



National Institute of Justice

Research in Brief

July 2001

Issues and Findings

Discussed in this Brief: An analysis of 4,032 incidents in which males assaulted their female intimate partners, comparing the number of repeat offenses when batterers are and are not arrested. The data in this study were obtained from five jurisdictions included in the National Institute of Justice–sponsored Spouse Assault Replication Program. This multisite analysis was cosponsored by the National Institute of Justice and the Centers for Disease Control and Prevention.

Key issues: Analysis of 314 incidents in the 1984 Minneapolis Domestic Violence Experiment found that when the assaulter was arrested, statistically significant reductions in subsequent offending were reported both in victim interviews and in official police records. Replication experiments began in the early 1990s. Five jurisdictions that used a diverse set of incidents and a variety of outcome measures reported that the use of arrest was only occasionally associated with statistically significant reductions in subsequent repeat offending. The results of the experiments varied by measures used and by the jurisdiction studied.

Key findings: Using consistent definitions of eligible cases across all five jurisdictions, a consistent set of five measures of repeat offending and appropriate statistical analyses for the combination

continued...

The Effects of Arrest on Intimate Partner Violence: New Evidence From the Spouse Assault Replication Program

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

After nearly 20 years of research designed to test the effects of arrest on intimate partner violence, questions persist on whether arrest is more effective at reducing subsequent intimate partner violence than such informal, therapeutic methods as on-scene counseling or temporary separation. The most important research efforts addressing this question were six experiments known collectively as the National Institute of Justice's (NIJ's) Spouse Assault Replication Program (SARP).¹ These field experiments, carried out between 1981 and 1991 by six police departments and research teams, were designed to test empirically whether arrests deterred subsequent violence better than less formal alternatives.

In the first of the six studies, the Minneapolis Domestic Violence Experiment (MDVE), Sherman and Berk found that arresting batterers reduced by half the rate of subsequent offenses against the same victim within a 6-month followup period.² Subsequently, after five replication experiments were completed,³ Schmidt and Sherman conducted a qualitative synthesis of MDVE and the five

replications.⁴ They reported that in three studies, offenders assigned to the arrest group had higher levels of repeat offending (recidivism) and that in the other three studies, a statistically significant but modest reduction was found among batterers assigned to arrest. Thus, rather than providing results that were consistent with MDVE, the published results from the five replication experiments produced inconsistent findings about whether arrest deters intimate partner violence.

Because of the inconsistent and conditional findings generated by the five replication experiments, scholars interested in the validity of deterrence theories⁵ and policymakers working to reduce intimate partner violence have become less confident about relying on arrest as the primary response to violence between intimates.⁶ The development of a coherent evaluation of the effectiveness of arrest based on the five experiments with published results was complicated by the differences across the experimental sites in case selection, incident eligibility rules, statistical analysis, and outcome

Support for this research was provided through a transfer of funds to NIJ from the Centers for Disease Control and Prevention.

CDC

Issues and Findings

...continued

of data from a multisite experimental study, this research finds that—

- Arrest is associated with less repeat offending in all five measures of repeat offending.
- Reductions in repeat offending are larger and statistically significant in the two measures that are derived from interviews with victims.
- Reductions in repeat offending are smaller and not statistically significant in the three measures that are derived from official police records.
- The effectiveness of arrest does not vary by jurisdiction.
- The size of the reduction in repeat offending associated with arrest is modest compared with the effect of other factors (such as the batterer's age and prior criminal record) on the likelihood of repeat offending.
- Regardless of whether or not the batterer was arrested, more than half of the suspects committed no subsequent criminal offense against their original victim during the followup period.
- A minority of suspects continue to commit intimate partner violence regardless of whether they were arrested, counseled, or temporarily separated from their partner. Future research needs to focus on identifying such offenders and the policies and practices that will prevent their partners from being victimized further.

Target audience: Criminal justice and public health researchers and practitioners; police managers; advocates for victims of domestic violence; and legislators, policy-makers, and domestic violence intervention planners at all levels of government.

measurements. With these differences, prior attempts to synthesize and understand the substantive diversities among and within the experiments proved difficult. Thus, the full potential of SARP to answer questions about the specific deterrent effect of arrest and the safety of victims has not been realized.

We have previously reviewed and compared the published data from the five replication sites that had reported final results to NIJ by 1993.⁷ In concluding our review, we cautioned readers not to use our synthesized results as the final conclusion on whether arrest deters repeat spouse assault. We pointed out that the comparisons were based on information drawn from different outcome measures, analytical models, and case selection criteria. Furthermore, we asserted that the inconsistency between sources and measures across sites was not necessarily because of limitations in the experimental designs, but because the SARP design called for multiple data sources and measures that could capture variations in the nature of the deterrent effect. We argued that conclusions about the deterrent effect of arrest therefore should wait until a more careful statistical analysis was completed, one based on data pooled from all five sites and using standardized measures of intervention and outcome. This Research in Brief summarizes the findings of such a statistical analysis.

We studied the deterrent effect of arrest, using an approach that addressed many problems faced by prior efforts to synthesize the results from SARP. Supported by NIJ and the Centers for Disease Control and Prevention (CDC), the project pooled incidents from the five replication experiments, computed comparable independent and outcome measures from common data intentionally embedded in each experiment, and

standardized the experimental designs and statistical models. Using the increased power of the pooled data, this study provides a more consistent, more precise, and less ambiguous estimation of the impact of arrest on intimate partner violence. Key results of this study include the following:

- Arresting batterers was consistently related to reduced subsequent aggression against female intimate partners, although not all comparisons met the standard level of statistical significance.
- Regardless of the statistical significance, the overall size of the relationship between arrest and repeat offending (i.e., the deterrent effect of arrest) was modest when compared to the size of the relationship between recidivism and such measures as the batterers' prior criminal record or age.
- The size of the reduction in subsequent intimate partner aggression did not vary significantly across the five sites. In other words, the benefit of arrest was about equal in regards to reducing aggression in all five sites.
- Regardless of the type of intervention, most suspects had no subsequent criminal offense against their original victim within the followup period, and most interviewed victims did not report any subsequent victimization by their batterer.
- This research found no association between arresting the offender and an increased risk of subsequent aggression against women.

About SARP

Historical background. In 1968, the *New York Times Magazine* reported on an innovative program in which some New York City police officers were trained to

use psychology to handle family crisis calls. A subsequent report published by the National Institute of Law Enforcement and Criminal Justice asserted the value of this alternative over traditional law enforcement approaches to domestic violence.⁸ However, by the late 1970s, many law enforcement officials and domestic violence advocates started believing that this nonpunitive and therapeutic practice of responding to violence against women was ineffective. States enacted arrest laws and police departments implemented policies that authorized police officers to make an arrest even if they did not witness a domestic violence incident or saw no evidence of a felonious act. These changes, largely fueled by groups seeking better protection of victims through the sanction and control of offenders,⁹ enabled the criminal justice system to initiate a more forceful response to less severe family problems.

In 1980, the Police Foundation received permission from the Minneapolis Police Department to test the efficacy of actions its police officers could take when responding to a domestic dispute that did not require an arrest.¹⁰ (See “From the Minneapolis Domestic Violence Experiment to the Five Replications.”) Using two sources of data—official police records of new offenses and interviews with victims—and several statistical tests, the researchers reported that arrest reduced by nearly 50 percent the rate of subsequent assaults during the 6-month followup period.¹¹ These results were subsequently argued by several scholars as among the most influential results ever generated by social science.¹²

After considerable public discussion about the findings, NIJ announced in 1986 that it would fund a multisite replication of the Minneapolis experiment. Five NIJ-funded sites completed their experiments: Charlotte, North Carolina; Colorado Springs, Colorado; Dade County, Florida; Milwaukee, Wisconsin; and, Omaha, Nebraska. Between 1986 and 1990, NIJ provided about \$750,000 for each of these research projects; in addition, the local police departments contributed substantial resources to these efforts over a several-year period.

Methodological background. NIJ and CDC supported the project described in this Brief to develop a more reliable estimation of the overall deterrent effect of arrest on intimate partner violence. The project design called for pooling incident-level data from the five completed replication experiments, computing comparable outcome and control measures across the five replication sites, and providing consistent analytical models that tested the various intervention protocols across different measures.

While we faced many problems completing each task, one of the most difficult was deciding which outcome measures to use from a database containing more than 300 potential outcome measures. After assessing the literature on the nature of intimate partner violence, each site’s raw data, and the published results from the five sites, we decided to employ two outcome measures, one from the criminal history database and one from the victim interviews. We believe these two outcomes provide the best available indicators of the overall extent and breadth of aggression by the batterer

against the original victim during the followup period (6 months to 3 years).

In addition to physical assaults, the two aggression measures captured incidents involving damage of property owned by the victim or the common household. The aggression measure from the victim interviews also captured verbal threats of physical or property damage made by the batterer against the victim. We calculated several dimensions of aggression within each of the two data sources. For the outcome measure based on criminal history, we computed a dichotomous recidivism measure (yes or no) that captured any incident within the first 6 months after the experimental incident, a count of the number of days that elapsed between the experimental incident (i.e., the one that included the suspect in the study) and the first subsequent police-recorded incident, and a measure of the annualized frequency of reported incidents of aggression (i.e., the number of incidents, adjusted to represent a per-year rate). We also computed two outcome measures based on the victim interviews: (1) a dichotomous (yes or no) measure that captured aggression occurring between the experimental incident and the last victim interview (which typically covered at least a 6-month window) and (2) the frequency of aggression between the experimental incident and the last interview.

We also collapsed the assigned intervention categories into two groups: nonarrest and arrest. The nonarrest group contains all suspects who were randomly assigned to one of seven alternative interventions. Exhibit 1 provides descriptive information on the experimental implementation by

site, as well as on the interview completion rates and on the suspect and incident characteristics.

These and other methodological issues are discussed in greater detail in the sidebar “About the Sample.”

What victim interviews suggest about whether arrest deters subsequent aggression

Exhibit 2 presents results from our statistical analysis of the relationship

between arrest and several dimensions of intimate partner aggression. The first analysis (prevalence) uses victim interview data to test for the association between arrest and any subsequent aggression during the period between the experimental incident

A From the Minneapolis Domestic Violence Experiment to the Five Replications

A 1970 report published by the National Institute of Law Enforcement and Criminal Justice^a recommended training police officers to calm down domestic violence situations—separating the parties, listening to the concerns of each disputant, and attempting to address the immediate problem underlying the current dispute—and to provide the victim with phone numbers for a variety of social services. Arresting one or both parties was not part of this approach, which was touted as integrating the psychologist’s knowledge of human behavior with the coercive authority of the law in a manner that promoted collaboration among the police and other social service agencies.^b

By the early 1980s, the effectiveness of this nonpunitive approach was being questioned, and police departments began introducing policies that changed the ways their officers responded to domestic violence by switching from the therapeutic models to more formal, certain, and punitive responses. Rooted in the assumptions of specific deterrence and incapacitation, these changes emphasized expanding the police officers’ legal powers and codifying when arrests could and should be made.

When police departments were beginning to make changes, there was little systematic knowledge and little reason to believe that arrest or any sanction could act as a specific deterrent or improve the safety of the victims.

Questions arose about whether new criminal penalties or civil actions were the most appropriate remedies and whether police officers should replace their peacekeeping efforts with formal sanctions.

Amidst this uncertainty, James Q. Wilson^c recommended that police departments systematically experiment with different methods of “reducing the chance that a dispute will lead to an assault and an assault to a homicide” within intimate settings.^d This suggestion led to the Minneapolis Domestic Violence Experiment (MDVE), the results of which were published in 1984. The study design called for officers in the Minneapolis Police Department (MPD) to carry out one of three responses when they had probable cause to believe a misdemeanor assault had occurred between cohabitants or spouses: (1) arrest the suspect, (2) order one party out of the residence, or (3) advise the couple on how to solve their problems at the scene.

The strength of MDVE was that the selection of a particular response in a particular incident of domestic violence was determined by an experimental design. This design made it easier to determine if differences in police responses were responsible for any differences in subsequent reoffending by the suspect.

In MDVE, researchers at the Police Foundation collected information on subsequent offenses for a period of 6 months from both official police records and from interviews with victims. Using data from 314 incidents, the researchers reported that when the suspect was arrested, there were statistically significant reductions in reoffending in the official records of all the cases and in the cases with victim interviews. Based on these results, the authors of MDVE recommended policies authorizing the use of arrest in misdemeanor domestic violence offenses.

A 1989 survey of local police departments concluded that the published results of MDVE may have substantially influenced over one-third of the police departments responding to their survey to adopt a proarrest policy.^e At the national level, the 1984 Attorney General’s Task Force on Family Violence, citing the MDVE results, recommended that “chief executives of every law enforcement agency establish arrest as the preferred response in incidents of family violence.”^f The MDVE results were published in the *New York Times* and in hundreds of other newspapers in the United States; three television networks reported the results during prime-time news programs; and numerous editorials and nationally syndicated columnists featured the study and its findings.^g

Support for replication of MDVE was widespread. The original authors urged

and the last time the victim was interviewed. This model estimated that if their batterers were arrested, about 25 percent fewer female victims than expected reported one or more incidents of aggression. In other words, when the likelihood of failure (reoffending)

is estimated for the typical case, about 36 percent of suspects in the arrest group reoffended, compared with 48 percent of suspects in the nonarrest group. This difference was statistically significant while controlling for differences among sites, the

length of time the researchers tracked the victims, and characteristics of the suspect and incident. When examining the rates or frequency of aggression, we again found a statistically significant reduction in subsequent aggression that is related to arrest. On

From the Minneapolis Domestic Violence Experiment to the Five Replications (continued)

replicationⁿ and early praise for the study's design among criminological scholars was tempered by a preference for replication.^l The Department of Justice task force recommending the adoption of a pro-arrest policy nationwide also recommended replication of the Minneapolis experiment.^l A multisite replication of the Minneapolis experiment was chosen because a single-site approach would provide only one additional data point, only slightly improving the generalizability of the Minneapolis findings.^k

In designing a program of replications, the National Institute of Justice (NIJ) required that each study involve (1) experimental comparisons of arrest and alternative police responses to misdemeanor spouse assault incidents and (2) measurements of victim safety using both official police records and victim interviews.^l Other aspects of the design were left to the preferences of the local teams of researchers and implementing police agencies. Seventeen law enforcement agencies competed to be part of the replication program even though this program did not provide additional financial resources to the department or to participating officers. The replication effort was research, not a demonstration program, and there were no Federal subsidies to the participating departments.

The characteristics of the five sites, the organizational structure of the research projects, and the requirements of the

solicitation led to similar studies, but not to exact replications of each other or MDVE. For example, each new study devised experimental designs that required the officers to report whether an incident was eligible for the study before they were told what the assigned treatment would be. The new studies included more cases than the Minneapolis study and some of them broadened the study eligibility to include female offenders, same-sex couples, and harassment offenses. Omaha conducted victim interviews at 6 and 12 months. Dade County used the same cases to conduct a second experiment studying police officer followup with victims after the initial incident. Milwaukee varied the number of hours arrested offenders would be detained in the jail. NIJ encouraged these innovations and variations among sites, but it also required that each site document the characteristics of victims, offenders, and police behavior so that common analyses using consistent eligibility requirements and outcome measures could be conducted.

a. Bard, Morton, *Training Police as Specialists in Family Crisis Intervention*, Washington, DC: U.S. Department of Justice, National Institute of Law Enforcement and Criminal Justice, 1970.

b. Garner, Joel H., and Christopher D. Maxwell, "What Are the Lessons of the Police Arrest Studies?" in *Program Evaluation and Family Violence Research*, ed. S.K. Ward and D. Finkelhor, Binghamton, NY: Haworth Publishers, 2000.

c. Sherman, Lawrence W., "The Influence of Criminology on Criminal Law: Evaluating Arrests for Misdemeanor Domestic Violence," *Journal of Criminal Law and Criminology* 83 (1) (Spring 1992): 1–45.

d. Wilson, James Q., "Foreword," in *Domestic Violence and the Police*, Washington, DC: Police Foundation, 1977: iii–vi.

e. Sherman and Cohen, "The Impact of Research on Legal Policy," 1989 (see note 11).

f. U.S. Attorney General's Task Force on Family Violence, *Final Report*, Washington, DC: U.S. Department of Justice, 1984.

g. Sherman, Lawrence W., and Ellen G. Cohen, "The Impact of Research on Legal Policy: The Minneapolis Domestic Violence Experiments," *Law & Society Review* 23 (1) (1989): 117–144.

h. Sherman, L.W., and R.A. Berk, "The Specific Deterrent Effects of Arrest for Domestic Assault," *American Sociological Review* 49 (1984b): 261–272.

i. Boffey, P.M., "Domestic Violence: Study Favors Arrest," *New York Times*, April 5, 1983; and Lempert, R., "From the Editor," *Law & Society Review* 18 (4) (1984): 505–513.

j. U.S. Attorney General's Task Force on Family Violence, *Final Report* (see note f).

k. Garner, Joel H., "Two Three ... Many Experiments: The Use and Meaning of Replication in Social Science Research," paper presented at the Annual Meeting of the American Society of Criminology, Baltimore, November 1990.

l. National Institute of Justice, *Replicating an Experiment in Specific Deterrence: Alternative Police Responses to Spouse Assault*. Washington, DC: U.S. Department of Justice, National Institute of Justice, 1985.

average, female victims whose batterers were arrested reported about 30 percent fewer incidences of subsequent aggression than expected over the followup period. Thus, we found a sizable reduction in subsequent

aggression reported by victims whose batterers were assigned to the arrest group. However, because these results are based on a subsample of interviewed victims, rather than on the entire sample of eligible cases, the

results from the victim interviews alone should be used with some caution because victims not interviewed may have been involved with suspects who responded differently to their intervention.

Exhibit 1. Experimental interventions, interview completion rates, and suspect and incident characteristics, by site

	Charlotte (N = 638)		Colorado Springs (N = 1,238)		Dade County (N = 906)		Milwaukee (N = 954)		Omaha (N = 296)		Total (N = 4,032)	
	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number
Incident Assignment Data												
Assigned to arrest	33	212	26	325	51	463	68	648	34	100	43	1,748
Actually arrested	40	256	28	349	60	546	68	648	34	100	47	1,898
Misassignment rate	13	84	7	86	10	91	0	0	3	8	3	270
Victim Interview Completion Rates												
Initial interview	64	411	83	1,026	65	593	60	574	79	235	70	2,839
Final interview	50	320	70	872	42	384	78	743	73	216	63	2,535
Suspect Characteristics*												
Employed	77	492	87	1,073	71	639	47	450	78	231	72	2,885
Prior arrest	31	196	43	530	12	110	62	590	65	192	40	1,618
Use of intoxicant	54	346	59	734	31	277	29	280	59	175	45	1,812
Race/ethnicity												
African-American	70	446	30	374	42	379	75	719	42	124	51	2,042
White	28	177	54	671	36	327	20	190	51	149	38	1,514
Hispanic	0	2	15	180	22	200	4	40	5	14	11	436
Asian/other	2	13	1	13	0	0	1	5	3	9	1	40
Relationship with victim												
Married	48	309	67	826	79	713	31	295	46	137	57	2,280
Separated	2	10	4	53	3	26	0	3	0	0	2	92
Divorced	0	2	1	8	2	18	1	7	1	4	1	39
Current or past intimate	50	317	28	351	16	149	68	649	52	155	40	1,621
Incident Characteristics												
Misdemeanor assault	97	621	37	464	100	906	100	954	100	296	80	3,241
Victim injured	84	534	55	683	92	830	100	951	73	215	80	3,213

Notes: All test values are from χ -square tests, except for the age comparison, which is an F-test. For all data, $p < .001$. Reported percent and counts reflect number after adjusting for missing information. N = Number of eligible incidents.

*The mean age of suspects was 32 total; 31 in Colorado Springs, Milwaukee, and Omaha; 33 in Charlotte; and 35 in Dade County.

What other factors are related to aggression?

Besides the consistent deterrent relationship between arrest and aggression, other factors were consistently related to aggression, but some factors were not. First, compared with the Omaha victims, a significantly smaller percentage of victims from the other sites (except Milwaukee) reported one or more victimizations by the suspect. On average, victims from these three sites also reported less frequent victimization. These differences in the base rates of aggression across the sites, however, did not translate into significantly different relationships between arrest and aggression in the different sites. In other words, the reduction we find in aggression reported by victims whose batterers were assigned an arrest is of about equal size in each site.

In addition to the comparisons we made across the sites, we looked for differences in aggression reported by the victims across several suspect characteristics. These comparisons found that the suspect’s age and race were consistently and significantly related to the frequency of subsequent aggression as reported by the victims. These victims reported significantly less aggression when the suspect was older and nonwhite. The suspects’ prior arrest records and their marital status with the victim were also consistently related to aggression, but only the prior record was significant in all but one of the analyses. Finally, several other suspect characteristics, such as employment and the use of intoxicants, were inconsistent in the direction of their relationship across the two dimensions of aggression (prevalence and frequency). For

example, about 2 percent more victims of employed suspects reported one or more incidents of aggression, though these same victims simultaneously reported about 21 percent fewer incidents of aggression over the followup period.

What official records suggest about whether arrest deters subsequent aggression

We next examined data collected by police departments to measure aggression by the suspect against the victim. The approach to testing whether arrest was related to officially recorded aggression follows the approach to the victim interviews, except we added a statistical analysis

that examined the timing of the first new aggressive incident. Overall, the results based on the police data regarding the effectiveness of arrest are consistent in direction with those based on the victim interview data: A consistent deterrent relationship exists between arrest of the suspect and later aggression while controlling for the differences across the sites, the victim interview process, and suspect characteristics (see exhibit 2). However, the police data show a far smaller reduction in aggression because of the arrest treatment than what was detected using victim interview data, and none of these relationships reached the traditional level of statistical significance. Specifically,

Exhibit 2. Deterrent effects of arrest on aggression against a female intimate partner

	Victim interviews		Official records		
	Prevalence	Frequency	Prevalence	Rate	Time-to-failure
Arrest	0.75*	0.70*	0.96	0.92	0.90
Site (compared with Omaha)					
Charlotte	0.66*	0.42*	1.78*	1.82*	2.21*
Colorado Springs	0.25*	0.23*	2.15*	1.86*	2.19*
Dade County	0.40*	0.37*	3.21*	3.44*	3.17*
Milwaukee	1.06	0.87	2.12*	3.56*	2.62*
Interview Exposure	2.36*	0.58*	1.71*	1.09	1.30*
Suspect Characteristics					
Age	0.68*	0.44*	0.61*	0.72*	0.70*
Use of intoxicant	0.99	1.43*	1.23*	1.10	1.10
White	1.26*	1.35*	0.68*	0.70*	0.70*
Married	1.05	1.11	0.99	1.10	0.90
Prior arrest	1.56*	1.12	2.50*	2.36*	3.30*
Employed	1.02	0.79*	0.94	0.85*	0.90

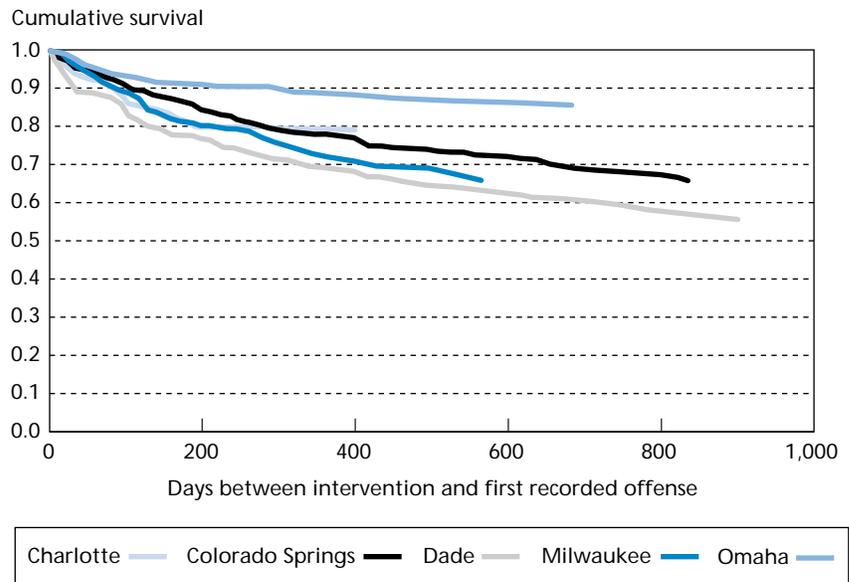
Note: Cell coefficients present odd ratios. Coefficients less than 1.0 indicate a decrease in the expected level of aggression due to treatment or other controlled measure; values larger than 1.0 indicate increases in aggression.

* $p < .05$

in the first analysis (prevalence), we found about 4 percent fewer than the expected percentage of male suspects in the arrest group with one or more incidents of subsequent aggression during the first 6 months of followup. The second analysis, which tested for the relationship between the intervention and the annual rate of aggression, found a reduction of about 8 percent from the expected number of incidents per year for suspects assigned to the arrest group. Finally, the last analysis, which examined the relationship between arrest and the timing of the first new incident, found that the expected risk of a new incident on any given day after arrest or nonarrest is reduced nearly 10 percent among the arrested suspects. Thus, depending on the dimension of the outcome, the average amount of reported aggression by the suspects dropped by between 4 and 10 percent if they were assigned to the arrest group.

Focusing more closely on the timing of the first subsequent incident of aggression, exhibits 3 and 4 display two “survival” graphs. Exhibit 3 displays, by site, the proportion of suspects with no officially recorded aggression against their intimate partner beyond a specified time (i.e., cumulative survival). The average survival rate throughout the followup period varied substantially by site. On the high end was Omaha, where nearly 90 percent of the suspects had not reoffended by the end of their observation period. On the low side was Dade County, where that figure (the cumulative survival rate) was slightly less than 60 percent. These differences between sites, however, did not result in differences in survival rates by intervention group when the five sites were pooled together. Exhibit 4 shows that throughout the followup period, which for some

Exhibit 3. Time to first police-recorded failure, by site

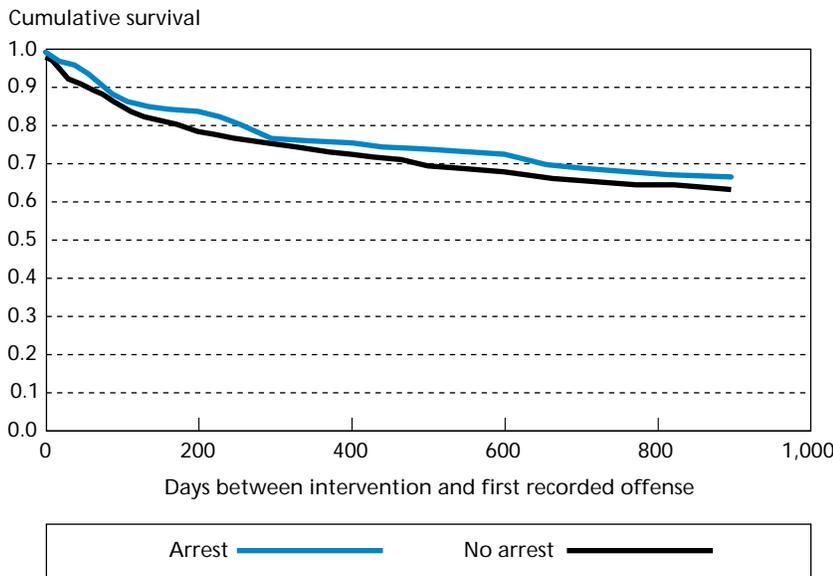


suspects lasted nearly 3 years, batterers who were assigned an arrest had a consistently greater rate of survival (nonoffending) than did those assigned an informal intervention.

This consistent, but small, difference in the survival rate by intervention is important because earlier analysis using data from Milwaukee suggested that arrest may have a significant long-term criminogenic effect.¹³ Our more detailed statistical analysis supports the visual evidence presented in these exhibits. During no particular observation period were the suspects assigned to an arrest more likely to batter their intimate partner than those in the control (nonarrest) group. Thus, among this larger sample of male intimate partner abusers, the survival rate for aggression among those assigned an arrest was never less than that of the control group, as earlier statistical analysis in one site had suggested.

Our statistical analysis also showed that the suspects’ age, race, employment status, and use of intoxicants at the time of the experimental incident were consistently and significantly related to subsequent aggression against the victim. Contrary to what we found with the victim interviews, white and employed suspects had lower levels of repeat offending according to the police records. Furthermore, suspects who were intoxicated at the time of the experimental incident and those with prior arrests for any crime had, on average, a greater likelihood of aggression recorded by the police. Only the measure of the suspect’s marital status with the victim was not consistently or significantly related to aggression. Similar to what we found with the victim interview data, marriage did not appear to provide notable protection against subsequent levels of aggression. Finally, we found that the longer the researchers were able to track the

Exhibit 4. Time to first police-recorded failure, by assigned treatment



victims for followup interviews, the more initial failures were reported to the police.

In addition to our findings about the relationship between arrest and aggression, we observed some patterns in the pooled data. First, we found a general pattern of cessation or termination of aggression that was only moderately related to the suspects’ assigned intervention. According to officially recorded data, less than 30 percent of the suspects, arrested or not, aggressed against the same victim during the followup period. Furthermore, only about 40 percent of the interviewed victims reported subsequent victimization of any measured type by the suspects. Other studies that specifically estimated the rate of desistance from intimate violence have also found similar rates over a 1- to 2-year period.¹⁴

A second pattern concerns the high concentration of repeat aggression

among a small number of batterers. During the 6-month followup, the 3,147 interviewed victims reported more than 9,000 incidents of aggression by the suspects since the initial incident. While most victims reported no new incidents of aggression, about 8 percent of them reported a total number of incidents that represented more than 82 percent of the 9,000 incidents. The same 8 percent also accounted for 28 percent of the 1,387 incidents recorded by the police that involved an interviewed victim.

Conclusion and policy implications

In 1998, the National Academy of Sciences report *Violence in Families* concluded that “arrest in all misdemeanor cases will not on average produce a discernable effect on recidivism.”¹⁵ Our early substantive assessment of the published reports was similar to their conclusion, but we

also argued that there was insufficient evidence in the site-specific and multisite publications to assess the effectiveness of arrest as a deterrent to spouse assault.¹⁶ Our multisite pooled analysis of the five replication experiments found good evidence of a consistent and direct, though modest, deterrent effect of arrest on aggression by males against their female intimate partners. The victim interviews indicate that the arrest of the suspect and any subsequent confinement, when compared with the alternative interventions collectively (see “About the Sample”), significantly reduced the expected frequency of subsequent aggression by 30 percent. Similarly, arrest may have reduced by a smaller amount the number of times the police responded to subsequent domestic violence incidents involving the same victim and suspect and may have extended the time between the initial incident and the first subsequent incident.

Our conclusion of a direct deterrent effect from arrest contradicts at least one assessment of findings from the original SARP publications. Berk, for example, argued that “the current balance of scientific evidence from the particular sites studied suggests that although arrest is not superior to a variety of other criminal justice interventions, one can on average do no better.”¹⁷ There are, however, various reasons the current statistical analysis should be preferred over prior individual and multisite analyses of SARP experiments for the following reasons:

- The consistent use of eligibility criteria across sites—we include only male offenders and female victims of intimate violence.

S About the Sample

Sample of suspects. The design of the Spouse Assault Replication Program (SARP) allowed each site to vary the eligibility of cases for its experiment. For example, for most of its implementation, the Dade County experiment only included married couples. The Milwaukee experiment included same sex couples and violent disputes between siblings. Several sites excluded incidents if the suspect had been included in a prior experimental case. Other sites included these repeat suspects.^a

Prior site-specific and multisite analyses of SARP data have not addressed these differences, but this study does. In this analysis, we include only cases where a male suspect committed violence against a female victim. Because some of the sites excluded cases with repeat suspects, we exclude all cases ($N = 248$) with repeat suspects from all sites to create a more consistent sample.

Overall, cases were most likely excluded because of the victim's or suspect's gender or because the suspect had appeared

previously in the experiment as a suspect (see exhibit A).

The site with the greatest percentage of cases excluded was Colorado Springs (25 percent); the exclusions were due mainly to the site's inclusion of some repeat suspects, male victims, female suspects, and incidents not involving an assault of the victim. Dade County had the fewest cases excluded (1 percent). The final count of incidents removed from our analysis was 760 (16 percent), which left 4,032 unique suspects for the cross-site

Exhibit A. Failure to meet eligibility criteria, by site

Eligibility criteria	Charlotte ($N = 686$)		Colorado Springs ($N = 1,660$)		Dade County ($N = 916$)		Milwaukee ($N = 1,200$)		Omaha ($N = 330$)		Total ($N = 4,792$)	
	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number
Relationship not spouselike	0.0	0	0.5	9	0.0	0	4.9	59	5.2	17	1.8	85
Suspect involved as a suspect in prior	5.2	36	7.6	126	0.7	6	6.7	80	0.0	0	5.2	248
Suspect not male	0.0	0	11.1	185	0.0	0	9.0	108	3.9	13	6.4	306
Victim not female	0.0	0	11.0	183	0.0	0	9.5	114	5.2	17	6.6	314
Victim and suspect same sex	0.0	0	0.1	2	0.0	0	4.0	48	1.8	6	1.2	56
Experimental incident did not involve an assault or victim injury	1.1	7	2.0	34	0.0	0	0.0	0	0.0	0	0.9	41
Victim under 18 years of age	0.6	4	0.3	5	0.1	1	0.5	6	0.3	1	0.4	17
Suspect under 18 years of age	0.2	1	0.1	1	0.0	0	0.0	0	0.0	0	0.0	2
Violated other site eligibility criterion	0.0	0	5.4	90	1.0	9	0.0	0	0.0	0	2.1	99
Missing treatment assignment code	0.0	0	0.1	2	0.0	0	0.0	0	0.0	0	0.0	2
Suspects eligible for multisite analysis	93.0	638	74.6	1,238	98.9	906	79.5	954	89.7	296	84.1	4,032

Note: Because some cases violated more than one eligibility rule, the percent columns will not necessarily add to 100%.

About the Sample (continued)

analysis. We think these selected suspects are representative of the majority of offenders reported to the police and represent cases of intimate partner violence that are likely to interest policy-makers and victim advocates. Mainly, the cases represent violence by an adult male suspect against a past or current female intimate partner.

Intervention comparisons. Our pooled analysis capitalizes on the features of the experimental design implemented within each site by using the suspects' assigned intervention, rather than the action actually taken by the police. This method preserves the integrity of the random assignment. The nonarrest group included such interventions as mediation counseling, a citation to appear in court, an order to leave the scene, a restraining order, or a warning about a future arrest if the officers were called back. The arrested group comprises only those suspects from each site who were randomly assigned to be arrested.^b

The sites differed in the percentages of suspects assigned to the arrest group and in the misassignment rates. These differences in the percentage of suspects assigned an arrest were expected, given the different number and types of interventions in each site. In Dade County, for example, only two interventions were each assigned one-half of the incidents; in Colorado Springs, suspects were assigned to one of four groups in equal proportions. Thus, in Colorado Springs, only one-fourth (26 percent) of the suspects were assigned to arrest and the other three-fourths to informal intervention. Overall, 43 percent of suspects were assigned to the arrest group.

The five sites differed in their rate at which suspects received an intervention different from the one they were

randomly assigned. The misapplication of intervention occurred when the police officers, after receiving the assigned intervention code from dispatch or the researcher team, chose to take a different action. In MDVE and the five SARP experiments, police officers could change the intervention while on the scene if one or more specific circumstances arose (such as the suspect assaulting or threatening the officers or the suspect assaulting or significantly threatening the welfare of the victim in the presence of the officers). Overall, misdelivery across the five sites occurred in less than 7 percent of the 4,032 cases, and in 90 percent of the misassigned cases, an arrest was made when nonarrest had been assigned.^c

Victim interviews. The rate at which initial and followup victim interviews were obtained varied within and between sites. For example, the completion rates of the final interview fluctuated from a high of nearly 80 percent in Milwaukee to a low of 42 percent in Dade. Nevertheless, in our sample, about 70 percent of the victims were interviewed during the initial followup period, and 63 percent were interviewed again after about 6 months. Although these completion rates are quite good considering the challenges of interviewing victims of intimate personal violence (such as locating them after 6 months), the less-than-100-percent completion rate poses difficulties for the cross-site analysis. Some researchers have suggested that the interviewing process itself could increase the likelihood of officially recorded failures and bias the estimation of the deterrent effect of arrest that is based on victim interview data.^d Based on this concern, our analysis includes an independent measure that partially captures the effects of the interviewing process. This measure accounts for the

variations across the victims in the length of time between the experimental incident and the last time they were interviewed, if at all.^e

Suspect characteristics. Substantial differences in suspect characteristics were found across the five sites for all demographic measures. Some notable differences include the large percentage of African-American suspects in Charlotte and Milwaukee but not in the other sites; the high percentage of married suspects in Colorado Springs and Dade County but not in the other sites; the small percentage of suspects in Dade County without a prior arrest; and the high rate of unemployment among suspects from Milwaukee. This substantial cross-site heterogeneity should be considered an asset, because it increases the overall generalizability of our findings. However, these variations also reinforce the need to control for the suspects' characteristics as factors in the experimental model; in previous criminal justice experiments, intervention effects may not have been found due to the large uncontrolled differences across the subjects among those deemed eligible for the experiment.^f

To assess the characteristics of the suspects in the two intervention groups, exhibit B provides the percentage of suspects assigned to and delivered an arrest, by categories of the suspects' demographic characteristics. The table also presents the proportions of suspects not receiving the assigned intervention, by their demographic characteristics. The characteristics of the suspects assigned to arrest versus the nonarrest control group and of those arrested versus those not arrested differed across several suspect characteristics. Among suspects assigned an arrest, more than expected were using drugs or alcohol at the time

About the Sample (continued)

of the incident; were not married, white, or employed; and had prior arrests. Similar differences were also found among those actually arrested when they were compared with those suspects not arrested, except for the increase in the percentage arrested among those using alcohol or drugs. This change in the assigned and delivered arrest rates among suspects who used alcohol or drugs is further confirmed when comparing suspects correctly assigned with those receiving something other than their assigned

treatment. Specifically, nearly 9 percent of those using intoxicants were misassigned, compared with 5 percent of those not using intoxicants.

The significant differences displayed in exhibit B, however, do not necessarily mean that the experimental protocols were systematically violated by the police officers. Rather, they suggest that when studies randomly assign different proportions of their cases to arrest, and when the demographically diverse sites were combined, the characteristics of suspects

in the arrest and nonarrest categories differ in many substantively important ways (more high-risk suspects in the arrest group than in the nonarrest group). This issue needs addressing in the outcome analysis. Therefore, properly controlling for these characteristics in the regression models is necessary to ensure that the intervention comparisons are fair. Prior single-site and multiple-site analyses of the SARP experiments have not controlled for variations in the timing of victim interviews, variations in misapplication rates, or demographic differences between

Exhibit B. Experimental intervention assigned, delivered, and misdelivered

	Assigned to arrest			Arrest delivered			Intervention misdelivered		
	Percent	Number	χ^2	Percent	Number	χ^2	Percent	Number	χ^2
Base rate	43.4	1,748		47.2	1,900		2.8	270	
Site			442***			436***			136***
Charlotte	33.2	212		40.1	256		13.2	84	
Colorado Springs	26.3	325		28.3	349		6.9	86	
Dade County	51.1	463		60.3	546		10.0	91	
Milwaukee	67.9	648		67.9	648		0.0	0	
Omaha	33.8	100		33.8	100		2.7	8	
Age			4.5			10.9***			13.8**
18 to 24	40.3	303		42.5	319		4.5	34	
25 to 28	42.9	361		45.8	385		5.9	50	
29 to 31	44.1	281		48.7	309		8.8	56	
32 to 37	44.0	400		49.4	449		8.0	73	
38 to 82	45.2	403		49.1	437		6.3	56	
Use of intoxicant			43.8***			10.7***			25.9***
No	48.0	1,066		49.5	1,098		4.9	108	
Yes	37.6	682		44.3	801		8.9	161	
Race			25.1***			17.4***			1.1
Nonwhite	46.4	1,168		49.7	1,250		6.4	160	
White	38.3	580		43.0	649		7.2	109	
Marital status			17.0***			11.9***			3.6
Nonmarried	47.2	784		50.4	836		5.8	96	
Married	40.7	964		44.9	1,063		7.3	173	
Prior arrest			22.1***			11.2***			5.4*
No	40.3	974		45.0	1,085		7.4	179	
Yes	47.8	774		50.4	814		5.6	90	
Employed			20.2***			25.6***			0.1
No	48.9	561		53.5	613		6.5	74	
Yes	41.1	1,187		44.7	1,286		6.8	195	

Note: $N = 4,032$

* $p < .05$; ** $p < .01$; *** $p < .001$

About the Sample (continued)

treatment groups that occurred when data from multiple sites are combined.

a. For a complete list of the eligibility rules, see Maxwell, Christopher D., *The Specific Deterrent of Arrest on Aggression between Intimates and Spouses* [diss.], Newark, NJ: Rutgers, the State University of New Jersey, 1998.

b. The comparison between arrest and nonarrest groups, instead of among all the treatment groups, was suggested by Binder and Meeker in their critique of the MDVE (Arnold Binder and James W. Meeker, "Experiments as Reforms," *Journal of Criminal Justice* 16 (4) (1988): 347–58). They argued that the best test of deterrence theory is one that compares

those who were punished with those who were not, rather than individually comparing the differences between arrest and each of the informal interventions. Besides testing for differences between arrest and nonarrest, our regression analysis controlled for differences in the average level of aggression in each site, the effect that the victim interviews may have had on the outcome, and several suspect and incident characteristics.

c. Although this is a fairly low rate of misassignment, the fact that it is not zero leaves the possibility that any effect of arrest is potentially biased upward or downward. See Richard A. Berk, Gordon K. Smyth, and Lawrence W. Sherman, "When Random Assignment Fails: Some Lessons from the Minneapolis Spouse Abuse Experiment,"

Journal of Quantitative Criminology 4 (3) (1988): 209–33.

d. Berk, Richard A., and Lawrence W. Sherman, "Data Collection Strategies in the Minneapolis Domestic Violence Experiment," in *Collecting Evaluation Data: Problems and Solutions*, ed. Leigh Burstein, et al., Beverly Hills, CA: Sage Publications, 1985: 35–48.

e. See Maxwell, 1998 (note a), for further information on how the interviewing process was addressed in the outcome analysis.

f. Weisburd, David, "Design Sensitivity in Criminal Justice Experiments," in vol. 17 of *Crime and Justice: A Review of Research*, ed. Michael Tonry, Chicago: University of Chicago Press, 1993: 337–80.

- The use of a consistent measure of repeat offending across all sites.
- The use of additional statistical controls for site and suspect differences between arrest and nonarrest groups.
- The use of longer followup periods and statistical controls for variability in followup periods.
- Increased statistical power from pooling cases from five sites.
- The consistent comparison of arrest with all other treatments combined.

The findings of this research have several implications for policy. First, our findings provide systematic evidence supporting the argument that arresting male batterers may, independent of other criminal justice sanctions and individual processes, reduce subsequent intimate partner violence. The size and statistical significance of the effect of arrest varied depending on whether the subsequent aggression

was measured by victim interviews or police records; even so, in all measures (prevalence, frequency, rate, and time-to-failure), arrest was associated with fewer incidents of subsequent intimate partner aggression. This finding exists during the first several days after the experimental incident regardless of the period of detention, as well as beyond 1 year. The arrested suspects were detained an average of 9 days, but the reduction in aggression associated with arrest did not vary by the length of the suspect's detention. Thus, our research finds no empirical support for the argument that arrest may eventually increase the risk for violence against women.

Second, our research showed that a minority of suspects continued to commit intimate partner violence, regardless of the intervention they received. While arrest reduced the proportion of suspects who reoffended and the frequency with which they reoffended, arrest did not prevent all batterers

from continuing their violence against their intimate partners. In fact, we found a small number of victims who have chronically aggressive intimate partners. Future research needs to build on preliminary efforts to accurately predict high-rate repeat offenders and to find methods of helping their victims before they are victimized further.

Third, our research showed that a majority of suspects discontinued their aggressive behaviors even without an arrest. This suggests that policies requiring arrest for all suspects may unnecessarily take a community's resources away from identifying and responding to the worst offenders and victims most at risk. Our research has documented the size of the specific deterrent effects of arrest, which, although consistent across sites and time, appeared modest compared with the overall percentage of suspects desisting from intimate partner violence. Although there may be other

benefits from policies requiring arrest that this research has not measured (including general deterrence), there are also likely costs of using arrests every time the police respond to an incident of intimate partner violence. Future research in this area needs to assess the benefits and costs of arresting all suspects before there can be a systematic conclusion of preferred or mandatory arrest policies.

Finally, it is unlikely that any single study can provide definitive answers to scientific questions or policy debates. Rather, a program of rigorous research involving many studies over time and place is necessary to provide sound bases for generating knowledge and improving policy.

Notes

- Garner, Joel, Jeffrey A. Fagan, and Christopher D. Maxwell, "Published Findings from the Spouse Assault Replication Program: A Critical Review," *Journal of Quantitative Criminology* 11 (1) (1995): 3–28; and Sherman, Lawrence W., "The Influence of Criminology on Criminal Law: Evaluating Arrests for Misdemeanor Domestic Violence," *Journal of Criminal Law and Criminology* 83 (1) (Spring 1992): 1–45.
- Sherman, Lawrence W., and Richard A. Berk, "The Specific Deterrent Effects of Arrest for Domestic Assault," *American Sociological Review* 49 (1) (1984): 261–72.
- A sixth site was funded, but the experiment was never completed.
- Schmidt, Janell D., and Lawrence W. Sherman, "Does Arrest Deter Domestic Violence?" *American Behavioral Scientist* 36 (5) (May/June 1993): 601–10.
- Berk, Richard A., "What the Scientific Evidence Shows: On the Average, We Can Do No Better Than Arrest," in *Current Controversies on Family Violence*, ed. Richard J. Gelles and Donileen R. Loseke, Newbury Park, CA: Sage Publications, Inc., 1993: 323–36; Fagan, Jeffrey A., "The Criminalization of Domestic Violence: Promises and Limits," paper presented at the Conference on Criminal Justice Research and Evaluation, Washington, DC, January 1996; and Sherman, "The Influence of Criminology on Criminal Law," 1992 (see note 1).
- Chalk, R.A., and P.A. King, eds., *Violence in Families: Assessing Prevention and Treatment Programs*, Washington, DC: National Academy of Science, Committee on the Assessment of Family Violence Intervention, Board on Children, Youth, and Families, National Research Council and Institute of Medicine, 1998; Clark, Jacob R., "Where to Now on Domestic Violence? Studies Offer Mixed Policy Guidance," *Law Enforcement News* 30 (April 1993): 1, 17; Frisch, Lisa A., "Research That Succeeds, Policies That Fail," *Journal of Criminal Law and Criminology* 83 (1) (Spring 1992): 209–16; Lerman, Lisa G., "The Decontextualization of Domestic Violence," *Journal of Criminal Law and Criminology* 83 (1) (Spring 1992): 217–40; and Mitchell, David B., "Contemporary Police Practices in Domestic Violence Cases—Arresting the Abuser: Is It Enough?" *Journal of Criminal Law and Criminology* 83 (1) (Spring 1992): 241–49.
- Garner, Fagan, and Maxwell, "Published Findings from the Spouse Assault Replication Program," 1995 (see note 1).
- Bard, M., and J. Zacker, "The Prevention of Family Violence: Dilemmas of Community Intervention," *Journal of Marriage and the Family* 33 (1971): 677–82; and Bard, Morton, *Training Police as Specialists in Family Crisis Intervention*, Washington, DC: U.S. Department of Justice, National Institute of Law Enforcement and Criminal Justice, 1970.
- Fagan, Jeffrey A., "Contributions of Family Violence Research to Criminal Justice Policy on Wife Assault: Paradigms of Science and Social Control," *Violence and Victims* 3 (3) (1988): 159–86.
- Bouza, Anthony V., *The Police Mystique: An Insider's Look at Cops, Crime, and the Criminal Justice System*, New York: Plenum Press, 1990.
- Sherman, L.W. and R.A. Berk, *The Minneapolis Domestic Violence Experiment*, Police Foundation Reports, no. 1, Washington, D.C.: Police Foundation, 1984.
- Fagan, Jeffery A., and Angela Browne, "Violence Against Spouses and Intimates," in *Understanding and Controlling Violence*, ed. A.J. Reiss, Jr., and J.A. Roth, vol. 3. Washington, DC: National Academy Press, 1994; Gelles, Richard J., "Constraints Against Family Violence: How Well Do They Work?" *American Behavioral Scientist* 36 (5) (1993): 575–86; Lempert, Richard, "Humility is a Virtue: On the Publication of Policy Relevant Research," *Law & Society Review* 23 (1) (1989): 145–61; and Sherman, Lawrence W., and Ellen G. Cohen, "The Impact of Research on Legal Policy: The Minneapolis Domestic Violence Experiments," *Law & Society Review* 23 (1) (1989): 117–44.
- Sherman, Lawrence W., et al., "The Variable Effects of Arrest on Crime Control: The Milwaukee Domestic Violence Experiment," *Journal of Criminal Law and Criminology* 83 (1992): 137–69.
- Feld, Scott L., and Murray Straus, "Escalation and Desistance of Wife Assault in Marriage," *Criminology* 27 (1) (February 1989): 141–61; Langan, Patrick A., and Christopher A. Innes, *Preventing Domestic Violence Against Women*, Special Report, Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics, 1986; and Quigley, Brian M., and Kenneth E. Leonard, "Desistance of Husband Aggression in the Early Years of Marriage," *Violence and Victims* 11 (4) (1996): 355–70.
- Chalk and King, *Violence in Families*, 1998 (see note 6).
- Garner, Fagan, and Maxwell, "Published Findings from the Spouse Assault Replication Program," 1995 (see note 1).
- Berk, "What the Scientific Evidence Shows," 1993 (see note 5).

This study was conducted by Christopher D. Maxwell, Ph.D., assistant professor, Michigan State University; Joel H. Garner, Ph.D., Joint Center for Justice Studies; and Jeffrey A. Fagan, Ph.D., director, Center for Violence Research and Prevention (Columbia University), and visiting professor, Columbia University Law School. The authors would like to acknowledge the important contribution made by the staff of the National

Archive of Criminal Justice Data and the technical assistance provided by Dr. Jordan Leiter, former manager of NIJ's Data Resources Program, and Dr. Angela Moore Parmley of NIJ.

Support for the study was provided by NIJ, the Centers for Disease Control and Prevention (grant number R49/CCR210534), and the Harry Frank Guggenheim Foundation. The writing of this Brief was funded solely by NIJ.

Findings and conclusions of the research reported here are those of the authors and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

The National Institute of Justice is a component of the Office of Justice Programs, which also includes the Bureau of Justice Assistance, the Bureau of Justice Statistics, the Office of Juvenile Justice and Delinquency Prevention, and the Office for Victims of Crime.

This and other NIJ publications can be found at and downloaded from the NIJ Web site (<http://www.ojp.usdoj.gov/nij>).

NCJ 188199

Quick Access to NIJ Publication News

For news about NIJ's most recent publications, including solicitations for grant applications, subscribe to JUSTINFO, the bimonthly newsletter sent to you via e-mail. Here's how:

- **Send an e-mail to listproc@ncjrs.org.**
- **Leave the subject line blank.**
- **Type subscribe justinfo your name. (e.g., subscribe justinfo Jane Doe) in the body of the message.**

Or check out the "Publications and Products" section on the NIJ home page (<http://www.ojp.usdoj.gov/nij>) or the "New This Week" section at the Justice Information Center home page (<http://www.ncjrs.org>).