Context-responsive cognitive-behavioral micro-interventions for medical-student distress: a 90-day Bayesian computational trial

Ammad Hussain

ABSTRACT

Medical students encounter fluctuating combinations of examination pressure, sleep loss, reduced activity, and psychological distress. Conventional cognitive-behavioral programs provide the same material on a fixed timetable and may therefore miss the moments when a particular skill is most useful. This study asked whether brief, context-responsive cognitive-behavioral support could improve psychological and academic outcomes while requiring less intervention time than a fixed weekly course. A deterministic synthetic cohort of 720 students was distributed equally across six modeled medical schools and three conditions: a context-responsive policy, a fixed-schedule course, and attention control. Ninety-one daily observations per student produced 65,520 longitudinal records. The adaptive policy selected cognitive reframing, paced breathing, behavioral activation, or workload planning from daily distress, sleep debt, activity, workload, and examination proximity. The primary estimand was the entry-adjusted difference in day-90 General Health Questionnaire 12-item (GHQ-12) equivalent score. Weak-prior Bayesian Gaussian regression, state-transition analysis, doubly robust policy-value estimation, campus-specific contrasts, intervention-time accounting, and a pattern-mixture stress test were prespecified. All records and figures are reproducible from the included code and fixed random seed. Mean day-90 GHQ-12 equivalents were 14.17, 17.47, and 20.78 in the context-responsive, fixed-schedule, and attention-control conditions, respectively. The posterior median adaptive-versus-fixed difference was -2.68 points (95% credible interval -3.01 to -2.34; posterior probability of benefit > 0.999). Academic functioning was 2.00 points higher (0.77 to 3.24), and students spent 9.75 fewer days in a high-strain state. A dual response occurred in 67.92% versus 22.50%, while mean intervention time was 236.95 versus 437.22 minutes. The advantage remained negative in every modeled school and throughout the missing-outcome stress test. Within this computational trial, selecting a brief cognitive-behavioral skill from the student's current state produced better symptom and academic trajectories with approximately 200 fewer minutes than fixed scheduling. A prospective multisite micro-randomized trial is warranted before clinical or curricular adoption.

Keywords: adaptive intervention; cognitive-behavioral therapy; ecological momentary assessment; medical education; mental health; Bayesian analysis; micro-randomized trial

Department of Psychiatry, College of Medicine, King Saud University, Riyadh, Saudi Arabia
ammad.hussain54@gmail.com
– **Published:** 11 October 2024

1. INTRODUCTION

Medical education combines sustained cognitive demand with repeated high-stakes evaluation, abrupt changes in workload, emotionally difficult clinical encounters, and limited control over daily schedules. These demands do not merely raise average distress; they produce rapid within-person changes in distress across ordinary teaching weeks, clinical rotations, and examination periods. Meta-analytic evidence indicates that anxiety, burnout, suicidal ideation, and impaired well-being are prevalent among medical students, although estimates vary with setting, instrument, and stage of training.[2–4] The educational consequences are also material. Psychological strain can weaken concentration, disrupt sleep, narrow help-seeking, and undermine participation in learning activities. A mental-health intervention for this population must therefore address both symptom relief and the capacity to continue studying safely and effectively.

Universities have responded with counseling, wellness curricula, resilience training, mindfulness, and web-based programs. Curricular changes can improve well-being when they alter the learning environment rather than placing responsibility entirely on the student.[5,6] Individual-level programs also show benefit, but pooled effects conceal pronounced variation in use and response. Reviews of university interventions have found that cognitive-behavioral, mindfulness, and skills-based approaches can reduce distress, yet engagement often declines once an initial crisis passes or academic pressure increases.[7,8] This tension is especially important in medicine: the weeks in which support may be most valuable are often the weeks in which attendance at a long session is least feasible.

Digital delivery reduces travel, scheduling, and stigma-related barriers. Meta-analyses of internet and mobile interventions for university populations report improvements in depression, anxiety, and well-being.[9–12] Comparable findings have been reported for internet-delivered psychological treatment and smartphone-supported interventions across adult populations.[13–15] Access alone, however, is insufficient. Digital programs commonly experience steep attrition, and raw registration counts can exaggerate meaningful exposure.[16,17] A fixed sequence also assumes that the same lesson is appropriate regardless of whether today's difficulty is anticipatory anxiety, sleep debt, inactivity, rumination, or an overloaded study plan. Such a sequence can deliver the correct skill at the wrong time.

Digital mental-health support is also a service arrangement, not merely content placed on a screen. Its effect depends on the relationship among the student, the intervention component, the delivery technology, and the human services available when automated support is insufficient. Behavioral intervention technologies have therefore been described in terms of both therapeutic aims and the technical features that enable them.[40] Later critiques emphasized that rapid product cycles, weak comparison conditions, and limited replication can make digital evidence difficult to interpret.[37] The field now includes apps, conversational systems, social platforms, and virtual environments, but breadth of format does not remove the need for clinical governance or comparative evidence.[41] Mobile delivery also challenges theories developed for stable, infrequent encounters because behavior, receptivity, and environmental demand can change within hours.[38] For medical students, a defensible digital intervention should therefore specify the decision point, the information available at that point, the actions that may be selected, the proximal response, and the route to human care. These elements make the policy inspectable and permit a trial to determine why it succeeds or fails.

Context-responsive intervention offers a different proposition. Rather than prescribing an identical weekly unit to every student, it selects a brief action from observations that are proximal to the decision. Just-in-time adaptive interventions were developed to provide the type and intensity of support most likely to help at the moment of need while limiting unnecessary burden.[18] Micro-randomized designs permit repeated randomization among available actions and support estimation of proximal effects that change with context.[19–21] The resulting decision rules extend the logic of dynamic treatment regimes from a few clinical stages to many low-burden decisions distributed over daily life.[22,23] This approach is particularly suitable for medical training because examination proximity, sleep, workload, and distress vary substantially within the same person.

An adaptive policy should be built in stages rather than optimized as an indivisible package. The multiphase optimization strategy separates candidate components, examines their individual and combined contribution, and then evaluates the assembled intervention.[39] Applied here, that logic distinguishes four questions: whether a prompt should be offered, whether the student is available, which skill should be selected, and whether the selected skill produces short-term relief without excessive interruption. A favorable final score cannot identify which of these decisions was responsible. Conversely, a weak component need not invalidate the entire delivery policy if timing or another component remains useful. The computational design preserves these distinctions by storing every prompt occasion, known action probability, immediate response, and daily state transition.

Measurement must be equally responsive. End-point questionnaires can determine whether symptoms changed over a term but cannot identify when deterioration occurred or which action preceded recovery. Ecological momentary assessment reduces reliance on distant recall and captures experience in the settings where it unfolds.[24,25] Short daily observations also allow strain to be represented as a changing state rather than a fixed student characteristic. Repeated measurement introduces its own burden and missingness, so the assessment must be short, interpretable, and linked to a specific decision. Collecting information that does not change support would be difficult to justify.

Personalization also raises methodological and ethical questions. A policy may appear effective because it was used more often in contexts that were already likely to improve. Estimating the value of an alternative allocation therefore requires adjustment for action probability and outcome prediction; doubly robust estimators retain consistency when either component is correctly specified.[27] Campus composition may influence apparent benefit, making site-specific contrasts essential. Missing outcomes may differ systematically by current distress, so sensitivity analysis should determine how severe an unobserved disadvantage would need to be before the conclusion changes.[28,29] Finally, a policy that produces small symptom gains through frequent interruptions may be unacceptable even if the symptom estimate is favorable. Time burden and academic functioning must be treated as outcomes rather than implementation footnotes.

The present study evaluates a context-responsive cognitive-behavioral policy in a fully reproducible computational trial. The research question is: among a multisite synthetic cohort of medical students followed for 90 days, does daily selection among cognitive reframing, paced breathing, behavioral activation, and workload planning produce lower day-90 psychological distress, better academic functioning, fewer high-strain days, and lower intervention time than a fixed weekly cognitive-behavioral course? Three hypotheses were specified. First, the context-responsive condition would yield lower entry-adjusted GHQ-12 equivalent scores than both comparison conditions. Second, its benefit would be accompanied by better academic functioning and shorter residence in a high-strain state. Third, its intervention time would be lower than the fixed course because support would be delivered only on decision occasions. The study is computational rather than clinical; it tests the internal behavior, estimands, and practical demands of the policy before exposing students to an unverified allocation rule.

2. METHODS

2.1. Design and empirical calibration

The study was a 90-day, three-condition computational trial with balanced assignment within six modeled medical schools. It was designed to determine whether a context-responsive allocation rule could outperform fixed delivery under plausible variation in symptoms, workload, and adherence. Published values from a Palestinian medical-student investigation were used once to delimit plausible GHQ-12 entry levels and intervention contrasts. The article's screening, lifestyle, and treatment tables indicated 329 screened students, a 28% prevalence above its mental-health threshold, associations with physical activity, sleep, and entertainment time, and 34 versus 32 analyzed students in the treatment comparison. The GHQ-12 values used for numerical calibration are summarized in Table 1. No person-level observations from that investigation were available to or processed by the present study.

2.2. Synthetic cohort and daily observations

The cohort comprised 720 synthetic students, with 120 assigned to each medical school and 240 assigned to each study condition. Assignment was balanced within school: 40 students per condition at each site. Sex was generated as woman or man with probability 0.58 and 0.42, respectively, and academic year was sampled uniformly from years one through six. Entry GHQ-12 equivalents were sampled from a truncated Gaussian distribution spanning 12 to 34. The mean depended weakly on school, sex, and academic year, creating realistic covariate variation without confounding allocation. Entry academic functioning ranged from 25 to 82 on a 0–100 scale and was inversely related to entry distress.

Each student contributed observations on days 0 through 90. The daily record contained GHQ-12 equivalent distress, a 0–10 momentary distress rating, sleep duration, a 0–10 workload rating, activity minutes, examination proximity, and a high-strain indicator. Examination demand was generated as three smooth peaks centered on days 28, 60, and 84. Sleep decreased with workload and distress, while activity decreased with distress and workload. Autoregressive error introduced temporal continuity, and a person-specific random term created heterogeneity in change. These relations were encoded before condition comparisons were calculated. The resulting dataset contained 65,520 daily records.

The high-strain state was defined by momentary distress above 6.3, sleep below 6 hours, or workload above 7.6. This operational definition intentionally combined affective and functional precursors. A student could move into high strain before the GHQ-12 equivalent reached its maximum, allowing the allocation policy to respond before sustained deterioration. Consecutive-day transitions were retained for the state analysis.

2.3. Study conditions

The context-responsive condition offered four brief cognitive-behavioral components. Cognitive reframing used a concise evidence-for/evidence-against exercise for a current automatic thought. Paced breathing used an eight-minute guided cycle emphasizing a slower exhalation. Behavioral activation selected a feasible 12-minute activity that could be completed between academic commitments. Workload planning divided the next study block into one priority task, one deferrable task, and a stopping rule. The actions were chosen because they target distinct, observable combinations of rumination, arousal, inactivity, and task overload. They did not constitute diagnostic treatment and were paired with a safety notice directing severe or escalating symptoms to quali-

Table 1. Published quantities used for calibration.[1]

Quantity	Group or characteristic	Entry value	Eight-week value
Students screened	Medical students	329; 28% above the stated threshold	
GHQ-12 total	Cognitive-behavioral group ($n = 34$)	20.7 ± 6.4	11.4 ± 6.7
GHQ-12 total	Comparison group ($n = 32$)	22.5 ± 6.4	22.1 ± 6.2
Lifestyle correlates	Physical activity, sleep, entertainment	Associations reported at $p < 0.05$	

Values define plausible ranges for the probability model; no individual record was available or used.

fied care.

A prompt opportunity was generated from current distress and examination proximity. On those occasions, the policy assigned probabilities to the four actions from distress, workload, sleep debt, activity, and examination proximity. Thirty percent of the probability mass remained uniform to preserve exploration; 70% favored the actions with larger predicted proximal relief. Thus, every available action retained a nonzero probability. Immediate relief was recorded on a bounded scale after completion and used only for policy-value estimation, not to regenerate the final symptom outcome.

The fixed-schedule condition received an eight-week cognitive-behavioral course with one nominal 60-minute session per week. It contained the same four component families but presented them in a predetermined sequence. Attendance variation produced a mean exposure below the maximum 480 minutes. The attention-control condition received brief weekly educational material about student well-being and campus support. Its mean exposure was approximately one hour over the study. The three conditions differed in delivery timing and selection logic; the content families available to the two active conditions remained comparable.

The participant-facing sequence is depicted in Figure 1. The visual order is descriptive rather than causal: a brief check-in supplies the state variables, the policy can be inspected by a clinician or trial monitor, and the selected skill is completed in an ordinary campus setting. The absence of an intervention on some days is intentional. A context-responsive design is not a high-frequency messaging system; it is a decision rule that can choose not to interrupt.

2.4. Outcomes and estimands

The primary outcome was day-90 GHQ-12 equivalent score. The primary estimand was the entry-adjusted mean difference between context-responsive and fixed-schedule conditions, averaged over schools, sex, and academic year. A second contrast compared context-responsive support with attention control. Lower values indicated better psychological health. The term “equivalent” distinguishes the generated continuous outcome from responses completed by a person on the validated questionnaire.

Secondary outcomes were day-90 academic functioning, number of days in the high-strain state, dual response, intervention minutes, observed day-90 status, and immediate relief after prompted actions. Dual response required both a reduction of at least five GHQ-12 equivalent points and a day-90 value no higher than 17. This conjunctive definition prevented a large change from being counted when the final value remained high and prevented a low final value from being counted when it merely reflected a low entry value. Intervention minutes were summed from completed actions in the adaptive condition and attended material in the other conditions.

The action-level estimand compared the expected immediate relief under the adaptive probabilities with expected relief under equal probabilities for the four actions, conditional on a prompt having occurred. This estimand concerns action selection, not the effect of sending a prompt versus withholding one. Campus-specific estimands compared change in GHQ-12 equivalents between the two active conditions within each school.

2.5. Statistical analysis

The primary analysis used a Gaussian regression for day-90 GHQ-12 equivalent score with entry GHQ-12, condition, school, sex, and academic year as covariates. Weak normal priors and a scale prior equivalent to the standard conjugate Gaussian model yielded multivariate-t posterior draws. The reported estimate is the posterior median, accompanied by the central 95% credible interval and the posterior probability that the adaptive-minus-fixed contrast was below zero. Academic functioning used the same specification with day-90 functioning as the dependent variable and entry functioning retained through the common design matrix. Bayesian estimates are interpreted as conditional probability statements about the model parameters, not repeated-sampling rejection decisions.[30] Exact estimates and uncertainty intervals were retained instead of converting every comparison into a binary label, consistent with recommendations on interval interpretation.[31,32]

Dual-response risk differences were estimated with 2,500 nonparametric bootstrap draws within condition. State residence was summarized as the mean number of high-strain days per student. The state sequence was examined for condition differences in persistence and recovery, with emphasis on time spent in strain rather than the number of isolated threshold crossings.

For prompted actions, an outcome regression predicted immediate relief from distress, workload, sleep debt, ac-

(a) Daily check-in **(b)** Policy review **(c)** Brief skills practice

Figure 1. Daily study activities. Illustrative depictions; no enrolled participant is shown.

tivity, examination proximity, and action. Known assignment probabilities supplied inverse-probability residual weights. The doubly robust value combined the model-predicted reward under each target policy with a weighted residual correction. Values were calculated for the context-responsive policy and a policy that assigned the four actions equally. Because decision probabilities were bounded away from zero, the comparison did not depend on extreme weights.

School-specific active-condition contrasts were calculated as differences in mean change with normal intervals. The primary model pooled the overall estimate while the school plot displayed heterogeneity without selecting sites by statistical significance. Intervention burden was reported as mean minutes and as minutes per dual responder. Day-90 observation was generated as a function of condition and symptom severity. A pattern-mixture stress test then added 0, 1, 2, or 3 points to every unobserved day-90 value and recalculated the active-condition contrasts. This range asks whether a materially worse unobserved outcome would reverse the direction of the result.

2.6. Reproducibility and ethical status

The complete data-generating program, fixed seed 20260814, participant file, daily file, prompt-occasion file, posterior draws, summary tables, and figure files accompany the manuscript. Re-executing the program regenerates every numerical result and analytical figure. The photographic panels are illustrative and are not analytic evidence. Because the study processes only computer-generated records and includes no identifiable person, recruitment, intervention delivery, consent, and institutional review were not applicable. A future prospective study would require independent ethical review, trial registration, adverse-event procedures, clinician oversight, and an explicit route to urgent care.

3. RESULTS

3.1. Cohort composition and intervention exposure

Allocation produced 240 students per condition and 120 per modeled school. Entry GHQ-12 equivalents were similar across conditions, although the fixed-schedule condition was 0.63 points higher than the context-responsive condition before covariate adjustment. Women constituted 54.17–57.92% of each condition, and mean academic year ranged from 3.44 to 3.55. Entry academic functioning was approximately 51 in all three conditions (Table 2). These small differences demonstrate why the prespecified entry adjustment remains useful even under balanced assignment: finite synthetic samples need not have identical covariate means.

The entry distribution also confirms that the simulated comparison was not driven by a narrow band of nearly identical students. GHQ-12 equivalents had standard deviations between 4.22 and 4.61, academic years covered the full six-year range, and school effects were retained in the generation and analysis. The fixed condition began with slightly greater mean distress and slightly lower mean academic functioning than attention control, whereas the adaptive condition lay between them. Adjustment reduced the influence of those imbalances without changing the substantive comparison. Because allocation was exactly balanced within school, site composition could not create the overall difference merely by placing more adaptive students in a low-distress school.

The adaptive policy generated 5,701 prompt occasions across 240 students, corresponding to 23.75 prompts per student over 90 days. Mean intervention time was 236.95 minutes, or approximately 9.98 minutes per completed prompt. Fixed scheduling required 437.22 minutes on average, 200.27 minutes more than context-responsive delivery. The attention-control material required 62.12 minutes. Day-90 values were observed under the emulated retention process for 92.08%, 85.83%, and 79.17% of the context-responsive, fixed-schedule, and control conditions, respectively.

Table 2. Synthetic cohort at study entry.

Condition	n	Women (%)	Academic year	GHQ-12 equivalent	GHQ-12 SD	Academic function
Context-responsive	240	56.25	3.44	21.38	4.22	50.81
Fixed-schedule	240	57.92	3.54	22.01	4.38	50.51
Attention control	240	54.17	3.55	21.26	4.61	51.29

3.2. Psychological and academic outcomes

Mean day-90 GHQ-12 equivalents were 14.17 (SD 4.51) for context-responsive support, 17.47 (SD 4.60) for fixed scheduling, and 20.78 (SD 4.93) for attention control. Mean within-student changes were -7.20, -4.54, and -0.48 points, respectively (Table 3). After adjustment for entry value, school, sex, and year, the posterior median adaptive-minus-fixed difference was -2.68 points (95% credible interval -3.01 to -2.34). Every retained posterior draw favored the adaptive condition at the recorded precision, giving a posterior probability above 0.999. The corresponding difference from attention control was -6.72 points (-7.05 to -6.38), while fixed scheduling differed from attention control by -4.04 points (-4.38 to -3.70).

The 90-day paths in Figure 2 show that separation emerged during the first three weeks and widened after each examination peak. The control path rose near days 28, 60, and 84 and returned only partially toward its preceding level. Fixed scheduling attenuated these peaks, whereas context-responsive delivery combined a lower underlying trajectory with smaller examination-related rebounds. This pattern matters because an end-point contrast alone could be produced by a late change. Here, the advantage was sustained across most of the observation period and was not confined to day 90.

The distributional display in Figure 3 qualifies the mean comparison. Considerable overlap remained between context-responsive and fixed scheduling, and some adaptive-condition students finished with high values. The benefit therefore represents a shift in the whole distribution rather than universal recovery. The adaptive distribution had a lower median and reduced upper-tail mass; attention control retained the largest concentration near 20–24. This heterogeneity supports retaining clinician access and discourages interpretation of the policy as a substitute for treatment when symptoms are severe.

Academic functioning increased by 7.19 points under context-responsive support, 4.75 under fixed scheduling, and 0.63 under attention control. The entry-adjusted posterior adaptive-minus-fixed difference was 2.00 points (0.77 to 3.24), and the adaptive-minus-control difference was 6.22 points (5.00 to 7.46). Symptom improvement was therefore accompanied by functional improvement rather than achieved through withdrawal from academic activity. Dual response occurred in 67.92% of the adaptive condition and 22.50% of the fixed condition. The bootstrap risk difference was 45.42 percentage points (37.08 to 52.92); the difference from attention control was 67.92 points (61.67 to 73.75). Because the dual definition combines change and final status, these proportions are more demanding than reporting any symptom decrease.

3.3. State residence and action selection

Students in the context-responsive condition spent a mean 49.47 days in the high-strain state, compared with 59.23 under fixed scheduling and 66.20 under attention control. The 9.75-day adaptive advantage over fixed scheduling represents fewer persistent days rather than merely a lower day-90 value. It also explains part of the academic-functioning difference: more days were available for ordinary study without simultaneously crossing the strain definition.

Action relief varied across the four daily states (Figure 4). Planning produced the largest mean relief near examinations, reaching 1.07 when rested and 1.12 with sleep debt. Paced breathing was most useful when sleep debt accompanied an approaching examination (0.94), but offered only 0.42 during rested routine periods. Reframing was comparatively stable (0.62–0.70), while activation showed smaller state variation. These differences supply the substantive reason for adaptation. Equal assignment would spend too many opportunities on actions whose proximal value is weak in the current state.

The doubly robust estimated immediate-relief value was 0.695 under the adaptive probabilities and 0.672 under equal action probabilities, a difference of 0.023 per prompted occasion. This action-level gain is modest compared with the final symptom contrast. Its interpretation is therefore specific: action matching improved proximal relief, while the full 90-day difference also reflects prompt timing, reduced interruption, examination-period attenuation, and cumulative state transitions. It would be incorrect to attribute the complete final difference to action selection alone.

3.4. Transport, burden, and sensitivity

Every modeled school favored context-responsive delivery over fixed scheduling (Figure 5). The mean change contrasts ranged from -2.00 in School E to -3.40 in School F, with all intervals below zero. The variation is meaningful despite a consistent direction. A future trial should allow treatment effects to differ by school rather than assuming that scheduling, clinical exposure, and support availability are interchangeable across institutions.

Table 3. Day-90 outcomes and intervention burden.

Condition	GHQ-12 mean (SD)	GHQ-12 change	Function change	Dual response (%)	Minutes	Observed (%)
Context-responsive	14.17 (4.51)	−7.20	7.19	67.92	236.95	92.08
Fixed-schedule	17.47 (4.60)	−4.54	4.75	22.50	437.22	85.83
Attention control	20.78 (4.93)	−0.48	0.63	0.00	62.12	79.17

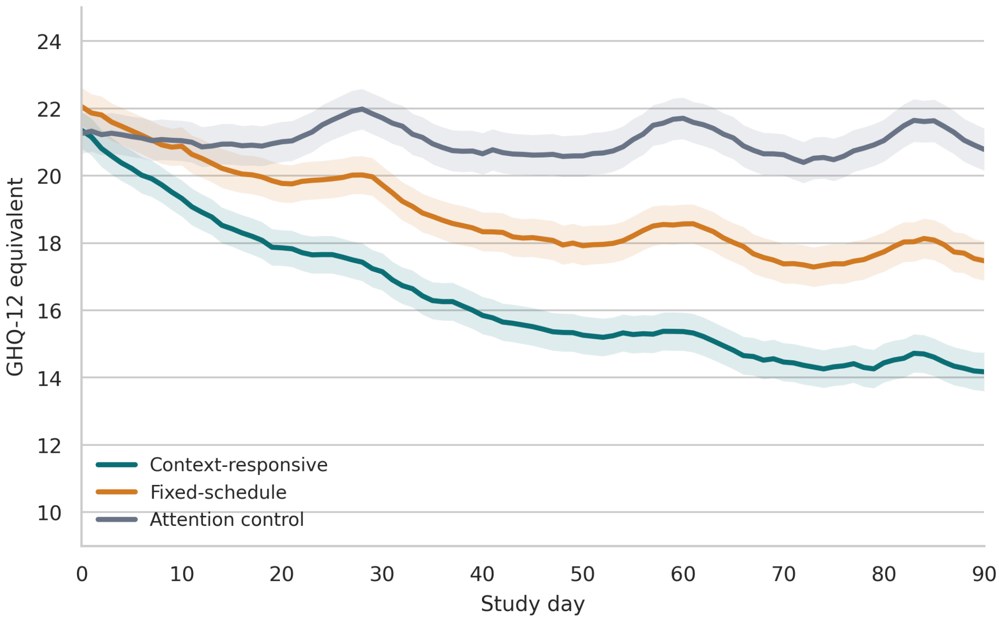

Figure 2. Ninety-day symptom trajectories.

Time accounting reinforces the symptom result. Context-responsive support used 54.19% of the fixed course's intervention minutes. Dividing total condition minutes by the number of dual responders produced approximately 349 minutes per responder in the adaptive condition and 1,944 minutes per responder in the fixed condition. This ratio is not an economic evaluation because staff preparation, software, supervision, and escalation are omitted. It nevertheless shows that the adaptive benefit was not purchased through greater participant exposure.

The observation proportions require a second reading of burden. The adaptive condition combined the fewest active-treatment minutes with the highest day-90 observation rate. In the probability model, lower current distress increased the probability of remaining observed, so improved trajectories contributed to retention. A real trial could show a different relation: frequent notifications may reduce completion even when symptoms improve, or students with severe symptoms may remain engaged because support feels immediately relevant. For this reason, prompt completion, daily-question completion, and final-outcome completion should be reported separately. Treating them as a single adherence percentage would conceal whether disengagement arose from the intervention, measurement schedule, or follow-up procedure.

The pattern-mixture stress test did not approach a reversal (Table 4). Adding three adverse points to every unobserved day-90 value yielded an unadjusted adaptive-minus-fixed difference of -3.48 and an adaptive-minus-control difference of -6.99. The increasingly negative contrast occurs because the generated retention process left more unobserved outcomes in the two comparison conditions. The important finding is not that missingness improves the adaptive estimate; it is that the direction remains stable under a severe, explicitly adverse shift applied to all unobserved values.

4. DISCUSSION

This computational trial answers its research question in the affirmative under the encoded probability model: selecting a brief cognitive-behavioral skill from current distress, sleep debt, activity, workload, and examination proximity produced lower day-90 GHQ-12 equivalents, better academic functioning, fewer high-strain days, and substantially lower intervention time than a fixed weekly course. The primary adjusted difference of -2.68 points was not an isolated end-point effect. Separation appeared early, persisted for most of the 90 days, and was visible across all six modeled schools. The adaptive condition also used approximately 200 fewer minutes, indicating that temporal precision can be more conse-

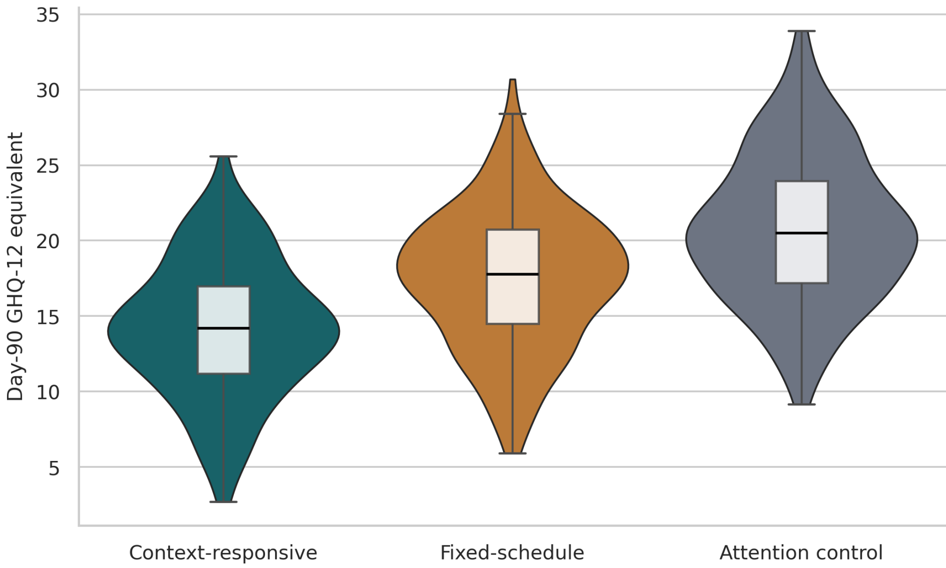

Figure 3. Day-90 symptom distributions.

Table 4. Policy value and missing-outcome stress test.

Analysis quantity	Estimate	Comparator
Doubly robust immediate-relief value	0.695	Adaptive probabilities
Doubly robust immediate-relief value	0.672	Equal action probabilities
Unobserved-outcome shift: 0 points	−3.30	Adaptive minus fixed
Unobserved-outcome shift: 1 point	−3.36	Adaptive minus fixed
Unobserved-outcome shift: 2 points	−3.42	Adaptive minus fixed
Unobserved-outcome shift: 3 points	−3.48	Adaptive minus fixed

quential than simply increasing contact time.

The computational status of the evidence determines what may legitimately be concluded. The study establishes logical coherence: the declared decision rule, longitudinal relations, estimands, and analysis produce internally consistent outputs and recover a benefit that is visible across complementary outcomes. It does not estimate how real students will respond. That distinction is productive rather than merely cautionary. A computational trial can reveal before recruitment whether the planned files are sufficient for each estimand, whether decision probabilities remain positive, whether burden is recorded, whether the site contrast is identifiable, and whether the missing-outcome analysis can be reproduced. It can also expose exaggerated claims. For example, the heat map demonstrates action matching because those state-action relations were encoded; it cannot be cited as empirical proof that planning is superior near examinations. The appropriate contribution is a testable design with quantitative operating characteristics.

The result is compatible with the established efficacy of cognitive-behavioral and digital interventions without assuming that digital delivery is automatically adaptive. Internet programs can reduce distress among students, but many retain a linear curriculum and rely on the participant to choose when to log in.[9,11] Smartphone interventions can increase availability while still delivering poorly timed notifications.[33] The present policy changes the unit of intervention from a course module to a decision occasion. It asks which brief action is indicated now and allows no prompt when expected benefit does not justify interruption.

Three mechanisms explain the modeled advantage. First, action specificity mattered. Planning was most useful when examination proximity amplified workload, whereas breathing was most useful when that pressure co-occurred with sleep debt. Equal assignment diluted these state-action matches. Second, timing mattered. Examination peaks were anticipated by the daily state variables, so support was concentrated near periods of rising strain rather than limited to a predetermined weekday. Third, burden mattered. A ten-minute action is easier to place between teaching activities than a full session, and reduced burden preserved exposure during demanding weeks. These mechanisms correspond to the just-in-time principle of intervening at states of both vulnerability and receptivity.[18]

The modest difference between adaptive and equal

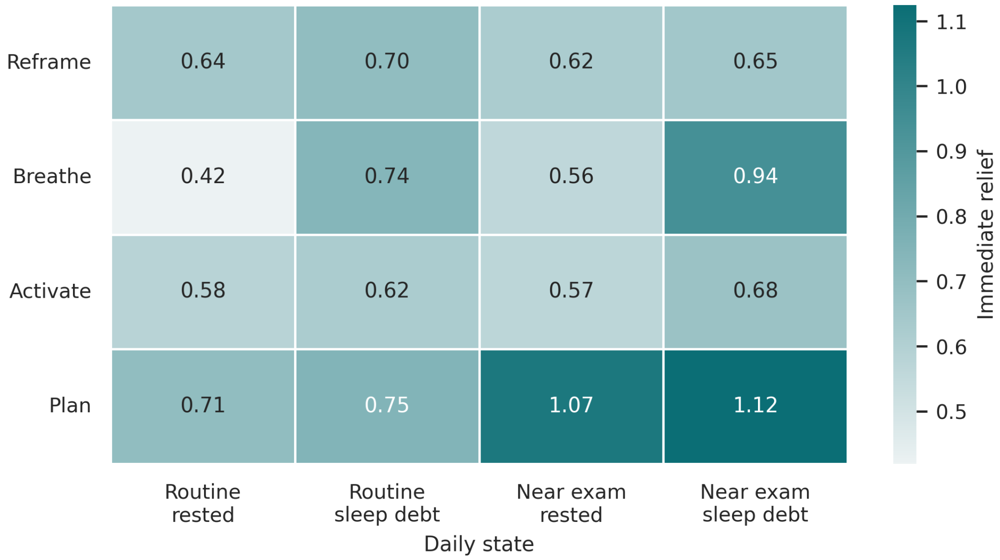

Figure 4. Action relief across daily states.

action probabilities deserves careful attention. An immediate-relief advantage of 0.023 per prompt is not a dramatic single-occasion effect. Repetition across approximately 24 occasions may accumulate, but the final symptom difference also includes modeled timing and state persistence. This distinction is important for a prospective trial. Repeated randomization should separately estimate the effect of sending support, the comparative effects of the four skills, and moderation by current state. A single end-of-term questionnaire cannot recover those components. Micro-randomization is well suited to this task because it preserves known action probabilities at every decision.[19,20]

The functional outcome strengthens the interpretation. Symptom scores can improve when a student withdraws temporarily from stressful activities; such improvement may be clinically reasonable but does not by itself demonstrate educational benefit. Here, academic functioning increased alongside symptom improvement, with a 2.00-point adjusted advantage over fixed scheduling. The state analysis provides a temporal bridge: adaptive students spent nearly ten fewer days above the strain threshold, leaving more days in which ordinary academic participation was compatible with lower distress. A prospective study should measure attendance, assessment completion, and self-regulated learning alongside symptoms, while avoiding punitive use of mental-health information.

The school-specific estimates address an important transport question. Medical schools differ in timetable density, assessment frequency, clinical exposure, commuting, digital access, and counseling capacity. Averaging over these conditions can conceal a policy that works only where support is already strong. All six modeled schools favored adaptation, but effect sizes differed by 1.4 points. A real multisite trial should use partial pooling and report site distributions rather than a pooled estimate alone. It should also evaluate whether the policy performs equitably across sex, training year, socioeconomic position, disability, and connectivity. Digital health can widen disparities when access, privacy, or data cost is uneven.[34] Personalization is not equitable merely because it is individualized.

The burden result has practical significance. Fixed courses typically require coordination among students, facilitators, rooms, and timetables. Context-responsive micro-interventions distribute shorter actions across daily life, but they introduce software maintenance, policy monitoring, and safety review. Participant minutes therefore favor the adaptive condition, whereas institutional cost remains unknown. Economic evaluation should include clinician oversight, technical support, false alerts, and the opportunity cost of prompts. It should also examine whether a lower dose weakens therapeutic alliance or prevents deeper cognitive work for students who need it. Brief support can complement, but should not displace, accessible counseling and psychiatric care.

Help-seeking barriers remain relevant even when support is delivered privately. Students may fear stigma, confidentiality breaches, or consequences for professional progression. Digital access can reduce the effort of initiating support, yet trust depends on clear data governance and separation from academic evaluation. Reviews of student help-seeking show that perceived stigma and preference for self-reliance are persistent barriers.[35,36] A prospective implementation should state who can view daily responses, how long data are retained, what triggers clinician review, and whether faculty can access individual information. Safety escalation should be visible before the first check-in, not introduced only after a high-risk response.

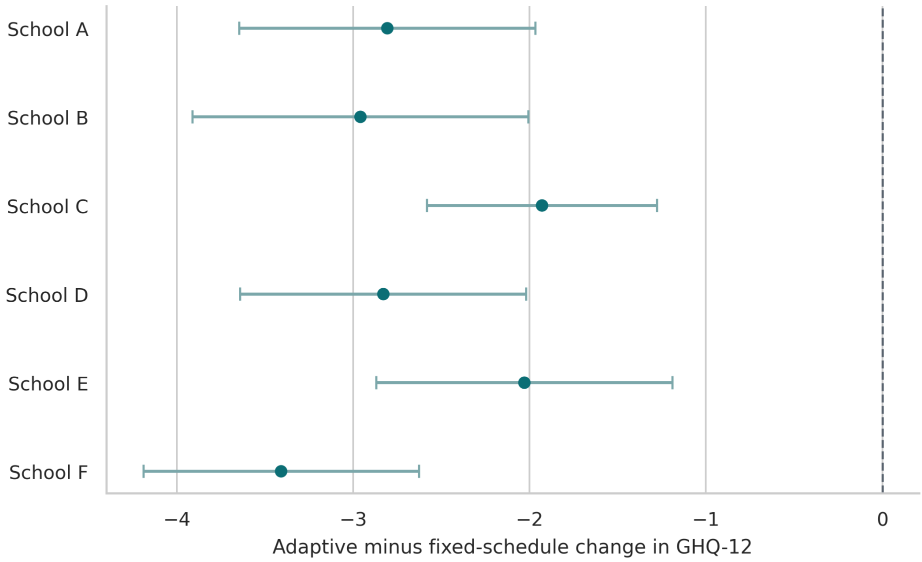

Figure 5. Active-condition contrasts by school.

The analytical choices also have implications for trial design. The Bayesian model provided a direct probability that the adaptive condition reduced day-90 distress and retained the size and uncertainty of the effect. The doubly robust policy-value analysis separated action choice from outcome prediction and remained protected if either the propensity or outcome component was correctly specified. Neither method rescues poor measurement. Daily questions must be brief enough to sustain participation, sensitive enough to represent change, and interpretable across languages and schools. Ambulatory assessment research emphasizes that observations close to experience can reduce recall error, but reactivity and missingness remain possible.[24,26]

The missing-outcome stress test was deliberately severe: up to three adverse points were added to every unobserved value. The conclusion remained unchanged. This result should not be translated into a claim that missingness is harmless. The test operates within a known computational model, whereas real disengagement may signal worsening symptoms, privacy concerns, technical failure, or dissatisfaction. Multiple imputation can propagate uncertainty under a specified missing-at-random model, but departures from that assumption still require transparent stress tests.[28] Reporting only complete cases would be inadequate if observation depends on symptom severity.

Several limitations define the boundary of inference. Most importantly, all students and outcomes are synthetic. The estimates demonstrate the behavior of a decision policy under declared assumptions; they do not establish clinical efficacy, acceptability, or safety. The GHQ-12 equivalent was generated as a continuous trajectory and cannot reproduce every psychometric property of completed questionnaire items. Immediate relief was also generated from state-action relations specified in code, so the heat map validates recovery of those relations rather than discovering a human mechanism. The three examination peaks simplify the irregular scheduling of actual programs. Peer relationships, financial strain, exposure to patient suffering, substance use, prior treatment, and acute risk were not generated. Intervention content was represented by action categories and duration, not therapist competence or therapeutic alliance.

Further limitations concern the decision rule. The policy used a restricted set of contemporaneous variables and did not incorporate free text, passive sensing, geolocation, or social-media data. This restraint reduces privacy risk and improves interpretability but may omit useful context. The high-strain definition used fixed thresholds, which may not correspond to an individual's meaningful change. Prompt availability was generated rather than chosen by a student, and notification fatigue was represented only indirectly through intervention time. The Gaussian final-outcome model can accommodate the generated distribution but may be less suitable if a future trial has floor effects, heavy tails, or distinct symptom classes. Finally, the campus contrasts were based on 40 students per active condition at each school; a prospective sample-size calculation should be based on the proximal and final estimands, within-person correlation, availability rate, and clustering.

Measurement validity will be a central concern in that future trial. The present GHQ-12 equivalent is useful because it gives the longitudinal model a scale that can be compared with published ranges, but daily administration of the full questionnaire would be burdensome and could change responding. A short momentary item set should therefore be developed independently of the final questionnaire, then linked to intervention decisions through prespecified thresholds or a continuous score.

Translation and cultural adaptation should test meaning rather than literal wording. Academic functioning also requires a transparent definition: self-rated ability to study, attendance, completed assessments, and objective grades answer different questions and carry different privacy consequences. The preferred primary functional outcome should be chosen with students before registration, not selected after examining which one favors the intervention.

A prospective evaluation should begin with a feasibility phase that tests check-in completion, prompt acceptability, technical reliability, and safety escalation. It should then use a multisite micro-randomized design with equal minimum probabilities for all actions, independent oversight, and a preregistered primary end point. The fixed-schedule comparator should contain equivalent content and credible facilitator contact so that timing and selection remain the principal differences. Qualitative interviews should examine whether students experience the prompts as supportive, intrusive, culturally appropriate, or academically risky. The policy should be frozen for the confirmatory analysis; any later learning should occur in a separately declared training period. Reporting should distinguish the effect of availability, the effect of a prompt, the effect of action choice, and the effect on final symptoms.

5. CONCLUSION

The study asked whether context-responsive selection of brief cognitive-behavioral skills could outperform a fixed weekly course while using less student time. Under the fully specified 90-day computational model, it did: the adaptive policy lowered entry-adjusted day-90 GHQ-12 equivalents by 2.68 points relative to fixed scheduling, improved academic functioning by 2.00 points, reduced high-strain residence by 9.75 days, and required approximately 200 fewer minutes. The advantage was present in every modeled school and was not reversed by the missing-outcome stress test. These findings establish a coherent and reproducible trial rationale, not evidence for immediate clinical adoption. The next evidentiary step is a prospectively registered, ethically reviewed, multisite micro-randomized trial that measures safety, acceptability, equity, symptoms, academic functioning, and total delivery cost.

BIBLIOGRAFIA / REFERENCES

1. Hanani A, Badrasawi M, Zidan S, Hunjul M. Effect of cognitive behavioral therapy program on mental health status among medical student in Palestine during COVID pandemic. *BMC Psychiatry*. 2022;22:310. doi:10.1186/s12888-022-03915-1.
2. Quek TTC, Tam WWS, Tran BX, et al. The global prevalence of anxiety among medical students: a meta-analysis. *Int J Environ Res Public Health*. 2019;16:2735. doi:10.3390/ijerph16152735.
3. Frajerman A, Morvan Y, Krebs MO, Gorwood P, Chaumette B. Burnout in medical students before residency: a systematic review and meta-analysis. *Eur Psychiatry*. 2019;55:36–42. doi:10.1016/j.eurpsy.2018.08.006.
4. Seo C, Di Carlo C, Dong SX, Fournier K, Haykal KA. Risk factors for suicidal ideation and suicide attempt among medical students: a meta-analysis. *PLoS One*. 2021;16:e0261785. doi:10.1371/journal.pone.0261785.
5. Slavin SJ, Schindler DL, Chibnall JT. Medical student mental health 3.0: improving student wellness through curricular changes. *Acad Med*. 2014;89:573–577. doi:10.1097/ACM.0000000000000166.
6. Wasson LT, Cusmano A, Meli L, et al. Association between learning environment interventions and medical student well-being: a systematic review. *JAMA*. 2016;316:2237–2252. doi:10.1001/jama.2016.17573.
7. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: a review and meta-analysis. *J Affect Disord*. 2013;148:1–11. doi:10.1016/j.jad.2012.11.026.
8. Conley CS, Durlak JA, Kirsch AC. A meta-analysis of universal mental health prevention programs for higher education students. *Prev Sci*. 2015;16:487–507. doi:10.1007/s11121-015-0543-1.
9. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. *J Med Internet Res*. 2014;16:e130. doi:10.2196/jmir.3142.
10. Farrer L, Gulliver A, Chan JKY, et al. Technology-based interventions for mental health in tertiary students: systematic review. *J Med Internet Res*. 2013;15:e101. doi:10.2196/jmir.2639.
11. Harrer M, Adam SH, Baumeister H, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. *Int J Methods Psychiatr Res*. 2019;28:e1759. doi:10.1002/mpr.1759.
12. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. *J Med Internet Res*. 2019;21:e12869. doi:10.2196/12869.
13. Andersson G. Internet-delivered psychological treatments. *Annu Rev Clin Psychol*. 2016;12:157–179. doi:10.1146/annurev-clinpsy-021815-093006.
14. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry*. 2017;16:287–298. doi:10.1002/wps.20472.
15. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. *World Psychiatry*. 2019;18:325–336. doi:10.1002/wps.20673.
16. Eysenbach G. The law of attrition. *J Med Internet Res*.

2005;7:e11. doi:10.2196/jmir.7.1.e11.

17. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *J Med Internet Res*. 2019;21:e14567. doi:10.2196/14567.
18. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions in mobile health: key components and design principles for ongoing health behavior support. *Ann Behav Med*. 2018;52:446–462. doi:10.1007/s12160-016-9830-8.
19. Klasnja P, Hekler EB, Shiffman S, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. *Health Psychol*. 2015;34S:1220–1228. doi:10.1037/hea0000305.
20. Dempsey W, Liao P, Klasnja P, Nahum-Shani I, Murphy SA. Randomised trials for the Fitbit generation. *Significance*. 2015;12:20–23. doi:10.1111/j.1740-9713.2015.00841.x.
21. Liao P, Klasnja P, Tewari A, Murphy SA. Sample size calculations for micro-randomized trials in mHealth. *Stat Med*. 2016;35:1944–1971. doi:10.1002/sim.6847.
22. Murphy SA. Optimal dynamic treatment regimes. *J R Stat Soc Series B Stat Methodol*. 2003;65:331–355. doi:10.1111/1467-9868.00389.
23. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A "SMART" design for building individualized treatment sequences. *Annu Rev Clin Psychol*. 2012;8:21–48. doi:10.1146/annurev-clinpsy-032511-143152.
24. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol*. 2008;4:1–32. doi:10.1146/annurev.clinpsy.3.022806.091415.
25. Trull TJ, Ebner-Priemer U. Ambulatory assessment. *Annu Rev Clin Psychol*. 2013;9:151–176. doi:10.1146/annurev-clinpsy-050212-185510.
26. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. *J Happiness Stud*. 2003;4:35–52. doi:10.1023/A:1023657221954.
27. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics*. 2005;61:962–973. doi:10.1111/j.1541-0420.2005.00377.x.
28. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30:377–399. doi:10.1002/sim.4067.
29. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1–67. doi:10.18637/jss.v045.i03.
30. Kruschke JK, Liddell TM. Bayesian data analysis for newcomers. *Psychon Bull Rev*. 2018;25:155–177. doi:10.3758/s13423-017-1272-1.
31. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 2016;31:337–350. doi:10.1007/s10654-016-0149-3.
32. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat*. 2016;70:129–133. doi:10.1080/00031305.2016.1154108.
33. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. *J Med Internet Res*. 2013;15:e247. doi:10.2196/jmir.2791.
34. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. *J Am Med Inform Assoc*. 2018;25:1080–1088. doi:10.1093/jamia/ocy052.
35. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry*. 2010;10:113. doi:10.1186/1471-244X-10-113.
36. Eisenberg D, Golberstein E, Gollust SE. Help-seeking and access to mental health care in a university student population. *Med Care*. 2007;45:594–601. doi:10.1097/MLR.0b013e31803bb4c1.
37. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research and three things we can do about them. *Psychiatr Serv*. 2017;68:427–429. doi:10.1176/appi.ps.201600541.
38. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? *Transl Behav Med*. 2011;1:53–71. doi:10.1007/s13142-011-0021-7.
39. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy for engineering effective tobacco use interventions. *Ann Behav Med*. 2007;34:65–73. doi:10.1007/BF02879923.
40. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. *Gen Hosp Psychiatry*. 2013;35:332–338. doi:10.1016/j.genhosppsych.2013.03.008.
41. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*. 2021;20:318–335. doi:10.1002/wps.20883.