npj | mental health research

Article

Article in Press

https://doi.org/10.1038/s44184-026-00251-0

ARES-SI: adaptive reinforcement-enhanced sampling for suicidal ideation

Check for updates

Ross Jacobucci[1] ✉, Taylor A. Burke[2,3], Ying Alison Cheng[4] & Brooke A. Ammerman[1,5]

Suicidal ideation (SI) fluctuates rapidly within individuals, but standard ecological momentary assessment (EMA) uses fixed schedules that miss critical risk windows. We present ARES-SI (Adaptive Reinforcement-Enhanced Sampling for Suicidal Ideation), a reinforcement learning system that personalizes EMA timing to individual risk patterns. ARES-SI combines random forest models to predict active SI probability and non-response likelihood with a Q-learning algorithm to determine optimal assessment intervals (30 min for high-risk states, 3 h for low-risk states, and next day to reduce burden). We trained the system using EMA data from 98 participants across three studies with varied sampling frequencies (4–15 assessments per day; 2359 positive active SI responses) over 28-day protocols. In validation testing ($N = 34$; 374 positive active SI responses), the random forest models achieved an AUC of 0.865 for active SI and 0.750 for non-response. ARES-SI demonstrated 65% higher action selection accuracy than random scheduling (0.781 vs. 0.474; macro F1 = 0.650 vs. 0.377), with stronger detection of high-risk windows (30-min recall = 0.810 vs. 0.392) and efficient low-risk scheduling (3-h precision = 0.972 vs. 0.876). These findings provide initial evidence that reinforcement learning can adapt assessment timing to momentary risk.

Suicide risk is highly dynamic, with transitions from ideation to behavior occurring within minutes to hours[1,2]. Yet, traditional suicide risk assessments remain largely reactive, relying on provider-administered screenings during clinical encounters, such as primary care visits or emergency department admissions[3,4]. These static, point-in-time screenings are further constrained by access barriers, as more than half of Americans live in mental health provider shortage areas[5], resulting in long intervals between assessments. Consequently, many acute risk elevations go undetected between visits, highlighting the need for scalable systems capable of capturing rapidly fluctuating suicide risk in daily life.

Ecological momentary assessment (EMA) has substantially advanced our ability to detect suicide risk in real-time, providing insights into short-term cognitive, affective, and interpersonal processes. While foundational in the knowledge provided about the proximal dynamics of suicide risk, EMA remains limited by self-report burden, compliance decay over time (often within 2–3 weeks)[6], and the fixed or semi-random sample schedules that miss risk states occurring between prompts. These limitations make sustained, high-frequency monitoring difficult to implement over clinically meaningful periods, posing a challenge for just-in-time interventions, which depend on accurate, continuous risk monitoring.

Most digital mental health research efforts have focused on the improving prediction of an outcome, such as suicidal ideation, using increasingly sophisticated models[7], but far fewer studies have addressed the complementary problem of *when* assessments should occur. Optimizing sampling schedules is essential to improve the capture of periods of heightened risk yet existing work has relied on heuristic or static statistical rules to determine sampling times[8,9]. These approaches do not incorporate sequential decision-making algorithms capable of learning optimal timing policies. Reinforcement learning (RL), in contrast, is built to make sequential decisions under uncertainty by continuously updating a policy to maximize expected outcomes based on prior experience. This makes RL particularly well-suited for identifying when assessments should occur in dynamic risk contexts. While RL has achieved remarkable success across other healthcare domains, including managing sepsis[10], physical activity promotion[11], and diabetes management[12], it tends to be absent from digital mental health: a systematic review found zero RL algorithms among 165 mental health AI studies from 2016 to 2021[13].

This gap is especially important given recent advances in adaptive assessment. Several methodological innovations now enable more flexible, data-driven approaches to real-time monitoring: adaptive EMA adjusts survey content or timing based on prior responses; passive sensing collects

[1]Center for Healthy Minds, University of Wisconsin – Madison, Madison, WI, USA. [2]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [3]Department of Psychiatry, Harvard Medical School, Boston, MA, USA. [4]Department of Psychology, University of Notre Dame, Notre Dame, IN, USA. [5]Department of Psychology, University of Wisconsin – Madison, Madison, WI, USA. ✉e-mail: jacobucci@wisc.edu



Fig. 1 | Temporal distribution of EMA data collection across studies. Stacked bar charts displaying weekly EMA response volumes for each of the four studies. Study 1 ($N = 45$) collected data from February to August 2021 using fixed-interval sampling (4 prompts/day). Study 2 ($N = 36$) collected data from August 2021 to February 2022 using variable-interval sampling (3–15 prompts/day). Study 3 ($N = 17$) piloted the initial ARES-SI system (v1.0) from February to November 2023. Study 4 ($N = 45$; $N = 34$ under final algorithm version) deployed ARES-SI v2.1 from February to November 2025. Studies 1 and 2 served as training data for the random forest classifiers; Study 3 was used for engineering validation; Study 4 provided independent evaluation of algorithm performance.

background data from smartphone sensors without active participant input; and micro-randomized trials randomly assign interventions at each decision point to support causal inference about time-varying treatments. Adaptive EMA can substantially reduce item burden while preserving measurement precision[14–16], and passive sensing shows promise for identifying periods of elevated suicide risk with better long-term engagement than self-report[17,18]. Micro-randomized trials have established rigorous causal frameworks for optimizing digital interventions, even with modest samples[19,20], underscoring a broader shift toward sequential, data-driven optimization of digital mental health tools. Yet these innovations have not been integrated with algorithms that can determine *when* assessments should occur, a necessary step for scalable, low-burden systems capable of identifying risk in real time.

Recent adaptive approaches in digital mental health have targeted burden through two mechanisms: within-assessment adaptivity, which uses computerized adaptive testing to reduce survey length while preserving measurement precision[16,21,22], and between-assessment adaptivity, which reduces the total number of prompts delivered. However, neither approach addresses the complementary problem of *when* assessments should occur. To our knowledge, no published system integrates RL-optimized sampling schedules with real-time risk prediction to optimize assessment timing. The Adaptive Reinforcement-Enhanced Sampling for Suicidal Ideation (ARES-SI) system was developed to address this gap by combining ensemble risk prediction (random forests) with reinforcement learning (Q-learning) to personalize assessment timing. This approach aims to concentrate assessments during high-risk windows while reducing prompts during stable periods, improving detection of proximal risk states while minimizing participant burden.

The present paper reports on the development and evaluation of ARES-SI, an RL-informed algorithm designed to optimize the timing of EMA for suicide risk. Using data from 143 participants across four EMA studies with varying sampling schedules, we trained ARES-SI to identify periods of elevated proximal risk (states in which the near-term probability of endorsing active SI is elevated relative to recent levels) and determine optimal assessment windows. We describe the development process and evaluate algorithm performance, in addition to initial indicators of acceptability.

Methods

Data were collected across four remote EMA studies over four distinct periods (see Fig. 1). Studies 1 and 2 served as training data, Study 3 piloted the initial version of the ARES-SI system, and Study 4 served as evaluation data for algorithm performance. All procedures were conducted remotely, requiring a meeting with study staff and identification verification. In Study 1, participants completed a phone session and self-report measures with study staff; and in Studies 2–4, participants completed a video call, including an interview and self-report measures, with study staff prior to the EMA period.

Ecological momentary assessment (EMA)

Four items assessed momentary (i.e., "At this moment…") SI, including two items assessing passive SI (i.e., "Life is not worth living for me"; "There are more reasons to die than to live for me") and active SI (i.e., "I think about taking my life"; "I want to die"), which were answered on a 5-point Likert scale and have been previously validated[23]. For the prediction models, a binary indicator of active SI was created: active SI was coded as present (1) when the sum of the two active SI items exceeded the floor (i.e., combined

Table 1 | Participant demographics

	Study 1[a]	Study 2	Study 3[b]	Study 4[c]
Age (*M*, SD)	26.58 (6.53)	33.61 (12.78)	33.00 (8.24)	36.82 (11.97)
Female (%)	68.3%	66.7%	81.3%	60.0%
Race (%)				
White	70.7%	61.1%	56.3%	51.1%
Black	4.9%	30.6%	18.8%	31.1%
Asian	12.2%	2.8%	6.3%	2.2%
American Indian/ Alaska Native	2.4%	2.8%	0.0%	0.0%
Multiracial	4.9%	2.8%	15.5%	13.3%
Hispanic/Latinx (%)	4.9%	13.9%	0.0%	20.0%
Suicide Attempt History	41.5%	63.6%	75.0%	70.2%

[a]2.4% of the sample identified as "other" race. 2.4% of sample did not provide racial information.
[b]Missing demographic information for 2.8% of sample.
[c]2.2% of sample participant identified as "other" race.

score >2, indicating any endorsement above the lowest anchor on at least one item) and absent (0) otherwise. As recent research has supported the distinction between passive and active SI[24], we selected active SI as our primary outcome for several reasons: (a) major suicide theories highlight a progression from passive to active SI, with active SI representing a more severe form of SI[25,26]; (b) active SI has been prospectively linked to suicide attempts and death[27,28]; and (c) clinical trial outcomes and treatment targets most often focus on more severe suicidal thinking[29]. Together, active SI was chosen because it offers greater specificity for more severe, action-oriented suicide risk, creating greater future clinical translation. If participants endorsed a score greater than 1 on any SI item, they additionally reported the duration of these thoughts (1–60 s, 2–15 min, 16–60 min, 1–3 h, or > 3 h). Across all Studies, when participants reported high levels of active suicidal ideation (≥4 out of 5), crisis resources were provided, and study staff reached out via phone to conduct a comprehensive suicide risk assessment. All participants provided informed consent, and procedures were approved by the Institutional Review Board (University of Notre Dame: #21-12-6965; UW-Madison: #2024-1031). The study was conducted in accordance with the Declaration of Helsinki.

The prompt-level prevalence of active SI varied across studies. Across the training data (Studies 1–3; 8919 completed prompts), 26.4% of responses included active SI (2359 positive responses), with study-level prevalence of 26.9% in Study 1, 34.0% in Study 2, and 16.1% in Study 3. In the validation sample (Study 4), 15.6% of completed prompts included active SI (374 of 2402 responses). The lower base rate in the validation sample underscores that the model was generalizing to a lower-prevalence context.

Momentary emotional state was assessed with 27 single-item adjectives (e.g., *interested, distressed, excited, upset, strong, guilty, scared, hostile, enthusiastic, proud, irritable, alert, ashamed, inspired, nervous, determined, attentive, jittery, active, afraid, tired, bored*). These items captured low- and high-arousal positive affect, low- and high-arousal negative affect, and affective states theoretically linked to perceived burdensomeness, thwarted belongingness, and meaning in life (*useless, like a burden, lonely, like I do not belong, like my life has meaning or purpose*). For each adjective, participants were asked to "indicate to what extent you feel this way in this moment" using a 5-point scale from 1 (*very slightly or not at all*) to 5 (*extremely*).

To capture proximal behavioral and contextual risk factors, each EMA prompt also assessed self-destructive behaviors and social context since the last prompt. Five items assessed recent engagement in or urges for alcohol use, non-alcohol substance use, nonsuicidal self-injury, verbal aggression (yelling or arguing with a family member, friend, or romantic partner), and physical aggression (throwing or hitting something or physically pushing or hitting someone). For each behavior, participants indicated whether it was *not present*, whether they *had the urge but did not* engage, or whether the behavior *occurred*; for alcohol use, substance use, and nonsuicidal self-injury, participants additionally reported whether they were alone at the time (yes/no). Social connectedness was assessed with a yes/no item asking whether participants had any social contact with another person since the last prompt, followed by check-all-that-apply items specifying the mode of contact (in person, phone call, text message, social media direct messaging, social media posts/comments). Participants also indicated whether they had experienced a stressful life event since the last prompt (yes/no) and, if yes, rated how well they felt able to handle this difficulty on a 1 (*not at all*) to 5 (*very much*) scale.

Participants in Study 1 were 45 adults who reported past-year suicidal thoughts or behaviors, assessed via self-report items during an initial phone screening, and were recruited via social media platforms (e.g., Reddit). See Table 1 for sample demographics. Participants received fixed-interval EMA prompts 4 times per day across a maximum possible 31 unique days ($M = 21.7$, SD = 9.2, range = 2–31). Compliance rates averaged 60.8% (SD = 30.3%), ranging from 3.6% to 100%.

Participants in Study 2 were 36 adults who reported past-month suicidal thoughts or behaviors, assessed using items from the Columbia Suicide Severity Rating Scale (C-SSRS) during a baseline screening interview, and were recruited via online platforms (e.g., Craigslist). See Table 1 for sample demographics. Participants received fixed- and variable-interval EMA prompts 3-15 times per day (randomized per day) over a maximum possible 32 unique days ($M = 26.2$, SD = 5.6), range = 2–32). Compliance rates averaged 76.7% (SD = 14.4%), ranging from 43.5% to 97.4%.

Participants in Study 3 were 17 adults who reported past-month suicidal thoughts or behaviors (C-SSRS) and were recruited via online platforms (e.g., Craigslist). Participants engaged with the ARES-SI system, receiving a personalized assessment schedule across a maximum possible 34 unique days ($M = 24.4$, SD = 8.2, range = 1–34). Compliance rates averaged 62.2% (SD = 27.1%), showing a range from 0% to 95.8%.

Participants in Study 4 were 45 adults who reported past-month suicidal thoughts or behaviors (C-SSRS) and were recruited via online platforms (e.g., Craigslist). Participants engaged with the ARES-SI system, receiving a personalized assessment schedule across a maximum possible 29 unique days ($M = 22.5$; SD = 10.1, range = 1–29). Average compliance rates were 69.2% (SD = 27.0%), ranging from 7.1% to 100%. Primary analyses are based on the 34 participants with data collected under the final algorithm version.

Across studies, lifetime rates of suicide attempts ranged from 41.5% to 75.0%. Lifetime attempt history was assessed via self-report during the baseline survey in Study 1 and via interview in Studies 2–4.

Supervised prediction model development

The preprocessing pipeline transformed raw EMA data into a format suitable for machine learning analysis while preserving the temporal structure and handling missing data appropriately. Initial data cleaning involved replacing sentinel values (−9) with NaN to properly represent missing responses, converting timestamps to datetime format, and sorting records by participant ID and notification time to ensure correct temporal ordering. The pipeline then engineered several time-based features critical for capturing assessment patterns: hours since each participant's first assessment, cumulative response count, and hours since the last completed response (calculated only for notifications where the participant actually responded). Behavioral and emotional state variables, originally coded on 1-based scales, were recoded to 0-based scales for consistency. Categorical variables (alcohol use, substance use, NSSI, interpersonal conflict, interpersonal contact, and violent behavior) were one-hot encoded while preserving NaN values to distinguish between "no occurrence" (0) and "missing data" (NaN), which is crucial for accurately modeling non-response patterns. Temporal context was captured through day-of-week indicators and hour-

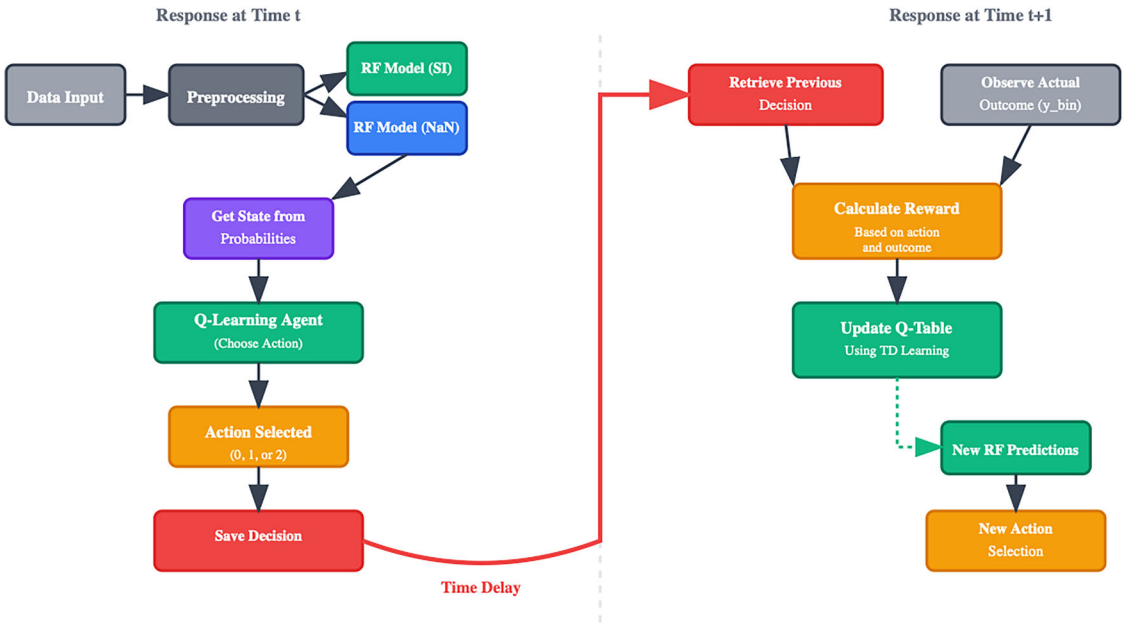


Fig. 2 | ARES-SI system architecture and reinforcement learning workflow. The flowchart illustrates the sequential decision-making process, divided into the action selection phase (Time t) and the policy update phase (Time $t + 1$). Left: Incoming EMA data is preprocessed and fed into dual Random Forest (RF) models to estimate the probability of active suicidal ideation (SI) and non-response. These probabilities are combined to define the current state, allowing the Q-learning agent to select and store an optimal assessment interval (Action 0, 1, or 2). Right: Following the time delay associated with the chosen action, the system retrieves the previous decision and compares it against the observed participant outcome (*y_bin*). A reward is calculated based on the predefined value structure (balancing clinical yield vs. burden), and the Q-table is updated via temporal-difference (TD) learning. The updated policy is then applied to new predictions for the subsequent decision cycle.

of-day features, recognizing that response patterns and symptom severity may vary systematically across time periods.

To enable the random forest classifiers to learn from temporal dependencies in the data, the pipeline created lagged features for all behavioral and emotional variables across four previous time points (lag 1–4). This allowed the model to capture how past states, behaviors, and response patterns influence future suicide risk and response likelihood. A lead version of the binary active SI target (defined above) was created by shifting values forward one time step within each participant, enabling the model to predict active SI at the next assessment. This comprehensive preprocessing ensured that the resulting dataset maintained the longitudinal structure necessary for sequential decision-making while appropriately handling the sparse and irregular response patterns characteristic of real-world EMA data.

The first stage of system development focused on ensuring the stability and discriminative performance of the supervised prediction model for momentary active suicidal ideation (SI) (presence vs. absence). Using data from Studies 1 and 2, a series of candidate models were systematically evaluated to identify the most accurate and computationally efficient algorithm for real-time risk estimation. In addition to Random Forest classifiers[30,31], we tested gradient boosting[32], XGBoost[33], elastic net regularization[34], Bayesian models with horseshoe priors[35], continuous-time structural equation models[36], multilevel Gaussian processes[37], and Facebook Prophet for time series forecasting[38], as well as a general AutoML framework[39]. All models were trained on an identical feature set comprising concurrent and short-lagged EMA responses, temporal indicators (e.g., time of day, day of week, hours since last response), and missingness features to capture non-response risk. Model training was performed on 80% of the sample, with evaluation on a 20% holdout set partitioned by participant to ensure out-of-person generalization.

Hyperparameter tuning was performed for each model class. For LightGBM, tuning included the number of leaves (31–127), tree depth (unlimited, 5, 10), learning rate (0.01–0.1), number of estimators (200–800), and sampling parameters. For Random Forest, tuning included the number of trees (500, 1000), maximum depth (none, 10, 20), and class weights. AutoML employed H2O's automated gradient boosting, XGBoost, GLM, and stacked ensemble methods[39]. Prophet models used five-period ahead hourly forecasts; CTSEM was fit using Bayesian estimation with advanced sampling parameters to ensure convergence; Gaussian process models used gamma-distributed priors to flexibly model temporal structure; and elastic net models were trained with glmnet regularization.

Performance was assessed using ROC–AUC (AUC) and PR–AUC (AUPRC) as primary metrics[40,41], with additional threshold-based confusion matrices to assess clinically relevant decision trade-offs. Across models, gradient boosting and ensemble methods achieved performance comparable to Random Forest, with AUCs typically in the 0.86–0.88 range for active SI prediction and 0.74–0.77 for non-response. Although boosting and ensemble approaches occasionally achieved slightly higher performance, these gains were inconsistent and associated with substantially greater computational cost. By contrast, Random Forest offered a strong balance of predictive accuracy, interpretability, and efficiency, allowing for nightly retraining and rapid inference at scale. This combination of performance and operational feasibility led to the selection of Random Forest as the final model for generating SI and non-response probability estimates used in the reinforcement learning state space. Finally, we evaluated post-hoc probability calibration (Platt scaling and isotonic regression) during development; calibration gains were modest and, because the Q-learning policy operates on discretized probability bins, would not have meaningfully affected state assignment. Calibration was therefore not included in the deployed system.

Reinforcement learning development

ARES-SI was developed across three versions. v1.0 integrated the Random Forest classifiers with rule-based decision logic and fixed probability thresholds (initially 0.11, later 0.234 for active SI); a 25% stochastic component was included to prevent participant habituation. v1.0 was piloted in Study 3 and used to inform the design of a reinforcement learning framework. v2.0 replaced the rule-based logic with tabular Q-learning over a 5×5 discretized state space and the three-action scheduler (30 min, 3 h, next day) described in the preceding section. v2.1 refined the v2.0 probability-bin cutpoints to better capture the mid-range of SI risk probabilities,

Table 2 | ARES-SI reward structure for Q-Learning policy optimization

Action	Participant response	Base reward	Confidence bonus condition	Confidence bonus value	Total possible reward
30 min	SI > 0	10	None	0	10
30 min	SI = 0	−1	None	0	-1
30 min	Non-response	−1	None	0	-1
3 h	SI > 0	−3	None	0	-3
3 h	SI = 0	3	Prob SI = 0 > 0.90	1	4
3 h	Non-response	−1	None	0	-1
Next day	Any response	−1	None	0	-1

Reward values assigned to the reinforcement learning agent based on the scheduled action and observed participant outcome. Positive rewards reinforce clinically desirable behavior: +10 for 30-min assessments that capture active suicidal ideation (SI > 0), and +3 for 3-h assessments during low-risk periods (SI = 0). Penalties discourage suboptimal timing: −3 for 3-h delays when SI was present, and −1 for non-responses or next-day deferrals. A confidence bonus (+1) is awarded when the system schedules a 3-h assessment with high certainty (>90% predicted probability of SI = 0), incentivizing efficient burden reduction during stable periods.

strengthened the reward function to provide clearer negative feedback for missed high-risk moments (notably increasing the penalty for 3-h assessments during elevated SI), and corrected an error in the temporal-difference update. Only data collected under v2.1 (July 22, 2025 onward) are reported in the present analyses; earlier versions served as engineering validation.

Algorithm development proceeded through iterative off-policy evaluation on logged EMA decision streams, with ROC-AUC and PR-AUC (AUPRC) as primary performance metrics and classification reports as secondary diagnostics. We conducted sensitivity analyses varying ε-decay schedule, discount factor, learning rate, state discretization, and action space. Expanding beyond the three core actions did not improve scheduling performance and increased participant burden, so the three-action space was retained. A linear function approximation variant was evaluated in parallel to tabular Q-learning but did not outperform the tabular approach on our state space.

The general algorithm workflow is depicted in Fig. 2. Extending the previous Random Forest system, the implementation employs Q-learning[42], a model-free reinforcement learning technique, to dynamically optimize assessment timing decisions based on cumulative experience across all participants:

$$Q(s_t, \alpha_t) \leftarrow Q(s_t, \alpha_t) + \alpha[r_{t+1} + \gamma_{\alpha'}^{\max} Q(s_{t+1}, \alpha') - Q(s_t, \alpha_t)]$$

This equation describes the temporal-difference update rule used in the Q-learning algorithm. At each time step t, the Q-value for the selected state–action pair (s_t, α_t) is incrementally updated toward the sum of the received reward r_t and the discounted value of the best action in the next state $\max_{\alpha'} Q(s_{t+1}, \alpha')$. The learning rate α determines how quickly new information is incorporated, and the discount factor γ controls the weight placed on future rewards. Over repeated interactions with the environment, this update allows the policy to converge toward the optimal state–action value function.

The system maintains a Q-table, which maps discrete state–action pairs to expected cumulative rewards. States are defined by the joint discretization of two continuous variables: the predicted SI probability and non-response probability, each partitioned into five bins based on the empirical distribution of model-predicted probabilities. The SI probability bins are [0.000,0.234,0.330,0.438,0.671,1.000] and the non-response probability bins are [0.000,0.216,0.312,0.402,0.562,1.000], yielding a 5×5 state space (25 states) and three possible actions: scheduling assessments in 30 min, 3 h, or the following day.

The Q-learning algorithm updates the state–action values through online temporal-difference learning, using a learning rate (α) of 0.15 and a discount factor (γ) of 0.95. After each participant interaction, the algorithm computes a temporal-difference error between the expected and observed rewards and updates the corresponding Q-value accordingly. The reward function (see Table 2) encodes three clinical priorities established by the study team. First, detecting active SI is the highest-value outcome from a safety perspective, warranting the largest reward (+10 for a 30-min prompt that captures SI > 0). Second, correctly deferring assessment during low-risk periods reduces participant burden and sustains long-term compliance, warranting a moderate positive reward (+3 for a 3-h interval when SI = 0). Third, failing to detect active ideation is more clinically costly than an unnecessary prompt or an unanswered one, motivating asymmetric penalties: −3 for scheduling a 3-h deferral when SI > 0, versus −1 for non-response or burden-related scheduling errors. A confidence bonus of +1 is awarded when the system schedules a 3-h assessment with >90% predicted probability of SI = 0, incentivizing efficient burden reduction during stable periods. These magnitudes were refined iteratively across system versions (v2.0 to v2.1), with the most consequential changes being the introduction of a − 3 penalty for missed high-risk windows and a − 1 penalty for 30-min prompts when SI = 0 (both previously 0), which produced clearer temporal-difference learning signals and more stable Q-table convergence[43,44].

To balance exploration and exploitation, the system employs an epsilon-greedy exploration policy with an initial $\varepsilon = 0.25$ (25% random actions), decaying multiplicatively by 0.999 with each update until reaching a floor of $\varepsilon = 0.15$ (15% random actions). This persistent exploration component ensures that all actions continue to be sampled, which both prevents premature convergence and supports subsequent comparison of the learned policy against a fully random scheduling baseline. Over time, this approach enables the system to adaptively tailor assessment timing to individual behavioral patterns while minimizing participant burden. Several critical operational constraints balance participant burden and data quality. The system enforces a strict daily limit of 10 assessments per participant. A minimum of 4 assessments must be completed before the algorithm can defer to the next day.

During daytime deployment, the RL system applies the most recently trained Random Forest models to estimate real-time SI and non-response probabilities for each participant. These models are updated nightly, incorporating all historical data (Studies 1-3) plus new data collected that day, ensuring that state representations reflect the most current behavioral patterns. The RL policy itself is updated online during the day based on observed rewards.

Algorithm performance was evaluated using an action-correctness classification framework. For each scheduling decision, the action taken by the system (30-min or 3-h assessment) was compared to the action that would have been appropriate given the observed participant outcome. Observed active SI (score >2 on the combined active SI items) was mapped to a desired action of 30 min (i.e., rapid follow-up was warranted), while both SI = 0 and non-response outcomes were mapped to a desired action of 3 h. Non-response was assigned to the 3-h class because, in the absence of an observed SI outcome, the 30-min action cannot be validated as correct; this conservative mapping ensures the system is only credited for rapid follow-

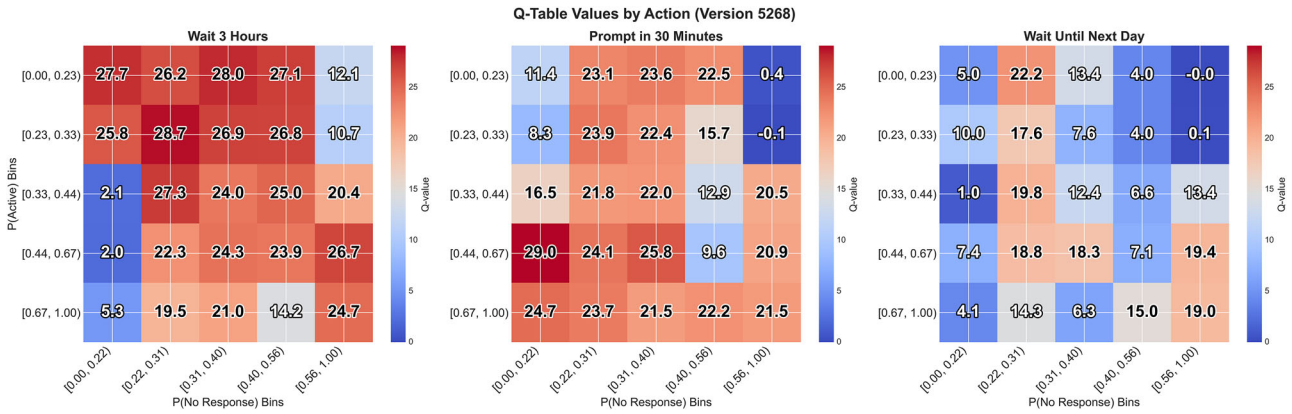


Fig. 3 | Visualization of the learned policy: final Q-table values. Heatmaps displaying the final state-action values ($Q(s, a)$) generated by the reinforcement learning agent after 5,268 cumulative updates (Version 5268). The state space is defined by the intersection of predicted active SI probability bins (y-axis, increasing from top to bottom) and predicted non-response probability bins (x-axis, increasing from left to right). Cell color indicates the expected cumulative reward for taking a specific action in a specific state, with red indicating high value (preferred action) and blue indicating low value. The distinct patterns demonstrate the agent's learned logic: high values for "Wait 3 h" are concentrated in the upper rows (low SI risk), whereas high values for "Prompt in 30 min" shift to the bottom rows (high SI risk), confirming the policy successfully optimized for risk-contingent sampling.

ups when active SI is actually observed. "Next day" actions were omitted from classification analyses because this action served primarily as a burden-limiting fallback and was rarely selected under the policy condition (17 of 2632 decisions). Because the resulting two action classes are substantially imbalanced (active SI was present in approximately 10% of evaluated decisions), we report macro-averaged precision, recall, and F1 alongside overall accuracy. Macro averaging computes each metric separately for the 30-min and 3-h classes and takes the unweighted mean, ensuring that performance on the rarer but clinically important high-risk class contributes equally to the summary metrics. Finally, because the action-correctness mapping treats a 30-min assessment during observed active SI as the target, it is aligned with the reward structure used to train the policy (Table 2). These metrics therefore quantify how consistently the learned schedule reproduces its optimization objective against observed momentary SI, and should be read as policy-consistency rather than as independent evidence of clinical benefit.

Sensitivity and stability analyses

To test whether the learned policy provides advantages beyond a simpler rule applied to the same predictions, we compared it to an optimized threshold rule on the 2615 policy-condition decisions (excluding next-day actions): decisions with predicted SI probability above a cutoff were assigned the 30-min action, otherwise the 3-h action, sweeping cutoffs from 0.00 to 1.00 in increments of 0.005 and selecting the threshold maximizing macro F1. To characterize stability and convergence, we computed classification metrics for chronological halves of the decision stream and in rolling windows, per-participant accuracy and macro F1 for the 33 participants with ≥10 policy-condition decisions (with Spearman correlations against study entry date), and 95% confidence intervals from 2000 participant-level bootstrap resamples. Full results appear in the Supplementary Information (Table S1; Figs. S1–S4).

Results

Results are reported for ARES-SI v2.1 only; earlier versions (v1.0, v2.0) were used for engineering validation.

ARES-SI v2.1 performance

From July 23 to November 4, 2025, the ARES-SI system generated a total of 3,526 EMA scheduling decisionsacross 34 unique participants (mean =103.71 notifications (SD = 55.36) per participant, range = 4–210).

The learned Q-values (Fig. 3) revealed clear patterns in the algorithm's assessment scheduling strategy. The system assigned highest values to 3-h intervals when both SI risk and non-response probability were low, while 30-min assessments received highest values in states with elevated SI risk but lower non-response likelihood. The "next day" action consistently received lower Q-values except in states with very high predicted non-response, demonstrating that the algorithm balanced prioritizing clinically urgent periods with avoiding assessment attempts when the probability of response was low.

Temporal trajectory of performance

To characterize how classification performance changed over the deployment period, we examined its longitudinal trajectory. Because the SI base rate declined substantially over the same window, we interpret apparent temporal gains cautiously (See Fig. 4). This analysis focused specifically on the alignment between the system's decision to trigger a rapid assessment (30 min) and the actual presence of risk (SI > 0) in completed surveys (excluding non-response). As illustrated in Fig. 4, performance followed a trajectory consistent with a transition from exploration to exploitation, though this pattern coincides with a declining SI base rate and cannot be attributed to policy learning alone. During the initial deployment phase (late July to early August), performance metrics were more volatile and lower. This is consistent with the higher early exploration rate (∈), but the early period also carried a substantially higher SI base rate, which depresses summary metrics such as overall accuracy even at a fixed level of performance. Nevertheless, summary performance metrics rose steadily over the deployment period. The 7-day moving average for Accuracy rose from approximately 0.60–0.70 in the initial weeks to the 0.9 s by the study's conclusion.

Similarly, Specificity and Recall showed robust upward trends, with Recall improving from < 0.50 in early August to consistently exceeding 0.80 in September and October. Of these, Recall is least sensitive to the declining base rate, since it is computed only over observed high-risk windows; the accuracy and specificity trends are more vulnerable to the prevalence shift and should be read accordingly.

Classification accuracy and decision outcomes

Across all decision points, a total of 2632 actions were assigned under the policy condition, with 17 "Next day" actions omitted from analysis, resulting in $n = 2615$ valid rows. Under the random condition, 114 "Next day" actions were omitted, yielding $n = 780$ valid rows. Under the policy condition, the overall accuracy (See Fig. 5) of action selection was 0.781, with a macro precision of 0.636, macro recall of 0.794, and a macro F1 of 0.650. Performance was markedly higher for the "3 h" action (precision = 0.972, recall =

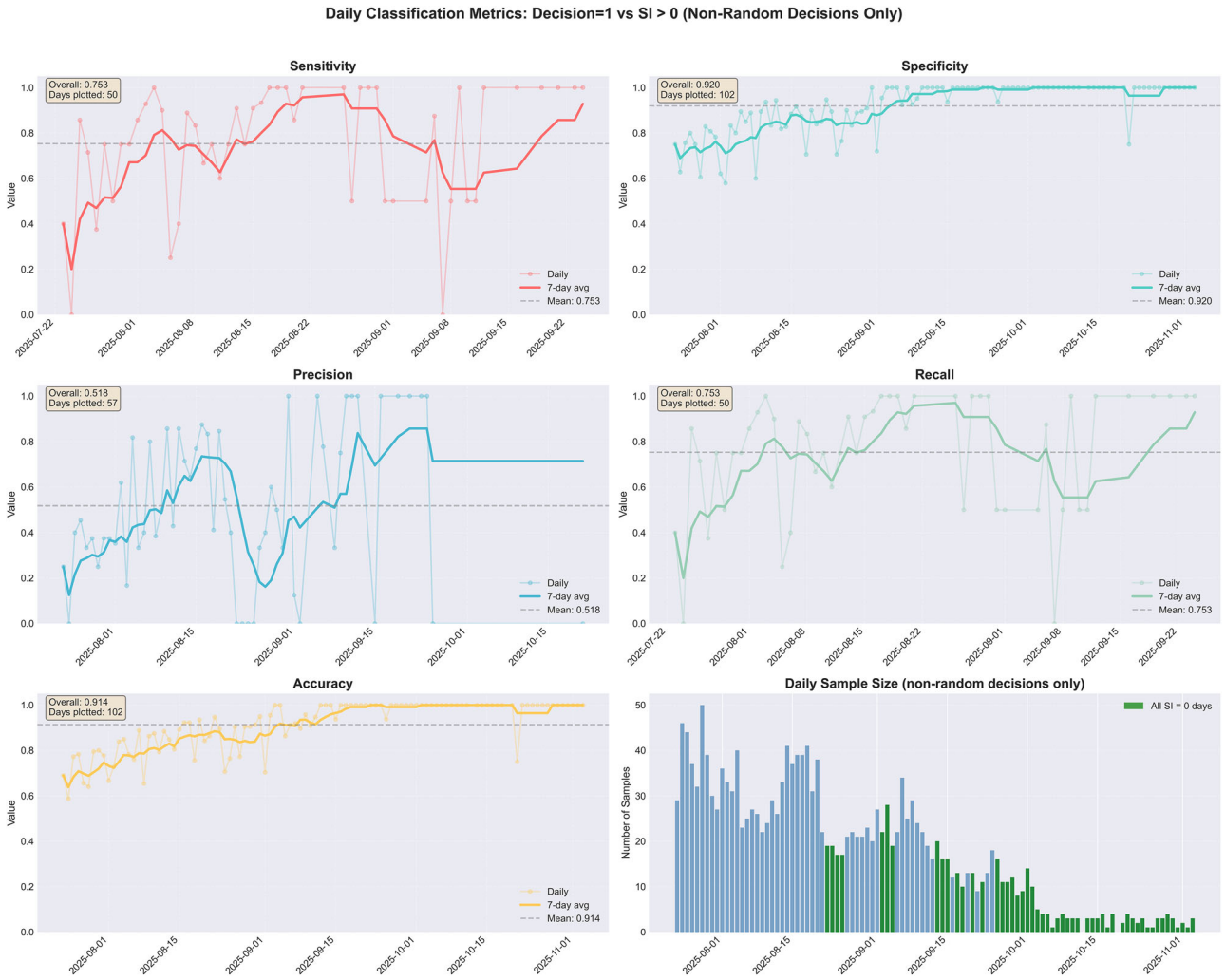


Fig. 4 | Longitudinal performance evolution of the ARES-SI policy. Daily classification metrics comparing the algorithm's decision to schedule a rapid assessment (Action = 30 min) against the observed presence of active suicidal ideation (SI > 0). The analysis is restricted to the Policy condition and excludes instances of non-response. Faint lines represent daily values; solid bold lines represent 7-day moving averages to visualize temporal trends. The bottom-right panel displays the daily sample size of completed assessments. Key metrics (Accuracy, Specificity, Recall) trend upward over the deployment period. Because the SI base rate declined over most of this period, the accuracy and specificity trends are difficult to separate from reduced task difficulty; Recall, computed only over observed high-risk windows, is least affected by the base rate.

0.777, F1 = 0.864) relative to the "30 min" action (precision = 0.299, recall = 0.810, F1 = 0.437), indicating that policy-based decisions were generally well aligned with the desired action, particularly for lower-risk time horizons.

In contrast, under the random condition, overall accuracy was 0.474, with a macro precision of 0.477, macro recall of 0.438, and a macro F1 of 0.377. Although precision for the "3 h" action remained moderate (0.876), recall was substantially lower (0.484), and performance for the "30 min" action was poor (precision = 0.079, recall = 0.392, F1 = 0.131).

Table 3 presents the participant states observed conditional on the assigned assessment interval. In the Policy condition, when a 30-min assessment was assigned ($n = 743$), the subsequent response indicated high risk ($SI \neq 0$) in 29.9% of cases, compared to 7.9% in the Random condition ($n = 393$). Non-response occurred significantly more frequently in 30-min assessments under the Policy condition (43.1%) compared to the Random condition (34.1%; $z = 2.94, p = .003, h = 0.18$). When a 3-h assessment was assigned in the Policy condition ($n = 1872$), 2.8% of outcomes indicated high risk, while 71.3% indicated low risk ($SI = 0$) and 26.0% were non-responses. In the Random condition ($n = 387$), 3-h assessments resulted in high risk in 12.4% of cases, low risk in 55.6%, and non-response in 32.0%.

Number of assessments

From July 23 to November 4, 2025 (3526 total scheduling decisions), the system most frequently selected 3-h prompts (2259; 64.1%), followed by 30-min prompts (1136; 32.2%) and, less often, "Next day" prompts (131; 3.7%). The percentage of 30-min assessments decreased over the study period (Fig. 6). Monthly means reflect the same pattern: August averaged 60.5% 3-h, 35.2% 30-min, and 4.3% next-day; September shifted to 74.5%/22.9%/2.6%; and October to 80.6%/14.4%/5.0%.

Acceptability

Of the 34 participants in the final evaluation sample, 33 completed the post-study feedback questionnaire. Participant responses (answered on 5-point Likert scale, 1 = not at all, 5 = extremely) demonstrated low levels of being overwhelmed ($M = 2.53$, SD = 1.23) or annoyed/frustrated ($M = 2.42$, SD = 1.27) by the number of surveys. Similarly, there were low levels of skipping surveys due to feeling overwhelmed by the number of surveys ($M = 2.19$, SD = 1.38). Participants reported that in future studies they would be willing answer, on average 6.69 surveys per day (SD = 7.07, Median = 5), which is greater than the average number of surveys administered in the study period. The timing of the surveys were reported to capture suicidal thinking moderately well ($M = 3.07$, SD = 1.07). Finally,

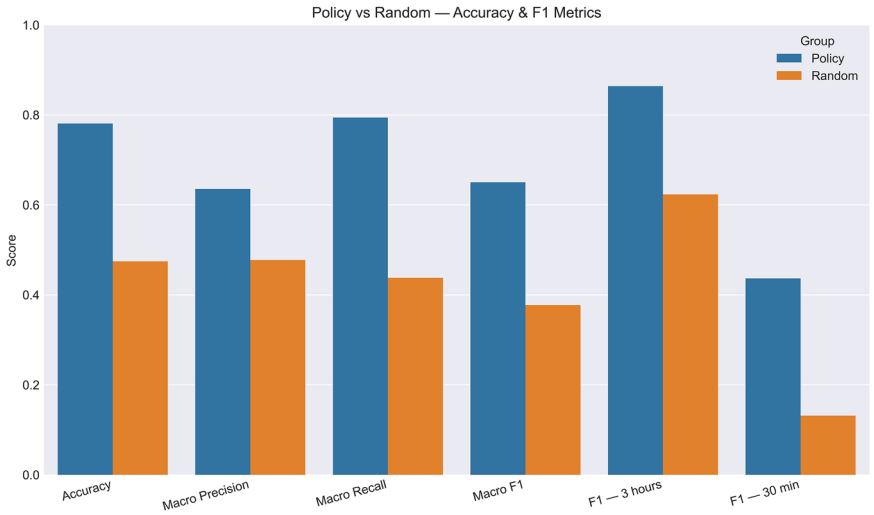


Fig. 5 | Comparative classification performance of the ARES-SI policy vs. random scheduling. The bar chart contrasts global performance metrics (Accuracy, Macro Precision, Macro Recall, Macro F1) and class-specific F1 scores between the Policy (blue) and Random (orange) conditions. The reinforcement learning policy showed stronger performance across all metrics. Notably, the performance gap is widest for the F1 – 30 min metric (representing the identification of high-risk states), where the Policy achieved substantial predictive utility ($F1 \approx 0.44$) compared to the negligible performance of the Random scheduler ($F1 \approx 0.13$).

Table 3 | Observed participant outcomes conditional on assigned assessment interval and study condition

	Random condition		Policy condition	
Outcome	**3-h**	**30-min**	**3-h**	**30-min**
SI = 0	215	228	1334	201
Row % (Sensitivity)	(48.5%)	(51.5%)	(86.9%)	(13.1%)
Col % (Precision)	[55.6%]	[58.0%]	[71.3%]	[27.1%]
SI ≠ 0	48	31	52	222
Row % (Sensitivity)	(60.8%)	(39.2%)	(19.0%)	(81.0%)
Col % (Precision)	[12.4%]	[7.9%]	[2.8%]	[29.9%]
Non-Response (NA)	124	134	486	320
Row % (Sensitivity)	(48.1%)	(51.9%)	(60.3%)	(39.7%)
Col % (Precision)	[32.0%]	[34.1%]	[26.0%]	[43.1%]
Total actions	387	393	1872	743

Frequencies (*N*) represent the participant state observed immediately following a scheduled assessment. Row % (in parentheses) indicates Sensitivity/Recall: the proportion of all instances of a specific state (e.g., *SI*≠0) that were assigned a specific action. Col % [in bold brackets] indicates Precision: the probability of a specific outcome occurring given that a specific action was assigned. The comparison highlights the Policy condition's increased precision in targeting high-risk states (SI≠0) with 30-min prompts compared to the Random condition.

informing the future utilization of ARES-SI, participants reported being very interested in having a smartphone application learn when they are most upset to inform the timing of delivery of mental health treatment information to their smartphone (1 = very disinterested, 5 = very interested; $M = 4.52$, SD = 0.67).

Discussion

This study introduces ARES-SI, the first reinforcement learning (RL) system designed to personalize and optimize the temporal precision of suicide risk monitoring using real-time data. By integrating momentary risk prediction with sequential decision-making, ARES-SI addresses a longstanding barrier in suicide prevention: the field's limited ability to detect rapid, within-person fluctuations in SI[1,2,21] while reducing participant burden. Across multiple evaluation metrics, the adaptive policy outperformed random scheduling, improving overall accuracy by 64.7% and nearly doubling the recall for high-risk states. These findings provide initial evidence that RL can effectively identify clinically meaningful high-risk windows and reduce unnecessary assessments during low-risk periods. Together, these advancements offer a scalable pathway to more efficient real-time risk detection and, ultimately, just-in-time intervention delivery. A supplemental analysis comparing the Q-learning policy to an optimized threshold rule on the random forest predictions confirmed that the state-dependent policy outperformed the best single-cutoff decision rule across all metrics (accuracy: 0.781 vs. 0.698; macro F1: 0.650 vs. 0.572; Supplemental Table and Fig. S1), suggesting that the Q-table's joint conditioning on risk and non-response probability provides advantages beyond what the underlying predictive model alone can achieve. Because the threshold rule was evaluated against outcomes observed under the actually implemented schedule rather than its own proposed timing, this comparison reflects action-selection quality rather than a fully prospective trial.

From a clinical standpoint, ARES-SI demonstrates that automated systems can meaningfully differentiate between periods of acute vulnerability and relative stability, a prerequisite for viable just-in-time interventions[45]. The policy triggered 30-min assessments during 81.0% of high-risk windows and concentrated longer 3-h intervals during periods when suicidal ideation was absent (86.9%). This pattern reflects clinically desirable behavior: intensifying surveillance when risk is elevated while deliberately reducing burden during stable periods. This balance directly responds to well-documented challenges in EMA research, where high-frequency sampling often generates participant burden, compliance decay, and rising missingness[6,46,47]. ARES-SI's adaptive strategy, especially its shift over time, suggests that RL can maintain strong risk detection without imposing the typical burden associated with fixed schedules.

The system's temporal performance trajectory is suggestive of the translational potential of RL for digital mental health. Performance metrics increased over time. We are cautious in attributing this to the agent's increasing ability to identify informative states, because the concurrent decline in SI prevalence offers a competing explanation the current design cannot rule out. This aligns with broader work showing that RL can support individualized optimization in other health domains, such as physical activity[11], chronic disease management[12], and medical treatment[10]. By suggesting that RL may adapt prospectively to idiosyncratic patterns of risk, this study provides initial evidence that adaptive algorithms can also be advantageous in populations experiencing suicidal thoughts. Supplemental stability analyses showed that overall accuracy improved from 0.725 to 0.836 across chronological halves and later-entering participants achieved higher accuracy (Spearman $r = 0.411$, $p = 0.018$), though this improvement was

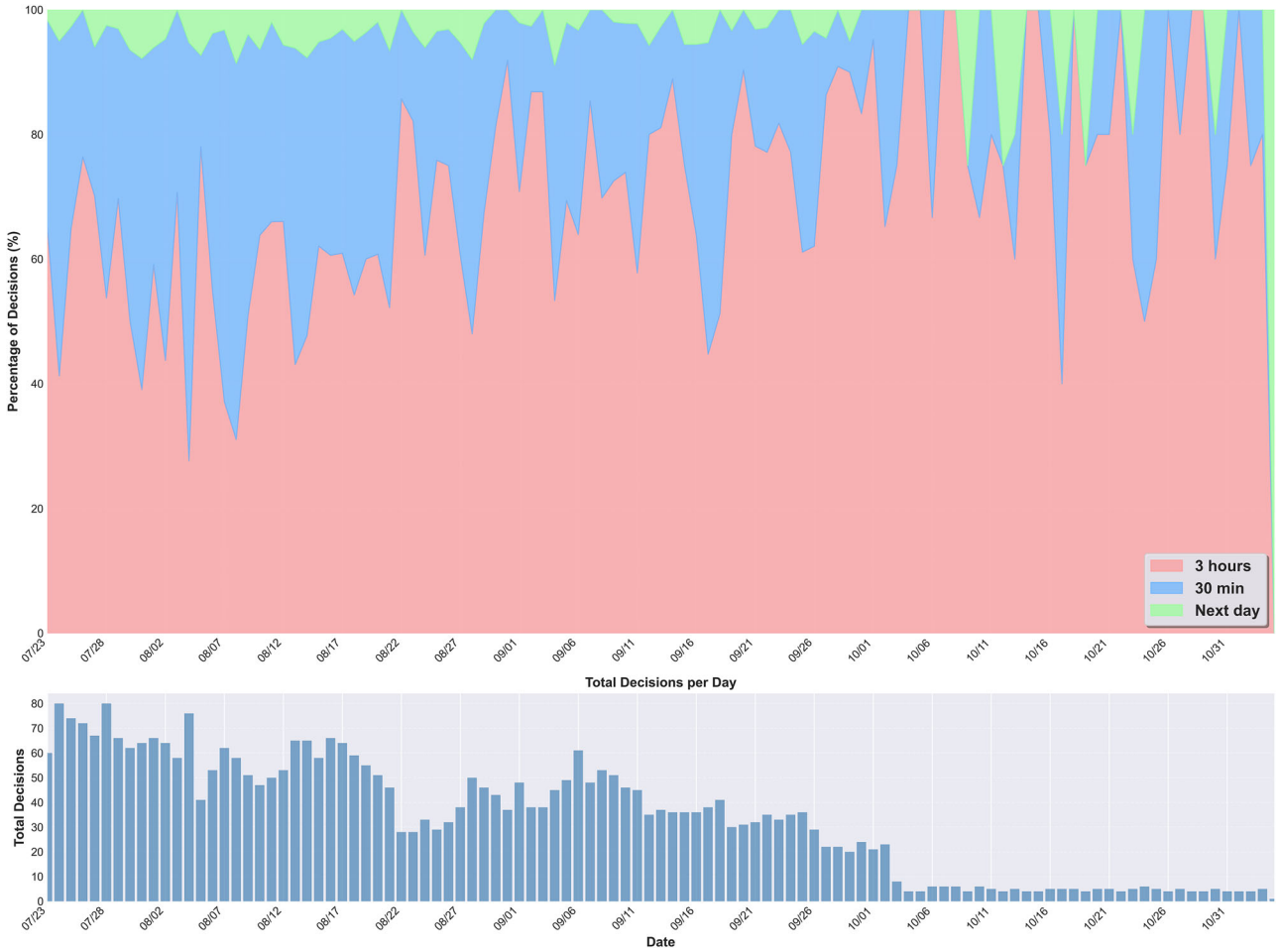


Fig. 6 | Longitudinal distribution of action selection. Top: Stacked area chart illustrating the daily proportion of scheduling decisions assigned by the policy: 3-h assessments (Pink), 30-min assessments (Blue), and Next-day deferrals (Green). Bottom: Daily volume of total decisions (sample size). The visualization reveals a temporal shift in policy behavior: earlier phases (July–August) exhibit a higher proportion of rapid 30-min assessments (reflecting higher initial exploration or risk volatility), whereas later phases (September–October) show a consolidation toward the 3-h interval as the default low-burden action, even as the total number of active participants declined.

partially confounded by a declining SI base rate over the study period (17.2–4.3% across halves). Bootstrap confidence intervals (accuracy: 0.782 [95% CI: 0.736, 0.825]; macro F1: 0.641 [95% CI: 0.554, 0.708]) confirmed that the policy's advantage over random scheduling was robust to variation in sample composition. Because policy and random decisions were interleaved throughout the deployment, both were exposed to this same declining base rate, indicating that the policy's advantage reflects scheduling quality rather than the shift in prevalence (Supplemental Figs. S2–S4).

A major strength of ARES-SI is its potential to serve as the "timing engine" for digital suicide intervention efforts, like JITs. JITs depend on optimizing *when* to deliver an intervention, with timing misalignment highlighted as a major barrier in the field[45]. ARES-SI helps address this problem by identifying windows when a micro-intervention may be most impactful and minimizing delivery during low-yield moments. This capability is particularly relevant as digital mental health research increasingly incorporates high-velocity passive data streams[18,48,49], which often outperform self-report in long-term engagement and prediction but require a reliable scheduling architecture capable of acting on moment-to-moment risk estimates. ARES-SI provides this architecture: a validated framework that can take in continuous risk estimates, generate an optimal action, and update its policy based on observed outcomes. In this way, ARES-SI lays the groundwork for future systems that proactively trigger micro-interventions[50] or safety supports based on converging sensor data.

Participant feedback supports the system's feasibility in suicide research. The system demonstrated improved accuracy over time, and participants reported low levels of burden and frustration despite relatively high survey volumes. Participants also expressed strong interest in future RL-driven mental health tools. Together, these findings suggest that RL-based monitoring can be implemented in ecologically valid contexts with individuals experiencing active SI, representing a meaningful step toward future clinical translation.

Several limitations should guide future work. The validation sample was modest, and although adequate for testing algorithmic performance, larger and more diverse samples will be needed to assess generalizability. The current system also relies on a shared Q-table updated across all participants, which may not fully accommodate individual differences in SI temporal dynamics; future implementations could incorporate participant-level policy adaptation (e.g., contextual bandits or hierarchical RL) to allow individualized parameters while leveraging shared structure. The 28-day window captured near-term fluctuations but did not allow examination of longer-term policy stability.

We emphasize a distinction between two questions. The reported metrics index policy-optimization performance, that is, how closely the learned schedule aligns with observed momentary SI, the objective it was trained toward. They do not constitute clinical-outcome validation, which would require showing that assessment or intervention delivered at RL-identified windows improves patient outcomes. The policy's advantage over

random scheduling on this same metric indicates that the alignment reflects genuine predictive signal rather than the definitional overlap between reward and evaluation; it does not, however, establish clinical utility, which will require an interventional design.

Additionally, although the reward structure was refined iteratively across system versions, a formal sensitivity analysis systematically varying reward parameters was not conducted; off-policy evaluation under counterfactual reward functions would require substantially more logged interactions than available here. Finally, while the system optimized *assessment* timing, the next step is to test whether *interventions* delivered at RL-identified high-risk windows can meaningfully reduce risk. Micro-randomized trials[19,20] provide the methodological infrastructure for testing whether RL-triggered micro-interventions can de-escalate SI in real time.

More broadly, all baseline contrasts here are off-policy: outcomes were observed under the deployed schedule rather than generated by the alternative strategies themselves. Although the random-scheduling actions were actually delivered through persistent exploration, both they and the threshold rule were evaluated within a decision stream produced by the deployed policy; the present results should therefore be read as an initial characterization of adaptive scheduling behavior rather than as evidence that the RL policy would prospectively outperform competing schedulers under independent deployment.

A further consideration concerns how non-response is treated when summarizing performance, as opposed to within the system itself. In the system, non-response is a distinct state: a dedicated random forest estimates non-response probability, which forms a separate axis of the 5×5 state space, and the reward function penalizes non-response equally (−1) whether a 30-min or 3-h assessment was scheduled (Table 2). The policy therefore conditions on unreachability rather than treating it as a low-risk signal. The collapsing arises only in the post-hoc action-correctness analysis, which maps each decision to a single desired action (30 min vs. 3 h); because a 30-min action cannot be validated without an observed SI outcome, non-response was assigned to the 3-h class for scoring. This is a measurement constraint, not a claim that non-response is benign, and it carries one consequence worth stating. If elevated risk increases disengagement or unreachability, a possibility examined in prior work on momentary missingness in suicide EMA[6], then high-risk-but-unreachable windows are invisible to high-risk recall (they never enter the observed SI > 0 denominator) and, when paired with a 3-h action, are scored as correct deferrals. Reported performance may therefore be optimistic in precisely the scenario of greatest clinical concern. The risk relevance of non-response is not unique to momentary research: in a large Army survey, not answering suicide-related questions was among the strongest predictors of later suicide attempt[50]. Future evaluations should treat non-response as its own outcome rather than folding it into the low-risk class, and deployed systems could pair high predicted non-response during elevated risk with active outreach or passive-sensing risk estimation rather than self-report alone.

Taken together, this study suggests that RL can be used to personalize the timing of suicide risk assessment in daily life, an essential step toward developing precision, low-burden digital mental health systems. By improving the detection of proximal risk windows while reducing unnecessary assessments, ARES-SI provides an initial blueprint for adaptive, sensor-informed suicide prevention platforms capable of delivering support at the exact moments when individuals need it most.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to the sensitive nature of individual-level data collected from participants experiencing suicidal thoughts and informed consent provisions that did not permit public data deposition, but are available from the corresponding author on reasonable request. Access to de-identified data will be subject to a data use agreement and approval from the relevant institutional review board.

Code availability

Code for the random forests or Q-learning scheduler is available upon request.

Received: 5 February 2026; Accepted: 28 September 2026;
Published online: 17 October 2026

References

1. Coppersmith, D. D. et al. Mapping the timescale of suicidal thinking. *Proc. Natl. Acad. Sci. USA* **120**, e2215434120 (2023).
2. Millner, A. J., Lee, M. D. & Nock, M. K. Single-item measurement of suicidal behaviors: validity and consequences of misclassification. *PLOS ONE* **12**, e0186613 (2017).
3. Bowers, A. et al. Suicide risk assessment in the emergency department: are there any tools in the pipeline? *Arch. Suicide Res.* **22**, 345–359 (2018).
4. Bryan, C. J., Rozek, D. C., Butner, J. & Rudd, M. D. Patterns of change in suicide ideation signal the recurrence of suicide attempts. *Psychotherapy* **60**, 1–12 (2023).
5. Health Resources & Services Administration. *Health Professional Shortage Areas* (U.S. Department of Health and Human Services, 2023); https://data.hrsa.gov/topics/health-workforce/shortage-areas.
6. Jacobucci, R., Ammerman, B. A. & McClure, K. Examining missingness at the momentary level in clinical research using ecological momentary assessment: implications for suicide research. *J. Clin. Psychol.* **80**, 2147–2162 (2024).
7. Kyron, M. J. et al. Dynamic prediction of inpatient self-harm and suicide attempts using machine learning: toward data-driven clinical monitoring. *J. Consult. Clin. Psychol.* **93**, 761 (2025).
8. Pielot, M. et al. Beyond interruptibility: predicting opportune moments to engage mobile phone users. In *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **1**, 1–25 (ACM, 2017).
9. Rathbun, S. L., Song, X., Neustifter, B. & Shiffman, S. Survival analysis with time varying covariates measured at random times by design. *J. R. Stat. Soc. Ser. C Appl. Stat.* **62**, 419–434 (2013).
10. Wu, T. et al. A value-based deep reinforcement learning model with human expertise in optimal treatment of sepsis. *npj Digit. Med.* **6**, 15 (2023).
11. Liao, P., Greenewald, K., Klasnja, P. & Murphy, S. Personalized HeartSteps: a reinforcement learning algorithm for optimizing physical activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **4**, 1–21 (2020).
12. Aguilera, A. et al. Effectiveness of a digital health intervention leveraging reinforcement learning: results from the Diabetes and Mental Health Adaptive Notification Tracking and Evaluation (DIAMANTE) randomized clinical trial. *J. Med. Internet Res.* **26**, e60834 (2024).
13. Tornero-Costa, R. et al. Methodological and quality flaws in the use of artificial intelligence in mental health research: systematic review. *JMIR Ment. Health* **10**, e42045 (2023).
14. Schneider, S. et al. Just-in-time adaptive ecological momentary assessment (JITA-EMA). *Behav. Res. Methods* **56**, 1365–1388 (2024).
15. Harrison, J. L. et al. Remote symptom monitoring with ecological momentary computerized adaptive testing: pilot cohort study of a platform for frequent, low-burden, and personalized patient-reported outcome measures. *JMIR mHealth uHealth* **11**, e48799 (2023).
16. McClure, K. et al. From repetition to relevance: initial development of a multidimensional computerized adaptive test for intensive longitudinal assessment of suicide risk. *JMIR Form. Res.* **9**, e76544 (2025).
17. Barrigón, M. L. et al. One-week suicide risk prediction using real-time smartphone monitoring: prospective cohort study. *J. Med. Internet Res.* **25**, e43719 (2023).

18. Porras-Segovia, A. et al. Real-world feasibility and acceptability of real-time suicide risk monitoring via smartphones: a 6-month follow-up cohort. *J. Psychiatr. Res.* **149**, 145–154 (2022).
19. Klasnja, P. et al. Micro-randomized trials: an experimental design for developing just-in-time adaptive interventions. *Health Psychol* **34**, 1220–1228 (2015).
20. Qian, M., Chakraborty, B., Maiti, R. & Klasnja, P. The micro-randomized trial for developing digital interventions: experimental design and data analysis considerations. *Psychol. Methods* **27**, 874–894 (2022).
21. Gibbons, C. J. Turning the page on pen-and-paper questionnaires: combining ecological momentary assessment and computer adaptive testing to transform psychological assessment in the 21st century. *Front. Psychol.* **7**, 1933 (2016).
22. Kleiman, E. M. et al. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. *J. Abnorm. Psychol.* **126**, 726 (2017).
23. Forkmann, T. et al. Assessing suicidality in real time: a psychometric evaluation of self-report items for the assessment of suicidal ideation and its proximal risk factors using ecological momentary assessments. *J. Abnorm. Psychol.* **127**, 758–769 (2018).
24. Wastler, H. M. et al. An empirical investigation of the distinction between passive and active ideation: understanding the latent structure of suicidal thought content. *Suicide Life Threat. Behav.* **53**, 219–226 (2023).
25. Van Orden, K. A. et al. The interpersonal theory of suicide. *Psychol. Rev.* **117**, 575 (2010).
26. Klonsky, E. D. & May, A. M. The three-step theory (3ST): a new theory of suicide rooted in the "ideation-to-action" framework. *Int. J. Cogn. Ther.* **8**, 114–129 (2015).
27. Franklin, J. C. et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol. Bull.* **143**, 187 (2017).
28. Wenzel, A. et al. Predictors of suicide relative to other deaths in patients with suicide attempts and suicide ideation: a 30-year prospective study. *J. Affect. Disord.* **132**, 375–382 (2011).
29. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. *Guidance for Industry: Suicidal Ideation and Behavior: Prospective Assessment of Occurrence in Clinical Trials* (Draft Guidance, Revision 1) (2012).
30. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
31. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
32. Ke, G. et al. LightGBM: a highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems* Vol. 30 (2017).
33. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016).
34. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).
35. Carvalho, C. M., Polson, N. G. & Scott, J. G. The horseshoe estimator for sparse signals. *Biometrika* **97**, 465–480 (2010).
36. Voelkle, M. C., Oud, J. H. L., Davidov, E. & Schmidt, P. An SEM approach to continuous time modeling of panel data: relating authoritarianism and anomia. *Psychol. Methods* **17**, 176–192 (2012).
37. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning* (MIT Press, 2006).
38. Taylor, S. J. & Letham, B. Forecasting at scale. *Am. Stat.* **72**, 37–45 (2018).
39. LeDell, E. & Poirier, S. H2O AutoML: scalable automatic machine learning. In *Proc. AutoML Workshop at ICML* (2020).
40. Davis, J. & Goadrich, M. The relationship between precision-recall and ROC curves. In *Proc. 23rd International Conference on Machine Learning* 233–240 (ACM, 2006).
41. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874 (2006).
42. Watkins, C. J. C. H. & Dayan, P. Q-learning. *Mach. Learn.* **8**, 279–292 (1992).
43. Gottesman, O. et al. Guidelines for reinforcement learning in healthcare. *Nat. Med.* **25**, 16–18 (2019).
44. Nock, M. K. et al. Risk factors for the transition from suicide ideation to suicide attempt: results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *J. Abnorm. Psychol.* **127**, 139–149 (2018).
45. Coppersmith, D. D. et al. Just-in-time adaptive interventions for suicide prevention: promise, challenges, and future directions. *Psychiatry* **85**, 317–333 (2022).
46. Jones, A. et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. *Addiction* **114**, 609–619 (2019).
47. Tonkin, M. et al. Evaluating declines in compliance with ecological momentary assessment in longitudinal health behavior research: analyses from a clinical trial. *J. Med. Internet Res.* **25**, e43826 (2023).
48. Auerbach, R. P. et al. Using smartphone GPS data to detect the risk of adolescent suicidal thoughts and behaviors. *JAMA Netw. Open* **8**, e2456429 (2025).
49. Jacobucci, R., Blacutt, M., Ram, N. & Ammerman, B. A. Smartphone screen time and suicide risk in daily life captured through high-resolution screenshot data. *npj Digit. Med.* **8**, 321 (2025).
50. Bentley, K. H. et al. Pilot microrandomized trial of a brief digital intervention for suicidal thoughts. *J. Consult. Clin. Psychol.* **93**, 690 (2025).

Acknowledgements

We gratefully acknowledge the Center for Research Computing at the University of Notre Dame for developing and maintaining the ecological momentary assessment delivery system used in this study.

Author contributions

R.J. and B.A.A. conceptualized the study; R.J., B.A.A., and T.A.B. discussed the reward structure. B.A.A. oversaw data collection; R.J. wrote the main manuscript text; R.J. ran primary analyses. R.J., B.A.A., Y.C., and T.A.B. edited the manuscript.

Funding

The authors received no specific funding for this work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44184-026-00251-0.

Correspondence and requests for materials should be addressed to Ross Jacobucci.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

Supplementary Information

Threshold-Rule Baseline Comparison

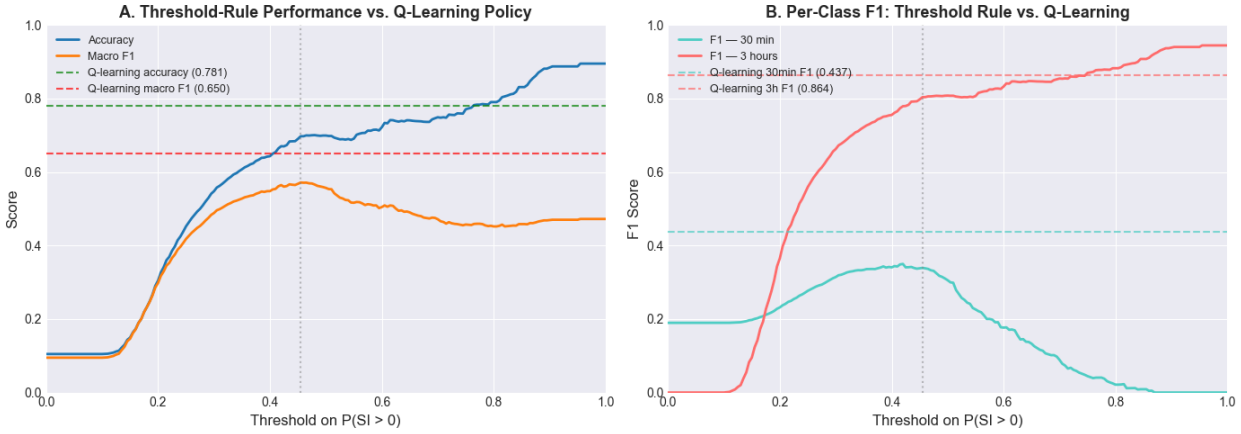
To evaluate whether the Q-learning policy provides advantages beyond those achievable by a simpler decision rule applied to the same random forest predictions, we compared the policy to an optimized threshold rule. For each scheduling decision in the Study 4 policy condition (n = 2,615, excluding next-day actions), we applied a threshold rule: if the random forest-predicted SI probability (prob_01) exceeded a cutoff, assign a 30-minute follow-up; otherwise, assign a 3-hour interval. We swept all thresholds from 0.00 to 1.00 in increments of 0.005 and computed accuracy and macro F1 at each cutoff.

The optimal threshold for macro F1 was 0.455, at which the threshold rule achieved accuracy = 0.698 and macro F1 = 0.572. Supplemental Table S1 presents a side-by-side comparison of the Q-learning policy, the best threshold rule, and random scheduling. The Q-learning policy outperformed the threshold rule across all metrics. Supplemental Figure S1 displays the full threshold sweep, with Q-learning performance plotted as reference lines. Because the threshold rule was never actually deployed, the outcomes it is evaluated against reflect the timing decisions implemented by the Q-learning policy (or random exploration), not the timing the threshold rule would have produced. This comparison should therefore be interpreted as an evaluation of action-selection quality given the same observed state-outcome pairs, rather than a prospective head-to-head trial.

Supplemental Table S1. Comparison of Q-Learning Policy, Optimal Threshold Rule, and Random Scheduling

Metric	Q-Learning	Best Threshold (0.455)	Random
Accuracy	0.781	0.698	0.474
Macro F1	0.650	0.572	0.377
Macro Precision	0.636	0.589	0.477
Macro Recall	0.794	0.717	0.438
F1 — 30 min	0.437	0.339	0.131
F1 — 3 hours	0.864	0.804	0.623

Note. Best threshold selected by maximizing macro F1. Random condition uses uniform random action selection. All metrics computed on policy-condition decisions (n = 2,615) excluding next-day actions.



Supplemental Figure S1. Threshold-rule performance across all probability cutoffs. Panel A displays accuracy and macro F1 as a function of the threshold on P(SI > 0), with Q-learning policy performance shown as dashed reference lines. Panel B displays per-class F1 scores (30-minute and 3-hour actions) across thresholds, illustrating the trade-off between sensitivity to high-risk states and correct low-risk scheduling.

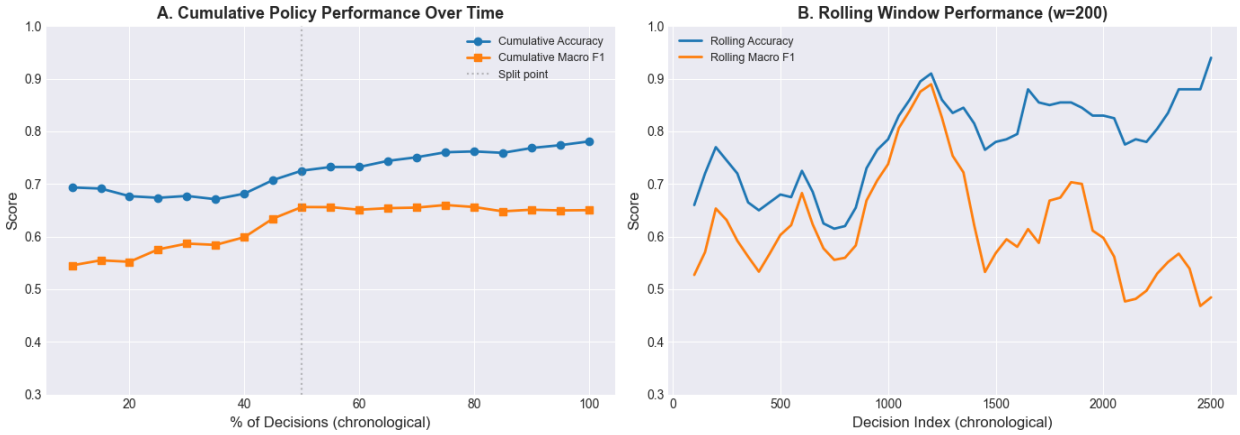
Policy Stability and Convergence Analyses

To evaluate the stability and convergence of the Q-learning policy, we conducted three sets of analyses on the 2,615 policy-condition decisions (excluding next-day actions).

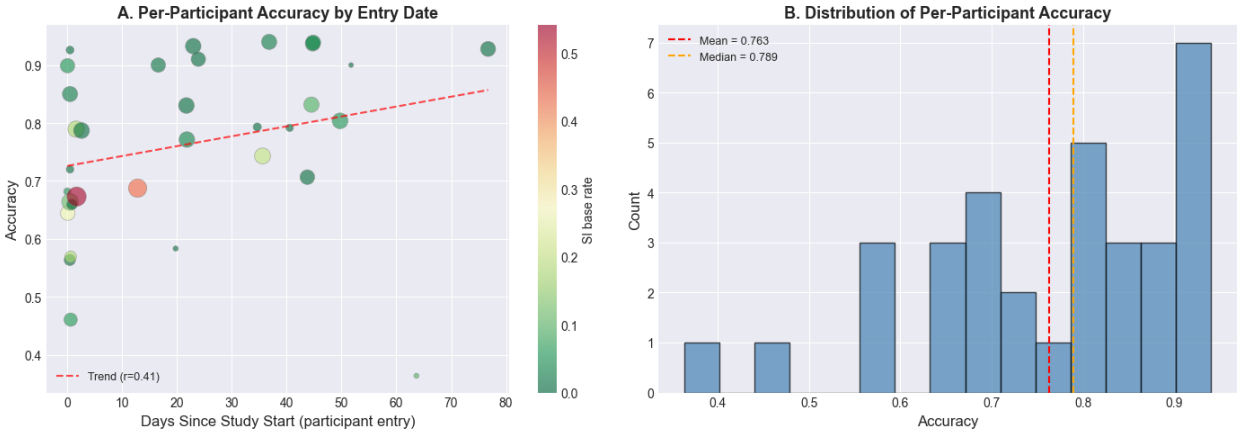
Temporal convergence. We divided decisions into chronological halves and computed classification metrics for each. Overall accuracy improved from 0.725 in the first half (through August 20, 2025; n = 1,307) to 0.836 in the second half (n = 1,308), and 3-hour F1 rose from 0.810 to 0.908. However, 30-minute F1 decreased from 0.502 to 0.277, and macro F1 declined from 0.656 to 0.592. This pattern is substantially confounded by a temporal decline in SI base rates: the split-half SI base rate dropped from 17.2% to 4.3%, with monthly prevalence falling from 16.4% in August to 4.2% in September and zero SI events in October and November (which contributed only 142 and 9 policy-condition decisions, respectively, reflecting the tail end of a small number of participants' protocols). The accuracy improvement therefore largely reflects an easier classification problem, and the 30-minute F1 decline is expected given the near-absence of positive cases in the later months. Disentangling Q-table learning from changing base rates is not possible with the current design. Supplemental Figure S2 displays cumulative and rolling-window performance metrics across the study period.

Per-participant heterogeneity. We computed per-participant accuracy and macro F1 for all 33 participants with at least 10 policy-condition decisions. Mean accuracy was 0.763 (SD = 0.147, range: 0.364–0.940, IQR: 0.673–0.900), and mean macro F1 was 0.522 (SD = 0.110, range: 0.342–0.734). Study entry date was significantly correlated with per-participant accuracy (Spearman r = 0.411, p = .018), but this association likely reflects both Q-table improvement and a declining SI base rate for later participants. Entry date was not significantly associated with macro F1 (r = 0.021, p = .907). Supplemental Figure S3 displays per-participant accuracy by entry date and the distribution of per-participant accuracy.

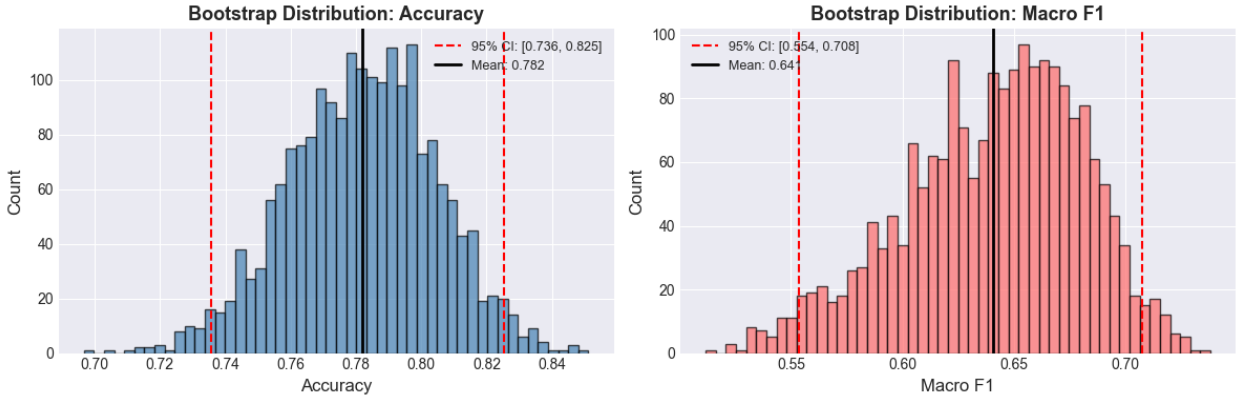
Bootstrap confidence intervals. To assess robustness, we computed 95% bootstrap confidence intervals by resampling participants with replacement (n = 2,000 iterations). The resulting intervals were accuracy: 0.782 [0.736, 0.825] and macro F1: 0.641 [0.554, 0.708]. These intervals do not overlap with the random condition's performance (accuracy: 0.474; macro F1: 0.377), confirming that the policy's advantage is robust to variation in sample composition. Supplemental Figure S4 displays the bootstrap distributions.



Supplemental Figure S2. Policy stability over time. Panel A displays cumulative accuracy and macro F1 as a function of the percentage of chronological decisions included. Panel B displays rolling-window (w = 200) accuracy and macro F1 across the decision sequence.



Supplemental Figure S3. Per-participant accuracy by study entry date. Panel A displays per-participant accuracy as a function of days since study start, with point size proportional to the number of policy-condition decisions and color indicating the participant's SI base rate. Panel B displays the distribution of per-participant accuracy.



Supplemental Figure S4. Bootstrap distributions for policy performance metrics. Histograms of accuracy (left) and macro F1 (right) across 2,000 bootstrap resamples (clustered by participant), with 95% confidence intervals indicated by dashed lines.