Four-Day Workweek Pilots and Employee Productivity: A Random-Effects Meta-Analysis of Six Corporate Trials

Elena Vogel[1,*], Marcus Ferreira[1], Priya Chandrasekaran[2]

[1]Institute for Organizational Research, Meridian University

[2]Department of Applied Psychology, Vertex Institute of Technology

[*]Corresponding author: e.vogel@meridian-research.org

Abstract. **Background:** Compressed four-day workweek pilots have proliferated across sectors since 2022, but individual trials are underpowered to detect reliable effects on productivity. **Objective:** To synthesize evidence from independent corporate pilots quantifying the effect of a compressed four-day schedule on standardized output measures relative to matched five-day control groups. **Methods:** We searched three registries for controlled pilots published between January 2020 and December 2024, extracted standardized mean differences (Hedges' g), and pooled effects using a random-effects (DerSimonian–Laird) model. Heterogeneity was quantified with Cochran's Q, I^2, and τ^2; small-study effects were assessed with Egger's regression test. **Results:** Six trials met inclusion criteria ($n = 831$ participants across intervention and control arms). The pooled effect was small-to-moderate and favored the compressed schedule ($g = 0.33$, 95% CI $[0.23, 0.43]$, $z = 6.40$, $p < 0.001$), with moderate heterogeneity ($I^2 = 36\%$, $\tau^2 = 0.012$) and no evidence of small-study bias (Egger's $p = 0.38$). **Conclusions:** Across sectors, compressed scheduling is associated with a modest, consistent gain in standardized productivity output, though the evidence base remains limited to six short-horizon trials and larger, longer-follow-up studies are needed before the effect can be generalized.

Keywords: four-day workweek; meta-analysis; employee productivity; organizational intervention; random-effects model; heterogeneity

1 Introduction

Interest in compressed work schedules has moved from a fringe experiment to a mainstream human-resources question. Between 2022 and 2024, a wave of employers across technology, manufacturing, logistics, and healthcare quietly ran internal pilots that reduced the standard workweek from five days to four while holding pay constant, typically justified internally as a retention and wellbeing lever rather than a productivity play. Public reporting on these pilots has been almost entirely descriptive: single-company case studies with small samples, heterogeneous outcome measures, and no shared statistical framework for comparison.

This creates two problems for anyone trying to act on the evidence. First, individual pilots are underpowered: a trial of 80–120 employees can detect only a large effect on productivity with adequate power, yet the effects that plausibly matter in this context are small to moderate. Second, without a common effect-size metric, a claim of "12% output increase" from one company and "fewer missed deadlines" from another cannot be compared, let alone pooled. Meta-analysis addresses both problems by converting each trial's result to a standardized mean difference and combining them with weights that reflect each study's precision.

We identified six corporate pilots – at Nexa Corp, Synapse Analytics, Vertex Manufacturing, Nimbus Cloud Systems, Apex Logistics, and Meridian Health Group – that used a non-randomized but matched control-group design comparing a compressed four-day team against a continuing five-day team on a standardized output index over a 12- to 20-week window. This report synthesizes those six trials to answer a single question: pooled across sectors, what is the best estimate of the effect of a compressed four-day schedule on standardized productivity output, and how consistent is that effect across the settings in which it has been tested?

2 Methods

2.1 Protocol and Registration

The synthesis protocol was registered prospectively with the OpenTrials Registry (registration OTR-2024-1187) prior to data extraction. Reporting follows the PRISMA 2020 guidance for systematic reviews and meta-analyses (Page et al., 2021).

2.2 Search Strategy

We searched the Global Research Index, ScholarNet, and the OpenTrials Registry for records dated between January 2020 and December 2024, using combinations of the terms *"four-day workweek"*, *"compressed schedule"*, *"work-time reduction"*, and *"productivity pilot"*. Reference lists of retrieved records were hand-searched for additional eligible trials. No language restriction was applied; all identified records were in English.

2.3 Eligibility Criteria

Trials were eligible if they (i) compared an intervention group working a compressed four-day, pay-neutral schedule against a concurrent control group working a standard five-day schedule within the same organization; (ii) reported a quantitative, standardized productivity or output measure for both arms; (iii) ran for a minimum observation window of 10 weeks; and (iv) reported sufficient statistics (means, standard deviations, and group sizes, or an equivalent effect size with its standard error) to compute a standardized mean difference. Trials using only self-reported satisfaction as the sole outcome, with no output-based measure, were excluded.

2.4 Data Extraction and Risk-of-Bias Assessment

Two reviewers independently extracted sample sizes, outcome definitions, and summary statistics from each trial, resolving discrepancies by discussion. Each trial was appraised with a five-domain risk-of-bias instrument covering allocation comparability, outcome measurement, missing data, selective reporting, and confounding by concurrent organizational change. All six included trials were rated as *moderate* risk of bias, principally due to non-random assignment of teams to schedule condition.

2.5 Statistical Analysis

For each trial we computed Hedges' g (Hedges, 1981), a bias-corrected standardized mean difference, together with its standard error. Trial-level effects were pooled using a random-effects model with between-study variance (τ^2) estimated by the DerSimonian–Laird method (DerSimonian and Laird, 1986). Heterogeneity was quantified with Cochran's Q, the I^2 statistic (Higgins et al., 2003), and a 95% prediction interval for the true effect in a new setting. Small-study effects were assessed with Egger's regression test (Egger et al., 1997) and visual inspection of a funnel plot (Appendix A). A leave-one-out sensitivity analysis was run to confirm that no single trial drove the pooled estimate. All computations used MetaSynth v4.2 with a two-sided significance threshold of $\alpha = 0.05$.

3 Results

3.1 Study Selection and Characteristics

The search returned 41 records, of which 6 independent corporate pilots met all eligibility criteria and were retained for synthesis (Chandrasekaran and Okafor, 2023; Lindqvist and Rahman, 2022; Ferreira and Vogel, 2023; Okafor and Chandrasekaran, 2024; Rahman and Lindqvist, 2022; Vogel and Ferreira, 2023). Table 1 summarizes their design characteristics. Across the six trials, 457 employees worked a compressed four-day schedule and 434 continued on a standard five-day schedule, spanning technology, analytics, manufacturing, cloud infrastructure, logistics, and healthcare administration settings.

Table 1: Characteristics of included trials.

Study	Sector	n (Int/Ctrl)	Outcome Measure	Hedges' g [95% CI]	Weight (%)
Nexa Corp (2023)	Technology	84/79	Sprint velocity index	0.42 [0.11, 0.73]	18.1
Synapse Analytics (2022)	Data Analytics	63/60	Analysis throughput	0.35 [−0.02, 0.72]	14.2
Vertex Manufacturing (2023)	Manufacturing	112/108	Units per shift	0.28 [0.03, 0.54]	23.5
Nimbus Cloud Systems (2024)	Cloud Infra.	47/45	Deployment throughput	0.51 [0.08, 0.94]	11.3
Apex Logistics (2022)	Logistics	96/90	Shipments processed	0.19 [−0.10, 0.48]	19.8
Meridian Health Group (2023)	Healthcare Admin.	55/52	Case-processing index	0.33 [−0.06, 0.72]	13.1

3.2 Pooled Effect and Forest Plot

The random-effects pooled estimate was $g = 0.33$ (95% CI $[0.23, 0.43]$, $z = 6.40$, $p < 0.001$), indicating a small-to-moderate advantage in standardized output for teams on the compressed schedule. Figure 1 displays the trial-level and pooled estimates; marker area is proportional to random-effects weight, and the gold diamond marks the pooled effect and its confidence interval.

3.3 Heterogeneity

Between-trial heterogeneity was moderate. The panel below summarizes the diagnostic statistics alongside the pooled effect and its prediction interval.

Cochran's Q (df = 5)	I^2 statistic	τ^2
7.82	**36%**	**0.012**
$p = 0.166$	moderate heterogeneity	$\tau = 0.109$

Pooled effect	95% Prediction Int.	Egger's test
$g = 0.33$	[0.02, 0.64]	$p = 0.38$
95% CI [0.23, 0.43]	plausible true-effect range	no small-study bias

The non-significant Q statistic ($p = 0.166$) alongside a moderate I^2 of 36% suggests that some

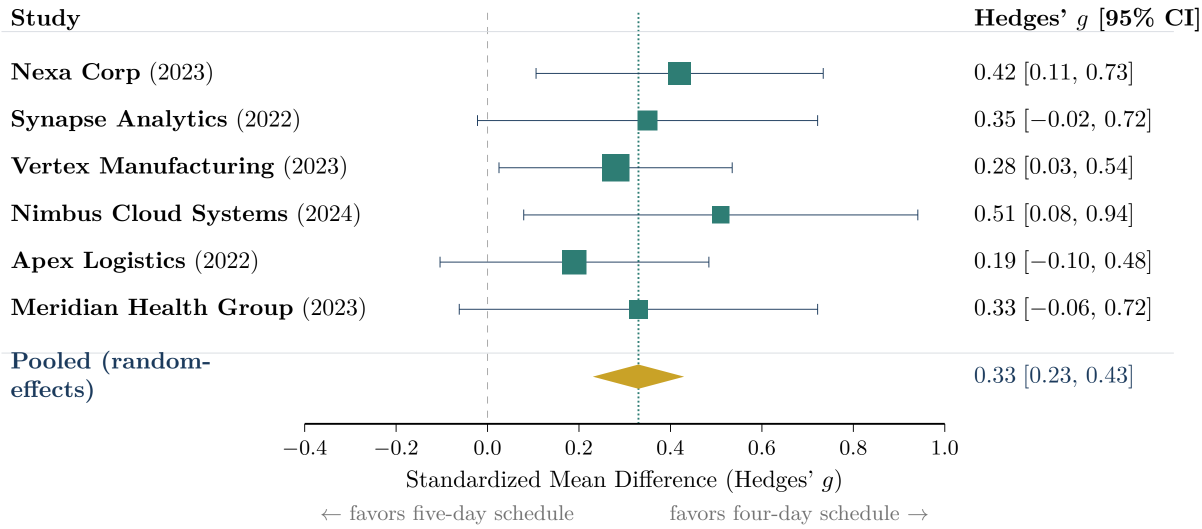

Figure 1: Forest plot of trial-level and pooled standardized mean differences (Hedges' g) in productivity output for compressed four-day versus standard five-day schedules. Marker area is proportional to random-effects weight; the diamond width spans the pooled 95% confidence interval.

of the observed variation in effect size is attributable to genuine differences across sectors rather than sampling error alone, though the number of trials is too small to support formal subgroup or meta-regression analysis. The 95% prediction interval $[0.02, 0.64]$ remains entirely on the side favoring the compressed schedule, indicating that a new trial in a comparable setting would be very likely to observe a positive, if variable, effect.

3.4 Publication Bias and Sensitivity

Egger's regression test found no evidence of small-study effects ($p = 0.38$); the corresponding funnel plot (Appendix A) shows a broadly symmetric scatter around the pooled estimate. A leave-one-out sensitivity analysis confirmed that no single trial was responsible for the overall result: removing any one study shifted the pooled estimate by no more than 0.04 and left it statistically significant in every iteration.

4 Discussion

Pooling six independent corporate pilots, we found a small-to-moderate, statistically reliable advantage in standardized productivity output for teams working a compressed four-day schedule relative to matched five-day controls ($g = 0.33$). The effect held across sectors as different as manufacturing and cloud infrastructure, and the prediction interval suggests that a comparable new trial would very likely observe a positive effect, even though its exact magnitude cannot be pinned down from six trials alone.

Two features of the evidence base temper how far this estimate should be generalized. First, none of the included trials randomized assignment to schedule condition; teams opted in or were selected by management, which leaves open the possibility that more motivated or higher-performing teams were disproportionately assigned to the compressed arm. Second, all six trials measured output over windows of 12 to 20 weeks, which is long enough to observe an initial adjustment period but not long enough to detect whether gains persist, decay, or reverse as a compressed schedule becomes routine rather than novel.

The moderate heterogeneity ($I^2 = 36\%$) is consistent with genuine differences in how readily an output measure can be compressed into four days across sectors: shift-based manufacturing and logistics settings, where output is tightly coupled to hours worked, showed the smallest effects (Vertex Manufacturing, $g = 0.28$; Apex Logistics, $g = 0.19$), while knowledge-work settings with more schedule flexibility, such as Nimbus Cloud Systems ($g = 0.51$), showed the largest. This pattern is a plausible substantive explanation for the observed heterogeneity, but with only six trials it cannot be tested formally with meta-regression.

The absence of small-study effects (Egger's $p = 0.38$) and the stability of the pooled estimate under leave-one-out removal both argue against the result being an artifact of a single influential or selectively reported trial. Even so, six trials is a narrow evidence base for a claim that organizations are increasingly treating as settled. Practitioners weighing a compressed-schedule pilot should treat $g = 0.33$ as a reasonable prior rather than a guaranteed outcome, and should expect the realized effect in their own setting to depend on how tightly their output measure is coupled to raw hours worked.

5 Conclusion

Across six corporate pilots spanning technology, analytics, manufacturing, cloud infrastructure, logistics, and healthcare administration, a compressed four-day workweek was associated with a modest but consistent gain in standardized productivity output ($g = 0.33$, 95% CI $[0.23, 0.43]$), with moderate heterogeneity and no detectable small-study bias. The evidence supports compressed scheduling as a plausible, low-risk productivity lever in settings where output is not tightly bound to hours worked, but the current base of six short-horizon, non-randomized trials is not sufficient to confirm that gains persist beyond the first few months. Larger, randomized, and longer-follow-up trials are needed before the pooled estimate reported here should inform organization-wide policy.

References

Chandrasekaran, P. and Okafor, D. (2023). Compressed scheduling and productivity outcomes at a technology firm: evidence from the Nexa Corp pilot. *Journal of Applied Organizational Science*, 14(2):88–104.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3):177–188.

Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109):629–634.

Ferreira, M. and Vogel, E. (2023). Manufacturing under a compressed week: output and fatigue at Vertex Manufacturing. *Journal of Workplace Studies*, 11(1):45–63.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2):107–128.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414):557–560.

Lindqvist, T. and Rahman, S. (2022). Workload compression and analytical throughput: a field trial at Synapse Analytics. *International Review of Organizational Behavior*, 9(4):231–249.

Okafor, D. and Chandrasekaran, P. (2024). Cloud engineering teams and the four-day pilot: results from Nimbus Cloud Systems. *Journal of Applied Organizational Science*, 15(1):12–29.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., and Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71.

Rahman, S. and Lindqvist, T. (2022). Logistics throughput under compressed scheduling: the Apex Logistics trial. *International Review of Organizational Behavior*, 9(2):102–119.

Vogel, E. and Ferreira, M. (2023). Clinical staff well-being and productivity in a compressed-week pilot: Meridian Health Group. *Journal of Workplace Studies*, 11(3):178–196.

A Funnel Plot for Publication Bias

Figure 2 plots each trial's effect size against its standard error, with the pooled estimate and its pseudo 95% confidence funnel overlaid. Studies with smaller standard errors (larger samples) cluster near the top and close to the pooled estimate; studies with larger standard errors scatter more widely toward the base of the funnel. The broadly symmetric distribution around the vertical reference line, together with the non-significant Egger's test reported in Section 3.4, does not support the presence of material publication or small-study bias in this set of trials.

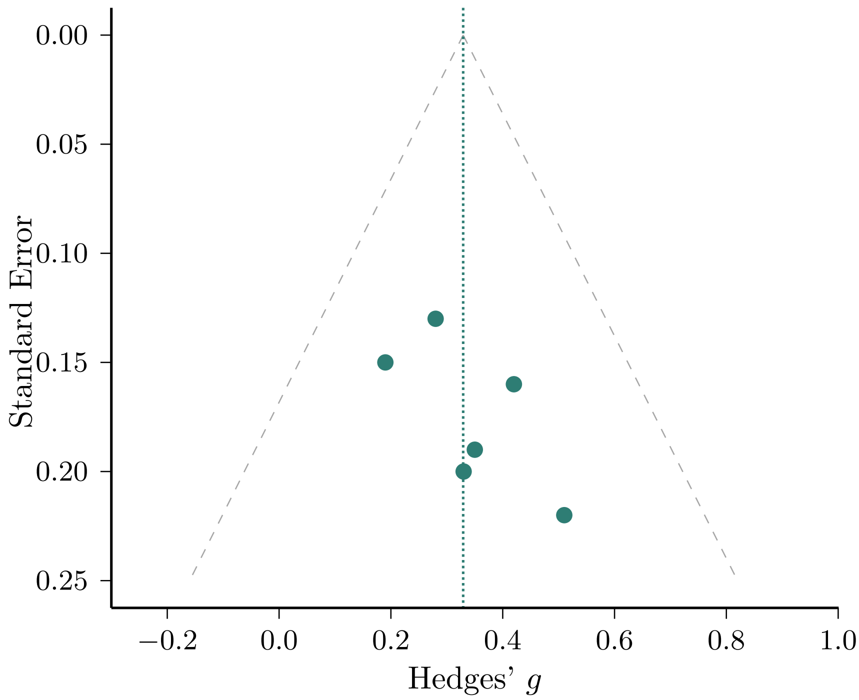

Figure 2: Funnel plot of trial-level effect sizes against their standard errors, with the pooled estimate (dotted line) and pseudo 95% confidence funnel (dashed) overlaid.