Prospective and Multicentric Validation of the ArgenSCORE in Aortic Valve Replacement Surgery. Comparison with the EuroSCORE I and the EuroSCORE II

ABSTRACT

Introduction
In patients with aortic stenosis and planned aortic valve replacement, an accurate stratification of surgical risk is mandatory to offer the best individual option. Preoperative risk scores have recovered a leading role in the assessment of these patients.

Objectives
The aim of this study was to perform a prospective, multicentric validation of the ArgenSCORE in patients with aortic valve replacement and compare its performance with the EuroSCORE I and the EuroSCORE II.

Methods
A total of 250 adult patients undergoing aortic valve replacement at four centers of the City of Buenos Aires were included in the study from February 2008 to December 2012. The ArgenSCORE was compared with the EuroSCORE I and the EuroSCORE II, evaluating model discrimination with the area under the ROC curve and calibration power comparing the relation between observed mortality and predicted mortality.

Results
The mean age of the validation population (n = 250) was 68.62 ± 13.3 years and overall mortality of was 3.6 %. The ArgenSCORE showed good discrimination power (area under the ROC curve of 0.82) and a good predictive capacity to allocate risk (relation between observed mortality: 3.6 % vs. predicted mortality: 3.39%; p = 0.471). The EuroSCORE I showed poor discrimination power (area under the ROC curve of 0.62) and risk overestimation (relation between observed mortality: 3.6 % vs. predicted mortality: 5.58 %; p < 0.0001). The EuroSCORE II showed an acceptable discrimination power (area under the ROC curve of 0.76), though lower than that of the ArgenSCORE, but a significant underestimation of predicted risk (relation between observed mortality: 3.6 % vs. predicted mortality: 1.64 %; p < 0.0001).

Conclusions
The ArgenSCORE evidenced adequate excellent ability to predict mortality in patients undergoing AVR aortic valve replacement surgery. This local model demonstrated good discrimination power and better calibration compared to the European models, as the EuroSCORE I overestimated and the EuroSCORE II underestimated predicted risk.

Key words
Cardiovascular surgical procedures - Risk assessment - Mortality

SEE RELATED ARTICLE: Rev Argent Cardiol 2014;82:3-4 - http://dx.doi.org/10.7775/rac.v82.i1.3388
INTRODUCTION
On-pump aortic valve replacement surgery (AVR) has proved to be a safe and effective intervention to improve quality of life and prognosis in severe symptomatic aortic stenosis. However, several series have documented that about 30% to 40% of patients with this valve disease cannot be operated using this conventional strategy due to old age, multiple comorbidities or on the recommendation of the treating physicians. (1-6)

In recent years, the development of transcatheter aortic valve replacement (TAVR) has become a new therapeutic alternative widely used in these patients. Although this new intervention appears as a reasonable option for patients with major comorbidities and high surgical risk, indications should be limited and well-defined as surgical AVR is still the current “gold standard”. (7) Thus, in patients with aortic stenosis and indication for valve replacement it is essential to make a correct operative risk stratification to offer the best possible choice in each case.

Predictive models or preoperative risk scores have been extensively used in the last three decades to achieve a better indication and optimize the outcome of cardiac surgery (8-13). Among the most commonly used risk models are those of the Society of Thoracic Surgeons (12), the European System for Cardiac Operative Risk Evaluation I (EuroSCORE I) (8, 9) and in recent years the EuroSCORE II (13).

Due to the poor performance of these models when applied to populations different from those used for their development, in 1999 we developed in our country an additive risk model for in-hospital cardiac surgical mortality, the Argentine System for Cardiac Operative Risk Evaluation (ArgenSCORE), subsequently validated and recalibrated in 2007 and published in 2009. (10, 11) The aims of this multicentric and prospective study were to validate the recalibrated ArgenSCORE in patients undergoing AVR and to compare its performance with the EuroSCORE I and the new EuroSCORE II.

METHODS
ArgenSCORE development and recalibration
The ArgenSCORE is a simple, additive risk model, with graphical representation, developed from the analysis of 2903 patients consecutively undergoing cardiac surgery at the Cardiology Institute of the Hospital Español of Buenos Aires from June 1994 to December 1999.

The method of model development and recalibration has been previously published in detail. (10) The model identifies 18 independent predictive variables of in-hospital mortality.

To improve its applicability in daily practice, a graphical method was developed allowing a simple estimation of death risk using only a pre-printed grid. Each variable included in the analysis was assigned a score calculated by multiplying by 10 the logistic regression coefficient. The estimated risk corresponds to the sum of the individual scores of each positive variable detected in the patient. Finally, to develop the pre-printed ArgenSCORE grid, a distribution curve was drawn to correlate absolute score values with predicted risk by multiple logistic regression.

In 2007 the first prospective, external and sequential ArgenSCORE validation was performed in 1807 patients. To optimize its performance, the original 1999-model was recalibrated (14-17), running it against a new logistic regression for in-hospital mortality, where the original 1999-model was the independent variable and in-hospital mortality the dependent variable. (17, 18) Figure 1 shows the recalibrated 2007-ArgenSCORE (10, 11) where estimated mortality is illustrated as a curve on a pre-printed grid, with its corresponding confidence intervals (CI), and is the recalibrated version of the model currently employed in daily practice. This pre-printed grid allows its simple, practical use, as the ArgenSCORE can be carried as a graphic pocket card score for its easy reference and application. (10)

RESULTS
From a total of 1908 adult patients consecutively undergoing cardiac surgery from February 1998 to December 2012 at four centers of the City of Buenos Aires (Instituto FLENI, Sanatorio Dr. Julio Mendez, Sanatorio de la Trinidad and Sanatorio Los Arcos), only patients with isolated AVR and AVR associated to coronary artery bypass grafting (CABG) were included in the analysis. Patients with AVR associated with ascending aortic replacement, other valve interventions, any other cardiac intervention and all patients operated-on with other types of cardiac surgery were excluded from the study. In-hospital mortality, defined as death occurring before discharge, was the study endpoint.

This multicentric population included 250 consecutive patients. The ArgenSCORE was externally and sequentially validated and its performance was compared with the EuroSCORE I and the EuroSCORE II. Discrimination of the three models was assessed calculating the area under the Receiver Operating Characteristic (ROC) curve. The calibration power of the three models was also evaluated comparing the relation between observed versus predicted mortality in all patients and through three risk tertiles. (14, 19, 20) The difference between mean observed mortality and mean predicted mortality was assessed using the t test. (21) A p value < 0.05 was considered statistically significant. The SPSS statistical software package, version 21.0 (SPSS Inc., Chicago, Ill.) was used to perform statistical analyses.

ArgenSCORE validation in aortic valve replacement
This study included 1908 adult patients consecutively undergoing cardiac surgery from February 1998 to December 2012 at four centers of the City of Buenos Aires (Instituto FLENI, Sanatorio Dr. Julio Mendez, Sanatorio de la Trinidad and Sanatorio Los Arcos), only patients with isolated AVR and AVR associated to coronary artery bypass grafting (CABG) were included in the analysis. Patients with AVR associated with ascending aortic replacement, other valve interventions, any other cardiac intervention and all patients operated-on with other types of cardiac surgery were excluded from the study. In-hospital mortality, defined as death occurring before discharge, was the study endpoint.

This multicentric population included 250 consecutive patients. The ArgenSCORE was externally and sequentially validated and its performance was compared with the EuroSCORE I and the EuroSCORE II. Discrimination of the three models was assessed calculating the area under the Receiver Operating Characteristic (ROC) curve. The calibration power of the three models was also evaluated comparing the relation between observed versus predicted mortality in all patients and through three risk tertiles. (14, 19, 20) The difference between mean observed mortality and mean predicted mortality was assessed using the t test. (21) A p value < 0.05 was considered statistically significant. The SPSS statistical software package, version 21.0 (SPSS Inc., Chicago, Ill.) was used to perform statistical analyses.

RESULTS
The external, sequential validation population consisted of 250 patients, consecutively undergoing isolated AVR or associated with CABG. Mean age was 68.62 ± 13.3 years and 36.8% were women. Isolated
AVR was performed in 67.2% of patients and AVR associated with CABG in 32.8%. Overall in-hospital mortality was 3.6%. Table 1 shows the characteristics and mortality of the validation population. The corrected valve disease was predominant aortic stenosis in 86.4% of patients and predominant aortic regurgitation in 13.6% (Table 2).

The external, sequential validation showed that the recalibrated ArgenSCORE had good ability to discriminate in-hospital mortality in AVR surgery with an area under the ROC curve of 0.82 (95% CI: 0.74 – 0.91; p < 0.001) (Figure 2).

The model also showed good ability to allocate surgical risk (calibration power) in the overall population, evidenced by the excellent relation between observed mortality of 3.6% and predicted mortality of 3.9% (p = 0.471). The analysis according to the risk tertiles indicated a more irregular calibration capacity (Table 3).

Assessment of the EuroSCORE I performance in this population of patients undergoing AVR showed poor ability to discriminate surgical mortality risk, with an area under the ROC curve of 0.62 (95% CI: 0.43 – 0.82; p = 0.186) (see Figure 2). Moreover, its capacity to allocate risk was inadequate as it significantly overestimated predicted risk with a relation between observed mortality of 3.6% versus predicted mortality of 1.64% (p < 0.0001) (see Table 3).

DISCUSSION
Calcific aortic stenosis is a frequent condition in the elderly, with an estimated 2% to 4% incidence in people over 65 years of age, and is the most common acquired valve disease in this population. (22, 23) Although surgical aortic valve replacement is considered the treatment of choice in these patients, associated comorbidities or old age can lead to surgical contraindication or significantly increase the operative risk in many elderly patients.

With the development and improvement of TAVR techniques, a possible alternative with lower operative risk in patients with aortic valve stenosis and high surgical risk has been established. However, due to complications associated with percutaneous procedures and unknown long-term outcome, surgical valve replacement remains today the intervention of choice in patients with moderate or low operative risk.

Faced with these treatment options, it is clinically important to properly assess severe aortic stenosis surgical risk. In this scenario, predictive models for cardiac surgery mortality are a useful and objective tool to complement the clinical criteria for surgical risk stratification, contributing to better treatment selection (8-13).

The ArgenSCORE is a predictive risk model for cardiac surgery developed in our country in 1999 and recalibrated in 2007. (10, 11) This model, which uses objective variables (12, 19, 20) and is easily implemented by means of graphical representation, can be applied to stratify risk simply and with adequate performance in clinical preoperative assessment (see Figure 1). The results of this study show an excellent performance of the ArgenSCORE applied in a population of patients undergoing AVR. The model showed
a good discriminatory power for mortality and a good ability to allocate risk in the overall population, with an excellent relation between observed mortality (3.6%) and predicted mortality (3.39%), although the lower tertiles evidenced a slight, albeit significant, mortality overestimation. Conversely, application of the EuroSCORE I showed poor ability both to discriminate the risk of surgical mortality and to allocate operative risk in the overall population, as it overestimated operative risk (relation between observed / predicted mortality 3.6% vs. 5.58%). In turn, the EuroSCORE II showed an acceptable ability to discriminate in-hospital mortality risk, although lower than the ArgenSCORE. However, regarding its calibration power, the EuroSCORE II showed a flaw in its performance due to its poor ability to assign risk of surgical mortality; with a significant undervaluation of estimated risk (relation between observed / predicted mortality 3.6% vs. 1.64%).

Previous experience evaluating the performance of different international risk models in isolated AVR sub-populations showed low performance of their predictive powers. The EuroSCORE I both in its logistic and additive versions overestimated risk in patients with isolated AVR, while the model of the Society of Thoracic Surgeons generally underestimated risk in these populations (7, 23-28). Moreover, a meta-analysis of all publications which intended to validate the EuroSCORE I in patients with valve surgery showed poor performance of this model both in the overall population and in patients with AVR (29).

The recently published EuroSCORE II was developed to improve the performance of the EuroSCORE I. Although, unlike its predecessor, it allows risk stratification according to the type of procedure to be performed, this new model in its original publication (13) and in recent external validations (30-32) under-

**Table 1. Population characteristics of patients undergoing AVR (validation population)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Validation population 2008 - 2012 (n: 250) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 60 years</td>
<td>18.8</td>
</tr>
<tr>
<td>60-69 years</td>
<td>22</td>
</tr>
<tr>
<td>70-79 years</td>
<td>41.6</td>
</tr>
<tr>
<td>≥ 80 years</td>
<td>17.2</td>
</tr>
<tr>
<td>Female gender</td>
<td>38.8</td>
</tr>
<tr>
<td>Hypertension</td>
<td>52.6</td>
</tr>
<tr>
<td>Diabetes</td>
<td>13</td>
</tr>
<tr>
<td>Insulin dependence</td>
<td>2.8</td>
</tr>
<tr>
<td>COPD</td>
<td>6.8</td>
</tr>
<tr>
<td>Preoperative renal failure</td>
<td>4.4</td>
</tr>
<tr>
<td>Preoperative renal dialysis</td>
<td>0.48</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>7.6</td>
</tr>
<tr>
<td>Active infective endocarditis</td>
<td>2.4</td>
</tr>
<tr>
<td>Previous AMI</td>
<td>8.8</td>
</tr>
<tr>
<td>AMI &lt; 90 days</td>
<td>1.6</td>
</tr>
<tr>
<td>FC III-IV angina</td>
<td>2.8</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>5.27</td>
</tr>
<tr>
<td>Pulmonary hypertension</td>
<td>2.0</td>
</tr>
<tr>
<td>Previous cardiac surgery</td>
<td>7.6</td>
</tr>
<tr>
<td>Elective surgery</td>
<td>92.4</td>
</tr>
<tr>
<td>Urgency surgery</td>
<td>7.6</td>
</tr>
<tr>
<td>Emergency surgery</td>
<td>0.0</td>
</tr>
<tr>
<td>Preoperative IABP</td>
<td>0.0</td>
</tr>
<tr>
<td>Isolated aortic valve replacement</td>
<td>67.2</td>
</tr>
<tr>
<td>Aortic valve replacement + CABG</td>
<td>32.8</td>
</tr>
<tr>
<td>One vessel disease</td>
<td>14.8</td>
</tr>
<tr>
<td>Two vessel disease</td>
<td>10.8</td>
</tr>
<tr>
<td>Three vessel disease</td>
<td>7.2</td>
</tr>
<tr>
<td>Moderate LV dysfunction</td>
<td>19.4</td>
</tr>
<tr>
<td>Severe LV dysfunction</td>
<td>5.7</td>
</tr>
<tr>
<td>Overall mortality</td>
<td>3.60</td>
</tr>
</tbody>
</table>

COPD = chronic obstructive pulmonary disease; AMI = acute myocardial infarction; CF = Functional class; IABP = Intraaortic balloon pump; CABG = Coronary artery bypass grafting; LV = Left ventricular

**Table 2. Valve pathologies corrected with aortic valve replacement (validation population)**

<table>
<thead>
<tr>
<th>Valve pathology</th>
<th>Validation population 2008 - 2012 (n: 250) n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe stenosis (total)</td>
<td>200 80</td>
</tr>
<tr>
<td>Severe pure stenosis</td>
<td>186 74.4</td>
</tr>
<tr>
<td>Severe stenosis + moderate heart failure</td>
<td>14 5.6</td>
</tr>
<tr>
<td>Moderate pure stenosis</td>
<td>16 6.4</td>
</tr>
<tr>
<td>Severe pure heart failure</td>
<td>31 1.4</td>
</tr>
<tr>
<td>Moderate pure heart failure</td>
<td>3 1.2</td>
</tr>
</tbody>
</table>

COPD = chronic obstructive pulmonary disease; AMI = acute myocardial infarction; CF = Functional class; IABP = Intraaortic balloon pump; CABG = Coronary artery bypass grafting; LV = Left ventricular

**Fig. 2.** Areas under the ROC curve obtained in the AVR population (n = 250). The recalibrated (2007) ArgenSCORE presented an area under the ROC curve of 0.82 (95% CI: 0.74 - 0.91, p < 0.001); the EuroSCORE I an area under the ROC curve of 0.62 (95% CI: 0.43 - 0.82, p = 0.186) and the EuroSCORE II an area under the ROC curve of 0.76 (95% CI: 0.65-0.87, p = 0.007).
estimated preoperative risk. Similarly, underestimation of operative risk was also observed in the present work. These findings support the benefits of developing and using local models in preoperative risk stratification. (20, 33, 34)

The application of international preoperative risk models intended for general use has shown performance limitations when employed in populations other than those on which they were developed. (33-36) This may be related to geographical and epidemiological differences in the risk profile, in the surgical strategies used and in the decision making of an eventual cardiac surgery between different countries and even between surgical centers in the same country (33-35, 37). These epidemiological differences were also observed in our experience, comparing a local population with the population used to develop the EuroSCORE I (11).

Some limitations of this study should be highlighted. External validation was only performed at four institutions of the city of Buenos Aires without extending it to a larger number of centers and other regions of our country. Although the event assessed was in-hospital mortality, complications such as different morbidities are important in the prognosis and quality of life and should be considered in the preoperative assessment (15, 19).

The risk / benefit of aortic valve surgery should be taken into account in the individual patient beyond the contribution of a surgical risk assessment based on statistical and mathematical methods, as is the case of these models, which only complement clinical criteria (6, 25). Risk models detect and focus on different backgrounds and comorbidities, but usually do not consider other variables indicative of the biological state and the fragility of the patient which have an impact on his/her outcome and postoperative prognosis (38, 39).

**CONCLUSIONS**

The ArgenSCORE represents the first externally validated risk stratification model of in-hospital mortality in cardiac surgery developed in our country. Its simple methodology and graphical representation allows easy risk estimation and model implementation. This local model has shown excellent performance in a prospective multicentred population of patients undergoing AVR surgery, revealing greater power of discrimination and better calibration compared with the EuroSCORE I and EuroSCORE II models.

**RESUMEN**

Validación prospectiva y multicéntrica del ArgenSCORE en la cirugía de reemplazo valvular aórtico. Comparación con el EuroSCORE I y el EuroSCORE II

**Introducción**

Los modelos de riesgo preoperatorios han recobrado un papel protagónico en la evaluación de pacientes para un eventual reemplazo valvular aórtico (RVA).

**Objetivos**

Validar el AgenSCORE en forma prospectiva y multicéntrica en pacientes con RVA y comparar su rendimiento con el EuroSCORE I y el EuroSCORE II.

**Material y métodos**

Se incluyeron 250 pacientes consecutivos con RVA en 4 centros de Buenos Aires, desde Febrero 2008 hasta Diciembre 2012. Se comparó el rendimiento del ArgenSCORE, EuroSCORE I y EuroSCORE II, evaluando la discriminación mediante el área bajo la curva ROC y el poder de calibración comparando la relación entre mortalidad observada / mortalidad predicha.

**Resultados**

La población de validación incluyó 250 pacientes, con edad media de 68.62 ± 15.3 años y mortalidad global del 3.6 %. El ArgenSCORE mostró buen poder de discriminación, curva ROC: 0,82, y buena capacidad para asignar riesgo, relación mortalidad observada (3,6 % versus mortalidad predicha 3,39 %; p = 0,471). El EuroSCORE I mostró bajo poder discriminativo, curva ROC: 0,62 y además, sobrevaloró el riesgo estimado, relación mortalidad observada (3,6 % versus mortalidad predicha 5,58 %; p <0,0001). El EuroSCORE II mostró aceptable capacidad discriminativa, curva ROC: 0,76, aunque menor al del ArgenSCORE, pero mostró una significativa subvaloración del riesgo estimado, relación mortalidad observada (3,6 % versus mortalidad predicha 1,64 %; p <0,0001)

**Conclusiones**

El ArgenSCORE demostró tener un excelente rendimiento en pacientes operados con RVA. Este modelo local mostró buen poder de discriminación y una mejor calibración comparado a los modelos europeos, ya que el riesgo estimado fue sobrevalorado por el EuroSCORE I y subvalorado por EuroSCORE II.

**Palabras clave**

Procedimientos quirúrgicos cardiovasculares - Medición de riesgo - Mortalidad

**Conflicts of interest**

None declared.
REFERENCES


