Can Road Conditions and Rurality Predict Road Accident Severity?

Introduction and Aims

According to the World Health Organisation (WHO, 2023), road accidents cause approximately 1.19 million deaths per year, and up to 50 million non-fatal injuries. Whilst developing countries are disproportionately affected by road accidents, making up around 90% of global road fatalities, road safety remains a relevant issue in the UK with over 29,000 people killed or seriously injured annually (Department for Transport, 2023a). As such, improving road safety is an essential endeavour. Since 2012, numerous safety regulations have been introduced. For example, in 2012 the UK Government introduced plans to toughen drug-driving legislation (Department for Transport, 2013) and the European New Car Assessment Programme (Euro NCAP) made Electronic Stability Control (ESC) compulsory for all new cars (AA, 2019). Recent technological advances have also progressed Advanced Driver Assistance Systems (ADAS) such as lane support systems and emergency brake assistance, representing significant road safety improvements (European Commission, 2018). It is important to investigate how road accidents have changed since 2012 to assess the true impacts of these technologies.

Whilst many safety technologies have experienced developments in recent years, there are some road characteristics which cannot be easily controlled. According to the Met Office (2023), the UK has experienced an increase in adverse weather conditions in recent decades, with 2011 to 2020 being 9% wetter than the entire period from 1961 to 1990. Adverse weather has been shown to consistently impact the prevalence of road accidents (e.g., Bergel-Hayat, 2013; Malin, 2019), which is unsurprising as wet road surfaces can cause aquaplaning and loss of traction. In fact, a report by the Department for Transport (2015) estimated that, had rainfall been average in 2014, there would have been 43 fewer fatalities on UK roads. It was suggested that adverse weather was responsible for the 4% increase in road fatalities in 2014. This demonstrates the catastrophic impact that adverse conditions may have on the severity of road accidents. Another contributing factor towards road accidents is rurality. Rural roads are often narrow and contain fast, blind bends. There is also a distinct lack of safety infrastructure on rural roads such as barriers. These have been suggested as some of the reasons why rural roads are more dangerous than urban roads (Brake.org, 2019). Whilst there is clear evidence of the impact that adverse conditions and rurality can have on the prevalence of road accidents, and on fatality rates, there is a distinct lack of research around the extent to which these variables impact the severity of road accidents. Research in this area is vital to understanding how existing road infrastructure may be adapted and improved, and where to direct resources. The present project aims to gain further understanding of the issues outlined to inform road improvements and ensure the safety of drivers and pedestrians. Specifically, the project seeks to address three key research questions:

1. How has the prevalence of severe road accidents changed between 2012 and 2023?
2. How does the severity of road accidents vary based on road conditions and rurality?
3. To what extent can road conditions and rurality predict the severity of road accidents in the UK?

Methodology

This project utilised the GOV.UK Road Safety Data from:

https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data .

The specific datasets used were named “Road Safety Data – Casualties 1979 – Latest Published Year”, “Road Safety Data – Vehicles 1979 – Latest Published Year”, and “Road Safety Data – Collisions 1979 – Latest Published Year”. These datasets contain comprehensive statistics of all recorded road accidents in the UK since 1979. The data in these sets was collated from the STATS-19 accident report form which is used by police officers responding to road accidents (https://assets.publishing.service.gov.uk/media/60d0cc548fa8f57ce4615110/stats19.pdf).

The datasets were imported into RStudio and joined by accident index using dplyr within the tidyverse package, so that each row in the merged dataset would contain all available data for a single accident. Since the STATS-19 form was updated in 2011 and 2024, and the present project focussed on changes that have occurred since 2012, the data was filtered to only include 2012 through 2023. Furthermore, the data was filtered to include only the variables of interest: accident year, accident severity, road surface conditions, and urban or rural area. Filtering the data in this way significantly reduced the size of the dataset, reducing processing speeds whilst retaining all relevant data. All variables were then recoded so that they reflected their corresponding labels as shown on the STATS-19 form. This ensured that all variables could be interpreted clearly throughout analyses without the need to constantly refer to the STATS-19 form. Much of the existing literature combines fatal and serious accidents into the same category (e.g., Wiratama et al., 2021). Similarly, in the present project, “Serious” and “Fatal” were merged into a category named “Severe” to ensure consistency with previous literature and to increase the number of instances in this category.

To answer the first research question (“How has the prevalence of severe road accidents changed between 2012 and 2023?”), a visualisation was produced. For this visualisation, a dataframe was created that contained the percentage change in both severity types between 2012 and 2023. A bar chart was created using the ggplot2 package, with severity on the x-axis and percentage change on the y-axis. A bar chart was chosen as it allows for direct comparison between the two categories. The scales package was used to manually fill the bars with colours of hexadecimal codes “#D55E00” and “#F0E442”. These colours were chosen because they are accessible (Okabe & Ito, 2008) and they represent an intuitive contrast whereby red is associated with higher severity than yellow (e.g., Elliot et al., 2015). For ease of presentation, the y-axis was scaled such that it ranged from -50% to 50% and the ggthemes package was used to provide a clean format to the chart.

To answer the second research question (“How does the severity of road accidents vary based on road conditions and rurality?”), further bar charts were produced. Bar charts were chosen again to allow for clear contrast between the categories. Only the most recent year, 2023, was used for these charts because this was assumed to be the most representative of current trends. One chart showed the percentage of accidents that were severe across different road conditions, whilst the other showed the percentage of road accidents that were severe in rural areas compared with urban areas. A plain orange colour scheme was chosen for the road conditions chart because there was no intuitive way of colour-coding road conditions that would improve interpretation of the chart, and with five different categories a colour scheme may have overcomplicated the chart unnecessarily. An accessible colour scheme was used for the chart of rurality as this only included two categories, so a colour distinction between these categories eased interpretation without busying the chart.

For the final research question ("To what extent can road conditions and rurality predict the severity of road accidents in the UK?"), a multinomial logistic regression model was produced in an attempt to predict accident severity from road conditions and rurality. A 70/30 train/test split was used such that the model was trained using a random 70% of cases, and the remaining 30% were used to test the model. This was an appropriate ratio as the high sample size meant a 30% test sample was sufficient. Due to the disproportionate number of cases of each severity type, whereby the dataset was overwhelmingly made up of "Slight" accidents, the model was reproduced after over-sampling the minority group and again after under-sampling the majority group. These models were evaluated to assess their accuracy, sensitivity, and specificity using the confusion matrix function in the caret package.

Results and Discussion

How has the prevalence of severe road accidents changed between 2012 and 2023?

Figure 1 is a bar chart that shows the percentage change in accidents from 2012 to 2023, separated by severity. The positive y-axis represents an increase in accident prevalence in 2023 compared with 2012, whilst the negative y-axis represents a decrease in accident prevalence. Crucially, this chart shows that, whilst slight accidents have experienced a significant decline of ~40%, severe accidents have risen almost 10%. The trend in slight accident reduction has been identified by the Department for Transport (Department for Transport, 2023a); they have acknowledged that slight accidents are consistently under-reported in the UK, but they also believe that this reduction is an accurate representation of reality. The rise in the severe category since 2012 is reported by the Department for Transport, but they warn that changes in the ways injuries are reported since 2013 may have skewed some of the severity data. However, they also report adjusted figures which still demonstrate a slight increase in total serious accidents from 2012 to 2023. Table 1 illustrates the extent of this pattern. Severe accidents have risen by 5,107 whilst slight accidents have dropped by 129,623. This means that the change in road accident prevalence is not consistent across severities, but it is unclear exactly what factors are contributing towards this pattern. It is possible that the safety technologies implemented since 2012 are more effective at preventing slight accidents than severe accidents (Abdel-Aty & Ding, 2024), or that worsening road conditions are contributing to the increase in severe accidents (Department for Transport, 2024; Middleton, 2024).

Figure 1: Percentage Change in UK Road Accidents by Severity, from 2012 to 2023

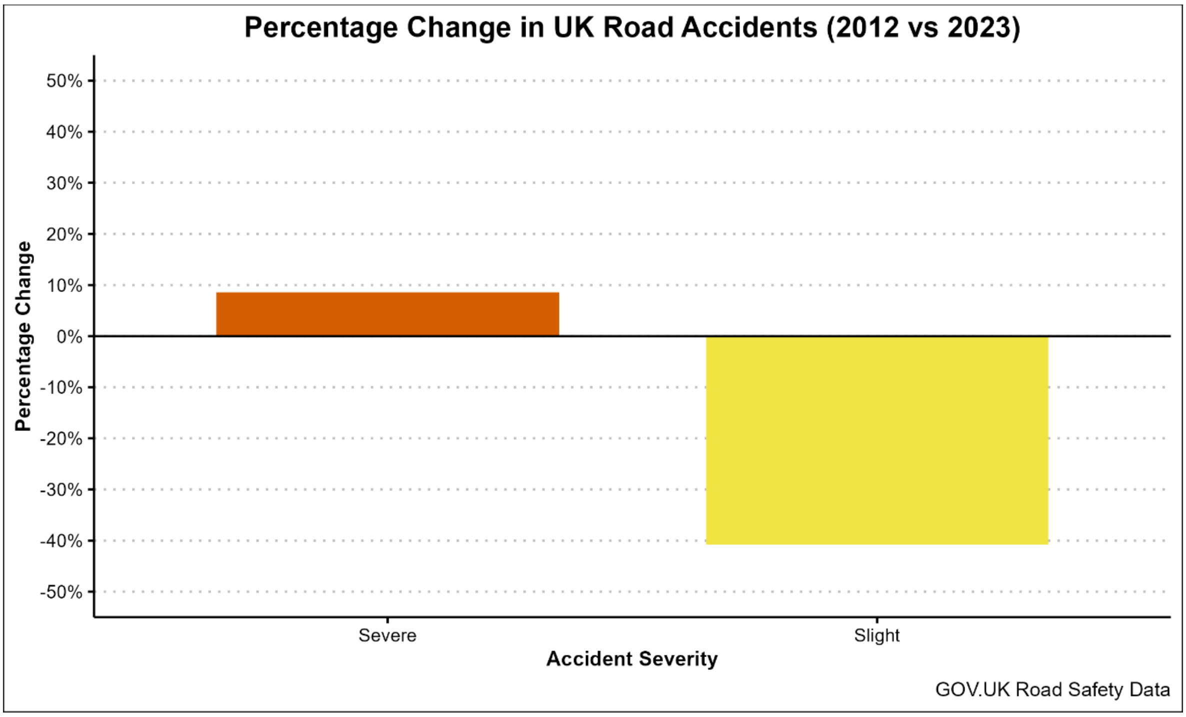

Table 1: Accident Count and Percentage Change by Severity, from 2012 to 2023

Severity	Accident Count (2012)	Accident Count (2023)	Count Change	Percentage Change
Severe	59,612	64,719	+5107	+8.57%
Slight	317,803	188,180	-129,623	-40.8%

How does the severity of road accidents vary based on road conditions and rurality?

Figure 2 is a bar chart showing the proportion of accidents that are severe across different road conditions. It shows that adverse road conditions are more likely to result in severe accidents compared with dry road conditions. Almost 40% of accidents in flooded conditions are severe, compared with ~25% of accidents in dry conditions. Table 2 provides additional context to this pattern by showing the counts and percentage change in severe accidents from 2012 to 2023, across different conditions. Combining the information from the graph and the table we can see that, whilst accidents in flooded conditions are most likely to be severe, floods represent the road condition with the fewest severe accidents overall, both in 2012 and 2023. However, there has been a drastic 56.3% increase in accidents under flooded conditions between 2012 and 2023, likely due to increased rainfall in the UK. The most interesting pattern visible from the table is that all other adverse weather conditions have experienced a decrease in severe accidents, but severe accidents under dry conditions have experienced a 14.2% increase. This suggests that the pattern seen in Figure 1 was not caused by increased adverse weather conditions as previously hypothesised, as the increase in severe accidents occurred overwhelmingly in dry conditions. Nevertheless, it can be concluded that the severity of road accidents does vary by road condition because accidents in adverse weather conditions are more likely to be severe than those in dry conditions, despite experiencing a reduction since 2012.

Figure 2: Percentage of Accidents that were Severe by Road Condition in 2023

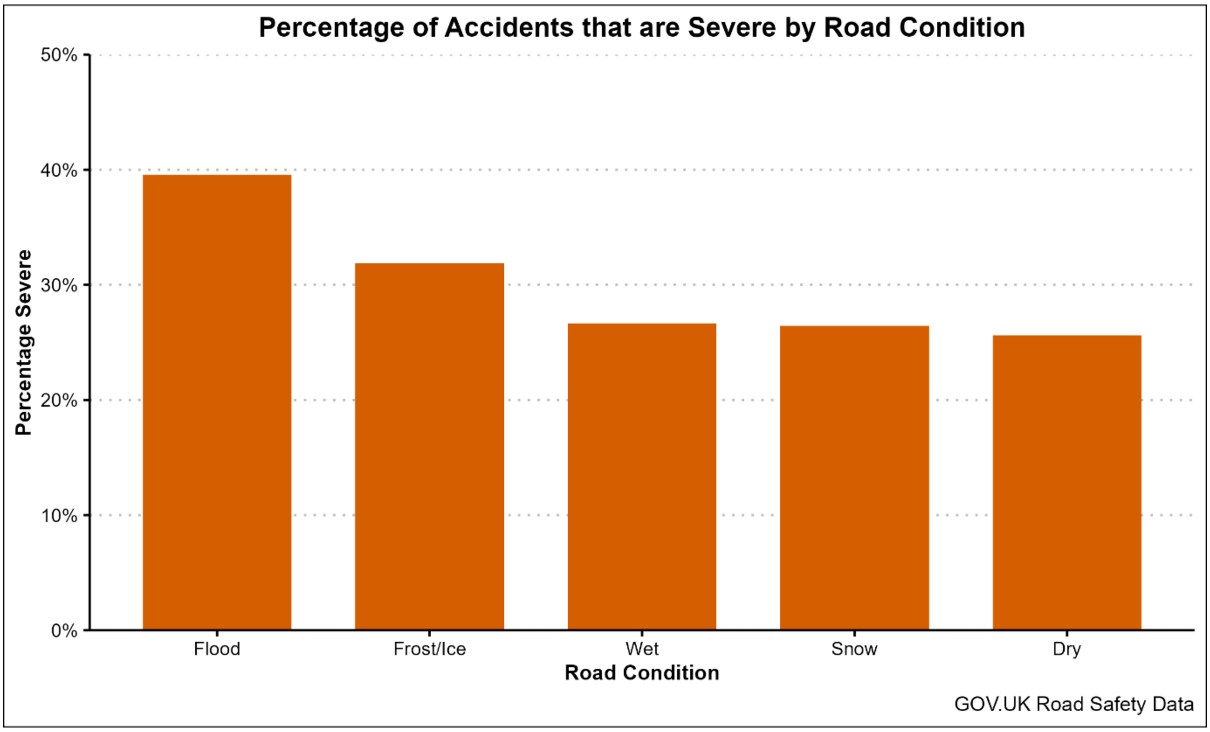

Table 2: Counts and Percentage Change in Severe Accidents by Road Condition

Road Condition	Severe 2012	Severe 2023	Percentage Change
Flood	103	161	+56.3%
Frost/Ice	1,297	1,149	-11.4%
Wet	18,366	17,731	-3.46%
Snow	201	185	-7.96%
Dry	39,549	45,165	+14.2%

Figure 3 is a bar chart visualisation showing the proportion of accidents that are severe in rural areas compared with urban areas. It shows that accidents in rural areas are more likely to be severe than those in urban areas, with ~30% of accidents in rural areas being severe, compared with ~20% in urban areas. As such, this visualisation support previous claims that rural roads are more dangerous than urban roads (e.g., Brake.org, 2019). Table 3 provides additional context by demonstrating how this pattern has changed from 2012 to 2023. Significantly, in 2012 there were more severe accidents in rural areas compared with urban areas, but by 2023 this pattern was reversed. This was caused by a 17.1% increase in the number of severe accidents recorded in urban areas, compared with just a 0.65% increase in rural areas. The increasing urbanisation of the UK (O'Neill, 2024) means that increasingly more miles are being driven in urban areas compared with rural areas (Department for Transport, 2023b), so the increase in severe accidents in urban areas could be a reflection of this. From these inferences, it can be concluded that the severity of road accidents does vary by rurality because accidents in rural areas are more likely to be severe than those in urban areas, despite urban areas experiencing a greater increase since 2012.

Figure 3: Percentage of Accidents that were Severe by Road Condition in 2023

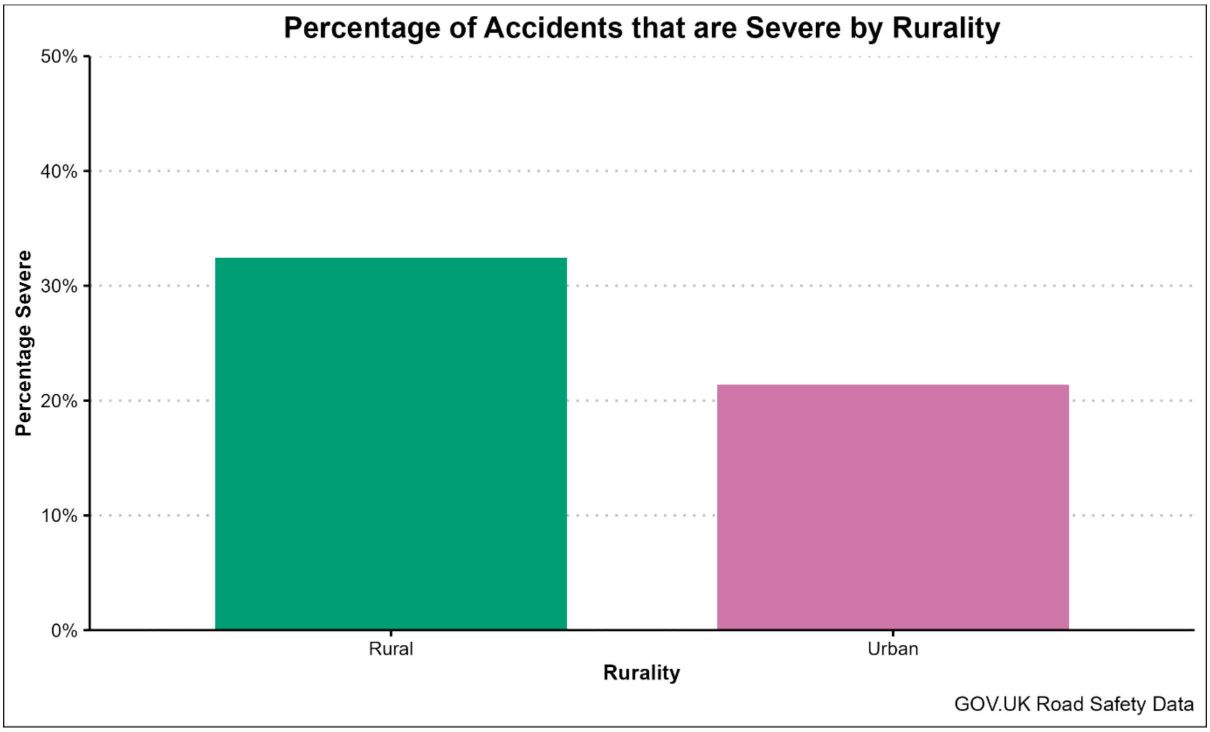

Table 3: Counts and Percentage Change in Severe Accidents by Rurality

Rurality	Severe 2012	Severe 2023	Percentage Change
Rural	30,985	31,185	+0.65%
Urban	28,627	33,528	+17.1%

To what extent can road conditions and rurality predict the severity of road accidents in the UK?

Multinomial logistic regression (MLR) was performed to model the relationship between the predictors (road conditions and rurality) and accident severity. Different sampling variations were used for each model, and the accuracy of each model is shown in Table 4. In Model 1, the default sampling technique was used. This model produced 71% accuracy. However, upon further inspection, sensitivity was found to be 100% and specificity was 0%. This occurred because the model was biased towards the majority class in the sample such that every instance was predicted to be of slight severity, thus revealing poor classification performance. In Model 2, the minority class (severe) was over-sampled to balance the dataset. This yielded a lower accuracy of 53.3%, but sensitivity and specificity were more balance at 50.1% and 61.4%, respectively. In Model 3, the majority class (slight) was under-sampled to balance the dataset, and this model revealed near identical metrics to Model 2. Whilst Models 2 and 3 were more balanced, they still struggled to identify severe instances, and yielded lower accuracy than Model 1. These results suggest that these models cannot accurately predict accident severity – the models are likely a major oversimplification of the myriad factors that are truly involved. Future research may benefit from incorporating additional variables such as weather conditions, speed limits, or driver age.

Table 4: Accuracy, Sensitivity and Specificity of MLR models

MLR Model	Sampling	Accuracy	Sensitivity	Specificity
1	Default	71%	100%	0%
2	Over-sampled Minority	53.3%	50.1%	61.4%
3	Under-sampled Majority	53.4%	50.1%	61.4%

GitHub

Further project details, including R code, can be found on the following GitHub profile: https://github.com/samkensett

The present project is under the INF6027 repository (https://github.com/samkensett/INF6027) which contains three files of R code and a README file. The README file summarises the project and provides instructions for carrying out the project using the provided R code.

Conclusion

This project aimed to answer three research questions relating to the severity of road accidents in the UK. In relation to these, it was found that:

- The prevalence of severe accidents in the UK has increased 8.57% from 2012 to 2023, whilst slight accidents have dropped 40.8%.
- Accidents in adverse weather conditions are more likely to be severe than those in dry conditions, and accidents in rural areas are more likely to be severe than those in urban areas.
- Road conditions and rurality did not accurately predict accident severity using multinomial logistic regression.

To provide context to these findings, it is also worth noting the following:

- The prevalence of severe accidents in dry conditions has risen 14.2% from 2012 to 2023, whilst those in adverse conditions have reduced despite worsening weather in the UK.
- The prevalence of severe accidents in urban areas has increased 17.1% from 2012 to 2023, whilst that in rural areas has increased marginally.
- In 2012, rural areas were responsible for more severe accidents than urban areas. By 2023, this pattern had reversed.

These findings suggest that, whilst rural areas and adverse road conditions are risk factors for road accident severity, more action must be taken to prevent severe accidents in urban areas and dry conditions.

The findings of this project should be interpreted in the context of the following limitations:

- The Department for Transport (Department for Transport, 2023a) has stated that there have been changes to the way road accident injuries are reported which may have impacted the statistical increase in severe accidents since 2012. This has the potential to invalidate the findings of this project that compare 2012 to 2023 because it is unclear whether our measure of severity is consistent across these years. However, the adjusted statistics reported by the Department for Transport still show a slight increase in severe accidents from 2012 to 2023, so there is some confidence that the measure used in this project is valid.
- It is unclear exactly what constitutes a “rural” or “urban” area. In geographic literature, this variable is being increasingly recognised as a spectrum rather than binary. For example, the Organisation for Economic Cooperation and Development has created a classification known as the “Degree of Urbanisation” which recognises an urban-rural continuum (OECD, 2021). The binary measure of rurality used in this project may be an oversimplification of the complexities of infrastructure and lacks interpretability as a result.
- The lack of an accurate multinomial logistic regression model does not mean that rurality and road conditions do not contribute to severity; it likely means that there are many other factors that also contribute. Furthermore, there are a range of alternative techniques that may be more sensitive to these predictor variables like decision trees and random forests. As such, the interpretability of this result is limited and further research is required to determine the extent to which rurality and road conditions can predict accident severity.
- The STATS-19 form does not provide detailed categorisation of the variables it includes. The form may lack reliability if reporting officers interpret variables like severity or road condition differently from one another. Like the rurality variable, these variables could instead be thought of as continuums, and the lack of clear differences between categories reduces their clarity and meaning. Improvements to the STATS-

19 form may assist further research by providing more detailed guidelines for accident categorisation. This would improve the reliability of this reporting measure and also provide researchers with a foundation upon which to interpret their findings.

The assumptions used in this project should also be evaluated:

- It was assumed that all accidents were reported and present in the data. Realistically, many accidents go unreported. This means the data may not be exactly representative of reality. However, the Department for transport has insisted that observed trends are truly representative as only a small minority of minor accidents go unreported.
- It was assumed that all categories were distinct. That is, the rural category is distinct from the urban category, wet conditions are distinct from snow conditions, and so on. In reality, these categorisations are not so definitive as these categories are not mutually exclusive. For example, a road could be partly wet and partly snow. Nevertheless, since there is no bias towards any particular category over another, it may be hypothesised that this would materialise as random variation that applies to all categories equally.

To develop these findings further, future research should:

- Develop a more sensitive regression model to predict accident severity.
- Include a wider range of variables in accident severity prediction.
- Improve the STATS-19 form to ensure interpretability and reliability.

References

Abdel-Aty, M., & Ding, S. (2024). A matched case-control analysis of autonomous vs human-driven vehicle accidents. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-48526-4

Bergel-Hayat, R., Debbarh, M., Antoniou, C., & Yannis, G. (2013). Explaining the road accident risk: Weather effects. *Accident Analysis & Prevention*, *60*, 456-465. https://doi.org/10.1016/j.aap.2013.03.006

Brake.org (accessed 2024, December 31). https://www.brake.org.uk/get-involved/take-action/mybrake/knowledge-centre/road-design/rural-roads

Department for Transport (2015). Actual and weather-adjusted fatalities in reported road. https://assets.publishing.service.gov.uk/media/5a75b13e40f0b67f59fcede6/rrcgb2014-03.pdf

Department for Transport. (2013, July 9). *Drug driving*. GOV.UK. https://www.gov.uk/government/speeches/drug-driving

Department for Transport. (2024). *Road conditions in England to March 2024*. GOV.UK. https://www.gov.uk/government/statistics/road-conditions-in-england-to-march-2024?utm_medium=email&utm_campaign=govuk-notifications-topic&utm_source=a2382564-14dd-4310-add1-b7b843c2474a&utm_content=immediately%5d

Department for Transport. (2023a). *Reported road casualties Great Britain, annual report: 2023*. GOV.UK. https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2023/reported-road-casualties-great-britain-annual-report-2023

Department for Transport. (2023b). *Road traffic estimates in Great Britain, 2023: Traffic in Great Britain by road type*. GOV.UK. https://www.gov.uk/government/statistics/road-traffic-estimates-in-great-britain-2023/road-traffic-estimates-in-great-britain-2023-traffic-in-great-britain-by-road-type

Elliot, A. J., Fairchild, M. D., & Franklin, A. (Eds.). (2015). Color symbolism and association. In *Handbook of Color Psychology* (pp. 317–432). part, Cambridge: Cambridge University Press.

European Commission (2018). Advanced Driver Assistance Systems https://road-safety.transport.ec.europa.eu/document/download/93f6a003-83dc-4e66-ae14-852d93236e57_en?filename=ersosynthesis2018-adas.pdf

Malin, F., Norros, I., & Innamaa, S. (2019). Accident risk of road and weather conditions on different road types. *Accident Analysis & Prevention*, *122*, 181-188. https://doi.org/10.1016/j.aap.2018.10.014

Met Office. (2019, July 31). *State of the UK Climate*. Met Office; Met Office. https://www.metoffice.gov.uk/research/climate/maps-and-data/about/state-of-climate

Middleton, N., & Middleton, N. (2024, December 24). *Road conditions in England continue to deteriorate, DfT data shows*. Great British Fleet Event. https://greatbritishfleetevent.co.uk/road-conditions-in-england-continue-to-deteriorate-dft-data-shows

O'Neill, A. (2024). *United Kingdom - Urbanization 2017.* Statista. https://www.statista.com/statistics/270369/urbanization-in-the-united-kingdom/

OECD (2024). Applying the Degree of Urbanisation. https://www.oecd.org/en/publications/applying-the-degree-of-urbanisation_4bc1c502-en.html

Okabe, M., Ito, K. (2008). Color Universal Design (CUD) – How to make figures and presentations that are friendly to colorblind people. *https://jfly.uni-koeln.de/color/#cudo*

The AA (accessed 2024, December 31). Guide to Euro NCAP crash tests and how they work. https://www.theaa.com/driving-advice/safety/euro-ncap

Wiratama, B. S., Chen, P.-L., Chen, L.-H., Saleh, W., Chen, S.-K., Chen, H.-T., Lin, H.-A., & Pai, C.-W. (2021). Evaluating the Effects of Holidays on Road Crash Injuries in the United Kingdom. *International Journal of Environmental Research and Public Health*, *18*(1), 280. https://doi.org/10.3390/ijerph18010280

World Health Organization. (2023, December 13). *Road traffic injuries*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries