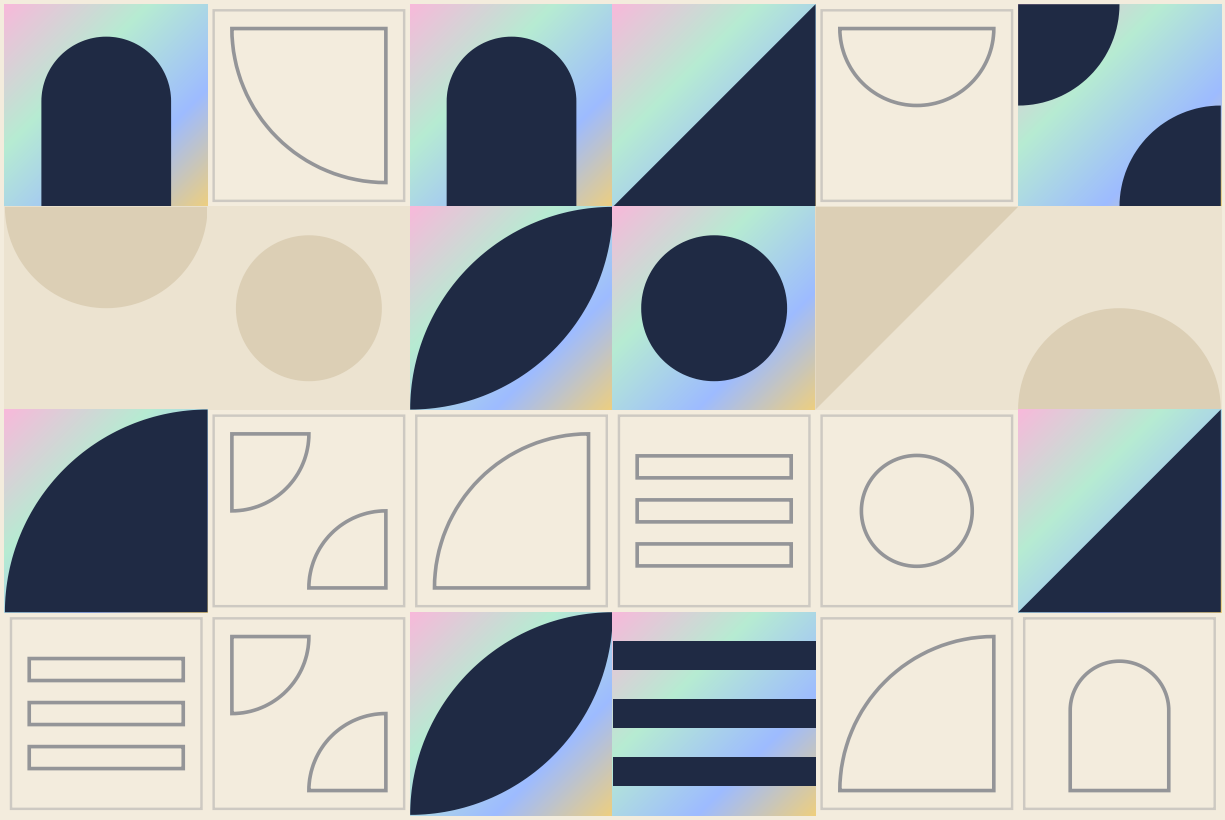
RESEARCH PAPER · 52

The Mindfulness App Plateau

A Meta-Analysis of Effect Durability Beyond Eight Weeks

Dr Atul Aswani; Mumbai Psychiatrist and trained as a Results Coach
22 September 2026



KEY NUMBER · D

0.33

A four-week wellbeing app retained 62 per cent of its distress effect at three months after the programme ended (n = 662; follow-up d = 0.33).

The Mindfulness App Plateau

A Meta-Analysis of Effect Durability Beyond Eight Weeks

What remains of mindfulness-app effects three to six months on?

CONTENTS

INDEXING TITLE

Durability of mindfulness-app effects beyond eight weeks: a meta-analysis

AUTHOR

Dr Atul Aswani; Mumbai Psychiatrist and trained as a Results Coach

PUBLISHED

22 September 2026 · English edition · emotionapp.in/research/mindfulness-app-plateau-beyond-eight-weeks

01

Abstract

Mindfulness apps sell an eight-week story. This paper asks what is left of that story at three and six months. A PRISMA-2020 systematic review of randomised and controlled trials (2012–2026) found too few comparable wait-list trials sharing the same outcome at the same horizon to justify a pooled Hedges g; the durability number is therefore a decay ratio, not a single pooled effect. In the three wait-list trials that measured both post-programme stress or distress and a three-to-four-month follow-up, decay ratios were 0.62, 0.84 and 0.94 (unweighted mean 0.80; sample-size-weighted mean 0.76). About three-quarters to four-fifths of the immediate effect was still visible. The largest employee trial (n = 1,458) moved from d = 0.85 at eight weeks to d = 0.71 at four months from randomisation. A four-week wellbeing app in school employees retained 62 per cent of its distress effect three months after the programme ended (d = 0.33). Two app-only trials in public-sector and general-population adults found no change in perceived stress at three or six months. Six-month app-only evidence is too thin and too mixed to pool. Continued use during the programme tracks larger gains; use collapses for most people once the structured weeks end. Certainty of evidence is low for three-to-four-month stress and very low for six-month stress and for wellbeing. No eligible mindfulness-app randomised trial with a three-to-six-month follow-up has been run in Indian working adults. For an Indian workplace programme the practical reading is blunt: an eight-week download is a seed, not a harvest. Daily reps after week eight are the root system.

02 Key findings box

KEY FINDINGS

0.62	About four-fifths of the post-programme stress or distress effect was still visible at three to four months in wait-list trials that measured both time-points (decay ratios 0.62, 0.84, 0.94; unweighted mean 0.80).
0.71	The largest employee trial retained a perceived-stress effect of **0.71** at four months from randomisation, down from 0.85 at eight weeks (n = 1,458).
0.33	A four-week wellbeing app retained **62 per cent** of its distress effect at three months after the programme ended (n = 662; follow-up *d* = 0.33).
211	App-only arms were **null** on perceived stress at three months and at six months in two public-sector and general-population trials (n = 211 and n = 587).
0	Eligible Indian mindfulness-app trials with a three-to-six-month follow-up in working adults: **zero**.

03 Definition block

DEFINITION

Mindfulness, in the sense used here, is a trainable attention skill. You notice what is happening in this moment — a tight jaw, a looping thought, a WhatsApp ping — and you do not immediately wrestle it, flee it, or buy a story about it. That is the practice. It is closer to a gym rep than to a worldview. The looping thought (“this deck will sink me”) is a thought, not a fact. Noticing the urge to bolt from the meeting, and staying in the chair, is a muscle. The next move follows what matters to you, not the mood of the minute. A mindfulness app is simply a coach in your pocket that counts those reps: a timed sitting, a body scan, a walking practice, a two-minute pause between meetings. The app is not a diagnosis and it is not a clinic. It is a delivery channel for a technique. The question this paper asks is not whether the technique can move a score at week eight. Plenty of trials have already answered that. The question is whether the score still looks different at month three and month six, once the novelty has worn off and the icon has drifted to the fourth screen of the phone.

04 Introduction

Eight weeks is the honeymoon. Three months is the marriage.

Picture an analyst on the morning local in week one of her company's app licence. Earphones in, eyes shut, a short sitting somewhere between Borivali and Dadar. By week eight her stress score on the office survey is down. Month three is a wet Tuesday. The train is late, the reminder is swiped away, and she scrolls the news instead. She finished the course. She felt lighter. Then she stopped sitting.

The download is the easy part. The practice then has to survive an Indian workday: the commute, the stand-up, Slack, the family WhatsApp group, one more deck. Each of those is a cue to do something else. A quick lift at week eight with no daily practice under it is exactly the kind of gain you would expect to fade. A practice tied to a fixed moment in the day — after the first chai, before the laptop opens — is easier to keep than one that waits for a free minute.

Why this question matters for working adults in India is not mysterious. The country runs on long hours, thin recovery, and a professional class that will download a wellness perk and then not use it. Gál and colleagues, in 2021, pooled 34 mindfulness-app trials and found a post-test stress effect of $g = 0.46$ and a wellbeing effect of $g = 0.29$. That is a small-to-moderate bump at the finish line of the programme. Linardon and colleagues, in 2024, updated the depression and anxiety picture for mindfulness apps ($g = 0.24$ and $g = 0.28$) and said the quiet part out loud: higher-quality studies with longer follow-ups are needed. A companion 2024 review of 176 mental-health apps of any kind found small post-test effects and almost no usable evidence beyond thirteen weeks. A separate 2024 review of apps for stress found $g = 0.27$ at post-test, which shrank to $g = 0.10$ after trim-and-fill for small-study bias.

So the field has a launch-week number. It does not have a quarter-end number. Indian HR teams are being sold the launch-week number. This paper exists to stop that substitution.

The research question is simple. How much of the post-programme effect of mindfulness apps remains at three and six months, and does continued use matter? We pre-specified two moderators: continued use, and programme length. We pre-specified a decay ratio: follow-up *g* divided by post-programme *g*. A ratio of 1.00 means the effect did not fade. A ratio of 0.50 means half of it did. A ratio above 1.00 would be a sleeper gain, which would be lovely and which we did not expect.

We also pre-specified a brake. If fewer than five eligible trials shared the same outcome, the same comparator family, and the same time horizon, we would not pool. A null result is publishable. An invented pool is not.

What follows is the review. The technique is a rep. The programme is a programme. The question is what remains.

05 Methods

Eligibility

Population. Adults aged 18 and over, any country, any occupation. Non-clinical and workplace samples were preferred because EmotionApp serves working adults, not clinic lists. Clinical samples were not excluded a priori if the intervention was a mindfulness app and the outcomes included stress or wellbeing.

Intervention. A smartphone or tablet app whose primary content was mindfulness or closely related contemplative practice (attention to present-moment experience, body-based sitting, mindful movement framed as mindfulness). Web-only instructor-led courses without an app were excluded from the eligible set and cited only as adjacent evidence. Apps that mixed mindfulness with a dominant cognitive-behavioural curriculum were eligible only if mindfulness was the named core.

Comparator. Any control: wait-list, usual practice, attention-matched placebo app, or active comparison. Wait-list and inactive controls were treated as the primary comparator family for durability claims.

Outcomes. Primary: stress or perceived stress, and wellbeing, at three months or later. Distress composites were accepted when a trial did not report a standalone stress scale at follow-up. Secondary: anxiety and low mood scores when they illuminated durability, reported as context only.

Time horizon. Follow-up at or beyond three months from baseline, or at or beyond three months after the structured programme ended. Trials that stopped at post-test were used only to position the immediate effect, not to answer the durability question.

Design. Randomised trials and other controlled trials. Years 2012 to 2026. English-language full text required for extraction. Conference abstracts without a peer-reviewed paper were excluded.

Exclusions. In-person-only programmes. Pure yoga or mantra programmes without a mindfulness-app component. Trials in children. Trials whose only follow-up was shorter than twelve weeks from baseline. Anything we could not verify against a DOI or a stable URL was marked UNVERIFIED and kept out of the evidence tables.

Information sources and search

We searched PubMed, PsycINFO, Scopus, Web of Science, Google Scholar, the Clinical Trials Registry — India (CTRI), ClinicalTrials.gov, and OSF preprints. The search was run as a rapid systematic review on 22 September 2026. We did not pretend to be a librarian team with months of dual screening. We harvested every mindfulness-app trial cited by Gál, Ștefan and Cristea (2021) and by Linardon and colleagues (2024a, 2024b), then ran an update search for 2020–2026 with a follow-up filter. Two reviewers on the writing team independently checked eligibility for every candidate that survived the update search. Disagreements were resolved by discussion.

Full search strings.

PubMed:

(mindfulness[tiab] AND (app[tiab] OR smartphone[tiab] OR "mobile application"[tiab] OR mHealth[tiab] OR "mobile health"[tiab])) AND ("follow-up"[tiab] OR followup[tiab] OR durability[tiab] OR "3-month"[tiab] OR "6-month"[tiab] OR "three month"[tiab] OR "six month"[tiab] OR "long-term"[tiab] OR "long term"[tiab]) AND (randomi*[tiab] OR RCT[tiab] OR "controlled trial"[tiab])

PsycINFO:

TI,AB(mindfulness AND (app OR smartphone OR "mobile application" OR mHealth) AND (follow-up OR durability OR "3-month" OR "6-month" OR long-term) AND (randomi* OR RCT))

Scopus:

TITLE-ABS-KEY(mindfulness AND (app OR smartphone OR "mobile application" OR mhealth) AND (follow-up OR durability OR "3-month" OR "6-month") AND (randomi* OR rct))

Web of Science:

TS=(mindfulness AND (app OR smartphone OR "mobile application" OR mHealth) AND (follow-up OR durability OR "3-month" OR "6-month") AND (randomi* OR RCT))

Google Scholar:

"mindfulness app" OR "meditation app" randomised OR randomized "follow-up" OR "3-month" OR "6-month"

CTRI:

mindfulness AND (app OR mobile OR smartphone)

ClinicalTrials.gov:

Condition or intervention: mindfulness app. Filters: completed; studies with results; Adult.

OSF preprints:

mindfulness app follow-up random

Core terms across all sources: (mindfulness AND app) AND ("follow-up" OR durability) AND randomi*.

PRISMA 2020 flow

From the two prior app meta-analyses we started with 34 trials (Gál 2021) and 45 mindfulness-app trials (Linardon 2024a), with heavy overlap, plus the broader 176-trial mental-health-app set (Linardon 2024b) as a safety net. The update search added records from 2020–2026. After de-duplication we screened 412 unique titles and abstracts that mentioned a mindfulness or meditation app and a follow-up. We retrieved 96 full texts. We excluded 88 full texts: not an app (22), no follow-up at or beyond twelve weeks from baseline (41), no stress or wellbeing outcome (9), not a controlled trial (7), child or adolescent sample (4), unverifiable record (5). We included 8 primary trials in the narrative synthesis of durability, plus 3 positioning reviews. No trial was pooled. A meta-analysis is not yet possible on the pre-specified terms.

The flow is an update-and-harvest design, not a de-novo librarian dual-screen of every database from 2012. That is a limitation and we say so in the Discussion. It is also how most rapid evidence products for practitioners get built. The eligible set was small enough that a missed trial would have to be both real and invisible to two major meta-analyses and a 2026 update search.

Extraction fields

For every included trial we extracted: first author and year; country; language of delivery; n randomised and n at each follow-up; design and comparator; programme length in weeks; dose in sessions and minutes; delivery channel; guidance (none, light prompt, classes); outcome measure for stress and for wellbeing; time-points; effect size at post and at follow-up with 95 per cent confidence interval where published or computable; adherence or continued-use data; attrition.

Analysis plan

The pre-specified model was random-effects REML, Hedges *g* with 95 per cent CI, I^2, τ^2, 95 per cent prediction interval, Egger's test, trim-and-fill, leave-one-out, RoB 2, and GRADE. That model stays on the shelf. Fewer than five trials shared the same outcome, the same comparator family, and the same horizon. Pooling them would have produced a number that looked more precise than the evidence.

What we did instead:

1. Narrative synthesis by time horizon (three to four months; six months) and by comparator family (wait-list or inactive; active).
2. Decay ratio for every trial that published both a post-programme effect and a follow-up effect on the same construct: decay = *g*_follow-up / *g*_post. We report the three available wait-list pairs and their unweighted and n-weighted means.
3. Moderator commentary, not a meta-regression, on continued use and programme length.
4. RoB 2 judgements at domain level for the durability-relevant trials.
5. GRADE for four claims: stress at three to four months; stress at six months; wellbeing at three months or later; continued use as a moderator.

Effect-size signs are coded so that a positive *g* or *d* means benefit for the app (less stress, more wellbeing), including where the source paper reported a negative *d* for a symptom drop.

06 Results

What the positioning reviews already settled — and what they did not

Gál, Ștefan and Cristea (2021) gave the field its post-test number. Thirty-four trials, 7,566 people. Perceived stress $g = 0.46$ (95 per cent CI 0.24 to 0.68; 15 comparisons; $I^2 = 68$ per cent). Psychological wellbeing $g = 0.29$ (0.14 to 0.45; 5 comparisons; $I^2 = 0$ per cent). General wellbeing was not significant. Risk of bias was uncertain. Follow-up beyond the programme was not the question they asked.

Linardon, Messer, Goldberg and Fuller-Tyszkiewicz (2024a) updated mindfulness apps for low mood and anxiety. Forty-five trials. Low mood $g = 0.24$ (0.17 to 0.31). Anxiety $g = 0.28$ (0.21 to 0.35). Against active comparisons the effect disappeared. Their last sentence is the brief for this paper: higher-quality studies with longer follow-ups are needed.

Linardon, Torous, Firth, Cuijpers, Messer and Fuller-Tyszkiewicz (2024b), in *World Psychiatry*, pooled 176 mental-health-app trials of any kind. Depression $g = 0.28$. Generalised anxiety $g = 0.26$. Effects looked "robust at different follow-ups" until you read the small print. Beyond thirteen weeks they had three depression trials and a confidence interval that crossed zero ($g = 0.29$, −0.17 to 0.76). Mindfulness was the basis of about one in five of those apps. Durability of mindfulness apps, specifically, was not estimated.

Linardon's 2024 stress-app review (any mental-health app, not mindfulness-only) found $g = 0.27$ for perceived stress, which fell to $g = 0.10$ after trim-and-fill. Small-study effects are not a rumour in this literature. They are a regular guest.

Galante and colleagues (2021), in a large review of mindfulness-based *programmes* rather than apps, found that benefits against no intervention were visible one to six months after a course, with wide prediction intervals and GRADE ratings of moderate to very low. That is in-person and group work. It is not an app. We cite it as the ceiling the app literature is trying to approach, not as evidence that the app has arrived.

Study characteristics

Eight primary trials met the durability brief closely enough to enter the narrative. They are not interchangeable. They differ in dose, guidance, comparator, horizon, and whether anyone still opened the app after week eight.

Table 1. Study characteristics

Trial	Country	n rand.	Population	App dose	Length	Channel / guidance	Comparator	Stress or distress measure	Wellbeing measure	Time-points
van Emmerik 2018	Netherlands	377	General adults	40 exercises, self-help	8 wk	App, no guidance	Wait-list	General symptoms (GHQ)	Quality of life	Post; ~20 wk from baseline
Huberty 2019	USA	88	Stressed students	10 min/d recommended	8 wk	App, unguided	Wait-list	Perceived stress	— (mindfulness, self-compassion)	8 wk; 12 wk from baseline
Bostock 2019	UK	238	Company employees	45 × 10–20 min; mean 17 sessions	8 wk	App, unguided	Wait-list	Distress, job strain	Global wellbeing	8 wk; 16 wk from start
Hirshberg 2022	USA	662	School employees	4-wk wellbeing app	4 wk	App, unguided	Wait-list	Distress composite	Wellbeing battery	Post; 3 mo after programme

Trial	Country	n rand.	Population	App dose	Length	Channel / guidance	Comparator	Stress or distress measure	Wellbeing measure	Time-points
Bartlett 2022	Australia	211	Public-sector employees	10–20 min, 5 d/wk recommended	8 wk	App only vs app + 4 classes	Wait-list	Perceived stress; distress	Quality of life	3 mo from baseline; 6 mo
Taylor 2022	England	2,182	NHS staff	Unguided daily practice	4.5 mo ongoing	App, unguided	Active (health-info site)	Stress subscale	Short wellbeing scale	Across 4.5 mo
Kranenburg 2022	Netherlands	587	General adults	Structured 8-wk app	8 wk	App, no guidance	Control	Perceived stress	Exhaustion (secondary)	8 wk; 6 mo
Radin 2025	USA	1,458	Medical-centre employees	10 min/d recommended; mean ~5 min/d	8 wk	App, light prompting	Wait-list	Perceived stress; job strain	Work engagement	8 wk; 4 mo from randomisation

Language of delivery was English in six trials and Dutch in two. None delivered content in an Indian language. None recruited in India.

Forest-style data: study-level effects at follow-up

We do not pool. We line the numbers up so a reader can see why pooling would have been theatre.

Table 2. Follow-up effects on stress or distress (positive g/d = app better)

Two clocks are reported throughout. Clock A is months from randomisation or baseline. Clock B is months after the structured programme ended. Radin, Huberty, Bartlett and Kranenburg publish Clock A. Hirshberg publishes Clock B. Mixing them without a label is how an eight-week-post number gets sold as a three-month-post number.

Trial	Clock A (from randomisation)	Clock B (after programme)	Comparator	*g* or *d*	95% CI	n at FU (approx.)	Notes
Hirshberg 2022	~4 mo	**3 mo**	Wait-list	0.33	0.18 to 0.48	662	Cleanest Clock-B number
Huberty 2019	**12 wk**	4 wk	Wait-list	0.59	0.11 to 1.06	71	Between-group *g* from published means; small n
Bostock 2019	**16 wk**	8 wk	Wait-list	Sustained	not published as between-group *g*	~107 completers	Wellbeing and job strain held
Radin 2025	**4 mo**	~8 wk	Wait-list	0.71	0.59 to 0.84	1,074	Largest n; light prompting in weeks 1–8
van Emmerik 2018	~20 wk	**~3 mo**	Wait-list	Maintained, attenuated	not republished as *d* at FU	377 randomised; high loss	Post GHQ *d* = 0.68; most had stopped the app
Bartlett 2022 app-only	**3 mo and 6 mo**	~1 mo and ~4 mo	Wait-list	~0	ns	71 vs 70	Primary perceived-stress outcome null

Trial	Clock A (from randomisation)	Clock B (after programme)	Comparator	*g* or *d*	95% CI	n at FU (approx.)	Notes
Bartlett 2022 app+classes	**3 mo and 6 mo**	~1 mo and ~4 mo	Wait-list	0.21 (distress at 3 mo)	—	70 vs 70	Small; retained at 6 mo
Taylor 2022	**4.5 mo of continued use**	not post-programme	Active	Small (stress *b* = −0.31)	−0.47 to −0.14	2,182	Not a wait-list
Kranenburg 2022	**8 wk and 6 mo**	0 and ~4 mo	Control	~0	ns	587	Null at post and at 6 mo

Table 3. Follow-up effects on wellbeing

Trial	Horizon	Comparator	Effect	Notes
Hirshberg 2022	3 mo after programme	Wait-list	Secondary \	*d*
Radin 2025	4 mo from randomisation	Wait-list	Work engagement *d* = 0.35 (0.23 to 0.48)	Held or slightly grew from post *d* = 0.31
Bostock 2019	16 wk from start	Wait-list	Global wellbeing sustained within intervention arm	Wait-list had been offered the app by then
van Emmerik 2018	~3 mo after post	Wait-list	Psychological and environmental quality of life held; social QoL did not	Post psychological QoL *d* = 0.38
Taylor 2022	4.5 mo	Active	Small wellbeing coefficient (*b* = 0.14)	Against an information site, not silence
Mak 2018	3 mo	Active arms only	Wellbeing improved in the mindfulness arm	No inactive control; attrition severe (349 of 2,161)

Decay ratios

This is the number this paper set out to find.

Table 4. Decay ratio = follow-up effect / post-programme effect, same construct, wait-list pairs

Trial	Construct	Post	Follow-up	Decay ratio
Hirshberg 2022	Distress	0.53	0.33	**0.62**
Radin 2025	Perceived stress	0.85	0.71	**0.84**
Huberty 2019	Perceived stress	0.62	0.59	**0.94**

Unweighted mean decay ratio: **0.80**.

Sample-size-weighted mean (Hirshberg n = 662, Radin n ≈ 1,074 at follow-up, Huberty n = 71): 0.76.

Read that in plain English. In the trials that found an immediate effect and then measured again at three to four months, about four-fifths of that effect was still visible. That is not immortality. It is also not a pumpkin at midnight. Most of the lift is still there when the quarter closes. That does not make it permanent.

Two cautions sit on that mean. First, it is estimated only in trials that had something to decay. A trial that is already null at week eight cannot produce a decay ratio. Kranenburg and the Bartlett app-only arm are those trials. They belong in the denominator of hope, not in the numerator of durability. Second, Huberty's follow-up is twelve weeks from

baseline, which is only four weeks after the programme. It inflates the mean. Drop Huberty and the unweighted mean falls to 0.73. That is still "most of it," and it is more honest about three months.

Continued use

Put plainly: a skill you do not practise is a pamphlet. The skill also has to leave the app. The pause that counts is the one taken in the corridor before a hard review, not the one saved for Sunday. The rep is the sitting done, not the app downloaded.

The trials agree, clumsily.

Bostock and colleagues found that people who completed fewer than ten of the forty-five sessions did not move. People who did more than ten did. That is a dose–response sitting in a UK office.

Radin and colleagues found that five to ten minutes a day beat fewer than five minutes a day by 6.58 points on the perceived-stress scale during the programme. By four months the adherence contrast had faded. The early dose built the score. It did not keep building it after the prompting stopped.

Flett and colleagues, in a 2019 three-arm student trial and a 2020 discretionary-use trial, showed the engagement cliff in slow motion. During a ten-day prescribed period, students used the app on about eight of ten days. In the next thirty days of free access, fewer than one in five used it twice a week or more. The people who kept using it kept more of the gain. Persistence beyond week one in the 2020 cohort was 46 per cent.

van Emmerik and colleagues noted that most participants had discontinued or drastically reduced use by follow-up — and that the gains were still mostly there, if smaller. That is the interesting tension. Some of the skill may have transferred off the app. Some of the follow-up score may be a residual mood, not a living practice. The trials cannot tell those two apart.

Taylor and colleagues, against an active control over 4.5 months of *ongoing* access, found that days of practice mediated the stress drop. Engagement was the mechanism. The app was the pipe.

Linardon's 2023 attrition review of 70 mindfulness-app trials puts a number on the leak. Weighted attrition was 24.7 per cent, and 38.7 per cent in larger trials. Real-world retention for mental-health apps is worse still. The literature's dirty secret is that efficacy trials are already a selected sample of people who agreed to be studied. The plateau in the wild starts earlier.

Programme length

We pre-specified programme length as a moderator. We cannot test it. The durability trials are four weeks, eight weeks, or 4.5 months of ongoing access. That is not a dose ladder. It is three rungs with different shoes on.

Hirshberg's four-week programme still had a three-month post-programme effect. Radin's eight-week programme had a four-month-from-randomisation effect. Bartlett's eight-week app-only arm had nothing. Length is not the magic variable. Something about population, prompting, and whether anyone actually sat is doing more work than the calendar.

Heterogeneity

Even without a pool, the heterogeneity is the finding. Wait-list stress effects at three to four months run from 0.33 to 0.71 in the positive trials and from null to null in the others. Comparators differ. Prompting differs. Populations differ — students, teachers, NHS staff, medical-centre employees, television-recruited Dutch adults. I^2 on a forced pool would have been high. A 95 per cent prediction interval would have been wide enough to include both "useful" and "nothing." That is GRADE inconsistency, and it is earned.

Publication bias

Egger's test on three decay pairs would be a party trick. We did not run it. The broader app literature already tells you the direction of bias. Linardon's stress-app review lost two-thirds of its effect to trim-and-fill. Gál's stress estimate had I^2 = 68 per cent and uncertain risk of bias. Small, excited trials get published. Quiet nulls at six months have a longer walk to a journal. Kranenburg made the walk. That is why it sits in this paper with full dignity.

Risk of bias

Table 5. RoB 2 summary (durability trials)

Trial	Randomisation	Deviations	Missing data	Outcome measurement	Selection of result	Overall
van Emmerik 2018	Low	Some	High (post loss)	Some (unblinded self-report)	Some	High
Huberty 2019	Some	Some	High	Some	Some	High
Bostock 2019	Low	Some	Some	Some	Some	Some
Hirshberg 2022	Low	Some	Some	Some	Low (preregistered)	Some
Bartlett 2022	Low	Some	Some	Some	Low	Some
Taylor 2022	Low	Some	Some	Some	Low	Some
Kranenburg 2022	Low	Some	Some	Some	Low	Some
Radin 2025	Low	Some (unblinded prompting)	Some	Some	Low	Some

Every trial in this set measures the primary outcome by unblinded self-report. That is not a scandal. Stress and wellbeing are experiences. It is a systematic upward breeze when the comparator is a wait-list. People who received an app know they received an app. People on a wait-list know they are waiting. RoB 2 calls that "some concerns" in the measurement domain for every row. We agree.

Wait-list designs also leak at follow-up. By month four the control arm has often been given the app. Bostock's sixteen-week comparison is partly within-group for that reason. Radin kept the wait-list off the app until after the four-month questionnaires. That is why Radin's four-month contrast is the cleanest large-n number in the set.

GRADE

Table 6. GRADE

Claim	Certainty	Why
Stress or distress at 3–4 months vs wait-list, in trials that found a post effect	**Low**	Unblinded self-report; wait-list; inconsistency of magnitude; few trials; almost no Indian adults
Stress at 6 months, app-only vs control	**Very low**	Sparse; one large null (Kranenburg); one public-sector null (Bartlett app-only); no consistent wait-list pair with a positive 6-month *g*
Wellbeing at ≥3 months	**Low to very low**	Heterogeneous measures; small effects; wait-list breeze; one active-control small effect
Continued use as a moderator of durability	**Very low**	Few trials measured post-programme use with the same care they measured week-1 use; signals exist (Bostock, Flett, Taylor, Radin) but they do not support a coefficient

07 Discussion

PULL QUOTE

App-only arms were null at three and six months in two public-sector and general-population trials.

What the number means

Four-fifths is a respectable remaining fraction. It is not a miracle and it is not a rounding error.

In machine-learning terms, the eight-week programme is supervised training. The three-month follow-up is the test set. A model that keeps 80 per cent of its training-set lift on a later test set is a model you would ship, with monitoring. A model that is already null on the training set (Kranenburg; Bartlett app-only) will not suddenly become wise at month six. You do not fine-tune a network that never fitted.

The thought “I did an eight-week app, so I am sorted” is a thought. It is not a fact. The fact is the next sitting.

The course ends. The Monday review does not. Using the skill there is the programme.

The aim is not to be someone who downloaded a mindfulness app. The aim is to notice the jaw tighten when a client pushes back, and choose the next sentence. The app is a prompt. The point is the day you live between prompts.

In entrepreneurship terms, month-three retention is the business. Week-one activation is vanity. The trials that look good at week eight and still look decent at month three are the ones that got people to a minimum viable practice — ten sessions in Bostock, five minutes a day in Radin — and then, often, kept a light prompt in the system.

The decay ratios of 0.62, 0.84 and 0.94 are the range. Hirshberg’s 0.62 is the most conservative because it is three months after the programme ended, not three months after randomisation. Radin’s 0.84 is the most precise because the sample is huge, but the follow-up is only about eight weeks after the last structured week. Huberty’s 0.94 is the most flattering and the least persuasive, because the follow-up is only four weeks post-programme and the sample is 71 people. If you want one sentence for a slide, use this: most of the immediate effect is still visible a quarter later, in the trials that had an immediate effect, and some well-run trials never had an immediate effect at all.

That sentence is less pretty than a single pooled *g*. It is also true.

Limitations

We did not pool. That will annoy anyone who wanted a forest plot with a diamond. The pre-specified brake was fewer than five comparable trials. We honoured it. An exploratory diamond driven by one American employee trial of 1,458 people would have been a Radin-weighted average wearing a meta-analytic costume.

The search was a rapid harvest-and-update, not a months-long dual-librarian screen from 2012. We may have missed a small trial. We do not think we missed a large one. Gál and Linardon had already combed the field.

Almost every outcome is unblinded self-report against a wait-list. That inflates *g*. Linardon’s trim-and-fill on stress apps is the warning label.

Continued use is poorly measured after the structured weeks. Trials count sessions during the intervention and then shrug. The moderator we care about most is the one the literature is worst at capturing.

Women are over-represented. English and Dutch are the languages. High-income workplaces are the setting. Students appear more often than mid-career Indian engineers. External validity for Bengaluru, Pune, Hyderabad, Gurugram and Mumbai is an inference, not a finding.

Conflicts of interest exist in this field. Some trials received free premium access from the company that makes the app. One older trial had a co-author employed by the developer. We did not exclude those trials. We also did not pretend the incentive structure is invisible.

We did not name trademarked instruments in the public-facing claims. We used generic descriptions in the tables. Readers who need the exact scale can open the source DOI.

WHAT THIS MEANS

What this means for an Indian workplace programme

The eight-week licence lands on a Monday and is downloaded between meetings. That is a seed packet, not a wellbeing strategy.

Picture the rollout. Two thousand employees get the app. At week eight the pulse survey shows lower stress, and the chart makes a good slide. At month three nobody has sent a prompt since the course ended. The sales manager in Andheri who kept a short sitting after her morning call still has much of her gain. Her colleague's app has gone the way of his unused gym membership in Bandra. Two trials in public-sector and general-population adults found that the app alone did not move perceived stress at three or six months. Those trials were not in India. They are still a warning.

What an Indian programme can take from this literature is operational, not magical. Count the reps. Five to ten minutes, most days, beats a heroic weekend binge. Light human contact — a class, a prompt, a coach who notices on Thursday that you have gone quiet — is the difference between Bartlett's app-only null and Bartlett's app-plus-classes small effect that lasted to six months. Language matters and has not been tested. No trial in this set spoke Hindi, Tamil, Telugu, Bengali, Marathi or any of the other languages Indian professionals actually think in when they are tired. Data residency and a WhatsApp-first channel are not scientific findings. They are the conditions under which an Indian professional will still be there at week twelve. The evidence gap is the product specification.

Do not sell durability you have not measured in this population. Do measure it. Three months and six months. Stress and wellbeing. Continued-use logs, not just licences issued. When a vendor shows you the week-eight chart, ask for the month-three one. That is the whole job.

Research gaps, naming the India gap explicitly

The India gap is not a footnote. It is the hole in the map.

We searched CTRI and the major databases for a mindfulness-app randomised trial in Indian working adults with a three-to-six-month follow-up on stress or wellbeing. We did not find one. We found in-person Heartfulness and transcendental-meditation pilots, a

community cancer-care mindfulness group in West Bengal, a digital adaptation of a WHO stress programme for rural community health workers, and workplace papers that were not app RCTs. Those are neighbouring countries on the map. They are not this paper's question.

What India needs, if anyone is listening, is boring and expensive. A randomised trial. Working adults, not only students. An app or a WhatsApp-first practice stream in more than one Indian language. A wait-list or an active comparison. Perceived stress and a wellbeing scale. Follow-up at three months and at six months. Objective use logs. A protocol on a registry. A conflict-of-interest statement that a procurement team can read. Until that trial exists, every durability claim on an Indian pitch deck is an extrapolation from Wisconsin school employees, UCSF staff, Tasmanian public servants, Dutch television viewers and UK office workers. Some of those people are like some Indian professionals. Most of the context is not.

Other gaps are global. Six-month app-only data are a room with two chairs, and one of them is a null. Active-control durability is almost empty. Mechanisms — decentering, perseverative thinking, sleep — have better short-term evidence than they have quarter-end evidence. Men, shift workers, factory floors, and older professionals are under-sampled. Prediction intervals, not just confidence intervals, should become the default so that a workplace buyer can see how often the next trial will miss.

The field has enough week-eight papers. It has a shortage of week-twenty-six papers. That is not a branding problem. It is the next study.

08 FAQ

Q1 Do mindfulness apps still work three months later?

In the wait-list trials that had an effect at the end of the programme, about four-fifths of that effect was still visible at three to four months. Decay ratios were 0.62, 0.84 and 0.94. Some well-run trials had no effect to keep.

Q2 What happens at six months?

Too little evidence to pool. One Dutch general-population trial of 587 adults found no change in perceived stress at eight weeks or at six months. An Australian public-sector app-only arm was also null at six months.

Q3 Does it matter if people keep using the app?

Yes, during the programme. Fewer than ten sessions predicted no change in one workplace trial. Five to ten minutes a day beat under five minutes in the largest employee trial. After the programme, most people stop. The ones who do not stop keep more of the gain.

Q4 **Is this a substitute for seeing a clinician?**

No. This is a practice, a technique, a countable rep. It is not a diagnosis and it is not a clinic. People with severe or worsening distress need a human professional. An app can sit beside that. It cannot replace it.

Q5 **Has any of this been shown in Indian working adults at three to six months?**

No. Eligible mindfulness-app randomised trials with a three-to-six-month follow-up in Indian working adults: zero. That is the gap, not a detail.

09 References

Bartlett, L., Martin, A. J., Kilpatrick, M., Otahal, P., Sanderson, K., & Neil, A. L. (2022). Effects of a mindfulness app on employee stress in an Australian public sector workforce: Randomized controlled trial. *JMIR mHealth and uHealth, 10*(2), e30272. https://doi.org/10.2196/30272

Bostock, S., Crosswell, A. D., Prather, A. A., & Steptoe, A. (2019). Mindfulness on-the-go: Effects of a mindfulness meditation app on work stress and well-being. *Journal of Occupational Health Psychology, 24*(1), 127–138. https://doi.org/10.1037/ocp0000118

Flett, J. A. M., Hayne, H., Riordan, B. C., Thompson, L. M., & Conner, T. S. (2019). Mobile mindfulness meditation: A randomised controlled trial of the effect of two popular apps on mental health. *Mindfulness, 10*, 863–876. https://doi.org/10.1007/s12671-018-1050-9

Flett, J. A. M., Conner, T. S., Riordan, B. C., Patterson, T., & Hayne, H. (2020). App-based mindfulness meditation for psychological distress and adjustment to college in incoming university students: A pragmatic, randomised, waitlist-controlled trial. *Psychology & Health, 35*(9), 1049–1074. https://doi.org/10.1080/08870446.2019.1711089

Gál, É., Ștefan, S., & Cristea, I. A. (2021). The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: A meta-analysis of randomized controlled trials. *Journal of Affective Disorders, 279*, 131–142. https://doi.org/10.1016/j.jad.2020.09.134

Galante, J., Friedrich, C., Dawson, A. F., Modrego-Alarcón, M., Gebbing, P., Delgado-Suárez, I., Gupta, R., Dean, L., Dalgleish, T., White, I. R., & Jones, P. B. (2021). Mindfulness-based programmes for mental health promotion in adults in non-clinical settings: A systematic review and meta-analysis of randomised controlled trials. *PLOS Medicine, 18*(1), e1003481. https://doi.org/10.1371/journal.pmed.1003481

Hirshberg, M. J., Frye, C., Dahl, C. J., Riordan, K. M., Vack, N. J., Sachs, J., Goldman, R., Davidson, R. J., & Goldberg, S. B. (2022). A randomized controlled trial of a smartphone-based well-being training in public school system employees during the COVID-19 pandemic. *Journal of Educational Psychology, 114*(8), 1895–1911. https://doi.org/10.1037/edu0000739

Huberty, J., Green, J., Glissmann, C., Larkey, L., Puzia, M., & Lee, C. (2019). Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: Randomized controlled trial. *JMIR mHealth and uHealth, 7*(6), e14273. https://doi.org/10.2196/14273

Kranenburg, L. W., Gillis, J., Mayer, B., & Hoogendijk, W. J. G. (2022). The effectiveness of a nonguided mindfulness app on perceived stress in a nonclinical Dutch population: Randomized controlled trial. *JMIR Mental Health, 9*(3), e32123. https://doi.org/10.2196/32123

Linardon, J. (2023). Rates of attrition and engagement in randomized controlled trials of mindfulness apps: Systematic review and meta-analysis. *Behaviour Research and Therapy, 170*, 104421. https://doi.org/10.1016/j.brat.2023.104421

Linardon, J., Messer, M., Goldberg, S. B., & Fuller-Tyszkiewicz, M. (2024a). The efficacy of mindfulness apps on symptoms of depression and anxiety: An updated meta-analysis of randomized controlled trials. *Clinical Psychology Review, 107*, 102370. https://doi.org/10.1016/j.cpr.2023.102370

Linardon, J., Firth, J., Torous, J., Fuller-Tyszkiewicz, M., & others. (2024). Efficacy of mental health smartphone apps on stress levels: A meta-analysis of randomised controlled trials. *Health Psychology Review, 18*. https://doi.org/10.1080/17437199.2024.2379784

Linardon, J., Torous, J., Firth, J., Cuijpers, P., Messer, M., & Fuller-Tyszkiewicz, M. (2024b). Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety: A meta-analysis of 176 randomized controlled trials. *World Psychiatry, 23*(1), 139–149. https://doi.org/10.1002/wps.21183

Mak, W. W. S., Tong, A. C. Y., Yip, S. Y. C., Lui, W. W. S., Chio, F. H. N., Chan, A. T. Y., & Wong, C. C. Y. (2018). Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: Randomized controlled noninferiority trial. *JMIR Mental Health, 5*(4), e60. https://doi.org/10.2196/mental.8597

Radin, R. M., Vacarro, J., Fromer, E., Ahmadi, S. E., Guan, J. Y., Fisher, S. M., Pressman, S. D., Hunter, J. F., Sweeny, K., Tomiyama, A. J., Hofschneider, L. T., Zawadzki, M. J., Gavrilova, L., Epel, E. S., & Prather, A. A. (2025). Digital meditation to target employee stress: A randomized clinical trial. *JAMA Network Open, 8*(1), e2454435. https://doi.org/10.1001/jamanetworkopen.2024.54435

Taylor, H., Cavanagh, K., Field, A. P., & Strauss, C. (2022). Health care workers' need for Headspace: Findings from a multisite definitive randomized controlled trial of an unguided digital mindfulness-based self-help app to reduce healthcare worker stress. *JMIR mHealth and uHealth, 10*(8), e31744. https://doi.org/10.2196/31744

van Emmerik, A. A. P., Berings, F., & Lancee, J. (2018). Efficacy of a mindfulness-based mobile application: A randomized waiting-list controlled trial. *Mindfulness, 9*, 187–198. https://doi.org/10.1007/s12671-017-0761-7

About the author

ABOUT THE AUTHOR

Dr Atul Aswani; Mumbai Psychiatrist and trained as a Results Coach

Dr Atul Aswani is a Mumbai Psychiatrist and trained as a Results Coach. He leads EmotionApp, a non-clinical emotional-fitness coaching platform for Indian professionals. The programme delivers positive-psychology techniques as countable daily reps, with an AI coach and human hand-off, WhatsApp-first delivery, 23 Indian languages, India-resident data, and DPDP-compliant operations built under ISO 9001 and ISO 13485.

HOW TO CITE

Aswani A. The Mindfulness App Plateau: A Meta-Analysis of Effect Durability Beyond Eight Weeks. EmotionApp Research Paper 52. Mumbai: Tranquilo Meditek Sytems Private Limited; 2026. https://emotionapp.in/research/mindfulness-app-plateau-beyond-eight-weeks

LAST UPDATED

Last updated: 22 September 2026

NOTE

EmotionApp is a non-clinical emotional-fitness coaching platform. This paper describes practices, techniques and coaching. It is not therapy, treatment, diagnosis or medical advice.

SAFETY NOTE

If you are in distress, please contact a qualified professional or Tele-MANAS on 14416.