

The Emperor Has No Clothes: Using Interrupted Time Series Designs to Evaluate Social Policy Impact

By Gary Kleck, Chester L. Britt & David J. Bordua

The most popular quasi-experimental strategy for evaluating the aggregate impact of changes in law and other social policies is the univariate interrupted time series design (ITSD). In practice, the internal validity of this approach has been greatly exaggerated and its users have largely ignored or minimized its flaws, including: (1) its general inability to rule out alternative explanations, (2) the use of a single or small number of arbitrarily chosen “control” or comparison jurisdictions, (3) arbitrary definition of the endpoints of the time series evaluated, (4) an inability to specify exactly when the intervention’s impact is supposed to be felt, raising problems of the falsifiability of the efficacy hypothesis, and (5) an atheoretical specification of the ARIMA impact model.

Data pertaining to the 1976 Washington, D.C., handgun ban are analyzed to illustrate these problems. Authors of a previous evaluation concluded that the ban reduced homicides; this conclusion collapses when any one of several valid changes in analytic strategy are made. Further, when “bogus intervention” points are specified, corresponding to nonexistent policy interventions, results as strong as those obtained by the original authors are obtained. Finally, when the same ITSD strategy is applied to an example of gun “decontrol,” a gun law repeal exactly opposite in character to that of the D.C. law, the same appearance of a homicide-reducing impact is generated. It is concluded that the univariate ITSD approach is of little value for policy assessment, because it can so easily be manipulated to generate results compatible with a researcher’s preconceived biases.

This is a revised version of a paper presented at the annual meeting of the American Society of Criminology in Phoenix, Arizona, October 30, 1993. A portion of this paper was published in 1996 in Law & Society Review.

Gary Kleck is professor at the School of Criminology and Criminal Justice at Florida State University; Chester L. Britt at the College of Human Sciences, Arizona State University – West; and David J. Bordua at the Department of Sociology, University of Illinois.

One of the most common general research designs used to assess whether a new law or change in public policy has affected the frequency of some behavior or outcome is the interrupted time series design (ITSD). In a typical application of this design, multiple observations of the target (dependent) variable (e.g., a count or rate of crime or violence) are analyzed to determine whether there was a shift in the level of the time series at the point when the new law or policy (labeled the “intervention” or the “treatment”) went into effect. Observations can be based on almost any unit of time, from hourly observations to annual ones, but past applications to crime most commonly have been based on monthly observations. Although not a necessary element of the basic research design itself, the

data analytic methods most commonly applied to the resulting observations have been versions of the Autoregressive Integrated Moving Average (ARIMA) time series methods developed by Box, Jenkins, and Tiao (Box and Tiao 1965; Box and Jenkins 1976). The analyses are almost always univariate, i.e. the only measured variable is the target variable.

This design is regarded by some as the strongest strategy for assessing the aggregate (population-wide) impact of policy interventions where, as is commonly the case, true experimentation is impossible or impractical (see, e.g., Campbell and Stanley, 1966; Cook and Campbell, 1979). The design has been applied to a variety of legal and policy issues, such as, the impact of changes in welfare policies (Hedrick and Shipman, 1981), drunk driving legislation (Ross et al., 1970, 1990), hotel room taxes (Bonham et al., 1992), child restraint laws (Rock, 1996), oil prices (on property crime) (Chamlin and Cochran 1998), and police patrol (Zimring, 1978; Cook, 1980).

The purpose of this paper is to show that the widespread faith in this design is unwarranted, and that it is a design prone to abuse when used for purposes of assessing the impact of policy interventions. To illustrate these problems, the literature on gun control impacts will be closely critiqued. The focus on gun control impacts serves a useful limiting purpose, since many of the more sophisticated applications of the design have been carried out in this area. If these more refined applications of the design have been misleading, then less skillful applications in other areas are likely to have generated even less reliable results. Thus, the paper's purposes are both methodological, with respect to the utility of the ITSD, and substantive, with respect to the validity of the findings of the gun control impact studies.

I. Applications of ITSD to Gun Control Impact

Table 1 lists the important studies using ITSD to evaluate the impact of gun control laws on crime and violence. [Note: all tables are printed at the end of this article.] These studies will be used to illustrate the key problems in applying ITSD to evaluate the hypothesis that a given policy change reduced the frequency or rate of some problematic behavior (e.g. crimes) or increased the frequency or rate of some desirable ones (e.g. police arrests). Two important patterns are evident in the table. First, only two types of gun control laws have received any significant attention, out of the dozens or hundreds of existing types of gun controls: laws providing mandatory penalties for unlawful carrying of weapons, and laws providing mandatory additional penalties when violent felonies are committed with guns. Second, the interventions evaluated were nearly all concentrated in a very brief segment of history, from 1974 to 1982. Both patterns suggest that any unique aspects or peculiarities of either the interventions or the time period may sharply restrict generalizability and distort findings, a suspicion that will be confirmed later.

II. Methodological Issues

A. The Inability to Rule Out Rival Explanations

The central problem in assessing the impact of policy changes on aggregates like cities or states is ruling out rival explanations

of observed trends in the target variable and thereby isolating the impact of the policy change (see, e.g., Lieberman, 1985). The simple interrupted time series design only allows the researcher to determine whether there was a systematic shift in the target variable time series around a given time point. It cannot identify the cause of that shift. There are innumerable confounding factors that could shift trends in a given target variable, and most of these are likely to be changing to at least some degree at the same time the policy change was implemented. While it is unlikely that large changes in the target variable are solely attributable to any one confounding factor, there is nothing implausible about even the largest changes in the target variable being due to modest changes in a combination of multiple rival factors.

Although multivariate time series methods are available (e.g., Tiao and Box 1981), ITSD applications to policy impact evaluation are almost invariably univariate (for an exception, see Ross et al.'s [1990] analysis of drunk driving behavior). Hence, there are no explicit controls for any other determinants of trends in the target variable other than the policy being evaluated. Simple ARIMA modelling of a time series cannot magically control for the influence of extraneous factors and thereby isolate the effect of the policy being studied. In a passage widely quoted but also widely ignored in practice, Hibbs (1977, p. 172) observed that "Box-Tiao or Box-Jenkins models are essentially models of ignorance that are not based in theory and, in this sense, are devoid of explanatory power." A group of analysts who approvingly quoted this passage and later applied the univariate ARIMA methods to crime data elaborated this observation as follows: "A univariate ARIMA model is a stochastic or probabilistic description of the outcome of a process operating through time. It provides no information about the inputs generating that process.... As in other areas of the social sciences, inference of a causal relationship in time series analysis can only be made through assessment of covariation between one or more explanatory variables and a dependent variable" (McCleary and Hay, with Meidinger and McDowall 1980, p. 111).

David McDowall has been coauthor to many of the applications of ARIMA models to gun law evaluations (see Table 1), so this caveat is especially noteworthy in light of the strongly worded causal inferences later drawn in those impact evaluations. For example, after evaluating one gun law, McDowall and colleagues flatly stated that "the law reduced gun-related suicides and homicides substantially and abruptly" (Loftin, McDowall, Wiersema and Cottey 1991, p. 1620). And in another ITSD study, the authors asserted that "The *only* plausible interpretation of the results is that the reductions in gun homicides are due to the announcement of the laws" (McDowall, Loftin, and Wiersema 1992, p. 390).

Given the extremely erratic shifts routinely observed for monthly crime counts for local areas like cities or counties, it would seem to be a reasonable working assumption that a large share of the causal determinants of these trends would also frequently exhibit similarly erratic shifts. If so, changes in laws or other public policies at any one point would generally be accompanied by nonsystematic changes in a large, though unknown, number of other factors that affected the target problem. Without explicit controls for these competing factors, the conclusion that the evaluated policy was responsible for an observed reduction in the problem amounts to little more than a guess.

B. Use of Control Series

The most common ITSD strategy for ruling out alternative explanations has been to use control series, which most commonly come in two varieties. First, trends in the intervention area (the area or jurisdiction where the evaluated policy change was implemented) may be compared with trends in some other area where no such intervention occurred. Second, trends in the targeted behavior in the intervention area may be compared with trends in a behavior which is similar (in some way) to that targeted by the intervention, but is not supposed to be affected by the intervention. In gun control studies, trends in counts or rates of 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). Table 1 indicates that five of the nine major studies of gun laws used control series -- four used only the gun/nongun comparison, and one used both kinds of control series.

1. Comparing Control Areas.

It is commonly hinted that the control area is similar enough to the intervention area to serve as a control analogous to control cases used in true experiments. However, the underlying logic for the selection of control areas is rarely made explicit. If one is comparing trends in the intervention and control areas, the necessary underlying assumption is this: "Trends in the intervention area would have been identical or similar to trends in the control area, had there been no intervention. Therefore, if the problematic target phenomenon (such as crime) decreases more (or increases less) in the intervention area than in the control area, it supports the claim that the intervention suppressed the problem, either reducing it or preventing a larger increase."

It is similarity of *trends* in the target variable between the intervention and control areas, not merely similarity in static *levels* of confounding factors, which should be especially pertinent to the adequacy of the comparison series as a control series. 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, it would obviously not be particularly meaningful that the intervention city enjoyed a post-intervention drop in crime while the control city experienced an increase. For another area to be useful as a control, it must show preintervention *trends* similar to those in the intervention area, and not just similarity in demographic characteristics. Yet, most applications of ITSD to social policy evaluation routinely cite only static cross-sectional similarity between intervention and control areas, or say nothing on the matter at all, allowing unwary readers to assume the similarity.

Pierce and Bowers (1981) did not report ARIMA results for any control areas, but they did report percentage changes in crime rates in a number of "control" cities before and after a new gun law was implemented in Boston. The cities were selected solely on the basis of being similar in population size and/or being located in the same region. Loftin, McDowall, Wiersema and Cottey (1991) compared homicide trends in Washington, D.C. with trends in the counties and independent cities in Maryland and Virginia surrounding the District. They did not explicitly justify this choice of a control area on the basis of either cross-sectional or cross-temporal similarity between D.C. and its suburbs.

In fact, 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 obviously 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. More importantly, preintervention trends in homicide were not similar in D.C. and in its suburbs. In the two years preceding the D.C. gun law, from 1974 to 1976, the homicide rate in D.C. decreased by 30%, while dropping less than 10% in the rest of the D.C. Standard Metropolitan Statistical Area (SMSA). From 1968 to 1976, the correlation of annual homicide rates between Washington and the rest of the D.C. metropolitan area was a statistically nonsignificant 0.31 (based on statistics in Table 2.)

None of the scholars applying ITSD to gun law evaluations has justified the selection of a control area based on similarity of its preintervention trends with those of the intervention area. Thus, the choices were made on arbitrary grounds unrelated to the logic underlying use of a control series. An alternative procedure would have been to systematically examine trend data in all cities (or states, counties, etc.), to identify those areas with the most similar preintervention trends in the target variable(s).

Further, although the results from an ITSD analysis with a single control area are stronger than those without any control areas at all, the results will nevertheless be inherently unstable, and can change radically with use of a different area, no matter how carefully the control area is chosen, due to eccentricities in trends in the control area. The use of multiple control areas, on the other hand, would permit inferences which would be more defensible than those based on use of a single comparison area. Nevertheless, the logical problems of not explicitly controlling for confounding factors would remain, since one still could not be confident that there were not other confounding factors operating just in the intervention area (or operating more strongly there) which caused the observed trends in the target variable.

2. Comparing Gun and Nongun Violence.

Perhaps in recognition of the difficulties of locating areas sufficiently similar to use as control jurisdictions, some authors have applied an alternative control strategy which uses a time series of events or behaviors similar to those targeted by the intervention, but which are not expected to be influenced (or at least not as much) by the intervention. For example, five of the ten studies in Table 1 compared trends in crimes committed with guns to trends in crimes 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. The underlying, usually unstated, rationale is that gun violence and nongun violence share the same set of causes (other than gun control policies), 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.

Advocates of the gun/nongun comparison strategy have argued that its value lies in somehow narrowing the set of rival explanations for observed violence trends, hinting 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, McDowall, Wiersema, and

Cottey 1991, pp. 1618-9). That is, they assume that few other factors, besides gun control laws, could selectively affect gun violence (Loftin et al 1983, p. 290). Putting it another way, McDowall, Loftin and Wiersema (1992, p. 381) stated that “Another causal variable would be confounded with the law only if it influenced gun and non-gun crimes differently, and if it changed markedly at the intervention point.”

There are two problems with the phrasing of this statement. First, the reference to “variable” in the singular is potentially misleading because it is unlikely that a *single* variable of any kind is responsible for most very large changes in crime rates. Second, if multiple variables were indeed responsible, it would not be necessary for any one of them to change “markedly” at the intervention point to produce large changes in the target variable, since modest changes in multiple variables would be sufficient.

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. First, trends in nongun violence cannot be used as a “control” in analyses of trends in gun violence because the two do not behave similarly in the absence of changes in gun control policies. One of the most conspicuous patterns evident in comparisons of gun and nongun homicide is that the former is far more volatile than the latter. For the period 1961-1990, national rates of gun homicide had a coefficient of relative variation of 25.8, compared to 18.2 for nongun homicide (computed from data in Table 3). By this measure, gun homicide rates were 42% more variable than nongun rates. When overall homicide is going down, gun homicide usually declines proportionally far more than nongun homicide. Conversely, when overall homicide is going up, gun homicide increases proportionally more than nongun violence. Consequently, by selectively studying interventions in periods of generally declining homicide, analysts can routinely expect to find bigger drops in gun homicide than in nongun homicide, regardless of whether they were accompanied by any new gun laws or other changes in gun-related public policies.

Patterns of larger declines in gun violence than in nongun violence were the dominant national trend from around 1973 through 1987. These patterns are documented in Table 3, which also shows that they were characteristic of all forms of crime involving guns, not just homicide. One simple way to detect a greater decline in gun violence than in nongun violence is to note trends in the percent of violent events which involved guns. When “percent gun” declines, it indicates a larger decline (or smaller increase) in gun violence than in the corresponding nongun violence category. For the U.S., the percent of violent crimes involving guns decreased from 1974 to 1983 for homicide (monotonically, if one excludes 1977), from 1975 to 1987 for robbery (monotonically, if one excludes 1979), and from 1973 to 1983 for aggravated assault.

The differences in gun and nongun trends can be even more extreme in the smaller local areas that typically are evaluated in ITSD studies than for the nation as a whole. For example, from 1975 to 1978, Baltimore experienced a 35% decrease in gun homicides, contrasted with only a 7% decrease in nongun homicides (analysis of FBI Supplementary Homicide Report computer tapes -ICPSR 1991). However, Baltimore did not have any new gun laws during this period; the restrictiveness of its gun laws remained unchanged and thus could not have caused the observed trends. At minimum, it is obvious that even enormous proportional drops in gun violence, accompanied by weaker or nonexistent drops in nongun violence, can occur in U.S. cities without new gun laws being

even partially responsible.

The reasons for these patterns need not concern us, beyond noting that they cannot be attributed to changes in gun control policy. One cannot argue that larger national declines in gun violence than in nongun violence were due to an increase in nationwide gun control strictness since there was no such increase during the 1973-1987 period.

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 a weakening of federal gun laws. The trend was the same in states and in local areas. During the 1973-1978 period, few new state gun restrictions were passed and these were often just minor revisions of existing controls (Jones and Reay 1980, Appendix III). For the period 1978-1987, the most important gun control trend was the passage, in nearly two-thirds of the states, of state preemption laws. 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 county level (*U.S. News and World Report* 4-25-88; Kleck 1991, pp. 332-3). Thus, if there was any noteworthy trend at all in gun control restrictiveness during this period, it was in a downward direction, opposite to that which could produce the observed trends in gun and nongun violence.

The trends in gun and nongun violence indicate that there obviously *are* other variables which routinely cause gun crime rates to decrease more than nongun crimes. Second, given the national prevalence of these patterns during this era, these covariates were clearly not minor factors which operated only under rare circumstances; instead, one could routinely expect them to be operating in most areas most of the time, including those times when a given legal jurisdiction happened to be implementing a new gun control policy. Given that gun and nongun violence trends so routinely diverge in the absence of new gun laws, it may well be that many or even most causes of violence have effects of different size on gun and nongun violence.

ARIMA methods address "drift," and thus would deal with gradual drops in percent gun which began before a given intervention. This is not the problem at issue here. Rather, the problem is that the relatively smooth national trends we have noted reflect widely scattered and very erratic local shifts which were often not at all gradual (e.g. trends in Baltimore and Louisville discussed later). These abrupt and seemingly erratic shifts in percent gun frequently occurred in places and at times where they could not have been due to new gun restrictions, since there were none.

The comparison of gun homicide with a nongun homicide "control" series does not allow the researcher to rule out any competing explanations of observed trends. Like the use of arbitrarily chosen control areas, the use of nongun violence trends for control purposes does not provide an adequate test of a gun policy's impact on the targeted behavior.

C. The Difficulties of Case Study Research

Policy impact studies using the ITSD approach are almost always case studies, assessing a single intervention in a single locale, or occasionally studying a small number (six or fewer) of similar interventions in a handful of different locales. Either variety suffers

from the obvious problem of generalizability. Even if one believes that a given intervention really produced a desirable impact in a given set of circumstances, there is no assurance that it would do so in another locale or at another time. This highlights the poor research efficiency of this approach: by applying a case study approach to single instances of a few types of gun control, it could take many decades before a large enough number of cases have been studied to permit generalizable conclusions about the effectiveness of any given type of gun control.

Another problem with evaluating a single instance (or small number of instances) of a type of intervention is identifying how it produces its effects. Even seemingly simple interventions are usually a complex bundle of elements, some very different from others. For example, many analysts evaluated the Bartley-Fox law as if it only established mandatory penalties for unlawful carrying (e.g. Deutsch and Alt 1977; Hay and McCleary 1979; Pierce and Bowers 1981), a measure opposed by the National Rifle Association (NRA), but it also established add-on penalties for committing crimes with a gun, a measure *supported* by the NRA.

Therefore, ITSD analysts using a case study approach usually cannot answer simple policy-relevant questions like “why or how did the intervention work?” or “what elements of the intervention worked?” Policy makers almost never adopt other jurisdictions’ policies *in toto*, unmodified in any way. Consequently, they run the risk of adopting a policy which worked elsewhere, yet omitting or distorting the key elements responsible for its success. Or, they run the risk of including the effective elements, but also needlessly including numerous other costly and ineffective or counterproductive elements as well, reducing the policy’s net effectiveness and efficiency. Thus, knowing exactly which elements really work is important.

The mechanisms by which the D.C. gun law supposedly reduced gun homicides are especially mysterious. The law mandated a ban on further handgun sales, a freeze on registering any more handguns, and a continuation of the existing ban on possession of unregistered handguns. Since existing registered guns could not be transferred, this effectively constituted a ban on handgun possession, but with already registered handguns “grandfathered” in as legal weapons. Thus, registered handguns continued to be legal and unregistered handguns continued to be illegal. The measure should have had little or no short-term impact on the supply of lawfully owned handguns, but should have eventually produced a gradual decrease in legal handguns, as lawful owners died or left the District. Nevertheless, Loftin et al. (1991) asserted that the law somehow produced an *abrupt* 25% reduction in gun homicides.

Even if one were willing to assume that the law somehow produced an abrupt rather than gradual drop in registered handguns, this could not have produced a 25% decline in gun homicides. The D.C. police chief reported that his department’s statistics for 1975 indicated that “less than 0.5% of the guns seized by police in connection with crimes were registered” (*Washington Post* 7-24-76, p. E3). If homicide guns were even approximately like other crime guns, even the instantaneous elimination of *all* registered handguns, never mind a mere freeze on additions to the registered handgun stock, could not have produced an abrupt 25% drop in gun homicides, since registered handguns simply were not used to commit any significant number of gun crimes in D.C.

The authors speculated that perhaps, for unstated reasons, “people voluntarily disposed of guns,” presumably including in this category violence-prone people getting rid of the *unregistered* handguns that actually predominated among D.C. gun crimes (p. 1619).

Local history, however, indicates that this is highly unlikely. Just one year before the handgun ban was passed, from April 6 to July 3, 1975, the D.C. police conducted an amnesty program in which residents could voluntarily turn in unregistered guns without fear of prosecution (*Washington Post* 4-3-75, p. D3). The first 17 days of the 90 day program yielded a grand total of 35 guns, evidently including long guns as well as handguns (*Baltimore Sun* 4-23-75). Even if this pace was maintained for the rest of the period, the program would have yielded only 185 guns, in a city where the police estimated there were 100,000 unregistered handguns (*Washington Post* 4-9-75, p. B3), plus an unknown additional number of unregistered rifles and shotguns. If this voluntary turn-in program yielded less than 0.2% of the stock of unregistered handguns, it is implausible that just one year later enough D.C. residents voluntarily disposed of their unregistered handguns to produce a 25% reduction in gun homicides, or any significant share thereof. Thus, neither the total elimination of registered handguns nor voluntary disposal of unregistered handguns is a plausible explanation of the drop in gun homicides.

Case studies also have the simple problem of being studies of a single case or a very small sample. The smaller the sample, the more likely it is that some local confounding factors could be responsible for whatever patterns are observed. For example, regarding the D.C. gun law study, even if one could have faith in the utility of the gun/nongun comparison, ignored the problems of using an unsuitable control area, and were willing to conclude that something gun-related was responsible for D.C.'s homicide trends, it would still be impossible to determine whether the new gun law was effective. As was pointed out over a decade before the Loftin et al. evaluation (in an article they cited), there were at least three other gun-related "interventions" going on in Washington at the same time its handgun ban ordinance was being debated and implemented (Jones 1981, pp. 144-5), none of which Loftin et al. mentioned to their readers. The federal Bureau of Alcohol, Tobacco and Firearms (BATF) conducted Operation CUE (Concentrated Urban Enforcement), a policy of intensified enforcement of existing federal gun laws, in the D.C. area and two other urban areas, from February 16, 1976 through 1977. It was devised with the express purpose of reducing illegal gun trafficking and thereby reducing gun violence (U.S. BATF, no date). Meanwhile, the D.C. handgun ban was approved in committee on 4-15-76, was passed by the D.C. City Council on 6-29-76, first went into effect on 9-24-76, and then, after legal challenges, went permanently into effect on 2-21-77 (*Washington Post* 4-16-76, p. C5; 6-30-76, p. 1A; Jones 1981). Thus Operation CUE completely overlapped the period in which the D.C. law was passed and implemented.

Further, in February 1976, the first of several undercover fencing operations in D.C. was announced to the public, operations which were responsible for, among other things, seizures of illegal guns and arrests of hundreds of criminals. Finally, the D.C. police, in cooperation with the U.S. Attorney for D.C., improved their efficiency in handling major criminal offenders (Jones 1981), who are disproportionately likely to use guns in their crimes (Kleck 1991, Chapter 5). Thus, even if one wanted to attribute homicide reductions to either gun control of some sort, or other criminal justice system activity, it would be impossible to confidently attribute it to the new D.C. gun law.

Oddly enough, the sponsors of Operation CUE cited some of the very same crime data used by Loftin et al. to support the D.C.

law, to support their claims that *Operation CUE* was responsible for violence reductions (U.S. BATF, no date; *Washington Post* 3-25-77, p. C1)! Thus, BATF and Loftin et al. were each implicitly in the peculiar position of having to assume that the other's preferred policy failed, in order to conclude that their own preferred policy succeeded. None of this prevented Loftin et al. from flatly stating that all alternative explanations of the gun homicide drop, other than attributing it to the local handgun ban, were "implausible" (p. 1618).

It would be nice to think that these sorts of confounding changes in the causal processes affecting crime trends were unique to Washington, but it is more realistic, and certainly more prudent, to assume that similar "unique" events or local disturbances are a routine feature of life in almost any large intervention area.

Indeed, a critical problem in using any longitudinal approach to evaluating the impact of public policy changes is that such changes are more or less continuous and omnipresent - governments are nearly always doing something intended to affect the frequency or severity of a given social problem. This is not merely true as a generalization about all policy-making, considered indiscriminantly in the aggregate, but also applies specifically to as narrow a category of policy as gun control; governments are nearly always modifying, or attempting to modify, gun policy in at least some minor ways.

Tamryn Etten's (1993) exhaustive examination of gun law making in Florida revealed that from 1949 to 1992, the Florida legislature considered a total of 641 gun control bills, passing 70 of them into law. Thus, an average of 14.6 were proposed and 1.6 were passed per year; better than one bill a month was introduced, and one became law about every seven months. Given that years can pass between a bill's initial introduction and its passage into law, this means that, even if one ignored bills that failed, the citizens of Florida and their elected representatives were virtually continually in the process of passing gun laws.

When Loftin and McDowall (1984) evaluated a Florida law which enhanced sentences for committing crimes with a gun, they did not note this near-continuous process of gun law-making, instead implicitly treating the passage of this particular law as an isolated event whose effects, if any, could not be confused with the effects of other gun laws being passed. However, even confining attention to a single narrow category of law, there were no less than *eight* sentence enhancement laws passed between 1961 and 1990, six of them between 1975 and 1990 (Etten 1993).

Thus, the making of gun laws was virtually continuous. Given the possibility of anticipation or "announcement" effects before a law's effective date, and of lagged effects after that date, every month in Florida was subject to the overlapping effects of multiple gun laws passed around the same time. How, under such circumstances, can one realistically expect to separate the effects of one particular gun law from those of other gun laws, never mind the effects of other laws and thousands of other variables influencing violence trends?

D. Selection of Intervention Sites

Even if more than one intervention site were evaluated, another problem which afflicts single-site case studies would still persist: the possibility of bias in the selection of sites. For example, Loftin, McDowall and Wiersema (1991) evaluated mandatory add-

on penalties for committing crime with guns in six cities, but noted that their six-city sample was selected because “there was publicity suggesting that [the gun laws these cities were subject to] had successfully reduced violent gun crime” (p. 17; see also McDowall, Loftin and Wiersema 1992, p. 391). Thus, the sample was biased to include cities with some *a priori* evidence that the policies were effective, so we should not be surprised by their finding significant reductions in gun homicides in four of the six cities.

E. Is There an Advantage for Determining Causal Order?

Longitudinal designs in general, including ITSD, use time-ordered observations, which can help in establishing causal order. When one is evaluating the impact of a discrete event, such as implementation of a new law or other public policy, time order is easy to establish: the policy’s implementation usually begins at a single known date and is then followed (or not) by a later change in the frequency of the targeted behavior.

Thus, one clear, potentially significant advantage of the univariate ITSD strategy over cross-sectional approaches is the former’s potential advantages in establishing causal order and disentangling possible reciprocal effects. With regard to policy impact evaluation, one might generally hypothesize that the magnitude of the target problem has a positive causal effect on the probability of any given potential public policy solution being adopted in the first place. Once adopted, the policy may then have the intended negative effect on the magnitude of the problem. Thus, with gun control, one might suppose that as gun violence increases, public and political support for stricter gun laws will rise, increasing the probability of a new law being passed. Then, once it is implemented, the law could reduce gun violence. Using time-ordered data could help address this possible reciprocal relationship.

This, however, is only an advantage when and where there actually is a two-way causal relationship to deal with. Regarding gun control, causal order is problematic only if violence rates have a net causal effect on passage of new gun laws. In fact, there is no empirical evidence this is true, and considerable evidence that it is not. Survey evidence indicates that public support for gun control is unrelated to crime rates in the cities where respondents live, to their own prior victimization experiences, or to their expressed level of fear of crime. Generally, public support for gun control is unrelated to crime (Kleck 1996). Further, survey evidence has indicated that nearly half of gun control supporters favor stricter gun laws even though they believe they will have no impact on crime or violence, suggesting that their support is not primarily based on concerns about crime (Kleck 1991, Chapter 9). Aggregate national survey data also indicate that crime rate increases in the 1960s and 1970s did not translate into increases in the level of support for gun control, because people who responded to crime trends by supporting gun control were balanced out by people who responded by getting a gun for self-defense, and consequently opposing gun control (Stinchcombe et al. 1980). Finally, historical evidence indicates that American gun laws, most of them tracing back to measures passed in the 19th and early 20th century, were passed primarily in response to concerns about racial and ethnic minorities, foreigners, labor organizers, political dissidents, and other groups unpopular with political elites and perceived to be dangerous, rather than concerns about ordinary crime (Kennett and Anderson 1975; Kates 1979)

The overt rationale for gun control is the reduction of crime and violence. Certainly legislators sponsoring gun laws will frequently cite crime statistics or individual violent incidents to justify the need for gun laws. However, even if one accepts the utilitarian premise that gun control, being a proposed solution for these problems, will become more popular as the perceptions of the seriousness of the problem increases, it still would not follow that increases in actual or measured violent crime rates make it more likely that new gun laws will be passed. Members of the general public do not have accurate perceptions of whether crime is going up or down. Increases in fear and the perception of crime and violence as serious problems are as likely to occur when violence is decreasing, as it did during the 1981-1986 period, as when violence is increasing, as it did in the 1964-1974 period (U.S. Bureau of Justice Statistics 1989).

These public perceptions may be driven instead by trends in news media coverage of violence and perhaps fictional mass media materials as well. The volume of news media coverage of crime, however, is largely unrelated to actual rates of crime (see reviews in Garafolo 1981 and Marsh 1989). Consequently, there is again no empirical basis for expecting measured or actual trends in crime or violence to affect the probability of gun laws being passed, and hence no basis for expecting a causal order problem in assessing the impact of gun laws on crime and violence rates. The usual advantage of longitudinal designs for helping address causal order problems appears to be irrelevant to this issue.

F. Arbitrary Definition of the Set of Time Points Analyzed

Another sampling issue pertains to the set of time points examined rather than the intervention or control sites evaluated. By definition, a time series is a continuous set of consecutive time points, and thus not a probability sample of all time points. In practice, the time series assessed in ITSD studies are arbitrarily defined segments of history, chosen primarily 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 (Kleck 1979; Cantor and Cohen 1980). Yet, in applications of ITSD to policy impact assessments, this issue is rarely empirically addressed by re-estimating models based on differing sets of time points. Instead, analysts commonly adopt the simplistic statistical stance that the longest time series, using all available time points, will yield the most stable parameter estimates (assuming the data are of constant quality). Since any other series would be shorter and thus statistically inferior, it is implied, only estimates based on the full series need be produced and reported. The longest time series also will not be influenced by short-term changes in the target variable and is more likely to detect cycles in the behavior that would be missed in a shorter series. These observations do not, however, dispose of the broader logical issue of whether findings will differ if a different series were used. If results change radically when varying subsets of time points are used, this lack of robustness is something which readers, not to mention the analysts themselves, ought to know about.

The impact of even small changes in the time series can be simply illustrated with analyses of the District of Columbia's handgun ban. Loftin et al. 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 (p. 1616), the post-intervention period ending in December 1987. However, if one adds 2 more years of data, covering 1988 and 1989, the post-intervention mean rises to 13.3, completely eliminating the apparent reduction in gun homicides. Adding in 1990 data boosts the post-intervention monthly mean to 15.1, implying a 16% *increase* in gun homicides after the law went into effect (computed from data in ICPSR 1991).

It should be stressed that since the D.C. law was a sort of “slow-motion” handgun ban, 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 1988-1990, rather than those immediately following the effective date.

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. Some analysts of a Massachusetts gun law rushed to study it within months of its implementation, so they had only 18 post-intervention data points to analyze, and could assess only short-term effects (Deutsch and Alt 1977). Others waited until more time had passed and they therefore had a longer and later series to work with (Pierce and Bowers 1981). Leaving aside why particular analysts timed their research as they did, it is possible for research outcomes to be manipulated merely by the timing of the study. 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. Anti-control researchers could do the reverse.

Even the choice of data sources to use, when multiple sources are available, can affect the finishing point of the time period studied in significant ways. The most common crime studied in ITSD gun control studies is homicide, which is counted by both the vital statistics system and the police. The national vital statistics system is far slower in generating usable statistics, with published and computer-readable data being released two to three years after police counts are available. In effect, this means that time series analysts can omit, on seemingly legitimate grounds of data availability, two or three years worth of data hostile to their preferred hypothesis simply by choosing to use vital statistics data rather than police data. Conversely, if the most recent time points favor their preferred hypothesis, they can include them by using the police counts.

G. The Intervention

1. Biased Selection of Interventions by Era

A similar but nevertheless distinct problem is the selection of interventions with respect to historical period. Not only can a given intervention be evaluated using an arbitrarily and possibly biased set of time points, but analysts can also choose to assess a biased sample of interventions which all occurred in the same unrepresentative historical period. Consider once again Table 1. Of 18 intervention-assessments (counting multiple assessments of the same intervention multiple times), 15 occurred in the brief period between 1974 and 1977, and all 18 occurred between 1974 and 1987. Were this period like any other, this would be of no consequence, but the data in Table 3 show that this period was not just like any other. As we noted above, it was an era when every

variety of gun violence was declining, and these declines were almost always greater than declines in nongun violence. In short, this was an era which favored pro-control conclusions regardless of the actual impact of gun control.

Loftin, McDowall and Wiersema's study of gun laws in six cities (1991; see also McDowall et al., 1992) provides an illustration of how results can be biased by the historical period of the intervention. They asserted that it was remarkable and highly significant that four of the six cities showed significant declines in gun homicide when gun laws were implemented, each larger than declines in nongun homicide, arguing that this "consistency" strongly buttressed their conclusion that the gun laws they were evaluating were responsible for these trends. However, in light of the national homicide trends presented in Table 3, it is quite possible that most any random sample of six cities examined for the 1975-1982 period would have shown the same gun/nongun patterns observed by Loftin et al. for at least four of the cities. Such consistency is neither remarkable nor necessarily a product of effective gun laws. Instead it was a commonplace pattern which was at least partly, and possibly entirely, attributable to other causal forces besides gun control, operating across the nation during this era.

The problems with Loftin et al.'s (1991) analysis also helps to reiterate a point we made above -- the need for multiple control sites that are similar to the intervention site both in terms of its demographic characteristics and its trends in crime. Had Loftin et al. used an appropriate group of control sites, their results likely would have called into question the effectiveness of the six gun laws, since cities without these laws experienced similar declines in gun homicides over the same period.

2. When Does the Intervention's Impact Begin?

One of the theoretical strengths of the interrupted time series design is the ability to test for differences in the level of the target variable before and after a *well defined intervention* occurred. In practice, however, this is much more difficult to specify in evaluating a legal or policy change. For example, when Massachusetts passed its law providing mandatory minimum sentences for illegal weapons carrying, most analysts simply assumed its impact would begin at its official effective date (e.g., Deutsch and Alt 1977; Hay and McCleary 1979). However, after Pierce and Bowers' (1981) ARIMA analysis failed to reveal any drop in gun violence in the month of the effective date, they searched for, and found, a drop in the month *preceding* the effective date.

While one might criticize them for *ex post facto* hypothesis testing, their rationale for looking for this pattern was perfectly reasonable. They argued that they had discovered an "announcement effect," and that prospective gun carriers had responded to publicity announcing the coming of the law, refraining from carrying before the law actually went into effect. Of course, it would be arbitrary to anticipate such an effect only for the month immediately preceding the effective data, since similar arguments could be made for almost any month between the law's initial legislative introduction through its effective date. Note, however, that if one concedes that laws not yet legally in effect can influence crime rates, what would prevent bills not yet passed from also affecting crime? And if these bills can have an impact, why not bills which are introduced (perhaps to much fanfare), but which will never be passed?

Conversely, one could also anticipate lagged effects of an intervention, on the assumption that people targeted by the law responded only after enough time has passed for news of the intervention to be communicated, or only after enough violators had been punished for “word to get out on the street.”

There are many points, often accompanied by a burst of publicity, at which 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 into law,
- (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, or
- (9) publicity about the law peaks.

Indeed, there would seem to be few time points anywhere near the “intervention point” which would *not* be plausible as points at which the intervention’s impact could begin. 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. Use of this date as the intervention point is therefore arbitrary. Nevertheless, the traditional ITSD analysis almost never considers any of these alternatives or tests for apparent “effects” when differing intervention points are used.

The peculiarities of D.C.’s handgun freeze highlight the difficulties of determining when an intervention’s impact is supposed to begin. Loftin et al. (1991) assumed that the law’s impact began at the law’s effective date of 9-24-76. However, even the effective date for this law was ambiguous because it took effect temporarily on 9-24-76, but then the deadline for owners of registered handguns to re-register their guns was extended, followed by legal challenges which resulted in the law being suspended for two months, with the law finally becoming fully effective on 2-21-77, five months after the initial effective date. Complicating matters further, the D.C. law did not necessarily 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 re-registered 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 have to be gradual.

Unfortunately, if analysts tested for all the more plausible impact points, they would run into the problem of “dredging the data” for supportive results through the use of multiple *ex post facto* hypothesis tests. This would artificially increase the chances of obtaining results indicating an apparent successful impact of the intervention, merely by increasing the number of tests performed (Selvin and Stuart 1966). Once it is realized how numerous the plausible alternative versions of the impact hypothesis are, the policy efficacy hypothesis begins to increasingly look like it is unfalsifiable through interrupted time series tests.

3. Specification of the Impact Model

If an intervention is going to have an effect on the targeted behavior, it is likely to take one of four forms:

- (1) abrupt and permanent, where there is an immediate effect of the policy change that has a long-term impact on behavior,
- (2) abrupt and temporary, where there is an immediate impact of the policy change, but its effect is short-lived,
- (3) gradual and permanent, where the policy change has only a minor effect on behavior shortly after it went into effect, but as time passes, there is an increasing impact on the target behavior, and
- (4) gradual and temporary, where the policy is slow to take effect, and then gradually diminishes in having any effect on the target behavior.

Unfortunately, much of the literature on the statistical modeling of the impact emphasizes the empirical results to the exclusion of any meaningful theoretical argument that calls for a specific type of intervention (see, e.g., Loftin et al., 1991).

The importance of this issue is related to how one interprets the statistical results. For example, if theory suggests that a law’s effect on the target behavior will be gradual (e.g., as in any law with a “grandfather” clause), but the gradual impact model did not fit the data very well, then in light of the nature of the intervention, one reasonable interpretation would be to conclude that the intervention did not have any impact, on the theoretically-based assumption that, if the law was effective, its impact had to be gradual.

Loftin et al. (1991) again provide an illustration of this problem in gun control research. They concluded that the D.C. law had an abrupt and permanent impact on gun homicides, since the ARIMA model specifying an abrupt impact 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 were clearly aware of this, since 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 policy interventions will allow such a clear-cut theoretically based choice of intervention impact patterns, yet Loftin et al. made a purely *ex post facto* choice of a less theoretically appropriate model solely because it fit the sample data better. This represents the triumph of technique over substance. If *a priori*

theory (or common sense) could play no role whatsoever in model specification in such a clear-cut case, it is hard to see how it could ever do so in any ITSD evaluation.

III. An Empirical Demonstration

So far, we have described problems that are inherent in almost any use of the ITSD, stressing especially flaws in the logic of the design. Logical argumentation alone, however, cannot indicate just how seriously astray the analyst can be led by this approach. In the following sections we illustrate these problems by replicating one of the more sophisticated applications of the strategy (i.e., Loftin et al., 1991) and then demonstrating how the conclusions reached by its users collapse when each of several features of the analysis is altered. Loftin et al. used both ARIMA methods and a simple before-after comparison of mean violence counts. Our analysis uses only the more sophisticated ARIMA methods commonly applied in ITSD studies.

The Loftin et al. study of Washington's handgun ban is arguably the most sophisticated of the ITSD analyses of gun laws. It certainly was the most highly publicized, as its publication in the *New England Journal of Medicine* was accompanied by a national press release and front page stories in newspapers across the nation. Though the article addressed suicides as well as homicides, it is sufficient for our purposes to confine the reanalysis to homicides.

The ARIMA models used in the following interrupted time series analyses were identified using standard model development procedures (e.g., McDowall et al. 1980; Wei 1990). In addition to visual inspection of the autocorrelations and partial autocorrelations, we used tests for the normality of the residuals and the Akaike Information Criteria (AIC) to assist our selection of the most appropriate time series model. Following the identification of the univariate ARIMA model, we then added the intervention parameter to test for a change in behavior that reflected a change in criminal law. Tables 4 through 9 present our results from this exercise.

The original D.C. study used vital statistics data on homicides. We used police-based data, derived from the FBI's Supplementary Homicide Reports program (ICPSR 1991), instead, for two reasons. First, Loftin and McDowall refused to provide us with a copy of their data, making a direct reanalysis of their published work impossible. Without their exact dataset, it would be impossible to determine whether any differences in results were due to differences in analytic procedures or to differences in the datasets produced in transferring data from vital statistics computer files to the files actually used for analysis.

Further, we believe that police-based local homicide counts are in any case superior to vital statistics data, for several reasons. First, the former properly exclude many justifiable civilian homicides which the latter do not (Kleck 1991). Second, the latter often erroneously includes negligent vehicular homicides which, being accidental, should not be grouped with intentional killings (Reidel 1990, p. 200). Because medical examiners and coroners rarely would know about homicides unknown to police (who are virtually the sole source of their caseload), this means that when vital statistics counts are higher than police counts, it is ordinarily attributable to this sort of vital statistics system classification error, rather than to superior coverage of homicide events.

Third, vital statistics data do not actually count the number of homicide attacks that occur in a given area, as police data do, but

rather the number of homicide *deaths* that occurred there. Thus, a victim who was shot outside the District of Columbia but who died at a hospital just the inside the border would be wrongly counted as a D.C. homicide and hence as a “failure” of the D.C. gun laws. This would be erroneous since there is no strong reason for D.C. laws to prevent shootings in areas not subject to its laws. And of course, the reverse error could also occur if the victim of a D.C. attack died in a nearby Virginia or Maryland hospital.

For large areas like states this would often be a minor concern, since only a small fraction of assaults occur close to the area’s borders. D.C., however, covers only 61 square miles, and no point is more than five miles from the nearest border with Virginia or Maryland. Seven of the District’s 14 highest homicide census tracts were within one mile of its Southeast border (Harries 1990, p. 111). All but one of D.C.’s certified hospital trauma centers are within three miles of the border (at most a six minute ambulance ride, even at 30 miles per hour), including Washington Hospital Center, which handles 30-40% of the city’s gunshot wound patients treated at trauma centers. Four others in Virginia and Maryland are also this close to the border (American Hospital Association 1990, pp. A88, A89; Webster et al. 1992). Consequently one cannot tell from vital statistics data how many homicide attacks occurred in D.C., or in any other city, county or similarly small area. (See Table 4.)

As it turns out, these flaws in the vital statistics homicide were apparently substantial enough to alter the evaluation of D.C. homicide trends. Loftin et al. based their favorable assessment of the law’s impact on homicides on two ARIMA findings: the “impact” parameter estimate was significant for gun homicides, and was not significant for nongun homicides, supposedly suggesting that there was something gun-related responsible for the pattern. Table 4 shows that analysis of D.C. Supplementary Homicide Report-based counts yields an “impact” estimate of -3.2321 in the gun homicide model, within 5% of the -3.4068 estimate produced by Loftin et al. using vital statistics. The SHR-based analysis, however, also finds a significant “impact” estimate for nongun homicides. These findings do not fit the gun/nongun comparison of Loftin et al. as well as their own findings, since they seem to suggest that something that affected nongun homicides as well as gun homicides was driving D.C. homicide trends during this period. (See Table 5.)

Instead of using the very dissimilar D.C. suburbs as a control area, we used the very similar Baltimore. Table 5 displays estimates from an analysis of Baltimore gun and nongun homicides over the 1968-1987 period, using an “intervention” point of October 1, 1976, the same as that used by Loftin et al. They show a negative, significant “impact” estimate for Baltimore gun homicides which is nearly as large (87%) as the corresponding estimate for D.C. Further, they show a far smaller, though significant, drop in nongun homicides, again the same as we found in D.C.

The problem with applying the Loftin et al. inferential logic here is that Baltimore had no new gun laws in or around October 1976. This demonstrates three points. First, it is certain that something other than new gun laws caused this pattern of ARIMA results in Baltimore, and that larger drops in gun violence than in nongun violence can be entirely due to causes other than new gun laws. Second, it is indisputable that using a more appropriate control area can alter and even reverse the conclusions implied by the analysis. If a similar area without a new gun law enjoyed a large drop in gun homicides, and a smaller drop in nongun homicides, it is perfectly possible that the identical D.C. pattern was also entirely due to factors other than its new gun law. Third, the gun/nongun comparison

cannot establish whether a new gun control policy caused drops in Baltimore's homicide, since this inferential logic would imply that Baltimore's homicide drops were due to such a legal change, an interpretation we know is impossible.

Table 6 makes explicit what Loftin et al. merely hinted at (1991, p. 1620). When the time series is extended to include just two more years' worth of time points, support for the gun law efficacy conclusion disappears. When the series covers 1968-1989 instead of 1968-1987, the impact estimate in the gun homicide equation is not significantly different from zero, nor significantly larger than the estimate in the nongun homicide model. Very likely, conflict linked with crack cocaine trafficking was an important confounding factor in 1988-1989, but then other confounding factors were also operating throughout the 1968-1987 period as well. If the univariate ARIMA model fails to deal with the effects of the crack combat during the later years, it also has the same flaw for other confounding factors in earlier years. The Loftin et al. results were extremely fragile, strongly dependent on use of a sharply, and (given availability of police-based homicide data for 1988-1989) needlessly limited set of time points.

When did the supposed "effect" of the D.C. handgun freeze begin? A univariate longitudinal impact assessment of any kind leans its case for an impact heavily on the temporal correspondence between the intervention and shifts in the target variable. Unfortunately, neither the D.C. study, nor any of the ITSD studies in Table 1, could actually establish when the favorable shifts in violence occurred. Instead, ITSD ARIMA analysts simply specify intervention models that assume that an intervention's impact, whether abrupt or gradual, began at a single particular time point (nearly always a law's "effective date"), comparing time points after this point with those before.

Table 7 shows that if one assumes that the intervention occurred in D.C. two years before the handgun law actually went into effect, one obtains the exact same combination of results as were obtained in the original analysis: a large significant drop in gun homicides and no significant change in nongun homicides. There was no new gun law introduced in D.C. in October of 1974 to produce this pattern of trends. Again, a nonexistent, or "bogus" intervention, generated as much apparent support for the policy efficacy hypothesis as the actual intervention. This exercise is a variation on the bogus intervention analysis of James Baron and Peter Reiss (1985, pp. 355-7).

We expanded this exercise by trying out many other bogus intervention points, at six month intervals before and after the actual intervention. The results, shown in Figure 1, indicate that *every one* of the bogus intervention points tested anywhere within four years of the actual intervention generated a significant negative impact estimate in the gun homicide equation. Indeed, the strongest estimated "effects" did not even coincide with the handgun ban. The largest estimates were for points 6-18 months *before* the ban. In this respect, the ARIMA analyses merely confirmed what was evident from a cursory visual examination of the simple gun homicides trend diagram in the original article - a decline in gun homicides was already underway well before the law went into effect or was even proposed (see Fig. 1, p. 1616 in Loftin et al. 1991). It is clear that the D.C. law simply did not correspond in time with the beginning of the decline in gun homicides, regardless of whether one uses ARIMA methods or simple visual inspection of the trends.

Thus, bogus intervention points, corresponding to nonexistent gun law changes, generate as much or more evidence of a

supposed “impact” as the actual intervention point. One could choose any of nearly a hundred different months as the “intervention” point, apply the Loftin et al. methods, and discover a policy “impact.” The tremendous flexibility of the method is disturbingly apparent. An incautious analyst could seize upon virtually any arguably violence-related development in D.C. occurring or beginning anytime during the 1972-1980 period, test for an impact using these methods, and come up with evidence indicating, according to the ITSD logic, that the policy caused a reduction in gun homicide.

Loftin and his colleagues specified an intervention model that assumed that the handgun ban exerted an abrupt impact, despite their knowledge that observers expected a gradual impact. Table 8 shows that it was necessary to specify an abrupt impact, however implausible, in order to obtain results supporting the efficacy hypothesis, since the more theoretically plausible specification of a gradual impact results in a gun homicide impact estimate not significantly different from zero. If the authors knew that the abrupt model fit the data better than the gradual model, this necessarily implies that they did estimate the gradual impact model and presumably obtained results very similar to those we obtained. They did not, however, report these unsupportive results to their readers. Instead, they flatly stated that the handgun ban had “truly preventive” effects (p. 1619) and that they had provided “strong evidence” that the law reduced homicides (p. 1620).

One further exercise is informative with regard to the utility of any ITSD evaluation of gun laws. The problems with the approach can be demonstrated by showing that interventions exactly opposite in character can yield precisely the same appearance of a beneficial “impact.” Scholars dispute whether the generally moderate existing gun controls reduce violence, but few have concluded that they increase violence. Empirical evidence instead generally indicates that existing moderate regulatory measures are merely ineffective (Kleck 1991, Chapter 10).

Therefore, *repealing* gun laws should either increase violence (if one assumes they suppressed violence while in effect) or have no impact (if one assumes they had no impact while in effect). There are many examples of gun laws being repealed in recent years. The NRA’s success in getting “state preemption” laws passed was arguably the dominant gun control trend of the 1980s. By passing such a law, the state preempts the field of gun control, accomplishing either or both of two things: it repeals existing local (municipal or county) gun controls, and forbids passage of other local controls in the future. Thus, passage of such a law is a sort of “anti-gun-control” or gun decontrol.

Louisville, Kentucky, a city with about 300,000 people in 1980, is illustrative. Before 1984 it had an extensive array of local gun controls, including: (1) a ban on handgun sales to members of various high-risk groups (criminals, minors, fugitives, etc.), (2) a ban on possession of handguns by such persons, (3) local gun dealer licensing, (4) a waiting period on handgun sales, and (5) local police registration of handgun sales and transfers. The last control was especially noteworthy because it covered private transfers as well as those involving licensed dealers, an uncommonly comprehensive feature (U.S. BATF 1984, pp. 55-6).

In 1984, however, Kentucky passed a state preemption bill that wiped out all local gun regulations, including Louisville’s. The relevant part of the Kentucky statutes reads: “Local firearms control ordinances prohibited. No city, county or urban-county

government may occupy any part of the field of regulation of the transfer, ownership, possession, carrying or transportation of firearms, ammunition, or components of firearms or combination thereof” (Kentucky 1990, p. 38 [Kentucky Revised Statutes 65.870]). The law’s effective date was July 13, 1984.

Table 9 shows the results of a univariate ITSD analysis of Louisville monthly gun and nongun homicide counts from January, 1976 to December, 1986, assuming the gun de-control intervention began on July 1, 1984. The impact estimates are significant and negative for the gun homicide model and insignificant for the nongun homicide model. Thus, following the methods and inferential logic of Loftin and his colleagues, one would have to conclude that *repealing* Louisville’s gun controls saved lives.

We do not believe that this is actually what happened. We suspect that it is more likely that the repeal of these controls had little or no impact, for good or ill. The point to this exercise is merely to demonstrate how easily the research design yields seemingly absurd results. Interventions exactly opposite in character can yield identical patterns of findings, leading to the unlikely conclusion that both passing handgun restrictions and repealing them reduces violence. We are not recommending a new round of ITSD analyses of state preemption laws to balance out the existing studies of new gun laws. Rather, we conclude that it is pointless to apply so dubious a methodology to the evaluation of any kind of intervention, no matter what its character.

In sum, three different kinds of “bogus” interventions all generated findings which appear to indicate an “impact” of policies, if one follows the methods and logic of ITSD approaches to gun control impact evaluation. There was a spurious appearance of an impact when the analysis assumed gun-related policy interventions for time points where there were no such interventions (the “bogus” intervention points), when the analysis was applied to an area (Baltimore) that had no such interventions, and when the analysis was applied to an actual intervention (state preemption in Kentucky) that was exactly opposite in character to laws restricting guns.

The authors of the ITSD studies summarized in Table 1 did not perform any of the tests for robustness that we have applied to the D.C. data. 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 or inferior to those applied to the D.C. data, are afflicted by the same flaws as the Loftin et al. D.C. study. Consequently, we believe that their results should be regarded, at least until these robustness tests are performed, as being at least as unreliable as those generated in the Loftin et al. D.C. study.

IV. Discussion

If the ITSD approach is so obviously inadequate, what accounts for its popularity? One explanation would be that if one is committed to determining whether one particular intervention in one particular site was effective, there often is no practical alternative to an ITSD case study. Rather than simply admitting that there are no sound, feasible methods for assessing whether a specific policy had an aggregate impact in a given city or state, many scholars would rather do the best they can, no matter how misleading their results might be, based on the dubious faith that some information is bound to be better than none.

Another explanation is simply that the approach is so easy. The univariate ITSD analyst does not have to learn anything about the causes of a phenomenon to apply univariate ARIMA analysis to it, since one does not have to devise an explanatory model. More importantly, one does not have to devote the hundreds or thousands of hours in tedious data gathering which multivariate researchers spend in measuring possible confounding factors (e.g. Kleck and Patterson 1993). There is always an attraction to getting something for nothing. ARIMA analysis is arguably the last major category of social science inquiry where univariate research is still considered respectable. This presumably is due to the faith that ARIMA modeling somehow “controls” for the “systematic” sources of variation in the series, leaving only a few sources of “nonsystematic” variation uncontrolled. Of course, if this were true, advocates of the approach would have little reason for bothering to analyze control series or developing multivariate ARIMA methods.

Finally, with respect to assessments of politically charged interventions, there is a strong ideological attraction to the ITSD approach. It is so flexible, so manipulable, that one can obtain almost any results one likes, merely by being careful in one’s selection of intervention type, historical era, intervention site, time series endpoints, intervention impact model, and control areas. The U.S. has thousands of legal jurisdictions, each with a different array of laws. One may choose from among hundreds of possible types of gun control, and for any given type, can often choose from among dozens or hundreds of different sites where the measures have been implemented. If one is opposed to gun control, one can simply select the weakest forms of control to assess, nonrandomly select sites where crime increased after the measures were implemented, or study time periods when gun and nongun violence trends were generally inconsistent with the gun control efficacy hypothesis. Conversely, if one were pro-control, one could make the opposite choices.

One can also vary the design details and inferential logic to suit one’s policy preferences. If a crude ITSD analysis without any control series yields the desired results, the analyst can stop there. If not, the analyst can add a control area which showed even worse (or better) trends in the target variable than the intervention area. Thus, if there is a decrease in gun violence around the time a law went into effect, one can conclude a gun law worked. However, even if there was an increase, or no change, in gun violence, one can then search for a comparison series which showed an even bigger increase and argue that the gun law had a “dampening” effect on violence, preventing it from being even worse than it otherwise would have been (for an example of this very line of reasoning, see O’Carroll, Loftin, Waller, McDowall, Bukoff, Scott, Mercy, and Wiersema 1991).

All of this would matter very little if ITSD studies yielded the same results as those generated by other approaches. In the gun control area, this is clearly not true. The results of ITSD studies stand out as anomalies. In general, the technically strongest research in the area indicates that all but a few types of gun control have no impact on the frequency of any form of violence, including homicide. Most of the exceptions to this generalization, however, used the ITSD approach. One review covering the pre-1990 research indicated that of 29 studies on gun control impact on crime, only three generally supported the hypothesis that gun laws reduced violence, with another eight providing some mixed or partial support, while 18 were consistently unresponsive. Among studies using non-ITSD methods, only 4 of 17 yielded results even partially supportive, while 7 of 12 studies using ITSD methods generated supportive results

(Kleck 1991, p. 417). The choice of research designs apparently does make a difference.

V. Conclusions

The ITSD approach is so deficient for purposes of policy impact assessment and hypothesis testing that it would not be an overstatement to describe it as “subscientific.” If one cannot rule out *any* rival explanations of trends in the target variable, then attributing them to an intervention amounts to little more than an idle guess, based on a very rough temporal coincidence. The Washington, D.C., study by Loftin et al. illustrates that ITSD findings are often so fragile that even the slightest changes in study design can completely overturn the conclusions. The appearance of a beneficial impact on homicide disappeared once any one of the following changes were made:

- (1) using a different source of homicide data,
- (2) using a more comparable control jurisdiction,
- (3) extending the time series by just two years, or
- (4) using a more theoretically appropriate impact model.

It is something of a mystery how univariate nonexperimental analysis of any kind, no matter how dressed up in statistical finery, can still be considered respectable at this late date. Perhaps ITSD studies enjoy a certain amount of unearned prestige from being labeled “quasi-experimental,” even though they are actually nonexperimental. This unfortunate label hints that the design has some of the significant features that make the internal validity of experiments so strong. However, the key feature of experimentation responsible for this strength is the ability of researchers to randomly assign or control treatments, i.e. to manipulate the cause or independent variable. The ITSD researcher does not enjoy this advantage. Scholars in general cannot do this with evaluations of new laws, and only rarely can do it with other public policies affecting large populations. Further, the fact that two groups, loosely labeled “experimental” and “control,” are sometimes used in ITSD studies does not make the research experimental in any sense. Even the use of time-ordered data is a minor secondary feature of experiments, usually unnecessary for drawing strong causal inferences.

It is time to acknowledge what should have been obvious, and recognize that this emperor has no clothes. What then is the alternative? Are we stuck with the ITSD approach on the premise that it is better than nothing? We would suggest that the approach can in practice be considerably worse than nothing, being so subject to illegitimate manipulation, so easily used to confirm a researcher’s preconceived biases, that use of the approach can be worse than no research at all. Sometimes it is better to simply say “we do not know” than to suggest that we can know, using methods which are prone to distortion and systematic error. More specifically, in cases like evaluation of the impact of new laws, where true policy experimentation is impossible, it may be best to say we simply have no sound way to assess whether a specific intervention worked in a particular locale.

This does not, however, imply that we cannot come to stronger conclusions about whether a *category* of interventions, such as a type of law, implemented in many different areas, has had an impact. One can, for example, assess whether laws requiring a waiting period before buying a gun, operating in dozens or hundreds of cities, have, on average, reduced crime. Once one shifts to a cross-sectional approach, comparing areas having a policy with areas lacking the policy, it is possible to use data from the Census and many other sources to measure and explicitly control for dozens or hundreds of possible confounding factors, and to estimate more realistic multivariate models (see Kleck and Patterson 1993 for an example). One will still be constrained by limits on both data and credible theory, but these same problems also afflict ITSD approaches, whether acknowledged or not. The main difference is that with a cross-sectional approach, the data constraints are much weaker and the analyst can explicitly rule out hundreds of specific rival explanations for observed associations between policies and target variables, while the univariate ITSD approach allows one to explicitly rule out none of them. Furthermore, as an empirical matter, it turns out that, in cross-sectional studies, specification of which control variables to include in the model is less consequential than analysts assumed. In contrast to the strong cross-temporal correlations found in time series studies, the presence or absence of gun laws has little or no correlation, across legal jurisdictions, with other known determinants of violence rates. Consequently, cross-sectional estimates of gun law impact are not substantially influenced by control variable specification decisions (Kleck and Patterson 1993).

Before-and-after comparisons are an essential part of how humans learn about how the world works. Often, our own personal experiences suggest the value of this general methodology for learning about our immediate environment; we take an action (the “intervention”) and observe the changes which immediately follow (the “impact”), and reasonably infer a connection between the two. Unfortunately, when one extends this same methodology to the evaluation of public policy impact, it is easy to overlook how drastically the application situation differs. Evaluating public policy impact involves assessing very remote causal effects on the “behavior” of aggregates composed of thousands or millions of individual persons, not the immediate impact of an individual action on a very constricted personal environment. In this light, the intuitive “common-sense” appeal of before-and-after comparisons becomes a danger because it short-circuits critical thinking.

Many of our criticisms have been stated in the scattered technical literature before (e.g. Cook and Campbell 1979). These prior statements, however, have evidently not been sufficiently influential on research practice, since these methods continue to be applied without users attempting to deal with the criticisms, and researchers continue to draw extremely strong conclusions that would not follow if the criticisms had been taken seriously. Consequently, we feel fully justified in our efforts, even if we have run the risk of going over some of the same ground as others have.

Skepticism about the long-accepted virtues of longitudinal research has been growing in recent years. For example, Gottfredson and Hirschi (1987) have questioned the value of longitudinal studies of delinquency causation, while Isaac and Griffin (1989) have challenged time series analyses of historical processes. It is time that this skepticism was extended to the use of ITSD for assessing public policy impact.

The problems with ITSD research are both so serious and so inherent in the logic of the research design (and in the severe, uncorrectable limits on availability of subnational time series data) that the approach appears to be unsalvageable. For now at least, the best course may be to abandon use of univariate time series analysis for hypothesis-testing purposes and confine its use to simple descriptive applications.

In any case, a superior alternative approach has recently become popular. The pooled cross-sections or multiple time series approaches exploit both cross-sectional and cross-temporal variation in the target variable, for large numbers of cross-sectional units. Marvell and Moody (1995) and Lott and Mustard (1997) have both used these designs to evaluate the impact of gun laws. Although these designs share with the ITSD design a limited ability to explicitly rule out rival explanations of trends in the target variable, they are far less subject to problems like biased selection of intervention and control areas and small sample size, since all relevant areas are typically studied.

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

Study	Location of Intervention	Date of Intervention	Control Nongun Series? ^b	Control Other Areas Series? ^b	Type of Intervention
Deutsch and Alt (1977)	Boston	4-1-75	No	No	Mandatory penalty for unlawful carrying
Hay and McCleary (1979)	Boston	4-1-75	No	No	Mandatory penalty for unlawful carrying
Deutsch (1981)	Boston	4-1-75	No	No	Mandatory penalty for unlawful carrying
Pierce and Bowers (1981)	Boston	4-1-75	No ^c	No ^c	Mandatory penalty for unlawful carrying
Loftin et al. (1983)	Detroit	1-1-77	Yes	No	Mandatory 2 year add-on penalty for felony w. gun
Loftin & McDowall (1984)	3 Florida cities	10-1-75	Yes	No	Mandatory minimum 3 years for gun possession during felonies
McPheters et al. (1984)	2 Arizona counties	8-1-74	No	No ^d	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
Loftin, McDowall, Wiersema and Cottey (1991)	Washington, D.C.	9-24-76	Yes	Yes	Ban on handgun possession, with "grandfather clause"
McDowall, Loftin and Wiersema (1992)	Detroit, Jacksonville, Tampa, Miami, Pittsburgh,	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

	Philadel phia				
--	------------------	--	--	--	--

Notes:

- a. Table covers published studies using ARIMA analytic methods. Simple before-and-after comparisons (e.g. Zimring 1975; Lucas and Ledgerwood 1978; Fife and Abrams 1989) are not covered. Also, where overlapping studies reported the same basic data twice (e.g. Loftin and McDowall 1981 and Loftin et al. 1983), only one is listed.
- b. Was gun crime series compared with corresponding nongun series (e.g. gun homicides compared with nongun homicides)? Was series in intervention area compared with series in nonintervention area?
- c. No ARIMA estimates were reported for nongun crime or for control areas; only simple before-and-after percentage changes.
- d. Control area was used for paired t-tests, but not for ARIMA analyses.

Table 2. Homicide Trends in Washington, D.C., Its Suburbs, and Baltimore, 1968-1990.

Year	Number of DC Homicides	DC Homicide Rate	Number of Homicides in SMSA for DC, excluding DC	Homicide Rate for DC suburbs	Number of Baltimore Homicides	Baltimore Homicide Rate
1968	178	22.19	52	2.76	200	22.23
1969	287	36.82	62	3.07	236	26.14
1970	221	29.21	105	4.99	231	25.50
1971	275	36.54	82	3.81	323	35.78
1972	245	32.58	122	5.54	330	37.12
1973	268	35.93	131	5.74	280	32.14
1974	277	38.43	131	5.65	293	33.63
1975	235	33.08	130	5.61	259	30.65
1976	188	26.85	121	5.10	200	24.43
Before and after division for DC handgun law						
1977	192	28.03	121	5.13	171	21.18
1978	189	28.19	106	4.48	197	24.87
1979	180	27.44	101	4.29	245	30.98
1980	200	31.33	126	5.24	216	27.46
1981	223	35.06	127	5.19	228	29.17
1982	194	30.74	140	5.68	227	29.44
1983	183	29.38	115	4.29	201	26.31
1984	178	28.58	107	3.87	215	28.33
1985	147	23.48	96	3.38	213	28.19
1986	194	30.99	104	3.61	240	30.63
1987	225	36.17	142	4.75	226	30.30
1988	369	59.81	178	5.76	234	31.14
1989	434	71.85	206	6.51	262	34.33
1990	472	77.77	212	6.39	305	41.44

Source: U.S. FBI, *Uniform Crime Reports*, annual issues for 1968-1990.

Notes: Figures for the remainder of the D.C. metropolitan area were obtained by subtracting D.C. figures from the D.C. SMSA crime and population counts.

D.C. gun law first became effective on 9-24-76.

Bivariate correlations of annual homicide rates, 1968-1976:

D.C. and rest of D.C. metro area: 0.313 ($p > .10$)

D.C. and Baltimore: 0.708 ($p < .05$)

Table 3. Trends in Gun and Nongun Violent Crime, U.S., 1961-1990.

Year	Murder & Nonnegligent Manslaughter Rate	% with guns	Rate of Murder & Nonnegligent Manslaughter with guns	Robbery Rate	Robbery % with Guns	Gun Robbery Rate	Aggravated Assault Rate	Assault % with Guns	Gun Assault Rate
1961	4.8	52.5	2.52	58.3			85.7		
1962	4.6	54.2	2.49	59.7			88.6		
1963	4.6	56.0	2.58	61.8			92.4		
1964	4.9	55.0	2.70	68.2			106.2	15	15.9
1965	5.1	57.2	2.92	71.7			111.3	17	18.9
1966	5.6	60.0	3.36	80.8			120.0	18.8	22.6
1967	6.2	63.6	3.94	102.8			130.2	20.9	27.2
1968	6.9	65.4	4.53	131.8			143.8	23.1	33.2
1969	7.3	64.5	4.73	148.4			154.5	23.8	36.8
1970	7.9	65.4	5.15	172.1			164.8	24.3	40.0
1971	8.6	65.1	5.61	188.0			178.8	25.1	44.9
1972	9.0	66.2	5.94	180.7			188.8	25.3	47.8
1973	9.4	67.0	6.27	183.1			200.5	25.7	51.5
1974	9.8	67.9	6.65	209.3	44.7	93.6	215.8	25.4	54.8
1975	9.6	65.8	6.33	218.2	44.8	97.8	227.4	24.9	56.6
1976	8.8	63.8	5.58	195.8	42.7	83.6	228.7	23.6	54.0
1977	8.8	62.5	5.52	187.1	41.6	77.8	241.5	23.2	56.0
1978	9.0	63.6	5.70	191.3	40.8	78.1	255.9	22.4	57.3
1979	9.7	63.3	6.17	212.1	39.7	84.2	279.1	23.0	66.7
1980	10.2	62.4	6.38	243.5	40.3	98.1	290.6	23.9	69.5
1981	9.8	62.4	6.13	258.7	40.1	103.7	289.7	23.6	68.4
1982	9.1	60.2	5.46	238.9	39.9	95.3	289.2	22.4	64.8
1983	8.3	58.3	4.81	216.5	36.7	79.5	279.2	21.2	59.2
1984	7.9	58.8	4.65	205.4	35.8	73.5	290.2	21.2	61.2
1985	7.9	58.7	4.67	208.5	35.3	73.6	302.9	21.3	64.5
1986	8.6	59.1	5.05	225.1	34.3	77.2	346.1	21.3	73.7
1987	8.3	59.1	4.88	212.7	33.0	70.2	351.3	21.4	75.2
1988	8.4	60.7	5.11	220.9	33.4	73.8	370.2	21.1	78.1
1989	8.7	62.4	5.40	233.0	33.2	77.4	383.4	21.5	82.4
1990	9.4	64.1	6.04	257.0	36.6	94.1	424.1	23.1	98.0

Sources: Total crime rates, 1961-1975: 1975 issue of *Uniform Crime Reports* (UCR), p. 49. Total crime rates, 1976-90, % gun, all years: each annual UCR issue for the corresponding year.

Notes: Gun rates were computed by multiplying the total crime rates (e.g. total robbery rate) by the corresponding % gun (e.g. % gun in robberies). Blank entries indicate that relevant data were not available.

Table 4. Replication With Police Data: District of Columbia Homicides, 1968-1987.

Panel A: District of Columbia Gun Homicides

Parameter	Replication (FBI Data)			Loftin et al. (Vital Statistics Data)			
	Coefficient Estimate	Standar d Error	Ratio	Parameter	Coefficient Estimate	Standar rd Error	Ratio
α	12.5706	.4989	25.20	α	13.1256	.5032	26.09
ϕ_1	.1367	.0649	2.11	ϕ_1	.1641	.0641	2.56
ϕ_2	.1357	.0651	2.08	ϕ_2	.1274	.0639	1.99
ω_o	-3.2321	.6649	-4.86	ω_o	-3.4068	.6650	-5.12

Q = 21.50, 22 df

Panel B: District of Columbia Non-Gun Homicides

Parameter	Replication (FBI Data)			Loftin et al. (Vital Statistics Data)			
	Coefficie nt Estimate	Standar rd Error	Ratio	Para met er	Coefficient Estimate	Standar d Error	Ratio
α	7.7429	.2754	28.11	α	7.3615	.3105	23.71
ϕ_1	.0587	.0656	.90	ϕ_1	.1288	.0645	2.00
ω_o	-1.1197	.3670	-3.05	ω_o	-.3915	.4126	-.95

Q = 23.68, 24 df

Table 5. Use of a More Appropriate Control Area: Baltimore Homicides, 1968-1987.

Panel A: Baltimore Gun Homicides

Parameter	Coefficient Estimate	Standard Error	Ratio
α	13.7050	.7979	17.18
ϕ_1	.2842	.0603	4.71
ϕ_4	.2427	.0605	4.01
ω_o	-2.8114	1.0532	-2.67

Q = 26.55, 22 df

Panel B: Baltimore Non-Gun Homicides

Parameter	Coefficient Estimate	Standard Error	Ratio
α	8.4670	.2900	28.36
ω_o	-1.1930	.3870	-3.08

Q = 22.20, 24 df

Table 6. Time Series Extended By Two Years: District of Columbia Homicides, 1968-1989.

Panel A: District of Columbia Gun Homicides

Parameter	Coefficient Estimate	Standard Error	Ratio
α	11.096	1.978	5.61
ϕ_1	.351	.060	5.85
ϕ_2	.271	.061	4.44
ϕ_3	.221	.060	3.67
ω_0	1.525	2.458	.62
Q = 24.33, 21 df			

Panel B: District of Columbia Non-Gun Homicides

Parameter	Coefficient Estimate	Standard Error	Ratio
α	7.7180	.5351	14.42
ϕ_1	.2357	.0606	3.89
ϕ_2	.2116	.0606	3.49
ω_0	-.5034	.6869	-.73
Q = 24.81, 22 df			

Table 7: “Bogus Intervention” at October 1974: District of Columbia Homicides, 1968-1987.

Panel A: District of Columbia Gun Homicides

Parameter	Coefficient Estimate	Standard Error	Ratio
α	12.5986	.6235	20.21
ϕ_1	.1683	.0645	2.61
ϕ_2	.1645	.0646	2.55
ω_o	-2.7948	.7649	-3.65

Q = 19.26, 22 df

Panel B: District of Columbia Non-Gun Homicides

Parameter	Coefficient Estimate	Standard Error	Ratio
α	7.8272	.3146	24.88
ω_o	-1.0787	.3865	-2.79

Q = 26.27, 24 df

Table 8. A Theoretically More Appropriate Gradual Impact Model: District of Columbia Homicides, 1968-1987.

Panel A: District of Columbia Gun Homicides

Parameter	Coefficient	Standard	Ratio
	Estimate	Error	
α	12.6795	.5009	25.32
ϕ_1	.1315	.0652	2.02
ϕ_2	.1427	.0658	2.17
ω_o	-1.4005	1.1702	-.65
δ	.5880	.6382	.92

Q = 22.08, 22 df

Panel B: District of Columbia Non-Gun Homicides

Parameter	Coefficient	Standard	Ratio
	Estimate	Error	
α	7.7829	.2757	28.22
ω_o	-2.0855	.7640	-2.73
δ	-.7981	.4227	-1.89

Q = 22.11, 24 df

Table 9. "Impact" of Gun Deregulation (July, 1984) on Louisville Homicides, 1976-1986.

Panel A: Louisville Gun Homicides

Parameter	Coefficient Estimate	Standard Error	Ratio
α	3.4024	.2740	12.42
ϕ_2	.2607	.0854	3.05
ω_0	-1.5529	.5632	-2.76
Q = 22.92, 23 df			

Panel B: Louisville Non-Gun Homicides

Parameter	Coefficient Estimate	Standard Error	Ratio
α	1.2255	.1200	10.21
ω_0	-.1745	.2518	-.69
Q = 26.64, 24 df			

References

- American Hospital Association (AHA). 1990. *American Hospital Association Guide to the Health Care Field*. Chicago: AHA.
- Baron, James N., and Peter C. Reiss. 1985. "Same Time, Next Year." *American Sociological Review* 50:347-63.
- Bonham, Carl, Edwin Fujii and Eric Im. 1992. "The Impact of the Hotel Room Tax: An Interrupted Time Series Approach (Hawaii)." *National Tax Journal* 45:433-41.
- Box, G. E. P., and Jenkins, G. M. 1976. *Time-series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Box, G. E. P., and Tiao, G. C. 1965. "A Change in Level of Nonstationary Time Series." *Biometrika* 52:181-192.
- Campbell, Donald T. and Julian Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Cantor, David, and Lawrence E. Cohen. 1980. "Comparing Measures of Homicide Trends." *Social Science Research* 9:121-145.
- Chamlin, Mitchell B., and John K. Cochran. 1998. "Causality, Economic Conditions, and Burglary." *Criminology* 36:425-440.
- Cook, Philip J. 1980. "Research in Criminal Deterrence: Laying the Groundwork for the Second Decade." Pp. 211-268 in *Crime and Justice: An Annual Review of Research*, Volume 2, edited by Norval Morris and Michael Tonry. Chicago: University of Chicago Press.
- Cook, Thomas D. and Donald T. Campbell. 1979. *Quasi-Experiments: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- Deutsch, Stephen Jay, and Francis B. Alt. 1977. "The Effect of Massachusetts' Gun Control Law on Gun-Related Crimes in the City of Boston." *Evaluation Quarterly* 1:543-68.
- Etten, Tamryn J. 1993. "Triggering criminal law." Paper presented at the annual meetings of the American Society of Criminology, Phoenix, Arizona, October 30, 1993.
- Fife, Daniel, and William R. Abrams. "Firearms' Decreased Role in New Jersey Homicides After a Mandatory Sentencing Law." *Journal of Trauma* 29:1548-51.
- Garafolo, James. 1981. "Crime and the Mass Media." *Journal of Research in Crime and Delinquency* 18:399-50.
- Gottfredson, Michael, and Travis Hirschi. 1987. "The Methodological Adequacy of Longitudinal Research on Crime." *Criminology* 25:581-614.
- Harries, Keith D. 1990. *Serious Violence: Patterns of Homicide and Assault in America*. Springfield, Ill.: Thomas.
- Hay, Richard, and Richard McCleary. 1979. "Box-Tiao Times Series Models for Impact Assessment." *Evaluation Quarterly* 3:277-314.
- Hedrick, Terry E. and Stephanie L. Shipman. 1988. "Multiple Questions Require Multiple Designs: An Evaluation of 1981 Changes to the AFDC Program." *Evaluation Review* 12:427-48.
- Hibbs, Douglas A., Jr. 1977. "On Analyzing the Effects of Policy Interventions: Box-Tiao vs. Structural Equations Models." In *Sociological Methodology 1977*, edited by Herbert L. Costner. San Francisco: Jossey-Bass.
- Inter-university Consortium for Political and Social Research (ICPSR). 1991. *Uniform Crime Reporting Program Data*. Study 9028,

- Supplementary Homicide Reports, 1975-1989. Federal Bureau of Investigation. Ann Arbor: Inter-University Consortium [distributor].
- Isaac, Larry W., and Larry J. Griffin. 1989. "Ahistoricism in Time-Series Analyses of Historical Processes." *American Sociological Review* 54:873-890.
- 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?" *The Annals* 455:138-149.
- _, and Marla Wilson Ray. 1980. *Handgun Control: Strategies, Enforcement and Effectiveness*. Unpublished report. Washington, D.C.: U.S. Department of Justice.
- Kates, Don B., Jr. 1979. "Toward a History of Handgun Prohibition in the United States." Pp. 7-30 in *Restricting Handguns: The Liberal Skeptics Speak Out*, edited by Don. B. Kates, Jr. Croton-on-Hudson, N.Y.: North River Press.
- Kennett, Lee, and James LaVerne Anderson. 1975. *The Gun in America: The Origins of a National Dilemma*. Westport, Conn.: Greenwood Press.
- Kentucky. 1990. *Kentucky Revised Statutes Annotated, 1990 Cumulative Supplement*, Volume 4. Charlottesville, Va.: The Michie Company.
- Kleck, Gary. 1979. "Capital Punishment, Gun Ownership, and Homicide." *American Journal of Sociology* 84:882-910.
- _. 1991. *Point Blank: Guns and Violence in America*. New York: Aldine.
- _. 1996. "Crime, Culture Conflict, and Support for Gun Control." *American Behavioral Scientist* 39(4):387-404.
- _, and E. Britt Patterson. 1993. "The Impact of Gun Control and Gun Ownership Levels on Violence Rates." *Journal of Quantitative Criminology* 9:249-287.
- Lieberman, Stanley. 1985. *Making It Count: The Improvement of Social Theory and Research*. Berkeley: University of California Press.
- Loftin, Colin, Milton Heumann, and David McDowall. 1983. "Mandatory Sentencing and Firearms Violence: Evaluating an Alternative to Gun Control." *Law & Society Review* 17:287-318.
- Loftin, Colin, and David McDowall. 1984. "The Deterrent Effects of the Florida Felony Firearm Law." *Journal of Criminal Law Criminology* 75:250-9.
- Loftin, Colin, David McDowall, and Brian Wiersema. 1991. "A Comparative Study of the Preventive Effects of Mandatory Sentencing Laws for Gun Crime." Discussion Paper 5, Violence Research Group, Institute of Criminal Justice and Criminology, University of Maryland.
- Loftin, Colin, David McDowall, Brian Wiersema, and Talbert J. Cottey. 1991. "Effects of Restrictive Licensing of Handguns on Homicide and Suicide in the District of Columbia." *New England Journal of Medicine* 325:1615-20.
- Lott, John, and David B. M. Mustard. 1997. "Crime, deterrence and right-to-carry concealed handguns." *Journal of Legal Studies* 26:1-68.
- Lucas, Charles E., and Anna M. Ledgerwood. 1978. "Mandatory Incarceration for Convicted Armed Felons." *The Journal of Trauma* 18:291-2.

- Marsh, Harry L. 1989. "Newspaper Crime Coverage in the U.S.: 1893-1988." *Criminal Justice Abstracts* 21:506-14.
- Marvell, Thomas B., and Carlisle E. Moody. 1995. "The impact of enhanced prison terms for felonies committed with guns." *Criminology* 33:247-281.
- McCleary, Richard, and Richard A. Hay, Jr., with Errol E. Meidinger and David McDowall. 1980. *Applied Time Series Analysis for the Social Sciences*. Beverly Hills: Sage.
- McDowall, David, Richard McCleary, Errol E. Meidinger, and Richard A. Hay, Jr. 1980. *Interrupted Time Series Analysis* Beverly Hills: Sage.
- McDowall, David, Colin Loftin, and Brian Wiersema. 1992. "A Comparative Study of the Preventive Effects of Mandatory Sentencing Law for Gun Crimes." *Journal of Criminal Law & Criminology* 83:378-94.
- McDowall, David, Brian Wiersema, and Colin Loftin. 1989. "Did Mandatory Firearm Ownership in Kennesaw Really Prevent Burglaries?" *Sociology and Social Research* 74:48-51.
- McPheters, Lee R., Robert Mann, and Don Schlagenhauf. 1984. "Economic Response to a Crime Deterrence Program." *Economic Inquiry* 22:550-70.
- O'Carroll, Patrick W., Colin Loftin, John B. Waller, David McDowall, Allen Bukoff, Richard O. Scott, James A. Mercy, and Brian Wiersema. 1991. "Preventing Homicide: An Evaluation of the Efficacy of a Detroit Gun Ordinance." *American Journal of Public Health* 81:576-81.
- Pierce, Glenn L. and William J. Bowers. 1981. "The Bartley-Fox Gun Law's Short-term Impact on Crime in Boston." *The Annals* 455:120- 137.
- Reidel, Marc. 1990. "Nationwide Homicide Data Sets." Pp. 175-205 in *Measuring Crime*, edited by Doris Layton MacKenzie, Phyllis Jo Baunach, and Roy R. Rolberg. Albany, N.Y.: State University of New York Press.
- Rock, Stephen M. 1996. "Impact of the Illinois Child Passenger Protection Act: A Retrospective Look." *Accident Analysis and Prevention* 28:487-492.
- Ross, H. Laurence, Donald T. Campbell, and Gene V. Glass. 1970. "Determining the Social Effects of a Legal Reform: The British 'Breathalyser' Crackdown of 1967." *American Behavioral Scientist* 13:493-509.
- Ross, H. Laurence, Richard McCleary, and Gary LaFree. 1990. "Can mandatory jail laws deter drunk driving? The Arizona case." *Journal of Criminal Law and Criminology* 81:156-180.
- Selvin, Hanan C., and Alan Stuart. 1966. "Data-dredging Procedures in Survey Analysis." *American Statistician* 20:20-33.
- Smith, Douglas A., and Patrick R. Gartin. 1989. "Specifying Specific Deterrence." *American Sociological Review* 54:94-106.
- Stinchcombe, Arthur, Rebecca Adams, Carol A. Heimer, Kim Lane Scheppele, Tom W. Smith, D. Garth Taylor 1980. *Crime and Punishment - Changing Attitudes in America*. San Francisco: Jossey-Bass.
- Tiao, G. C., and Box, G. E. P. 1981. "Modelling Multiple Time Series with Applications." *Journal of the American Statistical Association* 76:802-816.
- U.S. Bureau of Alcohol, Tobacco and Firearms (BATF). no date. *Concentrated Urban Enforcement*. Washington, D.C.: BATF.

- _. 1984. *State Laws and Published Ordinances, Firearms - 1984*. Washington, D.C.: BATF.
- U.S. Bureau of Justice Statistics. 1989. *Sourcebook of Criminal Justice Statistics, 1989*. Washington, D.C.: U.S. Government Printing Office.
- U.S. Federal Bureau of Investigation (FBI). 1969-1977. *Crime in the United States (year)*, annual issues covering 1968 to 1976. Washington, D.C.: U.S. Government Printing Office.
- Webster, Daniel W., Howard R. Champion, Patricia S. Gainer, and Leon Sykes. 1992. "Epidemiological Changes in Gunshot Wounds in Washington, DC, 1983-1990." *Archives of Surgery* 127:694-8.
- Wei, William W. S. 1990. *Time Series Analysis: Univariate and Multivariate Methods*. Redwood City, CA: Addison-Wesley.
- Wright, James D., and Peter H. Rossi. 1986. *Armed and Considered Dangerous: A Survey of Felons and Their Firearms*. N.Y.: Aldine.
- Zimring, Franklin E. 1975. "Firearms and Federal Law: The Gun Control Act of 1968." *Journal of Legal Studies* 4:133-98.
- __. 1978. "Policy Experiments in General Deterrence: 1970- 1975." In Blumstein, Alfred, Jacqueline Cohen, and Daniel Nagin (eds.), *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*. Washington, D.C.: National Academy of Sciences.