

# 21

## Quasi-experimental Research on Community Policing

Wesley G. Skogan

### INTRODUCTION

This chapter describes an evaluation of community policing in Chicago. Although Chicago was not the first city in the US to adopt community policing, it is among the largest. The city developed a program that reorganized the work of patrol officers to tie them more closely to the community, incorporated the public in identifying priority neighborhood problems, and coordinated the work of an array of city agencies in order to address those problems. It was dubbed 'CAPS', for Chicago's Alternative Policing Strategy. After a brief planning period the program began in 5 (of 25) test police districts in April 1993. By March 1995, important elements of the initiative were in place throughout the city. A team of academic researchers headquartered at Northwestern University, which is close to Chicago, became involved in evaluating this effort during its planning phase. Over the years more than 75 faculty students and researchers worked on the evaluation. Regular reports and three books resulted from the project. The first (Skogan and Hartnett, 1997) analyzes the political origins of the program, its planning and implementation, how the city staffed and paid for it, and the impact of community policing in the five test districts where it was first set in motion. The second (Skogan et al., 1999) assesses the program's problem-solving component. It presents a detailed, on-the-ground description of police and community problem solving efforts in 15 selected areas of the city. The final book (Skogan, 2006) describes the development of community policing in Chicago over a

12-year span, and examines trends in crime, fear and satisfaction with policing in the city's diverse neighborhoods.<sup>1</sup>

This was not the first evaluation of neighborhood-oriented policing, or something resembling it. In the early 1980s I was involved in studying a precursor to many of the community policing programs that came later. In that project, police in Houston, Texas, and Newark, New Jersey, tried to engage with the public by opening storefront offices, distributing newsletters, and going door to door to inquire about local problems. In Houston, a team of officers was trained in community organizing, and another provided new support for crime victims. Newark experimented with foot patrols, a neighborhood clean up, opening schools at night to provide recreational opportunities for youths, and aggressive order maintenance efforts on busses and busy street corners (Pate et al., 1986).

The research design and many of the survey questionnaires our team used in Chicago drew from that experience. So too did our interest in the impact of community-oriented policing on a broad range of outcomes, including fear of crime, disorder reduction, physical decay and community self-help. It was apparent in both Houston and Newark that engaging the public brought new issues to the fore, and that a narrow focus on traditional measures of crime victimization would miss many important aspects of 'public safety' more broadly conceived. The evaluation strategies employed in Houston and Newark influenced ensuing studies of community-oriented policing in a long list of cities. The findings of many of

these investigations, which took place between 1986 and 1990, are summarized in Skogan (1994).

Many more studies of community policing appeared in the 1990s, but they were released only after we had launched our work in Chicago. As we began, the Vera Foundation published an evaluation of the assignment of dedicated neighborhood officers in New York City (McElroy et al., 1993), and Sadd and Grinc (1993) issued a comparison of community policing projects in eight US cities. Studies then appeared describing projects in Australia, Britain and Canada. In the UK, Trevor Bennett (1991) evaluated a pioneering fear-reduction project that was conducted in both Birmingham and London. Fifteen years later, the Home Office was busily replicating many aspects of the program and the evaluation in communities all over England and Wales (Quinton and Tuffin, 2007). Meanwhile, an edited collection by Dennis Rosenbaum (1994) provided an outlet for a large number of studies that had been completed but were as yet unpublished.

As all of this research activity suggests, community policing was being widely adopted. The concept was proving immensely popular with the public, and thus with politicians, and as a result few police chiefs wanted to be caught without adopting something they could point to and call community policing. Chicago was no exception.

This chapter describes the evaluation and some of what we found. Early sections introduce the evaluation and the program, and review the kinds of data we gathered to assess its major components. There is a discussion of general issues that drove the design and execution of the project. A final section places the findings for Chicago in larger context, comparing them with the results of evaluations of other community policing initiatives.

## THE EVALUATION

Because of its magnitude, the developmental nature of the project, and its decentralized character, it was important to understand what the program actually looked like in the field, as well as to systematically assess how effectively it addressed the city's problems. These twin interests called for two kinds of evaluations, and we did both.

Process evaluations examine program design and implementation. They document a program's 'theory', or how its developers thought it was supposed to have an impact. Process evaluations also document the actual implementation of the program, for there is often a gap between plans and reality. Uncovering what really happened on the

ground is particularly important in the case of community policing. In Chicago, as in many cities, the program involved decentralizing authority and responsibility, moving it out of police headquarters and down the organizational hierarchy. Especially in the early years, this inevitably led the program in different directions, depending upon neighborhood conditions. It turned out that sometimes we knew more than the senior management about how key elements of the program were evolving. At the same time, community policing challenged many deeply rooted 'business as usual' aspects of the organization. The program was not initiated by the department; rather, it was imposed on it by the city's political leaders, who were responding to public concern about crime and dissatisfaction with policing in the city. Thus there was ample reason to fear that officers and commanders would fall back upon their familiar routines when no one was watching, and our early reports frequently concluded that the organization was not properly structured to ensure that the city's plan was actually being carried out. It almost collapsed in 1999, and our report that year warned that 'CAPS is dead in the water'. The ensuing crisis led to a successful re-engineering of supervision and accountability processes in the organization, and the program weathered the storm.

Impact evaluations analyze the effects that programs have on the problems that they target, and look to see if they have unexpected or unintended consequences as well. As I describe below, one thing we feared was that there would be unequal outcomes by race. The strength of an impact evaluation is dependent upon its design and how well it measures what the program might accomplish. This chapter describes how we took advantage of the trial run of the program in a few districts, plus a (rare) opportunity to gather benchmark data before the program actually began, to assess the impact of community policing in a diverse set of neighborhoods. Later, after the program expanded to encompass the entire city, the evaluation focused in detail on how specific components of the program were working, while it continued to monitor trends on a broad set of outcome measures.

## THE PROGRAM

Police departments embracing community policing typically adopt three inter-related organizational strategies: they (a) decentralize and adopt a local neighborhood orientation, (b) facilitate community input in defining and prioritizing their activities, and (c) adopt a broadly focused,

problem solving orientation toward their work. CAPS encompassed all three of these program elements.

### **Decentralization**

Chicago's first move was to reorganize the work of its patrol officers to support community policing. Small teams of officers (usually nine, plus part of the time of a supervising sergeant) were given round-the-clock responsibility for each of the city's 279 small police beats. The emergency 911 system was reconfigured, effectively restructuring the daily routines of thousands of patrol officers. The system prioritizes dispatching in a way that keeps team officers concentrated on their beat, often with some 'unassigned' time available for getting out of their cars to attend meetings and talk with residents, merchants and building managers. Officers assigned to beat teams are expected to engage in identifying and dealing with a broad range of neighborhood problems in partnership with neighborhood residents, community organizations and other city agencies. Beat team officers regularly attend public meetings that are held monthly with local residents. There they review their progress on problems and discuss emergent issues. Other police units were decentralized, so that local commanders had somewhat more control over investigators and gang-crime specialists and could integrate their efforts with plans being developed at the grassroots level.

One tool for evaluating how this new turf orientation was working was surveys of police officers. During the course of the evaluation we surveyed about 13,600 police officers, some undoubtedly more than once. To control our costs, questionnaires were distributed in group settings, either at the roll-call meetings held when officer's shifts began or when they gathered for training sessions. We found that officers surveyed before the program began were quite dubious about the community policing enterprise, as were their immediate supervisors, but that both groups grew more positive over time. At almost every point their reactions to community policing were tied to race; African-American officers were the most optimistic about CAPS, while white officers were more skeptical by a very large margin. White officers felt welcome at public meetings where most of the residents attending were white, but otherwise were uneasy. Latino officers felt less welcome at meetings hosting a large proportion of Spanish-speaking Latinos; apparently, Spanish-speakers in attendance were making heavy demands on them. On the other hand, the views of African-American officers were unaffected by the racial composition of the meetings they attended, and a large majority

felt welcome everywhere. Initially, most officers suspected that other city agencies would not help out with their problem-solving projects, but later reported that they did. Over time, the surveys revealed a host of program glitches; for example, that sergeants were not intervening to make the new dispatching system work, and the brief meetings that were to be held between team officers beginning and ending their work shifts were a waste of time.

We used data on individual police officers to monitor the department's ability to manage the process of linking beat officers with community residents. An anticipated feature of CAPS was that residents would come to know the officers who regularly attend beat meetings, and the officers would also be the ones that answered calls and could be seen around the neighborhood. On their part, beat team officers needed to stay in one place long enough for residents to meet and learn to trust them. Time would also help officers develop a personal sense of 'ownership' of their turf and develop an ethos of teamwork.

However, police in Chicago regularly bid for their district and shift assignment by seniority. In addition, personnel transfers, vacations, relief duties, court appearances, and compensatory time off made it hard to maintain staffing stability. The resulting turnover of officers from public meeting to public meeting did not go unnoticed, and both regular attendees and community organizations periodically complained about it. To assess the department's ability to make stable assignments happen, we examined the brief reports that are filed about each beat meeting and recorded the unique employee identification number of each officer listed as present. For example, during 1998, an average of five to seven police attended each beat meeting, and during the course of the year 4,650 different officers showed up for at least one meeting. Beats meet an average of ten times per year, so officers were classified as 'regular participants' if they attended at least five meetings in a beat during that period, a generous standard. By this standard, 14 percent of beats did not have any regular officers at all. We judged that only one-third of the beats had adequate personnel stability, with at least five officers attending meetings regularly. Continuity of officer participation in beat meetings was far from random. In particular, poor African-American areas least often had a full complement of regular participants; only 26 percent of those beats had five regularly attending officers. Beats in predominately Latino areas did not do much better (28 percent). Better-off African-American beats did better still; 42 percent of those beats saw a large complement of officers on a regular basis. On the other hand, more than 60 percent of predominately white, middle-income

beats had five or more police officers attending meetings on a regular basis.

We tried to use data on emergency calls from the dispatching center to assess stability of assignment, but it was in terrible shape. Until the late 1990s, information on each call and dispatch was handwritten and passed around the dispatching center on paper slips, then later punched on cards for bulk processing on a computer with vacuum tubes. Neither we nor the police could get a clear handle on the effectiveness of CAPS' reorganization of the dispatching of beat teams until they acquired new information-processing capacity. The goal was to keep beat teams on their beat 70 percent of the time, and the actual figure turned out to be 68 percent.

### ***Engaging with the public***

In Chicago, the primary mechanism that police have for engaging with the public is beat meetings. These are regular – usually monthly – gatherings of small groups of residents and a number of the officers who actually work in the beat. The meetings are held at night in church basements, park buildings and hospital cafeterias all over the city. During the evaluation an average of 6,700 residents attended about 250 meetings each month. They met with an average of five to seven police officers, most of whom regularly patrol in the area. Police serving in specialized units, such as gang teams or detectives, are often present as well, along with a representative of the district's neighborhood relations unit. Meetings are sometimes attended by representatives of the city's service departments and area community organizations, and the local aldermen's staff. Most meetings are co-led by a civilian recruited from the neighborhood. The meetings provide a forum for exchanging information and identifying, analyzing and prioritizing problems in an area. Local crime maps, lists of the most frequent crimes on the beat, and other informational materials are distributed at the meetings, following a printed agenda. There is always a discussion of what has happened with regard to issues raised at the last meeting, and this provides a bit of community oversight of police activity. The new business segment of the meetings focuses on identifying new issues and debating whether they are general problems or just the concern of one resident. Beat meetings are also a very convenient place to distribute announcements about upcoming community events, circulate petitions, and call for volunteers to participate in action projects. Importantly, they also provide occasions for residents and the officers who work in the area and will likely answer their calls to get acquainted.

One very important evaluation question was, 'What actually goes on at beat meetings?' This was a difficult and expensive question to answer. Because they are held night after night in hundreds of locations all over the city, it is hard even for the police department to know what is going on. An officer who attends fills out a brief form reporting the number of people who came and summarizes what was talked about, but not much information is recorded, and other observers might have a different view of what transpired. Our city-wide surveys asked respondents if they had attended a beat meeting in the past year, and in a typical year about 22 percent of adult Chicagoans indicated they had. The surveys found that a very large majority of them thought the meetings were useful and led to improvements in their neighborhood, but they could tell us few details about how the meetings actually functioned.

In order to examine the dynamics of beat meetings more closely, observers **working on the evaluation** attended hundreds of them between 1995 and 2002. They completed forms recording specific details about the sessions, including where it was located and how long it lasted. They counted the number of police and residents who attended, by race and sex. They reported on the languages that were spoken and the kinds of printed materials that were distributed. Using a checklist, they noted the issues that were raised by residents during the course of the meeting. They noted which participants identified problems and solutions to problems that were discussed. The observers were trained to make judgments about such matters as how effectively police and civilian leaders conducted the meetings, and they classified the roles played by police and other city employees. The previous description of what happens at beat meetings was drawn from the resulting data.

In addition, we surveyed participants at the meetings on several occasions. At a prearranged moment the observers rose, explained who they were, and distributed questionnaires to both the police and residents who were in attendance. In 1998, observers attended one meeting each in 253 beats and surveyed 5,293 residents and 1,050 officers. In 2002 they observed two or three (if we could) meetings in each of a random sample of 130 beats, and surveyed 3,495 residents and 643 officers. A 1995 study involved 161 beats clustered in selected districts. The surveys asked about conditions in the beat, the quality of local police service, resident involvement in CAPS-related activities, how frequently they attended the meetings, and assessments of the meetings they had attended. The questionnaires for residents were available in both English and Spanish.

In the field, we found that, at the average meeting, about one-third of those in attendance were there for the first time, another third reported attending two to three meetings in the past year, and the remainder were loyal participants. But because they came frequently, the latter group accounted for a very large percentage of those who showed up in the course of a year. Compared to census reports on the composition of the beats, frequent attenders were older, did not have as many children at home, were more often retired, and were more likely to be long-term residents and homeowners. In short, there was an 'establishment bias' of major proportions in beat meeting involvement. However, a comparison of participants' ratings of neighborhood problems with the findings of surveys of residents of the same beat found that the meetings actually represented the interests of their 'constituencies' fairly well. We concluded that this was because 'small is beautiful'. Chicago's beats are small in population and geographically compact. Some of the residents attending meetings may be college graduates and others high school dropouts, but they still walk to the same bus or train stop and shop at the same supermarket. They pass by the same graffiti-clad schools and abandoned automobiles. Living together, their fates are linked to those of their neighbors despite differences among them.

### ***Broadly focused problem solving***

One of the most interesting aspects of community policing is that police departments find that they must take on a much broader range of issues than they did before, a consequence of opening themselves up to public input. At community meetings residents complain about bad buildings, noise and people draining their car radiators at the curb, not just about burglary. If police reply, 'But that's not our responsibility', and try to move on, no one will come to the next meeting. As a result, they need to find partners that can help them. These frequently include city departments responsible for health, housing, garbage collection, and even street lighting. From the beginning, the delivery of services in Chicago was linked to community policing. Procedures were developed to link agencies to requests for assistance from the beat teams and in response to complaints lodged at beat meetings. Officers were able quickly to mobilize building inspectors, street cleaners and repair trucks, and other city resources. They could easily get abandoned cars towed away, empty buildings boarded up, and vacant lots cleared of trash. Department employees orchestrate neighborhood cleanups and graffiti paint-outs by volunteers and city workers, and take note of burned-out street

lights and trees that needed trimming. Police districts have 'problem-buildings officers' who inventory dilapidated and abandoned structures and track down property owners for civil prosecution over building, health, fire, sanitation, and business-license violations. This broad focus led us to include measures of physical deterioration on our checklist of program outcomes.

One of our tools for identifying changes in neighborhood conditions was a large-scale survey that involved interviewing, and later re-interviewing, randomly selected samples of residents. We were able to do this early on, before the program began, and then again while it was in operation in only a few police districts. The re-interviews thus enabled us to measure changes that took place following the introduction of the program, and in addition to compare those to over-time changes in areas where the program had yet to be introduced (there is more on this later). The surveys were conducted by telephone. This was principally a budgetary matter, but telephone surveys have advantages. We were able to re-contact sample numbers many times (as many as 22 times), enabling us to reach mobile and frequently-not-at-home individuals. We could also afford to re-contact households that initially refused to be involved in the study. Very importantly, it could be dangerous for interviewers to walk the streets of many of the neighborhoods we were studying, and evaluation staff safety considerations were always at the top of our checklists.

There are many advantages to two-wave surveys, in particular that they directly measure individual-level changes in attitudes and self-reported behavior. However, they typically suffer from significant attrition. Americans move frequently, especially so in high crime areas and immigrant communities, so the capacity of telephone surveys to successfully re-contact survey respondents after an extensive period of time is always limited. Across all areas we successfully contacted and re-interviewed 59 percent of the original survey respondents, which was about normal for paired interviews over an 18-month interval. We were least likely to find and re-interview Latinos, who are the youngest and most mobile component of the city's population. We adjusted the survey data to account for this attrition in several ways: by weighting the data we did collect using census estimates of the total population, and by controlling for the under-representation of groups in the statistical analysis.

A key role for the surveys was to gauge possible changes in the extent of crime, disorder and decay problems in each area. A survey is a good instrument for doing so, for many of the problems that concern residents are not captured by official record-keeping systems or are very poorly

recorded when they are. For example, street drug dealing appears in official statistics only when arrests are made, and arrest numbers simply did not reflect the wide-open street drug markets that plagued several of our study neighborhoods. Graffiti is only rarely reported to police (most people do not connect it with making an emergency 911 call), and before CAPS began it was not dealt with even when they did call. No one in Chicago kept official statistics on problems like junk-strewn vacant lots and loose garbage in the alleys, but these concerned residents of some of the study neighborhoods a great deal. Using surveys enabled us to gather uniform and comparable measures of a broad mix of problems. Respondents were quizzed about a long list of specific issues that we anticipated would be problems in various parts of the city. Neighborhood residents were asked to rate each of them as 'a big problem', 'some problem' or 'no problem'. In addition, the surveys also measured the victimization experiences of those we questioned. This enabled us to assess changes in crimes like burglary, vandalism, street robbery, and auto theft. Because personal crime and car theft could take place virtually anywhere, the surveys had asked victims if the incident took place in their neighborhood or not, so this could be accounted for.

We also examined time-series trends in officially reported crime. Surveys in the five prototype districts found that auto thefts, burglary, and assault were among the public's top concerns, and we could track both official crime figures and survey reports of victimization by these crimes. This was not easy. In the early 1990s police databases were a mess. Department records were maintained on an aging mainframe computer and stored on reels of computer tape that kept breaking. Information on such vital points as the address of the offence was keyed in haphazardly, with many variations in spelling and abbreviations. We had to individually clean up many thousands of computer records for the early 1990s, before the department made our life simpler by acquiring a modern computer and database system.

The surveys documented the disparate character of community priorities across the five prototype areas. Only two problems were of virtually universal concern: street drug dealing was nominated among the top four problems in every area, and 'shooting and violence by gangs' was a leading problem in four of the five prototypes. Gangs and drugs are challenging issues that lay near the core of the city's crime problems in the 1990s. They present a difficult target for community policing and, indeed, policing strategies of any style. Otherwise, a wide range of problems were identified as particularly troubling, and many

issues that loomed large in some areas were scarcely problems at all in other districts. In two areas car vandalism was near the top of the list, and in two others household vandalism ranked high. Auto theft, burglary, disruptions around schools, abandoned buildings and 'vacant lots filled with trash and junk' each stood near the top of the list in at least one district.

## ISSUES DRIVING THE EVALUATION DESIGN

### *Chicago is large yet divided*

The central city itself is home to almost three million residents, but most look to their immediate neighborhoods and identity groups when they decide how they are faring. The city is particularly segregated by race, and the racial composition of the city's neighborhoods provides a template that describes the distribution of almost every social ill. In 2009, African-Americans constituted about 35 percent of the population. A majority of them are poor, for beginning in the 1970s many better-off blacks migrated to close-in suburbs on the fringe of the city, searching for nicer, safer housing and schools. About 30 percent of the population is white. They are divided roughly evenly by lifestyle, with younger, more affluent and college-educated whites clustering by the lakeshore, while blue-collar and lower-middle class whites cling to the far reaches of the city proper, far from Lake Michigan. During the mid-2000s, Latinos passed whites to become the second largest group in the city. Along with a far-smaller number of Asians they are the only group that is growing, and it is quite possible that by 2012 or so Latinos will surpass African-Americans and become the largest group in the city. They are divided by immigration status. The numbers of newcomers and second-generation or more Latinos are both growing rapidly, but the flood of immigrants that came to Chicago from Mexico in the 1990s and 2000s has had the greatest impact on the growth of dense, Spanish language barrios (neighborhoods) around the city's core.

The contested nature of the city shaped the evaluation in fundamental ways. First, it was clear that we had to be able to be able to speak separately to the concerns of major population groups. They faced different problems, and the community activists and politicians who represented them would only be interested in findings that related to their constituencies. Instead of watching the program unfold in a few selected places, the evaluation needed to be based on large samples of residents, activists, patrol officers and police

beats, so that the findings could describe general trends in Chicago *and* have enough detail to break out what was happening to its largest groups. While in-depth studies of individual neighborhoods, police teams or specific problem solving efforts might have been informative, it would have been difficult for us to conduct them in a sufficient number and variety of locations. What people wanted to know was what happened in their neighborhoods, or at least in their kind of community.

The contested character of the city also made it imperative that the evaluation consider the *distributive* consequences of community policing, and not just its overall impact. I knew from experience that community policing threatens to become polarized. Policing by consent can be difficult in places where the community is fragmented by race, class and lifestyle. If, instead of trying to find common interests in this diversity, the police deal mainly with elements of their own choosing, they will appear to be taking sides. It is very easy for them to focus on supporting those with whom they get along best and whose outlook they share. But as a result, the 'local priorities' that they represent will be those of some in the community, but not all. My evaluation of community policing projects in Houston found that the way in which programs in various areas were run favored whites, homeowners and established interests in the community. Police worked well with members of those groups, but less affluent residents did not hear about the programs and did not participate in them. The positive effects of community policing turned out to be confined to whites and homeowners; African-Americans and Latinos saw no visible change in their lives. As a result of this experience, at every turn we considered the possibility of differential outcomes by race and class.

Concern about distributional issues proved to be well founded. The main findings of the evaluation were reflected in the subtitle of my 2006 book, *A Tale of Three Cities*. In a nutshell, blacks, whites and Latinos began the program facing different problems, and their communities went in different directions over the course of the next decade. Conditions were initially worst in the city's African-American neighborhoods, but things improved there a great deal and they saw the most benefits from the program. Beat meeting participation was highest in black neighborhoods, and crime there dropped most. Satisfaction with police performance rose by 14 percentage points among generally skeptical blacks. Whites, on the other hand, were already well off, and there was not much room for them to show many gains. They could already get things done when they needed, through strong neighborhood organizations and political channels, and they were already

friendly with the police. But over the course of the 1990s things grew worse in the city's predominantly Latino neighborhoods, and by the 2000s they were by some measures worse off than the blacks. The influx of immigrants from Mexico drove down wages and levels of education and home ownership. The Latino community cleaved in two during the 1990s. The poorer, more disorganized and immigrant segment grew faster and became more concentrated in new city barrios. Awareness of CAPS actually declined among Latinos, and few got involved in the program. In the end, the inability of the city to engage with its large and fast-growing immigrant Latino population was the biggest shortfall of community policing in Chicago. Police faced challenges thrown up by demographic turmoil and globalization of the economy in their own back yard, and they did not make much headway.

### ***The evaluation was not an 'experiment'***

Experiments, strictly speaking, demand a great deal of control on the part of the researcher. Experimenters must be able to control where or who gets the program, and they use this power to assign their subjects at random to program and control groups. They also control the intervention itself, determining who gets what 'treatment' and how much of it (the 'dosage') they receive. It is the ability of researchers to control who receives how much of what treatment, and thus to make unambiguous claims about their causal effects, that grants 'gold standard' status to randomized experiments in the social sciences. We had none of these powers.

What we had instead was a fairly brief time period – about 18 months – during which the police would be trying out their program plan in a few selected areas of the city, while in other places policing would continue as usual. They had developed a fairly elaborate plan for reorganizing the work of patrol officers and involving the public in identifying priority problems in their neighborhood, but they did not know if it would work. So, in a wise move, they decided to try out the plan first in five of the city's 25 police districts, a number that was selected because they thought they, and we, had the resources to monitor how well it would go there. These were officially dubbed 'prototype' districts, to connote that the program there was still a work in progress. The prototypes kept their existing personnel and leadership; in the words of one department executive, they did not 'stack the deck in favor of success', for they knew that once the program encompassed the entire city it had to work with the people and talent that the department already had.

However, the sheer size of the city ensured that this stage of the evaluation would still be large in scope. If just the five police districts in which the program was first tested had been independent, they would have been one of the larger cities in the US. Together, more people lived in the prototypes than in the city of Seattle, and they were only slightly outnumbered by the population of Boston. If the police officers in the five districts were to form their own department, it would have been far larger than the agencies serving the cities of Atlanta, Miami, Kansas City or New Orleans. The size and diversity of the prototype districts led us to anticipate that each district would evolve a somewhat different program over the course of the experimental period, because at the outset no one in Chicago had a clear idea about what community policing should look like. This was another reason why the evaluation needed to speak separately about each area, and about each area's problems.

While the evaluation was not an experiment, the decision to work first in the prototypes at least gave us both program districts and (for a while) 'business-as-usual' areas to work with. However, we had no power to determine *which* parts of the city would be involved at the outset. By a few months before the process was to begin, word about it had spread among community activists and political leaders, and everyone wanted to be first. The final call as to which districts would be among the prototypes was determined by City Hall. One district that was not on the candidate list submitted by the police department was inserted by the mayor after behind-the-scenes maneuvering by powerful politicians and by well-connected community activists who besieged him, demanding that their districts be selected. In this district, our first survey found that burglary was the top-ranked neighborhood problem, but that only 10 percent of residents gave it a high rating. There was not much room for improvement on many of our outcome measures in this fairly well-off community. However, between them the five districts did represent a range of communities that also included some of the city's worst-off African-Americans and dense concentrations of recent immigrants, giving us enough variation to work with.

A next step was to select sections of the city which closely matched the newly announced program areas. The matching factors for selecting these were race, home ownership, and features of the housing stock such as the percentage of residents living in single family or high-rise buildings. Crime rates were not used for matching purposes – it is bad practice to 'match on the dependent variable' – but each pair of evaluation areas turned out to have similar levels of officially

recorded crime. Because they were far from randomly selected, we dubbed these 'comparison' areas. In a true experiment the comparison areas would have been the 'control groups' against which changes in the program areas could be compared, and the areas would be almost exactly similar. However, none of these areas were randomly selected and they each had a distinctive character, population, and history. Because the match of each experimental and comparison area was imperfect, if the program was received differently among various social groups (say, by race or social class) the population mixes of the areas might account for some of the changes that we interpreted as effects of the program. So too could unique local events and other neighborhood factors that might influence either the experimental or comparison areas, but not both. We patched up these design flaws as well as we could, by monitoring events in the program and comparison areas and statistically controlling for some of the remaining race and class differences between residents of the paired communities. The comparison areas were selected so each included parts of at least three police districts, so that we could adjust at least some of our data for any independent community policing efforts that district commanders might decide to put in place during the course of the first year of the program. We could not control our 'control groups' either.

In the evaluation, the comparison areas were the 'counterfactual'. Changes over time in conditions in the comparison areas were used to represent what *would have happened* in the experimental districts *if* there had been no program. This is a far stronger research design than a simple 'before-and-after' description of trends over time in program areas. With before-and-after designs one has no idea what other factors may have contributed to changes that occurred; these might include the weather, general economic conditions, the coverage of crime and the police in the media, or even other programs or the efforts of other agencies. Changes over time in the matched comparison areas provided a benchmark against which changes in the program areas could be contrasted.

A significant feature of this evaluation is that we were afforded the luxury of being able to organize the evaluation and gather baseline data before the program began. This is often not the case. More typically, evaluations are almost an afterthought, and they are funded and begun after the agony and excitement of getting the program into the field has worn off. We had to hurry, but planning for the evaluation began almost four months before the program was launched, and we were able to gather enough money from private foundations and the federal government to staff



the study, select the comparison areas, and conduct large surveys in the program and comparison neighborhoods before it began. The result is that we could field something approximating a 'quasi-experimental' evaluation: there were before-and-after measures of the presumed outcomes of the program in both program and comparison areas (see Shadish et al., 2002).

We found modest but consistently positive effects of community policing. In the surveys, the extent of neighborhood physical decay was measured by ratings of the extent of vacant lots filled with trash and junk, abandoned cars, abandoned buildings and graffiti. Decay went down significantly in the three most troubled experimental areas. For two of these there was no parallel decline in matched comparison areas, and for the third district there was a smaller although still significant decline in the comparison area as well as the experimental area. In the worst-off area we studied, residents themselves prioritized two physical decay problems among their top four concerns. Both were the targets of concerted efforts by the police there, and both declined by a very substantial margin. The other two experimental areas were better off, and they were visibly cleaner and in better repair even before CAPS began. Respondents in one of those districts identified graffiti as a top-four problem, but it did not decline significantly there during the months after the program began. There were two measures of social disorder in the evaluation. Disorder in and around schools was identified by respondents in one district as a top-four problem, and over the course of the evaluation it declined noticeably but not significantly; there were no changes in the matched comparison area. Likewise, vandalism to parked cars declined but not significantly in another experimental area, but did not decline at all in its comparison area. The program led to a reduction in other forms of crime in three of the five areas. As measured by surveys, crime dropped in all five police districts, but did not all outpace trends in some of the matched comparison areas. Gang and drug problems declined significantly in three districts. For example, in the worst-off district, reports of serious street drug markets dropped from 62 to 49 percent, and from 66 to 53 percent in the next-worst areas. Officially recorded auto theft, burglary, robbery and assault generally declined (or remained stable) in parallel with the survey findings.

We also had to deal with the issue of displacement. This was the possibility that the introduction of new policing strategies to some extent pushed problems elsewhere, out of the program areas. While there was evidence of significant declines in a wide range of problems in Chicago's experimental police districts, there lingered the

possibility that some were simply displaced elsewhere rather than truly being resolved. The possibility of displacement questions the results of virtually every crime prevention program, but rarely are researchers in a good position to monitor this possible outcome. The possibilities for displacement are numerous. Depending on the problem, displacement might be geographical: the program may push it into another neighborhood, or somewhere down the highway. However, an effective program might instead displace a problem in time, to a period when residents or police are not patrolling; or, offenders might switch to another type of crime, with no guarantee that the neighborhood will be better off due to their taking up a new line of work.

Of all these possibilities, we were able to consider only the possibility of geographical displacement, and then only tentatively. To monitor displacement we identified the program areas in which problems that we judged to be the most displaceable – gang violence, street drug sales, and street crime – declined significantly. We then identified potential displacement zones around these program districts that were areas where we had also conducted evaluation surveys (they were parts of the comparison areas for this study). Those potential displacement areas were defined as the first two tiers of census tracts along the borders of the prototype areas. The survey responses of those who lived in potential displacement areas were contrasted with those of respondents living deeper in the comparison areas and were presumably less vulnerable to geographical displacement.

We then examined Wave 1–Wave 2 changes in gang violence, street drug sales, and street crime, accounting separately for changes in the prototypes, the potential displacement zones, and the (now shrunken) comparison areas. We examined change scores for the three areas and statistically combined all of the data. In no case was there evidence of a significant increase in drug, gang or street crime problems in a potential displacement zone. If anything, there was a hint of a diffusion of benefit rather than displacement of crime (see Clarke and Weisburd, 1998). In the aggregate, gang problems went up in the relevant comparison areas, but remained steady in the displacement zones near the prototypes. Street crime remained steady in the comparison areas, but went down in the displacement zones. There were no apparent shifts in drug problems, other than their significant decline in the prototype districts. While none of the gang or street crime changes in the displacement zones were statistically significant, the number of survey respondents living in the potential displacement areas was not large, so we were not tempted to claim that the *benefits* of

community policing spilled over from the prototypes into surrounding areas, but that hypothesis was at least as credible as fears about displacement.

### ***CAPS had few formal goals***

One goal of community policing is to open departments up to local input so that they can effectively discern differences in problems and priorities among neighborhoods, and to tailor their operations to respond appropriately. Chicago's policing plan did not specify particular targets for the program, or anticipate how much impact it might have. Our life as evaluators would have been much simpler if the department's leadership had announced goals and target figures – say, a 20 percent reduction in household burglary – but they did not. This reflected the bottom-up approach to policing they were developing. The planners' vision was that teams of officers meeting and working with the public would identify different kinds of problems, and differing priorities for them, in the city's many communities. The goal was to be 'responsive' to what they turned up, and to find effective ways to deal with it. Our evaluation in turn needed to be responsive to this agenda.

To establish whether CAPS successfully addressed priority neighborhood issues, the evaluation focused on the issues identified by neighborhood residents themselves as the most serious problems facing their communities. Our first analysis of the impact of the program examined the four biggest problems that residents of each area nominated in their first interview. We then compared them to ratings given the same issues more than a year later. This analysis let residents set the agenda for the evaluation through their expressions of concern about neighborhood conditions. Focusing on a fixed number of problems in each area helped protect us against random fluctuations in the measures. In the Chicago study there were five program areas, but several dozen outcome measures. If we looked at changes in every measure in every area, some differences almost inevitably would be due to chance fluctuations in the measures. This would be true whether the outcomes were measured by surveys or by official statistics, so disciplining ourselves by looking only at a clearly defined set of 'priority' problems was important.

A largely unstated goal of CAPS was to improve the image of the police in the city's poor neighborhoods. As in many cities, Chicago police faced a 'legitimacy deficit' of major proportions, and their more sophisticated executives knew that this was undermining their effectiveness. Shootings by

police and charges of brutality were commonplace in African-American neighborhoods, and our first survey found very large gaps between whites, blacks and Latinos in their evaluations of the quality of police service. Memories of the 1991 Rodney King episode – visions of a black man being pummeled by a large number of officers in Los Angeles while a video camera rolled – were still vivid in 1992 and 1993 when discussions took place in City Hall about what to do about policing in Chicago.

The surveys enabled us to track changes in the perceived quality of police service delivered to the experimental and comparison areas. We asked respondents to rate the performance of the police on several dimensions: on their effectiveness (at keeping order, reducing crime and helping victims), responsiveness (to community concerns and the problems that bother people, plus working together with residents) and demeanor (are they fair, polite, concerned and helpful?). We also questioned them about their recent contacts with the police, and asked those who had called or been stopped by the police to rate their effectiveness and the fairness with which they had been treated. Because views of the police are deeply divided by race and class, our analysis paid close attention to differences among social groups in this regard.

Before community policing began, almost two-thirds of the respondents already averaged a positive score on the police demeanor index. In general, whites perceived that police treated people well even in the early 1990s, and there was not much room for improvement. However, positive perceptions of police demeanor rose by about 10 percentage points among both Latino and African-American respondents and ended on a high note. Perceptions of police responsiveness to community concerns improved steadily. Perceived responsiveness went up the most among African-Americans and Latinos, rising by almost 20 percentage points between 1993 and 1999. The views of whites, which were more positive even before CAPS began, improved by about 10 percentage points. Finally, at the outset Chicagoans were mostly negative in their views of how well police performed their traditional tasks. Over time, the index measuring this aspect of police service improved significantly, rising from a low of 36 percent in 1994 to a high of 51 percent in 2003. Note, however, that for all of these changes, opinion *gaps* between the races closed not at all. While positive trends were apparent, racial polarization around policing issues changed hardly at all. A decade after the program began, the contrast between the general optimism of whites and the still-widespread pessimism of African-Americans was almost as large in 2003 as it had been in the beginning.

Our role as evaluators in *creating* goals for the program is best illustrated by our effort to evaluate beat meetings. Following a series of reports indicating that program implementation was faltering, new meeting requirements were announced that matched line-by-line the evaluation criteria that we had developed to assess them. Meetings were required to have printed agendas, and crime maps and city service reports were to be distributed each time. Distinct segments of the meetings were to be devoted to reviewing progress on problems identified at past meetings and on new crime and disorder problems. Participants were to identify tasks to be performed and a timetable, and officers sought volunteers from the group and assigned them jobs. Both the beat team sergeant and the managing lieutenant had to sign off on the official report that these things had taken place. If all the new requirements actually were met, a beat would have gotten a perfect score in our 1998 study, which concluded instead that they were meeting only half of their goals. This was no accident; beat meeting procedures were redesigned *in order* to get a perfect score from us, and also because our scoring criteria reflected the department's unrealized plan for them.

## CONCLUSION

The structure of the CAPS evaluation was at all points driven by the nature of the program and the character of the city. Because we got involved before the program plan was even finalized, our evaluation team could gather pre-program baseline information, a luxury that many studies of real-world, politically driven projects do not enjoy. The police department chose to phase in the implementation of the program so that they could fix any problems that emerged early, while they were still small in scale. This enabled us to contrast changes in the prototype districts with parallel trends in matched areas of the city, drawing on data collected after the program had been in operation in a few communities. The large and diverse character of Chicago compelled us to design an evaluation that could encompass the entire city. Instead of watching the program unfold in a few selected places, the evaluation would be based on samples of residents, activists, patrol officers, and police beats representing the entire city, so that the findings could describe general trends in Chicago and what was happening in its three major communities.

The evaluation continued, and over the years we learned a great deal about the operation of key elements of community policing. This included how to gain the attention of the community

through effective marketing and what it took to run a productive beat meeting; the organizational and leadership factors associated with effective neighborhood problem solving; and what did not work (for mostly it did not) when it came to involving the city's new immigrants in community policing. However, the strongest conclusions about whether the program itself caused changes in quality of life, fear of crime, and confidence in the police in the city's neighborhoods come from the first two years of the study.

Was Chicago's effort a success? In Skogan and Hartnett (1997) we put Chicago's accomplishments in context by comparing the results of our impact analyses with evaluation findings from other cities. By this accounting, Chicago's success rate was about that of other cities that have conducted carefully evaluated community policing programs.

The baseline for this comparison was a review of a large stack of largely unpublished community policing evaluations, which concluded that these interventions had a success rate of just over 50 percent (Skogan, 1994). The reviewed projects targeted victimization, fear of crime, drug markets, and the perceived quality of police service. They were carried out in experimental neighborhoods in Houston, Newark, Oakland, Birmingham, Madison and Baltimore. Each was evaluated using roughly the same approach that we employed in Chicago: matched comparison areas, two waves of resident surveys, and the analysis of census, crime, and other archival data. To give an example of the findings, fear of crime was a target in all of the projects, and it went down – probably as a result of the program – in half of them. Overall, compared to what happened in the comparison areas, positive changes were recorded in 27 of the 51 outcomes that were monitored. That constituted a 'success rate' of 53 percent.

In Chicago there was evidence of program impact on nine of the 20 top priority problems, for a success rate of 45 percent. As described earlier, we also assessed the impact of CAPS on four clusters of outcome measures for each district: drug and gangs problems, serious crime problems, physical decay problems, and perceived police responsiveness to community concerns. There was evidence of significant program effects for 10 of the 20 clusters, for a success rate of 50 percent.

Thus Chicago hit about the national mark. It fielded somewhat different community policing efforts of varying quality in five experimental districts; the programs in other cities also varied considerably in type and quality. Like in Chicago, some projects were well conceived and well executed, while others did not get very far. In the aggregate they succeeded about half the time.

Whether a success rate for community policing of 50 percent will be pronounced 'a success' or 'a disappointment' is a political rather than a research question. In Chicago it was viewed to be a considerable success, and by early 1995 the program had expanded to encompass the entire city. Running on its popularity, the mayor who insisted on reforming policing has since been re-elected several times, by widening margins and with increasing support among African-American and Latino voters. The democratic processes by which people who initiate programs are held accountable is another way to assess their effectiveness, and by that measure CAPS was successful indeed.

## NOTE

1 All of the reports and other published material from the evaluation can be found at: <http://skogan.org/ChicagoCAPS.html>

## RECOMMENDED READING

In addition to the material cited above, interested readers may wish to consult:

- Foster, A. and Jones, C. (2010) 'Nice to do' and Essential: Improving Neighbourhood Policing in an English Police Force' *Policing: A Journal of Policy and Practice*, 4, 395–402; for a study of the implementation of community policing in the UK.
- Skogan, W. and Frydl, K. (eds) (2004) *Fairness and Effectiveness in Policing: The Evidence*. (Washington D.C., National Academies Press) for a comprehensive review of the field of police research in the United States.
- Weisburd, D. and Braga, A. (eds) (2006) *Police Innovation: Contrasting Perspectives* (Cambridge: Cambridge University Press) for a set of eight 'debates' between leading police scholars over the implementation and effectiveness of recent innovations in policing.

## REFERENCES

- Bennett, Trevor. 1991. 'The Effectiveness of a Police Initiated Fear Reducing Strategy.' *British Journal of Criminology*, 31: 1–14.
- Clarke, Ronald V. and David Weisburd. 1994. 'Diffusion of Crime Control Benefits: Observations on the Reverse of Displacement.' In Ronald V. Clarke (ed.) *Crime Prevention Studies Vol. 2*, Monsey NY: Criminal Justice Press, pp. 165–183.
- McElroy, Jerome E., Colleen Cosgrove and Susan Sadd. 1993. *Community Policing: The CPOP in New York*. Newbury Park, CA: Sage Publications.
- Pate, Antony, Mary Ann Wycoff, Wesley G. Skogan and Lawrence W. Sherman. 1986. *Reducing Fear of Crime in Houston and Newark*. Washington DC: National Institute of Justice and the Police Foundation.
- Quinton, Paul and Rachel Tuffin. 2007. 'Neighbourhood Change: The Impact of the National Reassurance Policing Programme'. *Policing*, 1: 149–160.
- Rosenbaum, Dennis P. (ed.) 1994. *The Challenge of Community Policing: Testing the Hypotheses*. Newbury Park, CA: Sage Publications.
- Sadd, Susan and Randolph Grinc. 1993. *Issues in Community Policing: An Evaluation of Eight Innovative Neighborhood-Oriented Policing Projects*. New York City: Vera Institute of Justice.
- Shadish, William R., Thomas D. Cook and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.
- Skogan, Wesley G. 1994. 'The Impact of Community Policing on Neighborhood Residents: A Cross-Site Analysis.' In Dennis P. Rosenbaum (ed.) *The Challenge of Community Policing: Testing the Hypotheses*. Newbury Park, CA: Sage Publications, pp. 167–181.
- Skogan Wesley G. 2006. *Police and Community in Chicago: A Tale of Three Cities*. New York: Oxford University Press.
- Skogan, Wesley G. and Susan M. Hartnett. 1997. *Community Policing, Chicago Style*. New York: Oxford University Press.
- Skogan, Wesley G., Susan M. Hartnett, Jill DuBois, Jennifer T. Comey, Marianne, Kaiser and Justine H. Lovig. 1999. *On the Beat: Police and Community Problem Solving*. Boulder, CO: Westview Publishing Co.