The need to classify disease and predict outcomes is as old as medicine itself. Nearly 50 years ago, the advantage of applying multivariable statistics to these problems became evident. Since then, the increasing availability of databases containing often-complex clinical information from tens or even hundreds of millions of patients, combined with powerful statistical techniques and computing environments, has spawned exponential growth in efforts to create more useful, focused, and accurate prediction models. JAMA Network Open receives dozens of manuscripts weekly that present new or purportedly improved instruments intended to predict a vast array of clinical outcomes. Although we are able to accept only a small fraction of those submitted, we have, nonetheless, published nearly 2000 articles dealing with predictive models over the past 6 years.

The profusion of predictive models has been accompanied by the growing recognition of the necessity for standards to help ensure accuracy of these models. An important milestone was the publication of the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines nearly a decade ago. TRIPOD is a reporting guideline intended to enable readers to better understand the methods used in published studies but does not prescribe what actual methods should be applied. Since then, while the field has continued to advance and technology improve, many predictive models in widespread use, when critically evaluated, have been found to neither adhere to reporting standards nor perform as well as expected.

There are numerous reasons why performance of models falls short, even when efforts are made to adhere to methodologic standards. Despite the vast amounts of data that are often brought to bear, they may not be appropriate to the task, or they may have been collected and analyzed in ways that are biased. Additionally, that some models fall short may simply reflect the inherent difficulty of predicting relatively uncommon events that occur as a result of complex biological processes occurring within complex clinical environments. Moreover, clinical settings are highly variable, and predictive models typically perform worse outside of the environments in which they were developed. A comprehensive discussion of these issues is beyond the scope of this article, but as physicist Neils Bohr once remarked, “it is very difficult to predict—especially the future.”

Although problems with accuracy are well documented, hundreds of predictive models are in regular use in clinical practice and are frequently the basis for critically important decisions. Many such models have been widely adopted without subsequent efforts to confirm that they actually continue to perform as expected. That is not to say that such models are without utility, because even a suboptimal model may perform better than an unaided clinician. Nevertheless, we believe that a fresh examination of selected, well-established predictive models is warranted if not previously done. JAMA Network Open has published articles addressing prediction of relatively common clinical complications, such as recurrent gastrointestinal bleeding. We think there remains considerable opportunity for research in this vein. In particular, we seek studies that examine current performance of commonly applied clinical prediction rules. We are particularly interested in studies using data from a variety of settings and databases as well as studies that simultaneously assess multiple models addressing the same or similar outcomes.

We also remain interested in the derivation of new models that address a clear clinical need. They should utilize data that are commonly collected as part of routine care, or in principle can be readily extracted from electronic health records. We generally require that prediction models be validated with at least 1 other dataset distinct from the development dataset. In practice, this means data from different health systems or different publicly available or commercial datasets. We note...
that internal validation techniques, such as split samples, hold-out, k-fold, and others, are not designed to overcome the intrinsic differences between data sources and, therefore, are not suited to quantifying performance externally. While the population to which the models apply should be described explicitly, ideally any such models should be applicable to patients from the wide range of races, ethnicities, and backgrounds commonly encountered in clinic practice. Most importantly, we are interested in examples of models that have been evaluated in clinical settings, assessing their feasibility and potential clinical benefit. This includes studies with negative as well as positive outcomes.

Please see the journal’s Instructions for Authors for information on manuscript preparation and submission. This is not a time-limited call for studies on this topic.

ARTICLE INFORMATION
Published: April 12, 2024. doi:10.1001/jamanetworkopen.2024.9640
Open Access: This is an open access article distributed under the terms of the CC-BY License. © 2024 Fihn SD et al. JAMA Network Open.

Corresponding Author: Stephan D. Fihn, MD, MPH, Department of Medicine, University of Washington, 325 Ninth Ave, Box 359780, Seattle, WA 98104 (sfihn@uw.edu).

Author Affiliations: Department of Medicine, University of Washington, Seattle (Fihn, Rivara); Deputy Editor, JAMA Network Open (Fihn); Epidemiology, Rutgers The State University of New Jersey, New Brunswick (Berlin); Statistical Editor, JAMA Network Open (Berlin, Haneuse); Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts (Haneuse); Editor, JAMA Network Open (Rivara).

Conflict of Interest Disclosures: Dr Berlin reported receiving consulting fees from Kenvue related to acetaminophen outside the submitted work. No other disclosures were reported.

REFERENCES