

# Risk Stratification for Transcatheter Aortic Valve Replacement

Abdul Ahad Khan<sup>a, b</sup>, Ghulam Murtaza<sup>a</sup>, Muhammad F. Khalid<sup>a</sup>, Furqan Khattak<sup>a</sup>

## Abstract

Risk assessment models developed from administrative and clinical databases are used for clinical decision making. Since these models are derived from a database, they have an inherent limitation of being as good as the data they are derived from. Many of these models under or overestimate certain clinical outcomes particularly mortality in certain group of patients. Undeniably, there is significant variability in all these models on account of patient population studied, the statistical analysis used to develop the model and the period during which these models were developed. This review aims to shed light on development and application of risk assessment models for cardiac surgery with special emphasis on risk stratification in severe aortic stenosis to select patients for transcatheter aortic valve replacement.

**Keywords:** Risk stratification; Risk assessment model; Transcatheter aortic valve replacement; Society of Thoracic Surgeons Predicted Risk of Mortality; Surgical aortic valve replacement

## Introduction

Approximately 12% of the US population (2.7 million Americans) beyond 75 years age has aortic stenosis [1]. Over 67,000 surgical aortic valve replacements (SAVRs) procedures are performed in the USA annually [2]. An additional 290,000 elderly patients at high or prohibitive surgical risk could potentially be treated with transcatheter aortic valve replacement (TAVR) in Europe and North America. This number is expected to increase due to the ageing US population and the recent Food and Drug Administration (FDA) approval for TAVR in low- and intermediate-surgical-risk patients. Accurate risk assessment for cardiac procedures forms an essential component of patient centric care. Although risk models have been in use for many years, their importance has increased only as recently as 1986 after the Health Care Financing Administration

(HFCA), now known as the Center for Medicare and Medicaid Services (CMS) published the hospital outcomes data [3]. Such administrative databases were used to compare outcomes amongst care providers and hospitals. However, physicians rightly argued that such data did not take factors such as disease severity or patient heterogeneity into account. This led to the development of a number of high quality clinical databases and risk models related to cardiac procedures. Coronary artery bypass graft (CABG) surgery in particular became a focus of attention reflecting clinician's desire to not only accurately predict peri-operative mortality, perform better patient counseling, and operative decision making but also improve patient outcomes [4-6].

## Types of Databases

It is well recognized that no risk assessment model is better than the data from which it is devised [3]. Missing or incorrect data may underestimate surgical risk leading to devastating consequences in certain instances. Incorrect data may also lead patients to make misinformed decisions regarding providers while reviewing institutional quality of care outcomes. Additionally, a reliable model to predict mortality may be extremely helpful regarding both the informed consent of the patient and to give advice to the heart team [7].

## Administrative Databases

Primarily built for billing purposes, administrative databases contain clinical data, thus becoming one of the commonest sources for observational studies. They have been frequently used to formulate cardiac surgery risk models. The CMS database (MEDPAR) is one such example, which provides readily available and relatively inexpensive information on millions of patients [4, 8]. International Classification of Diseases Codes is used to collect patient-specific data and diagnoses. Based on the combination of codes, patients are further stratified into diagnosis-related groups for reimbursement purposes. Unfortunately, since these administrative databases are primarily built for billing purposes rather than for clinical studies, critical information such as the severity of comorbidities or ejection fractions is often missing. It is also often difficult to discern complication from comorbidity using these databases. They often have insufficient flexibility to properly classify certain comorbidities and sometimes exclude important variables that are not billable diagnoses [4, 9].

Manuscript submitted October 15, 2019, accepted November 5, 2019

<sup>a</sup>Division of Cardiovascular Medicine, East Tennessee State University, Johnson City, TN, USA

<sup>b</sup>Corresponding Author: Abdul Ahad Khan, Division of Cardiovascular Medicine, East Tennessee State University, 329 N State of Franklin Rd, Johnson City, TN 37604, USA. Email: khanaa1@etsu.edu

doi: <https://doi.org/10.14740/cr966>

Two important administrative databases in the USA are the Nationwide Inpatient Sample (NIS) and the University Health System Consortium (UHC) [10-12]. Founded in 1988 as part of the Healthcare Cost and Utilization Project (HCUP), NIS contains data on roughly 8 million hospital stays per year [11]. Similarly, UHC is an alliance of 302 affiliated hospitals representative of academic medical centers.

Despite all their flaws, administrative databases have shown similar discrimination capabilities as traditional clinical databases [13]. In a study of 152,523 isolated CABG surgeries in the UK between 1996 and 2004, risk prediction models derived from administrative databases were able to forecast operative mortality with similar discrimination compared to that of a risk model built on clinical datasets (European system for cardiac operative risk evaluation (EuroSCORE)); receiver operative characteristic curve score 0.77 versus 0.78) [13].

## Clinical Databases

Clinical databases provide detailed accounts of hospital admissions and operative procedures, and are frequently used as a source of data for risk modeling in cardiology and cardiac surgery [14]. Although various cardiac surgery databases such as the Northern New England Cardiovascular Disease Study Group [15] and the Veterans Affairs Administration previously existed, the release of HFCA outcome data triggered the Society of Thoracic Surgeons (STS) to formulate a newer, more inclusive database known as the STS National Cardiac Database. The database was formally founded in 1989. During the subsequent three decades, it has evolved to become one of the biggest and most robust single specialty databases. It contains data covering over 3 million procedures performed at over 1,000 cardiac surgery centers throughout the USA in the last decade alone [16]. Today, it is considered the gold standard for specialty outcome databases [14, 17, 18]. The number of variables collected by STS database has steadily increased to over 600 individual data points in its latest edition (version 2.81) [19]. Although, an attractive proposition, the increased number of data points is not only time consuming, but may also lead to increased chances of coding errors, affecting the quality and accuracy of the data entered [20]. Indeed, such databases require substantial financial investment on part of the participating institutions. Ideally, such databases entail employing professional data abstractors, who are adequately trained and intermittently recertified to remain well versed with updated data definitions [21].

Documentation of variables such as preoperative risk factors and patient demographics is fairly accurate in clinical databases [14, 22]. On the other hand, data abstraction of operative details and disease etiology becomes increasingly difficult especially in the absence of uniform definitions. Even a simple endpoint like mortality may be open to different interpretations such as: 1) Thirty days all-cause mortality; 2) In-hospital mortality, irrespective of when it occurs; 3) Operative mortality, defined as a combination of 1) or 2). The importance of clear and unambiguous definitions is re-emphasized by a recent study which showed only 89% median agreement be-

tween four abstractors gathering data on etiology of valve disease, procedural details and surgical acuity [23]. Interestingly, all the data abstractors were cardiothoracic surgery residents and fellows. The STS regularly updates variable definitions between successive database versions with the aim to improve diagnostic accuracy and clarity. However, frequent changes in definitions may often become problematic when it comes to data abstraction [20].

## Risk Assessment Model Development

Risk assessment model development entails the use of statisticians with considerable experience in this area, since different multivariable equations could be used by various statisticians for the same data [24]. Similarly, the type of modeling strategy and validation techniques may differ depending on whether the model is aimed at probabilistic prediction to aid in patient selection and counseling or merely meant to provide a comparison between providers and treatments [25, 26].

Three principal techniques have been used for developing cardiac risk models. Bayesian models were initially used for the STS database due to significant amount of missing data. By 1995, as data omissions resolved, logistic regression models became more favorable and are currently the most widely used statistical technique for cardiac risk modeling [27, 28]. Numerous risk models including the STS risk score, Veterans Affairs risk score [29] and the New York State risk score [30] currently utilize logistic regression. Although much more robust, the application of Bayesian and logistic regression models is limited at bedside since these models require complex calculations. This served as an impetus to form simpler models such as the EuroSCORE, which utilizes additive scores with weights derived from the logistic regression model. Some comparative studies have shown better overall performance of models based on logistic regression [31].

During risk model development, the study population is often randomly divided into a development or training sample and a validation or test sample. For instance, the STS isolated valve risk model was divided into 60% derivation/developmental sample and 40% validation sample. Data from the development sample are used to identify predictive variables and risk model coefficients. The validation sample is used to assess overall clinical performance and model fit, as defined by calibration (reliability) and discrimination (resolution). Calibration describes how closely the model estimated death rates align with observed mortality, quantified through the Hosmer-Lemeshow statistical test [32]. Discrimination is the model's ability to predict whether a specific patient will survive or not, measured using the area under the receiver operator curve (AUROC), also known as the c-index [3]. The c-index ranges between 0.5 and 1.0. Higher values of c-index imply better discrimination. On the other hand, values approaching 0.5 indicate the model's ability to discriminate is not better than random chance or "flip of a coin" [33]. For instance, risk models with low discrimination may accurately predict five out of 100 post-operative deaths for a particular set of risk factors, but fail to identify those five patients. A summary of usefulness of

**Table 1.** Advantages of Risk Prediction Models

Decision-making tool in valvular intervention to allow optimal operative decision making.
Predict outcomes including operative mortality, stroke, and renal failure in patients undergoing valve replacement.
Allow fair comparison of operative outcomes as they control for heterogeneous patient population, disease severity and clinical comorbidities.
Predict high-risk patients and allow selection of patients for transcatheter vs. surgical aortic valve replacement.
Derived from large-scale clinical databases to predict outcomes.
Serve as a valuable resource for quality improvement, clinical decision making, and patient counseling.
Addition of new variable such as prior cardiac intervention, in-hospital stroke and frailty improves prediction in high-risk patients to optimize selection of patients for TAVR.

TAVR: transcatheter aortic valve replacement.

risk prediction models particularly in TAVR patients is shown in Table 1.

### Limitations of Risk Adjustment Models

Important limitations (Table 2) need to be taken into account when employing risk adjustment models for assessing operative outcomes. Firstly, risk algorithms are most accurate in the population and time frame from which they were derived. For instance, the EuroSCORE that is widely popular in Europe does not predict outcomes accurately when applied to surgical practice in North America [34-36]. Similarly, model calibration can be lost over time leading to mis-estimation of operative outcomes, necessitating the need for periodic recalibrations. Loss of model calibration was one of the reasons why EuroSCORE had to be updated with the EuroSCORE II [37, 38]. The STS risk score regularly gets recalibrated as well, with the last update taking place as recently as 2008 [39].

Risk algorithms require homogeneity of treatment between the study groups and the population from which the algorithm is modeled. This has far reaching implications for patients with aortic stenosis. Although both SAVR and TAVR serve as viable treatment options for patients with aortic stenosis, risk algorithms based on SAVR outcomes may not be as accurate when applied to TAVR patients.

It is also important to underscore that risk adjustment loses accuracy at extremes of populations. The tail of the bell shaped curve often has too few patients on which to build a statistically valid model. This factor gains importance in the setting of TAVR patients where severe aortic stenosis patients lie at the end of the spectrum, often resulting in overestimation of mortality witnessed with many models [40-42].

Lastly, risk assessment models depend on the accurate data entry of a core set of variables. They often cannot account

for certain variables that have not been collected. Such omission of data primarily has two causes: 1) The occurrence of the condition is so infrequent that its impact is not measurable; 2) The causal factor cannot be measured accurately. For instance, frailty parameters have often been excluded from risk scores due to lack of a gold standard [43].

### Traditional Risk Assessment Models for Cardiac Surgery

The most frequently modeled outcome in cardiac procedures is mortality. Two most commonly used risk assessment models are the logistic EuroSCORE in Europe and the STS risk score in North America.

#### EuroSCORE

The EuroSCORE was developed in 1995 as a user-friendly system, in the hopes of encouraging as many units as possible to embark on programs of risk-adjusted quality monitoring. It was derived from data involving over 19,000 patients from 128 centers in eight European countries [44]. Originally, a simple additive score was published, enabling clinicians to calculate peri-operative mortality at bedside. Unfortunately, the model generally overestimated mortality at lower scores (EuroSCORE  $\leq 6$ ) and underestimated mortality at higher scores (EuroSCORE  $> 13$ ) [45].

As access to information technology increased and interest regarding quality control grew throughout Europe, the score was updated to a logistic regression model in 1999. In order to further improve the accuracy of the original EuroSCORE and counter calibration drift, the model was updated to EuroSCORE II, which was derived from more than 22,000

**Table 2.** Limitations of Risk Prediction Models

Most accurate in the population and time frame from which they are derived.
Model calibration can be lost over time leading to mis-estimation of operative outcomes, necessitating the need for periodic recalibrations.
Model derived from patients with a particular disease or procedure may not apply to another disease or procedure.
Accuracy is lost at extremes of population, i.e., very low-risk and very high-risk patients.
May not have all relevant parameters of "risk" or "severity" due to missing data in derivative population.

patients operated upon in 2010 in over 43 countries [46]. This model uses 18 covariates for predicting mortality in surgical aortic valve replacement. Whether the updated version fares better than the original version remains debatable. A recent multicenter validation study found that the EuroSCORE II has optimal calibration until 30% predicted mortality and consistently over predicts at higher risk [47].

### STS score

The STS risk score that is primarily used in the USA was developed in a later era (2002 - 2006). It provides highly predictive estimation of mortality in both the short and the long term [39]. Not only does it predict mortality in coronary artery bypass grafting, valve replacement procedures and combined cases, but also offers risk prediction for numerous other outcomes including stroke, deep sternal wound infection, renal dysfunction, reoperation, prolonged ventilation, and the length of hospital stay. The STS risk score is calculated through the STS website and continually adjusted to improve accuracy. The latest version 2.81 contains various risk factors not previously collected such as frailty, previous radiation exposure and liver disease.

The STS provides a quarterly report to participating institutions, whereby predicted risk is recalibrated to account for calibration drift due to improvements in surgical outcomes over time. Unlike the quarterly STS institutional reports, the STS online risk calculator which is used by most clinicians at bedside is still based on the 2008 STS isolated valve risk model and lacks the above mentioned recalibration factor [49]. Health care providers making clinical decisions using the online Calculator should bear these limitations in mind.

### Use of Traditional and Novel Risk Assessment Models in TAVR

The advent of TAVR, as a viable alternative to SAVR in patients with severe aortic stenosis who are deemed suboptimal for surgery has revolutionized valvular repairs [50, 51]. Naturally, it has sparked interest in developing risk models for management of patients with aortic stenosis. Below we will discuss the use of traditional and novel risk assessment models in patients undergoing TAVR.

### STS

The most commonly used model for this population is the Society of Thoracic Surgeons Predicted Risk of Mortality (STS-PROM), which was neither developed nor intended for use in this population. Supporters of TAVR cite the lower observed mortality with TAVR compared to predict mortality based on STS score, to promote TAVR. However, there are a number of reasons why these risk models do not fit well in TAVR patients. Firstly, these scores were primarily derived and validated in a totally different treatment group i.e. SAVR and are therefore highly likely to be invalid. Secondly, TAVR patients are mostly

at higher risk, lying on the extreme right of the bell curve, and the current models have very few patients at “high risk” for the model to offer robust discrimination [40-42]. Thirdly, some of the variables, such as sternal wound infection do not pertain to TAVR patients.

### EuroSCORE

It is generally agreed that the widely utilized EuroSCORE overestimates mortality in most patients [49-51]. Additionally, EuroSCORE II, although more accurate than the original logistic score is largely based on a dataset consisting of coronary procedures. It may therefore be less well adapted to aortic procedures [52-52].

### Edwards et al model based on the TVT Registry

Realizing the dearth of data related to TAVR procedures, the American College of Cardiology (ACC) in close partnership with STS and a number of other stakeholders including the FDA and CMS developed the Transcatheter Valve Therapy (TVT) Registry in 2011. The registry provides a rich source of information regarding what impact transcatheter heart valve procedures have on patient outcomes and how they are being deployed throughout the USA. The TVT Registry currently has more than 20,000 patient records with detailed demographic and post TAVR outcomes including procedural mortality. The national coverage determination from the CMS [56] specifies that reimbursement is contingent on participation in a national registry with a set number of specifications all of which are met by the TVT Registry. This increases the external validity of the TVT Registry sample to nearly a real-world TAVR experience [57]. As the number of patients enrolled in the TVT Registry increase, valuable information regarding the evolution of outcomes and clinical profiles of patients undergoing TAVR will emerge. It is expected the predicted risk will change over time and provide future analysis using the registry may provide useful insights regarding the clinical subsets of patients who might benefit from TAVR most. Conversely, it will also reveal subsets not suitable for TAVR. TVT Registry participant sites will be able to compare their individual risk adjusted outcomes data to national benchmarks as this information gets disseminated to them several times each year.

Using the TVT Registry data Edwards et al [57] developed a statistical model predicting in-hospital mortality after TAVR in 2016. The model is based on 13,718 consecutive US patients who underwent TAVR in 265 participating centers throughout the USA from 2011 to 2014. The model is unique in that it was developed to assess patient-level outcome predictions rather than center-level risk adjusted outcomes. The final model covariates used to predict to mortality included age, glomerular filtration rate, dialysis dependence, New York Heart Association (NYHA) class IV, severe chronic lung disease, non-femoral access site and patient acuity category. The model displayed a high degree of discrimination as measured by a C-statistic of 0.66. In comparison, The C-statistic was

**Table 3.** Summary of Commonly Used Risk Prediction Models for Aortic Valve Intervention (Advantages and Disadvantages of Risk Assessment Models)

Risk model	Year score created	Number of patients	Country	Advantage	Disadvantage
EuroSCORE	1995	19,030	128 centers from 8 European countries	Simple bedside tool to calculate mortality	Overestimates mortality at low score and underestimates at higher score
EuroSCORE II	2010	22,000	154 hospitals in 43 countries	Comprehensive model using multiple covariates for assessing mortality	Overestimates mortality at higher risk. Predicts only short-term mortality.
STS-PROM	2002 - 2006	24,222	USA	Highly predictive of short-term and long-term mortality	Developed for conventional surgical aortic valve replacement.
Edwards et al (STS/ ACC TAVR score)	2016	13,718	USA	Predicts only in-hospital mortality for patients undergoing TAVR	Lacks long-term outcomes.
German aortic valve score	2008	11,974	Germany	Comparison of predicted and observed mortality for SAVR vs. TAVR in all categories of risk groups	TAVR or SAVR were not tested as an independent risk factor. Low number of TAVR procedures included.

EuroSCORE: European system for cardiac operative risk evaluation; STS-PROM: Society of Thoracic Surgeons Predicted Risk of Mortality STS: Society of Thoracic Surgeons; ACC: American College of Cardiology; TAVR: transcatheter aortic valve replacement; SAVR: surgical aortic valve replacement.

recently reported to be 0.60 for the standard STS model and 0.53 for logistic EuroSCORE when applied to a population of TAVR patients [58].

The Edwards et al model is a small step in the long journey of risk prediction for TAVR patients. Based on initial data, the model appears to be much more accurate than the previous surgical models. The model does however have a number of limitations, including missing some key risk factors such as frailty and quality of life measures. This model only accounts for in-hospital mortality in contrast to a 30-day mortality model. This is important because even though a majority of TAVR patients survive their hospitalization, a large number eventually succumb within 30 days. Additionally, although the TVT Registry undergoes rigorous internal quality checks, the external audit process is still in its early stages, compromising the quality of the data forming the model [56]. Table 3 summarizes some of the commonly used risk prediction models for aortic valve intervention.

The linkage of clinical databases that capture short-term outcomes with administrative databases that capture long-term outcomes will allow risk models to be developed that capture long-term data. Efforts are underway to include data in the TVT Registry for both 30-day follow-up as well as 1-year survival information after TAVR. Currently this information is sparse in the database, although future models may have adequate information regarding these two important variables. In addition models to predict the probability of neurological deficit are also currently under development. Overall, approximately one-third of individuals aged > 80 years are considered to be frail indicating the high prevalence of this risk factor [59]. Variables such as patient frailty measured using standardized indices such as the preprocedural 5-m walk test, Katz index or Instrumental Activities of Daily Living Score are

also likely to be included as a covariate in future TVT Registry risk model [60-62]. The TVT Registry has the advantage of including quality of life metrics measured using the Kansas City Cardiomyopathy Questionnaire. Linking the information gained from the questionnaire with the clinical data in the TVT Registry will enable models to be developed that accurately estimate procedural benefit. Recently Thourani et al developed a risk model for in-hospital stroke after TAVR using STS TVT Registry using a cohort of 97,600 patients from 521 sites of TVT Registry. Their model developed prediction model to allow assessment of risk-relevant patient risk factors for in-hospital stroke risk. This data will be updated and transmitted to the TVT Registry [63].

Indeed, a risk model that can predict not only mortality but also stroke risk, and post-operative quality of life is on the horizon and will serve as a powerful adjunct in the process of appropriate patient selection.

#### **German aortic valve score (GAVS) based on German aortic valve registry (GARY)**

Since existing risk stratification models consistently overestimate the mortality observed in the German population, a new score was needed to estimate preoperative risk in the context of inter-hospital benchmarking and quality control [64]. First published in 2012, the GAVS I was based on a comprehensive national data pool from 2008 known as the GARY Registry [65, 66]. GARY is unique in that it includes all interventional and surgical treatment options for aortic valve disease that are currently available in Germany. Since this score was developed based on data from the early period of transcatheter techniques, it was limited in its applicability due to the small pro-

portion of TAVR procedures performed at the time (634 cases, 5.2% of the total patients enrolled in the database). Additionally, the score failed to account for hemodynamic results, long-term survival and quality of daily life. In order to avoid diminishing the number of transcatheter aortic valve implantation (TAVI) cases for calculation further, data were not divided into a learning and validation part, limiting the external validity of the data set. By 2012, the proportion of patients with aortic stenosis treated percutaneously with TAVR had grown almost 10-fold to 44.9%. Wolfgang Schiller et al re-calibrated GAVS I to account for the higher number of patients having TAVR forming the new GAVS II score [67]. The GAVS fits best to the German public health service, where the prevalence of TAVR is high compared to other countries. Unfortunately, the score can only be used to predict the short-term risk mortality and fails to benchmark SAVR and TAVR patients separately.

A rapidly increasing number of TAVR procedures have been observed over the last 5 years, particularly in countries with no restrictions on reimbursement, which requires further assessment and validation of these risk scores. While there is no specific scoring system tested in the Asian population, STS scoring is still widely used to select patients for TAVR in Asian countries. Several studies have shown good predictability of STS score in Asian population. Moreover, a recent objective clinical frailty score used in a Japanese cohort showed good discriminability to exclude frail patients who would not benefit from invasive therapy [68].

## Conclusions

Multiple risk assessment models are available to identify patients with severe aortic stenosis who are candidates for TAVR due to high predicted mortality with SAVR. Each model suffers from limitations e.g., EuroSCORE can overestimate surgical risk and misclassify patients, and focuses on in-hospital mortality and short-term outcomes. The STS score may offer better risk stratification due to addition of more relevant risk factors and ability to predict both short and long-term outcomes; however it is still not perfect due to being developed in the SAVR population. Novel risk assessment models such as Edwards et al and GAVS may improve selection of TAVR patients and predict procedural risk, but need long-term data and refinement.

## Acknowledgments

None to declare.

## Financial Disclosure

This research did not receive any financial or grant support.

## Conflict of Interest

The authors report no financial relationship or conflict of inter-

est regarding the content herein.

## Author Contributions

All authors are involved in manuscript writing, preparation, review and final approval.

## References

- Osnabrugge RL, Mylotte D, Head SJ, Van Mieghem NM, Nkomo VT, LeReun CM, Bogers AJ, et al. Aortic stenosis in the elderly: disease prevalence and number of candidates for transcatheter aortic valve replacement: a meta-analysis and modeling study. *J Am Coll Cardiol.* 2013;62(11):1002-1012.
- Clark MA, Duhay FG, Thompson AK, Keyes MJ, Svensson LG, Bonow RO, Stockwell BT, et al. Clinical and economic outcomes after surgical aortic valve replacement in Medicare patients. *Risk Management and Healthcare Policy.* 2012;5:117-126.
- Shahian DM, Blackstone EH, Edwards FH, Grover FL, Grunkemeier GL, Naftel DC, Nashef SA, et al. Cardiac surgery risk models: a position article. *Ann Thorac Surg.* 2004;78(5):1868-1877.
- Hannan EL, Racz MJ, Jollis JG, Peterson ED. Using Medicare claims data to assess provider quality for CABG surgery: Does it work well enough? *Health Serv Res.* 1997;31:659-678.
- Dubois RW, Rogers WH, Moxley JH, Draper D, Brook RH. Hospital inpatient mortality. Is it a predictor of quality? *N Engl J Med.* 1987;317(26):1674-1680.
- Griffith B, Hattler B, Hardesty R, Kormos R, Pham S, Bahnson H. The need for accurate risk-adjusted measures of outcome in surgery. Lessons learned through coronary artery bypass. *Ann Surg.* 1995;222(4):593-598; discussion 598-599.
- Van Mieghem NM, Serruys PW. The art of risk stratification in TAVI. *European Heart Journal.* 2013;34(25):1859-1861.
- Iezzoni LI. The risks of risk adjustment. *Jama.* 1997;278(19):1600-1607.
- Finlayson EV, Birkmeyer JD, Stukel TA, Siewers AE, Lucas FL, Wennberg DE. Adjusting surgical mortality rates for patient comorbidities: more harm than good? *Surgery.* 2002;132(5):787-794.
- Thalji NM, Suri RM, Greason KL, Schaff HV. Risk assessment methods for cardiac surgery and intervention. *Nat Rev Cardiol.* 2014;11(12):704-714.
- Healthcare and Utilization Project. Overview of the National (Nationwide) Inpatient Sample (NIS). n.d. [cited 2017 February 18]; Available from: Overview of the National (Nationwide) Inpatient Sample (NIS).
- Vizient. [cited 2019 20th November]; Available from: <https://www.vizientinc.com/about-us>.
- Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ.* 2007;334(7602):1044.

14. Shahian DM, Silverstein T, Lovett AF, Wolf RE, Normand SL. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation*. 2007;115(12):1518-1527.
15. Malenka DJ, O'Connor GT. The Northern New England Cardiovascular Disease Study Group: a regional collaborative effort for continuous quality improvement in cardiovascular disease. *Jt Comm J Qual Improv*. 1998;24(10):594-600.
16. The Society of Thoracic Surgeons. Adult Cardiac Surgery Database Executive Summary. [Online] 2015. [cited 2017 February 18]; Available from: [http://www.sts.org/sites/default/files/documents/2016Harvest1\\_ExecutiveSummary.pdf](http://www.sts.org/sites/default/files/documents/2016Harvest1_ExecutiveSummary.pdf).
17. Shahian DM, Edwards FH, Jacobs JP, Prager RL, Normand SL, Shewan CM, O'Brien SM, et al. Public reporting of cardiac surgery performance: Part 2—implementation. *Ann Thorac Surg*. 2011;92(3 Suppl):S12-23.
18. Bridgewater B. Society of Cardiothoracic Surgery in GB and Ireland. Cardiac registers: the adult cardiac surg.
19. The Society of Thoracic Surgeons. STS National Database. n.d. [cited 2017 February 18]; Available from: <http://www.sts.org/national-database>.
20. Schaff HV, Brown ML, Lenocho JR. Data entry and data accuracy. *J Thorac Cardiovasc Surg*. 2010;140(5):960-961.
21. Shahian DM, Edwards F, Grover FL, Jacobs JP, Wright CD, Prager RL, Rich JB, et al. The Society of Thoracic Surgeons National Adult Cardiac Database: a continuing commitment to excellence. *J Thorac Cardiovasc Surg*. 2010;140(5):955-959.
22. Mack MJ, Herbert M, Prince S, Dewey TM, Magee MJ, Edgerton JR. Does reporting of coronary artery bypass grafting from administrative databases accurately reflect actual clinical outcomes? *J Thorac Cardiovasc Surg*. 2005;129(6):1309-1317.
23. Brown ML, Lenocho JR, Schaff HV. Variability in data: the Society of Thoracic Surgeons National Adult Cardiac Surgery Database. *J Thorac Cardiovasc Surg*. 2010;140(2):267-273.
24. Naftel DC. Do different investigators sometimes produce different multivariable equations from the same data? *J Thorac Cardiovasc Surg*. 1994;107(6):1528-1529.
25. Krumholz HM. Mathematical models and the assessment of performance in cardiology. *Circulation*. 1999;99(16):2067-2069.
26. Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. *Stat Med*. 1986;5(5):421-433.
27. Edwards FH, Grover FL, Shroyer AL, Schwartz M, Bero J. The Society of Thoracic Surgeons National Cardiac Surgery Database: current risk assessment. *Ann Thorac Surg*. 1997;63(3):903-908.
28. Shroyer AL, Plomondon ME, Grover FL, Edwards FH. The 1996 coronary artery bypass risk model: the Society of Thoracic Surgeons Adult Cardiac National Database. *Ann Thorac Surg*. 1999;67(4):1205-1208.
29. Grover FL, Shroyer AL, Hammermeister KE. Calculating risk and outcome: the Veterans Affairs database. *Ann Thorac Surg*. 1996;62(5 Suppl):S6-11; discussion S31-12.
30. Hannan EL, Kilburn H, Jr., O'Donnell JF, Lukacik G, Shields EP. Adult open heart surgery in New York State. An analysis of risk factors and hospital mortality rates. *JAMA*. 1990;264(21):2768-2774.
31. Marshall G, Grover FL, Henderson WG, Hammermeister KE. Assessment of predictive models for binary outcomes: an empirical approach using operative death from cardiac surgery. *Stat Med*. 1994;13(15):1501-1511.
32. Dupuis JY. Predicting outcomes in cardiac surgery: risk stratification matters? *Curr Opin Cardiol*. 2008;23(6):560-567.
33. Ranucci M, Castelvechio S, Conte M, Megliola G, Speciale G, Fiore F, Guarracino F, et al. The easier, the better: age, creatinine, ejection fraction score for operative mortality risk stratification in a series of 29,659 patients undergoing elective cardiac surgery. *J Thorac Cardiovasc Surg*. 2011;142(3):581-586.
34. Brown ML, Schaff HV, Sarano ME, Li Z, Sundt TM, Dearani JA, Mullany CJ, et al. Is the European System for Cardiac Operative Risk Evaluation model valid for estimating the operative risk of patients considered for percutaneous aortic valve replacement? *J Thorac Cardiovasc Surg*. 2008;136(3):566-571.
35. Osnabrugge RL, Speir AM, Head SJ, Fonner CE, Fonner E, Kappetein AP, Rich JB. Performance of EuroSCORE II in a large US database: implications for transcatheter aortic valve implantation. *Eur J Cardiothorac Surg*. 2014;46(3):400-408; discussion 408.
36. Dewey TM, Brown D, Ryan WH, Herbert MA, Prince SL, Mack MJ. Reliability of risk algorithms in predicting early and late operative outcomes in high-risk patients undergoing aortic valve replacement. *J Thorac Cardiovasc Surg*. 2008;135(1):180-187.
37. Siregar S, Groenwold RH, de Heer F, Bots ML, van der Graaf Y, van Herwerden LA. Performance of the original EuroSCORE. *Eur J Cardiothorac Surg*. 2012;41(4):746-754.
38. Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, Lockowandt U. EuroSCORE II. *Eur J Cardiothorac Surg*. 2012;41(4):734-744; discussion 744-735.
39. Puskas JD, Kilgo PD, Thourani VH, Lattouf OM, Chen E, Vega JD, Cooper W, et al. The society of thoracic surgeons 30-day predicted risk of mortality score also predicts long-term survival. *Ann Thorac Surg*. 2012;93(1):26-33; discussion 33-25.
40. Iung B, Vahanian A. Towards improved risk scores: the quest for the grail continues. *Eur Heart J*. 2013;34(1):10-12.
41. Osswald BR, Gegouskov V, Badowski-Zyla D, Tochtermann U, Thomas G, Hagl S, Blackstone EH. Overestimation of aortic valve replacement risk by EuroSCORE: implications for percutaneous valve replacement. *Eur Heart J*. 2009;30(1):74-80.
42. Leontyev S, Walther T, Borger MA, Lehmann S, Funkat AK, Rastan A, Kempfert J, et al. Aortic valve replacement in octogenarians: utility of risk stratification with EuroSCORE. *Ann Thorac Surg*. 2009;87(5):1440-1445.

43. de Vries NM, Staal JB, van Ravensberg CD, Hobbelen JS, Olde Rikkert MG, Nijhuis-van der Sanden MW. Outcome instruments to measure frailty: a systematic review. *Ageing Res Rev.* 2011;10(1):104-114.
44. Kacila M, Granov N, Omerbasic E, Memisevic N. [Additive and logistic EuroSCORE in assessment of operative mortality in cardiac surgery]. *Med Arh.* 2008;62(5-6):298-299.
45. Gogbashian A, Sedrakyan A, Treasure T. EuroSCORE: a systematic review of international performance. *Eur J Cardiothorac Surg.* 2004;25(5):695-700.
46. Noyez L, Kievit PC, van Swieten HA, de Boer MJ. Cardiac operative risk evaluation: The EuroSCORE II, does it make a real difference? *Neth Heart J.* 2012;20(12):494-498.
47. Barili F, Pacini D, Capo A, Rasovic O, Grossi C, Alamanni F, Di Bartolomeo R, et al. Does EuroSCORE II perform better than its original versions? A multicentre validation study. *Eur Heart J.* 2013;34(1):22-29.
48. Vassileva CM, Aranki S, Brennan JM, Kaneko T, He M, Gammie JS, Suri RM, et al. Evaluation of the society of thoracic surgeons online risk calculator for assessment of risk in patients presenting for aortic valve replacement after prior coronary artery bypass graft: an analysis using the STS adult cardiac surgery database. *Ann Thorac Surg.* 2015;100(6):2109-2115; discussion 2115-2106.
49. Bhatti F, Grayson AD, Grotte G, Fabri BM, Au J, Jones M, Bridgewater B, et al. The logistic EuroSCORE in cardiac surgery: how well does it predict operative risk? *Heart.* 2006;92(12):1817-1820.
50. Bode C, Kelm M. EUROSCORE: still gold standard or less? *Clin Res Cardiol.* 2009;98(6):353-354.
51. Grossi EA, Schwartz CF, Yu PJ, Jorde UP, Crooke GA, Grau JB, Ribakove GH, et al. High-risk aortic valve replacement: are the outcomes as bad as predicted? *Ann Thorac Surg.* 2008;85(1):102-106; discussion 107.
52. Florath I, Rosendahl UP, Mortasawi A, Bauer SF, Dal-ladaku F, Ennker IC, Ennker JC. Current determinants of operative mortality in 1400 patients requiring aortic valve replacement. *Ann Thorac Surg.* 2003;76(1):75-83.
53. Ambler G, Omar RZ, Royston P, Kinsman R, Keogh BE, Taylor KM. Generic, Simple Risk Stratification Model for Heart Valve Surgery. *Circulation.* 2005;112(2):224-231.
54. Jin R, Grunkemeier GL, Starr A, Providence Health System Cardiovascular Study G. Validation and refinement of mortality risk models for heart valve surgery. *Ann Thorac Surg.* 2005;80(2):471-479.
55. Nowicki ER, Birkmeyer NJ, Weintraub RW, Leavitt BJ, Sanders JH, Dacey LJ, Clough RA, et al. Multivariable prediction of in-hospital mortality associated with aortic and mitral valve surgery in Northern New England. *Ann Thorac Surg.* 2004;77(6):1966-1977.
56. Carroll JD, Shuren J, Jensen TS, Hernandez J, Holmes D, Marinac-Dabic D, Edwards FH, et al. Transcatheter valve therapy registry is a model for medical device innovation and surveillance. *Health Aff (Millwood).* 2015;34(2):328-334.
57. Edwards FH, Cohen DJ, O'Brien SM, Peterson ED, Mack MJ, Shahian DM, Grover FL, et al. Development and Validation of a Risk Prediction Model for In-Hospital Mortality After Transcatheter Aortic Valve Replacement. *JAMA Cardiol.* 2016;1(1):46-52.
58. Beohar N, Whisenant B, Kirtane AJ, Leon MB, Tuzcu EM, Makkar R, Svensson LG, et al. The relative performance characteristics of the logistic European System for Cardiac Operative Risk Evaluation score and the Society of Thoracic Surgeons score in the Placement of Aortic Transcatheter Valves trial. *J Thorac Cardiovasc Surg.* 2014;148(6):2830-2837 e2831.
59. Ahmed N, Mandel R, Fain MJ. Frailty: an emerging geriatric syndrome. *Am J Med.* 2007;120(9):748-753.
60. Afilalo J, Eisenberg MJ, Morin JF, Bergman H, Monette J, Noiseux N, Perrault LP, et al. Gait speed as an incremental predictor of mortality and major morbidity in elderly patients undergoing cardiac surgery. *J Am Coll Cardiol.* 2010;56(20):1668-1676.
61. Cleveland JC, Jr. Frailty, aging, and cardiac surgery outcomes: the stopwatch tells the story. *J Am Coll Cardiol.* 2010;56(20):1677-1678.
62. Afilalo J, Alexander KP, Mack MJ, Maurer MS, Green P, Allen LA, Popma JJ, et al. Frailty assessment in the cardiovascular care of older adults. *J Am Coll Cardiol.* 2014;63(8):747-762.
63. Thourani VH, O'Brien SM, Kelly JJ, Cohen DJ, Peterson ED, Mack MJ, Shahian DM, et al. Development and Application of a Risk Prediction Model for In-Hospital Stroke After Transcatheter Aortic Valve Replacement: A Report From The Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry. *Ann Thorac Surg.* 2019;107(4):1097-1103.
64. Gummert JF, Funkat A, Osswald B, Beckmann A, Schiller W, Krian A, Beyersdorf F, et al. EuroSCORE overestimates the risk of cardiac surgery: results from the national registry of the German Society of Thoracic and Cardiovascular Surgery. *Clin Res Cardiol.* 2009;98(6):363-369.
65. Hamm CW, Mollmann H, Holzhey D, Beckmann A, Veit C, Figulla HR, Cremer J, et al. The German Aortic Valve Registry (GARY): in-hospital outcome. *Eur Heart J.* 2014;35(24):1588-1598.
66. Kottling J, Schiller W, Beckmann A, Schafer E, Dobler K, Hamm C, Veit C, et al. German Aortic Valve Score: a new scoring system for prediction of mortality related to aortic valve procedures in adults. *Eur J Cardiothorac Surg.* 2013;43(5):971-977.
67. Schiller W, Barnewold L, Kazmaier T, Beckmann A, Masseli F, Welz A, Szecsenyi J, et al. The German aortic valve score II. *Eur J Cardiothorac Surg.* 2017;52(5):881-887.
68. Tay ELW, Ngiam JN, Kong WK, Poh KK. Management of severe aortic stenosis: the Singapore and Asian perspective. *Singapore Med J.* 2018;59(9):452-454.