World Journal of Biology Pharmacy and Health Sciences
eISSN: 2582-5542
Cross Ref DOI: 10.30574/wjbphs
Journal homepage: https://wjbphs.com/

(RESEARCH ARTICLE)

# CONTINUED USE OF A SMARTWATCH-BASED SLEEP COMPANION AND TEN-MONTH DEPRESSION OUTCOMES: AN OBSERVATIONAL FOLLOW-UP ANALYSIS

Zaki Milhem [1] and Roland Stretea [2,*]

[1] WAKEZ HEALTH TECH ECOSYSTEM SRL, Cluj-Napoca, 400686, Romania.
[2] Department of Neurosciences, Psychiatry and Pediatric Psychiatry, "Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca, 400012, Romania.
* Corresponding Author

ORCID Details
Zaki Milhem: https://orcid.org/0000-0003-1898-3685
Roland Stretea: https://orcid.org/0000-0002-9433-3118

World Journal of Biology Pharmacy and Health Sciences, 2026, 28(02), 127–143

Publication history: Received on 02 October 2026; revised on 08 November 2026; accepted on 10 November 2026

Article DOI: https://doi.org/10.30574/wjbphs.2026.28.2.0541

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

## ABSTRACT

**Background:** Depression management is a public-health priority; disturbed sleep is a modifiable risk factor, and low-burden smartwatch sleep tools could support prevention at scale.

**Objective:** To assess whether continued use of WakeZ (smart alarm plus self-reflection) after a randomized trial was associated with new major depressive disorder (MDD) episodes and Montgomery–Åsberg Depression Rating Scale (MADRS) worsening, in the larger context of prevention of depression.

**Methods:** Observational 10-month follow-up of 103 trial completers (56 continued WakeZ, 47 discontinued; not randomized). New episodes were analysed in 49 participants without a current episode and worsening (MADRS increase ≥5) in all, using Fisher's exact tests (Holm-corrected) and exploratory adjusted models.

**Results:** New episodes occurred in 7/25 (28.0%) continuing users versus 15/24 (62.5%) discontinuers (risk ratio [RR] 0.45, 95% confidence interval 0.22–0.90; Holm p = 0.022; Firth-adjusted odds ratio 0.10, 0.02–0.43). Worsening occurred in 12/56 (21.4%) versus 22/47 (46.8%) (RR 0.46, 0.25–0.82; Holm p = 0.022), but not significantly among those with a current episode at baseline (p = 0.245). Current MDD at month 10 was 51.8% versus 68.1% (p = 0.110).

**Conclusions:** Continued use of a low-burden smartwatch sleep companion after symptomatic improvement was associated with less than half the risk of a new depressive episode, strongly hinting at depression prevention capabilities. These findings warrant a larger, randomized prevention trial, in the context of depression prevention.

**Keywords:** Depression Prevention; Major Depressive Disorder; Sleep; Wearable Electronic Devices; Digital Health; Observational Study

## 1 INTRODUCTION

Depressive disorders are a leading cause of disability worldwide. In the Global Burden of Disease 2019 analysis, mental disorders accounted for 125.3 million disability-adjusted life-years, with depressive disorders the largest single contributor [1]. In the United States, the incremental economic burden of adults with major depressive disorder (MDD) was estimated at US$326.2 billion in 2018, 61% of which arose from workplace costs [2]. MDD is common across cultures and frequently recurrent [3], and people with subthreshold depressive symptoms have a higher incidence of MDD than those without [4]. The period after symptomatic improvement is therefore a plausible window for prevention. Psychological prevention programmes, mostly based on cognitive behavioural therapy, reduce the one-year risk of onset by about 19% [5], but they are structured, guided interventions whose effects cannot be assumed for lower-intensity self-management tools.

Sleep disturbance is one of the most consistent prospective correlates of depression. Non-depressed people with insomnia have roughly twice the odds of developing depression [6, 7], and both short and long sleep duration have been associated with later depression [8]. The relationship is probably bidirectional [9]. Beyond duration, sleep regularity has attracted interest: among 2115 medical interns, wearable-measured variability in total sleep time and wake time was associated with more depressive symptoms [10], and reduced circadian rhythm amplitude was associated with lifetime MDD in 91,105 UK Biobank participants [11]. Depression is also associated with altered sleep architecture, including shortened rapid eye movement (REM) latency and increased REM density, and these features have been proposed as markers that may precede onset [12, 13]. Whether modifying REM sleep or the manner of awakening changes depression risk remains a hypothesis rather than an established mechanism.

Consumer smartwatches allow passive sleep estimation at scale, but with important limits. Against polysomnography, consumer devices detect sleep with high sensitivity but identify wakefulness less reliably (specificity 0.18–0.54), and their sleep-stage classification is inconsistent [14, 15]. Features that depend on stage estimates, such as a smart alarm intended to wake the user from lighter sleep to limit sleep inertia [16], therefore rest on imprecise inputs, and intense attention to device metrics can itself increase sleep-related anxiety [17]. Sustained engagement is a further challenge: median 30-day retention of mental-health apps is about 3% [18], and engagement typically declines quickly [19]. On the other hand, structured mood self-monitoring increased emotional self-awareness and reduced depressive symptoms in young people [20], smartphone interventions show modest symptom benefits overall [21], and sleep features from consumer wristbands are associated with depressive symptom severity [22].

WakeZ is a smartwatch-based sleep companion. In the parent randomized trial, participants were assigned to a REM-targeted haptic intervention (Test) or to sleep monitoring alone (Control) [23]. After the trial, participants could continue a consumer configuration of WakeZ, comprising a smart alarm and a structured self-reflection feature. Structured digital cognitive behavioural therapy for insomnia (CBT-I) has reduced depressive symptoms and, in some trials, incident depression [24–26], but those are therapeutic programmes rather than self-management companions. A tool that takes little time each day could, if acceptable, be used for long periods by people whose symptoms have recently improved. Whether continued use of such a tool is associated with later depression outcomes is relevant to deciding whether a prospective preventive trial is warranted. This follow-up is a first step in that direction: sustained, low-burden self-management is particularly relevant to the prevention of new depressive episodes, because preventive programmes that require repeated therapist contact are difficult to deliver at population scale.

This observational follow-up addressed two questions. First, among participants without a current depressive episode at follow-up baseline, was continued WakeZ use associated with fewer new major depressive episodes over 10 months than discontinuation? Second, across all participants, was continued use associated with a lower risk of clinically meaningful worsening of depressive symptoms and with lower month-10 MADRS scores? Both questions concern associations; neither tests the efficacy of WakeZ.

## 2 MATERIALS AND METHODS

### 2.1 Study design

This was an observational 10-month follow-up of participants who had completed the parent randomized trial of WakeZ, conducted in Cluj-Napoca, Romania, between September and November 2025 and retrospectively registered at ClinicalTrials.gov (NCT07891806; first posted 23 October 2026) [23]. Its primary results have been submitted for publication and are currently under review. The parent trial randomized 208 participants to the Test or Control arm. The present study is separate from that trial: the follow-up exposure groups were defined by participants' own continuation behaviour after the trial and were not randomized. Reporting follows the STROBE guideline for observational studies [27].

### 2.2 Participants

Of the 208 randomized participants, 140 completed the end-of-trial MADRS. All 140 were contacted at month 10, and the 103 who responded form the analysis cohort; the 37 non-responders contributed no follow-up data (Figure 1). Participants were contacted through the e-mail addresses and telephone numbers they had provided in the parent trial, and month-10 interviews were held either in person or in secure one-to-one online sessions. Responders and non-responders were compared on the parent-trial variables available for everyone (arm, entry MADRS and end-of-trial MADRS). Each follow-up record was linked to its parent-trial record through a unique parent identifier.

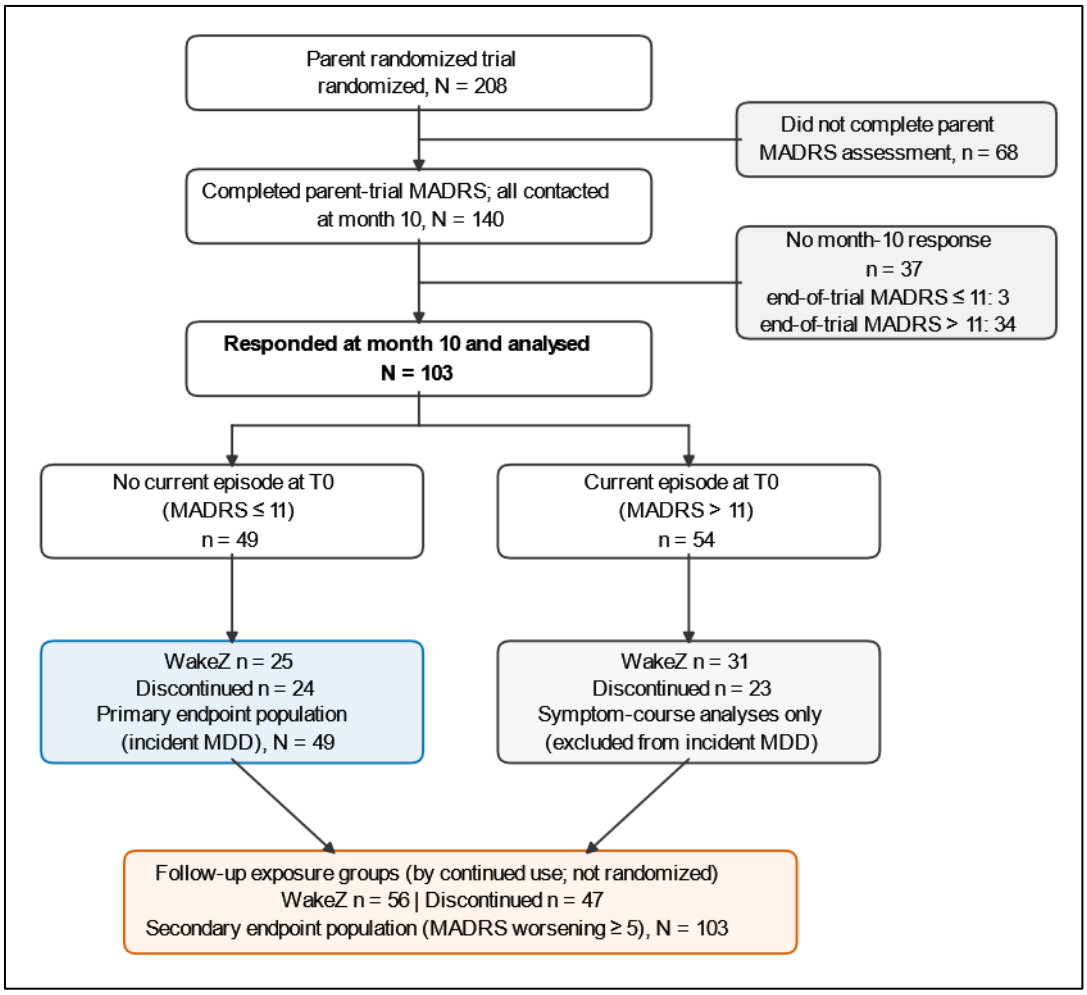



**Figure 1: Participant flow from the parent randomized trial to the 10-month follow-up. All 140 completers of the parent-trial MADRS were contacted at month 10, and the 103 responders were analysed. The follow-up exposure groups were formed by participants' continuation behaviour and were not randomized. The number randomized (N = 208) is taken from the parent trial. Current episode status was clinically reassessed at T0; the MADRS threshold defined the accompanying symptomatic-remission criterion. MADRS, Montgomery–Åsberg Depression Rating Scale; MDD, major depressive disorder; T0, follow-up baseline.**

### 2.3 Exposure

The exposure was the follow-up group. WakeZ denoted continued use of the system composed of the smart alarm and the self-reflection feature during follow-up; Discontinued denoted that use was not maintained. Continuation was ascertained from WakeZ logs, comprising recorded sleep data and completed surveys. In the analysis extract the logs contributed only this binary classification; individual-level sleep and feature-use metrics were not included, so usage frequency, duration and the timing of changes in engagement could not be reconstructed. The comparison category denotes participants who did not maintain the combined use of both features; it does not necessarily imply complete discontinuation of every WakeZ function. The current WakeZ configuration and module are distinct from the parent trial's REM-haptic research intervention, and parent arm was recorded as a covariate.

### 2.4 Timepoints and measures

Follow-up baseline (T0) was the parent end-of-trial assessment, and the T0 MADRS was the parent final MADRS (identical in all 103 records). Depressive symptoms were reassessed at month 10. The MADRS has 10 items scored 0–6

(total 0–60; higher scores indicate greater severity) [28]; conventional severity grades are 0–6 symptom-free, 7–19 mild, 20–34 moderate and ≥35 severe [29]. MADRS change was defined as the month-10 score minus the T0 score, so positive values indicate worsening.

At the conclusion of the parent interventional trial, participants underwent a clinical reassessment comprising clinician-administered MADRS scoring and a separate clinical evaluation of current major depressive episode status. Symptomatic remission was operationally defined as a MADRS total score of ≤11, consistent with published remission thresholds [30, 31], together with clinical confirmation that no major depressive episode was ongoing. This end-of-trial evaluation constituted baseline (T0) for the present observational follow-up. Participants without a current episode at T0 were eligible for the incident-episode analysis, irrespective of their previous depression history; those with an ongoing episode were included in analyses of symptom course.

At month 10, participants underwent another clinical interview by the research team, held in person or online, assessing current depressive episode status and the occurrence of episodes during the intervening period. MADRS was administered to quantify depressive symptom severity. Current episode status was determined by clinical assessment rather than by a MADRS score alone.

### 2.5 Outcomes

The primary outcome was a new major depressive episode (incident MDD) during the 10-month follow-up among participants without an ongoing episode at T0. Episode occurrence was ascertained retrospectively at the month-10 clinical interview, including episodes that had resolved before the follow-up visit. A participant who had an ongoing episode at month 10 was counted as an interval case if episode-free at T0; those with a clinically established episode at T0 were ineligible for the incident-episode analysis. Current MDD status at month 10 was recorded separately from interval episode occurrence.

The key secondary outcome was clinically meaningful symptom worsening, defined as a MADRS increase of at least 5 points. Published estimates of the minimal clinically important difference on the MADRS range from about 2 to 6 points depending on method and population [32, 33]. Further outcomes were month-10 MADRS (analysed with adjustment for T0), MADRS change, and clinically assessed current MDD at month 10 (exploratory). Current MDD at month 10 differs from incident MDD because a participant may develop an episode and recover before the month-10 visit.

### 2.6 Data validation and missing data

The analysis worksheet was imported as data, with identifiers kept as text and all other fields converted explicitly to numbers. Before analysis we checked record and identifier uniqueness, group and stratum sizes, the derivation of MADRS change and of the worsening indicator, value ranges (MADRS 0–60; binary fields 0/1), and linkage to the parent source for arm, entry MADRS and final MADRS. Incident MDD was the only variable with missing values: it was blank, by design, for the 54 participants with a current episode at T0, who were excluded from the incident-episode denominator rather than counted as non-events. No values were imputed. Clinician-recorded current episode status and interval episode occurrence were analysed as recorded and were not derived from MADRS scores; consistency checks on MADRS-derived variables do not independently confirm a clinical diagnosis.

### 2.7 Statistical analysis

Analysis populations were the 49 episode-free participants for incident MDD, all 103 participants for worsening and continuous outcomes, and the 49 episode-free participants and the 54 with a current episode for stratified sensitivity analyses. Balance between follow-up groups, and between responders and non-responders, was described with standardized mean differences (SMD) rather than hypothesis tests [34].

Binary outcomes were compared with two-sided Fisher's exact tests, which remain valid with small cell counts. Risks are reported with Wilson score 95% confidence intervals (CI) [35], the risk difference (WakeZ minus Discontinued) with Newcombe's hybrid score interval [36], the risk ratio (RR) with a Katz log interval, and the odds ratio (OR) as the conditional maximum-likelihood estimate with an exact CI. Because these interval methods differ from the Fisher test, an interval can occasionally disagree with the Fisher p-value near the 0.05 boundary.

The two principal unadjusted comparisons, incident MDD in the episode-free population and worsening in the full cohort, formed one family of two tests, and their p-values were adjusted with Holm's step-down procedure [37]. All other p-values are nominal and exploratory.

Exploratory adjusted logistic regression models included follow-up group, T0 MADRS (continuous) and parent arm, with Wald CIs. With 22 incident events and three covariates (7.3 events per variable), below conventional guidance [38, 39], Firth penalized-likelihood logistic regression [40, 41] was fitted alongside, with profile penalized-likelihood CIs and

likelihood-ratio p-values; because it reduces small-sample bias, the Firth estimate is reported as the main adjusted estimate. For worsening, a further specification added the baseline episode indicator; because clinically assessed episode status at T0 was concordant with the MADRS symptomatic-remission threshold (≤11) in this cohort, making the two variables closely related, variance inflation factors (VIF) are reported. The overlap between incident MDD and worsening in the episode-free stratum was tabulated.

Month-10 MADRS was analysed by analysis of covariance (ANCOVA) adjusting for T0 MADRS and parent arm [42], with HC3 heteroscedasticity-consistent CIs [43]. Diagnostics comprised the Shapiro–Wilk test of residuals, the Breusch–Pagan test, Cook's distance, leverage and externally studentized residuals, with a refit excluding observations with Cook's distance above 4/n. Sensitivity analyses added the baseline episode indicator, repeated the ANCOVA within each baseline stratum, and estimated the unadjusted difference in MADRS change. Because the episode-free stratum was selected on low baseline scores, floor effects and regression towards the mean were examined [44]. Time-to-event analyses were not possible because episode onset dates were not recorded, and formal interaction tests were not undertaken given the sample size.

Analyses used Python 3.13.6 with pandas 3.1.0, NumPy 2.5.4, SciPy 1.18.1 [45], statsmodels 0.15.0 [46] and matplotlib 3.11.2; Firth models were implemented in custom code. A single script regenerates the numerical results underlying the tables and all five figures from the analysis workbook. Tests were two-sided, and interpretation emphasizes estimates and CIs rather than significance thresholds [47].

## 3 RESULTS

### 3.1 Response, participants and baseline characteristics

Of the 140 parent-trial completers contacted, 103 (73.6%) responded at month 10 (Figure 1). Response was 49/52 (94.2%) among completers with an end-of-trial MADRS ≤11 and 54/88 (61.4%) among those above 11 (Fisher $p < 0.001$). Responders therefore had a lower mean end-of-trial MADRS than non-responders (13.42 vs 19.00; SMD −0.76) and were more often from the parent Test arm (54.4% vs 43.2%; SMD 0.22). Among completers with scores above 11, end-of-trial MADRS was similar in responders and non-responders (19.56 vs 19.91).

**Table 1: Cohort composition and baseline characteristics by follow-up group**

| Characteristic | Overall (N = 103) | WakeZ (n = 56) | Discontinued (n = 47) | SMD |
|---|---|---|---|---|
| *All participants (N = 103)* | | | | |
| Parent arm: Test (REM-haptic), n (%) | 56 (54.4) | 30 (53.6) | 26 (55.3) | −0.04 |
| Current depressive episode at T0, clinician-assessed, n (%) | 54 (52.4) | 31 (55.4) | 23 (48.9) | 0.13 |
| Parent entry MADRS, mean (SD) | 19.50 (5.03) | 19.25 (4.81) | 19.81 (5.31) | −0.11 |
| median [IQR] | 20 [17–23] | 19 [16.75–22] | 20 [17.5–23.5] | |
| T0 MADRS (= parent final MADRS), mean (SD) | 13.42 (8.02) | 13.52 (6.89) | 13.30 (9.27) | 0.03 |
| median [IQR] | 13 [7–19] | 13 [8–19] | 11 [6–19] | |
| range | 0–40 | 0–27 | 0–40 | |
| *No current episode at T0 (n = 49; 25 vs 24)* | | | | |
| T0 MADRS, mean (SD) | 6.65 (3.31) | 7.24 (3.31) | 6.04 (3.26) | 0.36 |
| Parent arm: Test, n (%) | 33 (67.3) | 16 (64.0) | 17 (70.8) | −0.15 |
| *Current episode at T0 (n = 54; 31 vs 23)* | | | | |
| T0 MADRS, mean (SD) | 19.56 (5.76) | 18.58 (4.34) | 20.87 (7.15) | −0.39 |
| Parent arm: Test, n (%) | 23 (42.6) | 14 (45.2) | 9 (39.1) | 0.12 |

The 103 responders comprised 56 WakeZ and 47 Discontinued participants; 49 had no current episode at T0 (25 and 24) and 54 had a current episode (31 and 23). The follow-up groups were similar on measured characteristics (Table 1): mean T0 MADRS was 13.52 versus 13.30 (SMD 0.03), 53.6% versus 55.3% had been in the parent Test arm, and

55.4% versus 48.9% had a current episode at T0 (SMD 0.13). Within strata, T0 MADRS was higher in the WakeZ group among episode-free participants (SMD 0.36) and lower among those with a current episode (SMD −0.39).

SMD, standardized mean difference (WakeZ minus Discontinued; pooled SD for continuous variables, pooled binomial variance for proportions); |SMD| < 0.10 indicates negligible imbalance. No hypothesis tests were performed for baseline comparisons. T0 MADRS equals the parent-trial final MADRS in all 103 records. Percentages use the column total of the relevant row as denominator. IQR, interquartile range; MADRS, Montgomery–Åsberg Depression Rating Scale (0–60); MDD, major depressive disorder; T0, follow-up baseline. Current episode status at T0 was clinician-assessed; participants without a current episode met the symptomatic-remission criterion (MADRS ≤11).

### 3.2 Data validation and parent linkage

All 103 records had unique follow-up and parent identifiers and linked to the parent source with matching arm, entry MADRS and final MADRS. T0 MADRS equalled the parent final MADRS, and the MADRS change and worsening indicators were correctly derived, in all 103 records. No values were out of range, the only missing data were the 54 structurally undefined incident-MDD values.

### 3.3 Incident MDD among participants without a current episode at baseline

Of the 49 episode-free participants, 22 had a new major depressive episode during follow-up: 7/25 (28.0%) (95% CI 14.3–47.6%) in the WakeZ group and 15/24 (62.5%) (42.7–78.8%) in the Discontinued group (Table 2; Figure 2). The risk difference was −34.5 percentage points (95% CI −55.8 to −6.7), the RR 0.45 (0.22–0.90) and the OR 0.24 (0.06–0.90); Fisher p = 0.022, Holm-adjusted p = 0.022. The Firth-adjusted OR was 0.10 (0.02–0.43; p = 0.001) and the conventional adjusted OR 0.07 (0.01–0.42); with 7.3 events per covariate, both are imprecise. Of the 22 participants with an incident episode, 19 were still in an episode at month 10 and 3 had recovered (1 WakeZ, 2 Discontinued); none of the 27 participants without an incident episode had a current episode at month 10.

**Table 2: Ten-month incident MDD among participants without a current episode at follow-up baseline (N = 49)**

| Measure | WakeZ (n = 25) | Discontinued (n = 24) | Estimate (95% CI) | p-value |
|---|---|---|---|---|
| Incident MDD, n | 7 | 15 | | |
| No incident MDD, n | 18 | 9 | | |
| Risk, % (95% CI)[a] | 28.0 (14.3–47.6) | 62.5 (42.7–78.8) | | |
| Risk difference, percentage points[b] | | | −34.5 (−55.8 to −6.7) | |
| Risk ratio[c] | | | 0.45 (0.22–0.90) | |
| Odds ratio, conditional MLE[d] | | | 0.24 (0.06–0.90) | Fisher 0.0222<br>Holm 0.0222 |
| Adjusted OR, logistic regression[e] | | | 0.07 (0.01–0.42) | 0.004 |
| Adjusted OR, Firth penalized[f] | | | 0.10 (0.02–0.43) | 0.001 |

Estimates compare WakeZ with Discontinued (reference). Participants with a current episode at T0 (n = 54) were ineligible and are excluded from the denominator. [a]Wilson score interval. [b]Newcombe hybrid score interval. [c]Katz log interval. [d]Exact conditional interval; the sample OR was 0.23 (Woolf 95% CI 0.07–0.78). Two-sided Fisher exact test; Holm adjustment across the two principal comparisons (this table and the full-cohort row of Table 3). [e]Adjusted for T0 MADRS and parent arm; Wald interval; 22 events, 7.3 events per covariate; exploratory. [f]Same covariates; profile penalized-likelihood interval and likelihood-ratio p-value. CI, confidence interval; MLE, maximum-likelihood estimate; OR, odds ratio.

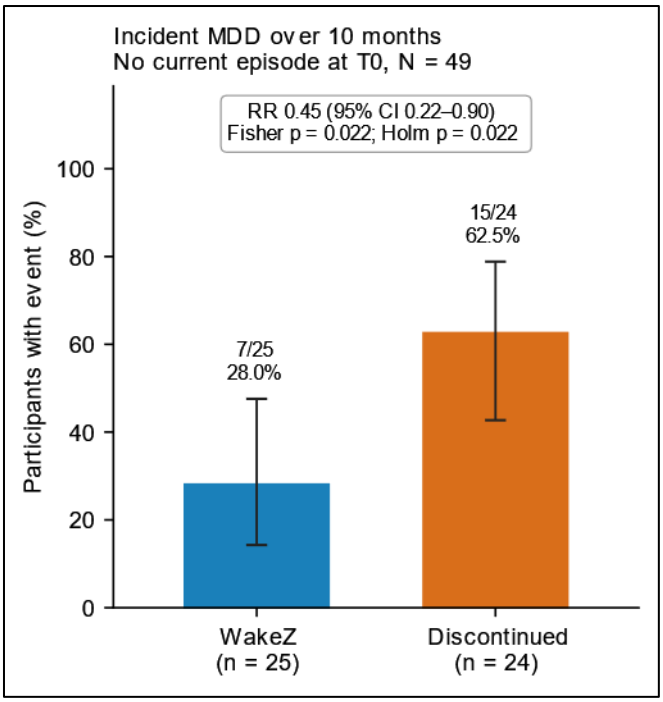



**Figure 2: Ten-month incident MDD among participants without a current episode at follow-up baseline (N = 49). Bars show observed cumulative proportions with Wilson 95% confidence intervals; labels give events/participants. RR, risk ratio (Katz log interval); Holm p, adjusted across the two principal Fisher tests.**

### 3.4 MADRS worsening

In the full cohort, worsening of at least 5 points occurred in 12/56 (21.4%) WakeZ and 22/47 (46.8%) Discontinued participants (Table 3; Figure 3): risk difference −25.4 percentage points (95% CI −41.8 to −7.1), RR 0.46 (0.25–0.82), Fisher p = 0.011, Holm-adjusted p = 0.022. The adjusted OR was 0.29 (0.12–0.71) and the Firth OR 0.31 (0.13–0.72); adding the baseline episode indicator gave 0.28 (0.11–0.72), with VIF of 2.9 for T0 MADRS and 2.9 for the indicator. Because this population includes the 54 participants with a current episode at T0, the result describes less symptom worsening, not prevention of new episodes.

Among the 49 episode-free participants, worsening occurred in 4/25 (16.0%) versus 12/24 (50.0%) (RR 0.32, 0.12–0.86; Fisher p = 0.016). This is not independent evidence: all 16 episode-free participants who worsened also had incident MDD, and 16 of the 22 incident cases also worsened. The adjusted OR in this subgroup was 0.05 (0.01–0.37; Firth 0.08, 0.01–0.39), an implausibly large and imprecise estimate from 16 events. Among the 54 participants with a current episode at T0, worsening occurred in 8/31 (25.8%) versus 10/23 (43.5%) (RR 0.59, 0.28–1.27; Fisher p = 0.245). This difference was not statistically significant, so the full-cohort result does not establish less worsening among participants who had a current episode at the start of follow-up.

**Table 3: MADRS worsening of at least 5 points over 10 months, by analysis population**

| Population | WakeZ n/N (%) | Discontinued n/N (%) | Risk difference, pp (95% CI) | RR (95% CI) | OR (95% CI) | Fisher p |
|---|---|---|---|---|---|---|
| All participants (N = 103); principal | 12/56 (21.4) | 22/47 (46.8) | −25.4 (−41.8 to −7.1) | 0.46 (0.25–0.82) | 0.31 (0.12–0.79) | 0.0110 (Holm 0.0221) |
| No current episode at T0 (N = 49); sensitivity | 4/25 (16.0) | 12/24 (50.0) | −34.0 (−54.9 to −7.7) | 0.32 (0.12–0.86) | 0.20 (0.04–0.84) | 0.0157 |
| Current episode at T0 (N = 54); symptom course, sensitivity | 8/31 (25.8) | 10/23 (43.5) | −17.7 (−40.8 to 7.3) | 0.59 (0.28–1.27) | 0.46 (0.12–1.66) | 0.2445 |

Worsening was defined as month-10 MADRS minus T0 MADRS ≥ 5. Estimates compare WakeZ with Discontinued (reference). The full-cohort difference was significant after Holm adjustment. In the episode-free stratum the difference was nominally statistically significant (p = 0.016), but all 16 events in that stratum were also incident-MDD cases. In the MDD-at-T0 stratum it was not statistically significant (p = 0.245), and the CIs for the risk difference and RR include the null. The full-cohort analysis includes participants with a current episode at T0 and indicates less symptom worsening, not prevention of new episodes. Risk difference: Newcombe hybrid score interval; RR: Katz log interval; OR: conditional MLE with exact interval. Adjusted models are reported in Table 5. pp, percentage points; RR, risk ratio; OR, odds ratio.

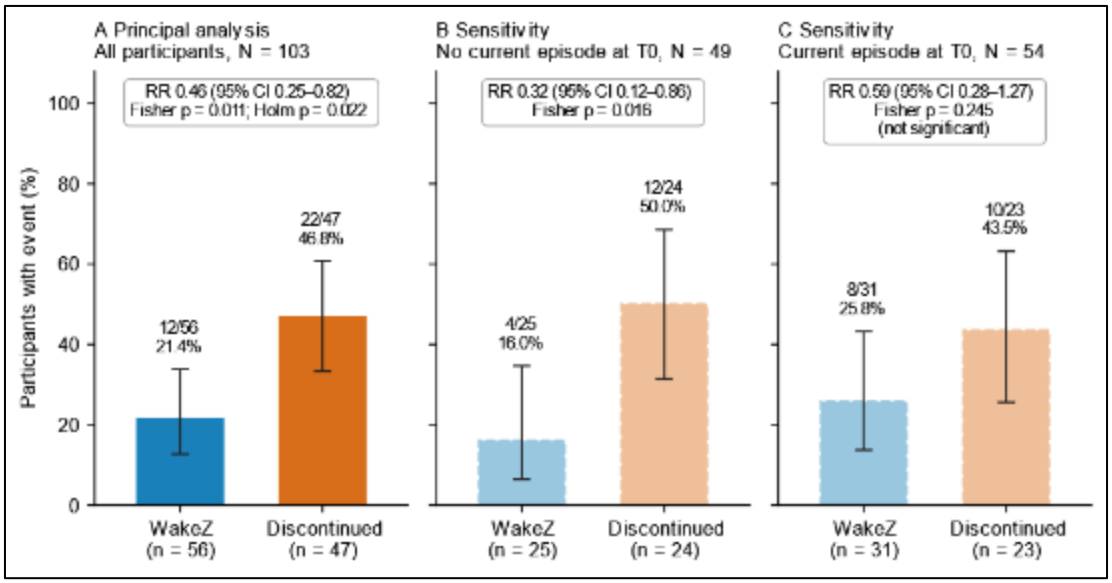



**Figure 3: Proportion of participants with MADRS worsening of at least 5 points over 10 months. (A) Principal analysis in all participants (N = 103), including 54 with a current episode at T0; this panel reflects symptom course, not prevention of new episodes. (B, C) Sensitivity analyses in participants without (N = 49) and with (N = 54) a current episode at T0, shown as lighter bars with dashed outlines; the difference was nominally statistically significant in B and not statistically significant in C. Error bars, Wilson 95% confidence intervals.**

### 3.5 Continuous MADRS outcomes

Mean MADRS in the WakeZ group was 13.52 at T0 and 13.38 at month 10 (mean change −0.14, SD 5.42); in the Discontinued group it was 13.30 and 16.87 (mean change 3.57, SD 5.50) (Table 4; Figure 4). The ANCOVA-adjusted difference in month-10 MADRS was −3.72 points (HC3 95% CI −5.93 to −1.51; p = 0.001; $R^2$ = 0.70). No studentized residual exceeded 3, the largest Cook's distance was 0.08, and the Shapiro–Wilk (p = 0.620) and Breusch–Pagan (p = 0.163) tests did not indicate marked non-normality or heteroscedasticity. Excluding 3 observations with Cook's distance above 4/n gave −4.54 (−6.57 to −2.51), and adding the baseline episode indicator gave −3.42 (−5.51 to −1.34; Breusch–Pagan p = 0.024, for which the HC3 interval is appropriate).

In exploratory stratified models, the adjusted difference was −3.81 (−6.06 to −1.56) among episode-free participants and −3.38 (−6.99 to 0.22; p = 0.065) among those with a current episode at T0; the intervals overlap and no interaction was tested. The episode-free stratum showed a floor pattern: median T0 MADRS was 7, 23 of 49 scored ≤6 and 4 scored 0, and mean change was 2.57 points compared with 0.63 in the MDD stratum.

**Table 4: Continuous MADRS outcomes by follow-up group and baseline-adjusted comparisons**

| Population / measure | WakeZ mean (SD) median [IQR] | Discontinued mean (SD) median [IQR] | Difference (95% CI)[a] | p-value |
|---|---|---|---|---|
| *All participants (N = 103; 56 vs 47)* | | | | |
| MADRS at T0 | 13.52 (6.89)<br>13 [8–19] | 13.30 (9.27)<br>11 [6–19] | | |
| MADRS at month 10 | 13.38 (8.39)<br>12 [6.75–20] | 16.87 (11.38)<br>16 [9–23] | | |

| | | | | |
|---|---|---|---|---|
| Change (month 10 − T0) | −0.14 (5.42)<br>0 [−4–2] | 3.57 (5.50)<br>4 [0–8] | | |
| ANCOVA, month-10 MADRS[b] | | | −3.72 (−5.93 to −1.51) | 0.001 |
| ANCOVA + baseline episode indicator | | | −3.42 (−5.51 to −1.34) | 0.002 |
| Unadjusted difference in change | | | −3.72 (−5.86 to −1.58) | <0.001 |
| *No current episode at T0 (N = 49; 25 vs 24); exploratory* | | | | |
| MADRS at T0 | 7.24 (3.31)<br>8 [6–10] | 6.04 (3.26)<br>6 [3.75–9] | | |
| MADRS at month 10 | 8.24 (5.55)<br>9 [5–10] | 10.25 (6.68)<br>12.5 [4.75–16] | | |
| Change (month 10 − T0) | 1.00 (3.43)<br>0 [−1–2] | 4.21 (4.51)<br>4 [0–8.25] | | |
| ANCOVA, month-10 MADRS[b] | | | −3.81 (−6.06 to −1.56) | 0.001 |
| *Current episode at T0 (N = 54; 31 vs 23); exploratory* | | | | |
| MADRS at T0 | 18.58 (4.34)<br>19 [15–21.5] | 20.87 (7.15)<br>19 [15.5–24] | | |
| MADRS at month 10 | 17.52 (8.05)<br>18 [11.5–23.5] | 23.78 (11.24)<br>24 [14–30] | | |
| Change (month 10 − T0) | −1.06 (6.52)<br>−4 [−5.5–4] | 2.91 (6.42)<br>4 [−1.5–8] | | |
| ANCOVA, month-10 MADRS[b] | | | −3.38 (−6.99 to 0.22) | 0.065 |

[a] WakeZ minus Discontinued, in MADRS points; negative values indicate lower (better) scores with continued use. HC3 heteroscedasticity-consistent 95% CIs. [b]Month-10 MADRS adjusted for T0 MADRS and parent arm. Change ranges: WakeZ −11 to 11, Discontinued −11 to 11. In this cohort, clinically assessed episode status at T0 was concordant with the MADRS symptomatic-remission threshold (≤11), so this specification is partly redundant. Stratified estimates are exploratory and were not tested for interaction. ANCOVA, analysis of covariance.

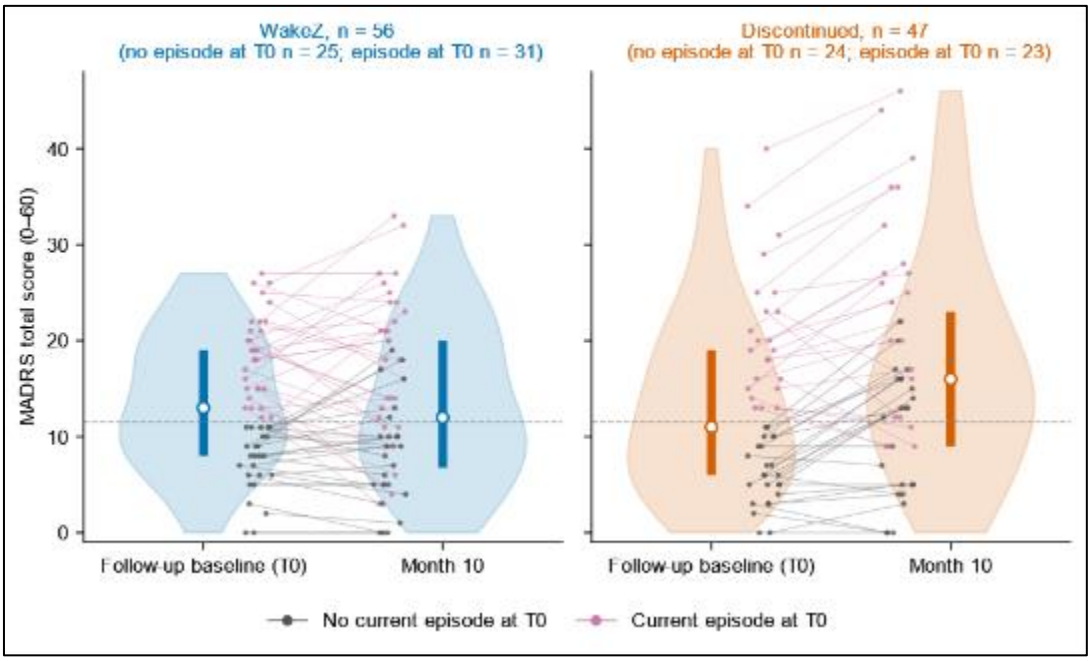



**Figure 4: Participant-level MADRS scores at follow-up baseline (T0) and month 10 by follow-up group. Shaded areas show score distributions, vertical bars the interquartile range and open circles the median. Lines connect each participant's two observed scores and are coloured by episode status at T0. Only two timepoints were observed, so lines do not represent intermediate trajectories. Dashed line, MADRS = 11, the upper bound of the symptomatic-remission criterion at T0; it is a score reference, not a diagnostic boundary.**

### 3.6 Depression status at month 10

Clinically assessed current MDD at month 10 was present in 29/56 (51.8%) WakeZ and 32/47 (68.1%) Discontinued participants (RR 0.76, 0.55–1.05; Fisher p = 0.110). Among participants with a current episode at T0 it was 23/31 (74.2%) versus 19/23 (82.6%) (RR 0.90, 0.68–1.19; p = 0.525), and among episode-free participants 6/25 (24.0%) versus 13/24 (54.2%) (RR 0.44, 0.20–0.97; p = 0.042). In the episode-free stratum, current MDD at month 10 is part of the interval-episode outcome, so that comparison is not independent of the primary result; point prevalence at month 10 and interval occurrence are distinct outcomes (Table 5; Figure 5).

### 3.7 Multiplicity and robustness

Both principal comparisons remained below 0.05 after Holm adjustment (p = 0.022 and 0.022). Table 5 summarizes all models and Figure 5 displays the estimates on their own scales.

**Table 5: Summary of unadjusted, adjusted and sensitivity models**

| Analysis (covariates) | n / events | Estimate (95% CI) | p-value | Notes and diagnostics |
|---|---|---|---|---|
| *Incident MDD, no current episode at T0 (primary)* | | | | |
| Fisher exact, unadjusted | 49 / 22 | OR 0.24 (0.06–0.90)<br>RR 0.45 (0.22–0.90) | 0.022<br>Holm 0.022 | Principal test 1 |
| Logistic (group, T0 MADRS, parent arm) | 49 / 22 | aOR 0.07 (0.01–0.42) | 0.004 | EPV 7.3; max VIF 1.04; unstable |
| Firth logistic (same covariates) | 49 / 22 | aOR 0.10 (0.02–0.43) | 0.001 | Main adjusted estimate; profile penalized-likelihood CI |
| *MADRS worsening ≥ 5 points* | | | | |
| Fisher exact, all participants | 103 / 34 | OR 0.31 (0.12–0.79)<br>RR 0.46 (0.25–0.82) | 0.011<br>Holm 0.022 | Principal test 2 |
| Logistic (group, T0 MADRS, parent arm) | 103 / 34 | aOR 0.29 (0.12–0.71) | 0.006 | EPV 11.3; max VIF 1.07 |
| Firth logistic (same covariates) | 103 / 34 | aOR 0.31 (0.13–0.72) | 0.006 | |
| Logistic + baseline episode indicator | 103 / 34 | aOR 0.28 (0.11–0.72) | 0.008 | VIF 2.9 (T0 MADRS), 2.9 (indicator); closely related (concordant in this cohort) |
| Fisher exact, no current episode at T0 | 49 / 16 | OR 0.20 (0.04–0.84)<br>RR 0.32 (0.12–0.86) | 0.016 | Sensitivity; all 16 events shared with incident MDD |
| Logistic, no current episode at T0 | 49 / 16 | aOR 0.05 (0.01–0.37) | 0.003 | EPV 5.3; exploratory; Firth aOR 0.08 (0.01–0.39) |
| Fisher exact, current episode at T0 | 54 / 18 | OR 0.46 (0.12–1.66)<br>RR 0.59 (0.28–1.27) | 0.245 | Symptom course; not significant |
| *Month-10 MADRS, ANCOVA (group, T0 MADRS, parent arm; HC3 CI)* | | | | |
| All participants | 103 | MD −3.72 (−5.93 to −1.51) | 0.001 | $R^2$ 0.70; Shapiro–Wilk p 0.620; Breusch–Pagan p 0.163; max Cook's D 0.08; no $\|r^*\| > 3$ |

| | | | | |
|---|---|---|---|---|
| Excluding Cook's D > 4/n | 100 | MD −4.54 (−6.57 to −2.51) | <0.001 | 3 observations removed |
| + baseline episode indicator | 103 | MD −3.42 (−5.51 to −1.34) | 0.002 | VIF 2.9 (T0 MADRS); Breusch–Pagan p 0.024 |
| No current episode at T0 | 49 | MD −3.81 (−6.06 to −1.56) | 0.001 | Exploratory; floor effects; $R^2$ 0.66 |
| Current episode at T0 | 54 | MD −3.38 (−6.99 to 0.22) | 0.065 | Exploratory; $R^2$ 0.61 |
| *Depression status at month 10 (exploratory, Fisher exact)* | | | | |
| Current MDD at month 10, all | 103 / 61 | RR 0.76 (0.55–1.05) | 0.110 | Clinically assessed at month 10 |
| Current MDD at month 10, no current episode at T0 | 49 / 19 | RR 0.44 (0.20–0.97) | 0.042 | Part of the interval-episode outcome |
| Current MDD at month 10, current episode at T0 | 54 / 42 | RR 0.90 (0.68–1.19) | 0.525 | Persistence or recurrence |

All estimates compare WakeZ with Discontinued. Only the two principal Fisher tests were Holm-adjusted; all other p-values are nominal and exploratory. Logistic-regression intervals are Wald intervals; Firth intervals are profile penalized-likelihood intervals; RR intervals are Katz log intervals. aOR, adjusted odds ratio; EPV, events per covariate; MD, adjusted mean difference in MADRS points; r*, externally studentized residual; RR, risk ratio; VIF, variance inflation factor.

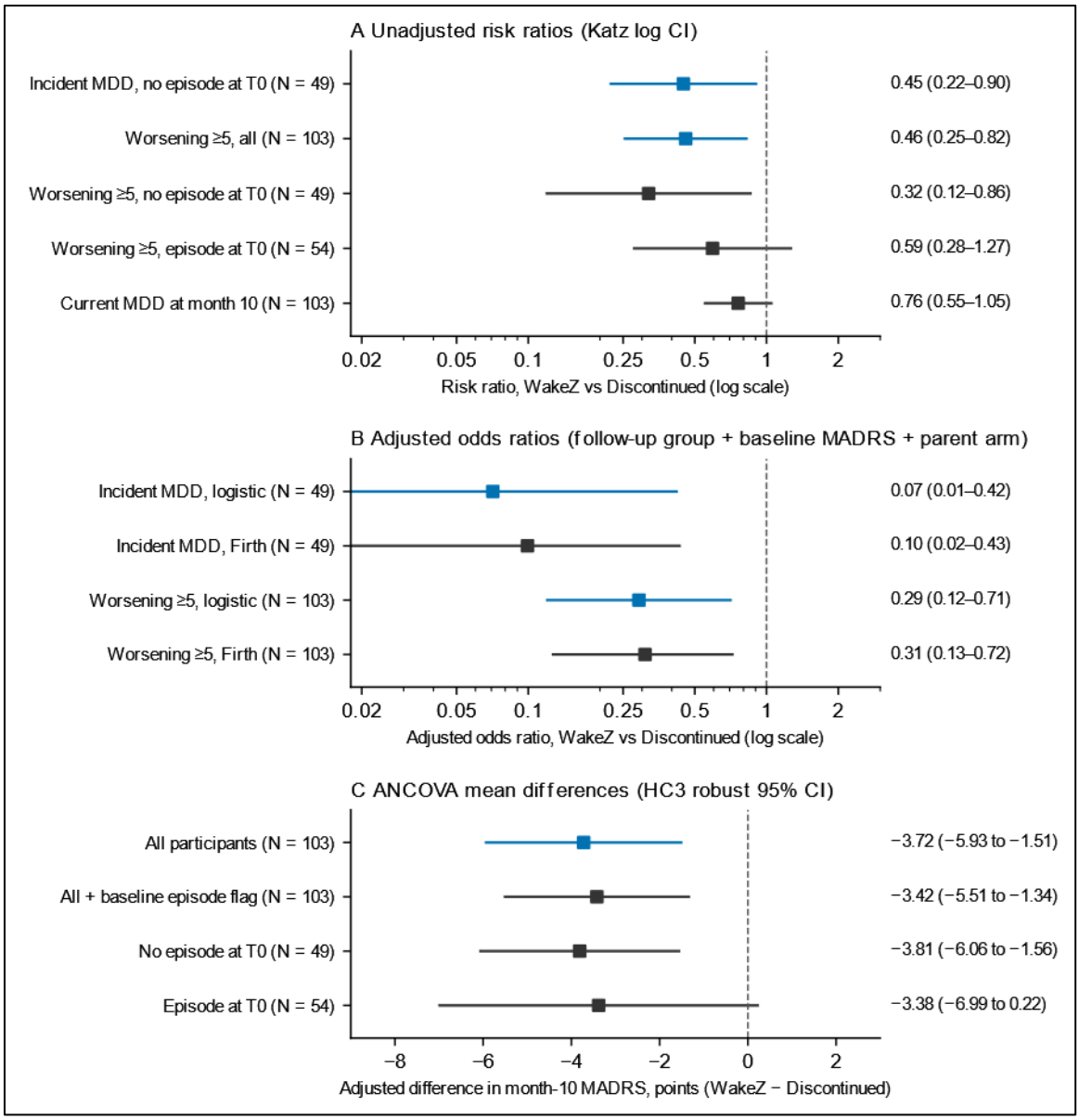



**Figure 5: Effect estimates for WakeZ versus Discontinued. (A) Unadjusted risk ratios, including current MDD at month 10; (B) adjusted odds ratios from conventional (Wald interval) and Firth penalized (profile interval) logistic regression; (C) ANCOVA-adjusted differences in month-10 MADRS, in points. Each panel uses its own measure and scale. Values left of the dashed reference line favour the WakeZ group. Blue, principal analysis populations; black, sensitivity or exploratory analyses.**

## 4 DISCUSSION

### 4.1 Principal findings

In this observational follow-up, participants who continued to use both the smart alarm and the self-reflection feature had a lower observed 10-month incidence of new major depressive episodes than those who discontinued (28.0% vs 62.5% among 49 episode-free participants) and were less likely to experience a MADRS increase of at least 5 points (21.4% vs 46.8% among all 103). Both unadjusted comparisons remained below 0.05 after Holm adjustment, and the adjusted logistic, Firth and ANCOVA estimates pointed in the same direction. The findings were not uniform, however. Worsening among episode-free participants largely repeated the incident-MDD result because the two outcomes shared most events. Among participants who already had a current episode at T0, neither worsening ($p = 0.245$), adjusted month-10 MADRS ($p = 0.065$) nor current MDD at month 10 ($p = 0.525$) differed significantly, and across all participants current MDD at month 10 was less frequent with continued use, but not significantly ($p = 0.110$). The data are therefore consistent with an association between continued use and fewer new episodes and less symptom deterioration, but provide no clear evidence of benefit for participants who had a current episode at T0 or of a lower prevalence of depression at the end of follow-up. Paving the way towards better prevention of depression, these observational data

are hypothesis-generating: if further confirmed in larger randomized studies, continued use of a low-burden sleep companion after symptomatic improvement could represent a scalable approach to preventing new depressive episodes.

### 4.2 Incident episodes and month-10 status

The interval-episode endpoint and current MDD status are related but not identical. Among participants without an ongoing episode at T0, 19 of the 22 interval episodes were ongoing at the month-10 assessment, whereas 3 had resolved, consistent with a median episode duration of about 3 months [48]; point prevalence at a single visit therefore understates interval incidence. The between-group difference in current MDD at month 10 was smaller than for interval episodes and not significant overall, because persistence or recurrence among participants with a current episode at T0 was similar in both groups (74.2% vs 82.6%). Both variables were obtained from clinical assessment, while MADRS provided a separate continuous measure of symptom severity. Consequently, episode outcomes and MADRS deterioration are correlated and should not be interpreted as independent demonstrations of benefit.

### 4.3 Possible explanations

Several mechanisms could link continued use with better outcomes, but none was measured. Regular use of an alarm may encourage more consistent wake times, and variability in wake time has been associated with depressive symptoms [10, 11]. Waking from lighter sleep may reduce sleep inertia [16], although consumer devices classify sleep stages imperfectly [14], so stage-targeted waking cannot be assumed to work as intended. Daily self-reflection may increase emotional self-awareness, which mediated symptom improvement in a mobile self-monitoring trial [20]. These remain hypotheses. The analysis extract contains no sleep-architecture, regularity, alarm-timing, reflection or usage-intensity data, so no mediation analysis was possible (Section 4.5).

### 4.4 Comparison with previous studies

We did not identify a published study of continued smart-alarm or reflection use with diagnostic depression outcomes. The nearest evidence comes from CBT-I. In a trial of digital CBT-I, incident moderate-to-severe depression at one year among participants with minimal baseline symptoms was approximately halved [24]; in older adults with insomnia, incident or recurrent major depression occurred in 12.2% after CBT-I versus 25.9% after sleep education over 36 months [26]; and the GoodNight trial reduced depressive symptoms but not diagnosed MDD at 6 months [25]. Across psychological prevention trials, the one-year RR for onset is about 0.81 [5]. Those interventions are structured therapeutic programmes evaluated by randomization; the direction of the association observed here is nonetheless consistent with a preventive effect of the kind reported for structured digital CBT-I [24–26]. The relative difference observed here (RR 0.45) is much larger than these randomized estimates, which, together with the wide CI, suggests that confounding, selection and chance may contribute substantially. Because engagement with mental-health apps usually declines quickly [18, 19], people who sustain use for many months are likely to differ systematically from those who stop.

### 4.5 Self-selection, non-response and confounding

The exposure groups were self-selected. Continuation is plausibly influenced by mood, motivation, perceived benefit and life circumstances, all of which also predict the course of depression. This is a form of healthy-adherer bias: good adherence even to placebo has been associated with lower mortality [49], showing that adherence can mark health-related behaviour rather than treatment effect [50]. Reverse causation is also plausible, because emerging symptoms such as loss of interest, fatigue and disturbed sleep may lead people to abandon a self-management tool. Exposure was classified over the same period in which outcomes accrued, and without time-stamped usage and symptom data the order of these events cannot be established. Psychiatric care, medication, psychotherapy and life events during follow-up were not available in the analysis extract and could confound the contrast, particularly for remission among participants with a current episode at T0.

Non-response is a second source of bias. Almost all completers with low end-of-trial scores responded, but 34 of 88 with higher scores did not, and their exposure and outcomes are unknown. If non-response among symptomatic completers was related both to continued use and to the course of depression, missing outcomes are not missing at random, and estimates in the current-episode stratum and the full cohort may be biased [51]; the loss of 68 randomized participants before the end-of-trial assessment further limits generalizability. Parent randomization does not extend to the follow-up contrast: parent arm was similarly distributed between follow-up groups and was adjusted for, but adjustment cannot randomize the later decision to continue, and these results provide no evidence on the efficacy of the parent REM-haptic intervention.

### 4.6 Outcome ascertainment, power and other limitations

Although depressive episode status was clinically assessed, interval events were reconstructed at a single retrospective month-10 interview, with no intermediate MADRS measurements; neither episode-onset dates nor sufficiently granular engagement histories were available to establish the temporal ordering of changes in WakeZ use and depressive symptoms. Recall of depressive episodes declines over time [52], and prospective assessment detects substantially more episodes than retrospective assessment [53]. If recall differed by exposure, for example because daily reflection improved recall of low-mood periods, misclassification could be differential. Episode status and MADRS ratings were assessed by members of the research team, who may have known participants' follow-up group. In the episode-free stratum, selection on low baseline scores means that regression towards the mean [44] contributes to the average increase in both groups and makes threshold-based worsening sensitive to small fluctuations.

The study was not prospectively powered for these comparisons. With 22 incident events, adjusted models had 7.3 events per covariate, below conventional guidance [38, 39], and the CI for the primary RR extended from 0.22 to 0.90. Statistically significant estimates from small samples tend to overstate effect sizes, and p-values of 0.01–0.03 provide only moderate evidence against the null hypothesis [47]. Exploratory analyses were not corrected for multiplicity, and no demographic, treatment, life-event or usage-intensity data were available in the analysis extract. Strengths include complete data for all responders, verified linkage to the parent trial, exact tests, explicit denominators and reporting of non-significant results.

### 4.7 Implications for future research

A confirmatory evaluation should randomize participants without a current depressive episode to continued WakeZ access or an appropriate control. It should capture time-stamped usage logs, repeated symptom assessments to date episode onset, structured diagnostic interviews by assessors blinded to exposure, concurrent treatment and life events, and strategies to limit differential loss to follow-up, with a prespecified primary endpoint analysed by time-to-event methods. Sample size should come from a separate planning exercise with conservative assumptions rather than from the estimates observed here. For example, detecting a reduction in 10-month incidence from 30% to 20% with 80% power at a two-sided α of 0.05 would require approximately 290 participants per group before allowance for attrition.

## 5 CONCLUSIONS

In this observational follow-up, continued use of WakeZ was associated with a lower 10-month incidence of new major depressive episodes among participants without a current episode at baseline (RR 0.45, 95% CI 0.22–0.90), less than half the risk observed with discontinuation, and with less MADRS worsening across the full cohort (RR 0.46, 95% CI 0.25–0.82); both comparisons remained significant after Holm adjustment, strongly hinting at depression prevention capabilities. The association with worsening was not significant among participants who had a current episode at T0, and current MDD at month 10 was less frequent with continued use but not significantly. These findings represent steps towards the prevention of new depressive episodes with low-burden digital tools and justify a larger, more adequately powered prospective randomized prevention trial with longitudinal measurement of use and outcomes and independent diagnostic assessment, which should also establish which component, the smart alarm or self-reflection, accounts for the observed benefit.

## STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

### *Acknowledgments*

The authors would like to acknowledge the contribution of all parties involved in the project.

### *Funding Source*

This research and project were financed by the North-West Regional Development Agency of Romania through the North-West Regional Program 2021–2027 from the European Regional Development Fund of the European Union, under grant(SMIS) number 308944, WAKEZ HEALTH TECH ECOSYSTEM SRL Beneficiary.

### *Disclosure of conflict of interest*

Dr. Milhem (affiliated with WAKEZ HEALTH TECH ECOSYSTEM SRL) was involved in providing proprietary software used in the study, developed by the company WAKEZ HEALTH TECH ECOSYSTEM SRL. No interpretation of data or formulation of conclusions was done by Dr. Milhem. Dr. Stretea served as an advisor to WAKEZ HEALTH TECH ECOSYSTEM SRL and is not employed by the company at the time of submission; he contributed to this research through his affiliation with the "Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca. The authors declare no other conflicts of interest.

### *Statement of ethical approval*

The parent trial protocol ("REM sleep disruption and its potential uses in depression management"), including all measures regarding the follow-up assessments, was reviewed by the Institutional Review Board – Public Health (IRB-PH) of Babeş-Bolyai University, Cluj-Napoca, Romania (IRB00005728), which granted it exempt status (protocol no. 2024-241218-002; approval date 23 December 2024). The study was conducted in accordance with the Declaration of Helsinki.

### *Statement of informed consent*

All participants gave informed consent in the parent trial, including consent to be contacted for follow-up assessments such as the present study.

### *Data availability*

The de-identified analysis dataset is available from the corresponding author on reasonable request, subject to data-protection requirements; keys linking study identifiers to participants are not shared.

### *Code availability*

The Python script that regenerates the numerical results underlying the tables and all five figures from the analysis workbook is available from the corresponding author on reasonable request.

## REFERENCES

[1] GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry. 2022;9(2):137-50.

[2] Greenberg PE, Fournier AA, Sisitsky T, et al. The economic burden of adults with major depressive disorder in the United States (2010 and 2018). Pharmacoeconomics. 2021;39(6):653-65.

[3] Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health. 2013;34:119-38.

[4] Cuijpers P, Smit F. Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies. Acta Psychiatr Scand. 2004;109(5):325-31.

[5] Cuijpers P, Pineda BS, Quero S, et al. Psychological interventions to prevent the onset of depressive disorders: a meta-analysis of randomized controlled trials. Clin Psychol Rev. 2021;83:101955.

[6] Baglioni C, Battagliese G, Feige B, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord. 2011;135(1-3):10-9.

[7] Li L, Wu C, Gan Y, et al. Insomnia and the risk of depression: a meta-analysis of prospective cohort studies. BMC Psychiatry. 2016;16(1):375.

[8] Zhai L, Zhang H, Zhang D. Sleep duration and depression among adults: a meta-analysis of prospective studies. Depress Anxiety. 2015;32(9):664-70.

[9] Riemann D, Krone LB, Wulff K, et al. Sleep, insomnia, and depression. Neuropsychopharmacology. 2020;45(1):74-89.

[10] Fang Y, Forger DB, Frank E, et al. Day-to-day variability in sleep parameters and depression risk: a prospective cohort study of training physicians. NPJ Digit Med. 2021;4(1):28.

[11] Lyall LM, Wyse CA, Graham N, et al. Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. Lancet Psychiatry. 2018;5(6):507-14.

[12] Palagini L, Baglioni C, Ciapparelli A, et al. REM sleep dysregulation in depression: state of the art. Sleep Med Rev. 2013;17(5):377-90.

[13] Baglioni C, Nanovska S, Regen W, et al. Sleep and mental disorders: a meta-analysis of polysomnographic research. Psychol Bull. 2016;142(9):969-90.

[14] Chinoy ED, Cuellar JA, Huwa KE, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep. 2021;44(5):zsaa291.

[15] de Zambotti M, Cellini N, Goldstone A, et al. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc. 2019;51(7):1538-57.

[16] Trotti LM. Waking up is the hardest thing I do all day: sleep inertia and sleep drunkenness. Sleep Med Rev. 2017;35:76-84.

[17] Baron KG, Abbott S, Jao N, et al. Orthosomnia: are some patients taking the quantified self too far? J Clin Sleep Med. 2017;13(2):351-4.

[18] Baumel A, Muench F, Edan S, et al. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res. 2019;21(9):e14567.

[19] Torous J, Nicholas J, Larsen ME, et al. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health. 2018;21(3):116-9.

[20] Kauer SD, Reid SC, Crooke AH, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res. 2012;14(3):e67.

[21] Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry. 2017;16(3):287-98.

[22] Zhang Y, Folarin AA, Sun S, et al. Relationship between major depression symptom severity and sleep collected using a wristband wearable device: multicenter longitudinal observational study. JMIR Mhealth Uhealth. 2021;9(4):e24604.

[23] WAKEZ HEALTH TECH ECOSYSTEM SRL. REM sleep disruption and its potential uses in depression management. In: ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US); 2000- [cited 2026 Nov 3]. Available from: https://clinicaltrials.gov/study/NCT07891806 NLM Identifier: NCT07891806.

[24] Cheng P, Kalmbach DA, Tallent G, et al. Depression prevention via digital cognitive behavioral therapy for insomnia: a randomized controlled trial. Sleep. 2019;42(10):zsz150.

[25] Christensen H, Batterham PJ, Gosling JA, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. Lancet Psychiatry. 2016;3(4):333-41.

[26] Irwin MR, Carrillo C, Sadeghi N, et al. Prevention of incident and recurrent major depression in older adults with insomnia: a randomized clinical trial. JAMA Psychiatry. 2022;79(1):33-41.

[27] von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453-7.

[28] Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382-9.

[29] Snaith RP, Harrop FM, Newby DA, et al. Grade scores of the Montgomery-Asberg Depression and the Clinical Anxiety Scales. Br J Psychiatry. 1986;148:599-601.

[30] Zimmerman M, Posternak MA, Chelminski I. Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. J Psychiatr Res. 2004;38(6):577-82.

[31] Hawley CJ, Gale TM, Sivakumaran T. Defining remission by cut off score on the MADRS: selecting the optimal value. J Affect Disord. 2002;72(2):177-84.

[32] Duru G, Fantino B. The clinical relevance of changes in the Montgomery-Asberg Depression Rating Scale using the minimum clinically important difference approach. Curr Med Res Opin. 2008;24(5):1329-35.

[33] Turkoz I, Alphs L, Singh J, et al. Clinically meaningful changes on depressive symptom measures and patient-reported outcomes in patients with treatment-resistant depression. Acta Psychiatr Scand. 2021;143(3):253-63.

[34] Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med. 2009;28(25):3083-107.

[35] Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc. 1927;22(158):209-12.

[36] Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med. 1998;17(8):873-90.

[37] Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6(2):65-70.

[38] Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-9.

[39] Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol. 2007;165(6):710-8.

[40] Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27-38.

[41] Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med. 2002;21(16):2409-19.

[42] Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. BMJ. 2001;323(7321):1123-4.

[43] Long JS, Ervin LH. Using heteroscedasticity consistent standard errors in the linear regression model. Am Stat. 2000;54(3):217-24.

[44] Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol. 2005;34(1):215-20.

[45] Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261-72.

[46] Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. In: Proceedings of the 9th Python in Science Conference. Austin, TX; 2010. p. 92-6.

[47] 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(4):337-50.

[48] Spijker J, de Graaf R, Bijl RV, et al. Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry. 2002;181:208-13.

[49] Simpson SH, Eurich DT, Majumdar SR, et al. A meta-analysis of the association between adherence to drug therapy and mortality. BMJ. 2006;333(7557):15.

[50] Shrank WH, Patrick AR, Brookhart MA. Healthy user and related biases in observational studies of preventive interventions: a primer for physicians. J Gen Intern Med. 2011;26(5):546-50.

[51] Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615-25.

[52] Patten SB, Williams JV, Lavorato DH, et al. Recall of recent and more remote depressive episodes in a prospective cohort study. Soc Psychiatry Psychiatr Epidemiol. 2012;47(5):691-6.

[53] Moffitt TE, Caspi A, Taylor A, et al. How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. Psychol Med. 2010;40(6):899-909.