Person-specific and pooled prediction models for binge eating, alcohol use and binge drinking in bulimia nervosa and alcohol use disorder

N. Leenaerts[1,2] iD , P. Soyster[3], J. Ceccarini[4], S. Sunaert[5], A. Fisher[3]
and E. Vrieze[1,2]

[1]Department of Neurosciences, KU Leuven, Leuven Brain Institute, Research Group Psychiatry, Leuven, Belgium; [2]Department of Neurosciences, Mind-Body Research, Research Group Psychiatry, KU Leuven, Belgium; [3]Department of Psychology, Idiographic Dynamics Lab, University of California, Berkeley, USA; [4]Department of Nuclear Medicine and Molecular Imaging, KU Leuven, Leuven Brain Institute, Research Nuclear Medicine & Molecular Imaging, Leuven, Belgium and [5]Department of Imaging and Pathology, Translational MRI, Biomedical Sciences Group, KU Leuven, Belgium

Abstract

Background. Machine learning could predict binge behavior and help develop treatments for bulimia nervosa (BN) and alcohol use disorder (AUD). Therefore, this study evaluates person-specific and pooled prediction models for binge eating (BE), alcohol use, and binge drinking (BD) in daily life, and identifies the most important predictors.
Methods. A total of 120 patients (BN: 50; AUD: 51; BN/AUD: 19) participated in an experience sampling study, where over a period of 12 months they reported on their eating and drinking behaviors as well as on several other emotional, behavioral, and contextual factors in daily life. The study had a burst-measurement design, where assessments occurred eight times a day on Thursdays, Fridays, and Saturdays in seven bursts of three weeks. Afterwards, person-specific and pooled models were fit with elastic net regularized regression and evaluated with cross-validation. From these models, the variables with the 10% highest estimates were identified.
Results. The person-specific models had a median AUC of 0.61, 0.80, and 0.85 for BE, alcohol use, and BD respectively, while the pooled models had a median AUC of 0.70, 0.90, and 0.93. The most important predictors across the behaviors were craving and time of day. However, predictors concerning social context and affect differed among BE, alcohol use, and BD.
Conclusions. Pooled models outperformed person-specific models and the models for alcohol use and BD outperformed those for BE. Future studies should explore how the performance of these models can be improved and how they can be used to deliver interventions in daily life.

Introduction

Bulimia nervosa (BN) and alcohol use disorder (AUD) are two psychiatric disorders that share a number of similarities (American Psychiatric Association, 2013). First, they can both be characterized by binge behavior where patients consume a large amount of food (i.e. binge eating [BE]) or alcohol (i.e. binge drinking [BD]) within a short period of time, while not being able to stop eating or drinking or not being able to control the amount they eat or drink. (American Psychiatric Association, 2013). Second, they can both have a significant impact on health with BN having a high mortality of 1.7 per 1000 person-years and with AUD being the largest risk factor for disease and disability among 15- to 49-year-olds (Arcelus, Mitchell, Wales, & Nielsen, 2011; Griswold et al., 2018). Third, both disorders can be difficult to treat, with up to 60% of patients who receive treatment not achieving remission (Fleury et al., 2016; Linardon & Wade, 2018). Taken together, the high impact and poor treatment outcomes highlight the need for more effective therapies for BN and AUD.

One promising new form of therapy is the just-in-time adaptive intervention (JITAI) (Nahum-Shani et al., 2018). In a JITAI, support is given 'just-in-time' or when a patient needs it the most. For example, a patient could report their emotions, behaviors, and context with a smartphone application, and an algorithm could evaluate the risk of BE or BD based on this information, after which an intervention could be sent out when this momentary risk is elevated. The support can also be adaptive, meaning that it can be tailored to a patient's in-the-moment needs. For instance, a patient could receive a text message alert when the estimated risk for BE or BD is moderate, but a phone call when the estimated risk is high. Because of its potential benefits, several researchers have developed and implemented JITAIs in recent years, but their results have been mixed (Carpenter, Menictas, Nahum-Shani, Wetter, &

Murphy, 2020; Hardeman, Houghton, Lane, Jones, & Naughton, 2019; Wang & Miller, 2020). One reason for this could be the limited adaptive nature of these JITAIs. Namely, the research designs of these JITAIs were primarily based on previous literature (e.g. which information should be gathered from participants and how it should be evaluated), which means that the decision to send out an intervention was static and based on findings from previous studies. However, if such decisions were to be based on the actual data provided by participants, a JITAI would be more adaptive and perhaps more effective.

This goal could be realized with the help of machine learning (ML) where statistical models and algorithms learn from data without explicit instruction (Shatte, Hutchinson, & Teague, 2019). A ML model could learn when individuals are at risk of BE or BD in daily life and then subsequently predict this risk when presented with new data in a JITAI. A ML model could also determine which of the many possibly assessed variables (e.g. momentary mood, location, social context, and time) are predictive of BE or BD and which ones are not. Several ML algorithms can examine a large number of variables and select only those that are most predictive of an outcome (Cai, Luo, Wang, & Yang, 2018). This kind of information could then provide targets for the interventions in a JITAI. However, researchers are confronted with specific challenges when using ML to predict daily life behavior. Namely, they need to decide whether they want to build person-specific or group-level (pooled) prediction models. On the one hand, person-specific models are trained with the data of an individual patient (Soyster, Ashlock, & Fisher, 2021). This type of model can be built more easily and can result in more person-specific information. On the other hand, pooled prediction models are trained with the data of multiple patients (Soyster et al., 2021). This model type is more difficult to build as more patients need to be included, but can result in a better model performance, particularly if the factors driving a momentary behavior are similar across the study participants.

In recent years, several studies have built person-specific and pooled models to predict BE, alcohol use, or BD in daily life (Arend et al., 2023; Bae, Chung, Ferreira, Dey, & Suffoletto, 2018; Bae et al., 2017; Levinson, Trombley, Brosof, Williams, & Hunt, 2022; Soyster et al., 2021; Walters et al., 2021). When it comes to disordered eating behavior, one study demonstrates that pooled models can predict future occurrences of BE, restriction, and purging in patients with an eating disorder (Levinson et al., 2022). By utilizing predictors related to disordered eating cognitions and behaviors, along with affect, the study shows that dietary restriction, weighing, and anxiousness are important predictors of subsequent BE. Additionally, another study indicates that person-specific models can perform well in predicting BE in patients with a binge-type eating disorder (Arend et al., 2023). This study, using a set of variables selected from feedback from clinicians and patients, reports that hunger and craving are the most common predictors of BE. When it comes to alcohol use, studies have demonstrated that pooled models can successfully use phone sensor data to predict non-heavy alcohol use as well as BD in individuals without AUD (Bae et al., 2017, 2018). In these studies, time of day, number of activities, and phone usage emerge as the most informative predictors. Furthermore, another study finds that both person-specific and pooled models can predict alcohol use in individuals without AUD utilizing predictors related to affect, craving, and recent alcohol use (Soyster et al., 2021). More specifically, it finds that craving and feeling pressured to drink are the most common predictors across individuals. Employing a similar set of variables, with

additional predictors related to social setting and location, a different study also concludes that pooled models can predict alcohol use in individuals without AUD (Walters et al., 2021). Here, craving, the availability of alcohol, and feeling that alcohol will improve mood were the most important predictors.

However, these studies have significant limitations. First, their generalizability to a broader clinical population is often limited. This is because only a few studies include a clinical sample and those that do, include a small number of participants for which they only have a limited number of observations. This is problematic as a small sample size can have serious methodological implications in ML (Way, Sahiner, Hadjiiski, & Chan, 2010). Indeed, most studies did not hold out data when training or tuning their ML models and therefore were not able to evaluate model performance on unseen data. This implies that their models run a serious risk of overfitting and might not generalize. Second, the majority of variables in these studies assess emotions or behaviors and do not look at the social or situational context of a patient. However, previous studies show the importance of context in BE, alcohol use, and BD (Allison & Timmerman, 2007; Clapp, Shillington, & Segars, 2009). Third, to our knowledge only one study evaluated both person-specific and pooled prediction model and did so only for alcohol use, leaving the question unanswered which model type performs best for BE and BD (Soyster et al., 2021).

Because of these limitations, it is still unclear to what extent BE, alcohol use, and BD can be predicted in the daily lives of patients with BN and/or AUD and which variables are important predictors. This study aims to fill that gap by collecting a large amount of data from a clinical sample on a variety of variables and make methodologically correct prediction models. We followed patients with BN and/or AUD over a period of 12 months during which we used the experience sampling method (ESM) to repeatedly assess the patient's emotions, behaviors, and contexts in daily life. We then used this data to fulfill the following two objectives. First, to build and evaluate person-specific and pooled prediction models for BE, alcohol use, and BD in daily life. Second, to identify the most important predictors of these behaviors.

Methods

Study sample

The participants were drawn from a larger ESM study that followed patients with BN and/or AUD as well as control volunteers without these diagnoses in daily life. In the current study, all the data of the patients with BN ($n = 50$), with AUD ($n = 51$) or with BN and AUD ($n = 19$) were used, after the elimination of one patient with BN and two patients with AUD due to insufficient data (i.e. not answering two consecutive assessments, causing the lagging procedure described below to fail). Recruitment happened in Flanders, Belgium through residential and ambulatory care centers, patient groups, universities, social media, and by handing out flyers on the street. Inclusion ran from September 2019 to February 2022. The inclusion criteria were: (1) being assigned female at birth; (2) understanding Dutch language; (3) being of age ⩾18 years; and (4) being of BMI ⩾18.5 kg/m2. It was decided to not include individuals assigned male at birth as the prevalence of BN in significantly lower in this population (Galmiche, Déchelotte, Lambert, & Tavolacci, 2019). Additional inclusion criteria for patients were: (5) meeting the criteria for BN or AUD of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013); (6) meeting those diagnostic criteria for a duration of ⩽5 years. This

maximum was set as the importance of certain predictors of BE, alcohol use, and BD are thought to change over the course of AUD and BN (Boness, Watts, Moeller, & Sher, 2021; Pearson, Wonderlich, & Smith, 2015). For example, it is thought that BE episodes start as a rash action during moments of high negative affect, and are reinforced by a subsequent decrease in negative affect (Pearson et al., 2015). However, it is then thought that after repeated cycles of emotional distress, urge, and BE, the behavior becomes habitual, where it persists, even when it is no longer reinforced by a decrease in negative affect (Pearson et al., 2015). Furthermore, it is thought that the role of negative affect and positive affect change of the course of AUD, whereby changes in positive affect are more predictive of alcohol use in the beginning of AUD, while the role of negative affect increases over the course of AUD (Koob & Le Moal, 1997). Participants with AUD also needed to display a pattern of repetitive BD according to the criteria of the National Institute on Alcohol Abuse and Alcoholism (i.e. drinking 4 units of alcohol within 2 h for women) (National Institute on Alcohol Abuse & Alcoholism [NIAAA], 2022). Participants were excluded for the following reasons: (1) major medical pathology (e.g. severe liver or kidney disease, uncontrolled diabetes, cancer or untreated hyper- or hypothyroidism); (2) chronic use of sedatives (i.e. more than three times in the past three month); (3) pregnancy; (4) presence of major psychiatric pathology (i.e. schizophrenia, autism spectrum disorder, bipolar disorder, substance use disorder other than AUD). All participants gave their written consent, and the study was approved by the ethical committee of the UZ/KU Leuven.

Study design

Potential participants were initially screened via telephone or email, after which they attended an in-person assessment. Here, a psychiatry resident confirmed an individual's eligibility to participate. The participants had their weight and height measured with a calibrated scale and stadiometer and completed clinical interviews and questionnaires. All participants underwent a briefing on the ESM questions and practiced the use of the mobile application. Then, the participants entered the ESM protocol on the first Thursday after the in-person assessment. An overview of the protocol can be seen in Fig. 1. It consisted of a repeated measurement design where seven bursts of data collection were spread out over a 12-month period. The bursts had a duration of three weeks and were separated by intervals of five weeks. During the bursts, data were only collected on Thursday, Friday, and Saturday to limit the protocol's impact on the participants. These specific days were selected to consecutively gather data on both week and weekend days. Then, participants were required to respond to eight signals on each day of data collection which were sent out on a signal-contingent (i.e. semi-random) basis. The participants received 20 eurocent per answered assessment. The ESM data were initially collected with the app MobileQ (Meers, Dejonckheere, Kalokerinos, Rummens, & Kuppens, 2020). When the development of the app was discontinued in October 2020, data collection continued using m-Path (Mestdagh et al., 2022). More information about the apps can be found in online Supplementary eMethods 1 and eTable 1 in the supplement.

Measures

Baseline measures

The Structured Clinical Interview for DSM-5 (SCID-5-S) was used to confirm the diagnosis of BN or AUD and to screen for other
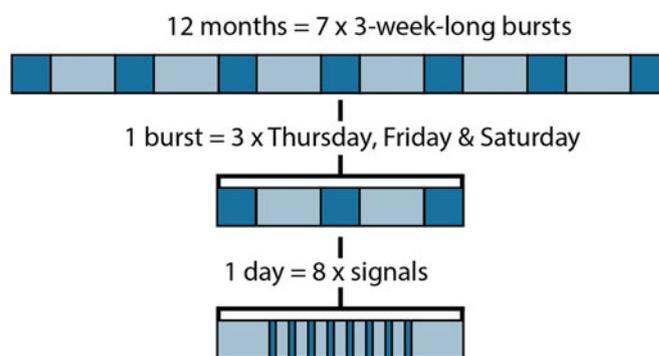
Figure 1. Experience sampling method protocol. The protocol consisted of seven bursts of data collection which were spread out over a 12-month period. The bursts had a duration of three weeks and were separated by intervals of five weeks. During the bursts, data were only collected on Thursday, Friday, and Saturday. On a given day of data collection, participants received eight signals which were sent on a signal-contingent (i.e. semi-random) basis.

psychiatric disorders (American Psychiatric Association [APA], 2017). BN and AUD severity were assessed using the Eating Disorder Examination Questionnaire (EDE-Q) and the Alcohol Use Disorders Identification Test (AUDIT) (Fairburn & Beglin, 1994; Saunders, Aasland, Babor, De La Fuente, & Grant, 1993).

ESM measures

At each assessment, the participants received questions evaluating different emotions, behaviors, and contexts. The exact number of items varied at each assessment as the presentation of some questions was conditional on a participant's response to a previous question. The full list of questions can be seen in Table 1. More information on the reliability and/or validity of the items can be found in the supplement (online Supplementary eMethods 2). Importantly, participants needed to indicate if they had eaten since the previous assessment. If so, they had to identify the eating event as undereating, normal eating, or overeating. Then, participants were asked if they experienced a loss of control over their eating behavior. As in previous studies, **BE** was defined as an episode of overeating with loss of control (Ambwani, Roche, Minnick & Pincus, 2015). The participants were trained to interpret overeating as eating an amount of food that is definitely larger than what most people would eat under similar circumstances. Furthermore, they were instructed to interpret loss of control as wanting to stop eating, but not being able to. Similarly, participants needed to indicate whether they drank alcohol since the previous assessment and if so, how many units they drank and if they experienced a loss of control over their drinking behavior. The participants were instructed on the definition of an alcohol unit. Here, **BD** was defined as having consumed at least four units of alcohol since the previous assessment while **alcohol use** was conceptualized as having consumed at least one unit since the previous assessment.

Data analysis

Data preparation

A figure providing an overview on the data preparation procedure can be found in the supplement (online Supplementary eFigure 1). Only assessments answered within 240 min of the prompt were used in the analyses. This window was chosen to include assessments which were answered later at night, where the likelihood that patients binge eat, drink alcohol, or binge

Table 1. Experience sampling method questions

Group	Question	Response
Negative affect	Right now, I feel insecure	Totally disagree 1 2 3 4 5 6 7 Totally agree
	Right now, I feel lonely	
	Right now, I feel afraid	
	Right now, I feel down	
	Right now, I feel guilty	
	Right now, I feel stressed	
Positive affect	Right now, I feel satisfied	
	Right now, I feel cheerful	
	Right now, I feel relaxed	
Rash action	Right now, I am doing something that could get me into trouble	
	Right now, I am doing something that I might regret later on	
	Since the last beep I did something risky	
	Since the last beep I acted without thinking	
	Since the last beep I said something without thinking	
Lack of perseverance	Right now, there is something I should be doing, but am not doing	
Motivation	Right now, I feel interested to do things	
Craving	Right now, I have … desire to have a binge eating episode	None 1 2 3 4 5 Overwhelming
	Right now, I have … desire to drink alcohol	
Activities	What am I doing? • Nothing • Resting • Eating/drinking • Passive leisure activity • Active leisure activity • Travel • Household chores/groceries • Work/studies • Self-care • Interacting/talking • Mobile phone/social media • Something else	Not present 0 1 Present
Activity-related stress	I can handle the current situation	Totally disagree 1 2 3 4 5 6 7 Totally agree
	The current situation takes effort	
	I would like to do something else	
Location	Where am I? • Home • With family or friends • At work or school • Bar/café/restaurant/party • In transit • Somewhere in a building • Somewhere outside	Not present 0 1 Present
Social context	Who am I with? • Nobody • Friends • Housemates • Family • Partner • Strangers • Coworkers • Acquaintances • People online	Not present 0 1 Present

(Continued)

Table 1. (*Continued.*)

Group	Question	Response
Social stress	I like this company	Totally disagree 1 2 3 4 5 6 7 Totally Agree
	I would rather be alone	
	It feels OK to be alone	
	I would like to be with others	
Most important event	The most important event since the last beep was …	Very unpleasant 1 2 3 4 5 6 7 Very pleasant
Positive events	Indicate which pleasant events happened since the last beep • I laughed or had a fun talk with someone • I felt part of a group • I enjoyed someone else's joy • I did something that made someone else happy • I performed well • I got positive feedback • I experienced bodily pleasure • I had luck • I did a fun activity • I thought about something pleasant that was going to happen • There was another pleasant event • There were no pleasant events	Not present 0 1 Present
Pleasantness positive events	How pleasant were these positive events all together?	Not at all 1 2 3 4 5 Extremely
Negative events	Indicate which negative events happened since the last beep • I had a dispute with someone • I felt excluded • I was burdened by someone else's suffering • I was under pressure • I did something wrong • I got negative feedback • I experienced physical discomfort • I had a setback • I did a boring or annoying activity • I thought about something annoying that was going to happen • There was another negative event • There were no negative events	Not present 0 1 Present
Stressfulness negative events	How stressful were these negative events all together?	Not at all 1 2 3 4 5 Extremely
Effort	Since the last beep, I made a mental effort	Totally disagree 1 2 3 4 5 6 7 Totally agree
	Since the last beep, I made a physical effort	
Substances	Since the last beep, I have used • Caffeine • Nicotine • Medication • Cannabis • Other drugs • None of the above	Not present 0 1 Present
Eating	Since the last beep, I have eaten	No 0 1 Yes
	My eating was …	Undereating (0) Normal eating (1) Overeating (2)
	I felt that I lost control over eating	No 0 1 Yes

(*Continued*)

Table 1. (*Continued.*)

Group	Question	Response
Alcohol use	Since the last beep, I have drunk alcohol	No 0 1 Yes
	I drank … units of alcohol	1 2 3 4 5 6 > 7 Units
	I felt that I lost control over drinking	No 0 1 Yes

drink could be higher. First, the scoring of the conditional ESM variables was corrected. A conditional ESM variable depended on a previous ESM answer (e.g. how stressful an event was, was only asked on the condition that a participant answered 'yes' on experiencing a stressful event). The conditional ESM variables therefore included missing values, when the condition was not met, which could be filled in with zeroes (i.e. indicating that past events were not stressful at all). Second, temporal variables were created that represented assessment number (i.e. 1–8), weekday (i.e. Thursday, Friday, and Saturday), time since starting participation in the study (linear, quadratic, and cubic) and cycles of 12 h, and 24 h frequency (Flury & Levri, 1999). These temporal variables have been used in previous studies predicting BE and alcohol use (Arend et al., 2023; Soyster et al., 2021). Third, to account for the varying levels of COVID-19 prevention measures throughout the study, a COVID-19 stringency variable was created based on the Oxford COVID-19 Government Response Tracker (Kira et al., 2022). More information on the temporal and COVID-19 stringency variables can be found in the supplement (online Supplementary eMethods 3). This brought the total possible number of predictors to 110. However, a predictor could not be entered in the prediction models when it had a variance of zero (i.e. meaning it always had the same response value). More specifically, for the person-specific models, the median number of predictors for BE was 97 (Q1–Q3: 90–100), while the median number of predictors for alcohol use was 96 (Q1–Q3: 90–99), and the median number of predictors for BD was 98 (Q1–Q3: 91–99). The pooled models used all predictors. Fourth, all ESM variables except for the outcome variables were lagged by one assessment, with time between assessments measuring 102 min on average. The variables could be lagged across days, but not across weeks. The temporal variables and the COVID-19 stringency variable were not lagged and therefore remained aligned in time with the outcome variables. Fifth, observations with missing values were removed from the data.

This resulted in a dataset which could be used to predict BE, BD, and alcohol use at a certain point in time in the future, based on the temporal variables and the COVID-19 stringency variable at that timepoint as well as the ESM variables at a previous timepoint. This was done to emulate how a machine learning-based JITAI would be used to treat a patient in daily life. For example, a model could predict whether a patient who reported to experience more stress is more likely to binge eat after two hours. As lagging across days was permitted, BE, BD and alcohol use episodes which happened at night but were reported in the morning could also be predicted.

Model training and evaluation

Person-specific as well as pooled prediction models were built for BE, alcohol use, and BD. Based on the definitions outlined under ESM measures, the moments of BD were also considered moments of alcohol use. This approach was taken because a JITAI would most likely focus on either alcohol use (i.e. drinking any alcohol) or BD (i.e. drinking more than 4 units in 2 h for women), rather than non-heavy alcohol use (i.e. alcohol use that is not BD). For BE, the data of the patients with BN and the patients with BN and AUD were used ($n = 69$). Similarly, for alcohol use and BD, the data of the patients with AUD and the patients with BN and AUD were used ($n = 70$). This meant that only the data of patients who displayed the behavior were included in the analyses for a specific outcome. The models were trained and evaluated with the *ensr*, *glmnet*, *pROC* and *caret* packages in R, version 4.1.1 (DeWitt, 2019; Friedman, Hastie, & Tibshirani, 2010; Kuhn, 2021; Robin et al., 2011). The scripts and data for the analyses can be found at https://rdr.kuleuven.be/dataset.xhtml?persistentId=doi:10.48804/OBLDWE. A brief description of the elastic net wrappers which were developed for this paper can be found in the supplement (online Supplementary eMethods 4). More detailed information can be found at https://github.com/nicolasleenaerts/NLML. *Person-specific* The models were fitted and evaluated on the data of each participant with nested k-fold cross-validation. A visual representation of this method can be seen in Fig. 2. More information on nested cross-validation can be found in the supplement (online Supplementary eMethods 5). For the outer loop, a stratified 5-fold cross-validation was used. Due to the stratification, the distribution of positive events was similar across folds. The allocation of observations to specific folds was random, meaning that the observations within each fold were not temporally contiguous. However, due to the lagging procedure described above, each instance of the dependent variables was only ever predicted by the independent variables for the immediately previous observation. The continuous variables of the training folds were standardized. This can increase performance of regression-based models and simplify comparisons between model estimates (Shahriyari, 2019). Additionally, the continuous variables from the test fold were also standardized, but with the mean and standard deviation from the training folds. This separated procedure transforms the testing data to the same scale as the training data but prevents any information from leaking.

For the inner loop, an elastic net regularized regression model was fitted to the training folds of the outer loop (Zou & Hastie, 2005). Though other machine learning techniques exist, elastic net was specifically chosen as the machine learning algorithm as it is especially suited for the objectives of the current manuscript. This is because it is a performant machine learning technique that reduces overfitting, can work with high-dimensional data (i.e. more predictors than observations), handles correlated variables, but also provides information on the strength and
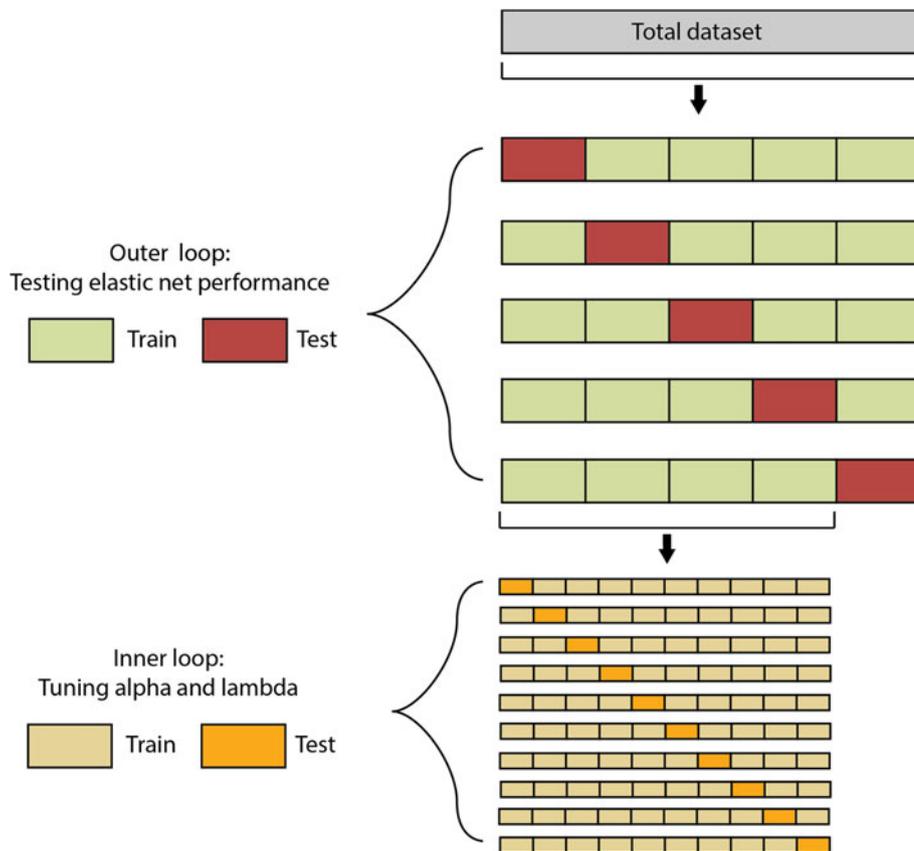
Figure 2. Nested cross-validation. In the outer loop, the total dataset was divided into five folds which were processed in five rounds. During each round, one fold was used as a test set while the other four folds were used a training set. In the inner loop, the training folds were used to select the most optimal alpha and lambda and to fit the elastic net model. A grid search of 10 alphas and 100 lambdas was performed with a 10-fold cross-validation. The combination with the lowest cross-validation error was used to fit the definitive elastic net model. This model was then evaluated on the test fold of the outer loop.

nature of the relation between a predictor and an outcome. It combines two regularization methods, ridge regression which shrinks model estimates and LASSO regression which removes variables that do not contribute to the model. The amount of ridge and lasso regression is expressed by a variable alpha which varies from 0 (exclusively ridge regression) to 1 (exclusively LASSO). The strength of the regularization is defined by a variable lambda with higher values leading to more shrinkage of the coefficients. The most optimal alpha and lambda were selected with a grid search of 10 alphas and 100 lambdas (i.e. the default settings of *ensr*). For each possible combination, a cross-validation error was calculated with 10-fold cross-validation. The combination with the lowest cross-validation error was then used to fit the definitive elastic net model on the training folds of the outer loop. This elastic net model was then used to predict BE, BD or alcohol use in the data of the test fold of the outer loop. The predictions were then compared with the actual BE, BD, and alcohol use events in the test fold to calculate the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, accuracy, positive predictive value (PPV) and negative predictive value (NPV). Due to the nested cross-validation, a participant needed to have a sufficient number of BE, BD or alcohol use events ($n > 4$) to be included in the analysis. *Pooled* The pooled models were also fitted and evaluated with nested k-fold cross-validation. However, they were trained on the pooled training data of all the participants and tested on the individual test data of each participant. More specifically, in the outer loop, the training folds were a combination of the standardized training folds of the participants. Due to the standardization at a participant-level, the values of the continuous variables represented a deviation from the within-person means. As no multilevel variant of elastic

net regularized regression exists, this accounted in part for the within-person nesting of the data (Soyster et al., 2021). In the inner loop, the most optimal alpha and lambda were again determined with 10-fold cross-validation and used to fit the final elastic net model. This model was then applied to the test fold of every individual participant to evaluate the AUC, sensitivity, specificity, accuracy, PPV, and NPV of the pooled model for each participant. As the training folds were pooled, a participant only needed to have one BE, BD, or alcohol use event in the test fold to be included in the analyses.

Model comparison

For each participant and each outcome, the AUC of the pooled models was subtracted from the AUC of the person-specific models. Then, to explore why some participants had a better performance with the person-specific model than with the pooled model, this difference in AUC was compared between the different analysis groups with Mann–Whitney U tests and correlated to age, BMI, EDE-Q scores, AUDIT scores, BE frequency, and BD frequency with Spearman correlations. Non-parametric tests were performed due to the non-normal distribution of the AUCs.

Validation analyses

First, to assess whether patients were more likely to binge eat, drink alcohol, or binge drink on certain days of the week (e.g. Thursday, Friday, and Saturday), generalized linear mixed models were constructed for each outcome of interest (i.e. binge eating, alcohol use, and binge drinking) with the aggregated data of all participants, with day of the week as a main effect and with a random intercept for the participants. Second, as there was an imbalance in the outcomes whereby patients typically did not display

BE, alcohol use or BD, several analyses were performed to assess the validity of the results. To begin, the results of the pooled and person-specific models were compared to those of models that always predict the majority class (i.e. not BE, not drinking alcohol, not BD). Additionally, the results of the person-specific and pooled models were compared to those of models where the imbalance in the outcomes was corrected with ROSE or SMOTE. Third, as some participants changed apps over the course of the study, the impact of app type on model performance was explored. More specifically, person-specific and pooled models were constructed with an additional variable indicating whether MobileQ of m-Path was used, after which their performance was compared to that of the original models. Additionally, the AUC of the original person-specific and pooled models was correlated to the percentage of observations that a participant reported through m-Path. Spearman correlations was performed due to the non-normal distribution of the AUCs. More information on these validation analyses can be found in the supplement (online Supplementary eResults 3–5).

Model predictors

For each outcome and each model type, the 10% best predictors were identified. This was based on the raw estimates for the pooled model (as only one estimate per variable existed) and the mean estimates over all participants for the person-specific models.

Results

Sample characteristics

The characteristics of the different patient groups can be found in Table 2. Additionally, the characteristics of the different analysis groups (i.e. for BE or BD/alcohol use) can be found in the supplement (online Supplementary eTable 2). Notably, the age of the patients with BN/AUD (mean = 20.4, S.D. = 1.7, CI 19.6–21.2) was lower than that of the patients with BN (mean = 22.4, S.D. = 4.1, CI 21.3–23.6). Also, the BMI of the patients with BN (mean = 25.6, S.D. = 5.9, CI 23.9–273) was higher than that of the patients with AUD (mean = 21.5, S.D. = 3.5, CI 20.5–22.4).

Data characteristics

In total, 41 (34.2%) participants (16 (32.0%) BN, 19 (37.3%) AUD, 6 (31.6%) BN/AUD) dropped out of the study before the end of the ESM protocol. For every participant group (AUD, BN, AUD/BN), there was no significant difference between patients who dropped out and those who did not when it came to age, BMI, illness duration, AUDIT scores, or EDE-Q scores. The mean compliance (percentage of signals answered) per participant during the first burst was 80.4% for the patients with BN, 75.2% for the patients with AUD and 73.6% for the patients with BN/AUD. This is similar to the compliance rates of previous cross-sectional ESM studies in patients with an eating disorder or AUD (Fischer, Wonderlich, Breithaupt, Byrne, & Engel, 2018; Jones et al., 2019; Schaefer et al., 2020). In total, the patients with BN answered 12 932 (61.5%) of their scheduled beeps, while the patients with AUD answered 12 328 (62.9%) and the patients with BN/AUD answered 3947 (51.2%). The overall compliance of this study fell in the range of the lengthier ESM studies on substance use (Jones et al., 2019). More information on the reasons for dropout and the compliance per burst can be found

in the supplement (online Supplementary eResults 1 and eTable 3). There was an imbalance in the outcomes whereby the median percentage of BE episodes that patients with BN and BN/AUD experienced was 14% (Q1–Q3: 6–20%), with a median percentage of alcohol use episodes that patients with AUD and BN/AUD experienced of 14% (Q1–Q3: 8–18%), while the median percentage of BD episodes that patients with AUD and BN/AUD experienced was 4% (Q1–Q3: 2–7%).

Model performance

The performance metrics had a skewed distribution within and between participants. Therefore, the median across folds and across participants was used to describe them. An extended overview can be found in Table 3. A visual summary can be seen in Fig. 3. The confusion matrices of the predictions can be found the supplement (online Supplementary eTables 4–9).

Binge eating

A person-specific model could be fitted and evaluated for 48 (69.6%) participants. The performance of the person-specific models was poor with a median AUC of 0.61 (Q1:0.53; Q3:0.73), sensitivity of 0.83 (Q1:0.67; Q3:1.00), specificity of 0.71 (Q1:0.56; Q3:0.78), PPV of 0.31 (Q1:0.22; Q3:0.43), and NPV of 0.97 (Q1:0.91;Q3:1.00). The pooled model could be evaluated on 66 (95.7%) participants. Its performance was adequate with a median AUC of 0.71 (Q1:0.60; Q3:0.78), sensitivity of 01.00 (Q1:0.75; Q3: 1.00), specificity of 0.75 (Q1:0.60; Q3:0.86), PPV of 0.33 (Q1:0.23;Q3:0.50), and NPV of 1.00 (Q1:0.94; Q3:1.00).

Alcohol use

There were 43 (61.4%) participants with a person-specific model. The performance of these models was good with an AUC of 0.80 (Q1:0.72; Q3:0.89), sensitivity of 1.00 (Q1:0.79; Q3:1.00), specificity of 0.80 (Q1:0.75; Q3:0.88), PPV of 0.38 (Q1:0.28; Q3:0.57), and NPV of 1.00 (Q1:0.96; Q3:1.00). The pooled model could be evaluated on 63 (90.0%) participants. It had an outstanding performance with an AUC of 0.90 (Q1:0.83; Q3:0.96), sensitivity of 1.00 (Q1: 1.00; Q3: 1.00), specificity of 0.88 (Q1:0.80; Q3:0.96), PPV of 0.50 (Q1:0.37; Q3:0.75), and NPV of 1.00 (Q1:1.00; Q3:1.00).

Binge drinking

A person-specific model could be fitted and evaluated for 13 (18.6%) participants. The performance of the person-specific models was good with a median AUC of 0.85 (Q1: 0.71; Q3: 0.93), sensitivity of 1.00 (Q1:0.75; Q3:1.00), specificity of 0.90 (Q1:0.78; Q3: 0.96), PPV of 0.28 (Q1:0.18;Q3:0.50), and NPV of 1.00 (Q1:0.98;Q3:1.00). The performance of the pooled model could be evaluated on 49 (70.0%) participants. Its performance was outstanding with an AUC of 0.93 (Q1: 0.87; Q3: 0.98), sensitivity 1.00 (Q1: 1.00; Q3: 1.00), a specificity of 0.93 (Q1: 0.84; Q3: 0.99), PPV of 0.50 (Q1:0.20; Q3:0.83), and NPV of 1.00 (Q1:1.00; Q3:1.00).

Model comparison

The difference between the AUC of the person-specific and pooled models did not depend on patient group, and did not correlate with age, BMI, EDE-Q scores, AUDIT scores, BE frequency and BD frequency. In other words, having a higher AUC for the

Table 2. Sample characteristics

	AUD (n = 51)		BN and AUD (n = 19)		BN (n = 50)	
	Mean (s.d.)	95% CI	Mean (s.d.)	95% CI	Mean (s.d.)	95% CI
Age	21.5 (3.5)	20.5–22.4	20.4 (1.7)	19.6–21.2	22.4 (4.1)	21.3–23.6
BMI	22.8 (2.0)	22.2–23.4	24.2 (3.0)	22.8–25.7	25.6 (5.9)	23.9–27.3
Illness duration BN (years)	0 (0)	0–0	2.7 (1.4)	2.1–3.4	2.4 (1.6)	2.0–2.9
Illness duration AUD (years)	3.1 (1.3)	2.7–3.5	2.2 (1.4)	1.5–2.9	0 (0)	0–0
Education (years)	14.7 (1.7)	14.3–15.2	13.9 (1.6)	13.2–14.7	14.7 (2.3)	14.0–15.3
AUDIT	15.2 (5.3)	13.7–16.7	16.6 (4.3)	14.6–18.7	4.5 (4.2)	3.3–5.7
EDE-Q						
Restraint	0.7 (0.8)	0.5–1.0	2.7 (1.6)	1.9–3.5	2.9 (1.5)	2.4–3.3
Shape concern	1.9 (1.4)	1.5–2.3	4.0 (1.4)	3.3–4.7	4.3 (1.5)	3.9–4.7
Weight concern	1.5 (1.3)	1.1–1.9	4.0 (2.0)	3.0–5.0	4.1 (1.5)	3.7–4.5
Eating concern	0.5 (0.7)	0.3–0.7	2.9 (1.4)	2.2–3.5	2.8 (1.5)	2.4–3.2
Total	1.2 (1.0)	1.0–1.5	3.5 (1.3)	2.8–4.1	3.6 (1.3)	3.3–4.0
Eating disorder symptoms (days/4 weeks)						
Binge eating	0 (0)	0–0	6.5 (5.5)	3.8–9.1	8.7 (7.2)	6.7–10.8
Fasting	0 (0)	0–0	7.6 (7.5)	4.0–11.2	8.0 (8.4)	5.6–10.4
Vomiting	0 (0)	0–0	1.9 (3.8)	0.1–3.8	2.2 (6.1)	0.5–4.0
Laxative use	0 (0)	0–0	0.1 (0.5)	0–0.3	0.4 (2.3)	0–1.0
Diuretic use	0 (0)	0–0	0 (0)	0–0	1.1 (5.3)	0–2.6
Compensatory exercise	0 (0)	0–0	6.7 (7.1)	3.3–10.1	6.5 (7.4)	4.4–8.6
	N (%)	95% CI	N (%)	95% CI	N (%)	95% CI
Binge drinking frequency						
Never	0 (0%)	0–0%	0 (0%)	0–0%	27 (54%)	42–68%
Annually	0 (0%)	0–0%	0 (0%)	0–0%	8 (16%)	4–30%
Semiannually	0 (0%)	0–0%	0 (0%)	0–0%	2 (4%)	0–18%
Three-monthly	5 (10%)	0–25%	1 (5%)	0–31%	6 (12%)	0–26%
Monthly	9 (18%)	6–33%	3 (16%)	0–42%	3 (6%)	0–20%
Biweekly	21 (41%)	29–56%	9 (47%)	32–74%	3(6%)	0–20%
Weekly	9 (18%)	6–33%	5 (26%)	11–53%	1 (2%)	0–16%
>Weekly	7 (14%)	2–29%	1 (6%)	0–31%	0 (0%)	0–0%
Therapy (BN or AUD)	1 (2%)	0–6%	6 (32%)	9–54%	8 (16%)	6–26%
Race and ethnicity						
Caucasian	48 (94%)	90–100%	19 (100%)	100–100%	43 (86%)	78–95%
Latina	1 (2%)	0–8%	0 (0%)	0–0%	0 (0%)	0–0%
Asian	0 (0%)	0–0%	0 (0%)	0–0%	3 (6%)	0–15%
Multiracial	2 (4%)	0–10%	0 (0%)	0–0%	0 (0%)	0–0%
Middle-Eastern	0 (0%)	0–0%	0 (0%)	0–0%	4 (8%)	0–17%
Black	0 (0%)	0–0%	0 (0%)	0–0%	0 (0%)	0–0%
Psychoactive medication	6(12%)	3–21%	3 (16%)	0–33%	8 (16%)	6–26%
Comorbidities						
MDD	3 (6%)	0–16%	4 (21%)	5–47%	5 (10%)	0–24%
PD	2 (4%)	0–14%	1 (4%)	0–31%	4 (8%)	0–22%

(*Continued*)

Table 2. (*Continued.*)

	AUD (*n* = 51)		BN and AUD (*n* = 19)		BN (*n* = 50)	
	N (%)	95% CI	N (%)	95% CI	N (%)	95% CI
SAD	1 (2%)	0–13%	1 (4%)	0–31%	5 (8%)	0–24%
ADHD	3 (6%)	0–16%	0 (0%)	0–0%	0 (0%)	0–0%
PTSD	0 (0%)	0–0%	4 (21%)	5–47%	5 (10%)	0–24%
AP	0 (0%)	0–0%	0 (0%)	0–0%	4 (8%)	0–22%

Abbreviations: ADHD, attention deficit hyperactivity disorder; AP, agoraphobia; AUD, alcohol use disorder; BMI, body mass index; BN, bulimia nervosa; CI, confidence interval; EDE-Q, Eating Disorder Examination Questionnaire; MDD, major depressive disorder; N, number; PD, panic disorder; PTSD, post-traumatic stress disorder; SAD, social anxiety disorder; SD, standard deviation.

person-specific model than for the pooled model was not related to any of the patient characteristics we recorded. A more detailed analysis on the relation between the number of assessments and the AUC for the person-specific models can be found in the supplement (online Supplementary eResults 2 and eFigure 2).

Validation analyses

There was no difference in the occurrence of BE or alcohol use across days. However, BD happened more frequently on Thursday (Thursday v. Saturday: $\beta = 0.452$, S.E. = 0.099, $p < 0.001$; Friday *v.* Saturday: $\beta = 0.093$, S.E. = 0.106, $p = 0.377$). The pooled and person-specific models exhibited lower overall accuracy than models always predicting the majority case, but they demonstrated a better predictive performance (online Supplementary eResults 3 and eTable 10). Despite the lower overall accuracy of the original models due to outcome imbalance, they exhibited higher NPV and PPV. This suggests that when the original models predicted the likelihood of an event (i.e. BE or no BE), these predictions were more correct. Moreover, adjusting for the outcome imbalance led to poorer performance overall, except for the person-specific models for BE (see online Supplementary eResults 4 and eTable 11). This aligns with recent findings indicating that correcting for an outcome

imbalance may introduce bias (Van Den Goorbergh, Van Smeden, Timmerman, & Ben Van Calster, 2022). Additionally, incorporating app type as a predictor did not enhance model performance (see online Supplementary eResults 5 and eTable 12). However, patients who reported a higher percentage of their responses through m-Path exhibited poorer performance in the person-specific models for alcohol use ($\rho = -0.424$, $p = 0.002$), but this was not observed in other models (see online Supplementary eResults 5).

Model predictors

A visual summary of the 10% best predictors for each outcome and each model type can be seen in Fig. 4.
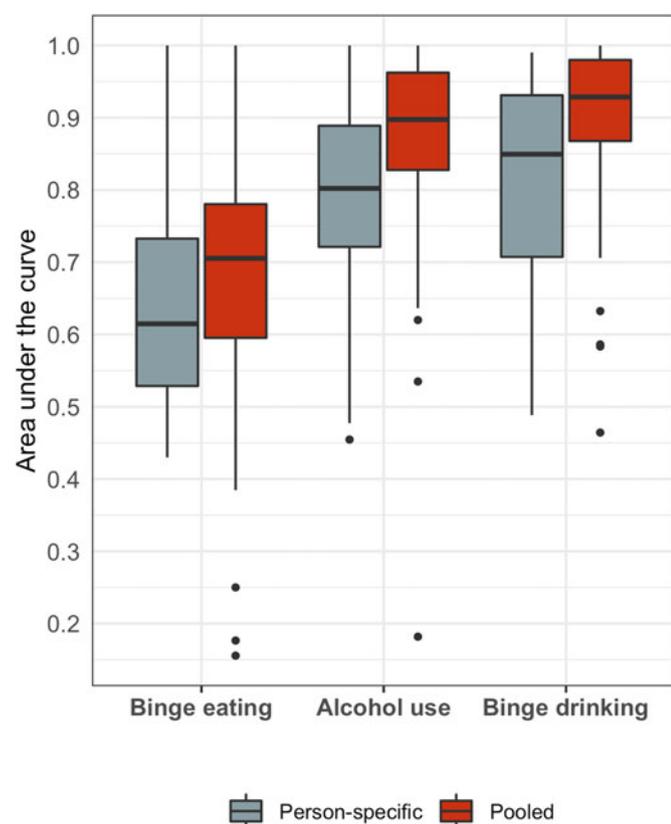
Binge eating

For the person-specific models, BE was positively predicted by craving for a BE episode, evening hours (i.e. ping 8), doing things that you regret, feeling down, and the pleasantness of the most important event. It was negatively predicted by feeling like you can handle the situation, feeling stressed, and night (i.e. ping 1). For the pooled model, the best positive predictors were evening (i.e. ping 7, ping 8) and craving for a BE episode. The best negative predictors were being alone, experiencing physical

Table 3. Model performance

Outcome	Type	CV aggregation	N (%)	AUC	Sensitivity	Specificity	Accuracy	PPV	NPV
Binge eating	Person-specific	Mean	48 (69.6%)	0.64	0.82	0.67	0.68	0.34	0.95
		Median		0.61	0.83	0.71	0.71	0.31	0.97
	Pooled	Mean	66 (95.7%)	0.69	0.88	0.73	0.73	0.41	0.96
		Median		0.71	1.00	0.75	0.74	0.33	1.00
Alcohol use	Person-specific	Mean	43 (61.4%)	0.80	0.90	0.80	0.80	0.43	0.98
		Median		0.80	1.00	0.80	0.80	0.38	1.00
	Pooled	Mean	63 (90.0%)	0.87	0.96	0.86	0.87	0.56	0.99
		Median		0.90	1.00	0.88	0.87	0.50	1.00
Binge drinking	Person-specific	Mean	14 (18.6%)	0.81	0.90	0.90	0.85	0.35	0.99
		Median		0.85	1.00	0.90	0.86	0.28	1.00
	Pooled	Mean	49 (70.0%)	0.90	1.00	0.89	0.90	0.54	1.00
		Median		0.93	1.00	0.93	0.93	0.50	1.00

Abbreviations: AUC, area under the curve; CV, cross-validation; N, number of participants with a successful model; NPV, negative predictive value; PPV, positive predictive value.
The performance metrics had a skewed distribution within and between participants. Therefore, they are best described by the median across folds and participants. To compare, the results after taking the mean across folds is also presented.

Figure 3. Model performance. Performance of the person-specific and pooled prediction models for binge eating, alcohol use, and binge drinking. Due to a skewed distribution of the performance metrics within participants, the median across folds was taken for the area under the curve.

discomfort, being under pressure, morning (ping 2), night (ping 1), being in a calm environment (i.e. passive leisure activities, experiencing no stressors), experiencing the suffering of others, and having laughed.

Alcohol use

For the person-specific models, alcohol use was positively predicted by craving alcohol, having drunk alcohol at the previous timepoint, evening (i.e. ping 7 and ping 8), pleasant company, experiencing pleasant events, and feeling cheerful. It was negatively predicted by wanting to do something. The best positive predictors for the pooled model were having drunk alcohol at the previous timepoint, evening (i.e. ping 7 and 8), smoking, and craving alcohol. The best negative predictors for the pooled model were morning/noon (ping 2, 3, 4, and 5), drinking caffeine, having eaten and being on your cellphone.

Binge drinking

For the person-specific models, BD was positively predicted by craving alcohol, being with friends, having drunk alcohol at the previous timepoint, evening (i.e. ping 8), night (i.e. ping 1), experiencing pleasant events, feeling satisfied, and doing groceries. However, BD was negatively predicted by a lack of perseverance, and feeling relaxed. For the pooled models, Important positive predictors were evening (i.e. ping 8), night (i.e. ping 1), having drunk alcohol at the previous timepoint, being with friends, smoking and experiencing no positive events. Important

negative predictors were noon (i.e. ping 3, 4, and 5), drinking coffee, studying/working, and having laughed.

Discussion

This study had two objectives. First, to build and evaluate person-specific and pooled prediction models for BE, alcohol use, and BD in patients with BN and/or AUD. Second, to identify the most important predictors of these behaviors.

Model performance

The AUC of the prediction models ranged from poor to outstanding, which was similar or slightly better than those in studies predicting eating behaviors and alcohol use in healthy volunteers (Goldstein et al., 2018; Soyster et al., 2021). Furthermore, the one study which implemented the use of ML in a JITAI has found that dietary lapses could be predicted and prevented with a ML model whose AUC was similar to that of the current study's pooled model for BE (Forman et al., 2019). This suggests that the pooled prediction models of the current study as well as the person-specific models for alcohol use and BD could be used in a JITAI for clinical use. However, this is nuanced by the median PPV of the different models in the current study, which ranges from 0.28 to 0.50. This means that patients engaged in binge eating, alcohol use, or binge drinking less than half of the times it was predicted by the models. This could be due to inaccuracies in the models or the possibility that patients were in an at-risk state but did not engage in these behaviors at the time (e.g. by exhibiting alternative behaviors or displaying the behaviors at a later time). However, having an NPV that is higher than the PPV is thought to be advantageous as the consequences of failing to send out an intervention are deemed to be more detrimental than intervening excessively. Notably, existing literature does not provide insights into the minimum PPV required for a JITAI to be effective without causing undue inconvenience by delivering excessive interventions, and this should be investigated by future studies.

In the current study, the performance of the prediction models for BE were lower than those for alcohol use and BD. This could be the result of how BE is defined, as it is more difficult to assess whether you have overeaten and lost control than whether you have drunk alcohol. This could be particularly evident in the current study, as no specific caloric amount was used to define overeating. Furthermore, it could be that some important predictors were not included in the current study. This could be remediated by basing the ESM items on the reported triggers of the individual patients, as was done in a previous study on BE (Arend et al., 2023). Lastly, it could be that other ML analysis techniques would be better suited. For example, it could be that elastic net regression is not well adapted to predict BE and that other ML techniques would result in a better predictive performance.

The results also showed that the pooled prediction models outperformed the person-specific ones. Indeed, some studies found that pooled models have a better predictive performance than person-specific ones (Ntekouli et al., 2022; Soyster et al., 2021). However, others reported the opposite to be true (Cheung et al., 2017; Rozet, Kronish, Schwartz, and Davidson, 2019). There could be several reasons why the current study finds that pooled models have a higher predictive performance than person-specific ones. On the one hand, the difference in performance could be
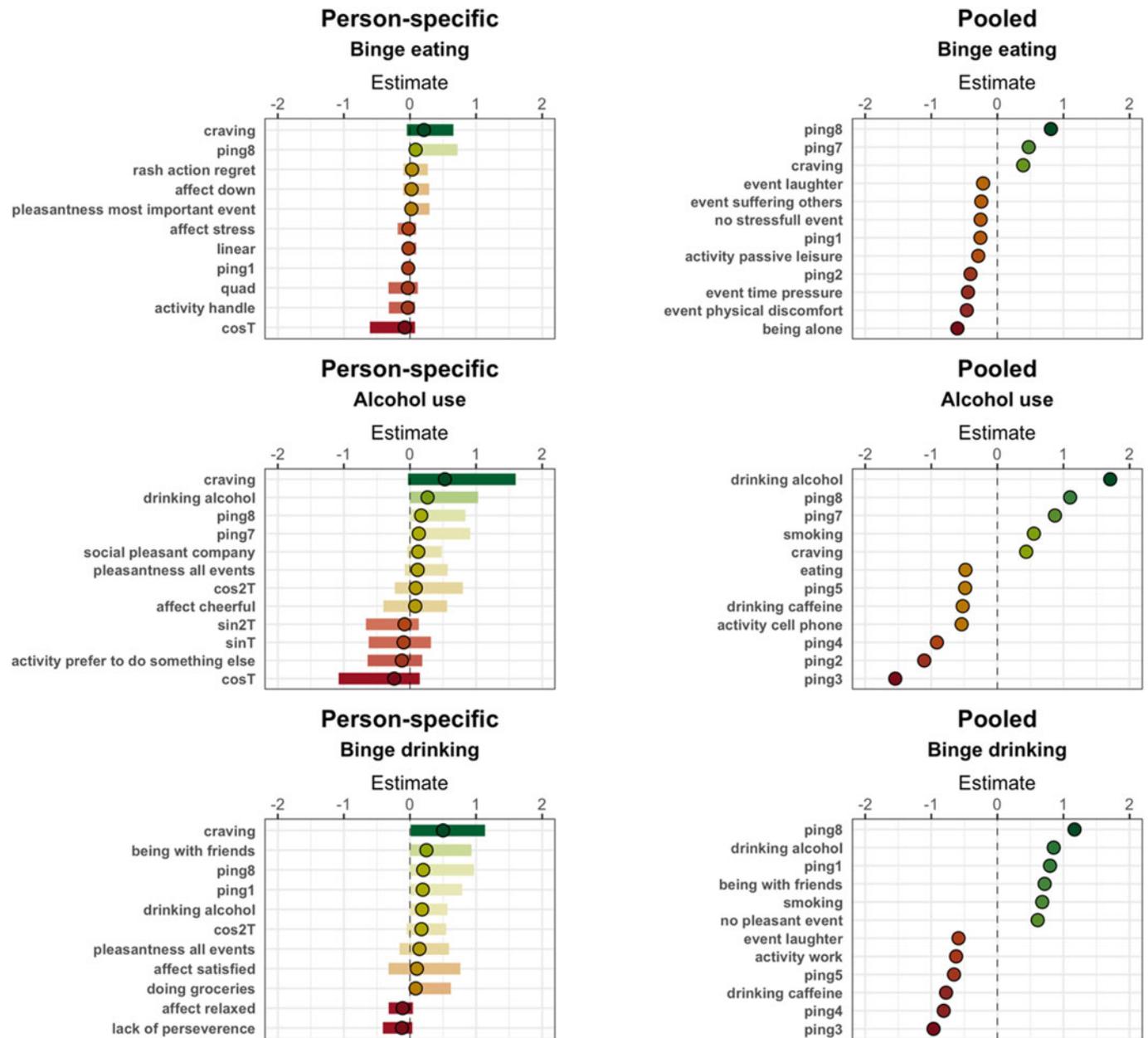
Figure 4. Model predictors. The predictors of the person-specific and pooled prediction models with the 10% highest estimates for binge eating, alcohol use, and binge drinking. For the person-specific models, the mean estimate and 95% interval across all participants is shown. For the pooled predictions models, the single estimate is displayed. Linear, quad, cosT, cos2T, sinT, and sin2T represent the linear, quadratic, cosinusoidal (24 h frequency), cosinusoidal (12 h frequency), sinusoidal (24 h frequency), and sinusoidal (12 h frequency) effect of time since participating in the study. Important to note, the drinking alcohol variable represents having drunk alcohol at the previous timepoint. Furthermore, craving meant craving for a binge eating episode for the binge eating outcome and craving for alcohol for the alcohol use and binge drinking outcomes.

the result of the substantially larger number of observations that were used to train the pooled models. Indeed, studies show that ML performance is related to dataset size (Althnian et al., 2021). One study also found that the performance of person-specific models increases with a greater number of observations until it is similar to or exceeds that of pooled models (Rozet et al., 2019). On the other hand, the better performance of the pooled models could be the result of the characteristics of the participants. Namely, previous studies show that group-level methods lend themselves well to samples that are homogenous (i.e. with a low inter-individual variability), which is the case

in the current study (Fisher, Medaglia, & Jeronimus, 2018; Molenaar, 2004).

Taken together, these findings raise the question whether pooled prediction models would translate well to a clinical setting, where it could be difficult to gather a large dataset and where more inter-individual variability is seen. However, building a person-specific model could present its own challenges. Namely, as the performance of prediction models is thought to be related to sample size, a patient would have to be followed for a considerable amount of time and need to answer a large number of assessments before a person-specific model could be fit.

However, in clinical practice, it might be difficult to observe patients for longer periods of time before moving on to treatment, and some patients may struggle to answer a large amount of assessments. In that case, it could be difficult to observe enough events of interest, which could be helped by aggregating small amounts of data from multiple patients in a pooled model. Therefore, as both person-specific and pooled models have their benefits and disadvantages, future studies should investigate a combination of both approaches. Indeed, studies bridging this gap show encouraging results, but further research is needed (Goldstein et al., 2018; Ren et al., 2023). Furthermore, studies should investigate which sampling frequency provides a treatment response with only a minimum of patient burden. In the current study, patients needed to answer eight assessments per day, which might be too demanding and excessive in the context of a JITAI.

Most important predictors

There were significant differences between the most important predictors of the person-specific and pooled prediction models. This is not unexpected as previous studies have shown that the agreement between person-level and group-level analyses is limited (Fischer et al., 2018). Furthermore, the differences between the types of predictors could have important implications for the development of JITAIs. Namely, it could be challenging to develop interventions that target the predictors of the pooled models as these mostly concern the time of day (e.g. evening or night) or recent events (e.g. experiencing something boring or being under pressure). It might be more valuable to focus on the predictors of the person-specific models as they deal with thoughts (e.g. craving), emotions (e.g. negative affect, positive affect), and behaviors (e.g. acting rash). This suggests that though pooled models might have a better performance, person-specific models could still be of value when it comes to tailoring daily life interventions.

The results also showed that there are both similarities and differences in the predictors for BE, alcohol use, and BD. First, it can be seen that craving was the most important predictor across the person-specific models of all behaviors. Though the relation between craving and alcohol use has been investigated by a large number of studies, this is less the case for BE (Cavicchioli, Vassena, Movalli, & Maffei, 2020; Novelle & Diéguez, 2018; Seo & Sinha, 2014). Future studies should therefore investigate the relationship between craving and BE in more depth. Second, positive events (i.e. the pleasantness of all events) and affect (i.e. feeling cheerful and satisfied) were important predictors of alcohol use and BD. This showcases the hypothesized link between positive emotions and alcohol consumption (Cooper, Frone, Russell, & Mudar, 1995). However, the pleasantness of the most important event was also a predictor of BE. Though studies have shown that positive affect often decreases before a BE episode, other studies indicate that patients who act more rashly when positive affect is high also have a higher BE frequency (Michael & Juarascio, 2021; Schaefer et al., 2020). Future studies should therefore explore whether positive emotions can also be a trigger for BE episodes in patients with BN. Third, BE was varyingly predicted by the different negative emotions. More specifically, it was positively predicted by feeling down and negatively predicted by feeling stressed. Similarly, a recent machine learning study in adolescents reports that loss of control eating is positively predicted by feeling lonely and feeling rejected, but negatively predicted

by feeling stressed (Hagan, Leenaerts, Walsh, & Ranzenhofer, 2024). Indeed, though studies show that negative emotions can trigger BE, others also have found that negative emotions can also lead to dietary restriction and that some emotions are more related to BE than others (Berg et al., 2013; Haedt-Matt & Keel, 2011; Leenaerts, Vaessen, Sunaert, Ceccarini, & Vrieze, 2023a; Mikhail, 2021). Furthermore, studies in the general population as well as rodents show that a strong acute stressor with pronounced physical responses decreases food intake while a mild chronic stressor increases consumption of energy-dense food (O'connor, Jones, Conner, Mcmillan, & Ferguson, 2008; Torres & Nowson, 2007). In the same line, the current study finds that experiencing physical discomfort and being under time pressure is related to a lower probability of BE. However, the results also highlight that stressors do play a role in BE, as experiencing no stressors is related to a lower chance of BE. Future studies should therefore explore when negative emotions lead to BE and when they lead to dietary restriction. Contrastingly, negative emotions or events were not included in the most important predictors of alcohol use or BD. Though the induction of negative affect has been shown to lead to increases in alcohol consumption in a laboratory context, a recent meta-analysis reports that this is not the case in daily life (Bresin, Mekawi, & Verona, 2018; Dora et al., 2022). This could be one reason why negative emotions were not an important predictor of alcohol use and BD in the current study. However, in our own recent work, we found that there was a relation between negative affect and alcohol use, but a non-linear one (Leenaerts, Vaessen, Sunaert, Ceccarini, & Vrieze, 2023b). This could be another reason why the elastic net regression did not retain negative emotions as an important predictor as it assumes a linear relation between variables. Interestingly, being alone was a negative predictor of BE in the pooled prediction model. This seems to contradict previous research showing that patients often binge eat when they are alone (Stickney, Miltenberger, & Wolff, 1999). However, it is important to consider that being alone in the current study was a predictor at the previous timepoint. It could therefore be that being alone at the previous timepoint is a negative predictor, as this means that the patients are not among other people, and are therefore less likely to experience interpersonal stressors, which are known to be linked to BE (Goldschmidt et al., 2014).

Limitations

This study has several limitations. Importantly, the sample mostly consisted of female participants who were Caucasian and had a short illness duration, resulting in a higher degree of homogeneity compared to the broader population of patients with AUD and/or BN. Consequently, the generalizability of the study's findings may be impacted. First, it's possible that predictors are differently associated with the outcomes in individuals from other backgrounds concerning sex, race, and illness duration. Second, a pooled model may not perform as effectively when applied to a more heterogeneous sample as inter-individual differences would make it more difficult for the machine learning algorithm to learn relations in the data. However, person-specific models might still demonstrate similar performance levels in a more heterogeneous sample, as their efficacy is not impacted by inter-individual differences. Third, individuals can binge eat, drink alcohol, and binge drink without meeting the criteria for AUD and/or BN. Future studies should therefore explore the efficacy of prediction models

in subclinical populations. Additionally, there are several methodological limitations. First, assessments were only sent out on Thursday, Friday, and Saturday. In the current study, there was no difference in the frequency of BE or alcohol use across days, but BD happened more often on Thursdays, which is in line with the findings of previous studies (Lavender et al., 2016; Van Damme et al., 2022). Nevertheless, participants could still have adapted their behavior on days when data was collected, which could have impacted the study results. Second, data were collected through more than one app. Though including app type as a predictor did not improve performance, participants who responded more through m-Path had a lower AUD in the person-specific models for AUD. This could indicate that app type was not directly related to the outcomes but could have an indirect impact on model performance (e.g. through answer times). Third, the nested nature of the data is partially accounted for in the pooled models by standardizing the continuous variables at a within-person level. Though this is a suboptimal way to handle nested data, this method was chosen as mixed effects elastic net regularized regression are not implemented in R. There are regression-based techniques that better account for multilevel data (i.e. mixed effects LASSO regression), but these techniques struggle with multi-dimensional data and highly correlated variables. Fourth, the majority of the predictors where based on ESM items, which are self-report items, and could have been biased. Future studies should explore the value of combining actively (i.e. ESM items) and passively gathered data (e.g. physiology). Fifth, the intervals between the assessments were uneven, and this is not accounted for by an elastic net model. Studies should therefore explore the use of continuous-time models which do not assume an even interval between assessments. Sixth, the cross-validation divided the data at random, which does not reflect how a JITAI would function, as a model would first be fit on the data of a patient, after which it would be applied to new registrations of a patient. Seventh, though no sample size calculations exist for elastic net regularized regression, studies on other regularized regressions techniques suggest that some person-specific models in the current study might suffer from a small sample size (Riley et al., 2020). Eight, there was a considerable amount of missing data which could have influenced the results. This was especially the case for the patient with both AUD and BN as they displayed a substantially lower compliance from the beginning of the study. Though no difference in the alcohol use or eating disorder characteristics was found between the patients with AUD/BN and the patients with either AUD or BN, it can be that the combination of experiencing difficulties with alcohol use and binge eating has a significantly larger impact on the patients, thereby making it more difficult for them to answer the ESM assessments. A popular technique to handle missingness is multiple imputation by chained equations (MICE), but this method struggles with the correlation between observations across time. However, there are promising deep learning-based methods, which future studies evaluate for clinical prediction models (Kazijevs & Samad, 2023). Ninth, only one individual assessed whether the participants met the criteria for the DSM-diagnoses, making it difficult to assess the reliability of the diagnoses. Tenth, it was not possible to assess the validity of several ESM items in the current study, nor was there any information available on the validity of these items from previous studies. It is necessary for future studies to validate their ESM items to enhance the robustness of their prediction models.

Conclusion

This study builds and evaluates person-specific and pooled prediction models for BE, alcohol use, and BD in patients with BN and/or AUD. The performances of the different models vary between poor and outstanding, but the pooled models outperform the person-specific ones and the models for alcohol use and BD outperform those for BE. This study also explores which variables are the most important predictors in the different models. Here, the predictors of the pooled models mostly concern the time of day and recent events, while those of the person-specific models mostly concern thoughts, emotions, and behaviors. Future studies should explore whether pooled and person-specific approaches could be combined and how BE, alcohol use, and BD can be impacted by interventions in daily life.

References

Allison, S., & Timmerman, G. M. (2007). Anatomy of a binge: Food environment and characteristics of nonpurge binge episodes. *Eating Behaviors*, 8(1), 31–38. https://doi.org/10.1016/J.EATBEH.2005.01.004

Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A. b., Alzakari, N., … Kurdi, H. (2021). Impact of dataset size on classification performance: An empirical evaluation in the medical domain. *Applied Sciences 2021*, Vol. 11, Page 796, 11(2), 796. https://doi.org/10.3390/APP11020796

Ambwani, S., Roche, M. J., Minnick, A. M. (2015). Negative affect, interpersonal perception, and binge eating behavior: An experience sampling study. *The International Journal of Eating Disorders*, 48(6), 715–726. https://doi.org/10.1002/eat.22410

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC.

American Psychiatric Association. (2017). *SCID-5-S Gestructureerd klinisch interview voor DSM-5 Syndroomstoornissen. Nederlandse vertaling van Structured Clinical Interview for DSM-5 Disorders- Clinician Version (SCID-5-CV)*. (1st ed.).

Arcelus, J., Mitchell, A. J., Wales, J., & Nielsen, S. (2011). Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. *Archives of General Psychiatry*, 68(7), 724–731. https://doi.org/10.1001/ARCHGENPSYCHIATRY.2011.74

Arend, A. K., Kaiser, T., Pannicke, B., Reichenberger, J., Naab, S., Voderholzer, U., … Blechert, J. (2023). Toward individualized prediction of binge-eating episodes based on ecological momentary assessment data: Item development and pilot study in patients With bulimia nervosa and binge-eating disorder. *JMIR Medical Informatics*, 11, e41513. https://doi.org/10.2196/41513

Bae, S., Chung, T., Ferreira, D., Dey, A. K., & Suffoletto, B. (2018). Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addictive Behaviors*, *83*, 42. https://doi.org/10.1016/J.ADDBEH.2017.11.039

Bae, S., Ferreira, D., Suffoletto, B., Puyana, J. C., Kurtz, R., Chung, T., & Dey, A. K. (2017). Detecting drinking episodes in young adults using smartphone-based sensors. Proceedings of the ACM on Interactive, Mobile, *Wearable and Ubiquitous Technologies*, *1*(2), 1–36. https://doi.org/10.1145/3090051

Berg, K. C., Crosby, R. D., Cao, L., Peterson, C. B., Engel, S. G., Mitchell, J. E., & Wonderlich, S. A. (2013). Facets of negative affect prior to and following binge-only, purge-only, and binge/purge events in women with bulimia nervosa. *Journal of Abnormal Psychology*, *122*(1), 111. https://doi.org/10.1037/A0029703

Boness, C. L., Watts, A. L., Moeller, K. N., & Sher, K. J. (2021). The etiologic, theory-based, ontogenetic hierarchical framework of alcohol Use disorder: A translational systematic review of reviews. *Psychological Bulletin*, *147*(10), 1075. https://doi.org/10.1037/BUL0000333

Bresin, K., Mekawi, Y., & Verona, E. (2018). The effect of laboratory manipulations of negative affect on alcohol craving and Use: A meta-analysis. *Psychology of Addictive behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, *32*(6), 617. https://doi.org/10.1037/ADB0000383

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, *300*, 70–79. https://doi.org/10.1016/J.NEUCOM.2017.11.077

Carpenter, S. M., Menictas, M., Nahum-Shani, I., Wetter, D. W., & Murphy, S. A. (2020). Developments in mobile health just-in-time adaptive interventions for addiction science. *Current Addiction Reports*, *7*(3), 280–290. https://doi.org/10.1007/S40429-020-00322-Y/TABLES/2

Cavicchioli, M., Vassena, G., Movalli, M., & Maffei, C. (2020). Is craving a risk factor for substance use among treatment-seeking individuals with alcohol and other drugs use disorders? A meta-analytic review. *Drug and Alcohol Dependence*, *212*. https://doi.org/10.1016/J.DRUGALCDEP.2020.108002

Cheung, Y. K., Hsueh, P. Y. S., Qian, M., Yoon, S., Meli, L., Diaz, K. M., … Davidson, K. W. (2017). Are nomothetic or ideographic approaches superior in predicting daily exercise behaviors? Analyzing N-of-1 mHealth data. *Methods of Information in Medicine*, *56*(6), 452. https://doi.org/10.3414/ME16-02-0051

Clapp, J. D., Shillington, A. M., & Segars, L. B. (2009). Deconstructing contexts of binge drinking Among college students. Http://Dx.Doi.Org/10.1081/ADA-100100596, *26*(1), 139–154. https://doi.org/10.1081/ADA-100100596

Cooper, M. L., Frone, M. R., Russell, M., & Mudar, P. (1995). Drinking to regulate positive and negative emotions: A motivational model of alcohol use. *Journal of Personality and Social Psychology*, *69*, 990–1005.

DeWitt P. (2019). ensr: Elastic Net SearcheR. https://CRAN.R-project.org/package=ensr

Dora, J., Piccirillo, M., Foster, K. T., Arbeau, K., Armeli, S., Auriacombe, M., … King, K. M. (2022). The daily association between affect and alcohol use: A meta-analysis of individual participant data. *Psychological Bulletin*, *149*(1–2), 1–24. https://doi.org/10.1037/bul0000387

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of Eating Disorders: Interview or Self-Report Questionnaire? https://doi.org/10.1002/1098-108X

Fischer, S., Wonderlich, J., Breithaupt, L., Byrne, C., & Engel, S. (2018). Negative urgency and expectancies increase vulnerability to binge eating in bulimia nervosa. *Eating Disorders*, *26*(1), 39–51. https://doi.org/10.1080/10640266.2018.1418253

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(27), E6106–E6115. https://doi.org/10.1073/PNAS.1711978115/-/DCSUPPLEMENTAL

Fleury, M. J., Djouini, A., Huỳnh, C., Tremblay, J., Ferland, F., Ménard, J. M., & Belleville, G. (2016). Remission from substance use disorders: A systematic review and meta-analysis. *Drug and Alcohol Dependence*, *168*, 293–306. https://doi.org/10.1016/J.DRUGALCDEP.2016.08.625

Flury, B. D., & Levri, E. P. (1999). Periodic logistic regression. *Ecology*, *80*(7), 2254–2260. https://doi.org/10.1890/0012-9658

Forman, E. M., Goldstein, S. P., Crochiere, R. J., Butryn, M. L., Juarascio, A. S., Zhang, F., & Foster, G. D. (2019). Randomized controlled trial of OnTrack, a just-in-time adaptive intervention designed to enhance weight loss. *Translational Behavioral Medicine*, *9*(6), 989–1001. https://doi.org/10.1093/TBM/IBZ137

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

Galmiche, M., Déchelotte, P., Lambert, G., & Tavolacci, M. P. (2019). Prevalence of eating disorders over the 2000–2018 period: A systematic literature review. *The American Journal of Clinical Nutrition*, *109*(5), 1402–1413. https://doi.org/10.1093/AJCN/NQY342

Goldschmidt, A. B., Wonderlich, S. A., Crosby, R. D., Engel, S. G., Lavender, J. M., Peterson, C. B., … Mitchell, J. E. (2014). Ecological momentary assessment of stressful events and negative affect in bulimia nervosa. *Journal of Consulting and Clinical Psychology*, *82*(1), 30. https://doi.org/10.1037/A0034974

Goldstein, S. P., Zhang, F., Thomas, J. G., Butryn, M. L., Herbert, J. D., & Forman, E. M. (2018). Application of machine learning to predict dietary lapses during weight loss. *Journal of Diabetes Science and Technology*, *12*(5), 1045. https://doi.org/10.1177/1932296818775757

Griswold, M. G., Fullman, N., Hawley, C., Arian, N., Zimsen, S. R. M., Tymeson, H. D., … Gakidou, E. (2018). Alcohol use and burden for 195 countries and territories, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *392*(10152), 1015–1035. https://doi.org/10.1016/S0140-6736(18)31310-2

Haedt-Matt, A. A., & Keel, P. K. (2011). Revisiting the affect regulation model of binge eating: A meta-analysis of studies using ecological momentary assessment. *Psychological Bulletin*, *137*(4), 660. https://doi.org/10.1037/A0023660

Hagan, K. E., Leenaerts, N., Walsh, B. T., & Ranzenhofer, L. (2024). Applying machine learning to ecological momentary assessment data to identify predictors of future loss-of-control eating and overeating in adolescents. https://doi.org/10.31234/OSF.IO/A5TS9

Hardeman, W., Houghton, J., Lane, K., Jones, A., & Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. *The International Journal of Behavioral Nutrition and Physical Activity*, *16*(1). https://doi.org/10.1186/S12966-019-0792-7

Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H. A., Wen, C. K. F., & Field, M. (2019). Compliance with ecological momentary assessment protocols in substance users: A meta-analysis. Addiction (Abingdon, England), *114*(4), 609. https://doi.org/10.1111/ADD.14503

Kazijevs, M., & Samad, M. D. (2023). Deep imputation of missing values in time series health data: A review with benchmarking. *Journal of Biomedical Informatics*, *144*, 104440. https://doi.org/10.1016/J.JBI.2023.104440

Kira, B., Saptarshi, M. L., Thayslene, M., Oliveira, M., Nagesh, R., Phillips, T., … Hallas, L. (2022). BSG Working Paper Series Variation in government responses to COVID-19. Retrieved from www.bsg.ox.ac.uk/covidtracker

Koob, G. F., & Le Moal, M. (1997). Drug abuse: Hedonic homeostatic dysregulation. *Science (New York, N.Y.)*, *278*(5335), 52–58. https://doi.org/10.1126/SCIENCE.278.5335.52/ASSET/52EECCCA-8686-4735-B10F-F57BEB0935F1/ASSETS/GRAPHIC/SE4075802004.JPEG

Kuhn M. (2021). caret: Classification and Regression Training. Retrieved from https://CRAN.R-project.org/package=caret

Lavender, J. M., Utzinger, L. M., Crosby, R. D., Goldschmidt, A. B., Ellison, J., Wonderlich, S. A., … Le Grange, D. (2016). A naturalistic examination of the temporal patterns of affect and eating disorder behaviors in anorexia nervosa. *The International Journal of Eating Disorders*, *49*(1), 77–83. https://doi.org/10.1002/EAT.22447

Leenaerts, N., Vaessen, T., Sunaert, S., Ceccarini, J., & Vrieze, E. (2023a). How negative affect does and does not lead to binge eating—The importance of craving and negative urgency in bulimia nervosa. *Journal of Psychopathology and Clinical Science*, *132*(5), 621–633. https://doi.org/10.1037/ABN0000830

Leenaerts, N., Vaessen, T., Sunaert, S., Ceccarini, J., & Vrieze, E. (2023b). The impact of negative and positive affect on craving, non-heavy alcohol use, and binge drinking in patients with alcohol use disorder and controls: An

experience sampling method study. https://doi.org/10.31234/OSF.IO/PR2SV

Levinson, C. A., Trombley, C. M., Brosof, L. C., Williams, B. M., & Hunt, R. A. (2022). Binge eating, purging, and restriction symptoms: Increasing accuracy of prediction using machine learning. *Behavior Therapy*, 54(2), 247–259. https://doi.org/10.1016/J.BETH.2022.08.006

Linardon, J., & Wade, T. D. (2018). How many individuals achieve symptom abstinence following psychological treatments for bulimia nervosa? A meta-analytic review. *The International Journal of Eating Disorders*, 51(4), 287–294. https://doi.org/10.1002/EAT.22838

Meers, K., Dejonckheere, E., Kalokerinos, E. K., Rummens, K., & Kuppens, P. (2020). MobileQ: A free user-friendly application for collecting experience sampling data. *Behavior Research Methods* 2019 52:4, 52(4), 1510–1515. https://doi.org/10.3758/S13428-019-01330-1

Mestdagh, M., Verdonck, S., Piot, M., Niemeijer, K., Tuerlinckx, F., Kuppens, P., & Dejonckheere, E. (2022). m-Path: An easy-to-use and flexible platform for ecological momentary assessment and intervention in behavioral research and clinical practice. https://doi.org/10.31234/OSF.IO/UQDFS

Michael, M. L., & Juarascio, A. (2021). Elevated cognitive dietary restraint mediates the relationship between greater impulsivity and more frequent binge eating in individuals with binge-spectrum eating disorders. *Eating and Weight Disorders: EWD*, 26(8), 2795–2800. https://doi.org/10.1007/S40519-021-01153-0

Mikhail, M. (2021). Affect dysregulation in context: Implications and future directions of experience sampling research on affect regulation models of loss of control eating. *Frontiers in Psychiatry*, 12. https://doi.org/10.3389/FPSYT.2021.747854

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology. This Time Forever, *Measurement*, 2(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, 52(6), 446. https://doi.org/10.1007/S12160-016-9830-8

National Institute on Alcohol Abuse and Alcoholism [NIAAA]. (2022). Drinking Levels Defined. Retrieved August 10, 2022, from https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking

Novelle, M. G., & Diéguez, C. (2018). Food addiction and binge eating: Lessons learned from animal models. *Nutrients*, 10(1). https://doi.org/10.3390/NU10010071

Ntekouli, M., Spanakis, G., Waldorp, L., & Roefs, A. (2022). Using explainable boosting machine to compare idiographic and nomothetic approaches for ecological momentary assessment data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13205 LNCS, 199–211. https://doi.org/10.1007/978-3-031-01333-1_16

O'connor, D. B., Jones, F., Conner, M., Mcmillan, B., & Ferguson, E. (2008). Effects of daily hassles and eating style on eating behavior. https://doi.org/10.1037/0278-6133.27.1(Suppl.).S20

Pearson, C. M., Wonderlich, S. A., & Smith, G. T. (2015). A risk and maintenance model for bulimia nervosa: From impulsive action to compulsive behavior. *Psychological Review*, 122(3), 516–535. https://doi.org/10.1037/A0039268

Ren, B., Balkind, E. G., Pastro, B., Israel, E. S., Pizzagalli, D. A., Rahimi-Eichi, H., … Webb, C. A. (2023). Predicting states of elevated negative affect in adolescents from smartphone sensors: A novel personalized machine learning approach. *Psychological Medicine*, 53(11), 5146–5154. https://doi.org/10.1017/S0033291722002161

Riley, R. D., Ensor, J., Snell, K. I. E., Harrell, F. E., Martin, G. P., Reitsma, J. B., … Van Smeden, M. (2020). Calculating the sample size required for developing a clinical prediction model. *BMJ*, 368. https://doi.org/10.1136/BMJ.M441

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(1), 1–8. https://doi.org/10.1186/1471-2105-12-77/TABLES/3

Rozet, A., Kronish, I. M., Schwartz, J. E., & Davidson, K. W. (2019). Using machine learning to derive just-In-time and personalized predictors of stress: Observational study bridging the Gap between nomothetic and ideographic approaches. *Journal of Medical Internet Research*, 21(4), e12910. https://doi.org/10.2196/12910

Saunders, J. B., Aasland, O. G., Babor, T. F. E., De La Fuente, J. R., & Grant, M. (1993). Development of the alcohol Use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction (Abingdon, *England*), 88(6), 791–804. https://doi.org/10.1111/J.1360-0443.1993.TB02093.X

Schaefer, L. M., Smith, K. E., Anderson, L. M., Cao, L., Crosby, R. D., Engel, S. G., … Wonderlich, S. A. (2020). The role of affect in the maintenance of binge-eating disorder: Evidence from an ecological momentary assessment study. *Journal of Abnormal Psychology*, 129(4), 387–396. https://doi.org/10.1037/ABN0000517

Seo, D., & Sinha, R. (2014). The neurobiology of alcohol craving and relapse. *Handbook of Clinical Neurology*, 125, 355–368. https://doi.org/10.1016/B978-0-444-62619-6.00021-5

Shahriyari, L. (2019). Effect of normalization methods on the performance of supervised learning algorithms applied to HTSeq-FPKM-UQ data sets: 7SK RNA expression as a predictor of survival in patients with colon adenocarcinoma. *Briefings in Bioinformatics*, 20(3), 985–994. https://doi.org/10.1093/BIB/BBX153

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, 49(9), 1426–1448. https://doi.org/10.1017/S0033291719000151

Soyster, P. D., Ashlock, L., & Fisher, A. J. (2021). Pooled and person-specific machine learning models for predicting future alcohol consumption, craving, and wanting to drink: A demonstration of parallel utility. *Psychology of Addictive Behaviors*, 36(3), 296–306. https://doi.org/10.1037/ADB0000666

Stickney, M. I., Miltenberger, R. G., & Wolff, G. (1999). A descriptive analysis of factors contributing to binge eating. *Journal of Behavior Therapy and Experimental Psychiatry*, 30(3), 177–189. https://doi.org/10.1016/S0005-7916(99)00019-1

Torres, S. J., & Nowson, C. A. (2007). Relationship between stress, eating behavior, and obesity. *Nutrition (Burbank, Los Angeles County, Calif.)*, 23(11–12), 887–894. https://doi.org/10.1016/J.NUT.2007.08.008

Van Damme, J., Thienpondt, A., Rosiers, J., Tholen, R., Soyez, V., Sisk, M., … Rapport, D. (2022). In hogere sferen Volume 5: Een onderzoek naar middelengebruik bij de Vlaamse studenten. Retrieved from www.vad.be

Van Den Goorbergh, R., Van Smeden, M., Timmerman, & D., & Ben Van Calster. (2022). The harm of class imbalance corrections for risk prediction models: Illustration and simulation using logistic regression. *Journal of the American Medical Informatics Association*, 29(9), 1525–1534. https://doi.org/10.1093/JAMIA/OCAC093

Walters, S. T., Businelle, M. S., Suchting, R., Li, X., Hébert, E. T., & Mun, E. Y. (2021). Using machine learning to identify predictors of imminent drinking and create tailored messages for at-risk drinkers experiencing homelessness. *Journal of Substance Abuse Treatment*, 127, 108417. https://doi.org/10.1016/J.JSAT.2021.108417

Wang, L., & Miller, L. C. (2020). Just-in-the-moment adaptive interventions (JITAI): A meta-analytical review. *Health Communication*, 35(12), 1531–1544. https://doi.org/10.1080/10410236.2019.1652388

Way, T. W., Sahiner, B., Hadjiiski, L. M., & Chan, H. P. (2010). Effect of finite sample size on feature selection and classification: A simulation study. *Medical Physics*, 37(2), 907. https://doi.org/10.1118/1.3284974

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, 67(2), 301–320.