

GETTING THE CRIME RATE DOWN: POLITICAL PRESSURE AND CRIME REPORTING

DAVID SEIDMAN *Princeton University*
and

MICHAEL COUZENS *University of California, Berkeley*

AUTHORS' NOTE: This article is a revision of a paper prepared for delivery at the 1972 Annual Meeting of the American Political Science Association, Washington, D.C., September 5-9, 1972. The original paper was prepared while we were at the Brookings Institution, and we gratefully acknowledge the Institution's support. However, the views expressed in this article are those of the authors and should not be taken to represent the views of the Brookings Institution, or necessarily to represent the views of any person associated with the Institution.

. . . [F]ear is being swept from the streets of some — though not all — American cities.

*John Mitchell, September, 1971
New York Times, 1971: IV, 16*

Enterprising police are making news rather than history in the preparation of their annual reports. Dry tabulations are giving way to charts, pictures, and other illustrations that catch the eye and tell the story of the problems and accomplishments of the department.

Federal Bureau of Investigation, 1966a: 58

INTRODUCTION

During the 1960's, "crime in the streets" emerged as a major political issue.¹ Consequently, measurement of changes in the amount of crime also became politically important. As a candidate in 1968, Richard Nixon, recognizing this, proposed the use of the District of Columbia as a site for an evaluation of his anti-crime proposals by means of crime statistics:

There is another area where the Federal Government has an opportunity to make a dramatic demonstration of its concern with the problem of crime, its commitment to new solutions and the efficacy of its proposals. That is in Washington, D.C.—the nation's capital where the authority of the Federal Government is great and its prerogatives many.

Washington, D.C. should be a model city as far as law enforcement is concerned—a national laboratory (*New York Times*, 1972).

Implicit in this suggestion is a sophisticated research strategy, a quasi-experimental time-series design.²

The emphasis on the District of Columbia was not new. Lyndon Johnson established the President's Commission on Crime in the District of Columbia in 1965, and the Commission produced a *Report* (1966) of nearly 2,000 pages late in 1966. There were at least two reasons for emphasizing the District. First, the District had been plagued by high crime rates for quite some time. Second, while the root causes of crime — which may well be such things as poverty, racism, and inequality — can be influenced by the national administration, a direct attack on these social problems is at best a long-term crime reduction strategy. Short-run strategies rely instead on the police, the courts, and, in recent years, narcotics control programs. These are primarily local matters, as noted in the Declarations and Purpose of the Omnibus Crime Control and Safe Streets Act of 1968:

Congress finds further that crime is essentially a local problem that must be dealt with by State and local governments if it is to be controlled effectively (82 STAT. 197, 42 U.S.C. §3701).

But in the District, state and local government is the national government. The Administration, after a campaign which promised improvements in the crime situation nationally, found in the District one place where it could directly implement programs to deal with crime. Consequently, having pledged to address a national problem, and lacking the means for improvement in the short run, the Nixon administration emphasized crime control in the District. It was to be a sort of pilot project, a demonstration of effective techniques which others could then imitate elsewhere.

Evaluation of the effectiveness of crime control programs typically depends upon crime statistics of the kind collected for the Uniform Crime Reporting (UCR) program administered by the Federal Bureau of Investigation.³ It is important, therefore, that the characteristics of the FBI data be clearly understood. These statistics constitute what Webb *et al.* (1966) refer to as "archival" data. They are collected in the normal course of the administrative functioning of police departments and the FBI, and therefore are presumably what Webb *et al.* would consider "non-reactive" measures. In fact, however, these statistics are neither adequate measures of the amount of crime⁴ nor notably non-reactive. As Campbell says:

Those who advocate the use of archival measures as social

indicators . . . must face up not only to their high degree of chaotic error and systematic bias, but also to the politically motivated changes in record keeping that will follow upon their public use as social indicators . . . (Campbell, 1969: 415).

This paper reviews widely recognized shortcomings of the FBI statistics as measures of crime with a focus on the scope for reactivity. Then we look at the "national laboratory" from the perspective of quasi-experimental research design. Patterns in District of Columbia crime data are examined in detail. Comparisons with other cities are made. The analysis concentrates on reactivity of the measures of the target behavior—crime. Implications of reactivity for analysis of the impact of public policy are discussed.

MEASURING CRIME

Media discussion, public awareness, and political controversy concerning the amount of crime focus on the periodic FBI reports or on monthly police reports.⁵ (The FBI reports are based on police statistics.) Police reports usually present statistics for the categories that the FBI uses. The reporting system, Uniform Crime Reporting, was developed by the FBI and the International Association of Chiefs of Police on the basis of standardized categories which allow interjurisdictional comparison (Thompson, 1968). The reason for this focus on police and FBI reports is simple: there is no other choice. Systematic knowledge of crime depends upon an organized means of knowing. The procedures of police departments are the only such means available.⁶

The UCR mechanism is straightforward.⁷ With some technical assistance from the FBI, local police departments collect their statistics and forward them to the FBI, which then groups them in various ways, changes them from a monthly to a quarterly or yearly frequency, and publishes the results. Not all crimes reported are publicized. Seven, the so-called "Index" crimes, receive most of the attention. The FBI describes its program and statistics:

The Uniform Crime Reporting Program employs seven crime classifications to establish an index to measure the trend and distribution of crime in the United States. These crimes—murder, forcible rape, robbery, aggravated assault, burglary, larceny \$50 and over in value, and auto theft—are counted by law enforcement agencies as the crimes become known to them. These crimes were selected for use in the Crime Index because, as a group, they represent the most common local crime problem. They are all serious crimes, either by their very nature or due to the volume in which they occur. Basic-

ally, they can be categorized as violent crimes, such as murder, forcible rape, robbery and aggravated assault, or as crimes against property, such as burglary, larceny \$50 and over in value, and auto theft.

It is believed desirable to point out that there is no way of determining the total number of crimes which are committed. Many criminal acts occur which are not reported to official sources. In light of this fact, the best source for obtaining a count of crime is the next logical universe, namely, crimes which come to police attention. The crimes used in the Crime Index are those considered to be most consistently reported to police and the computations of crime trends and crime rates are prepared using this universe—crimes known to police (Federal Bureau of Investigation, 1969: 4).

Several things are worth pointing out about this Crime Index: (1) An enormous number of crimes, many commonly considered serious, are not included in the Index. If a co-worker takes \$50 from my desk, that is larceny, and it may end up in the Index. If he embezzles a million dollars from the company, that is not larceny. If he forges checks, that is not larceny. If the company fixes prices illegally, violates health laws, or cheats on its taxes, these crimes are not counted in the Index. White collar crimes, both in Sutherland's (1949) sense of crimes committed by businesses and in the sense of crimes such as forgery and fraud committed by individuals, are not counted at all. The sins of organized crime are generally non-Index. The list could be extended. (2) Because data are collected originally by local law enforcement agencies, and not the FBI, the statistics are not really comparable across jurisdictions. Illustrations of this non-comparability appear below. (3) It may be true that the Index crimes are those "most consistently reported to the police," but evidence from surveys conducted for the President's Commission on Law Enforcement and the Administration of Justice suggests strongly that only a small and highly variable percentage of actual Index crime is in fact reported (Biderman, 1967). (4) The Index is a simple sum of the raw totals of the Index crimes; a murder counts equally with the theft of \$50. "The fundamental criticism here is that there is no weighting by seriousness when the Index is considered as a whole" (Mulvihill and Tumin, 1969: XI, 26). There is a simple solution to this problem, ignore the Index; and a complicated one, weight the crimes by seriousness. Neither is adequate: the first because of the appeal of a single number compared to an array, and the second mainly because it is too complicated for general use.⁸ And any scheme of weighting entails assigning degrees of seriousness to each category of

crime. Universal agreement on the relative seriousness of crimes is unlikely.

The Crime Index, then, is a measure — perhaps even an accurate measure — of formal police recognition of certain crimes. It bears an unknown relationship to the actual amount of “crime.”⁹ It seems reasonable that the statistics indicate a minimum possible value in most categories: there can be no less actual crime than is known to the police. It is plain that there is more, but we lack a way of knowing how much more. Perhaps we are justified in using the Index — or its separate categories — as a measure of crime, since we have nothing better. But perhaps knowledge of our ignorance is better than the error to which use of UCR statistics may lead.

Sources of Error in the Crime Statistics

Consider production of crime statistics as a sequential process with these stages:

1. An event occurs, which could be interpreted as a crime.
2. It, or its consequence, is observed, perhaps by the victim, perhaps by someone else. (For simplicity, ignore the near certainty that the perpetrator observes it.)
3. The observer notifies the police. (The observer may be a policeman, but this is not the usual case.)
4. The police decide whether the reported action is to be considered a crime¹⁰ and, if so, how it should be described.
5. Sometimes this description is reviewed at another point in the police hierarchy.
6. The police decide which of the FBI categories is appropriate. Typically, the way in which categories are defined by the laws of a jurisdiction does not correspond to the FBI-UCR definitions. For example, the FBI uses a \$50 cutoff figure to distinguish Index larcenies from minor larcenies, while states may use another figure to distinguish grand and petty larceny.
7. The statistics are made public.¹¹

Ideally, variation in the statistics released at stage 7 would be caused entirely by variation at stage 1. But extraneous variation may enter at any of the other stages. Not all the stages have been studied extensively. For example, it would be hard to gauge the proportion of crime observed only by the perpetrator.

In spite of the FBI's cheerful optimism, the proportion of observed crime which is not reported is large even for the Index crimes. The major national study of this problem indicates that about one-third of all robberies and two-thirds of all burglaries are never reported (Biderman, 1967: 17). Variation in these figures is to be expected. Biderman and Reiss (1967: 7) suggest one reason for this variation:

. . . [I]t is maintained that as larger proportions of the population become integrated into the dominant society and come to share its normative conceptions, more people mobilize the police to enforce public deportment.

Ramsey Clark (1971: 29-30) suggests others:

The better the police, the more they learn of crimes that are actually committed. . . . Whenever crime is not reported we know the police are not effective, trusted, or respected. . . . It should not be surprising that those who suffer crime most report it least.

And, of course, willingness to report to police can change.

A reactive source of variation in the proportion is "deviation amplification":

. . . [A]n initial rise in the officially recorded crime rate gives rise to an increased concern about crime in the press and broadcasting, and hence among the general public. This leads to a change in public attitude toward crime and criminals, which expresses itself in an increased propensity to report crimes. This in turn leads to a further increase in the officially recorded crime rate, and the process starts over again (Roshier, 1971: 502).

The police decision to treat an event as a crime is not automatic. Black (1970) suggests several sources of variation in treatment of events: desire of a complainant for legal action, "relational distance" separating the parties involved in the event, deference towards the police shown by the complainant, and complainant's social class. Thus the proportion of events which the police handle by informal means rather than by formally labeling the event as a crime can and does vary over time and between jurisdictions.

Police administrative procedures have a large impact on whether a complaint later turns up in official statistics. Therefore, changes in administrative procedures may produce large variations in reported crime totals. For example, a truly striking change in Chicago's larceny statistics (from about 10,000 yearly to about 30,000 yearly) is associated with a change in bookkeeping instituted by Orlando Wilson as part of his reform of the Chicago police (Campbell, 1969: 415). The effects are

rarely so striking, but administrative changes are common and tend to have substantial impact on the statistics. Examples of such changes are the introduction of a centralized complaint handling system, changes in the standard form for incident reports, revision of departmental guidelines for reporting specific offenses, and changes in training program curricula.¹²

These various sources of error¹³ mean that crime statistics are basically non-comparable across jurisdictions. A government study delicately suggested the magnitude of the problem:

Disparities as great as 17 to 1 between Newark and Jersey City, or 10 to 1 between St. Louis and Milwaukee, for certain offenses seem most unlikely in the absence of some reporting variation (President's Commission on Law Enforcement and Administration of Justice, 1967: 27).

Comparison over time within a single jurisdiction is a less clear case. Drift in many of the relevant sources of error may render long-term comparisons invalid. But excepting administrative procedures, most sources of error probably change slowly enough to allow use of crime statistics for evaluating trends in crime over a short period within a jurisdiction. The trick is to know when procedures change. Some procedural changes are publicly announced and noted in FBI quarterly reports. Others are not. Some, in fact, are vigorously denied.

Misclassification of Offenses

Determination of the correct UCR classification for an event can be complex. We limit our discussion to problems presented by three categories: larceny, burglary, and robbery. These three constitute the bulk of total Index offenses (about 73% in the District in 1970) and are extremely susceptible both to misclassification and to the detection of such misclassification, as shown below.

Larceny: The FBI defines this category as

. . . the unlawful taking or stealing of property or articles of value without the use of force or violence or fraud. . . . Auto theft . . . is excluded. . . . The Crime Index offense of larceny is limited to those thefts where the value of property stolen is \$50 or more (Federal Bureau of Investigation, 1969: 22).

The major problem is valuation of the stolen property. Under ideal conditions, appraisal of property may be inexact, and the valuation of stolen property is made more difficult by the absence of the property in question. The police, who make the determination, must appraise something which they cannot see under circumstances which may give the legal owner of the property incentive to claim an inflated value. That errors

might be large and frequent is understandable.¹⁴ It is also clear that should the police desire to reduce the number of reported Index crimes, all that is required is systematic under-valuation of property. Theft of property valued at \$50 is an Index crime, but should the value be only \$49, the proper category is a non-Index crime. Without a formal change in administrative procedure the police can, through property valuation, partially determine the level of the publicized Index crime statistics.¹⁵

Burglary:

. . . burglary is defined as the unlawful entry of a structure to commit a felony or theft, even though no force was used to gain entrance . . . (Federal Bureau of Investigation, 1969: 17).

There appears to be substantial misclassification of burglary simply because this definition does not correspond to that of many states (President's Commission on Law Enforcement and the Administration of Justice, 1967: 24; Federal Bureau of Investigation, 1966b: 27). Subtle questions of intent also present problems:

A forcible entry or unlawful entry where no theft occurs but where acts of vandalism, malicious mischief, etc., are committed, is not scored as a burglary *provided your investigation clearly establishes that the unlawful entry was for a purpose other than to commit a felony or theft* (Federal Bureau of Investigation, 1966b: 27).

Because no criteria are provided for "clearly establishes," some flexibility in classification is inherent in the definition. If the police want to reduce the number of burglaries, they can relax their standards for judging intent: "vandalism" and "malicious mischief" are available non-Index options.

The police can also simply ignore the unlawful entry (particularly if no force is involved) and record the crime as a larceny. Then the valuation of the stolen property determines whether the crime is an Index crime. Press reports assert that the Baltimore police have engaged in this practice for the purpose of reducing the number of Index crimes (Twigg, 1971).

Robbery:

Robbery is a special and vicious type of theft. It takes place in the presence of the victim (the owner or a person having custody of the property). To obtain the property or thing of value, the robber uses force or violence on the victim or puts the victim in fear by use of threats, weapons, etc. It is like

larceny but is aggravated by the element of force or threat of force (Federal Bureau of Investigation, 1966b: 20).

Assaults to rob or attempts are included (Federal Bureau of Investigation, 1969: 13).

Neither use of force nor threat of force is simply judged. How real was the threat of violence? A purse-snatching is a robbery if the thief uses "more force than is necessary to obtain the purse . . ." (Federal Bureau of Investigation, 1966b: 24). How much force is that? Was more force threatened? If the police decide that the threat was not real, or that the force was not more than necessary, a successful robbery may be counted as a larceny, which makes the valuation of the stolen property determine whether it is an Index crime. In the case of attempted robbery, the police may simply ignore the intent, and count the offense as an assault. Simple assault, as opposed to aggravated assault, is not an Index offense. Once again, should the police desire to reduce the reported number of Index crimes, small shifts in decisions will accomplish the result.

Scope for deliberate misclassification of robbery should not be exaggerated. In contrast to burglary and larceny, robbery implies a witness, the victim of attack, who may take considerable interest in the treatment of the event by the police. This presumably limits police discretion. On the other hand, the ability of the police to ignore the victim should not be entirely disregarded:

One police report . . . contains the account of a man who was robbed as he got out of a cab. The report says he was not injured and labels the case a simple larceny.

In fact, according to the victim . . . he was slugged repeatedly in the face by the robber. The police officer who compiled the report interviewed the 61-year-old man while he still wore a bloodstained bandage wrapped around his injured head, the victim said (Twigg, 1971: July 14).

In sum, the scope for accidental or deliberate misclassification of these three offenses is large. As a result the FBI index can be substantially misleading at times as a measure of changes in crime. Motivation for deliberate misclassification is suggested by Reiss' (1971: 542) comment that "On these crimes hinges the public's information about crime in the United States." The performance of the police, if measured by the crime statistics, can be improved by deliberate "downgrading" of offenses. Should this practice actually occur, it would be a clear example of the reactivity of archival data: the use of

crime statistics as a performance measure influences the level of that measure.

Larceny is probably the easiest of these three crimes for the police to downgrade. The only requirement is a little imagination in the valuation of property. Burglary is probably more difficult to downgrade, if only because, in the case of a successful burglary, a two-step process is involved: ignore the entry, misvalue the property. Robbery, because of the inevitable witness, seems substantially more difficult to downgrade than either of the other two.

THE NATIONAL LABORATORY

The Administration's anti-crime program in the District of Columbia has been highly controversial, though the most controversial features have almost certainly had little or no effect on crime in the District. An omnibus D.C. crime bill, which included such provisions as "no-knock" entry for the police and "preventive detention," was passed in July of 1970 (*Congressional Quarterly*, 1971). Preventive detention did not go into effect until February of the next year, and by November a reporter noted that in "its seven-month life, preventive detention . . . has been used in D.C. Superior Court about five times and about seven times in U.S. District Court . . ." (Kneece, 1971).

Other features of the program, financed by more than ten million dollars of Law Enforcement Assistance Administration funds, were more likely to have an impact. The *Washington Post* provides a partial catalogue:

Congress . . . authorized a massive increase in the size of the city's police force. The number of officers leaped from a low of 2,958 in mid-1968 to 5,100 by September, 1970. It now stands at about 4,900, giving the city the highest policeman-to-resident ratio in the country. . . .

The department acquired more scout cars, more scooters, more radios. Police became more visible and more mobile. An elaborate computer was installed at headquarters to provide almost instantly retrievable data on stolen autos, arrest records and other information.

The men and women on the force today are younger than ever before. They are better paid . . . more are black . . . more are enrolled in college courses; many have been placed in special police-community sensitivity programs (Valentine, 1971).

Residents of Washington are aware of another feature of the program: high-intensity street lights cover increasing segments of the city.

Narcotic treatment programs have been substantially increased in the District starting in late 1969. The Director of the city's Narcotics Treatment Administration wrote, in December, 1970:

One year ago, there were 10 patients in Washington's new treatment program, though only one was taking methadone. Today, over 2,300 participate in the City's program . . . the primary goal: reduced crime (Dupont, 1971: 4-5).

The quasi-experimental design most appropriate to an evaluation of the impact of this anti-crime package is what Campbell (1969: 416) calls the "interrupted time-series design." However,

with this weak design, *it is only abrupt and decisive changes that we have any chance of evaluating.* A gradually introduced reform will be indistinguishable from the background of secular change, from the net effect of the innumerable change agents continually impinging.

If Campbell's two primary examples, the Connecticut crackdown on speeding (Campbell and Ross, 1968) and the British Breathalyser crackdown (Ross, Campbell, and Glass, 1970), are the standard, the D.C. crackdown on crime is not abrupt and decisive. But in comparison with other forces influencing crime rates, the anti-crime package may be abrupt enough. In any case, we examined the data from the standpoint of interrupted time-series design to see if there had been a decrease in crime in the District.

Table 1 displays the yearly Index crime statistics for the District from 1968 through 1970. Table 2 shows the statistics for the first six months of each year from 1968 through 1971. These statistics indicate a decline in crime in the District, roughly coincident with at least some features of the Administration's program. There are fewer recorded Index crimes in 1970 than in 1969, and, using figures for January through June, 1971 has even lower statistics than 1970. The impression of progress in the "war on crime" is strengthened by inclusion of earlier data, as there was a rising trend for several years prior to 1968. Not *much* progress perhaps — the last year of the Johnson administration produced lower totals than did 1971 — but a long-term trend appears reversed.¹⁶

The interrupted time-series design has been referred to as a "weak" design because it does not guard well against threats to validity of the inferences drawn. In other words, many factors besides the Administration's anti-crime program could

TABLE 1: INDEX CRIMES, WASHINGTON, D.C., 1968-70

Year	Total	Murder	Rape	Robbery	Aggra- vated Assault	Burglary	Larceny G.T. \$50	Auto Theft
1970	59311	221	313	11816	4089	22348	9414	11110
1969	62229	287	336	12366	3609	22933	11508	11190
1968	49360	260	260	8622	3103	17950	7876	11354

Source: FBI, Uniform Crime Reports

TABLE 2: INDEX CRIMES, WASHINGTON, D.C., 1968-71, JANUARY-JUNE ONLY

Year	Total	Murder	Rape	Robbery	Aggra- vated Assault	Burglary	Larceny G.T. \$50	Auto Theft
1971	25876	125	241*	5624	2023	9778	3804	4281
1970	30999	116	135	5896	2027	11996	5381	5448
1969	26830	125	150	5096	1725	10107	4954	4673
1968	22013	88	100	3491	1489	8829	3138	4878

*This number should be disregarded because of a reporting change.

Source: 1968-70 FBI quarterly releases

1971 Statistical Reports, Metropolitan Police Department,
District of Columbia, January-June 1971

conceivably account for the pattern. Rather than take steps to guard against threats to the validity of the inference that the program produced a decline in crime, we will turn the design on its head. We will treat some other event as the interruption of the time-series, and consider its likely effects on the crime statistics. With the other event installed as the critical interruption, the anti-crime programs become threats to the validity of inferences concerning the new event, and we will attempt to control for these threats. The other event we selected is an internal change in the police department — the hypothesis being tested makes this administrative event the source of changes in the level of crime statistics, independent of the effects of the Administration's crime program.

Reactivity

Patrick Collins, a reporter for the *Washington Daily News*, examined police records on a number of larcenies. He found several cases of egregious misvaluation of property which downgraded serious larcenies to the lesser category of larcenies of less than \$50 in value. The police-determined value was often roughly \$49, while in some of these cases insurance companies paid the victim several hundred dollars more than the value the police recorded. Collins drew the inference that the District police were deliberately misvaluing property in order to produce a decline in the total of Index crimes (1971a). The police answer to this suggestion pointed to the overall deficiencies of the system: valuation of property is a difficult and imprecise task in which errors are to be expected. There always have been errors, and there always will be. Since there always have been errors, Collins' work says nothing that would suggest re-

cent misbehavior on the part of the police (Wilson, 1971).

This dispute between Collins and the D.C. police attracted our attention. It seemed to us that the pivotal question was whether there had been a change in the degree of police misvaluation of property. Several approaches could be taken to the question. One was to extend Collins' technique. But examining larceny reports covering a year or more, contacting victims, and attempting to revalue property was too large a task. Another technique was to interview policemen. This did not appear fruitful. (The *Washington Post* tried it, with ambiguous results. Its account of larceny valuation contained a welter of conflicting claims concerning "direct" and "indirect" pressures, complaint books, and lost dogs. The *Post* (Valentine, 1971) concluded that no official orders to manipulate crime statistics had been issued by headquarters). The approach we decided upon was an interrupted time-series design. As a *product* of police behavior, crime statistics can be used to *measure* that behavior.

The installation of Jerry V. Wilson as Chief of the District police in August 1969 provides the "abrupt and decisive" change required for the use of this design. The importance of Wilson's promotion to Chief is suggested by his reported threat that police commanders unable to reduce crime in their jurisdictions would be replaced by men who could (Mann, 1971). We focused on the relationship between larcenies of \$50 or more and larcenies of less than \$50. While the "less than" data do not form part of the FBI Index, police departments report them to the FBI, and they were therefore available to us.¹⁷

Consider the percentage of all larcenies which are greater than \$50 in value. It seems reasonable that this statistic will not be completely unstable over time. Behavior of thieves is not likely to vary much from month to month (barring seasonal fluctuations) or from year to year. The value of property is likely to vary. It rises in a period of inflation, and so we expect this statistic to increase as well; some property which had been worth only \$48 three years ago is now worth over \$50. Inflation being a chronic American problem, a steady rise in the proportion of larcenies "greater than" is likely.

Examining the relevant data, our expectations concerning this statistic are confirmed. Figure I shows this statistic aggregated for two groups of cities: the first group having populations of 500,000 to 1,000,000, and the second, populations over 250,000.¹⁸ There are movements both up and down, but in general the trend is up.

FIGURE I: PERCENTAGE OF ALL LARCENIES WHICH ARE \$50 OR MORE IN VALUE — NATIONAL DATA

Source: F.B.I. Uniform Crime Reports

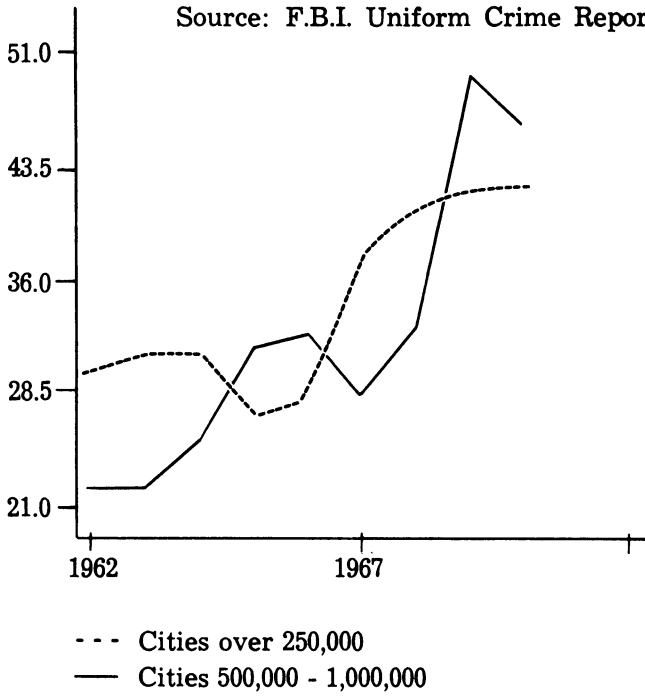
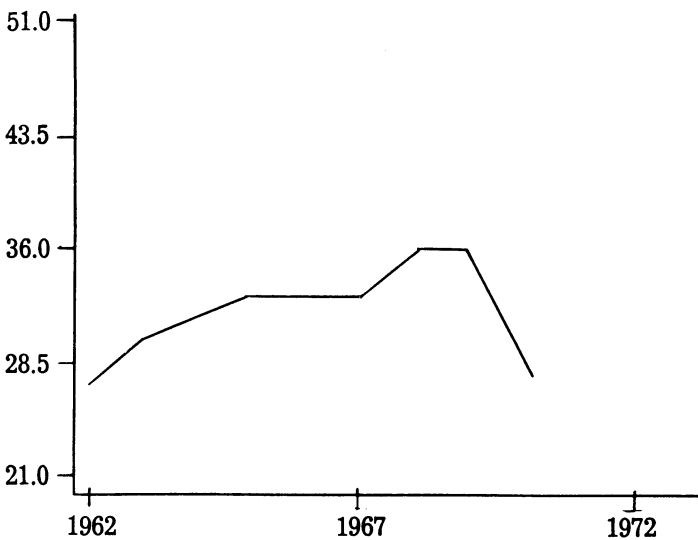


FIGURE II: PERCENTAGE OF ALL LARCENIES WHICH ARE \$50 OR MORE IN VALUE, DISTRICT OF COLUMBIA



The District of Columbia also conformed to this pattern through 1969 (Figure II). But 1970 represents a sharp break.¹⁹ This change is the first suggestion of a notable increase in police downgrading of larcenies.

A more detailed examination of the data can be made using monthly statistics. Figure III shows monthly data for the two types of larcenies from July 1967 through June 1971. Figure IV shows the proportion of all larcenies which were greater than \$50 in value. In both cases, the vertical line marks August 1969, the month of Jerry Wilson's installation. We notice roughly parallel movement of the two series prior to August 1969 and nearly oblique movement for at least a year following that date. (While we do not present the data here, we have examined these statistics as far back as January 1962: they move roughly in parallel throughout the period prior to August 1969.)

The visual impression is confirmed by statistical procedures. (The procedures themselves are discussed in Appendix I. Here we present only a quick overview of the results.) We seasonally adjusted the data from July 1967 through June 1971 and, using a Chow test, tested the null hypothesis that the coefficients of the regression of larceny less than \$50 in value on larceny greater than \$50 in value did not shift with Wilson's installation. The results are not entirely satisfactory — the Durbin-Watson statistics are unacceptable — but the F-test clearly rejects the null hypothesis. With 2,44 degrees of freedom, F equals 86.352, significant beyond the 1% level. (The regressions are presented in Table 3.) Of course, given the graphs, this is hardly surprising. It is noteworthy that the regression for the second period is not significant at the 5% level. This further suggests that something other than the actual level of crime has influenced the statistics.

Perhaps the Nixon anti-crime program was, for some reason, far more effective against big larcenies than against small ones. But an explanation of why this would happen is hard to develop: in many cases, no one even knows the size of a larceny until the police have valued the stolen property. This explanation of the pattern in the data cannot be completely refuted, but it appears less likely than the alternative, that misclassification increased substantially with Wilson's installation.

In the case of burglary or robbery, downgraded incidents get classified as minor larcenies or disappear into categories

FIGURE III: MONTHLY CRIME STATISTICS, DISTRICT OF COLUMBIA

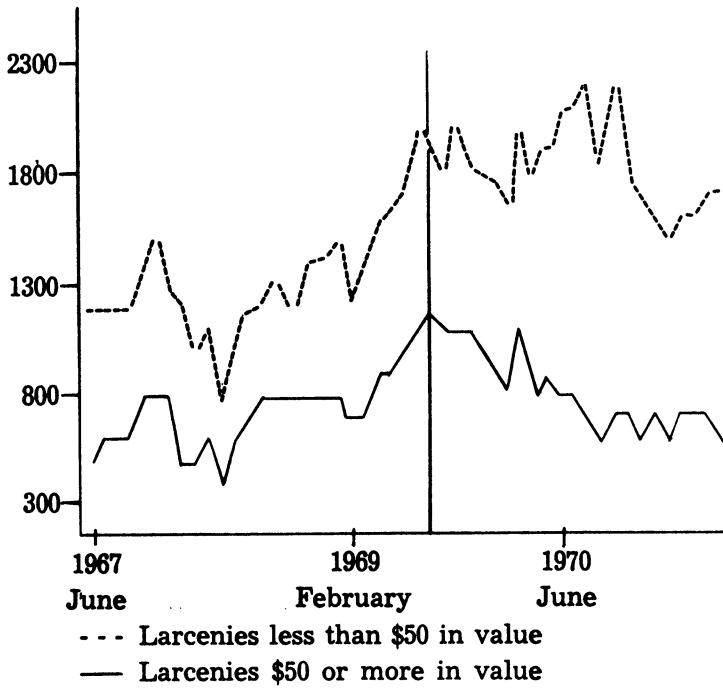
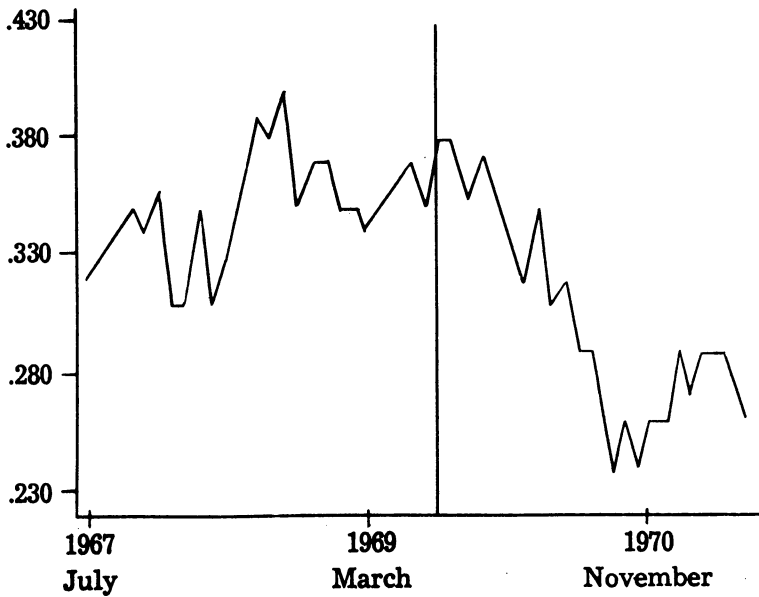


FIGURE IV: PROPORTION OF ALL LARCENIES WHICH ARE \$50 OR MORE IN VALUE, DISTRICT OF COLUMBIA



for which data are not readily available. Tests for downgrading are less clear-cut. The procedure we employ is to examine the relationship between the data for burglary and robbery

TABLE 3: DEPENDENT VARIABLE: LARCENIES LESS THAN \$50 IN VALUE, SEASONALLY ADJUSTED

<u>Variable</u>	<u>Coefficient</u>	<u>T-Statistic</u>	
PERIOD: July 1967-August 1969			
Constant	421.596	4.704	
Larceny Greater Than \$50	1.242	10.361	
			-2
			R = .810
			F _{1,20} = 107.36
			D = 1.183
PERIOD: September 1969-June 1971			
Constant	1583.638	11.066	
Larceny Greater Than \$50	0.306	1.720	
			-2
			R = .085
			F _{1,20} = 2.96
			D = .736
PERIOD: July 1967-June 1971			
Constant	768.655	4.474	
Larceny Greater Than \$50	1.040	4.690	
			-2
			R = .309
			F _{1,18} = 21.999
			D = .206

Chow Test: $F_{2,44} = 86.352$

and those for larcenies of \$50 or more in value. If the pattern for another crime closely matches that for larceny of \$50 or more in value, it can be inferred that this second crime has also been more frequently downgraded during Wilson's tenure. But the grounds for inference are weaker than with larceny.

Figure V shows that the pattern for burglary does in fact closely match that for larcenies \$50 or more in value for the period of the graph. (As before, the relationship shows a striking parallelism as far back as January 1962.) A Chow test, similar to the previous one, produces an F-statistic of 9.962 (Table 4). This value is statistically significant, and suggests that the relationship between burglaries and larcenies did shift with Wilson's installation. But the fact that the F-statistic is smaller than the one produced by the analysis of the two larceny series confirms the visual impression that the magnitude of the shift is not very large.

FIGURE V: MONTHLY CRIME STATISTICS, DISTRICT OF COLUMBIA

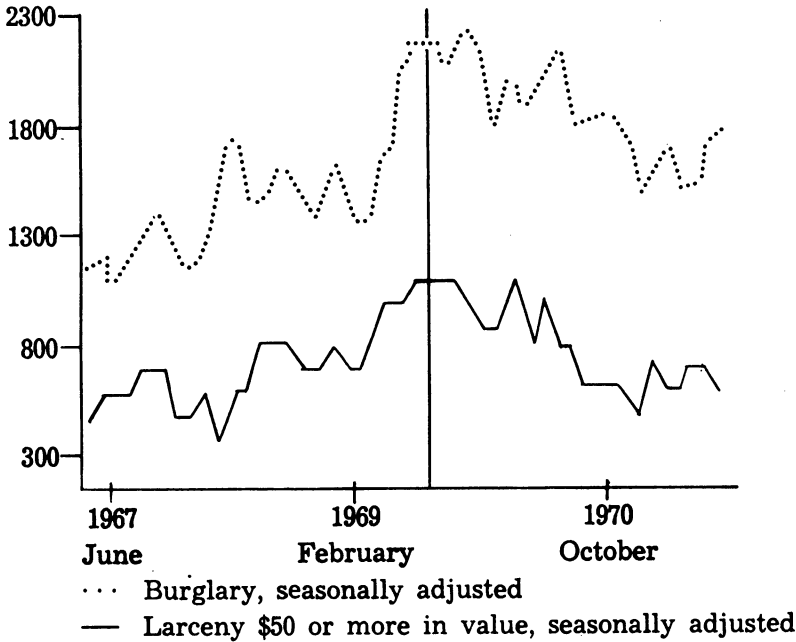


TABLE 4: DEPENDENT VARIABLE: BURGLARY, SEASONALLY ADJUSTED

Variable	Coefficient	T-Statistic
PERIOD: July 1967-August 1969		
Constant	627.945	3.691
Larceny Greater Than \$50	1.234	5.4263
Seasonally Adjusted		-2
		R = .532
		F _{1,24} = 29.445
		D = .962
PERIOD: September 1969-June 1971		
Constant	1071.561	7.010
Larceny Greater Than \$50	.957	5.041
		-2
		R = .538
		F _{1,20} = 25.413
		D = 1.763
PERIOD: July 1967-June 1971		
Constant	740.245	5.430
Larceny Greater Than \$50	1.223	6.948
		-2
		R = .501
		F _{1,48} = 48.269
		D = .910
Chow Test: F _{2,48} = 9.962		

Figure VI presents equivalent data for robbery. The value of the F-statistic in the Chow test (23.183) (see Table 5) confirms the visual impression that there is a rather sharp shift in the relationship at the time of Wilson's installation. The series for

robbery matches the series for major larcenies less well than does the burglary series.

FIGURE VI: MONTHLY CRIME STATISTICS, DISTRICT OF COLUMBIA

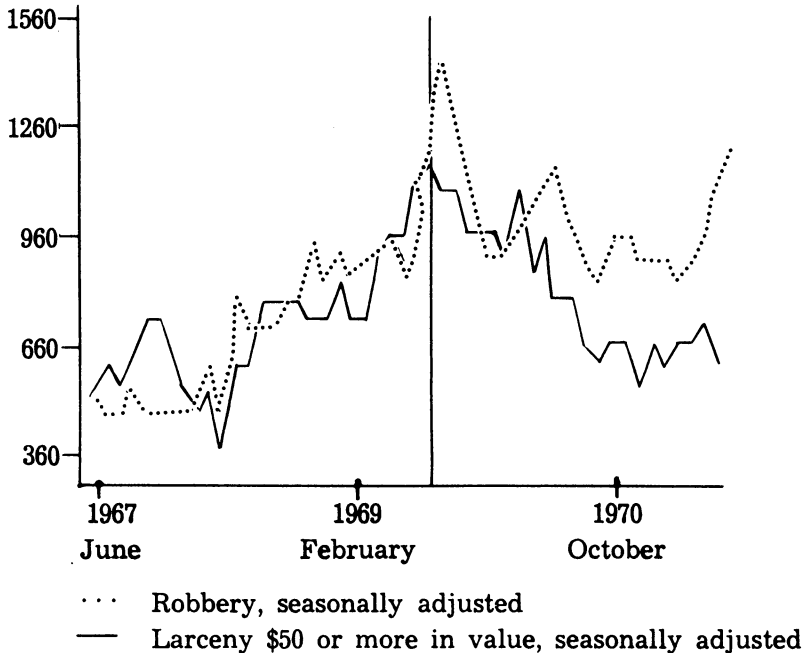


TABLE 5: DEPENDENT VARIABLE: ROBBERY, SEASONALLY ADJUSTED

Variable	Coefficient	T-Statistic
PERIOD: July 1967-August 1969		
Constant	36.372	0.349
Larceny Greater Than \$50 Seasonally Adjusted	.950	6.811
		-2
		R = .645
		F _{1,24} = 46.383
		D = .444
PERIOD: September 1969-June 1971		
Constant	657.282	4.808
Larceny Greater Than \$50	.455	2.681
		-2
		R = .228
		F _{1,20} = 7.188
		D = .695
PERIOD: July 1967-June 1971		
Constant	220.542	1.903
Larceny Greater Than \$50	.847	5.660
		-2
		R = .398
		F _{1,44} = 32.034
		D = .393
Chow Test: F _{2,44} = 23.183		

Accept the conclusion that downgrading of serious larcenies increased significantly following Wilson's installation. Accept, too, the testing procedure outlined above. Then it is reasonable to conclude that there has also been a significant increase in

the downgrading of burglaries. The same conclusion is not appropriate for robbery because of the highly significant change in the relationship. It is the absence of a large change in the relationship of another crime to serious larcenies which implies downgrading. (It was a large change which showed downgrading when we were dealing with the two larcenies. Now the assumption has been changed. Originally, minor larcenies served as a test of the honesty of serious larceny data. The subsequent tests of other crimes use serious larcenies as a standard of dishonesty.) Downgrading of robbery may or may not have increased. Our test cannot say.

District of Columbia: Conclusion

We conclude that at least part of the decline in the crime statistics for the District of Columbia is attributable to increased downgrading of larcenies and, to a lesser extent, of burglaries. This appears to be a pure case of the reactivity of a social indicator: the fact that the statistic is used as a measure of performance affects the statistic itself. The political importance of crime apparently caused pressures, subtle or otherwise, to be felt by those who record crime — pressures which have led to the downgrading of crimes.

This situation is unfortunate. Nixon's suggestion of the use of the District as a laboratory is not without merit. The proper conditions, however, were not met. While not all the shortcomings of the crime statistics as measures of crime could have been avoided, nor all the threats to both internal and external validity or inference overcome, the Nixon anti-crime program in the District could have illustrated

an experimental approach to social reform, an approach in which we try out new programs designed to cure specific social problems, in which we learn whether or not these programs are effective, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available (Campbell, 1969: 409).

Campbell (1969: 409) also suggests reasons for the failure:

It is one of the most characteristic aspects of the present situation that *specific reforms are advocated as though they were certain to be successful*. For this reason, knowing outcomes has immediate political implications. Given the inherent difficulty of making significant improvements by the means usually provided and given the discrepancy between promise and possibility, most administrators wisely prefer to limit the evaluations to those outcomes which they can control. (*Italics in original*).

Our data suggest that "administrators" did influence the level of Crime Index statistics in the District of Columbia. Increased misclassification brought the trend more closely into line with the stated goals of these administrators than anti-crime programs alone would have. The large element of discretion in the hands of the patrolman created the possibility. The stated goal of reducing crime announced the desired result. The linkages between managerial wish and the behavior of patrolmen need further investigation. But the existence of these linkages seems clearly demonstrated.

A Reaction

We extend the narrative beyond the analysis of reactivity of data to describe a reaction to the discovery of reactivity. A preliminary version of our analysis was presented in a Washington newspaper (Collins, 1971b). In response, the police provided, and the *Washington Post* printed, an alternative explanation of the peculiar pattern we had found in the larceny statistics:

. . . [I]n the past, numerous petty crimes, especially thefts involving \$5 or less, frequently were listed in a noncriminal complaint book along with property damage, dog bite and other minor civil complaints. The complaint books were kept in the outlying stationhouses and their criminal reports never absorbed into the department's central records.

In November, 1969, Chief Wilson ordered discontinuation of the complaint book. . . . Thus, the incidence of minor crimes, especially larcenies under \$50, would appear to have increased in proportion to the total of all offenses, police say (Valentine: 1971).

As the complaint books were "phased out" rather than suddenly eliminated, our findings are explained in plausible fashion.

While the *Post* obtained data on the number of larcenies of less than \$5 in value, it did not examine them. With the cooperation of the *Post*, we looked at these data. In our opinion, the gradual inclusion of these minor larcenies in the central statistics does not, in fact, explain the pattern we found. Subtracting these larcenies from larcenies less than \$50 in value yields a figure for larcenies less than \$50 but not less than \$5 in value. According to the complaint book theory, this statistic should be substantially unaffected by the phasing-out of the complaint books. Figure VII shows this statistic for the available time period together with larcenies less than \$50. It is evident that the subtraction of minor larcenies does not serious-

FIGURE VII: MONTHLY CRIME STATISTICS, DISTRICT OF COLUMBIA

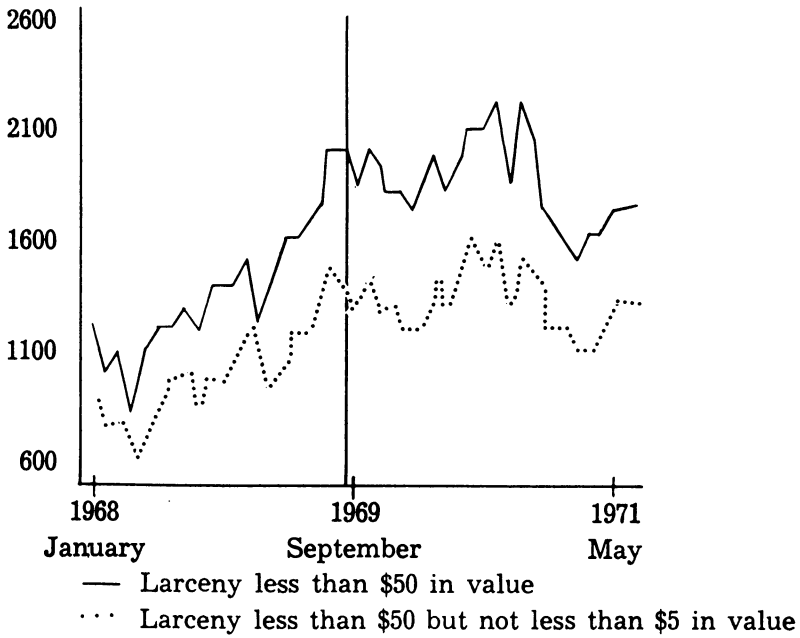
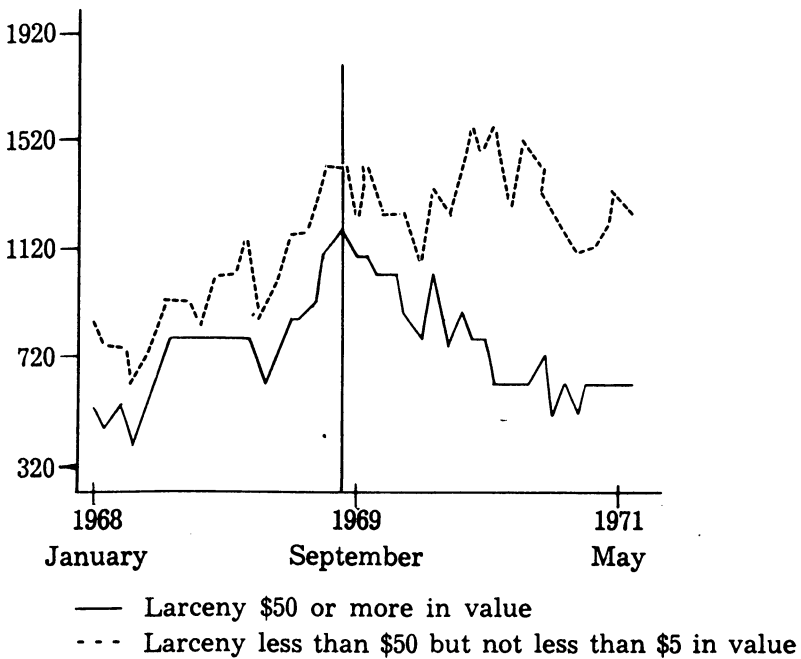


FIGURE VIII: MONTHLY CRIME STATISTICS, DISTRICT OF COLUMBIA



ly alter the pattern of the original series. The correlation between these two series is .989. Figure VIII is analogous to Figure III, substituting larcenies \$5-50 for larcenies less than \$50. Virtually the same pattern as found in Figure III emerges. The change in the relationship between the two types of larcenies may occur later in Figure VIII than in Figure III, but it does occur. A delay in this change is intuitively appealing: the impact of a new police chief might well not be instantaneous.

The police explanation brings to mind Niederhoffer's "principle of equilibrium": "It refers to an organizational imperative that requires the negation of any and all criticism. . . . The analysis unit is the main arm of defense in the struggle of the department to justify itself" (1969: 13-14).

ANALYSIS OF DATA FROM OTHER CITIES

The techniques we have used to investigate crime statistics, police behavior, and political pressures in the District of Columbia could be used equally well to investigate the statistics of other cities. But careful analysis depends upon a substantial body of contextual knowledge, including some awareness of day-to-day police administration. With limited resources, we did not obtain the information we would like to have before analyzing data from places other than the District. Nevertheless, we have examined monthly crime statistics from January 1962 through June 1971 for 29 other cities. They are not a random sample, but were selected on the basis of several criteria. We chose the six largest, a dozen in the District's range of population (250,000 to one million), and eleven others for assorted reasons, such as Law Enforcement Assistance Administration programs, geographical location, etc. (A complete list of these cities can be found in Appendix II.) The examination of these cities was quite casual compared to our analysis of the District, but the results give further indication of the sensitivity of crime statistics to factors other than crime. In most cases, we have found nothing in the pattern of the data which corresponds to the peculiarities in the District of Columbia data. But in some cases, the same peculiar relation emerges between statistics for larcenies of \$50 or greater in value and those for larcenies of less than \$50 in value. And one city's data suggest that there is sometimes reason for police to artificially *inflate* the statistics they report.²⁰ We will comment in some detail on two of these cities and very briefly on the rest.

Baltimore²¹

Manipulation of crime statistics is an old tradition in Baltimore. In 1965 the *Baltimore Sun* claimed that police had downgraded their reports of crime and even failed entirely to report some crimes. Shortly thereafter the police commissioner was fired. The International Association of Chiefs of Police sent a team of investigators, including Donald D. Pomerleau. The police department was reorganized and an "elaborate system of checks and balances to assure that police reports are accurately written and categorized" was instituted. Mr. Pomerleau was named as the new police commissioner.

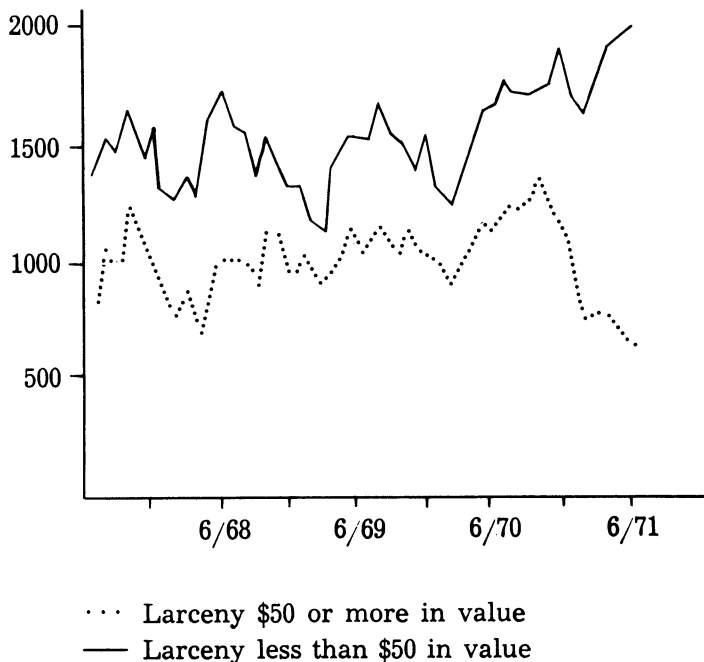
In July of 1971, the *Baltimore Sun* returned to the old theme with a series of articles quite similar to those by Patrick Collins in Washington discussed above. Police reports had been examined, and peculiarities in the valuation of stolen property were noted. Roger Twigg (1971) of the *Sun* claimed that he had interviewed patrolmen who said that pressure from their district commanders to reduce the crime rate had led to downgrading. Furthermore, "At least a half dozen patrolmen at Southern and Central districts as well as a sergeant in the detectives section . . . contend that orders were given to downgrade certain reports." Denials were issued by police officials. A deputy police commissioner was quoted as saying, "There is no competition to do this — no pressure brought to bear on their men to get the crime rate down." (Presumably he meant to say something other than what he did say.) Investigations were revealed, ordered, and discussed. Commissioner Pomerleau, obviously aware of the possibilities of crime statistics, is quoted as saying, "Humans — being humans — will interpret instructions: 'You've got to reduce crime,' differently."

The week before Twigg's initial story, Pomerleau "released statistics indicating a 15.1 per cent drop in major crimes during the first six months of this year, as compared with the same period in 1970." The parallel with the District of Columbia is striking.

The period covered by our data is too short to allow the type of statistical analysis employed with the District of Columbia data. Nor do we know what point should be taken as the "abrupt and decisive change" required for that analysis. We could let the data identify the point, but there is no real need for it. Visual inspection of the data will suffice for present purposes.

Figure IX presents the data for the two categories of larceny (not seasonally adjusted). It is striking that for the last four months the two series move in opposite directions for so long a period. These data, while not conclusive, support the *Sun's* assessment that the crime statistics in Baltimore reacted to increased pressures to reduce the number of recorded crimes. (We do not know why pressures might have increased at that time. It has been suggested to us that Commissioner Pomerleau was then planning to retire and wanted the record of his accomplishments to be clear.)

FIGURE IX: MONTHLY CRIME STATISTICS, BALTIMORE



We do not present the data for robbery and burglary here, but note that the pattern found in these data is similar to that in the high-value larceny data. Twigg found evidence that there was manipulation in the other categories as well.

Baltimore has been selected for the Impact Cities program of the Law Enforcement Assistance Administration. The major objective of this program is an immediate halt to increases in robbery and burglary, followed by a 20% decrease in these crimes over a five year period (Law Enforcement Assistance Administration, n.d.). We feel there is an excellent prospect

that the statistics will show the Impact Cities program to be successful in Baltimore.

Philadelphia

Philadelphia also has an established tradition of dubious crime statistics. Numerous conversations with professionals and academics concerned with crime statistics lead us to believe that the Philadelphia statistics are a standing joke in the field. Bell (1962: 152) relates some early history:

. . . [I]n 1953 the City of Brotherly Love reported 28,560 major crimes, as against 16,773 in 1951—a sudden jump of over 70 per cent. But there had been no invasion by criminals. Police Commissioner Thomas J. Gibbons, who assumed office in 1952 as part of the reform administration of Mayor Clark, had found that for years crime records, in order to minimize the amount of crime in the city, had been faked. One center-city district, he discovered, had handled 5,000 more complaints than it had recorded. A new central reporting system was installed, and as a result the number of “crimes” went up.

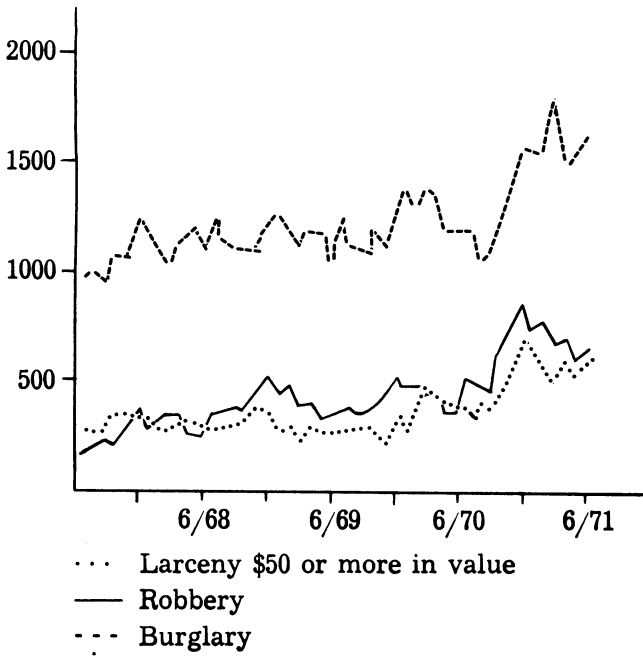
Either the Philadelphia police returned to their old ways sometime after that or the central reporting system was not a very efficient one. The fourth largest U.S. city, Philadelphia in 1970 reported fewer Index crimes than any other city among the ten largest. Indeed, Baltimore, with less than half the population of Philadelphia, reported over 60% more Index crimes in 1970 (Federal Bureau of Investigation, 1971).

It is well known that crime statistics are not comparable across jurisdictions, so these comparisons, extreme as they are, should not be a surprise. The analysis here, however, concerns the validity of comparisons across time within a jurisdiction. Figure X shows the relevant data for robbery, burglary, and larceny of property valued at \$50 or more. The period after September 1970 suggests that there has been a sharp increase in the average level of Index crimes in Philadelphia.

There is substantial room for “upgrading” of the Philadelphia statistics. The extent to which monthly minor larcenies outnumber serious larcenies is unusually great. In addition, the city’s large volume of arrests, perhaps partly attributable to arrest policies peculiar to Philadelphia, also could be an indication that reported crime is understated. Larger larcenies did not increase for a long period prior to 1970, as they did in most other cities. We are led to think that Philadelphia’s long-standing habit of understatement was abruptly changed during 1970.

The kind of analysis we attempted for the District of Columbia is not possible here, since it depends upon detailed

FIGURE X: MONTHLY CRIME STATISTICS, PHILADELPHIA



knowledge of police administration over the relevant time period. It has been suggested that the implementation of automated record-keeping by the department and the arrival of a new chief sometime after the noted change were both significant. In addition, the State of Pennsylvania established procedures for requiring reports from local departments and created a state audit capability for crime statistics on January 13, 1970. We must leave the question open; and we would be pessimistic about finding very solid answers, for here the "weak" design is threatened by several events which were abrupt and could have been decisive.

Other Cities

In the remaining 27 cities we discovered five cases of a pattern similar to the larceny pattern now familiar to the reader.²² While the existence of this pattern alone should be enough to cast doubt on the reliability of the crime statistics revealing it, we have no information about these cities which would suggest specific explanations for the emergence of the pattern at any particular time. Nor do we have any direct evidence, of the kind supplied by the *Washington Daily News*

or the *Baltimore Sun*, that the police are misvaluing property. But the number of cases in which we have found the pattern does suggest that the practice of altering established procedures for property valuation to shift the level of reported crime is not uncommon.

The failure of this analysis to discover a peculiar pattern should not be taken to indicate that the statistics in question are reliable. These techniques detect only some kinds of changes in reporting practices. They cannot, for example, detect changes which affect both larcenies equally. Nor can they detect "faking" of statistics which continues unchanged during the entire period covered by the data. So we can offer no comment on the state of crime statistics in those cities where we have detected no oddities.

CONCLUSION

Crime statistics, the Index crimes especially, are commonly used for several purposes. Crime became a significant concern of politicians and voters, and the statistics were widely used to evaluate the effectiveness of strategies of crime control. As the role of the police in the society became a matter of increased interest, crime statistics were used to evaluate the performance of police departments. Vast amounts of money are currently being made available for crime control programs, and crime statistics are used in some places to determine the allocation of these funds. The data generating system and the data generated are badly suited to these uses.

Each of these uses of crime statistics creates pressures to have the statistics show certain things. Sometimes the pressure is to show that crime is being reduced. Sometimes the pressure is to increase the number of crimes. These pressures impinge upon the data generating system, the police departments, and in some cases affect the statistics, entirely apart from the effects of the number of crimes which are actually committed. Consequently, those indicators almost invariably used for these purposes — the Index crimes of the Uniform Crime Reporting System — are highly misleading for what they are said to measure, in part simply because they are used as measures.

The techniques which we have used in this analysis can sometimes detect the nature of the reactivity of the measures. Sometimes they cannot. But even when they can, there is no way to tell with precision what the measures would have shown had they not been reactive. We conclude that the Uni-

form Crime Reporting System is useless as a tool for evaluation of social policy.

We do not suggest that it is impossible to use “crimes known to the police” as a measure of policy impact in all circumstances. Research might be designed carefully so that reported crimes can be a useful indicator of policy impact. But the Uniform Crime Reporting System does not constitute this kind of careful design.

A comment written more than forty years ago on another topic summarizes the reasons that Uniform Crime Reporting is not a useful tool for evaluation research:

The Government are very keen on amassing statistics—they collect them, add them, raise them to the n^{th} power, take the cube root and prepare wonderful diagrams. But what you must never forget is that every one of these figures comes in the first instance from the . . . (village watchman), who just puts down what he damn pleases (Stamp, 1929: 258-59).

AFTERWORD

Time does not stop to suit the convenience of those who analyze time-series data. Our stopping point, June 30, 1971, is arbitrary. While we resisted the temptation to redo the analysis each time the figures for an additional month became available, certain events subsequent to that date seem worth discussing. These are an audit of the District of Columbia crime statistics and a change in UCR categories.

The D.C. Audit

On October 13, 1971, after Collins' (1971a) initial article on larceny valuations had appeared, Chief Wilson emerged from a meeting with President Nixon and announced that an audit of the District's crime statistics would be conducted by an outside firm. He explained that the audit would occur because “there is always some skepticism about the reliability of crime statistics” (Angle, 1971). Months later—in March of 1972—the Metropolitan Police Department entered a contract for the audit with the consulting firm of Ernst & Ernst.

The Ernst & Ernst audit (1972) covered three fiscal years ending on June 30, 1972. The choice of that time period made it impossible for the audit to deal with the questions raised by our analysis: the baseline period we used falls outside Ernst & Ernst's period, so that our finding of change cannot be evaluated.

The report expresses doubt that officers could or would substantially downgrade crime reports because of risks involved in the practice:

When the original investigating officer arrives at the scene, he can report back to the Communications Center that no crime was detected, when in fact a crime was reported to him. He can incorrectly report back that the crime was less serious than originally reported by the caller. However, he takes the risk that complainants or prosecutors will later require a copy of the (offense report) which is available on request for insurance purposes or criminal proceedings. Further, the Field Inspection Division could uncover the deficient procedure through its review procedures (III-5).

In spite of this optimism, Ernst & Ernst conducted a special study of larceny reports. Samples of larcenies valued at between \$35 and \$65 were selected for each of the three fiscal years. Sample sizes and the division of the sample around the critical \$50 value are shown in Table 6. The steady rise in the proportion of the sample which falls below the critical point is con-

TABLE 6: ERNST & ERNST STUDY OF LARCENY VALUATIONS, RAW DATA

	Fiscal Year Ended June 30					
	1972		1971		1970	
	N	%	N	%	N	%
Larcenies valued between \$35 and \$50	96	64	77	54	61	41
Larcenies valued between \$50 and \$65	53	36	65	46	88	59
TOTAL	149	100	142	100	149	100

Source: Ernst & Ernst (1971: Table 8)

sistent with our analysis. Several different means were then used to estimate the correct value of the stolen property, and these estimates were compared with the official values to determine whether misvaluation had occurred. Of the 74 errors detected (16.8% of the sample), only one had no effect on classification of the incident. (The auditors were not interested in errors which had no effect.) Nine resulted in overstatement of seriousness, while 64 caused understatement. The distribution of these errors by fiscal years is shown in Table 7. There may

TABLE 7: ERRORS DETECTED IN LARCENY CLASSIFICATIONS

	Fiscal Year Ended June 30					
	1972		1971		1970	
	N	%	N	%	N	%
Caused Overstatement	3	8	2	15	4	17
Caused Understatement	35	92	11	85	18	79
No Effect	0	0	0	0	1	4
TOTAL	38	100	13	100	23	100

Source: Ernst & Ernst (1971: Table 8)

be an explanation of the unusually high figure for fiscal 1972:

In fiscal 1972, the MPD (Metropolitan Police Department) began inserting two stolen property valuations on larceny reports, the complainant's and the reporting officer's. Prior reports contained the reporting officer's only. This new information and the ability to contact complainants provided more information on which to base an opinion (Ernst & Ernst, 1972: Table 8).

The fiscal 1972 figures are probably a better indication of the extent of misclassification of larcenies than either of the two previous years.

If the figures of Table 6 are adjusted for misclassification, the clear trend in the proportion of the sample which falls below the critical \$50 value disappears (Table 8). It is also apparent that misclassification has a substantial effect on re-

TABLE 8: ERNST & ERNST STUDY OF LARCENY VALUATIONS, RAW DATA CORRECTED FOR MISCLASSIFICATION

	1972		1971		1970	
	N	%	N	%	N	%
Larcenies valued at less than \$50*	64	43	68	48	47	32
Larcenies valued at greater than \$50	85	57	74	52	102	68
TOTAL	149	100	142	100	149	100

*The value after revaluation is unknown, so that the original (Table 6) boundaries of \$35 and \$65 may be exceeded. The original source does not indicate in which category larcenies of exactly \$50 belong.

Source: Tables 6 and 7

ported larceny statistics. While the design of the Ernst & Ernst audit precludes examination of our hypothesis of an increase in the extent of misreporting coincident with the installation of Chief Wilson, the results of the audit disclose substantial misclassification during Chief Wilson's tenure. This is, of course, supportive of our findings.

Rather than quiet criticism of the Washington crime statistics, the audit, together with a *Washington Post* study, opened a new front. The *Post* discovered that the

basic failing of the system is that police do not file written reports on almost 40 per cent of citizen calls about major crimes—a violation of guidelines set by both the FBI and the International Association of Chiefs of Police (IACP).

Consequently, the *Post* concluded that "it is impossible to determine whether the city's crime rate is rising or falling" (Valentine, 1972a). Ernst & Ernst, commenting on the same phenomenon, note that

all of our tests and those of the Field Inspection Division indicated that the percentage of these kinds of errors would be very small in relation to the total offenses reported . . . (Ernst & Ernst, 1972: III-10).

Because we lack appropriate time series data, we cannot comment on the effect of this kind of nonreporting by the police on evaluations of the success of anti-crime programs in the District. Following the appearance of news stories concerning the rate of nonreporting, Chief Wilson called a news conference to deny that District crime statistics had been manipulated.

During that conference, he said:

I never had any doubt that my function in this city was to reduce crime and that if crime was not reduced I would probably be replaced as chief of police (Valentine, 1972b).

Reported crime was reduced. Wilson was not replaced.

Redefinition of Categories

Fifty dollars is no longer a critical point. In January 1973, the Federal Bureau of Investigation (1973) announced that beginning at that time all larcenies, regardless of value, would be included in the Index. Inflation had eroded the original significance of the \$50 figure.

This redefinition of Index larcenies has several effects. First, the incentive for misvaluation of stolen property is removed. Second, one technique for manipulating crime statistics has been eliminated. Third, the technique we have employed for detecting probable manipulation of crime statistics has become outdated. Fourth, the FBI's Crime Index now includes the theft of a 15¢ candy bar on a par with murder. Both the benefits and the costs of this redefinition appear to be small.

We might speculate that this minor change in the treatment of larcenies will be followed by more significant changes in government crime statistics. The death of J. Edgar Hoover has already led to change at the Federal Bureau of Investigation and more seems certain to come. Attorney General Richardson was not satisfied with the Uniform Crime Reporting System:

We have no mechanism in place for measuring the volume of crime committed . . . we have the FBI index, which measures the number of crimes that are recorded in police documents. But it is universally recognized that only a fraction of all crimes are reported to the police at all (Gentry, 1973).

We cannot now know whether there will be fundamental changes in the system of crime reporting. But any such change should be treated with extreme caution. The relationship between police performance and crimes reported remains poorly understood, but at least some of the properties of the Uniform Crime Reporting System are well known. It will take time to learn the shortcomings of any new system.

NOTES

- ¹ We know of no adequate study of the emergence of the crime issue. But see, *inter alia*, Scammon and Wattenberg (1970), Chester, Hodgson, and Page (1969), and Weisberg and Rusk (1970).
- ² For description and examples of this design, see Campbell (1969), Campbell and Ross (1968), and Campbell and Stanley (1966: 34-64).
- ³ These statistics are issued annually by the Federal Bureau of Investi-

- gation under the title *Crime in the United States: Uniform Crime Reports*, and quarterly in FBI press releases. The yearly publication as well as the system will be referred to as UCR.
- ⁴ Unrecorded crime, the "dark figure of crime," is the subject of a large literature. See especially Biderman and Reiss (1967) and the literature cited therein.
 - ⁵ Examples of police reports include the "Statistical Reports—Monthly Crime Index" of the Metropolitan Police Department, District of Columbia, and the "Monthly Report" of the Detroit Police Department.
 - ⁶ Frice (1966) attempts to use insurance rates to estimate crime rates. Victimization surveys are another approach to the problem. See Biderman (1967); Ennis (1967); Biderman, *et al.* (1967).
 - ⁷ Discussion of the Uniform Crime Reporting Program can be found in any UCR, or in the FBI's *Uniform Crime Reporting Handbook* (1966b).
 - ⁸ The most notable attempt at a weighting scheme is Sellin and Wolfgang (1964).
 - ⁹ Quotation marks are used because of complexities in the definition of "crime." This is not the place to discuss the issue, but as an indication of the nature of the problem, consider the classic question, "If a tree falls in the forest. . . ."
 - ¹⁰ Black (1970) discusses some of the factors relevant to this decision.
 - ¹¹ Revision of statistics subsequent to publication is not unknown. In late 1971, the District of Columbia police released a revision of the 1970 rape statistics. The new 1970 total was roughly twice the old (Taylor, 1971).
 - ¹² Additional illustrations may be found in President's Commission on Law Enforcement and the Administration of Justice (1967: 22-24). Some reporting changes are noted in the FBI quarterly press releases.
 - ¹³ Misclassification, a further source of variation in the statistics, will be discussed below.
 - ¹⁴ Insurance companies face a similar problem, and sometimes produce very different valuations. See Collins (1971a) and the discussion below.
 - ¹⁵ We assume that typically police desire to produce low crime rates (though a contrary case is discussed below), because evaluation of police performance often relies on crime rates. Another measure of performance is the "clearance rate," the percentage of crimes known to the police which the police believe have been "solved" (Skolnick, 1966: 168). There is an apparently simple relationship between the two measures: the lower the crime rate, the smaller the denominator of the clearance rate, and therefore the higher the clearance rate for any given number of "solutions". One instance of artful misclassification therefore improves both measures of performance. Caution must be observed, however, as it might be awkward to claim credit for the solution of a serious crime which has already been classified as non-serious.
- It might be thought that the desire to create resources for later plea bargaining would lead police to consider crimes as serious whenever possible. However, this presumes an arrest, and arrests are relatively rare for the three categories of crime we discuss.
- ¹⁶ The question of the statistical significance of the reversal need not concern us here. Various techniques are available for such significance tests. See, *e.g.* Box and Tiao (1965).
 - ¹⁷ Unless otherwise indicated, all monthly crime data were provided by the Uniform Crime Reporting Section of the FBI. We are extremely grateful for this assistance.
- Yearly statistics for larcenies less than \$50 in value are published in UCR, but quarterly figures have not been included in the quarterly FBI press releases. Until recently, the monthly Statistical Report of the District of Columbia police did not include the category either.
- ¹⁸ The number of cities in each category changes from year to year, in part because FBI population estimates change and in part because the FBI sometimes excludes cities which have misleading statistics.
 - ¹⁹ Prior to July, 1967, the District of Columbia police reported their statistics on the basis of a \$100 dividing line. This accounts for the sharp jump from 1967 to 1968.
 - ²⁰ This is not a novel suggestion. Bell (1962: 158) claims that the inflation of crime statistics in order to increase budget appropriations

“is not an uncommon feature of law enforcement in the United States today.”

²¹ The discussion of events in Baltimore is based entirely on Twigg (1971).

²² From January 1962 to June 1971, the pattern emerges, at some point and for some months, in Atlanta, Chicago, Cleveland, Oakland, and Seattle.

APPENDIX I: STATISTICAL PROCEDURES

The statistical question is: Was there a “real” (statistically significant) and continuing change in the relationship between the recorded monthly totals for two categories of crime at some particular time? To answer that question, we begin with a description of that relationship for the entire time period:

$$Y = A + BX + E$$

where X and Y are the two series;

A and B are unknown parameters; and

E is a random disturbance.

In other words, we assume that the number of recorded crimes Y in any given month is a linear function of the number of recorded crimes X and an “error,” which encapsulates other determinants of crime totals, random events, and so forth.

Least-squares regression (Johnston, 1963) applied to our data yields estimates of A and B. (These estimates are reported in the tables under the heading “Coefficient.”) There are several means of evaluating the adequacy of the assumed relationship, once the coefficients have been estimated. First, there is \bar{R}^2 , the (corrected) squared correlation coefficient, which indicates what proportion of the variation in Y is accounted for by the variation in X. (The remaining variation is attributed to random disturbances.) Second, the estimated values of A and B have associated t-statistics. Roughly speaking, if the t-statistic is greater than 2, the estimated value is statistically significant. (In other words, we are reasonably confident that the true value is not zero.) Third, the F-statistic provides a test of the statistical significance of the estimated equation as a whole, or, what amounts to the same thing, of \bar{R}^2 . If F is sufficiently large (and “sufficiently” depends upon the number of observations and the number of coefficients estimated), we are reasonably confident that the true correlation coefficient is not equal to zero. Finally, there is the Durbin-Watson statistic, D. One of the assumptions of the least-squares technique is that the errors are truly random. If knowledge of the value of the error in one time period provides information concerning the value of the error in the next time period, this assumption is violated. If D is approximately equal to 2, we can be reasonably confident

that this assumption is not violated in this way. As D departs from 2, we become less confident of the assumption. Thus a low D may indicate that there is some important variable which should have been included in the basic equation.

We next divide the data into two portions; one covering the period before the critical event, one covering the remaining time. The estimation procedure is then repeated for each of the two portions of the data, so that three separate pairs of estimates of A and B are obtained. If the relationship between X and Y did not change, all three pairs of estimates should be roughly the same. (They will not be exactly the same because of the random disturbances.) If they are, we should do almost as well explaining the variation in Y using all the data taken together as we do dividing the data into two portions. On the other hand, if the relationship did shift, if the true values of A and B changed at the critical point, explanation by pieces should work better than the single explanation (which will yield estimates falling between the original and the new A and B—estimates which are improper for both periods).

The Chow test compares the explanatory power of the single equation with that of the two separate equations. The procedure involves computing an F-statistic which, roughly speaking, is based on \bar{R}^2 from the three equations. If the value of that computed F is sufficiently large, we may with confidence reject the hypothesis that the relationship did not shift.

These statistical procedures, then, serve simply to confirm a judgment which, in this case, can be based on visual inspection of the graphs. Not all cases are so striking to the eye.

APPENDIX II: THE THIRTY CITIES

Atlanta, Georgia	Miami, Florida
Baltimore, Maryland	Milwaukee, Wisconsin
Berkeley, California	Minneapolis, Minnesota
Boston, Massachusetts	Newark, New Jersey
Buffalo, New York	New York, New York
Chicago, Illinois	Oakland, California
Cleveland, Ohio	Philadelphia, Pennsylvania
Columbia, South Carolina	Pittsburgh, Pennsylvania
Detroit, Michigan	Providence, Rhode Island
Flint, Michigan	Reading, Pennsylvania
Fresno, California	St. Louis, Missouri
Indianapolis, Indiana	St. Paul, Minnesota
Kansas City, Missouri	San Francisco, California
Los Angeles, California	Seattle, Washington
Louisville, Kentucky	Washington, D. C.

REFERENCES

We discovered, too late to discuss it in the text, an article which, drawing upon, *inter alia*, an earlier version of this paper, provides useful information on matters discussed here and related topics. Interested readers should consult "Note: Crime Statistics — Can They Be Trusted," 11 *Am. Crim. L. Rev.* 1045 (1973).

- ANGLE, Martha (1971) "Outside Firm to Audit D.C. Crime Data," *Washington Star*, October 13.
- BIDERMAN, Albert D. (1967) "Surveys of Population Samples for Estimating Crime Incidence," 374 *Annals* 16.
- BIDERMAN, Albert D. and Albert J. REISS, Jr. (1967) "On Exploring the 'Dark Figure' of Crime," 374 *Annals* 1.
- BIDERMAN, Albert D., *et al.* (1967) Report on a Pilot Study in the District of Columbia on Victimization and Attitudes Towards Law Enforcement. Washington, D.C.: U.S. Government Printing Office.
- BELL, Daniel (1962) *The End of Ideology* (rev. ed.). New York: Collier Books.
- BLACK, Donald J., "Production of Crime Rates," 35 *American Sociological Review* 733.
- BOX, G.E.P. and George C. TIAO (1965) "A Change in Level of a Non-stationary Time Series," 52 *Biometrika* 181.
- CAMPBELL, Donald T. (1969) "Reforms as Experiments," 24 *American Psychologist* 409.
- CAMPBELL, Donald T. and H. Laurence ROSS (1968) "The Connecticut Crackdown on Speeding," 3 *Law & Society Review* 33.
- CAMPBELL, Donald T. and Julian C. STANLEY (1966) *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- CHESTER, Lewis, Godfrey HODGSON, and Bruce PAGE (1969) *An American Melodrama*. New York: Viking Press.
- CLARK, Ramsey (1971) *Crime in America*. New York: Pocket Books.
- COLLINS, Patrick (1971a) "Are District Police Falsifying Reports to 'Reduce' Crime Rate?," *Washington Daily News*, August 25.
- (1971b) "D.C. Crime Dip! (Really?)," *Washington Daily News*, September 30.
- CONGRESSIONAL QUARTERLY (1971) "Crime: Nixon Approach, Its Costs and Results," 29 *Congressional Quarterly Weekly Report* 987.
- DUPONT, Robert (1970) "Narcotics Treatment as Crime Reduction and Human Resources Strategy," 1 *Trends* 4.
- ENNIS, Philip H. (1967) *Criminal Victimization in the United States*. Washington, D.C.: U.S. Government Printing Office.
- ERNST & ERNST (1972) Report of Work Performed under Contract NS 72332. Washington, D.C.: Ernst & Ernst.
- FEDERAL BUREAU OF INVESTIGATION (1966a) *Manual of Police Records*. Washington, D.C.: U.S. Government Printing Office.
- (1966b) *Uniform Crime Reporting Handbook*. Washington, D.C.: U.S. Government Printing Office.
- (1969) *Crime in the United States: Uniform Crime Reports—1968*. Washington, D.C.: U.S. Government Printing Office.
- (1971) *Crime in the United States: Uniform Crime Reports—1970*. Washington, D.C.: U.S. Government Printing Office.
- (1973) "Crime Index Offenses: Larceny," *Uniform Crime Reporting Newsletter*, January.
- GENTRY, Margaret (1973) "Richardson Seeks Crime Tools," *Washington Post*, August 6.
- JOHNSTON, J. (1963) *Econometric Methods*. New York: McGraw-Hill.
- KNEECE, Jack (1971) "Jury Is Still Out on Preventive Detention," *Washington Star*, October 4.
- LAW ENFORCEMENT ASSISTANCE ADMINISTRATION (n.d.) *The LEAA Impact Program*.

- MANN, Judy Luce (1971) "Suburban Law Expert Sees Varicous Figure Juggling Acts," *Washington Star*, August 27.
- MULVIHILL, Donald J. and Melvin M. TUMIN (1969) *Crimes of Violence*. Washington, D.C.: U.S. Government Printing Office.
- NEW YORK TIMES (1971) *Foam on the Crime Wave*, September 8, § IV, p. 16.
- (1972) *The Washington Crime Story*, January 17.
- NIEDERHOFFER, Arthur (1969) *Behind the Shield*. Garden City, New York: Doubleday Anchor.
- PRESIDENT'S COMMISSION ON CRIME IN THE DISTRICT OF COLUMBIA (1966) Report. Washington, D.C.: U.S. Government Printing Office.
- PRESIDENT'S COMMISSION ON LAW ENFORCEMENT AND THE ADMINISTRATION OF JUSTICE (1967) *Task Force Report: Crime and Its Impact—An Assessment*. Washington, D.C.: U.S. Government Printing Office.
- PRICE, James E. (1966) "A Test of the Accuracy of Crime Statistics," 14 *Social Problems* 214.
- REISS, Albert J. (1971) "Assessing the Current Crime Wave," in Melvin R. LEVIN (ed.) *Exploring Urban Problems*. Boston: Urban Press.
- ROSHIER, R. J. (1971) "Crime and the Press," 18 *New Society* 502 (September 16).
- ROSS, H. Laurence, Donald T. CAMPBELL, and Gene V. GLASS (1970) "Determining the Social Effects of a Legal Reform: The British 'Breathalyser' Crackdown of 1967," 13 *American Behavioral Scientist* 493.
- SCAMMON, Richard M. and Ben J. WATTENBERG (1970) *The Real Majority*. New York: Coward-McCann.
- SELLIN, Thorsten and Marvin WOLFGANG (1964) *The Measurement of Delinquency*. New York: Wiley.
- SKOLNICK, Jerome H. (1966) *Justice Without Trial*. New York: Wiley.
- STAMP, Sir Josiah (1929) *Some Economic Factors in Modern Life*. London: P.S. King & Son.
- SUTHERLAND, Edwin H. (1949) *White Collar Crime*. New York: Holt, Rinehart and Winston.
- TAYLOR, Walter (1971) "D.C. Figures Juggled: Crime Drop Reported," *Washington Star*, December 29.
- THOMPSON, John L. (1968) "Uniform Crime Reporting: Historical IACP Landmark," 35 *The Police Chief* 22 (February).
- TWIGG, Roger (1971) *Articles on Baltimore Crime Statistics*, *Baltimore Sun*, July 14-17, 24.
- VALENTINE, Paul W. (1971) "D.C. Crime Reports: What They Mean," *Washington Post*, October 26.
- (1972a) "No Reports Filed on 40% of Crime Calls," *Washington Post*, September 17.
- (1972b) "Wilson Denies Juggling Crime Figures," *Washington Post*, October 11.
- WASHINGTON POST (1971) "The Attorney General and Crime," September 11.
- WEBB, Eugene J., et al. (1966) *Unobtrusive Measures*. Chicago: Rand McNally.
- WEISBERG, Herbert F. and Jerrold G. RUSK (1970) "Dimensions of Candidate Evaluation," 64 *American Political Science Review* 1149.
- WILSON, Jerry V. (1971) Letter to Richard Hollander, *Washington Daily News*, August 27.