Measuring and Modeling Repeat and Near-Repeat Burglary Effects

M. B. Short · M. R. D'Orsogna · P. J. Brantingham · G. E. Tita

Abstract We develop a mathematical framework aimed at analyzing repeat and near-repeat effects in crime data. Parsing burglary data from Long Beach, CA according to different counting methods, we determine the probability distribution functions for the time interval τ between repeat offenses. We then compare these observed distributions to theoretically derived distributions in which the repeat effects are due solely to persistent risk heterogeneity. We find that risk heterogeneity alone cannot explain the observed distributions, while a form of event dependence (boosts) can. Using this information, we model repeat victimization as a series of random events, the likelihood of which changes each time an offense occurs. We are able to estimate typical time scales for repeat burglary events in Long Beach by fitting our data to this model. Computer simulations of this model using these observed parameters agree with the empirical data.

Keywords Repeat victimization · Burglary · Event dependence

Introduction

Repeat victimization has recently emerged as a central focus in criminology. Research has demonstrated that individuals who have been victims of personal or property crimes are more likely to be victimized again (Farrell and Pease 2001). In the case of residential burglary, which we shall focus on in this paper, repeat victimization is described in terms

M. B. Short (✉)
Department of Mathematics, UCLA, Box 951555, Los Angeles, CA 90095-1555, USA
e-mail: mbshort@math.ucla.edu

M. R. D'Orsogna
Department of Mathematics, CSUN, Los Angeles, CA 91330, USA

P. J. Brantingham
Department of Anthropology, UCLA, Los Angeles, CA 90095-1555, USA

G. E. Tita
Department of Criminology, Law & Society, UC-Irvine, Irvine, CA 92697, USA

of exact-repeat and near-repeat events (Johnson et al. 2007). Exact-repeat events are defined as consecutive burglaries occurring at the same location, separated by a time interval of any duration; near-repeat burglary events are instead classified as taking place within a set spatial neighborhood of a focal burglary point.

Repeat burglary victimization may be due to a variety of reasons, including persistent spatial heterogeneity of risk and/or event dependence tied to the specific activities of burglars (Tseloni and Pease 2003, 2004; Johnson 2008). When considering a complex urban environment, risk heterogeneity implies that some houses (or neighborhoods) are at higher risk than others, and that this difference in risk persists throughout time. Some houses may be at higher risk because they are physically soft targets (e.g., easily forced doors or windows) or because the routine activities of inhabitants leave them much less secure than other homes. By contrast, event dependence suggests that some aspect of the burglar's previous experience victimizing the house increases their preference to return. For example, a burglar may discover an abundance of items that could be targeted in a subsequent burglary, or they may simply prefer to return to a location where they know that their entry methods are guaranteed to work again, amongst other possible reasons (Farrell et al. 1995). In addition, it has been suggested that this elevated risk may spread to neighboring homes as well (Johnson et al. 1997; Townsley et al. 2003; Sagovsky and Johnson 2007), especially in areas where nearby homes are similar in layout and type of inhabitant. Notice, however, that in the case of event dependence, burglary risk is not persistent throughout time, but may change as the burglar's preferences, skills, and exposure to other opportunities change (Farrell et al. 1995).

The concept of such biased repeat burglary carries strong implications for the dynamics of crime pattern formation and for the development of prevention and resource allocation strategies (Bowers et al. 1998; Farrell and Pease 1993, 2001). Models based on the event dependence hypothesis show that individual crimes can establish positive feedbacks and nucleate into crime hotspots (Eck et al. 2005; Johnson and Bowers 2004; Short et al. 2008). Effective control strategies would pinpoint these pivotal sites, using past crimes as indicators of future ones, breaking the feedback loops and thus surgically halting the further spread of crime (Farrell et al. 2007). Therefore, simple and accurate methods of testing for the presence of event dependent repeat effects are of great importance.

Repeat burglary effects are often observed via the distribution of victimization order within a population of homes, where the victimization order is here defined as the number of times a home is burgled within some fixed temporal window. This distribution is typically inconsistent with a Poisson distribution, which is what would be expected if all homes had the same, persistent risk of burglary. In order to see whether event dependence may be responsible for this inconsistency, one often focuses on the distribution of time intervals τ between successive events that occurred at the same location, a procedure that has been performed using burglary data from a variety of cities worldwide (Johnson et al. 2007). In general, it is observed that the distribution of time intervals between burglary events is a rapidly decaying function, with short time intervals much more likely to occur than longer ones. This observation has been taken as evidence for the existence of event dependence, and that a house will exhibit an increased risk of being burgled after being victimized once. However, there has not yet been a rigorous discussion as to why exactly these decaying time interval distributions support this hypothesis. In fact, as we will show in this paper, this observation alone *does not necessarily* support the event dependence hypothesis at all, and the method of counting the time intervals is of critical importance when interpreting the distribution of τ.

Throughout the remainder of this paper, we will be performing analyses on a dataset which includes the geographic location and day on which each reported residential burglary for the years 2000–2005 occurred in Long Beach, CA. We consider only those burglaries that occurred at single family homes, since we do not possess data that is detailed enough to pinpoint specific units within multi-family housing. Here, the term "single family homes" refers to stand-alone housing units (i.e., detached houses) with unique physical addresses as opposed to "multi-family housing", which could be an apartment complex or condominium building where many separate units share the same physical address that belongs to the entire structure. In our analyses, we have ignored the influence of seasonality (Farrell and Pease 1994), specifically because the climate of Long Beach minimizes such fluctuations. However, all of the results and formulas can be modified to include seasonal effects in a straightforward manner. The dataset contains 9,042 events, and the distribution of victimization order across the homes is: 7,002 order one, 819 order two, 98 order three, 19 order four, 5 order five, and 1 order seven. According to the 2000 US census, there are between 70,000 and 80,000 occupied single family homes in Long Beach. Using this fact, and the house order distribution, we find that a simple Poisson distribution does not fit our Long Beach data well, indicating that something is indeed causing repeat victimization there.

The goal of this paper is to first describe a model in which this repeat victimization is due solely to risk heterogeneity; we refer to this as the random event hypothesis (REH) (see also Nelson 1980). We then show, following from some very reasonable assumptions, that the distribution of time intervals τ for exact-repeats in the REH is that of a sum of decaying exponentials, and that we should observe just this when using a moving-window counting method on our data (which will be describe later). Using our burglary data, we illustrate that the observed distribution of τ is completely compatible with that predicted by the REH, with the parameters of the fit interpreted as measures of risk heterogeneity. This compatibility, however, is not sufficient to prove the validity of the REH since other possible mechanisms of burglary dynamics might be equally compatible with the observed results. In fact, the parameters of the fit lead to a predicted distribution of home orders that is wildly different from that observed, indicating that the REH is insufficient to explain exact-repeat effects in our data. We then introduce a different method of counting exact-repeat time intervals that allows us to unequivocally differentiate data sets generated via the REH from those in which burglary events are in fact related via event dependence, using only the time interval distribution. Applying this novel analysis method to our data set, we find that there is, in fact, event dependence in Long Beach. We present a simple mathematical model with a straightforward criminological interpretation that explains the observations under both counting methods, and which reproduces both the time interval distributions and the home order distributions well in simulation. Finally, we extend some of these results to the measurement of event dependence in the near-repeat effect, finding that it too is present in our data.

The Random Event Hypothesis

The simplest assumption to make about burglary events at house i is that they occur entirely at random, defining a stochastic process where each burglary event is independent of all others. We will call this model the REH. In addition, it is obvious that two burglary events at house i cannot occur simultaneously, as they would then simply be thought of as one event. Mathematically, such a phenomenon can be modeled as a Poisson process

characterized by a rate parameter λ, representing the expected number of burglary events per unit time. The characteristics of Poisson processes are long-established (Feller 1968), but we present now a brief summary of the main results which are relevant to our analysis. For a Poisson process with rate parameter λ, the probability that one burglary occurs within a time interval t to $t + \delta t$ is given by

$$P_1(\delta t) = e^{-\lambda \delta t} \lambda \delta t \tag{1}$$

The probability that k burglaries occur is given by the general Poisson distribution

$$P_k(\delta t) = \frac{e^{-\lambda \delta t} (\lambda \delta t)^k}{k!} \tag{2}$$

The probability that no events occur within a time interval δt, then, is given by

$$P_0(\delta t) = e^{-\lambda \delta t}, \tag{3}$$

a monotonically decreasing function of time.

Consider now the time T_1 until the first burglary occurs at house i, as measured from a reference point with $t = 0$. In this case, T_1 will be greater than a given time τ only if there have been no events within the time interval from 0 to τ. Hence,

$$P(T_1 > \tau) = P_0(\tau) = e^{-\lambda \tau}. \tag{4}$$

Extending this result, we see that the probability that the first event occurs between times τ and $\tau + \delta t$ is

$$P(\tau < T_1 < \tau + \delta t) = P(T_1 > \tau) - P(T_1 > \tau + \delta t) = e^{-\lambda \tau}\left(1 - e^{-\lambda \delta t}\right). \tag{5}$$

If we divide this result by δt and take the limit as $\delta t \to 0$, we arrive at the standard Poisson process probability density function for the time interval τ between events:

$$p(\tau) = \lambda e^{-\lambda \tau}. \tag{6}$$

Therefore, if the REH is correct, the distribution of time intervals between exact-repeat events at a given home with rate constant λ should follow an exponential decay of the type shown in Eq. 6. Note that this distribution, which displays a much higher number of events at short time intervals than long, was derived without introducing any notion of correlation between burglary events. In fact, this distribution will only hold if the events are statistically independent, a notion that is completely contrary to the typical assumptions of the event dependence hypothesis.

We do not, however, expect every home within a city to display the same burglary rate λ, as it is well known that crime rates may vary spatially (Ratcliffe and McCullagh 1999) and we have already shown that our Long Beach data does not conform to a simple Poisson distribution. We therefore allow our homes to be divided into N groups, each of which is characterized by a particular burglary rate λ_i that is persistent in time for that group. If the fraction of homes exhibiting rate λ_i is defined to be w_i, then the composite distribution of waiting times should be given by

$$p(\tau) = \sum_{i=1}^{N} w_i \lambda_i e^{-\lambda_i \tau}, \tag{7}$$

which is just a weighted sum of the individual distributions for each group. Equation 7 might also be read as the mean waiting time for all houses, since the w_i sum to 1.

Equation 7 is therefore a mathematical representation of pure risk heterogeneity, in terms of the time intervals between exact-repeat events modeled via a compound Poisson process.

Testing the REH

The Moving-Window Method

In order to test the distribution predicted by the REH, one must first develop a proper counting scheme for the time intervals τ between exact-repeat events. Ideally, one would watch each burgled house within the city of interest until it is burgled again, and simply mark the time to repeat. However, this is clearly infeasible, as many homes will not be burgled again during a reasonable observation period. In fact, for our Long Beach data set, out of the 7,944 unique locations burgled, only 942 of them were burgled more than once. If we were to only use the time intervals from these relatively few locations, we would likely introduce a bias into our count because we would be systematically discarding many time intervals which were at longer timescales and were, therefore, never observed.

To count properly, then, we use a method that we will call here the *moving-window method*. The basic idea behind this method is to first choose a time window of interest, τ_{max}, and then to observe after each burglary event whether or not another event occurs at that same location within this time window; let us use as an example a τ_{max} of 727 days for our Long Beach data set. If an event does indeed occur, the time interval τ between the initial and secondary event is noted. Of course, any event which occurs within the last τ_{max} days for which we have data cannot be subsequently watched over the full τ_{max} window, as some of the window would clearly lie outside of the dates for which we have data. Therefore, we do not perform our observation following these events. We call this final τ_{max} period within our data the "buffer interval", which corresponds to the years 2004 and 2005 in our example. The final output of the count consists of the number of events N_o for which an observation was performed (this is equal to the total number of events in our dataset minus the number of events that occur within the buffer interval) and a list of time intervals observed. Note that the number of time intervals recorded will in general not be N_o, since not every home that is observed will be subject to another burglary within our τ_{max} window, as discussed above. Finally, we make a histogram of the observed τ, dividing the frequency for each histogram bin by N_o to arrive at a probability distribution that we can directly compare to Eq. 7. It is in this way that the homes burgled only once affect our count—they contribute no time intervals, but they do increase N_o and thereby influence the probabilities.

The results of such a moving-window count can be seen in Fig. 1, using our Long Beach data. Here we have, as in our example above, chosen a τ_{max} of 727 days, making the buffer interval roughly the years 2004 and 2005. The choice of 727 days is arbitrary, but was used because it is evenly divisible by our desired histogram bin width of 14 days. By adding together the probabilities for each histogram bin, we find that 10.8% of the events in our data set are followed by an exact repeat within 727 days.

Along with the observed τ distribution, we have plotted in Fig. 1 (the solid line) a curve of the type shown in Eq. 7 with parameters chosen to give the best fit to our data. Using $N = 3$, we find the best fit to be $w_1 = 0.915$, $\lambda_1 = 5.32 \times 10^{-5}$, $w_2 = 0.066$, $\lambda_2 = 2.45 \times 10^{-3}$, $w_3 = 0.019$, and $\lambda_3 = 8.41 \times 10^{-2}$. The choice of $N = 3$ was made simply because this was the smallest value for which a good fit of the curve to our data
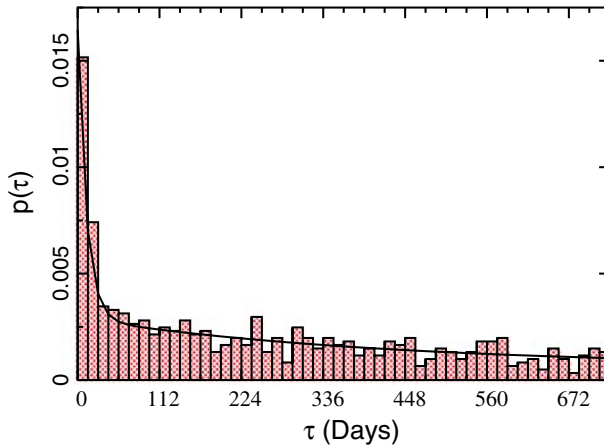
Fig. 1 Probability distribution $p(\tau)$ as measured from our Long Beach, CA data set using the moving-window method (*histogram bars*). Data has been binned by two week intervals, and the bars represent the total probability for each bin. Overall, 10.8% of the events in our data set are followed by an exact repeat within 727 days. The *black curve* represents the best fit of the form given in Eq. 7, with the parameters given in the text

could be found; both the $N = 1$ and $N = 2$ curves deviate too substantially from our data. For this choice, though, the REH curve fits our data rather well.

On the basis of this analysis, one might conclude that there is no event dependence effect in our data, since our observations of the distribution of τ are completely consistent with the REH, with the spatial heterogeneity of risk described by the w_i and λ_i used in the fit. Specifically, 91.5% of the houses have a time-invariant risk of being burgled once every 51 years, 6.6% of the houses have a higher time-invariant risk of being burgled once every 1.1 years, and 1.9% of the houses have an even higher time-invariant risk of being burgled once every 11.8 days (see Eqs. 11 and 12 below). However, one can perform further analyses that cast doubt upon this conclusion. If the distribution of risk and the rates for each were indeed as given by our w_i and λ_i above, the distribution of house order would be very different than what is observed: using 70,000 for the total number of single family homes in Long Beach, these parameters predict around 6,800 order one, 700 order two, 600 order three, 700 order four, 800 order five, 700 order six, 500 order seven, and many hundreds of homes with order greater than seven. It is clear from this subsequent analysis that spatial risk heterogeneity is not an accurate description of what is happening in Long Beach.

The Fixed-Window Method

Although the moving-window counting method is a valid approach, its corresponding null hypothesis curve as derived through the REH contains a large number of parameters, making it difficult to compare to observations in a meaningful way. In addition, as shown above, even if parameters can be chosen such that the REH curve fits the data very well, further calculations are needed to interpret these results. In order to more easily determine the validity of the REH, we develop a counting method for which the null hypothesis curve is completely parameter free and that can by itself definitively confirm or deny the REH; we term this the *fixed-window method*. We first remind the reader that each home

appearing within our data set can be classified by the number of times it was burgled in total over the D days of data available; we refer to this as the order of the house. The probability of any given home with burglary rate λ_i being of order k is given by Eq. 2, replacing δt with D. Note, however, that Eq. 2 is independent of the particular times at which the home was burgled, so long as there were a total of k events. This means, for example, that for order one homes, each of the D days is equally likely to be the day on which the one event occurred, assuming that λ_i is persistent in time (i.e., seasonality is ignored and there is no event dependence). Similarly, for order two homes, each possible *pair* of days that can be made from our D day interval is equally likely to be the observed pair.

Suppose that we isolate all order two homes from our data set and ask how the time intervals τ between the two events at each of these houses ought to be distributed, assuming validity of the REH. Although all pairs of days are equally likely to occur within our trimmed data set, all time intervals are not. For example, with a fixed window of 1 year ($D = 365$), there are very many pairs of days that will lead to a τ of just 1 day (Jan 1–Jan 2, Jan 2–Jan 3, etc.), while there is only one pair that exhibits the maximum time interval of 364 days (Jan 1–Dec 31). More generally, the number of pairs that will exhibit a time interval of τ is given by

$$N_{\text{pairs}}(\tau) = D - \tau, \tag{8}$$

one for each day that has at least τ days following it in the time window. We normalize these counts to one, so that the probability distribution of time intervals for order two homes is

$$p_2(\tau) = \frac{2(D - \tau)}{D(D + 1)}. \tag{9}$$

This particular distribution is specifically for order two houses. However, it can be shown that the probability distribution for houses of any order k is given by

$$p_k(\tau) = \frac{k}{D + k - 1} \prod_{l=0}^{k-2} \frac{D - \tau + l}{D + l}, \tag{10}$$

using similar arguments.

There are two important points to make concerning the fixed-window method of counting. First, as seen in the moving-window method, if the REH is true and houses do not experience any increased risk of burglary after an initial event (i.e., there is no event dependence), the distribution of time intervals that we count will still be heavily weighted toward the short end of the spectrum, and will disappear at the long end (Fig. 2). Therefore, we must be mindful of this combinatorial effect when interpreting the results of a fixed-window counting procedure. Second, and more positively, we see that the fixed-window method is an excellent way to test the REH, and thus the validity of risk heterogeneity for explaining repeat burglaries, because the expected probability distribution given by Eq. 10 has no free parameters. The fixed-window method therefore represents a simple and unambiguous way to determine whether or not the REH is in fact true, or if event dependence must be invoked.

To illustrate the usefulness of the fixed-window counting method in testing the REH, we present here the results of such a count, using our Long Beach data (Fig. 3). To perform this count, we first break up our data into six non-overlapping 364 day sets (we use $D = 364$ because it is a multiple of 14). For each set, we isolate the order two homes
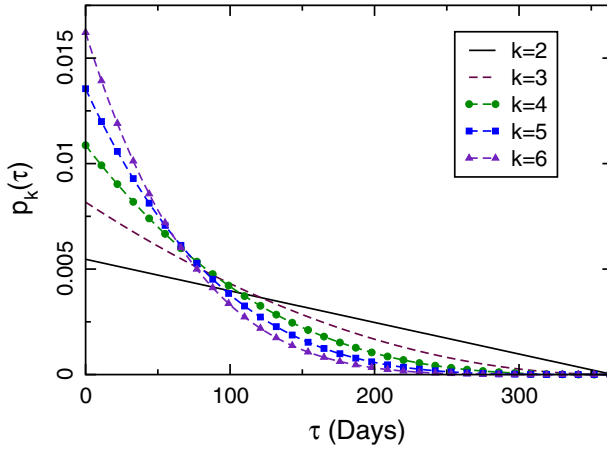
Fig. 2 Probability distributions $p_k(\tau)$ for various values of k, using $D = 365$ days. Note that they are all biased toward shorter time intervals, and that this bias becomes more pronounced as k increases

within that set, then count the time interval between the two burglaries for each of these homes. The resulting six τ lists are then combined into one, a histogram is made of this master list, and the histogram bins are each divided by the total number of order two homes used in the count to convert to a probability that will add up to 1. When this histogram is compared with the null hypothesis curve as derived above in Eq. 9, we see that the disagreement is very substantial, with many more events occurring at short time intervals than predicted by the REH, and fewer, therefore, at long time intervals. Thus, the REH is disproven in our data without the need for any further analysis, and it is clear that event dependence must be responsible for at least some of the repeat victimization effect.
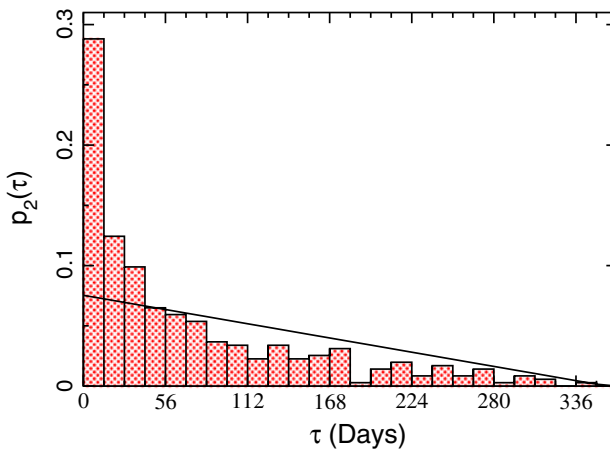

Fig. 3 Probability distributions $p_2(\tau)$ as measured from our Long Beach, CA data set using the fixed-window method (*histogram bars*) and as calculated via Eq. 9 using $D = 364$ days (*black line*). Note that the measured probability is much higher than expected for short time intervals, and that it falls below the expected values at long time intervals. This indicates that the REH is not consistent with our data set

333

Modeling Event Dependence

One may wonder why the time interval data seemed to agree with the REH when using the moving-window counting method, but clearly contradicts it when using the fixed-window method. The answer to this question is that the moving-window curve defined by Eq. 7 is not unique to the REH. An alternate model, based on the idea of event dependence, that would present the same moving-window curve can be described in the following way. Suppose that there exist some number N of different risk states that any given home can exhibit. Associated with each of these states i is a burglary rate λ_i and a weight w_i; these w_i form a probability distribution for the rates λ_i in a sense that shall be described shortly. Now, imagine that we have a large number of homes, each of which is initially assigned one of the rates λ_i, with the probability distribution for this assignment being the weights w_i. For each home, then, time goes by until that home is subject to a burglary event; the time elapsed will clearly depend upon which of the rates λ_i was initially assigned to the home. As the critical step, after this (and each subsequent) burglary event, the home is again assigned a rate λ_i according to the probability distribution w_i, thus leading to the possibility of a change in the Poisson rate at that home. This process repeats indefinitely, with a state assigned via the w_i distribution after each burglary event. Data generated via this process will lead to a moving-window curve as given in Eq. 7, but will not yield the fixed-window curve derived in Eq. 10. This is because the Poisson rate for any given house is no longer, in general, persistent over all time (thus altering the null hypothesis curve for the fixed-window method), though it is persistent between any two events at a home (giving it the same null hypothesis curve as the REH for moving-window counts). Hence, this event dependence based model is a good candidate for the process that is in reality generating our empirically observed burglary events.

The preceding model also has a simple criminological interpretation. Suppose $N = 2$, meaning that a home can exhibit two states, which we can refer to as the baseline state (that state with the smaller λ_1) and the excited state (that state with the larger λ_2). After any burglary event, the house will, with probability w_1, be assigned the baseline state, meaning that the offender who just burgled the home has no specific intention of returning there in the future (though he or another burglar of course may return simply by chance). Alternatively, the house will, with probability w_2, be assigned the excited state, meaning that the offender, possibly for reasons such as those described in the introduction, plans on returning to this specific home again in a relatively short period of time. Any number of possible states N greater than two simply correspond to more possible levels of excitement for returning offenders, who may range from only slightly interested to extremely interested in returning. Indeed, the good fit between Eq. 6 and our Long Beach data for $N = 3$ (see Fig. 2) suggests that in Long Beach, a house might be assigned one of three possible states after an event: the baseline state or one of two excited states.

The parameters of this model also have clear physical meanings. The various rates λ_i set timescales T_i over which repeat burglaries are most likely to occur, such that

$$T_i = \frac{1}{\lambda_i}. \tag{11}$$

For our Long Beach data (with $N = 3$), these timescales are approximately 50 years for the baseline state, 1 year for the first excited state, and 12 days for the second excited state. The weighting values w_i represent the probabilities of adopting any of these states after a burglary event, with the baseline rate being by far the most likely. In conjunction, T_i and w_i

predict steady state values s_i for the average fraction of homes exhibiting each state at a given instant in time. This is done according to the relation

$$s_i = w_i T_i \left/ \sum_{i=1}^{N} w_i T_i. \right.$$ (12)

For Long Beach, this predicts that at any given time, 99.842% of homes are in the baseline state, 0.156% are in the first excited state, and the remaining 0.002% are in the second excited state.

In order to verify how well a model such as this reproduces our Long Beach exact-repeat burglary data, a simple simulation was created using our best fit weights and burglary rates as found via the moving-window counting method above. The output of the simulation was then analyzed using both the moving-window method (Fig. 4a) and the fixed-window method (Fig. 4b). There is good agreement between our simulation based graphs and those constructed using the Long Beach data, lending credence to our model. In addition, one can use this model and its associated parameters to predict the distribution of home order. Again using a total of 70,000 homes and a time interval of 6 years, our model predicts that there should be around 6,800 order one, 870 order two, 100 order three, 10 order four, and 1 order five homes (leaving 62,200 order zero). This distribution is in good agreement with the data.

Near-Repeats

A near-repeat event occurs whenever two "nearby" houses are burgled within some period of time. Like exact-repeats, we can measure the time interval that lies between each event in a near-repeat pair, but in this case we must also make note of the physical separation of the two homes. This procedure allows us to examine separately the time interval histogram for near-repeat pairs that lie at varying physical distances from each other. It has been noted in previous studies that those near-repeat events that are relatively close in space
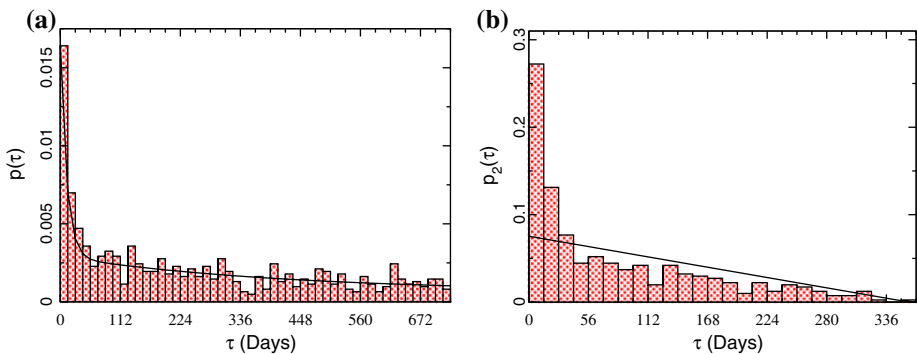

Fig. 4 Simulation output analyzed via the moving-window method (**a**) and the fixed-window method (**b**). The output from one run of the simulation is shown as the *histogram bars*, and the curves correspond to the expected results under the REH. As expected, using the simulation parameters that best fit our burglary data leads to simulation results that are similar to the graphs constructed from that data (Figs. 2 and 3). From the moving-window method, we see that in this simulation run 11.3% of simulated events were subject to an exact repeat within 724 days, as compared to 10.8% in the Long Beach data

tend to occur more closely in time as well, like exact-repeats, whereas those that are far apart seem to exhibit no temporal correlation. These previous studies use Monte Carlo algorithms to find the likelihood of the observed patterns happening if there were no correlation between the spatial and temporal distributions (Johnson et al. 2007; Ratcliffe and Rengert 2008), determining that this is highly unlikely. In this section, we instead test explicitly for near-repeat event dependence by extending our finite-window counting method used above for exact-repeats to the case of near-repeat events in our Long Beach data.

The first step in our derivation is actually a fact that we notice from our data, which is supported by our model. We observe, when each year of our dataset is examined separately, that order one homes exhibit approximately equal probability of being burgled on all of the 365 days available. This is illustrated in Fig. 5, where we see the flat nature of the distribution of order one burglary days. This observation is consistent with the theoretical curve predicted from our exact-repeat model (also plotted in Fig. 5) which fits the data well when using our previously estimated parameters. The observed pattern arises from the fact that about 91% of homes that are burgled only once in $D = 365$ days will enter into the time window with the lowest burglary rate, λ_1, and still exhibit that rate after the event. These houses should all have a completely flat line in this count, because their rates are not changing and so each day is equally likely. The remaining 9% of homes consists almost entirely of those that either entered the year with the middle rate, λ_2, and then switched to λ_1 or those that did the opposite. These homes explain the slight increases in the theoretical curve near the beginning and end of the year. In addition, Fig. 5 shows the lack of seasonality in our dataset.

To test for the presence of near-repeat event dependence, we first isolate in our data all order one homes. We then perform a fixed-window count on these events in a pair-wise fashion, measuring both the temporal separation and physical distance between the burglaries comprising each possible pair of events. Note that this is essentially the same procedure as was performed for the exact-repeats earlier, except that in that case the
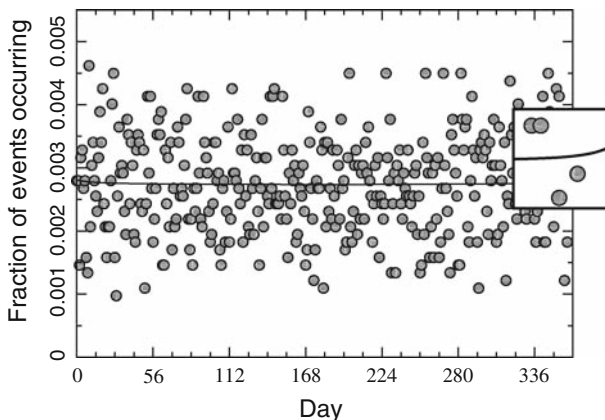

Fig. 5 Fraction of order one events occurring per day in Long Beach, counting each year separately (*black dots*). The curve predicted by our model is also shown (*black line*). The distribution is essentially flat, with only slight increases at the beginning and end of the interval (shown magnified on the right). Most homes that are burgled only once will exhibit a constant rate of λ_1, leading to overall flatness. A small number of order one homes will transition from λ_2 to λ_1, or vice versa, which happens most often near the beginning and end of the interval; this explains the deviations from flatness in these regions
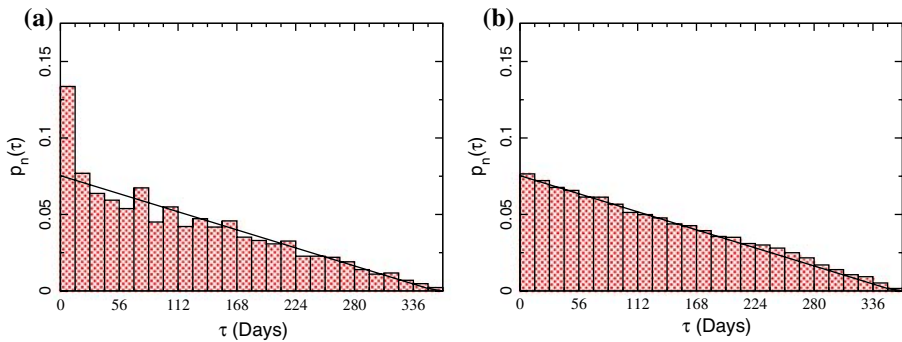
Fig. 6 Near-repeat burglary time intervals τ (*histogram bars*) as measured via the fixed-window method for homes within 100 m of each other (**a**) and within 3.9 and 4.0 km (**b**), both shown against the order two REH curve (*black line*). For the closer events, there is a clear deviation from the null hypothesis curve derived through the REH only at the shortest of time intervals. This deviation is not nearly as strong as that noted for exact-repeats, nor does it extend as long in time. For the further apart events, no particular deviation is noted, indicating that near-repeat event dependence does not extend to homes that are 4 km apart in Long Beach

physical distances were all zero and we used order two homes rather than order one. The fact that order one homes are approximately equally likely to be burgled on any day of our fixed interval means that the time intervals for near-repeat events should be distributed exactly as in Eq. 10 if no correlation between the two burglaries making up a pair exists (i.e., if there is no event dependence). This is because, since each of the homes is equally likely to be burgled on any given day, each of the possible pairs of days making up a near-repeat event ought to be equally likely as well, which is the condition that leads directly to Eq. 10.[1]

The results of such a count are presented in Fig. 6, where we show the time interval histograms for the distance bands of (0–100] meters and (3.9–4.0] kilometers. For the closer band, we clearly see that the measured number of near-repeats at the shortest time intervals is greater than that predicted by the null hypothesis, many of those in fact occurring at a time separation of zero days. At longer time intervals, though, the null hypothesis curve fits rather well. This is unlike the exact-repeat observations, for which the deviation was much greater and was spread over the entire curve. For the further distance band, no effect seems to be present at all, as the data and null hypothesis curves are in very good agreement. At distance bands between these two extremes, we observe an ever decreasing effect as the distance between the homes increases. These results indicate that there is in fact an event dependant near-repeat effect for nearby homes (e.g., those separated by less than 100 m) in Long Beach, but that this effect fades as the distance between homes increases, eventually disappearing completely for widely separated houses (e.g., those separated by distances greater than 3.9 km).

---

[1] Actually, there is a slight subtlety involved in this step that we should explain. To be precise, each pair of days d_1 and d_2 in which $d_1 \neq d_2$ is equally likely, and all of those pairs in which $d_1 = d_2$ are half as likely as that. This is due to the fact that each non-identical d_1, d_2 pair can be made in two ways: d_1 at house 1 and d_2 at house 2, or vice versa. In practice, we sidestep this issue by simply dividing the counted number of near-repeats by two if the time interval is greater than zero, so that our null hypothesis curve for near-repeats counted this way is continuous and precisely that of Eq. 10.

Discussion and Conclusions

Exact-repeat and near-repeat burglary patterns are often explained in terms of event dependence, in which burglars prefer targets with which they are familiar over targets that necessitate acquiring new information (Bernasco and Nieuwbeerta 2005). The daily routines of the residents may be known, as are the specific items that might be stolen, in a house that an offender has previously victimized. The risk of burglary is thought to spread to neighbors for many of the same reasons: even if the specific routines of residents, the goods to be stolen, or the general layouts of adjacent houses are not known exactly, they are likely to be similar. It is easy therefore to map a burglary strategy or set of cognitive scripts from one house to another (Wright and Decker 1994). Moreover, the expectation is that the risk of near-repeat burglary will spread more readily where housing types and the everyday routines of residents are spatially more homogeneous (Townsley et al. 2003; Johnson et al. 2007).

In this paper we have taken a new approach toward measuring and modeling event dependent repeat burglary effects, emphasizing the mathematical derivation of a null hypothesis model for the time interval between repeat burglaries. This model, termed the REH, assumes that burglaries at individual houses (or within some spatial neighborhood) are statistically independent events occurring at characteristic rates, and that persistent spatial heterogeneity of rates is solely responsible for any repeat victimization present. Equation 7 gives the probability of observing different time intervals between burglaries under the REH. We then empirically tested the REH using two different methods for counting burglaries. The moving-window method involves defining a temporal buffer equal to the maximum waiting time of interest, τ_{max}, between burglaries. The buffer ensures that we will count repeats only for houses burgled on days that can be subsequently observed for the full τ_{max} period. Using this counting method, the REH predicts that the probability distribution for the observed time intervals between exact-repeat burglaries decreases exponentially. Evidence from moving-window counting of residential burglaries in Long Beach, California is consistent with this formulation. This result is significant in suggesting that a decay-like distribution of waiting times between burglaries, measured using a moving-window counting procedure, is not necessarily indicative of a correlation between burglary events nor a preference for offenders to return to sites previously victimized (i.e., event dependence). However, further calculations based upon the moving-window count indicate that, though the time interval observations are consistent with the REH null hypothesis curve, the heterogeneity parameters derived from this count cannot explain the correct distribution of repeats within our data.

Because the moving-window counting method has a large number of parameters and does not seem to definitively test the REH, we explored an alternative method based on counting burglaries within a fixed time window. The resulting model gives the probability distribution of time intervals between events for houses burgled two or more times (see Eq. 10) within a set period of time. The fixed window counting procedure also predicts a decreasing probability of observing longer time intervals between burglaries, but in the case of houses victimized twice the relationship is linear with no free parameters (Eq. 9). Long Beach exact-repeat burglaries counted using the fixed-window method are inconsistent with the REH hypothesis. Repeat burglaries appear to be much more likely to occur in a short time interval after a first event than is expected given the REH, and it is clear from this counting method that burglary rates cannot be persistent in time. This result is significant in that it supports the view that there are temporal correlations between burglary

events and that offenders do preferentially return to previously victimized homes within a short period of time after an initial burglary.

Surprisingly, perhaps, different procedures for counting exact-repeat burglary events lead to different and seemingly contradictory test outcomes. However, it is possible to interpret Eq. 7 in a way such that both counting procedures yield results that are consistent with event dependant repeat burglary. Specifically, rather than referring to a persistent environmental heterogeneity in burglary risk, Eq. 7 may be taken as a description of individual houses stochastically transitioning between states of varying burglary risk following each burglary event. The probability of adopting state i after any event is given by w_i and the corresponding burglary rate for that state is λ_i. Using parameters fit to our Long Beach data, for example, a house in state $i = 1$, with a baseline burglary rate of $\lambda_1 = 5.32 \times 10^{-5}$ burglaries per day may, after an event, adopt an excited state $i = 2$ with probability $w_2 = 0.066$ and burglary rate $\lambda_2 = 2.45 \times 10^{-3}$ burglaries per day. We expect a repeat burglary to occur at this house on average within 408 days, which is 46 times faster than the expected interval between burglaries at houses that remain in the state $i = 1$. Similarly, we expect a house to adopt the even more excited state $i = 3$ with probability $w_3 = 0.019$ and a characteristic burglary rate of $\lambda_3 = 8.41 \times 10^{-2}$ burglaries per day. In this case, the expected time interval to a repeat burglary at the house is approximately 12 days, more than 1,530 times faster than a house in the baseline state, and 33 times faster than a house in the intermediate excited state. We have not specified the exact causes behind transition between different states, though it is reasonable to infer excited states (greater risk of repeat burglary) in some way correspond to burglars' preferences to return to previously victimized houses, as discussed previously.

Overall, the models and results presented here confirm that event dependence in the exact-repeat burglary effect exists and can be quantitatively distinguished from the REH using appropriate counting procedures. A similar conclusion is drawn when examining near-repeat burglary events, which are significantly different from a corresponding null hypothesis at short distances between events (e.g., 0–100 m), but are indistinguishable from the null hypothesis at greater distances (e.g., 3.9–4.0 km).

Most importantly, the general rate transition model described by Eq. 7 and our approach to testing for the presence of this behavior using a fixed-window counting procedure (Eq. 10) is potentially applicable to other types of repeat victimization. For example, we might hypothesize that the individuals in a given population may each be at risk of being a victim of a violent crime and, subsequent to being victimized for a first time, adopt according to some probability distribution one of several possible states, each with its own characteristic victimization rate. Some individuals will fall into a baseline group after victimization, while others might be more at risk after being victimized (Lauritsen and Quinet 1995). Similarly, Ratcliffe and Rengert (2008) analyze shootings in Philadelphia and find an elevated risk of near-repeat shootings occurring within 2 weeks and within one city block of previous incidents. Our model would suggest that the near-repeat effect observed by Ratcliffe and Rengert stems from the area around a previous shooting event transitioning to an excited state characterized by a higher rate of shooting events.

The results here reinforce the view that repeat and near-repeat victimization may play a role in the nucleation of crime patterns in space and time and, as a consequence, may be an appropriate basis for designing crime prevention strategies. However, we also note that there are a number of challenges yet to meet in designing optimized responses to repeat crimes. In particular, our results from analyses of exact-repeat burglaries in Long Beach suggests that at any given time only about 0.002% of houses exhibit the highest excited state (Eqs. 11 and 12) with an expected time to a repeat event of approximately 12 days.

This corresponds to only about 1 single family residence from a total of approximately 70,000 units. The challenge is to determine which house(s) belong to this very small set, which would allow preferential targeting of resources at these locations.

Open Access   This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Bernasco W, Nieuwbeerta P (2005) How do residential burglars select target areas? A new approach to the analysis of criminal location choice. Br J Criminol 45:296–315

Bowers KJ, Hirschfield A, Johnson SD (1998) Victimization revisited: a case study of non-residential repeat burglary on Merseyside. Br J Criminol 38:429–452

Eck JE, Chainey S, Cameron JG, Leitner M, Wilson RE (2005) Mapping crime: understanding hot spots. National Institute of Justice/NCJRS, Washington, DC

Farrell G, Pease K (1993) Once bitten, twice bitten: repeat victimization and its implications for crime prevention. Crime Prevention Unit Paper 46. Home Office, London

Farrell G, Pease K (1994) Crime seasonality: domestic disputes and residential burglary in Merseyside 1988-90. Br J Criminol 34:487–498

Farrell G, Pease K (eds) (2001) Repeat victimization. Criminal Justice Press, Monsey

Farrell G, Phillips C, Pease K (1995) Like taking candy: why does repeat victimization occur? Br J Criminol 35:384–399

Farrell G, Bowers KJ, Johnson SD, Townsley M (eds) (2007) Imagination for crime prevention. Crime prevention studies, vol 21. Criminal Justice Press, Monsey

Feller W (1968) An introduction to probability theory and its applications, 3rd edn., vol 1. Wiley, New York

Johnson SD (2008) Repeat burglary victimisation: a tale of two theories. J Exp Criminol 4:215–240

Johnson SD, Bowers KJ (2004) The stability of space-time clusters of burglary. Br J Criminol 44:55–65

Johnson SD, Bowers KJ, Hirschfield AFG (1997) New insights into the spatial and temporal distribution of repeat victimization. Br J Criminol 37:224–241

Johnson SD, Bernasco W, Bowers KJ, Elffers H, Ratcliffe J, Rengert G, Townsley M (2007) Space-time patterns of risk: a cross national assessment of residential burglary victimization. J Quant Criminol 23:201–219

Lauritsen JL, Quinet KFD (1995) Repeat victimization among adolescents and young-adults. J Quant Criminol 11:143–166

Nelson JF (1980) Multiple victimization in American cities: a statistical analysis of rare events. Am J Sociol 85:870–891

Ratcliffe JH, McCullagh MJ (1999) Hotbeds of crime and the search for spatial accuracy. Geogr Syst 1:385–398

Ratcliffe JH, Rengert GF (2008). Near-repeat patterns in Philadelphia shootings. Secur J 21:58–76

Sagovsky A, Johnson SD (2007) When does repeat burglary victimization occur? Aust N Z J Criminol 40:1–26

Short MB, D'Orsogna MR, Pasour V, Brantingham J, Tita G, Bertozzi A, Chayes L (2008) A statistical model of criminal behavior. Math Model Methods Appl Sci 18:1249–1267

Townsley M, Homel R, Chaseling J (2003) Infectious burglaries. A test of the near repeat hypothesis. Br J Criminol 43:615–633

Tseloni A, Pease K (2003) Repeat victimisation: 'boosts' or 'flags'? Br J Criminol 43:196–212

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

Wright RT, Decker SH (1994) Burglars on the job: streetlife and residential breakins. Northeastern University Press, Boston