

Do Carbon Wheels Make You Faster?

A Difference-in-Differences Analysis

Rodrigo Sánchez Sienna*

June 2026

Working paper, version 1.0. Not peer reviewed.

Abstract

Carbon road wheels are expensive and are widely believed to make their owner faster, a claim asserted often and tested rarely. I test it on a single cyclist, the author, using 553 rides recorded between 2021 and 2026 and the day I fitted carbon wheels to an otherwise unchanged road bike. A difference-in-differences against my indoor (Zwift) rides, which the wheels cannot affect, gives a gain of about +1.7 km/h at equal power and terrain. The design fails its own parallel-trends test, however: the pre-period gaps are seasonal rather than flat, and judged against the same season the effect falls to +0.4 km/h ($p \approx 0.6$). A second outdoor bike that never received carbon, the gravel bike, does not speed up while the road bike does, which points the other way, but it rests on four post-install rides. With a sample of one I cannot tell whether I bought speed or a placebo. The wheels are nice regardless.

1 Introduction

Whether expensive carbon wheels make a cyclist faster is a question amateur cyclists argue about endlessly and rarely settle with data. The wheels are not cheap, and the gains they promise, a few watts or a few seconds, are small enough to be hard to feel and easy to advertise. I record every ride I do, so instead of joining the argument I test the claim on my own data.

The difficulty is that the obvious comparison is confounded. A rider who buys fast wheels also tends to be fitter than he used to be and to ride in better weather, so a simple before-and-after comparison would credit the wheels with gains that have nothing to do with them. What is needed is a counterfactual: how fast I would have ridden over the same period without the wheels.

I construct one as follows. On 21 April 2026 I replaced the road bike's shallow aluminium wheels with deep-section carbon ones and left the rest of the bike untouched. Deep carbon rims are sold on aerodynamics: a deeper, shaped rim reduces drag, and because drag rises roughly with the square

*Data collection was performed by the author's legs; funding, involuntarily, by the author's wallet. I thank a smart trainer for serving as a patient control group. All errors are my own, as is the entire sample. Conflict of interest: the author owns the wheels.

of speed, any benefit should be largest on fast, flat riding. Because I also ride indoors on a smart trainer, where the wheels cannot matter, indoor rides serve as a control group that picks up my fitness and motivation over time while receiving no treatment. This is a difference-in-differences design, and I add a second outdoor bike that never received carbon as a placebo. The main lesson, as it turns out, is how treacherous an indoor trainer is as a stand-in for outdoor speed.

2 Data

The data consist of 553 of my own cycling activities recorded on Strava between April 2021 and June 2026. For each ride I observe average speed, average power (I use a power meter outdoors and a smart trainer indoors), distance, elevation gain, and heart rate.¹ The sample is, by construction, one (1) cyclist, so external validity is left to the reader.

Three outdoor setups appear in the data. The treated bike is a Canyon Endurance CF SLX road bike, ridden first on its original wheels, the DT Swiss Endurance LN (a shallow aluminium clincher, 22 mm internal width, with no aerodynamic pretensions), and then, from 21 April 2026, on deep-section carbon aero wheels (DT Swiss ARC 1400, around 55 mm deep). The wheels are the only change: the frame, groupset, power meter, and tyres are the same before and after. The placebo bike is a Canyon Grizl gravel bike, which keeps its aluminium wheels (DT Swiss G1800) throughout and never receives carbon. Indoor rides are done on a Wahoo Kickr Bike through Zwift and form the control group: they track the rider’s fitness over time but are mechanically unaffected by any wheel.²

3 Empirical strategy

My design is a difference-in-differences against the indoor control group. Index a ride by i , the time of the ride by t , and the bike group by g . The outcome Speed_{it} is the average speed Strava records for ride i , in kilometres per hour (distance over moving time): outdoors it comes from the bike computer’s GPS, and on the indoor trainer it is the virtual speed Zwift derives from the rider’s power output and the simulated course. With Outdoor_i a dummy for an outdoor road-bike ride and After_t an indicator for the period after the install, I estimate

$$\text{Speed}_{it} = \alpha_g + \delta_t + \beta(\text{Outdoor} \times \text{After})_{it} + \gamma'X_{it} + \varepsilon_{it}, \quad (1)$$

where α_g are bike-group fixed effects, δ_t are time fixed effects (season and calendar year in the main specification), and β is the carbon-wheel effect. The error ε_{it} is an idiosyncratic ride-level

¹The most recent rides were pulled from the Strava API; the rest come from a full account export. All cleaning and estimation is done in Stata.

²Road bike: Canyon Endurance CF SLX 8 Di2, Shimano Ultegra Di2 R8170, 4iiii power meter, Continental GP 5000 S TR tyres (30 mm front, 32 mm rear). Gravel bike: Canyon Grizl CF SL 8 (2022), Shimano GRX 800 (RX810), 4iiii power meter, DT Swiss G1800 wheels, Schwalbe G-One 45 mm tyres. Indoor: Wahoo Kickr Bike Pro; on Zwift the avatar rides a Canyon Aeroad (2024) with ENVE wheels throughout.

disturbance, and standard errors are heteroskedasticity-robust. The control vector X_{it} holds average power, climbing per kilometre, distance, and two Zwift indicators, one for drafting and one for racing,³ so the comparison is made at equal effort and terrain. I do not control for heart rate: better wheels plausibly lower heart rate at a given speed, which would make it a post-treatment variable and net out part of the effect I am trying to measure. Identification rests on parallel trends, that is, on the assumption that indoor and outdoor speeds would have moved together absent the wheels. Because the indoor trainer never sees weather or road, I also add a second outdoor bike, the gravel bike, as a placebo, and I test the parallel-trends assumption directly in Section 5.

4 Results

4.1 Difference-in-differences against the indoor control

Table 1 reports the difference-in-differences against the indoor control. The bare estimate (column 1) is +0.2 km/h and very noisy; adding the power, terrain, drafting, and racing controls (column 2) raises it to +2.7. Absorbing seasonality matters: season fixed effects alone (column 3) cut it to +1.1, year fixed effects alone (column 4) leave it at +1.9, and both together (column 5), the main specification, give +1.7 km/h. At equal power and terrain, then, the road bike still pulls away from the indoor control after the wheels go on. On its own this looks like a real effect.

³Drafting behind other riders on Zwift raises speed at a given power. I flag drafting events from my ZwiftPower history (races plus group, pace, and social rides), matched by date rather than name because I had renamed many rides; they are 56% of the indoor rides. Since races draft hardest, I add a separate name-based race indicator (19%) so they can carry an extra boost.

Table 1: Difference-in-differences: outdoor road versus indoor (Zwift) control.

	(1)	(2)	(3)	(4)	(5)
Outdoor road	-6.866*** (0.557)	-4.228*** (0.370)	-3.756*** (0.391)	-4.363*** (0.390)	-3.940*** (0.398)
After carbon	1.147 (1.204)	-1.189** (0.394)			
Outdoor \times After carbon	0.197 (1.381)	2.667*** (0.590)	1.106** (0.417)	1.927*** (0.446)	1.681*** (0.445)
Zwift draft		1.166*** (0.344)	1.291*** (0.340)	1.071** (0.344)	1.128*** (0.336)
Zwift race		2.141** (0.741)	1.891** (0.705)	2.200** (0.745)	1.854** (0.716)
Avg. power (W)		0.077*** (0.007)	0.078*** (0.007)	0.077*** (0.007)	0.080*** (0.007)
Climb per km (m)		-0.431*** (0.030)	-0.431*** (0.030)	-0.428*** (0.030)	-0.429*** (0.030)
Distance (km)		0.028*** (0.006)	0.030*** (0.006)	0.028*** (0.006)	0.028*** (0.006)
Constant	31.411*** (0.390)	19.782*** (1.311)	19.633*** (1.266)	19.445*** (1.204)	18.994*** (1.160)
Season FE	No	No	Yes	No	Yes
Year FE	No	No	No	Yes	Yes
Observations	389	388	388	388	388
R^2	0.204	0.886	0.890	0.886	0.893

Notes: Outdoor \times After carbon is the DiD estimate. Columns 3–5 add fixed effects in place of the “After” indicator: season alone (four meteorological seasons), calendar year alone, and both. Robust standard errors in parentheses. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.2 A second outdoor control: the gravel bike

The indoor trainer does not see weather, daylight, or road surface, so it may be a poor counterfactual for the road bike’s outdoor seasonality, a concern the event study in Section 5 makes concrete. As a second control that does share those conditions, I add the gravel bike: an outdoor bike, ridden on the same roads, that never received carbon. Under a pure-seasonality story both outdoor bikes should speed up together after the install. I estimate

$$\text{Speed}_{it} = \alpha_g + \delta_t + \beta_{\text{road}} (\text{Road} \times \text{After})_{it} + \beta_{\text{gravel}} (\text{Gravel} \times \text{After})_{it} + \gamma' X_{it} + \varepsilon_{it}, \quad (2)$$

where the bike groups g are now Zwift, gravel, and road. A genuine wheel effect implies $\beta_{\text{road}} > 0$ and a placebo $\beta_{\text{gravel}} \approx 0$, whereas pure seasonality implies $\beta_{\text{gravel}} \approx \beta_{\text{road}}$. Table 2 reports the estimates.

Table 2: Adding the gravel bike, a second outdoor (never carbon) control.

	(1)	(2)	(3)	(4)	(5)
Gravel	-9.905*** (0.661)	-2.575*** (0.547)	-2.558*** (0.542)	-2.465*** (0.550)	-2.431*** (0.545)
Road	-6.866*** (0.558)	-4.226*** (0.364)	-3.763*** (0.386)	-4.363*** (0.384)	-3.935*** (0.392)
After carbon	1.147 (1.206)	-1.144** (0.396)			
Road \times After carbon	0.197 (1.384)	2.603*** (0.584)	1.112** (0.414)	1.902*** (0.441)	1.655*** (0.444)
Gravel \times After (placebo)	1.903 (1.622)	-0.287 (0.686)	-1.448* (0.628)	-1.238* (0.590)	-1.211+ (0.637)
Zwift draft		1.100** (0.340)	1.212*** (0.335)	1.008** (0.338)	1.059** (0.329)
Zwift race		2.349** (0.712)	2.106** (0.681)	2.391*** (0.718)	2.066** (0.691)
Avg. power (W)		0.075*** (0.007)	0.075*** (0.006)	0.074*** (0.007)	0.077*** (0.006)
Climb per km (m)		-0.431*** (0.030)	-0.432*** (0.030)	-0.428*** (0.030)	-0.429*** (0.030)
Distance (km)		0.028*** (0.006)	0.030*** (0.006)	0.028*** (0.006)	0.028*** (0.006)
Constant	31.411*** (0.391)	20.209*** (1.229)	20.089*** (1.184)	19.820*** (1.136)	19.427*** (1.095)
Season FE	No	No	Yes	No	Yes
Year FE	No	No	No	Yes	Yes
Observations	418	417	417	417	417
R^2	0.278	0.894	0.897	0.894	0.900

Notes: Three bike groups (Zwift, gravel, road). Road \times After carbon is the DiD; Gravel \times After is the placebo, an outdoor bike that never received carbon. Column 2 adds the power, terrain, Zwift-drafting, and Zwift-race controls; columns 3–5 add season alone, calendar year alone, and both. Reference bike: Zwift. Robust standard errors in parentheses. $^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$.

At equal power and terrain (column 5) the road bike gains +1.7 km/h after the install while the gravel bike does not; if anything it slows (−1.2, marginally significant), so the difference between them is +2.9 km/h (95% CI 1.3 to 4.4, $p < 0.001$). That the other outdoor bike did not speed up is the strongest evidence here that the wheels do something, since it is harder for seasonality to explain, though on four post-install gravel rides the placebo is noisy.

5 Robustness and identification

5.1 Pre-trends and the event study

Figure 1 plots the event study: the road-minus-indoor (Zwift) speed gap in each month relative to the install, holding power, terrain, drafting, and racing fixed. If trends were parallel, the points before the install would lie near zero. They do not. The gap is deeply negative in the summer of 2025 and climbs back through autumn and winter into spring, a pre-trend that moves with the calendar rather than the flat one parallel trends require, and the first post-install point sits essentially on the

baseline. This seasonal swing is why the main specification absorbs season and year: a meaningful part of the raw indoor comparison is the rider’s own summer rather than the wheels.⁴

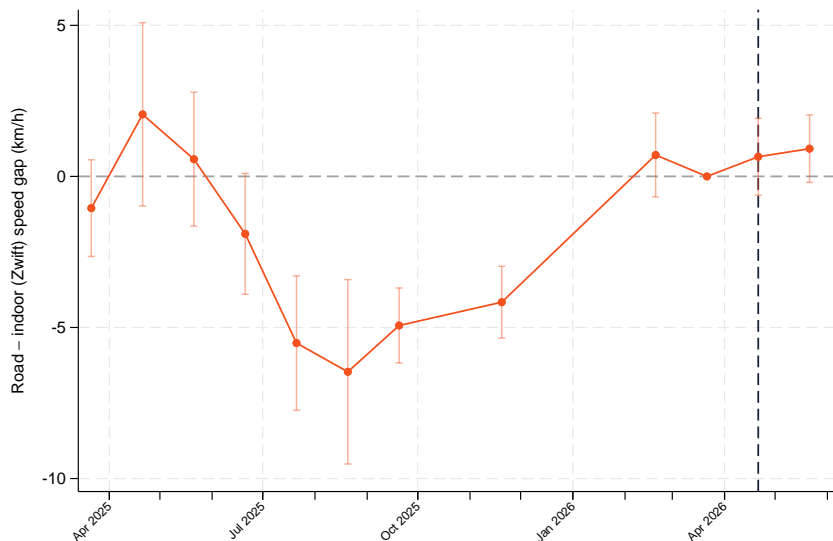


Figure 1: Event-study difference-in-differences. Each point is the road-minus-indoor (Zwift) speed gap in a calendar month, adjusted for power, terrain, drafting, and racing; the dashed line marks the carbon install (21 April 2026), and the month just before it is the omitted baseline. Bars are 95% confidence intervals.

A linear time trend does not fix this. The pre-trend is a seasonal U-shape, deep in summer and back near zero by spring, not a monotonic drift, so a straight line cannot absorb it: a treated-specific linear trend leaves the estimate essentially unchanged. Absorbing the swing instead calls for season fixed effects, as in the main specification, and for comparing like seasons directly, which I do next.

5.2 Same-season comparison

If seasonality drives the result, restricting to a matched calendar window should shrink it, because the spring post-period would then be set against the same months in earlier years rather than against the whole year. Table 3 re-estimates the controlled difference-in-differences on progressively warmer subsamples. All the post-install carbon rides fall in April to June 2026, so the treated-after observations are identical in every column and only the baseline changes.

⁴The dip is deepest in the summer training months for a mechanical reason: a watt buys far more speed indoors than out. Estimated separately, the power-to-speed slope is about 0.10 km/h per watt indoors against 0.03 on the road bike outdoors, because real roads carry the hills, wind, and surface that a smooth indoor course does not. The common power control, pulled toward the more numerous indoor rides, therefore over-predicts outdoor speed when outdoor power is highest, which is mid-season, deepening the road-minus-indoor gap precisely then.

Table 3: Robustness: the controlled DiD on alternative samples.

	(1)	(2)	(3)	(4)
Outdoor \times After carbon	2.667*** (0.590)	2.305*** (0.647)	0.425 (0.853)	0.838+ (0.460)
Sample	All months	Drop Dec–Feb	Apr–Jun	2026 only
Observations	388	249	103	141
R^2	0.886	0.872	0.891	0.951

Notes: Each column re-estimates the controlled road-versus-Zwift DiD (power, terrain, Zwift-drafting, and Zwift-race controls) on a different sample. Column 3 keeps only April–June, the months the post-install carbon rides actually span. Robust standard errors in parentheses. $^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$.

The estimate falls from +2.7 km/h on the full sample to +0.4 km/h, no longer significant, on the matched April-to-June window (column 3), and to +0.8 km/h when I instead restrict the sample to 2026 (column 4). Once the post-period is compared against the same season rather than against the rest of the year, most of the apparent gain disappears. A within-segment analysis in Appendix A, which uses no indoor data and instead treats the gravel bike as a same-season control, reaches the same conclusion with far more observations.

6 Limitations

The main limitation is external validity. The design has a sample size of one, so every estimate describes this rider, on these roads, with this fitness and position, and need not carry over to anyone else; being my own treated unit, I share its every unobservable characteristic, which keeps the comparison free of cross-person confounding but also confines it to a single case. Set against that, the controlled specifications are not thin: power, terrain, distance, bike, season, and year together explain the large majority of the variation in ride speed (R^2 around 0.9), so the comparisons are made at genuinely like-for-like effort and terrain.

7 Conclusion

I set out to measure whether an expensive set of carbon wheels made me faster, and the honest answer is that I cannot tell. The difference-in-differences against my indoor rides is positive and significant, about +1.7 km/h, but it rests on a parallel-trends assumption the event study rejects, and once the comparison is made within the same season the effect is essentially zero. The one result that points the other way, the gravel placebo, is also the one resting on the fewest rides.

This is what one would expect either if the wheels do nothing or if they help by an amount too small to detect with one rider and six weeks of data, which is also, roughly, the amount the wheels are marketed on. Settling the question would take a proper test: the same route at the same power, carbon and alloy back to back, repeated often enough to measure a difference of a few seconds. Until I run that, the most defensible conclusion is the one most cyclists reach by a more expensive route, which is to buy the wheels because they are nice and not because the data told you to.

Appendix

Do Carbon Wheels Make You Faster?

A Difference-in-Differences Analysis

Contents

A.	A within-segment design	A-2
A.1	Within-segment estimates	A-2
A.2	Segment-level difference-in-differences	A-3

A A within-segment design

The difference-in-differences in the body leans on the indoor trainer as the counterfactual, and Section 5 shows that counterfactual is weak: indoor and outdoor speeds do not move in parallel. This appendix takes a different route that does not use the trainer at all. Strava records a time for every named “segment”, a fixed stretch of road, each time I ride it, so the same short climb or kilometre of flat appears many times in the data, before and after the install. Comparing a segment with itself holds the route, the gradient, and the distance exactly fixed, and conditioning on power compares efforts at equal effort. The gravel bike, ridden on the same roads in the same weeks, then stands in for the trainer as a same-season control.

The pull covers 7,340 timed efforts. For the road bike, 211 segments were ridden both before and after 21 April 2026 with power recorded on each side, giving 606 post-install efforts against the 20 post-install rides available at the ride level. A further 53 segments were ridden by both the road and the gravel bike in both periods, enough for a segment-level difference-in-differences. Throughout, e indexes a segment effort, s a segment, and b a bike, and standard errors are clustered on segment and on ride.

A.1 Within-segment estimates

For the road bike alone I estimate

$$\text{Speed}_{es} = \alpha_s + \beta \text{Carbon}_{es} + \gamma' X_{es} + \varepsilon_{es}, \quad (3)$$

where α_s is a segment fixed effect, Carbon_{es} marks an effort ridden after the install, and X_{es} holds the Table 1 controls that still vary within a segment: power, and ride-level climbing and distance. The segment fixed effect absorbs the gradient and length of the road itself, and the Zwift drafting and racing controls do not apply because the sample is outdoor only.

Table A.1: Within-segment estimates: the same road before and after carbon.

	(1)	(2)	(3)
Carbon wheels	1.294** (0.404)	1.005*** (0.258)	1.035*** (0.270)
Segment FE	Yes	Yes	Yes
Controls	None	Power	Full
Observations	2,685	2,583	2,583
R^2	0.921	0.956	0.956

Notes: Road bike only. Outcome is segment speed (km/h). “Power” adds average power on the effort; “Full” adds ride-level climbing and distance. Standard errors clustered on segment and ride in parentheses. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

On the same road, carbon efforts are about +1.3 km/h faster raw, falling to +1.0 once power is controlled and holding there as the remaining controls are added (Table A.1). But “carbon” here is just “after”, so this gap still mixes the wheels with the spring-to-summer season, exactly as the raw

difference-in-differences did. Table A.2 bears this out. The effect is, if anything, slightly larger on flatter and longer segments, but when each segment is compared only against efforts from the same months of earlier years (column 4), it collapses to a precise zero and is no longer distinguishable from it. Same road, same season, same power, no gap.

Table A.2: Within-segment robustness, full controls throughout.

	(1)	(2)	(3)	(4)
Carbon wheels	1.035*** (0.270)	1.219*** (0.347)	1.158*** (0.270)	0.005 (0.328)
Sample	All overlap	Flatter half	Longer half	Same season
Observations	2,583	1,291	1,261	1,236
R^2	0.956	0.898	0.964	0.959

Notes: All columns are the full-control within-segment specification of Table A.1, column 3. “Flatter half” and “longer half” split segments at the median gradient and length; “same season” compares each segment only against efforts from April–June of other years. Standard errors clustered on segment and ride. $^+ p < 0.10$, $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$.

A.2 Segment-level difference-in-differences

The gravel bike gives a second way to remove the season. On the 53 segments ridden by both bikes I estimate

$$\text{Speed}_{esb} = \alpha_{sb} + \delta \text{After}_t + \beta (\text{Road} \times \text{After})_{esb} + \gamma' X_{esb} + \varepsilon_{esb}, \quad (4)$$

where α_{sb} is a segment-by-bike fixed effect and β , the road-times-after term, is the wheel effect net of anything that moved both bikes together.

Table A.3: Segment-level difference-in-differences: road versus gravel.

	(1)	(2)
Road \times After carbon	-0.971 (0.721)	0.391 (0.471)
After carbon	2.207*** (0.421)	0.540 (0.389)
Segment \times bike FE	Yes	Yes
Controls	None	Full
Observations	2,321	1,925
R^2	0.897	0.950

Notes: Segments ridden by both bikes before and after the install. Road \times After carbon is the difference-in-differences; “After carbon” is the common post-install shift. “Full” controls add power and ride-level climbing and distance. Reference bike: gravel. Standard errors clustered on segment and ride. $^+ p < 0.10$, $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$.

Both bikes speed up after the install, the “after” term in Table A.3 being positive, which is the season showing through on the road and the gravel bike alike. The road bike does not speed up by more: the difference-in-differences is +0.4 km/h at equal power and indistinguishable from zero, and slightly negative without the power control.

Two designs, then, each remove the season a different way, the same-season within-segment

comparison and the gravel difference-in-differences, and both land at essentially zero. With many more observations than the ride-level analysis they are also more precise, though they still cannot rule out a gain of around a kilometre per hour. The within-segment evidence tells the same story as the body: a clear raw gain that is mostly season, and no carbon effect that survives a same-season comparison.