Predictive Models in Heart Failure
Who Cares?

Robert M. Califf, MD; Michael J. Pencina, PhD

Despite a rich and growing history of predictive modeling in medicine, few such models have been successfully incorporated into routine practice and decision support at the point of care. In 1981, a young cardiologist predicted that point-of-care predictive models would be in widespread use in the near future.1 But more than 30 years later, even with the advent of electronic health records, clinicians continue to struggle through rounds and clinic with little practical decision support—a situation that contrasts sharply with the worlds of business, finance, and politics, where prediction models are routinely used for this purpose.

This relative lack of applied prediction modeling in medicine is due in part to the insufficient adoption of available, thoroughly validated models (eg, the Framingham risk algorithms in primary prevention).2 A more significant contributor, however, is the preponderance of insufficiently validated prediction models, as can be seen clearly in the arena of heart failure. A new systematic review of predictive models for patients with heart failure by Alba et al3 in this issue of Circulation: Heart Failure provides important insights into deficiencies that have hindered uptake and use of these models by the clinical community.

One key problem derives from the fact that the models with the greatest levels of validation are derived from clinical trials or populations with limited generalizability and, therefore, represent only a sliver of the population of patients with heart failure seen in practice. These models are certainly useful for understanding risks attributable to the measured factors for critical outcomes in heart failure, as well as for adjusting for differences in subgroup analyses within trial populations, but the degree to which such a model can be used to make accurate predictions in an unselected population is questionable.

This point is underscored by the variable discrimination and calibration of the Seattle Heart Failure Model and the Heart Failure Survival Score documented by Alba et al.3 In contrast, the massive analysis of the SHOCKED model in the Centers for Medicare and Medicaid Services heart failure data set provides a generalizable model for a large segment of those heart failure patients who undergo implantable cardioverter/defibrillator placement, but because it is based on claims data it lacks many key measurements that discriminate among patients with heart failure with different prognoses and is directed only at the minority of heart failure patients who receive implantable cardioverter/defibrillators.

Alba et al3 raise another important issue, namely, the fragmentation of efforts aimed at model development and validation. As the authors point out, many centers or groups develop their own models, each of which is necessarily based on a limited sample. In the course of their systematic review, the authors identified 20 different risk models developed for patients with heart failure. The field would benefit substantially if these efforts were united around the goal of creating a comprehensive, generalizable, well-validated model.

In general, attempts to obtain appropriate validation samples for these models have fallen short. However, the quality of the validation process itself is also a critical element in building a useful and accurate model. The use of samples from populations that differ in major ways from those used to create the original model to assess performance cannot be considered validation. Thus, it is not entirely clear to what extent the results pertaining to model validation gathered by Alba et al describe the application of a given model to the population for which it was designed. This problem has important ramifications, especially regarding the assessment of model calibration. We agree with Alba et al that calibration is essential for prognostic models and that in general it is insufficiently quantified. But we also note that the essential issue concerns the population and outcome of interest—if the model is not applied to a sample from the same population and does not predict the same outcome, there are few reasons to expect that it would calibrate well.

Another limitation of the validation process relates to the sensitivity of the c-statistic to the distribution of key predictors in the sample under study. It is conceivable that some of the differences in discrimination noted by Alba et al3 could be explained by the width of the age ranges in the validation samples. Given that the outcome of mortality is highly dependent on age, it is likely that wider age ranges result in higher discriminative ability for a given model. As the authors correctly point out, other measures of model performance may be less sensitive to this issue.4

Current prognostic models have too much in common with diagnostic models and miss the opportunities that the time-to-event nature of the data afford. Clinicians are well aware

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

From the Duke Translational Medicine Institute and the Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, NC (R.M.C.); and Duke Clinical Research Institute, Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC (M.J.P.).

Correspondence to Robert M. Califf, MD, MACC, Duke Translational Medicine Institute, Duke University School of Medicine, 200 Trent Dr, 1117 Davison Bldg, Durham, NC 27710. E-mail robert.califf@duke.edu

Circ Heart Fail is available at http://circheartfailure.ahajournals.org
DOI: 10.1161/CIRCHEARTFAILURE.113.000659

© 2013 American Heart Association, Inc.

877
that in severe illnesses, a patient’s changing situation may indicate that a prediction made at baseline is not as accurate as when that patient is seen on rounds on the second day of hospitalization or in clinic a month later. Models that are updated based on changes in the patient’s condition could potentially be more useful in decision making.5 In many major industries, big data are routinely applied to inform key decisions, together with frequent updates as conditions change. Such dynamic decision support is purported to have been a major factor in our most recent presidential election6: using data mining and analytics, campaign staff could predict people likely to vote and identify optimal ways to reach these potential voters and motivate them to get to the polls. The chronic heart failure models examined in the article by Alba et al3 take patients with a diagnosis of heart failure; in most cases, the inception point is the time of entry into a clinical trial. The model predicts a key outcome (death; death or rehospitalization) at an unspecified time in the future but does not predict differential outcomes depending on a decision.

We are entering an era in which almost all Americans have an electronic health record that is increasingly coordinated with disease registries and integrated with patient-reported outcomes. In the near future, personal mobile devices will record and feed physiological data to electronic health records, as well as information about quality of life and patient preferences. Unfortunately, the heavily regulated environment of electronic health record technology makes the incorporation of flexible decision support extremely challenging, but over time the technology will become more adaptable and fit for this purpose. The new data fabric will enable development of algorithms that can make predictions in real time about near-term and long-term prognosis and enable evaluation of the comparative effectiveness of choices about diagnosis, prevention, and treatment.

Ultimately, the review of Alba et al3 sheds needed light on how far we have to go in providing clinicians, administrators, and patients with useful predictive models. Such models should be gathered across generalizable populations to inform important decisions and may need to be updated over time to account for periodic ascertainment of follow-up information and, in the future, the continuous data supplied by remote monitoring. We encourage investigators seeking to create risk prediction models to work together to build models with the broadest possible applicability and to provide ample opportunities for rigorous validation.

Disclosures
For the period from 2010 to 2013, Dr Califf reports receiving research grants that partially support his salary from Amgen, Johnson & Johnson, Scios, Merck/Schering-Plough, Schering-Plough Research Institute, Novartis Pharma, Bristol-Myers Squibb Foundation, Aterovax, Bayer, Roche, Lilly, and Schering-Plough; all grants are paid to Duke University. Dr Califf also consults for Genentech, Medscape LLC/TheHeart.org, Johnson & Johnson, Scios, Kowa Research Institute, Nile, Parkview, Orexin Therapeutics, Pozen, WebMD, Bristol-Myers Squibb Foundation, AstraZeneca, Bayer-OrthoMcNeil, Bristol-Myers Squibb, Boehringer Ingelheim, Daiichi Sankyo, Gilead, GlaxoSmithKline, Li Ka Shing Knowledge Institute, Medtronic, Merck, Novartis, Sanofi-Aventis, XOMA, University of Florida, Pfizer, Roche, Servier International, DSI-Lilly, Janssen R&D, CV Sight, Regeneron and Gambro; all income from these consultancies is donated to nonprofit organizations, with most going to the clinical research fellowship fund of the Duke Clinical Research Institute. Dr Califf holds equity in Nitrox LLC, N30 Pharma, and Portola. Disclosure information for Dr Califf is also available at https://dcri.org/about-us/conflict-of-interest and at http://www.duke-health.org/physicians/robert_m_califf. Dr Pencina reports receiving consulting fees from McGill University Health Centre.

References

Key Words: Editorials ■ decision support techniques ■ heart failure ■ models, decision support ■ models, statistical
Predictive Models in Heart Failure: Who Cares?
Robert M. Califf and Michael J. Pencina

Circ Heart Fail. 2013;6:877-878
doi: 10.1161/CIRCHEARTFAILURE.113.000659
Circulation: Heart Failure is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2013 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-3289. Online ISSN: 1941-3297

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circheartfailure.ahajournals.org/content/6/5/877

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Heart Failure can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Heart Failure is online at:
http://circheartfailure.ahajournals.org//subscriptions/