







Cite this article as: Pollari F, Nardi P, Mikus E, Ferraro F, Gemelli M, Franzese I *et al.* Comparison of 4 mortality scores for surgical repair of type A aortic dissection: a multicentre external validation. *Eur J Cardiothorac Surg* 2024; doi:10.1093/ejcts/ezae005.

Comparison of 4 mortality scores for surgical repair of type A aortic dissection: a multicentre external validation

Francesco Pollari ^{a,*}, Paolo Nardi^b, Elisa Mikus^c, Francesco Ferraro^{d,e}, Marco Gemelli^f, Ilaria Franzese^g, Ilaria Chirichilli ^h, Claudia Romagnoniⁱ, Giuseppe Santarpino ^{a,j,k}, Salvatore Nicolardi^l, Roberto Scrofaniⁱ, Francesco Musumeci^h, Enzo Mazzaro^g, Gino Gerosa ^f, Massimo Massetti^{d,e}, Carlo Savini^c, Giovanni Ruvolo^b, Michele Di Mauro ^m, Luca Di Marcoⁿ, Fabio Barili^{o,p,q}, Alessandro Parolari^{r,s,t}, Theodor Fischlein ^{a,t} and on behalf of the GIROC (Italian Research Group on Outcome in Cardiac Surgery)[†]

^a Department of Cardiac Surgery, Klinikum Nürnberg–Paracelsus Medical University, Nuremberg, Germany

^b Department of Cardiac Surgery, Tor Vergata Policlinic University, Tor Vergata University of Rome, Rome, Italy

^c Department of Cardiac Surgery, Maria Cecilia Hospital, GVM Care & Research, Cotignola, Italy

^d Cardiovascular Sciences Department, Agostino Gemelli Foundation Polyclinic IRCCS, Rome, Italy

^e Cardiovascular Department, Catholic University of the Sacred Heart, Rome, Italy

^f Division of Cardiac Surgery, Department of Cardiac, Thoracic and Vascular Sciences, University of Padova, Padua, Italy

^g Division of Cardiac Surgery, Cardiothoracovascular Department, Azienda Sanitaria Universitaria Giuliano Isontina, Trieste, Italy

^h Department of Cardiac Surgery and Heart Transplantation, San Camillo Forlanini Hospital, Rome, Italy

ⁱ Cardiac Surgery Department, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy

^j Department of Cardiac Surgery, Città di Lecce Hospital, GVM Care & Research, Lecce, Italy

^k Cardiac Surgery Unit, Department of Experimental and Clinical Science, Magna Graecia University of Catanzaro, Catanzaro, Italy

^l Cardiac Surgery Unit, "Vito Fazzi" Hospital, Lecce, Italy

^m Department of Cardio-Thoracic Surgery, Cardiovascular Research Institute Maastricht, Maastricht, Netherlands

ⁿ Division of Cardiac Surgery, IRCCS, Azienda Ospedaliero Universitaria di Bologna, Bologna, Italy

^o Department of Biomedical and Clinical Sciences, Università degli Studi di Milano, Milan, Italy

^p University Cardiac Surgery Unit, IRCCS Ospedale Galeazzi Sant'Ambrogio, Milan, Italy

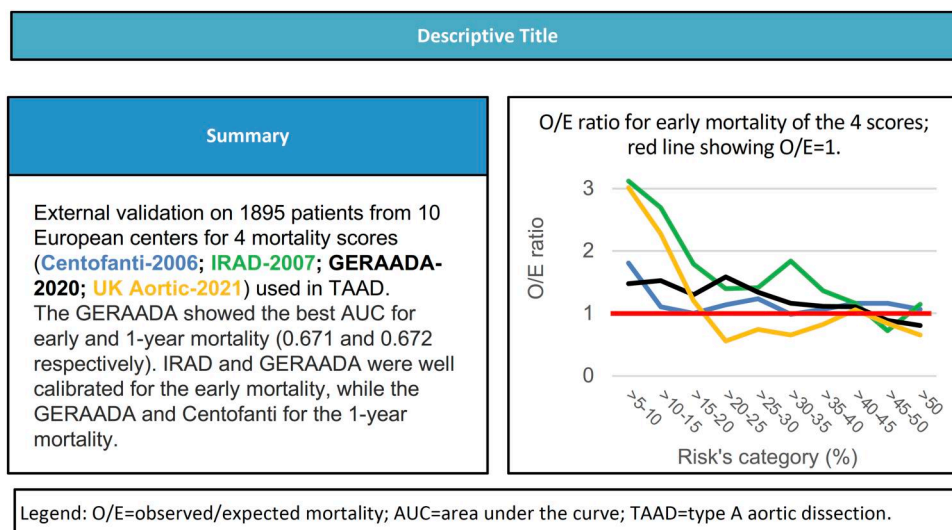
^q Harvard TH Chan School of Public Health, Boston, MA, USA

^r University Unit of Cardiac Surgery, IRCCS Policlinico S. Donato, S. Donato Milanese, Italy

^s Department of Biomedical Sciences, University of Milano, Milano, Italy

* Corresponding author. Department of Cardiac Surgery, Klinikum Nürnberg–Paracelsus Medical University, Breslauer Strasse 201, Nuremberg 90471, Germany. Tel: +49-911-3985441; fax: +49-911-3985443; e-mail: fpollari@gmail.com (F. Pollari).

Received 20 September 2023; received in revised form 21 December 2023; accepted 9 January 2024



[†]These authors should be considered both senior authors.

[†]A complete list of investigators in GIROC (Italian Research Group on Outcome in Cardiac Surgery) is available in the Acknowledgements section. Presented at the 37th EACTS Annual Meeting, Vienna, Austria, 5 October 2023.

Abstract

OBJECTIVES: In the last decades, 4 different scores for the prediction of mortality following surgery for type A acute aortic dissection (TAAD) were proposed. We aimed to validate these scores in a large external multicentre cohort.

METHODS: We retrospectively analysed patients who underwent surgery for TAAD between 2000 and 2020. Patients were enrolled from 10 centres from 2 European countries. Outcomes were the early (30-day and/or in-hospital) and 1-year mortality. Discrimination, calibration and observed/expected (O/E) ratio were evaluated.

RESULTS: A total of 1895 patients (31.7% females, mean age 63.72 ± 12.8 years) were included in the study. Thirty-day mortality and in-hospital mortality were 21.7% ($n = 412$) and 22.5% ($n = 427$) respectively. The German Registry of Acute Aortic Dissection Type A (GERAADA) score shows to have the best discrimination [area under the curve (AUC) 0.671 and 0.672] in predicting as well the early and the 1-year mortality, followed by the International Registry of Acute Aortic Dissection (IRAD) model 1 (AUC 0.658 and 0.672), the Centofanti (AUC 0.645 and 0.66) and the UK aortic score (AUC 0.549 and 0.563). According to Hosmer–Lemeshow and Brier tests, the IRAD model I and GERAADA, respectively, were well calibrated for the early mortality, while the GERAADA and Centofanti for the 1-year mortality. The O/E analysis showed a marked underestimation for patients labelled as low-risk for UK aortic score and IRAD model I for both outcomes.

CONCLUSIONS: The GERAADA score showed the best performance in comparison with other scores. However, none of them achieved together a fair discrimination and a good calibration for predicting either the early or the 1-year mortality.

Keywords: Scores • Outcomes • Aortic dissection • Long-term survival

ABBREVIATIONS

AUC	Area under the curve
CI	Confidence interval
IQ	Interquartile
IRAD	International Registry of Acute Aortic Dissection
O/E	Observed/expected
ROC	Receiver operating characteristic
TAAD	Type A acute aortic dissection

INTRODUCTION

Although the clear survival advantage in those patients who receive an emergency treatment, and the technological and anesthesiology improvements of last years, the surgical treatment of the type A aortic dissection (TAAD) is still burdened by a high mortality (16.9–23.9%) and morbidity rate [1–3], as well as a high resource consumption. This fact did trigger in the last decades the research for identifying risk factors and developing many different scores to predict the outcome. Centofanti *et al.* [4] in 2006, considering the relatively high (42%) survival rate of those patients treated conservatively, aimed to assess the risk of mortality for surgical treatment. Only 1 year later, Rampoldi *et al.* [3]—using the International Registry of Acute Aortic Dissection (IRAD) database—created a bedside tool helpful to assist surgeons who are considering whether or not to proceed with surgical correction in patients presenting with extreme clinical conditions. Recently, groups coordinated by Benedetto *et al.* [5] and Czerny *et al.* [2] presented 2 different scores derived from their national register, respectively, from the UK and Germany. The external validation of the above-mentioned scores is few, limited to single-centre experience or not reported at all. The aim of the present study was to perform a large external multicentre validation analysis of all above-mentioned scores specifically designed for the prediction of mortality in patients undergoing surgical repair of a TAAD.

METHODS

Ethics statement

This study was first approved by the Ethics Committee of Trieste (Comitato Etico Unico Regionale del Friuli Venezia Giulia, CEUR FVG; reference number CEUR-2022-Os-78 on the 26 April 2022, Trieste, Italy) and later approved by the local institutional study centres and was conducted in accordance with the principles of the Declaration of Helsinki. A written consent was waived from patients due to the retrospective nature of the study, and a verbal consent was obtained in case of telephonic follow-up.

Study design

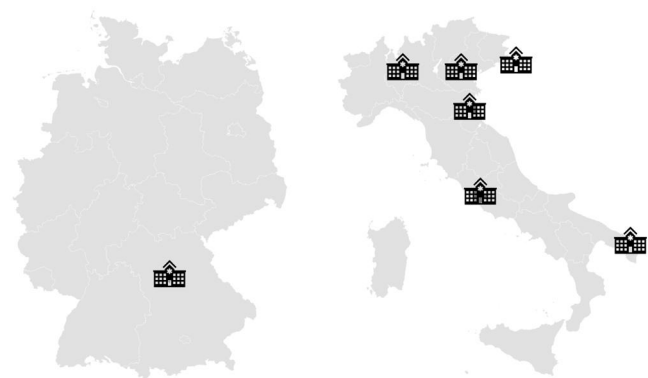
We conducted a multicentre international retrospective observational study of all consecutive patients who underwent aortic surgical replacement for type A acute aortic dissection between January 2000 and December 2020.

The study was proposed by the Klinikum Nürnberg (Germany) to the Italian Research Group on the Outcomes in Cardiac Surgery ('Gruppo Italiano di Ricerca sugli Outcome in Cardiocirurgia'; 'GIROC') and approved after revision and positive approval of the Ethic Committee. Included were all patients admitted in hospital because of a type A acute aortic dissection who underwent surgical repair. Details regarding GIROC, steering committee and study conduction are reported in the [supplementary material](#).

Exclusion criteria were patients participating to the database of the analysed scores, lacking or not retrievable data (regarding the primary outcome and variables needed for calculation of the 4 scores, see Table 1), pregnancy, sub-acute and chronic aortic dissection according to guidelines definition [1], dissections as acute complication of an elective surgical procedure. Finally, 10 centres from 7 cities located in Italy and Germany contributed with their data to the study population (Fig. 1). Operative techniques were at discretion of the single centres, however, always in accordance with the international guidelines [1].

Table 1: Description of variables and measured outcomes of the 4 scores analysed

GERAADA score	UK aortic score	Centofanti score	IRAD score I
Outcome			
30-Day mortality	In-hospital mortality	30-Day mortality	In-hospital mortality
Variables			
Age	Age	Age	Age
Sex	Malperfusion (any)	Shock	History of valve replacement
Resuscitation before surgery	Left ventricular ejection fraction	Redo	Shock
Previous cardiac surgery	Previous cardiac surgery	Coma	Migrating chest pain
Intubation/ventilation at referral	Preoperative ventilation	Acute renal failure	Tamponade
Catecholamines at referral	Preoperative resuscitation	-	Any pulse deficit
Aortic valve regurgitation	Need for intraoperative coronary artery bypass grafting	-	Electrocardiography's signs for infarct
Preoperative organ malperfusion (number of organs)	-	-	-
Preoperative hemiparesis	-	-	-
Extension of dissection	-	-	-
Location of primary entry tear within aortic arch	-	-	-

**Figure 1:** Geographical origin of the 10 participating centres from Italy and Germany.

Outcomes

The primary outcome of this retrospective analysis was the early mortality. This outcome was defined as death for all causes occurred during the hospital stay following the index procedure, i.e. in-hospital mortality, and/or after 30 days. All scores but the German Registry of Acute Aortic Dissection Type A (GERAADA) were calculated automatically starting from the published information, while the GERAADA score was calculated manually on the internet site of the Germany society [6].

The secondary outcomes were the length of stay in intensive care unit and in-hospital, as well as the 1-year mortality (defined as whole all-cause death at 1 year following the operative day). Lengths of stay were calculated starting from the admission date, independently if the surgery was started or ended at that day, up to the discharge for any cause, i.e. recovery or death. A follow-up, telephonic or using the records of outpatient's department, was attempted to assess the survival in all patients discharged alive.

Statistical analysis

Data were checked for normal distribution using the Shapiro test. Normally distributed continuous variables were presented as the mean \pm standard deviation, or as median and interquartile

(IQ) range. Categorical variables were presented as percentages. Intergroup comparisons were performed with 2-tailed Student's *t*-test or Mann-Whitney *U*-test for the continuous variables, or χ^2 test for the categorical variables. Correlation between continuous variables was explored with Pearson's test. The survival rate was calculated using the Kaplan-Meier method combined with the log-rank test. A *P*-value of <0.05 was considered significant.

The 4 aortic scores have been retrospectively calculated for every single patient. The ratio between observed/expected (O/E) death ratio has been calculated for every score, dividing the observed number of deaths (O) by the expected number of deaths (E); O/E ratios significantly >1 indicated a higher than expected mortality, whereas ratios <1 indicated a lower than expected mortality. Comparison of observed distribution with expected distribution was performed by binomial test. The predicting value of the scores was tested by creating receiver operating characteristic (ROC) curve. The area under the curve (AUC) values were used as measurement of the discriminative power: an AUC of 0.5 was considered as no discrimination, of 0.6–0.7 poor, 0.7–0.8 fair and >0.8 good. The goodness of fit (calibration) was assessed with the Hosmer-Lemeshow goodness-of-fit test. A large value of Chi-squared (with small *P*-value < 0.05) indicated poor fit and small Chi-squared values (with larger *P*-value) indicated a good logistic regression model fit. The Brier test (quadratic difference between predicted probability and observed outcome for each patient; the lower, the better, with 0 representing the perfect prediction) was also performed. Moreover, the calibration performance was evaluated by generating calibration plots that visually compared the prediction with the observed probability. The perfect calibrated predictions stay on the 45° line, while a curve below or above the diagonal, respectively, reflected the overestimation and the underestimation, respectively.

RESULTS

One thousand eight hundred ninety-five patients (31.7% females) were included in the analysis. The exact distribution of patients, outcomes and scores according to each centre are shown in [Supplementary Material, Fig. S1](#). Although differences in the number of patients, mortality rate and predicted scores

were homogeneous among the participating centres. The median age was 65 years (IQ, 55–74). All patients underwent emergency surgery, via a median sternotomy, because of an acute aortic dissection involving the ascending aorta (DeBakey type I or II). Baseline characteristics, predicted mortality, intraoperative and postoperative variables are shown in Table 2. In 118 patients (6.2%), death occurred during surgery or in the first 24 h following the procedure. Overall, 30-day mortality and in-hospital mortality were 21.7% ($n=412$) and 22.5% ($n=427$), respectively.

Primary outcome

The mortality rate prediction differed between the 4 scores, varying from a median value of 7.45% [IQ, 3.9–12.45] of the IRAD score model I to 24.3% [IQ, 15.96–32] of the UK aortic score. As shown in Table 2, each score predicted the outcome with a range of approximately between 3% and 90%. The O/E in-hospital and 30-day mortality ratios showed the Centofanti's and GERAADA scores having the best concordance throughout the categories (Supplementary Material, Table S1 and Supplementary Material, Fig. S2). In particular, the O/E ratio analysis in subcategories revealed a significant underestimation for patients classified as low-risk by IRAD score and UK aortic score.

Results of the ROC analysis and calibration tests are shown in Fig. 2 and Table 3. Results of the AUC substantially coincide between 30-day, in-hospital and early mortality. Comparing the 4 scores, the GERAADA score shows to have the best discrimination in predicting these 3 outcomes, followed by IRAD, Centofanti and UK aortic scores. However, the AUC showed a poor discrimination in the best case (0.679).

The calibration curves of all scores are shown in Fig. 3. The calibration analysis showed the GERAADA score having the better calibration in comparison with the other scores based on the Brier test; however, the Hosmer–Lemeshow test showed a better calibration for the IRAD model I.

Secondary outcomes

The overall median length of the follow-up was 1.46 years (IQ, 0.05–3.55), with a maximum follow-up period of 21.92 years. Of the 1440 patients who did not die neither in-hospital nor at 30 days of follow-up, 120 were not reachable, thus resulting in 1320 patients. Of these, 49 patients (3.7%) were died during the first year after the index procedure. Considering the whole population, the Kaplan–Meier curve (Supplementary Material, Fig. S3) shows an early steeper slope, indicating the high perioperative mortality rate, with a survival rate 88.1% [95% confidence interval (CI), 86.5–89.5]. Then, the curve has a flat area, indicating a period of relatively stable survival between 6 months (73.6%, 95% CI, 71.5–75.6) and 1 year after the index procedure (72.7%, 95% CI, 70.6–74.7).

As already shown for the primary outcome, also at 1 year, the O/E ratio revealed a significant underestimation for patients classified as at low risk by the IRAD and UK scores (Supplementary Material, Fig. S4), with the Centofanti's scores having the best concordance throughout the categories. Results of the ROC analysis and calibration tests are shown in Figs 2 and 3 and in Table 3. At 1 year, all but the UK aortic score reached a poor discrimination, with the GERAADA and IRAD showing the better AUC (0.672). According to the Hosmer–Lemeshow, the

Table 2. Baseline, intraoperative and postoperative characteristics of study population

Study population (N = 1895 patients)		
	Mean ± SD	Median [IQ]
Baseline characteristics		
Age (years)		65 [55–74]
Female, n (%)	600 (31.7%)	
Body mass index		26.54 [24.2–29.4]
Body surface area (m ²)	1.96 ± 0.23	
Marfan syndrome, n (%)	60 (3.1%)	
Other connective tissue disease, n (%)	20 (1%)	
Neurological deficit at admission, n (%)	518 (27.3%)	
Baseline serum creatinine (mg/dl)		1 [0.84–1.3]
Baseline dialysis, n (%)	16 (1%)	
Prior cardiac surgery, n (%)	101 (5%)	
Tamponade, n (%)	302 (16%)	
Resuscitation before surgery, n (%)	48 (2.5%)	
Preoperative shock status, n (%)	363 (19%)	
Aortic dissection DeBakey type 1, n (%)	828 (43.7%)	
GERAADA score (%)		14.30 [10.10–22.10]
UK aortic score (%)		24.3 [15.96–32]
Centofanti's score (%)		20.16 [16.1–25.95]
IRAD score model I (%)		7.45 [3.9–12.45]
Intraoperative characteristics		
Arterial cannulation site, n (%)		
Right axillary artery	655 (34.5%)	
Bilateral axillary artery	1 (0.05%)	
Carotid artery	39 (2%)	
Femoral artery	760 (40%)	
Innominate artery	173 (9.1%)	
Aortic arch	118 (6.2%)	
Combination of above	149 (7.8%)	
Extension of surgical repair, n (%)		
Aortic valve replacement	407 (21.5%)	
Aortic arch replacement	795 (41.9%)	
Frozen elephant trunk	43 (2.7%)	
Cardiopulmonary bypass time (min)		185 [139–24]
X-clamp time (min)		104 [77–142]
Early outcomes		
ICU stay (days)		5 [3–10]
In-hospital stay (days)		13 [9–21]
Intraoperative mortality, n (%)	118 (6.2%)	
In-hospital mortality, n (%)	427 (22.5%)	
30-Day mortality, n (%)	412 (21.7%)	

ICU: intensive care unit; IQ: interquartile; IRAD: International Registry of Acute Aortic Dissection; SD: standard deviation.

Centofanti showed the better calibration, while the Brier test showed similar results in Centofanti's and GERAADA's scores.

The overall median length of in-intensive care unit and in-hospital stay was 5 days (IQ, 3–10) and 13 days (IQ, 9–21), respectively. All scores showed a poor correlation with both hospitalizations (GERAADA $r=0.132$ and 0.075 ; UK aortic score $r=0.0682$ and 0.023 ; Centofanti $r=0.12$ and 0.019 ; IRAD model 1 $r=0.075$ and 0.05).

DISCUSSION

To assess whether a prediction model is accurate, different validation strategies, such as internal, temporal and external validation, can be distinguished [7]. As the performance of prediction models is generally poorer in new patients than in the

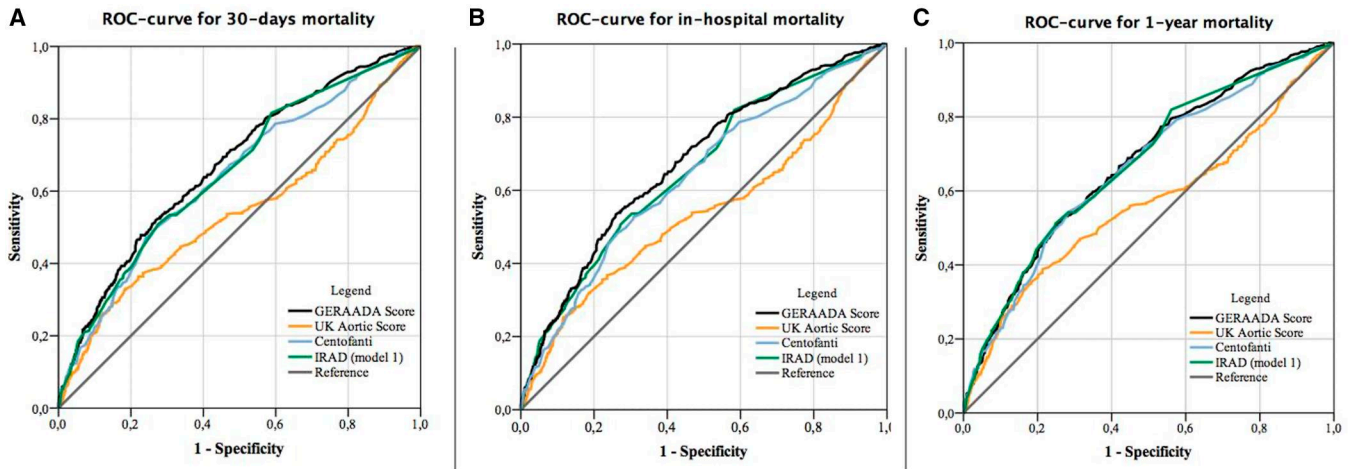


Figure 2: Receiver operating characteristic (ROC) curves for the 4 scores. (A) Referring to the 30-day mortality; (B) to the in-hospital mortality; and (C) to the 1-year mortality. Values of the area under the curves are expressed in Table 3.

Table 3: Area under the curve (with 95% CI) and calibration tests for the outcomes

	Area under the curve	Hosmer-Lemeshow tes	Brier test
30-Day mortality			
GERAADA	0.671 [0.64–0.7]	Chi-square 15.066, <i>P</i> -value = 0.05588	0.162 (0.151–0.172), <i>P</i> < 0.001
UK aortic score	0.541 [0.51–0.58]	Chi-square 54.478, <i>P</i> -value = <0.001	0.171 (0.160–0.181), <i>P</i> = 0.042
Centofanti	0.64 [0.61–0.67]	Chi-square 13.643, <i>P</i> -value = 0.02956	0.164(0.154–0.175), <i>P</i> < 0.001
IRAD model 1	0.652 [0.62–0.68]	Chi-square 7.7945, <i>P</i> -value = 0.4538	0.164(0.153–0.174), <i>P</i> < 0.001
In-hospital mortality			
GERAADA	0.676 [0.648–0.705]	Chi-square 23.691, <i>P</i> -value = 0.002582	0.164 (0.153–0.174), <i>P</i> < 0.001
UK aortic score	0.537 [0.503–0.571]	Chi-square 53.78, <i>P</i> -value = <0.001	0.173 (0.163–0.183), <i>P</i> = 0.061
Centofanti	0.636 [0.606–0.667]	Chi-square 12.449, <i>P</i> -value = 0.03623	0.167 (0.157–0.177), <i>P</i> < 0.001
IRAD model 1	0.656 [0.627–0.685]	Chi-square 7.7268, <i>P</i> -value = 0.4606	0.165 (0.155–0.175), <i>P</i> < 0.001
Early mortality			
GERAADA	0.679 [0.651–0.708]	Chi-square 24.837, <i>P</i> -value = 0.001657	0.171 (0.161–0.181), <i>P</i> < 0.001
UK aortic score	0.549 [0.516–0.582]	Chi-square 54.133, <i>P</i> -value = <0.0001	0.180 (0.170–0.190), <i>P</i> = 0.025
Centofanti	0.645 [0.615–0.674]	Chi-square 12.707, <i>P</i> -value = 0.1223	0.174 (0.164–0.184), <i>P</i> < 0.001
IRAD model 1	0.658 [0.629–0.686]	Chi-square 8.1737, <i>P</i> -value = 0.4167	0.172 (0.162–0.182), <i>P</i> < 0.001
1-Year mortality			
GERAADA	0.672 [0.644–0.700]	Chi-square 22.087, <i>P</i> -value = 0.004758	0.190 (0.180–0.199), <i>P</i> < 0.001
UK aortic score	0.563 [0.531–0.595]	Chi-square 60.651, <i>P</i> -value = <0.0001	0.199 (0.190–0.209), <i>P</i> = 0.00584
Centofanti	0.660 [0.632–0.688]	Chi-square 14.200, <i>P</i> -value = 0.07671	0.190 (0.181–0.200), <i>P</i> < 0.001
IRAD model 1	0.672 [0.645–0.699]	Chi-square 21.429, <i>P</i> -value = 0.00609	0.191 (0.182–0.201), <i>P</i> < 0.001

CI: confidence interval; IRAD: International Registry of Acute Aortic Dissection.

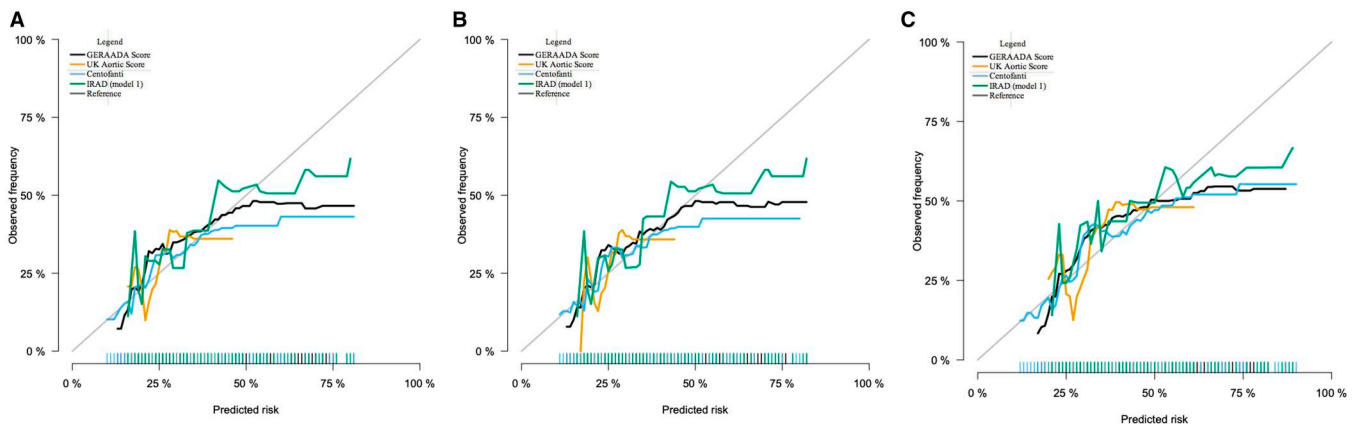


Figure 3: Calibration plots of GERAADA, UK aortic score, Centofanti’s score and IRAD model 1. The diagonal line at 45 degrees represents the perfect calibration. (A) 30-Day mortality; (B) in-hospital mortality; and (C) 1-year mortality.

developed population, models should not be recommended for clinical use before their external validity is established [8]. If we base clinical decisions on incorrect prediction models, this could have adverse effects on various patient outcomes. Moreover, external validation of existing models may help bridge the gap between the development and implementation of prediction models. External validation is necessary to assess a model's reproducibility (new individuals that are similar to the development population) and transportability (exploring whether the prediction tool is transportable to a separate population with different patient characteristics). This can therefore help to answer the question of whether the use of a score is justified in other countries, which perhaps do not have national databases suitable for generating their own risk predictive model.

Our external validation shows that none of the 4 scores achieved together a fair discrimination and a good calibration for predicting either the early or the late mortality. Comparing all the parameters, the GERAADA score seems to achieve the best performances in comparison with Centofanti, UK aortic, and IRAD scores. Especially, the latter 2 showed a marked underestimation in the patients labelled as at 'low-risk', with an observed mortality rate twice as predicted. However, even if better, also the GERAADA showed some limitation, such as an underestimation in low-risk-labelled patients and an overestimation in high-risk-labelled patients (>50%). Possible explanations may concern the presence of unbalanced data [9]. Indeed, when the data are highly imbalanced, the unequal distribution of classes will compound the problem of overlap and make classification (i.e. prediction) even more challenging. A hypothesis is that the number of patients at highest predicted risk was not included in the GERAADA register, because not operated at all.

By the comparison of the 4 scores, it should be highlighted that these were designed to assess different outcomes: the GERAADA and Centofanti scores, for example measure the 30-day mortality, while the other 2 the in-hospital mortality. For this reason, alongside the canonical 30-day and in-hospital mortality, we considered the early and 1-year mortality.

The regional differences, which theoretically could have biased these findings, are to be discussed. Each of the 4 scores has been developed in different countries, consequently reflecting the medical situation in each country, i.e. emergency system, infrastructures, time to transport in hospital, post-surgical care. This is not a limitation of the study, rather it fully falls within the definition of external validation; i.e. the validation cohort must structurally differ from the development cohort, such as in terms of region or county; sometimes termed 'geographic validation' [7]. In our study, <20% of the whole study population originated from Germany. However, the performances of GERAADA score were not penalized from this, resulting slight superior of the Centofanti's score, even though >80% of the study population come from Italian centres.

Although our results decree the superiority of the GERAADA, nonetheless other scores—in particular the IRAD and the Centofanti scores—have shown similar results, although inferior in an absolute sense. However, these scores have advantages compared to GERAADA and should not be excluded a priori: for example they can be calculated faster, as they require fewer variables, and even in the absence of computed tomography scan (this last aspect ideal for use outside the hospital).

To our knowledge, this is the largest external evaluation and comparison of scores specifically designed for patients undergoing surgery for type A acute aortic dissection. Starting from

the first score ever published, the Centofanti's score [4] was externally validated by Yu *et al.* [10] in a little study population of just 79 patients, who found low discrimination (AUC = 0.58). The second score ever published as well as the last one (i.e. the IRAD model 1 and the UK aortic score, respectively) have been never validated at the moment of the writing. On the contrary, the GERAADA score—published in 2020—stimulated several external validations in a relatively short time [11–14], still anyway in single-centre studies with little study populations. Moreover, different definitions of baseline characteristics could have contributed to different observed results [14].

The scope of the present analysis was to compare the performances and reliability of 4 Western mortality scores for patients affected by TAAD requiring emergency surgery. All these scores showed strengths, but also limitations, which should be kept in mind when we use these as prediction models. Generally, we recognize 2 possible uses of such models: supporting the physicians in the decision-making process and the comparison of performances between surgeons, hospitals or regions.

Regarding the decision-making process, surgeons should be awarded of our finding if basing their decision on one of these scores. For example, they should not deny surgery to patients categorized as high-risk based on the GERAADA and UK scores. On the other hand, they should not be too optimistic when operating on patients categorized as at low-risk based on the UK aortic or IRAD model 1 scores. As previously commented, we are convinced that the consequences for daily clinical practice should be at the discretion of the individual surgeon, always in agreement with the patient, or her/his relatives, when possible [15].

On the other side, the scores could be greatly valuable for comparing the performances between institutions or surgeons. Surgical outcomes in TAAD are well-known correlated to factors such as centre- or surgeon volume [16]. However, these differences could be biased by a selection process. Chikwe *et al.* [16] for example found that the mean Parsonnet score of patients undergoing surgery by high-volume surgeons was significant lower compared with those of low-volume surgeons (mean value of the score: 11.7 vs 12.9, $P = 0.003$). The use of a model specifically dedicated for TAAD can improve the outcome analysis for internal (training or quality assessment) or external (resources distribution, health care system interventions) purposes. For instance, we recommend the use of GERAADA score if comparing outcomes between Italian institutions or surgeons.

Strengths and limitations

Between limitations of the present study, there are to account for the retrospective nature of the data collection. Because it is a multicentre study, the time of enrolment of the patients could differ between the participating centres. Most of the enrolled patients, origin from Italy and Germany, and the performances of IRAD and UK aortic scores could be theoretically influenced by differences in the health care systems of these countries. However, the assessment of ones' score transportability outside from its contest, is one of the purposes of external validation.

Among the strengths of the study, there are the very large sample of patients, and the fact of having excluded patients included in the scores considered, in order to avoid overfitting.

CONCLUSIONS

This is the first and wider external validation of 4 scores specifically designed for surgical repair of TAAD. This comparison underlines the strengths but also the limitations of these scores. In our population, GERAADA score shows the better performances in comparison to others and should be recommended. Our results can be useful to translate these scores in the real world, as well as to implement the current prediction models.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *EJCTS* online.

FUNDING

No fundings have been received for this research.

Conflict of interest: none declared.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

Author contributions

Francesco Pollari: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Supervision; Validation; Visualization; Writing—original draft; Writing—review & editing. **Paolo Nardi:** Data curation; Formal analysis; Writing—review & editing. **Elisa Mikus:** Data curation; Formal analysis; Writing—review & editing. **Francesco Ferraro:** Data curation; Formal analysis; Writing—review & editing. **Marco Gemelli:** Data curation; Writing—review & editing. **Ilaria Franzese:** Data curation; Formal analysis; Methodology; Project administration; Writing—review & editing. **Ilaria Chirichilli:** Data curation; Formal analysis; Writing—review & editing. **Claudia Romagnoni:** Data curation; Formal analysis; Writing—review & editing. **Giuseppe Santarpino:** Data curation; Formal analysis; Writing—review & editing. **Salvatore Nicolardi:** Data curation; Formal analysis; Writing—review & editing. **Roberto Scrofani:** Data curation; Formal analysis; Writing—review & editing. **Francesco Musumeci:** Data curation; Formal analysis; Writing—review & editing. **Enzo Mazza:** Data curation; Formal analysis; Writing—review & editing. **Gino Gerosa:** Data curation; Formal analysis; Writing—review & editing. **Massimo Massetti:** Data curation; Formal analysis; Writing—review & editing. **Carlo Savini:** Data curation; Formal analysis; Writing—review & editing. **Giovanni Ruvolo:** Data curation; Formal analysis; Writing—review & editing. **Michele Di Mauro:** Formal analysis; Methodology; Software; Validation; Writing—original draft. **Luca Di Marco:** Conceptualization; Methodology; Supervision; Validation. **Fabio Barili:** Funding acquisition; Project administration; Resources; Supervision; Writing—review & editing. **Alessandro Parolari:** Funding acquisition; Project administration; Supervision; Validation; Writing—review & editing. **Theodor Fischlein:** Funding acquisition; Investigation; Project administration; Supervision; Writing—review & editing.

Reviewer information

European Journal of Cardio-Thoracic Surgery thanks Thomas Schachner, Sven Peters, and the other anonymous reviewers for their contribution to the peer review process of this article.

ACKNOWLEDGEMENTS

We thank all collaborators from each institution for their support in data collection. A complete list of them is provided in supplementary material.

REFERENCES

- [1] Erbel R, Aboyans V, Boileau C, Bossone E, Bartolomeo RD, Eggebrecht H *et al.* ESC Committee for Practice Guidelines. 2014 ESC Guidelines on the diagnosis and treatment of aortic diseases: document covering acute and chronic aortic diseases of the thoracic and abdominal aorta of the adult. The Task Force for the Diagnosis and Treatment of Aortic Diseases of the European Society of Cardiology (ESC). *Eur Heart J* 2014; 35:2873–926.
- [2] Czerny M, Siepe M, Beyersdorf F, Feisst M, Gabel M, Pilz M *et al.* Prediction of mortality rate in acute type A dissection: the German Registry for Acute Type A Aortic Dissection score. *Eur J Cardiothorac Surg* 2020;58:700–6.
- [3] Rampoldi V, Trimarchi S, Eagle KA, Nienaber CA, Oh JK, Bossone E *et al.*; International Registry of Acute Aortic Dissection (IRAD) Investigators. Simple risk models to predict surgical mortality in acute type A aortic dissection: the International Registry of Acute Aortic Dissection score. *Ann Thorac Surg* 2007;83:55–61.
- [4] Centofanti P, Flocco R, Ceresa F, Attisani M, La Torre M, Weltert L *et al.* Is surgery always mandatory for type A aortic dissection? *Ann Thorac Surg* 2006;82:1658–63; discussion 1664. discussion 1664.
- [5] Benedetto U, Dimagli A, Kaura A, Sinha S, Mariscalco G, Krasopoulos G *et al.* Determinants of outcomes following surgery for type A acute aortic dissection: the UK National Adult Cardiac Surgical Audit. *Eur Heart J* 2021;43:44–52.
- [6] https://www.dgthg.de/de/GERAADA_Score (15 September 2023, date last accessed).
- [7] Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2020;14:49–58.
- [8] Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015;68:25–34.
- [9] Pollari F, Di Mauro M. Commentary: if a patient's life is at stake, let's not stop at the surface of the curves. *J Thorac Cardiovasc Surg* 2023; 165:1447–8.
- [10] Yu PJ, Cassiere HA, Kohn N, Dellis SL, Manetta F, Esposito RA *et al.* Utility of established risk models to predict surgical mortality in acute type-A aortic dissection. *J Cardiothorac Vasc Anesth* 2016;30:39–43.
- [11] Luehr M, Merkle-Storms J, Gerfer S, Li Y, Krasivskiy I, Vehrenberg J *et al.* Evaluation of the GERAADA score for prediction of 30-day mortality in patients with acute type A aortic dissection. *Eur J Cardiothorac Surg* 2021;59:1109–14.
- [12] Nežić DG, Živković IS, Milčić MD, Milčić PA, Košević DN, Boričić MI *et al.* On-line risk prediction models for acute type A aortic dissection surgery: validation of the German Registry of Acute Aortic Dissection Type A score and the European System for Cardiac Operative Risk Evaluation II. *Eur J Cardiothorac Surg* 2022;61:1068–75.
- [13] Sugiyama K, Watanuki H, Tochii M, Futamura Y, Kitagawa Y, Makino S *et al.* Impact of GERAADA score in patients with acute type A aortic dissection. *J Cardiothorac Surg* 2022;17:127.
- [14] Gemelli M, Di Tommaso E, Natali R, Dixon LK, Mohamed Ahmed E, Rajakaruna C *et al.* Validation of the German Registry for Acute Aortic Dissection Type A Score in predicting 30-day mortality after type A aortic dissection surgery. *Eur J Cardiothorac Surg* 2023;63:ezad141.
- [15] Pollari F, Sirch J, Fischlein T. About usefulness of GERAADA score. *Eur J Cardiothorac Surg* 2021;60:1005.
- [16] Chikwe J, Cavallaro P, Itagaki S, Seigerman M, Diluozzo G, Adams DH. National outcomes in acute aortic dissection: influence of surgeon and institutional volume on operative mortality. *Ann Thorac Surg* 2013; 95:1563–9.