ABSTRACT

Title of Document:	LESS DEBATE, MORE ANALYSIS: A META ANALYSIS OF LITERATURE ON BROKEN WINDOWS POLICING
	Michael Distler, M.A., 2011
Directed By:	Associate Professor, Jean McGloin, Criminology and Criminal Justice

In their broken windows thesis, Wilson and Kelling (1982) propose that social and physical disorder leads to a breakdown in informal social controls, thereby allowing more serious crime to occur. This framework had a tangible impact on policy, though research has shown mixed results with regard to its effectiveness. This thesis conducts a meta-analysis of 66 effect sizes, nested within eleven studies, in order to better understand the effect of broken windows policing on crime according to the literature. Results show that broken windows policing does have an effect on crime and that methodological characteristics of the studies are related to the effect. The discussion section considers the relationship between these findings and other meta-analyses on policing innovations, such as hot spots and problem-oriented policing.

LESS DEBATE, MORE ANALYSIS: A META ANALYSIS OF LITERATURE ON
BROKEN WINDOWS POLICING

By

Michael Robert Distler

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Arts
2011

Advisory Committee:
Professor Jean McGloin, Chair
Professor Ray Paternoster
Professor Peter Reuter

Acknowledgements

I would like to start by acknowledging my thesis chair and academic advisor, Dr. Jean McGloin. Without her mentoring and assistance I would not be where I am today. I want to truly thank her for all of her help. I would also like to thank the rest of my committee, Dr. Ray Paternoster and Dr. Peter Reuter. I appreciate Dr. Paternoster, who made me think critically about the utility of meta-analysis and to help me realize the idea of triangulation in order to address any issue in criminology or criminal justice. I also appreciate Dr. Reuter, who made me critically think about my justifications for decisions I made throughout the whole process. I am also grateful to Dr. Travis Pratt, who was very responsive to questions I had regarding the technique of meta-analysis. I would also like to thank the authors of the studies included in the analysis who took the time to send me information I needed to be able to include their work in my analysis. These authors include Dr. Richard Rosenfeld, Dr. Charles Katz, Dr. David Schaefer, Dr. Anthony Braga, Dr. David Weisburd, and Cody Telep. I would also like to thank my family and friends for all of their support throughout the whole process. Finally, I am especially grateful to Mariel Alper. It was from her that I learned how to better structure my time and work towards completion of this document. I truly appreciate her love and support over the last six months.

Table of Contents

List of Tables

Chapter 1: Introduction

Almost thirty years ago, Wilson and Kelling (1982) offered a new thesis on crime that had a pervasive and tangible impact on policing policy. Their concept of "broken windows" (BW) proposed that social and physical disorder indirectly lead to serious crimes through a neighborhood breakdown of informal social controls. The implication of this perspective for police was to focus on low–level, misdemeanor offenses in order to prevent serious crime and improve quality of life. Soon thereafter, William Bratton adopted this perspective in the New York City subway system as the head of the Metropolitan Transportation Authority (MTA) and then imported it to the New York City Police Department (NYPD) when he became commissioner in the mid-1990s. He credited BW as the primary reason for the crime drop (Kelling & Bratton, 1998). Between the years of 1989 and 1998, arrests for violent crimes dropped from just under 150,000 to around 75,000, nearly a 50% decline (Kelling & Sousa, 2001). The homicide rate fell even more dramatically, demonstrating a 70% drop during the same time period (Kelling & Sousa, 2001). With the supposed success in the New York City subway and the early indications of successful crime reduction in the city as a whole, police departments in cities such as San Francisco, Boston, and Chicago turned attention towards BW policing and integrated it as part of their tactical strategies to reduce crime (Kelling & Coles, 1996).

Researchers remain unconvinced that this pervasive adoption was wise, however. Many studies have investigated the effectiveness of BW policing strategies in New York City and elsewhere, with overall results that are mixed (Braga et al., 1999; Braga & Bond, 2008; Corman & Mocan, 2005; Golub, Johnson, Taylor, &

Eterno, 2003; Golub, Johnson, Taylor, & Eterno, 2004; Harcourt & Ludwig, 2006; Harcourt & Ludwig, 2007; Katz, Webb, & Schaefer, 2001; Kelling & Sousa, 2001; Messner et al., 2007; Novak, Hartman, Holsinger, & Turner, 1999; Rosenfeld, Fornango, & Rengifo, 2007). According to some research, there appears to be a disjoint between the impact of BW policing and its continued and enthusiastic use among police departments. Many scholars have additionally claimed that BW leads to racism, police brutality, and attacks against the poor (discussed in Sousa & Kelling, 2006; see also Harcourt, 1998; Roberts, 1999). With arguments and studies falling on both sides of this debate, it appears that the empirical standing of BW as a policing strategy remains unclear. Can the implementation of a BW policing strategy, which involves a more focused approach to disorderly behavior and conditions, effectively lower crime rates?

In order to address this question, this thesis will evaluate the efficacy of BW policing through meta-analysis. Meta-analysis has been used more frequently in evaluations of criminological policies and theories (Pratt, 2002; Pratt & Cullen, 2000; Pratt, McGloin, & Fearn, 2006; MacKenzie, 2006). This technique is arguably a better indicator of the effectiveness of BW policing when compared to a literature review because it allows for objective and precise measurements of the effectiveness and significance of the policing strategy on a reduction in crime. Specifically, the benefits of evaluating the research in this manner include: (1) the ability to account for all studies of BW policing thus far; (2) an objective assessment of the effect size of BW policing on crime; and, (3) a determination of whether methodological variation in the studies is systematically related to the strategy's effectiveness.

Chapter 2 turns attention to the definition of BW as well as criticisms and empirical arguments about the effectiveness of the policing strategy. In Chapter 3 the technique of meta-analysis is described, including both strengths and weaknesses, and the inclusion/exclusion process is also discussed. Chapter 4 turns to the data used and results of the meta-analysis. Finally, Chapter 5 is a conclusion that provides a discussion of the policy implications and thoughts on future research.

Chapter 2: Literature Review

Wilson and Kelling (1982) root their theoretical notions of BW in the evaluation of a police foot patrol program implemented in Newark, NJ in the mid-1970s. This program, sparked by the Safe and Clean Neighborhoods Act, awarded funding to cities that empirically tested the effectiveness of foot patrol (Kelling, Pate, Ferrara, Utne, & Brown, 1981). The researchers used official statistics to test the viability of the strategy in reducing crime while also collecting data from surveys on victimization, residents' fear of victimization, residents' perceptions of the police, and police job satisfaction (Kelling et al., 1981). The results showed that, although official crime statistics were not affected by foot patrol, residents were less fearful and took less self-protective measures in foot-patrolled neighborhoods. Police who patrolled the streets also reported more positive job satisfaction when compared with officers who patrolled in cars (Kelling et al., 1981).

This work cast further doubt on other policing tactics which were empirically tested during the late 1960s and early 1970s. Randomized preventive car patrol and rapid response were the tactical foundations of the professional model of policing, yet evaluations showed them to be ineffective. In the evaluation of randomized patrol, findings revealed that more police patrol did not have an effect on crime rates, and more importantly, citizens did not even notice a difference in the level of patrol in their neighborhood (Kelling, Pate, Dieckman, & Brown, 1974). As for rapid response, results showed that the determining factor in whether a criminal is caught is not how fast the police can respond to a call for service, but rather how long citizens wait

before contacting the police, if the police are contacted at all (Kansas City Police Department, 1980; see also Spelman & Brown, 1981).

Moreover, the use of these tactics highlighted a clear disjoint between the community and the police that had taken root. Police-citizen interactions under the professional model of policing were seen as a business-client relationship and the police tried to avoid any attachment to the complainant (Kelling & Moore, 1988). The Newark Foot Patrol Experiment stood apart because it showed that unlike these other tactics, citizens noticed police who patrolled on foot and their presence made the citizens feel safer (Kelling et al., 1981).

George Kelling was one of the researchers involved in the Newark Foot Patrol study. Upon revisiting his notes from the evaluation, he was drawn to the interactions he observed between the officers and the citizens. These interactions were mentioned frequently in the BW article (Wilson & Kelling, 1982) because Kelling saw that on these foot-patrolled beats there were informal neighborhood rules of conduct that were understood by both citizens and officers. The results are similar to those noted by Egon Bittner (1967) in his qualitative study of "skid row". In his work, Bittner focused on peacekeeping, the task on which officers spent the majority of their time (Bittner, 1967). It was the officers' freedom to act and their ability to use discretion that was important on these difficult beats. These mostly white officers, working with mostly black citizens or street people, focused on managing low-level criminal behaviors and maintaining order (Bittner 1967). Police were not just law enforcement officials, but were called upon to mediate a broad range of issues and problems. Bittner recognized that by dealing with these issues early on through citizen

collaboration, it promoted safety and reduced the likelihood of further disputes which could then lead to more serious crime problems (Bittner, 1967). It was interesting to note that even though the race/ethnicity of the officers and neighborhoods differed, they were still able to work together to keep the peace in the area. In short, the discretion given to these officers, their intimate knowledge of their beats, and their attention to relatively "minor" issues (such as disorder) on skid row allowed them to manage these areas effectively.

From this previous work as well as the foot patrol experiences in Newark, Wilson and Kelling (1982; see also Kelling & Coles, 1996) argued that untended or disorderly behavior leads to a breakdown in community controls. This process, which they defined as urban decay, begins when physical and social disorder starts to crop up in the neighborhood. Abandoned property, panhandlers, rowdy teenagers, and litter are just a few of the problems that Wilson and Kelling (1982) describe. Citizens recognize that the informal social controls of the neighborhood are breaking down and they begin to withdraw from the neighborhood and engage in other self-protective behaviors (e.g., not leaving their house at night). Many families move out of the neighborhood, allowing those to move in who are not bound by the norms and rules of the neighborhood, leading to further breakdowns in informal social controls. Interactions between the police and citizens are also weakened. Citizens feel the police are not doing their jobs while the police are tired of listening to citizens' complaints. This is the situation where violent and serious crime is allowed to breed, because as 'orderly' citizens withdraw from public spaces within the neighborhood, criminals begin to claim these spaces as their own (Wilson & Kelling, 1982).

The metaphor goes that if one broken window is left untended, it will signal to criminals that neither the police nor the citizens care. This was first articulated by Stanford psychologist Phillip Zimbardo in 1969 when he conducted an experiment of human behavior. Leaving an automobile with its hood up in the Bronx, the car was vandalized after only ten minutes and was stripped of its parts soon thereafter (Wilson & Kelling, 1982). A similar automobile was left in Palo Alto, California and was left untouched for a week. However, once Zimbardo broke a window of the car, other people engaged in vandalism until it was fully destroyed. This experiment showed that once the level of informal controls was reduced–once one window was broken and left as such–people came to believe that disorderly behavior was acceptable (Wilson & Kelling, 1982).

An important aspect of BW policing is the focus on returning discretion to the patrol officers and allowing that their functions extend beyond enforcing laws, as articulated by Wilson and Kelling (1982). Discretion is seen as an important tool allowing officers to better address the important crime problems in the neighborhood. Through the advent of returning officers to patrolling the streets, officers would be given greater discretion in their attempts to keep the peace in their neighborhood.

It is in the decaying neighborhoods where citizens often fear disorder and lower-level offenses and these fears impact their behaviors (Wilson and Kelling, 1982). Therefore, officers should focus on the behaviors and problems that are harmful to the order of the neighborhood. Citizens should also work with the officers to help maintain order and enforce the rules and norms of the area (Wilson and Kelling, 1982). It is also important to understand that not all neighborhoods deal with

the exact same problems and the rules and norms need to be tailored to each individual neighborhood. Through the techniques of problem solving and a focus on minor offenses, Wilson and Kelling (1982) claim citizens and the police can coexist and keep order in the neighborhood. Thus, Wilson and Kelling (1982) introduce an idea which combines innovations that were taking hold in police departments across the country, such as problem-oriented policing and community policing, and adding to it this idea of a focus on minor offenses.

The first explicit attempt to implement these claims as a policing strategy was in the New York City subway system at the beginning of the 1990s. In the late 1980s, serious problems festered under the streets of New York City. Citizens feared entering the subway system, and its use began to decline. Problems in the subway of course included robbery, assaults, thefts, injuries and the like, but also graffiti on the subway cars and panhandling in the subway stations (Kelling & Coles, 1996). The initiative began with the Clean Car Program, where the NYC-MTA used problem-solving tactics in order to address the graffiti problem on subway cars (Kelling & Coles, 1996). After some research, they concluded that graffiti was a result of artists wanting to see their 'tags' on the subway cars. In response, the NYC-MTA pulled graffiti ridden cars and refused to return them to service until they were cleaned. This solved the graffiti problem, since artists no longer found benefit to tagging the clean cars (Kelling & Coles, 1996).

The NYC-MTA then moved on to disorderly behavior in the subway stations. The most serious of these problems included fare-beating and panhandling. Once again using a problem-solving technique, officials found that fare-beating could be

eliminated by stopping those who jump turnstiles and following through on punishing this behavior. As for panhandling, officials first tried to offer services and shelters to the homeless. This effort was unsuccessful and a campaign was then designed notifying patrons that giving money to panhandlers was harmful to their well-being because it allowed them to put off searching for employment. On top of this effort, arrests were also made in the subways for those who were panhandling (Kelling & Coles, 1996). After completion of the planning and initiation stages, it took only a few weeks before the subway stations became markedly safer; fare-beating and panhandling were all but eliminated. As a result of these initiatives, crime in the subway declined, citizens returned to the New York City subway system, and it is now viewed as one of the cleanest and safest in the country (Kelling & Coles, 1996). Also important was that officers noticed people whom they arrested for minor crimes such as fare-beating tended to be felons with outstanding warrants, or people carrying illegally concealed weapons (Kelling & Coles, 1996).

With the success in the subway, BW policing began to trickle out to the streets. It took a concerted effort from newly elected Mayor Rudolph Giuliani and newly appointed Police Commissioner William Bratton (who also led the NYC-MTA initiative) to implement policies that attempted to reduce disorderly behavior with the hope that they could also reduce more serious and violent crime (Kelling & Coles, 1996). After the successes experienced in the subway system, Bratton implemented similar policies in the NYPD. The NYPD started to focus their attention on panhandling in the streets and, more successfully, on squeegee men (Kelling & Coles, 1996).

With the supposed early success seen in New York City, many other police departments across the country took notice and began implementing similar policies. Departments such as Boston, San Francisco, and Chicago followed the New York City strategy (Kelling & Coles, 1996). Given the increased acceptance of this policy, many criminologists honed in on this area for empirical inquiry. Researchers also began to assess the tenants of the BW framework in order to analyze whether disorder does breed more serious crime.

Criticisms of BW Policing

With the apparent success of BW policing in New York City, scholars began to criticize the strategy on many levels. Specifically, these criticisms tended to focus on conceptual and philosophical issues, arguing that even if BW policing did reduce crime, it may not be "worth it." The primary critic of BW has been Bernard Harcourt. Harcourt has criticized BW theoretically (Harcourt, 1998) but has also conducted empirical analysis on the policing strategy (Harcourt & Ludwig, 2006, 2007). Harcourt's main critiques of BW policing include: this form of policing led to the creation of a disorderly class of people (Harcourt, 1998), racism (Harcourt, 1998; see also Roberts, 1999; Sousa & Kelling, 2006), mass surveillance (Harcourt, 1998), and police brutality (Harcourt, 1998).

Harcourt (1998) believed that BW policing socially constructed a new class of people that are defined as disorderly. According to the policing strategy, the definition of what is orderly and what is disorderly is left up to the officers on the beats. Harcourt (1998; see also Roberts, 1999) claimed that this definition was constructed by the policing strategy itself rather than a definition of disorderly agreed

upon by the community. This gives police too much discretion regarding who to arrest and "rough up." In essence, this allows police to focus on status rather than behavior, which can result in racist actions and brutality (Harcourt, 1998). Evidence shows that at the decision to arrest misdemeanor offenders, blacks are overrepresented compared to their proportion in the population (Harcourt, 1998). Another critic, Roberts (1999), claims the vagueness of some of the laws of order-maintenance policing affects minorities, especially those living in poorer neighborhoods. According to Roberts (1999), giving police the discretion to define disorderly behavior and then arrest for that behavior tends to affect minorities more than whites. She takes the example of a gang loitering law in Chicago which allowed for police to arrest groups that were congregating and seemed suspicious (Roberts, 1999). According to Roberts, a group of Black males congregating may seem more 'suspicious' to the police than a similar group of white males. She argues that the distinction between the orderly and disorderly is racialized, therefore leading to the enforcement of order as being racialized (Roberts, 1999). She concludes that this form of policing is used at the expense of Black residents' freedoms (Roberts, 1999; see also Harcourt, 1998).

Mass surveillance of the 'disorderly' is another problem highlighted by Harcourt (1998). He argues that this type of strategy allows the police to take into custody those people who appear 'suspicious', solely to run checks for outstanding warrants (Harcourt, 1998). This can also be thought of more negatively as police keeping mass surveillance on certain groups of people (minorities, juveniles) which can lead to problems of racism (discussed above) and police brutality (discussed

11

below). Harcourt does not necessarily see mass surveillance and checks of outstanding warrants as a bad strategy. He simply states that the police should be explicit in their true reasoning behind the use of these tactics (Harcourt, 1998).

Harcourt also discusses the increased police aggressiveness or increased citizen complaints of police brutality in response to the implementation of a BW strategy (Harcourt, 1998; see also Roberts, 1999; Herbert, 2001, Sousa & Kelling, 2006). An increase in police citizen encounters would be expected with the implementation of a BW policing strategy. These encounters would not necessarily be positive because when officers are given the discretion to go after those who seem 'suspicious', they come in contact with a broader range of citizens, according to Harcourt (1998; see also Roberts, 1999). Most of these citizens have done nothing wrong and they become frustrated with the police and given the relatively minor nature of the offense, these citizens also believe that the officer's aggression is unwarranted. Harcourt (1998) believes it is important to continue to pay attention to these problems, as it could be that they outweigh the benefits from these types of policing strategies. Although these are important problems that must be addressed if a strategy such as this is to be implemented, it is first important to disentangle whether the strategy is effective in reducing crime. After all, if the policy does not meet its primary goal, then it should not continue. In order to address this concern, it is important to look at the empirical research done in relation to BW policing and its effects on serious crime.

Empirical Tests of BW Policing as it Relates to Serious Crime

Given the self-declared success in New York City, many other police departments across the country implemented policies based on broken windows (Kelling & Coles, 1996). In response to the growing presence of these policies throughout the country, scholars began empirically assessing the effectiveness of the strategy. Speaking generally, results from these studies have been mixed. Some investigations have shown these policies to be effective at reducing serious crime, while others have found that BW policing does not affect the crime rate. Many of these studies have focused on the New York City crime drop during the 1990's, (Corman & Mocan, 2005; Golub et al., 2003; Golub et al., 2004; Golub et al., 2007; Harcourt & Ludwig, 2006; Harcourt & Ludwig, 2007; Kelling & Sousa, 2001; Messner et al., 2007; Rosenfeld et al., 2007). However, there has been some research conducted in other locations such as Chandler, Arizona (Katz et al., 2001), Lowell, Massachusetts (Braga & Bond, 2008), Jersey City, New Jersey (Braga et al., 1999), Texas (Jang, Hoover, & Lawton, 2008), and a 'major Midwestern industrial city' (Novak et al., 1999). Research is further distinguished by whether the study used an experimental design or whether a quasi-experimental design was employed. Looking at these studies will give us an idea on the standing of the literature regarding BW policing.

Kelling and Sousa (2001) evaluated the crime drop in New York City from 1989-1998, taking account of several factors likely influential in this decline. Specifically, they investigated the change in policing strategy (BW policing), drug trends, economic changes, and demographic changes with regard to changes on an

index of official homicides, rapes, felonious assaults, and robberies (Kelling & Sousa, 2001). Kelling and Sousa observed that precincts vary across multiple demographic factors, allowing them to compare each precinct with each other, as if each were a separate city. Every precinct received the same treatment from the police, so if the strategy was effective, it should emerge 'across the board.' Results showed that the policing strategy significantly reduced the amount of violent crime, even when controlling for the other variables, such as economic, demographic, and drug trends, which either had effects that were not significant or a significant effect opposite the theoretical expectation (Kelling & Sousa, 2001). This initially promising picture was complicated a few years later, however. Harcourt and Ludwig (2006) re-analyzed Kelling & Sousa's (2001) data because they believed that mean reversion[1] caused the dramatic results witnessed in New York City. Harcourt and Ludwig (2006) changed the modeling strategy that Kelling and Sousa (2001) used in order to account for mean reversion, and argued that this was the likely cause of the findings, not an effective policing strategy (Harcourt & Ludwig, 2006).

Building upon the work done by Kelling and Sousa, two other studies showed moderate support for the BW policing strategy in New York City. Rosenfeld, Fornango, and Rengifo (2007) addressed important limitations such as the use of the aggregate violent crime index, omissions of ordinance-violation arrests and complaints of disorder, ignoring the possibility of simultaneity, exclusion of imprisonment rates as influences on violent crime, and controls for spatial autocorrelation. They found that there was, at best, a modest impact of order-

---

[1] Mean reversion refers to the idea that those precincts which experienced the largest increases in crime during the late 1980s also saw the largest declines in crime thereafter. Those crime-ridden precincts also received the most BW policing during the 1990s (Harcourt & Ludwig, 2006).

maintenance policing on New York City crime reductions (Rosenfeld et al., 2007). Messner et al. (2007) pitted what they claim to be the two leading hypotheses for the crime drop in New York City against each other–BW policing and the decline in the crack-cocaine market–while also comparing the results for both gun-related homicides and non-gun-related homicides. Results showed that misdemeanor arrests and toxicity reports were statistically significant predictors for a reduction in gun-related homicides. However, misdemeanor arrests and toxicity reports where not statistically significant predictors for non-gun-related homicides, which began to decrease prior to BW policing and the decline in crack-cocaine use (Messner et al., 2007).

Looking at crime in New York City in a different fashion, Corman and Mocan (2005) analyzed monthly time series data from 1974-1999 to assess the effect of BW policing on serious crime. They used complaints for all Part 1 index crimes with the exception of arson. Economic conditions were measured using the unemployment rate and real minimum wage while felony arrests, the number of uniformed officers, and the prison population were also included in the analysis. Results showed that net of controls, misdemeanor arrests had a significant effect on motor vehicle theft, robbery and grand larceny. They concluded that the BW policing strategy did have a significant effect on certain crimes, and the strategy was more important than were the economic conditions (Corman & Mocan, 2005).

In an analysis of an initiative in Chandler, Arizona, Katz, Webb, and Schaefer (2001) also looked at the impact of Quality-of-Life (QOL) policing on serious crime through an interrupted time series. The initiative, known as Operation Restoration,

worked in cycles in four different areas located within the Chandler Redevelopment District. Only one area at a time received the treatment as the team moved from one area to the next. Two teams were developed for the project: a Neighborhood Service Unit that dealt with enforcing city code violations, and a Neighborhood Response Team of sworn police officers in charge of enforcing municipal codes and county laws (Katz et al., 2001). All four areas received two interventions, with a three month break where no areas were receiving the treatment. Results showed support for public morals and physical disorder calls across all four zones. Interestingly, one zone had many significant results opposite of the other three zones. The researchers speculated a dosage problem, whereby the units were worn out upon arriving at the last zone (Katz et al., 2001).

In Jersey City, New Jersey, Braga et al. (1999) conducted a randomized experiment on policing disorder in crime hot spots. This was an influential study because it showed how problem-oriented policing, hot spots, and BW policing could work in concert to affect crime rates. Twenty-four pre-determined hot spots were paired and officers were informed of which areas were to receive increased police attention and which were not. Police assessed the problems in these areas and the responses that were created were categorized as a 'policing disorder' type of strategy (Braga et al., 1999). Results showed that all incidents and calls had a negative relationship for treatment groups, with some effects being significant. They also found no evidence of either displacement or diffusion. The researchers concluded that these types of focused strategies are successful in cleaning up the area of disorder and also positively affecting the crime rates (Braga et al., 1999). Almost ten years later,

Braga and Bond (2008) conducted a similar study in Lowell, Massachusetts, finding similar results.

Novak et al. (1999) examined the effects of aggressive policing of disorder on serious crime. The study was conducted in an unnamed large Midwestern city and focused on a crackdown on disorder in one part of the city. The researchers used an interrupted time series (similar to Corman and Mocan, 2005) looking at the total number of crimes reported to the police during the fifty two weeks before and after the four week impact period (Novak et al., 1999). Results showed that aggressive enforcement of disorder does not have a significant effect on the number of crimes reported to the police. The researchers acknowledged that the lack of media coverage for the crackdown and the limited length of the crackdown might have affected the overall results (Novak et al., 1999).

These studies provide a sample of empirical research on BW policing and point to a general observation about this literature: the results do not clearly answer whether BW "works". This generic review of the literature does not give a clear picture as to whether BW policing is an effective strategy – instead, some objective means of balancing and analyzing the literature would be ideal. Under such circumstances, meta-analysis emerges as an appropriate technique.

Chapter 3: Data and Methods

Inclusion/Exclusion Criteria

Developing an explicit set of inclusion criteria is a crucial step in the meta-analysis process. The criteria must be explicitly stated so that any decisions are fully understood and can be replicated. Wilson and Lipsey (2001) discuss a wide range of what they term 'study eligibility' criteria. All of these criteria must be articulated before the population of studies can be gathered. With regard to this meta-analysis of BW policing, guidelines were established to differentiate studies that would be included as opposed to those that would not.

The first guiding factor was how the studies would be retrieved. Studies would be retrieved through an exhaustive search using electronic databases available from the University of Maryland's Library Services.[2] The chosen databases included Criminal Justice Abstracts, Criminal Justice Periodicals, National Criminal Justice Reference Service (NCJRS), Social Science Citations Index (SSCI), and ProQuest: Dissertations & Theses. Google Scholar was also searched for other academic research of BW policing. The first four databases and Google Scholar cover an extensive amount of different literature that has been published in peer-reviewed journals, trade journals, books, or other outlets. The ProQuest database allows for a search of studies that have not necessarily made it through to publication.

Once the databases were chosen, the next step was determining how to search for the BW policing literature. BW policing is too narrow a term to be able to locate all studies that assess initiatives based on Wilson and Kelling's (1982) ideals.

---

[2] Searches for eligible studies were conducted through March 16th, 2011.

Therefore, the search terms was expanded to include not just 'broken windows' but to also include 'quality-of-life,' 'order-maintenance,' and 'zero tolerance.' Using the 'and' function in the searches, 'policing' was also included to restrict the results to only those articles that talked about BW in a context of policing. This would exclude research that tested the theoretical underpinnings of BW.

Searches within the six databases with four different variations in the search terms yielded over one thousand results. Many of the results were book reviews and articles that were not empirical. The most important factor for inclusion was that it had to involve the empirical study of a BW policing strategy. Although precautions were already taken through the earlier process, each study determined to be a possible candidate for inclusion was read through fully to assess whether the research was empirically studying a BW policing strategy. As defined in the previous section, a BW policing strategy should follow the ideals of Wilson and Kelling (1982) and include some type of police action specifically taken against minor and/or disorderly behavior.[3] This strategy also had to be articulated by the police department where the evaluation was taking place.

Another key feature of the study that was important for inclusion into the meta-analysis was that the study had to assess the effects of the BW policing strategy on crime. This thesis hopes to address the overall impact of BW policing on crime[4], and therefore any other outcomes from the strategy are not considered. Only studies that used official crime outcomes were included in the analysis. Some studies

---

[3] From the outset, studies that analyzed zero-tolerance policing were going to be excluded from the analysis because they do not have second-order agreement with the tenets of BW. However, during the search for eligible studies, articles empirically assessing zero-tolerance were not found.

[4] Crime is defined as falling under violent or property crime. Outcomes that focused on drug crimes or other victimless crimes were excluded.

included in the analysis evaluated an index of serious and violent crimes (Corman & Mocan, 2005; Harcourt & Ludwig, 2006; Kelling & Sousa, 2001; Rosenfeld et al., 2007) while other studies focused on individual crimes such as murder (Messner et al., 2007). Because only official crime outcomes were included, surveys were excluded from the analysis. Surveys most often dealt with citizens' perceptions of the crime rates or how safe they feel in their neighborhoods and did not address whether crime actually decreased, which would result in wide heterogeneity in the meaning of effect estimates. Although such research is important to understanding the efficacy of the BW policing strategy, official statistics are the most reasonable place to start an assessment, given that it is often the "bottom line" in determining the effectiveness of police action. Furthermore, from a pragmatic view, there have been very few surveys which have assessed the effects of BW policing, with only three such studies utilizing this method (Golub et al., 2003; Golub et al., 2004; Golub et al., 2007).

A final issue that must be addressed is that the study must be reported in English. If the study was not reported in English, then it was not included in the analysis. Once these criteria were established, candidate studies were read through fully to assess whether the criteria were met. Upon completion of the process, the population included eleven studies that fit the criteria.[5]

Some critics might argue that eleven studies are too few to warrant a meta-analysis. Though there are only eleven studies, these studies do produce a large number of effect sizes, a total of 66, which are the units of analysis for this study and

---

[5] The eleven studies that were included in the analysis are: Braga & Bond, 2008; Braga et al., 1999; Corman & Mocan, 2005; Harcourt & Ludwig, 2006; Harcourt & Ludwig, 2007; Katz et al., 2001; Kelling & Sousa, 2001; Messner et al., 2007; Novak et al., 1999; Rosenfeld et al., 2007, Wagers, 2007.

greatly expand the number of "tests" of BW policing. Furthermore, other meta-analyses focused on policing found benefits in studying similar numbers of studies. For instance, Braga conducted a meta-analysis on the effects of hot spots policing on crime using nine studies (Braga, 2007), and Weisburd completed a meta-analysis on the effects of problem-oriented policing on crime and disorder using only ten studies (Weisburd at al., 2008). Given the use of meta-analysis with these policing strategies and the large number of effect sizes obtained from the eleven studies, use of meta-analysis to analyze BW policing seems appropriate.

Choosing Effect Sizes

After following the inclusion/exclusion criteria and conducting an exhaustive search to find all studies that should be included, the next requirement is to read through the studies to be able to successfully select the effects that will be used in the meta-analysis. In meta-analysis, an effect size estimate is the index used to represent the findings from the different studies (Lipsey & Wilson, 2001). The findings are chosen from the different studies to represent the relationship being analyzed; in this case, the relationship between BW policing and serious crime. Only statistics analyzing this relationship are included in the meta-analysis.

Each study included in the analysis needed to have enough statistical information to estimate the effect sizes. The statistical effects were first converted into a standardized effect size to allow for analysis and comparisons across the different studies. Formulae were used to convert the test statistics into the standardized correlation coefficient r. The standardized correlation coefficient was chosen over other effect size estimates, such as Cohen's d, for three main reasons: (1)

its ease of interpretation; (2) availability of formulae to convert the other test statistics used in the BW policing studies into an r; (3) the ability to use the Fisher's r to z transformation[6] (Wolf, 1986) to convert each r score into a z(r) score. Converting the effect size estimates to a z(r) score allows for use of the z-distribution, which is assumed to approach normality. All studies in this analysis used either a t-statistic[7] or included information on the beta coefficient and standard error so that a t-statistic could be calculated.[8] Once the statistics are converted into a standardized effect size, analysis will be conducted to test for the magnitude of the relationship between BW policing and serious crime.

A concern when drawing effects from the studies is whether multiple effect sizes should be drawn from the same study or whether only one effect size from each study should be used in the analysis (for a discussion of this concern see also Wolf, 1986). Lipsey and Wilson (2001) refer to this as the independence of effect size estimates. They conclude that only one effect size should be used because if more effect sizes are drawn from a single study, it will put more weight on that one study in the analysis (Lipsey & Wilson, 2001). The problem with this approach is the choice about which effect size estimate to use is left up to the researcher. It creates subjectivity in an analysis that is meant to be objective. Therefore, this thesis will follow the approach taken in other meta-analyses (Pratt & Cullen, 2000, Pratt et al., 2006) and will use multiple effect sizes from each individual study, while also

---

[6] $ES_{z_r} = .5 log_e \left[\frac{1+r}{1-r} \right]$ (Lipsey & Wilson, 2001).

[7] $r = \sqrt{\frac{t^2}{t^2 + df}}$ (Wolf, 1986).

[8] For those studies that did not provide enough statistical information, the authors were contacted and the necessary information was collected.

accounting for nesting within studies (i.e., that effect sizes are clustered within studies). Specifically, the current analysis was based on 66 effect sizes, nested within eleven studies. Table 1 indicates the studies included in the analysis and the number of effect size estimates drawn from each study.

Table 1: Studies Included in Meta-Analysis and Effect Size Estimates

Study	# of Effect Size Estimates
Braga and Bond (2008)	5
Braga et al. (1999)	8
Corman and Mocan (2005)	7
Harcourt and Ludwig (2006)	2
Harcourt and Ludwig (2007)	2
Katz et al. (2001)	30
Kelling and Sousa (2001)	1
Messner et al. (2007)	4
Novak et al. (1999)	2
Rosenfeld et al. (2007)	4
Wagers (2007)	1

Fixed Effects vs. Random Effects Modeling

The fixed effects model assumes that the effect size estimates from the BW policing studies are drawn from a single population and the differences noted between studies is associated with random error (Lipsey & Wilson, 2001; Wolf, 1986). A random effects model[9] is suggested if it is believed that the random differences are due to differences in procedures or settings or if it is assumed that the set of studies represent a sample rather than the entire population of studies on BW policing (Lipsey & Wilson, 2001). After an exhaustive review of the BW policing literature, it

---

[9] The equation for the random effects model is similar to that for the fixed effects model except for the inclusion of an estimate of the random effects variance (Lipsey & Wilson, 2001). The equation for the weighted random effects model is: $w = \frac{1}{SE_i^2 + \hat{v}_\theta}$ (Lipsey & Wilson, 2001). The equation for the estimate of v_θ is: $v_\theta = \frac{Q - (k-1)}{\sum w_i - (\frac{\sum w_i^2}{\sum w_i})}$ (Lipsey & Wilson, 2001).

is assumed that the chosen studies represent the totality of studies conducted on BW policing strategies.

However, a Q-test will be used to test for homogeneity of the effect size distribution (to test for whether a fixed effects model is appropriate). The Q statistic is distributed as a chi square with $k^{10} - 1$ degrees of freedom (Lipsey & Wilson, 2001). The computational formula for Q is given below:

$$Q = (\sum w_i ES_i^2) - \frac{(\sum w_i ES_i^2)}{\sum w_i},$$

where w_i is the inverse variance weight of the effect sizes and ES_i is the individual effect size (Lipsey & Wilson, 2001). If Q exceeds the critical value, the null hypothesis of homogeneity is rejected and a random effects model would be appropriate for this analysis.

Weighting of the overall effect size is also an issue discussed in the literature (Lipsey & Wilson, 2001; Wolf, 1986) and the technique has been used in previous meta-analyses in criminology (Cullen & Pratt, 2000; Pratt et al., 2006). The procedure of weighting takes into account the fact that these different effect sizes were extracted from studies with different sample sizes and differing variances. It allows for the understanding that not all studies have equal value in the analysis.

Lipsey and Wilson (2001) discuss two different weighting procedures, the first involving weighting each effect size by its sample size and the second procedure involves weighting the different effect size by the inverse of the variance. In weighting by sample size, the effect size is first multiplied by the sample size, all of the effect sizes are summed, and then this number is divided by the sum of the sample

---

[10] k stands for the number of effect sizes (Lipsey & Wilson, 2001).

sizes.[11] Lipsey and Wilson (2001) argue that this is a simple approach to weighting studies and it would instead be more appropriate to weight studies by the inverse variance. Weighting by the inverse variance involves the extra step of calculating the inverse variance.[12] Then the procedure is similar to that used for calculating the sample size weighted effect size. These two weighting techniques allow for a greater emphasis to be placed on effect sizes that were calculated from larger samples, thereby creating a more informative estimate (Lipsey & Wilson, 2001).

Methodological Control Variables

Beyond finding and converting effect size estimates for each study, coding of a number of other important distinguishing factors among studies is an important process. There could be various explanations accounting for the differences noted between effect sizes. By coding characteristics of each study, it allows for an assessment of the effects of the differences in the location and methods deployed. For each study in the analysis a number of important variables were coded.

The location of the study is an important factor to consider because of the great focus given by scholars to New York City with regard to BW policing. In short, there could be a New York effect, where the city is different from other cities in some unmeasured way. It is also possible that Bratton's familiarity with BW and his use of it early on in the NYC-MTA could have led to fidelity of treatment, which needs to be accounted for in the analysis. Of the 66 effect sizes examined, 20 came from studies conducted in New York, and will receive a value of '1' on this measure

---

[11] The formula for this procedure is: $\overline{ES} = \frac{\sum(n \cdot ES_i)}{\sum(n)}$ (Lipsey & Wilson, 2001).

[12] The inverse variance weight for Fisher's z(r) is calculated with the following formula: w = n-3 (Lipsey & Wilson, 2001). The formula for the procedure is: $\overline{ES} = \frac{\sum(w \cdot ES_i)}{\sum(w)}$ (Lipsey & Wilson, 2001).

(Corman & Mocan, 2005; Harcourt & Ludwig, 2006; Harcourt & Ludwig, 2007; Kelling & Sousa, 2001; Messner et al., 2007; Rosenfeld et al., 2007).

Research design for each of the different studies could also affect the overall results of the analysis. There is a growing endorsement in the field of criminology and criminal justice of randomized controlled experiments as the "gold standard" to test behavior (Farrington, 2003; Shadish, Cook, And Campbell, 2002; Sherman, 2010; Sherman et al., 1998; Weisburd, Lum, & Petrosino, 2001) because it accounts more fully for selection effects (Sherman, 2010). Thus, a dichotomous variable was created that assessed the type of statistical analysis used in the study, differentiating between experimental and quasi-experimental study designs. The variable was coded '1' if for experimental studies and '0' for quasi-experimental studies. There were 13 effect sizes from studies that employed randomized controlled experiments (Braga et al., 1999; Braga & Bond, 2008). The remaining 53 effect sizes came from studies that employed a quasi-experimental design (Corman & Mocan, 2005; Harcourt & Ludwig, 2006; Harcourt & Ludwig, 2007; Katz et al., 2001; Kelling & Sousa, 2001; Messner et al., 2007, Novak et al., 1999; Rosenfeld et al., 2007; Wagers, 2007).

Another important distinction between studies is whether the implementation of the BW policing strategy was the sole change in policing or whether it was implemented in concert with other policing innovations (such as problem-oriented or hot spots policing, Compstat, etc.). As discussed earlier, Wilson and Kelling (1982) never asserted that a BW strategy would be effective at reducing crime on its own. Rather, they noted that the strategy would work best along with problem-oriented policing and other policing innovations such as decentralization (see also Kelling &

Coles, 1996). Although disentangling the effect of BW policing on its own would not be possible, it was Wilson and Kelling (1982) who suggested that attempting to would be a fruitless exercise, as the strategy would be shown to be less effective on its own. The expectation is that if this strategy is implemented alongside other policing innovations, it would have more of an effect on serious crime. This information can also help inform policy and future directions in the use of the BW strategy. A dichotomous variable was created and coded '1' if the strategy was implemented alone and '0' if otherwise. There were 30 effect sizes from the one study (Katz et al., 2001) which was conducted as a stand-alone strategy. The other 36 effect sizes came from studies where the policing strategy was combined with various other policing tactics.

The dependent variable, crime, was also coded depending on the type of crime that was analyzed in the study. It is hypothesized that the BW policing strategy would be more effective at combating certain crimes such as robbery or property crimes (Wilson & Kelling, 1982). The strategy would theoretically have less of an effect on crimes of passion such as homicide or rape because these policing strategies will not alter individuals' psychological or biological makeup (Katz, 1988). Given this concern, a dichotomous variable was created to disentangle the effects of the differences in coding of the dependent variable. The variable was coded '1' if the dependent variable looked specifically at robbery or any type of property crime (indexes for property crime were also included) and '0' otherwise. Taking a conservative approach, effect sizes that used an index which included robbery with other violent crimes was coded '0', as it is difficult to estimate whether a specific

crime in the index is driving the results. There were 31 effect sizes that assessed

robbery, individually, or some type of property crime as the dependent variable.

Analog to the Analysis of Variance[13]

In order to study the possible differences among the 66 effect sizes, the

methodological control variables will be assessed individually. Bivariate relationships

will be assessed through an analog to the analysis of variance (ANOVA) This

technique will give additional information beyond the overall effect size estimate and

will support conclusions regarding which conditions BW policing is most effective at

reducing crime. With the analog to the ANOVA, the effect sizes are grouped by the

methodological control variables individually. Then effect sizes can be calculated on

the groups separately (Wilson & Lipsey, 2001). These effect sizes can also be

compared to each other to assess whether the values are significantly different from

each other (Lipsey & Wilson, 2001).[14]

Limitations of Meta-Analysis

Although meta-analysis is considered to be an improvement over a simple

literature review, it does not mean that the technique is without flaws, critiques, or

concerns. There are concerns that should be addressed prior to accepting the results as

accurate. The first criticism with meta-analysis is referred to as the "apples and

oranges problem" (Wolf, 1986; see also Weisburd et al., 2001). The concern here is

that by comparing and aggregating studies that use different measuring techniques

---

[13] In meta-analysis, the analog to the ANOVA is based on the results of the Q-test while with regular ANOVA is based on the chi-square (Hedges, 1982).

[14] In an attempt to examine the multivariate relationships, weighted least squares (WLS) regression was also run. However, due to issues of multicollinearity, the results are not discussed but are included in the Appendix. See footnote 20 for a more detailed description.

and definitions of variables, the results will not necessarily allow for a true analysis of the strategy's effectiveness (Wolf, 1986). With the current study, the vast differences in the composition of the cities and the different techniques used to evaluate the strategy are a concern. As discussed in the above section, the problem can be addressed by coding methodological variables, investigating any of these supposed differences. The concern still lies in whether there are important distinctions that have not been coded, although much thought and consideration has been given to this issue.

"Poorly" designed studies being weighted equally with "good" studies is another concern in meta-analysis. The problem lies in the fact that those studies which are poorly designed tend to find more statistically significant results than studies which have a better design (Wolf, 1986). By combining all studies and ignoring the quality of their design, the analysis results might show more significant findings than is actually the case (Wolf, 1986). This concern can be addressed by creating a system to rank the quality of the research design and can also be included as one of the methodological control variables. Within the current study, the research was all either quasi-experimental or randomized experiments and the quality of the research design was generally not of concern, though this has been controlled for dichotomously.

A final criticism is that there is an assumed published research bias which can affect the results from a meta-analysis. It is assumed that publishers favor results that are significant and will not publish studies that show insignificant results (Wolf, 1986). This problem has been referred to as the "file-drawer problem." Given the fact

that BW as a policing strategy has been a hotly debated topic in the field over the last twenty years, it could be assumed that any study, whether showing significant results or not, would find a home in some journal or published book. Even if this not the case, the current study used a dissertation and thesis repository as a way to check for unpublished sources. As another precaution, a fail-safe N[15] will be computed, which shows the number of additional studies that would need to be included in the meta-analysis to reverse the findings (Wolf, 1986). By taking these precautions, the chance of drawing incorrect conclusions is reduced.

Finally, one should also consider how a meta-analysis may or may not be an improvement over a review of the literature. First, a meta-analysis allows for inclusion of all studies evaluating BW policing thus far. Although inclusion criteria are still subjective with a meta-analysis, i.e. there are still decisions about which studies should be included in the analysis, these criteria are laid out explicitly, allowing the reader to understand and possibly replicate the results. Meta-analysis also includes objective and precise measurements of the effectiveness and significance of BW policing on a reduction in crime. When a literature review is conducted on a body of empirical research, the focus is on what the author believes to be important. A literature review uses the technique of 'vote counting' in which studies either show significant findings or they do not (Wolf, 1986). With meta-analysis, each study is weighted equally.[16] Studies or samples containing multiple findings can also be parsed out using meta-analysis. With few effect sizes, vote

---

[15] $N_{fs\,.05} = \left(\frac{\Sigma z}{1.645}\right)^{2} - N$ (Wolf, 1986).

[16] Studies can also be weighted by sample size or by the inverse variance, which is considered to give a better estimate of the effects (Lipsey and Wilson, 2001).

counting would seem most appropriate at gaining an understanding of the literature as a meta-analysis would seem unnecessary. However, when trying to understand the overall effects looking at 66 effect sizes, a meta-analysis will give us a more informative answer. Accurate vote counting becomes difficult with more effect sizes because overall effects tend to be hard to gauge as the number of effect sizes increases. With meta-analysis, a precise measurement of the effect can also be calculated and the technique will explicate whether it is significantly different from a finding of no effect.

Although using advanced statistical techniques to assess what previous studies have found regarding the effectiveness of BW policing on crime would seem to yield the best results, it is by no means perfect. To gain a precise measurement of the effect of BW policing on crime, a meta-analysis only looks at the correlation between the independent and dependent variable. Therefore, some detail is lost when coding this correlation. In the example of BW, there could be important contextual differences that are being ignored. With a narrative review, the author is allowed to address all of the intricacies of each study and to give appropriate attention to details that may be missed in a meta-analysis. Although coding of methodological variables allows for the meta-analyst to address some of these issues, there is not a perfect solution. With this criticism in mind, it is important to remember that what is lost in detail is gained in the ability to objectively assess the effectiveness of the strategy. A literature review can be a good supplement to any meta-analysis and would allow for a commentary on the status of the literature. Overall, meta-analysis is not a perfect measurement of the

results of a body of research, but the benefits of this type of technique outweigh the

costs and make it a contribution to the field.

Chapter 4: Results

The discussion and interpretation from the results of the meta-analysis will proceed in three stages. There will first be a brief discussion regarding the descriptive statistics of the effect size estimates. Attention will then be focused on results of the Q-test and the overall unweighted and weighted effect size estimates looking at the relationship between BW policing and crime. In the final section, the focus will be on the methodological control variables and the role they play in conditioning the effects of BW policing on crime by discussing the results of the analog to the ANOVA.

Descriptive Statistics

Table 2 presents a descriptive introduction of the effect sizes. Looking first at the location of the study, it is noteworthy that only thirty percent of the effect sizes are drawn from studies conducted in New York City. The remaining seventy percent of the effect sizes came from studies that were conducted in various other locales within the United States (Chandler, AZ; Jersey City, NJ; large Midwestern city; Lowell, MA). The driving factor behind these statistics could be the fact that thirty effect sizes were drawn from one location, Chandler, AZ. The common belief is that New York City dominates the literature and research conducted on BW policing (Beckett & Godoy, 2010; Weisburd & Eck, 2004), but it is evident that this is not the case.

Looking at the implementation of the BW policing strategy, there was a fairly even split between the strategy being implemented as a stand-alone initiative compared to instances in which the strategy was implemented as part of a package.

These data are a bit misleading, however, as the one study which was conducted as a stand-alone strategy (Katz et al., 2001) accounts for about half of the effect sizes in the model. Therefore, when comparing the effect sizes based on these differences below, it can be viewed as a test of the model without the inclusion of the Katz et al., (2001) study, which could be unduly driving the results.

Table 2: Descriptive Statistics

Methodological Controls	Proportion
Study Location	
New York City	0.303
Other	0.697
Study Implementation	
Alone	0.455
Package	0.545
Research Design	
Experimental	0.197
Quasi-Experimental	0.803
Dependent Variable	
Robbery/Property Crime	0.470
Serious Crime	0.530

According to Table 2, eighty percent of the studies did not use an experimental design and there was a fairly even split between the number of effect sizes that looked at robbery and property crimes and those which look at the most serious crimes (murder, rape, aggravated assault). These variables will be discussed further in the later sections of this chapter.

Overall Mean Effect Size Estimates

As was discussed in the previous chapter, a Q-test is routinely performed with a meta-analysis to determine whether the sample is homogenous. Table 3 presents the

Q-test results for the three models that were estimated; the unweighted model, the model weighted by sample size, and the model weighted by the inverse of the variance. As can be noted from the table, when looking at the unweighted mean effect size estimate, the Q-test is not significant, and actually has a P-value of 1.00. Therefore, the null hypothesis cannot be rejected and the conclusion is that the distribution of the effect size is homogeneous.

Table 3: Q-test

Statistic	Unweighted	Weighted: Sample Size	Weighted: Inverse Variance
Q-Test	6.41	505.53***	483.38***

*** p < .001

However, focusing on the Q-test for the weighted effect size estimates, both estimates are found to be statistically significant. In this case, the null hypothesis is rejected, meaning that the distribution is not homogeneous. According to these results, a random effects model is more appropriate for these data. Given these significant Q-tests, random effects models were calculated for all of the overall effect size estimates and are presented in Table 4.[17]

Looking at Table 4, it should be noted that the overall unweighted effect size is -.177. This modest effect size tells us that there is a negative relationship between BW policing and crime such that implementation of a BW policing strategy is followed by a reduction in crime. This is an important finding because it shows that this policing strategy does seem to have some overall effect when looking at violent and property crime. However, this statistic is not significant at the .05 level so the null

---

[17] Statistical analyses were conducted using Stata/SE 9.2 for Windows. Do files were obtained from Dr. David Wilson's website at http://mason.gmu.edu/~dwilsonb/ma.html. These Do files were used to conduct meta-analytical techniques in Stata.

hypothesis cannot be rejected, which in this case means that the notion that this effect size estimate is statistically different from zero cannot be rejected.

Table 4: Unweighted and Weighted Effect Size Estimates for Random Effects Model

Statistics	Unweighted	Weighted: Sample Size	Weighted: Inverse Variance
Mean Effect Size	-0.177	-0.102	-0.107
Standard Error	0.123	0.018	0.018
Z-score	-1.438	-5.805	-5.980
P-value	0.151	0.000	0.000
95% CI	-0.418	-0.137	-0.142
	0.064	-0.068	-0.072

Although the unweighted effect size does not give a clear answer to the research question, looking at the weighted effects tells a very different story. First, focusing on the sample size weighted effect size estimate, the magnitude of the effect has decreased, with an estimate of -.102. However, it should also be noted that the weighting procedure produced a much smaller standard error, and the effect is significantly different from zero. The confidence interval also shows that the true value of the overall effect size falls between -.137 and -.068, with ninety-five percent certainty, favoring the hypothesis that BW policing does have an effect on serious crime. This estimate differs from the unweighted effect size because the weighting procedure takes in to account the sample size of the specific studies. This then gives more weight to studies that found effects with larger sample sizes. These large differences can be accounted for by the fact that some of the largest magnitude individual effect sizes come from studies which used small sample sizes. Therefore, the weighted effect size estimates would take this into account and these studies would not weigh as heavily in the analysis.

Turning the attention to the effect size weighted by the inverse variance, similar results to those discussed with regard to the sample size weighted effect size are noted. The reason for these similar results is due to the formula that is used to calculate the inverse variance for a Fisher's z(r). The formula to calculate the inverse variance is simply the sample size minus three, so these two weighting techniques should not differ too much from each other. The inverse variance weighting yields an overall mean effect size of -.107 which is also significantly different from zero with a similar confidence interval. Given that the technique of weighting the effect size is considered to yield a more precise estimate of the true mean effect size (Lipsey & Wilson, 2001), results indicate that overall the implementation of a BW policing strategy is successful in reducing crime.[18]

Table 5: Unweighted and Weighted Effect Size Estimates without Katz et al. (2001)

Statistics	Unweighted	Weighted: Sample Size	Weighted: Inverse Variance
Mean Effect Size	-0.309	-0.256	-0.249
Standard Error	0.167	0.038	0.037
Z-score	-1.855	-6.780	-6.669
P-value	0.064	0.000	0.000
95% CI	-0.636	-0.330	-0.323
	0.018	-0.182	-0.176

A concern could be raised that the results are being driven by the large number of effect sizes that are drawn from a single study (Katz et al., 2001). Given this potential issue, the overall effect size was also calculated excluding the 30 effect sizes from Katz et al. (2001). The results of these calculations are presented in Table

---

[18] In response to the potential issue of the file-drawer problem, in which null findings may not be published, a fail-safe N was calculated to assess the extent of this problem. The fail-safe N estimate in this case is 1151.62, which means that there would have to be over 1150 null effect sizes to alter our estimates. Pratt has argued that the Fail Safe N is most informative when working with a small number of effect sizes (Pratt et al., 2006). Indeed, this large estimate does not appear to be realistic in terms of a publication bias affecting our results. Therefore, it is safe to assume that publication bias should not be an issue plaguing the study of BW policing.

5. Although the overall conclusion regarding the significance of the effect size estimates remains the same, it should be noted that the exclusion of this single study increases the magnitude of the overall unweighted effect size estimate from -.177 to -.309, a large increase magnitude. Similarly, the weighted effect size estimates increase in magnitude from -.10 to -.25. This indicates that the inclusion of the Katz et al. (2001) study in the analysis tempers the magnitude and is therefore a more conservative estimate of the effect of BW policing on violent and property crime.

Methodological Controls: Analog to the ANOVA

Beyond calculation of an overall estimate of the effect of BW policing on reducing crime, it is also important to control for certain circumstances under which this strategy can be most effective or when it might be less effective. Table 6 presents the results of the analog to the ANOVA estimation of the effect sizes for the methodologically controlled variables discussed in Chapter 3. Results are presented for the weighted effect sizes only.[19]

As Table 6 shows, the effect size estimates are larger for studies located in other cities compared with those located in New York City. For studies located elsewhere, the effect size estimate is -.158 and is significantly different from zero. The effect size for New York City studies is smaller and also non-significant. Although studies elsewhere are significantly different from zero while New York City studies are not, the difference between these two estimates are non-significant, meaning that it is not certain that location has a conditioning effect.

---

[19] Calculations were also conducted for the unweighted analog to the ANOVA, but none of these results were significant.

Table 6: Effect Size Estimates for Methodological Control Variables using Analog to the ANOVA

Methodological Controls	Sample Size Weighted	Inverse Variance Weighted
Study Location		
New York City	-0.095	-0.095
Other	-0.165***	-0.158***
Research Design†		
Experimental	-0.526***	-0.525***
Quasi-Experimental	-0.077**	-0.076**
Study Implementation†		
Alone	-0.018	-0.018
Package	-0.261***	-0.254***
Dependent Variable		
Robbery/Property Crime	-0.174***	-0.169***
Serious Crime	-0.113**	-0.109*

† difference between categories is statistically significant
* $p < .05$
** $p < .01$
*** $p < .001$

Looking at the research design variable, the effect size estimate for experimental studies is -.525 and is over seven times larger in magnitude than the quasi-experimentally designed studies, which is -.076. All of the weighted effect size estimates are significantly different from zero for at least an alpha level of .01. Results from the analog to the ANOVA also indicate that the two categories are significantly different at an alpha level of .001. This indicates that in a more highly controlled research environment, effects of BW policing on recurring crime become more pronounced.

As discussed in Chapter 3, the study implementation variable can be assessed on two levels. First, the variable was coded to compare studies in which BW policing was implemented as a stand-alone policy or as a package along with other policing

innovations. Given that the only study that assessed BW policing as a stand-alone policy also contributes forty-five percent of the effect size estimates, this variable can also be used to assess the effect size estimate with and without this one study (Katz et al., 2001). The results indicate that BW policing has a greater effect on crime when it is implemented as a package along with other policing innovations. The effect size estimate of -.254 is also significantly different from zero at an alpha level of .001. However, the effect size for those situations where BW policing is implemented as a stand-alone strategy is -.018 and non-significant. These effect size estimates are also significantly different from each other at an alpha level of .001. One should be cautious with this finding, however, as it could be a result of lack of studies that assess BW windows as a stand-alone strategy (only Katz et al, 2001).

Finally, Table 6 shows that the magnitude of the effect size appears to be greater when the outcome is robbery and/or property crime (-.169). The difference from the violent crime estimate (-.109), however, is not significant.[20]

---

[20] As discussed in Footnote 14, a regression model was run with these data. However, due to multicollinearity issues between the methodological control variables, the analysis is not discussed here. The Appendix includes the regression information. Based on the high multicollinearity between the location and study implementation variables, two models were run which excluded each of these respective variables individually. The results from these two models are consistent with the results from the analog to the ANOVA and are provided in the Appendix.

Chapter 5: Discussion and Conclusions

Since the introduction of the framework by Wilson and Kelling (1982) almost thirty years ago, scholars and practitioners have focused on the effectiveness of BW policing. The empirical literature commenting on the efficacy of this strategy is mixed, allowing skepticism regarding whether the implementation of a strategy focusing on fixing disorder to combat crime is wise to endure. Through meta-analysis, this thesis has attempted to give an overall clarity to the literature on BW policing. Does it work and under what circumstances does it yield the greatest benefit?

Looking at the mean effect size estimate for BW policing on violent and property crime, there is a significant relationship indicating that implementation of the policing strategy leads to a reduction in crime. Using the random effects model and weighting the individual effect size estimates by either the sample size or the inverse variance lead to a mean effect of -.10. But these numbers have little meaning without something to compare them to. The observed effect size estimate can first be compared to the scale created by Jacob Cohen in which a correlation of .5 is large, .3 is moderate, .1 is small, and anything less is considered trivial (Cohen, 1988). The results do not indicate that the implementation of this strategy has a huge effect on the reduction of crime. It should be cautioned that there is a difference between statistical significance and the magnitude of the effect; though the effect of broken windows on serious crime was found to be significantly different from zero, an effect size of -.1 is admittedly not dramatic. Even so, logic would dictate that crime is such a complex phenomenon that any policing strategy would be unlikely to have an exceptionally

high effect size. In short, this effect size suggests that broken windows policing is effective but it may not "solve" a crime problem, at least as studied thus far.

The general belief in the policing literature is that BW policing appears to be an ineffective strategy at reducing crime. According to Weisburd and Eck (2004), "While the common perception is that enforcement strategies (primarily arrest) applied broadly against offenders committing minor offenses lead to reductions in serious crime, research does not provide support for this proposition" (p. 50-51). Similarly, Harcourt and Ludwig (2006) assert, "there appears to be no good evidence that broken windows policing reduces crime" (p. 316). Finally, Kubrin (2008) notes, "some scholars question how much, if any, of the decline in crime can be attributed to order maintenance policing" (p. 203).

In comparison, policing innovations such as hot spots policing and problem-oriented policing have received more favorable evaluations. In discussing hot spots policing, Weisburd and Eck (2004) claim, "a strong body of evidence suggests that taking a focused geographic approach to crime problems can increase policing effectiveness in reducing crime and disorder" (p. 53). According to Braga and Weisburd (2006), "When police departments focus their efforts on identifiable risks, such as crime hot spots … they are able to prevent crime and disorder (p. 342). Mazerolle (2007) contends, "Recent research agrees with this 'hot spots policing'" (p. 643). Finally, Skogan and Frydl (2004) claims, "policing that is focused on hot spots can result in meaningful reductions in crime and disorder" (p. 238).

In terms of problem-oriented policing, it is Weisburd and Eck's (2004) assertion that, "There is a growing body of research evidence that problem-oriented

policing is an effective approach for reducing crime, disorder, and fear" (p. 55). Braga (2002) claims, "the available evaluation research suggests that problem-oriented policing is effective in dealing with a wide range of crime problems" (p. 2). Finally, according to Cordner and Biebel (2005), "Evaluations of the impact or effectiveness of POP have been generally positive" (p. 158).

There appears to be clear differences in the views of these different strategies, yet the empirical evidence tells a different story. In comparing the overall results of this thesis to Braga's (2007) meta-analysis of hot spots policing and Weisburd's (2008) meta-analysis of problem oriented policing, the magnitudes are found to be fairly consistent across studies. Both Braga (2007) and Weisburd (2008) use the standardized difference of means (Cohen's d) to calculate their overall effect sizes, which is around .12. When this is converted into a correlation coefficient,[21] the results are almost identical to those noted in this meta-analysis on BW.

It is also interesting to note that Braga et al. (1999) was included across all three meta-analyses, meaning that the intervention incorporated elements of BW policing, hot spots policing, and problem-oriented policing. In the current analysis, the Braga et al. (1999) study was responsible for some of the largest correlation coefficients, possibly showing that these three innovations can work in concert to more drastically reduce crime. Alternatively, it could also be the case that this one study could be having a large effect on all three meta-analyses.

Based on the analog to the ANOVA results, there also appears to be a significant difference between studies that used experimental designs when compared

---

[21] The formula to convert a standardized difference of means estimate to a correlation coefficient is:
$r = \frac{d}{\sqrt{d^2+4}}$ (Wolf, 1986).

with those which used quasi-experimental designs as well as studies that looked at BW policing as part of a package when compared to those looking at a strategy that comprised solely of BW policing. Using an experimental design to study the effects of BW policing on crime produces a larger mean effect size, according to the results in Table 6. As discussed previously, this magnitude was seven times greater when compared to quasi-experimental designs. The explanation behind these large differences could be accounted for by understanding the differences between efficacy and efficiency trials. An efficacy trial is a small scale experiment that tests the effectiveness of an intervention under highly controlled conditions while and efficiency trial is an experiment that tests how the policy can be implemented and whether it can be effective in a real-world setting. Usually, greater effects would be expected for efficacy trials when compared to efficiency trials. In the current analysis, the two studies that were experimental used highly controlled hot spots to assess the effect of BW policing on crime. Both studies had small sample sizes and can be considered efficacy trials. The remaining studies consisted of those that looked at the effect of a BW policing policy on crime and had larger sample sizes. They would be considered efficiency trials. Overall then, the results lend credence to the belief that BW policing is a highly effective innovation when tightly controlled, but it can still have a benefit when incorporated into a real-world setting.

According to the results, BW policing also works "better" as part of a package of policing innovations. The results from Katz et al. (2001), in which the police initiated a policy that would only focus on BW policing of a redevelopment district, were not as promising as the results from the rest of studies. The magnitude of the

effect size was near zero (-.02) and was insignificant. Without this study included in the analysis, the magnitude of the mean effect size would have been much larger and would have been considered modest by the standards set forth by Cohen (1988). Of course, this bivariate relationship hinges on the belief that there were no problems with the one study which accounted for all of the effect sizes looking at BW policing as a stand-alone strategy. If one accepts this finding, however, it reminds scholars and practitioners that BW policing should not be implemented as a singular stand-alone strategy, as Wilson and Kelling (1982; see also Kelling and Coles, 1996) suggest.

Possibly more surprising were the lack of significant findings with regards to the location of the study and crime type analyzed. Although the belief is that BW policing has been a New York City phenomenon (Beckett & Godoy, 2010; Weisburd & Eck, 2004), according to Table 2, the majority of the effect sizes came from locations outside of New York, though admittedly this imbalance is not as dramatic when considering studies rather than effect sizes. The results from the analog to the ANOVA also show that the benefits of BW policing in New York City were not as potent as they were in other location. This is in contrast to the hypothesis that studies in New York City would yield more positive benefits of BW policing on a reduction in crime. It should be noted that this finding could be a result of the two studies by Harcourt and Ludwig (2006, 2007) which found an increase in crime from BW policing. In one of the studies (Harcourt & Ludwig, 2006), the authors reanalyzed the data from an earlier study (Kelling & Sousa, 2001) attempting to find alternative explanations for the results.

It has also been hypothesized that BW policing is more effective at combating certain types of crime such as robbery or property crime, but would be less effective at combating more serious crime (Wilson & Kelling, 1982). According to the results, however, this hypothesis does not hold. In a comparison of effect sizes that assessed the effect of the strategy on serious violent crime with those that assessed the effect of the strategy on robbery or property crime, the results were similar and were not significantly different from each other. This means that results indicated BW policing can reduce violent and property crime and that this finding is significantly different from zero.

Overall, police departments can consider BW policing when interested in ways to improve crime rates. The results provide support for BW policing as a policy aimed at reducing violent and property crime. BW policing should be discussed with the likes of hot spots policing and problem-oriented policing in terms of its effectiveness at reducing crime. It also appears that successful policies would incorporate many of these different policing innovations.

The limitations of this analysis, already fully discussed in Chapter 3, are of importance as well. A meta-analysis can only inform the reader about what the literature says with regards to BW policing. Therefore, a meta-analysis is only as empirically strong and convincing as the research articles which have been included in the analysis. There is also the potential that research has been excluded, although many steps were taken to ensure that this was not the case. Furthermore, this analysis is essentially a commentary on eleven studies; as such, it should hardly be taken as the "final word" on the effectiveness of broken windows policing.

Another important limitation to this analysis is the exclusion of information regarding implementation fidelity amongst eleven included studies. To be clear, the studies did not provide much information on the treatment itself, meaning that its fidelity may vary considerably. Perry, Weisburd, and Hewitt (2010) recently highlighted that randomized controlled trials in criminology rarely provide adequate information on the treatment itself, focusing on the outcome almost exclusively. The articles used in the current analyses are consistent with that trend. This is an important concern because "intention to treat" is hardly the treatment of interest and because BW has the capacity to be misconstrued (Bowling, 1999; Cunneen, 1999; Greene, 1999). Future research on BW policing should take this into account and measures of implementation fidelity should be included as an important element of the study.

Perhaps the biggest direction to come from this study is a call for more research on BW policing. One of the most interesting findings of this meta-analysis was that only 11 studies of the effect of BW policing on crime were found. The literature that discusses the potential implications of BW policing considers the conceptual, rather than empirical, dimensions of this policing strategy and dwarfs the evaluation literature. Therefore it is still important that focus be directed on future evaluations of the efficacy of the strategy at reducing crime, taking into account the findings in this thesis showing that experimentally designed studies and studies looking at BW policing as part of a package are significantly more effective at reducing crime.

Future research must also be focused on the collateral consequences of a strategy focusing on disorder. Harcourt and Roberts draw our attention towards important considerations, such as racism and police brutality, and researchers must take these negative consequences into account (Harcourt, 1998; Roberts, 1999). Police legitimacy and trust in the police which has also received very little attention, can be seen as benefits to a BW policing strategy (Taylor, 2004; Taylor & Fagan, 2008). These positive consequences must also be taken into account to gauge the overall effectiveness of any BW policing strategy. By focusing in on these two areas for future research, the field can start to move beyond debating the efficacy of BW policing and analyze the strengths and weaknesses of this strategy.

Appendix

Table 7: Correlation Matrix for Methodological Control Variables

	Study Location	Research Design	Study Implementation	Dependent Variable
Study Location	1.0000			
Research Design	-0.0606	1.0000		
Study Implementation	**-0.9519**	-0.1746	1.0000	
Dependent Variable	-0.0862	0.0143	0.0655	1.0000

Table 8: Beta coefficients for WLS Regression with Methodological Control Variables (standard errors in parentheses)

Methodological Control Variables	Model 1		Model 2	
	Sample Size	Inverse Variance	Sample Size	Inverse Variance
Study Location	-0.039 (0.056)	-0.041 (0.054)	Excluded	Excluded
Research Design	-0.460*** (0.081)	-0.461*** (0.081)	-0.356*** (0.081)	-0.358*** (0.081)
Study Implementation	Excluded	Excluded	0.149** (0.050)	0.146** (0.049)
Dependent Variable	-0.048 (0.051)	-0.048 (0.049)	-0.053 (0.046)	-0.053 (0.045)

* $p < .05$
** $p < .01$
*** $p < .001$

Bibliography

Beckett, K., & Godoy, A. (2010). A tale of two cities: A comparative analysis of quality of life initiatives in New York and Bogotá. *Urban Studies, 47*(2), 277-301.

Bittner, E. (1967). The police on skid-row: A study of peace keeping. *American Sociological Review, 32*(5), 699-715.

Bowling, B. (1999). The rise and fall of New York murder: Zero tolerance of crack's decline? *British Journal of Criminology, 39*(4), 531-554.

Braga, A. A. (2002). *Problem oriented policing and crime prevention.* New York, NY: Criminal Justice Press.

Braga, A. A. (2007). *Effects of Hot Spots Policing on Crime.* A Campbell Collaboration systematic review, available at: http://www.aic.gov.au/campbellcj/reviews/titles.html.

Braga, A. A., & Bond, B. J. (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology, 46*(3), 577-607.

Braga, A. A., & Weisburd, D. (2006). Conclusion: Police innovation and the future of policing. In D. Weisburd & A. A. Braga (Eds.), *Policing innovation: Contrasting perspectives* (pp. 339-352). Cambridge, UK: University Press.

Braga, A. A., Weisburd, D. L., Waring, E. J., Mazerolle, L. G., Spelman, W., & Gajewski, F. (1999). Problem-oriented policing in violent crime places: A randomized controlled experiment. *Criminology, 37*(3), 541-580.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cordner, G. & Biebel, E. P. (2005). Problem-oriented policing in practice. *Criminology & Public Policy, 4*(2), 155-180.

Corman, H., & Mocan, N. (2005). Carrots, sticks, and broken windows. *Journal of Law and Economics, 48*(1), 235-266.

Cunneen, C. (1999). Zero tolerance policing and the experience of New York City. *Current Issues in Criminal Justice 10*(3), 299-313.

Farrington, D. P. (2003). A short history of randomized experiments in criminology: A meager feast. *Evaluation Review, 27*(3), 218-227.

Golub, A., Johnson, B. D., & Dunlap, E. (2007). The race/ethnicity disparity in misdemeanor marijuana arrests in New York City. *Criminology & Public Policy, 6*(1), 131-164.

Golub, A., Johnson, B. D., Taylor, A., & Eterno, J. (2003). Quality-of-life policing: Do offenders get the message? *Policing: International Journal of Police Strategies & Management, 26*(4), 690-706.

Golub, A., Johnson, B. D., Taylor, A., & Eterno, J. (2004). Does quality-of-life policing widen the net? A partial analysis. *Justice Research and Policy, 6*(1), 19-41.

Greene, J. A. (1999). Zero tolerance: A case study of police policies and practices in New York City. *Crime & Delinquency, 45*(2), 171-187.

Harcourt, B. E. (1998). Reflecting on the subject: A critique of the social influence conception of deterrence, the broken windows theory, and order-maintenance policing New York style. *Michigan Law Review, 97*, 291-389.

Harcourt, B. E., & Ludwig, J. (2006). Broken windows: New evidence from New York City and a five-city social experiment. *The University of Chicago Law Review, 73*(1), 271-320.

Harcourt, B. E., & Ludwig, J. (2007). Reefer madness: Broken windows policing and misdemeanor marijuana arrests in New York City, 1989-2000. *Criminology & Public Policy, 6*(1), 165-182.

Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics, 7*(2), 119-137.

Herbert, S. (2001). Policing the contemporary city: Fixing broken windows or shoring up neo-liberalism? *Theoretical Criminology, 5*(4), 445-466.

Jang, H., Hoover, L. T., & Lawton, B. A. (2008). Effects of broken windows enforcement on clearance rates. *Journal of Criminal Justice, 36*(6), 529-538.

Kansas City Police Department (1980). *Response time analysis volume II – Part I crime analysis.* Washington DC: U.S. Government Printing Office.

Katz, C. M., Webb, V. J., & Schaefer, D. R. (2001). An assessment of the impact of quality-of-life policing on crime and disorder. *Justice Quarterly, 18*(4), 825-876.

Katz, J. (1988). *Seductions of crime: Moral and sensual attractions in doing evil.* Washington, DC: Perseus Books Group.

Kelling, G. L., & Bratton, W. J. (1998). Declining crime rates: Insiders' view of the New York City story. *The Journal of Criminal Law & Criminology, 88*(4), 1217-1231.

Kelling, G. L., & Coles, C. M. (1996). *Fixing broken windows: Restoring order and reducing crime in our communities.* New York: Free Press.

Kelling, G. L., & Moore, M. H. (1988). The evolving strategy of policing. *Perspectives on Policing.* Washington DC: National Institute of Justice.

Kelling, G. L., & Sousa, W. H. (2001). *Do police matter? An analysis of the impact of New York City's police reforms*. New York: Manhattan Institute for Policy Research.

Kelling, G. L., Pate A., Ferrara, A., Utne, M., & Brown, C. E. (1981). *Newark foot patrol experiment.* Washington DC: The Police Foundation.

Kelling, G. L., Pate T., Dieckman, D., Brown, C. E. (1974). *The Kansas City preventative patrol experiment: A summary report.* Washington DC: The Police Foundation.

Kubrin, C. E. (2008). Making order of disorder: A call for conceptual clarity. *Criminology & Public Policy, 6*(1), 203-214.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications.

Mazerolle, L. (2007). Hot spots. In J. R. Greene (Ed.), *The encyclopedia of police science: Third edition* (pp. 640-644). New York, NY: Routledge.

McKenzie, D. L. (2006). *What works in corrections: Reducing the criminal activities of offenders and delinquents.* Cambridge, UK: Cambridge University Press.

Messner, S. F., Galea, S., Tardiff, K. J., Tracy, M., Bucciarelli, A., Piper, T. M., ... Vlahov, D. (2007). Policing, drugs, and the homicide decline in New York City in the 1990s. *Criminology, 45*(2), 385-414.

Novak, K. J., Hartman, J. L., Holsinger, A. M., & Turner, M. G. (1999). The effects of aggressive policing of disorder on serious crime. *Policing: An International Journal of Police Strategies & Management, 22*(2), 171-190.

Perry, A. E., Weisburd, D., & Hewitt, C. (2010). Are criminologists describing randomized controlled trials in ways that allows us to assess them? Findings from a sample of crime and justice trials. *Journal of Experimental Criminology, 6*(3), 245-262.

Pratt, T. C. (2002). Meta-analysis and its discontents: Treatment destruction techniques revisited. *Journal of Offender Rehabilitation, 35*(1), 23-40.

Pratt, T. C., & Cullen, F. T. (2000). The empirical status of Gottfredson and Hirschi's General Theory of Crime: A meta-analysis. *Criminology, 38*(3), 931-964.

Pratt, T. C., McGloin, J. M., & Fearn, N. E. (2006). Maternal cigarette smoking during pregnancy and criminal/deviant behavior. *International Journal of Offender Therapy and Comparative Criminology, 50*(6), 672-690.

Roberts, D. E. (1999). Foreword: Race, vagueness, and the social meaning of order-maintenance policing. *Supreme Court Review, 89*(3), 775-836.

Rosenfeld, R., Fornango, R., & Rengifo, A. F. (2007). The impact of order-maintenance policing on New York City homicide and robbery rates: 1988-2001. *Criminology, 45*(2), 355-384.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin Company.

Sherman, L. W., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P., & Bushway, S. (1998). *Preventing crime: What works, what doesn't, what's promising: A report to the United States Congress.* Washington, DC: National Institute of Justice.

Sherman, L. W. (2010). An introduction to experimental criminology. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (399-436). New York, NY: Springer.

Skogan, W., & Frydl, K. (Eds.). (2004). *Fairness and effectiveness in policing: The evidence.* Washington, DC: National Academic Press.

Sousa, W. H., & Kelling, G. L. (2006). Of "broken windows," criminology, and criminal justice. In D. Weisburd & A. A. Braga (Eds.), *Policing innovation: Contrasting perspectives* (pp. 77-97). Cambridge, UK: University Press.

Spelman, W., & Brown, D. K. (1981). *Calling the police: citizen reporting of serious crime*. Washington DC: Police Executive Research Forum.

Tyler, T. R. (2004). Enhancing police legitimacy. *The ANNALS of the American Academy of Political and Social Science, 593*, 84-99.

Tyler, T. R., & Fagan, J. (2008). Legitimacy and cooperation: Why do people help the police fight crime in their communities. *Ohio State Journal of Criminal Law, 6*(1), 231-276.

Weisburd, D., & Eck, J. E. (2004). What can police do to reduce crime, disorder, and fear? *The ANNALS of the American Academy of Political and Social Science, 593*, 42-65.

Weisburd, D., Lum, C. A., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *The ANNALS of the American Academy of Political and Social Science, 578*, 50-70.

Weisburd, D., Telep, C. W., Hinkle, J. C., & Eck, J. E. (2008). *The Effects of Problem-Oriented Policing on Crime and Disorder.* A Campbell Collaboration systematic review, available at: http://www.aic.gov.au/campbellcj/reviews/titles.html.

Wilson, J. Q., & Kelling, G. L. (1982). Broken windows: The police and neighborhood safety. *The Atlantic Monthly, 249*(3), 29-38.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis.* Beverly Hills, CA: SAGE Publications.