the basis of the type of health insurance and coverage provided, and details can be found in the online supplement of the authors' article. Patients with commercial insurance paid lower out-of-pocket costs than those with public coverage (commercial, U.S. $123 to U.S. $173/mo; Medicare Advantage, U.S. $434/mo). This would mean that patients would have to pay between U.S. $1,476 and U.S. $5,208 per year just to afford 1 year of prescriptions for the treatment of IPF. These out-of-pocket costs are staggering and possibly unaffordable for patients, especially when added to the costs of their other medications for the average of four comorbidities they may suffer from. Considering that, according to the U.S. Census Bureau, the median annual income for a family in the United States in 2019 was U.S. $68,703 (9), the out-of-pocket cost of antifibrotics could be between 2.1% and 7.6% of the total annual gross income of the household. Median earnings were lower for women than for men, which may contribute to the sex difference in medication initiation identified in this study.

This study has shown that antifibrotic uptake remains low in the United States and that discontinuation of treatment is high for those who do start medications. This may be in large part due to the high out-of-pocket cost for patients, but it is likely that other barriers and discrepancies exist but were not captured by this analysis. As more clinical trials looking at novel IPF treatments are combining drugs and looking at additive benefits of different medications, it will become even more difficult for patients who would benefit from those drugs to afford them. Exorbitant costs should not be a barrier toward a standard of care in a developed country with state-of-the-art medical advances and therapies. Policy changes to control the prohibitive costs of those medications are needed to ensure affordability for all who need them. Further barriers to access to care such as sex, race and ethnicity, or socioeconomic status also need to be identified and addressed in future studies to ensure appropriate care for all. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1 King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, *et al*.; ASCEND Study Group. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2083–2092.

2 Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, *et al*.; INPULSIS Trial Investigators. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2071–2082.

3 Ley B, Swigris J, Day BM, Stauffer JL, Raimundo K, Chou W, *et al*. Pirfenidone reduces respiratory-related hospitalizations in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2017; 196:756–761.

4 Albera C, Costabel U, Fagan EA, Glassberg MK, Gorina E, Lancaster L, *et al*. Efficacy of pirfenidone in patients with idiopathic pulmonary fibrosis with more preserved lung function. *Eur Respir J* 2016;48:843–851.

5 Ryerson CJ, Kolb M, Richeldi L, Lee J, Wachtlin D, Stowasser S, *et al*. Effects of nintedanib in patients with idiopathic pulmonary fibrosis by GAP stage. *ERJ Open Res* 2019;5:00127-2018.

6 Salisbury ML, Conoscenti CS, Culver DA, Yow E, Neely ML, Bender S, *et al*.; IPF-PRO Registry principal investigators as follows. Antifibrotic drug use in patients with idiopathic pulmonary fibrosis: data from the IPF-PRO registry. *Ann Am Thorac Soc* 2020;17:1413–1423.

7 Holtze CH, Freiheit EA, Limb SL, Stauffer JL, Raimundo K, Pan WT, *et al*. Patient and site characteristics associated with pirfenidone and nintedanib use in the United States; an analysis of idiopathic pulmonary fibrosis patients enrolled in the Pulmonary Fibrosis Foundation Patient Registry. *Respir Res* 2020;21:48.

8 Dempsey TM, Payne S, Sangaralingham L, Yao X, Shah ND, Limper AH. Adoption of the antifibrotic medications pirfenidone and nintedanib for patients with idiopathic pulmonary fibrosis. *Ann Am Thorac Soc* 2021; 18:1121–1128.

9 Semega J, Kollar M, Shrider EA, Creamer J. US Census Bureau. Income and poverty in the United States: 2019. Suitland, MD: U.S. Census Bureau; 2019 [accessed 2021 Feb 04]. Available from: https://www.census.gov/library/publications/2020/demo/p60-270.html.

A Bold First Toe into the Uncharted Waters of Evaluating Proprietary Clinical Prediction Models

Gary E. Weissman, M.D., M.S.H.P.

Palliative and Advanced Illness Research Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania
ORCID ID: 0000-0001-9588-3819 (G.E.W.).

Supported by National Institutes of Health grant K23HL141639.

DOI: 10.1513/AnnalsATS.202103-332ED

Would a clinician prescribe a new medication in the absence of any data about its efficacy or safety? Of course not. Regulatory authorities like the Food and Drug Administration (FDA) and good clinical judgment would prevent such a blunder. Then why would a health system deploy a clinical prediction model, designed to inform high-stakes decisions for patients at risk for critical

illness, without any evidence of efficacy or safety?

Although the FDA's regulatory strategy for clinical prediction models continues to mature and expand to include guidance around equity, transparency, and safety, significant gaps and uncertainties in oversight remain (1). For example, there are currently no federal regulatory standards for predictive clinical decision support (CDS) systems developed locally by hospitals (2). Those developed by private-sector companies for sale on the market may, in some cases, require FDA approval if they meet certain criteria (3). However, some of these criteria remain vague, and models released before these criteria were published have an uncertain fate.

The Epic Deterioration Index (EDI) is one such CDS system that may meet criteria for FDA regulation as a medical device and is reportedly in use in "hundreds of hospitals in the United States" (4). The EDI is a commercially available predictive CDS built by EPIC systems to identify patients at risk of clinical deterioration, was developed prior to the coronavirus disease (COVID-19) pandemic, and uses predictor variables such as patient age (but not race or sex), vital signs, nursing assessments, and laboratory values. However, the EDI is neither approved by the FDA nor had its performance, safety, or other important characteristics been reported in any peer-reviewed journal until now.

In this issue of *AnnalsATS*, Singh and colleagues (pp. 1129–1137) provided a public service by performing the first published evaluation of the EDI (5). Notably, none of the authors are affiliated with the FDA, and none disclosed any relationship to EPIC. The authors released a preprint of this study almost 1 year prior to this publication, thereby allowing substantial time for public comment and review (6). They studied the EDI's ability to predict a composite outcome of transfer to the intensive care unit, need for mechanical ventilation, or in-hospital death among ward patients with COVID-19 admitted to the University of Michigan's health system during the initial months of the COVID-19 pandemic. This is a particularly important population in which to study the EDI because the pandemic caused significant strain on many hospital wards, which may impair important care processes (7). Thus, under such strain, clinicians may rely more heavily on CDS systems, a scenario in which their efficacy, safety, and fairness become increasingly important.

This paper has several strengths that offer useful information to hospitals trying to decide if and how the EDI might be deployed. First, the authors found that among 392 patients who met inclusion criteria for the study, the area under the curve of the receiver operating characteristic was 0.79 (95% confidence interval, 0.74–0.84). In plain English, this means that if two randomly selected patients, one who did not experience the outcome and one who did, were compared with each other, the model would appropriately predict a higher risk for the latter patient 79% of the time. Figure 2 in Singh and colleagues article offers further insights into the lead time during which clinicians might respond to an alert based on the EDI's predictions. This information permits an assessment of whether or not there is sufficient time, in this case, a median of 24 hours, to respond to an alert that may vary by hospital depending on available resources.

Second, the authors identified clinically relevant classification thresholds corresponding to actual bedside care decisions that the EDI might inform. This is an insightful framing because many evaluations of clinical prediction models lack specific use cases, which precludes a necessary and pragmatic assessment. For example, at the high-risk threshold of an EDI score of 68.8, the positive predictive value was 75%, much higher than that for many early warning scores and with a very efficient number needed to evaluate of 1.4. However, at this threshold, the model only identified 39% of patients with the composite outcome.

Third, the authors provide some insight into the potential harms of the EDI while noting the disproportionate effects of COVID-19 on Black people. CDS systems such as the EDI risk reinforcing existing inequities as they focus resources on a patient in need, which may divert resources from other patients on the same ward (8). Thus, algorithmic equity—equivalent model performance across demographic subgroups—requires evaluation. No differences in the area under the curve of the receiver operating characteristic were detected between patients of different ages, genders, or races. However, the study may have been underpowered to detect such differences.

Fourth, Singh and colleagues chose to evaluate the model against a very reasonable and potentially actionable outcome to capture clinical deterioration. An early alert from the EDI might prompt expedited evaluation and attention for a patient in need. An inherent limitation to this choice, though through no fault of the authors, is that EPIC has never revealed the predicted outcome used to train the EDI in the first place. Thus, this evaluation is therefore limited in inferences that might be drawn about the "true" performance of the EDI model. At the same time, the authors' evaluation is pragmatic and appropriate and highlights the bizarre practice of selling and deploying clinical prediction models without explaining or understanding them.

The study should be interpreted in light of several additional limitations. First, Singh and colleagues reported the *in silico* performance of the model but not its direct effects on clinician decision-making or patient outcomes. The latter two outcomes would be best evaluated using a prospective randomized design that is outside the scope of this study but necessary for understanding how the EDI affects patient care. Second, the authors observed large fluctuations in the EDI every 15 minutes, as it is calculated when deployed. However, the authors reported performance measures using aggregations of the EDI at the hospitalization level. Although this practice is not uncommon in the reporting of clinical prediction models, it likely overestimates the true performance of the model and provides a less than real-world evaluation of how the model is used in practice. However, in the sensitivity analysis reported using a prediction-level evaluation, the positive predictive values were more modest and ranged from 5.5% to 24% over different time horizons.

We still don't know enough to evaluate the claim in the title of EPIC's news item on its own web page, "Artificial Intelligence Triggers Fast, Lifesaving Care." But to Singh and colleagues, a debt of gratitude is owed by the FDA, EPIC, "hundreds of hospitals," and the wider community of researchers and data scientists working to advance the field of clinical prediction models. Hospital leaders

are currently being faced with a barrage of incentives to roll out new predictive CDS systems. At the same time, hospitals that wouldn't approve of their clinicians prescribing new medications with no data behind them shouldn't themselves take up the same practice by deploying unvalidated clinical prediction models. If regulatory authorities don't step in, hospitals and independent researchers like Singh and colleagues will have to keep diving in to pick up the slack. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1 U.S. Food and Drug Administration. Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan. 2021 [accessed 2021 Mar 1]. Available from: https://www.fda.gov/media/145022/download.
2 Weissman GE. FDA regulation of predictive clinical decision-support tools: what does it mean for hospitals? *J Hosp Med* [online ahead of print] 19 Aug 2020; DOI: 10.12788/jhm.3450.
3 U.S. Food and Drug Administration. Clinical Decision Support Software: Draft Guidance for Industry and Food and Drug Administration Staff. Silver Spring, MD: FDA; 2019 [accessed 2021 Mar 1]. Available from: https://www.fda.gov/media/109618/download.
4 EPIC. Artificial intelligence triggers fast, lifesaving care for COVID-19 patients. 2020 [accessed 2021 Mar 1]. Available from: https://www.epic.com/epic/post/artificial-intelligence-epic-triggers-fast-lifesaving-care-covid-19-patients.
5 Singh K, Valley TS, Tang S, Li BY, Kamran F, Sjoding MW, *et al.* Evaluating a widely implemented proprietary deterioration index model among hospitalized patients with COVID-19. *Ann Am Thorac Soc* 2021;18:1129–1137.
6 Singh K, Valley TS, Tang S, Li BY, Kamran F, Sjoding MW, *et al.* Evaluating a widely implemented proprietary deterioration Index model among hospitalized COVID-19 patients [preprint]. *medRxiv*; 2020 [accessed 2020 Jun 20]. Available from: https://www.medrxiv.org/content/10.1101/2020.04.24.20079012v2.
7 Kohn R, Harhay MO, Bayes B, Mikkelsen ME, Ratcliffe SJ, Halpern SD, *et al.* Ward capacity strain: a novel predictor of 30-day hospital readmissions. *J Gen Intern Med* 2018;33:1851–1853.
8 Volchenboum SL, Mayampurath A, Göksu-Gürsoy G, Edelson DP, Howell MD, Churpek MM. Association between in-hospital critical illness events and outcomes in patients on the same ward. *JAMA* 2016;316:2674–2675.

Historic Abuses, Present Disparities, and Systemic Racism: Threats to Surrogate Decision-making for Critical Care Research Enrollment

Dustin C. Krutsinger, M.D., M.S.C.E.[1], Katherine R. Courtright, M.D., M.S.H.P.[2,3,4], and Paul A. Estabrooks, Ph.D.[5,6]

[1]Division of Pulmonary, Critical Care, and Sleep Medicine, [5]Department of Health Promotion; and [6]Center for Reducing Health Disparities, University of Nebraska Medical Center, Omaha, Nebraska; and [2]Palliative and Advanced Illness Research (PAIR) Center, [3]Department of Biostatistics, Epidemiology, and Informatics, and [4]Pulmonary, Allergy, and Critical Care Division, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

ORCID ID: 0000-0001-7362-7304 (D.S.K.).

DOI: 10.1513/AnnalsATS.202103-386ED

Anyone who has attempted to recruit critically ill patients into clinical trials recognizes the challenges that lie therein. Critically ill patients often lack decisional capacity and must rely on surrogate decision-makers (SDMs) to make both clinical and research enrollment decisions (1). The SDM role is both cognitively and emotionally burdensome (2, 3). Furthermore, it is frequently performed by a close family member who is already under the tremendous stress inherent in having a loved one in the intensive care unit (ICU). Therefore, it may be unsurprising that many SDMs suffer long-term psychological morbidity, including anxiety, depression, and symptoms of post-traumatic stress disorder (4). These effects may be exacerbated by being asked to consider enrollment into research (5). The reliance on SDMs for enrollment decisions may, in part, explain the low enrollment rates of critical care trials (6). To improve enrollment rates and reduce the burden on SDMs, an improved understanding of SDMs' decision-making processes surrounding clinical trial enrollment is imperative.

A previous study in this area identified three phases in SDMs' enrollment decision-making process: *1*) being approached, *2*) reflecting on participation, and *3*) making a decision (7). During each phase, SDMs reported factors related to decisions to move from one stage to the next. Although these findings provided some context for understanding SDM experiences and decision-making processes, the study was