Crime Science

RESEARCH

CrossMark

An information theory approach to hypothesis testing in criminological research

Gohar A. Petrossian[*] and Mike Maxfield

Abstract

Background: This research demonstrates how the Akaike information criterion (AIC) can be an alternative to null hypothesis significance testing in selecting best fitting models. It presents an example to illustrate how AIC can be used in this way.

Methods: Using data from Milwaukee, Wisconsin, we test models of place-based predictor variables on street robbery and commercial robbery. We build models to balance explanatory power and parsimony. Measures include the presence of different kinds of businesses, together with selected age groups and social disadvantage.

Results: Models including place-based measures of land use emerged as the best models among the set of tested models. These were superior to models that included measures of age and socioeconomic status. The best models for commercial and street robbery include three measures of ordinary businesses, liquor stores, and spatial lag.

Conclusions: Models based on information theory offer a useful alternative to significance testing when a strong theoretical framework guides the selection of model sets. Theoretically relevant 'ordinary businesses' have a greater influence on robbery than socioeconomic variables and most measures of discretionary businesses.

Keywords: Akaike information criterion, Information theory, Place and crime, Ordinary business

"A well-designed model is, after all, a judiciously chosen set of lies, or perhaps more accurately put, partial truths about reality, which have been chosen so as to permit us to reason more effectively about some issue than we otherwise could." (Baumol 1993, p. 55).

Background

Empirical criminological research relies heavily on testing null hypotheses of no difference. Rooted in statistical theory, decisions to reject a null hypothesis are keyed to finding statistically significant differences in relationships, or between outcome variables. Adopting conventions from previous research (Bushway et al. 2006; Sullivan and Mieczkowski 2008), we refer to this as null hypothesis significance testing (NHST). Despite its widespread use, researchers have identified a number of problems associated with the NHST approach as it is used in criminological research, and in other social sciences.

First is the reification of statistical significance as the most important outcome of quantitative research (Maltz 1994). Replicating analysis of papers published in the *American Economic Review* (McCloskey and Ziliak 1996; Ziliak and McCloskey 2004), Bushway and colleagues show that criminologists similarly more prominently report statistical significance than effect size. Second, scholars accept findings of no significance as evidence of no relationship (Weisburd et al. 2003), not always recognizing possible problems related to sample sizes, measurement error, or other features of research design. A related problem stems from modeling strategies when a large number of predictors are present. Third is the use of language such as "highly significant," "borderline significant," or "most significant," that mistakenly equates significance and effect size. Fourth, researchers with very large numbers of data points may find that all independent variables meet virtually any significance level in their relationship with dependent variables (Maltz 2006).

Setting aside these problems, NHST mandates a simplified approach to empirical research that assumes binary

*Correspondence: gpetrossian@jjay.cuny.edu
Department of Criminal Justice, John Jay College of Criminal Justice, 524 West 59th Street, New York, NY 10019, USA

increments to knowledge and often produces results of limited theoretical substance. Notably, the NHST requires that a researcher produce only one interesting research hypothesis and state the null. The research hypothesis, in essence, is never tested. Burnham and Anderson ask, "if there was little or no a priori belief in the null, what has been learned by its rejection?" (Burnham et al. 2011, p. 29).

This paper describes how an IT approach can guide selection of best-fitting statistical models. A key strength of this approach is its emphasis on testing a set of theory-based models against each other to identify the best among available models. What results is a more purposive comparison strategy in place of the somewhat arbitrary criterion of statistical significance, which plays virtually no role in AIC models.

We begin with a brief background discussion of an information theoretic approach that has become widely used in biology and psychology, but rarely guides criminological research. We then demonstrate the IT approach, using crime data from Milwaukee, Wisconsin, to examine how place-based measures of land use, together with measures of social disadvantage and age, are related to street robbery and commercial robbery. We use the Akaike information criterion (AIC) (Akaike 1973) to evaluate different models and aid in selecting the best models for two types of crime.

Akaike information criterion: a theoretical background

When building a theoretical model, information theorists posit that no model is a true model (Box 1976). This is largely because some percent of variance remains unexplained by all models. As such, any model built only approximates reality, or the unknown/unconstrained model.

However, Burnham and Anderson (2002) argue that it is possible to find the 'best approximation' to reality, or the distance between the unknown model and the model built to explain it, with a minimum loss of information. Kullback and Leibler (1951) developed a measure that became known as the Kullback–Leibler divergence, to represent this information loss associated with fitting a constraining model to the data.

Kullback and Liebler's (1951) paper quantified the meaning of "information", a concept related to Fisher's thinking about "sufficient statistics" (Burnham and Anderson 2004). Three decades later, Hirotugu Akaike's paper "Information Theory and an Extension of the Maximum Likelihood Principle" (Akaike 1973), proposed the Akaike information criterion (AIC), a method where Kullback–Leibler (K–L) divergence can be used to determine model suitability and selection.

The AIC approach computes goodness-of-fit (accuracy) and model variability (precision) to quantitatively rank different models in order to select the most *parsimonious* model (Saffron et al. 2006). Put somewhat differently, the AIC seeks to find "optimal complexity" (Garamszegi 2011, p. 2) by incorporating parsimony in model-selection. Among other things, this means that AIC model statistics are not defined for "full" models containing all possible variables.

Rooted in work by William of Occam (ca.1320), the parsimony principle states that the simplest competing description is the best (Anderson 2008; Saffron et al. 2006). Parsimony is used to determine how many parameters can be estimated and included to reach optimum model accuracy (Anderson 2008). Models with too few parameters are under-fitted and subject to bias due to the lack of information in the model. This is the familiar omitted-variables bias. Models with too many parameters are over-fitted and lack precision (McQuarrie and Tsai 1998; Burnham and Anderson 2002). Model selection, therefore, involves a trade-off between bias and variance, reflecting the statistical principle of parsimony (Burnham and Anderson 2004).

Models are often ranked based on conventional measures of goodness-of-fit, such as their R^2 values. Models that have increasing numbers of parameters end up with greater R^2 values, but at the expense of greater variability in how the model represents the data (Saffron et al. 2006). This is because every additional parameter captures a 'stochastic signal', and this decreased amount of information available for each calculation will lead to increased variation in parameter estimates (Rannala 2002; Lemmon and Moriarty 2004).

It may be argued that using adjusted R^2 value to report the fit of the model will achieve the same goal as AIC; the adjusted R^2 also has the penalty for each additional parameter when added to the model. However, Burnham and Anderson (2002) suggest that while adjusted R^2 values are useful as a measure of the proportion of explained variation in a model, these values should not be used for model selection and can be misleading (Burnham and Anderson 2002). Using an example of nine a priori models of avian species-accumulation curves from the Breeding Bird Survey (Flather 1996), they show that models with identical R^2 values of 0.99 had large differences in AIC values that yielded more precise statements about the "best" model (Burnham and Anderson 2002, p. 95). These comments also apply to measures such as pseuo-R^2, and others that center on proportion of variance explained.

The AIC includes a penalty for over-fitting the model, not allowing for an increase in the statistical bias when more parameters are fitted (Wilson et al. 2013). Another advantage of the AIC in model selection is that AIC is independent of the order in which models are computed (Anderson et al. 2001).

The Akaike information criterion, is calculated as

$$AIC = n\left[\ln\left(\frac{RSS}{n}\right)\right] + 2k \tag{1}$$

where n is the number of data points (i.e. sample size), *RSS* is the residual sum of squares, and k is the total number of estimated model parameters, which include both the model parameters and the constant.

Computationally the AIC is the sum of two so called "penalty terms" (Burnham and Anderson 2002), one for bias and one for uncertainty. This means that the smallest AIC values achieved among candidate models is deemed the preferred model. The addition of parameters will always increase the likelihood score, and this "penalty term" ensures that the over-parameterized model is not selected (Ripplinger and Sullivan 2008). In other words, models that have more fitted parameters will have higher AIC values, all other things being equal, and models that will be favored will be those with fewer parameters (Symonds and Moussalli 2011).

One of the strengths of building AIC models is the variety of methods that can be used to deal with model selection uncertainty (Garamszegi 2011). To compare models and determine relative support for each candidate model, several statistics can be calculated, which include the delta AIC (Δi), Akaike weights (wi) and evidence ratios.

Delta AIC (Δi) measures relative differences between a particular candidate model (*AICi)* and the Akaike 'best-ranked' model, the model with the smallest AIC value (*minAIC*). Delta AIC is used to evaluate relative support for other candidate models and is calculated as in Eq. 2.

$$\Delta i = AICi - minAIC \tag{2}$$

Burnham and Anderson (2002) suggest that models with $\Delta i \leq 2$ provide "substantial evidence" for the model, meaning these models are essentially as good as the best model. Models that have $4 \leq \Delta i \leq 7$ indicate "considerably less support" for the model, and $\Delta i > 10$ show that the model is "very unlikely" and should be rejected.[1]

It is important to note that AIC is a *relative* measure of how good a given model is among a candidate set of models, given the data. As such, even if essentially meaningless parameters or those that are poorly linked to the outcome variable are included, the AIC analysis will still produce a 'best' model among the candidate models examined.

Burnham et al. (2011) point out that such pitfalls can be avoided by *theory-based* selection of parameters.

Parsimony is a criterion for evaluating models with strong theoretical support, and is consistent with the goal of finding the best model among a set of possible models.

Akaike weights (Wi) are an essential next step after the AIC values for each proposed model have been calculated. These weights represent the ratio of delta AIC (Δi) value for each model relative to the whole set of candidate models (Burnham and Anderson 2002). The calculations of Akaike weights allow for an immediate ranking of all candidate models. Weights for the ith model in a set of R candidate models are calculated as shown in Eq. 3,

$$Wi = \frac{\exp\left(-\frac{1}{2\Delta i}\right)}{\sum_{r=1}^{R} \exp\left(-\frac{1}{2\Delta r}\right)} \tag{3}$$

where the denominator is simply the sum of the relative likelihoods for all candidate models. *Wi* is interpreted as the probability that the model is the Akaike 'best-ranked' among the set of candidate models (Burnham and Anderson 2002). For example, an Akaike weight of 0.80 for a given model indicates that this model has an 80% chance of being the Akaike 'best-ranked' model among the set of candidate models.

Lastly, Akaike weights can be used to determine the extent to which the 'best' model is better than other candidate models, expressed as evidence ratios:

$$\text{Evidence Ratio} = \frac{Wj}{Wi} \tag{4}$$

Equation 4 compares model *Wj* against model *Wi*, and any calculated value is interpreted such that model *j* is X times more likely than model *i* to be the 'best' in the set (Burnham and Anderson 2002). For example, an evidence ratio of 4 indicates that model *j* is four times better that model *i*. Evidence ratios allow researchers to express how much better the 'best' approximating model (or any given model in the set) is compared to the next best model or other models in the set (Symonds and Moussalli 2011). Evidence ratios can also be calculated relative to models other than the 'best' model, providing more evidence for the relative strength of all candidate models (Anderson 2008).

The calculation of Akaike weights across all models allows the researcher to evaluate the relative importance of many potential predictor variables within these models (Burnham and Anderson 2002). In fact, Lukacz et al. (2007) argue that model weights and their ability to account explicitly for model uncertainty are major reasons why IT approaches should be highly favored over NHST (Richards et al. 2011).

Other model selection approaches have been developed that aim at achieving the same goal as the Akaike

---

[1] In some instances, several models may compete for the 'best' model rank, as their Δj or evidence ratios are < 2. In this case, model-average estimates can be calculated, as well as the precision of these estimates. For more information, see Burnham and Anderson (2002).

information criterion: to identify the most parsimonious and theoretically relevant models. These approaches rely on different model selection strategies and use different criteria to evaluate model fit relative to its complexity. This diverse list of models includes Mallow's Cp method (Mallows 1973), Bayes information criterion (Schwartz 1978), Takeuchi's information criterion (Takeuchi 1976), generalized information criterion (Rao and Wu 1989), among others. The Akaike information criterion, however, has been receiving considerable attention in recent years (Garamszegi 2011). Many fields in behavioral, as well as life sciences, such as astronomy, cosmology, nuclear and particle science, medical physics, ecology, statistics and psychology, engineering and computer science, have turned to Akaike information theory to model relationships.

Using AIC in criminal justice research

Scholars in other disciplines have been quicker to recognize the limits and common misinterpretation of *p* values in significance testing. A statement by the American Statistical Association (Wasserstein and Lazar 2016) lists and summarizes many of these objections.

Analyzing very large numbers of cases with the NHST approach produces a type of parsimony problem that is common in criminological research. When very small effect sizes are reported as statistically significant, models can include coefficients that contribute little to the substantive understanding of research questions. For example, in their analysis of state sentences applied to convicted offenders in Florida, Feldmeyer et al. (2015) analyze 501,027 cases accumulated over a 7-year period. Each of 19 independent variables predicting a prison sentence is significant. Not surprisingly, this produces odds ratios that, in many cases, are not much different from 1.0.

Examples of sensitivity to the limits of NHST are emerging in criminological research. In their analysis of about 470,000 Pennsylvania defendants over seven years, Steffensmeier et al. (2016, p. 10) acknowledge that statistical significance is virtually assured: "As such, we place more emphasis on the direction and magnitude of the coefficients than on statistical significance….". Similarly, Bernasco et al. (2017) avoid discussing statistical significance in their analysis of the combined effects of time and types of places on robberies in 24,594 census blocks in Chicago. Instead, they examine how odds ratios bracketed by standard errors depart from 1.0 for different 2-h intervals within types of places. Using AIC-based models offers a tool for systematically assessing the relative importance of models irrespective of sample size.

A related phenomenon is that with many cases, more variables can be added, something that is sometimes done with minimal justification. Controlling for measures of social well-being, socioeconomic status, social disadvantage, known risk factors, and the like is the norm. This is partly because previous research includes such concepts, often with minimal theoretical justification. In any event, producing multiple models with staged introduction of predictor and control variables implicitly treats all as equally important or unimportant until proven otherwise.

Sullivan and Mieczkowski (2008) summarize how a Bayesian approach can be an alternative to NHST in applied criminal justice research. They describe an example that sequences research sites in a series of intensive probation experiments. Three sites are time-ordered, so that data collected from later sites draw on results for data from earlier sites in a cumulative analysis that "learns" from prior evidence. This contrasts with a NHST approach that would pool data from all three sites.

The most directly relevant example in criminology is Petrossian's (2015) analysis of illegal, unreported, and unregulated fishing in the waters of 53 countries. Her analysis of AIC values for models combining situational variables concluded that the best model included all predictor variables, rather than selected subsets. It's noteworthy that this analysis was published in *Biological Conservation*, a journal in which IT-based model selection is routine.

These examples notwithstanding, we are not aware of criminological research that uses an AIC approach to evaluate alternative theory-based models among a set of candidate models.[2] To illustrate how the AIC can be used, we examine how features of places are related to the distribution of two types of crime in Milwaukee, Wisconsin.

Crime and place

Criminological research has increasingly examined links between crime and place. The framework is theoretically rich, drawing on opportunity, crime pattern, and routine activity theories. That crime is concentrated at places, usually a small number of places, has been consistently demonstrated in a number of different cities. Weisburd (2015) offers a recent and comprehensive analysis showing this, to support his call for a new criminology of place. As noted by Weisburd (2015), and by Haberman and Ratcliffe (2015), empirical research has widely supported theoretical expectations about crime and place. Lee et al. (2017) present a systematic review showing the consistent links between crime and place.

---

[2] Karlis and Meligkotsidou (2007) include AIC and BIC in their comparison of different distributions of crime counts, but do not link their analysis to criminological theory.

An important example is research on how the presence of different kinds of businesses and facilities is related to crime patterns. Block and Block (1995) examined the presence of taverns and liquor stores near crime hotspots in Chicago. Bars and liquor stores are examples of crime attractors (Brantingham and Brantingham 1995), and have been the focus of much research on links between land use and crime (Groff 2014; Pridemore and Grubesic 2013; Gruenewald et al. 2006). Other types of undesirable but legal places, such as pawnshops, check cashing facilities and nightclubs, have also been examined in several cities. Such places are often referred to as "criminogenic," (Bernasco and Block 2011; Groff and Lockwood 2014; Haberman and Ratcliffe 2015) unpopular, or troublesome (Wilcox and Eck 2011).

Less common is research on how the presence of ordinary businesses and facilities is related to crime at places. An example is the analysis of robbery in Chicago by Bernasco and Block (2011). They describe how concentrations of businesses based mostly on small cash transactions (fast-food restaurants, grocery stores, barber and beauty shops) are associated with crime hot spots, in addition to such places as vice markets, bars, and pawnshops. Analyzing about 24,600 census blocks in Chicago, all facility types were significantly related to robbery. Haberman and Ratcliffe (2015) focus mostly on criminogenic places, but recognize how the kinds of facilities regularly used by large numbers of people can increase crime risks by serving as crime generators. Such places include corner stores, fast-food restaurants, ATMs, and mass transit stations.

Building on Haberman and Ratcliffe (2015), Bernasco et al. (2017) combine measures of place types with time of day and day of week to assess whether robbery increases for specific combinations of places and times in Chicago. They find little temporal variation except for the presence of high schools, and that robbery is higher in census blocks with a variety of small-scale retail places not normally viewed as criminogenic, such as restaurants, grocery stores, gas stations, and laundromats. Yu and Maxfield (2014, p. 314) similarly find that businesses, such as grocery stores, beauty parlors, and business services, are associated with higher rates of commercial and residential burglary. Their analysis concludes with discussion of different mechanisms at work in associations between the presence of ordinary businesses and burglary risk.

Our analysis builds on this research, and what Yu and Maxfield term "ordinary businesses." Unlike bars, liquor stores, pawnshops and the like, ordinary businesses are places that most people visit on a regular basis. Through such routines, "...innocuous or ordinary places play a role in exposing targets to an offender population." (Yu and Maxfield 2014, p. 314). Like Bernasco et al. (2017), and Haberman and Ratcliffe (2015), we examine robbery. Unlike previous research, we distinguish robbery of commercial places and street robbery, expecting that the presence of different kinds of facilities and businesses will be differently related to each type of robbery. The distinction is important, because commercial robberies target fixed places, while the victims of street robberies can be more mobile. It is possible that certain types of commercial places are more attractive targets of robbery. Similarly, street robbery victims may be targets because they visit certain types of establishments, or because they are on the street, visiting ordinary businesses.

Crime and place serves as a useful example to demonstrate the AIC approach to inference for two reasons. First is the strong theoretical and empirical framework that has been built up around crime and place. Bernasco et al. (2017) cite rational choice, routine activity, crime pattern theories and the geography of crime as complementary theoretical frameworks in understanding links between place and crime. Second, the role of ordinary businesses is inherently place-based, and the effects of ordinary businesses can be systematically compared to the effects of businesses described as criminogenic. Such specific theoretical expectations are best tested by an IT approach that evaluates different combinations of variables within a set of place types.

Because theories of place are comprehensive and have accumulated empirical support, the theoretical mechanisms at work are especially well-suited for comparing alternative models of robbery. Our analysis focuses on selecting the best among sets of models for commercial and street robbery. We then compare the AIC-ranked best models to models that include all variables under study.

Methods

Study site

Milwaukee is the 31st largest city in the United States, with a 2010 population of about 594,000. About 61% of the Milwaukee population is white, followed by 27% African American, and 3% Asian, with the remaining 9% comprising other races (American Community Survey 2013). As of 2013, Milwaukee ranks the 7th most dangerous city in America, with a violent crime rate of 587.1 per 100,000 people (FBI 2013).

Units of analysis

Considering the units of analysis that accurately capture the social process under investigation is an important first step in spatial analysis (Johnson et al. 2009). After examining the distribution and number of businesses in Milwaukee, as well as the overall distribution of the

crimes under investigation, we found the census tract level (N = 224) to be most appropriate.

We initially considered census blocks, but analyses revealed that about 90% of the census blocks remained unpopulated by the types of businesses examined here. Because drug stores, grocery stores, service stations and the like are common, we suspect their absence in the vast majority of census blocks reflects patterns of settlement in smaller Midwestern US cities like Milwaukee. Most research using census blocks has been conducted in larger, denser places like Chicago (Bernasco and Block 2011) or Philadelphia (Groff and Lockwood 2014; Haberman and Ratcliffe 2015). Moreover, past research has used census tracts as units of analysis to examine densities of businesses and violent crimes (e.g., Gruenewald et al. 2006; Livingston 2008; Zhu et al. 2004).

Data sources
Outcome variables
We obtained 2009 data on all crimes reported to police from the Milwaukee Police Department. Each record included the National Incident-Based Reporting System (NIBRS) code, address, time and date of the offense, type of location, and type of weapon(s) used. We selected *commercial robberies* and *street robberies* for further analyses. The Police Department provided the data in ArcGIS shapefile format, therefore, no further manipulations were necessary (such as geocoding addresses) to display the crime locations in ArcGIS.

Predictor variables
We used two sources to extract data for the predictor variables in this study. Data on demographic predictors aggregated at the census tract level, specifically, *percent below poverty, percent renter occupied, percent age 18–21*, and *percent age 22–29*, were obtained from the US Census Bureau (US Census 2000).

In this study, we distinguish between what we call discretionary places and ordinary places. Discretionary places are those that most people can choose whether to visit or not in the course of their normal activity. These include *drinking places*, *liquor stores*, and *places of amusement/recreation.* In contrast, ordinary places are businesses that most people patronize on a regular basis: *drug stores, grocery stores,* and *service (petrol) stations.*

Milwaukee data for the year 2009 were obtained from Infogroup, a company that provides data on businesses in the United States disaggregated by National Industry Classification codes. Infogroup's database contains information about all registered businesses in the United States, and includes such details as business address, size, sales volume, number of employees, type of industry under which the business is registered and the business's exact XY coordinate based on its registered address. The company contacts over 100,000 businesses daily (nationally) to verify the quality of the data in their database, as well as to ensure that the data are as current as possible (Infogroup 2015).

Data preparation
Demographic data in the form of ArcGIS shapefiles were directly downloaded from the US Census Bureau. The shapefiles were projected to match the projected coordinate system of the shapefiles containing data on crimes in Milwaukee. Crimes were then aggregated to 224 census tracts by spatially joining them to these tracts based on their location.

We used the XY coordinate information available in the Infogroup database to geocode the addresses of Milwaukee businesses used in the current study. Geocoding yielded a 100% match. We used the 'clip' tool in ArcGIS to select only the businesses that fell within the city boundary. We then aggregated these businesses to the 224 census tracts by spatially joining them to the census tracts. Table 1 presents descriptive statistics on businesses, crimes per census tract, age group, and social disadvantage.

Controlling for spatial autocorrelation
Spatial autocorrelation violates one of the important assumptions of traditional statistics-independence of observations. We found that spatial autocorrelation was present for each crime type.[3] As a result, we created spatial lags to represent the average values for neighboring areas (Anselin 2003), which can be either determined as those bordering the target census tract or those calculated based on a fixed distance from the centroid of the target census tract. In this research, we computed spatial lag based on the k-nearest neighbor method as the distance weight.

Multiple working hypotheses
This research considered two groups of theories: those based on traditional explanations of crime: the age-crime curve and social disadvantage; and those based on environmental criminology. The proposed hypotheses representing each model used in the analyses are listed in Table 2.

We use AIC models to test the empirical evidence for each of the hypotheses listed in Table 2 relative to the others in the set. In other words, each of these theoretically built models, which are considered *a priori*, are

---

[3] For *street robberies*—Moran's $I = 0.38$, $z = 18.40$, $p < 0.001$; *commercial robberies*—Moran's $I = 0.17$, $z = 8.29$, $p < 0.001$.

Table 1 Descriptive statistics for the variables at the census tract level

Variables	N per census tract		Mean	SD	Percent > 0
	Min	Max			
Business category					
Drinking places	0	18	1.84	2.60	67.9
Grocery stores	0	7	1.42	1.47	67.4
Service stations	0	5	0.80	1.04	49.6
Amusement/recreation	0	9	0.67	1.14	42.4
Liquor stores	0	3	0.44	0.62	37.5
Drug stores	0	4	0.35	0.68	25.9
Crime type					
Street robberies	0	34	8.67	8.07	88.8
Commercial robberies	0	36	5.43	4.72	90.2
Age and social disadvantage					
Percent age 18–21	2	76	7.18	8.21	100
Percent age 22–29	5	43	12.54	6.22	100
Percent renter occupied	0	96.5	43.63	21.39	99.6
Percent below poverty	1.9	97.5	23.58	16.30	100

Units = 224 census tracts

The extreme maximum values for the two age groups are census tracts that include Marquette University and the University of Wisconsin, Milwaukee

tested against the other competing models to evaluate their strength relative to their competitors.

Analyses and results

Steps to evaluating the models

Different modifications of AIC include AICc (or AIC corrected), QAIC (or quasi-AIC) and QAICc (see Symonds and Moussali 2011, for more information). To evaluate the fit of our models, we first determined which of these modifications of the AIC was most appropriate. We concluded that AICc is most appropriate given the small sample size (Anderson 2008). We proceeded to the following steps to estimate the models for each crime type using GLM (identity link function) and their associated AICc scores. These steps are shown in respective columns in Table 3.

A. Calculated the small sample corrected AIC (AICc) by (column 1)

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

where k is the total number of predictors in the model (including the constant and error), and n is the sample size.

B. Ranked the models from lowest to highest based on the AICc values. (Column 1)

C. Calculated the difference between the model with the lowest AICc and others in the set (i.e. Δi) by (column 2)

$$\Delta i = AICc_i - AICc_{min}$$

D. Calculated *relative likelihood* to evaluate the plausibility of each model by (column 3)

$$\mathcal{L}(g_i|y) \propto \exp\left(-\frac{1}{-2}\Delta i\right)$$

E. Calculated the *Akaike weights* for each model to normalize the relative likelihood values by (column 4)

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta i\right)}{\sum_{r=1}^{R} \exp\left(-\frac{1}{2}\Delta r\right)}$$

Table 2 Research hypotheses and constructs they measure

Construct	Hypotheses
Discretionary places	H1. *Street robberies* in Milwaukee are likely to be clustered around drinking places (DP), liquor stores (LS) and places of amusement/recreation (AR) H2. *Commercial robberies* in Milwaukee are likely to be clustered around drinking places (DP), liquor stores (LS) and places of amusement/recreation (AR)
Ordinary places	H3. *Street robberies* in Milwaukee are likely to be clustered around drug stores (DS), grocery stores (GS) and service stations (SS) H4. *Commercial robberies* in Milwaukee are likely to be clustered around drug stores (DS), grocery stores (GS) and service stations (SS)
Age-crime curve	H5. *Street robberies* in Milwaukee are likely to be clustered in census tracts with higher percent of the population between ages 18–21 and ages 22–29 H6. *Commercial robberies* in Milwaukee are likely to be clustered in census tracts with higher percent of the population between ages 18–21 and ages 22–29
Social disadvantage	H7. *Street robberies* in Milwaukee are likely to be clustered in census tracts with higher percent of the population below poverty and renter occupied H8. *Commercial robberies* in Milwaukee are likely to be clustered in census tracts with higher percent of the population below poverty and renter occupied

Table 3 Ranked models including all variables for commercial robbery

Models	AICc	ΔAIC	Relative likelihood of the model	Akaike weights	How much better is the first model compared to the competing models?
Discretionary + ordinary + SL	1118.96	0	1	0.78	
Ordinary + SL	1121.73	2.77	0.25	0.2	3.99
Ordinary + age + SL	1125.8	6.84	0.03	0.03	30.53
Discretionary +SL	1161	42.04	0	0	–
Discretionary + age + SL	1165.13	46.16	–	–	–
Discretionary + ordinary	1165.21	46.25	–	–	–
Spatial lag	1168	49.04	–	–	–
Ordinary	1174.16	55.2	–	–	–
Ordinary + age	1178.24	59.28	–	–	–
Discretionary	1217.48	98.52	–	–	–
Discretionary + age	1221.03	102.06	–	–	–
Age + social disadvantage	1224.03	105.07	–	–	–
Social disadvantage	1224.28	105.32	–	–	–
Intercept-only	1228.99	110.03	–	–	–
Age	1231.57	112.61	–	–	–

SL spatial lag

Results for commercial robbery (all variables)

Table 3 shows the results for commercial robberies. It lists all the models that test the theories in separate sets together with models that include different theoretical combinations (e.g. the model that combines the discretionary and ordinary variables). Models that include place types, age groups, and social disadvantage are also shown. Additionally, all theoretically built models are compared against the intercept-only model to determine if the predictor variables have merit when compared against the latter.

The columns in Table 3 correspond to the steps discussed above. Column 1 ranks each model using AICc. Here, based on the AICc value, the first model containing both the discretionary and ordinary variables, together with spatial lag, has been identified as the model most justified by data, also referred to as the AIC 'best-ranked' model. Akaike weights (column 4) show the weight of evidence that any given model is a plausible approximation given the data and the set of candidate models.

As indicated by the Delta AICc (column 2) and the relative likelihoods, the model that includes both discretionary and ordinary variables (plus *spatial lag*) was identified as having a 78% likelihood (column 4) of being the Akaike 'best-ranked' among the set. No other models were identified as strong competing candidates. The 'best' model is four times better than the second-ranked and 30 times better than the third-ranked model.

Results for commercial robbery (unpacked models)

To further examine whether we can discard models with uninformative parameters, we created the so-called unpacked models. Similar to Fondell et al. (2008), we retained only the AIC 'best-ranked' model from the previous step. We then considered a new set of models to determine if we could eliminate the least important parameters. Unpacked models consider individual business types within the grouped discretionary and ordinary categories. In this way, the set of all models considered includes different mixes of business types, based on the AIC 'best-ranked' model shown in Table 3. Results for the unpacked models are shown in Table 4.

The model that includes all ordinary business types, plus *liquor stores* and *spatial lag*, was identified as the AIC 'best-ranked' model. The Akaike weights indicate that this new model has an 83% likelihood of being the Akaike 'best-ranked' among the set, with no other models showing as possible strong candidates. The AIC 'best-ranked' mode is six times better than the second best model. Apart from these models, the remaining models are highly unlikely.

Results for street robbery (all variables)

Table 5 shows the results for street robberies. Similar to commercial robbery, we consider the theoretically constructed models separately, as well as in combination. The intercept-only model is included in this set as well.

Like results for commercial robbery, the model that includes both discretionary and ordinary variables (plus

Table 4 Ranked unpacked models for commercial robbery

Models	AICc	ΔAIC	Relative likelihood of the model	Akaike weights	How much better is the first model compared to the competing models?
(DS, SS, GS) LS, SL	1115.33	0	1	0.83	
Discretionary + ordinary + SL	1118.96	3.63	0.16	0.14	6.14
(DS, SS, GS) AR, SL	1123.1	7.77	0.02	0.02	48.71
(DS, SS, GS) DP, SL	1123.79	8.46	0.01	0.01	68.6
(DP, LS, AR) SS, SL	1131.14	15.81	0	0	2710.07
(DP, LS, AR) DS, SL	1144.92	29.59	–	–	–
(DP, LS, AR) GS, SL	1148.44	33.11	–	–	–
Intercept-only	1228.99	113.66	–	–	–

DS drug stores, *SS* service stations, *GS* grocery stores, *LS* liquor stores, *DP* drinking places, *AR* amusement/recreation, *SL* spatial lag

Table 5 Ranked models including all variables for street robbery

Models	AICc	ΔAIC	Relative likelihood of the model	Akaike weights	How much better is the first model compared to the competing models?
Discretionary + ordinary + SL	1275.28	0	1	0.47	
Ordinary + age + SL	1276.41	1.13	0.57	0.27	1.76
Ordinary + SL	1276.48	1.2	0.55	0.26	1.82
Discretionary + SL	1298.04	22.76	0	0	87,650.07
Discretionary + age + SL	1299.49	24.21	–	–	–
Spatial lag	1301.88	26.6	–	–	–
Age + social disadvantage	1389.66	114.38	–	–	–
Discretionary + ordinary	1392.81	117.53	–	–	–
Social disadvantage	1399.84	124.56	–	–	–
Ordinary	1399.92	124.64	–	–	–
Ordinary + age	1401.34	126.06	–	–	–
Discretionary	1430.9	155.62	–	–	–
Discretionary + age	1432.15	156.87	–	–	–
Intercept-only	1444.33	169.05	–	–	–
Age	1446.44	171.16	–	–	–

SL spatial lag

spatial lag) has a 47% likelihood of being the Akaike 'best-ranked' among the set. Two other models are candidates because Delta AICc are < 2. However, the 'best' model is almost two times better than the other competing models.

Because *age* was included in the second-best model, we added *age* to unpacked models in a separate analysis (not shown). Results indicated that the unpacked models that included *age* were not better than those with land-use variables only. In the interest of parsimony, we do not report the results of these unpacked model sets. Apart from these two competing models, the remaining models are highly unlikely.

Results for street robbery (unpacked models)

Similar to commercial robbery, we built unpacked models for street robbery. The results for the unpacked models are shown in Table 6.

As shown in Table 6, the model that includes all ordinary variables, plus *liquor stores* and *spatial lag*, has a 67% likelihood of being the Akaike 'best-ranked' model among the set, with no other models showing as possible strong candidates. The AIC 'best-ranked' mode is four and seven times better than the second and third best models, respectively. The remaining models have little support, producing results identical to those for commercial robbery.

Table 6 Ranked unpacked models for street robbery

Models	AICc	Δ AIC	Relative likelihood of the model	Akaike weights	How much better is the first model compared to the competing models?
(DS, SS, GS) LS, SL	1272.44	0	1	0.67	
Discretionary + ordinary + SL	1275.28	2.84	0.24	0.16	4.13
(DP, LS, AR) GS, SL	1276.46	4.01	0.13	0.09	7.44
(DS, SS, GS) AR, SL	1278.1	5.65	0.06	0.04	16.89
(DS, SS, GS) DP, SL	1278.49	6.05	0.05	0.03	20.57
(DP, LS, AR) SS, SL	1292.96	20.51	0	0	28,480.36
(DP, LS, AR) DS, SL	1295.51	23.06	–	–	–
Intercept-only	1444.33	171.88	–	–	–

DS drug stores, *SS* service stations, *GS* grocery stores, *LS* liquor stores, *DP* drinking places, *AR* amusement/recreation, *SL* spatial lag

Negative binomial regression results

Anderson (2008, p. 68) suggests that after the 'best-ranked' model has been identified, it is useful to assess the Akaike 'best-ranked' model using a goodness-of-fit test, such as residual analysis, R^2 or similar approaches. However, he cautions that these tests should be treated as descriptive statistics and run as *post hoc* tests only after the 'best-ranked' models have been identified.

Table 7 presents the negative binomial regression coefficients for variables in the models identified as the Akaike 'best-ranked' models. The final unpacked models for both robbery types identified as 'best', included all ordinary variables + *liquor stores + spatial lag.* As shown in the bottom panel of Table 7, adding all variables evaluated in the AIC analysis increments the pseudo-R^2 by only about 0.01 over that for the 'best' models (top panel).

Discussion and conclusions

Using the AIC to guide theory-based model selection, we find that the best models include mostly ordinary businesses, and one type of what we have termed "discretionary business."

Summary and discussion

If we had followed the traditional NHST approach, our analysis would look more like what is presented in the second panel of Table 7. That tacitly assumes place-based and socioeconomic variables are equally important. A traditional NHST analysis would cite theories of social disorganization or disadvantage and place-based theories as possible explanations of mechanisms related to the risk of robbery. Then measures, such as those shown in Table 7, would be included in successive models that are evaluated by assumptions about whether coefficients are statistically different from zero (Berk et al. 2010).

The information theoretic approach shown in Tables 3, 4, 5 and 6 and summarized in the top panel of Table 7 offers two insights. First, the best models for each type of robbery include ordinary and discretionary businesses and spatial lag (Tables 3, 5). Adding measures for two younger age groups and two measures of social disadvantage increases explanatory power, but not by enough to justify complicating the models when parsimony is considered. This claim is supported by the basic AIC modeling approach, in which easily computed changes in AIC from adding successive terms to a model balance added explanatory power against the number of terms in the model. In this sense, the AIC and related statistics express "criminological significance" rather than statistical significance.

Second, after unpacking models that included all types of ordinary and discretionary businesses, ordinary businesses plus liquor stores and spatial lag are the best models among those examined in Tables 4 and 6. Apart from liquor stores, the presence of discretionary businesses has no impact on commercial or personal robbery. Setting aside the models containing all "significant" variables allows us to focus more attention to the implied mechanisms at work in more parsimonious models.

Our expectations about possible differences in the effects of places by type of robbery were not supported. Both commercial and personal robberies are found in areas with a variety of businesses, most of them what we have called "ordinary". Drug stores, grocery stores, service stations, and liquor stores could be the targets of commercial robbery. For street robbery, it is likely that people visit these common places on a regular basis, thus exposing themselves to risk.

A substantive interpretation of the consistent impact of spatial lag is that robberies happen near other areas with robberies, a type of risk heterogeneity. This is consistent with recent work by Bernasco et al. (2017), suggesting that robbers work in fairly stable places where targets are to be found. These researchers also point to the role

Table 7 Negative binomial regression results for final (unpacked) models

Models	Street robbery		Commercial robbery	
Unpacked	**B**	**IRR**	**B**	**IRR**
Drug stores	0.021 *(0.072)*	1.021	0.162 *(0.064)*	1.176
Grocery stores	0.156 *(0.034)*	1.169	0.076 *(0.03)*	1.079
Service stations	0.092 *(0.044)*	1.096	0.2 *(0.04)*	1.221
Liquor stores	0.177 *(0.071)*	1.194	0.202 *(0.069)*	1.13
Spatial lag	0.1 *(0.008)*	1.105	0.122 *(0.016)*	1.13
Constant	0.616 *(0.103)*	1.852	0.527 *(0.107)*	1.694
Pseudo R^2	0.127		0.101	
All variables	**B**	**IRR**	**B**	**IRR**
Percent renter occupied	0.001 *(0.004)*	1.001	0.003 *(0.003)*	1.003
Percent below poverty	0.013 *(0.004)*	1.013	0.008 *(0.003)*	1.009
Percent age 18–21	− 0.016 *(0.008)*	0.984	− 0.005 *(0.007)*	0.995
Percent age 22–29	0.018 *(0.011)*	1.019	− 0.003 *(0.011)*	0.997
Drug stores	0.068 *(0.073)*	1.071	0.202 *(0.067)*	1.223
Grocery stores	0.118 *(0.033)*	1.125	0.041 *(0.032)*	1.042
Service stations	0.126 *(0.043)*	1.134	0.217 *(0.041)*	1.242
Liquor stores	0.199 *(0.067)*	1.220	0.205 *(0.068)*	1.228
Drinking places	0.005 *(0.019)*	1.005	− 0.011 *(0.019)*	0.989
Amusement/recreation	− 0.055 *(0.045)*	0.946	0.043 *(0.042)*	1.044
Spatial lag	0.088 *(0.008)*	1.092	0.114 *(0.017)*	1.121
Constant	0.249 *(0.152)*	1.282	0.298 *(0.160)*	1.347
Pseudo R^2	0.144		0.112	

The numbers in italics represent standard errors

of cash economies produced by businesses and facilities in attracting targets. Recalling place-based mechanisms, ordinary businesses both become and attract targets for robbery, and robberies tend to cluster near other places with robberies.

Concluding remarks

This paper has added to research on crime and place using an approach to modeling that we argue is preferable to traditional approaches in certain applications.

Theories of place offer guidance in how land-use may be related to the number of robberies. Following prior research on how robbery varies with the presence of different types of businesses, we successively modeled bundles of ordinary and discretionary businesses. Theory offered a clear guide to producing a set of models, and our analysis identified the best models among that set, considering both explanatory power and parsimony.

The complementary concepts of crime generators and crime attractors help explain the importance of ordinary

business. Though they mention potential victims, much of the discussion of generators by Brantingham and Brantingham (1995) refers to offenders: "Crime generators are particular areas to which large numbers of people are attracted for reasons unrelated to any particular level of criminal motivation they might have or to any particular crime they might end up committing" (1995, p. 7). Crime attractors create opportunities that are widely recognized by potential offenders (Brantingham and Brantingham 1995, p. 8). Cited examples are illegal markets, bars, and large shopping areas. While generators and attractors influence the behavior of potential offenders, they also affect the larger number of potential victims. As Yu and Maxfield (2014) note, not everyone chooses to visit a bar, pawn shop, or nightclub. But virtually all ambulatory people routinely visit and patronize certain retail establishments. Ordinary retail businesses are scattered around mostly residential areas, not entertainment districts. Everyone goes to grocery stores, and, in the Midwestern United States, most people end up near service stations. Service stations often include or are near small grocery stores or convenience stores. These are centers of behavioral routines for virtually everyone, not locations specializing in vice or drinking that appeal to more limited numbers of people.

Apart from these substantive findings, our approach departed from traditional NHST approaches in its consideration of sets of theory-based and socioeconomic variables. Theories of social disorganization and disadvantage permeate criminological research. One result is that researchers routinely include socioeconomic variables in multivariate analysis, regardless of the theoretical relevance or social processes under investigation. Socioeconomic variables, often inaccurately labeled "demographics," may be treated as controls, covariates, or predictor variables of interest. Analytic strategies often successively test models with and without different clusters of variables to see which combinations hold together.

While some theoretical rationale supports such strategies, what results are unduly complicated models that are often difficult to interpret, and do not address substantive significance. The consequences of this are most evident in analysis of large numbers of cases. Notably, the potential benefits of applying information theory are greater when analyzing large numbers of cases. Examining many cases can produce a kind of anti-parsimony by producing models where everything is statistically significant, yet little is said about substantive significance.

We recognize that our AIC approach is a substantial departure from methods long used in empirical criminology. Our approach also comes with certain limits and disadvantages. First, the AIC can be difficult to interpret, partly because it is not well-known. AIC does not consider that any of the tested models include the true model. These are all approximations to the true model and do not include the true model in the set. This is, however, founded on the assumption that all models are mere approximations, and no model can be treated as the 'true' model. A corollary of this is that AIC values are only indirectly related to effect size estimates for individual measures.

Second, although AIC will still produce the 'best-ranked' model among the set, much thought must be devoted to models a priori, primarily relying on theory. In other words, the results of the analyses are as good as the candidate set of models specified before the analyses are conducted (Mazerolle 2006). If all candidate models are poor fits, AIC will still produce the 'best-ranked' model. Similarly, the AIC analyses don't show if a better model exists other than the ones specified, unless that model is included in the set. Third, comparing AIC results across different studies can be difficult.

Finally, NHST can be more appropriate when it is difficult to specify a set of theory-based candidate models (Steidl 2006). In such case, NHST guides a statistical hypothesis rather than a substantive criminological hypothesis (Sleep et al. 2007). NHST is also preferable to AIC in the case of randomized experiments (Mazerolle 2006), where the null hypothesis of no difference is a straightforward baseline statement for framing analysis.

Future criminological research can use AIC in two ways. First, this approach can be used to build new models that not only aim at identifying the best among sets of models, but also to objectively assess competing models. Over 75 top-ranked journals in many fields that include astronomy, cosmology, nuclear and particle science, medical physics, ecology, statistics and psychology have published papers that used the AIC approach to model relationships. Criminologists have recently begun a more limited use of AIC and other information-theory criteria, but rarely to evaluate different models (Petrossian 2015; Groff 2014, are exceptions). The calculations of AIC are relatively easy. Many statistical software packages already produce AIC values within the goodness-of-fit tables. The subsequent calculations of delta AIC values (Δi) to assess the relative importance of all candidate models, as well as the calculations of Akaike weights (Wi) to evaluate the strength of evidence for these models, can be easily made in Microsoft Excel.

Second, this approach can be used to re-evaluate the models produced in previously published articles in order to weigh the importance of variables found to be statistically significant in these models. Criminological research offers examples where complex models built with tens or hundreds of thousands of cases are used to test the significance of large numbers of variables. Results may show

virtually every variable to be statistically significant. But what is the substantive importance of these variables? As Ziliak and McCloskey (2004) use the phrase "economic significance," and Sleep et al. (2007) propose the use of "biological hypothesis testing" to replace "statistical hypothesis testing", we might ask about "criminological significance" of low-performing predictor variables. AIC analysis of published research can re-evaluate such models with the goal of producing parsimonious explanations that are more theoretically sound.

Returning to the quote that opens this paper, "A well-designed model is, after all, a judiciously chosen set of lies, or… partial truths…." That is certainly true of the models we summarize in the top panel of Table 7. But the partial truths are consistent with theoretical expectations about people, places, and crime, and the models are parsimonious. Recalling a similar quote from Box (1976), "All models are wrong, but some are useful," we argue that empirically considering parsimony and relative theoretical support is more likely to produce useful models, than is empirically establishing statistical significance. Similarly, it's easier to evaluate a judiciously chosen, parsimonious set of lies than to sort through what untruths might underlie NHST-based models built with large numbers of cases and variables.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

American Community Survey. (2013). *State and county QuickFacts: Milwaukee County, City of Milwaukee*. Washington, D.C.: US Census Bureau.

Anderson, D. R. (2008). *Model based inference in the life sciences: A primer on evidence*. New York: Springer.

Anderson, D. R., Burnham, K. P. & White, G. C. (2001). Kullback-Leibler information in resolving natural resource conflicts when definitive data exist. *Wildlife Society Bulletin, 29*, 1260–1270.

Anselin, L. (2003). GeoDa 0.9 User's Guide. Urbana Champaign, IL: Spatial Analysis Laboratory, Department of Geography, University of Illinois, Center for Spatially Integrated Social Science.

Baumol, W. (1993). On my attitudes: Sociopolitical and methodological. In M. Szenberg (Ed.), *Eminent economists: Their life philosophies*. Cambridge: Cambridge University Press.

Berk, R., Brown, L., & Zhao, L. (2010). Statistical inference after model selection. *Journal of Quantitative Criminology, 26,* 217–236.

Bernasco, W., & Block, R. (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., Ruiter, S., & Block, R. (2017). Do street robbery locations vary over time of day or day of week? A test in Chicago. *Journal of Research in Crime and Delinquency, 54*(1), 244–275.

Block, R. L., & Block, C. R. (1995). Space, place, and crime: Hot spot areas and hot spot places of liquor-related crime. In J. E. Eck & D. David Weisburd (Eds.), *Crime and place. Crime prevention studies 4* (pp. 145–183). Monsey: Criminal Justice Press.

Box, G. E. P. (1976). Science and statistics. *Journal of American Statistical Association, 71*, 791–799.

Brantingham, P. L., & Brantingham, P. L. (1995). Crime generators and crime attractors. *European Journal of Cr: mmal Pohcy and Research, 3*(3), 5–26.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A Practical information-theoretic approach* (2nd ed., p. 2002). NY: Springer.

Burnham, K., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology, 65,* 23–35.

Bushway, S. D., Sweeten, G., & Wilson, D. B. (2006). Size matters: Standard errors in the application of null hypothesis significance testing in criminology and criminal justice. *Journal of Experimental Criminology, 2,* 1–22.

Feldmeyer, B., Warren, P. Y., Siennick, S. E., & Neptune, M. (2015). Racial, ethnic, and immigrant threat: Is there a new criminal threat in state sentencing?". *Journal of Research in Crime and Delinquency, 52*(1), 62–92.

Flather, C. (1996). Fitting species-accumulation functions and assessing regional land use impacts on avian diversity. *Journal of Biogeography, 23*(2), 155–168.

Fondell, T. F., Miller, D. A., Grand, J. B., & Anthony, R. M. (2008). Survival of dusky Canada goose goslings in relation to weather and annual nest success. *Journal of Wildlife Management, 72*(7), 1614–1621.

Garamszegi, L. Z. (2011). Information-theoretic approaches to statistical analysis in behavioral ecology: An introduction. *Behavioral Ecology and Sociobiology, 65,* 1–11.

Groff, E. (2014). Quantifying the exposure of street segments to drinking places nearby. *Journal of Quantitative Criminology, 30,* 527–548.

Groff, E., & Lockwood, B. (2014). Criminogenic facilities and crime across street segments in Philadelphia: Uncovering evidence about the spatial extent of facility influence. *ournal of Research in Crime and Delinquency, 51,* 277–314.

Gruenewald, P. J., et al. (2006). Ecological models of alcohol outlets and violent assaults: Crime potentials and geospatial analysis. *Addiction, 101,* 666–677.

Haberman, C. P., & Ratcliffe, J. H. (2015). Testing for temporally differentiated relationships among potentially criminogenic places and census block street robbery counts. *Criminology, 53*(3), 457–483.

Infogroup (2015). Our Company. Retrieved Apr 2, 2015, from http://www.infogroup.com/about-infogroup.

Johnson, S., Bowers, K., et al. (2009). Predictive mapping of crime by ProMap: Accuracy, units of analysis, and the environmental backcloth. In D. Weisburd, W. Bernasco, & G. Bruinsma (Eds.), *Putting crime in its place: Units of analysis in geographic criminology* (pp. 171–198). London: Springer.

Karlis, D., & Meligkotsidou, L. (2007). Finite mixtures of multivariate Poisson distributions with application. *Journal of Statistical Planning and Inference, 137*, 1942–1960.

Kullback, S. & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79–86

Lee, Y., Eck, J. E., Soohyun, O., & Martinez, N. N. (2017). How concentrated is crime at places? A systematic review from 1970 to 2015. *Crime Science, 6*, 6.

Lemmon, A. R. & Moriarty, E. C. (2004). The importance of proper model assumption in Bayesian phylogenetics. *Systematic Biology*, *53*, 265–277.

Livingston, M. (2008). Alcohol outlet density and assault: a spatial analysis. *Addiction, 103*, 619–628.

Lukacz, P. M., Thomson, W. L., Kendall, W. L., Gould, W. R., Doherty, P. F., Burnham, & Anderson, D. R. (2007). Concerns regarding a call for pluralims of information theory and hypothesis testing. *Journal of Applied Ecology*, *44*, 456–460.

Mallows, C. L. (1973). *Some comments on Cp. Technometrics* (Vol. 15, pp. 661–675). Estados Unidos: ASQ American Society for Quality.

Maltz, M. D. (1994). Deviating from the mean: the declining significance of significance. *Journal of Research in Crime and Delinquency*, *31*(4), 434–463.

Maltz, M. D. (2006). Some P-baked thoughts (P > 0.5) on experiments and statistical significance. *Journal of Experimental Criminology, 2*(2), 211–226.

Mazerolle, M. J. (2006). Improving data analysis in herpetology: Using Akaike's information criterion (AIC) to assess the strength of biological hypotheses. *Amphibia-Reptilia*, *27*(2), 169–180.

McCloskey, D. N., & Ziliak, S. T. (1996). The standard error of regressions. *Journal of Economic Literature, 34*, 97–114.

McQuarrie, A. D. R. & Tsai, C. L. (1998). Regression and time series model selection. New Jersey: World Scientific.

Petrossian, G. A. (2015). Preventing illegal, unreported and unregulated (IUU) fishing: A situational approach. *Biological Conservation, 189,* 39–48.

Pridemore, W. A., & Grubesic, T. H. (2013). Alcohol outlets and community levels of interpersonal violence: Spatial density, outlet type, and seriousness of assault. *Journal of Research in Crime and Delinquency, 50,* 132–159.

Rannala, B. (2002). Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Systematic Biology*, *51*, 754–760.

Richards, S. A., Whittingham, M. J. & Stephens, P. A. (2011). Model selection and model averaging in behavioral ecology: The utility of the IT-AIC framework. *Behavioral Ecology and Sociobiology*, *65*, 77–89.

Rao, C. R., & Wu, Y. (1989). A strongly consistent procedure for model selection in a regression problem. *Biometrika, 76,* 369–374.

Ripplinger, J., & Sullivan, J. (2008). Does choice in model selection affect maximum likelihood analysis? *Systematic Biology, 57,* 76–85.

Saffron, C. M., Park, J., Dale, B. E. & Voice, T. C. (2006). Kinetics of contaminant desorption from soil: comparison of model formulations using the Akaike information criterion. *Environmental Science & Technology*, *40*(24), 7662–7667.

Schwarz, G. (1978). Estimating the dimension of a model". *Annals of Statistics, 6,* 461–464.

Sleep, D. J. H., Drever, M. C., & Nudds, T. D. (2007). Statistical versus biological testing: Response to Steidl. *Journal of Wildlife Management, 71*(1), 2120–2121.

Steffensmeier, D., Painter-Davis, N., & Jeffrey Ulmer, J. (2016). Intersectionality of race, ethnicity, gender, and age on criminal punishment. *Sociological Perspectives*. https://doi.org/10.1177/0731121416679371.

Steidl, R. J. (2006). Model selection, hypothesis testing, and risks of condemning analytical tools. *Journal of Wildlife Management, 70*(6), 1497–1498.

Sullivan, C. J., & Mieczkowski, T. (2008). Bayesian analysis and the accumulation of evidence in crime and justice intervention studies. *Journal of Experimental Criminology, 4,* 381–402.

Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference, and model averaging in behavioral ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology, 65,* 13–21.

Takeuchi, K. (1976). Distribution of informational statistics and a criterion of model fitting. *Suri-Kagaku (Mathematical Sciences), 153,* 12–18. **(in Japanese)**.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *American Statistics, 70,* 129–133.

Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology, 53*(2), 133–157.

Weisburd, D., Lum, C. M., & Yang, S. M. (2003). When can we conclude that treatments or programs 'don't work?'. *The Annals of the American Academy of Political and Social Science, 587,* 31–48.

Wilcox, P., & Eck, J. E. (2011). Criminology of the unpopular: Implications for policy aimed at payday lending facilities. *Criminology & Public Policy, 10*(2), 473–482.

Wilson, D. K., Valente, D., Nykaza, E. T. & Pettit, C. L. (2013). Information-criterion based selection of models for community noise annoyance. *The Journal of the Acoustical Society of America*, *133*(3), EL195–EL201.

Yu, S. V., & Maxfield, M. G. (2014). Ordinary business: Impacts on commercial and residential burglary. *British Journal of Criminology, 54,* 298–320.

Zhu, L., Gorman, D. M., & Horel, S. (2004). Alcohol outlet density and violence: a geospatial analysis. *Alcohol and Alcoholism, 39*(4), 369–375.

Ziliak, S. T., & McCloskey, D. N. (2004). Size matters: The standard error of regressions in the American Economic Review. *The Journal of Socio-Economics, 33,* 527–546.