Testing Indicators of Risk Populations for Theft from the Person across Space and Time: The Significance of Mobility and Outdoor Activity

Song, Guangwen; Liu, Lin; Bernasco, Wim; Xiao, Luzi; Zhou, Suhong; Liao, Weiwei

published in
Annals of the American Association of Geographers
2018

DOI (link to publisher)
10.1080/24694452.2017.1414580

document version
Publisher's PDF, also known as Version of record

document license
Article 25fa Dutch Copyright Act

Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: https://www.tandfonline.com/loi/raag21

Testing Indicators of Risk Populations for Theft from the Person across Space and Time: The Significance of Mobility and Outdoor Activity

Guangwen Song, Lin Liu, Wim Bernasco, Luzi Xiao, Suhong Zhou & Weiwei Liao

To cite this article: Guangwen Song, Lin Liu, Wim Bernasco, Luzi Xiao, Suhong Zhou & Weiwei Liao (2018) Testing Indicators of Risk Populations for Theft from the Person across Space and Time: The Significance of Mobility and Outdoor Activity, Annals of the American Association of Geographers, 108:5, 1370-1388, DOI: 10.1080/24694452.2017.1414580

To link to this article: https://doi.org/10.1080/24694452.2017.1414580

Published online: 30 Jan 2018.

Testing Indicators of Risk Populations for Theft from the Person across Space and Time: The Significance of Mobility and Outdoor Activity

Guangwen Song,* Lin Liu ,† Wim Bernasco ,‡ Luzi Xiao,* Suhong Zhou,* and Weiwei Liao*

*Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, and Guangdong Key Laboratory for Urbanization and Geo-Simulation

†Center of Geo-Informatics for Public Security, School of Geographic Sciences, Guangzhou University, and Department of Geography, University of Cincinnati

‡Netherlands Institute for the Study of Crime and Law Enforcement and Department of Spatial Economics, School of Business and Economics, Vrije Universiteit Amsterdam

In recent years, it has increasingly been recognized that due to the uncertain geographic context problem caused by daily human mobility, the residential population is too static to serve as a valid measure of the population at risk for criminal victimization. Various alternative measures have been suggested instead. Guided by the routine activity approach, this study furthers the concept of crime risk population and its measurement across space and time. Using exceptionally comprehensive data sets on population mobility and on theft from the person in a large city in China, we select the best indicator of the risk population from the following four candidates: residential population, subway ridership, taxi ridership, and mobile phone users. Controlling for the potentially confounding effects of offender and guardian presence, we show that on both weekdays and weekends, the best indicators of risk population vary over the course of the day. In the morning, residential population outperforms other measures. In the afternoon and evening, taxi ridership and phone users are better indicators. Although the mobile phone user base forms an arguably more representative measure of ambient population, during some periods taxi ridership is superior because it provides a better indicator of outdoor (as opposed to indoor) activities. In terms of practical applications to security policy and law enforcement, these findings can help identify crime hot spots by calculating accurate crime risks. *Key Words: ambient population, residential population, risk population, theft from the person.*

由于人类日常活动所造成的不确定地理脉络问题，导致居住人口过于静态而无法作为有效测量犯罪被害的风险人口之有效方法之问题，已于近年来逐渐受到认识，并提出了各种另类的测量方法。本研究以日常活动方法为指引，推进犯罪风险人口的概念，及其于时空中的测量。我们特别运用在中国大城市中个人的人口移动与窃盗的综合数据集， 从以下四大候选人中选择最佳的风险人口指标:居住人口、地铁乘客、出租车乘客、以及手机使用者。我们控制犯法者与保护者在场的潜在混淆效应，同时展现在週间与週末, 风险人口的最佳指标随着日程而有所改变。在早上, 居住人口胜过其他测量。在下午与夜间, 出租车乘客和手机使用者是较佳的指标。尽管手机使用者可说是更能代表週遭人口，但在若干时期，出租车乘客却更具代表性， 因其对户外(相对于室内)活动提供了更佳的指标。在安全政策与执法的实务应用方面, 这些研究发现能够透过计算确切的犯罪风险, 协助指认犯罪热点。 *关键词： 週遭人口, 居住人口, 风险人口, 个人窃盗。*

Desde no hace mucho tiempo, se concede que, debido al incierto problema del contexto geográfico determinado por la movilidad humana diaria, la población residencial es demasiado estática para servir como medida válida de población en riesgo de victimización criminal. En sustitución, se han sugerido varias medidas alternativas. Guiado por el enfoque de la actividad rutinaria, este estudio impulsa el concepto de población en riesgo de crimen y su medición a través del espacio y el tiempo. Usando conjuntos de datos excepcionalmente detallados sobre movilidad de la población y sobre robo a la persona en una ciudad grande de China, seleccionamos el mejor indicador de población en riesgo entre los siguientes cuatro candidatos: población residencial, número de pasajeros del metro, número de pasajeros de taxi y usuarios de teléfonos celulares. Haciendo el control de potenciales efectos de confusión a la presencia del delincuente y del guardián, mostramos que, tanto en días normales como en los fines de semana, los mejores indicadores de la población en riesgo varían en el curso del día. En la mañana, la población residencial genera mejores resultados que otras medidas. En la tarde y durante la noche, el número de pasajeros de taxi y usuarios de teléfonos celulares son mejores indicadores. Aunque la base de usuarios de teléfonos celulares forma una medida supuestamente muy representativa de la población ambiente,

Annals of the American Association of Geographers, 108(5) 2018, pp. 1370–1388 © 2018 by American Association of Geographers
Initial submission, May 2017; revised submissions, September and November 2017; final acceptance, November 2017
Published by Taylor & Francis, LLC.

durante algunos períodos el número de pasajeros de taxi sirve mucho más debido a que esta medida proporciona un indicador de actividades fuera de la casa (como opuesto de actividades en el interior de la residencia). En términos de implicaciones prácticas a la política de seguridad y aplicación de la ley, estos descubrimientos pueden contribuir a identificar puntos calientes del crimen calculando riesgos exactos de criminalidad. *Palabras clave: población ambiente, población residencial, población en riesgo, robo a personas.*

The uncertain geographic context problem (Kwan 2012) has become a basic concern for researchers studying effects of area-based attributes on individual behaviors. Due to the daily mobility of the population and the resulting fluctuations in numbers of people physically present, static indicators of population size cannot reliably be used to assess the effects of geographical context on behavior. This limitation also applies to crime research. One of the major theoretical frameworks in crime research, the routine activity approach (Cohen and Felson 1979), claims that the convergence in space and time of a motivated offender and a suitable target in the absence of a capable guardian is a necessary and sufficient condition for predatory crime. This claim has been made explicit in a recent formalization and extension of the approach (Hipp 2016).

How to measure the activity or presence of suitable targets, motivated offenders, and capable guardians thus becomes a key issue in the explanation of spatial and temporal patterns of crime. In this article, our main aim is to improve the empirical assessment of the theory by addressing how one of these three necessary elements, the presence of suitable targets, can be adequately measured. Adequate measurement will minimize bias and thus reduce inferential errors. We focus on theft from the person, because it is one of the most frequent street crimes in China and because it is supposed to have a strong relationship with the physical presence of people.

In explaining the occurrence of crime, the routine activity approach prioritizes neither offenders, targets, nor guardians. Rather, it emphasizes that all three must be present at the same time and place (Cohen and Felson 1979). The extant literature that is inspired by the routine activity approach, however, strongly highlights the crucial role of targets, because potential targets are viewed as the population at risk of victimization. Although in our analysis we adopt a broader perspective and include measures of motivated offenders and of capable guardians, our theoretical focus and contribution to the literature is mainly about the presence and measurement of potential targets of crime.

The salience of crime targets is exemplified by the common method of calculating crime rates. Crime rates are calculated with the purpose of comparing the volume of crime between different periods, different locations, or different crime types. To standardize the volume of crime by its most important determinant, crime rates are defined as the ratio of the number of crimes and the number of potential targets.

The measurement of potential targets of crime is not a trivial issue. Traditionally, in calculating crime rates of areal units like countries, cities, or neighborhoods, the residential population has been chosen as the denominator of the crime rate (Mburu and Helbich 2016). One of the reasons for this choice is convenience, as residential population is a statistic that is generally available at multiple levels of spatial aggregation (Malleson and Andresen 2016).

Problems with Residential Population-Based Measures

In critique of the common method of calculating crime rates, Boggs (1965) suggested that the conventional use of residential population as a generic crime rate denominator for all types of crime can lead to biased crime rates because residential population is a poor proxy for criminal opportunity (i.e., for the presence of targets). She argued, for example, that central business districts have spuriously high crime rates because they are characterized by large numbers of potential crime targets (merchandise in stores, unattended parked cars, people on the streets) but small residential populations. Although subsequent research (Cohen, Kaufman, and Gottfredson 1985) has suggested that alternative denominators for burglary and auto theft targets do not yield very different conclusions, Boggs's general claim that crime rate denominators should reflect criminal opportunities has echoed in the literature for decades.

The discussion on crime rate denominators is not limited to making justified and fair comparisons between areal crime rates (e.g., to the question of whether Chicago is safer than Los Angeles). As indicated at the outset of this article, a correct measurement of the presence of suitable targets is essential for appropriate tests of routine activity theory and thus also for correct predictions of crime volumes: The notion of a crime rate already implies that the denominator is an important correlate of the numerator (as otherwise it would be redundant) and should thus be an important factor in the explanation of variations in crime volumes (Osgood 2000; Liu and Eck 2007).

Recent studies have readdressed the measurement of risk populations for crime. It has been suggested that crime rates of areas like neighborhoods should not be based on residential population but on ambient population. The ambient population in a given area is the number of individuals who are actually physically present in the area, rather than the number of individuals residing in it.

Nonresidential Population-Based Measures

Survey Data and LandScan Global Population Database

Based on the notion that cities vary in terms of inbound and outbound commuter flows, various studies have combined measures of resident populations with commuting surveys to calculate commuter-corrected estimates of the daytime population to predict crime intensities. Stults and Hasbrouck (2015) used commuter-corrected estimates of the daytime population to model variation in crime rates (homicide, aggravated assault, robbery, burglary, larceny, and auto theft) across 166 cities in the United States. Mburu and Helbich (2016) used commuter-harmonized estimates to model crime volumes (violence, disorder, theft and shoplifting, robbery, burglary, and vehicle crime) in small administrative areas in London, UK. Liu and Eck (2007) used commuter-corrected vehicle miles as the denominator for calculating rates of police vehicle stops for drivers with different ethnicities in the city of Cincinnati in the United States. In support of the routine activity approach, these three studies demonstrated that commuter-corrected estimates performed better than measures of residential population.

In addition to commuter flows, Boivin and Felson (2017) included mobility flows of trips made for the purposes of shopping, recreation, and education. Using data from a large city in eastern Canada, they showed that property and violent crimes were more strongly related to inflow for the purpose of recreation and, to a lesser extent, shopping than for the purpose of work or education. Their findings are supported by research that has explored the relations between types of land use and crime (Wilcox et al. 2004; Bernasco and Block 2011). Schools, retail businesses, restaurants, bars, parks, transit stations, and other places and facilities that attract large numbers of people are crime generators[1] that provide criminal opportunities because they are concentrations of potential victims and targets (Brantingham and Brantingham 1995; Kinney et al. 2008).

Other studies have used the LandScan Global Population Database to obtain an estimate of the ambient population in Vancouver, Canada (Andresen 2006, 2007, 2011; Andresen and Jenion 2010). These studies have demonstrated that using the LandScan Global Population Database to measure the ambient population in some circumstances provides better estimates of the risk population than using the residential population. At the very least it can be used to supplement the conventional measure that uses the residential population as the denominator of crime rates.

Big Data–Based Measures

Big data generated by social media and mobile phones can potentially generate more accurate measures of population at risk. A recent phenomenon embraced by social scientists is the advance of data from social media platforms. Some of these, in particular Twitter messages, are geotagged and thus contain information about where their authors are located. Various authors have attempted to use geotagged Twitter messages as a measure of ambient population to help predict crime levels in Leeds, UK (Malleson and Andresen 2015), and Charlottesville, Virginia (Wang, Gerber, and Brown 2012; Gerber 2014). It improves the ability of crime prediction compared to the residential population statistics.

The usefulness of Twitter messages for estimating ambient populations is limited, however, because the user base of Twitter is unlikely to be representative of the population and because only a minority of users post their messages with geotags. The utilization of location data from mobile phones seems more promising because of the widespread adoption and use of mobile phones and because geolocating phones is necessary for their proper functioning.

Many studies have measured ambient populations with data generated by cell phone use (e.g., Ratti et al.

2006), but very few have used it to model variations in crime volume. Hanaoka (2016) found that in the daytime, the number of street robberies had a weaker relation with the density of mobile phone use than in the nighttime in Osaka City, Japan. Bogomolov et al. (2014) found that mobile phone data, in combination with basic demographic information, could be used to predict crimes in London.

Using data from London, Malleson and Andresen (2016) compared the effectiveness of mobile phone use intensity with other measures of ambient population (including residential population and Twitter data) in the prediction of crime. They used correlation coefficients between the indicators and crimes to benchmark the indicators' performance in crime prediction. Although the comparative benchmark of different measures of ambient population is an important step forward, the study of Malleson and Andresen still has some limitations, including the assumption of linearity in the relation between indicators and crime, the exclusion of covariates other than ambient population in the crime equation, and, perhaps most important, the lack of time information in their crime data, which prevented them from studying the effects of temporal variations in the ambient population.

Summary

The four alternative measures of ambient population suggested here (mobility flows, Landscan Global Population Database, social media, and mobile phone location services) do not fully address the issue that Boggs identified. They are data-intensive attempts to estimate the presence of people at detailed spatial and temporal scales. They are, however, not necessarily the best indicators of the risk population for crime, of "suitable targets." Generally, most of the contributions to the unfolding literature on ambient populations and crime have implicitly or explicitly been based on four assumptions that are challenged in this article.

First, in the literature, it has generally been assumed that the chosen measure of ambient population equals the population at risk. Only a single study (Malleson and Andresen 2016) compared different measures of ambient population. We challenge the assumption that ambient population can be measured unambiguously and that it can be equated with the risk population. To that end, we argue that varying proportions of the ambient population might be physically present at a given location but not actually at risk for victimization. For example, while relaxing in their homes or working in their offices, individuals might be present in a place but not actually at risk for street robbery, pickpocketing, theft from the person, or other offenses perpetrated in public space. Messner and Blau (1987) demonstrated at the macrolevel that the frequency of leisure activities within households reduced crime rates but leisure activities outside households increased crime rates. Indeed, a key hypothesis of the routine activity approach is that "the dispersion of activities away from households and families increases the opportunity for crime and thus generates higher crime rates" (Cohen and Felson 1979, 588). Consequently, the analysis of crime for which opportunities exist in outdoor settings requires indicators that capture outdoor activities rather than indoor activities (in particular relating to home and work settings) and vice versa. One disadvantage of the data sources discussed previously, including mobile phone location data and Twitter data, is that they do not distinguish between outdoor and indoor activities.

Second, the time differentiation of the ambient population has typically not been fully addressed. Although most literature on ambient population has indeed noticed and emphasized that measures of the residential population fail to reflect the daily mobility patterns, due to limitations of the available data, most have still used static measures to represent the ambient population over the course of the day and the week. Temporal variation has therefore not been given the necessary attention. Accounting for the fact that crime is both spatially and temporally heterogeneous (Kwan 2012), we model crime risk as being dependent on measures of mobility (in particular, taxi and subway ridership and mobile phone locations) that do vary spatiotemporally over the course of the day and the week.

Third, the cited literature addressing crime and ambient population literature did not take motivated offenders and capable guardianship into account. These are two key concepts in the routine activity approach that could potentially confound the relation between ambient population and crime. Proximity to where offenders live has been demonstrated to increase crime rates (Bernasco and Luykx 2003), and lack of capable guardianship can lead to more crime (Reynald 2009). Helbich and Jokar Arsanjani (2015), for example, found that the Euclidean distance to police stations has a significant negative impact on nonviolent crime. Therefore, a rigorous assessment requires that we not only consider effects of potential targets but also account for the potential effects of motivated offenders and capable guardians.

Fourth, the large majority of work on the routine activity approach has been based on the empirical analysis of data in Western countries, in particular Canada, the United Kingdom, and the United States. Because there is no reason to geographically or culturally restrict the scope of routine activity theory, our analysis is based on data from a large city in China. The Chinese context allows us to test hypotheses that have been developed with the geography and culture of Western cities in mind on data from China. We thus add to a couple of recent studies that also applied elements of routine activity theory to crime in China, including Xu (2009), Zhang, Messner, and Liu (2007), Peng et al. (2011), Feng, Dong, and Song (2016), J. Chen et al. (2017), and Liu and Li (2017), albeit with different research objectives and empirical and analytical approaches.

In conclusion, to address the uncertain geographic context problem in crime research and motivated by limitations in the extant literature on ambient population and crime, we selected a large city in China to study indicators of the risk populations for theft from the person over the course of the day and the week by comparing four different measures, controlling for the presence of motivated offenders and capable guardianship. We aim to test the indicators of risk populations for theft from the person by improving the measures and by giving more consideration to temporal variation. In the remainder of this article, we discuss our data and methods, present our findings, and conclude and discuss the results.

Data and Methods

The area of this study is the central area of a metropolis in the southeast of China labeled ZG City in this article.[2] ZG City has a total population in excess of 10 million and is one of the most developed cities in China. The central area of ZG City is defined as the district within the beltway surrounding it and has an area of 203 km^2. It is an active area that attracts large numbers of visitors from outside. For this reason, the central area is an excellent test case for a study on the effect of ambient population measures on theft from the person. In addition, the area allows optimal measurement of mobile phone locations because it has a denser distribution of cell signal towers than other areas in ZG City.

Spatial and Temporal Selection and Units of Analysis

This study combines data on seven different variables: thefts from the person, mobile phone use, taxi ridership, subway ridership, residential population, offender residence locations, and police station locations, all of which were obtained from different departments, companies, and public sources. The first four variables are both georeferenced and time-stamped; we know where and when they took place. The last three are treated as stable over time and are thus only georeferenced.

Due to time limitations of the taxi ridership data, it was not possible to include 24/7 measures of the ambient population. Instead, we were able to use data from 7:00 a.m. to 10:00 p.m. for all days of the week and for each of the four measures.

Two main methods are available for categorizing time of day. One method is to distinguish a fixed number of equally sized time intervals. This method is straightforward, is easy to implement, and does not require any further assumptions, much like using a grid system for geographic data. An example of this method is a study by Bernasco, Ruiter, and Block (2017), who divided a full day into twelve two-hour intervals.

The second method is to create a limited number of categories based on some criterion of homogeneity. An example is a study by Haberman and Ratcliffe (2015), who used time use surveys to devise a categorization of four daily time intervals that maximally aligned with the timing of standard routine activities, such as morning rush hour and work or school. This method is complex, potentially generates time intervals with greater internal homogeneity in terms of main activities, and is also jeopardized by the fact that routine activities display a great deal of temporal variation. To make no unnecessary assumptions and to facilitate an effective comparison of the four measures of ambient population, we decided to apply the first method and distinguished between five intervals of three hours each: 7:00 a.m. to 10:00 a.m., 10:00 a.m. to 1:00 p.m., 1:00 p.m. to 6:00 p.m., 4:00 p.m. to 7:00 p.m., and 7:00 p.m. to 10:00 p.m. To capture different routine activity patterns during weekdays and weekends, we separated the weekdays (summed over Monday–Friday) from the weekends (summed over Saturday and Sunday).

Overlaying the study area with a 1 km × 1 km grid resulted in 205 grid cells, which are the spatial units of analysis in this study. This spatial resolution is the same as that in the study by Andresen (2011) cited earlier. All data were aggregated to the 205 grid cells for each of the five three-hour periods on weekdays and weekends, respectively.

Because it has been argued that land use features could operate at microscales (Weisburd, Bernasco, and Bruinsma 2009) and to minimize heterogeneity within units (Bernasco and Block 2011), there has been a tendency in the recent literature to use small spatial units of analysis, such as street blocks or street segments. Nevertheless, Boessen and Hipp (2015) demonstrated that many neighborhood factors have wide-ranging effects and that using spatial units of analysis that are too small leads to difficulties in modeling spatial spillover effects.

Many studies use census tracts, often with average area less than 2 km^2, as the spatial unit of analysis, and this spatial scale has proven to be small enough for social and physical attributes to be homogenous (Mburu and Helbich 2016; Boivin and Felson 2017). In China, the central area in ZG City is densely populated, and 1 km^2 grid cells seem to be an appropriate choice, as this scale is neither too large nor too small, and it is also a compromise between the required level of spatial detail and geocoding precision. We comprehensively explain the geocoding procedure in the "Geocoding Method" section.

Crime Data and Geocoding

Crime Data

Information on all cases of theft from the person in the year 2014 was obtained from the Public Security Bureau of ZG City. The data are independent of the mode of reporting to the police and therefore include crimes reported to the police in different ways, such as calling 110 (emergency number), going to a police station in person, or calling a local police station on the phone. For each theft from a person, information was provided on the date and approximate time and location where the crime took place. In the geocoding process, we assigned each theft from the person incident a specific pair of coordinates and subsequently aggregated these incident points to their corresponding 1 km^2 grid cell.

Besides crime incidents, the residential addresses of all offenders who were arrested for one or more thefts from the person in 2014 were geocoded and aggregated to their corresponding 1 km^2 grid cell. This offender count forms a measure of the local presence of offenders motivated and capable of committing thefts from the person. For each grid, the sum of offender counts of the eight neighboring grid cells is used to measure the presence of offenders in nearby areas. These two measures are obviously stable across the five time periods and across weekdays and weekends.

The police station (*Paichusuo* in Chinese) is the grassroots government unit of policing in China. In the study area, there were seventy-seven police stations. Their addresses were found using city maps. If a grid cell had a police station within its boundaries, its distance to the closest police station was coded 0. If not, the distance refers to the distance between the centroid of the grid cell and the nearest police station. It is supposed that because for police officers a police station is an anchor point that they repeatedly return to before, during, and at the end of their shift, on average they must spend more time (and exercise formal guardianship) in the proximity of the police station than further away from it.

Geocoding Method

The crime data included when and where the crime took place, but they were provided without geographic coordinates. We developed a method to geocode the address by using multiple Web-based geocoding services. It was based on an improved version of the work of Cui (2013), who compared and combined the Google Geocoding API,[3] the Yahoo Geocoding API, and the ESRI Address Coder to geocode and found that using multiple geocoding services to geocode user-generated locations is a time-saving, efficient method. In China, we used the top three map companies (Baidu Map, Tencent Map, and Gaode Map) to geocode the crime addresses (Liao et al. forthcoming).

The crime address includes four elements: district or town, police station, street, and nearest house number. We first used the APIs of the three map companies to obtain the coordinates of crimes and then checked the confidence level of the outcome provided by each map and used Dixon's Q test (Dean and Dixon 1951) to recognize any outliers. Finally, based on the performance of different maps, the final coordinates were chosen.

Subsequently, we used grid cell sizes of 0.25 km^2 and 1 km^2 to estimate the accuracy of the geocoding results. We chose 4,000 addresses randomly to compare their automatically constructed coordinates with the coordinates found manually. The results demonstrated that for grid cells of 500 m $\times$ 500 m, 92.5 percent of the addresses were geocoded correctly. For grid cells of 1 km × 1 km, the percentage correctly geocoded was 97.3 percent. Taking into account that 85 percent has been mentioned as a first estimate of a minimum reliable geocoding rate (Ratcliffe 2004) and that a 95 percent geocoding success rate has been rated as both acceptable and achievable (Ratcliffe 2010), we decided to use the 1 km × 1 km grid cells as the spatial unit of analysis for our study.

Four Measures of Population at Risk

Residential Population

The sixth census data in ZG City, collected in 2010, offered information about the number of residents in each census unit. To calculate the residential population per grid cell, we intersected the geographies of the grid cells and census units and calculated the surfaces of the resulting areas. Subsequently, we estimated the residential population in the grid cells by summing the residential populations of the census units, weighted by the proportion of the census unit that was located inside the grid cell. For example, a census unit located completely inside the grid cell received weight 1 and its population was completely assigned to the grid cell. If, however, only half the census unit was located inside grid cell, the census unit was assigned a weight of 0.50 and only half of the population was assigned to the grid cell.

Mobile Phone Users

The data of mobile phone users was offered by a major mobile phone service provider in China that has a market share of 22.5 percent (Chong, Teoh, and Qi 2015). There is no reason to assume any major systematic differences between the customers of different mobile phone providers, which suggests that our sample of mobile phone users is representative of the general population in ZG City.

The geographic data of mobile phones is based on the cellular signaling information and it includes the anonymized and aggregated total number of mobile phone users of the 2G and 3G networks per cell tower. Mobile phones will typically attempt to connect to the nearest tower. Note that cellular signaling data involves any behavior that creates a relation with the cell signal tower, such as Internet searches, messaging, and calls. Hourly statistics were collected for a complete week, from 12 through 18 May 2016 (including the weekend of 14 and 15 May). In the central area of ZG City, the density of the base stations is quite high and neighboring cell signal towers are within a distance of 500 m.

To aggregate the mobile phone user population of cell signal towers to the 205 grid cells in the study area, Thiessen polygons (also known as Voronoi polygons) were created with the cell tower locations as the seeds. For mobile phone users within the Thiessen polygon, the local cell signal tower is the nearest tower. Subsequently, the procedure described for the residential population of the census units was followed; that is, the Thiessen polygons were intersected with the grid cells and mobile phone users were assigned to the grid cells based on the proportion of the Thiessen polygon that was located inside the grid cell. Note that Global System for Mobile Communications signals do not distinguish between indoor and outdoor mobile phone use and that all users are thus included in the hourly data, irrespective of whether they are outside or inside buildings.

Taxi Ridership

The taxi is a popular mode of transportation for citizens in ZG City. There are about 20,000 taxis in the city and all of them are equipped with Global Positioning System (GPS) devices. Although the taxis belong to different companies, all GPS information is reported to a single official supervision department.

In the literature, taxi ridership data have previously been applied as an important source of information in the analysis of land use (Pan et al. 2013), of intracity human mobility (Li et al. 2012), and even of social functions (Qi et al. 2011). To our knowledge, taxi ridership has not been applied before in research on crime. In this study, taxi ridership is used to represent the volume of outdoor activity.

We obtained the taxi GPS data for a complete week, 23 through 29 March 2014 (23 March was a Sunday and 29 March was a Saturday). The taxi data include the longitude and latitude of the location of the taxi as well as its state of carrying passengers. For example, if there is a passenger in the taxi, the passenger state is 2 and it will return 1 when the passenger gets out. Therefore, from this state and its transition between 1 and 2, we identified the origins (1 to 2) and the destinations (2 to 1) of each journey. Subsequently, origins and destinations were aggregated to the grid cell in which they were situated.

Subway Ridership

The widespread implementation of smartcard-based fare payment in transportation systems has provided the transportation companies and geographers with large volumes of travel data (Bagchi and White 2005; Pelletier, Trépanier, and Morency 2011). Subway ridership data have been used for various purposes, including the spatial and temporal analysis of ridership volume (e.g., C. Chen, Chen, and Barry 2009). ZG City has a large subway transportation system that also uses smartcards for fare calculation and collection. It is one of the most important public transportation modes, and ridership is expected to be strongly correlated with the presence of people at or near subway stations. Subway ridership data were provided by the only subway company in ZG City. The data set provided by the subway company aggregates the number of passengers who entered and left the subway stations, separately by hour and for each subway station from 3 through 9 March 2014. This included the weekend of 8 and 9 March.

To aggregate the subway ridership to grid cells, we set a threshold of 1.5 km as the service radius of a subway station. A buffer of 1.5 km around each subway station was constructed and, similar to the procedure for mobile phone data, subway stations were used as the seeds (generators) to create the Thiessen polygons. The service circles were first intersected with the Thiessen polygons to identify the service area for every subway station. The passengers of subway stations were allocated to all resulting intersections, based on the proportion of the intersection surface that was located inside the whole service area. After that, the same procedure was used to aggregate the passenger volumes to the grid cells. Following this method, there will not be any passengers allocated to a grid cell if the grid cell is located further than 1.5 km from a subway station. If there are multiple subway stations within 1.5 km of the grid cell, only passengers from the nearest station are assigned to the grid cell.

Statistical Models

Our analysis focuses on explaining variation in the frequency of theft from the person across the 205 1 km^2 grid cells in the study area. Because the number of crimes is a count variable that can only take nonnegative integer values, a count regression method seems most appropriate. A regular Poisson regression model requires that the mean of the crime count equals its variance (equidispersion), an assumption that is likely too strict for the data. Many empirical count distributions are characterized by overdispersion, where the variance is larger than the mean (Osgood 2000; J. Chen et al. 2017).

The negative binominal regression model (Hilbe 2011) is a generalization of the Poisson model. It relaxes the equidispersion property, and for that reason it is often used as a preferred alternative to the Poisson model (e.g., Bernasco and Block 2011; Haberman and Ratcliffe 2015). Because the unconditional distributions of the frequency of theft from the person during the different three-hour periods were overdispersed, we decided to apply negative binominal regression models.

We estimated the same negative binomial regression model for each of the five three-hour periods and separately for weekdays and weekends. In correspondence with the routine activity approach, each model included a measure of potential targets, a measure of motivated offenders, and a measure of capable guardianship. The four alternative indicators of potential targets were included in separate models that are compared with respect to model fit. The presence and proximity of motivated offenders is viewed as a control variable and is included in all models. It is measured by the number of offenders residing in the grid cell and by the number of offenders living in adjacent grid cells. Capable guardianship is also treated as a control variable and included in every model. It is measured by proximity to the nearest police station.

Spatial autocorrelation occurs when characteristics at nearby locations are either positively correlated or negatively correlated. Spatial autocorrelation in the error terms violates standard statistical techniques that assume independence among observations.[4] Positive spatial autocorrelation is very common in geographic data on crime and other social phenomena. To address this issue, referring to Bernasco and Block (2011) and Boivin and Felson (2017), we used spatially lagged variables of the independent variables. These included, in particular, the number of motivated offenders residing in adjacent areas and the volume of the ambient populations in adjacent areas. The rationale is that effects of motivated offenders and ambient populations on crime do not stop at the boundary of an area but spill over into nearby ones. Adjacency was determined by the Queen criterion.

Taking lagged effects into account does not necessarily mean that any residual spatial autocorrelation disappears (Bernasco and Block 2011). To evaluate the reduction in spatial autocorrelation achieved, we compared residual autocorrelation of a null model with residual autocorrelation after estimating the models.

For model comparison, we used the Akaike's information criterion (AIC), as it is generally considered an appropriate benchmark to judge relative model fit between multiple nonnested negative binomial models (Hilbe 2011). The smaller the value of the AIC, the better the model fits. Although rules of thumb have been defined for judging AIC differences, a more rigorous comparison involves a bootstrapping procedure, in which a bootstrap sample (a sample with replacement of N cases from a data set of size N) is taken S times repeatedly from the original sample, and a frequency distribution is generated of the most preferred model across the S bootstrap replications (Burnham and Anderson 2002). Based on advice in Burnham and Anderson (2002), we selected a value of $S = 1{,}000$ bootstrap replications.

We calculated variance inflation factors (VIFs) to check for multicollinearity problems between the independent variables (Belsley 1991). As all VIF values were lower than three and all correlation coefficients, except the ones between values and lag values of risk population indicators, were below 0.60, multicollinearity was not considered an issue.

Findings

Descriptive Statistics

We show the descriptive statistics of variables for all periods in Table 1. The residential population, the distance to the closest police station, the number of offenders in the local grid cell population, and the number of offenders residing in contiguous grid cells are static measures and thus equal in all three-hour periods. Most thefts on weekdays and on weekends are in the period from 4:00 p.m. to 7:00 p.m. As for subway ridership, there are obvious peak hours (7:00 a.m.–10:00 a.m. and 4:00 p.m. to 7:00 p.m.) on weekdays, whereas on weekends the ridership is on a steady increase until 4:00 p.m. to 7:00 p.m. Taxi ridership also changes with time. Between 4:00 p.m. and 7:00 p.m. on weekdays, taxi ridership is quite low due to the handover of taxi drivers. The number of phone users in the central district reaches its top level in the period between 4:00 p.m. and 7:00 p.m.

From Table 2 and Table 3, we can see that taxi ridership has the strongest positive relationship with theft, both on weekdays and on weekends. The correlation with phone users and residential population is only slightly lower. On weekends, however, the relationship between (statically measured) residential population and theft is even stronger than that between (dynamically measured) phone users and theft. Theft is strongly positively related to the proximity of offender residences but is weakly negatively related to distance from the nearest police station.

To reveal the spatial pattern of theft from the person, we constructed kernel density maps of theft from the person on weekdays and weekends (Figure 1). Across the five different periods, there are four stable hot spots (from I to IV; Figure 1K, Figure 1L). Spot I is the old town and it features the traditional business street. It is larger on weekends than on weekdays, because more people go out for shopping on weekends, which generates more suitable targets for motivated offenders. Spots II, III, and IV are so-called urban villages or villages in the city. These areas attract settlements of rural–urban migrant workers, resulting in high densities of low-income people (Zhang, Zhao, and Tian 2003). There is no obvious difference of Spots II, III, and IV on weekdays and weekends. The spatial variation is shown for each time period from Figure 1A to Figure 1L.

Negative Binominal Regression Results

For each of the five three-hour time periods, we estimated four separate models with residential population, subway ridership, taxi ridership, and phone users as alternative indicators of risk population. There were thus twenty (4×5) models for weekdays (see Table 4) and also twenty models for weekends (see Table 5). An alternative approach would have been to estimate a large number of models with all possible subsets of the four risk population indicators. We decided against this alternative for two reasons. First, our key focus is on comparing the performance between alternative indicators rather than maximizing model fit. Second, because the alternative would involve estimating fourteen rather than four models per day and time unit, the risk of chance capitalization would be strongly increased (MacCallum, Roznowski, and Necowitz 1992).

Comparison of the residual spatial autocorrelation after estimating the models with the residual

Table 1. Descriptive statistics for all three-hour periods, on weekdays and weekends (study unit: 1 km × 1 km, 205 grids)

Period	Variable	*M*	*SD*	Minimum	Maximum	Period	Variable	*M*	*SD*	Minimum	Maximum
All periods	Residential population	22.932	17.280	0.494	86.931	All periods	Offenders local	7.756	9.006	0.000	56.000
	Distance to police	0.322	0.388	0.000	1.700		Offenders nearby	57.605	35.963	0.000	156.000
Weekday 7:00 a.m.–10:00 a.m.	Theft from person	19.468	18.264	0.000	92.000	Weekend 7:00 a.m.–10:00 a.m.	Theft from person	6.961	6.792	0.000	35.000
	Subway ridership	30.228	30.150	0.000	179.226		Subway ridership	6.478	5.654	0.000	27.953
	Taxi ridership	3.357	3.994	0.000	18.797		Taxi ridership	1.150	1.208	0.000	7.668
	Phone users	14.624	9.758	0.934	42.779		Phone users	5.241	3.300	0.312	14.733
Weekday 10:00 a.m.–1:00 p.m.	Theft from person	30.298	39.367	0.000	387.000	Weekend 10:00 a.m.–1:00 p.m.	Theft from person	13.259	16.830	0.000	122.000
	Subway ridership	17.458	19.670	0.000	117.114		Subway ridership	8.915	9.747	0.000	64.207
	Taxi ridership	3.612	4.303	0.000	27.535		Taxi ridership	1.383	1.531	0.000	10.711
	Phone users	16.729	12.898	1.203	63.378		Phone users	6.528	4.616	0.433	25.552
Weekday 1:00 p.m.–4:00 p.m.	Theft from person	28.483	36.485	0.000	256.000	Weekend 1:00 p.m.–4:00 p.m.	Theft from person	15.244	23.293	0.000	172.000
	Subway ridership	20.706	24.235	0.000	157.402		Subway ridership	11.413	14.074	0.000	102.096
	Taxi ridership	3.298	3.721	0.000	23.005		Taxi ridership	1.289	1.389	0.000	8.734
	Phone users	16.482	13.034	1.140	71.341		Phone users	6.480	4.935	0.438	29.580
Weekday 4:00 p.m.–7:00 p.m.	Theft from person	33.380	41.520	0.000	266.000	Weekend 4:00 p.m.–7:00 p.m.	Theft from person	17.224	26.220	0.000	178.000
	Subway ridership	33.306	35.953	0.000	228.637		Subway ridership	12.454	14.303	0.000	100.433
	Taxi ridership	2.632	2.514	0.000	12.323		Taxi ridership	1.211	1.190	0.000	6.249
	Phone users	18.278	13.862	1.355	69.544		Phone users	6.881	5.094	0.497	28.786
Weekday 7:00 p.m.–10:00 p.m.	Theft from person	31.356	38.499	0.000	190.000	Weekend 7:00 p.m.–10:00 p.m.	Theft from person	13.859	18.252	0.000	87.000
	Subway ridership	17.290	18.442	0.000	114.954		Subway ridership	7.488	8.314	0.000	55.823
	Taxi ridership	3.515	3.487	0.000	19.450		Taxi ridership	1.387	1.389	0.000	7.893
	phone users	15.265	10.340	1.297	46.340		Phone users	5.951	4.012	0.501	17.250

Note: Residential population, subway ridership, taxi ridership, and phone users per 1,000. Distance per kilometer.

Table 2. Correlation coefficients, 7:00 a.m. through 10:00 p.m. on weekdays

	Theft	Residential population	Offenders nearby	Offenders local	Distance to police	Subway ridership	Taxi ridership	Phone users
Theft	1.000							
Residential population	0.693*	1.000						
Offenders local	0.512*	0.578*	1.000					
Offenders nearby	0.574*	0.509*	0.331*	1.000				
Distance to police	−0.362*	−0.405*	−0.314*	−0.219	1.000			
Subway ridership	0.612*	0.511*	0.451*	0.296*	−0.383*	1.000		
Taxi ridership	0.736*	0.567*	0.466*	0.395*	−0.387*	0.739*	1.000	
Phone users	0.724*	0.535*	0.467*	0.499*	−0.337*	0.670*	0.754*	1.000

*$p < 0.001$ (two-tailed).

spatial autocorrelation of a null model (i.e., spatial autocorrelation of the dependent variable) demonstrates that the Moran's *I* values decrease significantly after inclusion of the covariates and thus that the covariates account for a large part of the spatial autocorrelation.

Potential Offenders and Guardians

As predicted by the routine activity approach, indicators of the presence and proximity of motivated offenders have significant positive effects on theft from the person in most models, with the single exception of the 7:00 a.m. to 4:00 p.m. period on weekends. In virtually all models we also find a positive effect of lagged offender presence, except the models of weekends between 7:00 a.m. and 10:00 a.m., 10:00 a.m. and 1:00 p.m., and 1:00 p.m. and 4:00 p.m., which indicates that the effects of the proximity to motivated offenders spill over to adjacent grid cells.

In contrast to the expectation, the distance to the nearest police station negatively affects crime, with the two exceptions of subway ridership models between 7:00 p.m. and 10:00 p.m. on both weekdays and weekends. Thus, proximity to a police station appears to increase crime. Possible reasons for this finding are discussed in the concluding section.

Potential Targets

The results clearly demonstrate that, for all measures, for all three-hour periods and both on weekends and on weekdays, the number of potential targets is an important risk factor for theft from the person, irrespective of whether the measurement is based on residential population, taxi ridership, and phone users. The single exception appears to be subway ridership between 7:00 a.m. and 7:00 p.m. on weekdays and 7:00 a.m. and 4:00 p.m. on weekends, as this measure appears to be unrelated to theft.

Most of the lagged values of the ambient population indicators seem to have no significant impact on thefts from the person, except maybe for subway ridership during weekdays and some time periods on weekends.

Table 3. Correlation coefficients, 7:00 a.m. through 10:00 p.m. on weekends

	Theft	Residential population	Offenders local	Offenders nearby	Distance to police	Subway ridership	Taxi ridership	Phone users
Theft	1.000							
Residential population	0.652*	1.000						
Offenders local	0.475*	0.578*	1.000					
Offenders nearby	0.554*	0.509*	0.331*	1.000				
Distance to police	−0.325*	−0.405*	−0.314*	−0.219	1.000			
Subway ridership	0.621*	0.484*	0.424*	0.306*	−0.370*	1.000		
Taxi ridership	0.681*	0.602*	0.474*	0.430*	−0.393*	0.679*	1.000	
Phone users	0.634*	0.586*	0.475*	0.586*	−0.309*	0.443*	0.582*	1.000

*$p < 0.001$ (two-tailed).

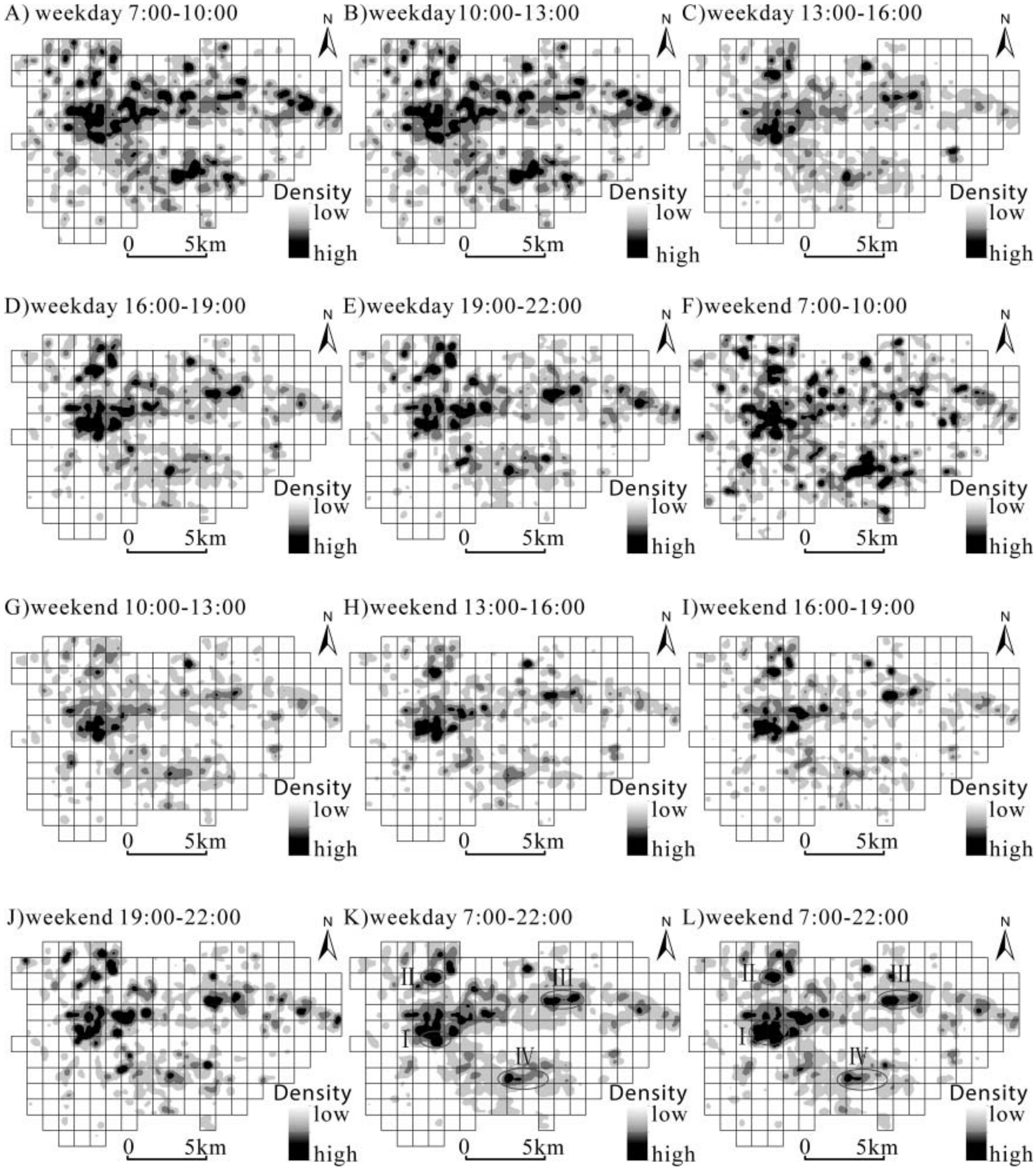

Figure 1. Kernel densities of theft from the person across five three-hour periods on weekdays and weekends.

This demonstrates that there are no or only very limited spatial spillover effects between the 1 km^2 grid cells.

Model Performance

The AIC value and bootstrap result are used as measures of relative model fit. Note that the outcomes of the two alternative ways of evaluation (either select model with lowest AIC value or most preferred bootstrap model) fully correspond with each other. Thus, of the four models being compared, the model with the lowest AIC value (e.g., residential model during the weekend from 7:00 a.m. to 10:00 a.m.) is consistently the model with the highest percentage in the bootstrap procedure (the taxi and phone models on weekdays between 7:00 a.m. and 10:00 p.m. can be considered ties, as they are virtually equal on both accounts).

In Table 4 (models for weekdays), except during the period from 7:00 a.m. to 10:00 a.m., when residential population is the best predictor, taxi ridership and

Table 4. Negative binominal regression models in different periods on weekday

	Weekday 7:00 a.m.–10:00 a.m.				Weekday 10:00 a.m.–1:00 p.m.				Weekday 1:00 p.m.–4:00 p.m.			
	Residential	Subway	Taxi	Phone user	Residential	Subway	Taxi	Phone user	Residential	Subway	Taxi	Phone user
Constant	1.683***	1.720***	1.807***	1.627***	1.968***	2.008***	2.087***	1.872***	2.071***	2.050***	2.056***	1.880***
Distance to police	−0.302*	−0.300*	−0.320*	−0.416**	−0.663***	−0.545***	−0.620***	−0.700***	−0.990***	−0.785***	−0.838***	−0.919***
Offenders local	0.032***	0.050***	0.049***	0.034***	0.035***	0.050***	0.044***	0.035***	0.034***	0.044***	0.035***	0.027***
Lag_offenders	0.005**	0.006***	0.006***	0.007***	0.005*	0.006***	0.007***	0.008***	0.006*	0.006**	0.007***	0.008***
Indicator of risk population	0.031***	0.003	0.115***	0.045***	0.024***	0.003	0.152***	0.036***	0.021***	0.007	0.194***	0.041***
Lag_risk population	−0.001	0.001*	−0.005	−0.001	0.001	0.003***	−0.006	0.000	0.001	0.002**	−0.009*	0.000
AIC	1464.92	1496.27	1469.56	1466.73	1639.08	1649.71	1608.87	1625.37	1628.84	1623.99	1586.13	1590.42
Bootstrap (n = 1,000)	41.4%	0.0%	29.9%	28.7%	2.7%	0.0%	70.2%	27.1%	0.2%	0.0%	55.3%	44.5%
Moran's I of TFP volume	0.415***	0.415***	0.415***	0.415***	0.379***	0.379***	0.379***	0.379***	0.403***	0.403***	0.403***	0.403***
Moran's I of residuals	0.166***	0.126***	0.122***	0.243***	0.098**	0.082**	0.059*	0.12***	0.112***	0.051	0.043	0.086**

	Weekday 4:00 p.m.–7:00 p.m.				Weekday 7:00 p.m.–10:00 p.m.			
	Residential	Subway	Taxi	Phone user	Residential	Subway	Taxi	Phone user
Constant	2.045***	2.019***	1.987***	1.890***	1.584***	1.559***	1.620***	1.338***
Distance to police	−0.717***	−0.586***	−0.649***	−0.702***	−0.384*	−0.245	−0.296*	−0.429**
Offenders local	0.037***	0.051***	0.043***	0.036***	0.042***	0.058***	0.047***	0.037***
Lag_offenders	0.006**	0.007***	0.007***	0.007***	0.009***	0.009***	0.009***	0.008***
Indicator of risk population	0.026***	0.005	0.240***	0.038***	0.031***	0.016**	0.198***	0.055***
Lag_risk population	0	0.001*	−0.006	0	0	0.002	−0.006	0.001
AIC	1671.85	1674.4	1643.22	1647.95	1631.44	1627.73	1596.04	1595.77
Bootstrap (n = 1,000)	1.3%	0.0%	64.0%	34.7%	1.1%	0.0%	49.3%	49.6%
Moran's I of TFP volume	0.462***	0.462***	0.462***	0.462***	0.422***	0.422***	0.422***	0.422***
Moran's I of residuals	0.159***	0.071*	0.094*	0.142***	0.245***	0.092**	0.091**	0.202***

Note: AIC = Akaike's information criterion; TFP = theft from the person. Bootstrap (n = 1,000) lists percentage of replications where the column is the most preferred model.
$^{*}p < 0.05$ (two-tailed, based on robust standard errors).
$^{**}p < 0.01$.
$^{***}p < 0.001$.

Table 5. Negative binominal regression models in different periods on weekends

	Weekend 7:00 a.m.–10 a.m.				Weekend 10:00 a.m.–1:00 p.m.				Weekend 1:00 p.m.–4:00 p.m.			
	Residential	Subway	Taxi	Phone user	Residential	Subway	Taxi	Phone user	Residential	Subway	Taxi	Phone user
Constant	0.967***	0.975***	1.044***	0.922***	1.168***	1.169***	1.289***	1.099***	1.215***	1.182***	1.210***	1.019***
Distance to police	−0.484**	−0.487**	−0.501**	−0.579***	−0.501**	−0.396*	−0.487**	−0.601***	−0.743***	−0.526**	−0.675***	−0.698***
Offenders local	0.040***	0.052***	0.047***	0.036***	0.037***	0.055***	0.046***	0.035***	0.041***	0.052***	0.039***	0.026**
Lag_offenders	0.003	0.004*	0.005**	0.006**	0.004	0.005**	0.006**	0.007**	0.005	0.006**	0.009***	0.009***
Indicator of risk population	0.025***	0.002	0.253***	0.097***	0.028***	0.007	0.395***	0.090***	0.020**	0.013*	0.570***	0.110***
Lag_risk population	−0.001	0.005	−0.015	−0.004	0.000	0.006***	−0.019	0.000	0.001	0.004**	−0.034*	0.000
AIC	1078.59	1103.083	1088.673	1086.575	1323.109	1336.441	1313.173	1322.862	1368.846	1361.164	1330.745	1338.326
Bootstrap ($n = 1{,}000$)	59.5%	0.0%	24.8%	15.7%	23.8%	0.5%	56.7%	19.0%	0.9%	0.8%	62.1%	36.2%
Moran's *I* of TFP volume	0.321***	0.321***	0.321***	0.321***	0.392***	0.392***	0.392***	0.392***	0.421***	0.421***	0.421***	0.421***
Moran's *I* of residuals	0.120***	0.132***	0.128***	0.182***	0.071*	0.063**	0.051*	0.134***	0.124***	0.078*	0.086**	0.164***

	Weekend 4:00s p.m.–7:00 p.m.				Weekend 7:00 p.m.–10:00 p.m.			
	Residential	Subway	Taxi	Phone user	Residential	Subway	Taxi	Phone user
Constant	1.336***	1.283***	1.303***	1.168***	0.939***	0.901***	1.005***	0.668***
Distance to police	−0.909***	−0.704***	−0.847***	−0.885***	−0.439*	−0.304	−0.414**	−0.462**
Offenders local	0.041***	0.053***	0.046***	0.032***	0.041***	0.056***	0.045***	0.031***
Lag_offenders	0.005*	0.006**	0.008***	0.007***	0.008***	0.008***	0.009***	0.008***
Indicator of risk population	0.023***	0.016**	0.484***	0.108***	0.028***	0.036***	0.509***	0.146***
Lag_risk population	0.001	0.003**	−0.013	0.000	0.000	0.003	−0.026*	0.001
AIC	1387.868	1374.826	1365.167	1350.823	1327.222	1315.65	1291.919	1282.174
Bootstrap ($n = 1{,}000$)	3.7%	1.8%	17.2%	77.3%	1.7%	0.4%	25.0%	72.9%
Moran's *I* of TFP volume	0.432***	0.432***	0.432***	0.432***	0.384***	0.384***	0.384***	0.384***
Moran's *I* of residuals	0.157***	0.079*	0.145***	0.175***	0.130***	0.030	0.074*	0.114**

Note: AIC = Akaike's information criterion; TFP = theft from the person. Bootstrap ($n = 1{,}000$) lists percentage of replications where the column is the most preferred model.
*$p < 0.05$ (two-tailed, based on robust standard errors).
**$p < 0.01$.
***$p < 0.001$.

phone users are better indicators than residential population and subway ridership. During the periods from 10:00 a.m. to 1:00 p.m., 1:00 p.m. to 4:00 p.m., and 4:00 p.m. to 7:00 p.m., taxi ridership proves to be a better indicator of suitable targets. Only in the evening, the period between 7:00 p.m. and 10:00 p.m., does phone usage outperform the other indicators.

Reviewing the results of the models for the weekend (Table 5), it can be concluded that they are very similar to those for weekdays. The residential population model performs best in the early morning (7:00 a.m.–10:00 a.m.), taxi ridership dominates the afternoon (10:00 a.m.–4:00 p.m.), and phone usage dominates the evening (4:00 p.m.–10:00 p.m.).

Conclusion and Discussion

Based on the routine activity perspective, and controlling for the presence and proximity of potential offenders and of guardians, this study focused on the third element of the crime triangle, the targets or victims. Our contribution aimed to identify the best indicator of populations at risk for theft from the person in a large city in China at different times of the day and different days of the week. We compared the predictive efficacy of four alternative measures (residential population, subway ridership, taxi ridership, and phone users).

As for the key issue of this study, with the exception of the early morning (7:00 a.m.–10:00 a.m.) when residential population is the best predictor both on weekdays and on weekends, taxi ridership and mobile phone users provide the best measures of the risk population, with mobile phone usage being the best indicator during the evening (7:00 p.m.–10:00 p.m.). This finding confirms that, due to daily mobility patterns, the size of the residential population fails to continuously reflect the ambient population. This finding is consistent with recent literature that speaks to the same issue (Hanaoka 2016; Mburu and Helbich 2016). In the morning, however, both on weekdays and during the weekend, residential population outperforms the other measures. Arguably, the morning is the period of the day when people are gradually starting to perform outdoor activities but when the majority of them, including the offenders, have not yet traveled far away from their area of residence, so that residential population fares relatively well as a measure of the population at risk. This finding demonstrates that a time-constant measure like residential population does not necessarily always perform worse than time-varying measures like taxi and subway ridership and mobile phone users and that time is very critical in the assessment of risk populations, an issue that has only recently been appreciated in the literature. In terms of subway ridership, it always ranked last in predicting theft from the person, mainly due to its sparse distribution across urban space and failure in measuring where the people were.

The contrasting effect of taxi ridership and phone users in determining risk is also very interesting but somewhat surprising. During the time slots 10:00 a.m. to 1:00 p.m., 1:00 p.m. to 4:00 p.m., and 4:00 p.m. to 7:00 p.m. on weekdays and 10:00 a.m. to 1:00 p.m. and 1:00 p.m. to 4:00 p.m. on weekends, taxi ridership performs better than mobile phone users. Arguably, during these periods most people are performing activities indoors, such as learning at school, working at the workplace, or performing household chores at home. As their mobile phone activity will be monitored both indoors and outdoors, the mobile phone measure becomes biased because it includes people who are staying indoors where they have low or nonexistent risks of becoming victims of theft from the person. Taxi ridership, however, is likely a better indicator of outdoor activities. In the evenings and nights, phone use performs better as a crime predictor because at those times of the day the general amount of outdoor activity increases. Thus, although mobile phone use is a superior measure of ambient population because almost everybody carries a mobile phone, the measure cannot distinguish between indoor and outdoor presence, corresponding with the concern of the algorithmic uncertainty of mobile phone data that are based on the cell tower (Kwan 2016). Taxi ridership, to the contrary, represents only a small proportion of the population, but because taxis are widely used in China, taxi use is common enough to be representative of outdoor activities that make people vulnerable and thus suitable targets for theft from the person. Thus, taxi ridership can be an important indicator of risk populations for outdoor crimes, whereas phone use data might be further refined to represent the people on the street.

Consistent with the predictions of the routine activity framework, our results further demonstrate that proximity to the homes of motivated offenders increases the volume of theft from the person. Distance to the police station had a negative effect, as reported by Helbich and Jokar Arsanjani (2015). Victims of theft from the person might be more likely to notify the police if victimization takes place near a police station. These findings show that it is important to control the effects of motivated offenders and capable guardianship while assessing optimal indicators of the population at risk.

In conclusion, the contributions of this study are not only that we have taken the proximity of motivated offenders and capable guardianship into account but more important that we considered changes of the optimal indicators of population at risk over the course of the day and the week. The findings give a better way to solve the uncertain geographic context problem in crime research, despite the fact that it still cannot delineate the spatial and temporal characteristics of the true geographic context to 100 percent. Besides, by adopting direct measurements of motivated offenders, potential targets, and capable guardianship, we confirmed the applicability of the routine activity framework in the Chinese urban context.

This study provides a deeper understanding of crime generators, criminogenic places where many people come together and therefore attract potential offenders. Our findings further the concept by considering the presence of people more precisely, in particular regarding the distinction between being indoors and outdoors. At a practical level, our findings provide new insights regarding the population at risk. Residential populations or ambient populations are not necessarily the targets of offenders. It depends on where people are, in particular whether they find themselves indoors or outdoors, in private or in public space. The presence of people changes with time. This implies that in allocating their resources in preventing theft from the person, the police could focus on the significance of mobility and outdoor activities. Crime prevention strategies of police should be based on victimization risk, which is dynamic over the course of the day and the week. Besides, the crime prevention strategy should be based not only on where potential targets are but also on where offenders live. Places closer to their residential addresses are more likely to be targeted. As police resources to prevent crimes are limited, the use of police strength can be more specific and more efficient with accurate identification of crime hot spots.

There are some limitations to this study, and to the interpretation of its findings, that should be highlighted here. First, although taxi ridership proved to be a good measure of the population at risk in the central district, it might be not a good indicator in the suburban areas of the city where the utilization of taxis is sparse. Phone use appears to be a good overall choice to measure the population at risk, but it needs adjustment to sort out indoor users who might not be at risk. By using the mobile phone data to assess the mobility of population, we can further estimate the volume of outdoor activities. In this study, two minor limitations are that we could only use data of a single mobile phone service company and that there is no 100 percent guarantee that mobile phones will always contact the closest station, which could negatively affect the precision of the geotracking data.

Second, the concept of capable guardianship could not be completely measured. Although Helbich found that the Euclidean distance to police stations has a significant negative impact on nonviolent crime (Helbich and Jokar Arsanjani 2015), distance to the police station might not be strongly related to the frequency, direction, and length of police patrols. In addition, guardianship is not limited to formal guardianship (police, monitor camera, etc.) but also includes informal guardianship (effects of "eyes on the streets," vigilant bystanders who prevent thefts, etc.), an element of guardianship not included in our study.

Third, due to limitations in the available data, we could not compare the indicators of risk populations for the whole day, so that we miss the nighttime between 10:00 p.m. and 7:00 a.m. In addition, only theft from the person is addressed in this study and other types of crime will possibly require other indicators. Regarding the modeling strategy, a purpose of future research could be trying to find an optimal combination of indicators to predict crime. In this study, we did not want to take a data-driven approach but preferred to focus on revealing the significance of mobility and outdoor activity in understanding the spatial and temporal pattern of crimes.

In closing, we conclude that this study has generated insights on how different indicators of risk population might contribute to the incidence of theft from the person during different periods of the day and the week. The insights gained have the potential to assist police departments in allocating their limited resources more effectively.

Acknowledgments

The authors thank Professor Mei-Po Kwan and three anonymous reviewers for their valuable comments and suggestions. Guangwen Song is grateful for the support of the China Scholarship Council (CSC). He conducted parts of the research while doing research internships at Vrije Universiteit Amsterdam and the Netherlands Institute for the Study of Crime and Law Enforcement. Lin Liu served as the corresponding author for this article.

Funding

This work was funded by the National Natural Science Key Foundation of China (No. 41531178), the Natural Science Foundation Research Team Project of Guangdong Province (No. 2014A030312010), and the Outstanding Youth Fund of the National Natural Science Foundation of China (No. 41522104).

ORCID

Lin Liu http://orcid.org/0000-0002-7202-3418
Wim Bernasco http://orcid.org/0000-0002-3385-0883

Notes

1. Four other categories include crime attractors, crime detractors, crime-neutral places, and fear generators, but these are not relevant for the argument presented here.
2. Access to crime data was granted by the police authorities on the condition that the real name of the city would not be mentioned in publications.
3. API is an acronym for application programming interface. Here, it indicates a set of definitions and routines that allows a computer program to access structured address information from a Web site.
4. Temporal autocorrelation occurs if observations that are nearby in linear time are positively or negatively correlated. In a regression context, it can lead to underestimated standard errors. Note that in this analytical design, where we aggregated TPF incidents of a full year, incidents observed in adjacent time periods of the day (say 1:00 p.m. to 4:00 p.m. and 4:00 p.m. to 7:00 p.m.) did not necessarily take place shortly after each other in linear time, as they could have taken place on any day during the year 2014. Thus, our dependent variable is not sequentially ordered in real time and therefore temporal autocorrelation does not affect the estimates or the standard errors of our estimates.

References

Andresen, M. A. 2006. Crime measures and the spatial analysis of criminal activity. *British Journal of Criminology* 46 (2):258–85.

———. 2007. Location quotients, ambient populations, and the spatial analysis of crime in Vancouver, Canada. *Environment and Planning A* 39 (10):2423–44.

———. 2011. The ambient population and crime analysis. *The Professional Geographer* 63 (2):193–212.

Andresen, M. A., and G. W. Jenion. 2010. Ambient populations and the calculation of crime rates and risk. *Security Journal* 23 (2):114–33.

Bagchi, M., and P. R. White. 2005. The potential of public transport smart card data. *Transport Policy* 12 (5):464–74.

Belsley, D. A. 1991. *Conditioning diagnostics: Collinearity and weak data in regression*. New York: Wiley.

Bernasco, W., and R. Block. 2011. Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors and offender anchor points. *Journal of Research in Crime and Delinquency* 48 (1):33–57.

Bernasco, W., and F. Luykx. 2003. Effects of attractiveness, opportunity and accessibility to burglars on residential burglary rates of urban neighborhoods. *Criminology* 41 (3):981–1001.

Bernasco, W., S. Ruiter, and R. Block. 2017. Do street robbery location choices vary over time of day or day of week? A test in Chicago. *Journal of Research in Crime and Delinquency* 54 (2):244–75.

Boessen, A., and J. R. Hipp. 2015. Close-ups and the scale of ecology: Land uses and the geography of social context and crime. *Criminology* 53 (3):399–426.

Boggs, S. L. 1965. Urban crime patterns. *American Sociological Review* 30 (6):899–908.

Bogomolov, A., B. Lepri, J. Staiano, N. Oliver, F. Pianesi, and A. Pentland. 2014. Once upon a crime: Towards crime prediction from demographics and mobile data. In *Proceedings of the 16th International Conference on Multimodal Interaction*, 427–34. Istanbul, Turkey: ACM.

Boivin, R., and M. Felson. 2017. Crimes by visitors versus crimes by residents: The influence of visitor inflows. *Journal of Quantitative Criminology* 4:1–16.

Brantingham, P. J., and P. L. Brantingham. 1995. Criminality of place: Crime generators and crime attractors. *European Journal of Criminal Policy and Research* 3 (3):5–26.

Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer.

Chen, C., J. Chen, and J. Barry. 2009. Diurnal pattern of transit ridership: A case study of the New York City subway system. *Journal of Transport Geography* 17 (3):176–86.

Chen, J., L. Liu, S. Zhou, L. Xiao, G. Song, and F. Ren. 2017. Modeling spatial effect in residential burglary: A case study from ZG City, China. *ISPRS International Journal of Geo-Information* 6 (5):138.

Chong, S. C., W. M. Y. Teoh, and Y. Qi. 2015. Comparing customer satisfaction with China mobile and China telecom services: An empirical study. *The Journal of Developing Areas* 49 (6):247–62.

Cohen, L. E., and M. Felson. 1979. Social change and crime rate trends: A routine activity approach. *American Sociological Review* 44 (4):588–608.

Cohen, L. E., R. L. Kaufman, and M. R. Gottfredson. 1985. Risk-based crime statistics: A forecasting comparison for burglary and auto theft. *Journal of Criminal Justice* 13 (5):445–57.

Cui, Y. 2013. A systematic approach to evaluate and validate the spatial accuracy of farmers market locations using multi-geocoding services. *Applied Geography* 41:87–95.

Dean, R. B., and W. J. Dixon. 1951. Simplified statistics for small numbers of observations. *Analytical Chemistry* 23 (4):636–38.

Feng, J., Y. Dong, and L. Song. 2016. A spatio-temporal analysis of urban crime in Beijing: Based on data for property crime. *Urban Studies* 53 (15):3223–45.

Gerber, M. S. 2014. Predicting crime using Twitter and kernel density estimation. *Decision Support Systems* 61:115–25.

Haberman, C. P., and J. H. Ratcliffe. 2015. Testing for temporally differentiated relationships among potentially criminogenic places and census block street robbery counts. *Criminology* 53 (3):457–83.

Hanaoka, K. 2016. New insights on relationships between street crimes and ambient population: Use of hourly population data estimated from mobile phone users' locations. *Environment and Planning B: Planning and Design*. Advance online publication. doi:10.1177/0265813516672454

Helbich, M., and J. Jokar Arsanjani. 2015. Spatial eigenvector filtering for spatiotemporal crime mapping and spatial crime analysis. *Cartography and Geographic Information Science* 42 (2):134–48.

Hilbe, J. M. 2011. *Negative binomial regression*. 2nd ed. Cambridge, UK: Cambridge University Press.

Hipp, J. R. 2016. General theory of spatial crime patterns. *Criminology* 54 (4):653–79.

Kinney, J. B., P. L. Brantingham, K. Wuschke, M. G. Kirk, and P. J. Brantingham. 2008. Crime attractors, generators and detractors: Land use and urban crime opportunities. *Built Environment* 34 (1):62–74.

Kwan, M.-P. 2012. The uncertain geographic context problem. *Annals of the Association of American Geographers* 102 (5):958–68.

———. 2016. Algorithmic geographies: Big data, algorithmic uncertainty, and the production of geographic knowledge. *Annals of the American Association of Geographers* 106 (2):274–82.

Li, X., G. Pan, Z. Wu, G. Qi, S. Li, D. Zhang, W. Zhang, and Z. Wang. 2012. Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science* 6 (1):111–21.

Liao, W., L. Liu, S. Zhou, G. Song, Q. Li, and K. Liu. Forthcoming. A model for optimizing Chinese addresses' geocoding results from multiple map APIs based on clustering and classifying. *Tropical Geography*.

Liu, L., and J. Eck. 2007. Analyzing of police vehicle stops in Cincinnati: A geographic perspective. *Geography Research Forum* 27:29–51.

Liu, L., and J. C. M. Li. 2017. Progress and future directions of crime research in China with selected case studies. *Journal of Research in Crime and Delinquency* 54 (4):447–53.

MacCallum, R. C., M. Roznowski, and L. B. Necowitz. 1992. Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin* 111 (3):490–504.

Malleson, N., and M. A. Andresen. 2015. Spatio-temporal crime hotspots and the ambient population. *Crime Science* 4 (1):10.

———. 2016. Exploring the impact of ambient population measures on London crime hotspots. *Journal of Criminal Justice* 46:52–63.

Mburu, L. W., and M. Helbich. 2016. Crime risk estimation with a commuter-harmonized ambient population. *Annals of the American Association of Geographers* 106 (4):804–18.

Messner, S. F., and J. R. Blau. 1987. Routine leisure activities and rates of crime: A macro-level analysis. *Social Forces* 65 (4):1034–52.

Osgood, W. 2000. Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology* 16 (1):21–43.

Pan, G., G. Qi, W. Zhaohui, D. Zhang, and S. Li. 2013. Land-use classification using taxi GPS traces. *IEEE Transactions on Intelligent Transportation Systems* 14 (1):113–23.

Pelletier, M.-P., M. Trépanier, and C. Morency. 2011. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies* 19 (4):557–68.

Peng, C., S. Xueming, Y. Hongyong, and L. Dengsheng. 2011. Assessing temporal and weather influences on property crime in Beijing, China. *Crime, Law and Social Change* 55 (1):1–13.

Qi, G., X. Li, S. Li, G. Pan, Z. Wang, and D. Zhang. 2011. Measuring social functions of city regions from large-scale taxi behaviors. In *Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 384–88. New York: IEEE Press.

Ratcliffe, J. H. 2004. Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science* 18 (1):61–72.

———. 2010. Crime mapping: Spatial and temporal challenges. In *Handbook of quantitative criminology*, ed. A. R. Piquero and D. Weisburd, 5–24. New York: Springer.

Ratti, C., D. Frenchman, R. M. Pulselli, and S. Williams. 2006. Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design* 33 (5):727–48.

Reynald, D. M. 2009. Guardianship in action: Developing a new tool for measurement. *Crime Prevention and Community Safety* 11 (1):1–20.

Stults, B. J., and M. Hasbrouck. 2015. The effect of commuting on city-level crime rates. *Journal of Quantitative Criminology* 31 (2):331–50.

Wang, X., M. S. Gerber, and D. E. Brown. 2012. Automatic crime prediction using events extracted from Twitter posts. In *Social computing, behavioral–cultural modeling and prediction*, ed. S. Yang, A. Greenberg, and M. Endsley, 231–38. Berlin: Springer.

Weisburd, D., W. Bernasco, and G. J. N. Bruinsma, eds. 2009. *Putting crime in its place: Units of analysis in geographic criminology*. New York: Springer.

Wilcox, P., N. Quisenberry, D. T. Cabrera, and S. Jones. 2004. Busy places and broken windows? Toward defining the role of physical structure and process in

community crime models. *Sociological Quarterly* 45 (2):185–207.
Xu, J. 2009. The robbery of motorcycle taxi drivers (Dake Zai) in China: A lifestyle/routine activity perspective and beyond. *British Journal of Criminology* 49 (4):491–512.
Zhang, L., S. F. Messner, and J. Liu. 2007. A multilevel analysis of the risk of household burglary in the city of Tianjin, China. *British Journal of Criminology* 47 (6):918–37.
Zhang, L., S. X. B. Zhao, and J. P. Tian. 2003. Self-help in housing and chengzhongcun in China's urbanization. *International Journal of Urban and Regional Research* 27 (4):912–37.

GUANGWEN SONG is a PhD Candidate in the School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, P.R. China. E-mail: songgwen@126.com. His research interests include spatial and temporal patterns of crime, offenders' target location choice, and human mobility.

LIN LIU is Professor of Geography, University of Cincinnati, and Center of Geo-Informatics for Public Security, School of Geographic Sciences, Guangzhou University, Guangzhou 510006, P.R. China. E-mail: lin.liu@uc.edu. His research interests include crime mapping and analysis, geographic information systems, and remote sensing.

WIM BERNASCO is Senior Researcher at the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands, and Professor in the Department of Spatial Economics at VU University Amsterdam, Leiden NL-3200, The Netherlands. E-mail: wbernasco@nscr.nl. His research interests focus on spatial aspects of crime, including offender travel behavior and target selection.

LUZI XIAO is a PhD Candidate in the School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, P.R. China. E-mail: xiaoluzi@mail2.sysu.edu.cn. Her research interests include spatiotemporal patterns of crime and crime location choice.

SUHONG ZHOU is Professor in the School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, P.R. China. E-mail: eeszsh@mail.sysu.edu.cn. Her research interests include spatial and temporal human behavior, urban planning, and crime.

WEIWEI LIAO is a Master's Candidate in the School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, P.R. China. E-mail: liaoww3@sina.com. Her research interests include crime prediction and spatiotemporal data mining.