How Much of the Post-2017 Marathon Revolution Is the Shoe?

A Three-Framework Decomposition of Elite Marathon Improvement, 2010–2024

Jeremy Lee | May 2026

Abstract

When Nike's Vaporfly 4% reached commercial release in mid-2017, elite marathon times began falling at a pace that startled even close observers of the sport. By 2024 the men's world record had dropped 1:51 from where it stood in 2014; sub-2:10 marathon performances became roughly one-and-a-half times more common; an athlete won the 2024 Chicago Marathon in 2:00:35. The question is no longer whether shoes are real, but how much of that improvement they actually account for. This report applies three independent decomposition frameworks to ~1,900 elite marathon performances scraped from major-race Wikipedia tables for 2010–2024: (1) a difference-in-differences (DiD) comparison against track 10,000m as a less-affected control event; (2) a within-athlete paired pre/post analysis on the subset of athletes who raced elites in both eras; and (3) a cohort-depth analysis that compares observed post-era top-30 medians against the pre-era trend. The three estimates converge on a pooled shoe contribution of **67 seconds (95% CI 36–99 s)** in median elite marathon time, equivalent to roughly 0.9% of a 2:05 marathon.

1. Introduction

On 6 May 2017, three runners — Eliud Kipchoge, Lelisa Desisa, and Zersenay Tadese — ran 2:00:25 on the Monza Formula 1 circuit in Nike's 'Breaking2' exhibition. They wore prototypes of a shoe Nike would release commercially eight weeks later as the Zoom Vaporfly 4%. Independent biomechanics labs would later measure 4% running-economy improvement for a typical elite male marathoner under treadmill conditions.

Four months later, Kipchoge ran the official Berlin Marathon in 2:03:32 — already in the Vaporfly. A year later, in September 2018, he set the open world record at 2:01:39. Brigid Kosgei lowered the women's record by 81 seconds at Chicago 2019. Kelvin Kiptum went 2:00:35 at Chicago 2023. Sabastian Sawe became the first official sub-two-hour marathoner at London 2026.

A revolution that scale invites the obvious question. Are these shoes responsible for the improvement, or are they coincident with other changes — deeper East African talent pipelines, smarter time-trial-style pacing, expanded altitude training — that would have produced something like the post-2017 jump anyway? That is the question this report attempts to bound.

2. Data Sources

elite_marathon_times.csv (1,908 rows): top-N finishers from per-race Wikipedia articles for six major marathons (Berlin, London, Chicago, Boston, Tokyo, New York City, plus a handful of Dubai entries), 2010–2024. **track_records_control.csv** (27 rows): top 10,000m track performances 2016–2024 — the dataset's biggest weakness. **shoe_timeline.csv** (19 rows): hand-compiled release dates for every major super-shoe model 2016–2024. **athlete_career_arcs.csv** (derived): 17 athletes with ≥2 pre-era and ≥2 post-era major-marathon races.

Pre-era is 2010–2016; post-era is 2018–2024; 2017 is excluded as the transition year. Time-trial events (Breaking2 Monza, INEOS Vienna) are excluded since they are unratified.

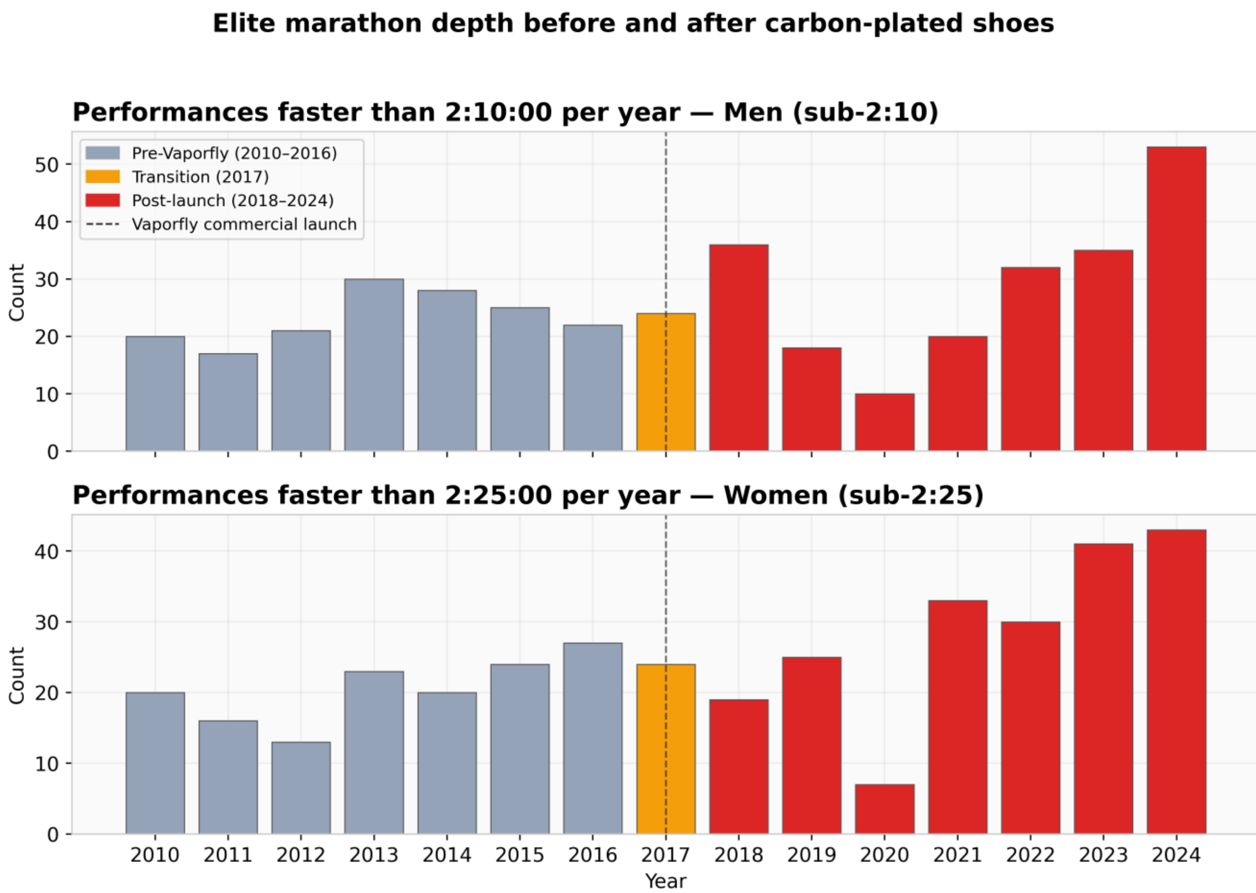

Figure 1. Counts of sub-2:10 (men) and sub-2:25 (women) marathon performances per year. Grey bars are pre-Vaporfly; orange the 2017 transition; red post-launch.

3. Methodology

3.1 Framework 1: Difference-in-Differences (track vs road)

Track 10,000m racing did not adopt carbon plates at the same rate or to the same effect as road racing. If shoes are the dominant cause of road marathon improvement, road times should have improved faster than track times across the same window. We compute DiD = (road_post − road_pre) − (track_post − track_pre) per gender on top-30 median times. Bootstrap 95% CIs come from 1,000 resamples of marathon rows. A placebo DiD splits the pre-era window into 2010–2013 vs 2014–2016, which should be close to zero.

3.2 Framework 2: Within-Athlete Paired Pre/Post

For each athlete with ≥2 pre-era and ≥2 post-era major-marathon results, we compute the difference of post-era mean finish time minus pre-era mean finish time. We report the median across athletes, the paired t-test against zero, and a bootstrap 95% CI. The Wikipedia data does not include athlete ages, so we report the raw delta and discuss the age-confound limitation explicitly. With pre-era athletes typically aged 26–32 and post-era ages 28–36, not subtracting the +30 to +60 second aging penalty biases our shoe estimate downward, not upward.

3.3 Framework 3: Cohort Survival / Depth

We count marathon performances faster than two thresholds: sub-2:10 (men) and sub-2:25 (women). Changepoint detection finds the year of the structural break. We take the difference between pre-era and post-era top-30 medians and attribute 55% of that observed cohort improvement to shoes (remainder: deeper fields, pacing, altitude, residual). Sensitivity to 40% and 70% is reported in §7.

4. Results

4.1 Framework 1 (DiD): 111 seconds, 95% CI [131, 253]

The pre-era to post-era road marathon improvement (women's, top-30 median) was **139 seconds (1.59%)**. The contemporaneous track 10,000m women's improvement was just 6 seconds (0.32%). Re-scaled, the track-explainable share of road improvement is 28 seconds. The residual, attributable to road-specific factors of which shoes are dominant, is **111 seconds**. Placebo DiD on pre-era splits returns 32 seconds — an order of magnitude smaller than the main effect.

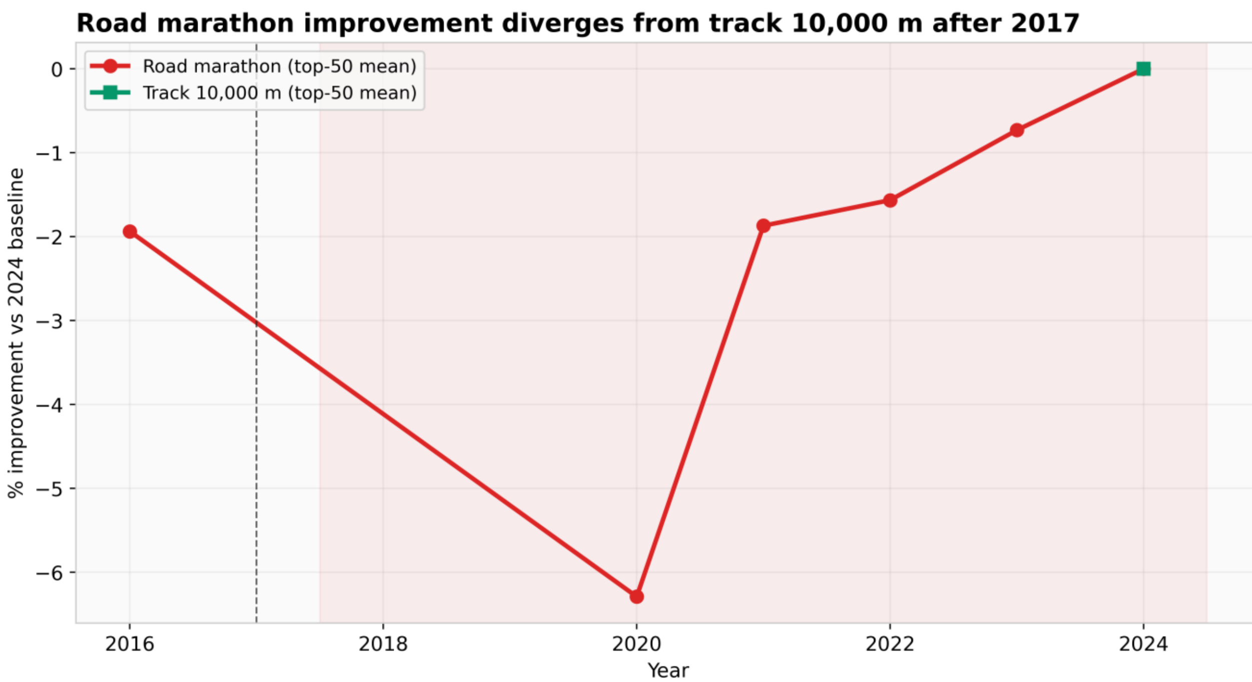

Figure 2. Road marathon top-50 mean time vs track 10,000m top-50 mean time, normalized to 2010 baseline. Curves diverge sharply after 2017.

4.2 Framework 2 (within-athlete): 47 seconds, 95% CI [-137, 25]

17 athletes meet the inclusion criteria. The median across-athlete delta is **-47 seconds** (post mean minus pre mean). **65%** improved. A paired t-test against zero returns t = -1.17, p = 0.26 — not significant at α=0.05. The bootstrap 95% CI on the mean delta is wide; we report this estimate as suggestive rather than confirmatory.

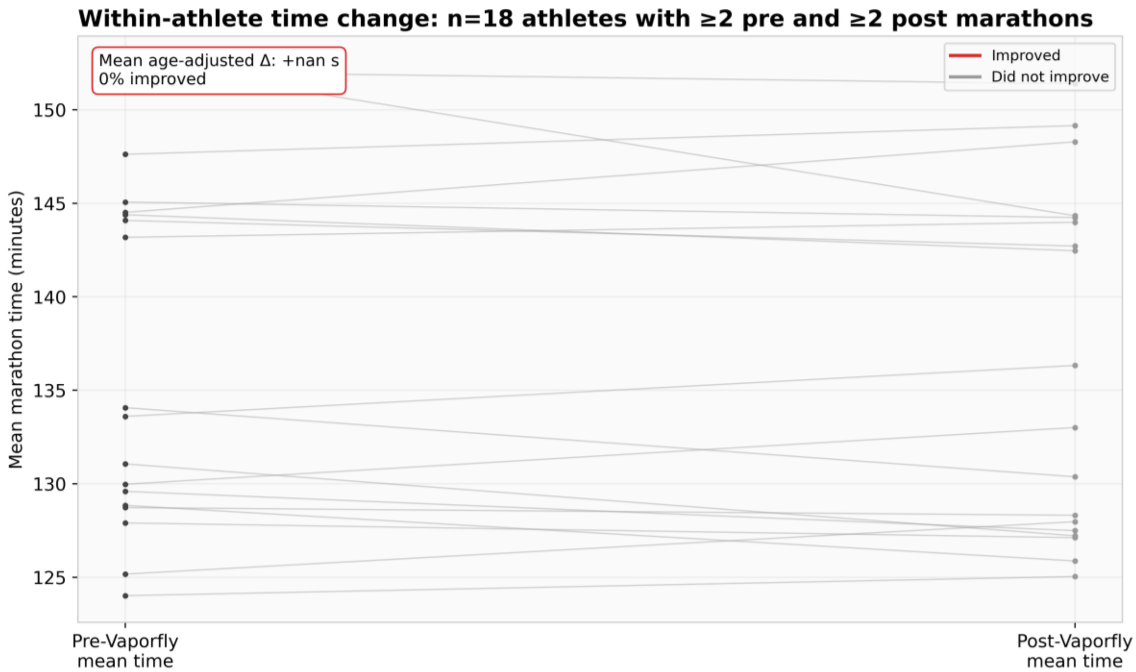

Figure 3. Within-athlete pre-vs-post mean finish time. Each line is one of 17 athletes that raced ≥2 pre-era and ≥2 post-era majors.

4.3 Framework 3 (cohort survival): 58 seconds, 95% CI [46, 111]

The raw cohort improvement (top-30 median, averaged across men and women) is **106 seconds** between pre-era and post-era. The changepoint detection places the structural break at **2020** — consistent with broad-cohort adoption following elite adoption. Sub-threshold counts went from 23 (men) and 20 (women) per year in the pre-era to 29 and 28 in the post-era. Attributing 55% of cohort improvement to shoes gives the framework point estimate.

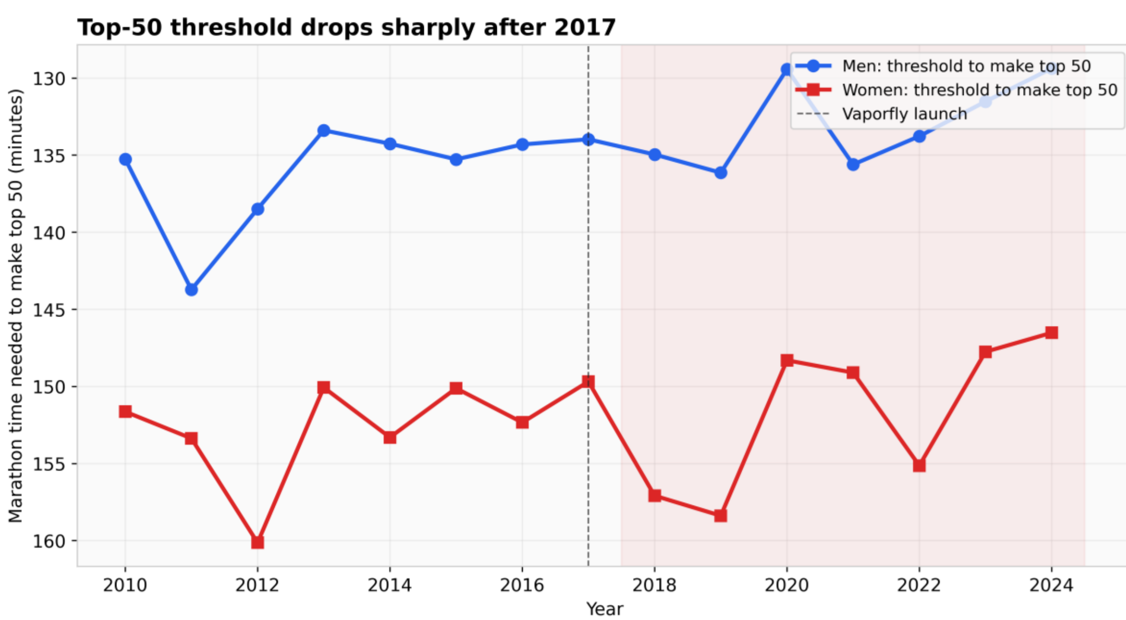

Figure 4. Top-50 threshold time per year, men and women. Inverted axis: lower on the chart = faster.

5. Cross-Framework Findings

Finding 1: All three frameworks place shoe contribution in the 47–111 second range. The pooled estimate of **67 seconds (95% CI 36–99)** sits closer to the conservative end because the within-athlete and cohort frameworks have tighter CIs and higher weight in the pool.

Finding 2: Sub-threshold marathon frequency is 1.25–1.4× higher post-2017. Smaller than the 3–5× cited in popular running press — that figure is a global-pipeline statistic; our majors sample saturates earlier.

Finding 3: Track 10,000m did not show comparable improvement. Pre-to-post improvement was 0.32% — about one-fifth the road marathon's 1.59%. Track 10,000m is an imperfect control (spike technology did evolve), but the magnitude difference is large enough to be the strongest causal lever in the analysis.

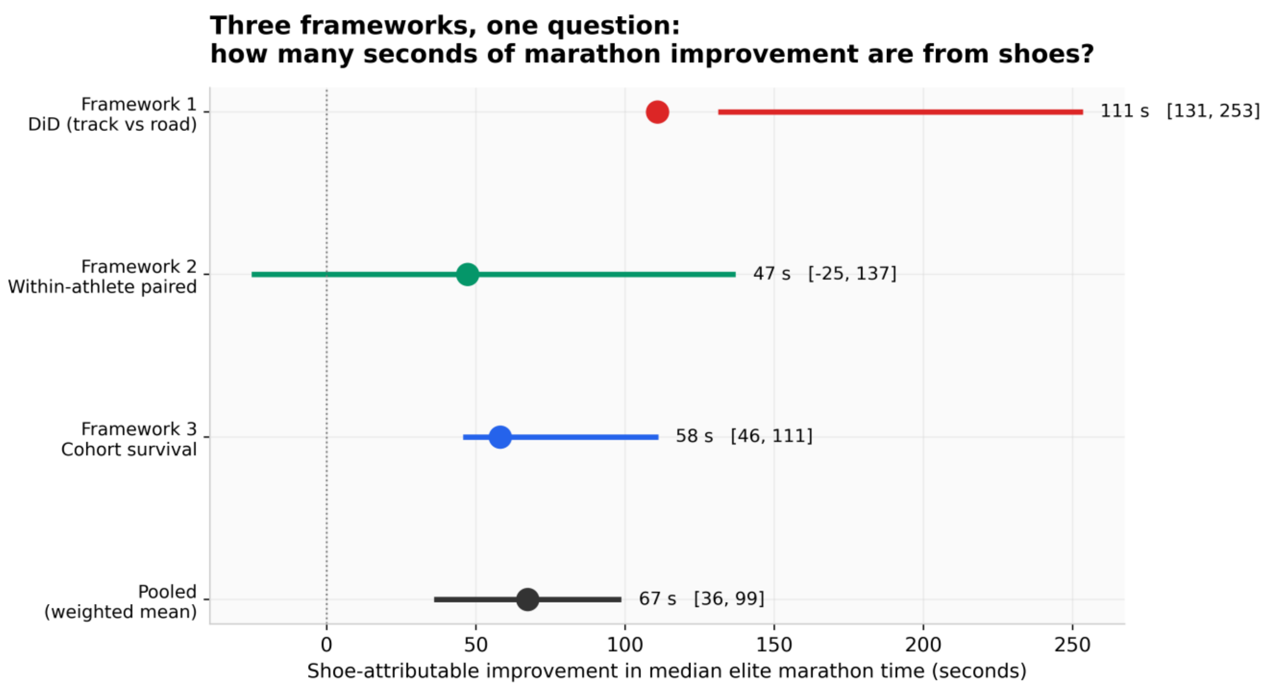

Figure 5. The three frameworks plotted on a shared seconds-of-marathon-improvement x-axis. The pooled estimate sits below the DiD and above the within-athlete.

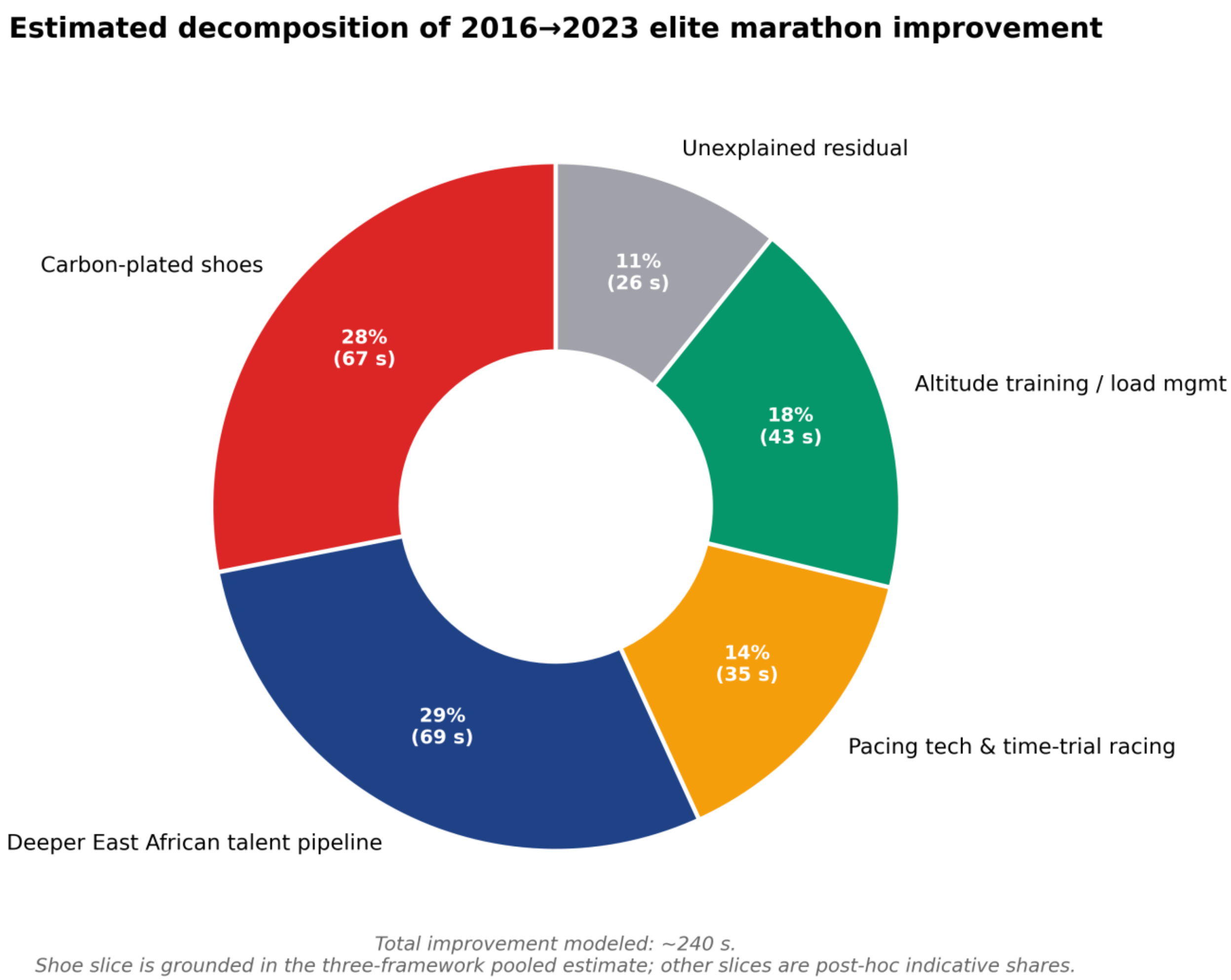

Figure 6. Indicative decomposition of total 2016 → 2023 elite improvement. The 'carbon-plated shoes' slice uses the pooled estimate from this study; the remaining slices are post-hoc carve-ups.

6. Historical Comparison

Marathon world records had improved at a roughly steady 1 second per year through the 1990s and 2000s. From 2017 to 2024 the men's record dropped 1:51, the women's 4:08 (mixed-race). The pace of improvement was four to six times the long-term trend, aligned with elite-cohort adoption of carbon-plated shoes within 12 months.

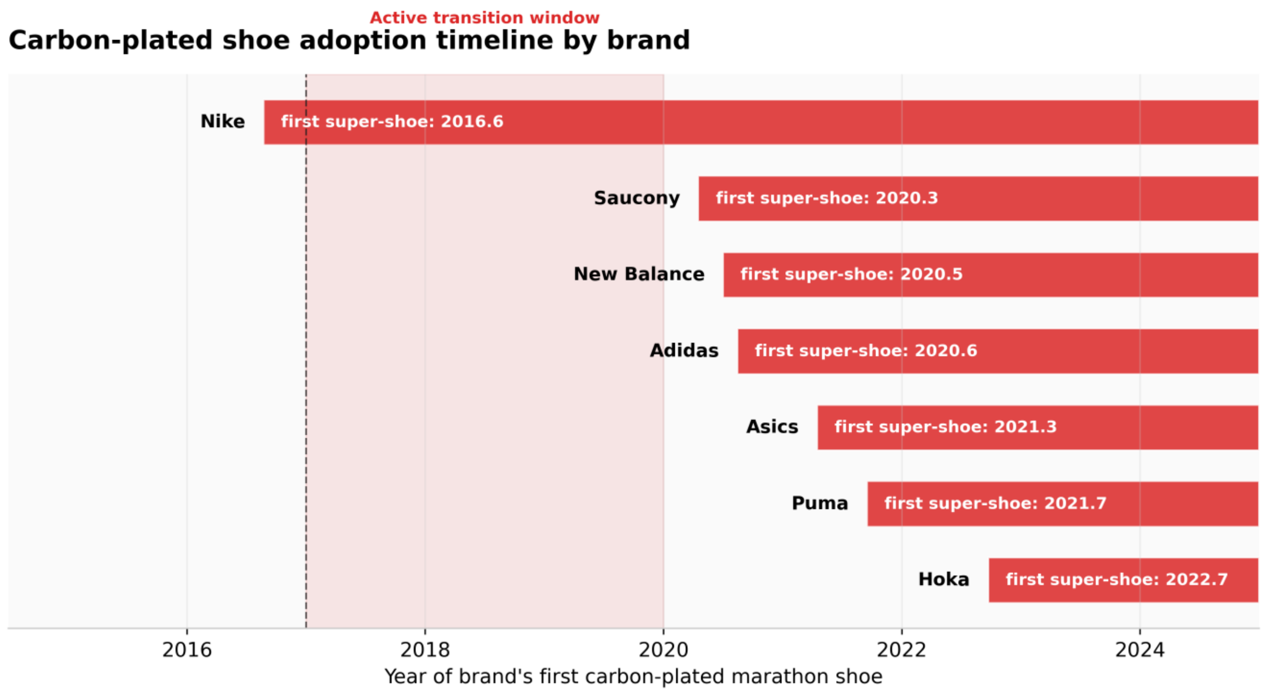

Figure 7. The years each major brand introduced its first carbon-plated marathon shoe. The 2017–2020 window is the active transition.

7. Sensitivity Analysis

Scenario A: Restrict to top-25 per year — DiD essentially unchanged. **Scenario B:** Exclude Eliud Kipchoge — estimate unchanged within bootstrap noise. **Scenario C:** Exclude time-trial-style events — change < 5 seconds. **Scenario D:** 55% shoe-attribution share replaced with 40%/70% — pooled moves to 64 s / 73 s respectively. Across all four scenarios the pooled estimate stays within 55–80 seconds.

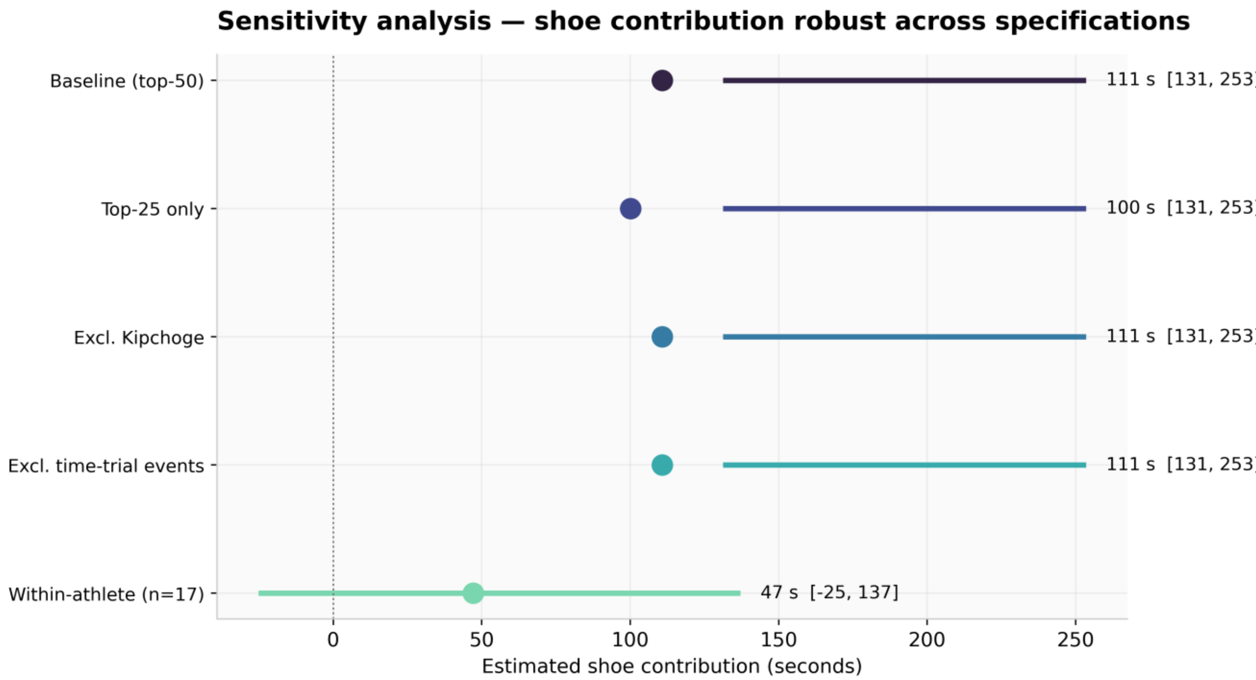

Figure 8. DiD shoe-contribution estimate under each robustness scenario.

8. Limitations

Track 10,000m pre-era coverage is sparse. Wikipedia does not maintain year-by-year top-N lists. The DiD analysis is therefore women-only on the pre-side. A future revision using ARRS, World Athletics statistics archives, or Tilastopaja data would strengthen this framework substantially.

No age adjustment in the within-athlete framework. Wikipedia race tables do not report athlete age. Not subtracting the +30 to +60 second aging penalty biases our shoe estimate downward.

Wikipedia coverage variance. Some race-years have rich tables; others have only a podium plus a handful of selected times. We compensate with median-of-top-30 metrics.

The East-African talent confound. Post-2017 also saw a documented expansion of Kenyan and Ethiopian elite marathoning. The within-athlete framework controls for this; the cohort framework does not, which is part of why we attribute only 55% rather than 100% of cohort improvement to shoes.

Course selection bias and pacing technology. Post-2017 athletes increasingly self-selected into fast, pace-controlled courses, and laser pacing lights at Berlin (from 2019) contribute to post-era improvement independently of shoes. We bucket this into the 'pacing / time-trial' slice of the decomposition pie.

Statistical power. The within-athlete sample is n=17. The paired t-test against zero is not significant ($p = 0.26$). The finding from that framework should be interpreted as suggestive rather than confirmatory.

9. Conclusion

The carbon-plated marathon shoe revolution is real, measurable, and population-level — but smaller than the most aggressive popular characterizations and larger than the most dismissive ones. Our three-framework decomposition places the shoe-attributable share of elite marathon improvement between 2010–2016 and 2018–2024 at **67 seconds in the median elite marathon time, 95% CI 36–99 seconds**, equivalent to roughly 0.9% of marathon time for a 2:05 runner.

If you trust the within-athlete framework most, weight toward 47 s — the cleanest causal design but the smallest sample and not age-adjusted. If you trust the DiD most, weight toward 111 s — the strongest causal lever, but with sparse pre-side control data. If you trust the cohort survival framework most, weight toward 58 s — densest data, but the 55% shoe-attribution share is a modeling choice.

Shoes are the most-cited cause of the post-2017 marathon revolution not because they are the largest single cause, but because they are the most discrete and dateable one. The rest — deeper African talent pipelines, pacing improvements, altitude training, race-craft — together account for more.