Statistical methods for comparing road traffic collision and casualty rates: proposed approach

Contents

1 Foreword

Our roads play a vital part in people's lives: connecting people to their work, family and friends. We want to make sure every person gets to where they want to go, safely and reliably. That's why we prioritise the safety of people who travel and work on our roads above everything else.

We can't eliminate every risk from our road network, or from our work. But we can recognise those risks, assess them, and do everything we can to protect people from them. That includes ensuring the way that we use the data collected on our network helps us to make decisions about improving it.

Agreeing robust principles for how we analyse data and use statistics is key. In this report we describe statistical methods that we propose can be used to compare road traffic collision rates and casualty rates.

We have developed methods for calculating confidence intervals for both collision and casualty rates, and for comparing collision and casualty rates using hypothesis tests. These methods help people to interpret the statistics appropriately, and to not jump to conclusions when observed differences may be due to natural variability in the data rather than a difference in safeness. As far as we are aware, there are no existing methods available to compare casualty rates in this way.

During the development of these methods, we sought advice and feedback from the Department for Transport, the Methodology Advisory Service of the Government Statistical Service and Professor Jonathan Tawn of Lancaster University. However, while the proposed methods have been carefully developed, they are not yet finalised. We want our methods to be scrutinised and reviewed by the statistical community. We welcome further feedback on the methods described in this document

Through this work we look to contribute to delivering our ambition that no one should be killed or injured when travelling on our roads.

Mark Clements

Chief Analyst

2 Introduction

The aim of this document is to describe proposed confidence interval and hypothesis testing methods. These can be used by National Highways, and others, to compare road traffic collision and casualty rates derived from STATS19[1] data.

These methods have been developed in response to the Office of Road and Rail (ORR) quality assurance of all lane running motorway data report [13]. This noted that 'undertaking significance testing on the headline figures [casualty rates] in future would help explain the levels of uncertainty around the results. We recommend that this is developed' and 'Including information about the level of uncertainty associated with the high-level statistics [overall casualty rates for different road types], through statistical significance testing, would add important context to any conclusions'. However, the proposed methods have wider applicability beyond only comparing motorway types. For example, they could be used to compare collision and casualty rates of any two roads or road sections, or of a single road from year to year.

We have developed these methods to be used with collision and casualty rates derived from data collected using the STATS19 system, managed by the Department for Transport. That is, a collision is defined as one which occurs on the public highway, involves at least one vehicle, becomes known to the police within 30 days, and causes personal injury. This necessitates that at least one casualty arises from every collision – this dataset does not record damage only road traffic collisions.

In the remainder of this document, we describe our proposed methods to compare collision rates and casualty rates of two roads, which we call Road 1 and Road 2. Note that these could equally be two road types or a single road in two consecutive years. We then show a worked example of applying these methods to fictitious data, and finish with some guidance on important things to consider when using these methods.

2.1 Engagement and feedback

During the development of these methods, National Highways sought advice and feedback from the Department for Transport, the Methodology Advisory Service of the Government Statistical Service, and Professor Jonathan Tawn of Lancaster University. Alternative approaches were considered for many of the methods outlined in this publication, and we are content with the methods we are proposing.

However, while the proposed methods have been carefully developed, they are not yet finalised. National Highways welcomes further feedback on the methods described in this document. We are keen to hear from you if you uncover a problem with our proposed methods or can suggest an alternative approach which would be a substantial improvement. However, at this stage in the process we are not seeking feedback which would lead to only minor improvement. Feedback can be given to networkanalysisandstatistics@nationalhighways.co.uk by the end of August 2022.

[1] https://www.gov.uk/government/collections/road-accidents-and-safety-statistics

2.2 Natural variability

Before describing the proposed methods in the remainder of this document, we first clarify where the uncertainty in collision and casualty rates arises. We assume that the statistics we work with – that is the number and rate of road traffic collisions and the number and rate of the resulting casualties – have been correctly recorded. Therefore, in this case uncertainty does not arise from partial observation of events, for example as it does when working with survey data collected from a sample of individuals. Instead, uncertainty arises from natural variability. That is, the fundamental unpredictability of the natural world and, in particular, of rare events like road traffic collisions [15].

Due to this natural variability, we can treat the statistics – for example the observed collision rate of a road - as a single estimate of an underlying hypothetical quantity that we cannot measure directly – for example the ‘true’ underlying collision rate of a road.

Suppose we have observed road traffic collision rates of two different roads, say a rate of 5 casualties per hundred million vehicle miles travelled on Road 1 and a rate of 10 casualties per hundred million vehicle miles travelled on Road 2. We can easily say which of these casualty rates is larger: these are just numbers and 10 is larger than 5. However, the question we are actually interested in answering is which of the underlying casualty rates is larger. Considered along with other evidence, the answers to these types of questions helps us draw conclusions about the underlying safeness of the roads.

3 Proposed methods for comparing road traffic collision rates

The Poisson distribution is already used by both the Department for Transport (DfT) [7] and National Highways [8] to model road traffic collisions. Based on this statistical model and the assumption that the road traffic level is sufficiently large, currently a Z-test is used to calculate a p-value to statistically compare underlying collision rates. This Z-test is also known as a chi-squared test with 1 degree of freedom. In the following section, we describe the statistical methods we propose to compare the number of collisions that do not make the assumption that the road traffic levels are sufficiently large. We then follow by describing how confidence intervals for the underlying collision rate can be calculated using this statistical model. We conclude this section by describing how a formal hypothesis test can be conducted to determine whether there is sufficient evidence to suggest that the underlying collision rates of two roads are different.

3.1 Statistical model

We assume that road traffic collisions occur according to a non-homogeneous Poisson process [11] with the rate dependent on the observed road traffic. As for the existing Z-test approach, the collisions are assumed to be independent; this means that the occurrence of one collision has no influence on the occurrence of another collision. The independence assumption is thought to be valid in most cases. We use maximum likelihood techniques [5] to estimate the underlying collision rates.

In the existing Z-test approach, it is assumed that daily counts of collisions follow a Poisson distribution. This implies that there is a constant expected daily rate of road traffic collisions through time. On the other hand, we assume that road traffic collisions occur according to a non-homogeneous Poisson process; this implies that the rate of collisions is not constant. These are different starting points, but ultimately lead to the same statistical distribution for the number of collisions, as under our assumption it can be shown that the daily number of collisions still follows a Poisson distribution, but with a rate dependent on the daily vehicle miles observed. We feel the non-homogeneous Poisson process better justifies the assumption of the Poisson distribution for the number of collisions given the level of road traffic, as when we have loooked at data, we see that the rate of collisions is not constant.

For Road i, we have data collected over a time period during which N_i collisions are observed and the amount of road traffic observed is v_i vehicle miles. The observed collision rate, R_i, is given by:

$$R_i = \frac{N_i}{v_i}. \tag{1}$$

We assume that the rate at which collisions occur is governed by the intensity function, $\lambda_i(v)$, where v is the road traffic. Therefore, the number of collisions for Road i given the level of road traffic follows the following Poisson distribution:

$$N_i \sim \text{Poisson}\left(\int_0^v \lambda_i(\nu)d\nu\right). \tag{2}$$

We re-parameterise the Poisson distribution in terms of γ_i, the underlying expected number of collisions per vehicle mile, where $\gamma_i = \frac{1}{v_i}\int_0^{v_i} \lambda_i(\nu)d\nu$ so that $N_i \sim \text{Poisson}(\gamma_i v_i)$.

Given the two roads that we wish to compare, Road 1 and Road 2, we calculate estimates for the underlying collision rates using maximum likelihood techniques as follows. If we assume the number of collisions observed on Road 1 and Road 2 are independent, the likelihood function is given by:

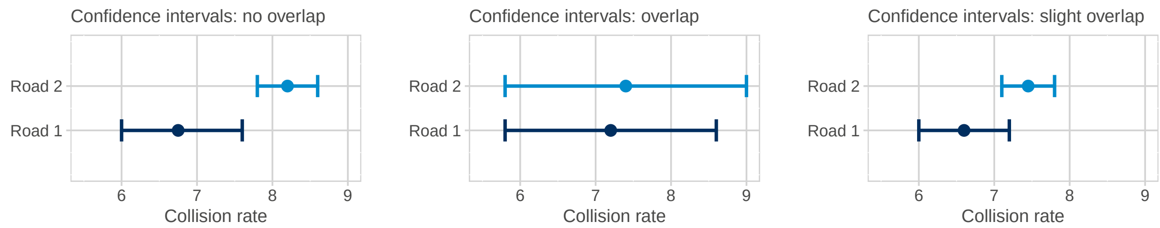
$$L(\gamma_1, \gamma_2) = \frac{e^{-\gamma_1 v_1}(\gamma_1 v_1)^{N_1}}{N_1!}\frac{e^{-\gamma_2 v_2}(\gamma_2 v_2)^{N_2}}{N_2!}. \tag{3}$$

Assuming the underlying collision rates of Road 1 and Road 2 are different, the maximum likelihood estimate for Road i is given by $\hat{\gamma}_i = \frac{N_i}{v_i}$ for $i = 1, 2$. The maximum likelihood estimates provide a single value for the estimate of the underlying collision rates. In the next section we describe how to compute confidence intervals to provide a measure of uncertainty in these estimates.

3.2 Confidence intervals

We propose calculating a confidence interval for $\hat{\gamma}_i$ using a parametric bootstrap method [4], given by Algorithm 1 in Appendix A.1.

After calculating the confidence intervals, we may find ourselves in one of three situations illustrated in Figure 1. Figure 1a shows that the confidence intervals do not overlap, which may indicate that the underlying collision rates for Road 1 and Road 2 are different. Figure 1b shows large overlap between the confidence intervals, which may indicate that the underlying collision rates for Road 1 and Road 2 are similar. Finally, Figure 1c shows some overlap between the confidence intervals. Here, in particular, a formal statistical hypothesis test can help us determine if the evidence suggests that the underlying collision rates are different. In the following section, we state the hypotheses considered and describe how to compute a p-value to test the hypothesis.



(a) Situation 1 (b) Situation 2 (c) Situation 3

Figure 1: An illustration of the three situations that could be observed when the confidence intervals for the underlying collision rates are computed. The points show the observed collision rates and the lines indicate the extent of the confidence intervals.

3.3 Hypothesis test

Following the observations made using the confidence intervals, we can formally test the null hypothesis that the underlying collision rates are the same for Road 1 and Road 2 as follows. The null hypothesis, H_0, and alternate hypothesis, H_1, are defined as:

$$\begin{aligned} H_0 &: \gamma_1 = \gamma_2 \\ H_1 &: \gamma_1 \neq \gamma_2. \end{aligned} \tag{4}$$

There are several test procedures described in the literature that can be used to test this hypothesis, including the Z-test or chi-squared test with 1 degree of freedom described previously. However, several of these test procedures rely on the asymptotic distribution of the test statistic [12]. When levels of road traffic are low, the distribution of the test statistic may be far from the asymptotic distribution of the test statistic, meaning that the p-values produced by these methods may be misleading. The C-test proposed by Przyborowski & Wilenski [14] does not rely on asymptotic distributions but did not perform as favorably as the numerical approach (E-test) proposed by Krishnamoorthy and Thomson [9]. However, what is known by the Neymann-Pearson lemma [2] is that the optimal way to obtain a test statistic in the single parameter hypothesis testing case is to use the likelihood ratio as the basis of constructing a test statistic. Therefore, we conduct a likelihood ratio test and propose calculating a p-value using a Monte-Carlo approach [1]. Appendix A.2 describes this calculation in detail.

In this section we have shown how we propose to model the number of collisions given the road traffic, estimate the underlying collision rates, measure the uncertainty in these estimates and conduct a hypothesis test to determine whether there is sufficient evidence to suggest that collision rates are different. In the following section, we consider how to model the number of casualties and measure uncertainty in the estimates of the underlying casualty rates.

4 Proposed methods for comparing casualty rates

Casualties are a result of road traffic collisions. Therefore, the total number of casualties is dependent on the total number of collisions and the number of casualties that result from each collision.

Let $X_{i,j}$ be the observed number of casualties that result from collision j of Road i. Recall that the observed number of collisions for Road i is N_i, and the observed road traffic is v_i. The observed casualty rate for Road i, Q_i, is dependent on the number of collisions and the number of casualties, $C_i = \sum_{j=1}^{N_i} X_{i,j}$ as follows:

$$Q_i = \frac{C_i}{v_i} = \frac{\sum_{j=1}^{N_i} X_{i,j}}{v_i}. \tag{5}$$

4.1 Statistical model

We assume that collisions occur according to a non-homogeneous Poisson process as described in Section 3.1, so that the number of collisions given the level of road traffic can be shown to follow a Poisson distribution. Since the number of casualties is given by the sum of the number of casualties per collision, where the number of collisions is assumed to follow a Poisson distribution, we assume that the number of casualties can be well-modeled by a compound Poisson process [10]. Here, we assume that the number of casualties per collision are independent and come from an unspecified probability distribution.

In the development of these methods we have explored fitting a Poisson, a negative-binomial and a geometric distribution to the number of casualties per collision but found the fit of all distributions to be poor. We also explored truncating the aforementioned distributions at zero and inflating the distributions at one [3], which made only marginal improvements. We hypothesise that the poor fit of these distributions could be due to several factors such as, but not limited to, the combination of the types of vehicle involved in collisions and the number of occupants in a vehicle. These factors suggest that a mixture of distributions may be more suitable to model the number of casualties per collision parametrically. However, determining the exact form of the mixture distribution is non-trivial, and so we propose proceeding with a non-parametric approach.

4.2 Confidence intervals

Confidence intervals for the underlying casualty rates of Road 1 and Road 2 can be calculated to informally compare them. We propose doing this using a two step process to simulate a compound Poisson process. Initially, the number of collisions are simulated from a Poisson distribution. Then, the number of casualties per collision is sampled from the observed casualties per collision distribution. The $100(1-\alpha)\%$ confidence intervals for the casualty rates are computed by taking the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ simulated quantiles of the number of casualties and dividing through by the observed road traffic. Full details of this are given by Algorithm 2 in Appendix B.1.

The relative positions of the confidence intervals of Road 1 and Road 2 give us some idea whether the evidence suggests that the underlying casualty rates are different. However, to formally compare the underlying casualty rates, we propose conducting statistical hypothesis tests. To determine whether the evidence suggests that the casualty rates for Road 1 and Road 2 are different, we need to compare both the collision rates, using the methods proposed in Section 3, and the number of casualties per collision for Road 1 and 2. In the following section we describe our proposed approach for the latter.

4.3 Hypothesis test

In Section 3 we described a test procedure to determine if the evidence suggests a difference in the underlying collision rates. The same methodology cannot by used to compare the underlying casualty rates as the non-homogeneous Poisson process used to model collisions assumes independence and that only one collision can occur at any given level of road traffic. However, more than one casualty can result from a single collision and the injuries are sustained at exactly the same time. The variability in the number of casualties resulting from each collision means that, in comparison to the variability in the number of collisions, the uncertainty in the total number of casualties will be higher. Therefore, if the collision rate methodology was used to compare the casualty rates, unrealistically confident conclusions could be drawn by underestimating the uncertainty.

We assume that the number of casualties can be well modeled by the compound Poisson process. It can be shown that the expected casualty rate is given by the product of the expected collision rate, and the first moment of the number of casualties per collision distribution. Therefore, when seeking evidence to compare the casualty rates for two roads, it is sufficient to consider the expected collision rate and the first moment of the number of casualties per collision distribution.

We have already discussed how to test for a difference in the underlying collision rates and now focus on testing for a difference in the first moment of the casualty per collision distribution. That is, testing the following hypothesis:

$$\begin{aligned} H_0 &: \mathbb{E}[X_1] = \mathbb{E}[X_2] \\ H_1 &: \mathbb{E}[X_1] \neq \mathbb{E}[X_2]. \end{aligned} \tag{6}$$

In Section 4.1 we discussed the difficulty in specifying a suitable parametric distribution for the number of casualties per collision, hence we propose to use a non-parametric approach to calculate a p-value.

These proposed methods assume that the collisions are independent, the number of casualties per collision is independent and the number of collisions is independent to the number of casualties per collision. The test statistic used in our our non-parametric test is the absolute difference in the two sample means. Assuming the null hypothesis holds, our non-parametric approach uses bootstrap re-sampling [4] without replacement to estimate the distribution of the test statistic. The procedure proposed to compute a p-value to test our hypothesis is given in Appendix B.2.

If only a small number of collisions for a road are observed, we do not have much information about the distribution of the number of casualties per collision. In this case, it may be that we cannot conduct this hypothesis test comparing the first moment of the number of casualties per collision between two roads or that we should treat the conclusions with additional caution.

Considering the outcomes of the two hypothesis tests together will determine whether we should conclude that the underlying casualty rates are different. For example, if we conclude that it is likely the underlying collision rates are different, but there is no evidence to suggest that the first moment of the casualties per collision are different, then we conclude that the underlying casualty rate is different. If we conclude that it is likely the underlying collision rates are different and the first moment of the casualties per collision are different then the underlying casualty rate could be different, or they could be the same in the surprising case where the two differences act in opposite ways and cancel out. In this case, we need to return to the confidence intervals calculated to assess whether the evidence suggests that the underlying casualty rates are different or not.

4.4 Formally combining p-values

In Section 3 we proposed a hypothesis test for the difference in the underlying collision rates between the two roads, and in Section 4.3 we proposed how to test for a difference in the first moment of the casualties per collision. We then proposed using the outcomes of those two individual hypothesis tests to draw conclusions about the underlying casualty rates. In this section we describe a method for formally combining the information from the hypothesis tests described previously to produce a single p-value. Note that this is still not a single p-value to test differences in casualty rates directly.

The test considers the following hypotheses:

$$\begin{aligned} H_0 &: \gamma_1 = \gamma_2 \text{ and } \mathbb{E}[X_1] = \mathbb{E}[X_2] \\ H_1 &: (\gamma_1 \neq \gamma_2 \text{ and } \mathbb{E}[X_1] = \mathbb{E}[X_2]) \text{ or} \\ &\quad (\gamma_1 = \gamma_2 \text{ and } \mathbb{E}[X_1] \neq \mathbb{E}[X_2]) \text{ or} \\ &\quad (\gamma_1 \neq \gamma_2 \text{ and } \mathbb{E}[X_1] \neq \mathbb{E}[X_2]). \end{aligned} \tag{7}$$

Here, the null hypothesis states that the underlying collision rate and the expected number of casualties per collision are the same. The alternative hypothesis states that at least one of them are different. We propose to use the method proposed by Fisher [6] to determine a p-value for testing (7). Let the p-values obtained in Section 3 and 4 be denoted as p_1 and p_2 respectively. Then Fisher's test statistic is given by,

$$T_F = -2\log(p_1) - 2\log(p_2). \tag{8}$$

Under the null hypothesis, the test statistic follows a chi-squared distribution with four degrees of freedom, $T_F \sim \chi^2_4$, and hence a p-value for testing (7) can be easily obtained.

If the p-value obtained from testing (7) is sufficiently small, we are inclined to believe that there is a difference in at least one of the collision rates or first moments of the number of casualties per collision distribution. This suggests that there is some difference in at least one of the two underlying processes leading to casualties.

5 Worked example

Now that we have described our proposed methods for comparing the underlying collision and casualty rates, we will apply these methods to a fictitious example.

Suppose that the observed statistics for Road 1 and Road 2 are given in Table 1 and the histogram of the number of casualties per collision, the $X_{i,j}$, are as shown in Figure 4. Note that these statistics are based on fictitious data simply for this worked example.

Table 1: Observed statistics of Road 1 and Road 2 for this example. Recall that N_i, v_i, R_i, C_i, Q_i and $\bar{x}_i$ are the number of collisions, the road traffic, the collision rate, the number of casualties, the casualty rate, and the observed first moment of the number of casualties per collision respectively.

i	N_i	v_i	R_i	C_i	Q_i	$\bar{x}_i$
1	117	25	4.680	204	8.160	1.744
2	382	58	6.586	726	12.517	1.901

5.1 Comparing collision rates

We are first interested in determining whether there is sufficient evidence to conclude that the underlying collision rates of Road 1 and Road 2, γ_1 and γ_2, are different. First, we calculate confidence intervals for each collision rate, shown in Figure 2. The confidence intervals do not overlap which may suggest that the underlying collision rates for Road 1 and Road 2 are different. However, they are reasonably close so we formally test the following hypotheses by conducting the likelihood ratio test:

$$\begin{aligned} H_0 &: \gamma_1 = \gamma_2 \\ H_1 &: \gamma_1 \neq \gamma_2. \end{aligned} \tag{9}$$

Using Monte-Carlo simulation, as described in Section 3, we calculate the p-value as 0.001 (to 3 decimal places). The small p-value indicates that we have sufficient evidence to reject the null hypothesis and conclude that the evidence suggests the underlying collision rates for Road 1 and Road 2 are statistically different. In this case, the evidence suggests that the underlying collision rate for Road 2 is higher than the underlying collision rate of Road 1.

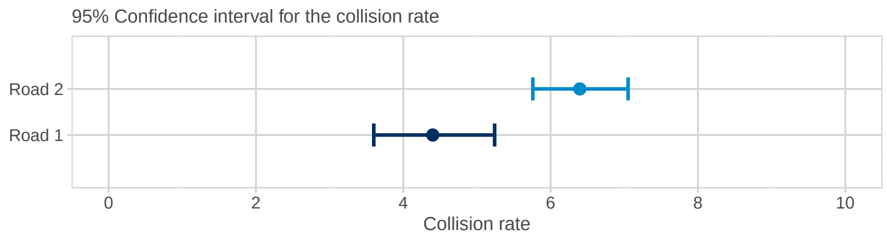


Figure 2: 95% Confidence intervals for each of the collision rates.

5.2 Comparing casualty rates

We are now interested in determining whether there is evidence to suggest that the underlying casualty rates are different for Road 1 and 2. Initially, we calculate confidence intervals for each casualty rate, shown in Figure 3. There is no overlap in the confidence intervals which may suggest that the underlying casualty rates are different. This seems plausible as we previously concluded the underlying collision rates are likely to be different, and casualties arise from collisions.

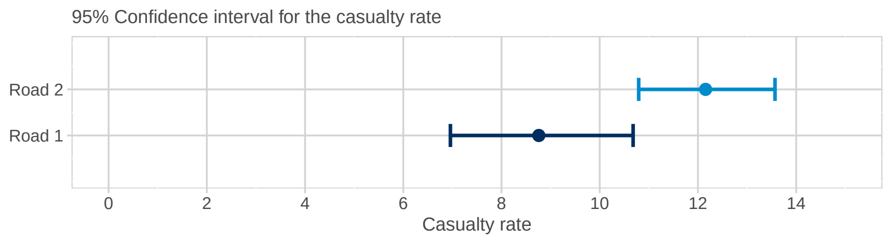
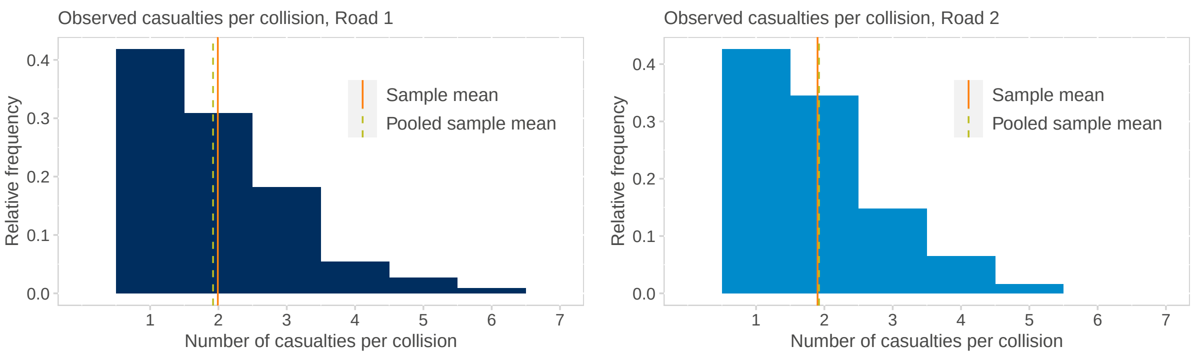


Figure 3: 95% Confidence intervals for each of the casualty rates.

We now formally test for a difference in the first moments of the casualties per collision distributions. Before conducting a hypothesis test on the first moments, we inspect the observed number of casualties per collision for outliers. These could adversely affect the conclusions of the hypothesis test. By eye, inspecting the histograms in Figure 4, the number of casualties per collision appear similar for Road 1 and Road 2 and without any large outliers. However, Road 2 does have a relatively small number of collisions that result in five and six casualties, whereas all collisions for Road 1 result in fewer than five casualties.



(a) Road 1

(b) Road 2

Figure 4: Histograms of the number of casualties per collision for Road 1 and Road 2. The solid orange vertical lines indicate the mean number of casualties per collision for each road and the dashed yellow vertical lines indicate the pooled sample mean, where the mean is calculated using the number of casualties per collision for both roads.

We test the following hypotheses using the non-parametric test described in Section 4,

$$\begin{aligned} H_0 &: \mathbb{E}[X_1] = \mathbb{E}[X_2] \\ H_1 &: \mathbb{E}[X_1] \neq \mathbb{E}[X_2]. \end{aligned} \tag{10}$$

The p-value is 0.112 (to 3 decimal places). This relatively large p-value suggests that there is not enough evidence to reject the null hypothesis. We cannot conclude that there is a difference in the first moment of the casualties per collision for Road 1 and Road 2.

Finally, we can combine the p-values as discussed in Section 4. The combined p-value obtained is 0.001 (to 3 decimal places). This also helps us to conclude that there is a difference in at least one of the underlying collision rates or the first moment of the casualties per collision distributions.

Overall, all of this evidence taken together supports the conclusion that the casualty rates for Road 1 and Road 2 are different.

6 Guidance

In this section we provide some general guidance for using the methods proposed in this document. We cover a broad range of topics that consider how the data is collected, anomalous values that may affect the conclusions drawn, and how to ensure that conclusions are not misleading.

Often, statistical methods cannot be simply used 'out of the box', but the practitioner must make decisions that could affect the conclusions of the analysis. This is the case for these proposed methods, and examples of such decisions will be discussed in more detail in Sections 6.2, 6.3, and 6.4.

6.1 Observing a low number of collisions

The methods proposed in this document for comparing collision rates have been selected to perform favorably in situations where few collisions have been observed. However, there are still some cases where the proposed collision rate methods cannot be used as presented in this document.

Consider the example where zero collisions are observed on Road 1. Here, the estimate of the underlying collision rate for Road 1 will be zero. The probability mass function for the Poisson distribution with rate equal to zero has point mass at zero, meaning that the corresponding random variable will always take the value zero and a confidence interval for the collision rate of Road 1 cannot be estimated. Providing at least one collision is observed on Road 2, confidence intervals for $\hat{\gamma}_2$ can be calculated to determine how close the interval lies to zero. Further, conducting the hypothesis test will formalise the comparison between the collision rates of Road 1 and Road 2 despite observing zero collisions on Road 1.

When zero collisions are observed, we cannot calculate confidence intervals for the casualty rate for the same reasons described above. We also cannot conduct a hypothesis test on the first moment for the number of casualties per collision as no information will exist for this statistic. Further, when the observed number of collisions is only small, we have little information about the distribution of the number of casualties per collision. Therefore, when a low number of collisions has been observed it is likely to not be appropriate to use the methods as presented in Section 4 or, if used, the conclusions should be treated with additional caution.

6.2 Outliers in the number of casualties per collision

Sometimes the data may contain unusually large observations for the number of casualties per collision. These outliers should be investigated. Suppose a high number of casualties per collision were observed as a consequence of an extreme weather event. Then, the estimate of the casualty rate may be biased by this event. It should be carefully considered whether some feature of the road could have reduced or increased the effects of this event, and what the purpose of the analysis is.

Depending on the conclusions drawn, it may be or may not be appropriate to remove outliers. Consider the following example.

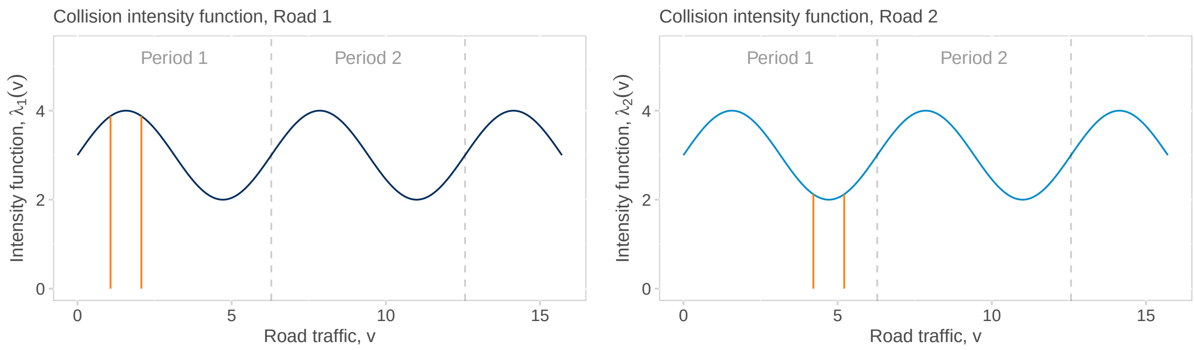
Suppose a high number of casualties were observed as a result of heavy snowfall. If the road was at a high altitude, then high snowfall events may be common. While it may not be possible to change the altitude of a road, there may be some other interventions that could prevent collisions that could reduce the casualty rate, such as road closures for example. Here, removing the outliers could lead to missed opportunities to improve safety. Conversely, if the road was not at a high altitude and the heavy snowfall was a one-off event, and the analysis was not aiming to understand the safety impact of the snowfall, then removing the outliers may be appropriate.

We acknowledge that we cannot describe all possible scenarios. In scenarios where it is not clear, it may be sensible to conduct the analysis with and without the outliers.

6.3 Data collection time period

We assume that the collisions occur according to a non-homogeneous Poisson process. Ultimately, this allows us to assume that the collision rate can vary over time, which we have observed in data during the development of these methods. It also seems to be a reasonable assumption as there are many temporal factors, such as the weather and daylight hours, that may impact collisions. However, this means that the time intervals that we compare should be thought about carefully. Ideally, they should be the same. Consider the following examples.

Figure 5 shows an example where the collision intensity functions for Road 1 and Road 2 are identical, but the intensity functions are periodic. However, here we compare data for Road 1 and Road 2 collected at different times within the period. In particular, the data for Road 1 is collected during a time when the collision intensity function for Road 1 is higher than Road 2. Consequently, it is likely that the data will suggest that the underlying collision rate for Road 1 is higher than that for Road 2. This could be misleading if it is not made clear that the two datasets were collected over different times.

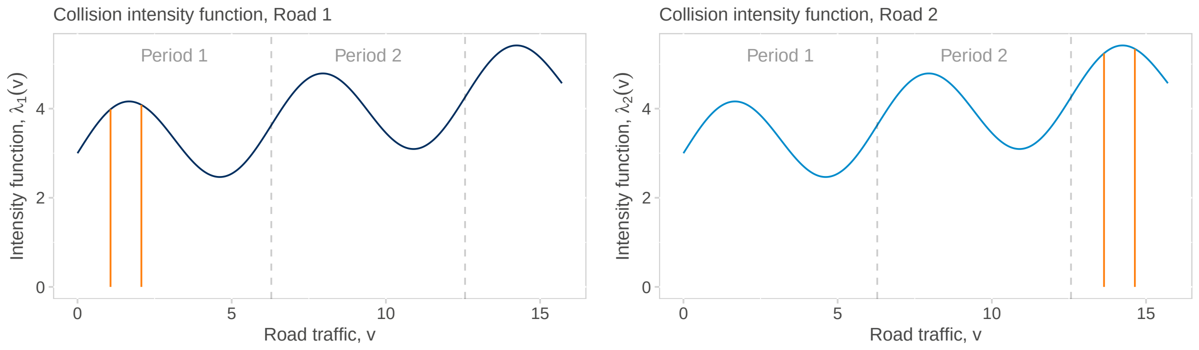


(a) Road 1

(b) Road 2

Figure 5: Collision intensity function for Road 1 and Road 2. The solid vertical (orange) lines indicate the time periods of data collection. The dashed vertical lines indicate when the behaviour of the collision intensity function repeats.

Figure 6 shows another example where the collision intensity functions for Road 1 and Road 2 are identical. Here, the collision intensity functions have periodic behaviour and trend. Whilst the data is collected during the same time within each period, in the sense that they both correspond to times where the collision rates are highest within each period, the trend in the collision rate intensity function could again produce misleading conclusions.



(a) Road 1 (b) Road 2

Figure 6: Collision intensity function for Road 1 and Road 2. The solid vertical (orange) lines indicate the time periods of data collection. The dashed vertical lines indicate when the behaviour of the collision intensity function repeats.

Given these examples, we suggest that the methods proposed in this document are only used when the data used to compare two roads is collected over the same period of time, or a single road between two time periods of equal duration and position in the year. If this is not possible, it should be made clear in the conclusions that the time periods are not the same and the possible implications of this taken into account.

6.4 Sensitivity analysis

It is important to consider the quality of the data being used to draw conclusions. For example, it is possible that collisions occur that are not reported to the police and therefore are not included in STATS19. It is also possible that road traffic estimates are inaccurate, as the road network is a complex system which spans a large geographical area. Biases may exist in any dataset so when comparing statistics between two roads or road types, you should consider whether it is likely that any bias is inherent to feature being compared (road or road type).

A sensitivity analysis could be conducted to consider how the collision and casualty rates change, along with the associated confidence intervals and p-values, when the collision, casualty, and road traffic estimates are varied by some percentage. For example, if the analysis suggests that there is a difference in the underlying collision and casualty rates, then a sensitivity analysis would determine the percentage change in road traffic that leads to the opposite conclusions with similar p-values, or a much smaller or larger p-value.

6.5 How to interpret p-values

When undertaking a formal hypothesis test, we begin by assuming that the null hypothesis is true and then calculate a p-value. A p-value is defined as the probability of getting a test statistic at least as extreme as that observed, if the null hypothesis was really true. Since the p-value is a probability, it will lie between zero and one.

A p-value close to zero suggests that it is unlikely to observe a test statistic as extreme as the one we observed given the null hypothesis is true. So we have evidence to reject the null hypothesis and can conclude it is unlikely that the null hypothesis is true. Conversely, a p-value close to one suggests that it is not unusual to observe a test statistic like the one we have given the null hypothesis is true. Here, we do not have evidence to reject the null hypothesis but we cannot use the analysis as evidence that the null hypothesis is true. The null hypothesis is simply the working assumption, and hypothesis testing only allows us to look for evidence to reject it rather than accept it as truth.

In our application, when comparing collision and casualty rates between roads, the null hypothesis is that the underlying collision or casualty rates are the same. We guide users of these proposed methods to be careful to not interpret a large p-value as 'proof' that this is the truth.

We have talked about p-values being close to zero or close to one, but what do we mean by close? Traditionally, when undertaking a formal hypothesis test, practitioners have compared the p-value to a threshold of 0.05. From this, they have either rejected the null hypothesis when the calculated p-value is smaller than the threshold or not when it is larger. However, it is now widely established that using a, somewhat arbitrary, threshold to draw binary conclusions is not appropriate and that a p-value of 0.049 should not lead to such a different conclusion as a p-value of 0.051. [16, 17] We recommend reporting p-values as they are calculated, rather than only in comparison to a threshold, and interpreting them on a continuous scale from zero to one.

6.6 Practical significance

The methods proposed in this document can be used to quantify the amount of evidence supporting statistical differences, but they cannot be used to determine whether the size of the differences are meaningful. Small differences can be statistically different, but this does not mean that they are practically different or important, or that the difference is large enough to justify interventions. The size of any differences, and the size of the difference relative to the rates, should be considered carefully.

It is also important to remember that the size of a p-value is relative to the data collection period. For example, if the underlying collision rates for two roads are fixed, on average, we'd expect the p-value to decrease in size as the data collection period increases. When small p-values are not observed, the expense, the time needed, and the practical importance of the currently observed differences should be carefully considered to determine whether it is suitable to continue collecting data to be used as evidence.

6.7 Other uses

In the Introduction we described that we have developed these methods to be used to compare collision and casualty rates derived from data collected using the STATS19 data collection system. We anticipate them being useful to compare rates between different geographical areas (such as between regions, roads, road types and road sections) and between different time periods of a single geographical area (such as comparing one year to the next). It may be suitable to also apply these methods to other rate statistics, but we have not formally reviewed this yet.

The statistical model we propose for the number of collisions is a non-homogeneous Poisson process. If, in some alternative application, it is still reasonable to assume that events occur according to a non-homogeneous Poisson process, then it might be suitable to apply the methods described in Section 3 to these statistics. Similar considerations apply to the methods described in Section 4 and the compound Poisson process assumption. One example that may be suitable, given a suitable data source, is the rate of breakdowns given by dividing the number of breakdowns that occur by the vehicle miles for a given road. The rate at which breakdowns occur may vary over time, the breakdowns may be independent, and a single breakdown should only occur at one time (when measured by an infinitely small period of time). However, the suitability of the methods to this application still needs to be thoroughly investigated.

There may also be applications where these methods cannot be applied as presented in this report. Consider using the methods developed to compare casualty rates for severity-adjusted STATS19 casualty data.[2] In particular, consider comparing adjusted slight or serious casualty rates between two roads. Here, an alternative definition of casualty rate would be needed, as casualties are not either *slight* or *serious* in the adjusted data. Instead, the adjustments are estimates of how casualty severity may have been recorded prior to the change in the STATS19 reporting system. In this case, we are unable to analyse, for example, the distribution of casualties per collision as we propose.

Comparing daylight collision rates against nighttime collision rates may be another example where the methods cannot be used as presented here. In this example, the collision process is no longer observed continuously due to the data collection period alternating between daylight and nighttime over the data collection period. It may be possible to amend the likelihood function given in Section 3.1 to address this, but this would need further investigation.

Finally, we have presented methods for comparing rates for two roads. It is possible to extend these methods to compare rates among more than two roads. Details are given in Appendix C.

[2] https://www.gov.uk/government/publications/guide-to-severity-adjustments-for-reported-road-casualty-statistics/guide-to-severity-adjustments-for-reported-road-casualties-great-britain

7 Conclusions and next steps

In this document we have described statistical methods that we propose can be used to compare road traffic collision rates and casualty rates derived from STATS19 data. The methods we propose to compare underlying collision rates build on statistical assumptions similar to those already used by the DfT and National Highways. However, the method we propose can be used irrespective of the level of road traffic observed.

To the best of our knowledge, a method to compare casualty rates did not exist. We showed how the casualty rate depends on both the collision rate and the expected number of casualties per collision. Therefore, we propose to test for a difference in casualty rates by testing for both a difference in the underlying collision rates and a difference in the underlying first moments of the distributions for the number of casualties per collision. The conclusions of those two pieces of analysis can then be used to determine whether the evidence suggests that there is a difference in the underlying casualty rates.

We have included a worked example to demonstrate these methods, and in the Guidance section we have described some factors to take into account when using these methods.

While we have rigorously and carefully developed these proposed methods, they are not yet finalised. Before establishing these methods as best practice across National Highways, we welcome feedback and advice. In particular, we are keen to hear from you if you uncover a problem with our proposed methods or can suggest an alternative approach which would be a substantial improvement. Feedback can be given to networkanalysisandstatistics@nationalhighways.co.uk by the end of August 2022, after which the methods will be updated if necessary and finalised. In the meantime, these proposed methods will be trialled in a comparison of collision rates and casualty rates between different motorway types. This will test these proposed methods being applied to real data.

A Appendix: Methods for comparing collision rates

This section describes the methods proposed to compare collision rates.

A.1 Algorithm 1: confidence interval

A parametric bootstrap method for calculating a confidence interval for the underlying collision rate. For Road i:

1. Calculate $\hat{\gamma}_i = \frac{N_i}{v_i}$
2. For $s = 1, \ldots, S$ where S is sufficiently large
   (a) Simulate $n_s \sim$ Poisson$(\hat{\gamma}_i v_i)$
3. Order the n_s such that $n_{(1)} \leq n_{(2)} \leq \ldots \leq n_{(S)}$
4. The $100\alpha\%$ confidence interval is given by

$$\left[\frac{n_{(\frac{S\alpha}{2})}}{v_i}, \frac{n_{(S(1-\frac{\alpha}{2}))}}{v_i}\right]. \tag{11}$$

A.2 Calculating a p-value to compare two collision rates

To test hypotheses 4 we construct the likelihood ratio test statistic:

$$t = 2\left(\log(L(\hat{\gamma}_1, \hat{\gamma}_2)) - \log(L(\tilde{\gamma}, \tilde{\gamma}))\right), \tag{12}$$

where $\tilde{\gamma} = \frac{N_1+N_2}{v_1+v_2}$ is the maximum likelihood estimate of the collision rate under H_0 and $\hat{\gamma}_i = \frac{N_i}{v_i}$ for $i = 1, 2$, is the maximum likelihood estimate for the (different) underlying collision rates under H_1. The p-value can be estimated as follows:

1. Calculate $\tilde{\gamma} = \frac{N_1+N_2}{v_1+v_2}$
2. For $s = 1, \ldots, S$ where S is sufficiently large
   (a) Simulate $n_1 \sim$ Poisson $(\tilde{\gamma} v_1)$ and $n_2 \sim$ Poisson $(\tilde{\gamma} v_2)$
   (b) Calculate $\tilde{\gamma}_s = \frac{n_1+n_2}{v_1+v_2}$
   (c) Calculate $\hat{\gamma}_{1,s} = \frac{n_1}{v_1}$ and $\hat{\gamma}_{2,s} = \frac{n_2}{v_2}$
   (d) Calculate $t_s = 2\left(\log(L(\hat{\gamma}_{1,s}, \hat{\gamma}_{2,s})) - \log(L(\tilde{\gamma}_s, \tilde{\gamma}_s))\right)$
3. Calculate $\hat{\gamma}_1 = \frac{N_1}{v_1}$ and $\hat{\gamma}_2 = \frac{N_2}{v_2}$
4. Calculate $t = 2\left(\log(L(\hat{\gamma}_1, \hat{\gamma}_2)) - \log(L(\tilde{\gamma}, \tilde{\gamma}))\right)$
5. Calculate $p = \frac{1}{S}\sum_{s=1}^{S} \mathbb{I}_{t_s \geq t}$.

Here, $\mathbb{I}_{t_s \geq t} = 1$ if $t_s \geq t$ and 0 otherwise. We propose to use $S = 1 \times 10^6$. To determine if S is large enough, the test can be run a second time to ensure that the p-value returned is the same to three decimal places.

B Appendix: Methods for comparing casualty rates

This section describes the methods proposed to compare casualty rates.

B.1 Algorithm 2: confidence interval

A parametric bootstrap method to calculate confidence intervals for the underlying casualty rates. For Road i:

1. Calculate $\hat{\gamma} = \frac{N_i}{v_i}$
2. For $s = 1, \ldots, S$ where S is sufficiently large
   (a) Simulate $n \sim$ Poisson$(\hat{\gamma} v_i)$
   (b) Sample $\{Y_1, \ldots, Y_n\}$ from $\{X_{i,1}, \ldots, X_{i,N_i}\}$ with replacement
   (c) Calculate $\zeta_s = \frac{\sum_{j=1}^{n} Y_j}{v_i}$
3. Order the ζ_s such that $\zeta_{(1)} \leq \zeta_{(2)} \leq \ldots \leq \zeta_{(S)}$
4. Calculate the $100\alpha\%$ confidence interval of the casualty rate of Road i as

$$\left[\zeta_{(\frac{S\alpha}{2})}, \zeta_{(S(1-\frac{\alpha}{2}))}\right]. \tag{13}$$

B.2 Calculating a p-value for first moment

The bootstrap procedure proposed to calculate a p-value for testing Hypotheses 6 is as follows:

1. Calculate $d = |\frac{1}{N_1} \sum_{j=1}^{N_1} X_{1,j} - \frac{1}{N_2} \sum_{j=1}^{N_2} X_{2,j}|$
2. For $s = 1, \ldots, S$ where S is sufficiently large
   (a) Randomly select $\mathcal{A} = \{A_1, \ldots, A_{N_1}\}$ from $\mathcal{X} = \{X_{1,1,}, \ldots, X_{1,N_1}, X_{2,1}, \ldots, X_{2,N_2}\}$ and let $\{B_1, \ldots, B_{N_2}\} = \mathcal{X} \quad \mathcal{A}$
   (b) Calculate $d_s = \left|\frac{1}{N_1} \sum_{j=1}^{N_1} A_j - \frac{1}{N_2} \sum_{j=1}^{N_2} B_j\right|$
3. Calculate $p = \frac{1}{S} \sum_{s=1}^{S} \mathbb{I}_{d_s \geq d}$

where $\mathbb{I}_{d_s \geq d} = 1$ if $d_s \geq d$ and 0 otherwise.

C Appendix: Analysis for comparing more than two roads

Suppose interest lies in comparing the collision rates between M roads. Let N_i, v_i and γ_i denote the number of observed collisions, road traffic, and the expected number of collisions per 100 million vehicle miles for Road i where $i = 1, \ldots, M$. Then considering the following hypotheses:

$$\begin{aligned} H_0 &: \gamma_1 = \gamma_2 = \ldots = \gamma_M \\ H_1 &: \gamma_1 \neq \gamma_2 \neq \ldots \neq \gamma_M, \end{aligned} \tag{14}$$

we can estimate the different collision rates under H_0 and H_1. Note, under the alternate hypothesis, each of the collision rates can assume a different value. If the null hypothesis is rejected, we cannot say for sure that each of the underlying collision rates takes a different value, just that there is some difference amongst them. The likelihood function for the data is:

$$L(\gamma_1, \gamma_2, \ldots, \gamma_M) = \frac{e^{-\gamma_1 v_1}(\gamma_1 v_1)^{N_1}}{N_1!} \frac{e^{-\gamma_2 v_2}(\gamma_2 v_2)^{N_2}}{N_2!} \ldots \frac{e^{-\gamma_M v_M}(\gamma_M v_M)^{N_M}}{N_M!}. \tag{15}$$

Under H_0, the collision rate for all roads is given by $\tilde{\gamma} = \frac{\sum_{i=1}^{M} N_i}{\sum_{i=1}^{M} v_i}$ which can be obtained from maximising $L(\tilde{\gamma}, \tilde{\gamma}, \ldots, \tilde{\gamma})$. Under H_1, the collision rates are given by $\hat{\gamma}_i = \frac{N_i}{v_i}$ for $i = 1, \ldots, M$ by maximising $L(\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_M)$. Confidence intervals for the collision rates can be calculated using Algorithm 1. The p-value for testing (14) can be calculated by substituting $L(\tilde{\gamma}, \tilde{\gamma})$ with $L(\tilde{\gamma}, \tilde{\gamma}, \ldots, \tilde{\gamma})$ and $L(\hat{\gamma}_1, \hat{\gamma}_2)$ with $L(\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_M)$ in the method described in Section 3.

Similarly, the expected number of casualties per collision can be compared across M roads. Let X_i denote the random variable for the number of casualties per collision for Road i for $i = 1, \ldots, M$. Then, the hypotheses considered here are:

$$\begin{aligned} H_0 &: \mathbb{E}[X_1] = \mathbb{E}[X_2] = \ldots = \mathbb{E}[X_M] \\ H_1 &: \mathbb{E}[X_1] \neq \mathbb{E}[X_2] \neq \ldots \neq \mathbb{E}[X_M]. \end{aligned} \tag{16}$$

Consider the null hypothesis where $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \ldots = \mathbb{E}[X_M]$. Then, under the null hypothesis, if a large number of collisions were observed for each road, the sample mean of each road should be close to the value of the mean casualties per collision of all roads. Therefore, we calculate a p-value for testing (16) by considering the distribution of the following test statistic:

$$d = \frac{1}{M} \sum_{i=1}^{M} \left| \frac{\sum_{i=1}^{M} \sum_{j=1}^{N_m} X_{i,j}}{\sum_{i=1}^{M} N_i} - \frac{1}{N_m} \sum_{j=1}^{N_m} X_{i,j} \right|. \tag{17}$$

The p-value is calculated as follows:

1. Calculate d
2. For $s = 1, \ldots, S$, where S is sufficiently large

(a) Randomly select $\tilde{\mathcal{X}}_1 = \{\tilde{X}_{1,1}, \dots, \tilde{X}_{1,N_1}\}$ from $\mathcal{X} = \{X_{1,1,} \dots, X_{1,N_1}, \dots, X_{M,1}, \dots, X_{M,N_M}\}$ without replacement and $\tilde{X}_i = \{\tilde{X}_{i,1}, \dots, \tilde{X}_{i,N_i}\}$ from $\tilde{X} \quad \bigcup_{m=1}^{i-1} \tilde{X}_m$ without replacement for $i = 2, \dots, M$.

(b) Calculate $d_s = \frac{1}{M} \sum_{i=1}^{M} \left| \frac{\sum_{i=1}^{M} \sum_{j=1}^{N_m} X_{i,j}}{\sum_{i=1}^{M} N_i} - \frac{1}{N_m} \sum_{x \in \mathcal{X}_m} x \right|$

3. Calculate $p = \frac{1}{S} \sum_{i=1}^{S} \mathbb{1}_{d_i \geq d}$.

Confidence intervals for the casualty rates can be obtained by modifying Algorithm 2.

In addition to the definition of d given in (17), we considered an alternative, $\tilde{d}$ where:

$$\tilde{d} = \sup_i \left| \frac{\sum_{i=1}^{M} \sum_{j=1}^{N_m} X_{i,j}}{\sum_{i=1}^{M} N_i} - \frac{1}{N_m} \sum_{j=1}^{N_m} X_{i,j} \right|. \tag{18}$$

It is thought that $\tilde{d}$ would be a suitable choice if there was concern that there exists a large difference in the mean number of casualties per collision for a single road. In contrast, d should capture variation in the mean number of casualties per collision.

References

[1] *Basic Large Sample Theory*, pages 419–481. Springer New York, New York, NY, 2005.

[2] *Uniformly Most Powerful Tests*, pages 56–109. Springer New York, New York, NY, 2005.

[3] Dankmar Böhning and Peter G. M. van der Heijden. The identity of the zero-truncated, one-inflated likelihood and the zero-one-truncated likelihood for general count densities with an application to drink-driving in Britain. *The Annals of Applied Statistics*, 13(2):1198 – 1211, 2019.

[4] Frederik Michel Dekking, Cornelis Kraaikamp, Hendrik Paul Lopuhaä, and Ludolf Erwin Meester. *The bootstrap*, pages 269–284. Springer London, London, 2005.

[5] Frederik Michel Dekking, Cornelis Kraaikamp, Hendrik Paul Lopuhaä, and Ludolf Erwin Meester. *Maximum likelihood*, pages 313–327. Springer London, London, 2005.

[6] R. A. Fisher. *Statistical Methods for Research Workers*, pages 66–70. Springer New York, New York, NY, 1992.

[7] Department for Transport. Testing for statistically significant changes in road accident and casualty numbers. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/995114/testing-for-statistically-significant-changes.pdf.

[8] National Highways. POPE Methodology Manual: Post-opening project evaluation for major projects. https://nationalhighways.co.uk/media/exypgk11/pope-methodology-note-jan-2022.pdf.

[9] K. Krishnamoorthy and Jessica Thomson. A more powerful test for comparing two poisson means. *Journal of Statistical Planning and Inference*, 119(1):23–35, 2004.

[10] Günter Last and Mathew Penrose. *Compound Poisson Processes*, pages 153–165. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2017.

[11] Günter Last and Mathew Penrose. *Poisson Processes on the Real Line*, pages 58–68. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2017.

[12] Hon Keung Tony Ng, K. Gu, and Man-Lai Tang. A comparative study of tests for the difference of two poisson means. *Computational Statistics & Data Analysis*, 51(6):3085–3099, 2007.

[13] Office of Road and Rail. Quality assurance of all lane running motorway data. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1015501/orr-quality-assurance-of-all-lane-running-motorway-data.pdf.

[14] J. Przyborowski and H. Wilenski. Homogeneity of results in testing samples from poisson series: With an application to testing clover seed for dodder. *Biometrika*, 31(3/4):313–323, 1940.

[15] David Spiegelhalter. *The Art of Statistics: Learning from Data*. Penguin Books Ltd, London, 2019.

[16] Ronald L. Wasserstein and Nicole A. Lazar. The asa statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2):129–133, 2016.

[17] Ronald L. Wasserstein, Allen L. Schirm, and Nicole A. Lazar. Moving to a world beyond ”p < 0.05”. *The American Statistician*, 73(sup1):1–19, 2019.

If you need help accessing this or any other National Highways information, please call **0300 123 5000** and we will help you.

This document is also available on our website at **www.nationalhighways.co.uk**

For an accessible version of this publication please call **0300 123 5000** and we will help you.

If you have any enquiries about this publication email **info@highwaysengland.co.uk** or call **0300 123 5000***. Please quote the National Highways publications code **PR81/22**.

National Highways creative job number BRS22_0016

*Calls to 03 numbers cost no more than a national rate call to an 01 or 02 number and must count towards any inclusive minutes in the same way as 01 and 02 calls.

These rules apply to calls from any type of line including mobile, BT, other fixed line or payphone. Calls may be recorded or monitored.

Printed on paper from well-managed forests and other controlled sources when issued directly by National Highways.

Registered office Bridge House, 1 Walnut Tree Close, Guildford GU1 4LZ

National Highways Limited registered in England and Wales number 09346363