See Article page 176.

Commentary: Universal surgical risk models: The answer you receive depends on the question you ask

Oliver K. Jawitz, MD, and
Charles M. Wojnarski, MD, MS

Oliver K. Jawitz, MD, and Charles M. Wojnarski, MD, MS

CENTRAL MESSAGE

Patient- and procedure-specific risk models likely outperform “universal” aggregate perioperative risk models and should be sought out whenever possible to accurately predict individual patient risks.

In this issue of the *Journal*, Mori and colleagues¹ provide an expert commentary regarding the pitfalls associated with “universal” surgical risk models. They highlight 2 commonly used examples of such models—the National Surgical Quality Improvement Program surgical risk model and European System for Cardiac Operative Risk Evaluation II—which were both devised and validated using a heterogeneous sampling of surgical procedures, albeit to varying degrees. Although these “universal models” may perform well using a large and diverse patient population, these performance values may be misleading to the physician who is attempting to risk stratify their single patient for a single surgical procedure. Instead, most important to the practicing clinician are data regarding model performance for a particular procedure type in question, data that are rarely available or presented.

Mori and colleagues¹ elegantly illustrate their point by fitting a multivariable logistic regression model for in-hospital mortality using data from the National Inpatient Sample database composed of an unlikely combination of surgeries across the spectrum of risk and average surgeon volume: coronary artery bypass grafting, esophagectomy, and cholecystectomy. This example “universal” model, which included key demographic characteristics in addition to the well-established Elixhauser comorbidity index, performed well overall (c-statistic 0.864, Brier score 0.011). When applied to patients undergoing individual

procedure types, however, the model’s performance suffered significantly, thus demonstrating the importance of critically evaluating “universal” risk prediction models before applying them to individual patients.

There are several important lessons to be learned from this work beyond the shortcomings of “universal” models. It is vital to remember that the strength and utility of a risk prediction model is only as robust as the applicability of the covariates to the sample population studied. Many widely used comorbidity indices including the Charlson and Elixhauser scores were originally developed from a diverse population of inpatients and thus do not contain the granularity necessary to accurately predict the risk associated with highly specific populations, such as those with esophageal cancer undergoing esophagectomy or coronary artery disease requiring coronary artery bypass grafting.² Indeed, the original article describing the development of the Elixhauser index wisely concluded that “[covariates] had independent effects on outcomes and probably should not be simplified as an index because they affect outcomes differently among different patient groups.”³ Instead, risk indices developed and validated using specific surgical cohorts such as the Society of Thoracic Surgeons Cardiac Surgery Risk Models are more appropriate for effectively risk stratifying patients undergoing individual surgical procedures.⁴

As we move into the era of “big data” and machine learning, we may find that the problems we see today with traditional regression analyses might not apply to the

From the Division of Cardiovascular and Thoracic Surgery, Department of Surgery, Duke University Medical Center, Durham, NC.

Dr Jawitz received funding provided by National Institutes of Health T-32 Grant 5T32HL069749.

Disclosures: Authors have nothing to disclose with regard to commercial support.

Received for publication Jan 6, 2020; accepted for publication Jan 6, 2020; available ahead of print Jan 22, 2020.

Address for reprints: Charles M. Wojnarski, MD, MS, Duke University Medical Center, 2301 Erwin Rd, Durham, NC 27710 (E-mail: charles.wojnarski@duke.edu).

J Thorac Cardiovasc Surg 2020;160:184-5

0022-5223/\$36.00

Copyright © 2020 by The American Association for Thoracic Surgery

<https://doi.org/10.1016/j.jtcvs.2020.01.009>

future of prediction modeling.⁵ Newer methodologies are able to improve model performance for procedures with low prevalence, but do so at the expense of understanding interactions between specific covariates and their associations with an outcome. These modern techniques should continue to be used in concert with more time-honored methods to help researchers answer the specific questions being asked of the data available.

Ultimately, Mori and colleagues' analysis¹ reminds us of the limitations inherent to conventional risk prediction models when it comes to guiding clinical decision making. Risk models are formulated using historic patient populations under unique circumstances with varying degrees of overlap with the patient we are trying to counsel. As such, these models are simply tools in the surgeon's armamentarium and should never replace the art of examining a patient, reviewing their history, and facilitating

a patient-centric discussion regarding the risks and benefits of operative intervention.

References

1. Mori MS, Shahian DM, Huang C, Li S, Normand ST, Geirsson A, et al. Surgeons: buyer beware—does 'universal' risk prediction model apply to patients universally? *J Thorac Cardiovasc Surg.* 2020;160:176-9.e2.
2. Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative data bases. *Int J Epidemiol.* 2000; 29:891-8.
3. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care.* 1998;36:8-27.
4. Shahian DM, O'Brien SM, Filardo G, Ferraris VA, Haan CK, Rich JB, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1—coronary artery bypass grafting surgery. *Ann Thorac Surg.* 2009; 88(1 Suppl):S2-22.
5. Corey KM, Kashyap S, Lorenzi E, Lagoo-Deenadayalan SA, Heller K, Whalen K, et al. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data (Pythia): a retrospective, single-site study. *PLoS Med.* 2018;15: e1002701.