DOCUMENT RESUME

ED 174 743 UD 019 723

AUTHOR Fowler, Floyd J., Jr.
TITLE The Evaluation of the Hartford Experiment: A Rigorous, Multi-Method Effort to Learn Something. Revised.
SPONS AGENCY Law Enforcement Assistance Administration (Dept. of Justice), Washington, D.C.
PUB DATE 24 Jan 78
GRANT 73-NI-99-0044-G; 75-NI-99-0026
NOTE 17p.; Paper prepared for the Second National Workshop on Criminal Justice Evaluation (November 20-21, 1978); Prepared at the Center for Survey Research, Boston, Massachusetts

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Community Attitudes; Community Programs; Crime; *Evaluation Methods; *Program Evaluation; *Research Methodology; *Social Science Research; Urban Areas
IDENTIFIERS *Hartford Project

ABSTRACT

This paper focuses on problems faced and lessons learned in the evaluation of a community crime-reduction program. The program ("Hartford Project"), designed to reduce residential burglary, street robbery, and the fear of those crimes, is briefly described. A multimethod approach to evaluation, which considered each variable in at least two different ways and utilized both quantitative and qualitative information, is explained as it evolved in response to the characteristics of the Hartford Project. Demographic factors which might have had an impact on the evaluation results are also considered. Although the Project is tentatively concluded to have been a success, limitations of the evaluation are pointed out. The importance of good methodology in social science evaluation research is stressed. (GC)

The Evaluation of the Hartford Experiment: A Rigorous, Multi-Method Effort to Learn Something

by
Floyd J. Fowler, Jr.

Center for Survey Research
a facility of
The University of Massachusetts-Boston and the
Joint Center for Urban Studies of M.I.T. and Harvard University

Prepared for Second National Workshop on Criminal Justice Evaluation
November 20-21, 1978

This project was supported by grant numbers 73-NI-99-0044-G and 75-NI-99-0026 awarded by The Law Enforcement Assistance Administration, U.S. Department of Justice, under the Omnibus Crime Control and Safe Streets Act of 1968, as amended. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Revised - 1/24/78

Introduction

Evaluation means many different things. The goals of the evaluation of a program can include:

a) describing the activities;

b) assessing the impact of the program, the way things are different because of the program;

c) learning about the reasons for the program's success or failure.

Usually some information is gathered or collated. The amount and type of information collected, as well as the methodological rigor, varies, of course, from project to project.

The Hartford project was complex, as is usual for environmental design programs; therefore, it was relatively difficult from an evaluation design point of view. The goals of the evaluation included all three of those listed above: detailed description of the programs implemented, an assessment of the program impact on crime and fear, and, most important, an effort to further general knowledge about crime reduction or control. The design was comparatively elaborate and the methods were comparatively rigorous.

For these reasons, the evaluation of the Hartford experiment provided an unusual opportunity to learn about some strategies for evaluation that were successful and may be useful in other evaluations. The purpose of this paper is to present some of the lessons that can be learned.

The Nature of the Program

In order to understand the research, it is first necessary to understand the program.

The Hartford Project was an experiment in how to reduce residential burglary and street robbery/pursesnatch and the fear of those crimes in an urban, residential neighborhood. Its most distinctive feature was its integrated approach to crime control: police, community organization, and physical design changes were all used

to increase the willingness and ability of residents to control the neighborhood to reduce criminal opportunities.

The initial planning for this project occurred in 1973. Analysis of the crime in the area was undertaken by an interdisciplinary team. Its task was to understand the way residents, potential offenders, police and the physical environment interacted to create criminal opportunities; and to design inexpensive strategies that could be quickly implemented to intervene in the pattern of rising crime.

A principal conclusion of the analysis was that a number of features of the physical environment were working to destroy the residential character of the neighborhood. Cars and pedestrians passing through the area dominated the streets and depersonalized them. The streets belonged more to outsiders than to residents, creating an ideal environment for potential offenders.

Based on this analysis, a lengthy planning and implementation period ensued. In 1976, a three-part program was fully implemented that included:

a. closing and narrowing streets as a main strategy for reducing outside traffic and increasing the residential character of the neighborhood.

b. instituting a neighborhood police unit with strong relationships with the residents, and

c. creating and encouraging area organizations to work with the police and to initiate resident efforts to improve the neighborhood and reduce criminal opportunities.

Five features of the experiment are particularly important because they complicated the evaluation.

1. The program was implemented in only one neighborhood area, which had a population of approximately 5,000 people. Therefore, there was only one test of the concepts and ideas

2. As noted above, one essential component of the Hartford experiment was its multi-faceted nature. Perhaps the cornerstone of the project was the street changes, by which the planners hoped to limit vehicular traffic in the neighborhood. However, the police and community organization components of the project were important

as well. Each was seen as a potential catalyst to resident initiatives to crime control, both formal and informal. Describing the implementation and, more importantly, assessing the significance of each program component added considerably to the complexity of the project.

3. A related but different point is that the way the program was supposed to reduce crime and fear was complex and involved a chain of events. The fundamental premise of the program was that the residents themselves, through their informal efforts, could reduce crime, and thereby fear, by taking control of events in their neighborhood. Each of the program components was intended to increase the ability or willingness of the residents to control the neighborhood. Such a model is complicated conceptually and analytically.

The best example of this complexity is the role of the street closings in crime control. Many residents, and even some of the police, could never get over the notion that the purpose of the street closings was to keep out offenders. Properly skeptical that anyone who wanted to enter the neighborhood would be deterred, such people could not believe that the program would have any effect on crime. They failed to grasp a chain of logical steps: that the effect of a lot of traffic in residential areas was to depersonalize them; that a reduction in traffic would make the outside spaces more pleasant and attractive for use by residents; that if residents used the outside spaces more, it would increase the likelihood that they would take an interest in and become involved in what went on in the public and semi-private spaces near their homes; that such an interest would make it less likely that offenders would lurk in the neighborhood, waiting for criminal opportunities.

In essence, the street changes were one important part of an effort to restore the residential character of the neighborhood and give the area back to the residents. Part of the evaluation goal was to learn more about whether the hypothesized chain of events really worked. The analytic complexities of accomplishing that were considerable.

4. The planning and implementation of the program took place over a three-year period. This is fairly typical of environmental design programs. However, such a time period provides considerable opportunity for other, unplanned events to occur to further confuse the evaluation.

5. The program, including the physical changes, was in place less than a year when its impact was evaluated. Timing has considerable effect on evaluation. On the one hand, an early evaluation can show the effects of attention, regardless of the content of the program (Hawthorne Effect). On the other hand, some of the goals of the program, such as increased commitment to the neighborhood, might well take longer than a year to develop.

Each of the above points basically meant that the program was complicated to evaluate. In order to evaluate a complicated program, one is likely to need a complicated evaluation scheme.

Types of Measures

Two goals guided the research design. First, an attempt was made to measure each important concept or variable in at least two different ways using different methods. Second, although there was a commitment to quantitative evidence regarding the program, the design provided a variety of opportunities for qualitative feedback as well.

The multi-method approach to measurement is cited as desirable in almost any text on methodology. It is well known that any particular way of measuring something has its limits and likely biases. Conclusions based on different ways of measuring the same thing are likely to be sounder because they transcend the limits of any particular method. A distinctive characteristic of the Hartford experiment was not that the multi-method approach was valued but rather the extent to which the project team was successful in finding more than one way to measure the same phenomena.

Victimization rates and fear were measured by a sample survey of residents. Since the purposes of the program were primarily to produce improvements in crime

and fear of crime, some sort of resident survey was essential. However, the survey also was used to measure a wide range of resident perceptions and behaviors. In fact, for almost every aspect of the program and its effects that were studied, a useful set of measures came out of the resident survey.

Fear of crime was one of the few variables for which a second source of quantitative data was not developed. It is hard to measure fear except by talking to people. However, the views and observations of a panel of community leaders were solicited via semi-structured interviews to supplement the survey data.

With respect to crime, a second available source of information is, of course, police records. In this regard, the Hartford experience provides a good example both of the value of a multi-method approach to measurement and, in particular, of how essential victimization surveys are in assessing crime control programs.

It has long been known that a considerable portion of crimes that occur are not reported to police. Rates of burglary and robbery/pursesnatch derived from surveys are routinely two or three times the comparable rates derived from police records. However, it has been argued that for the measurement of trends over time, police records will provide a meaningful indicator of whether crimes are going up or down.

In Hartford, there was an opportunity to carry out victimization surveys over a five year period; and to compare the figures from the victimization surveys with comparable figures from police records. The results of this comparison are not surprising to those who have studied factors which affect police record estimates. However, they provide a warning to those who would rely on police record data alone as indicators of rates of crime.

During the five-year period in which Hartford crime was monitored, the study showed not one but two different occasions when, for reasons which had nothing to do with the rate of crime, the trends in crime based on police record data were very misleading.

The first case parallels a classic police anecdote. The introduction of a new Chief of Police in Hartford in 1974 was accompanied by an apparently massive increase

in crime. Victimization survey data showed that the increase was largely due to improved reporting practices on the part of police officers.

Three years later, the police record data showed a city-wide drop in burglary, while the victimization survey showed an increase. Some further research revealed that one of the symptoms of some continuing contract negotiation problems between the police and the city had been a sharp decline in the rate at which calls for service had yielded reports of actual crimes.

This experience illustrates two points. First, what shows up in the police records as a reported crime is dependent both on the behavior of citizens and the behavior of police officers. Extraneous factors which affect the behavior of either can have important affects on police record data and, consequently, on comparisons over time based on such figures. Although victimization survey estimates are not perfect by any means, the sources of bias or error should be consistent from time to time if a survey is properly done. Comparative statements based on victimization surveys should be reliable.

The second point to note is the value of the multi-method approach. In this case, the survey and the police record data did not produce the same conclusion. When this is the case, the discrepancy can make the researcher do further investigation. If only one method is used, the results are likely to be taken as accurate. Many evaluation studies, unfortunately, provide little potential for seeing inconsistency because of the lack of overlapping measures. Obviously, the more such overlap can be built in, the less likely the researcher is to make an error; and the more convincing will be the conclusions based on the research.

Measuring the use of spaces proved to be one of the most complex parts of the evaluation. In their initial analysis of the area, the urban designers had made numerous observations about the relationships between residents, non-residents and the spaces in the area: The neighborhood is depersonalized. Strangers dominate the streets. There does not appear to be any social cohesion. The parks are not used in an appropriate way.

Changing such things was an essential intermediate goal of the program. Therefore, it was incumbent upon the evaluation team to be able to make statements about whether and how much such changes occurred. To do that, it was necessary to quantify, or at least systematize, the observations of the urban design team.

Counts of vehicular traffic on Asylum Hill streets, which entail only the placement of counting machines for 24 hours, were one obvious source of information about vehicular traffic. The pattern of pedestrians' use of those streets was quantified by using human counters stationed at strategic spots for five different hour-long periods during the day. Days were standardized in that they had to be at least minimally attractive for walking; i.e., the temperature had to be above 50 degrees with no precipitation. Counters not only counted the number of persons passing their spot; they also coded them into sex, age, and ethnic categories by observation.

A third important source of information about the use of the neighborhood came from the survey residents, of course. Their perceptions of the vehicular and pedestrian traffic as well as their reports of their own behaviors were important input into understanding of how the neighborhood was being used.

Finally, the urban design team attempted to codify their observations. Based on a series of systematic walking trips through the area at specified times of day, they put on maps the people observed and their activities. The goal was not necessarily to produce a statistical basis for conclusions, but to systematize their observations, to provide some basis against which to compare observations at a later point.

In fact, there were significant problems in actually reaching conclusions based on changes in their coded observations from one time to another. Relatively little analytic use was made of these data. However, figuring out some way to codify observations of use of space is important to studies of environmental design programs. More work is needed to figure out how to do it well.

In summary, analysis of the way the land was used and how that might have changed as a result of the program was based qualitatively on the observations of the urban designers and the reports of people in the community; it was based quantitatively on

traffic and pedestrian counts and standardized survey measures administered before and after implementation.

Data on police were gathered in a similar way. Qualitative information was available on police operations from at least two sources. First, on a routine basis, the team leaders met with Hartford Institute[1] staff to review plans and problems. The Hartford Institute staff, in turn, produced routine summaries of significant happenings with respect to policing in the area. In addition, an outside monitor, experienced in police operations, spent a couple of days every two months visiting with the police team: talking with leaders and patrol officers, riding in patrol cars and reviewing record data. Both of these were extremely important to having an accurate, up-to-date picture of the police component of the program.

In addition, there were three more quantitative sources of information about the police. First, the police officers themselves filled out a questionnaire shortly after the police team was established and again near the end of the evaluation period. The resident survey included a number of questions both about resident perceptions of the police and about their own behavior with respect to the police. Included were items about reporting crimes to police, the amount and quality of contacts with police as well as citizen perceptions of response time, responsiveness and police effectiveness.

Finally, the police department's own records provide a quantitative indicators of police activity. Calls for service, arrests, and reported crimes all provide information which can be useful to an overall analysis.

The activities of the community groups that were formed in Acylum Hill were monitored in several ways. The Hartford Institute provided a good deal of information about these groups. Staff members attended most early meetings and had frequent con-

[1]The Hartford Institute for Criminal and Social Justice was responsible for implementation of the projects.

tact with the groups throughout the project. Their knowledge about activities and problems was periodically summarized.

In addition, a set of people knowledgeable about the community was interviewed in a semi-structured way on two occasions. Officers and leaders of the formal organizations in Asylum Hill were among those in the panel; and one of their particular contributions was to provide additional information about the groups and their activities.

Finally, of course, the resident survey once again was an invaluable source of information about residents participation in and knowledge of the community organizations that were trying to help them.

Thus, for each component of the program, the evaluation was able to draw on multiple sources of information. In some cases, exactly comparable measures were available from two different sources. In other cases, the data were complementary. In almost all cases, however, the fact that there were multiple sources of information significantly reduced the likelihood of an inadvertant error about what was going on and significantly increased the strengths of the conclusions that could be reached.

Analysis Strategies

There were two basic kinds of analytic conclusions that the evaluation was asked to come up with. The first question to be answered was whether or not the program was successful in reducing burglary and robbery/pursesnatch in Asylum Hill and the fear of those crimes. Second, regardless of the outcome, was there something to be learned from the experience in Hartford that would help others to design a crime reduction program in existing neighborhoods?

The impact analysis actually turned out to be two questions. Did crime and fear improve in Asylum Hill? and, was the program responsible for the improvement?

It is evident from the fact that the second question had to be asked that the answer to the first question was affirmative: at the end of a year, burglary and the fear of burglary had dropped to a level of approximately half of what one would have expected without intervention. Statistically, that was a highly unlikely chance

event. In addition, although the data on robbery and pursesnatch were less conclusive because of the comparatively low rates of those crimes, the odds were better than 2 out of 3 that those crimes and the fear of those crimes had also improved.

But was it the program that was responsible for this reduction, or was something else at work? It turns out to be extremely difficult in social science to prove that there is not a mysterious unidentified factor responsible for results. However, in this situation, the presence of the extensive Hartford data base was a tremendous asset in making alternative hypotheses less plausible.

One set of hypotheses was ruled out by analysis of city-wide data. The harshness of the winter, a change in economic climate or the inception of a city-wide offender work program all could have been plausible alternative reasons for a reduction in burglary. However, they would have affected the city as a whole. The decline observed in Asylum Hill occurred in the context of an overall 10 percent increase in crime throughout Hartford.

Having data on Asylum Hill in 1973, 1975, 1976 and 1977 helped to address other hypotheses. The improvement that was observed occurred in the experimental year of 1976-1977, not before. Prior to the experimental year, crime rates and fear in Asylum Hill had been rising steadily. Only events that would not have affected the crime prior to 1976 but then would have had a dramatic affect just during that year needed to be considered as plausible alternatives.

This logic was quite important in addressing one of the most compelling alternative ideas: that the offender population that had worked in Asylum Hill had moved away. A public housing project which had produced a disporportionate number of criminals working in Asylum Hill had been "thinned out". There also had been quite a bit of abandonment and demolition in an area north of Asylum Hill where offenders had been known to live. It was, of course, not known exactly how many offenders had moved, nor whether they had moved far. However, that at least some of them had moved somewhere was almost certain.

There were, however, two facts which argued against this change being a major factor in the observed reductions in crime in Asylum Hill. First, the thinning out of the public housing project and the housing abandonment had been going on for at least a year prior to the experimental year. One would have expected to see effects of this prior to the 1976-1977 year if it was significant. Second, detailed victimization data on areas around Asylum Hill did not show declines in burglary and robbery such as those found in North Asylum Hill. Since these areas were within reach of the same offenders who worked in North Asylum Hill, one would expect a significant change in the offender population to have affected these adjacent areas as well. Thus, the data permitted one to rule out a change in the offender population as a significant factor in the observed crime reduction with a considerable degree of confidence. Had the data been less rich, that hypothesis might well have seriously undermined confidence in the conclusion that the program affected crime.

The above deals with negative arguments, trying to rule out alternative hypotheses. Another approach is to produce documentation that the program produced changes which could plausibly reduce crime.

It will be recalled that the key to crime reduction was thought to be increased resident control over the neighborhood. There was considerable evidence that things had moved in a positive direction in this respect: vehicular traffic had clearly been restructured and reduced overall; there had been some reduction of pedestrian traffic on residential streets, though that was not always the case; residents reported that they were doing significantly more walking in the area and were using the parks more; they reported that their stranger recognition had improved; they reported more frequent arrangements with neighbors to watch out for one another's houses.

These changes, most of them statistically significant, helped to buttress the notion that the program had succeeded in starting a chain of events that plausibly could lead to crime reduction. On the other hand, there were some changes

that were expected but not observed. Optimism about the neighborhood's future had not improved. While fear of the target crimes had gone down, there were a number of neighborhood problems which, in the view of residents, had not improved.

Of course, data alone, no matter how good, do not eliminate the role of judgement. Were the changes observed dramatic enough to have produced a 50 percent reduction in burglary? Some reviewers will be more convinced than others. However, because of the extensive data base, critics of the conclusion that the program reduced crime and fear during its first year have a difficult case to make. The possible alternatives identified by the research team do not hold up under scrutiny. Could there have been an heretofore unnoticed event that occurred at roughly the same time as the street closings, affected North Asylum Hill but not surrounding areas, and had the exact effect the program was designed to have?

In social science, it is difficult to prove anything definitively. However, the case <u>for</u> a program impact seems much stronger than the case <u>against</u>.

<u>To produce generalizable knowledge</u> was the other analytic goal of the evaluation. Based on one demonstration, there is no statistical basis for generalizing. The foundation on which one generalizes from a single experiment is conceptual rather than statistical. It is in this context, again, that the complex data base developed in Hartford both before and after program implementation was critical to the value of that experiment to others.

There are two kinds of questions that a person considering the Hartford model would want answered. First, was the situation identified in North Asylum Hill sufficiently similar that one could apply the analysis to another community? Second, did the apparent success of the intervention in North Asylum Hill say anything about the likely success or failure of other similar interventions? Through detailed description of the "before" situation, a good evaluation should enable a person to answer the first question. Through analysis of the dynamics of the intervention, and detailed description of what was implemented and with what effect, a reader should be able to begin to address the second question.

The analytic value of good, comprehensive data was once again demonstrated in connection with the question of the role of the three components - physical changes, police and community organizations - in the program's success. Fortunately, two unplanned natural experiments occurred that permitted a fairly definitive answer.

In the target area, the police and community organization components were begun a year before the street changes were made. However, it was only after the street changes that crime and fear declined.

An area adjacent to the target area was served by the Asylum Hill police team and also developed a significant crime-oriented community organization. However, no street changes were made in this area; and no decreases in crime or fear occurred.

Although the role of the other components cannot be assessed fully, it is clear that the physical design changes were necessary to the success of the program. Being able to make that statement is very important to those who would learn from the Hartford experience. The answers will seldom be definitive or unassailable. However, the better the quality of description and understanding that an evaluation produces, the more likely it is to be useful to others.

Conclusion

The evaluation of the experiment in Hartford was unusually full and complete. Even so, there were desirable steps not taken because of limited funding. There is always some limit to funding. For example, although offender interviews were conducted in the planning stages of the project, none were done after implementation. There were ways in which the monitoring of some of the community activities was not as detailed as it could have been. More money and more time would have reduced the number of gaps in the analysis, but clearly would not have eliminated them all. Social science evaluations do not produce certainty very often; and this one was no exception.

Having made that point, perhaps it is appropriate to close with a more general comment about the importance of good methodology in evaluation research.

The jumping-off point for evaluation research was probably the experimental designs outlined by Campbell and Stanley many years ago. Those faced with the task of evaluating real projects soon found that the conditions for true experiments were seldom met. Moreover, it was observed that often the results of even careful evaluations were inconclusive.

There have always been those who considered research a waste of time and money. There have always been practicing researchers who, through lack of sophistication or for other reasons, did methodologically weak research. Such people have found support from methodologists who focus on the limits of evaluation and understate the achievements, both real and potential. From the statement that definitive conclusions are unlikely to result from evaluations, it is an easy leap to decide that the quality of an evaluation does not matter.

There are many programs that are so poorly conceived or implemented that they warrant little or no investment in evaluation. However, at any point in time, there is extant a set of ideas about how to deal with a certain kind of problem, in this case, community crime control. When a program is implemented which provides the opportunity to learn something about the validity of those ideas and how to apply them, a serious, careful research evaluation effort is a very good investment. There is no possibility that even a tiny fraction of the funds spent on poor or ineffective programs will ever be spent on research.

To criticize evaluations that do not meet strict statistical requirements for experimental generalization is to hold up an artificial standard. The goal of evaluation research is to learn. Learning means to reduce uncertainty about the way things are and the way things work. It does matter how well a research evaluation is carried out; whether the effort be large or modest, the better the methodology, the more uncertainty will be reduced.

The Hartford project was not a perfect evaluation. It was a good one. Most important, the rigorous and comprehensive approach to evaluation that was utilized was essential to the general value that can be derived from the project. It was a serious attempt to learn something important. More such efforts are needed.