SYNTAX, STS and EuroSCORE – How good are they for risk estimation in atherosclerotic heart disease?

Bernhard Metzler1; Bernhard Winkler2
1Department of Internal Medicine III / Cardiology, Innsbruck Medical University, Austria; 2Department of Cardiology, King’s College, London, UK

Summary
Tests that enable prediction of adverse outcome after surgical or non-surgical intervention in cardiac patients are of great importance since they can help guide clinical decision making. The new evolving percutaneous therapeutic techniques combined with the currently available risk scoring systems require improved prediction models. In the context of steadily improving surgical techniques and perioperative care, on the one hand, and the inadequacy of regional patient data sets to provide generally applicable risk prediction base, on the other, there is need for adaption and recalibration of scoring systems some of which are partly outdated but still widely in use. The accuracy of predictive models depends on their proper application as well as the knowledge of their individual strengths and weaknesses. The EuroSCORE and the STS score take into consideration some risk factors associated with mortality, whereas the SYNTAX score relies solely on coronary anatomy and lesion characteristics. A combination of selected score components from the EuroSCORE, assessing the mortality risk, and those from the SYNTAX score, reflecting the coronary artery disease complexity, can be expected to yield more accurate results in estimating risk in individual patients. In this review, the predictive ability of the SYNTAX score, the STS score and the EuroSCORE will be discussed.

Keywords
Surgery, cardiology, risk factors

Introduction
Current epidemiologic predictions show that the number of people at high cardiovascular risk is increasing dramatically. In Western countries, those without any risk factors comprise only a minority of people; moreover, greater than 70% of at-risk individuals have multiple risk factors. Atherosclerosis, which underlies most cardiovascular diseases (CVD), is rarely the result of one single risk factor, but more usually the result of combined risk factors, such as hypertension, smoking, abnormal lipids, diabetes mellitus and lack of regular physical activity. Cardiovascular risk factors show a continuous association with overall cardiovascular risk with no minimum threshold for disease (1,2). Several models, such as the SCORE, which is recommended by the 2007 European Society of Cardiology (ESC) Guidelines on cardiovascular disease prevention in clinical practice (3,4) have been established to calculate the estimated CVD risk. In daily clinical practice, it is of great importance to have risk stratification methods to identify CVD patients as those suitable for non-surgical or surgical intervention. This has led to the development of a number of risk stratification models or scores in the past few years. With the help of these scores, a heart team comprising a cardiologist and a cardiac surgeon can choose the best management strategy for an individual patient. There are several risk prediction models of which a list comprising 19 was drawn up in a study (5); in Europe the EuroSCORE is the most frequently used surgical score. In general, predictive models can accurately predict only overall mortality in a patient group, but are inaccurate to predict mortality of individual patients (6).

In this review the focus is on the more widely used scores, in particular the European System for Cardiac operative Risk Evaluation (EuroSCORE), the model provided by the Society of Thoracic Surgeons (STS), and the angiographic SYNergy between PCI with TAXUS™ and Cardiac Surgery Score (SYNTAX).

Principle of scores, discriminative power, and accuracy of predictive models
A model can have a good discriminating power even if it is not accurate per se. That means, even if a model underestimates or overestimates the risk of mortality, it still can be reliable in discriminating which of the patients will live and which of them will die (6). The discriminative power for a model, including the EuroSCORE, the SYNTAX score, and the STS score, is often shown with a receiver operating characteristic (ROC) curve. For binary outcomes, the area under the ROC curve is identical to the c-index or c-statistic, with c standing for concordance (7). If the area under the ROC

Thrombosis and Haemostasis 108.6/2012
If a risk score is used for the prediction not only of the short-term but also the one-year mortality, the ROC area will be smaller in the latter case, because the proportion of cardiovascular fatalities among all causes of mortality is usually lower at one year than at 30 days after surgery/intervention (5).

Mortality is the most frequently reported outcome parameter in evaluating heart surgery risk scores. Because morbidity comprises heterogeneous parameters, obviously the predictive values for mortality were considerably higher than for morbidity in these scoring systems (10).

To devise a risk score, most score-databases were randomly divided into two equal parts by using a specific application of a statistical analysis program. The first part is used to develop the risk score. The second part is used to validate the score.

### EuroSCORE

The EuroSCORE was published in 1999 (11). By assessing 17 risk factors (patient-related, cardiac-related, and surgery-related), the predicted postoperative mortality among patients undergoing open heart surgery can be calculated (Table 1). It was developed from a large European data set comprising more than 13,000 patients with a wide range of risk factors associated with mortality (7, 12). The data base was divided into developmental and validation subsets. There are two available methods of calculating predicted outcome with the EuroSCORE: First, the original additive model, and second, the more recent and also more complex one to obtain the logistic model. The additive model allocates points to each risk factor that results in a score used to predict mortality (11, 13). Thresholds were defined to distinguish low risk (EuroSCORE 1–2), medium risk (EuroSCORE 3–5), and high-risk groups (EuroSCORE 6 plus) (11). This easy-to-apply model has some weaknesses, namely possible overprediction of mortality in low-risk cases (14) and poor prediction of mortality in high-risk patients (14–16). Driven by the demand for improved predictive accuracy

### Table 1: Risk factors and numeric values for additive EuroSCORE.

The risk factors, their definitions and the weight allocated to them are shown. The system is additive: to calculate the predicted risk for a patient, the scores for existing risk factors are added to give an approximate percentage predicted mortality figure.

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Descriptor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient-related factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Per 5 years greater than 60</td>
<td>1</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>1</td>
</tr>
<tr>
<td>Chronic pulmonary disease</td>
<td>Use of bronchodilators or steroids</td>
<td>2</td>
</tr>
<tr>
<td>Extra cardiac arteriopathy</td>
<td>Any one more of claudication, carotid occlusion or &gt; 50% stenosis; previous or upcoming intervention for abdominal aorta, limb arteries, or carotids</td>
<td>2</td>
</tr>
<tr>
<td>Neurological disease</td>
<td>Affect ambulation or daily function</td>
<td>2</td>
</tr>
<tr>
<td>Previous cardiac surgery</td>
<td>Requires reopening of pericardium</td>
<td>3</td>
</tr>
<tr>
<td>Serum creatinine</td>
<td>More than 200 μM</td>
<td>2</td>
</tr>
<tr>
<td>Active endocarditis</td>
<td>Receiving antibiotics at time of surgery</td>
<td>3</td>
</tr>
<tr>
<td>Critical preoperative state</td>
<td>Any one or more of ventricular tachycardia, fibrillation, or aborted sudden death; cardiac massage; preoperative Ventilation; inotropic support; intraaortic balloon; or preoperative renal failure</td>
<td>3</td>
</tr>
<tr>
<td><strong>Cardiac-related factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstable angina</td>
<td>Angina at rest</td>
<td>2</td>
</tr>
<tr>
<td>Left ventricle dysfunction</td>
<td>Moderate (LVEF 30% to 50%)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Severe (LVEF &lt;30%)</td>
<td>3</td>
</tr>
<tr>
<td>Recent myocardial infarction</td>
<td>Less than 90 days previous</td>
<td>2</td>
</tr>
<tr>
<td>Pulmonary hypertension</td>
<td>Systolic pulmonary artery pressure &gt;60 mm Hg</td>
<td>2</td>
</tr>
<tr>
<td><strong>Operation-related factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emergency</td>
<td>Carried out on referral before next working day</td>
<td>2</td>
</tr>
<tr>
<td>Other than CABG</td>
<td>Major cardiac procedure in addition to or other than CABG</td>
<td>2</td>
</tr>
<tr>
<td>Surgery on thoracic aorta</td>
<td>Ascending, arch, or descending aorta</td>
<td>3</td>
</tr>
<tr>
<td>Postinfarct septal rupture</td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

EuroSCORE = European System for Cardiac Operative Risk Evaluation; LVEF = left ventricular ejection fraction; CABG = coronary artery bypass grafting. aDerived from the EuroSCORE Web site.
of the EuroSCORE, a logistic model was developed (17). The logistic model is more extensive and requires a computer to derive a score (18). However, even with this more sophisticated model, there are still concerns about overprediction of mortality in many risk groups (13, 19). Curiously, the logistic values are even higher than those of the additive score (20). In general, the EuroSCORE is easy to use and is available as online calculator (www.euroscore.org). Especially in Europe, the EuroSCORE is somewhat the gold standard against which other risk stratification models are measured. Furthermore, it was shown that the EuroSCORE predicts not only the length of stay in intensive care unit, but also the costs for open heart surgery (21).

**STS score**

The STS score was published 2007 (22) and like the EuroSCORE, is an easy-to-use online calculator for predicting the postoperative mortality in patients undergoing open heart surgery (www.sts.org). Similar to the EuroSCORE, the study population was divided into a developmental and a validation group. In contrast to the EuroSCORE, the required data entry is more detailed (Table 2). As in other models, mortality is the primary outcome. However, the STS database has also been used to produce models with other endpoints than mortality such as the calculation of the composite of morbidity and 30-day mortality, length of stay, neurologic injury, deep sternal wound infection, prolonged ventilation, renal failure and reoperation (23). However, information on this detailed outcome is only available for isolated coronary artery bypass graft (CABG) procedures. Published STS mortality estimations are available for isolated CABG (24), valvular surgery (25), and combined procedures (26). In contrast to the EuroSCORE, the STS score seems to slightly underestimate the perioperative risk (27).

**SYNTAX score**

The SYNTAX score is an angiographic tool for grading the complexity of coronary artery disease (28, 29) and has been developed to perform a randomised allcomer study of patients with significant lesions in left main stem and/or three epicardial coronary arteries suitable for both revascularisation strategies. The absence of grading of the severity of coronary artery disease and the lack of comparison of lesion complexity based on pre-treatment angiographic criteria between various patient groups severely limits the interpretation of the results of the earlier trials performed for investigating the difference between CABG and percutaneous coronary intervention (PCI) in multivessel disease (28, 29). An anatomical scoring model addressing these specific issues thus became necessary. The SYNTAX score ranges from 0 to 83 and higher scores indicate a more complex coronary artery disease. For the selection of the optimal revascularisation strategy for patients with three vessel and/or left main stem disease, there are three major requirements (28):

- Performance of an “allcomer” study in such patients
- Consensus between the interventional cardiologist and the cardiothoracic surgeon for the treatment plan
- The quantification of the complexity of coronary artery disease, taking into account not only the number of significant lesions and their location, but also the complexity of each lesion independently.

The SYNTAX score has been developed based on a number of existing classifications such as the AHA classification of the coronary tree segments (30), the Leaman Score (31), the ACC/AHA lesion classification system (32), the total occlusion classification system (33), the Duke and ICPS classification system for bifurcation lesions (34), and in consultation with experts.

The SYNTAX score is calculated by a computer program consisting of sequential and interactive self-guided questions (Table 3). The algorithm consists of 12 main questions that can be divided in two groups: the first three determine the dominance, the total number of vessel segments, and the number of segments involved per lesion, whereas the last nine questions refer to adverse lesion characteristics (Table 4) and are repeated for each lesion.

The SYNTAX score reflects neither patient characteristics nor the treatment strategy, but only the coronary anatomy.
Scores for PCI

Sianos et al. published in 2005 the SYNTAX score as a new angiographic score (28). As already mentioned above, this score relies solely on lesion characteristics and coronary anatomy, but does not take any patient characteristics into account (Tables 3 and 4). This score has been shown to predict cardiac mortality and major adverse cardiovascular events (MACE) in patients undergoing percutaneous revascularisation (35). c-indices for SYNTAX score in terms of cardiac death and MACE were 0.83 and 0.64, respectively (35).

The EuroSCORE risk model, which was established for the prediction of early mortality following open-heart surgery was also recently shown in two independent studies to predict the in-hospital and also the long-term prognosis of patients undergoing multivessel-PCI (36, 37).

Lehmann et al. recently reported that the logistic EuroSCORE even outperforms the SYNTAX score in predicting the long-term survival following multivessel-PCI (38).

The predictive accuracy of the SYNTAX score can be improved by combining the SYNTAX score with a simple clinical risk score incorporating age, ejection fraction, and creatinine clearance to produce the clinical SYNTAX score (39). Herewith the c-statistic for the five-year mortality can be improved from 0.62 with the SYNTAX score to 0.69 with this modified “clinical” SYNTAX score.

A limitation of the clinical application of the SYNTAX score is the moderate intra-observer reproducibility, with a kappa value of 0.51 when patients were grouped into tertiles (40). The kappa value is the degree of agreement beyond the level of chance (moderate: 0.4<kappa≤0.6, almost perfect: 0.8<kappa≤1.0).

Taking into account the above-mentioned findings, it appears that the combination of selected score components from the EuroSCORE, assessing the mortality risk, and those from the SYNTAX score, reflecting the coronary artery disease complexity, could lead to better accuracy in estimating the risk of individual patients (41, 42).

A new EuroHeart score model incorporating 16 independent patient or lesion characteristics for the evaluation of in-hospital mortality in patients undergoing PCI (n=46,064) divided into a training and validation group demonstrated good discriminative power (c-index=0.90) (43).

Scores for CABG

Although a high SYNTAX score is indicative for a worse long-term outcome after PCI, it is not an independent predictor for MACE in patients treated with CABG (29, 44, 45). Holzhey et al. analysed the predictive ability of the SYNTAX score in a retrospective analysis of 200 consecutive in CABG-treated patients. Despite the fact that the SYNTAX score in this study ranged between 2 and 52, a complex coronary pathology did not affect the long-term outcome after CABG in this study. These findings were confirmed by a study analysing a cohort of 320 patients with multivessel disease treated with CABG (46). These results can be explained by the fact that variables describing the coronary anatomy and complexity do not appear to be important for CABG outcome, whereas the strong predictors for cardiovascular events after surgery, such as age, diabetes, left ventricular ejection fraction, history of pulmonary disease, and renal failure are not included in the SYNTAX score.

A recent study found no difference in the performance of the logistic and the additive EuroSCORE in predicting in-hospital mortality in patients undergoing off-pump coronary artery bypass grafting (OPCAB; n=1,140) (47). Also a meta-analysis of eight studies representing more than 19,000 patients revealed a similar performance of the logistic and additive EuroSCORE. Both risk models, however, significantly overestimated mortality (47).

The 2008 STS risk model for coronary artery bypass grafting included a study population consisting of 77,881 isolated CABG procedures, performed on adult patients aged 20 to 100 years. Plots of observed versus predicted event rates revealed acceptable calibration both in the overall population as well as in the numerous subgroups. When patients were grouped into risk categories, the absolute difference between the observed and the expected event rates was less than 1.5% for each endpoint. This 2008 STS model
Scores for valve surgery

Similar to risk assessment in the setting of isolated CABG, and combined CABG and valve surgery-procedures, the STS model was also used to assess risk in each of the three types of isolated valve surgery: aortic valve replacement, mitral valve replacement, or mitral valve repair (25). Using STS data from 2002 to 2006, isolated valve surgery risk models were developed for operative mortality, permanent stroke, renal failure, prolonged ventilation, deep sternal wound infection, reoperation for any reason, a major morbidity or mortality, prolonged postoperative hospital stay, and short postoperative hospital stay (25).

Scores for combined procedures

In a recent analysis, the performance of the EuroSCORE in predicting short- and mid-term mortality in combined aortic valve replacement and coronary artery bypass procedures was investigated (51). In this retrospective study including 233 patients with a mean follow-up period of 2.2 ± 1.7 years, the mean additive and logistic EuroSCOREs were 8.77 and 16.1, respectively. The observed mortality was 9.44%. The area under the ROC curves for additive EuroSCORE was 0.76 and for logistic EuroSCORE was 0.75. So both, additive and logistic EuroSCOREs were accurate in predicting operative mortality in combined procedures. However, only additive EuroSCORE was predictive of mid-term mortality in this patient group.

The 2008 STS model for assessing risk involved in valve plus coronary artery bypass grafting surgery consisted of a study population undergoing 101,661 procedures, including aortic valve replacement plus CABG, mitral valve replacement plus CABG, or mitral valve repair plus CABG between January 1, 2002 and December 31, 2006 (26). The c-index for mortality prediction for the overall valve plus CABG population was 0.75. The same nine endpoints that were analysed by the STS risk model in isolated CABG and the isolated valve intervention were used. These STS models...
are the first to assess risk in valve plus CABG intervention that also include risk prediction of individual major morbidities, composite major morbidity or mortality, and short and prolonged length of stay (26).

**Scores for transcatheter aortic valve implantation (TAVI)**

The relationship between the logistic EuroSCORE and the STS score in predicting mortality in patients implanted with a CoreValve ReValving System was recently described (52). In this two-centre study (Bern-Rotterdam), the observed mortality was 11.1%. Notably, the mean logistic EuroSCORE was three times higher than the mean STS score (20.2 ± 13.9% vs. 6.7 ± 5.8%). The authors concluded that although the STS score outperformed the logistic EuroSCORE, both models had suboptimal discriminatory power. Some concerns arise, if the patient selection for TAVI based on a high EuroSCORE, as it presently frequently is applied, can be suggested, because of its obvious lack of discrimination power (20). This practice can lead to complacency on the part of the treating physician, especially when the score grossly overestimates the actual mortality risk of TAVI patients (53).

**Conclusion**

Accurate prediction of the individual mortality risk for a given cardiac procedure is of increasing importance in daily clinical practice. In particular, since the SYNTAX study has shown identical MACE outcomes in low, as well as in intermediate risk patients, treated with either CABG or PCI, the scoring seems mandatory (45). All of the three scoring systems listed above have different strengths and weaknesses in risk prediction, because they were devised for different procedures and different study populations. In particular, the SYNTAX score is solely an angiographic score relying on coronary anatomy and lesion characteristics, but does not take any patient characteristics into account. A combination of this scoring system with some clinical factors, such as age, ejection fraction, and creatinine clearance can be expected to improve the accuracy of SYNTAX score in individual risk prediction (39).

Fellahi et al. showed that preoperative B-type natriuretic peptide (NT-proBNP) measurement is a strong, independent, and even more accurate predictor of adverse outcome than EuroSCORE in patients undergoing cardiac surgery (54). So further works could now determine the utility of BNP in combination with cardiac risk scores.

Attention was recently drawn to the importance of taking into consideration frailty as an indirect sign of the “biological age” of elderly patients undergoing cardiac surgery (55). Despite the epidemiological differences between (European) countries, an analysis performed with 18,676 patients showed a good-to-excellent discriminative power for the EuroSCORE with an area under the ROC curve between 0.74 and 0.87 for the different European countries (56).

The constant overprediction of mortality, in most of the published studies by a factor of about 2, with the additive and even more with the logistic EuroSCORE, suggest a need for a recalibration or reworking of the EuroSCORE. It must be kept in mind that the EuroSCORE was developed from a data set of patients operated almost one and a half decades ago; in the meantime, there has been a steady improvement in both the surgical techniques employed and the perioperative care of patients. Against this background, the recommendation of Choon et al. of a recalibration ratio of 0.5 of the logistic EuroSCORE appears reasonable (57).

In summary, outdated scores should be constantly recalibrated and adjusted to reflect new findings. A new score combining selected score components from the EuroSCORE, assessing mortality risk, those from the SYNTAX score, reflecting the coronary artery disease complexity, and additional aspects of a frailty-test, could lead to improved accuracy in estimating the individual risk of a specific cardiac intervention in CVD patients.

**Conflicts of interest**

None declared.

**References**


© Schattauer 2012

Thrombosis and Haemostasis 108.6/2012

Scoring in thrombosis and cardiovascular disease

Downloaded from www.thrombosis-online.com on 2018-04-21 | ID: 1001066444 | IP: 54.70.40.11
For personal or educational use only. No other uses without permission. All rights reserved.