CrossMark

ORIGINAL PAPER

Measuring the Distribution of Crime and Its Concentration

Rafael Prieto Curiel[1] · Sofía Collignon Delmar[2] ·
Steven Richard Bishop[1]

Abstract

Objectives Generally speaking, crime is, fortunately, a rare event. As far as modelling is concerned, this sparsity of data means that traditional measures to quantify concentration are not appropriate when applied to crime suffered by a population. Our objective is to develop a new technique to measure the concentration of crime which takes into account its low frequency of occurrence and its high degree of concentration in such a way that this measure is comparable over time and over different populations.
Methods This article derives an estimate of the distribution of crime suffered by a population based on a mixture model and then evaluates a new and standardised measurement of the concentration of the rates of suffering a crime based on that distribution.
Results The new measure is successfully applied to the incidence of robbery of a person in Mexico and is able to correctly quantify the concentration crime in such a way that is comparable between different regions and can be tracked over different time periods.
Conclusions The risk of suffering a crime is not uniformly distributed across a population. There are certain groups which are statistically immune to suffering crime but there are also groups which suffer chronic victimisation. This measure improves our understanding of how patterns of crime can be quantified allowing us to determine if a

✉ Rafael Prieto Curiel
rafael.prieto.13@ucl.ac.uk

Sofía Collignon Delmar
a.delmar.12@ucl.ac.uk

Steven Richard Bishop
s.bishop@ucl.ac.uk

[1] Department of Mathematics, University College London, Gower Street, London WC1E 6BT, UK

[2] Department of Political Science, University College London, Gower Street, London WC1E 6BT, UK

prevention policy results in a crime reduction rather than target displacement. The method may have applications beyond crime science.

Introduction

The risk of suffering a crime is not uniformly distributed over a region (Johnson 2010) and nor is it uniformly distributed across members of the same community (Grove et al. 2012), with some regions and some population groups more affected by crime than others (Farrell 2015). In the case of burglary, for example, it has been shown that houses in deprived areas suffer a higher risk of being the target of a crime, whilst other regions appear to be immune to that type of crime (Bowers et al. 2005). Equally, whether it is determined by race, age, income or many other factors, it has been found that the risk of suffering a robbery of a person is not uniformly distributed across all the members of the same population. Thus, just as there are places in which crime is concentrated (Freeman 1996), there are population groups with a higher risk of suffering a crime (Farrell and Pease 1993).

As a result of this risk heterogeneity, crime is highly concentrated in certain population groups. For example, a landmark analysis of victimisation data from the British Crime Survey showed that 2% of people who suffer the highest number of personal crimes, in fact, suffered 66% of the total reported for that type of crime (Pease 1998). This has often been attributed to the *attractiveness* of a place or a person (Brantingham and Brantingham 2010), the interaction in space and time of motivated offenders with suitable targets and the absence of any deterrence system, such as a police, a security guardian or perhaps even ordinary citizens (Hindelang et al. 1978; Cohen and Felson 1979) or based on some other theory (Stark 1987). But, how far away from a homogeneous distribution is the crime suffered by the population in a region and how can we quantify it? Does the distribution of crime, as well as its concentration, change over time? These questions are highly relevant for decision-makers interested in designing policies to reduce the levels of crime.

Measuring the number of crimes that a person suffers, quantifying its concentration and understanding the reasons why one person is victimised more frequently than others, has a counterpart in the number of crimes committed by potential offenders. Relevant questions include, how many criminals are there in a region? And if crime increases, does it mean that there are more criminals, or that a few individuals have become more active? These questions have been of interest to criminologists for years. For example, Wolfgang's classic study of a birth cohort in Philadelphia found that the majority of the population had no contact with the police, but at the same time there was a small group (less than 7% of the population) that was responsible for the majority of the crimes committed by that cohort (Wolfgang et al. 1987). Also, by considering families and not only the individuals within a family, it was shown that less than 5% of the families account for more than 30% of the arrests (Farrington et al. 2001). Thus, crime is highly concentrated in the regions in which it is executed, in the population that suffers it and in the people that commit it.

Unfortunately, current measures for the concentration of crime do not take into account the relative rarity of these events and so do not provide particularly useful information that would be necessary for policy design and decision-making. This results in misleading measures that underestimate crime concentration and which are not comparable over time,

potentially leading to vulnerable groups being wrongly targeted while others are overlooked.

The traditional and most frequent measurement of the concentration of crime suffered by a population is provided by the average number of crimes suffered by a single victim. Formally, we consider a population of size N and let V be the number of people within the population who suffered a particular type of crime during a given time period (usually one year). The *victimisation rate* (v), also known as the *prevalence*, is then defined as

$$\text{victimisation rate} = v = \frac{V}{N}. \tag{1}$$

We can interpret this by saying that if we take a person at random, then v is the probability that they suffered that type of crime of during the time period being considered. Now, let C be the number of times that a particular type of crime was committed on the whole population during the same time period. The *crime rate* (c), also known as *incidence*, is defined by

$$\text{crime rate} = c = \frac{C}{N}, \tag{2}$$

thus, c is defined as the proportion of the number of crimes to the population size.

Both, the crime rate c and the victimisation rate v are usually reported with regards to a population of 100,000 individuals and, based on these two measures, a frequently used metric is the *concentration of crime* (H), given by the ratio of the crime rate c to the victimisation rate v, that is

$$H = \frac{\text{crime rate}}{\text{victimisation rate}} = \frac{c}{v} = \frac{C}{V}. \tag{3}$$

Assuming that each crime is assigned to a single victim, $H \geq 1$, and so H is a measure which can be interpreted as the average number of crimes suffered by the victims of that type of crime. H has been used, for example, to measure the concentration of burglary, referred to in that case as the *burglary concentration* (Tseloni et al. 2004).

Although H is frequently used to measure the concentration of crime, it has severe flaws because it does not help us determine if crime is, in fact, more or less concentrated. To illustrate this, here we use a simple example to show that it is a poor summary of the crime suffered. Consider two populations, A and B, both with a population size of $N = 100,000$, and both suffering the same crime rate of $c = 0.1$ and the same victimisation rate of $v = 0.05$ so that in both populations we have the same number of victims ($V = 5,000$) and experience the same number of crimes ($C = 10,000$), hence we obtain the same concentration of crime $H_A = H_B = 2$ (where the subscript denotes the population to which it is applied) . However, consider the artificial situation in which for the population A each victim suffered exactly two crimes, but in the population B there were 4,000 victims suffering a single crime each and 1,000 victims suffering 6 crimes each. Clearly this single measure of $H_A = H_B = 2$ does not help us differentiate the construction of these two distinct scenarios where, in the population B, the 1% of the population who suffers the highest amount of crime actually suffers 60% of the crime which is a completely different behaviour than that observed in the population A.

The level in which crime is concentrated should have an impact on the design of policies. For example, if we know that 1% of the population suffers 60% of the crime, as in the example case of the population B, then efforts might be better directed towards that particular population group, both in terms of crime prevention and victim support. Hence, a

better quantitative approach needs to be developed to measure the level in which crime is concentrated.

Measures of variance, such as the coefficient of variation or the standard deviation, might be applied directly to the number of crimes suffered by each person or might be applied to the number of crimes suffered only by the actual victims. Results in the first case tend to be highly dependant on the crime rate c and results in the second case ignore whether 50 or 95% of the population did not suffer any crime. Other measurements of the level of concentration of crime have been proposed, for example, by taking into account the time between repeat offences (Short et al. 2009). However, this results in a measurement that cannot be compared between different regions, over different time periods or types of crime, due to the different crime intensities. A more detailed discussion on the issues the most frequently used measures can be found in the Appendix of ESM.

In the specific case of a measure of spatial concentration of crime, a common technique is to consider the street segments in which the crime was committed (including, sometimes, the intersections) and then to determine the amount of crime concentrated in the top 5% of the segments Andresen et al. (2016) or any other top $P\%$. A similar type of metric is often used when the number of crimes suffered by the most victimised people is reported (Pease and Ignatans 2016; Pease 1998), or the most criminal individuals (Wolfgang et al. 1987) or families (Farrington et al. 2001). This metric, however, has some severe issues, such as the lack of agreement on the percentage that gets reported (Fox and Tracy 1988); the metric might not be comparable between different cities (Hipp and Kim 2016); it might be the result of a certain degree of randomness (Levin et al. 2016) and it does not work as an adequate metric when the data is extremely sparse. Consider, for example, the number of street segments of The Hague and the number of sexual offences registered by the police between 2007 and 2009 in that city (Bernasco and Steenbeek 2016). The extremely low frequency of this type of crime (only 430 cases) distributed over the large number of street segments (14,375 segments) means that taking the top 5% of streets is not even properly defined since, at most, 3% of the segments concentrate all the crimes. Taking the top 5% street segments, victims or criminals is a weak way of measuring crime concentration based on an artificial cutoff point which is blind, not only to the other 95% of the observations, but also, it is blind to the distribution inside the 5% being considered, which is extremely relevant when the events are rare, as crime usually is.

An adequate metric to determine the statistical dispersion of crime is to consider the Lorenz curve and its corresponding Gini coefficient G (Fox and Tracy 1988) since it is a global metric that does not depend on an artificial cutoff point. In the artificial case that one individual suffers all the crime or all the events are concentrated on one street segment, or in the real case of the sexual offences in The Hague, the Gini coefficient gives a value close to one, indicating a high degree of concentration. However, there is still a major drawback in using the Gini coefficient directly from the number of crimes suffered by the population. The Gini coefficient is a valid metric with distinct values in the context where a variable, such as income, is distributed across most of the members of the population, but in the case of crime, the majority of the population suffers zero crimes and so the coefficient, computed directly from the number of crimes suffered by the population (or the number of crimes on each street segment) overestimates the level of concentration (Bernasco and Steenbeek 2016). In the previous example of populations A and B, the Gini coefficient of the number of crimes is fairly similar given by $G_A = 0.95$ and $G_B = 0.97$ respectively, which reveals that crime is highly concentrated, but nothing more, providing little additional information to distinguish between the concentration observed in A and B.

When there are more individuals than crimes, as it is almost always the case, or more street segments than crimes, as it sometimes occurs, then an arithmetic adjustment to the Gini coefficient has been proposed (Bernasco and Steenbeek 2016) which compares against a case of maximum equality (termed the generalised Gini coefficient denoted as G'). Although the generalised Gini coefficient is a clever way to correct the traditional Gini coefficient, it still has one major drawback. Consider two populations, B and C, where, as before, population B has a size of $N = 100,000$ individuals, and has 4,000 victims suffering a single crime each and 1,000 victims suffering 6 crimes each while the population C has only $N = 10,000$ individuals (that is only 10% of the size of B) where the population C has (just like population B) 4,000 victims suffering a single crime each and 1,000 victims suffering 6 crimes each. We observe that populations B and C suffer crime under a different pattern and have a completely different concentration of crime since 95% of the population of B did not suffer any crime, whereas 50% of the C population suffered at least one crime. However, in this artificial but illustrative example, B and C have the same line of "maximal equality" and the same corrected Gini coefficient $G'_B = G'_C = 0.7$, precisely highlighting the main weakness of this arithmetic correction of the Gini coefficient: in the case of the population B, it *corrects* the traditional Gini coefficient by ignoring 90% of the population, but in the case of the population C it takes all of its individuals into account. In fact, any population with 4,000 victims suffering a single crime each and 1,000 victims suffering 6 crimes each, with a population size of $N \geq 10,000$ gives the same corrected Gini coefficient $G' = 0.7$ regardless of whether the population has only 10,000 inhabitants or millions. Thus, the arithmetic correction to the Gini coefficient creates another issue that the original Gini coefficient did not have (since $G_B = 0.97$ and $G_C = 0.7$).

Very recent developments, responding perhaps to issues raised by the *law of crime concentration* (Weisburd 2015), have highlighted the need for more specific tools in the field of crime science. Here, we construct a new measure of the concentration of crime rates suffered by a population which overcomes problems encountered when using other measures and descriptive statistics as metrics. This new measure can be used to compare different regions and it is also tractable across various time periods and therefore it can be used for purposes such as policy design, policing and crime prevention (Laycock and Farrell 2003).

A Probabilistic Approach to the Crime and Victimisation Rates

Assuming that the number of crimes suffered by the individuals within a population is independent and that suffering a crime does not affect the probability of a person being a victim in the future, then the number of crimes suffered by the i-th individual during a period of time (usually a year) might be modelled as a Poisson distribution with rate $\lambda_i \geq 0$. Although other distributions could be used for modelling the random component of suffering crime, such as a Negative Binomial (Park and Eck 2013), the Poisson distribution allows to focus on a single parameter (the rate λ), and so it is frequently used in crime science (Maltz 1996).

What is relevant about this approach for the distribution of the crime rates is that it is probabilistic: even when a person has a rate of $\lambda_j > 0$ of suffering a crime, the probability that he or she does not suffer any crime is not negligible, given by $\exp(-\lambda_j)$, therefore, even when a person did not suffer a crime, it does not necessarily mean that he or she has a crime rate of $\lambda = 0$, and vice versa, if a person suffers many crimes, it could be the result of

a small rate and bad luck. By focusing on the rates λ_i rather than on the observed frequency of crime, we consider not only the population who actually suffered crime, but also the people who did not suffer any crime but were lucky, in the sense that given that their rate is greater than zero they were fortunate not to suffer any crime.

The assumptions required to model the crime counts as a Poisson distribution (independence of the crime suffered by individuals, independence between past and future victimisation and a constant rate) are quite problematic, especially since we know that past victimisation actually helps predict future victimisation (Tseloni and Pease 2003), crime suffered by individuals might be strongly correlated, for example, if individuals live nearby and finally, certain types of crime are more likely to occur at specific times of the day. These assumptions are thus too strong and hence the results are not really useful if we use the Poisson distribution to forecasting the number of crimes that a person will suffer, for instance. However, the objective here is to construct a global metric of the concentration of crime and so these assumptions, although apparently unrealistic, help us measure the concentration based on the least possible aggregated observations, usually from large populations but for which the number of crimes suffered is quite small. Other scenarios are discussed later in this paper, where different assumptions are made.

Since we consider a distinct crime rate for each individual, here the number λ_i refers to the *individual crime rate* and it represents the rate or "speed" at which the i-th individual suffers crime. The reasons why individuals experience different rates have been considered in depth by others and explanations go from individual attributes, which cause an increase in the attractiveness, to a boost on the probability of suffering a second crime after suffering a first one (Johnson et al. 2009). We referred to this as a population which suffers an inhomogeneous distribution of crime.[1]

The causal mechanism that leads to a population suffering an inhomogeneous rate has been studied before (Tseloni and Pease 2004), but the focus of this study is the distribution itself, so here we assume different individual rates, without going any further into this topic. Considering a probabilistic approach means that the actual number of crimes suffered by the i-th individual is an observation from the Poisson distribution, so we can analyse the individual crime rates rather than the observations. The expected number of crimes suffered by the population is simply the sum of the rates λ_i, from which

$$C = \sum_{i=1}^{N} \lambda_i, \tag{4}$$

which also means that the average rate $\bar{\lambda}$ is the population crime rate $c = C/N$.

The probability that the i-th person actually suffers a crime is $(1 - \exp(-\lambda_i))$, which means that the number of victims V follows a Poisson-Binomial distribution with parameters $p_i = 1 - \exp(-\lambda_i)$, with $i \in 1, 2, \ldots, N$. The Poisson-Binomial distribution is closely related to a Binomial distribution, in which each observation is allowed to have a different success probability (Chen and Liu 1997). The expected number of victims is

$$V = \sum_{i=1}^{N} p_i = N - \sum_{i=1}^{N} \exp(-\lambda_i). \tag{5}$$

Since the number of crimes C and the number of victims V from the population is a fixed (observed) number, we thus have three restrictions for the rates λ_i:

---

[1] We use the term "inhomogeneous" here rather than "heterogeneous" for consistency purposes with other applications of a Poisson process.

$$\text{Restriction I}: \quad \lambda_i \geq 0 \text{ for all } i = 1, 2, \ldots, N.$$
$$\text{Restriction II}: \quad \sum_{i=1}^{N} \lambda_i = C.$$
$$\text{Restriction III}: \quad N - \sum_{i=1}^{N} \exp(-\lambda_i) = V.$$

The restrictions II and III refer to the expected value of the number of crimes C and victims V (so the left-hand side is not necessarily a whole number), which means that both are considered to be satisfied in an interval around each integer C and V. Additional restrictions could also be considered by taking into account the observed number of people who suffered exactly two crimes, or exactly three crimes and so on against the theoretical (expected) outcome. Different distributions of the individual rates λ_i, with $i = 1, 2, \ldots, N$ which satisfy these restrictions could be the distribution of the crime rates over the whole population, and a goodness of fit test could help us accept or reject the distribution of the individual rates λ_i.

A homogeneous distribution of the rates means that λ_i is the same for all $i = 1, 2, \ldots, N$, from which, due to the Restriction II, we see that $\lambda_i = c$. If (and only if) c and v are such that $(1 - \exp(-c)) = v$, then a homogeneous distribution of the individual rates might be accepted. However, this is rarely the case since crime is far more concentrated than random events would predict (Osborn and Tseloni 1998); usually v is much smaller than $(1 - \exp(-c))$, which tells us that a homogeneous distribution is far from being the observed one and other distribution of the rates λ_i needs to be considered. Here, we present a potential distribution of the individual crime rates λ_i.

Although the methodology presented here, by modelling the distribution of the victimisation rates, is designed for the number of crimes suffered by individuals, with certain precautions it could also be applied to other aspects of crime, for instance, the concentration of criminality (the observations could also be the number of crimes executed by a person) or the spatial concentration of crime (so the observations could be the number of crimes committed on a street segment). In the case of the crime committed by the population, the assumption that past and future events are independent is strong, since it has been found that the rate at which an individual executes crime tends to increase as they commit more crimes (Ferguson 1952) meaning that a constant rate is dubious. In the case of the spatial distribution, crime suffered in street segments (Weisburd 2015) or regions of a city (Mohler et al. 2012) tends to be highly concentrated and a hot spot pattern usually emerges, indicating a geographic clustering of crime (Short et al. 2008), thus, assuming independence between the observations might not be adequate. The methodology presented could be used for measuring the concentration of crime in the other two aspects (the location and the criminal) but we focus here on the crime suffered by the victims.

Inhomogeneous Distribution of the Crime Rates

The individual crime rates λ_i depend on many factors (Tseloni 2000), such as the habits of the i-th person, their lifestyle, the region in which he or she usually commutes, physical attributes (such as gender or age), and perhaps that rate is of a similar value to other individuals who live under the same circumstances. To model the inhomogeneous distribution of the crime rates, we assume that the population can be divided into $k \geq 1$ distinct groups, where group j say, has Q_j members, all of whom suffer the same crime rate λ_j, with $j = 1, 2, \ldots, k$. Each of the N members of the whole population belongs to one and only one group so that $Q_1 + Q_2 + \ldots + Q_k = N$. The proportion of the population who suffer the crime rate λ_j is $q_j = Q_j/N$, so that $\sum_{j=1}^{k} q_j = 1$. To avoid ambiguous definitions, we order

the groups by their crime rate in increasing order, $\lambda_1 < \lambda_2 < \ldots < \lambda_k$, so we label the groups according to their rate.

If we consider a random person from the population, the distribution of the number of crimes that he or she suffered can be expressed as

$$q_1\text{Pois}(\lambda_1) + q_2\text{Pois}(\lambda_2) + \ldots + q_k\text{Pois}(\lambda_k), \tag{6}$$

which means that the person is assigned into any of the k groups and suffers a Poisson distribution with the corresponding rate.

Assuming that the population can be divided into $k \geq 1$ distinct groups where all of the members of each group suffer the same rate is a common technique that simplifies the crime suffered by a population of perhaps millions of people into only a few parameters (Short et al. 2009; Brame et al. 2006; Nagin and Land 1993). The number of groups, k, is crucial for the mixture model (Böhning et al. 1992), and distributions with a larger number of groups are less useful since for each additional group, its size and its rate need to be estimated, so this increases the number of parameters of the model. The (non-parametric) maximum likelihood estimator (*mle* or *npmle*) helps us compare between models and to pick the best (Böhning et al. 1998), since in our case we have no prior information on the number of groups (McLachlan and Peel 2004). Although other techniques to estimate the number of groups are also available, for example, by using bootstrapping (Schlattmann 2005), the *mle* is used here, which includes an estimate $\hat{k}$ of the number of population groups, easily computed using the statistical package CAMAN (Computer Assisted Analysis of Mixtures) by considering the observed number of crimes suffered by each of the individuals, C_i. A similar procedure, using a mixture model, has been used in different scenarios (Böhning 1998), such as traffic accidents, the number of accidents in a factory and even the number of criminal acts from a set of persons considered to have deviant behaviour.

The results obtained using CAMAN and estimating the *mle* are:

- the optimal number of groups in which the population might be divided $\hat{k}$,
- the size of each population group relative to the total population $\hat{q}_j$, expressed as a vector as $\underline{q}$, and
- the corresponding rate for each group $\hat{\lambda}_j$, also expressed as a vector $\underline{\lambda}$.

As an example, we consider again the previous populations A and B, both with $N = 100,000$, $c = 0.1$ and $v = 0.05$. In population A (where 10,000 crimes are suffered uniformly by the 5,000 victims), the *mle* gives $\hat{k} = 2$ groups, with $\underline{q} = (0.937, 0.063)$ and $\underline{\lambda} = (0, 1.594)$, which means that 93.7% of the population has a crime rate of $\hat{\lambda}_1 = 0$ and 6.3% of the population has a crime rate of $\hat{\lambda}_2 = 1.594$. On the other hand, in the population B (where 6,000 crimes are suffered by 1,000 victims and 4,000 crimes are suffered by 4,000 victims), the *mle* gives $\hat{k} = 3$ groups, with $\underline{q} = (0.409, 0.580, 0.010)$ and $\underline{\lambda} = (0, 0.069, 5.897)$, which means that the mixture model tells us that indeed 1% of the population suffers crime with rate $\hat{\lambda}_3 = 5.897$, but also that 58% of the population suffers crime at a rate of $\hat{\lambda}_2 = 0.069$, so that if we randomly select 15 individuals from that group, we would expect to find only one victim. Results indicate that a large proportion of the population (more than half) suffers crime at a very small rate, but there is a particular group, (formed by only 1% of the individuals) whose members suffer a considerably large crime rate ($\hat{\lambda}_3 = 5.89$). In the population B, efforts might be much better oriented towards

the small population group who expect to suffer $\hat{\lambda}_3 = 5.89$ crimes during that time period rather than the large population group who suffer the lower rate $\hat{\lambda}_2 = 0.069$.

The mixture model process depends not only on c and v, but it also changes based on the number of people who suffered exactly 0, 1, 2, 3 or more crimes. It gives a more accurate distribution of the crime rates, considering the random component of suffering a crime, hence it provides a better understanding of the distribution of crime in the population since it estimates that a group of relative size $\hat{q}_j$ suffers crime at a rate $\hat{\lambda}_j$. This distribution is useful since it allows us firstly, to compare different regions or different time periods, secondly, to simulate crimes in a population, and thirdly, to determine the expected departures that natural variability gives to the number of crimes suffered by the population. The mixture model is useful from a macro perspective giving an approximate distribution of the number of crimes over the whole population in terms of only a few parameters. It is not useful, though, from an individual perspective. For example, results for population B are that nearly 60% suffers a crime rate $\lambda > 0$ but it comes from a population where 95% suffered zero crimes, meaning that if a person suffered no crimes we would not be able to tell whether they belong to the group that suffers no crime or whether they belong to a group which suffers a small rate $\hat{\lambda}_2 = 0.069$ or even less likely, but not impossible, they belong to the group who suffers a large rate $\hat{\lambda}_3 = 5.89$ and yet they were lucky and suffered no crime.

Two particular cases of the results of the mixture model are worth further comment. The first is when $\hat{k} = 1$ (which means that there is only one group, with rate $\hat{\lambda}_1 = c$), in which case the mixture model tells that crime is uniformly suffered by the population. The second case is when $\hat{k} = 2$ and $\hat{\lambda}_1 = 0$, which indicates that the population is divided into two groups, one of them of relative size $\hat{q}_1$ which does not suffer crime and the second group, of relative size $\hat{q}_2 = 1 - \hat{q}_1$, suffers all the crime within that population. This type of model is also known as a Zero-Inflated Poisson Model (Böhning 1998), frequently used to model heterogeneity in the rates (Bushway and Tahamont 2016) and for count data in which the number of zeros is frequent, such as here in which we consider the count of the number of people who suffered zero crimes. These two cases might result from data after fitting the mixture model, which means that the mixture model lets the data adjust to the most suitable distribution of the crime rates, without assuming anything about the uniformity of crime suffered by the whole population.

Immunity and Chronic Victimisation

As previously noted (Sparks 1981; Hope and Norris 2013), there is usually a population group which is immune to victimisation and the mixture model allows us to detect the existence of such a population group and determine its size. After analysing the data, if the results show that there are $\hat{k} \geq 2$ groups and $\hat{\lambda}_1 = 0$, then it means that indeed there is a group who is immune to crime and its relative size is given by $\hat{q}_1$. It is important to note that the existence of an immune group is the result of the data and the model rather than by an assumption. Equally, results might reveal that there is no immune group, in which case the smallest crime rate would be $\hat{\lambda}_1 > 0$.

Population groups that suffer chronic victimisation have also been noted previously (Hope and Trickett 2008), where again, their existence might be tested using the results from the mixture model. Results from the mixture model might also show that a population group which suffers a rate higher than $\lambda = 2$. Suffering two or more crimes per year is a

persistent and perhaps habitual victimisation and so groups which suffer a rate $\hat{\lambda}_k \geq 2$ are considered to suffer chronic victimisation.

In the previous example of populations A and B (with $N = 100,000$, $c = 0.1$ and $v = 0.05$), for the population A (where 10,000 crimes are suffered uniformly by the 5,000 victims), results of the mixture model showed that 93.7% of the population has a crime rate of $\hat{\lambda}_1 = 0$, which means that a large portion of the population is statistically immune to crime. For the population B (where 6,000 crimes are suffered by 1,000 victims and 4,000 crimes are suffered by 4,000 victims), results of the mixture model showed that 40.9% are immune to crime, but also, 1% of the population has a crime rate of $\hat{\lambda}_3 = 5.89$, which means that they expect to suffer chronic victimisation of almost six crimes each year.

Concentration of Crime Metric

Although the distribution of the rates $(\underline{q}, \underline{\lambda})$ is powerful by itself, the *Rare Event Concentration Coefficient* (*RECC*) (Prieto Curiel and Bishop 2016) is a new and standardised summary statistic from the mixture model, based on the Lorenz curve (Marsh and Elliott 2008; Hope and Norris 2013) and the Gini coefficient (Dorfman 1979) of the distribution of the crime rates. It is important to note that it is not the Gini coefficient computed directly from the number of crimes suffered by each member of the population [previously used to measure the concentration of crime (Tseloni and Pease 2005; Fox and Tracy 1988; Bernasco and Steenbeek 2016)], but rather it is the Gini coefficient of the rate at which individuals suffer crime. The *RECC* is given by

$$RECC = \frac{1}{2 \sum_{i=1}^{\hat{k}} \hat{\lambda}_i \hat{q}_i} \sum_{i=1}^{\hat{k}} \sum_{j=1}^{\hat{k}} \hat{q}_i \hat{q}_j |\hat{\lambda}_i - \hat{\lambda}_j|, \tag{7}$$

which is the Gini coefficient of a stepwise distribution and can be interpreted in a similar way to how the Gini coefficient is used in the case of the distribution of wealth: a smaller value of the *RECC* means a more homogeneous distribution of the crime rates across the population, and a value closer to one means that crime is more concentrated in some population groups. The *RECC* is comparable between different time periods and different regions and different types of crime.

Using again the example of populations A and B (both with $N = 100,000$, $c = 0.1$ and $v = 0.05$), for the population A (where 10,000 crimes are suffered uniformly by the 5,000 victims) the mixture model says that 93.72% of the population is considered immune to crime and so, the $RECC_A = 0.9372$. On the other hand, for the population B (where 6,000 crimes are suffered by 1,000 victims and 4,000 crimes are suffered by 4,000 victims) the $RECC_B = 0.7546$ which means that crime is suffered more homogeneously in the population B, perhaps an expected result since 59% of the population has a crime rate greater than zero.

Coefficient Interval

Is observing $RECC_A = 0.9372$ statistically different from $RECC_B = 0.7546$? We construct an interval for the *RECC* of each population based on a Monte Carlo method (Mooney 1997) which allows us to incorporate a level of uncertainty. This method assumes that the distribution $(\underline{q}, \underline{\lambda})$ is the *true* distribution of the crime rates and so that we can simulate N individuals which suffer crime with the distribution given in Eq. 6. Each one of the N

simulated individuals represents the number of crimes that a person taken at random from the population might suffer, given the true distribution of crime and by simulating N individuals we are thus considering the departures from the true distribution that the number of crimes the population could experience. By computing the mixture model of the simulated crimes and then considering its corresponding $RECC_{sim}$ we are thus taking into account just how low, or high, the $RECC$ could be, given the exact same distribution.

Following the same procedure a sufficient number of times (100 in our case) results in a simulated sample of potential values of the $RECC$. We subsequently consider the 95% intervals to avoid the extreme simulated values (Greenland 2004).

The results, in terms of the simulated $RECC$, are given in Table 1 in terms of the 95% lower and upper bound intervals.

These results show that with the true distribution observed in the population A, the $RECC$ does not achieve values as low as the ones obtained for population B. Therefore, with the simulated intervals, we can reject a Null Hypothesis that both populations have the same concentration of crime and so, thanks to the simulated intervals we have a statistical justification that both populations suffer a different concentration of crime. A more detailed explanation of the simulated intervals for the $RECC$ is included in the Appendix of ESM.

A relevant observation from the simulations is that the number of groups $\hat{k}_{sim}$ might change and also the sizes of the immune group and the chronic group might also change since, for example, suffering a small rate of $\hat{\lambda} = 0.01$ is almost the same as $\hat{\lambda} = 0$ but this difference would change the size of the immune group. Therefore, for comparing two different populations or comparing the same population over different time periods, a global metric, such as the $RECC$, provides more stable results.

Case Study

We use the case of Mexico to apply the mixture model. Its territory is divided into 32 states with a wide variety in terms of population size —some states have a population of just above 700,000 inhabitants (a population size similar to Luxembourg), while at the same time, there is a state with a population size nearly 23 times larger, of more than 16 million inhabitants (a population size similar to the Netherlands)— and this

Table 1 Group sizes $\underline{q}$, crime rate $\underline{\lambda}$ and intervals for the $RECC$ for the populations A and B

Population		Group			$RECC$	Lower bound	Upper bound
		1	2	3			
A	$\hat{q}_j$	**0.937**	0.063		0.9372	0.9369	0.9428
	$\hat{\lambda}_j$	*0*	1.594				
B	$\hat{q}_j$	**0.409**	0.580	***0.010***	0.7546	0.7117	0.7948
	$\hat{\lambda}_j$	*0*	0.069	<u>5.897</u>			

Immune population size are given in bold

Zero crime rate $\lambda_1 = 0$ are given in italic

Chronic population size is given in bold italic

Chronic crime rate $\lambda_j \geq 2$ is underlined

sub-division also considers Mexico City as a separate state, allowing us to detect whether crime tends to be more concentrated in regions with more inhabitants (Glaeser and Sacerdote 1996).

Data was obtained from an annual victimisation survey conducted yearly between 2011 and 2016 in Mexico. Thus, six years of data are available which allows us to measure the concentration of crime across time. Although previous victimisation surveys were also conducted in Mexico, the survey used here provides comparable data over different years and also provides the most up to date data (INEGI 2016), with micro-data available on-line.[2] For each year, more than 80,000 surveys were conducted, and its sampling method allows separate data for each of the 32 states. The survey contains an expansion factor, used to establish an estimate for the number of people who are represented by each survey respondent so that every person older than 18 years in the country is represented by a single survey respondent.

For different types of crime, such as robbery of a person, car theft or burglary, the person is asked whether he or she suffered that type of crime and the number of times that it occurred during the previous year. We mainly use the case of robbery of a person in our studies since it has the highest variability, from a crime rate as low as $c_{BCS} = 0.007$ (where the subscript denotes the state) to a crime rate as high as $c_{MEX} = 0.471$, nearly 68 times larger, so this particular type of crime allows us to detect whether higher crime rates are also associated with a higher concentration of crime. Also, the analysis of the crime rates from the 32 states in Mexico allows us to compare population groups, so detecting, for example, an immune group in the states with high crime rates implies that they live under better conditions, in terms of crime and security than some of the groups from states with low crime rates. Thus, living in a state with a lower crime rate is not necessarily preferable from an individualistic viewpoint. More specific details on why we focus on the case of robbery of a person in Mexico is included in the Appendix of ESM.

Results

At a national level, the Table 2 gives the number of crimes suffered by the survey respondents in Mexico for 2016, as well as the national estimate, considering the expansion factor for each survey. The data shows that 91.9% of the population did not suffer a robbery of a person during 2015, but also, it is estimated that more than 100,000 persons suffered at least four robberies during that year.

Is the crime suffered the result of a homogeneous distribution? We use one state (Guerrero) and over one year (2016) to test against a random distribution of crime (Park and Eck 2013). With the observed number of crimes, $c = 0.069$ we would expect, from Eq. 5, a victimisation rate of $v = 0.066$ and values between 0.065 and 0.068 would support this hypothesis. However, the observed victimisation rate ($v = 0.053$) is far from this interval so that, in this state, crime is far from being homogeneously suffered by the population and there are much fewer victims than the homogeneous distribution would indicate. Similar results occur for other states and so there is, indeed, a high concentration of crime.

The distribution of the rates for each state and for each year have been computed (R Core Team 2014; Schlattmann et al. 2015) based on data from the victimisation surveys and the concentration *RECC* and its corresponding intervals are displayed in Table 4. The

---

Table 2 Crimes suffered by the population, in thousands, estimated by considering the expansion factor for each survey respondent

Crimes suffered	0	1	2	3	4	5	6	7 or more
Respondents	81,672	3486	455	95	22	10	1	23
Population[a]	75,992.6	5604.5	769.5	193.1	72.7	23.1	3.9	5.9
%	91.9	6.8	0.9	0.2	0.1	0.0	0.0	0.0

[a] Population expressed in thousands meaning that 75.9 million people did not suffer any crime, 5.6 million people suffered exactly one crime and so on

full distribution of the rates for the latest two years (2015 and 2016) in Tables and and the remaining years can be found in the Appendix of ESM. The results show that crime has a completely different pattern across the 32 states from Mexico. For example, in some states (Baja California Sur in 2015 or Aguascalientes in 2016) the population can be divided into just two groups, the immune and the victimised. However, in Morelos, for example, for 2015 and 2016 the model gives 4 groups, which means that a more complex distribution of crime is needed.

Results for the 32 states, over the six years of data available, are displayed in Table 7 and they show that there is usually a large group which have a rate $\hat{\lambda}_1 = 0$, forming the group which statistically is immune to crime and its relative size reaches a value as high as 95.6% in the state of Campeche in the year 2015. This means that during that year, in that state, less than 5% of the population actually expected to suffer crime, but with a rate of $\hat{\lambda}_2 = 0.853$, higher than most of the groups from all the other states. Actually, considering the population size of each state, this small group from Campeche suffered a higher rate than 96.8% of the whole country. The results obtained support the theory of the existence of an immune population group (Hope and Trickett 2008) and its size on average through the six years considered, was 61.5% of the whole population.

Results also show that there are some states (15 out of the 32 states in 2015 and only 9 in 2016) with a group which suffers chronic victimisation, so they expect to suffer two or more crimes in a year. For example, in Estado de México in 2015, there is a small group (which represents approximately only 0.2% of the population), but which has a crime rate of $\hat{\lambda}_4 = 7.8$, which sadly means that they expect to suffer one robbery roughly every seven weeks.

Figure 1 shows the results from Mexico City during 2014 simply as an illustration of the crime rates suffered by the population. The upper diagram, part (a), gives the rates from the mixture model so it provides the *Victimisation Profile* of Mexico City during 2014. The lower diagram, part (b), gives the cumulative rates and the Lorenz curve, where the area shaded in colour grey is the distance to a uniform distribution of the rates.

Crime Rates and Crime Concentration

Another observation of the results is that lower (or higher) crime rates do not necessarily mean a lower (or higher) concentration of crime. For example, the state of Chiapas (CHIS) between 2015 and 2016 suffers a similar crime rate ($c_{CHIS,2015} = 0.037$ and $c_{CHIS,2016} = 0.034$ respectively) with an opposite pattern for the concentration of crime ($RECC_{CHIS,2015} = 0.897$ and $RECC_{CHIS,2016} = 0.228$). In general, though there is a national decrease in the concentration of crime between 2011 and 2016. Between 2011 and 2013 the average $RECC$, weighted by the population size of each state, was nearly 0.8 but it has gradually decreased to an average of $RECC = 0.688$ by 2015 and $RECC = 0.563$ by 2016.
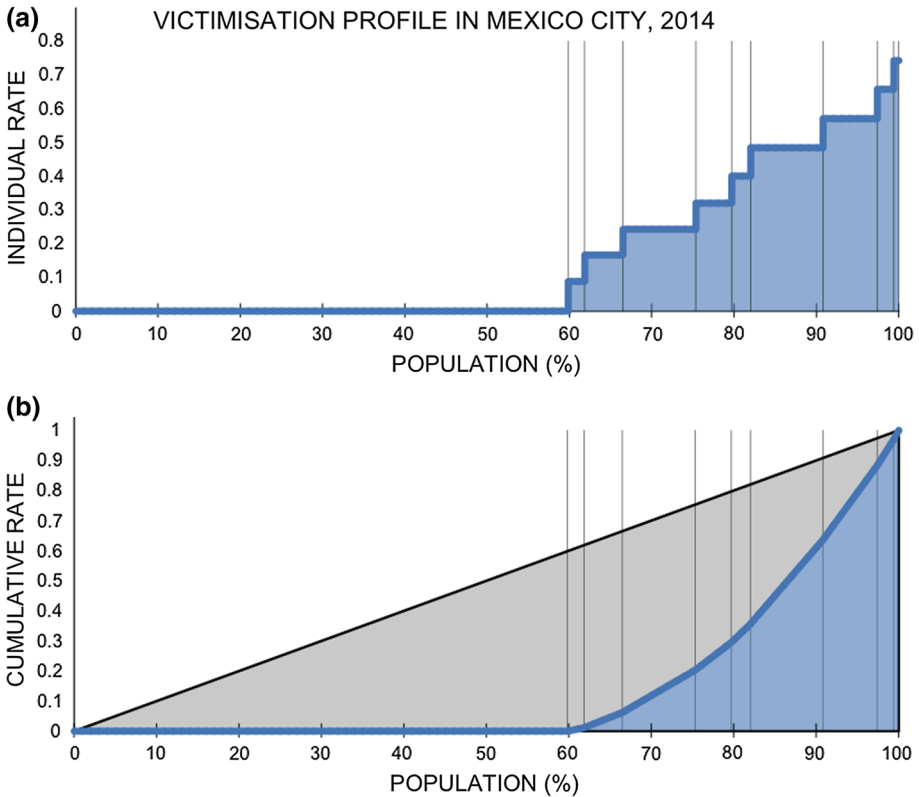
Fig. 1 **a** Victimisation Profile (individual crime rates $\underline{\lambda}$ and group sizes $\underline{q}$) and **b** Lorenz curve of the individual crime rates for Mexico City in 2014

At the same time, crime rates have increased slightly through this period, from a national rate near $c = 0.084$ to an average of $c = 0.105$. Thus, for this particular type of crime, there are now more robberies which are being suffered by more people.

Less or More Concentrated?

Having the estimated distribution of crime and its corresponding *RECC* enables us to understand the degree of concentration of crime and, using data from the victimisation survey, allows us to obtain quantitative results for the different states in Mexico during the years considered. What is not clear is what degree of concentration of crime is preferable.

Crime prevention strategies might result in some displacement of crime (Guerette and Bowers 2009; Johnson et al. 2014) from one place to the other, to an alternative victim (which is referred to as *target displacement*), to different times of the day, to a different tactic or to a different type of crime (Bowers and Johnson 2003), which has an effect on the levels of concentration of crime, but then this promotes the question: is it desirable to have less concentrated crime? Clearly, a population with overall less crime is desirable, but let us compare two populations with the same number of crimes. On the one hand, a high degree of concentration of crime means that fewer people suffer crime, that is, fewer

victims, but those victims suffer usually much more than a single crime. With a high degree of concentration of crime, resources might be better targeted to those who suffer most crime in terms of prevention and victim support. On the other hand, a low degree of concentration of crime means more victims, which makes policies less efficient, and it might deteriorate the perception of security in a particular region. Suffering a crime when there is a low degree of concentration of crime becomes a matter of *bad luck* and not a matter of being socially deprived, a minority, a female or any other attribute which perhaps increases the chances of suffering a crime, therefore it could be considered as a more fair distribution of crime (Bowers and Johnson 2003) as opposed to a population with a high degree of concentration of crime.

As an example, we analyse the particular case of Mexico City between 2011 and 2012. We focus on this particular example for three reasons. Mexico City is different from other states as it is the only one which is also a single metropolitan area so that any security programs are easily identifiable within its constraints. The majority of other states have three levels of police officers (federal, state and local) but in Mexico City, all security efforts are coordinated by the state police. Secondly, between 2010 and 2012, Mexico City started a security program which consisted of installing more than 13,000 CCTVs across the city and utilising several hundreds of police officers to perform surveillance[3] with a real-time police allocation strategy, investing nearly 500 million dollars in this program alone.[4] Thirdly, it has one of the most drastic changes of the *RECC* between two consecutive years meaning that the concentration of crime rates across its population significantly changed between these two years.

The most frequently used metrics for the concentration of crime in Mexico City were computed (Table 3) which show that with most of the traditional metrics we obtain almost the same results, for example, the Gini coefficient is $G_{CDMX,2011} = 0.898$ and $G_{CDMX,2012} = 0.872$ respectively, which only highlights the high concentration of crime, but a barely perceptible change and the reason is that in both years more than 85% of the population (that is, more than 85% of the observations) did not suffer any crime. The arithmetic correction of the Gini coefficient G' and the average number of crimes suffered by the victims H does show a slight difference between the two years. However, the individual rates for these two years are displayed in Fig. 2 where a drastic change is noticeable.

Using our method, for 2011 it was estimated that around 5% of the population suffered a rate of $\hat{\lambda} = 1.35$ and, perhaps due to a policy oriented to reduce the crime suffered by that population group, for 2012 this high rate was reduced to $\hat{\lambda} = 0.32$, which appears to be a good result. However, during 2011 more than 50% of the population were immune to crime, because their rate was exactly zero, but by 2012, only 28% of the population were immune to crime, and in fact, the global crime rate increased from $c_{CDMX,2011} = 0.170$ to $c_{CDMX,2012} = 0.173$.

Between 2011 and 2012, some people suffered a lower crime rate while some suffered a higher rate, which means that comparing individual rates does not provide much information (Fig. 2). However, for 2011 the $RECC_{CDMX,2011} = 0.680$, shows a much higher degree of concentration than for 2012, where the result was $RECC_{CDMX,2012} = 0.367$ and Table 4 shows that the values are statistically different. Thus, between 2011 and 2012, the results obtained in Mexico City indicate that there was not a reduction in the crime rates,

---

Table 3 Metrics of the concentration of crime observed in Mexico City in 2011 and 2012

Metric	Victim rate	Crime rate	Std. dev.	CoV.	Concentration	Gini	General Gini	Rare Event
	v	c	σ	CoV	H	G	G'	$RECC$
2011	0.127	0.170	0.507	2.990	1.330	0.898	0.401	0.680
2012	0.148	0.173	0.514	2.976	1.166	0.872	0.257	0.367
2011/ 2012	0.860	0.982	0.986	1.005	1.141	1.031	1.560	1.853
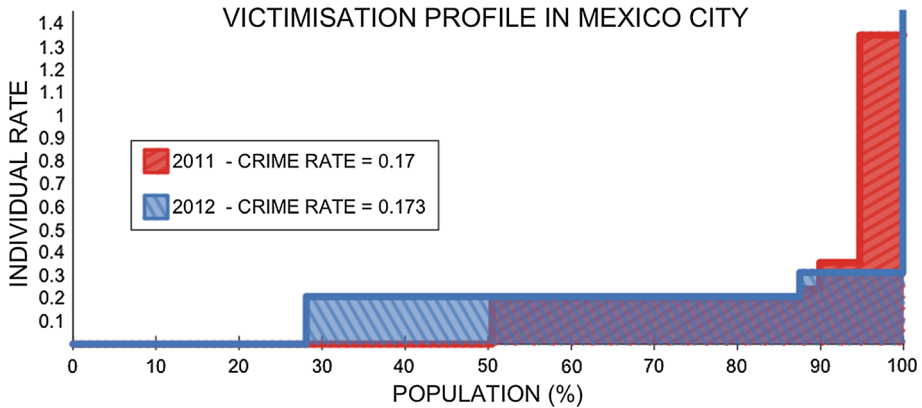
Fig. 2 Victimisation Profile (individual crime rates $\underline{\lambda}$) and group sizes $\underline{q}$) for Mexico City between 2011 and 2012

but rather it was merely a target displacement with a lower level of concentration of crime. In this period, the crime rate did not change, in fact increasing slightly, meaning that the surveillance program with a real-time police allocation strategy might have simply induced a target displacement.

Extensions of the Rare Event Concentration Coefficient

The *RECC* is a metric designed to detect changes in the concentration of crime based on the number of crimes suffered by the individuals. It considers that the observed numbers might be the result of luck and that fluctuations might be present so that it gives a probabilistic approach to the concentration of crime. It is based on some assumptions: individuals suffer crime independently, at a constant rate and suffering a crime does not affect future victimisation. Based on these assumptions, the number of crimes suffered by an individual might then be modelled as a Poisson distribution.

However, looking at the individual observations, such as the crimes suffered by a particular person, the crimes executed by an offender or even its arrests, is not consistent with a constant rate (Bushway and Tahamont 2016). Also, looking at a more aggregate type of data means reducing the number of observations, from millions of individuals into a few thousand street segments or regions and it reveals a pattern in terms of the space, the time or both, in which crime occurs (Osgood 2000). For more aggregate data, assuming

Table 4 Crime rates and *RECC* for the 32 states in Mexico between 2011 and 2016

State	Year	2011	2012	2013	2014	2015	2016
AGS	Crime rate	0.028	0.019	0.041	0.025	0.041	0.042
	RECC	0.848 ± 0.006	0.905 ± 0.004	0.927 ± 0.003	0.901 ± 0.006	0.853 ± 0.007	0.81 ± 0.009
BC	Crime rate	0.048	0.049	0.056	0.053	0.071	0.035
	RECC	0.696 ± 0.004	0.676 ± 0.004	0.744 ± 0.009	0.806 ± 0.009	0.866 ± 0.019	0.792 ± 0.013
BCS	Crime rate	0.007	0.011	0.021	0.016	0.019	0.024
	RECC	0.855 ± 0.005	0.922 ± 0.007	0.963 ± 0.014	0.834 ± 0.003	0.871 ± 0.01	0.89 ± 0.004
CAMP	Crime rate	0.018	0.022	0.035	0.039	0.038	0.036
	RECC	0.907 ± 0.008	0.936 ± 0.003	0.95 ± 0.014	0.803 ± 0.007	0.956 ± 0.001	0.871 ± 0.004
COA	Crime rate	0.018	0.034	0.038	0.043	0.04	0.036
	RECC	0.757 ± 0.055	0.944 ± 0.009	0.911 ± 0.004	0.901 ± 0.005	0.918 ± 0.003	0.926 ± 0.008
COL	Crime rate	0.016	0.011	0.022	0.018	0.015	0.018
	RECC	0.693 ± 0.019	0.927 ± 0.006	0.875 ± 0.009	0.937 ± 0.005	0.862 ± 0.002	0.735 ± 0.003
CHIS	Crime rate	0.035	0.026	0.029	0.043	0.037	0.034
	RECC	0.916 ± 0.002	0.818 ± 0.003	0.873 ± 0.005	0.851 ± 0.006	0.897 ± 0.003	0.228 ± 0.108
CHIH	Crime rate	0.068	0.061	0.049	0.062	0.043	0.045
	RECC	0.755 ± 0.042	0.718 ± 0.01	0.801 ± 0.026	0.842 ± 0.01	0.763 ± 0.005	0.898 ± 0.009
CDMX	Crime rate	0.17	0.173	0.165	0.152	0.208	0.218
	RECC	0.68 ± 0.029	0.367 ± 0.057	0.619 ± 0.003	0.699 ± 0.01	0.48 ± 0.02	0.321 ± 0.035
DGO	Crime rate	0.042	0.025	0.035	0.04	0.049	0.045
	RECC	0.644 ± 0.004	0.897 ± 0.004	0.857 ± 0.002	0.62 ± 0.004	0.898 ± 0.007	0.55 ± 0.056
GTO	Crime rate	0.048	0.059	0.074	0.066	0.093	0.079
	RECC	0.853 ± 0.006	0.738 ± 0.011	0.84 ± 0.005	0.808 ± 0.011	0.526 ± 0.098	0.603 ± 0.01
GRO	Crime rate	0.069	0.071	0.081	0.066	0.074	0.069
	RECC	0.867 ± 0.009	0.729 ± 0.013	0.817 ± 0.008	0.875 ± 0.004	0.854 ± 0.012	0.88 ± 0.003
HGO	Crime rate	0.047	0.049	0.045	0.051	0.041	0.029
	RECC	0.835 ± 0.005	0.88 ± 0.002	0.729 ± 0.005	0.865 ± 0.005	0.384 ± 0.076	0.847 ± 0.004

Table 4 continued

State	Year	2011	2012	2013	2014	2015	2016
JAL	Crime rate	0.065	0.063	0.083	0.088	0.068	0.091
	RECC	0.896 ± 0.002	0.811 ± 0.007	0.846 ± 0.017	0.849 ± 0.003	0.524 ± 0.072	0.547 ± 0.085
MEX	Crime rate	0.169	0.222	0.298	0.471	0.389	0.292
	RECC	0.573 ± 0.02	0.669 ± 0.019	0.626 ± 0.003	0.66 ± 0.019	0.672 ± 0.016	0.214 ± 0.078
MICH	Crime rate	0.029	0.04	0.043	0.049	0.04	0.031
	RECC	0.871 ± 0.007	0.809 ± 0.006	0.797 ± 0.006	0.87 ± 0.002	0.603 ± 0.056	0.813 ± 0.005
MOR	Crime rate	0.09	0.097	0.104	0.123	0.12	0.119
	RECC	0.76 ± 0.013	0.848 ± 0.017	0.806 ± 0.039	0.84 ± 0.012	0.727 ± 0.018	0.71 ± 0.017
NAY	Crime rate	0.023	0.02	0.02	0.021	0.022	0.02
	RECC	0.916 ± 0.004	0.897 ± 0.006	0.965 ± 0.002	0.887 ± 0.006	0.635 ± 0.083	0.245 ± 0.082
NL	Crime rate	0.053	0.064	0.096	0.087	0.079	0.053
	RECC	0.796 ± 0.007	0.695 ± 0.007	0.831 ± 0.007	0.846 ± 0.004	0.82 ± 0.003	0.404 ± 0.077
OAX	Crime rate	0.074	0.08	0.046	0.057	0.073	0.065
	RECC	0.905 ± 0.002	0.637 ± 0.009	0.902 ± 0.006	0.849 ± 0.012	0.65 ± 0.062	0.763 ± 0.004
PUE	Crime rate	0.058	0.086	0.081	0.073	0.079	0.071
	RECC	0.818 ± 0.005	0.836 ± 0.009	0.825 ± 0.014	0.791 ± 0.017	0.509 ± 0.082	0.561 ± 0.048
QRO	Crime rate	0.023	0.029	0.033	0.032	0.041	0.043
	RECC	0.756 ± 0.007	0.904 ± 0.004	0.885 ± 0.006	0.655 ± 0.004	0.912 ± 0.012	0.747 ± 0.009
QR	Crime rate	0.079	0.094	0.089	0.053	0.069	0.07
	RECC	0.853 ± 0.005	0.893 ± 0.003	0.9 ± 0.003	0.857 ± 0.026	0.675 ± 0.07	0.605 ± 0.075
SLP	Crime rate	0.036	0.069	0.073	0.078	0.06	0.035
	RECC	0.808 ± 0.006	0.904 ± 0.001	0.874 ± 0.016	0.917 ± 0.004	0.433 ± 0.093	0.447 ± 0.053
SIN	Crime rate	0.058	0.065	0.079	0.062	0.051	0.034
	RECC	0.793 ± 0.008	0.832 ± 0.025	0.83 ± 0.02	0.881 ± 0.004	0.863 ± 0.004	0.842 ± 0.002
SON	Crime rate	0.035	0.037	0.032	0.027	0.032	0.043
	RECC	0.903 ± 0.003	0.949 ± 0.002	0.795 ± 0.003	0.811 ± 0.003	0.912 ± 0.004	0.873 ± 0.002

Table 4 continued

State	Year	2011	2012	2013	2014	2015	2016
TAB	Crime rate	0.055	0.052	0.076	0.069	0.08	0.074
	RECC	0.611 ± 0.067	0.79 ± 0.01	0.85 ± 0.005	0.82 ± 0.007	0.308 ± 0.091	0.324 ± 0.108
TAMPS	Crime rate	0.031	0.047	0.051	0.035	0.057	0.025
	RECC	0.697 ± 0.005	0.843 ± 0.004	0.825 ± 0.017	0.875 ± 0.003	0.808 ± 0.029	0.498 ± 0.076
TLAX	Crime rate	0.035	0.054	0.045	0.046	0.077	0.056
	RECC	0.92 ± 0.003	0.906 ± 0.002	0.706 ± 0.032	0.88 ± 0.007	0.8 ± 0.007	0.85 ± 0.009
VER	Crime rate	0.034	0.032	0.041	0.043	0.035	0.035
	RECC	0.92 ± 0.003	0.802 ± 0.002	0.823 ± 0.009	0.826 ± 0.009	0.751 ± 0.009	0.562 ± 0.053
YUC	Crime rate	0.02	0.017	0.03	0.019	0.029	0.025
	RECC	0.844 ± 0.01	0.855 ± 0.007	0.824 ± 0.015	0.94 ± 0.004	0.932 ± 0.002	0.915 ± 0.002
ZAC	CRIME RATE	0.02	0.03	0.028	0.018	0.034	0.014
	RECC	0.899 ± 0.003	0.674 ± 0.059	0.923 ± 0.005	0.834 ± 0.004	0.938 ± 0.008	0.803 ± 0.003

Table 5 Group sizes $\underline{q}$ and crime rate $\underline{\lambda}$ for the 32 states in Mexico, based on data for robbery of a person, 2015

Year 2015		Population group			
		1	2	3	4
AGUASCALIENTES	q_j	**0.839**	0.159	*0.002*	
	λ_j	*0*	0.232	2.399	
BAJA CALIFORNIA	q_j	**0.635**	0.261	0.099	*0.006*
	λ_j	*0*	0.035	0.474	2.651
BAJA CALIFORNIA SUR	q_j	**0.871**	0.129		
	λ_j	*0*	0.146		
CAMPECHE	q_j	**0.956**	0.044		
	λ_j	*0*	0.853		
COAHUILA	q_j	**0.912**	0.086	0.002	
	λ_j	*0*	0.418	1.747	
COLIMA	q_j	**0.623**	0.353	0.024	
	λ_j	*0*	0.013	0.433	
CHIAPAS	q_j	**0.896**	0.012	0.092	
	λ_j	*0*	0.323	0.359	
CHIHUAHUA	q_j	**0.696**	0.296	0.008	
	λ_j	*0*	0.11	1.272	
CIUDAD DE MEXICO	q_j	**0.348**	0.071	0.557	0.025
	λ_j	*0*	0.036	0.31	1.361
DURANGO	q_j	**0.888**	0.111	*0.001*	
	λ_j	*0*	0.391	5.184	
GUANAJUATO	q_j	**0.379**	0.613	*0.008*	
	λ_j	*0*	0.114	2.861	
GUERRERO	q_j	**0.796**	0.097	0.095	*0.012*
	λ_j	*0*	0.249	0.262	2.092
HIDALGO	q_j	**0.033**	0.948	0.019	
	λ_j	*0*	0.027	0.832	
JALISCO	q_j	**0.11**	0.82	0.069	*0.001*
	λ_j	*0*	0.038	0.491	2.583
ESTADO DE MEXICO	q_j	**0.446**	0.383	0.168	*0.002*
	λ_j	*0*	0.302	1.545	7.833
MICHOACAN	q_j	**0.096**	0.878	0.026	
	λ_j	*0*	0.019	0.919	
MORELOS	q_j	**0.713**	0.028	0.257	*0.002*
	λ_j	*0*	0.382	0.4	3.532
NAYARIT	q_j	**0.201**	0.783	0.016	
	λ_j	*0*	0.012	0.762	
NUEVO LEON	q_j	**0.82**	0.18		
	λ_j	*0*	0.439		
OAXACA	q_j	**0.354**	0.616	0.03	
	λ_j	*0*	0.059	1.231	

Table 5 continued

Year 2015		Population group			
		1	2	3	4
PUEBLA	q_j	**0.014**	0.941	0.045	
	λ_j	*0*	0.038	0.973	
QUERETARO	q_j	**0.904**	0.095	***0.001***	
	λ_j	*0*	0.395	<u>3.872</u>	
QUINTANA ROO	q_j	**0.121**	0.825	0.052	*0.003*
	λ_j	*0*	0.026	0.807	<u>2.241</u>
SAN LUIS POTOSI	q_j	**0.003**	0.981	0.016	
	λ_j	*0*	0.034	1.7	
SINALOA	q_j	**0.854**	0.146	***0.001***	
	λ_j	*0*	0.33	<u>7.734</u>	
SONORA	q_j	**0.903**	0.096	***0.002***	
	λ_j	*0*	0.296	<u>2.322</u>	
TABASCO	q_j	**0.015**	0.966	0.018	***0.001***
	λ_j	*0*	0.057	1.208	<u>3.48</u>
TAMAULIPAS	q_j	**0.769**	0.229	***0.002***	
	λ_j	*0*	0.202	<u>4.821</u>	
TLAXCALA	q_j	**0.8**	0.2		
	λ_j	*0*	0.387		
VERACRUZ	q_j	**0.735**	0.264	***0.001***	
	λ_j	*0*	0.124	<u>2.531</u>	
YUCATAN	q_j	**0.922**	0.075	0.003	
	λ_j	*0*	0.325	1.681	
ZACATECAS	q_j	**0.938**	0.062		
	λ_j	*0*	0.545		

Immune population size are given in bold

Zero crime rate $\lambda_1 = 0$ are given in italic

Chronic population size are given in bold italic

Chronic crime rate $\lambda_j \geq 2$ are underlined

independence or a constant rate for each observation might not be appropriate, as most of the observations are not zero and so crime, within this more macro data frame, might not even be a rare event. Therefore, the *RECC* is not convenient to determine the concentration of crime within this context and more adequate tools for estimating the rate of crime of each of the *k* observations exist, for example, by considering the rate as a function of time *t* which might consider space, past victimisation, or even the topology of the street network (Rosser et al. 2016).

Provided that the rate of the *i*-th region (or a street segment, for instance) is modelled as a function of time $\hat{\lambda}_i(t)$ then measuring the concentration of crime within a more macro data set can be carried out by considering the *Event Concentration Coefficient* (*ECC*) (Prieto Curiel and Bishop 2016) which is constructed similarly, by computing the Gini

Table 6 Group sizes $\underline{q}$ and crime rate $\underline{\lambda}$ for the 32 states in Mexico, based on data for robbery of a person, 2016

Year 2016		Population group			
		1	2	3	4
AGUASCALIENTES	q_j	**0.81**	0.19		
	λ_j	0	0.219		
BAJA CALIFORNIA	q_j	**0.521**	0.387	0.092	
	λ_j	0	0.022	0.284	
BAJA CALIFORNIA SUR	q_j	**0.647**	0.317	0.034	*0.001*
	λ_j	0	0.017	0.453	2.652
CAMPECHE	q_j	**0.869**	0.024	0.107	
	λ_j	0	0.253	0.284	
COAHUILA	q_j	**0.919**	0.011	0.069	*0.001*
	λ_j	0	0.392	0.404	3.726
COLIMA	q_j	**0.726**	0.273	*0.001*	
	λ_j	0	0.063	2.605	
CHIAPAS	q_j	**0.013**	0.982	0.005	
	λ_j	0	0.027	1.666	
CHIHUAHUA	q_j	**0.871**	0.041	0.084	*0.004*
	λ_j	0	0.265	0.284	2.496
CIUDAD DE MEXICO	q_j	**0.254**	0.732	0.014	
	λ_j	0	0.265	1.681	
DURANGO	q_j	**0.11**	0.843	0.046	
	λ_j	0	0.024	0.532	
GUANAJUATO	q_j	**0.097**	0.778	0.125	
	λ_j	0	0.031	0.441	
GUERRERO	q_j	**0.87**	0.115	0.015	
	λ_j	0	0.482	0.895	
HIDALGO	q_j	**0.832**	0.167	0.001	
	λ_j	0	0.156	1.962	
JALISCO	q_j	**0.017**	0.933	0.051	
	λ_j	0	0.04	1.055	
ESTADO DE MEXICO	q_j	**0.001**	0.964	*0.035*	
	λ_j	0	0.228	2.1	
MICHOACAN	q_j	**0.793**	0.205	0.002	
	λ_j	0	0.135	1.771	
MORELOS	q_j	**0.633**	0.029	0.307	0.031
	λ_j	0	0.231	0.253	1.12
NAYARIT	q_j	**0.008**	0.989	*0.002*	
	λ_j	0	0.016	2.129	
NUEVO LEON	q_j	**0.056**	0.904	0.041	
	λ_j	0	0.034	0.534	
OAXACA	q_j	**0.256**	0.657	0.087	
	λ_j	0	0.02	0.595	

Table 6 continued

Year 2016		Population group			
		1	2	3	4
PUEBLA	q_j	**0.467**	0.521	0.012	
	λ_j	*0*	0.109	1.156	
QUERETARO	q_j	**0.463**	0.41	0.127	
	λ_j	*0*	0.025	0.259	
QUINTANA ROO	q_j	**0.101**	0.798	0.102	
	λ_j	*0*	0.029	0.464	
SAN LUIS POTOSI	q_j	**0.029**	0.957	0.014	
	λ_j	*0*	0.02	1.114	
SINALOA	q_j	**0.825**	0.174	***0.001***	
	λ_j	*0*	0.175	<u>6.803</u>	
SONORA	q_j	**0.868**	0.131	***0.001***	
	λ_j	*0*	0.313	<u>6.786</u>	
TABASCO	q_j	**0.035**	0.905	0.06	
	λ_j	*0*	0.052	0.448	
TAMAULIPAS	q_j	**0.147**	0.837	0.017	
	λ_j	*0*	0.017	0.648	
TLAXCALA	q_j	**0.846**	0.153	***0.001***	
	λ_j	*0*	0.361	<u>3.5</u>	
VERACRUZ	q_j	**0.194**	0.765	0.041	
	λ_j	*0*	0.023	0.439	
YUCATAN	q_j	**0.914**	0.006	0.08	
	λ_j	*0*	0.263	0.297	
ZACATECAS	q_j	**0.766**	0.125	0.109	
	λ_j	*0*	0.042	0.08	

Immune population size are given in bold

Zero crime rate $\lambda_1 = 0$ are given in italic

Chronic population size are given in bold italic

Chronic crime rate $\lambda_j \geq 2$ are underlined

coefficient of the k individual rates, even in the case in which they were estimated using a different model and under different assumptions. In this case, the *ECC* is constructed as

$$ECC(t) = \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} |\hat{\lambda}_i(t) - \hat{\lambda}_j(t)|}{2k \sum_{i=1}^{k} \hat{\lambda}_i(t)}, \tag{8}$$

which would also be a function of time. Thus, the *ECC*(*t*) gives a measure of the concentration of the rates under a different data frame and with different assumptions, but its interpretation is the same as the *RECC*, so values closer to 1 means a higher concentration of the rates and values closer to zero means a more homogeneous distribution. With the *ECC*(*t*), periods of the week in which crime is more evenly distributed or more concentrated in a few places, could be detected.

Table 7 Immune and chronic percentage of the population for the 32 states in Mexico between 2011 and 2016

State	Year (%)	2011	2012	2013	2014	2015	2016
AGUASCALIENTES	Immune	81.3	86.6	86.5	87.4	83.9	81
	Chronic	0	0	0	0	0.2	0
BAJA CALIFORNIA	Immune	66.2	60.7	66.8	78.6	63.5	52.1
	Chronic	0.2	0	0	0	0.6	0
BAJA CALIFORNIA SUR	Immune	85.5	90.2	91.5	80.4	87.1	64.7
	Chronic	0	0	0	0.1	0	0.1
CAMPECHE	Immune	87.6	89.7	90.4	77.7	95.6	86.9
	Chronic	0	0	0	0	0	0
COAHUILA	Immune	75.7	90.3	86	85.8	91.2	91.9
	Chronic	0	0.3	0	0	0	0.1
COLIMA	Immune	69.3	90.6	85.3	92.2	62.3	72.7
	Chronic	0	0	0	0.1	0	0
CHIAPAS	Immune	84	78.4	86.2	81.2	89.6	1.3
	Chronic	0	0	0.1	0	0	0
CHIHUAHUA	Immune	67.9	63.8	74.9	78	69.6	87.1
	Chronic	0	0	0.5	0	0	0.4
CIUDAD DE MEXICO	Immune	50.4	28.1	54.2	59.8	34.8	25.4
	Chronic	0	0.1	0.1	0	0	0
DURANGO	Immune	64	87.4	81.1	62	88.8	11
	Chronic	0	0	0.2	0	0.1	0
GUANAJUATO	Immune	77.5	68.3	73.4	75	37.9	9.7
	Chronic	0	0	0	0	0.8	0
GUERRERO	Immune	83.1	64.1	75.9	82.4	79.6	87
	Chronic	0.3	0	0	0	1.2	0
HIDALGO	Immune	78.5	80.5	67.5	78.7	3.3	83.2
	Chronic	0	0	0	0	0	0
JALISCO	Immune	83.5	75.7	81.1	78.1	11	1.7
	Chronic	0	0	0.8	0	0.1	0
ESTADO DE MEXICO	Immune	42.6	48.6	36.3	49.5	44.6	0.1
	Chronic	0	2.1	0.1	0.2	0.2	3.5
MICHOACAN	Immune	82.8	77	77.1	81.8	9.6	79.3
	Chronic	0	0	0	0.2	0	0
MORELOS	Immune	69.4	79.1	74.4	74.5	71.3	63.3
	Chronic	0.3	0.5	0.6	0	0.2	0
NAYARIT	Immune	86.9	87.3	93	86.7	20.1	0.8
	Chronic	0	0	0	0	0	0.2
NUEVO LEON	Immune	72.5	62.8	72.4	78.7	82	5.6
	Chronic	0	0	0	0	0	0
OAXACA	Immune	85.9	54.4	87	82.4	35.4	25.6
	Chronic	0	0	0.1	0.4	0	0
PUEBLA	Immune	76.3	78.1	73.2	72	1.4	46.7
	Chronic	0	0.2	0	0	0	0

Table 7 continued

State	Year (%)	2011	2012	2013	2014	2015	2016
QUERETARO	Immune	75.6	85.1	83.4	65.5	90.4	46.3
	Chronic	0	0	0	0	0.1	0
QUINTANA ROO	Immune	78.7	85.4	84.8	80.2	12.1	10.1
	Chronic	0	0.1	0.1	0.2	0.3	0
SAN LUIS POTOSI	Immune	76.5	81.6	81.3	87.4	0.3	2.9
	Chronic	0	0	0	0.2	0	0
SINALOA	Immune	71	78.8	76	83.3	85.4	82.5
	Chronic	0	0.3	0	0	0	0.1
SONORA	Immune	85.9	88.9	77.3	78.9	90.3	86.8
	Chronic	0	0	0	0	0.2	0
TABASCO	Immune	57.4	73.3	80.9	75.6	1.5	3.5
	Chronic	0.2	0	0	0	0.1	0
TAMAULIPAS	Immune	65.2	80.6	76.8	84.9	76.9	14.7
	Chronic	0	0	0	0.2	0.2	0
TLAXCALA	Immune	86.6	86.9	66.1	82.5	80	84.7
	Chronic	0	0	0.2	0	0	0
VERACRUZ	Immune	87.5	78.5	76.7	77.7	73.5	19.4
	Chronic	0	0	0	0	0.1	0
YUCATAN	Immune	83.5	84.6	78.1	90.3	92.2	91.4
	Chronic	0	0	0	0	0	0
ZACATECAS	Immune	88.5	60.1	87.6	78.4	93.8	76.6
	Chronic	0.1	0.3	0	0	0	0

Conclusions

Although this probabilistic approach to studying crime and victimisation rests on limited assumptions (i.e., that crime occurs independently, and that being victimised once does not affect the likelihood of being victimised again), by considering the number of crimes that each person suffers and modelling its random component, we can gain several pieces of valuable information about the crime problem. First, we can reject the notion that every person suffers the crime at the same rate, and accept that a homogenous distribution of crime in the population is far from being observed. Second, a mixture model provided a distribution of crime rates for an entire the population. The model has only a small number of parameters and so the distribution is useful for simulating the number of crimes that a population might suffer. Thirdly, the model allows us to detect whether a group exists which expects to suffer more than two crimes in a year, so they suffer from chronic victimisation and it also allows us to detect the existence of a group which is immune to crime. The existence of this immune group and the chronic group is based on the distribution of the crime rates and not directly on the number of crimes so, for example, a person who suffered no crime during a year is not necessarily immune to crime. Instead, they might have a positive rate of suffering crime but it was just lucky that they experienced no crime and a similar reasoning applies to a person who suffered two or more crimes, which could also be observed with a small rate.

The model presented here captures the general behaviour of the distribution of crimes. It is easily displayed through the Victimisation Profile, a novel way to graphically display the

distribution of the crime rates suffered by the population, which is comparable between different populations and over a different time period and so it also provides a versatile tool for dissemination purposes.

Finally, the model also allows us to construct a global measure for the concentration of crime (the *RECC*), which is a comparable metric between different time periods or different populations. This is an adequate tool for measuring the concentration of crime than others widely used in the field since it takes into account that crime is rare, highly concentrated and depends on random elements.

By using data from a victimisation survey in Mexico, and considering a different distribution of crime over its 32 states over 6 years, this study revealed that in most cases there is a considerably large population group which suffers a crime rate equal to zero so that they are statistically immune to crime. Also, some states have a small population which suffers chronic victimisation, meaning that they expect to suffer two or more crimes. These sorts of questions could also be answered using other data sources, for instance, the National Crime Victimization Survey (NCVS) (Bureau of Justice Statistics 2016), the Crime Survey for England and Wales (Office for National Statistics 2017) or perhaps by examining crime calls at addresses or street segments (Hipp and Kim 2016).

Using the specific case of the robbery of a person suffered in Mexico City between 2011 and 2012 allowed us to show a real scenario for which the traditional tools for measuring crime concentration failed to detect any major changes between one year and the next one, but the Victimisation Profile and the corresponding *RECC* show a drastic change in the observed pattern in which that crime is suffered.

A similar analysis could be conducted by considering the number of crimes committed by the population. With proper data we could also consider the rates at which the population commits crime and divide the population into the non-criminal group, the one-time offenders and the chronic offenders (Wolfgang 1983) which would be helpful to determine the effect of a preventive program.

This work goes towards the provision of adequate tools in the field in response to the need highlighted by the *law of crime concentration* (Weisburd 2015). Having adequate tools to measure the global concentration of crime, both from the crime that is suffered and the crime that is executed by the population, or whether is a spatial metric or based on the time of the crime, can help researchers and policy makers better understand the crime problem and what can be done to fix it.

Compliance with Ethical Standard

References

Andresen MA, Linning SJ, Malleson N (2016) Crime at places and spatial concentrations: Exploring the spatial stability of property crime in Vancouver BC, 2003–2013. J Quant Criminol. doi:10.1007/s10940-016-9295-8

Bernasco W, Steenbeek W (2016) More places than crimes: Implications for evaluating the law of crime concentration at place. J Quant Criminol. doi:10.1007/s10940-016-9324-7

Böhning D (1998) Zero-inflated Poisson models and C.A.MAN: a tutorial collection of evidence. Biometric J 40(7):833–843

Böhning D, Dietz E, Schlattmann P (1998) Recent developments in computer-assisted analysis of mixtures. Biometrics 525–536

Böhning D, Schlattmann P, Lindsay B (1992) Computer-assisted analysis of mixtures (C.A. MAN): Statistical algorithms. Biometrics 283–303

Bowers K, Johnson S, Pease K (2005) Victimisation and re-victimisation risk, housing type and area: a study of interactions. Crime Prevent Commun Safe Int J 7(1):7–17

Bowers KJ, Johnson SD (2003) Measuring the geographical displacement and diffusion of benefit effects of crime prevention activity. J Quant Criminol 19(3):275–301. doi:10.1023/A:1024909009240

Brame R, Nagin DS, Wasserman L (2006) Exploring some analytical characteristics of finite mixture models. J Quant Criminol 22(1):31–59. doi:10.1007/s10940-005-9001-8

Brantingham P, Brantingham P (2010) Criminality of place. Eur J Crim Pol Res 3(3):5–26

Bureau of Justice Statistics: National Crime Victimization Survey (NCVS). http://bjs.ojp.usdoj.gov/index.cfm?ty=dcdetail&iid=245. Accessed May 2016

Bushway SD, Tahamont S (2016) Modeling long-term criminal careers what happened to the variability? J Res Crime Delinq 53(3):372–391

Chen SX, Liu JS (1997) Statistical applications of the Poisson–Binomial and conditional Bernoulli distributions. Stat Sin 7(1):875–892

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 588–608

Dorfman R (1979) A formula for the Gini coefficient. Rev Econ Stat 146–149 (1979)

Farrell G (2015) Crime concentration theory. Crime Prevent Commun Safe 17(4):233–248

Farrell G, Pease K (1993) Once bitten, twice bitten: repeat victimization and its implications for crime prevention. Police Res Gr Crime Prevent Unit Paper 46(1):1–32

Farrington DP, Jolliffe D, Loeber R, Stouthamer-Loeber M, Kalb LM (2001) The concentration of offenders in families, and family criminality in the prediction of boys' delinquency. J Adolescen 24(5):579–596

Ferguson T (1952) The young delinquent in his social setting. A Glasgow study. The Young Delinquent in his Social Setting, A Glasgow Study

Fox JA, Tracy PE (1988) A measure of skewness in offense distributions. J Quant Criminol 4(3):259–274. doi:10.1007/BF01072453

Freeman S (1996) The spatial concentration of crime. J Urban Econ 40(1):216–231

from INEGI OW (2016) Victimisation survey. http://www3.inegi.org.mx/sistemas/microdatos/encuestas.aspx?c=34517&s=est. Accessed Oct 2016

Glaeser EL, Sacerdote B (1996) Why is there more crime in cities?. Technical report, National Bureau of Economic Research

Greenland S (2004) Interval estimation by simulation as an alternative to and extension of confidence intervals. Int J Epidemiol 33(6):1389–1397

Grove L, Farrell G, Farrington D, Johnson S (2012) Preventing repeat victimization: a systematic review, 1st edn. The Swedish National Council for Crime Prevention, Stockholm

Guerette RT, Bowers KJ (2009) Assessing the extent of crime displacement and diffusion of benefits: a review of situational crime prevention evaluations. Criminology 47(4):1331–1368

Hindelang MJ, Gottfredson MR, Garofalo J (1978) Victims of Personal crime: an empirical foundation for a theory of personal victimization. Ballinger, Cambridge

Hipp JR, Kim YA (2016) Measuring crime concentration across cities of varying sizes: Complications based on the spatial and temporal scale employed. J Quant Criminol. doi:10.1007/s10940-016-9328-3

Hope T, Norris PA (2013) Heterogeneity in the frequency distribution of crime victimization. J Quant Criminol 29(4):543–578. doi:10.1007/s10940-012-9190-x

Hope T, Trickett A (2008) The distribution of crime victimisation in the population. Int Rev Victimol 15(1):37–58

Johnson S, Summers L, Pease K (2009) Offender as forager? a direct test of the boost account of victimization. J Quant Criminol 25(2):181–200

Johnson S (2010) A brief history of the analysis of crime concentration. Eur J Appl Math 21(1):349–370

Johnson S, Guerette R, Bowers K (2014) Crime displacement: what we know, what we dont know, and what it means for crime reduction. J Exp Criminol 10(1):549–571

Laycock G, Farrell G (2003) Repeat victimization: lessons for implementing problem-oriented policing. Crime Prevent Stud 15(1):213–237

Levin A, Rosenfeld R, Deckard M (2016) The law of crime concentration: an application and recommendations for future research. J Quant Criminol. doi:10.1007/s10940-016-9332-7

Maltz MD (1996) From Poisson to the present: applying operations research to problems of crime and justice. J Quant Criminol 12(1):3–61. doi:10.1007/BF02354470

Marsh C, Elliott J (2008) Exploring data: an introduction to data analysis for social scientists. Polity, United Kingdom

McLachlan G, Peel D (2004) Finite mixture models. John Wiley & Sons, Canada

Mohler GO, Short MB, Brantingham PJ, Schoenberg FP, Tita GE (2012) Self-exciting point process modeling of crime. J Am Stat Assoc 106:100–108

Mooney CZ (1997) Monte Carlo simulation. First edition edn. Sage Publication Inc, California

Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. Criminology 31(3):327–362

Office for National Statistics: Crime Survey for England and Wales, 1996-2016 (CSEW). http://doi.org/10.5255/UKDA-SN-7280-6. Accessed May 2017

Osborn DR, Tseloni A (1998) The distribution of household property crimes. J Quant Criminol 14(3):307–330

Osgood DW (2000) Poisson-based regression analysis of aggregate crime rates. J Quant Criminol 16(1):21–43

Park SM, Eck JE (2013) Understanding the random effect on victimization distributions: a statistical analysis of random repeat victimizations. Vict Offend 8(4):399–415. doi:10.1080/15564886.2013.814612

Pease K (1998) Repeat victimisation: taking stock. Home Office Police Research Group London, United Kingdom

Pease K, Ignatans D (2016) The global crime drop and changes in the distribution of victimisation. Crime Sci 5(1):1–6. doi:10.1186/s40163-016-0059-4

Prieto Curiel R, Bishop SR (2016) A measure of the concentration of rare events. Scientific Reports 6

R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2014). R Foundation for Statistical Computing. http://www.R-project.org/

Rosser G, Davies TP, Bowers KJ, Johnson SD, Cheng T (2016) Predictive crime mapping: arbitrary grids or street networks? J Quant Criminol 1–26 (2016)

Schlattmann, P., Hoehne, J., Verba, M.: C.A.MAN.: Finite Mixture Models and Meta-Analysis Tools - Based on C.A.MAN. (2015). R package version 0.73. http://CRAN.R-project.org/package=CAMAN

Schlattmann P (2005) On bootstrapping the number of components in finite mixtures of Poisson distributions. Stat Comput 15(3):179–188

Short MB, D'Orsogna MR, Pasour VB, Tita GE, Brantingham PJ, Bertozzi A, Chayes LB (2008) A statistical model of criminal behavior. Math Models Methods Appl Sci 18(supp01):1249–1267

Short MB, D'Orsogna MR, Brantingham PJ, Tita GE (2009) Measuring and modelling repeat and near-repeat burglary effects. J Quant Criminol 25(3):325–339

Short MB, D'Orsogna MR, Brantingham PJ, Tita GE (2009) Measuring and modeling repeat and near-repeat burglary effects. J Quant Criminol 25(3):325–339. doi:10.1007/s10940-009-9068-8

Sparks RF (1981) Multiple victimization: evidence, theory, and future research. J Crim Law Criminol 72(2):762–778

Stark R (1987) Deviant places: a theory of the ecology of crime. Criminology 25(4):893–910

Tseloni A (2000) Personal criminal victimization in the United States: fixed and random effects of individual and household characteristics. J Quant Criminol 16(4):415–442

Tseloni A, Wittebrood K, Farrell G, Pease K (2004) Burglary victimization in England and Wales, the United States and the Netherlands a cross-national comparative test of routine activities and lifestyle theories. Br J Criminol 44(1):66–91

Tseloni A, Pease K (2003) Repeat personal victimization. Boosts or flags? Br J Criminol 43(1):196–212

Tseloni A, Pease K (2004) Repeat personal victimization random effects, event dependence and unexplained heterogeneity. Br J Criminol 44(6):931–945

Tseloni A, Pease K (2005) Population inequality: the case of repeat crime victimization. Int Rev Victimol 12(1):75–90

Weisburd D (2015) The law of crime concentration and the criminology of place. Criminology 53(2):133–157. doi:10.1111/1745-9125.12070

Wolfgang ME (1983) Delinquency in two birth cohorts. In: Prospective studies of crime and delinquency, pp. 7–16. Springer, United States (1983)

Wolfgang ME, Figlio RM, Sellin T (1987) Delinquency in a birth cohort. University of Chicago Press, Chicago