Metropolitan Police Service. (2011) Four Days in August. London: Metropolitan Police Service.
Reiner, R. (1992) *The politics of the police* (2nd ed.). London: Harvester Wheatsheaf.
Jennings, W.G., Lynch, M.D., & Fridell, L.A. (2015) Evaluating the impact of police officer body-worn cameras (BWCs) on response-to-resistance and serious external complaints: Evidence from the Orlando police department (OPD) experience utilizing a randomized controlled experiment. *Journal of Criminal Justice*, 43(6), 480–486.
Stanko, B., & Dawson, P. (2013) Reflections on the Offending Histories of those Arrested during the Disorder. *Policing*, 7(1), 3–11.
Stanko, E., & Dawson, P. (2015) Police Use of Research Evidence: Recommendations for Improvement, Springer.

Biographies

Paul Dawson is the Research Manager in Evidence and Insight, Mayor's Office for Policing And Crime in London. He is a Chartered Psychologist and Associate Fellow of the British Psychological Society. He previously worked within the National Health Service and the Home Office prior to joining the Metropolitan Police Service in 2008. He holds a Ph.D. in Psychology from the University of Birmingham investigating the empirical contribution of weapons to the study of serious sexual offending.

Elizabeth A. Stanko is Head, Evidence and Insight, Mayor's Office for Policing And Crime in London. For over a decade, she worked inside the London Metropolitan Police Service, establishing a social research function alongside performance analysis. In her first life, she was a professor of criminology. She was awarded an OBE for her services to policing in the 2014 Queen's Birthday Honours.

Legal Information Management, 16 (2016), pp. 71–77
© The Author(s) 2016. Published by British and Irish Association of Law Librarians
doi:10.1017/S1472669616000207

Analysing Police-Recorded Data[1]

Abstract: The quarterly bulletins on crime statistics in England and Wales are compiled from two sets of data: crime survey and police-recorded crime. Whilst the former is considered to give the most reliable trends, the latter has a greater level detail for a fuller spectrum of crimes types. This paper by Allan Brimicombe, explores the advantages and problems of analysing police-recorded data for the insights they contain. This is illustrated by examples from an analysis of domestic violence.
Keywords: domestic violence; police-recorded crime; crime statistics; data analysis

INTRODUCTION

Statistics on Crime in England & Wales are published quarterly by the Office of National Statistics (ONS)[2]. Each bulletin contains statistics that are about 15 months in arrears to allow for compilation, quality checks, analysis and publication. They are compiled from two main sources of data: the Crime Survey for England and Wales (CSEW)[3] and police-recorded crime (PRC). CSEW is a face-to-face victimisation survey carried out on an annual rolling basis with a sample of about 36,000 households. Whilst CSEW is viewed as giving the most accurate comparative trend over time for a range of crime types, the figures are given nationally (for England and Wales) and are only for crimes experienced by individuals and households. PRC derives from the transactions databases of individual police forces, can represent more detailed geographies and time periods, and record both crimes against the individual and crimes against the state[4]. There are therefore certain advantages of working with PRC, but at the same time there are important drawbacks and pitfalls. This paper explores these issues of working with police-recorded data.

https://doi.org/10.1017/S1472669616000207 Published online by Cambridge University Press

TERMINOLOGY

Broadly speaking two types of events are recorded by the police: offences and incidents. Offences are in two categories: those that are notifiable to the Home Office and include all offences triable by jury or either-way, and those less serious offences that are dealt with exclusively by the Magistrates' Courts. Offences are categorised as a crime type according to National Crime Recording Standards (NCRS) and Home Office Counting Rules (HOCR). The Notifiable Offences List (NOL)[5] which forms part of the HOCR is revised from time to time to reflect changing legislation and forms the basis for the compilation of the ONS statistics on police-recorded crime. Some crimes, such as domestic abuse, are not statutory offences and are recorded according to the appropriate notifiable offences category (e.g. assault with injury) and are marked in the police database using a flag. Incidents are recorded according to the National Standards for Incident Recording (NSIR) and associated counting rules. Incidents are all manner of events reported to the police where the public has cause for concern (e.g. traffic accidents, missing persons, anti-social behaviour). Only three categories of incident are notifiable to the Home Office and these deal with reports of rape which, for specific reasons, are not crimed. There are two practical points to note: there are many more incidents recorded than offences and that a full picture of victimisation needs to include data on both offences and non-crime incidents. Because there are these two categories of events – offences and incidents – and because PRC refers generally to offences only, the more general term of police-recorded data is used here.

ADVANTAGES OF POLICE-RECORDED DATA

Police databases record the facts and evidence pertinent to an offence or incident. This includes details of the location where the event occurred, victim(s), informants and witnesses, perpetrator(s) (either as suspects or accused) and the modus operandi (MO). Because there can, for example, be more than one victim and/or perpetrator to an offence, the database has many-to-many relations. This adds complexity in organising the data for analysis, but also adds richness and extraordinary granularity given that each event is a separate entry. A researcher can expect to be working with tens to hundreds of thousands of records. In general, the objective of research using police-recorded data is not to focus on specific individuals *per se* but to identify aggregate patterns that provide new understanding and suggest interventions or solutions to problems. Two different documented examples of this type of approach are: a) an analysis of the upsurge in SatNav theft from the beginning of 2006 given in Brimicombe (2012), and b) an analysis of offender histories of those arrested in the London riots of 2011 given in Stanko and Dawson (2012).

Police-recorded data allows research into all crime and incident types rather than being restricted to individual and household victims as is the case with CSEW or restricted to the NOL. For a start, a face-to-face survey can't include experience of homicide and there are practical and ethical issues of including questions such as on child abuse. Then there are the so-called victimless crimes that are classed as crimes against the state (e.g. possession of drugs, perjury). In the NOL there are currently 1069 categories of crime against the state versus 275 categories of victim-based crime[6]. But as already observed there are many more incidents than offences, and the categories of offences dealt with by Magistrates' Courts are not included in the NOL. The CSEW, important as it is, only covers a minority of offence categories whereas police-recorded data covers the totality of what is reported to or detected by the police.

The granularity of police-recorded data is the highest available. Events are recorded with the time and date at which they are reported to the police, and where the actual occurrence of an event is uncertain (e.g. a burglary while away on holiday) a beginning and ending time and date are also recorded. Events reported retrospectively (such as child abuse reported later in life) are similarly handled. This uncertainty in the actual time and date of occurrence for some crime types can be offset to some extent by aggregation, such as by the day or week, using the mid-point of the date range or it can be handled probabilistically (Ratcliffe, 2002). Police-recorded data also has a high level of geographical granularity in that many events are recorded against an address, either where the event happened (e.g. a burglary), to the nearest place where it happened (e.g. a violent altercation outside a pub) or, say, at a road junction. This forms the basis of 'hotspot' mapping allowing police to focus their resources on those locations that have highest intensities of crime (Harries, 1999; Brimicombe, 2012). The data on victims and perpetrators can be further filtered by gender, age, ethnicity and offence/incident category in order to study specific patterns of activity. The volume of data records and the number of variables make these data particularly amenable to predictive analytics and techniques of machine learning (see for example Berk, 2013).

Notwithstanding some of the difficulties of working with police-recorded data discussed in the next section, it is feasible to identify and analyse patterns of repeat victimisation and repeat offending (prolific offenders). Whilst it is technically quite difficult to reliably identify and profile repeats (see Brimicombe, forthcoming), these are at the heart of informed crime reduction (Pease & Tseloni, 2014), yet for example, one third of police forces were found to have no data on repeat victims of domestic abuse (HMIC, 2014a). Whilst CSEW includes data on repeat victimisation (though obviously not repeat offenders), the number of victimisations is capped for statistical reasons at five events in the past twelve months (ONS, 2015). Police-recorded data holds out the possibility of longer victimisation chronologies (not just the last

https://doi.org/10.1017/S1472669616000207 Published online by Cambridge University Press

twelve months) and the full number of repeat occurrences in chronic cases.

Finally, of growing importance is data linkage. This is where individuals are matched across data systems in order to look at patterns of multiple service use. An example would be victims who report violent crimes to the police and report to accident and emergency (A&E). There may be occasions when violence is reported without attendance at A&E and other occasions when A&E is attended as a result of an assault but not reported to the police. Linking the two up can reveal informative patterns of activity. Some government initiated programmes such as Troubled Families Phase 2[7], require local authorities to link social services, housing, education, benefits, mental health and police data in order to identify families classed as 'troubled' and thus qualifying for interventions on a payment by results model.

ISSUES IN USING POLICE-RECORDED DATA

Having discussed the advantages of using police-recorded data for a range of analyses that can inform policy, strategy and operations, the use of such data is not without its problems.

Historically, the 43 police forces in England and Wales have independently developed their IT systems and designed their own database schemas for recording events reported to them. There are at least 88 data centres (PASC, 2011) and some 2,000 IT systems (CPA, 2012). There can be separate databases for call and dispatch (999 calls), incidents and offences, details of victims and details of perpetrators/accused and so on. There may not be unique keys that connect these databases because of the many-to-many relationships that occur in crime events and keys meant to achieve greater integration may not be assiduously copied across due to time and effort. As already mentioned, some offences are not statutory crimes (such as domestic abuse) and may be identified by a flag (or several different flags) in police databases, marked in a separate register, or be mirrored in a separate database specific for that type of offence. This all adds complexity when trying to retrieve all the relevant data for analysis.

The personal details present in police-recorded data means that they are subject to the provisions of the 1998 Data Protection Act and disclosure must be prevented. Breaches reported to the Information Commissioner's Office (ICO), such as loss or theft of sensitive personal data will attract a financial penalty of up to £500,000[8]. Not surprisingly high security IT systems operating within a well-founded information security framework are required for the storage, analysis and eventual archiving or destruction of such data in order to prevent breaches. Individuals with access to the data also usually need to be vetted by the relevant police force(s) and may be asked to sign the Official Secrets Act. These conditions are not so severe for anonymised or aggregated police data, but then granularity and analytical depth are lost.

The public's trust in crime statistics is fundamental in a transparent democracy but had declined in the late 1990s and 2000s due in large part to political 'spin'. Despite subsequent work to repair the trust (UKSA, 2010), the quality and reliability of PRC has come in for heavy criticism at Parliamentary Committee (PASC, 2014) and the UK Statistics Authority subsequently withdrew its National Statistics designation (UKSA, 2014). At the same time a divergence was noted between the amount of crime as measured through CSEW and that given for comparable police-recorded crime (ONS, 2013). Whilst the Audit Commission had carried out regular checks of police data quality from 2003/04 following the introduction of the NCRS, they were discontinued after 2006/07. By 2011, quality concerns lead Her

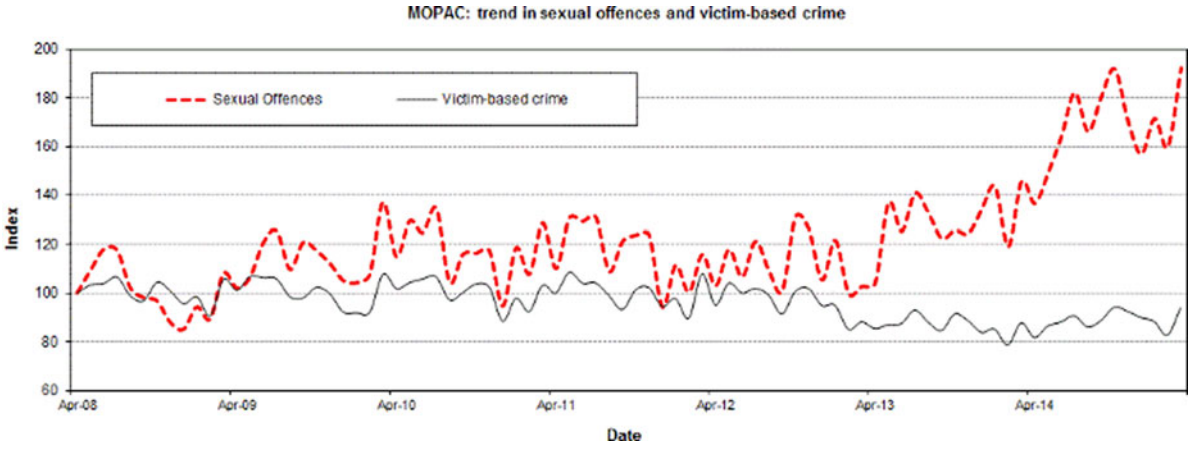


Figure 1: Marked divergence in indexed monthly counts of all victim-based crime and sexual offences from April 2013 coterminous with HMIC inspections which led to criticisms of the accuracy in recording sexual offences (data available from: http://data.london.gov.uk/dataset/metropolitan-police-service-recorded-crime-figures-and-associated-data).

https://doi.org/10.1017/S1472669616000207 Published online by Cambridge University Press

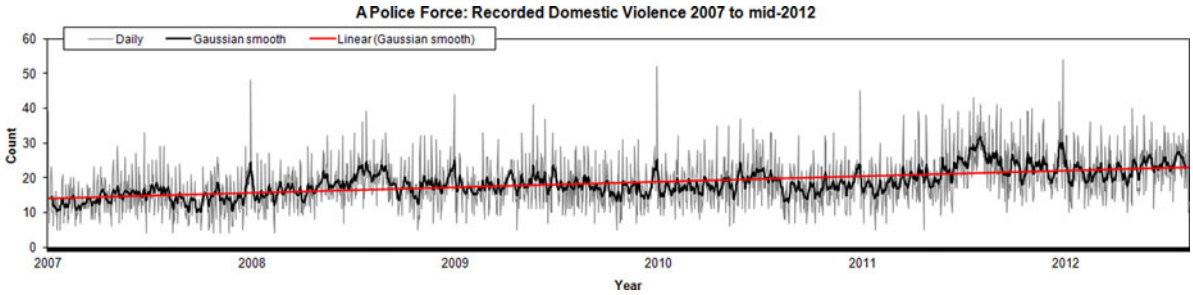


Figure 2: Daily counts of domestic violence with 28-day Gaussian smoothing and trend line 2007 to mid-2012.

Majesty's Inspectorate of Constabulary (HMIC) to carry out a series of reviews of police crime and incident reports in England and Wales. As might be expected, the inspections have resulted in changes in police-recorded crime which can introduce marked discontinuities in the data series. By way of illustration, HMIC criticised the lack of accuracy in recording sexual offences (HMIC, 2014b); Figure 1 shows for London how monthly counts (indexed to 100 at the start of the data series) for all victim-based crimes and sexual offences which roughly track each other for five years and then markedly diverge with a steep increase in recorded sexual offences once HMIC begins their review. The 80% increase in recorded sexual offences between April 2013 and mid-2014 is more likely to be the result of better recording practices than an actual increase in the amount of sexual offences in London. The National Statistics designation of quality is unlikely to be restored until changes in PRC consistently reflects real changes in the amount of crime.

Whilst much of the quality debate has focused on adherence to the NCRS and HOCR in determining if an event is an offence or incident and the correct classification of offences by crime type, quality problems also concern the accuracy, consistency and completeness with which fields in the databases are populated with data. Whilst it needs to be recognised that there is no such thing as the perfectly correct database, the recording of names, addresses and other particulars of events, often in difficult, tense situations, are subject to inadvertent errors, gaps and lack of consistency. Victims do not always give accurate responses. Extensive data cleaning is required to maximise the analytical use of the data (Brimicombe *et al.*, 2007). Furthermore, it is well understood from CSEW that 20–60% of crime, depending on crime type, is not reported to the police and that police-recorded data are always an undercount of the true level of activity. Whilst CSEW is accepted as giving a consistently reliable trend of personal and household victimisation at a national scale, and notwithstanding the issues of data quality discussed above, police-recorded data are the only means of understanding sub-national activity down to police command unit and Community Safety Partnership (CSP) level – and even down to neighbourhood geographies - as a means to compare the performance of command units and CSP and what works in preventing crime.

EXAMPLES OF ANALYSING POLICE-RECORDED DATA

Presented in this section are a few examples of the types of analyses that can be carried out from police-recorded data. This represents only a fraction of what can be extracted from this type of transactions database. The data for these

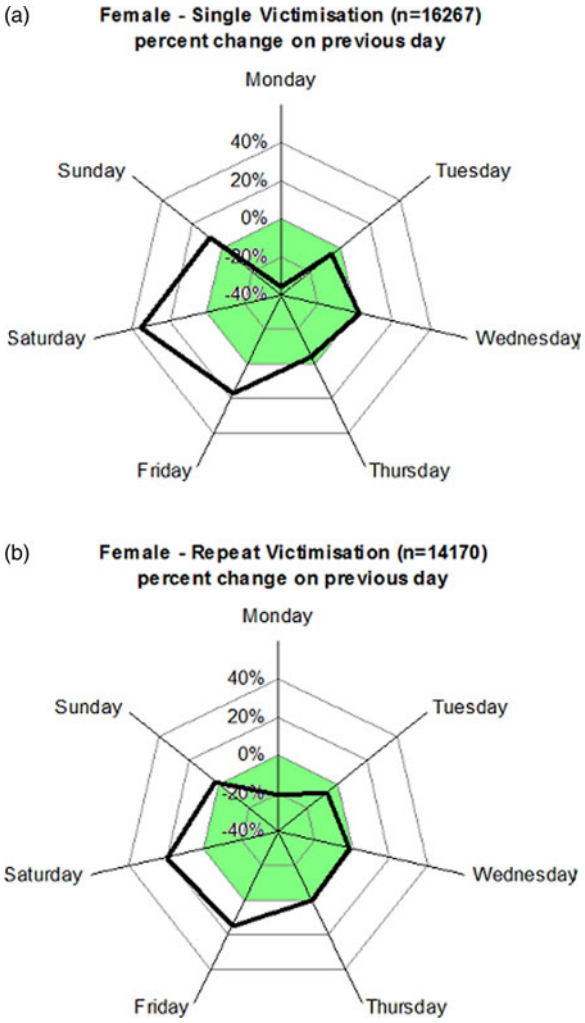


Figure 3: Comparison of aggregated weekly cycle of domestic violence (percentage change on previous day) for females having reported only one event to the police and for female repeat victims.

https://doi.org/10.1017/S1472669616000207 Published online by Cambridge University Press

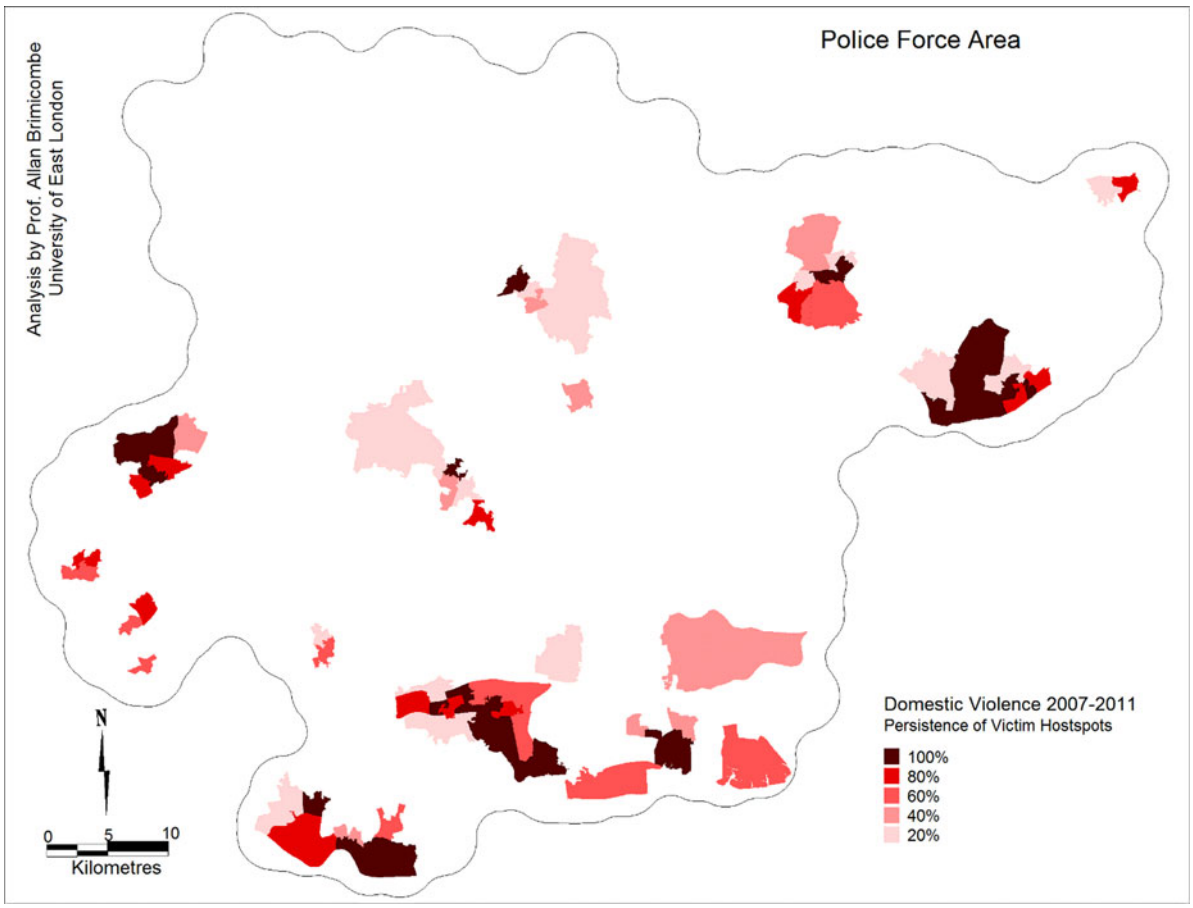


Figure 4: Persistence of statistically significant hotspots of domestic violence by victim home address.

examples are for domestic violence offences over a five and a half year period from 2007 to mid-2012 (and thus predates the methodological difficulty introduced by the HMIC inspections as illustrated in Figure 1) for an entire police force (county level) with a mixture of urban and rural areas.

Underlying most approaches to analyses that inform problem-solving or evidence-based approached to crime reduction is the understanding that crimes tend to form patterns (Brantingham & Brantingham, 1984). Such patterns are the product of processes (actions) which, once the drivers or risk factors are understood, can be modified or stopped through an appropriate set of interventions (legal, social, or situational) that disrupts the patterns of crime. The patterning of crime occurs in one or more key dimensions: spatially (location), temporally, the victims, the offenders and their modus operandi. The patterns of most interest are those that cause clustering (as in a crime hotspot) or a degree of regularity and thus have a degree of predictability. The results of analyses that cause surprise are more likely to change thinking about crime and its prevention.

Figure 2 plots the daily occurrence for domestic violence for the case study area. This appears to have chaotic oscillations from day to day. Two sets of smoothing have been applied to aid interpretation: one is a 28-day Gaussian smoothing and the other is a best fit linear trend line. Apart from the fact that police-recorded domestic violence is rising over the period and that on no single day is there no record of domestic violence, the most eye-catching elements of this graph are the spikes that occur once a year, every year – this is new year's day, signalling for a significant number of households the end of the season of peace and goodwill. Some of the other spikes may have a discernable cause, but some are difficult to determine. For example Brimicombe and Cafe (2012) were able to conclusively attribute spikes in the national rate of domestic violence when the England team either lost or won (but not drew) their matches in the 2010 FIFA World Cup.

The apparent chaotic oscillations in Figure 2 become clearer when aggregated into weekly patterns whereby it becomes evident that that counts of domestic violence are much higher at the weekend. Figure 3 illustrates this effect by plotting percentage change on the previous day in a spider diagram showing the whole week in a cyclical arrangement. By way of contrast Figure 3(a) is for events suffered by female victims who have only reported to the police once, while Figure 3(b) is for female repeat victims. Whereas in Figure 3(a) reporting of domestic violence starts to increase on Fridays, increases by a further 35% on Saturdays, increases a further 10% on Sundays to then decline dramatically on a Monday, the weekend effect in Figure 3(b) is less

https://doi.org/10.1017/S1472669616000207 Published online by Cambridge University Press

pronounced and is indeed statistically significantly different. This would suggest that the process of, or events leading up to victimisation is subtly different between the two and merits more detailed analysis of the MO.

A geographical view of the data is given in Figure 4. This is by victim home address (not necessarily the same as the location at which the crime was perpetrated as it may have been carried out at the perpetrator's address if the perpetrator and victim are not currently co-habiting; domestic violence also takes place in public locations such as shopping centres and pubs), aggregated to lower super output areas (LSOA)[9]. For each full year of data (2007–11), the count of domestic violence victims by LSOA has been tested for statistical significance[10] and if the count of victims is above the 95% confidence interval (i.e. this level of activity is unlikely to have arisen by chance), then the LSOA has been deemed to be a 'hotspot'. How often in each of the 5 years an LSOA is a statistical hotspot then becomes a measure of how persistent these hotspots are, with the distribution of these persistences given in Figure 4. The map shows the majority of LSOA are not statistical hotspots even though domestic violence is reported to the police everywhere. This is contrasted with those areas that are hotspots and these are geographically clustered with a number of areas that are hotspots every or nearly every year. These tend to be urban areas with multiple deprivation, have a high level of predictability regarding the number of domestic violence victims, and should thus be the focus of police and partnership intervention.

CONCLUSIONS

Police-recorded data are of value for their granularity (level of detail) across all crime types and the ability to reveal, through analysis, patterns of activity likely to be of interest for policy, strategic planning and operations. The few examples presented in this paper demonstrate this. Key problems of working with such data are quality issues: their consistency and completeness, and the fact that currently, changes in the amount of some crimes may not be real changes but changes in recording practices. A barrier to their use by analysts outside policing is the level of confidentiality which police data attract, but police analysts are often too busy on routine activities to undertake speculative research to find new insights. The more in-depth analyses undertaken using police-recorded data, including using new and novel techniques, the greater will be the recognition of their value beyond routine operations and worth the effort by front-line officers to invest in data quality, which would in turn further enhance the value of the data. Police-recorded data can then become a reliable source of evidence in experimental and quasi-experimental evaluation of what works in policing and crime prevention.

Footnotes

[1] this paper is based on a presentation given at the IALS/British Library/SLSA/BSC National Training Day 2015: Sources and Methods in Criminology and Criminal Justice.

[2] e.g. http://www.ons.gov.uk/ons/rel/crime-stats/crime-statistics/crime-in-england-and-wales---year-ending-september-2015/stb-crime-sept-2015.html.

[3] formerly the British Crime Survey.

[4] a general guide to the use of crime statistics in England and Wales is available at: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/crime-statistics-methodology/user-guide-to-crime-statistics.pdf.

[5] available at: https://data.gov.uk/dataset/recorded-crime-counting-rules/.

[6] there are a further 162 categories crime that are both victim-based and against the state (e.g. causing danger to road users).

[7] https://www.gov.uk/government/publications/financial-framework-for-the-expanded-troubled-families-programme.

[8] The Data Protection (Monetary Penalties) (Maximum Penalty and Notices) Regulations 2010.

[9] LSOA are a neighbourhood census reporting area equivalent to an average population of 1,600 residents – for further information see http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html.

[10] spatial randomness is usually modelled using the Poisson distribution, but for data where only counts of those affected are known (and not those unaffected and at risk) a truncated Poisson distribution is used to test for significance.

References

Berk, R. (2013) Algorithmic criminology. *Security Informatics* 2: 5. Available at: http://www.security-informatics.com/content/2/1/5

Brantingham, P. and Brantingham, P. (1984) *Patterns in Crime*. New York: Macmillan.

Brimicombe, A.J. (2012) Did GIS start a crime wave? SatNav theft and its implications for geo-information engineering. *The Professional Geographer* 64: 430–445.

Brimicombe, A.J. and Cafe, R. (2012) Beware win or lose: domestic violence and the World Cup. *Significance* 9(5): 32–35. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2012.00606.x/pdf.

https://doi.org/10.1017/S1472669616000207 Published online by Cambridge University Press

Brimicombe, A.J. (forthcoming) Mining police recorded offences and incidents data to inform a definition of repeat victimisation of domestic abuse for statistical purposes. *Policing*.

Brimicombe, A.J., Brimicombe, L.C. and Li, Y. (2007) Improving geocoding rates in preparation for crime data analysis. *International Journal of Police Science & Management* 9: 80–92.

CPA (2012) *Mobile Technology in Policing*. London: House of Commons Committee of Public Accounts. Available at: http://www.parliament.uk/documents/TSO-PDF/committee-reports/129.pdf.

Harries, K. (1999) *Mapping Crime, Principles and Practice*. Washington DC: National Institute of Justice. Available at: https://www.ncjrs.gov/pdffiles1/nij/178919.pdf.

HMIC (2014a) *Everyone's business: Improving the police response to domestic abuse*. London: Her Majesty's Inspectorate of Constabulary. Available at: https://www.justiceinspectorates.gov.uk/hmic/wp-content/uploads/2014/04/improving-the-police-response-to-domestic-abuse.pdf.

HMIC (2014b) *Crime recording: Making the victim count*. London: Her Majesty's Inspectorate of Constabulary. Available at: https://www.justiceinspectorates.gov.uk/hmic/wp-content/uploads/crime-recording-making-the-victim-count.pdf.

ONS (2015) *High frequency repeat victimisation in the Crime Survey for England and Wales*. Newport: Office of National Statistics. Available at: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/crime-statistics-methodology/methodological-notes/high-frequency-repeat-victimisation.pdf.

PASC (2011) *Government and IT- "a recipe for rip offs": time for a new approach*. London: House of Commons Public Administration Select Committee. Available at: http://www.publications.parliament.uk/pa/cm201012/cmselect/cmpubadm/715/715i.pdf.

PASC (2014) *Caught red-handed: why we can't count on police recorded crime statistics*. London: House of Commons Public Administration Select Committee. Available at: http://www.publications.parliament.uk/pa/cm201314/cmselect/cmpubadm/760/760.pdf.

Pease, K. and Tseloni, A. (2014) *Using Modelling to Predict and Prevent Victimization*. New York: Springer.

Ratcliffe, J.H. (2002) Aoristic signatures and the temporal analysis of high volume crime patterns. *Journal of Quantitative Criminology* 18: 23–43.

Stanko, B. and Dawson, P. (2012) Reflections on the offending histories of those arrested during the disorder. *Policing* 7: 3–11.

Biography

Allan Brimicombe is Professor of Geo-Information at the University of East London. For the past two decades he has focused on knowledge discovery in large transactions databases including police-recorded crime, accident and emergency admissions and a wide range of business data. He is a Director of the British Society of Criminology and a Fellow of the Academy of Social Sciences.

https://doi.org/10.1017/S1472669616000207 Published online by Cambridge University Press