**Supplemental Online Content**

The following items were included to give the readers more information

**P2.** eMethods 1. App characteristics

**P2.** eTable 1. App characteristics

**P3.** eMethods 2. Psychometric properties of the experience sampling method questions

**P6.** eFigure 1. Data preprocessing

**P7.** eMethods 3. Additional predictors

**P7.** eMethods 4. Elastic net wrappers

**P8.** eMethods 5. Nested k-fold cross-validation

**P9.** eTable 2. Sample characteristics per analysis group

**P10.** eResults 1. Reasons for dropout

**P10.** eTable 3. Mean participant compliance (%) per burst per participant group

**P10.** eTable 4. Confusion matrix of the person-specific models for binge eating

**P10.** eTable 5. Confusion matrix of the pooled models for binge eating

**P11.** eTable 6. Confusion matrix of the person-specific models for alcohol use

**P11.** eTable 7. Confusion matrix of the pooled models for alcohol use

**P11.** eTable 8. Confusion matrix of the person-specific models for binge drinking

**P11.** eTable 9. Confusion matrix of the pooled models for binge drinking

**P11.** eResults 2. Relation between number of assessments and area under the curve for the person-specific models

**P12.** eFigure 2. Relation between number of assessments and area under the curve for the person-specific models

**P12.** eResults 3. Comparing model performance with a model which always predicts the majority case

**P13.** eTable 10. Comparing model performance with a model which always predicts the majority case

**P13.** eResults 4. Correcting for an imbalance in the outcome

**P13.** eTable 11. Correcting for an imbalance in the outcome

**P14.** eResults 5. The impact of app type on model performance

**P14.** eTable 12. The impact of app type on model performance

**P15.** eReferences.

**eMethods 1. App characteristics**

Both MobileQ and m-Path were open source and developed at KU Leuven, Belgium. The apps were installed on the smartphone of the participant. If the app didn’t function on their smartphone, a compatible smartphone was lent out. As the development of MobileQ was discontinued in October 2020, data collection continued using m-Path. The participants received the same questions and responded using the same answers (e.g., yes-no questions, multiple choice questions) through m-Path as they did through MobileQ. However, the assessment schedule differed as MobileQ randomly dispersed the pings between the average wake-up and bed times of the participants, while m-path randomly sent out a ping within a measurement window that is predefined by the user. Furthermore, MobileQ requires users to answer the ping immediately with the ping ending in approximately 90 seconds, while m-Path allows users to answer the ping until the next ping.

|  |
| --- |
| **eTable 1. App characteristics** |
|  | MobileQ | m-Path |
| Internet access required | No | Yes |
| Begin time | Average wake-up time  | 8:30 |
| End time | Average bed time  | 22:00 |
| Start time ping 1, median (Q1-Q3) | 9:00 (8:56-11:00) | 9:21 (8:56-9:49) |
| Start time ping 2, median (Q1-Q3) | 11:29 (10:33-13:20) | 11:16 (10:37-12:34) |
| Start time ping 3, median (Q1-Q3) | 13:37 (12:19-15:47) | 13:10 (12:35-15:07) |
| Start time ping 4, median (Q1-Q3) | 15:38 (14:12-17:42) | 15:04 (14:10-17:13) |
| Start time ping 5, median (Q1-Q3) | 17:15 (16:00-19:10) | 16:52 (15:51-18:49) |
| Start time ping 6, median (Q1-Q3) | 18:40 (17:48-19:57) | 18:29 (17:30-19:49) |
| Start time ping 7, median (Q1-Q3) | 19:59 (19:18-21:32) | 19:46 (19:00-20:50) |
| Start time ping 8, median (Q1-Q3) | 21:44 (21:21-22:07) | 20:54 (20:26-21:27) |
|  |  |  |
| Operating system | Android | Android, iOS |
| Retrospective entry | No | No |
| Time interval between signals, median (Q1-Q3) | 111.67 (82.28-136.80) minutes | 97.37 (65.33-128.30) minutes |
| Answer time, median (Q1-Q3) | 1.19 (0.58-2.29) minutes | 12.1 (1.45-46.9) minutes |

**eMethods 2. Psychometric properties of the experience sampling method questions**

For each construct, the source of the items is mentioned. When the construct consisted of multiple items, a between-subject Cronbach’s alpha and a within-subject Cronbach’s alpha was calculated to assess reliability. Whenever data was available on a questionnaire which measured the same construct, a Pearson correlation was calculated between the average value of the ESM items per participant and their score on the questionnaire to assess convergent validity.

Negative affect: The six items concerning negative affect were based on previous research of Collip et al. (2011), Lataster et al. (2013), and Rintala et al. (2020). The items had an excellent between-person reliability with a Cronbach’s alpha of 0.96, and an acceptable within-person reliability with a median Cronbach’s of 0.79. Furthermore, the items had an adequate convergent validity as there was a correlation of 0.60 between the average negative affect score of a participant and their score on the negative affect scale of the Leuven Affect Pleasure Scale (Demyttenaere et al., 2019).

Positive affect: The three items concerning positive affect were based on previous research of Collip et al. (2013). The items had an excellent between-person reliability with a Cronbach’s alpha of 0.95, and an acceptable within-person reliability with a median Cronbach’s of 0.79. Furthermore, the items had an adequate convergent validity as there was a correlation of 0.58 between the average positive affect score of a participant and their score on the positive affect scale of the Leuven Affect Pleasure Scale (Demyttenaere et al., 2019).

Rash action: The five items concerning rash action were based on previous research of Sperry et al. (2018). The items had an excellent between-person reliability with a Cronbach’s alpha of 0.95, and an acceptable within-person reliability with a median Cronbach’s of 0.72. Convergent validity could not be assessed in the current study as there was no data available on a relevant questionnaire measuring rash action or urgency. However, the measures displayed an adequate convergent validity in the study of Sperry et al. (2018) by showing a moderate association with the negative urgency subscale of the UPPS-P.

Lack of perseverance: The one item concerning lack of perseverance was on previous research of Sperry et al. (2018). Convergent validity could not be assessed as there was no data available on a relevant questionnaire measuring lack of perseverance. However, the measure displayed an adequate convergent validity in the study of Sperry et al. (2018) by showing a moderate association with the lack of perseverance subscale of the
UPPS-P.

Motivation: The one item concerning motivation (i.e., Right now, I feel interested) was based on the Positive and Negative Affect Scale (PANAS) (Watson et al., 1988). Convergent validity could not be assessed as there was no data available on a relevant questionnaire measuring motivation. However, the measure was still included a lack of interest/boredom has been associated with emotional eating and alcohol use (Koball et al., 2012; Orcutt, 1984).

Craving for a binge eating episode: The item concerning craving for a binge eating episode was based on previous research of Wonderlich et al. (2017). However, the question was adapted from assessing craving for food to craving for a binge eating episode as previous research has shown that patients can have a specific craving for a binge eating episode (Ferriday & Brunstrom, 2011; Gluck et al., 2004). There was an adequate convergent validity as there was a correlation of 0.53 between the average craving score per participant in daily life and their score on the FCQ-T (Vander Wal et al., 2007).

Craving for alcohol: The item concerning craving for alcohol was created to match the item concerning craving for a binge eating episode. There was an adequate convergent validity as there was a correlation of 0.53 between the average craving score per participant in daily life and their score on the OCDC-C and a correlation of 0.52 between the average craving score per participant in daily life and their score on the OCDC-O (Anton, 2000).

Activities and activity-related stress: The item concerning activities and the three items concerning activity-related stress were based on previous research (Collip et al., 2011; Gevonden et al., 2016; Glaser et al., 2006; Kasanova et al., 2018; Myin-Germeys, Van Os, et al., 2001). The items had a good between-person reliability with a Cronbach’s alpha of 0.83, and a good within-person reliability with a median Cronbach’s of 0.81. Convergent validity could not be assessed in the current study, but these measures are associated with changes in daily life negative affect, suggesting that they do possess predictive validity (Myin-Germeys et al., 2001).

Location: The item concerning location was based on previous research from (Glaser et al., 2006). It has been used to study the relation between social context and psychotic experiences (Akcaoglu et al., 2024; Myin-Germeys, Nicolson, et al., 2001). Though no information on its validity is available, it includes the most important locations where patients binge eat, drink alcohol, or binge drink (Crowther et al., 1984; Grüne et al., 2017). Furthermore, as to our knowledge, no studies have explored the validity of ESM measures concerning location.

Social context and social stress: The item concerning social context and the four items concerning social stress were based on previous research (Collip et al., 2011; Gevonden et al., 2016; Glaser et al., 2006; Kasanova et al., 2018; Myin-Germeys, Van Os, et al., 2001). The items had a good between-person reliability with a Cronbach’s alpha of 0.85, and an excellent within-person reliability with a median Cronbach’s of 0.91. Convergent validity could not be assessed in the current study, but these measures are associated with changes in daily life negative affect, suggesting that they do possess predictive validity (Myin-Germeys et al., 2001).

Most important event: The item concerning the most important event (i.e., The most important event since the last beep was) was based on previous research (Collip et al., 2011; Gevonden et al., 2016; Glaser et al., 2006; Kasanova et al., 2018; Myin-Germeys, Van Os, et al., 2001). Convergent validity could not be assessed in the current study, but the measure are associated with changes in daily life negative affect, suggesting that it does possess predictive validity (Myin-Germeys et al., 2001).

Negative events and their stressfulness: The item concerning negative events and the item on their combined stressfulness was based on previous research from Bastiaansen et al. (2018). Their validity in the current study could not be assessed, and no information on their validity has been published. However, items measure similar events which have been found to predict binge eating (interpersonal stress, work stress, daily hassles) and alcohol use (i.e., negative work, health, and nonwork events) (Carney et al., 2000; Goldschmidt et al., 2014).

Positive events and their pleasantness: The item concerning positive events and the item on their combined pleasantness was based on previous research from Bastiaansen et al. (2018). Their validity in the current study could not be assessed, but they are the opposite of the negative events mentioned previously.

Effort: The two items concerning effort have been used in the SMILE study (<https://gbiomed.kuleuven.be/english/research/50000666/50000673/cpp/research-1/stress-reward/smile>) to investigate the role of task engagement in stress-reactivity. Though no information on the validity of these items is available, they were included to account for a possible relation between task engagement and binge eating, alcohol use, and binge drinking.

Substances: The item concerning substance use came from an internal item repository, which has not been used in previously published research. However, it was included in the current study to inventorize when participants used substances (other than alcohol) which could impact their thoughts, emotions, and behaviors such as smoking, using narcotics or taking medication.

Eating: The three items concerning eating were based on previous research of Ambani et al. (2015). There was an adequate convergent validity as there was a correlation of 0.50 between the average number of binge eating episodes of a participant in daily life and their the number of binge eating episodes they reported with the EDE-Q (Fairburn & Beglin, 1994).

Alcohol: The first two items concerning alcohol use (i.e., Since the last beep, I have drunk alcohol; I drank …) were based on previous research of Gorka et al (2017). The last item (i.e., I felt that I lost control over drinking) was added to match the items concerning eating. There was an adequate convergent validity as there was a moderate Pearson correlation of 0.59 between the average number of alcohol use episodes of a participant in daily life and their score on the AUDIT (Saunders et al., 1993).

**eFigure 1. Data preprocessing**

The original ESM data, which consisted of the outcomes (binge eating, alcohol use, and binge drinking) and the ESM predictors, was expanded with temporal predictors as well as a COVID-19 stringency index based on the Oxford COVID-19 Government Response Tracker. Afterwards the final data set was created by keeping the outcomes, the temporal predictors, and the COVID-19 stringency index at the current timepoint, while the ESM data were lagged by one timepoint. This made it possible to predict the outcomes 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. Abbreviations: AU, alcohol use; BD, binge drinking; BE, binge eating; ESM, experience sampling method; T-1, previous timepoint; T0, current timepoint.

****

**eMethods 3. Additional predictors**

Temporal variables

The linear, quadratic, and cubic time since starting the study variables were included to account for long-term changes in behavior as it can be presumed that the frequency at which patients binge eat, drink alcohol, binge drink does not stay constant over longer periods of time. For example, a patient could be increasingly less likely to binge eat (e.g., due to treatment), and in this case, a negative linear term would be included in the prediction model. The 12 hr and 24 hr frequency variables were created by calculating sine and cosine terms with a 12 hr or 24 hr period based on the linear time since starting the study variable, which made it possible to capture cyclical changes in the outcome. In a clinical setting, the temporal variables would not be based on the time since the start of the study, but rather on the time since the start of the JITAI. More specifically, in a person-specific model, a patient would first gather data before interventions would be sent out, and the temporal predictors fit on that data could inform a JITAI on changes in binge eating, alcohol use, and binge drinking over time in that patient. Additionally, in a pooled model, the temporal predictors would inform a JITAI on changes in binge eating, alcohol use, and binge drinking over time across all patients.

Oxford COVID-19 Government Response Tracker

This stringency variable ranged from 0 (i.e., no restrictive measures) to 100 (i.e., the strictest measures) and was based on several metrics tracking a government’s response to COVID (e.g., the restriction of public gatherings or internal movements).

**eMethods 4. Elastic net wrappers**

The elastic net wrappers are two R functions designed to construct accurate person-specific and pooled prediction models using experience sampling method data. The process involves several steps. First, data preprocessing, where the splitTools package is utilized to divide the data into training and testing sets (Mayer, 2022). This division can be achieved through either a train-test split or a k-fold split, with options for stratified (accounting for outcome distribution) or non-stratified splitting. Additionally, data standardization can be applied at this stage. Second, model fit, where, the ensr package is employed to determine the optimal alpha and lambda values for elastic net regularized regression (Peter DeWitt, 2019). This selection is done using k-fold cross-validation on the training data. Following this, the glmnet package is used to fit the elastic net regularized regression model (Friedman et al., 2010). Third, model evaluation, where the pROC package is used to compute key performance metrics, including sensitivity, specificity, and the area under the curve for the model on the test set (Robin et al., 2011). The caret package is also utilized to calculate additional metrics such as accuracy, negative predictive value, and positive predictive value (Kuhn, 2021).

**eMethods 5. Nested k-fold cross-validation**

A k-fold cross-validation splits an entire dataset into a k number of folds. Then, a model is trained on all but one of the folds and tested on the remaining one. This is repeated a k number of times so that each fold is used as a test fold once. The performances on the test folds are then averaged to calculate a mean performance for the entire dataset.

K-fold cross-validation typically results in a reliable estimate of a model’s performance. Nevertheless, problems can arise when a machine learning model includes hyperparameters. These parameters (e.g., alpha and lambda for elastic net) can be tuned to improve performance. However, there are no rules dictating which values they should take for a specific dataset. Therefore, they are often tuned to a dataset using cross-validation. For example, a grid search can be done to test the performance of different models with different sets of hyperparameters. Then, the model with the best performance can be selected.

However, using the same data to tune hyperparameters and to evaluate model performance will result in overfitting. Therefore, the cross-validation to tune hyper-parameters needs to be nested in the cross-validation to evaluate model performance.

First, there needs to be an outer loop of cross-validation which splits the data into a number of folds. Second, on the training folds of the outer loop, there needs to be an inner loop of cross-validation which tunes hyperparameters and fits the model. Third, the performance of this model is needs to be evaluated on the test fold of the outer loop.

|  |
| --- |
| **eTable 2. Sample characteristics per analysis group** |
|  | BE (n=69) | BD/alcohol use (n=70) |
|  | Mean (SD) | 95% CI | Mean (SD) | 95% CI |
| Age | 21.9 (3.7) | 21.0-22.8 | 21.2 (3.2) | 20.4-21.9 |
| BMI | 25.2 (5.3) | 24.0-26.5 | 23.2 (2.4) | 22.6-23.7 |
| Illness Duration BN (years) | 2.5 (1.5) | 2.2-2.9 | 0.8 (1.5) | 0.5-1.2 |
| Illness Duration AUD (years) | 0.6 (1.2) | 0.3-0.9 | 2.9 (1.4) | 2.5-3.2 |
| Education (years) | 14.5 (2.1) | 13.9-15.0 | 14.5 (1.7) | 14.1-14.9 |
| AUDIT | 7.8 (6.9) | 6.2-9.5 | 15.8 (5.1) | 14.4-16.8 |
| EDE-Q Restraint Shape Concern Weight Concern Eating Concern Total | 2.8 (1.5)4.2 (1.4)4.1 (1.6)2.8 (1.4)3.6 (1.3) | 2.5-3.23.9-4.63.7-4.52.5-3.23.3-3.9 | 1.3 (1.4)2.4 (1.7)2.2 (1.9)1.1 (1.4)1.8 (1.3) | 0.9-1.62.0-2.81.7-2.60.8-1.51.5-2.2 |
| Eating disorder symptoms (days/4 weeks) Binge eating Fasting Vomiting Laxative use Diuretic use Compensatory exercise | 8.1 (6.8)7.9 (8.1)2.1 (5.5)0.3 (1.9)0.8 (4.5)6.5 (7.3) | 6.5-9.85.9-9.70.8-3.50-0.80-1.84.8-8.3 | 1.8 (4.0)2.1 (5.1)0.5 (2.1)0 (0.2)0 (0)2.0 (4.8) | 0.9-2.80.8-3.30-1.00-0.10-00.8-3.1 |
|  | N (%) | 95% CI | N (%) | 95% CI |
| Binge drinking frequency Never Annually Semiannually Three-monthly Monthly Biweekly Weekly >Weekly | 27 (39%)8 (12%)2 (3%)7 (10%)6 (9%)12 (17%)6 (9%)1 (1%) | 29-52%1-24%0-15%0-23%0-21%7-30%0-21%0-14% | 0-0%0 (0%)0 (0%)6 (9%)12 (17%)30 (43%)14(20%)8 (11%) | 0-0%0-0%0-0%0-20%6-29%31-55%9-32%0-23% |
| Therapy (BN or AUD) | 14 (20%) | 11-30% | 12 (17%) | 8-26% |
| Race and ethnicity Caucasian Latina Asian  Multi-racial Middle-Eastern Black | 62 (90%)0 (0%)3 (4%)0 (0%)4 (6%)0 (0%) | 84-96%0-0%0-11%0-0%0-12%0-0% | 67 (96%)1 (1%)0 (0%)2 (3%)0 (0%)0 (0% | 93-100%0-6%0-0%0-8%0-0%0-0% |
| Psychoactive medication | 11 (16%) | 7-25% | 9 (13%) | 5-21% |
| Comorbidities MDD PD SAD ADHD PTSD AP | 9 (18%)5 (10%)6 (12%)0 (0%)9 (18%)4 (8%) | 6-33%0-25%0-27%0-0%6-33%0-23% | 7 (10%)3 (4%)2 (3%)3 (4%)0 (0%)0 (0%) | 3-20%0-14%0-13%0-14%0-0%0-0% |
| Abbreviations: ADHD, attention deficit hyperactivity disorder; AP, agoraphobia; BD, binge drinking; BE, binge eating; BMI, body mass index; 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. |

**eResults 1. Reasons for dropout**

Out of the 41 participants that dropped out of the ESM protocol, 4(9.58%) did so because of app incompatibility, 11 (26.8%) due to the intensity of the protocol and 26 (63.4%) because of unknown reasons. For every participant group (AUD, BN, AUD/BM), there was no significant difference in age, BMI, illness duration, AUDIT scores, or EDE-Q scores between patients who dropped out and those who didn’t. Importantly, it can be seen in eTable3 that though drop-out was relatively similar across the different participant groups, the patients with AUD/BN 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.

|  |
| --- |
| **eTable 3. Median compliance and drop-out percentage per burst and per participant group** |
|  | AUD (n=51) | AUD/BN (n=19) | BN (n=50) |
| Burst | Compliance (%) | Drop-out (%) | Compliance (%) | Drop-out (%) | Compliance (%) | Drop-out (%) |
| 1 | 81.9 | 0 | 77.8 | 0 | 86.8 | 0 |
| 2 | 77.7 | 9.8 | 48.6 | 5.3 | 77.8 | 0.2 |
| 3 | 68.1 | 15.7 | 44.4 | 10.5 | 69.4 | 0.8 |
| 4 | 62.5 | 23.5 | 34.7 | 21.1 | 68.1 | 16.0 |
| 5 | 56.9 | 25.5 | 19.4 | 26.3 | 55.6 | 20.0 |
| 6 | 61.8 | 29.4 | 43.1 | 26.3 | 56.9 | 20.0 |
| 7 | 67.4 | 37.3 | 36.1 | 31.6 | 54.2 | 32.0 |

**eTable 4. Confusion matrix of the person-specific models for binge eating**

|  |  |
| --- | --- |
|  | Original |
| Not binge eating | Binge eating |
| Prediction | Not binge eating | 4856 | 385 |
| Binge eating | 2495 | 945 |

**eTable 5. Confusion matrix of the pooled models for binge eating**

|  |  |
| --- | --- |
|  | Original |
| Not binge eating | Binge eating |
| Prediction | Not binge eating | 5734 | 347 |
| Binge eating | 2376 | 1033 |

**eTable 6. Confusion matrix of the person-specific models for alcohol use**

|  |  |
| --- | --- |
|  | Original |
| No alcohol use | Alcohol use |
| Prediction | No alcohol use | 5508 | 166 |
| Alcohol use | 1526 | 855 |

**eTable 7. Confusion matrix of the pooled models for alcohol use**

|  |  |
| --- | --- |
|  | Original |
| No alcohol use | Alcohol use |
| Prediction | No alcohol use | 6955 | 132 |
| Alcohol use | 1274 | 963 |

**eTable 8. Confusion matrix of the person-specific models for binge drinking**

|  |  |
| --- | --- |
|  | Original |
| Not binge drinking | Binge drinking |
| Prediction | Not binge drinking | 2251 | 245 |
| Binge drinking | 572 | 231 |

**eTable 9. Confusion matrix of the pooled models for binge drinking**

|  |  |
| --- | --- |
|  | Original |
| Not binge drinking | Binge drinking |
| Prediction | Not binge drinking | 5067 | 17 |
| Binge drinking | 769 | 236 |

**eResults 2. Relation between number of assessments and area under the curve for the person-specific models.**

For binge eating, the median number of assessments per participant for the person-specific models was 163 (range: 31-459). There was no significant correlation between the number of assessments and the area under the curve (rspearman=0.04, p=0.780). For alcohol use, the median number of assessments per participant for the person-specific models was 169 (range: 34-419). There was no significant correlation between the number of assessments and the area under the curve (rspearman=0.18, p=0.240). For binge drinking, the median number of assessments per participant for the person-specific models was 169 (range: 34-419). There was no significant correlation between the number of assessments and the area under the curve (rspearman=0.01, p=0.970).

**eFigure 2. Relation between number of assessments and area under the curve for the person-specific models**

Relationship between the area under the curve in person-specific models for binge eating, alcohol use, and binge drinking and the number of assessments completed by participants. The left side of the figure presents bar plots depicting the average area under the curve values for different brackets of the number of assessments completed by participants. The right side displays smoothed loess curves showing the average area under the curve values for a continuous measure of the number of assessments completed by participants.



**eResults 3. Comparing model performance with a model which always predicts the majority case**

The results of the pooled and person-specific models of the current study were compared to those of models that always predicted the majority class (i.e., not BE, not drinking alcohol, not BD). From the results in eTable 4, it can be seen that the pooled and person-specific models of the current study had a higher negative predictive value, even though the general accuracy was lower. However, comparing the AUC, sensitivity, and PPV is diificult as always predicting the majority class (i.e., the absence of an event) leads to an AUC of 0.5, a sensitivity of 0, and a non-existent PPV.

|  |
| --- |
| **eTable 10. Comparing model performance with a model which always predicts the majority case** |
| Outcome | Type | Original/Majority | AUC | Sensitivity | Specificity | Accuracy | PPV | NPV |
| Binge eating | Pooled | Original | 0.71 | 1.00 | 0.75 | 0.74 | 0.33 | 1.00 |
| Majority | 0.5 | 0 | 1 | 0.91 | NA | 0.91 |
| Person-specific | Original | 0.61 | 0.83 | 0.71 | 0.71 | 0.31 | 0.97 |
| Majority | 0.5 | 0 | 1 | 0.85 | NA | 0.85 |
| Alcohol use | Pooled | Original | 0.90 | 1.00 | 0.88 | 0.87 | 0.50 | 1.00 |
| Majority | 0.5 | 0 | 1 | 0.91 | NA | 0.91 |
| Person-specific | Original | 0.80 | 1.00 | 0.80 | 0.80 | 0.38 | 1.00 |
| Majority | 0.5 | 0 | 1 | 0.88 | NA | 0.88 |
| Binge drinking | Pooled | Original | 0.93 | 1.00 | 0.93 | 0.93 | 0.50 | 1.00 |
| Majority | 0.5 | 0 | 1 | 0.91 | NA | 0.91 |
| Person-specific | Original | 0.85 | 1.00 | 0.90 | 0.86 | 0.28 | 1.00 |
| Majority | 0.5 | 0 | 1 | 0.85 | NA | 0.85 |

**eResults 4. Correcting for an imbalance in the outcomes**

To investigate whether correcting for an imbalance in the outcomes would improve model performance, adaptations of the original wrappers were created whereby ROSE or SMOTE was integrated in the workflow. These wrappers can be found at <https://github.com/mikojeske/NLML/tree/main/Elastic%20Net/Unbalanced%20Data>. In these adapted wrappers, the data is split into a train and test set, after which the balance in the outcome is corrected using either ROSE or SMOTE, and afterwards the elastic net model is fit and evaluated. From the results in eTable 4, it can be seen that a correction typically resulted in a worse model performance, except for the person-specific models for BE.

|  |
| --- |
| **eTable 11. Correcting for an imbalance in the outcomes** |
| Outcome | Type | Original | ROSE | SMOTE |
| Binge eating | Pooled | 0.71 | 0.70 | 0.70 |
| Person-specific | 0.61 | 0.64 | 0.64 |
| Alcohol use | Pooled | 0.90 | 0.89 | 0.89 |
| Person-specific | 0.80 | 0.80 | 0.77 |
| Binge drinking | Pooled | 0.93 | 0.87 | 0.92 |
| Person-specific | 0.85 | 0.80 | 0.68 |

**eResults 5. The impact of app type on model performance**

To investigate the impact of app type on model performance, several analyses were performed. First, person-specific and pooled models were fit and evaluated with an additional app type predictor which was coded 0 when the observation was recorded through MobileQ, and coded 1 when it was recorded through m-Path. The results from this analysis can be seen in eTable 12. Second, the percentage of responses of a participant recorded through m-Path was correlated the AUC of the original person-specific and pooled model. These were Spearman correlations due to the non-normal distribution of the AUCs. Here, a higher percentage of responses through m-Path was correlated with the AUC of the person-specific models for alcohol use (ρ=-0.424, p=0.002), but not in the other models.

|  |
| --- |
| **eTable 12. The impact of app type on model performance** |
| Outcome | Type | Original/Majority | AUC | Sensitivity | Specificity | Accuracy | PPV | NPV |
| Binge eating | Pooled | Original | 0.71 | 1.00 | 0.75 | 0.74 | 0.33 | 1.00 |
| App type | 0.69 | 1.00 | 0.75 | 0.74 | 0.33 | 1.00 |
| Person-specific | Original | 0.61 | 0.83 | 0.71 | 0.71 | 0.31 | 0.97 |
| App type | 0.61 | 0.83 | 0.69 | 0.71 | 0.31 | 0.97 |
| Alcohol use | Pooled | Original | 0.90 | 1.00 | 0.88 | 0.87 | 0.50 | 1.00 |
| App type | 0.90 | 1.00 | 0.90 | 0.88 | 0.50 | 1.00 |
| Person-specific | Original | 0.80 | 1.00 | 0.80 | 0.80 | 0.38 | 1.00 |
| App type | 0.80 | 1.00 | 0.80 | 0.80 | 0.38 | 1.00 |
| Binge drinking | Pooled | Original | 0.93 | 1.00 | 0.93 | 0.93 | 0.50 | 1.00 |
| App type | 0.94 | 1.00 | 0.94 | 0.94 | 0.50 | 1.00 |
| Person-specific | Original | 0.85 | 1.00 | 0.90 | 0.86 | 0.28 | 1.00 |
| App type | 0.85 | 1.00 | 0.90 | 0.86 | 0.29 | 1.00 |

**eReferences**

Akcaoglu, Z., Vaessen, T., Velthorst, E., Lafit, G., Achterhof, R., Nelson, B., McGorry, P., Schirmbeck, F., Morgan, C., Hartmann, J., Van Der Gaag, M., De Haan, L., Valmaggia, L., McGuire, P., Kempton, M., Steinhart, H., Klippel, A., Viechtbauer, W., Batink, T., … Myin-Germeys, I. (2024). The temporal association between social isolation, distress, and psychotic experiences in individuals at clinical high-risk for psychosis. Psychological Medicine. https://doi.org/10.1017/S0033291723003598

Ambwani, S., Roche, M. J., Minnick, A. M., & Pincus, A. L. (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

Anton, R. F. (2000). Obsessive-compulsive aspects of craving: development of the Obsessive Compulsive Drinking Scale. Addiction (Abingdon, England), 95 Suppl 2(8). https://doi.org/10.1080/09652140050111771

Bastiaansen, J. A., Meurs, M., Stelwagen, R., Wunderink, L., Schoevers, R. A., Wichers, M., & Oldehinkel, A. J. (2018). Self-monitoring and personalized feedback based on the experiencing sampling method as a tool to boost depression treatment: a protocol of a pragmatic randomized controlled trial (ZELF-i). BMC Psychiatry, 18(1). https://doi.org/10.1186/S12888-018-1847-Z

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

Carney, M. A., Armeli, S., Tennen, H., Affleck, G., & O’Neil, T. P. (2000). Positive and negative daily events, perceived stress, and alcohol use: A diary study. Journal of Consulting and Clinical Psychology, 68(5), 788–798. https://doi.org/10.1037/0022-006X.68.5.788

Collip, D., Nicolson, N. A., Lardinois, M., Lataster, T., Van Os, J., & Myin-Germeys, I. (2011). Daily cortisol, stress reactivity and psychotic experiences in individuals at above average genetic risk for psychosis. Psychological Medicine, 41(11), 2305–2315. https://doi.org/10.1017/S0033291711000602

Crowther, J. H., Lingswiler, V. M., & Stephens, M. A. P. (1984). The topography of binge eating. Addictive Behaviors, 9(3), 299–303. https://doi.org/10.1016/0306-4603(84)90024-8

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

Demyttenaere, K., Mortier, P., Kiekens, G., & Bruffaerts, R. (2019). Is there enough “interest in and pleasure in” the concept of depression? The development of the Leuven Affect and Pleasure Scale (LAPS). CNS Spectrums, 24(2), 265–274. https://doi.org/10.1017/S1092852917000578

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

Ferriday, D., & Brunstrom, J. M. (2011). “I just can’t help myself”: effects of food-cue exposure in overweight and lean individuals. International Journal of Obesity (2005), 35(1), 142–149. https://doi.org/10.1038/IJO.2010.117

Gevonden, M., Myin-Germeys, I., Wichers, M., Booij, J., van den Brink, W., van Winkel, R., & Selten, J. P. (2016). Reactivity to social stress in ethnic minority men. Psychiatry Research, 246, 629–636. https://doi.org/10.1016/J.PSYCHRES.2016.10.076

Glaser, J. P., van Os, J., Portegijs, P. J. M., & Myin-Germeys, I. (2006). Childhood trauma and emotional reactivity to daily life stress in adult frequent attenders of general practitioners. Journal of Psychosomatic Research, 61(2), 229–236. https://doi.org/10.1016/J.JPSYCHORES.2006.04.014

Gluck, M. E., Geliebter, A., Hung, J., & Yahav, E. (2004). Cortisol, hunger, and desire to binge eat following a cold stress test in obese women with binge eating disorder. Psychosomatic Medicine, 66(6), 876–881. https://doi.org/10.1097/01.PSY.0000143637.63508.47

Goldschmidt, A. B., Wonderlich, S. A., Crosby, R. D., Engel, S. G., Lavender, J. M., Peterson, C. B., Crow, S. J., Cao, L., & 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

Gorka, S. M., Hedeker, D., Piasecki, T. M., & Mermelstein, R. (2017). Impact of alcohol use motives and internalizing symptoms on mood changes in response to drinking: An ecological momentary assessment investigation. Drug and Alcohol Dependence, 173, 31–38. https://doi.org/10.1016/J.DRUGALCDEP.2016.12.012

Grüne, B., Piontek, D., Sleczka, P., Kraus, L., & Pogarell, O. (2017). Drinking Location and Drinking Culture and Their Association With Alcohol Use Among Girls and Boys in Europe. Https://Doi.Org/10.15288/Jsad.2017.78.549, 78(4), 549–557. https://doi.org/10.15288/JSAD.2017.78.549

Kasanova, Z., Ceccarini, J., Frank, M. J., van Amelsvoort, T., Booij, J., Heinzel, A., Mottaghy, F. M., & Myin-Germeys, I. (2018). Daily-life stress differentially impacts ventral striatal dopaminergic modulation of reward processing in first-degree relatives of individuals with psychosis. European Neuropsychopharmacology : The Journal of the European College of Neuropsychopharmacology, 28(12), 1314–1324. https://doi.org/10.1016/J.EURONEURO.2018.10.002

Koball, A. M., Meers, M. R., Storfer-Isser, A., Domoff, S. E., & Musher-Eizenman, D. R. (2012). Eating when bored: Revision of the emotional eating scale with a focus on boredom. Health Psychology, 31(4), 521–524. https://doi.org/10.1037/A0025893

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

Lavender, J. M., Utzinger, L. M., Crosby, R. D., Goldschmidt, A. B., Ellison, J., Wonderlich, S. A., Engel, S. G., Mitchell, J. E., Crow, S. J., Peterson, C. B., & 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

Lataster, T., Valmaggia, L., Lardinois, M., Van Os, J., & Myin-Germeys, I. (2013). Increased stress reactivity: a mechanism specifically associated with the positive symptoms of psychotic disorder. Psychological Medicine, 43(7), 1389–1400. https://doi.org/10.1017/S0033291712002279

Myin-Germeys, I., Nicolson, N. A., & Delespaul, P. A. E. G. (2001). The context of delusional experiences in the daily life of patients with schizophrenia. Psychological Medicine, 31(3), 489–498. https://doi.org/10.1017/S0033291701003646

Myin-Germeys, I., Van Os, J., Schwartz, J. E., Stone, A. A., & Delespaul, P. A. (2001). Emotional Reactivity to Daily Life Stress in Psychosis. Archives of General Psychiatry, 58(12), 1137–1144. https://doi.org/10.1001/ARCHPSYC.58.12.1137

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

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2020). Momentary predictors of compliance in studies using the experience sampling method. Psychiatry Research, 286. https://doi.org/10.1016/J.PSYCHRES.2020.112896

Orcutt, J. D. (1984). Contrasting Effects of Two Kinds of Boredom on Alcohol Use. Http://Dx.Doi.Org/10.1177/002204268401400112, 14(1), 161–173. https://doi.org/10.1177/002204268401400112

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

Sperry, S. H., Lynam, D. R., & Kwapil, T. R. (2018). The convergence and divergence of impulsivity facets in daily life. Journal of Personality, 86(5), 841–852. https://doi.org/10.1111/JOPY.12359

Vaessen, T., van Nierop, M., Decoster, J., Delespaul, P., Derom, C., de Hert, M., Jacobs, N., Menne-Lothmann, C., Rutten, B., Thiery, E., van Os, J., van Winkel, R., Wichers, M., & Myin-Germeys, I. (2017). Is sensitivity to daily stress predictive of onset or persistence of psychopathology? European Psychiatry, 45, 167–173. https://doi.org/10.1016/J.EURPSY.2017.07.002

Vander Wal, J. S., Johnston, K. A., & Dhurandhar, N. V. (2007). Psychometric properties of the State and Trait Food Cravings Questionnaires among  overweight and obese persons. Eating Behaviors, 8(2), 211–223. https://doi.org/10.1016/j.eatbeh.2006.06.002

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. Journal of Personality and Social Psychology, 54(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wonderlich, J. A., Breithaupt, L. E., Crosby, R. D., Thompson, J. C., Engel, S. G., & Fischer, S. (2017). The relation between craving and binge eating: Integrating neuroimaging and ecological momentary assessment. Appetite, 117, 294–302. https://doi.org/10.1016/J.APPET.2017.07.005