nature medicine

LETTERS

https://doi.org/10.1038/s41591-020-0951-z

A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor

Katharina Schultebraucks[1,2,3] ✉, Arieh Y. Shalev[1], Vasiliki Michopoulos[4,5], Corita R. Grudzen[6], Soo-Min Shin[6], Jennifer S. Stevens[4], Jessica L. Maples-Keller[4], Tanja Jovanovic[7], George A. Bonanno[8], Barbara O. Rothbaum[4], Charles R. Marmar[1,9], Charles B. Nemeroff[10,11], Kerry J. Ressler[4,12] and Isaac R. Galatzer-Levy[1,13]

Annually, approximately 30 million patients are discharged from the emergency department (ED) after a traumatic event[1]. These patients are at substantial psychiatric risk, with approximately 10–20% developing one or more disorders, including anxiety, depression or post-traumatic stress disorder (PTSD)[2–4]. At present, no accurate method exists to predict the development of PTSD symptoms upon ED admission after trauma[5]. Accurate risk identification at the point of treatment by ED services is necessary to inform the targeted deployment of existing treatment[6–9] to mitigate subsequent psychopathology in high-risk populations[10,11]. This work reports the development and validation of an algorithm for prediction of post-traumatic stress course over 12 months using two independently collected prospective cohorts of trauma survivors from two level 1 emergency trauma centers, which uses routinely collectible data from electronic medical records, along with brief clinical assessments of the patient's immediate stress reaction. Results demonstrate externally validated accuracy to discriminate PTSD risk with high precision. While the predictive algorithm yields useful reproducible results on two independent prospective cohorts of ED patients, future research should extend the generalizability to the broad, clinically heterogeneous ED population under conditions of routine medical care.

Previous studies identified multiple trauma-related predictive signals of PTSD risk[7,12–17], including aspects of the biological stress response[18–23], immune response[24–26], threat perception, psychophysiological arousal[15,19,27] and psychosocial determinants of clinical risk[28]. Many indicators related to these biological systems and psychosocial indicators are routinely collected in the ED and logged in the electronic medical records (EMRs), making them viable as candidate predictors of risk. Some factors, such as self-reported psychological stress, are not yet part of the medical routine and only about 7% of level 1 trauma centers routinely screen for PTSD[29].

Notably, PTSD comes with long-term clinical and pecuniary costs to both the individual and the healthcare system. While empirically validated treatments are effective in reducing the risk for PTSD[6,8,9], early prevention strategies are typically not implemented due to the lack of established methods for timely and reliable risk identification[11]. The ED visit is often the sole contact of trauma survivors with the healthcare system and the time immediately after trauma opens a critical window to prevent the development of PTSD[11,30]. Accurate identification of risk for PTSD during ED evaluation using algorithms running on accessible data sources would provide new opportunities for cost-effective and scalable methods of risk assessment and intervention to reduce the prevalence of PTSD without posing high additional burden for ED personnel[11].

The use of predictive models to integrate multiple post-traumatic stress (PTS) risk indicators has demonstrated moderate to strong predictive accuracy on a proof-of-concept level[12–14]. However, the frequent lack of external validation in the literature obscures the generalizability of model performance[31,32], ultimately hampering the implementation of such algorithms in clinical practice. Recognizing the high clinical need, the US National Institute of Mental Health has funded a large multi-site consortium that will start to collect data from independent sites[33] suitable to evaluate the reliability of emerging predictive models of PTS course. Independently of this effort, large hospital systems are actively working to identify novel methods that can be integrated into the standard of care to improve patient outcomes and decrease long-term costs to the hospital system[34]. Together, there is an indication of both the necessary research and clinical interest in the development and deployment of data-driven approaches to predict the clinical risk of psychopathology in the context of ED healthcare.

We set out to develop and test the prediction of PTSD symptom development in a reproducible way across independent samples. At two independent sites (Supplementary Figs. 1 and 2), ED patients who reported experience of a traumatic event according to trauma criterion A[35] were enrolled in a prospective longitudinal cohort.

[1]Department of Psychiatry, New York University Grossman School of Medicine, New York, NY, USA. [2]Vagelos School of Physicians and Surgeons, Department of Emergency Medicine, Columbia University Irving Medical Center, New York, NY, USA. [3]Data Science Institute, Columbia University, New York, NY, USA. [4]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA. [5]Yerkes National Primate Research Center, Atlanta, GA, USA. [6]Ronald O. Perelman Department of Emergency Medicine, New York University Grossman School of Medicine, New York, NY, USA. [7]Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI, USA. [8]Department of Counseling and Clinical Psychology, Teachers College, Columbia University, New York, NY, USA. [9]Center for Alcohol Use Disorder and PTSD, New York University Grossman School of Medicine, New York, NY, USA. [10]Dell Medical School, Department of Psychiatry, University of Texas at Austin, Austin, TX, USA. [11]Institute for Early Life Adversity Research, University of Texas at Austin, Austin, TX, USA. [12]McLean Hospital, Harvard Medical School, Boston, MA, USA. [13]AiCure LLC, New York, NY, USA. ✉e-mail: ks3796@cumc.columbia.edu

Table 1 | Inclusion and exclusion criteria

	Model development sample (Grady Memorial)	External validation sample (Bellevue Hospital)
Inclusion criteria	•→Adults 18–65 years of age •→Evidence of acute traumatic event exposure defined by DSM-4 PTSD criterion A •→Fluency in English •→have blood obtained by ED care staff	•→Adults 18–70 years of age •→Evidence of acute traumatic event exposure defined by DSM-5 PTSD criterion A •→Fluency in English, Spanish or Mandarin •→Living in New York tri-state area
Exclusion criteria	•→Evidence of traumatic brain injury indicated by a Glasgow Coma Scale score <15 •→Current or past history of mania, schizophrenia, other psychoses •→Prominent suicidal ideation in the last month •→Intoxication •→Severe pain •→Active labor •→Respiratory distress •→Admission to an intensive care unit, admission or surgery, medical instability or hemodynamic compromise. •→No phone or home address •→Evidence of an inability to understand study procedures, risks or otherwise unable to give informed consent	•→Evidence of traumatic brain injury indicated by a Glasgow Coma Scale score <13 •→Evidence of present or past psychotic symptoms •→Evidence of ongoing traumatic exposure (for example, domestic violence) •→Evidence of homicidality/suicidality •→Adults in police custody or Department of Correction patients •→Admission to an intensive care unit, admission or surgery, medical instability or hemodynamic •→No phone or home address •→Evidence of an inability to understand study procedures, risks or otherwise unable to give informed consent

Detailed sample characteristics are presented in Supplementary Tables 1–3 and Table 1.

We first built a cross-validated predictive algorithm of non-remitting PTSD symptom progression over 12 months following ED discharge based on a multi-layer ensemble machine-learning approach in a prospective longitudinal cohort ($n=377$; Grady Memorial Hospital, Atlanta) of ED patients who had experienced a traumatic event (Supplementary Fig. 1). Second, we externally validated the algorithm in an independent prospective longitudinal cohort ($n=377$; Bellevue Hospital Center, New York City) of ED patients admitted after trauma (Supplementary Fig. 2). Additionally, we determined the predictive accuracy of distinct symptom trajectories beyond non-remitting PTSD symptoms ('one-versus-rest' classification). Finally, we tested the prediction of provisional PTSD diagnosis at 12 months after ED admission (Table 2). Extended Data Fig. 1 shows a schematic overview of the study design.

Heterogeneity in the symptom progression over 12 months following ED discharge was statistically described using latent growth mixture modeling (LGMM). Two unconditional LGMM models were independently fitted for the two longitudinal cohorts to obtain PTS symptom trajectories as outcomes that are statistically independent of each other to guarantee the integrity of the external validation procedure (Supplementary Tables 4 and 5). The symptom trajectories are qualitatively described as 'non-remitting symptoms', 'recovery', 'worsening symptoms' (external validation cohort only) or 'resilience' (Supplementary Fig. 3). The non-remitting versus the resilient trajectory was used as one outcome for the algorithm. An additional outcome was the discrimination between the non-remitting symptom trajectory versus all other trajectories (Table 3).

After the preprocessing of the data, 70 variables were included at both sites as candidate predictors (Supplementary Table 2). With these data, the algorithm achieved high discriminatory accuracy (area under the curve (AUC) = 0.84) to classify ED patients on a non-remitting symptom trajectory versus resilient ED patients in the model development sample (Table 3). The high discriminatory performance was reproduced (AUC = 0.83) on the external validation dataset (Table 3).

To benchmark the predictive model, we compared the algorithm with a scikit-learn DummyClassifier (Supplementary Fig. 4) and a logistic regression (external validation set; weighted-average precision = 0.76, recall = 0.69, f1 score = 0.63, AUC = 0.62; Supplementary Table 6 and Supplementary Fig. 5). Moreover, we applied the algorithm on a subset of candidate predictor variables using only biomarker data from EMRs that are routinely collectible in the ED (Supplementary Table 7). As might be expected, the performance decreased but demonstrates that the prediction is not exclusively driven by self-reports and that the accuracy of the prediction based on EMR alone is informative and in the range of 'fair' discriminatory ability of the classifier on the external validation dataset (AUC = 0.72). We compared these results to only using specific psychometric data that were collected for research purposes in the ED (Supplementary Table 8), consisting of the Immediate Stress Reaction Checklist (ISRC)[36] and Peritraumatic Dissociative Experiences Questionnaire (PDEQ)[37]. The algorithm yielded accurate and reproducible results (Table 3) when applied on a subset of the candidate predictors that consists of routinely collected EMR data plus four items of the ISRC (item 6, item 7, item 26 and item 27). The receiver operating characteristic (ROC) curve shows the specificity and the sensitivity of the binary predictions on the model development dataset (AUC = 0.85) and the external validation dataset (AUC = 0.86) and is accompanied by a calibration plot for the predicted probabilities (Extended Data Fig. 2a,b).

In addition, we proceeded to test how well the algorithm discriminates the non-remitting trajectory from all other PTS symptom trajectories, not only from the resilient one (Supplementary Table 9). While the results are less stable, the algorithm achieved very high discriminatory accuracy on the model development dataset

Table 2 | Prediction of provisional PTSD diagnosis at 12 months after ED admission

	Provisional PTSD diagnosis		
	1 month	6 months	12 months
Model development sample	32.3% (109 of 337)	19.3% (53 of 274)	15.5% (40 of 258)
External validation sample	30.4% (35 of 115)	30.7% (31 of 101)	22% (17 of 75)

The number of ED patients in the sample who screened positive for provisional PTSD diagnosis (PTSD Checklist for DSM-5 (PCL-5) score ≥33 or modified PTSD Symptom Scale (mPSS) score ≥21) and number of ED patients for which available scores were available. Percentages and frequencies in brackets are based on complete case analysis of fully available PCL-5 or mPSS scores at the respective time of measurement.

Table 3 | Discrimination between the non-remitting symptom trajectory versus all other trajectories

	EMR data plus ISRC and PDEQ (non-remitting versus resilient)				
	Precision	Recall	f1 score	ROC-AUC	Positive events[a]/total events[b]
Training set	0.86	0.84	0.84	0.84	27 of 164
Internal validation	0.83	0.64	0.69	0.70	14 of 89
External validation	0.86	0.85	0.85	0.83	38 of 93
	EMR data plus the four ISRC items[c] (non-remitting versus resilient)				
	Precision	Recall	f1 score	ROC-AUC	Positive events[a]/total events[b]
Training set	0.88	0.85	0.85	0.85	27 of 164
Internal validation	0.85	0.65	0.71	0.69	14 of 89
External validation	0.86	0.86	0.86	0.86	38 of 93
	EMR data plus ISRC and PDEQ non-remitting versus all other PTS trajectories				
	Precision	Recall	f1 score	ROC-AUC	Positive events[a]/total events[b]
Training set	0.87	0.84	0.84	0.96	35 of 320
Internal validation	0.82	0.68	0.74	0.59	6 of 57
External validation	0.83	0.75	0.78	0.78	38 of 221

[a]Positive events, LGMM class of ED patients with non-remitting PTS symptoms; [b]Total events, all participants included in the analysis (LGMM class of ED patients with non-remitting PTS symptoms + LGMM class of resilient ED patients); [c]ISRC item 6: "I felt like I was not there, like I was not part of what was going on,", ISRC item 7: "I felt confused,", ISRC item 26: "I get upset when something reminds me of what happened," and ISRC item 27: "I feel hyper or like I can't stay still." Weighted-average performance metrics of several experiments during predictive model development, validation on an internal test set (Grady Memorial Hospital) and external validation set (Bellevue Hospital Center). We explored the performance of the algorithm using EMR data along with two brief validated psychometric assessments in the ED (Supplementary Table 2) to predict the primary outcome of non-remitting stress symptoms versus resilience as indicated by LGMM and of non-remitting symptoms versus all other PTS trajectories.

(AUC = 0.96) and reasonably good performance (AUC = 0.78) on the independent external validation set (Table 3).

Finally, we determined the number of ED patients who screened positive for PTSD caseness at 12 months based on a cutoff for provisional PTSD diagnosis (Table 2). Across the combined samples, the 12-month prevalence of this provisional PTSD diagnosis was 17.12% (57 of 333). The algorithm predicted provisional PTSD diagnosis at 12 months with high discriminatory accuracy of AUC = 0.87 on the external validation dataset (Supplementary Fig. 4).

In summary, our results demonstrate the development and internal and external validation of a predictive algorithm with high discriminatory accuracy for PTS pathology. External validation of predictive models is the gold standard to evaluate the generalizability of the performance[31] but is, to our best knowledge[32], still lacking in the literature of PTSD prediction models. The externally validated results of the proposed prediction algorithm presented here are within the range of existing benchmark models without external validation and are encouraging for further research[12,13].

Out of all patients who are predicted to manifest non-remitting PTSD symptoms through 12 months, 90% were presenting such non-remitting symptoms. Only 5% of all patients who were 'resilient' through 12 months were falsely predicted to manifest non-remitting PTSD symptoms (Supplementary Table 10). Out of all patients who are predicted to have no or low PTSD symptoms through 12 months (resilient trajectory), 29% would develop non-remitting PTSD symptoms through 12 months. As it is often clinically more useful to predict non-remitting symptoms than resilience, the algorithm is promising to support early clinical screening for PTS long-term risk in the population of ED patients who experienced a traumatic event consistent with PTSD criterion A[35] (Supplementary Table 3) and who undergo blood sampling, as examined in this study.

To examine the key markers predicting PTS risk, we calculated SHAP (SHapley Additive exPlanation) values to rank the features in the order of importance for the prediction[38]. The rank order informs which feature values influence the prediction the most while accounting for the influence of all other feature values and controlling for the order of adding the features to the model[38]. Figure 1a,b shows the 20 most influential predictors when using all data and Fig. 1c,d when using only EMR biomarker data plus the four most predictive ISRC items. The variable importance for the prediction of non-remitting symptoms versus all other trajectories is presented in Supplementary Fig. 6. SHAP values can also visualize which variables mainly drive correct classifications or misclassification (Supplementary Fig. 7b,c). As the prediction of the algorithm is an achievement of all variables together, importance ranking (Fig. 1a–d and Supplementary Fig. 7a–d) should not be interpreted causally. Casual inferences would require randomized controlled trials, whereas this study was designed to test the reproducibility of the predictive accuracy on out-of-sample data. Nevertheless, the post hoc variable importance ranking may generate new hypotheses to design future prospective studies to determine the relevance of the predictor variables as potential risk factors or to guide research on prevention measures.

We demonstrated the development and internal and external validation of a predictive algorithm with high discriminatory accuracy for PTS pathology. The predictive model is based on the psychological stress response and biological data collected in the ED directly after a traumatic event. We achieved accurate out-of-sample prediction of PTS risk with good predictive performance in two independent clinical samples. External validation is the gold standard to evaluate the generalizability of a predictive model's performance but is still lacking in the literature of PTSD prediction models. The next step is to evaluate the net utility of the implementation of the algorithm as an automated clinical read-out available for ED clinicians at discharge planning. The nature of the predictor variables and the timing of their measurement make this goal amenable for further investigation and would have significant implications for both the early identification of PTS following trauma, as well as integration of acute care providers managing such patients in the near term. For the current algorithm to be used in healthcare, important additional steps are required to promote the interoperability and implementation of the predictive model into diverse medical infrastructures. We recommend that ED patients with suspected criterion

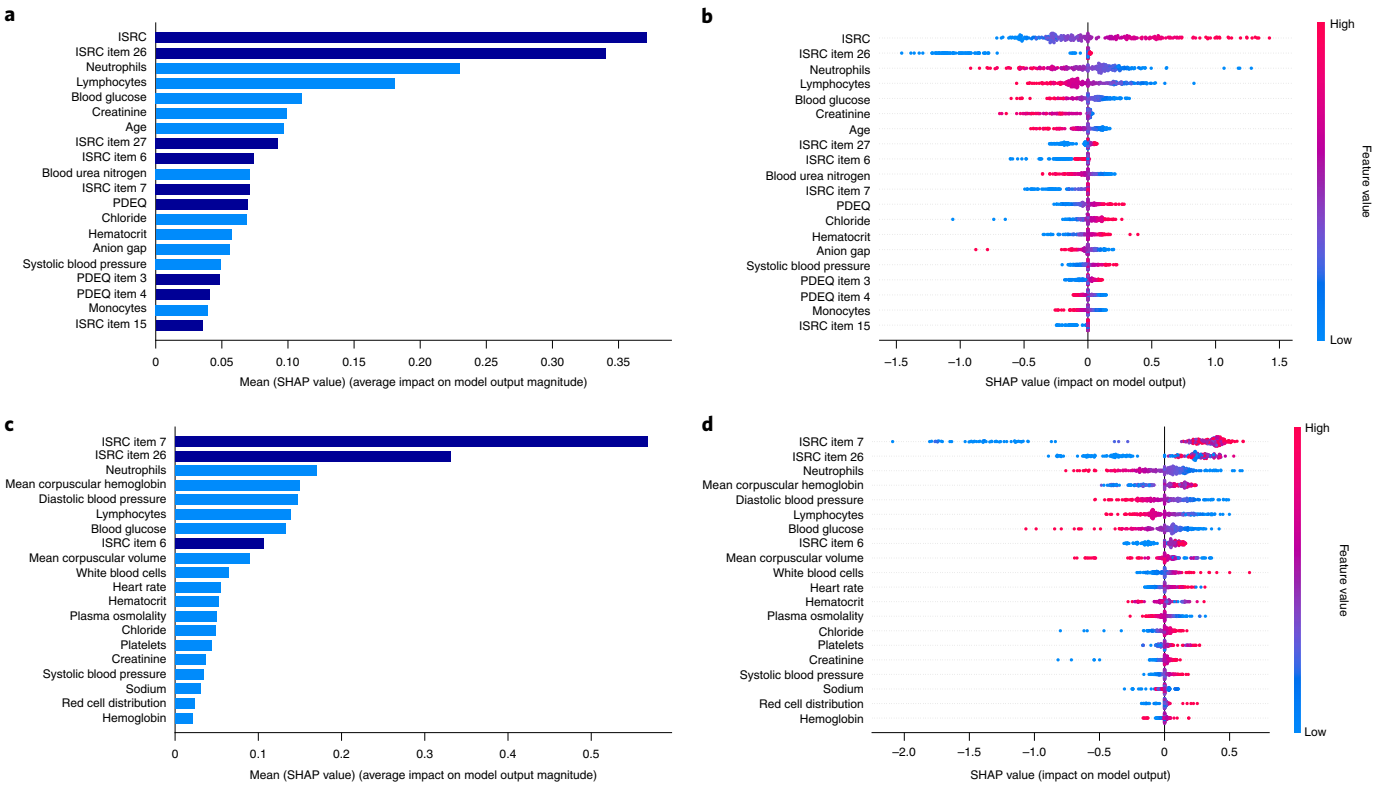


Fig. 1 | Variable importance for the training set. a, Variable importance using SHAP for the training set based on EMR data plus ISRC and PDEQ. The SHAP value is calculated for each feature by comparing what the model's prediction would be without the feature and with the feature in every possible order of adding the feature to the model. The bar plot shows the mean absolute SHAP value per feature. The larger the SHAP value, the more important the feature is to discriminate between the non-remitting and resilient trajectory. **b**, SHAP summary dot plot (for the same analysis as in **a**) displaying features that influence model predictions of positive outcome (non-remitting class) the most. The higher the SHAP value of a feature, the higher the log odds of a non-remitting PTSD trajectory. Features are first sorted by their global impact (*y* axis). For every individual in the sample, a dot represents the attribution value for each feature from low (blue) to high (red). The density of the plot shows that the ISRC is the most important predictor and that higher ISRC levels give rise to higher SHAP values (higher probability to be in the non-remitting class) displayed on the *x* axis. Chloride (which is the most important predictor in the external test set; Supplementary Fig. 7) shows that a higher score increases the likelihood (a log odds ratio) of being assigned to the non-remitting trajectory by the model (higher SHAP value). **c**, SHAP variable importance for the training set using EMR data plus the four most predictive ISRC items. **d**, SHAP summary dot plot for the respective analysis from **c**. Self-report measures are colored in dark blue in **a**–**c**.

A traumatic stress exposure are assessed with brief psychometric instruments, consistently with recommendations of embedded screenings for depression, suicidality and substance abuse[29]. In the proposed algorithm, the inclusion of four items of the ISRC considerably increased the accuracy of the prediction (Table 3).

A limitation to note is the use of PTSD self-report measures rather than gold-standard structured clinical interviews, such as the clinician-administered PTSD scale of the Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 (ref. [39]). Further, the algorithm evaluation should be extended beyond the defined inclusion and exclusion criteria (Table 1) to test the ecological validity for the broadest and most heterogeneous ED population at significant risk for PTSD, including trauma types other than those in our sample (Supplementary Table 3), such as the traumatic experience of cardiovascular events[40].

However, in urban public hospitals, patients may utilize the ED for primary care related to routine illness or injury without meeting DSM criteria A. Since every screening tool will yield a certain false positive rate, it is best to avoid unnecessary screening of patients who can by definition not develop PTSD. Accordingly, the intended target population of the algorithm are ED patients after suspected trauma.

Similarly, the current algorithm was built using patients for whom blood was drawn (Table 1). This limits the generalizability to patients with more severe illness or injury, requiring clinical blood testing. However, patients who experience less severe injuries that are psychologically perceived to be life-threatening are also at risk for PTS.

Ultimately, such algorithms will only become clinically actionable when directly integrated into care, thereby informing translational research with clinicians' long-term experience as a stream of feedback on model utility. Ecological validity in diverse clinical settings requires continued evaluation and vastly more independent data. The current effort provides a starting point for new technologies to be built into the existing healthcare systems for actionable prediction of PTS risk after trauma.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-020-0951-z.

Received: 16 July 2019; Accepted: 22 May 2020;
Published online: 06 July 2020

References

1. DiMaggio, C. J., Avraham, J. B., Lee, D. C., Frangos, S. G. & Wall, S. P. The epidemiology of emergency department trauma discharges in the United States. *Acad. Emerg. Med.* **24**, 1244–1256 (2017).
2. Wiseman, T. A., Curtis, K., Lam, M. & Foster, K. Incidence of depression, anxiety and stress following traumatic injury: a longitudinal study. *Scand. J. Trauma Resusc. Emerg. Med.* **23**, 29 (2015).
3. Sullivan, E. et al. The association between posttraumatic stress symptoms, depression, and length of hospital stay following traumatic injury. *Gen. Hosp. Psychiatry* **46**, 49–54 (2017).
4. Fakhry, S. M. et al. Continuing trauma: the unmet needs of trauma patients in the postacute care setting. *Am. Surgeon* **83**, 1308–1314 (2017).
5. Shalev, A. Y. et al. Estimating the risk of PTSD in recent trauma survivors: results of the International Consortium to Predict PTSD (ICPP). *World Psychiatry* **18**, 77–87 (2019).
6. Rothbaum, B. O. et al. Early intervention following trauma may mitigate genetic risk for PTSD in civilians: a pilot prospective emergency department study. *J. Clin. Psychiat.* **75**, 1380 (2014).
7. Galatzer-Levy, I. R. et al. Early PTSD symptom trajectories: persistence, recovery, and response to treatment: results from the Jerusalem Trauma Outreach and Prevention Study (J-TOPS). *PLoS ONE* **8**, e70084 (2013).
8. Shalev, A. Y. et al. Long-term outcome of early interventions to prevent posttraumatic stress disorder. *J. Clin. Psychiat.* **77**, e580–e587 (2016).
9. Shalev, A. Y. et al. Prevention of posttraumatic stress disorder by early treatment: results from the Jerusalem Trauma Outreach And Prevention study. *Arch. Gen. Psychiatry* **69**, 166–176 (2012).
10. Roberts, N. P., Kitchiner, N. J., Kenardy, J. & Bisson, J. I. Early psychological interventions to treat acute traumatic stress symptoms. *Cochrane Database Syst. Rev.* **17**, CD007944 (2010).
11. Shalev, A. Y. & Barbano, A. C. PTSD: risk assessment and early management. *Psychiatr. Ann.* **49**, 299–306 (2019).
12. Galatzer-Levy, I. R., Karstoft, K. I., Statnikov, A. & Shalev, A. Y. Quantitative forecasting of PTSD from early trauma responses: a machine-learning application. *J. Psychiatr. Res.* **59**, 68–76 (2014).
13. Galatzer-Levy, I. R., Ma, S., Statnikov, A., Yehuda, R. & Shalev, A. Y. Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD. *Transl. Psychiatry* **7**, e1070 (2017).
14. Karstoft, K.-I., Galatzer-Levy, I. R., Statnikov, A., Li, Z. & Shalev, A. Y. Bridging a translational gap: using machine learning to improve the prediction of PTSD. *BMC Psychiatry* **15**, 30 (2015).
15. Shalev, A. Y. et al. A prospective study of heart rate response following trauma and the subsequent development of posttraumatic stress disorder. *Arch. Gen. Psychiatry* **55**, 553–559 (1998).
16. Yehuda, R., McFarlane, A. & Shalev, A. Predicting the development of posttraumatic stress disorder from the acute response to a traumatic event. *Biol. Psychiatry* **44**, 1305–1313 (1998).
17. Papini, S. et al. Ensemble machine-learning prediction of posttraumatic stress disorder screening status after emergency room hospitalization. *J. Anxiety Disord.* **60**, 35–42 (2018).
18. Ressler, K. J. Molecular signatures of stress and posttraumatic stress disorder: an overview. *Biol. Psychiatry* **83**, 792–794 (2018).
19. Hinrichs, R. et al. Increased skin conductance response in the immediate aftermath of trauma predicts PTSD risk. *Chronic Stress* **3**, 2470547019844441 (2019).
20. Heim, C., Schultebraucks, K., Marmar, C. R. & Nemeroff, C. B. in *Post-Traumatic Stress Disorder* (eds Nemeroff, C. B. & Marmar, C.) 331 (Oxford Univ. Press, 2018).
21. Morris, M. C., Hellman, N., Abelson, J. L. & Rao, U. Cortisol, heart rate, and blood pressure as early markers of PTSD risk: a systematic review and meta-analysis. *Clin. Psychol. Rev.* **49**, 79–91 (2016).
22. Van Zuiden, M. et al. Glucocorticoid receptor pathway components predict posttraumatic stress disorder symptom development: a prospective study. *Biol. Psychiatry* **71**, 309–316 (2012).
23. Schultebraucks, K. et al. Heightened biological stress response during exposure to a trauma film predicts an increase in intrusive memories. *J. Abnorm. Psychol.* **128**, 645 (2019).
24. Michopoulos, V. et al. Association of prospective risk for chronic PTSD symptoms with low TNF-α and IFN-γ concentrations in the immediate aftermath of trauma exposure. *Am. J. Psychiatry* **2019**, 19010039 (2019).
25. Mellon, S. H., Gautam, A., Hammamieh, R., Jett, M. & Wolkowitz, O. M. Metabolism, metabolomics, and inflammation in posttraumatic stress disorder. *Biol. Psychiatry* **83**, 866–875 (2018).
26. Michopoulos, V., Powers, A., Gillespie, C. F., Ressler, K. J. & Jovanovic, T. Inflammation in fear-and anxiety-based disorders: PTSD, GAD, and beyond. *Neuropsychopharmacology* **42**, 254 (2017).
27. Nugent, N. R., Christopher, N. C. & Delahanty, D. L. Emergency medical service and in-hospital vital signs as predictors of subsequent PTSD symptom severity in pediatric injury patients. *J. Child Psychol. Psychiatry* **47**, 919–926 (2006).
28. Shalev, A., Liberzon, I. & Marmar, C. Post-traumatic stress disorder. *New Engl. J. Med.* **376**, 2459–2469 (2017).
29. Love, J. & Zatzick, D. Screening and intervention for comorbid substance disorders, PTSD, depression, and suicide: a trauma center survey. *Psychiatr. Serv.* **65**, 918–923 (2014).
30. Vermetten, E., Zhohar, J. & Krugers, H. J. Pharmacotherapy in the aftermath of trauma; opportunities in the 'golden hours'. *Curr. Psychiatry Rep.* **16**, 455 (2014).
31. Altman, D. G., Vergouwe, Y., Royston, P. & Moons, K. G. Prognosis and prognostic research: validating a prognostic model. *Br. Med. J.* **338**, b605 (2009).
32. Schultebraucks, K. & Galatzer-Levy, I. R. Machine learning for prediction of posttraumatic stress and resilience following trauma: an overview of basic concepts and recent advances. *J. Trauma Stress* **32**, 215–225 (2019).
33. McLean, S. A. et al. The AURORA study: a longitudinal, multimodal library of brain biology and function after traumatic stress exposure. *Mol. Psychiatry* **25**, 283–296 (2020).
34. Horwitz, L. I., Kuznetsova, M. & Jones, S. A. Creating a learning health system through rapid-cycle, randomized testing. *New Engl. J. Med.* **381**, 1175 (2019).
35. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (American Psychiatric Publishing, 2013).
36. Fein, J. A., Kassam-Adams, N., Vu, T. & Datner, E. M. Emergency department evaluation of acute stress disorder symptoms in violently injured youths. *Ann. Emerg. Med.* **38**, 391–396 (2001).
37. Marmar, C. R, Weiss, D. S. & Metzler, T. J. The Peritraumatic Dissociative Experiences Questionnaire. in *Assessing Psychological Trauma and PTSD* 2nd edn (eds Wilson, J. P. & Kean, T. M.) 144–167 (Guilford Press, 2004).
38. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 4765–4774 (2017).
39. Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K. & Domino, J. L. The posttraumatic stress disorder checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. *J. Trauma. Stress* **28**, 489–498 (2015).
40. Schultebraucks, K., Wen, T., Kronish, I. M., Willey, J. & Chang, B. P. Post-traumatic stress disorder following acute stroke. *Curr. Emerg. Hosp. Med. Rep.* **8**, 1–8 (2020).

Methods

Participants. At two US sites, patients who were admitted to the ED of a level 1 trauma center after experiencing a traumatic event were enrolled into one of two separate longitudinal study cohorts.

The first sample ($n=377$) of ED patients was prospectively enrolled from 2012 to 2017 at the Marcus Trauma Center of the Grady Memorial Hospital and was used for model development (discovery sample). A second independent sample ($n=221$) of ED patients was prospectively enrolled from 2012 to 2016 at Bellevue Hospital Center and was used for external validation of the predictive model (validation sample).

The participants were approached based on information of the ED's 'White Board', trauma surgery discharge rounds or the team's rounding sheet. Potential eligible patients were contacted by the study personnel in the ED and our inclusion and exclusion criteria were assessed (Table 1). A flow chart details the patient flow for both samples (Supplementary Figs. 1 and 2).

Across both samples, all participants had experienced a traumatic event such as a life-threatening accident, assault or attack satisfying the DSM-5 trauma criterion A of PTSD. Further inclusion criteria across both sites were capacity to give informed consent, age between 18–65 years (Grady Memorial sample) or 18–70 years of age (Bellevue sample), US residency and fluent in English, Spanish and Mandarin (Bellevue sample) or fluency in speaking English (Grady Memorial sample). Exclusion criteria for the Bellevue sample were risk for ongoing traumatic exposure (such as domestic violence), evidence of homicidal/suicidal behavior, present or past psychotic symptoms, custody of police or Department of Correction, open head injury, survivors in coma or evidence of traumatic brain injury indicated by a Glasgow Coma Scale score <13 or no reliable access to email or telephone. Similarly, exclusion criteria for the Grady Memorial sample were current intoxication, suicidal ideation or suicide attempts during the previous 3 months, a history of schizophrenia, psychosis or mania, Glasgow Coma Scale score <15, respiratory distress or medical instability, for example, hemodynamic (Table 1). ED patients received treatment and diagnostic assessments as usual, which included blood draws.

The studies were approved by the ethics committee of New York University (Bellevue study) and the Emory Institutional Review Board and the Grady Hospital Research Oversight Committee (Grady Memorial study) to be in line with the most recent version of Declaration of Helsinki[41]. All participants signed informed consent.

Outcomes. The primary outcome of the predictive model was the classification of ED patients based on longitudinal symptom trajectories of non-remitting PTS symptoms versus resilience as identified by LGMM. As a secondary outcome, we classified those on a non-remitting trajectory versus all other trajectories. For model development, the outcome measure was the mPSS[42] prospectively collected at 1, 3, 6 and 12 months after ED discharge at Grady Memorial Hospital. The mPSS is a 17 item self-report psychometric instrument to assess PTSD symptom severity, with a cutoff of ≥21 screening positive for provisional PTSD diagnosis[42,43]. The specifier 'provisional' can be used "when there is a strong presumption that the full criteria will ultimately be met for a disorder but not enough information is available to make a firm diagnosis" at this moment in time[35].

The outcome measure used for external validation of the predictive model was PTS symptom severity and was measured with PCL-5 (ref. [44]). Data were prospectively collected at ED admission, within 7 d thereafter (phone screen interview) and at 1, 3, 6 and 12 months after ED admission at Bellevue Hospital. The PCL-5 is a 20-item self-report measure of PTSD and ranges from 0 to 80, with a cutoff score of 33 for screening positive for a provisional PTSD diagnosis[44].

As an additional outcome, we predicted provisional PTSD diagnosis at 12 months after ED admission based on an mPSS score ≥21 (model development)[42] and a PCL-5 score ≥33 (model validation)[44]. The PCL-5 and mPSS are well-established and cost-effective alternatives to structured clinical interviews for research purposes as the mPSS[45] and PCL-5 show a "good diagnostic utility for predicting a CAPS-5 PTSD diagnosis" and "good structural validity and sensitivity to clinical change comparable to that of a structured interview"[46]. Both measures have good reliability, convergent, concurrent, discriminant and structural validity[45,46].

Candidate predictor variables. Data collection was performed prospectively at the two sites, independently from each other, before the outcome of interest was determined. Both longitudinal cohort studies were designed to examine factors associated with PTSD in ED trauma survivors and the collected data were based on previous theory on biological risk factors for PTSD[18–21,24,25] such as psychophysiological stress response and threat perception[18,20–23,25], psychophysiological arousal[15,19], immune and inflammatory markers[24–26], as well as psychosocial determinants of PTS[28]. For the proposed prediction model, a subset of all collected data was selected as candidate predictor variables based on the consideration that data should be readily available and routinely collected in the ED and thus, at least partially, available at both sites. Additionally, we included psychometric instruments that measure the acute stress response in a brief and noninvasive way. We considered 83 variables as candidate predictors that were available at both sites, 13 variables were removed due to missing values of more than 45%. In total, 70 variables were included for building and validating the model (Supplementary Table 2).

Electronic medical records. Data routinely collectible from EMRs comprised demographic variables, including patient age, sex, race and body mass index. Furthermore, it contained biomarkers, (vital signs of heart rate and blood pressure at ED arrival as well as data from blood draws such as hemogram and metabolomics analyses). Moreover, data collection contained self-reported pain level ratings and whether opiates were administered in the ED. Furthermore, the information collected includes whether a computed tomography or magnetic resonance imaging scan was indicated, the evaluation of loss of consciousness in the case of head injury and a record of coma.

Psychometric measures. We collected data about the participants' acute responses to the traumatic event using the ISRC[36] and dissociative experiences during the traumatic event using PDEQ[37].

Statistical analysis. For outcome definition, LGMM in Mplus v.7 (ref. [47]) identified heterogeneous trajectories based on PTS symptoms through 12 months after trauma. Individuals were assigned to trajectories based on their most likely class membership. For identifying the best-fitting model we followed recommendations from the literature[48] (Supplementary Information). For the model development sample, mPSS scores of $n=377$ ED patients and for the external validation sample, PCL-5 scores of $n=221$ ED patients, were included in the LGMM. Trajectory analyses were run on the sample of participants who had at least one follow-up assessment. The best-fitting LGMM model in the discovery sample was a three-class solution with a linear slope (Akaike Information Criteria (AIC) = 8,111.160, Bayesian Information Criteria (BIC) = 8,158.347, sample-size adjusted Bayesian Information Criterion (SSBI) = 8,120.274, Vuong–Lo–Mendell–Rubin Likelihood test (VRLT) = 0.0036, Lo–Mendell–Rubin Adjusted LRT test (LRT) = 0.0046) with high entropy of 83%. The three LGMM classes were qualitatively described as non-remitting, recovery and resilient symptom development over time. The most common symptom trajectory was resilient ($n=212$; 56.23%) followed by recovery ($n=124$; 32.89%). The non-remitting symptom trajectory was the least common response of ED patients ($n=41$; 10.88%).

The best-fitting model in the external validation sample was a four-class solution with a quadratic slope (AIC = 7,915.391, BIC = 7,983.354, SSBI = 7,919.973, VRLT = 0.0085, LRT = 0.0104 and entropy of 95.7%). One trajectory describes ED patients with non-remitting symptoms ($n=38$; 17.2%), one with ED patients whose symptoms recover over time ($n=90$; 40.72%), one was a trajectory of worsening symptoms ($n=38$; 17.2%) and one was a symptom trajectory of resilient ED patients ($n=55$, 24.89%). The LGMM model selection (Supplementary Tables 4 and 5) was guided by the Guidelines for Reporting on Latent Trajectory Studies checklist[48].

Model development. Data were preprocessed using the R package, caret[49] (Supplementary Information). An ensemble of machine-learning classification algorithms (Deep Super Learner)[50] in Python (scikit-learn 0.19.1)[51] was applied to predict the probability of ED patients belonging to the non-remitting symptom trajectory throughout 12 months. This algorithm is an ensemble of base learners. Each base learner is specialized for a specific task required for an accurate prediction of the outcome of interest. Random forest (RandomForestClassifier)[52], AdaBoost (AdaBoostClassifier)[53], logistic regression (LogisticRegression)[53] and support vector machines (SVCs)[53] were used as base learners and optimized to minimize log loss[50], an established performance measure in machine-learning classification models[54]. This algorithm was selected because of evidence of favorable performance in relatively small samples utilizing small features to sample ratios as in this study, especially compared to deep neural nets[50,55].

External model validation. Model generalizability was assessed in an independent sample of ED patients. Thus, the finalized predictive model developed on the sample of Grady Memorial Hospital was applied once and unaltered to the independent sample from Bellevue Hospital Center. Measures of the predictive model's performance on the external validation sample were precision, recall, f1 score and ROC-AUC, which are all established performance measures for machine-learning classification tasks[49,56]. All phases of model development, validation and reporting have been informed by respective guidelines and recommendations as applicable[31,48,57–59].

Predictor importance ranking. We report methods for explainable machine learning using SHAP feature importance ranking based on Shapley values[38]. SHAP values are useful for examining machine-learning predictions and to critically appraise which features the model mainly relies on to arrive at individual prediction outcomes. Kernel-based SHAP values were used to rank the variables for their ability to predict the non-remitting symptoms versus the resilient PTS trajectory (Fig. 1)[38]. This is an additive feature attribution method using kernel functions, enabling consistent and locally faithful explanation of feature importance[38,60].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All requests for raw and analyzed data and related materials, including programming code, will be reviewed by our legal departments (New York University Grossman School of Medicine and Emory University School of Medicine) to verify whether the request is subject to any intellectual property or confidentiality constraints. Any data and materials that can be shared will be released via a material transfer agreement for noncommercial research purposes. Request should be addressed to the corresponding author (K.S.) or the Principal Investigators of the two study sites (K.J.R. and I.R.G.-L.).

Code availability

The programming code is based on Scikit-learn (https://scikit-learn.org/stable/) and SHAP (https://github.com/slundberg/shap) and the core algorithm can be obtained from https://github.com/KSchultebraucks/DeepSuperLearner. Requests should be addressed to the corresponding author (K.S.).

References

41. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *J. Am. Med. Assoc.* **310**, 2191 (2013).
42. Foa, E. B., Riggs, D. S., Dancu, C. V. & Rothbaum, B. O. Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *J. Trauma. Stress* **6**, 459–473 (1993).
43. Falsetti, S. A., Resnick, H. S., Resick, P. A. & Kilpatrick, D. G. The modified PTSD symptom scale: a brief self-report measure of posttraumatic stress disorder. *Behav. Ther.* **16**, 161–162 (1993).
44. Weathers, F. W., et al. The PTSD checklist for DSM-5 (PCL-5). *Department of Verterans Affairs* http://www.ptsd.va.gov (2013).
45. Ruglass, L. M., Papini, S., Trub, L. & Hien, D. A. Psychometric properties of the modified posttraumatic stress disorder symptom scale among women with posttraumatic stress disorder and substance use disorders receiving outpatient group treatments. *J. Trauma. Stress Disord. Treat.* https://doi.org/10.4172/2324-8947.1000139 (2014).
46. Weathers, F. W. Redefining posttraumatic stress disorder for DSM-5. *Curr. Opin. Psychol.* **14**, 122–126 (2017).
47. Muthén, L. K. & Muthén, B. O. *Mplus User's Guide: Statistical Analysis with Latent Variables* (Muthén & Muthén, 1998–2017).
48. van de Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S. & Vermunt, J. K. The GRoLTS-checklist: guidelines for reporting on latent trajectory studies. *Struct. Equ. Modeling* **24**, 451–467 (2017).
49. Kuhn, M. & Johnson, K. *Applied Predictive Modeling* (Springer, 2013).
50. Young, S., Abdou, T. & Bener, A. Deep Super Learner: a deep ensemble for classification problems. in *Advances in Artificial Intelligence. Canadian Conference on Artificial Intelligence (Canadian AI 2018)* (eds Bagheri, E. & Cheung, J. C. K.) 84–95 (2018).
51. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
52. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
53. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning*, (Springer, 2001).
54. Ferri, C., Hernández-Orallo, J. & Modroiu, R. An experimental comparison of performance measures for classification. *Pattern Recognit. Lett.* **30**, 27–38 (2009).
55. Zhou, Z.-H. & Feng, J. Deep forest. *Nat. Sci. Rev.* **6**, 74–86 (2018).
56. Fawcett, T. ROC graphs: notes and practical considerations for researchers. *Mach. Learn.* **31**, 1–38 (2004).
57. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med.* **13**, 1 (2015).
58. Luo, W. et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J. Med. Internet Res.* **18**, e323 (2016).
59. Moons, K. G. M., Royston, P., Vergouwe, Y., Grobbee, D. E. & Altman, D. Prognosis and prognostic research: what, why, and how? *Br. Med. J.* **338**, b605 (2009).
60. Štrumbelj, E. & Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* **41**, 647–665 (2014).

Acknowledgements

K.S. was supported by the German Research Foundation (SCHU 3259/1–1). The study was also supported by K01MH102415 (I.R.G.-L.), R01MH094759 (C.B.N.) and R01MH094757 (K.J.R.).

Author contributions

I.R.G.-L., K.J.R, C.B.N., A.Y.S., C.R.M., B.O.R., V.M., T.J. and K.S. substantially contributed to the design of the study and developed the study concept. S.-M.S., I.R.G.-L., V.M., J.S.S., J.L.M.-K., C.R.G. were involved in the data collection process. K.S. developed the data analytical plan and performed data analysis. G.A.B. and I.R.G.-L. provided supervision. K.S. wrote the first draft of the manuscript and all co-authors reviewed and revised the manuscript critically for important intellectual content. All co-authors approved the version of the manuscript to be published.

Competing interests

B.O.R. has funding from Wounded Warrior Project, Department of Defense Clinical Trial Grant No.W81XWH-10-1-1045, National Institute of Mental Health grant no. 1R01MH094757-01 and McCormick Foundation. B.O.R. also receives royalties from Oxford University Press, Guilford, APPI and Emory University and received advisory board payments from Genentech, Jazz Pharmaceuticals, Sophren, Nobilis Therapeutics, Neuronetics and Aptinyx. C.R.M. serves on the scientific advisory board and has equity in Receptor Life Sciences. He also serves on the PTSD advisory board for Otsuka Pharmaceutical. He receives support from the National Institute on Alcohol Abuse and Alcoholism, National Institute of Mental Health, Department of Defense, US Army Congressionally Directed Medical Research Program, the Steven & Alexander Cohen Foundation, Cohen Veterans Bioscience, Cohen Veterans Network, Home Depot Foundation, McCormick Foundation, Robin Hood Foundation and the City of New York. C.B.N. discloses the following: research/grants from National Institutes of Health and Stanley Medical Research Institute; consulting (last three years) at Xhale, Takeda, Taisho Pharmaceutical Inc., Bracket (Clintara), Fortress Biotech, Sunovion Pharmaceuticals Inc., Sumitomo Dainippon Pharma, Janssen Research & Development LLC, Magstim, Inc., Navitor Pharmaceuticals, Inc., TC MSO, Inc. and Intra-Cellular Therapies, Inc.; stockholder of Xhale, Celgene, Seattle Genetics, Abbvie, OPKO Health, Inc., Antares, BI Gen Holdings, Inc., Corcept Therapeutics Pharmaceuticals Company, TC MSO, Inc. and Trends in Pharma Development, LLC; scientific advisory boards of American Foundation for Suicide Prevention (AFSP), Brain and Behavior Research Foundation, Xhale, Anxiety Disorders Association of America (ADAA), Skyland Trail, Bracket (Clintara) and Laureate Institute for Brain Research Inc.; board of directors of AFSP, Gratitude America, ADAA and Xhale Smart, Inc.; income sources or equity of USD$10,000 or more from American Psychiatric Publishing, Xhale, Bracket (Clintara), CME Outfitters, Takeda, Intra-Cellular Therapies, Inc., Magstim and EMA Wellness; patents for the method and devices for transdermal delivery of lithium (US 6,375,990B1), the method of assessing antidepressant drug therapy via transport inhibition of monoamine neurotransmitters by ex vivo assay (US 7,148,027B2) and compounds, compositions, methods of synthesis and methods of treatment (CRF receptor binding ligand) (US 8,551,996 B2). K.J.R. performs consulting for Janssen, Verily, Alkermes and Biogen, Inc. on matters unrelated to this manuscript. I.R.G.-L. receives salary and stock options from AiCure. All other authors declare no competing interests.

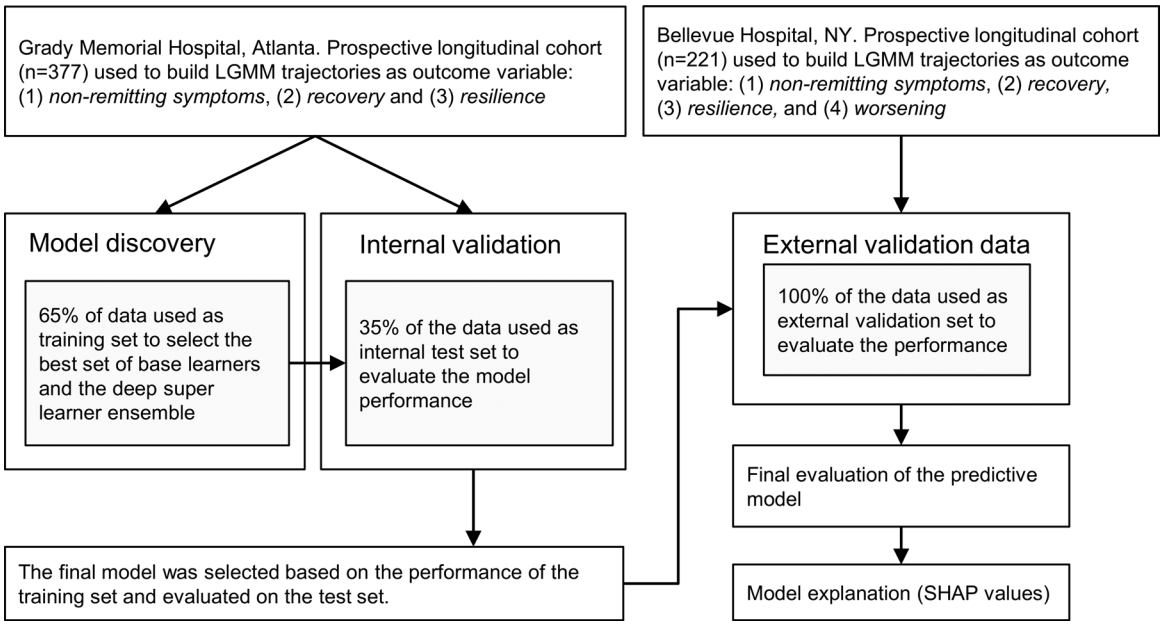
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-020-0951-z.

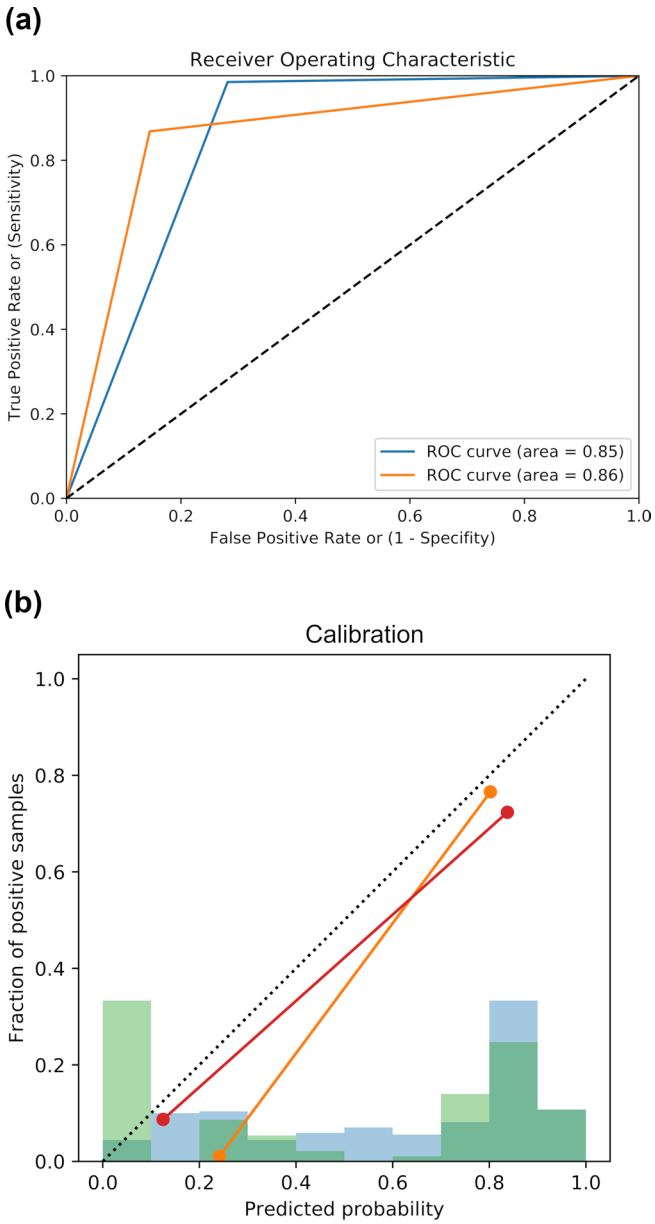
Supplementary information is available for this paper at https://doi.org/10.1038/s41591-020-0951-z.

Correspondence and requests for materials should be addressed to K.S.

Peer review information Kate Gao was the primary editor on this article, and managed its editorial process and peer review in collaboration with the rest of the editorial team.



Extended Data Fig. 1 | Schematic overview of the study design. Displayed are the basic steps of the model development and model validation.



Extended Data Fig. 2 | Predictive performance in terms of discrimination and calibration. In panel **a**, the ROC curve shows the specificity and the sensitivity of the predictions on the training set (blue line) and the external validation set (orange line) and is accompanied by a calibration plot for the predicted probabilities on the training set (orange line and blue bars) and the external validation set (red line and green bars) in panel **b**. The bars in the calibration plot in panel (**b**) displays the predicted probabilities in 10 bins [0, 10%], (10%, 20%],…, (90%, 100%], whereas the lines visualize the predicted probabilities in two bins (low vs. high probability).

nature medicine

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41591-020-0951-z

In the format provided by the authors and unedited.

A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor

Katharina Schultebraucks[1,2,3]✉, Arieh Y. Shalev[1], Vasiliki Michopoulos[4,5], Corita R. Grudzen[6], Soo-Min Shin[6], Jennifer S. Stevens[4], Jessica L. Maples-Keller[4], Tanja Jovanovic[7], George A. Bonanno[8], Barbara O. Rothbaum[4], Charles R. Marmar[1,9], Charles B. Nemeroff[10,11], Kerry J. Ressler[4,12] and Isaac R. Galatzer-Levy[1,13]

[1]Department of Psychiatry, New York University Grossman School of Medicine, New York, NY, USA. [2]Vagelos School of Physicians and Surgeons, Department of Emergency Medicine, Columbia University Irving Medical Center, New York, NY, USA. [3]Data Science Institute, Columbia University, New York, NY, USA. [4]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA. [5]Yerkes National Primate Research Center, Atlanta, GA, USA. [6]Ronald O. Perelman Department of Emergency Medicine, New York University Grossman School of Medicine, New York, NY, USA. [7]Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI, USA. [8]Department of Counseling and Clinical Psychology, Teachers College, Columbia University, New York, NY, USA. [9]Center for Alcohol Use Disorder and PTSD, New York University Grossman School of Medicine, New York, NY, USA. [10]Dell Medical School, Department of Psychiatry, University of Texas at Austin, Austin, TX, USA. [11]Institute for Early Life Adversity Research, University of Texas at Austin, Austin, TX, USA. [12]McLean Hospital, Harvard Medical School, Boston, MA, USA. [13]AiCure LLC, New York, NY, USA. ✉e-mail: ks3796@cumc.columbia.edu

Supplementary Information

Outcome definition

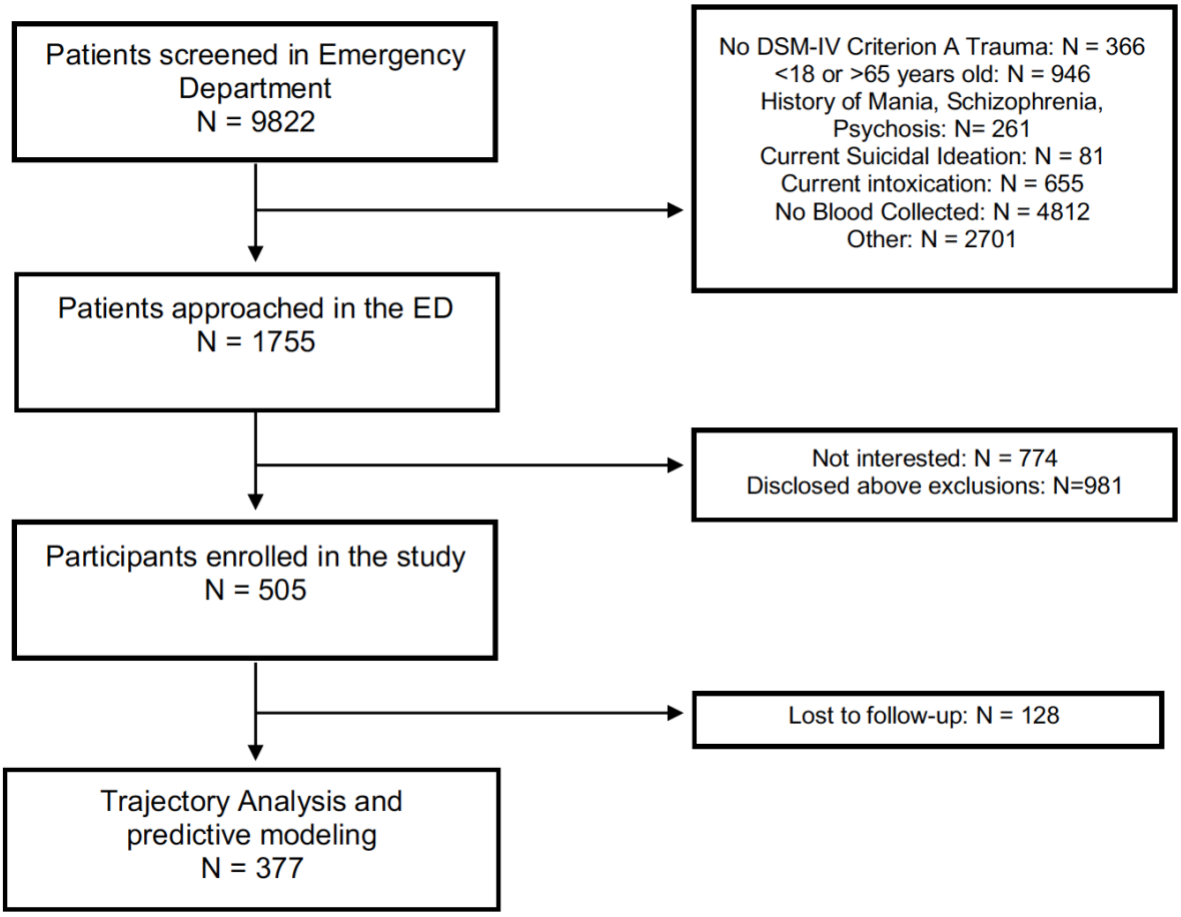
As recommended in guidelines for LGMM[1,2], the best-fitting model was selected through a nested model evaluation based on multiple criteria. These criteria are entropy, reduction in Information Criteria, i.e. Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), sample-size adjusted Bayesian Information Criterion (SSBI), and significance indicated by the Vuong-Lo-Mendell-Rubin Likelihood test (VRLT) and the Lo-Mendell-Rubin Adjusted LRT test (LRT) along with parsimony and interpretability (see supplementary Table 4 and 5 for model fits). To identify the best fitting trajectory shape we examined a linear and quadratic slope.

Model development

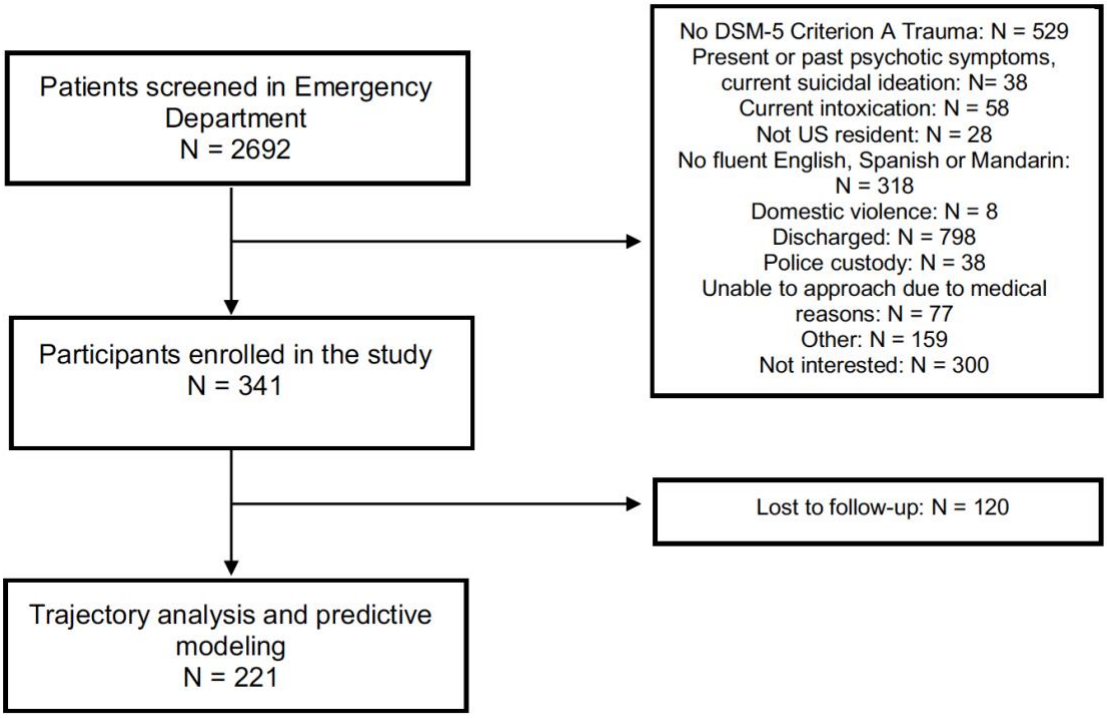
All categorical variables were dummy coded into binary numerical values ("one-hot encoding") and all numeric variables were normalized to the range [0;1] using R package *caret*[3,4]. All candidate predictor variables were manually curated to detect threats to data integrity such as typos (e.g. factor names) or inconsistent coding (e.g. missing values) and to align variable names and coding of identical variables across both samples. Variables not available at both sites or variables with more than 45% missing values were removed before bootstrap aggregation (bagged) tree imputation[3,5]. From the 83 variables available at both sites, 13 variables were excluded from analysis due to missing values of more than 45% of the data. 70 variables were included in the final model (Supplementary Table 2**)**.

Afterward, a *deep super learner* is built by calculating the optimally weighted composite as the average across the three individual base learners.[6,7] To account for the threat of "overfitting", k-fold cross-validation with 3 folds and 10 repeats was applied for internal validation during model development.[6,7]

To deal with the imbalance in the numbers of ED patients per LGMM class, minority oversampling of data points was performed using the SMOTE algorithm (only for model development), which is a common and effective method in machine learning to deal with unbalanced data of rare events.[8] The performance of the external validation dataset is reported "as is" without using over- or undersampling.



Supplementary Fig. 1. Flow diagram – visualizes the inclusion and exclusion of participants in the model development sample (Grady Memorial Hospital Atlanta)



Supplementary Fig. 2. Flow diagram – visualizes the inclusion and exclusion of participants in the external validation sample (Bellevue Hospital NY)

Supplementary Table 1. Sample characteristics of relevant differences between the sample used for the development of the predictive model and the sample used for external validation.

	Model development sample (N = 377)	Model validation sample (N = 221)	Descriptive statistics
Age (mean +/-SD)	36.05 (12.87)	36.69 (13.46)	t(595) = .58, p=.56
Gender (% Female)	47.1	37.1	$X_2(1) = 5.63$, p = .02
ISRC (mean +/-SD)	24.5 (10.91)	23.58 (10.26)	t(523) = -.97, p=.33
PDEQ (mean +/-SD)	23.79 (9.30)	24.96 (10.26)	t(320) = 1.05, p=.30
Hispanic (%)	3.7	26.1	$X_2(1) = 65.04$, p <.001

Supplementary Table 2. Descriptive statistics of the 70 features included in the final model. Shown are the routinely collectible data extracted from the electronic medical records (vital signs, biomarkers from blood draws, information about potential head injury, medication treatment, and demographic variables). Moreover, the sums cores and items of the Immediate Stress Reaction Checklist and Peritraumatic Dissociative Experiences Questionnaire are shown. The final prediction model presented in the main text (Results section) also included the items of these two psychometric assessments.

Data type	Feature group	Feature name	Grady Memorial Study (N=377)	Bellevue Study (N=221)
Extracted from electronic medical records	Vital Signs	Systolic Blood pressure (mean +/-SD)	139.27 (22.36)	135.62 (23.04)
		Diastolic Blood pressure (mean +/-SD)	84.87 (16.74)	81.91 (15.80)
		Heart rate (mean +/-SD)	84.51 (17.69)	80.45 (16.41)
	Biomarkers markers from blood draw	White Blood Count (10^9/L) (mean +/-SD)	8.40 (6.14)	9.95(3.93)
		Red Blood Count (10^12/L) (mean +/-SD)	5.06 (5.13)	4.71 (.60)
		Neutrophils (percentage) (mean +/-SD)	62.02 (14.43)	58.75 (15.14)
		Lymphocytes (percentage) (mean +/-SD)	25.95 (11.78)	29.28 (12.08)
		Monocytes (percentage) (mean +/-SD)	7.80 (3.30)	6.84 (2.21)
		Mean Corpuscular Hemoglobin (g/dL) (mean +/-SD)	33.10 (2.36)	33.21 (.75)
		Red Blood Cell Width (percentage) (mean +/-SD)	15.79 (18.97)	13.10 (2.31)
		Hemoglobin (g/dL) (mean +/-SD)	13.89 (3.80)	14.10 (1.66)
		Hematocrit (percentage) (mean +/-SD)	40.98 (6.80)	42.45 (4.87)
		Anion Gap (mean +/-SD)	8.96 (9.19)	10.28 (5.24)
		Creatinine (mg/dL) (mean +/-SD)	.99 (.61)	.95 (.26)
		Sodium (mmol/L) (mean +/-SD)	136.09 (14.28)	139.57 (3.00)
		Calcium (mg/dL) (mean +/-SD)	9.31 (.50)	9.25 (.64)
		Potassium (mmol/L) (mean +/-SD)	5.19 (10.63)	4.13 (.72)
		Glucose (mg/dL) (mean +/-SD)	114.21 (74.62)	124.56 (46.68)
		Carbon dioxide (mmol/L) (mean +/-SD)	25.04 (3.34)	24.45 (3.91)
		Osmolality (mosm/L) (mean +/-SD)	273.00 (29.31)	280.07 (6.18)
		Urea Nitrogen (mg/dL) (mean +/-SD)	12.81 (5.23)	13.01 (4.42)
		Mean Corpuscular Volume (f/L) (mean +/-SD)	87.14 (8.42)	90.31 (5.16)
		Platelet Count (10^9/L) (mean +/-SD)	234.61 (79.55)	242.17 (66.76)

		Chloride (mmol/L) (mean +/-SD)	103.11 (10.74)	105.11 (3.27)
	Potential Head injury	Loss of consciousness (% of those who lost consciousness)	18.7%	34.5%
	Medication	Opiate (% of those who received Opiate)	34.8%	59.4%
	Demographic variables	Gender (% Female)	47.1%	37.1%
		Age (mean +/-SD)	36.05 (12.87)	36.69 (13.46)
Actively collected psychometric assessment to measure the experienced stress reactivity to the traumatic event		Immediate Stress Reaction Checklist - ISRC (sum score) (mean +/-SD)	24.5 (10.91)	23.58 (10.26)
		ISRC item 1: “My mind went blank” (mean +/-SD)	1.03 (.88)	1.09 (.87)
		ISRC item 2: “I did things that I did not even know I was doing.” (mean +/-SD)	.50 (.75)	.66 (.79)
		ISRC item 3: “Things seemed to happen really slowly” (mean +/-SD)	.68 (.86)	.73 (.79)
		ISRC item 4: “Things seemed to happen really fast” (mean +/-SD)	1.40 (.84)	1.13 (.85)
		ISRC item 5: “What was happening seemed unreal to me – like I was in a dream or watching a movie” (mean +/-SD)	1.37 (.82)	1.27 (.84)
		ISRC item 6: “I felt like I was not there – like I was not part of what was going on” (mean +/-SD)	.74 (.86)	.71 (.84)
		ISRC item 7: “I felt confused” (mean +/-SD)	1.17 (.87)	1.18 (.83)
		ISRC item 8: “I felt numb – like I did not have any feelings” (mean +/-SD)	.82 (.87)	.87 (.86)
		ISRC item 9: “People like my family or friends seemed like strangers to me” (mean +/-SD)	.24 (.60)	.21 (.54)
		ISRC item 10: “Everything seemed weird, not normal” (mean +/-SD)	.96 (.85)	1.08 (.82)
		ISRC item 11: “At times I was not sure where I was or what time it was” (mean +/-SD)	.74 (.86)	.68 (.79)
		ISRC item 12: “There were times when I did not feel any pain even when I was hurt.” (mean +/-SD)	.76 (.87)	.86 (.87)
		ISRC item 13: “I felt really scared” (mean +/-SD)	1.54 (.74)	1.36 (.79)
		ISRC item 14: “I wanted to make it stop happening, but I could not” (mean +/-SD)	1.62 (.69)	1.40 (.79)

	ISRC item 15: “I felt sick because what was happening seemed so horrible” (mean +/-SD)	1.04 (.92)	.94 (.87)
	ISRC item 16: “I can’t remember some parts of what happened” (mean +/-SD)	.58 (.78)	.88 (.87)
	ISRC item 17: “I can’t stop thinking about what happened” (mean +/-SD)	1.47 (.78)	1.22 (.75)
	ISRC item 18: “I don’t want to think about what happened” (mean +/-SD)	1.19 (.87)	.91 (.82)
	ISRC item 19: “I feel jumpy” (mean +/-SD)	.64 (.82)	.63 (.78)
	ISRC item 20: “My feelings are numb – I feel cut off from my emotions” (mean +/-SD)	.51 (.77)	.53 (.74)
	ISRC item 21: “When I think about what happened, I feel really upset” (mean +/-SD)	1.33 (.82)	1.26 (.79)
	ISRC item 22: “I am trying not to remember or think about what happened to me” (mean +/-SD)	.86 (.88)	.63 (.76)
	ISRC item 23: “I am having a hard time concentrating or paying attention” (mean +/-SD)	.56 (.76)	.79 (.75)
	ISRC item 24: “I feel spacey or out of touch with the world around me” (mean +/-SD)	.52 (.76)	.57 (.71)
	ISRC item 25: “Pictures or sounds from what happened keep popping into my mind” (mean +/-SD)	1.26 (.86)	.97 (.81)
	ISRC item 26: “I get upset when something reminds me of what happened” (mean +/-SD)	.66 (.84)	.69 (.79)
	ISRC item 27: “I feel hyper or like I can’t stay still” (mean +/-SD)	.38 (.69)	.51 (.71)
	Peritraumatic Dissociative Experiences Questionnaire – PDEQ (sum score) (mean +/-SD)	23.79 (9.30)	24.96 (10.26)
	PDEQ item 1: “I had moments of losing track of what was going on. I blanked out or spaced out or in some way felt that I was not part of what was going on” (mean +/-SD)	2.52 (1.57)	2.99 (1.60)
	PDEQ item 2: “I found that I was on automatic pilot. I ended up doing things that I later	2.02 (1.46)	2.35 (1.47)

	realized I had not actively decided to do" (mean +/-SD)		
	PDEQ item 3: "My sense of time changed. Things seemed to be happening in slow motion" (mean +/-SD)	2.32 (1.50)	2.65 (1.46)
	PDEQ item 4: "What was happening seemed unreal to me, like I was in a dream, or watching a movie or play" (mean +/-SD)	3.26 (1.58)	3.02 (1.52)
	PDEQ item 5: "I felt as though I was a spectator, watching what was happening to me as if I were floating above the scene or observing it as an outsider" (mean +/-SD)	2.08 (1.42)	2.02 (1.37)
	PDEQ Item 6: "There were moments when my senses of my own body seemed distorted or changed. I felt disconnected from my own body, or that it was unusually large or small" (mean +/-SD)	2.15 (1.48)	2.09 (1.34)
	PDEQ item 7: "I felt as though things that were actually happening to others, were happening to me – like I was in danger when I really was not" (mean +/-SD)	1.46 (1.25)	1.50 (1.11)
	PDEQ item 8: "I was surprised to find afterwards that a lot of things happened at the time that I was not aware of, especially things that I ordinarily would have noticed" (mean +/-SD)	2.50 (1.59)	2.47 (1.55)
	PDEQ item 9: "I felt confused; That is, there were moments when I had difficulty making sense of what was happening" (mean +/-SD)	2.93 (1.54)	2.83 (1.46)
	PDEQ item 10: "I felt disoriented; That is, there were moments when I felt uncertain about where I was or what time it was" (mean +/-SD)	2.50 (1.65)	2.53 (1.49)

Supplementary Table 3. Trauma type. All potential participants were approached by study personnel and ask if they felt that their life was in danger and that they could have been seriously injured or killed. In the affirmative case, a brief interview was performed to establish trauma exposure according to DSM.[9] The IRB approved wording for the first contact was: "We are conducting a survey on how people react to an event where they felt they could have died, been seriously injured, or their life or physical integrity or that of others was threatened. These people may experience intense fear, hopelessness, or horror. Does that sound like what happened to you today?"

Trauma types	**Model development dataset (N=377)**	**External validation dataset** (N=220)
Gunshot wound	17	2
Pedestrian versus car	38	40
Motor vehicle collision	199	34
Motorcycle collision	16	3
Bike accident	12	47
Fall	14	41
Sexual assault	23	0
Non-sexual assault	26	17
Others	32	36

Note: X_2 (8) = 167.03, $p \leq .001$

Supplementary Table 4. Shown are the results of the Latent Growth Mixture Modeling with fixed variance for intercept and slope. The final model selected for the Grady Memorial Hospital dataset (model developing sample) is highlighted in bold font type. LGMM was used to empirically identify the number and shape of longitudinal trajectories. The best-fitting model was selected through a nested modeling approach based on entropy, reduction in Information Criteria, i.e. Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), sample-size adjusted Bayesian Information Criterion (SSBI), and significance indicated by the Vuong-Lo-Mendell-Rubin Likelihood test (VRLT) and the Lo-Mendell-Rubin Adjusted LRT test (LRT) along with parsimony and interpretability.

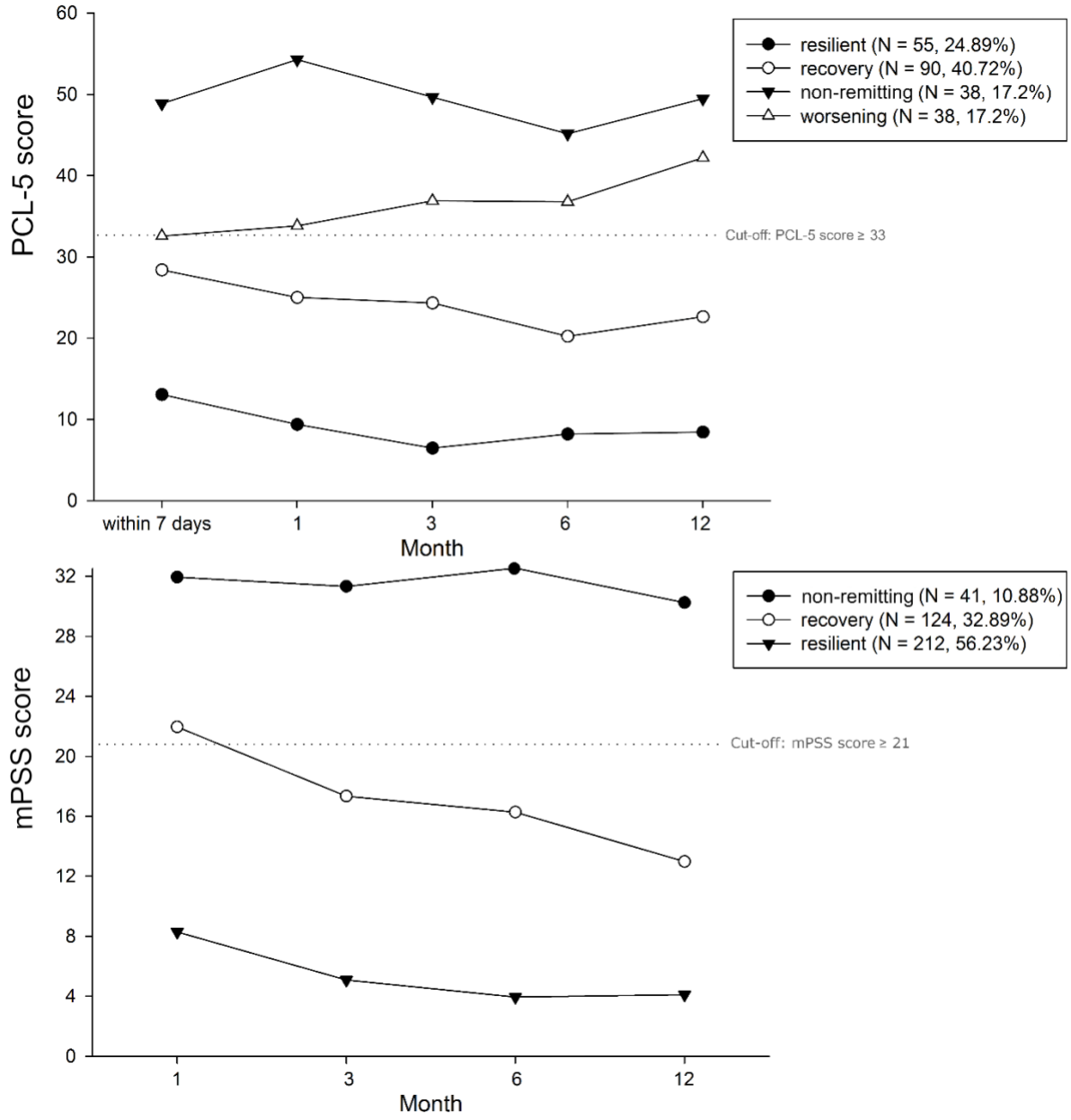
Fit	Linear weights					Linear and quadratic weights				
indices	1 class	2 class	**3 class**	4 class	5 class	1 class$_f$	2 class$_f$	3 class$_f$	4 class$_f$	5 class$_f$
AIC	8894.640	8335.608	**8111.160**	8053.352	8022.029	8889.182	8323.185	8086.142	8030.327	7991.707
BIC	8918.233	8370.998	**8158.347**	8112.336	8092.809	8916.707	8366.440	8145.126	8105.040	8082.149
SSBI	8899.197	8342.443	**8120.274**	8064.745	8035.699	8894.498	8331.539	8097.534	8044.757	8009.176
Entropy	–	.847	**.830**	.836	.857	–	.848	.845	.846	.859
VRLT	–	.0087	**.0036**	.1494	.0319	–	.0080	.4219	.0700	.3868
LRT	–	.0106	**.0046**	.1581	.0363	–	.0094	.4300	.0729	.3972

Note: AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; SSBI: Sample-size adjusted BIC; VRLT: Vuong-Lo-Mendell-Rubin Likelihood Ration test; LRT: Lo-Mendell-Rubin Adjusted LRT test; $_f$: To avoid computational issues the variance around the quadratic trend has been fixed to zero.

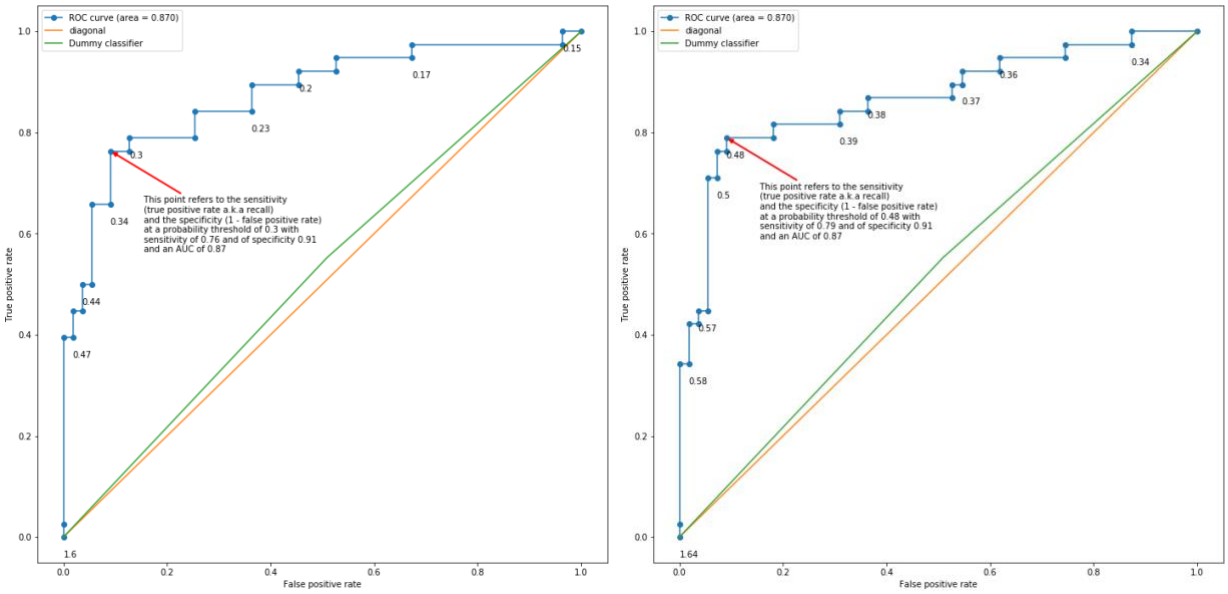
Supplementary Table 5. Final model selection in bold for the Bellevue Hospital Center (external model validation sample) results of the Latent Growth Mixture Modeling with random variance for intercept and slope. LGMM was used to empirically identify the number and shape of longitudinal trajectories. The best-fitting model was selected through a nested model based on entropy, reduction in Information Criteria, i.e. Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), sample-size adjusted Bayesian Information Criterion (SSBI), and significance indicated by the Vuong-Lo-Mendell-Rubin Likelihood test (VRLT) and the Lo-Mendell-Rubin Adjusted LRT test (LRT) along with parsimony and interpretability

Fit indices	Linear weights					Linear and quadratic weights				
	1 class	2 class	3 class	4 class	5 class	1 class$_f$	2 class$_f$	3 class$_f$	4 class$_f$	5 class$_f$
AIC	9203.624	8465.123	8112.013	7969.663	7883.365	9196.630	8441.477	8071.746	**7915.391**	7812.809
BIC	9227.411	8499.104	8156.190	8024.034	7947.930	9223.815	8482.255	8126.117	**7983.354**	7894.365
SSBI	9205.228	8467.414	8114.992	7973.329	7887.719	9198.463	8444.226	8075.412	**7919.973**	7818.308
Entropy	–	0.949	0.953	0.952	0.924	–	0.952	0.957	**0.957**	0.930
VRLT	–	<.0001	<.0001	0.0050	0.2423	–	<.0001	<.0001	**0.0085**	0.2612
LRT	–	<.0001	<.0001	0.0066	0.2633	–	<.0001	<.0001	**0.0104**	0.2759

Note: AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; SSBI: Sample-size adjusted BIC; VRLT: Vuong-LoMendell-Rubin Likelihood Ration test; LRT: Lo-Mendell-Rubin Adjusted LRT test; $_f$: To avoid computational issues the variance around the quadratic trend has been fixed to zero.



Supplementary Fig. 3. LGMM – the unconditional model of the latent trajectories of the longitudinal PTSD symptom severity progression. The upper figure represents the trajectories of the model validation sample (Bellevue Hospital sample) and the lower figure represents those from the model development sample (Grady Memorial Hospital sample)

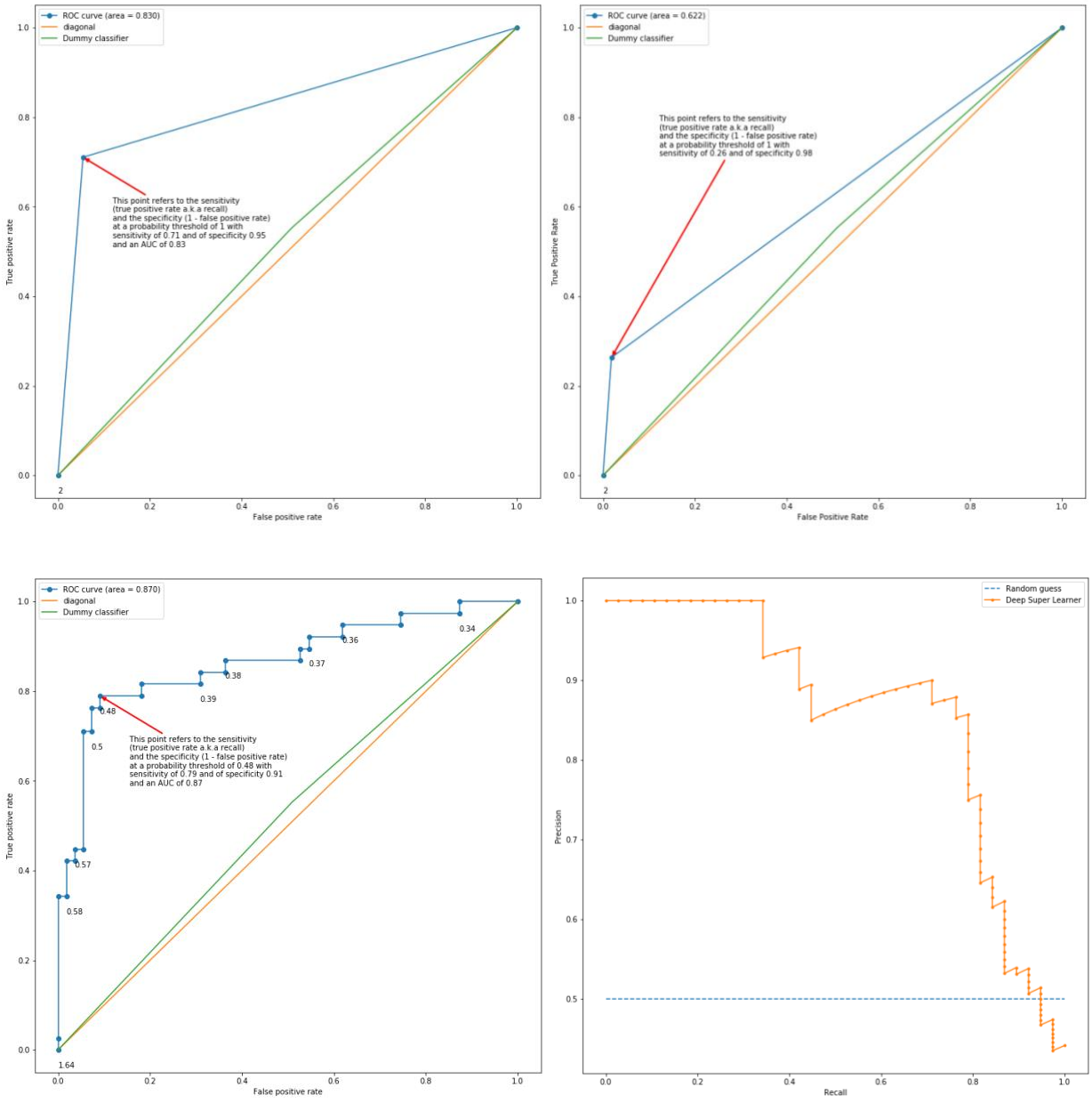


Supplementary Fig. 4. Displayed is the performance of the deep super leaner to predict PCL-5 score ≥33 (left) and LGMM (right) evaluated on the external test set. Please also refer to Supplementary Fig. 5 for the comparison with the ROC curve for a single threshold of 0.5.

Supplementary Table 6. Weighted average performance for predictive model development and validation on an internal test set (Grady Memorial Hospital) and external validation set (Bellevue Hospital Center, NY). Presented are the results of one base learner (logistic regression) using data collected at ED to predict the primary outcome of non-remitting stress symptoms versus resilience as indicated by LGMM. The model uses all the information shown in Supplementary Table 2.

	Logistic regression				
	Precision	Recall	f1-sore	ROC-AUC	Positive events*/total events**
Trainings set	.81	.69	.73	.68	27/164
Internal validation	.77	.63	.68	.58	14/89
External validation	.76	.69	.63	.62	38/93

Note: * positive events = LGMM class of ED patients with non-remitting PTSD symptoms; ** total events = all participants included in the analysis (LGMM class of ED patients with non-remitting PTSD symptoms + LGMM class of resilient ED patients)



Supplementary Fig. 5. Displayed is the performance of the algorithm (upper left) compared to a logistic regression model as evaluated on the external test set (upper right). The upper ROC curve displays the performance of a “hard classifier” that assigns each patient to exactly one discrete class (“non-remitting” versus “resilient”). The lower left ROC curve displays the ROC curve of a “soft classifier” providing for each participant the probability of being in class “non-remitting” and also the probability of being class “resilient” and then allows to select the optimal threshold to make the class assignment. The lower left ROC curve shows that for our algorithm the optimal threshold (0.48) would be slightly below 0.5 (the typical default threshold; reported in the paper) which is rewarded with a slightly improved specificity and sensitivity. To further asses the performance on the external validation set, we can also assess the precision-recall curve and compare the model against a random guess (dotted blue line).

Supplementary Table 7. Shown is the weighted average performance of a deep super learner using biomarker data collected at ED presentation. The outcome is the non-remitting stress symptoms versus resilience as indicated by LGMM. The model only uses **objective biomarkers** extracted from electronic medical records (see column "data type" in Supplementary Table 2).

	Precision	Recall	f1-sore	ROC-AUC	Positive events*/Total events**
Trainings set	.92	.90	.91	.91	27/164
Internal validation	.82	.67	.72	.64	14/89
External validation	.72	.67	.61	.72	38/93

Note: * positive events = LGMM class of ED patients with non-remitting PTS symptoms; ** total events = all participants included in the analysis (LGMM class of ED patients with non-remitting PTS symptoms + LGMM class of resilient ED patient

Supplementary Table 8. Weighted average performance for predictive model development and validation on an internal test set (Grady Memorial Hospital) and external validation set (Bellevue Hospital Center, NY). Presented is a deep super learner using **psychometric data** collected at ED (see column "data type" in Supplementary Table 2) to predict the primary outcome of non-remitting stress symptoms versus resilience as indicated by LGMM.

	Deep super learner (to predict LGMM profiles based on "early PTSD-related symptoms)					
	Precision	Recall	f1-sore	ROC-AUC	Positive events*	events*/Total events**
Trainings set	.88	.85	.85	.85		27/164
Internal validation	.85	.71	.75	.74		14/89
External validation	.66	.63	.56	.56		38/93

Note: * positive events = LGMM class of ED patients with non-remitting PTSD symptoms; ** total events = all participants included in the analysis (LGMM class of ED patients with non-remitting PTSD symptoms + LGMM class of resilient ED patients)

weaker than what was described in text as fair, by heuristic poor or failed and was essentially at chance

Supplementary Table 9. Weighted average performance metrics of several experiments during predictive model development and validation on an internal test set (Grady Memorial Hospital, Atlanta) and external validation set (Bellevue Hospital Center, NY).

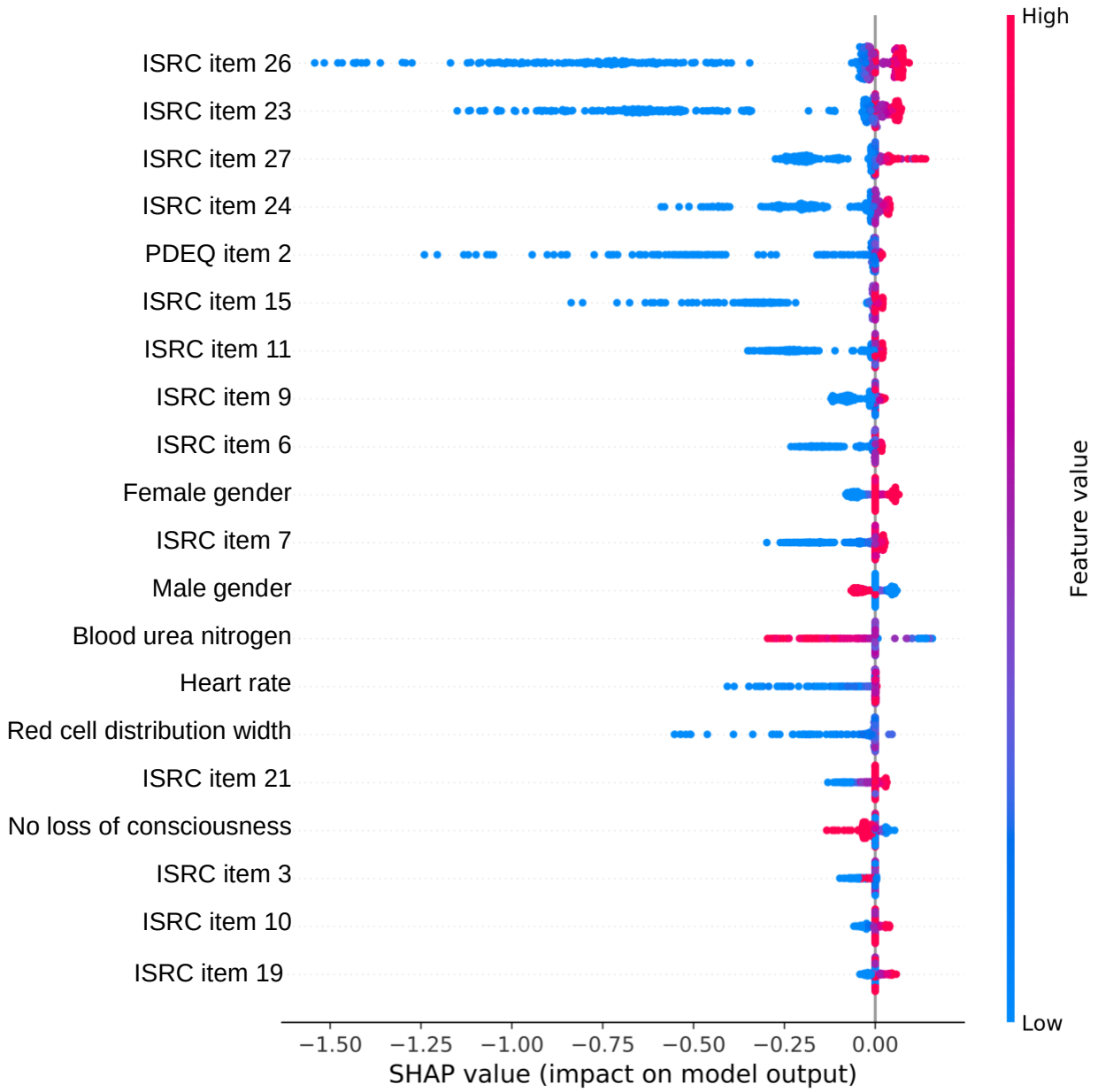
	Classification of ED patients PTS symptom trajectories				
	"non-remitting symptoms" versus "resilience"				
	Precision	Recall	f1-sore	ROC-AUC	Positive events*/Total events**
Trainings set	.86	.84	.84	.84	27/164
Internal validation	.83	.64	.69	.70	14/89
External validation	.86	.85	.85	.83	38/93
	"recovery of symptoms" versus "resilience"				
	Precision	Recall	f1-sore	ROC-AUC	Positive events***/Total events****
Trainings set	.81	.73	.71	.73	77/218
Internal validation	.66	.58	.58	.62	47/118
External validation	.77	.73	.73	.75	90/145
	"non-remitting" versus "recovery of symptoms"				
	Precision	Recall	f1-score	ROC-AUC	Positive events*/Total events*****
Trainings set	.87	.83	.82	.83	23/107
Internal validation	.61	.53	.55	.54	18/58
External validation	.77	.74	.75	.75	38/128
	"non-remitting symptoms" versus rest (the three other trajectories)				
	Precision	Recall	f1-sore	ROC-AUC	Positive events*/Total events******
Trainings set	.87	.84	.84	.96	35/320
Internal validation	.82	.68	.74	.59	6/57
External validation	.83	.75	.78	.78	38/221

Note: * positive events = LGMM class of ED patients with non-remitting PTS symptoms; total events = all participants included in the analysis (LGMM class of ED patients with non-remitting PTS symptoms + LGMM class of resilient ED patients); *** positive events = LGMM class of ED patients with recovery PTS symptoms; **** total events = all participants included in the analysis (LGMM class of ED patients with recovery PTS symptoms + LGMM class of resilient ED patients); ***** total events = all participants included in the analysis (LGMM class of ED patients with non-remitting PTS symptoms + LGMM class of recovery ED patients); ****** total events = all participants included in the analysis (all participants from all LGMM trajectories).

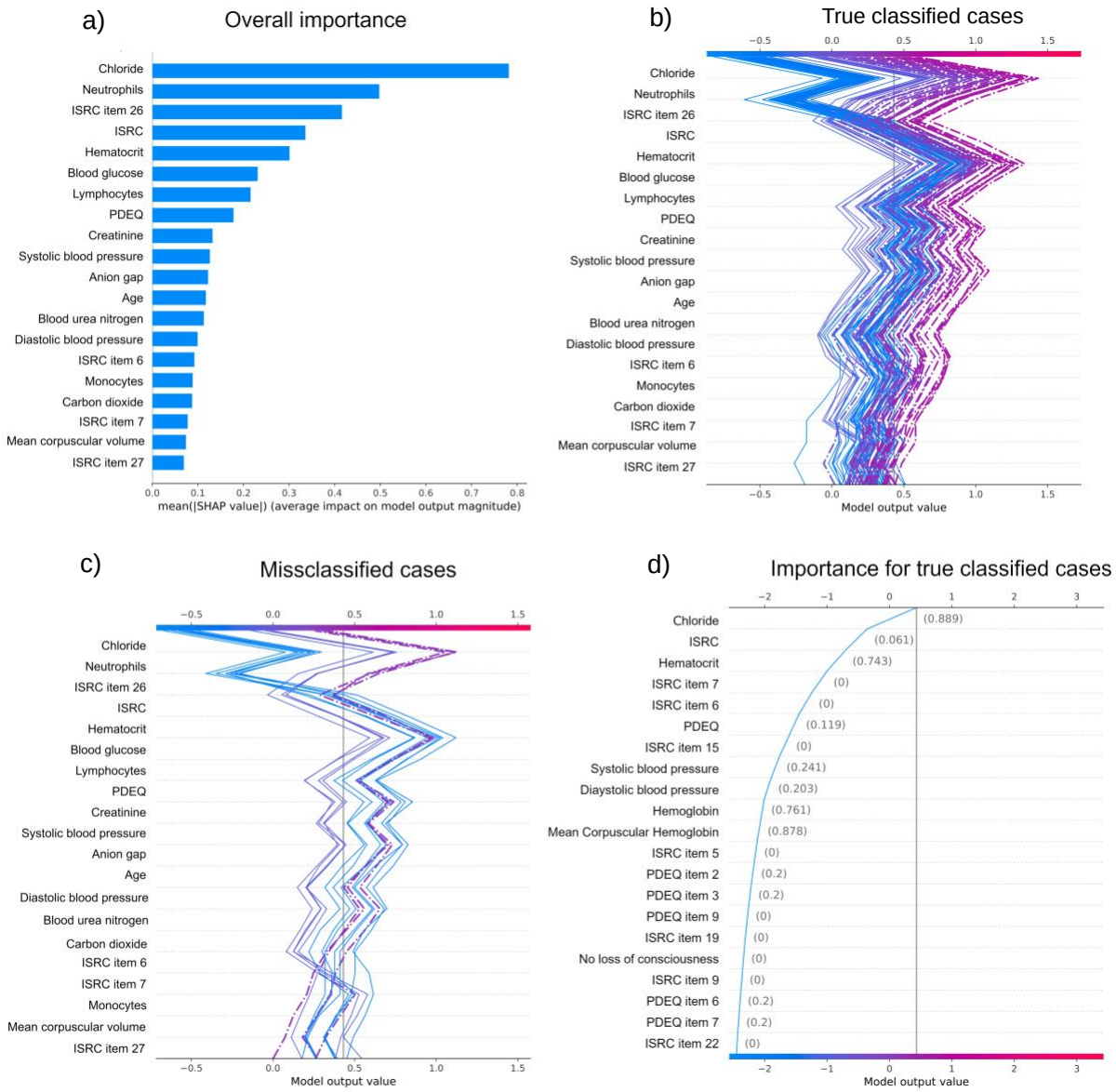
Supplementary Table 10. Shown is the performance per class on the external validation set for the prediction of “non-remitting” symptom trajectory vs. “resilience”.

	Deep super learner					
	Precision	Recall (TPR)	1-Recall (FNR)	Specificity (TNR)	1-Specificity (FPR)	Number of events/ Total events
non-remitting PTSD symptoms	0.90	0.71	0.29	0.95	0.05	38/93
Resilient ED patients	0.83	0.95	0.05	0.71	0.29	55/93
Weighted average	0.86	0.85	0.25	0.81	0.19	93/93

Note: * TPR = True Positive Rate, FNR = false negative rate, TNR = true negative rate, FPR = false positive rate



Supplementary Fig. 6. Variable importance for the training set using Electronic medical record data plus ISRC and PDEQ using SHAP values for the prediction of non-remitting symptoms versus rest, i.e. all three other trajectories (see Supplementary Table 9).



Supplementary Fig. 7. Supplementary Fig 7a Variable importance for the external validation set using SHAP (SHapley Additive exPlanations).[10] Note, only the rank order is different between training and external validation set (Figs. 1a and Supplementary Fig. 7a). The decision plot in Supplementary Figs 7b and 7c show how the variables change specific types of model outputs. Supplementary Fig 7b displays the decision path for correctly classified observations (dotted lines represent the non-remitting trajectory); Supplementary Fig 7c shows the decision path of the model for misclassified observations (dotted lines represent the non-remitting trajectory); Supplementary Fig 7d shows the mean impact of the top-ranked variables for the correctly classified observation.

Supplementary Discussion

As an exploratory part of this study, we examined possible "reasons" why the algorithm made a given prediction. We used posthoc calculated SHAP values that rank the features in the order of importance for the model's prediction. SHAP values allow assessing what information influences the prediction the most while accounting for the influence of all other features and controlling for the order of adding features to the model[10]. SHAP values allow examining the contribution of each predictor variable on each predicted outcome for each individual. SHAP values can be calculated for both the model development as well as the external validation data set. The Figs. 1a and 1b display the contribution of the top 20 most influential variables on the mean predicted outcome of the algorithm applied to the training dataset used for model development. Figs. 1c and 1d show the variable importance for the algorithm being applied only to the data extracted from the EMR plus 4 ISRC items. The SHAP values can also visualize how the model arrives at correct classifications (Supplementary Fig. 7b) or misclassifications on the external validation data (Supplementary Fig. 7c). The prediction of each observation is always an achievement of all variables together and the variable importance ranking of the predictive model should not be interpreted causally. While causal explanation requires randomized controlled trials,[11] our study design was optimized for predictive accuracy on out-of-sample data. Nevertheless, the posthoc assessment of the variable importance may generate novel hypotheses. Such new hypotheses will require additional prospective studies to determine their relevance as risk factors or to guide prevention measures.

Interestingly, the proposed prediction model uses probabilistic information of peripheral immune markers such as neutrophils, lymphocyte, monocytes, which were assessed directly in the ED after the traumatic event took place. This is consistent with previous studies showing that inflammatory and immune markers are important predictors of posttraumatic stress responses[12-14]. Moreover, we identified some predictive potential of blood markers, such as hematocrit and MCV, for discriminating the resilient and the non-remitting posttraumatic stress symptom trajectories. Altered hematocrit levels are linked to PTSD and acute stress response[15-17]. Recently, platelets and red blood cell counts were reported to be associated with PTSD and inflammation[18].

Concerning the symptoms manifested directly after trauma, we identified predictive information in acute stress disorder symptoms measured with the ISRC and PDEQ. This is in line with recent work of the International Consortium to Predict PTSD showing that the initial Clinician-Administered PTSD Scale (CAPS) score is predictive for the development of PTSD

But the psychometric by itself yielded only AOC .56 for external validation!? Thats not good

symptoms[19]. Previous results showed that early symptoms measured with the Posttraumatic Adjustment Scale collected at the end of the hospital discharge (on average 8.27 (SD=10.61) days after severe injury) are predictive for PTSD development[20]. Similarly, peritraumatic dissociation is associated with PTSD[21].

Previous studies examining early risk factors after trauma found age and physiological markers of the stress response, such as blood pressure or pulse, as predictors for PTSD symptoms[22-25]. Based on the SHAP values and the computationally complex machine learning algorithm, we identified several metabolic markers that are new and not yet well-established as potential predictors of the posttraumatic stress symptom course after traumatic stressors. For instance, blood glucose was a relevant predictor in our samples. This is consistent with previous studies showing that acute stress affects glucose metabolism[26-28]. Similarly, CO_2 is a relevant predictor consistent with previous studies showing that CO_2 induction is associated with acute panicogenic and anxiogenic effects[29] and panic attacks[30]. Elevated arousal during hyperventilation is reported to be associated with PTSD symptoms[31]. Intriguingly, blood chloride levels were highly discriminatory between the patients on a “non-remitting” and a “resilient” trajectory in our sample. Notably, anion gap was another potential predictor. Chloride and anion gap are critical measures of the body’s pH (acid-base) status, and as with CO_2, are associated with acidosis, which is associated with panic attacks and fear in numerous studies[32-36]. Whether or not chloride levels are mediated by potential blood or NaCl infusion warrants further investigation. Finally, recent work from our group has identified nausea as a predictor for later PTSD development[37]. Nausea was not included in the current analyses, however, nausea is also associated with altered chloride, anion gap, and CO_2 status[38].

Importantly, the identified markers, while likely not mechanistically, relate to biological components of biological systems that have been examined either in the preclinical or the clinical literature as mechanistically related to stress and trauma responses. For example, CO_2 production is mechanistically related to ghrelin production[39], a hormone that is activated by stress via the hypothalamic-pituitary-adrenal axis[40]. Similarly, hematocrit changes alongside with cortisol warrant further investigation. Cortisol is a well-studied predictor and correlate of PTSD[41] and there is a mechanistic relationship between lymphocytes and cortisol[41]. This indicates that there is a potential for further research on accessible proxy markers that provide probabilistic information consistent with the mechanistic functioning of the well-studied stress response system. Furthermore, some molecular modulators of the stress system have demonstrated effects as treatments or prophylactics for PTSD[42-44]. Ultimately, these approaches

may be better matched to the individual based on data-driven predictive algorithms that capitalize on these data sources.

Supplementary References

1. van de Schoot, R., Sijbrandij, M., Winter, S.D., Depaoli, S. & Vermunt, J.K. The GRoLTS-Checklist: Guidelines for Reporting on Latent Trajectory Studies. *Structural Equation Modeling: A Multidisciplinary Journal* **24**, 451-467 (2017).
2. Muthén, L.K. & Muthén, B.O. *Mplus User's Guide: Statistical analysis with latent variables* (Muthén & Muthén, Los Angeles, CA, 1998-2017).
3. Kuhn, M. & Johnson, K. *Applied predictive modeling*, (Springer, 2013).
4. Kuhn, M. Caret package. (2008).
5. Breiman, L. Bagging predictors. *Machine Learning* **24**, 123-140 (1996).
6. Young, S., Abdou, T. & Bener, A. Deep Super Learner: A Deep Ensemble for Classification Problems. in *Advances in Artificial Intelligence: 31st Canadian Conference on Artificial Intelligence, Canadian AI 2018, Toronto, ON, Canada, May 8–11, 2018, Proceedings 31* 84-95 (Springer, 2018).
7. Young, S., Abdou, T. & Bener, A. Deep Super Learner: A Deep Ensemble for Classification Problems. *arXiv preprint arXiv:1803.02323* (2018).
8. Chawla, N.V., Bowyer, K.W., Hall, L.O. & Kegelmeyer, W.P. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* **16**, 321-357 (2002).
9. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)*, (American Psychiatric Pub, 2013).
10. Lundberg, S.M. & Lee, S.-I. A unified approach to interpreting model predictions. in *Advances in Neural Information Processing Systems* 4765-4774 (2017).
11. Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science* **12**, 1100-1122 (2017).
12. Mellon, S.H., Gautam, A., Hammamieh, R., Jett, M. & Wolkowitz, O.M. Metabolism, metabolomics, and inflammation in post-traumatic stress disorder. *Biological psychiatry* (2018).
13. Michopoulos, V., Powers, A., Gillespie, C.F., Ressler, K.J. & Jovanovic, T. Inflammation in fear-and anxiety-based disorders: PTSD, GAD, and beyond. *Neuropsychopharmacology* **42**, 254 (2017).
14. Michopoulos, V*., et al.* Association of Prospective Risk for Chronic PTSD Symptoms With Low TNFα and IFNγ Concentrations in the Immediate Aftermath of Trauma Exposure. *American Journal of Psychiatry*, appi. ajp. 2019.19010039 (2019).
15. Austin, A.W., Patterson, S.M. & von Känel, R. Hemoconcentration and hemostasis during acute stress: interacting and independent effects. *Annals of behavioral medicine* **42**, 153-173 (2011).
16. Hammamieh, R*., et al.* Murine model of repeated exposures to conspecific trained aggressors simulates features of post-traumatic stress disorder. *Behavioural brain research* **235**, 55-66 (2012).
17. Kudielka, B.M. & Wüst, S. Human models in acute and chronic stress: assessing determinants of individual hypothalamus–pituitary–adrenal axis activity and reactivity. *Stress* **13**, 1-14 (2010).

18. Lindqvist, D.*, et al.* Increased circulating blood cell counts in combat-related PTSD: Associations with inflammation and PTSD severity. *Psychiatry research* **258**, 330-336 (2017).
19. Shalev, A.Y.*, et al.* Estimating the risk of PTSD in recent trauma survivors: results of the International Consortium to Predict PTSD (ICPP). *World Psychiatry* **18**, 77-87 (2019).
20. O'Donnell, M.L.*, et al.* A predictive screening index for posttraumatic stress disorder and depression following traumatic injury. *Journal of Consulting and Clinical Psychology* **76**, 923 (2008).
21. Lensvelt-Mulders, G.*, et al.* Relations among peritraumatic dissociation and posttraumatic stress: A meta-analysis. *Clinical psychology review* **28**, 1138-1151 (2008).
22. Papini, S.*, et al.* Ensemble machine learning prediction of posttraumatic stress disorder screening status after emergency room hospitalization. *Journal of anxiety disorders* **60**, 35-42 (2018).
23. Galatzer-Levy, I.R., Karstoft, K.I., Statnikov, A. & Shalev, A.Y. Quantitative forecasting of PTSD from early trauma responses: A machine learning application. *Journal of psychiatric research* **59**, 68-76 (2014).
24. Galatzer-Levy, I.R., Ma, S., Statnikov, A., Yehuda, R. & Shalev, A.Y. Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD. *Translational psychiatry* **7**, e1070 (2017).
25. Shalev, A.Y., Sahar, T., Freedman, S. & et al. A prospective study of heart rate response following trauma and the subsequent development of posttraumatic stress disorder. *Archives of General Psychiatry* **55**, 553-559 (1998).
26. Pitman, R.K.*, et al.* Biological studies of post-traumatic stress disorder. *Nature Reviews Neuroscience* **13**, 769-787 (2012).
27. Nowotny, á.*, et al.* Effects of acute psychological stress on glucose metabolism and subclinical inflammation in patients with post-traumatic stress disorder. *Hormone and Metabolic Research* **42**, 746-753 (2010).
28. Yehuda, R., McFarlane, A. & Shalev, A. Predicting the development of posttraumatic stress disorder from the acute response to a traumatic event. *Biological psychiatry* **44**, 1305-1313 (1998).
29. Muhtz, C., Yassouridis, A., Daneshi, J., Braun, M. & Kellner, M. Acute panicogenic, anxiogenic and dissociative effects of carbon dioxide inhalation in patients with post-traumatic stress disorder (PTSD). *Journal of psychiatric research* **45**, 989-993 (2011).
30. Cougle, J.R., Feldner, M.T., Keough, M.E., Hawkins, K.A. & Fitch, K.E. Comorbid panic attacks among individuals with posttraumatic stress disorder: Associations with traumatic event exposure history, symptoms, and impairment. *Journal of Anxiety Disorders* **24**, 183-188 (2010).
31. Nixon, R.D. & Bryant, R.A. Induced arousal and reexperiencing in acute stress disorder. *Journal of Anxiety Disorders* **19**, 587-594 (2005).
32. Quagliato, L.A., Freire, R.C. & Nardi, A.E. The role of acid-sensitive ion channels in panic disorder: a systematic review of animal studies and meta-analysis of human studies. *Translational psychiatry* **8**, 185 (2018).
33. Vollmer, L., Strawn, J. & Sah, R. Acid–base dysregulation and chemosensory mechanisms in panic disorder: a translational update. *Translational psychiatry* **5**, e572 (2015).
34. Smoller, J.W.*, et al.* The human ortholog of acid-sensing ion channel gene ASIC1a is associated with panic disorder and amygdala structure and function. *Biological psychiatry* **76**, 902-910 (2014).
35. Wemmie, J.A. Neurobiology of panic and pH chemosensation in the brain. *Dialogues in clinical neuroscience* **13**, 475 (2011).

36. Peskind, E.R.*, et al.* Sodium lactate and hypertonic sodium chloride induce equivalent panic incidence, panic symptoms, and hypernatremia in panic disorder. *Biological psychiatry* **44**, 1007-1016 (1998).
37. Michopoulos, V.*, et al.* Nausea in the peri-traumatic period is associated with prospective risk for PTSD symptom development. *Neuropsychopharmacology*, 1 (2018).
38. Hertford, J., McKenna, J. & Chamovitz, B. Metabolic acidosis with an elevated anion gap. *American family physician* **39**, 159-168 (1989).
39. Wren, A.*, et al.* Ghrelin enhances appetite and increases food intake in humans. (2001).
40. Spencer, S.J.*, et al.* Ghrelin regulates the hypothalamic-pituitary-adrenal axis and restricts anxiety after acute stress. *Biological psychiatry* **72**, 457-465 (2012).
41. Thomson, S.P., McMahon, L.J. & Nugent, C.A. Endogenous cortisol: a regulator of the number of lymphocytes in peripheral blood. *Clinical immunology and immunopathology* **17**, 506-514 (1980).
42. Feder, A.*, et al.* Efficacy of intravenous ketamine for treatment of chronic posttraumatic stress disorder: a randomized clinical trial. *JAMA psychiatry* **71**, 681-688 (2014).
43. Sijbrandij, M., Kleiboer, A., Bisson, J.I., Barbui, C. & Cuijpers, P. Pharmacological prevention of post-traumatic stress disorder and acute stress disorder: a systematic review and meta-analysis. *The Lancet Psychiatry* **2**, 413-421 (2015).
44. Vermetten, E., Zhohar, J. & Krugers, H.J. Pharmacotherapy in the aftermath of trauma; opportunities in the 'golden hours'. *Current psychiatry reports* **16**, 455 (2014).

nature research

Corresponding author(s): Katharina Schultebraucks

Last updated by author(s): May 15, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	IBM SPSS Statistics Version 22 REDCap (Version 10/2014-02/2017; Harris et al., 2009) PA Harris, R Taylor, R Thielke, J Payne, N Gonzalez, JG. Conde, Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support, J Biomed Inform. 2009 Apr;42(2):377-81.
Data analysis	Mplus version 7, R version 3.5.1 in Rstudio 1.1.456, Python version 3.7

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data is available upon reasonable request but for research purposes only. Please send requests to the Principal Investigators of the two study sites (Kerry Ressler, McLean Hospital, Harvard Medical School; Isaac Galatzer-Levy, NYU School of Medicine).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Since we did not perform classical significance testing, performing a traditional power calculation based on pre-defined significance level, effect size, and power level is not adequate. However, previous literature in this area has established that a sample size of 100 subjects or more is required to build multivariate predictive models to detect a modest multivariate/composite predictive signal (AUC ~ 0.7). As opposed to methods such as logistic regression that operate under the assumption that the underlying relationship between predictors and outcome is a linear function, supervised learning does not make this assumptions and instead depends on logical if-then conditions. This makes supervised learning suitable for sample sizes in the hundreds as are often observed in biomedicine, where sampling is very cost-intensive. In light of the expected conversion rate to PTSD of about 15% in our observational cohorts, we defined that sample size must be large enough so that in the training set of the supervised learning algorithm at least 25 independent samples exist for each class to predict.
Data exclusions	Inclusion and exclusion criteria for eligible patients were pre-established. All eligible cases with follow-up data on the outcome variable (see flow chart in Supplementary Tables 1 and 2) were included in the analysis. The exclusion of predictor variables from the data collected did not lead to the exclusion of cases from the sample but a smaller set of candidate predictors. The exclusion of predictor variables was not pre-specified but based on established principles from the literature. Variables not available at both sites or variables with more than 45% missing values were removed and assumed not to have a favorable information-to-noise ratio. We also removed predictors with near-zero-variance, i.e., predictors that have very few unique values relative to the number of samples and the ratio of the frequency of the most common value to the frequency of the second most common value is large (such predictors are theoretically implausible to yield unbiased predictions). All other missing values were replaced with statistically plausible estimates using bootstrapped aggregation (bagged) tree imputation.
Replication	We used a model development sample and tested our algorithm in an external validation sample on independent data. External validation of a predictive model is the gold standard to evaluate the generalizability of a predictive model.
Randomization	We included all participants who fulfilled the inclusion and exclusion criteria in the time frame and who consent to participate in our study. There were no treatment conditions so included patients were not assigned to treatment groups and therefore no radom allocation procedure was used.
Blinding	Patients received no treatment but were enrolled into a prospective observational cohort. As there were no treatment groups, no concealment of group allocation was performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	We examined N=377 trauma survivors (mean age = 36.05 +/-12.87, 47.1% female) in the model development sample (Grady Memorial Hospital) and the external validation sample (Bellevue Hospital Center) comprised N=221 trauma survivors (mean age = 36.69 +/-13.46, 37.1% female).
Recruitment	The subjects have been patients who arrived at the Emergency Department following a potential trauma. The participants were initially screened and approached based on information of the Emergency Department's "White Board", trauma

surgery discharge rounds, or the team's rounding sheet. All potential participants were approached by study personnel and ask if they felt that their life was in danger and that they could have been seriously injured or killed. In the affirmative case, a brief interview was performed to establish trauma exposure according to DSM. The IRB approved wording for the first contact was: "We are conducting a survey on how people react to an event where they felt they could have died, been seriously injured, or their life or physical integrity or that of others was threatened. These people may experience intense fear, hopelessness, or horror. Does that sound like what happened to you today?" Patients who met the study inclusion criteria but no exclusion criterion and who consented to participate in the study were eligible to enroll in the study and have been asked to complete a phone screen interview seven days after their emergency room triage, to determine their full eligibility. The enrolled participants completed the follow-up assessment (1, 3, 6, and 12 months) after their Emergency Department admission. The selection process can give rise to selection bias. Therefore, the inclusion and exclusion criteria presented in Table 1 are important in order to understand to which target population the results may generalize. In addition, we did not include all participants who met the inclusion and exclusion criteria but only those who were willing to provide informed consent. This could give rise to self-selection bias insofar as participants who are willing to participate may differ compared to those participants who refuse participation. We did not test for significant differences regarding sample characteristics, such as gender, age, or disease severity, in order to comply with the ethical principles of the IRB that did require informed consent for any data collection and analysis. However, because in this prospective study, the enrollment and data collection of candidate predictors was completed before the outcome was determined, it is very unlikely that the outcome status has directly affected the decision for or against study participation (self-selection bias). The inclusion and exclusion criteria depended on NIH grant peer review to help agudicate how to best build samples to support the goal of building predictive models that would be of high value to move this area of research forward.

Ethics oversight	The ethics committee of New York University (Bellevue study) and the Emory Institutional Review Board and the Grady Hospital Research Oversight Committee (Grady Memorial study).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	This is a prospective observational longitudinal cohort study without treatment or intervention and not within the scope of FDAAA 801. The protocol was not prospectively registered in a public database, which are not currently tailored and most data fields are not appropriate for the purpose to register development and external validatation of prognostic models.
Study protocol	The protocol is available on request to the Principal Investigators of the study sites (Kerry Ressler, McLean Hospital, Harvard Medical School; Isaac Galatzer-Levy, NYU School of Medicine).
Data collection	At two US sites, trauma survivors who were admitted to the ED of a Level 1 trauma center after experiencing a DSM criterion A trauma were enrolled into a longitudinal study cohort. The first sample (n=377) of ED patients was prospectively enrolled from 2012 to 2017 at the Marcus Trauma Center of the Grady Memorial Hospital, Atlanta, GA and was used for model development. A second sample (n=221) of ED patients was prospectively enrolled from 2012 to 2016 at Bellevue Hospital Center, New York City, NY and was used for external validation of the predictive model. At both sites, data was prospectively collected at ED admission, within 7 days (only at Bellevue Hospital Center) thereafter (phone screen interview), and at 1, 3, 6, and 12 months after ED. admission .
Outcomes	The primary outcome is the non-remitting longitudinal trajectory of posttraumatic stress symptoms of ED patients identified by Latent Growth Mixture Modeling (LGMM) vs. resilient ED patients. As a second outcome we classified those on a non-remitting trajectory versus all other trajectories. The outcome is measured with the well-established, validated clinician administered psychometric instruments (Modified PTSD Symptom Scale, mPSS; PTSD Checklist for DSM-5, PCL-5). As secondary outcome we predicted provisional PTSD diagnosis at 12 months after ED admission based on pre-defined criteria of a mPSS score ≥21 (model development) and a PCL-5 score ≥33 (model validation). Additionally, we examined the prediction a non-remitting vs. a recovery trajectory and also a recovery vs. a resilience trajectory.