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.

General Hospital Psychiatry 74 (2022) 9–17

Contents lists available at ScienceDirect

General Hospital Psychiatry

journal homepage: www.elsevier.com/locate/genhospsych

Research paper

Longitudinal validation of an electronic health record delirium prediction model applied at admission in COVID-19 patients

Victor M. Castro [a,b], Kamber L. Hart [a], Chana A. Sacks [c], Shawn N. Murphy [b,d], Roy H. Perlis [a], Thomas H. McCoy Jr [a,*]

[a] Center for Quantitative Health, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA
[b] Research Information Science and Computing, Mass General Brigham, 399 Revolution Drive, Somerville, MA 02145, USA
[c] Department of Medicine, Massachusetts General Hospital, 100 Cambridge Street, Boston, MA 02114, USA
[d] Department of Neurology, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA

ARTICLE INFO

Keywords:
Delirium
Predictive modeling
Machine learning
Electronic health records
Replication
COVID-19

ABSTRACT

Objective: To validate a previously published machine learning model of delirium risk in hospitalized patients with coronavirus disease 2019 (COVID-19).
Method: Using data from six hospitals across two academic medical networks covering care occurring after initial model development, we calculated the predicted risk of delirium using a previously developed risk model applied to diagnostic, medication, laboratory, and other clinical features available in the electronic health record (EHR) at time of hospital admission. We evaluated the accuracy of these predictions against subsequent delirium diagnoses during that admission.
Results: Of the 5102 patients in this cohort, 716 (14%) developed delirium. The model's risk predictions produced a c-index of 0.75 (95% CI, 0.73–0.77) with 27.7% of cases occurring in the top decile of predicted risk scores. Model calibration was diminished compared to the initial COVID-19 wave.
Conclusion: This EHR delirium risk prediction model, developed during the initial surge of COVID-19 patients, produced consistent discrimination over subsequent larger waves; however, with changing cohort composition and delirium occurrence rates, model calibration decreased. These results underscore the importance of calibration, and the challenge of developing risk models for clinical contexts where standard of care and clinical populations may shift.

1. Introduction

People who are hospitalized with COVID-19 develop wide ranging neuropsychiatric symptoms [1–5] including the acute confusional state of delirium [6–9], which has been shown to occur in 10 to 50% of COVID-19 patients [7,8]. More generally, delirium is the most common neuropsychiatric syndrome encountered in the general hospital setting and a condition of particular concern to psychiatrists working with medical and surgical patients [10–12]. Delirium is associated with a wide range of adverse outcomes including increased critical care utilization, longer length of stay, increased rate of institutional discharge, worse functional outcomes, increased rate of readmission, and increased mortality [13–34]. In addition to poorer clinical outcomes, care of patients with delirium is associated with greater caregiver and clinician burden [35–42]. Fortunately, delirium can be prevented through multicomponent interventions [43–47]; however, optimal allocation of these multicomponent prevention resources, as with any other scarce resource, requires a way of identifying individuals at greatest risk for experiencing delirium [48].

The COVID-19 pandemic, with its associated increased demands on caregivers and hospital clinical resources, has increased the potential consequences of a case of delirium, as critical care services and individual caregivers are already overburdened. It has also increased the difficulty of delivering multicomponent interventions that could reduce delirium risk, by necessitating additional infection control precautions and protective equipment conservation efforts [49–51]. This combination of circumstances, which developed during care of the various surges of COVID-19 patients, increased the importance of zero-contact stratification of delirium risk.

Although delirium is often underdiagnosed and under coded,

* Corresponding author at: Simches Research Building, Massachusetts General Hospital, 185 Cambridge St, 6th Floor, Boston, MA 02114, USA.
E-mail address: thmccoy@partners.org (T.H. McCoy).

https://doi.org/10.1016/j.genhosppsych.2021.10.005
Received 30 August 2021; Received in revised form 25 October 2021; Accepted 27 October 2021
Available online 2 November 2021
0163-8343/© 2021 Elsevier Inc. All rights reserved.

electronic health records (EHR) data can be used to study both the epidemiology and biology of delirium [52–59]. Building on this potential, EHR data produced during routine care have been repurposed for secondary use in developing clinical prediction models of delirium risk [60–67]. We previously reported the development and contemporaneous external validation of a machine learning model of delirium risk among patients with COVID-19 presenting during the spring of 2020 [68]. As the characteristics, treatments, and outcomes of COVID-19 have changed over the course of the pandemic, and clinical model predictive accuracy can vary over time, we sought to replicate this model using the later waves of pandemic patients. In other words, we aimed for a replication which is longitudinal with respect to the patients included in the study but not with respect to an individual patient [69–74]. This work builds on a readily adopted and freely available model, applicable during emergency department care, based on facts already present in the EHR prior to inpatient hospitalization. It seeks to evaluate the extent to which the predictive quality of the model remained stable over successive waves of pandemic care despite evolving patient characteristics and treatment approaches.

2. Materials and methods

2.1. Conceptual approach

We have previously shown that EHR data can be used at the time of hospital admission to predict delirium diagnosis over the subsequent hospital stay in patients with COVID-19. This prediction is possible using a statistical model developed through machine learning which makes a prediction based on diagnostic history, lab test results, medication orders, and other patient characteristics. This development occurred during the first wave of COVID-19 cases; however, over the intervening waves of COVID-19 cases, the composition of patients and treatment approaches changed a great deal [74,75]. The aim of this study was to re-validate the model in these subsequent waves of patients to establish whether the quality of statistical predictions had decreased. The overall methodological approach below outlines who was studied, how the data in their health records were made presentable to the statistical model, and finally the multiple approaches taken to establishing the quality of the predictions made by the statistical model.

2.2. Cohort development

This study included all adults hospitalized with PCR-confirmed SARS-CoV-2 between June 1, 2020, and April 30, 2021, across six hospitals, including two academic medical centers and four community hospitals. The original development of the previously published model occurred during the first surge of patients and included those hospitalized prior to May 31, 2020. This validation is longitudinal in that it occurs in a cohort who were cared for after – in calendar time – those in whom the model was originally developed and validated. As such, it provides an estimate of how the model would have performed had it been deployed clinically during this time. While this study is longitudinal with respect to cohort and evolution of the pandemic, it is only minimally longitudinal with respect to the individual patients included as the span of individual time considered is merely that between emergency room presentation and hospital discharge.

All data for this study—including the demographic, laboratory, diagnostic, medication, and other clinical data—were extracted from the health system EHR [68]. These included date of birth for calculation of age at admission, home zip code for area deprivation index calculation, body mass index (BMI), and lifetime smoking status [76]. The study protocol was approved by the Mass General Brigham 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 informed consent.

2.3. Clinical data handling and feature encoding

To make statistical predictions from the data in the EHR, those chart-like clinical facts must be systematically reshaped into spreadsheet-like data which are suitable for predictive modeling. This reshaping of EHR data into spreadsheet-like data used the same steps as those of the predictive model's initial development and validation [68]. The 34 features – columns in the notional spreadsheet – included in the predictive model (which was used in exactly the previously published form without modification of any kind) cover a range of diagnostic codes, medications, laboratory results, and registration characteristics. Those features and the coefficients, which come from the original model development and are what would be required to implement this model, are included in Supplemental Table 1. The included features are notable for the relative paucity of protective factors (that is features which are associated with a lower risk of delirium) relative to risk factors (features which are associated with greater risk of delirium when present). Clinical features associated with reduced predicted risk of delirium include serum albumin and body mass index. Clinical features associated with increased predicted risk include laboratory reported high troponin and low absolute lymphocyte count. Of the 34 total features included in the model and presented in Supplemental Tables 1, 16 features are based on current lab tests results, 6 are based on prior diagnosis, 8 are based on medication prescription, and the remainder are patient and demographic characteristics (e.g., age or smoking history).

For this replication study, as in the original development and validation study, BMI, lifetime smoking status, and age at admission were all drawn from predefined structured fields within the EHR. Prior clinical diagnostic codes (both billing or non-billing problem list) were aggregated from native International Classification of Disease (ICD) codes to the second level of the Healthcare Utilization Project (HCUP) Clinical Classification Software (CCS) hierarchy as log-transformed counts of the total number of codes assigned within that category [77]. Medications were encoded as at the Unified Medical Language System (UMLS) RxNorm ingredient as the log-transformed count of orders and prescriptions over the 30 days preceding admission [60,78,79]. For clinical laboratory test results, both continuous features in laboratory supplied units and laboratory supplied flags (logical true or false variables) of abnormally high and low values were extracted as required by the model from the first occurrence within the episode of care (i.e., the first emergency room collection of a given test for those who had multiple instances of the test prior to admission). For count encoded features, medications or diagnostic codes which did not occur were counted zero times. Missing continuous values (e.g., laboratory values) were replaced through medium imputation from the original training sample and thus a prediction was possible on every admission over the study period.

To facilitate cohort description and evaluation of model performance within clinical and demographic substrata, we extracted additional clinical and demographic data from the EHR, including data which occurred after prediction time. These data were not made available to the original model training and played no part in prediction; thus, these evaluative features were unable to contaminate the antecedent predictions (sometimes referred to as 'peeking' forward in time [80]). Instead, these features were merely used in secondary analysis of predictive accuracy. Given the importance and frequency of delirium in the critical care setting and delirium superimposed on dementia, dementia diagnosis and subsequent critical care were selected [81–85]; however, it is notable that the former could conceptually be excluded at the time of prediction based on prevalent dementia diagnosis whereas the latter eventual incident cases of critical illness could not. To evaluate model performance over demographic strata, we extracted race and gender from structured EHR demographics tables. Recognizing the clinical importance of critical illness and cognitive impairment in delirium, we extracted both subsequent need for critical care and a previously described cognitive impairment feature from the EHR [86].

For our delirium outcome, we extracted coded delirium diagnosis

made during the subsequent hospital admission using the same coded definitions used in initial model development and also applied a natural language processing (NLP) delirium definition to the discharge summaries of the hospital admission as a sensitivity analysis on the coded definition [54,57]. The full bag of codes used to define delirium based on the Medicare General Equivalence Mappings was: F01.51, F03.90, F05, F06.0, F06.1, F06.2, F06.30, F06.4, F06.8, F10.231, F15.920, F19.921, F19.939, F19.950, F19.951, F19.97, G92, G93.40, G93.41, G93.49, R40.0, R40.4, and R41.82. The NLP-based definition produced equivalent results to the primary analysis done against the coded delirium definition and thus these duplicative results are not shown. Conceptually, it is important to note that within a given patient the timing of features used can be divided into three distinct eras of chart time: (1) features used for the statistical prediction of subsequent delirium risk (all of which were available at the time of admission; e.g. emergency department lab tests and past diagnostic history), (2) the clinical diagnosis of delirium used as the reference outcome in evaluating predictive accuracy (which was only available at the time of hospital discharge), and (3) information which played no part in the prediction or outcome but was merely used for post hoc secondary description of predictive accuracy in specific subgroups (e.g. need for critical care).

2.4. Model evaluation and statistical analysis

We characterized the COVID-19 patient cohort and contexts in which patients received care using descriptive statistics (i.e., counts and percentages for discrete variables, means and standard deviations for continuous measures) stratified by eventual diagnosis of delirium. Thereafter, we evaluated the discrimination – the extent to which those who went on to develop delirium were consistently predicted as higher risk than those who did not – and calibration – the extent to which all those predicted to have a given risk of delirium did develop delirium at the predicted rate – of delirium risk predictions made at the time of admission by reference to the presence or absence of an eventual coded diagnosis of delirium over the course of the subsequent hospitalization. We evaluated model discrimination, both in the pooled cohort and by clinical and demographic strata, using area under the receiver operating characteristic (ROC) curve [87]. We evaluated model calibration using the quantile-by-quintile comparison of predicted and observed outcome rate, Spiegelhalter's z statistic, and inspection of calibration plots including both logistic and flexible loess curves [88–91]. We also calculated the familiar measures of diagnostic testing accuracy (e.g. sensitivity, specificity, negative predictive value, and positive predictive value) by dichotomizing risk predictions at the high-risk threshold identified in the initial model validation cohort [92]. To quantify the expected benefits from intervention in any given case we completed a decision curve analysis of this cohort [93–95]. Parallel sensitivity check analysis using the NLP augmented delirium outcome identified only a small number of additional cases and the prediction results were consistent with the primary analysis done by reference to the coded diagnosis definition and thus the duplicative NLP results are not shown. All analysis was conducted using R version 4 [96].

3. Results

The cohort included 5102 patients of whom 716 were diagnosed with delirium (14%). The mean age on admission was 62.7 years old (SD 18.73) and 2520 (49.4%) patients were less than 65 years of age on admission. The cohort included 285 (5.6%) patients with a prior diagnosis of dementia and 2525 (49.5%) were treated at community hospitals. Full demographic and clinical characteristics of the cohort are shown in Table 1.

The area under the ROC curve (AUC) for the COVID-19 delirium model in this longitudinal replication cohort was 0.75 (95% CI 0.73–0.77; Fig. 1, left), similar to that previously reported for this model (AUC 0.75 [95% CI 0.71–0.79] [68]). In this replication cohort, the observed and predicted prevalence of delirium differed significantly in quantile-by-quintile comparison of predicted and observed outcome rate ($\chi^2(23) = 56.05$, $p < 0.001$). Spiegelhalter's Z was significant ($Z = -2.29$, $p = 0.02$) and calibration plot showed lower expected outcome rates in those with high predicted risk (Fig. 1, right). As such, in this model those who did develop delirium likely had higher risk scores whereas those patients at a specific risk score were less likely to go on to develop delirium at exactly that predicted rate. Model predictions were broadly consistent across clinical and demographic contrasts (Fig. 2, Supplemental Table 2); however, predictive discrimination was generally worse in high-risk patients 65 years and older (AUC 0.67 95% CI 0.64–0.70) and those with a history of dementia (AUC 0.58 95% CI 0.52–0.65).

Table 1
Sociodemographic and clinical characteristics of the patient cohort stratified by delirium diagnosis.

	Overall	Delirium case	Non-case
n	5102	716	4386
Male gender (%)	2603 (51.0)	406 (56.7)	2197 (50.1)
Race (%)			
Black	614 (12.0)	78 (10.9)	536 (12.2)
Other	1259 (24.7)	165 (23.0)	1094 (24.9)
White	3229 (63.3)	473 (66.1)	2756 (62.8)
Age (decade, %)			
<30	304 (6.0)	15 (2.1)	289 (6.6)
30s	461 (9.0)	18 (2.5)	443 (10.1)
40s	444 (8.7)	21 (2.9)	423 (9.6)
50s	804 (15.8)	66 (9.2)	738 (16.8)
60s	1000 (19.6)	141 (19.7)	859 (19.6)
70s	1036 (20.3)	181 (25.3)	855 (19.5)
80+	1053 (20.6)	274 (38.3)	779 (17.8)
Prior dementia history (%)	285 (5.6)	112 (15.6)	173 (3.9)
ICU care (%)	795 (15.6)	255 (35.6)	540 (12.3)
Community hospitals (%)	2525 (49.5)	293 (40.9)	2232 (50.9)

Patients with higher risk scores were enriched for eventual delirium diagnosis: the top decile of risk scores included 27.7% of the total delirium cases; if all patients were ranked from highest risk to lowest risk the 10% of patients with the highest risk scores at admission ultimately accounted for 27.7% of delirium cases by discharge. The model produced a lift of 2.4 in the highest risk quintile which captured 48.5% of all delirium cases with a delirium occurrence rate of 34%. The lowest risk quintile included only 25 delirium cases for a case rate of 2.4% and lift of 0.17. At the previously identified optimal predicted risk cut point of 0.15 – that is treating those with a predicted risk greater than 0.15 as "positive" cases and those less than the cut off as "negative" cases – the predictions produced a sensitivity of 0.62 (0.58–0.65), specificity of 0.75 (0.73–0.76), negative predictive value of 0.92 (0.91–0.93), and positive predictive value of 0.28 (0.26–0.31) relative to the eventual diagnosis of delirium. Beyond this specific cut off point for determination of high-risk patients, decision curve analysis (Fig. 3) showed a wide range of risk thresholds over which risk stratification by the predictive model produced superior results to either the naive allocation strategy of universal intervention (i.e., intervening to prevent delirium in every admission) or that of never-intervention (i.e., providing no intervention regardless of risk).

4. Discussion

This replication of a previously developed, externally validated delirium prediction model based on EHR data available at the time of hospital admission showed comparable discrimination and diminished calibration in this cohort of more than 5000 COVID-19 patients. While the model was not applied clinically during this period, the present evaluation reflects what would have occurred during a silent evaluative deployment – i.e., no changes were made in the model features, all of which are available at time of hospital admission, and no patients were excluded. The AUC in this replication cohort of subsequent waves of

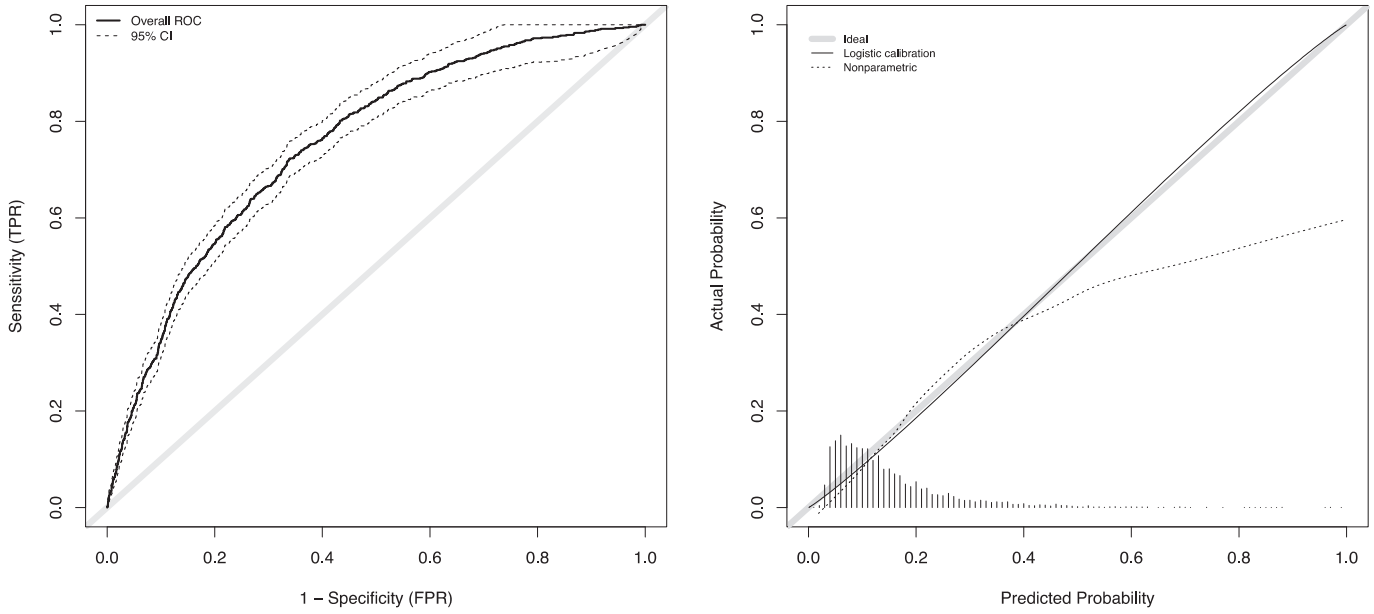

Fig. 1. Overall receiver operating characteristic (ROC) curve (left) and model calibration plot (right) for the machine learning model in the full longitudinal replication cohort.

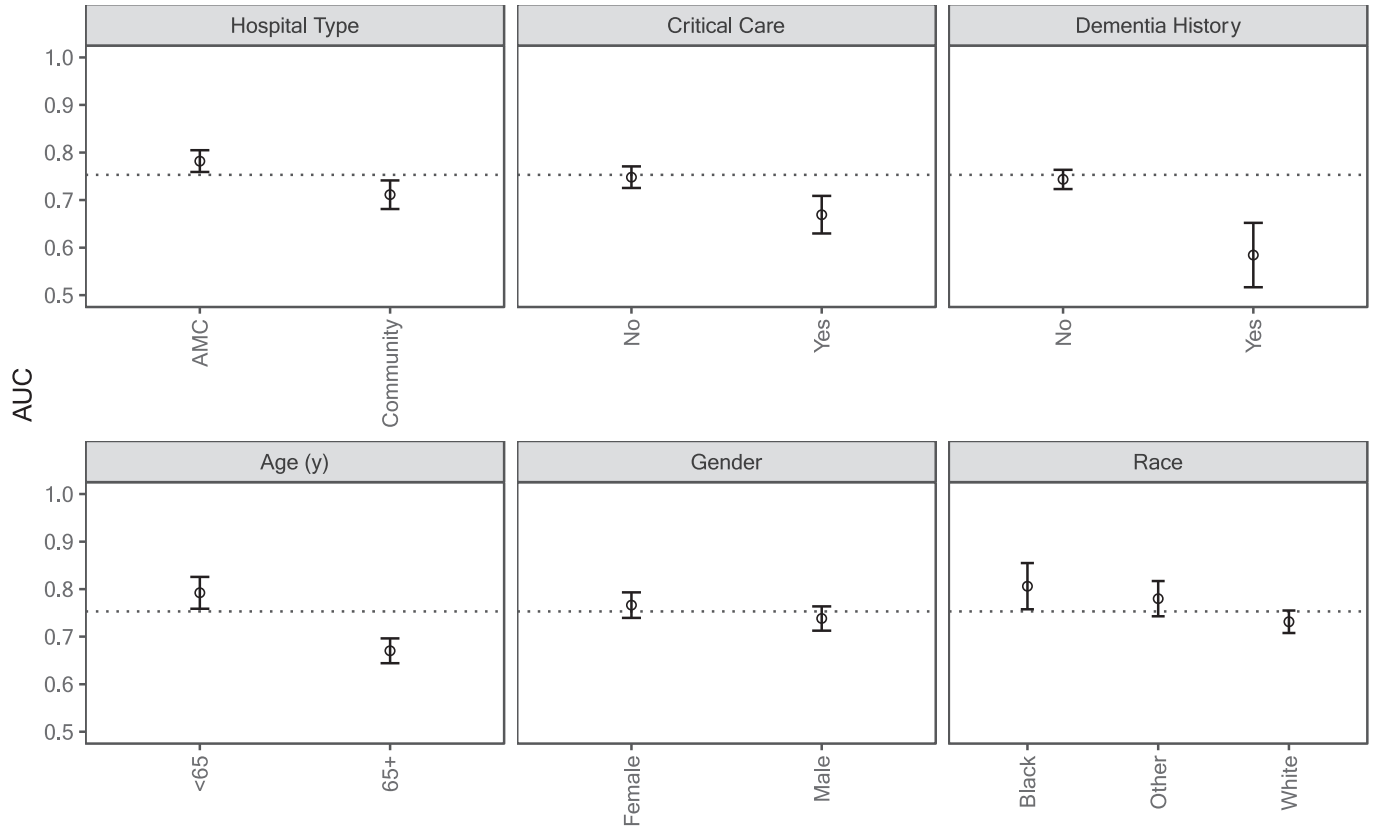

Fig. 2. Area under the ROC curve (AUC) and 95% confidence intervals within clinical and demographic strata of the cohort compared to the overall full cohort AUC (horizontal dotted line).

patients with COVID-19 was 0.75, similar to that previously reported for this model in initial external validation cohort patients with COVID-19 during the initial wave of cases and more broadly consistent with those of diverse delirium prediction models (independent of COVID-19) summarized in a recent systematic review of the topic [64,68]. Similarly, this wide range of results falls within the range of results seen for predictions made in patients with COVID-19 about outcomes other than delirium [97]; however, the most directly comparable application of these methods to outcomes other than delirium in patients with COVID-19 produced superior discrimination to that seen in this

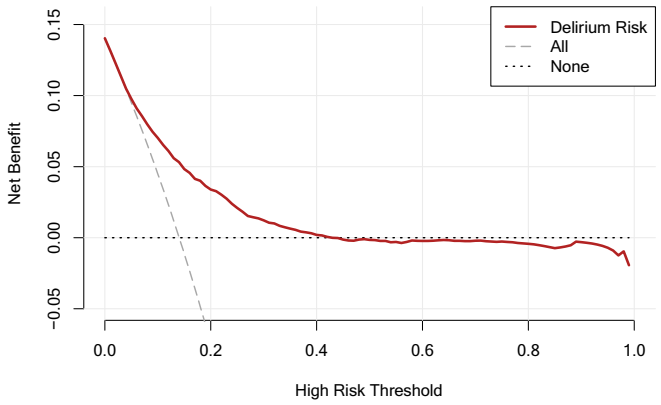

Fig. 3. Decision curve analysis of net benefit of risk stratification using the delirium predictive model as compared to intervening in all patients or intervening in no patients strategies.

replication cohort of delirium as an outcome [98]. This stability over subsequent waves of patients with COVID-19 is reassuring evidence of model robustness given the rapidly evolving patient population and treatment norms over the course of consecutive surges in patient volume [74,99–102]. Model calibration – how closely the observed probability of the outcome matches the predicted probability of the outcome [103,104] – declined in this longitudinal validation cohort relative to that previously reported in the initial cross sectional external validation cohort. Nevertheless, patients with high predicted risk were highly enriched for actual observed cases with 27.7% of cases occurring in the highest decile of predicted risk scores. In light of emerging concern for risk of bias in machine learning models, it is notable that the overall sample AUC fell within or below the confidence interval for AUCs stratified by both race and gender [105]; however, clinical characteristics such as older age, dementia history, need for critical care, and treatment at a community hospital were associated with reduced discriminative accuracy.

The longitudinal reduction in calibration we observe is consistent with the change in observed delirium rate from 18.9% of patients in whom this model was developed to 14.0% in this subsequent replication cohort. Although the initial cohort from the spring of 2020 surge in whom the model was developed and this subsequent longitudinal replication cohort had similar average ages (62.9 vs 62.7 years respectively) the prevalence of preexisting dementia diagnosis dropped from 11.3% in the initial surge of patients to 5.6% in this replication cohort as did the proportion of patients requiring critical care, which dropped from 22.5% in the initial surge to 15.6% in this longitudinal cohort. This reduction in critical care is consistent with the declining rates of mechanical ventilation and fluctuating mortality reported elsewhere [73,99]. It is possible that the reduction in dementia diagnostic history represents a shift from nursing home residents to community dwelling older adults over subsequent waves [106,107]. These underlying changes in patient characteristics are an instance of concept drift and demonstrate the associated compromise of internal validity consistent with a rapidly evolving pandemic [108–111]. That model discrimination held despite this drift is reassuring; however, the impact on calibration highlights the importance of automated methods for continuous model validation and potential approaches to recalibration [112,113]. In sum, this model could be used to support population risk stratification to direct a scarce resource (e.g., delirium prevention programs) toward those patients who are most likely to be diagnosed with delirium. Whether risk predictions made by this model are adequately calibrated for counseling of individual patients about their specific predicted risk would be context specific but is of secondary relevance to the question of optimal allocation of a limited resources in the setting of pandemic associated scarcity.

This result must be viewed in the context of its strengths and weaknesses. The reliance on EHR data is a strength in that it facilitates deployment to clinical practice without expectation of any change in routine operations or real time clinical data generation; however, the use of an EHR covering an open system means that care rendered elsewhere, as in another health system, cannot be considered. The absence of relevant prior data – for example, dementia diagnosis made elsewhere – likely bias prediction results toward the null, as relevant clinical detail that would indicate increased risk may not be available. Similarly, as delirium is underdiagnosed in routine practice [114–117], the use of a clinical diagnosis instead of active bedside screening for model evaluation likely biases results toward more conservative estimates of discrimination. The impact of this discordance between delirium diagnosis and the syndrome of delirium may not be uniformly distributed over delirium motoric subtypes in this COVID-19 population [118,119]. Outside of pandemic illness where contact and resource scarcity present unique constraints, it is possible that predictive approaches of this kind might be used to direct delirium assessments to improve on the current problem of under recognition [120]. While granting that delirium is underdiagnosed in claims data, potentially by an order of magnitude relative to formal assessment studies, secular trends in delirium diagnosis suggest increasing rates of recognition in routine practice. The 14% occurrence rate here is within the range of rates observed in studies of delirium in COVID [7,8,58,114,121]. Whereas previous research using free text identified more evidence of delirium than reliance on coded diagnosis alone that was not the case in this analysis in which sensitivity checks of our primary coded diagnosis outcome using a previously described outcome which included discharge summary text was not substantially changed and thus is not shown [57]. Given this uncertainty around the quality of coded diagnosis in COVID-19 patients during the pandemic the role and relevance of otherwise common underdiagnosis is an important area of uncertainty for future work. In addition to uncertainty about the rate of underdiagnosis in the EHR this study does not address questions about post hospital outcomes of individual patients with COVID-19 as it is longitudinal with respect to timing of cohorts not individual clinical course. The stratified analysis by dementia presents unique challenges as well as dementia, like delirium, is often underdiagnosed [122–125] and when present delirium superimposed on dementia presents an ambiguous clinical picture [81,126].

A key strength of this work is that it is a longitudinal replication using an unmodified previously published predictive model and thus it is protected against overfitting and associated threats to replication [127–129]. This replication provides important information on the likely robustness of predictive models developed in the early phases of a pandemic for application in later phases of a pandemic illness as would be required in any attempt to do algorithmic allocation [130]; however, it does so in the setting of a single geographic region, thus this result represents a replication over time but not a replication over place. Nevertheless, complete replications of this sort are rare and an important assessment of model robustness [131]. Additionally, this model makes a single risk prediction at the moment of inpatient admission about risk of delirium diagnosis over the course of the subsequent hospitalization – this is both a strength and a limitation. The strength of this setup is that the prediction is based on facts which were already available in the EHR at the time of admission and thus the reported predictions could be made in the operational clinical EHR and doing so would be well-timed to stratify delirium preventative resources; on the other hand, the limitation of this prediction setup is that it is incapable of targeting the specific moment – over a potentially lengthy hospitalization – at which the fluctuation course of delirium would produce the greatest symptom burden. Delirium has been formulated in a diathesis stress model. Within that frame, this pre-admission risk prediction is attending primarily to the diathesis [132]. Finally, it is important to note that this model of all patients with COVID-19 did not achieve equivalent discrimination to that produced by models developed for targeted clinical context like surgical and critical care [133,134]. This difference

within clinical populations is of particular importance as improvements at the upper limit of AUC are more challenging [135].

4.1. Conclusions

This evaluation of a previously developed and validated machine learning model predicting delirium risk at the time of hospital admission with COVID-19 demonstrates accurate risk stratification through stable model discrimination over the course of the second surge of COVID-19 patients, in spite of substantial changes in the underlying patient characteristics. This stability of model discrimination provides evidence that tools of this sort could be developed in the initial phases of a pandemic and applied thereafter; however, close monitoring would be required to assess model performance over subsequent waves in the setting of changing treatment patterns or patient cohorts. Nevertheless, it is reassuring that the performance of this model was stable despite changes in the patient population and treatment approaches over the months following initial model development. Beyond the pandemic setting, this longitudinal replication provides evidence that predictive models applied to EHR data on admission could support optimal allocation of scarce resources like prevention efforts or formal diagnostic efforts to improve recognition.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.genhosppsych.2021.10.005.

Data availability statement

The IRB approval under which this individual health information was used does not allow redistribution of the clinical data.

Disclosures

Mr. Castro and Ms. Hart report 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 Brain and Behavior Research Foundation, National Institute of Mental Health, National Institute of Nursing Research, National Human Genome Research Institute Home, and Telefonica Alfa.

Data availability

The authors do not have permission to share data.

Acknowledgements

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

The authors would like to thank the Partners Healthcare research computing and data repository team for their support in developing health records databases to study the COVID-19 pandemic.

References

[1] Nepal G, Rehrig JH, Shrestha GS, Shing YK, Yadav JK, Ojha R, et al. Neurological manifestations of COVID-19: a systematic review. Crit Care 2020;24:421. https://doi.org/10.1186/s13054-020-03121-z.
[2] Ahmad I, Rathore FA. Neurological manifestations and complications of COVID-19: a literature review. J Clin Neurosci 2020;77:8–12. https://doi.org/10.1016/j.jocn.2020.05.017.
[3] Asadi-Pooya AA, Simani L. Central nervous system manifestations of COVID-19: a systematic review. J Neurol Sci 2020;413:116832. https://doi.org/10.1016/j.jns.2020.116832.
[4] Whittaker A, Anson M, Harky A. Neurological manifestations of COVID-19: a systematic review and current update. Acta Neurol Scand 2020;142:14–22. https://doi.org/10.1111/ane.13266.
[5] Favas TT, Dev P, Chaurasia RN, Chakravarty K, Mishra R, Joshi D, et al. Neurological manifestations of COVID-19: a systematic review and meta-analysis of proportions. Neurol Sci 2020;41:3437–70. https://doi.org/10.1007/s10072-020-04801-y.
[6] Cipriani G, Danti S, Nuti A, Carlesi C, Lucetti C, Di Fiorino M. A complication of coronavirus disease 2019: delirium. Acta Neurol Belg 2020;120:927–32. https://doi.org/10.1007/s13760-020-01401-7.
[7] Ticinesi A, Cerundolo N, Parise A, Nouvenne A, Prati B, Guerra A, et al. Delirium in COVID-19: epidemiology and clinical correlations in a large group of patients admitted to an academic hospital. Aging Clin Exp Res 2020;32:2159–66. https://doi.org/10.1007/s40520-020-01699-6.
[8] Pun BT, Badenes R, Heras La Calle G, Orun OM, Chen W, Raman R, et al. Prevalence and risk factors for delirium in critically ill patients with COVID-19 (COVID-D): a multicentre cohort study. Lancet Respir Med 2021;9:239–50. https://doi.org/10.1016/S2213-2600(20)30552-X.
[9] Paterson RW, Brown RL, Benjamin L, Nortley R, Wiethoff S, Bharucha T, et al. The emerging spectrum of COVID-19 neurology: clinical, radiological and laboratory findings. Brain 2020;143:3104–20. https://doi.org/10.1093/brain/awaa240.
[10] Maldonado JR. Acute brain failure: pathophysiology, diagnosis, management, and sequelae of delirium. Crit Care Clin 2017;33:461–519. https://doi.org/10.1016/j.ccc.2017.03.013.
[11] McCoy TH. Mapping the delirium literature through probabilistic topic modeling and network analysis: a computational scoping review. Psychosomatics 2019;60:105–20. https://doi.org/10.1016/j.psym.2018.12.003.
[12] Nisavic M, Shuster JL, Gitlin D, Worley L, Stern TA. Readings on psychosomatic medicine: survey of resources for trainees. Psychosomatics 2015;56:319–28. https://doi.org/10.1016/j.psym.2014.12.006.
[13] Cole MG, Primeau FJ. Prognosis of delirium in elderly hospital patients. Can Med Assoc J 1993;149:41–6.
[14] Crocker E, Beggs T, Hassan A, Denault A, Lamarche Y, Bagshaw S, et al. Long-term effects of postoperative delirium in patients undergoing cardiac operation: a systematic review. Ann Thorac Surg 2016;102:1391–9. https://doi.org/10.1016/j.athoracsur.2016.04.071.
[15] Girard TD, Thompson JL, Pandharipande PP, Brummel NE, Jackson JC, Patel MB, et al. Clinical phenotypes of delirium during critical illness and severity of subsequent long-term cognitive impairment: a prospective cohort study. Lancet Respir Med 2018;6:213–22. https://doi.org/10.1016/S2213-2600(18)30062-6.
[16] Gleason LJ, Schmitt EM, Kosar CM, Tabloski P, Saczynski JS, Robinson T, et al. Effect of delirium and other major complications on outcomes after elective surgery in older adults. JAMA Surg 2015;150:1134–40. https://doi.org/10.1001/jamasurg.2015.2606.
[17] Goldberg TE, Chen C, Wang Y, Jung E, Swanson A, Ing C, et al. Association of delirium with long-term cognitive decline: a meta-analysis. JAMA Neurol 2020. https://doi.org/10.1001/jamaneurol.2020.2273.
[18] Khouli H, Astua A, Dombrowski W, Ahmad F, Homel P, Shapiro J, et al. Changes in health-related quality of life and factors predicting long-term outcomes in older adults admitted to intensive care units. Crit Care Med 2011;39:731–7. https://doi.org/10.1097/CCM.0b013e318208edf8.
[19] Kiely DK, Marcantonio ER, Inouye SK, Shaffer ML, Bergmann MA, Yang FM, et al. Persistent delirium predicts increased mortality. J Am Geriatr Soc 2009;57:55–61. https://doi.org/10.1111/j.1532-5415.2008.02092.x.
[20] Koster S, Hensens AG, van der Palen J. The long-term cognitive and functional outcomes of postoperative delirium after cardiac surgery. Ann Thorac Surg 2009;87:1469–74. https://doi.org/10.1016/j.athoracsur.2009.02.080.
[21] Leslie DL, Inouye SK. The importance of delirium: economic and societal costs. J Am Geriatr Soc 2011;59:S241–3.
[22] 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.
[23] McCusker J, Cole M, Abrahamowicz M, Primeau F, Belzile E. Delirium predicts 12-month mortality. Arch Intern Med 2002;162:457–63.
[24] McCusker J, Cole M, Dendukuri N, Belzile É, Primeau F. Delirium in older medical inpatients and subsequent cognitive and functional status: a prospective study. Can Med Assoc J 2001;165:575–83.
[25] Pandharipande PP, Girard TD, Ely EW. Long-term cognitive impairment after critical illness. N Engl J Med 2014;370:185–6. https://doi.org/10.1056/NEJMc1313886.
[26] 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. https://doi.org/10.1016/j.ahj.2015.04.013. 79–86.e1.
[27] Salluh JI, Wang H, Schneider EB, Nagaraja N, Yenokyan G, Damluji A, et al. Outcome of delirium in critically ill patients: systematic review and meta-analysis. BMJ 2015;350:h2538.
[28] Schubert M, Schürch R, Boettger S, Garcia Nuñez D, Schwarz U, Bettex D, et al. A hospital-wide evaluation of delirium prevalence and outcomes in acute care patients - a cohort study. BMC Health Serv Res 2018;18:550. https://doi.org/10.1186/s12913-018-3345-x.
[29] 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;32:539–47. https://doi.org/10.1002/gps.4491.

[30] Vasilevskis EE, Chandrasekhar R, Holtze CH, Graves J, Speroff T, Girard TD, et al. The cost of ICU delirium and coma in the intensive care unit patient. Med Care 2018;56:890–7. https://doi.org/10.1097/MLR.0000000000000975.

[31] Weinrebe W, Johannsdottir E, Karaman M, Füsgen I. What does delirium cost? Z Gerontol Geriatr 2016;49:52–8.

[32] Wolters AE, van Dijk D, Pasma W, Cremer OL, Looije MF, de Lange DW, et al. Long-term outcome of delirium during intensive care unit stay in survivors of critical illness: a prospective cohort study. Crit Care 2014;18:R125. https://doi.org/10.1186/cc13929.

[33] Zhang Z, Pan L, Ni H. Impact of delirium on clinical outcome in critically ill patients: a meta-analysis. Gen Hosp Psychiatry 2013;35:105–11. https://doi.org/10.1016/j.genhosppsych.2012.11.003.

[34] van den Boogaard M, Schoonhoven L, Evers AWM, van der Hoeven JG, van Achterberg T, Pickkers P. Delirium in critically ill patients: impact on long-term health-related quality of life and cognitive functioning. Crit Care Med 2012;40:112–8. https://doi.org/10.1097/CCM.0b013e31822e9fc9.

[35] Breitbart W, Gibson C, Tremblay A. The delirium experience: delirium recall and delirium-related distress in hospitalized patients with cancer, their spouses/caregivers, and their nurses. Psychosomatics 2002;43:183–94. https://doi.org/10.1176/appi.psy.43.3.183.

[36] Bruera E, Bush SH, Willey J, Paraskevopoulos T, Li Z, Palmer JL, et al. Impact of delirium and recall on the level of distress in patients with advanced cancer and their family caregivers. Cancer 2009;115:2004–12. https://doi.org/10.1002/cncr.24215.

[37] Fong TG, Racine AM, Fick DM, Tabloski P, Gou Y, Schmitt EM, et al. The caregiver burden of delirium in older adults with Alzheimer disease and related disorders. J Am Geriatr Soc 2019;67:2587–92. https://doi.org/10.1111/jgs.16199.

[38] Grossi E, Lucchi E, Gentile S, Trabucchi M, Bellelli G, Morandi A. Preliminary investigation of predictors of distress in informal caregivers of patients with delirium superimposed on dementia. Aging Clin Exp Res 2020;32:339–44. https://doi.org/10.1007/s40520-019-01194-7.

[39] Morandi A, Lucchi E, Turco R, Morghen S, Guerini F, Santi R, et al. Delirium superimposed on dementia: a quantitative and qualitative evaluation of informal caregivers and health care staff experience. J Psychosom Res 2015;79:272–80. https://doi.org/10.1016/j.jpsychores.2015.06.012.

[40] Morita T, Akechi T, Ikenaga M, Inoue S, Kohara H, Matsubara T, et al. Terminal delirium: recommendations from bereaved families' experiences. J Pain Symptom Manage 2007;34:579–89. https://doi.org/10.1016/j.jpainsymman.2007.01.012.

[41] Mossello E, Lucchini F, Tesi F, Rasero L. Family and healthcare staff's perception of delirium. Eur Geriatr Med 2020;11:95–103. https://doi.org/10.1007/s41999-019-00284-z.

[42] Toye C, Matthews A, Hill A, Maher S. Experiences, understandings and support needs of family carers of older patients with delirium: a descriptive mixed methods study in a hospital delirium unit. Int J Older People Nurs 2014;9:200–8. https://doi.org/10.1111/opn.12019.

[43] Hshieh TT, Yang T, Gartaganis SL, Yue J, Inouye SK. Hospital elder life program: systematic review and meta-analysis of effectiveness. Am J Geriatr Psychiatry 2018;26:1015–33. https://doi.org/10.1016/j.jagp.2018.06.007.

[44] Hshieh TT, Yue J, Oh E, Puelle M, Dowal S, Travison T, et al. Effectiveness of multicomponent nonpharmacological delirium interventions: a meta-analysis. JAMA Intern Med 2015;175:512–20.

[45] Khan A, Boukrina O, Oh-Park M, Flanagan NA, Singh M, Oldham M. Preventing delirium takes a village: systematic review and meta-analysis of delirium preventive models of care. J Hosp Med 2019;14. https://doi.org/10.12788/jhm.3212. E1–7.

[46] 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;13:384–91. https://doi.org/10.1097/SPC.0000000000000458.

[47] Wang Y-Y, Yue J-R, Xie D-M, Carter P, Li Q-L, Gartaganis SL, 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 2020;180:17. https://doi.org/10.1001/jamainternmed.2019.4446.

[48] Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: development and validation of the 4C mortality score. BMJ 2020;370:m3339. https://doi.org/10.1136/bmj.m3339.

[49] 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 2020;24:176. https://doi.org/10.1186/s13054-020-02882-x.

[50] O'Hanlon S, Inouye SK. Delirium: a missing piece in the COVID-19 pandemic puzzle. Age Ageing 2020;49:497–8. https://doi.org/10.1093/ageing/afaa094.

[51] Anmella G, Arbelo N, Fico G, Murru A, Llach CD, Madero S, et al. COVID-19 inpatients with psychiatric disorders: real-world clinical recommendations from an expert team in consultation-liaison psychiatry. J Affect Disord 2020;274:1062–7. https://doi.org/10.1016/j.jad.2020.05.149.

[52] Hope C, Estrada N, Weir C, Teng C-C, Damal K, Sauer BC. Documentation of delirium in the VA electronic health record. BMC Res Notes 2014;7:208. https://doi.org/10.1186/1756-0500-7-208.

[53] Coombes CE, Coombes KR, Fareed N. A novel model to label delirium in an intensive care unit from clinician actions. BMC Med Inform Decis Mak 2021;21:97. https://doi.org/10.1186/s12911-021-01461-6.

[54] 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;31:631–40. https://doi.org/10.1007/s10877-016-9873-1.

[55] 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;53:312–8. https://doi.org/10.1111/j.1532-5415.2005.53120.x.

[56] Kim DH, Lee J, Kim CA, Huybrechts KF, Bateman BT, Patorno E, et al. Evaluation of algorithms to identify delirium in administrative claims and drug utilization database: delirium identification in claims data. Pharmacoepidemiol Drug Saf 2017;26:945–53. https://doi.org/10.1002/pds.4226.

[57] McCoy Jr 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;58:113–20. https://doi.org/10.1016/j.psym.2016.10.007.

[58] McCoy TH, Hart KL, Perlis RH. Characterizing and predicting rates of delirium across general hospital settings. Gen Hosp Psychiatry 2017;46:1–6. https://doi.org/10.1016/j.genhosppsych.2017.01.006.

[59] 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. https://doi.org/10.1016/j.neurobiolaging.2018.03.008.

[60] McCoy TH, Castro VM, Hart KL, Perlis RH. Stratified delirium risk using prescription medication data in a state-wide cohort. Gen Hosp Psychiatry 2021;71:114–20. https://doi.org/10.1016/j.genhosppsych.2021.05.001.

[61] Hercus C, Hudaib A-R. Delirium misdiagnosis risk in psychiatry: a machine learning-logistic regression predictive algorithm. BMC Health Serv Res 2020;20:151. https://doi.org/10.1186/s12913-020-5005-1.

[62] Lee A, Mu JL, Joynt GM, Chiu CH, Lai VKW, Gin T, 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;118:391–9. https://doi.org/10.1093/bja/aew476.

[63] Lee S, Harland K, Mohr NM, Matthews G, Hess EP, Bellolio MF, et al. Evaluation of emergency department derived delirium prediction models using a hospital-wide cohort. J Psychosom Res 2019;127:109850. https://doi.org/10.1016/j.jpsychores.2019.109850.

[64] Lindroth H, Bratzke L, Purvis S, Brown R, Coburn M, Mrkobrada M, et al. Systematic review of prediction models for delirium in the older adult inpatient. BMJ Open 2018;8:e019223. https://doi.org/10.1136/bmjopen-2017-019223.

[65] Menzenbach J, Guttenthaler V, Kirfel A, Ricchiuto A, Neumann C, Adler L, 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. https://doi.org/10.1016/j.conctc.2019.100501.

[66] Wassenaar A, Schoonhoven L, Devlin JW, van Haren FMP, Slooter AJC, Jorens PG, et al. Delirium prediction in the intensive care unit: comparison of two delirium prediction models. Crit Care 2018;22:114. https://doi.org/10.1186/s13054-018-2037-6.

[67] Wassenaar A, van den Boogaard M, van Achterberg T, Slooter AJC, Kuiper MA, Hoogendoorn ME, et al. Multinational development and validation of an early prediction model for delirium in ICU patients. Intensive Care Med 2015;41:1048–56. https://doi.org/10.1007/s00134-015-3777-2.

[68] Castro VM, Sacks CA, Perlis RH, McCoy TH. Development and external validation of a delirium prediction model for hospitalized patients with coronavirus disease 2019. J Acad Consult Liaison Psychiatry 2021. https://doi.org/10.1016/j.jaclp.2020.12.005.

[69] Roth GA, Emmons-Bell S, Alger HM, Bradley SM, Das SR, de Lemos JA, et al. Trends in patient characteristics and COVID-19 in-hospital mortality in the United States during the COVID-19 pandemic. JAMA Netw Open 2021;4:e218828. https://doi.org/10.1001/jamanetworkopen.2021.8828.

[70] Sarnovsky M, Kolarik M. Classification of the drifting data streams using heterogeneous diversified dynamic class-weighted ensemble. PeerJ Comput Sci 2021;7:e459. https://doi.org/10.7717/peerj-cs.459.

[71] McCoy TH, Pellegrini AM, Perlis RH. Assessment of time-series machine learning methods for forecasting hospital discharge volume. JAMA Netw Open 2018;1:e184087. https://doi.org/10.1001/jamanetworkopen.2018.4087.

[72] Beyene AA, Welemariam T, Persson M, Lavesson N. Improved concept drift handling in surgery prediction and other applications. Knowl Inf Syst 2015;44:177–96. https://doi.org/10.1007/s10115-014-0756-9.

[73] Yang W, Kandula S, Huynh M, Greene SK, Van Wye G, Li W, et al. Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis. Lancet Infect Dis 2021;21:203–12. https://doi.org/10.1016/S1473-3099(20)30769-6.

[74] The RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with Covid-19. N Engl J Med 2021;384:693–704. https://doi.org/10.1056/NEJMoa2021436.

[75] Chen Y-T. The effect of vaccination rates on the infection of COVID-19 under the vaccination rate below the herd immunity threshold. IJERPH 2021;18:7491. https://doi.org/10.3390/ijerph18147491.

[76] Knighton AJ, Savitz L, Belnap T, Stephenson B, VanDerslice J. Introduction of an area deprivation index measuring patient socio-economic status in an integrated health system: implications for population health. EGEMs 2016;4(9). https://doi.org/10.13063/2327-9214.1238.

[77] Healthcare Cost and Utilization Project (HCUP). Clinical Classifications Software (CCS) for ICD-10-PCS. Rockville, MD: Agency for Healthcare Research and Quality; 2017. https://www.hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp. [Accessed 2 July 2020].

[78] Bennett CC. Utilizing RxNorm to support practical computing applications: capturing medication history in live electronic health records. J Biomed Inform 2012;45:634–41. https://doi.org/10.1016/j.jbi.2012.02.011.

[79] 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. https://doi.org/10.1136/bmjopen-2016-012189.

[80] Diciotti S, Ciulli S, Mascalchi M, Giannelli M, Toschi N. The "peeking" effect in supervised feature selection on diffusion tensor imaging data. AJNR Am J Neuroradiol 2013;34:E107. https://doi.org/10.3174/ajnr.A3685.

[81] Fick DM, Agostini JV, Inouye SK. Delirium superimposed on dementia: a systematic review. J Am Geriatr Soc 2002;50:1723–32. https://doi.org/10.1046/j.1532-5415.2002.50468.x.

[82] Fick D, Foreman M. Consequences of not recognizing delirium superimposed on dementia in hospitalized elderly individuals. J Gerontol Nurs 2000;26:30–40. https://doi.org/10.3928/0098-9134-20000101-09.

[83] Fick DM, Hodo DM, Lawrence F, Inouye SK. Recognizing delirium superimposed on dementia: assessing Nurses' knowledge using case vignettes. J Gerontol Nurs 2007;33:40–9. https://doi.org/10.3928/00989134-20070201-09.

[84] Voyer P, Cole MG, McCusker J, Belzile É. Prevalence and symptoms of delirium superimposed on dementia. Clin Nurs Res 2006;15:46–66. https://doi.org/10.1177/1054773805282299.

[85] Fiest KM, Soo A, Hee Lee C, Niven DJ, Ely EW, Doig CJ, et al. Long-term outcomes in ICU patients with delirium: a population-based cohort study. Am J Respir Crit Care Med 2021;204:412–20. https://doi.org/10.1164/rccm.202002-0320OC.

[86] 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 2020. https://doi.org/10.1016/j.jalz.2019.09.084. j.jalz.2019.09.084.

[87] Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett 2006;27:861–74. https://doi.org/10.1016/j.patrec.2005.10.010.

[88] Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

[89] Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016;74:167–76. https://doi.org/10.1016/j.jclinepi.2015.12.005.

[90] Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. Stat Med 1986;5:421–33. https://doi.org/10.1002/sim.4780050506.

[91] Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken, New Jersey: Wiley; 2013.

[92] Dankers FJWM, Traverso A, Wee L, van Kuijk SMJ. Prediction modeling methodology. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of clinical data science. Cham: Springer International Publishing; 2019. p. 101–20. https://doi.org/10.1007/978-3-319-99713-1_8.

[93] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.

[94] 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;8:1.

[95] 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;172:729–48. https://doi.org/10.1111/j.1467-985X.2009.00592.x.

[96] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

[97] Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328. https://doi.org/10.1136/bmj.m1328.

[98] Castro VM, McCoy TH, Perlis RH. Laboratory findings associated with severe illness and mortality among hospitalized individuals with coronavirus disease 2019 in eastern Massachusetts. JAMA Netw Open 2020;3:e2023934. https://doi.org/10.1001/jamanetworkopen.2020.23934.

[99] Yeates EO, Nahmias J, Chinn J, Sullivan B, Stopenski S, Amin AN, et al. Improved outcomes over time for adult COVID-19 patients with acute respiratory distress syndrome or acute respiratory failure. PLoS One 2021;16:e0253767. https://doi.org/10.1371/journal.pone.0253767.

[100] Greene DN, Jackson ML, Hillyard DR, Delgado JC, Schmidt RL. Decreasing median age of COVID-19 cases in the United States—changing epidemiology or changing surveillance? PLoS One 2020;15:e0240783. https://doi.org/10.1371/journal.pone.0240783.

[101] Horwitz LI, Jones SA, Cerfolio RJ, Francois F, Greco J, Rudy B, et al. Trends in COVID-19 risk-adjusted mortality rates. J Hosp Med 2021;16:90–2. https://doi.org/10.12788/jhm.3552.

[102] Dennis JM, McGovern AP, Vollmer SJ, Mateen BA. Improving survival of critical care patients with coronavirus disease 2019 in England: a National Cohort Study, march to June 2020*. Crit Care Med 2021;49:209–14. https://doi.org/10.1097/CCM.0000000000004747.

[103] Walsh CG, Sharman K, Hripcsak G. Beyond discrimination: a comparison of calibration methods and clinical usefulness of predictive models of readmission risk. J Biomed Inform 2017;76:9–18. https://doi.org/10.1016/j.jbi.2017.10.008.

[104] Lindhiem O, Petersen IT, Mentch LK, Youngstrom EA. The importance of calibration in clinical psychology. Assessment 2020;27:840–54. https://doi.org/10.1177/1073191117752055.

[105] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366:447–53. https://doi.org/10.1126/science.aax2342.

[106] McMichael TM, Currie DW, Clark S, Pogosjans S, Kay M, Schwartz NG, et al. Epidemiology of Covid-19 in a long-term care facility in King County, Washington. N Engl J Med 2020;382:2005–11. https://doi.org/10.1056/NEJMoa2005412.

[107] Shen K, Loomer L, Abrams H, Grabowski DC, Gandhi A. Estimates of COVID-19 cases and deaths among nursing home residents not reported in Federal Data. JAMA Netw Open 2021;4:e2122885. https://doi.org/10.1001/jamanetworkopen.2021.22885.

[108] Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google flu: traps in big data analysis. Science 2014;343:1203–5. https://doi.org/10.1126/science.1248506.

[109] Jung K, Shah NH. Implications of non-stationarity on predictive modeling using EHRs. J Biomed Inform 2015;58:168–74. https://doi.org/10.1016/j.jbi.2015.10.006.

[110] Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA Jt Summits Transl Sci Proc 2020;2020:191–200.

[111] Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. J Am Med Inform Assoc 2017;24:1052–61. https://doi.org/10.1093/jamia/ocx030.

[112] Davis SE, Greevy RA, Fonnesbeck C, Lasko TA, Walsh CG, Matheny ME. A nonparametric updating method to correct clinical prediction model drift. J Am Med Inform Assoc 2019;26:1448–57. https://doi.org/10.1093/jamia/ocz127.

[113] Vieira DM, Fernandes C, Lucena C, Lifschitz S. Driftage: a multi-agent system framework for concept drift detection. Gigascience 2021;10. https://doi.org/10.1093/gigascience/giab030. giab030.

[114] McCoy TH, Snapper L, Stern TA, Perlis RH. Underreporting of delirium in statewide claims data: implications for clinical care and predictive modeling. Psychosomatics 2016;57:480–8. https://doi.org/10.1016/j.psym.2016.06.001.

[115] Milisen K, Foreman MD, Wouters B, Driesen R, Godderis J, Abraham IL, et al. Documentation of delirium in elderly patients with hip fracture. J Gerontol Nurs 2002;28:23–9. https://doi.org/10.3928/0098-9134-20021101-07.

[116] Vollmer CM, Bond J, Eden BM, Resch DS, Fulk L, Robinson S, et al. Incidence, prevalence, and under-recognition of delirium in urology patients. Urol Nurs 2010;30:235.

[117] Lastrapes K, Dang M, Cassel JB, Orr T, Proffitt T, Del Fabbro E. Delirium documentation in hospitalized pediatric patients with cancer. Palliat Support Care 2021;19:283–6. https://doi.org/10.1017/S1478951521000171.

[118] Krewulak KD, Stelfox HT, Leigh JP, Ely EW, Fiest KM. Incidence and prevalence of delirium subtypes in an adult ICU: a systematic review and meta-analysis*. Crit Care Med 2018;46:2029–35. https://doi.org/10.1097/CCM.0000000000003402.

[119] Pandharipande P, Cotton BA, Shintani A, Thompson J, Costabile S, Truman Pun B, et al. Motoric subtypes of delirium in mechanically ventilated surgical and trauma intensive care unit patients. Intensive Care Med 2007;33:1726–31. https://doi.org/10.1007/s00134-007-0687-y.

[120] Raman R, Chen W, Harhay MO, Thompson JL, Ely EW, Pandharipande PP, et al. Dealing with missing delirium assessments in prospective clinical studies of the critically ill: a simulation study and reanalysis of two delirium studies. BMC Med Res Methodol 2021;21:97. https://doi.org/10.1186/s12874-021-01274-1.

[121] Ryan DJ, O'Regan NA, Caoimh RÓ, Clare J, O'Connor M, Leonard M, et al. Delirium in an adult acute hospital population: predictors, prevalence and detection. BMJ Open 2013;3:e001772. https://doi.org/10.1136/bmjopen-2012-001772.

[122] Amjad H, Roth DL, Sheehan OC, Lyketsos CG, Wolff JL, Samus QM. Underdiagnosis of dementia: an observational study of patterns in diagnosis and awareness in US older adults. J Gen Intern Med 2018;33:1131–8. https://doi.org/10.1007/s11606-018-4377-y.

[123] Lang L, Clifford A, Wei L, Zhang D, Leung D, Augustine G, et al. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open 2017;7:e011146. https://doi.org/10.1136/bmjopen-2016-011146.

[124] Shao Y, Zeng QT, Chen KK, Shutes-David A, Thielke SM, Tsuang DW. Detection of probable dementia cases in undiagnosed patients using structured and unstructured electronic health records. BMC Med Inform Decis Mak 2019;19:128. https://doi.org/10.1186/s12911-019-0846-4.

[125] Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. Aging Ment Health 2011;15:978–84. https://doi.org/10.1080/13607863.2011.596805.

[126] Leonard MM, Agar M, Spiller JA, Davis B, Mohamad MM, Meagher DJ, et al. Delirium diagnostic and classification challenges in palliative care: subsyndromal delirium, comorbid delirium-dementia, and psychomotor subtypes. J Pain Symptom Manage 2014;48:199–214. https://doi.org/10.1016/j.jpainsymman.2014.03.012.

[127] LaPlante DA. Replication is fundamental, but is it common? A call for scientific self-reflection and contemporary research practices in gambling-related research. Int Gambl Stud 2019;19:362–8. https://doi.org/10.1080/14459795.2019.1672768.

[128] Wacker J. Increasing the reproducibility of science through close cooperation and forking path analysis. Front Psychol 2017;8:1332. https://doi.org/10.3389/fpsyg.2017.01332.

[129] Wicherts JM, Veldkamp CLS, Augusteijn HEM, Bakker M, van Aert RCM, van Assen MALM. Degrees of freedom in planning, running, analyzing, and reporting

psychological studies: a checklist to avoid p-hacking. Front Psychol 2016;7:1832. https://doi.org/10.3389/fpsyg.2016.01832.
[130] Rubin EB, Knipe RS, Israel RA, McCoy TH, Courtwright AM. Existing crisis standards of care triage protocols may not significantly differentiate between patients with coronavirus disease 2019 who require intensive care. Crit Care Explor 2021;3:e0412. https://doi.org/10.1097/CCE.0000000000000412.
[131] Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol 2015;68:25–34. https://doi.org/10.1016/j.jclinepi.2014.09.007.
[132] Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. Lancet 2014;383:911–22. https://doi.org/10.1016/S0140-6736(13)60688-1.
[133] Kim MY, Park UJ, Kim HT, Cho WH. DELirium prediction based on hospital information (Delphi) in general surgery patients. Medicine 2016;95:e3072. https://doi.org/10.1097/MD.0000000000003072.
[134] Van den Boogaard M, Pickkers P, Slooter A, Kuiper M, Spronk P, Van Der Voort P, 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.
[135] Pencina MJ, D'Agostino RB, Massaro JM. Understanding increments in model performance metrics. Lifetime Data Anal 2013;19:202–18. https://doi.org/10.1007/s10985-012-9238-0.