QUANTIFYING GANG LOCATIONS: SYSTEMATICALLY TESTING VALIDITY USING A PARTIAL TEST OF MESSICK'S UNIFIED PERSPECTIVE

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

By

Amber Perenzin Askey

December 2017

Examining Committee Members:

Jerry Ratcliffe, PhD, Advisory Chair, Department of Criminal Justice
Caterina Roman, PhD, Department of Criminal Justice
Ralph Taylor, PhD, Department of Criminal Justice
P. Jeff Brantingham, PhD, External Member, University of California Los Angeles

ABSTRACT

Gangs pose a serious problem in 21st century policing. After spending forty years studying them, Klein describes why we still have a lot to learn about gangs. Specifically, he suggests: “Too little attention has been paid to the communities in which gangs appear. Observing and understanding neighborhoods is far more complex than studying their gangs; yet it is communities that spawn gangs and must inevitably be the proximal focus for controlling them” (Klein 2007: xiv). The neighborhood complexity to which Klein refers has not been adequately addressed. The field lacks strong theoretical development to guide decisions about conceptualizing and operationalizing gangs from an ecological perspective. As a reflection of that lack of guidance, researchers today employ several alternate indicators of gang ecologies.

This study seeks to identify the consequences of variations in how gangs are conceptualized and measured at an ecological level. Researchers model gangs in substantially different ways, using dissimilar indicators, spatial scales, and levels of measurement. These variations may generate disparities in empirical results and different estimates of gang ecologies without a clear rationale for, or understanding of, the implications of their selection.

The analysis examines two central research questions: (1) Do indicators of gang ecologies identify gangs in similar ways and with results that are consistent across spatial scales? (2) Can indicators predict the presence versus absence of gangs—as a binary outcome—as well as predict continuous gang outcomes, such as the number of gangs present or the geographic size of the gang ecology, with results consistent across spatial scales?

Arrest data and gang data provided by the Philadelphia Police Department (PPD) included information on 3,996 gang members who belonged to 113 gangs. PPD data indicated (1) where gang members live, (2) gang arrests (N=7,488 from 2012-2015), (3) crime incidents that involve a gun (N=26,865 from 2012-2015), and (4) PPD defined gang set space boundaries.

This study examined the validity of each of these indicators when each is used to define gang ecologies. The analysis plan was guided by Messick's unified perspective of construct validity (Messick 1995) and included two types of analyses.

The first analysis employed a series of 60 regression models. Model comparisons tested various aspects of construct validity as proposed by Messick. The second analysis developed an algorithm creating gang set space polygons using either the locations where gang members live or the locations of gang-related crime. The set space polygons created by this algorithm were compared to the PPD set space polygons. The degree to which the gang set space polygons created by the algorithm overlapped with the gang set space polygons defined by PPD functioned as another validity test.

The results of the regression analysis revealed the home address and arrest variables better explained the spatial distribution of the PPD gang set space locations than the gun crime variable. The link between the gang indicators and the PPD identified set space polygons, however, was complex. Oftentimes, the home address data and the arrest variables significantly predicted a binary gang outcome—whether one or more gangs existed in an area; but those variables could not significantly estimate a continuous gang outcome, i.e. how many gangs were present, or geographic size of gang set space. This means the home address and arrest variables have limited ability to explain the spatial distribution of the gang set space boundaries defined by PPD.

The spatial analysis used an algorithm to approximate gang set space locations. The results indicate that locations of gang members' homes and of gang arrests both can approximate the PPD reported gang set space locations equally well. However, the spatial overlap between the PPD reported set space and the approximated set space locations proved relatively small. Although the approximated gang set space polygons usually did overlap with the PPD reported gang set space polygons to some extent, the mean overlap was 10% using the home address locations and 7% using the arrest data. The policy or practice usefulness of the indicators used here could be minimal. Although the overlap wasn't perfect, the algorithm was able to identify the general locations where gangs exist using only the home address of gang members or gang arrest data.

This study contributes to our knowledge about gang measurement at the ecological level. Conceptualizing and measuring gangs in a theoretically driven way is critical to the development of effective policies to control and prevent violence, fear, and other social harms caused by gangs. The results of this study will pave the way for future research to build on our understanding of how gangs link to crime, how community level dynamics work to foster or prevent gang activity, and ultimately, how to reduce and prevent gang problems.

STATEMENT OF FUNDING

This dissertation was supported by the National Institute of Justice's Graduate Research Fellowship Program in the Social and Behavioral Sciences (award number 2017-IJ-CX-0014). The National Institute of Justice is part of the Office of Justice programs and the U.S. Department of Justice. The findings, conclusions, and recommendations expressed in this dissertation are the author's and do not necessarily represent those of the National Institute of Justice, Office of Justice Programs, or the U.S. Department of Justice.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of the amazing individuals in Temple University's Criminal Justice Department. I feel lucky to have been part of such an incredible department. Thank you for making graduate school enjoyable, challenging, and rewarding.

Words cannot begin to express how grateful I am for my amazing committee: Dr Ratcliffe, Dr. Roman, Dr Taylor and Dr. Brantingham. You are remarkable mentors, and you probably do not realize how much you have taught me. I will be forever grateful for your support, guidance, patience, and encouragement.

I would also like to thank the Graduate School at Temple University for their financial support through the Graduate School Fellowship and the National Institute of Justice for supporting my dissertation research through the Graduate Research Fellowship program.

Most importantly, I owe a huge thank you to my family. Mom and Dad, thank you for going out of your way to provide me with every opportunity to be successful. Your sacrifices, endless support, and love have lead me to where I am. Dad, I know how much you loved telling me to do my homework from elementary school into my 30s, but I am officially done my homework, forever! Corey, I don't know what I did to deserve such a loving and supportive husband. You are an inspiration. I love you. This dissertation is dedicated to you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: PROBLEM STATEMENT

Gangs pose a serious problem in 21st century policing. The most recent statistics from the National Youth Gang Survey estimate there are approximately 30,700 gangs in the United States with a total of 850,000 active members. Approximately 85% of large cities reported experiencing gang activity between 2008-2012 (Egley, Howell, and Harris 2014). It is estimated that gang-related homicides increased nearly 25% in 2012 compared to the previous five year average (Egley, Howell, and Harris 2014). The association between gangs and violent crime makes gangs a high priority for law enforcement and academic researchers who share the overarching goal of understanding and preventing crime.

To better understand the link between gangs and crime, a body of research has developed that examines gangs from a spatial perspective. This is accomplished by aggregating gang data to geographic (ecological) units. The problem with this research is twofold. First, a variety of indicators are used to measure the presence of gangs at the ecological level. The ecological level is where data are measured at spatial units, such as census block groups or census tracts. This is problematic because we do not know if these different indicators are capturing gang ecologies in similar or different ways. Second, gang data have been aggregated to spatial units varying drastically in size. It is not yet known if gang presence measured at micro-spatial units, such as 500ft grid cells, are conceptually similar to gang presence measured at meso- or macro-spatial units, such as census tracts or at the city level, respectively.[1] This study investigates the construct validity of different indicators of gang ecologies at a variety of spatial units, with gang ecologies simply referring to any measurement of gangs at spatial units of any size.

[1] A discussion of these issues begins on page 9 and is summarized in Table 1 on page 13.

Construct validity refers to how much a measure actually taps the qualities of the intended construct (Maxfield and Babbie 2011). Many researchers have expressed concerns with construct validity in gang research (Ball and Curry 1995; Decker and Kempf-Leonard 1991; Esbensen, Winfree, He, and Taylor 2001). Definitional issues plaguing gang research is one such reason for this concern. Gangs have been defined in a number of ways (see Ball and Curry 1995; Decker and Kempf-Leonard 1991; Esbensen, Winfree, He, and Taylor 2001). Variations in how gangs, gang members and gang incidents are defined exist not only between researchers and state legislatures but among law enforcement agencies as well. “Definitions in different states and communities determine, in part, whether we have a large, small, or even no problem; whether more or fewer gangs and gang members exist” (Spergel 1995: 17). These inconsistent definitions have added a layer of complexity to gang research because they could be spawning gang ecology indicators of varying validity.

It is important that researchers and practitioners alike identify gang ecologies using valid indicators. For researchers, valid measures are needed in order to develop and test theories that explain gang behavior. For practitioners, measures have implications for how the gang problem is diagnosed and treated. In the words of Decker & Kempf-Leonard, “the formulation of effective policy responses to gangs depends on a reliable and valid foundation of knowledge of the ‘gang problem’” (Decker and Kempf-Leonard 1991: 272). Such knowledge is necessary to develop effective gang intervention programs.

Following the concerns raised by many researchers regarding how gangs are conceptualized and operationalized (Ball and Curry 1995; Decker and Kempf-Leonard 1991; Esbensen, Winfree, He, and Taylor 2001) this study examined the construct validity of different indicators used to identify gang ecologies. Time and financial constraints prohibited an exhaustive comparison of all gang ecology indicators that have been used in the literature; however, four dominant and widely-used indicators were compared. These indicators include the home address of gang members, gang arrests, crime incidents that involve a gun, and practitioner-defined gang set space. Gang set space is a geographically small sub-region of a territory where gang members “come together as a sociological group to ‘hang out’” (Tita, Cohen, and Engberg 2005: 272). The practitioner-defined gang set space will serve as the criterion variable while the home address, arrest and gun crime incidents will serve as independent variables, modeled as counts.

Gang research in general is fraught with inconsistent terminology. The terms set space, turf, and territory have all been used to describe ecologies—places associated with gangs. Kennedy and colleagues (1997) use the terms territory and turf interchangeably. In recent years, some scholars have attempted to add clarity to the terms. Tita and colleagues (2005) coined the term “set space” and described set space as follows: “…a gang may ‘claim’ an entire neighborhood as its domain or ‘turf’ but set space is the actual area within the neighborhood where gang members come together as a gang” (Tita et al 2005: 280). Tita and colleagues, therefore, distinguish gang set space from gang turf. This distinction is important for policing because smaller geographic areas have been shown to be most effective in achieving crime reduction outcomes. “A strong body of evidence suggests that taking a focused geographic approach to crime problems can increase the effectiveness of policing” (National Research Council 2004: 247). Police initiatives that target the precise locations where gang members hang out, therefore, would be more effective than initiatives that target the entire turf.

As will be shown in the literature review, gang ecologies can also be places where gang crime concentrates. While some researchers have referred to these locations as gang activity areas (Block, 2000) the term sphere of influence (Huddleston, Fox, and Brown 2012) has also been used.

But as reflected in a number of studies (e.g. Block 2000; Curry and Spergel 1988; Papachristos and Kirk 2006; Rosenfeld, Bray, and Egley 1999), scholars are often using concentrations of gang crime or associated gang activity as places associated with gangs. The purpose of this dissertation is to determine to what extent these gang ecology indicators converge or don't converge to provide further clarification as to what gang set space actually is.

In this study, I chose to use gang set space, as defined by PPD, as the criterion variable. This dataset is maintained by the criminal intelligence unit and is useful because it identifies the locations where gang members are often observed by the police. The decision to use the PPD set space polygon file as my criterion variable was a choice made based on the data available in Philadelphia. I do not argue here that gang set space is the be-all and end-all of gang research from a geographic perspective. The goal here is to determine to what extent other gang ecology measures align with the set space data.

Construct validation is a process, and there are a number of ways construct validity can be investigated. This work adopts portions of Messick's (1995) unified approach to construct validation to examine different aspects of construct validity. This will be accomplished using two types of analyses. The first compares a series of zero-inflated beta regression models using different combinations of gang indicators. These models are used for two reasons. First, they allow modeling an outcome variable that is a proportion. Second, these models are hurdle models that can model places without gang presence *separately from* places that contain some degree of gang presence. To the best of my knowledge, this is the first study to apply these types of models to gang research. The second analysis involves creating gang set space boundaries quantitatively using different gang indicators, then comparing the spatial congruence of set space defined using different indicators. The degree to which these set space locations overlap will serve as an assessment of construct validity.

The validity of gang ecology indicators has implications for theory testing and theory development. The underlying assumption is that the same neighborhood-level social processes that drive crime—such as broken windows, collective efficacy or social disorganization theory for instance (Shaw and McKay 1942 [1969])[2]—are also responsible for generating the spatial patterning of gang ecologies, and those *same* neighborhood-level process can be captured with any of the four gang ecology indictors that will be used in this study. If impacts of *different* social processes are reflected with each indicator, this implies that gang ecologies are multi-dimensional with interrelated attributes that either link to separate gang dynamics or reflect impacts of different non-gang social processes. This study sheds light on the dimensionality of gang ecologies similar to what has been done with the construct of self-control (Ward, Nobles, and Fox 2015).[3] Before competing mechanisms driving various gang dimensions can be considered, methodological limitations need to be taken into account.

Identifying gang ecologies proves relevant to practitioners, and particularly for law enforcement officials, who are tasked with developing initiatives to prevent and reduce crime. Recognizing the importance of neighborhoods in combating the gang problem, many agencies have chosen to adopt crime reduction strategies that focus on changing community level dynamics. These holistic, community-level interventions are often geographically focused and target areas

[2] Social disorganization was developed by Shaw & McKay to explain the spatial patterning of delinquency rates. They noticed that delinquency rates, among other outcomes, were spatially patterned in a way that aligns with urban expansion and the concentric zones as described by Burgess. Shaw and McKay conclude that delinquency rates are highest in economically disadvantaged areas that are marked by residential instability and heterogeneity. Areas with these social characteristics are unable to regulate themselves and are regarded as socially disorganized. This disorganization explains why delinquency rates are higher in these locations.

[3] Ward and colleagues examined whether self-control as a concept is separate from the six conceptually and empirically distinct elements that comprise self-control (risk-seeking, impulsivity, physical activities, temper, self-centeredness, and simple tasks).

that have high concentrations of crime (Boyle, Lanterman, Pascarella, and Cheng 2010; Corsaro, Brunson, and McGarrell 2009; Papachristos, Meares, and Fagan 2007). However, the literature provides no supporting evidence to suggest that crimes can be used to identify the locations that generated the gang problem, or that crime locations are the most appropriate indicators to identify the locations to focus interdiction efforts. If a gang develops in a school, and those members then go to a local park to commit crime, the park is not the source of the gang problem and arguably should be reconsidered as the focus of a gang intervention. This is important, because "The results of ecological studies of gangs [are needed] to formulate economic and social policies that *address the root causes of gangs rather than simply providing triage to the problems caused by gangs*." (Tita 1999: 130, emphasis added). The current study will clarify whether gang problems overlap spatially with set space and gang residence locations. The results can be used to identify the locations where gang interventions are the most applicable and appropriate.

Additionally, the findings of this study could have implications for practitioners who spend considerable time and resources mapping gang ecologies qualitatively. If quantitative indicators that are commonly available to most police department can replicate similar results to the qualitative methods, this process could be automated, saving time and resources for law enforcement agencies and increasing objectivity in the process. This last point is increasingly relevant because legal scholars have started to question the accuracy of police mapping efforts when law enforcement officers use areas such as 'gang territory' or 'crime hotspot' as part of their totality of circumstances when justifying a constitutionally-reasonable search of an individual or property (Ferguson 2011).

CHAPTER 2
LITERATURE REVIEW

Why study gangs?

A number of studies have found that gang membership increases criminality and deviant behavior at the individual level. In two studies, offending rates of gang members were collected at three time periods: before, during and after gang membership. Both studies found that individuals are involved in more crime and delinquency while they are in the gang compared to before they join or after they leave (Esbensen and Huizinga 1993; Thornberry, Krohn, Lizotte, and Chard-Wierschem 1993). Specifically, longitudinal studies have shown that male gang members had individual offending rates that were two to three times greater compared to the males who were not in a gang (Esbensen and Huizinga 1993). Using data from a sample of students, Brattin and colleagues found that gang members commit more crime compared to delinquents who were not associated with a gang (Brattin, Hill, Abbott, Catalano, and Hawkins 1998). This study also found that the relationship between gang membership and delinquency remained statistically significant even after controlling for the influence of friends who were delinquent but were not in a gang (Brattin et al. 1998). Together, these studies demonstrate a strong link between gangs and crime at the individual level.

In addition to crime, gangs are have been linked to other types of social problems. For instance, gang members are more likely to engage in risky sexual behaviors and are more likely to abuse alcohol and other substances compared to delinquents who are not in a gang (Brattin, Hill, Abbott, Catalano, and Hawkins 1998; Harper and Robinson 1999). At the community level, gangs have been associated with high levels of public fear and intimidation (Lane and Meeker 2000) which negatively affects various aspects of community such as the academic performance of students in schools (Oehme 1997), the ability of local businesses to sustain themselves (Decker and Van Winkle 1996; Venkatesh 2008), and, in the case of witness intimidation, the effectiveness of the criminal justice system (Anderson 2007). The negative effects associated with gangs make gangs a high priority for law enforcement and academic researchers who share the overarching goal of reducing the negative effects of these groups.

Gang ecology

In a recent review of 719 studies published in the journal *Criminology* between 1990 and 2014, Weisburd (2015) found that individuals are the unit of analysis in over 66% of the published studies. Geographic areas at the micro-, meso- and macro-level *together* made up less than 24% of criminology literature. These findings clearly indicate that criminology research is dominated by studies focusing on individuals. Richard Rosenfeld spoke to the implications of this unbalance in his 2010 address to the American Society of Criminology. He stated that criminology studies are lacking "a theoretical framework to organize macro level research in a cumulative body of knowledge, disclose unresolved research puzzles, and point to productive areas of future inquiry" (Rosenfeld 2011: 7). The dominance of individual level studies has also been noted specifically in gang research (Decker, Melde, and Pyrooz 2013). Literature focusing on gang members at the individual level is far more abundant than literature at the ecological level. "Even the resurgence of neighborhood-level research over the past quarter-century has not resulted in considerable attention toward gangs" (Decker, Melde, and Pyrooz 2013: 387).

Although ecological studies of gangs appear less frequently in the literature compared to individual level studies, studies have found that the negative effect of gangs at the individual level translates to the community level. Communities with gangs experience more crime compared to communities that do not have gangs (Huebner, Marin, Moule, Pyrooz, and Decker 2014).

The importance of ecological level studies has been stressed by a number of researchers who have found evidence to suggest that gangs may possess qualities that are supra-individual in nature and cannot be measured exclusively at the individual level (Braga, Kennedy, Waring, and Morrison-Piehl 2001; Klein 2004; Papachristos 2009; Spergel 2007). Using data from Chicago, Papachristos (2009) found that spatial patterns of gang conflict were stable from 1994 through 2002, despite the fact that gang membership has been described as extremely transitory in nature (Decker and Curry 2002; Sullivan 2005). “Gang members come and go, but their patterns of behavior create a network structure that persists and may very well provide the conduit through which gang values, norms, and culture are transmitted to future generations” (Papachristos 2009: 119). These findings suggest that gang problems cannot be completely reduced to the individuals who comprise gang membership.

In a similar vein, Klein (2004) describes how the gang is rooted at the community level and is likely to re-generate after a take-down initiative by law enforcement if the community itself does not change. “You break up a gang not by busting its leaders, but by working on its group processes and its neighborhood context” (Klein 2004: 111). This neighborhood context is where gang behavior is normalized, so changing these neighborhood level dynamics can lead to reductions in gang crime.

These findings have encouraged researchers to examine gangs from an ecological perspective by applying neighborhood level theories to explain gang presence. “More comparisons of gang joining and gang behavior across neighborhoods and across cities are necessary for the field to move forward. The way ahead in these areas is clearly through multimethod, multisite studies of gangs that integrate dimensions of time as well as Short’s three-level framework.” (Decker, Melde, and Pyrooz 2013: 394). Short’s three-level framework is also referred to as the level of explanation problem (Short 1985; Short 1998). This idea recognizes the distinction between sociological, psychological and group level processes that are used to explain crime outcomes. Short points out that researchers and theorists typically do not integrate the three levels of explanation in their work. Such integration is important to advance the field. In order to design studies to address the neighborhood level dynamics of gangs, valid indicators of ecologies are needed.

Gang ecology indicators

Overall, gang research at the ecological level falls into three camps: (1) understanding the relationship between gangs and crime or other outcomes (Block and Block 1993; Braga, Hureau, and Papachristos 2014; Hall, Thornberry, and Lizotte 2006; Huebner et al. 2014; Papachristos and Kirk 2006; Pyrooz 2012; Rosenfeld, Bray, and Egley 1999; Taniguchi, Ratcliffe, and Taylor 2011; Tita and Radil 2011; Tita and Ridgeway 2007), (2) understanding why gangs operate in some areas but not others (Brantingham, Tita, Short, and Reid 2012; Cartwright and Howard 1966; Decker, van Gemert, and Pyrooz 2009; Katz and Schnebly 2011; Thrasher 1927; Tita, Cohen, and Engberg 2005) and (3) understanding how to prevent and reduce gang activity through community-level gang initiatives; focused deterrence strategies, such as Operation Ceasefire (Braga, Kennedy, Waring, and Morrison-Piehl 2001) and Project Safe Neighborhoods (Papachristos, Meares, and Fagan 2007).

While the research questions in these studies vary widely, each of them relies on an ecological indicator of gangs. These indicators include: social observation of gang members; the home address of gang members; hand-drawn territories by practitioners, gang members or community residents that are not in a gang; the use of crime data such as gang-related crime, gang-motivated crime, or gun-involved crime; or the locations of gang graffiti. Studies that rely on these various ecological gang indicators are summarized in Table 1[4] (see page 14). These researchers did not intend to operationalize a new construct with each indicator; rather, they used different geographic indicators based on availability.

Despite the fact that various indicators have been used to study gang ecologies, previous ecological studies on gangs have not compared these gang indicators against each other. The goal of this study is to determine how similar these indicators actually are. This study will address two main research questions: (1) Do different indicators of gang ecologies reflect the same construct? (2) Can indicators predict the presence versus absence of gangs (binary gang outcome) as well as continuous measures of gang ecologies akin to the number of gangs present or the geographic size of the gangs' set space?

Many researchers who study gangs from an ecological perspective acknowledge the limitations of the indicator they chose to adopt in their study (Table 1). Bynum and Varano (2003) intended to measure the extent of the gang problem in their study. Although they chose gun-involved crime incidents to model the gang problem, they acknowledge that other measures, such as gang-related crime, have been used to measure the same construct. Similarly, Blasko and colleagues (2015) measure gang presence by asking practitioners to identify gang set space locations as has been done by many other gang researchers (Block and Block 1993; Brantingham, Tita, Short, and Reid 2012; Kennedy, Braga, and Piehl 1997). They recognize that their measure was not validated by gang members themselves as was done by Tita et al. (2005).

[4] The search criteria used to identify these studies included keyword searches, forward searches using foundational studies and review of prominent scholar CVs. Keyword searches included gang territory, gang set space, gangs and community, gangs and spatial analysis, gangs and ecology. The forward searches to identify works that cited Ley & Cybriwsky (1974), Tita, Cohen & Engberg (2005) and Cartwright and Howard (1966). CV searches included Papachristos, Tita, Spergel, Kennedy, Braga, Maxson, Klein and Huff.

Sometimes researchers discuss the limitations of other indicators and describe how the indicator they chose to adopt is superior to these other methods Table 1. They do not, however, argue that their indicator captures a *different* construct. For instance, Katz and Schnebly (2011) argue that binary gang indicators differentiating locations with versus without gangs, as done also in Tita et al. (2005), are not the best measures of gang ecologies. Instead, they propose an indicator capturing the magnitude of the gang problem such as the residential addresses of gang members. They suggest the home address indicator "[more] accurately reflect[s] the realities of neighborhood-level gang problems" compared to binary indicators of gang presence (Katz and Schnebly 2011: 386). Curry and Spergel (1988) justify their use of gang homicide to gauge the extent of gang problems in communities in a similar way: "A key methodological issue is the operational definition of the gang problem. With a variety of options available, some empirical if not theoretical consensus exists that the measure should be based on official crime statistics" (Curry and Spergel 1988: 384).

Beyond variation in gang ecology indicator used is variation in the size of the geographical units used in each study in Table 1. The spatial units include 150 meter grid cells (Block 2000), census block groups (Blasko, Roman, and Taylor 2015; Tita, Cohen, and Engberg 2005), census tracts (Katz and Schnebly 2011), neighborhood clusters (Chicago: Papachristos and Kirk 2006), community areas (Chicago: Curry and Spergel 1988) and police precincts (Detroit: Bynum and Varano 2003). In two studies, the spatial distribution of the data were studied without aggregating the data to a spatial unit (Ley and Cybriwsky 1974; Thrasher 1927). The use of spatial units varying drastically in size indicates that gang researchers have been operating under the homology assumption. "Under the homology assumption, the same relationship…has a homologous relationship across levels because the 'laws' or processes connecting the cause and outcome are ‘formally identical’” (Taylor 2015: 94-95). In other words, under the homology assumption, it doesn’t matter if you aggregate gang data to 150 meter grid cells or census tracts because the relationships you find will be driven by the same theoretical dynamics regardless of spatial scale.

Since different indicators and spatial units are used in this body of research, it is difficult to make meaningful comparisons across studies. Such comparisons are important for the advancement of criminological theory using ecological units of analysis (Taylor 2011) and to articulate the current state of gang research to inform the development of effective policies to reduce gang violence.

The remainder of this review will describe each of gang indicators mentioned in Table 1 in more detail. These gang ecology indicators are grouped into three categories: (1) gang territory/set-space, (2) indicators using the home address of gang members, and (3) indicators using officially recorded crime data. The strengths and weaknesses of each indicator will be discussed in terms of both theory and practice. Concluding the review is a summary of construct validity as it pertains to gang ecology indicators.

Table 1: Summary of gang ecology indicators

Research question	Unit of analysis	Gang indicator	Was the indicator validated?	Noted limitations	Findings
(Thrasher 1927)					
What are the conditions that give rise to gangland?	Geocoded locations (data were not aggregated)	Social observation	No	There are different types of gang hang-outs; some are indoors, some are outdoors. The hang-out may or may not be where the gang boys live.	Gangs are most likely to be found in the “zone of transition” (Burgess), or areas marked by disorganization and instability and rapid change.
(Katz and Schnebly 2011)					
Which community structural indicators explain variability in the gang phenomenon?	Census tracts	Home address	No	There are different ways to measure gang presence (binary vs continuous variables). Limitations concerning how gang members are documented are also noted.	Gangs flourish in disadvantaged neighborhoods marked by some degree of residential stability.
(Blasko, Roman and Taylor 2015)					
What are the psychological and ecological impacts of gangs on perceived neighborhood incivilities?	Level: 1 distance to gang set space Level 2: census block groups	Hand-drawn maps by practitioners	No - but participants included both local and federal law enforcement (LE). Also, all levels of LE management were represented.	The authors note that their measure of gang set-space has not been validated by gang-members	Residents’ negative perceptions of rowdy teen groups increased as distance from gang set-space decreased.

Table 1: continued

Research question	Unit of analysis	Gang indicator	Was the indicator validated?	Noted limitations	Findings
(Tita, Cohen, and Engberg 2005)					
What are the features of communities that foster gang set-space?	Census bock groups	Hand-drawn maps by gang members	Yes - using non-gang participants, street workers and the presence of gang graffiti	None	Communities marked by low levels of social control and underclass features are the most salient indicators of gang presence.
(Curry and Spergel 1988)					
Do poverty and social disorganization variables explain gang crime and delinquency outcomes in similar ways?	Chicago community areas (n=75)	Gang-related homicide	No	There are different definitions of gang behaviors. Gang homicide is just one of many different types of gang problems.	Community gang problems are ecologically distinct from non-gang delinquency.
(Papachristos and Kirk 2006)					
Can the systemic social disorganization theory explain gang behaviors at the neighborhood level?	Chicago neighborhood clusters (N=343)	Gang-motivated homicide	No	They acknowledge that their measure of gang behavior is one of many measures that could have been used.	The effect of immigrant concentration and residential stability do not explain gang homicide and non-gang homicide in similar ways.

Research question	Unit of analysis	Gang indicator	Was the indicator validated?	Noted limitations	Findings
(Block 2000)					
How do gang activity areas influence crime patterns?	150 meter cells	Gang-motivated crime incidents	No	Gang activity areas need to be validated by police officers, community residents and gang members.	Violent crime, property crime and drug crime are higher where gang activity areas overlap. However, this relationship is less pronounced for property crime.
(Bynum and Varano 2003)					
Was the Detroit anti-gang initiative successful in reducing gang behavior?	Detroit police precincts	Gun involved crime	No	Gun-related incidents are used to approximate gang-related incidents. Furthermore, police officers reported that some gang members are more likely to carry firearms compared to others.	The anti-gang initiative included saturation patrol and suppression efforts. A time-series analysis reveals gun crime declined after the program was implemented in both precincts. This decline, however, was not statistically significant.
(Ley and Cybriwsky 1974)					
What does graffiti tell us about the social order of Philadelphia neighborhoods?	Geocoded locations (data were not aggregated)	Gang graffiti	No - but the authors mention that the residences of gang members were likely to overlap with the gang turf	Graffiti is an accurate indicator of gang turf but some gang territories are better defined compared to others. They conclude that "territorial indistinctness mirrors social weakness" (497).	Graffiti can be used to identify turf ownership and locations of gang conflict. The intensity of gang graffiti can be used as an indicator of the strength of territorial claims.

Note: The studies in this table are not intended to be exhaustive of all studies that examine gangs from an ecological perspective. Rather, this table serves to provide an overview of gang ecology indicators that have been used in the literature with an example of how each indicator was used in the study.

Theoretical relevance of different gang indicators

Researchers rely on both qualitative and quantitative methods to identify gang ecologies. Qualitative methods include identifying gang ecologies through interviews, focus groups or social observation. Quantitative methods include geocoding the locations where gang members live or where they commit crimes and then aggregating those locations to spatial units. This has also been done using types of crime that are associated with gang activity, such as shooting locations or the locations of gun-involved crime more generally, when a list of documented gang members is not available. Each of these data sources, and how they are used in gang research, will be described next. The strengths and weaknesses of each indicator will also be described in the context of theory and practice.

Qualitative methods

Qualitative methods, including the use of interviews, focus groups, gang audits or social observation, have been used to identify where gangs are located. Gang audits are becoming more popular as the number of jurisdictions implementing focused deterrence strategies are increasing. Tita and colleagues identified gang ecologies using interviews. With the help of the local police department, they recruited a sample of 50 gang members and ten community residents and asked them to identify locations where gang members come together as a gang. Tita and colleagues refer to these locations as gang set space—geographically small sub-regions of neighborhoods where gang members "come together as a sociological group to 'hang out'" (Tita, Cohen, and Engberg 2005: 272). Set space locations were delineated by providing all participants with a detailed map of their neighborhood and asking them to use markers to draw gang set space locations. Gang members were also asked to identify locations of rival gangs' set spaces. The data were validated across respondents. When inconsistent boundaries were drawn, the research team made site visits to look for the presence of graffiti.

Another source of gang information comes from focus groups. Participants include practitioners who work in criminal justice agencies and related fields. Police officers, probation officers and social workers come into contact with gang members on a regular basis. Their experiences allow them to identify locations where gangs operate. Many researchers have used gang boundaries delineated by practitioners to identify gang locations (Block and Block 1993; Brantingham, Tita, Short, and Reid 2012; Kennedy, Braga, and Piehl 1997; Roman and Chalfin 2008).

Lastly, Thrasher (1927; 1928) relied on informal social observation to study gangs from a human ecology perspective. Although the process through which he built rapport with gang boys took several weeks, once rapport was established, he was able to observe the gang in a variety of settings to document many aspects of gang life (Thrasher 1928). Over the course of his research, he was able to document the location of over 1,200 gang hang-outs. Most of these hang-outs were located either outdoors on street corners or indoors in clubhouses. Thrasher describes these "hang-out" locations separately from the "home territory" of gangs (Thrasher 1927: 124-126) which suggests these two locations are conceptually distinct from one another. Thrasher characterized "gangland" as a "geographically and socially interstitial area in the city" (Thrasher 1927: 22).

Strengths

A number of ethnographic studies have suggested gangs influence community life (Jankowski 1991; Suttles 1968; Venkatesh 2008; Whyte 1955). These studies show that gangs engage with community residents and sometimes operate as a social institution within the community (Jankowski 1991; Suttles 1968). If this is true, community residents would be aware of who the gang members are and where they hang out. Furthermore, research has shown that gangs do not operate in isolation from one another (Papachristos 2006; Papachristos, Hureau, and Braga 2013; Radil, Flint, and Tita 2010; Tita and Radil 2011). The existence of inter-gang social networks lends support to the use of gang members to identify the set space of other gangs in their community. Additionally, this method allows for cross validation of responses between subjects which improves reliability. Finally, understanding where gang members hang out, even if they do not commit crime there, is also important for studying social incivilities related to gang presence, such as fear and disorder. In other words, our ability to identify where gang members hang out is important in and of itself.

Weaknesses

Several weaknesses of qualitative methods deserve to be noted. The sampling method used to identify participants could influence the results. If a community resident is not an active community member or is new to the area, he/she may not be able to identify gang ecologies. That person may not even be aware that a gang operates in their neighborhood, or alternatively, may exaggerate gang boundaries.

Data generated by practitioners also have limitations. Successful gangs operate under the radar of law enforcement. It is possible that practitioners are unaware of active gang members and gang areas because gang activities have not been detected by law enforcement. Klein recalls bringing a list of 100 gang members to a local police department (Klein 1971). To his surprise, the police's documented list of gang members for that particular gang included only 20 people.

Qualitative data collection is also very time consuming. It can take several months or even years (in the case of Thrasher 1927) to collect, collate and validate gang data that are collected qualitatively. Furthermore, the information gathered from focus groups and interviews is cross-sectional. When gang areas and members change, which some argue is quite frequently (Decker and Curry 2002; Sullivan 2005), the process of defining gang ecologies has to start all over again.

Home address

Quantitative methods also can be used to identify gangs at the ecological level. One such method involves counting the number of gang members in a given area. In one example, Pyrooz and colleagues relied on data from the National Youth Gang Survey to identify the number of known gang members in each of the 100 largest cities in the U.S. (Pyrooz, Fox, and Decker 2010). In a similar vein, the home address of gang members has been used to identify the presence of gangs at the neighborhood level. In one study, a gang area was classified as such if at least 70% of the gang resided within that geographic unit (Cartwright and Howard 1966). In another study, home address locations were used to identify the number of gang members within each census tract (Katz and Schnebly 2011). More recently, Huebner and colleagues also used gang member home address locations to calculate a gang membership rate for each census tract in St. Louis (Huebner et al. 2014).

Strengths

The use of home addresses data as an indicator of gang ecologies is consistent with crime pattern theory (Brantingham and Brantingham 1993). This theory would classify these locations as activity nodes that form an anchor for both activity and awareness spaces of gang members. Activity spaces are locations people visit frequently; they may include where a person lives or where they work or go to school. Awareness spaces are places a person traverses to get to their activity spaces. People are familiar with the places along the routes they travel to get to and from work, the grocery store, the doctor's office, etc. The theory suggests individuals are most likely to engage in criminal or deviant acts in locations they are familiar with, i.e. within their activity or awareness space. If this is true of gang members, home address locations have a potentially theoretical foundation as anchoring gang set space. Porteous (1977) used the activity spaces of gang members in Victoria, British Columbia to identify gang turf. Activity spaces included where each gang member lived, went to school and worked. He also used meeting places (e.g. a diner,) and the locations where they committed crime. He found that "Clearly, the objective social space of the Burnside Gang cannot be contained within a single all-embracing boundary line. However…it is evident that a territorial core may be demarcated around which residences, schools, and meeting places are located, and within which most delinquent acts are committed" (Porteous 1977: 252).

Weaknesses

Two problems arise when data are aggregated to spatial units. The first problem concerns how large the spatial units are, the second concerns where the boundaries of each unit are drawn. Both problems are described as the modifiable areal unit problem (MAUP) (Fotheringham and Wong 1991; Openshaw 1984).

> "The MAUP is in reality composed of two separate but closely related problems. The first of these is the well-known *scale problem* which is the variation in results that can often be obtained when data for one set of areal units are progressively aggregated into fewer and larger units for analysis…Although scale differences are a most obvious manifestation of the MAUP there is also the problem of alternative combinations of areal units at equal or similar scales. Any variation in results due to the use of alternative units of analysis when the number of units is held constant is termed the *aggregation problem*" (Openshaw 1984: 8).

This study attempts to isolate the effects of the *scale* problem described by Openshaw. Aggregating home address locations to census block groups, for instance, may lead to one set of results, but different results may emerge when data are aggregated to zip code boundaries. In one study, crime and delinquency data were aggregated at two different units of analysis, the census tract level (N=495) and the neighborhood level (N=84). The two units of analysis produces slightly different results; specifically, the exploratory power of the predictor variables was stronger in the neighborhood level models compared to the models using data aggregated to census tracts (Ouimet 2000). Because of this problem, it is difficult to determine if statistical effects at different scales are true effects, or if they are statistical artifacts related to the MAUP.

Furthermore, the discontinuity thesis suggests that each spatial scale is conceptually different from the others. "For a range of analytic and theoretical reasons, *different types of processes* are likely to be involved at different spatial scales" (Taylor 2010: 461). Aggregation bias and disaggregation effects must be considered in order to avoid fallacies of the wrong level (Taylor 2010: 461-462).

Using home address locations to define gang ecologies is problematic for another reason. The underlying assumption is that gang members live within their gang's set space, however, some research has found that gang members do not necessarily live within the territories they defend as their turf (Moore, Vigil, and Garcia 1983). This creates conceptual slippage in terms of what constitutes a gang area. Is it where a gang operates or where the gang members live?

Another limitation of this method is that it not only requires the researcher to have an exhaustive list of gang members, but it requires they have an accurate home address location for each member as well. To the first point, previous research is mixed concerning the ability of law enforcement agencies to successfully document gang members (Katz, Webb, and Schaefer 2000;

Zatz 1987). Members of the Philadelphia Police Department, for example, are extremely hesitant to classify a person as a gang member if they have never been arrested. This is because the State of Pennsylvania includes criminal or delinquent acts in its definition of what is a gang (18 Pennsylvania § 5131 Recruiting Criminal Gang Member). Even when the Philadelphia Police Department has acquired evidence or intelligence that suggests a person is a gang member, this person will *not* be identified as a gang member until they have been arrested for a crime; this is particularly the case with juveniles. In his 1997 address to the American Society of Criminology, Short points out that most police agencies and gang researchers classify gang members in a similar way which requires a criminal element in gang definitions. This is done despite the fact that "playgroups" and "unsupervised peer groups" form the bedrock of classical gang research (Short 1998). Excluding known gang members that have no prior arrests from their official gang list underestimates the number of gang members operating in the city. This is a limitation of gang ecology indicators that use home address data.

To the second point, regarding the accuracy of home address information, these data are often self-reported by gang members when they are stopped by the police. A gang member might report living with his/her mother to law enforcement when he/she actually lives with a cousin or friend, or spends time in multiple residences. Some researchers have found that gang members tend to be unreliable, either exaggerating information or outright lying. "The only thing worse than the young reporter's description of a gang incident is his acceptance of the gang participant's statement about it. The gang member is often the worst informant about gang affairs, a fact…too often overlooked by news reporter and social scientist alike" (Klein 1971: 18).

Crime locations associated with gang activity

Crime data can be used to identify gang locations in a similar way that home addresses data are used. This can be done either with or without a documented list of gang members.

If a list of documented gang members is available there are two ways an incident can be classified as a gang incident: one uses a member-based definition, the other uses a motive-based definition (Maxson and Klein 1990). Using the member-based definition, an incident is classified as a gang incident if the suspect or victim is a documented gang member. This is a rather broad definition because it includes incidents that may be completely unrelated to gang activity but involve documented gang members. Member-based definitions are often referred to as the Los Angeles definition since this is the definition used by the Los Angeles Police Department and the Los Angeles Sheriff's Office. Other agencies that classify gang incidents using member-based definition include the New York, Detroit and Miami Police Departments (Maxson and Klein 1990). Motive-based definitions, however, are more restrictive compared to the member-based definition. Using this definition, an incident is only classified as a gang incident if the incident was actually motivated by gang activities such as territoriality, retaliation or recruitment. The motive-driven definition is often referred to as the Chicago definition, but it is also used in other agencies including the Philadelphia and Seattle Police Departments (Maxson and Klein 1990).

Curry and Spergel operationalize the presence of a "gang problem" at the community level using homicide locations (Curry and Spergel 1988: 384). Specifically, they calculate the homicide rate in each of Chicago's 75 community areas, but only using homicides that were determined to be gang-motivated. Similarly, Papachristos & Kirk identify "gang behavior" locations using the number of gang-motivated homicides in a given area using the more restrictive Chicago definition (Papachristos and Kirk 2006: 68). Rosenfeld and colleagues classified gang homicides using the motive-based definition as well as the member-based definition (Rosenfeld, Bray, and Egley 1999). Homicide incident counts were summed in each census block group and the results generated by the two definitions were compared.

The studies described above each use homicide data to classify locations as gang areas. Instead of relying only on only homicide data, Block included all crime incidents that were classified as gang-motivated to identify gang locations (Block 2000). Crime incident counts were summed within each grid cell.

If a list of documented gang members is not available, gang locations can be identified by geocoding types of crime that are traditionally associated with gang activity. Bynum and Varano (2003) did not have access to crime data that identified crime incidents as gang-motivated or gang-related. Instead, the authors used homicide, robbery and non-domestic aggravated assault incidents that involved a gun as a proxy measure of gang activity. These incidents were then analyzed spatially and temporally to examine the effectiveness of a gang intervention. Unfortunately, the authors were unable to verify that gun-crimes directly link to gang-crime.

Strengths

Since many crimes go unreported, criminal justice research is often plagued by the fact that crime reported to the police does not reflect a true figure of crime. Homicide data, arguably, most accurately reflect the true figure of murder since homicides are more likely to be reported and intensely investigated compared to other crime types, thereby overcoming this issue. Using homicide locations to quantify gang areas is not likely to be influenced by issues related to non-reporting which makes homicide data an attractive indicator of gang ecologies.

There are also benefits to using a sub-set of crime incidents, such as gun-crime, to model gang activity. The primary benefit is it does not require a list of documented gang members. Gang lists are problematic for a number of reasons. First, and as will be described under the limitations section starting on page 160, previous research is mixed regarding the ability of law enforcement agencies to document gang members (Katz, Webb, and Schaefer 2000; Klein 1971; Zatz 1987). Second, there are a number of ways to classify a crime incident as a gang incident (Maxson, Gordon, and Klein 1985). Lastly, researchers have found that definitions used to classify incidents as gang incidents vary between police jurisdictions, making it difficult to compare studies across agencies (Maxson 1999; Papachristos and Kirk 2006). If we can identify where gangs are operating without these lists, these issues are avoided.

Another benefit of using crime data to identify gang areas is the ease at which the data can be geocoded and spatially analyzed. As mentioned earlier, identifying gang ecologies qualitatively through focus groups or interviews is an extremely time-consuming process. The process of geocoding crime data and creating summed counts or density maps is a much less labor-intensive process compared to qualitative methods.

Weaknesses

A number of limitations concerning the use of official crime data deserve to be noted. One limitation of using crime data to identify gang locations is it sometimes requires that researchers have an exhaustive list of gang members and a reliable method of coding incidents as gang-motivated, gang-related, or not connected in any way to a gang. This is problematic because lists of gang members from the police may not be accurate (Klein 1971; Zatz 1987) and the method of coding incidents has not been standardized across agencies (Maxson 1999; Papachristos and Kirk 2006).

Another limitation concerns the fact that some researchers have found most gang activity is non-criminal in nature. More often than not, gang members spend their time socializing with one another, hanging out, or partying (Klein 2004; Suttles 1968; Venkatesh 2008; Whyte 1955). Since these types of incidents are not recorded in official crime data, the use of crime data to identify gang ecologies could misrepresent the locations where gangs are active.

There is also evidence to suggest that when gang members do engage in criminal activity, they tend to engage in a wide range of offenses. This has been described as "cafeteria-style offending" (Klein 1995: 68). Klein found that most gang members are involved in different types of crime, not just violent crime or gun-involved crime. Identifying gang areas using only gun-involved crime may overcome the definitional issues related to gang lists, but it ignores the fact that gang members engage in a wide range of offenses. Furthermore, when gang behavior is approximated using particular types of crime, such as homicide, the researcher is making the assumption that *all* of these types of crimes are related to gang activity. This often is not the case. In 2012, approximately 16% of all homicides in the United States were related to gang activity (Egley, Howell, and Harris 2014). Given the size of the gang population, this represents a substantial percentage of all homicide incidents. However, it is clear from this figure that not all homicides are related to gang activity.

Finally, the results of some studies suggest the non-reporting rate for gang crime may be higher compared to non-gang crime. As the next paragraphs explain, there are a number of reasons why gang crime may be under-reported.

Jankowski (1991) described how gangs need the acceptance of the community in order to operate without attracting the attention of the police. To create this acceptance, the gang operates as an institution that actually strengthens the community and helps to insulate it from crime. Mary Pattillo describes such a scenario in a middle class black neighborhood: "Ironically, having an organized gang in the neighborhood has, in some respects, translated into fewer visible signs of disorder, less violence, and more social control" (Pattillo 1998: 767).

A gang can protect members of their community from crime in a number of ways. One such way is by identifying outsiders who come into the community to victimize residents, then threatening those people to keep them out of the community. Another way the gang can protect the community from crime is by offering to escort residents when they need to go somewhere so they do not have to walk alone. Jankowski reported that 84% of the 31 gangs in his study offered some kind of escort service to residents, such as the elderly, to protect them from becoming victims of robbery. The gangs in this study would also provide security services to local business owners to prevent their stores from being robbed.

Once the gang has acceptance from the community, reported crime may decrease for two reasons: (1) residents report crime to the gang leader who may be able to identify the suspect and mediate the situation in a more efficient and effective way than the criminal justice system, or (2) when the gang is responsible for generating crime, residents chose not to call the police because they know that the gang provides social services to the community. If the police arrest gang members or disrupt the gang, these services go away. These types of social services are also described by Venkatesh (2008) who reported that the gang provided financial support to the community by funding back-to-school parties and barbeques. The gang also provided protection to the community by offering escort services. Overall the gang helped to maintain order within the housing project.

Another reason why gang crime may be underreported is if victims fear the gang will harm them if the police are contacted. Research on fear of gangs is not prevalent. There is evidence, however, that gangs create fear that is separate from the general fear of crime (Lane and Meeker

2000). Felson (2006: 316) includes intimidation in his definition of gangs when he defines a gang as “a very local group of youths who intimidate others with overt displays of affiliation.” Melde and Robinson point out that some researchers have interpreted this to mean that “gangs are only functional, or perhaps even a gang at all, when they provide members with the ability to intimidate others” (Melde and Rennison 2010: 622). If a crime victim is afraid that the gang will harm him/her for reporting a crime to the police, the crime may go unreported. Melde and Rennison (2010) tested the hypothesis that violence by perceived gang members is more likely to go unreported to the police. Using a sample of 29,511 cases from the National Crime Victimization Survey, results show that this may be true for some types of crime, but not others. Victims of robbery were less likely to report the crime if they believed the suspect was a gang member compared to victims who did not believe the suspect was in a gang. Victims were not, however, less likely to report incidents of simple assault, aggravated assault, or rape.

Lastly, violent crime by gang members may not be reported if the violence is related to defending the family’s honor (Horowitz 1983; Horowitz 1987). In communities where defending honor is important, violence may go unreported when the use of violence is viewed as necessary. Some research has shown that parents who are aware of their child’s delinquent behavior are be embarrassed to call the police. “…to call in outside agents (police) to control gang violence is not a viable solution because to question publicly a son’s moral character is to question the honor of the family” (Horowitz 1987: 449). In her study, Horowitz found that parents would prefer to send their children to live with an out-of-state relative than report their child to the police. Parents justify this decision because they feel that their child will “grow out of it” arguing “boys will be boys” (Horowitz 1987: 444).

Summary

This section summarized the theoretical relevance of qualitative and quantitative methods that have been used by researchers to identify gang ecologies. Qualitative methods include identifying gang ecologies through interviews, focus groups or social observation. Quantitative methods include geocoding the locations where gang members live or where they commit crimes and then aggregating those locations to spatial units. This has also been done using types of crime that are associated with gang activity, such as shooting locations or the locations of gun-involved crime more generally, when a list of documented gang members is not available.

While each of these indicators are different, with varying strengths and weaknesses, the stance taken here is that researchers do not intend to operationalize a new construct with each indicator they use. Many researchers who study gangs from an ecological perspective acknowledge the limitations of the indicator they chose to adopt in their study (see Table 1). Sometimes researchers discuss the limitations of other indicators and describe how the indicator they chose to adopt is superior to these other methods. They do not, however, argue that their indicator captures a *different* construct. For instance, Katz and Schnebly (2011) advocate for the use of home address data which "[more] accurately reflect[s] the realities of neighborhood-level gang problems" compared to binary indicators of gang presence (Katz and Schnebly 2011: 386). Curry and Spergel (1988) promote the use of gang homicide data to gauge the extent of gang problems in communities: "A key methodological issue is the operational definition of the gang problem. With a variety of options available, some empirical if not theoretical consensus exists that the measure should be based on official crime statistics" (Curry and Spergel 1988: 384). There are no studies, to date, which directly compare each of these methods against each other as this dissertation does.

Such comparisons are needed because various methods may generate disparities in empirical results and significantly different estimates of gang ecologies.

Construct validity

Gang ecologies are difficult to observe or measure directly. The previous section described a number of indicators that gang researchers have used to measure gang ecologies. It is important to remember, however, that these indicators do not define what gang ecologies are, rather they are simply measures of a gang construct.

Constructs are abstract ideas. Only through a process of operationalization, followed by assessment can they be studied empirically (Maxfield and Babbie 2011: 122; Taylor 1994: 40-41). “Scientists do empirical studies with specific instances of units, treatments, observations, and settings; but these instances are often of interest only because they can be defended as measures of general constructs” (Shadish, Cook, and Campbell 2002: 65).

Constructs help researchers interpret the empirical connections observed between different indicators to theory. Constructs connect these operations used in a study to the language of policymakers thereby allowing them to use the results to inform practical action; this is precisely why construct validity is important (Shadish, Cook, and Campbell 2002: 65). When construct validity problems arise, they call into question what the results of studies really mean and what the implications are for policy (Goldkamp 2010).

Confusion about the links between measurements of constructs and the constructs themselves has been noted in community criminology (Taylor 2011). For example, many studies have made claims of testing social disorganization theory but have failed to incorporate measures of social disorganization concepts in their analyses (Taylor 2011: 76-77). Similarly, it is possible that, *measures* of gang ecologies have been confused with *definitions* of gang ecologies. Empirical indicators of gang ecologies do not define what gang ecologies represent. Conceptualization and operationalization are distinct operations (Taylor 1994: 40-41). Operationalized indicators represent chosen ways out of many possible to capture features of the underlying construct.

Possible sources of invalidity in gang research

The validity of gang ecology indicators deserves investigation. Gangs have been described as a complex and dynamic phenomena (Hughes 2005) riddled with definitional problems (Ball and Curry 1995). Hughes (2005) has suggested that indicators of gang ecologies may be too narrow and may fail to capture other dynamics and dimensions within the larger gang ecology domain. More broadly, this is a potential threat to construct validity known as "construct underrepresentation" (Messick 1995: 742). This occurs when the assessment tool "is too narrow and fails to include important dimensions or facets of the construct" (Messick 1995: 742). The purpose of this study is to investigate if such dimensions Messick mentions exist within the gang ecology domain. If different gang ecology indicators prove somewhat independent and link to gang set space in different ways, this may be because they tap into different facets of gang ecologies.

Related to construct underrepresentation is the idea of content validity. Content validity concerns how a well a variable measures the more general domain of interest (Taylor 1994). When a general domain has sub-areas within it, it is important that a variable measuring that domain is able to capture all of these sub-areas. With respect to the measurement of gang ecologies, content validity refers to the ability of different indicators, such as gang member home address locations or gang arrests, to represent the entire gang ecology domain. If these indicators *do* measure the entire domain, they will identify gang ecologies in similar ways. If they do *not* measure the entire domain, each indicator may actually be measuring sub-areas or dimensions within the domain.

Another potential threat to construct validity is something referred to as "construct-irrelevant variance" (Messick 1995: 742). This happens when the measurement tool either makes it too difficult to identify gang areas by setting the standards too high, or by making it too easy to identify gang areas by setting the standards too low. Messick (1995: 742) describes these concepts as "construct-irrelevant difficulty" and "construct-irrelevant easiness." Evidence of this in gang research can be found in the work of Decker & Kempf-Leonard (1991) whose results indicate that police and juveniles have very different definitions and criteria for gangs. Police consistently used a more narrow definition of gang membership and gang activity and were less likely to describe a scenario as "gang related" compared to a juvenile sample of gang and non-gang members. For instance, when asked how many gang members there were in the city of St. Louis, the mean response of police officers was 438, while the mean response of self-identified gang members was 2,553. It is possible that data generated by law enforcement agencies exhibit some forms of "construct-irrelevant difficulty" as described by Messick (1995) and under-represent gangs, while data generated by gang members exhibit some forms of "construct-irrelevant easiness" which leads to the over-representation of gang prevalence. Similarly, using arrest data to define gang membership may be too narrow, creating construct irrelevant difficulty, especially after considering definitions used by other gang scholars, such as Thrasher or Short, which do not require a person to have an arrest record to be classified as a gang member.

Summary

Ecological research on gangs has received less attention compared to research at the individual level. Gangs have been described as a complex and dynamic phenomena (Hughes 2005) that are defined in a number of ways (Decker and Kempf-Leonard 1991) which makes operationalizing gangs at the ecological level a difficult task. The limited numbers of studies that use ecological data rely on a variety of methods and data sources to quantify gang ecologies. Nobody, to date, investigated the construct validity of these indicators. The findings of prior research that find the spatial distribution of gang related crime, gang set space and gang member home address locations may be different suggest such an assessment is necessary.

This study tests the ability of each of these data sources to measure the entire domain of gang ecologies (Taylor 1994). If indicators generate different results, this could be interpreted as evidence of a content domain sampling issue where individual indicators may not adequately represent the entire conceptual domain of gang ecologies. If some degree of construct relevant variance is present in each indicator, but there is also some construct irrelevant variance, there will be some degree of spatial congruence between methods; the overlap, however will not be perfect.

This study helps us understand the extent to which different indicators of gang ecologies generate similar results. This, in turn, may help future researchers decide if the method used to identify gang locations should be selected based on the particular goals of the study, or alternatively, if all methods are equally appropriate. This study may produce evidence that gang ecologies are too complex to be adequately measured using a single data source. If this is the case, research designs that make use of multiple data sources, or a mix of both qualitative and quantitative methods, may prove to be necessary in ecological gang research.

CHAPTER 3
METHODS

Research questions

Previous ecological studies on gangs have not compared ecological gang indicators against each other. The goal of this study is to determine how similar these indicators actually are. This study will address two main research questions: (1) Do different indicators of gang ecologies reflect the same construct? (2) Can indicators predict the presence versus absence of gangs (binary gang outcome) as well as continuous measures of gang ecologies, such as the number of gangs present or the geographic extent of gang set space?

Both of these questions will be addressed using focus group data generated by practitioners as the criterion variable. These data represent the cumulative knowledge of practitioners who interact with gang members in a variety of capacities. Of course, the focus group data may not identify gang locations with 100% accuracy. Nevertheless, for the purposes of this project, these data at least reflect the situation as perceived by a collective of law enforcement practitioners.

Study site

The study site is Philadelphia, PA. Philadelphia had an estimated population of 1,560,297 in 2014 (United States Census Bureau 2015). Of those people, approximately 35.6% identify as white non-Hispanic, 41.2% identify as black non-Hispanic and 13.6% identify as Hispanic. Compared to the national average of $53,657 the median household income in Philadelphia is just $39,043 and nearly 27% of people live below the poverty line (United States Census Bureau 2015).

Philadelphia experiences exceptionally high levels of crime. Based on the most recent city level crime data published by the FBI, Philadelphia ranks 5th in the number of homicides experienced in 2013 behind Chicago, New York, Detroit and Los Angeles. After controlling for the residential population, Philadelphia's homicide rate ranks 15th (15.90 homicides per 100,000 people) among cities with a population of 250,000 or more (Uniform Crime Reports 2015). A number of programs have been implemented in Philadelphia to reduce crime—particularly violent crime. In recent years, these programs include, among others, Operation Safe Streets (Lawton, Taylor, and Luongo 2005) and Cure Violence (Butts, Roman, Bostwick, and Porter 2015). Additionally, in 2013 Mayor Nutter announced the Youth Violence Prevention Strategic Plan: a federally supported program to reduce youth violence in the 22nd Police District through the use of a multi-disciplinary taskforce (Nutter 2013).

Researchers have studied gangs in Philadelphia using data from the Philadelphia Police Department for decades (Ley and Cybriwsky 1974; Twist 2013). In 2013, the Philadelphia Police Department's Criminal Intelligence Unit (CIU) revamped their data collection efforts and began to solicit information on gangs and gang members from each district Captain on a quarterly basis. Prior to this, if district Captains provided gang information to CIU, it was provided to them sporadically. Soon after 2013, a gang audit was instituted as part of a research strategy to evaluate one of the violence reduction programs in the city. The research strategy was designed as a research-practitioner partnership. Audits were conducted jointly by Temple University researchers and the Philadelphia Police Department's Criminal Intelligence Unit (CIU).

Data sources

The data for this study were supplied by the Philadelphia Police Department (PPD). These data include a list of known gang members and gang set space boundaries supplied by the Criminal Intelligence Unit (CIU). Four years of gang member arrest records and citywide incident records (from January 1, 2012 through December 31, 2015) are also used as data sources.

Two sources of data, the list of gang members and PPD set space locations, were collected as part of a city-wide gang audit. Starting in the summer of 2013, PPD conducted a gang audit throughout the City of Philadelphia. Gang officers assigned to CIU vetted intelligence to identify every gang, their set space, and all known gang associates throughout the city. Every four months since the audit was completed, CIU updates the gang database using information supplied by each of the 21 police districts. A packet of information summarizing the current knowledge on each gang operating within that district is delivered to each district captain. Captains are asked to update the existing information if any new information has been vetted since the last quarter. Specifically, captains are asked about gang set space boundaries and recent gang activity. New information is then validated by the intelligence unit's five-squad before it is entered into the database. Five-squad is a type of unit within the Philadelphia Police Department. These units are generally comprised of 15-25 veteran officers who are used in a variety of capacities under the discretion of the supervising captain. Validated data include the name of each gang, the gang set space boundaries and the type of gang (specified as gang, corner drug sales, or a drug trafficking organization).

Over the course of 17 months, from June 2014 through October 2015, focus groups were organized by Dr. Caterina Roman to further validate and update existing gang information. Focus group meetings were held at the Philadelphia Police Department and lasted approximately four hours. Separate meetings were scheduled for each of the six police divisions in the city. Officers at all ranks in the Philadelphia Police Department, from patrol officers through captains, were asked to attend these meeting to confirm if existing gang information was correct and to fill in the gaps where information was missing. Members of the PPD Intelligence Unit were also present at each meeting. In addition to PPD staff, practitioners from other agencies, including juvenile probation, corrections agencies and the Philadelphia District Attorney's office, were also invited to validate these data.

The meetings began with a brief overview of all of the gangs that currently operate within the police division. Each individual gang was then discussed in more detail to validate the list of documented gang members as well as the geographic boundaries of the gang's set space. Sometimes, these discussions included everyone in the room. Other times, a smaller break out group of 5-6 participants with knowledge of that particular gang would emerge in the middle of the room to discuss the gang under question. When this happened, Dr. Roman would reconvene the people in the room afterwards to share the information that was discussed in this smaller group.

To validate the list of documented gang members, attendees were able to see the name, date of birth, total number of arrests, and photo (either a booking photo or a state ID photo) of each gang member. The Philadelphia Police Department uses the Pennsylvania definition of a gang, defined as "a formal or informal ongoing organization, association, or group, with or without an established hierarchy that has as one of its primary activities the commission of criminal or delinquent acts and that consists of three or more persons" (18 Pennsylvania § 5131 Recruiting Criminal Gang Member). Each member was discussed individually to determine if they were still active in the gang. Once each member was reviewed, meeting attendees were collectively asked if there were any individuals in the gang that were not yet discussed. Once an exhaustive list of gang members was generated, the gang's set space became the focus of the meeting.

Gang set space was defined as the location where gang members hang out or operate. To validate the gang's set space, a map was projected onto a screen depicting the existing boundaries. The group was collectively asked if these boundaries were in the correct locations. The units of analysis in these discussions were streetblocks and intersections. Streetblocks include both sides of a street between two intersections. After the set space boundaries were validated or updated, the next gang was discussed. These meetings were designed similarly to those described in Roman and Chalfin (2008) and Blasko et al. (2015).

The information in these meetings were recorded by Dr. Roman and three to four research assistants who were seated in different parts of the room. Multiple research assistants were used to record information in side conversations between participants. A set of notes from the meetings were given to a lead research assistant to collate into one final document which was forwarded along to the Philadelphia Police Department's Criminal Intelligence Unit to be further validated and vetted. The list of documented gang members and the set space boundaries used in this study is the information that was vetted by the intelligence unit as a result of this process.

Given the covert nature of gang activity, it is not assumed that the gang data are exhaustive of every gang in Philadelphia, nor is the information assumed to be 100% accurate. The gang audits did not include motorcycle gangs, since those gangs are not place-based. Nevertheless, the multi-agency effort to cross-validate information and the vetting process used by the intelligence unit make it unlikely that the database is missing a significant number of criminally active gangs or gang members that are tied to places.

For the current study, PPD provided a list of 3,996 gang members who belonged to 113 gangs. The size of the 113 gangs ranged from two members to 191 members, with a median of 26 members (see Table 2). Only nine gangs had more than 100 documented members (see Table 3). About 55% of the gangs (n=62) had 10-39 members. Although PPD uses the definition of a gang defined in Pennsylvania Statute Title 18 —defined as "a formal or informal ongoing organization, association, or group, with or without an established hierarchy that has as one of its primary activities the commission of criminal or delinquent acts and that consists of three or more persons"

(18 Pennsylvania § 5131 Recruiting Criminal Gang Member)—the data supplied by PPD included two gangs with fewer than three members.

Table 2: Gang membership descriptive statistics, N=113 gangs

N	Min	Max	Mean	Median	SD
113	2	191	36.95	26	33.80

Table 3: Gang membership frequency table, N=113 gangs

Gang size	Frequency	Percent
more than 100 members	9	8%
50-99 members	21	19%
40-49 members	5	4%
30-39 members	13	12%
20-29 members	22	19%
10-19 members	27*	24%
3-9 members	14	12%
less than 3 members	2	2%
Total	*113*	*100%*

*Note: PPD did not provide the set space location for one gang in the 10-19 membership category; therefore, there were 112 gangs where both membership and set space data were available.

In addition to the list of gang members, PPD also supplied gang set space information for 224 gangs.[5] Of these 224 gangs, about 70% of them were active gangs (n=160 in Table 4). The remainder of the gangs were classified as inactive (n=54) or deactivated (n=10).

Table 4: Gang activity level frequency table, N=224

Gang activity level	Frequency	Percent
Active-hot	78	35%
Active-warm	82	37%
Inactive-cold	54	24%
Deactivated	10	4%
Total	*224*	*100%*

[5] PPD did not supply membership information for all 224 gangs identified in the set space file. Membership information was provided for 112 gangs. These gangs were classified into one of three gang typologies (corner drug sales, drug trafficking organization, street gang, see Table 5 on page 40).

PPD also classifies gangs into one of three typologies, to include corner drug sales, drug trafficking organizations or traditional gangs. According to PPD, corner drug sale gangs are defined as:

> "Areas (street corners, abandoned lots, etc.) where drug dealers and their buyers congregate and make transactions. This type of illegal activity can lead to other crimes; including crimes of opportunity and quality of life issues. These investigations are primarily handled by District NETS officers and the Narcotics Strike Force." (Source: Philadelphia Police Department Criminal Intelligence Unit)

Alternatively, drug trafficking organizations are defined as:

> "Complex organizations with highly defined command-and-control structures that produce, transport and/or distribute large quantities of one or more illicit drugs. These investigations are primarily handled by the Department's IDIS, Narcotics Field Unit and/or Federal agencies." (Source: Philadelphia Police Department Criminal Intelligence Unit)

Finally, traditional gangs are defined as:

> "Groups or associations of three or more persons with a common identifying sign, symbol or name; the members of which individually or collectively engage in criminal activity that creates an atmosphere of fear and intimidation." (Source: Philadelphia Police Department Criminal Intelligence Unit)

Of the 224 gang set spaces, roughly 17% were classified as corner drug sales (n=39), 10% were classified as drug trafficking organizations (n=22), and 23% were classified as street gangs. Half of the 224 gangs were not classified into a particular typology (see Table 5).

Table 5: Gang typologies frequency table, N=224

Gang typology	Frequency	Percent
Corner drug sales	39	17%
DTO	22	10%
Street gang	51	23%
Not classified	112	50%
Total	*224*	*100%*

In addition to the gang database, arrest data and incident data from January 1, 2012 and December 31, 2015 were also used in this study. The arrest data were queried by PPD to include only arrests of gang members. The data supplied by PPD included the location where the gang member lived at the time of their arrest (aggregated to the census block level to protect the identity of the gang member) and the location where the crime took place.

The PPD provided a list of 3,996 gang members who belonged to 113 gangs. Of those 3,996 members, 2,859 of them were arrested between 2012 and 2015. The arrested gang members belonged to 111 different gangs. Many of these gang members only belonged to one gang (n=2,797); however, there were 59 individuals who belonged to two gangs, two individuals who belonged to three gangs, and one individual who belonged to eight gangs. There were 1,137 individuals who were identified as a gang member but were not arrested during the four-year period (2012-2015).

The gang arrest file contained 7,488 arrest records. Narcotics and drug law violations accounted for 47% of all gang arrests (n=3,550). Part I violent crime incidents, which include homicide, rape, robbery and aggravated assault, were the next largest crime type which cumulatively accounted for 17% of all arrests (n=1,284). Other offenses, which included violations of probation and contempt of court, made the third largest category which accounted for 9% of all gang arrests (n=699). Part I property crime, which includes burglary, theft, motor vehicle theft accounted for about 8.6% of all arrests (n=642). Finally, weapons violations also accounted for a sizable proportion of gang arrests (6%, n=470). These frequencies, along with frequencies for other crime types, are summarized in Table 6.

Table 6: Gang arrest crime type frequency distribution, N=7,488 arrest incidents

Crime type	Frequency	Percent
Homicide	87	1.16%
Rape	36	0.48%
Robbery	609	8.13%
Aggravated assault	552	7.37%
Burglary	201	2.68%
Theft	250	3.34%
Motor vehicle theft	191	2.55%
Assault	432	5.77%
Arson	3	0.04%
Forgery	3	0.04%
Fraud	31	0.41%
Embezzlement	4	0.05%
Receiving stolen property	18	0.24%
Vandalism	29	0.39%
Violation of uniform firearms act	470	6.28%
Prostitution	9	0.12%
Other sex offenses	12	0.16%
Narcotics/drug law violations	3,540	47.28%
Offenses against family and children	2	0.03%
DUI	179	2.39%
Public drunkenness	1	0.01%
Disorderly conduct	20	0.27%
Loitering	4	0.05%
Other (contempt of court, probation violation)	699	9.33%
Investigate persons	2	0.03%
Traffic violations	22	0.29%
Hospital cases	1	0.01%
Investigations	80	1.07%
Vehicular accidents	1	0.01%
Total	*7,488*	*100.00%*

In addition to arrest data, crime incident data from the same time period (January 1st, 2012 through December 31st, 2015) are also used in this study. This was done because some researchers do not have access to gang membership data and therefore cannot use gang arrests or residence information to study gangs from an ecological perspective. Even when membership information is available, the use of crime incidents related to gang activity overcomes the issues with maintaining gang lists as was discussed on page 25. Incident data were queried to include all gun incidents. Again, these incidents were not restricted to gang members; all gun incidents recorded in Philadelphia during the time period were included. Gun incidents included any homicide, robbery or aggravated assault incident that involved a firearm, as well as weapons violations. A list of PPD UCR codes that meet this definition can be found in Appendix A.

From 2012 through 2015, there were 26,865 gun incidents in Philadelphia. A summary of gun crime frequencies appears in Table 7. The number of gun involved incidents in Philadelphia declined over the time period. In each year, robbery incidents are the most prevalent crime category. Of the 26,865 gun incidents that were reported to the police, 11,079 (41%) of those incidents resulted in an arrest. Incident locations, regardless of whether or not an arrest was made, are used to model gun crime incidents in this study.

Table 7: Gun crime frequencies by year, 1/1/12 – 12/31/15

	2012	2013	2014	2015
Homicide	277	199	211	230
Robbery	3,436	3,172	3,099	2,887
Aggravated assault	2,479	2,325	2,140	2,280
Weapons violation	1,132	980	1,074	944
Yearly totals	*7,324*	*6,676*	*6,524*	*6,341*

Units of analysis

Previous studies that investigate gangs from an ecological perspective quantify gang areas at different spatial units. These units are as big as cities (Decker and Pyrooz 2010; Pyrooz 2012; Pyrooz, Fox, and Decker 2010) or as small as 150 meter grid cells (Block 2000).

The unit of analysis chosen to represent gang ecologies is an important methodological and theoretical decision which has implications for how gangs are conceptualized and studied. Tita and colleagues (2005) found that gang set space areas are often smaller than census tracts, suggesting that smaller units of analysis are more appropriate than large units of analysis when it comes to approximating gang set space. This is not to say, however, that gangs are not relevant at larger units of analysis. Gangs may link to neighborhood level or city level dynamics. For these reasons, the current study compared gang ecologies across a variety of spatial units.

Previous ecological studies involving gangs have been conducted at a range of spatial scales. The spatial scales include:

- 150 meter grid cells (Block 2000),
- thiessen polygons around street intersections (Taniguchi, Ratcliffe, and Taylor 2011),
- census block groups (Blasko, Roman, and Taylor 2015; Rosenfeld, Bray, and Egley 1999; Tita, Cohen, and Engberg 2005; Tita and Radil 2011; Tita and Ridgeway 2007),
- census tracts (Hall, Thornberry, and Lizotte 2006; Katz and Schnebly 2011),
- neighborhood clusters (Chicago; Papachristos and Kirk 2006),
- community areas (Chicago; Curry and Spergel 1988) ,
- police precincts (Detroit; Bynum and Varano 2003), and
- cities (Decker and Pyrooz 2010; Pyrooz 2012; Pyrooz, Fox, and Decker 2010).

In this study, each of these spatial units were explored except for the city level since data were only available for Philadelphia. Various spatial units were included because gang ecologies may reflect different constructs at different spatial scales, and therefore, the indicators of those constructs would potentially be different. Gang ecologies at the street segment level may take on a different meaning than gang ecologies at the community or city level. For instance, the locations where gang members live may prove to be a valid measure of gang ecologies at micro spatial scales (grid cells and thiessen polygons) but they may not function as an adequate representation of gang ecologies at larger spatial scales, such as the census tract or the neighborhood level. Since the gang ecology construct may take on a different meaning at different spatial scales, six different spatial scales are examined here. These include: grid cells, thiessen polygons, census block groups, census tracts, neighborhood clusters and community areas. Each of these spatial units are described next.

In this study, census geographies (census block groups and census tracts) reflect the boundaries at the time of the 2010 census. In Philadelphia, there are 1,336 census block groups and 384 census tracts in the city. Using the five-year estimates from the 2013 American Community Survey, census block groups in Philadelphia had an average residential population of 1,150 people. Philadelphia census tracts are of course much larger with an average residential population of 4,000 people.

The grid cells and thiessen polygons that are used in this study were created using ArcGIS. Grid cells were created using the *create fishnet* tool to generate 500ft by 500ft cells. To completely cover the geographic boundaries of the City of Philadelphia, 16,419 grid cells were created. Thiessen polygons identify all locations that are closest to a given point. Following previous research (Taniguchi, Ratcliffe, and Taylor 2011), thiessen polygons were created around street intersections. Highways, highway ramps, footpaths and driveways were excluded from the analysis.[6] The *dissolve* tool in ArcGIS was used to collapse streets with the same name into one feature class, then the *intersect* tool was used to generate a point dataset that identified each street intersection. Using these intersection points, the *create thiessen polygons* tool was used to generate the thiessen feature class. This process generated 22,396 thiessen polygons. Since the street

[6] To eliminate these types of street segments, streets were queried using the 'Class' field in the street network dataset. Streets that were coded as 2, 3, 4, or 5 (n=39,540) were retained to generate the thiessen polygons.

network in the center city neighborhood is extremely dense, the thiessen polygons in this area of the city are actually smaller than the 500ft by 500ft grid cells.

In Chicago, gang research at the neighborhood level has also been conducted at two relatively unique spatial units; these include groups of census tracts referred to as neighborhood clusters (Papachristos and Kirk 2006), and larger geographies known as community areas (Cartwright and Howard 1966; Curry and Spergel 1988). To create the neighborhood clusters in Chicago, 847 census tracts were grouped together to generate 343 neighborhood clusters. Each neighborhood cluster has a population of about 8,000 Chicago residents. Each neighborhood cluster is "relatively homogeneous with respect to racial/ethnic mix, socioeconomic status, housing density, and family structure" (Sampson 2012: 79). The 77 community areas however, are much larger with an average residential population of about 40,000 people; these boundaries reflect areas that are socially meaningful to residents and "have both political force and symbolic value that have been reinforced over time in a kind of self-fulfilling prophecy" (Sampson 2012: 78). Using similarly sized geographies, comparable units of analysis can be defined in Philadelphia.

There is no universally agreed upon definition of what constitutes a neighborhood (Hunter 1979). Since the City of Philadelphia does not formally recognize neighborhood boundaries, an official list of neighborhoods and their boundaries does not exist. There are a number of ways neighborhood boundaries can be delineated in a similar way that was done in Chicago. In this study, I relied on two datasets that define Philadelphia neighborhoods in different ways. Both datasets were informed by culturally and historically defined neighborhoods in Philadelphia.

To approximate geographies similar in size to the neighborhood clusters used in Chicago, a dataset of 158 Philadelphia neighborhoods was used. These neighborhood boundaries were generated as the result of a project between Azavea Inc. and the Philadelphia Department of

Records. Neighborhood boundaries that were originally delineated in 2006 by the Philadelphia Department of Records were updated by Azavea Inc. in 2013 to reflect the most current culturally defined neighborhoods. This was done using data published on community association websites and maps published by other local organizations. These data are publicly available on opendataphilly.org and ArcGIS Online.

Census data were used to estimate the residential populations of the 158 neighborhoods.[7] Each neighborhood has an average population of 9,700 people. The Chicago neighborhood clusters had an average population of about 8,000 people. Although the Philadelphia neighborhoods have a slightly larger average residential population compared to the Chicago neighborhood clusters, these neighborhood boundaries will be used because they are the closest comparable dataset.

To approximate the larger community areas (areas with a residential population of about 40,000 people) a dataset of 45 Philadelphia communities was used. This dataset is maintained by the Philadelphia Public Health Management Corporation and is used to administer a biannual survey to Philadelphia residents. Each community area has an average population of 33,000 people which make them similar in size to the 77 community areas used in Chicago. The 45 community areas in Philadelphia have been used across a variety of disciplines to include not only in the field of public health but the fields of sociology (Mennis, Dayanim, and Grunwald 2013) and criminology as well (Garcia, Taylor, and Lawton 2007; Mennis, Dayanim, and Grunwald 2013; Mennis, Harris, Obradovic, Izenman, Grunwald, and Lockwood 2011).

[7] Neighborhood populations were estimated by proportionally allocating census data collected in the American Community Survey five year estimates (2009-2013) using data at the census block group level. The same process was used to estimate the population of the 45 community areas in Philadelphia which are described in the next paragraph.

Variables

Academic researchers and practitioners currently use a number of indicators to define gang ecologies. The current study compares four such indicators: (1) the home address of gang members, (2) gang arrests, (3) the location of gun-involved crime incidents, and (4) qualitative methods that rely on law enforcement intelligence, or knowledge provided by criminal justice practitioners (such as probation officers or street workers), or community residents. To determine if each of these indicators are valid measures of the larger gang ecology domain, gang ecologies in Philadelphia were identified using all four of these indicators originating from the gang audit data and official police data from 2012-2015.

Data are aggregated to six different units of analysis to determine if the indicators of gang ecologies are consistent across units of analysis. These units include: 500' by 500' grid cells, thiessen polygons around street intersections, census block groups, census tracts, neighborhood clusters and community areas.

Dependent variable

The dependent variable in this study was modeled using the qualitatively derived PPD gang set space locations. As described earlier, gang set space locations were identified using focus group meetings that included practitioners who work in law enforcement, the court system and corrections agencies in Philadelphia. Gang locations were primarily identified at the streetblock level, but sometimes included street corners.

In using the set space data as the criterion variable I am making three assumptions: (1) that the data are relevant in defining the entire domain of gang ecologies, (2) that they are reliable, and (3) that they are practical for decision making (Taylor 1994: 143-144). These data have been validated by practitioners from a variety of agencies who work with gang members in a variety of capacities which lends support to the relevance and practicality of the data. Since the focus group data represent the cumulative knowledge accessible to law enforcement and have been critically analyzed by PPD staff of all ranks, the data are assumed to be valid for the purposes of this dissertation. Of course, there are several reasons why the dataset may not be valid, despite the extensive vetting process and cross-validation by practitioners from different agencies. A discussion of the limitations surrounding the use of official police data to study gang locations begins on page 161 in the limitations section.

Previous published studies that aggregate gang set space to other units of analysis do not specify *how much* of a unit of geography needs to be covered by gang set space to classify that unit of analysis as a gang area. Tita and colleagues dichotomized gang set space as locations that contain some set space or locations that do not contain set space. They do not, however, describe what criteria were used to quantify "some set space" (Tita, Cohen, and Engberg 2005: 282). In later work, Tita and Ridgeway provide descriptive statistics on how many census block groups in Pittsburg were classified as "carriers of set space" but they do not explain what cutoff point was used for classifying a census block group as such (Tita and Ridgeway 2007: 218). Blasko and colleagues modeled the influence of gangs using two methods: one psychological, one ecological. The ecological variable was defined as spatial units that "straddled or touched" a census block group (Blasko, Roman, and Taylor 2015: 24). A specific degree of overlap, however, was never mentioned in any of the above mentioned studies. The articles suggest that *any* amount of overlap, even just one square foot perhaps, is enough to classify a geographic unit as containing gang presence.

Gang ecologies have traditionally been dichotomized by separating geographies that don't have any gang activity (coded as 0) from locations that have any degree of gang activity (coded as

1). This dichotomy does not recognize the range of values *within* the gang classification. Some gangs are large, other gangs are small. If there is a variation *within* the gang classification, it is possible that indicators reflecting the spatial extent of one or more gangs within a geography are different from the indicators that measure gang ecologies as a binary outcome, meaning they either do or do not contain set space for one or more gangs.

To examine the variation within the gang classification, the outcome variable in this study is modeled at each of the six units of spatial aggregation—grid cells, thiessen polygons, block groups, tracts, neighborhood clusters, and community areas—as a continuous variable representing the *proportion* of the area that is covered by gang set space, which is different from what has been done previously. This operationalization attempts to measure gangs in terms of ecological predominance within a spatial unit, or ecological magnitude, and not merely whether they exist in a location or not.

The outcome variable in the current study was also modeled at each of the six units of spatial aggregation as a continuous variable representing the *number* of gangs also that are present within a geography. This is another way to measure gangs in terms of magnitude.

The Philadelphia Police Department provided me with membership data for 113 gangs. They also provided set space information, which contained information for 224 gangs. Those 224 gangs were queried to only include gangs with at least three known gang members. This was done to match the definition of a gang used by the state of Pennsylvania. Applying this criterion reduced the 224 gang file down to 110 gangs. Of these 110 gangs, about 93% of them were considered to be active gangs (Active-Hot + Active Warm in Table 8). The remainder of the gangs were classified as inactive (n=7) or deactivated (n=1).

Table 8: Gang activity level frequency table, N=110 gangs each with at least three known gang members

Gang activity level	Frequency	Percent
Active-hot	53	48%
Active-warm	49	45%
Inactive-cold	7	6%
Deactivated	1	1%
Total	*110*	*100%*

Of the 110 gangs with a membership of three or more members, roughly 35% were classified as corner drug sales (n=39), 20% were classified as drug trafficking organizations (n=22), and about 45% (n=49) were classified as street gangs (see Table 9).

Table 9: Gangs typology frequency table, N=110 gangs

Gang typology	Frequency	Percent
Corner drug sales	39	35%
DTO	22	20%
Street gang	49	45%
Total	*110*	*100%*

Descriptive statistics for the dependent variable, modeled as a proportion based on the size of the gang in terms of area, are listed in Table 10. It is evident that this variable is constrained to the [0,1] interval and is not normally distributed. Frequency tables of these proportions are found in Table 11. These tables demonstrate that a majority of locations do not contain gang territories (proportion = 0). This pattern generally holds at each of the six units of analysis.

Table 10: Dependent variable modeled as a proportion, descriptive statistics

Spatial unit	N	Min	Max	Median	Mean	SD	Skewness	Kurtosis
Grid cells	16,419	0	1.00	0	0.03	0.15	5.63	34.36
Thiessens	22,396	0	1.00	0	0.06	0.21	3.63	15.07
Block groups	1,336	0	1.00	0	0.07	0.21	3.34	13.46
Tracts	384	0	1.00	0	0.06	0.16	3.48	16.25
Neighborhood	158	0	0.70	0	0.05	0.12	3.31	15.10
Community areas	45	0	0.46	0.01	0.05	0.08	3.06	15.05

Note: Dependent variable is measured as the proportion of each spatial unit that is covered by gang set space.

Table 11: Dependent variable modeled as a proportion, frequency table

Proportion gang area	Grid cells		Thiessens		Census block groups		Census tracts		Neighborhoods		Community areas	
0	15,449	*94.1%*	20,386	*91.0%*	974	*72.9%*	234	*60.9%*	88	*55.7%*	15	*33.3%*
< = .09	225	*1.4%*	87	*0.4%*	197	*14.7%*	89	*23.2%*	39	*24.7%*	21	*46.7%*
.1-.19	109	*0.7%*	85	*0.4%*	28	*2.1%*	16	*4.2%*	18	*11.4%*	8	*17.8%*
.2-.29	70	*0.4%*	128	*0.6%*	21	*1.6%*	17	*4.4%*	6	*3.8%*	0	*0.0%*
.3-.39	68	*0.4%*	127	*0.6%*	10	*0.7%*	7	*1.8%*	1	*0.6%*	0	*0.0%*
.4-.49	52	*0.3%*	171	*0.8%*	11	*0.8%*	7	*1.8%*	2	*1.3%*	1	*2.2%*
.5-.59	49	*0.5%*	259	*1.2%*	27	*2.0%*	5	*1.3%*	2	*1.3%*	0	*0.0%*
.6-.69	36	*0.2%*	207	*.9%*	17	*1.3%*	3	*0.8%*	1	*0.6%*	0	*0.0%*
.7-.79	45	*0.3%*	114	*0.5%*	12	*0.9%*	3	*0,8%*	1	*0.6%*	0	*0.0%*
.8-.89	33	*0.2%*	66	*0.3%*	3	*0.2%*	0	*0.0%*	0	*0.0%*	0	*0.0%*
.9-.99	64	*0.4%*	43	*0.2%*	4	*0.3%*	1	*0.3%*	0	*0.0%*	0	*0.0%*
1	219	*1.33%*	723	*3.2%*	32	*2.4%*	2	*0.5%*	0	*0.0%*	0	*0.0%*
Total	16,419	*100%*	22,396	*100%*	1,336	*100%*	384	*100%*	158	*100%*	45	*100%*

Note: The dependent variable is measured as a proportion of the spatial unit that is covered by gang set space. Frequencies and column percentages are displayed.

Table 11 demonstrates that a large share of the dependent variable reflects cases where there is no gang present within the geography (proportion = 0). If values of zero are excluded, the distribution of the dependent variable changes drastically. In Table 12, cases that do not contain any gang set space (proportion=0) have been *removed* from the dataset resulting in a much smaller sample. These values are bound on the (0,1] interval. Removing the cases with a proportion value equal to zero improved skewness and kurtosis values.

Table 12: Dependent variable modeled as a proportion excluding values of zero, descriptive statistics

Spatial unit	N	Min	Max	Median	Mean	SD	Skewness	Kurtosis
Grid Cells	970	1.88E-6	1.00	0.41	0.49	0.39	0.18	1.40
Thiessens	2010	6.65E-7	1.00	0.65	0.67	0.31	-0.37	1.91
Block groups	362	3.81E-6	1.00	0.06	0.26	0.33	1.15	2.93
Tracts	150	1.36E-5	1.00	0.05	0.16	0.22	1.90	6.37
Neighborhoods	70	1.48E-4	0.70	0.07	0.12	0.15	2.08	7.17
Community areas	30	5.44E-4	0.46	0.05	0.07	0.09	2.65	11.77

Although the distribution of the proportion variable improves when values that are equal to zero are removed, it may not be advantageous to remove those values. The zero values seem to be a dominant feature of the dataset. This may be because the functional role, or niche,[8] of a geography that does not contain gang set space is qualitatively different from a geography that is partially covered by gang set space. If this is true, the variables driving the spatial patterning of values that are zero may be different from the variables that are driving the patterning of non-zero values. Stated differently, the variables that describe the *presence* of gangs and the variables that describe the *magnitude*, or ecological spatial predominance, of gang presence may be two different sets of variables. To address this in the current study, both distributions are modeled using a

[8] For a discussion of the functional niche of a community and how it relates to human ecology, see Hawley, Amos H. 1950. *Human Ecology: A Theory of Community Structure*. New York: Ronald Press.

regression technique that creates separate parameter estimates for each distribution. These models will be explained on page 74.

The dependent variable is also modeled as a count variable. This was done to measure gang ecological intensity in a slightly different way than the proportion variable. Rather than measuring gangs in terms of the geographic area their set space covered, a count variable was created to measure how many gangs were present within each geography. Descriptive statistics for the dependent variable when it is modeled as a count are listed in Table 13. This variable ranges from zero to 11 gangs and is not normally distributed. Frequency distributions of these counts can be found in Table 14. These frequency distributions demonstrate that a majority of locations do not contain gang set space, as was the case when this same variable was modeled as a proportion. This pattern generally holds at each of the six units of analysis.

Table 13: Dependent variable modeled as counts, descriptive statistics

Spatial unit	N	Min	Max	Mean	SD	Median	Skew	Kurtosis	Variance
Grid Cells	16,419	0	5	0.07	0.29	0	5.42	45.10	0.09
Thiessens	22,396	0	5	0.10	0.35	0	4.58	35.44	0.12
Block groups	1,336	0	5	0.38	0.72	0	2.36	10.11	0.52
Tracts	384	0	6	0.69	1.08	0	1.91	6.85	1.17
Neighborhoods	158	0	8	1.22	1.74	0	1.50	4.73	3.04
Community areas	45	0	11	3.09	3.26	2	0.80	2.53	10.63

Note: Dependent variable is measured as the number of gangs present in each spatial unit. Gangs that are in more than one spatial unit are counted more than once.

Table 14: Dependent variable displayed as counts per spatial unit

Gang Count	Grids		Thiessen		Census block groups		Census tracts		Neighborhoods		Community areas	
0	15,449	*94.09%*	20,386	91.03%	974	72.90%	234	60.94%	88	55.70%	15	33.33%
1	843	*5.13%*	1,791	8.00%	253	18.94%	83	21.61%	17	10.76%	6	13.33%
2	117	*0.71%*	198	0.88%	85	6.36%	38	9.90%	21	13.29%	3	6.67%
3	4	*0.02%*	1	0.00%	18	1.35%	18	4.69%	14	8.86%	2	4.44%
4	2	*0.01%*	9	0.04%	2	0.15%	6	1.56%	7	4.43%	4	8.89%
5	4	*0.02%*	11	0.05%	4	0.30%	4	1.04%	6	3.80%	5	11.11%
6	0	*0.00%*	0	0.00%	0	0.00%	1	0.26%	3	1.90%	4	8.89%
7	0	*0.00%*	0	0.00%	0	0.00%	0	0.00%	1	0.63%	0	0.00%
8	0	*0.00%*	0	0.00%	0	0.00%	0	0.00%	1	0.63%	1	2.22%
9	0	*0.00%*	0	0.00%	0	0.00%	0	0.00%	0	0.00%	3	6.67%
10	0	*0.00%*	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	2.22%
11	0	*0.00%*	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	2.22%
Total	16,419	*100%*	22,396	100%	1,336	100%	384	100%	158	100%	45	100%

Note: The dependent variable is the number of gangs present in each spatial unit. Frequencies and column percentages are displayed.

It is evident from Table 14 that the dependent variable includes a large share of cases where there is no gang present within the geography (gang count = 0). If values of zero are excluded, the distribution of the dependent variable shifts. In Table 15, cases that do not contain any gang set space were removed from the dataset. Removing the cases slightly reduced skewness and kurtosis values. It is evident that the table (Table 14) that even after removing values of zero, cases still cluster at values of 1 and decline as count values increase. The values are still distributed as one would expect for a count variable.

Table 15: Dependent variables displayed as counts excluding values of zero, descriptive statistics

Spatial unit	N	Min	Max	Mean	SD	Median	Skew	Kurtosis	Variance
Grid Cells	970	1	5	1.15	0.45	1.0	4.28	29.17	0.67
Thiessens	2010	1	5	1.13	0.46	1.0	5.05	35.87	0.68
Block groups	362	1	5	1.40	0.71	1.0	2.32	9.83	0.84
Tracts	150	1	6	1.75	1.06	1.0	1.58	5.28	1.03
Neighborhoods	70	1	8	2.74	1.63	2.0	1.05	3.68	1.28
Community areas	30	1	11	4.63	2.95	4.5	0.49	2.30	1.72

Independent variables

Three different independent variables are used to predict the dependent variables. These variables were created using official police data from 2012-2015. The independent variables include the home address of gang members, crime incidents involving gang members and gun-crime incidents. Each of these are discussed below.

Home address of gang members

The home address of gang members has been used as an indicator of gang ecologies (Cartwright and Howard 1966; Katz and Schnebly 2011). In this study, home address data that are recorded when a person is arrested are used to measure where gang members live. PPD queried all gang arrests from 1/1/2012 through 12/31/2015 (N=7,573). To protect the identity of the gang members, PPD aggregated the home address locations to census blocks. The census block feature class was converted to a point file, with each point corresponding to the centroid of each census block polygon. For a short discussion concerning the error introduced into the analysis due to the use of census block centroids over the precise address locations, see page 161. This point file was used to model the home address location of each gang member.

If a gang member was arrested multiple times from 1/1/2012 through 12/31/2015, there are duplicate home address records for that individual in this dataset. When an individual has multiple arrest records, the home address for the most recent arrest was retained. Once these duplicated records were removed, 2,859 records remained in the dataset. Of these 2,859 gang members, 2,735 members had a home address that was geocoded (95.7% geocoding hit rate). This difference reflects the fact that some of the home address locations could not be geocoded or were geocoded to locations outside the Philadelphia city limits. When this happened, or the gang member did not have a home address location specified in the arrest record (if the gang member was homeless, for instance) the home address location did not have a census block associated with it. Once these records were removed, 2,735 addresses remained in the dataset.

Another data set was created which accounts for the fact that some individuals belonged to more than one gang. Of the 2,859 gang members who were arrested between 2012 and 2015, 62 of them belonged to more than one gang. Duplicate home address records were generated so there was a home address record to which each gang an individual belonged. This step was important for the spatial analysis which attempted to create gang set space boundaries quantitatively. For that analysis, gangs are the unit of analysis. When a person belongs to more than one gang, it is important that their information is used for each gang to which each belongs. These duplicated records brought the home address dataset up from 2,735 records (one home address per person) to 2,801 records (duplicate records if gang member belonged to more than one gang).

There are some notable limitations to using arrest records to identify where gang members live. First, this method only included the addresses of gang members who are known to the police. Some gang members fly under the radar of law enforcement and therefore were not in the database. A related issue concerns how the PPD defines gang membership (see page 38) which requires an arrest. Known gang members who have not been arrested were not included in the gang database. This under-represented the number of known gang member residential addresses. Lastly, some of the address data may contain errors that prevent PPD from being able to geocode them.

Gang arrests

Researchers have used gang-motivated homicide locations to define gang ecologies (Curry and Spergel 1988; Papachristos and Kirk 2006). Other types of crime have also been used; Block (2000) used all crime incidents that were determined to be gang-motivated.

In the current study, all crime types were used to model gang-related crime, similar to the Block (2000) study. A member-based definition of gang crime was used to define gang-related incidents instead of a motive-based definition.[9] This means, all incidents that lead to the arrest of a gang member were used to model gang-related crime. Arrest incidents were included regardless of whether or not the incident related specifically to the activities of the gang. This was done because the arrest database maintained by PPD does not identify which arrest incidents are actually motivated by gang activity.

[9] The Philadelphia Police Department maintains a dataset of shooting incidents that identify whether the shooting was gang-motivated or not. In 2014, an initiative began which increased information sharing between the Criminal Intelligence Unit and the police districts. As a result of this, the number of documented gang members, and therefore the number of gang-motivated shootings, increased drastically. These data cannot be updated retroactively. The shootings that were coded prior to 2014 are somewhat limited since they are based on incomplete gang lists.

From 1/1/2012 through 12/31/2015, there were 7,406 crime incidents that lead to the arrest of a gang member that were able to be geocoded; only 51 incidents were not able to be geocoded. This reflects a 99.3% geocoding hit rate. As was done with the home address data, a second data set was created which accounts for the fact that some individuals belonged to more than one gang. Duplicate arrest records were generated to create an arrest record for each gang to which an individual belonged. These duplicated records brought the incident arrest dataset up from 7,406 records (one incident per person) to 7,925 records (duplicate records if gang member belonged to more than one gang).

Gun-crime incidents

Gun-crime has been used as a proxy measure of gang activity when a list of gang members is not available (Bynum and Varano 2003). In this study, gun crime was used to generate a gang ecology indicator without using a list of documented gang members. Incident data were queried to include 26,865 gun incidents from January 1, 2012 through December 31, 2015. These incidents included any homicide, robbery or aggravated assault incident that involved a firearm, as well as weapons violations.[10] A list of PPD UCR codes that meet this definition can be found in Appendix A. A summary of gun crime frequencies appears in Table 16. The number of gun-involved incidents in Philadelphia declined from 2012-2015. In each year, robbery incidents are the most prevalent crime category. These numbers reflect incidents that were reported to the police and do not reflect arrests for these crimes.

[10] Weapons violations of the uniform firearms act are commonly referred to as VUFAs. A person is charged with a weapons violation if (s)he carries a gun on a public street without a firearm license, or if s(he) is a convicted felon in possession of a firearm.

Table 16: Gun crime frequencies, 1/1/12 – 12/31/15

	2012	2013	2014	2015
Homicide	277	199	211	230
Robbery	3,436	3,172	3,099	2,887
Aggravated assault	2,479	2,325	2,140	2,280
Weapons violation	1,132	980	1,074	944
Yearly totals	*7,324*	*6,676*	*6,524*	*6,341*

Gun incidents were geocoded by PPD. Of the 26,865 gun incidents that were reported to the police, 26,625 were able to be geocoded (99.1% hit rate).

Descriptive statistics for each of the three independent variables at each spatial unit are displayed in Table 17.

Table 17: Independent variables, descriptive statistics for all spatial units

Spatial unit	variable	N	Sum	Min	Max	Mean	SD	Median	Skewness	Kurtosis
Grid cell	home address	16,419	2,735	0	25	0.17	0.71	0	9.04	158.70
	gang arrests	16,419	7,406	0	179	0.45	2.81	0	30.09	1451.17
	gun incidents	16,419	26,625	0	49	1.62	3.41	0	3.49	21.49
Thiessen	home address	22,396	2,735	0	11	0.12	0.50	0	6.56	67.30
	gang arrests	22,396	7,406	0	104	0.33	1.81	0	25.68	1063.75
	gun incidents	22,396	26,625	0	37	1.19	2.09	0	3.53	25.18
Census block groups	home address	1,336	2,735	0	41	2.05	3.29	1	3.85	28.52
	gang arrests	1,336	7,406	0	225	5.54	13.92	2	9.24	117.23
	gun incidents	1,336	26,625	0	132	19.93	18.04	15	1.80	8.04
Census tracts	home address	384	2,735	0	73	7.12	10.04	3	2.83	14.16
	gang arrests	384	7,406	0	353	19.29	36.90	8	5.33	39.75
	gun incidents	384	26,625	0	345	69.34	62.46	51	1.21	4.40
Neighborhoods	home address	158	2,735	0	132	17.31	26.07	5	2.29	8.57
	gang arrests	158	7,406	0	740	46.87	85.31	17.5	4.47	31.10
	gun incidents	158	26,625	0	991	168.51	213.45	70.5	1.76	5.56
Community areas	home address	45	2,735	0	258	60.78	53.19	50	1.27	5.28
	gang arrests	45	7,406	1	1280	164.58	202.32	125	3.83	21.66
	gun incidents	45	26,625	64	1808	591.67	444.12	493	0.93	3.20

Control variables

Control variables are used to specify statistical models for two reasons: (1) to satisfy theoretical assumptions, or (2) to satisfy mathematical assumptions (Sweet and Grace-Martin 2010). Theoretical assumptions include the use of control variables to identify the unique effect of independent variables after holding constant the effect of other variables that could influence the dependent variable. Mathematical assumptions include the use of control variables to satisfy underlying assumptions of the regression model. Violations of these assumptions influence the robustness of parameter estimates, so correcting these violations is important.

There are a number of theoretically appropriate control variables that could be used in this study. The use of community structural characteristics could be used to increase the explanatory power of the model (Bursik and Grasmick 1993: see Chapter 5). Including such variables, however, would change how the independent variables are interpreted; results of the model would represent the unique effect of each gang indicator *holding constant* these other variables. The purpose of this study is to identify if these indicators are appropriate measures of gang ecologies *on their own* as is currently done in gang research. To include additional control variables would change the research question this study proposes to answer.

The current study made use of two control variables to address mathematical assumptions of the regression model. To satisfy the assumption of independence of observations, a spatially lagged dependent variable was included in the model. To control for the size of each geographic unit, a variable representing area of each spatial unit was included in the model as an exposure variable.[11]

[11] When variables are treated as exposure variables, they are entered into the model after being log transformed. The coefficient of the log transformed variable is set to 1, meaning a parameter is not estimated for this variable. Since a parameter is not estimated for the exposure variable this increases available degrees of freedom to fit the model. Exposure variables can be defined by using the exposure(x) sub command in Stata or offset(log(x)) in R.

Examining construct validity

Construct validity can be assessed a number of ways. While there are strengths and weaknesses of using different indicators, researchers have not widely accepted the use of one particular assessment over others (Western and Rosenthal 2003). The choice of assessment is largely up to the researcher. The following section will briefly describe how construct validity has been assessed and why Messick's (1995) unified perspective of construct validity was chosen.

One way to look at construct validity is with the multitrait-multimethod matrix proposed by Campbell and Fiske (Campbell and Fiske 1959). This method separates trait variance from method variance and simultaneously tests convergent and discriminant validity. Trait variance simply refers to the existence of multiple traits within a construct, while method variance refers to the ability to measure these traits using different measurement procedures. For instance, let's assume where gang members commit crime and where they live are two distinct traits of gang territories. These two traits can be measured using official police data, but they could also be measured by interviewing gang members. Both traits can be measured using both methods. The multitrait-multimethod matrix allows researchers to compare trait variance and method variance in construct validity assessments. It also can be used to gauge convergent and discriminant validity. Convergent validity is reflected when different measures of the same concept correlate strongly with one another. Discriminant validity is reflected when an indicator of different concepts correlate more weakly than indicators of the same concept.

Self-control research proves illustrative (Gottfredson and Hirschi 1990). The self-control construct has six sub-domains commonly captured using the Grasmick scale (Grasmick, Tittle, Bursik, and Arneklev 1993). This scale includes four measures for each of the six traits of self-control. Since there are multiple traits—six traits within the construct of self-control—and multiple ways to measure each of these traits—four measures for each trait—we have both trait variance and item variance in this example. We would expect that the four measures for one particular trait would produce similar scores, or at least that these four sets of scores would be highly enough correlated to suggest the four measures reflected the same trait. If these correlations are strong, that suggests convergent validity at the sub-domain level within the broader self-control construct. We would also expect that the four measures of one trait, impulsivity for example, would produce *different* scores than the four indicators that measure temper. Impulsivity and temper are two separate traits—or sub-domains within the broader self-control construct— that are conceptually distinct from one another. If the between-trait correlations are weaker than the within-trait correlations, that suggests discriminant validity.

The multitrait-multimethod matrix requires that a researcher is able to measure multiple traits using multiple methods or instruments. For example, in gang research, Decker and Pyrooz (2010) examined correlations between two traits—gang homicides and total homicides—and three data sources—data published by the Uniform Crime Reports, the National Gang Center, and supplemental homicide reports. The supplemental homicide reports and National Gang Center data measuring gang homicide correlated strongly (.94) while correlations between total homicides, measured using UCR data, and gang homicides, using supplemental homicide reports and National Gang Center data, proved to be much weaker (.59 and .58, respectively).

This dissertation relies on a single data type, law enforcement data, to measure potentially different qualities of gang ecologies. The multitrait-multimethod matrix is not a feasible option for measuring construct validity in this study unless other data sources can be used to generate gang indicators. Other methods could include social observation in suspected gang neighborhoods or interviews of gang members and community members. With over 100 documented gangs in the city of Philadelphia, these data collection strategies would take a considerable amount of time for one researcher to complete. Time constraints and lack of funding prevent these methods from being used in this study.

Another way to look at construct validity involves the use of contrast analyses (Western and Rosenthal 2003). Since this method is based on correlation coefficients, it is relatively simple to compute and understand. These correlations, however, may not be appropriate to model gang ecologies if there is a qualitative distinction between places that have gangs and places that do not have gangs. Correlation analyses would not be able to capture what a value of zero (no gang) truly represents in this context because it would assume that the process that is generating the zero values is the same process that is generating non-zero (positive) values. This method may be appropriate when comparing indicators of gang magnitude, but what is needed in gang research is a method of distinguishing places that have gangs from places that do not have gangs while *simultaneously* accounting for differences in the magnitude of the gang problem. In other words, the primary goal of this study is to compare indicators that identify where gangs are and where they are not while acknowledging there may be a continuum of values within the category of gang presence. To do this, regression models with a hurdle component are most appropriate.

Having ruled out a number of approaches, this study relies on Messick's unified perspective on construct validation. Messick advanced the argument that validity is a "unified concept" and discusses six distinguishable aspects of construct validity (Messick 1995). Although the framework was developed for educational and psychological measurement, it has been used to examine ecological construct validity in criminology literature (Blasko, Roman, and Taylor 2015; Taylor 2011). In these studies, a series of multiple regression models were run. To examine various aspects of construct validity, parameter estimates and explained variance (R^2) are compared across models (Blasko, Roman, and Taylor 2015; Decker, Pyrooz, Sweeten, and Moule 2014; Taylor 2011). The details of Messick's perspective and how they apply to this study are described next.

Messick's unified perspective on construct validity

The unified perspective of construct validity proposes that traditional views on validity, which separate construct validity from criterion and content validity, are fragmented. Messick proposed that a unified perspective of construct validity is more appropriate. This unified perspective integrates six distinguishable aspects of validity; these include the (1) content, (2) substantive, (3) structural, (4) generalizable, (5) external and (6) consequential aspects of construct validity. While each of these aspects of validity are relevant to gang ecologies, I was not able to evaluate all six aspects of construct validity as proposed by Messick; however, I was able to evaluate three of them. These three aspects of validity include the content, structural and generalizable aspects of construct validity. Each of the six aspects of construct validity and how they relate to gang ecologies will be described next.

1) **Content**

> "A key issue for the content aspect of construct validity is the specification of the boundaries of the construct domain to be assessed—that is, determining the knowledge, skills, attitudes, motives, and other attributes to be revealed by the assessment tasks" (Messick 1995:745).

The main research question in this study refers to the content relevance and representativeness of ecological gang indicators. The focus of this aspect of construct validity is to ensure that all aspects of the trait (gang ecologies) are included in the measure. If gang ecologies are one-dimensional, then each indicator is redundant and does not add anything new to explain the construct. If this is the case, any of the indicators are sufficient to use on their own to model gang areas.

The empirical relevance and representativeness of each indicator can be examined using zero-inflated beta regression models (these models will be described starting on page 74). If each measure is relevant to gang areas, then each measure will explain variance in the gang outcome variable over and above control variables.[12] If each measure individually captures the representativeness of all characteristics associated with gang ecologies, then the explained variance will be relatively high, and will not drastically improve after adding additional indicators. Improvement will be assessed by considering differences in the Bayesian Information Criterion (Raftery 1995), pseudo R^2 values, and likelihood ratio tests. These metrics are described starting on page 79.

If gang areas are multi-dimensional, it would be important to know which indicator of gangs is the best indicator. Gang research is difficult to conduct as it is, and researchers may not have access to multiple measures of gang ecologies. When this happens, this study could inform choices regarding which indicator best approximates gang ecologies.

2) **Substantive**

> "The substantive aspect of construct validity emphasizes the role of substantive theories and process modeling in identifying the domain process to be revealed in assessment tasks" (Messick 1995: 745).

This aspect of construct validity concerns *how* or *through what process* a respondent arrives at an answer to a question in an assessment. A researcher should have a theoretical rationale for how the respondent generated each response. To use educational testing as an example, a

[12] Since the dependent variable in this study is a proportion, zero inflated beta regression models will be used. Unlike ordinary least squares regression, this model uses maximum likelihood estimation to generate parameter estimates. Models that are estimated using maximum likelihood do not generate an R^2 value to assess explained variance. Instead, these models generate a log likelihood (specifically the difference between 2 log likelihood values multiplied by -2) so the PRE measure for these models are interpreted as "the proportional reduction in the absolute value of the log-likelihood measure, where the -2LL or the absolute value of the log likelihood—the quantity being minimized to select the model parameters—is taken as a measure of "variation", not identical but analogous to the variance in OLS regression" Menard, Scott. 2002. *Applied Logistic Regression Analysis*. Thousand Oaks, CA: Sage Publications, Inc.

question on an exam may require a student to find the area of a circle. A student can arrive at the correct answer a variety of ways: (1) they understand and can apply the formula $A=\square r^2$, (2) they could have memorized the answer from a practice question, or (3) they could copy the answer from a student sitting next to them, or many other possibilities. Is the correct answer in each of these scenarios reflective of the student's understanding of the formula? This aspect of validity is included in the substantive aspect of construct validity. To quantify the domain process in educational testing, student's responses to questions can be timed, or students can be asked to show their work.

In the context of this study, home address locations and crime locations may be relevant to the identification of gang areas for different theoretical reasons. The substantive aspect of validity concerns the process through which these variables are relevant indicators of gang ecology. The process driving the spatial patterning of gang ecologies may or may not be the same process driving the spatial patterning of the indicators we used to identify gang ecologies.

Unfortunately, the data provided by the police department do not contain information allowing specification of processes driving the patterning of gang ecologies. This process could be uncovered qualitatively through the systematic social observation of gang members, or through ethnographic research. Using these methods would take an extraordinary amount of time and resources to gather these data on every gang in the city of Philadelphia. Capacity constraints prevented the analysis of the substantive aspect of construct validity in this study.

3) **Structural**

> "Ideally, the manner in which behavioral instances are combined to produce a score should rest on knowledge of how the process underlying those behaviors combine dynamically to produce effects" (Messick 1995: 746).

This aspect of construct validity is also referred to as structural fidelity. In the context of this study, this applies to how data are aggregated to spatial units. For an example, some researchers have used the number of gang members or the number of crime incidents to model gangs at the ecological level, while others have used a binary measure representing the presence or absence of gangs.

Another structural aspect of construct validity concerns the unit of analysis to which the data are aggregated. The ecological effects of gangs have been analyzed at different spatial units of analysis including 150 meter grid cells, thiessen polygons, census block groups, census tracts, neighborhood clusters, community areas, and cities.

In the context of the current study, this aspect of construct validity has been described by many as the modifiable areal unit problem (MAUP). This problem addresses the two aspects of structural validity discussed above: (1) how data should be aggregated to spatial units and (2) how large the units should be (Weisburd, Bruinsma, and Bernasco 2009: 5). The consequence of the MAUP was tested by Fotheringham & Wang (1991) and Ouimet (2000) who both found that the results change depending on how the data are aggregated. There are two different theoretical frames for explaining why the results differ: the homology assumption and the discontinuity thesis.

Under the homology assumption, the inconsistent results found by Fotheringham & Wang (1991) and Ouimet (2000) were purely attributed to aggregation bias (Taylor 2015). Under this perspective, any differences that are found are simply a "statistical artifact" generated by different spatial units (Hannan 1991: 3). Under the homology assumption, the underlying theory, or process, that explains relationships between variables is relevant at *any* spatial unit. Another perspective, however, might attribute these conflicting results as evidence that different theoretical processes are operative at different spatial scales. Under the discontinuity thesis, different results found by Fotheringham & Wang (1991) and Ouimet (2000) would be expected because relationships and processes change when connections are examined at different spatial units (Taylor 2015).

To address the MAUP issue, the analysis undertaken in this dissertation was run at multiple spatial scales to examine how effects change across spatial units. Variations in scoring were also examined by modeling the outcome as a binary variable as well as a continuous variable.

4) **Generalizable**

> "The concern that a performance assessment should provide representative coverage of the content and process of the construct domain is meant to insure that the score interpretation not be limited to the sample of assessed tasks but be broadly generalizable to the construct domain...setting the boundaries of score meaning is precisely what generalizability evidence is meant to address" (Messick 1995: 746).

The generalizability of gang ecology indicators may be dependent on the type of gang under investigation. In other words, score meaning here may depend on gang type. The gang typologies have been discussed at length (Klein 2004). In the current study, the data supplied by the Philadelphia Police Department do contain data on gang typologies. Each gang is classified as either a traditional gang, a corner drug crew or a drug trafficking organization. It is possible that drug trafficking organizations are more organized and are more successful in flying under the radar of law enforcement compared to traditional gangs or corner drug crews. If this is the case, the members of these types of gangs may not be well represented on official gang lists compared to other types of gang members. Furthermore, if these types of gangs are exceptionally good at avoiding the police, they may not have arrest records that are as extensive as other gang members. Taking these two points together, official crime data may prove to be a better indicator of some types of gang ecologies but not others. Using the PPD typologies, the generalizable aspect of construct validity can be assessed.

Messick's description of generalizability concerned the degree to which scores are representative of the broader construct domain and whether the results are generalizable across tasks, across time and across observers. External validity, which is related to Messick's description of generalizability, also considers the extent to which findings are generalizable across settings. It is possible that relationships between the variables included here depend heavily on the Philadelphia context. Philadelphia gangs have been described as "less organized" with more fluid membership compared to gang in other cities such as Boston and New York (Fader 2013: 31). Perhaps gang structures within Philadelphia vary more than they do in other cities. Data from other cities would allow investigating the degree to which Philadelphia gang ecology linkages apply to other cities. Hopefully, future researchers with access to gang ecology indicators from other cities will assess the external validity of what is found here.

5) **External**

> "Both convergent and discriminant correlation patterns are important, the convergent pattern indicating a correspondence between measures of the same construct and the discriminant pattern indicating a distinctness from measures of other constructs" (Messick 1995: 746).

The multitrait-multimethod matrix proposed by Campbell and Fiske (1959) is a popular method to test for convergent and discriminant validity. To test this aspect of construct validity, a researcher must have access to at least two methodologies or measurement tools. In order to use the multitrait-multimethod matrix to test convergent and discriminant validity, another method of identifying features of gang ecologies is needed. As was previously mentioned, such methods are not available, precluding the assessment of convergent multi-method validity in this study.

6) **Consequential**

> "The consequential aspect appraises the value implications of score interpretation as a basis for action as well as the actual and potential consequences of test use, especially in regard to sources of invalidity related to issues of bias, fairness, and distributive justice"(Messick 1995: 745).

The consequence of identifying a location as a gang area can include increased law enforcement attention in that area. According to the consequential aspect of validity, this increased attention from law enforcement should not have any bearing on how gang ecologies are identified in the future.

If appropriate data were available, this aspect of validity could be tested using a longitudinal dataset. When a gang area is identified, the police may decide to implement an initiative that is designed to dismantle the gang. As a result of the initiative, the police might exclude this gang from future audits because as far as they know, the gang has successfully been dismantled. The effect of that initiative on the future identification of active gang areas is an example of consequence as validity evidence.

The effect of law enforcement actions on the identification of gang areas cannot be assessed in the current study since the data are cross-sectional. Future research could use longitudinal data to examine these effects.

Summary

In the current study, I was not able to evaluate all six aspects of construct validity as proposed by Messick; however, I was able to evaluate three of them. These three aspects of validity include the content, structural and generalizable aspects of construct validity. Gang research at the ecological level has been conducted at spatial units that vary in size. This is problematic because many researchers have suggested that different spatial scales may be conceptually and empirically distinct from each other (Fotheringham and Wong 1991; Ouimet 2000; Taylor 2010; Weisburd, Bruinsma, and Bernasco 2009). Given these findings, the structural aspect of construct validity is particularly important to this area of research.

Analysis plan

Two different analyses are proposed to answer the two main research questions: (1) Do indicators of gang ecologies identify the same construct? (2) Do indicators predict gang presence in the same way they predict gang magnitude?

The first analysis uses a series of regression models to answer both research questions and assess various aspects of construct validity as proposed by Messick. The second analysis uses an algorithm to approximate gang set space boundaries that were defined in the gang audits. The degree to which these approximated boundaries overlap with the criterion variable informs the first research question and addresses different aspects of construct validity as proposed by Messick.

Regression model specification

To examine the construct validity of various indicators of gang ecologies, a series of hurdle regression models are specified. Zero inflated beta regression models are used to model the proportion outcome while zero adjusted negative binomial models are used to model the count outcome. Independent variables reflect the home address of gang members, gang-related crime, and gun-crime. Relationships between the independent variables and the dependent variables are assessed after controlling for nearby gang presence and the area of each spatial unit.

Given the large number of cases with proportion or count values equal to zero (see Table 11 for proportion distribution and Table 14 for the count distribution), a selection issue may be a factor with this type of analysis. It is possible that the variables that influence whether or not a gang exists in an area are different from the variables that influence how large a gang set space is or how many gangs are present in a given spatial unit. From a statistical perspective, the variables that predict the probability of a non-zero value may not be the same variables that predict the mean of the positive values. It is important to choose a statistical modeling technique that can accommodate this possibility, otherwise, the model will be mis-specified and the parameters estimates will be biased (Heckman 1979).

Hurdle models overcome a selection issue by allowing the user to estimate separate models for zero and non-zero values. These models—specifically zero inflated models—have been used to model data when zeros are an important feature of the data (Cook, Kieschnick, and McCullough 2008; Ospina and Ferrari 2012; Smithson and Verkuilen 2006). Following other researchers, I used hurdle models to explore whether variables predict the zero and non-zero values in similar ways.

If the results of this study suggest a selection issue is present in the data, a theory of human ecology can be used to explain the sociological process driving the selection issue. According to this theory, spatial units with a proportion value equal to zero not only reflect the absence of gang presence, but reflect areas with a different functional role, or “niche” as described by Hawley (1950: 44-45), compared to spatial units that do contain some amount of gang presence.

Zero-inflated hurdle models available through an R package called GAMLSS (Generalized Additive Models for Location, Scale and Shape) were used in the current study. These models generate separate parameter estimates (*mu* and *nu*) for each of the independent variables. The parameter estimates that reflect the ability of a variable to discriminate between zero and non-zero values are referred to as *nu* parameters in the GAMLSS package. A separate set of parameter estimates, *mu* parameters, model the mean of the non-zero values (0,1).[13] Furthermore, these models generate separate parameter estimates for the dispersion parameter. These parameter estimates, sometimes referred to as scale parameters, are labeled as *sigma* parameters in the regression output (regression output can be found in Appendix B which starts on page 183).

Hurdle models were specified using two different distributions to accommodate the distribution of the dependent variable when it is modeled as a proportion and when it is modeled as a count. When dependent variables are proportions, fractions or rates, beta regression models

[13] My dependent variable falls on the closed interval [0,1] which includes the values of zero and one, but the beta distribution will only model values on the open interval (0,1) which excludes values of zero and one. A hurdle model will be used to model the zero values separately (nu parameter). Since the beta distribution does not include values that are equal to one, the non-zero values were transformed using the formula y=(y*N-1)+.5)/N as proposed by Smithson, Michael and Jay Verkuilen. 2006. "A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-Distributed Dependent Variables." *Psychological Methods* 11:54-71.

are most appropriate (Baum 2008). When the dependent variable is modeled as a count, negative binomial or poisson distributions can be used. Negative binomial distributions are most appropriate when overdispersion is present. The count variable was modeled using both negative binomial and poisson distributions; the negative binomial distribution fits the data best (using AIC as a criteria), and thus became the chosen model in this study.

While hurdle models are flexible and can be applied to a variety of applications (Ferrari and Cribari-Neto 2004), these models have inherent assumptions that must be met before parameter estimates can be interpreted. Specifically, multicollinearity and outlier cases can influence the estimation of parameters. Additionally, these models assume that observations are independent.

Multicollinearity

As with many types of regression, beta and negative binomial regression models will produce biased parameter estimates and standard errors if predictor variables are highly correlated with each other. This has been described as the partialling fallacy (Gordon 1968). If the four indicators of gangs that are used in this study are in fact measures of the same construct, they will be collinear. To avoid issues related to the partialling fallacy, variables were entered into models independent of one another (these models are summarized in Table 18 on page 83). What is of more concern is collinearity with the control variable and each indicator. To test for multicollinearity, tolerance and Variance Inflation Factor (VIF) scores were generated for each variable. A value of tolerance smaller than .1 or VIF values greater than ten were used to flag issues of multicollinearity.

Outliers

Multivariate outliers generate models with poor fit. To screen for multivariate outliers, standardized residuals and the Cook's D influence statistic were generated.[14] Values greater than one were flagged and further evaluated to ensure the data were not miscoded. If values greater than one were not the result of miscoded data they were removed from the dataset. Models were run both with and without these outlier cases to determine what effect, if any, outlier cases have on the results.

Independence of errors (spatial adjacency analysis)

Lastly, regression models assume that observations are independent of one another. If observations are not independent, error terms will be correlated. When this assumption is violated, correlated errors will inflate the likelihood of a Type I error. Given that "near things are more related than distant things" (Tobler 1970: 236), spatial data frequently violate this assumption.

The Global Moran's Index is used to determine if model residuals are spatially clustered. Statistically pseudo significant spatial clustering in the data ($p<0.05$) indicate the assumption of independent observations has been violated. The Global Moran's Index is a frequently used statistic to determine if data are spatially dependent (Tita and Radil 2011).

Global Moran's Index values can be generated using the spdep package in R. This statistic measures spatial autocorrelation by simultaneously considering feature values and geographic location. This tool calculates the average distance between each feature in the dataset, then uses that average distance to determine if features with high values cluster with other high values, or if

[14] The GAMLSS R package generates normalized (randomized) quantile residuals as defined by Dunn, Peter K and Gordon K Smyth. 1996. "Randomized quantile residuals." *Journal of Computational and Graphical Statistics* 5:236-244. Their definition "produces residuals which are exactly normal, apart from sampling variability in the estimated parameters, by inverting the fitted distribution function for each response value and finding the equivalent standard normal quantile. Our definition includes some randomization to achieve continuous residuals when the response variable is discrete" (236).

features with low values cluster with other low values. When clustering is detected, the Global Moran's Index will be positive. The tool generates a z-score and pseudo p-value for the Index value. If the Global Moran's Index pseudo p-value is statistically significant ($p<0.05$), a spatially smoothed predictor variable should be created from the outcome to account for spatial dependence in the data.

Generating a spatially smoothed variable involves specifying a spatial weights matrix defining neighboring features. A spatially lagged variable averages values from neighboring features. There are a number of ways to specify a spatial weights matrix. Some have argued that researchers have not made use of theoretically appropriate spatial weights matrices (Tita and Greenbaum 2009). They encourage researchers to think more critically about *how* spatial features might be related, then generate spatial weights matrix based on the theoretical process driving the spatial relationship (Tita and Greenbaum 2009). Here spatial dependence is controlled using a contiguity based spatial weights matrix that reflects diffusion processes generating the spatial patterning of gang ecologies.

The diffusion process suggests gang presence in ecological units spreads through local social interactions. Cultural values and norms characterizing communities with gang presence can be transmitted through local social networks and thereby diffuse into nearby communities. This diffusion process has been used to explain the spatial distribution of drug crime (Rengert 1996), homicides (Loftin 1986), and specifically, gang homicides (Cohen and Tita 1999).

Based on the diffusion process, a variety of spatial weights matrices could be used to correct spatial dependence in the data. The most popular spatial weights matrices are contiguity based (rook, queen, or k nearest neighbors), distance based (inverse/fixed distance weighting) or are based on social networks (For examples using social networks, see Tita and Greenbaum 2009;

Tita and Radil 2011). Each of these methods, with the exception of inverse distance weighting, traditionally use row standardization, meaning every neighbor is given equal weight. In this study, first order queen contiguity was used to define neighbors in the spatial weights matrix. Using a contiguity based adjacency rule means the extent of the hypothesized diffusion process will vary across different size ecological units. The extent will increase as the spatial unit becomes larger.

Model diagnostics

This study runs five regression models on each of the six dependent variables (one dependent variable at each of the six units of analysis). Since multiple models were run on the same dependent variable, alpha values were adjusted to reduce the possibility of a Type I error. A Type I error occurs when the null hypothesis is erroneously rejected. There are different ways to adjust probability values (p-values) to minimize the likelihood of committing a Type I error. In this study, a Bonferroni adjustment is used to account for the five regression models run on each dependent variable (see Equation 1).

Equation 1: Bonferroni adjusted probability value

$$\frac{\alpha}{n} = \frac{.05}{5} = .01$$

To explain notable changes in explained variance in the outcome variable, global deviance measures (-2 * log likelihood) were used. One such statistic based on deviance measures is the Bayesian Information Criterion (BIC). The BIC approximates the Bayes Factor. The Bayes Factor is used when comparing two models to determine which model "will, on average, give better out-of-sample predictions" (Raftery 1995: 130).

BIC values can be used to identify improvement in model fit across nested models. BIC values gauge strength of prediction while handicapping the model based on its complexity (the number of predictor variables in the model). A model with seven independent variables might have a higher R^2 value compared to a model with only three independent variables, however, the seven predictor model would not necessarily have a higher BIC value if six of those variables are not contributing unique explained outcome variance.

The formula specified in Equation 2 calculated BIC values (Raftery 1995: 133-134). In this equation, L_k^2 is the deviance (-2 * log likelihood) for model k, df_k is the number of degrees of freedom in model k, and n is the sample size.

Equation 2: BIC formula

$$BIC_k = L_k^2 - df_k(log\ n)$$

Models were also assessed using likelihood ratio tests. This is an omnibus test that each of the new variables has an impact (b weight) of zero. The test statistic and corresponding p-value tests the null hypothesis that the parameter estimate for the new variable in the model is equal to zero—the addition of this variable adds nothing beyond what the existing variables already explain. If the test statistic is determined to be statistically significant (p-value < .01), this can be interpreted as evidence that the new variable improves the model. This test builds on the BIC comparisons because it tests whether improvement from one model to the next is statistically significant; this is something the BIC does not tell you. In other words, this is the only test of different models that has a clear null hypothesis accept/reject decision.

The formula specified in Equation 3 was used to calculate the likelihood ratio test statistic (Long 1997: 94). In this formula, 2 ln L(M_u) is the global deviance (log likelihood) for the unconstrained model (the model with a new variable) and 2 ln L(M_c) is the global deviance for the constrained model (the model without the variable). The constrained model is nested within the unconstrained model.

Equation 3: Likelihood ratio chi-square test

$$G^2(M_c|M_u) = 2 \ln L(M_u) - 2 \ln L(M_c)$$

Finally, a global goodness-of-fit measure was calculated using a pseudo R^2 statistic. The pseudo R^2 value was calculated using the formula in Equation 4 (Cox and Snell 1989). This coefficient of determination is interpreted similarly to the R^2 value in linear regression models as explained variance. However, this coefficient does not have an associated p-value; it is purely descriptive. When comparing models using the pseudo R^2, there are no significance tests that will determine if improvement in R^2 is statistically significant. In this formula, $L(\hat{\theta})$ is the log-likelihood of the null model (the model without any predictors) and $L(\hat{\theta}^H)$ is the log-likelihood of the alternative model (the model with predictor variables).

Equation 4: Pseudo R^2

$$p^2 = 1 - \left(\frac{L(\hat{\theta})}{L(\hat{\theta}^H)}\right)^{\frac{2}{N}}$$

Applying Messick's unified approach

The various aspects of construct validity outlined by Messick were examined using output from the regression models. To test the two primary research questions, output from five regression models were analyzed across all six dependent variables (summarized in Table 18). These models include a baseline model which only includes the control variable and exposure variable followed by three subsequent models that include each of the independent variables into the regression series independently. A final model included all three independent variables together in one model. Each of the independent variables were entered into the regression series individually to determine how well each independent variable performs on its own. Interaction terms were not included in the regression series since I was interested in understanding the extent to which each indicator contributes its own unique portion of construct relevant variance. Interaction terms are used when a researcher has reason to believe that the effect of one variable depends on the value of another independent variable. There is nothing in the gang research that would suggest the effect of gang member home address locations on gang set space, for instance, is dependent on the amount of gun crime in an area. It is possible that this is true, which would open up interesting opportunities for future research that address different research questions. In my assessment of the validity of different gang indicators, interaction terms do not allow me to capture the unique effect of each variable, which is why they were not used in the current study.

The following section explains how the results of these five regression models (across all six dependent variables) address different aspects of construct validity as proposed by Messick.

Table 18: Example output from hurdle regression models

	Baseline model	Model A	Model B	Model C	Full model
Mu model (magnitude)					
Intercept					
Gang indicators					
Home address		x			x
Gang crime			x		x
Gun crime				x	x
Control variable					
Spatial lag	x	x	x	x	x
Area (exposure)	-	-	-	-	-
Nu model (zero)					
Intercept					
Gang indicators					
Home address		x			x
Gang crime			x		x
Gun crime				x	x
Control variable					
Spatial lag	x	x	x	x	x
Area (exposure)	-	-	-	-	-
Model diagnostics					
-2 log likelihood	LL	LL	LL	LL	LL
Pseudo R^2	R^2	R^2	R^2	R^2	R^2
BIC	BIC	BIC	BIC	BIC	BIC
LR test					
Baseline		$LR_{baseline}$	$LR_{baseline}$	$LR_{baseline}$	
Full		LR_{full}	LR_{full}	LR_{full}	

Notes: This table was generated for each of the six dependent variables reflecting the six levels of spatial aggregation (five models * six dependent variables = 30 total models for proportion outcome and 30 models for count outcome). The models were estimated using the logit function to estimate the mean of the beta distribution. The *nu* models were estimated using the logit function to predict the zero inflated part of the model. Each x in the *mu* and *nu* models represents a parameter estimate, standard error and p value. $LR_{baseline}$ and LR_{full} represents the likelihood ratio test statistic and accompanying p value.
Abbreviations: LL=-2 * log likelihood, BIC=Bayesian Information Criterion.
LR test=Likelihood ratio test.

Table 19: Application of Messick's model

Research question	Empirical test	Messick
1: Do indicators of gang ecologies identify the same construct?		
A: Do models with a single indicator have better fit than the baseline model?	BIC, LR test	Content
B: Does the full model have better fit than the single indicator models?	BIC, LR test Pseudo R^2	Content
C: Do models with a single indicator each fit the data equally well?	BIC, Pseudo R^2	Content
D: Is the significance pattern generated in the analyses from 1A, 1B and 1C consistent across levels of spatial aggregation?	BIC, LR test Pseudo R^2	Structural
E: How much do quantitatively derived gang polygons overlap with the practitioner defined polygons?	Analysis of variance	Content
F: Does the percentage of overlap vary by type of gang?	Analysis of variance	Generalizable
2: Are indicators able to predict gang presence (binary outcome) and gang magnitude (proportion outcome) in similar ways?		
A: Do significance patterns in the *nu* models (binary outcome) match the significance patterns in the *mu* models (proportion outcome)?	*mu* parameters vs *nu* parameters	Structural
B: Is the significance pattern generated by the analysis in 2A consistent across levels of spatial aggregation?	*mu* parameters vs *nu* parameters	Structural

To determine if indicators of gang ecologies are measuring the same construct, BIC values and likelihood ratio test results were compared across all six models using three separate tests. First, if each indicator is relevant, they should improve model fit when added to the baseline model. To determine if each indicator improves model fit over the baseline model (baseline = only control variables) the BIC values generated by Models A, B and C were individually compared to the baseline model. These models should have BIC values that are ten units better than the baseline model. Similarly, the likelihood ratio test should also indicate that Model A, B and C offer significant improvement over the baseline model. This part of the analysis addressed Messick's **content** aspect of construct validity.

Second, if all three indicators of gang ecologies are measuring the same construct, adding all three indicators to the same model would be redundant. If this is the case, then Models A, B, and C will have better fit compared to the full model (the model with all three indicators). If the BIC values for Model A, B, and C are at least ten units better than the full model, this would suggest that the full model is redundant; the variables are not modeling anything that the single variable model is not already picking up. Similarly, the likelihood ratio test that compares Models A, B and C with the full model should *not* produce a test statistic that is large enough to reject the null hypothesis if the full model is redundant. Finally, if each gang ecology indicator is measuring the same underlying construct, pseudo R^2 values for the single indicator models will be approximately the same as the full model. This part of the analysis also addressed Messick's **content** aspect of construct validity.

Third, if all three indicators of gang ecologies are measuring the same construct, each single indicator model will fit the data equally well. If the BIC values for Model A, B, and C are approximately equivalent to each other (difference of less than ten), this would suggest all three models fit the data equally well. Furthermore, pseudo R^2 values will be approximately the same across Models A, B and C. Since these models are not nested, likelihood ratio tests cannot be used to compare model fit. This part of the analysis addressed Messick's **content** aspect of construct validity.

Fourth, if indicators of gang ecologies are equally relevant across units of spatial aggregation, the significance patterns detected in the analysis described above should be consistent across all six versions of the dependent variable (grid cells, thiessen polygons, census block groups, census tracts, neighborhood clusters, community areas). If significant patterns are *not* consistent across all six dependent variables, this would suggest that indicators of gang ecologies at one unit of analysis may not be as effective at other units of analysis. This part of the analysis addressed Messick's **structural** aspect of construct validity.

To address the second research question, regression output were analyzed to determine if indicators of gang ecologies are able to predict gang presence, modeled as a binary variable, and gang magnitude, modeled as a proportion variable, in similar ways. To answer this research question, significance patterns in the *mu* models were compared with the significance patterns in the *nu* models. These comparisons were made *within* each model. For instance, if the home address variable is significant in the *mu* models *and is also significant* in the *nu* models, this would suggest that this indicator can be used to predict gang presence as well as gang magnitude. Comparisons were made across Models A, B and C. These comparisons were not made in the full model to avoid issues related to the partialling fallacy.[15] This part of the analysis addressed Messick's **structural** aspect of construct validity.

---

[15] When correlated indicators are entered into the same model, B weights and standard error values cannot be used to evaluate the model See Gordon, Robert A. 1968. "Issues in Multiple Regression." *The American Journal of Sociology* 73:592-616.

Next, a subsequent analysis examined whether the significance pattern of the *mu* and *nu* models are consistent across levels of spatial aggregation. If the significance pattern is consistent across all six dependent variables, this again would imply that gang indicators operate in the same way across all units of spatial aggregation. This analysis also addresses Messick's **structural** aspect of construct validity.

In addition to the regression analyses, I created an algorithm to approximate gang set space locations. A comparison of the approximated gang set space locations and the official PPD gang set space was used to answer these research questions in a different way. The details of this analysis are described next.

Algorithm specification

The goal of this analysis was to examine how well each indicator was able to approximate gang set space locations. This was done by approximating gang set space locations quantitatively using the home address and arrest locations of gang members for each individual gang. The degree to which these quantitatively derived territories overlap with the set space locations (identified qualitatively using police and practitioner knowledge) served as yet another test of the relevance of each indicator.

To conduct this analysis, two of the independent variables (home address, gang crime) were used to create polygons representing the gang set space locations. The set space boundaries were delineated using a raster analysis. This involved a two-step process described next.

First, the independent variables (represented as a point data file) were aggregated to raster cells. This was done using the *kernel density* tool in ArcGIS. The *kernel density* tool calculates density values by taking the weighted sum of all of the points that fall within a specified distance of each raster cell. The tool weights incidents that are close to the cell more heavily compared to points that are further away; this accounts for the fact that points that are in close proximity are more related than points that are far apart. The weighting is based on the function in Equation 5. This equation was taken from figure 4.5 in Silverman (1986: 75). In this study, 100 foot grid cells were specified with a 1,320 foot buffer.[16]

Equation 5: Kernel Function

$$K_2(x) = \begin{cases} 3\pi^{-1}(1 - x^T x)^2 & if\ xT\ < 1 \\ 0 & otherwise \end{cases}$$

Once density values were assigned to each raster cell, statistically significant clusters of high density cells were used to specify the set space boundaries. Statistical significance was calculated using the Getis-Ord Gi* statistic.

The Getis-Ord Gi* statistic identified clusters of high or low values by calculating z-scores and p-values, indicative of spatial homogeneity, for each feature in the dataset. Large positive z-scores indicate spatial clustering of high values, while negative z scores indicate spatial clustering of low values. Values close to zero indicate no spatial clustering is present in the data. This tool allows the user to specify a weighting mechanism so that nearby raster cells are weighted more than distant cells. Furthermore, a cutoff distance can be specified to exclude the effect of outliers. This method was developed to detect "pockets" of spatial clustering that may not be evident with the use of global measures (Getis and Ord 1992: 189).

Once the point data were aggregated to raster cells, gang set space boundaries were delineated using results from the Getis-Ord Gi* analysis. Using the Getis-Ord Gi*, statistically significant *positive* z score values can be used to identify each gang set space. These significant

[16] Previous research has found that a quarter mile buffer (1,320 feet) is an appropriate length to specify "near" events. See Groff, Elizabeth R. 2011. "Exploring 'near': Characterizing the spatial extent of drinking place influence on crime." *Australian & New Zealand Journal of Criminology* 42:156-179. The Block (2000) study used a buffer of 750 meters, which is about 2,400 feet.

clusters can then be used to generate a polygon feature class for each individual gang. Positive Getis-Ord Gi* values that are statistically significant at the .001 level using the inverse weighting function were used to define the boundaries of gang areas.[17] An example of the output generated by this process is available in Figure 1.

The *kernel density* tool creates output that not only are spatially smoothed but also contain enough variation to run the Gi* tool. The spatially smoothed values are important because without the spatial smoothing (if grid cell counts are used for instance) the Gi* statistic is likely to identify *individual* grid cells that are statistically significnt rather than *groups* of cells that are clusterd together. It is possible to create these smoothed values using incident counts that fall within a distance band of each grid cell, but this method removes variation in the data which is also necessary for the Gi* tool to work properly. The *kernel density* tool uses inverse distance weighting to overcome this limitation. The *kernel density* tool was used in this study because it is a tool that creates the spatial smoothing and variation that is necessary to identify statistically significant clusters of high value grid cells using the Gi* statistic.

[17] Since the Getis-Ord Gi* tool can only be used on features with integer values, the kernel density values were converted to whole numbers. This was done using the *Map Algebra* tool. Since kernel density values were very small, these values were multiplied by 10,000,000 to create integer values that are between 0 and 100. Once this was done, statistically significant Gi* values ($p<.001$) were combined using the dissolve tool.

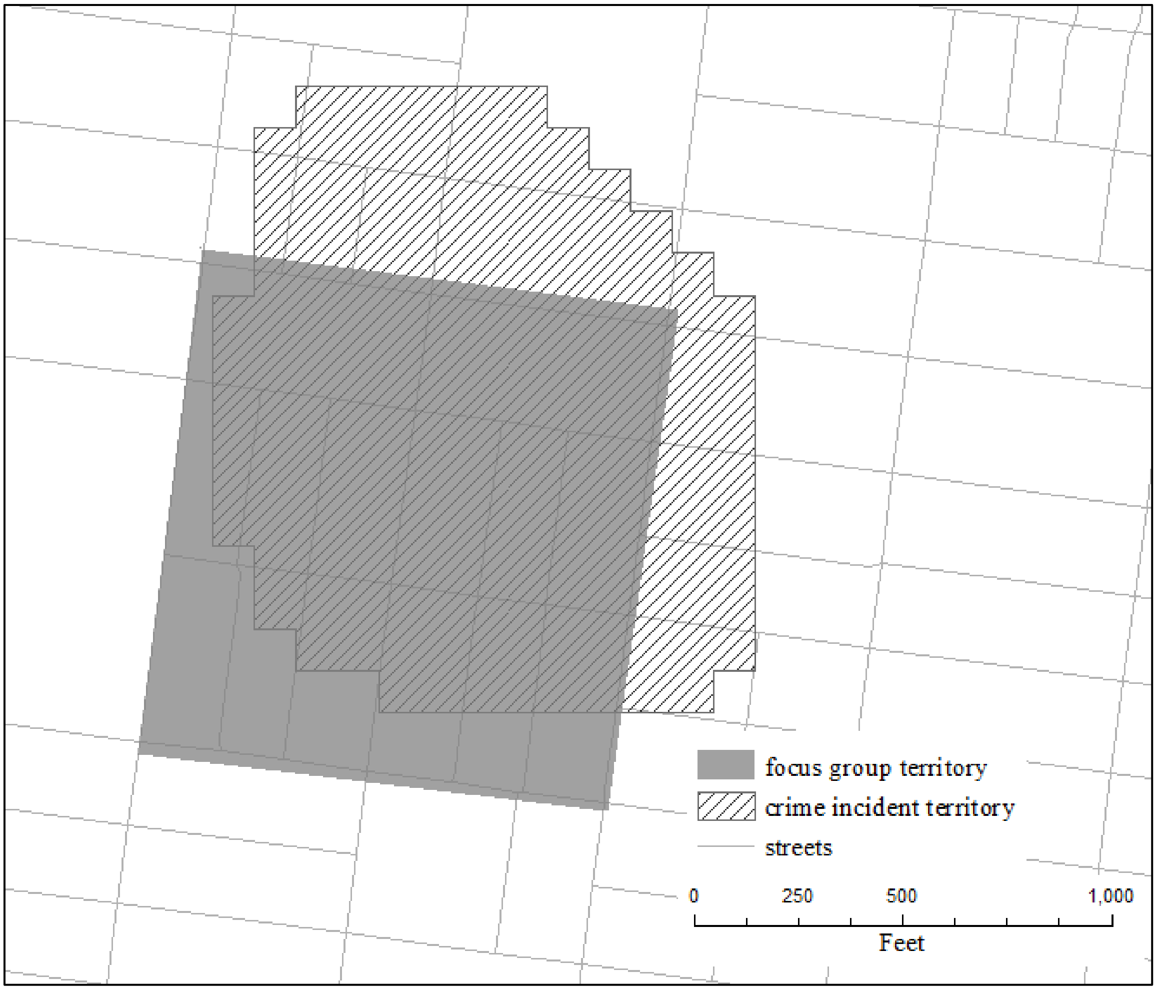


Figure 1: Map displaying set space overlap, PPD set space vs gang arrest set space

Note: A 1,320 foot bandwidth was specified for the kernel density function. This tool produced grid cell values that are 100 square feet. The kernel density values were converted to integer values using the *Map Algebra* tool. Kernel density values were multiplied by 10,000 to generate integer values. Raster cells were then converted to polygon features so the Getis-Ord Gi* analysis could be run. This was calculated using the inverse distance squared weighting function. Cells with p-values <.001 were retained to generate the gang set space polygons.

Significance testing

Once the point data were aggregated to polygons reflecting the approximated gang areas, the last step of this analysis involved calculating the degree to which gang areas produced by each variable overlapped with the qualitatively derived territories. Overlapping areas were calculated using the *union* tool in ArcGIS. Each of these areas were then compared to the criterion variable—the PPD set space locations delineated through focus groups—and were assessed using a coefficient of areal correspondence (Garson and Biggs 1992). This coefficient compares the area of the two polygons that is shared to the total area of both polygons (see Equation 6). This coefficient ranges from zero to one, where values of zero indicate the two areas do not overlap at all (for an example, see image A in Figure 2) and a value of one indicates the two areas overlap 100% (see image C in Figure 2). The indicator that created the largest amount of overlap with the criterion variable was interpreted as the most accurate indicator of gang ecologies.

Equation 6: Coefficient of areal correspondence

$$Coefficient\ of\ areal\ correspondence = \frac{area\ covered\ jointly\ by\ both\ polygons}{total\ area\ of\ both\ polygons}$$

An example of this analysis for one gang is depicted in Figure 1. The map includes the qualitatively derived gang set space that was identified through the focus group meetings and a quantitatively derived polygon that was generated using arrest locations of gang members between January 1, 2012 and December 31, 2014. The coefficient of areal correspondence for these two features indicates that about 49% of the area of the two polygons overlap. The analysis produced a separate coefficient of areal correspondence for each gang. In other words, two coefficient of areal correspondence values were generated for each gang: one was generated using the gang arrest indicator and the other was generated using the home address indicator.

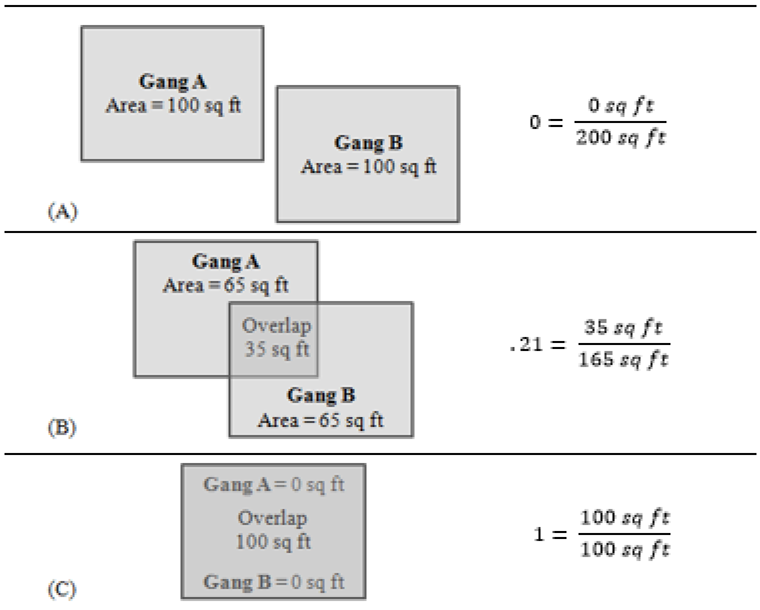


Figure 2**: Calculating the coefficient of areal correspondence**

Applying Messick's unified approach.

To answer the first research question, the coefficient of correspondence was used to measure the degree to which each indicator can be used to identify the criterion variable. If each indicator is equally relevant, the coefficient of correspondence values should be approximately the same across all three indicators. To test this research question, an Analysis of Variance (ANOVA) was used to test the differences in mean coefficient values across these indicators. If each of the indicators are equally effective in identifying gang ecologies, the test statistic should *not* reach statistical significance. When test statistic did reach the critical value, post hoc tests were used to determine which indicators are preferred. This part of the analysis examined Messick's **content** aspect of construct validity.

The last analysis compared coefficient of areal correspondence values across the three PPD gang typologies. Separate ANOVA models were run for each typology. If each indicator is equally relevant for each type of gang, the test statistic should not reach a critical value in any of the three models. If some methods are better at predicting some types of gangs compared to other types, post hoc tests were used to determine which indicators are most relevant. This final analysis examined Messick's **generalizability** aspect of construct validity.

CHAPTER 4: RESULTS

This study addressed two main research questions: (1) Do indicators of gang ecologies identify the same construct? (2) Are indicators able to predict gang presence (binary outcome) and gang magnitude (proportion/count outcome) in similar ways?

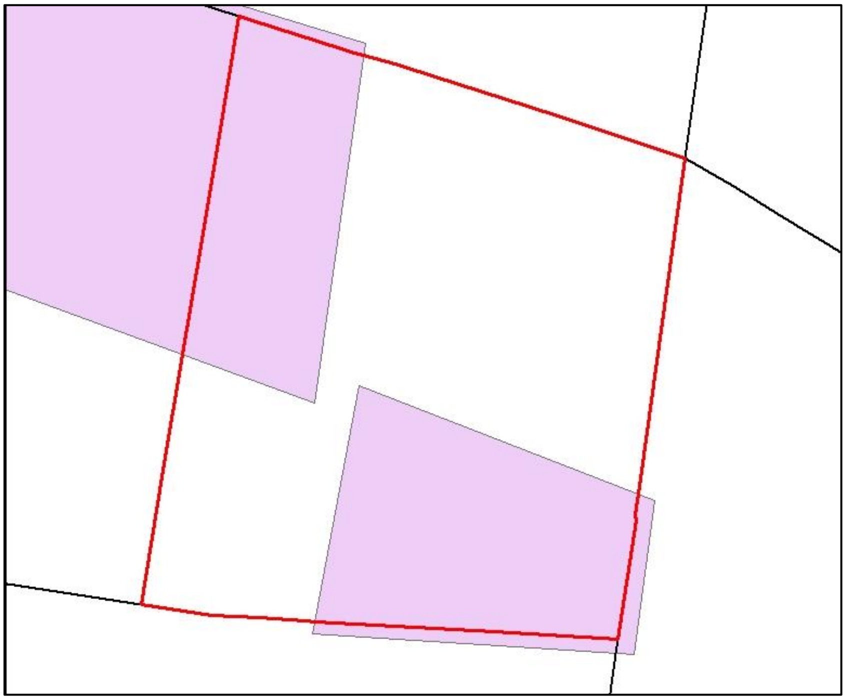
The data were analyzed using two analysis techniques. The first technique involved running a series of regression models that used the home address, gang arrest, and gun crime variables to predict PPD set space defined in focus group meetings. In the second analysis, I developed an algorithm that created gang set space polygons using either the locations where gang members live or gang arrest data. The gun crime variable was not used in this analysis because gun incidents could not be linked to an individual gang. The set space polygons that were created by the algorithm were compared to the official gang set space file provided by the PPD Criminal Intelligence Unit (CIU).

The remainder of this chapter is dedicated to describing the results of these analyses. The regression analyses are described first, followed by the results of the algorithm output.

Regression models

To examine the construct validity of various indicators of gang ecologies, a series of hurdle regression models were specified. The dependent variable was operationalized two different ways. Both versions of the dependent variable were measured using a continuous scale. The first version operationalized gangs in terms of area—the amount of gang set space covering each geographic unit measured as a proportion. The second version of the dependent variable was quantified using the number of gangs present in a given area. An example of this can be found in Figure 3. In Figure 3 one census tract is highlighted in red. Gang set space locations are represented by the shaded polygons. 40% of the red census tract is covered by gang set space; therefore, the proportion value for this record is .40. However, there are two different gangs present within this census tract, so the gang count value is two.

Figure 3: Quantifying magnitude using count and proportion values



Note: The shaded polygons reflect gang set space. The thick black lines reflect census tract boundaries.

Beta regression models were used when the dependent variable was a proportion and negative binomial models were used when the dependent variable was a count. Independent variables included the home address of gang members, gang arrests, and gun-crime. Relationships between the independent variables and the dependent variable were assessed after controlling for nearby gang presence (spatially lagged dependent variable) and the size of each spatial unit (measured in square feet).

Assumption checks

Multicollinearity: VIF and tolerance values were created at each spatial unit. All VIF values are less than ten, and all tolerance values are greater than 0.1. These results suggest the independent variables and the control variable (area) are not collinear. These values indicate the proportion of the variable that is independent of the other variables. What is notable about these results is that each variable (with the exception of the control variable, area) gets progressively less unique at larger geographic scales. In other words, the variables begin to get more tightly tied together at larger spatial scales. This may simply be driven by aggregation patterns, or, it may be that these process each of these indicators are tapping into become less defined and begin to fuse together at larger spatial scales. The reason for this pattern is unclear.

Table 20: VIF values

Variable	Grid cell	Thiessen polygon	Census block group	Census tract	Neighborhood	Community
Home address	1.25	1.02	1.94	2.69	5.32	6.12
Gang arrests	1.26	1.10	1.80	2.31	3.67	4.09
Gun incidents	1.23	1.10	1.33	1.76	2.74	2.77
Area	1.01	1.00	1.02	1.03	1.05	1.09

Table 21: Tolerance values

Variable	Grid cell	Thiessen polygon	Census block group	Census tract	Neighborhood	Community
Home address	0.80	0.98	0.51	0.37	0.19	0.16
Gang arrests	0.79	0.91	0.56	0.43	0.27	0.24
Gun incidents	0.81	0.91	0.75	0.57	0.37	0.36
Area	0.99	1.00	0.98	0.97	0.95	0.91

Outliers: Cook's distance values were created for each of the six spatial units. Cook and Weisberg (1982) suggest that values greater than one suggest that case may be overly influential in the model. All of the Cook's Distance values were below one, however, there was one case in the Thiessen model had a value of .99.

Table 22: Five highest Cook's distance values

Highest Cook's distance values	Grid cell	Thiessen polygon	Census block group	Census tract	Neighborhood	Community
1	0.76	0.99	0.56	0.34	0.24	0.51
2	0.73	0.36	0.27	0.14	0.11	0.17
3	0.66	0.20	0.14	0.05	0.10	0.10
4	0.17	0.13	0.07	0.04	0.06	0.07
5	0.07	0.10	0.06	0.03	0.05	0.04

The thiessen polygon with a Cook's distance value of .99 was a polygon where 16.8% of the area was covered by PPD gang set space. No gang members lived within that thiessen polygon, but 90 gang arrests occurred there. Only one other thiessen contained more gang arrests than this particular case; the highest arrest value was 104 incidents. Further examination of these 90 incidents revealed that 82 of them occurred prior to June 1, 2013. Additionally, all of those 82 cases were arrests that involved members in the same gang. The remaining eight arrests were by members of five different gangs. The regression analyses were run with and without this particular case.

Independence of errors (spatial adjacency analysis): The Global Moran's Index is used to determine if there is significant spatial clustering in model residuals. Statistically significant global spatial clustering in the data ($p<0.05$) indicate the assumption of independent observations has been violated.

Global Moran's Index values were generated using R using the spdep package. A number of models were run to determine the most appropriate way to model the data. A baseline model was run first. This model did not account for the spatial nature of the data. The baseline model was followed by three subsequent models that accounted for spatial dependence in different ways. The first included a spatially lagged version of the dependent variable as an independent variable. The lagged variable was created by specifying a first order queen contiguity neighbor matrix. The second model also included a spatially lagged version of the dependent variable, but this model specified neighboring areas using second order queen contiguity to define neighbors. The final model accounted for spatial dependencies in the data using Gaussian Markov Random Fields. These models account for spatial dependence by "bringing fitted values from neighbouring regions closer together (rather than shrink them towards the overall mean as is the case of a simple random effect model term)" (Bastiani, Rigby, Stasinopoulous, Cysneiros, and Uribe-Opazo 2016: 6). Units that are not defined as neighbors are assumed to be conditionally independent. Neighbors were defined using a first order queen contiguity matrix. Each of the regression models were estimated using maximum likelihood. Following Bastiani et al. (2016), a global measure of model fit—the Bayesian Information Criterion (BIC)—was used to identify which models fit the data best. BIC values were calculated separately using four different models. These models included a full model, with all three independent variables, and three subsequent models each with only one independent variable. Results of these models are listed in Table 23.

Table 23: BIC comparisons to address spatial autocorrelation

	Grid cells				Thiessen polygon			
	full	home	incident	gun	full	home	incident	gun
Baseline	2681	3255	3149	3134	2398	3050	2813	2983
Lag 1st	-2580	-2371	-2442	-2433	-9811	-9606	-9751	-9749
Lag 2nd	-1222	-954	-1030	-1002	-7449	-7127	-7338	-7275
GMRF	-	-	-	-	-	-	-	-

	Census block group				Census tract			
	full	home	incident	gun	full	home	incident	gun
Baseline	485	586	568	649	23	55	18	106
Lag 1st	-79	-16	-18	-4	-65	-55	-72	-13
Lag 2nd	193	271	266	317	-26	-10	-34	33
GMRF	485	586	568	649	23	55	18	106

	Neighborhood				Community areas			
	full	home	incident	gun	full	home	incident	gun
Baseline	-4	20	11	40	-34	-39	-39	-21
Lag 1st	-12	-7	-7	10	-36	-38	-37	-20
Lag 2nd	-4	1	2	20	-32	-35	-33	-16
GMRF	-4	20	11	40	-34	-39	-39	-21

Note: Lag 1st = spatially lagged version of dependent variable included as a predictor where neighbors are specified using first order queen contiguity. Lag 2nd = spatially lagged version of dependent variable included as a predictor where neighbors are specified using second order queen contiguity. GMRF=Gaussian Markov Random Fields models where neighbors are specified using first order queen contiguity. Smallest BIC values are highlighted in green. – symbol indicates the model would not converge, therefore BIC values could not be generated.

Based on the results in Table 23 it is evident that the first order queen lag models fit the data best at five of the six spatial units. These units include data aggregated to grid cells, thiessen polygons around street intersections, census block groups, census tracts and neighborhoods. At the community spatial unit, all of the models fit equally well (difference in BIC is less than ten in each model). Based on these findings, the results that follow include a spatially lagged variable using first order queen contiguity to define neighbors to model spatial dependence for each of the six spatial units.

Research Question 1: Do indicators of gang ecologies identify the same construct?

The main research question in this study is whether different gang indicators reflect the same general construct to the same degree. This question is relevant to the content relevance and representativeness of ecological gang indicators according to Messick. Three empirical tests were used to answer three sub-questions.

Sub-question A: Do models with a single indicator have better fit than the baseline model?

If each of the three gang indicators used in this study (where gang members live, gang arrests, and gun crime) are relevant to the identification of gang set space, then each variable should explain variance in the gang set space variable over and above control variables (control variables include the size of each spatial unit and a spatially lagged version of the dependent variable). In other words, each of the single indicator models should improve model fit over the baseline model. In this study, improvement in model fit was assessed using two methods: likelihood ratio tests that compared the single indicator models to the baseline model and BIC comparisons.

Likelihood ratio tests that compare the single indicator models with the baseline models were always significant at the .01 level. This means that the single indicator models always improve model fit over the baseline model. This was true for all three independent variables. The only exception to this was Model A at the thiessen polygon level using the number of gangs as the outcome variable (see Table 53 on page 223). At this spatial unit, a model using only the home variable to predict the gang count outcome did not generate better model fit than the baseline model.

BIC comparisons were also used to compare the single indicator models to the baseline model. This is a more conservative test because the BIC accounts for model complexity, while the likelihood ratio test does not. The BIC value will only improve if each of the variables add something *new* to the model. If each of the variables are not adding something new, the BIC will not improve when the extra variables are included in the model. BIC results comparing the single indicator models with the baseline models were largely consistent with the results of the likelihood ratio tests. The BIC values for Models A, B and C were usually at least ten units smaller than the baseline model. This means that even after controlling for additional model complexity, the single indicator models were better than the baseline models. These results are displayed in Table 24 on page 102.

There were, however, some models that proved to be an exception to this general trend. For an example, in four cases, while the single indicator models did have smaller BIC values compared to the baseline model, the difference, less than ten units, was not large enough to conclude the difference was sizable. This happened in Model A—the home only model—at the thiessen polygon level and the grid cell level, and in Model C—the gun only model—at the community level. These cells are highlighted in Table 24. Additionally, there was one model where the single indicator model produced a *larger* BIC value than the baseline model, meaning the baseline model was preferred over this model. This means that in this model (Model A at the thiessen polygon level using gang counts as the outcome variable) the home address variable did not add anything above and beyond what the control variables measured. This model is highlighted in bold text in Table 24.

Table 24: BIC comparisons of single indicator models to the baseline model

		Model A: home address	Model B: gang arrests	Model C: gun crime
Proportion outcome	Grid cell	-188	-259	-250
	Thiessen polygon	-1	-146	-144
	Census block group	-146	-199	-184
	Census tract	-114	-131	-72
	Neighborhood	-91	-91	-74
	Community	-22	-23	-6
Count outcome	Grid cell	-4	-45	-53
	Thiessen polygon	**20**	-44	-84
	Census block group	-55	-60	-44
	Census tract	-61	-69	-27
	Neighborhood	-66	-70	-48
	Community	-23	-19	-5

Note: Values in this table represent the difference in the BIC value generated in the single indicator models versus the BIC in the baseline model. Positive values (highlighted in bold text) indicate the baseline model was preferred. Negative values indicate the single indicator model was preferred. Cells are highlighted if the absolute value of the difference was less than ten which means the difference was not statistically significant.

In summary, both the likelihood ratio tests and the BIC comparisons generally demonstrated the single indicator models offered an improvement over the baseline models. This means that all three variables (home address, arrest, and gun crime) each explain variance in the gang outcome variables over and above the control variables. The next question explored whether each measure individually captures all of the variance associated with the dependent variables.

Sub-question B: Does the full model have better fit than the single indicator models?

If each individual measure is equally representative of all characteristics associated with gang ecologies, then three things should be true. First, the likelihood ratio test comparing the single indicator models with the full model should *not* find statistically significant improvement in model fit. This is an omnibus test that each of the new variables has an impact (b weight) of zero. In this study, this is the only test of model comparisons that has a clear null hypothesis accept/reject decision. The test statistic and corresponding p-value tests the null hypothesis that the parameter estimate for the new variables in the model are equal to zero—the addition of the new variables add nothing beyond what the existing variable already explains. If the each of the variables are measuring the same thing, the model would *not* improve once all of the variables are included together. Each of the variables should, on their own, measure everything that the other variables are measuring. Second, the BIC should identify the simpler models—the single indicator models—as the preferred model. Third, the explained variance will be relatively high, and will not drastically improve after adding additional indicators in the full model.

Turning to the results generated in the 12 sets of regression models, the likelihood ratio test did *not* indicate the single indicator models have similar model fit compared to the full model. In fact, this test usually reveals that the full models are preferred over the single indicator models. This is the opposite of what was expected if each of the three gang indicators all measure the same general construct. Instead, these results indicated that each variable is adding something *new* to the model. In other words, each variable is making a unique contribution—the model that includes all three variables is the preferred model. The only exception to this proved to be the gang arrest model at the census tract level using gang counts as the dependent variable. In this model, the likelihood ratio test was not significant at the .01 level.

BIC values were also compared to address sub-question B. If each of the variables are measuring the same thing, the single indicator models should always have better fit (smaller BIC values) than the full model. The BIC comparisons identified results that were similar to the likelihood ratio tests. Once again, contrary to what was initially expected, the BIC values for the full model were usually *smaller* than the BIC produced in Models A, B and C, which suggests the full models were preferred over the single indicator models. In Table 25, cells highlighted in light grey identify instances when the full model was preferred over the single indicator models. In these models, each of the three variables are contributing something new, or are explaining different parts of the dependent variable.

There were three instances where the single indicator models were actually preferred over the full model. These models are highlighted in dark grey in Table 25. It is also worth noting that there were 13 instances where the BIC values for the single indicator models were very similar to the full model (difference less than ten units). This means that each of the single indicator models performed equally well when compared to the full model—neither model was preferred.

Other patterns emerged in Table 25. The gun crime variable appears to overlap least with the other indicators; in ten out of 12 cases, the full model was preferred over the gun crime model (Model C). The home address variable appears to have slightly more overlap with the other indicators. The full model outperformed the home address model in only six out of 12 models. Finally, the gang arrest model appears to overlap the most with the other indicators. The full model outperformed the gang arrest model in only four out of 12 models.

Finally, across all three indicators, the full model is most likely to outperform the single indicator models at small spatial scales. At the grid cell, thiessen polygon and census block group geographies, the full model outperformed the single indicator models 13 out of 18 times.

Table 25: BIC comparisons of single indicator models to the full model

		Model A: home address	Model B: gang arrest	Model C: gun crime
Proportion outcome	Grid cell	**209**	**138**	**147**
	Thiessen polygon	**205**	**60**	**62**
	Census block group	**63**	**61**	**75**
	Census tract	**10**	-7	**52**
	Neighborhood	**5**	5	**22**
	Community	-2	-1	**16**
Count outcome	Grid cell	**38**	-4	-12
	Thiessen polygon	**84**	**21**	-19
	Census block group	**3**	-2	**14**
	Census tract	-6	-14	**28**
	Neighborhood	**3**	-1	**21**
	Community	-5	-1	**13**

Note: Values in this table represent the difference in the BIC value generated in the single indicator models and the full model. Positive values (highlighted with bold text) indicate the full model was the preferred model. Negative values indicate the single indicator model was preferred. Cells are highlighted if the absolute value of the difference was greater than ten which means the difference was statistically significant.

In summary, both the likelihood ratio tests and the BIC comparisons generally demonstrated the full models were preferred over the single indicator models. This is the opposite of what was expected. Instead, these results indicated that each variable is adding something *new* to the model—this is more likely to be true for the gun crime indicator and across all three indicators at small spatial scales. In other words, each variable is making a unique contribution in explaining gang ecologies. The model that includes all three variables was usually the preferred model. The next question explored the strength of each indicator in explain gang ecologies.

Sub-question C: Do models with a single indicator fit the data equally well?

This question investigates whether each of the single indicator models are similar when compared against each other. If all three indicators are measuring the same general construct, each single indicator model should fit the data equally well. BIC and pseudo R^2 comparisons should indicate Model A, B, and C are approximately equivalent to each other. Since these models are not nested, likelihood ratio tests cannot be used to compare model fit.

Comparisons of BIC values across the three single indicator models revealed the three indicators were *not* equivalent in terms of model fit (see Table 26). Across all spatial scales and both versions of the dependent variable (proportion and count version), at least one model was always identified as the preferred model (lowest BIC by ten units or more). In other words, there was *never* a case where all three single indicator models performed equally well. Preferred models are highlighted in dark gray in Table 26. Across the 12 sets of regression models, the home address model was the preferred model eight times, the gang arrest model was preferred nine times, and the gun crime model was the preferred model only three times. It is also worth noting that the gun crime model was the worst performing model in nine out of 12 model comparisons. The home address model was the worst model three times, but the gang arrest model *never* had the lowest BIC value.

Based on these results, gang arrests appear to be the most adequate representation of gang ecologies across both outcomes. The gang arrest model was the preferred model in nine out of 12 model comparisons and was never the worst performing model. These results also indicate that gun crime may be the least adequate representation of the gang ecology construct, relative to the home address and arrest variables. The gun crime model was the preferred model in only three out of 12 model comparisons and was most likely to be the worst performing model relative to the other two. This model had the largest BIC value in nine out of 12 model comparisons.

It is also worth noting that sometimes, according to the results the best and the worst gang indicators depend on spatial scaling. For instance, at the thiessen polygon geography, the gun crime model is the preferred model and the home address model is the worst model, in terms of model fit. However, at larger spatial scales, such as the neighborhood and community levels, these variables swap places as the best and worst performing models; the home address model becomes the best model and the gun crime model becomes the worst model. The fact that the best and the worst model depend on spatial scale suggest the reasons why these variables link to the construct may change across spatial units. In short, the gang ecology construct, and the indicators used to measure that construct, may vary by spatial scale. The meaning behind gang ecologies at the community level may be different from the meaning at the thiessen polygon level.

Table 26: BIC values for single indicator models

		Model A: home address	Model B: gang arrests	Model C: gun crime
Proportion outcome	Grid cell	-2371	-2442	-2433
	Thiessen polygon	-9606	-9751	-9749
	Census block group	-16	-18	-4
	Census tract	-55	-72	-13
	Neighborhood	-7	-7	10
	Community	-38	-37	-20
Count outcome	Grid cell	2695	2654	2646
	Thiessen polygon	4234	4170	4130
	Census block group	1465	1460	1476
	Census tract	614	606	648
	Neighborhood	378	374	396
	Community	185	189	203

Note: BIC values for each of the single indicator models. Model comparisons are made *within* spatial scales and separately for each version of the dependent variable. Preferred models, defined as models with a BIC values at least ten units smaller than another model are highlighted in dark gray. The worst performing models (the highest BIC value) are highlighted in light gray.

Pseudo R^2 values were also used to compare model fit for the full model across the single indicator models. If each of the three indicators predict the gang outcome in similar ways, the pseudo R^2 values should be relatively consistent across all three models. Generally speaking, the results demonstrated this was true. At small spatial scales, the values were identical. The fact that the pseudo R^2 values are identical at small spatial scales might suggest that the gang ecology domain is unidimensional at small spatial scales and is represented equally well with any of the three indicators used here. At larger spatial scales, however, the home address and gang arrest models have larger values than the gun crime model. Once again, model comparisons are revealing the gun crime model is the least adequate representation of the gang ecology construct. However, comparisons across pseudo R^2 values are difficult to interpret since there isn't a significance test associated with them.

Table 27: Pseudo R^2 values across single indicator models

		Model A: home address	Model B: gang arrest	Model C: gun crime
Proportion Outcome	Grid cell	0.33	0.33	0.33
	Thiessen polygon	0.44	0.44	0.44
	Census block group	0.43	0.43	0.42
	Census tract	0.47	0.49	0.40
	Neighborhood	0.43	0.43	0.36
	Community	0.38	0.37	0.09
Count Outcome	Grid cell	0.29	0.29	0.29
	Thiessen polygon	0.39	0.39	0.39
	Census block group	0.43	0.43	0.42
	Census tract	0.54	0.55	0.50
	Neighborhood	0.53	0.54	0.47
	Community	0.60	0.56	0.40

Note: Largest pseudo R^2 values are highlighted in dark gray. Smallest pseudo R^2 values are highlighted in light gray.

In summary, these results indicate the three variables are *not* equivalent in their ability to explain PPD gang set space locations. The BIC comparisons revealed there was *never* a case when all three single indicator models performed equally well. The pseudo R^2 values generated similar results at smaller spatial scales. Overall, both the BIC comparisons and the pseudo R^2 values reveal the gun crime indicator was the weakest indicator. The next question explored the issue of spatial scaling.

Sub-question D: Are the significance pattern generated in the analyses from sub-questions A, B and C consistent across all levels of spatial aggregation?

Sub-question A focused on whether each of the single indicator models performed better than the baseline model. As Table 24 demonstrates, this was generally true across all six spatial scales. The single indicator models performed better than the baseline model regardless of which spatial scale was examined.

Sub-question B explored whether the single indicator models performed better than the full model. Contrary to what was expected, the full model usually was the preferred model; however, at large spatial scales (census tract, neighborhood and community level) differences in BIC values became less pronounced. The home address and gang arrest models were similar to the full model at large spatial scales.

Finally, sub-question C examined whether the single indicator models performed similarly when compared against each other. BIC comparisons indicated that one model is always stronger compared to the others; however, which variable was preferred depended on the spatial scale at which it was measured. At small spatial scales, the gun crime model was the preferred model. The home address model was more likely to be the preferred model at larger spatial scales. The gang arrest variable was usually preferred at all spatial scales except for the grid cell and thiessen polygon geographies. The pseudo R^2 comparisons revealed the indicators were more likely to be similar at small spatial scales. At the grid cell and thiessen polygon spatial units, the pseudo R^2 values were exactly the same across all three models. At large spatial scales, differences began to emerge. The largest differences were found at the community level.

In summary, spatial scaling does appear to play a role in the adequacy of each indicator to represent the gang ecology construct. At small spatial scales, all three indicators are fairly similar in their ability to represent the construct. However, at larger spatial scales (census block group and up) the gun crime variable proves inadequate compared to the other two. These difference suggest there is variation in the gang ecology construct and how it is measured across spatial scales.

Research Question 2: Are indicators able to predict gang presence and gang magnitude in similar ways?

To address the second research question, regression output was analyzed to determine if the gang indicators are able to predict gang presence, modeled as a binary variable, and gang magnitude, modeled either as a proportion or a count, in similar ways. This question was broken down into two sub-questions.

Sub-question A: Do significance patterns in the mu models match the significance patterns in the nu models?

The hurdle models generated two separate parameter estimates for a single variable in each model—*mu* parameters and *nu* parameters. *Mu* parameters estimate the mean of the non-zero values (i.e. gang magnitude) while *nu* parameters estimate the probability of gang presence. To answer this research question, significance patterns in the *mu* models were compared with the significance patterns in the *nu* models. These comparisons were made within each model. For instance, if the home address variable is significant in the *mu* models and is also significant in the *nu* models, this would suggest that this indicator can be used to predict gang presence as well as gang magnitude.

The results of 36 regression models are displayed in Table 28. Cells in this table are highlighted only if the significance pattern in the *mu* and *nu* sections of the model are *different*, meaning one coefficient was significant while the other was not. If a cell is highlighted, this means that the variables were *not* predicting gang presence and gang magnitude in similar ways. In 16 of the 36 models, the coefficient was significant in the *nu* section of the model but was not significant in the *mu* section of the model. For the home address variable, this inconsistency occurred in three models, for the gang arrest model, this happened in six models, and for the gun crime variable this happened in seven models. But, at the grid cell level, there is complete consistence for all models and outcomes. Overall, these results demonstrate that gang indicators are *not* able to predict both gang presence and gang magnitude. Oftentimes, the indicators were only able to predict the binary gang outcome (*nu* parameters) in a statistically significant way. These results are highlighted in Table 29. In this table, all cells are highlighted if the p-value for the estimated coefficient was less than .01. Generally speaking, these variables are significant when they are used to predict whether or not a gang exists in an area (*nu* parameters). These variables usually are not significant when they are used to predict how many gangs are present or how large the set space is in terms of square footage.

Table 28: Significance pattern for *mu* and *nu* sections of each model

	Model A: home address		Model B: gang arrest		Model C: gun crime	
	mu	*nu*	*mu*	*nu*	*mu*	*nu*
	gang magnitude	*gang presence/ absence*	*gang magnitude*	*gang presence/ absence*	*gang magnitude*	*gang presence/ absence*
Proportion outcome						
Grid cell	0.19*	-0.52*	0.03*	-0.24*	0.04*	-0.18*
Thiessen polygon	0.08*	-0.22*	0.02*	-0.26*	-0.02	-0.24*
Census block group	0.05*	-0.38*	0.01	-0.14*	-0.01*	-0.06*
Census tract	0.02	-0.26*	-2.6E-03	-0.14*	-2.0E-03	-0.02*
Neighborhood	-0.01	-0.17*	0.00	-0.08*	-1.9E-03*	-0.01*
Community	2.6E-03	-0.17	0.00	-0.06	1.8E-04	-0.01*
Count outcome						
Grid cell	0.08*	-0.18*	0.02*	-0.11*	0.05*	-0.09*
Thiessen polygon	-0.01	0.02	0.02	-0.19*	1.7E-03	-0.19*
Census block group	0.05*	-0.25*	0.01*	-0.09*	0.01	-0.04*
Census tract	0.03*	-0.21*	0.01	-0.11*	2.3E-03	-0.02*
Neighborhood	7.3E-04	-0.16*	6.1E-04	-0.08*	-1.2E-03	-0.01*
Community	0.01	-0.19	1.7E-03	-0.05	1.6E-04	-0.01

Note: Coefficients for *mu* and *nu* section of the models. *=p<0.01. Cells are highlighted only if *mu* and *nu* significance patterns are different. Significant coefficient is highlighted in dark gray, while the non-significant coefficient is highlighted in light gray.

Table 29: Significance pattern for *mu* and *nu* sections of each model

	Model A: home address		Model B: gang arrests		Model C: gun crime	
	mu	*nu*	*mu*	*nu*	*mu*	*nu*
	gang magnitude	*gang presence/ absence*	*gang magnitude*	*gang presence/ absence*	*gang magnitude*	*gang presence/ absence*
Proportion outcome						
Grid cell	0.19*	-0.52*	0.03*	-0.24*	0.04*	-0.18*
Thiessen polygon	0.08*	-0.22*	0.02*	-0.26*	-0.02	-0.24*
Census block group	0.05*	-0.38*	0.01	-0.14*	-0.01*	-0.06*
Census tract	0.02	-0.26*	-2.6E-03	-0.14*	-2.0E-03	-0.02*
Neighborhood	-0.01	-0.17*	0.00	-0.08*	-1.9E-03*	-0.01*
Community	2.6E-03	-0.17	0.00	-0.06	1.8E-04	-0.01*
Count outcome						
Grid cell	0.08*	-0.18*	0.02*	-0.11*	0.05*	-0.09*
Thiessen polygon	-0.01	0.02	0.02	-0.19*	1.7E-03	-0.19*
Census block group	0.05*	-0.25*	0.01*	-0.09*	0.01	-0.04*
Census tract	0.03*	-0.21*	0.01	-0.11*	2.3E-03	-0.02*
Neighborhood	7.3E-04	-0.16*	6.1E-04	-0.08*	-1.2E-03	-0.01*
Community	0.01	-0.19	1.7E-03	-0.05	1.6E-04	-0.01

Note: Coefficients for *nu* and *mu* section of the models. *=p<0.01. Cells are highlighted only if the coefficient is significant.

In summary, these results indicate that gang indicators generally are *not* able to predict both gang presence and gang magnitude. Although this was not always the case, more often than not, the indicators were only able to predict the binary gang outcome (*nu* parameters) in a statistically significant way. These tables can also be used to examine whether the results hold across spatial scales.

Sub-question B: Is the significance pattern generated by the analysis in 2A consistent across all levels of spatial aggregation?

Regression output was compared across levels of spatial aggregation. It is evident from Table 28 that the coefficient significance patterns are consistent at the grid cell level and the community level. At extremely small and extremely large spatial scales, all three indicators are able to consistently predict binary gang outcomes and continuous gang outcomes and they are consistent and significant at the grid cell level. Across the remaining four spatial scales, the variables are inconsistent. Generally speaking, this is an indication that these variables operate in different ways across spatial scales. Across all six spatial scales, the variables are more likely to be significant when they are used to predict a binary gang outcome compared to a continuous gang outcome.

In summary, the home address variable and the gang arrest variable consistently produced similar BIC and pseudo R^2 values in the regression analysis (see Table 26 on page 107 for BIC results and Table 27 on page 108 for the pseudo R^2 results). This was true across spatial scales. Furthermore, gun crime consistently had the smallest effect size. Sometimes, this variable behaved in the opposite direction as the home address and arrest variables, meaning there is less gun crime in more gang predominated locations. The second question asked whether the variables were able to predict both gang magnitude and gang presence. Oftentimes, the indicators were only able to predict the binary gang outcome (*nu* parameters) in a statistically significant way. But at the grid cell level, both presence and magnitude were significantly predicted by all three indicators. These results were generally consistent across spatial scales.

Algorithm analysis

To supplement the results of the regression models, an algorithm was created to examine how well each indicator was able to approximate PPD gang set space locations. The algorithm used either the home address or gang arrest locations to approximate gang set space boundaries. The degree to which these quantitatively derived set space polygons overlap with the PPD set space locations serve as yet another construct validity assessment.

As discussed on page 57, PPD provided a list of 4,153 individuals who were associated with 110 gangs. PPD does not classify a person as a gang member unless they have an arrest record. In addition to this list, PPD also supplied a list of gang members who were arrested between January 1, 2012 and December 31, 2015. These arrest incident locations are also used to approximate gang set space boundaries. Descriptive statistics are reported in Table 30.

Gangs are the unit of analysis for this part of the study. The 110 gang territories were approximated using either the home address of gang members or gang arrest data. It is important to note that some gang members belonged to more than one gang. In these cases, duplicate records were created.

An algorithm was used to develop approximated gang set space locations using the home address of gang members and gang arrests as inputs. A description of this process begins on page 87. The gang set space polygons that were generated by the algorithm were compared to the official PPD set space polygon file. A coefficient of areal correspondence was used to calculate the degree of overlap for each individual gang. The home address data produced coefficient of areal correspondence values that ranged from 0 (no overlap) to .66 (66% of the area is shared) with a mean value of .10. The arrest data produced coefficient values that ranged from 0 (no overlap) to .56 (56% of the area is shared) with a mean value of .07. These results are displayed in Table 30. What follows is a detailed description of these results.

Table 30: Coefficient of areal correspondence, descriptive statistics

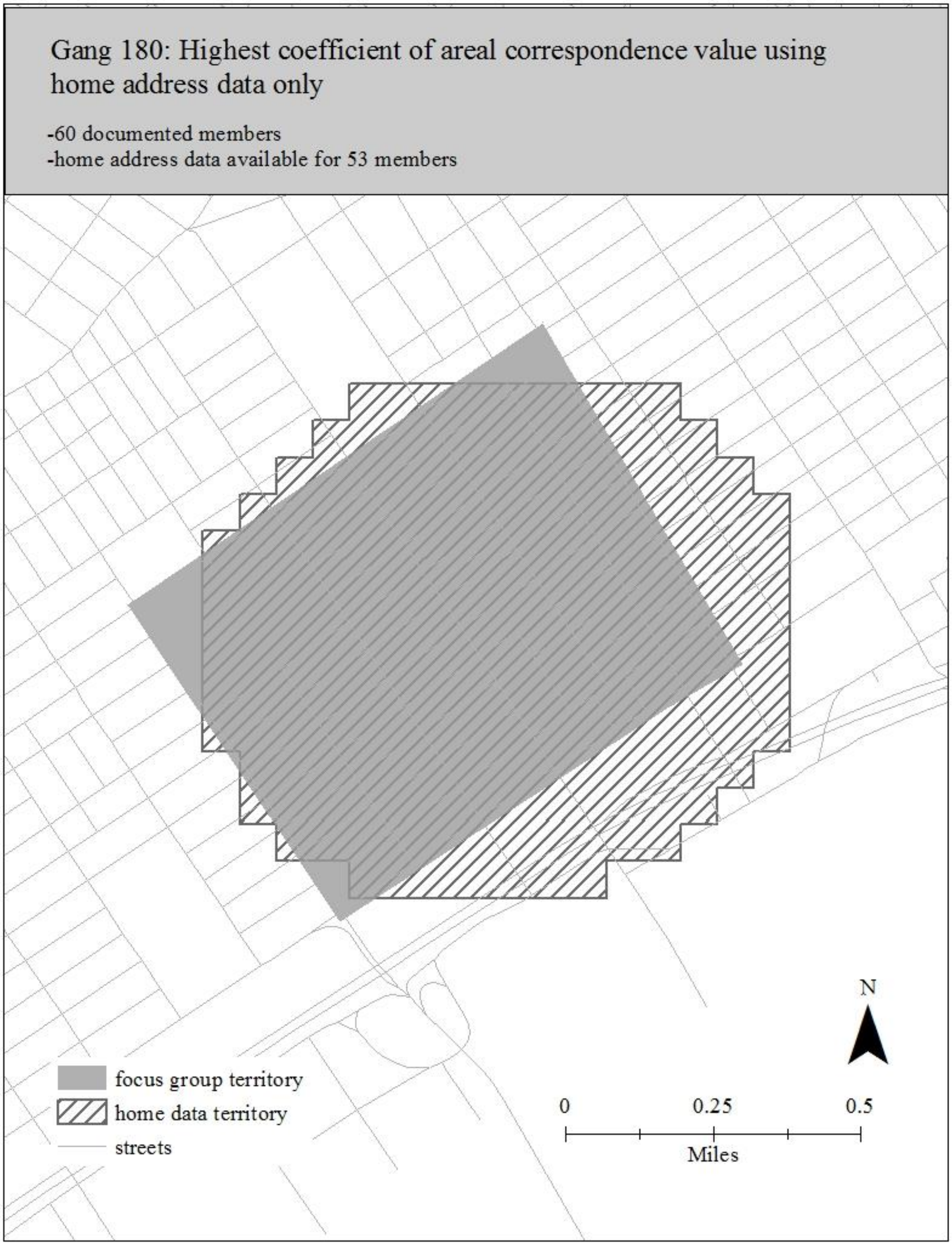
Coefficient of areal correspondence	N	Min	Max	Mean	SD	Median
Using only home address locations	110	0	0.66	0.10	0.15	0.03
Using only arrests	110	0	0.56	0.07	0.11	0.03

Home address results

Home address data were used to approximate gang set space locations for each of the 110 gangs with at least three documented members. The home address data produced coefficient of correspondence values that ranged from 0 (no overlap) to .66 (66% of the area is shared) with a mean value of .10 (Table 30). For 29 of the 110 gangs, the set space polygon that was created by the algorithm using home address data did not overlap with the official PPD set space locations (coefficient of areal correspondence = 0).

The home address data approximated the set space of gang 180 the best. This gang consisted of 60 documented members. Of those members, valid home address locations were available for 53 members (88.3%). The set space polygon that was created using these addresses, and the actual PPD set space location that was identified qualitatively during the focus group meetings, are displayed in Figure 4.

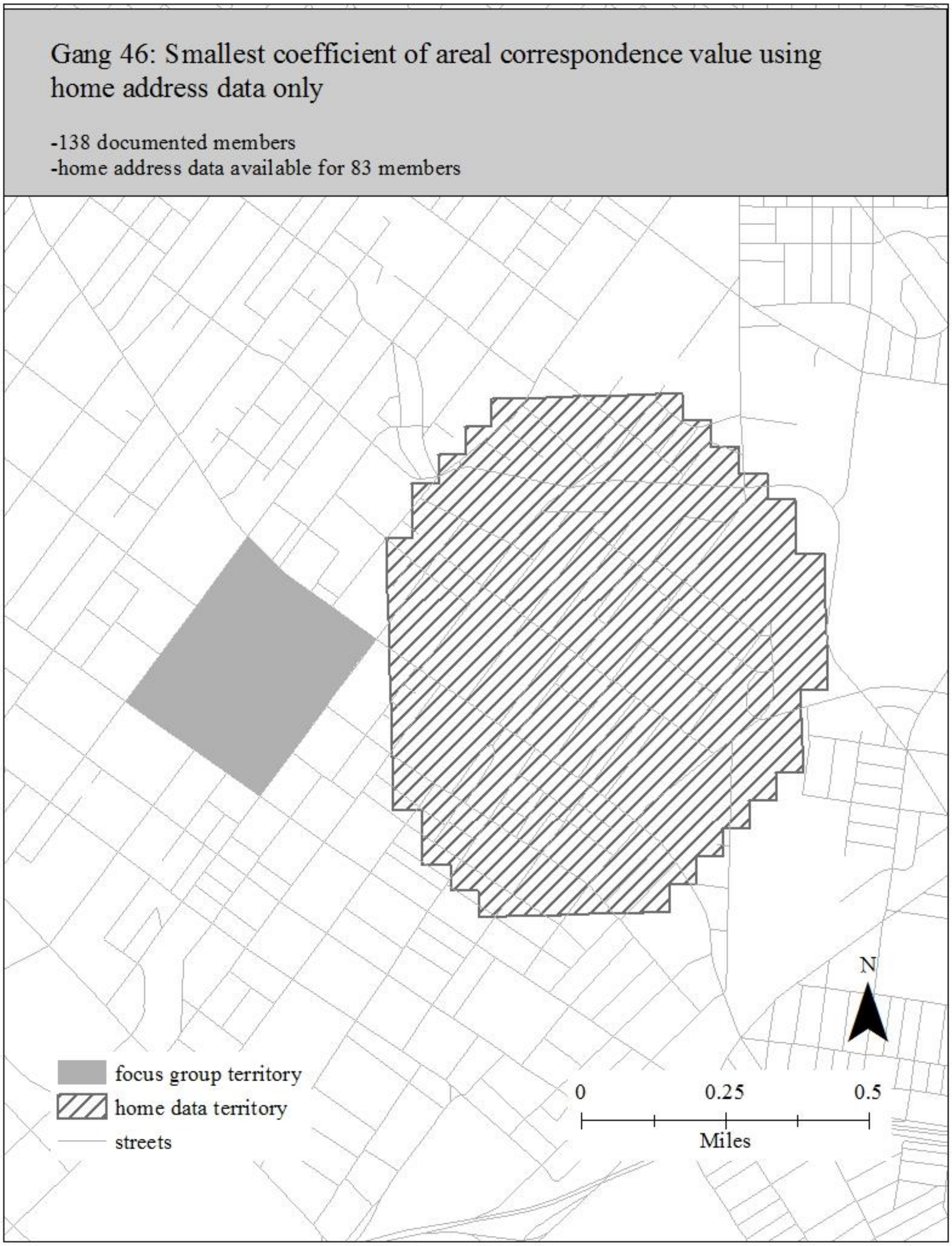
Figure 4: Gang with highest coefficient of areal correspondence value using home address data, gang 180



For 29 gangs, the set space location defined by the algorithm did not overlap with the PPD set space location. There are a number of reasons why this happened. First, for some of these gangs, the GI* statistic never reached a significant level, and therefore no cells were selected; this happened for 18 gangs. Second, for two gangs, none of the gang members were arrested between 2012 and 2015 and therefore there were no home address data available for those gangs members; this happened for gang 186 and 239. Third, for another two gangs, only one member was arrested (gang 13 and 230). The kernel density tool used in this analysis cannot generate output when there is only one record, therefore, the algorithm did not create a set space polygon for these two gangs. Finally, for the remaining seven gangs, a set space polygon was created but the location did not overlap with the PPD set space location.

Gang 46 is one of the gangs where the home address data did not create a set space polygon that overlapped with the PPD set space location. This gang was comprised of 138 documented members. Of those members, valid home address locations were available for 83 members (60.1% of all members). The set space that was created using these addresses, and the PPD set space that was identified qualitatively during the focus group meetings, are displayed in Figure 5.

Figure 5: Gang with lowest coefficient of areal correspondence value using home address data, gang 46



Gang arrest results

Gang arrest data were also used to approximate gang set space locations for each of the 110 gangs with at least three documented members. Overall, the arrest data produced coefficient of correspondence values that ranged from 0 (no overlap) to .56 (56% of the area is shared) with a mean value of .07 (Table 30 on page 115). For 16 of the 110 gangs, the set space that was created with the incident locations created a set space polygon that did not overlap with the PPD set space location (coefficient of areal correspondence = 0).

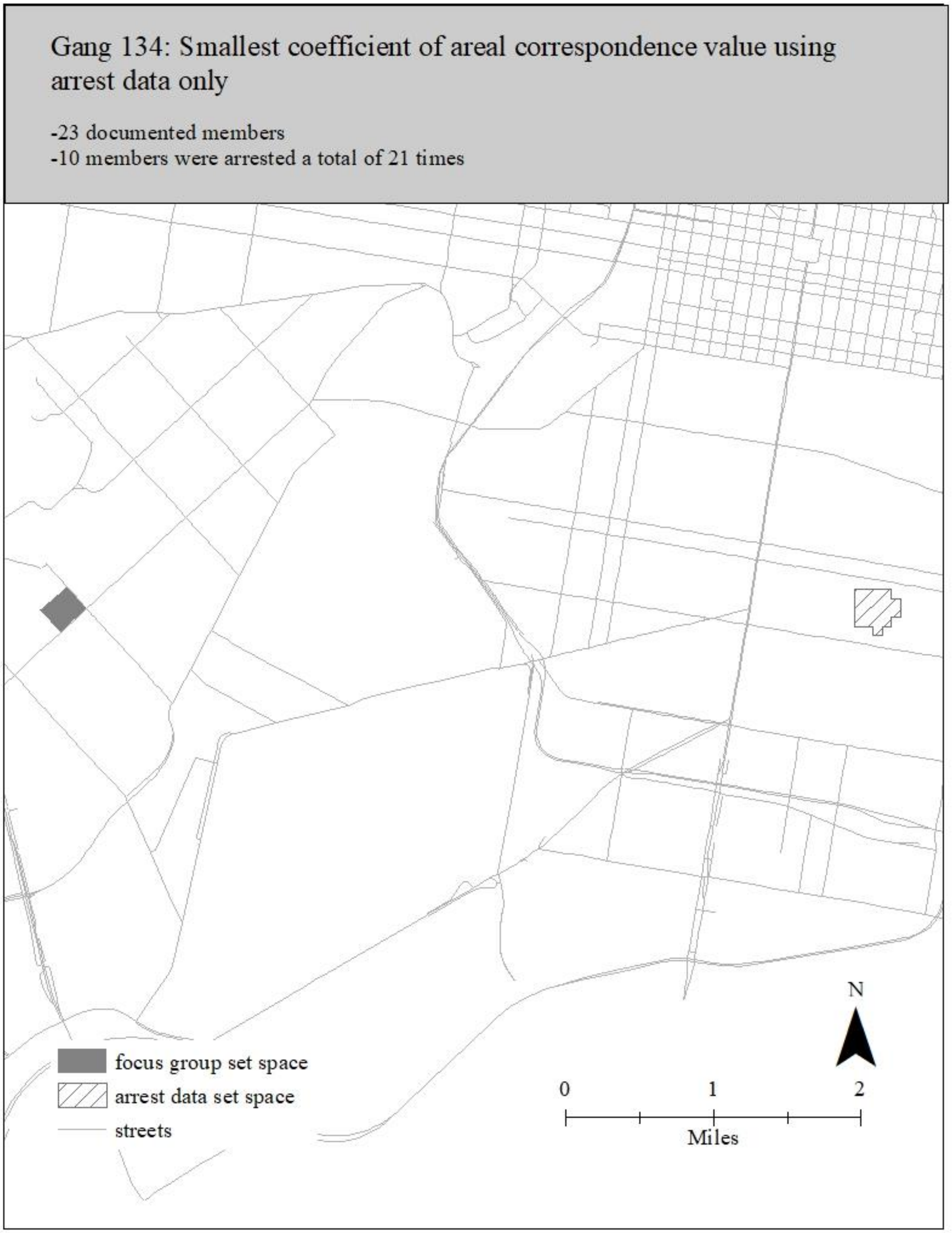
The arrest data approximated the set space of gang 180 the best. This was also the gang that the home address data identified the best. This gang had a total of 60 documented members. Of those members, 54 members were arrested at least once between 2012 and 2015 (90% of all members). Those 54 members were arrested a total of 180 times. Those 180 arrest incidents were used to approximate the set space boundaries. The set space polygon that was created using these arrest data and the PPD set space location are displayed in Figure 6.

Figure 6: Gang with highest coefficient of areal correspondence value using arrest data, gang 180

For 16 gangs, the set space polygon generated using the arrest data did not overlap with the PPD set space locations. This happened for various reasons. For eight of these gangs, the GI* statistic never reached a significant level, and therefore no cells were selected. For two gangs (gang 186 and 239), none of the documented members were arrested between 2012 and 2015 and therefore there were no arrest data associated with those gangs. For one gang (gang 13), there was only one arrest associated with that gang. Once again, the kernel density tool cannot generate output when there is only one record, so a set space polygon was not approximated for this gang. For the remaining five gangs, a set space polygon was created but it did not overlap with the PPD set space location.

Gang 134 was one such gang where the set space generated by the arrest data did not overlap at all with the PPD set space (see Figure 7). In fact, the approximated set space polygon was about 5.5 miles away from the PPD set space location. This gang consisted of 23 documented members. Of those 23 members, ten of them were arrested 21 times from 2012-2015. The large distance between the approximated set space polygon and the actual PPD set space location may be related to two peculiarities of the approximated location. First, the approximated set space location is just outside a high school. It is possible that the 21 gang arrests that cluster in this area are related to the school. What is also notable about this location is its proximity to four other gang set spaces. It is possible that the 21 incidents that created the approximated gang set space polygon were related to rivalries with these other gangs.

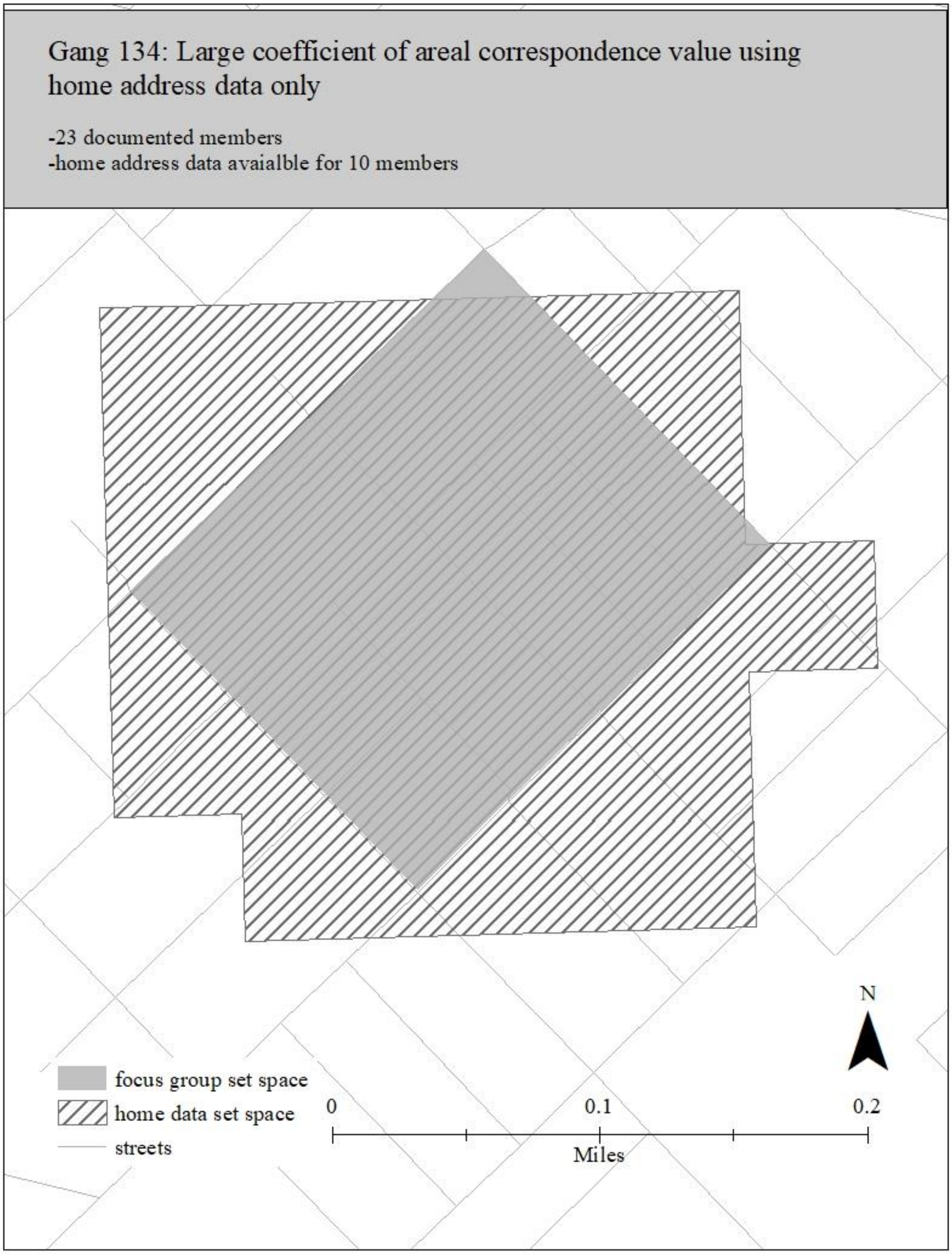
Figure 7: Gang with lowest coefficient of areal correspondence value using gang arrest data, gang 134



Oddly enough, for this particular gang, the home address data were very good at modeling the set space of gang 134 (see Figure 8). The coefficient of areal correspondence generated when comparing these two set space polygons was .49, which was among the highest values in the entire dataset (n=110 gangs). While arrest data were not effective in modeling this gang set space, the home address data were extremely effective.

The fact that the algorithm generated two distinct set space polygons for this gang is surprising. Not only did the home address and the arrest points cluster in two different locations for this gang, those points were clustered in a statistically significant way as per the Gi* statistic. These two distinct locations suggest, at least for some gangs, that gangs may have multiple ecologies which may or may not be spatially overlapping. In other words, there may be different types of ecologies for some types of gangs. This finding is important for law enforcement officials tasked with coordinating crime reduction strategies targeting gangs. For some gangs, the places where they commit crime are nowhere near the places they defend as their turf or where they live. An intervention that targets crime related to this gang would need to be focused in the location where the crimes are occurring. From a theoretical perspective, the large distance between these two locations highlights the possibility that gang members' awareness spaces can extend to a variety of locations that may not be related to gang set space.

Figure 8: Set space created for gang 134 using home address data



Research Question 1: Do indicators of gang ecologies identify the same construct?

Research question 1 was also addressed using the results of the algorithm analysis. These sub-questions are summarized next.

Sub-question E: How much do the quantitatively derived gang polygons overlap with the practitioner defined polygons?

The coefficient of areal correspondence was used to measure the degree to which each indicator identified the qualitatively derived set space boundaries. If the home address data and arrest data are equally relevant, the coefficient of correspondence values should be approximately the same across both indicators. To test this research question, a t-test was used to test the differences-in-mean coefficient values across these two groups. If both indicators are equally effective in identifying gang ecologies, the test statistic will *not* reach statistical significance indicating the mean values are not significantly different. This part of the analysis examined Messick's **content** aspect of construct validity.

A Shapiro-Wilk normality test indicated the home address and arrest overlap values are not normally distributed (home variable: $W=0.69$, $p<.001$; incident variable: $W=.67$, $p<.001$). As a result, a dependent samples Wilcoxon Signed-Rank Test was used to determine if the mean overlap values for the home address data and arrest data were the same. Results show that the mean overlap for the home address data was greater than the mean overlap using the arrest data. This difference was statistically significant. Although the variables were not normally distributed, the results of a dependent samples t-test was run to supplement the non-parametric test. This test proved to be significant as well (see Table 31).

Table 31: T-test output comparing mean overlap for home address and arrest variables

Sample estimates				
Home address mean	arrest mean	t-value	df	p-value
0.096	0.070	2.696	109	0.008

Sub-question F: Does the percentage overlap vary by type of gang?

Coefficient of areal correspondence values are also compared across the three PPD gang typologies. Separate ANOVA models were run for each typology. If each indicator is equally relevant (from a statistical perspective) for each type of gang, the test statistic should not reach the critical value in any of the models. If methods are better at predicting some types of gangs compared to other types, post hoc tests are used to determine which indicators are most relevant. This analysis examines Messick's **generalizability** aspect of construct validity. Descriptive statistics and 95% confidence intervals around the sample mean are displayed in Table 32 and Figure 9 using the home address data, and Table 33 and Figure 10 using arrest data.

Table 32: Home address variable overlap, descriptive statistics by typology

	Mean	95% CI	N	Min	Max	SD
Corner drug sales	0.059	0.034	39	0	0.48	0.10
DTO	0.029	0.019	22	0	0.14	0.04
Street gang	0.156	0.054	49	0	0.66	0.19

Figure 9: Home address variable overlap, 95% confidence intervals by typology

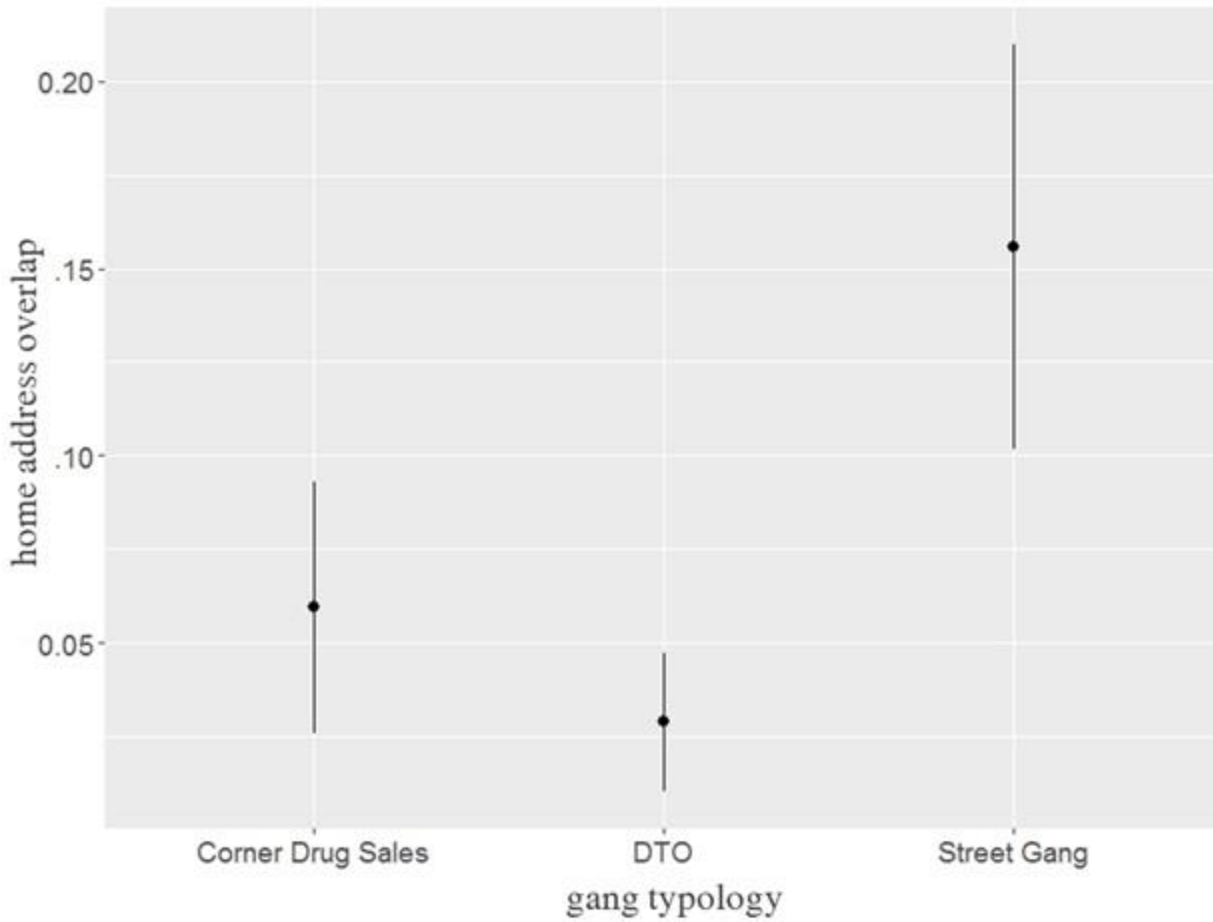
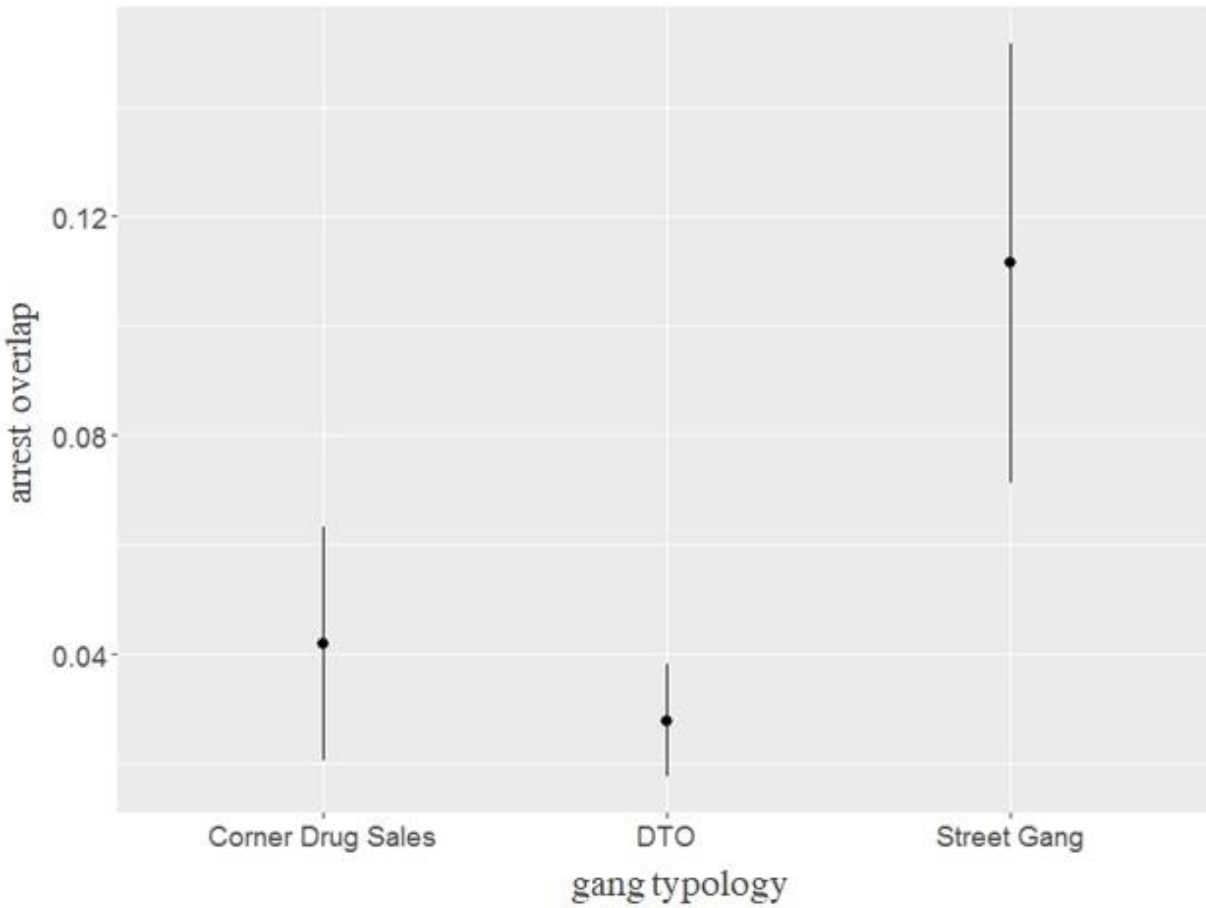


Table 33: Arrest variable overlap, descriptive statistics by typology

	Mean	95% CI	N	Min	Max	SD
Corner drug sales	0.042	0.021	39	0	0.301	0.065
DTO	0.028	0.010	22	0	0.097	0.023
Street gang	0.111	0.040	49	0	0.565	0.140

Figure 10: Arrest variable overlap, 95% confidence intervals by typology



The Levene's test for homogeneity of variance revealed the variances of different typologies were not equal. Using the home address data, the Levene's test was significant at the .01 level, $F(2,107)=8.586$, $p<0.01$. Using the arrest data, this test was also significant at the .01 level, $F(2,107)=8.340$, $p<0.01$. Based on these results, it can be assumed that the population variances are not equal across groups. A robust ANOVA analysis was run which makes adjustments for differences in group variances. Specifically, Welch's F was applied to the data.

Using the home address variable, Welch's $F(2, 69.956)=10.434$, $p<0.01$, indicates that the mean overlap values do vary across different typologies. The same results were found using the arrest data, Welch's $F(2, 68.033)=8.479$, $p<0.01$.

Post-hoc tests reveal the significant differences existed between the street gang typology the other two typologies. These differences were significant at the .01 level and they hold for both the home address data (Table 34) and the arrest data (Table 35). This means that both indicators, home address and arrest, were able to identify the gang set space locations of street gangs better than they were able to identify the set space of corner drug crews or drug trafficking organizations (DTOs). This is an indication that the generalizability of gang ecology indicators may be dependent on the type of gang under investigation. The home address and arrest variables appear to be adequate representations of some types of gangs (street gangs) but are less adequate in representing corner drug crews or DTOs. These differences may because there are different facets of gang ecologies that are linked to gang typologies.

Table 34: ANOVA Post-hoc tests for home address variable, comparisons across typologies

Bonferroni adjusted p values		
	Corner drug sales	DTO
DTO	1	
Street gang	**0.0059**	**0.0021**
Holm adjusted p values		
	Corner drug sales	DTO
DTO	0.4176	
Street gang	**0.004**	**0.0021**
Benjamini-Hochberg adjusted p values		
	Corner drug sales	DTO
DTO	0.4176	
Street gang	**0.003**	**0.0021**

Table 35: ANOVA Post-hoc tests for arrest variable, comparisons across typologies

Bonferroni adjusted p values		
	Corner drug sales	DTO
DTO	1	
Street gang	**0.0058**	**0.0056**

Holm adjusted p values		
	Corner drug sales	DTO
DTO	0.6065	
Street gang	**0.0056**	**0.0056**

Benjamini-Hochberg adjusted p values		
	Corner drug sales	DTO
DTO	0.6065	
Street gang	**0.0029**	**0.0029**

The results of the ANOVA analysis indicate that both indicators are able to identify street gangs better than corner drug crews or DTOs. However, this test does not tell us if one indicator works better than the other for each typology. For instance, does the home address variable predict the street gang typology gang set space locations better than the arrest variable does? To answer this question, a dependent samples t-test was used to see if the home address or arrest mean overlap values were the same *within* each typology. The results indicate that mean values within each typology do not differ for corner drugs sales or DTOs—both indicators identify gang set space to the same degree. However, the mean overlap values produced by the home address variable are larger than the arrest variable and this difference proved to be statistically significant (see Table 36). The Wilcoxon Signed Rank test, the non-parametric version of this test, identified statistically significant differences between the home address and arrest variables across the corner drug sale and street gang categories.

Table 36: T-tests comparing mean overlap values across typologies

	Home mean	Incident mean	t value	df	p-value
Corner drug sales	0.059	0.042	1.622	38	0.113
DTO	0.029	0.028	0.145	21	0.886
Street gang	0.156	0.111	2.261	48	0.028

Sub-question G: Does the percentage overlap vary by gang size?

The analysis described above was also run after categorizing gangs by size. It is possible that the home address or arrest indicators are better at identifying large gangs compared to small gangs. Mean overlap values and the 95% confidence intervals are presented in Table 37 and Figure 11 using the home address variable and Table 38 and Figure 12 using the arrest variable.

Table 37: Home variable overlap, descriptive statistics by gang size

	Mean	95% CI	N	Min	Max	SD
20 members or less	0.038	0.027	41	0	0.421	0.087
21 to 75 members	0.143	0.024	56	0	0.664	0.181
More than 75 members	0.081	0.026	13	0	0.259	0.092

Figure 11: Home address variable overlap, 95% confidence intervals by gang size

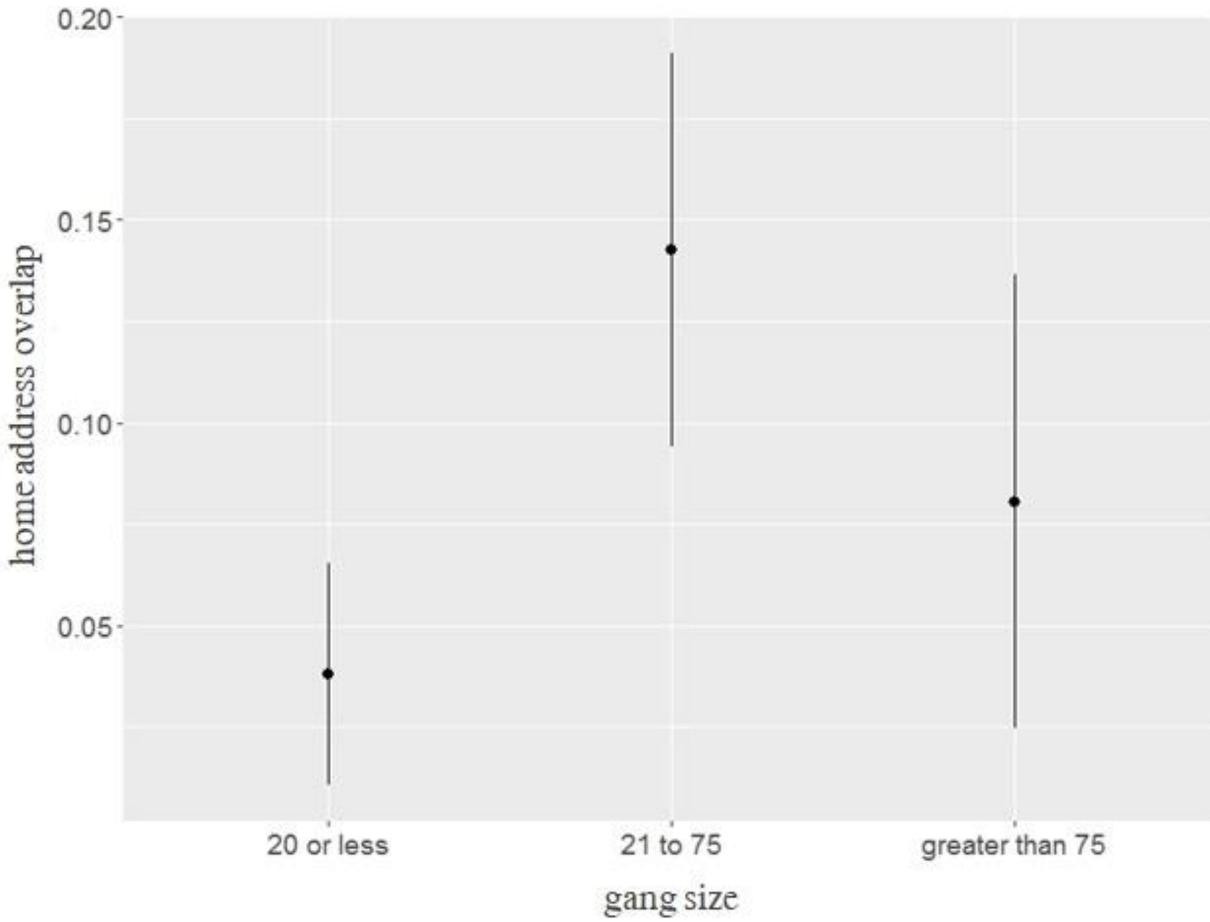
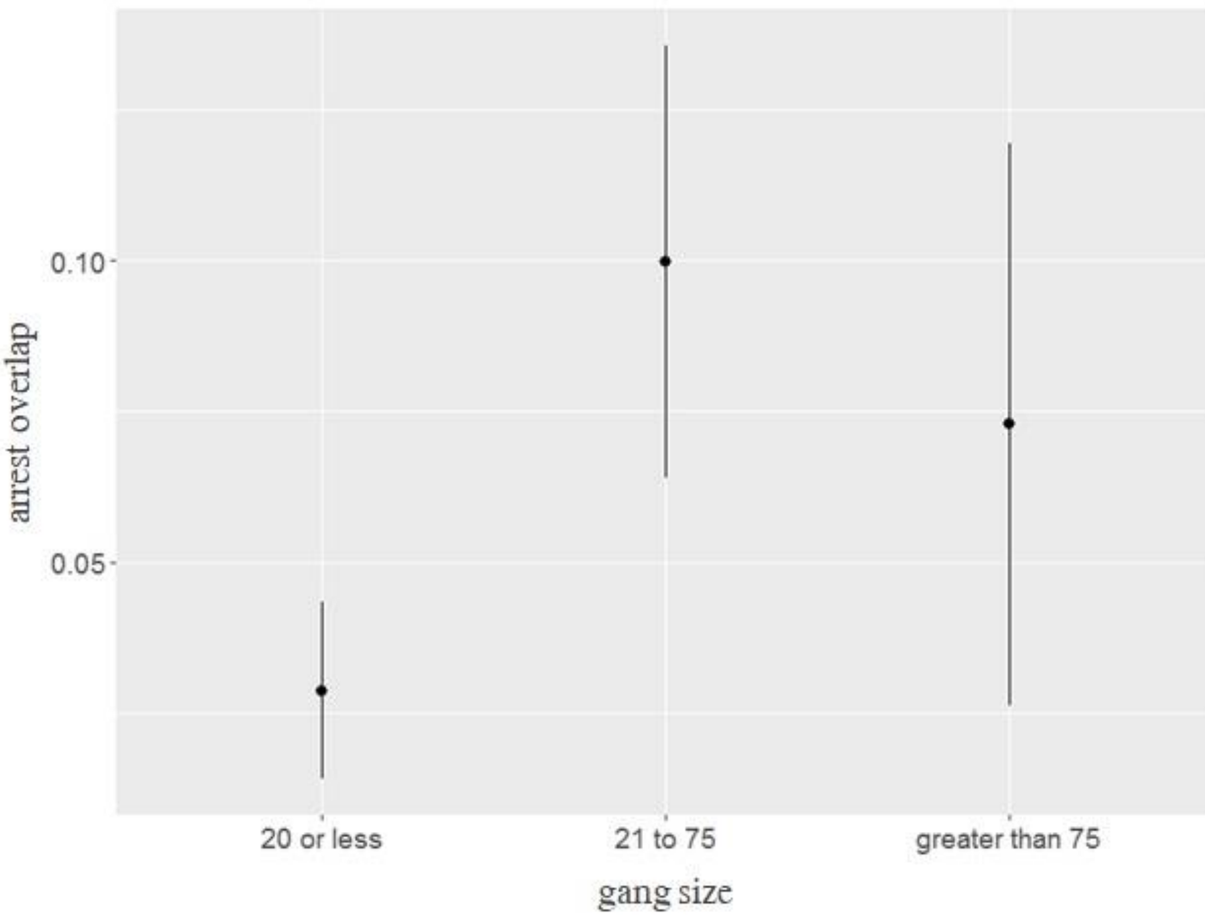


Table 38: Arrest variable overlap, descriptive statistics by gang size

	Mean	95% CI	N	Min	Max	SD
20 members or less	0.029	0.015	41	0	0.216	0.046
21 to 75 members	0.100	0.036	56	0	0.565	0.134
More than 75 members	0.073	0.047	13	0	0.206	0.077

Figure 12: Arrest variable overlap, 95% confidence intervals by gang size



The Levene's test for homogeneity of variance revealed the variances of each group were not equal for the home address data but were equal for the arrest data. Using the home address data, the Levene's test was significant at the .01 level, F(2,107)=6.224, p=0.003. Using the arrest data, the Levene's test was not significant at the .01 level, F(2,107)=4.769, p=0.011. Based on these results, it can be assumed that the population variances are not equal across groups for the home address data but are equal across groups for the arrest variable. A robust ANOVA analysis was run which makes adjustments for differences in group variances in the home data. Specifically, Welch's F was applied to the data.

Using the home address variable, Welch's F(2, 37.095)=7.171, p=0.002, which means the mean overlap values do vary across gangs of different sizes. The same results were found using the arrest data, F(2,107)=5.606, p=0.005. Although the roust test was not needed for the arrest data, the robust version was used to generate results that are consistent with the home address data. Using the robust test, Welch's F(2, 31.733)=7.783, p=0.002, which means the mean overlap values using arrest data also vary across gangs of different sizes.

Post-hoc tests reveal the significant differences existed between gangs with 20 members or less and the 21 to 75 member gangs. These differences were significant at the .01 level and they hold for both the home address data (Table 39) and the arrest data (Table 40). This means that both indicators, home address and arrest, were able to identify the medium sized gangs (21-75 members) better than they were able to identify the territories of small gangs. The difference was not significant for large gangs.

Table 39: ANOVA Post-hoc tests for home address variable, comparisons across gang size

Bonferroni adjusted p values		
	20 or less	21 to 75
21 to 75	**0.002**	
Greater than 75	1.00	0.493
Holm adjusted p values		
	20 or less	21 to 75
21 to 75	**0.002**	
Greater than 75	0.354	0.329
Benjamini-Hochberg adjusted p values		
	20 or less	21 to 75
21 to 75	**0.002**	
Greater than 75	0.354	0.247

Table 40: ANOVA Post-hoc tests for arrest variable, comparisons across gang size

Bonferroni adjusted p values		
	20 or less	21 to 75
21 to 75	**0.003**	
Greater than 75	0.548	1.00

Holm adjusted p values		
	20 or less	21 to 75
21 to 75	**0.003**	
Greater than 75	0.365	0.399

Benjamini-Hochberg adjusted p values		
	20 or less	21 to 75
21 to 75	**0.003**	
Greater than 75	0.274	0.399

The results of the ANOVA analysis suggest both indicators are able to identify gangs with 21-75 members better than gangs with 20 or fewer members. However, the results of the ANOVA analysis do not tell us if one indicator works better than the other for each category. For instance, does the home address variable predict the territories of gangs with 21-75 members better than the arrest variable does? To answer this question, a dependent samples t-test was used to see if the home address or arrest mean overlap values were the same *within* each category. The results indicate that mean values within each typology do not differ across gangs with fewer than 20 members or more than 75 members (Table 41). Moderately sized gangs, however, were better approximated using the home address data. The same significance patterns were generated using the non-parametric Wilcoxon Signed Rank test.

Table 41: T-tests comparing mean overlap values across gang size

	Home address mean	Arrest mean	t value	df	p-value
Less than 20	0.038	0.029	0.831	40	0.411
20 to 75	0.143	0.100	2.523	55	0.015
Greater than 75	0.081	0.073	1.300	12	0.218

Summary of the results

Gang set space boundaries used in this study were provided by the Philadelphia Police Department (PPD) Criminal Intelligence Unit (CIU). These boundaries were initially validated over the summer of 2013 as part of a gang audit organized by PPD. The data were audited a second time from June 2014 through October 2015 by Dr. Caterina Roman in a collaborative multi-agency gang audit with the PPD using a series of focus groups.

In addition to the PPD gang set space boundaries, four years of gang member arrest records were also used as data sources (January 1, 2012 through December 31, 2015). This dataset included the x and y coordinates of the arrest incident location as well as the census block where the gang member lived. Philadelphia Police incident data were also used as a data source, including all crime incidents (gang related or not) that involved a firearm.

These datasets allowed me to compare four different sources of information that have been used to identify gang locations:

- official gang set space locations identified by practitioners
- where gang members live
- gang-related crime (gang arrests)
- gun crime

The data were analyzed using two analysis techniques. The first technique involved running a series of regression models that used the home address, arrest and gun crime variables to predict the set space locations provided by CIU. In the second analysis, an algorithm created gang set space polygons using either the locations where gang members live or the locations of gang arrests. The set space polygons created by the algorithm were compared to the PPD set space locations supplied by CIU.

The results of the regression analysis comparing the single indicator models revealed that the home address and arrest variables were better at explaining the spatial distribution of the CIU gang set space locations than the gun crime variable. However, the link between these gang indicators and the CIU gang set space locations was complex. Oftentimes, in the single indicator models, the home address and the arrest variables were significant when predicting a binary gang outcome (whether or not there was any gang presence in an area), but they were not significant when estimating a continuous gang outcome, i.e. how many gangs were present, or how large the territories were in terms of square footage. This means the home address and arrest variables were limited in their ability to explain the spatial distribution of the CIU set space boundaries.

The spatial analysis used an algorithm to approximate gang set space locations. The results of this analysis show the locations were gang members live and gang arrests are able to approximate the actual gang set space locations equally well. However, the degree of overlap compared to the size of both set space polygons (the PPD set space *and* the set space generated by the algorithm) was relatively small. The mean overlap was only 10% using the home address locations and was just 7% using the arrest data.[18] Despite this small degree of overlap, it is worth noting that 80% of the time, the approximated gang set space did at least partially overlap with the actual gang set space to some extent.

[18] The median overlap was .03 for both indicators. The interquartile range was .11 for the home variable and was .08 for the incident variable.

Overall, this study found that gang set space locations can only be identified quantitatively using official police data to a modest degree. However, the regression analysis and the algorithm output both highlight limitations with this method of gang identification. On average, these data sources only created 7-10 percent shared overlap with the actual gang set space locations. Figure 13 provides an example of what 10 percent shared area looks like. In this figure, it is evident that the home address data created an approximated gang set space polygon that was much larger than the actual gang set space identified by CIU. In fact, the approximated set space measured .53 square miles, while the actual gang set space was only .05 square miles—the approximated set space was nearly 11 times the size of the PPD set space. While the approximated gang set space covered 100% of the PPD gang set space, the shared area was only 10 percent. Levels of overlap between 7 and 10 percent are probably of questionable value from a policy perspective.

Figure 13: Example of 10% overlap

The regression analysis highlights the fact that home address and arrest data are only able to explain some, but not all, of the spatial distribution of the CIU gang set space. Other variables may be at play which were not included in my models. This analysis does clearly highlight the inadequacy of using gun crime incidents in explaining gang locations. This is discussed further in the discussion chapter (see page 145).

The spatial analysis revealed gang set space identification may be automated to a certain extent; however, the algorithm was able to identify some types of gangs better than others. Specifically, it was able to identify street gangs better than it was able to identify corner drug crews or drug trafficking organizations. If police departments have the desire to identify gang locations using an automated process it is important they are aware of this limitation. Furthermore, given that formal identification of a gang area might have legal ramifications, an overlap measure of between 7 and 10 percent might be of questionable value to police departments.

CHAPTER 5: DISCUSSION

This goal of this dissertation was to systematically evaluate the validity of various gang ecology indicators using Messick's unified approach to construct validation. Specifically, this study investigated the extent to which different variables and methods are able to identify gangs from a geographic perspective.

Two main research questions were addressed:

1. Do indicators of gang ecologies identify gangs in similar ways?
2. Are indicators able to predict binary gang outcomes in the same way they are able to predict continuous gang outcomes?

This chapter will begin with a summary of the results and possible explanations for why these results emerged. The implication of the findings are then discussed in the context of theory and policy. The chapter concludes with a discussion of the limitations of the study and directions for future research.

Research Question 1

Do different indicators of gang ecologies reflect the same construct?

This research question was addressed using a number of statistical tests to evaluate the results of regression analyses as well as output from an algorithm that was used to approximate gang set space locations. Two main findings emerged.

(1) The home address variable and the arrest variable consistently produced similar BIC and pseudo R^2 values in the regression analysis (see Table 26 on page 107 for BIC results and Table 27 on page 108 for the pseudo R^2 results). This was true across all six spatial scales. Similar results emerged when these variables were used to approximate gang set space locations using an algorithm. T-tests reveal the accuracy of both

indicators were similar (see Table 31 on page 127). **Generally speaking, both variables perform equally well based on the results of both analyses.**

(2) The gun crime variable consistently had the smallest effect size. Sometimes, this variable behaved in the opposite direction as the home address and arrest variables, meaning there is less gun crime in gang dominated locations. In other words, it did not model gang areas as well as the gang member home address and gang arrest measures.

Why would the home address and gang arrest variables reflect the same construct?

In this study, the locations where gang members live and gang arrests were found to be similar in their ability to predict PPD gang set space. PPD gang set space represent the places where law enforcement see gang members. Two explanations may account for the similarity of these two indicators. First, while some differences emerged across PPD gang typologies (i.e. DTO, street gang, corner drug sales), PPD gang set space generally align with gang member living places. Second, the PPD gang set space also align with gang criminal behavior. Again, some differences emerged across PPD gang typologies, which will be discussed in the paragraphs that follow.

Crime pattern theory (Brantingham and Brantingham 1993) would support both of these possibilities. This theory would classify home address locations as activity nodes that form an anchor for awareness spaces of gang members. Awareness spaces are areas people are familiar with; they may include where a person lives or where they travel to for work, school or recreation. The theory suggests individuals are most likely to engage in criminal or deviant acts in locations they are familiar with, i.e. within their awareness space. If this is true of gang members, home address locations have a potentially theoretical foundation as anchoring gang set space and gang arrests.

The algorithm output identified some exceptions to this finding. For some gangs, arrest locations and home address locations clustered in different places. For example, the set space polygon generated by the arrest data for gang 134 was over five miles away from the set space identified in PPD's gang audit (see Figure 7 on page 123). The large distance between these two locations highlights the possibility that gang members' awareness spaces can extend to a variety of locations that may not be related to gang set space. For instance, it is possible that the high school near the approximated set space polygon functioned as an activity node for members of gang 134, and the arrest incidents that clustered in this location were related to the school, not activities of the gang. In this dissertation, gang crime was operationalized using a member-based definition (Maxson and Klein 1990). An arrest incident was classified as a gang incident if the person arrested for the crime was a documented gang member. This is a rather broad definition because it includes incidents that may be completely unrelated to gang activity. In the example described previously, the incidents that clustered around the school involved gang members but may have been related to school activities, not gang activities. In the case of gang 134, the approximated gang set space polygon may have been in the correct location had a more restrictive motive-based definition of gang crime been used. Using this definition, an arrest would only be classified as a gang incident if the incident was actually *motivated by gang activities* such as territoriality, retaliation or recruitment. Some police departments do have protocols in place to classify an incident as gang-motivated. The Philadelphia Police Department is currently identifying gang-motivated shooting incidents. Determining whether an incident is motivated by gang activity can be a difficult task, especially when it is unclear if other people were involved in the incident and their identity as gang members.

Why would gun crime be a weak predictor of gang territories?

Another finding related to the first research question was the weak and sometimes theoretically inconsistent behavior of the gun crime variable in predicting practitioner defined gang set space boundaries. When the gun crime variable reached statistical significance in the regression analysis, it consistently had the smallest effect size compared to the other two variables.

The question then becomes, why would the gun crime variable behave in a different manner than the home address and the gang arrest variables? Some research has demonstrated that when gang members engage in criminal activity, they tend to engage in a wide range of offenses. This has been described as "cafeteria-style offending" (Klein 1995: 68). Klein found that most gang members are involved in different types of crime, not just violent crime or gun-related crime. Identifying gang areas using only gun-involved crime ignores the fact that gang members engage in a wide range of offenses. This was particularly true of the gang data used in this study. The gang arrests in this study involved both violent and property crime (see Table 6 on page 43 for a list of gang related offenses in this study). In fact, all Part I offenses were represented in the gang arrest file. In addition to those, many other types of offenses were also represented, including forgery, vandalism, prostitution, traffic violations, loitering, and even embezzlement. Gang members in Philadelphia engage in a wide variety of offenses; therefore, using just one type of crime to represent all gang activity may not be the best way to model gang locations.

A related explanation for these findings considers the fact that the arrest incidents and the gun crime incidents were very different from each other in this study. Of the 7,488 arrests used in this study, only 1,260 (~17%) of those incidents involved a firearm. When gangs are approximated using particular types of crime, such as homicide or gun involved crime, the researcher is making the assumption that *all* of these types of crimes are related to gang activity. This is not necessarily the case. In 2012, approximately 16% of all homicides in the United States were related to gang activity (Egley, Howell, and Harris 2014). Given the size of the gang population, this represents a substantial percentage of all homicide incidents. However, it is clear from this figure that not *all* homicides are related to gang activity. This was also true of gun-related crime used in the current study.

I chose to use all reported gun crime incidents (n=26,865) to explain the locations of gang set space in this study. The gun crime variable included all incidents that involved a firearm, regardless of whether or not the incident involved a gang member or was related to gang activities. This was done because some researchers have chosen to use certain types of crime, such as homicide or gun crime, to measure gang activity when lists of gang members or gang crime incidents are not available. In the current study, I investigated whether *gun crime* as a general crime category related to gang activities could be used to identify gang set space to approximate methods used by researchers who do not have access to a list of documented gang members. This indicator proved to be a very weak predictor of gang set space.

In this study I was able to query how many of the gun incidents met the member-based definition of gang crime. Of the 26,865 gun incidents used in this study, only 1,260 gun incidents (~5%) led to the arrest of a documented gang member; meaning, only 5% of all gun crime incidents were actually gang-related. Based on these figures, the link between *all* reported gun crime and gang locations would be weak since most of the gun crime is *not* related to gang members, at least as the known offender. The 26,865 gun incidents, therefore, were not suitable predictors of gang territories.

Another reason why the gun crime variable was only weakly correlated with gang ecologies in this study may be related to the idea that gangs may insulate communities from violent crime to some degree. Some researchers have found evidence that gang members provide social control within communities (Taylor 2001: 287; Venkatesh 2008). Mary Pattillo describes such a scenario in a middle class black neighborhood: “Ironically, having an organized gang in the neighborhood has, in some respects, translated into fewer visible signs of disorder, less violence, and more social control” (Pattillo 1998: 767). Bursik and Grasmick (1993) suggest that gangs develop parochial and public networks of social control which result in lower levels of crime.

Other researchers, however, have found evidence to suggest that communities with gangs experience more crime compared to communities that do not have gangs (Huebner et al. 2014; Taniguchi, Ratcliffe, and Taylor 2011). It is possible that gangs link to higher levels of crime in some contexts but lower levels of crime in other contexts. Spatial scaling may also have a role to play. Taniguchi and colleagues aggregated their data to small spatial units defined using thiessen polygons around street intersections, Huebner and colleagues used census tracts, while Bursik and Grasmick were looking at larger spatial units at the neighborhood and community level. This dissertation did not attempt to identify which contexts would produce a positive link between gangs and crime and those that would produce a negative link. Disentangling this link could be an avenue for future research to explore.

Research Question 2

Can indicators predict the presence versus absence of gangs (binary gang outcome) as well as the magnitude of gang ecologies (continuous gang outcome).

This research question was addressed using a study of various gang indicators aggregated to different spatial units. The data were analyzed using regression models; specifically, hurdle models that allow the variables to predict zero and non-zero values separately. One main finding emerged from this analysis: The home address, arrest and gun crime variables were more effective in predicting gang presence outcomes compared to the magnitude (continuous) gang outcome. This was true across all spatial scales.

Why would the gang indicators predict presence outcomes better than magnitude outcomes?

The variables used in this study were able to predict whether or not a gang exists in an area, but they were not significant when used to predict how large a gang set space is (i.e. proportion) or how many gangs (i.e. count) exist in an area (see Table 28 on page 112). These findings suggest gang presence is easier to quantify than gang magnitude. A number of theories, which were not directly tested in this dissertation, can be used to explain this finding.

One explanation for these findings can be found in the field of ecology through the concept of an ecological niche. The functional role, or niche, of places that have gangs may be qualitatively different from places that do not have gangs. This concept has been applied in a number of ways, both in the fields of sociology (Hawley 1950) and biology (Hutchinson 1957; Hutchinson 1965).

In the field of sociology, the idea of an ecological niche was introduced by Amos Hawley (1950) who explained that neighborhoods serve functions for the larger society. "A community may be viewed as an organization of niches, since the activities of each class of organism influences the activities of every other class in the association" (Hawley 1950: 44). These functional niches are found in many aspects of life, including culture, the economy and politics. In many cities, for example, you might find an area that has been designated, or politically zoned, as the sports district or a financial district. Many cities also have a bar district with a high concentration of drinking facilities. Hawley explains how areas can be zoned by non-political means and that some of these functions can be related to criminality. Niches can be encouraged, sustained, or grown through political (Logan and Molotch 1987) and cultural means as well

(Molotch, Freudenburg, and Paulsen 2000). For an example, a certain area of the city may function as a drug market or a red light district; people who are looking to engage in those activities know where they need to go because that is the ecological niche of that particular area. Similarly, with respect to the current study, gangs may serve a functional role within communities. This role could be, for instance, to supply an area with drugs, provide camaraderie and social support for youth or to provide protection or social services to communities that are neglected by local government. Using Hawley's version of a functional niche, it is possible that locations that have gangs operating within them are qualitatively different from locations that do not have gangs because the gang is providing some kind of functional role for the larger community related to self-governance for instance. Delgado (2008) proposes that police initiatives to control crime can generate negative reactions from the community which lead to a countering response. "A countering response may take the form of indigenous movements aiming to nullify the official policing effort and substitute radically different norms" (Delgado 2008: 1204). Delgado cites the "anti-snitching movement" as an example of how communities choose to self-govern rather than involve the police. In this dissertation, the qualitative differences between gang and non-gang areas could be related to the functional role, or niche, that gangs provide to communities in Philadelphia.

The concept of an ecological niche was applied in a slightly different manner by George Hutchinson (1957; 1965). Hutchinson developed the idea that an ecological niche not only applied to the physical, or environmental, conditions that a necessary to support life, but that it also applied to interactions with the environment. He describes the *fundamental niche* as "a state of the environment which would permit the species…to exist indefinitely" (Hutchinson 1957: 416). He also describes a *realized niche* which considers the conditions under which an organism can survive after you consider competition for resources with other organisms. The concept of a *realized niche* considers when two species can coexist in the same environment and is also used to explain why a species is not found in locations where the environment can support them.

In this study, the independent variables were able to predict locations that did or did not contain gang presence but they were not able to predict how many gangs were present, or how much set space was contained in a particular geography. Based on Hutchinson's conceptualization of the ecological niche, there may be two distinct sets of characteristics that identify whether or not any gang can survive in an area—Hutchison's fundamental niche—and whether a gang can thrive and grow—the realized niche. The results of this study support the idea that the independent variables correlate only with one of these characteristics; specifically, the fundamental niche that determines whether or not a gang can survive in an area.

Thrasher referred to 'habitats' where gangs thrive (Thrasher 1927: 9). The work of Jankowski (1991) suggests community support is one such characteristic that is fundamental to gang emergence and persistence in an area. This community support, therefore, is a defining characteristic of a gang's habitat, to use Thrasher's terminology. Without this support, the gang simply cannot persist. Another characteristic of the gang habitat may be the presence of an unsupervised youth population, or a sense in the community that criminal justice agencies do not care about local residents (Anderson 1999). In other words, a number of factors, including demographic, political, cultural, and locational factors might make places more likely to become gang ecologies.

In the context of the current study, there probably are places in Philadelphia that are suitable to support gangs, but gangs struggle to survive there. This is because habitats can support a variety of organisms which all interact with each other. Gangs, for instance, might compete with an after school program or a community center to attract unsupervised youth in a community. This competition constrains the gang. While the environment allows the gang to exist, competition prevents the gang from growing. Without competition from the community center, the gang might be able to grow its territory. It is also possible that police interdiction limits the capacity of a gang to flourish, even when community factors enable some gang presence.

In the context of the current study, the hurdle models allowed me to examine if predictors of gang presence are different from predictors of gang magnitude. An ecological perspective can be used to understand the social environment where (1) gangs are able to emerge and (2) where they are able to grow and sustain themselves. The variables I used to model gangs correlated with the fundamental niche of gang locations as described by Hutchinson, but they did not correlate with the realized niche (i.e. how many gangs are present or how large their territories are).

Implications

Community level dynamics play a critical role in understanding and preventing gang problems. This study contributes to our knowledge about gang measurement at the ecological level which is necessary to establish validity and develop this area of research. Construct validity is critical to understanding what the results of studies really mean and how they can be used to guide policy (Goldkamp 2010). The results of this study can inform future research that builds on our understanding of how gangs link to crime, how community level dynamics foster gang activity, and perhaps eventually, how to reduce and prevent gang problems. The implications of this study for both theory and policy are discussed next.

Theoretical implications

A variety of indicators are used in ecological gang research. These indicators include measures of gang presence as a dichotomous variable or as variables that measure the magnitude of the gang problem using crime rates or gang membership rates. The current research begins to unpack whether these indicators are measures of the same underlying gang construct, or if these indicators are each measuring different dimensions of gang ecologies. Researchers have noticed that the results of gang studies sometimes contradict each other (Katz and Schnebly 2011). This is certainly the case with respect to the applicability of social disorganization theory in explaining gang outcomes.

A number of researchers have examined how community structural variables can be used to explain gang outcomes at the ecological level (Curry and Spergel 1988; Katz and Schnebly 2011; Mares 2010; Papachristos and Kirk 2006; Rosenfeld, Bray, and Egley 1999; Tita, Cohen, and Engberg 2005). These studies use similar independent variables, including economic disadvantage, residential stability, racial heterogeneity and family structure. Many of these studies find that economic disadvantage links in a positive and statistically significant way to gang outcomes. Other variables, including residential stability, the percent vacant land, population density and racial heterogeneity variables, exhibit inconsistent relationships with gang indicators. Some studies conclude that overall, community structure is relevant in explaining gang outcomes while others argue that a different social process is operating in gang neighborhoods.

These inconsistent findings may be due at least in part to the notion that gang ecologies are a multi-dimensional phenomenon. Some aspects of gang ecologies may be explained with the social disorganization theory while other aspects may not be. Katz and Schnebly suggest this may be the case when they compare studies that examine the role of community structural variables on gang presence (measured as a binary variable) with studies that examine gang activity (measured as a continuous variable). They point out that "…the extant literature typically has examined gang presence through binary measures and has not more fully examined whether community structural factors are associated with the magnitude of a community's gang problem " (Katz and Schnebly 2011: 382).

Similar issues have been highlighted in other sub-fields of criminology. Researchers evaluating self-control theory have noted inconsistent findings between studies and have suggested that these inconsistencies may arise from how self-control is measured. "Research has shown the elements [of self-control, such as impulsivity and risk taking] may have distinct relationships with outcomes of interest...[it is not yet known] whether it is self-control that is influencing an outcome, the elements of self-control, or some combination of the two." (Ward, Nobles, and Fox 2015: 598). An assessment of the construct validity of different gang ecology indicators is needed to shed light on this issue and to add depth to our understanding of the community level dynamics that are correlated with gangs at the ecological level.

It is possible that inconsistent findings in gang research are not in fact related to conceptual differences in gang indicators. Instead, these differences may be driven by variations in units of analysis. Researchers that use social disorganization theory to explain gang outcomes aggregate data to a variety of spatial units. These units are as small as census block groups (Rosenfeld, Bray, and Egley 1999) or as large as the 75 community areas in Chicago (Curry and Spergel 1988). Many scholars have recognized the difficulty of choosing an appropriate unit of analysis to study gangs. Few, however, explicitly recognize that different social *processes* may be occurring at different units of analysis. Tita and Radil (2011) prove to be one exception. In their study, they heed the advice of John Hipp who advised "In part, the question of the appropriate aggregation depends on the spatial component of the process being studied" (Hipp 2007: 662). These ideas are consistent with the discontinuity thesis. Under this theoretical frame, inconsistent findings across units of spatial aggregation are expected because the fundamental processes driving relationships between variables become more complex at larger spatial units. In this dissertation, results demonstrate that the relevance of gang ecology indicators change across units of spatial aggregation. Future research can consider whether these different results are generated by different social processes that are operating at different spatial scales or if they are simply the result of measurement error.

The results of this study also have the potential to advance the theoretical implications of gang typologies. After analyzing the structural dimensions of gangs using data from 59 cities, Klein & Maxson (2006) were able to identify five distinct gang typologies. These typologies included traditional, neotraditional, compressed, collective and specialty gangs.[19] These typologies are relevant to theory testing because the group processes that link gangs to crime may be different for each typology. The results of the current study suggest both variables performed equally well in predicting PPD gang set space across three PPD typologies, however, these variables were able to predict the traditional gang category better than they were able to predict the corner drug sales or drug trafficking organization. There were no differences between the two drug-related gang typologies.

This study also has implications for how construct validity is assessed in criminology more generally. A number of methods can be used to assess validity. This study relies on Messick's unified approach to examine various aspects of construct validity simultaneously. This model can be used to test the validity of other constructs in criminological research. Currently, there are no guidelines concerning the degree to which two or more indicators should overlap to come to a conclusion about validity. At what point is a correlation strong enough to suggest two indicators are measuring the same thing? It is possible that this question is related to the process driving the relationship between two variables, in which case there would not be a single criterion that would

[19] In the Klein & Maxson text, see Table 5.1 on page 176 for a description of each of these typologies.

apply across different types of studies. Taylor (2011) has raised the concern that researchers need to better specify which construct they are modeling at the ecological level to avoid indicator/construct ambiguity and stressed that this distinction is important for theory testing. Competing theories cannot be tested unless indicators are clearly distinguished from the constructs they are attempting to identify.

Policy implications

The results of this study have implications for geographically focused gang interventions, how gang set space is identified by police departments, how gang typologies are studied, and how an automated process of gang mapping can be used in tactical, organizational and strategic decision making in policing.

Geographically focused interventions. Recognizing the importance of neighborhoods in combating the gang problem, many agencies have chosen to adopt crime reduction strategies that focus not only on individual gang members but on community level dynamics as well. One such example is focused deterrence. Although focused deterrence has been implemented in different ways, the original program that was implemented in Boston used law enforcement crackdowns to send a deterrence message to gang members. “The Ceasefire crackdowns were not designed to eliminate gangs or stop every aspect of gang activity but to control and deter serious violence” (Braga, Kennedy, Waring, and Morrison-Piehl 2001: 200). The deterrence message that was generated by these crackdowns changed the “gang values, norms, and culture”, as described by Papachristos (2009), by disrupting the cycle of violence that gang members were sustaining. In another community-level gang intervention, the Little Village Gang Project, Spergel also emphasized that the goal of the intervention was not necessarily to eliminate gangs. Instead, the primary goal was to reduce youth violence by changing the gang culture (Spergel 2007: 25).

These holistic, community-level interventions are often geographically focused and target areas that have high concentrations of crime (Boyle, Lanterman, Pascarella, and Cheng 2010; Corsaro, Brunson, and McGarrell 2009; Papachristos, Meares, and Fagan 2007). However, the literature provides no supporting evidence to suggest that crime locations can be used to identify the places that generated the gang problem, or that crime locations are the most appropriate indicators to identify the places to focus interdiction efforts. If a gang develops in a school, and those members then go to a local park to commit crime, the park is not the source of the gang problem and arguably should be reconsidered as the focus of a gang intervention if targeting the initial cause of the gang is the primary goal. Opportunity theories and crime pattern theory are certainly useful for understanding why the gang activities are taking place in the park, but they also highlight the possibility that gang activities are taking place in a variety of locations such as vacant lots, street corners, or residential neighborhoods. It would be more efficient for police to target a single location that is likely the root cause of the gang problem (the school) than to target the myriad of places where opportunities for gang crime emerge. This is important, because "The results of ecological studies of gangs [are needed] to formulate economic and social policies that address the root causes of gangs rather than simply providing triage to the problems caused by gangs." (Tita 1999: 130). This study has produced evidence to suggest that gang problems overlap spatially with set space and gang residence. However, the use of gang-related crime locations was more capable of identifying some types of gangs than others. Specifically, it was better at identifying traditional gangs rather than drug trafficking organizations. As others have already pointed out, focused deterrence programs that use gang-related crime to identify target locations may be inadvertently identifying some types of gang structures and missing others. More specifically, they may be inadvertently excluding gangs with a drug nexus.

Identification of gang set space: The findings of this study have implications for focused deterrence programs not only at the community level but at the individual level as well. In addition to focusing on community level dynamics, many programs make use of custom notification to communicate the goals of the focused deterrence program to individual gang members (Kennedy and Friedrich 2014). The current study found evidence to suggest the locations where gang members live cluster in a statistically significant way. As such, these custom notifications could be delivered more efficiently if these notifications were distributed based on proximity.

The findings of this study also have implications for crime analysis and intelligence units in law enforcement agencies that currently spend a considerable amount time mapping gangs qualitatively. Quantitative indicators that are commonly available to most police department were able, to a certain extent, to replicate the results generated by the focus groups. Although the degree of overlap was relatively small (ranging from .07 to .10), these findings suggest this process could be refined and automated in the future, and if sufficient improvements could be made there may be a saving in time and resources for law enforcement agencies as well as increased objectivity in the process. This last point is increasingly relevant because legal scholars have started to question the accuracy of police mapping efforts when law enforcement officers use areas such as 'gang territory' or 'crime hotspot' as part of their totality of circumstances when justifying a constitutionally-reasonable search of an individual or property (Ferguson 2011).

Admittedly, the home address and gang arrest variables produced extremely small coefficient of areal correspondence values, with mean values around .1 (results reported in Table 30 on page 115). This degree of overlap is not sufficient to completely automate gang set space identification using official police data. Although more time intensive, qualitative methods of gang set space identification are necessary to identify gang set space. Studies that are missing this qualitative piece may have mis-specified gang set space.

The small coefficient of areal correspondence values, however, should not be interpreted as the proportion of the gang set space that was correctly identified using the automated gang mapping process. The denominator in this statistic is the area of both gang set spaces, the one created through automation and the official PPD set space. To calculate the proportion of the gang set space that was covered by the automated process, the denominator would need to be modified to only include the area of the PPD set space. However, this metric would not be useful because it would not consider the size of the approximated set space polygon. Automated set space boundaries that cover the entire city of Philadelphia would achieve 100% overlap with the official PPD set space but obviously would not produce a result that is useful for policing micro-areas. The coefficient of areal correspondence was used to overcome this limitation.

PPD gang typologies: Finally, this study builds on our understanding of gang typologies (Klein and Maxson 2006). The Philadelphia Police Department classifies drug trafficking organizations and corner drug crews as two distinct types of gangs. The results of this study did not reveal the empirical indicators of these gang types are different; however, the indicators were able to predict the locations of traditional gangs better than they were able to predict the locations of the "specialty gangs". Within the specialty gang category, there were no differences between gangs classified as the corner drug sales and drug trafficking organizations. This was surprising, because many would argue that there are conceptual differences between drug gangs with hierarchical structure and those without this structure which lends support to the PPD choice to separate these two categories of gangs in the first palace. Instead, the results of this study indicate that the "specialty gang" category proposed by Klein and Maxson is an appropriate way to categorize gangs; there does not appear to be variation in the social processes linking gangs to crime within their specialty gang classification.

Across the three PPD typologies, the home address and gang arrest variables were least effective in explaining the locations of drug trafficking organizations. It is possible that these types of gangs have a weaker association to place compared to traditional street gangs. Another explanation for these findings may concern how much information is known about these types of gangs. Given the covert nature of drug trafficking organizations, it is possible that less information is available for members of these types of organizations, which makes it difficult to identify these types of gangs using home address or arrest data associated with individual gang members. PPD recognizes the difficulty in gathering information about drug trafficking organizations; as such, they do not use the set space data in the same way for each type of gang. The results of this dissertation highlight these difficulties.

Social justice: In Illinois v Wardlow, the Supreme Court ruled characteristics of locations, such as high crime areas, are relevant contextual considerations informing and officer's decision to conduct a pedestrian stop. In this context, there may be legal ramifications for how gang locations are identified. Legal scholars have questioned how we define high crime areas (Ferguson, 2011), and similar arguments can be made for how we identify gang areas. The low percentage overlap values (7%-10%) that were generated by the home address and gang arrest variables in the analyses that investigated the generalizable aspect of validity raise questions about the appropriateness of using characteristics of places in establishing reasonable suspicion.

More often than not, the approximated set space polygons that were generated using the gang arrest and home address variables were far larger than the PPD defined set space polygons. The coefficient of areal correspondence values were greatly reduced due to this discrepancy alone.

However, while the algorithm used to approximate these locations could be improved by adjusting the parameters of the kernel density tool and significance levels specified in the Gi* analyses, the findings presented here suggest there is a substantial amount of variation in police defined set space locations that *cannot* be explained using the locations where gang members live or gang arrest locations; at least, not how they were modeled in this study. Given these variations, it may be inappropriate to use gang locations, as defined by the police, as grounds for establishing reasonable suspicion of a crime.

Tactical, organizational and strategic decision making: The results of this study also have implications for how tactical, organizational and strategic decisions are made. The influence of crime mapping to inform these types of decisions was explained in depth by Chainey and Ratcliffe (2005). Tactical decisions primarily concern front-line level investigations with a case specific or offender focus. Operational level decisions are made by managers of the front-line officers who are tasked with deploying resources in an effective manner. Finally, the goal of a strategic crime control strategy is to deal with chronic crime problems to accomplish long-term crime reduction goals. Valid indicators of gang locations are vital to each of these levels of decision making.

The use of gang mapping at the tactical level not only needs to be valid, but it needs to be precise as well. If line officers are investigating one particular gang, deploying resources to an area that is even just a few blocks away from the true locations the gang is operating would not be beneficial to the investigation. Based on the results of this study, the use of an automated gang mapping process may not be helpful for tactical operations because the mapping process relies on historical crime patterns. In the case of gang investigations, the people involved in an organization can change on a day-to-day basis; people leave the gang, new people join, some get arrested and are incarcerated for an extended period of time, others are killed. While an automated gang mapping process may identify the general area the gang is operating in, this information would be more useful for operational decision making than it would be for tactical initiatives.

The results of this study suggest an automated process of gang mapping may be well suited for operational decision making because it can be used to identify the general locations gangs are operating in. These locations can be used to inform deployment decisions. While the degree of overlap between the automated gang mapping process and the police-identified gang set space was small, it is worth noting that 80% of the time, the approximated gang set space did overlap with the PPD set space to some extent. While the overlap wasn't perfect, the algorithm was able to identify the general locations where gangs exist 80% of the time.

Furthermore, at the operational level timely information is extremely important. When it comes to mapping gangs, operational decisions need to be made with up to date information because gang locations may change over time. Qualitative methods are too time consuming to provide this up to date information to police commanders. Mapping gang locations using an automated process would be beneficial to informing operational level decisions because new maps can be generated with up to date information in a timely manner. Related to this, it would be helpful if police analysts could figure out why things are changing geographically. Predicting changes in gang ecologies is another area of research that would be beneficial to operational decision making.

Lastly, valid gang maps are crucial to the success of crime reduction programs at the strategic level. When crime analysts are able to identify the communities suffering from chronic gang problems, researchers can analyze the root causes of gang emergence and persistence so programs can be designed to change these communities. The results of this study suggest that various gang indicators are able to distinguish between places that have gangs and places that do not which may be beneficial to strategic level decision making.

Limitations

A number of limitations deserve to be noted. One limitation concerns the fact that this study relies only on official police data. Other limitations include measurement error, definitional issues surrounding gangs, temporal scaling and quantifying gang magnitude. Each of these limitations are discussed next.

Official police data

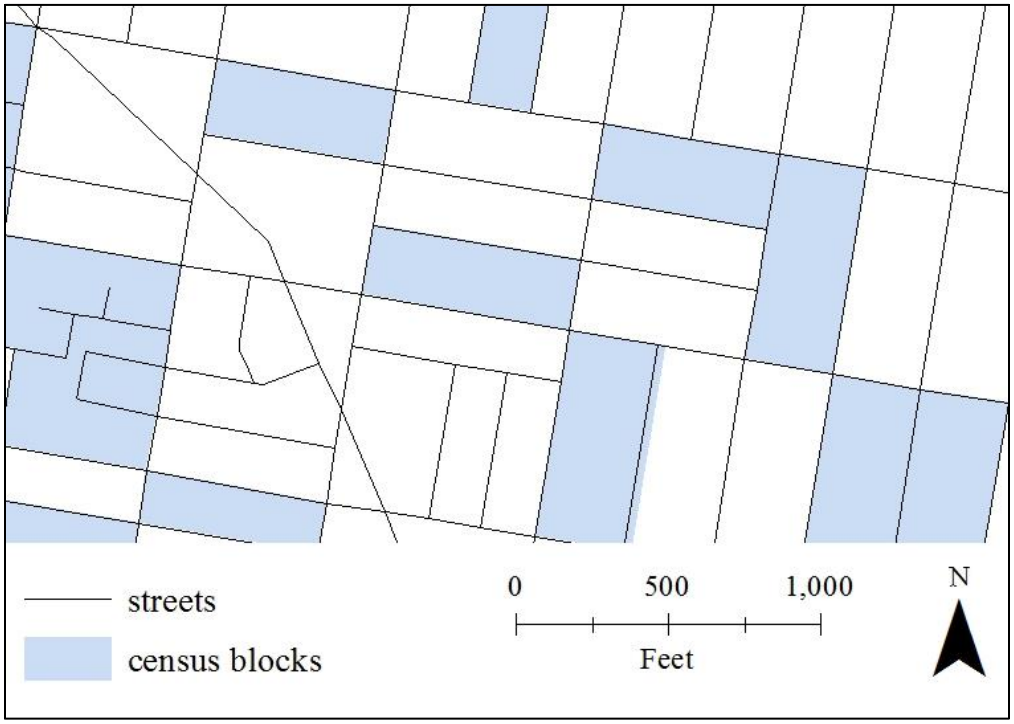
One main limitation of this study is that it relies only on official police data to examine gang ecologies. The perceptions of gang members and local community residents are not reflected in any of the indicators I reviewed. This is problematic for several reasons. First, although the PPD gang set space data were extensively vetted by CIU and were cross-validated by criminal justice practitioners from a variety of agencies, it is possible that the set spacec dataset is incomplete. Some gangs fly under the radar of law enforcement and successfully avoid interaction with the criminal justice system. These gangs, if they exist in Philadelphia, would not be represented in official police data. Community members may be aware of gangs that the police are not privy to. Second, the set space locations used in this study were identified by the police and other criminal justice practitioners. If a separate set of focus group meetings were held for community members, or even gang members themselves, a different set of gang set space locations may have emerged. Future research could consider using community members and/or gang members as a source of information for identifying gang ecologies.

The quality of police record keeping poses another limitation to this study. With the exception of the PPD set space locations, the data used in this study were recorded at the address level. Anytime address level data are used, geocoding errors become a potential issue. When addresses recorded in police reports include street names that are misspelled, are missing relevant address information, or simply do not exist, these locations will not be geocoded and therefore will not be included in subsequent analyses. The home address data used in this study achieved a geocoding hit rate of 96% while the gang arrest and gun-related crime variables both achieved a geocoding hit rate of 99%. These geocoding hit rates are extremely high, so geocoding issues are not likely to influence the results in a substantial way. However, it is worth noting that there is some error present which has the potential of influencing the results.

Error associated with home address data

The Philadelphia Police Department supplied home address data for each gang member in their gang database. These home address locations, however, were aggregated to census blocks to protect the identity of the gang members. Aggregating the home address locations to census blocks obscured the location of the home address by up to four streets. This is because census blocks often include four street segments, as seen in Figure 14, which depicts census blocks where gang members live.

Figure 14: Philadelphia census blocks



Gang members belonged to 2,670 census blocks across Philadelphia. Census blocks are extremely small units of analysis. The 2,670 census blocks where gang members live in Philadelphia are on average 256,613 square feet in size (.009 square miles). These descriptive statistics can be found in Table 42.

Table 42: Size of Philadelphia census block groups, descriptive statistics

N	Min	Max	Mean	SD	Median
2,670	18,037 sq ft	14,813,229 sq. ft	256,613 sq. ft	575,878 sq. ft	171,198 sq.ft

To determine how much error was introduced to the analysis with the use of census block centroids, the distance from each centroid to the furthest point of the census block was calculated. The maximum distances from the census block centroids to the farthest point of the census block ranged from a minimum of 40 feet to a maximum of 1,130 feet. This means that at worst, the most error that could be introduced to the analysis is 1,130 feet. Across all 2,670 census blocks, the average maximum distance was just 129 feet. These descriptive statistics, along with statistics on the minimum and mean distance, are reported in Table 43.

Table 43: Minimum, maximum and mean distances from census block centroid to farthest point of census block

	N	Min	Max	Mean	SD	Median
Minimum distance	2,670	3.13	479.49	85.56	35.04	89.91
Maximum distance	2,670	40.31	1130.38	129.16	81.08	115.32
Mean distance	2,670	32.95	713.95	110.93	45.32	105.72

Based on these figures, the use of census block centroids to model the home address location of gang members introduced a modest amount of error into the analysis. It is unlikely that this error influenced the analysis in a statistically significant way.

Definitional issues

Defining gang members: Definitional issues in gang research make it difficult to define who is a gang member (Ball and Curry 1995). After reviewing state and federal legislation governing gangs, Barrows & Huff (2009) found that there is a great deal of variability in gang membership criteria across states. Because of these variations, a person could be classified as a gang member in one state, but that same person may not be classified as a gang member in another state. This study relies on a list of gang members that is maintained by the intelligence unit at the Philadelphia Police Department. Previous research is mixed regarding the ability of law enforcement agencies to document gang members (Katz, Webb, and Schaefer 2000; Zatz 1987). There are a number of reasons why the gang list supplied by the intelligence unit may be incomplete which may create a content domain sampling issue in two of the gang ecology indicators.

First, it is possible that some gangs operating in the city have flown under the radar of law enforcement. Klein (1971: 19) recalls bringing a list of 100 gang members to the local police department. To his surprise, the police's documented list of gang members for that particular gang had just 20 people identified. It is possible that a law enforcement intelligence unit may not be aware of a criminal gang operating on the street simply because the gang has successfully avoided law enforcement interaction.

Second, some researchers have noted that gang membership is transitory in nature (Decker and Curry 2002; Sullivan 2005). Although the list maintained by the intelligence unit is updated and validated every three months, the transitory nature of gang membership may hinder their ability to maintain this list.

Third, the strict standard used by PPD to identify gang members—requiring a person to have at least one prior arrest to be classified as a gang member—will underestimate the number of gang members in the city. This strict definition may create construct irrelevant difficulty, making it irrelevantly difficult to classify a person as a gang member.

Lastly, the criminal intelligence unit may be unaware of gang members operating in Philadelphia because other officers in the department are unwilling to share their knowledge with the intelligence unit. Consider a scenario where a patrol officer observes what they suspect to be a drug crew operating on a street corner. If they report this to the intelligence unit, the narcotics team may set up surveillance on this group and ultimately make an arrest once they have enough evidence. If the officer keeps the intelligence to his/her self, that officer can potentially make the arrest independently and get credit for the arrest and the ability to get overtime in court (Ratcliffe 2008).

To improve their ability to maintain an accurate list of gang members, the intelligence unit updates their list of gang members on a quarterly basis, thereby reducing the likelihood that transitory nature of gang membership will affect the accuracy of this list. Furthermore, efforts to facilitate a culture of information sharing within the police department have been made to encourage district captains to share gang information with the intelligence unit. These efforts, however, do not guarantee compliance. While the list of gang members used in this study has been extensively and continuously validated by the intelligence unit, the aforementioned limitations should be noted.

Defining gang-related crime. Another limitation of this study concerns the method used to classify crime incidents as gang-related incidents. There are a number of ways to classify a crime incident as a gang incident (Maxson 1999; Papachristos and Kirk 2006). The Los Angeles definition is a member-based definition. This definition includes all incidents where the victim or suspect is a known gang member. The Chicago definition however, requires a gang motive.

It is difficult to discern the extent to which the use of a member-based definition instead of a motive-based definition is a limitation in this study. After comparing homicide cases using both the member-based and motive-based definitions, Maxson and Klein (1990) found that the characteristics of both types of incidents are comparable. After analyzing 231 gang-related homicides—137 of which were determined to be gang-motivated—they conclude: "The characters of motive-defined and member-defined gang homicides are quite similar…When contrasting gang with non-gang homicide incidents, it does not matter which definitional approach is used for purposes of describing the settings and participants of each" (Maxson and Klein 1990: 90-91). However, as the title of their article highlights, using a gang-related definition of crime produced nearly twice as many gang incidents as the gang-motivated definition. Rosenfeld and colleagues (1999) also find that the two definitions generate empirically different results. The data supplied by the Philadelphia Police Department in this study cannot be disaggregated from "gang-related" to “gang-motivated”. The inconsistent results of previous research make it difficult to determine how significant this limitation is.

Temporal scaling

Another limitation of this study concerns temporal scaling. In this study, I used a list of gang members that was validated over a 17 month period from June 2014 through October 2015. To measure gang-related crime I relied on arrest records that list these gang members as a suspect using records from January 2012 through December 2015. Creating the gang-related variable this way assumes that the gang members that were identified as late as October 2015 were active in the gang throughout the 2012-2015 time period. If a person identified in the audit was not a gang member in 2012, then any crime incidents they were involved in at that time should *not* be coded as a gang-related incident. One way to address this issue is to use fewer years of official crime data, but doing so risks excluding crime incidents by gang members who *were* active in the gang three years prior. Regardless of what time period is used, there will be some degree of error associated with this indicator. To overcome this limitation, the gang database could be updated with data that reflect when individual gang members joined the gang and when they left the gang.

A related limitation is this study is cross sectional. Connections between gang ecology indicators might look different if changes over an extended period of time were analyzed.

Quantifying gang magnitude

Finally, quantifying the magnitude of gang ecologies is difficult to do. This study relied on a number of continuous variables to measure gang magnitude. For each geographic unit, these continuous variables included proportion values that correspond with the size of the gang set space, the number of gangs operating there, the number of gang members living there, the number of gang arrests, and the number of gun-involved crime incidents. One important limitation to note is that higher values on any of these variables do not necessarily correspond to more harmful gangs or a more severe gang problem. Some of the worst gangs in Philadelphia—gangs that cause the most social harm to communities—may have small territorial claims (small proportion values), have only a handful of documented gang members (low levels of residency), or may not have lengthy arrest records (low levels of gang-related crime). In these instances, the magnitude of the gang problem cannot be measured with any of the variables in this study. An in-depth qualitative study on the quality of life in Philadelphia neighborhoods may uncover the magnitude of the gang problem in these communities.

Summary

Gangs pose a serious problem in 21st century policing. After spending forty years studying gangs, Malcolm Klein admits that we still have a lot to learn about gangs: "Too little attention has been paid to the communities in which gangs appear. Observing and understanding neighborhoods is far more complex than studying their gangs; yet it is communities that spawn gangs and must inevitably be the proximal focus for controlling them" (Klein 2007: xiv).

To better understand the relationship between gangs and crime, researchers and practitioners have developed a number of indirect measures to document where gangs are located. These methods have not been compared against each other, until now. This dissertation systematically examined various aspects of construct validity as they relate to the spatial distribution of gang ecologies.

A number of analyses were used to approach this inquiry from different perspectives. Overall, this study found that gang set space locations can be identified quantitatively using official police data. However, the regression analysis highlighted the fact that home address and gang arrest data are only able to explain some, not all, of the spatial distribution of police identified gang set space. Other variables are at play which were not included in this study. The algorithm output revealed gang set space identification can be automated to a certain extent. However, the algorithm was able to some types of gangs better than others. Specifically, it was able to identify street gangs better than it was able to identify corner drug crews or drug trafficking organizations. If practitioners have the desire to automate how gang set space is identified to identify gang locations in real-time, it is important to be aware of this finding; some types of gangs are more difficult to identify compared to others.

REFERENCES CITED

Anderson, Elijah. 1999. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. New York: W.W. Norton & Company.

Ball, Richard A. and G. David Curry. 1995. "The Logic of Definition in Criminology: Purposes and Methods for Defining "Gangs"." *Criminology* 33:225-245.

Barrows, Julie and C. Ronald Huff. 2009. "Gangs and public policy: Constructing and deconstructing gang databases." *Criminology and Public Policy* 8:675-703.

Bastiani, Fernanda De, Robert A. Rigby, Dimitrios M. Stasinopoulous, Audrey H.M.A. Cysneiros, and Miguel A. Uribe-Opazo. 2016. "Gaussian Markov random field spatial models in GAMLSS." *Journal of Applied Statistics*.

Baum, Christopher F. 2008. "Stata tip 63: Modeling Proportions." *The Stata Journal* 8:299-303.

Blasko, Brandy, Caterina Roman, and Ralph Taylor. 2015. "Local gangs and residents' perceptions of unsupervised teen groups: Implications for the incivilities thesis and neighborhood effects." *Journal of Criminal Justice* 43:20-28.

Block, Carolyn and Richard Block. 1993. "Street gang crime in Chicago." U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, Washington, D.C.

Block, Richard. 2000. "Gang Activity and Overall Levels of Crime: A New Mapping Tool for Defining Areas of Gang Activity Using Police Records." *Journal of Quantitative Criminology* 16:369-383.

Boyle, Douglas J., Jennifer L. Lanterman, Joseph E. Pascarella, and Chia-Cherng Cheng. 2010. "The Impact of Newark's Operation Ceasefire on Trauma Center Gunshot Wound Admissions." *Justice Research and Policy* 12:105-123.

Braga, Anthony A., David M. Hureau, and Andrew V. Papachristos. 2014. "Deterring Gang-Involved Gun Violence: Measuring the Impact of Boston's Operation Ceasefire on Street Gang Behavior." *Journal of Quantitative Criminology* 30:113-139.

Braga, Anthony A., David M. Kennedy, Elin J. Waring, and Anne Morrison-Piehl. 2001. "Problem-Oriented Policing, Deterrence, and Youth Violence: An Evaluation of Boston's Operation Ceasefire." *Journal of Research in Crime and Delinquency* 38:195-225.

Brantingham, Jeffrey, George Tita, Martin Short, and Shannon Reid. 2012. "The Ecology of Gang Territorial Boundaries." *Criminology* 50:851-885.

Brantingham, Patricia L. and Paul J. Brantingham. 1993. "Nodes, Paths and Edges: Considerations on the complexity of crime and the physical environment." *Journal of Environmental Psychology* 13:3-28.

Brattin, Sara R., Karl G. Hill, Robert D. Abbott, Richard F. Catalano, and J. David Hawkins. 1998. "The Contribution of Gang Membership to Delinquency Beyond Delinquent Friends." *Criminology* 36:93-115.

Bursik, Robert and Harold Grasmick. 1993. *Neighborhoods and Crime: The Dimensions of Effective Community Control*. New York: Lexington Books.

Butts, Jeffrey A., Caterina Gouvis Roman, Lindsay Bostwick, and Jeremy R. Porter. 2015. "Cure Violence: A Public Health Model to Reduce Gun Violence." *Annual Review of Public Health* 36:39-53.

Bynum, Timothy S. and Sean P. Varano. 2003. "The Anti-gang Initiative in Detroit: An Aggressive Enforcement Approach to Gangs." in *Policing Gangs and Youth Violence*, edited by S. Decker. Belmont, CA: Wadsworth Thomson.

Campbell, Donald T. and Donald W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56:81-105.

Cartwright, Desmond S. and Kenneth I. Howard. 1966. "Multivariate Analysis of Gang Delinquency: Ecologic Influences." *Multivariate Behavioral Research* 1:321-371.

Chainey, Spencer and Jerry Ratcliffe. 2005. *GIS and Crime Mapping*. West Sussex, England: Wiley.

Cohen, Jaqueline and George Tita. 1999. "Diffusion in Homicide: Exploring a General Method for Detecting Spatial Diffusion Process." *Journal of Quantitative Criminology* 15:451-493.

Cook, Douglas O., Robert Kieschnick, and B. D. McCullough. 2008. "Regression analysis of proportions in finance with self selection." *Journal of Empirical Finance* 15:860-867.

Cook, R. Dennis and Sanford Weisberg. 1982. *Residuals and influence in regression*. London ; New York: Chapman and Hall.

Corsaro, Nicholas, Rod K. Brunson, and Edmund F. McGarrell. 2009. "Problem-Oriented Policing and Open-Air Drug Markets: Examining the Rockford Pulling Levers Deterrence Strategy." *Crime & Delinquency* 59:1085-1107.

Cox, D. R. and E. J. Snell. 1989. *The Analysis of Binary Data*. London: Chapman & Hall/CRC.

Curry, G. David. and Irving A. Spergel. 1988. "Gang Homicide, Delinquency, and Community." *Criminology* 26:381-405.

Decker, Scott H. and G. David Curry. 2002. "Gangs, gang homicides and gang loyalty: Organized crimes or disorganized criminnals." *Journal of Criminal Justice* 30:343-352.

Decker, Scott H. and Kimberly Kempf-Leonard. 1991. "Constructing Gangs: The Social Definition of Youth Activities." *Criminal Justice Policy Review* 5:271-291.

Decker, Scott H., Chris Melde, and David C. Pyrooz. 2013. "What do we know about gangs and gang members and where do we go from here?" *Justice Quarterly* 30:369-402.

Decker, Scott H. and David C. Pyrooz. 2010. "On the Validity and Reliability of Gang Homicide: A Comparison of Disparate Sources." *Homicide Studies* 14:359-376.

Decker, Scott H., David C. Pyrooz, Gary Sweeten, and Richard K Jr Moule. 2014. "Validating Self-Nomination in Gang Research: Assessing Differences in Gang Embeddedness Across Non-Current, and Former Gang Members." *Journal of Quantitative Criminology* 30:577-598.

Decker, Scott H., Frank van Gemert, and David C. Pyrooz. 2009. "Gangs, Migration, and Crime: The Changing Landscape in Europe and the USA." *International Migration & Integration* 10:393-408.

Delgado, Richard. 2008. "Law Enforcement in Subordinated Communities: Innovation and Response." *Michigan Law Review* 106:1193-1210

Dunn, Peter K and Gordon K Smyth. 1996. "Randomized quantile residuals." *Journal of Computational and Graphical Statistics* 5:236-244.

Egley, Arlen, James C. Howell, and Meena Harris. 2014. "Highlights of the 2012 National Youth Gang Survey." U.S. Department of Justice, Washington D.C.

Esbensen, Finn-Aage and David Huizinga. 1993. "Gangs, Drugs and Delinquency in a Survey of Urban Youth." *Criminology* 31:565-589.

Esbensen, Finn-Aage, L. Thomas Jr. Winfree, Ni He, and Terrance J. Taylor. 2001. "Youth Gangs and Definitional Issues; When is a Gang a Gang, and Why Does it Matter?" *Crime & Delinquency* 47:105-130.

Fader, Jamie J. 2013. *Falling Back: Incarceration and Transitions to Adulthood among Urban Youth*. New Brunswick, NJ: Rutgers University Press.

Felson, Marcus. 2006. *Crime and Nature*. Thousand Oaks, CA: Sage.

Ferguson, Andrew Guthrie. 2011. "Crime Mapping and the Fourth Amendment: Redrawing "High-Crime Areas"." *Hastings Law Journal* 63:101-154.

Ferrari, Silvia L.P. and Francisco Cribari-Neto. 2004. "Beta Regression for Modeling Rates and Proportions." *Journal of Applied Statistics* 7:799-815.

Fotheringham, A. Stewart and D. W. S. Wong. 1991. "The Modifiable Areal Unit Problem in Multivariate Analysis." *Environment and Planning A* 23:1025-1044.

Garcia, R. Marie, Ralph B. Taylor, and Brian A. Lawton. 2007. "Impacts of Violent Crime and Neighborhood Structure on Trusting Your Neighbors." *Justice Quarterly* 24:679-704.

Garson, G. David and Robert S. Biggs. 1992. "Analytic Mapping and Geographic Databases." in *Quantitative Applications in the Social Sciences*, vol. 07-087. Newbury Park: Sage.

Getis, Arthur and J. K. Ord. 1992. "The Analysis of Spatial Association by use of Distance Statistics." *Geographical Analysis* 24:189-206.

Goldkamp, John S. 2010. "Construct Validity: The Importance of Understanding the nature of the Intervention Under Study." Pp. 455-479 in *Handbook of Quantitative Criminology*, edited by A. R. Piquero and D. Weisburd. New York: Springer.

Gordon, Robert A. 1968. "Issues in Multiple Regression." *The American Journal of Sociology* 73:592-616.

Gottfredson, Don M. and Travis Hirschi. 1990. *A General Theory of Crime*. Stanford: Stanford Univerisity Press.

Grasmick, Harold G., Charles R. Tittle, Robert J. Bursik, and Bruce J. Arneklev. 1993. "Testing the Core Empirical Implications of Gottfredson and Hirschi's General Theory of Crime." *Journal of Research in Crime and Delinquency* 30:5-29.

Groff, Elizabeth R. 2011. "Exploring 'near': Characterizing the spatial extent of drinking place influence on crime." *Australian & New Zealand Journal of Criminology* 42:156-179.

Hall, Gina Penly, Terence P. Thornberry, and Alan J. Lizotte. 2006. "The Gang Facilitation Effect and Neighborhood Risk: Do Gangs Have a Stronger Influence on Delinquency in Disadvantaged Areas?" Pp. 47-61 in *Studying Youth Gangs, Violence Prevention and Policy Series*, edited by J. F. J. Short and L. A. Hughes. New York: AltaMira Press.

Hannan, Michael T. 1991. *Aggregation and Disaggregation in the Social Sciences*. Lexington, MA: Lexington Books.

Hawley, Amos H. 1950. *Human Ecology: A Theory of Community Structure*. New York: Ronald Press.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153-161.

Hipp, John R. 2007. "Block, Tract, and Levels of Aggregation: Neighborhood Structure and Crime and Disorder as a Case in Point." *American Sociological Review* 72:659-680.

Horowitz, Ruth. 1983. *Honor and the American Dream: Culture and Identity in a Chicano Community*: Rutgers University Press.

—. 1987. "Community Tolerance of Gang Violence." *Social Problems* 34:437-450.

Huddleston, Samuel H., Jon Fox, and Donald E. Brown. 2012. "Mapping Gang Spheres of Influence." *Crime Mapping: A Journal of Research and Practice* 4:39-67.

Huebner, Beth M., Kimberly Marin, Rickard K. Jr. Moule, David C. Pyrooz, and Scott H. Decker. 2014. "Dangerous Places: Gang Members and Neighborhood Levels of Gun Assault." *Justice Quarterly*:1-27.

Hughes, Lorine A. 2005. "Studying Youth Gangs: Alternative Methods and Conclusions." *Journal of Contemporary Criminal Justice* 21:98-119.

Hunter, Albert. 1979. "The Urban Neighborhood: Its Analytical and Social Contexts." *Urban Affairs Quarterly* 14:267-288.

Hutchinson, G. Evelyn. 1957. "Concluding Remarks." *Cold Spring Harbor Symposia on Quantitative Biology* 22:415-427.

—. 1965. *The ecological theater and the evolutionary play*. New Haven,: Yale University Press.

Jankowski, Martin. 1991. *Islands in the Street: Gangs and American Urban Society*. Berkeley and Los Angeles: University of California Press.

Katz, Charles and Stephen Schnebly. 2011. "Neighborhood Variation in Gang Member Concentrations." *Crime & Delinquency* 57:377-407.

Katz, Charles, Vincent Webb, and David Schaefer. 2000. "The Validity of Police Gang Intelligence Lists: Examining Differences in Delinquency Between Documented Gang Members and Nondocumented Delinquent Youth." *Police Quarterly* 3:413-437.

Kennedy, David M., Anthony A. Braga, and Anne M. Piehl. 1997. "The (Un)Known Universe: Mapping Gangs and Gang Violence in Boston." Pp. 219-262 in *Crime Mapping and Crime Prevention*, vol. Crime Prevention Studies, edited by D. Weisburd and T. McEwen. New York: Criminal Justice Press.

Kennedy, David M. and Michael A. Friedrich. 2014. "Custom Notifications: Individualized Communication in the Group Violence Intervention." Office of Community Oriented Policing Services, Washington, DC.

Klein, Malcolm W. 1971. *Street Gangs and Street Workers*. Englewood Cliffs, N.J.: Prentice-Hall.

—. 1995. *The American Street Gang: Its Nature, Prevalence and Control*. New York: Oxford University Press.

—. 2004. *Gang Cop: The Words and Ways of Officer Paco Domingo*. Walnut Creek, CA: AltaMira Press.

—. 2007. *Chasing After Street Gangs: A Forty-Year Journey*. Upper Saddle River, NJ: Pearson Prentice Hall.

Klein, Malcolm W. and Cheryl L. Maxson. 2006. *Street Gang Patterns and Policies*. Oxford: Oxford University Press.

Lane, Jodi and James W. Meeker. 2000. "Subcultural Diversity and the Fear of Crime and Gangs." *Crime & Delinquency* 46:497-521.

Lawton, Brian A., Ralph B. Taylor, and Anthony J. Luongo. 2005. "Police Officers on drug corners in Philadelphia, drug crime, and violent crime: Intended, diffusion and displacement impacts." *Justice Quarterly* 22:427-451.

Ley, David and Roman Cybriwsky. 1974. "Urban Graffiti as Territorial Markers." *ANNALS of the Association of American Geographers* 64:491-505.

Loftin, Colin. 1986. "Assaultive violence as a contagious social proces." *Bulletin of the New York Academy of Medicine* 62:550-555.

Logan, John R. and Harvey L. Molotch. 1987. *Urban Fortunes: The Political Economy of Place*. Berkeley: University of California Press.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.

Mares, Dennis. 2010. "Social Disorganization and Gang Homicides in Chicago: A Neighborhood Level Comparison of Disaggregated Homicides." *Youth Violence and Juvenile Justice* 8:38-57.

Maxfield, Michael G. and Earl R. Babbie. 2011. *Research Methods for Criminal Justice and Criminology*. Belmont, CA: Wadsworth.

Maxson, Cheryl L. 1999. "Gang Homicide: A Review and Extension of the Literature." Pp. 239-254 in *Homide: A Sourcebook of Social Research*, edited by M. D. Smith and M. A. Zahn. Thousand Oaks, CA: Sage.

Maxson, Cheryl L., Margaret A. Gordon, and Malcolm W. Klein. 1985. "Differences Between Gang and Nongang Homicides." *Criminology* 23:209-222.

Maxson, Cheryl L. and Malcolm W. Klein. 1990. "Street gang violence: Twice as great, or half as great?" Pp. 71-100 in *Gangs in America*. Newbury Park, CA: Sage.

Melde, Chris and Callie Marie Rennison. 2010. "Intimidation and Street Gangs: Understanding the Response of Victims and Bystanders to Perceived Gang Violence." *Justice Quarterly* 27:619-666.

Menard, Scott. 2002. *Applied Logistic Regression Analysis*. Thousand Oaks, CA: Sage Publications, Inc.

Mennis, Jeremy, Suzanne Lashner Dayanim, and Heidi Grunwald. 2013. "Neighborhood Collective Efficacy and Dimensions of Diversity: A Multilevel Analysis." *Environment and Planning A* 45:2176-2193.

Mennis, Jeremy, Philip W. Harris, Zoran Obradovic, Alan J. Izenman, Heidi E. Grunwald, and Brian Lockwood. 2011. "The Effect of Neighborhood Characteristics and Spatial Spillover on Urban Juvenile Delinquency and Recidivism." *The Professional Geographer* 63:174-192.

Messick, Samuel. 1995. "Validity of Psychological Assessment: Validation of Inferences From Persons' Responses and Performances as Scientific Inquiry Into Score Meaning." *American Psychologist* 50:741-749.

Molotch, Harvey, William Freudenburg, and Krista E. Paulsen. 2000. "History Repeats Itself, But How? City Character, Urban Tradition, and the Accomplishment of Place." *American Sociological Review* 65:791-823

Moore, Joan, Diego Vigil, and Robert Garcia. 1983. "Residence and Territoriality in Chicano Gangs." *Social Problems* 31:182-194.

National Research Council, . 2004. *Fairness and Effectiveness in Policing: The Evidence*. Washington, D.C.: The National Academies Press.

Nutter, Michael A. 2013. "Philadelphia's Strategic Plan to Prevent Youth Violence." City of Philadelphia.

Openshaw, Stan. 1984. "The Modifiable Areal Unit Problem." *Concepts and Techniques in Modern Geography* 38:41.

Ospina, Raydonal and Silvia L.P. Ferrari. 2012. "A general class of zero-or-one inflated beta regression models." *Computational Statistics and Data Analysis* 56:1609-1623.

Ouimet, Marc. 2000. "Aggregation bias in ecological research: How social disorganization and criminal opportunities shape the spatial distribution of juvenile delinquency in Montreal." *Canadian Journal of Criminology* 42:135-156.

Papachristos, Andrew. 2006. "Social network analysis and gang research: Theory and methods." Pp. 99-116 in *Studying Youth Gangs:*, edited by J. F. J. Short and L. A. Hughes. Oxford: Altamira Press.

—. 2009. "Murder by Structure: Dominance Relations and the Social Structure of Gang Homicide." *American Journal of Sociology* 115:74-128.

Papachristos, Andrew, David Hureau, and Anthony Braga. 2013. "The Corner and the Crew: The Influence of Geography and Social Networks on Gang Violence." *American Sociological Review* 78:417-447.

Papachristos, Andrew and David Kirk. 2006. "Neighborhood Effects on Street Gang Behavior."

Papachristos, Andrew, Tracey L. Meares, and Jeffrey Fagan. 2007. "Attention Felons: Evaluating Project Safe Neighborhoods in Chicago." *Journal of Empirical Legal Studies* 4:223-272.

Pattillo, Mary E. 1998. "Sweet Mothers and Gangbangers: Managing Crime in a Black Middle-Class Neighborhood." *Social Forces* 76:747-774.

Porteous, John Douglas. 1977. *Environment and Behavior: Planning and Everyday Urban Life*   Reading, MA: Addison_Wesley.

Pyrooz, David C. 2012. "Structural Covariates of Gang Homicide in Large U.S. Cities." *Journal of Research in Crime and Delinquency* 49:489-519.

Pyrooz, David C., Andrew Fox, and Scott H. Decker. 2010. "Racial and Ethnic Heterogeneity, Economic Disadvantage, and Gangs: A Macro-Level Study of Gang Membership in Urban America." *Justice Quarterly* 27:867-892.

Radil, Steven, Colin Flint, and George Tita. 2010. "Spatializing Social Networks: Using Social Network Analysis to Investigate Geographies of Gang Rivalry, Territoriality and Violence in Los Angeles." *Annals of the Association of American Geographers* 100:307-326.

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111-163.

Ratcliffe, Jerry H. 2008. *Intelligence-Led Policing*. Cullompton: Willan Publishing.

Rengert, George F. 1996. *The Geography of Illegal Drugs*. Boulder, CO: Westview Press.

Roman, Caterina G. and Aaron Chalfin. 2008. "Fear of Walking Outdoors: A multi-level ecologic analysis of crime and disorder." *American Journal of Preventive Medicine* 34:306-312.

Rosenfeld, Richard. 2011. "The Big Picture: 2010 Presidential Address to the American Society of Criminology." *Criminology* 49:1-26.

Rosenfeld, Richard, Timothy Bray, and Arlen Egley. 1999. "Facilitating Violence: A Comparison of Gang-Motivated, Gang-Affiliated, and Nongang Youth Homicides." *Journal of Quantitative Criminology* 15:495-516.

Sampson, Robert. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago: University of Chicago Press.

Shadish, William R., Thomas D. Cook, and Donald Thomas Campbell. 2002. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Shaw, Clifford Robe and Henry D. McKay. 1942 [1969]. *Juvenile delinquency and urban areas. A study of rates of delinquency in relation to differential characteristics of local communities in American cities*. Chicago: University of Chicago Press.

Short, James F. Jr. 1985. "The Level of Explanation Problem in Criminology." Pp. 51-72 in *Theoretical Methods in Criminology*, edited by R. F. Meier. Beverly Hills, CA: Sage.

—. 1998. "The Level of Explanation Problem Revised-The American Society of Criminology 1997 Presidential Address." *Criminology* 36:3-36.

Silverman, B. W. 1986. "Density Estimation for Statistics and Data Analysis." New York: Chapman and Hall.

Smithson, Michael and Jay Verkuilen. 2006. "A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-Distributed Dependent Variables." *Psychological Methods* 11:54-71.

Spergel, Irving A. 1995. *The Youth Gang Problem: A Community Approach*. New York, NY: Oxford University Press.

—. 2007. *Reducing Youth Gang Violence: The Little Village Gang Project in Chicago*, Edited by M. S. Fleisher. Lanham, MD: Rowman & Littlefield Publishers.

Sullivan, Mercer L. 2005. "Maybe We Shouldn't Study "Gangs" Does Reification Obscure Youth Violence?" *Journal of Contemporary Criminal Justice* 21:170-190.

Suttles, Gerald. 1968. *Social Order of the Slum*. London: University of Chicago Press, Ltd.

Sweet, Stephen A. and Karen A. Grace-Martin. 2010. *Data Analysis with SPSS: A first course in applied statistics*: Pearson.

Taniguchi, Travis, Jerry Ratcliffe, and Ralph B. Taylor. 2011. "Gang Set Space, Drug Markets, and Crime around Drug Corners in Camden." *Journal of Research in Crime and Delinquency* 48:327-363.

Taylor, Ralph B. 1994. *Research Methods in Criminal Justice*. New York, NY: McGraw-Hill.

—. 2001. *Breaking Away From Broken Windows : Baltimore Neighborhoods and the Nationwide Fight Against Crime, Grime, Fear, and Decline*. Boulder, Colo. ; Oxford: Westview Press.

—. 2010. "Communities, Crime, and Reactions to Crime Multilevel Models: Accomplishments and Meta-Challenges." *Journal of Quantitative Criminology* 26:455-466.

—. 2011. "Communities and Crime Theories: Construct Validation Deficits, a Paucity of Comparisons, and a boudon-Coleman Metamodel Resolution." Pp. 71-100 in *Measuring Crime and Criminality: Advances in Criminological Theory*, vol. 17, edited by J. MacDonald. New Brunswick: Transaction Publishers.

—. 2015. *Community Criminology: Fundamentals of Spatial and Temporal Scaling, Ecological Indicators, and Selectivity Bias*. New York: New York University Press.

Thornberry, Terence P., Marvin D. Krohn, Alan J. Lizotte, and Deborah Chard-Wierschem. 1993. "The Role of Juvenile Gangs in Facilitating Delinquent Behavior." *Journal of Research in Crime and Delinquency* 30:55-87.

Thrasher, Frederic M. 1927. *The Gang: a study of 1,313 gangs in Chicago*. Chicago: University of Chicago Press.

—. 1928. "How to Study the Boys' Gang in the Open." *American Sociological Association* 1:244-254.

Tita, George. 1999. "Mapping the Set Space of Urban Street Gangs." Pp. 125-131 in *Atlas of Crime: Mapping the Criminal Landscape*, edited by L. S. Turnbull, E. H. Hendrix, and B. D. Dent. Phoenix, AZ: Oryx Press.

Tita, George, Jacqueline Cohen, and John Engberg. 2005. "An Ecological Study of the Location of Gang "Set Space"." *Social Problems* 52:272-299.

Tita, George and Robert Greenbaum. 2009. "Crime, neighborhoods and units of analysis: Putting space in its place." Pp. 145-170 in *Putting crime in its place: Units of Analysis in Geographic Criminology*, edited by D. Weisburd, W. Bernasco, and G. Bruinsma. New York: Springer.

Tita, George and Steven Radil. 2011. "Spatializing the Social Networks of Gangs to Explore Patterns of Violence." *Journal of Quantitative Criminology* 27:521-545.

Tita, George and Greg Ridgeway. 2007. "The Impact of Gang Formation on Local Patterns of Crime." *Journal of Research in Crime and Delinquncy* 44:208-237.

Tobler, W. R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46:234-240.

Twist, Peter J. 2013. "Identity and Empowerment: Gangs in Nineteenth Century Philadelphia." Department of History, Temple University.

Uniform Crime Reports, . 2015. "Offenses Known to Law Enforcement, by State by City, 2013." *Crime in the United States, 2013*.

United States Census Bureau, . 2015. "2014 American Community Survey." Pp. http://ftp2.census.gov: U.S. Census Bureau's American Community Survey Office.

Venkatesh, S. 2008. *Gang Leader for a Day*. New York, NY: Penguin.

Ward, Jeffrey T., Matt R. Nobles, and Kathleen A. Fox. 2015. "Disentangling Self-Control from Its Elements: A Bifactor Analysis." *Journal of Quantitative Criminology* 31:595-627.

Weisburd, David. 2015. "The Law of Crime Concentration and the Criminology of Place." *Criminology* 53:133-157.

Weisburd, David, Gerben J.N. Bruinsma, and Wim Bernasco. 2009. "Units of Anslysis in Geographic Criminology: Historical Development, Critical Issues, and Open Questions." Pp. 3-31 in *Putting Crime in its Place*, edited by D. Weisburd, W. Bernasko, and G. J. N. Bruinsma. New York, NY Springer.

Western, Drew and Robert Rosenthal. 2003. "Quantifying Construct Validity: Two Simple Measures." *Journal of Personality and Social Psychology* 84:608-618.

Whyte, William Foote. 1955. *Street Corner Society: The Social Structure of an Italian Slum*: University of Chicago Press.

Zatz, Majorie S. 1987. "Chicano youth gangs and crime: the creation of a moral panic." *Contemporary Crises* 11:129-158.

APPENDICES

APPENDIX A. CRIME CODES USED TO IDENTIFY GUN-CRIME INCIDENTS

UCR	Description
111	Homicide: Handgun
112	Homicide: Rifle
113	Homicide: Shotgun
300	Robbery: Highway: Handgun
301	Robbery: Highway: Shotgun
302	Robbery: Highway: Rifle
310	Robbery: Commercial Establishment: Handgun
311	Robbery: Commercial Establishment: Shotgun
312	Robbery: Commercial Establishment: Rifle
316	Robbery: Cargo Hijack of Vehicle or Theft of Contents by Force: Handgun
317	Robbery: Cargo Hijack of Vehicle or Theft of Contents by Force: Shotgun
318	Robbery: Cargo Hijack of Vehicle or Theft of Contents by Force: Rifle
320	Robbery: Drug Store: Handgun
321	Robbery: Drug Store: Shotgun
322	Robbery: Drug Store: Rifle
330	Robbery: Gas Station: Handgun
331	Robbery: Gas Station: Shotgun
332	Robbery: Gas Station: Rifle
340	Robbery: Chain Store: Handgun
341	Robbery: Chain Store: Shotgun
342	Robbery: Chain Store: Rifle
350	Robbery: Residence: Handgun
351	Robbery: Residence: Shotgun
352	Robbery: Residence: Rifle
360	Robbery: Bank, Including Savings & Loan and Credit Unions: Handgun
361	Robbery: Bank, Including Savings & Loan and Credit Unions: Shotgun
362	Robbery: Bank, Including Savings & Loan and Credit Unions: Rifle
370	Robbery: Taxi Cab: Handgun
371	Robbery: Taxi Cab: Shotgun
372	Robbery: Taxi Cab: Rifle
376	Robbery: Taproom, State Store or Liquor Licensed Establishments: Handgun
377	Robbery: Taproom, State Store or Liquor Licensed Establishments: Shotgun
378	Robbery: Taproom, State Store or Liquor Licensed Establishments: Rifle
380	Robbery: Grocery Store/Delicatessen: Handgun
381	Robbery: Grocery Store/Delicatessen: Shotgun
382	Robbery: Grocery Store/Delicatessen: Rifle
388	Robbery: Vehicle: Handgun
389	Robbery: Vehicle: Shotgun
390	Robbery: Miscellaneous: Handgun
391	Robbery: Miscellaneous: Shotgun

392 Robbery: Miscellaneous: Rifle
396 Robbery: Vehicle: Rifle
401 Aggravated Assault: On a Student by School Employee: Handgun
402 Aggravated Assault: On a Student by School Employee: Shotgun
403 Aggravated Assault: On a Student by School Employee: Rifle
407 Aggravated Assault: Domestic Abuse: Handgun
408 Aggravated Assault: Domestic Abuse: Shotgun
409 Aggravated Assault: Domestic Abuse: Rifle
411 Aggravated Assault: Handgun
412 Aggravated Assault: Shotgun
413 Aggravated Assault: Rifle
421 Aggravated Assault: On a Philadelphia Police Officer: Handgun
422 Aggravated Assault: On a Philadelphia Police Officer: Shotgun
423 Aggravated Assault: On a Philadelphia Police Officer: Rifle
431 Aggravated Assault: On a Teacher/Employee, Public School: Handgun
432 Aggravated Assault: On a Teacher/Employee, Public School: Shotgun
433 Aggravated Assault: On a Teacher/Employee, Public School: Rifle
441 Aggravated Assault: On a Teacher/Employee, Private School: Handgun
442 Aggravated Assault: On a Teacher/Employee, Private School: Shotgun
443 Aggravated Assault: On a Teacher/Employee, Private School: Rifle
451 Aggravated Assault: On a Student, Public School: Handgun
452 Aggravated Assault: On a Student, Public School: Shotgun
453 Aggravated Assault: On a Student, Public School: Rifle
461 Aggravated Assault: On a Student, Private School: Handgun
462 Aggravated Assault: On a Student, Private School: Shotgun
463 Aggravated Assault: On a Student, Private School: Rifle
471 Aggravated Assault: On Other Law Enforcement: Handgun
472 Aggravated Assault: On Other Law Enforcement: Shotgun
473 Aggravated Assault: On Other Law Enforcement: Rifle
491 Aggravated Assault: Child Abuse: Handgun
492 Aggravated Assault: Child Abuse: Shotgun
493 Aggravated Assault: Child Abuse: Rifle
1501 Weapon Violations: Uniform Firearms Act (Adults): Handgun
1502 Weapon Violations: Uniform Firearms Act (Adults): Shotgun
1503 Weapon Violations: Uniform Firearms Act (Adults): Rifle
1504 Weapon Violations: Uniform Firearms Act (Adults): Carrying on Public Street w/o License
1505 Weapon Violations: Uniform Firearms Act (Adults): Possession of by Convict
1506 Weapon Violations: Uniform Firearms Act (Adults): All Other Firearms
1531 Weapon Violations: Uniform Firearms Act (on School Property/Business): Handgun
1532 Weapon Violations: Uniform Firearms Act (on School Property/Business): Shotgun
1533 Weapon Violations: Uniform Firearms Act (on School Property/Business): Rifle
1534 Weapon Violations: Uniform Firearms Act (on School Property/Business): Other Firearms
1541 Weapon Violations: Uniform Firearms Act (Juveniles): Handgun

1542	Weapon Violations: Uniform Firearms Act (Juveniles): Shotgun
1543	Weapon Violations: Uniform Firearms Act (Juveniles): Rifle
1544	Weapon Violations: Uniform Firearms Act (Juveniles): All Other Firearms

APPENDIX B. REGRESSION MODEL RESULTS AND TABLES

Proportion model results

The results from zero inflated beta regression models are estimated using the logit-link function. Incident rate ratios are calculated by exponentiating the estimated coefficients in beta regression models as they are done in other models that use the logit-link function (Ferrari and Cribari-Neto 2004; Long 1997: 360).

Adjusted probability values are used to determine statistical significance. Since multiple models were run on the same dependent variable, effects with p-values less than .01 are considered statistically significant.

Separate parameter estimates model zero values and non-zero values of the dependent variable separately. *Nu* parameters estimate the ability of the independent variables to discriminate between zero and non-zero values. In other words, this part of the model predicts whether or not the location contains any gang set space. Alternatively, *mu* parameters estimate the non-zero values; if there is a gang set space present, this part of the model estimates how much gang set space is present. The coefficients in these two models are interpreted in slightly different ways.

The *nu* parameter estimates reflect ability to predict a response of zero. A positive parameter estimate—an exponentiated coefficient greater than one—indicates that the variable *increases* the odds that the case will be zero. A negative parameter estimate—an exponentiated coefficient less than one—indicates that the variable *decreases* the odds that that case will be zero.

The *mu* parameter estimates reflect the ability to predict the sample mean for the non-zero values. In this part of the model, a positive parameter estimate reflects an *increase* in the expected proportion value of the dependent variable for each one-unit increase in the predictor variable. Negative coefficients reflect a *decrease* in the expected proportion value.

A series of five regression models are reported for each spatial unit. These five models include

- Baseline model: This model includes only the control variables as predictor variables.
- Model A: This model only includes the home variable and the control variables as predictor variables.
- Model B: This model only includes the incident variable and the control variables as predictor variables.
- Model C: This model only includes the gun variable and the control variables as predictor variables.
- Full model: This model includes all three independent variables and the control variables as predictor variables.

The results of the beta regression models at each of the six spatial units follow. In each of these models, the dependent variable is measured as the proportion of gang set space coverage. A spatially lagged version of the dependent variable was included to control for spatial autocorrelation. Neighbors were specified using first order queen contiguity.

Grid cell results. At the grid cell level, the Global Moran's Index p-value was less than .05. The null hypothesis for this test is that the values are randomly distributed. This is an indication that the residuals of these models were spatially clustered in a statistically significant way. Attempts were made to reduce or eliminate the spatially auto-correlated residuals. Gaussian Markov Random Fields[20] models would not converge at this spatial unit so spatially lagged

[20] These models account for spatial dependence by "bringing fitted values from neighbouring regions closer together (rather than shrink them towards the overall mean as is the case of a simple random effect model term)" Bastiani, Fernanda De, Robert A. Rigby, Dimitrios M. Stasinopoulous, Audrey H.M.A. Cysneiros, and Miguel A. Uribe-

versions of the dependent variable were the only available option to control for spatial autocorrelation. Spatial lag variables were created using first order queen contiguity through seventh order queen contiguity to define neighbors. In each of these models, the residuals remained spatially clustered at the global level. The BIC values are reported in Table 44.

Table 44: Addressing auto-correlated residuals, grid cell spatial unit BIC values

	Full model	Model A: Home	Model B: Incident	Model C: Gun
baseline	2681	3255	3149	3134
1st order queen lag	-2580	-2371	-2442	-2433
2nd order queen lag	-1222	-954	-1030	-1002
3rd order queen lag	-379	-69	-131	-91
4th order queen lag	166	510	441	509
5th order queen lag	561	936	854	939
6th order queen lag	868	1274	1180	1271
7th order queen lag	1082	1512	1409	1504

In the analyses that follow, results from the first order queen contiguity lag models are reported since the BIC values identified these models as fitting the data best.

Models with a single independent variable (Models A, B and C) all perform better than the baseline model. This is evident through the Likelihood ratio (LR) tests (control) and BIC comparisons. The LR test chi square coefficients are all significant at the .01 level. Furthermore, BIC values for Models A, B and C are all at least ten units smaller than the baseline model (BIC for baseline = -2183).

The Full Model outperforms Models A, B and C, which is also confirmed by both the LR test (full) and comparisons of BIC values. LR chi square values that compare the Full Model to Models A, B and C are all significant at the .01 level. The BIC for the Full Model (BIC=-2580) is at least ten units smaller than the BIC produced in Models A, B and C.

Opazo. 2016. "Gaussian Markov random field spatial models in GAMLSS." *Journal of Applied Statistics*.. Units that are not defined as neighbors are assumed to be conditionally independent.

The pseudo R^2 value is consistent across all 5 models. The baseline model had the smallest pseudo R^2 value (R^2 =.32) while the full model had the largest R^2 value (R^2 =.34)

Finally, the significance patterns in the *mu* and *nu* section of the hurdle models are consistent across Models A, B and C. In each of these models, the independent variables link in a positive and significant way in predicting the non-zero values (*mu* section). In Model A, each additional gang member living in an area links to a 21% increase in the expected proportion value. In Model B, each additional gang incident links to a 3% increase in the expected proportion value. In Model C, each additional gun incident links to a 4% increase in the expected proportion value.

In the *nu* models, each of the three independent variables reduce the likelihood of being in the no-gang group. The presence of a gang member living in a grid cell reduced the likelihood of that area having a proportion value of 0 by 40%. Similarly, the presence of a gang incident reduced the likelihood of that area being classified as a no-gang area by 21%, while the presence of a gun incident reduced the likelihood by 16%.

Turning attention to the significance patterns of each variable, the results of the Full Model are slightly different than the single indicator models. In the Full Model, the home and gun variables exerted a statistically significant effect at the modified alpha level ($p<.01$) in the *mu* section after controlling for the other independent variables. Each additional gang member living in an area increased the proportion value by 17%, even after controlling for gang incidents and gun crime. Similarly, each additional gun incident increased the proportion value by 2%. Turning to the *nu* section, all three independent variables decreased the odds of being in the no-gang group. Each additional gang member living in an area reduced the odds of that area being in the no-gang group by 19%, the unique effect of the incident variable reduced the odds by 11% and the unique effect of the gun variable reduced the odds by 12%. Another interesting finding concerns the size of the home variable coefficient. In the *mu* models, the home coefficient is much larger than the other two variables, but in the *nu* section, all three variables are somewhat comparable.

Across all five models, the spatial lag variable is statistically significant at the .01 level. A one unit increase in the spatial lag variable (nearby proportions increase from 0 to 1) results in an increase both in the expected proportion value and the likelihood that the case will contain some degree of gang presence (reduces the likelihood of being in the no-gang group).

Table 45: Grid cell regression output, proportion outcome (n=16,419)

	Model A: home only			Model B: incident only			Model C: gun only		
	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu (proportion)									
(Intercept)	-14.98			-14.81			-14.99		
Home	**0.19*** **	**1.21**	**0.02**						
Incident				**0.03*** **	**1.03**	**0.01**			
Gun							**0.04*** **	**1.04**	**0.01**
Spatial lag	**5.30*** **	**200.53**	**0.17**	**5.18*** **	**176.81**	**0.17**	**5.23*** **	**186.05**	**0.18**
Nu (zero)									
(Intercept)	-7.78			-7.62			-7.30		
Home	**-0.52*** **	**0.60**	**0.05**						
Incident				**-0.24*** **	**0.79**	**0.02**			
Gun							**-0.18*** **	**0.84**	**0.01**
Spatial lag	**-24.86*** **	**1.6E-11**	**0.68**	**-25.16*** **	**1.2E-11**	**0.70**	**-24.76*** **	**1.8E-11**	**0.70**
Diagnostics									
Global deviance	-2439			-2510			-2501		
BIC	-2371			-2442			-2433		
Pseudo R^2	0.33			0.33			0.33		
Moran's I	**0.02*** **			**0.02*** **			**0.02*** **		
LR test (control)	**207.52*** **			**278.95*** **			**269.10*** **		
LR test (full)	**247.47*** **			**176.04*** **			**185.90*** **		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 45: Grid cell regression output, proportion outcome (n=16,419) continued

	Baseline Model			Full Model		
	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu (proportion)						
(Intercept)	-14.62			-15.20		
Home				**0.16*** **	**1.17**	**0.02**
Incident				0.01*	1.01	4.6E-3
Gun				**0.02** **	**1.02**	**0.01**
Spatial lag	**4.92*** **	**137.28**	**0.17**	**5.51*** **	**247.51**	**0.18**
Sigma						
(Intercept)	0.19			0.34		
Nu (zero)						
(Intercept)	-7.98			-7.23		
Home				**-0.22*** **	**0.81**	**0.06**
Incident				**-0.12*** **	**0.89**	**0.02**
Gun				**-0.12*** **	**0.88**	**0.01**
Spatial lag	**-25.89*** **	**5.7E-12**	**0.68**	**-24.40*** **	**1.5E-11**	**0.71**
Diagnostics						
df	5			11		
Global deviance	-2231			-2686		
BIC	-2183			-2580		
Pseudo R^2	0.32			0.34		
Moran's I	**0.02*** **			**0.01*** **		

Note: *=p<.05, **=p<.01, ***=p<.001

Thiessen polygon results. At the thiessen polygon level, the Global Moran's Index p value was less than .05. This is an indication that the residuals of these models were spatially clustered in a statistically significant way. Attempts were made to reduce or eliminate the spatially auto-correlated residuals. The Gaussian Markov Random Fields models would not converge, so various spatial lag models were run. Lag variables were created using first order queen contiguity through seventh order queen contiguity to define neighbors. In each of these models, the residuals remained spatially clustered at the global level.

Table 46: Addressing auto-correlated residuals, thiessen polygon unit BIC values

	Full model	Model A: Home	Model B: Incident	Model C: Gun
Baseline	2398	3050	2813	2983
1st order queen lag	-9811	-9606	-9751	-9749
2nd order queen lag	-7449	-7127	-7338	-7275
3rd order queen lag	-5578	-5167	-5408	-5302
4th order queen lag	-4153	-3687	-3934	-3802
5th order queen lag	-3128	-2623	-2865	-2716
6th order queen lag	-2310	-1776	-2012	-1856
7th order queen lag	-1587	-1023	-1260	-1097

In the analyses that follow, results from the first order queen contiguity lag models are reported since the BIC values identified these models as fitting the data best at both spatial units.

Using data aggregated to thiessen polygons around street intersections, models with a single independent variable (Models A, Band C) all perform better than the Baseline Model. This is evident through the LR tests (control) and BIC comparisons. The LR test chi square coefficients are all significant at the .01 level. Furthermore, BIC values for Models A, B and C are all at least ten units smaller than the Baseline Model (BIC for baseline =-9605).

The Full Model outperforms all Models A, B and C as well, which is also confirmed by both the LR test (full) and comparisons of BIC values. LR chi square values that compare the Full

Model to Models A, B and C are all significant at the .01 level. The BIC for the Full Model (BIC=-9811) is at least ten units smaller than the BIC produced in Models A, B and C.

The pseudo R^2 value is consistent across all five models; all models have a pseudo R^2 value of .44.

The significance patterns in the *mu* and *nu* section are *not* consistent across Models A, B and C. Only the home and incident variable link in a positive and significant way in predicting the non-zero values (mu section). In Model A, each additional gang member living in the area linked to a 9% increase in the expected proportion value. In Model B, each additional gang incident linked to a 2% increase in the expected proportion value. The effect of the gun variable, however, was not significant at the .01 level.

In the *nu* models, increases in all three variables reduced the likelihood of being in the no-gang group. Each additional gang member living in an area reduced the likelihood of that area being classified as a no-gang area by 20%. Furthermore, each additional gang incident reduced the likelihood of that area being classified as a no-gang area by 23%, while the presence of a gun incident reduced the likelihood by 21%.

Turning attention to the significance patterns of each variable, the results of the Full Model are slightly different. In the Full Model, all three variables exerted a statistically significant effect in the *mu* section after controlling for the other independent variables. Each additional gang member living in an area increased the proportion value by 8%, while each additional gang incident increased the proportion value by 2%. The gun variable, however, worked in the opposite direction; this variable *reduced* the proportion value by 3%. Turning to the *nu* section, gang incidents and gun crime decreased the odds of being in the no-gang group. The unique effect of the incident variable and the gun variable both reduced the odds of being in the no gang group by 17%. The home variable was not significant at the .01 level.

Across all five models, the spatial lag variable was statistically significant. A one unit increase in the spatial lag variable (nearby proportions increase from 0 to 1) resulted in an increase both in the expected proportion value and the likelihood that the case will contain some degree of gang presence (reduces the likelihood of being in the no-gang group).

Table 47: Thiessen polygon regression output, proportion outcome (n=22,396)

		Model A: home			Model B: incident			Model C: gun		
		B	Exp(b)	Std. Error	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu	(Intercept)	-13.94			-13.96			-13.85		
	Home	**0.08*****	**1.09**	**0.02**						
	Incident				**0.02*****	**1.02**	**4.6E-3**			
	Gun							-0.02	0.98	0.01
	Spatial lag	**6.08*****	**434.89**	**0.12**	**6.12*****	**456.41**	**0.12**	**6.02*****	**412.66**	**0.12**
Sigma	(Intercept)	0.83			0.83			0.82		
Nu	(Intercept)	-6.42			-6.24			-5.97		
	Home	**-0.22****	**0.80**	**0.07**						
	Incident				**-0.26*****	**0.77**	**0.02**			
	Gun							**-0.24*****	**0.79**	**0.02**
	Spatial lag	**-20.86*****	**8.7E-10**	**0.46**	**-20.88*****	**8.5E-10**	**0.47**	**-21.06*****	**7.1E-10**	**0.48**
Diagnostics										
	Df	7			7			7		
	Global deviance	-9676			-9821			-9819		
	BIC	-9606			-9751			-9749		
	Pseudo R^2	0.44			0.44			0.44		
	Moran's I	**0.02*****			**0.02*****			**0.01****		
	LR test (control)	**20.79*****			**165.75*****			**163.70*****		
	LR test (full)	**245.40*****			**100.43*****			**102.48*****		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 47: Thiessen polygon regression output, proportion outcome (n=22,396) continued

		Baseline Model			Full Model		
		B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu	(Intercept)	-13.90			-13.92		
	Home				**0.07****	**1.08**	**0.02**
	Incident				**0.02*****	**1.02**	**5.0E-3**
	Gun				**-0.03****	**0.97**	**0.01**
	Spatial lag	**6.05*****	**425.85**	**0.12**	**6.09*****	**442.55**	**0.12**
Sigma	(Intercept)	0.82			0.84		
Nu	(Intercept)	-6.46			-5.90		
	Home				-0.14	0.87	0.07
	Incident				**-0.18*****	**0.83**	**0.02**
	Gun				**-0.19*****	**0.83**	**0.02**
	Spatial lag	**-21.02*****	**7.5E-10**	**0.46**	**-20.92*****	**8.3E-10**	**0.48**
Diagnostics							
	df	5			11		
	Global deviance	-9655			-9921		
	BIC	-9605			-9811		
	Pseudo R^2	0.44			0.44		
	Moran's I	**0.02*****			**0.02*****		

Note: *=p<.05, **=p<.01, ***=p<.001

Census block group results. Using data aggregated to the census block group level, models with a single independent variable (Models A, Band C) all perform better than the baseline model. This is evident through the LR tests (control) and BIC comparisons. The LR test chi square coefficients are all significant at the .01 level. Furthermore, BIC values for Models A, B and C are all at least ten units smaller than the Baseline Model (BIC for baseline =130).

The Full Model outperforms all Models A, B and C as well, which is also confirmed by both the LR test (full) and comparisons of BIC values. Likelihood ratio chi square values that compare the Full Model to Models A, B and C are all significant at the .01 level. The BIC for the Full Model (BIC=-158) is at least ten units smaller than the BIC produced in Models A, B and C.

The pseudo R^2 value is consistent across all five models. The Baseline Model had the smallest pseudo R^2 value (R^2 =.36) while the Full Model had the largest pseudo R^2 value (R^2 =.47)

The significance patterns in the *mu* and *nu* section are *not* consistent across Models B and C, but are consistent across Model A. In the *mu* section, the home variable and the gun variable links in a significant way in predicting the non-zero values (mu section). In Model A, each additional gang member living in an area increased the likelihood of the expected proportion value by 5%. In Model C, however, each additional gun incident linked to a 1% *decrease* in the expected proportion value. The effect of the incident variable was not significant at the .01 level.

In the *nu* models, all three variables reduced the likelihood of being in the no-gang group. Each additional gang member living in the area reduced the likelihood of that area being classified as a no-gang area by 31%. Each additional gang incident reduced the likelihood of that area being classified as a no-gang area by 13%, while the presence of a gun incident reduced the likelihood by 5%.

Results of the Full Model are slightly different than the single indicator models. In the Full Model, only the home variable and the gun variable exerted a statistically significant effect in the *mu* section after controlling for the other variables. Each additional gang member living in an area increased the expected proportion value by 7%, even after controlling for gang incidents and gun crime. However, the gun variable worked in the opposite direction; each additional gun incident *reduced* the expected proportion value by 2%. Turning to the *nu* section, all three variables decreased the odds of being in the no-gang group. The unique effect of the home variable reduced the odds of being in the no-gang group by 17%, even after controlling for the other independent variables. The incident variable reduced the odds of being in the no-gang group by 6% and the gun variable reduced the odds by 3%.

Across all five models, the spatial lag variable is statistically significant. A one unit increase in the spatial lag variable (nearby proportions increase from 0 to 1) results in an increase both in the expected proportion value and the likelihood that the case will contain some degree of gang presence (reduces the likelihood of being in the no-gang group).

Table 48: Census block group regression output, proportion outcome (n=1,336)

	Model A: home			Model B: incident			Model C: gun		
	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu (proportion)									
(Intercept)	-16.20			-16.04			-15.43		
Home	**0.05****	**1.05**	**0.02**						
Incident				0.01	1.01	3.0E-3			
Gun							**-0.01*****	**0.99**	**3.6E-3**
Spatial lag	**4.26*****	**70.72**	**0.41**	**4.16*****	**63.92**	**0.41**	**3.58*****	**35.91**	**0.41**
Sigma									
(Intercept)	-0.33			-0.34			-0.35		
Nu (zero)									
(Intercept)	-11.27			-11.39			-10.79		
Home	**-0.38*****	**0.69**	**0.04**						
Incident				**-0.14*****	**0.87**	**0.01**			
Gun							**-0.06*****	**0.95**	**0.01**
Spatial lag	**-21.01*****		**1.42**	**-20.37*****	**1.4E-9**	**1.38**	**-21.81*****	**3.4E-10**	**1.39**
Diagnostics									
Df	7			7			7		
Global deviance	-67			-69			-54		
BIC	-16			-18			-4		
Pseudo R^2	0.43			0.43			0.42		
Moran's I	**0.05****			0.02			**0.04****		
LR test (control)	**160.27*****			**162.06*****			**148.01*****		
LR test (full)	**91.63*****			**89.83*****			**103.88*****		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 48: Census block group regression output, proportion outcome (n=1,336) continued

	Baseline Model			Full Model		
	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu (proportion)						
(Intercept)	-15.93			-15.66		
Home				**0.07*****	**1.07**	**0.02**
Incident				2.7E-3	1.00	4.1E-3
Gun				**-0.02*****	**0.98**	**4.0E-3**
Spatial lag	**3.99*****	**48.35**	**0.40**	**3.88*****		**0.42**
Sigma						
(Intercept)	-0.36			-0.29		
Nu (zero)						
(Intercept)	-12.20			-10.56		
Home				**-0.19*****	**0.83**	**0.04**
Incident				**-0.06*****	**0.94**	**0.02**
Gun				**-0.03*****	**0.97**	**0.01**
Spatial lag	**-22.36*****	**1.9E-10**	**1.37**	**-20.86*****	**8.7E-10**	**1.43**
Diagnostics						
Df	5			11		
Global deviance	94			-158		
BIC	130			-79		
Pseudo R^2	0.36			0.47		
Moran's I	**0.054*****			**0.028***		

Note: *=p<.05, **=p<.01, ***=p<.001

Tract results. Using data aggregated to the tract level, models with a single independent variable (Models A, B and C) all perform better than the Baseline Model. This is evident through the LR tests (control) and BIC comparisons. The likelihood ratio test chi square coefficients are all significant at the .01 level. Furthermore, BIC values for Models A, B and C are all at least ten units smaller than the Baseline Model (BIC for baseline =59).

The Full Model outperforms all Models A, B and C, which is also confirmed by both the LR test (full) and comparisons of BIC values. Likelihood ratio chi square values that compare the Full Model to Models A, B and C are all significant at the .01 level. The BIC for the Full Model (BIC=-65) is at least ten units smaller than the BIC produced in Models A, B and C.

The pseudo R^2 values are relatively consistent across Models A, B and C. Models A and B are very similar, with R^2 values of .47 and .49, respectively. Model C, however, had a smaller R^2 value of .40. The Baseline Model had the smallest pseudo R^2 value (R^2=.26) while the Full Model had the largest R^2 value (R^2=.51)

The significance patterns in the *mu* and *nu* section are *not* consistent across Models A, B and C. None of the variables are significant in predicting the non-zero values (*mu* section). However, in the *nu* models, all three variables reduced the likelihood of being in the no-gang group. Each additional gang member living in the area reduced the likelihood of that area being classified as a no-gang area by 23%. Each additional gang incident reduced the likelihood of that area being classified as a no-gang area by 13%, while the presence of a gun incident reduced the likelihood by 2%.

The significance patterns generated by the Full Model are different from the individual indicator models. In the Full Model, none of the variables exerted a statistically significant effect in the *mu* section after controlling for the other variables. Turning to the *nu* section, only the incident variable decreased the odds of being in the no-gang group in a statistically meaningful way ($p<.01$). The unique effect of the incident variable reduced the odds of being in the no-gang group by 9%, even after controlling for the other independent variables.

Across all five models, the spatial lag variable is statistically significant. A one unit increase in the spatial lag variable (nearby proportions increase from 0 to 1) results in an increase both in the expected proportion value and the likelihood that the case will contain some degree of gang presence (reduces the likelihood of being in the no-gang group).

Table 49: Census tract regression output, proportion outcome (n=384)

	Model A: home			Model B: incident			Model C: gun		
	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu (proportion)									
(Intercept)	-18.36			-18.20			-17.77		
Home	0.02	1.02	0.01						
Incident				-2.6E-3	1.003	1.8E-3			
Gun							-2.0E-3	0.998	1.6E-3
Spatial lag	**5.21*****	**182.42**	**0.74**	**5.05*****	**155.66**	**0.73**	**4.48*****	**88.08**	**0.75**
Sigma									
(Intercept)	0.46			0.43			0.40		
Nu (zero)									
(Intercept)	-12.87			-12.71			-12.87		
Home	**-0.26*****	**0.77**	**0.03**						
Incident				**-0.14*****	**0.87**	**0.02**			
Gun							**-0.02*****	**0.98**	**2.9E-3**
Spatial lag	**-18.01*****	**1.5E-8**	**2.74**	**-15.36*****	**2.1E-7**	**2.62**	**-20.08*****	**1.9E-9**	**2.70**
Diagnostics									
Df	7			7			7		
Global deviance	-97			-114			-54		
BIC	-55			-72			-13		
Pseudo R^2	0.47			0.49			0.40		
Moran's I	0.01			-0.05			0.03		
LR test (control)	**126.10*****			**143.53*****			**83.96*****		
LR test (full)	**33.83*****			**16.41*****			**75.97*****		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 49: Census tract regression output, proportion outcome (n=384) continued

	Baseline Model			Full Model		
	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu (proportion)	-18.05			-17.93		
(Intercept)				0.02	1.02	0.01
Home				-2.1E-3	1.002	2.7E-3
Incident				-4.4E-3*	0.996	1.9E-3
Gun	**4.80*** **	**121.73**	**0.71**	**4.76*** **	**116.37**	**0.75**
Spatial lag						
Sigma	0.41			0.46		
(Intercept)						
Nu (zero)	-14.64			-12.44		
(Intercept)				-0.12*	0.89	0.05
Home				**-0.10*** **	**0.91**	**0.02**
Incident				-3.0E-4	0.99	4.4E-3
Gun	**-20.60*** **	**1.1E-9**	**2.54**	**-15.87*** **	**1.3E-7**	**2.70**
Spatial lag						
Diagnostics						
df	5			11		
Global deviance	30			-130		
BIC	59			-65		
Pseudo R^2	0.26			0.51		
Moran's I	-4E-3			-0.02		

Note: *=p<.05, **=p<.01, ***=p<.001

Neighborhood results. Using data aggregated to the neighborhood level, models with a single independent variable (Models A, Band C) all perform better than the Baseline Model. This is evident through the LR tests (control) and BIC comparisons. The LR test chi square coefficients are all significant at the .01 level. Furthermore, BIC values for Models A, B and C are all at least ten units smaller than the Baseline Model (BIC =84).

There is mixed evidence concerning whether the Full Model outperforms all Models A, B and C. The LR chi square values that compare the Full Model to Models A, B and C are all significant at the .01 level. However, the BIC for the Full Model (BIC=-12) is not at least ten units smaller than the BIC produced in Models A and B. The BIC produced from Models A and B was -7. The BIC produced by Model C, however, is at least ten units greater than the Full Model. This means that the Full Model is preferred over Model C, but it is not substantially preferable over Models A and B.

The pseudo R^2 values are not consistent across Models. Models A and B are the same with R^2 values of .43. Model C, however, had a smaller R^2 value of .36. The Baseline Model had the smallest pseudo R^2 value (R^2=-.09) while the Full Model had the largest R^2 value (R^2=.51)

The significance patterns in the *mu* and *nu* section are *not* consistent across Models A, B and C. Only the gun variable is significant in predicting the non-zero values (mu section). In Model C, each additional gun incident linked to a .19% *decrease* in the expected proportion value. The effect of the home and the incident variables were not significant at the .01 level. However, in the *nu* models, all three variables reduced the likelihood of being in the no-gang group. Each additional gang member living in the area reduced the likelihood of that area being classified as a non-gang area by 16%. Each additional gang incident reduced the likelihood of that area being classified as a no-gang area by 8%, while the presence of a gun incident reduced the likelihood by 1%.

Results of the Full Model are different from the single indicator models. In the Full Model, only the gun variable exerted a statistically significant effect in the *mu* section after controlling for the other variables. Each additional gun incident *reduced* the expected proportion value by 0.28%. Turning to the *nu* section, only the home variable was significant in predicting the no-gang group after controlling for the other independent variables. Each additional gang member living in an area reduced the likelihood of being in the no-gang group by 20%.

The spatial lag variable is statistically significant at the .01 level in the *nu* section of all five models. It is not, however, consistently significant at the .01 level in the *mu* section of the models. Even in instances when the coefficient was not significant, the direction of the coefficient is always in the expected direction. A one unit increase in the spatial lag variable (nearby proportions increase from 0 to 1) results in an increase both in the expected proportion value and the likelihood that the case will contain some degree of gang presence (reduces the likelihood of being in the no-gang group).

Table 50: Neighborhood regression output, proportion outcome (n=158)

	Model A: home			Model B: incident			Model C: gun		
	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu (proportion)									
(Intercept)	-19.40			-19.58			-18.80		
Home	-0.01	0.99	4.4E-3						
Incident				-8.2E-4	0.99	1.2E-3			
Gun							**-1.9E-3*** **	**0.998**	**5.5E-4**
Spatial lag	**7.43*****	**1678.53**	**2.11**	**7.89*****	**2677.03**	**2.10**	5.11*	166.25	2.24
Sigma									
(Intercept)	1.19			1.19			1.20		
Nu (zero)									
(Intercept)	-13.38			-13.33			-13.85		
Home	**-0.17*****	**0.84**	**0.03**						
Incident				**-0.08*****	**0.92**	**0.01**			
Gun							**-0.01*****	**0.99**	**1.6E-3**
Spatial lag	**-22.35*****	**2.0E-10**	**5.45**	**-18.59****	**8.4E-9**	**5.57**	**-24.04*****	**3.6E-11**	**5.15**
Diagnostics									
df	7			7			7		
Global deviance	-42			-42			-25		
BIC	-7			-7			10		
Pseudo R^2	0.43			0.43			0.36		
Moran's I	-0.08			-0.03			-0.05		
LR test (control)	**101.32*****			**101.22*****			**83.99*****		
LR test (full)	**25.20*****			**25.30*****			**42.53*****		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 50: Neighborhood regression output, proportion outcome (n=158) continued

	Baseline Model			Full Model		
	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu (proportion)						
(Intercept)	-19.68			-18.84		
Home				4.9E-3	1.005	0.01
Incident				1.5E-3	1.002	1.9E-3
Gun				**-2.8E-3*** **	**0.997**	**7.7E-4**
Spatial lag	**8.17*** **	**3537.75**	**2.07**	4.87*	129.78	2.25
Sigma						
(Intercept)	1.19			1.24		
Nu (zero)						
(Intercept)	-15.55			-13.13		
Home				**-0.23** **	**0.80**	**0.08**
Incident				-0.06*	0.94	0.02
Gun				0.01	1.01	0.01
Spatial lag	**-27.55*** **	**1.1E-12**	**5.11**	**-19.65** **	**2.9E-9**	**5.91**
Diagnostics						
df	5			11		
Global deviance	59			-68		
BIC	84			-12		
Pseudo R^2	-0.09			0.51		
Moran’s I	-0.08			-0.02		

Note: *=p<.05, **=p<.01, ***=p<.001

Community results. Using data aggregated to the community level, there is mixed evidence concerning whether Models A, B and C outperform the Baseline Model. The LR chi square values that compare Models A, B and C to the Baseline Model are all significant at the .01 level. However, the BIC for Model C is not at least ten units smaller than the BIC produced in the Baseline Model. The BIC produced from Model C is -20; the BIC for the Baseline Model is -14. The BIC produced by Models A and B (-38 and -37, respectively), however, *are* at least ten units smaller than the Baseline Model. This means that Models A and B are preferred over the Baseline Model, but Model C is not preferred over the Baseline Model.

There is mixed evidence concerning whether the Full Model outperforms all Models A, B and C. The LR chi square values that compare the Full Model to Models A, B and C are all significant at the .01 level. However, the BIC for the Full Model (BIC=-36) is not at least ten units smaller than the BIC produced in Models A and B. The BIC produced from Models A and B are -38 and -37, respectively. The BIC produced by Model C, however, is at least ten units greater than the Full Model. This means that the Full Model is preferred over Model C, but it is not preferred over Models A and B.

The pseudo R^2 values are not consistent across models. Models A and B are similar, with R^2 values of .38 and .37, respectively. Model C, however, had a much smaller R^2 value of .09. The baseline model had the smallest pseudo R^2 value (R^2=-.26) while the full model had the largest R^2 value (R^2=.54)

The significance patterns in the *mu* and *nu* section are consistent across Models A, B and C. None of the variables are significant in predicting the non-zero values (*mu* section). This largely true in the *nu* models as well; only the gun variable is significant at the .01 level. Each additional gun incident reduces the odds of being in the no-gang group by 5%. I the Full Model,none of the variables are significant in predicting either part of the model (*mu* or *nu*).

The spatial lag variable is not significant in any of the models.

Table 51: Community regression output, proportion outcome (n=45)

	Model A: home			Model B: incident			Model C: gun		
	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu (proportion)									
(Intercept)	-21.62			-21.44			-21.52		
Home	2.6E-03	1.00	3.3E-3						
Incident				3.5E-4	1.00	7.0E-4			
Gun							1.8E-04	1.00	5.5E-4
Spatial lag	8.27*	3887.80	3.90	7.87*	2623.89	3.85	8.18	3582.28	4.10
Sigma									
(Intercept)	1.70			1.67			1.68		
Nu (zero)									
(Intercept)	-12.19			-13.65			-15.95		
Home	-0.17*	0.84	0.07						
Incident				-0.06*	0.94	0.02			
Gun							**-0.01****	**0.995**	**2.0E-3**
Spatial lag	-22.51	1.7E-10	16.44	-17.49	2.5E-8	14.87	-20.51	1.2E-9	13.75
Diagnostics									
df	7			7			7		
Global deviance	-64			-63			-47		
BIC	-38			-37			-20		
Pseudo R^2	0.38			0.37			0.09		
Moran's I	4.0E-3			0.08			0.07		
LR test (control)	**31.79*****			**30.82*****			**14.45*****		
LR test (full)	**13.30*****			**14.28*****			**30.64*****		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 51: Community results (n=45) continued

	Baseline Model			Full Model		
	B	Exp(b)	Std. Error	B	Exp(b)	Std. Error
Mu (proportion)						
(Intercept)	-21.34			-21.75		
Home				0.01	1.01	0.01
Incident				-8.3E-4	0.99	1.5E-3
Gun				-1.4E-4	0.99	5.8E-4
Spatial lag	7.70*	2210.09	3.81	8.53	5057.14	5.03
Sigma						
(Intercept)	1.65			1.72		
Nu (zero)						
(Intercept)	-18.60			357.70		
Home				-8.21	2.7E-4	9.2E+3
Incident				-2.38	0.09	3.0E+3
Gun				0.22	1.25	5.9E+2
Spatial lag	-20.28	1.6E-9	12.17	-1.1E+3	0.00	1.5E+6
Diagnostics						
df	5			11		
Global deviance	-33			-78		
BIC	-14			-36		
Pseudo R^2	-0.26			0.54		
Moran's I	0.03			-0.03		

Note: *=p<.05, **=p<.01, ***=p<.001

Count model results

Another set of models were run which modeled the gang dependent variable in a slightly different way. This was done in an attempt to overcome issues with spatial autocorrelation which were a persistent issue in the proportion models. It is possible that the use of proportion overlap values may not be the best way to model gang magnitude; as an alternative, gang counts are used instead. A visual highlighting the difference between the proportion values and the count values can be found on page 95 in Figure 3. In this figure, one census tract is highlighted in red. 40% of this census tract is covered by gang set space; therefore, the proportion value for this record is .40. However, there are two different gangs present within this census tract, so the gang count value is two. While the previously discussed results all used the proportion values to model the dependent variable, the results reported next all use the number of gangs present within each spatial unit.

The transition from a proportion outcome to a count outcome required an adjustment in the modeling technique. The zero inflated beta regression models were replaced with zero adjusted negative binomial models. As was done in the beta regression models, a first order queen contiguity lag variable was included in these models as well.

The results from zero adjusted negative binomial models are also estimated using the logit-link function. Incident rate ratios are calculated by exponentiating the estimated coefficients. Furthermore, adjusted probability values are used to determine statistical significance. Since multiple models were run on the same dependent variable, effects with p-values less than .01 are considered statistically significant.

These models are interpreted in the same way the proportion models are interpreted. The *nu* parameters estimate the ability of the independent variables to discriminate between zero and non-zero values; if the area contains a gang set space. The *mu* parameters estimate how much gang set pace there is. In these models, this “how much” question is modeled as the number of gangs in a location, not the amount of space gang occupies as was the case with the proportion outcome.

A series of five regression models are reported for each spatial unit. These five models include the Baseline Model, Model’s A-C and a Full Model; models are specified exactly as they were in the beta regression models.

The results of the zero adjusted negative binomial regression models at each of the six spatial units follow. In each of these models, the dependent variable is measured as the number of gangs present. A spatially lagged version of the dependent variable was included to control for spatial autocorrelation. Neighbors were specified using first order queen contiguity.

Grid cell results. At the grid cell level, evidence suggesting models with a single independent variable (Models A, B and C) perform better than the Baseline Model is mixed. The LR test chi square coefficients are all significant at the .01 level. In line with these findings, BIC values for Models B and C are all at least ten units smaller than the Baseline Model (BIC = 2696). The BIC for Model A, however, is only one unit smaller than the Baseline Model and therefore is not preferred after considering added complexity.

Evidence suggesting the Full Model outperforms the single indicator models is also mixed. While the LR chi square values that compare the Full Model to Models A, B and C are all significant at the .01 level, the BIC comparisons tell a different story. The BIC for the Full Model (BIC=2658) is at least ten units smaller than the BIC produced in Model A which means that the Full Model is preferred over the home only model, even after controlling for model complexity. This is not the case, however, for Models B and C. The BIC values for Models B and C are actually smaller than the BIC for the Full Model. This means, that after controlling for model complexity, the single indicators models using incidents or gun crime are actually better than the Full Model.

This improvement is only four units for Model B, which is not compelling evidence to select Model B over the Full Model. For Model C, however, the difference in BIC is 12 units, which means the gun only model does significantly better than the Full Model after controlling for model complexity.

The pseudo R^2 value is consistent across all five models. All five models produced a pseudo R^2 value of .29

The significance patterns in the *mu* and *nu* section are consistent across Models A, B and C. In each of these models, each of the independent variables link in a positive way in predicting the non-zero values (*mu* section), but only the incident variable (Model B) reaches statistical significance at the .01 level. In this model, each additional gang incident that occurred in an area links to a 2% increase in the expected count value. Each additional gang member living in an area links to a 9% increase in the expected gang count, while each additional gun incident links to a 5% increase.

In the *nu* models, each of the three independent variables reduce the likelihood of being in the no-gang group in a statistically significant way. The presence of a gang incident reduced the likelihood of that area being classified as a no-gang area by 10%, while the presence of a gun incident reduced the likelihood by 9%. The home variable reduced the likelihood by 16%.

The significance patterns generated by the Full Model are slightly different than the single indicator models. In the Full Model, none of the variables exerted a statistically significant independent effect in the *mu* section after controlling for the other two independent variables. Turning to the *nu* section, all three independent variables decreased the odds of being in the no-gang group, but only the gun and incident variables rose to statistical significance at the .01 level. The unique effect of the incident variable reduced the odds of being in the no gang group by 6%

while the effect of the gun variable reduced the odds by 7% even after controlling for the other two variables.

Across all five models, the spatial lag variable is statistically significant. A one unit increase in the spatial lag variable (nearby counts increase from 0 to 1) results in an increase both in the expected count value and the likelihood that the case will contain some degree of gang presence (reduces the likelihood of being in the no-gang group). The Global Moran's Index value is not significant in any of the five models, which means the residuals are not spatially clustered in any of the models.

Table 52: Grid cell regression output, count outcome (n=16,419)

	Model A: home			Model B: incident			Model C: gun		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)									
(Intercept)	-15.79			-15.95			-16.06		
Home	**0.08****	**1.09**	**0.03**						
Incident				**0.02*****	**1.02**	**0.00**			
Gun							**0.05****	**1.05**	**0.01**
Spatial lag	**1.43*****	**4.16**	**0.19**	**1.56*****	**4.74**	**0.26**	**1.51*****	**4.53**	**0.09**
Sigma									
(Intercept)	-5.06			-2.51			-7.9E+13		
Nu (zero)									
(Intercept)	-7.39			-7.33			-7.15		
Home	**-0.18****	**0.84**	**0.05**						
Incident				**-0.11*****	**0.90**	**0.02**			
Gun							**-0.09*****	**0.91**	**0.01**
Spatial lag	**-10.46*****	**2.8E-05**	**0.27**	**-10.39*****	**3.1E-05**	**0.27**	**-10.23*****	**3.5E-05**	**0.27**
Model diagnostics									
Df	7			7			7		
Global deviance	2627			2586			2578		
BIC	2695			2654			2646		
Pseudo R^2	0.29			0.29			0.29		
Moran's I	-0.01			-0.01			0.00		
LR test (control)	**20.11*****			**61.16*****			**69.62*****		
LR test (full)	**76.17*****			**35.12*****			**26.65*****		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 52: Grid cell regression output, count outcome (n=16,419) continued

	Baseline Model			Full Model		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)						
(Intercept)	-15.83			-16.11		
Home				0.05	1.06	0.04
Incident				0.01*	1.01	0.00
Gun				0.04*	1.04	0.02
Spatial lag	**1.55*** **	**4.69**	**0.21**	**1.47*** **	**4.34**	**0.09**
Sigma						
(Intercept)	-2.803		2.416	-1.6E+14		
Nu (zero)						
(Intercept)	-7.42			-7.16		
Home				-0.03	0.97	0.06
Incident				**-0.06** **	**0.94**	**0.02**
Gun				**-0.07*** **	**0.93**	**0.01**
Spatial lag	**-10.67*** **	**2.3E-05**	**0.27**	**-10.15*** **	**3.9E-05**	**0.28**
Model diagnostics						
df	5			11		
Global deviance	2647			2551		
BIC	2696			2658		
Pseudo R^2	0.29			0.29		
Moran's I	0.00			-0.01		

Note: *=p<.05, **=p<.01, ***=p<.001

Thiessen results. At the thiessen polygon level, evidence that models with a single independent variable (Models A, Band C) perform better than the Baseline Model is mixed. This is evident through the LR tests (control) and BIC comparisons. The LR test control chi-square values are significant at the .01 level for Model B and Model C. In line with these findings, BIC values for Models B and C are all at least ten units smaller than the Baseline Model (BIC = 4214). The LR test and BIC for Model A, however, suggest that Model A does not outperform the Baseline Model. The BIC for Model A, is 20 units greater than the Baseline Model and therefore is not preferred after considering added complexity. The chi-square test statistic is not significant (χ^2=0.15, 2df, p=0.927).

Evidence about whether the Full Model outperforms the single indicator models is also mixed. While the LR chi square values that compare the Full Model to Models A, B and C are all significant at the .01 level, the BIC comparisons tell a different story. The BIC for the Full Model (BIC=4149) is at least ten units smaller than the BIC produced in Models A and B. Which means that the Full Model is preferred over the home only model (Model A) and the incident only model (Model B), even after controlling for model complexity. This is not the case, however, for Model C. The BIC value for Model C is actually smaller than the BIC for the Full Model. This means, that after controlling for model complexity, the gun model is actually preferred over the Full Model (BIC improvement is 19 units over the Full Model).

With the exception of the Baseline Model, the pseudo R^2 values are consistent across models. All models produced a pseudo R^2 value of .39. The Baseline Model, however, produced a pseudo R^2 value less than zero[21].

[21] This happened because the log-likelihood of the null model was slightly larger than the log-likelihood of the alternative model. The actual pseudo R^2 value was -.000034.

The significance patterns in the *mu* and *nu* section are not not significant across Models A, B and C. In each of these models the coefficient of the home value is negative, suggesting each additional gang member living in an area reduces the predicted mean gang count (*mu* section); however, this effect does not reach statistical significance. While the incident and gun variables link in a positive way in predicting the non-zero values, neither variable reaches statistical significance.

Turning to the *nu* models, the coefficient for the home variable is not significant. The coefficients of the incident and the gun variables are each negative, which means each variable reduces the likelihood the area is in the no-gang group. Both of these variables reached statistical significance at the .01 level. The presence of a gang incident reduced the likelihood of that area being classified as a no-gang area by 17%, while the presence of a gun incident reduced the likelihood by 11%.

In the Full Model, as in the single predictor models, none of the variables exerted a statistically significant independent effect on the count (*mu*). Turning to the absence (*nu*) section, the incident and gun variables remained significant and negative, suggesting these two variable decreased the odds of no gangs being present. The presence of a gang incident reduced the likelihood of that area being classified as a no-gang area by 11%, while the presence of a gun incident reduced the likelihood by 15%.

Across all five models, the spatial lag variable is statistically significant. A one unit increase in the spatial lag variable (nearby counts increase from 0 to 1) results in an increase both in the expected count value and the likelihood that the case will contain some degree of gang presence (reduces the likelihood of being in the no-gang group). The Global Moran's I value in all five models is not significant, which means we can accept the null hypothesis that the residuals are not spatially clustered in any of the models.

Table 53: Thiessen polygon regression output, count outcome (n=22,396)

	Model A: home			Model B: incident			Model C: gun		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)									
(Intercept)	-19.40			-19.49			-19.41		
Home	-0.01	0.99	0.09						
Incident				0.02	1.02	0.02			
Gun							1.7E-03	1.00	0.03
Spatial lag	**3.24*** **	**25.44**	**0.34**	**3.24*** **	**25.61**	**0.33**	**3.23*** **	**25.40**	**0.34**
Sigma									
(Intercept)	2.52			2.57			2.52		
Nu (zero)									
(Intercept)	-6.12			-6.00			-5.75		
Home	0.02	1.02	0.07						
Incident				**-0.19*** **	**0.83**	**0.02**			
Gun							**-0.19*** **	**0.82**	**0.02**
Spatial lag	**-11.12*** **	**1.5E-05**	**0.23**	**-11.01*** **	**1.6E-05**	**0.23**	**-11.08*** **	**1.5E-05**	**0.23**
model diagnostics									
df	7			7			7		
Global deviance	4164			4100			4060		
BIC	4234			4170			4130		
Pseudo R^2	0.39			0.39			0.39		
Moran's I	-3.3E-03			-1.1E-03			1.3E-03		
LR test (control)	0.15			**64.01*** **			**104.01*** **		
LR test (full)	**125.17*** **			**61.32*** **			**21.31*** **		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 53: Thiessen polygon regression output, count outcome (n=22,396) continued

	Baseline Model			Full Model		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)						
(Intercept)	-19.40			-19.37		
Home				-0.02	0.98	0.09
Incident				0.02	1.02	0.02
Gun				-0.01	0.99	0.04
Spatial lag	**3.23***	**25.32**	**0.34**	**3.22***	**25.04**	**0.35**
Sigma						
(Intercept)	2.52		1.39	2.51		
Nu (zero)						
(Intercept)	-6.12			-5.75		
Home				0.06	1.06	0.07
Incident				**-0.12***	**0.89**	**0.03**
Gun				**-0.16***	**0.85**	**0.02**
Spatial lag	**-11.11***	**1.5E-05**	**0.23**	**-11.06***	**1.6E-05**	**0.23**
Model diagnostics						
df	5			11		
Global deviance	4164			4039		
BIC	4214			4149		
Pseudo R^2	-3.4E-05			0.39		
Moran's I	-6.9E-04			-0.01		

Note: *=p<.05, **=p<.01, ***=p<.001

Census block group results. At the census block group level, LR tests and BIC comparisons both suggest models with a single independent variable (Models A, B and C) perform better than the Baseline Model. The LR test chi square coefficients are all significant at the .01 level. In line with these findings, BIC values for Models A, B and C are all at least ten units smaller than the Baseline Model (BIC = 1520).

Evidence suggesting the Full Model outperforms the single indicator models, however, is mixed. The LR chi square values say the Full Model does better than Models A, B or C ($p<.01$). level, the BIC comparisons tell a different story: the Full Model (BIC=1462) only does better than Model C, the gun only model. Models A and B and Full perform equally well, after controlling for model complexity.

The pseudo R^2 value are similar across all five models. All five models produced a pseudo R^2 value that ranged from 0.40 (Baseline Model) to 0.44 (Full Model). Models A and B produced an R^2 of 0.43 while Model C produced an R^2 value that was slightly smaller (0.42).

The significance patterns in the *mu* and *nu* sections are consistent across Models A, B. In both of these models, the independent variables link in a positive way in predicting the mean gang count (*mu* section). Each additional gang member living in an area increased the mean predicted gang count by 5%, while each additional gang incident increased the expected mean count by 1%. The gun variable impact, however, was not significant. .

In the *nu* models, each of the three independent variables reduce the likelihood of being in the no-gang group ($p<.01$). Each additional gang member living in an area reduced the likelihood of that area being classified as a no-gang area by 22%. The presence of a gang incident reduced the likelihood by 9%, while the presence of a gun incident reduced the likelihood by 4%.

Significance patterns generated by the Full Model are slightly different from the results of the single indicator models. None of the variables exerted a statistically significant independent effect on the predicted gang count mean (*mu*). Turning to the *nu* section, only the gun variable decreased the odds of being in the no-gang group at the adjusted alpha (.01) level. The unique effect of the gun variable reduced the odds of being in the no-gang group by 2%, even after controlling for the other two variables.

Across all five models, the spatial lag variable was statistically significant. A one unit increase in the spatial lag variable (nearby counts increase from 0 to 1) results in an increase both in the expected count value and the likelihood that the case will contain some degree of gang presence (reduces the likelihood of being in the no-gang group). The Global Moran's I value in all five models is not significant, which means the residuals are not spatially clustered in any of the models.

Table 54: Census block group regression output, count outcome (n=1336)

	Model A: home			Model B: incident			Model C: gun		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)									
(Intercept)	-17.70			-17.68			-17.80		
Home	**0.05****	**1.05**	**0.02**						
Incident				**0.01****	**1.01**	**3.6E-03**			
Gun							0.01	1.01	0.01
Spatial lag	**1.56*****	**4.74**	**0.20**	**1.63*****	**5.10**	**0.20**	**1.67*****	**5.30**	**0.21**
Sigma									
(Intercept)	-0.10			-0.07			0.13		
Nu (zero)									
(Intercept)	-11.04			-11.11			-10.71		
Home	**-0.25*****	**0.78**	**0.03**						
Incident				**-0.09*****	**0.91**	**0.01**			
Gun							**-0.04*****	**0.96**	**0.00**
Spatial lag	**-4.04*****	**0.02**	**0.26**	**-3.96*****	**0.02**	**0.26**	**-4.14*****	**0.02**	**0.26**
Model diagnostics									
df	7			7			7		
Global deviance	1415			1410			1426		
BIC	1465			1460			1476		
Pseudo R^2	0.43			0.43			0.42		
Moran's I	-0.03			-0.02			-0.02		
LR test (control)	**69.07*****			**74.15*****			**58.26*****		
LR test (full)	**32.26*****			**27.18*****			**43.07*****		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 54: Census block group regression output, count outcome (n=1336) continued

	Baseline Model			Full Model		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)						
(Intercept)	-17.55			-17.65		
Home				0.03	1.03	0.03
Incident				0.01	1.01	4.7E-03
Gun				-1.8E-3	1.00	0.01
Spatial lag	**1.63*** **	**5.11**	**0.20**	**1.57*** **	**4.80**	**0.20**
Sigma						
(Intercept)	0.04			-0.14		
Nu (zero)						
(Intercept)	-11.44			-10.68		
Home				-0.12*	0.89	0.04
Incident				-0.04*	0.96	0.02
Gun				**-0.02*** **	**0.98**	**0.01**
Spatial lag	**-4.44**	**0.01**	**0.26**	**-3.88*** **	**0.02**	**0.27**
Model diagnostics						
df	5			11		
Global deviance	1484			1383		
BIC	1520			1462		
Pseudo R^2	0.40			0.44		
Moran's I	-0.02			-4.1E-3		

Note: *=p<.05, **=p<.01, ***=p<.001

Tract results. At the census tract level, results suggest that models with a single independent variable (Models A, Band C) perform better than the Baseline Model. This is evident through the LR tests (null) and BIC comparisons. The LR test chi square coefficients are all significant at the .01 level. In line with these findings, BIC values for all three models are all at least ten units smaller than the Baseline Model (BIC = 675).

Results concerning whether the Full Model outperforms the single indicator. The LR chi square values that compares the Full Model to Models A, B and C is only significant at the .01 level for Models A and C; this means that the Full Model is preferred over Models A and C, but is not preferred over Model B. The BIC comparisons tell a similar story. The BIC for the Full Model (BIC=620) is at least ten units smaller than the BIC produced in Model C which means that the Full Model is preferred over the gun only model after controlling for model complexity. This is not the case, however, for Models A and B. The BIC value for Models A and B is actually smaller than the BIC for the Full Model. This means, that after controlling for model complexity, the single indicators models are actually better than the Full Model. This improvement is only four units for Model A, which is not compelling evidence to select Model A over the Full Model. For Model B, however, the difference in BIC is 14 units, which means the incident only model does significantly better than the Full Model after controlling for model complexity.

The pseudo R^2 values range from 0.45 in the Baseline Model up to 0.56 in the Full Model. Of the single indicator models, Model C had the lowest Pseudo R^2 value (0.50). Models A and B had similar pseudo R^2 values of 0.54 and 0.55, respectively.

The significance patterns in the *mu* and *nu* section are not consistent across Models A, B and C. In each of these models, the independent variables link in a positive way in predicting the mean gang count (*mu* section), but only the home variable reaches statistical significance at the

.01 level. In the *nu* models, however, each of the three independent variables reduce the likelihood of being in the no-gang group and all of the variables reached statistical significance at the .01 level. Each additional gang member living in an area reduced the likelihood of that area being classified as a non-gang area by 19%. The presence of a gang incident reduced the likelihood of that area being classified as a no-gang area by 10%, while the presence of a gun incident reduced the likelihood by 2%.

The significance patterns generated in the Full Model are slightly different than the single indicator models. In the Full Model, none of the variables exerted a statistically significant effect in the *mu* section after controlling for the other two independent variables. Although not significant, the coefficient for the gun variable is actually negative, suggesting additional gun incidents reduce the number of gang territories in an area. Turning to the *nu* section, all three independent variables decreased the odds of being in the no-gang category, but only the incident variable rose to statistical significance at the .01 level. The unique effect of the incident variable reduced the odds of being in the no-gang category by 8%, even after controlling for the other two variables.

Across all five models, the spatial lag variable is statistically significant. A one unit increase in the spatial lag variable (nearby counts increase from 0 to 1) results in an increase both in the expected count value and the likelihood that the case will contain some degree of gang presence (reduces the likelihood of being in the no-gang group). The Global Moran's I value in all five models is not significant, which means the residuals are not spatially clustered in any of the models.

Table 55: Census tract regression output, count outcome (n=384)

	Model A: home			Model B: incident			Model C: home		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)									
(Intercept)	-18.53			-18.48			-18.60		
Home	**0.03****	**1.03**	**0.01**						
Incident				0.01*	1.01	2.1E-03			
Gun							2.3E-03	1.00	2.1E-03
Spatial lag	**1.21*****	**3.35**	**0.19**	**1.28*****	**3.59**	**0.19**	**1.34*****	**3.80**	**0.21**
Sigma									
(Intercept)	-0.66			-0.52			-0.33		
Nu (zero)									
(Intercept)	-12.62			-12.59			-12.53		
Home	**-0.21*****	**0.81**	**0.03**						
Incident				**-0.11*****	**0.90**	**0.02**			
Gun							**-0.02*****	**0.98**	**3.0E-03**
Spatial lag	**-2.22*****	**0.11**	**0.31**	**-1.96*****	**0.14**	**0.31**	**-2.49*****	**0.08**	**0.30**
Model diagnostics									
df	7			7			7		
Global deviance	573			565			606		
BIC	614			606			648		
Pseudo R^2	0.54			0.55			0.50		
Moran's I	0.02			1.4E-03			-0.02		
LR test (null)	**72.98*****			**80.84*****			**39.61*****		
LR test (full)	**17.67****			9.81*			**51.04*****		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 55: Census tract regression output, count outcome (n=384) continued

	Baseline Model			Full Model		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)						
(Intercept)	-18.31			-18.41		
Home				0.02*	1.02	0.01
Incident				2.5E-03	1.00	2.7E-03
Gun				-1.3E-03	1.00	2.3E-03
Spatial lag	**1.31*** **	**3.71**	**0.21**	**1.21*** **	**3.35**	**0.19**
Sigma						
(Intercept)	-0.29			-0.72		
Nu (zero)						
(Intercept)	-13.64			-12.41		
Home				-0.09	0.91	0.05
Incident				**-0.08*** **	**0.92**	**0.02**
Gun				-4.1E-04	1.0E+00	4.4E-03
Spatial lag	**-2.86*** **	**0.06**	**0.29**	**-1.91*** **	**0.15**	**0.31**
Model diagnostics						
df	5			11		
Global deviance	646			555		
BIC	675			620		
Pseudo R^2	0.45			0.56		
Moran's I	3.5E-04			-0.01		

Note: *=p<.05, **=p<.01, ***=p<.001

Neighborhood results. At the neighborhood level, both the LR tests and BIC comparisons suggest models with a single independent variable (Models A, Band C) perform better than the Baseline Model. The LR test chi square coefficients are all significant at the .01 level. In line with these findings, BIC values for all three models are all at least ten units smaller than the Baseline Model (BIC = 444).

Evidence suggesting the Full Model outperforms the single indicator models is mixed. While the LR chi square values that compare the Full Model to Models A, B and C are all significant at the .01 level, the BIC comparisons tell a slightly different story. The BIC for the Full Model (BIC=375) is at least ten units smaller than the BIC produced in Model C, which means that the Full Model is preferred over the gun only model, even after controlling for model complexity. This is not the case, however, for Models A and B. The BIC for the Full Model is only three units smaller than the BIC for Model A, which is not compelling evidence to suggest the Full Model is preferred over Model A. The BIC value for Model B is actually smaller than the BIC for the Full Model. This means, that after controlling for model complexity, the incident only model is actually better than the Full Model. This improvement, however, is only one unit, which is not compelling evidence to select Model B over the Full Model.

The pseudo R^2 values range from 0.23 for the Baseline Model up to 0.59 for the Full Model. The pseudo R^2 value is 0.47 for Model C. Model A and Model B have similar pseudo R^2 values, 0.53 and 0.54 respectively.

The significance patterns in the *mu* section are not consistent across Models A, B and C. While the independent variables link in a positive way in predicting the non-zero values for Models A and B, the coefficient is negative in Model C. This means that the effect of gun incidents reduces the expected gang count. This variable, along with the variables in Model A and B, are not significant at the .01 level.

In the *nu* models, each of the three independent variables reduce the likelihood of being in the zero group in a statistically meaningful way ($p<0.01$). Each additional gang member living in an area reduces the odds of that area being classified as a no-gang area by 14%. The presence of a gang incident reduced the likelihood of that area being classified as a no-gang area by 8%, while the presence of a gun incident reduced the likelihood by 1%.

The significance patterns generated by the Full Model are different from the significance patterns generated in the single indicator models. In the Full Model, only the gun variable exerted a statistically significant effect in the *mu* section after controlling for the other two independent variables. Turning to the *nu* section, the home variable and the incident variable decreased the odds of being in the zero group, but the gun variable had an opposite effect. Only the home and incident variables, however, were significant at the 0.01 level.

Across all five models, the spatial lag variable is statistically significant. A one unit increase in the spatial lag variable (nearby counts increase from 0 to 1) results in an increase both in the expected count value and the likelihood that the case will contain some degree of gang presence (reduces the likelihood of being in the no-gang group). However, it is important to note that the Global Moran's I value in Models A and B were less than 0.05, which means the residuals are spatially clustered in these models. The p value for the Moran's I statistic was 0.01 in Model A and was 0.04 in Model B.

Table 56: Neighborhood regression output, count outcome (n=158)

	Model A: home			Model B: incident			Model C: gun		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)									
(Intercept)	-17.63			-17.68			-17.10		
Home	7.3E-04	1.00	4.2E-03						
Incident				6.1E-04	1.00	1.1E-03			
Gun							-1.2E-03*	1.00	5.3E-04
Spatial lag	**0.58*** **	**1.79**	**0.15**	**0.59*** **	**1.80**	**0.15**	**0.54*** **	**1.72**	**0.13**
Sigma									
(Intercept)	-0.46			-0.48			-0.80		
Nu (zero)									
(Intercept)	-13.16			-13.13			-13.44		
Home	**-0.16*** **	**0.86**	**0.03**						
Incident				**-0.08*** **	**0.92**	**0.01**			
Gun							**-0.01*** **	**0.99**	**1.6E-03**
Spatial lag	**-1.20*** **	**0.30**	**0.30**	**-0.97** **	**0.38**	**0.30**	**-1.34*** **	**0.26**	**0.27**
Model diagnostics									
df	7			7			7		
Global deviance	342			339			361		
BIC	378			374			396		
Pseudo R^2	0.53			0.54			0.47		
Moran's I	**0.12** **			**0.08* **			0.017		
LR test (control)	**75.98*** **			**79.27*** **			**57.64*** **		
LR test (full)	**23.07*** **			**19.78*** **			**41.41*** **		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 56: Neighborhood regression output, count outcome (n=158) continued

	Baseline Model			Full Model		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)						
(Intercept)	-17.59			-17.20		
Home				0.01	1.01	0.01
Incident				2.0E-3	1.00	1.7E-3
Gun				**-2.2E-3****	**1.00**	**6.6E-4**
Spatial lag	**0.58*****	**1.78**	**0.14**	**0.55*****	**1.74**	**0.12**
Sigma						
(Intercept)	-0.46			-1.07		
Nu (zero)						
(Intercept)	-14.74			-13.04		
Home				**-0.22****	**0.80**	**0.08**
Incident				**-0.07****	**0.93**	**0.03**
Gun				0.01*	1.01	0.01
Spatial lag	**-1.61*****	**0.20**	**0.25**	-0.96**	0.38	0.32
Model diagnostics						
df	5			11		
Global deviance	418			319		
BIC	444			375		
Pseudo R^2	0.23			0.59		
Moran's I	-0.07			-0.05		

Note: *=p<.05, **=p<.01, ***=p<.001

Community results. At the community level, evidence suggesting models with a single independent variable (Models A, Band C) perform better than the Baseline Model is mixed. This is evident through the LR tests (control) and BIC comparisons. The LR test chi square coefficients are significant at the .01 level for all three models. In line with these findings, BIC values for Models A and B are all at least ten units smaller than the Baseline Model (BIC = 208). The BIC for Model C, however, is only five units smaller than the Baseline Model and therefore is not preferred after considering added complexity.

Evidence suggesting the Full Model outperforms the single indicator models is also mixed. The LR chi square values that compare the Full Model to Models A, B and C are all significant at the .01 level. However, the BIC comparisons tell a different story. The BIC for the Full Model (BIC=190) is at least ten units smaller than the BIC produced in Model C which means that the Full Model is preferred over the gun only model, even after controlling for model complexity. This is not the case, however, for Models A and B. The BIC value for Models B and C is actually smaller than the BIC for the full model. This means, that after controlling for model complexity, the single indicators models are actually better than the Full Model. This improvement is only five units for Model A and one unit for Model B, which is not compelling evidence to select Models A or B over the Full Model.

The pseudo R^2 values range from 0.21 in the Baseline Model up to 0.68 in the Full Model. Of the single indicator models, Model C had the lowest pseudo R^2 value (0.40), followed by Model B (0.56) and Model A (0.60).

The significance patterns in the *mu* and *nu* section are consistent across Models A, B and C. In each of these models, the independent variables link in a positive way in predicting the non-zero values (*mu* section), but none of them reach statistical significance at the .01 level. In the *nu* models, each of the three independent variables reduce the likelihood of being in the no-gang group. However, none of them reached statistical significance at the .01 level.

Significance patterns generated in the Full Model slightly different than the single indicator models. In the Full Model, none of the variables exerted a statistically significant effect in the *mu* section after controlling for the other two independent variables. Turning to the *nu* section, the incident and the home variables both decreased the odds of being in the zero group, but neither of these variables were statistically significant. The gun variable again worked in the opposite direction, however, this variable was not statistically significant.

Across all five models, the spatial lag variable is not statistically significant. Additionally, the Global Moran’s I value in all five models is not significant, which means the residuals are not spatially clustered in any of the models.

Table 57: Community regression output, count outcome (n=45)

	Model A: home			Model B: incident			Model C: gun		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)									
(Intercept)	-18.63			-18.50			-18.15		
Home	0.01	1.01	4.9E-03						
Incident				1.7E-03	1.00	1.3E-03			
Gun							1.6E-4	1.00	5.0E-04
Spatial lag	0.26	1.30	0.15	0.27	1.31	0.15	0.25	1.29	0.16
Sigma									
(Intercept)	-0.09			-0.04			0.02		
Nu (zero)									
(Intercept)	-10.02			-13.97			-14.63		
Home	-0.19*	0.83	0.08						
Incident				-0.05*	0.96	0.02			
Gun							-0.01*	0.99	2.2E-03
Spatial lag	-0.85	0.43	0.43	-0.44	0.64	0.36	-0.68*	0.51	0.29
Model diagnostics									
df	7			7			7		
Global deviance	158			163			176		
BIC	185			189			203		
Pseudo R^2	0.60			0.56			0.40		
Moran's I	0.07			-0.07			-0.01		
LR test (control)	**30.73*****			**26.45*****			**12.72****		
LR test (full)	**10.39****			**14.68*****			**28.40*****		

Note: *=p<.05, **=p<.01, ***=p<.001

Table 57: Community regression output, count outcome (n=45) continued

	Baseline Model			Full Model		
	Estimate	exp(b)	Std. Error	Estimate	exp(b)	Std. Error
Mu (count)						
(Intercept)	-18.02			-18.47		
Home				0.01	1.01	0.01
Incident				1.1E-3	1.00	1.8E-3
Gun				-5.8E-4	1.00	7.6E-4
Spatial lag	0.25	1.28	0.16	0.27	1.31	0.15
Sigma						
(Intercept)	0.03			-0.14		
Nu (zero)						
(Intercept)	-17.65			189.10		
Home				-11.31	1.2E-05	343.27
Incident				-2.54	0.08	73.40
Gun				0.98	2.67	22.93
Spatial lag	-0.60	0.55	0.23	-37.31	6.2E-17	882.24
Model diagnostics						
df	5			11		
Global deviance	189			148		
BIC	208			190		
Pseudo R^2	0.21			0.68		
Moran's I	-0.09			-0.13		

Note: *=p<.05, **=p<.01, ***=p<.001