

Kent R. Kerley (ed.) Policing and Program Evaluation.
Upper Saddle River, NJ: Prentice Hall, 2005

Evaluating Community Policing in Chicago*

Wesley G. Skogan



INTRODUCTION

This chapter examines the impact of Chicago's community policing experiment on neighborhood problems. It comes at a time when there is interest around the world in developing new approaches to policing and in systematically assessing how well they work. These twin interests call for two kinds of evaluations. *Process evaluations* examine program design and implementation. They document a program's "theory," or how its developers thought it was supposed to work. Process evaluations also document the actual implementation of the program, for there is often a gap between plans and reality. *Impact evaluations* analyze the effects that a program had on the problems that it targeted, and if the program had unexpected or unintended consequences as well. The strength of an impact evaluation is dependent on its design and how well it measures what the program might accomplish.

This chapter describes an evaluation of a community policing program in Chicago. It was both a process and an impact evaluation, although this chapter will focus on what our team (made up of academic researchers headquartered at Northwestern University) found out about the consequences of the program for the city's residents. The first section describes the program and the evaluation. The next documents the analytic approach that was adopted that enabled us to assess the impact of community

*This project was supported in part by grants from the National Institute of Justice, the Illinois Criminal Justice Information Authority, and the John D. and Catherine T. MacArthur Foundation. Points of view or opinions contained within this document are those of the author and do not necessarily represent those organizations' official positions or policies.

policing in five city neighborhoods. The third section presents what we found about the impact of the program. It looks at the impact of community policing on a variety of community problems and illustrates how different ways of measuring those problems pointed to the same conclusions. The next section deals with a recurring problem in crime prevention projects: displacement. It examines whether crimes were actually prevented or if they just moved somewhere else. The final section summarizes the findings and comments on the general features of evaluation projects.

COMMUNITY POLICING IN CHICAGO

Beginning in May 1993, Chicago's first community policing program, Chicago's Alternative Policing Strategy (CAPS), was field-tested in five diverse experimental districts. During the program development period, those districts' commanders, senior department executives, and civilian planners developed the policies and procedures necessary to make community policing a workable concept in Chicago. From the beginning, they adopted a problem-solving orientation. The police department's official description of the program stated:

It is this focus on prevention through a stronger government-community partnership that holds real hope for addressing some of the City's most difficult neighborhood problems—and for doing so in a way that is far less expensive than constantly reacting to those problems after the fact. (Chicago Police Department, 1993: 12)

Organizational changes were made to support problem solving. In the experimental districts, patrol officers were divided on a rotating basis into beat teams and rapid response units. Tasks were assigned to these units using formulas designed to free up sufficient time for beat officers to attend meetings in their area and work with community organizations and agencies. Changes in 911 dispatching allowed them to spend most of their time working on their beats, while rapid response officers took over charge of excess 911 calls. Beat meetings were held on a monthly basis in each of the 54 small police beats in the five experimental areas. These meetings were regular public gatherings of small groups of residents and the beat team members. They met in church basements and schoolrooms to identify and discuss neighborhood problems and (hopefully) to develop joint police-citizen action plans for resolving them. Other police units were being decentralized so that local commanders had control over various plainclothes officers and youth officers and could integrate their efforts with plans being developed at the grassroots level. Each of the police districts formed a civilian advisory committee to advise the commanders and to help mobilize problem-solving resources in the community. Recognizing that problem-solving policing needs the support of a wide range of agencies, procedures were developed to link the delivery of city services to requests for assistance from the beat teams. Beat officers were able quickly to mobilize building inspectors, street cleaners, repair trucks, and other city resources. They could easily get abandoned cars towed away, empty buildings boarded up, graffiti painted over, and vacant lots cleared of trash.

Most of the important elements of Chicago's community policing program to this day were developed first in the experimental districts. In the experimental areas, the coordination of city services seemed to work. Beat officers had the time they needed to focus

on identifying local problems and the resources they needed to solve them and to work with neighborhood residents. All of the officers serving in the experimental districts received four days of special training, the district advisory committees got to work in all five areas, and hundreds of public meetings were held in the police beats in the experimental districts. In part because representatives of police and civilian employee unions were included in making many important decisions, the program avoided becoming an object of labor-management controversy. Then, in the winter of 1994–1995, the program expanded to encompass the entire city. Over time the city's program developed into one of the decade's most substantial and closely watched experiments in policing.

Our team of academic researchers has been involved in evaluating Chicago's community policing effort since its planning phases. A great deal of our effort has focused on understanding the origins of the program, its initial planning, and the ways it has been implemented. Two books (Skogan & Hartnett, 1997; Skogan, et al., 1999) describe the program in detail, including such important details as how the city paid for it. This chapter describes our efforts to evaluate the impact of the program on neighborhood problems. Following the program from its inception through its implementation in the experimental areas, we used a quasi-experimental approach to evaluate its impact at the district level.

RESEARCH DESIGN

Some of the evidence about the program's impact was drawn from survey interviews with random samples of residents of Chicago's neighborhoods. The first survey took place during April and May 1993, and it was completed before the program began. Respondents to this survey were reinterviewed 18 months later in order to measure changes in their perceptions of crime and other neighborhood problems. Some respondents lived in the experimental police districts, and others resided in matched areas that were not involved in the new program. The effects of the program were assessed based on differential changes in the views and experiences of the two groups over time, assuming those changes could be logically linked to actual elements of the program. This quasi-experimental approach to evaluating police programs has been used in many countries, including Australia (Criminal Justice Commission of Queensland, 1995), Britain (Bennett, 1991), Canada (Hornick, Burrows, Phillips, & Leighton, 1993), and the Netherlands (Spickenheuer, 1983).

To prepare for the surveys, 1990 census data were used to select sections of the city that closely matched the demography of the five newly announced experimental areas. The matching factors for selecting these comparison areas were race, home ownership, and percentage of residents living in large buildings. Crime rates were not used for matching purposes, but each pair of evaluation areas turned out to have similar levels of officially recorded crime.

In this evaluation, changes in conditions in the comparison areas during the course of the project were used to represent what would have happened in the experimental districts if there had been no program because community policing was not extended to cover the rest of the city until the end of the experimental period. This is one of the critical differences between quasi-experimental evaluations and simple "before-and-after" descriptions of crime trends in program areas. With before-and-after designs, one has no idea what other factors may have contributed to changes that occur besides the addition to the program. Examples might include extremes in weather (crime is very seasonal), changing economic

conditions, coverage of crime and the police in the media, and programs developed by other agencies. Changes over time in the matched comparison areas provided a benchmark against which changes in the program areas could be contrasted.

In a true experiment, the comparison areas would have been the “control groups” against which changes in the program areas could be compared. However, none of these areas were randomly selected (a feature of a true experiment), 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, political forces, and other neighborhood-level factors that influenced either the experimental or comparison area, but not both. To a certain extent these factors can be controlled for statistically, but only if they can be recognized and measured accurately in advance.

In a true experiment, we also would have controlled the intervention, determining what happened in each district and how much “treatment” each area received. However, this study took place in the real world. Local police and political leaders selected the experimental districts, and there was hot competition to be on the list. The police department then developed a workable program in those areas, so that what the intervention actually looked like and how effectively it was delivered were important evaluation issues rather than factors controlled by the evaluators. And of course there were only five pairs of evaluation and comparison areas, not the many subjects usually involved in experiments. As a result, I have tried to interpret the results modestly. They provide a background against which to judge the extent of the changes we observed in the evaluation areas, and in every case we were not confident that the program had an effect unless we understood how the specific program activities that took place there could have produced the effects that we observed in our evaluation data.

The first evaluation survey (“Wave 1”) was conducted in the prototype districts and matched comparison areas before the program began. Residents were then reinterviewed (“Wave 2”) in another 14 to 17 months, depending on the area. All of the interviews were conducted by telephone in English or Spanish. A total of 1,506 people were interviewed on both occasions; there were an average of 180 reinterviews each in the prototype districts and 150 in their comparison areas. Because this was a telephone survey, households without phones could not be included. As a result, the survey underrepresented those who were poor and less educated and people who rented rather than owned their homes.

The completion rate for the first survey was 59 percent, the reinterview rate when we attempted to recontact the original respondents was 60 percent. An analysis of who we succeeded in recontacting indicates that men, Hispanics, younger respondents, those with less education, and renters were less likely than others to be recontacted successfully. Like those who did not have telephones at all, these groups are generally less supportive of the police and less likely than most to get involved in community policing, so their underrepresentation might exaggerate the apparent effectiveness of the program. However, both telephone ownership and survey nonresponse patterns were similar for both the target and the comparison areas. Since any sample bias was about the same for both groups, it could not account for the program effects documented by the surveys.

The respondents were selected the first time using a combination of sampling techniques to reach respondents who were living in these relatively small areas. Half of the

respondents were selected at random from telephone directory listings of households in each of the targeted areas. The other half were contacted by calling randomly generated telephone numbers. The second approach ensured that households that did not have listed phone numbers would be included in the data. The decision to employ a “dual sample frame” reflected the difficulty and expense of randomly contacting households concentrated in small areas. Computer-generated telephone numbers would have produced a more accurate picture of conditions there, but a huge proportion of the calls would have contacted households that were not actually in the target areas.

The analyses that follow compare the results of the two waves of interviews. We used the data in several ways. First, before-and-after data for the experimental and comparison areas monitored changes in each pair of places. When there was a change in a prototype but no comparable shift in its comparison area—or vice versa—it can be evidence that community policing made a difference, assuming we could reasonably link it to specific elements of the program. Second, we were concerned throughout with issues of race and class, and we examined the impact of the program within subgroups in the population. To do this, respondents in all of the analysis areas were combined, and the data were examined for evidence of race or class differences in changes in the prototype and comparison areas. As will be explained below, we used two different approaches for selecting the neighborhood problems to be examined. First, we examined the impact of the program on what respondents identified as the four biggest problems in each area. But because this presented a methodological problem, we also examined index scores combining clusters of similar problems.

Most of the discussion in this chapter focuses on survey ratings of the extent of various problems in each community. This is appropriate because many of the problems that concerned neighborhood residents could not easily be counted as discrete incidents, and most either are not captured by any official record-keeping system or are very poorly recorded when they are. For example, street drug dealing appears in official statistics only when arrests are made, and those numbers simply did not reflect the wide-open drug dealing that plagued several of our study neighborhoods. Graffiti was only rarely reported to police (most people do not connect it with making a 911 call) and was not dealt with very intensively when it was. No one in Chicago kept official statistics on problems such as junk-strewn vacant lots, but these lots concerned residents of some of the study neighborhoods a great deal.

Using surveys enabled us to employ uniform and comparable measures of the impact of a broad mix of problems and to go directly to the public rather than rely on secondhand sources of information about the quality of their lives. There is some evidence that individuals are fairly reliable informants about the extent of these problems (Skogan, 1990). However, for some issues it is also possible to examine alternative measures. For example, our surveys included systematic measures of the victimization experiences of those we questioned. Respondents were asked a list of 22 screening questions, plus some follow-ups, about crimes that had affected them and their households. This enabled us to assess changes in crimes such as burglary, vandalism, street robbery, and auto theft. We also examined time-series trends in officially reported crime at the district level when they could be compared to the problems identified by neighborhood residents as priority issues.

The surveys also 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, most importantly on how

responsive police were to community concerns and the problems that really bothered people who lived there. We also screened for recent contacts with the police, and we asked people 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.

IMPACT ON NEIGHBORHOOD PROBLEMS

The first look at our evaluation data focuses on the issues identified by neighborhood residents as the most serious problems facing their communities. This was appropriate because Chicago's community policing plan did not specify particular targets for the program or speculate about how much impact it might have. Rather, mechanisms were created to encourage police to be more responsive to conditions in the small areas that they served. Chicago's plan assumed that the problems they identified would be different in different places and that the problems would vary in magnitude from place to place. One goal of community policing is to open departments up to local input so that they can effectively discern these kinds of differences among neighborhoods and tailor their operations to respond to them. Likewise, our evaluation also needed to be responsive to these local variations.

Focusing on a fixed number of problems in each area also protected us against random error in the measures. In the Chicago study, there were five program areas but several dozen outcome measures. If we looked for 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 to look only at a clearly defined set of problems was important.

In the surveys, respondents were quizzed about 18 specific issues that the evaluators thought—before the program began—might 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.” Our first analysis of the impact of the program examined the 4 biggest problems that residents of each area nominated in their first interview. Figure 3–1 presents those and illustrates the diversity of both the nature of the problems and their intensity across the five experimental districts. Our evaluation then tracked the ratings given these issues a year or so later when we reinterviewed residents. This analysis let residents set the agenda for the evaluation through their expressions of concern about neighborhood conditions.

Two problems on the list were of virtually universal concern: “Street drug dealing” was one of the top-ranked problems in every area we studied, and “shooting and violence by gangs” was one of the leading problems in four of the five prototypes. Ratings of drug and gang violence clustered together tightly when people assessed conditions in their neighborhoods; at the individual level, the correlation between the two measures was 0.72. Gangs and drugs are challenging issues that were 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 other style.

Otherwise, a wide range of problems was identified as particularly troubling. In two areas car vandalism was near the top of the list, and in two others household vandalism ranked high. Problems with “people being attacked or robbed” were also rated high in two

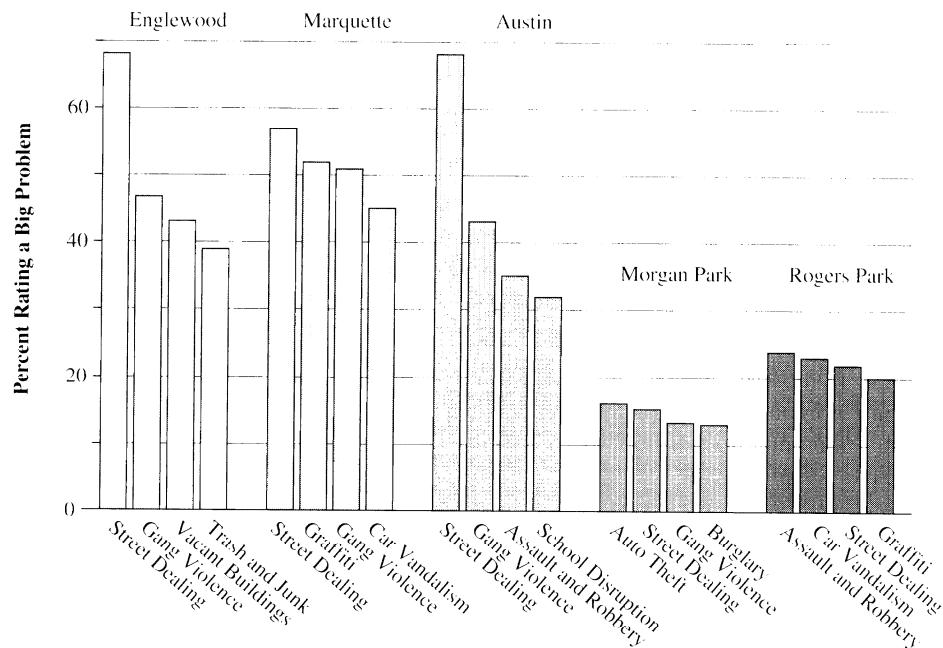


FIGURE 3-1 Top-rated problems in the experimental districts.

areas. 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 one district.

It is important to note that the initial level of these biggest problems also varied considerably from district to district. Issues that loomed large in some areas were scarcely problems at all in other districts. For example, street drug dealing was rated a big problem by 60 percent or more of residents of Englewood, Marquette, and Austin. On the other hand, only about 13 percent of the residents of Morgan Park and 20 percent of those we interviewed in Rogers Park thought this was a big problem, even though it was one of the area’s top-ranked issues before community policing was initiated. In Morgan Park, burglary was a top-ranked problem, but only 10 percent of residents gave it a high rating. Morgan Park was a fairly well-off community, and there was not much room for improvement on many of our outcome measures.

Figure 3-2 presents illustrative findings for one of Chicago’s experimental areas: Englewood. Englewood is an extremely poor and largely African-American neighborhood. During the early 1990s, it had one of the highest homicide rates in the city. Before the program began, more than 60 percent of the residents of Englewood rated street drug dealing a big problem, and gang violence was the number three problem there. But the two other most highly ranked problems turned out to be quality-of-life issues. Problems with “vacant lots filled with trash and junk” stood near the top of the list, and so did the large number (600, by the district commander’s estimate) of abandoned buildings that plagued the district. In fact, in four of the five experimental areas, two of the top four problems were quality-of-life concerns rather than serious criminal offenses.

Figure 3-2 presents problem ratings for Englewood and its comparison area, for both the Wave 1 and Wave 2 surveys. The statistical significance of each change over time is presented

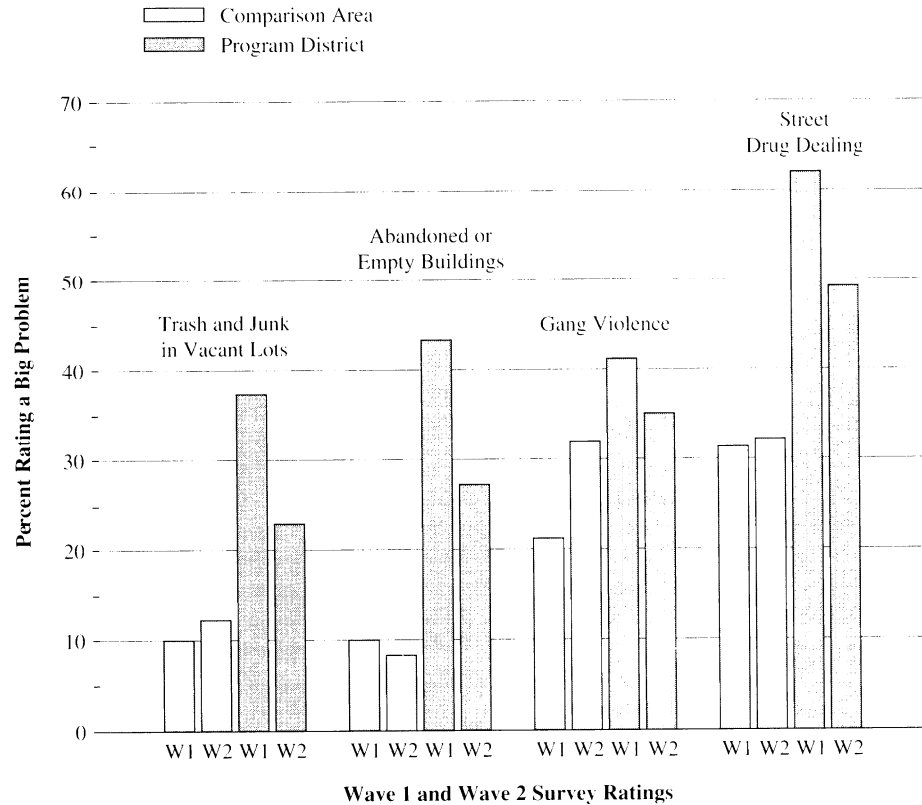


FIGURE 3-2 Changes in Englewood's problems.

as well. However, the conclusions about changes over time presented in this chapter are based on the results of statistical analysis of the data using repeated-measures analysis of variance, which focuses on the significance of differential changes in the means of the outcome measures over time. This analysis used the full range of the measures, not just the "big problem" percentages that are illustrated in Figure 3-2.

The findings for Englewood can be summarized as follows. All four of the community's biggest problems declined while none went down significantly in Englewood's comparison area. The problem of street drug sales was ranked a big problem by 62 percent of Englewood residents in 1993 but by only 49 percent in 1994. Abandoned building problems dropped from 43 percent to 27 percent, and problems with litter from 37 to 23 percent. Gang violence was down only modestly (the percentage who thought it was a big problem declined from 41 to 35 percent), but it increased significantly in Englewood's comparison area.

These findings reflect the relative vigor with which Englewood used the city service request process. Englewood residents and police were extremely successful at mobilizing city services to respond to both of its decay problems. During the 16 months ending in August 1994, they generated 1,314 service requests to attend to abandoned buildings and 2,379 requests for special service from the Department of Streets and Sanitation. In both cases Englewood's service request count ranked number one among the five experimental

areas, both absolutely and relative to the size of their populations. There was one request to deal with an abandoned building for every 85 Englewood residents. As Figure 3–2 illustrates, perceptions of the extent of both problems went down significantly in Englewood during that period. In addition, gang and drug problems were the focus of marches by community members, which were organized by a coalition of local churches. The district's commander was easily the most charismatic and energetic of the group.

When common crimes were involved, there were two alternative measures of the extent of neighborhood problems that could be directed toward the analysis of change in the experimental districts. Common crimes were among the top-ranked issues in Morgan Park (auto theft and burglary), Rogers Park (robbery and assault), and Austin (robbery and assault). In each case, both official crime figures and survey reports of victimization could be compared to the problem ratings people gave these crimes. For official crime statistics, we examined the average monthly crime rate for the 17 months preceding the program (that is, from December 1991 through April 1993) as compared to the 17 months following the inauguration of community policing in the experimental districts. To measure crime victimization, we calculated the percentage of survey respondents who were victimized during the year preceding each wave of surveys. Because assault, robbery, and car theft could take place virtually anywhere, we asked victims if the incident took place in their neighborhood. Only crimes that took place in the experimental district are examined here.

Table 3–1 compares the results of three different measures of trends for the two common crimes that were identified as among the biggest problems in Morgan Park. In the case of auto theft, popular perceptions of the extent of the problem went down by one-third, victimization dropped to more than half its original rate, and officially recorded crime dropped by 26 percent. On the basis of this, it appears likely that auto theft went down in Morgan Park. On the other hand, we are not as confident that burglary declined. The victimization measure dropped to 60 percent of its former figure, but neither popular impressions of the extent of the problem nor the official crime count dropped much at all. In contrast, similar data for Morgan Park's comparison area indicated no decline in crime.

TABLE 3–1 Three Measures of Crime Trends in Morgan Park

Crime Type	Percent of "Big Problem"	Official Crimes per Month	Percent of Victims
Auto theft			
Before	15	146	8.0
After	10	108	3.2
	p = 0.02	–26%	p = 0.02
Burglary			
Before	10	107	10.1
After	8	102	6.0
	p = 0.11	–5%	p = 0.11

Note: Official crimes per month average a 17-month period before CAPS and 17 months following CAPS implementation. Tests of significance are for before-and-after changes in problem ratings and victimization. Percentage change is given for monthly recorded crime.

IMPACT ON PROBLEM CLUSTERS

The approach to evaluating the impact of community policing employed above—following the trend of each experimental area’s most highly ranked problems—risks a methodological flaw. Because we examined only problems that were ranked highly in the Wave 1 survey, there was a potential bias toward concluding that they would decline in the next survey. This bias is called a regression artifact, and in this case it threatened to cause us to come to incorrect conclusions about the impact of the program.

Because every social measurement includes a significant random error component, some of the most important neighborhood problems that we identified in Wave 1 may have been highly ranked due to chance fluctuation. If this was true, there was a high likelihood that chance factors would then bring those scores down when they were measured again. This is called regression (e.g., movement) back toward the mean (see Cook & Campbell, 1979). Because problem measures for the comparison areas were not selected for their extreme values, they were less likely to decline (other things being equal) by chance. This tendency toward misleading differential change threatens evaluations that focus on changes in program and comparison areas over time.

A related problem is that selecting the highest-rated problems in the program areas for examination means that often they will be much less of a problem in the comparison areas. In the language of evaluation research, the pairs of areas will be “unmatched at the pretest.” This can be seen in Figure 3–2: In every case, Englewood’s biggest problems overshadowed those in its comparison area.

One defense against regression artifacts is to examine the fate of all of the problems examined in the survey, not just the most extreme. This section reports the results of such an analysis. However, this approach to gauging the impact of community policing on neighborhood problems is conservative and actually biased against finding any effect of the program (for reasons discussed below), while avoiding the potential bias raised by examining only an area’s worst problems. In this section, results are reported of summary scores that combine assessments of multiple closely related issues into measures of general clusters of problems. Through statistical analysis, three clusters of neighborhood problems were identified: we dubbed these major crimes, gang and drug problems, and signs of physical decay. The clusters encompass almost all of the neighborhood problems that were assessed in the surveys. Because we did not select the component problems that were particularly high at Wave 1, they should not be biased downward at Wave 2. The actual measures should also be more statistically robust because they combine responses to multiple questions; this helps cancel out random measurement error.

This approach brings with it other problems, however. An analysis of the impact of community policing on general clusters of problems is useful for drawing general lessons about the impact of community policing, but the summary scores take into account topics that were not the focus of public concern. At Wave 1, some of the problems we asked about were rated scarcely above “no problem” in many of the experimental areas. They were problems elsewhere in Chicago, but not in those districts. To use them in an evaluation presents a difficult test because problems that were not very large in the first place do not have very much room to show improvement. This is called a floor effect, because at Wave 2 the measures could not drop below the bottom of the survey scale (a rating that an issue was “no problem”). For some problems in some areas, this was not very far away. This biased the analysis a bit against identifying any program effects.

The problem clusters and their specific components were as follows:

- **Major crimes.** This cluster included problems with car vandalism, auto theft, burglary, street crime, and rape or other sexual assaults. The reliability of this measure was 0.85.
- **Gangs and drugs.** This group included problems with street drug dealing as well as shootings and violence by gangs. The correlation between these two measures was 0.72.
- **Physical decay.** This cluster included problems with vacant lots filled with trash and junk, abandoned cars in the streets and alleys, abandoned houses or other empty buildings, and graffiti. The reliability of this measure was 0.75.

To examine the impact of community policing on these problems, we employed the quasi-experimental analysis strategy described in the previous sections. Instead of using responses to questions about individual problems, however, we averaged responses to all of the questions in each cluster and used the summary score. Statistical change measures were created by controlling for each individual's Wave 1 score, measures that were collected before the program began. Ten personal and economic characteristics of each respondent were controlled for as well. The analysis then examined differences in Wave 2 scores between the experimental and matched comparison areas. For graphical display, Figure 3–3 presents problem cluster scores for the major crime problems scale for each of the experimental and comparison areas, for both the Wave 1 and Wave 2 surveys.

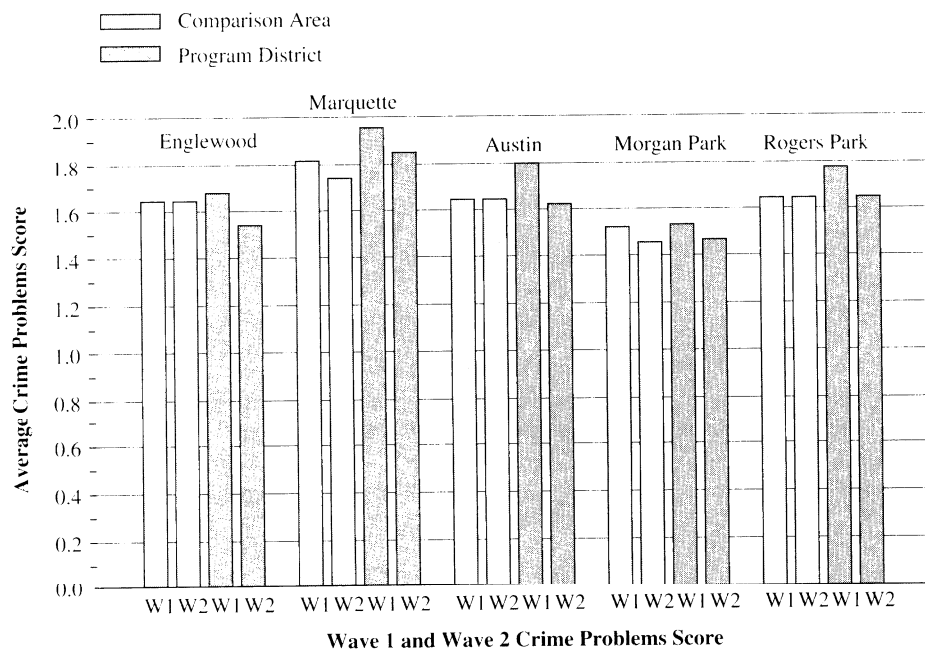


FIGURE 3–3 Changes in crime problems.

Figure 3–3 depicts the average major crime score in each area before community policing began and again 14 to 17 months later. By this accounting, crime problems declined significantly in each of the five experimental areas during that period. The decline was smallest in Morgan Park, where issues on the list presented above already ranked relatively low in intensity (averaging about halfway between “no problem” and “some problem”). Statistically, the apparent decline of major crime problems in Marquette and Morgan Park did not greatly outpace parallel shifts in their comparison area, so it is chancy to attribute these declines to community policing. In the three remaining areas, statistical analyses confirm that declines in major crime problems in the prototypes outweighed any changes in their comparison areas, and the declines could well have been caused by the program.

WERE PROBLEMS JUST DISPLACED?

While there is evidence of significant declines in a wide range of problems in Chicago’s experimental police districts, there lingers the possibility that some or all of them were simply displaced elsewhere rather than truly being resolved. In fact, the possibility of displacement questions the results of virtually every crime prevention program, but rarely are researchers in a good position to assess this outcome. The possibilities for displacement are wide ranging. Depending on the problem, displacement might be geographic—the program may push it into another neighborhood or somewhere down the highway. Geographic displacement is the most widely discussed alternative, but the intervention might instead displace it in time, to a period when residents or police are not patrolling, or displace it in activity, when offenders might switch to another activity, with no guarantee that the neighborhood will be better off due to their new “line of work.”

Of all these possibilities, we were able to consider only the possibility of geographic displacement, and even then only tentatively. To monitor displacement, we identified the instances in which problems that we judged to be the most displaceable—gang violence, street drug sales, and street crime—declined significantly. This occurred in Englewood, Austin, and Rogers Park. We then identified potential displacement zones around these experimental districts that were areas where we had also conducted evaluation surveys (these 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 Englewood, Austin, and Rogers Park. In our surveys, 78 respondents lived in zones that were potential displacement areas for drug problems and gang problems, and 77 lived in areas that were the potential targets of displaced street crime. Respondents living in areas deeper in the comparison areas continued to be used as “control group” cases.

We then examine Wave 1-Wave 2 changes in the three “displaceable crimes,” accounting separately for changes in the prototypes, the displacement areas, and the (now shrunken) comparison areas. We examined simple change scores for the three areas and used multiple regression combining all 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 a displacement of crime. In the aggregate, gang problems went up in the relevant comparison areas but remained steady in the displacement areas near the prototypes. Street crime remained steady in the

comparison areas but went down in the displacement areas adjoining the prototypes. There were no apparent shifts in drug problems other than their significant decline in Austin and Englewood. None of the gang or street crime changes in the displacement zones were statistically significant, and the number of survey respondents living in the displacement areas was not large, so we are not tempted to claim that the benefits of community policing spilled over from the prototypes into surrounding areas, but in this case that hypothesis is at least as credible as fears about displacement.

CONCLUSION

This evaluation examined the impact of Chicago's community policing program using a quasi-experimental design that tracked trends in problems at the level of police districts. We did this for several reasons. One was that the program was designed to facilitate the identification and resolution of neighborhood problems locally, and we knew that the problems that concerned neighborhood residents would vary considerably from place to place. We also anticipated that each district would devise a somewhat different program over the course of the experimental period because no one in Chicago had a clear idea about what community policing should look like when the experiment began. Therefore, our evaluation needed to speak separately about each district and about each district's problems.

We also had to conduct an evaluation that could encompass a very large program at an affordable cost. One of the most significant features of Chicago's experiment is that it was a very large one. If the five districts had been independent, they would have been one of the largest cities in the country. 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 a bit bigger than Cleveland's and far larger than the departments serving Atlanta, Miami, Kansas City, or New Orleans. While micro-level studies of individual police teams or specific problem-solving efforts might have been informative, it would have been difficult for us to conduct them in sufficient number or in enough locations to generalize about the areas that we were studying, at the level at which project management was focused. What people wanted to know was what happened in the districts.

Instead, we examined trends both in the problems that residents of the experimental districts identified as their greatest concerns and in summary measures that we created through statistical clustering. There was evidence of improvement in the lives of residents of every program area. As illustrated above, a survey-based measure of the extent of major crime problems went down in all five experimental districts, although they declined in two comparison areas as well. The victimization component of the survey pointed to decreases in auto theft in one district and street crime in another. Reports of drug and gang problems declined in two of the worst areas, as did perceptions of the extent of physical decay. Graffiti went down significantly in the area where it presented the biggest problem.

Was Chicago's problem-solving effort a success? In the book *Community Policing, Chicago Style* (Skogan & Hartnett, 1997), we attempt to put Chicago's accomplishments in context by comparing the results of all of our impact analyses with those from other cities. By our accounting, Chicago's success rate was about equal to that of other cities that have conducted carefully evaluated community policing programs.

The baseline for this comparison was a review of community policing evaluations (Skogan, 1994). This review found that the programs had a success rate of just over 50 percent. Those projects targeted victimization, fear of crime, casual social disorder (e.g., loitering, panhandling, public drinking, and street harassment), drug availability, and 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 collection of census, crime, and other official 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 we evaluated the impact of the program in two different ways, so we calculated two success rates. Both of those approaches have been described in this chapter. Our first approach was to monitor the fate of the 4 biggest problems in each of the five districts, as identified by the people who lived there. There was evidence of program impact on 9 of those 20 problems, for a success rate of about 45 percent. Our second approach to assessing the impact of community policing on the neighborhoods was to examine changes over time in four clusters of outcome measures: drug and gang problems, serious crime problems, physical decay problems, and perceived police responsiveness to community concerns. This also (quite accidentally) totaled 20 outcome measures. There was evidence of significant program effects in 10 instances, for a success rate of 50 percent. Interestingly, even though tracking individual problems risked regression biases and tracking cluster scores risked floor effects, the two approaches came to the same overall conclusion.

Thus, Chicago seemed to be at about the national mark. Chicago fielded somewhat different community policing efforts of varying quality in five experimental districts; in other cities, some projects were well conceived and well executed, while others did not get very far. In the aggregate, they all 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 as a considerable success.

This judgment reflected in part the sheer difficulty in implementing these programs in the first place. There are many obstacles to innovation in policing (Sherman, 1986). Adopting problem-oriented or community policing calls for a reallocation of resources. Officers will be asked to perform new tasks, often at different hours of the day and in different neighborhoods. Many will not be interested in doing so, and the new styles of policing risk being marginalized in the organization. Because a problem-solving orientation calls for new work styles, all of the officers will have to be extensively trained. They will also need time to work with the community; to have this time, they will have to be freed from responding to emergency calls. Working closely with the community and attempting to respond to its priorities will inevitably change the nature of their work, for the public is concerned about many issues that do not fall into traditional crime categories (Moore, 1992).

Chicago faced all of these problems and succeeded well enough in its experimental areas that the program later was expanded to encompass the entire city. The city’s mayor and the department’s senior managers made it clear that community policing was their first priority, and that it was going to be the stance of the entire organization. Almost a thousand new police officers were hired to staff the city’s community policing program and to ensure

that emergency calls would be handled at the same time. All of the city's uniformed officers were trained in problem solving, and their supervisors received more management training. The close integration of police with the city's other service agencies ensured that officers throughout the city could respond to a broad range of community problems by working with those partner agencies.

REFERENCES

- BENNETT, T. 1991. The effectiveness of a police-initiated fear-reducing strategy. *British Journal of Criminology*, 31, 1–14.
- CHICAGO POLICE DEPARTMENT. 1993. *Together we can*. Chicago: Chicago Police Department.
- COOK, T. D., & CAMPBELL, D. T. 1979. *Quasi-experimentation: Design and analysis issues*. New York: Houghton Mifflin.
- CRIMINAL JUSTICE COMMISSION OF QUEENSLAND. 1995. *Beat policing: A case study*. Brisbane, Australia: Criminal Justice Commission of Queensland.
- HORNICK, J. P., BURROWS, B. A., PHILLIPS, D. M., & LEIGHTON, B. 1993. An impact evaluation of the Edmonton neighborhood foot patrol program. In J. Chacko & S. E. Nancoo (Eds.), *Community policing in Canada* (pp. 311–332). Toronto: Canadian Scholars' Press.
- MOORE, M. H. 1992. Problem solving and community policing. In M. Tonry & N. Morris (Eds.), *Modern policing—crime and justice: A review of research*, vol. 15 (pp. 99–158). Chicago: University of Chicago Press.
- SHERMAN, L. 1986. Policing communities: What works? In A. J. Reiss & M. Tonry (Eds.), *Communities and crime—crime and justice: A review of research*, vol. 8 (pp. 343–386). Chicago: University of Chicago Press.
- SKOGAN, W. G. 1990. *Disorder and decline: Crime and the spiral of decay in American cities*. New York: Free Press.
- SKOGAN, W. G. 1994. The impact of community policing on neighborhood residents: A cross-site analysis. In D. P. Rosenbaum (Ed.), *The challenge of community policing: Testing the promises* (pp. 167–181). London: Sage.
- SKOGAN, W. G., & Hartnett, S. M. 1997. *Community policing, Chicago style*. New York: Oxford University Press.
- SKOGAN, W. G., et al. 1999. *On the beat: Police and community problem solving*. Boulder, CO: Westview.
- SPICKENHEUER, H. 1983. *Foot patrols and crime prevention instruction in Amsterdam-Ostdorp*. The Hague: Ministry of Justice, Research and Documentation Centre.