

# An investigation of the relationship between speed limit and various demographic, environmental and spatial factors on road accident severity in Great Britain

Group 9

27/02/2023

## Introduction

According to the World Health Organisation, there are approximately 1.3 million fatalities globally each year due to road traffic accidents [1]. Indeed, road traffic injuries are the primary cause of death for those aged between five and 29 years old [1]. Vulnerable road users such as pedestrians, cyclists and motorcyclists account for over half of all road traffic accident fatalities globally [1]. Although most deaths occur in low-to middle-income countries, it is nevertheless an issue of significant concern for high-income countries also.

It is well-documented that driving at an increased speed is related to both the risk of a road traffic accident and the severity of the accident, should it occur. It is reported that for every 1% increase in speed, there is a corresponding 3% increase in risk of serious crashes and 4% increase in the risk of fatal crashes. [1] Other risk factors for accident severity include demographic factors; approximately three quarters (73%) of all road traffic fatalities involve males under 25 years old [1]. In addition, it can be hypothesised that variables such as the time of day that the accident occurred, the condition of the road surface where the accident occurred e.g. the presence of ice which could lead to skidding or water which could result in a longer stoppage time, and the location of the accident in terms of whether it occurred close to a junction or not, could also contribute to the severity of a road traffic accident.

An increased understanding of the impact of these factors on road accident severity could aid policymakers in decision making and be used to inform awareness campaigns aimed to promote safe driving practices. A clearer understanding of the impact of contributing factors to increased road accident severity could result in the introduction of more targeted preventative measures and a more optimal use of finite resources, aimed at increasing road safety.

The primary aim of this analysis is to determine the relationship between severity of road traffic accidents in Great Britain and the speed limit in place where the accident occurred. Secondary aims include investigating the effects of environmental, spatial and demographic factors on road accident severity. More specifically, the relationship between accident severity and road surface conditions, time of day, casualty type (e.g. pedestrian, cyclist, car occupant), junction location and involvement of males under 25 years old, will be explored as secondary aims.

Information on the data source and methods used to analyse the data will be provided in the subsequent Methods section. The Results section will present the main findings, including figures and tables. The Conclusions and Discussion section will elaborate on the findings and discuss the limitations of the analysis.

## Methods

### Data Sources

This retrospective analysis utilised road safety data for the year 2020, which is available online from the UK's Department of Transport [gov.uk](https://gov.uk). The data which is published provides details of road accidents which occurred on public roads in Great Britain, involved at least one vehicle and which resulted in human injury. Information is available on the circumstances of the accident, such as weather conditions, time of day, visibility, speed limit, road type and surface conditions. In addition, data is available on the types of vehicles and characteristics of the casualties involved.

The outcome variable selected for this analysis was road accident severity. The available data reports accident severity as fatal, serious or slight. For this analysis, the fatal and serious groups were combined, so that accidents were sub-divided into two categories: 'fatal or serious' and 'slight.' In addition, a subset of the other available variables were extracted from the data for analysis, based on their likelihood to impact accident severity. An overview of the selected variables and how they were coded for analysis is provided in the Appendix, Table 1.

### Selection of variables

There are too many variables included in the data sets to use all of them so a selection of variables were selected. These variables were selected to show a wide range of factors that could impact on accident severity.

- Speed limit (accidents data): The relationship between speed limit and accident severity is the main focus of the study. Speed limit can be considered as a categorical variable as there is only a set number of speed limit options. However, it is impossible to tell from the data whether a vehicle is travelling at, above or below the speed limit. Thus, the speed limit will be considered as a continuous, quantitative variable. Moreover, this will allow increased flexibility when predicting accident severity at different speeds.
- Road surface conditions (accidents data): This is often considered to be a key factor in accidents and it is vital to consider if this impacts on accident severity. The original data set had seven categories. However, five of these categories contained only a total of 2.4% of all accidents. It was therefore considered to be reasonable to reduce this to two variables. These were dry and wet/damp which contain 68.7% and 28.8%, respectively, of all accidents.
- Time of day (accidents data): time of day was considered important to include as it allows for the impact of the tiredness of the driver to be included in the model, as well as other associated driving issues such as drunk drivers. The original data records the actual time of each accident. This was reduced to the subcategories of night off peak (7pm-7am), morning peak (7am-10am), day off peak (10am-4pm) and evening peak (4pm-7pm). There was a fairly even split of accidents between each category: 24.9%, 14.2%, 37% and 23.9% respectively. There was no data missing for this variable.
- Casualty type (casualty data sheet): The type of casualty was considered to be an interesting variable. The raw data contains many types of casualty which has been collapsed to car occupant (55.5%), cyclist (14.1%), motorcyclist (11.4%), pedestrian (12.8%) and all other vehicles (6.2%). Although some groups contain more than others, each group contains enough entries to justify their inclusion in the analysis.
- Age and gender of driver (vehicle data sheet): Young, male drivers are often cited as causing the highest number of accidents. It was therefore considered vital to examine the impact that these drivers have on accident severity. As accidents often involve multiple drivers it is difficult to analyse their impact. In each accident the age of the youngest driver was recorded, and it was possible to isolate all accidents that contained a driver who was male and under 25 years old. Thus allowing the impact of the driver's age on accident severity to be assessed.
- Proximity to a junction (vehicle data sheet): The proximity to a junction could impact the severity of an accident. The raw data contained many categories which were collapsed into two categories. These are "at a junction/roundabout" or "not at a junction/roundabout." Each category contains ...% and ...% of results.

- There were some entries that have recordings of “not available” or “unknown.” These were only a very small proportion of the data so it was considered valid to discount these from our analysis.

## Data analysis

**Checking tyhe Variable Selection** After having selected the initial variables that were thought of having an impact on accident severity some exploratory data analysis was conducted to see if this was in fact the case. This will allow better decision making to be made when deciding which terms should and should not be included in the model. Additionally, this will provide some indications of what outcomes would be expected from the model.

- The data will be checked to verify whether there are any combinations of variables where all the accidents are either severe or slight. This will be done by calculating the mean of severe accidents for each combination of the variables. This will allow the proportion of severe responses to be seen.
- Contingency tables will be created that allows the comparison of the variables, and will allow, by inspection, to see which variables affect the accident severity.
- Box plots were constructed to show if there are any interaction effects between the speed limit and other variables. These will be created with all of the variables.
- The mean speed limit will be included so that easy comparisons can be made.

**Creating a regression model** A Generalised Linear Model will be created to allow predictions to be made about how accident severity changes for different speed limits given a certain set of conditions - such as road surface, driver’s age and so on. As we have binary outcomes - severe or not severe, we will use a specific type of regression analysis called logistic regression. We will use a multiple regression model as it allows us to consider more than one factor that might affect accident severity.

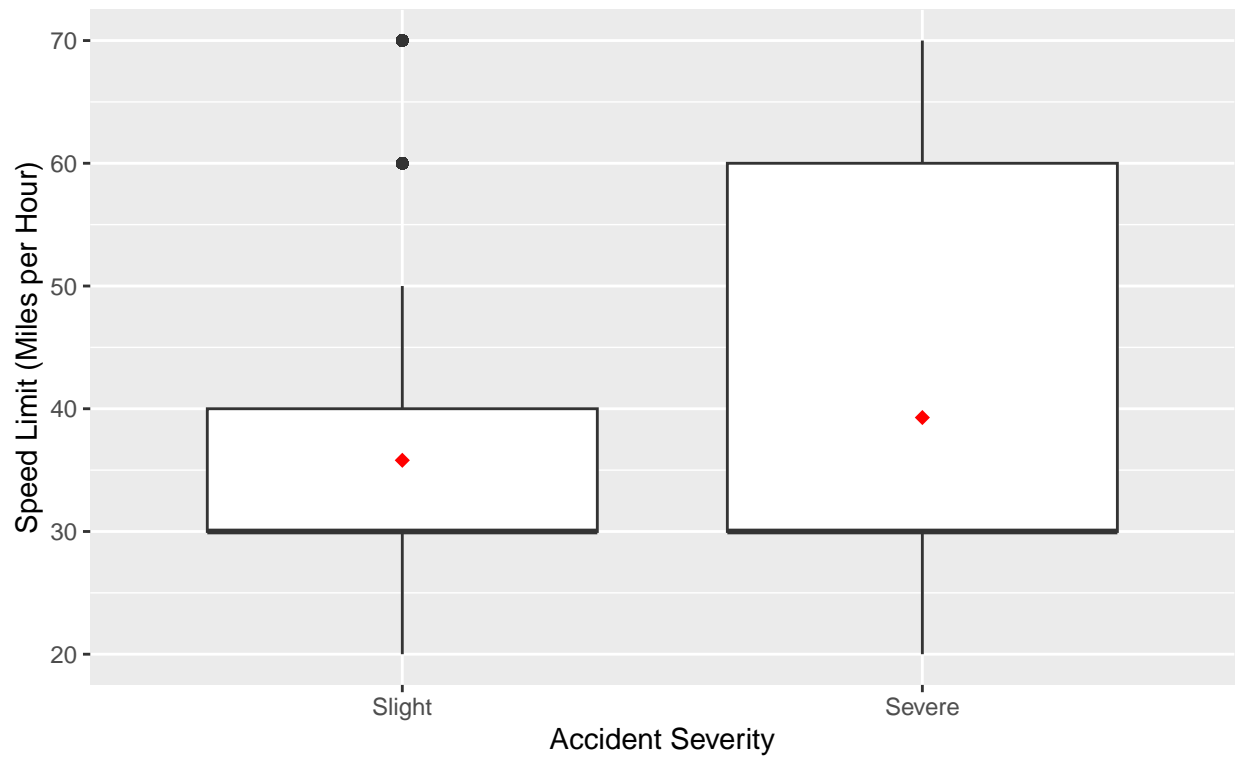
The value of each variable will then be reassessed to consider its significance in predicting accident severity. If it is not considered to be relevant, it will be removed from the model. We will consider a 5% significance level to be appropriate.

## Results

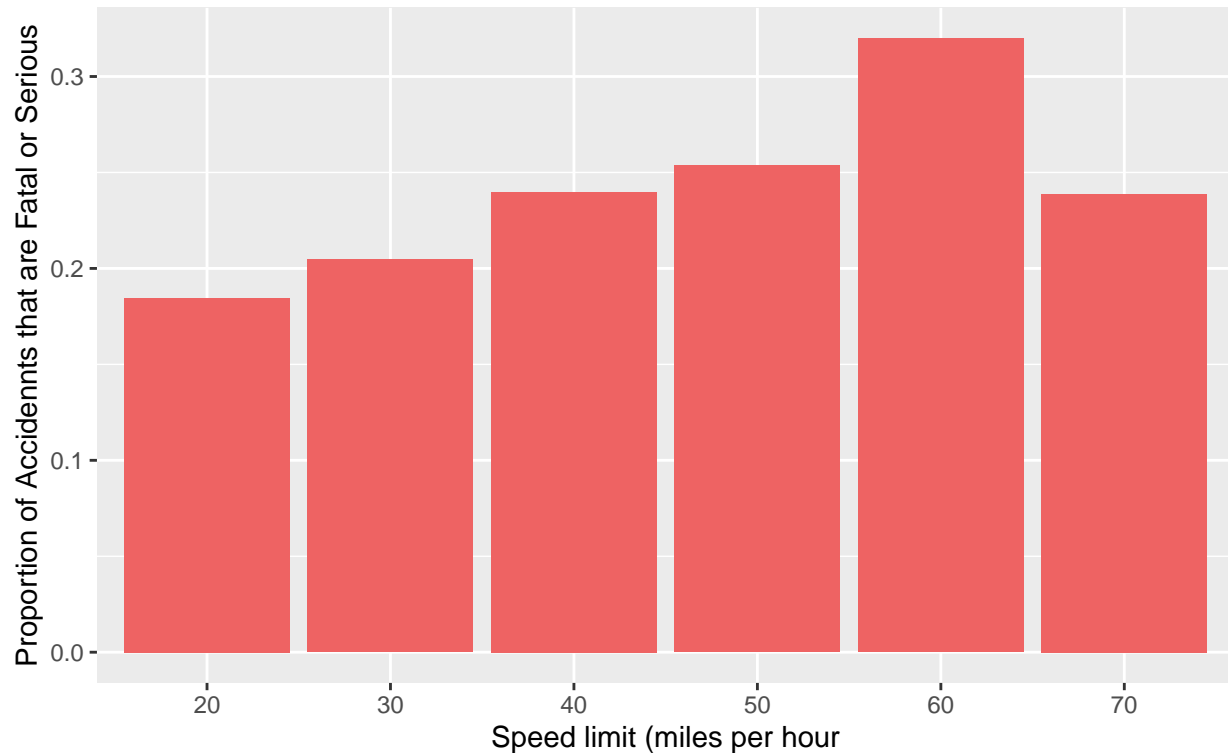
### Graphs

Description and text to follow, linking to eventual regression

The speed limit in place tends to be higher for fatal or serious accidents compared to slight accidents. The red diamond indicates the mean speed limit.



The proportion of serious or fatal accidents increases with increase in speed  
 Bar graph of speed limit versus proportion of fatal or severe accidents



Plot is static and needs implementing in rmarkdown

## Regression Model

In this analysis we use a specific type of regression analysis for binary outcomes called a logistic regression.

In addition we have included more than one variable, creating a multiple regression. This allows us to “control” for variables.

- Provide better explanation as to what a logistic and multiple regression is and why it is appropriate
- Explain in semi-lay terms how to interpret the results/coefficients ( exponentiate for odds?)
- Discuss the “significance”

```
##
## Call:
## glm(formula = accident_severity ~ speed_limit + factor(road_surface_conditions) +
##      factor(time_of_day) + casualty_pedestrian + male_u25 + junction,
##      family = binomial(), data = dat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3054  -0.7319  -0.6352  -0.5597   2.0735
##
## Coefficients:
##                                Estimate Std. Error z value
## (Intercept)                   -2.1237078  0.0341473 -62.192
```

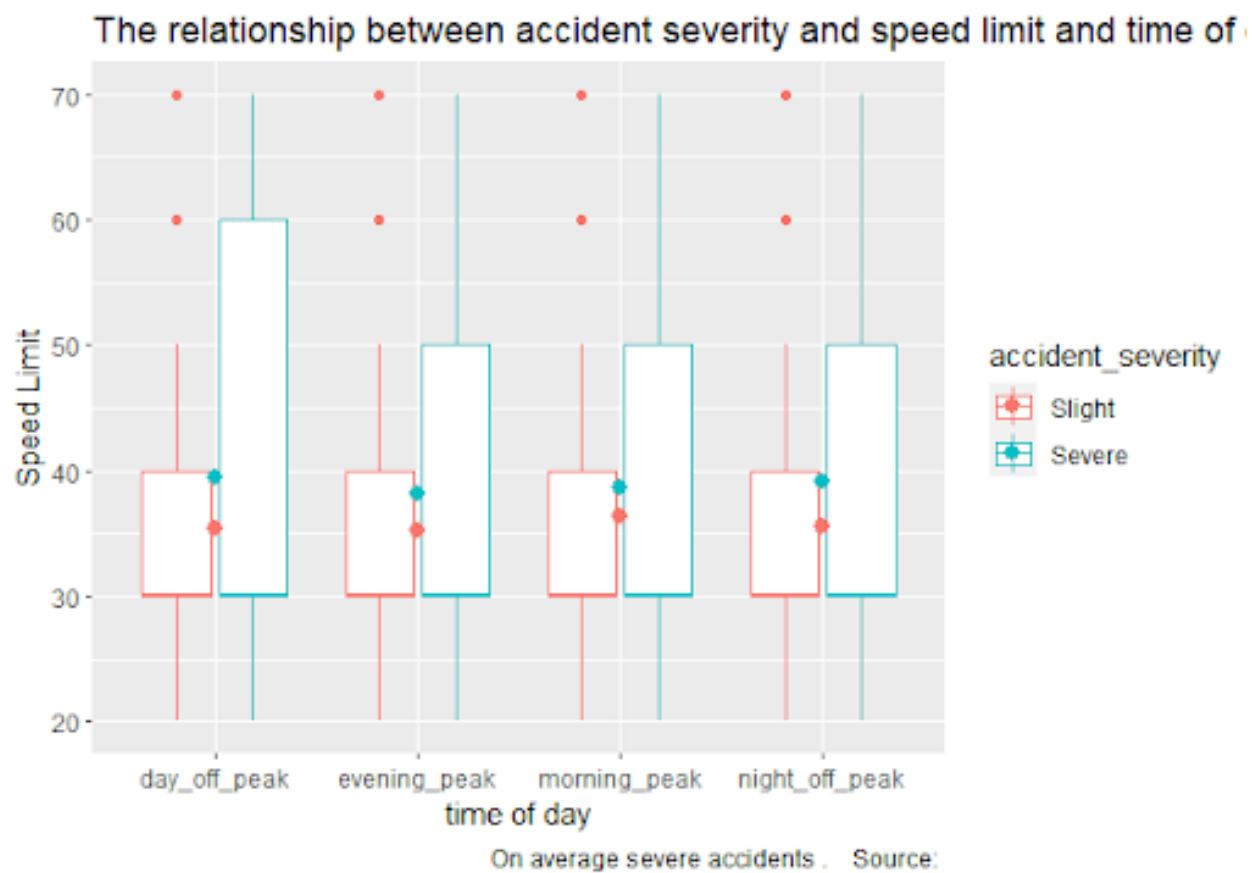


Figure 1: timeOfDay

```

## speed_limit                0.0174541  0.0007509  23.245
## factor(road_surface_conditions)Other -0.3225039  0.0947809  -3.403
## factor(road_surface_conditions)Wet or damp -0.1169852  0.0223723  -5.229
## factor(time_of_day)evening_peak      0.0442646  0.0265670   1.666
## factor(time_of_day)morning_peak     -0.1032896  0.0321944  -3.208
## factor(time_of_day)night_off_peak    0.2401133  0.0263477   9.113
## casualty_pedestrianTRUE             0.8416050  0.0283482  29.688
## male_u25                         0.1320840  0.0238221   5.545
## junction                        0.1161186  0.0216394   5.366
##                               Pr(>|z|)
## (Intercept)                     < 2e-16 ***
## speed_limit                      < 2e-16 ***
## factor(road_surface_conditions)Other 0.000667 ***
## factor(road_surface_conditions)Wet or damp 1.70e-07 ***
## factor(time_of_day)evening_peak      0.095683 .
## factor(time_of_day)morning_peak      0.001335 **
## factor(time_of_day)night_off_peak    < 2e-16 ***
## casualty_pedestrianTRUE              < 2e-16 ***
## male_u25                           2.95e-08 ***
## junction                           8.05e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 61666  on 57683  degrees of freedom
## Residual deviance: 60225  on 57674  degrees of freedom
## AIC: 60245
##
## Number of Fisher Scoring iterations: 4

```

## References

[1] who, “Who.”