MASTERARBEIT | MASTER'S THESIS

Titel | Title

Cycling Infrastructure and Cyclist Safety in Vienna: A Staggered Difference-in-Differences Evaluation
Evidence from Substantive Project Completions, 2015–2024

verfasst von | submitted by

Francesco Antonio Gaudino

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 913
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Applied Economics
Betreut von | Supervisor:	Univ.-Prof. Philipp Schmidt-Dengler PhD

Eidesstattliche Erklärung | Declaration of Authorship

Deutsch

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Diese Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

English

I hereby declare on oath that I have completed this thesis independently and without the use of any aids other than those indicated. Ideas taken directly or indirectly from external sources are identified as such.

This thesis has not been submitted to any other examining authority in the same or similar form and has not yet been published.

Wien, den ______________ ____________________________

(Francesco Gaudino)

Use of AI Tools | Verwendung von KI-Werkzeugen

In accordance with the University of Vienna's guidelines on the use of artificial intelligence in academic work, I declare that AI tools (specifically, large language models for literature search, text editing, and code debugging) were used as auxiliary tools in the preparation of this thesis. All AI-generated content was critically reviewed, verified, and edited by the author. AI tools were used exclusively as a "copilot" to support the research process; all intellectual contributions, interpretations, and conclusions are my own.

Gemäß den Richtlinien der Universität Wien zur Verwendung künstlicher Intelligenz in wissenschaftlichen Arbeiten erkläre ich, dass KI-Werkzeuge (konkret: große Sprachmodelle zur Literaturrecherche, Textbearbeitung und Code-Debugging) als Hilfswerkzeuge bei der Erstellung dieser Arbeit verwendet wurden. Alle KI-generierten Inhalte wurden vom Autor kritisch geprüft, verifiziert und bearbeitet. KI-Werkzeuge wurden ausschließlich als "Copilot" zur Unterstützung des Forschungsprozesses eingesetzt; alle intellektuellen Beiträge, Interpretationen und Schlussfolgerungen stammen vom Autor selbst.

Contents

Zusammenfassung

Verbessert die Fertigstellung größerer Radinfrastruktur-Projekte die Sicherheit von Radfahrer:innen? Diese Arbeit untersucht die zeitlich gestaffelte Fertigstellung der größeren Radinfrastruktur-Projekte Wiens (über 800 Meter Länge) zwischen 2015 und 2024 mit einem heterogenitätsrobusten Difference-in-Differences-Design nach Callaway und Sant'Anna (2021b) auf einem monatlichen Panel aller 23 Wiener Bezirke. Es zeigt sich kein signifikanter Effekt der Fertigstellungen auf die Radunfallrate pro Einwohner ($-0{,}005$ Unfälle pro 1.000 Einwohner und Monat; $p = 0{,}815$). Die Schätzung ist allerdings unpräzise: Das Konfidenzintervall umfasst moderate Rückgänge ebenso wie moderate Anstiege. Der Befund ist robust über alternative Kontrollgruppen, Kohortengewichtungen und Outcome-Definitionen.

Schlagwörter: Radverkehrsinfrastruktur, Verkehrssicherheit, Difference-in-Differences, staggered adoption, Wien

Abstract

Does completing substantive cycling infrastructure make cycling safer? This thesis studies the staggered completion of Vienna's larger cycling-infrastructure projects (over 800 metres in length) between 2015 and 2024, using the heterogeneity-robust difference-in-differences design of Callaway and Sant'Anna (2021b) on a monthly panel of all 23 districts. No significant effect of project completions on cyclist-accident rates per resident is found (-0.005 accidents per 1,000 residents per month; $p = 0.815$). The estimate is imprecise, however: the confidence interval admits moderate reductions as well as moderate increases. The finding is robust across alternative comparison groups, cohort weightings, and outcome definitions.

Keywords: Cycling infrastructure, cyclist safety, difference-in-differences, staggered adoption, Vienna

List of Figures

List of Tables

List of Abbreviations

Abbreviation	**Meaning**
ATT	Average Treatment Effect on the Treated
CI	Confidence Interval
DiD	Difference-in-Differences
FE	Fixed Effects
OGD	Open Government Data
SE	Standard Error
SUTVA	Stable Unit Treatment Value Assumption
TWFE	Two-Way Fixed Effects

1. Introduction

Cities increasingly invest in cycling infrastructure as a road-safety lever and as a complement to broader sustainable-mobility policy. Vienna provides a useful setting for empirical evaluation because the city's larger cycling-infrastructure projects were completed across districts in multiple waves rather than simultaneously, generating quasi-experimental variation in treatment timing that can be exploited with modern difference-in-differences (DiD) methods.

This thesis evaluates whether the staggered completion of cycling-infrastructure projects in Vienna affected cyclist-accident rates at the district level. The unit of analysis is the district-month, and the panel covers all 23 districts over 120 months from January 2015 to December 2024. Treatment is defined at the district level: a district enters treatment in the calendar month in which its first cycling-infrastructure project of length at least 800 m is completed, irrespective of the intervention type, where the completion date is taken from the project register in three steps – from the completion-date field where available, otherwise from the recorded end of construction, otherwise from a documented status change to "completed" – and projects without any of these three dates are excluded. This yields 9 treated districts and 14 never-treated districts. The estimator is the staggered DiD design of Callaway and Sant'Anna (2021b), implemented in a doubly-robust specification with district-clustered bootstrap inference.

1.1. Research Question

The thesis addresses one principal question and two related sub-questions:

(Q1) What is the average treatment effect on the treated (ATT) of completing a cycling-infrastructure project of at least 800 m in a Vienna district on the monthly cyclist-accident rate per 1,000 residents, identified under parallel trends and SUTVA?

(Q2) How do treatment effects evolve relative to adoption within the observed event-time window?

(Q3) Is there evidence of heterogeneity by treatment timing across cohorts?

The outcome is policy-relevant because it captures changes in cyclist-accident incidence at the resident-population scale, but it deliberately does not separate changes in per-trip risk from changes in cycling exposure. I return to this distinction in Chapter 6.

1.2. Contribution and Preview of Findings

The thesis contributes a transparent and fully reproducible application of modern staggered-DiD methods to the Vienna setting – a combination still uncommon in the applied cycling-safety literature – paired with a defensive treatment of the data limitations, in particular the inability to normalise the outcome by cycling exposure, that constrain causal interpretation. All estimation code, data provenance, and output mappings are documented in the replication package described in Appendix C.2.

No significant effect of infrastructure completions on cyclist-accident rates is found: the estimated overall average treatment effect on the treated is -0.005 accidents per 1,000 residents per month (SE 0.022, 95% CI [-0.047, $+0.037$], $p = 0.815$). The estimate is imprecise, however: measured against the treated districts' own baseline rate of about 0.045 per 1,000 residents, the confidence interval spans roughly -106% to $+83\%$. Cohort-specific estimates differ by adoption timing – negative but insignificant for early cohorts and positive and pointwise significant for late cohorts – but six of seven cohorts contain a single district and the late cohorts overlap mechanically with the post-COVID period, so I treat the heterogeneity pattern descriptively, not causally.

1.3. Roadmap

Chapter 2 situates the design relative to the cycling-safety literature and the staggered-DiD methodological literature. Chapter 3 describes the data sources and the construction of the district-month panel and the treatment variable. Chapter 4 sets out the empirical strategy. Chapter 5 presents the main results, dynamic effects, cohort heterogeneity, and the compact robustness summary. Chapter 6 interprets the findings, addresses confounding and exposure measurement, and lists the design's limitations. Chapter 7 concludes.

2. Literature and Identification Context

2.1. Cycling Infrastructure and Cyclist Safety

A central tension in the cycling-safety literature is the distinction between *risk per unit of cycling exposure* and *aggregate accident incidence* at the area level. Route-level case–crossover studies such as Teschke et al. (2012) and Harris et al. (2013) report substantially lower injury risk on cycle tracks and similar protected facilities than on major streets without infrastructure, but they identify *segment-conditional* risk by construction and do not capture the net effect of a citywide rollout on aggregate accident counts. They also do not capture exposure responses: if infrastructure simultaneously lowers per-trip risk and raises cycling volumes, the two channels can offset one another at the area level. Jacobsen (2003) and Elvik and Bjørnskau (2017) treat such "safety-in-numbers" adjustments as a distinct mechanism, but a credible empirical test requires district-by-time exposure data that are typically unavailable.

2.2. Identification in Staggered Policy Rollouts

Difference-in-differences identifies the average treatment effect on the treated under parallel trends in the absence of treatment (Angrist & Pischke, 2009). The standard two-way fixed-effects (TWFE) implementation, however, breaks down under staggered adoption with heterogeneous treatment effects: the resulting coefficient combines many two-by-two comparisons and can place negative weights on some of them, so it need not recover a meaningful average effect (de Chaisemartin & D'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun & Abraham, 2021). Callaway and Sant'Anna (2021b) address this by defining group-time average treatment effects $\mathrm{ATT}(g, t)$ as the building block, with an explicitly valid comparison group (never-treated or not-yet-treated units), and aggregating into interpretable summary parameters (overall ATT, event-study path, cohort-specific effects). The standard credibility check is inspection of pre-treatment event-study coefficients; the joint Wald-style pre-test provided in the accompanying software requires a non-singular covariance matrix that is not available when most cohorts contain a single unit, and Roth (2022) cautions that even when available such tests have limited power.

2.3. Contribution

The route-level evidence and the staggered-DiD literature meet at a clear methodological tension: route-level case–crossover studies recover *segment-conditional* risk but not the net effect of a citywide programme on aggregate incidence, while the heterogeneity-robust staggered-DiD literature is largely methodological and leaves the applied work of bringing such estimators to administrative municipal infrastructure data comparatively underdeveloped. This thesis contributes to closing that gap with a transparent, fully reproducible application: it implements the Callaway and Sant'Anna (2021b) estimator on an administrative district-month panel for Vienna covering all 23 districts and 120 months (2015–2024), with treatment defined as the completion of the first cycling-infrastructure project of at least 800 m in the district. The contribution is empirical rather than methodological, and I make the value of that explicit: the estimator is applied by the book, every data-construction and dating decision is documented and reproducible from the replication package, and the binding limitations – above all the inability to normalise by cycling exposure – are stated rather than glossed. The aim is a credible, replicable benchmark for one city, not a claim of methodological novelty.

3. Data and Treatment

3.1. Data Sources and Panel Construction

The analysis combines six administrative and project-level datasets for Vienna covering January 2015 through December 2024 (Table 3.1); most of them are published as Open Government Data (OGD) by the City of Vienna. The unit of analysis is the district-month, yielding a balanced panel of $23 \times 120 = 2{,}760$ observations. Full documentation of the source files, the variable definitions, and the merge sequence is provided in the replication package (Appendix C.2).

Table 3.1.: Data Sources

Source	Content	Use
Statistik Austria / MA 23	Cyclist-accident records, monthly, by district	Outcome (numerator)
Stadt Wien (OGD)	Annual district population, 2015–2024	Outcome (denominator)
Radlobby Wien project register	Cycling-infrastructure projects: type, length, completion	Treatment definition
Stadt Wien (OGD)	Automatic cyclist counting stations, monthly	Descriptive context + counter-based risk robustness
Stadt Wien (Mobilitätsreports, press releases)	City-wide annual cycling modal share, 2015–2024	Modal-share risk proxy (Appendix A.3 only)
GeoSphere Austria	Monthly temperature and precipitation, city-wide	Absorbed by time FE

Notes: Accident microdata were obtained on written request from MA 23 on 5 September 2025 (not OGD-published). Counting-station data are not used in baseline estimation because coverage is incomplete (11 of 23 districts). Weather data are city-wide and therefore absorbed by the time fixed effects implicit in the Callaway–Sant'Anna design.

3.2. Treatment Definition

Treatment is defined at the district level: a district enters treatment in the calendar month in which its first cycling-infrastructure project of at least 800 metres in length is completed within the 2015–2024 sample window, irrespective of the intervention type.[1] The 800-metre cutoff removes 89.6% of the completed, dateable projects in the register (Figure 3.1) and concentrates treatment on the upper tail, which is overwhelmingly composed of substantive physical-infrastructure projects (*Radwege* and structurally protected facilities). A narrower sensitivity that restricts treatment to explicitly physical infrastructure categories – dedicated cycle paths, structural build-outs, and protected bike lanes – is reported in Appendix A.3.

Only projects recorded as completed (*fertiggestellt*) qualify. The completion date is determined in three steps: it is taken from the project's completion-date field where that field is filled in; otherwise from the recorded end of construction; and only where both are missing, from the date on which the register's status-change log documents the switch to "completed" – provided that the log also shows a preceding construction phase, which excludes purely administrative status updates without a real construction history. Projects for which none of the three dates is available are excluded.

The 800-metre threshold isolates the upper tail of the project-length distribution (Figure 3.1): of the 144 projects with status *fertiggestellt* and a valid completion date, the median length is 307 m, the 90th percentile is about 808 m, and 89.6% fall below the cutoff. The threshold-sensitivity check (Figure 3.2) shows the design implied by each cutoff: lower thresholds admit progressively smaller projects and shrink the never-treated comparison group (at 700 m, only 10 districts remain never-treated), while higher thresholds leave too few treated units. I therefore set the cutoff at 800 m – essentially the 90th percentile of the length distribution – where 9 districts are treated and 14 remain never-treated across seven cohorts between October 2017 and December 2024, preserving a comparatively large and stable never-treated control pool for the Callaway–Sant'Anna design; 700 m is a defensible alternative that trades control-group size (10 vs. 14 never-treated districts) for four additional treated districts. Reported project lengths are validated against the project geometries in Appendix B.1; the maximum absolute discrepancy is below 4 m, so the threshold assignment is unambiguous.

[1] "First" refers to the first qualifying completion within the sample window. The register also records a small number of completed projects of at least 800 metres before 2015, in both later-treated and never-treated districts; treatment timing is defined by 2015–2024 completions only, so the estimand concerns the first substantive completion of the 2015–2024 rollout rather than the first in a district's history.

Two interpretive consequences of the length threshold deserve emphasis. First, never-treated districts are not free of cycling-infrastructure activity: the project register records completed sub-threshold projects in every one of the 14 never-treated districts. The estimand therefore contrasts districts after their first *substantive* completion with districts experiencing only routine, sub-threshold works – an infrastructure-intensity contrast, not an infrastructure-versus-none comparison. Second, the ATT is not re-estimated at alternative length thresholds, because moving the cutoff re-partitions the treated set and the never-treated comparison pool simultaneously (at 700 m, for example, four districts would switch from control to treated), so each threshold defines a different estimand on a different comparison group rather than a robustness check on the same one; the narrower physical-infrastructure definition (Appendix A.3) provides the controlled variation in the treatment definition instead.

Of the 15 projects above the threshold, five are dated by their completion-date field, nine by the recorded end of construction, and one via the status-change log. The date source of every project is documented automatically on each run of the analysis code as part of the replication package (Appendix C.2).

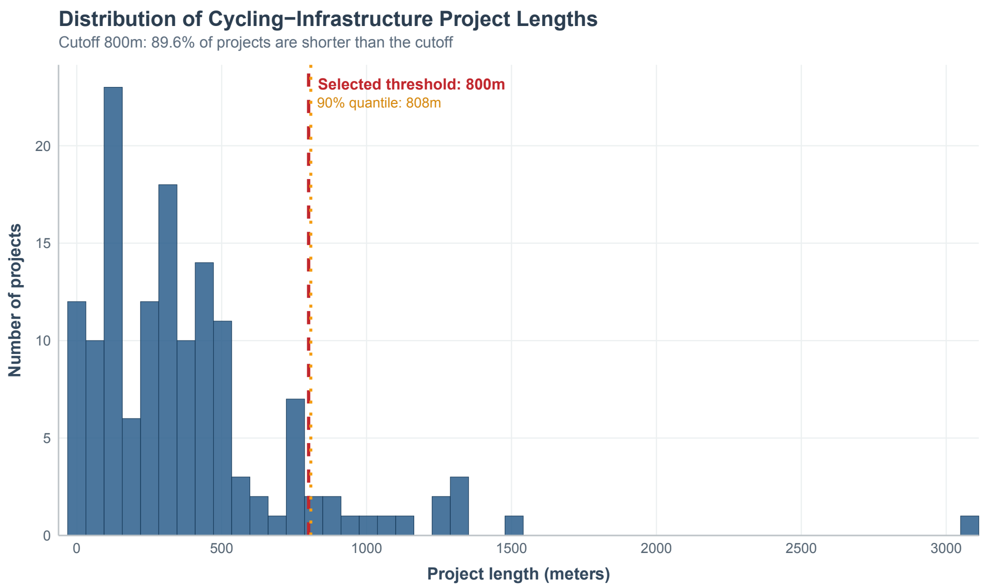

Figure 3.1.: Distribution of Cycling-Infrastructure Project Lengths

Notes: The figure shows a histogram of completed cycling-infrastructure project lengths ($n =$ 144 with valid completion date). The dashed red line marks the selected 800-metre treatment threshold; the dotted orange line marks the 90th percentile ($\approx$ 808 m). 89.6% of projects fall below the cutoff. *Source:* Radlobby Wien project register; own calculations.

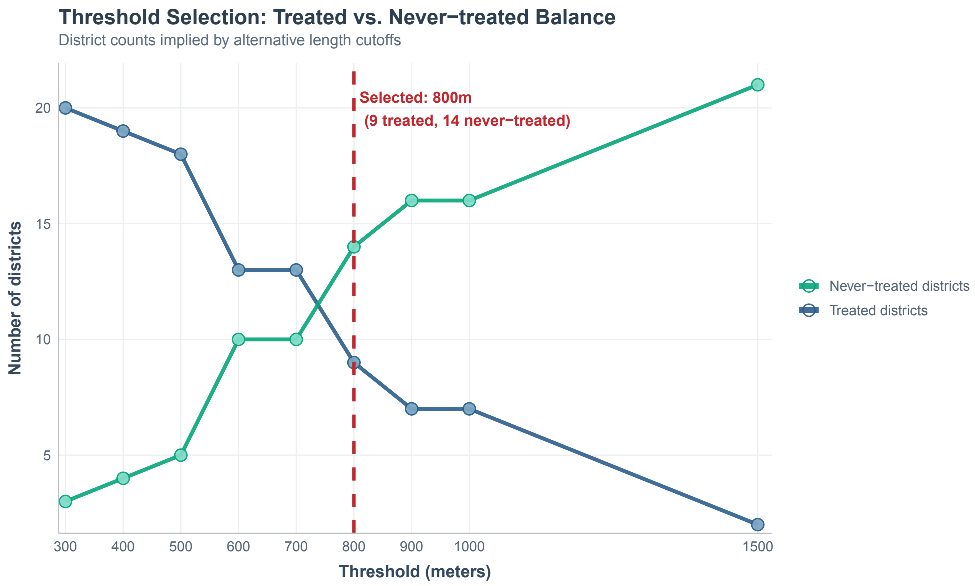

Figure 3.2.: Threshold Sensitivity: Treated vs. Never-Treated Balance

Notes: The figure shows the number of treated and never-treated districts implied by alternative project-length thresholds. The dashed vertical line marks the selected 800-metre cutoff. *Source:* Radlobby Wien project register; own calculations.

3.3. Outcome and Cohort Structure

The primary outcome is the cyclist-accident rate per 1,000 residents,

$$Y_{it} = \frac{\text{Accidents}_{it}}{\text{Population}_{it}} \times 1{,}000, \tag{3.1}$$

where the numerator counts records in the official Austrian accident statistics with at least one cyclist as an involved party and the denominator is the resident population of district i in the corresponding year (population is annual and held constant within calendar years). Pedestrian-only and motor-vehicle-only accidents are excluded. The outcome is policy-relevant at the resident-population scale but deliberately does not normalise by cycling exposure; I return to the implications in Chapter 6.

Table 3.2 lists the treatment month, the corresponding period index G (where 1 denotes January 2015 and 120 December 2024), and the length of the qualifying project for each treated district. The cohort structure is staggered with two early cohorts ($G \leq 48$, covering Oct 2017–Dec 2018, two districts), one mid-sample cohort (Donaustadt, $G = 69$, September 2020), and four late cohorts ($G \geq 82$, covering Oct 2021–Dec 2024, six districts). Six of the seven cohorts contain a single district; the November 2024 cohort

($G = 119$) bundles three districts. Post-treatment follow-up varies from 87 months for Penzing ($G = 34$) to 1–2 months for the November/December 2024 cohorts. This wide variation in follow-up length has important implications for cohort-specific inference (Chapter 4, Section 4.5). Figure 3.3 visualises this staggered adoption pattern, with point size proportional to the length of each district's first qualifying project.

Table 3.2.: Treatment Timing by District

District	Treatment month	G (period)	Length (km)
Penzing (14)	Oct 2017	34	1.51
Favoriten (10)	Dec 2018	48	3.08
Donaustadt (22)	Sep 2020	69	0.86
Ottakring (16)	Oct 2021	82	1.27
Margareten (5)	Oct 2022	94	1.09
Wieden (4)	Nov 2024	119	1.13
Rudolfsheim-Fünfhaus (15)	Nov 2024	119	1.34
Floridsdorf (21)	Nov 2024	119	0.83
Döbling (19)	Dec 2024	120	0.82

Notes: G indexes the treatment month on the panel's monthly time axis, where 1 denotes January 2015 and 120 denotes December 2024. "Length" refers to the first qualifying project under the three-step dating rule of Section 3.2, that is, to the project that triggers treatment, not to the cumulative length of all subsequent projects in the district. *Source:* Radlobby Wien project register; own calculations.

3.4. Pre-Treatment Comparisons

Table 3.3 reports descriptive pre-treatment differences between ever-treated and never-treated districts over January–November 2015, the first eleven months of the sample period, which predate the first treatment (October 2017) by almost two years. The two groups differ substantially at baseline: treated districts are about 1.7 times larger in population and carry a baseline accident rate roughly 57% below that of the never-treated districts (0.045 versus 0.105 per 1,000 residents). These are level differences, and the difference-in-differences design identifies the ATT from the equality of untreated outcome *trends*, not levels, so a level gap is not by itself a violation of the identifying assumption. I nonetheless take the gap seriously for two reasons. First, large baseline differences raise the possibility that the groups were on different underlying trajectories; I therefore inspect pre-treatment trends directly (Figure 3.4 and the event-study leads in Section 5.2), and find no pre-treatment divergence detectable at conventional levels, though I read this as the absence of detectable violations under limited power rather than as proof of parallel

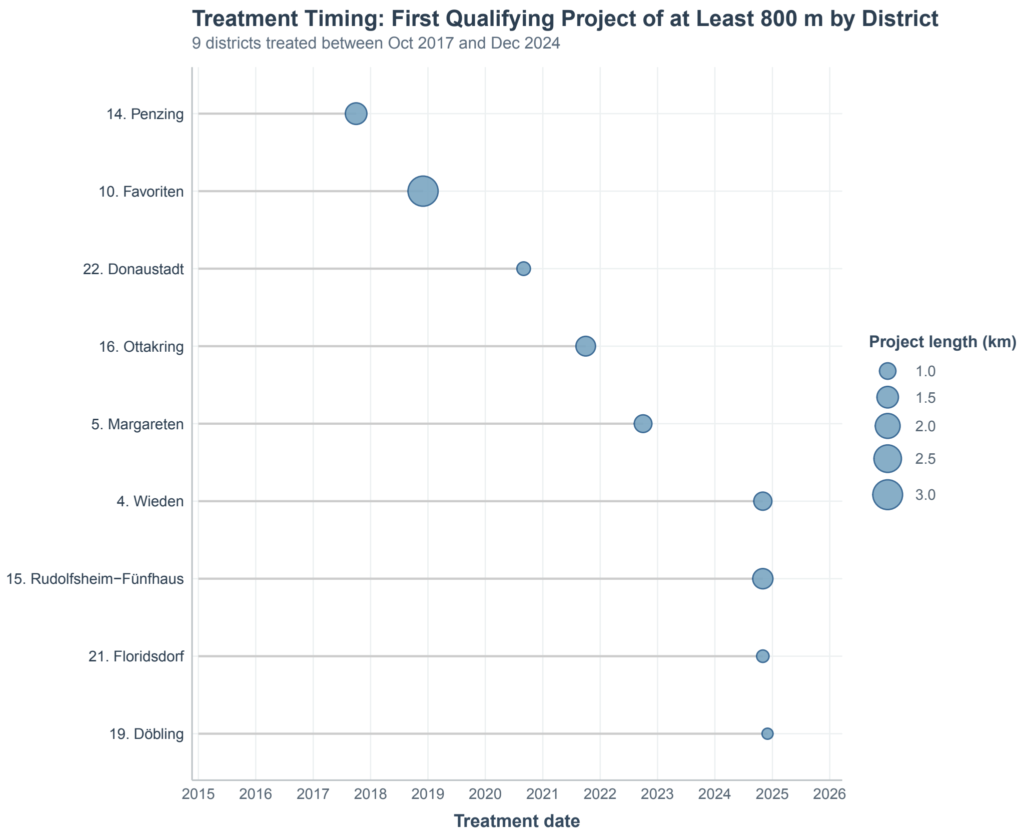

Figure 3.3.: Treatment Timing and First-Qualifying-Project Length by District

Notes: The figure visualises Table 3.2. Each point marks the completion month of the first qualifying project of at least 800 metres in a treated district; point size is proportional to the project length in kilometres. Districts are ordered by treatment date; never-treated districts are not shown. *Source:* Radlobby Wien project register; own calculations.

trends (Section 4.5). Second, a treated baseline as low as 0.045 leaves little room for further reductions, so any true safety improvement would have to operate against a floor effect that mechanically compresses the scope for a negative estimate; this is a reason to interpret the resident-normalised null cautiously. A city-wide aggregated counting-station series (Vienna cycling activity 2015–2024) is reported as descriptive context in Appendix B.2; it is not used in any estimation because counting-station coverage is restricted to 11 of 23 districts. This same series provides the broad city-level cycling-volume trend against which the resident-normalised results should be read (Figure B.1): counted cyclist volumes rise moderately across 2015–2024 with strong seasonality and a temporary spring-2020 dip, indicating that cycling exposure was generally increasing over the sample period. Because station placement is non-random and coverage is partial, the series characterises this general trend but does not support a district-level exposure correction.

Table 3.3.: Baseline Descriptive Statistics (Jan–Nov 2015)

Variable	Treated mean	Never-treated mean	Difference
Accidents (count)	3.98	3.68	+0.30
Accident rate (per 1,000)	0.045	0.105	−0.060
Population	104,092.44	61,464.64	42,627.80

Notes: The baseline window is January–November 2015, the first eleven months of the sample period; it predates the first treatment (October 2017) by almost two years. *Source:* Statistik Austria / MA 23 and Stadt Wien (OGD); own calculations.

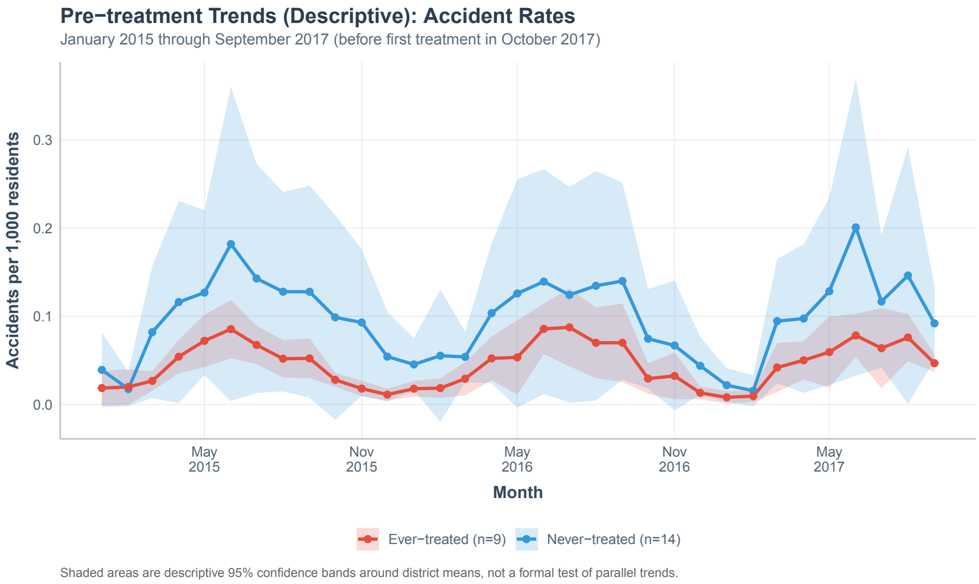

Figure 3.4.: Pre-Treatment Accident Rates by Treatment Status

Notes: The figure shows mean monthly accident rates per 1,000 residents by treatment status from January 2015 through September 2017, that is, before the first treatment in October 2017 (Penzing). It is purely descriptive and does not constitute a formal test of parallel trends. *Source:* Statistik Austria / MA 23 and Stadt Wien (OGD); own calculations.

4. Empirical Strategy

4.1. Overview

I estimate the average treatment effect on the treated using the staggered difference-in-differences design of Callaway and Sant'Anna (2021b), motivated by the well-documented bias of two-way fixed-effects estimators under heterogeneous treatment timing (de Chaisemartin & D'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun & Abraham, 2021). Treatment is defined at the district level: a district enters treatment in the calendar month in which its first cycling-infrastructure project of at least 800 metres in length is completed. Group-time average treatment effects $\mathrm{ATT}(g, t)$ are estimated with the doubly-robust estimator using the 14 never-treated districts as the comparison group, and aggregated into an overall ATT, an event-study path, and cohort-specific summaries. Identification rests on unconditional parallel trends between treated cohorts and never-treated districts in the absence of treatment, together with SUTVA. Inference relies on a district-clustered multiplier bootstrap with 1,000 iterations. I do not include covariates, so identification proceeds under unconditional, not conditional, parallel trends.

4.2. Identification

Let $Y_{it}(0)$ denote district i's cyclist-accident rate per 1,000 residents in month t in the absence of treatment, and $Y_{it}(g)$ the corresponding potential outcome if first treated in month g. The observed outcome is $Y_{it} = D_{it}Y_{it}(g) + (1 - D_{it})\, Y_{it}(0)$, where $D_{it} = 1$ indicates that district i is in a post-treatment period at t. Let $G_i \in \{g_1, \ldots, g_7, \infty\}$ denote i's first treatment month, with $G_i = \infty$ indicating a never-treated district. The estimand is the group-time average treatment effect on the treated,

$$\mathrm{ATT}(g, t) \;=\; \mathbb{E}[\, Y_{it}(g) - Y_{it}(0) \,|\, G_i = g \,], \qquad t \geq g, \tag{4.1}$$

which I identify under the (unconditional) parallel-trends assumption

$$\mathbb{E}[Y_t(0) - Y_{t'}(0) \mid G_i = g] \; = \; \mathbb{E}[Y_t(0) - Y_{t'}(0) \mid G_i = \infty] \,, \qquad t \geq g > t', \tag{4.2}$$

and SUTVA. Parallel trends is the central substantive assumption: it requires that, absent treatment, treated cohorts and never-treated districts would have evolved on the same trajectory. It is not directly testable for post-treatment periods because untreated potential outcomes for treated districts are unobserved after treatment; I therefore rely on pre-treatment diagnostics described below.

SUTVA requires no interference: one district's treatment does not affect outcomes in other districts. Vienna's districts are spatially contiguous and cyclists routinely cross district boundaries, so the completion of larger cycling infrastructure in a treated district can plausibly redistribute cycling traffic and accident risk across neighbouring districts. The data available here do not allow me to identify the sign or magnitude of such spillovers; I therefore interpret the estimated effects as net effects inclusive of any local spillovers and flag the issue in the discussion.

4.3. Estimator and Control-Group Choice

Group-time effects are estimated with the doubly-robust estimator implemented in the *did* package of Callaway and Sant'Anna (2021a), which combines outcome regression and inverse-propensity weighting and is consistent if either component is correctly specified (Bang & Robins, 2005). Because I do not include additional covariates, identification proceeds under unconditional parallel trends, and the propensity component reduces to group membership; the doubly-robust machinery therefore offers no additional protection here over a plain outcome-regression comparison of treated and never-treated districts, and I adopt it for consistency with the standard workflow of the estimator rather than as an active safeguard. The omission of covariates is a data constraint rather than a modelling preference: the available candidates are either city-wide time series (weather, the aggregated counter index), which the group-time design absorbs through its time dimension, or essentially time-invariant district characteristics over the window, which the group contrast absorbs, while resident population already enters the outcome denominator. No credible district-by-month covariates remain that would make a conditional parallel-trends specification better identified, and with 23 districts a propensity model on district characteristics would risk overfitting more than it corrects.

I adopt the 14 never-treated districts as the baseline comparison group; the never-treated set provides a stable, constant-composition control panel across the sample period and avoids the dual role of late-cohort districts as both potential treated and potential controls. The not-yet-treated alternative – which identifies $\text{ATT}(g,t)$ under the analogous parallel-trends assumption that conditions on units not yet treated by t (i.e. $G_i > t$) rather than on the never-treated, and therefore relies on a slightly different, larger but time-varying comparison set – yields a numerically very similar estimate (-0.006 vs. -0.005; Section 5.4). Anticipation is set to zero in the baseline; no pre-treatment event-study coefficient is significant at conventional levels, but I still report anticipation sensitivities at three and six months in Appendix A.4 as a defensive check. The event-study base period is varying.

4.4. Aggregation and Inference

Three aggregations of the group-time effects are reported: an overall ATT, the event-study path $\text{ATT}(e)$ at event time $e = t - g$ over the window $e \in [-24, +24]$, and cohort-specific summaries. The overall aggregate is a weighted average over all post-treatment $\text{ATT}(g,t)$ with $t \geq g$, with weights proportional to cohort size *and* to the number of available post-treatment periods (Callaway & Sant'Anna, 2021b); this gives long-observed early cohorts such as Penzing and Favoriten more weight. As a check on that weighting, I also report a cohort-balanced overall aggregate that weights each cohort equally (Section 5.4). Pointwise (not simultaneous) confidence intervals are reported. Standard errors come from a district-clustered multiplier bootstrap with 1,000 iterations and deterministic random seeds across specifications.

4.5. Diagnostics and Their Limits

Parallel trends cannot be tested directly for post-treatment periods. I rely on three complementary diagnostics: (i) inspection of the event-study path $\text{ATT}(e)$ at pre-treatment event times $e < 0$; (ii) descriptive pre-treatment comparisons between ever-treated and never-treated districts in Chapter 3; (iii) the robustness summary reported in Section 5.4, which varies the comparison group, the cohort weighting, and the cohort composition. I note at the outset that diagnostic (i) is itself weakened by the cohort structure: because six of the seven cohorts contain a single district, the pre-treatment coefficients that underpin the event-study check rest on the same thin within-cohort support as the unavailable joint test, so a clean pre-trend should be read as the absence of detectable violations under low power rather than as a strong endorsement of the assumption.

A formal joint Wald-style pre-test of parallel trends is not available in this application: the covariance matrix of the pre-treatment group-time effects is singular, so the test statistic cannot be computed. The cause is structural, not implementation-specific: six of the seven treatment cohorts contain a single district ($n_g = 1$), so the cluster bootstrap cannot produce a full-rank covariance matrix for the pre-treatment group-time effects required by the joint test. The same small-group constraint affects cohort-specific inference more generally; aggregated estimates that pool across cohorts (overall ATT, event-study path) are less affected because they do not depend on within-cohort variation, but cohort-specific p-values for $n_g = 1$ cohorts should be read as Wald-type approximations rather than as resampling-based tests. These limits constrain what the diagnostics can establish, but they do not leave the identifying assumption untested: the credibility of the design rests on the convergent reading of the pre-treatment event-study path, the descriptive pre-trend comparison, and the stability of the overall ATT across comparison groups, cohort weighting, and cohort exclusion (Section 5.4). It is the *pooled* estimates, which do not depend on within-cohort variation, that I interpret; the singleton structure cautions against over-reading any single cohort, not against the aggregate conclusion.

4.6. Implementation

All estimation is carried out in R with the *did* package (Callaway & Sant'Anna, 2021a), using the doubly-robust estimator, the never-treated comparison group, zero anticipation, a varying base period, and a district-clustered multiplier bootstrap with 1,000 iterations. Deterministic random seeds ensure exact reproducibility of all standard errors and p-values. The exact estimator settings are documented in Appendix C.1; the replication package is described in Appendix C.2.

5. Results

This chapter presents the empirical results in five steps. Section 5.1 reports the headline overall treatment effect and gauges its precision. Section 5.2 turns to the identification diagnostics, inspecting the pre-treatment event-study path in place of the unavailable joint pre-test. Section 5.3 examines heterogeneity across adoption cohorts, and Section 5.4 subjects the headline conclusion to a battery of robustness checks. Section 5.5 closes with three caveats that qualify the reported inference.

5.1. Overall Treatment Effect

Figure 5.1 plots mean accident rates over the full sample period from 2015 to 2024 for descriptive context; the ever-treated and never-treated series track each other broadly. Because this is a raw calendar-time comparison that pools treated and not-yet-treated district-months under staggered adoption, it is not by itself identifying evidence – the formal parallel-trends check is the event-study path in Section 5.2 – and I use it only to describe the levels.

Table 5.1 reports the headline overall ATT. The point estimate is -0.005 accidents per 1,000 residents per month, with a standard error of 0.022, a 95% pointwise confidence interval of $[-0.047, +0.037]$, and $p = 0.815$; no significant effect is found at conventional levels. To gauge magnitude, the interval can be scaled by a baseline rate. Relative to the never-treated baseline of approximately 0.105 accidents per 1,000 residents per month, the bounds correspond to about -45% and $+36\%$. Relative to the lower treated-group baseline of approximately 0.045 (Table 3.3), which is arguably the more appropriate scale for an effect on the treated, the same interval is far wider, roughly -106% to $+83\%$. Either way, the estimate is imprecise: it is compatible with moderate reductions as well as moderate increases. In design terms, the implied minimum detectable effect at conventional power (80%, 5% two-sided) is approximately $2.8 \times \text{SE} \approx \pm 0.061$ accidents per 1,000 residents – about 58% of the never-treated baseline and larger than the treated-group baseline itself – so the study was, ex post, powered only against implausibly large average effects.

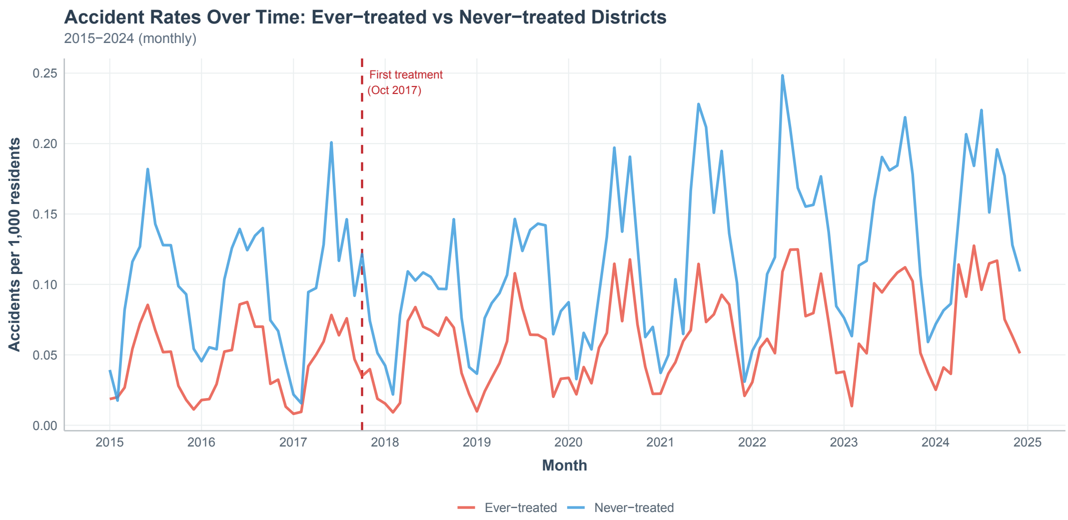

Figure 5.1.: Accident Rates Over Time, Ever-Treated vs. Never-Treated Districts

Notes: The figure shows monthly mean accident rates per 1,000 residents for ever-treated and never-treated districts from 2015 to 2024. The dashed vertical line marks the first treatment (October 2017). The series are raw group means and are not composition-adjusted. *Source:* Statistik Austria / MA 23 and Stadt Wien (OGD); own calculations.

Table 5.1.: Main Result: Overall ATT on Cyclist Accident Rates

Estimator	ATT	SE	95% CI	p-value
Callaway–Sant'Anna, overall aggregate	−0.0051	0.0216	[−0.0475, 0.0374]	0.815

Notes: The outcome is cyclist accidents per 1,000 residents per month. The estimation sample covers 23 districts × 120 months (2,760 observations), with 9 treated and 14 never-treated districts. The table reports doubly-robust group-time effects with the never-treated comparison group and cluster-bootstrap standard errors (district clusters, 1,000 iterations); the confidence interval is pointwise. *Source:* Own calculations.

5.2. Identification Diagnostics

The formal joint pre-test for parallel trends is unavailable in this application: six of seven cohorts contain a single district, which renders the joint covariance matrix of the pre-treatment group-time effects rank-deficient, so the test statistic cannot be computed (Section 4.5). I rely on the pre-treatment event-study path as the credibility check instead.

Of the 24 pre-treatment event-study coefficients (event times $e = -24, \ldots, -1$), zero are statistically significant at the 5% pointwise level and none are marginally significant at the 10% level; under the null, approximately 2.4 rejections at the 10% level would be expected from 24 tests, so the absence of marginal pre-rejections is consistent with the null of parallel trends. The full event-study table is reproduced in Appendix A.1.

Figure 5.2 shows the event-study aggregation over $e \in [-24, +24]$. Post-treatment coefficients fluctuate around zero with two pointwise-significant points at $e = +1$ ($+0.035$, $p = 0.004$) and $e = +14$ ($+0.049$, $p = 0.040$); given 25 post-treatment estimates ($e = 0, \ldots, +24$), approximately 1.25 rejections at the 5% level would be expected under the null, so the isolated pattern is consistent with multiple-testing variability rather than evidence of a systematic short-run effect.

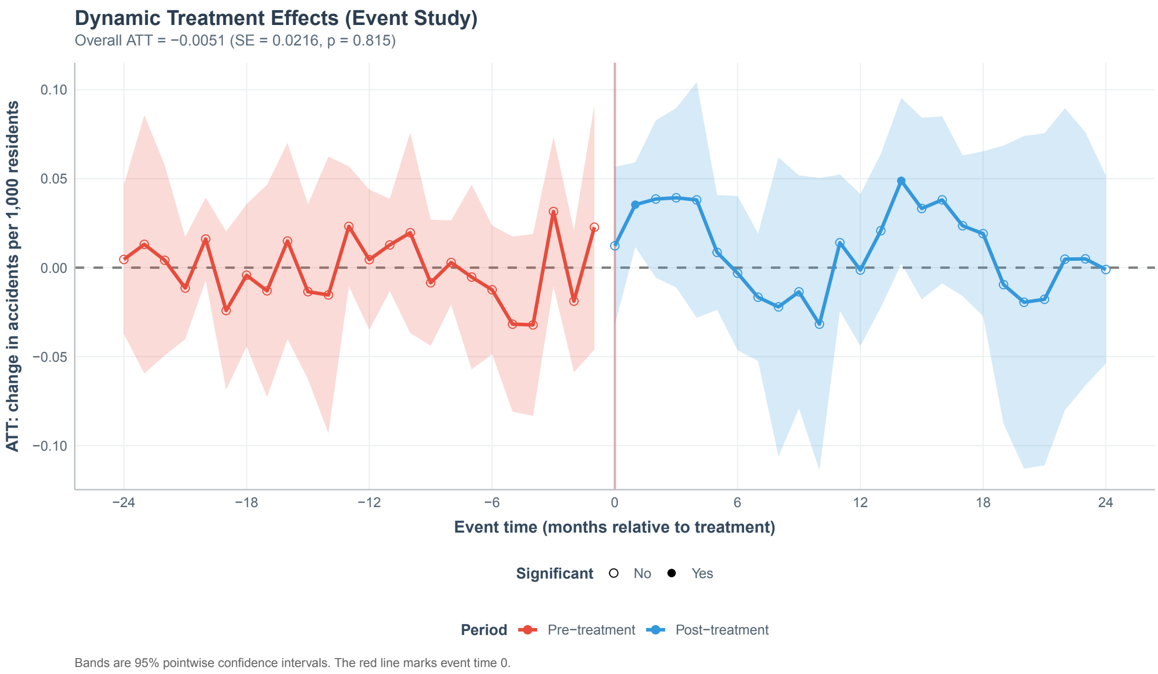

Figure 5.2.: Event-Study: Dynamic Treatment Effects, $e \in [-24, +24]$

Notes: The figure shows ATT(e) by event time relative to treatment. Bands are 95% pointwise confidence intervals from the cluster bootstrap. The cohort composition is held constant over the full event window, so the November and December 2024 cohorts ($G = 119, 120$) are excluded from this figure but included in the overall ATT in Table 5.1. *Source:* Own calculations.

5.3. Cohort Heterogeneity

Aggregating cohorts into early ($G \leq 48$, two districts: Penzing and Favoriten) and late ($G \geq 82$, six districts) groups reveals a sign reversal in the point estimates: the early-cohort aggregate is -0.029 ($p = 0.390$), while the late-cohort aggregate is $+0.043$ ($p = 0.007$, pointwise significant) (Table 5.2). I flag at the outset – and develop in Section 5.5 – that these cohort-specific results rest on singleton cohorts with fragile resampling support, so the pointwise significance of the late-cohort estimate should not be read as robust causal evidence. Excluding the November/December 2024 cohorts (which contribute only one to two post-treatment months) leaves the late-cohort estimate at $+0.044$ ($p = 0.029$). The positive signal therefore does not collapse when the short-horizon cohorts are removed, but it continues to rest on single-district cohorts all beginning in or after October 2021, coinciding with the post-COVID period. Pandemic-related shifts in cycling volumes, mobility composition, and enforcement cannot be separated from this timing within the present design; Chapter 6 therefore reads the late-cohort estimate as more consistent with sample composition and timing than with causal evidence of worsening. The mid-sample Donaustadt cohort ($G = 69$, September 2020) returns an ATT close to zero ($+0.002$, $p = 0.912$).

The full cohort-by-cohort table (seven cohorts, six of which contain a single district; the December 2024 cohort, Döbling, has a numerically degenerate p-value due to a single post-treatment observation) is reported in Appendix A.2. Figure 5.3 visualises the same estimates.

5.4. Robustness

Table 5.2 reports the compact robustness summary. The headline conclusion – a null indistinguishable from zero – survives every check below.

Comparison group. Replacing the never-treated districts with the not-yet-treated comparison group moves the overall ATT only from -0.005 to -0.006 ($p = 0.758$), so the result does not depend mechanically on which units serve as controls.

Pre-COVID restriction. Restricting the sample to periods up to December 2019, with post-2019 adopters recoded as never-treated, yields -0.015 ($p = 0.316$), also statistically indistinguishable from zero.

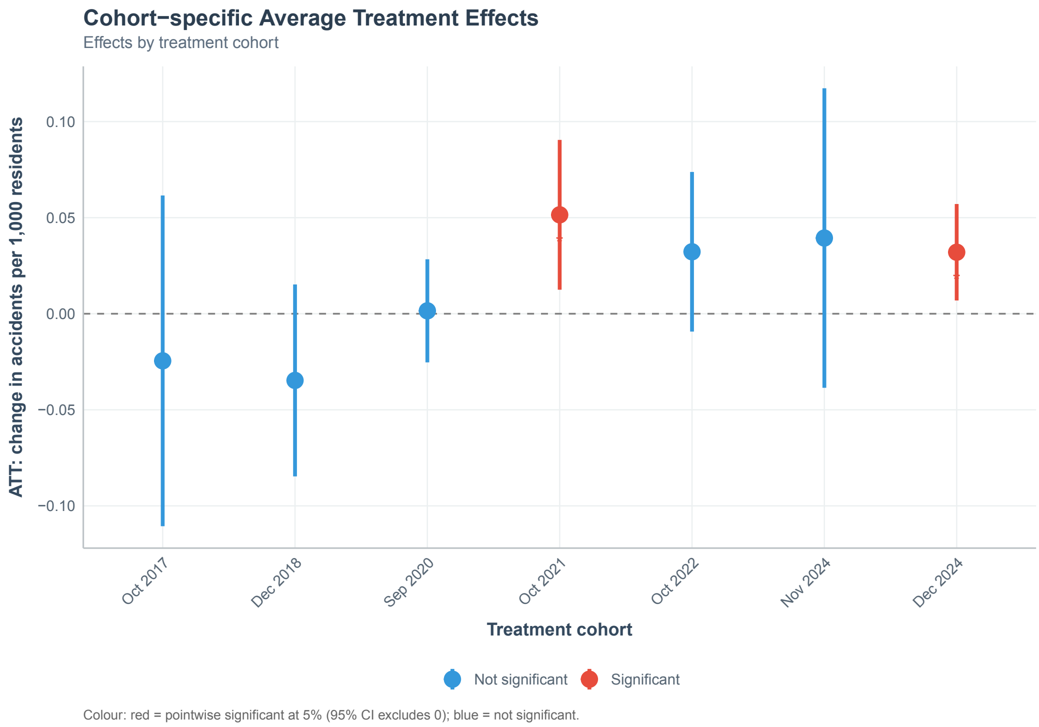

Figure 5.3.: Cohort-Specific Average Treatment Effects

Notes: The figure shows cohort-specific average treatment effects; bars are 95% pointwise confidence intervals. Six of seven cohorts contain a single district, so cohort-specific inference should be read as a Wald-type approximation rather than as a resampling-validated rejection. The full table is reported in Appendix A.2. *Source:* Own calculations.

Counter-based risk outcome. Rescaling the denominator by monthly Vienna cyclist-counter trips (see Section 6.2) gives a point estimate that is likewise statistically indistinguishable from zero ($\hat{\tau} = +0.60$, $p = 0.40$). Because the counter index is city-wide and time-specific, it is largely absorbed by the time fixed effects implicit in the design and does not isolate a per-trip exposure correction; I therefore read it only as a same-direction null on a different scale, not as evidence that exposure is unconfounded (Section 6.2).

Narrower treatment definition. Restricting treatment to projects explicitly recorded as physical cycling infrastructure – dedicated cycle paths, structural build-outs, or protected bike lanes – yields an estimate very close to zero (+0.003, $p = 0.888$), and both log-transformed outcomes have 95% confidence intervals that include zero (full results in Appendix A.3).

Cohort weighting. The headline aggregation weights group-time effects by cohort size and post-treatment exposure (Section 4.4), so the two long-observed early cohorts (Penzing and Favoriten, with 87 and 73 post-treatment months) carry most of the weight. To check whether the near-zero headline is an artefact of this weighting, I also aggregate the cohort-specific effects with equal cohort weight. This cohort-balanced aggregate is +0.020 (SE 0.014, $p = 0.155$): the point estimate changes sign relative to the exposure-weighted -0.005, which confirms that the headline magnitude is sensitive to how cohorts are weighted, but it too remains statistically indistinguishable from zero. The qualitative conclusion – no precisely estimated average effect in either direction – therefore holds under both weighting schemes, even though the point estimate itself is not weighting-invariant.

Figure 5.4 visualises the eight resident-rate specifications – the seven resident-rate rows of Table 5.2 together with the narrower physical-infrastructure definition reported in Appendix A.3: the headline point estimate is essentially zero with a wide pointwise interval, the not-yet-treated and pre-COVID variants line up with the main result, and the only pointwise-significant rejections come from the two late-cohort sub-samples, consistent with the cohort-composition reading developed in Section 5.3.

A leave-one-out exercise (Figure 5.5) yields a maximum absolute deviation from baseline of 0.0102 – the largest single shift comes from excluding the December 2018 cohort (Favoriten, $G = 48$), which moves the overall ATT to approximately +0.005 – and all leave-one-out re-estimates remain statistically indistinguishable from zero, so the headline conclusion is not driven by any single cohort.

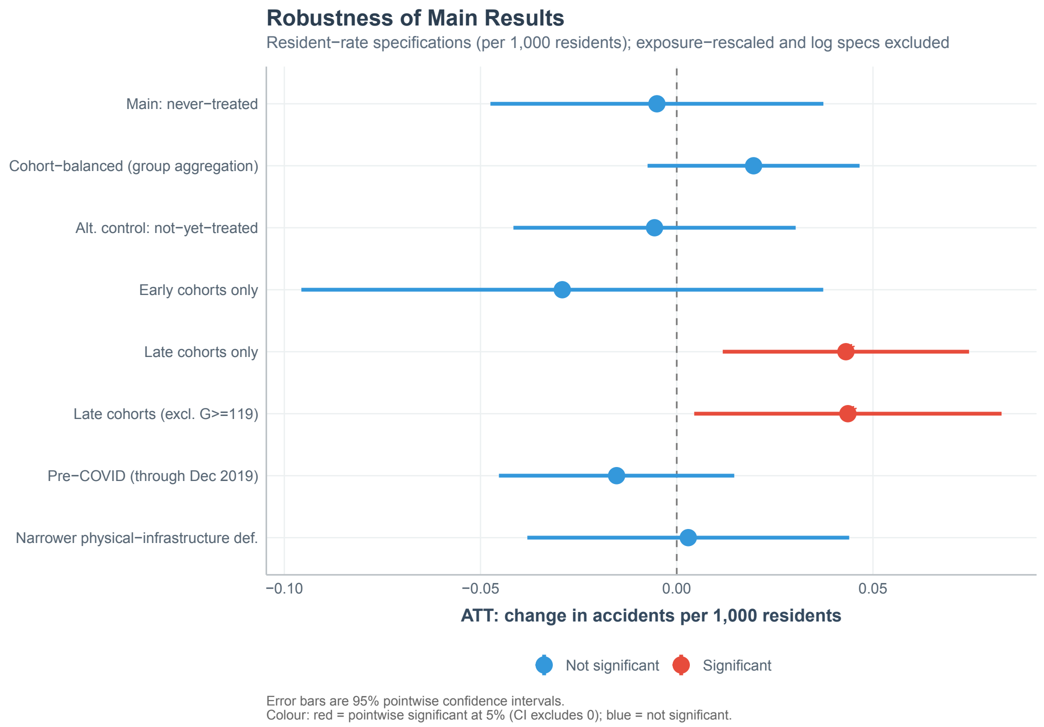

Figure 5.4.: Robustness of Main Results across Resident-Rate Specifications

Notes: The figure shows a forest plot of the eight resident-rate ATT specifications: the seven resident-rate rows of Table 5.2 (the main never-treated specification, the cohort-balanced aggregate, the not-yet-treated alternative, the early- and late-cohort sub-samples, the late-cohort sub-sample excluding $G \geq 119$, and the pre-COVID restriction) plus the narrower physical-infrastructure treatment definition from Table A.3. Bars are 95% pointwise confidence intervals from the district-clustered multiplier bootstrap; all specifications use the exposure-weighted aggregation except the cohort-balanced row, which weights each cohort equally. The counter-based risk outcome (shown in Table 5.2) and the log-transformed outcomes are omitted because they are on different scales; their estimates are reported in Appendix A.3. *Source:* Own calculations.

Table 5.2.: Robustness Summary: Overall ATT under Alternative Specifications

Specification	ATT	SE	95% CI	p-value
Main: never-treated controls	−0.0051	0.0216	[−0.0475, 0.0374]	0.815
Cohort-balanced (group aggregation)	+0.0196	0.0138	[−0.0074, 0.0466]	0.155
Alt. control: not-yet-treated	−0.0057	0.0183	[−0.0416, 0.0303]	0.758
Early cohorts only ($G \leq 48$)	−0.0292	0.0339	[−0.0957, 0.0373]	0.390
Late cohorts only ($G \geq 82$)	+0.0431	0.0160	[0.0117, 0.0745]	0.007*
Late cohorts (excl. $G \geq 119$)	+0.0436	0.0200	[0.0045, 0.0828]	0.029*
Pre-COVID ($t \leq$ Dec 2019)†	−0.0153	0.0153	[−0.0453, 0.0147]	0.316
Counter-based risk‡	+0.604	0.712	[−0.79, 2.00]	0.396

Notes: All specifications estimate the overall ATT with the exposure-weighted aggregation, except the cohort-balanced row, which weights each cohort equally (Section 5.4). Standard errors are district-clustered bootstrap standard errors with 1,000 iterations and deterministic random seeds. †The pre-COVID specification recodes cohorts first treated after December 2019 as never-treated. ‡The counter-based risk outcome is accidents per 10^6 city-wide cyclist-counter trips, aggregated over the 9 districts with continuous 2015–2024 counting-station coverage; it is on a different scale than the resident-rate specifications and is discussed in Section 6.2 and Appendix B.3. *Pointwise significant at the 5% level. Narrower physical-infrastructure, log-outcome, leave-one-out, and anticipation sensitivities are reported in Appendix A.3 and Appendix A.4. *Source:* Own calculations.

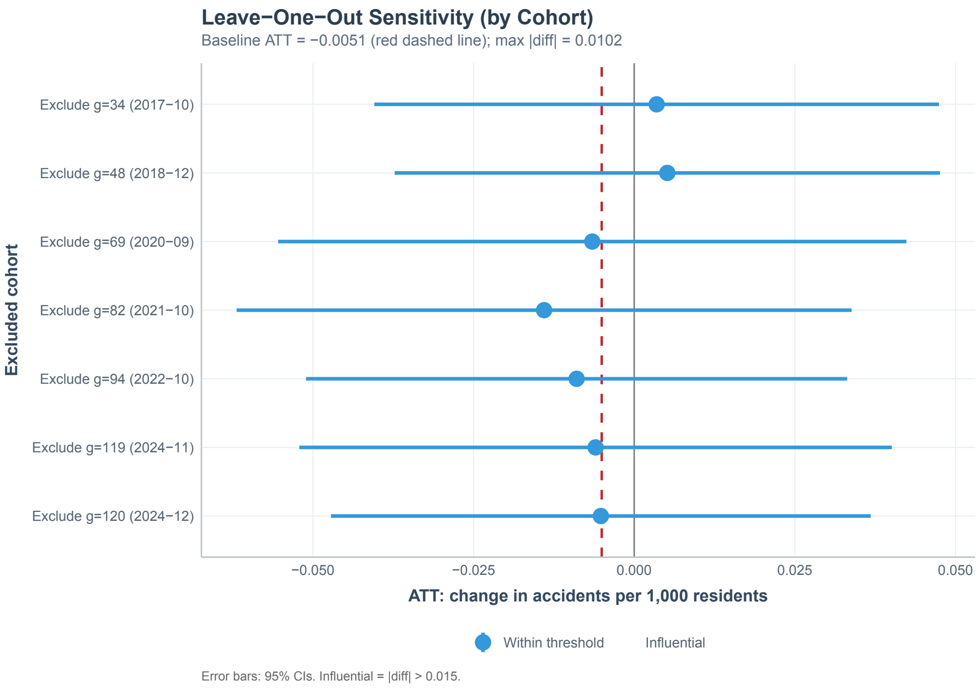

Figure 5.5.: Leave-One-Out Sensitivity by Treatment Cohort

Notes: The figure shows the overall ATT re-estimated after sequentially excluding each treatment cohort. The dashed line marks the baseline overall ATT (−0.005); the maximum absolute deviation from the baseline is 0.0102. *Source:* Own calculations.

5.5. Inference Caveats

Three caveats qualify the reported standard errors. First, the design contains only 23 districts (9 treated, 14 never-treated), below the conventional threshold of roughly thirty clusters for asymptotically valid cluster-robust inference (Cameron, Gelbach, & Miller, 2008); standard errors and p-values should therefore be read as approximations. Second, p-values for the aggregated effects rest on a normal approximation rather than directly on the resampling distribution. Third, six of the seven treatment cohorts consist of a single district, so cohort-specific inference is fragile; the aggregated overall ATT is less affected but still inherits limited resampling support. For these reasons, no conclusion in this thesis rests on a borderline significance result, including the late-cohort estimates ($p = 0.007$ and $p = 0.029$).

6. Discussion

6.1. Interpreting the Headline Result

Taken together, the evidence neither supports the claim that Vienna's cycling-infrastructure rollout produced a measurable reduction in district-level cyclist-accident incidence per resident over the sample period, nor the opposite claim. The 95% confidence interval contains both moderate reductions and moderate increases, and the point estimate is dominated by sampling noise rather than a precise zero. The not-yet-treated control specification produces a numerically very similar estimate (-0.006 vs. -0.005), so the design does not depend mechanically on the choice of comparison group, and the leave-one-out exercise shows a maximum absolute deviation from baseline of 0.0102, with all re-estimates statistically indistinguishable from zero. The cohort-by-cohort pattern is suggestive but fragile: six of the seven cohorts contain a single district, the late cohorts overlap mechanically with the post-COVID period, and several late cohorts contribute only one or two post-treatment months. I read the positive late-cohort point estimate as more consistent with sample composition and timing than with causal evidence of worsening.

6.2. Confounding, Exposure, and the Counterfactual

The outcome measures cyclist accidents per 1,000 residents, not accidents per cyclist trip. If the completion of cycling infrastructure simultaneously raises cycling volumes and lowers per-trip injury risk, a near-zero estimate on the resident-normalised outcome is fully consistent with a positive per-trip safety effect; conversely, an unchanged per-trip risk combined with a rise in volumes can mechanically increase the resident-normalised rate. Disentangling these channels would require district-level exposure measures (cyclist trips or vehicle-kilometres) at a monthly frequency, which are not available: Vienna's permanent counting stations cover only 11 of 23 districts and are placed at trafficked corridors that are not representative of district-wide cycling exposure. An aggregated city-wide counting-station series (Appendix B.2) shows a moderate upward trend with strong seasonality, consistent with rising exposure over the sample period, but it does not support a district-level exposure correction and is therefore used as descriptive context only. I accordingly identify the average effect of cycling-infrastructure completions on *aggregate* cyclist-accident incidence per resident, and I cannot rule out that this estimate masks offsetting movements in exposure and per-trip risk. The same data constraint also implies that I cannot empirically test "safety-in-numbers" mechanisms within the present design.

I emphasise that the present design does not resolve this exposure confound; it is one of the study's two binding limitations (the other being the singleton-cohort structure, Section 6.3) rather than a gap I am able to close. The counter-based risk outcome reported in the robustness summary (ATT +0.60, $p = 0.40$) should not be read as an exposure correction. It rescales the denominator by a *city-wide* monthly counter index, which – being common to all districts in a given month – is largely absorbed by the time fixed effects implicit in the Callaway–Sant'Anna design and therefore primarily changes the outcome scale rather than isolating per-trip risk; its agreement with the headline reflects a same-direction null on a different scale, not evidence that exposure is unconfounded. A genuinely district-specific correction is infeasible here for a more basic reason: of the nine treated districts, only four (Donaustadt, Margareten, Wieden, and Döbling) have continuous counting-station coverage, and three of those (Margareten, Wieden, and Döbling) belong to the post-October-2021 cohorts – with Wieden and Döbling contributing only one to two post-treatment months, and Wieden being the sole covered district in its three-district cohort. A district-by-month exposure-normalised estimator would thus rest on an even thinner and more selectively composed treated sample than the resident-normalised headline, so the resident-population denominator is the more defensible choice given the available data, not merely a convenient one. Strava Metro, bike-sharing trip data, and the Google/Apple mobility reports were also assessed and rejected for the reasons documented in Appendix B.3.

A second confound is timing-correlated: the three latest cohorts ($G = 94$, 119, 120, comprising five treated districts) begin treatment after October 2021, mechanically overlapping with the post-COVID recovery of cycling traffic, with changes in vehicle composition (e.g. delivery and ride-hail vehicles), and with shifts in enforcement and parking management. None of these can be measured at the district-month level in the data, and none can be separated from the late-cohort treatment timing within the present design.

6.3. Limitations

Two limitations are of first order. The first is the exposure-measurement gap discussed in Section 6.2: the per-trip safety effect of infrastructure cannot be identified without district-level cyclist-exposure data, which constrains *what* the design can answer. The second is the cohort structure: six of seven treatment cohorts contain only one district, which constrains *how well* it answers it. This singleton structure is not merely an inference nuisance – it removes the joint Wald-style pre-test entirely, leaves the event-study pre-trend check with thin within-cohort support (Section 4.5), and makes the headline magnitude sensitive to cohort weighting: the exposure-weighted overall ATT (-0.005) and the cohort-balanced aggregate ($+0.020$) differ in sign even though both are statistically indistinguishable from zero (Section 5.4). I therefore treat the two as co-equal binding limitations rather than ranking one above the other. Four further limitations remain, in addition to the post-COVID timing confound already discussed in Section 6.2. First, the three-step dating rule (Section 3.2) yields a smaller treated sample (9 districts versus 14 never-treated) by excluding projects without a credible completion date. Second, Vienna's districts are spatially contiguous, so SUTVA may be violated by network spillovers that I cannot measure. Third, anticipation horizons longer than six months cannot be tested in the present design because the singleton-cohort structure leaves too few pre-treatment periods to support them; the sensitivity reported in Appendix A.4 therefore covers only $a = 3$ and $a = 6$. Fourth, the outcome counts all cyclist accidents equally. The district-month extract does report accident-level severity classes (property-damage-only, slight, severe, fatal), but these code the most severe outcome across all involved parties rather than cyclist-specific injury severity, and I deliberately estimate on total counts; severity-shifting effects – for instance, infrastructure converting severe into slight accidents while leaving total counts unchanged – therefore remain untested in this design and are left to future work.

7. Conclusion

This thesis estimates the average treatment effect of larger cycling-infrastructure project completions in Vienna on district-monthly cyclist-accident rates per resident, using the staggered difference-in-differences design of Callaway and Sant'Anna (2021b). No significant overall effect is found (-0.005, $p = 0.815$). The qualitative conclusion – no precisely estimated average effect in either direction – holds across the choice of comparison group, the cohort-balanced aggregation, the exclusion of individual cohorts, and the log-transformed outcomes, because none of these specifications is statistically significant at conventional levels; I stress that this is consistency in statistical insignificance, not in the point estimate, which moves and even changes sign across the cohort-balanced and log specifications (Section 5.4). The late-cohort positive estimate is not robust enough – given singleton cohorts and post-COVID timing – to support a policy claim of worsening, and the early-cohort negative estimate is too imprecise to support a policy claim of improvement.

The most direct implication for future research is the need for district-level cyclist-exposure measurement. Without it, the per-trip safety effect of infrastructure cannot be identified, and policy conclusions about whether cycling has become safer per kilometre travelled remain outside the scope of any design built on resident-normalised accident counts. Two further extensions would strengthen inference: a panel covering more treated cohorts or multiple cities to address the small-cluster constraint, and project-level data on intersection treatment, network continuity, and complementary measures to move from average to conditional effects.

The most honest summary is that, in this design and on this outcome, the average effect is too imprecise to support a policy claim either way. The 95% confidence interval spans roughly -45% to $+36\%$ of the never-treated baseline rate (and wider still, about -106% to $+83\%$, against the lower treated-group baseline), remaining consistent with both moderate increases and moderate reductions in cyclist-accident incidence per resident. What the design establishes is therefore not a number but a boundary: with these data and this outcome, Vienna's larger cycling-infrastructure completions did not produce a district-level change in resident-normalised cyclist-accident incidence large enough to be detected, and identifying the per-trip safety effect that policy ultimately cares about will require the district-level exposure data this study did not have.

A. Complete Estimation Results

This appendix collects the complete estimation output behind the results reported in Chapter 5: the full event-study table, the cohort-by-cohort breakdown, the full robustness summary, and the anticipation sensitivity.

A.1. Full Event-Study Results

Table A.1 reports dynamic treatment-effect estimates for all event times $e = -24, \ldots, +24$, from the Callaway and Sant'Anna (2021b) doubly-robust estimator under the baseline specification (never-treated comparison group, district-clustered bootstrap standard errors). Event time 0 is the month in which the first qualifying project of at least 800 metres is completed; the aggregation holds the cohort composition constant over the full window, so the November and December 2024 cohorts ($G = 119, 120$) do not contribute (cf. Figure 5.2).

Table A.1.: Complete Dynamic Event Study Results (All Available Event Times)

Event Time	ATT	SE	t-stat	p-value	Sig	95% CI
-24	0.0046	(0.0213)	0.216	0.829		[-0.0372, 0.0464]
-23	0.0131	(0.0370)	0.353	0.724		[-0.0595, 0.0857]
-22	0.0041	(0.0272)	0.152	0.879		[-0.0493, 0.0575]
-21	-0.0115	(0.0146)	-0.782	0.434		[-0.0402, 0.0173]
-20	0.0161	(0.0119)	1.352	0.176		[-0.0072, 0.0394]
-19	-0.0241	(0.0227)	-1.063	0.288		[-0.0685, 0.0203]
-18	-0.0043	(0.0204)	-0.210	0.834		[-0.0443, 0.0357]
-17	-0.0130	(0.0304)	-0.428	0.668		[-0.0727, 0.0466]
-16	0.0150	(0.0281)	0.533	0.594		[-0.0401, 0.0700]
-15	-0.0135	(0.0251)	-0.540	0.590		[-0.0626, 0.0356]
-14	-0.0153	(0.0397)	-0.386	0.699		[-0.0931, 0.0624]
-13	0.0232	(0.0172)	1.348	0.178		[-0.0105, 0.0569]
-12	0.0045	(0.0201)	0.224	0.823		[-0.0349, 0.0439]
-11	0.0128	(0.0132)	0.964	0.335		[-0.0132, 0.0387]
-10	0.0196	(0.0288)	0.682	0.495		[-0.0367, 0.0760]
-9	-0.0085	(0.0181)	-0.470	0.638		[-0.0439, 0.0269]
-8	0.0029	(0.0121)	0.239	0.811		[-0.0208, 0.0265]
-7	-0.0053	(0.0265)	-0.200	0.841		[-0.0572, 0.0466]
-6	-0.0124	(0.0184)	-0.675	0.500		[-0.0486, 0.0237]
-5	-0.0317	(0.0251)	-1.265	0.206		[-0.0809, 0.0174]
-4	-0.0322	(0.0260)	-1.235	0.217		[-0.0832, 0.0189]
-3	0.0315	(0.0214)	1.472	0.141		[-0.0105, 0.0735]
-2	-0.0188	(0.0204)	-0.926	0.355		[-0.0587, 0.0210]
-1	0.0227	(0.0351)	0.647	0.517		[-0.0461, 0.0915]
+0	0.0123	(0.0226)	0.542	0.588		[-0.0321, 0.0566]
+1	0.0354	(0.0121)	2.917	0.004	*	[0.0116, 0.0591]
+2	0.0385	(0.0225)	1.711	0.087		[-0.0056, 0.0827]
+3	0.0392	(0.0257)	1.526	0.127		[-0.0112, 0.0897]
+4	0.0380	(0.0338)	1.127	0.260		[-0.0281, 0.1042]
+5	0.0086	(0.0165)	0.521	0.603		[-0.0237, 0.0408]
+6	-0.0031	(0.0221)	-0.141	0.888		[-0.0464, 0.0402]
+7	-0.0166	(0.0183)	-0.909	0.363		[-0.0524, 0.0192]
+8	-0.0220	(0.0429)	-0.514	0.607		[-0.1060, 0.0620]
+9	-0.0136	(0.0334)	-0.407	0.684		[-0.0790, 0.0519]
+10	-0.0318	(0.0419)	-0.758	0.448		[-0.1138, 0.0503]
+11	0.0140	(0.0196)	0.717	0.473		[-0.0243, 0.0523]
+12	-0.0013	(0.0218)	-0.060	0.952		[-0.0440, 0.0413]
+13	0.0207	(0.0220)	0.943	0.346		[-0.0224, 0.0639]
+14	0.0488	(0.0238)	2.054	0.040	*	[0.0022, 0.0954]
+15	0.0332	(0.0260)	1.276	0.202		[-0.0178, 0.0843]
+16	0.0381	(0.0239)	1.592	0.111		[-0.0088, 0.0850]
+17	0.0235	(0.0202)	1.167	0.243		[-0.0160, 0.0630]
+18	0.0191	(0.0236)	0.808	0.419		[-0.0272, 0.0654]
+19	-0.0095	(0.0399)	-0.239	0.811		[-0.0877, 0.0686]
+20	-0.0194	(0.0477)	-0.408	0.683		[-0.1129, 0.0740]
+21	-0.0178	(0.0476)	-0.374	0.709		[-0.1111, 0.0755]
+22	0.0048	(0.0433)	0.110	0.913		[-0.0801, 0.0896]
+23	0.0049	(0.0363)	0.136	0.892		[-0.0662, 0.0760]
+24	-0.0011	(0.0269)	-0.040	0.968		[-0.0538, 0.0517]

Notes: The asterisk indicates that the 95% pointwise confidence interval excludes zero. Estimation details are given in the chapter text above. *Source:* Own calculations.

A.2. Cohort-Specific ATTs

Table A.2 reports the full cohort-by-cohort breakdown referenced in Section 5.3. Six of the seven cohorts contain a single district ($n_g = 1$), so cluster-bootstrap inference has no within-cohort variation and the reported p-values should be read as Wald-type approximations; the December 2024 entry ($G = 120$, Döbling) rests on a single post-treatment month, so its near-zero p-value is numerically degenerate.

Table A.2.: Cohort-Specific ATTs (Full Breakdown)

Cohort (first treatment month)	n_g	ATT	SE	95% CI	p-value
Oct 2017	$1^{\dagger}$	−0.0245	0.0439	[−0.1106, 0.0615]	0.577
Dec 2018	$1^{\dagger}$	−0.0347	0.0255	[−0.0846, 0.0152]	0.173
Sep 2020	$1^{\dagger}$	+0.0015	0.0137	[−0.0253, 0.0283]	0.912
Oct 2021	$1^{\dagger}$	+0.0515	0.0199	[0.0125, 0.0905]	0.010^{*}
Oct 2022	$1^{\dagger}$	+0.0322	0.0212	[−0.0093, 0.0738]	0.128
Nov 2024	3	+0.0394	0.0398	[−0.0385, 0.1173]	0.322
Dec 2024	$1^{\dagger}$	+0.0320	0.0128	[0.0069, 0.0571]	$0.012^{*,\ddagger}$

Notes: The table reports cohort-specific ATTs under the baseline specification, aggregated with equal weight within each cohort. *Pointwise significant at the 5% level (confidence interval excludes zero); consistent with the * convention used in Tables 5.1, 5.2, and A.3. †Singleton cohort ($n_g = 1$); the cluster bootstrap has no within-cohort variation, so p-values are Wald-type approximations. ‡Numerically degenerate: the estimate rests on a single post-treatment observation. *Source:* Own calculations.

A.3. Full Robustness Summary

Table A.3 reproduces the full robustness summary, adding to the resident-rate specifications already shown in Section 5.4 the secondary exposure-rescaled outcomes and the two log-transformed outcome specifications.

The main-specification treatment is defined by length alone (at least 800 metres, dated by the three-step rule of Section 3.2), without an intervention-type filter; the narrower sensitivity in Table A.3 additionally requires the project to be recorded as physical cycling infrastructure – a dedicated cycle path, a structural build-out, or a protected bike lane – and its estimate of +0.0029 (SE 0.021, $p = 0.888$) reinforces the headline conclusion. The full leave-one-out output is part of the replication package (Section C.2); the maximum absolute deviation from the baseline overall ATT is 0.0102 (excluding the December 2018 cohort, Favoriten, which moves the estimate to approximately +0.0052), and all leave-one-out re-estimates remain statistically insignificant.

Table A.3.: Full Robustness Summary

Specification	ATT	SE	95% CI	p-value
Main: never-treated controls	−0.0051	0.0216	[−0.0475, 0.0374]	0.815
Cohort-balanced (group aggregation)¶	+0.0196	0.0138	[−0.0074, 0.0466]	0.155
Alt. control: not-yet-treated	−0.0057	0.0183	[−0.0416, 0.0303]	0.758
Early cohorts only ($G \leq 48$)	−0.0292	0.0339	[−0.0957, 0.0373]	0.390
Late cohorts only ($G \geq 82$)	+0.0431	0.0160	[0.0117, 0.0745]	0.007*
Late cohorts (excl. $G \geq 119$)	+0.0436	0.0200	[0.0045, 0.0828]	0.029*
Pre-COVID ($t \leq$ Dec 2019)†	−0.0153	0.0153	[−0.0453, 0.0147]	0.316
Narrower physical-infrastructure def.	+0.0029	0.0209	[−0.0381, 0.0439]	0.888
Risk proxy (modal-share)‡	+6.20	16.47	[−26.09, 38.49]	0.707
Counter-based risk§	+0.604	0.712	[−0.79, 2.00]	0.396
log(accidents + 1)	+0.0430	0.0773	[−0.1085, 0.1945]	0.578
log(acc_rate + 0.01)	+0.0660	0.0950	[−0.1203, 0.2523]	0.487

Notes: All specifications estimate the overall ATT with the exposure-weighted aggregation, except the cohort-balanced row. Standard errors are district-clustered bootstrap standard errors with 1,000 iterations and deterministic random seeds. ¶The cohort-balanced aggregate weights each cohort equally rather than by cohort size and post-treatment exposure; it is reported as a weighting-robustness check (Section 5.4). †The pre-COVID specification recodes cohorts first treated after December 2019 as never-treated. ‡The risk-proxy outcome is accidents per 100,000 (population × city-wide annual cycling modal share); it is reported as a secondary exposure rescaling for completeness. The wide confidence interval reflects amplification of variance from dividing by the small (≈ 0.07–0.11) modal-share value; the counter-based outcome is therefore featured in the main text instead. §The counter-based risk outcome is accidents per 10^6 city-wide cyclist-counter trips, aggregated over the 9 districts with continuous 2015–2024 counting-station coverage; see Section B.3. Both exposure-rescaled outcomes are on different scales than the resident-rate specifications and are not directly comparable in levels; the log specifications are also on a different scale. Because the two exposure-rescaled outcomes carry very wide confidence intervals, their agreement with the headline null reflects their low precision at least as much as a substantive robustness (Section B.3). *Pointwise significant at the 5% level. *Source:* Own calculations.

A.4. Anticipation Sensitivity

The baseline specification sets anticipation to zero periods. Because cycling-infrastructure projects in Vienna are typically announced and visible during construction, Table A.4 re-estimates the overall ATT under anticipation horizons of three and six months as a defensive check, holding every other estimator setting identical to the baseline. The estimates move on either side of the baseline (+0.041 for $a = 3$ and −0.034 for $a = 6$) – the small-sample sensitivity expected when only seven cohorts contribute to identification – but both 95% confidence intervals contain zero, so neither horizon overturns the qualitative conclusion.

Table A.4.: Anticipation Sensitivity

Anticipation (months)	ATT	SE	95% CI	p-value
$a = 0$ (baseline)	−0.0051	0.0216	[−0.0475, 0.0374]	0.815
$a = 3$	+0.0413	0.0352	[−0.0276, 0.1102]	0.240
$a = 6$	−0.0341	0.0242	[−0.0815, 0.0133]	0.158

Notes: The anticipation horizons are re-estimated with estimator settings identical to the baseline (never-treated controls, doubly-robust estimation, varying base period, 1,000 cluster-bootstrap iterations, district clusters). *Source:* Own calculations.

B. Data and Measurement Notes

This appendix documents three measurement issues referenced in Chapters 3 and 6: the geometric validation of the reported project lengths, the coverage of Vienna's cyclist-counting stations, and the alternative exposure measures assessed during the design phase.

B.1. Length Validation

To verify that the reported project-length values are accurate enough to make the 800-metre treatment threshold unambiguous, the analysis code cross-checks them against geometric distances computed from the projects' route geometries, projected to the Austrian national grid (MGI/Austria GK East). Across all projects with available geometry, the maximum absolute deviation between reported and geometric length is approximately 3.3 m (at the longest project, Favoritenstraße: 3,081 m reported vs. 3,078 m geometric), the mean absolute deviation is below 0.5 m, and larger relative errors (up to about 2.2%) occur only on a few very short projects whose absolute errors remain below 4 m. The threshold assignment is therefore robust to length-measurement noise. The numerical results, and a note on the optional geospatial software this check requires, are part of the replication package (Section C.2).

B.2. Cycling Exposure Data

Vienna's automatic cyclist-counting-station network provides monthly volume measurements for 11 of the 23 districts. Coverage is unbalanced across treated and never-treated districts, and station placement is non-random (typically on trafficked corridors rather than residential streets), so incorporating these data into the baseline panel would require imputation or sample restriction; the baseline outcome is consequently scaled by resident population. Figure B.1 reports the aggregated city-wide series for descriptive context: a moderate upward trend with strong seasonality and a temporary drop during the spring of 2020, consistent with the broader rise in cycling activity in Vienna over the sample period. It is not used in any estimation; Section 6.2 of the main text discusses the implications of resident-normalised outcomes for the interpretation of the headline result.

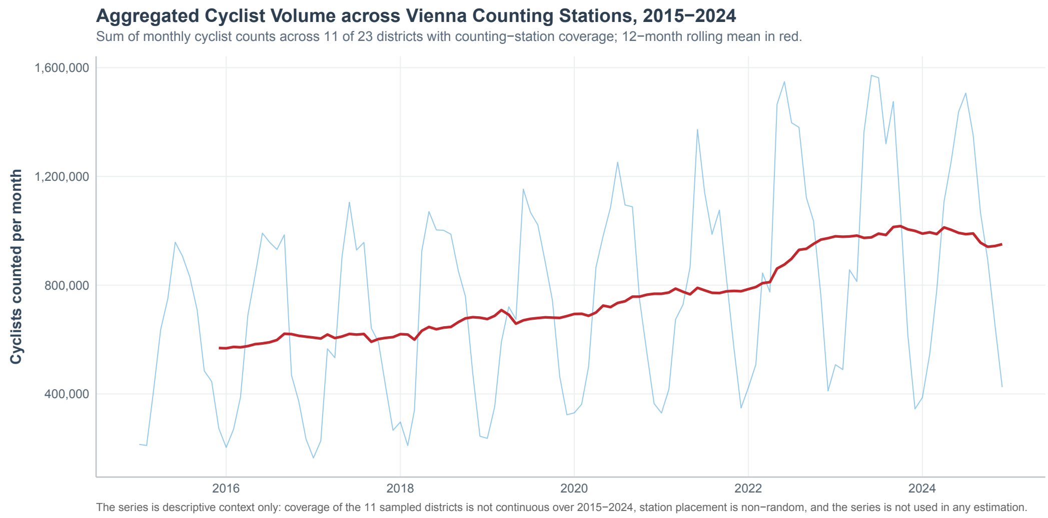

Figure B.1.: Aggregated Cyclist Volume across Vienna Counting Stations, 2015–2024

Notes: The figure shows the sum of monthly cyclist counts across the 11 of 23 districts with counting-station coverage, with a 12-month rolling mean overlaid. Station placement is non-random; the series serves as descriptive context only and is not used in any estimation. *Source:* Stadt Wien (OGD); own calculations.

B.3. Alternative Exposure Measures

Because the resident-normalised outcome does not isolate per-trip risk, several alternative exposure data sources were assessed during the design phase; none supports a district-by-month exposure correction within the present sample:

- **Strava Metro:** anonymised, district-resolved cycling volumes widely used in cycling-safety research, but academic researcher access was paused at the time of writing, so the data could not be obtained.
- **Bike-sharing trip data (Citybike Wien / WienMobil Rad):** the mid-2022 operator change creates a structural break within the 120-month panel, station coverage is concentrated in inner districts, and – most fundamentally – bike-sharing usage is itself plausibly responsive to infrastructure investment, making the trip count a “bad control” in the sense of Angrist and Pischke (2009).

- **Mobility-app reports (Google, Apple):** both programmes were discontinued after 2022 and neither resolves the data to the Vienna-district level.
- **Google Trends and *Mikrozensus Verkehr*:** the former is a sentiment proxy that does not measure trips; the latter is too infrequent and too geographically coarse to construct a district-month panel.

The closest available substitute is the counter-based risk outcome reported in Tables 5.2 and A.3, which rescales the denominator by an index of monthly cyclist counts aggregated over the 9 districts with continuous 2015–2024 counting-station coverage (district codes $1, 2, 4, 5, 7, 13, 19, 22, 23$):

$$\text{risk}_{i,t} \;=\; \frac{\text{accidents}_{i,t}}{\text{city-wide counter trips}_t} \;\times\; 10^6.$$

Because the counter index is city-wide and time-specific, it is largely absorbed by the time fixed effects implicit in the Callaway–Sant'Anna design; the specification therefore primarily changes the outcome scale rather than identifying a per-trip risk effect (overall ATT +0.604, SE 0.712, $p = 0.396$). The modal-share rescaling reported in Table A.3 (+6.20, SE 16.47, $p = 0.707$) behaves analogously, with variance amplified by the small ($\approx$ 0.07–0.11) modal-share value. Both agree with the headline null, but given their very wide confidence intervals this convergence reflects their shared imprecision at least as much as a substantive robustness; none of the three specifications constitutes a genuine district-month exposure correction. Such a correction remains infeasible with currently published Vienna data; constructing one is a natural extension for future research.

C. Estimation and Replication Details

C.1. Estimator Settings and Inference

The empirical analysis uses the Callaway–Sant'Anna framework as implemented in the `did` package. The baseline specification estimates group-time effects via `att_gt()` with the doubly-robust estimator, the never-treated comparison group, zero anticipation, and a varying base period. Aggregated effects use `aggte()`: `type = "simple"` for the overall ATT, `type = "dynamic"` for the event-study path, and `type = "group"` both for the cohort-specific summaries and for the cohort-balanced overall aggregate that serves as the weighting-robustness check in Section 5.4. The `"simple"` aggregate weights group-time effects by cohort size and the number of post-treatment periods, whereas the `"group"` overall aggregate averages the cohort-specific effects with equal cohort weight.

All confidence intervals are Wald-type intervals of the form $[\hat{\tau} - 1.96\,\widehat{\mathrm{SE}},\ \hat{\tau} + 1.96\,\widehat{\mathrm{SE}}]$, where $\widehat{\mathrm{SE}}$ is the district-clustered multiplier bootstrap standard error. p-values for cohort-specific and aggregated effects are computed as Wald-type normal approximations, $p = 2(1 - \Phi(|\hat{\tau}/\widehat{\mathrm{SE}}|))$, because `did::aggte()` does not return resampling-based p-values directly. Under small group sizes the bootstrap distribution may deviate from normality and the normal approximation may be imprecise; this caveat applies in particular to cohorts with $n_g \leq 2$ (Cameron et al., 2008).

The exact baseline estimator call is

```
att_gt(yname = "acc_rate", tname = "period", idname = "district_id",
       gname = "G", data = analysis_data, panel = TRUE,
       control_group = "nevertreated", est_method = "dr",
       base_period = "varying", anticipation = 0,
       bstrap = TRUE, biters = 1000, clustervars = "district_id",
       cband = FALSE, alp = 0.05).
```

No covariates are passed to the estimator, so identification proceeds under unconditional parallel trends. Treatment is defined upstream in the analysis script as the first project completion of at least 800 metres per district (no intervention-type filter). Deterministic random seeds are set before every bootstrap-based estimation, so all reported standard errors and p-values are exactly reproducible; the seed scheme is documented in Section C.2.

C.2. Replication Information

The replication package is contained in the `R-Code/` directory and includes a self-contained R analysis script that produces all empirical outputs (tables, figures, CSV exports) automatically. Six input data files in `Data/`, a `README.md` with step-by-step reproduction instructions and a complete output-to-thesis mapping, a `DATA_CODEBOOK.md` with variable definitions and data provenance, and `Output/sessionInfo.txt` documenting the R version (4.5.2) and all package versions used round out the package. Diagnostic logfiles regenerated on each run are written to `Output/logs/`. The complete package is available from the author upon request.

Bootstrap seeds. The R script sets individual deterministic random seeds before each bootstrap-based `att_gt()` estimation: `set.seed(2025001)` through `set.seed(2025009)`, plus `set.seed(2025011)` and `set.seed(2025012)`, for the eleven main and robustness specifications; `set.seed(2025010L + a)` for the anticipation-sensitivity horizon a (used: $a = 3 \Rightarrow$ `2025013`, $a = 6 \Rightarrow$ `2025016`); and `set.seed(2025100 + i)` for leave-one-out iteration i. The value `2025010` is therefore reserved as the anticipation base and is not used as a stand-alone specification seed. The `aggte()` aggregation steps draw from the deterministic random-number stream that follows each seeded estimation, so for the fixed pipeline order all standard errors, confidence intervals, and p-values are exactly reproducible, provided the same R version and package versions are used. Point estimates are deterministic and not affected by bootstrap randomness.

Code outputs and thesis tables. The R script generates the LaTeX table fragments and supporting CSV files. The full event-study table in Section A.1 is included directly from this generated output; the remaining thesis tables were re-set and annotated manually in the thesis source, so their footnote text and column formatting differ from the raw code output. For numerical verification, the CSV exports (`att_overall.csv`, `att_dynamic.csv`, `att_by_cohort.csv`, `robustness_summary.csv`, `robustness_anticipation.csv`) are the definitive machine-readable reference.

Data sources. Three of the six input files are from publicly available Austrian and Viennese open government data sources. The cycling-infrastructure project data are maintained by Radlobby Wien and are publicly accessible but not part of the OGD portal. The accident microdata are not published via Open Government Data; they were obtained on request from the City of Vienna's municipal statistical office (MA 23, `anfragen@ma23.wien.gv.at`) on 5 September 2025. The aggregated district-month extract used in the analysis is included in the replication package with the permission of MA 23; the underlying accident-level microdata remain with MA 23 and Statistik Austria.

Bibliography

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, *61*(4), 962–973. doi: 10.1111/j.1541-0420.2005.00377.x

Callaway, B., & Sant'Anna, P. H. C. (2021a). *did: Difference in differences*. https://bcallaway11.github.io/did/. (R package version 2.3.0)

Callaway, B., & Sant'Anna, P. H. C. (2021b). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. doi: 10.1016/j.jeconom.2020.12.001

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, *90*(3), 414–427. doi: 10.1162/rest.90.3.414

de Chaisemartin, C., & D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–2996. doi: 10.1257/aer.20181169

Elvik, R., & Bjørnskau, T. (2017). Safety-in-numbers: A systematic review and meta-analysis of evidence. *Safety Science*, *92*, 274–282. doi: 10.1016/j.ssci.2015.07.017

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. doi: 10.1016/j.jeconom.2021.03.014

Harris, M. A., Reynolds, C. C. O., Winters, M., Cripton, P. A., Shen, H., Chipman, M. L., . . . Teschke, K. (2013). Comparing the effects of infrastructure on bicycling injury at intersections and non-intersections using a case-crossover design. *Injury Prevention*, *19*(5), 303–310. doi: 10.1136/injuryprev-2012-040561

Jacobsen, P. L. (2003). Safety in numbers: More walkers and bicyclists, safer walking and bicycling. *Injury Prevention*, *9*(3), 205–209. doi: 10.1136/ip.9.3.205

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, *4*(3), 305–322. doi: 10.1257/aeri.20210236

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199. doi: 10.1016/j.jeconom.2020.09.006

Teschke, K., Harris, M. A., Reynolds, C. C. O., Winters, M., Babul, S., Chipman, M., ... Cripton, P. A. (2012). Route infrastructure and the risk of injuries to bicyclists: A case-crossover study. *American Journal of Public Health*, *102*(12), 2336–2343. doi: 10.2105/AJPH.2012.300762