Psychology of Violence

ISSN: 2152-0828

2025, Vol. 15, No. 4, 425–434
https://doi.org/10.1037/vio0000556

Factors Associated With Momentary Acts of Aggression: An Investigation Using Machine Learning Approaches in Ecological Momentary Assessment Data

Yongtian Cheng[1], K. V. Petrides[2], Anastasia Ushakova[3], Denis Ribeaud[4], Manuel Eisner[5], and Aja Murray[6]

[1] Department of Clinical, Health, and Educational Psychology, Division of Psychology and Language Sciences, University College London
[2] Department of Psychology and Psychometrics, University College London
[3] Department of Biostatistics, Lancaster University
[4] Jacobs Center for Productive Youth Development, University of Zurich
[5] Institute of Criminology, University of Cambridge
[6] Department of Psychology, University of Edinburgh

Objective: Model the associations between aggressive behavior and potential precursors. Little research exists that can illuminate the most proximal factors to momentary aggression as they occur in daily life and against the background of an individual's profile of relevant traits (e.g., their self-control levels). ***Method:*** This study used data from the combined longitudinal cohort and ecological momentary assessment (EMA) study, Decades-to-Minutes, with machine learning techniques to find the most important factors associated with "in-the-moment" aggressive behavior. Two types of models fitted by elastic net were examined: one with momentary data from the EMA component of the study and the other with both EMA and sociodemographic and trait data from the longitudinal survey component. ***Results:*** The best models fitted by elastic net achieved balanced accuracies of .76 and .79, while traditional methods achieved balanced accuracies of .63 and .64. ***Conclusions:*** Findings provide proof-of-concept evidence for the ability of elastic net to extract more important factors associated with aggression captured via short smartphone-based surveys and for the advantage of the elastic net method over stepwise regression for this purpose. The proposed models provide a step toward "in-the-moment" interventions to prevent aggressive behavior. Researchers are encouraged to apply the feature selection method used in this study for further research, such as exploring it in the context of smartphone applications for early prevention of aggressive behavior.

Keywords: aggressive behavior, momentary ecological assessment, elastic net, supervised machine learning, feature selection

Supplemental materials: https://doi.org/10.1037/vio0000556.supp

Aggressive behavior can cause severe physical (Chen et al., 2010) and psychological harm (Inoue et al., 2006; Richter & Berger, 2006). As such, finding the most important features associated with when and by whom aggressive behavior is perpetrated is important from a prevention perspective (Poldrack et al., 2018). Machine learning techniques have previously proven valuable in extracting the most important factors associated with momentary behaviors such as suicidality (Grendas et al., 2022) or diabetes management behaviors (e.g., self-monitoring of blood glucose; Zhang et al., 2022) and may thus also be promising for fitting models with high performance in identifying associations with momentary acts of aggression. In this study, we, therefore, applied machine learning techniques (i.e., elastic net) on ecological momentary assessment (EMA) data to extract the most important factors and to examine the associations of these factors with momentary acts of aggression based on an individual's background characteristics and momentary experiences. We also compared the logistic regression model with the elastic net model for this task.

EMA and Its Application

EMA is a method that uses repeated collection of near real-time data on participants' behavior and experiences in their natural environments (Shiffman et al., 2008). EMA is valuable for measuring a wide range of momentary symptoms and behaviors, including pain, mood, anxiety, bipolar disorder symptoms, and aggressive behavior (Thiele et al., 2002). For instance, Fried et al. (2022) conducted a study on the measurement of depressive symptoms, anxiety, and loneliness among students during the COVID-19 lockdown period in

This article was published Online First November 11, 2024.
Yongtian Cheng https://orcid.org/0009-0000-9569-6263
Denis Ribeaud https://orcid.org/0000-0002-1668-8319
Manuel Eisner https://orcid.org/0000-0001-5436-9282
Aja Murray https://orcid.org/0000-0002-9068-3188
The authors have no conflicts of interest to disclose. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
Correspondence concerning this article should be addressed to Aja Murray, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, United Kingdom. Email: Aja.Murray@ed.ac.uk

the Netherlands. Using a multilevel model, they observed decreases in anxiety, loneliness, and COVID-19-related concerns despite rapidly increasing rates of infections and deaths, with stress levels remaining stable. Testa et al. (2020) conducted an EMA study with couples who were asked to report perpetrating or experiencing physical or verbal aggression (e.g., yelling, making threats, insulting, pushing/grabbing/shoving, and throwing/kicking/hitting something). Their study found a nonsignificant relationship between depletion and aggressive behavior and a significant relationship between anger/arguing and aggressive behavior when analyzed with multivariate multilevel modeling using Bayesian estimation. Yeater et al. (2022) used an EMA design to test the correlations of sexual assault with factors such as regretted hookups, unprotected sex, drinking with peers, and peer-pressured sex. They found a significant correlation between these factors and sexual assault using generalized linear mixed models.

Machine Learning Models Examining Associations With Aggressive Behavior

When the goal is to fit a model that maximizes the associations between a set of factors and aggressive behavior, some machine learning techniques can offer advantages over traditional regression methods used in psychological research. In particular, machine learning methods can often provide stronger associations between independent variables (IVs) and dependent variable (DV) than traditional regression methods (Dwyer et al., 2018).

In previous studies, machine learning methods have been applied to examine associations with aggression. For example, Chatzakou et al. (2017) used several tree classifiers to categorize whether a Tweet is normal, aggressive, bullying, or spam. The decision tree is a type of model used in supervised learning that makes decisions by splitting data into subsets based on feature values, resembling a tree structure (Myles et al., 2004). Gutiérrez-Esparza et al. (2019) used the random forest method, which is based on multiple rule-based trees, to categorize cyberaggression and bullying in cyberspace. Random forests build multiple decision trees and combine their model results, aiming to improve performance and reduce overfitting by aggregating results (Cutler et al., 2012). Some studies have also addressed the identification of factors associated with real-world aggressive behavior. Hofmann et al. (2022) used a support vector machine, a method based on the division of a hyperplane, to classify outcomes like patient aggressive behavior during hospitalization using demographic data, childhood/youth experiences, psychiatric history, and other features in a group of 370 patients with schizophrenia spectrum disorders.

However, research such as the Hofmann et al. (2022) study has tended to focus only on the categorical classification of whether patients will engage in aggressive behavior during their hospitalization. This means their proposed models have not included the frequency and timing of aggressive behavior. Understanding the associations with momentary aggressive behavior can have considerable value for real-world prevention practices. For example, identifying when aggression is more likely to occur can be useful for preventing physical confrontations (Chen et al., 2010; Grumm et al., 2011; Poldrack et al., 2018) and psychological damage (Inoue et al., 2006; Richter & Berger, 2006).

To develop a model that can identify associations with momentary aggression, researchers need to identify and measure relevant feature sets (i.e., sets of IVs). There are some theories of aggressive behavior that can help guide this process. For example, the I^3 theory proposed by Finkel (2014) is well-suited for examining comprehensive associations of momentary aggression because of its focus on both background characteristics and momentary influences. According to the I^3 theory, there are three important features for understanding the likelihood of an individual's behavior (e.g., aggressive behavior) regarding a given target object in a particular context.

These features are instigation, impellance, and inhibition. Instigation is the net strength of the immediate environmental stimuli (e.g., an insult), impellance is the net strength of situational or dispositional qualities that influence how strongly the instigator for this individual in this situation fosters a proclivity to aggress (e.g., the presence of peers), and inhibition is the net strength of situational or dispositional qualities that influence how strongly the proclivity to enact an aggressive response manifests in aggressive behavior (e.g., self-control). According to the "perfect storm theory," which is derived from the I^3 model, the highest likelihood or intensity of behavior emerges when instigation and impellance are strong and inhibition is weak (Finkel, 2014; Finkel & Hall, 2018).

Numerous studies have employed this theory, adopting various methods to measure immediate environmental stimuli related to aggressive behavior. For example, the daily diary approach (Gunthert & Wenze, 2012) offers repeated assessments in natural settings that are close to real time. This method enables researchers to examine individuals' experiences, behaviors, and circumstances in their natural environments, relying on participant self-reports (Lischetzke & Könen, 2020). Several investigations using this approach have identified correlations between dating violence and alcohol consumption (Shorey, Stuart, McNulty, & Moore, 2014), dating violence and marijuana consumption (Shorey, Stuart, Moore, & McNulty, 2014), and domestic violence with alcohol consumption (Derrick & Testa, 2017).

In multiple studies demonstrating strong associations, machine learning has been applied to EMA data sets (e.g., Gee et al., 2020; Mikus et al., 2018). A number of supervised machine learning methods that train a model on a labeled data set are capable of modeling complex nonlinear relationships, making them suitable for tasks where data is not linearly separable, which leads to a better performance model than linear regression (Dwyer et al., 2018).

However, supervised machine learning methods do not outperform simple linear regression in all studies. In the task of identifying associations of depression using EMA Actiwatch data, Kim et al. (2019) found that logistic regression performed better than random forest, boosting trees, and decision trees. Therefore, though various machine learning methods have the potential to learn or simulate all features with high accuracy (Cybenko, 1989), only applying supervised machine learning methods to EMA data sets is not necessarily sufficient to achieve an optimal model. Researchers should include both models from machine learning methods and simple regression models in their studies.

The Present Study

The present study examines the association between momentary aggression and trait and state features. We hypothesize that (1) momentary aggression is associated with state measures alone; (2) a model incorporating both state and trait measures will show stronger associations with momentary aggression compared to a model using trait measures alone; and (3) the elastic net model fitting procedure will outperform the ordinary least squares model in terms of association strength and the number of features included in the model, providing more parsimonious models.

Method

Participants

Data for the study are drawn from the "Decades-to-Minutes" (D2M) EMA study (Murray, Speyer, et al., 2022) and *z*-proso longitudinal cohort study, within which D2M is embedded (Ribeaud et al., 2022). We will refer to the D2M EMA study as the "EMA study" and the *z*-proso main survey study as the "main study." The total target sample for the main study was 1,675, with 1,571 contributing data for at least one wave (Eisner et al., 2019). Participants for the embedded EMA were recruited following the age 20 main data collection wave of the main study. The EMA add-on was designed to provide a deeper understanding of the links between long-term developmental processes and short-term "momentary" processes. A convenience sampling method was employed with the goal of including as many of the same participants from the multiple substudies of the main study as possible. Two hundred sixty participants were recruited from the main study participants poll, with 255 having complete data on the most relevant measures. The remaining five participants were excluded from the data analyses.

Procedure

The main study is a longitudinal cohort study of development from childhood to adulthood based in Zurich, Switzerland. The *z*-proso study began in 2004 when participants were entering primary school at age 7 and was then followed up at ages 8, 9, 10, 11, 12, 13, 15, 17, and 20 (Ribeaud et al., 2022). Participants were selected at baseline using a stratified sampling procedure with schools as the sampling unit, and stratification was used to ensure adequate representation of schools in different geographical regions. In this study, we mainly used data from the age 20 main survey wave of the *z*-proso study. Additionally, some demographic information comes from the wave at age 13.

For the EMA study, through a mobile application, participants received a notification four times per day between 10 a.m. and 10 p.m. over a 14-day period that directed them to the EMA survey. Multiple repeated measures were collected from each participant.

As reported by Murray, Speyer, et al. (2022), compared with the main study sample, there were more females in the EMA subsample (62% female in EMA compared with 49% in the main sample), and their socioeconomic status based on the maximum household International Socio-Economic Index is significantly higher ($p < .001$) than in the main study cohort. They were also slightly lower in self-reported aggression based on age 20 aggression questionnaires, $t(516.7) = -2.92, p = .004$, and higher on stress, $t(440.48) = 2.78, p = .006$, but showed no difference in Attention deficit hyperactivity disorder symptoms, $t(434.85) = 1.40$, $p = .16$. There is also no significant difference between the EMA subsample and the main study sample in internalizing problems and alcohol use. Overall, previous studies using the data have concluded that the EMA subsample is only slightly selective with respect to some of the constructs addressed within the study based on such comparisons.

Materials

Trait Measures From the Age 20 Main z-Proso Survey

As mentioned above, for the main study data set, the total target sample was 1,675, with 1,571 contributing data for at least one wave (Eisner et al., 2019). We decided to consider all features to reach the maximum performance of the model, subject to the limitations of the main study data set. Therefore, we included the results of multiple psychometric scales in the main study, which are used as information on inhibition and impellance to examine associations with aggressive behavior. All of the psychometric scales were administered in Swiss German, reflecting the official language of the location of the study. A full list of psychometric scales from the main study included in this study can be found in Supplemental Materials named "supplementary table" with references to the publications detailing the development of these psychometric scales. There are 41 features used in the model fitting from the main study, which is made up of 214 items. All of the data from the main study are used as IVs (i.e., features) in this study.

Measures of Momentary States/Situations From the EMA Study

Twenty-three items from the EMA study were used in this study as the immediate environment/states (i.e., instigation and impellance in I^3 theory) to examine associations of aggressive behavior. Two EMA items were one-hot coded for analysis. Each question asked the participants to reflect on the last 30 min and was completed on the participant's own smartphone. All of the DVs and IVs in the EMA measurements were originally in Swiss German, reflecting the official language of the location of the study. For the details of the scales, please refer to the Supplemental Materials named "supplementary items information."

Aggression was measured by Aggression Experience Sampling Abbreviated in D2M (Murray, Eisner, et al., 2022), which captures momentary aggressive behavior in four items. Responses were recorded on a 4-point Likert-type scale from 1 (*strongly agree*), 2 (*agree*), 3 (*disagree*), to 4 (*strongly disagree*). Most of the responses (97.8%) were "strongly disagree" or "disagree." In the rest 2.2%, most of them were "agree" (76.8%). As a result, the items were coded such that if a respondent answered "agree" or "strongly agree" to any item at a given time point, indicating they engaged in some level of aggression, they were assigned a score of 1, whereas if they answered "disagree" or "strongly disagree," they were assigned a score of 0. A similar design has been used by other studies examining associations with aggressive behavior (e.g., McConville & Cornell, 2003). The binary variable created with this procedure was used as the DV of this study.

Several features were also derived from the EMA: context adapted from a study by Juslin and Västfjäll (2008), provocations (Borah et al., 2021), negative affect measured by Positive Affect and Negative Affect Scale Expanded (Watson & Clark, 1994), stress measured by Perceived Stress Scale (Cohen et al., 1983), and substance use. These features were also used in this study as IVs. Only the most common substances were included in the EMA since EMA is best suited to capturing relatively frequent events that occur within the short time frame (in this case, 2 weeks) of the study.

Data Analysis Plan

There were two stages of analysis. In the first stage, only the data set from the EMA study was included in the model fitting. In the second stage, the data set from the main study was also included in the model fitting. We expected that the performance of the model would be improved by including the extra impellance and inhibition features. Listwise deletion was used for data cleaning before further analysis. We chose not to apply a normality transformation to the data set because the original data are more interpretable (Lee, 2020). Additionally, we wanted to avoid potential reductions in the correlations between variables that such a normality transformation could cause (Qiu et al., 2005).

Sample Size Planning

Because this study was conducted after the data collection, sample size planning was not conducted in this study. The original sample size planning was based on a resource constraint approach (Murray, Speyer, et al., 2022).

However, the current analysis is justified in terms of adequate sample size. For sample size planning using logistic regression, events per variable (EPV) is an applied criterion for sample size planning (Austin & Steyerberg, 2015). It is calculated as rows of data divided by the number of the variables included. While most researchers view an EPV > 10 as providing a sufficient sample size (Peduzzi et al., 1996), some researchers believe an EPV > 30 is necessary for reliable performance (van Smeden et al., 2016). Stage 1 has an EPV of 289.34 for linear logistic regression model fitting. As suggested by the developer, the elastic net method needs a smaller sample size than a regression for a similar level of performance (Zou & Hastie, 2005). This is because the elastic net tends to include fewer variables in the model. In addition, the cross-validation procedure and training/testing division provide more reproducible quantitative results than using the full data set collected in a study (Yarkoni & Westfall, 2017).

After the listwise data-cleaning procedure, 58.9% of the data remained in the data set. There were 7,126 rows of data with DV = 0 and 501 rows of data with DV = 1 from 255 participants. Between participant divisions, 6,045 (79%) rows of data were assigned to the training data set, and the remaining 1,582 (21%) rows of data were assigned to the testing data set. This made the EPV = 95.95 for model fitting in the second stage of analysis.

Data Set Separation

The total data set was split into a training data set and a testing data set to ensure the model's generalizability and to avoid overfitting. Overfitting occurs when a model performs well on the data set used for fitting but poorly on an independent data set (e.g., a data set collected by independent researchers).

A common practice of data separation is to randomly let 80% of the data set be a training data set, and the remaining 20% be a testing data set (Joseph, 2022). In this study, some participants were thus randomly assigned to the training data set and the rest to the testing data set to ensure that all rows of each participant were assigned to either the training or testing data set but not both. With this design, we can ensure that the testing data set and training data set are independent of each other to avoid leakage (Cawley & Talbot, 2010).

Method Selection

The machine learning method of the elastic net proposed by Zou and Hastie (2005) was used in the model fitting. It is robust to multicollinearity (Altelbany, 2021). The loss function of the elastic net is:

$$\widehat{w^*_{\text{ElasticNet}}} = \text{Logloss} + \lambda_1 \left|\left|w\right|\right|_2^2 + \lambda_2 \left|\left|w\right|\right|_1, \tag{1}$$

in which:

$$\text{Log loss} = \sum_{(x,y)\in D} -y\log(y') - (1-y)\log(1-y'). \tag{2}$$

Log loss is the loss function of logistic regression, in which x is the set of features (IVs), y is the true value of the DV, and y' is the response of the DV. The $\lambda_1 \left|\left|w\right|\right|_2^2$ and $\lambda_2 \left|\left|w\right|\right|_1$ are penalty terms that penalize complexity. λ_1 and λ_2 are hyperparameters. The hyperparameters of elastic models were determined by 10-fold cross-validation, which is a common design for machine learning model fitting and model selection (Koul et al., 2018).

As previously mentioned, various machine learning methods, such as random forest and support vector machine, have been utilized to identify associations with aggressive behavior. These models might exhibit strong performance for this task. However, their interpretability is poorer. In contrast, the elastic net offers a straightforward model reminiscent of linear regression, as exemplified by Yoo (2018). Researchers can compute outcomes or associations using the elastic net model just like they would with a simple logistic regression. Researchers can also interpret the coefficients in the elastic net model just as they interpret the coefficients in the logistic regression model: An item with a coefficient of 1 in the elastic net regression model means a one-unit increase in the item will result in an increase of an exp(1) change in odds. This user-friendly nature of the elastic net is a primary rationale behind its selection over other machine learning methods.

Aside from providing results with high interpretability, this study also aims to provide a model that has fewer items. In EMA studies, shorter questionnaires have advantages in terms of reducing participant burden, dropout, and low-quality data (Wrzus & Neubauer, 2023). As previously mentioned, the elastic net allows a reduced, more parsimonious set, including only the most critical features, to be measured in future studies where space can be scarce.

Based on the reasons mentioned above, the elastic net was chosen as the machine learning model used in the study. In addition, a stepwise logistic regression was also included as a control method to make a comparison and test whether the elastic net can provide better performance than this commonly used method. The stepwise regression method is commonly used in research to identify associations with aggressive behavior (e.g., Ersan, 2020; Gómez-Leal et al., 2022; Lickiewicz et al., 2020) and therefore represents a suitable comparison method.

Rebalancing of Data Set

As the data were collected from a community sample, it is reasonable to assume the data may be imbalanced, given that aggression is relatively infrequent. In addition to utilizing the raw data set, both undersampling and oversampling methods were used in the training data set to rebalance the distribution of the DV. In undersampling methods, some cases with DV = 0 were excluded from the model training. In the oversampling method, the Synthetic Minority Oversampling Technique (SMOTE; Chawla et al., 2002) was used to simulate more cases with DV = 1.

Both resampling methods were only applied to the training data set and did not influence the testing data set. These two balancing methods, together with the analyses of the original data set, resulted in 2 × 3 = 6 fitted models with elastic net and logistic regression. Whether better model adequacy metrics are achieved with under- versus oversampling in the context of this study is not substantively important; however, both methods were explored with the goal of optimizing model adequacy and checking the consistency of findings across different approaches.

Metrics

In line with the goal of pursuing a high-accuracy model, we will present only the parameters from the model that demonstrates the strongest associations (i.e., maximum accuracy) on the testing data set. Traditional accuracy is not an apt metric for tasks with significant data imbalance because high accuracy can be achieved merely by always guessing that a case belongs to the majority class (e.g., did not aggress). Instead, we utilized balanced accuracy, sensitivity, and specificity as our evaluation metrics. Balanced accuracy offers a more comprehensive insight into a model's performance, especially in imbalanced binary classification scenarios (García et al., 2010). Balanced accuracy, calculated as the mean of sensitivity and specificity, can estimate the accuracy of a model in identifying associations within a population where researchers have no prior knowledge about the distribution of events (i.e., aggressive acts in daily life) in a population.

The area under the curve is also reported for the final model. However, it was only reported as a reference as it is biased under imbalanced data (Jeni et al., 2013; Saito & Rehmsmeier, 2015). All the metrics reported in this study are based on the model's performance on the testing data set. The models were compared based on their performance on the metrics, and the model with the best performance in each stage was selected.

Platform and Selection of Package

The analysis was conducted in R (R Core Team, 2013) and R Studio (RStudio Team, 2020). The caret package (Kuhn, 2022) was used to implement cross-validation and model fitting. DMwR (Torgo, 2016) was used for the confusion matrix analysis. The car package (Fox & Weisberg, 2019) was used for variance inflation factor (VIF) calculation, and the rms package (Harrell, 2022) was used for logistic regression model fitting. The glmnet package (Friedman et al., 2010) was used for elastic net model fitting. The package MLmetrics (Yan, 2016) was used for metric calculation. The package pROC (Robin et al., 2011) was used for area under the curve. The code is provided in Supplemental Materials titled "R code."

Results

In this section, we provide a comparison of the six models across two stages, including detailed information on the proposed model. Furthermore, we present the detailed results of stepwise logistic regression with no specific sampling method applied to contrast the machine learning models with common practice methods.

Results for Stage 1 Analysis

Table 1 presents the performance of the six models, including (1) the stepwise logistic regression model fitted using the original training data set; (2) the stepwise logistic regression model fitted using the SMOTE oversampling method on the training data set; (3) the stepwise logistic regression model fitted using undersampling on the training data set; (4) the elastic net model fitted using the original training data set; (5) the elastic net model fitted with the SMOTE oversampling method in the training data set; and (6) the elastic net model fitted with the undersampling in the training data set.

Model 1 served as the control in this study, and we anticipated that models employing machine learning methods would exhibit superior performance compared to the control model.

Both Models 2 and 3 utilized rebalanced training data sets using logistic regression without adjusting for the bias–variance trade-off. We anticipated that they might deliver a more balanced accuracy. However, we also expected the number of items in these models to remain consistent with the control model since no penalties for complexity are included in the fitting function.

On the other hand, Model 4 was designed to address the bias–variance trade-off. We expected that it would encompass fewer items than the control model but would achieve the same or marginally improved performance relative to the control.

Both Model 5 and Model 6 utilized rebalanced training data sets without adjusting for the bias–variance trade-off. We anticipated that they might deliver a more balanced accuracy. However, we also expected the number of items in these models to remain consistent with the control model. In other words, these two models were expected to have the best-balanced accuracy with fewer items than the control model and the other models.

In the first stage of analysis, only the 23 EMA items were included. Multinomial items are one-hot coded. After the listwise data-cleaning procedure, about 65.9% of the data remained in the data set. There were 7,990 rows of data reporting no aggressive behavior (i.e., DV = 0) and 544 rows of data reporting aggressive behavior (i.e., DV = 1) from 255 participants with the EMA design.

Table 1
Models Performance on Testing Data Set in the Stage 1 Analysis

Model	LG	LGS	LGU	EN	ENS[a]	ENU
Sensitivity	0.99	0.85	0.83	0.99	0.85	0.85
Specificity	0.27	0.68	0.65	0.27	0.68	0.64
Balanced accuracy	0.63	0.76	0.74	0.63	0.76	0.75
AUC	0.84	0.79	0.79	0.84	0.83	0.85

Note. LG = logistic regression model; LGS = logistic regression model fitted by the SMOTE oversampling method in the training data set; LGU = logistic regression model fitted by undersampling the training data set; EN = elastic net model fitted by the original training data set; ENS = elastic net model fitted by the SMOTE oversampling method in the training data set; ENU = elastic net model fitted by undersampling the training data set; AUC = area under the curve; SMOTE = Synthetic Minority Oversampling Technique.
[a] The model is selected as the final proposed model.

With participant divisions, 6,655 (78%) rows of data were assigned to the training data set, and the remaining 1,879 (22%) rows of data were assigned to the testing data set.

Both the training data set and testing data set were highly imbalanced. In the training data set, there were 6,228 rows of data with DV = 0 and 427 rows of data with DV = 1. With the SMOTE method, more aggressive behavior cases were simulated with IV features in the training data set, resulting in 6,228 rows of data with DV = 0 and 6,228 rows with DV = 1 for the rebalanced data set. With the undersampling method, some DV = 0 cases were excluded, resulting in 427 rows of data with DV = 0 and 427 rows of data with DV = 1 for the rebalanced data set. In the testing data set, there were 1,762 rows of data with DV = 0 and 117 rows of data with DV = 1.

Table 1 displays the results of the stepwise logistic regression for the original data set. The logistic regression provided a model with pseudo $R^2 = 0.343$, $\chi^2(13) = 926.50$, $p < .001$. All of the items have VIFs less than 5, which means that the multicollinearity is within a reasonable level. The balanced accuracy for this model on the testing data set is .6302.

Across all models, the highest balanced accuracies came from the logistic regression model fitted with SMOTE oversampling training data set, which has a balanced accuracy of .7635, and the elastic net model fitted with the data set with SMOTE oversampling training, which has a balanced accuracy of 0.7604.

On balance, two elastic net models were selected as the optimal models. This is because the performances of these two models are similar. Yet, SMOTE oversampling stepwise logistic regression uses less IVs in the model (22 vs. 35 after one-hot coding). The coefficient for SMOTE oversampling stepwise logistic regression and the coefficient of the proposed elastic net model are provided in Supplemental Materials.

Results for Stage 2 Analysis

In the second stage of analysis, the EMA and main study data sets were combined to implement a further round of models, following the same structure as the model fitting with only the EMA data. There were 63 IVs in this stage of analysis. Table 2 displays the performance of the six models, which is the same as the Stage 1 analysis, that is, the stepwise logistic regression model fitted using the original training data set, the stepwise logistic regression model fitted using the SMOTE oversampling in the training data set, the stepwise logistic regression model fitted using undersampling on the training data set, the elastic net model fitted using the original training data set, the elastic net model fitted with the SMOTE oversampling method in the training data set, and the elastic net model fitted with the undersampling training data set.

Table 2
The Models Performance on Testing Data Set in Stage 2 Analysis

Model	LG	LGS	LGU	EN	ENS	ENU[a]
Sensitivity	0.96	0.84	0.79	0.99	0.83	0.82
Specificity	0.32	0.57	0.66	0.23	0.67	0.76
Balanced accuracy	0.64	0.71	0.72	0.61	0.75	0.79
AUC	0.81	0.78	0.74	0.78	0.83	0.85

Note. LG = logistic regression model; LGS = logistic regression model fitted by the SMOTE oversampling method in the training data set; LGU = logistic regression model fitted by undersampling the training data set; EN = elastic net model fitted by the original training data set; ENS = elastic net model fitted by the SMOTE oversampling method in the training data set; ENU = elastic net model fitted by undersampling the training data set; AUC = area under the curve; SMOTE = Synthetic Minority Oversampling Technique.
[a] The model is selected as the final proposed model.

Both the training data set and testing data set were still highly imbalanced. In the training data set, there were 5,652 rows of data with DV = 0 and 393 rows of data with DV = 1. In the testing data set, there were 1,474 rows of data with DV = 0 and 108 rows of data with DV = 1. With the SMOTE method, more DV = 1 cases were simulated, resulting in 5,652 rows of data with DV = 0 and 5,652 rows of data with DV = 1 for the rebalanced data set. With the undersampling method, some IV = 0 cases were excluded, resulting in 393 rows of data with IV = 0 and 393 rows of data with IV = 1 for the rebalanced data set.

Table 2 displays the results for the traditional stepwise regression. The logistic regression provided a model with pseudo $R^2 = 0.411$, $\chi^2(32) = 1032.32$, and $p < .001$. All items have VIF less than 5, which means there were no issues with multicollinearity. The balanced accuracy for this model in the testing data set is .6417. Details of this model are presented in Supplemental Materials.

The elastic net model, fitted using undersampling on the training data set, showed the best-balanced accuracy of .7874 among all six models in the testing data set. In the best models, there are three trait measure items and seven state measure items. The three trait measures in the elastic net model are the Political and Religious Violence Scale (0.0474; Nivette et al., 2017), Psychopathy Scale (0.1180; Jones & Paulhus, 2014), and the Self-Reported Indirect Aggression Scale (0.0713; Tremblay et al., 1991). The state measure items in the elastic net model are (coefficients in parentheses) “In the last 30 min (I felt) … hostile” (−0.0970); “upset” (−0.3901); “unable to control the important things in my life” (0.0167); “someone insulted me” (−0.2586); “someone prevented me from doing something I wanted” (−0.2473); “I thought about a time when someone had annoyed me” (−0.4498); and “someone tried to start an argument with me” (−0.0363). The full details of all models are presented in the Supplemental Materials.

Discussion

This study explored the extent to which momentary aggressive behavior can be associated with data collected in the course of people’s daily lives using brief smartphone-based surveys, with and without background trait data.

As shown in Table 1, the immediate proximal factors are associated with momentary aggressive behavior. Therefore, Hypothesis 1, which states that momentary aggression is associated with state measures alone, is supported. Comparing Table 1 with Table 2, the set of models that used both immediate proximal factors and the traits of the participants had better performance than the set of models that used only immediate environmental stimuli. Therefore, Hypothesis 2, which states a model incorporating both state and trait measures will show stronger associations with momentary aggression compared to a model using trait measures alone, is supported. As we mentioned above, the elastic model with oversampling method has about the same performance (i.e., balanced accuracy = .76) as the logistic regression with the rebalancing method in Stage 1 analysis with fewer numbers of IV after one-hot coding (22 vs. 35). Based on the information provided in Table 2, the elastic model with undersampling has the best performance across all other models in Stage 2 analysis, with far fewer IVs (10) included in the model than logistic models. Based on these findings, Hypothesis 3, which states the elastic net model fitting procedure will outperform the ordinary least squares model in terms of association strength and the number of features included in the model, providing more parsimonious models, is generally supported.

Our analyses drew on existing established theories and were not focused on identifying novel factors associated with momentary aggression. Indeed, our findings replicated many associations identified in past research. Future research could potentially improve the extent to which the variation in momentary aggression can be explained by additionally examining novel risk factors.

The innovation in this study lies in the feature selection, which means the significance of a feature, even the significance of a feature with control on other features to the DV, does not have a close relationship with the inclusion/exclusion of the feature in a model finding the maximum association with the model.

When comparing the two models with only EMA items, we observed that including nonsignificant items from the logistic regression can enhance the performance of the elastic net-fitted model. For example, alcohol consumption, though not significant in logistic regression and hence excluded from that model, improved performance when included in identifying associations in an independent data set (i.e., testing data set). Conversely, when both EMA and main study IVs were considered, we found that including significant IVs from the logistic regression does not necessarily enhance the performance of an independent data set. For instance, while participant gender and education status were significant in the regression model, they were not incorporated into the elastic net model. However, the elastic net model, which performed well, did include self-reported indirect aggression, a nonsignificant IV in the control model. From the above comparisons, it is evident that an IV's significance in prediction or association tasks does not necessarily imply its inclusion will yield better performance and vice versa.

While there is theoretical support for these IVs' associations with aggressive behavior, the exclusion of other factors from the elastic net model with similar theoretical relationships posited between the other items and aggressive behavior merits discussion. A plausible explanation, from a machine learning perspective, hinges on the elastic net's bias–variance trade-off. It is conceivable that the measurements of the excluded factors exhibit high variance, leading to their omission from the model based on hyperparameters determined through cross-validation. This illustrates a key advantage of using cross-validation when the goal is to find maximum associations that are likely to generalize to future applications.

Limitations

It is also important to consider the limitations of the present study. We did not apply an experimental design in this study; therefore, we cannot conclude that the associations identified are causal.

In addition, the EMA subsample is a convenience sample from the main study. As Ribeaud et al. (2022) noted, the data set for the main study comprises young people growing up in an urban environment in one of the most affluent cities in the world (i.e., Zurich, Switzerland). In addition, this study also suffers from some common limitations of EMA studies. As suggested by Murray, Eisner, et al. (2022), participants may feel less motivated to respond when in a negative affective state, or they may be less likely to respond if their attention is captured in an interpersonal conflict. Finally, the data collection platform is an application for smartphones. Although most young people have smartphones, this can still cause some minor bias in representativeness.

Another limitation of this study is the use of a listwise data-cleaning procedure for missing data. Although this method is commonly used in examining associations of aggressive behavior (e.g., Bentley et al., 2021; Jensen et al., 2019; Rath et al., 2019), we are aware this method can cause bias as data may not be missing completely at random. However, a method like full information maximum likelihood (e.g., Murray et al., 2023) was not applied in this study because we were concerned about the issue of leakage. As we mentioned above, the training and testing data sets should be completely independent to avoid a boost in performance, and using the information from the full data set in missingness treatments can cause leakage. Currently, available implementations of machine learning with full information maximum likelihood and cross-validation do not provide a practical-to-implement solution to this issue.

Future Research Directions

Two future research directions can be derived from the present study. First, the design could be replicated in other samples that are more representative of the general population or of populations exhibiting higher levels of aggression where the prediction of aggression may be of considerable interest (e.g., forensic groups, high alcohol use groups). This design would also help determine the predictability of aggression across different contexts with experimental design. Predictability and specific key IVs could vary across contexts; for example, environmental constraints and triggers could vary according to populations' levels of aggression and developmental stage. We would also encourage researchers to apply the model fitting method used in this study to fit models when parsimony and high accuracy are required. The design and the code of this study can be used as a reference.

In addition, we believe a better model to identify associations with aggressive behavior can be fitted when further information is provided. For example, biological information such as heart rate can be passively collected by a smartwatch and could contribute to aggression prediction (Wilson & Scarpa, 2014). This approach will be explained in the next section.

Prevention Implications

While our results provide a first step toward utilizing trait and momentary data to prevent acts of aggression, there are a number of further steps and challenges to be overcome before this could be implemented in practice. For example, it will be necessary to assess the feasibility of utilizing the collection of momentary data in populations at high risk of aggression and to evaluate the extent to which a high degree of adherence can be achieved in these populations, as the biases caused by noncompliance from self-reported momentary features remain a recognized problem that may hurt the performance of the model (Markowski et al., 2021). This challenge could potentially be mitigated by using passive data collection for markers (e.g., physiological data, locations, voice tone) that may be associated with aggression (Ben-Zeev et al., 2013; Gustafson et al., 2014), but this will require future research. It will also be valuable to validate the findings with other measures of aggression (e.g., using body-worn cameras or sound recorders, or informant reports) to ensure that the findings identified via self-report are accurate. Ultimately, it will be necessary to select interventions that can be triggered when the risk of aggression is indicated to be high and to evaluate the effectiveness of utilizing these interventions in combination with momentary aggression risk data. Finally, there will be ethical issues to consider, such as those associated with triggering an intervention based on the potential for an aggressive act to occur, rather than on an aggressive act itself.

References

Altelbany, S. (2021). Evaluation of ridge, elastic net and lasso regression methods in precedence of multicollinearity problem: A simulation study. *Journal of Applied Economics and Business Studies*, *5*(1), 131–142. https://doi.org/10.34260/jaebs.517

Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of Clinical Epidemiology*, *68*(6), 627–636. https://doi.org/10.1016/j.jclinepi.2014.12.014

Ben-Zeev, D., Kaiser, S. M., Brenner, C. J., Begale, M., Duffecy, J., & Mohr, D. C. (2013). Development and usability testing of FOCUS: A smartphone system for self-management of schizophrenia. *Psychiatric Rehabilitation Journal*, *36*(4), 289–296. https://doi.org/10.1037/prj0000019

Bentley, K. H., Coppersmith, D. L., Kleiman, E. M., Nook, E. C., Mair, P., Millner, A. J., Reid-Russell, A., Wang, S. B., Fortgang, R. G., Stein, M. B., Beck, S., Huffman, J. C., & Nock, M. K. (2021). Do patterns and types of negative affect during hospitalization predict short-term post-discharge suicidal thoughts and behaviors? *Affective Science*, *2*(4), 484–494. https://doi.org/10.1007/s42761-021-00058-6

Borah, T. J., Murray, A. L., Eisner, M., & Jugl, I. (2021). Developing and validating an experience sampling measure of aggression: The aggression-es scale. *Journal of Interpersonal Violence*, *36*(11–12), NP6166–NP6182. https://doi.org/10.1177/0886260518812068

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, *11*, 2079–2107. https://doi.org/10.5555/1756006.1859921

Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., & Vakali, A. (2017). Mean birds: Detecting aggression and bullying on twitter. *Proceedings of the 2017 ACM on web science conference* (pp. 13–22) .

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, W.-C., Huang, C.-J., Hwang, J.-S., & Chen, C.-C. (2010). The relationship of health-related quality of life to workplace physical violence against nurses by psychiatric patients. *Quality of Life Research*, *19*(8), 1155–1161. https://doi.org/10.1007/s11136-010-9679-4

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 385–396. https://doi.org/10.2307/2136404

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In C. Zhang & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications* (pp. 157–175). Springer. https://doi.org/10.1007/978-1-4419-9326-7_5

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, *2*(4), 303–314. https://doi.org/10.1007/BF02551274

Derrick, J. L., & Testa, M. (2017). Temporal effects of perpetrating or receiving intimate partner aggression on alcohol consumption: A daily diary study of community couples. *Journal of Studies on Alcohol and Drugs*, *78*(2), 213–221. https://doi.org/10.15288/jsad.2017.78.213

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*(1), 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Eisner, N. L., Murray, A. L., Eisner, M., & Ribeaud, D. (2019). A practical guide to the analysis of non-response and attrition in longitudinal research using a real data example. *International Journal of Behavioral Development*, *43*(1), 24–34. https://doi.org/10.1177/0165025418797004

Ersan, C. (2020). Early language development and child aggression. *World Journal of Education*, *10*(1), 1–11. https://doi.org/10.5430/wje.v10n1p1

Finkel, E. J., & Hall, A. N. (2018). The I^3 Model: A metatheoretical framework for understanding aggression. *Current Opinion in Psychology*, *19*, 125–130. https://doi.org/10.1016/j.copsyc.2017.03.013

Finkel, E. J. (2014). The I3 model: Metatheory, theory, and evidence. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 49, pp. 1–104). Academic Press.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage Publications. https://www.john-fox.ca/Companion/

Fried, E. I., Papanikolaou, F., & Epskamp, S. (2022). Mental health and social contact during the covid-19 pandemic: An ecological momentary assessment study. *Clinical Psychological Science*, *10*(2), 340–354. https://doi.org/10.1177/21677026211017839

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

García, V., Mollineda, R. A., & Sánchez, J. S. (2010, August). Theoretical analysis of a performance measure for imbalanced data. *2010 20th International Conference on Pattern Recognition* (pp. 617–620). IEEE.

Gee, B. L., Han, J., Benassi, H., & Batterham, P. J. (2020). Suicidal thoughts, suicidal behaviours and self-harm in daily life: A systematic review of ecological momentary assessment studies. *Digital Health*, *6*, 1–38. https://doi.org/10.1177/2055207620963958

Gómez-Leal, R., Megías-Robles, A., Gutiérrez-Cobo, M. J., Cabello, R., & Fernández-Berrocal, P. (2022). Personal risk and protective factors involved in aggressive behavior. *Journal of Interpersonal Violence*, *37*(3–4), NP1489–NP1515. https://doi.org/10.1177/0886260520926322

Grendas, L. N., Chiapella, L., Rodante, D. E., & Daray, F. M. (2022). Comparison of traditional model-based statistical methods with machine learning for the prediction of suicide behaviour. *Journal of Psychiatric Research*, *145*, 85–91. https://doi.org/10.1016/j.jpsychires.2021.11.029

Grumm, M., Hein, S., & Fingerle, M. (2011). Predicting aggressive behavior in children with the help of measures of implicit and explicit aggression. *International Journal of Behavioral Development*, *35*(4), 352–357. https://doi.org/10.1177/0165025411405955

Gunthert, K. C., & Wenze, S. J. (2012). Daily diary methods. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 144–159). Guilford Press.

Gustafson, D. H., McTavish, F. M., Chih, M. Y., Atwood, A. K., Johnson, R. A., Boyle, M. G., Levy, M. S., Driscoll, H., Chisholm, S. M., Dillenburg, L., Isham, A., & Shah, D. (2014). A smartphone application to support recovery from alcoholism: A randomized clinical trial. *JAMA Psychiatry*, *71*(5), 566–572. https://doi.org/10.1001/jamapsychiatry.2013.4642

Gutiérrez-Esparza, G. O., Vallejo-Allende, M., & Hernández-Torruco, J. (2019). Classification of cyber-aggression cases applying machine learning. *Applied Sciences*, *9*(9), Article 1828. https://doi.org/10.3390/app9091828

Harrell, F. E., Jr. (2022). *Rms: Regression modeling strategies* [R package Version 6.3-0]. https://CRAN.R-project.org/package=rms

Hofmann, L. A., Lau, S., & Kirchebner, J. (2022). Advantages of machine learning in forensic psychiatric research—Uncovering the complexities of aggressive behavior in schizophrenia. *Applied Sciences*, *12*(2), Article 819. https://doi.org/10.3390/app12020819

Inoue, M., Tsukano, K., Muraoka, M., Kaneko, F., & Okamura, H. (2006). Psychological impact of verbal abuse and violence by patients on nurses working in psychiatric departments. *Psychiatry and Clinical Neurosciences*, *60*(1), 29–36. https://doi.org/10.1111/j.1440-1819.2006.01457.x

Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing imbalanced data–recommendations for the use of performance metrics. *2013 Humaine association conference on affective computing and intelligent interaction* (pp. 245–251).

Jensen, M., George, M., Russell, M., & Odgers, C. (2019). Young adolescents' digital technology use and mental health symptoms: Little evidence of longitudinal or daily linkages. *Clinical Psychological Science*, *7*(6), 1416–1433. https://doi.org/10.1177/2167702619859336

Jones, D. N., & Paulhus, D. L. (2014). Introducing the short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment*, *21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining*, *15*(4), 531–538. https://doi.org/10.1002/sam.11583

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559–575. https://doi.org/10.1017/S0140525X08005293

Kim, H., Lee, S., Lee, S., Hong, S., Kang, H., & Kim, N. (2019). Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: Observational study on older adults living alone. *JMIR MHealth and UHealth*, *7*(10), Article e14149. https://doi.org/10.2196/14149

Koul, A., Becchio, C., & Cavallo, A. (2018). Cross-validation approaches for replicability in psychology. *Frontiers in Psychology*, *9*, Article 1117. https://doi.org/10.3389/fpsyg.2018.01117

Kuhn, M. (2022). *Caret: Classification and regression training* [R package Version 6.0-92]. https://CRAN.R-project.org/package=caret

Lee, D. K. (2020). Data transformation: A focus on the interpretation. *Korean Journal of Anesthesiology*, *73*(6), 503–508. https://doi.org/10.4097/kja.20137

Lickiewicz, J., Piotrowicz, K., Hughes, P. P., & Makara-Studzińska, M. (2020). Weather and aggressive behavior among patients in psychiatric hospitals—An exploratory study. *International Journal of Environmental Research and Public Health*, *17*(23), Article 9121. https://doi.org/10.3390/ijerph17239121

Lischetzke, T., & Könen, T. (2020). Daily diary methodology. In F. Maggino (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 1–8). Springer.

Markowski, K. L., Smith, J. A., Gauthier, G. R., & Harcey, S. R. (2021). Patterns of missing data with ecological momentary assessment among people who use drugs: Feasibility study using pilot study data. *JMIR Formative Research*, *5*(9), Article e31421. https://doi.org/10.2196/31421

McConville, D. W., & Cornell, D. G. (2003). Aggressive attitudes predict aggressive behavior in middle school students. *Journal of Emotional and Behavioral Disorders*, *11*(3), 179–187. https://doi.org/10.1177/10634266030110030501

Mikus, A., Hoogendoorn, M., Rocha, A., Gama, J., Ruwaard, J., & Riper, H. (2018). Predicting short term mood developments among depressed patients using adherence and ecological momentary assessment data. *Internet Interventions*, *12*, 105–110. https://doi.org/10.1016/j.invent.2017.10.001

Murray, A. L., Brown, R., Zhu, X., Speyer, L. G., Yang, Y., Xiao, Z., Ribeaud, D., & Eisner, M. (2023). Prompt-level predictors of compliance in an ecological momentary assessment study of young adults' mental health. *Journal of Affective Disorders*, *322*, 125–131. https://doi.org/10.1016/j.jad.2022.11.014

Murray, A. L., Eisner, M., Ribeaud, D., & Booth, T. (2022). Validation of a brief measure of aggression for ecological momentary assessment research: The aggression-es-a. *Assessment*, *29*(2), 296–308. https://doi.org/10.1177/1073191120976851

Murray, A. L., Speyer, L. G., Brown, R., Zhu, X., Yang, Y., Eisner, M., & Ribeaud, D. (2022). *Advancing multi-timeframe developmental research through combining long-term cohort and ecological momentary assessment studies: The decades-to-minutes (d2m) study*. PsyArxiv.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *18*(6), 275–285. https://doi.org/10.1002/cem.873

Nivette, A., Eisner, M., & Ribeaud, D. (2017). Developmental predictors of violent extremist attitudes: A test of general strain theory. *Journal of Research in Crime and Delinquency*, *54*(6), 755–790. https://doi.org/10.1177/0022427817699035

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*(12), 1373–1379. https://doi.org/10.1016/S0895-4356(96)00236-3

Poldrack, R. A., Monahan, J., Imrey, P. B., Reyna, V., Raichle, M. E., Faigman, D., & Buckholtz, J. W. (2018). Predicting violent behavior: What can neuroscience add? *Trends in Cognitive Sciences*, *22*(2), 111–123. https://doi.org/10.1016/j.tics.2017.11.003

Qiu, X., Brooks, A. I., Klebanov, L., & Yakovlev, N. (2005). The effects of normalization on the correlation structure of microarray data. *BMC Bioinformatics*, *6*, Article 120. https://doi.org/10.1186/1471-2105-6-120

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rath, D., de Beurs, D., Hallensleben, N., Spangenberg, L., Glaesmer, H., & Forkmann, T. (2019). Modelling suicide ideation from beep to beep: Application of network analysis to ecological momentary assessment data. *Internet Interventions*, *18*, Article 100292. https://doi.org/10.1016/j.invent.2019.100292

Ribeaud, D., Murray, A., Shanahan, L., Shanahan, M. J., & Eisner, M. (2022). Cohort profile: The zurich project on the social development from childhood to adulthood (z-proso). *Journal of Developmental and Life-Course Criminology*, *8*(1), 151–171. https://doi.org/10.1007/s40865-022-00195-x

Richter, D., & Berger, K. (2006). Post-traumatic stress disorder following patient assaults among staff members of mental health hospitals: A prospective longitudinal study. *BMC Psychiatry*, *6*(1), Article 15. https://doi.org/10.1186/1471-244X-6-15

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, Article 77. https://doi.org/10.1186/1471-2105-12-77

RStudio Team. (2020). *Rstudio: Integrated development environment for r*. https://www.rstudio.com/

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, *10*(3), Article e0118432. https://doi.org/10.1371/journal.pone.0118432

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Shorey, R. C., Stuart, G. L., McNulty, J. K., & Moore, T. M. (2014). Acute alcohol use temporally increases the odds of male perpetrated dating violence: A 90-day diary analysis. *Addictive Behaviors*, *39*(1), 365–368. https://doi.org/10.1016/j.addbeh.2013.10.025

Shorey, R. C., Stuart, G. L., Moore, T. M., & McNulty, J. K. (2014). The temporal relationship between alcohol, marijuana, angry affect, and dating violence perpetration: A daily diary study with female college students. *Psychology of Addictive Behaviors*, *28*(2), 516–523. https://doi.org/10.1037/a0034648

Testa, M., Wang, W., Derrick, J. L., Crane, C., Leonard, K. E., Collins, R. L., Hanny, C., & Muraven, M. (2020). Does state self-control depletion predict relationship functioning and partner aggression? An ecological momentary assessment study of community couples. *Aggressive Behavior*, *46*(6), 547–558. https://doi.org/10.1002/ab.21915

Thiele, C., Laireiter, A.-R., & Baumann, U. (2002). Diaries in clinical psychology and psychotherapy: A selective review. *Clinical Psychology & Psychotherapy*, *9*(1), 1–37. https://doi.org/10.1002/cpp.302

Torgo, L. (2016). *Data mining with R: Learning with case studies* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315399102

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivée, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology*, *19*(3), 285–300. https://doi.org/10.1007/BF00911232

van Smeden, M., de Groot, J. A., Moons, K. G., Collins, G. S., Altman, D. G., Eijkemans, M. J., & Reitsma, J. B. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology*, *16*(1), Article 163. https://doi.org/10.1186/s12874-016-0267-3

Watson, D., & Clark, L. A. (1994). *The PANAS-x: Manual for the positive and negative affect schedule-expanded form* [unpublished manuscript]. University of Iowa.

Wilson, L. C., & Scarpa, A. (2014). Aggressive behavior: An alternative model of resting heart rate and sensation seeking. *Aggressive Behavior*, *40*(1), 91–98. https://doi.org/10.1002/ab.21504

Wrzus, C., & Neubauer, A. B. (2023). Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment*, *30*(3), 825–846. https://doi.org/10.1177/10731911211067538

Yan, Y. (2016). *Mlmetrics: Machine learning evaluation metrics* [R package Version 1.1.1]. https://CRAN.R-project.org/package=MLmetrics

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yeater, E. A., Witkiewitz, K., Testa, M., & Bryan, A. D. (2022). Substance use, risky sex, and peer interactions predict sexual assault among college women: An ecological momentary assessment (ema) study. *Journal of Interpersonal Violence*, *37*(7–8), NP5094–NP5115. https://doi.org/10.1177/0886260520958720

Yoo, J. E. (2018). Timss 2011 student and teacher predictors for mathematics achievement explored and identified via elastic net. *Frontiers in Psychology*, *9*, Article 317. https://doi.org/10.3389/fpsyg.2018.00317

Zhang, P., Fonnesbeck, C., Schmidt, D. C., White, J., Kleinberg, S., & Mulvaney, S. A. (2022). Using momentary assessment and machine learning to identify barriers to self-management in type 1 diabetes: Observational study. *JMIR MHealth and UHealth*, *10*(3), Article e21959. https://doi.org/10.2196/21959

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Received June 28, 2023
Revision received July 16, 2024
Accepted July 19, 2024 ■