The American Society of Thoracic Surgery Score versus EuroSCORE I and EuroSCORE II in Israeli Patients Undergoing Cardiac Surgery

Ayelet Shapira-Daniels MD[1]*, Orit Blumenfeld PhD[2]*, Amit Korach MD[1], Ehud Rudis MD[1], Uzi Izhar MD[1] and Oz M. Shapira MD[1]

[1]Department of Cardiothoracic Surgery, Hadassah–Hebrew University Medical Center, Ein Kerem Campus, Jerusalem, Israel
[2]Israel Center for Disease Control, Division of Medical Technologies and Research, Ministry of Health, Sheba Medical Center, Tel Hashomer, Israel

ABSTRACT: **Background:** Recently, Israel established the first national-level adult cardiac surgery database, which was linked to the American Society of Thoracic Surgeons (STS).

Objectives: To validate and compare the STS predicted risk of mortality (PROM) to logistic EuroSCORE I (LESI) and EuroSCORE II (ESII) in Israeli patients undergoing cardiac surgery.

Methods: We retrospectively studied 1279 consecutive patients who underwent cardiac surgeries with a calculable PROM. Data were prospectively entered into our database and used to calculate PROM, LESI, and ESII. Scores were normalized and correlated using linear regression and Pearson's test. To examine model calibration, we plotted the total observed versus expected mortality for each score and across five risk-score subgroups. Model discrimination was assessed by measuring the area under the receiver operating curves.

Results: The observed 30-day operative mortality was 1.95%. The median (IQ1; IQ3) PROM, LESI, and the ESII scores were 1.45% (0.69; 3.22), 4.54% (2.28; 9.27), and 1.88% (1.18; 3.54), respectively, with observed over expected ratios of 0.63 (95% confidence interval [95%CI] 0.42–0.93), 0.59 (95%CI 0.40–0.87), and 0.24 (95%CI 0.17–0.36), respectively, (STS vs. ESII P = 0.36, STS vs. LESI P = 0.0001). There was good correlation among all scores. All models overestimated mortality. Model discrimination was high and similar for all three scores. Model calibration of the STS, PROM, and ESII were more accurate than the LESI, particularly in higher risk subgroups.

Conclusions: All scores overestimated mortality. In Israeli patients, the STS, PROM, and ESII risk-scores were more reliable metrics than LESI, particularly in higher risk patients.

IMAJ 2019; 21: 671–675

KEY WORDS: adult cardiac surgery, risk prediction

Clinical data repository is an essential component of cardiothoracic surgery practice. It allows accurate and fair assessment of clinical performance as well as procedural cost-effectiveness analysis. To achieve these goals, assessment and reporting of outcomes must be risk-adjusted, while considering patient profile and procedural details [1]. Linking of an institutional database to a well-established, large-scale clinical database provides a reliable benchmark and affords the opportunity to utilize sophisticated quality metrics and risk prediction models.

In 2011, the American Society of Thoracic Surgeons (STS) leadership expanded participation in the STS Adult Cardiac Surgery Database (ACSD) internationally to facilitate a worldwide quality collaborative in cardiac surgery [2]. The Department of Cardiothoracic Surgery at Hadassah–Hebrew University Medical Center was among the first international sites to join this collaboration. Recently, the Israeli Society of Cardiothoracic Surgery and Israel Ministry of Health established the first national-level adult cardiac surgery database, linked to that of the STS [3]. The most widely used alternative risk prediction scores for mortality to that of the STS are the European System for Cardiac Operative Risk Evaluation (EuroSCORE) I and II [4,5]. The EuroSCORE I was introduced in 1999 and was based on a dataset from 19,000 patients, most of whom had isolated coronary bypass graft grafting (CABG). The EuroSCORE I can be calculated in additive or logistic fashion using Logistic EuroSCORE I (LESI) [4]. The EuroSCORE II (ESII) was developed in 2012 based on a wider variety of patients and revised definition of its variables [5]. These risk models have been validated in different countries and a variety of cardiac operations with inconsistent results [6-9].

Fundamental social, economic, and cultural differences among the United States, Europe, and Israel, in addition to marked differences in healthcare systems, infrastructure, patient profiles, referral, and practice patterns have all raised concerns in regard to applicability and validity of STS ACSD and the European risk-prediction models and quality assessment tools in Israeli patients. Validation of the risk models is a mandatory step in the process of advancing from an institutional to a national database. Using data we have accumulated over a period of 8 years in our STS ACSD-linked database, we aimed to compare the accuracy of the STS PROM to that of the LESI and ESII.

*The first and second authors contributed equally to this study

Table 1. Baseline clinical and surgical profile of the study cohort

	Entire cohort (n=1279)	Isolated CABG (n=798)	Isolated CABG STS* (n=116, 231)
Age, years	64 ± 12 (median 64, Q1 56, Q3 73)	63 ± 10 (median 63, Q1 56, Q3 71)	64 (median 65, Q1 58, Q3 72)
Gender (male / female)	929 (73%) / 350 (27%)	654 (82%) / 144 (18%)	87,754 (75.5%) / 28,477 (24.5%)
Hypertension	818 (64%)	540 (68%)	103,910 (89.4%)
Diabetes mellitus	553 (43%)	399 (50%)	57,302 (49.3%)
Preoperative stroke	135 (11%)	79 (10%)	9182 (7.9%)
Preoperative myocardial infarction	643 (50%)	524 (66%)	62,765 (54%)
Preoperative renal failure requiring dialysis	40 (3.1%)	17 (2.1%)	3719 (3.2%)
Preoperative peripheral vascular disease	158 (11%)	112 (14%)	16,156 (13.9%)
LVEF (%)	51 ± 12 (median 50, Q1;41, Q3;60)	49 ± 12 (median 50, Q1:40, Q3 :60)	52 (median 55, Q1:45, Q3:60)
Extent of coronary artery disease			
None	286 (22%)	0	0
Single vessel disease	121 (10%)	45 (6%)	4417 (3.8%)
Double vessel disease	253 (20%)	206 (26%)	22,084 (19.0%)
Triple vessel disease	619 (48%)	547 (68%)	88,684 (76.3%)
Left main disease	268 (21%)	239 (30%)	36,613 (31.5%)
Operative procedure			
CABG	798 (62%)		
AVR	214 (17%)		
AVR ± CABG	92 (7%)		
MVR, r	126 (10%)		
MVR, r ± CABG	49 (4%)		
Priority of surgery			
Elective	665 (52%)	271 (34%)	43,470 (37%)
Urgent	606 (47%)	487 (61%)	67,995 (58%)
Emergent or emergency-salvage	8 (1%)	40 (5.0%)	4765 (4%)
Redo surgery	81 (6%)	19 (2%)	2789 (2%)
STS predicted 30-day mortality (%)	3.71 ± 2.20 (median 1.45, Q1 0.69, Q3 3.23)	2.19 ± 4.04% (median 0.98, Q1 0.52, Q3 2.03	
Logistic EuroSCORE I (%)	7.97 ± 9.98 (median 4.54, Q1 2.28, Q3 9.27)		
EuroSCORE II (%)	3.31 ± 4.72 (median 1.88, Q1 1.18, Q3 3.54)		
Observed 30-day mortality	25 patients (1.95%)	11 (1.4%)	2673 (2.3%)

*Data extracted from the STS ACSD 2017 4th quarter report

ACSD = Adult Cardiac Surgery Database, AVR = aortic valve replacement, CABG = coronary bypass graft grafting, LVEF = left ventricular ejection fraction, STS = Society of Thoracic Surgeons

PATIENTS AND METHODS

PATIENTS

The study population consisted of consecutive patients (n=1279) who underwent five cardiac surgical procedures with a calculable STS PROM at Hadassah Hebrew University Medical Center between January of 2008 and December of 2015. The procedures included isolated CABG scores, aortic valve replacement (AVR) with or without CABG, and mitral valve repair or replacement (MVr, MVR), with or without CABG.

DATA COLLECTION

This retrospective study was designed to analyze data that were prospectively collected and entered into our STS-linked departmental database using the STS collection tool and definitions [10]. We used the most recent report of the STS ACSD as a benchmark to compare the baseline characteristics of our study cohort [11]. Procedural mortality was defined as mortality occurring within 30 days after the surgery or during the same hospitalization. Three versions of the STS ACSD collection tool were utilized during the study period: 2.61, 2.73, and 2.81. The STS PROM was updated periodically. We used the most updated version of the risk algorithm to calculate STS PROM for each patient throughout the study. We used the specific risk algorithm for each of the five surgical procedures. We calculated the LESI and the ESII using the publicly available online calculators [12]. The creatinine clearance was calculated using the Cockroft-Gault formula [13]. The information necessary to commplete the variable "poor mobility" for the EuroSCORE calculation was obtained directly from the patient medical records.

STATISTICAL ANALYSIS

Categorical variables were expressed as absolute numbers and percentages. Continuous variables were expressed as means ± standard deviation, median and interquartile (Q1, Q3) range. We categorized the STS PROM into five sub-groups: 0–0.99%, 1.0–1.99%, 2.0–2.99%, 3.0–4.99%, and ≥ 5.0%. For each of the risk models and each sub-group we calculated the observed over expected mortality ratio (O/E ratio). To normalize the distribution of the STS PROM, ESII and LESI, we transformed the data into *ln* and verified it using the Kolmogorov–Smirov test. We then compared the O/E ratios of each model by calculating the exponent of Wald confidence intervals for Poisson regression [14]. OE ratios were expressed as values with 95% confidence intervals (CI). Correlation among the scores was assessed by plotting separate scattered graphs of *ln* of the STS PROM against the *ln*LESI and the *ln*ESII and calculating Pearson's correlation coefficients and linear regression. To assess the discriminative power of each risk model, we plotted the receiver operating characteristic curve (ROC) and calculated the area under the curve (AUC). A *P* value of < 0.05 was considered significant. The statistical analyses were performed using IBM Statistical Package for the Social Sciences statistics software, version 24 (SPSS, IBM Corp, Armonk, NY, USA).

RESULTS

PATIENTS

The study cohort consisted of 1279 consecutive patients who underwent cardiac surgery at our medical center between January 2008 and December 2015. The five surgical procedures with a calculable STS PROM that were included in the study comprised 78% of the operative volume in our institution during the study period. The baseline clinical profile of the study cohort is summarized in Table 1. The clinical profile was typical for a contemporary cohort of patients undergoing cardiac surgery. The average age was 64 years of age (range 20–90) with 73% males. Of the patients, 43% had been diagnosed with diabetes mellitus and 50% sustained prior myocardial infarction. Coronary artery disease was found in 78% of patients–three-vessel disease: 48%, left main disease: 21%. The largest group of our study cohort consisted of patients who underwent isolated CABG (n=798, 62%). The baseline clinical profile of our CABG patients was comparable to that of the STS ACSD [Table 1]. The operative profile of the study cohort is summarized in Table 1. Of the surgeries, 48% were non-elective and 6% were re-do procedures. The operative profile of our patients who had isolated CABG was similar to that of the STS ACSD. The observed CABG mortality in our study was lower than that of the STS ACSD [Table 1].

CORRELATION AMONG THE SCORES

The correlation among the scores is depicted in Figure 1. We observed good correlation between the *ln*STS PROM and both the *ln*LESI (red) and the *ln*ESII (blue). The Pearson correlation coefficients (R values) between the *ln*STS PROM, *ln*LESI and *ln*ESII were of 0.74 and 0.67, respectively, with P values of < 0.001 for both correlation coefficients. We observed higher residuals in between *ln*STS PROM and *ln*LESI compared with *ln*STS PROM and *ln*ESII in Linear regression, suggesting better correlation between the STS PROM and ESII than LESI.

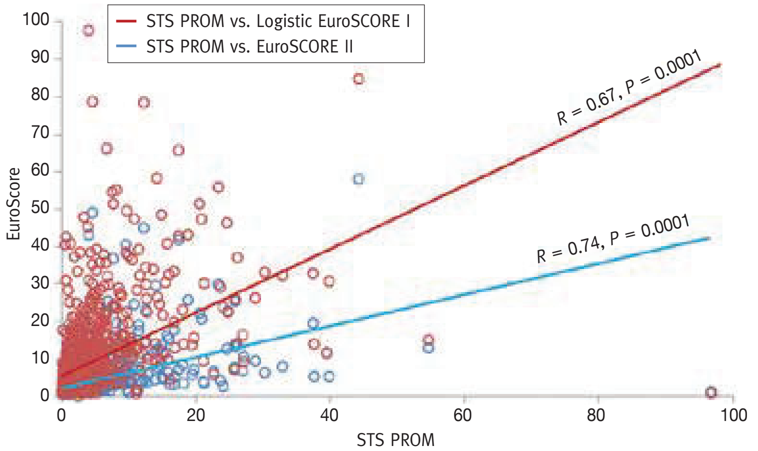
MODEL DISCRIMINATION

To assess model discrimination, we plotted the ROC curves and calculated the AUC for each model. The plots are depicted in Figure 2A. Model discrimination was excellent and similar for all three scores with AUC values of 0.83 (95% confidence interval [95%CI] 0.74–0.91), 0.81 (0.72–0.89), and 0.82 (0.73–0.90) for the STS PROM, LESI, and ESII, respectively ($P = 0.361$).

MODEL CALIBRATION

To examine model calibration, we calculated the observed versus expected mortality for the three models across the entire cohort and the five risk-score subgroups within each model [Table 2, Figure 2B]. The observed 30-day mortality was 1.95% (25 patients). All three models overestimated mortality to a certain extent. Model calibration of the STS PROM and ESII was similar and superior to that of LESI. The LESI model overestimated mortality, particularly in higher risk patients The median (IQ1;IQ3) STS PROM, LESI, and the ESII scores were 1.45% (0.69;3.22), 4.54% (2.28;9.27), and 1.88% (1.18;3.54) respectively. The O/E ratios of the STS PROM, LESI, and ESII were 0.63 (95%CI 0.42–0.93), 0.24 (0.17–0.36), and 0.59 (0.40–0.87), respectively [Table 2]. As shown in Figure 2B, the O/E ratios of the STS PROM and the ESII were similar with overlapping 95%CIs. In contrast, the LESI O/E ratio for the entire cohort was significantly lower compared to those of the STS PROM and ESII. (STS vs. ESII $P = 0.365$, STS vs. LESI $P < 0.001$). [Table 2, Figure 2B].

Figure 1. Correlation among the *ln*STS PROM, *ln*LES I (red) and *ln*ESII (blue). There was a good correlation between the STS PROM and both the LESI and the EUS. There were higher residuals in between the STS PROM and LESI compared with the STS PROM and ESII



PROM = predicted risk of mortality, STS = Society of Thoracic Surgeons

Table 2. Model calibration. Observed versus expected mortality for the entire cohort and the five risk sub-groups

Risk sub-group	Risk model	Observed mortality (%)	Expected mortality (%)	O/E ratio	95% confidence interval
Entire Cohort	STS PROM	1.95	3.12	0.63	0.42–0.93
	LESI	1.95	7.97	0.24	0.17–0.36
	ESII	1.95	3.31	0.59	0.40–0.87
1	STS PROM	0.64	0.56	1.13	0.36–3.50
	LESI	0	0.88	0	
	ESII	0.44	0.77	0.57	0.08–4.03
2	STS PROM	0	1.43	0	
	LESI	0.54	1.57	0.34	0.05–2.44
	ESII	0.67	1.44	0.47	0.15–1.45
3	STS PROM	1.32	2.41	0.53	0.14–2.17
	LESI	0	2.41	0	
	ESII	1.00	2.45	0.41	0.10–1.64
4	STS PROM	3.32	3.89	0.85	0.35–2.03
	LESI	0.71	3.87	0.18	0.04–0.73
	ESII	1.99	3.77	0.53	0.20–1.41
5	STS PROM	7.81	12.10	0.64	0.39–1.07
	LESI	3.69	14.16	0.26	0.17–0.39
	ESII	7.32	10.59	0.69	0.42–1.14

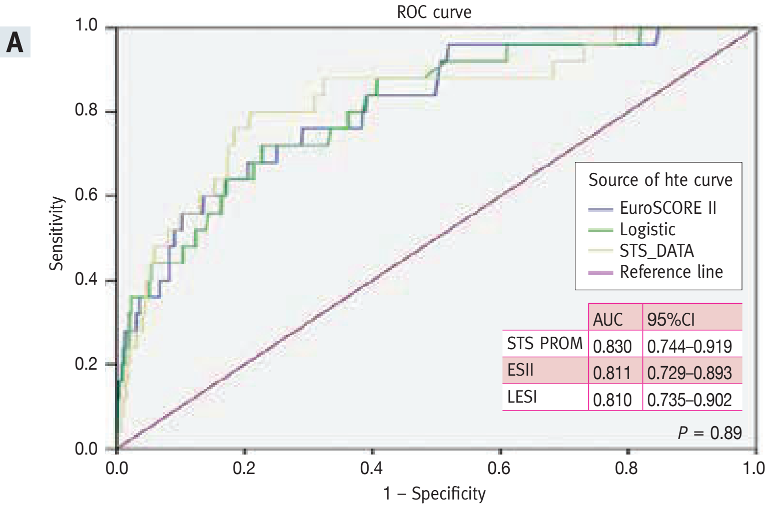
ESII = EuroSCORE II, LESI = logistic EuroSCORE I, O/E = observed over expected, PROM = predicted risk of mortality, STS = Society of Thoracic Surgeons

COMMENT

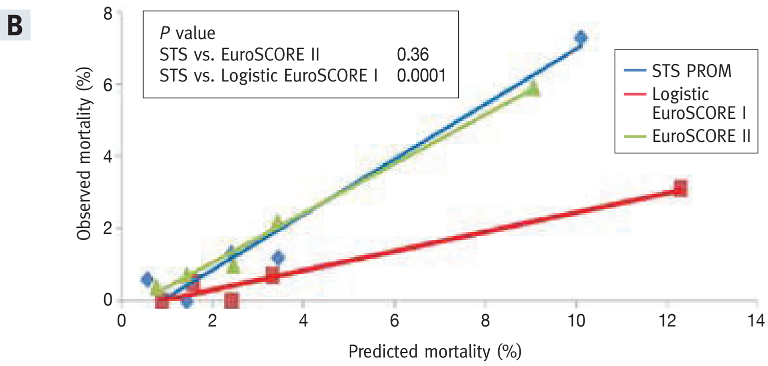
Early postoperative outcomes, particularly procedural risk-adjusted mortality, have become invaluable tools that are used for multiple purposes in cardiac surgery. They are used by the

Figure 2. Model discrimination and model calibration. All three models overestimated mortality. Model calibration was similar for both the STS PROM and the ESII. The LESI model overestimated mortality in higher-risk patients

[A] To assess model discrimination, we plotted the ROC curves and calculated the AUC for each model. Model discrimination was excellent and similar for all three scores



[B] To examine model calibration, we calculated the observed versus expected mortality ratios for the entire cohort and across five risk-score subgroups within each model. Model calibration of the STS PROM and the EuroSCORE II was similar and superior to that of the EuroSCORE I



AUC = area under the curve, ESII = EuroSCORE II, LESI = logistic EuroSCORE I, O/E = observed over expected, PROM = predicted risk of mortality, ROC = receiver operating characteristic, STS = Society of Thoracic Surgeons, 95%CI = 95% confidence interval

surgeon, cardiologist, and other caregivers to assess risk-benefit ratio and for patient selection. They are increasingly included in the discussion with the patient as an essential part of the informed consent process. Finally, operative outcomes are used by all stakeholders in healthcare as a component of procedural cost-effectiveness analysis [15].

Global sharing of medical knowledge among the entire spectrum of healthcare stakeholders is increasingly recognized, not only as a fundamental requisite to improve healthcare, but also as an ethical commitment to our patients [16]. In line with this understanding, and due to interest by growing number of international STS members, the STS expanded its ACSD participation to include international sites in 2012 [2]. Given our small single institutional case volume, we realized that it is essential to link our clinical database to a much more robust database that provides well-recognized and validated risk-prediction models. The most frequently utilized risk prediction models, other than that of the STS and ACSD, are LESI and ESII.

There are important differences between the STS and the two EuroSCOREs. The EuroSCORE models are simple, derived from a limited set of variables and applicable to the entire spectrum of cardiac surgical procedures. However, the EuroSCORE models are based on a fixed and relatively small number of patients operated years ago [4,5]. Therefore, the EuroSCORE models may fail to capture the on-going changes in patient profiles and practice patterns. As a result, the weights of specific variables within the risk prediction models may become inaccurate over time. In contrast, the STS ACSD risk-prediction models are available for only five surgical procedures (although accounting for 70–80% of the average practice), are based on large amount of data, and are updated periodically [15,16]. The most recent version of the STS ACSD risk algorithms were developed based on 775,000 operative records. In addition to mortality, the STS ACSD provides important additional risk-adjusted predictive outcomes scores, including overall morbidity and specific complications affording the opportunity to calculate comprehensive procedural composite scores [16]. In contrast, the two EuroSCORES provide prediction of procedural mortality only [4,5].

We observed a good correlation and model discrimination among the three scores. These findings suggest that all three models performed equally well in the cohort tested. However, one must exercise caution since, by themselves, these two tests may not be sensitive enough when examined exclusively. For example, the AUC methodology to determine model discrimination requires very large cohorts of patients. The AUC methodology may also fail to detect changes between updated and old versions of the same model such as the LESI and ESII [8].

Model calibration assessed by the observed versus expected mortality ratios revealed that all three models overestimated mortality to a certain extent. This observation can be explained simply because the surgeons at Hadassah performed better than the average. Alternative and plausible explanations could be that overestimation of mortality reflects the natural lag between the most updated risk algorithms and the continuous improvement in surgical outcomes. In a sense, it is continuously changing. Supporting this hypothesis and consistent with previous studies [9,17], model calibration was less accurate for the LESI in our cohort. LESI overestimated mortality compared to both the STS PROM and ESII. This difference was more pronounced in higher risk patients. The inferior performance of the LESI may be related to the fact that it is the oldest model, not undergoing periodical updates and based largely on patients who underwent

CABG, whereas, in our study, about one-third of patients had valve surgery. Another pitfall common to all risk assessment models is the small number of patients in the high-risk categories. The small number of patients jeopardizes the reliability of any statistical method at the high end of the spectrum.

LIMITATIONS

An important limitation of the present study is the relatively small and heterogeneous cohort for this type of investigations. The small number of patients precluded separate sub-group analyses with respect to each of the five cardiac surgical procedures. Previous studies demonstrated that performance of the three scores varies considerably across different surgical procedures, suggesting that one risk-prediction model does not fit all situations, even in the same institution [9]. Perhaps the only way to circumvent this problem is to establish a national or, preferably, a truly global clinical database and use the vast amount of data to develop uniform risk prediction models. We used three different versions of the STS ACSD risk-prediction models during the study period. Comparing a continuously updated model against fixed older models may have created bias. In addition, this study reflects a single academic tertiary-care institution experience. Applicability of our findings to other types of institutions in Israel remains to be investigated.

CONCLUSIONS

Our data suggest that although derived from entirely different patient cohorts and healthcare systems, both the STS PROM and the ESII were reasonably accurate and reliable quality metrics in a contemporary cohort of Israeli patients undergoing cardiac surgery. Further multi-center studies on much larger cohorts are required to validate the mortality scores nationally and evaluate additional quality metrics in Israeli patients undergoing cardiac surgery.

Acknowledgments

An abstract of this paper was presented at the 1st STS/EACTS Latin America Cardiovascular Conference, September 2017, Cartagena, Columbia

The authors thank Ms. Rita Dichtiar, Head of the Department of Statistics, Israel Center for Disease Control, for her assistance in the extensive revision of the statistical analyses

The authors thank Ms. Lea Behar for her invaluable assistance ensuring complete and accurate data collection

Correspondence

Dr. O.M. Shapira
Dept. of Cardiothoracic Surgery Hadassah–Hebrew University Medical Center, Ein Kerem, Jerusalem 9112001, Israel
Phone: (972-2) 677-6960
Fax: (972-2) 643-8005
email: ozshapira@hadassah.org.il

References

1. Wasfy JH, Borden WB, Secemaky EA, Mccabe JM, Yeh RW. Public reporting in cardiovascular medicine: accountability, unintended consequences, and promise for improvement. *Circulation*, 2015; 131; 1518-27.
2. Shapira OM, Badhwar V, Shahian D, et al. International participation in the Society of Thoracic Surgeons National Database–Mapping the journey from dream to reality. *Ann Thorac Surg* 2014; 97: 1127-30.
3. Shapira OM, Blumenfeld O, Bolotin G, Grover FL, Shahian DM. International participation in the Society of Thoracic Surgeons Adult Cardiac Surgery Database–From Institutional to National. Editorial, *Ann Thorac Surg* 2017; 103: 1683-6.
4. Roques F, Nashef SAM, Michel P, 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: 816-23.
5. Nashef SA, Roques F, Sharples LD, et al. EuroSCORE II. *Eur J Cardiothorac Surg* 2012; 41: 734-45.
6. Kunt AG, Kurtcephe M, Hidiroglu M, et al. Comparison of original EuroSCORE, EuroSCORE II and STS risk models in a Turkish cardiac surgical cohort. *Interac Cardiovasc Thorac Surg* 2013; 16: 625-9.
7. Osnabrugge RL, Speir AM, Head SJ, et al. Performance of the EuroSCORE II in a large US database: implications for transcatheter aortic valve implantation. *Eur J Cardiothorac Surg* 2014; 46: 400-8.
8. Kieser TM, Rose MS, Head SJ. Comparison of logistic EuroSCORE and EuroSCORE II in predicting operative mortality in 1125 total arterial operations. *Eur J Cardiothorac Surg* 2016; 50: 509-18.
9. Ad N, Holmes SD, Patel J, Pritchard G, Shuman DJ, Halpin L. Comparison of EuroSCORE II, original EuroSCORE, and the Society of Thoracic Surgeons risk score in cardiac surgery patients. *Ann Thorac Surg* 2016; 102: 573-9.
10. The Society of Thoracic Surgeons. Adult Cardiac Surgery Database Data Collection. [Available from http://www.sts.org/sts-national-database/database-managers/adult-cardiac-surgery-database/data-collection#data]. [Accessed 4 December 2017].
11. Duke University Medical Center. DCRI - Registry System. [Available from https://outcomes.dcri.duke.edu/registry]. [Accessed 8 May 2018].
12. EuroSCORE Interactive Calculator. [Available from http://www.euroscore.org/calc.html]. [Accessed 4 December 2017].
13. Cockroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron* 1976; 16: 31-41.
14. Rodríguez, G. Poisson Models for count data. In Lecture Notes on Generalized Linear Models, 9-10. Princeton. University of Princeton New Jersey press, 2007. [Available from http://data.princeton.edu/wws509/notes. [Accessed 25 July 2018].
15. Lehman R. Sharing as the future of medicine. *JAMA Int Med* 2017; 177: 127-8.
16. Shahian DM. The Society of Thoracic Surgeons National Database: "What`s past is prologue". *Ann Thorac Surg* 2016; 101: 841-5.
17. Shanmugam G, West M, Berg G. Additive and Logistic EuroSCORE performance in high risk patients. *Interac Cardiovasc Thorac Surg* 2005; 4: 299-303.

"There are stars whose radiance is visible on Earth though they have long been extinct. There are people whose brilliance continues to light the world though they are no longer among the living. These lights are particularly bright when the night is dark. They light the way for humankind"

Hannah Senesh (1921–1944), poet and Special Operations Executive (SOE) member, one of 37 Jewish SOE recruits from Mandate Palestine parachuted by the British into Yugoslavia during the World War II to assist anti-Nazi forces and ultimately in the rescue of Hungarian Jews about to be deported to the German death camp at Auschwitz