Interactive CardioVascular and Thoracic Surgery 16 (2013) 625–629
doi:10.1093/icvts/ivt022 Advance Access publication 12 February 2013

Comparison of original EuroSCORE, EuroSCORE II and STS risk models in a Turkish cardiac surgical cohort[†]

Ayse Gul Kunt[a,*,†], Murat Kurtcephe[b,†], Mete Hidiroglu[a], Levent Cetin[a], Aslihan Kucuker[a], Vedat Bakuy[c], Ahmet Ruchan Akar[d] and Erol Sener[a]

[a] Cardiovascular Clinics, Ataturk Education and Research Hospital, Ankara, Turkey
[b] Department of Computer Engineering, Bilkent University, Ankara, Turkey
[c] Cardiovascular Clinics, Bakirkoy Education and Research Hospital, Istanbul, Turkey
[d] Department of Cardiovascular Surgery, Ankara University School of Medicine, Ankara, Turkey

* Corresponding author. CK-1 Blok, No: 61, Yukarı Yurtçu Mah, Eskişehir Yolu, 06810 Ankara, Turkey. Tel: +90-532-4918912; fax: +90-332-3236723; e-mail: ataturkkvc@yahoo.com.tr (A. Gul Kunt).

Received 13 September 2012; received in revised form 18 December 2012; accepted 28 December 2012

Abstract

OBJECTIVES: The aim of this study was to compare additive and logistic European System for Cardiac Operative Risk Evaluation (EuroSCORE), EuroSCORE II and the Society of Thoracic Surgeons (STS) models in calculating mortality risk in a Turkish cardiac surgical population.

METHODS: The current patient population consisted of 428 patients who underwent isolated coronary artery bypass grafting (CABG) between 2004 and 2012, extracted from the TurkoSCORE database. Observed and predicted mortalities were compared for the additive/logistic EuroSCORE, EuroSCORE II and STS risk calculator. The area under the receiver operating characteristics curve (AUC) values were calculated for these models to compare predictive power.

RESULTS: The mean patient age was 74.5 ± 3.9 years at the time of surgery, and 35.0% were female. For the entire cohort, actual hospital mortality was 7.9% ($n = 34$; 95% confidence interval [CI] 5.4–10.5). However, the additive EuroSCORE-predicted mortality was 6.4% ($P = 0.23$ vs observed; 95% CI 6.2–6.6), logistic EuroSCORE-predicted mortality was 7.9% ($P = 0.98$ vs observed; 95% CI 7.3–8.6), EuroSCORE II- predicted mortality was 1.7% ($P = 0.00$ vs observed; 95% CI 1.6–1.8) and STS predicted mortality was 5.8% ($P = 0.10$ vs observed; 95% CI 5.4–6.2). The mean predictive performance of the analysed models for the entire cohort was fair, with 0.7 (95% CI 0.60–0.79). AUC values for additive EuroSCORE, logistic EuroSCORE, EuroSCORE II and STS risk calculator were 0.70 (95% CI 0.60–0.79), 0.70 (95% CI 0.59–0.80), 0.72 (95% CI 0.62–0.81) and 0.62 (95% CI 0.51–0.73), respectively.

CONCLUSIONS: EuroSCORE II significantly underestimated mortality risk for Turkish cardiac patients, whereas additive and logistic EuroSCORE and STS risk calculators were well calibrated.

Keywords: Coronary artery bypass grafting • Risk prediction model • Mortality • EuroSCORE • Society of Thoracic Surgeons risk calculator

INTRODUCTION

Scoring systems are an essential part of current cardiac surgical practice in assessing operative mortality and morbidity. Changes in cardiac surgery case mix, surgical techniques and clinical outcomes continuously lead investigators to improve and modify currently available risk-stratification systems. All improvements and modifications require further validation tests for different patient populations globally.

A European System for Cardiac Operative Risk Evaluation (EuroSCORE) additive (standard) model from eight European countries was developed between 1995 and 1999 [1, 2] and has gained wide acceptance in Europe, North America [3] and Asia [4, 5]. However, this model generally overestimates mortality in low-risk patients (EuroSCORE ≤6) and underestimates it in high-risk patient groups (EuroSCORE >13) [6]. The logistic EuroSCORE model was developed to improve the predictive performance in high-risk patients [7, 8]. The investigators of the EuroSCORE have continued their work creating a new version called EuroSCORE II [9, 10] to improve the accuracy of the original EuroSCORE for estimating the risk of death after cardiac surgery. The Society of Thoracic Surgeons (STS) database was first established in 1989 to assess operative mortality after cardiac surgery in adults [11–13]. This risk-stratification model has been widely used in North America and comprised over 40 clinical parameters [14].

EuroSCORE and the STS risk algorithms have been compared for coronary bypass grafting [15, 16]. However, similar comparisons

[†]Presented at the 26th Annual Meeting of the European Association for Cardio-Thoracic Surgery, Barcelona, Spain, 27–31 October 2012.
[†]Both authors contributed equally to this work.

© The Author 2013. Published by Oxford University Press on behalf of the European Association for Cardio-Thoracic Surgery. All rights reserved.

ORIGINAL ARTICLE

https://academic.oup.com/icvts/article/16/5/625/683392 by ANKARA user on 11 December 2021

between the EuroSCORE, EuroSCORE II and STS in the patient populations for coronary bypass grafting are limited. We, therefore, aimed to compare these three risk models on a prospectively collected data from the Turkish cardiac surgical population stored in the TurkoSCORE database.

MATERIALS AND METHODS

Data collection

Turkish patients ($n = 428$) undergoing isolated CABG at the Ataturk Education and Research Hospital between June 2004 and March 2012 were extracted from the TurkoSCORE database. Patients who underwent isolated coronary artery bypass grafting and older than 70 years were included in this study. Patients were excluded from the study if concomitant other cardiac procedures were performed. A total of 546 variables for each patient data set including demographic and administrative ($n = 33$), preoperative risk factors and medications ($n = 77$), preoperative evaluations ($n = 112$), laboratory findings including genetic risk factors ($n = 60$), intraoperative ($n = 160$), postoperative data and complications ($n = 59$), follow-up ($n = 32$), and mortality and morbidity ($n = 13$) have been recorded in this database [5]. For all patients already having additive and logistic EuroSCORE and STS values, the new version of EuroSCORE was applied to the original data set. Observed and predicted mortalities were compared for the additive and logistic EuroSCORE, EuroSCORE II and STS risk calculator.

Receiver operating characteristic curve analysis

The area under the receiver operating characteristic (ROC) curve (AUC) is the value used to compare different risk-estimation methods [16]. Each risk-estimation method provides an ordering for the patient population based on the risk of mortality. Then, a ROC curve is obtained based on this ordering. Each ordering results in an AUC value between 0.5 and 1.0 where the 0.5 AUC value indicates a random ranking of the patients being alive and deceased. The higher the AUC value, the better the performance of the risk-assessment method. In this study, AUC values were calculated for additive and logistic EuroSCORE, EuroSCORE II and STS models to compare the predictive performance of these risk-assessment methods.

Statistical analysis

In prevalence analysis, categorical variables were presented as numbers and/or percentages, and continuous variables presented as mean ± standard deviation (SD). In ROC analysis, a non-parametric method was employed to calculate standard error for area under ROC (AUC) and to provide 95% confidence intervals (CIs). One-sample *t*-test was employed during the comparison of observed and predicted mortality rates. A two-sided *P*-value of <0.05 was considered to indicate statistical significance between the observed mortality and the predicted mortality rates. Data were analysed using SPSS 16.0 (SPSS, Inc., Chicago, IL, USA) for Windows (Microsoft Corp., Redmond, WA, USA).

RESULTS

Patients' characteristics

The clinical features of the study group are reported in Table 1. The mean patient age was 74.5 years at the time of surgery. The segment of age between 70–74 years was 13.1% in the Turkish population ($n = 8018$) and 17.9% in the European population

Table 1: Prevalence of risk factors in EuroSCORE, EuroSCORE II and STS risk-stratification algorithms

	EuroSCORE	EuroSCORE II	STS	Mean (SD) or %
Age (years)	√	√	√	74.5 ± 3.9
Gender (female)	√	√	√	35.0%
Chronic pulmonary disease	√	√	√	26.4%
Extracardiac arteriopathy	√	√	√	21.5%
Neurological dysfunction	√	√		2.1%
Previous cardiac surgery	√	√	√	1.9%
Creatinine >200 μmol/l	√	√	√	5.6%
Critical preoperative state	√	√	√	0.1%
Unstable angina	√	√	√	22.7%
Left ventricular function	√	√	√	50.7 ± 12.2
Recent myocardial infarction	√	√		29.7%
Myocardial infarction >8–21 days			√	14.7%
Pulmonary hypertension	√			15.7%
Pulmonary hypertension—moderate		√		12.4%
Emergency	√	√	√	3.70%
Diabetes control—insulin		√	√	14.7%
NHYA class 1		√	√	76.4%
Hypertension			√	60.5%
Three-vessel disease			√	67.3%

STS: Society of Thoracic Surgeons.

Downloaded from https://academic.oup.com/icvts/article/16/5/625/683392 by ANKARA user on 11 December 2021

Table 2: Predicted and observed mortalities by additive and logistic EuroSCORE risk levels for the entire patient cohort

	Patients (deaths)	Observed mortality rate (95% CI)	Predicted mortality rate (95% CI)	Recalibration	*P*-value
EuroSCORE additive					
Low risk (0–3)	26 (1)	3.8 (–4.1–11.8)	3.0 (3.0–3.0)	1.28	0.23
Medium risk (4–6)	219 (9)	4.1 (1.5–6.8)	5.0 (4.9–5.1)	0.83	
High risk (7+)	183 (24)	13.1 (8.3–18.1)	8.6 (8.3–8.8)	1.53	
Total	428 (34)	7.9 (5.4–10.5)	6.4 (6.2–6.6)		
EuroSCORE logistic[a]					
Low risk	143 (5)	3.5 (0.4–6.5)	3.0 (2.9–3.1)	1.15	0.98
Medium risk	143 (7)	4.9 (1.3–8.5)	6.0 (5.7–6.2)	0.82	
High risk	142 (22)	15.5 (9.5–21.5)	14.9 (13.6–16.2)	1.04	
Total	428 (34)	7.9 (5.4–10.5)	8.0 (7.3–8.6)		

[a]Patients were divided into three approximately equal risk tertiles for logistic EuroSCORE analysis.

Table 3: Predicted and observed mortalities by EuroSCORE II and STS risk levels for the entire patient cohort

EuroSCORE II	Patients (deaths)	Observed mortality rate (95% CI)	Predicted mortality rate (95% CI)	Recalibration	*P*-value
Low risk	143 (4)	2.1 (0.1–5.5)	0.9 (0.87–0.91)	3.14	0.00
Medium risk	143 (6)	4.2 (0.9–7.5)	1.4 (1.4–1.5)	2.93	
High risk	142 (24)	16.9 (10.7–23.1)	2.8 (2.6–3.1)	5.94	
Total	428 (34)	7.9 (5.4–10.5)	1.7 (1.6–1.8)		
STS risk model					
Low risk	143 (10)	7.0 (2.8–11.2)	1.8 (1.7–1.9)	3.88	0.10
Medium risk	143 (6)	4.2 (0.9–7.5)	4.9 (4.8–5.1)	0.85	
High risk	142 (18)	12.7 (7.1–18.2)	10.7 (10.1–11.4)	1.18	
Total	428 (34)	7.9 (5.4–10.5)	5.8 (5.4–6.2)		

STS: Society of Thoracic Surgeons.

($n = 19\,030$) and the segment of age between 65–74 years was 31.63% in the STS population ($n = 245\,132$) [2, 5, 13].

Patients' outcomes

The discrepancies between observed and expected mortality rates were apparent in our patient population. For the entire cohort, actual hospital mortality was 7.9% ($n = 34$; 95% CI 5.37–10.5). However, additive predicted mortality was 6.4% ($P = 0.23$ vs observed; 95% CI 6.15–6.59), logistic-predicted mortality was 8.0% ($P = 0.98$ vs observed; 95% CI 7.32–8.61), predicted mortality was 1.7% ($P = 0.00$ vs observed; 95% CI 1.61–1.83) for EuroSCORE II and STS-predicted mortality was 5.8% ($P = 0.10$ vs observed; 95% CI 5.4–6.24). As shown in Tables 2 and 3, EuroSCORE overestimated predicted mortality for the entire cohort.

Receiver operating characteristic analysis

The ROC curves for the entire cohort were given in Fig. 1. The predictive performance of the models for the entire cohort was fair, with a 0.70 (95% CI 0.60–0.79) AUC value for additive EuroSCORE, 0.70 (95% CI 0.59–0.80) AUC value for logistic EuroSCORE, 0.72 (95% CI 0.62–0.81) AUC value for EuroSCORE II and 0.62 (95% CI 0.51–0.73) AUC value for STS risk calculator.

DISCUSSION

This study evaluates the validity and applicability of the current gold standard risk-prediction models namely additive EuroSCORE, logistic EuroSCORE, EuroSCORE II and STS risk calculator in a patient population from a single institution. The results of our study suggest that EuroSCORE II significantly underestimated the mortality risk for Turkish cardiac patients, whereas additive and logistic EuroSCORE and STS risk calculators were well calibrated in this cohort. Although the discriminatory ability of the models was very satisfactory, we are unable to identify whether the failure of the EuroSCORE II model in our study is due to factors unique to Turkish surgical patients or factors newly added in the model.

EuroSCORE has gained wide acceptance in Europe and also in Turkey in predicting in-hospital mortality in cardiac surgery. In fact, the use of the EuroSCORE model in the adult Turkish cardiac surgical population is obligatorily practiced by the national health authority and Turkish Social Security Agency. Additive and logistic EuroSCORE were introduced in 1999 and 2003, respectively [1, 2, 7, 8]. The additive EuroSCORE is designed by using the β coefficients as weights for each risk factor, whereas the logistic EuroSCORE model aimed to improve the predictive performance of high-risk patients with a desire to use the full logistic equation of EuroSCORE rather than the approximation. However, the present study analysed the performance of the additive and logistic EuroSCORE models on a Turkish cardiac surgical cohort that was dependent on the original

ORIGINAL ARTICLE

https://academic.oup.com/icvts/article/16/5/625/683392 by ANKARA user on 11 December 2021

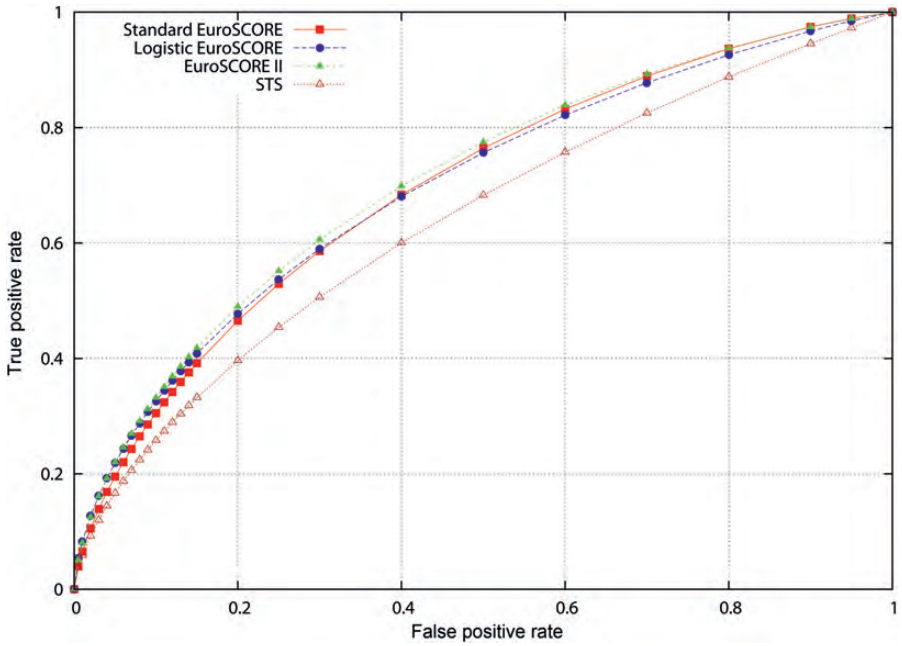


Figure 1: Receiver operating characteristic curve analysis of the entire cohort.

databases. In agreement with the EuroSCORE study group, the additive EuroSCORE model predicted very well in our patient cohort.

Nashef *et al.* [1] analysed 14 799 patients and found that the observed mortality was 4.37 ± 5.06 and predicted mortality was 4.72 ± 4.95 for the entire cohort. The group identified the superiority of the logistic EuroSCORE in predicting mortality in high-risk cardiac surgical patients [7]. They reported that the logistic model was closer to actual mortality than the additive model [4.84% (4.72–4.94) vs 4.21 (4.21–4.26)]. They realized that most of the difference was due to high-risk patients where actual mortality was 11.18% and the predicted was 7.8% (additive) and 11.2% (logistic). However, we previously showed that the original EuroSCORE risk models overestimated mortality at all risk subgroups in a Turkish cardiac surgical population ($n = 9443$) from 2005 to 2010 [5].

Subsequently, Nashef *et al.* [10] updated the EuroSCORE risk model on 22 381 consecutive patients undergoing major cardiac surgery in 154 hospitals in 43 countries over a 12-week period (May–July 2010) to compare with the original 1995 EuroSCORE database. Recently developed EuroSCORE II included additional surgical risk predictors such as poor mobility, diabetes on insulin, New York Heart Association, Canadian Cardiovascular Society Class 4 angina and weight of the intervention; and excluded neurological dysfunction, unstable angina and post-infarct septal rupture. They validated EuroSCORE II on a data subset of 5553 patients (actual mortality: 4.2% and predicted: 4.0%), and very good discrimination was maintained with an area under the ROC curve of 0.81. EuroSCORE II also improved on the original logistic EuroSCORE, though mainly for combined aortic valve replacement and CABG cases in a cohort of patients from UK [17]. In contrast to recent reports, prospectively collected data from the Society for Cardiothoracic Surgery in Great Britain and Ireland database revealed that the model is poorly calibrated for isolated coronary artery bypass grafting and in both the highest- and lowest-risk patients [18]. In our study, the EuroSCORE II significantly underestimated mortality risk in our elderly cardiac cohort (observed: 7.9% and predicted: 1.7%). Recently introduced predictors in EuroSCORE II may have an impact on underestimated mortality risk in our Turkish surgical cohort. The information regarding poor mobility secondary to musculoskeletal dysfunction could have been missed because of the retrospective study design. However, there was no incidence of missing data in this surgical cohort.

The first version of The Society of Thoracic Surgeons National Adult Cardiac Surgery Database (STS NCD) was developed nearly two decades ago. The STS risk calculator predicts morbidity factors such as stroke, renal dysfunction, reoperation, prolonged ventilation, deep sternal wound infection and the length of hospital stay. All STS risk models have undergone periodic revisions. The STS risk model on coronary artery bypass grafting originated from 774 881 isolated CABG procedures from 819 centres performed between 2002 and 2006. Nilsson *et al.* [15] compared EuroSCORE and STS risk algorithms in a cohort of 4497 CABG operations from Sweden. The actual 30-day mortality was 1.9% (95% CI 1.5–2.3%) in their cohort. The investigators reported that the area under the ROC curve was 0.84 (95% CI 0.80–0.88) for EuroSCORE and 0.71 (95% CI 0.66–0.77) for STS; the discriminatory power (area under the ROC curve) was significantly larger for EuroSCORE compared with STS ($P < 0.00005$). They concluded that the additive EuroSCORE algorithm had a significantly better discriminatory power to predict 30-day mortality than the STS risk algorithm for patients undergoing coronary artery bypass in the Swedish population. In our Turkish cardiac cohort, the original EuroSCORE and STS risk models were well calibrated.

The major limitations of this study are the sample size and single-centre design, which limit generalized conclusions for the Turkish cardiac population.

CONCLUSION

In our practice, the original additive and logistic EuroSCORE risk models and STS risk calculator estimated mortality very well in

Downloaded from https://academic.oup.com/icvts/article/16/5/625/683392 by ANKARA user on 11 December 2021

the cardiac Turkish population compared with EuroSCORE II. Ethnicity, seasonal variations and single-centre study should be kept in mind.

Funding

This study was supported by the Turkish Society of Cardiovascular Surgery.

Conflict of interest: none declared.

REFERENCES

[1] Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardiothorac Surg 1999;16:9–13.
[2] Roques F, Nashef SA, Michel P, Gauducheau E, de Vincentiis C, Baudet E *et al*. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. Eur J Cardiothorac Surg 1999;15:816–22; discussion 22–3.
[3] Nashef SA, Roques F, Hammill BG, Peterson ED, Michel P, Grover FL *et al*. Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. Eur J Cardiothorac Surg 2002;22:101–5.
[4] Kawachi Y, Nakashima A, Toshima Y, Arinaga K, Kawano H. Risk stratification analysis of operative mortality in heart and thoracic aorta surgery: comparison between Parsonnet and EuroSCORE additive model. Eur J Cardiothorac Surg 2001;20:961–6.
[5] Akar AR, Kurtcephe M, Sener E, Alhan C, Durdu S, Kunt AG *et al*. Validation of the EuroSCORE risk models in Turkish adult cardiac surgical population. Eur J Cardiothorac Surg 2011;40:730–35.
[6] Shanmugam G, West M, Berg G. Additive and logistic EuroSCORE performance in high risk patients. Interact CardioVasc Thorac Surg 2005;4:299–303.
[7] Michel P, Roques F, Nashef SA. Logistic or additive EuroSCORE for high-risk patients? Eur J Cardiothorac Surg 2003;23:684–7; discussion 87.
[8] Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. Eur Heart J 2003;24:881–2.
[9] Sergeant P, Meuris B, Pettinari M. EuroSCORE II, illum qui est gravitates magni observe. Eur J Cardiothorac Surg 2012;41:729–31.
[10] Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR *et al*. EuroSCORE II. Eur J Cardiothorac Surg 2012;41:734–44; discussion 44–5.
[11] O'Brien SM, Shahian DM, Filardo G, Ferraris VA, Haan CK, Rich JB *et al*. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. Ann Thorac Surg 2009;88:S23–42.
[12] Shahian DM, O'Brien SM, Filardo G, Ferraris VA, Haan CK, Rich JB *et al*. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 3—valve plus coronary artery bypass grafting surgery. Ann Thorac Surg 2009;88:S43–62.
[13] Shahian DM, O'Brien SM, Filardo G, Ferraris VA, Haan CK, Rich JB *et al*. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1—coronary artery bypass grafting surgery. Ann Thorac Surg 2009;88:S2–22.
[14] Ad N, Barnett SD, Speir AM. The performance of the EuroSCORE and the Society of Thoracic Surgeons mortality risk score: the gender factor. Interact CardioVasc Thorac Surg 2007;6:192–5.
[15] Nilsson J, Algotsson L, Hoglund P, Luhrs C, Brandt J. Early mortality in coronary bypass surgery: the EuroSCORE versus The Society of Thoracic Surgeons risk algorithm. Ann Thorac Surg 2004;77:1235–9; discussion 39–40.
[16] Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007;115:928–35.
[17] 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;43:1–7.
[18] 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.

APPENDIX. CONFERENCE DISCUSSION

Dr P. Sergeant *(Leuven, Belgium)*: The authors have created an interesting manuscript, clean and easy to read. The manuscript has a number of positive and negative elements. The authors study hospital mortality; this observation interval is inappropriate for patients who are 80 years of age. When you discharge an 80-year-old, the first pneumonia he or she gets at home will kill the patient. And it is only by observing that interval that we will be able to give the true risk to the patient and inform the patient correctly.

Now, your data set has only 428 records, but (luckily for the patients) there were only 34 events. So this means that all your analysis is based on 34 events. And in an area under the curve analysis, there are a certain number of rules of thumb. Very often the number 50 is used as the minimum number of events needed for a stable area under the curve analysis. So any kind of area under the curve analysis with only 34 events becomes very fragile.

Your average age of the data set is 74 years. Most of us consider an elderly population above 80 and beyond. So the word 'elderly' is rather arbitrarily chosen on the lower line.

Second, the authors identify the very bad discrimination. Indeed, an area under the curve of 0.7 for these scoring systems is not really fair. We have to classify 0.7 as a bad classification system. This would have been very visible if the authors of the scoring systems had used other visualization methods, as, for example, misclassification plots.

The authors cite the lack of calibration of EuroSCORE II and, in fact, that is one of the problems of the Hosmer-Lemeshow test. When it is statistically significant, it is a sign of bad calibration, but when it is not statistically significant, it is no proof of good calibration.

So my proposal to you is for the final version of the manuscript to look into different ways of visualization, outside of the area under the curve and the Hosmer-Lemeshow test. Have you tested other visualization methods to look at calibration and discrimination?

Dr Kunt: No.

Dr Sergeant: Once you start identifying the problem with the area under the curve and the limited number of events, any kind of conclusion becomes very insecure.

https://academic.oup.com/icvts/article/16/5/625/683392 by ANKARA user on 11 December 2021