173 (2025) 209685

Contents lists available at ScienceDirect

Journal of Substance Use and Addiction Treatment

journal homepage: www.journals.elsevier.com/journal-of-substance-use-and-addiction-treatment

A longitudinal observational study with ecological momentary assessment and deep learning to predict non-prescribed opioid use, treatment retention, and medication nonadherence among persons receiving medication treatment for opioid use disorder

Michael V. Heinz [a,b,*], George D. Price [c], Avijit Singh [c], Sukanya Bhattacharya [c], Ching-Hua Chen [e], Asma Asyyed [g], Monique B. Does [f], Saeed Hassanpour [a,d], Emily Hichborn [a], David Kotz [a,h], Chantal A. Lambert-Harris [a], Zhiguo Li [e], Bethany McLeman [a], Varun Mishra [i,j], Catherine Stanger [a], Geetha Subramaniam [k], Weiyi Wu [a,d], Cynthia I. Campbell [f,l,m], Lisa A. Marsch [a], Nicholas C. Jacobson [a,d]

[a] Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States
[b] Department of Psychiatry, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States
[c] Quantitative Biomedical Sciences Program, Dartmouth College, Hanover, NH, United States
[d] Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States
[e] Center for Computational Health, International Business Machines (IBM) Research, Yorktown Heights, NY, United States
[f] Division of Research, Kaiser Permanente Northern California, Oakland, CA, United States
[g] The Permanente Medical Group, Northern California, Addiction Medicine and Recovery Services, Oakland, CA, United States
[h] Department of Computer Science, Dartmouth College, Hanover, NH, United States
[i] Khoury College of Computer Sciences, Northeastern University, Boston, MA, United States
[j] Department of Health Sciences, Bouvé College of Health Sciences, Northeastern University, Boston, MA, United States
[k] Center for Clinical Trials Network, National Institute on Drug Abuse, Bethesda, MD, United States
[l] Department of Psychiatry and Behavioral Sciences, University of California San Francisco, San Francisco, CA, United States
[m] Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, United States

ARTICLE INFO

Keywords:
Ecological momentary assessment
EMA
Deep learning
Opioid treatment
Opioid relapse
Opioid addiction
Opioid use disorder
Relapse prediction
Dense longitudinal time series

ABSTRACT

Background: Despite effective treatments for opioid use disorder (OUD), relapse and treatment drop-out diminish their efficacy, increasing the risks of adverse outcomes, including death. Predicting important outcomes, including non-prescribed opioid use (NPOU) and treatment discontinuation among persons receiving medications for OUD (MOUD) can provide a proactive approach to these challenges. Our study uses ecological momentary assessment (EMA) and deep learning to predict momentary NPOU, medication nonadherence, and treatment retention in MOUD patients.

Methods: Study participants included adults receiving MOUD at a large outpatient treatment program. We predicted NPOU (EMA-based), medication nonadherence (Electronic Health Record [EHR]- and EMA-based), and treatment retention (EHR-based) using context-sensitive EMAs (e.g., stress, pain, social setting). We used recurrent deep learning models with 7-day sliding windows to predict the next-day outcomes, using Area Under the ROC Curve (AUC) for assessment. We employed SHapley additive ExPlanations (SHAP) to understand feature latency and importance.

Results: Participants comprised 62 adults with 14,322 observations. Model performance varied across EMA subtypes and outcomes with AUCs spanning 0.58–0.97. Recent substance use was the best performing predictor for EMA-based NPOU (AUC = 0.97). Life-contextual factors were best performers for EMA-based medication nonadherence (AUC = 0.68) and retention (AUC = 0.89), and substance use risk factors (e.g., nicotine and alcohol use) and self-reported MOUD adherence performed best for predicting EHR-based medication nonadherence (AUC = 0.79). SHAP revealed varying latencies between predictors and outcomes.

* Corresponding author at: Dartmouth College, 46 Centerra Parkway Suite 315, Lebanon, NH 03766, United States.
E-mail address: michael.v.heinz@dartmouth.edu (M.V. Heinz).

https://doi.org/10.1016/j.josat.2025.209685
Received 15 May 2024; Received in revised form 6 January 2025; Accepted 18 March 2025
Available online 22 March 2025

Conclusions: Findings support the effectiveness of EMA and deep learning for forecasting actionable outcomes in persons receiving MOUD. These insights will enable the development of personalized dynamic risk profiles and just-in-time adaptive interventions (JITAIs) to mitigate high-risk OUD outcomes.

1. Introduction

The United States (US) has seen an exponential increase in drug overdose deaths since the 1990s, with opioids involved in over 80,000 overdose deaths in 2022, a 19 % increase from 2020 (CDC WONDER). The cause of this increase is likely multifactorial, driven in part by the availability of higher potency synthetic opioids (Mattson et al., 2021). Opioid use disorder (OUD), characterized by frequent, non-prescription use of opioids impacting daily functioning and leading to symptoms such as cravings, withdrawal symptoms, and failure to fulfill role obligations (American Psychiatric Association, 2013), affects nearly 7 million adults in the US (Keyes et al., 2022); further, OUD accounts for over 1 billion in US spending annually (Luo, 2021). Apart from the societal economic burden, OUD causes tremendous individual suffering, associated with elevated rates of co-occurring mental, physical, and psychosocial problems, such as anxiety, depression, serious infections, loss of employment, and termination of parental rights (American Psychiatric Association, 2013; Howard & Guastaferro, 2019; White et al., 2020). Further, persons with OUD have a 20 times higher risk of death than those without (National Academies of Sciences, Engineering, and Medicine, et al., 2019).

Importantly, effective and lifesaving medications for OUD (MOUD) exist, including FDA-approved products made with buprenorphine, methadone, and naltrexone. MOUD has robust evidence supporting its benefit, including reduced deaths by overdose (Ma et al., 2019), improved treatment retention, reduced risk of relapse, reduced risk of serious infection (MacArthur et al., 2012), and improvements in quality of life (Ponizovsky & Grinshpoon, 2007). Indeed, persons who start an MOUD have an 80 % reduction in all cause and overdose crude mortality rates (Ma et al., 2019).

For most, successfully maintaining recovery requires long-term treatment with MOUD (Kraus et al., 2011; Weinstein et al., 2017), with discontinuation increasing the risk of adverse outcomes, including relapse and death (Ma et al., 2019). A cascade of care framework for OUD, spanning detection in the early risk stage to maintenance and recovery, shows significant drop-off between stages (e.g., treatment retention to sustained recovery) (Yedinak et al., 2019). Unfortunately, the natural course of OUD is unstable and often characterized by low treatment retention and repeated relapse and remission (Fishman et al., 2020). The causes of relapse and suboptimal retention in treatment are manifold, driven by complex psychosocial, physiologic, and environmental factors, with relapse itself being a risk factor for treatment discontinuation (Weinstein et al., 2017) in OUD. Other known substance-use risk factors include patient contextual factors, such as environmental cues and stressful life events, negative mood states, and withdrawal symptoms (Chalana et al., 2016; Chang & Raynor, 2021; Preston et al., 2017). It is of considerable public health importance to understand factors which contribute to relapse and treatment discontinuation.

While research to date has investigated risk factors for MOUD discontinuation and relapse (Chang & Raynor, 2021; Heiwe et al., 2011; Naji et al., 2016; Samples et al., 2018), many studies have focused on predominantly static, group-level factors, often assessed retrospectively. For example, Naji et al. (2016) found that the age onset of OUD and benzodiazepine use predicted relapse among patients receiving methadone; Samples et al. (2018) found that male sex, younger age, and minority race/ethnicity predicted discontinuation of buprenorphine treatment. While these findings are important in stratifying individual risk according to static, historical factors, they do not account for fluctuating risk states or account for the dynamic nature of addiction, which may be characterized by rapid temporal changes and influenced by contextual factors, mood states, and physiologic changes. Further, understanding dynamic risk factors, which predict relapse or treatment discontinuation is a prerequisite for just-in-time adaptive interventions (JITAIs), which aim to deliver the right intervention at the right time, proximal to points of receptivity (Mishra et al., 2021; Perski et al., 2022).

Ecological momentary assessment (EMA), the repeated sampling of a person's experiences and behavior in their natural environment (Shiffman et al., 2008) provides a promising approach to explore dynamic risk factors for relapse, treatment retention, and medication nonadherence among persons with OUD. Indeed, many randomized controlled trials and observational studies to date have demonstrated the feasibility and acceptability of frequent EMA among persons receiving MOUD (Heinz et al., 2024; Alexander et al., 2023) and have begun to disentangle the complex dynamic relationships between opioid use outcomes, such as non-prescribed opioid use, and risk factors, such as cravings (Huhn et al., 2016; Mun et al., 2021), stress (Bertz et al., 2022), mood and anxiety (Fatseas et al., 2018), and pain (Mun et al., 2021). To our knowledge, most EMA studies to date have focused on only a subset of potential opioid risk factors and have been conducted outside of a predictive framework, important for developing personalized risk models.

EMA combined with *deep learning methods*, which have the capacity to model highly dimensional time series (Petneházi, 2019), such as that generated by multi-item EMA, have the potential to identify personalized markers that predict important outcomes in persons with OUD. In the present analysis, we use comprehensive app-delivered EMA prompts regarding contextual, psychological, and physiologic factors paired with deep recurrent neural networks (RNNs) to predict four outcomes: (1) self-reported non-prescribed opioid use (NPOU), (2) self-reported medication nonadherence, (3) objective medication nonadherence, and (4) retention in OUD treatment, the latter two outcomes objectively measured by data from the participant's Electronic Health Record (EHR). For interpretability, we trained a model for each outcome, for each of the twelve EMA subtypes (e.g., Withdrawal Symptoms, Last Hour Substance Use, Mood Symptoms). We hypothesized that EMA would predict outcomes with moderate performance (AUC $>$ 0.70). Lastly, we used the SHapley Additive exPlanations (SHAP) method to explore the most influential input features across the 48 machine-learning models.

2. Methods

2.1. Participants

Participants (N = 62) were drawn from the National Institute on Drug Abuse (NIDA) Clinical Trials Network's "Digital Health to Understand Clinical Trajectories in Medication Treatment for Opioid Use Disorder (DTECT)" study (Marsch et al., 2022), approved with CTN-0084-A2 DTECT Study protocol; primary outcomes have been published previously (Campbell et al., 2023), including baseline demographics, which are summarized in Table 1. The DTECT study recruited adults who were enrolled in Kaiser Permanente Northern California (KPNC) Addiction Medication Recovery Services (AMRS) and undergoing outpatient treatment with buprenorphine MOUD for at least

Table 1
Aggregate summary of baseline participant characteristics. N: Number of participants; + due to small cell sizes, we do not report gender identification other than male and female.

	N	%
All participants	62	100.00 %
Gender identity +		
Female	29	46.77 %
Male	31	50.00 %
Age (years), mean (SD)	37.2 (13.3)	
Ethnicity		
Hispanic, Latino or of Spanish Origin	13	20.97 %
Not Hispanic, Latino or of Spanish Origin	49	79.03 %
Race		
White	44	70.97 %
Non-white/More than one race	18	29.03 %

two weeks in the date range of June 2020 to January 2021. Participants provided informed consent to participate in a 12-week study during which time they completed EMA prompts and provided permission for researchers to access EHR data regarding MOUD dispensation.[1]

2.2. Ecological momentary assessment

Participants were prompted with EMAs related to mental health, psychological state, and substance use three times daily. All EMA subtypes were conducted three times daily, unless otherwise specified.[2] EMA subtypes included those relevant to sleep, MOUD adherence, stress, pain, substance cravings, substance withdrawal symptoms, substance use risk context, mood, locale and social context, substances used in the last hour, substances used over the last day, and momentary self-regulation.[3] For a full list of EMA subtypes with respective abbreviations, to be used hereafter in the text, please refer to Supplemental Fig. 1. For a full list of EMA subtypes and their corresponding questions, please refer to Supplemental Table 1. All EMA responses were scaled to be between 0 and 1, and dummy coding was used where necessary to create binary variables.

2.3. Outcomes

We analyzed four outcomes: NPOU, nonadherence to MOUD (measured by EMA), nonadherence to MOUD (measured by EHR), and treatment retention. NPOU was derived from the EMA: "Substance Use Over the Last Hour," where participants were asked three times daily whether they had used any of nine substances over the last hour. If the substance was an opioid, including street methadone or street buprenorphine, fentanyl, or heroin, a positive flag was set for NPOU for that EMA instance. EMA-based nonadherence to MOUD was derived from the EMA item, MED-ADHERE (reference Supplemental Table 1). EHR-based nonadherence to MOUD was derived from EHR data[4] about medication dispensations. If a participant had no remaining medication on a particular day, based on previous dispensations and the expected daily consumption, they were flagged as non-adherent. If they still had medication left, they were labeled as adherent. Finally, regarding treatment retention, participants were considered to have been retained in treatment for 84 days if they had no long gap(s) in their MOUD coverage during the study. A one-day gap in coverage occurred when, on a specific day, a participant ran out of their buprenorphine supply, given past medication disbursements and expected daily intake. A "long gap" is a contiguous gap of 30 or more days in MOUD; if such gap(s) occurred, their treatment retention duration was defined to be the number of days prior to the first such gap.

2.4. Data preprocessing

We preprocessed the EMA data for each participant to construct a 7 day (21 EMA) sliding window per person. Missing EMA information was coded with a missing value indicator (−1) in the modeling pipeline. A sliding window of 21 or less EMA responses for days 1–84 (EMA responses 1–252) resulted in a total of 231 windows for each person for each EMA type. For each EMA type, we stacked corresponding time windows for each participant. Thus, for each EMA type, each time window for all participants was represented by an array of size (62 [participants] x 21 [EMA measurement points] x n [EMA features]), where n represents the total number of EMA questions for a particular EMA type). Please reference Supplemental Fig. 1 for a schematic representation of such a dataframe.

2.5. Missing data among input features

Given our goal to preserve the temporal dynamics in EMA data, we did not impute missing data among the input features. Rather, we inserted a missing value indicator (−1) for each missing *input* data point prior to running the model. The missing value indicator approach allowed us to use all available data, without discarding incomplete time windows.

2.6. Missing data among outcomes

We replaced missing values in our NPOU outcome with positive use (1). This practice is common in OUD treatment studies because it provides the most conservative estimate of treatment efficacy and NPOU (Biondi et al., 2020). In the same way, we replaced missing values in our EMA-based medication nonadherence outcome with non-adherence (1). However, to mitigate the undue influence of replaced outcomes on either parameter tuning or performance metrics, we imposed sample weighting[5] during model training. The application of sample weights during model training reduced the importance of observations, i.e., individual participant time windows, to an extent directly proportional to the amount of missing outcome values for that participant at that time window. For example, if a given participant was missing 10 % of their outcome values for a given time window, the sample weight for this observation would be 0.90.

Further, to avoid biasing our performance results, we excluded outcome predictions that corresponded to missing EMA-based outcomes when calculating performance metrics, including Area Under the Receiver Operating Characteristic Curve (AUC), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Sensitivity, Specificity, and F1 Score (F1). This means we did not consider the model's predictions for any imputed outcome label in the performance metric calculation. The remaining two outcomes, MOUD nonadherence and treatment retention did not require these considerations because none of their data points were missing; they were derived from the participants' EHR records.

[1] The interested reader can refer to Marsch et al. (2022) for further information regarding the DTECT study protocol.

[2] The day was divided into 3 periods–morning, afternoon, and evening–and the prompts were delivered randomly within each time period. The prompts could be delayed an hour if the participant was preoccupied. Participants were also able to provide self-initiated EMAs, however these were excluded due to low volume.

[3] It is important to emphasize that the self-regulation questions were intended to measure in-the-moment self-regulation and were not an overall measure of trait self-regulation.

[4] The interested reader can refer to Marsch et al. (2022) for more details about the KPNC AMRS treatment programs and the KPNC EHR. The health system from which participants were recruited maintained pharmacy information in the Virtual Data Warehouse.

[5] We used the Keras model training API and applied the sample weight argument within the model.fit() method.

2.7. Modeling approach and evaluation

Given our time-series data format, combined with our aim to use the raw data without extensive preprocessing, we used a deep recurrent neural network (RNN) architecture for modeling, specifically an RNN with gated recurrent units (GRU). Deep RNNs with GRUs have been shown to effectively model long and short temporal dependencies in time series data (Chung et al., 2014) and have been successfully used to model time-series data in mental health (Tlachac et al., 2022). We used a 'sequence to sequence' modeling approach, where the input sequence comprised the predictor EMA subtype of shape [62 (persons) x 21 (EMA time points) x n (EMA features, where n was the number of individual EMA questions)] and the output comprised a sequence [62 (persons) x 21 (EMA time points)] for NPOU or [62 (persons) x 7 (EHR-based time points, 1 per day)] for treatment retention, EMA-based MOUD adherence, and EHR-based MOUD adherence. Toward our exploratory predictive aim, we created a one-day latency between predictors and outcome; given the small sample size, we alternated between training and evaluation across each successive time window, beginning with training and ending with evaluation. Evaluation was only performed on data previously unseen by the model to ensure generalizability and prevent overfitting.

For example (see Fig. 1.B), we trained our model using EMA features from the first time window (days 1–7) and outcomes from the second time window (days 2–8). We then evaluated our model using EMA predictors from days 2–8 and outcomes from days 3–9. Saving all model weights, we then updated our model weights by training on EMA inputs from days 2–8 and outcomes from days 3–9. We then evaluated our model on EMA inputs from days 3–9 and outcomes from days 4–10. We continued this process for the remaining time windows, comprising 84 total days of participant EMA completion. Fig. 1.B. displays a schematic of our modeling pipeline. We used AUC as the primary performance metric, given its balance of true positive and false negative rate, and utility in binary classification tasks as compared to other performance metrics (Halimu et al., 2019; Ling et al., 2003). To provide a comprehensive performance report, we also computed sensitivity, specificity, PPV, NPV, and F1.

2.8. Model interpretability

Although our methods permitted moderate interpretability by the nature of our analytic design using EMA subtypes, we additionally used SHapley additive ExPlanations (SHAP) analysis for finer model interpretability (Lundberg & Lee, 2017). In the context of our modeling approach, we used SHAP to determine the marginal contribution of each individual EMA item and hour of the day on the model's prediction. SHAP enables this interpretation by perturbing the values of a feature and assessing its influence on the model's predictions. SHAP values were calculated, the absolute value was taken, and these values were subsequently averaged across persons resulting in a single value per feature per time point. This value corresponds to the relative influence of the feature at a particular time point on the model's prediction of either the negative or positive class.

3. Results

3.1. Study Participants

Participants comprised 62 adults receiving MOUD (Observations = 14,322), with the following self-reported baseline demographic characteristics: 50.0 % self-reported male, 46.8 % self-reported female, 3.2 % are not reported[6]; mean age 37.2 years; 71.0 % White, 29.0 % Non-white or more than one race; 79.0 % Not of Hispanic, Latino or of Spanish Origin, 21.0 % of Hispanic, Latino, or Spanish Origin. NPOU was reported on 410 participant-days, attributable to 32 participants. There were a total of 191 reports of EMA-based medication non-adherence attributable to 34 participants. There were a total of 630 EHR-based nonadherence events attributable to 29 participants. The mean percentage of days retained in treatment was 92.20 [86.37, 98.04] for all participants. As part of their treatment, participants engaged in scheduled group or individual encounters with the Addiction Medicine and Recovery Services Department with a mean length of 74 min. On average, participants completed a total of 13.16 visits (median = 13, SD = 10.35) during the study. The majority of these were video virtual visits (mean = 6.21, median = 3, SD = 8.58) and scheduled telephone visits (mean = 5.98, median = 5, SD = 5.56).

3.2. EMA missing data

Across EMA subtypes, the proportion of missing EMA data ranged from 0.29 to 0.32 for EMA-based medication nonadherence and sleep, respectively. The interested reader may reference Supplemental Table 2 for detailed data by EMA subtype.

3.3. Modeling performance

3.3.1. Non-prescribed opioid use model performance

According to AUC, we found NPOU to be the most strongly predicted on average by all EMAs (AUC range = [0.62–0.97]) compared to other outcomes. Self-reported substance use over the last hour performed the best (AUC = 0.97 [0.96, 0.98]). Sleep and self-reported substance use over the last day perform the most poorly (AUC = 0.62 [0.59, 0.64]). Additionally, the negative predictive value for EMA-based NPOU was much higher than the positive predictive value across all EMA types. The PPV ranged from 0.10 to 0.77, while the NPV ranged from 0.94 to 0.99. All performance metrics for all EMAs are displayed in Table 2.

3.3.2. MOUD EMA-based nonadherence model performance

According to AUC, we found EMA-based medication nonadherence to be the least strongly predicted (AUC range = [0.58–0.68]) compared to all other outcomes. The CTXT and substance use over the last hour EMA subtypes performed the best (AUC = 0.68 [0.64, 0.73]). The SLEEP and substance use over the last day EMA subtypes performed the most poorly (AUC = 0.58 [0.53, 0.63]). Additionally, the negative predictive value for EMA-based treatment nonadherence was much higher than the positive predictive value across all EMA types. The PPV ranged from 0.09 to 0.18, while the NPV ranged from 0.96 to 0.97. All performance metrics for all EMAs are displayed in Table 3.

3.3.3. MOUD EHR-based nonadherence model performance

According to AUC, we found the EHR-based MOUD nonadherence to be moderately predicted by EMA (AUC range = [0.73, 0.79]), compared to other outcomes. The SURF and medication adherence EMA subtypes performed the best (AUC = 0.79 [0.76, 0.81]). The SLEEP, WITHDWL, STRESS, and CRAVING EMA subtypes performed the most poorly (AUC = 0.73 [0.71, 0.76]). Additionally, the negative predictive value for EHR-based treatment nonadherence was much higher than the positive predictive value across all EMA types. The PPV ranged from 0.30 to 0.71, while the NPV ranged from 0.92 to 0.94. All performance metrics for all EMAs are displayed in Table 4.

[6] Two participants (3.2 %) self-reported "Non-Binary" or "Prefer not to report". Summary statistics are not reported for participants who reported these gender identities because there were fewer than 5 participants who selected either of these categories.

A. EMA Legend

EMA Short	EMA Full	EMA Short	EMA Full
SLEEP	Sleep Quality	SURF	Substance Use Risk Factors
MED-ADHERE	Buprenorphine Adherence	MOOD	Mood
STRESS	Stress Level	CTXT	Location and People
PAIN	Pain Level	HR-USE	Substance Use (Last Hour)
CRAVE	Substance Cravings	TODAY-USE	Substance Use (Today)
WTHDWL	Withdrawal Symptoms	SR	Self Regulation

B. Pipeline Schematic

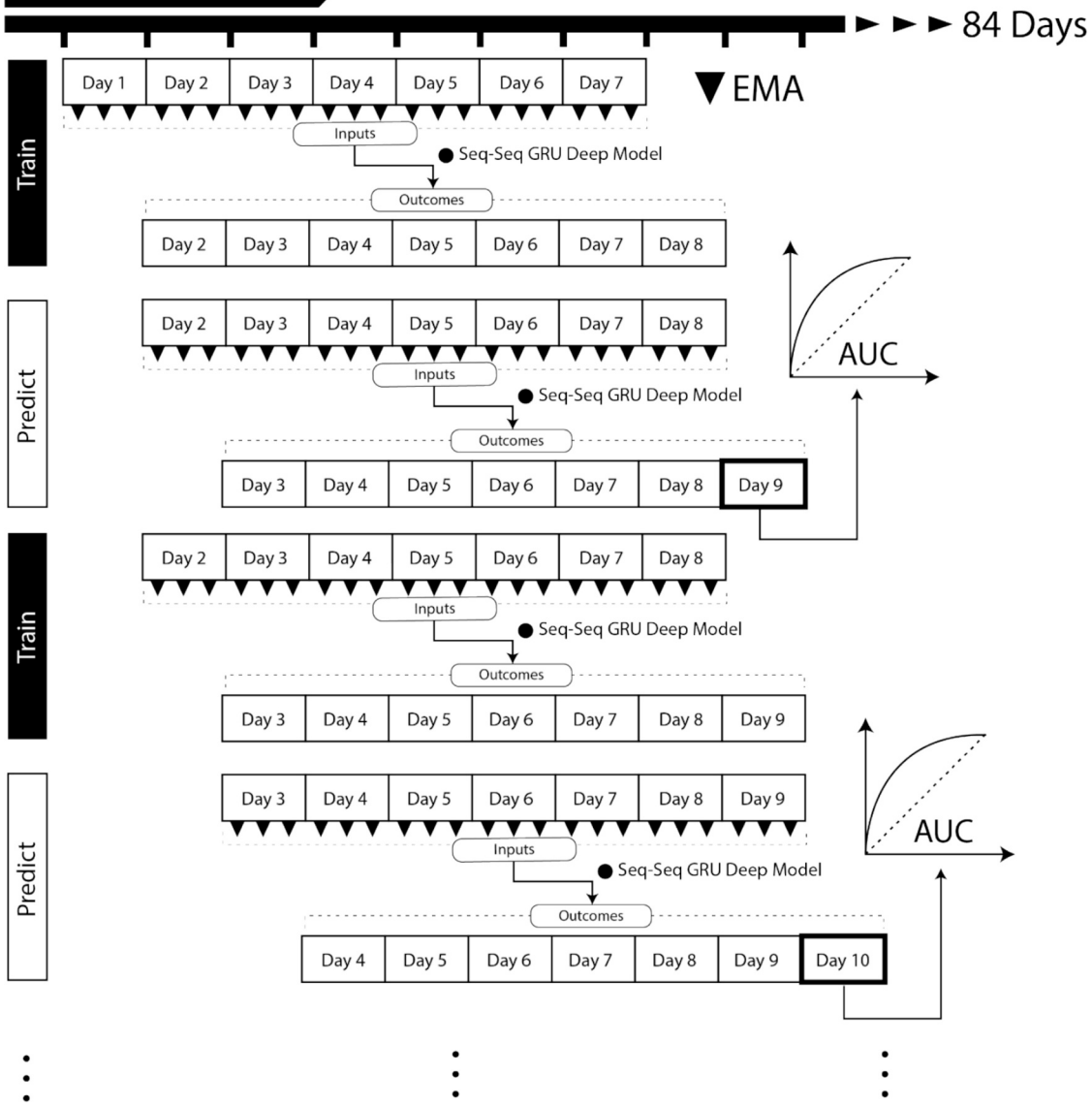

Fig. 1. **(A)** The EMA legend, describing the 12 EMA subtypes with their respective abbreviations. **(B)** Displays the schematic for our modeling pipeline. Note that a sequence to sequence gated recurrent unit (GRU)-based model is run alternately in training and prediction mode in order to train and evaluate performance. Note that the final sequence prediction (denoted by wide border) is removed and used for the performance evaluation (AUC is noted in the figure). Finally, note that the prediction used in evaluation always comprises unseen test data, which has not been used in model training at the time of prediction. The schematic drawn above shows the most general version of our modeling pipeline; note that for the EMA outcome, NPOU, our time window moved forward at the EMA level, rather than the day level.

Table 2
Model performance metrics for NPOU prediction. AUC: Area under the receiving operator characteristic curve; C.I.: confidence interval; NPOU: Non-Prescribed Opioid Use; PPV: Positive predictive value; NPV: Negative Predictive value; F1: F1 Score.

EMA type	Test set performance (NPOU)						
	AUC	95 % C.I.	Sensitivity	Specificity	PPV	NPV	F1
SLEEP	0.62	(0.60, 0.64)	0.74	0.45	0.10	0.95	0.18
TODAY_USED	0.62	(0.59, 0.64)	0.54	0.64	0.12	0.94	0.19
MED_ADHERE	0.65	(0.63, 0.67)	0.75	0.46	0.11	0.95	0.19
PAIN	0.79	(0.77, 0.81)	0.59	0.84	0.24	0.96	0.34
WITHDWL	0.79	(0.77, 0.81)	0.57	0.93	0.43	0.96	0.49
STRESS	0.80	(0.78, 0.82)	0.71	0.76	0.21	0.97	0.32
CRAVING	0.82	(0.80, 0.84)	0.74	0.75	0.20	0.97	0.32
CTXT	0.90	(0.89, 0.92)	0.82	0.86	0.34	0.98	0.48
MOOD	0.91	(0.90, 0.93)	0.81	0.89	0.38	0.98	0.52
SR	0.92	(0.91, 0.93)	0.81	0.92	0.46	0.98	0.59
SURF	0.93	(0.91, 0.94)	0.84	0.91	0.44	0.98	0.57
HR_USED	0.97	(0.96, 0.98)	0.92	0.98	0.77	0.99	0.84

Table 3
Model performance metrics for nonadherence to buprenorphine medication prediction, based on participant self-report. AUC: Area under the receiving operator characteristic curve; C.I.: confidence interval; PPV: Positive predictive value; NPV: Negative Predictive value; F1: F1 Score.

EMA Type	Test set performance (medication nonadherence via EMA)						
	AUC	95 % C.I.	Sensitivity	Specificity	PPV	NPV	F1
SLEEP	0.58	(0.53, 0.62)	0.42	0.75	0.09	0.96	0.15
TODAY_USED	0.58	(0.54, 0.63)	0.33	0.84	0.10	0.96	0.16
PAIN	0.60	(0.55, 0.64)	0.38	0.81	0.10	0.96	0.16
STRESS	0.62	(0.57, 0.66)	0.43	0.78	0.10	0.96	0.16
WITHDWL	0.63	(0.58, 0.67)	0.45	0.80	0.11	0.96	0.18
CRAVING	0.63	(0.58, 0.67)	0.51	0.76	0.11	0.96	0.18
MOOD	0.65	(0.60, 0.69)	0.57	0.71	0.10	0.97	0.17
SURF	0.66	(0.62, 0.71)	0.52	0.77	0.11	0.97	0.19
SR	0.66	(0.61, 0.70)	0.58	0.70	0.10	0.97	0.17
HR_USED	0.68	(0.64, 0.73)	0.46	0.86	0.16	0.97	0.24
CTXT	0.68	(0.64, 0.73)	0.50	0.87	0.18	0.97	0.26

Table 4
Model performance metrics for nonadherence to buprenorphine medication prediction, based on health record. AUC: Area under the receiving operator characteristic curve; C.I.: confidence interval; PPV: Positive predictive value; NPV: Negative Predictive value; F1: F1 Score.

EMA type	Test set performance (medication nonadherence via health record)						
	AUC	95 % C.I.	Sensitivity	Specificity	PPV	NPV	F1
SLEEP	0.73	(0.71, 0.76)	0.50	0.89	0.40	0.92	0.45
WITHDWL	0.73	(0.71, 0.76)	0.56	0.83	0.33	0.93	0.42
STRESS	0.73	(0.71, 0.76)	0.52	0.85	0.35	0.92	0.42
CRAVING	0.73	(0.71, 0.76)	0.51	0.91	0.45	0.93	0.48
CTXT	0.75	(0.72, 0.77)	0.52	0.97	0.71	0.93	0.60
PAIN	0.76	(0.74, 0.79)	0.61	0.84	0.36	0.94	0.45
TODAY_USED	0.76	(0.74, 0.79)	0.59	0.85	0.37	0.93	0.46
MOOD	0.77	(0.74, 0.79)	0.59	0.87	0.40	0.94	0.48
SR	0.77	(0.75, 0.80)	0.59	0.88	0.41	0.94	0.49
HR_USED	0.78	(0.76, 0.81)	0.64	0.86	0.40	0.94	0.49
MED_ADHERE	0.79	(0.76, 0.81)	0.59	0.90	0.46	0.94	0.52
SURF	0.79	(0.76, 0.81)	0.67	0.77	0.30	0.94	0.42

3.3.4. Treatment retention model performance

According AUC, we found the treatment retention outcome to be second to NPOU in overall modeling performance with EMA (AUC range = [0.73, 0.89]) with a narrower performance distribution than the first. The CTXT EMA subtype performed the best (AUC = 0.89 [0.88,0.90]) of all EMA subtypes. The WITHDWL EMA subtype performed the most poorly (AUC = 0.73 [0.70, 0.75]), albeit still with moderate performance. Additionally, the positive predictive value for EHR-based retention was much higher than the negative predictive value across all EMA types. The PPV ranged from 0.95 to 0.98, while the NPV ranged from 0.25 to 0.60. All performance metrics for all EMAs subtypes are displayed in Table 5.

3.4. Results on model interpretability

3.4.1. Non prescribed opioid use SHAP analysis

Two primary trends in the relationship between individual EMA responses across time and NPOU were revealed via SHAP analysis. The first trend showed an short latency between EMA response and NPOU, most notably for CTXT ('Now Walking' and 'Now at the Bar/Club' items), SR ('Thinking' and 'Enjoy Risk' items), SURF ('What if I Used'), and MOOD ('Lively', 'Bored', 'Exhausted', and 'Uneasy' items). Alternatively, the second trend showed a clear gradient for a higher model influence the longer the duration since a response about STRESS (Stress) and PAIN (Level). Refer to Supplemental File 2 for SHAP plots of the items discussed herein.

Table 5
Model performance metrics for retention in treatment prediction. AUC: Area under the receiving operator characteristic curve; C.I.: confidence interval; PPV: Positive predictive value; NPV: Negative Predictive value; F1: F1 Score.

EMA type	Test set performance (treatment retention)						
	AUC	95 % C.I.	Sensitivity	Specificity	PPV	NPV	F1
WITHDWL	0.73	(0.70, 0.75)	0.91	0.54	0.95	0.35	0.93
STRESS	0.73	(0.71, 0.75)	0.91	0.54	0.95	0.35	0.93
MED_ADHERE	0.74	(0.72, 0.76)	0.91	0.6	0.96	0.39	0.94
TODAY_USED	0.76	(0.74, 0.78)	0.90	0.59	0.96	0.36	0.93
PAIN	0.77	(0.75, 0.79)	0.80	0.73	0.97	0.25	0.88
SLEEP	0.80	(0.78, 0.82)	0.90	0.62	0.96	0.38	0.93
SURF	0.81	(0.79, 0.83)	0.87	0.65	0.96	0.33	0.92
CRAVING	0.81	(0.79, 0.83)	0.89	0.68	0.97	0.37	0.93
MOOD	0.83	(0.82, 0.85)	0.86	0.75	0.97	0.33	0.91
SR	0.85	(0.83, 0.86)	0.89	0.74	0.97	0.39	0.93
HR_USED	0.86	(0.84, 0.87)	0.91	0.76	0.98	0.43	0.94
CTXT	0.89	(0.88, 0.90)	0.96	0.74	0.97	0.60	0.96

3.4.2. *EMA-based MOUD nonadherence SHAP analysis*

SHAP analysis for the individual EMA prompts revealed that the most influential features for predicting EMA-based MOUD nonadherence had mixed latencies. Of particular note were SR ('planning' item was most influential at higher latency), MOOD ('boredom', 'exhaustion', and 'contentment' items had higher latencies) and CRAVING ('cocaine', 'meth', and to a lesser extent 'heroin' were most influential at shorter latencies). Refer to Supplemental File 3 for SHAP plots of the items discussed herein.

3.4.3. *EHR-based MOUD nonadherence SHAP analysis*

SHAP analysis for the individual EMA prompts revealed that a majority of influential features for predicting EHR-based MOUD nonadherence had long latencies. Specifically CRAVING ('cocaine' item), SR ('thinking', 'planning', 'cautious', and 'finish' items), MOOD ('bored', 'exhausted', and 'content' items) and HR_USED (any non-prescribed opioid, crack cocaine, benzodiazepines, alcohol, and marijuana items) exhibited a high-latency relationship with the outcome. SURF ('nicotine' item) and PAIN ('level' item) showed low latencies. Refer to Supplemental File 4 for SHAP plots of the items discussed herein.

3.4.4. *Treatment retention SHAP analysis*

SHAP analysis for the individual EMA prompts revealed that the most influential features for predicting treatment retention generally had a high latency, similar to EHR-Based Nonadherence. Specifically, CRAVING ('Cocaine' and 'Meth' items), SR ('Thinking', 'Automatic', 'Planned' and 'Finish' items), SURF (Nicotine), PAIN ('Level', 'Thoughts', and 'Interference' items), MOOD ('Bored' and 'Exhausted' items) and CTXT ('Used Drugs') exhibited this high-latency relationship with the outcome. Refer to Supplemental File 5 for SHAP plots of the items discussed herein.

4. Discussion

Using dense EMA time series, our study was the first to examine dynamic risk factors associated with important outcomes among patients receiving MOUD. Our work is of considerable public health importance given the high prevalence of OUD and the challenges associated with successful MOUD treatment and retention. Although research to date has explored risk factors for opioid relapse and treatment dropout, much of this research has explored immutable or static baseline risk factors, lacking sufficient account for the dynamic and fluctuating course of OUD and its risk factors. In the present analysis, we used novel recurrent deep learning approaches aimed at preserving the dynamic temporal dependencies embedded in raw EMA data, accounting for the dynamic nature of important OUD risk factors. We divided the data according to EMA *subtypes* (e.g., self-regulation, sleep) to understand those OUD-related behaviors most important to relapse, treatment retention, and medication nonadherence. For fine-grained interpretability, we used a robust iterative approach (SHAP) to determine which features drove EMA outcome prediction on the EMA question-level within each EMA subtype.

Our results demonstrated that each EMA subtype had the capacity to predict NPOU, medication nonadherence (measured via both EHR and EMA), and treatment retention better than chance up to 7 days in the future. Each EMA also showed the capacity to predict both self-reported (NPOU and EMA-based medication nonadherence) and objectively assessed (EHR-based medication nonadherence and retention) outcomes for persons receiving MOUD. Notably, substance use within the last hour (HR_USED) EMA responses appeared in the top three most important EMA subtypes across *all* outcomes by AUC (see Tables 2–5). HR_USED also had the highest PPV for NPOU (by far) and treatment retention, the former perhaps not a surprise given the propensity of past use to predict future use. General context variables (CTXT) variables showed importance across outcomes (see Tables 2–5), ranking as the highest by PPV for both EMA- and EHR-based nonadherence. Self-regulation (SR) EMA responses were highly important for predicting NPOU, Treatment Retention, and EMA-based medication non-adherence (see Tables 2–3, 5).

It stands to reason that recent substance use behaviors would predict future substance use. Further, it is not surprising that the self-regulation (SR) EMA subtype is among the most important features, given prior evidence for the relationship between the constructs measured by the SR, such as sensation seeking and emotion regulation (Scherer et al., 2022), and opioid use. Specifically, emotional regulation (Aaron et al., 2020; Lutz et al., 2018; Riquino et al., 2018), notably "nonacceptance of emotional responses" (Gold et al., 2020), sensation seeking (Franques et al., 2003), and impulsivity (Marino et al., 2013) have been associated with opioid use. Further evidence supporting the importance of SR is the demonstrated efficacy of self-regulatory practices on opioid misuse (Garland et al., 2019).

In contrast with the highest performing EMA subtypes, we found more variability across outcomes in the lower performers. We found the lowest performing EMA subtypes were SLEEP, TODAY-USED, and MED-ADHERE for predicting NPOU (see Table 2); SLEEP, TODAY_USED, and PAIN were the lowest for predicting EMA-based nonadherence (see Table 3); STRESS, WITHDWL, and SLEEP were the lowest for predicting EHR-based medication nonadherence (see Table 4); and, WITHDWL, STRESS, and MED-ADHERE were the lowest for predicting treatment retention (see Table 5). While prior evidence supports important substance use risk factors from each of these subtypes (Marsch et al., 2022), we suspect that the low performance of MED-ADHERE, SLEEP, and TODAY-USED may have been negatively impacted by their relatively low temporal density, owing to a single daily measurement. Further, TODAY-USED had sparse positive values, given its use as a backup item for heretofore unreported substance use.

Although we found model performance above chance across EMA subtypes, we noted that PPV was low to moderate in NPOU and medication nonadherence, and NPV was low to moderate in predicting retention. Considering a scenario where such models were deployed for predicting real world patient outcomes, these findings would suggest a need for a careful, statistically- and clinically-informed approach. Specifically, we would expect a relatively high false discovery rate for NPOU and nonadherence models and a relatively high false omission rate for treatment retention models. Thus, considering the risk of downstream effects would be a crucial prerequisite to deployment. An unintended consequence of a false positive in the case of NPOU or medication nonadherence, for instance, could be outcomes similar to false positives on urine drug screens (UDS), with the potential for damage to therapeutic rapport and unnecessary follow up measures.

Beyond a basic understanding of the most predictive EMA subtypes permitted by model performance metrics across EMA subtypes, SHAP analysis allowed for a more granular understanding of particular EMA items and associated temporal latencies most influential in model prediction. As these results flow from an exploratory feasibility study with many EMA questions, an exhaustive discussion of all features for every outcome is outside the scope of this work. Rather, we provide a narrative overview of broad themes observed in the SHAP plots and provide commentary on this unique explanatory method for multivariate time series within a deep learning framework. We further provide a subset of potentially useful EMA prompt subtypes, which have promise for future research aimed at predicting important outcomes among persons receiving MOUD; these include substance use over the last hour (HR_USED), general contextual variables (CTXT), and self-regulation (SR). We refer the interested reader to Supplemental Files 2–5 for all 48 SHAP plots (4 outcomes with 12 EMA subtypes each).

We identified substance use over the last hour as an important predictor for all outcomes. In particular, SHAP analysis revealed cannabis use as influential for predicting both NPOU and retention outcomes, consistent with existing literature identifying cannabis use as a risk factor for OUD (Olfson et al., 2018). Further, we identified cocaine and methamphetamine cravings as influential for prediction treatment retention and, to a lesser extent, MOUD nonadherence. This finding is in line with studies which highlight the considerable co-occurrence of non-prescribed opioid use and psychostimulant use (Ellis et al., 2018; Fischer et al., 2021), charged with particular importance due to the elevated risk for overdose among persons using both opioids and psychostimulants (Palis et al., 2022). Not surprisingly, mood was predictive of future OUD outcomes (Supplemental Figs. 2–5), with *boredom* and/or *exhaustion* showing particular importance across outcomes. Perhaps less obviously, *boredom* and *exhaustion* showed considerably higher temporal latency for nonadherence and retention outcomes compared to NPOU (Supplemental Figs. 2–5). Substance Use Risk Factors (*SURF*) also showed predictive importance across multiple outcomes, with nicotine use showing influence for retention and EHR-based nonadherence (Supplemental Figs. 4–5). This finding may be contextualized in robust literature showing shared high co-occurrence and shared neurobiologic pathways of nicotine and opioid use (Morris & Garver-Apgar, 2020), as well as improved OUD treatment outcomes with nicotine cessation (Lichenstein et al., 2019).

Many EMA predictors indicative of EHR-based nonadherence and retention often showed high temporal latency, observable approximately 5–8 days preceding the outcome (Supplemental Figs. 4–5). This latency in outcomes related to objective measures of nonadherence and retention, in contrast to the NPOU outcome, may be attributable to a mediating role played by *NPOU* between risk factors for substance use and instances of treatment discontinuation or MOUD nonadherence. Specifically, individuals might initially have engaged in NPOU, which, in turn, amplified their likelihood of discontinuing treatment and/or not adhering to MOUD. This possibility aligns with existing literature, indicating that recent NPOU use and relapse escalate the risk of premature treatment discontinuation (Gottlieb et al., 2022; Marcovitz et al., 2016).

Taken together, our results demonstrate promise in the use of EMA to predict clinically meaningful and actionable outcomes in persons receiving MOUD. EMA, when it included physiological, psychological and contextual aspects, proved to be a robust tool in predicting not only important self-reported outcomes (i.e., NPOU, EMA-based medication nonadherence), but also *objective* EHR-derived outcomes (i.e., medication nonadherence and treatment retention). Further, our use of recurrent deep learning methods with dense EMA time series permitted precision in understanding temporal latencies between particular predictors and OUD outcomes; identification of these temporal latencies suggested the existence of potentially salient time windows for intervention. Thus, using personalized and temporally sensitive risk factors for important OUD outcomes could permit the effective delivery of *JITAIs*, (Perski et al., 2022) aimed at preventing relapse and/or maintaining medication adherence and retention. To date, smartphone-based digital interventions have shown feasibility, acceptability, and real-world effectiveness (Maricich et al., 2021) in promoting abstinence and retention in treatment. Considering a cascade of care framework for OUD (Williams et al., 2019), JITAIs harnessing existing evidence based interventions informed by real-time insights into an individual's likelihood of NPOU, MOUD nonadherence and retention could reduce drop-off between stages and increase the proportion of individuals achieving sustained remission.

Although our study used robust methods to investigate dynamic risk factors of NPOU, treatment retention, and medication nonadherence among persons with OUD, our results must be contextualized among several important limitations. First, designed as a feasibility study (Marsch et al., 2022), our sample size is modest. Somewhat mitigating this limitation is the tremendous number of observations within and across persons (Observations = 14,322). Even so, our demographic diversity is limited, impacting the generalizability of our results. During initial data collection there were seven possible categories for race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Pacific Islander, White, Other, or Prefer Not to Answer). However, due to the modest sample size and subsequent low counts for some racial categories, they were collapsed into White and Non-white, limiting more fine-grained demographic analysis. Further, we acknowledge that our sample represents a treatment-enrolled population and is most likely not representative of the larger OUD population, many of whom face barriers, such as internal and external stigma, to treatment engagement (McCurry et al., 2023).

Second, our EMA input features, the NPOU outcome, and the EMA-based medication nonadherence outcome are based on self-reports, which are susceptible to inaccurate reporting, and cognitive biases. Even so, EMA has advantages when compared to extended interval retrospective self-reporting (Stone et al., 2007), such as commonly used biweekly symptom inventories (Kroenke et al., 2001). Third, the intensive and active nature of EMA data collection resulted in some missing data. Our decision to replace missing responses with NPOU or non-adherence, while common practice, may have overestimated the actual prevalence of the self-report based outcomes. We mitigated this by downweighting for missing outcome values during performance evaluation. As a recommendation for future research aimed at further improving predictive power for real-world implementation, we suggest testing more direct EMA questions aimed at measuring treatment satisfaction, e.g., “Are you satisfied with your current MOUD treatment?”

Despite the limitations, our work is the first to use personalized, naturalistic features to predict clinically relevant outcomes in persons receiving MOUD. Further, despite the subjective nature of EMA, our work provides evidence for its capacity to predict important and objective EHR-based outcomes. In so doing, the work highlights the capacity of recurrent deep learning models to inform temporally nuanced predictions of relapse and MOUD treatment attrition. Such an understanding of temporally proximal, personalized risk factors will

allow for the building of timely clinical interventions, which hold tremendous potential in bolstering approaches to curb relapse and avert treatment attrition.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.josat.2025.209685.

CRediT authorship contribution statement

Michael V. Heinz: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **George D. Price:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Avijit Singh:** Writing – review & editing, Writing – original draft. **Sukanya Bhattacharya:** Writing – review & editing, Writing – original draft. **Ching-Hua Chen:** Writing – review & editing, Project administration, Investigation, Data curation, Conceptualization. **Asma Asyyed:** Project administration, Writing – review & editing. **Monique B. Does:** Writing – review & editing, Project administration. **Saeed Hassanpour:** Writing – review & editing, Investigation, Conceptualization. **Emily Hichborn:** Writing – review & editing. **David Kotz:** Writing – review & editing, Investigation, Conceptualization. **Chantal A. Lambert-Harris:** Writing – review & editing, Writing – original draft, Project administration, Investigation, Data curation, Conceptualization. **Zhiguo Li:** Writing – review & editing. **Bethany McLeman:** Writing – review & editing, Project administration. **Varun Mishra:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Catherine Stanger:** Writing – review & editing, Investigation, Conceptualization. **Geetha Subramaniam:** Writing – review & editing, Project administration, Investigation. **Weiyi Wu:** Writing – review & editing, Data curation. **Cynthia I. Campbell:** Writing – review & editing, Project administration, Conceptualization. **Lisa A. Marsch:** Writing – review & editing, Supervision, Conceptualization. **Nicholas C. Jacobson:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization.

Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Funding statement

This work was supported by the National Institute on Drug Abuse (NIDA) Clinical Trials Network grant UG1DA040309 (Northeast Node) (M.H., G.P., A.S., S.B., C-H.C., S.H., E.H., D.K., C.L-H., Z.L., B.M., V.M., C.S., W.W., L.M., N.J.), UG1DA040314 (Health Systems Node) (A.A., C. C., M.D.), NIDA Center grant P30DA029926 (M.H., G.P., A.S., S.B., C-H. C., S.H., E.H., D.K., C.L-H., Z.L., B.M., V.M., C.S., W.W., L.M., N.J.), and National Institute of Health (NIH) T32 CA134286 (M.H.).

Declaration of competing interest

Authors C-H.C. and Z.L. are employed by IBM Research. C.C. and M. D. have received support managed through their institution from the Industry PMR Consortium, a consortium of companies working together to conduct postmarketing studies required by the Food and Drug Administration that assess risks related to opioid analgesic use. L.M. is affiliated with Square2 Systems, Pear Therapeutics and Click Therapeutics. These relationships are extensively managed by her academic institution, Dartmouth College. N.C.J. has received a grant from Boehringer-Ingelheim. NCJ has edited a book through Academic Press and receives book royalties; N.C.J. and M.V.H. receive speaking fees related to their research. M.V.H. has served as a private consultant for ChatNexus and Artisight. G.S. is an employee of the National Institute on Drug Abuse (NIDA), which is the funding agency for the study. She was substantially involved in her role as NIDA Scientific Officer on the study funded by the NIDA Cooperative grant awards. She had no substantial involvement in the other cited grants. This manuscript reflects the views of the authors and may not reflect the opinions, views, and official policy or position of the U.S. Department of Health and Human Services or any of its affiliated institutions or agencies. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

Not applicable.

References

Aaron, R. V., Finan, P. H., Wegener, S. T., Keefe, F. J., & Lumley, M. A. (2020). Emotion regulation as a transdiagnostic factor underlying co-occurring chronic pain and problematic opioid use. *American Psychologist, 75*(6), 796–810. https://doi.org/10.1037/amp0000678

Alexander, K., Sanjuan, P., & Terplan, M. (2023). The use of ecological momentary assessment methods with people receiving medication for opioid use disorder: A systematic review. *Current Addiction Reports, 10*(3), 366–377. https://doi.org/10.1007/s40429-023-00492-5

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Association.

Bertz, J. W., Panlilio, L. V., Stull, S. W., Smith, K. E., Reamer, D., Holtyn, A. F., … Preston, K. L. (2022). Being at work improves stress, craving, and mood for people with opioid use disorder: Ecological momentary assessment during a randomized trial of experimental employment in a contingency-management-based therapeutic workplace. *Behaviour Research and Therapy, 152*, Article 104071. https://doi.org/10.1016/j.brat.2022.104071

Biondi, B. E., Zheng, X., Frank, C. A., Petrakis, I., & Springer, S. A. (2020). A literature review examining primary outcomes of medication treatment studies for opioid use disorder: What outcome should be used to measure opioid treatment success? *Am. J. Addict., 29*, 249–267. https://doi.org/10.1111/ajad.13051

Campbell, C. I., Chen, C.-H., Adams, S. R., Asyyed, A., Athale, N. R., Does, M. B., … Marsch, L. A. (2023). Patient engagement in a multimodal digital phenotyping study of opioid use disorder. *Journal of Medical Internet Research, 25*, Article e45556. https://doi.org/10.2196/45556

Centers for Disease Control and Prevention, National Center for Health Statistics. National Vital Statistics System, Mortality 2018-2023 on CDC WONDER Online Database, released in 2024. Data are from the Multiple Cause of Death Files, 2018-2023, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed at http://wonder.cdc.gov/mcd-icd10-expanded.html on Jan 16, 2025 5:09:03 PM.

Chalana, H., Kundal, T., Gupta, V., & Malhari, A. S. (2016). Predictors of relapse after inpatient opioid detoxification during 1-year follow-up. *Journal of Addiction, 2016*, Article 7620860. https://doi.org/10.1155/2016/7620860

Chang, Y.-P., & Raynor, T. (2021). Factors associated with relapse in individuals with opioid use disorder receiving suboxone in rural areas. *Journal of Addictions Nursing, 32*(1), 20. https://doi.org/10.1097/JAN.0000000000000381

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling (arXiv:1412.3555). arXiv. doi:10.48550/arXiv.1412.3555.

Ellis, M. S., Kasper, Z. A., & Cicero, T. J. (2018). Twin epidemics: The surging rise of methamphetamine use in chronic opioid users. *Drug and Alcohol Dependence, 193*, 14–20. https://doi.org/10.1016/j.drugalcdep.2018.08.029

Fatseas, M., Serre, F., Swendsen, J., & Auriacombe, M. (2018). Effects of anxiety and mood disorders on craving and substance use among patients with substance use disorder: An ecological momentary assessment study. *Drug and Alcohol Dependence, 187*, 242–248. https://doi.org/10.1016/j.drugalcdep.2018.03.008

Fischer, B., O'Keefe-Markman, C., Lee, A., Min, H., & Daldegan-Bueno, D. (2021). 'Resurgent', 'twin' or 'silent' epidemic? A select data overview and observations on increasing psycho-stimulant use and harms in North America. *Substance Abuse Treatment, Prevention, and Policy, 16*, 17. https://doi.org/10.1186/s13011-021-00350-5

Fishman, M., Vo, H. T., Burgower, R., Ruggiero, M., Rotrosen, J., Lee, J., & Nunes, E. (2020). Treatment trajectories during and following a medication trial for opioid use disorder: Moving from research as usual to treatment as usual. *Journal of Addiction Medicine, 14*(4), 331–336. https://doi.org/10.1097/ADM.0000000000000592

Franques, P., Auriacombe, M., Piquemal, E., Verger, M., Brisseau-Gimenez, S., Grabot, D., & Tignol, J. (2003). Sensation seeking as a common factor in opioid dependent subjects and high risk sport practicing subjects. A cross sectional study. *Drug and Alcohol Dependence, 69*(2), 121–126. https://doi.org/10.1016/S0376-8716(02)00309-5

Garland, E. L., Hanley, A. W., Riquino, M. R., Reese, S. E., Baker, A. K., Salas, K., … Howard, M. O. (2019). Mindfulness-oriented recovery enhancement reduces opioid misuse risk via analgesic and positive psychological mechanisms: A randomized

controlled trial. *Journal of Consulting and Clinical Psychology, 87*(10), 927–940. https://doi.org/10.1037/ccp0000390
Gold, A. K., Stathopoulou, G., & Otto, M. W. (2020). Emotion regulation and motives for illicit drug use in opioid-dependent patients. *Cognitive Behaviour Therapy, 49*(1), 74–80. https://doi.org/10.1080/16506073.2019.1579256
Gottlieb, A., Yatsco, A., Bakos-Block, C., Langabeer, J. R., & Champagne-Langabeer, T. (2022). Machine Learning for Predicting Risk of Early Dropout in a Recovery Program for Opioid Use Disorder. *Healthcare, 10*(2), 223. https://doi.org/10.3390/healthcare10020223
Halimu, C., Kasem, A., & Newaz, S. H. S. (2019). Empirical comparison of area under ROC curve (AUC) and Mathew correlation coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification. In *Proceedings of the 3rd international conference on machine learning and soft computing* (pp. 1–6). https://doi.org/10.1145/3310986.3311023
Heinz, M. V., Lekkas, D., Abreu, V., Lee, C., Marsch, L. A., & Jacobson, N. C. (2024). Evaluating a mobile app's effects on depression and anxiety in medication-treated opioid use disorder. *Npj Mental Health Research, 3*(1), 1–14. https://doi.org/10.1038/s44184-024-00086-7
Heiwe, S., Lönnquist, I., & Källmén, H. (2011). Potential risk factors associated with risk for drop-out and relapse during and following withdrawal of opioid prescription medication. *European Journal of Pain, 15*(9), 966–970. https://doi.org/10.1016/j.ejpain.2011.03.006
Howard, H., & Guastaferro, W. F. (2019). Mothers and opioid use disorder: Moving toward justice, wellness, and engagement. *Journal of the American Academy of Matrimonial Lawyers, 32*, 29.
Huhn, A. S., Harris, J., Cleveland, H. H., Lydon, D. M., Stankoski, D., Cleveland, M. J., … Bunce, S. C. (2016). Ecological momentary assessment of affect and craving in patients in treatment for prescription opioid dependence. *Brain Research Bulletin, 123*, 94–101. https://doi.org/10.1016/j.brainresbull.2016.01.012
Keyes, K. M., Rutherford, C., Hamilton, A., Barocas, J. A., Gelberg, K. H., Mueller, P. P., … Cerdá, M. (2022). What is the prevalence of and trend in opioid use disorder in the United States from 2010 to 2019? Using multiplier approaches to estimate prevalence for an unknown population size. *Drug and Alcohol Dependence Reports, 3*, Article 100052. https://doi.org/10.1016/j.dadr.2022.100052
Kraus, M. L., Alford, D. P., Kotz, M. M., Levounis, P., Mandell, T. W., Meyer, M., … Wyatt, S. A. (2011). Statement of the American Society of Addiction Medicine Consensus Panel on the use of buprenorphine in office-based treatment of opioid addiction. *Journal of Addiction Medicine, 5*(4), 254. https://doi.org/10.1097/ADM.0b013e3182312983
Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine, 16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
Lichenstein, S. D., Zakiniaeiz, Y., Yip, S. W., & Garrison, K. A. (2019). Mechanisms and clinical features of co-occurring opioid and nicotine use. *Current Addiction Reports, 6*(2), 114–125. https://doi.org/10.1007/s40429-019-00245-3
Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: A better measure than accuracy in comparing learning algorithms. In Y. Xiang, & B. Chaib-draa (Eds.), *Vol. 2671. Advances in artificial intelligence* (pp. 329–341). Berlin Heidelberg: Springer. https://doi.org/10.1007/3-540-44886-1_25.
Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 4765–4774.
Luo, F. (2021). State-level economic costs of opioid use disorder and fatal opioid overdose—United States, 2017. *MMWR. Morbidity and Mortality Weekly Report, 70*. https://doi.org/10.15585/mmwr.mm7015a1
Lutz, J., Gross, R. T., & Vargovich, A. M. (2018). Difficulties in emotion regulation and chronic pain-related disability and opioid misuse. *Addictive Behaviors, 87*, 200–205. https://doi.org/10.1016/j.addbeh.2018.07.018
Ma, J., Bao, Y.-P., Wang, R.-J., Su, M.-F., Liu, M.-X., Li, J.-Q., … Lu, L. (2019). Effects of medication-assisted treatment on mortality among opioids users: A systematic review and meta-analysis. *Molecular Psychiatry, 24*(12), 1868–1883. https://doi.org/10.1038/s41380-018-0094-5
MacArthur, G. J., Minozzi, S., Martin, N., Vickerman, P., Deren, S., Bruneau, J., … Hickman, M. (2012). Opiate substitution treatment and HIV transmission in people who inject drugs: Systematic review and meta-analysis. *BMJ (Clinical Research Ed.), 345*, Article e5945. https://doi.org/10.1136/bmj.e5945
Marcovitz, D. E., McHugh, R. K., Volpe, J., Votaw, V., & Connery, H. S. (2016). Predictors of early dropout in outpatient buprenorphine/naloxone treatment. *Am. J. Addict., 25*, 472–477. https://doi.org/10.1111/ajad.12414
Maricich, Y. A., Xiong, X., Gerwien, R., Kuo, A., Velez, F., Imbert, B., … Williams, K. (2021). Real-world evidence for a prescription digital therapeutic to treat opioid use disorder. *Current Medical Research and Opinion*. https://doi.org/10.1080/03007995.2020.1846023
Marino, E. N., Rosen, K. D., Gutierrez, A., Eckmann, M., Ramamurthy, S., & Potter, J. S. (2013). Impulsivity but not sensation seeking is associated with opioid analgesic misuse risk in patients with chronic pain. *Addictive Behaviors, 38*(5), 2154–2157. https://doi.org/10.1016/j.addbeh.2013.01.020
Marsch, L. A., Chen, C.-H., Adams, S. R., Asyyed, A., Does, M. B., Hassanpour, S., … Campbell, C. I. (2022). The feasibility and utility of harnessing digital health to understand clinical trajectories in medication treatment for opioid use disorder: D-TECT study design and methodological considerations. *Frontiers in Psychiatry, 13*. https://doi.org/10.3389/fpsyt.2022.871916
Mattson, C. L., Tanz, L. J., Quinn, K., Kariisa, M., Patel, P., & Davis, N. L. (2021). Trends and geographic patterns in drug and synthetic opioid overdose deaths—United States, 2013–2019. *Morbidity and Mortality Weekly Report, 70*(6), 202–207. https://doi.org/10.15585/mmwr.mm7006a4
McCurry, M. K., Avery-Desmarais, S., Schuler, M., Tyo, M., Viveiros, J., & Kauranen, B. (2023). Perceived stigma, barriers, and facilitators experienced by members of the opioid use disorder community when seeking healthcare. *Journal of Nursing Scholarship, 55*(3), 701–710. https://doi.org/10.1111/jnu.12837
Mishra, V., Künzler, F., Kramer, J.-N., Fleisch, E., Kowatsch, T., & Kotz, D. (2021). Detecting receptivity for mHealth interventions in the natural environment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5*(2), 74. https://doi.org/10.1145/3463492
Morris, C. D., & Garver-Apgar, C. E. (2020). Nicotine and opioids: A call for co-treatment as the standard of care. *The Journal of Behavioral Health Services & Research, 47*(4), 601–613. https://doi.org/10.1007/s11414-020-09712-6
Mun, C. J., Finan, P. H., Epstein, D. H., Kowalczyk, W. J., Agage, D., Letzen, J. E., … Preston, K. L. (2021). Craving mediates the association between momentary pain and illicit opioid use during treatment for opioid-use disorder: An ecological momentary assessment study. *Addiction, 116*(7), 1794–1804. https://doi.org/10.1111/add.15344
Naji, L., Dennis, B. B., Bawor, M., Plater, C., Pare, G., Worster, A., … Samaan, Z. (2016). A prospective study to investigate predictors of relapse among patients with opioid use disorder treated with methadone. *Substance Abuse: Research and Treatment, 10*, 9–18. https://doi.org/10.4137/SART.S37030
National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, & Committee on Medication-Assisted Treatment for Opioid Use Disorder. (2019). In M. Mancher, & A. I. Leshner (Eds.), *Medications for opioid use disorder save lives*. National Academies Press (US). http://www.ncbi.nlm.nih.gov/books/NBK538936/.
Olfson, M., Wall, M. M., Liu, S.-M., & Blanco, C. (2018). Cannabis use and risk of prescription opioid use disorder in the United States. *The American Journal of Psychiatry, 175*(1), 47–53. https://doi.org/10.1176/appi.ajp.2017.17040413
Palis, H., Gan, W., Xavier, C., Desai, R., Scow, M., Sedgemore, K., … Slaunwhite, A. (2022). Association of opioid and stimulant use disorder diagnoses with fatal and nonfatal overdose among people with a history of incarceration. *JAMA Network Open, 5*(11), Article e2243653. https://doi.org/10.1001/jamanetworkopen.2022.43653
Perski, O., Hébert, E. T., Naughton, F., Hekler, E. B., Brown, J., & Businelle, M. S. (2022). Technology-mediated just-in-time adaptive interventions (JITAIs) to reduce harmful substance use: A systematic review. *Addiction (Abingdon, England), 117*(5), 1220–1241. https://doi.org/10.1111/add.15687
Petneházi, G. *Recurrent neural networks for time series forecasting*. (2019, January 1). arXiv.Org https://arxiv.org/abs/1901.00069v1.
Ponizovsky, A. M., & Grinshpoon, A. (2007). Quality of life among heroin users on buprenorphine versus methadone maintenance. *The American Journal of Drug and Alcohol Abuse, 33*(5), 631–642. https://doi.org/10.1080/00952990701523698
Preston, K. L., Kowalczyk, W. J., Phillips, K. A., Jobes, M. L., Vahabzadeh, M., Lin, J.-L., … Epstein, D. H. (2017). Context and craving during stressful events in the daily lives of drug-dependent patients. *Psychopharmacology, 234*(17), 2631–2642. https://doi.org/10.1007/s00213-017-4663-0
Riquino, M. R., Priddy, S. E., Howard, M. O., & Garland, E. L. (2018). Emotion dysregulation as a transdiagnostic mechanism of opioid misuse and suicidality among chronic pain patients. *Borderline Personality Disorder and Emotion Dysregulation, 5*, 11. https://doi.org/10.1186/s40479-018-0088-6
Samples, H., Williams, A. R., Olfson, M., & Crystal, S. (2018). Risk factors for discontinuation of buprenorphine treatment for opioid use disorders in a multi-state sample of Medicaid enrollees. *Journal of Substance Abuse Treatment, 95*, 9–17. https://doi.org/10.1016/j.jsat.2018.09.001
Scherer, E. A., Kim, S. J., Metcalf, S. A., Sweeney, M. A., Wu, J., Xie, H., … Marsch, L. A. (2022). Momentary self-regulation: Scale development and preliminary validation. *JMIR Mental Health, 9*(5), Article e35273. https://doi.org/10.2196/35273
Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415
Stone, A. A., Shiffman, S., Atienza, A., & Nebeling, L. (2007). *Science of real-time data capture, The: Self-reports in health research*. Oxford University Press.
Tlachac, M., Flores, R., Reisch, M., Houskeeper, K., & Rundensteiner, E. A. (2022). DepreST-CAT: Retrospective smartphone call and text logs collected during the COVID-19 pandemic to screen for mental illnesses. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 6*(2), 1–32. https://doi.org/10.1145/3534596
Weinstein, Z. M., Kim, H. W., Cheng, D. M., Quinn, E., Hui, D., Labelle, C. T., … Samet, J. H. (2017). Long-term retention in office based opioid treatment with buprenorphine. *Journal of Substance Abuse Treatment, 74*, 65–70. https://doi.org/10.1016/j.jsat.2016.12.010
White, A. G., Spittle, T., Fernan, C., Billmyer, E., Marrett, E., Kwong, W. J., & Rossiter, L. F. (2020). Assessment of work loss associated with prescription-related opioid use disorder: A retrospective analysis of claims data. *Journal of Occupational and Environmental Medicine, 62*(3), 217. https://doi.org/10.1097/JOM.0000000000001802
Williams, A. R., Nunes, E. V., Bisaga, A., Levin, F. R., & Olfson, M. (2019). Development of a cascade of care for responding to the opioid epidemic. *The American Journal of Drug and Alcohol Abuse, 45*(1), 1–10. https://doi.org/10.1080/00952990.2018.1546862
Yedinak, J. L., Goedel, W. C., Paull, K., Lebeau, R., Krieger, M. S., Thompson, C., … Marshall, B. D. L. (2019). Defining a recovery-oriented cascade of care for opioid use disorder: A community-driven, statewide cross-sectional assessment. *PLoS Medicine, 16*(11), Article e1002963. https://doi.org/10.1371/journal.pmed.1002963