Exposure-Normalized Bicycle Crash Risk Along Berlin Routes

Eric Berger [*] **Edward Eichhorn** [*] **Liaisan Faidrakhmanova** [*] **Luise Grasl** [*] **Tobias Schnarr** [*]

Abstract

We investigate bicycle crash risk on Berlin's urban street network, addressing a key limitation of many safety analyses: raw crash counts conflate danger with demand and fail to distinguish intrinsically risky locations from high-use roads. We combine police-reported crashes with a city-wide dataset of measured bicycle volumes to compute exposure-adjusted risk at the street-segment and junction levels. Risk estimates, available at monthly resolution, can be conditioned on contextual factors (e.g., time of day, weather) to capture temporal variation. Aggregating risk to arbitrary routes enables comparisons that trade off safety against convenience. The result is a reproducible framework for context-aware, exposure-controlled bicycle safety analysis and routing.

/edit, when we
nal results

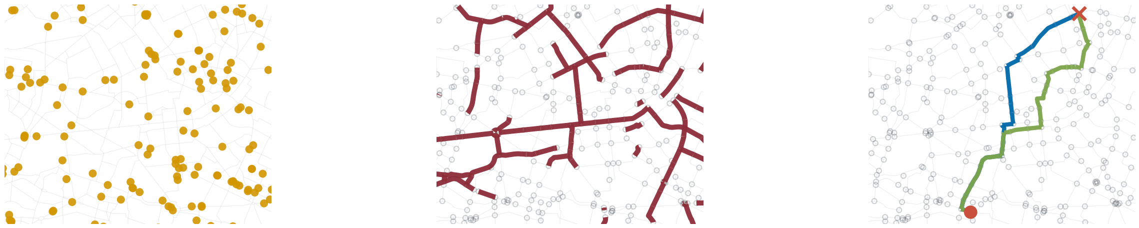
1. Introduction

Cycling safety analyses often rely on raw crash counts, which conflate danger with demand: streets that attract many cyclists tend to accumulate more incidents even when per-rider risk is low (Lücken, 2018). This obscures intrinsically risky locations and limits both targeted interventions and everyday route choice, especially in dense urban networks such as Berlin (Uijtdewilligen et al., 2024). We address this by estimating exposure-adjusted crash risk on the street network via a relative-risk formulation that separates cyclist demand from intrinsic danger and remains stable under sparse or unevenly distributed observations. The resulting estimates are transformed into expected crash costs and propagated to route-level scores suitable for navigation. Our study combines police-recorded crashes from the German Unfallatlas (Berlin subset) (German Federal Statistical Office, 2025) with a city-wide dataset of measured bicycle volumes at the street-segment level (Kaiser, 2025). We compute exposure-adjusted relative risk for individual street segments and, motivated by the concentration of crashes at intersections, derive junction-level risk by aggregating volumes from adjoining segments. Risk is estimated at monthly resolution and can be conditioned on contextual factors such as time of day and weather to capture temporal variability. We then aggregate segment- and junction-level expected crash costs to evaluate arbitrary routes and compare alternatives that reduce estimated risk while maintaining comparable convenience in terms of distance or travel time, see Figure 1. We address the question: how can exposure-adjusted relative crash risk be estimated from measured bicycle volumes and integrated into context-aware routing? Our work makes the following contributions: (i) a reproducible pipeline for estimating exposure-adjusted relative crash risk at street and junction levels from measured cyclist volumes, with support for context-conditional analysis; and (ii) a route-scoring method that integrates network-level risk into safety-aware routing under convenience constraints.

Related Work. To avoid conflating danger with demand, prior work normalizes bicycle crashes by cyclist exposure (Lücken, 2018). City-scale studies show that exposure-normalized risk yields more informative spatial patterns than raw counts and that finer temporal resolution improves inference, while noting persistent under-reporting in police records (Uijtdewilligen et al., 2024). A central challenge is obtaining reliable exposure: some approaches extrapolate city-wide volumes from sparse counters using learning-based models and multi-source features, with short measurement campaigns improving predictions at new locations (Kaiser et al., 2025a), whereas more recent efforts provide street-segment datasets of measured bicycle volumes, enabling downstream safety analyses without modeling exposure (Kaiser et al., 2025b). At the network level, studies define risk as crashes per unit exposure on links and address practical issues such as spatial snapping of crashes, allocating events near intersections, and integrating safety metrics into routing under convenience constraints (Wage et al., 2022). Intersection safety is critical: case analyses document strong crash concentrations at junctions and stress controlling for exposure when comparing infrastructure types or locations (Medeiros et al., 2021). We estimate exposure-adjusted crash risk for Berlin at street-segment and junction levels, enable context-conditional risk analysis, and integrate network-level risk into routing decisions.

[*]Equal contribution. Correspondence to: TS <tobias-marco.schnarr@student.uni-tuebingen.de>.

Project report for the "Data Literacy" course at the University of Tübingen, Winter 2025/26 (Module ML4201). Style template based on the ICML style files 2025.



(a) Data

(b) Risk estimation

(c) Safety routing

Figure 1. **Safety-aware routing pipeline for the Berlin cycling network.** Panels (a–c) show an example month (June 2021) and are zoomed in for readability; see Section 3 for formal definitions and notation. **(a):** police-recorded bicycle crashes (points) and street segments with measured cyclist exposure (lines, used as the base network in all panels). **(b):** monthly segment risk (accidents per 10,000 trips); high-risk segments in red are the top 90th percentile of $r_{s,t}$ in the displayed month. Circles mark junctions (degree ≥ 3), for which we also estimate junction risk from exposure aggregated over incident segments. **(c):** shortest path (blue) versus a safer alternative (green) chosen to reduce route risk under a distance-detour constraint. Filled circle and cross denote origin and destination, respectively; circles show junctions for reference.

The paper is organized as follows: we present our data in Section 2, describe our methods in Section 3, report the results in Section 4, and conclude in Section 5.

2. Data

We combine police-recorded bicycle crashes with measured cyclist exposure for the city of Berlin. Crash data are drawn from the Berlin subset of the German *Unfallatlas* (German Federal Statistical Office, 2025) and filtered for bicycle-related incidents. Exposure comes from a city-wide dataset of measured bicycle volumes aggregated at the street-segment level (Kaiser et al., 2025b). The street network is represented as polyline segments with associated monthly cyclist counts. The resulting dataset spans 2019–2023 and covers 4,958 street segments and 2,924 junctions, with 33,181 recorded bicycle crashes. At monthly resolution the data are sparse: in a typical month fewer than 5% of segments and about 3% of junctions record at least one crash, and some periods include segments with zero measured exposure. This sparsity motivates the exposure-normalized risk estimation and small-count stabilization described in Section 3.

Preprocessing. To enable network-scale analysis, all layers are harmonized to a common street-network topology and a projected coordinate reference system. Crash locations are matched to segments using nearest-segment assignment. To capture the concentration of crashes at intersections, we identify junctions as nodes where at least three segments meet; crashes within a fixed radius are assigned to the nearest junction, and junction exposure is computed from the exposures of incident segments. Matched crashes and exposures are aggregated to monthly resolution, and segments and months with zero exposure are identified and dropped. To obtain stable risk estimates under sparse observations, monthly aggregates are pooled to yearly totals and an Empirical Bayes approach is applied as described in Section 3. The resulting yearly segment- and junction-level risk estimates serve as inputs to all routing analyses.

3. Methods

Empirical Bayes relative risk. For each street segment s and month t, let $A_{s,t}$ denote the number of police-recorded bicycle crashes and $E_{s,t}$ the measured cyclist exposure. We assume that, within a given month, crash incidence is proportional to exposure under a "no special risk" baseline, yielding the expected number of crashes

$$\widehat{A}_{s,t} = A_t \frac{E_{s,t}}{E_t}, \qquad A_t = \sum_s A_{s,t}, \;\; E_t = \sum_s E_{s,t}.$$

Although crashes and exposure are aggregated at monthly resolution in preprocessing, sparse observations motivate estimating relative risk at yearly resolution. For each year y comprising months $\mathcal{T}(y)$, we pool monthly aggregates to yearly totals

$$A_{s,y} = \sum_{t \in \mathcal{T}(y)} A_{s,t}, \qquad E_{s,y} = \sum_{t \in \mathcal{T}(y)} E_{s,t},$$

and define $A_y = \sum_s A_{s,y}$ and $E_y = \sum_s E_{s,y}$. The corresponding baseline expectation is

$$\widehat{A}_{s,y} = A_y \frac{E_{s,y}}{E_y}.$$

To obtain stable estimates under sparse observations, we introduce a latent relative-risk multiplier $\theta_{s,y}$ and model

$$A_{s,y} \mid \theta_{s,y} \sim \text{Poisson}(\widehat{A}_{s,y}\, \theta_{s,y}),$$
$$\theta_{s,y} \sim \text{Gamma}(\alpha, \alpha),$$

where the Gamma distribution is parameterized in shape–rate form, enforcing a unit prior mean $\mathbb{E}[\theta_{s,y}] = 1$. The hyperparameter α is estimated via empirical Bayes by maximizing the marginal likelihood pooled across all segments and years. By conjugacy, the posterior mean

$$r_{s,y} = \mathbb{E}[\theta_{s,y} \mid A_{s,y}, \widehat{A}_{s,y}] = \frac{A_{s,y} + \alpha}{\widehat{A}_{s,y} + \alpha}$$

serves as our smoothed segment-level relative risk. This estimator shrinks extreme values toward 1, with stronger shrinkage for segments with low expected crash counts.

To capture the concentration of crashes near intersections, we also estimate junction-level relative risk. A junction is defined as any network node with degree ≥ 3; crashes within a fixed radius of its centroid are assigned to that junction. Because a traversal typically contributes exposure to two incident segments (entering and exiting), we approximate junction exposure by a half-sum of incident segment exposures,

$$E_{v,y} = \tfrac{1}{2} \sum_{s \in \mathcal{I}(v)} E_{s,y}, \tag{1}$$

which avoids double-counting. Aggregating exposure over incident links is common when turning movements are unavailable (Hakkert et al., 2002; Wang et al., 2020). The expected number of junction crashes is defined as $\widehat{A}_{v,y} = A_y E_{v,y}/E_y$, and junction relative risk $r_{v,y}$ is estimated using the same Empirical Bayes formulation and hyperparameters as for segments. All risks are estimated independently by year.

Routing graph. For each year y, we build an undirected graph $G_y = (V_y, E_y)$ from the street network: nodes represent segment endpoints and edges represent street segments with length ℓ_e. Each graph edge e corresponds to a street segment s and inherits its estimated yearly risk, denoted $r_{e,y} = r_{s,y}$. Junction identifiers and risks are mapped to nodes via spatial snapping in a projected coordinate system, yielding a year-specific, risk-annotated network.

Safety-aware routing. We compare shortest-distance routes with alternatives that reduce estimated crash risk under a bounded detour. For year y, the length of a route P is

$$L(P) = \sum_{e \in P} \ell_e.$$

To account for both segment- and junction-level exposure, the risk contribution of edge $e = (u, v)$ is defined as

$$\rho_{e,y} = r_{e,y} + \eta \, \frac{r_{u,y} + r_{v,y}}{2},$$

where $r_{u,y}$ and $r_{v,y}$ are junction risks (set to zero for non-junction nodes) and $\eta \geq 0$ weights the contribution of junction risk. We interpret these relative-risk weights as an *additive surrogate* for cumulative route risk, not as probabilities.

Given an origin–destination pair, the baseline route P_{dist} minimizes $L(P)$. The safety-aware route solves

$$\begin{aligned} P_{\text{safe}} = \arg\min_P R(P) = \sum_{e \in P} \rho_{e,y} \\ \text{s.t. } L(P) \leq (1 + \varepsilon)\, L(P_{\text{dist}}), \end{aligned} \tag{2}$$

where ε is the allowable relative detour. We approximate this constrained problem using a weighted-sum sweep: for $\lambda \in \Lambda$,

$$P(\lambda) = \arg\min_P \sum_{e \in P} \big(\rho_{e,y} + \lambda\, \ell_e\big),$$

then select among feasible candidates the route with minimal $R(P)$, breaking ties by shorter $L(P)$. Shortest paths are computed using Dijkstra's algorithm. Routing is performed independently by year.

Evaluation metrics. For each origin–destination pair, we report the relative length increase

$$\Delta_L = \frac{L(P_{\text{safe}}) - L(P_{\text{dist}})}{L(P_{\text{dist}})}$$

and the relative risk reduction

$$\Delta_R = \frac{R(P_{\text{dist}}) - R(P_{\text{safe}})}{R(P_{\text{dist}})}.$$

Pairs with $R(P_{\text{dist}}) = 0$ are excluded from Δ_R due to an undefined denominator. These metrics quantify how bounded detours trade distance for reductions in relative crash risk.

4. Results

5. Discussion and Conclusion

We provide implementation details, hyperparameters, and supplementary material, available at `https://github.com/ytobiaz/data_literacy`.

References

German Federal Statistical Office. Unfallatlas. https://unfallatlas.statistikportal.de/, 2025. Interaktive Kartenanwendung zu Straßenverkehrsunfällen mit Personenschaden in Deutschland.

Hakkert, A. S., Braimaister, L., and Van Schagen, I. The uses of exposure and risk in road safety studies. Technical report, SWOV Institute for Road Safety SWOV, Leidschendam, 2002.

Kaiser, S. K. Data from: Spatio-temporal graph neural network for urban spaces: Interpolating citywide traffic volume, May 2025. URL https://doi.org/10.5281/zenodo.15332147.

Kaiser, S. K., Klein, N., and Kaack, L. H. From counting stations to city-wide estimates: data-driven bicycle volume extrapolation. *Environmental Data Science*, 4:e13, 2025a. doi: 10.1017/eds.2025.5.

Kaiser, S. K., Rodrigues, F., Azevedo, C. L., and Kaack, L. H. Spatio-temporal graph neural network for urban spaces: Interpolating citywide traffic volume, 2025b. URL https://arxiv.org/abs/2505.06292.

Lücken, L. On the variation of the crash risk with the total number of bicyclists. *European Transport Research Review*, 10(2):33, 2018. doi: 10.1186/s12544-018-0305-9. URL https://doi.org/10.1186/s12544-018-0305-9.

Medeiros, R. M., Bojic, I., and Jammot-Paillet, Q. Spatiotemporal variation in bicycle road crashes and traffic volume in berlin: Implications for future research, planning, and network design. *Future Transportation*, 1(3):686–706, 2021. ISSN 2673-7590. doi: 10.3390/futuretransp1030037. URL https://www.mdpi.com/2673-7590/1/3/37.

Uijtdewilligen, T., Ulak, M. B., Wijlhuizen, G. J., Bijleveld, F., Geurs, K. T., and Dijkstra, A. Examining the crash risk factors associated with cycling by considering spatial and temporal disaggregation of exposure: Findings from four dutch cities. *Journal of Transportation Safety & Security*, 16(9):945–971, 2024. doi: 10.1080/19439962.2023.2273547. URL https://doi.org/10.1080/19439962.2023.2273547.

Wage, O., Bienzeisler, L., and Sester, M. Risk analysis of cycling accidents using a traffic demand model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B4-2022:427–434, 2022. doi: 10.5194/isprs-archives-XLIII-B4-2022-427-2022. URL https://isprs-archives.copernicus.org/articles/XLIII-B4-2022/427/2022/.

Wang, K., Zhao, S., and Jackson, E. Investigating exposure measures and functional forms in urban and suburban intersection safety performance functions using generalized negative binomial - p model. *Accident Analysis & Prevention*, 148:105838, 2020. ISSN 0001-4575. doi: https://doi.org/10.1016/j.aap.2020.105838. URL https://www.sciencedirect.com/science/article/pii/S0001457520316584.