TRL THE FUTURE OF TRANSPORT

Modelling the impacts of Bikeability training on KSIs: Final project deliverable

N Harpham, I Marino, L Smith, G Beard

September 2024

Contents

- Executive summary
- Introduction and overview of this project
- Data review, collection and processing
- Model development
  - Exploratory analysis
  - Modelling
- Explanation of results
- Interpretation and Conclusions
- Limitations and Recommendations
- Appendix A – Variables Considered for Modelling
- Appendix B – Additional information
- Appendix C – Model details and performance
- Appendix D - References

Executive Summary

TRL developed a series of **collision prediction models** for the Bikeability Trust with the aim of answering the question:

Can statistical models be used to determine associations between levels of Bikeability training and road traffic KSIs in English local authorities?

Models were developed at **local authority level** and used data from a **10-year period** from 2013 to 2022.

L2 and L3 Bikeability training were included as potential explanatory variables. Four models were developed:

1. L2 (explanatory variable), KSIs (dependent variable)
2. L3 (explanatory variable), KSIs (dependent variable)
3. L2 (explanatory variable), cyclist KSIs (dependent variable)
4. L3 (explanatory variable), cyclist KSIs (dependent variable)

	KSIs	Cyclists KSIs
L2	✓	✓
L3	X	X

✓ = significant association
X = no significant association

- The models indicated a significant association between L2 training and KSIs, whereby increased levels of Bikeability training delivery is associated with fewer KSIs, and fewer cyclist KSIs, at the local authority level.
- Because of the low L3 delivery numbers, statistical power for the analyses was low, which meant an association did not emerge. This may be reflective of the generally lower levels of uptake of L3 training across the local authorities studied.
- These initial indications are informative - building on this with a larger dataset or more granular analysis would strengthen the conclusions that can be made on the impact of Bikeability, particularly L3 training.

Introduction and overview of this project

- The Bikeability Trust has a requirement for **data-led evidence** on the **impact of Bikeability training**, which will support their application for future funding.
- To generate a valuable source of evidence for this purpose, the Trust commissioned TRL to develop models assessing the **association** between **Bikeability training** and road traffic **KSIs***.
- For this purpose, TRL developed a series of **collision prediction models** – a standard approach for explaining the relationship between several different variables and collision or casualty risk.
- National models were developed considering factors at the **Local Authority** level.
- This report outlines the **model development** process and **key findings** associated with this work.

- The **deliverables** for this project were:
  - This PowerPoint report
  - The R code and associated outputs used to develop the models
  - The processed data used for the models

The key question
Can statistical models determine association between Bikeability training and casualties (specifically KSIs)?

*KSIs = Killed and seriously injured casualties, an important measure in road safety

Data review, collection and processing - overview

Data covered England (not including London) at a **Local Authority level**. Models therefore aimed to explain KSI risk by Local Authority.

- After processing and consolidation of some local authorities we had useable data for 109 Local Authorities, representing all parts of England excluding London.
- The period considered was 2013 – 2022 to align with the Bikeability training delivery data available (see slide 10).

Next slides give details of:

- **Data identification**
- **Processing** – how the datasets were processed
- **Outcome variables (Dependent Variables)** – the variables that we are interested in explaining or predicting
- **Predictors (Independent Variables)** – the variables that we use to observe their effect on the dependent (outcome) variable. These variables are used to explain variations in the dependent variables.
- **How we measured levels of Bikeability training delivery**

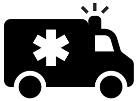
Data review, collection and processing – data identification

- Various data sets were identified by the Bikeability Trust and TRL that might help to explain variance in KSIs between local authorities
  - This was based on knowledge of variables that influence KSIs
  - Data was required at local authority level
- Data that was included (see slides 8 to 12) for more details:
  - Infrastructure variables: road length
  - Traffic variables: traffic, flow
  - Area type (rural or urban)
  - Demographics: population, IMD, % young males
  - Bikeability data
  - Cycle specific variables: cycleway length, ATE funding, % of population cycling to work and school
- Data that was identified but not included:
  - Driving and motorcycle tests – these did not match well with the local authority areas
  - Deaths due to alcohol and drugs – not specifically related to road collisions
  - Road conditions – some missing data
- Note that the occurrence of collisions and severity outcomes is complex and therefore no statistical model can include variables that fully explain the number of KSIs.
  - E.g. differences in road types, enforcement practices, employment, public transport use, etc are all factors.

Data review, collection and processing – processing

- All the data were linked together using the government local authority codes. These codes are used for most data published by the Office for National Statistics (ONS)
- Local Government reorganisation meant that some data were not available for all of the current local authorities. In particular:
  - Somerset and North Yorkshire became unitary authorities - the geographical extents were the same and therefore the old and new areas were matched
  - Bournemouth, Christchurch and Poole have merged to become one unitary authority
  - The following county councils have split into the corresponding unitary authorities – all data were combined to historic county council level
    - Bedfordshire: Bedford + Central Bedfordshire
    - Northamptonshire: North Northamptonshire + West Northamptonshire
    - Cheshire: Cheshire East + Cheshire West and Chester
    - Cumbria: Cumberland + Westmorland and Furness
  - Bikeability data for Liverpool included Sefton, Knowlsey, Wirral and St Helens and therefore all data for these authorities were combined.

Data review, collection and processing – outcome variables



The outcome variable of interest is the number of casualties in each local authority.

- Based on STATS19 database of reported injury collisions
- Killed and Seriously Injured (KSI) data were used:
  - The number of killed casualties is too small for robust analysis
  - The reporting of slight casualties may not be consistent over time and across different police forces
- Two outcome variables for modelling:
  - Number of **KSIs** in road traffic collisions
  - Number of **cyclist KSIs** in road traffic collisions
- Exploratory analysis also used the risk of KSIs i.e. the number of KSIs divided by the total traffic (veh-km)
  - **KSI rate**
  - **Cyclist KSI rate**

Data review, collection and processing – predictors (1)

To develop models for explaining casualty risk, data was obtained at **local authority level** for a set of key predictor variables. It is known for example, from previous research (Wallbank, Harpham and Fletcher, 2023 – see Reference in Appendix D), that traffic and deprivation are key factors associated with casualty risk. The predictor variables considered were categorised as follows:

- **Infrastructure variables** – Total road length (including and excluding motorways)
  - This is normally included in models as areas with more road length will intuitively have more collisions. This was included in collision prediction models developed for Transport Infrastructure Ireland (TII) (Wallbank et al, 2023)
- **Traffic variables** – ‘Flow’ (number of vehicles passing a fixed point on an average day) and ‘Traffic’ (a density measure combining flow and road length)
  - Flow is normally included in models as areas with higher flow will typically have more collisions. This was included in the previous TII models (Wallbank et al, 2023)
  - Traffic is the exposure which is used as a comparison of casualty risks
- **Area type –** whether the area is urban or rural
  - This can help to explain variation in casualty numbers. Rural areas generally have higher vehicle speeds, which can influence casualties. Urban areas have more opportunities for conflicts, especially with pedestrians and cyclists which can also influence casualties. This was included in the previous TII models (Wallbank et al, 2023)

See **Appendix A** at the end of this slide deck for a full list of variables considered, along with their definitions

Data review, collection and processing – predictors (2)

- **Demographic variables** – These help to explain differences in the populations of each local authority
  - **Index of Multiple Deprivation (IMD)** - See, for example, collision prediction models developed in Wallbank et al (2023), and the links between deprivation and casualty numbers reported here: [Reported road casualties Great Britain: Casualties and deprivation - GOV.UK (www.gov.uk)]()
  - **Population size** – authorities with higher population are likely to have a higher number of casualties. This is likely to be correlated with the road length and traffic
  - **Proportion of young males in the population** – research has shown that young males have a high collision risk. Therefore, authorities with a higher proportion of young males may have a higher collision rate

Bikeability variables – Amount of Level 2 and level 3 places delivered, as a proportion of the Local Authority population

  - This is the main variable that this project was interested in.
  - Two predictor variables were explored (see Slide 10)

Other cycling specific variables – these may help to understand variation in casualties between local authorities and could be used to compare any effects with Bikeability

  - **Amount of Active Travel England (ATE) funding** - this was included as it could be considered that local authorities with a high amount of ATE funding might have more schemes relating to improving cycling and walking. Note that this budget can cover many parts of a scheme; for example, consultation, planning, implementation.
  - **Length of cycleways** – this was calculated as a percentage of road length. More cycleways might encourage more cycling, and more cycleways might suggest a safer level of infrastructure. The length of cycleways may also relate to ATE funding.
  - **Proportion of people cycling to work and school** – this was taken from Census data. Local authorities with large numbers of people cycling to work or school might have a higher number of casualties (due to a larger exposure). Note that there may be a ‘safety in numbers’ effect whereby more cyclists yield a lower cycling casualty risk. The amount of cycling to work or school may also relate to the length of cycle ways in an area (for example if there are safe routes then people might be more likely to use them).

See **Appendix A** at the end of this slide deck for a full list of variables considered, along with their definitions

Data review, collection and processing – measuring Bikeability

Measuring 'Bikeability exposure' for each Local Authority

In order to capture the amount of 'Bikeability exposure' for each Local Authority area, data was used on the number of L2 and L3 training places offered. This data was available from 2013 to 2022, and hence data for the other variables were selected to align with this time period.

Two variables were created (for each of L2 and L3) to represent 'Bikeability exposure':

1. Total Bikeability places offered as a proportion of the population each year who could have taken part in Bikeability training

   E.g. The population aged 9-11 in Essex was 46,669 in 2011, increasing to 55,613 in 2022

2. Total Bikeability places offered (2013-22) as a proportion of the local authority population in 2023 of the appropriate age to have taken part in the training over the past 10 years

   E.g. level 2 training is normally for ages 9-11. Those aged 11 in 2013 would be aged 21 in 2023; so therefore the population aged 9 to 21 in 2023 were considered.

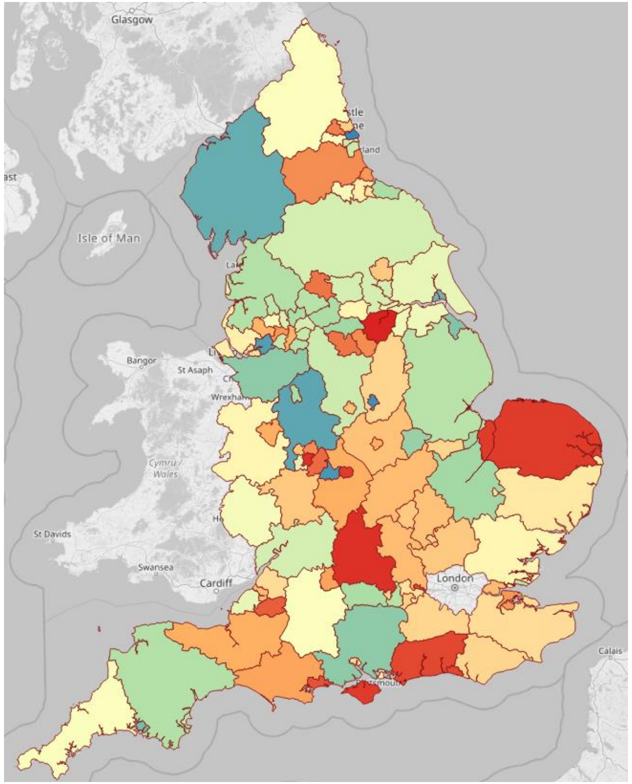
Models tested for each of these variables produced very similar results. The second of these metrics was used in the final modelling as it performed better in the univariate models. Appendix A outlines the calculations for these variables.

Data review, collection and processing – measuring Bikeability

L2 Bikeability divided by 2023 population aged 9-21 for local authorities used in analysis

- Values ranged from 12% in Doncaster to 85% in south Tyneside
- Map shows all local authorities used in study on coloured scale shown below

L2 Bikeability percentage	Example local authority
12 %	Doncaster
26%	Dorset
40%	Northumberland
52%	Devon
63%	Cumbria
85%	South Tyneside



© OpenStreetMap contributors

Model development - overview

The most appropriate modelling approach for this work was identified to be **Generalised Linear Models (GLMs)** of the Poisson and Negative Binomial form. These models are used to assess the relationship between predictor variables and an outcome variable represented as a count (such as casualty numbers).

Before developing the models, **exploratory analysis** aimed to understand the relationships between the variables. This is important to:

- Ensure that the **key predictors** of KSIs are selected in the models
- Ensure that **highly correlated variables** are **not used** in the same models (the presence of multicollinearity can lead to unreliable coefficient estimates)

The GLMs were developed by testing different combinations of variables to assess the best model fit. Various metrics were used to assess the **quality of the models** during the selection process.

The focus was to understand **whether Bikeability training delivery levels within local authorities are(statistically) significantly associated with KSIs**, and the extent of any relationship. We also wanted to compare Bikeability with other key predictors.

See slides 12 and 13 for detail on the exploratory analysis.

See slides 14-20 for detail on the modelling approach and key outputs.

Exploratory analysis

1 star = significant at 5% level,
2 stars = significant at 1% level,
3 stars = significant at 0.1% level

- A **correlation matrix** was produced to understand the correlations between the variables.
- The cells show the result of a correlation test between the relevant variables, and the stars represent the level of statistical significance in the relationship (3 star being strongest).
- The complete matrix is provided as an excel worksheet separate from this presentation; however, a section of the matrix for some of the key variables is shown here.

KSIs

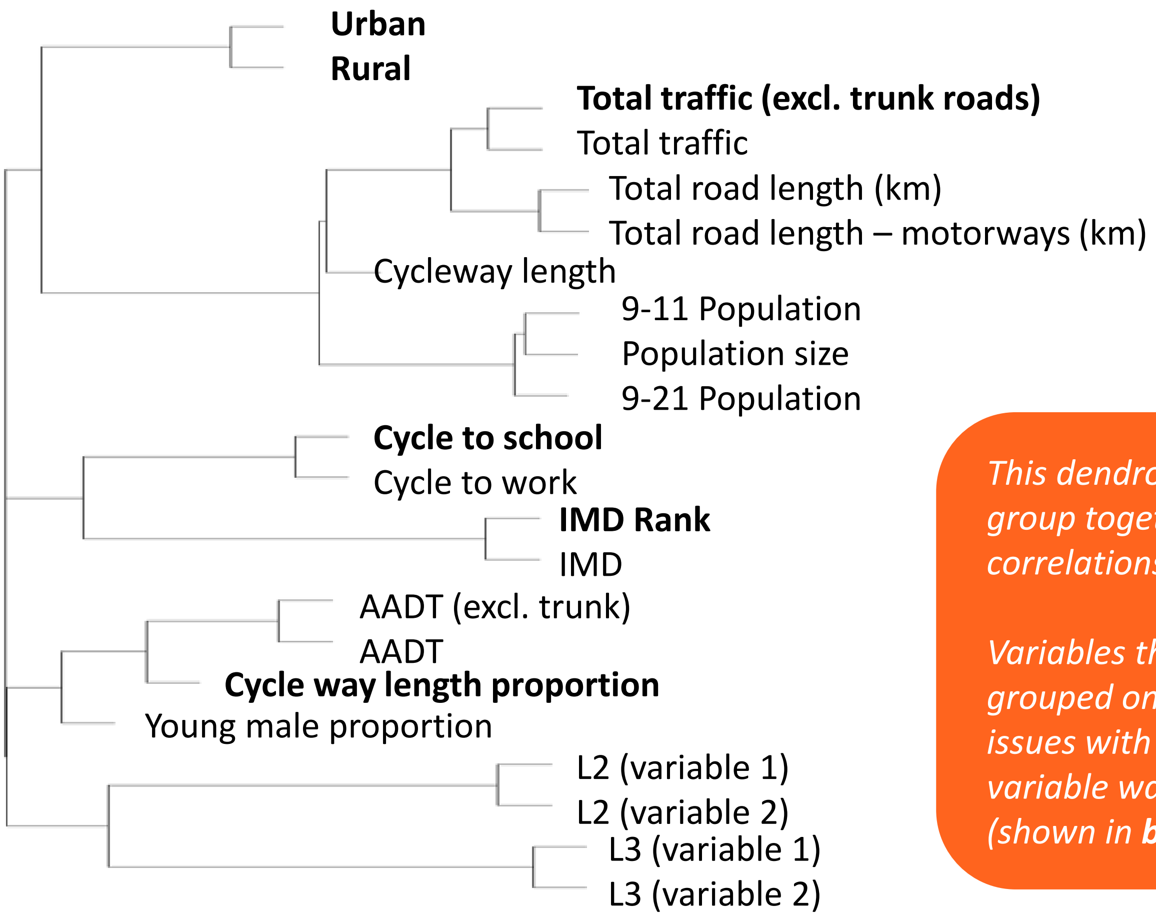
KSI rate

	Cyclists.KSI.2013.2022.adj	IMD_Extent_2019	KSI.2013.2022.adj	L2.bikeability...2012.2022.yr9.11.pop	L3.bikeability...2012.2022.yr11.14.pop	Population_Size	Proportion_Young_Males	Total.road.length..km.	X2011.school.census...cycle.to.school	X2013.2022.Total.traffic.excl.trunk..mveh.km.
Cyclists.KSI.2013.2022.adj	1.00	-0.20 *	0.93 ***	0.01	-0.09	0.87 ***	0.05	0.63 ***	0.20 *	0.82 ***
IMD_Extent_2019	-0.20 *	1.00	-0.24 *	0.02	0.05	-0.24 *	0.36 ***	-0.45 ***	-0.48 ***	-0.38 ***
Killed.2013.2022rate.per.bvkm	0.26 **	0.10	0.29 **	-0.05	-0.04	0.27 **	-0.01	0.23 *	-0.09	0.20 *
KSI.2013.2022.adj	0.93 ***	-0.24 *	1.00	-0.01	-0.12	0.95 ***	-0.09	0.79 ***	0.08	0.93 ***
KSI.2013.2022.adjrate.per.bvkm	0.28 **	0.37 ***	0.17	-0.04	0.05	0.12	0.24 *	-0.09	-0.12	-0.02
L2.bikeability...2012.2022.yr9.11.pop	0.01	0.02	-0.01	1.00	0.47 ***	-0.06	-0.15	0.04	0.00	-0.03
L3.bikeability...2012.2022.yr11.14.pop	-0.09	0.05	-0.12	0.47 ***	1.00	-0.09	-0.01	-0.07	-0.04	-0.11
Population_Size	0.87 ***	-0.24 *	0.95 ***	-0.06	-0.09	1.00	-0.07	0.79 ***	0.06	0.96 ***
Proportion_Young_Males	0.05	0.36 ***	-0.09	-0.15	-0.01	-0.07	1.00	-0.28 **	-0.07	-0.19
Total.road.length..km.	0.63 ***	-0.45 ***	0.79 ***	0.04	-0.07	0.79 ***	-0.28 **	1.00	0.13	0.89 ***
X2011.school.census...cycle.to.school	0.20 *	-0.48 ***	0.08	0.00	-0.04	0.06	-0.07	0.13	1.00	0.13
X2013.2022.Total.traffic.excl.trunk..mveh.km.	0.82 ***	-0.38 ***	0.93 ***	-0.03	-0.11	0.96 ***	-0.19	0.89 ***	0.13	1.00

Key variables significantly correlated with KSIs at an LA level were:
population size, road length, traffic and deprivation level.
Modelling is needed to understand the relationships in more detail. For example, when accounting for traffic, only IMD is still significant, and proportion of young males becomes a significant factor.

Exploratory analysis (2)

TRL



This dendrogram shows how variables group together based on their correlations.

*Variables that are highly correlated are grouped on the same branch - to avoid issues with multicollinearity, only one variable was selected within a group (shown in **bold**)*

Modelling

Why do we need models?

Statistical models aim to explain the relationship between variables and are particularly useful when there are several influencing factors.

In this case, models were used to understand any relationship between Bikeability training and KSIs, accounting for traffic, demographics and other key factors.

What models did we use and why?

Poisson and Negative Binomial models were developed. These are regression models used when the outcome of interest (in this case casualty numbers) takes integer values.

Poisson and negative binomial models are similar; however, negative binomial models are more flexible as they allow for greater variance in the data. Poisson models assume that the mean and variance of the data is the same.

Overdispersion occurs when the observed variance in the data is greater than what is expected under a given statistical model. In the context of count data models like the Poisson regression, overdispersion is a common issue. An ‘overdispersion test’ was used to see which model is more suitable. If the test indicates overdispersion, the negative binomial model is preferred as it can accommodate the extra variability in the data.

To select the variables for the models we tried 3 different approaches. The ‘Supervised’ models were chosen as they provided the best performance.

Stepwise is a method used in statistical modeling to select variables (from a pre-defined list) for a model. It is particularly useful when you have many potential variables and want to identify the most significant ones. [1]

1. Manual
The variables were chosen based on previous experience (e.g. picking variables that have been significant in similar work)

2. Stepwise
The variables were selected by the automated stepwise process, with no manual intervention.

3. ‘Supervised’
Stepwise was implemented on a starting model which had the traffic variable (known to be a significant predictor of risk) as minimum.

Supervised is a combination of Manual and Stepwise, combining statistical justification with theoretical experience

Diagnostics were assessed to choose the best performing and most theoretically sensible models, based on experience.

1. Stepwise should be used with caution, as it can sometimes lead to models that are overly sensitive to the specific data used for selection. To make sure this was not the case we checked the coefficients to make sure they were sensible from a theoretical point of view.

Variable selection for approach 3 (supervised approach)

The variables pre-selected based on theoretical expertise and the correlation analysis were:

Total traffic (excl. trunk)
Urban/Rural
Cycling to school %
IMD Rank
Cycle way length proportion
Bikeability variable (L2 or L3)

Included in all models

Tested for inclusion in each of the models according to the stepwise process

Model 1: KSIs (outcome variable), Bikeability L2 (predictor variable)

> Bikeability L2 was identified as a significant predictor

Variable	Coefficient	Signif.	Interpretation
(Intercept)	6.98E+00	***	
Total traffic (excl. trunk)	3.74E-05	***	The number of KSI (Killed or Seriously Injured) incidents is more likely to increase with higher traffic levels (indicating that local authorities with more traffic have more KSI incidents)
IMD Rank	-2.97E-03	***	The number of KSIs is less likely to increase with a higher IMD Rank (indicating that less deprived areas have fewer KSI incidents)
Bikeability L2	**-6.28E-01**	**	**The number of KSIs is less likely to increase with a higher number of Bikeability L2 training places provided (indicating that areas with higher levels of Bikeability L2 training have fewer KSIs)**
Cycle way length proportion	-9.05E-04		The number of KSI is less likely to increase with a higher proportion of cycle ways

Significance Codes for different p-values:

```
‘***’ 0.001
‘**’  0.01
‘*’   0.05
‘.’   0.1
‘ ’   1
```

Bikeability L2 was significant at the 1% level (p-value less than 0.01)

Cycle to school % and urban/rural were not identified as adding value to the models and so were not selected as variables. Cycle way length was selected by the stepwise process but the variable was not significant.

More information on this model (including performance diagnostics) is given in Appendix C

Model 2: KSIs (outcome variable), Bikeability L3 (predictor variable)

> Bikeability L3 contributes to the final model, but was not statistically significant

Variable	Coefficient	Signif.	Interpretation
(Intercept)	6.72E+00	***	
Total traffic (excl. trunk)	3.80E-05	***	The number of KSI (Killed or Seriously Injured) incidents is more likely to increase with higher traffic levels (indicating that local authorities with more traffic have more KSI incidents)
IMD Rank	-3.21E-03	***	The number of KSI is less likely to increase with a higher IMD Rank (indicating that less deprived areas have fewer KSI incidents)
Bikeability L3	**-1.89E+00**		**The variable improved the overall performance of the model, but was not a significant predictor on its own**

Significance Codes for different p-values:

```
'***'  0.001
'**'   0.01
'*'    0.05
'.'    0.1
' '    1
```

Bikeability L3 was selected but found to be **not** statistically significant at 10% level

More information on this model (including performance diagnostics) is given in Appendix C

Cycle to school %, cycle way length and urban/rural were not identified as adding value to the models and so were not selected as variables.

Model 3: Cyclist KSIs (outcome variable), Bikeability L2 (predictor variable)

> Bikeability L2 was identified as a significant predictor

Variable	Coefficient	Sign	Interpretation
(Intercept)	4.59E+00	***	
Total traffic (excl. trunk)	3.68E-05	***	The number of KSI (Killed or Seriously Injured) incidents is more likely to increase with higher traffic levels (indicating that local authorities with more traffic have more KSI incidents)
Urban	3.48E-01	*	Urban areas are positively associated with the increase of Cyclists KSIs compared to Rural areas
Rural	1.47E-01		
Cycle to school	9.88E+00	***	The number of KSI (Killed or Seriously Injured) incidents is more likely to increase with higher proportion of people cycling to school
IMD Rank	-3.42E-03	*	The number of KSI is less likely to increase with a higher IMD Rank (indicating that less deprived areas have fewer KSI incidents)
Bikeability L2	**-6.31E-01**	*	**The number of KSI is less likely to increase with a higher number of Bikeability L2 provided (indicating that areas with higher levels of Bikeability L2 training have fewer KSIs)**

Significance Codes for different p-values:

```
'***' 0.001
'**'  0.01
'*'   0.05
'.'   0.1
' '   1
```

Bikeability L2 was significant at the 5% level (p-value less than 0.05)

More information on this model (including performance diagnostics) is given in Appendix C

Model 4: Cyclist KSIs (outcome variable), Bikeability L3 (predictor variable).

> Bikeability L3 contributes to the final model, but was not statistically significant

Variable	Coefficient	Sign	Interpretation
(Intercept)	4.40E+00	***	
Total traffic (excl. trunk)	3.64E-05	***	The number of KSI (Killed or Seriously Injured) incidents is more likely to increase with higher traffic levels (indicating that local authorities with more traffic have more KSI incidents)
Urban	3.66E-01	**	Urban areas are positively associated with the increase of Cyclists KSIs
Rural	1.48E-01		
Cycle to school	9.74E+00	**	The higher the proportion of people cycling to school the higher are the registered cases
IMD Rank	-3.35E-03	*	The number of KSI is less likely to increase with a higher IMD Rank (indicating that less deprived areas have fewer KSI incidents)
Bikeability L3	**-2.66E+00**		**The variable improved the overall performance of the model, but was not a significant predictor on its own**

Significance Codes for different p-values:

```
'***' 0.001
'**'  0.01
'*'   0.05
'.'   0.1
' '   1
```

Bikeability L3 was selected by the model, however it was found to **not** be statistically significant at the 10% level

More information on this model (including performance diagnostics) is given in Appendix C

Model Performance

The models developed in this analysis had a **modest level of explanatory power**, as indicated by the resulting McFadden's R^2 values which fell between 0.099 and 0.116. Unlike the R^2 metric used in linear regression, McFadden's R^2 values are generally lower and values between 0.2 to 0.4 are considered indicative of a model with an excellent fit.

Given the McFadden's R^2 values observed in this work, it is clear that while the developed models do capture some of the variability in the outcome variables, there is still a significant portion of the variability that remains unexplained. This could be due to several factors:

- **Complexity of the Phenomenon**: The relationship between the predictors and the outcome may be inherently complex and influenced by factors not included in the model. For instance, road safety outcomes can be affected by a wide range of variables such as weather conditions, driver behavior, and enforcement of traffic laws, which may not be fully captured by the available data.

- **Data Quality and Aggregation**: The level of data aggregation and potential measurement errors can also impact model performance. Aggregated data can obscure individual-level variations and interactions that are crucial for understanding the true nature of the relationships between variables.

- **Unobserved Heterogeneity**: There may be unobserved factors that influence the outcome variables, leading to unexplained variability. These could include socio-economic factors, local policies, and cultural attitudes towards road safety.

While the McFadden's R^2 values indicate that our models have limited explanatory power, they still provide valuable insights into the factors associated with the outcome variables. The models can be used as a starting point for further investigation and refinement.

Explanation of results

	KSIs	Cyclists KSIs
L2	✓	✓
L3	X	X

For the Bikeability L2 models, the Bikeability variable was identified as being statistically significantly associated with both KSIs and cyclist KSIs.

For the Bikeability L3 models, the Bikeability variable was selected as part of the stepwise process, but was <u>not</u> found to be statistically significantly associated with either KSIs or cyclist KSIs.

How can we interpret the results of the models?

- **Bikeability L2**: Total KSIs and cyclist KSIs were lower in local authorities in which Bikeability training delivery levels were higher.
- **Bikeability L3**: The relationship between delivery of L3 training and KSIs is similar to that for L2 training. However, given the statistically non-significant nature of this relationship, further analysis is required (see Limitations & recommendation for suggestions of further research).
- The coefficients for the other selected variables align with theoretical expectation and past experience, and show a greater overall impact than Bikeability. They are also very consistent across all of the models:
    - Increased traffic is associated with an increase in KSIs
    - Increased deprivation is associated with an increase in KSIs
    - Urban areas see more KSIs than rural area

> This increases our confidence in the validity of the models and the conclusions we can draw.

Interpretation and Conclusions

The question this research aimed to address was: **Can statistical models be used to determine associations between levels of Bikeability training and road traffic KSIs in English local authorities?**

The key conclusions are as follows:

- The models indicated a significant association between L2 training and KSIs (and cyclist KSIs). Increased training is associated with fewer KSIs and fewer cyclist KSIs at the local authority level.
- Because of the low L3 delivery numbers, statistical power for the analyses was low, which meant an association did not emerge. However, the relationship between L3 delivery levels and KSIs was similar to that for L2 delivery – i.e., as delivery levels increase, KSIs reduce.
- Whilst the models indicate an association, we cannot infer causation for any of the selected variables. It is possible that Bikeability training has a direct impact on KSI rates; however, another hypothesis is that this association is reflective of a more robust approach to road safety being taken in certain local authority areas. For example, local authorities with more Bikeability training may also, by proxy, take road safety more seriously, and therefore may implement a greater number of other road safety interventions, such as road improvements, enforcement and education, which collectively lead to fewer incidents on the roads.
- L2 Bikeability is not as strongly significant as traffic and deprivation, commonly understood predictors of risk, which are significant at the 0.1% level; however, it is notable that L2 Bikeability is also contributing to the models significantly.
- These initial indications are informative; however, a larger dataset or more granular analysis (see next slide) would support with developing more robust models and therefore strengthening the conclusions that can be made.

Limitations & recommendations for further research

- The main limitation with these models is that the sample size was small with only just over 100 data points used, and data aggregated over a 10-year period. If the number of data points could be increased substantially this would increase the reliability of the conclusions. Models could then be created with a rigorous test/validation set up, which was not feasible here. Options for achieving this include:
  - Including more local authority areas (for example borough data for London could be added on a next iteration)
  - Splitting up the data into different time periods (this would potentially require more complex models that account for changes over time)
  - Collecting more granular data (for example at school level)
- The differences between L2 and L3 training are perhaps indicative of this sample size issue. If more data was available for L3 it is possible that an association might be realised, as for L2. The negative coefficients for L3 hint at a similar relationship.
- As mentioned on the previous slide, these models cannot be used to attribute causation – a general limitation of this type of analysis. Alternative research designs, such as a longitudinal study, that enables tracking of young riders over time, should be considered to support the findings from this analysis.
- Through this analysis the potential long-term impacts of Bikeability training were examined - in this case, on KSIs. It is recommended that these results are considered alongside additional data gathered as part of the Trust's wider M&E activities, which are typically focussed on identifying shorter term outcomes.

Appendix A1 – Variables Considered for Modelling (casualty variables)

VARIABLE IN DATA (R)	DEFINITION	SOURCE
Killed.2013.2022	Number of fatalities 2013-22	Sourced from Stats19 database of reported road injury collisions, summarised here: [Reported road collisions, vehicles and casualties tables for Great Britain - GOV.UK]
KSI.2013.2022.adj	Number of Killed and seriously injured casualties 2013-22	
All.2013.2022	Number of casualties 2013-22,	
Cyclists.Killed.2013.2022	Number of cyclist fatalities 2013-22	
Cyclists.KSI.2013.2022.adj	Cyclist KSIs 2013-22 (adjusted)	
Cyclists.All.2013.2022	Cyclists in collisions 2013-22	
Killed.2013.2022rate.per.bvkm	Fatality rate = Fatalities / traffic 2013-22	Calculated by dividing the number of casualties by traffic
KSI.2013.2022.adjrate.per.bvkm	KSI rate = KSIs (adjusted) / traffic 2013-22	
All.2013.2022.rate.per.bvkm	All casualty rate = All casualties / traffic 2013-22,	

Appendix A2 – Variables Considered for Modelling (infrastructure, traffic, area type)

VARIABLE IN DATA (R)	DEFINITION	SOURCE
Total.road.length..km.	Length of road in km	Road lengths
Total.road.length...motorways..km.	Length of road (excluding motorways) in km	
X2013.2022.Total.traffic..mveh.km.	Total traffic (veh-km) 2013 - 22	Road traffic estimates (TRA) - GOV.UK
X2013.2022.Total.traffic.excl.trunk..mveh.km.	Total traffic (veh-km) 2013 - 22 excluding Strategic road network (SRN)	
Average.AADT.calculated.from.traffic.and.length	Average flow (number of vehicles passing a point on an average day)	Calculated using $AADT = \frac{traffic}{road\ length \times 365}$
Average.AADT.exclu.trunk.calculated.from.traffic.and.length	Average flow (number of vehicles passing a point on an average day) not including SRN	
Broad.RUC11 (not in correlation matrix)	Urban rural classification (three options) from 2011 census data	Rural Urban Classification (2011) of Local Authority Districts in EN

Appendix A3 – Variables Considered for Modelling (demographics)

VARIABLE IN DATA (R)	DEFINITION	SOURCE
Population_Size	Population size as of 2021 census	Sex by single year of age - Office for National Statistics
Proportion_Young_Males	Males 15-25 years old divided by total population (census 2021)	
Population.mye2022.Aged.9.to.21	2022 mid year estimated population aged 9 to 21 (so population that might have been exposed to Bikeability between 2012 and 2023)	Population estimates for the UK, England, Wales, Scotland, and Northern Ireland - Office for National Statistics
X2012.2022.sum.9.11.year.population.mye	Sum of ages 9-11 in population each year between 2012 and 2022	Estimates of the population for England and Wales - Office for National Statistics
IMD_Extent_2019	Proportion of a local authority's district living in most deprived 30% LSOAs in the country - from 2019	English indices of deprivation 2019
IMD._Rank_of_average_rank	Ranking of LAs by index of multiple deprivation	

Appendix A4 – Variables Considered for Modelling (Bikeability)

VARIABLE IN DATA (R)	DEFINITION	SOURCE
L2.bikeability...2012.2022.yr9.11.pop	L2 Bikeability /(X2012.2022.sum.9.11.year.population.mye / 3)	Data from the Bikeability Trust divided by relevant population
L2...2023.ages9.21.pop	Number of level 2 places / 9-21 population in 2023	
L3.bikeability...2012.2022.yr9.11.pop	L3 Bikeability /(X2012.2022.sum.11.14.year.population.mye / 3)	
L3...2023.ages9.21.pop	Number of level 3 places /11-25 population in 2023	

Appendix A5 – Variables Considered for Modelling (other cycling data)

VARIABLE IN DATA (R)	DEFINITION	SOURCE
cycleway.length..km.	Length of cycle way	Calculated from open street map data (Geofabrik Download Server) roads labelled as 'cycleway' within each LA boundary (Counties and Unitary Authorities (December 2023) Boundaries UK BFC | Counties and Unitary Authorities (December 2023) Boundaries UK BFC | Open Geography Portal (statistics.gov.uk)) in QGIS using processing toolbox sum line lengths
cycle_way_length_prop_ofroadlength	Cycle way length / total road length	
X2011.census...of.16..who.full.time.work.to.cycle.to.work	Percentage of full time workers who cycle to work, 16+, 2011 census	Cycling to work (at local authority level) (CW090)
X2011.school.census...cycle.to.school	% of students cycling to school from 2011 school census	Propensity to Cycle Tool - Northamptonshire (pct.bike) (Click on National data – commuting layer (local authorities)
ATE.funding.2020.21.to.23.24	2020/21 to 2024 funding. Joint authorities split according to population. Funding may be for planning, consultation, design or implementation.	Active Travel Fund: final allocations
ATE_Funding_over_population	ATE funding / population size (2021)	

Appendix B – Additional information

Potential 'lag analysis'

- To supplement the analysis performed here, we suggest that a 'lag analysis' could be conducted to determine the time gap between Bikeability interventions and observed changes in KSI.
- This would involve shifting the Bikeability data by different time lags and analysing the correlations with KSIs. This approach would be more insightful with more granular data than we had available here, such as school data and catchment area perimeters.
- In addition, other relationships over time could also be considered, not just the effect of Bikeability on KSIs, as there could be indirect effects involved.

Appendix B – Additional information

Other research

There is little published research of a similar kind to this work, assessing the relationship between cycle training and road traffic casualties. However, as part of this work we found the following research pieces that discuss (to some extent) the relationship between cycle training and road safety:

- The impact of cycle proficiency training on cycle-related behaviours and accidents in adolescence: findings from ALSPAC, a UK longitudinal cohort: IJPDS (2017) Issue 1, Vol 1:118, Proceedings of the IPDLN Conference (August 2016) | International Journal of Population Data Science
- An examination of the relationship between cycle training, cycle accidents, attitudes and cycling behaviour among children: Ergonomics: Vol 45 , No 9 - Get Access (tandfonline.com)
- Prevention of bicycle-related injuries in children and youth: a systematic review of bicycle skills training interventions | Injury Prevention (bmj.com)
- A review of evaluations of bicycle safety education as a countermeasure for child cyclist injury (trb.org)

Appendix C – Model details and performance

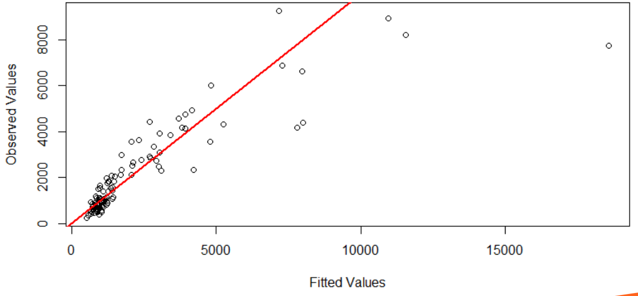
Model 1: KSIs (dependent variable), Bikeability L2 (explanatory variable)

```
Coefficients:
                                         Estimate Std. Error z value Pr(>|z|)
(Intercept)                             6.983e+00  1.251e-01  55.829  < 2e-16 ***
X2013.2022.Total.traffic.excl.trunk..mveh.km.  3.742e-05  1.741e-06  21.490  < 2e-16 ***
IMD._Rank_of_average_rank              -2.974e-03  7.777e-04  -3.824 0.000131 ***
L2...2023.ages9.21.pop                 -6.283e-01  2.162e-01  -2.906 0.003663 **
cycle_way_length_prop_ofroadlength     -9.048e-04  6.484e-04  -1.395 0.162881
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(9.7121) family taken to be 1)

    Null deviance: 736.41  on 108  degrees of freedom
Residual deviance: 110.88  on 104  degrees of freedom
AIC: 1663.8
```

Bikeability L2 found to be significant at 1%



McFadden R^2: 0.116

The McFadden's R-squared value was 0.12 suggesting this model isn't particularly strong. This can be attributed to both the level of aggregation over 10 years and the limited data points available.

*A McFadden R^2 between 0.2 - 0.4 is typically indicative of a model with an **excellent fit**, with a higher index denoting better overall model performance.*
Source: Hensher and Stopher, 1979 (see References in Appendix D)

Appendix C – Model details and performance

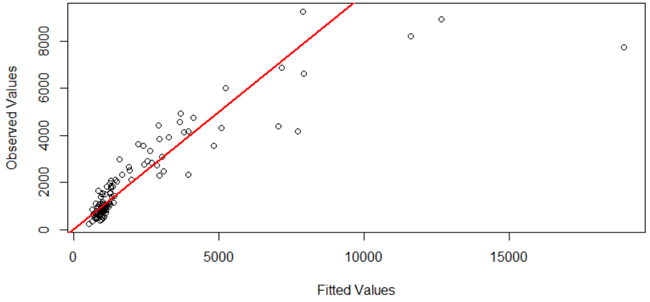
Model 2: KSIs (dependent variable), Bikeability L3 (explanatory variable)

```
Coefficients:
                                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                                  6.718e+00  7.168e-02  93.723  < 2e-16 ***
X2013.2022.Total.traffic.excl.trunk..mveh.km.  3.800e-05  1.699e-06  22.370  < 2e-16 ***
IMD._Rank_of_average_rank                   -3.210e-03  7.945e-04  -4.040 5.34e-05 ***
L3...2023.ages11.25.pop                     -1.888e+00  1.175e+00  -1.607    0.108
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(9.1627) family taken to be 1)

    Null deviance: 694.99  on 108  degrees of freedom
Residual deviance: 111.01  on 105  degrees of freedom
AIC: 1668.3
```

Bikeability L3 found to be not significant at 5%



The McFadden’s R-squared value was 0.11 suggesting this model isn’t particularly strong. This can be attributed to both the level of aggregation over 10 years and the limited data points available.

McFadden R^2: 0.112

*A McFadden R^2 between 0.2 - 0.4 is typically indicative of a model with an **excellent fit**, with a higher index denoting better overall model performance.*
Source: Hensher and Stopher, 1979 (see References in Appendix D)

Appendix C – Model details and performance

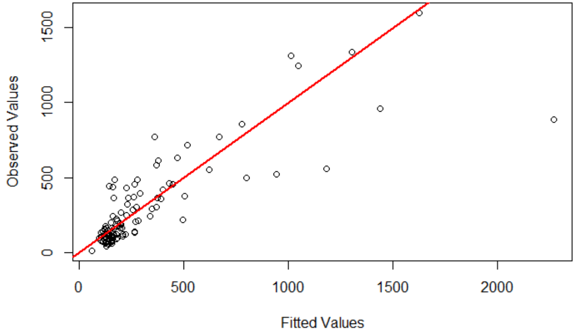
Model 3: Cyclist KSIs (dependent variable), Bikeability L2 (explanatory variable)

```
Coefficients:
                                           Estimate Std. Error z value Pr(>|z|)    
(Intercept)                               4.591e+00  2.065e-01  22.236  < 2e-16 ***
X2013.2022.Total.traffic.excl.trunk..mveh.km.  3.681e-05  2.526e-06  14.575  < 2e-16 ***
Broad.RUC11Predominantly Urban            3.476e-01  1.351e-01   2.572  0.01011 *  
Broad.RUC11Urban with Significant Rural   1.470e-01  1.524e-01   0.965  0.33478    
X2011.school.census...cycle.to.school     9.878e+00  2.928e+00   3.374  0.00074 ***
IMD._Rank_of_average_rank                -3.421e-03  1.418e-03  -2.412  0.01588 *  
L2...2023.ages9.21.pop                   -6.312e-01  3.031e-01  -2.082  0.03731 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(4.9563) family taken to be 1)

    Null deviance: 399.02  on 108  degrees of freedom
Residual deviance: 112.70  on 102  degrees of freedom
AIC: 1318
```

Bikeability L2 found to be significant at 5%



The McFadden’s R-squared value was 0.10 suggesting this model isn’t particularly strong. This can be attributed to both the level of aggregation over 10 years and the limited data points available.

McFadden R^2: 0.101

*A McFadden R^2 between 0.2 - 0.4 is typically indicative of a model with an **excellent fit**, with a higher index denoting better overall model performance.*
Source: Hensher and Stopher, 1979 (see References in Appendix D)

Appendix C – Model details and performance

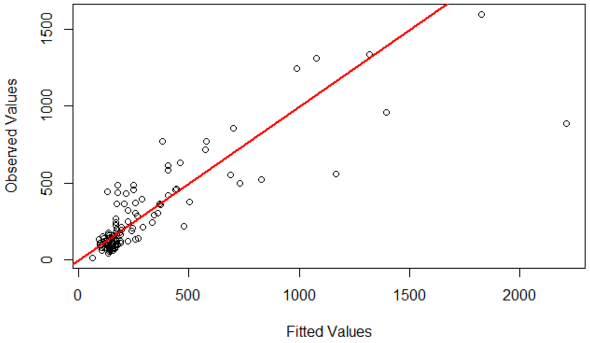
Model 4: Cyclist KSIs as dependent variable, testing Bikeability L3

```
Coefficients:
                                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                                 4.403e+00  1.787e-01  24.645  < 2e-16 ***
X2013.2022.Total.traffic.excl.trunk..mveh.km. 3.642e-05  2.555e-06  14.256  < 2e-16 ***
Broad.RUC11Predominantly Urban              3.658e-01  1.365e-01   2.679  0.00738 **
Broad.RUC11Urban with Significant Rural     1.475e-01  1.537e-01   0.960  0.33712
X2011.school.census...cycle.to.school       9.742e+00  2.962e+00   3.290  0.00100 **
IMD._Rank_of_average_rank                  -3.349e-03  1.430e-03  -2.342  0.01916 *
L3...2023.ages11.25.pop                    -2.657e+00  1.645e+00  -1.615  0.10634
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(4.8708) family taken to be 1)

    Null deviance: 392.29  on 108  degrees of freedom
Residual deviance: 112.83  on 102  degrees of freedom
AIC: 1320.1
```

Bikeability L3 found to be not significant at 5%



The McFadden’s R-squared value was 0.10 suggesting this model isn’t particularly strong. This can be attributed to both the level of aggregation over 10 years and the limited data points available.

McFadden R^2: 0.099

*A McFadden R^2 between 0.2 - 0.4 is typically indicative of a model with an **excellent fit**, with a higher index denoting better overall model performance.*
Source: Hensher and Stopher, 1979 (see References in Appendix D)

Appendix D - References

Hensher DA and Stopher PR (Eds.) (1979) Behavioural Travel Modelling. Taylor & Francis

Wallbank C, Harpham N and Fletcher J (2023) *TII268 Lot1 Collision Prediction Model for the Irish National Road Network – Phase 2 Report*, PPR2031. Crowthorne: Transport Research Laboratory

(Link: TRL Report (tii.ie))