



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## Original Research Article

# Development and External Validation of a Delirium Prediction Model for Hospitalized Patients With Coronavirus Disease 2019



Victor M. Castro, M.S., Chana A. Sacks, M.D., Roy H. Perlis, M.D., Thomas H. McCoy, M.D.

**Background:** The coronavirus disease 2019 pandemic has placed unprecedented stress on health systems and has been associated with elevated risk for delirium. The convergence of pandemic resource limitation and clinical demand associated with delirium requires careful risk stratification for targeted prevention efforts. **Objectives:** To develop an incident delirium predictive model among coronavirus disease 2019 patients. **Methods:** We applied supervised machine learning to electronic health record data for inpatients with coronavirus disease 2019 at three hospitals to build an incident delirium diagnosis prediction model. We validated this model in three different hospitals. Both hospital cohorts included academic and community settings. **Results:** Among 2907 patients across 6 hospitals, 488 (16.8%) developed delirium. Applying the predictive model in the external validation cohort of 755 patients, the c-index was 0.75

(0.71–0.79) and the lift in the top quintile was 2.1. At a sensitivity of 80%, the specificity was 56%, negative predictive value 92%, and positive predictive value 30%. Equivalent model performance was observed in subsamples stratified by age, sex, race, need for critical care and care at community vs. academic hospitals. **Conclusion:** Machine learning applied to electronic health records available at the time of inpatient admission can be used to risk-stratify patients with coronavirus disease 2019 for incident delirium. Delirium is common among patients with coronavirus disease 2019, and resource constraints during a pandemic demand careful attention to the optimal application of predictive models.

(Journal of the Academy of Consultation-Liaison Psychiatry 2021; 62:298–308)

**Key words:** delirium, cohort study, COVID-19, machine learning, electronic health records, predictive modeling, crisis standard of care.

## INTRODUCTION

The neuropsychiatric consequences of coronavirus disease 2019 (COVID-19) are increasingly evident and include a consequential incidence of delirium.<sup>1–8</sup> Delirium is a heterogeneous neuropsychiatric syndrome characterized by acute changes in cognition and awareness, leading to fluctuations in attention, memory, or consciousness.<sup>9–11</sup> Delirium, regardless of underlying illness, is associated with multiple serious adverse clinical and functional outcomes, including increased rates of medical comorbidity, longer hospital stays, higher health care costs, and higher risk of

postdischarge mortality.<sup>12–24</sup> Although treatment of delirium is symptomatic, multicomponent prevention is possible.<sup>25–30</sup> Given the consequence and potential preventability of delirium, recognition is essential;

Received August 17, 2020; revised November 27, 2020; accepted December 9, 2020. From the Center for Quantitative Health (V.M.C., R.H.P., T.H.M.), Massachusetts General Hospital, Boston, MA; and Department of Medicine (C.A.S.), Massachusetts General Hospital, Boston, MA. Send correspondence and reprint requests to Thomas H. McCoy, Jr, MD, Simches Research Building Massachusetts General Hospital 185 Cambridge St, 6th Floor Boston, MA 02114; e-mail: [thmccoy@partners.org](mailto:thmccoy@partners.org)

© 2020 Academy of Consultation-Liaison Psychiatry. Published by Elsevier Inc. All rights reserved.

however, in routine care, delirium is underrecognized.<sup>31–34</sup>

Systematic screening is an important approach to recognition of prevalent delirium, whereas risk stratification for incident delirium is a key component of a targeted delirium prevention program.<sup>35–37</sup> An electronic health record (EHR)–based approach is one means of both retrospectively studying delirium and, through clinical prediction, routing prevention efforts toward those who are vulnerable to delirium.<sup>38–41</sup> Although these efforts are hampered by variability in delirium occurrence rates across patient and provider characteristics,<sup>31,42–44</sup> delirium can be studied in the EHR to provide insights into both epidemiology and biology.<sup>38,45–51</sup>

COVID-19–associated delirium recapitulates the challenges of delirium in the context of any other underlying illness—including a more challenging posthospital course, under-recognition, and treatment uncertainty.<sup>52–55</sup> Pandemic circumstances compound the challenges inherent in routine delirium management and prevention. The convergence of increased difficulty of implementing a multicomponent prevention program in the context of respiratory illness isolation precautions juxtaposed with the increased need for prevention in the setting of increase in demand for hospital beds creates a confluence of need for precision targeting of delirium prevention efforts. Recognizing both the clinical novelty of COVID-19–associated delirium and the unique dynamism of prevention efforts in the context of a pandemic, we sought to extend previous work on EHR-based delirium prediction and present an EHR-based machine learning approach to predict COVID-19–associated delirium.

## MATERIALS AND METHODS

### Overview of Approach

The aim of this study is to develop a statistical model that can be applied to EHR data to predict the probability of a hospitalized patient with COVID-19 developing delirium. To achieve this aim, we extracted health records data from the EHR of 3 hospitals and applied a basic form of machine learning that both selects the most important variables and defines a regression model through which those variables can be mathematically combined to produce a risk score. After developing the model by identify the most predictive variables and the weights associated with each, we applied it to health records data drawn from 3 separate

hospitals and evaluated how well it identified delirium in this second “testing” group. That evaluation consisted of 2 separate notions of performance - the ability to rank patients as higher and lower risk and to predict the risk in a given patient accurately.

### Subject Cohort

The study used 2 cohorts—one for model development and the second for model evaluation—each composed of 1 academic medical center (AMC) and 2 community hospitals. These cohorts were assigned in a cluster randomized fashion to ensure each cohort had 1 of the 2 AMCs. Although drawn from different hospitals, data handling and cohort definitions were the same in both cohorts. Both cohorts included all adults hospitalized between March 1, 2020 and May 31, 2020 at the 6 hospitals with polymerase chain reaction-confirmed severe acute respiratory syndrome coronavirus within 5 days of admission. For all individuals in the study, all clinical data, including vital signs, diagnostic codes, laboratory results, and medications, up through and including the hospital encounter, were extracted from the hospitals’ EHRs and used to generate a study data mart.<sup>56</sup> Registration and demographic data were also extracted including body mass index at the time of hospitalization, lifetime smoking status, and zip code as area deprivation index.<sup>57,58</sup>

The study protocol was approved by the Partners Health Care Human Research Committee. No participant contact was required in this study which relied on secondary use of data produced by routine clinical care, allowing waiver of requirement for informed consent as detailed by 45 CFR 46.116.

### Data Handling and Encoding

For laboratory tests and vital signs, we considered the first instance of each feature that occurred during the admission, inclusive of any time in the emergency department before admission. As an aggregate measure of comorbidity, Charlson comorbidity index was calculated using diagnostic codes from all available records before hospitalization.<sup>59</sup> To represent differential risk associated with specific co-occurring medical conditions, prior diagnostic codes were collapsed to the second level of the Healthcare Utilization Project Clinical Classification Software hierarchy.<sup>60</sup> As with our prior approach to encoding clinical history, the log-transformed count of the total number of prior

## EHR Prediction of COVID-19 Delirium

diagnostic codes within each category was added as a possible feature.<sup>61,62</sup>

As a wide range of laboratory tests were available, only those tests available in 80% of individuals were included in subsequent analysis as continuous measures. This limited subset of continuous values was constrained at the 99th percentile to diminish the impact of extreme values but otherwise used in native units without transformation. In addition to continuous laboratory values, laboratory-specific reference flags of normality or abnormality were included as logical predictors.

Vital signs, body mass index, smoking status, age, race, and ethnicity were all taken from predefined structured fields within the health record and those features which are numerical were analyzed in their clinical units (e.g., kg/m<sup>2</sup> for body mass index). Consistent with prior work,<sup>63,64</sup> medications were encoded as at the UMLS RxNorm ingredient level as the log-transformed count of ingredient orders and prescriptions over the 30 days before admission.<sup>65,66</sup> Recognizing the particular importance of dementia as a risk factor for delirium in hospitalized patients, a previously published specific dementia feature was calculated for descriptive and secondary analysis.<sup>67</sup> Finally, the primary outcome of delirium at any point during the studied COVID-19 hospitalization was ascertained based on a previously published EHR definition, ported using the Centers for Medicare & Medicaid Services General Equivalence Mappings.<sup>39,68</sup> Because delirium can be under reported in coded data, an alternative delirium definition which includes natural language processing of provider authored notes to identify patients who were delirious during their hospital course was used as a secondary outcome definition for sensitivity analysis on the main result.<sup>31,38</sup>

### Study Design and Analysis

Solely for purposes of description, the 2 cohorts of patients were pooled. This pooled cohort was characterized using appropriate summary statistics with differences between the delirium group and nondelirium group tested through univariate comparison (i.e., chi square test for binary variables, Student's *t*-test for continuous measures). For the primary predictive analysis, 2 distinct cohorts were used (one for model development or training, the other for independent validation). Each cohort included a single academic

medical center and 2 community medical centers. All model development occurred in the training cohort. Once the final model was specified and fitted, that model's performance was characterized in the independent testing cohort.

To identify a compact model that nevertheless considered the full range of available predictor features, we used L1-penalized regression—also known as the least absolute shrinkage and selection operator (Lasso).<sup>69</sup> This form of linear regression allows selection of the most relevant features for clinical prediction from among a large set of potential predictors and ultimately produce a simple linear model that can be readily reimplemented from tabled coefficients. Model fitting used all individuals at the training sites through median imputation of missing data. In the testing sites, participants with complete data were considered. The accuracy of the final model was characterized using conventional quantile-by-quantile comparison of predicted and observed outcome rate using the Hosmer-Lemeshow method,<sup>70,71</sup> area under the receiver operating characteristic curve discrimination,<sup>72–74</sup> evaluation of cumulative probability distributions,<sup>75,76</sup> decision curve analysis,<sup>77–79</sup> and analysis of the classification confusion matrix.<sup>80</sup> Optimization of Youden's index was used for reproducible selection of high risk thresholds that balance sensitivity and specificity except where otherwise noted.<sup>81</sup>

Given the novelty of COVID-19, wide range of technical capacity across medical centers, and the wide range of area under the receiver operating characteristic curves (AUCs) found in delirium prediction studies,<sup>47</sup> we elected to take 2 approaches to developing references for comparisons: raw age and a logistic regression including only age and dementia history. First, and primarily, we used raw age in years on admission—older age is reliably identified as a strong risk factor for delirium<sup>82–84</sup>—as a trivial risk ranking for comparison. Use of age for risk stratification is within any clinician's instantaneous ability to apply on the fly in the ward as it requires only knowledge of birthdays. Although simple, age has merit as age offers much greater granularity of risk stratification than systems based on one, or a small number of, categorical traits (e.g., history of a given diagnosis) without requiring any computational work to implement. As a secondary comparator, we developed a logistic regression model with 2 predictors: prior diagnosis of dementia and age. This 2 feature regression model has the advantage of

**TABLE 1. Sociodemographic characteristics of the cohorts used for testing and training as well as a pooled cohort merging the testing and training hospitals for descriptive purposes**

	Pooled (N = 2907)	Test (N = 755)	Train (N = 2152)
Hospital type			
Academic medical centers	1557 (53.6%)	406 (53.8%)	1151 (53.5%)
Community hospitals	1350 (46.4%)	349 (46.2%)	1001 (46.5%)
Age group			
<30	126 (4.3%)	19 (2.5%)	107 (5.0%)
30–39	261 (9.0%)	58 (7.7%)	203 (9.4%)
40–49	324 (11.1%)	53 (7.0%)	271 (12.6%)
50–59	492 (16.9%)	127 (16.8%)	365 (17.0%)
60–69	560 (19.3%)	187 (24.8%)	373 (17.3%)
70–79	510 (17.5%)	135 (17.9%)	375 (17.4%)
80+	634 (21.8%)	176 (23.3%)	458 (21.3%)
Male Sex	1536 (52.8%)	365 (48.3%)	1171 (54.4%)
Race			
Black	496 (17.1%)	256 (33.9%)	240 (11.2%)
Other	481 (16.5%)	209 (27.7%)	699 (32.5%)
White	1503 (51.7%)	290 (38.4%)	1213 (56.4%)
Charlson comorbidity index	2.55 (3.35)	2.67 (3.56)	2.5 (3.27)
Prior Dementia Diagnosis	328 (11.3%)	75 (9.9%)	253 (11.8%)
ICU Care	655 (22.5%)	189 (25.0%)	466 (21.7%)
Delirium	488 (16.8%)	143 (18.9%)	345 (16.0%)
Alternative Delirium	516 (17.8%)	366 (48.3%)	150 (7.0%)
Death	462 (15.9%)	143 (18.9%)	319 (14.8%)

calibrated predictions over raw age rank but requires a smaller number of feature to be engineered than the full Lasso model does, beyond number of features, the implementation effort is equivalent. All analysis used R, version 4.

## RESULTS

The training cohort included 2152 patients; of whom, 1151 were treated at the academic medical center (Table 1). The testing cohort included 755 patients; of whom, 406 were treated at the academic medical center. The training cohort included 345 cases of delirium for a 16.0 case rate. The testing cohort included 143 cases of delirium for an 18.9 case rate.

Pooling the 2 otherwise separate cohorts for descriptive purposes only (Table 2) yields an average age of 62.9 years in a cohort which is 52.8% men (n = 1536) and 51.7% white (n = 1503). In the pooled cohort, delirium cases had an average age of 71.5 years, whereas those without a diagnosis of delirium had an

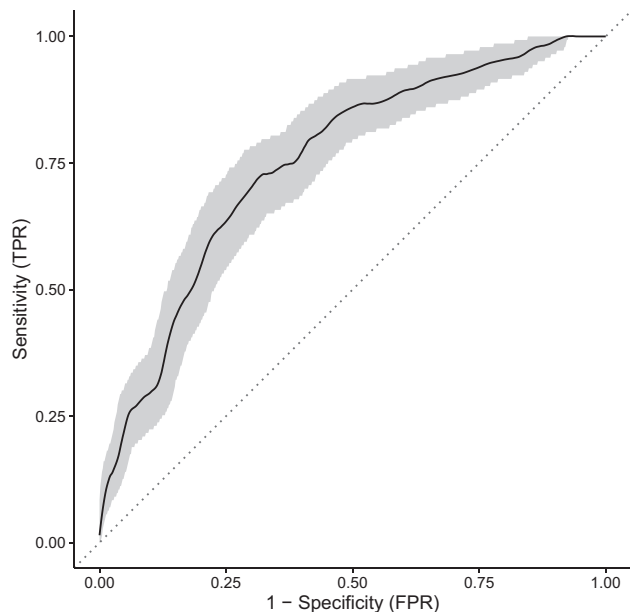
**TABLE 2. Sociodemographic characteristics contrasted by the presence or absence of a delirium diagnosis pooling data from both the testing and the training hospitals**

	Delirium = (N = 488)	No delirium (N = 2419)
Hospital type		
Academic medical centers	268 (54.9%)	1289 (53.3%)
Community hospitals	220 (45.1%)	1130 (46.7%)
Age group		
<30	9 (1.8%)	117 (4.8%)
30–39	12 (2.5%)	249 (10.3%)
40–49	21 (4.3%)	303 (12.5%)
50–59	61 (12.5%)	431 (17.8%)
60–69	95 (19.5%)	465 (19.2%)
70–79	116 (23.8%)	394 (16.3%)
80+	174 (35.7%)	460 (19.0%)
Male sex	256 (52.5%)	1280 (52.9%)
Race		
Black	88 (18.0%)	408 (16.9%)
Other	124 (25.4%)	784 (32.4%)
White	276 (56.6%)	1227 (50.7%)
Charlson comorbidity index	3.58 (3.96)	2.346 (3.17)
Prior Dementia Diagnosis	138 (28.3%)	190 (7.9%)
ICU Care	173 (35.5%)	482 (19.9%)
Death	132 (27.0%)	330 (13.6%)

average age of 61.1 years (Supplemental Figure 1). In the pooled cohort, 35.5% (n = 173) of delirium cases required care in an intensive care unit (ICU), whereas only 19.9% (n = 482) of cases without a diagnosis of delirium required ICU care. Similarly, 28.3% (n = 138) of the pooled cohort delirium cases had a prior diagnosis of dementia, whereas only 7.9% (n = 190) of those who did not develop delirium had a diagnosis of dementia.

We used L1-penalized regression to train a delirium prediction model based on admission patient characteristics, vital signs, laboratory values, and medication and diagnostic history in the training cohort which consisted of 1 academic medical center and 2 community hospitals. Of the 783 features entered into the Lasso, 34 had nonzero coefficients. Although coefficients of a penalized model are of limited explanatory use, the final selected variables of the full model are tabled in Supplemental Table 1 and clinically notable for the inclusion of a wide range clinically plausible factors including central nervous system diagnostic codes, antibiotic orders, and age. We then applied the resulting model in the independent testing sample. This wholly independent testing sample was used to evaluate

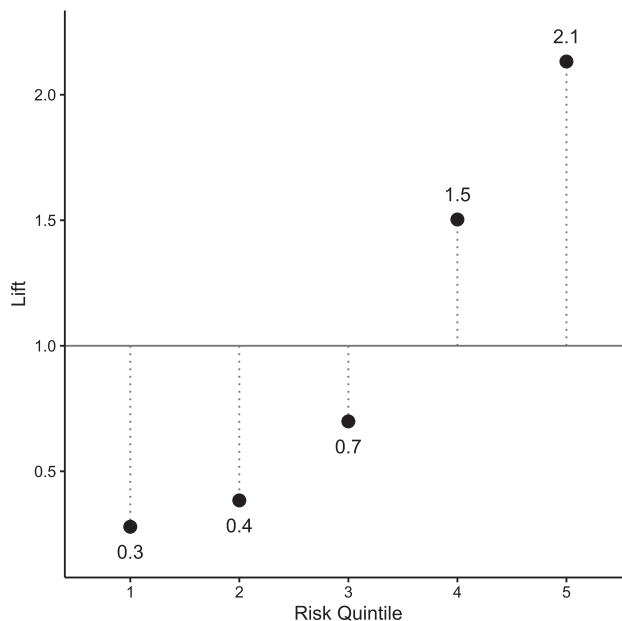
**FIGURE 1.** Receiver operating characteristic (ROC) curve for delirium prediction in the independent testing cohort.



the performance of the delirium predictive model developed in the training cohort.

In the independent test set, AUC for the delirium prediction model was 0.75 (95% confidence interval 0.71–0.79; [Figure 1](#)). The model produced a lift of 2.1 in the highest risk quintile ([Figure 2](#)) which captured 42.7% of all delirium cases with a delirium occurrence rate of 40.4%. The lowest risk quintile included only 8 delirium cases for a case rate of 5.3% and lift of 0.3. The predicted and observed prevalence were equivalent in quantile-by-quantile calibration ( $\chi^2(23) = 27.44$ ,  $P = 0.24$ ) consistent with a well-calibrated model. The optimal predicted risk cut point occurred at 0.15 and produced a sensitivity of 0.73 (0.65–0.80), specificity 0.69 (0.65–0.73), negative predictive value of 0.92 (0.89–0.94), and positive predictive value of 0.35 (0.30–0.41). A secondary approach to selection of the optimal risk cut point, maximization of the separation delirium, and nondelirium cumulative distribution functions also occurred at 0.15 ([Supplemental Figure 2](#)). Specificity of 80% occurred at a cut point of 0.12 and produced a specificity of 0.56 (0.52–0.60), negative predictive value of 0.92 (0.89–0.95), and positive predictive value of 0.30 (0.25–0.35). Sensitivity analysis suggested the observed result in the testing set was insensitive to the particular outcome definition as

**FIGURE 2.** Lift by quintile of predicted delirium risk in the independent testing cohort.



the alternative delirium definition which included natural language processing yielded an AUC of 0.74 (0.70–0.79).

To contextualize the quality of the primary predictive model developed here, we treated age on admission as a naïve risk score. Age on admission produced an AUC of 0.65 (0.60–0.70). The optimal high-risk age cut point occurred at 63 years of age. Using 63 years of age as the risk threshold produced a sensitivity of 0.76 (0.68–0.83) and specificity of 0.46 (0.42–0.50). A sensitivity of 0.80 was achieved with a risk cut point of 61 years of age which produced a specificity of 0.42 (0.38–0.46). Age naturally lends itself to ranking higher and lower-risk patients, whereas it does not have a natural calibration for risk of delirium at a given age and thus only discrimination was evaluated as context. To contextualize the benefit of additional feature engineering work, a logistic regression model using age on admission and prior diagnosis of dementia was fitted in the training set and used to predict delirium risk in the independent test set. In the independent test set, the model was calibrated to the observed case rate ( $\chi^2(23) = 23.0$ ,  $P = 0.46$ ) in quantile-by-quantile comparison and resulted in an AUC of 0.68 (0.63–0.72). The optimal high-risk age cut point by

Youden index occurred in the age and dementia-only regression model at 0.13. Using 0.13 as the risk threshold produced a sensitivity of 0.79 (0.71–0.85) and specificity of 0.46 (0.42–0.50).

In secondary analysis of the primary predictive model stratified by patient characteristic, to understand whether it was likely to perform differentially in clinical cohorts with different patient distributions, AUCs were similar across subgroups (Figure 3). Among patients who were less than 65 years of age, the AUC was 0.77 (0.69–0.85), whereas among those who were 65 years of age or older, the AUC was 0.69 (0.63–0.75; Figure 3A). Among male patients, the AUC was 0.74 (0.68–0.80), whereas among female patients, the AUC was 0.76 (0.70–0.82; Figure 3B). Among black patients, the AUC was 0.76 (0.68–0.84), among white patients, the AUC was 0.71 (0.64–0.79), and among those of any other race, the AUC was 0.80 (0.73–0.86; Figure 3C). Among patients who required management in an ICU, the AUC was 0.67 (0.59–0.76), whereas the AUC was 0.77 (0.71–0.82) among those who did not require ICU care (Figure 3D). The AUC was 0.74 (0.69–0.80) among those patients cared for at a community hospital, whereas the AUC was 0.75 (0.68–0.81) among those patients cared for at the testing set academic medical center (Figure 3E).

Given the potential for rapidly evolving circumstances in the setting of pandemic illness and possible need to implement crisis standards of care, a decision curve analysis considered the full range of risk thresholds under both opt-in (Supplemental Figure 3) and opt-out models (Supplemental Figure 4) for the full model in the independent testing set.

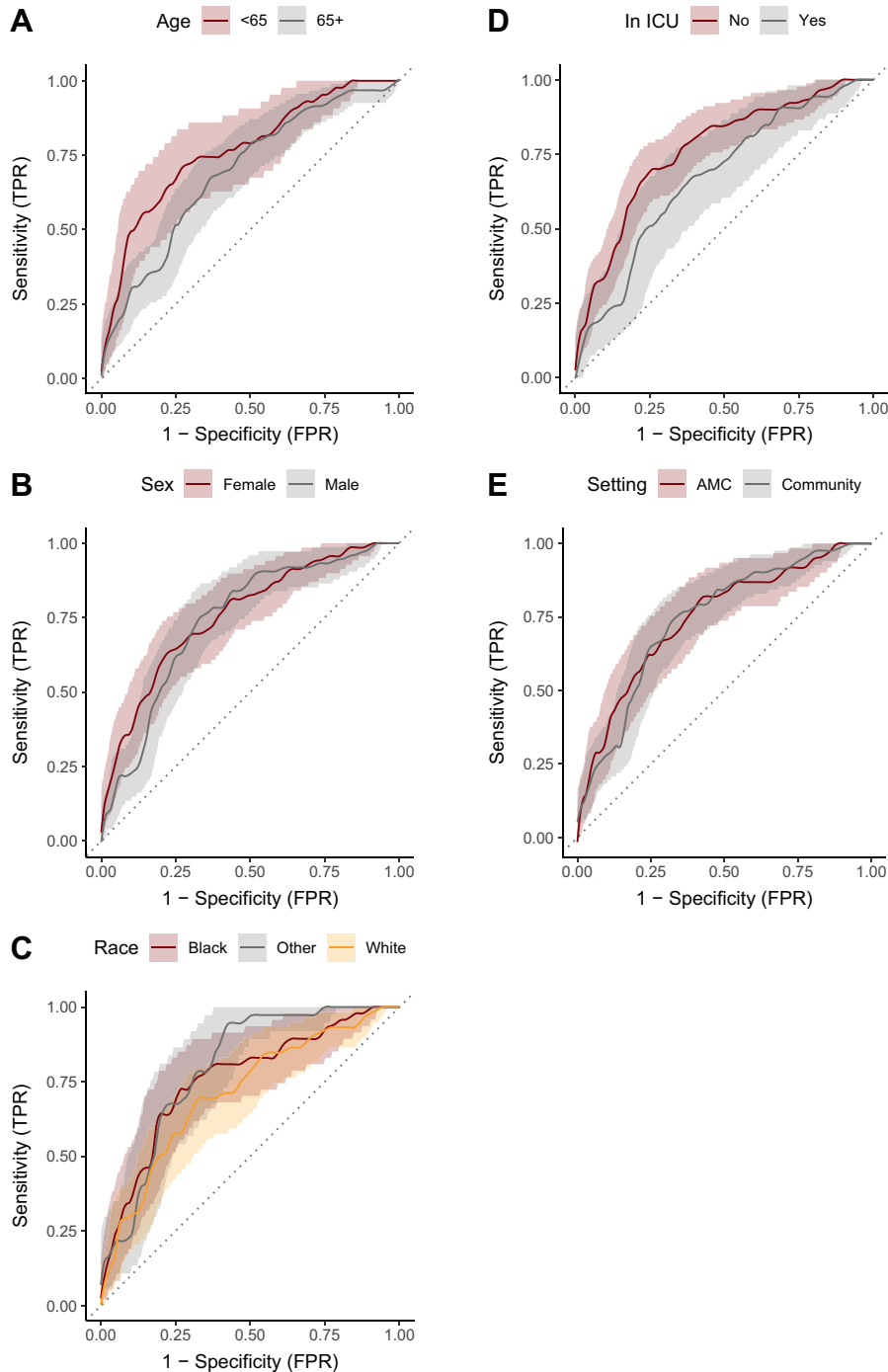
## DISCUSSION

In this study of individuals with COVID-19, we developed a risk prediction model using clinical data available at time of inpatient hospitalization which yielded promising discrimination and calibration in an independent cohort. The AUC of 0.75 reached in this independent evaluation cohort with COVID-19 falls squarely in the previously reported range (0.52 to 0.94) of delirium prediction captured in a recent systematic review of the topic.<sup>47</sup> Given the previously described risk of predictive model bias,<sup>85,86</sup> it is also important to note that model performance was consistent across patient characteristics (age, sex, and race) and clinical

context (ICU vs. ward and tertiary care vs. community hospital). The full machine learning model consistently outperformed raw age as a simple means risk ranking for comparison. However, the naïve age model's AUC of 0.65 is similarly within the band of previously reported delirium prediction models which raises the possibility that age may be of differing importance as a delirium risk factor in different underlying disease states. This possibility warrants further focused research as widely varying delirium rates would need to be considered.<sup>31,42,44</sup> Although the direct use of age as a trivial means of ranking higher and lower risk between patients comes at the expense of a prediction calibration, and thus, the use of age as a predictive reference is limited, Supplemental Figure 1 provides a clinically applicable intuition for it of potential epidemiological value to the consulting psychiatrist caring for patients with COVID-19. As a middle ground between uncalibrated direct use of raw age and the full predictive model, the logistic regression using only age and dementia history produces intermediate results with an AUC of 0.68. The full model and the limited age and dementia-only models had more similar sensitivities and their respective optimized cut points (0.73 for the full model vs 0.79 for the limited model), whereas the specificity of the full model was superior (0.69 vs 0.46).

We note multiple limitations in the present work. The data are drawn from an open system, and as such, we cannot exclude the possibility of a patient receiving medical care—before the studied COVID-19 admission—outside the studied network of hospitals. These missing data likely limit the predictive accuracy of this model which is—as a matter of applicability—limited to before, and early, hospital facts. The delirium outcome was based on health records, not a gold standard reference diagnosis or structured screening approach. The literature suggests delirium is under recognized; as such, the use of recognition through routine care as an outcome likely biases results toward the null, that is, the noncase group is likely to contain true cases that were not recognized at the time. Important work remains to be performed characterizing the true rate of delirium, as established by reference diagnosis, among patients with COVID-19 as well as establishing both the feasibility and accuracy of screening during pandemic contact minimization efforts. Finally, because the model building process used L1-penalized regression, the resulting coefficients cannot be interpreted as coefficients in a conventional

**FIGURE 3.** A: Receiver operating characteristic (ROC) curves for delirium prediction in the independent test cohort stratified by patient age. B: Sex. C: Race. D: Intensive care unit (ICU) care. E: Academic medical center vs community hospital admission.



prespecified regression model. On the other hand, the list of variables selected and operation of the model are wholly inspectable such that many of the objections to black box artificial intelligence do not apply. Beyond

limitations, the present approach may not maximize predictive accuracy as it is possible that a focused ICU prediction effort would be more accurate than the pooled equivalency found here.<sup>48,87,88</sup>



The present work has important practical strengths. In the setting of a pandemic, the value of zero-contact EHR-based risk stratification increases. Contactless evaluation eliminates the need for patient exposure and with it any associated infection exposure risk or consumption of scarce personal protective equipment. Identification of low-risk patients has the potential to further minimize exposure risk. At the other end of the risk spectrum, identification of high-risk individuals targets both staff exposure and associated personal protective equipment utilization to those patients with the greatest potential to benefit from multicomponent delirium prevention efforts (efforts which may in turn reduce length of stay and free scarce acute care resources in the setting of pandemic surge). Concretely, selecting a high-risk cutoff by maximization of Youden's index and applying this model in the test cohort would have allowed 72.7% of cases to be intervened on through contact with only 38.9% of the cohort. As the selection of risk threshold is independent of risk scoring, application of a model of this kind could be modified in real time to track shifting resource scarcity—from scarce acute care space which might favor broader intervention to scarce staff and protective equipment which might favor a more stringent intervention threshold.

We note in particular 2 aspects of the present approach which were motivated by the desire for simplicity, scalability, and generalizability. First, the data considered for prediction were limited to those available shortly after hospital admission. This allows early risk stratification, at the cost of missing the opportunity to retarget individuals as high risk based on data that became available over the course of a hospitalization. The practical importance of this limitation, and associated predictive accuracy penalty, has been previously noted.<sup>89</sup> Second, the model used is a simple linear model. The advantage of this approach is ease of reimplantation in the production transactional EHR (a practical requirement if value is to be realized); however, limiting to simple linear models likely sacrifices optimal predictive accuracy and thus further work on more sophisticated approaches to prediction is warranted. Although the linear model is algorithmically simple to reimplement (encode the features as described, apply the coefficients from [Supplemental Table 1](#), and sum), the number of features and lack of statistical equivalence to conventional predefined explanatory linear models limits direct clinical

interpretation of the result model. Regardless of the approach taken, careful application and validation within a new application context are critical components of predictive medicine.<sup>90,91</sup> Finally, we note that the choice to evaluate the model in independent clinical sites—instead of pooling and randomly dividing—is a strength of the present work as past experience suggests that stable calibration and accuracy across clinical sites is a more significant challenge than random partitions within a pooled group of sites or time-based cohorts.<sup>63,64,67,92</sup>

---

## CONCLUSION

Machine learning applied to EHR facts available on inpatient admission can be used to predict the subsequent emergence of delirium among patients hospitalized with COVID-19. This prediction, developed in a cohort drawn from 3 hospitals and validated in a separate cohort of 3 different hospitals, was found to be well calibrated to observed occurrence rates, discriminate better than direct use of patient age, and demonstrate clinical utility over a wide range of risk thresholds.

---

## SUPPLEMENTARY DATA

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jaclp.2020.12.005>.

*Conflicts of Interest:* Mr. Castro reports no conflict of interest. Dr. Sacks has received research funding from the Carney Family Foundation. Dr. Perlis has received consulting fees from Burrage Capital, Genomind, RID Ventures, and Takeda. He holds equity in Outermost Therapeutics and Psy Therapeutics. Dr. McCoy has received research funding from the Stanley Center at the Broad Institute, the Brain and Behavior Research Foundation, National Institute of Mental Health, National Institute of Nursing Research, National Human Genome Research Institute Home, and Telefonica Alfa.

*Funding:* This study was funded by the National Institute of Mental Health (grant numbers 1R01MH120991 and 5R01MH116270). The sponsors had no role in study design, writing of the report, or data collection, analysis, or interpretation.

# EHR Prediction of COVID-19 Delirium

*Acknowledgments: The authors would like to thank the full Partners Healthcare PRDR and Analytics*

*Enclave teams for their support in making EHR data available to study the COVID-19 pandemic.*

## References

1. Eliezer M, Hautefort C, Hamel A-L, et al: Sudden and complete Olfactory Loss function as a possible Symptom of COVID-19. *JAMA Otolaryngol Head Neck Surg* 2020; 146:674–675
2. Politi LS, Salsano E, Grimaldi M: Magnetic Resonance Imaging Alteration of the brain in a patient with coronavirus disease 2019 (COVID-19) and Anosmia. *JAMA Neurol* 2020; 77:1028–1029
3. Helms J, Kremer S, Merdji H, et al: Neurologic features in Severe SARS-CoV-2 infection. *N Engl J Med* 2020; 382:2268–2270
4. Mao L, Jin H, Wang M, et al: Neurologic Manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan, China. *JAMA Neurol* 2020; 77:683–690
5. Orsucci D, Ienco EC, Nocita G, Napolitano A, Vista M: Neurological features of COVID-19 and their treatment: a review. *Drugs Context* 2020; 9:2020-5-1
6. Sher Y, Rabkin B, Maldonado JR, Mohabir P: COVID-19-Associated hyperactive intensive care Unit delirium with Proposed pathophysiology and treatment: a case report. *Psychosomatics* 2020; 61:544–550
7. Garg RK: Spectrum of Neurological Manifestations in covid-19: a review. *Neurol India* 2020; 68:560–572
8. Soysal P, Kara O: Delirium as the first clinical presentation of the coronavirus disease 2019 in an older adult. *Psychogeriatr Off J Jpn Psychogeriatr Soc* 2020; 20:763–765
9. Maldonado JR: Delirium pathophysiology: an updated hypothesis of the etiology of acute brain failure. *Int J Geriatr Psychiatry* 2018; 33:1428–1457
10. Maldonado JR: Pathoetiological model of delirium: a comprehensive understanding of the neurobiology of delirium and an evidence-based approach to prevention and treatment. *Crit Care Clin* 2008; 24:789–856,ix
11. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Washington, D.C: American Psychiatric Publishing; 2013. DSM-5. 5 edition
12. Goldberg TE, Chen C, Wang Y, et al: Association of delirium with Long-term cognitive Decline: a meta-analysis. *JAMA Neurol* 2020; 77:1–9
13. Salluh JIF, Wang H, Schneider EB, et al: Outcome of delirium in critically ill patients: systematic review and meta-analysis. *BMJ* 2015; 350:h2538
14. Leslie DL, Inouye SK: The importance of delirium: economic and societal costs. *J Am Geriatr Soc* 2011; 59(suppl 2):S241–S243
15. Leslie DL, Marcantonio ER, Zhang Y, Leo-Summers L, Inouye SK: One-year health care costs associated with delirium in the elderly population. *Arch Intern Med* 2008; 168:27–32
16. Vasilevskis EE, Chandrasekhar R, Holtz CH, et al: The cost of ICU delirium and coma in the intensive care Unit patient. *Med Care* 2018; 56:890–897
17. Weinrebe W, Johannsdottir E, Karaman M, Füsigen I: What does delirium cost? An economic evaluation of hyperactive delirium. *Z Gerontol Geriatr* 2016; 49:52–58
18. Tropea J, LoGiudice D, Liew D, Gorelik A, Brand C: Poorer outcomes and greater healthcare costs for hospitalised older people with dementia and delirium: a retrospective cohort study. *Int J Geriatr Psychiatry* 2017; 32:539–547
19. McCusker J, Cole M, Dendukuri N, Belzile E, Primeau F: Delirium in older medical inpatients and subsequent cognitive and functional status: a prospective study. *CMAJ Can Med Assoc J J Assoc Medicale Can* 2001; 165:575–583
20. Cole MG, Primeau FJ: Prognosis of delirium in elderly hospital patients. *CMAJ Can Med Assoc J J Assoc Medicale Can* 1993; 149:41–46
21. McCusker J, Cole M, Abrahamowicz M, Primeau F, Belzile E: Delirium predicts 12-month mortality. *Arch Intern Med* 2002; 162:457–463
22. Crocker E, Beggs T, Hassan A, et al: Long-Term effects of postoperative delirium in patients Undergoing cardiac operation: a systematic review. *Ann Thorac Surg* 2016; 102:1391–1399
23. Kiely DK, Marcantonio ER, Inouye SK, et al: Persistent delirium predicts greater mortality. *J Am Geriatr Soc* 2009; 57:55–61
24. Pauley E, Lishmanov A, Schumann S, Gala GJ, van Diepen S, Katz JN: Delirium is a robust predictor of morbidity and mortality among critically ill patients treated in the cardiac intensive care unit. *Am Heart J* 2015; 170:79–86,86.e1
25. Hshieh TT, Yue J, Oh E, et al: Effectiveness of multicomponent nonpharmacological delirium interventions: a meta-analysis. *JAMA Intern Med* 2015; 175:512–520
26. Girard TD, Exline MC, Carson SS, et al: Haloperidol and Ziprasidone for treatment of delirium in critical illness. *N Engl J Med* 2018; 379:2506–2516
27. Wang Y-Y, Yue J-R, Xie D-M, et al: Effect of the Tailored, Family-Involved hospital Elder Life program on postoperative delirium and function in older adults: a randomized clinical trial. *JAMA Intern Med* 2019:e194446
28. Skelton L, Guo P: Evaluating the effects of the pharmacological and nonpharmacological interventions to manage delirium symptoms in palliative care patients: systematic review. *Curr Opin Support Palliat Care* 2019; 13:384–391
29. Barbateskovic M, Krauss SR, Collet MO, et al: Pharmacological interventions for prevention and management of delirium in intensive care patients: a systematic overview of reviews and meta-analyses. *BMJ Open* 2019; 9:e024562
30. Friedman JI, Soleimani L, McGonigle DP, Egol C, Silverstein JH: Pharmacological treatments of non-substance-withdrawal delirium: a systematic review of prospective trials. *Am J Psychiatry* 2014; 171:151–159
31. McCoy TH, Snapper L, Stern TA, Perlis RH: Under-reporting of delirium in Statewide claims data: Implications

- for clinical care and predictive modeling. *Psychosomatics* 2016; 57:480–488
32. Lange PW, Lamanna M, Watson R, Maier AB: Undiagnosed delirium is frequent and difficult to predict: results from a prevalence survey of a tertiary hospital. *J Clin Nurs* 2019; 28:2537–2542
  33. Ruangratsamee S, Assanasen J, Praditsuwan R, Srinonprasert V: Unrecognized delirium is prevalent among older patients Admitted to general medical wards and Lead to higher mortality rate. *J Med Assoc Thai Chotmaihet Thangphaet* 2016; 99:904–912
  34. Ely EW, Siegel MD, Inouye SK: Delirium in the intensive care unit: an under-recognized syndrome of organ dysfunction. *Semin Respir Crit Care Med* 2001; 22:115–126
  35. Alosaimi FD, Alghamdi A, Alsuhaibani R, et al: Validation of the Stanford Proxy test for delirium (S-PTD) among critical and noncritical patients. *J Psychosom Res* 2018; 114:8–14
  36. Wong CL, Holroyd-Leduc J, Simel DL, Straus SE: Does this patient have delirium?: value of bedside instruments. *JAMA* 2010; 304:779–786
  37. Grover S, Kate N: Assessment scales for delirium: a review. *World J Psychiatry* 2012; 2:58–70
  38. McCoy TH, Chaukos DC, Snapper LA, Hart KL, Stern TA, Perlis RH: Enhancing delirium case definitions in electronic health records using clinical free Text. *Psychosomatics* 2017; 58:113–120
  39. Bui LN, Pham VP, Shirkey BA, Swan JT: Effect of delirium motoric subtypes on administrative documentation of delirium in the surgical intensive care unit. *J Clin Monit Comput* 2017; 31:631–640
  40. Inouye SK, Leo-Summers L, Zhang Y, Bogardus ST, Leslie DL, Agostini JV: A chart-based method for identification of delirium: validation compared with interviewer ratings using the confusion assessment method. *J Am Geriatr Soc* 2005; 53:312–318
  41. Kim DH, Lee J, Kim CA, et al: Evaluation of algorithms to identify delirium in administrative claims and drug utilization database. *Pharmacoepidemiol Drug Saf* 2017; 26:945–953
  42. Ryan DJ, O'Regan NA, Caoimh RÓ, et al: Delirium in an adult acute hospital population: predictors, prevalence and detection. *BMJ Open* 2013; 3:e001772
  43. Menzenbach J, Guttenthaler V, Kirfel A, et al: Estimating patients' risk for postoperative delirium from preoperative routine data - Trial design of the PRe-Operative prediction of postoperative DELirium by appropriate SCreening (PROPDESC) study - a monocentre prospective observational trial. *Contemp Clin Trials Commun* 2020; 17:100501
  44. McCoy TH, Hart KL, Perlis RH: Characterizing and predicting rates of delirium across general hospital settings. *Gen Hosp Psychiatry* 2017; 46:1–6
  45. McCoy TH, Hart K, Pellegrini A, Perlis RH: Genome-wide association identifies a novel locus for delirium risk. *Neurobiol Aging* 2018; 68. 160.e9–160.e14
  46. Hercus C, Hudaib A-R: Delirium misdiagnosis risk in psychiatry: a machine learning-logistic regression predictive algorithm. *BMC Health Serv Res* 2020; 20:151
  47. Lindroth H, Bratzke L, Purvis S, et al: Systematic review of prediction models for delirium in the older adult inpatient. *BMJ Open* 2018; 8:e019223
  48. Lee A, Mu JL, Joynt GM, et al: Risk prediction models for delirium in the intensive care unit after cardiac surgery: a systematic review and independent external validation. *Br J Anaesth* 2017; 118:391–399
  49. Wassenaar A, Schoonhoven L, Devlin JW, et al: Delirium prediction in the intensive care unit: comparison of two delirium prediction models. *Crit Care Lond Engl* 2018; 22:114
  50. Wassenaar A, van den Boogaard M, van Achterberg T, et al: Multinational development and validation of an early prediction model for delirium in ICU patients. *Intensive Care Med* 2015; 41:1048–1056
  51. Lee S, Harland K, Mohr NM, et al: Evaluation of emergency department derived delirium prediction models using a hospital-wide cohort. *J Psychosom Res* 2019; 127:109850
  52. Kotfis K, Williams Roberson S, Wilson JE, Dabrowski W, Pun BT, Ely EW: COVID-19: ICU delirium management during SARS-CoV-2 pandemic. *Crit Care Lond Engl* 2020; 24:176
  53. O'Hanlon S, Inouye SK: Delirium: a missing piece in the COVID-19 pandemic puzzle. *Age Ageing* 2020; 49:497–498
  54. Mcloughlin BC, Miles A, Webb TE, et al: Functional and cognitive outcomes after COVID-19 delirium. *Eur Geriatr Med* 2020; 11:857–862
  55. Anmella G, Arbelo N, Fico G, et al: COVID-19 inpatients with psychiatric disorders: real-world clinical recommendations from an expert team in consultation-liaison psychiatry. *J Affect Disord* 2020; 274:1062–1067
  56. Murphy SN, Weber G, Mendis M, et al: Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc JAMIA* 2010; 17:124–130
  57. Knighton AJ, Savitz L, Belnap T, Stephenson B, VanDerslice J: Introduction of an area deprivation index measuring patient Socioeconomic status in an Integrated health system: Implications for population health. *EGEMS Wash DC* 2016; 4:1238
  58. Singh GK: Area deprivation and widening inequalities in US mortality, 1969-1998. *Am J Public Health* 2003; 93:1137–1143
  59. Charlson M, Szatrowski TP, Peterson J, Gold J: Validation of a combined comorbidity index. *J Clin Epidemiol* 1994; 47:1245–1251
  60. Healthcare cost and utilization Project: clinical classifications Software (CCS) for ICD-10-PCS (beta version). Agency for healthcare research and quality. Available at: <https://www.hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp>. [Accessed 2 July 2020]
  61. McCoy TH, Castro VM, Snapper LA, Hart KL, Perlis RH: Efficient genome-wide association in biobanks using topic modeling identifies multiple novel disease loci. *Mol Med Camb Mass* 2017; 23:285–294
  62. McCoy TH, Castro VM, Snapper L, et al: Polygenic loading for major depression is associated with specific medical comorbidity. *Transl Psychiatry* 2017; 7:e1238

## EHR Prediction of COVID-19 Delirium

63. Castro VM, McCoy TH, Cagan A, et al: Stratification of risk for hospital admissions for injury related to fall: cohort study. *BMJ* 2014; 349:g5863
64. McCoy TH, Castro VM, Cagan A, Roberson AM, Perlis RH: Validation of a risk stratification tool for fall-related injury in a state-wide cohort. *BMJ Open* 2017; 7:e012189
65. Bennett CC: Utilizing RxNorm to support practical computing applications: capturing medication history in live electronic health records. *J Biomed Inform* 2012; 45:634–641
66. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO: The Unified medical language system: an informatics research collaboration. *J Am Med Inform Assoc JAMIA* 1998; 5:1–11
67. McCoy TH, Han L, Pellegrini AM, Tanzi RE, Berretta S, Perlis RH: Stratifying risk for dementia onset using large-scale electronic health record data: a retrospective cohort study. *Alzheimers Dement J Alzheimers Assoc* 2020; 16:531–540
68. Butler R: The ICD-10 general equivalence Mappings. Bridging the translation gap from ICD-9. *J AHIMA* 2007; 78:84–85
69. Tibshirani R: Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B Methodol* 1996; 58:267–288
70. Hosmer DW, Lemeshow S, Sturdivant RX: *Applied Logistic Regression*, 3 edition. Hoboken, New Jersey: Wiley; 2013
71. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L: A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inform Assoc JAMIA* 2020; 27:621–633
72. Green DM, Swets JA: *Signal detection theory and psychophysics*. New York: Wiley; 1966
73. Fawcett T: An introduction to ROC analysis. *Pattern Recognit Lett* 2006; 27:861–874
74. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29–36
75. Marsaglia G, Tsang WW, Wang J: Evaluating Kolmogorov's distribution. *J Stat Softw* 2003; 8:1–4
76. Stephens MA: EDF statistics for Goodness of Fit and Some comparisons. *J Am Stat Assoc* 1974; 69:730–737
77. Baker SG, Cook NR, Vickers A, Kramer BS: Using relative utility curves to evaluate risk prediction. *J R Stat Soc Ser A Stat Soc* 2009; 172:729–748
78. Vickers AJ, Elkin EB: Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak Int J Soc Med Decis Mak* 2006; 26:565–574
79. Vickers AJ, Cronin AM, Elkin EB, Gonen M: Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak* 2008; 8:53
80. Dankers FJWM, Traverso A, Wee L, van Kuijk SMJ: Prediction modeling methodology. In: Kubben P, Dumontier M, Dekker A, (eds) *Fundamentals of Clinical Data Science*. Cham (CH): Springer; 2019
81. Youden WJ: Index for rating diagnostic tests. *Cancer* 1950; 3:32–35
82. McCoy TH: Mapping the delirium literature through Probabilistic topic modeling and network analysis: a computational Scoping review. *Psychosomatics* 2019; 60:105–120
83. Ahmed S, Leurent B, Sampson EL: Risk factors for incident delirium among older people in acute hospital medical units: a systematic review and meta-analysis. *Age Ageing* 2014; 43:326–333
84. Zaal IJ, Devlin JW, Peelen LM, Slooter AJC: A systematic review of risk factors for delirium in the ICU. *Crit Care Med* 2015; 43:40–47
85. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G: Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med* 2018; 178:1544–1547
86. Obermeyer Z, Powers B, Vogeli C, Mullainathan S: Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019; 366:447–453
87. van den Boogaard M, Pickkers P, Slooter AJC, et al: Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: observational multicentre study. *BMJ* 2012; 344:e420
88. van den Boogaard M, Schoonhoven L, Maseda E, et al: Recalibration of the delirium prediction model for ICU patients (PRE-DELIRIC): a multinational observational study. *Intensive Care Med* 2014; 40:361–369
89. Green C, Bonavia W, Toh C, Tiruvoipati R: Prediction of ICU delirium: validation of Current delirium predictive models in routine clinical Practice. *Crit Care Med* 2019; 47:428–435
90. Pencina MJ, Goldstein BA, D'Agostino RB: Prediction models - development, evaluation, and clinical application. *N Engl J Med* 2020; 382:1583–1586
91. Wynants L, Van Calster B, Collins GS, et al: Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020; 369:m1328
92. McCoy TH, Pellegrini AM, Perlis RH: Assessment of time-Series machine learning methods for Forecasting hospital discharge Volume. *JAMA Netw Open* 2018; 1:e184087