Commentary
Artificial neural networks as prediction tools in the critically ill
Gilles Clermont

Co-Director, The CRISMA (Clinical Research, Investigation, and Systems Modeling of Acute Illness) Laboratory, Department of Critical Care Medicine, and Medical Director of The Center for Inflammatory and Regenerative Modeling, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Corresponding author: Gilles Clermont, clermontg@upmc.edu

See related research by Jaimes *et al*. in this issue [http://ccforum.com/content/9/2/R150]

Abstract

The past 25 years have witnessed the development of improved tools with which to predict short-term and long-term outcomes after critical illness. The general paradigm for constructing the best known tools has been the logistic regression model. Recently, a variety of alternative tools, such as artificial neural networks, have been proposed, with claims of improved performance over more traditional models in particular settings. However, these newer methods have yet to demonstrate their practicality and usefulness within the context of predicting outcomes in the critically ill.

Introduction

The science of outcome prediction is particularly useful in the setting of the emergency room – the entry point of many to acute care. In this issue of *Critical Care* Jaimes and coworkers [1] evaluate the usefulness of artificial neural networks (ANNs) in predicting hospital mortality in patients presenting to the emergency room with suspected sepsis. The construction of a prediction tool is a difficult undertaking; it requires careful methodological consideration and validation before the predictions can be deemed valid and reliable in naïve patients [2,3]. These tools identify the presence of associations between the outcome of interest and empiric risk factors that contribute to this outcome. A well designed tool will typically possess three qualities: discrimination (the ability to identify accurately those patients who will reach the outcome from those who will not), goodness of fit (the ability to match accurately predicted and actual outcomes, such as mortality rate, in subgroups of patients), and the ability to achieve these predictions in cohorts of patients similar to those in which the tool was developed [4,5].

Artificial neural networks as prediction tools

Most predictive tool use logistic regression – a well vetted statistical technique that is applicable to situations in which the outcome is binary (e.g. survival/death), measured at a predetermined time in the future [6]. The technique can precisely quantify the relative contribution of each risk factor to outcome, typically crystallized as the odds ratio (i.e. the relative odds that a patient with the risk factor has of reaching the outcome as compared with a patient without the risk factor). ANNs represent an alternative technique for achieving predictions. The key difference between the two techniques is that the contribution of each risk factor is not as rigidly dictated with ANNs as it is in a logistic regression model. ANNs can improve predictions by extracting information drawn from unforeseen interactions between predictors. Arguably, if a modeler had foresight of the important interactions present, then they could design an accurate standard logistic regression model without the heavy price associated with the use of an ANN.

Often considered as 'black boxes', ANNs suffer from several shortcomings. There is no clear prescription to construct an ANN. Off the shelf software is readily available, but there is little guidance as to how ANN parameters, such as number of intermediate (hidden) layers of neurons, number of neurons in those layers, learning rate, activation functions, and several other tuning parameters, should be chosen and tuned for optimal performance by a novice user. ANNs, by virtue of the presence of numerous weights linking the neurons, can accommodate nonlinearity but they also include very large numbers of parameters. There are no clear techniques that provide limits of confidence for those parameters. Consequently, the relative contribution of each input risk factor to the outcome is difficult to quantify. Because of the large number of parameters, it is very easy for an ANN to overfit the development dataset. In other words, the predictions offered by the ANN will be overly optimistic in the original population of patients, and this good performance will not generalize well to populations to which the ANN is naïve.

ANN = artificial neural network.

Jaimes and coworkers [1] do not offer a prediction in a naïve population, nor do those investigators clearly indicate whether techniques were applied to minimize the risk of overfitting. Thus, their conclusion may be overstated. Limited measures guarding against overfitting can be applied within a development set, without specific recourse to an independent validation set. Finally, preparing input data for an ANN requires some amount of *a priori* knowledge. For example, white blood cell count or fever typically have a 'U'-shaped relationship to outcome. Both low and high values portend a bad outcome, whereas intermediate values are normal. The predictions of an ANN will be significantly improved if this unintuitive relationship between risk factor and outcome is already 'known' to the modeler ahead of time.

ANNs have been used by many investigators to predict outcomes in the critically ill [7–11]. Clearly, ANNs – as prediction tools – require significant expertise to build, have significant shortcomings, and should be developed and validated as rigorously as logistic regression models [12]. ANNs do offer greater flexibility and may allow the identification of unforeseen interactions. ANNs can predict continuous outcomes, and thus offer an alternative to multivariable regression. Similarly, ANNs are easily scalable to ordinal or categorical outcomes, and to survival analysis [13], which is much less familiar territory for the critical care physician. However, ANNs can easily be misused, typically because their limitations are not well understood [14,15].

Conclusion

Therefore, in settings in which familiar tools such as logistic regression can be applied, ANNs should be reserved for those situations where standard models do not perform well; where one suspects the presence of intense but poorly characterized interactions between risk factors; where one does not particularly care about quantifying the relative contributions of risk factors; where appropriate validation is possible; and, most importantly, where proper expertise is readily available.

The science of developing prediction models is best left to the expert, but the clinician can contribute invaluable content knowledge to the process.

Competing interests

Dr Clermont is a paid consultant to Amgen, Altana, and Immunetrics, LLC. The current manuscript was realized without any financial or other material support from commercial entities.

References

1. Jaimes F, Farbiarz J, Alvarez D, Martinez C: **Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room.** *Crit Care* 2005, **9:**R150-R156.
2. Lemeshow S, Klar J, Teres D: **Outcome prediction for individual intensive care patients: useful, misused, or abused?** *Intensive Care Med* 1995, **21:**770-776.
3. Clermont G, Angus DC: **Severity scoring systems in the modern intensive care unit.** *Ann Acad Med Singapore* 1998, **27:**397-403.
4. Lemeshow S, Teres D, Avrunin JS, Pastides H: **Predicting the outcome of intensive care unit patients.** *J Am Stat Assoc* 1988, **83:**348-356.
5. Schuster DP: **Predicting outcome after ICU admission. The art and science of assessing risk.** *Chest* 1992, **102:**1861-1870.
6. Lemeshow S, Hosmer DW Jr: **A review of goodness of fit statistics for use in the development of logistic regression models.** *Am J Epidemiol* 1982, **115:**92-106.
7. Buchman TG, Kubos KL, Seidler AJ, Siegforth MJ: **A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit.** *Crit Care Med* 1994, **22:**750-762.
8. Doyle HR, Dvorchik I, Mitchell S, Marino IR, Ebert FH, McMichael J, Fung JJ: **Predicting outcomes after liver transplantation. A connectionist approach.** *Ann Surg* 1994, **219:**408-415.
9. Doig GS, Inman KJ, Sibbald WJ, Martin CM, Robertson JM: **Modeling mortality in the intensive care unit: comparing the performance of a back-propagation, associative-learning neural network with multivariate logistic regression.** *Proc Annu Symp Comput Appl Med Care* 1993, 361-365.
10. Hotchkiss JR, Broccard AF, Crooke PS: **Artificial neural network prediction of ventilator-induced lung edema formation.** *Crit Care Med* 2003, **31:**2250.
11. Clermont G, Angus D, DiRusso S, Griffin M, Linde-Zwirble W: **Predicting hospital mortality for patients in the intensive care unit: a comparison of artificial neural networks with logistic regression models.** *Crit Care Med* 2001, **29:**291-296.
12. Randolph AG, Gordon HG, Calvin JE, Doig G, Richardson WS: **Understanding articles describing clinical prediction tools.** *Crit Care Med* 1998, **26:**1603-1612.
13. de Laurentiis M, Ravdin PM: **A technique for using neural network analysis to perform survival analysis of censored data.** *Cancer Lett* 1994, **77:**127-138.
14. Wyatt J: **Nervous about artificial neural networks?** *Lancet* 1995, **346:**1175-1177.
15. Schwarzer G, Vach W, Schumacher M: **On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology.** *Stat Med* 2000, **19:**541-561.