*Reminder of RQs:*

*1. Spatial and Temporal Patterns*

* 1. Are there spatial patterns in missing incidents?
  2. Are there temporal patterns in missing incidents?
  3. What did Covid-19 mean for missing incidents?

*2. Police Response*

* 1. How has the handling of missing person calls changed from 2015-2020 over, grade, origin, response time and classification?
  2. Are there any significant associations between these variables?

*3. Environmental/Neighbourhood Factors*

* 1. Is there an association between rate of incidents and area-level deprivation?
  2. Is there an association between rate of incidents and mental health at the neighbourhood level?

# Methods

## 

## Datasets

A total of 6 datasets were used in the analysis with the first being the CFS data received by Cheshire Police. It contains routinely recorded calls by police forces and public safety professionals in England and Wales containing both public and non-public emergencies. The dataset covers crime calls from 2015 to 2020 with detailed geographical, descriptive and time-date variables covering 74 types of incidents. CFS provide one of the most robust datasets for crime analysis due to the specificity of variables covering times and location, but also creates an opportunity to evaluate the effectiveness of police departments strategy (Makin et al., 2021). Earlier research as identified the importance of CFS that can hold as much importance as official statistics and victimisation data (Warner and Pierce, 1993). Using this method of data-collection has also allowed for the effects to be situated in real-world behaviour with the benefit of a large sample size allowing for generalisability. Moreover, with a wide timeframe more robust methods of analysis can be used that allow for the temporal trends, noise and spatial patterns to be explored in the patterns of missing incident reports (Weston et al., 2019) . The CFS data was then joined to a shapefile containing the geometric attributes for the four main local authorities (LAs); Cheshire East, Cheshire West and Chester, Warrington and Halton. After transforming the object to the correct coordinate reference system (CRS), using the **EPSG:4326 (WGS 84),** we now have a *simple features* object that can display the thematic cartography.

The second and third datasets were obtained from the *open geography portal (source accreditation to ONS -* Office for National Statistics licensed under the Open Government Licence v.3.0)*.* The first being the lookup table for the LSOA names and codes in England and Wales 2011. These were added to the CFS data as they only contained the LSOA names. The second were the rural/urban classifications of LSOAs (2011), these were again joined to the calls dataset and provided further contextual knowledge about how missing incidents are distributed across different area. The classification variable was based off the revised version the 2001 census. All output areas (OAs) were coded as urban if they were allocated to a 2011 built-up-area with a population of 10,000 people or more, the remaining were codded as ‘rural’. The LSOA classification is based on the aggregation of OAs, but also identifies smaller hub towns with a population between 10,000 to 30,0000. There were 4 types of smaller hub rural areas and 4 types of smaller hub urban areas but were grouped and recoded into a binary variable representing just rural and urban areas. These will help represent the spatial regimes in the spatial regression models (*See Appendix A for original rural/urban classification*)

The fourth datasets were population statistics also obtained from the Census 2011, specifically obtaining the total LSOA population (n/total population\*1000) and residential population (n/residential population\*1000) to compare different rates of missing incidents. Typically, crime rate is expressed by 100,000 inhabitants but as my unit of analysis is a small area statistic, a rate of 1000 better represents the average population across LSOAs in the U.K. Using rate reduces statistical bias and reduces effect of the modifiable areal unit problem (Wong, 2004). Median Age was also obtained due to literature addressing important differences in age. There were 5 LSOAs in Cheshire East that did not match with the census data and produced NAs, thus imputation of the mean residential population from Cheshire East were used to fulfil the dataset (Little and Rubin, 2002).

In order to examine how environmental and neighbourhood covariates affect missing incidents small area level statistics were used that measure deprivation and mental health separately. The Index of Multiple Deprivation or IMD (2019) provides area level statistics for a relative measure of deprivation across England ranking from 1 (most deprived) to 32, 844 (least deprived areas); additionally. The relative measure for LSOAS is based on seven different domains of deprivation (income, employment, education, health/disability, crime, housing and environment) and are referred to as neighbourhood-level indices. Decile scores are also provided by ranking the 32,844 LSOAs and dividing them into equal size groups from 10, where 1 is most deprived 10% and decile 10 is least deprived 10% of areas nationally.

The Small Area Mental Health Index or SAMHI (2020) also provides an annual measure of population mental health in each LSOA in England from 2013 – 2018. It combines data from multiple health sources including NHS, prescribing data (antidepressants, QQF DWP), incapacity benefit and employment support allowance into a single index. The index was individually standardised via z-scores (using mean of 0 and standard deviation of 1) to obtain the intra-correlations between all indicators. Similarly, decile scores were provided alongside the index from 1 – 10. For the analysis the relative decile scores of both the IMD and SAMHI were used as it was calculated using exponential distribution thus allowing for comparison between areas.

## 

## Measures

### Study Area and Unit of Analysis

The total amount of calls recovered from 2015-2020 were 1,749,123, with missing incidents accounting for 42,547 calls over 903 LSOAs. As 528 calls across 251 LSOAs were mapped outside of the main LAs, they were removed as they would fail to represent the true count of missing incidents in these areas (*view Appendix B).* The total sample was therefore reduced to 42,019 missing incident calls across 652 LSOAs from the four main local authorities.

To clarify, two main datasets were used to explore the spatial and temporal trends. For Theme 1 And Theme 3, the CFS joined to the shapefile and then combines the 6 datasets above. The study’s Unit of Analysis is the LSOA (the total number of missing incidents per LSOA from 2015-2020) and the dependent variable is the rate of missing person (n/residential population \* 1000). Whereas for Theme 2 the empty CFS is used to produce the descriptive statistics, where the dependent variable is the count of missing persons.

### Variables

12 variables from the CFS were used;

**LSOA name**: LSOAs were used to create a simple features object that will explore the spatial trends of missing incidents across the four local authorities. **Local Authority:** LAs used to help identify the boundary districts for the 4 local authorities. **Date and Times**: This variable included the year, date, month and time. (hour/minute/second) and was converted to a date object that allowed to access for any temporal patterns. **Median Response Time:** In order to obtain the median response time, all calls that had been attended were selected and the difference between the *earliest deployed date time* and the *earliest arrived date time* was calculated creating a new time variable representing the average minutes it took to deploy and to arrive to a missing person call. This will help to access the similarities to the GMPs incident response time policy. **Attendance:** This binary variable represents whether the call was attended or not with 1 meaning Yes and 0 meaning no. **Call Origin**: Call Origin, a nominal variable, identified the source of the call and originally included 11 categories (*view appendix C*), however due to low numbers these were categorised into 5 main themes. *Public-non emergency* (which included public-non emergency, ANPR, social media, single online home, alarm company, helpdesk and email), 999, Police Generated, Other Emergency Services and Unknown. **Final classification**: Final Classification, a nominal variable, identified the final description of the missing incident type, and were displayed across over 30 incident types (*view appendix D*). However, these were categorised into just three groups representing *missing persons, absent person* and *other.* The later represents all of the incident’s types that received counts smaller than 30 to reduce noise. This variable will be used to contributed to the discursive platform towards the definitions of missing persons.

**Initial and current grade**: Both grade variables are ordinal. They are used by dispatchers to represent the urgency of the call by giving a score from 1 -5. Grade 1 represent an emergence response (attendance within 15 minutes), Grade 2 is a priority response (within 1 hour), Grade 3 is a routine response (within 4 hours), Grade 4 is a scheduled response (attendance or resolution within 48 hours) and Grade 5 represent a telephone solution. Initial grade would typically relate to resource allocation, whereas current would reflect the actual response required of the call in which used to gain an understanding toward police practices. Grater Manchester Polices’ Incident Response Policy (2017) were used to conceptualise the grades from the CFS.

**Grade Change:** This binary variable was created to highlight whether the grade changed between *initial grade* and the *current grade,* measuredas either True = 1, or False = 0*.* Following the incident response policy above (*ibid,* 2017) the grade typically only changes on two grounds; when 1) the circumstances of the incident change or 2) when the OCB supervisor perceives incorrect grading. This will help provided contextual knowledge about why certain types of calls were re-graded and how certain techniques of police response are most affected.

## Models and Analysis

This paper uses a series of models to analyse underlying concepts amongst the three main themes; these were all conducted using the software R.

**Theme 1 explores both the spatial and temporal trends of missing incidents.** Local indicators of spatial autocorrelation (LISA) maps will be used to indicate the extent to which a significant spatial clustering of homogenous values exist in the study area. LISA has two characteristics; the first it provides a statistic for each location while providing an assessment of statistical significance and secondly, it provides a statistic for the relative relationship between the sum of the local statistic and its corresponding global statistic. Spatial autocorrelation statistics combines a measure of attribute similarity between two observations, using spatial weights which act as an indicator for locational similarity. All Moran statistics are based on a *contiguity-based spatial matrix,* in this paper I use the ‘*the queens’ criteria* as it accounts for all areas in a polygon whereas the ‘*rook criteria’* does not include shared corners reducing the distances in our LISA maps (Anselim, 1995). The LISA maps make connection between significance and the Moran scatter plot, which give you four types of spatial autocorrelation known as high-high and low-low (hot spots and cold spots; spatial clusters) and high-low and low-high (spatial outliers). To examine temporal trends decompositions methods will be used to examine the seasonality and noise components of the data, specifically examining seasonal variation and changes in the autocorrelation. Lastly seasonal autoregressive moving average (SARIMA) models will be used to answer whether the pandemic has affected the rate of missing incidents as a result of changes to routine activities. This involved forecasting data in order to make predictions about what would have happened in the absence of Covid-19 accounting for overall trends and noise.

**Theme 2 will analyse the changes in police responses through univariate, bivariate and advanced time-series models.** Univariate statistics examine how each characteristic of police responses vary in their mean, median and range before examining the association to other variables. Bivariate analysis will help to establish whether there is a statistically linear relationship between police response and the number of missing incidents. These were assessed using Chi-square to test for homogeneity. Time series models were used to highlight how police responses vary over time. The interval of time was measured at weeks as using daily counts created too much noise; these were represented using a conditional mean (Hyndman and Athanasopoulos, 2018).

**Theme 3 will use explore the effects of the neighbourhood covariates through Poisson and spatial regression model.** Negative binomial regression, a form of Poisson regression that allow for the overdispersion of variance, will provide rate ratios on how more or less common the different deciles are compared to within the deprivation statistics and within the mental health statistics. A spatial regression model allowed for the examination of how missing incident rates varies systematically over space due to its correlation with the environmental correlates. In spatial regression models, the standard properties of OLS do not hold as the assumption states that values of the coefficients of the variable’s areas constant across the spatial elements, which would confine any spatial variation to the error term. However, a model assuming that observations are independent are incorrect therefore spatial regression models allow for spatial autocorrelation. Spatial autocorrelation accounts for the level of spatial randomness; measured both through global/local estimates and spatial regression (testing for heterogeneity). Two spatial regression model were examined, one over the deprivation statistics including the median age of each LSOA and one over the mental health statistics again including the median age. From both models the urban and rural LSOAs (spatial regimes) were examined as research would indicate increased clustering in the more densely populated areas so we can test to see if this is true with the presence of missing incident rates.

Before running the spatial regression models for spatial heterogeneity, two non-spatial regression model was run where the residuals (difference between observed and predicted value) provide a picture of whether missing incident rates are over or under predicted in our study area. Moran’s I test were used on the regression residuals to test for statistical significance, if the residuals (errors) are significant as in related systematically among themselves, then these provides guidance in whether spatial regression models are appropriate amongst the urban/rural LSOAS (also known as spatial regimes). In order to produce better models, spatial dependence was incorporated in the regression models. The Lagrange Multiplier Test will provide the most appropriate model; either a spatial error model which incorporates spatial dependence through attributional dependence, or a spatial lagged model that treats spatial dependence by adding a spatially lagged variable to the regression equation

## Limitations

The biggest challenge with spatial analysis is that the accuracy of spatial data is obscured by geo-masking techniques that served to protect the location privacy of the victims. In the CFS dataset for example only partial postcodes are given, although location coordinates have been obtained via joining to the LSOA shapefiles, we are still restricted to spatial point patterns in an area/polygon rather than analysing spatial point patterns along networks. Typically, in criminal justice applications crime is geocoded alongside a linear street network, therefore sometimes non-random spatial point process is masked by the two-dimensional space (Tompson et al., 2015). It is necessary to address the issues with geo-masking as the reliability of the findings can be questioned. The use of examining spatial regimes however, aim to add substances by contextualising the study area over environmental correlates.

Another weakness with the CFS data is the failure to capture repeated measures reducing the opportunity to build a case review on repeat location. The UKMPU (2016) reported that 48.5% of reports are attributable to ‘repeat missing’ in the U.K. with a higher proportion of reports among children. For a child they will be classed as ‘repeat missing’ if they have been reported missing three times or more in a 90-day period. These repeated reports raise the issue of the duty of care and responsibilities from police forces. Although unable to study the effect of repeat missing incidents on the proportion of calls received, this paper will still address the importance of GIS in the understanding of missing incident trends and highlights the importance of quantitative studies in adding to the overall picture of missing persons. Moreover, this study is based on data from only Cheshire Police and raises questions about generalisability across other police forces, specifically in regard to differences in police response to missing incidents. This is not to say that these findings cannot resonate nationally, but should be with caution. Moreover, CFS, as with crime rate data, are still affected by reporting statistics among certain demographics of the population. Graber and Stern (2018) highlighted that to call the police is a privilege of being white, additionally police legitimacy can also affect the willingness to call the police (Taylor et al., 2015). Therefore, not all populations are likely to call the police so reports of missing incidents raise questions about reliability from our sample size. Arguably, there is no perfect measure of crime incidence but CFS provide a dataset that is not just measuring crime, but also questioning the practices police and public services.

The SAHMI dataset also has some limitations, the first being that years do not overlap with the years in the CFS removing the possibility to examine 2019 and 2020 which may have been crucial in understanding the effect of mental health over the pandemic (Brewin and DePierro, 2020). Additioannly, its index measure is only created from four sources specifically from those who receive medical help from the NHS or incapacity benefits, largely ignoring a large proportion of the population who are undiagnosed. Mental health research is challenging given the apparent limitations of both ethical principles and data-collection. Nevertheless, SAMHI offers one of the first longitudinal placed-based data resources replacing the Indices of Deprivation Mood and Anxiety. Disorders. The IMD has been used constantly in amending public policy, guiding resource distribution and contributed to the housing markets. However, increasing literature is drawing debate on its elements of methodology that could systematically underscore the level of deprivation in big cities due to techniques such as double counting and weighting applications, but also due to the disbursement of urban police regeneration resources (Deas et al., 2003).

## Ethics

The variables from the CFS dataset were requested as part of an N8 research project named *understanding changing demand for police during the coronavirus pandemic,* the dataset had become available due to my presence in the project as a research assistant. Following this, all ethics have been considered and approved by The University of Manchester; reference approved by *2020-11031-17216.*  Although geo-masking is a problem for spatial accuracy, by anonymising postcodes victims are provided a level of privacy that protects them among academic and policy reports. On a conceptual level this paper and the CFS can be critiqued for its use of the definitions surrounding ‘missing’. The final categorisation variable still includes the category of absent person even though this was removed from policy in 2017, additionally many of the categories has been sub-grouped under other incident types. When analysing missing incidents, it is important to view missing persons as individual beings that do not have their individual experience removed following Biehal’s (2003) definition of ‘going missing’.

(p.s I then include a frequency table for the CFS variables at the start of findings section)