Low Traffic Neighbourhoods in London: Interrupted time series analysis of sensor count data

Dr Jamie Furlong[1], Professor Ben Armstrong[2], Professor Rachel Aldred[1], Dr Phil Edwards[3]

Study objectives

The key study objective of this part of the NIHR research project is to understand the impacts of Low Traffic Neighbourhood (LTN) implementation on motor vehicle traffic and active travel on roads within their boundaries. The primary research question that this statistical analysis plan relates to is as follows:

In London, what is the impact of introducing LTNs on the following outcomes, as compared to matched control streets or areas:

- Volume of active travel (walking and cycling) inside LTNs?
- Motor vehicle volumes inside LTNs?

Study design and intervention/controls

Sampling unit: LTNs

The primary sampling unit of this research is the LTN, and we have selected 8 LTN schemes for inclusion in this study from a total of 6 London boroughs. Local authorities find it challenging to plan and implement schemes, and initial lists of planned schemes do not always materialise. We therefore did not randomly sample proposed LTNs but purposively selected schemes that local officers were confident will happen. There are two schemes in Lambeth and Newham and then one scheme in Hackney, Haringey, Camden and Islington.

Choosing control areas

For each scheme, we identified a suitable control area in the same borough based on a range of criteria. These were: size and demographic similarity, suitability for an LTN intervention in principle (but without one planned), not adjacent to the study scheme, and likely to contain sites with roughly similar travel patterns to sites selected from the study area. Where possible, we matched by trip generating destinations including schools, parks, and local high streets, i.e. seeking a control site with similar destination types. There is high demographic similarity between intervention and control sites, with similar profiles for all characteristics. This enhances the internal validity of the comparisons we will make.

Selecting observation points inside LTNs and control areas

Within each LTN and control area, we identified road segments that were travel desire lines. We chose to focus on travel desire lines (rather than e.g., already quiet cul-de-sacs) because we wanted to study streets where a) motor traffic should decrease considerably

[1] Active Travel Academy, School of Architecture and Cities, University of Westminster, 35 Marylebone Road, London, NW1 5LS
[2] London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London, WC1H 9SH
[3] London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT

if/when an LTN was introduced and b) any subsequent area-wide increase in walking and cycling would likely be captured. We then purposively identified two observation points in each area that a) were >200m crow-fly distance from each other, b) covered different desire lines (e.g., North-South and East-West, or two different East-West lines), to minimise double counting, and c) were on lampposts suitable for installing sensors. Advice from local stakeholders on choosing segments in control areas helped us identify segments that they felt had comparable walking, cycling, and motor traffic flows to segments in the intervention area (generally there was no existing data on such flows for these streets, so this had to be based on their local knowledge). Where intervention sites had key destinations likely to affect traffic flows (e.g., a park) we either matched a nearby segment to a similar location in a control area or, if this was not possible, chose a segment away from that destination. We mapped boundary roads for each LTN and control area, defining these as the closest surrounding road links that might experience traffic displacement from the LTN.

Primary data collection: VivaCity sensor data on active travel and traffic volumes

To measure active travel, we installed a VivaCity sensor on each observation point inside LTNs and control areas. These sensors film the streets and use artificial intelligence to classify street users into detailed modes (e.g., pedestrian, bicycle, car, van etc). Footage is classified in the sensor and deleted after classification. Our study uses the latest version of the technology which also classifies sub-modes, e.g., cargo bikes as a subcategory of bicycles; wheelchair users or people with a pushchair as subcategories of pedestrian. These sensors record 24/7, providing rich data. Such sensors are increasingly used in the UK to collect traffic data, including across London. They will collect data continuously throughout the follow-up period for 38 months, until July or October 2024, depending on the LTN.

Data handling

Data collection

In the data collection, all data for every sensor a is periodically downloaded to be cleaned. As the data cleaning process is labour and time intensive, it is completed periodically throughout the time period. The first data download and clean, of predominantly baseline data, has occurred, and began with from the data validation date (that is the date at which the sensor was validated, normally a couple of weeks after installation) to the day of download – June 2022. The start date for baseline data has been amended where, in the first few days or weeks from validation date, there still appears to be significant fluctuation. Any sensors that are road crossings have been dropped from the analysis alongside data from the left-hand-side pavement on one of the control sensors in Camden Square.

Missing data handling

The first task after downloading the data for all time periods and noting the validation/start dates, is to identify any periods of zero count data where the sensor is likely to have stopped recording data. An automated script examines counts from each direction for each sensor and identifies days when there were either a) zero counts for a mode type; b) a very low count (less than three) across all modes. Sensor data is removed from the analysis (and subsequent analyses) if it meets any of the following criteria:

- **Zero count day:** On this day, there were zero counts for all modes for this countline direction
- **Very low count day:** On this day, the counts were so low for all modes that it appears there was a sensor malfunction

- **Between zero count days or very low count days:** The day falls between two zero count days or two very low count days. E.g. the day before and day after both have zero counts, making it likely the data on this day is unreliable.
- **Buffer before zero count/very low count day:** A 'buffer before' period is a time period on the day prior to the zero count day on which data is not considered reliable. This is either 'buffer before time', which sets the start of the buffer as the start of the very last 3 hour period of time in which there was a non-zero/non very low count.
- **Buffer after zero count/very low count day:** This is the same principle as above but after the zero count or very low count day.

In subsequent analysis, if one of the sensors in an LTN-control set has missing data, data is removed for all of the sensors (LTN and control) at the same time period. This is to ensure that the comparison is precisely like-for-like and that any findings are not distorted by the exclusions of time periods from one but not all sensors. The same principle is applied to cases where anomalous data has been removed, as is discussed in the following section.

Night-time data: detecting anomalies

The first stage of detecting anomalies uses a method of time series decomposition that combines short and longer-term trends to identify anomalies in one-hour periods at night-time (between 10pm and 6am). Known as STL decomposition ('Seasonal and Trend decomposition using Loess'), this relatively robust approach is formed on a decomposition of two parts – seasonal and trend – and uses a LOESS estimate to calculate the trend component (Cleveland, Cleveland, McRae & Terpenning, 1990).

By analysing hourly data over an extended time period, the STL method can account for patterns that arise in the count data e.g., fluctuations at different points in the day, weekend versus weekday changes, seasonal changes. More typical statistical tests of clustering methods would fail to incorporate any of these time-based changes so would perform poorly in identifying outliers. In short, the time series decomposition uses the seasonal trend and the longer-term trend to essentially create a predicted value for each count data point for each mode separately. The additive model used is specified as follows:

$$Y_t = T_t + S_t + R_t$$

Where Y_t refers to the time series values at time period t

T_t refers to the trend component.] at time period t

S_t refers to the seasonal component at time period t

And R_t denotes the remainder (or error) component at time period t.

The seasonal and longer-term trends are automatically designated based on the scale of the time series. As this is likely always daily data, the scale is set so that the season is 7 days, and the long-term trend is a month. A 'remainder' value is created that is the difference between the observed value and predicted values. Anomalies are detected when, based on the information from the long-term trend, short-term trend and the 'remainder' values, they fall outside of pre-set upper and lower limits, based on the interquartile range (IQR)[4]. The IQR of the 'remainder' values is defined as Q3-Q1 where Q3 is the 75th percentile of the

[4] For more details see: https://cran.r-project.org/web/packages/anomalize/anomalize.pdf and https://towardsdatascience.com/why-1-5-in-iqr-method-of-outlier-detection-5d07fdc82097

remainder values and Q1 the 25th percentile of the reminder values. We apply an IQR Factor of 6 to this, such that observations are considered outliers if either:

1) $y_t < Q1 - 6 \times IQR$ or
2) $y_t > Q3 + 6 \times IQR$

Using such a large IQR Factor ensures that a narrow definition of an anomaly is used. In addition, we apply an additional filter to ensure that anomalies are only detected when the 'remainder' – that is the difference between the estimated and observed values, is greater than 15 in the one-hour period. The total proportion of anomalies has also been capped for each sensor at a maximum of 5% of all hourly data points.

Night-time data: imputation of anomalies

The reason for using such a narrow definition with a filter and cap is because we subsequently replace anomalous data with the predicted values from the time series decomposition. It is important that there is a high degree of certainty that a) the observed data is not real (i.e. it is a malfunction or a misclassification associated with the sensor); b) the observed data could have a significant impact on the analysis. For instance, by only excluding data points where the remainder is greater than 15, we ensure that a case where, say, a group of 8 cyclists on a road at 11pm are not removed from the data because there is a predicted value of less than 1 cyclist per hour.

This process of anomaly detection is replicated for each mode and sensor separately. For anomalies detected in the time series decomposition that also meet the criteria outlined above, they are removed from the data and the values are replaced with the predicted values that account for seasonal and short-term trends in the data. Night-time data is imputed rather than removed because we felt that during the night, with such a narrow classification of anomalies, the risk of replacing real values was low. This, in combination with the lower variation in night-time values across all modes means we are more confident that the imputed values are likely to be good estimates for the actual values. The night of 31st December is removed from the anomaly detection due to increases in all modes that are highly likely to be real. Overall, early analysis of data held up to June 2022 indicates that the proportion of night-time data identified as anomalous and then re-imputed is low: just 0.36% of all night-time one-hour periods have been imputed in this time period.

Day-time data: anomalous data handling

After imputing the anomalous night-time count data, anomalies are detected specifically during the daytime (6am-10pm) using the same time series decomposition approach, with some modifications. Firstly, because of the greater variation in daytime count data, we are not replacing anomalies with missing values, as there is a higher risk that the observed data is real (due to higher day-time variance in hourly count data) rather than the result of a sensor malfunction/miscounting. Instead, the identified anomalies are manually reviewed before deciding whether they should be included/excluded from any analysis, as described in the next section. Secondly, anomalies are detected in the daytime at the day-level i.e., we are identifying anomalies for each mode for each day, rather than at the hourly level. The reason for this is to make the subsequent manual process more manageable across a small team – there would simply be too many anomalies to review manually if they were detected at the hourly-level. Thirdly, the narrow classification of anomalies used at night-time is replicated, except that a filter is not applied to the value of the 'remainder'. The remainder values and the variation in counts are generally much larger for day-time counts meaning any filter would be likely both unnecessary and ineffective.

After day-time anomalies are identified, the next stage of the process is twofold:

1) To identify whether pre-identified day-time anomalies should be excluded or included in the final, cleaned data for analysis.

2) To identify any other longer-term trends (rather than short-term, daily anomalous points/fluctuations) that are unusual and come to a decision about whether this data should be included or excluded.

This is a manual task completed by several reviewers. Each reviewer has access to the identified anomalies, the predicted values, weather and bank holiday data, the tracks images on the Vivacity dashboard for each sensor, the day-time daily counts for each sensor (including anomalies) and a daytime count graph of pedestrians in the road. Using a combination of this data, in respect to objective 1 above, the reviewer is tasked with deciding whether each anomalous data point should be including or excluding in the data for analysis. Table 1 below was given to each reviewer with instructions for how to make the decision and how to code the decision for future reference/analysis. The logic behind the table was that:

1) All obvious sensor malfunctions/miscounting should be removed
2) Anomalies likely to be caused by one-off or non-repeated events should be removed
3) Anomalies likely to be caused by roadworks either in the LTN/control area or nearby should be removed.
4) Anomalies likely to be caused by bank holidays, weather conditions or recurring events should be retained.
5) Anomalies that appear to have some other cause that is real (e.g. children playing) should be retained.
6) All other anomalies without an explanation should be removed.

While we accept that there is a degree of subjectivity to this manual process, anomalies can be caused by multiple factors that simply cannot be identified by a statistical model. In addition, we have put validation methods in place. Firstly, prior to the reviewers conducting this manual process, meetings take place between reviewers to carefully consider examples of each type of ‘inclusion’ and ‘exclusion’ anomaly and how they can be identified from the tracks or any other data. Secondly, after each reviewer conducts a review of a set of sensor anomalies, a discussion takes place of any uncertain cases that were flagged with other reviewers to come to a collective decision. Thirdly, the final validation is that a non-reviewer completes the same process with identical instructions on a sample of anomalous data points. Their decisions to include or exclude data points are compared with those made by the reviewer. So far, this has been completed for the baseline data, and has shown a high level of agreement between the reviewers and the non-reviewer.

Table 1 Instructions given to reviewers for deciding to exclude or include anomalous data points

Decision	Description
Exclude	anything that is an obvious sensor malfunction e.g., a **completely implausible** count.
	an obvious case of mis-counting a parked motor vehicle or pedestrian, as shown by a circle/group of tracks in one location/on one spot.
	one-day or short-term spikes that do not seem as if they could be repeatable i.e. if there is no reason to believe that this might be the result of some event that might happen again next year or a bank holiday etc. In practice this means the default for 'anomalous for unknown reasons' is exclude.
	anomalous days we think may be roadworks. These are not an artefact in that it captures something that happened in the world, but are not expected to be consistent between the LTN/control area or pre-/post, and are arbitrarily based on where we happen to put a specific sensor. *Roadworks on the road itself are likely identifiable by a fall in all forms of motor traffic (not just cars), though there could also be a very high number of LGV/OGV as a van/digger works near the countline. Possibly also the pedestrian tracks show construction workers concentrated on a site. Also, look at the* ***pedestrians in the road graph****: there could be an increase during periods of construction. Be careful not to confuse school holidays for roadworks.*
	one-off unusual days due to highly localised one-off things e.g. a non-repeating street party. I suggest exclude on basis that are similar to road-works in being highly localised and potentially sporadic? plus in practice we may find it hard to tell the difference.
Include	unusual days due to school holidays/bank holidays etc. Also unusual days due to unusual weather, e.g. snow supressing cycling. These fluctuations are real and may be expected to be consistent pre/post and between LTN/control, and also to have been seen regardless of where the sensor is.
	periodic / recurrent unusual days, e.g. a market every Saturday. Since should see this pre/post consistently.
	Any recurrent but low counts e.g. a bus which is normally zero being 2.
	Other reasons to include might be seeing something like a spike in cyclists in school holidays or summer months. If tracks are legitimate, we might conclude that this is likely children playing.

Why have all anomalies not been removed or retained?

It is important to note that anomalies have not simply been removed because of their extremeness. Rather, they have been removed when they are likely a measurement error caused by the sensor malfunction/misclassification or a likely one-off event that could distort findings. If these values were left in the dataset, it could violate key model assumptions and lead to false conclusions. This risk is especially high when we are comparing data across two LTN and two control sensors, rather than aggregating across all sensors. On the contrary, removing all anomalous values could lead to the exclusion of both 'real' and valuable data and reduce the statistical power of any models by leading to larger periods of missing data. As explained in the following section, a sensitivity analysis will assess the extent to which changes in the effect sizes of LTN interventions are associated with the inclusion or exclusion of anomalies.

Data sources
The primary outcome data – count data by time and mode - comes from the vivacity sensors – two in each LTN and two in each control area. Data related to the LTNs, including the boundaries, implementation dates and modal filter locations, has been attained from the websites of each borough council. We have then defined our own boundaries for each LTN and ensured that each sensor is placed within these agreed boundaries.

Statistical analysis

The following section outlines the initial descriptive analysis and the provisional planned core models. Specific details of the statistical models will depend on preliminary, descriptive analyses and *may change* depending on data availability and suitability.

Step 1. Initial descriptive analysis
Before specifying the exact specifications of the models outlined in Step 2 and Step 3 of this section, we will conduct some initial, preliminary analysis of the data. For the most part, this will consist of some simple before-and-after descriptive comparisons of the sensor count data across a) selected time periods and b) LTN and control areas. For example, we are likely to compare the mean or median daily counts, pre- and post-intervention, for each transport mode across LTN and control pairs. This initial analysis will not only help validate any findings from later models but will be useful, and could be expanded on, in cases where interrupted time series models are not possible.

Step 2. Interrupted time series analysis
Once there is aggregated sensor data to provide daily totals, e.g. bicycles per day, for each sensor, we will first conduct an interrupted time series (ITS) analysis separately for each LTN sensor only. The benefits of conducting initial analysis for each sensor will allow us to examine the heterogeneity in the effects associated with LTN interventions across different locations, since we expect areas to adapt differently to schemes. ITS is used here because we have data over time, and we want to understand the extent to which the outcome variables – counts of pedestrians, cyclists and motor vehicles – have changed as a result of implementation.

Three separate models will be created for each LTN sensor (rather than aggregated for each site) at this stage because it is possible that at each site, the two LTN sensors could be differently affected by the LTN intervention. It would also be far more complicated to separate each sensor in the second part of the analysis - the controlled interrupted time series – as each LTN sensor is not necessarily matched to a specific control sensor (rather than the sensors being matched, the 'sites' are matched). As the outcome variable is always a count, three regression models will be specified, the core of which is as follows:

$$Y = b_0 + b_1T + b_2D + b_3P + e$$

Where *Y* is the outcome variable (model 1: daily count of pedestrians, model 2: daily count of cyclists; model 3: daily count of motor vehicles, all logged). If daily counts appear low (<10) on an appreciable number of days for some outcomes, the model will be fit as a quasi-Poisson regression model with log link, otherwise as a standard linear regression model. (See also sensitivity analyses.)

T is a continuous variable which indicates the number of days passed from the first day of observations (the validation date)

D is a dummy variable indicating whether the observation was collected before (0) or after (1) the LTN intervention

P is a continuous variable that is the number of days that have passed since the intervention. Prior to the intervention occurring *P* = 0.

The inclusion of *D* and *P* in the models is dependent on their significance and the impact that they have on the power/precision of the models. For example, our *a priori* hypothesis is that there is more likely to be a significant immediate impact of the LTN. In such a case, the 'slope' estimated by *P* could be insignificant and negatively impact on the precision of the *D* estimate. In the case of the models predicting cycling and walking counts, the opposite may be the case, as we hypothesise that the LTNs may have a longer-term, gradual impact. To begin with, we will specify the models to include both *D and* P, but if either is insignificant ($P<0.05$) and there is a negative impact on power or precision, then the model will be changed accordingly.

This model essentially allows us to assess the impact of LTN interventions by comparing counts in the pre-intervention period with counts in the post-intervention period for each LTN sensor. The pre-intervention period will act as the control in this model. While the underlying trend will be accounted for in these models and population change over the time period is likely to be minimal, the ITS models will not be able to account for the confounding effects of other time-varying risk factors: the possibility that there are underlying changes in the area that may be mis-specified as an effect of the LTN intervention (Bernal et al., 2018).

Step 3. Controlled Interrupted Time Series

One of the weaknesses of the ITS models conducted solely for each sensor is that they are subject to threats to internal validity. Observed changes in count data pre- and post-intervention may reflect underlying trends that occur more widely in the area, perhaps due to other events such as the expansion of the Ultra-Low Emission Zone, Covid-19, petrol prices and so on. To address this issue, we will make use of the control area sensors that are not subject to the intervention. This will enable us to be more confident that any observed effects are the result of the policy intervention itself. Three controlled interrupted time series models will be created for each LTN-control pair where the outcome variable is a ratio of counts (daily LTN sensor count/daily control sensor count). The count data will be aggregated for each pair, such that we take the total counts from two LTN sensors as our intervention counts and the total counts from two control sensors as the control counts. The data will therefore have two observations per each day and will include a dummy variable to differentiate between the control (0) and treatment group (1).

The regression models will have as the outcome variables the ratio of pedestrian/cyclist/motor vehicle counts in the LTN area to pedestrian/cyclist/motor vehicle counts in the control area respectively.

$$Y = b_0 + b_1T + b_2D + b_3P + +e$$

Where *Y* is the outcome variable, specified as log(intervention count/control count) I.e. the ratio of intervention to control counts.

T is a continuous variable which indicates the number of days passed from the first day of observations (the validation date)

D is a dummy variable indicating whether the observation was collected before (0) or after (1) the LTN intervention

P is a continuous variable that is the number of days that has passed since the intervention. Prior to the intervention occurring *P* = 0

As with the ITS models, *P and D* will both be included in the models unless sensitivity analyses show that they are insignificant and negatively impacting on the precision of each other.

Where daily counts are overwhelmingly greater than 10. This will be a standard linear model. Otherwise, it will be a quasi-binomial model with logit link.

Methodological considerations

1) Time-varying factors

Although having a control group unaffected by the intervention reduces the problem of internal validity, there could still be time-varying cofounders - variables that could change across time and could be plausibly related to the intervention and the outcome variables, in particular for the ITS. In the daily count model, it is anticipated that this will include day of the week (weekday or weekend), local school holiday (yes or no), public holidays (yes or no) and season. Seasonal patterns may in part be explained by weather, such that some weather variables will be included: length of daylight, daily rainfall and daily temperature[5]. These explanatory variables will be included in the models irrespective of their statistical significance for the uncontrolled ITS, but subject to review for the controlled ITS.

Another potential source of time-varying confounding is in other infrastructural changes that affect either the LTN or the control area. We have already outlined how we have addressed the problem associated with anomalous counts that are thought to be roadworks. It could also be the case that there are other changes made to road layouts or infrastructural changes e.g., pavement widening, one-way street creation that could impact on the outcome variable. While it is almost impossible for us to detect every change that could affect traffic levels in the wider area, we did choose control areas after lengthy discussion with local authorities in which they confirmed no plans across the time period for any substantial infrastructural changes in the area. We maintain contact with stakeholders in this regard and will address further issues should they arise.

Finally, there is likely to be a time-varying effect of the Covid-19 pandemic, as restrictions have changed across the time period under study. Some of the pre-intervention data is likely to be affected by the Covid-19 pandemic. We will incorporate in the ITS the Government Stringency Index (GSI) from the Oxford COVID-19 Government Response Tracker in our models. This index is scaled from 0 to 100 with 0 being the least strict measures and 100 the strictest. For the time period up until October 2022 (when this was no longer being updated), we will also include a measure of residential mobility taken from the Google Mobility Index. This shows the difference in median time spent at home for a region (e.g., Greater London) compared to the same weekday in the pre-pandemic period.

2) Seasonality

Travel behaviour in the UK generally exhibits a seasonal pattern in which people are more likely to travel actively in warmer, drier and lighter months and more likely to drive in colder, wetter and darker periods of the year. This seasonal pattern could cause problems at two levels. Although much of this pattern is likely to be explained by weather variables which we will be controlling for, some might not be. Firstly, if the distribution of months of the year in the before and after data is not equal, the results could be biased. There might, for instance, be a higher proportion of summer months in the pre-intervention data than the post-intervention data. Secondly, if outcomes of consecutive months are more similar than those that are different points in the year, there will be autocorrelation present in the data (Bernal et al., 2017). The problem of autocorrelation is discussed in more detail below.

To address these problems caused by seasonality in the ITS models, with ample pre and post-intervention data, we could simply include an indicator variable in the model for each

[5] The inclusion of these and other numerical covariates (linear, curved, grouped, etc) will be decided based on model fit. Their functional form will also be determined in the same way, starting with linear terms and adding polynomial terms if model fit is significantly improved by doing so.

month or season. However, here data is limited, so to gain power and precision we plan to model seasonality as annual period fourier (sine-cosine) functions (Bhaskaran et al., 2013).

Seasonality is not likely to be an issue in the CITS analysis as season will impact similarly on intervention and control areas.

3) Time-invariant factors

While in most forms of transport-based regression modelling, common confounding variables that are often controlled for might be area-level demographic characteristics or socioeconomic characteristics, this is generally not necessary in interrupted time series. These characteristics would only have a confounding effect on the time series results if they both predicted the outcome variable and changed according to the time of the intervention (Wagner et al., 2002). Given the short time period of data collection, we consider it unlikely that there would be significant population change within the LTN or control areas that would impact on the results of the study. Therefore, risk factors such as area-level demographic characteristics are effectively time-invariant and do not need to be included as confounders in this study.

An additional factor that could lead to a different temporal pattern in the outcome variable is proximity from the sensors to the nearest school, or the number of schools/pupils in proximity to a sensor. As well as including a dummy variable in the models to account for school holidays (which should account for most variance), we will investigate such effect modification by presence of primary and secondary schools in the area. A variable will be created that counts the number of primary and secondary school pupils at any schools in the LTN/control area and within a specific walking distance from the sensor. Any models will test the extent to which there are changes to any effects associated with including or excluding these variables as interaction terms.

4) Over-dispersion

If we specify Poisson or binomial time series models, a further issue that might affect the time series models is over-dispersion. That is, a situation where the variance in the outcome variable (counts in the Poisson ITS models or ratio in the binomial CITS models) is higher than that which is predicted by the model. This can lead to the incorrect estimating of standard errors (Bernal et al., 2017). If this is found to be the case, a scaling adjustment will be made to attain more appropriate standard errors, as outlined in Bernal et al. (2018) and Bhaskaran et al. (2013).

5) Autocorrelation

An assumption in the regression models will be that the observations are independent from one another. In time series data, there is frequently autocorrelation present in the data, in which observations that are closer in time (i.e., consecutive) are more similar than those that are further apart in time. It is likely that, after introducing controls to the model for seasonality, long-term trends and the intervention itself, residual autocorrelation in the outcome data will be significantly reduced. Nonetheless, residual autocorrelation will be assessed through an examination of a residual plot against time and (for linear models) by performing suitable tests such as the Breusch-Godfrey test or generating a Durbin-Watson statistic. If residual autocorrelation remains in the data, it may be necessary to adjust the model by (for Poisson models) including an estimated autocorrelation parameter (Brumbach

2000). Alternatively, if a linear model is specified, an autoregressive integrated moving average (ARIMA) could be used (see Nelson, 1998).

6) Allowing for delayed intervention effects

One further issue in the models is that there may be delayed intervention effects. Some LTN interventions may not be implemented on one specific day and in many cases local authorities do not begin fining drivers of motor vehicles for driving through camera-controlled modal filters for a few weeks after implementation. In addition, in the short-term following implementation, many drivers are unaware about where they can and cannot drive. Subsequently, it is likely that there are quite unusual travel patterns e.g., many drivers driving up to a modal filter and having to turn around and drive out of the LTN area. To allow for this if there is evidence for it, we can introduce a delay to the intervention dates in the model or potentially exclude dates immediately following the intervention and re-execute the model to assess if the trends hold true with or without the delayed intervention.

7) Sensitivity analyses

The sections above have outlined the potential problems that may affects the ITS and CITS models and have outlined how they will be tested for and their impact on the models assessed. The following bullet points summarise the sensitivity analyses we will execute:

- The impact of replacing linear regression models for log(counts) and log(ratios) by quasi-Poisson and quasi-binomial (logistic) will also explored, in particular for measures where counts are lowest (e.g. bicycles).
- We will assess the impact on model power and precision of including/excluding the *P* ('slope') and *D* ('step') variables from the models.
- The effect of night-time anomaly imputation will be explored by comparing models with imputed data with those with anomalies retained or removed. This can be repeated for day-time anomalies.
- There will be a validation exercise in the manual decision-making process on day-time anomalies: a non-reviewer will complete the same steps with identical instructions on a sample of anomalous data points. The agreement between reviewers' identification of anomalies and inclusion/exclusion will be assessed.
- Models will be executed that include and exclude various time-varying variables and controls for seasonality, to test whether these have significant effects on any conclusions.
- Models will be assessed for over-dispersion, autocorrelation and delayed intervention effects, with appropriate adjustments applied where necessary.
- We will test whether the inclusion or exclusion of outliers/anomalies has an impact on the intervention effects.
- We will test for any effects associated with changing the controls for seasonality and the specification of confounding variables or outcome variables.

The main model choices may be modified in the light of the sensitivity analyses o experience in model fitting. Ultimately, the results will be assessed to determine how sensitive they are to some of the key decisions made throughout the process.

Reference list

Armstrong, B. (2006), Models for the Relationship Between Ambient Temperature and Daily Mortality. *Epidemiology*, 17:6, 624-631

Bernal, J.L, Cummins, S. and Gasparrini, A. (2018), The use of controls in interrupted time series studies of public health interventions, *International Journal of Epidemiology*, 47:6, 2082–2093

Bernal, J.L, Cummins, S. and Gasparrini, A. (2017), Interrupted time series regression for the evaluation of public health interventions: a tutorial, *International Journal of Epidemiology*, 46:1, 348-355

Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L. and Armstrong, B. (2013), Time series regression studies in environmental epidemiology, *Int J Epidemiol,* 42:4, 1187-95.

Brumback, B.A.,Ryan,L.M.,Schwartz,J.D.,Neas,L.M.,Stark,P.C.,Burge,H.A. (2000), Transitional regression models, with application to environmental time series. *J. Am.Stat.Assoc.* 95,16–27.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., AND Terpenning, I. (1990), STL: A seasonal-trend decomposition. *Journal of official statistics,* 6(1), 3-73.

Dennis, J., Ramsay, T., Turgeon, A.F. and Zarychanski, R. (2013), Helmet legislation and admissions to hospital for cycling related head injuries in Canadian provinces and territories: interrupted time series analysis, *BMJ*, 346:f2674

Nelson, B.K. (1998), Time series analysis using autoregressive integrated moving average (ARIMA) models, *Acad Emerg Med*, 5, 739-44

Wagner, A.K., Soumerai, S.B., Zhang, F. and Ross-Degnan, D. (2002), Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics*, 27, 299-309