RoSPA
accidents don't have to happen

Cycle Smart Brum Project
Summary Report

Department for Transport
Produced with the support of the Department for Transport

SEE.SENSE®

Contents

Introduction

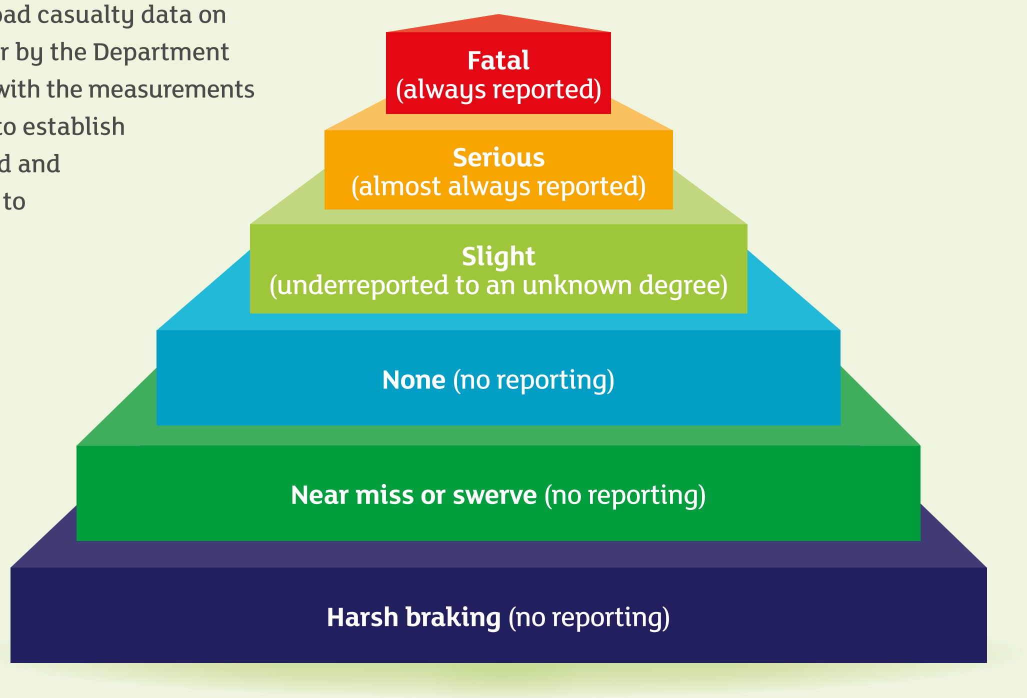
In 2019, 100 cyclists were killed, 4,333 were seriously injured and a further 12,451 were slightly injured in reported collisions on Great Britain's roads. These figures reflect the number of personal injury accidents reported to the police in 2019. Incidents that do not result in a fatality or serious injury are often unreported. For each reported collision, it can be considered that there are many unreported collisions that have taken place (DfT, 2019).

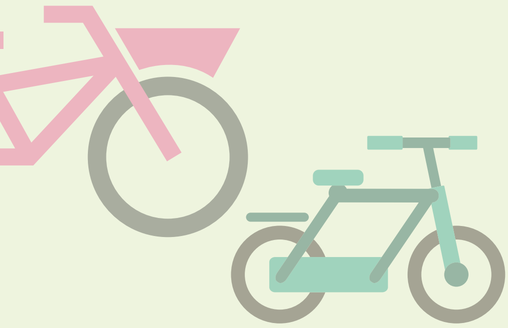
To gain a better understanding of reported and unreported cycling collisions, near misses and the conditions under which they occur, The Royal Society for the Prevention of Accidents (RoSPA), a leading accident prevention charity, and See.Sense, a cycling tech and data startup, have undertaken Cycle Smart Brum, an innovative, data-driven project.

The project examined reported road casualty data on pedal cycles, published each year by the Department for Transport, and compared this with the measurements collected by the See.Sense light to establish the relationship between reported and unreported incidents with a view to developing a predictive model.

It has been theorised that if the number of collisions of a lower severity are reduced, this will have a direct impact on reducing higher severity collisions too (Heinrich, 1931). The generation of a predictive model allows us to understand cycling conditions "on the ground" and determine the areas in which collisions are likely to occur. Understanding of the lower levels of the collision triangle fills a significant gap in our cycle safety knowledge and offers potential progress towards reducing injury rates across all levels of severity.

Through the creation of this predictive model, our aim is that data insights can be used to understand collisions as symptomatic of their environment and road design rather than the attribution of blame to an individual cyclist.





Assumptions

For this study, the STATS19 data from 2016 to 2019 was used. The following assumptions were made:

- STATS19 is generally considered to have a Poisson distribution - thus the events occur at random, with the probability of an event occurring being dependent on the underlying rate of occurrence (not on how long it has been since a previous event, nor upon the number of events that have occurred in a recent period).
- STATS19 location data is subject to error or interpretation by the recording officer. We have assumed that the location given is accurate and is the point of the initial collision.
- Changes to road design are infrequent, therefore despite the STATS19 reporting period and the cyclist data collection period being different, the vast majority of the area under study will not have experienced any change.
- The cyclists recruited for the project were regular commuter cyclists pre-COVID19. They are generally experienced and frequent cyclists. Motor traffic levels have also reduced due to COVID19. Therefore the probability of an event is perhaps lower than pre-COVID19.
- The cyclist's behaviour we measured was considered to be a result of poor environmental factors, such as compromises in road design. We are not indicating that the cyclist's behaviour is always a factor in STATS19 incidents.

SEE.SENSE®

Methodology

Data Collection

See.Sense intelligent and connected bike lights use patented technology, edge processing and artificial intelligence (AI), to generate highly granular **sensor** data, which is collated by a companion app before being sent to the See.Sense Data Lake. Use of sensor data has the advantage in that it is passively collected, removing bias of perception of the user. Qualitative **perception data** was also collected using the app paired with the lights, whereby participants could report location of perceived close pass, near-miss, and obstructions and other concerns, and provide commentary.

The Cycle Smart Brum project distributed See.Sense ACE rear bike lights to 196 volunteer participants in Birmingham. Cyclists were selected on the basis of regularly riding in the central Birmingham area. In exchange for a heavily discounted light, cyclists agreed to participate in the project by sharing aggregated sensor data collected by the light, as well as providing qualitative data in the form of in-app post-ride surveys.

Data collection commenced in June 2020 through to December 2020. Data from a total of 235 cyclists was used (including both project participants and See.Sense retail customers who have opted in to share data). Over the period of the study, 42,161 km travelled, representing 798,292,700 individual sensor readings for each characteristic.

Due to the impact of COVID-19, fewer participants commuted to work during the project period, therefore we extended our area of study from the city centre to a wider area, based on the most frequently cyclied areas. A bounding box of 9.5km wide and 10km tall allowed us to get a good density of cyclists and cycling journeys.

Type of Data Collected

It is important to understand what data the See.Sense ACE measures and how it is processed. This is primarily telemetry information, including location, speed and acceleration. We also include the rate of change of acceleration, known as the 'jerk'. Thus the standard units we use are:

1 POSITION **the location of an object**

2 VELOCITY **the rate of change of Position**

3 ACCELERATION **the rate of change of Velocity**

4 JERK **the rate of change of Acceleration**

Position, velocity and acceleration are generally well understood. The Jerk is more easily understood if we consider the example of braking in a car. If the brakes are applied very suddenly, as in an emergency stop, the driver and passengers will be thrown forwards due to the high rate of change of deceleration, or high Jerk value. Conversely it is possible to apply the brake progressively, to achieve the same deceleration, without throwing forwards the occupants. This is because the rate of deceleration or Jerk is a smaller value over a longer period.

In the case of a cyclist, braking (deceleration) will be seen in many places, such as roundabouts and road junctions. Very high values of deceleration may be interesting, but as cyclists seldom brake as hard as the bike will allow (when the front brake is applied hard enough so that the rear wheel is just about to lift off) the peak value will vary significantly per cyclist.

Braking jerk, or how quickly the brakes are applied, is much more useful. The See.Sense ACE profiles each individual rider to determine what their normal brake jerk value is. It then compares the jerk value to what is normal for that cyclist and records a value relative to that normal value. Generally this means that an experienced rider, who looks further ahead and anticipates more will have a lower average jerk, while a beginner cyclist will be more abrupt in their inputs and have a higher average jerk. Up to 800 acceleration readings per second are used to calculate the jerk value and the result is stored at one second intervals. The key benefit of this approach is that by profiling each cyclist, the jerk dataset is standardised, with individual variances being removed.

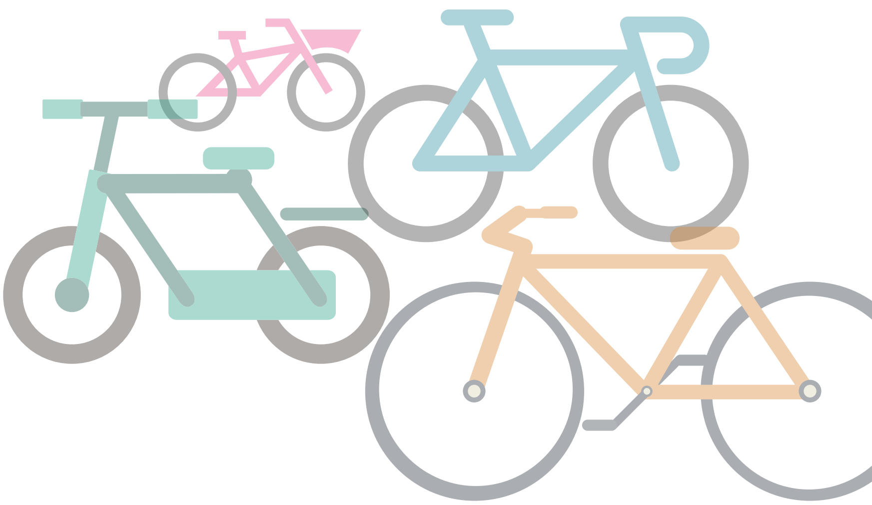
The swerve and swerving jerk is similar. A cyclist travelling quickly around a roundabout will experience a high swerving force (centripetal acceleration). However, the rate of swerve (jerk) as they steer and lean into the manoeuvre is small. The swerve jerk value is high where the cyclist makes a sudden swerve, such as when avoiding an obstacle or pothole with little warning.

The data collected from See.Sense light and app is geospatially located, including speed, dwell time, as well as proprietary data fields as follows:

- **Swerve jerk** - this is relative to the cyclist's average value such that normal behaviour scores close to zero and the more extreme the swerve behaviour, the higher the score that is generated.
- **Braking jerk** - like swerving, we compare the normal behaviour of the cyclist to their current behaviour to determine how their current behaviour compares. The faster the brakes are applied, the higher the score.
- **Surface** - we process sensor data, taking into consideration the different characteristics, such as bike type, frame material and tyre width and pressure, to create a standardised score for the road surface. This provides an indication of the road surface roughness at a highly granular scale. It does not generally indicate large defects, such as potholes, which cyclists will typically steer around. This score has been validated with AECOM, verifying that it has comparable differential capability to a qualified visual inspection.

Each of the data fields above has 1 second granularity and is an integer which ranges from 0 (representing normal riding or a good surface) to approximately 100 (representing an extreme swerve or brake event, or a very poor road surface).

Further qualitative data was collected using Facebook. At the beginning of the project, a private Cycle Smart Brum Facebook group was set up for participants, where they could interact and have discussions about their riding experiences around central Birmingham. These posts, combined with the feedback from the post-ride surveys and light data, were useful for finding areas of concern in Birmingham. Surveys were also held throughout the project, and a focus group was held towards the end of the project.



Case Study Approach

When combining the See.Sense and STATS19 data using mapping software, we are able to develop a clear visualisation of where reported accidents have occurred and the overlap with the See.Sense data assessment in many areas.

Towards the end of the project, See.Sense carried out quantitative analysis of the cycle light data to investigate the extent to which the data coverage overlapped with STATS19 incidents. It was found that there were five locations covered by the light data that had multiple STATS19 incidents with exactly the same latitude and longitude. These five 'case studies' of potentially highly dangerous areas were then investigated, using a range of sources from the project and beyond, including STATS19 dataset attributes, Google Streetview, post-ride surveys, and knowledge of local Birmingham infrastructure.

Data Analysis Approach

The project examined reported road casualty data on pedal cycle collisions, published from the period 2016 to 2019 (3 years inclusive) by the Department for Transport, and compared this with the data collected by the See.Sense lights. With nearly 800 million data points for each of Swerving, Braking and Surface, it was not practical to consider each individually. We therefore split the map into 10 metre squares in order to create 172,471 aggregate data collections, which can be mapped to visualise and analyse the data.

When we consider the contributing factors to a STATS19 event, it's possible that the indicators, such as swerving or braking could occur prior to the location recorded for the event. However, using a 10x10m tile, gives us a reasonable balance between distance covered from the event without too high a risk of including false positives - for example, we are not likely to include indicators from neighbouring roads, particularly at junctions where various routes converge.

Findings

In order to ensure a sufficient data density, we only included tiles with 3 or more cyclists and 15 or more journeys. At this level of coverage, we were able to study 453 (81%) out of 559 STATS19 events. The image below shows the total coverage area and STATS19 locations.

Due to the initial selection criteria targeting city centre cyclists, we had less cycling coverage in the North and East of the study area.

Since not every cycling journey will include an event causing the cyclist to react abnormally, we compared the normal behaviour (across the whole network) against the maximums seen when abnormal behaviour was observed in each tile. We can clearly see that for any given measure, we see significantly larger values in tiles with recorded STATS19 events.

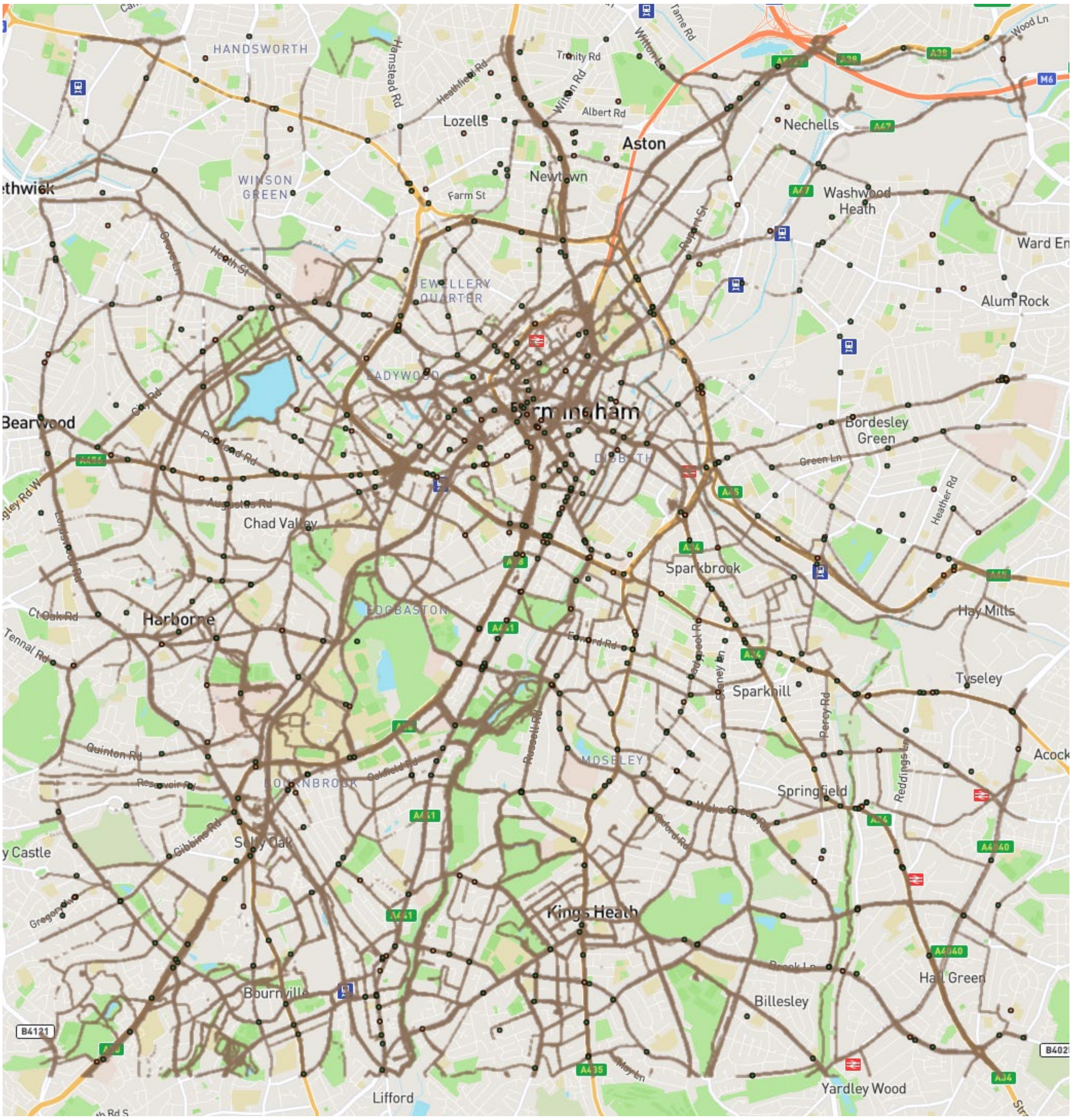


Figure 1: The coverage (brown) and STATS19 locations (dots).

SEE.SENSE®

Much like STATS19, the swerve jerk and brake jerk values can be considered a Poisson distribution, that is, a randomly occurring count which has a fixed underlying probability. Our theorem is that sudden avoidance behaviours will have a higher probability at STATS19 locations, and generally lower occurrences elsewhere.

The basic comparison of the average values of the largest jerk seen in each tile strongly supports this, with all statistical measures pointing to higher values for tiles in which STATS19 events occurred.

	Swerve Jerk		Brake Jerk	
	non-STATS19	STATS19	non-STATS19	STATS19
Mean	1.99	18.23	5.05	33.4
Mode	0	8	0	32
Median	1	16	2	32

In order to further analyse this, we evaluated the distribution and count of each swerve jerk and brake jerk for all tiles, comparing those in which STATS19 events occurred with those that did not.

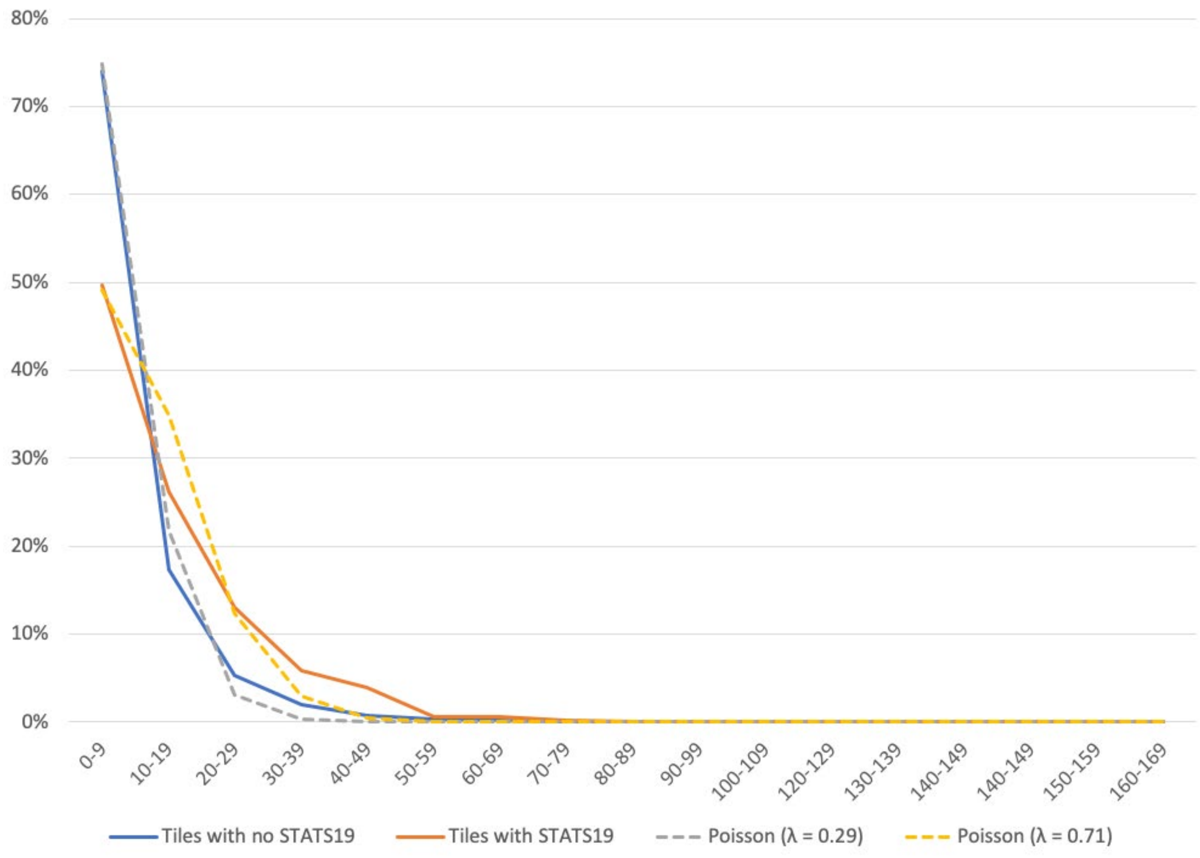


Figure 2: Histogram of Swerve Jerk Distribution.

We can see that the swerve jerk is an infrequent event. The distribution shows that the values for tiles which contain STATS19 events is likely to be higher than those which do not contain STATS19 events.

We performed a very basic statistical analysis of this distribution and found that they have a reasonable fit with standard Poisson distribution. We did not assess the full goodness of fit (p-value) and a different curve may be better. However, using this method gives us lambda values which we can compare. The lambda is the probability of an event occurring.

Fitting these curves to Poisson distributions, we find that the curves have quite different lambda values:

- **Tiles with STATS19 - lambda = 0.71 (1 journey in 141)**
- **Tiles without STATS19 - lambda = 0.29 (1 journey in 345)**

Lambda is the probability of a swerve jerk occurring. Hence comparing the rate (0.71 / 0.29) we can see that based on this model, the swerve jerk is 2.4 times more likely to occur in a tile with prior STATS19 events.

Therefore, this is indicating the probability of something happening at a STATS19 locations remains higher and requries some form of change to take place in order for it to be reduced.

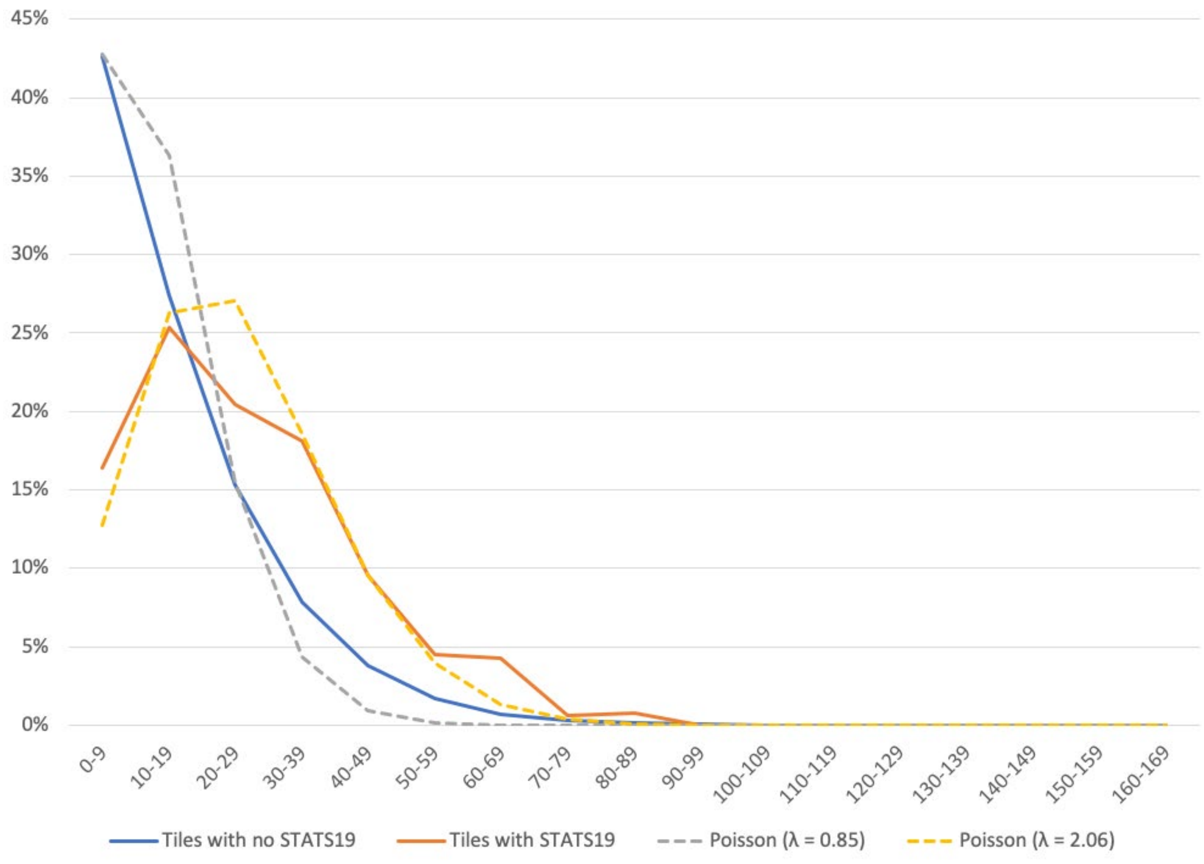


Figure 3: Histogram of Brake Jerk Distribution.

SEE.SENSE®

Compared with the swerve jerk, we see that the brake jerk is a somewhat more frequent occurrence. Again, we see the distribution of tiles with STATS19 appear to the right of those tiles without STATS19.

Fitting these curves to Poisson distributions, we find that the curves have quite different lambda values:

- Tiles with STATS19 - lambda = 2.06 (1 journey in 49)
- Tiles without STATS19 - lambda = 0.85 (1 journey in 118)

Comparing the rate of occurrence, we again see that brake jerk events are 2.4 times more likely in tiles with STATS19 events than those without.

Comparing the rates of brake jerk versus swerve jerk, we see that cyclists are 2.9 times more likely to brake jerk than they are to swerve jerk. This seems intuitively correct as the primary response to a sudden event is an initial sharp brake with fewer events causing an initial sharp swerve.

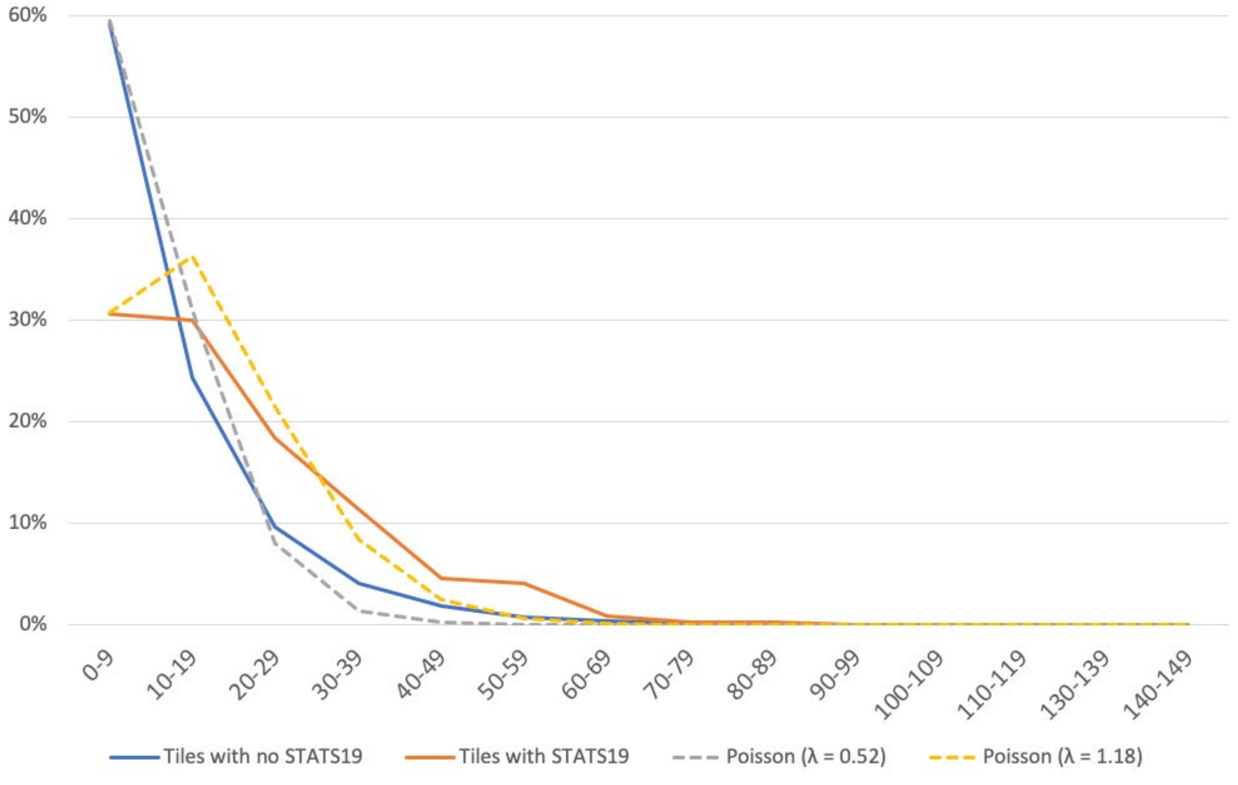


Figure 4: Distribution Histogram of Road Surface Texture

Again when fitting a Poisson distribution and comparing the rates of occurrence, we see more than 2x the occurrence of a poorer road texture.

It is interesting that we see a similar pattern, however we also know that road surface is generally a contributing factor, rather than the primary cause of a STATS19 collision.

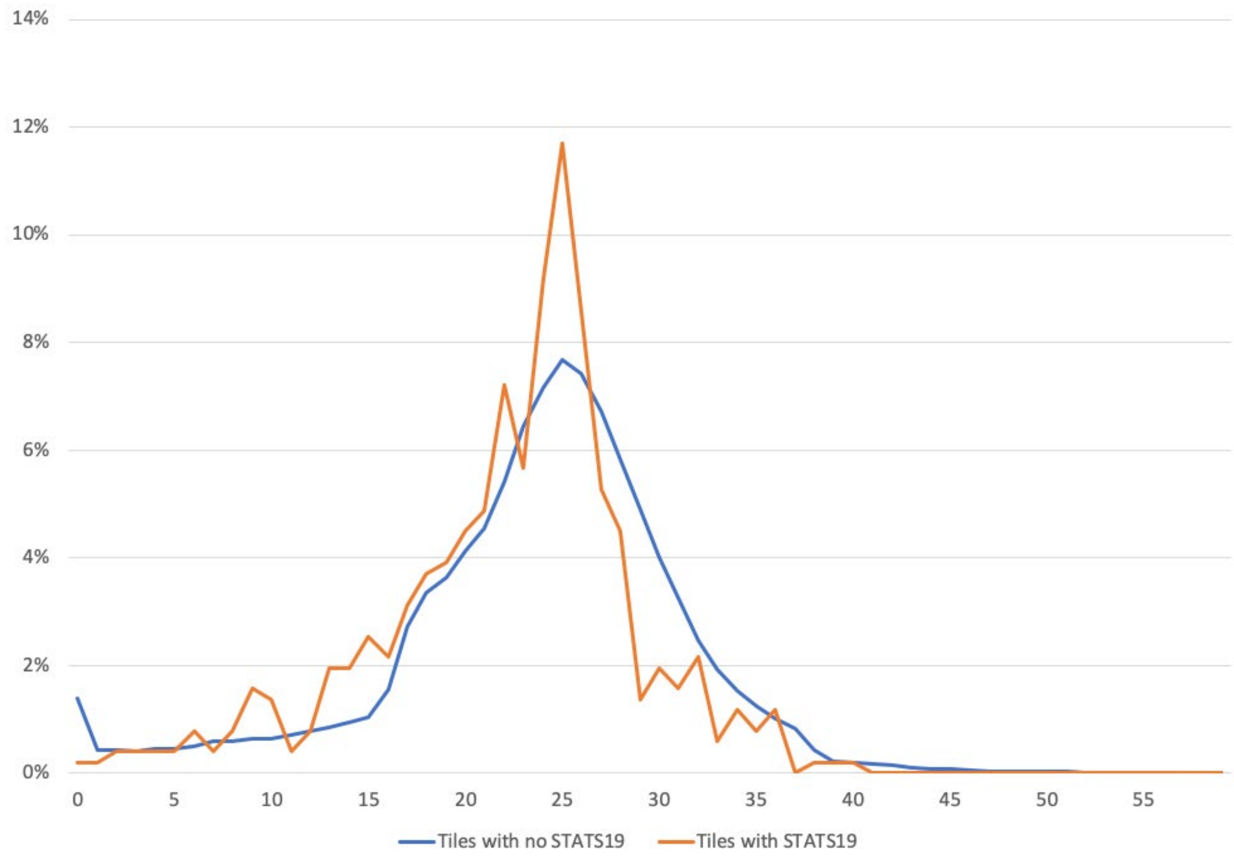


Figure 5: Speed (km/h) distribution histogram.

As we can see from the histogram above, there is no significant deviation between the average speeds at tiles with STATS19 events versus those without.

Using this baseline statistical analysis, we studied in detail a number of areas, several of which are included in Appendix A. This enabled us to develop an indicative layer, based on a combination of braking jerk and swerving jerk. Our aim was to highlight areas with the highest risk of a STATS19 event, based on having the largest quantity of high jerk values.

The indicative layer is constructed as follows:

- The minimum magnitude value of brake jerk and swerve jerk was calculated so that only the largest 5% of all jerks was included. This allowed us to look at only the most indicated sites.
- Jerks were counted within each tile and only tiles with three or more jerks from two or more distinct cyclists were included. This allowed us to rule out single events, or single cyclist anomalies.
- We assumed that tiles with both brake jerks and swerve jerks represented a higher risk than one measure alone.

SEE.SENSE®

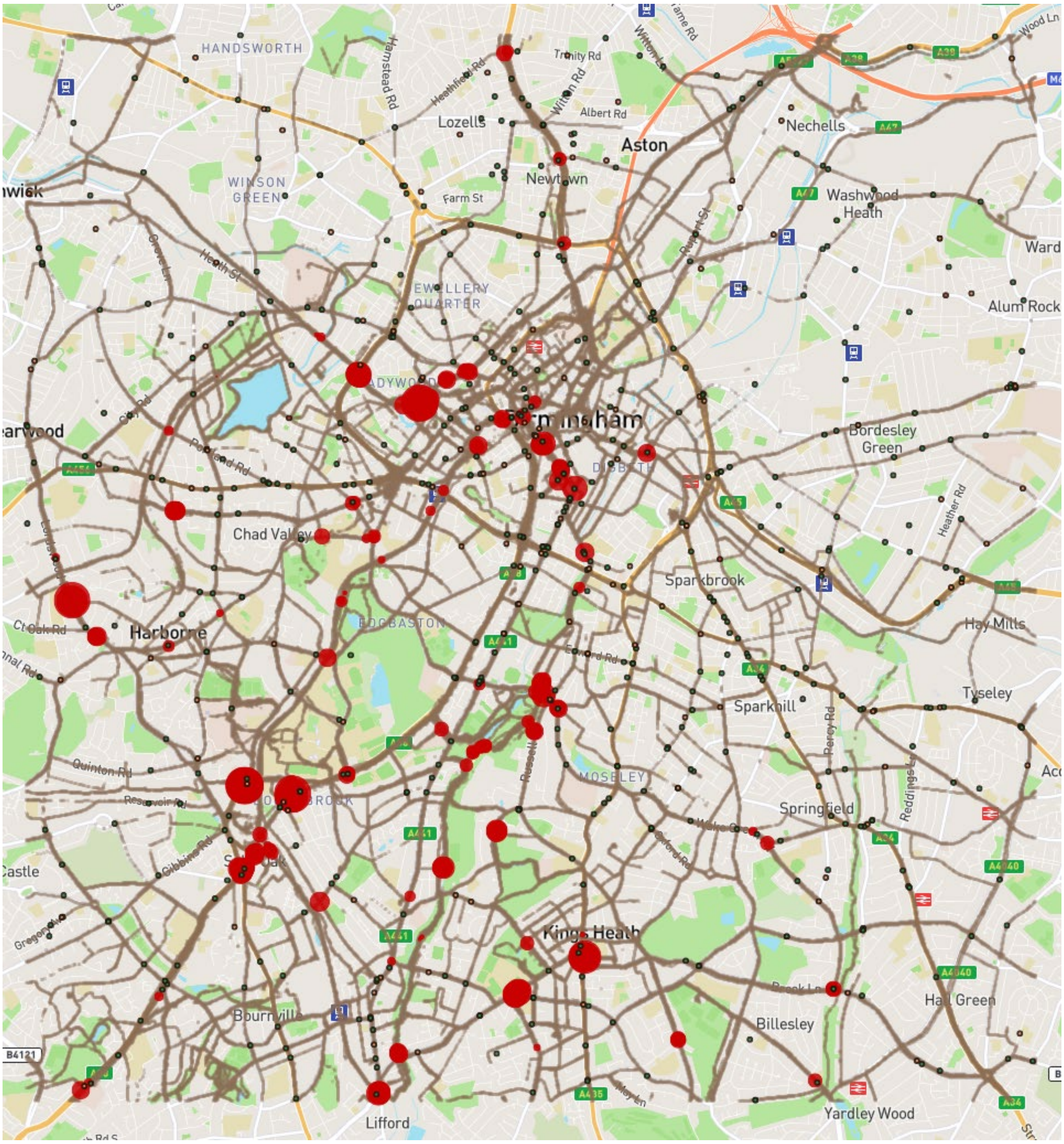
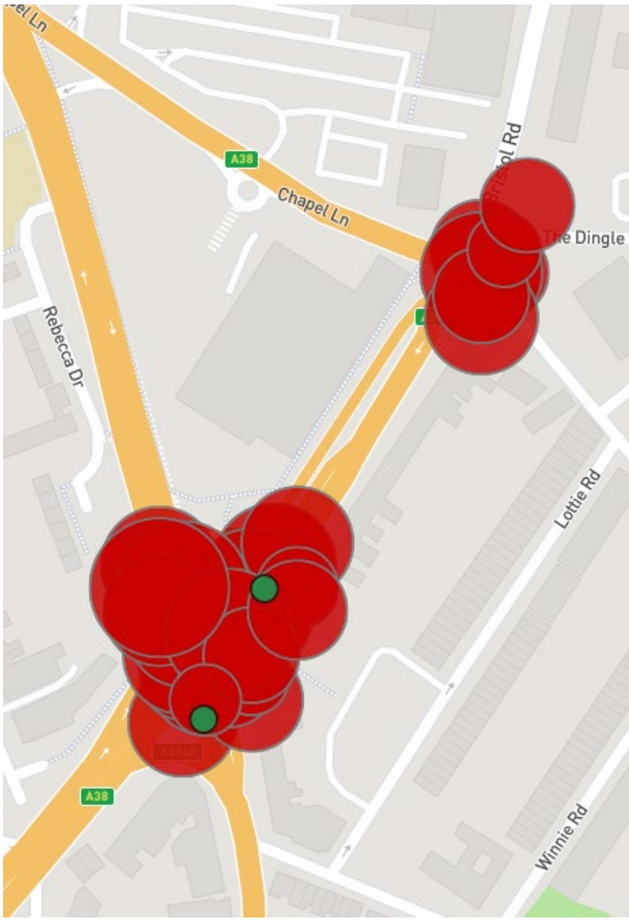


Figure 6: Indicative Red Dots where high brake jerk and swerve jerk events occur.

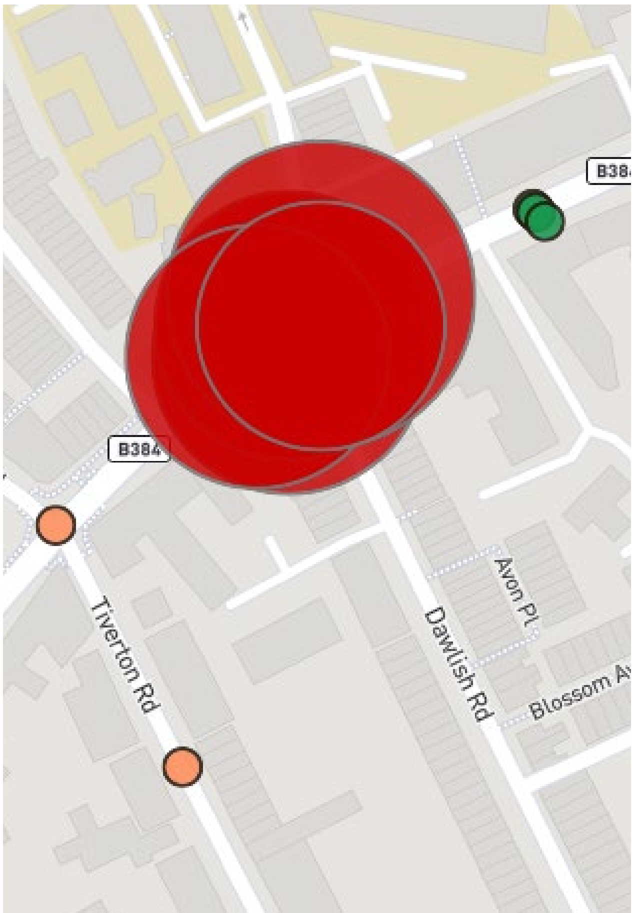


Key:

- Severity 3 STATS19 event
- Severity 2 STATS19 event
- High brake and swerve jerk indication (larger is a higher count value)

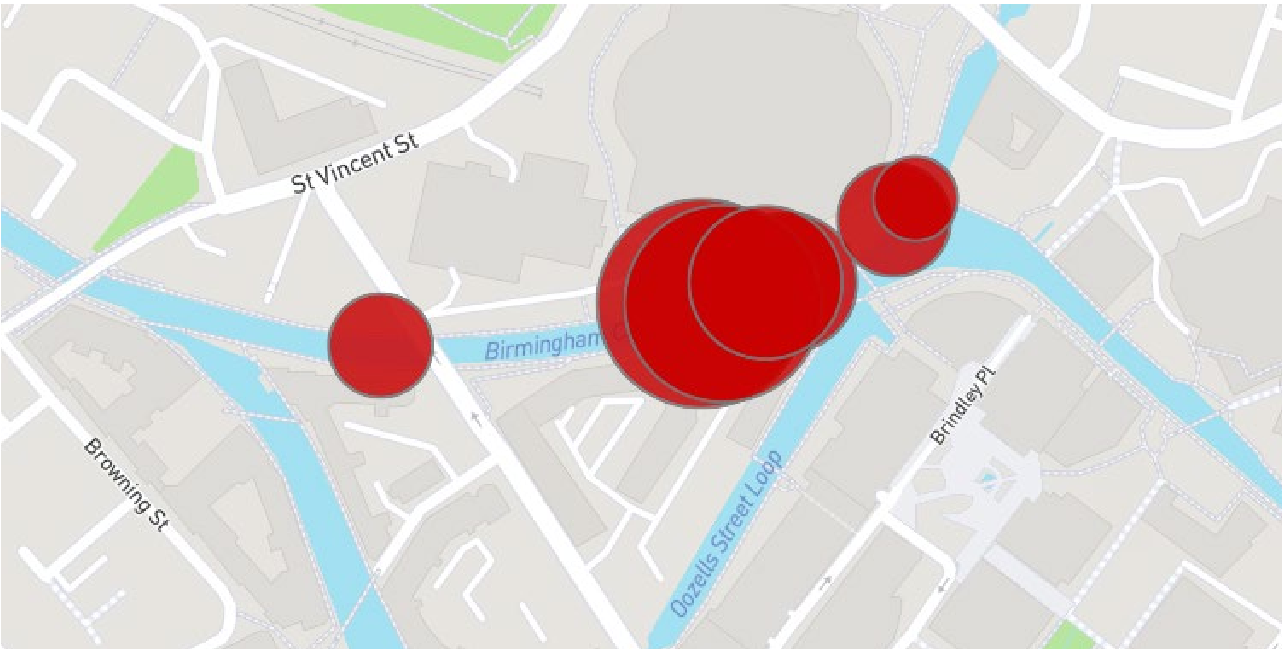
Example 1: Selly Oak, Birmingham B29 6JE. 52.439407, -1.940065

Here we see an example of a junction with two STATS19 events, surrounded by a cloud of high risk indicators. To the north east, we see a nearby junction which displays a very similar indication in the model.



Example 2: 556 Bristol Rd, Selly Oak, Birmingham B29 6BE. 52.445764, -1.933191

In this area, we see seven STATS19 events, two of which resulted in serious injury. This appears to be a fast and dangerous road. We see a very distinct cluster of risk indication at the junction with Dawlish Road, which is worthy of further study.



Example 3: Birmingham B16 8AF. 52.478776, -1.915202

Although there are no recorded STATS19 events in this area, we see a high risk score close to where the canal path narrows near shared pedestrian footbridges.

Conclusion

In areas where we have 3 or more cyclists making 15 or more journeys, we have observed that cyclists are 2.4 times more likely to experience a brake jerk or swerve jerk event in the immediate vicinity of recorded STATS19 events. Further study of these areas has enabled us to build an indicative model, based on counting the most extreme brake jerk and swerve jerk events. Where they occur most frequently, we see various potential causes, which are often worthy of further investigation, such as:

- Traffic flow disruption due to a bus stop, leading to errors in judgement on merging traffic
- Merging between different infrastructure types (e.g. a cycle lane merging into a shared lane)
- Cycling-only contraflows potentially leading to a vehicle driver making an error of judgement
- Unexpected change in the surface (e.g. potholes and slippery cobbles on canal path bridges)

We would propose that this could form the basis for a useful investigative tool to quickly identify the most hazardous cycling areas. Or alternatively, this could be used as a tool to analyse an area based on other indicating data, such as reports from cyclists.

There are a number of avenues of potential further research, based on these initial findings. These include:

- extending the data collection, and increasing the number of cyclists to further enhance the data available for analysis - this could be particularly useful for before and after analysis of areas undergoing infrastructure change
- exploration of indicated areas on the ground to determine what the ‘ground truth’ is
- the comparison of this dataset with another location
- tuning of the model based on further analysis various locations
- **conducting an analysis of the root cause** for indications, for example including other dimensions or layers, to enable deeper analysis, for example
    - The location of bus stops
    - Details of road widths and ease of traffic flow
    - Details of cycle infrastructure types, such as cycle lanes

accidents don't have to happen

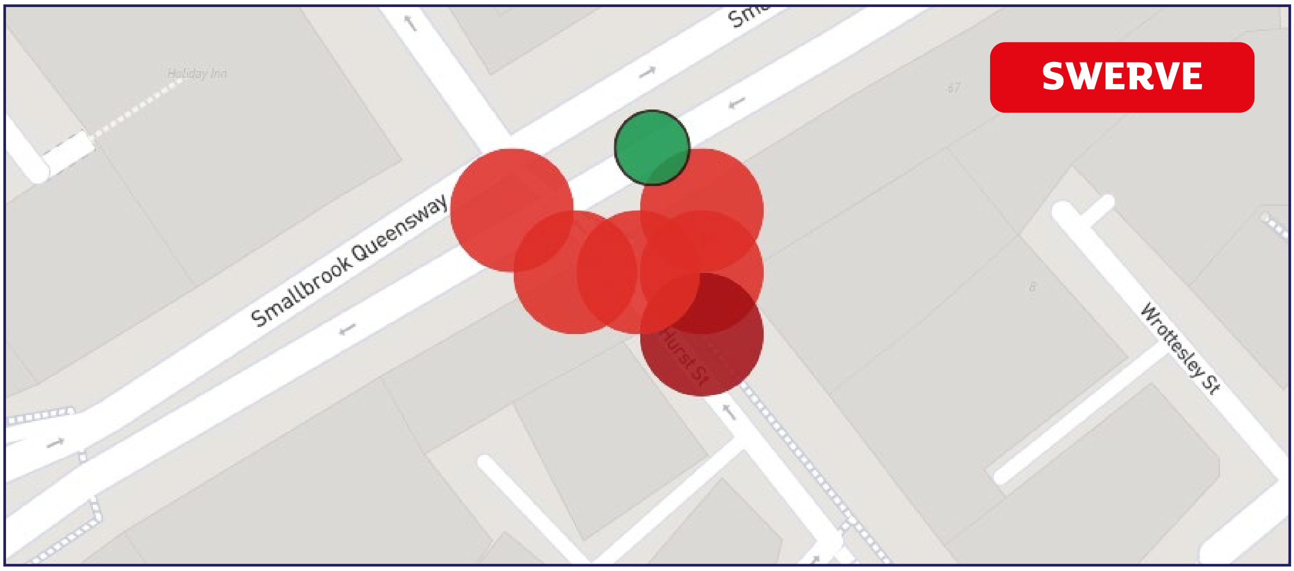
Appendix

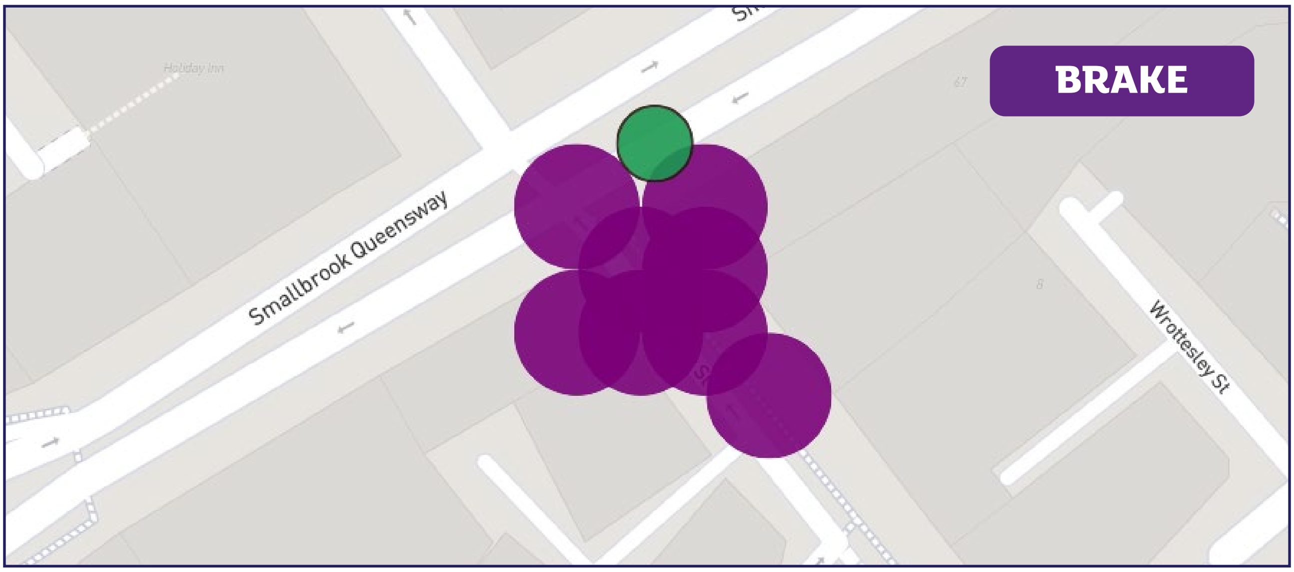
Case Study Area 1: Hurst Street

Post ride surveys from the study group suggest that motorists may be carelessly pulling out from side streets into the path of oncoming cyclists. From STATS19, we observe 4 events on this road, one at each end and two near the crossing with Kent Street.

This first study area is the intersection where Hurst Street meets Smallbrook Queensway. There is one STATS19 event at this location and we see indications of high values of both brake jerk and swerve jerk.

Case 1a: Hurst St, Birmingham B5 4HQ. 52.475593, -1.898539

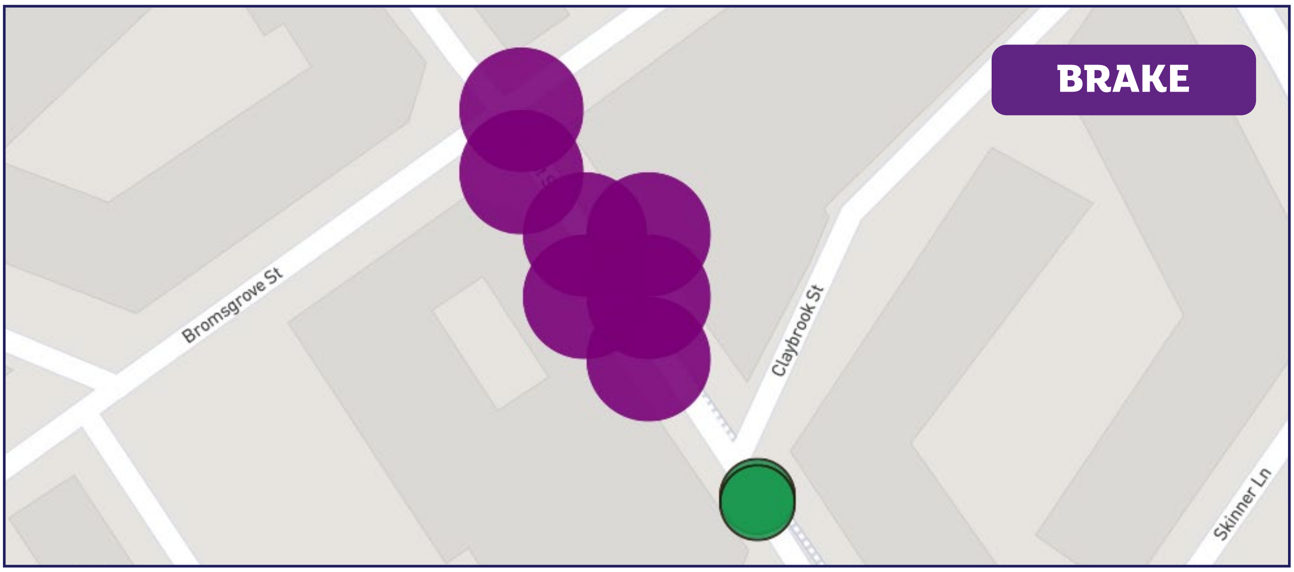




SEE.SENSE®

Case 1b: 136-150 Hurst St, Birmingham B5 6SD 52.472832, -1.895248

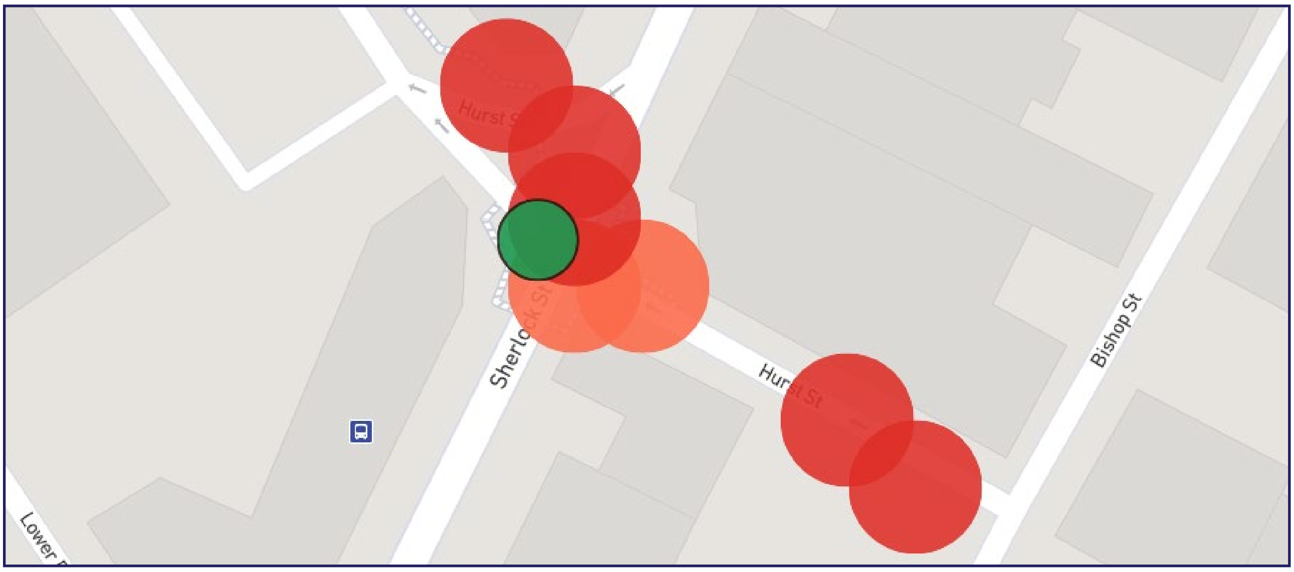
In this section of Hurst Street, we have a flow of one way motorised traffic with a contraflow cycling lane. We observe brake jerk events in what appears to be a reaction to the expectation of the motorised traffic moving through the contraflow cycle lane into Claybrook Street. There only nearby swerve jerk indication, is probably unrelated. Two STATS19 events are recorded where these roads meet.

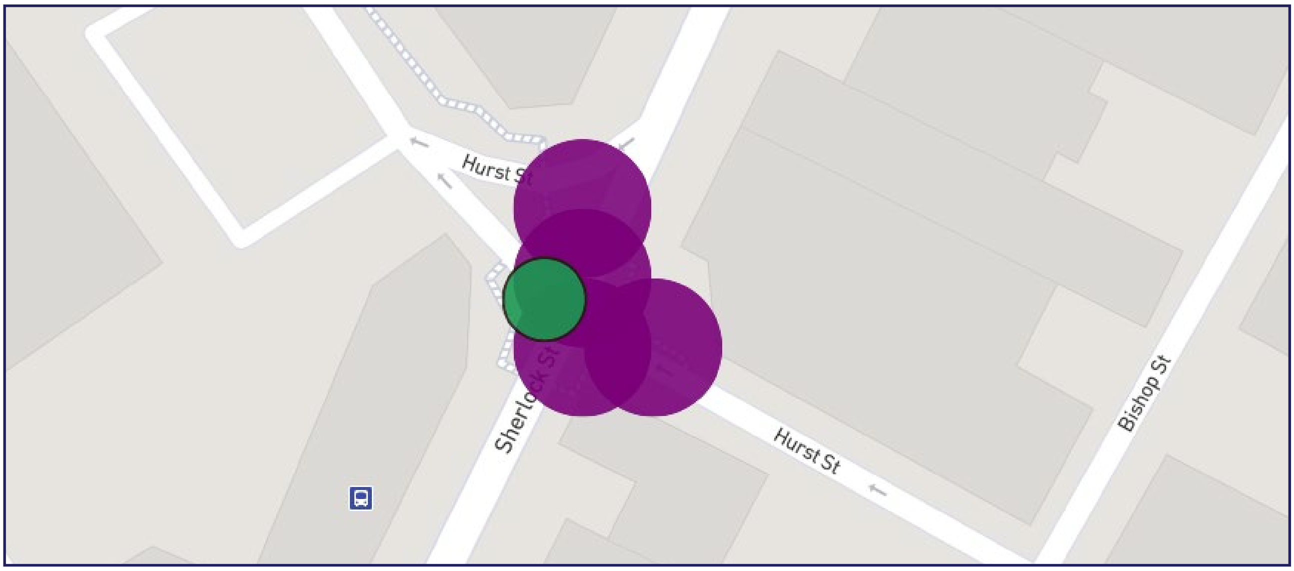


Key:
- Severity 3 STATS19 event
- Severity 2 STATS19 event
- Brake jerk indication (darker colour is a higher jerk value)
- Swerve jerk indication - (darker colour is a higher jerk value)

Case 1c: 152-164 Hurst St, Birmingham. 52.471760, -1.893936

Here we explore the junction of Hurst Street and Sherlock Street. Here we see a very complex combination of road, cycle lanes and pedestrian footways. There are a large number of points where cyclists, pedestrians and motor traffic cross in different combinations and directions. Here we observe one STATS19 event and a large degree of both brake jerk and swerve jerk.

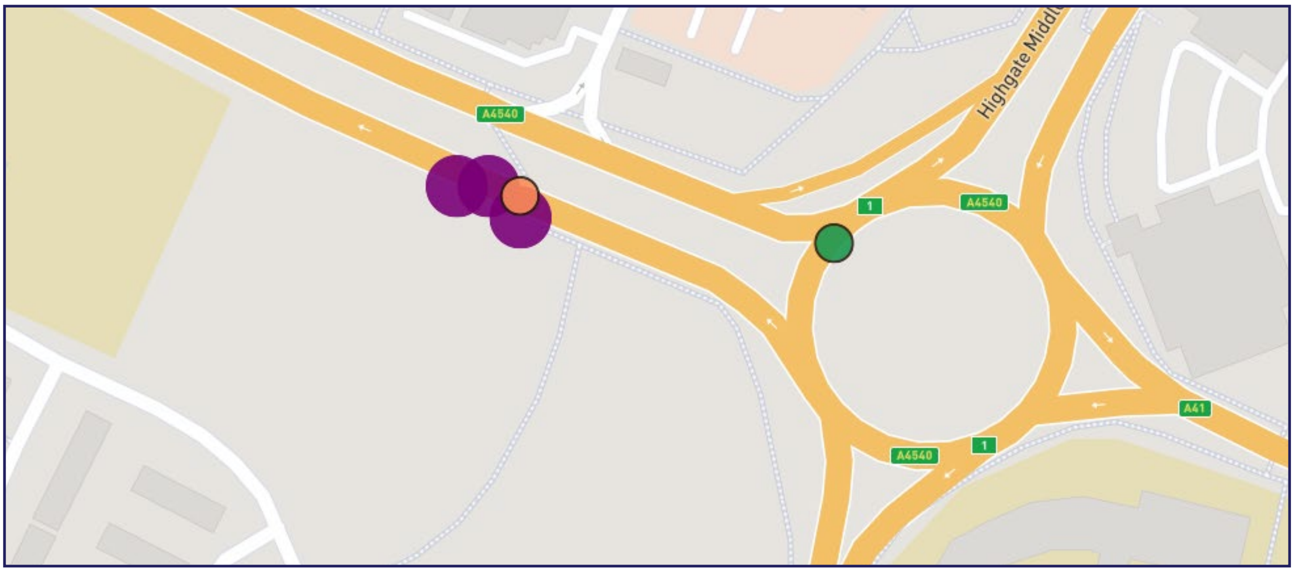


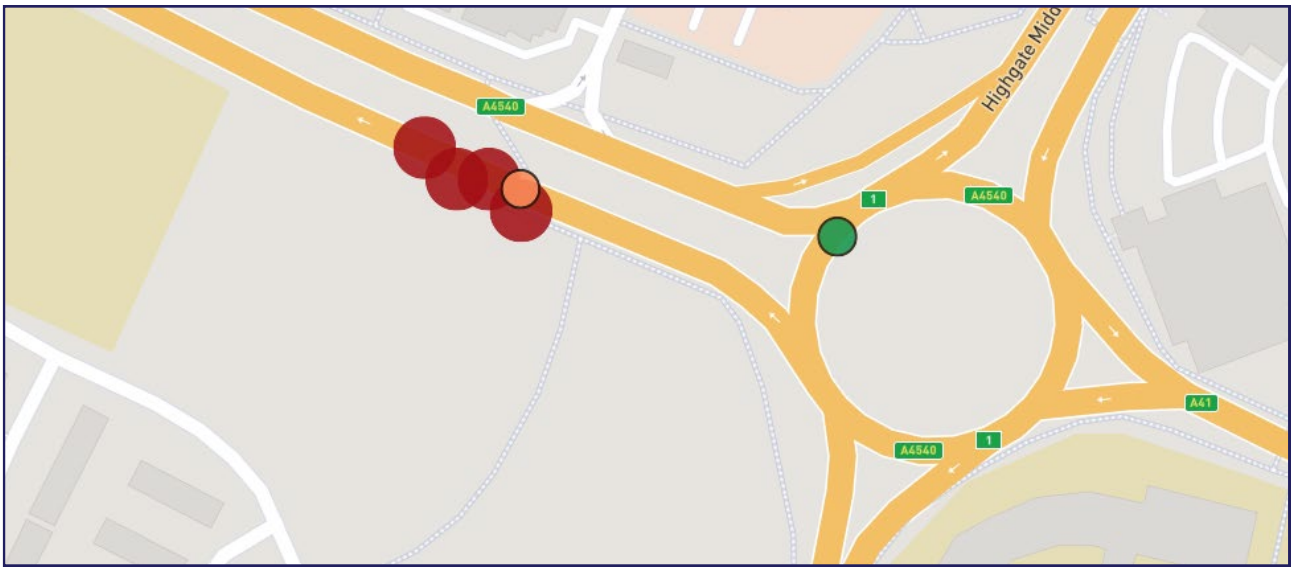


SEE.SENSE®

Case Study Area 2: Highgate Roundabout

The previous study fits the expected profile of accidents occurring at or near road junctions. In this example we explore the exit from Highgate Roundabout along Belgrave Middleway. Here we have a three lane road, with traffic travelling in one direction. We see a STATS19 event with serious injury a reasonable distance away from the roundabout. (This is the orange circle, the green, minor STATS19 event on the roundabout is not related.) Here we see both brake jerk and swerve jerk indications close to the STATS19 event. When we view the local area, we see that this coincides with a bus stop which likely precipitates non laminar traffic flow.

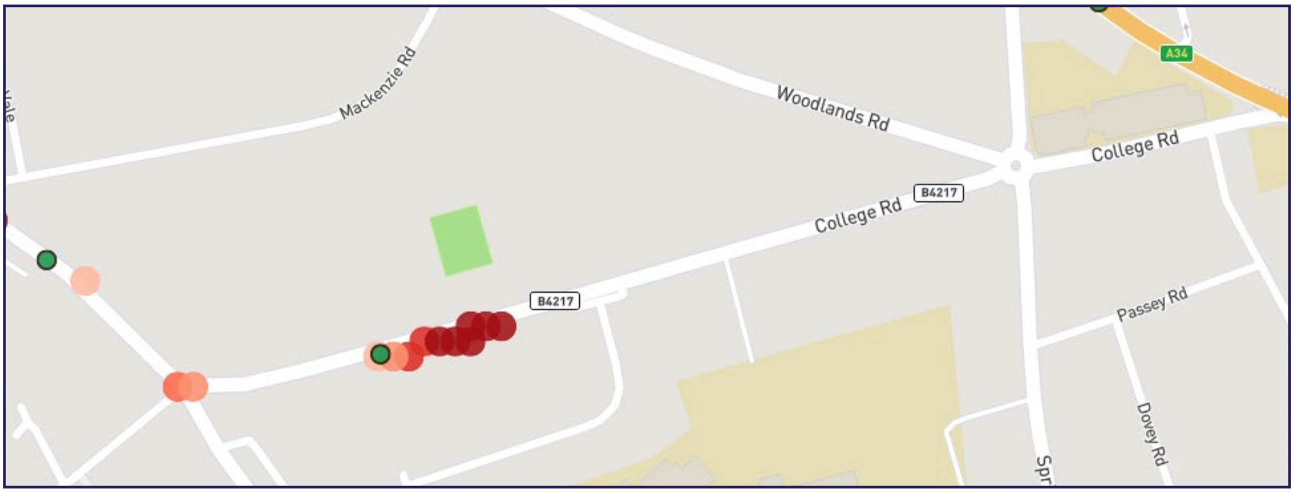


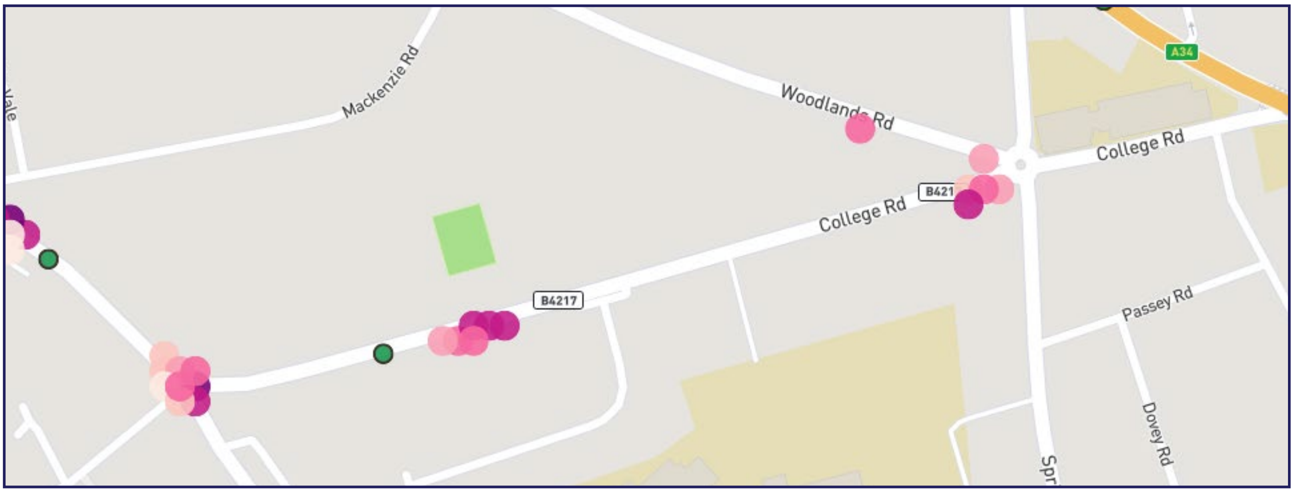


Case Study **Area 3: College Road**

157 College Rd, Birmingham B13 9LH. 52.442003, -1.865098

This street is busy due to the access it offers to the A34 Stratford Road. Motorised traffic also includes local busses. It is also narrow with cars straddling the footway and road to park. The most restricted section is where we see one STATS19 event. We see high swerve jerk and brake jerk in the vicinity. However, we do see a higher indication near to a social club and postal box. An on the ground survey might be required to fully understand what is happening.





SEE.SENSE®

RoSPA
accidents don't have to happen

RoSPA House
28 Calthorpe Road
Edgbaston
Birmingham
B15 1RP, UK

www.rospa.com

SEE.SENSE®