# A Report on "The Effect of Exercise on Suicidal Ideation and Behaviors: A Systematic Review and Meta-Analysis of Randomized Controlled Trials" by Fabiano et al. (2023)

Reviewer 2

February 07, 2026

v1

# Disclaimer

This report was generated by large language models, overseen by a human editor. It represents the honest opinion of The Catalogue of Errors Ltd, but its accuracy should be verified by a qualified expert. Comments can be made here. Any errors in the report will be corrected in future revisions.

> I am wiser than this person; for it is likely that neither of us knows anything fine and good, but he thinks he knows something when he does not know it, whereas I, just as I do not know, do not think I know, either. I seem, then, to be wiser than him in this small way, at least: that what I do not know, I do not think I know, either.
>
> Plato, *The Apology of Socrates*, 21d

> To err is human. All human knowledge is fallible and therefore uncertain. It follows that we must distinguish sharply between truth and certainty. That to err is human means not only that we must constantly struggle against error, but also that, even when we have taken the greatest care, we cannot be completely certain that we have not made a mistake.
>
> Karl Popper, 'Knowledge and the Shaping of Reality'

# Overview

**Citation:** Fabiano, N., Gupta, A., Fiedorowicz, J. G., Firth, J., Stubbs, B., Vancampfort, D., Schuch, F. B., Carr, L. J., and Solmi, M. (2023). The Effect of Exercise on Suicidal Ideation and Behaviors: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Journal of Affective Disorders*, Vol. 330, pp. 355–366.

**URL:** https://doi.org/10.1016/j.jad.2023.02.071

**Abstract Summary:** This systematic review and meta-analysis of 17 randomized controlled trials investigated the effect of exercise on suicidal ideation and behaviors in subjects with mental or physical conditions, finding no significant decrease in suicidal ideation or mortality but a significant decrease in suicide attempts.

**Key Methodology:** Systematic review and random-effects meta-analysis of 17 randomized controlled trials (RCTs) comparing exercise to control groups in subjects with mental or physical conditions.

**Research Question:** What is the effect of exercise on suicidal ideation and risk of suicidal behavior in patients with mental or physical illness?

# Summary

## Is It Credible?

This systematic review and meta-analysis by Fabiano et al. sets out to determine the impact of exercise on suicidal ideation and behaviors in individuals with mental or physical illnesses. The authors synthesize data from 17 randomized controlled trials (RCTs) to assess whether physical activity can serve as a viable intervention for suicidality. The article presents two primary headline claims: first, that "exercise did significantly decrease suicide attempts" compared to inactive controls (p. 355); and second, that there was "no significant difference" in suicidal ideation between exercise and control groups (p. 358). While the review addresses a critical gap in the literature, the credibility of these conclusions is strained by data limitations, statistical heterogeneity, and interpretative overreach regarding the safety and feasibility of the interventions.

The most striking finding—that exercise significantly reduces suicide attempts (Odds Ratio = 0.23)—rests on a precarious foundation. As the authors acknowledge, this conclusion is derived from a meta-analysis of only two studies (p. 358). While the effect size appears large, the reliance on such a "minimal amount of data points" (p. 360) makes the finding highly susceptible to the specific characteristics of those two trials rather than representing a generalized effect. Furthermore, there is a mathematical inconsistency in the reporting of this result. The authors report a *p*-value of 0.04 alongside a 95% confidence interval of 0.09–0.67 (p. 358). A confidence interval that far from the null value of 1.0 would typically correspond to a much lower *p*-value (approximately 0.004), suggesting a reporting error. While the result remains statistically significant in either case, this discrepancy, combined with the extremely small number of studies included in this specific analysis, suggests the finding should be treated as a signal for future research rather than a

definitive clinical proof.

The second major claim—that exercise has no significant effect on suicidal ideation—is compromised by extreme statistical heterogeneity. The meta-analysis of post-intervention suicidal ideation scores yielded an $I^2$ statistic of 94% (p. 360), indicating that nearly all the variance in the results is due to fundamental differences between the studies rather than chance. Pooling data with such high heterogeneity often results in a summary estimate that lacks clinical meaning. This issue is compounded by the conflation of measurement methods. The authors combined studies that actively screened for ideation using standardized scales with studies that merely reported it as an adverse event (p. 358). Passive reporting is known to substantially underestimate incidence compared to active screening. Consequently, the null result may not reflect a lack of efficacy, but rather the noise generated by mixing incompatible data sources.

Finally, the authors' assertion regarding the feasibility of exercise is an over-interpretation of the data. Based on a non-significant difference in discontinuation rates between exercise and control groups (OR = 0.85), the authors conclude that this "demonstrates that adherence... are not as infeasible as often assumed" (p. 365). However, the confidence interval for this outcome is wide (0.38–1.94) (p. 358), meaning the data are compatible with a scenario where exercise nearly doubles the dropout rate. A failure to find a statistically significant difference does not prove equivalence, and the data are too imprecise to "demonstrate" feasibility or "eliminate physicians' hesitation" (p. 365). Additionally, the classification of control groups appears inconsistent; while the article claims 82% of studies used inactive controls (p. 357), Table 1 lists several active comparators such as CBT and yoga (pp. 359–360), further muddying the clarity of the comparisons.

## The Bottom Line

The claim that exercise significantly reduces suicide attempts is mathematically significant but fragile, relying on only two studies and containing a numerical reporting inconsistency. The finding that exercise does not impact suicidal ideation is inconclusive due to extreme statistical heterogeneity and the conflation of active screening with passive adverse event reporting. Furthermore, the authors' conclusion that exercise is demonstrably feasible overstates the evidence, as the wide confidence intervals around discontinuation rates do not rule out substantial differences in adherence.

# Potential Issues

**Extreme statistical heterogeneity in the primary meta-analysis:** The article's primary meta-analysis on suicidal ideation is characterized by extreme statistical heterogeneity, which may render the pooled result uninterpretable. The authors report an $I^2$ statistic of 94% for the analysis of post-intervention suicidal ideation scores, indicating that nearly all of the variability in the results across studies is due to genuine differences between the studies rather than random chance (p. 360). Methodological guidance often considers pooling data to be inappropriate when heterogeneity is this high. While the authors use a random-effects model, the resulting summary estimate (Standardized Mean Difference = -1.09) represents an average across what appear to be fundamentally different studies. The accompanying 95% confidence interval is exceptionally wide (-3.08 to 0.90), spanning from a potentially large benefit to potential harm, which reflects this underlying inconsistency. The authors acknowledge this as a limitation, but the decision to present a single pooled estimate from such disparate data is conceptually questionable and risks providing a clinically meaningless average (p. 364).

**Inconsistent classification of control groups:** The article's classification of study control groups into "active" and "inactive" categories appears to be inconsistent with the data presented in its own table. The authors state that three studies (18%) used an active control and 14 studies (82%) used an inactive control (p. 357). However, a review of the control group descriptions in Table 1 suggests that at least five studies used what would typically be considered active controls: Abdollahi et al. used Cognitive Behavioral Therapy (CBT); Martiny et al. used wake therapy; Nyer et al. compared two doses of yoga; Taylor et al. compared boxfit training to trauma-informed yoga; and Haussleiter et al. used encouraged physical activity as a comparator (pp. 359–360). This discrepancy between the authors' summary count and the details in their table undermines the validity of any subgroup analyses based on

control type and raises questions about the interpretability of the main comparison of exercise versus "inactive controls."

**Conflation of disparate measurement methods for suicidality:** The review combines data on suicidal ideation that were collected using two fundamentally different methods: active screening with validated scales and passive reporting of adverse events. This introduces a significant and unquantified measurement bias. The authors acknowledge this limitation, stating that "The majority of studies which reported suicidal ideation as a binary outcome only recorded this as an adverse event, and did not actively screen for suicidal ideation among participants, which likely resulted in measurement error" (p. 358). Data from passive reporting is known to substantially underestimate the true incidence of events compared to active, systematic screening. The authors suggest this likely led to a "nondifferential misclassification" that would bias results toward the null hypothesis, but this remains an assumption (p. 358). Pooling data from such different measurement approaches introduces a severe form of heterogeneity that threatens the validity of the meta-analytic results for binary suicidality outcomes.

**Mathematical inconsistency in the sole significant clinical finding:** The article's only statistically significant finding on a clinical outcome—that exercise reduces suicide attempts—contains a notable mathematical inconsistency. The authors report an Odds Ratio (OR) of 0.23, a 95% Confidence Interval (CI) of 0.09 to 0.67, and a $p$-value of 0.04 (p. 358). However, a standard statistical calculation based on the reported OR and CI yields a $p$-value of approximately 0.004. This calculated $p$-value is an order of magnitude smaller than the reported $p$-value of 0.04. While the finding remains statistically significant in either case, as the CI does not include 1.0, this discrepancy suggests a reporting error in either the $p$-value or the confidence interval, reducing confidence in the precision of the article's most impactful result.

**Over-interpretation of inconclusive discontinuation data:** The article appears to over-interpret a non-significant finding on discontinuation rates to make a strong

clinical recommendation about adherence. The meta-analysis found no statistically significant difference in all-cause discontinuation between exercise and control groups (OR = 0.85, CI 0.38–1.94, $p = 0.86$, k = 12) (p. 358). Based on this, the authors conclude, "This demonstrates that adherence for those with mental or physical illness are not as infeasible as often assumed, which should eliminate physicians' hesitation to prescribe exercise" (p. 365). The use of the word "demonstrates" is not supported by the data. The confidence interval is very wide, indicating that the data is compatible with both a 62% reduction and a 94% increase in the odds of discontinuation in the exercise group. Such an imprecise result is inconclusive; it fails to show a significant difference but does not prove equivalence or feasibility. The data are insufficient to "demonstrate" feasibility and may not be strong enough to "eliminate" physician hesitation.

**Clerical and numerical inconsistencies:** The article contains several minor clerical and numerical errors that, while not altering the main conclusions, reduce confidence in the precision of the final reporting. First, there is an internal inconsistency in the participant counts listed in Table 1: the sum of the "Sample size" column is 1021, matching the total reported in the text, but the sum of the individual exercise and control arms within the same table is only 1011 (pp. 359–360). Second, there is a one-person discrepancy between the sample size for the main suicidal ideation analysis as reported in the text (n=248) and the sum of participants in the corresponding forest plot (n=247) (pp. 358, 360). Finally, there are minor inconsistencies between reported $p$-values and those calculated from the provided odds ratios and confidence intervals for outcomes like all-cause discontinuation, where the reported $p$-value is 0.86 but the calculated value is approximately 0.70 (p. 358).

# Future Research

**Standardization of suicidality measurement:** Future trials must adopt a uniform approach to measuring suicidality to resolve the issue of measurement conflation identified in this review. Research should prioritize the use of validated, active screening tools (e.g., the Columbia-Suicide Severity Rating Scale) as a primary or secondary efficacy outcome, rather than relying on passive adverse event reporting which introduces significant measurement error and bias.

**Powered trials for rare events:** To validate the preliminary finding regarding suicide attempts, future work requires large-scale RCTs specifically powered to detect changes in low-frequency events like suicide attempts. Given the low base rate of these behaviors, multi-site collaborations are necessary to achieve sufficient sample sizes, moving beyond the "minimal amount of data points" currently available.

**Delineation of control conditions:** Future meta-analyses and trials should rigorously distinguish between active and inactive control groups to ensure accurate effect size estimation. Research designs should explicitly compare exercise against defined active controls (e.g., psychotherapy, medication) separately from inactive controls (e.g., waitlist), avoiding the inconsistent classification that complicates the interpretation of current aggregate data.

isitcredible.com