

2002

The Preventive Effects of Arrest on Intimate Partner Violence: Research, Policy and Theory

Christopher D. Maxwell
Michigan State University

Joel H. Garner
Joint Centers for Justice Studies, Inc.

Jeffrey A. Fagan
Columbia Law School, jfagan@law.columbia.edu

Follow this and additional works at: https://scholarship.law.columbia.edu/faculty_scholarship



Part of the [Criminal Law Commons](#), and the [Law Enforcement and Corrections Commons](#)

Recommended Citation

Christopher D. Maxwell, Joel H. Garner & Jeffrey A. Fagan, *The Preventive Effects of Arrest on Intimate Partner Violence: Research, Policy and Theory*, 2 CRIMINOLOGY & PUB. POL'Y 51 (2002).

Available at: https://scholarship.law.columbia.edu/faculty_scholarship/2570

This Article is brought to you for free and open access by the Faculty Publications at Scholarship Archive. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Scholarship Archive. For more information, please contact cls2184@columbia.edu.

THE PREVENTIVE EFFECTS OF ARREST ON INTIMATE PARTNER VIOLENCE: RESEARCH, POLICY AND THEORY*

CHRISTOPHER D. MAXWELL

Michigan State University

JOEL H. GARNER

Joint Centers for Justice Studies, Inc.

JEFFREY A. FAGAN

Columbia University

Research Summary:

This research addresses the limitations of prior analyses and reviews of five experiments testing for the specific deterrent effect of arrest on intimate partner violence by applying to individual level data consistent eligibility criteria, common independent and outcome measures, and appropriate statistical tests. Based on 4,032 cases involving adult males who assaulted their female intimate partners, multivariate regression analyses show consistent but modest reductions in subsequent offenses targeting the original victim that is attributable to arresting the suspect. Although the reductions attributable to arrest are similar across all five studies, other factors, such as the suspect's prior arrest record, are stronger predictors of subsequent offenses. The effect of arrest is also modest compared with the general decline in offenses toward the same victim during the follow-up period.

Policy Implications:

These results lend limited support for policies favoring arrest over informal police responses to intimate partner violence. However, the analyses also show that despite police intervention, a minority of suspects repeatedly victimize their partners and that factors other than formal sanctions play larger roles in explaining the cessation or continuation of aggressive behavior between intimates. These findings suggest that new policies replacing or enhancing arrest that target

* The National Institute of Justice (Grant 93-IJ-CX-0021), the Centers for Disease Control and Prevention (Grant R49/CCR210534), and the Harry Frank Guggenheim Foundation supported this research. Points of view in this publication are those of the authors and do not necessarily represent the official position or policy of the U.S. Department of Justice, the U.S. Department of Health and Human Services, the Harry Frank Guggenheim Foundation, Michigan State University, the Joint Centers for Justice Studies, or Columbia University. Please direct all correspondence about this publication to Christopher Maxwell (e-mail: cmaxwell@msu.edu).

potential repeat offenders might produce larger reductions in intimate partner violence.

KEYWORDS: Deterrence, Intimate partner violence, Mandatory arrest, Policing, Arrest

In the past quarter-century, many alternatives for the appropriate law enforcement response to intimate partner violence have been proposed, studied, recommended, adopted as policy, and enacted in federal and state laws. These alternatives have varied from doing nothing to on-scene counseling, temporary separation, and more formal criminal justice sanctions such as arrest, restraining orders, and coerced treatment (Fagan, 1996). The rationales for these policies were based on theories about deterrence, rehabilitation, incapacitation, victim empowerment, officer safety, and a general concern for the efficacy of criminal law regarding intimate private relationships (Fagan and Browne, 1994:3; Zimring, 1989:11). Until the 1980s, the empirical base for assessing the extent to which the alternative policies fulfilled the promises of their theoretical rationales was thin. In the foreword to a domestic violence research report that showed domestic violence was repetitive and highly visible to police, James Q. Wilson asserted that the criminal justice field lacks “reliable information as to the consequences of following different approaches” when responding to intimate partner violence. He argued that “gathering such information in a systematic and objective manner ought to be a high-priority concern for local police and prosecutors” (Wilson, 1977:v).

For the past 25 years, the law enforcement and research community has addressed Wilson’s challenge by gathering systematic and objective information about alternative police responses to intimate partner violence. However, gathering information alone has not led to a clear understanding of the consequences of alternative policies or to the strength of the theories underlying those policies (Davis and Smith, 1995; Fagan, 1996). To alleviate this shortcoming, we use common data and consistent measures from 4,032 incidents of misdemeanor assault compiled in five jurisdictions to test the preventive effects of arrest on intimate partner violence. We begin by reviewing the published results from six field experiments that tested for the deterrent effects of arrest on intimate partner violence. Then we describe our methods for pooling data and conducting our analyses of the five coordinated field experiments known collectively as the Spouse Abuse Replication Program (SARP). Next, we present the results of analyses using multiple data sources, methods, and measures. We conclude with a discussion of the policy implications of our re-analysis.

BACKGROUND

In 1980, the Minneapolis Police Department and the Police Foundation accepted Wilson's challenge and proposed to compare three alternative police responses to partner violence: arrest, on-scene counseling, and separation (Sherman, 1980). This proposal was innovative in using arrest as the tested sanction rather than conviction or prison time. However, it took a more traditional approach to confirm specific deterrence theory by testing for a negative relationship between the use of a formal sanction against an individual and that person's subsequent illegal behavior. In this study, volunteer Minneapolis officers carried out one of the three alternative responses based on an experimental design. Sherman and Berk (1984b) reported that when police did not arrest the suspect during a misdemeanor spouse assault incident, 21% re-offended within six months according to official records, a rate 50% higher than the 14% re-offending rate of arrested suspects. Results were similar when re-offending was measured by victim interviews. Thus, the experiment designed to test a specific deterrence theory found consistent, statistically significant, and supportive findings for what was by 1984 becoming the preferred policy option among domestic violence reform advocates—arresting the suspect.

POLICY IMPACT OF THE MINNEAPOLIS EXPERIMENT

Although the results of this experiment received extensive coverage on national television and in newspapers, the actual impact of this research is difficult to gauge. The policy debate about police response to domestic violence shifted quickly during the 1980s from one in which many jurisdictions did not authorize police officers to make arrests in misdemeanor assault unless they occurred in the officer's presence, to laws and policies that encouraged the use of arrest, to laws and policies that mandated arrest in at least some circumstances (Hirschel and Hutchison, 1991:3). Sherman and Berk (1984a) interpreted their Minneapolis's findings as support for using arrest but not necessarily for the mandated use of arrest. Nevertheless, several indications show that the Minneapolis experiment influenced the policy debate about the appropriate police response to domestic violence (Boffey, 1983; Lempert, 1984; Sherman and Cohn, 1989; U.S. Attorney General's Task Force on Family Violence, 1984). What is less clear is whether this experiment's impact stems from its grounding in theory, experimental design, consistent findings, visibility of the research results, or compatibility of its pro-arrest findings with growing public support for more formal sanctions for domestic violence.

THE SPOUSE ASSAULT REPLICATION PROGRAM

Support for replication of the Minneapolis experiment was widespread

among researchers and policy makers. Sherman and Berk (1984b) urged replication, and some academics' early praise for the study also was tempered by others preference for replication (Boffey, 1983; Lempert, 1984). The U.S. Attorney General's Task Force on Family Violence also encouraged replicating the Minneapolis experiment (1984). By 1986, six new experiments were initiated in Atlanta, Charlotte, Colorado Springs, Dade County, Milwaukee, and Omaha. Each new study involved experimental comparisons of arrest with alternative police responses to misdemeanor spouse assault incidents and measured victim safety using official police records and victim interviews (Garner and Maxwell, 2000).

These new experiments became known as the Spouse Assault Replication Program (SARP), but that name is a misnomer because the designers of the new experiments changed several crucial aspects of the Minneapolis design. For instance, in each new experiment, police officers determined case eligibility before the researchers assigned an alternative treatment to carry out. This method of determining eligibility without knowing the randomized treatment is preferred for experimental studies. The Minneapolis experiment, however, permitted officers to know the treatment before they decided case eligibility (Sherman and Berk, 1984b). In addition, the Minneapolis experiment attempted to interview victims by phone every two weeks. In the design of the SARP experiments, victim interviews were to occur twice, once within a month of the experimental incident and once at six months after the experimental incident.¹ The SARP experiments also developed a series of common measures about suspects, victims, treatments, and outcomes. Overall, the SARP experiments built on the Minneapolis design, increasing the number of sites and experimental incidents, enhancing the rigor of the random assignment, archiving the research data, and promoting commonality among the new experiments at the expense of commonality with the original Minneapolis experiment (Garner and Maxwell, 2000; National Institute of Justice, 1985).

SYNTHESIZING SARP FINDINGS

The published findings from the SARP experiments generated a complex mixture of deterrence, null, and escalation effects. Where there once had been one experiment with two statistically significant and consistent results, there now were six experiments with their own set of internally

1. In Omaha, victims were to be interviewed three times over a year. In Colorado Springs, three-quarters of the victims were to be interviewed every two weeks for the first three months, and all were interviewed at six months. In Milwaukee, 25% of victims were not interviewed until six months after the experimental incident. For more details about the SARP designs, see Maxwell (1998) and Garner and Maxwell (2000).

and externally inconsistent findings. Seven prior efforts have tried synthesizing the substantive findings from these experiments. These efforts found deterrent effects where the original authors did not (Zorza and Woods, 1994), identified the inconsistent and incomplete nature of the published findings (Garner et al., 1995), produced deterrent effects from a meta-analysis of prevalence findings (Sugarman and Boney-McCoy, 2000), reported mixed effects by site in a review of each experiment including Minneapolis (Sherman, 1992b), asserted that the effects of arrest vary by the marital and employment status of suspects (Berk et al., 1992; Schmidt and Sherman, 1992), and made an expert assessment that "arrest in all misdemeanor cases will not, on average, produce a discernable effect on recidivism" (Chalk and King, 1998:176). Thus, prior efforts at synthesis vary almost as much as do the published reports from the individual sites.

Although alternative approaches to synthesizing a large body of research, such as qualitative literature reviews and meta-analysis of published findings, have contributed to our understanding of this and other bodies of research, the secondary analysis of case-level data provide the most rigorous method for combining information across a variety of studies (Cooper and Hedges, 1994). Although efforts at secondary analysis of the archived data from the SARP experiments (see Berk et al., 1992; Sherman et al., 1992) have provided insights into the conditions under which arrest may or may not improve women's safety, they still are incomplete for several reasons. First, they use only the official records as a measure of repeat offending and do not consider information generated by thousands of victim interviews. Second, Sherman et al. (1992a) considered only the frequency of re-offending and Berk et al. (1992) considered only the prevalence of re-offending. Finally, Sherman et al. (1992a) reported two single site analyses (Milwaukee and Omaha) and Berk et al. (1992) did not use information from the Charlotte experiment. For these and other reasons, the published syntheses of the SARP experiments cannot be the definitive assessments of the average effect of arrest on subsequent offending.

SARP'S POLICY IMPACT

The inconsistencies in the site-specific and the multisite analyses reported in the various SARP reviews generated some ambivalence among researchers (Berk, 1993; Fagan, 1996; Sherman, 1992b) and policy makers (Clark, 1993; Frisch, 1992; Lerman, 1992; Mitchell, 1992) about the efficacy of arrest as the primary mechanism to control intimate partner violence. However, this ambivalence does not appear to have influenced police practices. We know of no jurisdiction that revised its policy to reflect the concerns these scholars raised. Furthermore, under the authority of the 1994 Violence Against Women Act, the U.S. Department of Justice initiated the Grants to Encourage Arrest Policies Program. To

support this program, Congress appropriated nearly \$120 million between 1994 and 1996 for the Violence Against Women Office to help local jurisdictions "implement mandatory arrest or proarrest programs and policies in police departments, including mandatory arrest programs and policies for protection order violations" (Violence Against Women Grants Office, 1996:5). Thus, presently, the findings (and the interpretation of findings) from the SARP experiments are not as closely connected to current policies and practices as the findings from the Minneapolis experiment were in the 1980s and 1990s.

METHODS

This study builds on a body of research that addresses the specific deterrent effect of arrest on the subsequent aggressive behavior by intimate partners. We synthesize the original data generated by the SARP experiments and conduct analyses that differ from the site-specific analyses in several ways. First, in contrast to prior secondary analyses (see Berk et al., 1992; Sherman et al., 1992a), our design conforms to the program's original plans for a multisite analysis of the case-level data.² Second, we use information about subsequent offenses from all victim interviews and from all official police records. Third, we use common measures about suspects, victims, incidents, and treatments to apply consistent case eligibility standards across the five experiments in which arrest was one possible treatment. Finally, we address (1) the complexities of combining data from five independent studies with systematic design differences; (2) the variability in the existence, number, and timing of victim interviews; and (3) the differences in the collection of the official data. The SARP experiments, by design, drew cases from different populations. They varied in size from 330 to 1,600 cases. They randomly assigned arrest to two-thirds (Milwaukee), one-half (Dade), one-third (Omaha and Charlotte), and one-fourth (Colorado Springs) of the eligible cases. In Dade County, the experiment was initially limited to married couples; in Milwaukee, the experiment included assaults between siblings and gathered cases only from selected neighborhoods. Although there were other differences in incident eligibility rules between sites, this is the first effort to synthesize the SARP experiments that addresses these issues (see Maxwell, 1998 for a detailed listing of differences among sites).

SELECTION OF CASES FROM THE POOLED DATA

We used common measures about suspects and experimental incidents (e.g., the incident at which the treatment was assigned and delivered) to

2. The idea for this design was originally proposed by Albert J. Reiss, Jr. and Robert F. Boruch.

select a research sample that best represents the archetypical male-on-female assaults that drive much of the policy debate about controlling intimate partner violence. The five sites collected data about 4,792 experimental incidents; we use information from the 4,032 incidents involving adult male suspects who assaulted their female intimate partner. To arrive at 4,032 cases, we excluded 306 incidents that involved a female suspect and 314 incidents that involved a male victim. Other experimental incidents excluded in our study involved victims and offenders whose relationships were not spouse-like, such as brothers and sisters ($n = 85$), and experimental incidents that did not involve an assault or victim injury ($n = 34$) (see Maxwell et al., 2001) for additional details about sample selection).

TREATMENTS ASSIGNED AND COMPARED

One of the SARP requirements was that arrest be one of the alternative police treatments tested. The nature of the alternatives to arrest and the proportion of cases assigned to arrest and nonarrest treatments were left to the implementing teams of researchers and police agencies. Of the 4,032 suspects in the research sample, 43.4% were assigned to the arrest treatment and the remaining 56.6% were assigned to a variety of nonarrest treatments (see Table 1). In our analyses, we compare the arrest treatment with all of the nonarrest treatments. Binder and Meeker (1988) suggested this comparison of the formal sanction of arrest with the informal alternatives in their critique of the original Minneapolis analysis. This method also was used in subsequent analysis of the Minneapolis experiment (see Berk and Sherman, 1988), and in some original SARP analyses (see Pate and Hamilton, 1992; Sherman et al., 1992a).

The designs in each of the SARP sites allowed officers to avoid using the randomly assigned treatment and apply an alternative treatment under certain conditions, such as an assault on the victim in the officers' presence or an assault on the officers. Consequently, the treatments delivered differed from the treatment assigned in 6.7% of the incidents.³ We chose to compare suspects based on the treatment randomly assigned. This is the only comparison for which we have a statistical basis for assuming uncorrelated error terms for both measured and unmeasured characteristics of suspects (Armitage, 1996:13; Heckman and Robb, 1986). This choice also was consistent with the method used in the original Minneapolis experiment, the SARP design, each of the original SARP analyses, and clinical

3. The majority of decisions to avoid the random assignment (78% of misdelivered treatments) involved suspects who were randomly assigned to nonarrest treatments but were arrested. Officers chose not to arrest when arrest was assigned in 59 (22%) of all misdelivered treatments.

TABLE 1. SAMPLE CHARACTERISTICS, TREATMENTS ASSIGNED, VICTIM INTERVIEW RATES BY SITE

	Charlotte		Col. Springs		Dade County		Milwaukee		Omaha		All Sites	
	Row %	N	Row %	N	Row %	N	Row %	N	Row %	N	Row %	N
Research Sample	15.8	638	30.7	1,238	22.5	906	23.7	954	7.3	296	100.0	4,032
	Col. %	N	Col. %	N	Col. %	N	Col. %	N	Col. %	N	Col. %	N
	100.0	638	100.0	1,238	100.0	906	100.0	954	100.0	296	100.0	4,032
Treatment Assigned												
Arrest	33.2	212	26.3	325	51.1	463	67.9	648	33.8	100	43.4	1,748
Nonarrest	66.8	426	73.7	913	48.9	443	32.1	306	66.2	196	56.6	2,284
Suspect Characteristics												
Mean Age	32.8		30.8		35.1		31.0		31.4		32.2	
Use of Intoxicant	54.2	346	59.3	734	30.6	277	29.4	280	59.121	175	44.9	1,812
Race/Ethnicity												
African-American	69.9	446	30.2	374	41.8	379	75.4	719	41.9	124	50.6	2,042
White	27.7	177	54.2	671	36.1	327	19.9	190	50.3	149	37.5	1,514
Hispanic	0.3	2	14.5	180	22.1	200	4.2	40	4.7	14	10.8	436
Asian/Other	2.0	13	1.1	13	0.0	0	0.5	5	3.0	9	1.0	40
Relationship with Victim												
Married	48.4	309	66.7	826	78.7	713	30.9	295	46.3	137	56.5	2,280
Separated	1.6	10	4.3	53	2.9	26	0.3	3	0.0	0	2.3	92
Divorced	0.3	2	0.6	8	2.0	18	0.7	7	1.4	4	1.0	39
Current/Past Intimate	49.7	317	28.4	351	16.4	149	68.0	649	52.4	155	40.2	1,621
Prior Arrest	30.7	196	42.8	530	12.1	110	61.8	590	64.9	192	40.1	1,618
Employed	77.1	492	86.7	1,073	70.5	639	47.2	450	78.0	231	71.6	2,885
Extent of Victim Interviews Completion												
Initial	64.4	411	82.9	1,026	65.5	593	60.2	574	79.4	235	70.4	2,839
Final	50.2	320	70.4	872	42.4	384	77.9	743	73.0	216	62.9	2,535
Any Victim Interview	64.4	411	87.9	1,088	65.7	595	85.7	818	79.4	235	78.1	3,147

trials in medical research.⁴

Sample characteristics vary by site and by treatment assigned. For instance, Colorado Springs contributes 30.7% of the research sample and Milwaukee provides 23.7% (see Table 1). The Omaha experiment contributed only 7.3% of the total sample. Thus, the larger sites contributed three or four times as many experimental incidents as did the smallest site. Approximately 19% of subjects in the research sample were less than 24 years of age, and almost 45% were older than 31. At the time of the experimental incident, either the police determined or the victim reported that the suspect was using an intoxicant—either alcohol or illegal drugs—in 45% of the incidents. In 37.5% of the research sample, the suspect's race was white. At the time of the experimental incident, most suspects were married to the victim (58.8%), most were employed (71.5%), and 40% had at least one prior arrest. As displayed in Table 2, the proportion of cases assigned to the arrest treatment varied significantly by site as well as by suspect's race, marital status, arrest record, employment condition, and use of intoxicants at the time of the experimental incident. All of these differences are statistically significant ($p < 0.05$). Suspect age was the only characteristic with similar distributions between the arrest and the nonarrest cases.

CHARACTERISTICS OF THE POOLED SAMPLE

Thus, Table 2 established that the research sample has unequal proportions in the characteristics of suspects assigned to arrest and nonarrest treatments. Unequal proportions in the treatment and control groups are unlikely in a single site experiment, but when we merged the data from the five experiments, three differences emerged: (1) Both criminal histories and social characteristics of suspects vary by site, (2) the sites have different numbers and types of incidents, and (3) the sites assigned different proportions of suspects to the arrest treatment. Accordingly, the unequal proportions of cases, the uneven distribution of victim and suspect characteristics, and the uneven allocation of cases to treatment groups are factors we must address in our multisite analyses that test for the effects of arrest on subsequent criminal behavior.

VICTIM INTERVIEWS

The SARP design called for initial victim interviews within a month of the experimental incident and a second interview six months after the experimental incident. However, the SARP researchers were not able to

4. A logistic regression model found that the only suspect characteristics that predicted misdelivery of treatment are intoxication ($b = 0.57$; $p < 0.001$) and unemployment ($b = 0.30$; $p < 0.05$).

TABLE 2. SITE, SUSPECT AND INCIDENT CHARACTERISTICS BY TREATMENT ASSIGNED ($N = 4,032$)

		Treatment Assignment				Total	
		Non-Arrest		Arrest			
		Row %	<i>N</i>	Row %	<i>N</i>	Row %	<i>N</i>
Selected Incidents		56.6	2,284	43.3	1,748	100	4,032
		Col. %	<i>N</i>	Col. %	<i>N</i>	Col. %	<i>N</i>
Selected Incidents		100	2,284	100	1,748	100	4,032
Site	Charlotte	18.7	426	12.1	212	15.8	638
	Colorado Springs	40.0	913	18.6	325	30.7	1,238
	Dade Co.	19.4	443	26.5	463	22.5	906
	Milwaukee	13.4	306	37.1	648	23.7	954
	Omaha	8.6	196	5.7	100	7.3	296
Age	18 to 24	19.7	449	17.3	303	18.7	752
	25 to 28	21.1	481	20.7	361	20.9	842
	29 to 31	15.6	356	16.1	281	15.8	637
	32 to 37	22.3	510	22.9	400	22.6	910
	38 to 82	21.4	488	23.1	403	22.1	891
Use of intoxicant		49.5	1,130	39.0	682	44.9	1,812
Race	Non-White	59.1	1,350	66.8	1,168	62.5	2,518
	White	40.9	934	33.2	580	37.5	1,514
Marital status							
	Non-married	38.4	877	44.9	784	41.2	1,661
	Married	61.6	1,407	55.1	964	58.8	2,371
Prior arrest		36.9	843	44.3	774	40.1	1,617
Employed		74.3	1,697	67.9	1,187	71.5	2,884

interview all of the victims; those victims interviewed were rarely interviewed according to the plan, and several sites added additional interviews or deviated from the basic design for some of their cases.⁵ Initial victim interviews were completed in more than 70% of the research sample (see Table 1), but the rate varied from 60% in Milwaukee (where, by design, initial interviews were not attempted in 25% of the cases) to almost 80% in Omaha. In just less than 63% of the research sample, a final interview

5. The Colorado Springs design called for interviews in the first and sixth month for 25% of the cases and for monthly interviews for 75% of the sample. The archived data from this site did not identify the actual date of any interview. To avoid losing the Colorado Springs data, we used the dates the interview was scheduled to occur.

was obtained. Milwaukee had the highest proportion of final interviews, with more than 79%, and Charlotte, at 50%, had the lowest. In more than 78% of the research sample ($n = 3,147$), at least one interview took place with the victim. Besides site differences in the proportion interviewed, the timing of the initial interviews varied from 1 to 776 days after the experimental incident; the mean number of days was 39.2. The actual time to the final interviews varied from 12 to 674 days, and the mean number of days was 280. Thus, the actual exposure time covered by the final interviews was, on average, 97 days longer than the planned 183 days.

We addressed the methodological issues created by the difference in victim interview rates and length of follow-up in several ways. First, we extended the basic Heckman selection model (Heckman, 1979) and produced a time-dependent latent-hazard selection measure. This approach used as its dependent measure the maximum length of time that the researchers tracked each victim during the study, rather than the traditional dichotomized measure of interview completion or noncompletion. For the 22% of victims who never were interviewed, their interview exposure time was set to one day. For the remaining victims with one or more interviews, their follow-up time was set to equal the number of days between the experimental incident and the date of their last interview. We then modeled the length of all 4,032 victim interview times as a function of the site and victim characteristics using a semiparametric maximum-likelihood Cox regression model. Using this process, we then produced a latent interview exposure rate for every victim. In our analyses of repeat offending, we used this measure as one means of addressing measured differences between the interviewed and noninterviewed victims, as well as to control for the different lengths of victim follow-up. This measure also permitted us to take advantage of the information about new victimizations gathered from all interviews regardless of the number or timing of interviews, rather than just those victimizations reported in the final interviews.

In the model predicting the length of victim interview follow-up, two of the seven measures tested, the site and the victim's age, predicted variation in the time covered by a victim's interviews. Dade County, Milwaukee, and Omaha had longer follow-up periods on average compared with Charlotte. In all sites, older victims also were observed over a longer follow-up period. The suspect's assigned treatment, the victim's relationship with the suspect, the victim's employment status, race, and the timing of any subsequent criminal offenses filed with the police involving the suspect did not significantly predict differences in whether the victim was interviewed or the average length of the follow-up period.

OUTCOME MEASURES

The SARP design called for collecting data that would permit the computation of the prevalence, frequency, and time-to-failure dimensions of the criminal career paradigm (Blumstein et al., 1986). In addition, the design called for computing each of these dimensions separately for violent offenses, property damage, and other types of offenses against the same victim, other victims, and any victim. This approach resulted in potentially hundreds of outcome measures derived from the official records and victim interviews. However, after review of the raw data and each site's data collection instruments and protocols, we found that not all of the data needed to calculate all of these measures were available for each of the five SARP experiments (see Maxwell, 1998 for details on the diversity and commonality of available measures in the archived data).

Using the available information in the victim interviews, we constructed composite outcome measures that capture incidents of subsequent assaults, verbal threats of assault, or property damage by the suspect against the original victim. From local law enforcement criminal history records, we constructed a measure that captured any reported offense against the same victim after the experimental incident. Using this measure, we then calculated a six-month prevalence rate, an annual incident rate, and a time-to-first-failure rate. From the victim interview data, we also calculated prevalence and six-month incident rates. Because sufficient information to determine the date of subsequent incidents reported in victim interviews was not always present in the victim interviews, we were unable to calculate a time-to-first-failure rate for all five sites. Both measures also are limited to incidents involving the offender and the victim identified in the original experimental incident. The interview measure captures threats, but police records typically do not include threats. Of course, many victims do not report offenses to the police (Bachman and Coker, 1995), and police do not always document all citizen complaints (Klinger and Bridges, 1997). Thus, we expected that victim interviews would identify more frequent victimization than would the official records. Therefore, we preferred the victim interview information over the official records for a measure of repeat offending and victim safety.

MULTIVARIATE ANALYTICAL MODELS

We estimated the effects of arrest on the recurrence of intimate partner violence in a series of models using (1) the treatment and site measures; (2) treatment, site, and the interview exposure measures; (3) treatment, site, interview exposure, and site by treatment interaction measures; and (4) treatment, site, interview exposure, and six suspect characteristics measurements thought to be associated with increased risk of re-offending.

The choice of the appropriate estimation routine for the five outcome measures reflected the three measurement dimensions: The prevalence of any new victimization is a dichotomy, the rate of aggressive incidents is a count, and the time-to-first-offense is a right-hand censored interval measure of days between the experimental incident and the first officially recorded offense. For dichotomous dependent measures of prevalence, models were estimated using logistic regression methods. To estimate the number of incidents, models were tested using Negative Binomial Regression.⁶ We estimated the time-to-first-failure rates using the Cox semiparametric regression.⁷ For each regression procedure, we report the unstandardized coefficients, the coefficient's standard errors, and the odds ratios. We use the odds ratio as a rough measure of the relative size of the effect of arrest.

RESULTS

OFFENSES AND VICTIMIZATIONS BASE RATES

Table 3 reports prevalence and mean incident rates of new incidents reported in the victim interviews and found in the official police records. In the official police records, 23.1% of the suspects in the research sample had one or more reported offenses after the experimental incident. The annual incident rate averaged 0.39 offenses per suspect. Among 3,149 victims interviewed at least once, 42.5% reported at least one new victimization by the suspect through the final victim interview. These same victims reported 9,009 incidents (an average of 2.86 incidents per suspect) during this period after the experimental treatment. Thus, similar to what was found in other domestic violence research (see Feld and Straus, 1989; Langan and Innes, 1986; Quigley and Leonard, 1996), analysis of the official criminal history records indicated no new offenses against three-quarters of all suspects, and almost three-fifths of the interviewed victims reported no new victimizations. However, when victims report at least one new incident, the average number of victimizations was 6.7, or greater than one victimization per month.

6. We compared a poisson regression with a negative binomial regression, but in every instance, there was too much overdispersion to justify using the results based on the Poisson Regression model. This result is consistent with what has been found with other criminal justice data of similar format (Land et al., 1996).

7. We tested the Cox regression assumption of equal or proportional hazards by introducing a time-dependent covariate that indicates whether the effect of arrest is dependent on the passage of time. While controlling for site effects, we found that the time-dependent covariate is not significantly associated ($b = 0.00$; $p = 0.08$) with the hazard rate.

TABLE 3. BASE RATES OF FAILURE BY
TREATMENT ASSIGNED

<u>Official Records Sample</u>	<u>Non-Arrest</u>		<u>Arrest</u>		<u>Total</u>	
	<u>(N = 2,284)</u>		<u>(N = 1,748)</u>		<u>(N = 4,032)</u>	
	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>
Six Month Prevalence of Recidivism	21.1	481	25.7	450	23.1	931
	Chi-Sqr. = 0.725					
Mean Annual Frequency of Recidivism	0.32		0.48		0.39	
	F-Value = 25.397 ***					
Mean Survival Time (days)	865		816		850	
	Log-Rank = 15.8 ***					
<u>Victim Interview Sample</u>	<u>Non-Arrest</u>		<u>Arrest</u>		<u>Total</u>	
	<u>(N = 1,789)</u>		<u>(N = 1,358)</u>		<u>(N = 3,147)</u>	
	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>
Prevalence of Victimization	42.3	756	42.9	583	42.5	1,339
	Chi-Sqr. = 0.14					
Mean 6 Month Frequency of Victimization	2.74		3.03		2.86	
	F-value = 1.075					

EFFECTS OF ARREST

With multiple outcome measures and sources, our criteria for judging across the five models for the existence of an effect for arrest are a combination of the consistency of direction, its relative size compared with other measures in the models, and its statistical significance. Among these three criteria, we place greater stock in the consistency of direction effects and in the size of effects, and we de-emphasize statistical significance tests. As other criminologists have noted, the use of statistical significance tests is technically not appropriate for nonprobability samples (Sampson and Laub, 1993), such as those used in the Minneapolis and SARP experiments. Nagin and Farrington (1992:519) similarly argue that "empirical regularities" such as the consistent direction of effects are the "grist for useful theory," rather than one or two tests for statistical significance. Yet, like Sampson and Laub (1993), we also report statistical significance tests to help avoid type I errors.

As reported in Table 4, arrest reduced the prevalence of new victimization by 25% and the incidents of victimizations by 30%. In the official criminal history data, arrest also was associated with reductions of 4% in the prevalence and 8% in the incidence of recidivism, as well as a 12%

TABLE 4. MULTIVARIATE TESTS OF THE EFFECTS OF ARREST ON INTIMATE PARTNER VIOLENCE

Independent Variable:	Victim Reports (n = 3,147)						Official Reports (n = 4,032)								
	Prevalence ^a			6-Month Rate ^b			6-Month Prevalence ^a			Annual Rate ^b			Time-to-Failure ^c		
	B	S.E.	Exp(B)	B	S.E.	Exp(B)	B	S.E.	Exp(B)	B	S.E.	Exp(B)	B	S.E.	Exp(B)
Arrest	-0.28	0.08	0.75 ***	-0.35	0.08	0.70 ***	-0.04	0.08	0.96	-0.08	0.07	0.92	-0.03	0.06	0.88
Site (Charlotte)															
Col. Springs	-0.97	0.13	0.38 ***	-0.49	0.15	0.61 ***	0.28	0.12	1.33 *	-0.02	0.13	0.98	-0.02	0.10	0.76 **
Dade Co.	-0.49	0.14	0.61 ***	-0.01	0.16	0.99	0.64	0.13	1.89 ***	0.64	0.12	1.89 ***	0.35	0.11	1.61 ***
Milwaukee	0.48	0.13	1.61 ***	0.96	0.16	2.62 ***	0.23	0.13	1.26	0.67	0.12	1.96 ***	0.16	0.11	1.10
Omaha	0.42	0.17	1.52 *	-1.01	0.20	0.36 ***	-0.54	0.19	0.58 **	-0.60	0.19	0.55 ***	-0.80	0.16	0.39 ***
Interview exposure	0.86	0.11	2.36 ***	-0.55	0.09	0.58 ***	0.54	0.09	1.71 ***	0.08	0.10	1.09	0.26	0.07	1.28 ***
Suspect's															
Age	-0.39	0.15	0.68 **	-0.83	0.14	0.44 ***	-0.50	0.15	0.61 ***	-0.32	0.13	0.72 *	-0.41	0.13	0.71 **
Use of Intoxicant	-0.01	0.08	0.99	0.36	0.07	1.43 ***	0.21	0.08	1.23 **	0.10	0.08	1.10	0.14	0.06	1.12
White	0.23	0.08	1.26 **	0.30	0.07	1.35 ***	-0.39	0.08	0.68 ***	-0.36	0.07	0.70 ***	-0.33	0.07	0.74 ***
Married	0.04	0.09	1.05	0.11	0.07	1.11	-0.01	0.08	0.99	0.10	0.08	1.10	-0.03	0.07	0.90
Prior Arrest	0.45	0.08	1.56 ***	0.11	0.09	1.12	0.92	0.08	2.50 ***	0.86	0.08	2.36 ***	0.78	0.07	3.34 ***
Employed	0.02	0.09	1.02	-0.23	0.08	0.79 ***	-0.06	0.09	0.94	-0.16	0.07	0.85 *	-0.07	0.07	0.94
Constant	0.20	0.53		3.98	0.47		-0.07	0.52		-0.11	0.46				
Negative Binomial Scalar				5.02	0.17					2.12	0.13				
Initial Likelihood	4292.51			-20158.7			4735.61			-4224.08					17656.49
Final Likelihood	3949.93			-5537.95			4476.31			-3618.03					17384.65

^a = Logistic Regression; ^b = Negative Binomial Regression; ^c = Cox Regression.
 * = p [0.05; ** = p <0.02; *** = p <0.001.

reduction in the hazard rate. In other words, based on five outcome measures from two sources, there were consistently smaller rates of subsequent victimization and recidivism among the suspects assigned to the arrest treatment versus the nonarrest interventions. For the two outcome measures based on victim interview data, these negative effects were statistically significant at the traditional $p < 0.05$ level.⁸ Using the three outcome measures derived from the criminal history information, the arrest treatment also was associated with a reduction in the recidivism rates, but none of the differences between the two treatment groups were statistically significant.

The results reported in Table 4 also show that there were statistically significant differences in the base rates of failures across the five sites for all five outcome measures. We therefore tested to determine whether the effects of arrest were related to the site for any of the five measures, but did not find any evidence that was the case (see Maxwell, 1998 for detailed results of these tests). Thus, because the effects for arrest were in the same direction across all five measures and in the same direction in each of the five sites, as well as statistically significant in two of the five models, the results support the notion that, compared with nonarrest interventions, arrest provides additional safety to female victims of intimate partner assault. This finding is consistent with the specific deterrence hypothesis that these studies originally were testing.⁹

EFFECT OF THE VICTIM INTERVIEW PROCESS

As mentioned above, we had to address the issue of missing victim interviews. We did so by including in the outcome models our interview

8. We report findings from the models that include controls for suspect characteristics because we believe these provide the most precise estimate of the effect of arrest.

9. In addition to arrest, some suspects received additional criminal justice sanctions or controls, such as restraining orders, conviction, probation, fines, or incarceration. In the Milwaukee experiment, the researcher found that 60% of the suspects appeared at the prosecutor's office for charging, 3.2% were charged, and 2.7% were required to attend counseling. The researchers also asked the victims about the presence of restraining orders and found that 12% reported having one against the suspect (Bousa et al., 1990). In the Charlotte site, about 25% of the suspects either pled guilty or were found guilty regardless of their assigned treatment. The Omaha and Colorado sites also collected data on court disposition, but no disposition information was collected in Dade. Although it may have been valuable to include measures of additional sanctions, there were several reasons we could not do so in our multisite analysis. First, these subsequent treatments were not randomly assigned. Thus, any analysis would be confounded with selection biases. Second, we do not know the timing of subsequent sanctions so we would not know whether re-offending started before or after the sanction. Third, although some of the experiments collected and archived data about some of these issues, these items were not common data elements in the SARP design.

exposure measure to assess the impact of the interview process on the outcome measures and on the effect of arrest. As reported in Table 4, we found higher victimization rates among those whose last interview occurred later in time after the initial experimental incident in four of the five comparisons. The one effect that was not significant was the relationship between exposure and failure measured by the victim's six-month victimization rate. This measure showed a negative association between the length of follow-up and the frequency of offenses reported by the victims.

Because our approach to controlling for missing victim interviews and time covered by the interviews was innovative, we also tested for relationships between arrest and our five outcomes without specifying the victim interview exposure measure. For the five models reported in Table 4, the exclusion of the exposure measure did not change the direction or statistical significance of either the treatment or site measures and the size of the arrest coefficients increased by an average of 3%. As an additional test of the interview selection process, we then modeled the criminal history data using only the suspects that had at least one victim interview ($n = 3,147$). The substantive results were similar to those reported in Table 4 for the full sample. Among this subsample, arrest led to a 7% reduction in the odds of any subsequent recidivism, but again this reduction was not significant ($p = 0.44$). The timing of the first subsequent incident ($b = 0.001$; $p = 0.988$) and the difference in the incident rates between the arrest and nonarrest groups were also not significant ($b = -0.01$; $p = 0.839$). Lastly, we compared the prevalence and incident rates of re-offending among those with and without a victim interview and found that suspects with noninterviewed victims have significantly smaller rates of recidivism across all three measures. This finding is contrary to our concern that noninterviewed victims in the SARP experiments might have had higher rates of repeat victimizations. From these results, we conclude that discrepancies in the size and statistical significance of the arrest coefficients from the victim interviews and from the official records were due less to differences in the interviewed and noninterviewed samples, and more to the ability of the victim interviews to capture subsequent victimizations not included in official police reports.

EFFECTS OF SUSPECT CHARACTERISTICS

In Table 4, we also included statistical tests for the suspect's age, use of intoxicants at the experimental incident, race, marital status, prior arrest record, and employment status. Older suspects were less likely to aggress against their female intimate partners (cf., Farrington, 1986:7). Intimate partner aggression is perhaps also resistant to low-cost legal sanctions: Having one or more prior arrests for any offense against any victim was

consistently associated with greater quantity of incidents, and this relationship was statistically significant in four of our five regression models (cf., Moffitt et al., 2000). The size of the relationship between prior arrest and new incidents ranged from about a 12% increase in the odds of a new victimization to a 234% increase in the odds for the time-to-first-failure in the official criminal history records. Alcohol use increased the risk of intimate partner aggression: The use of intoxicants at the time of the experimental incident was similarly associated with increased failures in four out of five outcome measures, and two of these positive relationships were statistically significant (the prevalence of victimization and the prevalence of recidivism) (cf., Fagan and Browne, 1994:3; Kantor and Straus, 1987).¹⁰

The suspect's race had a substantial and statistically significant relationship to subsequent failures, but the direction of the association depended on whether the data came from the victim interviews or official records. Based on victim interview data, victims were more likely to report new offenses if the suspect was white. Based on the criminal history information, white suspects were less likely to recidivate compared with all other suspects. The sizes of these relationships was substantial, from a 35% increase in the odds of victimization to a 30% decrease in the odds of recidivism. The role of the suspect's marital relationship with the victim also varied in size and direction by outcome measure. By three outcome measures, married suspects were more likely to fail, and by two measures, they were less likely to fail. However, only for the prevalence of new victimization was the increase statistically associated with married suspects. This one relationship was nearly a 5% increase. Finally, in four out of five models, employed suspects were less likely to commit additional incidents (cf., Sherman et al., 1992a). Two of the four negative relationships were statistically significant and ranged from 15% to 21% decreases in the odds of new incidents.

The results regarding the additional covariates suggest that the preventive effect of arrest was modest relative to the size of the relationships between suspect and victim characteristics and failure rates. We base this assessment primarily on comparing the sizes of Menard's (1995) standardized logistic regression coefficients, which corresponds to a one standard deviation increase in the independent measure for every b standard deviation change in logit (Y). The standardized coefficients based on the official criminal history data were arrest = -0.01; interview exposure = 0.09; prior arrest = 0.19; white = -0.08; age = -0.05; intoxication = 0.04; employed = 0.01; and, married = 0.00. The standardized coefficients for the victim interview data were arrest = -0.06; interview exposure = 0.15; prior record

10. The high correlation between victim and suspect demographics precluded us from including both in the same model.

= 0.10; arrest = -0.06; age = -0.05; white = 0.05; married = 0.01; intoxication = 0.00; and, employment = 0.00.

DISCUSSION AND CONCLUSION

The design and implementation of the Spouse Assault Replication Program approaches the standards for criminological research put forth by the National Academy of Sciences (Blumstein et al., 1978; Sechrest et al., 1979). The five experiments used a common protocol that included random assignment of treatment after selecting incidents, documentation of suspect and victim characteristics, collection of outcome data from multiple sources, and data elements that allowed for the construction of multiple dimensions of subsequent offending. Earlier efforts to synthesize the SARP results have failed to capitalize on these important characteristics. Instead, these earlier syntheses relied on qualitative methods (Schmidt and Sherman, 1993), tested models using data from only a few sites (Sherman et al., 1992a), modeled just one outcome from one data source (Berk et al., 1992), or relied on published results (Sugarman and Boney-McCoy, 2000). To address these shortcomings, our research integrated original data from all five experiments, standardized cases according to the nature of incidents, employed three dimensions of outcomes, controlled for natural variation between suspects and incidents, used multiple sources of information about outcomes, and controlled for variability in victim interview rates and timing.

POLICY IMPLICATIONS

Our findings of a consistent reduction in the incidents of victimization due to arrest, independent of other criminal justice sanctions and individual processes, support the continued use of arrests as a preferred law enforcement response for reducing subsequent victimization of women by their intimate partners.¹¹ Although the size and statistical significance of the effect of arrest varied depending on whether victim interviews or law enforcement records measured the suspect's subsequent aggression, in all five measures, arrest is associated with fewer incidents of subsequent intimate partner aggression. This finding exists during the first several days after the experimental incident as well as beyond one year. Thus, our

11. One alternative explanation for our findings is that arrest does not change suspect offending as much as it changes the victim's willingness to report offenses to the police and to interviewers. Two recent empirical studies do not support this speculation. Using data from SARP's Metro-Dade County Experiment, Hickman (2000) found that the use of arrest in the experimental incident was not related to reporting subsequent incidents to the police. Felson and Ackerman (2001), using NCVS data, found that the existence of prior domestic assaults increased the respondent's willingness to sign a police complaint.

research does not find that arrest will eventually increase the risk for violence against women.

In several ways, the results we generated from the Spouse Assault Replication Program provide stronger support for pro-arrest policies than do the results from the Minneapolis Domestic Violence experiment. First, our findings stem from five jurisdictions with a total sample of more than 4,000 male suspects, whereas the Minneapolis findings were based on 314 incidents collected in one jurisdiction. Second, the random assignment procedures used in SARP were more rigorous than the one used in Minneapolis. Third, SARP researchers interviewed almost 80% of the female victims, whereas only about 60% were interviewed in Minneapolis. Finally, our analyses control for missing victim interviews, the variability in timing of victim interviews, and suspect characteristics. Sherman and Berk's (1984b) published findings consider only the assigned treatment effects.

In addition, our data and methods may underestimate the current empirical support for arrest for a number of reasons. First, although Sherman and Berk's (1984b) Minneapolis results provide strong support for the deterrent effect for arrest, we could not incorporate cases from Minneapolis into our analyses. If the archived Minneapolis data were sufficiently complete and similar to SARP's common data, their use would likely enhance the evidence for the effectiveness of arrest. Second, our analyses compared cases assigned to arrest with those cases not assigned to arrest. However, the police arrested about 10% of the cases assigned to an informal treatment. Thus, the comparison between the arrest and the nonarrest treatments is diluted. Had the treatment assignment been implemented perfectly, the size of our reported deterrent effect might have been even larger.

On the other hand, because the size of the deterrent effects found in our analyses is smaller than those reported in the Minneapolis experiment, and because the results from our analyses of the official criminal history data do not reach statistical significance, our findings provide weaker support for pro-arrest policies than the Minneapolis findings provided. Nevertheless, on balance, we believe that the predominate weight of the empirical regularities favors a conclusion that arrest has a modest preventive effect on intimate partner violence.

Unlike most criminological research, the Minneapolis Domestic Violence Experiment and SARP were part of a continuing program of research that focused on a theory-driven policy addressing an important social problem. The results of these efforts have identified some benefits of arrest, specifically, the reduction of victimization of female intimate partners. However, as extensive as this program was, it does not provide a complete basis for a systematic examination of the costs and the benefits

of the use of arrest to address violence against women. For instance, some academics (e.g., Stark, 1993) have argued for arresting batterers no matter its deterrent benefits because the forced separation provides immediate, presumably incapacitative, protection for the victims. Others have pondered whether arrest may create negative outcomes when other effects on the victims, the suspects, and their families are considered. Binder and Meeker (1992) suggest that arrest may have both positive and negative collateral consequences on children in the household, the likelihood that a spouse will call the police in the future, the stability of the marriage, and the suspect's employment status (see also Berk and Sherman, 1985; McCord, 1992). In addition, both fiscal and resource expenditures are attached to arresting suspects as well as possible reductions in expenditures due to law enforcement agencies responding to fewer subsequent encounters with suspects. However, neither SARP nor other systematic research has generated evidence about other possible costs or benefits of arrests. Although this type of assessment is routine in the development of environmental, health, and safety regulations (Cohen, 2000), we are unaware of any attempt to address the question of whether the benefits we find outweigh the cost of arrest to society. Therefore, a more thorough assessment of a policy promoting or mandating arrest needs to capture both the major costs and benefits of arrest.

The findings from the SARP experiment also suggest real limitations in the effectiveness of arrest in reducing violence against women. The evidence from this study shows that regardless of the treatment assigned and irrespective of the data source, most victims reported no subsequent victimization by their male partner during follow-up periods that ranged from six months to more than two years. The mere physical presence of a police officer may redefine the parameters of the violence from an interpersonal struggle for power to one that now involves, formally or informally, outsiders. Although the presence of an outsider may last from a few minutes to an hour, it may be enough to convince many suspects that the victim and the police mean business. In other words, the threat of arrest may suffice as the best specific deterrent for most suspects.¹²

However, victimizations also persisted for about 40% of victims. We estimate that the average suspect with at least one subsequent incident had committed about an average of seven new incidents of aggression against the same victim within just the first six months of follow-up. Apparently, some women continue to be victimized multiple times by

12. The argument that there is desistance from calling the police is consistent with Bowker (1984) and Dutton et al.'s (1991) claim that social disclosure alone deters further domestic violence. Although these scholars were speaking about disclosure originating out of an arrest, the same effect seems likely to occur from just calling the police.

their intimate partners, even after the police have responded to a request for help. For these reasons, the SARP experiments show that arresting suspects, although effective on average, is not a panacea for all victims of intimate partner violence. This suggests that other policies, either replacing or enhancing the use of arrest, that focus on identifying potential repeat offenders and either treating, sanctioning, or incapacitating them might produce larger reductions in intimate partner violence. The challenge is twofold: developing plausible policies and carrying out a long-term, systematic research program that rigorously tests the underlying theories of those policies.

TESTING THEORY

Sherman (1980) designed the Minneapolis Domestic Violence Experiment as a test of specific deterrence theory. Its apparent impact on public policy might be a coincidence or a single example that illustrates a point—successful efforts at policy relevance need not be atheoretical. We believe that the use of theory in the Minneapolis and SARP experiments should enhance their policy relevance because the theoretical framework provides a basis for generalizing the results beyond a few jurisdictions at one point in time. However, the theoretical formulation used in the Minneapolis and SARP experiments was not fully developed. It merely asserted that there would be some reduction in criminal behavior with use of a formal sanction. This simple formulation makes specific deterrence theory infallible; because with any contrary finding, researchers can assert that the sanction was not swift, certain, or severe enough.

With our research, the field now has systematic evidence about how much reduction in subsequent violence is associated with arresting suspects for intimate partner violence. However, we agree with Lempert (1989) that additional sanctions need to be tested in ways that experimentally compare variations in the certainty, celerity, and severity across a variety of offenses, offenders, and victims. SARP contributes to this kind of theory-testing effort by examining one sanction for one offense in a variety of contexts. One of our concerns is that others might instead overgeneralize our results to support the use of more severe sanctions, particularly to control intimate partner violence, despite other research reporting no gains for domestic violence victims from more punitive practices such as prosecution (Davis et al., 1998; Ford, 1991; Steinman, 1990; Thistlethwaite et al., 1998). For these reasons, we advise against simple extrapolations of the preventive effects of arrest to other criminal justice sanctions, such as the restraining orders or incarceration of batterers without further systematic research and evaluation.

In addition, we suggest that future research obtain measurements of the offender's immediate and long-term cognitive reactions to sanctions and

treatments. Subsequent research also would be stronger if it measured secondary outcomes like changes in offenders' employment status and familial relationships, or the victims' and children's welfare.¹³ These measurements could further our understanding of the "black box" of specific deterrence (Manning, 1993:641) and permit an assessment of whether there are unintended consequences of sanctions on those not directly punished. As McCord (1992:233) argued, further research on domestic violence needs to go beyond the concept of deterrence by also assessing whether sanctions lead to outcomes like "loss of support for children or loss of shelter." Except for interviewing some arrestees in Milwaukee (Sherman et al., 1990), the SARP experiments did not measure the suspects' attitudes and perceptions and therefore cannot address the variety of underlying mechanisms suggested by Gibbs (1975) that might account for the association between increased sanctions and reduced offending found here. At this time, the available research can only estimate the amount of subsequent aggression reduced because of arrest. We need further research to understand more completely why and when sanctions deter and whether secondary consequences of arrest exist.

SUSPECT CHARACTERISTICS

Several suspect characteristics are significantly related to the prevalence, frequency, and timing of the first new incident of victimization and recidivism. For instance, the odds of new victimization were 30% to 60% less for each additional year of age. Also, according to the official data, suspects with prior arrests for any offense are from 250% to 330% more likely to commit new acts of intimate partner violence. Unlike the consistent effects for suspect's age and prior record, the contradictory findings regarding their race in victim interviews and official records present a conundrum for this and subsequent research. From official records, white suspects are less likely to re-offend after the experimental incident. The effect size is large, with a 30% reduction in the odds for both the frequency and timing of an officially recorded failure. However, the victim interview data yielded the opposite result: White suspects are 30% more likely than are nonwhite suspects to continue victimization. There are several plausible explanations for this contradiction, including race interactions with the severity of violence, the different treatments, or the willingness of victims to report incidents to the police.¹⁴ Future research

13. See, for instance, Paternoster and Brame's (1997) test of the effect of police procedural justice on recidivism using information from interviews of arrested offenders from the Milwaukee Domestic Violence Experiment.

14. Similar to our finding concerning the relationship between aggression and race, Bachman and Coker (1995) found evidence using the National Crime Victimization Survey that victimized African-American women were more likely to call the

will need to address these possibilities to unravel the complex role of race and should also incorporate the suspect's age and arrest record.

REFERENCES

- Armitage, Peter
1996 The design and analysis for clinical trials. In Subir Ghosh and Caly-ampudi Radhakrishna Rao (eds.), *Design and Analysis of Experiments*, Vol. 13. *Handbook of Statistics*. Amsterdam: Elsevier.
- Bachman, Ronet and Ann L. Coker
1995 Police involvement in domestic violence: The interactive effects of victim injury, offender's history of violence, and race. *Violence and Victims* 10:91-106.
- Berk, Richard A.
1993 What the scientific evidence shows: On the average, we can do no better than arrest. In Richard J. Gelles and Donileen R. Loseke (eds.), *Current Controversies on Family Violence*. Newbury Park, Calif.: Sage.
- Berk, Richard A. and Lawrence W. Sherman
1985 Data collection strategies in the Minneapolis domestic violence experiment. In Leigh Burstein, Howard E. Freeman, and Peter H. Rossi (eds.), *Collecting Evaluation Data: Problems and Solutions*. Beverly Hills, Calif.: Sage.
- Berk, Richard A., Alec Campbell, Ruth Klap, and Bruce Western
1992 The deterrent effect of arrest in incidents of domestic violence: A Bayesian analysis of four field experiments. *American Journal of Sociology* 57:698-708.
1988 Police responses to family violence incidences: An analysis of an experimental design with incomplete randomization. *Journal of the American Statistical Association* 83:70-76.
- Binder, Arnold and James Meeker
1988 Experiments as reforms. *Journal of Criminal Justice* 16:347-358.
1992 Arrest as a method to control spouse abuse. In Eve S. Buzawa and Carl G. Buzawa (eds.), *Domestic Violence: The Changing Criminal Justice Response*. Westport, Colo.: Greenwood.
- Blumstein, Alfred, Jacqueline Cohen, and Daniel Nagin (eds.)
1978 *Estimating the Effects of Criminal Sanctions on Crime Rates*. Washington, D.C.: National Academy Press.
- Blumstein, Alfred, Jacqueline Cohen, Jeffrey A. Roth, and Christy A. Visher (eds.)
1986 *Criminal Careers and "Career Criminals."* Washington, D.C.: National Academy Press.
- Boffey, Philip M.
1983 Domestic violence: Study favors arrest. *New York Times* (5 April):C1.

police compared with victimized white women. They also found that if the police were called to the scene, the officers were more likely to arrest the suspect if the victim and suspect were both African-American. Together, these two findings suggest an explanation for the relationship between the suspect's race and the likelihood of officially recorded failure found in SARP.

- Bousa, Dominick, Sherman, Janell D. Schmidt, Dennis Rogan, and Patrick Gartin
1990 Codebook and Frequencies for the Milwaukee Domestic Violence Experiment: The Criminal Justice Processing File. Washington, D.C.: Crime Control Institute.
- Bowker, Lee N.
1984 Coping with wife abuse: Personal and social networks. In Albert R. Roberts (ed.), *Battered Women and Their Families*. New York: Springer.
- Chalk, Rosemary A. and Patricia A. King, (eds.)
1998 Violence in Families: Assessing Prevention and Treatment Programs. Committee on the Assessment of Family Violence Intervention, Board on Children, Youth, and Families, National Research Council and Institute of Medicine. Washington, D.C.: National Academy of Science.
- Clark, Jacob R.
1993 Where to now on domestic-violence? Studies offer mixed policy guidance. *Law Enforcement News* (30 April):1,17.
- Cohen, Mark A.
2000 Measuring the Costs and Benefits of Crime and Justice. David Duffee (ed.), *Criminal Justice 2000*, 4, no. NCJ 182411. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.
- Cooper, Harris and Larry V. Hedges
1994 Research synthesis as a scientific enterprise. In Harris Cooper and Larry V. Hedges (eds.), *The Handbook of Research Synthesis*. New York: Russell Sage.
- Davis, Robert C. and Barbara Smith
1995 Domestic violence reforms: Empty promises or fulfilled expectations? *Crime & Delinquency* 41:541–552.
- Davis, Robert C., Barbara E. Smith, and Laura B. Nickles
1998 The deterrent effect of prosecuting domestic violence misdemeanors. *Crime & Delinquency* 3:434–442.
- Dutton, Donald G., Stephen G. Hart, Leslie W. Kennedy, and Kirk R. Williams
1991 Arrest and the reduction of repeat wife assault. In Eve Buzawa and Carl Buzawa (eds.), *Domestic Violence: The Changing Criminal Justice Response*. Westport, Conn.: Greenwood.
- Fagan, Jeffrey A.
1996 The Criminalization of Domestic Violence: Promises and Limits. Presented at the Conference on Criminal Justice Research and Evaluation, Washington, D.C., National Institute of Justice.
- Fagan, Jeffery A. and Angela Browne
1994 Violence against spouses and intimates. Panel on the Understanding and Control of Violent Behavior, Committee on Law and Justice, Commission on Behavioral and Social Science and Education, National Research Council. In Albert J. Reiss, Jr. and Jeffrey A. Roth (eds.), *Understanding and Controlling Violence*, Vol. 3. Washington, D.C.: National Academy Press.

Farrington, David P.

- 1986 Age and crime. In Michael Tonry and Noval Morris (eds.), *Crime and Justice: An Annual Review of Research*, Vol. 7. Crime and Justice. Chicago, Ill.: University of Chicago Press.

Feld, Scott L. and Murray Straus

- 1989 Escalation and desistance of wife assault in marriage. *Criminology* 27:141-161.

Felson, Richard B. and Jeff Ackerman

- 2001 Arrest for domestic violence and other assaults. *Criminology* 39:655-675.

Ford, David

- 1991 Prosecution as a victim power source: A note on empowering women in violent conjugal relationships. *Law & Society Review* 25:313-334.

Frisch, Lisa A.

- 1992 Research that succeeds, policies that fail. *Journal of Criminal Law & Criminology* 83:209-216.

Garner, Joel H. and Christopher D. Maxwell

- 2000 What are the lessons of the police arrest studies. *Journal of Aggression, Maltreatment & Trauma* 4(1):83-114.

Garner, Joel H., Jeffrey A. Fagan, and Christopher D. Maxwell

- 1995 Published findings from the Spouse Assault Replication Program: A critical review. *Journal of Quantitative Criminology* 11:3-28.

Gibbs, Jack P.

- 1975 *Crime, Punishment and Deterrence*. New York: Elsevier.

Heckman, James J.

- 1979 Sample selection bias as a specification error. *Econometrica* 47:153-161.

Heckman, James J. and Richard Robb

- 1986 Alternative methods for solving the problem of selection bias in evaluating the impact of treatment outcomes. In Howard Wainer (ed.), *Drawing Inferences from Self-Selected Samples*. Papers from a conference sponsored by Education Testing Services. New York: Springer-Verlag.

Hickman, Laura Jean

- 2000 Exploring the impact of police behavior on the subsequent reporting of domestic violence victims. Ph.D. dissertation, University of Maryland at College Park.

Hirschel, J. David and Ira W. Hutchison III

- 1991 Police-preferred arrest policies. In Michael Steinman (ed.), *Women Battering: Policy Responses*, Vol. 3. Highland Heights, Ky. and Cincinnati, Ohio: Anderson Publishing Co. and Academy of Criminal Justice Science.

Kantor, Glenda K. and Murray A. Straus

- 1987 The "drunken bum" theory of wife beating. *Social Problems* 34:213-321.

Klinger, David A. and George S. Bridges

- 1997 Measurement error in calls-for-service as an indicator of crime. *Criminology* 35:705-726.

Land, Kenneth C., Patricia L. McCall, and Daniel S. Nagin

- 1996 A comparison of Poisson, negative binomial, and semiparametric mixed regression models with empirical applications to criminal careers data. *Sociological Methods and Research* 24:387-442.

Langan, Patrick A. and Christopher A. Innes

- 1986 Preventing Domestic Violence Against Women. Special Report. Washington D.C.: U.S. Government Printing Office.

Lempert, Richard

- 1984 From the editor. *Law & Society Review* 18:505-513.
1989 Humility is a virtue: On the publication of policy relevant research. *Law & Society Review* 23:145-161.

Lerman, Lisa G.

- 1992 The decontextualization of domestic violence. *Journal of Criminal Law & Criminology* 83:217-240.

Manning, Peter K.

- 1993 The preventive conceit: The black box in market context. *American Behavioral Scientist* 36:639-650.

Maxwell, Christopher D.

- 1998 The specific deterrent effect of arrest on aggression between intimates and spouses. Ph.D. dissertation, Newark, N.J.: Rutgers, the State University of New Jersey.

Maxwell, Christopher D., Joel H. Ganer, and Jeffrey A. Fagan

- 2001 The Effects of Arrest on Intimate Partner Violence: New Evidence from the Spouse Assault Replication Program. Research In Brief. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

McCord, Joan

- 1992 Deterrence of domestic violence: A critical view of research. *Journal of Research in Crime and Delinquency* 29(2):229-239.

Menard, Scott

- 1995 Applied Logistic Regression Analysis. Quantitative Applications in the Social Sciences, 07-106. Michael S Lewis-Beck (ed.) Thousand Oaks, Calif.: Sage.

Mitchell, David B.

- 1992 Contemporary police practices in domestic violence cases: Arresting the abuser: Is it enough? *Journal of Criminal Law & Criminology* 83:241-249.

Moffitt, Terrie E., Robert F. Krueger, Avshalom Caspi, and Jeffrey Fagan

- 2000 Partner abuse and general crime: How are they the same? How are they different. *Criminology* 38:199-232.

Nagin, Daniel S. and David P. Farrington

- 1992 The onset and persistence of offending. *Criminology* 30:501-523.

National Institute of Justice

- 1985 Replicating an Experiment in Specific Deterrence: Alternative Police Responses to Spouse Assault. Washington, D.C.: National Institute of Justice.

- Pate, Anthony and Edwin E. Hamilton
 1992 Formal and informal deterrents to domestic violence. *American Sociological Review* 57:691-697.
- Paternoster, Raymond and Robert Brame
 1997 Multiple routes to delinquency? A test of developmental and general theories of crime. *Criminology* 35:49-84.
- Quigley, Brian M. and Kenneth E. Leonard
 1996 Desistance of husband aggression in the early years of marriage. *Violence and Victims* 11:355-370.
- Sampson, Robert J. and John H. Laub
 1993 Structural variations in juvenile court processing: Inequality, the underclass, and social control. *Law and Society Review* 27:285-311.
- Schmidt, Janell D. and Lawrence W. Sherman
 1993 Does arrest deter domestic violence? *American Behavioral Scientist* 36:601-610.
- Sechrest, Lee, Susan O. White, and Elizabeth Brown (eds.)
 1979 *The Rehabilitation of Criminal Offenders: Problems and Prospects*. Washington, D.C.: National Academy of Sciences.
- Sherman, Lawrence W.
 1980 Specific deterrent effect of spouse assault. Proposal submitted to the National Institute of Justice Crime Control Theory Program. U.S. Department of Justice, Washington, D.C.
 1992 *Policing Domestic Violence: Experiments and Dilemmas*. New York: Free Press.
- Sherman, Lawrence W. and R. A. Berk
 1984a *The Minneapolis Domestic Violence Experiment*. Police Foundation Reports, No. 1. Washington, D.C.
 1984b The specific deterrent effects of arrest for domestic assault. *American Sociological Review* 49:261-272.
- Sherman, Lawrence W. and Ellen G. Cohn
 1989 The impact of research on legal policy: The Minneapolis Domestic Violence Experiment. *Law and Society Review* 23:117-144.
- Sherman, Lawrence W., Douglas A. Smith, Janell D. Schmidt, and Dennis P. Rogan
 1992a Crime, punishment, and stake in conformity: Legal and informal control of domestic violence. *American Sociological Review* 57:680-690.
- Sherman, Lawrence W., Janell D. Schmidt, Dennis P. Rogan, Douglas A. Smith, Patrick Gartin, Ellen G. Cohen, Dean J. Collins, and Anthony R. Bacich
 1992b The variable effects of arrest on crime control: The Milwaukee Domestic Violence Experiment. *Journal of Criminal Law & Criminology* 83:137-169.
- Sherman, Lawrence W., Janell D. Schmidt, Dennis P. Rogan, Patrick R. Gartin, Dean J. Collins, Anthony Bacich, and Ellen G. Cohn
 1990 *The Milwaukee Domestic Violence Experiment. Final Report*. Washington, D.C.: Crime Control Institute.
- Stark, Evan
 1993 Mandatory arrest of batterers: A reply to its critics. *American Behavioral Scientist* 36:651-680.

- Steinman, Michael
1990 Lowering recidivism among men who batter women. *Journal of Police Science and Administration* 17:124-132.
- Sugarman, David B. and Sue Boney-McCoy
2000 Research synthesis in family violence: The art of reviewing the research. *Journal of Aggression, Maltreatment & Trauma* 4:55-82.
- Thistlethwaite, Amy, John Wooldredge, and David Gibbs
1998 Severity of disposition and domestic violence recidivism. *Crime & Delinquency* 3:388-398.
- U.S. Attorney General's Task Force on Family Violence
1984 Attorney General's Task Force on Family Violence. Final Report. Washington, D.C.: U.S. Government Printing Office.
- Violence Against Women Grants Office
1996 Grants to Encourage Arrest Polices Pograms. Proposed Regulations No. 28 CFR Part 90 [OJP No. 1019] RIN 1121-AA35. Washington, D.C.: Office for Justice Programs, U.S. Department of Justice. 13.
- Wilson, James Q.
1977 Forward. In Police Foundation (ed.), *Domestic Violence and the Police*. Washington D.C.: Police Foundation.
- Zimring, Franklin E.
1989 Toward a jurisprudence of family violence. In Lloyd Ohlin and Michael Tonry (eds.), *Family Violence, Vol. 11. Crime and Justice: A Review of Research*. Chicago, Ill.: University of Chicago Press.
- Zorza, Joan and Laurie Woods
1994 Analysis and Policy Implications of the New Police Domestic Violence Studies. New York: National Center on Women and Family Law.

Christopher D. Maxwell is Assistant Professor in the School of Criminal Justice at Michigan State University. His research interests include the social control and criminal justice processing of intimate partner violence, the efficacy of aggression and delinquency prevention programs, and the impact of social and ecological contexts on patterns of delinquency, crime, and criminal justice decision making. Dr. Maxwell is a graduate of Rutgers University (Ph.D., 98, MA, 94) and Indiana University-Bloomington (BA, 90).

Joel H. Garner is Director of Research at the Joint Centers for Justice Studies, Inc. He received his Ph.D. from The George Washington University in 1986. He was the government program manager for the Minneapolis Domestic Violence Experiment and the Spouse Assault Replication Program. His current research interests include police use of force, Federal firearm regulations, and organizational compliance with environmental laws.

Jeffrey Fagan is a Professor of Law and Public Health at Columbia University. For over two decades, his research and scholarship has focused on crime, law and social policy. He currently chairs the National Policy Committee of the American Society of Criminology, and serves on the Committee on Law and Justice of the National Research Council.