This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: The Economics of Crime: Lessons for and from Latin America

Volume Author/Editor: Rafael Di Tella, Sebastian Edwards, and Ernesto Schargrodsky, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-15374-6 (cloth); 0-226-79185-8 (paper)
ISBN13: 978-0-226-15374-2 (cloth); 978-0-226-79185-2 (paper)

Volume URL: http://www.nber.org/books/dite09-1

Conference Date: November 29-30, 2007

Publication Date: July 2010

Chapter Title: Does Arrest Deter Violence? Comparing Experimental and Nonexperimental Evidence on Mandatory Arrest Laws

Chapter Authors: Radha Iyengar

Chapter URL: http://www.nber.org/chapters/c11851

Chapter pages in book: (421 - 452)

12

Does Arrest Deter Violence? Comparing Experimental and Nonexperimental Evidence on Mandatory Arrest Laws

Radha Iyengar

12.1 Introduction

For two decades, there has been considerable debate in economics about the use and validity of experimental approaches in the development and design of public policy.[1] In criminology, there has been a marked increase in the use of randomized experiments, signaling a widespread acceptance of experiments as an effective means of determining policy-relevant parameters. Recently, a new set of papers in economics by Deaton (2009) and Imbens (2009), in addition to ongoing work by Heckman and coauthors (e.g., Heckman and Urzua 2009; Heckman, Urzua, and Vytlacil 2006) have revived the debate on the usefulness of experiments in estimating parameters of economic and policy importance in both first-world and developing-world contexts. The debate partitions empirical research into two broad categories: design-based studies (DBS), which focus on the empirical evaluation design, and theory-based studies (TBS), which focus on the underlying theory and fit new or existing estimates explicitly within that framework. Included in DBS are both experimental evaluations, such as randomized controlled trials (RCT) and field experiments, and nonexperimental evaluations focused on establishing causal relationships, such as difference-in-

Radha Iyengar is assistant professor of economics at London School of Economics, and a faculty research fellow of the National Bureau of Economic Research.

I am grateful to David Autor, Josh Angrist, Francine Blau, Rafael Di Tella, Faruk Gul, Hank Farber, Lawrence Katz, Steve Pischke, Jesse Rothstein, and participants at numerous seminars for helpful comments and insightful suggestions. Financial support from the Robert Wood Johnson Foundation is gratefully acknowledged. Any remaining errors are entirely my own.

1. Meyers (1992) provides useful overview on this debate. For additional details, see, for example, Ashenfelter and Card (1985), Lalonde (1986), Heckman and Robb (1985), Friedlander and Robins (1995).

difference estimators, matching methods, and instrumental variables. Much of the TBS literature uses "structural estimation" in which the parameters of an explicit theoretical model are identified by imposing restrictions across various parameters in a generalized method of moments or maximum likelihood framework. This set of studies can also include models of sample selections and usually require explicit assumptions on optimality, rational expectations/behavior, and parametric or functional form. In the context of crime research, most empirical research has fallen into the DBS category; the emphasis on experimental evidence combined with the importance of establishing causation makes this methodological debate critical in directing future of crime policy research. This study considers the conflict between DBS and TBS in the context of an important social experiment: arrest laws for spousal abuse.

The most prominent and arguably influential theory among crime scholarship is that of deterrence, which predicts that an increase in the cost of a crime will reduce participation in that crime. While there existed some simple theoretical models (e.g., Becker 1968; Ehrlich 1972) and some limited calibrated models (Flinn 1986), there was limited evidence on the efficacy of deterrence. To determine if any such significant deterrence effect existed, a prominent set of experiments funded by the National Institute of Justice tested the effect of increasing the probability of arrest on future incidences of spousal abuse (thus called the Spousal Assault Replication Program [SARP]). The initial experiment, run in 1981 in Minneapolis, Minnesota, suggested that arrests were effective at deterring future violence, reducing future violence by more than 50 percent.[2] These results were used to justify laws requiring the warrantless arrests of individuals police believe to be responsible for misdemeanor assault of an intimate partner. Recent quasi-experimental evidence using Federal Bureau of Investigation (FBI) homicide data (Iyengar 2009) finds that laws aimed at increasing arrests for spousal abuse appear to have increased intimate partner homicides. The apparent contradiction between the experimental and the nonexperimental results can be understood by placing the experimental results in the broader context of a behavioral model of spousal abuse. The distinction this simple model highlights is that the deterrence effect is a product of both the probability of arrest conditional on reporting and the probability of reporting conditional on assault. By measuring only one of these parameters, the experiment could not accurately extrapolate to the unconditional effect of increased penalty (i.e., increased probability of arrest) on future violent incidents.

These experiments and the subsequent nonexperimental evidence provide a stark example of a common situation where experimental evidence was

2. Subsequent replication produced what appeared to be a range of results from deterrence to escalation. However, a recent reanalysis of the six experiments suggests that in all cases, arrest reduced the probability of future violence. This will be discussed in detail in section 12.2.

necessary but not sufficient. It naturally begs the question: why do experiments occupy such an important place in policy debates when they are likely to be insufficient to answer the policy question? While there are a range of detailed and nuanced arguments on the relative efficacy of either experimental/DBS and TBS, the crux of the issue lies in the trade-off between internal and external validity.[3] The benefit of DBS is that they focus on identifying a parameter with great internal validity. This means that in experimental, quasi-experimental, or instrumental variable settings, we may obtain an unbiased estimate of the local average treatment effect (LATE) with minimal assumptions that are easier to test and to believe. The cost of this approach is that the parameter we estimate may not be parameter of interest. This is because we obtain the LATE rather than the average treatment effect (ATE). As its name suggests, the LATE is locally unbiased (or in the case of instrumental variables, consistent) for the group over whom the estimate is constructed.[4] Instrumental variables (IV) typically obtain the LATE because the estimation uses variation from only portion of the sample—in the language of IV, the compliant subpopulation (Angrist and Imbens 1995). Experimental studies will only obtain ATE if the sample under investigation is representative of the general population (or at least the population for whom the experimental results will be generalized).[5] Because most social experiments also rely on a voluntary subject population, by the same logic as IV, experiments with selected samples will tend to identify only the LATE, in this case local to the voluntary sample, rather than the average treatment effect (as in the case of a randomly selected sample).[6]

Theory-based studies, on the other hand, offer a detailed underlying

3. Following the definition in Burtless (1995), I define an "internally valid" estimate as an unbiased measure of a given treatment effect in the chosen sample. An "externally valid" estimate is a treatment-effect estimate that can be validly extended beyond the chosen sample to some other external group.

4. As Deaton (2009, 10) suggests, "this goes beyond . . . looking [at] an object where the light is strong enough to see; rather we have control over the light but choose to let it fall where it may." Of course, this is a bit extreme because the experimental parameters are not entirely separate from the desired parameter. As Imbens (2009, 6) notes, "even if simple average effects of these interventions are not directly answering questions about plausible economic policies, they are often closely related to the effects of such policies and therefore viewed as quantities of interest."

5. This point ignores noncompliance among the subject population after the start of the experiment. In the case of experimental noncompliance, two approaches are used. One uses the original assignment and compares the assigned treatment and assigned control groups, regardless of actual treatment status. This estimate is the intent-to-treat (ITT) estimate and may also be obtained in some quasi-experimental settings. The second approach is to use the assigned treatment as an instrument for actual treatment status. In this case, the experiment reverts to the IV case.

6. This is related to the point made by Heckman (1992) regarding experiments under the Job Training Partnership Act (JTPA). Experiments may not obtain the average treatment effect, even when properly administered, due to sample contamination—that is, the sample chosen is not the general population that would experience the treatment in the absence of the program.

framework for both analyzing and interpreting results. This allows studies to take observational and even experimentally obtained parameters and formally extrapolate the results to a wide range of settings. These studies also make explicit both the underlying behavioral mechanism that the policy or program affects and the assumptions necessary to apply the estimates to other settings. In this sense, the TBS design is built on the importance of external validity. The cost of such a method, however, is lack of transparency and credibility due to the often complex and detailed assumptions required for estimation. As such, these methods may lack the internal validity necessary to make policymakers and even other nonspecialists confident in the estimates obtained.

Thus, the trade-off, agreed upon by all sides, is as follows: when done properly, randomized experiments can precisely isolate the effect of a specific intervention—that is, experiments can estimate and statistically bound a targeted behavioral parameter. Unfortunately, there is no guarantee that the parameter estimated is of particular interest or relevant for a related policy. On the other hand, while structural estimates (and generally TBS) provide a framework for extrapolating results, the estimates obtained from these methods often lack internal validity and credibility.

Thus, the use of the spousal abuse arrest experimental results to justify policies that had perverse effects illustrates both the danger of treating the LATE as the ATE and the importance of having clear, credible information available to policymakers during the policy design. It is for this reason that this paper rejects the dichotomy between internally valid DBS and externally valid TBS as a false one. At the heart of the issue is a serious information problem regarding most behavioral outcomes related to desired policies. Thus, we require several components: a credible measure of the average effect of each behavioral parameter of interest;[7] an estimate of the relative magnitude of these multiple parameters; and the relationship between the distribution of treatment effects and relevant population characteristics. When considering the policy decision problem as a whole, we can see the issue is not the hierarchy of evidence based on methodological selection (although surely the quality of any particular study is important) but rather how to aggregate information from multiple sources.[8] Based on this insight, this study offers two contributions to the literature on the relative value of experimental, quasi-experimental, and structural estimates.

First, this article wishes to diverge with the general tendency on both sides to dismiss quasi-experimental methods as both atheoretical (relative

7. Note that in general an experimental "treatment effect" need not be the composite response parameter anticipated in a policy (we shall return to this point later).

8. A procedure suggested to aggregate experimental, quasi-experimental, and structural evidence is presented in section 6 of Imbens (2009). Imbens suggests that one may use experimental evidence to "pin down some combination of the structural parameters." It is in this spirit that this article suggests combining experimental and quasi-experimental evidence.

to structural estimates) and potentially biased (relative to experimental estimates). Indeed, all sides in the methodological debate seem to agree that "quasi-experimental" approaches are "second-best" alternatives to either extreme of structural estimation or experimental evaluations (e.g., Imbens 2009, 15; Deaton 2009, p 23). Actually, quasi-experimental studies play a critical role in bridging the gap between experimental and structural models; they provide the best possible means of measuring the relative magnitude of competing behavioral parameters conditional on a set of both observable and unobservable characteristics typically based on an underlying theory of economic or individual behavior. As always, there is a trade-off and, in particular, such evaluations require additional assumptions (as noted in Imbens 2009, 15). However, one might argue that these evaluations provide the only feasible means to test the full effects of a policy and, thus, provide greater, rather than less, external validity than randomized experiments. At the same time, the more simple, intuitive assumptions made by quasi-experimental methods provide a more transparent and credible estimates. Thus, quasi-experimental findings can more useful for informing policy decisions than comparable structural model–based estimates. This point is especially salient when applied to crime policy research where state-level variation in laws and procedures facilitates a wide range of DBS approaches. Crime researchers especially should ignore the value of quasi-experimental studies even in the presence of a wide range of potential experimental approaches.

Second, this article attempts to reconcile the two apparently oppositional methodological positions: DBS and TBS. If the goal is gathering sufficient and accurate information to inform policy decisions, then there is a great deal of room for agreement between the two camps. As Imbens (2009, 3) notes, "conditional on the question of interest being one for which randomized experiment is feasible, randomized experiments are superior to all other designs in terms of statistical reliability." This statement is almost indisputable, but that is largely because of the initial conditioning statement. At issue is that in many cases, the policy option cannot be perfectly replicated in an experiment. Rather, the experiment must be used to isolate some relevant behavioral parameter that we must than use to extrapolate. It is in this sense that as Deaton (2009, 4) notes a "randomized controlled trial has no special priority." Indeed, one may agree that RCT is the "gold standard" for obtaining internally valid estimate of some behavioral parameters, while noting that in general the experimental results themselves will be insufficient to answer the question of interest. As such, no matter how internally valid the estimate may be, it is of little value to the policy question under debate. Randomized control trials specifically and DBS more generally must be subject to the same external validity scrutiny as their TBS counterparts. This study illustrates that it is not always the case that "better LATE than nothing" if the LATE estimate provide misleading information to policymakers. In the end, these estimates are not useful and may even be counterproductive if assumptions regarding the applicability of the results are not made *explicit* and appropriate caution is not taken in interpreting and presenting these results.

In order to more precisely discuss these points, this article will present and discuss both the experimental and the nonexperimental evidence on spousal arrests. Section 12.2 discusses the experimental evaluations of arrest policy as an illustration for both the uses and limitations of experimental methods. Section 12.3 discusses the theoretical complexity in translating the spousal assault arrest experiment to an arrest policy. Section 12.4 discusses the use of quasi-experimental methods to resolve theoretical ambiguities and to contextualize evidence obtained experimentally. Section 12.5 returns to the broader debate, discusses issues the results from previous sections, and concludes.

12.2 Experimental Evaluations of Arrests for Spousal Abuse

The use of arrests for spousal abuse and, in particular, the laws which mandated arrests were premised on the results of a series of randomized experiments conducted by National Institute of Justice (NIJ). The experiments were carried out over a ten-year period, from 1981 to 1991, in six different cities with different police departments. The initial and most influential of these studies was the Minnesota Domestic Violence Experiment (MDVE).[9] The experiment was motivated by basic deterrence theory (suggesting that the increased penalty for a crime reduces the incidence of that crime), which was tested using an innovative design. This experiment, funded by the Minnesota Police Department, the Police Foundation, and the Department of Justice, was run by randomly assigning a police response to domestic violence calls (Sherman 1992). The objective was to determine if arrests were more effective at reducing future violence. Police applied one of three possible treatments: (a) advising and counseling the couple, (b) separating the individuals, or (c) arresting the suspect. Researchers then interviewed the victims shortly after police involvement and then followed up every two weeks for six months. The original results found that arresting the suspect resulted in substantially less future violence than did either advising or counseling (Sherman 1992). Indeed, the initially estimated effect sizes suggested that future violence was reduced by nearly 50 percent.[10]

The experiment, while excellently designed and extremely well done,

9. Evidence that MDVE was discussed when passing these laws can be found in Wanless (1996).

10. An in-depth evaluation of the results by Tauchen and Witte (1995) found that arrest resulted in significantly more deterrence than either advising or separating the couple, consistent with the original findings of the experiment. However, unlike the original findings, Tauchen and Witte use a dynamic setting that found that most of the deterrent effect of arrest occurs within two weeks of the initial arrest.

suffered from a standard experimental problem: compliance. As noted in Angrist (2006), officers deviated from their randomly assigned responses for largely one of three reasons: first, officers may have determined that it was inappropriate to advise or separate the couple because doing so may put the victim at risk. This was most often the case when the suspect attempted to assault the officer and when both parties were injured. Second, victims sometimes persistently demanded an arrest. Third, officers occasionally forgot to bring their report forms. When police were randomly assigned to arrest a suspect, they did so 98.9 percent of the time; when they were assigned to separate, they did so 77.8 percent of the time; and when they were assigned to counsel, they did so 72.8 percent of the time (Sherman and Berk 1984). When estimating using treatment assigned as an instrument (rather than estimating the intent-to-treat effect), the effect size appears even larger. Put differently, among compliers, the effect of arrest is nearly 80 percent reduction in recidivism. The direction of the bias gives us a great deal of information about noncompliers. Those who were arrested when assigned to a less severe police response were likely also subject to the "treatment" and may have reduced their violence dramatically. This dilutes the effect of the treatment group in the intent-to-treat measure resulting in larger IV estimates.

The MDVE was replicated by five other experiments over the next decade in Charlotte, North Carolina; Colorado Springs, Colorado; Metro-Dade County, Florida; Milwaukee, Minnesota; and Omaha, Nebraska. These replications were not exactly the same as the original experiment in two dimensions. First the precise treatments differed, though all assigned at least one "arrest" and one "nonarrest" police response. Table 12.1 reports the results from the randomized experiments. The most commonly held view is that the replications failed to show that arrest deterred and indeed provided some evidence of escalation. This does not appear to be the case in a simple comparison of means in which the only significant effects (Miami Dade survey, Omaha Police reports, and the original MDVE results) indicate that the arrest disposition reduced recidivism relative to any other alternative disposition.[11] An analysis of the pooled data from all six sites by Maxwell, Garner, and Fagan (2001) finds that "arresting batterers was consistently related to reduced subsequent aggression against female intimate partners." While not all the effect sizes were statistically significant, there did not appear to be any association between arresting the offender and escalation of violence against the victim.

Before turning to the limitations of these experiments, it is worth explicitly considering the value such estimates provide. First, prior to the results from the MDVE, there was a tendency for police to offer nonpunitive and "therapeutic" responses to spousal abuse. The MDVE provided clear and

11. For a detailed and thoughtful review of why these differences cannot well be interpreted given the data available, see Garner, Fagan, and Maxwell (1995).

Table 12.1 Estimated treatment effects from six experiments on spousal assault

Site	N initial (N final)	Experimental comparison	% noncompliers	Recidivism: Survey reports	Recidivism: Police reports
Minneapolis	314	1. Arrest	17.8	–0.131*	–0.083***
	(202)	2. Advise		(0.073)	(0.012)
		3. Separate			
Charlotte	686	1. Arrest	13.2	–0.039	.026
	(650)	2. Advise/Separate		(0.056)	(.031)
		3. Issue citation for court appearance			
Colorado Springs	1,660	1. Arrest + protective order	6.9	—	–0.001
	(1,658)	2. Protective order + counseling			(0.022)
		3. Protective order			
		4. Restore order at the scene			
Metro-Dade County	916	1. Arrest	10.0	–0.139***	–0.015
	(907)	2. Nonarrest		(0.047)	(0.028)
Milwaukee	1,200	1. Short arrest	0.0	0.014	–0.005
	(927)	2. Overnight arrest		(0.036)	(0.052)
		3. Warning that next instance will result in arrest			
Omaha[a]	577	1. Arrest	2.7	–0.026	–0.080*
		2. Separate		(0.033)	(0.049)
		3. Mediate			

Source: Garner, Fagan, and Maxwell (1995).

Notes: Differences calculated comparing arrest to nonarrest dispositions. All reported differences report *assigned* treatment. Estimates are calculated using reported "Recidivism" in each study. Treatment and control effect sizes are not adjusted for any covariates. Recidivism is defined as committing a new offense at least once in the six-month period after treatment. Raw effect sizes based on sample size and number of failures. Dashed cells indicate no information on specified variable available from site.

[a]The Omaha experiment actually consisted of two experiments, one in which offenders were arrested if present and the other issued an arrest warrant if the offender was not out. These are combined here with arrest warrant treated as the arrest disposition.

***Significant at the 1 percent level.

*Significant at the 10 percent level.

convincing evidence that such approaches were substantially less effective at reducing violence than if the abuser had been subjected to an arrest. The authors of the initial MDVE report were even able to separate the deterrence effect from a simple incapacitation effect. Because more than 40 percent of offenders were released within one day and over 85 percent were released within one week, there was very little incapacitation due to arrest and imprisonment. This finding was among the first clear tests of deterrence theory, and the behavioral parameter identified in this setting is of significant academic and policy import. Second, although less frequently discussed, the data from these studies also provided detailed information as to the importance of police response *relative* to other offender characteristics (such as prior criminal history, race, age, or employment status). Table 12.2 shows the set of covariates available in all six experiments, illustrating the rich individual level data obtained on a sample of calls to the police for domestic violence. It appears that the effect size from arrests is quite modest when compared to the effect of other factors, such as age or prior criminal record. This type of analysis helps contextualize the results from the experiment and provides some hints as to why it may be difficult to extrapolate from the data. Third, there appears to be some variation in the intensity of the treatment effect by demographic characteristics. While the average effect across the sites was statistically similar, the variation in the estimated effect of arrest on recidivism varied greatly. This was in part due to the different distribution of covariates in the control and treatment groups.

While the experimental evidence on arrests advanced our understanding of how batterers respond to arrest, there are several issues that make it difficult to generalize from even these well-conducted experiments.[12] First is the issue of sample contamination; this is the concern that the sample of individuals who are in the experiment are not the same as those who would be in place if a policy that increased the likelihood of arrest were implemented. In part, this is due to some of the sample restrictions from the experiments that excluded serious cases (i.e., felonies) as determined by the officer. In addition, officers may not have included cases they deemed to not be misdemeanor assault. The unknown true distribution of intimate partner abuse case severity and case type makes it difficult to sign the bias on this type of sample selection. In part, the sample is by construction different from the policy sample. The experimental designed ensured the initial reporting of an offense because the sample frame is based on domestic violence cases that request police presence. This means that all experimental results are the effect of arrest *conditional on reporting,* and this condition is quite meaningful in the case of domestic violence. The sample could not measure whether

12. The subsequent discussion should not be read as a criticism of the methodology of these experiments. Indeed, the initial MDVE and the five replications are among the most carefully run experiments in the crime literature.

Table 12.2 **Characteristics of subjects from six experiments on spousal assault**

	Minneapolis	Charlotte	Colorado Springs	Metro-Dade County	Milwaukee	Omaha
Number of subjects	205	638	1238	906	954	296
A. Incident assignment data						
Assigned to arrest (%)	29	33	26	51	68	34
Actually arrested (%)	43	40	28	60	68	34
Initial victim interview completion rates (%)	62	64	83	65	60	79
Final victim interview completion rates (%)	49	50	70	42	78	73
B. Suspect characteristics						
Age (average)	32	33	31	35	31	31
Employed (%)	40	77	87	71	47	78
Prior arrest (%)	59	31	43	12	62	65
Use of intoxicant (%)	61	54	59	31	29	59
Race/Ethnicity						
African American (%)	23	70	30	42	75	42
White (%)	58	28	54	36	20	51
Hispanic (%)	1	0	15	22	4	5
Asian/Native American/other (%)	18	2	1	0	1	3
Relationship with victim						
Married (%)	35	48	67	79	31	46
Separated or divorced (%)	3	2	5	5	1	1
Unmarried (%) (current or former)	47	50	28	16	68	52
Son, brother, roommate (%)	15					
C. Incident characteristics						
Misdemeanor (%)	98	97	37	100	100	100
Victim injured (%)	—	84	55	92	100	73

Sources: Minneapolis: Sherman and Berk (1984) and data provided by Angrist (2006); Charlotte, Colorado Springs, Metro-Dade County, Omaha: Maxwell, Garner, and Fagan (2001). Dashed cells indicate no information on specified variable available from site.

Notes: Differences calculated comparing arrest to nonarrest dispositions. All reported differences report *assigned* treatment. Categories are pooled when appropriate to facilitate comparability across studies. Omaha sample uses only the Offender-Present sample.

victims would be more or less likely to report to the police because it was based on an initial report.

Second is the issue of police cooperation. The departments chosen to conduct the experiments tended be those for whom compliance would be a minimal issue (Maxwell, Garner, and Fagan 2001). In addition, among the officers conducting experiment, most of the sample was collected by only a few officers.[13] This lends itself to ensuring high quality experimental design but not necessarily generalizability. In particular, the ways in which police may choose to avoid compliance must be explicitly considered when determining how to implement a policy to achieve the experimental results.

Finally, because of the relatively small number of cases, experiments will in general *not* identify the most serious, low-probability events. Serious injury and death are relatively rare occurrences but so undesirable that policies wish to avoid any increases in these high-cost outcomes.[14] Experimental results will in general lack the power to detect these low-probability events and even structural models may not be able to determine the realistic probability of such occurrences.[15]

Thus, the experimental results leave open several questions. First, given relatively similar average effects but differences in the variance of these effects across sites, can we generalize from arrests in these six sites to the efficacy of arrest relative to nonarrest in other locations? Second, if arrest, relative to other nonarrest police responses, reduces recidivism under experimental conditions, what is the effect of a policy that increases arrest? While the parameters estimated in the experiments are informative about both of these questions, they cannot answer them definitively without additional assumptions.

12.3 Arrest Laws and Criminal Behavior in a Repeat Interaction Setting

Although the experimental evidence from the six spousal assault experiment sites did not directly answer a policy questions, the experimental evidence on the efficacy of arrest encouraged many states to pass policies that encourage or require arrest of domestic abusers. These policies play a prominent role in the government's attempt to combat domestic violence.[16]

13. For example, in the MDVE, 9 percent of the officers produced 28 percent of the cases (Sherman and Berk 1984).

14. This statement should not suggest to the reader that intimate partner homicide is at acceptable levels. Indeed, women are more likely to be killed by a current or former intimate partner than anyone else. It is rather a statistical statement about the likelihood of observed intimate partner homicide given the total sample size.

15. A similar problem is discussed in the context of suicide due to antidepression drugs by Ludwig and Marcotte (2005).

16. For a detailed discussion of the emergence of these statutes, see Iyengar (2009). For a discussion on the role of the experimental evidence in influencing policy, see Maxwell, Garner, and Fagan (2001, 4–5).

Currently, twelve states and the District of Columbia have passed mandatory arrest laws. These laws requires police to arrest a suspect without a warrant if there is probable cause to suspect that an individual has committed some form of assault (either misdemeanor or felonious) against an intimate partner or family member. An additional ten states have recommended arrest laws, which specify arrest as recommended but not required when confronted with probable cause that an intimidate partner or familial assault has occurred. States in both of these groups are reported in table 12.3. These laws were implemented as an explicit strategy to increase the fraction of domestic violence cases in which police arrest the suspect. Many economists may immediately note that this is not a direct application of the experimental results.[17] However, this also represents a realistic setting in which experimental evidence, when it is the *only* source of information regarding a policy, may not produce the desired outcome.

To illustrate how mandatory arrest laws changed interactions between abusers and victims, consider how they changed the nature of interaction in the repeated setting of intimate partnerships. Mandatory arrest laws increase the cost of choosing violence. This effect is largely the reason why mandatory arrest laws were originally advocated. Ideally, this increase in cost would result in a situation where violence is never chosen. Unfortunately it is typically difficult to sustain such a situation, in particular because it requires victims to report whenever there is a violent incident. Because both anecdotal and sociological evidence suggest victims dislike reporting, many situations arise where batters are violent but victims do not report. Given then that reporting is uncertain, we cannot assume that violence is never chosen, and, as such, we will observe some level of violence.

Arrest laws also change the cost of reporting abuse to the police. The problem with this is that because abusers have more freedom in adjusting their behavior than do victims, this increased cost is borne almost entirely by the victims. To illustrate this, supposed mandatory arrest laws increase the utility to victims from reporting so that, all other things equal, reporting would be more desirable after the law change. In such a situation, an abuser could adjust the probability of violence if the victim does not report relative to the probability of violence if the victim does report. This essentially allows abusers to adjust the probability of each outcome to fully account for any utility gains from reporting to the victim. Similarly, suppose arrest laws decrease the utility to victims from reporting so that reporting is really a punishment strategy taken by the victim to induce better stream of behavior by the abuser in the future. In such a situation, an abuser can adjust the probability of violence if the victim reports such that any punishment strategy

17. Indeed, the authors of the original study note that mandating arrests may not be appropriate because of potential heterogeneity in offender responses (Sherman and Berk 1984, 270).

Table 12.3 **Mandatory arrest laws by state**

	Year passed	Code/Statute
Recommended arrest states		
AZ	1991	Ariz. Rev. Stat. Ann. §13-3601(B)
CA	1993	Cal. Penal Code §836(c)(1)
KS	2000	Kan. Stat. Ann. §22-2401(c)(2)
MS	1995	Miss. Code Ann. §99-3-7(3)(a)
MO	1989	Mo. Ann. Stat. §455.085(1)
NY	1994	N.Y. Crim. Proc. Law §140.10(4)
OH	1994	Ohio Rev. Code Ann. §2935.032(A)(1)(a)
SC	2002	S.C. Code Ann. §16-25-70(B)
Mandatory arrest states		
AK	1996	Alaska Stat. §18.65.530(a)
CO	1994	Colo. Rev. Stat. Ann. §18-6-803.6(1)
CT	1987	Conn. Gen. Stat. §46b-38b(a)
DC	1991	D.C. Code Ann. §16-1031(a)
IA	1990	Iowa Code §236.12(3)
ME	1995	Me. Rev. Stat. Ann. tit. 19-A, §4012(6)(D)
NV	1989	Nev. Rev. Stat. Ann. §171.137(1)
NJ	1991	N.J. Stat. Ann. §2C:25-21(a)
OR	2001	Or. Rev. Stat. §133.055(2)(a)
RI	2000	R.I. Gen. Laws §12-29-3(c)(1)
SD	1998	S.D. Codified Laws §23A-3-2.1
UT	2000	Utah Code Ann. §7-36-2.2(2)(a)
VA	2002	Va. Code Ann. §19.2-81.3(B)
WI	1996	Wis. Stat. Ann. §968.075(2)(a)
WA	1999	Wash. Rev. Code Ann. §10.31.100(2)

Source: http://www.westlaw.com.

Notes: Mandatory arrest states are defined as states where officers have no discretion as to whether to make a warrantless arrest when an intimate partner offense is reported. Recommended arrest states are defined as states where officers are instructed but not required to make a warrantless arrest when an intimate partner offense is reported. For specific information on coverage, see data appendix available on author's Web site (http://personal.lse.ac.uk/iyengarr/ma_appendix.pdf).

is unsustainable (i.e., reporting becomes too costly to make the utility gains from nonviolence in the future worth seeking).

Thus, the abuser can essentially be nicer to the victim if she does not report him after a violent incident, thus encouraging her not to report after violence (i.e., reducing the probability of violence conditional on not reporting). Indeed, this is often referred to by domestic violence advocates as the "honeymoon period," where abusers are extra attentive and loving. In addition, the abuser can take retributive action after reporting (i.e., increasing the probability of violence after reporting). Thus, the counterintuitive result of arrest policies for intimate partner violence is that they may indeed increase intimate partner violence because batters have a greater ability to determine the outcomes. Thus, the abusers are better able to shift the burden of arrest onto the victims, deterring reporting rather than deterring abuse.

To illustrate how changes in the level of homicides can be linked to the total number of abusive incidents, consider a model where with some small probability, p, domestic abuse escalates to murder. For n intimate partner incidents, the number of homicides in a jurisdiction is then pn. There are two main theories on potential responses that may result in increased homicides after arrest laws: reprisal (abuser response) and reporting (victim response). I will consider each in turn.

A different explanation for the response could be that abusers respond to arrest by punishing victims, and this increases intimate partner homicides. Suppose p increases because violence becomes more severe. This could occur if abusers are very angry when returning home after arrest and so more frequently commit violence against their partners. Thus, for a given n intimate partner incidents, the number of homicides pn increases. Note that if there is no deterrence effect, that is, n is constant, then, once again, the effect on the law is to increase violence. However, if there is a decrease in incidents (n declines), then the overall effect of mandating arrest is ambiguous. This response is consistent with evidence on victim fears. As discussed in the preceding, fear of reprisal is the most commonly cited reason for not reporting. To the extent that the fear is rational, this is consistent with the reprisal hypothesis. Moreover, when a victim leaves her relationship, she is at the greatest risk from her partner.[18] If arrest allows women to leave, then reprisal rates may increase. Evidence against this hypothesis comes from the experimental evidence, which found no significant increase in reprisal, though this may be because the abusers did not blame their victims for their subsequent arrest (instead blaming police officers).

While the experimental results may provide evidence against the reprisal hypothesis, the results are silent in the context of the reporting hypothesis. The reporting hypothesis suggests that victims are less willing to report an incident if their abuser will be arrested. Suppose that the probability of reporting given violence decreases after the passage of mandatory arrest laws. Because police presence, regardless of the police response, can disrupt a violent incident and prevent escalation to homicide, this failure to report to the police can increase the rate of intimate partner homicides. Thus, the victim's decision not to notify the police may increase p. This is not the overall effect of the law because the threat of arrest deters (as it did in the MDVE); then n, may decrease. In this case, the effect of the law on homicides is ambiguous. While arrest, conditional on reporting, deters violence, the unconditional effect of arrest on violence may be small or zero if victims substantially reduce their reporting.

Domestic violence victims may decide not to report for several reasons.

18. See Tjaden & Thoennes (2000).

First, there is a psychological and emotional component of intimate partner abuse that often generates victims who remain committed to their abuser and do not wish to send him to prison. Thus, the victim's guilt may increase her or his own costs of reporting as well as the abusers.[19] Second, if abusers are arrested but no further legal action is taken, they may return home within a day of their arrest and further terrorize their victim. In a nonexperimental evaluation of mandatory arrest as a policy, Lyon (1999) used a logistic model to compare the likelihood of arrest under mandatory arrest laws versus proarrest laws in two cities in Michigan. She found that once a victim calls the police to report an incident, she is significantly less likely to call again. She posits this was likely because police intervention in the form of an arrest resulted in retribution by the abuser, deterring future reporting.[20] Third, in many cases, arrest laws resulted in the victim also being arrested if there was evidence that she (or he) physically assaulted her (or his) partner. In many areas, women constitute nearly 20 percent of domestic violence arrests, a far higher percentage than the estimated proportion of female abusers.[21] Over half of these female arrestees can be identified as previous victims of intimate partner violence (Martin 1997). Anecdotal evidence from some battered women advocates suggests that these "dual arrests" are the most serious problem with mandatory arrest.[22] Dual arrests have serious implications for victims who are immigrants and may be deported if convicted of assault. In addition, those who have children face potential loss of custody during the arrest period. This latter response can be viewed as a method by police to avoid complying with the intentions of the mandatory arrest laws.

19. Recent research finds that many women do not perceive any benefit from mandatory arrest laws, no-drop policies, (requiring prosecution conditional on reporting) and mandatory medical reporting, and these laws may make them less willing to report in the future (Smith 2000).

20. Rennison (2000) found that fear of reprisal from the abuser was the most commonly cited cause for not reporting a domestic violence incident. This is a hotly contested claim. Mills (1998), based on research by Sherman and Berk (1984), claims that arrests actually increase reassaults. More recent work by Maxwell, Garner, and Fagan (2001) find that there is no significant change in the risk of assault.

21. For example, in Phoenix, Arizona, 18 percent of domestic violence arrests are women (http://www.azcasa.org). Women are thought to be abusers in less than 5 percent of intimate partner violence cases (Dobash et al. 1992). Though some work suggests there is a surprisingly high rate of female-on-male abuse (see Strauss and Geller 1980), this work is problematic and, for the most part, ignores the severity and context of the violence (see Blau 1998). This is particularly relevant in the case of intimate partner abuse. For example, suppose a husband spent years beating his wife severely. At the time of the survey, the husband shoved his wife, and she immediately threatened him with a knife. The conflict tactics scale (CTS) treats the wife's behavior as aggressive when it is, in context, clearly defensive. Moreover, the CTS fails to properly differentiate acts of violence that constitute severe abuse. When severity of abuse is considered, men typically have the higher rates of the most dangerous behaviors, such as firing a gun, repeating their violence more often, and doing more physical harm. For a greater discussion, see DeKeserdy and Schwartz (1998).

22. This statement is based on conversations with individuals at battered women's coalitions in New Jersey, Arizona, New York, California, Connecticut, and Illinois.

This, thus, represents a second way in which implementation of policies that encourage arrests may differ from experimental evaluations of arrest policies. All of these costs may result in an increased unwillingness to report abuse to the police.

Thus, there is a potential that the results of a policy to increase arrest are the exact opposite of those in the experiment, even when the experimental parameter is an unbiased estimate of the deterrence effect and is externally valid to groups outside the original sample. This can be explained by recognizing that the experimental results estimated the effect of an actual arrest conditional on reporting, while the estimates presented in this study estimate the unconditional effect of the certainty of arrest. The spousal abuse arrest experiments held constant the probability of reporting given violence because all cases in the experiment required an initial report of domestic abuse to the police. Thus, the spousal abuse arrest experiments estimated the effect of a decrease in batterer's utility, when they abuse and are reported, on their probability of choosing violence in the future. Unfortunately, if the victim also faces an increased cost from the increased penalty, then the overall effect of these laws on abuse is theoretically ambiguous (and empirically these laws appear to increase levels of abuse).

Finally, it is worth noting that heterogeneity in responses is especially important in this case. Even if most offenders respond to arrest by reducing future violence, the risk that some may respond to arrest by increasing the severity of violence (in the extreme, committing murder) highlights the importance of considering both the mean and the variance of the treatment effect. If, as in the case of domestic violence, substantial variance in response to treatment may result in very serious outcomes, then the local average treatment effect, as estimated by the experiment (which becomes local by the nonrandomness of sample selection), may be insufficient for policy-making purposes.

12.4 Estimates on the Effectiveness of Mandating Arrest

To test the effectiveness of mandatory arrest laws, I consider the effect of these laws on intimate partner abuse. This requires special attention to the total number of incidents of domestic violence, not simply the number of reported incidents, because the fraction of incidents that are reported to the police is potentially affected by this policy.[23] If I cannot observe unreported incidents, changes in the number of reported incidents and change in the total number of incidents (both reported and unreported) are observation-

23. The National Incident Based Reporting System (NIBRS), which does provide identification of the victim-offender relationships, is, therefore, ill-suited to the purposes of this study. Because the NIBRS solely comprises reported incidents, analysis of this data is not useful for measuring the true incidence of domestic violence.

ally equivalent.[24] In part, because I can observe victim-offender relationships, and in part because these crimes are almost perfectly reported, I use a measure of intimate partner homicides as a way to measure intimate partner abuse. Assuming that police intervention can reduce the probability of violence changes in the intimate partner homicide measure may provide insight into the impact of mandatory arrest laws on intimate partner violence.[25]

To construct a data set of intimate partner homicides, I use the FBI Uniform Crime Reports, Supplementary Homicide Reports (1976–2003), which provide data for all homicides that took place in the years 1976 to 2003 in all fifty states and the District of Columbia with additional descriptive variables about the victim, offender, and the nature of the crime (Federal Bureau of Investigation). I define an intimate partner homicide to include any homicide committed against a husband, wife, common-law husband, common-law wife, ex-husband, or ex-wife.[26] The data are constructed at the incident level with about 6.5 percent of the sample (36,442 observations) being intimate partner homicides.[27] I constructed a count of the number of relevant homicides by aggregating the incidents of intimate partner homicide, as defined in the preceding, in a given state for each year from 1976 until 2003. I also aggregated the number of intimate partner homicides by the race of the victim and offender and by sex of the victim and offender. Estimates are then scaled using census estimates for state population.[28]

A plot of the trend in various types of homicides before and after mandatory arrest laws suggests that these laws may have had a significant impact on intimate partner abuse. Figure 12.1 shows the rate of intimate partner and

24. An ideal data source for this type of analysis would be the National Crime Victimization Survey (NCVS) with state-level identifiers. Although previous researchers were able to access geo-coded versions of this data (see, for example, Farmer and Tiefenthaler 2003), recent changes in the administration and management of the NCVS make such access no longer possible. Some analysis using this data previously obtained suggests that mandatory arrest laws may reduce intimate partner violence but also reduce the number of cases that are reported in the system (Dugan 2003). Additional information about reasons why NCVS access is no longer possible is available upon request.

25. The linkage between misdemeanor assault prevalence and intimate partner homicide is well established. See, for example, Gwinn and O'Dell (1993). Moreover, the underlying causes are linked; see Mercy and Saltzman (1989).

26. The specific coverage of each law is reported in the legal appendix of Iyengar (2009). The distribution of samples across all groups is also described there.

27. There is some measurement error in the victim-offender relationship variable. About 1.25 percent of female victims reported as having a relationship to their offender that would imply she's a man and about .43 percent of male victims reported as having a relationship to their offender that would imply he's a woman. Together, these account for about 200 observations and less than 1 percent of the total sample. This is due to the classification of multiple homicides. In multiple victim homicides, the first victim-offender relationship is recorded for all of the victims. Because the selection of the "first" victim tends to be arbitrary, and this constitutes a very small fraction of the overall sample, these cases are excluded from analysis.

28. This scaling by population seems the appropriate deflater as arrest laws often apply to unmarried couples; however, the subsequent analysis has been repeated with a number of married couples with little qualitative effect on the coefficients.

Fig. 12.1 Intimate partner and familial homicide rates in mandatory arrest law states

Notes: Means based on author's own calculations using Supplementary Homicide Reports 1976–2003. Intimate partner homicides include homicides of husbands, wives, ex-husbands, ex-wives, common-law husbands, and common-law wives. Mandatory arrest states are defined as states where officers have no discretion as to whether to make a warrantless arrest when an intimate partner offense is reported.

family homicide rates as a function of time since the arrest law change. There appears to be a discrete increase of about 0.4 intimate partner homicides per 100,000. There is only a small decline in the number of family violence homicides. In contrast, figure 12.2 shows that recommended arrest laws have relatively little effect on intimate partner or familial homicides.

Comparing intimate partner homicides in states with and without arrest laws before and after the passage of these laws, I estimate a linear regression of the impact of mandatory arrest laws on the number of intimate partner homicides per 100,000 inhabitants. Column (1) of table 12.4 reports some coefficients from this regression. The *mandatory arrest effect* variable is defined as 1 in states that passed mandatory arrest laws in the years after the law was passed. Similarly, *recommended arrest effect* variable equals 1 in states that passed recommended arrest laws in the years after the law was passed. The results suggest that mandatory arrest laws are responsible for an additional 0.8 murders per 100,000 people. This corresponds to a 54 percent increase in intimate partner homicides.

There does not appear to be a significant effect in recommended arrest law states. Although the coefficient is negative, it is measured relatively imprecisely. Estimates in columns (2) and (3) of table 12.4 include controls for some other state characteristics and crime rates. Because these laws are between the previous discretionary arrest system and the mandatory arrest, we might expect a smaller but positive effect on homicide rates. There are

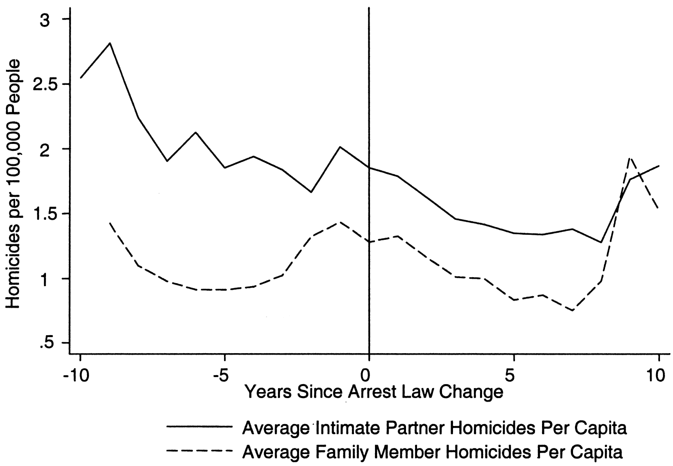

Fig. 12.2 Intimate partner and familial homicide rates in recommended arrest law states

Notes: Means based on author's own calculations using Supplementary Homicide Reports 1976–2003. Intimate partner homicides include homicides of husbands, wives, ex-husbands, ex-wives, common-law husbands, and common-law wives. Recommended arrest states are defined as states where officers are instructed, but not required, to make a warrantless arrest when an intimate partner offense is reported.

several reasons why this might not happen: first, if the arrest is perceived by abusers as discretionary, then they may not blame the victim for being arrested, reducing the reprisal rate. Second, because officers have discretion, victims may be more willing to call the police hoping to get an intermediate response. Finally, police themselves may not have changed their behavior much, opting to retain discretion and fill out paperwork rather than simply arresting.

There are several potential state-year factors that may be associated with both increased arrests and increased domestic violence. One important factor is the state crime rate, which may indicate how crime prone society is as well as the other crimes police must deal with. To measure the violent crime rate, I used the number of rape, robbery, and assault crime reports per 100,000 people from the FBI's Uniform Crime Reports. Column (2) of table 12.4 reports these results. Another concern might be state economic conditions that may increase domestic violence. I use average annual state unemployment rate derived from the Current Population Survey to control for this effect. There appears to be little effect of these limited covariates on the mandatory arrest law effect.

Column (3) of table 12.4 includes a more rich set of covariates including state-year level variables on demographics, economic conditions, and social policies. Because of racial differences in crime rates, I include some demographic controls (such as fraction of population that is black or white). I

Table 12.4 **Difference-in-difference estimates of mandatory and recommended arrest laws**

	All intimate partner homicides per 100,000 inhabitants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable Mean				1.48			
Mandatory arrest law effect	0.83**	0.81**	0.76**	1.15***	1.15***	1.10***	0.47
(=1 in MA law states after law change)	(0.33)	(0.33)	(0.34)	(0.40)	(0.41)	(0.39)	(0.30)
Recommended arrest law effect	–0.61	–0.66	–0.62	–0.31	–0.30	–0.36	–0.96*
(=1 in RA law states after law change)	(0.61)	(0.59)	(0.47)	(0.64)	(0.63)	(0.63)	(0.54)
Unemployment rate		0.01	0.04	0.019	0.019	0.02	0.02
		(0.07)	(0.09)	(0.07)	(0.07)	(0.07)	(0.07)
1 year postlaw change							0.23
							(0.27)
2 year postlaw change							0.14
							(0.31)
3 or more years postlaw change							0.41
							(0.42)
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Controls for other violent crime rates[a]	No	Yes	Yes	Yes	Yes	Yes	Yes
Controls for unemployment rate[b]	No	Yes	Yes	Yes	Yes	Yes	Yes
State-year demographic variables[c]	No	No	Yes	Yes	Yes	Yes	Yes
State-year economic and social controls[d]	No	No	Yes	Yes	Yes	Yes	Yes

Linear trend	No	No	No	Yes	No	No	No
Quadratic trend					Yes	No	No
State-specific trend controls	No	No	No	Yes	Yes	Yes	No
Postlaw interaction effects	No	No	No	No	No	No	Yes
State-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.6125	0.6127	0.6214	0.6275	0.6284	0.6252	0.6746

Notes: All regressions include 992 observations. The dependent variable for each column is the column title per 100,000 inhabitants. Robust standard errors, clustered by state, are reported in parentheses. Intimate partner homicides include homicides of husbands, wives, ex-husbands, ex-wives, common-law husbands, and common-law wives. Mandatory arrest (MA) states are states that require an arrest conditional on a report of domestic violence. Recommended arrest (RA) states are states where officers are instructed but not required to make a warrantless arrest when an intimate partner offense is reported.

[a]Crime rate controls use FBI Uniform Crime reports for the number of crimes per 100,000 inhabitants. Indexed crimes included in the violent crime variable are murder, robbery, assault, and rape. Indexed crimes included in the nonviolent crime count are burglary, larceny, motor vehicle theft, and drug crimes.

[b]Unemployment estimates are based on the March *Current Population Survey.*

[c]State demographic controls are based on the March *Current Population Survey* and include variables for the fraction of the population that is black, white, and other race, as well as age composition indicating share of prison population that is aged fourteen–nineteen, twenty–forty-nine, and fifty or older.

[d]State economic control variables are based on the March *Current Population Survey* and include the variables log state personal income per capita, and female-to-male employment ratio. State social policy controls include max Aid to Families with Dependent Children/Temporary Assistance to Needy Families (AFDC/TANF) for a family of three, unilateral divorce laws indicators (based on classification in Stevenson and Wolfers 2006), and indicators for whether the state has the death penalty.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

also include share of prison population in the state that is aged under twenty, twenty to thirty-five, thirty-six to forty-nine and fifty or older, which may be indicative of police behavior and crime enforcement levels in a given state. In addition to the unemployment rate used in the previous specification, I include economic covariates of crime, such as state-year average log personal income and male-female employment ratio. Finally, the state social policy controls that are related to crime generally include whether the state has the death penalty and the Aid to Families with Dependent Children/Temporary Assistance to Needy Families (AFDC/TANF) max for a family of three. I also included a control for when the state passed unilateral divorce laws based on Stevenson and Wolfers (2006). After including these covariates, the coefficient on the effect of mandatory arrest laws on intimate partner homicides shrinks to about 0.76, which is slightly smaller but similar in magnitude to the estimates from previous specification.

Because there was a significant secular trend in the domestic violence homicide rates, I estimated several specifications with trend variables. Column (4) reports the results when including a linear trend, and column (5) includes the results when including a quadratic trend. The inclusion of trend controls appears to increase the coefficient, suggesting that declining rates of intimate partner violence were inducing an underestimate of the full effect of the law. Column (6) reports the results from including state-specific linear trends. The coefficient is still larger than the estimates without a trend, consistent with the downward trend biasing the ordinary least squares (OLS) estimates.

To estimate the effect of the adoption of these laws over time, the specifications reported in column (7) include a time since law change interaction effect. Combined with the year-fixed effects, this both controls for any differences at a given point in time (year-fixed effect) as well as differences generated from the duration of the law (years since law change). The main effect of mandatory arrest laws corresponds to the effect of mandatory arrest laws in the initial year of passage. This effect is about half the size of previously estimated effect and insignificant. However, while the effect in the initial year is not significantly different than zero, the effect in the second year (the mandatory arrest law main effect plus the one-year postlaw change effect) is about 0.7 and is significant with a p-value of 0.02 (joint test statistics not reported in the table). The total law effect in later years is similarly significant (although the two-year postlaw effect is significant only at the 10 percent level), and there does not appear to be a significantly different effect of these laws over time. The effect does not appear to grow significantly over time (although there does appear to be a slight lag in the effect, which is to be expected). It is somewhat surprising that the effect of the law does not grow over time. There are several potential reasons why this may be the case. First is the annual nature of the data, which means that monthly growth over the first and second years may be missed and is aggregated into a single point estimate. Second, because the outcome variable is homicides, it may be less sensitive to the more subtle changes over time and, thus, is a relatively blunt outcome to measure the temporal diffusion of behavior. Finally, this is consistent with a reprisal story where the behavioral response is a one-time adoption immediately after the law change. If that is the case, and most police agencies adopted the law relatively rapidly, then we would not expect to see the effect grow over time.[29]

Because mandatory arrest laws were an important means by which domestic violence became represented and treated as a criminal justice issue (as compared to a family or community problem), we might be concerned that these laws will have a disparate impact on communities that have greater mistrust of the criminal justice system. In particular, some studies have shown that African American women may be especially reluctant to report crimes to the police, preferring instead to handle instances within their own communities.[30] To evaluate the effect on different subgroups of interest, columns (2), (3), and (4) of table 12.5 compare the effect of mandatory arrest laws on intimate partner homicides committed between white couples, African American couples, and Asian couples, respectively. The point estimate for white and blacks are similar although blacks have a larger increase in percentage terms (based on mean homicides rates reported in the first row of each column, whites have a base rate of 0.81 intimate partner homicides per 100,000, and blacks have a rate of 0.51 per 100,000). This provides some evidence that the negative effect of mandatory arrest laws is disproportionately strong in certain communities.

This is certainly consistent with some of the heterogeneity in the experimental evidence. For example, in Milwaukee, evidence suggested that African American men were more likely to escalate (rather than be deterred). In Dade County, experimental results suggested that unemployment increased the risk that the response was escalation rather than deterrence (Pate and

29. Thus far, little attention has been paid to the bias that unknown homicides might introduce into evidence. Underlying this is the assumption that it is less likely that family homicides are not likely to be unsolved as the offender would be a known individual (as opposed to a stranger-on-stranger crime in which the offender may be entirely unknown to the police). This assumption is not entirely accurate and, indeed, may produce some bias in measuring homicide rates (see Riedel 1999). However, a broad range of studies (e.g., Williams and Flewelling 1987; Pampel and Williams 2000) have suggested that family homicides need substantially less adjustment than intimate partner homicides, and, as such, this assumption may not be too harmful. For sensitivity analysis, see Iyengar (2009).

30. This point is highly contested. Evidence from the National Crime Victimization Survey suggests that African American women report intimate partner violence at higher rates than do their white or Asian counterparts (see, for example, Rennison 2002). However, surveys and outreach workers cite general mistrust of the police, mistreatment of the police, and concerns that reporting will send partners with criminal records back to prison as reasons why underreporting may be more prevalent in African American communities (see Hampton, Oliver, and Magarian 2003).

Table 12.5 **Estimates of the effect of mandatory and recommended arrest laws on intimate partner homicides for various subgroups**

	All intimate partner homicides per 100,000 inhabitants (1)	Intimate partner homicides with white victims and perpetrator (2)	Intimate partner homicides with black victims and perpetrators (3)	Intimate partner homicides with Asian victims and perpetrators (4)	Homicides of females by male intimate partners (5)	Homicides of males by female intimate partners (6)
Dependent variable mean	1.48	0.81	0.59	0.01	0.89	0.59
Mandatory arrest law effect	1.1525***	0.5080**	0.6208***	0.0144*	0.6023**	0.5502***
(=1 in MA law states after law change)	(0.4067)	(0.2339)	(0.1981)	(0.0077)	(0.2459)	(0.1863)
Recommended arrest law effect	–0.2960	–0.1225	–0.1342	0.0118	–0.0738	–0.2222
(=1 in RA law states after law change)	(0.6348)	(0.2785)	(0.3435)	(0.0124)	(0.3251)	(0.3270)
Unemployment rate	0.0196	0.0181	0.0066	–0.0000	0.0037	–0.0159
	(0.0741)	(0.0339)	(0.0403)	(0.0005)	(0.0366)	(0.0383)
Controls for other crime rates[a]	Yes	Yes	Yes	Yes	Yes	Yes
Controls for unemployment rate[b]	Yes	Yes	Yes	Yes	Yes	Yes
State-year demographic variables[c]	Yes	Yes	Yes	Yes	Yes	Yes
State-year economic and social controls[d]	Yes	Yes	Yes	Yes	Yes	Yes
State-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear trend	Yes	Yes	Yes	Yes	Yes	Yes
Year-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.6284	0.6733	0.5643	0.4536	0.6448	0.5910

Note: See table 12.4 notes and footnotes.

Hamilton 1992). Given the differences in unemployment rates across subgroups, this is also consistent with observed differences in estimate effects. While the results appear to hold for all groups, the heterogeneity may be particularly prominent in the certain communities. If we believe that certain communities may be less willing to report to police, the reporting effect might be stronger in those communities. In this case, I find some evidence of this, which may suggest that reporting by victims could explain the rise in homicides. The larger (in percent terms) effect among blacks and Asians provides support for the reporting effect over the reprisal effect if aversion to the police in general makes the response of minority communities stronger than the response in white communities.

Thus far, I have given little attention to the question of fault when constructing these counts. This is relevant because the intimate partner homicide count used thus far likely includes some homicides that are eventually (but not initially) classified as self-defense or justifiable. While I cannot identify "self-defense" killings from murders, homicides of males by their female intimate partner may more closely approximate the subset of cases for which self-defense is a plausible future classification. Column (5) of table 12.5 presents estimate of intimate partner homicides with only female victims killed by male intimate partners. Column (6) presents estimates of intimate partner homicides committed against males by their female intimate partners. Intimate partner homicides of females increase about 50 percent, a similar percent increase to the main, unrestricted estimate (presented in column [1] of table 12.5). Similarly, homicides of males by their female intimate partners are significantly affected by mandatory arrest laws—in fact, the effect is larger in percent terms. Overall, these results are consistent with either a reporting or reprisal effect in response to the law change. Indeed, to the extent that police intervention facilitates some flight or escape by victims, murder of the abuser may be a substitute for other improved outside options. This evidence is consistent with studies that suggest that battered women who kill their husbands do so more often when they have fewer extralegal opportunities.[31]

In an effort to verify the difference-in-difference framework, I test the effect of mandatory arrest laws on various sets of uncovered homicides. If the difference-in-difference estimates find a significant effect of mandatory arrest laws on homicides between individuals who should be unaffected by domestic and family violence laws, then it is likely the differences identified in the preceding may be unrelated to the passage of these laws. For the purposes of these falsification tests, I define a class of homicides called "other homicides," which includes homicides committed against employees, employers,

31. See, for example, O'Keefe (1997). This is also consistent with evidence that finds female perpetrated abuse is affected not by criminal justice options but by outside extralegal resources (e.g., shelters). (See Browne and Williams 1989.)

friends, other known individuals, and strangers.[32] These homicides should be unaffected by mandatory arrest laws. I estimate two specifications, one with only state- and year-fixed effects and one with the full set of covariates described in the preceding. The results from these regressions are reported in table 12.6, columns (1) and (2). In both specifications, neither mandatory arrest laws nor recommended arrest laws have a significant effect on the homicide level of uncovered homicides. To more closely approximate the homicides of females, I also estimate these two specifications on a count of "other homicides," which have female victims. The results from these two regressions are reported in columns (3) and (4) of table 12.6, and again, there appears to be no significant effect of these laws on homicide rates. Finally, I test the effect of arrest laws on intimate and familial homicides that are uncovered by arrest laws. These include homicides committed by boyfriends, girlfriends, homosexual partners, and nonnuclear family relatives. Because the SHR data is from police reports, the distinction between cohabiting or common-law married partners is somewhat blurred. While some states do treat cohabiting and common-law married partners differently, the question of whether mandatory arrest laws are enforced in cases of cohabiting intimate partner violence is unclear, and to date, there does not appear any systematic evidence to answer the question. In the case of noncohabiting intimate partners, the law will only be enforced if these groups are specifically covered.[33] The results are reported in columns (5) and (6) of table 12.6. These results suggest that there is no significant effect of these laws on uncovered homicides, and the estimated effects are significantly smaller.[34]

12.5 Discussion and Concluding Remarks

Experimental evidence from the Minnesota Domestic Violence Experiment and replication at five other sites encouraged many states seeking better responses to the problem of domestic violence to pass laws requiring the arrest of individuals believed to abuse their spouses. An evaluation of these laws suggests that they have increased level of intimate partner homicide.

32. I have excluded homicides committed by individuals of "unknown relationship." While it is likely that these homicides were not committed by immediate family members or intimate partners, it is was not possible to estimate the subset of these homicides that would be covered, and, thus, all are excluded.

33. Including boyfriends and girlfriends in the intimate partner counts for states in which these groups are not explicitly covered does not significantly change the results. They remain a relatively small fraction of all intimate partner homicides, and while there does not seem to be a significant effect on this group, the results for common-law married, married, and formerly married couples are robust to their inclusion.

34. The Fisher test for equality between mandatory and recommended arrest law coefficients is rejected at the 0.02 level. The comparison is between specifications reported in column (6) of table 12.3 and column (6) of table 12.5.

Table 12.6 Falsification tests of difference-in-difference estimates of the effect of mandatory and recommended arrest laws

	"Other homicides" per 100,000 inhabitants		"Other homicides" with female victims		Intimate partner homicides uncovered by arrest laws	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable mean	10.37		1.41		0.73	
Mandatory arrest law effect	4.2531	3.9985	0.5064	0.4631	0.2929	0.2656
	(2.7014)	(2.6749)	(0.4168)	(0.4036)	(0.2103)	(0.2071)
Recommended arrest law effect	0.5980	0.2595	–0.2492	–0.3104	0.1237	0.0911
	(3.8689)	(3.8360)	(0.5675)	(0.5531)	(0.2128)	(0.2050)
		–0.1151		–0.0167		–0.0153
		(0.4438)		(0.0569)		(0.0233)
Controls for other crime rates[a]	No	Yes	No	Yes	No	Yes
Controls for unemployment rate[b]	No	Yes	No	Yes	No	Yes
State-year demographic variables[c]	No	Yes	No	Yes	No	Yes
State-year economic and social controls[d]	No	Yes	No	Yes	No	Yes
Linear trend	Yes	Yes	Yes	Yes	Yes	Yes
State-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.6434	0.6454	0.6011	0.6050	0.6669	0.6710

Notes: All regressions include 994 observations. The dependent variable for each column is the column title per 100,000 inhabitants. Robust standard errors, clustered by state, are reported in parentheses. "Other homicides" include homicides committed against employees, employers, other (nonimmediate) family, friends, other known individuals, and strangers. Intimate partner homicides uncovered by law refers to relationships that are classified as intimate partner but were not specified in the state's arrest law statute. See Legal Appendix for detailed coverage by state. Mandatory arrest (MA) states are states which require an arrest conditional on a report of domestic violence. Recommended arrest (RA) states are states where officers are instructed but not required to make a warrantless arrest when an intimate partner offense is reported.

[a]Crime rate controls use FBI Uniform Crime reports for the number of crimes per 100,000 inhabitants. Indexed crimes included in the violent crime variable are murder, robbery, assault, and rape. Indexed crimes included in the nonviolent crime count are burglary, larceny, motor vehicle theft, and drug crimes.

[b]Unemployment estimates are based on the March *Current Population Survey.*

[c]State demographic controls are based on the March *Current Population Survey* and include variables for the fraction of the population that is black, white, and other race, as well as age composition indicating share of prison population that is aged fourteen–nineteen, twenty–forty-nine, and fifty or older.

[d]State economic control variables are based on the March *Current Population Survey* and include the variables log state personal income per capita, and female-to-male employment ratio. State social policy controls include max Aid to Families with Dependent Children/Temporary Assistance to Needy Families (AFDC/TANF) for a family of three, unilateral divorce laws indicators (based on classification in Stevenson and Wolfers 2006), and indicators for whether the state has the death penalty.

This may be because abusers escalate violence after arrest as retribution for their punishment or because abuse victims may be less likely to contact the police in the face of an increased likelihood the abuser will be arrested. While experimental evidence rejected the former theory (on retribution), it could not test the latter reporting theory. However, this failure to contact the police results in fewer interventions, risking an increased probability of escalating violence. The differences between the experimental and quasi-experimental results on arrests for spousal abuse thus raise three important issues related to the broader discussion on the value of design-based versus theory-based studies.

First, if the experimental conditions had been better replicated in the policy, would the experimental results have generalized? Put differently, if holding reporting fixed we could increase arrest rates, would violence decrease as predicted by experimental evidence. To test this, I considered the effect of mandatory arrest laws on homicides committed against members of the immediate family. Because mandatory arrest laws required the arrest of an abuser in a domestic situation, familial abuse was also covered by these laws. However, unlike for adults, children typically do not report their own physical abuse to police. Instead, abuse is usually detected by an outside adult (such as a teacher or a doctor).[35] In this case, the probability of reporting may not be affected by increased penalties for the abusers.[36] Under these conditions, a law mandating arrests may more closely replicate the experimental conditions and, therefore, probability of severe violence to children by family members. To test this further, I restrict attention to only homicides of school-aged children (i.e., age six to seventeen). It is likely that abuse of quite small children may rely more on the reporting by an individual within the household and thus be subject to the same transference of costs as direct victims of intimate partner violence. In contrast, school-aged children are likely to see teachers, doctors, and nurses on a regular basis. As such, heightened abuse of these children is mostly likely to generate an increased likelihood of third-party reporting.[37] Comparing states with and without mandatory arrest laws before and after the laws change, I find a nearly 75 percent reduction in homicides of these children.[38] These results are consistent with the model, suggesting that once arrest laws do not rely on reporting by the abused, these

35. More specifically, of the nearly 2.8 million child abuse cases reported to child protective services agencies in 2000, 56.1 percent of all reports were from law enforcement, educators, medical and mental health professionals, social services personnel, child care providers, and other mandated reporters. See U.S. Department of Health and Human Services, Administration on Children, Youth, and Families, *Child Maltreatment 2000*, (Washington, DC: U.S. Government Printing Office, 2002).

36. Actually, many professionals have legal requirements to report suspected abuse, which can compensate for any potential costs they might incur from reporting abusers in their community.

37. This mandated reported is believed to be related to the decline in familial homicides. For discussion of this trend, see Durose et al. (2005).

38. For detailed analysis, see Iyengar (2009). This specification controls for the full set of covariates included in table 12.4 as well as state- and year-fixed effects.

laws appear to function as predicted, reducing harm to victim by imposing costs on abuse. The effect size is quite similar to the IV estimates presented for the MDVE in Angrist (2006). This comparison suggests that perhaps the experimental results were inapplicable, but were insufficient to determine in what contexts the arrest may be an effective deterrence.

The second issue raised by the difference between the experimental and nonexperimental results highlights the concept stated succinctly by Deaton: "heterogeneity is *not* a technical problem" (2009, page 10). In particular, the concept of heterogeneity is intimately tied to the theory by which we extrapolate from experimental evidence. While understanding the mean effect is critical, determining the variance is crucial from determining how broadly effective a policy might be. In the case of arrest for spousal abuse, while intimate partner homicides may have increased, it is not certain that this corresponds to increased levels of intimate partner abuse. If the intimate partner homicides and intimate partner abuse are negatively correlated, then arrest laws may decrease abuse while increasing homicides. The theory of how arrests affect violence levels is thus critical in determining what to take away from both the experimental and the nonexperimental evidence. Understanding the nature of this heterogeneity is critical for determining how effective arrest policies may be. For example, if heterogeneity implies that low-level violence is deterred but for some small set of individuals, arrests increase escalation, leading to homicide, then a policy of more stringent lethality assessment and greater nonlegal resources for victims may be most appropriate. If, on the other hand, heterogeneity implies that violence and homicides increase for many victims, even if they decrease for others, encouraging arrest may not be an effective response to intimate partner violence. Thus, at the heart of the policy question is the extent to which the treatment population is heterogeneous, and that aspect is as important as accurately estimating the mean treatment effect. This argument should include the caveat that the reverse is true as well. That experimental studies may be limited in their ability to fully characterize the distribution of treatment effects does not undermine the value of what they can contribute: a credible, unbiased measure of the treatment effect—which is no small thing and also critical to policy decisions.

This leads to the final point highlighted by the experimental and nonexperimental results: the important role for quasi-experimental studies. It is clear that experimental evidence is both necessary and often insufficient for determining the full effect of many policies. In the case of violence, low probability but high-costs events like homicide are unlikely to be detected by small-sample experiments but critical for decision making by policymakers. A similar claim can be made about theory-based designs, which do not lend themselves naturally to transparency for policymakers. In addition, often theory alone can have ambiguous predictions of the overall effect of a policy. Quasi-experimental designs, especially those with transparent designs, have an important role to play, not as a second-best alternative, but as an important contribution to the overall information about the efficacy of a policy.

In conclusion, this article takes some issue with the debate that appears to force economists to take a stance on the primacy of either internal or external validity. In the end, such a distinction is not helpful because failure of either internal or external validity is problematic for both academics and policymakers. Instead of imposing a hierarchy of methods, I propose viewing the information provided by each method as complementary components to the knowledge necessary to make informed decisions about policy efficacy. I also wish to emphasize the point that in empirical research, humility is a virtue. There is a sad irony that a mandatory arrest law intended to deter abuse actually increases intimate partner homicides, which provides an important cautionary tale. Suggesting that implementing policies with *only* experimental evidence, absent theory and some confirming nonexperimental studies, may be not only ineffective but counterproductive, hurting the very people the policy seeks to help. Thus, rather than view experiments, quasi-experiments, or structural estimation as procedures at odds with each other, this paper highlights the value that an integrated approach, which explicitly links randomized controlled trials, quasi-experimental studies, and structural modeling, may provide to more fully understand the effects of a desired policy intervention.

References

Angrist, J. 2006. Instrumental variables methods in experimental criminological research: What, why and how. *Journal of Experimental Criminology* 2:23–44.

Angrist, J., and G. Imbens. 1995. Two staged least squares estimates of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90 (430): 431–42.

Ashenfelter, O., and D. Card. 1985. Using the longitudinal structure of earnings to estimate the effect of training programs. *Review of Economics and Statistics* 67 (4): 648–60.

Becker, G. S. 1968. Crime and punishment: an economic approach. *Journal of Political Economy* 76 (2): 169–217.

Blau, F. D. 1998. Trends in the well-being of American women, 1970–1995. *Journal of Economic Literature* 36 (1): 112–65.

Browne, A., and K. Williams. 1989. Exploring the effect of resource availability and the likelihood of female-perpetrated homicides. *Law and Society Review* 23 (1): 75–94.

Burtless, G. 1995. The case for randomized field trials in economic and policy research. *Journal of Economic Perspectives* 9 (2): 63–84.

Buzawa, E. S., and C. G. Buzawa. 1996. *Domestic violence: The criminal justice response.* 2nd ed. Thousand Oaks, CA: Sage.

Deaton, A. 2009. Instruments for development: Randomization in the topics, and the search for the elusive keys to economic development. Paper presented at Keynes Lecture, British Academy, London.

DeKeserdy, W. D., and M. Schwartz. 1998. *Measuring the extent of woman abuse in intimate heterosexual relationships: a critique of the conflict tactics scales.* Harrisburg, PA: National Online Resource Center on Violence Against Women. Available at http://new.vawnet.org/assoc_files_vawnet/ar_ctscrit.pdf.

Dobash, R. P., E. E. Dobash, M. Wilson, and M. Daly. 1992. The myth of sexual symmetry in marital violence. *Social Problems* 391:71–91.

Dugan, L. 2003. Domestic violence legislation: Exploring its impact on the likelihood of domestic violence, police involvement, and arrest. *Criminology and Public Policy* 2:283.

Durose, M., C. Harlow, P. Langan, M. Motivans, R. Rantala, and E. Smith. 2005. Family violence statistics: Including statistics on strangers and acquaintances. Bureau of Justice Statistics Report no. NCJ 207846. Washington, DC: Bureau of Justice Statistics.

Ehrlich, I. 1972. The deterrent effect of criminal law enforcement. *Journal of Legal Studies* 1 (2): 259–276.

Farmer, A., and J. Tiefenthaler. 2003. Explaining the recent decline in domestic violence. *Contemporary Economic Policy* 21 (2): 158–72.

Federal Bureau of Investigation. 1999. *Uniform Crime Reports, Supplementary Homicides Reports 1976–1999.* Washington, DC: GPO.

Federal Bureau of Investigation. 1999. *Uniform Crime Reports, Offenses Known and Clearances 1965–1999.* Washington, DC: GPO.

Flinn, C. 1986. Dynamic models of criminal careers. In *Criminal careers and "career criminals."* Vol. 2, ed. A. Blumstein. Washington, DC: National Academy Press 356–79.

Friedlander, D., and P. K. Robins. 1995. Evaluating program evaluations: New evidence on commonly used nonexperimental methods. *American Economic Review* 85 (4): 923–37.

Garner, J., J. Fagan, and C. Maxwell. 1995. Published findings from the Spouse Assault Replication Program: A critical review. *Journal of Quantitative Criminology* 11 (1): 3–28.

Gwinn, C. G., and A. O'Dell. 1993. Stopping the violence: the role of the police officer and the prosecutor. *Western State University Law Review* 20:297–317.

Hampton, R., W. Oliver, and L. Magarian. 2003. Domestic violence in the African American community. *Violence Against Women* 9 (5): 533–57.

Heckman, J. J. 1992. Randomization and social policy evaluation. In *Evaluating welfare and training programs,* ed. C. F. Manski and I. Garvinkel. Cambridge MA: Harvard University Press, 207–30.

Heckman, J. J., and R. Robb. 1985. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics* 30:239–67.

Heckman, J. J., and S. Urzua. 2009. Comparing IV with structural models: What simple IV can and cannot identify. NBER Working Paper no. 14706. Cambridge, MA: National Bureau of Economic Research.

Heckman, J. J., S. Urzua, and E. J. Vytlacil. 2006. Understanding instrumental variables in models with essential heterogeneity. *Review of Economic and Statistics* 88 (3):389–432.

Imbens, G. 2009. Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). NBER Working Paper no. 14896. Cambridge, MA: National Bureau of Economic Research.

Iyengar, R. 2009. Does the certainty of arrest reduce domestic violence: Evidence from mandatory and recommended arrest laws. *Journal of Public Economics* 93:85–98.

Lalonde, R. 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76 (4): 604–20.

Ludwig, J., and D. Marcotte. 2005. Anti-depressants, suicide and drug regulation. *Journal of Policy Analysis and Management* 24 (2): 249–72.

Lyon, A. D. 1999. Be careful what you wish for: An examination of arrest and prosecution patterns of domestic violence cases in two cities in Michigan. *Michigan Journal of Gender and Law* 5:272–97.
Martin, M. 1997. Double your trouble: Dual arrest in family violence. *Journal of Family Violence* 12 (2): 139–57.
Maxwell, C. D., J. H. Garner, and J. A. Fagan. 2001. The effects of arrest on intimate partner violence: New evidence from the Spouse Assault Replication Program. Research in Brief. NCJ 188199. Washington, DC: U.S. Dept. of Justice, National Institute of Justice.
Mercy, J. A., and L. E. Saltzman. 1989. Fatal violence among spouses in the United States. *American Journal of Public Health* 79 (5): 595–99.
Meyers, B. 1992. Natural and quasi-experimental in economics. *Journal of Business and Economic Statistics* 13 (2): 151–61. JBES Symposium on Program and Policy Evaluation (April 1995).
Mills, L. G. 1998. Mandatory arrest and prosecution policies for domestic violence: A critical literature review and the case for more research to test victim empowerment approaches. *Criminal Justice and Behavior* 25:306–18.
Neville, H. A., and A. O. Pugh. 1997. General and culture-specific factors influencing African American women's reporting patterns and perceived social support following sexual assault. An exploratory investigation. *Violence Against Women* 3 (4): 361–81.
O'Keefe, M. 1997. Incarcerated battered women: A comparison of battered women who killed their abusers and those incarcerated for other offenses. *Journal of Family Violence* 12 (1): 1–19.
Pampel, F. C., and K. R. Williams. 2000. Intimacy and homicide: Compensating for missing data in the SHR. *Criminology* 38 (2): 661–79.
Pate, A., and E. Hamilton. 1992. Formal and informal deterrents to domestic violence: The Dade County spouse assault experiment. *American Sociological Review* 57:691–97.
Pollak, R. A. 2004. An intergenerational model of domestic violence. *Journal of Population Economics*
Rennison, C. 2002. Rape and sexual assault: Reporting to police and medical attention 1992–2000. Washington, DC: Bureau of Justice Statistics.
Sherman, L. 1992. *Policing domestic violence.* New York: Free Press.
Sherman, L., and R. A. Berk. 1984. The specific deterrent effects of arrest for domestic assault. *American Sociological Review* 49 (1): 261–72.
Smith, A. 2000. It's my decision, isn't it? A research note on battered women's perceptions of mandatory intervention laws. *Violence Against Women* 6 (12): 1384–1402.
Stevenson, B., and J. Wolfers. 2006. Bargaining in the shadow of the law: Divorce laws and family distress. *Quarterly Journal of Economics* 121 (1): 267–88.
Strauss, M. A., and R. J. Gelles. 1980. *Behind closed doors: A survey of family violence in America.* New York: Double day.
Tauchen, H., and A. D. Witte. 1995. The dynamics of domestic violence. *American Economic Review* 85 (2): 414–18.
Tjaden, P., and N. Thoennes. 2000. *Extent, nature, and consequences of intimate partner violence: findings from the National Violence Against Women Survey.* Washington, DC: Dept. of Justice, National Institute of Justice.
Williams, K., and R. L. Flewelling. 1987. Family, acquaintance, and stranger homicide: Alternative procedures for rates calculations. *Criminology* 25 (3): 543–60.