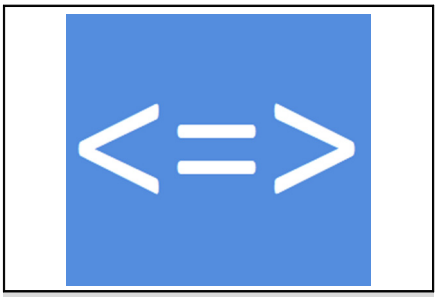
See Article page 176.

# Commentary: To underfit and to overfit the data. This is the dilemma

Umberto Benedetto, MD, PhD, and Arnaldo Dimagli, MD



The right balance between overfitting and underfitting in risk modeling.

**CENTRAL MESSAGE**

Risk modeling should always consider the drawback of overfitting due to the inclusion of a large number of parameters.

We read with great interest the paper from Mori and colleagues[1] on the potential risk of poor performance of prediction models designed to be applied to heterogeneous groups of surgical patients (the so-called universal model). They trained and tested a model including cholecystectomy, coronary artery bypass graft, and esophagectomy. They concluded that the model performance was reduced when applied to a specific subset of procedures, in particular with esophagectomy.

However, the authors' conclusions highlight possible limitation of these models and suggest that poor representation of low-volume case, model performance changes by the included case types, and variable effect sizes of unobserved covariates between case types can explain the poor performance observed in specific subset.

This manuscript looks at the oldest dilemma in risk modeling just from a different angle: the bias variance tradeoff.[2] In fact, a universal model will focus on a restricted number of variables that are common among different procedures. This model may be too simple and with very few parameters (underfitting); then, it may have high bias (difference between the average prediction of our model and the correct value which we are trying to predict). In contrast, if our model has large number of parameters to capture all possible aspects of individual procedures (overfitting), it will perform very well on training data but will have high error rates on test data (high variance).

The poor performance of the universal model tested by the authors can partially be attributed to the fact that variables were included in the model without a variable selection process such as (ie, recursive feature elimination). We also should note that the poor performance of the universal model can be simply related to risk overestimation in a subgroup at greater risk frequently observed with the risk model based on logistic regression.[3] In fact, the universal model developed by the authors poorly performs in esophagectomy, which is the procedure with the greatest observed mortality. Moreover, the authors have not specified the dataset time period, and the poor performance can be partially explained by model calibration drift due to improvement in quality of care over the time or chance in case mix.[4]

Broadly speaking, a universal model is very appealing because its implementation would allow the comparison of center and surgeon performance across a wide spectrum of surgical procedures. The statistical challenge remains and applies to any dataset: the balance between overfitting and underfitting the data.

## 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. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc Natl Acad Sci USA*. 2019;116:15849-54.
3. Provenchère S, Chevalier A, Ghodbane W, Bouleti C, Montravers P, Longrois D, et al. Is the EuroSCORE II reliable to estimate operative mortality among octogenarians? *PLoS One*. 2017;12:e0187056.
4. Hickey GL, Grant SW, Murphy GJ, Bhabra M, Pagano D, McAllister K, et al. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *Eur J Cardiothorac Surg*. 2013;43:1146-52.

From the Bristol Heart Institute, University of Bristol, Bristol, United Kingdom.

Disclosures: Authors have nothing to disclose with regard to commercial support.
Received for publication Dec 19, 2019; accepted for publication Dec 19, 2019; available ahead of print Jan 11, 2020.

Address for reprints: Umberto Benedetto, MD, PhD, Bristol Royal Infirmary, Office Room 84, Level 7, Upper Maudlin St, BS2 8HW Bristol, United Kingdom (E-mail: umberto.benedetto@bristol.ac.uk).

J Thorac Cardiovasc Surg 2020;160:183
0022-5223/$36.00



https://doi.org/10.1016/j.jtcvs.2019.12.079