Reducing Cycling Injury Risk While Cycling Grows

A report to the Road Safety Trust by Professor Rachel Aldred, University of Westminster

Acknowledgements

I would like to thank colleagues who worked on this project, primarily Georgios Kapousizis, who worked on this project as a Research Associate, Dr. Anna Goodman who provided expert statistical input, and Irena Itova, who assisted with aspects of the GSV and Google Streetview analysis. Thanks to the Road Safety Trust for supporting the project, DfT and others (e.g. Basemap) for providing data, and our stakeholder group for providing useful input first in-person and later virtually.

Despite Covid-related challenges, we have completed the research successfully, and generated some novel and important findings. As dissemination has been delayed, we used part of our no-cost extension to conduct additional research building on initial findings confirming the importance of junction status. Our additional research paired control and injury sites depending on junction status, and hence generated differential results for the two types of site.

I am also able here to draw on some results from a collaboration with Thomas Adams (TfL), who has conducted related research in London. I have drafted a third paper with him which has been published in *Transport Findings* (in addition to the two submitted papers based on this research, under review with *Accident Analysis and Prevention* and *International Journal of Environmental Research and Public Health*). The collaboration helps contextualise our findings around cycle infrastructure, primarily by providing a comparison with higher quality cycling infrastructure built in London.

About the Project

This project was funded by the Road Safety Trust. It began in January 2019, and incorporating a nine-month no-cost Covid extension, will finish in March 2021.

The study uses a case-crossover method to identify correlates of cycling injury risk in Britain for weekday commuters in 2017, by comparing the characteristics of injury sites with matched 'control' sites. It is innovative in its use of such a methodology. While rarely used in road safety research, this approach provides a way to control for exposure (i.e. amount of cycling in different types of location), which otherwise will tend to skew results. For instance, without controlling for exposure, infrastructure that tends to attract more cyclists may appear riskier, simply because more cyclists will tend to mean more injuries, even where the per-cyclist risk is lower.

The case-crossover method has the additional merit of controlling for differences between cyclists, because injured individuals act as their own controls (here by selecting a site at which they were *not* injured). Hence for instance, where we find a type of infrastructure increases risk, it cannot be explained by individuals who behave in a riskier manner disproportionally using that type of infrastructure, as control sites relate to the same individuals who were injured.

The project is additionally innovative in using an algorithm to predict cyclist routes. Routes are required by the case-crossover method, as control sites are generated by random selection of points from the route the cyclist was following before their injury. However, we did not have cyclist routes, although we did have the injury sites (from Stats19) and home postcodes (provided by DfT), which for people injured in the weekday morning peak are likely (unless far from the injury site) to be their start point. We used a cyclist journey planner (Cyclestreets) to route our cyclists between origin and destination, and then randomly generated our control points from this algorithmically generated route.

As shown in the summary table below, we have identified twelve key findings from these two papers, which we will share with policy-makers in a series of short briefings, once this report has been approved by the Trust. Many are novel (e.g. on high streets, or guard railing, or mini-roundabouts) and we expect that the planned series of one- or two-page briefings (accompanied by short video recorded explanations, and presented at conferences and webinars as the opportunity arises, and at our own events) will prove popular and contribute to policy and practice.

I would like to thank the Road Safety Trust again for supporting this research and look forward to sharing our findings.

Professor Rachel Aldred, 20th January 2021

Summary of findings

The table below highlights twelve key findings, context needed to understand them further, and the implications for policy that we derive from them. While we are yet to move into the full dissemination phase, the below provide an indication of the key messages we will seek to convey through practitioner-focused briefings. As mentioned above, we also draw on recent findings in a study by *Adams (2020)*[1], whose recent study uses related methods but covers only London and hence can explore the impact of higher-quality cycling infrastructure than generally found across Britain as a whole.

The table also includes references and links to the fuller discussion later in the report.

Finding	Context/meanings	Policy Implications
1. Intersections are much riskier for cyclists than non-intersection locations, as are main roads and wider roads. *[2. Effects of road type variables]*	While not a novel finding, this confirms other research.	Measures to protect cyclists (e.g. high-quality infrastructure) are needed most on main roads, multi-lane roads, and at junctions. While routes away from main roads can help reduce risk, given directness and destination proximity many cycle trips will still require main road usage.
2. Where there is an intersection, presence of a roundabout (whether a standard or a mini-roundabout) substantially raises risk, compared to a signalised junction. *[6. Examination of differential effects between non-intersection versus intersection points]*	Other research has suggested roundabouts are uniquely dangerous for cyclists (while they can reduce injury risk for other road users). Our research confirmed this but found a similar elevation in risk related to presence of a mini-roundabout, which have often been installed in UK residential areas as a traffic calming measure.	Mini-roundabouts should be re-considered as a traffic calming measure, as it appears that like conventional roundabouts, they may increase risk to cyclists. At conventional roundabouts, conversion to traffic lights and/or measures to protect cyclists should be considered.
3. High streets (with clusters of retail and related activity) are riskier for people cycling. *[1. Effects of area-type variables]*	While main roads are established as risky for cyclists, more place-focused high streets have so far been overlooked as concentrating risk. This is likely because of competing kerbside activities.	High streets should not be assumed to be benign for cycling, where high volumes of conflicting movements are likely. High streets should like main roads be considered for high-quality cycle infrastructure, although other measures such as (part)pedestrianisation may be appropriate too depending on the context.
4. Sites with bus lanes are riskier for people cycling than sites without, although this is partly mitigated around bus stops.	There is relatively little research on safety of shared bus lanes, partly because although they are a common form of cycle infrastructure across Britain, this is less the case in other countries. These findings suggest that bus lanes put cyclists at	Shared bus lanes, while preferable to cyclists not being permitted to use bus lanes, are not a form of safe cycle infrastructure. Safer forms of cycle infrastructure should be implemented along busy bus corridors. Where cyclists must in

[1] Presented at the European Transport Conference: due to the synergy between the two projects, Adams and I have since published a co-authored paper based on his research, which I see as a sister paper to the two nationally-focused papers from this research.

[3. Effects of street infrastructure variables]	higher risk, even controlling for their being on wider, busier roads. The national picture to some extent contrasts with London, shown by Adams' work. This may be because London's bus lanes may have longer operational times, are more likely to prohibit car parking (via Red Route Clearways) and/or could be related to the differing organisation of buses in London (the franchise model).	the short to medium term continue to share bus lanes, lessons should be learnt from London bus lane operation.
5. Sites with painted (usually 'advisory') cycle lanes are riskier for people cycling than sites without. *[3. Effects of street infrastructure variables]*	Some research has found similar results, although other studies have found a protective effect. Adams' London study, like ours, found that advisory painted lanes put cyclists at higher risk, with a similar coefficient to that found here – even though the quality of painted lanes in London was probably higher than among our nation-wide sample.	Painted cycle lanes should not generally be seen as a form of safe cycle infrastructure.
6. We found no protective effect from what we called 'cycle tracks' (in practice, largely shared footways), and the presence of this infrastructure was riskier at junctions. *[3. Effects of street infrastructure variables]*	This finding is contrary to other research, and to Adams' recent (2020) London study which finds a substantial protective effect from higher-quality 'kerb-segregated' or 'stepped track' infrastructure (note: this does not cover shared footways).	Current 'protected cycle infrastructure' (or what existed in 2017) across Britain fails to protect cyclists and may at times put them at higher risk, although higher-quality protected infrastructure in London does offer protection. Cycle infrastructure must be built to higher quality standards, learning from London's achievements and the new LTN 1/20 Cycle Infrastructure Design Guidance which specifies this level of quality.
7. The presence of petrol stations and car parks raise injury risk for people cycling. *[3. Effects of street infrastructure variables]*	Such sites are likely to lead to frequent turning conflicts between straight on cyclists and turning drivers.	Protective infrastructure and/or other measures to reduce risk are particularly important at such sites.
8. Presence of on-street parked cars was found to raise risk in one analysis but not another. *[4. Effects of travel behaviour variables]*	Our first analysis, not matching for intersection status, found parked cars raised risk, but this was not the case when matching for intersection status (although there was a non-significant trend towards higher risk at non-intersection sites). However, in both types of model the odds ratio increased when adjusting for other factors – i.e. parked cars tend to be more common in safer	More research is needed on presence of parked cars and injury risk, but they should not be assumed to be a form of traffic calming benign to cyclists. Our measure of parked car presence was crude (anywhere within our four Google images per site) and would have been affected

	(residential) streets but tend to reduce the protective effect of those streets.	by time of day and day of week the Streetview car passed.
9. Guard railing raises risk for people cycling. *[3. Effects of street infrastructure variables]*	Confirms the result of one Transport for London study looking at all road users (rather than only cyclists).	Guard railing should be removed, as has been done in many parts of London.
10. Lower speed limits are associated with reduced risk in univariate analysis, but this association disappears when adjusting for road type, width, etc. *[2. Effects of road type variables]*	Lowering speed limits alone may not reduce risk to people cycling, if road geometry remains the same. (Note however that when comparing all and KSI injuries, there was a borderline significant interaction – p=0.09 – suggesting that lower speed limits may protect more against the more serious injuries).	Lowering speed limits should happen alongside changes in road design where these are needed, to improve compliance.
11. Increased congestion (measured by lower peak time morning speeds) raises risk. However, at intersections, the opposite is true for the intersecting road (i.e. higher speeds *increase* risk). *[4. Effects of travel behaviour variables]*	Congested roads may simply mean that there are more potential collision partners. (Note that when comparing all and KSI injuries, there was a borderline significant interaction – p=0.11 – suggesting that congestion is more of a problem for slight injuries than KSI injuries, perhaps because of relatively low collision speeds).	Improved infrastructure and protection for people cycling is important at congested sites.
12. We found a 'safety in numbers' effect, i.e. that more cyclists on a section of route reduced injury risk. *[4. Effects of travel behaviour variables]*	This effect has been found in other research although its causation remains debated.	Building infrastructure that attracts more cyclists should also improve safety for them through the 'safety in numbers' effect, in addition to any safety effect from the infrastructure itself. This generates a road safety argument for building the types of infrastructure shown to increase cycling.

Methods

As described above, this project has used a case-crossover method to quantify cycling injury risk in relation to different types of route environment and infrastructure. The case-crossover method requires us to select control points representing where an individual might have been injured, had risks been identical along the whole of their route prior to injury. We did not have actual routes; however, using injury points, home locations, and an algorithm, we were able to estimate likely routes followed for commuters in the morning weekday peak.

From routes longer than 100m but shorter than 25km we randomly selected control points, ensuring that we only selected control points that were on or adjacent to the highway network (as injuries taking place for instance on a canal greenway would not be in scope for Stats19 police injury data; therefore neither should our control points be away from the highway).

Our first analysis compared these control points (one per route) with injury points. An important finding was the substantially higher risk experienced at junctions. Therefore, we decided to use some of our no-cost extension time (due to Covid-19) to investigate this further, and generated control points matched to intersection status (i.e. if an injury happened at an intersection, we selected a control point from another intersection on that route, and vice versa for non-intersection points). I describe below the procedure used in this second analysis, which in other respects was similar to the first set of analyses (e.g. the building of regression models to separate out the impact of different factors).

Data Sources

We obtained home postcode data from Department for Transport, for all cyclists injured in Britain during 2017[2]. For many trips the start location is a person's home, and this can be accurately predicted based on trip timing given that that >95% of cycle trips during the morning peak start from home. We used home location data alongside publicly available Stats19 injury data, which includes a range of variables from injury location to involvement of other vehicles, casualty gender and age group.

Generation of routes and control points

In Britain between 5am and 9:59am, Monday to Friday, 4,303 cyclists were injured during 2017. Of these 3,507 (81.5%) had full home postcode data. We used the Cyclestreets API (fastest-route option) to model routes from home postcode area centroids to injury points. We excluded points associated with routes longer than 25km (137 routes, or 3.9%). We excluded 29 points where injury occurred <100m from home, as this did not give sufficient scope for the control and injury point to differ in their characteristics. See Appendix 1: Selection of routing method for more details of the sensitivity testing we conducted related to this algorithm.

We initially generated one control point randomly from each of the 3,341 remaining routes, using ArcGIS Random Points. There were 62% injuries among 4,282 intersection points, and 29% injuries among 2,400 non-intersection points. The univariable odds ratio was 4.42 (95% CI 3.90, 5.00), and the odds ratio after adjusting for area, road, street infrastructure and travel behaviour variables was 3.43 (2.99, 3.93). In analyses restricted only to KSI injuries, the effect was 3.77 (2.68, 5.29). These strong effects matched our expectation that intersection status would be a major predictor of odds of injury, supporting our decision to generate a set of control points matched on intersection status for these analyses.

Route environment data

We sourced route environment data in a range of ways. This included datasets provided by partners (e.g. Basemap) or available online (e.g. OpenStreetMap) and use of Google Street View.

[2] While we did also have data from Northern Ireland, this represented only ~1% of all cycle injuries, and much route environment datasets only covered GB. Hence we decided to only cover GB in this analysis.

We assigned each point the following route environment characteristics, grouped *a priori* into four different categories:

1. Area type: urban/rural status, high street status, average small area deprivation.
2. Road type: road class, road width, road gradient, speed limit, motor connectivity ranking.
3. Nearby street infrastructure: Bicycle infrastructure, guard railing, bus lane, bus stop, metro/rail/tram stop, petrol station/car park, intersection status.
4. Travel behaviour: average AM peak speed, parked cars, cycle commuter flow.

Appendix 2: Route environment data sources presents details of how each of these variables was calculated.

Statistical modelling

We used conditional logistic regression, matching each injury point to its sampled control point matched on intersection status. We analysed our data guided by our four-category classification of environmental correlates into area type, road type, nearby street infrastructure, and travel behaviour.

We fitted the adjusted regression models using a hierarchal modelling structure, starting with the categories of variables conceptualised as most distal to the outcome, and continuing with categories of variables we saw as mediating more distal factors. In stratified analyses restricted to intersection points, we included variables on traffic signals and roundabouts, as additional elements of street infrastructure. We included road type and travel behaviour variables for the intersecting road where available. We conducted sensitivity analyses restricted to KSI casualties, and present results for tests for interaction between each predictor and whether the injury was a KSI or not.

Because our study is focusing on injuries occurring during the morning commute, control points will be closer to home and further from work than injury points, and on average places where people work are less residential and more commercial. Hence, we expected that injury points would generally have a higher workplace density than control points, as an artefact. This was indeed observed: workplace density was higher in the injury point for 1155 participants (34.6%), in the control point for 735 participants (22.0%), and similar (within 0.05) for 1451 participants (43.4%). To reduce confounding, we included workplace density in all adjusted models as a covariate.

Note also that the Propensity to Cycle Tool (PCT) route network (used to look up cycling volume, see Appendix 1) was created using an algorithm (Cyclestreets) to route cyclists between origin and destination, based largely but not only on directness. By contrast injuries can happen anywhere that cyclists travel. Our method therefore means control points are less likely to be 'off the PCT network', so less likely to get zero or very low cycle volume value. For this reason, when modelling cycle volume as a continuous variable we simultaneously entered a binary dummy variable identifying whether the route contained 0-5 versus 6+ cyclists.

We examined crude associations to guide how continuous variables should be entered into our model. Motor connectivity ranking was highly correlated with road class and other road type variables, so we entered it as a categorical variable. Otherwise where possible we entered continuous variables as linear terms, to increase power and avoid complications of interpretation. To limit the effect of outliers, we capped road width at 15m (276 higher values, or 4.1%, rounded to 15), average peak speed at 50 miles/hr (303 higher values, or 4.5%, rounded to 50) and the number of cycle commuters at 1000 (88 higher values, or 1.3%, rounded to 1000). After this, all continuous variables showed an approximately linear relationship in visual inspection, with no evidence of non-nonlinearity as judged by the inclusion of a quadratic term (all $p>0.05$ in adjusted analyses).

The proportion of variables with missing data ranged from 0 to 6.2% with respect to the road on which the crash happened. At intersections, the proportion with missing data ranged from 0 to 12.6% with respect to the second, intersecting road. We imputed this data using multiple imputation (25 imputations) under an assumption of Missing at Random. We confirmed that results were similar using a complete case analysis on the 2589 participants (77.5%) with complete data for both injury and control points.

Limitations

This study is limited in a variety of ways. We were only able to include weekday morning peak journeys, and because lacking journey start location we needed to use home postcode as a proxy. We had to exclude those injured cyclists for whom home postcode was not known. Our use of a modelling algorithm to route the cyclists could lead to bias, for instance, if cyclists in practice make more use of residential roads than is suggested by the algorithm. However, use of a relatively direct route (the Cyclestreets 'fast route' algorithm prioritises directness, but does avoid the very busiest roads where possible) is, we believe, likely to represent well enough cyclist routes, especially at commuting times. We were limited in route environment data sources available, and use of current streetview images may introduce bias, if for instance infrastructure has been built post-2017 (which might be particularly likely in more dangerous environments). Our data predominantly relates to slight injuries, these being most injuries recorded by the police.

Strengths

We used national data and controlled for cyclist volume and individual characteristics, through the case-crossover approach used. This is unusual and represents an innovative use of secondary data, allowing the research to be conducted without potentially intrusive and time-consuming primary data collection.

Results

Results below refer unless otherwise stated to our second set of analyses that matched on control and intersection group status. With the exception of analyses looking specifically at intersection characteristics (which we did not carry out in the first set of analyses), the direction and magnitude of most of the results are similar to those found in the unmatched analysis. Where this is not the case, it has been noted, and unmatched results are in Appendix 3: Additional results.

Sample characteristics

Characteristics of the 3341 individuals in our sample are shown in Table 1. The large majority were from England. 77% were male, 73% aged 25-59, and people living in the richest two-fifths of areas were somewhat underrepresented. Of injuries, 82% were slight, 17% serious and 0.4% fatal. The large majority, 91%, involved cars, taxis, or vans. Most injuries occurred when it was light (as expected given during the morning commute), in fine weather with dry road conditions.

Table 1: Characteristics of individuals and of their crash

Characteristic	Level	N (%)
Full sample	-	3341 (100%)
Country	England Scotland Wales	3159 (94.6%) 131 (3.9%) 51 (1.5%)
Sex	Male Female	2579 (77.2%) 762 (22.8%)
Age	0-15 16-24 25-39 40-59 60-74 75+	293 (8.9%) 415 (12.5%) 1276 (38.5%) 1139 (34.4%) 155 (4.7%) 34 (1.0%)
Small-area deprivation of home	Fifth 1 (richest) Fifth 2 Fifth 3 Fifth 4 Fifth 5 (poorest)	546 (17.3%) 569 (18.0%) 642 (20.3%) 778 (24.6%) 623 (19.7%)
Injury severity	Fatal Serious Slight	14 (0.4%) 578 (17.3%) 2749 (82.3%)
Striking vehicle	No other vehicle Cyclist HGV Bus Other motor vehicle, mostly cars	188 (5.6%) 20 (0.6%) 70 (2.1%) 38 (1.1%) 3025 (90.5%)
Light conditions	Light Dark	2933 (87.8%) 408 (12.2%)
Weather conditions	Fine, no high winds Other	2708 (85.4%) 464 (14.6%)
Road surface conditions	Dry Other	2401 (74.3%) 832 (25.7%)

Numbers add to less than 3341 for some variables due to missing data: in these cases, the % is calculated relative to those who non-missing data.

Effects of area, road, street infrastructure and travel behaviour

1. Effects of area-type variables

Being urban and on a high street were both significantly associated with an increased odds of injury in univariable analyses, but there was no association with area deprivation. The impact of being in an urban area attenuated and became no longer significant after mutual adjustment for presence of a high street (adjusted model 1), and then further attenuated upon additional adjustment. This suggests the univariable urban effect reflected the types of roads found in urban areas plus the higher concentration of high streets. The impact of being on a high street was somewhat attenuated after adjusting for road type, street infrastructure and travel behaviour, but a significant independent effect remained in the final adjusted model (3), suggesting some risk posed by aspects of the high street not captured in other variables (odds ratio 1.32, or a 32% increase in the odds of injury).

2. Effects of road type variables

All five variables were significantly associated with the odds of injury in univariable analyses. After mutual adjustment plus adjusting for area type and nearby street infrastructure (adjusted model 2), injury was independently predicted by primary road type (double the injury odds of residential roads, for instance); greater road width (10% rise in injury odds per 1m increase); and a lower gradient value (4% reduction in injury odds for every 1% increase in gradient). There was no longer evidence in adjusted analysis of an independent effect of speed limit or motor connectivity.

3. Effects of street infrastructure variables

Five of six variables were significantly associated with odds of injury in univariate analyses, the exception being a nearby bus stop. After mutual adjustment, plus adjusting for area type and road type (adjusted model 2), injury was independently predicted by the presence of a bicycle lane (54% increase in injury odds) or a bicycle track plus a lane (246% increase in injury odds; with wide confidence intervals as relatively few cases). These associations with bicycle infrastructure type changed little after adjusting for travel behaviour. Increased odds of injury was independently predicted by presence of a bus lane (84% increase in injury odds), guardrail (18% increase), petrol station or car park (43% increase). Again, none of these associations changed much after adjusting for the travel behaviour variables.

As mentioned above and as found in other studies, intersection locations substantially raised the odds of injury, with an odds ratio of 3.43 found in our unmatched models.

4. Effects of travel behaviour variables

Higher average speed was associated with a lower odds of injury in both univariate and adjusted analyses (22% reduction for a 10mph increase). Parked cars were associated with lower odds of injury in univariable analyses, but this effect disappeared in adjusted models, in particular after adjusting for road type (parked cars are more common on residential streets). This is the only variable in both sets of models (matched for intersection status vs. not matched) where the unmatched models gave different results: in that case, the odds ratio was 1.35, a strongly statistically significant result ($p<0.0001$).

In univariable analyses there was no association between odds of injury and volume of cycling, but after adjustment for other factors a higher volume of cyclists was associated with lower odds of injury (6% lower for every increase of 100 cyclists).

5. Examination of differential effects between slight injuries versus KSI

We conducted stratified analyses comparing the 2749 individuals with a slight injury to the 592 individuals who were killed or seriously injured (KSI) (see Appendix 3: Additional results). In general, the point estimates were similar between the two injury types, although less often statistically significant for KSI because of the much smaller sample size. There was never evidence of an interaction between any of the 17 predictor variables and KSI status (all $p\geq0.09$).

6. Examination of differential effects between non-intersection versus intersection points

We conducted stratified analyses comparing the 684 individuals who were not injured at an intersection to the 2657 individuals who were injured at an intersection. The adjusted results are shown in Table 2. As the number of non-intersection injuries was relatively small, the confidence intervals are fairly wide and there is low power for testing for interactions, and in 15 of the 17 variables tested, there was little or no evidence of an interaction in adjusted analyses (all $p \geq 0.09$). There was, however, strong evidence of an interaction between intersection status and road class ($p<0.001$) such that the protective effect of being on a secondary road and in particular tertiary road was stronger at intersection than at non-intersection points. There was strong evidence of an interaction between intersection status and average speed ($p=0.004$), such that the protective effect of higher average speeds (= lower congestion) was more pronounced at intersections than non-intersections.

In Table 2, Adjusted model 2 for intersection points includes some variables applying to the second road. After adjusting for the characteristics of the first road, there was a significantly increased odds of injury if the second road was a minor (i.e. not primary) road, and a significantly increased odds of injury if the second road was wider. Further exploratory analyses indicated an interaction between the road class of the first road and the second road ($p=0.003$), such that injury odds were specifically increased if first road were a primary road and the second road a minor road specifically (adjusted OR 2.41, 95% CI 1.88 to 3.08, compared to first road and second road both primary: see table in Appendix 2)

There was no effect of having a traffic signal present at an intersection, with or without an ASL. There was, however, substantially higher odds of injury if the intersection involved a roundabout or a mini roundabout, with similar effects of these two sorts of roundabouts (298% and 355% increase in injury odds respectively, compared to other junction locations). Finally, there was evidence that the odds of injury increased as average speed on the second road increased.

Table 2: Predictors of injury, all points

Category	Predictor	Level	N points	% injury points	Univariable	Adjusted 1	Adjusted 2	Adjusted 3
Area Type	**Urban**	**Rural**	464	47%	1*	1	1	1
		Urban	6218	50%	1.41 (1.01, 1.96)	1.31 (0.94, 1.83)	1.15 (0.80, 1.66)	1.19 (0.82, 1.74)
	High Street	**No**	5953	49%	1***	1***	1***	1**
		Yes	729	61%	1.85 (1.55, 2.20)	1.58 (1.32, 1.89)	1.48 (1.22, 1.80)	1.32 (1.08, 1.62)
	Average deprivation	**Change per standard deviation increase**	-	-	1.03 (0.96, 1.11)	1.04 (0.97, 1.12)	1.02 (0.94, 1.10)	1.01 (0.93, 1.09)
Road type	**Road class**	**Primary**	2561	58%	1***		1***	1***
		Secondary	745	49%	0.54 (0.44, 0.66)		0.67 (0.53, 0.84)	0.68 (0.54, 0.86)
		Tertiary	1215	45%	0.43 (0.36, 0.51)		0.55 (0.45, 0.67)	0.55 (0.45, 0.67)
		Residential or other	2160	44%	0.44 (0.38, 0.51)		0.60 (0.49, 0.74)	0.50 (0.40, 0.63)
	Road width	**Change per 1m increase**	-	-	1.16 (1.14, 1.19)***		1.11 (1.08, 1.14)***	1.10 (1.07, 1.13)***
	Gradient	**Change per 1% increase in incline (downhill = negative)**	-	-	0.97 (0.94, 0.99)*		0.96 (0.94, 0.99)**	0.96 (0.93, 0.98)**
	Speed limit	**20mph or less**	1244	47%	1**		1	1
		30mph	4633	51%	1.34 (1.12, 1.61)		0.95 (0.77, 1.18)	0.95 (0.77, 1.18)
		40mph	**395**	52%	1.51 (1.13, 2.03)		0.90 (0.64, 1.26)	1.10 (0.77, 1.57)
		over 40mph	347	50%	1.31 (0.95, 1.82)		0.91 (0.62, 1.32)	1.10 (0.74, 1.62)
	Connectivity rank	**0-24%**	310	42%	1***		1	1
		25-49%	622	43%	1.06 (0.80, 1.40)		1.04 (0.77, 1.40)	1.09 (0.80, 1.47)
		50-74%	1246	47%	1.31 (1.01, 1.70)		1.17 (0.89, 1.55)	1.33 (1.00, 1.76)
		75-100%	4217	53%	1.72 (1.34, 2.20)		0.96 (0.72, 1.28)	1.17 (0.87, 1.58)
Nearby street infra-structure	**Bicycle infrastructure**	**None**	5203	48%	1***		1***	1***
		Track (no lane)	571	53%	1.29 (1.07, 1.56)		1.19 (0.97, 1.46)	1.18 (0.96, 1.45)
		Lane (no track)	626	60%	1.86 (1.53, 2.26)		1.48 (1.20, 1.84)	1.54 (1.24, 1.91)
		Track and Lane	84	69%	2.79 (1.70, 4.56)		2.46 (1.45, 4.16)	2.46 (1.44, 4.22)
		Other, e.g. sign	142	50%	1.13 (0.80, 1.59)		1.23 (0.85, 1.78)	1.39 (0.95, 2.03)
	Guardrail	**No**	5598	49%	1***		1**	1*
		Yes	1028	58%	1.54 (1.33, 1.78)		1.25 (1.07, 1.46)	1.18 (1.01, 1.39)
	Bus lane	**No**	6267	49%	1***		1***	1***
		Yes	359	68%	2.51 (1.95, 3.23)		1.81 (1.37, 2.39)	1.84 (1.39, 2.44)
	Bus stop	**No**	6016	50%	1		1**	1**
		Yes	666	47%	0.89 (0.76, 1.05)		0.75 (0.63, 0.90)	0.77 (0.64, 0.92)
	Metro/rail/ tram stop	**No**	6642	50%	1*		1	1
		Yes	40	70%	2.60 (1.25, 5.39)		1.72 (0.79, 3.76)	1.52 (0.68, 3.36)
	Petrol station or car park	**No**	6259	49%	1***		1**	1**
		Yes	423	58%	1.47 (1.19, 1.81)		1.48 (1.18, 1.85)	1.43 (1.14, 1.79)

Travel behaviour	**2-way average morning peak speed**	**Change per 10mph increase**	-	-	0.81 (0.77, 0.86)***			0.78 (0.73, 0.84)***
	Parked cars	**No** **Yes**	2649 3977	52% 49%	1** 0.86 (0.78, 0.96)			1 1.00 (0.88, 1.14)
	No. cycle commuters on segment	**Change per 100 cyclists increase**	-	-	**0.99 (0.95, 1.03)**			**0.94 (0.90, 0.99)***

†p<0.1, *p<0.05, **p<0.01, ***p<0.001 in tests for heterogeneity. Numbers in the N' column add to less than 6682 points for some variables due to missing data. In all other columns all 6682 points are used, using multiple imputation. All adjusted models additionally adjust for workplace density, as linear and quadratic terms, and when examining number of commuters on the segment we additionally included a dummy variable '0-5 cycle commuters versus 6+'. Control point selected after matching for intersection status : see Appendix 3 for equivalent analyses using control point selected without regard for intersection status .

Discussion

High street status was associated with an elevated injury risk in final adjusted models, while urban area status was not, an initial effect becoming attenuated when adjusting for other variables. In adjusted models, injury risk was independently predicted by road type being primary, and by a more downhill gradient. Lower speed limits and lower motor traffic connectivity were initially associated with lower injury risk, but these effects were no longer statistically significant when adjusting for other variables. Increased road width was associated with increased injury risk in all models.

Our findings confirmed that main roads and wider roads (likely to have more traffic lanes) are riskier for people cycling. Adjusting for these factors meant that the impact of speed limits became statistically insignificant. This suggests that in practice road characteristics such as width and/or number of lanes may be more important in injury risk than formal measures to reduce speeds alone. Our modelling of actual motor traffic speeds in the morning peak suggested that congestion increases injury risk, with roads with very low motor traffic speeds seeing higher risks; although at intersections, congested second roads conversely decreased risk. The finding for guard railing suggests that this (anti)pedestrian infrastructure may help to create a perception among drivers that they will not encounter conflict with non-motorised users.

When separating out intersection and non-intersection points, we found that type of intersection mattered: both roundabouts and mini-roundabouts raised injury odds threefold at intersection locations. Signals, with or without on-road infrastructure of Advanced Stop Lines ('bike boxes') were not associated with increase or decrease in injury risk. At intersections, the negative impact of main roads and of congestion (low morning traffic speeds) were heightened.

The negative impact of environments with conflicting motor traffic movements appears clear in most cases; away from intersections this is likely to particularly relate to kerbside activity, with route environment characteristics such as high street activities, car parks, and bus lanes all raising injury risks. Restrictions on car parking and hence better visibility for people cycling might then account for the somewhat protective effect of bus stops (without a bus lane, which has a larger negative impact). As in other studies, we found a safety in numbers impact from other cyclists being present on the road segment; there did not appear to be a negative impact from conflicting movements in relation to other cyclists.

Our findings in relation to cycle infrastructure are contrary to other literature, which generally finds a protective impact from high quality separated tracks (less so for cycle lanes); including work by Thomas Adams (2020) using a similar method in London alone. Assuming our algorithm has not introduced bias (e.g. cyclists are in reality more likely to use roads with cycle infrastructure than predicted by the Cyclestreets direct routing), we believe the explanation likely lies in the quality of the cycle infrastructure typically existing across Britain in 2017. Figure 1 illustrates two images of typical cycle lanes (top) and cycle tracks (bottom): both were characteristically narrow, frequently disappeared suddenly and/or gave way to side roads and entrances, particularly tracks. Usually, any protection disappeared at junctions. None of this is inherent to the design of cycle infrastructure.

Figure 1: typical examples of cycle lanes and tracks in our dataset

England's new Cycle Infrastructure Design Guidance (LTN 1/20) suggests that infrastructure quality may start to improve. In London, where a similar update to guidance was published six years ago, studies such as Thomas Adams' already show a reduction in risk from what we believe is on average higher quality cycle infrastructure.

This high-quality infrastructure is most needed in contexts with higher existing risks. If roundabouts are to remain, higher-quality designs are needed, drawing on research from contexts such as the Netherlands where roundabouts are safer for cyclists than in the UK. Main roads, high streets, and roads with bus lanes or high congestion are all risky for cyclists, yet often serve key desire lines and destinations. Such routes should be prioritised for higher quality cycling infrastructure, ensuring high quality design at intersections. As cyclists are also at high risk on main roads when passing side road junctions, these designs should not just focus on protecting cyclists at primary-primary junctions, but also reducing risk at side roads (for instance, reducing the number and speeds of turning movements into and out of side roads). Making quieter streets more attractive and pleasant for cycling, for instance through low traffic neighbourhood-type schemes restricting through motor traffic, can also help to provide safe alternative cycle routes.

Table 2: Results separating intersection and non-intersection sites, and additional results for intersection points

Category	Predictor	Level	Non-intersection points (N=1,366 points)			Intersection points (N=5,312 points)				P-value for interaction with intersection status, Adjusted 1 models
			N points	% injury points	Adjusted	N points	% injury points	Adjusted 1	Adjusted 2	
Area type	Urban	Rural Urban	177 1191	48% 50%	1 1.91 (0.91, 3.99)	287 5027	47% 50%	1 1.04 (0.66, 1.64)	1 1.11 (0.68, 1.80)	p=0.57
	High Street	No Yes	1284 84	49% 68%	1* 1.79 (1.01, 3.19)	4669 645	49% 60%	1* 1.28 (1.03, 1.59)	1** 1.44 (1.15, 1.80)	p=0.43
	Average deprivation	Change per standard deviation	-	-	0.89 (0.74, 1.07)	-	-	1.03 (0.95, 1.13)	1.04 (0.95, 1.14)	p=0.18
Road type, first road	Road class	Primary Secondary Tertiary Residential or other	434 185 271 477	56% 51% 53% 42%	1* 0.85 (0.51, 1.40) 1.15 (0.73, 1.81) 0.52 (0.31, 0.88)	2127 560 944 1683	58% 48% 42% 45%	1*** 0.63 (0.49, 0.82) 0.44 (0.35, 0.56) 0.47 (0.37, 0.60)	1*** 0.44 (0.33, 0.58) 0.34 (0.27, 0.44) 0.40 (0.31, 0.53)	p<0.001
	Road width	Change per 1m increase	-	-	1.04 (0.95, 1.12)	-	-	1.11 (1.08, 1.15)***	1.07 (1.04, 1.11)***	p=0.19
	Gradient	Change per 1% increase in incline	-	-	0.97 (0.91, 1.03)	-	-	0.95 (0.92, 0.98)**	0.95 (0.91, 0.98)**	p=0.75
	Speed limit	20mph or less 30mph 40mph over 40mph	218 885 97 147	48% 50% 54% 52%	1 0.83 (0.47, 1.49) 0.87 (0.38, 2.02) 1.04 (0.45, 2.43)	1026 3748 298 200	47% 51% 52% 48%	1 0.97 (0.77, 1.23) 1.10 (0.74, 1.65) 1.05 (0.66, 1.66)	1 0.96 (0.73, 1.27) 1.04 (0.63, 1.69) 1.14 (0.64, 2.02)	p=0.83
	Connectivity rank	0-24% 25-49% 50-74% 75-100%	65 126 260 801	40% 42% 46% 54%	1 1.03 (0.52, 2.03) 1.26 (0.66, 2.42) 1.65 (0.82, 3.32)	245 496 986 3416	42% 43% 48% 52%	1* 1.08 (0.77, 1.53) 1.36 (0.98, 1.88) 1.07 (0.76, 1.50)	1 1.07 (0.75, 1.52) 1.28 (0.92, 1.78) 0.98 (0.69, 1.39)	p=0.16
Road type, second road	Road class	Primary Not primary	- -	- -		885 4429	53% 49%		1*** 2.04 (1.63, 2.54)	-
	Road width	Change per 1m increase	-	-		-	-		1.08 (1.05, 1.12)***	-
	Speed limit	20mph or less 30mph 40mph over 40mph	- - - -	- - - -		1100 3193 192 158	49% 50% 53% 48%		1 1.00 (0.77, 1.29) 0.92 (0.55, 1.56) 0.75 (0.42, 1.33)	-
Nearby street	Bicycle infrastructure	None Track (no lane)	1144 103	50% 44%	1 0.81 (0.49, 1.35)	4059 468	48% 55%	1*** 1.31 (1.04, 1.65)	1*** 1.31 (1.03, 1.67)	p=0.09

infra-structure		**Lane (no track)**	74	62%	1.68 (0.92, 3.05)	552	59%	1.52 (1.20, 1.92)	1.60 (1.25, 2.05)	
		Track and Lane	9	78%	11.84 (0.88, 159.8)	75	68%	2.23 (1.28, 3.90)	2.34 (1.31, 4.18)	
		Other, e.g. sign	13	31%	0.47 (0.12, 1.87)	129	52%	1.50 (1.00, 2.24)	1.36 (0.90, 2.05)	
	Guardrail	**No**	1201	49%	1	4397	48%	1	1	p=0.80
		Yes	142	59%	1.31 (0.85, 2.00)	886	58%	1.18 (0.99, 1.41)	1.14 (0.94, 1.37)	
	Bus lane	**No**	1288	49%	1	4979	49%	1***	1***	p=0.47
		Yes	55	64%	1.84 (0.88, 3.84)	304	68%	1.87 (1.37, 2.54)	1.89 (1.38, 2.58)	
	Bus stop	**No**	1212	51%	1**	4804	50%	1	1	p=0.24
		Yes	156	44%	0.57 (0.39, 0.84)	510	49%	0.82 (0.66, 1.00)	0.90 (0.72, 1.12)	
	Metro/rail/ tram stop	**No**	1361	50%	[omitted]	5281	50%	1	1	p=0.99†
		Yes	7	100%		33	64%	1.20 (0.52, 2.76)	1.67 (0.70, 3.98)	
	Petrol station or car park	**No**	1314	49%	1	4945	49%	1*	1*	p=0.54
		Yes	54	63%	1.73 (0.92, 3.22)	369	57%	1.38 (1.08, 1.77)	1.34 (1.03, 1.74)	
	Traffic signal	**No**	-	-		4833	49%		1	-
		Yes, no ASL	-	-		303	61%		1.14 (0.84, 1.55)	
		Yes, with ASL	-	-		178	59%		1.26 (0.87, 1.83)	
	Roundabout	**None**	-	-		4559	47%		1***	-
		Roundabout	-	-		557	69%		2.98 (2.25, 3.95)	
		Mini-roundabout	-	-		198	69%		3.55 (2.39, 5.27)	
Travel behaviour, first road	**2-way average morning peak speed**	**Change per 10mph increase**	-	-	0.93 (0.79, 1.10)	-	-	0.76 (0.70, 0.82)***	0.78 (0.72, 0.85)***	p=0.004
	Parked cars	**No**	584	51%	1	2065	52%	1	1	p=0.87
		Yes	759	49%	1.01 (0.76, 1.34)	3218	49%	0.99 (0.86, 1.14)	1.06 (0.91, 1.23)	
	No. cycle commuters on segment	**Change per 100 cyclists increase**	-	-	0.95 (0.83, 1.08)	-	-	0.94 (0.90, 0.99)*	0.94 (0.89, 0.99)*	p=0.86
Travel behaviour, second road	**2-way average morning peak speed**	**Change per 10mph increase**	-	-		-	-		1.17 (1.08, 1.27)***	-

†p<0.1, *p<0.05, **p<0.01, ***p<0.001 in tests for heterogeneity. ASL=advanced stop lane. Numbers in the N' column add to less than 1368/5314 points for some variables due to missing data. In all other columns all points are used, using multiple imputation. All models additionally adjust for workplace density, as linear and quadratic terms, and a dummy variable '0-5 cycle commuters versus 6+'. †From interaction test in univariable analysis, as multivariable model could not converge. 4 points, from 2 injuries, are excluded because it was not possible to sample a control point matched for intersection status (e.g. as the injury occurred at the first intersection after the participant's house).

Appendix 1: Selection of routing method

Our approach to modelling routes was informed by analysis of 230 tracks from early-morning commuters provided to us by Beeline, a new company which produces a navigation app like a compass for cyclists. 107 from those points were for the London area and 123 for the rest of the country. We excluded training, leisure rides and incorrect data (straight lines): in London, 7 tracks were recognized as training, being routes that consisted of circles around parks rather than routes from one place to another, 40 as leisure trips and 14 were straight lines. Outside of the London area, there were 36 leisure rides, 16 training rides and 19 straight lines. This left 98 valid tracks (46 for London and 52 out of London) that were used to decide which algorithm we should use for routing.

From these routes, we used the start and the end points to create new routes through two alternative methods in order to find which corresponded better to observed routes. These alternatives were ArcGIS and Cyclestreets API.

1. To model our routes into ArcGIS, we used the tool Network Analyst which is based on Dijkstra's algorithm. Dijkstra is an algorithm for finding the shortest paths between nodes in a graph, or in our case road network. It picks the unvisited vertex with the lowest distance, calculates the distance through it to each unvisited neighbour, and updates the neighbour's distance if smaller. We set up some new parameters on GIS regarding the Dijkstra algorithm, so that road types such as motorway and footway were excluded from the road network that the algorithm could use.
2. Cyclestreets.net is a journey planner for cyclists, which uses a related algorithm but incorporating other variables (for instance, likely cycling speed on different types of segment). Cyclestreets offers different options (Fast, Balanced and Quiet) which trade off directness against route comfort. Initial investigations (and evidence from Meade and Stewart 2018[3]) suggested that only the Fast option was likely to plausibly represent commuter cycling behaviour, with other options creating relatively long detours due to the paucity of cycling infrastructure in much of Britain. Thus our sensitivity testing compared only the Fast routes generated to the Dijkstra algorithm.

The random point tool in ArcGIS was used to create 20 random points for each route per source (20 points*98 tracks*3 sources of tracks, i.e. Beeline, Cyclestreets and Dijkstra), resulting in 5880 points in total. Every point corresponded to a road segment of each track. The approach aimed to draw out information from the road network segment for each one of these 5880 points, by using the spatial join tool in ArcGIS, since the aim was not to model exactly where people went, but rather represent well the types of the routes they chose.

Comparing the subsequent road types across the three route types (actual, Dijkstra, and Cyclestreets), we found that Cyclestreets provided the closer comparison to the actual routes followed. For instance, 27% of the actual route points were located on residential or unclassified streets, compared to 30% for the Cyclestreets algorithm but only 20% for the Dijkstra algorithm. This informed our decision to use CycleStreets for our algorithmic routing.

[3] https://www.sciencedirect.com/science/article/pii/S235214651830303X

Appendix 2: Route environment data sources

Table 3: List of Route Environment Data Sources

Sequence number	Variables and contributing factors	Value	Type of variable	Operationalisation of variables	Dataset name, owner, and date	Data location
			A. Area type			
1	Urban	1 Rural 2 Urban	Polygon	We matched the Rural Urban Classification with the boundaries for England, Wales and Scotland using the Lower Layer Super Output Areas code. Then, we identified where the injury and control points are located within the boundaries of LSOA.	Rural Urban Classification, GOV.UK, Department for Environment, Food & Rural Affairs, January 2020 Urban Rural Classification, Scottish Government, Geographic Information Science & Analysis Team, January 2020	https://data.gov.uk/dataset/b1165cea-2655-4cf7-bf22-dfbd3cdeb242/rural-urban-classification-2011-of-lower-layer-super-output-areas-in-england-and-wales https://statistics.gov.scot/data/urban-rural-classification
2	High Street	0 Not on or close to a high street 1 On or close to a high street	Point	We used the POIs catalogue but only some of the categories. These were Retail, Eating and drinking, Education and health, Sport and entertainment, Attractions, Commercial services. Once we selected the classification, we matched them with the corresponding data from the whole POIs dataset and we created polygon the clustering based on the point data using ArcGIS. Then we selected the road network from OSM within the polygon cluster. At the final step, we selected all the injury and control points that are located 25 m near of the selected road network.	Points of Interest, Ordnance Survey, November 2018 dataset used	https://www.ordnancesurvey.co.uk/documents/product-support/support/points-of-interest-classification-scheme.pdf https://digimap.edina.ac.uk/
3	Average deprivation	Change per standard deviation increase	Polygon	We located injury and control points inside each zone and looked up the deprivation levels per household for Lower Super Output Areas and Data Zones.	Classification of household deprivation (Great Britain) 2011 - Lower Super Output Areas and Data Zones, UK Data service, dataset used December 2019	https://www.statistics.digitalresources.jisc.ac.uk/
4	Workplace density		Polygon	The workplace population with the boundaries has been matched. Then we located injury and control points inside each zone and looked up the workplace density.	Classification of Workplace Zones, Consumer Data Research Centre, dataset used Janury 2020	https://data.cdrc.ac.uk/

		B. **Road Type**				
5	Road class(hierarchy)	Primary Secondary Tertiary Residential or other 6	Line	We mapped injury and control points to the nearest OSM road segment. As vector datasets represent roads as lines, and injuries are more frequent on major than minor roads, matching off-network points by distance tends to disproportionately allocate the points to minor roads, at intersection locations (Aldred et al, 2018). We similarly found that when comparing our initial distance-based matching of injury points to route segments, only 31.6% were matched to major roads, compared to an allocation of 43.3% by the police for the same set of points. While police data is not always completely accurate, this disparity suggests that at or close to intersections, our matching was biased towards minor roads. Hence, we carried out the following process. Points lying within 10m of an intersection (267 locations) were reclassified to a major road, where they had initially been assigned to a minor road. In total, this then gave 1472 injury points located at a major road, a number that represents 43.7% of injury points, close to the proportion recorded by the police (43.3%). Primary Secondary — Motorway, A road	Great Britain (England, Scotland and Wales) datasets, Open Street Map, dataset used March 2019	https://www.geofabrik.de/data/download.html
6	Road width	Change per 1m increase	Line	We used the OS Mastermap road network. Then the nearest roads on a range (buffer zone) of 20 m. of the injury and control points were selected. We used the average road width classification from the dataset.	Highways Network Road, Ordnance Survey, November 2019 dataset used	https://www.basemap.co.uk/
7	Gradient	Change per 1% increase in incline	API	The elevations and the distances from the Cyclestreets API have been used. We used road segments up to 250 meters before the injury and control points with the same slope in order to calculate the gradient.	Cyclestreers API, Cyclestreets, journey planner system, APi used March 2020	https://www.cyclestreets.net/api/
8	Speed limit	1 20 mph or less 2 30 mph 3 40 mph 4 over 40 mph	Line	We selected the nearest road on a range (buffer zone) of 20 m. of the injury and control points.	Basemap (the creator of the dataset) directly provided speed limit data from 2017 to us; speed limit data is also now	Basemap (the creator of the dataset) directly provided speed limit data from 2017 to us; speed limit data is

					available via Ordnance Survey Public Sector Mapping Agreement: https://www.ordnancesurvey.co.uk/business-government/products/mastermap-highways-speed-data	also now available via Ordnance Survey Public Sector Mapping Agreement https://www.ordnancesurvey.co.uk/business-government/products/mastermap-highways-speed-data
9	Connectivity rank	0-24% 25-49% 50-74% 75-100%	Line	The SpaceSyntax dataset has been used. It is a linear dataset and we used the 10km Choice Rank classification. The nearest road segment on a range of 20 meters of the injury and control points has been used.	Space Syntax OpenMapping, Spacesyntax, dataset used January 2020	https://spacesyntax.com/openmapping/
		C. Nearby street infrastructure				
10	Bike infrastructure	0 No bicycle infrastructure 1 Track (no lane) 2 Lane (no track) 3 Track and Lane 4 Other, e.g. sign	GSV	Lookups to see whether any bicycle infrastructure was present at any of the four streetview images that were downloaded for each point (where available). Then coding of the bicycle infrastructure type, separating lanes (on-road, paint-based) from tracks (off-road, separated from motor vehicles in some way).	Google Street View images, Google API, API used November 2019-March 2020	https://rrwen.github.io/google_streetview/ https://developers.google.com/maps/documentation/streetview/intro
11	Bus lane	0 Not bus lane 1 Yes, bus lane	GSV	GSV lookups to see whether a bus lane was visible in any of the four lookup images.	Google Street View images, Google API, API used November 2019-March 2020	https://rrwen.github.io/google_streetview/ https://developers.google.com/maps/documentation/streetview/intro
12	Bus stops	0 No, bus stops in a range of 20m 1 Yes, bus stops in a range of 20m	Point	We used data from NAPTAN. We created the point based on the coordinates and then used a 20 m range (buffer zone) from injury and control points in order to select all the relative points (Bus stops)	National Public Transport Access Nodes, Department for Transport, dataset used December 2019	https://data.gov.uk/dataset/ff93ffc1-6656-47d8-9155-85ea0b8f2251/national-public-transport-access-nodes-naptan
13	Metro/rail/tram stops	0 No, bus stops in a range of 20m 1 Yes, bus stops in a range of 20m	Point	We used data from NAPTAN. We created the point based on the coordinates and then used a 20 m range (buffer zone) from injury and control points in order to select all the relative points (Metro/rail/tram stops)	National Public Transport Access Nodes, Department for Transport, dataset used December 2019	https://data.gov.uk/dataset/ff93ffc1-6656-47d8-9155-85ea0b8f2251/national-public-transport-access-nodes-naptan

14	Petrol station or car park	0 Without Petrol station or car park on a range of 20 m 1 Within Petrol station or car park on a range of 20 m	Point, polygon	Data from OSM was used. Then we selected all the points that are related to the petrol station or car park in a range (buffer zone) of 20 m from injury and control points.	Great Britain (England, Scotland and Wales) datasets, Open Street Map, dataset used January 2020	https://www.geofabrik.de/data/download.html
15	Intersection	0 Without an intersection in 20 m range 1 Within an intersection in 20 m range	Line, point	The OSM road network was used. We identify the intersections using ArcGIS. Then we used a range (buffer zone) of 20 m of the injury and control points that lay near to an intersection.	Great Britain (England, Scotland and Wales) datasets, Open Street Map, dataset used March 2019	https://www.geofabrik.de/data/download.html
			D. Travel behaviour			
18	2-way average morning peak speed	Change per 10mph increase	Line	We used the average speed based on 2017 from basemap. We matched the speed data with the Master map network based on the TOID number. Then the nearest road on a range (buffer zone) of 20 m. of the injury and control points were selected	TOIDs (based on 2017) which have the average speed for the morning peak, Basemap (the creator of the dataset) directly provided the average speed data for the morning peak from 2017 to us.	https://www.basemap.co.uk/
19	Parked cars	0 Not on or close to cars parked 1 On or close to cars parked	GSV	GSV lookups to see whether parked cars were visible in any of the four lookup images.		https://www.geofabrik.de/data/download.html https://rrwen.github.io/google_streetview/ https://developers.google.com/maps/documentation/streetview/intro
20	Cycle commuters on segment	Change per 100 cyclists increase	Line	We used the PCT tool which uses Census origin-destination data to allocate commuter cyclists across the route network within England and Wales. The nearest road segment on a range (buffer zone) of 20 meters of the injury and control points has been used. As the PCT does not cover Scotland, we used the stplanr package in R (developed for the PCT) to create cycling volume using data from Census 2011	Cycle commuters, Propensity to Cycle Tool (PCT), dataset used December 2019 Census Scotland 2011, National Records of Scotland, dataset used January 2020	https://www.pct.bike/ https://github.com/ropensci/stplanr https://www.scotlandscensus.gov.uk/

Appendix 3: Additional results

Table 4: Results stratified by KSI (killed and seriously injured) status

Category	Predictor	Level	Adjusted, slight only (N= 5498 points)	Adjusted, KSI only (N= 1184 points)	P for Interaction with KSI status
Area type	Urban	Rural Urban	1 1.14 (0.72, 1.81)	1 1.26 (0.64, 2.50)	p=0.95
	High Street	No Yes	1* 1.28 (1.03, 1.59)	1 1.51 (0.90, 2.55)	p=0.66
	Average deprivation	Change per standard deviation increase	0.99 (0.91, 1.08)	1.10 (0.91, 1.33)	p=0.42
Road type	Road class	Primary Secondary Tertiary Residential or other	1*** 0.67 (0.52, 0.87) 0.51 (0.40, 0.63) 0.47 (0.36, 0.60)	1 0.70 (0.42, 1.18) 0.74 (0.45, 1.20) 0.67 (0.39, 1.13)	p=0.66
	Road width	Change per 1m increase	1.10 (1.06, 1.13)***	1.12 (1.04, 1.20)**	p=0.39
	Gradient	Change per 1% increase in incline	0.96 (0.93, 0.99)*	0.94 (0.88, 1.00)	p=0.41
	Speed limit	20mph or less 30mph 40mph over 40mph	1 0.86 (0.67, 1.09) 1.04 (0.70, 1.55) 0.87 (0.55, 1.38)	1 1.56 (0.91, 2.69) 1.36 (0.61, 3.07) 2.21 (0.94, 5.19)	p=0.09
	Connectivity rank	0-24% 25-49% 50-74% 75-100%	1 1.05 (0.75, 1.46) 1.25 (0.91, 1.71) 1.08 (0.77, 1.51)	1 1.32 (0.64, 2.73) 1.85 (0.92, 3.69) 1.87 (0.91, 3.86)	p=0.42
Nearby street infra-structure	Bicycle infrastructure	None Track (no lane) Lane (no track) Track and Lane Other, e.g. sign	1*** 1.18 (0.94, 1.49) 1.62 (1.28, 2.06) 2.16 (1.20, 3.88) 1.36 (0.88, 2.08)	1 1.10 (0.67, 1.81) 1.19 (0.69, 2.05) 4.90 (1.13, 21.27) 1.74 (0.74, 4.09)	p=0.73
	Guardrail	No Yes	1 1.13 (0.95, 1.35)	1 1.43 (0.94, 2.18)	p=0.21
	Bus lane	No Yes	1*** 1.74 (1.28, 2.37)	1** 2.79 (1.30, 6.00)	p=0.29
	Bus stop	No Yes	1* 0.80 (0.66, 0.97)	1* 0.63 (0.40, 1.00)	p=0.51
	Metro/rail/ tram stop	No Yes	1 1.86 (0.73, 4.73)	1 0.80 (0.15, 4.33)	p=0.45
	Petrol station or car park	No Yes	1** 1.40 (1.09, 1.80)	1 1.69 (0.97, 2.93)	p=0.77
Travel behaviour	2-way average morning peak speed	Change per 10mph increase	0.77 (0.71, 0.83)***	0.83 (0.70, 0.98)*	p=0.11
	Parked cars	No Yes	1 1.04 (0.90, 1.19)	1 0.82 (0.60, 1.13)	p=0.13
	No. cycle commuters on segment	Change per 100 cyclists increase	0.95 (0.90, 0.99)*	0.97 (0.85, 1.10)	p=0.74

†p<0.1, *p<0.05, **p<0.01, ***p<0.001 in tests for heterogeneity. All models additionally adjust for workplace density, as linear and quadratic terms, and a dummy variable ‘0-5 cycle commuters versus 6+’.

Table 5: Predictors of injury, all points – with controls selected not matching for intersection status

Category	Predictor	Level	N points	% injury points	Univariable	Adjusted 1	Adjusted 2	Adjusted 3
Area Type	**Urban**	**Rural**	490	45%	1***	1***	1	1
		Urban	6,192	50%	2.02 (1.43, 2.85)	1.88 (1.33, 2.66)	1.37 (0.90, 2.08)	1.40 (0.90, 2.17)
	High Street	**No**	6,014	48%	1***	1***	1***	1**
		Yes	668	67%	2.52 (2.08, 3.06)	2.15 (1.77, 2.62)	1.75 (1.40, 2.20)	1.48 (1.17, 1.86)
	Average deprivation	**Change per standard deviation increase**	-	-	1.09 (1.01, 1.16)*	1.09 (1.01, 1.17)*	1.04 (0.95, 1.13)	1.03 (0.95, 1.13)
Road type	**Road class**	**Primary**	2,511	59%	1***		1***	1***
		Secondary	744	49%	0.51 (0.41, 0.63)		0.68 (0.53, 0.87)	0.70 (0.54, 0.90)
		Tertiary	1,210	45%	0.41 (0.34, 0.49)		0.59 (0.47, 0.73)	0.59 (0.47, 0.73)
		Residential or other	2,216	43%	0.40 (0.34, 0.46)		0.64 (0.51, 0.81)	0.50 (0.39, 0.65)
	Road width	**Change per 1m increase**	-	-	1.23 (1.20, 1.26)***		1.12 (1.09, 1.16)***	1.11 (1.08, 1.15)***
	Gradient	**Change per 1% increase in incline (downhill = negative)**	-	-	0.96 (0.94, 0.99)**		0.97 (0.94, 1.00)*	0.96 (0.93, 1.00)*
	Speed limit	**20mph or less**	1,257	47%	1***		1	1
		30mph	4,582	51%	1.39 (1.16, 1.67)		0.92 (0.73, 1.17)	0.93 (0.73, 1.19)
		40mph	424	49%	1.21 (0.92, 1.61)		0.87 (0.60, 1.26)	1.10 (0.74, 1.62)
		over 40mph	382	45%	0.94 (0.68, 1.30)		0.99 (0.65, 1.50)	1.38 (0.88, 2.16)
	Connectivity rank	**0-24%**	327	40%	1***		1	1
		25-49%	620	43%	1.14 (0.85, 1.51)		1.09 (0.78, 1.52)	1.20 (0.85, 1.68)
		50-74%	1,281	46%	1.33 (1.03, 1.72)		1.19 (0.88, 1.62)	1.43 (1.04, 1.96)
		75-100%	4,170	53%	1.93 (1.51, 2.46)		1.03 (0.74, 1.42)	1.39 (1.00, 1.94)
Nearby street infra-structure	**Bicycle infrastructure**	**None**	5,209	48%	1***		1***	1***
		Track (no lane)	571	53%	1.32 (1.09, 1.59)		1.13 (0.90, 1.41)	1.17 (0.93, 1.48)
		Lane (no track)	627	60%	1.84 (1.51, 2.23)		1.34 (1.06, 1.69)	1.39 (1.10, 1.76)
		Track and Lane	66	88%	9.55 (4.34, 21.0)		5.99 (2.55, 14.0)	6.35 (2.61, 15.4)
		Other, e.g. sign	131	54%	1.34 (0.94, 1.91)		1.14 (0.75, 1.73)	1.22 (0.80, 1.88)
	Guardrail	**No**	5,704	47%	1***		1***	1***
		Yes	900	66%	2.33 (1.99, 2.73)		1.57 (1.31, 1.89)	1.48 (1.23, 1.79)
	Bus lane	**No**	6,250	49%	1***		1**	1**
		Yes	354	69%	2.56 (1.99, 3.29)		1.58 (1.17, 2.14)	1.60 (1.18, 2.17)
	Bus stop	**No**	5,987	51%	1*		1*	1*
		Yes	695	45%	0.81 (0.69, 0.95)		0.82 (0.68, 0.99)	0.82 (0.68, 1.00)
	Metro/rail/ tram stop	**No**	6,640	50%	1*		1	1
		Yes	42	67%	2.00 (1.05, 3.80)		1.30 (0.60, 2.84)	1.17 (0.53, 2.57)
	Petrol station or car park	**No**	6,289	49%	1***		1**	1**
		Yes	393	63%	1.82 (1.46, 2.27)		1.50 (1.16, 1.95)	1.47 (1.13, 1.92)

	Intersection	**No** **Yes**	2,400 4,282	29% 62%	1*** 4.42 (3.90, 5.00)		1*** 3.59 (3.14, 4.10)	1*** 3.43 (2.99, 3.93)
Travel behaviour	**2-way average morning peak speed**	**Change per 10mph increase**	-	-	0.71 (0.67, 0.75)***			0.76 (0.70, 0.83)***
	Parked cars	**No** **Yes**	2,832 3,772	48% 51%	1* 1.15 (1.03, 1.28)			1*** 1.35 (1.17, 1.55)
	No. cycle commuters on segment	**Change per 100 cyclists increase**	-	-	1.00 (0.96, 1.04)			0.95 (0.90, 0.99)*

†p<0.1, *p<0.05, **p<0.01, ***p<0.001 in tests for heterogeneity. Numbers in the N' column add to less than 6682 points for some variables due to missing data. In all other columns all 6682 points are used, using multiple imputation. All adjusted models additionally adjust for workplace density, as linear and quadratic terms, and when examining number of commuters on the segment we additionally included a dummy variable '0-5 cycle commuters versus 6+'

Table 6: Predictors of injury, among points at intersections (N=5314), according to the combination of the road class of the first road and the second road

Predictor	**Level**	**N points**	**% injury points**	**Adjusted for area, road type, nearby infrastructure and travel behaviour**
Road class combination	**Primary 1st * Primary 2nd**	729	54%	1***
	Primary 1st * Minor 2nd	1,398	60%	2.41 (1.88, 3.08)
	Minor 1st * Primary 2nd	156	48%	0.67 (0.44, 1.02)
	Minor 1st * Minor 2nd	3,031	44%	0.83 (0.63, 1.09)

Adjusted model includes the same variables as 'Adjusted 3' in Table 3 of the main text

Appendix 4: Literature review from first academic article

Much analysis of the infrastructural causes or correlates of cycling injuries uses an outcome variable as being either injury numbers or injury severity (e.g. Chen et al, 2015). However, predicting crash frequency without including a measure of bicycle volume or distance travelled means that it is not possible to separate the risk that a (type of) location poses to each individual cyclist from the number of cyclists using that (type of) location. Hence, while analysis can identify characteristics of sites with high numbers of injuries or where injuries are relatively severe, it may fail to identify route characteristics that keep cyclists safer but simultaneously attract more cyclists (or conversely, route characteristics that raise injury risk but simultaneously deter cycling).

One reason for this methodological limitation is a lack of cycling flow data, on which exposure calculations could be based (Dozza, 2107); another is the traditionally poor spatial data on characteristics of the street environment that might be associated with injury risk. Because of these limitations, analysis that does seek to control for exposure has frequently focused on only a small number of sites (e.g. Lusk et al 2011) as this facilitates collection of count and infrastructure data that may not be available across a wider network. There are relatively few global analyses taking in a range of sites and a range of infrastructure types, with some (e.g. (Vandenbulcke et al., 2009) using area-level data, which is limited by an inability to link risk directly to route segment characteristics.

The following sub-sections review evidence related to route environment factors that may affect cycling injury risk. As per above, much of the evidence relates to injury numbers or injury severity; however, some work does cover risk in relation to exposure and we focus on these results in the summary below.

Intersections and Road Hierarchy

A disproportionate amount of cycling injuries take place at or near intersections (DfT 2017), where conflicting movements occur. In a study incorporating GPS-derived measures of cyclist flows, Strauss et al (2015) found that signalized intersections, which are often located at the intersection of major arterials, witness 4 times more injuries and 2.5 times greater risk than non-signalized intersections. Similarly, Strauss et al (2015) found arterial roads have higher risk than do minor roads, a finding replicated by other studies (e.g. Williams et al 2017, Aldred et al 2018).

Motor Traffic Volumes and Speeds

Higher motor traffic volumes have been found to be associated with higher injury risk, with Aldred et al (2018) finding that this has an impact independent of road class (arterial roads would generally be expected to carry more motor traffic thasn residential roads). Speed is established as a risk factor for injury severity (Chen et al 2010).

Speed limits

Some papers have examined the impact of speed limits on cycling injury risk, as opposed to actual speeds. In London, Aldred et al (2018) found a reduction in cycling injury odds of 21% for 20mph compared to 30mph streets. Kaplan et al (2014) found similar results for Denmark.

Topography

Teschke et al (2012) found downhill route gradients were associated with elevated injury risk, while Vandenbulcke et al's (2009) area-level study found that hilly areas had higher risks.

Land Use

Studies in the USA (Cho et al 2009), China (Ma et al 2010), and New Zealand (Williams et all 2018) have found relationships between land use and cyclist injuries. Some have highlighted mixed land use and/or high street locations as a risk; while Chen and Shen (2019) found that mixed land use areas have less severe injuries.

Guard railing

In Britain many cities and towns have installed 'guard railings' between footways and roads, particularly at 'desire lines' without controlled crossings, and close to junctions and crossings. A 2017 report by Transport for London found that removing guard railing reduced collisions for pedestrians and all users, however.

Cycle infrastructure

Studies that control for cycling volume (often higher on cycle lanes and tracks) generally find that cycle infrastructure plays a protective role, although with some conflicting findings regarding infrastructure type, and differences by context. Strauss et al's (2015) Canadian study found that while there were more cyclist injuries where there were cycle tracks, this was due to higher cycling volumes, and hence the risk per cyclist was lower than on streets without cycle tracks. Again in Canada, Teschke et al (2012) found a nine-fold reduction in cycling injury odds (albeit with large confidence intervals) for cycle tracks compared to major roads with parked cars, although they did not find a similar reduction for painted cycle lanes; nor for off-road routes. However, Williams et al (2017) found cycle lanes (on-road, painted) in New Zealand were associated with reduced injury risk.

While Li et al's (2017) London study found no difference between cyclist injury risk on London Cycle Superhighways and other roads, their results showed 'that it is much safer to cycle on CS3' [Cycle Superhighway 3, which was then the only Superhighway in the study largely consisting of separate cycle tracks] than on roads with painted or no cycle infrastructure. Adams and Aldred (2020) found similar results, with separated cycle infrastructure in London associated with a 40-65% reduction in injury odds, whereas painted lanes increased risk. By contrast, Jensen et al (2007) found that introduction of cycle tracks in Copenhagen during 1978-2003 was associated with a 10% increase in cycling injury risk.

Summary

The discussion above highlights some key findings and areas of debate in the literature. As mentioned above, more evidence is still needed, especially covering a global network as opposed to (for instance) several intersection sites. Methods to control for exposure are needed to separate out the impacts of increased risk and increased usage. Where exposure data exists case-control methods can be used (e.g. Aldred et al 2018, Williams et al 2017, Vandenbulcke et al 2014); however, at a national level this is rarely available.

Adams, T. and Aldred, R. (2020). Cycling Injury Risk in London: impacts of road characteristics and infrastructure. Transport Findings.

Aldred, R., Goodman, A., Gulliver, J., & Woodcock, J. (2018). Cycling injury risk in London: A case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits. Accident Analysis and Prevention, 117(November 2017), 75–84. https://doi.org/10.1016/j.aap.2018.03.003

Aldred, R., Goel, R., Woodcock, J. and Goodman, A. (2019) Contextualising Safety in Numbers: a longitudinal investigation into change in cycling safety in Britain, 1991–2001 and 2001–2011, Injury Prevention, 25 (3), pp. 236-241

ESRI (2020). Random Points Tool. 25th November 2020. https://pro.arcgis.com/en/pro-app/tool-reference/data-management/create-random-points.htm

ESRI (2020a). Network Analyst extension. 23rd November 2020. https://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/algorithms-used-by-network-analyst.htm

Beeline. Beeline. 11th June 2019. https://beeline.co/.

Buehler, R. and Pucher, J. (2017) Trends in Walking and Cycling Safety: Recent Evidence From High-Income Countries, With a Focus on the United States and Germany, *American Journal of Public Health* 107, no. 2 (February 1, 2017): pp. 281-287.

Chen, P. (2015). Built environment factors in explaining the automobile-involved bicycle crash frequencies: A spatial statistic approach. Safety Science, 79, 336–343. https://doi.org/10.1016/j.ssci.2015.06.016

Chen, P., & Shen, Q. (2019). Identifying high-risk built environments for severe bicycling injuries. Journal of Safety Research, 68, 1–7. https://doi.org/10.1016/j.jsr.2018.11.002

Chen, Y., Yang, J., & Otte, D. (2010). Load and Impact Conditions for Head Injuries in Car-to-Pedestrian and Car-to-Cyclist Accidents – A Comparison of Real Accidents and Simulations.

Cho, G., Rodríguez, D. A., & Khattak, A. J. (2009). The role of the built environment in explaining relationships between perceived and actual pedestrian and bicyclist safety. Accident Analysis and Prevention, 41(4), 692–702. https://doi.org/10.1016/j.aap.2009.03.008

CycleStreets.net. CycleStreets.net. n.d. https://www.cyclestreets.net/.

Department for Transport (2017) Pedal Cycle Factsheet. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/686969/pedal-cycle-factsheet-2017.pdf

Department for Transport. data.gov.uk. 20th January 2019. https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data.

Dozza, M. (2017). Crash risk: How cycling flow can help explain crash data. Accident Analysis and Prevention, 105, 21–29. https://doi.org/10.1016/j.aap.2016.04.033

Li, H., Graham, D.J. and Liu, P. (2017) Safety effects of the London cycle superhighways on cycle collisions, Accident Analysis & Prevention, 99, A., 90-101.

Jeffrey S., Stone, D.H., Blamey, A., et al (2009) An evaluation of police reporting of road casualties, Injury Prevention 2009;15:13-18.

Jensen, S.U. (2007) Bicycle Tracks and Lanes: a Before-After Study, available at https://www.researchgate.net/publication/237524182_Bicycle_Tracks_and_Lanes_a_Before-After_Study/link/5a548377458515e7b732688e/download

Kaplan, S., Vavatsoulas, K., & Prato, C. G. (2014). Aggravating and mitigating factors associated with cyclist injury severity in Denmark. Journal of Safety Research, 50, 75–82. https://doi.org/10.1016/j.jsr.2014.03.012

Lusk, A. C., Furth, P. G., Morency, P., Miranda-Moreno, L. F., Willett, W. C., & Dennerlein, J. T. (2011). Risk of injury for bicycling on cycle tracks versus in the street. Injury Prevention, 17(2), 131–135. https://doi.org/10.1136/ip.2010.028696

Ma, M., Yan, X., Abdel-Aty, M., Huang, H., & Wang, X. (2010). Safety Analysis of Urban Arterials Under Mixed-Traffic Patterns in Beijing. Transportation Research Record: Journal of the Transportation Research Board, 2193, 105–115. https://doi.org/10.3141/2193-13

Meade, S., & Stewart, K. (2018). The Visualisation of SiN in Edinburgh. Scottish Transport Application Research, 2020, 1–15.

Office for National Statistics. Office for National Statistics. 6th November 2019. https://geoportal.statistics.gov.uk/datasets/national-statistics-postcode-lookup-latest-centroids.

Strauss, J., Miranda-Moreno, L. F., & Morency, P. (2015). Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts. Accident Analysis and Prevention, 83, 132–142. https://doi.org/10.1016/j.aap.2015.07.014

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. https://doi.org/10.2105/AJPH.2012.300762

Transport for London (2017) Collisions Before and After the Removal of Pedestrian Railings at 70 Junctions and Crossings on the Transport for London Road Network. Report no. SB257. London: TfL.

Vandenbulcke, G., Thomas, I., de Geus, B., Degraeuwe, B., Torfs, R., Meeusen, R., & Int Panis, L. (2009). Mapping bicycle use and the risk of accidents for commuters who cycle to work in Belgium. Transport Policy, 16(2), 77–87. https://doi.org/10.1016/j.tranpol.2009.03.004

Williams, T., Page, S., & Doscher, C. (2018). Spatial characteristics of bicycle–motor vehicle crashes in Christchurch, New Zealand: A case-control approach. Journal of Transport and Land Use, 11(1), 849–864. https://doi.org/10.5198/jtlu.2018.1147