Increased death rates of domestic violence victims from arresting vs. warning suspects in the Milwaukee Domestic Violence Experiment (MilDVE)

Lawrence W. Sherman · Heather M. Harris

Abstract

Objectives We explored death rates from all causes among victims of misdemeanor domestic violence 23 years after random assignment of their abusers to arrests vs. warnings.

Methods We gathered state and national death data on all 1,125 victims (89 % female; 70 % African-American; mean age=30) enrolled by Milwaukee Police in 1987–88, after 98 % treatment as randomly assigned.

Results Victims were 64 % more likely to have died of all causes if their partners were arrested and jailed than if warned and allowed to remain at home (*p*=.037, 95 % CI= risk ratio of 1:1.024 to 1:2.628). Among the 791 African-American victims, arrest increased mortality by 98 % (*p*=.019); among 334 white victims, arrest increased mortality by only 9 % (95 % CI=RR of 1:0.489 to 1:2.428). The highest victim death rate across four significant differences found in all 22 moderator tests was within the group of 192 African-American victims who held jobs: 11 % died after partner arrests, but none after warnings (*d*=.8, *p*=.003). Murder of the victims caused only three of all 91 deaths; heart disease and other internal morbidity caused most victim deaths.

Conclusions Partner arrests for domestic common assault apparently increased premature death for their victims, especially African-Americans. Victims who held jobs at the time of police response suffered the highest death rates, but only if they were African-American. Replications and detailed risk factor studies are needed to confirm these conclusions, which may support repeal or judicial invalidation of state-level mandatory arrest laws.

L. W. Sherman · H. M. Harris (✉)
Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA
e-mail: hmh@umd.edu

L. W. Sherman
Institute of Criminology, University of Cambridge, Sidgwick Avenue, CB3 9DA Cambridge, UK
e-mail: ls434@cam.ac.uk

Introduction

Mandatory arrest for *misdemeanor* domestic violence—mostly without injury—was
adopted by over half of U.S. state legislatures in the 1980s (Mills 2003; Sherman
1992). These laws required arrests for slaps, kicks, punches, or other acts unlikely to
cause death or serious injury (as distinct from aggravated assault, for which arrest was
not made mandatory). They were based in part on a randomized experiment that found
arrest for these minor assaults caused less re-offending for 6 months after police
response (Sherman and Berk 1984).

Yet by the early 1990s, five replications of the original experiment created a more
complex body of evidence: arrests consistently reduced repeat offending for employed
suspects, but among unemployed suspects arrest doubled recidivism (Pate and Hamilton
1992; Sherman and Smith 1992) or, at least, was ineffective (Berk et al. 1992). A pooled
sample analysis of all six trials found a small average benefit but confirmed the different
effects on jobless versus employed suspects (Maxwell et al. 2002). Although these
recidivism findings suggested that many victims, like their suspect counterparts, might
be harmed by mandatory arrest, the research had little impact on state laws or police
policies—partly because of measurement issues in both low-response rate victim
interviews and selective reporting of domestic violence to police.

A quarter-century after these experiments, a clearer outcome can be exam-
ined to measure the impact of mandatory arrest on victims: death. This unam-
biguous measure provides a more powerful focus for debate about the wisdom
of current state laws. While no experiment on cases can provide evidence on
whether a general deterrent effect of mandatory arrest lowers domestic assault
across entire cities, mortality data can suggest whether more victims are dying
after arrests than after warnings. Since the majority of domestic homicides
occur without police contact of any kind with the victim prior to these murders
(Sherman 1991; Sherman and Strang 1996), knowledge of the effects of arrests
on mortality of those victims who do have contact with police may provide the
most concrete guidance for policy.

One case–control design has reported that for domestic homicide, arrest is a
protective factor (Campbell et al. 2003). No prospective-longitudinal or experimental
design, however, has previously examined the association between partner arrest and
mortality of domestic violence victims.

We began this long-term follow-up by attempting to discover whether arrest de-
terred—or increased—future domestic homicide. As our study protocol indicated, we
were also open to all possible effects of arrest on mortality. Our intention in this
analysis was to explore any possible connection between arrest and long-term mortality
rates. Milwaukee offered the experiment with the highest success rate (98 %) of the six
1980s trials in delivering treatments as randomly assigned (Sherman 1992). This trial
arguably provides an opportunity to discover causal links of arrest to death even
without directly measuring any causal pathways triggered by the randomly assigned
treatments.

Methods

Sample and trial design

The 1987–88 Milwaukee Domestic Violence Experiment (MilDVE) was reported in detail shortly after its completion (Sherman 1992; Sherman and Smith 1992), including eligibility criteria, procedures, and treatment fidelity. No long-term mortality outcomes for victims have been reported from MilDVE to date (Sherman and Harris 2013).

MilDVE was a three-armed (1:1:1) randomized trial comparing "short" and "long" arrests of suspected assailants to a warning. In the latter arm, police left the scene while suspects and victims were usually still present together. In "short" arrests, police released suspects in a mean of 4.5 h after booking at police headquarters; in "long" arrests release averaged 11.1 h after booking. About half of the victims in the arrest arm met with a prosecutor at the courthouse the next business day, but only 5 % of the arrests were prosecuted.

Cases of serious injury were ineligible, as were attempts to inflict serious injury. While 13 % of the eligible cases had an injury for which victims sought medical attention, only 5 % of victims went to hospital after the police came (Sherman 1992). Both suspects and victims were required to be older than 18 and in a domestic relationship. A team of 36 officers screened, enrolled, and accepted random assignment of 1,200 cases from April 6, 1987, through August 8, 1988 (Sherman 1992).

Mortality follow-up study

In 2012–2013, the present authors obtained the list of victim names and dates of birth in MilDVE, then purchased mortality data from the Wisconsin Office of Vital Statistics that we cross-checked and expanded through searches of the Social Security Death Index (SSDI), which features national coverage of all persons who have ever received Federal welfare benefits of any kind. Tests of the comprehensiveness of its Death Master File (DMF) show that for the people aged 25–54 (the age most victims in our sample entered the study) the percentage of publicly available death records compared to the national (confidential) list of death certificates was 75 % in 1997, having risen steadily for over a decade (Hill and Rosenwaike 2001).

Since only 10 % of the DMF cases found in our study lacked records in Wisconsin, a small but unknown number of cases were probably undetected during mortality follow-up, but were arguably not "lost" in CONSORT terms (Fig. 1) because we looked for them in the same way that we found the other deaths. To avoid errors, multiple dates of birth, cities in which decedents had resided, and other checks were used to determine matches. Cause of death is not known for the nine deaths (out of 91) identified only by the DMF. There is no evidence that any differential movement out of state was created by the randomly assigned treatment. Errors in detecting death would likely have the same structure in both groups.

While there was no national registry of crime and justice trials in 1987, the primary MilDVE publications (Sherman 1992; Sherman and Smith 1992) report that approval for the random assignment of arrest was granted by a vote of the Milwaukee City Council and an Institutional Review Board of the Crime Control Institute of Washington DC, both in 1986. Approval for our 23-year follow-up study was obtained

Trial Profile: Victims As First Treated

Assessed: n=1,979*

Ineligible: n=854*
Randomized: n=1,125

Enrollment

13 victims were treated
as suspects before they
were treated as victims

Randomized to Arrest: n=754
Received Arrest: n=750
Warned: n=4

Allocation

Randomized to Warning: n=371
Received Warning: n=369
Arrested: n=2

Lost to Follow-Up: n=0

Follow -Up

Lost to Follow-Up: n=0

Excluded from Analysis: n=0

Analysis

Excluded from Analysis: n=0

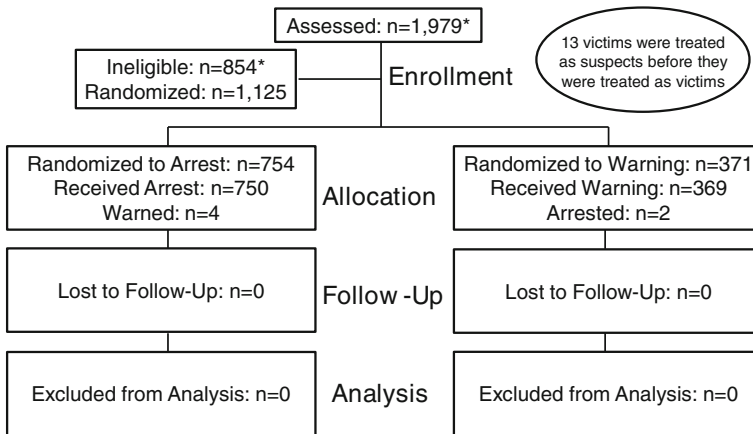* Estimated assuming all ineligible cases included unique individuals

Fig. 1 CONSORT diagram for MilDVE victims as first treated

from the Institutional Review Board of the University of Maryland in 2012, title 334834–2.

Randomization and masking

Suspects were randomly assigned to arrest or warning in a sequence determined by an independent statistician (Sherman 1992), without consent of either victims or suspects, consistent with ethical standards and law for experimentation in criminal sanctions (Federal Judicial Center 1981). Assignments were enclosed in sequentially numbered, opaque, sealed envelopes opened one at a time by the independent research team. Details of the sequence remained unknown to all police and members of the research team until police called researchers to report identifying details for both suspect and victim, and each envelope was opened while police officers remained on the phone and researchers communicated the assigned treatment. Because randomization assigned suspects to different legal statuses, it was impossible to mask the treatment from the research staff or participants.

Procedures

After deciding that there was probable cause to believe that an eligible suspect had assaulted an eligible victim, MilDVE police officers called the warrant desk to insure there were no outstanding arrest warrants for the suspect. Police then obtained the randomly assigned treatment and arrested or warned the suspects accordingly, using a standard script for the warning. Police handcuffed suspects in 94 % of arrests, doing so in front of the victims in 73 %. No victims in any case were arrested.

MilDVE randomly assigned arrests or warnings to cases, rather than to people. A CONSORT diagram for victims appears below as Fig. 1. The experiment allowed re-enrolment of the same persons in multiple cases, based on a primary concern with

short-term repeat offending (Sherman 1992). For purposes of assessing long-term differences in mortality, we analyzed the random assignment of intention-to-treat in the first case in which each individual appeared as a victim, even if they were subsequently treated as an offender or as a victim whose partner received a different treatment. Only 1 % of victims (13 of 1,125) had previously been treated in the experiment as suspects.

Statistical analyses

Most analyses below pool data from the two arms of the trial in which victims' partners were arrested. This procedure under-estimates the effect of the "standard" arrest in Milwaukee at the time, described here as "long" arrest, since it was only "short" arrest that was adopted for the unique purpose of the trial; "long" arrest was the standard arrest condition before and after the trial. We combine the two arrest categories to both increase the power of moderator analyses, and provide more external validity to a range of time periods in police custody—something that can vary widely across police agencies operating under identical mandatory arrest laws.

Using the intention-to-treat principle, we calculate relative risk ratios and their confidence intervals between mortality for victims whose partners were arrested or warned in their first enrolment as a victim. We examine both main effects and subgroup effects conditional on theoretically relevant moderator factors. As a sensitivity analysis, we tested the effect of partner arrest on victims who had at any point had their partner arrested by random assignment in the experiment, and not just at the point of their initial enrolment as victims; we found no different results from that procedure. We also used Cox hazard models to adjust for slight imbalances in known baseline covariates that could predict outcomes, and again for race as a strong moderator. We used Stata Version 13 for all analyses.

Results

Of the 2,054 cases assessed by the 35 police officers, 1,200 cases were randomly assigned to arrest or warning, with 1,125 unique victims whose partners were treated in one or more of the cases. The CONSORT diagram shows the treatment pathways of this sample with 98 % compliance on intent-to-treat.

Table 1 shows victims' characteristics at baseline, which are similar but not identical for the two treatment groups. Both groups were about 70 % African-American, 10 % male, and averaging 30 years of age. Of eight characteristics examined, however, three show significant differences between treatment groups. Victims whose partners were arrested were more likely than those whose partners were warned to have had partners who were employed, partners with a prior arrest, and a prior arrest themselves. The combined effect of these differences yields an under-estimate of the effect of partner arrest on mortality, based on the moderator analyses reported in Table 3. We adjust for these differences in a Cox proportional hazard model estimated in Table 4 and displayed (after correction) in Fig. 6 below.

Table 1 Victim characteristics at baseline

	Missing		Arrested		Warned		t test (α=0.10)	p value
Variable	N	%	n	%	n	%		
Treatment	0	0.00	754	67.02	371	32.98		
Victim female	0	0.00	682	90.45	332	89.49	−0.0218	0.98
Victim black	0	0.00	529	70.16	262	70.62	−0.1589	0.87
Victim employed	3	0.27	224	29.71	106	28.57	−0.3117	0.76
Victim had prior arrest**	0	0.00	217	28.78	79	21.29	−2.6871	0.01
Suspect employed***	15	1.33	364	48.28	129	34.77	−4.2892	0.00
Suspect had prior arrest**	0	0.00	502	66.58	217	58.49	−2.6615	0.01
Couple married at treatment	2	0.18	230	30.50	105	28.30	−0.7453	0.46
	Mean	SD	Mean	SD	Mean	SD	t test (α=0.10)	p value
Age	0	0.00	29.79	8.76	29.55	8.48	−0.4063	0.68

= significant at 0.01 level; *= significant at <0.001 level

Main effects

The primary outcome for all cases showed 64 % more deaths among victims whose suspects had been arrested (Fig. 2). The arrested-partner group suffered 92.8 deaths per 1,000 victims from all causes, compared to 56.6 per 1,000 for the victims whose partners had been warned ($p=0.02$).

The relative risk ratio was even larger (but not significant) in the first 5 years after random assignment, with victim deaths in that period over three times higher in the partner-arrest group. The 23-year trend in main effect size is displayed in Fig. 3.

Disaggregating outcomes by cause of death (Table 2) showed only two categories with even marginally significant differences: heart disease ($p=0.122$), which caused twice as many deaths for partner-arrested victims as for partner-warned, and "other" internal causes ($p=0.06$), for which arrest raised the risk of death by 183 % ($d=0.4$). Homicide rates were
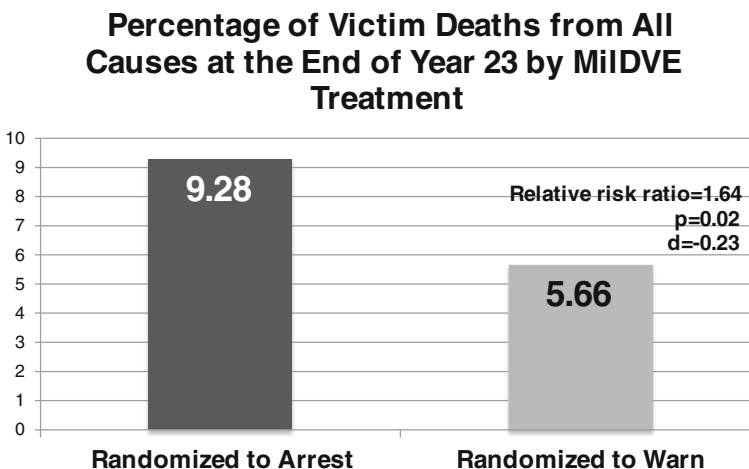

Fig. 2 Main effects of suspect arrest on 23-year victim mortality
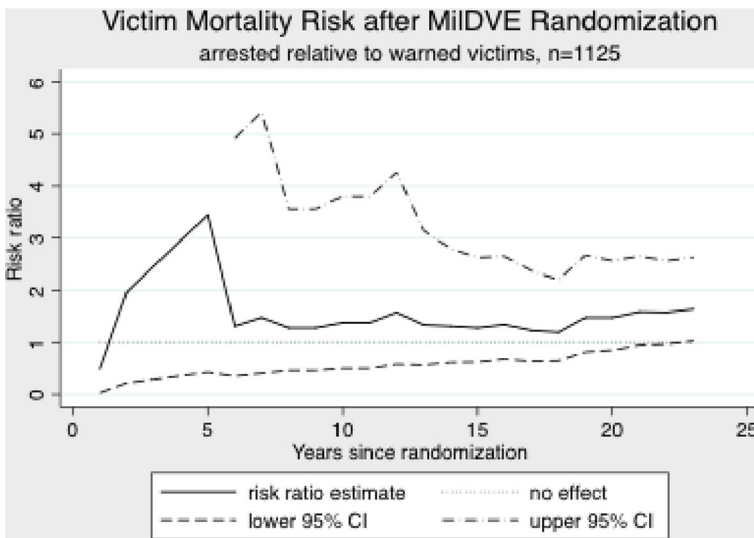
Fig. 3 Victim mortality rates by suspect arrest by year

identical for the two treatment groups at 2.7 per thousand. Differences in group death rates from cancer, alcohol, and drugs were also indiscernible.

Moderator analyses

We conducted 22 moderator analyses (Table 3), of which four (18 %) were significant, almost four times what would have been expected by chance. The moderator

Table 2 Main effects of partner arrest on victim mortality by cause of death

Cause	Arrest (n=754)		Warn (n=371)		Risk ratio	95 % CI	Fisher's exact two-sided p value	Cohen's d
	Dead (n)	Dead (%)	Dead (n)	Dead (%)				
All deaths*	70	9.28	21	5.66	1.6401	1.024-2.628	0.037	0.230
Homicide	2	0.27	1	0.27	0.9841	0.09-10.818	1.000	−0.008
Alcohol/drug-related	8	1.06	5	1.35	0.7873	0.259-2.39	0.768	−0.118
Cancer	7	0.93	3	0.81	1.1481	0.299-4.414	1.000	0.064
Heart disease	19	2.52	4	1.08	2.3372	0.801-6.821	0.122	0.339
Other internal causes	23	3.05	4	1.08	2.8292	0.986-8.121	0.060	0.397
Other external causes	5	0.66	1	0.27	2.4602	0.288-20.982	0.670	0.349
Unknown/ unspecified	6	0.80	3	0.81	0.9841	0.247-3.913	1.000	−0.008

*=significant effect

Table 3 Victim mortality after 23 years of suspect treatment and 22 moderators

Moderator	Risk ratio	95 % CI	Fisher's exact two-sided p	Cohen's d	N	Arrest N	Arrest Dead (n)	Warn N	Warn Dead (n)	Dead (%) Arrest	Dead (%) Warn
Main effect	1.6401	1.024-2.628	0.037	0.230	1,125	754	70	371	21	9.28	5.66
Male	2.7169	0.631-11.692	0.217	0.423	124	83	11	41	2	13.25	4.88
Female	1.5272	0.926-2.518	0.103	0.199	1,001	671	59	330	19	8.79	5.76
Black*	1.9811	1.099-3.572	0.019	0.305	791	529	52	262	13	9.83	4.96
White	1.0900	0.489-2.428	1.000	0.043	334	225	18	109	8	8.00	7.34
Couple never married	1.5517	0.9-2.675	0.131	0.208	788	523	49	265	16	9.37	6.04
Couple ever married	1.9174	0.743-4.946	0.192	0.284	335	230	21	105	5	9.13	4.76
Victim no prior	1.5950	0.903-2.816	0.120	0.221	829	537	44	292	15	8.19	5.14
Victim has prior	1.5776	0.675-3.689	0.397	0.202	296	217	26	79	6	11.98	7.60
Suspect no prior*	2.2917	1.079-4.869	0.034	0.386	406	252	30	154	8	11.91	5.20
Suspect has prior	1.3301	0.726-2.436	0.437	0.133	719	502	40	217	13	7.97	5.99
Suspect unemployed	1.5873	0.912-2.763	0.116	0.231	617	381	41	236	16	10.76	6.78
Suspect employed	2.5694	0.921-7.168	0.065	0.344	493	364	29	129	4	7.97	3.10
Victim unemployed	1.3529	0.817-2.24	0.290	0.148	792	530	52	262	19	9.81	7.25
Victim employed*	4.2589	1.007-18.02	0.027	0.508	330	224	18	106	2	8.04	1.89
Victim unemployed and black	1.3964	0.762-2.56	0.348	0.160	597	404	38	193	13	9.41	6.74
Victim unemployed and white	1.2778	0.514-3.175	0.806	0.126	195	126	14	69	6	11.11	8.70
Victim employed and black*	.	.	0.003	0.807	192	125	14	67	0	11.20	0.00
Victim employed and white	0.7879	0.154-4.129	1.000	-0.118	138	99	4	39	2	4.04	5.13
Victim unemployed, Suspect unemployed	1.4249	0.783-2.592	0.271	0.180	499	311	33	188	14	10.61	7.45
Victim unemployed, Suspect employed	1.5023	0.529-4.264	0.612	0.169	283	215	19	68	4	8.84	5.88
Victim employed, Suspect unemployed	2.6857	0.596-12.094	0.312	0.454	117	70	8	47	2	11.43	4.26
Victim employed, Suspect employed	.	.	0.065	0.668	208	149	10	59	0	6.71	0.00

*=significant moderator

candidates were selected for examination largely on the basis of previous moderators found for arrest effects on repeat domestic assault (Sherman et al. 1992; Sherman and Smith 1992). In this analysis, victim employment and race were both powerful moderators.

The overall 64 % difference between the arrest and warned group was almost entirely concentrated among the African-American victims (Fig. 4). The white sub-group had only a 9 % higher death rate after partner arrest than after a partner warning. For the African-American victims, the rate of death after a partner's arrest was 98 % higher than it was after a partner's warning ($p=.03$).

Further moderator analyses showed, in both races combined, that the association of partner arrest with victim mortality was much higher if the victim had a job (RR= 4.26:1; $d=-0.51$; $p=0.03$). That effect, however, was also entirely isolated among African-American victims, since there was no discernible treatment effect on mortality among white victims who were employed at time of enrolment. Not one of 67 employed black victims (0 %) whose partners were warned had died after 23 years, compared to 14 of the 125 employed black victims (11 %) whose partners had been arrested ($p=0.003$). That large effect size ($d=-0.81$) was the largest among all 22 moderator and main effects examined, and compares to a small effect size (but in the same direction) of suspect arrest for all other cases (Fig. 5).

A final significant moderator was among victims whose abuser had no prior arrest, which raised the risk of victim death after partner-arrest by 129 % over a partner-warning ($p=0.03$, $d=-0.39$).

Survival analyses

In order to increase precision in the estimate of our main effect, we estimated a Cox hazard model by adjusting for the predictive influence of three imbalanced covariates at baseline (Fig. 6). The adjustment was informed by Table 4, which shows the predicted effect of having each imbalance equalized. Based on the moderator analyses (Table 3), the arrested group had slightly more prior victim arrests, which lowered their predicted
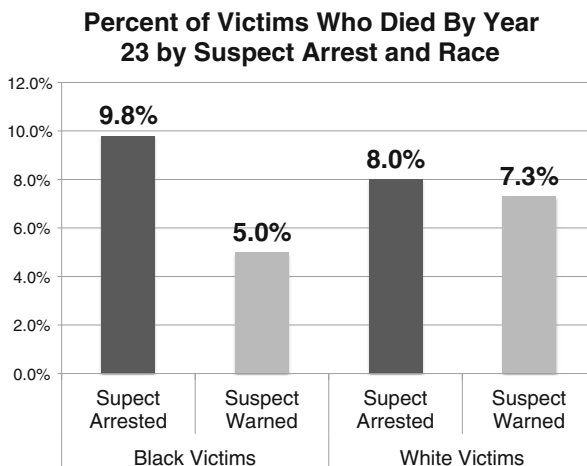

Fig. 4 Effect of suspect arrest on victim mortality by race of victim
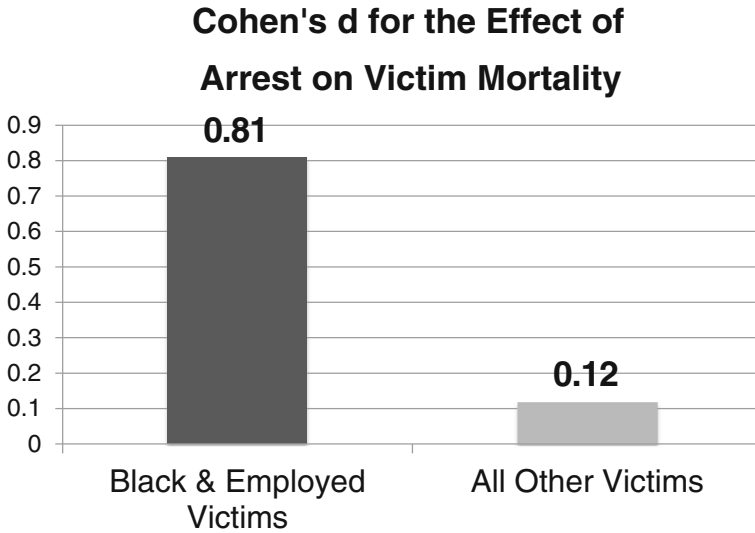
Cohen's d for the Effect of
Arrest on Victim Mortality

Fig. 5 Standardized mean difference (Cohen's *d*) of effect of suspect arrest on all causes of mortality for employed African-American victims vs. all other victims

mortality; more employed suspects, which raised predicted victim mortality; and more suspects with priors, which lowered predicted victim mortality. The results slightly increase the main effect of partner arrest on victim mortality over the unadjusted raw increase of 64 %, with the adjusted hazard rate showing victims in the partner-arrest group 85 % more likely to die than those in the partner-warned group. Table 5 performs further adjustments beyond baseline imbalances, incorporating key moderators including race and employment.
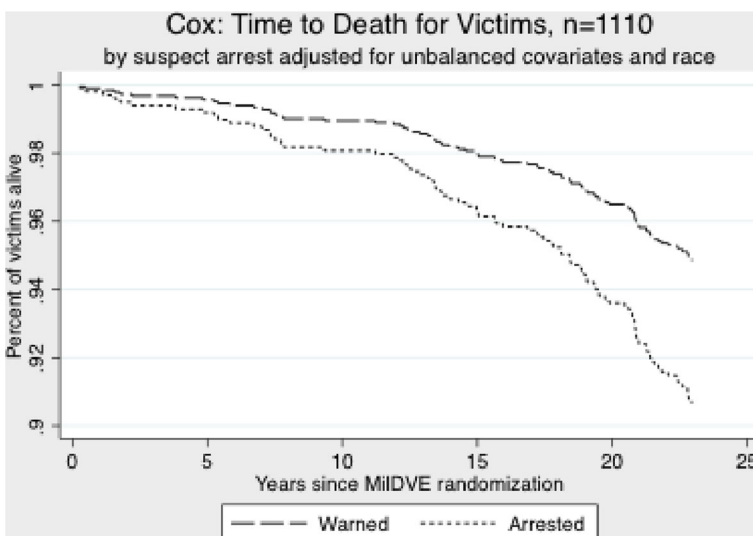

Fig. 6 Cox proportional hazard regression models with unbalanced covariates and race

Table 4 Adjusted Cox proportional hazard regression models for unbalanced covariates

Variable	Hazard ratio	Robust SE	p value
Suspect arrested	1.8496	0.4780	0.02
Victim has prior arrest	1.5042	0.3447	0.08
Suspect employed	0.6613	0.1500	0.07
Suspect has prior arrest	0.6819	0.1507	0.08

Table 5 Cox proportional hazard regression models with unbalanced covariates, race, and gender

	[1]	[2]	[3]	[4]	[5]
Suspect arrested					
Hazard ratio	1.8496	1.8503	1.8517	1.8524	1.8204
Robust SE	0.478	0.478	0.479	0.478	0.472
p value	0.017	0.017	0.017	0.017	0.021
Victim has prior					
Hazard ratio	1.5042	1.5061	1.4701	1.4720	1.4342
Robust SE	0.345	0.345	0.330	0.330	0.317
p value	0.075	0.074	0.086	0.085	0.103
Suspect has prior					
Hazard ratio	0.6819	0.6828	0.6926	0.6935	0.6861
Robust SE	0.151	0.151	0.158	0.159	0.157
p value	0.830	0.085	0.108	0.110	0.100
Suspect employed					
Hazard ratio	0.6613	0.6573	0.6622	0.6583	0.7080
Robust SE	0.150	0.156	0.151	0.157	0.174
p value	0.068	0.077	0.070	0.079	0.161
Victim race					
Hazard ratio		1.0359		1.0349	1.0864
Robust SE		0.254		0.254	0.272
p value		0.886		0.889	0.741
Victim sex					
Hazard ratio			0.8893	0.8933	0.8088
Robust SE			0.273	0.274	0.252
p value			0.711	0.713	0.495
Victim employed					
Hazard ratio					0.6925
Robust SE					0.196
p value					0.194

Randomly assigned time in custody

Table 6 presents the main effects on victim mortality separately by all three randomly assigned treatments, including the two different randomly assigned lengths of time in custody for arrested partners. The results show higher mortality for the standard length of time in custody than for the artificially reduced time. Analyzing only the "long" arrest condition against the warn condition, the raw relative risk ratio is 1:1.87, compared to only 1:1.40 for the "short" arrest effect.

Other analyses

We used Stata to calculate the population attributable fraction of deaths (Rockhill et al. 1998) related to arrest as a risk factor in this analysis. In the absence of arrest, 30 % of the observed mortality in this population over 23 years would have been avoided, holding all other factors constant. For black victims, 40 % of deaths would have been avoided; for whites it would have been 6 %.

    Finally, since death was relatively rare in the sample overall, we performed a permutation test as a sensitivity analysis. This test shuffled the time order of deaths 1,125 times (creating 1,125 new variables). We then ran the Cox model 1,125 times and saved each outcome. Figure 7 plots the actual order of deaths when time is real in relation to the order when time is shuffled. It shows that the point estimate of arrest effects in the actual sequence (RR 1:1.85) lies in the far right tail of the distribution of all estimates. This result provides strong evidence that the relationship between victim death and partner arrest is causal, rather than a merely chance difference between the groups that was not overcome by random assignment with very high treatment fidelity and little potential attrition.

Discussion

The analysis presented herein is an inductive exploration, not a deductive hypothesis test. We did not predict our findings, and were surprised by their direction and magnitude. This is not a reason to dismiss them: similar serendipity has characterized important scientific discoveries that withstood further testing, such as the effect of penicillin on infections (Haven 1994). Nevertheless, the present findings lack an

Table 6 Short versus full arrest effects on victim mortality, all causes of death

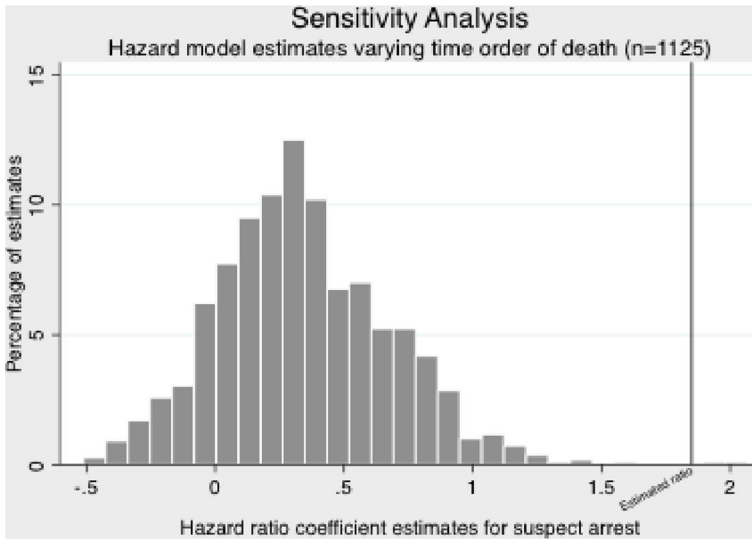| Variables	Assigned (N)	Deaths (N)	Deaths (%)	p value	Relative risk ratio
Warn	371	21	5.66		
Short arrest	377	30	7.95		
Full arrest	377	40	10.61		
Short vs. full arrest				0.26	1.33
Warn vs. short arrest				0.25	1.40
Warn vs. full arrest				0.02	1.87

Fig. 7 Permutation test distribution of 1,125 estimates, fluctuating the time order of death

empirical context: They are both the result of the longest-ever follow-up of a randomized trial of a criminal justice sanction *and* the first study to examine the victim-mortality effects of suspect arrest for domestic violence.

This discussion seeks to provide a larger context for these findings. Doing so requires that three key issues be addressed. One is the statistical risk that the findings are unique to this sample, and not indicative of a more general pattern. The second is the theoretical pathway by which these results can be understood, especially the differences between black and white victims and by victim employment status. The third is the potential policy implications of these findings, both for Wisconsin and elsewhere.

Statistical interpretation

While our results remained robust through multiple sensitivity tests, it remains possible that they are not causal. Imbalances in *measured* covariates suggest that, despite random assignment, our results could contain spurious imbalances in *unmeasured* factors predisposing mortality. Additionally, our findings may not generalize beyond the Milwaukee sample, even if they generalize fully to the Milwaukee population from which they were sampled. Fortunately, because similar experiments were conducted at the same time, our analytic process might be replicable in three more samples (Charlotte, Miami, and Omaha). Those replications, if accomplished, would provide a valuable check on the external validity of our results.

Despite the robustness of our results, some may question whether random assignment can plausibly "last" so long and over a life-course in which other causes of death could intervene. For example, this analysis does not examine common proximate causes of death, such as cigarette smoking and obesity, but there is no time limit on the causal interpretation of randomized experiments. Whatever the proximate causes of death may be, they are logically connected to differences in causal pathways influenced

by the exogenous instrument of random assignment itself. If, for example, the arrest group developed more obesity or a higher prevalence of smoking by year 23, such differences are likely to have been shaped by a causal pathway attributable to the randomly assigned treatment.

Theoretical interpretation

These results pose three related challenges of explanation:

1. By what causal pathways can arresting a suspect cause a victim to die?
2. What could explain why those pathways produced such different results for black and white victims?
3. Why does victim employment double the lethal effect of suspect arrest among black victims, but not among white victims?

Causal pathways Currently, only hypotheses, not conclusions, can be drawn regarding the particular causal pathways that may have extended from the arrest of MilDVE suspects to the deaths of their victims. We suggest the interactions between arrest, race, and employment provide clues as to the direction those causal pathways took. We also suggest that those clues point toward differential post-traumatic stress manifestations in victims whose partners were arrested.

Extensive studies of mortality and employment hierarchy in the British civil service demonstrate that a large portion of the variation in mortality remains unexplained by smoking, alcohol, or biological risk factors (Marmot et al. 1991, 1997; Kuper and Marmot 2003; Kunz-Ebrecht et al. 2004). That variance can, however, be explained by psychosocial causes like the degree of autonomy people have in their work, the extent of their social isolation (Stringhini et al. 2012), and depression, which could result from an arrest that increases stress by limiting current or even future intimate and other social relationships.

By the last wave of MilDVE victim interviews (77 % response rate at 6 months after enrolment), a large minority of victims in all three arms (45 % long-arrest; 44 % short-arrest; 38 % warned) had already separated from their partners (Sherman 1992). Yet, with no subsequent interviews, we cannot know whether arrest had long-term effects on their then-current or future relationships, on their loneliness and social isolation, or on any other potential stressor on a causal pathway to higher mortality. A new wave of MilDVE victim interviews could explore those stressors and causal pathways by examining, for example, whether post-traumatic stress differences between victims whose partners are warned or arrested create a biochemical pathway that reduces victims' life-spans.

Given evidence of elevated levels of post-traumatic stress symptoms among women victims of domestic violence in general (Silva et al. 1997; Golding 1999), the added trauma of an arrest could be more dangerous than witnessing an arrest at lower baseline levels of PTSS. Chronic but low-to-moderate elevation of post-traumatic stress symptoms have been reported to have a strong prospective link to premature mortality from coronary heart disease (Kubzansky et al. 2007; Boscarino 2006), possibly via hyperactivity of the central CRH [corticotrophin-releasing hormone] systems with underactivity of the pituitary-adrenal axis (Kasckow et al. 2001). If some domestic violence

victims experience partner arrest as traumatic, that stress-related response chain could trigger an increase in coronary heart disease and other morbidity leading to premature death. As a first step toward examining the viability of this pathway as an explanation of our results, new interviews could use PTSS assessment instruments to determine whether victims in the partner-arrest group have higher current levels of PTSS than those in the partner-warned group. Even better would be new experiments in which PTSS measures were taken within a few weeks after random assignment.

Explaining racial differences A major obstacle to applying a PTSS theory to these results lies with the racial differences. For the PTSS theory to fit the facts, African-American domestic victims should presumably have higher baseline levels of PTSS than white victims of domestic violence, on average. Yet, even a brief examination of the literature shows, at worst, that black domestic violence victims suffer *less* PTSS than whites, evincing greater resilience in relation to numerous symptoms and measures of PTSS. At best, the literature, which includes only a handful of studies, is mixed (Dutton 2009; Lily and Graham-Bermann 2009; Wright et al. 2010; Iverson et al. 2013).

A careful examination of that literature, however, shows that those findings may be confounded with employment because unemployed African-American domestic violence victims dominated the samples. Not all of the studies report employment data, but in the two that did unemployment amongst black domestic violence victims was greater or statistically similar to high unemployment among whites. In one study of 132 domestic violence victims in southern Michigan, in which 31 % of respondents were black and 59 % white, 63 % of the black domestic violence victims were unemployed, whereas only 31 % of the white victims were unemployed (Lily and Graham-Bermann 2009). In the second study of 204 residents in a domestic violence shelter, unemployment was slightly, although not significantly, higher for white (76 %) than for black (72 %) victims (Wright et al. 2010). In the MilDVE sample, unemployment predominated amongst all victims at 70.6 %, but varied considerably by race. 75.5 % of black victims were unemployed, whereas only 58.4 % of white victims were. Statistically speaking, this significant racial difference in employment opens the door for a theoretical explanation that may indicate a paradox of *social context* of the kind that Sampson (2013) calls "contextual causality."

The African-American victims in the MilDVE resided in the context of the most hyper-segregated metropolitan area in the US (Massey and Denton 1993). Most of the black victims lived in areas of concentrated unemployment, while whites lived in mixed income neighborhoods dominated by employment rather than unemployment (Wilson 1981, 1996). This social context was also associated with a divergence in the structure of intimate partner relationships between blacks and whites. Fertility in the absence of marriage was more prevalent and accelerating far faster among black women than among white women (Wilson 1981; Wojtkiewicz et al. 1990). These contextual differences within the MilDVE victim sample may have led to equally divergent identity constructs that Sampson (2013) calls "identities of place" (p. 2) that could theoretically create a paradox in the relationship between race, employment, and victims' vulnerability to PTSS mediating their risk of death in response to arrest.

The paradox would be present if there was greater resilience (and less PTSS) among black victims who were *unemployed* than among those who were *employed*, and if

unemployed black victims also showed greater resilience than white victims whether they were employed or not. That paradox may relate to the extent to which employed African-American women, in particular, are more vulnerable to stress stemming from a threat to their livelihood or a change in their status and identity resulting from the threat of loss of employment, in comparison to unemployed women. The latter not only had no job to lose; many could also predictably depend on welfare benefits (such as aid for dependent children) for their financial support. Only a minority of women in Milwaukee's concentrated poverty neighborhoods were employed "strivers," pursuing what Anderson (1999) calls a "decent" code of conduct. They could reasonably have feared a far greater loss of status and respectability from their partner being arrested than most of the unemployed women, who had a different "code" or self-defined identity unrelated to employment.

For employed white women in a neighborhood context of relatively high employment, the impact of a partner's arrest could have been very different. White victims were more likely to have ever been married to their abusers than were black women (42 vs. 26 %), with higher rates of employment among their partners. In particular, only 16 % of the partners of employed black victims also had jobs, whereas 40 % of the partners of employed white victims, 38 % of the partners of unemployed black victims, and 30 % of the partners of unemployed white victims had jobs. White women were therefore more likely to have been second earners, whose own employment provided a buffer against a husband's or partner's loss of work, rather than the sole means of support in a single-parent household, as appears to have been the case for employed black women. The psycho-social (and possibly biomedical) meaning of employment and its contribution to the identities of married white women, living in predominantly white neighborhoods of higher employment, could be quite different from that for employment of unmarried African-American women in struggling neighborhoods. Employment for white victims could even provide a protective effect against PTSS in response to a partner's arrest. Unemployment for white victims, however, could leave them more economically vulnerable to a loss of partner's earnings—or even the partner entirely—in the event of an arrest, thus increasing their stress levels.

A race–employment paradox? The possibility that employment for black domestic violence victims may lead to more PTSS from partner arrest, but make white victims more resilient, implies that their death rates should correspond to this paradox of employment and identity. Arrest should cause higher death rates for employed (vs. unemployed) black victims, but have the opposite interaction for white victims, regardless of their employment. As Table 3 shows, that is generally the case. For employed black victims whose suspects were arrested, the 23-year death rate per 1,000 was 112; for unemployed black victims, the comparable figure was 94—less, but admittedly not lower than for employed white victims. For white unemployed victims, the death rate when partners were arrested was 111 per 1,000, almost the same as for employed black victims; for white employed victims, it was only 40, and was actually lower for those whose partners were arrested than for those whose partners were warned (which was 50).

A further nuance in the findings is the different rates of death by race and employment after partners are *warned*. For unemployed white women, the death rate after partner warning was close to their death rate after partner arrest (87 vs. 111 per 1,000), but for

employed black women, the randomly assigned warning was a lifesaver: not one of them had died after their partner was warned. If differences in neighborhood context help shape different death rates (by race) from similar treatments, then these findings may be a prime example of Sampson's (2013) concept of "contextual causality."

Although the results from this analysis failed to provide perfectly symmetrical support for it, the race employment paradox may still be true. It is simply one hypothesis for which the evidence provides limited support. Suffice it to say that evidence on PTSS among black and white victims of different employment statuses remains quite thin, and future evidence may still show that PTSS levels differ in direct proportion to the differences in death rates by demographic subgroups, consistent with differences in social context.

There are two ways to continue to investigate and develop these theoretical possibilities. One is gathering new evidence in the field from the Milwaukee sample (and the samples of the companion experiments). The other is to analyze the data not by race but by neighborhood context. In further analyses we will examine the Census tract data for all victims, not only at random assignment but even at later periods of the follow-up, as may be indicated by change of address data from the Wisconsin drivers' license records we have obtained.

Policy implications

Regardless of the reasons for the racial disparity we observe, the evidence is clear: African-American victims of domestic violence are disproportionately likely to die after partner arrests relative to white victims. The magnitude of the disparity strongly indicates that mandatory arrest laws, however well-intentioned, can create a racially discriminatory impact on victims. While 6 % of the deaths amongst the white partners of arrested suspects could have been avoided, 40 % of the deaths amongst the black partners of arrested suspects could have been avoided, if only their partners had been warned instead.

Wisconsin state law requires police to make arrests when they have probable cause to believe that a misdemeanor assault has occurred in a domestic relationship, irrespective of race. This law was not intended to place black victims at a disadvantage to white victims, or to deprive them of life. Like many other "race-neutral" policies, however, this policy generates racially disparate outcomes. A key question is whether the unintended potential harm generated by that disparity is outweighed by the intended potential benefit of the policy.

In the realm of labor law, employment policies often require a high school diploma or passage of tailored exams to qualify for certain jobs. Those policies exemplify seemingly race-neutral policies that have resulted in racial disparity. In considering whether those policies could be considered justified, the US Supreme Court has focused on whether the policy under examination has sufficient evidence to justify its use as an employment criterion (e.g., *Griggs v. Duke Power Co.*, 401 U.S. 424, 1971; *Ricci v. DeStefano*, 557 U.S. 557, 2009).

In the realm of criminal law, there is no comparable body of US case law regarding unintentional disparate impacts of criminal justice policies on victims of crime. In fact, there has been no prior evidence of victim mortality differences apparently caused by

any criminal sanction. Thus, it is unclear how US federal courts would rule if a Constitutional challenge to the Wisconsin mandatory arrest statute were filed by surviving families of deceased African-American domestic violence victims whose partners had been arrested.

Since 1871, Section 1983 of the U.S. Code has prohibited states from depriving people of certain rights "under color of" statutes (or other acts or omissions). This code was enacted to enforce the 14th Amendment to the U.S. Constitution, which says (in part) that "No State shall make or enforce any law which shall… deny to any person within its jurisdiction the equal protection of the laws." Whether equal protection is defined by equal enforcement or equal effect is a theme of increasing discussion by legal scholars, especially with respect to racially disparate effects of incarceration rates and the collateral consequences of incarceration. This has led to proposals for "demographic impact statements" regarding sentencing of offenders and to one state (Minnesota) reconsidering a potential change to its sentencing regime based on the projection that it would differentially impact African-Americans (Reitz 2009).

That legislatures should consider the potential impact of legislation before enacting new sentencing laws could apply equally to repealing existing arrest laws. The present article can be seen as an example of what could be called a "*demographic* victim impact statement," as distinct from the case-by-case use of "victim impact statements" before sentencing (e.g., Davis and Smith 1994). Such statements are also relevant to proposals to repeal existing laws governing police conduct.

Perhaps the key fact in any "balancing test" of evidence for or against a law that legislatures or courts might consider is the lack of any discernible evidence of a longevity benefit for victims, caused by the arrest of abusive partners. Using the stringently unambiguous criterion of victim life expectancy, none of the 22 moderator analyses found any significant increase in longevity associated with arrest. Earlier findings on short-term misdemeanor recidivism (Pate and Hamilton 1992; Sherman and Smith 1992) had demonstrated opposite effects for different subgroups, with victims of employed abusers gaining a benefit of less violence from arrest, at the cost of victims of unemployed abusers experiencing a higher rate of domestic violence if the abusers had been arrested. When death is the test of benefit, the result is unambiguous: no victims appear to benefit.

The potential benefit of a general deterrent effect is much harder to examine, and remains untested with respect to overall mortality of domestic violence victims. Other benefits, such as moral condemnation of domestic violence, also remain untested. If nothing else, the present findings may justify a shift in the burden of proof onto those who would prefer to see mandatory arrest continued. For those who caution no policy change (at least in Wisconsin) until there is more evidence, we must recall that the evidence on which the Wisconsin statute was based was a 6-month follow-up of an experiment in Minnesota (Sherman and Berk 1984). For reasons of both internal and external validity, the present study covering 23 years arguably provides much more policy-relevant evidence for Wisconsin.

Whatever effect these findings may have on future research or policy decisions, they provide strong evidence for proposals to test criminal sanctions in the same way that medical treatments are tested. Evidence from controlled trials indicating a substantial cause of premature death for African-American victims should be taken seriously unless and until further strong evidence shows otherwise.

References

Anderson, E. (1999). *Code of the street: decency, violence, and the moral life of the inner city.* New York: W.W. Norton.

Berk, R. A., Campbell, A., Klap, R., & Western, B. (1992). The deterrent effect of arrest in incidents of domestic violence: a Bayesian analysis of four field experiments. *American Sociological Review, 57,* 698–708.

Boscarino, J. A. (2006). Posttraumatic stress disorder and mortality among US Army veterans 30 years after military service. *Annals of Epidemiology, 16,* 248–256.

Campbell, J. C., Webster, D., Koziol-McLain, J., Block, C., Campbell, D., Curry, M. A., et al. (2003). Risk factors for femicide in abusive relationships: results from a multisite case control study. *American Journal of Public Health, 93,* 1089–1097.

Davis, R. C., & Smith, B. E. (1994). The effects of victim impact statements on sentencing decisions: a test in an urban setting. *Justice Quarterly, 11*(3), 453–469.

Dutton, M. A. (2009). Pathways linking intimate partner violence and posttraumatic disorder. *Trauma, Violence & Abuse, 10,* 211–224.

Federal Judicial Center. (1981). *Experimentation in the law: Report of the Federal Judicial Center Advisory Committee on experimentation in the law.* Washington: Advisory Committee on Experimentation in the Law.

Golding, J. M. (1999). Intimate partner violence as a risk factor for mental disorders: a meta-analysis. *Journal of Family Violence, 14,* 99–132.

Haven, K. F. (1994). *Marvels of Science.* Littleton: Libraries Unlimited. ISBN 1-56308-159-8.

Hill, M. E., & Rosenwaike, I. (2001). The social security administration's death master file: the completeness of death reporting at older ages. *Social Security Bulletin, 64,* 45–51.

Iverson, K. M., Bauer, M. R., Shipherd, J. C., Pineles, S. L., & Harrington, E. F. (2013). Differential associations between partner violence and physical health symptoms among Caucasian and African-American help-seeking women. *Psychological Trauma, Theory, Research, Practice, and Policy, 5,* 158–166.

Kasckow, J. W., Baker, D., & Geracioti, T. D., Jr. (2001). Corticotropin-releasing hormone in depression and post-traumatic stress disorder. *Peptides, 22,* 845–851.

Kubzansky, L. D., Spiro, A., Sparrow, D., Koenen, K., & Vokonas, P. (2007). Prospective study of posttraumatic stress disorder symptoms and coronary heart disease in the normative aging study. *Archives of General Psychiatry, 64,* 109–116.

Kunz-Ebrecht, S. R., Kirschbaum, C., Marmot, M., & Steptoe, A. (2004). Differences in cortisol awakening response on work days and weekends in women and men from the Whitehall II cohort. *Psychoneuroendocrinology, 29,* 516–528.

Kuper, H., & Marmot, M. (2003). Job strain, job demands, decision latitude, and risk of coronary heart disease within the Whitehall II study. *Journal of Epidemiology and Community Health, 57,* 147–153.

Lily, M. M., & Graham-Bermann, S. A. (2009). Ethnicity and risk for symptoms of posttraumatic stress following intimate partner violence. *Journal of Interpersonal Violence, 24*(1), 3–19.

Marmot, M. G., Bosma, H., Hemingway, H., Brunner, E., & Stansfeld, S. (1997). Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *The Lancet, 350,* 235–239.

Marmot, M. G., Stansfeld, S., Patel, C., North, F., Head, J., White, I., et al. (1991). Health inequalities among British civil servants: the Whitehall II study. *The Lancet, 337,* 1387–1393.

Massey, D., & Denton, P. (1993). *American apartheid.* Cambridge: Harvard University Press.

Maxwell, C. D., Garner, J. H., & Fagan, J. A. (2002). The preventive effects of arrest on intimate partner violence: research, policy and theory. *Criminology & Public Policy, 2,* 51–80.

Mills, L. G. (2003). *Insult to injury: Rethinking our responses to intimate abuse*. Princeton: Princeton University Press.

Pate, A. M., & Hamilton, E. E. (1992). Formal and informal deterrents to domestic violence: The Dade County spouse assault experiment. *American Sociological Review, 57*, 691–697.

Reitz, K. R. (2009). Demographic impact statements, O'Connor's warning, and the mysteries of prison release: topics from a sentencing reform agenda. *Florida Law Review, 61*, 683–703.

Rockhill, B., Newman, B., & Weinberg, C. (1998). Use and misuse of population attributable fractions. *American Journal of Public Health, 88*, 15–19.

Sampson, R. J. (2013). The place of context: a theory and strategy for criminology's hard problems. The 2012 ASC presidential address. *Criminology, 51*(1), 1–31.

Sherman, L. W. (1992). *Policing domestic violence: Experiments and dilemmas*. New York: Free Press.

Sherman, L., & Berk, R. (1984). The specific deterrent effects of arrest for domestic assault. *American Sociological Review, 49*, 261–272.

Sherman, L. W., & Harris, H. M. (2013). Increased homicide victimization of suspects arrested for domestic assault: a 23-year follow-up of the Milwaukee Domestic Violence Experiment (MilDVE). *Journal of Experimental Criminology, 9*, 491–514.

Sherman, L. W., & Smith, D. A. (1992). Crime, punishment and stake in conformity: legal and informal control of domestic violence. *American Sociological Review, 57*, 680–690.

Sherman, L. W. & Strang, H. (1996). *Policing domestic violence: The problem-solving paradigm.* Canberra: Australian Institute of Criminology. http://www.aic.gov.au/media_library/aic/rjustice/rise/sherman-strang.pdf

Sherman, L. W., Schmidt, J. D., Rogan, D. P., & Smith, D. A. (1992). Variable effects of arrest on criminal careers: The Milwaukee Domestic Violence Experiment. *Journal of Criminal Law & Criminology, 83*, 137.

Sherman, L.W. (1991). Predicting Domestic Homicide: Prior Police Contact and Gun Threats. pp. 73–94 in Michael Steinman, ed., *Woman battering: Policy responses*. An Academy of Criminal Justice Sciences Book. Cincinnati, Oh.: Anderson Publishing.

Silva, C., McFarlane, J., Soeken, K., Parker, B., & Reel, S. (1997). Symptoms of posttraumatic stress disorder in abused women in a primary care setting. *Journal of Women's Health, 6*, 543–552.

Stringhini, S., Berkman, L., Dugravot, A., Ferrie, J. E., Marmot, M., Kivimaki, M., et al. (2012). Socioeconomic status, structural and functional measures of social support, and mortality: the British Whitehall II Cohort Study, 1985–2009. *American Journal of Epidemiology, 175*, 1275–1283.

Wilson, W. J. (1981). The black community in the 1980s: questions of race, class, and public policy. *Annals of the American Academy of Political and Social Science, 454*, 26–41.

Wilson, W. J. (1996). *When work disappears: The world of the new urban poor*. NY: Alfred A. Knopf.

Wojtkiewicz, R. A., McLanahan, S. S., & Garfinkel, I. (1990). The growth of families headed by women: 1950–1980. *Demography, 27*, 19–30.

Wright, C. V., Perez, S., & Johnson, D. M. (2010). The mediating role of empowerment for African-American women experiencing intimate partner violence. *Psychological Trauma: Theory, Research, Practice and Policy, 2*(4), 266–272.

Lawrence W. Sherman is Distinguished University Professor in the Department of Criminology and Criminal Justice at the University of Maryland and Wolfson Professor of Criminology at the University of Cambridge, where he directs the Jerry Lee Centre of Experimental Criminology.

Heather M. Harris is currently a doctoral candidate in the Department of Criminology and Criminal Justice at the University of Maryland. Her previous publications addressed the effect of arrest on death rates of misdemeanor domestic violence offenders, and the idea of "collateral damage" by the criminal justice system.