Evaluation of the E-PRE-DELIRIC prediction model for ICU delirium: a retrospective validation in a UK general ICU

Authors:

1. Sarah L. Cowan, Addenbrooke's Hospital, Cambridge, UK
2. Jacobus Preller, Addenbrooke's Hospital, Cambridge, UK
3. Robert J. B. Goudie, MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, UK

Institutions:

1. Addenbrooke's Hospital, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK
2. MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, CB2 0SR, UK

Corresponding author:

Robert J. B. Goudie

MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, CB2 0SR, UK

+44 (0) 1223 330370

robert.goudie@mrc-bsu.cam.ac.uk

Keywords: clinical scoring systems, critical care, delirium, intensive care, prediction model, validation

Word count: 598

Introduction

E-PRE-DELIRIC is a point-of-admission ICU delirium risk-prediction tool [1], with reported good or moderate performance [2-4]. In this study we assessed its performance in a large UK teaching hospital general ICU using routinely-collected data, as approved by the local Research Data Governance Committee.

Methods

We retrospectively analysed data for 2445 consecutive ICU admissions (November 2014 to June 2017). Patients were routinely assessed for delirium, using twice-daily Confusion Assessment Method for the ICU (CAM-ICU) assessment [5]. As in previous E-PRE-DELIRIC studies [1-4], delirium was defined as any positive CAM-ICU assessment, or antipsychotics initiation while on ICU.

Table 1: Patient characteristics in this study, the E-PRE-DELIRIC development dataset [1] and other validation studies [2-4].

Factor	This study	Development dataset [1]	DECISION study [2-3]	Green et al [4]
Admissions during study period, *n*	2445	–	2802	803
Included in analysis, *n* (%)	1746 (71.4)	1962 (–)	2178 (77.7)	455 (56.7)
Delirium, *n* (%)	763 (43.7)	481 (24.5)	466 (21.4)	160 (35.2)
Age (years), mean (Q1-Q3, min/max)	58.6 (47.0-71.8, 18/94)	61.7 (53-74, 18/95)	62.1 (–)	66.7 (49.0-77.3, –/–)
Male, *n* (%)	1010 (57.8)	1166 (59.4)	1324 (60.8)	241 (53.0)
Admission category, *n* (%)				

Surgery	813 (46.6)	1019 (51.9)	1079 (49.5)	–
Medicine	837 (47.9)	683 (34.8)	859 (39.3)	–
Trauma	33 (1.9)	90 (4.6)	86 (4.0)	–
Neurology/neurosurgery	63 (3.6)	170 (8.7)	157 (7.2)	–
Urgent admission, *n* (%)	1534 (87.9)	1163 (59.3)	1345 (61.8)	–
APACHE II	20.0 (mean)	–	17.4 (mean)	16 (median)
ICU LoS (days), median (Q1-Q3, min/max)	4.5 (2.4-10.0, 1.0/184.0)	2.0 (1-6, 1/133)	3.0 (2-6, 1/96)	2.6 (1.5-4.4, –/–)
ICU mortality, *n* (%)	210 (12.0)	–	–	17 (3.7)

– indicates the figure was not reported.

We adopted the original E-PRE-DELIRIC exclusion criteria [1], excluding 683 ICU admissions for: ICU stay <24 hours (425 admissions), incomplete CAM-ICU data (152), delirium on admission (50), comatose throughout entire ICU stay (47), age under 18 (9). 16 admissions were excluded due to missing E-PRE-DELIRIC components. 1746 admissions (1569 unique patients) remained for analysis; this 71.4% inclusion rate is consistent with previous studies (Table 1).

Results and Discussion

763 delirium cases were identified (43.7% of ICU admissions), a higher incidence than reported previously (Table 1). This is likely due to differences in the study population compared to previous studies: more patients were classified as urgent; the mean APACHE II score was higher; and median length of stay (LoS) was longer (Table 1).

[Figure 1 here]

The mean E-PRE-DELIRIC score was 0.269 (Q1-Q3; 0.154–0.371). The histogram of E-PRE-DELIRIC scores shows extensive overlap between patients who did and did not develop delirium (Figure 1a). The Receiver Operator Characteristic (ROC) curve (Figure 1b), and the Precision Recall (PR) curve (Figure 1c), showing precision (positive predictive value, PPV) against recall (sensitivity), both indicate moderate-to-poor discriminative performance. The Area under the ROC (AUROC) was 0.628 (95% CI 0.602–0.653). The Area Under the PR Curve (AUPRC) was 0.534. For sensitivity >0.1, PPV was between 0.437 and 0.585, indicating only around half of patients predicted to develop delirium actually did, in a population with 43.7% incidence. Refitting the E-PRE-DELIRIC logistic regression model to our data hardly improved discrimination: AUROC was 0.648 (95% CI 0.622–0.673) and AUPRC was 0.566.

The calibration plot, of predicted risk against observed delirium rate, shows the risk of delirium is considerably underestimated, especially in patients with predicted risk of delirium less than 0.5 (Figure 1d). Poor calibration is corroborated by the calibration slope model logit(probability of delirium) = alpha + beta * logit(p), where p is the E-PRE-DELIRIC score [6]. The estimated slope beta=0.58 (95% CI 0.46–0.71) is significantly below 1, indicating the predicted probabilities are overly variable; and the estimated intercept alpha=0.84 (95% CI 0.74–0.95) is significantly above 0 when fixing beta=1, indicating the predicted probabilities are predominantly too low. E-PRE-DELIRIC is particularly poorly calibrated for the surgical patients in the study, many of whom have major intra-abdominal pathology: those with predicted risk <10% had an observed incidence of 26%.

Of 763 delirium cases, 563 were CAM-ICU-positive and 200 were included due to antipsychotic initiation. When including only CAM-ICU-positive delirium, calibration was improved (alpha=0.29) but remained overly variable (beta=0.52), while discrimination was similar (AUROC 0.615; AUPRC 0.396, with 32.2% observed incidence).

While E-PRE-DELIRIC is intended as a point-of-admission score, some of its exclusion criteria are retrospective (LoS; CAM-ICU completeness; comatose throughout). To assess real-world performance we repeated our analysis without these criteria. AUROC (0.615) and AUPRC (0.423, with 35.0% observed incidence) remained similar.

Conclusion

In this population, the E-PRE-DELIRIC score is not as discriminative or as well-calibrated as previously reported. PPV was only slightly higher than delirium incidence, meaning the utility of E-PRE-DELIRIC for guiding clinical decision-making in this population is limited.

List of abbreviations

AUROC: Area Under the ROC curve

AUPRC: Area Under the PR Curve

CAM-ICU: Confusion Assessment Method for the ICU

E-PRE-DELIRIC: Early Prediction model for Delirium in ICU patients

ICU: Intensive Care Unit

LoS: Length of Stay

PPV: Positive Predictive Value

PR curve: Precision-Recall curve

ROC: Receiver Operator Characteristic

References

1. 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;41:1048-1056.
2. 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.
3. Wassenaar A, Schoonhoven L, Devlin JW, van Haren FMP, Slooter AJC, Jorens PG, et al. External validation of two models to predict delirium in critically ill adults using

either the Confusion Assessment Method-ICU or the Intensive Care Delirium Screening Checklist for delirium assessment. Crit Care Med. 2019;47:e827-e835.

4. 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.
5. Ely EW, Margolin R, Francis J, May L, Truman B, Dittus R, et al. Evaluation of delirium in critically ill patients: Validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). Crit Care Med. 2001;29:1370-1379.
6. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015;162:W1–W73.

Declarations

Ethics approval and consent to participate: The use of the anonymous data used in this study was approved by the Cambridge Clinical Informatics Research Data Governance Committee.

Consent for publication: Not applicable.

Availability of data and materials: The data that support the findings of this study are available from Cambridge Clinical Informatics, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. The data are anonymised but to preserve patient confidentiality and privacy the Data Use Agreement states that the data cannot be deposited into open access repositories of any kind. Anyone wishing to access data, must submit and receive approval for access to these data from the Cambridge Clinical Informatics Research Data Governance Committee.

Competing interests: The authors declare that they have no competing interests.

Funding: This study was supported by the UK National Institute for Health Research (NIHR) through the Cambridge Biomedical Research Centre (BRC), with data provided by Cambridge Clinical Informatics (Led by Drs. Afzal Chaudhry and Lydia Drumright). RJBG was supported by the UK Medical Research Council [programme code MC_UU_00002/2]. The funders had no role in the design, collection, analysis, interpretation or writing of the manuscript.

Authors' contributions: SLC interpreted and guided the statistical analyses, and drafted the manuscript. JP conceived the study, interpreted the study results and substantively revised the manuscript. RJBG designed the study, conducted and interpreted the statistical analyses and substantively revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements: This study was supported by the UK National Institute for Health Research (NIHR) through the Cambridge Biomedical Research Centre (BRC), with data provided by Cambridge Clinical Informatics (Led by Drs. Afzal Chaudhry and Lydia Drumright). RJBG was supported by the UK Medical Research Council [programme code MC_UU_00002/2]. We are grateful to Vince Taylor for his careful work extracting data for this study.

Figures

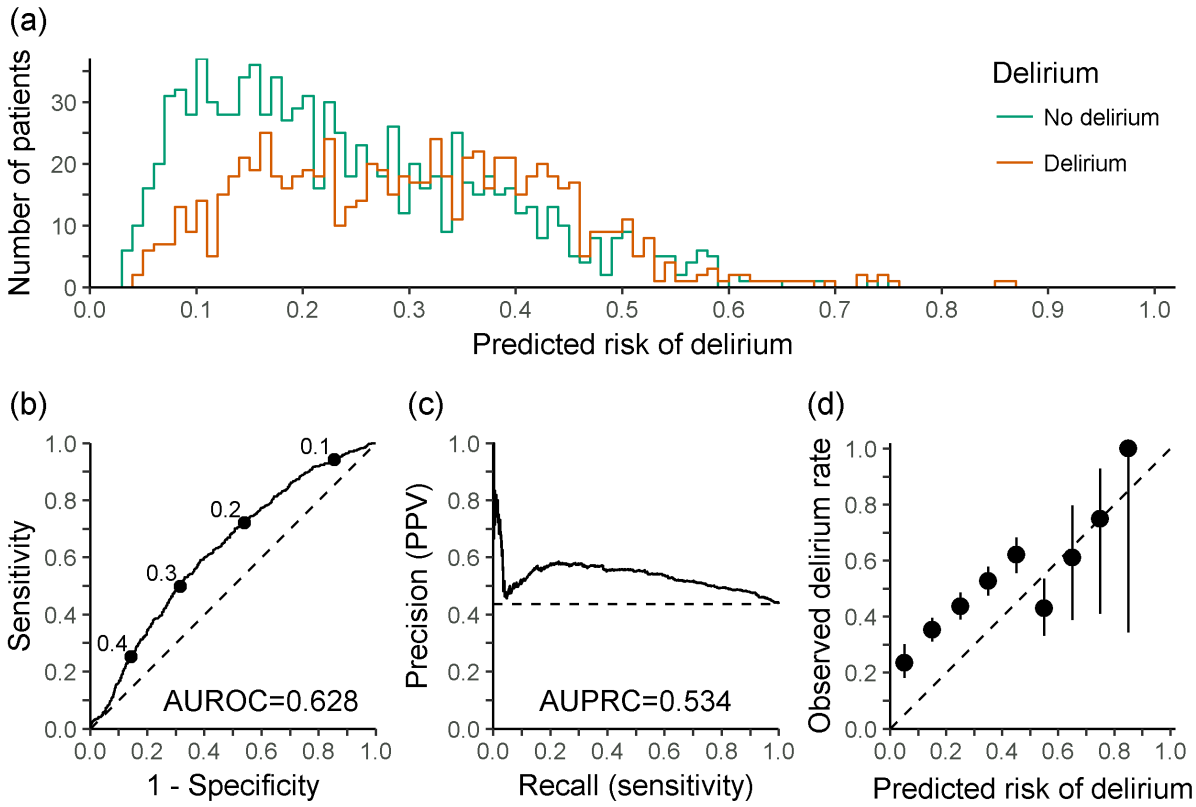

Figure 1. (a) Histogram of predicted risk of delirium by delirium status; (b) Receiver Operator Characteristic plot, with labels indicating the corresponding threshold and the dashed line indicating the line of no discrimination; (c) Precision-Recall plot, with the 43.7% observed incidence indicated by dashed line; (d) Calibration plot (with 95% CI), by tenths of predicted risk, with the dashed line indicating perfect calibration.