JScholar
Fostering Scholarly Communication

Editorial

Beware of Medical Decisions That are Automatic

Robert Poston[*]

MD, Division of Cardiothoracic Surgery, SUNY Downstate Health Sciences University, Brooklyn, New York, United States

*****Corresponding author:** Robert Poston, MD, Division of Cardiothoracic Surgery, SUNY Downstate Health Sciences University, Brooklyn, New York, United States, Tel: 6176100126, E-mail: postonlab@gmail.com

Citation: Robert Poston (2021) Beware of Medical Decisions That are Automatic. J Anesth Surg Care 1: 1-4.

Abstract

In 1990, the Society of Thoracic Surgery developed a national cardiac surgical database as an effort to improve surgical quality [1]. For the past 2 decades, extensive data has entered this database from >95% of cardiac surgical procedures all across the US. As a result of its size, the STS database provides a powerful way to measure the quality of a surgeon or surgical program by comparing their average patient outcomes against national averages. More recently, this database has also been used to predict the risk that a given patient will experience a bad outcome [2]. This prediction uses statistical models to add up the impact of a variety of different risk factors on the risk of death. Death after heart surgery is mostly caused by the trauma of the procedure leading to the failure of important organs like the lungs, liver or kidneys. That means that the strongest predictors of operative death are those that signal that these organ systems are vulnerable [3]. Any surgeon asked to operate on a patient whose only problem is severe lung dysfunction immediately recognizes the risk. However, it is a more challenging task to recognize the risk of death caused by a variety of modest risk factors, such as mild dysfunction in 3 or 4 organ systems in an elderly diabetic. Humans don't have enough cognitive bandwidth in our working memory to consider the impact of multiple variables at once [4]. A computer armed with the right statistical models is far more capable than the human mind of considering how these multiple variables influences surgical mortality. The risk score it provides augments the surgical team's ability to select appropriate cases that are not too high risk for a successful outcome.

Keywords: STS Database; Cardiac Surgery; Risk Predictors

Introduction

Despite its strengths, there are also important weaknesses of the STS risk calculator. [5] First, it is highly accurate at predicting that a surgeon operating on 100 patients with similar risk is likely to have 5 patients die, but much less precise in discriminating exactly who are those 5 patients. Second, it is not good at the extremes of a population, i.e., very high-risk patients. The tail of the bell-shaped curve often has too few patients on which to build a statistically valid model with a high level of discrimination. Third, several important risk factors are not included in the STS risk calculation, such as severe calcification of the aorta, a history of chest radiation, liver dysfunction, cognitive impairment, nutrition level, frailty, pulmonary hypertension and severe CHF as illustrated by B-type natriuretic peptide. Because these factors independently increase surgical risk, the models often assume these "unmeasured confounders" are present in a high risk group even when they are not. Finally, the risk of mortality improves over time, particularly for high risk patients, and the database models must be recalibrated to reflect this change [6]. However, the STS online risk calculator used by clinicians for a bedside risk estimate is still based on the 2008 STS models with no recalibration since that time. [7] Evidence has shown that all these issues cause the STS tool to overestimate risk for mortality in high-risk cases.

Based on the above, it is logical to conclude that the strengths of the STS database for risk predictions outweigh its weakness except perhaps for one clinical scenario – using the STS online risk calculator tool to try to discriminate a patient that is above a high-risk score cutoff. When a patient has a score deemed to be low-risk, our confidence in this estimate can be high and the patient can be confidently reassured. However, a score that comes back as high risk for surgery should be viewed skeptically, at least initially. Such an adverse assessment may very well be accurate and provide us with useful information. However, the known inaccuracies of the model within this patient subset obliges us to exercise due diligence, particularly when it leads to a conclusion that a patient is too high risk for surgery. We demonstrate this by asking the following:

1. Are the general impressions of the clinical team of the patient's risk favorable (i.e. the patient "passes the eyeball test")? [8]
2. Is the patient free from any important unmeasured risk factors?
3. Can the typical approach to surgery be modified to reduce mortality risk?
4. Are the patient and family highly motivated to accept risk?

When the answers to these questions are all "yes", it is likely that the risk score is overestimated. It is unfair to use an overestimate as a sole reason to exclude a patient from the benefit of a life-saving operation. Unfortunately, this is common practice for case selection committees at many hospitals. They exclude patients from surgical consideration if there is a single machine generated estimate of risk for mortality that exceeds 8.0%, often with no opportunity for even a discussion of these cases. High risk cases are the exact ones which benefit the most from the judgments and experience of the multidisciplinary members of the cardiac program that attend these meetings. A score >8.0% can put the final decision on autopilot with no opportunity for change.

There is tremendous value of using computerized risk assessment as an aide to choose high risk cases wisely. But "the devil is in the details". The damning problem with the rigid protocol is not that it is based on a poor statistical understanding of how databases work. More importantly, it needlessly pits the risk assessments of the STS calculator against human judgment, creating an imaginary conflict of machine vs. man like in Terminator or The Matrix. One envisions hospital administrators preparing for the day when clinicians eventually band together behind Schwarzenegger or Reeves to stop the STS machine from oppressing the judgment of its haggard physicians.

Instead of that comic book scenario, maybe we can learn from another high reliability field struggling with their own man v. machine dilemma: airline pilots and their use of autopilot. [9] Autopilot improves overall airline safety, but some pilots cancel out its benefits by misusing it. Many crash investigations have documented the problems that come from when a pilot's attitude about autopilot is "set it and forget it". The pilots of Korean Airlines 214, Continental 3407 and Aeroflot 593 all put their blind trust in this tool, causing them to idly stand by as it led to a crash. If the machine says its so, it must be true.

Both medicine and aviation would be better served by reframing their challenges to automation not as man vs. machine but instead as man plus machine. A high performing team views the STS score and autopilot as key teammates. Like any teammate, the point of their automated outputs is to challenge our judgments. However, it is also our job to challenge theirs.

Everyone, even the most brilliant teammate on earth, is fallible. We are not being good teammates if we accept anything on blind faith.

Most important and above all else, humans (not machines) get the final say. The many problems that arise when that rule is not followed are bizarre and tragic. Boeing designed an autopilot software program (MCAS) that was able to intervene on the flight of its 737 MAX jets without pilot input. Seemingly out of the blue, that MCAS software thrust two separate jets downward directly into the earth, killing everyone on board, based on faulty input signals suggesting an abnormal angle of those two planes that was obviously incorrect to both pilots. Likewise, a recent risk score of >8% triggered an autopilot decision at our hospital exclude a salvageable patient from surgical consideration. Like a 737 MAX jet, this unlucky patient soon crashed from untreated coronary artery disease while our Heart Team remained willfully blind to the clear inaccuracies of the patient's assigned risk score.

If we recognize that all team members have their limitations, we will use the automated risk scores when they are likely to be accurate and engage in multidisciplinary debate about the best course of action when they aren't.

References

1. Caceres M, Braud RL, Garrett Jr HE (2010) A short history of the Society of Thoracic Surgeons national cardiac database: perceptions of a practicing surgeon. The Annals of thoracic surgery 89: 332-9.

2. Khan AA, Murtaza G, Khalid MF, Khattak F (2019) Risk Stratification for Transcatheter Aortic Valve Replacement. Cardiol Res. 10: 323-30.

3. Němec P, Fila P, Štěrba J, Černošek J (2015) Value of autopsy in cardiac surgery. Cor et Vasa 57: e91-4.

4. Cowan N (2016) Working memory capacity: Classic edition. Routledge; 2016 Apr 14.

5. Khan AA, Murtaza G, Khalid MF, Khattak F (2019) Risk Stratification for Transcatheter Aortic Valve Replacement. Cardiol Res 10: 323-30.

6. Hansen LS, Hjortdal VE, Andreasen JJ, Mortensen PE, Jakobsen CJ (2015) 30-day mortality after coronary artery bypass grafting and valve surgery has greatly improved over the last decade, but the 1-year mortality remains constant. Ann Card Anaesth 18: 138-42.

7. Vassileva CM, Aranki S, Brennan JM, Kaneko T, He M, et al. (2015) 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. The Annals of thoracic surgery 100: 2109-16.

8. Jain R, Duval S, Adabag S (2014) How accurate is the eyeball test? A comparison of physician's subjective assessment versus statistical methods in estimating mortality risk after cardiac surgery. Circulation: Cardiovascular Quality and Outcomes 7: 151-6.

9. Konnikova M (2014) The hazards of going on autopilot. The New Yorker.