Encrypted prediction: A hacker's perspective

Tara Karamlou, MD, MSc,^a Daniel A. Velez, MD,^a and John J. Nigro, MD^b

From the ^aDivision of Pediatric Cardiac Surgery, Phoenix Children's Hospital, Phoenix, Ariz; and ^bDivision of Pediatric Cardiac Surgery, Rady Children's Hospital and University of California, San Diego, San Diego, Calif. Disclosures: Authors have nothing to disclose with regard to commercial support.

Received for publication Aug 27, 2017; accepted for publication Aug 30, 2017; available ahead of print Sept 30, 2017.

Address for reprints: Tara Karamlou, MD, MSc, The Mayo Clinic and Phoenix Children's Hospital, 1919 E Thomas Rd, Heart Center, Main Building, Phoenix, AZ 85016 (E-mail: tkaramlou@phoenixchildrens.com). J Thorac Cardiovasc Surg 2017;154:2038-40

0022-5223/\$36.00

Copyright © 2017 by The American Association for Thoracic Surgery

<https://doi.org/10.1016/j.jtcvs.2017.08.112>

Tara Karamlou, MD, MSc

Central Message

Advanced analytic techniques must be accessible to clinicians. Moreover, development of models for complex conditions should adhere to the general principle that parsimony is preferred to complexity.

See Article page 2030.

In this issue of the *Journal*, Gupta and colleagues¹ regale us with their rendition of a prediction model for outcomes (specifically preoperative and postoperative in-hospital mortality and “poor composite outcome”) for children undergoing the Norwood operation. Gupta and colleagues¹ used the Single Ventricle Reconstruction Trial public-use data set, and the risk models were constructed with complex statistical methods, including Bayesian conditional probit regression and Monte Carlo simulations. The article describes the creation of these prediction models and also their validation, which unfortunately used the same data set. To be sure, in the era of personalized or precision medicine, prediction of prevalent or incident diseases and their expected convalescent course has become more critical for health care providers. For complex congenital heart surgery such as the Norwood operation, prediction of expected outcomes is critical to inform care pathways, facilitate shared decision making between parents and physicians, and reduce heterogeneity of practice across centers. It is therefore laudable that Gupta and colleagues¹ are moving the analytic needle vis-à-vis prediction by using these advanced techniques. In our opinion, however, the article by Gupta and colleagues¹ has so densely encrypted their results in extensive statistical jargon without adequate description or context that it has missed the target.

The underlying criteria of a successful prediction tool are that it has utility for the generalized audience and that it is effective in directing future therapy. In other words, is it practical to apply the models to clarify expectations for children undergoing the Norwood operation? Unfortunately, because the collective emphasis in the report of Gupta and colleagues¹ is concentrated on methodologic minutiae, neither the utility of the tool nor its potential impact is evident. Moreover, although Gupta and colleagues¹ are correct in their statement that individual patient-level prediction has not been described for patients undergoing the Norwood operation, a review of the most recent 10 articles (of the 160 indexed on PubMed) that elucidate risk factors for mortality or morbidity after the Norwood operation reveal nearly *all* the same factors uncovered by the

prediction model in the current report.²⁻¹¹ So, we were left wondering whether the use of these innovative analytics provided an important return on the investment.

Not to fall into the same convoluted trap here, let us spend just a few moments (we promise) deconstructing the methodology. First, what is a clinical prediction tool (or rule)? Clinical prediction tools may be classified into 3 distinct groups: (1) diagnostic, (2) prognostic, and (3) prescriptive.^{12,13} Studies that focus on predictive factors related to a specific diagnosis (or procedure) are diagnostic tools. Tools that are designed to predict an outcome, such as mortality or survival, are considered prognostic. Finally, tools that are designed to identify the most effective interventions are prescriptive. Not surprisingly, the last sort require prospective, longitudinal, randomized, controlled trials that compare outcomes after selected interventions for subjects with similar scores for the clinical prediction tool. In the current report, Gupta and colleagues¹ define their prediction tool as prognostic in some sections, as predictive in other sections, and as prescriptive (insofar as identification of procedural factors [interventions]) in others. The first step in decryption needed here is to define clearly the purpose and scope of the elucidated models and restrict potential import to these defined realms.

Second, what are the purported advantages of Bayesian methodology to construct prediction tools? Bayesian inference is a method for formulating beliefs about uncertainty. One of the central concepts of Bayesian statistics is that

“probability is orderly opinion, and that inference from data is nothing other than the revision of such opinion in the light of relevant new information.”¹⁴ In simpler terms, individuals innately form decisions on the basis of existing (*prior*) beliefs when faced with uncertainty. Such beliefs are unique to each individual and represent collective “historic” experiences. Rational individuals will integrate new information in an iterative manner, updating their previous beliefs. The new belief, called the *posterior belief*, is therefore based on the prior belief conditioned on the new information. The ability to update beliefs is fundamental to the Bayesian methodology, and is described in Bayes theorem.¹⁴ Gupta and colleagues¹ have leveraged the ability of Bayesian techniques to account for model uncertainty and compute the entire probability distribution for individual patient risk. Further, the methods used obviate the need to describe covariables as a single point estimate, which introduces flexibility to account for inferential uncertainties. The ability to incorporate latent (unobserved) variables in a hierarchic framework is another important advantage of the methods described. The provision of relevance, however, was not adequately presented by Gupta and colleagues.¹ Had they injected sufficient discussion to deconstruct these principles, the potential import of their findings could have been clarified.

The third issue obscured in this report was the utility of these advanced analytics to improve outcome discrimination. Most readers are familiar with basic criteria for assessing model performance, such as the concept of explained variance (R^2) for linear models or the c-statistic for logistic regression models. Gupta and colleagues¹ report a “measure of the variability in outcomes captured by the predictor variables” within the online supplement, but not within the body of the article. Further, the areas under the curve for the mortality model and for the “composite poor outcome” were 0.77 and 0.72, respectively. Although these are reasonable, they are not incrementally better than previously described risk models for post-Norwood outcomes.²⁻¹¹ It is likely that machine learning, random forests, and neural networks may considerably improve the performance of any clinical prediction tool (including the one described here), but any mention of these techniques is glaringly absent within this article.

Finally, although a philosophical digression about the merits of simplicity is beyond the scope of this commentary, one of the fundamental concepts in statistical analysis is that parsimonious models are generally preferable to more complex models. This concept is particularly relevant to an analysis of the models described by Gupta and colleagues,¹ because of the limited sample size (indeed only 88 patients died in hospital, and although 241 patients achieved the composite poor outcome, this was composed of a heterogeneous group of events with very different prevalences within the cross-sectional in-hospital end point) and the

limited number of variables available for inclusion. A limited data set and sample size assume greater importance within the structural framework used by Gupta and colleagues,¹ because this framework incorporates latent (“hidden” or unobserved) variables through which inferences are made on the basis of shared variance with the *observed* variables. Parsimonious models can mitigate these issues because they reduce overfitting and generalization error, both of which are relevant concerns with the data set used in the report of Gupta and colleagues.¹ Although the current emphasis on big data analytics and machine learning, which are feasible because of the seemingly infinite computing power available today, may alter our perspectives on the issue of complexity versus parsimony, the case for simplicity (or realism) is still worth elucidating.

It is evident that Gupta and colleagues¹ have taken an important formative step in their application of advanced statistical methods to a highly visible public data set. The creation of the online application is another helpful mechanism to facilitate adoption and render these complex concepts more accessible. One important obstacle to the advancement of statistical analysis within our field, however, is dispelling the popular notion articulated by, among many others, Mark Twain, “There are lies, damned lies, and statistics.”¹⁵ It is incumbent on investigators presenting these techniques to make them understandable and inspire confidence in results derived from their use. I suspect that had Gupta and colleagues¹ aligned their presentation of the methods and results with this objective in mind, their message would be better received—“no decryption required.”

References

- Gupta P, Chakraborty A, Gossett JM, Rettiganti M. A prognostic tool to predict outcomes in children undergoing the Norwood operation. *J Thorac Cardiovasc Surg*. 2017;154:2030-7.e2.
- Tabbutt S, Ghanayem N, Ravishankar C, Sleeper LA, Cooper DS, Frank DU, et al; Pediatric Heart Network Investigators. Risk factors for hospital morbidity and mortality after the Norwood procedure: a report from the Pediatric Heart Network Single Ventricle Reconstruction trial. *J Thorac Cardiovasc Surg*. 2012;144:882-95.
- Chowdhury SM, Graham EM, Atz AM, Bradley SM, Kavarana MN, Butts RJ. Validation of a simple score to determine risk of hospital mortality after the Norwood procedure. *Semin Thorac Cardiovasc Surg*. 2016;28:425-33.
- Hornik CP, He X, Jacobs JP, Li JS, Jaquiss RD, Jacobs ML, et al. Relative impact of surgeon and center volume on early mortality after the Norwood operation. *Ann Thorac Surg*. 2012;93:1992-7.
- Ohye RG, Gaynor JW, Ghanayem NS, Goldberg CS, Laussen PC, Frommelt PC, et al; Pediatric Heart Network Investigators. Design and rationale of a randomized trial comparing the Blalock-Taussig and right ventricle-pulmonary artery shunts in the Norwood procedure. *J Thorac Cardiovasc Surg*. 2008;136:968-75.
- Ohye RG, Sleeper LA, Mahony L, Newburger JW, Pearson GD, Lu M, et al; Pediatric Heart Network Investigators. Comparison of shunt types in the Norwood procedure for single-ventricle lesions. *N Engl J Med*. 2010;362:1980-92.
- Pasquali SK, Jacobs JP, Xe H, Hornik CP, Jaquiss RD, Jacobs ML, et al. The complex relationship between center volume and outcome in patients undergoing the Norwood operation. *Ann Thorac Surg*. 2012;93:1556-62.
- Wilder TJ, McCrindle BW, Phillips AB, Blackstone EH, Rajeswaran J, Williams WG, et al. Survival and right ventricular performance for matched children after stage-I Norwood: modified Blalock-Taussig shunt versus

right-ventricle-to-pulmonary-artery conduit. *J Thorac Cardiovasc Surg.* 2015; 150:1440-50.

9. Wilder TJ, McCrindle BW, Hickey EJ, Ziemer G, Tchervenkov CI, Jacobs ML, et al; Congenital Heart Surgeons' Society. Is a hybrid strategy a lower-risk alternative to stage 1 Norwood operation? *J Thorac Cardiovasc Surg.* 2017;153: 163-72.e6.
10. Kulkarni A, Neugebauer R, Lo Y, Gao Q, Lamour JM, Weinstein S, et al. Outcomes and risk factors for listing for heart transplantation after the Norwood procedure: an analysis of the Single Ventricle Reconstruction Trial. *J Heart Lung Transplant.* 2016;35:306-11.
11. Cross RR, Harahsheh AS, McCarter R, Martin GR, National Pediatric Cardiology Quality Improvement Collaborative. Identified mortality risk factors associated with presentation, initial hospitalisation, and interstage period for the Norwood operation in a multi-centre registry: a report from the National Pediatric Cardiology-Quality Improvement Collaborative. *Cardiol Young.* 2014;24: 253-62.
12. Tetreault L, Le D, Côté P, Fehlings M. The practical application of clinical prediction rules: a commentary using case examples in surgical patients with degenerative cervical myelopathy. *Global Spine J.* 2015;5:457-564.
13. Lee Y, Bang H, Kim DJ. How to establish clinical prediction models. *Endocrinol Metab (Seoul).* 2016;31:38-44.
14. Edwards W, Lindman H, Savage LJ. Bayesian statistical inference for psychological research. *Psychol Rev.* 1963;70:193-242.
15. Twain M, Smith HE, Griffin B, Fischer V, Frank MB, eds. *Autobiography of Mark Twain: the Complete and Authoritative Edition, Volume 1.* Berkeley: University of California Press; 2010:228.