

# THE IMPACT OF RESEARCH ON LEGAL POLICY: THE MINNEAPOLIS DOMESTIC VIOLENCE EXPERIMENT

LAWRENCE W. SHERMAN  
ELLEN G. COHN

Experimental research has recently shown a powerful impact on legal policy. An experiment demonstrating a deterrent effect of arrest on domestic violence has shaped public policy. Stimulated by efforts to publicize the results of the Minneapolis Domestic Violence Experiment (Sherman and Berk, 1984a), police departments were persuaded to adopt an arrest policy for misdemeanor domestic violence. Over one-third of respondents from U.S. police departments in 117 cities said their policy had been influenced by the experiment, although respondents from some departments that adopted an arrest policy did not recognize the experiment or its results. Lempert (1987, 1984), citing medical research as precedent, suggests that this impact is premature and inappropriate until replications are completed. However, we find no indication that medical research employs a standard of delaying adoption of research results prior to replication. Our analysis suggests that publicity can encourage replication of legal research at other sites and thus improve the knowledge base for policy recommendations.

When should researchers refrain from publicizing results and thus possibly influencing legal policy? Until recently, this question was at most an interesting hypothetical, because most research was ignored by policy makers. Lately, however, some social science research has shown a demonstrable effect on legal policy (Petersilia, 1987). Responding to this apparent influence, some researcher-critics have attacked the published social research as being too flawed or limited to serve as a reliable guide to legal decision making.

The Martinson (1974) and Lipton *et al.* (1975) research on the

---

This article was supported in part by grant 86IJCXK043 to the Crime Control Institute from the National Institute of Justice. Opinions expressed in this paper are solely those of the authors, and do not reflect the official position of the National Institute of Justice. We wish to thank Charles Wellford, Edwin E. Hamilton, Albert J. Reiss, Jr., Patrick R. Gartin, and the staff of the Crime Control Institute. Joel Garner provided valuable comments on an earlier draft, which was presented to the Australian Institute of Criminology's National Conference on Domestic Violence, Canberra, November 1985. We also wish to thank Shari Diamond, Richard Lempert, and an anonymous peer reviewer for their suggestions on the revised draft.

LAW & SOCIETY REVIEW, Volume 23, Number 1 (1989)

ineffectiveness of rehabilitation, for example, was widely cited by policy-making opponents of rehabilitative programs, yet attacked by correctional researchers for concluding too much from too little evidence (see Sechrest, *et al.*, 1979: 34). The Greenwood (1982) scale of predictive factors for selective incapacitation was apparently quite influential among criminal sentencing decision makers, but researchers criticized it as being too often inaccurate (Blumstein *et al.*, 1986: 180). The Kansas City Preventive Patrol Experiment (Kelling *et al.*, 1974) has been widely discussed in decisions about police staffing levels but criticized by researchers as being too weak a test of the deterrent value of patrol (Larson, 1975; Fienberg *et al.*, 1976; Sherman, 1986a: 362–364). All three studies received substantial press attention when they were released.

In all of this criticism, however, we can find no suggestion that the studies should not have been publicized. Yet Lempert (1984) has suggested in this journal just that about another highly publicized study, the Minneapolis Domestic Violence Experiment (Sherman and Berk, 1984b). The experiment, a randomized test of police actions in response to misdemeanor domestic assault, found that arrest produced less repeat violence than did two nonarrest alternatives.

Lempert (1987: B15) describes the Minneapolis study as “arguably the best field experiment on a criminal justice policy problem done to date.” But he also (1984: 509) suggests that the results have been “prematurely and unduly publicized, and that police departments that have changed their arrest practices in response to this research may have adopted an innovation that does more harm than good.”

Lempert’s principal argument against publicizing the Minneapolis experiment is that publicity should await replication, at least when the theoretical basis for the research is poorly understood and when there may be substantial variations in effects across jurisdictions—conditions that apply to virtually all sociological research. He suggests that legal research should follow medical practice, which, he says (*ibid.*, p. 510), “painstakingly tests new drugs for safety and effectiveness before putting them into general distribution.” Lempert’s specific claim (1987: B15) about the Minneapolis experiment is that “if an anti-cancer drug had been tested instead of an anti-crime drug, substantial additional testing would have been required before the drug was made available for general distribution.” He also suggests (1984: 510) a need for more systematic study of how research is publicized and influences policy, a general issue made all the more important by the apparently growing influence of research on legal behavior.

The first part of this article describes how the Minneapolis experiment was publicized and traces its apparent impact on urban police policy, drawing on participants’ recollections and a three-wave panel survey of urban police departments. The second part

examines Lempert's claims that the publicity and influence were premature, given the limitations of the Minneapolis research, the evidentiary basis for the police practices that preceded that experiment, and the actual practices in medical research. The article concludes that legal effectiveness research should be actively publicized when it is completed, if only to speed the process of producing the multi-site replications Lempert properly advocates.

## I. PUBLICITY AND POLICY INFLUENCE

### A. *Publicizing the Research*

Lempert claims that the Minneapolis experiment was *unduly* publicized, which suggests that somewhat less publicity might have been acceptable. This claim implies that researchers—or any who attempt to publicize a story—can tightly control the amount of coverage the story receives. Yet the available social research on news organizations indicates that while it may be possible to prevent a story from becoming news through rigid secrecy, it is very hard for newsmakers to obtain less (or more) coverage for a story once it is released (Sigal, 1973; Roshco, 1975; Gans, 1979).

If a topic is intrinsically “hot” enough, simply presenting a paper at an academic meeting can produce national publicity. A 1988 paper on the negative effects of full-time child care merely presented at an academic conference on infant studies was reported prominently in the *Washington Post* (Evans, 1988). More often, however, attempts to publicize fail to attract as much coverage as the publicizers seek. There is a finite amount of newspaper space and broadcasting time available, and the competition for it is intense.

The Minneapolis experiment was announced in Washington, a city in which 535 members of Congress, all federal agency heads, and thousands of special interest organizations compete for national and local news coverage. In this context, little scientific news attracts press attention. Although federal agencies supporting research routinely put out press releases summarizing results, many are never reported or at most receive a few paragraphs buried deep in the *New York Times*, the *Washington Post*, or wire service coverage that editors in cities around the country may or may not ignore.

One purpose of doing policy research, as promised in the grant proposal for the experiment (Sherman, 1980), is to influence decision-making elites. Press maneuvers are designed for that target audience (Sigal, 1973: 143). Press coverage also helps justify continued or increased funding for research programs (*Footnotes*, 1987: 2). The Washington culture of social science policy research takes news-seeking for granted and views news-getting as very good luck.

The Minneapolis experiment received extensive coverage.

While it is probably true, as Lempert (personal communication, 1988) suggests, that the experiment would not have received as much publicity if it had not been actively promoted, the mere attempt to promote is no guarantee of success. The interesting question is why there was *so much* news coverage. Several factors may account for the amount of attention.

**1. Intrinsic Factors.** The substance of the research touched two topics of eternal human interest: sex and violence. But it also addressed the major changes in domestic power relationships produced over the preceding two decades by the feminist movement. The very idea of using police to redress the imbalance of physical power between men and women in love relationships was, in Gans's (1979: 146) terms, a highly "suitable" story for editors. The suitability extended to several editorial domains: national news, local police news, science news, and women's news. It thus satisfied a major suitability criterion of impact on large numbers of people (*ibid.*, p. 151). And, as Lempert (1984) suggests, the results were part of a larger story about the increasing punitiveness of American criminal justice.

**2. Extrinsic Factors.** The researchers, one of whom has advocated that social scientists should generally view publicity-seeking as a duty rather than a vice (Sherman, 1986b), attempted to orchestrate the release of the experimental results for maximum press coverage. Such attempts do not usually succeed, but the strong intrinsic factors made the decisions about how to "manage" the story more important. Those decisions included persuading the Minneapolis area public television station to film a documentary on the research during the field phase of the experiment, which provided action footage for several national television news shows covering the announcement of the experimental results many months later—by which time there would have been no experiment to film.

More important was Sherman's decision to release the story in two waves: the preliminary results in April 1983 and the final results in May 1984. This decision entailed the risk of violating the press's "repetition taboo" (Gans, 1979: 169), which can apply for twelve to twenty-four months. But the preliminary results were released through an exclusive story in the *New York Times* Tuesday "Science" section (Boffey, 1983), and not through a press release. The *Times* story, which the reporter agreed to write only if the research was legitimated by comments of three distinguished senior scholars who had read the preliminary report, was picked up by 107 newspapers nationally,<sup>1</sup> including editorials in the *Washington Post* and *New York Times*. It was also covered by the

<sup>1</sup> Newspaper statistics are derived from the Burelle's Clipping Service count for the Police Foundation, which paid Burelle's on a continuing basis to

CBS "Morning News" on May 6, 1983, in several hundred papers by columnist Ellen Goodman (1983), and in a story in the *Times* of London (Lamb, 1983), which was cited in a successful effort to legislate greater arrest powers for domestic violence in New South Wales, Australia. The story was not picked up by the Associated Press or United Press International wire services for general dissemination in the United States.

When the study was complete, the research was published in the *American Sociological Review* (Sherman and Berk, 1984b), a column summarizing the project was accepted for publication in the *Wall Street Journal* (Sherman and Bouza, 1984), and a simplified version of the report was mailed by the Police Foundation to some 3,000 policy makers. Sherman then decided to seek additional publicity through a Police Foundation press release. As a result of internal battles at the foundation, however, permission was refused. Sherman then recommended that the National Institute of Justice (NIJ) staff put out a press release. The manager of the research program funding the experiment, Joel Garner (who was later picked to manage its subsequent replications), agreed.<sup>2</sup>

A third key decision was to release the final results on the Sunday of Memorial Day weekend. This timing increased the chances that the story would face less competition on a "slow news day" (Gans, 1979: 146), even more so than Sundays without a three-day weekend. The wire services and over 300 newspapers carried the story. The PBS "McNeil-Lehrer News Hour," which had been notified well in advance, also ran a ten-minute segment on the study, followed by CBS "Evening News" three months later. In the nine months after the NIJ press release, the Police Foundation received over 2,500 requests for the report from around the country, many of them from citizen's groups, battered women's shelters, and the like. The foundation had received only some 500 requests in the year after the release of the preliminary findings.

The reason for publicizing this research was not to convey the detailed technical content of the report but to get the attention of the key audiences affecting police department policies. The research summary distributed by the Police Foundation was explicitly *not* pro-arrest for its own sake but rather pro-research, with a commitment to see research have an influence on policy. One specific objective was to have the key audiences at least *recognize* the message out of the masses of information they receive every day. Another objective was to have the audiences recall the information with *accuracy* and then use this information in thinking about the relevant policy issues. At the most ambitious level, Sherman

---

read about 1,000 newspapers nationwide and clip any story mentioning the foundation.

<sup>2</sup> Garner had to overcome opposition, however, by colleagues who thought the story was old news.

wanted the audiences to be influenced by the recommendations of the research and to become more willing to conduct replications and randomized experiments in general. The data presented below provide some indication of the effects of that publicity strategy on the key audiences' recognition and accuracy of recall of the findings as well as the findings' influence on police policy.

### *B. The Impact of the Publicity*

A national telephone survey of police departments serving cities of over 100,000 people was completed just before the second wave of the Minneapolis results was released (Sherman with Hamilton, 1984).<sup>3</sup> The survey was repeated in June 1985 (Sherman and Cohn, 1986) and in June 1986 (Cohn and Sherman, 1987).

**1. Methods.** Both the 1984 and 1985 telephone surveys used the same universe of cities as the sampling frame, with the minor exception that the 1984 survey identified 173 cities of 100,000, while the 1985 survey identified three more cities inadvertently omitted in 1984. The 1984 survey obtained 146 responses, for a response rate of 84 percent. The 1985 survey obtained 173 responses, for a response rate of 98 percent. The 1985 survey also obtained 143 responses from among the 146 cities responding in 1984, for a 98 percent response rate in that panel. The 1986 survey reached all 176 cities, for a 100 percent response rate.<sup>4</sup>

The same interview schedule was used in all three surveys, with minor modifications. The schedule began with an introduction that made no mention of the Minneapolis experiment but did mention the organizational affiliation of the interviewer (Police Foundation in 1984 and University of Maryland in 1985 and 1986). The first question asked the respondents to describe their department's current policy for dealing with minor domestic violence. Subsequent questions asked whether domestic violence arrests were increasing and tested respondent recognition of the research and accuracy of recall for the principal findings. Other items probed the impact of the study on department policy, and the interview closed by exploring the department's willingness to replicate the experiment.

Interviewers asked to speak with the head of the planning and research unit in the police department. Since not all cities of over 100,000 have such a unit, the interviewers often had to accept re-

---

<sup>3</sup> It would have been better, of course, to have completed the first wave of the survey of police practices before the preliminary 1983 report was released.

<sup>4</sup> Unfortunately, the raw data for part of the 1984 wave were lost before this article was prepared, and only 117 responses survived, for a 66% response rate. Since that is the maximum base we could use for a panel analysis of changes over time, the displayed data are limited to the 117 cities with a 100% response rate at all three waves. The aggregate trends, however, vary little between the panel and full sample data.



**Table 1.** Recognition of the Minneapolis Experiment, Domestic Violence Policies and Arrest Trends, and Respondent's Position in 117 Police Departments in Cities of over 100,000 (in percent) ( $N = 117$ )

Variable	1984	1985	1986
<b>Recognition</b>			
Yes	31	71	71
Maybe	8	1	1
No	59	26	28
Missing data	3	2	0
	<u>101</u>	<u>100</u>	<u>100</u>
<b>Current Policies</b>			
Arrest	10	30	43
Mediation	33	18	19
Sending suspect away	5	5	5
Officer's discretion	40	43	33
No policy	11	3	0
Missing data	0	2	0
	<u>99</u>	<u>101</u>	<u>100</u>
<b>Respondent's position</b>			
Planning director	33	11	27
Planning office	26	22	22
Chief's office	12	9	15
Training department	3	3	6
Operations	21	12	23
Major crimes	0	1	6
Missing data	4	42	1
	<u>99</u>	<u>100</u>	<u>100</u>

sponses from other spokespersons. Moreover, in 1985 the interviewers did not clearly identify the unit in which 42 percent of the respondents worked; Table 1 shows those respondent positions that were identified. We were able to track forty-seven (40 percent) of the respondents across the three years, even when they changed positions.<sup>5</sup>

**2. Results.** The reported use of arrest for minor domestic violence increased from 1984 to 1986. Recognition of the study and accuracy of recall were high in all three waves, at least by the standards of general public knowledge of news events. The study appeared to be considered in policy making in the first wave, and its apparent effect increased in the second and third waves. To es-

<sup>5</sup> When respondents differed from year to year, some inconsistencies appeared. For example, the 1984 Houston respondent said that the department's policy had been influenced by the Minneapolis experiment, but the 1985 Houston respondent said it had not. This discrepancy represents the larger problem of determining the policies and practices of police departments through telephone surveys.

**Table 2.** Accuracy of Recall, Influence of Research, and Willingness to Replicate of Respondents Who Recognized the Minneapolis Experiment (in percent)

Variable	1984 ( <i>N</i> = 45)	1985 ( <i>N</i> = 84)	1986 ( <i>N</i> = 84)
<b>Accuracy of recall</b>			
Arrest	49	73	80
Mediation	5	5	0
Separation	9	5	2
Don't know	38	18	18
	<u>101</u>	<u>101</u>	<u>100</u>
<b>Influence of Research</b>			
Yes	7	32	45
No	84	62	51
Don't know	5	6	4
Missing data	5	0	0
	<u>101</u>	<u>100</u>	<u>100</u>
<b>Willingness to replicate</b>			
Yes	27	31	NA
No	13	23	NA
Maybe	42	44	NA
Don't know	18	2	NA
	<u>100</u>	<u>100</u>	NA

estimate the change reliably, only results based on cities measured in all three waves are displayed.

**3. Recognition.** As Table 1 shows, almost one-third of the respondents reported recognition of the Minneapolis experiment after only the preliminary findings were released. (The question was, "Have you ever heard of the Minneapolis Domestic Violence Experiment announced in 1983: yes, no, or think so?") After both the second and third round of publicity, double that number reported hearing of the experiment.

**4. Accuracy.** As Table 2 shows, respondents on the second and third waves both reported and demonstrated greater familiarity with the experiment. If they claimed to have heard of the study, they were then asked, "Of the three methods of handling domestic violence—arrest, separation, and mediation—do you recall which one the Minneapolis experiment found to produce the least repeat violence: arrest, separation, mediation, or don't recall?" In 1984, almost half of the respondents who claimed recognition were able to identify the results of the experiment correctly; most of the others admitted that they could not recall. By the second wave of the survey, after the second round of publicity, 73 percent of those claiming to have heard of it correctly recalled



**Table 3.** Department Arrest Policy by Respondent's Awareness of the Experiment\* (in percent)

Policy	1984-1985		1985-1986	
	Unaware ( <i>N</i> = 22)	Aware ( <i>N</i> = 73)	Unaware ( <i>N</i> = 30)	Aware ( <i>N</i> = 82)
Change to arrest	23	26	13	33
No change or change to a nonarrest policy	77	74	87	67
	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>

\*  $X^2$  ( $df = 1$ ; corrected for continuity) = 3.29;  $p < .10$

that arrest was found to be the most effective treatment in the experiment. By 1986, accuracy of recall among those who claimed awareness had risen to 80 percent.

**5. Policies.** The best test of the influence of the research may be to look at the policies themselves. Respondents were asked, "In general, which is the preferred policy in your department for dealing with minor domestic assaults: arrest, mediation, sending the suspect away, letting the officer decide, or no policy?" In 1984, few police departments had a policy of arrest (Table 1). By 1986, nearly half of the departments reported arrest as the leading policy. A number of factors in addition to the Minneapolis experiment probably contributed to this rapid increase, including the recommendations of the Attorney General's Task Force on Family Violence (1984), the television movie "The Burning Bed," and a growth in the antifamily violence movement. Hence we used additional measures to assess the influence of the experiment itself.

**6. Influence.** To test the likely effect of the experiment, we asked respondents who said they had heard of the experiment whether the study had influenced policy in their departments. While a majority in all three waves said it had not, by the third wave nearly half claimed influence (Table 2). This finding is consistent with the evidence in Table 3 which compares the changes in reported policy in departments according to whether their spokespersons were aware of the experiment in 1984-85 and 1985-86. The 1984-85 changes show no difference in changes to an arrest policy between departments whose spokespersons claimed in both years to be aware of the experiment and those whose spokespersons did not. In 1985-86 there was twice as much change to arrest in the "aware" group as in the "unaware" group, although the difference was not significant.

The reported influence of the research was not always a

**Table 4.** Policy Changes Among Those Who Said the Experiment Found That Arrest Reduced Repeat Violence, by Accuracy of Findings (in percent)

Policy	Accurately Reported Results of Experiment			
	1984-85*		1985-86**	
	Yes	No	Yes	No
Change to arrest	34	5	27	8
No change or change to a nonarrest policy	66	95	56	92
	100	100	100	100

\*  $X^2$  ( $df = 1$ , corrected for continuity) = 5.18;  $p < .05$

\*\*  $X^2$  ( $df = 1$ , corrected for continuity) = 8.47;  $p < .01$

change to a policy of arrest. Twelve of the thirty-six departments in the full 1985 sample ( $N = 173$ , data not displayed) that reported such influence described it as "a greater emphasis on arrest when mediation fails" (even though the preferred policy was mediation) that took the place of either the previous policy of viewing arrest as a last resort, new training, or greater attention to the problem of domestic violence in general.

Recording mere awareness of the experiment, however, may not be sufficient to assess its impact. Theoretically, one would expect a change to a policy of arrest only when the department accurately perceives the experiment to recommend arrest. In such instances, as Table 4 demonstrates, the correlation between accurate knowledge and a change to an arrest policy is clear, with a 700 percent relative increase in the likelihood of change to arrest the first year and 300 percent the second.

Even this correlation may not be causal, of course. If a purely political decision is made to "get tough" with wife beaters, there may be a motivation to review the literature and justify the decision with the most relevant study. In this situation policy change would cause accuracy of recognition rather than vice versa.

It is therefore plausible to read these results as quite modest, depending on one's standard for a powerful effect of social research. One reviewer has suggested that "the Minneapolis experiment probably reached the high water mark of research impact: in view of the publicity that the research received and the climate in which it was released, one can probably expect that social research will seldom have as much impact as this experiment did." How much impact is revealed by the fact that forty-three police departments reported a change to an arrest policy when respondents accurately knew the experimental results whereas twelve reported such a change when respondents did not accurately know the results? As a percentage of the possible universe, the impact is

quite modest. Yet if one considers that many departments may, like Milwaukee, make an additional 5,000 or more arrests for domestic violence for years to come, the impact becomes quite substantial.

**7. Lawsuits.** Another reason for the increase in arrest policies may be the fear of lawsuits for failure to make arrests. A 1985 verdict for \$2.6 million (later settled for \$1.9 million) against a Connecticut police department (*Thurman v. Torrington, D., Conn.*) was widely publicized and may have prompted some departments to change their policies. Nonetheless, the 1986 survey showed more problems with traditional false arrest litigation for making domestic violence arrests than with lawsuits for failing to make them. Of all 176 departments, only 5 percent reported being sued for failure to make arrests. Fifteen percent of the respondents said their departments had been sued for false arrests for domestic violence.

**8. Arrest Trends.** Whatever their departments' policies, the majority of respondents in all three waves reported that the actual number of arrests for domestic violence had either increased or stayed about the same over the preceding year. The percentage reporting increased arrests over the prior year rose from one-fourth in 1984 to one-third in 1985 and to almost half in 1986 (Table 1).

**9. Willingness to Replicate.** Given Lempert's argument that replication should precede publicity, it is interesting to note the impact of publicity on the feasibility of replication. The willingness to replicate more than doubled from 1984 to 1985, from 12 departments to 26 in the panel of 117. While the proportion willing to replicate changed very little among those respondents who claimed to have heard of the study (from 27 percent to 31 percent), the number in the panel who had heard of the experiment had almost doubled (from 45 to 84; see Table 2).

It seems, then, that more publicity (or the passage of time or both) helps to make replication possible. Without any publicity, it may not have been possible to replicate the experiment in more than one or two cities. But when the National Institute of Justice solicited proposals for replication in early 1986, nineteen police departments participated in or filed their own replication proposals (a total of six replications were funded). It is not surprising that publicity should have this effect for this type of research design; when a research procedure is potentially controversial among practitioners and the public, publicity about its use in another city documents a precedent that provides a measure of protection for decision makers bold enough to replicate.

The number of departments willing to replicate before the second wave of publicity was initially reported as twenty-two

**Table 5.** Knowledge of the Experiment and Arrest Policy from Same and Different Respondents (in percent)

Respondent	1984		1985		1986	
	Same Respondent ( <i>N</i> = 47)	Different Respondent ( <i>N</i> = 70)	Same Respondent ( <i>N</i> = 47)	Different Respondent ( <i>N</i> = 70)	Same Respondent ( <i>N</i> = 47)	Different Respondent ( <i>N</i> = 70)
<b>Knowledge of experiment</b>						
Yes	21	17	70	40	72	47
No	79	83	30	60	28	53
	100	100	100	100	100	100
<b>Policy</b>						
Arrest	15	7	36	26	51	37
Other	85	93	64	74	49	63
	100	100	100	100	100	100

(Sherman with Hamilton, 1984), before some of the raw data on all 146 departments were lost. In fact, the chiefs of all twenty-two departments (and a few others) were invited to an NIJ-sponsored conference on possible replications in Minneapolis in the summer, of 1984. However, only about half of the chiefs left the conference willing to replicate. Those who were willing faced the issues of departmental management and commitment to conducting the replications properly. Thus a telephone indication that a respondent is willing to replicate should not be equated with a fully agreeable, feasible site for a replication.

**10. Testing Effect.** As in any panel, there is a danger of testing or “practice” effects. The potential influence of testing is shown in Table 5, which compares the responses for cities in which the same forty-seven respondents answered questions in all three years with those for cities in which the respondent changed. There is some evidence of testing effects: Respondents who appeared in all three waves more accurately reported the results of the experiment on waves two and three. Moreover, nearly all respondents (96 percent) at least claimed they had heard of the experiment when questioned in 1985 and 1986 (in contrast to 28 percent who reported hearing of it in 1984). These findings suggest that the interviews themselves may have at least supported efforts to publicize the experiment. A testing effect cannot, however, explain the entire increase. As Table 5 indicates, arrest policy grew in cities with both the same and different respondents across the three years.

**11. Summary.** This study of the impact of the Minneapolis experiment is quite limited. It has completely omitted the question of the experiment's influence on several state laws imposing mandatory arrest on local police departments (thereby making replication impossible in those states, contrary to the Sherman and Berk [1984b] recommendation), local mandatory arrest laws (Lempert, 1987), and other laws increasing police powers to arrest (Petersilia, 1987: 22–23). It has also omitted any close analysis of actual changes in police behavior (*Law Enforcement News*, 1987; *New York Times*, November 15, 1987: 65) or failures of implementation (Ferraro, 1985) after police departments adopt arrest policies. Most important, the longitudinal design of the case study cannot control for the influence of other factors on police policy besides the Minneapolis research.

Nonetheless, this study supports Lempert's basic premise. The publicity about the Minneapolis experiment did reach a large number of police departments and may have had a substantial—although far from universal—influence on policy. Thus it is important to ask whether the publicity and influence were premature.

## II. WAS THE PUBLICITY PREMATURE?

Lempert (1984) claims that the publicity was premature until the experiment had been replicated. This claim raises three questions about the Minneapolis experiment and its relationship to policy:

1. How severe were the threats to the internal and external validity of the experiment as published in 1984?
2. How does the quality of evidence in the experiment compare to the evidence supporting police practices prior to the publicity over the experiment?
3. What is the practice in medical research with respect to the approval of drugs or treatments, which Lempert argues should be the standard for social research on legal processes?

### A. *The Limitations of the Minneapolis Experiment*

The Minneapolis experiment found that misdemeanor offenders randomly assigned to arrest for domestic violence were significantly less likely to re-offend than those not randomly assigned to arrest—including some who were actually arrested in pre-approved exceptions to the randomization protocol (Sherman and Berk, 1984b; Berk and Sherman, 1988). The measurement of recidivism included both official records of 100 percent of the sample and at least one (and up to twelve) interviews with 65 percent of the victims.

As in any field experiment involving humans, a number of

factors could not be controlled. A review of the internal and external validity issues is appropriate to evaluate the policy implications the researchers drew.

### *B. Internal Validity*

**1. Randomization.** Despite substantial evidence that the officers in the Minneapolis experiment did not seriously compromise the randomization (Sherman and Berk, 1984b), it is at least possible that they decided to exclude certain nonarrest cases from the study after arriving at the scene. (Any unrandomized arrests were detected by independent monitoring by Police Foundation researchers.) The agreement with the Minneapolis police did not allow the experimental officers to receive all domestic calls, so it was very difficult for observers to ride with those officers and witness the rare case in which violence had occurred and the offender was still present.<sup>6</sup>

**2. Differential Victim Reporting by Treatment.** One rival hypothesis for the lower rate of repeat violence among arrestees is that they “intimidated” their victims into neither calling the police during future violence nor telling the interviewers about it in their biweekly interviews. Sherman and Berk (1984b: 269) concluded that the intimidation hypothesis was unlikely for two reasons. First, in 55 percent of all cases randomized in the experiment, someone other than the victim called the police. It seems unlikely that the offender managed to intimidate all who might call the police.<sup>7</sup> Second, the victims’ initial response rates to the interviewers did not vary according to the police action against their offenders. This makes *differential* intimidation unlikely, since it is more plausible that an intimidated woman would not agree to or keep an appointment for an interview (conducted at her home or any other location she designated in advance) than that she would keep the appointment and lie. That is, intimidators would probably try to discourage the contact entirely rather than to shape the

<sup>6</sup> The Milwaukee replication of the experiment solved this problem by giving most domestic calls to “domestic cars” staffed by volunteer officers selected as elite “clinical investigators” who could be efficiently observed. The Milwaukee experiment reduced treatment failures to under 2% of its 1,200 cases (from 18% of the 314 cases in Minneapolis) by screening cases in the field prior to randomization and then calling the research office, where a sealed envelope containing a pre-randomized treatment was opened. See Sherman and Berk (1984b) and Berk and Sherman (1988) for a discussion of the analytic procedures used to correct the 18% post-randomization treatment failure rate.

<sup>7</sup> Unfortunately, Sherman and Berk (1984b) were unable to collect data on who called the police in the recidivism cases as distinct from the randomized entry cases. A reduction in the percentage of cases in which victims called police could show victim intimidation. But it would have to have been a 100% elimination of victim calls to account for the overall 50% lower magnitude of recidivism among the arrest cases compared to nonarrest. This is possible but seems unlikely.

content of the contact. Nonetheless, we cannot discount intimidation entirely as a rival hypothesis.

**3. Sample Size.** The sample size and base rate of repeat violence were both large enough to detect main effects of the alternative treatments. Neither the sample size nor the distributions of subpopulations were adequate, however, for a thorough testing for interaction effects. It is entirely possible that arrest may backfire for some types of offenders, increasing their propensity to violence against the same or other potential victims. As Sherman (1984) points out, advances in personal computers would make it easy for police to process this kind of detailed information in determining what action to take. But larger samples are needed to produce reliable predictions on which such software would be based.

**4. Analysis.** The Sherman and Berk (1984b) analysis was limited to two methods: (1) analysis of the prevalence (or percentage) of offenders who had at least one repeat incident, and (2) analysis of the time to "failure," or the length of time between police intervention and the first repeat incident. This analysis omits two other policy-relevant questions: (1) What was the difference across treatments in the *frequency* of offending, or the average number of repeat offenses per offender over the follow-up period (cf. Blumstein and Cohen, 1979; Blumstein *et al.*, 1986)?, and (2) What was the difference in the *seriousness* of repeat offending, measured by injury and hospitalization? If total incidents or average seriousness was lower for any of the nonarrest treatments than for the arrest group, that would have seriously complicated the policy implications. But the Sherman and Berk (1984b) analysis did not address those issues.

**5. Follow-Up Period.** Although there is no evidence that a longer follow-up period would alter the results, it is possible that the longer-term results are different. In a reanalysis of the interview data, Witte and her colleagues (personal communication, 1986) have found that most of the deterrent effect disappears by the end of the standard six-month follow-up period. If this trend were to continue, arrest could actually *enhance* overall violence beyond six months. If, on the other hand, the effect just disappeared, the findings would still support more arrests, even if only for a six-month reduction in violence. Berk and Sherman (1988), however, find a continuing deterrent effect with official data beyond six months.

**6. Displacement.** As Reiss (1985: 173) has pointed out, it is possible that arrest of an offender merely displaces the offender's violence on to another victim so that no net reduction in domestic violence is achieved. Alternatively, victims may enter into new re-



relationships where they become abused. While the Minneapolis data showed no difference across treatment groups in *rate* of relationships breaking up, the timing was not clear; moreover, the response rate was too low to rule out the Reiss hypothesis. Testing this theory requires tracking both victims and offenders separately, and not just their relationship, with both official records and interviews. The Milwaukee replication has gathered such official data.

### C. *External Validity*

More troubling than the possible threats to the internal validity of the results were the clearly established threats to external validity. These were also the principal concerns of Lempert (1984) and others who ask whether policy decisions in other cities should be made on the basis of the Minneapolis results.

1. **Jail Time.** As Sherman and Berk (1984a) point out, Minneapolis may be unusual in jailing suspects arrested for domestic assault for at least one night. Thus, the treatment actually tested in Minneapolis was arrest plus immediate, but brief, jail time, rather than arrest and immediate release. The effects of arrest in cities practicing immediate release might be very different, including the offender's possible return to the victim while he is still in a drunken rage, whereupon he might inflict even more serious damage.<sup>8</sup>

2. **Mediation Quality.** It has been argued that Minneapolis was not a fair test of the effects of mediation, since the participating police officers did not have special training in family crisis intervention (Reiss, 1984: 106). All had the standard, rather minimal, patrol officer's training in those skills. The skills of a highly trained special unit devoting most of its time to domestic violence might be much more effective and thus might produce better results than arrest. The Minneapolis findings may not generalize to cities where such units are in operation; they should, however, generalize to the larger number of cities without such units.

3. **Interaction of Interviews and Arrest.** Police Foundation interviewers made extensive efforts to contact all victims, which included up to twenty phone calls and visits and lasted up to six months. These efforts cannot account for any differences between treatments, since they were apparently equal across the treatments, as were the response rates (Sherman and Berk, 1984b: 269). They do raise questions about how effective arrest would be in the absence of this form of "surveillance," which may have had some

---

<sup>8</sup> The Milwaukee replication randomized the amount of time in jail in order to compare the effects of 2 hours and 8 hours or more.

deterrent effect on the offenders. As Sherman and Berk (1984a) concede, it is an open question whether the same effects would be found without the interviews.

**4. Absence of Theory: The Black Box Problem.** As Lempert (1984: 509) suggests and Berk and Newton (1985: 262) emphasize, the theoretical basis for the observed deterrent effects of arrest is poorly understood. In the language of Cook and Campbell (1979), Sherman and Berk report only on the “molar” relationships, not on their “micro-mediation,” or the links in the causal chain between police action (e.g., victim empowerment through alliances with the police) and the likelihood of further violence. Documenting such links would probably require interviews with offenders that were guided by some model of causation to focus the questions.

**5. Victim Perception of Officer.** One example of such possible micro-mediation is the victim empowerment argument, in which the victim perceives an alliance with police against the suspect. The deterrence effect appeared to be enhanced when victims thought that officers took the time to listen to their side of the story (Sherman and Berk, 1984a; Sherman and Bouza, 1984). This finding was based on a multiple regression analysis in which the victim interview response on the listening item was entered as a control variable. After further consultation with statisticians, Berk (personal communication, 1984) concluded that the borderline statistical significance of the finding was not sufficient. If that judgment was correct, the manner in which police officers treat victims in domestic violence situations should not affect the external validity of the results.

**6. City Context.** The setting of the experiment itself presents a major external validity issue. Minneapolis is not America, nor is any city truly comparable to any other. Variations in weather, ethnic composition, age structure, prevailing crime rates, and general sanctioning levels (Sampson, 1986) might all affect the reactions of offenders to alternative police actions for domestic violence. Minneapolis is an extreme case on many of these dimensions. For at least six months a year, its weather prevents men ordered out of the house from spending the night out of doors. It has one of the smallest minority populations of any major city, one of the largest Native American populations, and probably the largest proportion of white Protestants. Both its homicide rate and its imprisonment rate are among the lowest in the country. Whether arrest would work as well in Miami, San Diego, or Pittsburgh is an open question. As Lempert recommends, the only way to find an answer is to replicate the experiment in as many cities as possible, cities that represent different points on these many contextual dimensions.

**7. Alternative Procedures.** The Minneapolis conclusion that arrest is the most effective response to domestic violence clearly cannot be generalized to comparisons with other procedures not included in the research design. Different findings might result, for example, by extending the design to include offenders who had already left the scene by randomly picking some to be tracked down; by randomly assigning victim “empowerment” within each treatment, with officers either paying great or little attention to the victims; by randomly altering the aftermath of the intervention, with police making follow-up visits to some offenders and not others or with some arrested offenders being diverted to mandatory counseling. All of these complexities are possible and commonplace in many cities and could fruitfully be examined in future replications. Moreover, the external validity of the “arrest-works-best” conclusion is limited to the two other comparison treatments only.

**8. Summary.** The list of possible threats to the internal and external validity of the “arrest-works-best” finding is clearly quite extensive. But this can probably be said for any single piece of research. As Lempert points out, it is not the conduct of the experiment that is at issue but rather the fact that it was a single experiment. In comparison to most other policy studies and even to other randomized experiments, the Minneapolis experiment actually suffered quite minor threats to validity (see, e.g., Pocock *et al.*, 1987). The question Lempert raises is really, *How much evidence is enough to change policy or practice?*

To answer this question for police policy on domestic violence, the appropriate test is a comparison of the evidentiary strength of the recommendations derived from the Minneapolis experiment with the strength of the evidence in support of the pre-experiment status quo.

#### *D. The Existing Knowledge Base for Practice*

When the Minneapolis findings were published, the knowledge base for police practice was virtually nonexistent. Most police officers were guided by their experience in handling such cases, which gave them an intuitive judgment about what works best. As Sherman (1984: 62) points out, this kind of experience suffers from highly selective feedback. Unlike craftsmen, who can see the results of their efforts, big city police may never see the “product” of their work (defined as the subsequent rates of crime after police intervention). If there is a recurrence while the officer is on the same shift and the same beat, the feedback may occur. Otherwise, feedback is rare.

This point is important because of the hyperscientific tendency to criticize any and all experiential learning as a basis for

policy decisions (see Lindblom and Cohen, 1979). We do not dismiss all experience as a knowledge base for police decision making. But given the flawed and erratic system of feedback to officers for their handling of minor domestic violence, this kind of experiential learning must be discounted as a reliable knowledge base for making policy decisions.

The research literature available in 1983 was even less help as a guide to practice. For all of the ink expended in describing, explaining, or criticizing police practices in minor domestic violence (e.g., Parnas, 1972; Potter, 1979; Black, 1980), we found only one study that claimed to provide any systematic empirical evidence evaluating the effectiveness of any alternatives for reducing subsequent violence.<sup>9</sup>

When the Minneapolis experiment was undertaken, then, the existing level of knowledge about the consequences of alternative police actions was not only low but also misunderstood. There were no prospective, nonrandomized follow-ups of offenders or households treated with different methods. Virtually no one had reported any data on what happened after police left the scene (but see Meyer and Lorimer, 1977), regardless of the approach police employed.

Consequently, there was no reliable basis for choosing any policy. The Minneapolis experiment, with all its limitations, represented a substantial advance over the existing knowledge. Had similar evidence become available about a medical treatment, with its stronger assumptions about external validity, it would not have been subjected to further testing before approval, as Lempert (1984) suggests. Rather, any doctor who failed to use a treatment proven effective with this level of evidence could have been sued for malpractice.

### *E. The Medical Model of Research and Practice*

Lempert's comparison of arrest policies for domestic violence to a new cancer drug (1984: 510; 1987: B15) breaks down on several points. First, contrary to his assertion, there is no requirement or standard practice for replicating randomized clinical trials. Second, arrest was not a new "drug" in the sense of no prior general use; police had been making arrests for domestic violence for centuries, since legislatures had passed laws against assault. Third, the theoretical rigor in medicine is often no higher than it is in the sociology of law. Fourth, Lempert's premise that the delayed implementation of research is more cautious may apply to construction projects but not to ongoing life-and-death human problems that cannot be delayed pending further research.

---

<sup>9</sup> This study (Bard, 1970) used a weak, after-only comparison design (Cook and Campbell, 1979) and was arguably interpreted incorrectly (Liebman and Schwartz, 1973).

**1. Replications in Medicine.** Proposals for new drugs in medical practice are heavily scrutinized and regulated, but this regulation is a relatively recent development. Before World War II, drugs could be marketed freely without testing (Pocock, 1983: 26). Even after the war, tests were largely limited to toxicity levels. In the wake of the thalidomide and DES disasters, however, the Food and Drug Administration (FDA) raised the level of evidence required to prove the safety and effectiveness of new drugs, but it was not until 1969 that the FDA required evidence from randomized clinical trials to obtain marketing approval for new drugs (ibid.).

The *Bureau of Drugs Clinical Guidelines* (U.S. FDA, 1977) established four phases of human drug testing following positive results among animals (ibid., pp. 2–4). Phase I uses healthy volunteers to measure the toxicity, or unhealthy side effects, of the drug. Phase II tests the drug's effectiveness on a small number of patients, who are monitored very closely. Phase III is a randomized controlled trial (RCT) on a large number of patients. If this *single* trial is successful, the drug can be approved. Phase IV is postmarketing surveillance, including long-term, large-scale follow-up of morbidity and mortality (such as the heart attack rate associated with the birth control pill).

In theory, additional trials could be conducted after approval is obtained, but this seems to be quite rare. To the contrary, many doctors would find it unethical to withhold a treatment already found effective in one RCT. As a medical journalist points out, “in principle, randomized trials are performed only when there is no clear evidence as to which therapy is best” (Keller, 1985: 22). It is not at all clear, despite the limitations of the Minneapolis experiment discussed below, that physicians would find it ethical even to conduct replications, let alone to withhold treatment until the replications are completed. In fact, in the early stages of the Minneapolis replications, several victim's rights groups opposed the replications for precisely this reason.

The standard medical research goal seems to be to conduct one large RCT that “settles” a question, or at least a narrowly defined segment of it. Problems may develop that limit the RCT's conclusiveness, but the ideal is *not* to have to replicate. Consider, for example, the RCT of the vaccine for hepatitis B, described as the “finest clinical trial in the history of medicine” (Goodfield, 1984: 48). The fifteen-year time span between the discovery of an agent and the licensing of the vaccine was reportedly the shortest ever. The RCT was preceded by tests on chimpanzees and on 200 human volunteers from the Merck Company (which developed and manufactured the vaccine), none of whom developed any adverse effects. The investigators recruited 1,083 high-risk volunteers from the homosexual community in Greenwich Village, 96 percent of whom returned for the second injection (of either vac-

cine or a placebo drug) and 85 percent of whom stayed in touch with the study over the follow-up period. Apparently basing the decision on the finding that the vaccine was 81 percent effective in preventing infection, the FDA licensed the vaccine for general use.

More recently, the controversy over the clinical trial of the experimental AIDS drug AZT demonstrates the enormous pressure on medicine to rush treatments into practice without ideally thorough testing. Doctors treating AIDS patients actually lobbied the Congress to force the National Institutes of Health (NIH) to abandon the 282-patient RCT before it was completed, since an uncontrolled Phase II study of 15 patients had found some life-prolonging effects. They argued that the FDA should approve the drug for more general use based on the initial research. Since then, the *Wall Street Journal* has regularly attacked the alliance of regulators and researchers for playing God in withholding new treatments from general use (e.g., see December 7, 1988: A14). Both the controversy and the experiment were cut short when 19 placebo patients and 1 AZT patient had died by the twenty-fourth week of the study (Fischl *et al.*, 1987), and the FDA released the drug for general use (*Washington Post*, September 20, 1986: A1). It is clear there will never be a replication of the trial.

Nor will there be a replication of the aborted 1985 clinical trial of the heart attack victim's lifesaving drug, tissue plasminogen activator (TPA), which the National Heart, Blood and Lung Institute of the NIH stopped because "investigators felt that TPA was so superior to an alternative drug, streptokinase, that they could not ethically withhold TPA from patients in the trial who were not receiving it" (Boffey, 1987: A17). This conclusion was reached despite an initial finding that the drug caused severe brain damage in 2 percent of the cases, which led the FDA to withhold approval until late 1987, when further Phase II studies showed the brain damage rate to be 0.4 percent.

**2. Theoretical Rigor in Medicine.** Lempert might favor adopting medical treatments with limited controlled experimentation because the theoretical basis for predicting the effectiveness of a treatment is presumably stronger than comparable theoretical developments in social science. As Lempert (1984: 509) argues, "We should remember that the key to generalizing in science is theory. . . . If we simply assume that what has occurred in one setting will occur in another, our generalizations will rest on shaky ground whenever the settings differ in important particulars. Making policy on the basis of a single study is always dangerous in part because one study is almost never sufficient to develop a reliable theory."

But precisely the same point is debated in medical research. Dr. Emil Freirich, a professor of medicine who claims that randomized controlled trials are widely overused in medical research,



argues that because the biological heterogeneity of human populations is “enormous” (quoted in Keller, 1985), it may well be inappropriate to generalize to populations beyond those selected for any medical study. Despite the assumption by social scientists that human bodies are less variable in their biochemistry than cities are in their criminology, one could argue that a vaccine that “works” on a homosexual population in New York might not work in other settings where diet, water content, sanitation, the epidemiology of the disease, and other factors are different. Yet, the hepatitis B vaccine was approved for national use after a single RCT on this unique population.

Similarly, it is not clear that cities vary more than bodies. When the Milwaukee replication was proposed to the city council, one feminist organization took the position that cities do not vary. This group lobbied against approval of the replication on the grounds that the effectiveness of arrest had already been proven in Minneapolis and that the creation of a control group within the city’s mandatory arrest policy was an unethical withholding of treatment from victims at risk of further violence. Sherman responded that cities were probably too varied to make that assumption and that the replication would better serve Milwaukee victims in the long run by providing local officials with information specific to their own city.

The important point is that medical research and policy practices are not irrelevant to legal research merely because bodies are assumed to be more consistent than cities. Moreover, many medical treatments, such as antibiotics, are used because they work, without a full or even rudimentary theoretical understanding about why they work.

**3. Testing New Versus Existing Treatments.** RCTs are not used to test all new or existing treatments in medicine (Hiatt, 1986), although some doctors advocate the RCT as the standard evaluation procedure. Many medical treatments grow “like Topsy” and are only later (if ever) subjected to tests to settle debates—much like criminal justice treatments and sanctions. A good example is the medical response to breast cancer that is detected early, when the tumor is still confined to the breast, which encompasses some 60 percent of all breast cancer patients. In the early 1960s, when many such patients refused to suffer the disfigurement of a full mastectomy, Harvard physicians developed the now famous lumpectomy, cutting out the tumor but leaving the breast intact and following the surgery with radiation. Prospective follow-up of 357 women treated from 1968 to 1978 showed only 18 cases of recurrence. This evidence (as well as even less evidence at earlier stages) was sufficient for many physicians to recommend that their patients undergo lumpectomies, while others remained adamantly opposed to anything less than radical mastectomy. Ultimately, an



RCT found no difference in five-year survival and recurrence rates for such patients (Fisher *et al.*, 1985).

The lumpectomy/mastectomy RCT also illustrates the division in medicine over the issue Lempert raises about when to publicize. The results were submitted for publication almost a year before they appeared in the *New England Journal of Medicine* (*ibid.*). Although the reason for the delay was not made public, it was widely speculated that, upon peer reviewer recommendations, the *Journal* decided to ask for a longer follow-up period to increase the statistical power of the data (Bishop, 1984). The delay upset many physicians who wanted the formal guidance of published results to help them decide which treatment to recommend. The delay was all the more upsetting because of the widespread rumors that oral presentations of the findings at medical conferences had reported no difference in treatment outcomes. This example illustrates the complexity of the timing issue that Lempert raises.

Delaying the announcement of RCT results until the evidence is stronger is not necessarily a more cautious approach. The physicians who had to operate on over 100,000 breast cancer patients during the year delay in publication may, as it turns out, have unnecessarily disfigured many of those patients. On the other hand, had they prematurely recommended lumpectomies and had later evidence showed higher rates of cancer recurrence, they would have increased the risk of death. These problems of Type I and Type II error have no obvious solution but require the kind of judgment that will always be debatable.

**4. The Costs of Delay.** There are costs and controversies in delayed as well as premature publicity. Medical practices, like crime control practices, deal with ongoing human problems. Unlike decisions to build a dam or a highway, for example, decisions about such problems cannot be postponed until a careful study of environmental impact and other issues has been completed. The cost of delay for such construction projects is usually just the delay of a new benefit, rather than the positive imposition of harm or ineffectiveness at saving lives. In attacking the FDA's two-year delay in approving the heart attack drug TPA, for example, the *Wall Street Journal* accused the FDA of having "sacrificed thousands of American lives on the altar of pedantry" (Specter, 1987: A16).

Moreover, in both medical and crime control practice, policies can be changed much more quickly than a dam or highway can be torn down. Publication practices in medical research assume that physicians will continue to follow new research, that science is constantly evolving, and that practice should continue to change with new knowledge. The question in medicine is never, "What is the final and ultimate truth?" but rather, "What is the best available knowledge at this point in time?"

To be sure, misdemeanor assaults produce a lower risk of

death than AIDS or breast cancer, so the costs of delay may be relatively smaller in relation to the potential costs of premature change in practices. Moreover, the criminal justice community may not be as diligent as the medical community in keeping up with current research. But recent discussions with both doctors and police administrators suggest that the differences are not great and that doctors may have been slower to respond to the lumpectomy RCT than police have been to the Minneapolis experiment.

**5. Consensus from Conflicting Results.** Medical research, like social science research, is often plagued with conflicting evidence. One solution medicine has adopted may be applicable to social science as well. Since 1977, the National Institutes of Health have sponsored a "Consensus Development Program" (Perry, 1987). The program convenes a distinguished group of experts without conflicts of interest to review the evidence on controversial questions of medical practice and produce recommendations for doctors. Over sixty topics have been considered, including breast cancer screening, Caesarean childbirth, and liver transplantation. At least four other countries have adopted similar programs.

Perhaps legal practices could benefit from similar consensus development panels, especially where legislatures have allowed officials to exercise discretion. But even considering such an approach opens the broader questions of how to regulate legal research and practice that Lempert raises.

### III. CONCLUSION: RESEARCH AND REGULATION

This article has considered whether the Minneapolis Domestic Violence Experiment was "unduly and prematurely publicized," how much policy influence the experiment had, and whether that influence was inappropriate. The evidence suggests that the researchers tried quite hard to publicize the project and apparently produced some impact on policy. The data also suggest that the policy influence was not premature or inappropriate, at least not by medical standards. Legal researchers seem fully justified in releasing and publicizing research results as soon as they have been favorably reviewed.

Both legal practice and medical practice can benefit from research on two issues: new, unused technologies (such as the Greenwood scale or electronic monitors for house arrest) and technologies long in use but never tested (such as preventive patrol or arrests for domestic violence). One might argue that arrests for domestic violence were only used in extreme cases and that widespread use is more a new invention than an old technology. But Minneapolis officers reported strong *individual* variation in practices, with some of them always arresting on probable cause and

others never doing so. Under these circumstances, at least in Minneapolis, the experiment was testing an old, unevaluated practice.

Such practices can be more or less effective, and research on them can be more or less reliable. The question Lempert raises is when, if ever, the process of decisions to change policies should be regulated, which seems implicit in his question about *how much* publicity a peer-reviewed study should receive in any given case.

By criticizing the publicity about and the influence of the Minneapolis experiment, Lempert may imply that *researchers* should either be regulated or regulate themselves to prevent policy from being changed on insufficient research grounds. This clearly has no basis in the medical model, since there is no restraint (other than normal peer review) on publication or other efforts to seek publicity for medical research. Rather, strict application of the medical model would imply some regulation of *practitioners* based on a standard of research.

Civil litigation, especially against police, has increasingly employed research to regulate legal practice, just as it regulates medical malpractice. But there is no FDA or NIH consensus program for new or existing legal practices. Perhaps such structural changes are warranted. But they seem most likely to happen, as they did in medicine, after sufficient development and funding for the research enterprise makes the need for them obvious.

A related issue is the kind of publicity a study receives. Mass media often fail to convey the full complexity of a study's findings, which readers need to interpret the results properly. But the media are the only form of publicity that can be sure to reach top policy makers, since it is the media that shapes the editorial and political pressures to which they must respond. To advocate publicity solely through professional channels may be to advocate burying research results so that they can have little useful effect on either current practice or the conduct of replications.

As of 1988, the Minneapolis Domestic Violence Experiment is being replicated in six cities under funds from the National Institute of Justice. Neither these federal funds nor the willingness of these six police departments to randomize arrest would have been as likely had there not been so much mass media publicity about the original experiment (Joel Garner, personal communication, 1986). The NIJ staff was interested in replicating the experiment even before the second wave of publicity, but the initial funding commitment was not made until after this wave. Even then the planned funding was too low to provide an adequate measure of external validity—that is, a multi-site test. Only the continuing press attention, including editorials more than a year after the second wave, supported what has become the first large-scale, multi-site replication of a controlled experiment in the NIJ's history. The publicity supported the scientific argument for replication with an institutional argument for the funding agency, tempered

by the risk that replications might contradict the earlier "good news" and make the NIJ look foolish.

Lempert is right to criticize single studies as the final word or proof of a panacea in either legal or medical research. The practical question is how to get around the strong tradition of single-site, unreplicated studies.

For those who would use research for improving legal effectiveness—hardly a universal goal among social scientists of law (Black, 1972)—the Minneapolis experiment suggests a broader lesson: Extensive publicity about new research can speed up the process by which practice can be more reliably tested in multiple sites. Since no study is final, publicity about each new study can focus attention and funding on further research. Should further studies reach different conclusions, publicity about them can influence policies to change yet again. If legal policies are to be based on science, they must be able to change along with the constant evolution of scientific knowledge. Publicity can hasten this evolution.

## REFERENCES

- ATTORNEY GENERAL'S TASK FORCE ON FAMILY VIOLENCE (1984) *Report*. Washington, DC: U.S. Department of Justice.
- BARD, Morton (1970) *Training Police as Specialists in Family Crisis Intervention*. Washington, DC: U.S. Department of Justice.
- BERK, Richard A., and Phyllis NEWTON (1985) "Does Arrest Really Deter Wife Battery? An Effort to Replicate the Findings of the Minneapolis Spouse Abuse Experiment," 50 *American Sociological Review* 253.
- BERK, Richard A., and Lawrence W. SHERMAN (1988) "Police Responses to Family Violence Incidents: An Analysis of an Experimental Design with Incomplete Randomization," 83 *Journal of the American Statistical Association* 70.
- BISHOP, Jerry E. (1984) "Report Validating Lump Removal in Breast Cancer Awaits Publication," *Wall Street Journal* (December 4).
- BLACK, Donald (1980) *The Manners and Customs of the Police*. New York: Academic.
- (1972) "The Boundaries of Legal Sociology," 81 *Yale Law Journal* 1086.
- BLUMSTEIN, Alfred, and Jacqueline S. COHEN (1979) "Estimation of Individual Crime Rates from Arrest Records," 70 *Journal of Criminal Law and Criminology* 561.
- BLUMSTEIN, Alfred, Jacqueline S. COHEN, Jeffrey A. ROTH, and Christy VISHNER (1986) *Criminal Careers and "Career Criminals"*, Vol. 1. Washington DC: National Academy of Sciences.
- BOFFEY, Philip M. (1987) "U.S. to Back Gene-Engineered Drug for Clotting," *New York Times* (November 13) A1.
- (1983) "Domestic Violence: Study Favors Arrest," *New York Times* (April 5) C1.
- COHN, Ellen G., and Lawrence W. SHERMAN (1987) *Police Policy on Domestic Violence, 1986: A National Survey*. Crime Control Reports, No. 5. Washington, DC: Crime Control Institute.
- COOK, Thomas D., and Donald T. CAMPBELL (1979) *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand-McNally.
- EVANS, Sandra (1988) "Study Shows Negative Effects of Full-Time Child Care," *Washington Post* (April 23) A10.
- FERRARO, Kathy (1985) Round Table presentation, American Sociological Association, Washington, DC (August).

- FIENBERG, Stephen E., Kinley LARNTZ, and Albert J. REISS, Jr. (1976) "Redesigning the Kansas City Preventive Patrol Experiment," 3 *Evaluation* 124.
- FISCHL, Margaret A., et al. (1987) "The Efficacy of Azidothymidine (AZT) in the Treatment of Patients with AIDS and AIDS-Related Complex," 317 *New England Journal of Medicine* 185.
- FISHER, B. et al. (1985) "Five Year Results of a Randomized Trial Comparing Total Mastectomy and Segmental Mastectomy with or without Radiation in the Treatment of Breast Cancer," 312 *New England Journal of Medicine* 665.
- FOOTNOTES (1987) "Sociologist-Journalist Sought for Clearinghouse," *Footnotes: American Sociological Association* (November) 2.
- GANS, Herbert J. (1979) *Deciding What's News*. New York: Pantheon.
- GOODFIELD, June (1984) "Vaccine on Trial," *Science* 84 (March) 78.
- GOODMAN, Ellen (1983) "Using 'Muscle' Against Wife Beaters," *Washington Post* (April 19), p. A21.
- GREENWOOD, Peter (1982) *Selective Incapacitation*. Santa Monica: The Rand Corporation.
- HIATT, Howard (1986) "Will Your Next Hospital Stay Be Necessary?," *Wall Street Journal* (November 18) 32.
- KELLER, Allen S. (1985) "Treatment by Chance: Some Doctors Raise Ethical Questions About Randomized Studies," *Washington Post* (June 26), p. H22.
- KELLING, George, Tony PATE, Duane DIECKMAN, and Charles BROWN (1974) *The Kansas City Preventive Patrol Experiment*. Washington, DC: Police Foundation.
- LAMB, Rosemarie Wittman (1983) "Wife Beaters to Be Arrested" *Times of London*, April 18, p. 10.
- LARSON, Richard C. (1975) "What Happened to Patrol Operations in Kansas City?," 3 *Journal of Criminal Justice* 267.
- LAW ENFORCEMENT NEWS (1987) "Dallas Spouse-Abuse Arrests Soar," *Law Enforcement News* (October 27) 3.
- LEMPERT, Richard O. (1987) "Spouse Abuse: Ann Arbor Rushed into Arrest Ordinance Without Studying Side Effects," *Ann Arbor News* (June 21).
- (1984) "From the Editor," 18 *Law & Society Review* 505.
- LIEBMAN, Donald A., and Jeffery SCHWARTZ (1973) "Police Programs in Domestic Crisis Intervention: A Review," in J. R. Snibbe and H. M. Snibbe (eds.), *The Urban Policeman in Transition*. Springfield, IL: Charles C. Thomas.
- LINDBLOM, Charles, and David COHEN (1979) *Usable Knowledge: Social Science and Social Problem-Solving*. New Haven: Yale University Press.
- LIPTON, Douglas, Robert MARTINSON, and Judith WILKS (1975) *The Effectiveness of Correctional Treatment*. New York: Praeger.
- MARTINSON, Robert (1974) "What Works?—Questions and Answers About Prison Reform," 35 *The Public Interest* 22.
- MEYER, Jeanie Kenney, and Theodore P. LORIMER (1977) "Police Intervention Data and Domestic Violence: Exploratory Development and Validation of Prediction Models." Report to National Institute of Mental Health from the Kansas City, Missouri, Police Department.
- PARNAS, Raymond I. (1972) "The Police Response to the Domestic Disturbance," in L. Radzinowicz and M. E. Wolfgang (eds.), *The Criminal in the Arms of the Law*. New York: Basic Books.
- PERRY, Seymour (1987) "The NIH Consensus Development Program: A Decade Later," 317 *New England Journal of Medicine* 485.
- PETERSILIA, Joan (1987) *The Influence of Criminal Justice Research*. Santa Monica: The Rand Corporation.
- POCOCK, Stuart J. (1983) *Clinical Trials: A Practical Approach*. London: John Wiley.
- POCOCK, Stuart J., Michael D. HUGHES, and Robert J. LEE (1987) "Statistical Problems in the Reporting of Clinical Trials: A Survey of Three Medical Journals," 317 *New England Journal of Medicine* 426.
- POTTER, Jane (1979) "The Police and the Battered Wife: The Search for Understanding," 1 *Police Magazine* 40.
- REISS, Albert J., Jr. (1985) "Some Failures in Designing Data Collection That

- Distort Results," in L. Burstein, H. E. Freeman, and P. H. Rossi (eds), *Collecting Evaluation Data: Problems and Solutions*. Beverly Hills: Sage Publications.
- (1984) "Consequences of Compliance and Deterrence Models of Law Enforcement for the Exercise of Police Discretion," 47 *Law and Contemporary Problems* 83.
- ROSHCO, Bernard (1975) *Newsmaking*. Chicago: University of Chicago Press.
- SAMPSON, Robert (1986) "Crime in Cities: The Effects of Formal and Informal Social Control," in A. J. Reiss, Jr., and M. Tonry (eds.) *Communities and Crime: Crime and Justice. A Review of Research*, Vol. 8. Chicago: University of Chicago Press.
- SECHREST, Lee, Susan O. WHITE, and Elizabeth BROWN (eds.) (1979) *The Rehabilitation of Criminal Offenders: Problems and Prospects*. Washington, DC: National Academy of Sciences.
- SHERMAN, Lawrence W. (1986a) "Policing Communities: What Works?," in A. J. Reiss, Jr., and M. Tonry (eds.), *Communities and Crime: Crime and Justice, A Review of Research*, Vol. 8. Chicago: University of Chicago Press.
- (1986b) "Publicizing Research Results" Presented at the Panel on Research and the Press, American Society of Criminology, Atlanta (November).
- (1984) "Experiments in Police Discretion: Scientific Boon or Dangerous Knowledge?," 47 *Law and Contemporary Problems* 61.
- (1980) "The Specific Deterrent Effects of Arrests for Spouse Assault: A Field Experiment." Proposal submitted to the National Institute of Justice.
- SHERMAN, Lawrence W., and Richard A. BERK (1984a) *The Minneapolis Domestic Violence Experiment*. Washington, DC: Police Foundation.
- (1984b) "The Specific Deterrent Effects of Arrest for Domestic Assault," 49 *American Sociological Review* 261.
- SHERMAN, Lawrence W., and Anthony V. BOUZA (1984) "The Need to Police Domestic Violence," *Wall Street Journal* (May 22) 28.
- SHERMAN, Lawrence W., and Ellen G. COHN (1986) *Police Policy on Domestic Violence, 1985: A National Survey*. Crime Control Reports, No. 1. Washington, DC: Crime Control Institute.
- SHERMAN, Lawrence W., with Edwin E. HAMILTON (1984) "The Impact of the Minneapolis Domestic Violence Experiment: Wave 1 Findings." Washington, DC: Police Foundation, unpublished manuscript.
- SIGAL, Leon V. (1973) *Reporters and Officials: The Organization and Politics of Newsmaking*. Lexington, MA: D. C. Heath.
- SPECTER, Michael (1987) "FDA to Approve New Drug for Heart Attacks: Experts Say Anticlotting Medication May Save Thousands of Lives," *Washington Post* (November 13) A1.
- U.S. FOOD AND DRUG ADMINISTRATION (1977) *Bureau of Drugs Clinical Guidelines*. Rockville, MD: The Food and Drug Administration.