ORIGINAL ARTICLE

Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD

IR Galatzer-Levy[1,2], S Ma[3], A Statnikov[4], R Yehuda[5] and AY Shalev[1,2]

To date, studies of biological risk factors have revealed inconsistent relationships with subsequent post-traumatic stress disorder (PTSD). The inconsistent signal may reflect the use of data analytic tools that are ill equipped for modeling the complex interactions between biological and environmental factors that underlay post-traumatic psychopathology. Further, using symptom-based diagnostic status as the group outcome overlooks the inherent heterogeneity of PTSD, potentially contributing to failures to replicate. To examine the potential yield of novel analytic tools, we reanalyzed data from a large longitudinal study of individuals identified following trauma in the general emergency room (ER) that failed to find a linear association between cortisol response to traumatic events and subsequent PTSD. First, latent growth mixture modeling empirically identified trajectories of post-traumatic symptoms, which then were used as the study outcome. Next, support vector machines with feature selection identified sets of features with stable predictive accuracy and built robust classifiers of trajectory membership (area under the receiver operator characteristic curve (AUC) = 0.82 (95% confidence interval (CI) = 0.80–0.85)) that combined clinical, neuroendocrine, psychophysiological and demographic information. Finally, graph induction algorithms revealed a unique path from childhood trauma via lower cortisol during ER admission, to non-remitting PTSD. Traditional general linear modeling methods then confirmed the newly revealed association, thereby delineating a specific target population for early endocrine interventions. Advanced computational approaches offer innovative ways for uncovering clinically significant, non-shared biological signals in heterogeneous samples.

Translational Psychiatry (2017) **7,** e1070; doi:10.1038/tp.2017.38; published online 21 March 2017

INTRODUCTION

Risk factors for emerging psychopathology may vary between individuals. Post-traumatic stress disorder (PTSD) represents a good example of this principle as diverse risk factors, associated with the development of the disorder and are not shared by all individuals affected.[1,2] In addition, biological risk factors (for example, genetics, epigenetic and neuroendocrine) with demonstrated associations with PTSD often failed to replicate.[3]

Such inconsistencies, however, may not necessarily reflect a weak or absent relationship between biological factors and responses to trauma. Other putative sources of inconsistencies include inadequate computational modeling, biological heterogeneity of PTSD, the disorder's multicausal etiology and the reliance on categorical outcome measures (for example, diagnostic categories).

The reliance on general linear models constitutes a barrier to replication because these models are not designed to examine complex interactions between variables of various distributional patterns, identify factors that concern a subpopulation and infer complex predictive relationships. Offering remedy to those, machine learning (ML) predictive modeling is increasingly used to develop algorithms that classify individuals with varying, unequally distributed risk factors (for example, in predicting the course of malignancies from tissue biomarkers[4]).

Early prediction and the identification of early intervention targets are of significant interest, under the assumption that mechanisms underlying the development of the disorder can be manipulated to prevent its development. Pathways to PTSD, however, involve distinct genetic, endocrine, demographic and environmental factors that are not shared by all PTSD patients, suggesting that efficient early treatment may have to address individual-specific pathogenic pathways.[1,5–9] Graph induction models, also known as network models, offer new ways to identify individual paths to illness. Graph induction algorithms are variants of unsupervised ML that identify and map a network of connections within complex multicausal matrices[10–12] and derive specific causal paths within these networks. The use of classification models and graph models both may provide unique and equally important information as classification models focus on the identification of sets of features that provide the strongest classification accuracy but may obfuscate upstream features that are mechanistically important.[13] Conversely, graph models can identify mechanistically important that may not be identified in classification models because they do not contribute to prediction

[1]Department of Psychiatry, New York University School of Medicine, New York, NY, USA; [2]Steven and Alexandra Cohen Center for the Study of Post-Traumatic Stress and Traumatic Brain Injury, New York, NY, USA; [3]Department of Medicine, University of Minnesota, Minneapolis, MN, USA; [4]Department of Medicine, New York University School of Medicine, New York, NY, USA and [5]Traumatic Stress Studies Division, Mount Sinai School of Medicine, James J Peters VA Medical Center, Bronx, NY, USA. Correspondence: Dr IR Galatzer-Levy, Department of Psychiatry, New York University School of Medicine, 1 Park Avenue, New York, NY 10016, USA.
E-mail: Isaac.Galatzer-Levy@nyumc.org

2

over and above features that are more proximal to the outcome. By conducting both sets of analyses, both the predictive accuracy and the mechanistic role of the endocrine response to trauma can be identified.

The selection of outcome is of equal importance in any research related to the development of PTSD. Psychiatric research traditionally relies on end point diagnosis cued to the Diagnostic and Statistical Manual of Mental Disorders (DSM) or International Classification of Disease systems of classification. Diagnostic categories, however, suffer from significant underlying heterogeneity,[14] ambiguous distributional properties[15] and limited empirical accuracy for predictive models.[16] To palliate this shortcoming, latent growth mixture modeling (LGMM) and allied methods have emerged as an empirically driven alternative method for identifying healthy and pathological responses to traumatic events.[17–22] LGMM identifies individuals' clustering into symptom progression trajectories and classifies participants belonging to these trajectories based on the posterior probability of class membership.[23] Relying on early PTSD symptoms templates or categories (for example, acute stress disorder) as predictors similarly constrain predictive models, whereas empirically derived predictive features (including PTSD symptoms but also other informing antecedents) have been shown to better predict PTSD and PTSD symptom trajectories.[16]

In the current investigation, we apply supervised and unsupervised ML methods to examine the relationship between background, environmental and neuroendocrine risk factors and PTSD development over 5 months that follow trauma exposure. Specifically, we apply LGMM trajectory identification, supervised ML to infer predictive features related to these trajectories and graph induction algorithms to examine causal paths to data from a large longitudinal study of the neuroendocrine response to traumatic event,[24,25] that previously failed to demonstrate group-wide association between ER endocrine markers and PTSD status 5 months later using classical general linear model statistics.

MATERIALS AND METHODS

Participants
As previously described,[24,25] adult survivors of potentially traumatic were evaluated during their emergency room (ER) admission following the traumatic incident and re-evaluated 10 days, 1 month and 5 months later. Subjects who completed their ER evaluation and at least two additional evaluations were included in the current study ($n = 152$; for further participant, recruitment and retention information see Supplementary Information).

Instruments
Assessment of trauma exposure. Characteristics evaluated during ER admission include exposure to death and injury of strangers or relatives (dichotomous yes/no answers), perceived threat during the event, subjective severity of the incident, subjective distress during the incident and negative perception of one's own responses (single item for each, scored 0 (not at all) to 4 (extreme)).

Personal information. Information acquired from ER records and the 10-day interviews included age, gender, trauma type, body mass index, heart rate (HR), blood pressure (BP), hour of admission and length of ER stay (ER records; HR and BP were evaluated upon ER admission, re-evaluated upon obtaining subject's consent and averaged) and education level (years of schooling), income, prior military service, past trauma history, smoking habits (yes/no; number smoked per day) and psychiatric treatment upon study enrollment (yes/no; 10 days records).

Neuroendocrine assessment. Blood and saliva samples were collected in the ER, and on the morning of each follow-up assessment (10 days, 1 month, 5 months). Assays performed included plasma cortisol, norepinephrine (NE) and adrenocorticotropic hormone (ACTH), lymphocytes' glucocorticoid-receptor density, saliva cortisol and, during ER admission only, urine cortisol and epinephrine, based on 4-h urine

collection during ER admission. For a full description of biological sample collection, storage and assays (see Supplementary Material and Videlock et al.[24] and Shalev et al.[25]).

Diagnostic and psychometric instruments. The Clinician-Administered Scale[26] evaluated DSM-IV PTSD symptom severity and diagnostic status at 1 and 5 months.

The *Structured Clinical Interview for DSM-IV*[27] Global Assessment of Functioning (GAF; single item; range 1-100) was administered at 1 and 5 months).

The *Impact of Events Scale—Revised*[28] measured PTSD severity at 10 days, 1 month and 5 months. Subscales of intrusions, avoidance and hyperarousal were derived following the instrument's specifications.[28]

The *Beck Depression Inventory (BDI*[29]) and the *State-Trait Anxiety Inventory*[30] evaluated depression and anxiety (10 days, 1 month and 5 months).

The *Peritraumatic Dissociative Experience Questionnaire*[31] administered at 10 days evaluated dissociation symptoms upon trauma exposure.

The *Trauma History Questionnaire*[32] (1 month) evaluated lifetime exposure to traumatic events. The Trauma History Questionnaire queries exposure to each of 13 distinct traumatic events, and for each event, the presence of fear, helplessness or horror (as per DSM-IV PTSD Criterion A: A traumatic event). Two dichotomous variables are: (a) exposure to childhood trauma and (b) exposure to adult life trauma.

Data analytic procedures
Identification of trajectories of post-traumatic stress. LGMM[28] was employed to identify PTSD symptom severity trajectories (Impact of Events Scale—Revised total scores) from 10 days to 5 months using MPlus 7.2.[33] The best-fitting model was selected through a nested model approach where progressive numbers of classes are fit until model fit indices no longer favor additional classes following recommendations from the literature[34] (see Supplementary Information for statistical selection criteria). In the current study, results were validated by specifying 500 random starting values, resulting in consistent replicating of the best log likelihood value.[35,36] To further guard against overfitting, the sample was randomly split into two halves, and the best analytic procedure was executed on both the random halves to determine whether the same number of classes were recovered in both random splits. The split-half cross-validation was conducted by comparing the Lo-Mendell Rubin likelihood ratio Test (LMR-LRT) because of evidence from the literature that it is an accurate statistic for class enumeration.[37]

Individuals were assigned to classes based on the highest posterior probability of class membership in identified latent trajectories. Class membership was used for analysis as the outcome variable in subsequent ML predictive modeling and graph analyses. The predictive accuracy of trajectories was compared to PTSD diagnostic status established through clinician assessment using the CAPS.

Prediction of trajectory membership. First, data were transformed and prepared for modeling (see Supplementary Materials for full description of procedures). Next, multiple models were built with data from progressive time points integrated to determine the accuracy of predicting trajectories. Specifically, predictive models were constructed including (a) pre-trauma exposure variables only (that is, demographics including age, gender, prior trauma history and so on); (b) pre-trauma + ER data; (c) pre-trauma + ER + 1-week data; (d) pre-trauma + ER + 1-week data + 1 month data. All models were also constructed on (1) clinical/demographic data alone; (2) neuroendocrine and physiological data alone; (3) clinical/demographic and neuroendocrine data together. Linear support vector machines (SVMs) with default parameterizations (that is, C parameter = 1) were utilized as the primary method to classify individuals into trajectories because prior comparisons across ML methods demonstrated that linear SVMs performed equivalently or outperformed other classification methods.[16] Specifically, SVMs identify a linear hyperplane in high-dimensional space (each predictor variable is a dimension) that accurately separates the sample into two or more previously specified populations.[38] For completeness, we also compare the performance of five other 'best in class' classification algorithms.[38] The results of that comparison are presented in Appendix A.[39,40]

SVM-based recursive feature elimination (RFE) was employed to select the smallest subset of features that most accurately classifies the outcome.[39,40] Feature selection algorithms, such as RFE, are commonly utilized in tandem with classification algorithms to reduce dimensionality and to guard against overfitting.
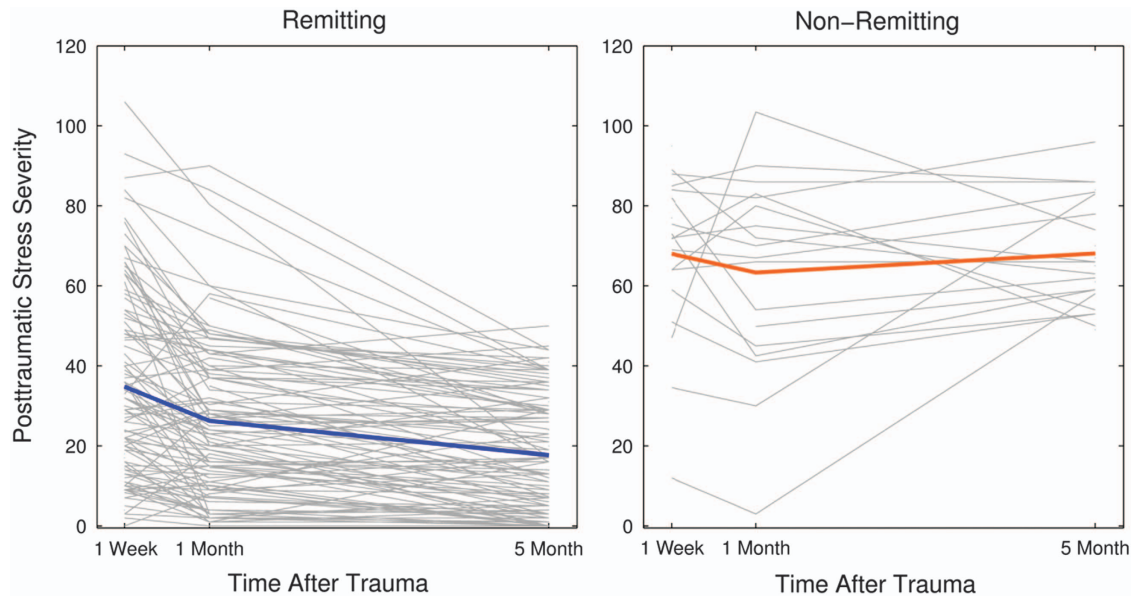
Figure 1. Latent growth mixture model identified trajectories of remitting and non-remitting PTSD. Two classes represented as the mean trajectories with random effects around LGMM identified trajectories of remitting (82.8%) and non-remitting (17.2%) post-traumatic stress based on repeated measures of the IES-R (*n* = 155). Gray lines represent individual observations. Colored lines represent the mean trajectory. IES-R, Impact of Events Scale Revised; LGMM, latent growth mixture modeling; PTSD, post-traumatic stress disorder.

Prediction and classification accuracy are provided as the mean area under the receiver operator characteristic curve (AUC) across cross-validation runs (see below) with RFE feature selection. Accuracy without RFE is presented in Appendix A. Results were cross-validated through 5×10-fold cross-validation to guard against overfitting resulting in a mean AUC across cross-validation runs. Cross-validation provides a method of guarding against false discovery akin to corrections for multiple comparisons, which is utilized in a null hypothesis-testing framework. As ML methods are not based on a null hypothesis framework, guards against overfitting are achieved by replication in random parts of the data (and ultimately in independent data sets). AUC of 0.50–0.60 indicates prediction at chance; 0.60–0.70: poor prediction; 0.70–0.80: fair prediction; 0.80–0.90: good prediction; 0.90–1.0: excellent prediction[41] (for a full description of SVMs, RFE and cross-validation procedures see Supplementary Materials). In addition, we examined sensitivity, specificity and positive predictive validity (PPV) at different cutpoints (along with 95% confidence intervals (CIs)) with and without the addition of feature types (clinical alone, neuroendocrine alone, clinical and neuroendocrine combined). PPV was calculated at different cutpoints corresponding to the precision-recall curve, calculated as $\text{Precision} = \frac{tp}{tp+fp}$; $\text{Recall} = \frac{tp}{tp+fn}$, where tp = true positives, fp = false positives and fn = false negatives.[42]

This provides more direct information about how the algorithm is performing with response to detecting true-positive cases of non-remitting PTSD.

Graph model of relationships between all study variables. A network encompassing all probabilistically relevant features was identified (global network) by inducing sets of adjacencies around each feature (local network). This analysis was executed using the GLL-PC algorithm.[12,43] Specifically, a network encompassing all non-redundant and statistically relevant features was identified (global network) by inducing sets of adjacencies around each feature (local network) using the GLL-PC algorithm[12,43] for *local-to-global* learning HITON-PC instantiated through the Causal Explorer Toolkit[44] in MATLAB R2914b,[45] which identifies the minimal set of variables associated with an outcome variable of interest (target variable), conditioned on which all other variables are probabilistically independent. This approach has been shown to have favorable properties for inductive analysis including a low false discovery rate.[43] Edges are defined as lines that connect variables in a graph. Directed edges signify the causal relationships between variables.[46] The current study only uses directed edges between variables that occur at different times. Undirected edges, indicating a causal relationship of unknown directionality, are utilized cross-sectionally. Data with ⩾ 30% missing values for pairwise comparisons were dropped from analyses. Models were run

across 100 bootstrap draws and only edges that replicated in ⩾ 30% of bootstrap draws were retained. We sought to induce the global network encompassing all relationships between all available subject information to accurately identify the local network around PTSD trajectory membership.

RESULTS

Symptom trajectory results

Progressive nested trajectory models with freely estimated intercept and slope parameters were compared. A two-class model demonstrated a large reduction in the information criteria ($\text{AIC}_{\text{reduction}} = 32.81$; $\text{BIC}_{\text{reduction}} = 23.48$; $\text{SSABIC}_{\text{reduction}} = 32.98$), a highly significant LMR-LRT ($P \leqslant 0.001$), Vuong-Lo-Mendell-Rubin likelihood ratio test ($P \leqslant 0.01$) and Bayesian likelihood ratio test ($P \leqslant 0.001$) and good entropy (0.87). The addition of a third class resulted in minor reductions in the information criteria ($\text{AIC}_{\text{reduction}} = 15.06$; $\text{BIC}_{\text{reduction}} = 5.72$; $\text{SSABIC}_{\text{reduction}} = 15.21$), non-significant LMR-LRT ($P = 0.19$) and Vuong-Lo-Mendell-Rubin likelihood ratio test ($P = 0.21$), while the Bayesian likelihood ratio test remained highly significant ($P \leqslant 0.001$). Entropy decreased slightly (0.85). Next, solutions for the two random splits of the data were estimated, demonstrating that in both random splits a two-class solution best fit the data (Two Class solution for Random Split 1: LMR-LRT $P = 0.052$; Three Class solution for Random Split 1: LMR-LRT $P = 0.33$; Two Class solution for Random Split 2: LMR-LRT $P = 0.001$; Three Class solution for Random Split 2: LMR-LRT $P = 0.11$). On the basis of converging evidence, we retained the two-class model (Figure 1).

The two identified trajectories included: *Recovery* (82.8% of the sample) was characterized by an initial intercept at the 1-week time point, approximating the clinical cutoff for PTSD diagnostic status ($E_{\text{st}} = 32.38$; s.e. = 2.07; $P \leqslant 0.001$), a significant negative slope over time ($E_{\text{st}} = -3.83$; s.e. = 0.36; $P \leqslant 0.001$) and a high average latent class probability for most likely latent class membership (0.97), and *non-remitters* (17.2%) demonstrated an initial intercept in the severe range ($E_{\text{st}} = 62.47$; s.e. = 4.91; $P \leqslant 0.001$), a nonsignificant slope indicating no significant change in PTSD scores over time ($E_{\text{st}} = 0.47$; s.e. = 1.41; $P = 0.74$) and a high latent class probability for most likely latent class membership (0.92). Classes were compared on gender, age, depression, global

4

Table 1. Comparison of individuals in LGMM identified symptom trajectory classes on clinical variables at 1 and 5 months

Class demographics/ symptoms	Remitting mean (s.d.)	Non-remitting mean (s.d.)
Demographics		
Age	30.75 (10.71)	32.07 (12.67)[*]
Gender (male)	59.12%	50.00%[*]
1 Month		
BDI	6.00 (5.98)	17.80 (12.24)
GAF	79.99 (12.46)	62.21 (12.36)
PTSD total	19.51 (15.73)	58.22 (29.12)
5 Months		
BDI	4.81 (5.54)	19.16 (13.43)
GAF	79.99 (12.46)	62.21 (12.36)
PTSD total	12.67 (15.53)	62.67 (30.57)

Abbreviations: BDI, Beck Depression Inventory; CAPS IV, Clinical Assessment of Posttraumatic Stress IV; DSM, Diagnostic and Statistical Manual of Mental Disorders; GAF, Global Assessment of Functioning; LGMM, latent growth mixture modeling; PTSD, post-traumatic stress disorder; SCID, Structured Clinical Interview for DSM-IV. Note: Depression was measured using the BDI-II. PTSD symptom frequency and intensity were measured using the CAPS IV cued to DSM-IV criteria; GAF is based on the SCID. All comparisons were conducted using one-way analysis of variance with the exception of the frequency of gender, which was assessed using χ^2. All comparisons were significant at the $P < 0.001$ level with the exception of those which did not approach significance, which are marked [*].

functioning and PTSD severity measured with the CAPS using one-way analysis of variance or χ^2-test for categorical variables. Results indicated that classes did not differ on demographics but non-remitters demonstrated elevated PTSD and depression scores and lower functioning scores at all time points (Table 1).

Predictive model results

Prediction of symptom trajectories. SVMs provided better than chance predictive accuracy based on background information alone (AUC = 0.64 (CI = 0.61–0.67)), good accuracy based on background and ER variables (AUC = 0.82 (CI = 0.80–0.85)), good based on data through 1 week (AUC = 0.88 (CI = 0.86–0.91)) and excellent accuracy for data through 1 month (AUC = 0.93 (CI = 0.91–0.95)). Neuroendocrine and physiological features alone provided the following predictive accuracy: ER alone (AUC = 0.67 (CI = 0.63–0.71)); ER + 1 week (AUC = 0.66 (CI = 0.61–0.70)); and ER + 1 week + 1 month (AUC = 0.70 (CI = 0.67–0.74)). Clinical/demographic variables alone provided the following predictive accuracy: ER alone (AUC = 0.77 (CI = 0.72–0.80)); ER + 1 week + 1 month (AUC = 0.90 (CI = 0.88–0.92)). These results are confirmed by examining sensitivity and specificity at different thresholds, which demonstrates that at moderate sensitivity (0.70) all features together from the ER provide relatively high specificity (0.75 (CI = 0.71–0.79)), while neuroendocrine features alone provide weaker specificity (0.58 (CI = 0.53–0.64)) as do clinical features alone (0.70 (CI = 0.63–0.77)). Combined features demonstrate improved performance over clinical features alone and neuroendocrine features alone through 1 month (see Supplementary Table 1). PPV corresponding to the precisio–recall curve revealed that combined features at 1 month at 0.70 sensitivity provided the strongest model (0.70 (CI = (0.62–0.77); for compete results see Supplementary Table 2). Results indicate that neuroendocrine information improves predictive ability from the ER but provides only trivial predictive over directly observed clinical indicators beyond the ER (Figure 2).

To determine the contribution of specific features within a multivariate predictive model, we outputted a measure of the percentage of times any given feature was selected through RFE across 5 × 10-fold cross-validation runs. This was done for all of the above models to determine features' stability as predictors in the context of other sources of information. We define a stable feature as one that is selected in ⩾ 55% for RFE runs. Importantly, feature stability changed depending on what other features were in the model (all sources of information together; biological alone; clinical alone) and what time epochs were included in the model (ER alone; +1 week; +1 month), indicating complex dependencies between features. Features that were stable across all models included prior trauma exposure and heightened sense of danger and urinary cortisol collected in the ER.

For completeness, three additional sets of analyses were conducted and are presented in Supplementary Material: (1) all SVM-based predictive models were conducted using end point PTSD diagnostic status to determine whether empirical trajectories or diagnosis are more amendable to predictive modeling; (2) predictive results for trajectories are compared to other common classification algorithms for completeness (Supplementary Figure 1); (3) the stability of all features across all models (Supplementary Figures 2–4). Results comparing different parameters (that is, linear compared to polynomial parameters) did not improve fit, indicating that nonlinear models did not provide better fit.

Graph analysis

The local graph, or direct connections, around a particular variable (terms the target variable) makes the target variable statistically independent of all other variables in the graph (see Figure 3a for the local-to-global graph around PTSD non-remission). The graph model reveals four features included in the local graph around trajectory membership. Included are two features from the 5-month time epoch (PTSD symptom severity and the NE/cortisol ratio measured in plasma. This indicates that these two features are most closely associated with non-remission of PTSD. Depression was shown to be associated with trajectory membership indirectly though PTSD severity.

Further features in the local graph around trajectory membership included ER hourly urinary cortisol and avoidance symptomatology at 1 month. The connections to these features revealed two key pathways to trajectory membership (see Figure 3b). Low levels of ER urinary cortisol were dependent on self-reported childhood trauma exposure. The second pathway through avoidance symptomatology at 1 month was dependent on avoidance at 1 week, indicating that this symptom cluster is impactful early in the disorder development.

An examination of the larger graph reveals that the emergence of avoidance symptomatology was dependent on individual's subjective assessment of the severity of the trauma, which in turn was dependent on increased salivary cortisol, the NE/cortisol ratio measured in plasma, injury severity and multiple subjective assessments of the event. This pathway was further shown to be associated with increased pulse, HR and plasma NE levels, and was less likely to be dependent on childhood trauma. Increased plasma NE was found to be negatively associated with glucocorticoid receptor cell receptor density, indicating that regulation of the hypothalamic pituitary axis is relevant for individuals on this pathway as well as those on the other pathway. The two pathways intersect through ER hourly urinary NE. Together, results indicate that factors across time epochs have an impact on the development and maintenance of PTSD.

Post hoc confirmatory analysis

To corroborate the identified graph relationship through traditional methods, we first compared the mean differences between the two classes (remitting/non-remitting) through independent samples t-test, demonstrating significant differences in ER cortisol
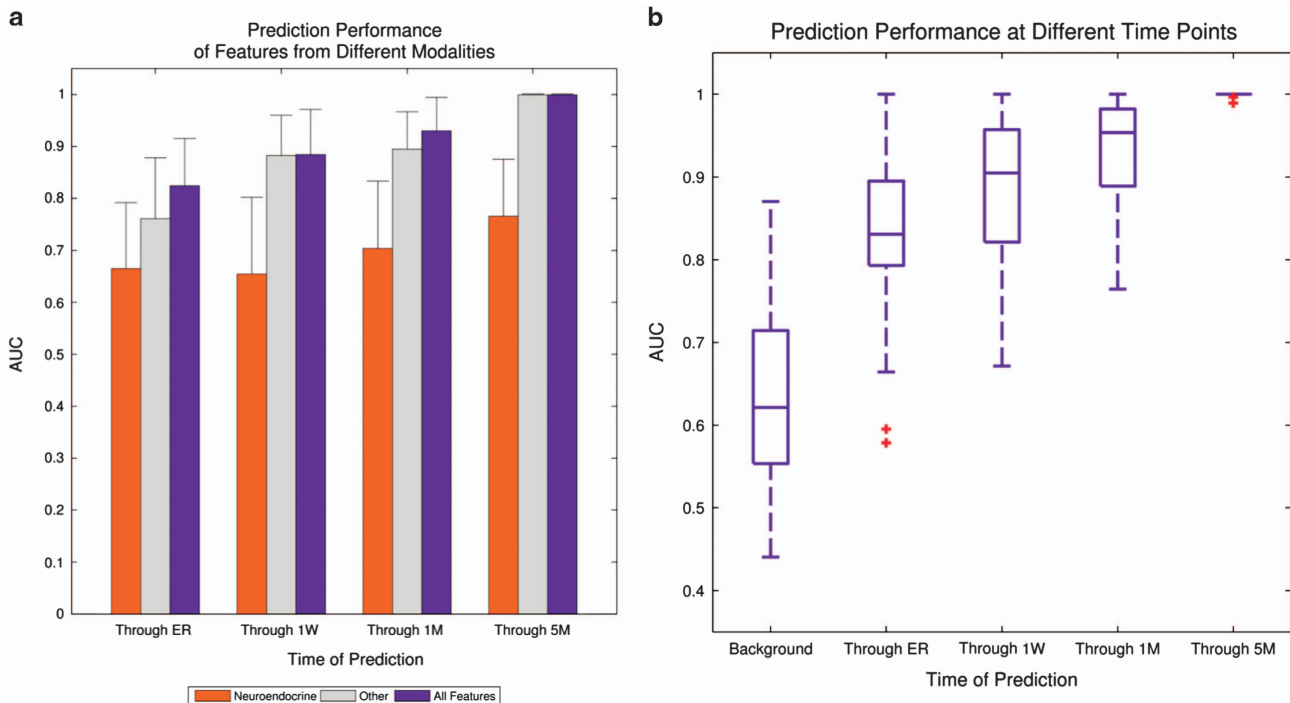
Figure 2. Classification results based on mean AUC for support vector machines with recursive feature elimination across 5 × 10-fold cross-validation. (**a**) Classification accuracy based on AUCs for (1) neuroendocrine features alone; (2) clinical and demographic features alone; (3) neuroendocrine, clinical and demographic features together. Each time point represents the AUC inclusive of features from the previous time point. (**b**) Means and s.d.'s of AUCs across 5 × 10-fold cross-validation runs. AUC, area under the receiver operator characteristic curve.
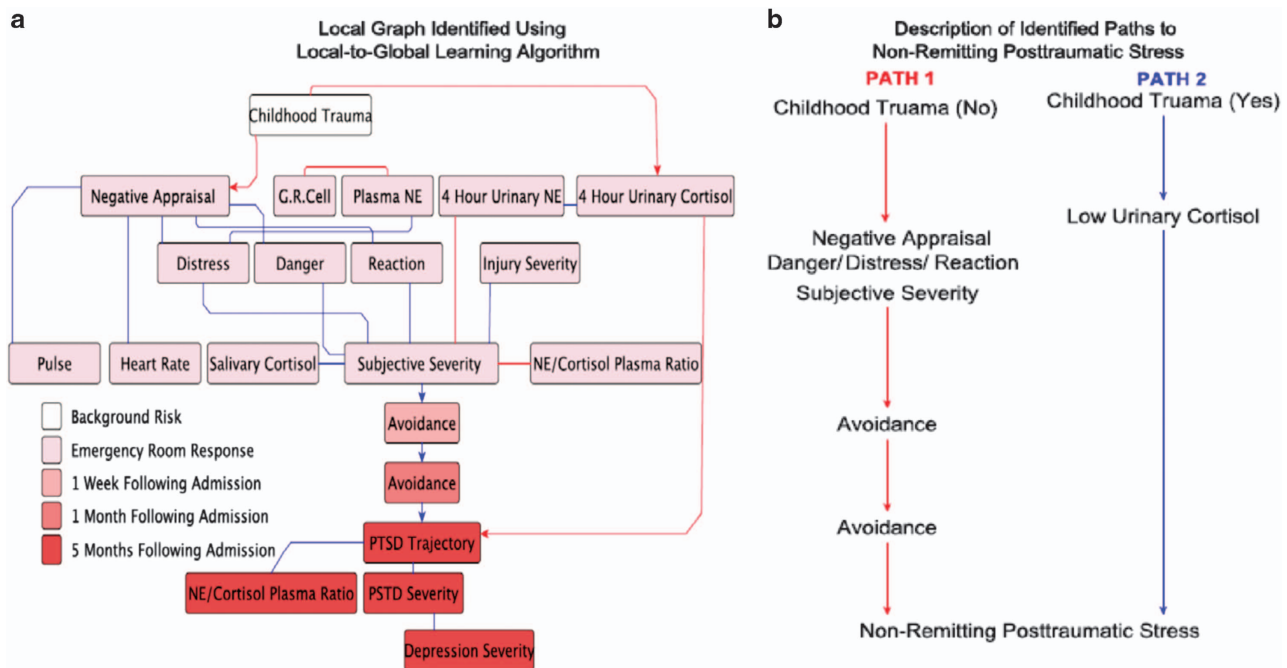
Figure 3. Causal graph around non-remitting PTSD trajectory membership derived using local-to-global learning algorithm. Note: see full description of features in methods section. (**a**) In graph, red lines represent negative relationships and blue lines represent positive relationships. The graph represents the local network surrounding the target variable PTSD Trajectory, identified using the HITON-PC algorithm for local-to-global learning. The local network represents the set of variables that renders all other variables in the model statistically independent of the target variable. The local network includes 4-h Urinary Cortisol, avoidance symptoms at 1 month, NE/cortisol plasma ratio and PTSD severity at 5 months. Expanding the network reveals that trajectory membership is associated with depression via PTSD severity. Further, two pathways to non-remitting post-traumatic stress are identified. (**b**) Path 1 indicates that individuals who do not report childhood trauma experience high sympathetic arousal and negative appraisals in the emergency room leading to the emergence of avoidance symptomatology at 1 week and 1 month, leading to non-remitting PTSD trajectory membership. Path 2 indicates that report of childhood trauma has a causal effect on PTSD non-remission through low urinary cortisol response in the emergency room. The two pathways are connected through 4-h Urinary NE. NE, norepinephrine; PTSD, post-traumatic stress disorder.

6

by class (t (1,96) = − 3.23; $P < 0.002$; mean (s.d.) $_{remitting}$ = 5.82 (4.79); mean (s.d.)$_{non-remitting}$ = 12.31(14.51)). We then examined differences in ER urinary cortisol by childhood trauma as a mean (s.d.) $_{ECT}$ = 4.81 (4.78); mean (s.d.)$_{non-ECT}$ = 7.49 (7.94)). Finally, χ^2-test of independence demonstrated a significant relationship between early childhood trauma and non-remitting trajectory membership (χ^2 (1,165) = 4.70; $P < 0.05$). Finally, logistic regression with class membership as the dependent variable and early childhood trauma and ER cortisol as independent variables revealed a significant overall model (χ^2 (2,96) = 7.92; $P < 0.05$), a nonsignificant effect of early childhood trauma (Wald = 0.66; $P = 0.41$) and a significant effect of ER urinary cortisol (Wald = 5.57; $P = 0.02$). These findings confirm that the relationship between childhood trauma and trajectory membership is fully mediated by the ER cortisol.

Finally, as both predictive and graph models revealed a roll for hourly cortisol measured in the ER, we examined the univariate classification accuracy of these variables for differentiating identified trajectories by calculating the AUC for hourly cortisol alone. Results revealed that this variable alone does not classify trajectory membership better than chance (AUC = 0.58), indicating that measurements of cortisol in the ER alone will not provide adequate classification accuracy to identify those at risk.

DISCUSSION

The current findings demonstrate that the neuroendocrine response contributes to an accurate predictive signal of PTSD trajectory of response to trauma. Further, cortisol provides a stable predictive signal when measured in conjunction with other related neuroendocrine and clinical sources of information. This indicates that the neuroendocrine response to trauma is an important biomarker panel to capture for predictive tools.

Network analyses revealed that decreased levels of cortisol was associated with increased risk for PTSD non-remission. As this measurement is an average over 4 h after exposure, this result can be interpreted to have reduced elevation in response to the trauma rather than lower diurnal levels. The reduced cortisol response in the ER was dependent on report of early childhood trauma exposure. This finding is consistent with evidence from animal models that early life abuse produces a blunted cortisol response to trauma[3] and findings that a blunted cortisol response to trauma is predictive of abnormal stress responses.[21] Further, this finding points to genomic and molecular mechanisms for further investigation as abnormal cortisol responses to trauma and PTSD are associated multiple genetic, epigenetic and proteomic elements including FKBP5 among others,[47] many of which can serve as accessible therapeutic targets in emergency medical contexts including military personnel deployed to war zones.

A second general pathway was identified that was probabilistically less likely to be associated with childhood trauma. This pathway consisted of a general sense of heightened arousal and danger in response to the trauma. This response influenced the development of avoidance symptoms and ultimately PTSD non-remission. Interestingly, this pathway was associated with elevated salivary cortisol, indicating that cortisol levels may confer differential risk for individuals on different pathways with distinct background risk. This possibly explains contradictory findings regarding cortisol response to trauma and risk for PTSD as results show that the same marker has different effects based on how it is measured as well as the population it is related to.

Both network and classification algorithms indicate that the interaction between multiple individual factors confers risk for non-remission. Classification models consistently selected multiple features to build an accurate classifier, indicating complex dependencies between features in classifying trajectory membership. Network models further provided evidence of complex dependencies and interactions between variables such as the

identified interaction between childhood trauma and low cortisol in the ER integrating to increase the propensity for non-remission. Knowledge of these pathways may be directly relevant to clinical care of individuals identified following trauma in the general medical ER. Specifically, cortisol-based interventions have been explored as a potential method to prevent the development of PTSD.[48] The current results indicate that such interventions may only be relevant to a subset of individuals, particularly those who report early childhood trauma, indicating that such interventions may need to be targeted to particular subpopulations to be effective.

It is important to note that ML-based network models are an inherently exploratory data analytic method, and as such might be seen as 'hypotheses generating'. While such an approach is informative in situations where complex relationships cannot be proposed and tested *a priori*, such an approach also presents with inherent limitations as a high number of relationships are estimated simultaneously introducing a non-trivial probability of false discovery. To overcome this in the current study, the key relationships of interest were confirmed in a more rigorous null hypothesis framework through the use of logistic regression. While this improves confidence in the findings, both the graph and logistic regression results are bound to the current sample and will require further validation in independent samples.

Limitations to implementing systematic neuroendocrine assessment following trauma exposure are its cost, the small increment in predicting PTSD over initial symptom severity measure and its yet unknown sensitivity and specificity. As such, this work does not sustain a use of early neuroendocrine responses as risk indicators, but rather as a way to identify likely candidates for early neuroendocrine interventions and to understand current intervention-limited effect size in unselected aggregates.

Despite these limitations, this work identifies the substantive role of cortisol and the broader neuroendocrine response in predictive modeling of post-traumatic stress responses. Further, this work identifies subpopulations for which the cortisol response following trauma is most relevant. Together, results indicate that the neuroendocrine response to trauma should be both utilized as a biomarker in predictive tools and should be considered as a target for intervention that is most relevant to a subpopulation. Further, this work provides a methodology that is relevant across psychiatry and other behavioral sciences that transcend the limitations of commonly utilized data analytic tools to match the complexity of the current state of theory in these fields.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1 Brewin CR, Andrews B, Valentine JD. Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *J Consult Clin Psychol* 2000; **68**: 748–766.

2 Ozer EJ, Best SR, Lipsey TL, Weiss DS. *Predictors of PTSD Symptoms in Adults: a Meta-analysis*. International Society for Traumatic Stress Studies: Washington, DA, USA, 1998.

3 Michopoulos V, Norrholm SD, Jovanovic T. Diagnostic biomarkers for posttraumatic stress disorder: promising horizons from translational neuroscience research. *Biol Psychiatry* 2015; **78**: 344–353.

4 Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics* 2006; **2**: 59.

5 Glover EM, Jovanovic T, Mercer KB, Kerley K, Bradley B, Ressler KJ *et al*. Estrogen levels are associated with extinction deficits in women with posttraumatic stress disorder. *Biol Psychiatry* 2012; **72**: 19–24.

6 Ressler KJ, Mercer KB, Bradley B, Jovanovic T, Mahan A, Kerley K et al. Post-traumatic stress disorder is associated with PACAP and the PAC1 receptor. Nature 2011; 470: 492–497.

7 Binder EB, Bradley RG, Liu W, Epstein MP, Deveau TC, Mercer KB et al. Association of FKBP5 polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults. J Am Med Assoc 2008; 299: 1291–1305.

8 Klengel T, Mehta D, Anacker C, Rex-Haffner M, Pruessner JC, Pariante CM et al. Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. Nat Neurosci 2013; 16: 33–41.

9 Sawamura T, Klengel T, Armario A, Jovanovic T, Norrholm SD, Ressler KJ et al. Dexamethasone Treatment Leads to Enhanced Fear Extinction and Dynamic Fkbp5 Regulation in Amygdala. Neuropsychopharmacology 2016; 41: 832–846.

10 Pearl J. Causality: Models, Reasoning and Inference. Cambridge University Press: Cambridge, UK, 2000.

11 Spirtes P, Glymour C, Scheines R. Causation, Prediction, and Search. The MIT Press: Cambridge, MA, 2000.

12 Aliferis CF, Statnikov A, Tsamardinos I, Mani S, Koutsoukos XD. Local causal and markov blanket induction for causal discovery and feature selection for classification part I: algorithms and empirical evaluation. J Mach Learn Res 2010; 11: 171–234.

13 Ray B, Henaff M, Ma S, Efstathiadis E, Peskin ER, Picone M et al. Information content and analysis methods for multi-modal high-throughput biomedical data. Sci Rep 2014; 4: 4411.

14 Galatzer-Levy IR, Bryant RA. 636,120 Ways to have posttraumatic stress disorder. Perspect Psychol Sci 2013; 8: 651–662.

15 Galatzer-Levy I. Empirical characterization of heterogeneous posttraumatic stress responses is necessary to improve the science of posttraumatic stress. J Clin Psychiatry 2014; 75: e950.

16 Galatzer-Levy IR, Karstoft K-I, Statnikov A, Shalev AY. Quantitative forecasting of PTSD from early trauma responses: a machine learning application. J Psychiatr Res 2014; 59: 68–76.

17 Bonanno GA, Kennedy P, Galatzer-Levy IR, Lude P, Elfstom ML. Trajectories of resilience, depression, and anxiety following spinal cord injury. Rehabil Psychol 2012; 57: 236–247.

18 Galatzer-Levy IR, Ankri Y, Freedman S, Israeli-Shalev Y, Roitman P, Gilad M 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 2013; 8: e70084.

19 Galatzer-Levy IR, Bonanno GA. Beyond normality in the study of bereavement: Heterogeneity in depression outcomes following loss in older adults. Soc Sci Med 2012; 74: 1987–1994.

20 Galatzer-Levy IR, Madan A, Neylan TC, Henn-Haase C, Marmar CR. Peritraumatic and trait dissociation differentiate police officers with resilient versus symptomatic trajectories of posttraumatic stress symptoms. J Traumatic Stress 2011; 24: 557–565.

21 Galatzer-Levy IR, Steenkamp MM, Qian M, Inslicht S, Henn-Haase C, Otte C et al. Cortisol response to an experimental stress paradigm prospectively predicts long-term distress and resilience trajectories in response to active police service. J Psychiatr Res 2014; 56: 36–42.

22 Bonanno GA, Mancini AD, Horton JL, Powell TM, LeardMann CA, Boyko EJ et al. Trajectories of trauma symptoms and resilience in deployed US military service members: prospective cohort study. Br J Psychiatry 2012; 200: 317–323.

23 Muthén B. The potential of growth mixture modelling. Infant Child Dev 2006; 15: 623–625.

24 Videlock EJ, Peleg T, Segman R, Yehuda R, Pitman RK, Shalev AY. Stress hormones and post-traumatic stress disorder in civilian trauma victims: a longitudinal study. Part II: the adrenergic response. Int J Neuropsychopharmacol 2008; 11: 373–380.

25 Shalev AY, Videlock EJ, Peleg T, Segman R, Pitman RK, Yehuda R. Stress hormones and post-traumatic stress disorder in civilian trauma victims: a longitudinal study. Part I: HPA axis responses. Int J Neuropsychopharmacol 2008; 11: 365–372.

26 Blake DD, Weathers FW, Nagy LM, Kaloupek DG, Gusman FD, Charney DS et al. The development of a clinician-administered PTSD scale. J Traumatic Stress 1995; 8: 75–90.

27 First M, Spitzer R, Williams J, Gibbon M. Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV), 4th edn, New York State Psychiatric Institute, Biometrics Research: New York, NY, USA, 1996.

28 Weiss DS. The impact of event scale: revised. In: Wilson JP, So-Kum T, Catherine C (eds). Cross-cultural Assessment of Psychological Trauma and PTSD. Springer: New York, NY, 2007, pp 219–238.

29 Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961; 4: 561–571.

30 Maruish ME. The use of psychological testing for treatment planning and outcomes assessment. In: Spielberger CD, Sydeman SJ, Owen AE, Marsh BJ, Maruish ME (eds). Measuring Anxiety and Anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI), 2nd edn. Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 1999, pp 993–1021.

31 Marmar CR, Weiss DS, Metzler TJ The peritraumatic dissociative experiences questionnaire. In: Wilson JP, Keane TM (eds). Assessing Psychological Trauma and PTSD: a Handbook for Practitioners. Guilford Press: New York, NY, USA, 1997, pp 412–428.

32 Hooper L, Stockton P, Krupnick J, Green B. Development, use, and psychometric properties of the Trauma History Questionnaire. Journal of Loss and Trauma 2011; 16: 258–283.

33 Muthen LK, Muthen BO. Mplus User's Guide. 3rd edn, Muthen & Muthen: Los Angeles, CA, USA, 1998.

34 Nylund K, Asparouhov T, Muthén B. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equation Model 2007; 14: 535–569.

35 Nylund. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equation Model 2007; 14: 535–569.

36 Muthen B. Statistical and substantive checking in growth mixture modeling: comment on Bauer and Curran (2003). Psychol Methods 2003; 8: 369–377, discussion 384-393.

37 Guerra-Peña K, Steinley D. Extracting spurious latent classes in growth mixture modeling with nonnormal errors. Educ Psychol Meas 2016, 0013164416633735.

38 Statnikov A, Aliferis CF, Hardin DP, Guyon I. A Gentle Introduction to Support Vector Machines in Biomedicine: Case Studies. World Scientific Publishing Co., Inc.: River Edge, NJ, USA, 2011.

39 Statnikov A, Aliferis CF, Hardin DP. A Gentle Introduction to Support Vector Machines in Biomedicine: Theory and Methods. World Scientific Publishing Co., Inc.: River Edge, NJ, USA, 2011.

40 Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine Learn 2002; 46: 389–422.

41 Swets JA. Measuring the accuracy of diagnostic systems. Science 1988; 240: 1285–1293.

42 Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd international conference on Machine learning; 25–29 June 2006; Pittsburgh, PA, USA. ACM, New York, NY, USA, 2006, pp 233-240.

43 Aliferis CF, Statnikov A, Tsamardinos I, Mani S, Koutsoukos XD. Local causal and markov blanket induction for causal discovery and feature selection for classification part II: analysis and extensions. J Mach Learn Res 2010; 11: 235–284.

44 Aliferis CF, Tsamardinos I, Statnikov AR, Brown LE. Causal explorer: a causal probabilistic network learning toolkit for biomedical discovery. METMBS 2003, 371–376.

45 Guide MUs. The MathWorks Inc. Mathworks: Natick, MA, 1998; 4: 382.

46 Spirtes P, Glymour C, Scheines R. Causation, Prediction, and Search. The MIT Press: Cambridge, MA, USA, 2001.

47 Boscarino JA, Erlich PM, Hoffman SN, Zhang X. Higher FKBP5, COMT, CHRNA5, and CRHR1 allele burdens are associated with PTSD and interact with trauma exposure: implications for neuropsychiatric research and treatment. Neuropsychiatr Dis Treatment 2012; 8: 131.

48 Kearns MC, Ressler KJ, Zatzick D, Rothbaum BO Early interventions for PTSD: a review. Depress Anxiety 2012; 29: 833–842.

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)

APPENDIX A

Probability of each feature being selected as information from progressive time points is integrated based on all available information

	Probability of feature being selected				
Features	Pre-ER	ER	Through 1 week	Through 1 month	Through 5 months
Background #cigarettes	0.7	0.5	0.18	0.14	
Adult trauma	0.34	0.2	0.24	0.28	
Age	0.76	0.22	0.14	0.12	
Car accident	0.5	0.18	0.14	0.46	
Child trauma	0.68	0.26	0.16	0.2	
Education	0.46	0.18	0.16	0.18	
Gender	0.34	0.32	0.2	0.4	
Holocaust	0.56	0.42	0.16	0.04	
Income	0.76	0.32	0.16	0.22	
Military service	0.26	0.2	0.4	0.28	
Other trauma	0.66	0.84	0.76	0.84	
Psych treatment	0.52	0.56	0.58	0.56	
Smoking	0.58	0.62	0.42	0.4	
A1 ER		0.16	0.28	0.3	
A2 ER		0.68	0.74	0.78	
Appraisal ER		0.92	0.78	0.88	
Arrival time ER		0.38	0.3	0.3	
Blood time ER		0.2	0.24	0.18	
Blood pressure ER		0.14	0.2	0.12	
Danger ER		0.4	0.52	0.36	
Distress ER		0.14	0.16	0.26	
GR cell ER		0.12	0.22	0.1	
Heart rate ER		0.48	0.14	0.16	
Lymphocytes ER		0.48	0.48	0.52	
NE/cort plasma ratio ER		0.12	0.14	0.2	
Subjective severity ER		0.54	0.46	0.48	
Plasma ACTH ER		0.18	0.36	0.24	
Plasma Cort ER		0.66	0.58	0.72	
Plasma NE ER		0.18	0.12	0.08	
Pulse ER		0.32	0.58	0.62	
Reaction ER		0.62	0.6	0.78	
Saliva Cort ER		0.16	0.22	0.32	
Severity ER		0.34	0.16	0.16	
Time from trauma ER		0.22	0.38	0.34	
Urine Cort ER		0.4	0.42	0.54	
12-h urine Cort ER		0.72	0.58	0.34	
Urine NE ER		0.2	0.1	0.08	
12-h NE ER		0.14	0.12	0.12	
Arousal W1			0.7	0.44	
Ascort W1			0.24	0.14	
Avoidance W1			0.22	0.22	
Blood time W1			0.36	0.28	
Bscort W1			0.34	0.44	
Cscort W1			0.32	0.28	
Depression W1			0.92	0.7	

	Probability of feature being selected				
Features	Pre-ER	ER	Through 1 week	Through 1 month	Through 5 months
Dscort W1			0.12	0.04	
Exposure W1			0.22	0.24	
GR cell W1			0.26	0.28	
Intensity W1			0.12	0.22	
Intrusion W1			0.44	0.3	
Lymphocytes W1			0.54	0.48	
NE/Cort plasma ratio W1			0.34	0.12	
PDEQ W1			0.46	0.22	
PEDS W1			0.14	0.12	
Plasma ACTH W1			0.22	0.26	
Plasma Cort W1			0.12	0.14	
Plasma NE W1			0.2	0.24	
PTSD score W1			0.26	0.08	
Reaction W1			0.14	0.12	
Saliva Cort W1			0.2	0.1	
State Anx W1			0.18	0.08	
Support W1			0.62	0.52	
Trait Anx W1			0.16	0.08	
Urine Cort W1			0.12	0.06	
12-h urine Cort W1			0.3	0.2	
Urine NE W1			0.38	0.34	
12 Hour Urine NE W1			0.14	0.1	
Arousal M1				0.34	
Avoidance M1				0.16	
Blood time M1				0.04	
Depression M1				0.62	
GR cell M1				0.16	
GAF M1				0.92	
Intrusion M1				0.38	
Lymphocytes M1				0.38	
NE/Cort plasma ratio M1				0.38	
Plasma ACTH M1				0.22	
Plasma Cort M1				0.06	
Plasma NE M1				0.56	
PTSD criteria M1				0.92	
PTSD score M1				0.42	
Reaction M1				0.2	
Saliva Cort M1				0.36	
State Anx M1				0.1	
Support M1				0.24	
Trait Anx M1				0.14	
Urine Cort M1				0.1	
12-h Urine Cort M1				0.08	
Urine NE M1				0.28	
12-h Urine NE M1				0.06	
Arousal M5					0.04
PTSD Score M5					1