London murders: a predictable pattern?

Four stabbings to death in a single day. Ninety murders in 7 months. Shocking figures—or are they? Knife crime makes the headlines almost daily but are Londoners at increased risk of being murdered? **David Spiegelhalter** and **Arthur Barnett** investigate—and find a predictable pattern of murder.

Violence in London attracts headlines. After four people were murdered in separate incidents in London on July 10th, 2008, BBC correspondent Andy Tighe said "To have four fatal stabbings in one day could be a statistical freak"[1]. But could it? And on July 28th *thelondonpaper* had the front page headline: "London's murder count reaches 90".

Of course every single murder is a tragedy for those whom it affects, and it is little comfort to them what patterns we may detect in the figures. But for the benefit of the police service and broader society we may wish to consider whether these "shocking" numbers in the media are evidence that things are really getting worse.

Each year the London Metropolitan Police record around 170 homicides, and this has been stable for the last 5 years[2]. Box 1 explains what is covered by "homicide"—here we shall use "murder" as an equivalent term. We shall focus on all murders, whatever the method of killing: Box 2 provides details about the cause of death and briefly considers stabbings alone. Each of these murders is an individual crime that cannot be predicted. It may appear strange, but this very randomness means that the overall pattern of murders is, in some ways, quite predictable. Using some basic probability theory, and assuming that the level of violence remains the same, we can answer the questions coming out of the two stories given above: first, four murders in a day is unusual but not extraordinary. We would expect it to occur around once every 3 years. And, second, we would have expected, on statistical grounds, 93 murders by July 28th—so the count of 90 is not at all surprising.

We can also answer questions such as how often would we expect days with no murders? (The answer is: on around 64% of days.) How often could we expect three or more murders in one day? (On around 4 days a year.) And how often should we expect there to be long gaps between murders? (A gap of 7 days should occur around six times a year.)

Why should anyone care about this? By making full use of the kind of data collected by the Home Office Specialist Crime & Police Resources Team, both the public and police forces can be alert to real changes in the incidence of crime, but not be either unnecessarily alarmed or falsely reassured by apparent variability. This analysis is part of our contribution to the work of the Risk and Regulation Advisory Council (`http://rrac.intelligus.net`) in trying to stimulate public interest in a debate about "risk" in policing following on from Sir Ronnie Flanagan's review of policing[3].

Box 1. What counts as a "homicide"?

The terminology used in this article reflects reporting by the popular press rather than Home Office statistical definitions and nomenclature[2]. Home Office statistics report "homicides" not "murders", and homicide by "sharp instrument" not "stabbings".

The term "homicide" covers the offences of murder, manslaughter and infanticide. Murder and manslaughter are common law offences that have never been defined by statute, although they have been modified by statute. The offence of infanticide was created by the Infanticide Act 1922 and refined by the Infanticide Act 1938 (s1). Homicide offences are shown according to the year in which the police initially recorded the offence as homicide. This is not necessarily the year in which the incident took place nor the year in which any court decision was made.

Box 2. Four stabbings in a day?

The main body of this article addresses the question "How surprising is four murders in one day in London?" A full answer to Andrew Tighe's question of whether four stabbings can be considered a "statistical freak" is not so straightforward and requires assumptions that have not been tested using the micro-data. Nevertheless some basic perspectives on the question are provided here using aggregate data for London for 2006–2007 kindly provided by the Home Office.

Table 1 shows the method of killing in 167 homicides recorded in London in 2006–2007: these counts are consistent with Table 1.03 from the report of Coleman *et al.*[2]. Stabbing—or, more correctly, killing by sharp instrument—is the most frequent method, accounting for 41% of the incidents, with shooting the next most frequent, comprising 17% of homicides.

If we assume that the causes of death are independent for multiple murders on the same day, and that the observed rates can be considered as estimates of the risks for each murder, then we can calculate the probability of any particular combination occurring, *given* that there are four murders in a day. These calculations can also be checked by simulation.

We can therefore estimate that four stabbings has a probability of $0.41^4 = 0.028$, which turns out to be only the eighth most likely combination. However all specific combinations are unlikely: the estimate of the probability of the most likely four-murder combination, which includes two stabbings, is only 0.060. The probability of at least three stabbings out of four murders is estimated to be 0.20, for two or more stabbings 0.54, and for one or more stabbings it is 0.88. So although four murders in one day is not a particularly surprising event when taken over a period of time, the fact that they were all stabbings is more notable.

Table 1. Method of killing for homicides recorded by the Metropolitan Police in 2006–2007

Method	*Rank*	*Number 2006–2007*	*%*
Sharp instrument	1	68	41
Shooting	2	29	17
Not known	3	29	17
Hitting, kicking, etc.	4	15	9
Blunt instrument	5	8	5
Other	6	6	4
Strangulation	7	5	3
Poison or drugs	8	5	3
Burning	9	1	0.5
Drowning	10	1	0.5
Motor vehicle	11	0	0
Explosion	12	0	0
Total		167	100

Four murders also took place on April 8th 2006, when *no* stabbings were recorded: this is not a surprising pattern of four murders and has estimated probability $(1-0.41)^4 = 0.12$. However, the *particular* combination of four methods used on that day did not occur once in the simulation of 10 000 combinations—equivalent to approximately 30 000 years of murders. This suggests either a rare event or a need to question our simplistic model.

We have looked at the pattern of murders in London between April 2004 and March 2007. Over these 1095 days a total of 508 homicides were recorded by the Metropolitan Police. Twelve occurred on an unknown date, and 13 were the result of the bombings on July 7th, 2005 (other victims were recorded by the British Transport Police), which is an exceptional "cluster" that should be dealt with separately: removing these leaves 483 homicides on known dates. Figure 1 shows the number of homicides each day over this 3-year period: there was one day on which four murders were ecorded (April 8th, 2006), but also one gap of 12 days (February15th–26th, 2007) in which no murders were recorded at all.

How predictable is the overall pattern of murders? Overall there are around 13 murders a month, which is three a week or, on average, one every 54 hours. We can start by pretending that murders just happen as unconnected events with around a 1 in 50 chance of one happening each hour. Of course murders do not occur in this way and, in particular, we know that the risk of any specific individual being murdered can depend strongly on age, gender, geography, ethnicity and so on. But here we are dealing with the total over the whole London population of over 7 million, and we can use probability theory to tell us what we would expect to see if murders did occur as "random" events, and check whether this fits with the observed pattern. The idea is that we can make many predictions about the pattern of murders, based only on knowing the overall number over 3 years.

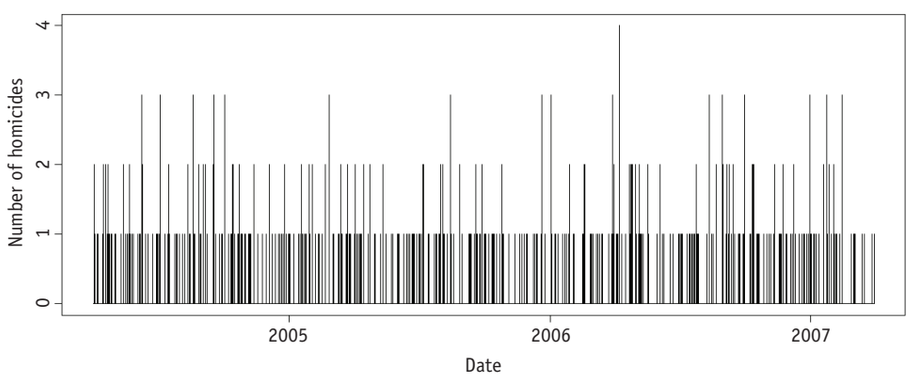

Figure 1. Number of homicides recorded each day in London between April 2004 and March 2007

Can we predict how many days there are no murders, one murder and so on?

First we can look at how many murders happen each day. If murders happened as random events, the number of murders each day would follow a Poisson distribution with a mean of 0.44. Over 3 years we would expect the pattern shown in Table 2 and Figure 2.

The actual numbers of days with 0, 1, 2 and so on homicides are shown in Table 1 and drawn in Figure 2: it is clear that there is a very close match between the observed and expected distribution. In particular we would

Table 1. Observed and expected number of days with 0, 1, 2 etc. homicides occurring in London on 1095 days between April 2004 and March 2007

	Number of days with the following numbers of homicides per day					
	0	*1*	*2*	*3*	*4*	*5 or more*
Expected	705.2	310.3	68.3	10.0	1.1	0.1
Observed	713	299	66	16	1	0

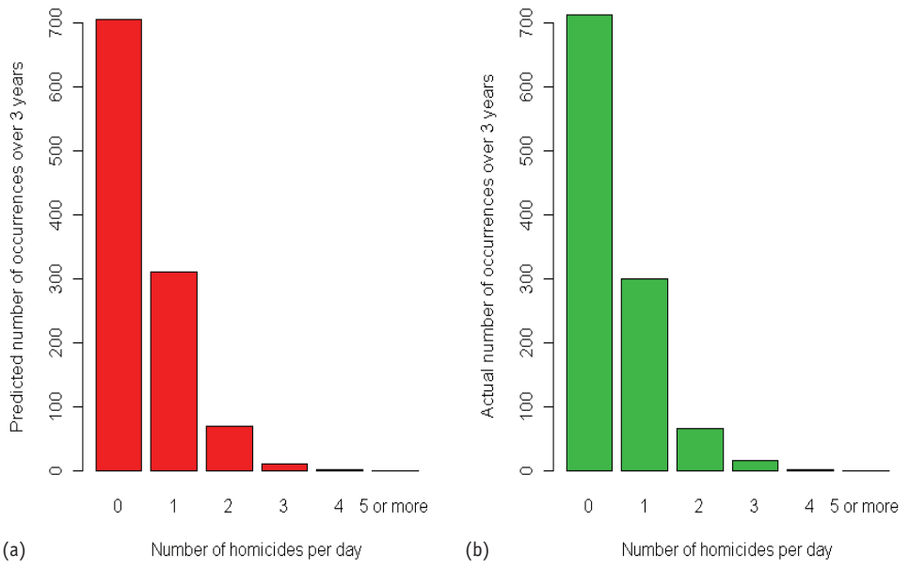

Figure 2. Frequency of occurrence of days with different numbers of homicides recorded in London between April 2004 and March 2007: (a) predicted; (b) observed

have expected one day with four murders—and we observed one; we would have expected 64% (705.2/1095) of days to have had no murders—and the observed proportion was 65% (713/1095). Formally, a χ^2 test shows no evidence against the Poisson distribution (simulation-based p-value of 0.44), and reminds us of the famous 19th-century example[4], which showed a similar excellent fit of the Poisson distribution to the number of Prussian officers kicked to death by their horses each year.

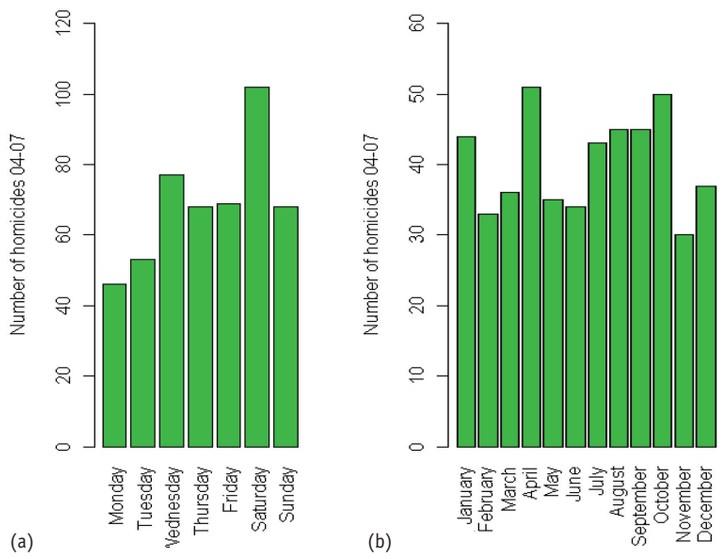

Figure 3. Frequency of occurrence of homicides on (a) different days and (b) different months in London between April 2004 and March 2007

Figure 3 shows the distribution of homicides by day and month: a Poisson regression reveals that there is no evidence that homicide rates depend on the month, but there is a significant "Saturday effect" of around a 60% increase in homicide rate compared with all other days of the week combined. However, taking this effect into account does not make much difference to the predictions or to the quality of the fit.

Can we predict how often there will be long gaps between murders?

The average gap between murders is 54 hours. But they do not happen in a regular pattern and the gap can be much less and much more than 54 hours. We can use our probability model to predict how often there will be murders on consecutive days, how often there will be a day without a murder, and so on. The predicted pattern for 2004–2007 is shown in Figure 4: we predict that there will be 135 occasions on which there are murders on consecutive days, but also 18 occasions in which there is a clear gap of at least 7 days between murders, and two occasions in which there is a gap of 12 days.

The observed pattern for 2004–2007 is also shown in Figure 4, and follows the predicted pattern reasonably well: there were actually 143 occasions on which there were murders on consecutive days, 19 occasions in which there was a gap of at least 7 days between murders, and one occasion in which there was a gap of 12 days (between February 15th and February 26th, 2007).

Did we predict correctly how many murders there would be in London in 2008?

If we assume that there is no major change in circumstances year on year, then we can predict ahead of time how many murders will be recorded in London during any particular year.

How accurate will our predictions be? Using only data from the years 2004–2007, we tried to predict the number of London murders that 2008 would see.

We did the exercise in October 2008. Figure 5 shows that we expected there to be a total of 161 murders during 2008 and, although it was unlikely to be exactly that number, we could be 95% sure that the total would be between 137 and 187. We can also read off the totals for specific dates: for example by July 28th we would expect around 93 murders, and be 95% sure there will be between 74 and 112. In fact on July 28th the headline

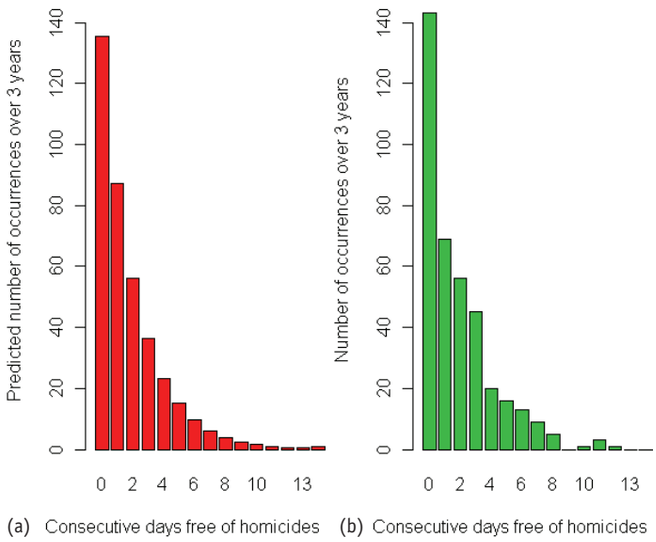

Figure 4. Frequency of different length intervals between homicides recorded in London between April 2004 and March 2007: (a) predicted; (b) actual

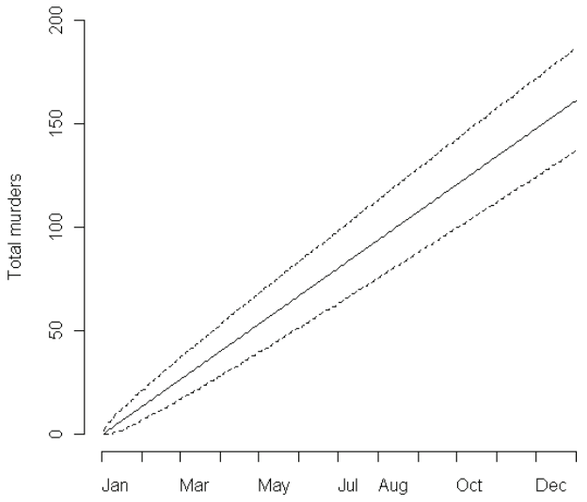

Figure 5. Predicted number of murders in London during 2008 (with 95% prediction limits)

"London's murder count reaches 90" was on the front page of a London paper—very close to the prediction we can make based on the historical data.

And how many murders were there in fact? We approached the Metropolitan Police press office on December 31st 2008 and they told us their latest estimate was 152 homicides in the calendar year 2008. Their spokesman was delighted: "We haven't seen so few for several years", he said. Does this mean that, despite the headlines, London is actually getting safer? A statistician would be cautious. The number is indeed lower than in any of the years that we used as data; but it is still well within our 95% prediction limits. Also the definitive measures for crime statistics are presented by the Home Office on an April/March basis and not a calendar year: these showed a small decrease from a revised 166 in 2006/07 to 162 in 2007/08, but 2008/09 is not yet over as this goes to print. We must keep in mind the UK Statistics Authority's recent warnings about the dangers of selective use of data, however well intentioned. Our assumption that there has been no major change in London has not been disproved.

So we cannot predict individual murders, but their pattern is highly predictable. This should mean that we can be ready for events that appear to be good (a long gap between murders) or bad (three or more murders on the same day)—both events are to be expected by chance alone. But, by knowing what pattern to expect, then we should also be able to spot when something really unusual is happening. The 2008 figure was low but it was not really unusual.

No undue panic; no false reassurance

As statisticians we should be well placed to provide this kind of analysis, but for it to be most useful we need to think carefully about the most appropriate way to present the data for different audiences. For the media and public consumption, our comments must remain firmly based on the data and should always seek to avoid either unnecessary alarm or unwarranted complacency. For policymakers, the analysis should rather be focused to support proportionate risk-based decision-making, avoiding unnecessary new initiatives and making best use of existing resources. For example, after the four murders in one day, resources could perhaps best be deployed in supporting those affected, catching the perpetrators and reassuring the public.

The use of statistical control chart methodology such as cumulative sums may be appropriate to identify as rapidly as possible when rates are either going up or, hopefully, going down. Such techniques are being increasingly used in a health context and might have been effective in the early detection of serial murderer Harold Shipman had suitable data been available[5]. The potential for careful statistical analysis stretches far beyond policing into other public risk arenas that policymakers are asked to regulate, even if the relative scarcity of some events presents a considerable challenge.

Success is likely to need cooperation between statisticians within and outside government, and their respective policy customers, to identify relevant problems and appropriate analytic approaches. In this way statisticians may be able to play a valuable role in ensuring that tragedies such as London's violent deaths are not forgotten but nor are they the cause of a disproportionate response—whether that be undue panic or false reassurance.

References

1. BBC News Online (2008) Brown pledge to tackle stabbings. (Available from `http://news.bbc.co.uk/1/hi/uk/7502569.stm`.)

2. Coleman, K., Kaiza, P. Hoare, J. and Jansson, K. (2008) Homicides, firearm offences and intimate violence 2006/07. *Home Office Statistical Bulletin*, Jan. 31st, 2008. (Available from `http://www.homeoffice.gov.uk/rds/pdfs08/hosb0308.pdf`.)

3. Flanagan, R. (2008) *Independent Review of Policing by Sir Ronnie Flanagan—Final Report*. London: Home Office. (Available from `http://police.homeoffice.gov.uk/publications/police-reform/Review_of_policing_final_report/`.)

4. Preece, D. A., Ross, G. J. S. and Kirby, P. J. (1988) Bortkewitsch's horse-kicks and the generalised linear model. *The Statistician*, **37**, 313–318.

5. Spiegelhalter, D. J. and Best, N. G. (2004) Shipman's statistical legacy. *Significance*, **1**, 10–12.

David Spiegelhalter is Winton Professor for the Public Understanding of Risk at the Statistical Laboratory in the University of Cambridge, and Senior Scientist in the MRC Biostatistics Unit Cambridge. Email david@statslab.cam.ac.uk

Arthur Barnett is a statistician in the Department for Business Enterprise and Regulatory Reform (BERR) Strategic Policy Analysis directorate (SPA) working with the Risk and Regulation Advisory Council (RRAC). email arthur.barnett@berr.gsi.gov.uk