ArgenSCORE versus EuroSCORE II

To the Editor

We have read the article by Carosella et al. in the previous issue of the Journal. (1) The authors have been working on risk scores for years and are qualified researchers on this topic. In this article, they compare the results of the ArgenSCORE with the EuroSCORE I and II in patients with (isolated or combined) aortic valve replacement, and conclude with the advantages of the local score. While it is important to develop local models, there are limitations in this analysis that deserve to be commented on. Calibration was performed by comparing the differences between observed and predicted mortality, overall and in risk tertiles. The authors did not use the usual Hosmer-Lemeshow test, which allows for global and regional assessment (by tertiles). Despite its limitations, most studies report this statistic. Although the ArgenSCORE ability to allocate risk seems globally adequate (3.6% versus 3.4%; p = 0.471), the analysis by tertiles in Table 3 shows that the ArgenSCORE predicted mortality was significantly different from the mortality observed in two of the three tertiles; thus, it was adequate only for patients with higher risk (upper tertile). Another question on the classification by tertiles is why the second tertile has only 77 operated patients and the upper tertile 89, when, by definition, each tertile should have between 83 and 84 subjects for n = 250. Another controversial aspect is related to the comparison "between" models, since informal rather than formal comparisons are made between ROC curves or likelihood ratios. The Hanley-McNeil test would have allowed evaluating the differences between them. However, comparing confidence intervals between ROC areas of the ArgenSCORE and the EuroSCORE II suggests no significant differences, indicating that models are equivalent and that an advantage could only be determined by increasing the sample size.

It should also be pointed out that the use of an international score offers comparative advantages and insertion in the world, something difficult to achieve with the ArgenSCORE. Beyond overestimation of risk in the EuroSCORE I, the validity of the EuroSCORE II predictions are being analyzed; however, the EuroSCORE II validations present ROC areas between 0.760 and 0.990. (2-5)

For comparative purposes, we calculated the prospective results with the EuroSCORE II in 192 isolated or combined aortic valve replacements performed in 2012-2013. The ROC area was 0.808 and the Hosmer-Lemeshow test showed a good calibration (p = 0.656). While the EuroSCORE II is more accurate in high risks than its predecessor, it is extremely demanding with low risks. Thus, young patients without morbidities undergoing coronary surgery will have an expected risk of about 0.5%. This new standard of quality requires addressing the problem of progressively improving surgical outcomes in our environment. In the case of measuring the performance of risk scores, the limit of predicting 3-sigma probability events (≈ 0.5%) lies in the assumption that, at this level of expected probability, death becomes an almost unpredictable fact.

Raúl A. Borracci MTSAC
Universidad Austral, Argentina
Javier Mariani MTSAC
Hospital El Cruce, Florencio Varela, Buenos Aires, Argentina

REFERENCES

1. Carosella VC, Mastantuono C, Golovonevsky V, Cohen V, Grancelli H, Rodríguez W, et al. Validación prospectiva y multicéntrica del ArgenSCORE en la cirugía de reemplazo valvular aórtico. Comparación con el EuroSCORE I y el EuroSCORE II. Rev Argent Cardiol 2014;82:5-11. http://doi.org/rr7

2. Borracci RA, Rubio M, Celano L, Ingino CA, Allende NG, Ahuad Guerrero RA. Prospective validation of EuroSCORE II in patients undergoing cardiac surgery in Argentinean centres. Interact CardioVasc Thorac Surg 2014;1-5. Epub ahead of print

3. Grant SW, Hickey GL, Dimarakis I, Trivedi U, Bryan A, Treasure T, et al. How does EuroSCORE II perform in UK cardiac surgery; an analysis of 23 740 patients from the Society for Cardiothoracic Surgery in Great Britain and Ireland National Database. Heart 2012;98:1568-72. http://doi.org/pvt

4. Chalmers J, Pullan M, Fabri B, McShane J, Shaw M, Mediratta N, et al. Validation of EuroSCORE II in a modern cohort of patients undergoing cardiac surgery. Eur J Cardiothorac Surg 2013;43:688-94. http://doi.org/rh9

5. Carnero-Alcázar M, Silva Guisasola JA, Reguillo Lacruz FJ, Maroto Castellanos LC, Cobiella Carnicer J, Villagrán Medinilla E, et al. Validation of EuroSCORE II on a single-centre 3800 patient cohort. Interact CardioVasc Thorac Surg 2013;16:293-300. http://doi.org/pvv

Authors' Reply

We thank Dr. Borracci and Dr. Mariani for their comments on our work, and we would like to make the following observations.

Regarding the use of the Hosmer-Lemeshow test, similar to our work, several publications have evaluated the calibration to validate models in cardiac surgery and intensive care by comparing overall observed and estimated mortality (OM/EM) and among different risk groups using other tests. (1, 2) The Hosmer-Lemeshow test does not evaluate by tertiles (as stated) but by deciles of risk. Although it is often used, it has its limitations, and other researchers have warned about its problems to calibrate models, highlighting the importance of evaluating the OM/EM ratio with other tests in risk groups.

The ArgenSCORE ability to allocate risk was adequate in the overall population (3.6% vs. 3.39%; p = 0.471); the EuroSCORE I overestimated risk (3.6%

versus 5.58%; p < 0.0001) and the EuroSCORE II underestimated it (3.6% versus 1.64% (p < 0.0001). In the analysis by tertiles, the three models presented significant differences in two of the three tertiles, but with the advantage that the ArgenSCORE showed a good prediction of overall risk, not observed with the other two models.

The comment on the number of cases per tertiles is striking. These categories are built by ordering the sample according to its value (estimated mortality); then, the corresponding value for the desired partition (33.33%) is calculated, and the value it corresponds to in the distribution is observed. Since the sample has many repeated figures and the same value cannot be in two different categories, the difference observed in the number of cases by tertiles is perfectly possible.

Clearly, the interpretation of our results and conclusions has created some confusion. The article states that the EuroSCORE II showed an acceptable ability to discriminate risk (ROC curve: 0.76, CI: 0.65-0.87), slightly lower than the ArgenSCORE (ROC curve: 0.82, CI: 0.74-0.91), although this difference does not seem to be significant in terms of confidence intervals. As commented above, we do conclude that the EuroSCORE II significantly underestimated the predicted risk.

Although the EuroSCORE II improved the performance of its predecessor, it requires future validations to determine its full ability. Dr. Borracci and Dr. Mariani cite studies assessing the EuroSCORE II validation but only describe its discrimination ability (ROC area), which should not be confused with calibration (allocation of risk). Coincidentally, three of the articles cited (and others not cited) reported inadequate calibration of the EuroSCORE II because it underestimated the predicted risk, similar to our experience. (3-5)

We agree that these models are "a standard of quality to improve surgical outcomes", but it is also important to estimate operative risk in the majority of the population, precisely those who are at low risk. We do not see this comparison between models as a competition ("ArgenSCORE versus EuroSCORE"); it only seeks to evaluate how our score works compared with established models. Although an international score like the EuroSCORE can have "greater global insertion", the development and validation of local models such as the ArgenSCORE can contribute to improve risk stratification in our population.

Victorio C. Carosella, César Cárdenas and Hugo Grancelli, MTSAC
Instituto FLENI

REFERENCES

1. Bhatti F, Grayson AD, Grotte G, Fabri BM, Au J, Jones M, et al; North West Quality Improvement Programme in Cardiac Interventions. The logistic EuroSCORE in cardiac surgery: how well does it predict operative risk? Heart 2006;92:1817-20. http://doi.org/d6wxpr

2. Metnitz PG, Lang T, Vesely H, Valentin A, Le Gall JR. Ratios of observed to expected mortality are affected by differences in case mix and quality of care. Intensive Care Med 2000;26:1466-72. http://doi.org/cws9wf

3. Carnero-Alcázar M, Silva Guisasola JA, Reguillo Lacruz FJ, Maroto Castellanos LC, Cobiella Carnicer J, Villagrán Medinilla E, et al. Validation of EuroSCORE II on a single-centre 3800 patient cohort. Interact Cardiovasc Thorac Surg 2013;16:293-300. http://doi.org/pvv

4. Grant SW, Hickey GL, Dimarakis I, Trivedi U, Bryan A, Treasure T, et al. How does EuroSCORE II perform in UK cardiac surgery; an analysis of 23740 patients from the Society for Cardiothoracic Surgery in Great Britain and Ireland National Database. Heart 2012;98:1568-72. http://doi.org/pvt

5. Chalmers J, Pullan M, Fabri B, McShane J, Shaw M, Mediratta N, et al. Validation of EuroSCORE II in a modern cohort of patients undergoing cardiac surgery. Eur J Cardiothorac Surg 2013;43:688-94. http://doi.org/rh9

Rev Argent Cardiol 2014;82:154-155. http://dx.doi.org/10.7775/rac.v82.i2.3936
SEE RELATED ARTICLE:
Rev Argent Cardiol 2014;82:6-12. http://dx.doi.org/10.7775/rac.v82.i1.3388

Pre-ejection period during the tilt test

"There are three kinds of lies: real lies, false lies, and statistics."
Attributed to **Benjamin Disraeli**

I congratulate the authors of the article "Fast Tilt Test: a New Paradigm in the Management of Reflex Syncope" for their ability and creativity in finding easy solutions to difficult problems. Left ventricular pre-ejection period (PEP) reminds us of our efforts to evaluate it in the 1960s. I will discuss some experiences, in the hope that they will be helpful to this methodology.

Electrical stimulation requires the recruitment of a certain number of myocytes to exert the necessary force to raise intraventricular pressure, close the mitral valve, and then open the aortic valve. All these constitute the PEP, which includes the electromechanical phase, from QRS to mitral valve closure, and the isovolumic systolic (IS) phase, from that point to the opening of the aortic sigmoid valves. (2)

In general, the electromechanical delay is constant, so the PEP variations are caused by the IS phase. The electrocardiographic lead that expresses the onset of the QRS is recommended. DII is commonly used, but it is not always the right one. (3)

Isovolumic systole depends on multiple variables, which can be summarized in preload, afterload, heart rate (HR) and myocardial function. (3)

Reduced preload, maintaining HR and fixed afterload, increase it and vice versa.

Increased afterload with fixed preload and HR, prolong it and vice versa.

Increased HR with stable preload and afterload, reduce it.

Depression of the inotropic state with the other variables fixed, prolong it.

Physiological variables that influence IS, such as breathing, needing the average of two cycles, should be considered, as well as circadian variations, patient position (standing reduces preload), and age and female gender which prolong it. Its duration is shortened with inotropic and sympathomimetic drugs, and it is prolonged with beta-blockers and vasodilators. In addition, numerous pathologies affect IS. (4)

Nitrites cause reduction in preload, blood pressure, and reflex tachycardia, a technique with too many variables to establish constants.

Controls should be performed in basal conditions and adequate environments, with 12-hour fasting, and no smoking or caffeine. Transducers should be validated, as well as absorption and resonance. The study does not mention the system characteristics or validation. (1) The small deflection at the beginning of the pulse waveform is absent in Figure 2, perhaps due to low sensitivity, and the pulse in Figure 4 is technically poor.

PEP has been evaluated after ergometric and isometric exercise, variations in position, heart diseases, and with different drugs. It has also been measured with echocardiography. (5)

Vasogenic syncopes occur for different reasons that do not occur under normal conditions during the examination, reducing enthusiasm regarding the reproducibility of the tilt test.

In our experience, based on thousands of registries, PEP measurement has proved useful for individual follow-up and somewhat useful for group evaluation, but not in the isolated patient due to the wide range of variables affecting it.

The study (1) has provided statistical support to the results, generating logical conclusions and hopes. However, figures do not always reflect reality, and a more detailed analysis and rigorous technology are necessary to obtain conclusive and quasi-lapidary assertions.

Ricardo J. Esper MTSAC

REFERENCES

1. Villamil AM, Torres Y, Mariani JA, Perona C, Tajer CD, Piccinini JM y cols. Fast tilt test: tras un nuevo paradigma en el manejo del síncope reflejo. Rev Argent Cardiol 2014;82:42-9. http://doi.org/rr8

2. Esper RJ. Tiempos sistólicos del ventrículo izquierdo. Rev Argent Cardiol 1973;41:451-6.

3. Esper RJ, Madoery RJ. Evaluación de la función ventricular mediante el fonomecanocardiograma. En: Esper RJ, Madoery RJ, editores. Progresos en Auscultación y Fonomecanocardiografía. Buenos Aires: López Libreros Editores; 1974. p. 279-91.

4. Weissler AM. Interpreting systolic time intervals in man. J Am Coll Cardiol 1983;2:1019.

5. Esper RJ, Madoery RJ. El pulso arterial. Los tiempos sistólicos del ventrículo izquierdo. En: Esper RJ, Madoery RJ, editores. Examen físico cardiovascular. Buenos Aires: Ed Promedicina; 1987. p. 42-59.

Authors' reply

We thank Dr. Ricardo Esper for his interest in our work (1) and his critical contributions. We believe, however, that your reading has not captured the contribution of our research, and your critique on our contribution to statistics in scientific research is a bit surprising.

We have used the best technology available, maximized the repetitions and validations of the method, and applied the best recommended statistics. Since the works of Pierre Alexander Louis in the XIX century in France, which –by means of a controlled study– ruled out the use of bleeding, the statistical method has been a key tool of scientific methodology. In your letter, we have found no comments on our statistical methodology applied to diagnostic methods (sensitivity, specificity, area under the ROC curve), and statistical thinking does not seem to require our defense.

In reply to your methodological criticisms, in 2009 we published our first work on the parameter behavior (2), and we received your first letter to the editor (3). In that letter, based on your experience, you spoke at length on the assessment of the pre-ejection period using phonomechanocardiography in the 1960s and 1970s, which you have repeated now with the same emphasis and similar arguments.

At that time, our reply was (4): "We have analyzed the dynamic behavior of the interval obtained in dorsal recumbent position and during tilt at 70 degrees, considering the onset of the QRS at the same monitoring point in both positions." This reduces the relevance of the onset of electrical activity criterion -provided it is the same one-, because the parameter arises from the numerical difference and not from the absolute static value. Let us recall that the carotid transducer has (4) "a bandwidth of 1 Hz to 330 Hz from mechanical sensing to digitalization of the signal measured at -3 db drop", and it is validated.

Regarding the pathophysiology of the parameter, we already discussed it in 2009 and we reinforce it now: (1) simultaneous evaluation with Doppler echocardiography during tilt test in a group of patients with vasovagal syncope and parameter prolongation showed that pre-ejection period was not significantly modified after tilt. It suggests that the phenomenon is due to ejection time prolongation, probably showing a particular distensibility response in patients with reflex crises.

Beyond the pathophysiological considerations, our premise is pragmatic in the sense that the parameter, measured as it was, predicts the result of the tilt test quite well, with the benefit of being simple and reproducible, not requiring a tilt table, and being independent from the operator, in line with "an easy solution to a difficult problem".

P.S. The quote attributed to Benjamin Disraeli –but coined by the American humorist Mark Twain– is mistranslated. It originally refers to three types of progressively discrediting lies: lies, damned lies, and statistics, and not "real lies, false lies, and statistics", as you quoted. If false lies existed, as your mistranslation proposes,

they would turn into truths, which would introduce us into the labyrinth of self-referential paradoxes (example: “this phrase is false”, it cannot be true or false), a complete nonsense.

Alejandro Villamil MTSAC, **Javier Mariani** MTSAC, **Carlos Tajer** MTSAC

REFERENCES

1. Villamil AM, Torres Y, Mariani JA, Perona C, Tajer CD, Piccinini JM y cols. Fast tilt test: tras un nuevo paradigma en el manejo del síncope reflejo. Rev Argent Cardiol 2014;82:42-9. http://doi.org/rr8

2. Villamil AM, Perona C, Carnero GS, Torres Y, Mariani JS, Tajer CD y cols. Evaluación del comportamiento del intervalo desde el inicio del QRS al inicio de la onda del pulso radial y carotídeo con el resultado del tilt test. Rev Argent Cardiol 2009;77:347-53.

3. Esper R. Evaluación del comportamiento del intervalo desde el inicio del QRS al inicio de la onda de pulso radial y carotídeo con el resultado del tilt test. Carta de Lectores. Rev Argent Cardiol 2009;78:185-6.

4. Villamil A, Torres Y, Tajer C. Evaluación del comportamiento del intervalo desde el inicio del QRS al inicio de la onda de pulso radial y carotídeo con el resultado del tilt test. Respuesta de los autores. Rev Argent Cardiol 2009;78:186.

Rev Argent Cardiol 2014;82:155-157. http://dx.doi.org/10.7775/rac.v82.i2.4053