THE SPECIFIC DETERRENT EFFECTS OF ARREST FOR DOMESTIC ASSAULT*

LAWRENCE W. SHERMAN
University of Maryland, College Park and Police Foundation

RICHARD A. BERK
University of California, Santa Barbara

with

42 Patrol Officers of the Minneapolis Police Department,
Nancy Wester, Donileen Loseke, David Rauma, Debra Morrow, Amy Curtis,
Kay Gamble, Roy Roberts, Phyllis Newton, and Gayle Gubman

The specific deterrence doctrine and labeling theory predict opposite effects of punishment on individual rates of deviance. The limited cross-sectional evidence available on the question is inconsistent, and experimental evidence has been lacking. The Police Foundation and the Minneapolis Police Department tested these hypotheses in a field experiment on domestic violence. Three police responses to simple assault were randomly assigned to legally eligible suspects: an arrest; "advice" (including, in some cases, informal mediation); and an order to the suspect to leave for eight hours. The behavior of the suspect was tracked for six months after the police intervention, with both official data and victim reports. The official recidivism measures show that the arrested suspects manifested significantly less subsequent violence than those who were ordered to leave. The victim report data show that the arrested subjects manifested significantly less subsequent violence than those who were advised. The findings falsify a deviance amplification model of labeling theory beyond initial labeling, and fail to falsify the specific deterrence prediction for a group of offenders with a high percentage of prior histories of both domestic violence and other kinds of crime.

Sociologists since Durkheim ([1893] 1972:126) have speculated about how the punishment of individuals affects their behavior. Two bodies of literature, specific deterrence and labeling, have developed competing predictions (Thorsell and Klemke, 1972). Durkheim, for example, implicitly assumed with Bentham that the pains of punishment deter people from repeating the crimes for which they are punished, especially when punishment is certain, swift and severe. More recent work has fostered the ironic view that punishment often makes individuals more likely to commit crimes because of altered interactional structures, foreclosed legal opportunities and secondary deviance (Lemert, 1951, 1967; Schwartz and Skolnick, 1962; Becker, 1963).

Neither prediction can muster consistent empirical support. The few studies that allege effects generally employ weak designs in which it is difficult, if not impossible, to control plausibly for all important factors confounded with criminal justice sanctions and the rule-breaking behavior that may follow. Thus, some claim to show that punishment deters individuals punished (Clarke, 1966; F.B.I., 1967:34–44; Cohen and Stark, 1974:30; Kraut, 1976; Murray and Cox, 1979; McCord, 1983), while others claim to show that punishment increases their deviance (Gold and Williams, 1969; Shoham, 1974; Farrington, 1977; Klemke, 1978). Yet all of these studies suffer either methodological or conceptual flaws as tests of the effects of punishment (Zimring and Hawkins, 1973; Gibbs, 1975; Hirschi, 1975; Tittle, 1975), especially the confounding of incarceration with attempts to rehabilitate and the frequent failure to differentiate effects for different types of offenders and offenses (Lempert, 1981–1982).

Perhaps the strongest evidence to date comes from a randomized experiment conducted by Lincoln et al. (unpubl.). The experiment randomly assigned juveniles, who had already been apprehended, to four different treatments ranked in their formality: release; two types of diversion; and formal

* Direct all correspondence to: Lawrence W. Sherman, Police Foundation, 1909 K Street N.W., Washington, D.C. 20006.

This paper was supported by Grant #80–IJ–CX–0042 to the Police Foundation from the National Institute of Justice, Crime Control Theory Program. Points of view or opinions stated in this document do not necessarily represent the official position of the U.S. Department of Justice, the Minneapolis Police Department, or the Police Foundation.

We wish to express our thanks to the Minneapolis Police Department and its Chief, Anthony V. Bouza, for their cooperation, and to Sarah Fenstermaker Berk, Peter H. Rossi, Albert J. Reiss, Jr., James Q. Wilson, Richard Lempert, and Charles Tittle for comments on an earlier draft of this paper.

charging. The more formal and official the processing, the more frequent the repeat criminality over a two-year follow-up period. This study supports labeling theory for arrested juveniles, although it cannot isolate the labeling or deterrent effects of arrest per se.

In all likelihood, of course, punishment has not one effect, but many, varying across types of people and situations (Chambliss, 1967; Andenaes, 1971). As Lempert (1981–1982:523) argues, "it is only by attending to a range of such offenses that we will be able to develop a general theory of deterrence." The variables affecting the deterrability of juvenile delinquency, white-collar crime, armed robbery and domestic violence may be quite different. Careful accumulation of findings from different settings will help us differentiate the variables which are crime- or situation-specific and those which apply across settings.

In this spirit, we report here a study of the impact of punishment in a particular setting, for a particular offense, and for particular kinds of individuals. Over an eighteen-month period, police in Minneapolis applied one of three intervention strategies in incidents of misdemeanor domestic assault: arrest; ordering the offender from the premises; or some form of advice which could include mediation. The three interventions were assigned randomly to households, and a critical outcome was the rate of repeat incidents. The relative effect of arrest should hold special interest for the specific deterrence–labeling controversy.

POLICING DOMESTIC ASSAULTS

Police have been typically reluctant to make arrests for domestic violence (Berk and Loseke, 1981), as well as for a wide range of other kinds of offenses, unless victims demand an arrest, the suspect insults the officer, or other factors are present (Sherman, 1980). Parnas's (1972) qualitative observations of the Chicago police found four categories of police action in these situations: negotiating or otherwise "talking out" the dispute; threatening the disputants and then leaving; asking one of the parties to leave the premises; or (very rarely) making an arrest.

Similar patterns are found in many other cities. Surveys of battered women who tried to have their domestic assailants arrested report that arrest occurred in 10 percent (Roy, 1977:35) or 3 percent (see Langley and Levy, 1977:219) of the cases. Surveys of police agencies in Illinois (Illinois Law Enforcement Commission, 1978) and New York (Office of the Minority Leader, 1978) found explicit policies against arrest in the majority of the agencies surveyed. Despite the fact that violence is reported to be present in one-third (Bard and Zacker, 1974) to two-thirds (Black, 1980) of all domestic disturbances police respond to, police department data show arrests in only 5 percent of those disturbances in Oakland (Hart, n.d., cited in Meyer and Lorimer, 1977:21), 6 percent of those disturbances in a Colorado city (Patrick et al., n.d., cited in Meyer and Lorimer, 1977:21) and 6 percent in Los Angeles County (Emerson, 1979).

The best available evidence on the frequency of arrest is the observations from the Black and Reiss study of Boston, Washington and Chicago police in 1966 (Black, 1980:182). Police responding to disputes in those cities made arrests in 27 percent of violent felonies and 17 percent of the violent misdemeanors. Among married couples (Black, 1980:158), they made arrests in 26 percent of the cases, but tried to remove one of the parties in 38 percent of the cases.

An apparent preference of many police for separating the parties rather than arresting the offender has been attacked from two directions over the last fifteen years. The original critique came from clinical psychologists, who agreed that police should rarely make arrests (Potter, 1978:46; Fagin, 1978:123–24) in domestic assault cases, and argued that police should mediate the disputes responsible for the violence. A highly publicized demonstration project teaching police special counseling skills for family crisis intervention (Bard, 1970) failed to show a reduction in violence, but was interpreted as a success nonetheless. By 1977, a national survey of police agencies with 100 or more officers found that over 70 percent reported a family crisis intervention training program in operation. While it is not clear whether these programs reduced separation and increased mediation, a decline in arrests was noted for some (Wylie et al., 1976). Indeed, many sought explicitly to reduce the number of arrests (University of Rochester, 1974; Ketterman and Kravitz, 1978).

By the mid-1970s, police practices were criticized from the opposite direction by feminist groups. Just as psychologists succeeded in having many police agencies respond to domestic violence as "half social work and half police work," feminists began to argue that police put "too much emphasis on the social work aspect and not enough on the criminal" (Langley and Levy, 1977:218). Widely publicized lawsuits in New York and Oakland sought to compel police to make arrests in every case of domestic assault, and state legislatures were lobbied successfully to reduce the evidentiary requirements needed for police to make arrests for misdemeanor domestic assaults. Some legislatures are now

considering statutes requiring police to make arrests in these cases.

The feminist critique was bolstered by a study (Police Foundation, 1976) showing that for 85 percent of a sample of spousal homicides, police had intervened at least once in the preceding two years. For 54 percent of the homicides, police had intervened five or more times. But it was impossible to determine from the cross-sectional data whether making more or fewer arrests would have reduced the homicide rate.

In sum, police officers confronting a domestic assault suspect face at least three conflicting options, urged on them by different groups with different theories. The officers' colleagues might recommend forced separation as a means of achieving short-term peace. Alternatively, the officers' trainers might recommend mediation as a means of getting to the underlying cause of the "dispute" (in which both parties are implicitly assumed to be at fault). Finally, the local women's organizations may recommend that the officer protect the victim (whose "fault," if any, is legally irrelevant) and enforce the law to deter such acts in the future.

RESEARCH DESIGN

In response to these conflicting recommendations, the Police Foundation and the Minneapolis Police Department agreed to conduct a randomized experiment. The design called for random assignment of arrest, separation, and some form of advice which could include mediation at the officer's discretion. In addition, there was to be a six-month follow-up period to measure the frequency and seriousness of domestic violence after each police intervention. The advantages of randomized experiments are well known and need not be reviewed here (see, e.g., Cook and Campbell, 1979).

The design only applied to simple (misdemeanor) domestic assaults, where both the suspect and the victim were present when the police arrived. Thus, the experiment included only those cases in which police were empowered (but not required) to make arrests under a recently liberalized Minnesota state law; the police officer must have probable cause to believe that a cohabitant or spouse had assaulted the victim within the last four hours (but police need not have witnessed the assault). Cases of life-threatening or severe injury, usually labeled as a felony (aggravated assault), were excluded from the design for ethical reasons.

The design called for each officer to carry a pad of report forms, color coded for the three different police actions. Each time the officers encountered a situation that fit the experiment's criteria, they were to take whatever action was indicated by the report form on the top of the pad. We numbered the forms and arranged them in random order for each officer. The integrity of the random assignment was to be monitored by research staff observers riding on patrol for a sample of evenings.

After police action was taken, the officer was to fill out a brief report and give it to the research staff for follow-up. As a further check on the randomization process, the staff logged in the reports in the order in which they were received and made sure that the sequence corresponded to the original assignment of treatments.

Anticipating something of the victims' background, a predominantly minority, female research staff was employed to contact the victims for a detailed face-to-face interview, to be followed by telephone follow-up interviews every two weeks for 24 weeks. The interviews were designed primarily to measure the frequency and seriousness of victimizations caused by the suspect after the police intervention.[1] The research staff also collected criminal justice reports that mentioned the suspect's name during the six-month follow-up period.

CONDUCT OF THE EXPERIMENT

As is common in field experiments, implementation of the research design entailed some slippage from the original plan. In order to gather data as quickly as possible, the experiment was originally located in the two Minneapolis precincts with the highest density of domestic violence crime reports and arrests. The 34 officers assigned to those areas were invited to a three-day planning meeting and asked to participate in the study for one year. All but one agreed. The conference also produced a draft order for the chief's signature specifying the rules of the experiment. These rules created several new situations to be excluded from the experiment, such as if a suspect attempted to assault police officers, a victim persistently demanded an arrest, or if both parties were injured. These additional exceptions, unfortunately, allowed for the possibility of differential attrition from the separation and mediation treatments. The im-

[1] The protocols were based heavily on instruments designed for an NIMH-funded study of spousal violence conducted by Richard A. Berk, Sarah Fenstermaker Berk, and Ann D. Witte (Center for Studies of Crime and Delinquency, Grant #MH-34616–01). A similar protocol was developed for the suspects, but only twenty-five of them agreed to be interviewed.

plications for internal validity are discussed later.

The experiment began on March 17, 1981, with the expectation that it would take about one year to produce about 300 cases (it ran until August 1, 1982, and produced 330 case reports.) The officers agreed to meet monthly with the project director (Sherman) and the project manager (Wester). By the third or fourth month, two facts became clear: (1) only about 15 to 20 officers were either coming to meetings or turning in cases; and (2) the rate at which the cases were turned in would make it difficult to complete the project in one year. By November, we decided to recruit more officers in order to obtain cases more rapidly. Eighteen additional officers joined the project, but like the original group, most of these officers only turned in one or two cases. Indeed, three of the original officers produced almost 28 percent of the cases, in part because they worked a particularly violent beat, and in part because they had a greater commitment to the study. Since the treatments were randomized by officer, this created no internal validity problem. However, it does raise construct validity problems to which we will later return.

There is little doubt that many of the officers occasionally failed to follow fully the experimental design. Some of the failures were due to forgetfulness, such as leaving the report pads at home or at the police station. Other failures derived from misunderstanding about whether the experiment applied in certain situations; application of the experimental rules under complex circumstances was sometimes confusing. Finally, from time to time there were situations that were simply not covered by the experiment's rules.

Whether any officers intentionally subverted the design is unclear. The plan to monitor randomization with ride-along observers broke down because of the unexpectedly low incidence of cases meeting the experimental criteria. The observers had to ride for many weeks before they observed an officer apply one of the treatments. We tried to solve this problem with "chase-alongs," in which the observers rode in their own car with a portable police radio and drove to the scene of any domestic call dispatched to any officer in the precinct. Even this method failed.

Thus, we are left with at least two disturbing possibilities. First, police officers anticipating (e.g., from the dispatch call) a particular kind of incident, and finding the upcoming experimental treatment inappropriate, may have occasionally decided to void the experiment. That is, they may have chosen to exclude certain cases in violation of the experimental design. This amounts to differential attrition, which is clearly a threat to internal validity. Note that if police officers blindly decided to exclude certain cases (e.g., because they did not feel like filling out the extra forms on a given day), all would be well for internal validity.

Second, since the recording officer's pad was supposed to govern the actions of each pair of officers, some officers may also have switched the assignment of driver and recording officer after deciding a case fit the study in order to obtain a treatment they wanted to apply. If the treatments were switched between driver and recorder, then the internal validity was again threatened. However, this was almost certainly uncommon because it was generally easier not to fill out a report at all than to switch.

Table 1 shows the degree to which the treatments were delivered as designed.[2] Ninety-nine percent of the suspects targeted for arrest actually were arrested, while only 78 percent of those to receive advice did, and only 73 percent of those to be sent out of the residence for eight hours were actually sent. One explanation for this pattern, consistent with the experimental guidelines, is that mediating and sending were more difficult ways for police to control the situation, with a greater likelihood that officers might resort to arrest as a fallback position. When the assigned treatment is arrest, there is no need for a fallback position. For example, some offenders may have refused to comply with an order to leave the premises.

Such differential attrition would potentially bias estimates of the relative effectiveness of arrest by removing uncooperative and difficult offenders from the mediation and separation treatments. Any deterrent effect could be underestimated and, in the extreme, artifactual support for deviance amplification could be found. That is, the arrest group would have too many "bad guys" *relative* to the other treatments.

We can be more systematic about other factors affecting the movement of cases away from the designed treatments. The three delivered treatments represent a polychotomous outcome amenable to multivariate statistical analysis. We applied a multinominal logit formulation (Amemiya, 1981:1516–19; Maddala, 1983:34–37), which showed that the designed treatment was the dominant cause of the treatment actually received (a finding suggested by Table 1). However, we also found that five other variables had a statistically sig-

[2] Sixteen cases were dropped because no treatment was applied or because the case did not belong in the study (i.e., a fight between a father and son).

Table 1. Designed and Delivered Police Treatments in Spousal Assault Cases

Designed Treatment	Delivered Treatment			Total
	Arrest	Advise	Separate	
Arrest	98.9% (91)	0.0% (0)	1.1% (1)	29.3% (92)
Advise	17.6% (19)	77.8% (84)	4.6% (5)	34.4% (108)
Separate	22.8% (26)	4.4% (5)	72.8% (83)	36.3% (114)
Total	43.4% (136)	28.3% (89)	28.3% (89)	100% (314)

nificant effect on "upgrading" the separation and advice treatments to arrests: whether police reported the suspect was rude; whether police reported the suspect tried to assault one (or both) of the police officers; whether police reported weapons were involved; whether the victim persistently demanded a citizen's arrest; and whether a restraining order was being violated. We found no evidence that the background or characteristics of the suspect or victim (e.g., race) affected the treatment received.

Overall, the logit model fit the data very well. For well over 80 percent of the cases, the model's predicted treatment was the same as the actual treatment (i.e., correct classifications), and minor alterations in the assignment threshold would have substantially improved matters. Moreover, a chi-square test on the residuals was not statistically significant (i.e., the observed and predicted treatments differed by no more than chance). In summary, we were able to model the assignment process with remarkable success simply by employing the rules of the experimental protocol (for more details, see Berk and Sherman, 1983).

We were less fortunate with the interviews of the victims; only 205 (of 330, counting the few repeat victims twice) could be located and initial interviews obtained, a 62 percent completion rate. Many of the victims simply could not be found, either for the initial interview or for follow-ups: they either left town, moved somewhere else or refused to answer the phone or doorbell. The research staff made up to 20 attempts to contact these victims, and often employed investigative techniques (asking friends and neighbors) to find them. Sometimes these methods worked, only to have the victim give an outright refusal or break one or more appointments to meet the interviewer at a "safe" location for the interview.

The response rate to the bi-weekly follow-up interviews was even lower than for the initial interview, as in much research on women crime victims. After the first interview, for which the victims were paid $20, there was a gradual falloff in completed interviews with each successive wave; only 161 victims provided all 12 follow-up interviews over the six months, a completion rate of 49 percent. Whether paying for the follow-up interviews would have improved the response rate is unclear; it would have added over $40,000 to the cost of the research. When the telephone interviews yielded few reports of violence, we moved to conduct every fourth interview in person, which appeared to produce more reports of violence.

There is absolutely no evidence that the experimental treatment assigned to the offender affected the victim's decision to grant initial interviews. We estimated a binary logit equation for the dichotomous outcome: whether or not an initial interview was obtained. Regressors included the experimental treatments (with one necessarily excluded), race of the victim, race of the offender, and a number of attributes of the incident (from the police sheets). A joint test on the full set of regressors failed to reject the null hypothesis that all of the logit coefficients were zero. More important for our purposes, none of the t-values for the treatments was in excess of 1.64; indeed, none was greater than 1.0 in absolute value. In short, while the potential for sample selection bias (Heckman, 1979; Berk, 1983) certainly exists (and is considered later), that bias does not stem from obvious sources, particularly the treatments. This implies that we may well be able to meaningfully examine experimental effects for the subset of individuals from whom initial interviews were obtained. The same conclusions followed when the follow-up interviews were considered.

In sum, despite the practical difficulties of controlling an experiment and interviewing crime victims in an emotionally charged and violent social context, the experiment succeeded in producing a promising sample of 314 cases with complete official outcome measures and an apparently unbiased sample of responses from the victims in those cases.

RESULTS

The 205 completed initial interviews provide some sense of who the subjects are, although the data may not properly represent the characteristics of the full sample of 314. They show the now familiar pattern of domestic violence cases coming to police attention being disproportionately unmarried couples with lower than average educational levels, disproportionately minority and mixed race (black male, white female), and who were very likely to have had prior violent incidents with police intervention. The 60 percent suspect unemployment rate is strikingly high in a community

Table 2. Victim and Suspect Characteristics: Initial Interview Data and Police Sheets

A. Unemployment		
Victims 61%		
Suspects 60%		
B. Relationship of Suspect to Victim		
Divorced or separated husband	3%	
Unmarried male lover	45%	
Current husband	35%	
Wife or girlfriend	2%	
Son, brother, roommate, other	15%	
C. Prior Assaults and Police Involvement		
Victims assaulted by suspect, last six months	80%	
Police intervention in domestic dispute, last six months	60%	
Couple in Counseling Program	27%	
D. Prior Arrests of Male Suspects		
Ever Arrested For Any Offense	59%	
Ever Arested For Crime Against Person	31%	
Ever Arrested on Domestic Violence Statute	5%	
Ever Arrested On An Alcohol Offense	29%	
E. Mean Age		
Victims 30 years		
Suspects 32 years		
F. Education	Victims	Suspects
< high school	43%	42%
high school only	33%	36%
> high school	24%	22%
G. Race	Victims	Suspects
White	57%	45%
Black	23%	36%
Native American	18%	16%
Other	2%	3%

N = 205 (Those cases for which initial interviews were obtained)

with only about 5 percent of the workforce unemployed. The 59 percent prior arrest rate is also strikingly high, suggesting (with the 80 percent prior domestic assault rate) that the suspects generally are experienced lawbreakers who are accustomed to police interventions. But with the exception of the heavy representation of Native Americans (due to Minneapolis' unique proximity to many Indian reservations), the characteristics in Table 2 are probably close to those of domestic violence cases coming to police attention in other large U.S. cities.

Two kinds of outcome measures will be considered. One is a *police-recorded* "failure" of the offender to survive the six-month follow-up period without having police generate a written report on the suspect for domestic violence, either through an offense or an arrest report written by any officer in the department, or through a subsequent report to the project research staff of a randomized (or other) intervention by officers participating in the experiment. A second kind of measure comes from the *interviews with victims,* in which victims were asked if there had been a repeat incident with the same suspect, broadly defined to include an actual assault, threatened assault, or property damage.

The two kinds of outcomes were each formulated in two complementary ways: as a dummy variable (i.e., repeat incident or not) and as the amount of time elapsed from the treatment to either a failure or the end of the follow-up period. For each of the two outcomes, three analyses were performed: the first using a linear probability model; the second using a logit formulation; and the third using a proportional hazard approach. The dummy outcome was employed for the linear probability and logit analyses, while the time-to-failure was employed for the proportional hazard method.[3]

Given the randomization, we began in traditional analysis of variance fashion. The official measure of a repeat incident was regressed on the treatment received for the sub-

[3] In addition to the linear probability model, the logit and proportional hazard formulations can be expressed in forms such that the outcome is a probability (e.g., the probability of a new violent incident). However, three slightly different response functions are implied. We had no theoretical basis for selecting the proper response function, and consequently used all three. We expected that the substantive results could be essentially invariant across the three formulations.

Table 3. Experimental Results for Police Data

Variable	Linear Coef	Linear t-value	Logistic Coef	Logistic t-value	Proportional Hazard Rate Coef	Proportional Hazard Rate t-value
Intercept (separate)	0.24	5.03*	−1.10	−4.09*	—	—
Arrest	−0.14	−2.21*	−1.02	−2.21*	−0.97	−2.28*
Advise	−0.05	−0.79	−0.31	−0.76	−0.32	−0.88
	F = 2.01		Chi-square = 5.19		Chi-square = 5.48	
	P = .07		P = .07		P = .06	
N = 314						

* p < .05, two-tailed test.

set of 314 cases (out of 330) that fell within the definition of the experiment. Compared to the baseline treatment of separation, which had the highest recidivism rate in the police data, the arrest treatment reduced repeat occurrences by a statistically significant amount (t = −2.38). Twenty-six percent of those separated committed a repeat assault, compared to 13 percent of those arrested. The mediation treatment was statistically indistinguishable from the other two. To help put this in perspective, 18.2 percent of the households failed overall.

The apparent treatment effect for arrest in this conventional analysis was suggestive, but there was a danger of biased estimates from the "upgrading" of some separation and advise treatments. In response, we applied variations on the corrections recommended by Barnow et al. (1980: esp. 55). In brief, we inserted instrumental variables in place of the delivered treatments when the treatment effects were analyzed. These instruments, in turn, were constructed from the multinomial logit model described earlier.[4]

Table 3 shows the results of the adjusted models. The first two columns report the results for the linear probability approach. Again, we find a statistically significant effect for arrest (t = −2.21). However, it is well known that the linear probability model will produce inefficient estimates of the regression coefficients and biased (and inconsistent) estimates of the standard errors. Significance tests, therefore, are suspect. Consequently, we also estimated a logit model, with pretty much the same result. At the mean of the endogenous variable (i.e., 18.2 percent), the logit coefficient for arrest translates into nearly the same effect (i.e., −.15) found with the linear probability model (t = −2.21).

One might still object that the use of a dummy variable outcome neglects right-hand censoring. In brief, one cannot observe failures that occur after the end of the experimental period, so that biased (and inconsistent) results follow. Thus, we applied a proportional hazard analysis (Lawless, 1982: Ch. 7) that adjusts for right-hand censoring. In this model the time-to-failure dependent variable is transformed into (roughly) the probability at any given moment during the six-month follow-up period of a new offense occurring, given that no new offenses have yet been committed. The last two columns of Table 3 indicate that, again, an effect for arrest surfaces (t = −2.28). The coefficient of −0.97 implies that compared to the baseline of separation, those experiencing an arrest were less likely to commit a new battery by a multiplicative factor of .38 (i.e., e raised to the −0.97 power). If the earlier results are translated into comparable terms, the effects described by the proportional hazard formulation are the largest we have seen (see footnote 4). But the major message is that the arrest effect holds up under three different statistical methods based on slightly different response functions. Overall, the police data indicate that the separation treatment produces the highest recidivism, arrest produces the lowest, with the impact of "advise" (from doing nothing to mediation) indistinguishable from the other two effects.

Table 4 shows the results when self-report data are used. A "failure" is defined as a new assault, property destruction or a threatened assault. (Almost identical results follow from a definition including only a new assault.) These results suggest a different ordering of the effects, with arrest still producing the lowest recidivism rate (at 19%), but with advice producing the highest (37%).

Overall, 28.9 percent of the suspects in Table 4 "failed." Still, the results are much the same as found for the official failure measure. However, given the effective sample of 161, we are vulnerable to sample selection bias. In response, we applied Heckman's (1979) sample selection corrections. The results were virtually unchanged (and are therefore not reported).

[4] We did *not* simply use the conditional expectations of a multinomial logit model. We used an alternative procedure to capitalize on the initial random assignment. The details can be found in Berk and Sherman (1983).

Table 4. Experimental Results for Victim Report Data

	Linear		Logistic		Proportional Hazard Rate	
Variable	Coef	t-value	Coef	t-value	Coef	t-value
Intercept (advise)	0.37	5.54*	−0.53	−1.70	—	—
Arrest	−0.18	−2.00*	−0.94	−2.01*	−0.82	−2.05*
Separate	−0.04	−0.35	−0.15	−0.10	−0.27	−0.09
	F = 2.31		Chi-square = 4.78		Chi-square = 4.36	
	P = .10		P = .09		P = .11	

N = 161 (Those cases for which *all* follow-up interviews were obtained)

* $p < .05$, two-tailed test.

An obvious rival hypothesis to the deterrent effect of arrest is that arrest incapacitates. If the arrested suspects spend a large portion of the next six months in jail, they would be expected to have lower recidivism rates. But the initial interview data show this is not the case: of those arrested, 43 percent were released within one day, 86 percent were released within one week, and only 14 percent were released after one week or had not yet been released at the time of the *initial* victim interview. Clearly, there was very little incapacitation, especially in the context of a six-month follow-up. Indeed, virtually all those arrested were released before the first follow-up interview. Nevertheless, we introduced the length of the initial stay in jail as a control variable. Consistent with expectations, the story was virtually unchanged.

Another perspective on the incapacitation issue can be obtained by looking at repeat violence which occurred shortly after the police intervened. If incapacitation were at work, a dramatic effect should be found in households experiencing arrest, especially compared to the households experiencing advice. Table 5 shows how quickly the couples were reunited, and of those reunited in one day, how many of them, according to the victim, began to argue or had physical violence again. It is apparent that *all* of the police interventions effectively stopped the violence for a 24-hour period after the couples were reunited. Even the renewed quarrels were few, at least with our relatively small sample size. Hence, there is again no evidence for an incapacitation effect. There is also no evidence for the reverse: that arrested offenders would take it out on the victim when the offender returned home.

DISCUSSION AND CONCLUSIONS

The experiment's results are subject to several qualifications. One caution is that both kinds of outcome measures have uncertain construct validity. The official measure no doubt neglects a large number of repeat incidents, in part because many of them were not reported, and in part because police are sometimes reluctant to turn a family "dispute" into formal police business. However, the key is whether there is *differential* measurement error by the experimental treatments; an undercount randomly distributed across the three treatments will not bias the estimated experimental effects (i.e., only the estimate of the intercept will be biased). It is hard to imagine that differential undercounting would come solely from the actions of police, since most officers were not involved in the experiment and could not have known what treatment had been delivered.

However, there might be differential undercounting if offenders who were arrested were less likely to remain on the scene after a new assault. Having been burned once, they might not wait around for a second opportunity. And police told us they were less likely during the follow-up period (and more generally) to record an incident if the offender was not present. For example, there would be no arrest forms since the offender was not available to arrest. If all we had were the official outcome

Table 5. Speed of Reunion and Recidivism by Police Action

	Time of Reunion					
Police Action	Within One Day	More than One Day but Less Than One Week	Longer or No Return	(N)	New Quarrel Within A Day	New Violence Within A Day
Arrested (and released)	38%	30%	32%	(N=76)	(2)	(1)
Separated	57%	31%	10%	(N=54)	(6)	(3)
Advised	—	—	—	(N=72)	(4)	(1)

N = 202 (Down from the 205 in Table 2 due to missing data)

measures, there would be no easy way to refute this possibility. Fortunately, the self-report data are *not* vulnerable on these grounds, and the experimental effects are found nevertheless.

It is also possible that the impact for arrest found in the official outcome measure represents a reluctance of *victims* to call the police. That is, for some victims, the arrest may have been an undesirable intervention, and rather than face the prospect of another arrest from a new incident, these victims might decide not to invoke police sanctions. For example, the arrest may have cost the offender several days' work and put financial stress on the household. Or the offender may have threatened serious violence if the victim ever called the police again. However, we can again observe that the self-report data would not have been vulnerable to such concerns, and the experimental effects were found nevertheless. The only way we can see how the self-report data would fail to support the official data is if respondents in households experiencing arrest became more hesitant to admit to *interviewers* that they had been beaten a second time. Since there was no differential response rate by treatment, this possibility seems unlikely. If the arrested suspects had intimidated their victims more than the other two treatment groups, it seems more likely that such intimidation would have shown up in noncooperation with the interviews than in differential underreporting of violence in the course of the interviews.

This is not to say that the self-report data are flawless; indeed there is some reason to believe that there was undercounting of new incidents. However, just as for the official data, unless there is differential undercounting by the experimental treatments, all is well. We can think of no good reasons why differential undercounting should materialize. In summary, internal validity looks rather sound.

The construct validity of the treatments is more problematic. The advice and separation interventions have unclear content. Perhaps "good" mediation, given consistently, would fare better compared to arrest. The more general point is that the treatment effects for arrest are only relative to the impact of the other interventions. Should their content change, the relative impact of arrest could change as well.

Likewise, we noted earlier that a few officers accounted for a disproportionate number of the cases. What we have been interpreting, therefore, as results from different intervention strategies could reflect the special abilities of certain officers to make arrest particularly effective relative to the other treatments. For example, these officers may have been less skilled in mediation techniques. However, we re-estimated the models reported in Tables 3 and 4, including an interaction effect to capture the special contributions of our high-productivity officers. The new variable was not statistically significant, and the treatment effect for arrest remained.

Finally, Minneapolis is hardly representative of all urban areas. The Minneapolis Police Department has many unusual characteristics, and different jurisdictions might well keep suspects in custody for longer or shorter periods of time. The message should be clear: external validity will have to wait for replications.

Despite these qualifications, it is apparent that we have found no support for the deviance amplification point of view. The arrest intervention certainly did not make things worse and may well have made things better. There are, of course, many rejoinders. In particular, over 80 percent of offenders had assaulted the victims in the previous six months, and in over 60 percent of the households the police had intervened during that interval. Almost 60 percent of the suspects had previously been arrested for something. Thus, the counterproductive consequences of police sanction, if any, may for many offenders have already been felt. In labeling theory terms, secondary deviation may already have been established, producing a ceiling for the amplification effects of formal sanctioning. However, were this the case, the arrest treatment probably should be less effective in households experiencing recent police interventions. No such interaction effects were found. In future analyses of these data, however, we will inductively explore interactions with more sensitive measures of police sanctioning and prior criminal histories of the suspects.

There are, of course, many versions of labeling theory. For those who theorize that a metamorphosis of self occurs in response to official sanctions over a long period of time, our six-month follow-up is not a relevant test. For those who argue that the development of a criminal self-concept is particularly likely to occur during a lengthy prison stay or extensive contact with criminal justice officials, the dosage of labeling employed in this experiemnt is not sufficient to falsify that hypothesis. What this experiment does seem to falsify for this particular offense is the broader conception of labeling implicit in the prior research by Lincoln et al. (unpubl.), Farrington (1977) and others: that for every possible increment of criminal justice response to deviance, the more increments (or the greater the formality) applied to the labeled deviant, the greater the likelihood of subsequent deviation. The absolute strength of the dosage is irrelevant to this hypothesis, as long as some variation in dosage

is present. While the experiment does not falsify all possible "labeling theory" hypotheses, it does at least seem to falsify this one.

The apparent support for deterrence is perhaps more clear. While we certainly have no evidence that deterrence will work in general, we do have findings that swift imposition of a sanction of temporary incarceration may deter male offenders in domestic assault cases. And we have produced this evidence from an unusually strong research design based on random assignment to treatments. In short, criminal justice sanctions seem to matter for this offense in this setting with this group of experienced offenders.

A number of police implications follow. Perhaps most important, police have historically been reluctant to make arrests in domestic assault cases, in part fearing that an arrest could make the violence worse. Criminal justice sanctions weakly applied might be insufficient to deter and set the offender on a course of retribution. Our data indicate that such concerns are by and large groundless.

Police have also felt that making an arrest was a waste of their time: without the application of swift and severe sanctions by the courts, arrest and booking had no bite. Our results indicate that only three of the 136 arrested offenders were formally punished by fines or subsequent incarceration. This suggests that arrest and initial incarceration alone may produce a deterrent effect, regardless of how the courts treat such cases, and that arrest makes an independent contribution to the deterrence potential of the criminal justice system. Therefore, in jurisdictions that process domestic assault offenders in a manner similar to that employed in Minneapolis, we favor a *presumption* of arrest; an arrest should be made unless there are good, clear reasons why an arrest would be counterproductive. We do not, however, favor *requiring* arrests in all misdemeanor domestic assault cases. Even if our findings were replicated in a number of jurisdictions, there is a good chance that arrest works far better for some kinds of offenders than others and in some kinds of situations better than others.[5] We feel it best to leave police a loophole to capitalize on that variation. Equally important, it is widely recognized that discretion is inherent in police work. Simply to impose a requirement of arrest, irrespective of the features of the immediate situation, is to invite circumvention.

[5] Indeed, one of the major policy issues that could arise from further analysis of the interaction effects would be whether police discretion should be guided by either achieved or ascribed relevant suspect characteristics.

REFERENCES

Amemiya, Takeshi
1981 "Qualitative response models: a survey." Journal of Economic Literature 19:1483–1536.

Andenaes, Johannes
1971 "Deterrence and specific offenses." University of Chicago Law Review 39:537.

Bard, Morton
1970 "Training police as specialists in family crisis intervention." Washington, D.C.: U.S. Department of Justice.

Bard, Morton and Joseph Zacker
1974 "Assaultiveness and alcohol use in family disputes—police perceptions." Criminology 12:281–92.

Barnow, Burt S., Glen G. Cain and Arthur S. Goldberger
1980 "Issues in the analysis of selectivity bias." Pp. 53–59 in Ernst W. Stromsdorfer and George Farkas (eds.), Evaluation Studies Review Annual, Volume 5. Beverly Hills: Sage.

Becker, Howard
1963 The Outsiders. New York: Free Press.

Berk, Richard A.
1983 "An introduction to sample selection bias in sociological data." American Sociological Review, 48:386–98.

Berk, Richard A. and Lawrence W. Sherman
1983 "Police responses to family violence incidents: an analysis of an experimental design with incomplete randomization." Unpublished manuscript, Department of Sociology, University of California at Barbara.

Berk, Sarah Fenstermaker and Donileen R. Loseke
1981 "Handling family violence: situational determinants of police arrest in domestic disturbances." Law and Society Review 15:315–46.

Black, Donald
1980 The Manners and Customs of the Police. New York: Academic Press.

Chambliss, William
1967 "Types of deviance and the effectiveness of legal sanctions." Wisconsin Law Review 1967:703–19.

Clarke, Ronald V. G.
1966 "Approved school boy absconders and corporal punishment." British Journal of Criminology: 6:364–75.

Cohen, Lawrence E. and Rodney Stark
1974 "Discriminatory labeling and the five-finger discount." Journal of Research in Crime and Delinquency 11:25–39.

Cook, Thomas D. and Donald T. Campbell
1979 Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

Durkheim, Emile
[1893] Selected Writings. Edited with an Introduction by Anthony Giddens. [Selection from Division of Labor in Society, 6th edition, 1960 (1893)] Cambridge: Cambridge University Press.
1972

Emerson, Charles D.
1979 "Family violence: a study by the Los

Angeles County Sheriff's Department." Police Chief 46(6):48–50.

Fagin, James A.
1978 "The effects of police interpersonal communications skills on conflict resolution." Ph.D. Dissertation, Southern Illinois University Ann Arbor: University Microfilms.

Farrington, David P.
1977 "The effects of public labeling." British Journal of Crininology 17:112–25.

Federal Bureau of Investigation
1967 Uniform Crime Reports. Washington, D.C.: U.S. Department of Justice.

Gold, Martin and Jay Williams
1969 "National study of the aftermath of apprehension." Prospectus 3:3–11.

Gibbs, Jack P.
1975 Crime, Punishment and Deterrence. New York: Elsevier.

Heckman, James
1979 "Sample selection bias as a specification error." Econometrica 45:153–61.

Hirschi, Travis
1975 "Labeling theory and juvenile delinquency: an assessment of the evidence." Pp. 181–203 in Walter R. Gove (ed.), The Labeling of Deviance. New York: Wiley.

Illinois Law Enforcement Commission
1978 "Report on technical assistance project—domestic violence survey." (Abstract). Washington, D.C.: National Criminal Justice Reference Service.

Ketterman, Thomas and Marjorie Kravitz
1978 Police Crisis Intervention: A Selected Bibliography. Washington, D.C.: National Criminal Justice Reference Service.

Klemke, Lloyd W.
1978 "Does apprehension for shoplifting amplify or terminate shoplifting activity?" Law and Society Review 12:391–403.

Kraut, Robert E.
1976 "Deterrent and definitional influences on shoplifting." Social Problems 23:358–68.

Langley, Richard and Roger C. Levy
1977 Wife Beating: The Silent Crisis. New York: E. P. Dutton.

Lawless, Jerald F.
1982 Statistical Models and Methods for Lifetime Data. New York: Wiley.

Lemert, Edwin M.
1951 Social Pathology. New York: McGraw-Hill.
1967 Human Deviance, Social Problems and Social Control. Englewood Cliffs, NJ: Prentice-Hall.

Lempert, Richard.
1981–1982 "Organizing for deterrence: lessons from a study of child support." Law and Society Review 16:513–68.

Lincoln, Suzanne B., Malcolm W. Klein, Katherine S. Teilmann and Susan Labin
unpubl. "Control organizations and labeling theory: official versus self-reported delinquency." Unpublished manuscript, University of Southern California.

Maddala, G. S.
1983 Limited, Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.

McCord, Joan
1983 "A longitudinal appraisal of criminal sanctions." Paper presented to the IXth International Congress on Criminology, Vienna, Austria, September.

Meyer, Jeanie Keeny and T. D. Lorimer
1977 Police Intervention Data and Domestic Violence: Exploratory Development and Validation of Prediction Models. Report prepared under grant #RO1MH27918 from National Institute of Mental Health. Kansas City, Mo., Police Department.

Murray, Charles A. and Louis A. Cox, Jr.
1979 Beyond Probation. Beverly Hills: Sage.

Office of the Minority Leader, State of New York
1978 Battered Women: Part I (Abstract). Washington, D.C.: National Criminal Justice Reference Service.

Parnas, Raymond I.
1972 "The police response to the domestic disturbance." Pp. 206–36 in Leon Radzinowicz and Marvin E. Wolfgang (eds.), The Criminal in the Arms of the Law. New York: Basic Books.

Police Foundation
1976 Domestic Violence and the Police: Studies in Detroit and Kansas City. Washington, D.C.: The Police Foundation.

Potter, Jane
1978 "The police and the battered wife: the search for understanding." Police Magazine 1:40–50.

Roy, Maria (ed.)
1977 Battered Women. New York: Van Nostrand Reinhold.

Schwartz, Richard and Jerome Skolnick
1962 "Two studies of legal stigma." Social Problems 10:133–42.

Sherman, Lawrence W.
1980 "Causes of police behavior: the current state of quantitative research." Journal of Research in Crime and Delinquency 17:69–100.

Shoham, S. Giora
1974 "Punishment and traffic offenses." Traffic Quarterly 28:61–73.

Thorsell, Bernard A. and Lloyd M. Klemke
1972 "The labeling process: reinforcement and deterrent." Law and Society Review 6:393–403.

Tittle, Charles
1975 "Labeling and crime: an empirical evaluation." Pp. 157–79 in Walter R. Gove (ed.), The Labeling of Deviance. New York: Wiley.

University of Rochester
1974 "FACIT—Family Conflict Intervention Team Experiment—Experimental Action Program." (Abstract). Washington, D.C.: National Criminal Justice Reference Service.

Wylie, P. B., L. F. Basinger, C. L. Heinecke and J. A. Reuckert
1976 "Approach to evaluating a police program

of family crisis interventions in sex demonstration cities—Final report." (Abstract). Washington, D.C.: National Criminal Justice Reference Service.

Zimring, Franklin E. and Gordon T. Hawkins
1973 Deterrence: The Legal Threat in Crime Control. Chicago: University of Chicago Press.

AN APPROACH TO SENSITIVITY ANALYSIS IN SOCIOLOGICAL RESEARCH*

JAE-ON KIM
University of Iowa

Sociologists, in particular, and social scientists, in general, are plagued by both weak theory and weak data. When the existing theory is insufficient to make all the relevant parameters of a statistical model identified, we often introduce dubious assumptions to make the model identifiable. In a similar way, when the variables under examination do not meet the measurement levels required by the statistical model, we sometimes pretend as if they do. As a means of evaluating the extent to which the parameter estimates are sensitive to such potential errors in model specification and data, a new approach to sensitivity analysis is proposed. The sensitivity index, which is the key to the new approach, has been widely used in different disciplines under a variety of different names.

As we employ increasingly powerful statistical tools in sociological research, we are often faced with problems caused by weak theory and weak data. On one hand, we may not have sufficient theory to fully specify the statistical model, and on the other, we may have to rely on data containing a variety of nonrandom errors. Faced with these problems, we could refrain from using these techniques and thereby forgo important explorations which only such powerful methods can provide. Or, we could proceed as if there were no problems. For instance, we may introduce enough arbitrary assumptions to make the model identifiable and go on to interpret the results as usual; likewise we could treat the variables that are measured, at best, on an ordinal scale as if they met more stringent measurement requirements.

A sensitivity analysis, which evaluates how sensitive the conclusions of a statistical analysis are to possible misspecifications in the model and errors in the data, can alleviate the problem. Such an analysis cannot in itself solve inherent problems in weak theory or data, but it can minimize the chance of making false conclusions. Better still, it may occasionally provide us with the happy finding that some parameter estimates are insensitive to minor inadequacies in model specification and data.

The main objectives of this paper are to propose a general approach to sensitivity analysis, to apply it to a sociological problem, and to persuade researchers to use it as an indispensable part of their data analysis. Sociologists (e.g., Duncan, 1969; Hauser, 1971; Jencks et al., 1972) have used a variety of sensitivity analyses, which Land and Felson (1978) reviewed. These approaches, however, address only the problems that arise when weak theory forces the researcher to introduce arbitrary assumptions into the model, but not the problem of errors in data. The approach proposed here can deal with both misspecification in the statistical model and possible errors in data.

The key to this new approach is an index of sensitivity, which has been used in other disciplines, under different names, and in slightly different contexts.[1] By introducing a new

* Direct all correspondence to: Jae-On Kim, Department of Sociology, University of Iowa, Iowa City, IA 52242.

The author is grateful to Jane Hansen, Barry Markovsky, Fred Pampel, Toby Parcel, David Parton, Jay Teachman, and Diana Hix Warner, who have read earlier versions of the paper and given valuable comments, and to two anonymous reviewers of *ASR* for their comments and advice.

[1] The mathematical foundation for it is readily available in standard calculus texts under the heading of implicit function theorem or total differential (see, e.g., Courant and John, 1974). Readers interested in pursuing the topic further may consult literature dealing with the Lagrange multiplier or undetermined coefficients in optimization, comparative statics and input-output analysis in economics, and