A Reassessment of the D.C. Gun Law: Some Cautionary Notes on the Use of Interrupted Time Series Designs for Policy Impact Assessment

Chester L. Britt
Gary Kleck
David J. Bordua

Interrupted time series designs are commonly used to assess the impact of gun control legislation, as well as other legal and policy changes. Three common problems in the use of these designs—(1) selection of an appropriate control series, (2) specification of the intervention model, and (3) specification of the time series studied—raise questions about the validity of the conclusions reached in research on the impact of gun control. We illustrate these problems with a critical reassessment of Loftin et al.'s (1991) evaluation of the 1976 District of Columbia Gun Law. We then use monthly homicide data from the Federal Bureau of Investigation and the National Center for Health Statistics to illustrate how careful consideration of these three design issues results in a significantly different conclusion about the effectiveness of the District of Columbia Gun Law.

Interrupted time series designs provide one of the most common means for assessing the impact of a change in law or in social policy. Recently, interrupted time series designs have been used to assess the impact of changes in drunk driving legislation on the incidence of driving under the influence (e.g., Ross et al. 1990), of changes in gun control legislation on homicides (e.g., McDowall et al. 1992), of changes in plea bargaining policy on court caseloads (Holmes et al. 1992), of executions on homicides (Cochran et al. 1994), and of Supreme Court decisions on police use of deadly force (Tennenbaum 1994). The key characteristic to this form of analysis is repeated observations of a variable (often referred to as the target variable) before and after some form of "intervention." The analytical goal is to test for a shift in

We would like to thank Michael Gottfredson, Richard McCleary, and David McDowall for their input on earlier versions. A fuller version of this paper was presented at the 1993 Annual Meetings for the American Society of Criminology, Phoenix, AZ, and can be obtained from Gary Kleck (Kleck, Britt, & Bordua 1993); write to him at School of Criminology and Criminal Justice, Florida State University, Tallahassee, FL 32306. Address other correspondence to Chester L. Britt, Crime, Law and Justice Program, Department of Sociology, Pennsylvania State University, University Park, PA 16802.

the level of the target variable at the point when a new law or a policy change became effective.

Our concerns in this article are twofold. First, we hope to make policy researchers aware of the relative strengths and weaknesses of the interrupted time series design as it has been used to test for the impact of gun control legislation (see Table 1 for a list of important gun control studies that have used interrupted time series designs). Following a brief description of the interrupted time series design, we highlight what we see as three key methodological issues that emerge from the research literature on the impact of gun control legislation. Second, and more directly, we illustrate these concerns with a reassessment of Loftin et al.'s (1991) evaluation of the impact of the 1976 District of Columbia gun law on homicides in Washington.

Interrupted Time Series Designs

An extensive methodology literature documents the widespread use of interrupted time series designs for testing policy impact (see, e.g., Campbell & Stanley 1966; Cook & Campbell 1979). Interrupted time series designs come in two basic forms: single and multiple. Single interrupted time series designs involve taking repeated observations on some group (e.g., a classroom or a city) at regular intervals for a specified period of time. Multiple interrupted time series designs involve taking repeated observations on two or more groups at regular intervals for a specified period of time. In practice, at least one of these groups should not be exposed to the intervention. Classic methodology texts (Campbell & Stanley 1966 and Cook & Campbell 1979, among others) have argued that interrupted time series designs (both single and multiple) provide some of the strongest quasi-experimental designs available to researchers.

Interrupted time series designs are generally viewed as having strong internal validity but weak external validity. Because they involve the taking of multiple observations on the same population for an extended period, these designs are able to rule out a variety of hypotheses that the (un)observed effect is not an artifact of the research design. The remaining threat to the internal validity of the single interrupted time series design is history—it is impossible to rule out the possibility that some unmeasured and uncontrolled factor is responsible for the observed change in the target variable. For example, if the target variable correlates with the business cycle and a change in the business cycle coincided with the intervention, a change would be observed in the target variable, even though the intervention may have had no effect. Multiple interrupted time series designs are less susceptible to this threat to internal validity because there are observations on the same target variable in an area not subject to the intervention. The use of an appropriate control group or control site should rule out historical effects if the unmeasured factors that influence the target variable (e.g., the business cycle) operate the same way in different groups or areas.

Although the internal validity of interrupted time series designs may be strong, there is weak external validity—the results may not be generalizable to other groups or areas. The reasons for this are twofold. First, there will be an interaction between the selection of the observation area and the intervention. An area will likely only be studied after it has made some legal or social policy change. The area is not randomly selected to implement the policy change, and in all likelihood, there will be unique social and historical circumstances responsible for the change in policy. Second, policies are rarely adopted in toto by other jurisdictions. There may be broad similarities, for example, in increased penalties for drunk driving offenders, but the penalties for each jurisdiction may have some differences. Thus, if different patterns are found in areas that implemented similar but not identical policies, it cannot be known why the observed pattern appeared.

Ideally, the use of interrupted time series designs should provide us with a strong means for evaluating the impact of legal and social policy changes on some population's behavior. However, we have observed three common problems in the use of these designs testing for the impact of gun control legislation that raise questions about the validity of substantive conclusions reached in this area of research. The first problem concerns the selection of control series. One of the strengths of the multiple interrupted time series design is the ability to rule out competing hypotheses by using time series data from a group that was not exposed to the intervention. Unfortunately, many applications of the interrupted time series design to the study of gun control legislation either use no control series or use inappropriate control series (see Table 1), raising the possibility of incorrect conclusions. The second problem concerns the specification of the intervention-impact model. Does the law have an immediate and permanent impact, an immediate but temporary impact, or a gradual but permanent impact? Theory, prior research, and an understanding of the law may suggest a certain specification, but many applications of interrupted time series design have focused solely on the statistical evidence (i.e., the model that best fits the observed data). The third problem concerns the specification of the time series. Often, the length of the time series analyzed is determined solely by data availability, and it is rare for researchers to investigate whether differing substantive results would be achieved with use of a time series with different start and end points or of a time series of a different length. In short, we are concerned with the robustness of the results. Presumably, if an intervention has a meaningful effect on some targeted behavior, then minor changes in the length of the time series should not substantially alter the apparent intervention effect.

Table 1. Major Interrupted Time Series Evaluations of the Impact of Gun Control Laws

Study	Intervention Location	Date of Intervention	Control Series[a] Nongun	Control Series[a] Other Areas	Type of Interveition
Deutsch & Alt (1977)	Boston	4/1/75	No	No	Mandatory penalty for unlawful carrying
Hay & McCleary (1979)	Boston	4/1/75	No	No	Mandatory penalty for unlawful carrying
Pierce & Bowers (1981)	Boston	4/1/75	No	No	Mandatory penalty for unlawful carrying
Loftin et al. (1983)	Detroit	1/1/77	Yes	No	Mandatory 2-year add-on penalty for felony with gun
Loftin & McDowall (1984)	3 Florida cities	10/1/75	Yes	No	Mandatory minimum 3 years for gun possession during felonies
Loftin et al. (1991)	Washington, DC	9/24/76	Yes	Yes	Ban on handgun possession, with "grandfather clause"
McDowall et al. (1992)	Detroit Jacksonville Tampa Miami Pittsburgh Philadelphia	1/1/77 10/1/75 10/1/75 10/1/75 6/1/82 6/1/82	Yes	No	Mandatory add-on penalties for committing crimes with guns
McPheters et al. (1984)	2 Arizona counties	8/1/74	No	No[b]	Mandatory minimum sentence for robbery with a deadly weapon
O'Carroll et al. (1991)	Detroit	1/10/87	Yes	No	Mandatory penalty for unlawful carrying

NOTE: Studies covered used ARIMA analytic methods. Simple before-and-after comparisions (e.g., Zimring 1975; Lucas & Ledgerwood 1978; Fife & Abrams 1989) are not included.
[a] Was gun crime series compared with corresponding nongun series (e.g., gun homicide compared with nongun homicides)? Was series in intervention area compared with series in nonintervention area?
[b] Control area was used for paired t-tests but not for ARIMA analyses.

Methodological Issues

Selection of Control Series

The strength of the multiple interrupted time series design lies in the use of one or more control series, allowing the researcher to rule out some historical effects and to test whether the intervention had a real and measurable effect on the target variable. For the researcher to conclude the intervention had an effect, there should be no similar pattern observed in the control series. If the intervention site shows a significant change in the target variable, then the control site should show either no effect or an opposite effect. If similar and significant effects are observed in both the intervention and control sites, the ability to conclude the intervention had a meaningful impact is weakened.

For the multiple interrupted time series designs to have strong internal validity, the characteristics of the control group and its similarity to the intervention group are crucial. Campbell and Stanley (1966:47) note that the control and the intervention groups should "constitute naturally assembled collectives . . . as similar as availability permits." In policy impact studies, this implies that the intervention site should be as similar as possible to the control site. The practical difficulty raised here is the determination of whether two or more cities are similar enough to justify a comparison. The strongest test for an intervention effect is given when there is both cross-sectional and cross-temporal similarity between the intervention and control areas. For example, if two matched cities were identical in every respect at the 1980 Census, yet the intervention city was trending downward in crime before the intervention while the control city was trending upward prior to the intervention, it would be much more difficult to establish that a post-intervention drop in crime in the intervention city was not an artifact of the pre-intervention trends. However, if one needs to choose between cross-sectional and cross-temporal similarity, Campbell and Stanley suggest that cross-sectional similarity is relatively more important, since there are ways to adjust statistically for different trends in the two sites prior to the intervention. The result should be a fair and rigorous test of whether the intervention had an impact on the target variable.

Relatively few gun control studies have used control areas, and in those studies which have included a control site (see Table 1), the underlying logic for its selection has rarely been made explicit. This problem is perhaps best illustrated by Loftin et al.'s (1991) study of the District of Columbia gun law, in which they compared homicide trends in Washington, D.C., with trends in the counties and the independent cities in Maryland and Virginia surrounding the District. This choice of control area was not justified on the basis of either cross-sectional or cross-temporal similarity between D.C. and its suburbs. Unfortunately, there was neither kind of similarity. There are few pairs of areas less similar than these two in a cross-sectional comparison. D.C. is a high violence city, with a very poor, predominantly black, and exclusively urban population, while its suburbs constitute one of the nation's wealthiest areas, with low violence rates and an overwhelmingly white, largely suburban or rural population.

Interestingly, a previous evaluation of the D.C. gun law (U.S. Conference of Mayors 1980) had been criticized for being "deficient in its choices and use of control jurisdictions" (Jones 1981:145). Jones noted that "the appropriate control for comparisons of changes in D.C. crime rates are other urban jurisdictions" (ibid.). Jones analyzed data from an obviously more appropriate control area—Baltimore. Located just 45 miles from D.C. and with a 1980 population of 692,000 (compared with D.C.'s 607,000), Baltimore is enough like D.C. to be considered a virtual sister city, closely resembling D.C. in all those respects on which the D.C. suburbs differed greatly from the city (i.e., urban area, violence rates, demographic composition). Furthermore, Baltimore's pre-intervention trends in homicide were much more like Washington's than like those in the D.C. suburbs.

From 1975 (the last complete year before the D.C. law was implemented in late September 1976) through 1977 (the first full year after the new law), the Baltimore homicide rate dropped 31% compared with a drop of 15% for D.C. These changes in homicide rates raise the possibility that Loftin et al.'s (1991) conclusion that the D.C. gun law was responsible for a 25% drop in monthly homicide counts in the District is mistaken. We explore this possibility in more detail below.

Comparing Trends in Gun and Nongun Violence

Several researchers studying the effects of gun control have chosen for their "control" series data on nongun violence in the intervention area. In these studies (see Table 1), trends in gun crimes (e.g., homicides committed with guns) are compared with trends in the corresponding nongun version of the same crime (e.g., homicides committed without guns). The basic idea is to study, within the intervention area itself, a time series of some category of events or behaviors otherwise similar to those targeted by the intervention but which are not expected to be influenced (or at least not as much) by the intervention. Thus, trends in homicides committed with guns have been compared with trends in homicides committed without guns. If gun violence decreases more (or increases less) than nongun violence after a new gun law is implemented, this pattern is supposed to be strongly supportive of the hypothesis that the gun law suppressed violence.

Justification for the gun-nongun quasi-experimental design has suggested that its value lies in narrowing the set of rival explanations for observed violence trends, meaning that there are few (and perhaps no) other likely explanations for a greater drop (or smaller increase) in gun violence than nongun violence, other than effective gun controls (e.g., Loftin et al. 1991:1618–19). The underlying rationale is that gun violence and nongun violence share the same set of causes (other than gun control efforts) and are influenced by these causes to the same degree, so that gun violence would trend the same way as nongun violence were it not for changes in gun control policies (Loftin et al., 1983).

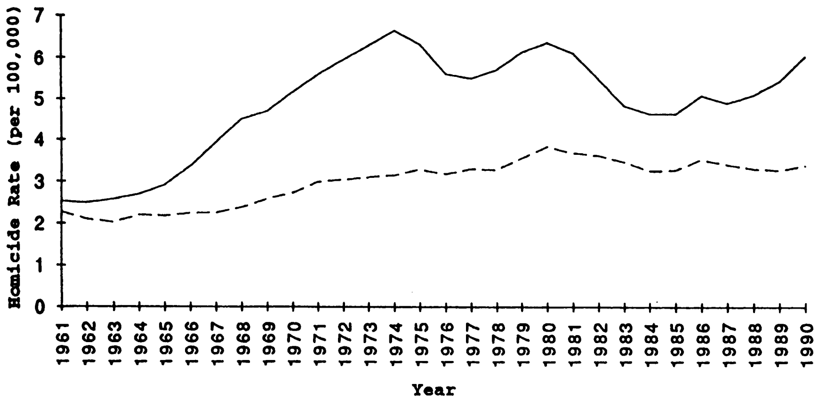

Figure 1. Trends in gun and nongun homicide rates in the United States, 1961–1990. *Solid line* = Gun homicide rate; *dashed line* = Nongun homicide rate.

The assumption that few or no other factors besides new gun laws could produce more decrease (or less increase) in gun violence than in nongun violence is implausible. Figure 1 displays the gun and nongun homicide rates in the United States from 1961 to 1990 (Federal Bureau of Investigation 1961–91). The trends in gun and nongun homicide rates indicate that there obviously are other variables that routinely cause gun homicide rates to decrease more and to increase more than nongun homicide rates. The most conspicuous pattern in gun and nongun homicide rates over this period is that gun homicide rates are much more volatile than nongun homicide rates. For the period 1961–90, national rates of gun homicide had a coefficient of relative variation of 25.8, compared with 18.2 for rates of nongun homicide. By this measure, gun homicide rates were about 42% more variable than nongun homicide rates. Thus, when overall homicide rates are declining, gun homicide rates decline proportionally more than nongun homicide rates. The periods 1974–77 and 1980–85 provide good illustrations of this pattern.

The reasons for these patterns need not concern us, though we should note that they cannot be attributed to changes in gun control policy. It is difficult to argue that larger national declines in gun violence were due to an increase in nationwide gun control strictness, since there was no such increase during the 1973–87 period.[1] Unfortunately, the use of a nongun homicide

[1] In fact, no significant new federal gun laws were passed between the 1968 Gun Control Act and the 1986 Firearms Owners' Protection Act, the latter being an NRA-sponsored bill widely interpreted as weakening federal gun laws. The trend was the same

series does not allow the researcher to rule out competing explanations of observed trends, since gun and nongun homicides appear to have different covariates and routinely diverge in the absence of new gun laws.

Specification of the Intervention Model

Perhaps as critical to the internal validity of the interrupted time series design is the specification of the impact model. There are two issues of concern here. One issue is focused on the timing of the intervention: When is the change in law or policy supposed to have an effect on the target variable? Following a determination of when the law is supposed to have an effect, a second issue is centered on how the law influences the target variable: What is the mechanism that results in a change in the target variable? Did the law have an effect on the target variable that was immediate and permanent? Immediate but temporary? Or gradual but permanent? One of the assumed strengths of the interrupted time series design has been the expectation of a well-specified intervention. For example, Campbell and Stanley (1966: 41–42) assert: "If such time series are to be interpreted as experiments, it seems essential that the experimenter must specify *in advance* the expected time relationship between the introduction of the experimental variable and the manifestation of an effect" (emphasis added).

The most common intervention point chosen in policy impact research is the official effective date of the law. The difficulty here is that many points, often accompanied by a burst of publicity, could be designated as the time when a new law's impact might plausibly begin. These would include the time when:

1. The law is first publicly proposed or introduced
2. The law is passed by a legislative committee
3. The law is passed by each house of the legislature
4. The law is signed by the executive
5. The law's effective date arrives
6. The first violator is arrested, convicted, or sentenced
7. A large enough number of violators are punished so "word gets out on the streets"
8. Publicity about the law begins in earnest

in state and local areas. During the 1973–78 period, few new state gun restrictions were passed and these were often just minor revisions of existing controls (Jones & Ray 1980:App. III). For the period between 1978 and 1987, the most important gun control trend has been the passage, in nearly two-thirds of the states, of state preemption laws (26 states during the 1982–90 period). These measures declare that the state government preempts some or all of the field of gun regulation, typically repealing existing local gun ordinances and/or forbidding future passage of new gun controls at the municipal or at the county level (U.S. News & World Report 1988; Kleck 1991:332–33). Thus, if there was any noteworthy trend at all in gun control restrictiveness during the period from 1973 to 1987, it was in a downward direction, opposite to that which could produce the observed trends in gun and nongun homicide.

9. Publicity about the law peaks

The term "effective date" is just a legalism; it has no special claim to being the point at which new laws will actually begin to have an effect on the target variable. Use of this date as the intervention point is legally relevant but socially arbitrary. Policy impact assessments have rarely considered alternatives or tested for apparent effects when differing intervention points may be appropriate.

The peculiarities of D.C.'s handgun freeze highlight the issue of specification. Loftin et al. (1991) assumed the D.C. gun law's impact began at the law's "effective date" of 24 September 1976. However, even the effective date for this law is ambiguous. The law became effective temporarily on 24 September 1976, but then the deadline for owners of registered handguns to reregister their guns was extended, followed by legal challenges which resulted in the law being suspended for two months. The D.C. gun law finally became fully effective on 21 February 1977, five months after the initial "effective date." In addition, the D.C. law did not immediately change the legal status of *any* handguns—the illegal (unregistered) handguns remained illegal, and the legal ones, due to the grandfather clause, could be reregistered under the new law and thus remain legal. In the long run, all legal handgun ownership in the District would disappear as legal owners died or moved away, but it was unknown how long it would be before this could exert an impact on gun homicides. It was only clear that any effects on the level of legal handgun ownership would be gradual.

Assuming that it is possible to specify the time at which a change in law or social policy had an effect on the target variable, a related difficulty with this research concerns the specification of the impact model (i.e., *how* the law influenced the target variable). In many instances, it may not be possible for the analyst to specify whether the change in law should have an abrupt and permanent, abrupt and temporary, or gradual and permanent effect on the target variable. In those cases where the analyst does have some expectation about how the law *should* affect the target variable, say, on the basis of theory, prior research, and/or an understanding of the law, then that specification should be chosen and tested (McCain & McCleary 1979). If the model does not fit—the hypothesis of an intervention of a specified form is rejected—a reasonable implication is that the law may not have had an effect on the target variable.

Loftin et al.'s (1991) study of the D.C. gun law again provides an illustration of this problem. They concluded that the law had an "abrupt" impact on gun homicides, based solely on results showing that the model with an abrupt and permanent impact specification fit the data better than one specifying a gradual impact. On a priori theoretical grounds, however, it would be hard to imagine an intervention whose impact (if any) was more likely to be gradual. By effectively banning future legal handgun acquisitions but allowing existing legal handguns to remain legal, the D.C. law was virtually designed to have only a gradual effect. The authors acknowledged this when they noted that "observers expected the gun-licensing law to have limited or gradual effects because it 'grandfathered' previously registered handguns and did not directly remove existing guns from their owners" (p. 1619).

Few interventions will allow such a clear-cut, theoretically based choice of intervention impact patterns, yet Loftin et al. (1991) made a purely ex post facto choice of a theoretically less appropriate model solely because it fit the data better. If the gradual impact model did not fit the data very well, then in light of the nature of the D.C. gun law, it may have been more appropriate to conclude that the law did not have an effect on homicides, since the impact of the law had to be gradual. Put another way, the pattern of change in homicide trends suggested that something other than the handgun ban was responsible for the observed decline in gun homicides.

Specification of the Time Series

Although less threatening to internal validity, a third problem with policy impact assessment studies using interrupted time series designs pertains to the beginning point and the end point and the length of the time series examined, rather than the characteristics of the control and the intervention sites evaluated. By definition, a time series is a continuous set of observations at consecutive time points and thus not a random sample of all time points. In practice, the time series assessed in policy impact studies are arbitrarily defined chunks of history, generally chosen on the basis of data availability. It has routinely been observed that the results of time series regression studies can vary sharply, depending on exactly which set of time points is used, especially when, as is usually the case, the sample size is fairly small (see, e.g., Kleck 1979; Cantor & Cohen 1980). Yet, in policy impact assessments using time series data, this issue is rarely empirically addressed by reestimating models based on differing sets of time points. Instead, the commonly held view is that the longest time series, using all available time points, will yield the most accurate parameter estimates. Since any other series would be shorter and thus statistically inferior, it is implied, only estimates based on the full series need to be produced and reported. This argument ignores the broader logical issue of whether findings will vary if a series with a different length was used. If results differ radically when varying subsets of time points are used, this lack of robustness is something readers ought to know about.

The pronounced impact of even small changes in the specification of the time series can be illustrated simply with analyses of the District of Columbia handgun ban. Loftin et al. (1991:1616, Table 1) reported that gun homicides averaged 13.0 per month in the 105 months before D.C.'s handgun ban and 9.7 per month in the first 135 months after the ban, the post-intervention period ending in December 1987. However, when we added one year of data, the post-intervention mean increased to 10.7 and by the end of 1989, the post-intervention mean had increased to 12.1, nearly eliminating the apparent reduction in gun homicides. Since the D.C. law was a sort of "slow-motion" handgun ban, as described above, one would expect its impact to be most apparent a number of years after its effective date. Thus, the years most crucial to an assessment of this particular law's impact would be *later* years, including 1988 and 1989, rather than those immediately following the effective date.

Similarly, determination of the end point of a time series to be studied is often arbitrarily determined simply by when analysts choose to study a given intervention. For example, Deutsch and Alt's (1977) analysis of a Massachusetts gun-carrying law was performed within months of its implementation, resulting in only 18 post-intervention data points to analyze, meaning they could assess only short-term effects. Other evaluations of this law were performed after more time had passed, and the researchers had a longer and later series to work with (see, e.g., Pierce & Bowers 1981). Since the evaluation of a law's effectiveness may vary with the specific time series used, there is the potential for research outcomes to be manipulated merely by the timing of the study. This issue has special relevance to policy impact studies of socially and politically sensitive issues such as gun control. For example, pro-control analysts could hurry to begin analysis of a law which was followed by crime drops the analysts suspected would be short-lived, or, if the law was followed by crime increases, could delay analysis until violence trends turned around and showed a decline. And anti-control researchers could do the reverse.

A Reassessment of the D.C. Gun Law

We use both major sources of homicide data in the United States in our following analyses to test for an effect of the D.C. gun law: police-based data, derived from the Federal Bureau of Investigation's (FBI's) Supplementary Homicide Reports program (Inter-University Consortium for Political and Social Research (ICPSR), 1991) and vital statistics data, derived from the National Center for Health Statistics (NCHS) *Mortality Detail Files*

(ICPSR 1985, 1993). We constructed monthly gun and nongun homicide counts for Washington, D.C., and Baltimore to cover the period January 1968 through December 1987 (the same period covered in Loftin et al.'s (1991) analysis of the D.C. gun law). We are using Baltimore as a control site because of the factors discussed above.

The ARIMA models used in the following interrupted time series analyses were selected using standard model development procedures (see, e.g., Box & Jenkins 1976; Box & Tiao, 1965; Hay & McCleary, 1979; McCain & McCleary 1979; McDowall et al. 1980; Wei 1990). Following the development of the univariate ARIMA model, we then included the intervention parameter (denoted as ω_0 in the tables to follow) to test for an effect of a change in behavior that reflected a change in criminal law. Tables 2–6 present our results from this exercise.

Table 2 presents the results[2] for testing for an intervention effect of the D.C. gun law on gun homicides. Panels A and B present the results for gun homicides using the FBI and NCHS data, respectively. We used the law's first effective date—24 September 1976—as the intervention point in both cities, where October 1976 is used as the first post-intervention month. The results in panels A and B show statistically significant drops in monthly homicides in both D.C. (–3.232 and –3.145 for the FBI and the NCHS data, respectively) and Baltimore (–2.811 and –4.308 for the FBI and the NCHS data, respectively). Interestingly, the magnitude of the drop is far larger in Baltimore than it is in D.C. using the NCHS data, but the pattern is reversed when the FBI data series is analyzed, showing a slightly larger drop in D.C. Irregardless, these results call into question Loftin et al.'s (1991) conclusion about the efficacy of the D.C. gun law. Since Baltimore had no new gun laws in or around October 1976, there is something other than new gun laws responsible for the drop in homicides in Baltimore. Given the similarity of Baltimore to D.C., these results imply that something other than the D.C. gun law was responsible for the drop in homicides in D.C. Thus, the use of a more appropriate control site, as we noted above, can provide very different conclusions about the nature of the effect of a legal or policy change.

Table 3 displays the results for nongun homicides in D.C. and Baltimore.[3] The results from the FBI homicide data in Panel A show statistically significant drops in nongun homicides in both D.C. (–1.121) and Baltimore (–1.193). However, the results from the NCHS data in panel B show a significant drop in

[2] Our statistical analyses were performed with SAS.

[3] Our nongun homicide series for D.C. differs from that used by Loftin et al. (1991), who apparently had included deaths by legal intervention in their nongun homicide series. In correspondence, Loftin et al. claim that they have not included such deaths, but we were unable to replicate their results until we included these values.

nongun homicides only in Baltimore. The absolute magnitudes of the impact parameters for nongun homicides are substantially smaller than are the corresponding parameters for gun homicides. Traditionally, the pattern of results found in Tables 2 and

Table 2. The Impact of the D.C. Gun Law: Gun Homicides in Washington, D.C., and Baltimore, 1968–1987

A. FBI Data

Washington, D.C.				Baltimore			
Parameter	Coefficient Estimate	Standard Error	Ratio	Parameter	Coefficient Estimate	Standard Error	Ratio
α	12.571	.499	25.20	α	13.705	.798	17.18
ϕ_1	.137	.065	2.11	ϕ_1	.284	.060	4.71
ϕ_2	.136	.065	2.08	ϕ_4	.243	.061	4.01
ω_0	−3.232	.665	−4.86	ω_0	−2.811	1.053	−2.67
	$Q = 21.50$, $df = 22$				$Q = 26.55$, $df = 22$		

B. NCHS Data

Washington, D.C.				Baltimore			
Parameter	Coefficient Estimate	Standard Error	Ratio	Parameter	Coefficient Estimate	Standard Error	Ratio
α	12.891	.517	24.91	α	14.963	.749	19.98
ϕ_1	.185	.065	2.84	ϕ_1	.238	.062	3.81
ϕ_2	.127	.065	1.96	ϕ_4	.160	.063	2.56
ω_0	−3.145	.689	−4.56	ω_0	−4.308	.993	−4.34
	$Q = 18.14$, $df = 22$				$Q = 24.30$, $df = 22$		

Table 3. The Impact of the D.C. Gun Law: Nongun Homicides in Washington, D.C., and Baltimore, 1968–1987

A. FBI Data

Washington, D.C.				Baltimore			
Parameter	Coefficient Estimate	Standard Error	Ratio	Parameter	Coefficient Estimate	Standard Error	Ratio
α	7.743	.275	28.11	α	8.467	.290	29.19
ω_0	−1.121	.367	−3.05	ω_0	−1.193	.387	−3.08
	$Q = 23.68$, $df = 24$				$Q = 22.20$, $df = 24$		

B. NCHS Data

Washington, D.C.				Baltimore			
Parameter	Coefficient Estimate	Standard Error	Ratio	Parameter	Coefficient Estimate	Standard Error	Ratio
α	6.886	.261	26.40	α	8.238	.300	27.48
ω_0	−.271	.348	−.78	ω_0	−1.068	.400	−2.67
	$Q = 31.64$, $df = 24$				$Q = 18.42$, $df = 24$		

3 has been taken as an indicator of the efficacy of gun control legislation—that gun homicides decreased proportionally more than nongun homicides. In light of the national trends in gun and nongun homicide rates presented in Figure 1, however, these results are not surprising. We should expect that when total homicides are decreasing, gun homicides will decrease by a greater proportion than will nongun homicides. We find these results uninformative in regard to testing for an impact of the D.C. gun law.

Table 4. Alternative Intervention Date—Second Effective Date: Impact on Gun Homicides in Washington, D.C., and Baltimore

A. FBI Data

Washington, D.C.				Baltimore			
Parameter	Coefficient Estimate	Standard Error	Ratio	Parameter	Coefficient Estimate	Standard Error	Ratio
α	12.430	.505	24.61	α	13.508	.822	16.43
ϕ_1	.145	.065	2.24	ϕ_1	.296	.060	4.95
ϕ_2	.151	.065	2.33	ϕ_4	.256	.060	4.27
ω_0	−3.099	.685	−4.52	ω_0	−2.560	1.103	−2.32
	$Q = 19.72$, $df = 22$				$Q = 26.50$, $df = 22$		

B. NCHS Data

Washington, D.C.				Baltimore			
Parameter	Coefficient Estimate	Standard Error	Ratio	Parameter	Coefficient Estimate	Standard Error	Ratio
α	12.799	.512	25.00	α	14.678	.782	18.78
ϕ_1	.187	.065	2.88	ϕ_1	.257	.062	4.15
ϕ_2	.134	.065	2.05	ϕ_4	.178	.062	2.87
ω_0	−3.096	.694	−4.46	ω_0	−3.951	1.055	−3.74
	$Q = 18.20$, $df = 22$				$Q = 24.50$, $df = 22$		

Table 4 presents the results for testing for a second intervention point, again using Baltimore as a control site. This second intervention may be considered the D.C. gun law's second "effective date" that occurred after legal challenges to the law, with the law becoming fully effective on 21 February 1977 (we use March 1977 as the first month following the intervention). The results in Table 4 are virtually indistinguishable from the results in Table 2. There are significant impact parameters in D.C. as well as Baltimore. Thus, it again appears that the D.C. gun law had little, if any, impact on homicides in D.C., since a nearly identical pattern was observed in a city that implemented no new gun laws at this time.

A second issue in the specification of the intervention model concerns how the law would impact homicides. As we discussed above, the D.C. gun law was expected to have a gradual impact on gun availability and therefore gun homicides. Table 5 presents the results for testing a gradual impact model in D.C. and Baltimore. Since we have a priori justification for assuming that the law, if it was to have any effect, must be gradual, this model would seem to provide the critical test of the D.C. gun law. What we find in Table 5, regardless of the data source, is a complete lack of statistical significance for the gradual intervention model in both D.C. and Baltimore.[4] Again, contrary to the conclusion of Loftin et al., we do not find any evidence that the D.C. gun law had any effect—either abrupt or gradual on monthly homicide counts in D.C.

Finally, with respect to the specification of the time series, we extended each of the homicide time series by adding two years of data, thereby extending the period covered from January 1968 through December 1989. Table 6 presents the results from this analysis. We find no evidence of a statistically significant intervention effect in D.C. with either the FBI or the NCHS data, but we continue to see a significant drop in monthly homicides in Baltimore in both the FBI and the NCHS data. We find these results puzzling. In D.C., where the gun law was passed and implemented, the significant intervention effect disappears by adding just two years of data, but in Baltimore, which had no new gun laws, there continued to be a significant drop in monthly homicide counts from October 1976 through December 1989. These results not only call into question any lasting effect (if any) of the D.C. gun law, but reveal how fragile these results can be. The results in Table 6 appear to corroborate the results in Tables 2 and 4, suggesting that something other than gun laws was influencing monthly homicide counts over this time period.

The authors of the studies using interrupted time series designs, summarized in Table 1, did not perform any of the tests for robustness that we have performed for this article. In the absence of information to the contrary, we believe the prudent assumption at this point is that these very similar studies, using methods either identical to or inferior to those used here, may suffer from the same kinds of flaws as Loftin et al.'s (1991) evaluation of the D.C. gun law. Consequently, we believe that the results of these studies should be regarded as unreliable, at least until tests for the robustness of the findings are performed.

[4] Although the gradual impact parameter (δ) for D.C., using the NCHS data, is statistically significant ($p < .05$), discussions of interrupted time series models argue that the abrupt impact parameter (ω_0) must also be statistically significant in order for the graduate impact parameter to be meaningful (see, e.g., McDowall et al. 1980).

Table 5. Test for a Gradual Intervention Effect: Gun Homicides in Washington, D.C., and Baltimore

A. FBI Data

Washington, D.C.				Baltimore			
Parameter	Coefficient Estimate	Standard Error	Ratio	Parameter	Coefficient Estimate	Standard Error	Ratio
α	12.680	.501	25.32	α	13.463	.827	16.29
ϕ_1	.132	.065	2.02	ϕ_1	.294	.060	4.89
ϕ_2	.143	.066	2.17	ϕ_4	.258	.061	4.27
ω_0	−1.401	1.170	−0.65	ω_0	−3.522	2.651	−1.33
δ	.588	.638	0.92	δ	−.391	.976	−0.40
	$Q = 22.08$, $df = 22$				$Q = 26.41$, $df = 22$		

B. NCHS Data

Washington, D.C.				Baltimore			
Parameter	Coefficient Estimate	Standard Error	Ratio	Parameter	Coefficient Estimate	Standard Error	Ratio
α	13.018	.514	25.32	α	14.931	.765	19.51
ϕ_1	.172	.065	2.65	ϕ_1	.240	.063	3.82
ϕ_2	.135	.066	2.06	ϕ_4	.165	.063	2.63
ω_0	−.844	1.199	−0.70	ω_0	−2.285	3.590	−0.64
δ	.751	.355	2.12	δ	.472	.830	0.57
	$Q = 18.21$, $df = 22$				$Q = 24.11$, $df = 22$		

Discussion and Conclusion

The use of multiple site interrupted time series designs provides analysts of policy impact a useful tool for assessing change in the target variable that may be a result of a change in legal or social policy. Our analysis of the D.C. gun law and its lack of effect on monthly homicides in D.C. provides a strong illustration of what we see as the three major problems in the use of interrupted time series designs for assessing social policy impact. Had we followed common practice in this area and (1) used nongun homicides as the statistical control, (2) not used an appropriate control group-control site, (3) discarded information about the implementation of the D.C. gun law, and (4) ignored the possibility of temporal artifacts, we would have reached conclusions similar to those in other published research: namely, that the D.C. gun law had an immediate and statistically significant "effect" that reduced monthly homicide counts in D.C.

We have attempted to strengthen the internal validity of our results by using an appropriate control area for the District of Columbia (Baltimore) and, in so doing, were unable to document an impact of the D.C. gun law on homicides. In addition, we tested for the robustness of the significant impact parameter and found it to be quite fragile in D.C.—that with the addition of

Table 6. Homicide Time Series Data Extended by Two Years: Gun Homicides in Washington, D.C., and Baltimore, 1968–1989

A. FBI Data

Washington, D.C.				Baltimore			
Parameter	Coefficient Estimate	Standard Error	Ratio	Parameter	Coefficient Estimate	Standard Error	Ratio
α	11.096	1.978	5.61	α	13.731	.777	17.66
ϕ_1	.351	.060	5.85	ϕ_1	.302	.058	5.24
ϕ_2	.271	.061	4.44	ϕ_4	.217	.058	3.71
ϕ_5	.221	.060	3.67				
ω_0	1.525	2.458	.62	ω_0	−2.587	.995	−2.60
	$Q = 24.33,\ df = 21$				$Q = 26.19,\ df = 22$		

B. NCHS Data

Washington, D.C. (Gun Homicides Logged)				Baltimore			
Parameter	Coefficient Estimate	Standard Error	Ratio	Parameter	Coefficient Estimate	Standard Error	Ratio
α	2.400	.124	19.42	α	14.969	.769	19.47
ϕ_1	.275	.061	4.50	ϕ_1	.268	.059	4.53
ϕ_2	.248	.062	3.99	ϕ_4	.148	.060	2.47
ϕ_3	.191	.062	3.07				
ω_0	−.002	.158	−.02	ω_0	−3.810	.987	−3.86
	$Q = 27.15,\ df = 21$				$Q = 25.05,\ df = 22$		

just two years of monthly homicide data, the effect vanished. Throughout, we have compared the results for D.C. with Baltimore, which allowed us to make note of the similarity of patterns and the implication that gun laws were likely not responsible for changes in monthly homicide counts in either D.C. or Baltimore.

Although the focus of our discussion has been the use of interrupted time series designs to test for effects of gun control legislation, our concerns apply to other legal and policy areas. For example, Ross et al.'s (1990) analysis of the impact of DUI legislation and Cochran et al.'s (1994) analysis of the impact of executions on homicides both fail to include a control site; thus it will remain unclear whether their results are unique to the areas studied or if similar patterns appeared elsewhere.

Thus, while interrupted time series designs have the potential to provide researchers and policy analysts with some evidence about the apparent impact of a legal or a policy change, the execution of these studies often leaves much to be desired. We hope that future efforts using these models will take into account the following considerations:

1. Having control sites that are as much like the intervention site as possible, both cross-sectionally and temporally.

2. Considering a number of possible intervention times—when the law can be enforced, when the law is enforced, when people are punished due to the law, etc.
3. Drawing on knowledge about the law's implementation to develop the intervention model. Rather than simply relying on statistical tests showing one model to fit the data better than another model, we hope that researchers will use their knowledge and understanding of the process under study as well as theoretically relevant information to justify their choice of model.
4. The need to test for the robustness of the intervention results. Specifically, are the results due to the unique nature of the time series being studied or do they hold up when additional data are added to or removed from the series being studied? It seems that if the results are so fragile that they change with minor variations in the length of the time series, the validity of a significant intervention is called into question.

Finally, with respect to assessments of politically charged interventions, such as gun control, there may be a strong temptation to skirt some of these issues if they cast doubt on the efficacy of the policy change. We hope that researchers evaluating social policy changes will resist these temptations and produce scientifically and technically sound research on which to base sensible public policy.

References

Box, George E. P., & Gwilym M. Jenkins (1976) *Time-Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Box, G. E. P., & G. C. Tiao (1965) "A Change in Level of Nonstationary Time Series," 52 *Biometrika* 181.

Campbell, Donald T., & Julian C. Stanley (1966) *Experimental and Quasi-experimental Designs for Research.* Boston: Houghton Mifflin.

Cantor, David, & Lawrence E. Cohen (1980) "Comparing Measures of Homicide Trends," 9 *Social Science Research* 121.

Cochran, John K., Mitchell B. Chamlin, & Mark Seth (1994) "Deterrence or Brutalization? An Impact Assessment of Oklahoma's Return to Capital Punishment," 32 *Criminology* 107.

Cook, Thomas D., & Donald T. Campbell (1979) *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Boston: Houghton Mifflin.

Deutsch, Stephen Jay, & Francis B. Alt (1977) "The Effect of Massachusetts' Gun Control Law on Gun-related Crimes in the City of Boston," 1 *Evaluation Q.* 543.

Federal Bureau of Investigation (1961–91) *Uniform Crime Reports.* Washington: GPO.

Fife, Daniel, & William R. Abrams (1989) "Firearms' Decreased Role in New Jersey Homicides after a Mandatory Sentencing Law," 29 *J. of Trauma* 1548.

Hay, Richard, & Richard McCleary (1979) "Box-Tiao Time Series Models for Impact Assessment," 3 *Evaluation Q.* 277.

Holmes, Malcolm D., Howard C. Daudistel, & William A. Taggart (1992) "Plea Bargaining Policy and State District Court Caseloads: An Interrupted Time Series Analysis," 26 *Law & Society Rev.* 139.

Inter-University Consortium for Political and Social Research (ICPSR) (1991) *Uniform Crime Reporting Program Data.* Study 9028, Supplementary Homicide Reports, 1975–1989. Ann Arbor, MI: ICPSR.

——— (1985) *Mortality Detail File: External Cause Extract, 1968–1980.* Study 8224. Ann Arbor, MI: ICPSR.

——— (1993) *Mortality Detail Files, 1981–1989.* Study 7632. Ann Arbor, MI: ICPSR.

Jones, Edward D., III (1981) "The District of Columbia's 'Firearms Control Regulations Act of 1975': The Toughest Handgun Control Law in the United States—Or Is It?" 455 *Annals* 138.

Jones, Edward D., III, & Marla Wilson Ray (1980) "Handgun Control: Strategies, Enforcement and Effectiveness." Unpub. report. Washington: U.S. Department of Justice.

Kleck, Gary (1979) "Capital Punishment, Gun Ownership, and Homicide," 84 *American J. of Sociology* 882.

——— (1991) *Point Blank: Guns and Violence in America.* New York: Aldine de Gruyter.

Kleck, Gary, Chester L. Britt, & David J. Bordua (1993) "The Emperor Has No Clothes: Using Interrupted Time Series Designs to Evaluate Social Policy Impact." Presented at American Society of Criminology Annual Meetings, 30 Oct. 1993.

Loftin, Colin, Milton Heumann, & David McDowall (1983) "Mandatory Sentencing and Firearms Violence: Evaluating an Alternative to Gun Control," 17 *Law & Society Rev.* 287.

Loftin, Colin, & David McDowall (1984) "The Deterrent Effects of the Florida Felony Firearm Law," 75 *J. of Criminal Law & Criminology* 250.

Loftin, Colin, David McDowall, Brian Wiersema, & Talbert J. Cottey (1991) "Effects of Restrictive Licensing of Handguns on Homicide and Suicide in the District of Columbia," 325 *New England J. of Medicine* 1615.

Lucas, Charles E., & Anna M. Ledgerwood (1978) "Mandatory Incarceration for Convicted Armed Felons," 18 *J. of Trauma* 291.

McCain, Leslie J., & Richard McCleary (1979) "The Statistical Analysis of the Simple Interrupted Time-Series Quasi-Experiment," in Cook & Campbell 1979.

McCleary, Richard, & Richard A. Hay, Jr., with Errol E. Meidinger & David McDowall (1980) *Applied Time Series Analysis for the Social Sciences.* Beverly Hills, CA: Sage Publications.

McDowall, David, Richard McCleary, Errol E. Meidinger, & Richard A. Hay, Jr. (1980) *Interrupted Time Series Analysis.* Beverly Hills, CA: Sage Publications.

McDowall, David, Colin Loftin, & Brian Wiersema (1992) "A Comparative Study of the Preventive Effects of Mandatory Sentencing Law for Handgun Crimes," 83 *J. of Criminal Law & Criminology* 378.

McPheters, Lee R., Robert Mann, & Don Schlagenhauf (1984) "Economic Response to a Crime Deterrence Program," 22 *Economic Inquiry* 550.

O'Carroll, Patrick W., Colin Loftin, John B. Waller, David McDowall, Allen Bukoff, Richard O. Scott, James A. Mercy, & Brian Wiersema (1991) "Preventing Homicide: An Evaluation of the Efficacy of a Detroit Gun Ordinance," 81 *American J. of Public Health* 576.

Pierce, Glenn L., & William J. Bowers (1981) "The Bartley-Fox Gun Law's Short-Term Impact on Crime in Boston," 455 *Annals* 120.

Ross, H. Laurence, Richard McCleary, & Gary LaFree (1990) "Can Mandatory Jail Laws Deter Drunk Driving? The Arizona Case," 81 *J. of Criminal Law & Criminology* 156.

Tennenbaum, Dr. Abraham N. (1994) "The Influence of the *Garner* Decision on Police Use of Deadly Force," 85 *J. of Criminal Law & Criminology* 241.

U.S. Conference of Mayors (1980) "The Analysis of the Firearms Control Act of 1975: Handgun Control in the District of Columbia." Washington: U.S. Conference of Mayors.

U.S. News & World Report (1988) "Even the NRA Can Have a Bad Day (Handgun Control Law Passed)," *U.S. News & World Report*, p. 15 (25 April).

Wei, William W. S. (1990) *Time Series Analysis.* Redwood City, CA: Addison-Wesley Publishing Co.

Zimring, Franklin E. (1975) "Firearms and Federal Law: The Gun Control Act of 1968," 4 *J. of Legal Studies* 133.