RESEARCH ARTICLE

Tracking Violent Crime with Ambulance Data: How Much Crime Goes Uncounted?

Alex Sutherland[1] · Lucy Strang[2] · Martin Stepanek[3] · Chris Giacomantonio[4] · Adrian Boyle[5] · Heather Strang[6]

Abstract

Research Question What proportion of ambulance records documenting injuries caused by criminal violence is included in police records for violent crimes occurring in the same area at the same dates and times as incidents found in ambulance records?

Data We analysed subsets of three datasets during matched time periods: West Midlands Ambulance Service records of all 36,639 incidents of violent injuries from January 2012 to March 2017; 132,317 West Midlands Police records of violent crimes from January 2012 to December 2015; and 9083 records of treatment of violent injuries as recorded in hospital Emergency Department (ED) records covering September 2013 to March 2016.

Methods We compared all incidents in the ambulance dataset and ED data to corresponding locations and times in incidents recorded in police datasets.

Findings Approximately 90% of cases in the ambulance dataset did not have a corresponding case in the police dataset. The proportion was even lower in the Emergency Department dataset, where less than 5% of cases were successfully matched to a police record. These data suggest that adding the medical data to the police data could add 15 to 20% more violent offences to the totals recorded by the police.

Conclusions Tracking identified ambulance data can add substantial numbers of serious violent crimes, over and above those reported to the police. These added cases can increase the targeting of police and public health resources to prevent harm against victims, at places, and by offenders at highest risk of serious violence.

Keywords Violence · Injury · Ambulance · Reported crime · Crime statistics

✉ Heather Strang
hs404@cam.ac.uk

Extended author information available on the last page of the article

🖄 Springer

Introduction

The use of medical records has long been known as a potential source of assessing the completeness of reported crime data. Yet that potential remains undeveloped across most police agencies. The rise of evidence-based targeting of highest-harm locations, offenders and victims (Sherman 2007, 2013) makes the value of comprehensive records more important than ever. The cost of creating new systems for checking police data against medical records becomes minor in comparison to their value in selecting the most important targets for prevention of serious violence. More complete data on violent crime could, in principle, literally help police to save lives.

This article demonstrates a proof of concept for using ambulance data to expand the "catchment" of all violent events into police records, by names of victims and place of occurrence of injury. The fact that the data can be identified is a crucial step forward in police use of medical records for violence prevention in the UK, which has been limited largely to de-identified victimisations reported in identified locations (Shepherd, 2004).

The potential for identifying victims most at risk of violent crime harm has recently been shown to be substantial, with under 4% of victims in Dorset experiencing 85% of the crime harm recorded by police in 1 year (Dudfield et al., 2017). Given the very high Crime Harm Index scores (Sherman et al., 2016) of some violent offenders (Liggins et al., 2019), the omission of even one crime from the list of crimes against one individual can skew the rank-order of victims by total harm. That, in turn, could mean the omission of a highly harmed victim from the list of people receiving the greatest attention from police units engaged in proactive safeguarding.

This article shows that a very high proportion of violent injuries documented in rescue service records in one urban area was not recorded in police records. It reports on the methods used to reach that conclusion. The same methods could be applied to the ongoing tracking of all ambulance records of violence against police records that should correspond to each and every case. Tracking all ambulance cases is a standard that could be set for police agencies themselves. Tracking a sample of ambulance cases is a method that could be deployed by policing oversight bodies, such as Her Majesty's Inspectorate of Constabulary, Fire and Rescue Service, which to date has apparently not used this method of assessing the completeness of police crime records.

Research Question

The research question for this study is this:

> What proportion of ambulance records documenting injuries caused by criminal violence is included in police records for violent crimes occurring in the same area at the same dates and times as incidents found in ambulance records?

Data

In May 2016, West Midlands Ambulance Service (WMAS) and West Midlands Police (WMP) provided to RAND Europe 3 years of data (2012–2013 to 2014–2015), relating to all violence-related calls for ambulance service in Birmingham, Coventry and Wolverhampton. WMP also provided an Emergency Department (ED) dataset covering the same period. All three datasets were limited to incidents involving interpersonal violence.

The WMAS provided both the computer-aided dispatch dataset, produced automatically at each ambulance call-out, and the Patient Report Form dataset, which is recorded on paper by paramedics on the scene of an incident and then read by text recognition software. The data for police recorded crime and the ambulance data cover 2012–2017 in Birmingham, Coventry and Wolverhampton. ED data was from September 2013 to March 2016.

WMAS (Ambulance) Dataset

This dataset allocates each incident to a specific time and postcode. Given that each postcode covers approximately 15 addresses, the referencing is broadly at the same level of precision as geospatial coordinates, but the actual geospatial area covered by a postcode varies by the density of housing in a given location.

All reported cases are classified as assault (domestic as well as public), stabbing or shooting, and subdivide by severity (though classification of severity has changed over time and its comparison is therefore limited). Overall, the dataset covers 36,639 incidents over the January 2012 to March 2017 period and includes the categories shown in Table 1.

In terms of data quality, ambulance data is consistent over time, automatically generated during 999 calls for service. With a few exceptions, most variables were

Table 1 Ambulance dataset variables

Variable	Description
Date	Date of incident
CAD_ID	Automatically generated case ID
Age	Age of injured person
Gender	Gender of injured person
Postcode	Full postcode of place where the ambulance attended
Chief complaint	Type of incident: assault, firearm, gunshot or stabbing
Symptom group	The nature of the injury[1]
Priority	Category of threat to life
CCG name	Clinical Commissioning Group area where the ambulance attended
Caller	Which service initiated the call to the ambulance[2]

[1] Over 20 separate categories of injury type, such as abdominal pain or eye injury were found in the dataset

[2] Mostly a police service, although some other categories include the 111 service, Emergency Operations Centre (representing emergency calls by the public) or other agencies

recorded consistently. Specifically, no incidents had missing data on postcodes, type of incident or severity classification. While there may be a delay between an incident happening and call for ambulance, the time of connecting the call is captured precisely from a digital clock, giving a reliable indication of when an incident was reported. Data on gender and age were less consistent, with approximately 14.4% and 34.5% of data missing, respectively. Similarly, the "Caller" variable is often missing (in 44.2% of cases). A significant unanswered question is whether ambulance pick-up location is the same or similar to the location of where the violent incident actually occurred; this question needs to be explored in future research.

The Emergency Department Dataset

The ED dataset contains 9083 records covering the September 2013 to March 2016 period. Similar to ambulance data, it has information on patients' age and gender, and also contains data on ethnicity, type of weapon used and whether alcohol played a role in the incident. On the other hand, it lacks information on severity of injuries. The full list of available variables is presented in Table 2.

The main issues with the ED dataset stem from records being self-reported and manually logged. There is also a time lag between incidents taking place and hospital reception, and inconsistency between place of incident and place of reporting. In particular, there may be up to several days between an incident happening and the patient being treated at the ED, which naturally leads to possible errors in reporting incident time and location. Furthermore, incident day is not captured as actual date but rather only as day of the week; there is therefore no information on whether the incident

Table 2 ED dataset variables

Category	Description
ID	Patient ID (only for identification within the dataset)
Age	Age of patient
Gender	Gender of patient
Ethnicity	Ethnicity of patient (Asian, black, mixed or white, with additional groups)
ED department	Name of department assisting the patient
Location	Type of incident location (over 20 categories such as pub, car or street)
Address	Address of location where incident happened, including open text description such as "at home" or "in kitchen"
Date logged	Date of treatment at ED
Incident day	Day (Monday–Sunday) of incident
Incident time	Time of incident
Alcohol	Indication whether patient consumed alcohol prior to the incident
Reported police	Indication whether the incident was reported to police or not
Weapon	Weapon used in the incident (over 10 categories such as knife or body part)
Relationship	Relationship between the patient and the attacker (e.g. stranger, friend)
Residential postcode	Outer part of patient's residential postcode (e.g. CB4)

happened even in the same week as it was reported. The open text field describing nature of the incident may also be filled differently depending on the ED receptionist.

From the perspective of using ED data for analysis, the most problematic issue is the high proportion of missing location data (over 25%) and imprecision even when location is known. The notes often indicate patients were not certain *where* an incident actually happened. Combined with a degree of uncertainty about *when* the incident took place, the records are therefore difficult to be matched with the police dataset. The initial analysis of the ED, ambulance and police datasets suggested very low overlap between the ED and other datasets (less than 5% of cases in the ED dataset could be assigned to a matching incident in the police dataset) and we therefore do not include the ED dataset in the final results.

Police Recorded Crime Dataset

The police dataset, covering January 2012 to December 2015, consists of 132,317 records from the WMD CRIMES system, which were automatically created whenever a police patrol was called to an incident. The reports could have been made by victims, witnesses, police officers and staff or other third parties, and include recorded crimes as defined by the Home Office Counting Rules (i.e. not necessarily violence as described by other organisations such as the World Health Organisation). Unlike the other two datasets, each record includes an exact location of the incident entered as geographic coordinates, allowing more detailed analysis. The dataset available for this project consisted of the variables given in Table 3.

As it is automatically generated and without any form of open text, the WMAS dataset was consistently recorded. However, it did not provide any details on the number of assailants in incidents or their personal profiles that could be matched with incidents in other datasets. Hence, the only means of identification of overlapping cases were location and time of incident. Offences involving more than one victim were recorded as separate cases.

Data in all datasets were anonymised before being shared. In addition, data in all analytical outputs were carefully assessed in terms of specificity to prevent possible identification of particular victims; all summary results with geographical and time

Table 3 Police call-out dataset variables

Category	Description
Police department	Location of responsible police department
Crime reference	Unique identifier of incident
Date	Date of incident
Coordinates	Geographic coordinates of incident location
Offence	Detailed description of offence (over 230 individual categories, subsets of broader categories such as assault or rape)
Offence type	Broad category of offence
Weapon used	Indicator of weapon use
Time	Approximate time period (start and end time) of incident and exact time of reporting to the police

information were deleted if they contained less than three observations per location, i.e. whenever it was possible to identify a specific location in a specific time period, such as a postcode in any given month, all data points consisting of less than three observations (e.g. the number of armed assaults) were not reported.

Although all datasets generally covered the same geographical areas, the exact locations were not identical. In particular, the location of incidents reported at the ED was dispersed beyond the city limits (with several cases in different regions), perhaps as a result of individuals seeking treatment at their local ED sometime after the incident. There were also differences in volume and geographical distribution of cases between the datasets. In particular, there are more than four times as many incidents recorded in the police dataset as in the ambulance dataset—more than 130,000. Overall, ambulance and police data mostly cover the same areas, whereas ED data is more sparsely distributed.

Methods

During the first year of the project, the primary obstacle was agreeing upon a satisfactory approach to data sharing between WMAS, WMP and RAND Europe that would allow the project to proceed. Underpinning these discussions was a concern around sharing patient-identifiable data with the police, as this could be used to inform criminal investigations or to pursue arrests/charges. It was agreed that the RAND Europe research team would act as processor of potentially patient-identifiable data to create anonymised/aggregated outputs for use by WMP and WMVPA. Furthermore, the final outputs were datasets aggregated to neighbourhood level,[1] with low counts (below three) omitted. Following approximately 10 months of discussions, a data sharing agreement was established between WMAS and RAND Europe in April 2016.

As part of the project, RAND Europe provided outputs from the WMAS dataset to police analysts and West Midlands Violence Prevention Alliance (WMVPA) partners over a 6-month period to explore its utility as a source of intelligence in violence reduction or prevention activities in the West Midlands.

As a first step in the analysis, the available data was cleaned and recoded. Thereafter, the main data processing tasks consisted of the following:

- *Conversion of geographical information* was done to allow postcode referencing. Each set of coordinates in the police data was assigned a postcode using publicly available online resources[2] and analogously each postcode in the ambulance data was assigned a set of geographical coordinates corresponding to its centre (where postcode information was available). This means that for every incident we collected postcode information and a single set of coordinates relating to the middle of the postcode the incident was recorded in.

---

[1] Lower Layer Super Output Areas (LSOAs).
[2] Grid Reference Finder and Postcodes.io online applications, available at https://gridreferencefinder.com/ and https://postcodes.io, respectively.

- *Classification of time* into 1- to 4-h categories, i.e. essentially reclassification of the time variable from continuous (hours and minutes) into categorical (hours or longer time periods) in order to simplify later analyses.
- *Location filtering* so that all remaining data points would be within the city limits. For postcodes, this was done using the outer parts of the postcode; any postcode not starting with "B", "CV" or "WV" (standing for Birmingham, Coventry and Wolverhampton, respectively) was excluded from the analysis. For coordinates, this was done either after matching each coordinate with a postcode in the same way, or manually using QGIS software[3] by deleting all data points visibly outside of the city boundaries
- *Classification of variables* into consistent categories. This involved particularly categorisation of offences in the police dataset into a more limited number of categories.[4]
- *Removal of missing data and duplicates* where necessary. Since the police and ambulance datasets were consistent, less than 0.1% of data needed to be excluded. Cases with missing/clearly wrong date or location were excluded (less than 0.1% of data).

Subsequently, the police and ambulance datasets were combined with coordinates, postcode, date and time of incident available for all remaining observations, though dataset-specific variables such as type of offence or gender were available only in the respective parts. The combined dataset spanned the period of January 2012 to December 2015—the largest timeframe available in both individual datasets.

Analysis

Analyses were undertaken in three steps: *geographical analysis*, *analysis of frequencies* and *analysis of overlap*. The geographical analysis consisted principally of data visualisation using QGIS software. Analysis of frequencies then looked separately at the three sites as well as individual postcodes or their groups and showed the number of incidents across the following dimensions: time (day of week, time of day, month and year), location, type of offence, type of weapon used, severity of injuries and characteristics of the injured individuals. The number of incidents per location was then used in identification of so-called hot postcodes reported in the ambulance dataset and in the police dataset within the same time period.[5] These hot postcodes indicated that, regardless of overlap between the two datasets, some of the incidents were not reported to the police.

The analysis of overlap was done in two ways: using postcode referencing and spatial nets ("as the crow flies" distance matching).[6] The matching variables consisted of time, location and number of observations, but differed across iterations of the analysis:

---

[3] See QGIS (2017).
[4] The original 231 categories (including categories such as abduction of child by other person) were recategorised into 79 broader categories such as assault, attempted rape and attempted robbery.
[5] That is, counting two and more patients treated in relation to one incident as a single case.
[6] Both analyses were done in Stata 13 using the "merge 1:1" command, which looks for matching pairs of data in two specified datasets and records whether any data point in either of the two datasets was matched or not.

- *Time matching*: Incidents that happened within the same 1-, 2- or 3-h periods were considered "matched". We also tested longer time periods—up to 24 h either side of the incident—but these did not yield improved results and led to increased probability of matching unrelated incidents. Additionally, we looked for incidents that happened at the same time but on the previous or next day (if not matched on the same day). This is because an incident in one dataset on Monday at 11.59 p.m. might appear in another as 12.01 a.m. on Tuesday.
- *Distance matching*: Incidents within the same postcode or within 30 m × 30 m, 100 m × 100 m and 200 m × 200 m spatial nets were considered matched. Note that the 200 m × 200 m areas are relatively large and may cover multiple postcodes particularly in densely populated areas.
- *Number of observations*: In both the police and ambulance datasets, each incident was recorded separately. In some cases, there was more than one incident per specified timeframe and area within a dataset; each of these was assigned a unique identifier so that if there were more matching cases in one dataset than in the other one, only the correct proportion would be matched. In other words, for an incident "A" from the ambulance dataset, we may find two matching incidents in the police dataset. Hence, one is considered matching and the other one is not. Now if incident "B" in the ambulance dataset happened nearly at the same time and place as "A", only the unmatched incident in the police dataset would be considered appropriate.

Depending on the choice of geographical and time windows, any two records that happened at the same time and place (i.e. within the same spatial and time window) were considered "matched". Given that the police dataset should, in principle, contain all cases of assault, the matching was done in one direction only, looking for cases in the police dataset matching those in the ambulance dataset.

Interestingly, over one thousand cases in the police data (approximately one-third of all cases that were matched with incidents in the ambulance dataset) could not be matched with an incident in the ambulance dataset within a specified time frame on the same day, yet they could be matched with another incident that happened 1 day earlier or 1 day later that were not matched to any other record in the police dataset. Eighty per cent of these cases were recorded 1 day later in the police dataset, suggesting a date reporting error in either the police or ambulance data. We recommend that some more detailed work is done by the police and ambulance service to check overlap on a sample of cases individually, but note this could be automated with the right investment.

Methodological and Technical Issues with Using Ambulance and Police Datasets

Given the data limitations described above, the main methodological issues using the ambulance and police datasets consisted of (i) assigning postcodes to grid references and (ii) specifying matching terms. Specifically, the Grid Reference Finder online tool offers a batch conversion method, allowing theoretically unlimited numbers of postcodes to be transformed into geographical coordinates. In practice, this is limited to several hundreds of postcodes per attempt due to technical limitations. A reversed batch conversion tool exists

only in API format[7] and therefore requires programming techniques to connect automatically to the server and request postcode information for a prepared set of locations. This was done in Excel VBA but the technical requirement for this task might be a barrier for police analysts (which could be overcome if coordinate data is provided for ambulance incidents, or a mid-way technical solution would allow police to see but not download ambulance data.)

Specifying matching terms was therefore a methodological rather than technical issue, as there is no single correct answer as to the best way the area or time should be specified. There is a trade-off between matching success rate and reliability of matches. Increasing the geographical and/or time frame around incidents for matching will lead to additional cases in the matching dataset being considered and therefore a greater chance of matching but there is an increased risk of matching two unrelated incidents. This risk of incorrect matching is higher in areas with higher crime rates, because multiple unrelated incidents could happen within a short time period of each other.

Findings

The purpose of this analysis was to assess the extent of overlap in the location and timing of incidents recorded in each of the three datasets. The primary reason for doing so was to understand whether ambulance data makes a unique contribution to understanding where and when violence occurs, over and above existing data.

Geographical Coverage of Different Datasets

The geographical allocation of recorded incidents in Wolverhampton city centre in Fig. 1 reveals several interesting findings. First, the overall geographical coverage seems to be very consistent across datasets, with higher frequency of incidents in the densely populated areas. Second, ambulance call-outs seem to be more evenly distributed than the offences attended to by police, with blue dots highly concentrated around the main city buildings and in several other neighbourhoods, whereas the green squares are less consistently distributed around the city centre. Lastly, there are some instances of incidents reported in the ambulance data that do not have a corresponding record in the police dataset, i.e. not only the incidents do not match in the time-geographical space but they are also isolated geographically. This may be due to errors in reporting but otherwise suggests that some crime-related areas may not be captured in police data.

Descriptive Statistics and Trends

The number of incidents occurring on weekdays is substantially lower than on weekends in both the police and the ambulance datasets, yet there is an interesting difference in the distribution for Saturdays and Sundays. Slightly more incidents in the ambulance dataset happen on Saturday than on Sunday (and there are also more incidents on Friday than on other weekdays), but the opposite trend can be seen from the police dataset.

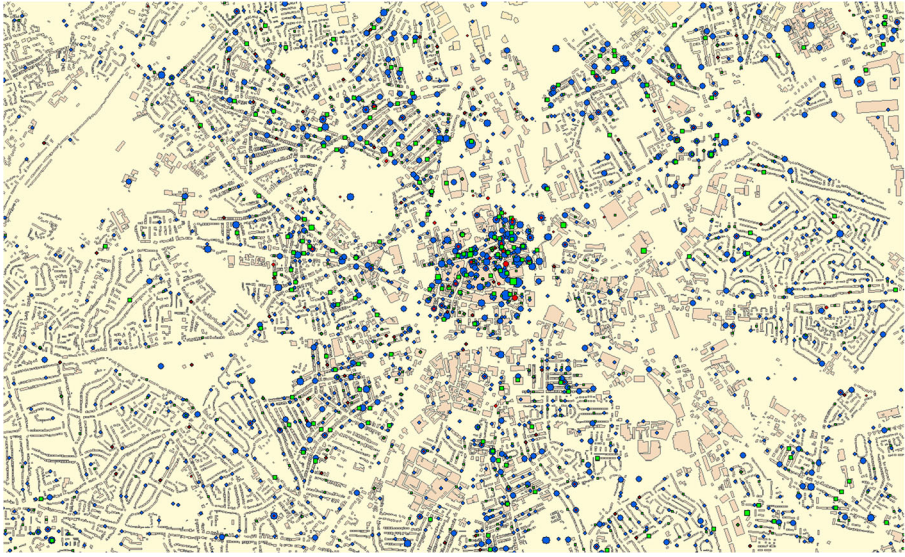
[7] See Bulk Reverse Geocoding, Postcodes (2017).

Fig. 1 Representation of cases in Wolverhampton city centre 2012–2015; police (blue), ambulance (green) and ED (red) datasets

This discrepancy may be a result of multiple factors, particularly many incidents happening at around midnight and being reported and/or recorded at slightly different times to the police and ambulance. There may also be different types of incidents being reported to the police, or even a systematic error in reporting of incidents in either of the datasets. The differences may also partially explain the emerging pattern in many incidents in the ambulance dataset being matched with police records from the next day.

We compared the data with breakdown of incidents by time of day: unsurprisingly, the distribution of incidents in the ambulance dataset is substantially smoother than the police dataset, suggesting that the call times are reported precisely. On the other hand, there are large differences in the number of incidents at different times, with just a very few incidents being recorded in the morning and most of the incidents happening at night. The distribution of offences in the police dataset by contrast is less uniform with several large spikes around lunchtime, mid-afternoon and midnight. (These correspond with, for example, the school day finishing at around 3 p.m. and pubs closing at around 11 p.m.–12 a.m.)

The distributions over a calendar year look similar in both datasets, with slightly more incidents occurring during the spring and summer months but generally rather uniformly distributed. Overall, the results appear mostly consistent. Yet there are several notable differences between them. These differences suggest either that (1) there are some systematic errors in reporting, or (2) the ambulance dataset provides substantial value added to the analysis or (3) both. In particular, the relative uniformity of distribution of incidents recorded in the ambulance dataset (compared to spikes in the police dataset) suggests that either the two datasets capture different incidents or that incidents in one of the datasets (arguably the police dataset given the large spike) are not captured at the time they happened.

Finally, the number of incidents in the original police, ED and ambulance datasets over time is depicted in Fig. 2, showing a slight decrease in ambulance call-outs and broadly stable number of assaults recorded in the ED dataset, yet an increasing trend in the police dataset in the period that data was available. These data suggest that adding the medical data to the police data could add 15 to 20% more offences to those recorded by police.

Looking specifically at incidents from the April 2016 to March 2017 period in the ambulance dataset, depicted in Fig. 3 and Fig. 4, we can see that the majority of assaulted individuals were men (over 60%), and the most frequently assaulted age group is those aged 16–30. The share of missing gender information is relatively low, at around 11%, though the share of missing age information is higher, above 30%. Similarly, the severity of incident and classification of assault categories are dominated by "C3 Red" category and general assaults, followed by stabbing in 10% of cases and very few gunshots.

To What Extent Do Police, Ambulance and ED Incidents Overlap?

Previous research on overlaps between police data with ED and ambulance data has produced a small number of findings on the degree to which data may be matched. A pilot Injury Surveillance Unit in NHS Lanarkshire found that 49% of violence-related injuries recorded by the ED were not found in police data (Reform, 2014). Another study analysed hotspots of community violence–related calls for service for the police and ambulance service in Peterborough, England, over a 12-month period (Boyle et al., 2013); that study found that only 62% of ambulance calls were present in the police data, and on average a 50% overlap in the two services' respective hotspots of violence. In addition, only 8–50% of ambulance call-outs were found to have been transported to hospital, indicating that there are a large number of incidents not recorded by the police or the ED. Similarly, a cross-sectional study conducted by the Trauma and Injury Intelligence Group (TIIG) in North West England examined the nature, extent and characteristics of ambulance call-outs for violent incidents for the period April 2013 to March 2015. TIIG used information collected by the North West Ambulance Service and found that around one-third of ambulance calls for service in the region were not
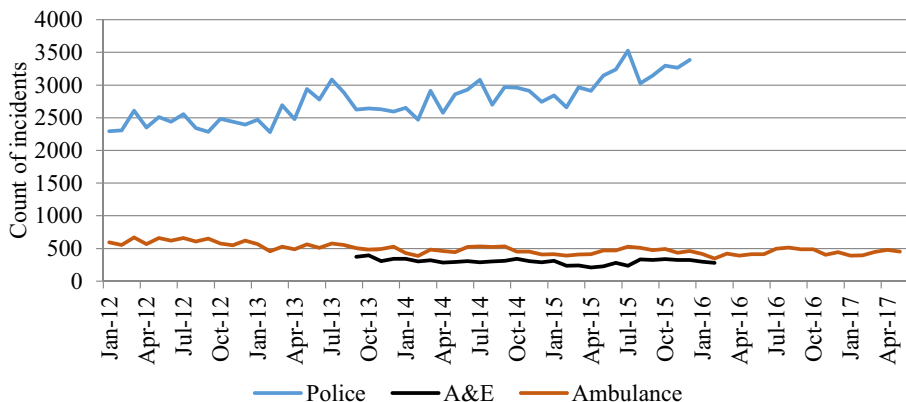

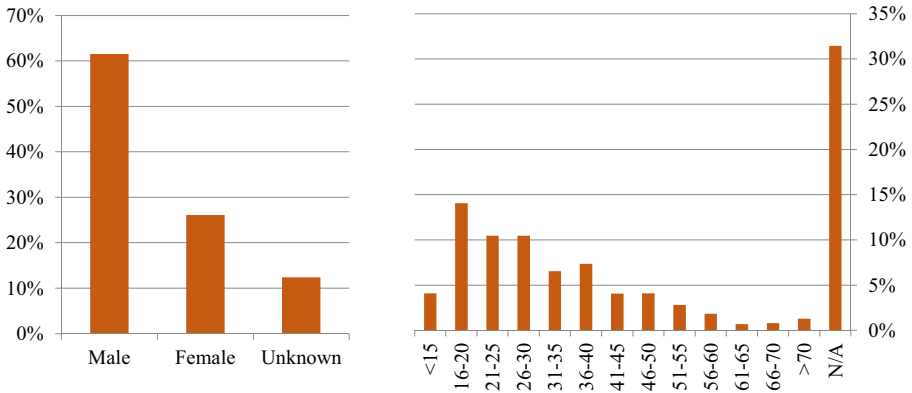**Fig. 2** Number of incidents in police, ED and ambulance datasets over time

Fig. 3 Gender and age distribution in the ambulance dataset, April 2016 to March 2017

transported to the hospital. This also suggests that a substantial proportion of assaults are not recorded in ED data. There are many ways to assess overlap between incidents, and that is part of the challenge of cross-validating these data sources. The approach taken here was to select an incident in the ambulance data and then attempt to find a match in the police and/or ED data.

Our approach was to look at both the temporal (time) and spatial (location) overlap in different ways. First, we analysed overlaps using 6-h time nets (3 h either side of the incident time recorded in the ambulance dataset) and postcode referencing. This resulted in only 10.2% overlap (as percentage of all incidents in the ambulance dataset) between the ambulance and police datasets, meaning that approximately 90% of cases in the ambulance dataset did not have a corresponding case in the police dataset using this approach. The proportion was even lower in the ED dataset, where less than 5% of cases were successfully matched. Further detail is provided in Table 4.

Interestingly, increasing the time range for matching does not result in better results (e.g. increasing the time to 6 h to either side increases the overlap by just 0.4%), but it does increase the probability of (incorrectly) capturing two independent incidents.
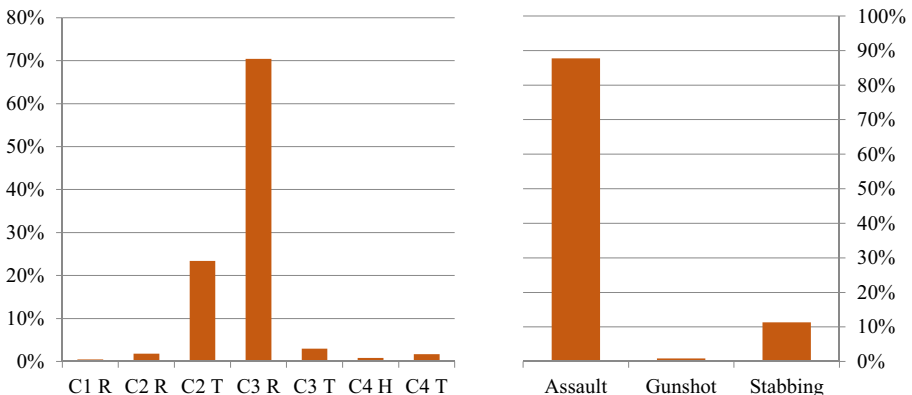
Fig. 4 Severity of injuries and type of offence in the ambulance dataset, April 2016 to March 2017

Table 4 Overlap between the police, ambulance and ED datasets using postcode referencing and 6-h time nets

	ED in police	Ambulance in police
All	4.6%	10.2%
Birmingham	4.3%	10.4%
Coventry	6.5%	10.1%
Wolverhampton	4.8%	9.6%

Recognising that the postcode information in the ambulance and police datasets may differ slightly for the same incident,[8] we then used an alternative geographical referencing looking for incidents that happened within a given distance (measured "as the crow flies"). We also tried different geographical specifications for the search, expanding the geographical "nets" around each location from 15 m (i.e. using squares of 30 m × 30 m), up to 100 m (resulting in 200 m × 200 m squares). The declining marginal increase in the overlapping incidents with increasing size of spatial nets suggests that a net with cell size of 100 to 200m provides the best trade-off between reliability and matching. This is represented in Table 5.

It remains unclear whether the increasing percentage of overlap is really capturing the same incidents or independent events that share similar date/time/locations. Combined with the fact that most incidents happen at weekends and during the night suggests that accidental overlap could be quite high. Leaving these points aside, even at 34% the overlap between police and ambulance data is relatively low. This suggests that the ambulance dataset may provide large amount of new information on otherwise unreported incidents but also that there may be systematic errors in the reporting that need to be identified.

One source of potential "error" that needed further attention was police-initiated calls for service. Approximately 55% of calls in the ambulance dataset originate from a police source, but the research team was only able to link up to 34% of ambulance cases to incidents in the police dataset. As such, we need to understand more about police-initiated calls and how they might affect the overlap.

During a workshop with stakeholders from the WMVPA, the police gave several reasons for why there might be so many police-initiated ambulance call-outs that were not logged in police data. For example, police may be called to an incident to find an injured person and then initiate a call for an ambulance, or an individual may be injured resisting arrest. Another reason given was that calls could originate from police stations as a result of people already in custody requiring medical attention.

To check where police-initiated calls were located, we mapped these against the location of police stations in the data using police station postcodes. This revealed that of the 55% police-initiated calls (17,813 of the total 32,429 incidents that happened within the city boundaries, collected as of June 2017), 6.1% (1087 incidents) were located in the same postcode as a police station. This provides evidence *against* the suggestion that police-initiated calls are because of injuries discovered in police

---

[8] For example, due to an incident happening close to a postcode boundary.

Table 5 Proportion of ambulance and ED datasets incidents found in the police dataset using varying spatial nets

Size of spatial net	Dataset	Overlap
30 m × 30 m	Amb	9.5%
	ED	3.1%
100 m × 100 m	Amb	24.3%
	ED	11.5%
200 m × 200 m	Amb	34.0%
	ED	20.0%

Note: Spatial net size with 6-h time window

custody. It does, however, suggest further work with both ambulance and police is needed to understand this.

How Can Ambulance Data Be Used by Police? The Example of "Hotspots"

Over the past 25 years, the concept of crime "hotspots", or spatio-temporal concentrations of criminal activity, has gained traction in criminal justice research, with evidence accumulating that the majority of crimes are committed in concentrated areas such as a street corner (Sherman et al., 1989; Braga and Weisburd, 2010; Weisburd et al., 2010; Weisburd et al., 2014). One of the best known of these studies found that over half of all police calls for service in Minneapolis were for fewer than 4% of addresses within the city limits (Sherman et al., 1989).

As a result of these findings, police forces have increasingly sought to focus their resources towards these hotspots rather than general, community-wide policing activities. The benefits of this approach are twofold: police response times may be minimised as patrolling officers are more likely to be close to the location of the call for service; and the enhanced visibility of the police in areas of high crime may act as a deterrent to potential offenders and prevent crime from taking place (Sherman and Weisburd, 1995). Indeed, recent meta-analyses have indicated that hotspot interventions by police are effective as a crime prevention strategy, with between a 15 and 25% reduction in crime (Braga et al., 2012; Braga, Welsh et al. 2014). Further research has also established that reductions are also found when utilising Police Community Safety Officers in hotspot interventions (Ariel et al., 2016).

Research using ED attendance and hospital admissions data for assault-based injuries has produced similar findings (e.g. Warburton and Shepherd, 2004), although there is not yet a body of evidence on this approach. [9] It has been noted in the literature that ambulance call-out data holds significant potential for the identification of violent crime hotspots, as ambulance services typically collect detailed and highly accurate information about the location of such incidents (Weinborn et al., 2017; Taylor et al., 2016). While the use of ambulance data for this kind of analysis is still emerging, and

---

[9] See also ISTV https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/isb1594-information-sharing-to-tackle-violence-minimum-dataset

little is known about the efficacy of data sharing between emergency services, a 12-month study of hotspots in police and ambulance calls for service in Peterborough found that these locations were equally concentrated in a small number of locations, with at least half of ambulance hotspots unknown to the police (Ariel et al., 2016).

Following these research findings, we have calculated the extent of crime concentration across the police and ambulance datasets in our study for Birmingham, Coventry and Wolverhampton (Table 6). For the purpose of this analysis, we pooled data from the geographical locations for each dataset (e.g. the figures for police are for all three sites). For both police and ambulance data, 10% of postcodes accounted for around 40% of incidents.

Figure 5 shows visualisation of "hot postcodes" on a map of Wolverhampton, plotting all incidents involving victims aged 21–25 in the city, differentiating frequency of incidents by size of circles on the map. Such a subgroup analysis may be particularly useful in targeting, for example, a certain vulnerable group of individuals. Note that each location represents a centroid of the relevant postcode as the ambulance dataset does not offer detailed coordinate data. Analysis of this type may be particularly helpful in providing a general overview of the main locations accounting for high volumes of incidents for a specific age group (in this example), but the data could be divided up in many different ways. Comparing this with an equivalent map based on the police dataset then helps to see if there are any obvious areas underrepresented in the police dataset. Moreover, restricting the analysis to a particular subgroup in terms of characteristics of individuals, types of assault or severity as it is done in the top panel of Fig. 5 may provide better information on the most serious, rather than most frequently repeating, locations (note that police dataset does not have a gender or age variable that would allow comparable subgroup analysis).

As an alternative approach, we also defined "hot" postcodes as those having more incidents recorded in the ambulance dataset than in the police dataset. Note that given the large difference in the total observation counts in both datasets, even postcodes with at least half as many incidents in the ambulance dataset as in the police dataset may be considered important as they, on average, show abnormally high concentration of incidents. We therefore suggest running similar analyses using different cut-offs.

The top 10 hot postcodes (defined by the largest difference between the number of incidents recorded in the ambulance and police datasets) show another interesting pattern, as over half of the postcodes refer to the location of police stations. Arguably, this is due to individuals going to police station to report an offence and police officers calling an ambulance while noting the actual location of the incident, but it is worth further investigation whether some other factors may play a role as well.

Table 6 Spatial concentration of violent crime by postcode

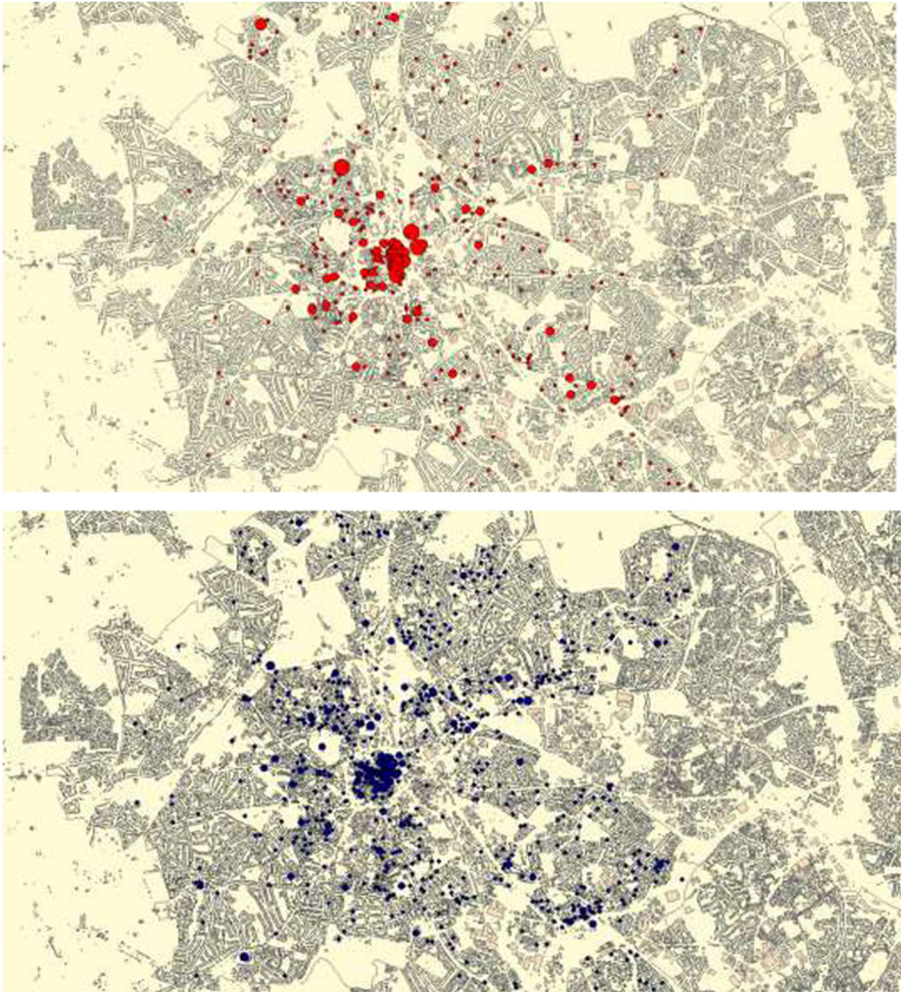|	Share of all incidents in the X% postcodes		
	Top 10%	Top 20%	Top 50%
Police	41.8%	58.7%	85.0%
Ambulance	39.5%	54.0%	80.2%

Fig. 5 Assaults within the 21–25 years old age group in Wolverhampton from the ambulance dataset, 2012–2015 (top panel); police dataset of all incidents, 2012–2015 (bottom panel). Scale for circles was determined automatically using the Jenks natural breaks classification; the number of incidents per postcode is not comparable

In addition, it is worth investigating further the typology of incidents in identified hotspots, for example, whether some hotspots show an abnormally high share of incidents involving adolescents or the use of guns, for example, or whether the time patterns of incident occurrence differ. Many of the incidents, particularly at specific postcodes, may also be related to bigger sporting or cultural events, demonstrations, etc. These unusual spikes in offences should be reflected also in the police dataset, which may help to identify the events.

For practical purposes, hotspot analysis at the postcode level may not be particularly useful as it may cover a large geographical area depending on various factors such as housing density. As an alternative, a similar analysis can be done at the street level, matching each data point with the closest street and subsequently identifying streets
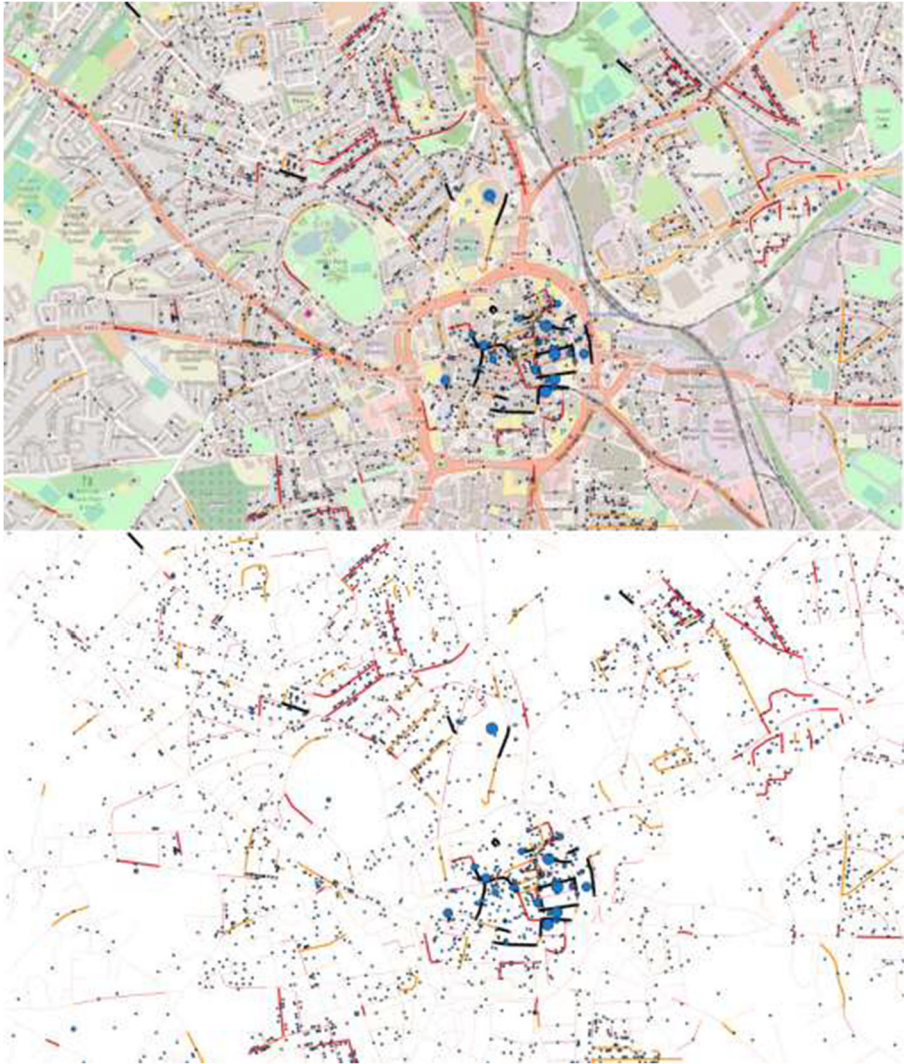
Fig. 6 Hot street segments in Wolverhampton with and without background in police dataset, 2012–2015. Note: Blue dots represent individual incidents recorded in the police dataset with size representing their frequency (the larger the dot the more frequent incidents that spot); hot streets are differentiated by colour (red and black representing the hot streets with most incidents, followed by progressively lighter orange meaning fewer and fewer incidents associated with a given street)

with multiple/most incidents occurring within a given timeframe.[10] This may then allow better specification of problematic areas, especially when combined with analysis

---

[10] MMQGIS plug-in for QGIS software was used for the analysis. Note that the analysis used streets as defined in the Ordnance Survey OpenData (https://www.ordnancesurvey.co.uk/opendatadownload), which were of different lengths and shapes. Longer street segments thus have by default higher probability of being associated with more incidents. This is similar to Weinborn et al. (2015) who identify street segments as parts of streets between two intersections. However, in order to make the street segments comparable to a greater extent it would be useful to limit the maximum length of a street.

of day and time patterns. For instance, one may then create a detailed patrol schedule so that the streets with most incidents are monitored at the most critical times. Alternatively, one may determine "hot corners" instead (i.e. intersections of two or more hot streets) or broader hotspots, although in a more systematic way than using hot postcodes.

The ambulance dataset does not include information on exact location of an incident, only a postcode that was subsequently assigned a geographical location during our data processing. As shown in Fig. 5, postcodes—particularly in densely populated areas—cover relatively small geographical areas and the results may be quite accurate.

In contrast, the following analysis is done using the police dataset, which contains precise coordinates of all incidents. The main purpose of this analysis is thus a proof of concept, i.e. presentation of the methodology and discussion of its application in practice. The results are depicted in Fig. 6.

We can see from the figures that while the overall picture remains unchanged, with the city centre showing by far the highest concentration of incidents, it is possible to better identify the most affected segments. This type of analysis may also be useful for easier patrol planning as some of the street segments may be connected. The application of this approach to ambulance data means it would be possible to identify much more specific locations that are "high frequency" in terms of calls for service, and aid with violence prevention efforts in those locations and the surrounding area. The other benefit of such targeting—assuming that coordinate level data could be used for incidents in the ambulance data—is that it is more equitable and cost-effective; the additional precision means a specific junction or location could be the focus rather than a wider "neighbourhood".

Conclusions

From this study, we may draw five conclusions. First, *ambulance records contain substantial new information on violence*, with between 66 and 90% of ambulance incidents not found in police data. Therefore, police are not aware of the location of a substantial proportion of violent incidents. Ambulance data can be used to inform patrol officers about where violence hotspots are developing and when the areas are "active", both of which they need to know for preventing further incidents.

Second, the volume of ambulance call-outs for public violence, averaging 16 per day in the West Midlands, means *ambulance data can offer high volume data that is not typically recorded by the police or ED*.

Third, as ambulance data is collected automatically and includes location data for each call, this data *does not require substantial additional work to be collated and shared*. This means that should ambulance data prove effective in reducing crime, it is easily scalable, particularly in light of the planned Ambulance Data Set (ADS)[11] that may allow for nation-wide comparisons to police data in England.

Fourth, *a substantial proportion (around 55%) of ambulance calls for service were from police officers*, although the research team were only able to link up to 34% of ambulance cases to incidents in the police dataset. Only 6% of calls by police were

---

[11] https://www.england.nhs.uk/urgent-emergency-care/improving-ambulance-services/ambulance-data-set/

located in the same postcode as a police station, indicating that the majority of these calls were not a result of people already in custody requiring medical attention. These figures require further research to establish the reason or reasons for the apparent discrepancy in the datasets.

Fifth, while ambulance data may have value for violence prevention or reduction activities, *it is still unproven and further study is required*. At a time when serious violence is being taken very seriously in the UK, the need for more research and development with these methods seems clearly justified.

References

Ariel, B., Weinborn, C., & Sherman, L. W. (2016). "Soft" policing at hot spots—do police community support officers work? A randomized controlled trial. *Journal of Experimental Criminology, 12*(3), 277–317.

Boyle, A., Ariel, B., & Weinborn, C. (2013). Can ambulance data be used to reduce community violence? Exploratory record linkage study. *Emergency Medicine Journal, 30*(10), 870–870.

Braga, A. A., & Weisburd, D. (2010). *Policing problem places: Crime hot spots and effective prevention*. Oxford University Press on Demand.

Braga, A., Papachristos, A., & Hureau, D. (2012). Hot spots policing effects on crime. *Campbell Systematic Reviews, 8*(1), 1–96.

Dudfield, G., Angel, C., Sherman, L. W., & Torrence, S. (2017). The "power curve" of victim harm: Targeting the distribution of crime harm index values across all victims and repeat victims over 1 year. *Cambridge Journal of Evidence-Based Policing, 1*(1), 38–58.

Liggins, A., Ratcliffe, J. H., & Bland, M. (2019). Targeting the most harmful offenders for an English police agency: Continuity and change of membership in the "Felonious Few". *Cambridge Journal of Evidence-Based Policing, 3*(3), 80–96.

Reform. 2014. 'Violence reduction unit, Scotland'. As of 20 July 2017: http://www.reform.uk/wp-content/uploads/2014/11/Preventative_criminal_justice_in_Glasgow_Scotland.pdf.

Shepherd, J. (2004). *Effective NHS contributions to violence prevention – the Cardiff Model*. Cardiff: Cardiff University.

Sherman, L.W. (2007). The power few: experimental criminology and the reduction of harm. *Journal of Experimental Criminology, 3*(4), 299–321.

Sherman, L. W. (2013). The rise of evidence-based policing: Targeting, testing, and tracking. *Crime and Justice, 42*(1), 377–451.

Sherman, L. W., & Weisburd, D. (1995). General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. *Justice Quarterly, 12*(4), 625–648.

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology, 27*(1), 27–56.

Sherman, L. W., Neyroud, P. W., & Neyroud, E. (2016). The Cambridge crime harm index: Measuring total harm from crime based on sentencing guidelines. *Policing: A Journal of Policy and Practice, 10*(3), 171–183.

Taylor, A., Boyle, A., Sutherland, A., & Giacomantonio, C. (2016). Using ambulance data to reduce community violence: critical literature review. *European Journal of Emergency Medicine, 23*(4), 248–252 http://www.ncbi.nlm.nih.gov/pubmed/26629765.

Warburton, A. L., & Shepherd, J. P. (2004). Development, utilisation, and importance of accident and emergency department derived assault data in violence management. *Emergency Medicine Journal, 21*(4), 473–477.

Weinborn, C., Ariel, B., Sherman, L. W., & O'Dwyer, E. (2017). Hotspots vs. harmspots: Shifting the focus from counts to harm in the criminology of place. *Applied Geography, 86*, 226–244.

Weisburd, D., Groff, E. R., & Yang, S. M. (2012). The criminology of place: Street segments and our understanding of the crime problem. Oxford University Press.

Weisburd, D., Telep, C. W., Braga, A. A., & Groff, E. R. (2010). *The importance of place in policing: Empirical evidence and policy recommendations* (pp. 1–68). Stockholm: Brottsförebyggande rådet (BRÅ).

Welsh, B. C., Braga, A. A., & Sullivan, C. J. (2014). Serious youth violence and innovative prevention: on the emerging link between public health and criminology. *Justice Quarterly, 31*(3), 500–523.

Affiliations

Alex Sutherland[1] · Lucy Strang[2] · Martin Stepanek[3] · Chris Giacomantonio[4] · Adrian Boyle[5] · Heather Strang[6]

[1]   The Behavioural Insights Team, London, UK

[2]   Policy Institute at King's, King's College, London, UK

[3]   The Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republic

[4]   Pier Labs, Halifax, Canada

[5]   Emergency Department, Cambridge University Hospitals Foundation Trust, Cambridge, UK

[6]   Institute of Criminology, Cambridge University, Cambridge, UK