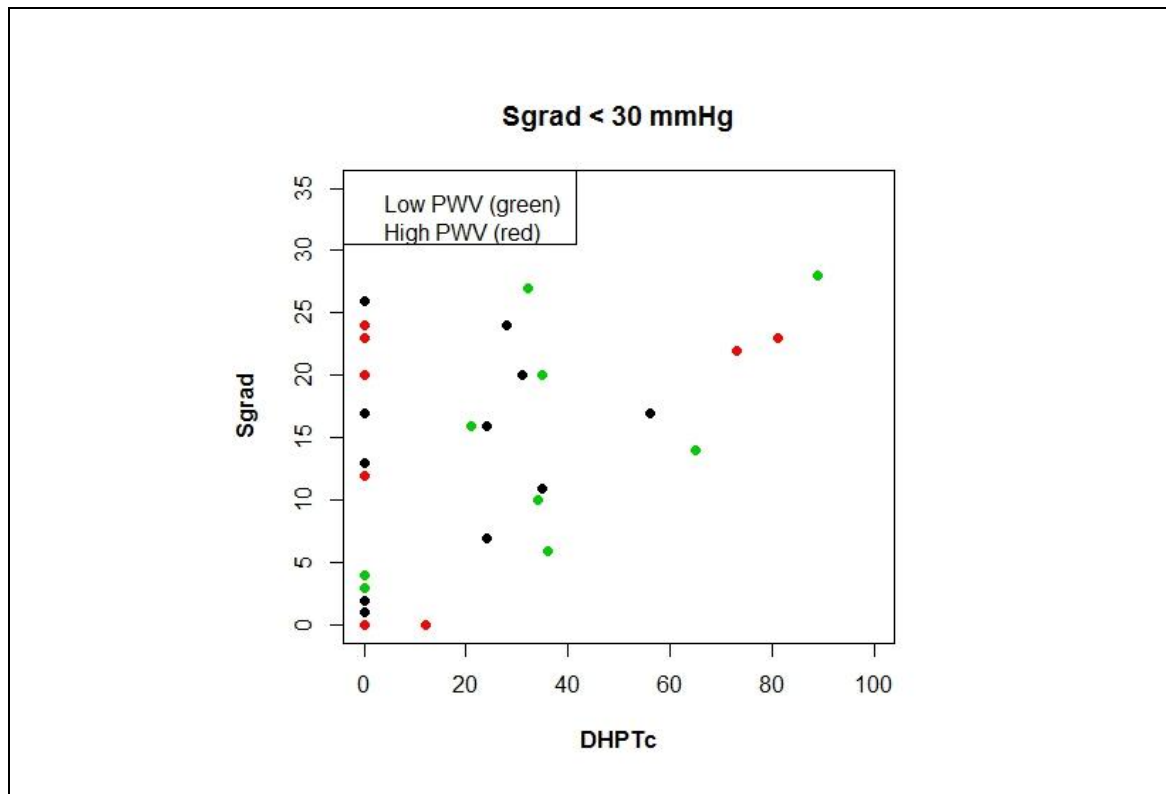


Figure 19: Scatter plot between Sgrad DHPTc, in the group of patients with Sgrad bellow 30 mmHg. Patients with higher PWV seem to have lower values of DHPTc. The black dots are from patients without measurement of PWV (missing data).

Because DHPTc was equal to zero in most patients with higher PWV, the variable DHPTc was turned into a binary variable named diastolic tail (DTail). DHPTc higher than zero means that there is a diastolic tail in the Doppler flow pattern. Therefore, DTail was equal to 1 if DHPTc was higher than zero and equal to 0 if DHPTc was also equal to zero. The boxplot for the relation between PWV and this variable was obtained (Figure 20).

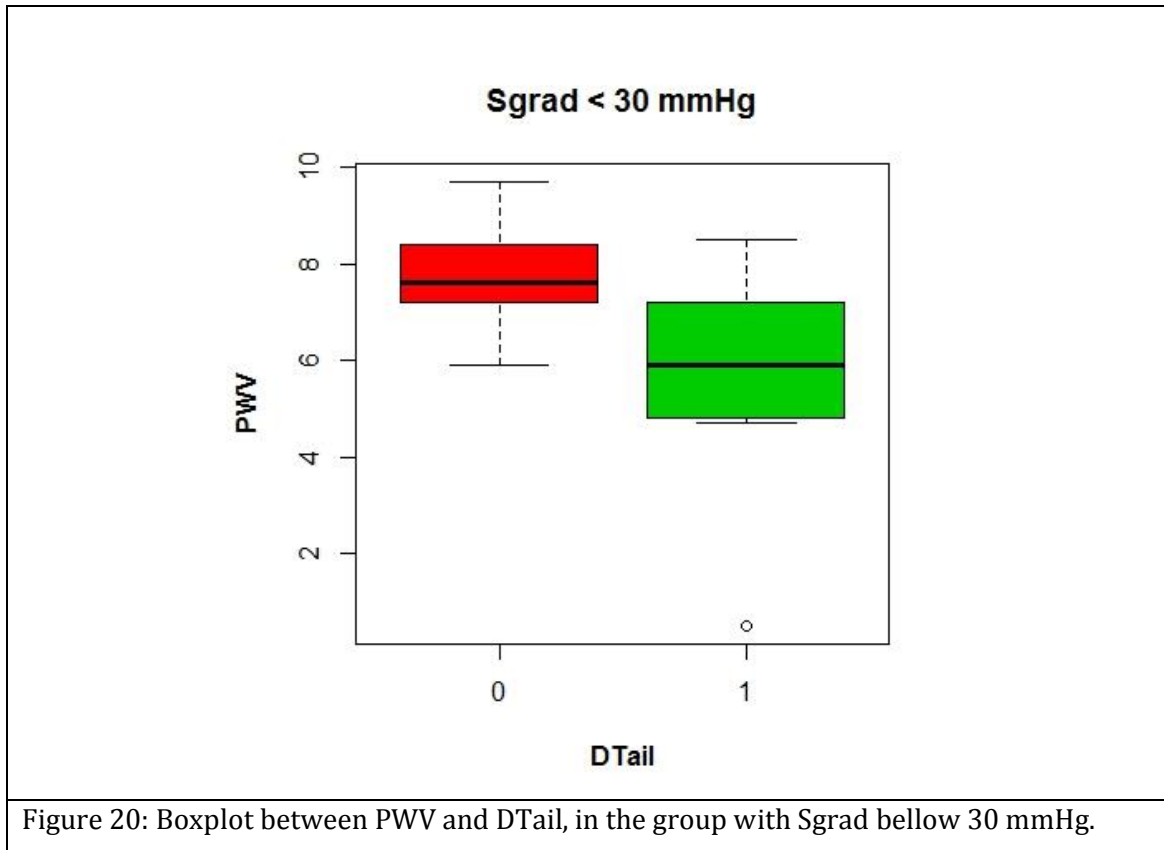


Figure 20: Boxplot between PWV and DTail, in the group with Sgrad bellow 30 mmHg.

The simple logistic regression model with DTail as the independent variable and PWV as the independent variable was obtained (Figure 21) and provided the formula:

$$\ln\left(\frac{p_{DTail}}{1-p_{DTail}}\right) = 6.69 - 0.94 \times PWV .$$

Using the level of significance of 5%, a significant relation was defined as a *p-value* for the likelihood-ratio test inferior to 0.05. The model had a negative coefficient β_1 and a *p-value* = 0.0495 for the likelihood-ratio test, which means that lower values of PWV are more likely to be associated with DTail = 1, in the group of patients with Sgrad bellow 30 mmHg.

The OR of this model (Figure 20) explained that if PWV is 1 m/s lower, patients have a 2.57 times higher chance of DTail = 1, in the group of patients with Sgrad bellow 30 mmHg.

```

> m=glm(DTail~PWV,binomial,base1);summary(m)

Call:
glm(formula = DTail ~ PWV, family = binomial, data = base1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.67813  -0.86487   0.06307   0.75320   1.76632

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.6907     3.4650   1.931  0.0535 .
PWV          -0.9429     0.4800  -1.965  0.0495 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 26.287  on 18  degrees of freedom
Residual deviance: 19.514  on 17  degrees of freedom
AIC: 23.514

Number of Fisher Scoring iterations: 5

> 1/exp(coef(m)[2])
      PWV
2.567437

```

Figure 21: Simple logistic regression model with DTail as the independent variable and PWV as the independent variable (top). OR of DTail for each unit of PWV (bottom).

4 Discussion

Using scatter plot analysis and simple regression models, it was possible to identify a relation between CoAo invasive gradients, or Sgrad, and all individual TTE variables: Dgrad, DHPTc, SHPTc, DVT, DVQ and VRc.

Because Dgrad aims to estimate the value of Sgrad through echocardiography, a significant relation between these two variables was expected. But Dgrad does not estimate precisely Sgrad, most of the times it overestimates Sgrad by about 10 to 20 mmHg. Thus, it was important to study the relation between Sgrad and other TTE variables in order to find if they may resolve the inaccuracy of Dgrad in CoAo patients.

SHPTc and DHPTc reflect the presence of diastolic tail in the CoAo Doppler flow pattern. A higher SHPTc means that the peak systolic pressure takes a longer time to decrease, so blood flow during diastole is more likely. DHPTc is easier to understand. DHPTc can assume the value zero because in a normal aortic arch there is no diastolic tail. In a CoAo patient, a longer time for the peak diastolic pressure to decrease (higher DHPTc) reflects a more significant diastolic tail. The significant relation found between Sgrad and both SHPTc and DHPTc suggested that diastolic tail is related to CoAo invasive gradients. This finding is in accordance to what was described by Carvalho (1990), but with CoAo severity estimated using invasive gradients instead of diameter ratios.

DVT and DVQ also reflect the presence of a diastolic tail. A higher the DVT means that the diastolic tail is more significant. An observation of DVQ above zero means that diastolic tail is present even at end diastole. The significant relation found between Sgrad and both DVT and DVQ also suggested that diastolic tail is related to CoAo invasive gradients. This finding is in accordance to what was described by Tan (2005), but also using invasive gradients instead of diameter ratios.

VRc was the variable used in *in-vitro* and computer model studies to describe the presence of a diastolic tail. Because VRc measures the time for the maximum velocity to decrease to 33% its value, higher VRc suggests a more significant diastolic tail. Significant relation between VRc and Sgrad was found using simple

linear regression. This finding also supports that diastolic tail is related to CoAo invasive gradients.

Using multiple linear regression, the best model was represented by the formula $Sgrad = -4.61 + 0.75 \times Dgrad + 0.06 \times DHPTc$. This showed that the variables that should be used to predict Sgrad were Dgrad and DHPTc. This model had a better fit for the data in comparison to the model using only Dgrad ($R^2 = 0.77$ including Dgrad and DHPTc versus $R^2 = 0.74$ including only Dgrad). Thus, DHPTc seems to resolve part of the inaccuracy of Dgrad in predicting Sgrad.

The limitation of this model was that residuals seemed to be unbiased but heteroscedastic. Multiple generalized regression models did not improve the residual's distribution.

A logistic model that predicts the odds of a significant invasive gradient may be useful in selecting the patients that would benefit from treatment. The variable Sign was built with Sign = 0 if Sgrad < 20 mmHg and Sign = 1 if Sgrad ≥ 20 mmHg. The use of a cutoff value of 20 mmHg was due to the fact that surgical or percutaneous treatment of CoAo is usually performed if Sgrad is equal or above 20 mmHg.

Sign showed a significant relation with all TTE variables using simple logistic regression. Dgrad, DHPTc, SHPTc, DVT, DVQ and VRc are more likely to be associated with Sign = 1. The first significant result was found by noticing that all patients with DVQ superior to zero had Sign = 1. This means that if there is persistence of blood flow at the end of diastole, Sgrad is always above 20 mmHg and treatment should be considered. For the rest of the TTE variables, these findings are clarified in the same way explained in their relation with Sgrad.

Using multiple logistic regression, the best model was represented by the formula

$$\text{Ln} \left(\frac{p_{\text{Sign}}}{1-p_{\text{Sign}}} \right) = -4.70 + 0.12 \times Dgrad + 0.06 \times DHPTc - 0.0008 \times Dgrad \times DHPTc.$$

The variables included for the multiple logistic model are the same included in the multiple linear model. But in the logistic case, the best model included the interaction between Dgrad and DHPTc. The interaction term means that the effect of one independent variable on Sign is different for different values of the other independent variable. In this model, although the odds for Sign increases with both

independent variables, the effect of Dgrad was higher if DHPTc was lower, as well as the effect of DHPTc was higher if Dgrad was lower. When we look at the Doppler flow pattern, this finding makes sense. When Dgrad is higher, the diastolic slope that follows tends to be sharper, thus the time for the diastolic pressure to decrease becomes slightly shorter, which translates to a lower DHPTc value. In the same way, a lower value of Dgrad tends to have a diastolic slope a little less sharp, therefore the time for the diastolic pressure to decrease becomes slightly longer, thus DHPTc is higher.

The multiple logistic model had an area under the ROC curve of 93.4%, reflecting most of the data is correctly predicted by this model. The best cutoff value choice for $Ln\left(\frac{p_{Sign}}{1-p_{Sign}}\right)$ was 0.34, resulting in a sensibility of 96% and specificity of 74%.

With this cutoff value, only 4% of patients that need treatment are not identified (false negatives) with the cost that 26% of patients are incorrectly identified with a significant invasive gradient (false positives). The consequence of an untreated CoAo has a substantial impact on the patient's health as it often evolves into arterial hypertension. A higher number of false positives has only the consequence of submitting patients to angiography, an invasive diagnostic technique with low likelihood of complications. Therefore, when estimating CoAo severity by TTE, it is more beneficial for the multiple logistic model to identify almost all patients that need treatment.

When comparing the ROC curve of the selected model to the model that included only the independent variable Dgrad, a lower area under the ROC curve was found for the simple regression model (91.5%), also suggesting that DHPTc seems to resolve part of the inaccuracy of Dgrad in predicting significant Sgrad (Sign).

By simulating the same CoAo severity, previous *in-vitro* and computed model studies suggested that diastolic tail increases with decreasing arterial stiffness. The aim was to find if and in what way arterial stiffness influences the models obtained earlier to predict Sgrad.

As expected, no significant relation was found between PWV and the variables Sgrad, Dgrad or DHPTc in the 28 patients with PWV measurement. Absence of a significant relation between PWV and both Sgrad and Dgrad showed that higher

arterial stiffness does not mean higher degrees of CoAo severity. The absence of relation between PWV and DHPTc showed that, when all degrees of CoAo severity are included, PWV does not have a significant influence on diastolic tail. Therefore, the next step was to choose a group of patients with the same degree of CoAo severity.

In the group of patients with Sgrad bellow 30 mmHg (19 patients), scatter plot reflected that higher PWV seemed to have lower values of DHPTc. Therefore, representing patients with mild and moderate CoAo severity, this became the group for studying how PWV may influence DHPTc. But in this group of patients with Sgrad bellow 30 mmHg, higher PWV had a lot of DHPTc values equal to zero, hence the need for the variable DTail.

The simple logistic model with the formula $Ln\left(\frac{p_{DTail}}{1-p_{DTail}}\right) = 6.69 - 0.94 \times PWV$ explained that PWV 1 m/s lower had a 2.57 times higher chance of DTail = 1, in the group of patients with Sgrad bellow 30 mmHg. This finding is in agreement with previous *in-vitro* and computed model studies. A compliant artery increases in diameter when the pulse wave passes through. After the pulse wave has passed, the artery recoils. In the presence of a CoAo, the recoil in the begging of diastole may be responsible for the late blood flow that results in the diastolic tail. On the other hand, a stiff artery does not have a physiological increase in diameter when the pulse wave passes and no significant recoil in diastole. Therefore, the absence of a diastolic tail in a significant CoAo may be due to an increased arterial stiffness. The significant relation between PWV and DTail suggests that low arterial stiffness may contribute for the persistence of flow in diastole in CoAo, but only in patients with mild or moderate stenosis. In severe stenosis, a significant obstacle may incite a diastolic tail, whether the patient has high or low PWV.

This result was obtained from few data due to the fact that this is a rare disease and tonometry is not in everyday use for studying CoAo patients. These findings should be considered as suggesting this phenomenon and more arterial stiffness studies in CoAo patients should be performed.

In conclusion, Dgrad and DHPTc are TTE variables obtained by spectral Doppler analysis that can be used to predict invasive gradients, Sgrad, which quantifies the

severity of CoAo. But in patients with mild or moderate CoAo severity, the arterial stiffness may influence the DHPTc value. Higher values of DHPTc are more likely in patients with lower arterial stiffness, and vice versa.

These finding where possible by use of regression models. Parametric and non-parametric statistical tests alone could have provided information about the presence or absence of a statistical significant relation between variables. However, regression models improve the same information by providing the quantification of the influence of those statistical significant variables. With these quantifications it is possible to obtain mathematical formulas that can be applied when studying patients.

References

1. Carvalho, J. (2008). Half peak systolic velocity deceleration time: a new index for umbilical artery Doppler, but not a new Doppler parameter for waveform quantification. *Ultrasound Obstet Gynecol*, 31, 716-717.
2. Carvalho, J., Redington, A. N., Shinebourne, E. A., Rigby, M. L., Gibson, D. (1990). Continuous wave Doppler echocardiography and coarctation of the aorta: gradients and flow patterns in the assessment of severity. *Br Heart J*, 64, 133-7.
3. Degroff, C.G., Orlando, W., Shandas, R. (2003). Insights into the effect of aortic compliance on Doppler diastolic flow patterns seen in coarctation of the aorta: a numeric study. *J Am Soc Echocardiogr*, 16: 162-169.
4. Faraway, J. (2002) *Practical Regression and Anova using R*. Available from: cran.r-project.org/doc/contrib/Faraway-PRA.pdf.
5. Gomes, J. (2011). *Regressão Linear*. Lisboa: DEIO/FCUL.
6. Gomes, J. (2011). *Regressão Binária – O Modelo Logístico*. Lisboa: DEIO/FCUL.
7. Hsieh, F. Y., Bloch, D., Larsen, M. (1998). A simple method of sample size calculation for linear and logistic regression. *Statist Med*, 17: 1623-1634.
8. Kenny, D., Polson, J.W., Martin, R.P., Wilson, D.G., Caputo, M., Cockcroft, J.R., Paton, J.F., Wolf, A.R. (2010). Surgical approach for aortic coarctation influences arterial compliance and blood pressure control. *Ann Thorac Surg*. 90, 2, 600-4.
9. Lai, W., Mertens, L., Cohen, M., Geva, Tal. (2009). *Echocardiography in Pediatric and Congenital Heart Disease from Fetus to Adult*. Chichester: Wiley-Blackwell.
10. Laurent, S., Cockcroft, J., Van Bortel, L., Boutouyrie, P., Giannattasio, C., Hayoz, D., Pannier, B., Vlachopoulos, C., Wilkinson, I., Struijker-Boudier, H. (2006). Expert consensus document on arterial stiffness: methodological issues and clinical applications. *European Heart Journal*, doi:10.1093/eurheartj/ehl254.

11. Lombardi, K.C., Northrup, V., McNamara, R.L., Sugeng, L., Weismann, C.G. (2013). Aortic stiffness and left ventricular diastolic function in children following early repair of aortic coarctation. *Am J Cardiol.* 112, 11, 1828-33.
12. Oliveira, A. G. (2009). *Bioestatística, Epidemiologia e Investigação - Teoria e Aplicações*. Lisboa: LIDEL.
13. Ou, P., Celermajer, D.S., Jolivet, O., Buyens, F., Herment, A., Sidi, D., Bonnet, D., Mousseaux, E. (2008). Increased central aortic stiffness and left ventricular mass in normotensive young subjects after successful coarctation repair. *Am Heart J.* 155, 1, 187-93.
14. Pennsylvania State University (2015). Lesson 41: One-Factor Analysis of Variance: The ANOVA Table. Available in: <https://onlinecourses.science.psu.edu/stat414/node/218>.
15. Pereira, M., Sousa, L., Grosso, A. (2014). *Intragenic Initiation in SETD2 Deficient Cells*. (unpublished Master's degree thesis). FCUL, Portugal.
16. Tacy, T.A., Baba, K., Cape, E.G. (1999). Effect of aortic compliance on Doppler diastolic flow pattern in coarctation of the aorta. *J Am Soc Echocardiogr*, 12: 636-642.
17. Tan, J.L., Babu-Narayan, S.V., Henein, M.Y., Mullen, M., Li, W. (2005). Doppler echocardiographic profile and indexes in the evaluation of aortic coarctation in patients before and after stenting. *J Am Coll Cardiol*, 46: 1045-1053.

A R Scripts

The scripts corresponding to the data processing are shown below. CoAo invasive gradients study was performed using linear regression (script A.1) and logistic regression (script A.2). Study of arterial stiffness influence using the variable PWV was also performed (script A.3).

A.1 Invasive Gradients - Linear Regression

```
#####
### Tese - Susana Cordeiro ###
##### Linear Regression #####
#####

coao=read.table("CoAoDiast.txt",h=T)
names(coao)
length(coao$age)

### Dependent variable :
# Sgrad = sgrad --- invasive pressure gradient

### Independent variables :
# Dgrad = gradc ---- Doppler method to estimate Sgrad
# DHPTc = dthpc ---- Diastolic half pressure time
# SHPTc = sthpc ---- Systolic half pressure time
# DVT = dvt ----- velocity at the end of T wave
# DVQ = dvq ----- velocity at wave
# VRc = drc --- time from Vmax to decrease to 33% Vmax

names(coao)
vari=c("drc","dvt","dvq","sthpc","dthpc","gradc")
coao=coao[,c("sgrad",vari)]
summary(coao)
```

```
#####
```

```
### SIMPLE LINEAR REGRESSION:
```

```
### Dgrad = gradc
```

```
plot(sgrad~gradc,coao,xlim=c(0,100),ylim=c(0,100));abline(mgradc)
mgradc=lm(sgrad~gradc,coao);summary(mgradc)
abline(a=0,b=1,col=2,lty=2)
# Dgrad slightly overestimates Sgrad
```

```
### DHPTc = dthpc
```

```
plot(sgrad~dthpc,coao)
mdthpc=lm(sgrad~dthpc,coao);summary(mdthpc)
abline(mdthpc)
```

```
### SHPTc = sthpc
```

```
plot(sgrad~sthpc,coao)
msthpc=lm(sgrad~sthpc,coao);summary(msthpc)
abline(msthpc)
```

```
### DVT = dvt
```

```
plot(sgrad~dvt,coao,col=3)
mdvt=lm(sgrad~dvt,coao[coao$dvt!=0,]);summary(mdvt)
abline(mdvt,col=2)
```

```
### DVQ = dvq
```

```
plot(sgrad~dvq,coao)
mdvq=lm(sgrad~dvq,coao);summary(mdvq)
points(fitted(mdvq)~dvq,col=2,pch=19,coao)
# Because of a lot of zeros --- new variable was tested
coao$dvqc=as.factor(with(coao,ifelse(dvq==0,0,1)))
```

```
### VRc = drc
```

```
plot(sgrad~drc,coao[coao$sgrad>30,],ylim
=c(0,100),xlim=c(0,1200),col=3)
points(sgrad~drc,coao[coao$sgrad<30,],pch=19,col=2)
# drc > 400 then sgrad > 30
# there is a different behaviour for sgrad > 30 mmHg
# new variable drcc
coao$drcc=with(coao,ifelse(drc<400,1,0))
boxplot(sgrad~drcc,coao)
plot(sgrad~drc,coao)
mdrc=lm(sgrad~drcc+I(drc*drcc),coao);summary(mdrc)
points(fitted(mdrc)~coao$drc,pch=19,col=3)
```

```
#####
```

```
### Final Graph
```

```
par(mfrow=c(3,2))
```

```
plot(sgrad~gradc,coao,xlim=c(0,100),ylim=c(0,100),font.lab=2,pch=19,
```

```

col=2,ylab="Sgrad",xlab="Dgrad"); abline(mgradc);
abline(a=0,b=1,col="gray70",lty=2)

plot(sgrad~dthpc, coao, font.lab=2, pch=19, col=4, ylab="Sgrad",
xlab="DHPTc"); abline(mdthpc)

plot(sgrad~sthpc, coao, font.lab=2, pch=19, col="cadetblue", ylab="Sgrad",
xlab="SHPTc"); abline(msthpc)

plot(sgrad~dvt, coao, font.lab=2, pch=19, col="forestgreen", ylab="Sgrad",
xlab="DVT"); abline(mdvt)

plot(sgrad~dvq, coao, font.lab=2, pch=19, col=3, ylab="Sgrad", xlab="DVQ");
abline(mdvq)

plot(sgrad~drc, coao[coao$sgrad<30, ], font.lab=2, ylim
=c(0,100), xlim=c(0,1200),
col="darkgoldenrod2", ylab="Sgrad", xlab="VRc")
points(sgrad~drc, coao[coao$sgrad>30, ], pch=19, col="darkgoldenrod2")
points(fitted(mdrc)~coao$drc, col="gray40")

#####

### MULTIPLE LINEAR REGRESSION

mod_nulo=lm(sgrad~1, coao)
mod_completo=lm(sgrad~., coao)
m1=lm(sgrad~drc+dvt+dvq+dthpc+sthpc+gradc, coao); summary(m1)
library(MASS)
form=~drc+dvt+dvq+dthpc+sthpc+gradc
m5=stepAIC(mod_nulo, scope=list(lower=~1, upper=form),
direction="forward");
m3=stepAIC(mod_nulo, scope=list(lower=~1, upper=form),
direction="both");
m4=step(m1, direction="backward");
summary(m5); summary(m3); summary(m4)
#Second try: drcc+dvqc replaced drc+dvq, producing the same best model

coao$Sgrad=coao$sgrad
coao$Dgrad=coao$gradc
coao$DHPTc=coao$dthpc
model=lm(Sgrad~Dgrad+DHPTc, coao); summary(model)

#####

### Muticolinearity
X=model.matrix(model)
XX=cbind(X, Y=coao$sgrad)
cor(XX[, -1])
library(car)
vif(model) # vif = variance inflata factor

```

```
### Residuals
par(mfrow=c(1,2))
plot(residuals(model),main="Residuals Plot",xlab="",
ylab="",pch=19,col="cyan4"); abline(0,0)
#plot(density(rstudent(model)))
qqnorm(residuals(model), main="Residuals \n Normal Q-Q Plot"
,pch=19,col="cyan4"); qqline(residuals(model))
mean(rstudent((model)));sd(rstudent((model)))
ks.test(rstudent(model),"pnorm",0,1)
```

```
#####
```

```
### Generalized Regression Model
```

```
coao$sgrad[coao$sgrad==0]=1
m8=glm(sgrad ~ gradc + dthpc,gaussian,coao)
m9=glm(sgrad ~ gradc + dthpc,Gamma,coao)
summary(m8);summary(m9)
```

A.2 Invasive Gradients - Logistic Regression

```
#####
### Tese - Susana Cordeiro ###
##### Logistic Regression #####
#####

coao=read.table("CoAoDiast.txt",h=T)
cut=20
coao$sign=with(coao,ifelse(sgrad>=cut,1,0))
summary(as.factor(coao$sign))

### Dependent variable - Sign:
# Sign = 1 if invasive pressure gradient above 20 mmHg
# Sign = 0 if invasive pressure gradient bellow 20 mmHg

### Independent variables :
# Dgrad = gradc ---- Doppler method to estimate Sgrad
# DHPTc = dthpc ---- Diastolic half pressure time
# SHPTc = sthpc ---- Systolic half pressure time
# DVT = dvt ----- velocity at the end of T wave
# DVQ = dvq ----- velocity at wave
# VRC = drc --- time from Vmax to decrease to 33% Vmax

names(coao)
vari=c("drc","dvt","dvq","sthpc","dthpc","gradc")
coao=coao[,c("sign","sgrad",vari)]
summary(coao)

#####

### SIMPLE LOGISTIC REGRESSION:

par(mfrow=c(3,2))

### Dgrad = gradc
boxplot(gradc~sign,coao,col="coral3",ylab=expression(bold("Sign")),
xlab=expression(bold("Dgrad")))
mgradc=glm(sign~gradc,binomial,coao);summary(mgradc)

### DHPTc = dthpc
boxplot(dthpc~sign,coao,col="cornflowerblue",
ylab=expression(bold("Sign")),
xlab=expression(bold("DHPTc")))
mdthpc=lm(sign~dthpc,coao);summary(mdthpc)
```

```

### SHPTc = sthpc
boxplot(sthpc~sign,coao,col="cadetblue",ylab=expression(bold("Sign")),
xlab=expression(bold("SHTPc")))
msthpc=lm(sign~sthpc,coao);summary(msthpc)

### DVT = dvt
boxplot(dvt~sign,coao,col="forestgreen",ylab=expression(bold("Sign")),
xlab=expression(bold("DVT")))
mdvt=lm(sign~dvt,coao[coao$dvt!=0,]);summary(mdvt)

### DVQ = dvq
boxplot(dvq~sign,coao,col="darkolivegreen3",
ylab=expression(bold("Sign")),
xlab=expression(bold("DVQ")))
mdvq=lm(sign~dvq,coao);summary(mdvq)
# When dvq>0, sign is always equal to 1

### VRc = drc
boxplot(drc~sign,coao,col="darkgoldenrod2",
ylab=expression(bold("Sign")),
xlab=expression(bold("VRc")))
mdrc=lm(sign~drc,coao);summary(mdrc)

### Likelihood-ratio test ###

pval=c(0);nome=c(0)
tab=list();prop=list()
vari=c("drc","dvt","sthpc","dthpc","gradc") #dvq was removed
k=length(vari)
for (i in 1:k)
{
  g=glm(sign~coao[,vari[i]],binomial,coao)
  anova(g,test="Chisq")
  (pval[i]=anova(g,test="Chisq")$"Pr(>Chi)"[2])
}
cbind(vari,pval)
# age and pwv are not significant
# all 5 TTE variables have p-value < 0.05

#####

### MULTIPLE LOGISTIC REGRESSION:

form=sign~drc+dvt+sthpc+dthpc+gradc
modcomp=glm(form,binomial,coao)
mod_nulo=glm(sign~1,binomial,coao)

library(MASS)
mfor=stepAIC(mod_nulo,scope=list(lower=~1,upper=form),

```



```

direction="forward");
mboth=stepAIC(mod_nulo,scope=list(lower=~1,upper=form),
direction="both");
mback=step(modcomp,direction="backward");
cbind(coao$sign,fitted(mback)) #there are estimated values equal to 1
summary(mfor);summary(mboth);summary(mback)

# all 3 methods produce the same model mselec
mselec=glm(sign~gradc+dthpc+dvt,binomial,coao);summary(mselec)
# but dvt coefficient negative and opposing to the simple regression,
so dvt was removed

coao$Sign=coao$sign
coao$Dgrad=coao$gradc
coao$DHPTc=coao$dthpc

m1=glm(Sign~Dgrad,binomial,coao)
m2=glm(Sign~Dgrad+DHPTc,binomial,coao)
m3=glm(Sign~Dgrad*DHPTc,binomial,coao)
AIC(m1,m2,m3)
# The model with inferior AIC includes the interaction
summary(m3)

model=glm(Sign~Dgrad*DHPTc,binomial,coao);summary(model)

#####

### ROC Curve

library(ROCR)
fit<-model$fitted
par(mfrow=c(1,2))
pred <- prediction(fit,coao$sign)

# Sensibility and Specificity
sensibilidadade <- performance(pred,"tpr")
plot(sensibilidadade, main="Sensibility and Specificity",col=4,lwd=3)
especificidade <- performance(pred,"tnr")
plot(especificidade,add=T,col=2,lwd=3)
legend("bottomright", "Sensibility (blue) \nSpecificity (red) ")

# ROC Curve
area <- performance(pred, "auc")
area
plot(sensibilidadade@y.values[[1]]~1-
especificidade@y.values[[1]],t="l",lty=1,
lwd=3, xlab="1-Specificity", ylab="Sensibility", main= "ROC Curve",)
abline(1,-1, lty=2, lwd=2)

```

```
# Values of Dgrad and DHPTc for Sign == 0 and == 1
cbind(fitted(m2),sg=coao$sgrad,model.matrix(m2)[,-
c(1,4)])[coao$sign==0,]
cbind(fitted(m2),sg=coao$sgrad,model.matrix(m2)[,-
c(1,4)])[coao$sign==1,]

# Values for Sensibility and Specificity
cbind(sensibilidade@x.values[[1]],sensibilidade@y.values[[1]],
especificidade@y.values[[1]])
#Choosing a cut of 0.34, the model has a sensibility of 96% and
specificity 74%

### ROC curve for the model with only Dgrad as the independent
variable

model2=glm(Sign~Dgrad,binomial,coao)

fit<-model2$fitted
pred <- prediction(fit,coao$sign)
area <- performance(pred, "auc")
area

# Values for Sensibility and Specificity
cbind(sensibilidade@x.values[[1]],sensibilidade@y.values[[1]],especifi
cidade@y.values[[1]])
#Choosing a cut of 0.26, the model has a sensibility of 93% and
specificity 65%
```

A.3 Arterial Stiffness

```
#####
### Tese - Susana Cordeiro ###
##### Arterial Stiffness #####
#####

coao=read.table("CoAoDiast.txt",h=T)
coao$sign=with(coao,ifelse(sgrad>=20,1,0))

### Variables :
# PWV = pwv -- pulse wave velocity, measurement for arterial stiffness
# Sgrad = sgrad --- invasive pressure gradient
# Dgrad = gradc ---- Doppler method to estimate Sgrad
# DHPTc = dthpc ---- Diastolic half pressure time
# DTail ---- presence or absence of DHPTc

#####

### Graph analysis

mean(coao$pwv,na.rm=TRUE)
#PWV bellow and above mean value was chosen as high or low PWV

# PWV and Sgrad
par(mfrow=c(1,2))
plot(sgrad~gradc,t="n",coao,ylab=expression(bold("Sgrad")),
xlab=expression(bold("Dgrad")))
points(sgrad~gradc,pch=19,coao)
points(sgrad~gradc,pch=19,coao[coao$pwv>7.18,],col=2)
points(sgrad~gradc,pch=19,coao[coao$pwv<=7.18,],col=3)
legend("topleft", "Low PWV (green) \nHigh PWV (red)")
# As expected, PWV does not seem to be associated with Dgrad

# PWV and DHPTc
plot(sgrad~dthpc,t="n",coao,ylab=expression(bold("Sgrad")),
xlab=expression(bold("DHPTc")))
points(sgrad~dthpc,pch=19,coao)
points(sgrad~dthpc,pch=19,coao[coao$pwv>7.18,],col=2)
points(sgrad~dthpc,pch=19,coao[coao$pwv<=7.18,],col=3)
abline(30,0)
legend("topleft", "Low PWV (green) \nHigh PWV (red)")
#Under 30 mmHg, there seems to be association

# PWV and DHPTc, in Sgrad<30mmHg
par(mfrow=c(1,1))
plot(sgrad~dthpc,t="n",coao,ylab=expression(bold("Sgrad")),
xlab=expression(bold("DHPTc")),ylim=c(0,35),xlim=c(0,100),
```

```

main="Sgrad < 30 mmHg")
points(sgrad~dthpc,pch=19,coao[coao$sgrad<30,])
points(sgrad~dthpc,pch=19,coao[coao$pwv>7.18&coao$sgrad<30,],col=2)
points(sgrad~dthpc,pch=19,coao[coao$pwv<=7.18&coao$sgrad<30,],col=3)
legend("topleft", "Low PWV (green) \nHigh PWV (red)")

#Boxplot
dthpcc=as.factor(iffelse(coao$dthpc>0,1,0))
boxplot(coao$pwv[coao$sgrad<30]~dthpcc[coao$pwv!="NA"&coao$sgrad<30],
col=(c(2,3)),ylab=expression(bold("PWV")),
xlab=expression(bold("DTail")),main="Sgrad < 30 mmHg")

#####

#Linear regression - All patients and all Sgrad
summary(lm(sgrad~pwv,coao))
summary(lm(gradc~pwv,coao))
summary(lm(dthpc~pwv,coao))

#####

### PWV and DTHPc in patients with Sgrad < 30 mmHg ###

base1=coao[,c("sign","sgrad","dvq","pwv","drc","dvt","sthpc",
"dthpc","gradc)];base1=subset(base1,pwv>=0)
base1[order(base1$pwv),]
base1$DTail=with(base1,iffelse(dthpc==0,0,1))
base1=subset(base1,sgrad<30)
length(base1$pwv)
base1$PWV=base1$pwv

### Binary Regression Model:
m=glm(DTail~PWV,binomial,base1);summary(m)

# ODDS
1/exp(coef(m)[2])
1/exp(confint(m))[2,]

```