Agreement between the new EuroSCORE II, the Logistic EuroSCORE and the Society of Thoracic Surgeons score: Implications for transcatheter aortic valve implantation

Dimitri Arangalage, Claire Cimadevilla, Soleiman Alkholder, Andrea Chiampan, Dominique Himbert, Eric Brochet, Bernard lung, Patrick Nataf, Jean-Pol Depoix, Alec Vahanian, David Messika-Zeitoun

Department of Cardiology, Assistance publique–Hôpitaux de Paris, Bichat Hospital, Paris, France
University Paris 7, Faculté de Médecine Paris-Diderot, Paris, France
Department of Cardiac Surgery, Assistance publique–Hôpitaux de Paris, Bichat Hospital, Paris, France
Department of Anesthesiology, Assistance publique–Hôpitaux de Paris, Bichat Hospital, Paris, France
INSERM U1148, Bichat Hospital, Paris, France

Received 10 January 2014; received in revised form 27 April 2014; accepted 6 May 2014
Available online 1 July 2014

Keywords
Aortic valve stenosis;
Background

Since its introduction in 2002 [1], transcatheter aortic valve implantation (TAVI) has become increasingly popular for the treatment of severe aortic stenosis. It is now considered a valuable alternative to surgical aortic valve replacement for patients with contraindications to surgery or those considered to be at high surgical risk [2]. In order to assess patients’ operative risk, several scoring systems have been developed. The European System for Cardiac Operative Risk Evaluation (EuroSCORE) [3,4] and the Society of Thoracic Surgeons (STS) predicted risk of mortality score [5,6] are the most commonly used. The European society recommends to consider performing a TAVI when the Logistic EuroSCORE is ≥ 20% and/or the STS score is ≥ 10% [7].

Recently, the EuroSCORE II has been proposed as an updated version of the Logistic EuroSCORE in order to provide a better assessment of the perioperative mortality risk of patients undergoing open heart surgery, especially heart valve surgery [8]. However, comparisons between the new EuroSCORE II and the Logistic EuroSCORE and the STS score in the setting of TAVI are rare. There are currently limited data regarding the EuroSCORE II in high-risk patients and no threshold value has been proposed to define...
Agreement between the EuroSCORE II, the Logistic EuroSCORE and the STS score

high-risk patients. Thus, the aims of the present study were to:
- compare the EuroSCORE II to the Logistic EuroSCORE and the STS score;
- determine a EuroSCORE II value corresponding to the Logistic EuroSCORE threshold of 20% and the STS score threshold of 10%;
- evaluate the agreement between these scoring systems with regards to the risk classification for surgery.

Methods

Study population

Consecutive patients with severe symptomatic aortic stenosis who underwent a TAVI using the Edwards-Sapien (Edwards Lifesciences, Irving, CA, USA) or the Medtronic CoreValve (Medtronic, Minneapolis, MN, USA) prostheses in our cardiology department were included in a prospective single-centre registry. All procedures were performed through retrograde transfemoral, subclavian, or retroperitoneal or anterograde transapical approaches, under local or general anaesthesia [7].

Patients were considered candidates for TAVI based on a Logistic EuroSCORE ≥ 20% or because of frailty, presence of severe comorbidities or contraindication to surgical aortic valve replacement. The decision to perform a TAVI was validated by our medico-surgical heart team. All patients provided signed informed consent for subsequent data collection and analysis for research purposes.

Clinical and echocardiographic characteristics

All patients were initially screened in order to confirm the diagnosis of severe symptomatic aortic stenosis. Severe aortic stenosis was defined as an aortic valve area < 1.0 cm² and/or a mean transvalvular gradient > 40 mmHg or a peak velocity > 4 m/sec [9]. Data concerning history of coronary artery disease, peripheral artery disease, cerebrovascular stroke, renal impairment or chronic lung disease and previous cardiac surgery were collected. Porcelain aorta was defined as an extensive circumferential calcification of the ascending aorta. A comprehensive transthoracic echocardiography was performed before TAVI.

Risk score calculations

Predicted operative mortality was calculated using the EuroSCORE II, the Logistic EuroSCORE and the STS score. For each patient, the Logistic EuroSCORE [10] and the STS score [11] were calculated prospectively using web-based systems. The EuroSCORE II was calculated retrospectively, based on a prospective data collection of items, using an online calculator [10].

Statistical analysis

Variables are expressed as mean ± standard deviation (SD) or number (%). Comparisons between scores were performed using paired t-tests and linear regressions. The diagnostic value of the EuroSCORE II for the diagnosis of high operative risk (defined as Logistic EuroSCORE ≥ 20% or STS score ≥ 10%) was analyzed. Sensitivity, specificity, and positive and negative predictive values were determined for various thresholds of the EuroSCORE II. A contingency analysis was performed to assess the agreement between the different scores with regards to risk classification for surgery and expressed by kappa values. A high kappa value demonstrates that the risk classification of patients is close or equivalent using the two scoring systems. Statistical analyses were performed using JMP 7 software (SAS Institute, Cary, NC, USA). P < 0.05 was considered statistically significant.

Results

Patient characteristics

From October 2006 to June 2011, 272 consecutive patients underwent a TAVI in the cardiology department of Bichat Hospital, Paris. Demographic, clinical and echocardiographic characteristics of the population are summarized in Table 1. Briefly, mean age was 82 ± 9 years and 44% of patients were female. All patients had severe aortic stenosis. A large proportion of patients had severe symptoms, including 228 patients (84%) in New York Heart Association (NYHA) class III/IV. TAVI was performed using the transfemoral approach in 166 patients (61%), transapical in 88 patients (32%), subclavian in 15 (6%) and retroperitoneal in 3 patients (1%); under general anaesthesia in 202 patients (74%) or local anaesthesia in 70 patients (26%).

Comparison between the EuroSCORE II, the Logistic EuroSCORE and the STS score

The mean Logistic EuroSCORE was 23 ± 14%, STS score 10 ± 9% and EuroSCORE II 9 ± 8%. The EuroSCORE II was significantly lower than the logistic EuroSCORE (P < 0.01), but not significantly different to the STS score (P = 0.10). The mean EuroSCORE II was significantly higher in patients with a Logistic EuroSCORE ≥ 20% (n = 158 [58%]) than in those with a Logistic EuroSCORE < 20% (n = 114) (12% vs. 5%; P = 0.0001). Similarly, the mean EuroSCORE II was significantly higher in patients with an STS score ≥ 10% (n = 101 [37%]) than in those with an STS score < 10% (n = 171) (12% vs. 8%; P = 0.0001). However, the EuroSCORE II was only moderately correlated with the Logistic EuroSCORE (r = 0.61, P < 0.001) (Fig. 1A) and poorly with the STS score (r = 0.25, P < 0.001) (Fig. 1B).

Determination of the EuroSCORE II cut-off value for selecting high-risk patients

Sensitivity, specificity, negative predictive values and positive predictive values of various EuroSCORE II thresholds are presented in Tables 2 and 3. A EuroSCORE II ≥ 7% (observed in 138 patients [51%]) provided the highest diagnostic value (defined as the sum of the sensitivity and specificity) for predicting high operative risk based on a Logistic EuroSCORE ≥ 20% (70% sensitivity and 75% specificity) and an STS score ≥ 10% (69% sensitivity and 60% specificity).
Table 1 Demographic, clinical and echocardiographic characteristics of the study population.

<table>
<thead>
<tr>
<th></th>
<th>All patients (n = 272)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>82 ± 9</td>
</tr>
<tr>
<td>Female</td>
<td>119 (44)</td>
</tr>
<tr>
<td>History of smoking</td>
<td>91 (33)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>204 (75)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>69 (25)</td>
</tr>
<tr>
<td>Insulin-dependent diabetes mellitus</td>
<td>17 (6)</td>
</tr>
<tr>
<td>Hypercholesterolaemia</td>
<td>141 (52)</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>119 (44)</td>
</tr>
<tr>
<td>Prior heart valve surgery</td>
<td>15 (6)</td>
</tr>
<tr>
<td>Prior cerebrovascular accident</td>
<td>26 (10)</td>
</tr>
<tr>
<td>Carotid artery stenosis</td>
<td>50 (18)</td>
</tr>
<tr>
<td>Peripheral artery disease</td>
<td>46 (17)</td>
</tr>
<tr>
<td>Chronic pulmonary obstructive disease</td>
<td>78 (29)</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>103 (38)</td>
</tr>
<tr>
<td>Creatinine clearance (mL/min)</td>
<td>47 ± 24</td>
</tr>
<tr>
<td>Renal failure</td>
<td>80 (29)</td>
</tr>
<tr>
<td>Porcelain aorta</td>
<td>40 (15)</td>
</tr>
<tr>
<td>NYHA class III/IV</td>
<td>228 (84)</td>
</tr>
<tr>
<td>Canadian Cardiovascular Society class ≥ 2</td>
<td>53 (19)</td>
</tr>
<tr>
<td>Mean aortic gradient (mmHg)</td>
<td>51 ± 17</td>
</tr>
<tr>
<td>Aortic valve area (cm²)</td>
<td>0.73 ± 0.19</td>
</tr>
<tr>
<td>Left ventricular ejection fraction &lt; 50%</td>
<td>100 (37)</td>
</tr>
<tr>
<td>Systolic pulmonary artery pressure (mmHg)</td>
<td>50 ± 15</td>
</tr>
</tbody>
</table>

NYHA: New York Heart Association. Data are mean ± standard deviation or number (%).

Table 2 Kappa statistic and contingency analyses between the EuroSCORE II and the Logistic EuroSCORE.

<table>
<thead>
<tr>
<th>EuroSCORE II threshold values (%)</th>
<th>Logistic EuroSCORE ≥ 20%</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kappa values</td>
<td>Similarly classified</td>
</tr>
<tr>
<td>≥ 5</td>
<td>0.41</td>
<td>198 (73)</td>
</tr>
<tr>
<td>≥ 6</td>
<td>0.45</td>
<td>200 (74)</td>
</tr>
<tr>
<td>≥ 7</td>
<td>0.44</td>
<td>196 (72)</td>
</tr>
<tr>
<td>≥ 8</td>
<td>0.43</td>
<td>193 (71)</td>
</tr>
<tr>
<td>≥ 9</td>
<td>0.38</td>
<td>184 (68)</td>
</tr>
<tr>
<td>≥ 10</td>
<td>0.34</td>
<td>177 (65)</td>
</tr>
</tbody>
</table>

EuroSCORE: European System for Cardiac Operative Risk Evaluation. Data are number (%).

Figure 1. Correlations between the European System for Cardiac Operative Risk Evaluation (EuroSCORE II) and the Logistic EuroSCORE (A) and the Society of Thoracic Surgeons score (B).

Agreement between the EuroSCORE II, the Logistic EuroSCORE and the STS score

Using a EuroSCORE II threshold value of ≥ 7% to define high-risk, the agreement with the Logistic EuroSCORE was modest (kappa = 0.44) (Table 2). Contingency analyses showed that 196 patients (72%) were similarly classified as being either...
at high or low operative risk by both the EuroSCORE II and the Logistic EuroSCORE, whereas the risk assessment was different in 76 patients (28%) (Table 2 and Fig. 2). Using the same threshold value of ≥ 7%, the agreement between the EuroSCORE II and the STS score was poor (kappa = 0.27) (Table 3). Contingency analyses showed that only 173 patients (64%) were similarly classified by both the EuroSCORE II and the STS score, whereas risk assessment was different in 99 patients (36%) (Table 3 and Fig. 2). It is worth noting that the agreement between the Logistic EuroSCORE and the STS score using the recommended threshold values of 20% and 10%, respectively, was also poor (kappa = 0.23). The number of similarly classified patients by both the Logistic EuroSCORE and the STS score was only 163 (60%) (Fig. 2).

### Discussion

In this series of patients with severe aortic stenosis referred for TAVI, we found that the mean EuroSCORE II was significantly lower than the Logistic EuroSCORE, and that the correlation between these scoring systems was only moderate. In contrast, although the mean values of the EuroSCORE II and STS score were not significantly different, the correlation between the two scores was poor. A EuroSCORE II cut-off value of 7% corresponded to a Logistic EuroSCORE of 20% or an STS score of 10%. However, only approximately half of our population was considered at high-risk based on these thresholds and, more importantly, the agreement between the EuroSCORE II and the other two scores regarding risk classification (high or low surgical risk) and thus the selection of patients for TAVI was only modest, with 28% (EuroSCORE II vs. Logistic EuroSCORE) to 36% (EuroSCORE II vs. STS score) of patients classified differently.

The assessment of surgical risk has recently gained new interest with the development of TAVI and the need for objective and quantitative criteria to identify patients at high surgical risk and thus to guide clinical decision-making [12]. Multivariable scoring systems are currently the only means of reducing subjectivity of risk estimation. The EuroSCORE was elaborated more than 15 years ago using data from 19,030 patients who underwent cardiac

<table>
<thead>
<tr>
<th>EuroSCORE II threshold values (%)</th>
<th>STS score ≥ 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kappa values</td>
</tr>
<tr>
<td>≥ 5</td>
<td>0.15</td>
</tr>
<tr>
<td>≥ 6</td>
<td>0.23</td>
</tr>
<tr>
<td>≥ 7</td>
<td>0.27</td>
</tr>
<tr>
<td>≥ 8</td>
<td>0.25</td>
</tr>
<tr>
<td>≥ 9</td>
<td>0.22</td>
</tr>
<tr>
<td>≥ 10</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Table 3. Kappa statistic and contingency analyses between the EuroSCORE II and the STS score.

EuroSCORE: European System for Cardiac Operative Risk Evaluation; STS: Society of Thoracic Surgeons. Data are number (%).

![Figure 2](image_url) Proportion of patients considered at high surgical risk based on the European System for Cardiac Operative Risk Evaluation (EuroSCORE) II (≥ 7%), Logistic EuroSCORE (≥ 20%) and Society of Thoracic Surgeons score (≥ 10%) (A) and agreement between the three scoring systems with regard to the risk stratification based on the above-mentioned thresholds (B).
surgery — mainly coronary artery bypass graft (CABG) — in eight European countries \cite{3,4}. The logistic version of the EuroSCORE, published in 2003, provided an improved calibration compared to the original additive model, but is known to markedly overestimate the in-hospital mortality of contemporary surgery \cite{13－15}. The STS score, however, was derived from a larger dataset of patients, operated on in a more recent era, and specifically considers the type of surgery. It also includes more covariates than the Logistic EuroSCORE, is the most detailed risk model available and is updated on a regular basis.

The EuroSCORE II was developed to address some of the limitations of previous scoring systems in order to improve the accuracy for predicting the operative mortality of contemporary cardiac surgery. It was elaborated from a contemporary (2010) large series of patients \((n=22,381)\) who underwent CABG or valvular surgery in similar proportions \cite{8}. The EuroSCORE II incorporates new parameters in its algorithm, e.g. NYHA and Canadian Cardiovascular Society classes. The degree of renal impairment has been further refined, with a classification based on the creatinine clearance level. The ejection fraction, the pulmonary artery systolic pressure and the degree of urgency have also been categorized. New variables, such as reduced mobility due to musculoskeletal dysfunction or insulin-dependent diabetes, have been included. The new version also distinguishes four different types of surgical procedures, with emphasis on the weight of the intervention.

In the present study, we showed that predicted mortality is significantly lower using the EuroSCORE II than using the Logistic EuroSCORE. Previous studies have clearly demonstrated that the Logistic EuroSCORE significantly overestimated the operative mortality and that the mis-calibration was particularly significant in high-risk patients \cite{14,16－19}. The lower EuroSCORE II found in the present study suggest a better calibration (ability to accurately predict the operative mortality) than the Logistic EuroSCORE, and indeed a good calibration and discrimination (ability to differentiate between low- and high-risk patients) have been reported in patients with isolated aortic valve replacement or aortic valve replacement + CABG \cite{20,21}. In contrast, the EuroSCORE II and the STS score were not significantly different, but the correlation was poor, suggesting that they may provide different calibration.

Using cut-off values currently recommended to define high surgical risk, we were able to define a corresponding threshold value for EuroSCORE II. Compared to both the Logistic EuroSCORE and the STS score, a threshold of \(\geq 7\%\) provided the best diagnostic value. To the best of our knowledge, this is the first study that searched for correspondence between these scoring systems and proposes a threshold for high-risk for the EuroSCORE II. Since surgical risk factors such as frailty or porcelain aorta are not adequately represented in all scoring systems, 42\% and 63\% of our patients referred for TAVI did not reach the recommended thresholds (Logistic EuroSCORE \(\geq 20\%\) or STS score \(\geq 10\%), respectively). Similarly, almost half of our patients did not reach the above-mentioned threshold of \(\geq 7\%\) for the EuroSCORE II. More importantly, the agreement between these scoring systems was only modest, and risk assessment was different in 28\% (EuroSCORE II vs. Logistic EuroSCORE) and 36\% (EuroSCORE II vs. STS score) of patients. Thus, not only are current scoring systems only partially capturing factors associated with high surgical risk, they are also selecting different subsets of patients. It is interesting to note that both the logistic EuroSCORE and the STS score were not different in the PARTNER A and PARTNER B trials \cite{22,23}, so between patients considered inoperable and those considered ‘only’ at high-risk but operable (29－30\% for the Logistic EuroSCORE and 11－12\% for the STS score).

In a recent study \cite{20}, the EuroSCORE II performed well in the first tertile of risk (the majority of the study group), but its discrimination dramatically deteriorated for higher tertiles, where its performance was not significantly better than the Logistic EuroSCORE \cite{20}. This poor discrimination of the EuroSCORE II in high-risk patients is a major limitation for risk assessment in candidates for a TAVI.

Our results deserve further validation in larger or multicentre studies, but clearly highlight the limitations of the use of current scoring systems in the setting of TAVI. These systems should not be used as isolated tools, which is in agreement with the recently updated recommendations of the European Society of Cardiology (ESC) and the European Association of Cardio-Thoracic Surgery (EACTS) \cite{2}, which state that ‘in the absence of a perfect quantitative score, the risk assessment should mostly rely on the clinical judgment of the heart team, in addition to the combination of scores.'

The present study has several limitations. First, the three scoring systems were developed to predict perioperative mortality rather than in-hospital mortality after TAVI. Consequently, we did not compare the effective in-hospital mortality after TAVI in our population to the score-predicted mortality. It is well known that the performance of the Logistic EuroSCORE system is poor in TAVI patients \cite{24} and specific risk scores need to be developed. Second, the Logistic EuroSCORE and the STS scores were calculated prospectively whereas the EuroSCORE II was calculated retrospectively. However, all the data needed for its calculation were recorded prospectively in our database. Third, we enrolled patients who underwent a TAVI and were considered at high-risk based on scores and clinical judgment. Therefore, one may argue that our results are not unexpected or tautological. However, as many as half of our population did not qualify for TAVI based on scores alone, and one main result of our study is to show that the three scoring systems (EuroSCORE II, Logistic EuroSCORE and STS score) often disagree and lead to selection of different subsets of patients for TAVI. This disagreement is a major limitation for patient selection in the setting of TAVI. However, we are not implying that scoring systems should be disregarded or abandoned. Rather, risk estimation should be encouraged, but it is important to keep in mind the limitations of current scoring systems in high-risk patients considered for TAVI and these scores should be integrated into — rather than substitute — clinical judgement.

**Conclusions**

In this series of high-risk patients referred for TAVI, correlations between the new EuroSCORE II, the Logistic EuroSCORE and the STS score were at best modest. A EuroSCORE II threshold value of \(\geq 7\%\) corresponded to
a Logistic EuroSCORE \( \geq 20\% \) and an STS score \( \geq 10\% \), but approximately half of our patients did not reach these threshold values and agreements between the three scoring systems was poor. Our results highlight the limits of current scoring systems and reinforce current ESC/EACTS guidelines on the management of valvular heart disease recommending an integrative approach based on a combination of scores and the clinical judgment of a heart team. Specific scores aimed at predicting outcomes after TAVI need to be developed and validated.

**Contributorship statement**


Drafting the article or revising it critically for important intellectual content: D. A., D. H., B. I., A. V., D. M.-Z.


Contributors responsible for the overall content as guarantors: D. A., D. M.-Z.

**Disclosure of interest**

D. A., C. C., S. A., A. C., J.-P. D. declare that they have no conflicts of interest concerning this article.

D. H. proctor for Edwards Lifesciences Inc and Medtronic Inc.

E. B. receives lectures fees from Edwards Lifesciences Inc and Medtronic Inc.

B. I. has received consultant fees from Servier, Boehringer Ingelheim, Bayer, Valtech and Abbott, and speaker’s fees from Edwards Lifesciences Inc, St. Jude Medical and Sanofi-Aventis.

P. N. received proctoring/consultant fees from Edwards Lifesciences Inc and Medtronic Inc.

A. V. member of advisory boards for Medronic, Abbott, Valtech and Boehringer Ingelheim, and has received speaker’s fees from Edwards Lifesciences Inc and Siemens.

D. M.-Z. received consultant/lecture fees from Philips, Valtech, Symetis, Abbott and Edwards Lifesciences Inc.

**Acknowledgements**

None.

**Sources of funding:** the present work was partially funded by the COFRASA and GENERAC studies (Assistance publique—Hôpitaux de Paris; PHRC National 2005 and PHRC regional 2007, respectively).

**References**


