The Siren Song of Simple Tools That Predict Mortality

> It is in my opinion a most excellent thing for the physician to practice forecasting. — Hippocrates

Hippocrates recognized that determining a patient's prognosis is one of physicians' most important responsibilities.[1] Although research on prognosis has been outpaced by that on etiology and therapy, scientists have made great strides in characterizing the risk factors for death across all areas of medicine.[2] However, this deeper understanding of prognosis comes at the expense of simplicity; in many cases providers must now integrate dozens of factors—implicitly or explicitly—to accurately estimate an individual's prognosis. Since surmounting this complexity is challenging, providers often use heuristics or rules of thumb to reduce prognostic assessments to simple judgments.[3] These heuristics, although helpful in many situations, may lead to biased estimates of prognosis.[3,4] For example, a provider may turn to a recent memory of a patient with pneumonia who fared well when estimating the risk of death for a patient with pneumonia she is caring for today (ie, availability bias).[3,4] Even if a provider avoids heuristics and assimilates the necessary information needed to determine an individual's risk of death, he must then overcome his own biases (eg, age, experience, religion), which may unduly influence his estimate of the risk of death.[5-7] For these reasons, providers are in great need of tools, such as predictive models or decision aids that can objectively incorporate complex information and provide mortality estimates that are reproducible and immune to bias. This need for objective measures is particularly relevant to critical care where illness is highly complex, evolves rapidly, and commonly results in poor outcomes.

See the Original Study on Page 420

To address this need, critical care investigators have developed several of the most widely recognized prognostic models, including the Acute Physiology and Chronic Health Evaluation (APACHE)[8] and the Simplified Acute Physiology Score (SAPS).[9] These models, and their more recent versions, can consistently and accurately estimate the risk of death for populations of critically ill patients throughout the world. But this strength—accuracy for a *general* intensive care unit (ICU) population—is also a weakness. When these models are applied to populations with a *specific* disease, such as cancer or acute renal failure, among others, their ability to provide accurate estimates decreases.[10,11]

One strategy investigators use to improve mortality prediction beyond general prognostic measures, such as those highlighted above, is to develop a disease-specific model, one that incorporates variables known to be particularly relevant to the disease of interest. This is the approach taken by Villar et al[12] in an analysis reported in this issue of Respiratory Care. With a cohort of 170 patients with acute respiratory distress syndrome (ARDS) enrolled in a multicenter observational study in Spain, they developed and validated a predictive model for death specific to ARDS. They first divided 11 measured continuous variables of interest (eg, age and airway plateau pressure) into tertiles and labeled each tertile as either low-risk, medium-risk, or high-risk, depending on its association with death. Among the 11 variables, each now represented by its tertiles, they determined that age, airway plateau pressure, and P_{aO_2}/F_{IO_2} were significantly associated with mortality. They further simplified their analysis by then counting the total number of high-risk tertiles for these 3 variables present in each patient. This final variable, the high-risk tertile count, was remarkably predictive of mortality. Ninety percent of patients with 3 high-risk tertiles died, whereas only 12% of patients with no high-risk tertiles died. Moreover, the probability of death predicted by the total number of high-risk tertiles was better able to discriminate patients who lived from those who died than that predicted by the APACHE II model.

Villar and colleagues are not the first group of investigators to seek an accurate prognostic index for ARDS patients. In fact, at least a dozen prior studies aimed to fully characterize the predictors of mortality among patients with acute lung injury (ALI) or ARDS or to integrate predictors of death into a prognostic index specific to this disease.[13-16] Villar et al are also not the first to identify age, airway plateau pressure, and oxygenation as predictors of mortality in ARDS. Several studies previously identified these variables as independently associated with death.[15,17,18]

What, then, does this study[12] add to the literature? Villar and colleagues deviate from most, but not all,[16] prior studies by developing and validating an extremely simple prediction tool that retains excellent performance. For prediction models to be useful to clinicians, they must be accurate,

but they must also be easy to use. Because prediction tools such as the APACHE model and most ARDS models rely on dozens of variable inputs to predict mortality, a calculator or computer is needed to estimate the risk of death.[15] The need for a computer is trivial when these models are used in the research context for risk-adjustment or for benchmarking ICU performance, but can be impractical when employed clinically. Only the simplest tools stand a chance at being used at the bedside, given the time constraints placed on critical care practitioners. Yet simple models often sacrifice accuracy because they omit predictive variables. Not so for the model generated by Villar et al, which is simple yet highly predictive.[12]

Although we commend Villar et al for developing and validating a simple model that retains accuracy, there are several reasons their model should not yet form the basis of patient decisions, such as the timing of ICU discharge, as Villar et al[12] suggest, or withdrawal of life-sustaining therapy. First, even though their model performed well, 10% of patients with 3 high-risk tertiles survived and 12% of patients with no high-risk tertiles died. These rates are not close enough to zero to be helpful for decisions. Second, the development and validation sample included only 170 and 50 patients, respectively, which may be inadequate to support strong inference. Specifically, confidence intervals around the mortality estimates for each "tertile count" are quite wide, indicating considerable uncertainty in the mortality estimate for an individual. For example, if the study by Villar et al were repeated 100 times we would expect the average mortality of all patients with 3 high-risk tertiles to fall between 67% and 98%. In the face of uncertainty in the mortality estimates from this model, most providers would not be persuaded to change the course of care for a patient with ARDS. This limitation is not unique to the model of Villar et al; most prognostic models will never predict patient outcomes with 100% certainty, and are therefore much less useful for decision making at the bedside.[19,20]

Finally, although Villar et al[12] validated the prediction model in a novel cohort of patients, they did not test if the information provided by the model influences care. Future research should quantify whether the model's use in daily practice improves decision making and, ultimately, patient outcomes, using a comparative design (known as an impact study[20]) prior to its application in decision making.

If not yet ready for clinical decision making, how can we use the Villar et al model? If the model is validated in larger cohorts, it may be useful in several contexts. First, although it should not be used as the sole means to estimate a patient's risk of death, providers could use the model to inform their own bedside predictions of a patient's risk of death. For example, if a provider using her own judgment and available patient data believes the risk of death for an ARDS patient is high, this model may either reinforce or draw into question such a belief. At least one prior study found that the combination of model-based and physician-based estimates of mortality often outperform either in isolation.[21] Second, the model may be useful in patient-level decisions of lesser consequence. For example, clinical trials with patients with ARDS may utilize this score to stratify patients into each study arm to ensure that patients with similar risk of death are balanced between treatment and control groups. Finally, the model may prove useful for future research efforts in patients with ARDS, in which the model is used to evaluate an ARDS population rather than to make patient-level decisions. Severity of disease, as captured by this model, could be evaluated as a modifier of treatment effect in clinical trials of therapies thought to have greater impact among sicker patients. Analysis of clinical, observational, and translational studies may be able to use this score to better adjust for severity of illness as a confounding variable when exploring novel clinical, biologic or genetic data for their associations with mortality.

As we strive to provide the most objective estimates of prognosis for our patients with ARDS and their loved ones, we must find novel ways to deal with our increasingly nuanced understanding of the risk factors for death in this disease. Until published data and real-time clinical information are seamlessly integrated into accurate bedside decision aids, clinicians will rely on heuristics and simpler models to estimate mortality for critically ill patients. Yet even if we are able to build a more perfect model, we must remember the limitations intrinsic to these prognostic tools.[20]

Colin R Cooke MD MSc
Division of Pulmonary and Critical Care Medicine
and
Robert Wood Johnson Foundation
Clinical Scholars Program
University of Michigan
Ann Arbor, Michigan

REFERENCES

1. Hippocrates. The genuine works of Hippocrates. Baltimore: Williams & Wilkins; 1939.
2. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009; 338:b375.
3. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science 1974;185(4157):1124-1131.
4. Bornstein BH, Emler AC. Rationality in medical decision making: a review of the literature on doctors' decision-making biases. J Eval Clin Pract 2001;7(2):97-107.
5. Christakis NA, Asch DA. Biases in how physicians choose to withdraw life support. Lancet 1993;342(8872):642-646.

6. Christakis NA, Asch DA. Medical specialists prefer to withdraw familiar technologies when discontinuing life support. J Gen Intern Med 1995;10(9):491-494.
7. Christakis NA, Asch DA. Physician characteristics associated with decisions to withdraw life support. Am J Public Health 1995;85(3): 367-372.
8. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985; 13(10):818-829.
9. Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3–From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. Intensive Care Med 2005;31(10):1336-1344.
10. Mehta RL, Pascual MT, Gruta CG, Zhuang S, Chertow GM. Refining predictive models in critically ill patients with acute renal failure. J Am Soc Nephrol 2002;13(5):1350-1357.
11. den Boer S, de Keizer NF, de Jonge E. Performance of prognostic models in critically ill cancer patients - a review. Crit Care 2005; 9(4):R458-R463.
12. Villar J, Pérez-Méndez L, Basaldúa S, Blanco J, Aguilar G, Toral D, et al; Hospitales Españoles Para el Estudio de la Lesión Pulmonar (HELP) Network. A risk tertiles model for predicting mortality in patients with acute respiratory distress syndrome: age, plateau pressure, and P_{aO_2}/F_{IO_2} at ARDS onset can predict mortality. Respir Care 2011;56(4):420-428.
13. Heffner JE, Brown LK, Barbieri CA, Harpel KS, DeLeo J. Prospective validation of an acute respiratory distress syndrome predictive score. Am J Respir Crit Care Med 1995;152(5 Pt 1):1518-1526.
14. Monchi M, Bellenfant F, Cariou A, Joly LM, Thebert D, Laurent I, et al. Early predictive factors of survival in the acute respiratory distress syndrome. A multivariate analysis. Am J Respir Crit Care Med 1998;158(4):1076-1081.
15. Cooke CR, Kahn JM, Caldwell E, Okamoto VN, Heckbert SR, Hudson LD, et al. Predictors of hospital mortality in a population-based cohort of patients with acute lung injury. Crit Care Med 2008;36(5): 1412-1420.
16. Cooke CR, Shah CV, Gallop R, Bellamy S, Ancukiewicz M, Eisner MD, et al. A simple clinical predictive index for objective estimates of mortality in acute lung injury. Crit Care Med 2009;37(6):1913-1920.
17. Luhr OR, Antonsen K, Karlsson M, Aardal S, Thorsteinsson A, Frostell CG, et al; The ARF Study Group. Incidence and mortality after acute respiratory failure and acute respiratory distress syndrome in Sweden, Denmark, and Iceland. Am J Respir Crit Care Med 1999;159(6):1849-1861.
18. Ware LB, Koyama T, Billheimer DD, Wu W, Bernard GR, Thompson BT, et al. Prognostic and pathogenetic value of combining clinical and biochemical indices in patients with acute lung injury. Chest 2010;137(2):288-296.
19. Herridge MS. Prognostication and intensive care unit outcome: the evolving role of scoring systems. Clin Chest Med 2003;24(4):751-762.
20. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606.
21. Rocker G, Cook D, Sjokvist P, Weaver B, Finfer S, McDonald E, et al. Clinician predictions of intensive care unit mortality. Crit Care Med 2004;32(5):1149-1154.

Dr Cooke is the statistical consultant for RESPIRATORY CARE. He has disclosed no other conflicts of interest.

Correspondence: Colin R Cooke MD MSc, Division of Pulmonary and Critical Care Medicine, University of Michigan, 6312 Medical Science Building 1, 1150 W Medical Center Drive, Ann Arbor MI 48109. E-mail: cookecr@umich.edu.

DOI: 10.4187/respcare.01281