

The Connecticut Crackdown on Speeding: Time-Series Data in
Quasi-Experimental Analysis

Donald T. Campbell; H. Laurence Ross

Law & Society Review, Vol. 3, No. 1. (Aug., 1968), pp. 33-54.

Stable URL:

[http://links.jstor.org/sici?sici=0022-9216\(196808\)2:3;1-0;FTCCOSTW3F2.0.CO;3B2-D](http://links.jstor.org/sici?sici=0022-9216(196808)2:3;1-0;FTCCOSTW3F2.0.CO;3B2-D)

Law & Society Review is currently published by Law and Society Association.



Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/lawsa.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

<http://www.jstor.org>
Tue Aug 28 19:57:27 2007

THE CONNECTICUT CRACKDOWN
ON SPEEDING
Time-Series Data in
Quasi-Experimental Analysis

DONALD T. CAMPBELL

Northwestern University

and

H. LAURENCE ROSS

University of Denver College of Law

SOCIAL RESEARCH FREQUENTLY encounters the task of evaluating change produced in nonrandomly selected groups by events which are beyond the researcher's control. The social scientist must verify that there has in

AUTHORS' NOTE: The preparation of this paper has been supported in part by the National Science Foundation (Grant GS 1309x), the U. S. Office of Education (Project C-998, Contract 3-20-001), the U. S. Bureau of Public Roads (CPR 11-5981), the National Institutes of Health, the U. S. Public Health Service (RG-5359), and the Automotive Safety Foundation (as an aspect of Experimental Case Studies of Traffic Accidents conducted at Northwestern University). A brief version of it appears as H. L. Ross & D. T. Campbell, The Connecticut Speed Crackdown: A Study of the Effects of Legal Change, in PERSPECTIVES ON THE SOCIAL ORDER: READINGS IN SOCIOLOGY 30-35 (2d ed. H. L. Ross ed. 1968).

fact been a change, and that the indicated event is its cause. Illustrations are manifold: a state terminates capital punishment, and proponents of this type of punishment predict an increase in the murder rate; a school is integrated, and supporters of the reform expect to find an increase in the positive self-evaluation of Negro pupils; a natural disaster occurs in a community, and altruistic behavior is expected to increase. Because in these situations the investigator has no control over the assignment of individuals or groups to "experimental" and "control" situations, the logic of the classical experiment must be reexamined in a search for optimal interpretative procedures.

This paper introduces, in the context of a problem in applied sociology and the sociology of law, a mode of analysis designed to deal with a common class of situations in which research must proceed without the benefit of experimental control. The general methodology expounded here is termed "quasi-experimental analysis." The specific mode of analysis is the "interrupted time-series design." Perhaps its fundamental credo is that lack of control and lack of randomization are damaging to inferences of cause and effect only to the extent that a systematic consideration of alternative explanations reveals some that are plausible. More complete explications of quasi-experimental analysis have appeared elsewhere;¹ this paper will merely illustrate its use in a situation where a series of observations has been recorded for periods of time both prior and subsequent to the experience of the specific event to be studied. Such data are quite commonly available, yet they are seldom fully utilized and investigators often confine themselves unnecessarily to much less satisfactory methodologies. The 1955 crackdown on speeding in the State of Connecticut furnishes an apt example of the potentialities of such quasi-experimental analysis.

1. *E.g.*, D. T. Campbell & J. S. Stanley, *Experimental and Quasi-Experimental Designs for Research on Teaching*, in *HANDBOOK OF RESEARCH ON TEACHING* 171-246 (N. L. Gage ed. 1963) reprinted as *EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH* (1963); D. T. Campbell, *From Description to Experimentation: Interpreting Trends as Quasi-Experiments*, in *PROBLEMS IN MEASURING CHANGE* (C. W. Harris ed. 1963); D. T. Campbell & K. N. Clayton, *Avoiding Regression Effects in Panel Studies of Communication Impact*, in *STUDIES IN PUBLIC COMMUNICATION* 99-118 (Dept. of Sociology, University of Chicago, No. 3, 1961) reprinted in Bobbs-Merrill Reprints in Sociology as S-353. For an application of this type of analysis to legal impact, see R. Lempert, *Strategies of Research Design in the Legal Impact Study*, 1 *L. & Soc'y Rev.* 111 (1966).

A PROGRAM FOR REDUCING HIGHWAY FATALITIES

In 1955, 324 people were killed in automobile accidents on the highways of Connecticut. Deaths by motor vehicle accidents had reached a record high for the decade of the fifties as the usually hazardous Christmas holidays approached. Two days before Christmas, Governor Abraham Ribicoff of Connecticut initiated an unparalleled attempt to control traffic deaths by law enforcement, and announced his crackdown on speeders in that state.

Ribicoff believed, along with many safety specialists, that excess speed was the most common contributing factor in traffic deaths, and that control of speed would result in diminished fatalities. He believed that previous efforts to control speeding under the usual court procedures and by the existing "point system" had been inadequate. In a study of three months' records of the police court in Hartford, it was noted that no more than half the persons originally charged with speeding were so prosecuted, the charge often being diminished to a less serious one. Ribicoff wanted to initiate a program with reliable procedures and strong sanctions as a means to control speeding and thus to reduce traffic deaths.

On December 23, 1955, Governor Ribicoff announced that in the future all persons convicted of speeding would have their licenses suspended for thirty days on the first offense. A second violation was to mean a sixty-day suspension, and a third conviction for speeding would result in indefinite suspension of the driver's license, subject to a hearing after ninety days.

The decree was put into force through the Governor's power of appointment over local judges. Under Connecticut practice, the Motor Vehicle Department was suspending licenses on the recommendation of police court judges. The judges were appointed by the Governor, who threatened loss of reappointment in 1957 to judges who appeared lax in the conviction of speeders, or who did not recommend suspension of licenses to the Motor Vehicle Department.

In the first three months of 1956, license suspensions for speeding numbered 2,855, an increase of almost 2,700 over the corresponding period in 1955. There were ten fewer fatalities, and 765 fewer arrests for speeding. The Governor was reported "encouraged" by the drop in violations and in fatalities. The press quoted him as saying, "This is

positive proof that operators are not only driving slower, but are driving better."

By late May, deaths had declined from 122 in 1955 to 107 in 1956. Suspensions for speeding numbered 4,559, as against 209 in 1955. Speeding arrests had dropped 53 per cent. The Governor received a telegram of commendation for the program from the National Safety Council.

At the end of June there were twenty-two fewer fatalities than in the first six months of 1955, representing a 15 per cent reduction. Suspensions for speeding in the first six months of the year had risen from 231 to 5,398, and arrests had declined from 4,377 to 2,735. Ribicoff announced:

Connecticut has succeeded in stopping the upward surge in highway deaths, and in the first six months of this year, contrary to the national trend, we have saved lives. Fewer people died on the highways this year than in the same period last year, in Connecticut. We did it by enforcing the law, something the safety experts said couldn't be done because the people wouldn't be behind it.

In July, a new State Police program, using unmarked police cars and making extensive use of radar, was inaugurated. The police issued a report stating that 2 per cent of the cars observed by radar on July 4 were found to be speeding; at a later date, it was claimed that no speeders were found among 53,000 cars similarly observed.

In the late summer, however, Connecticut experienced a very high number of traffic fatalities. By the beginning of September, 194 people had been killed, a number almost equal to the 195 of the comparable period in the previous year. The accident "epidemic" was embarrassing to the authorities, who retreated to defending the speeding crackdown on the grounds (a) that the fatality rate remained low in comparison with the national trend, which showed a 7 per cent increase; (b) that exposure to accidents in the State had increased by 100 million vehicle miles without an increase in deaths; and (c) that the total accident rate had risen, thereby lowering the proportion of fatal accidents to total accidents.

Fatalities were fewer in the fall of 1956, and by the end of the year Connecticut could count 284 deaths in traffic as against 324 in 1955. The Governor stated, "With the saving of forty lives in 1955, a reduction of 12.3 per cent from the 1955 motor vehicle death toll, we can say the program is definitely worthwhile."

The crackdown on speeding is still in effect in Connecticut, although it is no longer the subject of newsworthy comment. It was not entirely a political asset for the Democratic Governor. From the start, there were problems with neighboring states, which originate a substantial share of Connecticut traffic, and which at first refused to suspend licenses of drivers convicted of speeding in Connecticut. More important, many powerful individuals and groups within Connecticut resented the direct effects of the crackdown. Members of the Republican Party wanted the program "tempered with justice." The Teamsters sponsored a bill to eliminate compulsory license suspension on a first offense, and other legislation granting restricted driving permits for "hardship" cases was introduced. These efforts were not successful in officially moderating the crackdown policy.

The people of Connecticut and their officials are paying what in many instances appears to be a high price for the continuation of the crackdown on speeding. Few will feel the price is too high if it can be shown that as many as forty lives per year are being saved. However, the question must be raised as to whether the results claimed for the program in 1956 are valid in the light of both formerly and more recently available statistics on highway fatalities.

QUASI-EXPERIMENTAL ANALYSIS

Before-and-After Measures

Traffic fatalities in Connecticut for 1956, compared with 1955, are presented in Figure 1. These are the data upon which Governor Ribicoff relied in claiming success for the crackdown on speeding. Skillfully presented, such results can look impressive, but can also be fundamentally misleading.

We can speak of the evidence presented in Figure 1 as a quasi-experiment: there is a "pretest" (the 1955 figures), an "experimental treatment" (the crackdown), and a "posttest" (the 1956 figures). A substantial change is noted which one would like to ascribe to the "experimental treatment." In quasi-experimental analysis this interpretation is held to be legitimate, provided consideration is given to plausible rival explanations of the differences, with supplementary analyses being added to eliminate these where possible. In the language of quasi-experimental analysis, the data of Figure 1 constitute a One-Group Pretest-Posttest Design. This design

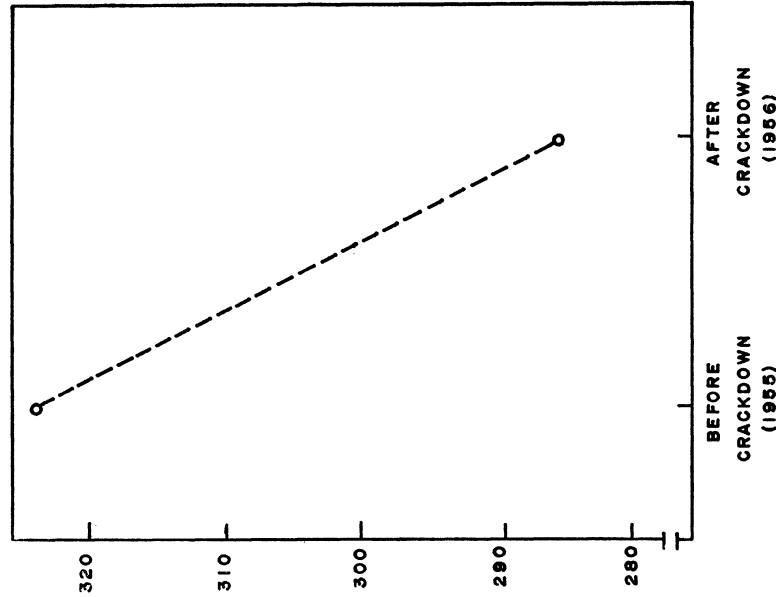


Figure 1. Connecticut Traffic Fatalities, 1955-1956

fails to control for the six common threats to the validity of experiments specified below:

1. *History.* This term denotes specific events, other than the experimental treatment, occurring between the pretest and posttest, which might account for the change. It furnishes a "rival hypothesis" to the experimental hypothesis, a competing explanation of the before-to-after change that must be eliminated as implausible, by one means or another, before full credence can be given to the experimental hypothesis. For instance, 1956 might have been a particularly dry year, with fewer accidents due

to rain and snow, or there might have been a dramatic improvement of the safety features on the 1956-model cars. In fact, neither of these is a particularly plausible rival hypothesis in this instance, and we have not encountered more likely ones, so this potential weakness may not be crucial here.

2. *Maturation.* This term originates in studies of individuals, where it refers to regular changes correlated with the passage of time, such as growing older, more tired, more sophisticated, etc. It is distinguished from history in referring to processes, rather than to discrete events. Thus, one could classify here the general long-term trend toward a reduction in automobile mileage death rates, presumably due to better roads, increased efficacy of medical care, etc. The better designs discussed below provide evidence concerning this trend in Connecticut in previous years, and in other states for the same year.

3. *Testing.* A change may occur as a result of the pretest, even without the experimental treatment. In the present instance, the assessment of the traffic death rate for 1955 constitutes the pretest. In this case it is conceivable that the measurement and publicizing of the traffic death rate for 1955 could change driver caution in 1956.

4. *Instrumentation.* This term refers to a shifting of the measuring instrument independent of any change in the phenomenon measured. In the use of public records for time-series data, a shift in the government agency recording the fatality statistics could account for such a shift. For example, suicide statistics increased a dramatic 20 per cent in Prussia between 1882 and 1883, when record keeping was transferred from the local police to the national civil service.² Similarly, Orlando Wilson's reforms of the police system in Chicago led to dramatic increases in rates for most crimes, due presumably to more complete reporting.³ In earlier versions of the present study, the death rate per hundred million vehicle miles is computed by using the number of gallons of gasoline sold in the state to estimate the number of miles driven. The latter figure is obtained by multiplying the former by an empirically-derived constant. A decrease in the actual miles obtained

2. Cited in C. SELTZ, M. JAHODA, M. DEUTSCH, & S. W. COOK, RESEARCH METHODS IN SOCIAL RELATIONS 323 (1959).

3. J. SWEEN & D. T. CAMPBELL, A Study of the Effect of Proximally Autocorrelated Error on Tests of Significance for the Interrupted Time Series Quasi-Experimental Design 31-32, Figs. 11 & 12 (mimeographed Research Report, Dept. of Psychology, Northwestern University, 1965). These figures also will appear in D. T. Campbell, Reforms as Experiments, AM. PSYCHOLOGIST (to be submitted).

per gallon, as through engines of larger horsepower or driving at higher speeds could masquerade as a lower mileage death rate through inflating the estimate of miles driven. Conversely, if the crackdown actually reduced driving speeds, this would increase the miles-per-gallon actually obtained, leading to an underestimate of mileage driven in the post crackdown period, and consequently an overestimate of the fatality rate.

5. *Instability.*⁴ A ubiquitous plausible rival hypothesis is that the change observed is due to the instability of the measures involved. Were Figure 1 to show fatality rates for a single township, with the same 12.3 per cent drop, we would be totally unimpressed, so unstable would we expect such rates to be. In general, as is made explicit in the models for tests of significance, the smaller the population base, the greater the instability. In the uncontrolled field situation sample size is only one of many sources of instability. Much instability may be due to large numbers of change-producing events of the type which, taken individually, we have called history.

6. *Regression.* Where a group has been selected for treatment just because of its extreme performance on the pretest, and if the pretest and posttest are imperfectly correlated, as they almost always are, it follows that on the average the posttest will be less extreme than the pretest. This regression is a tautological restatement of the imperfect correlation between pretest and posttest, as it relates to pretest scores selected for their extremity. The r of the correlation coefficient actually stands for the percentage of regression toward the mean. An analogous regression problem exists for time-series correlations.

Selection for extremity (and resultant retest regression) can be seen as plausibly operating here in two ways: (a) of all states in 1955, this treatment was most likely to be applied to one with an exceptionally high traffic casualty rate; (b) for Connecticut, the most likely time in which a crackdown would be applied would be following a year in which traffic fatalities were exceptionally high.

In the true experiment, the treatment is applied randomly, without relation to the prior state of the dependent variable: the correlation between pretest scores and exposure to treatment is zero. Likewise, in the

4. Instability has not been singled out as a specific threat to validity in previous discussions of quasi-experimental design, although the discussion of tests of significance in such situations has implied it. Tests of significance obviously do not provide "proof" relevant to the many other sources of invalidity, but they are relevant to this one plausible rival hypothesis even where randomization has not been used.

most interpretable of quasi-experiments, the treatment is applied without systematic relationship to the prior status of the group. Thus, an analysis of the effects of a tornado or an earthquake can be made with confidence that the pretreatment values did not cause the tornado or the earthquake. Not so here: the high 1955 rates can plausibly be argued to have caused the treatment. That 1956 was less extreme would then be expected because of regression.⁵

Interrupted Time-Series Analysis

Figure 2 plots traffic fatalities for five years before and four years after the crackdown. This mode of quasi-experimental analysis has been labeled "Interrupted Time-Series" to distinguish it from the time-series analysis of economics. In the latter, the exogenous variable to which cause is imputed is a continuously present variable, occurring in different degrees. In the Interrupted Time-Series, the "causal" variable is examined as an event or change occurring at a single time, specified independently of inspection of the data.

The Interrupted Time-Series design represents a use of the more extensive data which are often available even when only before-and-after measures are reported. Some potential outcomes of such a time-series analysis greatly reduce the plausibility of certain threats to validity. If the preexposure series shows but minor point-to-point fluctuations and no trend anticipating a big transtreatment shift, then maturation may not be plausible, for in most instances the plausible maturation hypothesis would have predicted shifts of the same order as the transtreatment shift in each of the pretreatment stages. Reasonable models of the testing effect would have the same implications. (In our instance, this would be on condition that the annual fatality rates had been given equal publicity.) The outcome in Figure 2 is not of this readily interpretable sort, although the trend is perhaps generally upward prior to the treatment, and steadily downward subsequently.

Judgments of the plausibility of instrumentation effects must be based upon other than time-series data. However, notice should be taken here of a frequent unfortunate confounding: the administrative reform which is meant to produce a social change very frequently is accompanied by

5. This issue is extremely complex. In ordinary correlation, the regression is technically toward the mean of the second variable, not to the mean of the selection variable, if these means differ. In time-series, the regression is toward the general trend-line, which may of course be upward or downward or unchanging. A more expanded analysis of the regression problem in correlation across persons is contained in Campbell & Clayton (1961) and in Campbell & Stanley (1963), both *supra* note 1.

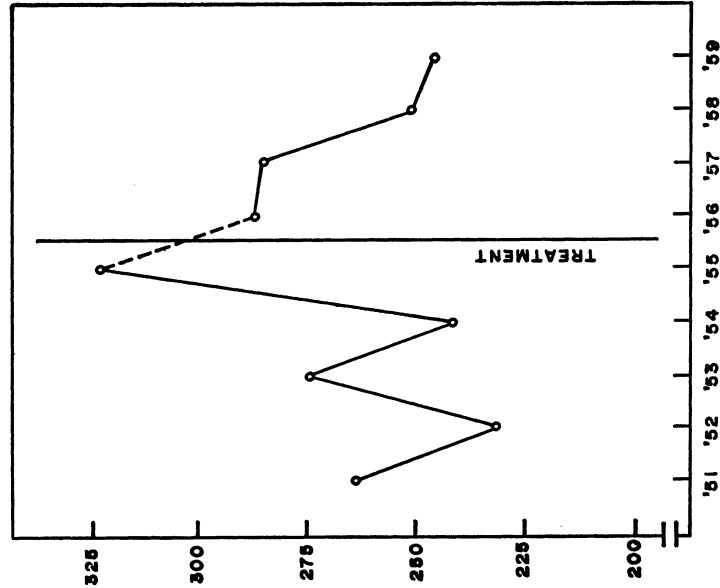


Figure 2. Connecticut Traffic Fatalities, 1951-1959

a coincident reform of the record keeping, ruling out valid inferences as to effects. The Chicago police reform cited above is a case in point. In the present instance, we have found no evidence of a change in record keeping or index computing of the type that would produce a pseudo-effect.

The likelihood of regression, or of selection for "treatment" on a basis tending to introduce regression, is supported by inspection of the time-series data. The largest change of any year is not the one after the crackdown, but is instead the upswing in the series occurring in 1954-55, just prior to the crackdown. In terms of crude fatality rates, 1955 is strikingly the highest point reached. It thus seems plausible

• 42 •

that the high figure of 1955 caused the crackdown, and hence it seems much less likely that the crackdown caused the low figure of 1956, for such a drop would have been predicted on regression grounds in any case.

The graphic presentation of the precrackdown years provides evidence of the general instability of the accidental death rate measure, against which the 1955-56 shift can be compared. This instability makes the "treatment effect" of Figure 1 now look more trivial. Had the drop following the treatment been the largest shift in the time series, the hypothesis of effect would have been much more plausible. Instead, shifts that large are relatively frequent. The 1955-56 drop is less than half the magnitude of the 1954-55 gain, and the 1953 gain also exceeds it. It is the largest drop of the series, but it exceeds the drops of 1952, 1954, and 1958 by trivial amounts. Thus the unexplained instabilities of the series are of such a magnitude as to make the 1955-56 drop understandable as more of the same. On the other hand, it is noteworthy that after the crackdown, there are no year-to-year gains, and in this respect, the character of the time-series has changed. The plausibility of the hypothesis that instability accounts for the effect can be judged by visual inspection of the graphed figures, or by qualitative discussion, but in addition it is this one threat to validity which can be evaluated by tests of significance. These will be discussed later, and they do find some evidence of change exceeding that which the pretreatment instability would lead one to expect.

Multiple Time-Series

In many situations, time-series involving but a single experimental unit will be all that are available. In these situations, analyses on the above model are a great improvement over the usual before-and-after study. However, it is in the spirit of quasi-experimental analysis to make use of *all* available data that could help to rule out or confirm any plausible rival hypothesis. In a setting such as this, no randomly assigned control group is available. But in quasi-experimentation, even a non-equivalent control group is helpful. It provides the only control for history (for those extraneous change agents that would be expected to affect both the experimental and control group), and assists in controlling maturation, testing, and instrumentation. For Connecticut, it was judged that a pool of adjacent and similar states—New York, New Jersey, Rhode Island and Massachusetts—provided a meaningful comparison. Figure 3 plots the death rates for the control states alongside

• 43 •

Connecticut, all data being expressed on a per 100,000 population base to bring the figures into proximity. The control data are much smoother, due to the much larger base, *i.e.*, the canceling out of chance deviations in the annual figures for particular states.

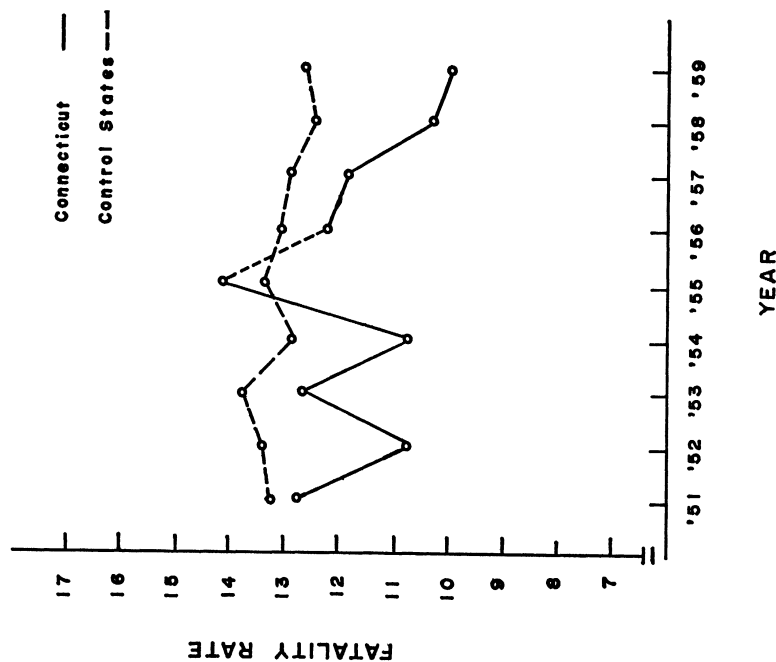


Figure 3. Connecticut and Control States Traffic Fatalities, 1951-1959
(per 100,000 population)

While in general these data confirm the single time-series analysis, the differences between Connecticut and the control states show a pattern supporting the hypothesis that the crackdown made a difference. In the pretest years, Connecticut's rate is parallel or rising relative to the con-

trol, exceeding it in 1955. In the posttest years, Connecticut's rate drops faster than does the control, steadily increasing the gap. While the regression argument applies to the high point of 1955 and to the subsequent departure in 1956, it does not plausibly explain the steadily increasing gap in 1957, 1958, and 1959.

Figure 4 shows the comparison states individually. Note that four of the five show an upward swing in 1955, Connecticut having the largest. Note that all five show a downward trend in 1956. Rhode Island is most similar to Connecticut in both the 1955 upswing and 1956

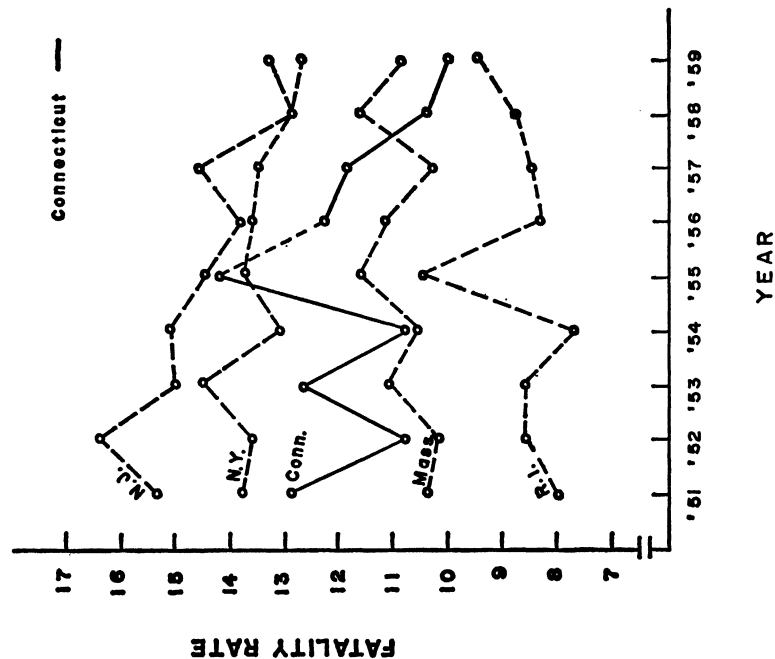


Figure 4. Traffic Fatalities for Connecticut, New York, New Jersey, Rhode Island, and Massachusetts (per 100,000 persons)

downswung, actually exceeding Connecticut in the latter—in a striking argument against the hypothesis of a crackdown effect. However, the trend in 1957, 1958, and 1959 is steadily upward in Rhode Island, steadily downward in Connecticut, supporting the concept of effect.

The list of plausible rival hypotheses should include factors disguising experimental effects as well as factors producing pseudo-effects. Thus, to the list should be added *diffusion*, the tendency for the experimental effect to modify not only the experimental group, but also the control group. Thus the crackdown on speeding in Connecticut might well have reduced traffic speed and fatalities in neighboring states. Dodd reports such an effect in his classic experiment on community hygiene in Syria.⁶ The comparison of posttreatment levels of Connecticut and the neighboring states might thus be invalid, or at least underestimate the effects. Conceivably one might for this reason prefer the single time-series analysis to the multiple time-series one. If highly similar remote states were available, these would make better controls, but for matters of either weather or culture adjacency and similarity are apt to be strongly associated.

Tests of Significance

Our position in regard to tests of significance is an intermediate one. On the one hand, we would agree that they are overly honored and are often useful in ruling out that one threat and should be used for that purpose. They are appropriate even where randomization has not been used because even there it is a relevant threat to validity to be able to argue that even had these data been assigned at random, differences this large would be frequent.⁷

The simplest tests conceptually are those testing for a difference in slope or intercept between pretreatment and posttreatment observations. As applied here these assume linearity and independence of error. It has been shown that the "proximally autocorrelated" error typical of natural situations (in which adjacent points in time share more error than non-adjacent ones) biases the usual tests in the direction of finding too many significant differences.⁸ Unaffected by this bias is a t-test by Mood which compares a single posttreatment point with a value ex-

6. S. C. Dodd, A CONTROLLED EXPERIMENT ON RURAL HYGIENE IN SYRIA (1934).

7. D. T. Campbell, *Quasi-Experimental Design* in 5 INT'L ENCYC. SOC. SCI. 259 (Sills ed. 1968).

8. J. Sween & D. T. Campbell, *supra* note 3. The tests thus biased include tests of slope and intercept provided by H. M. WALKER & J. LEV, STATISTICAL INFERENCE 390-95,

trapolated from the pretreatment series.⁹ None of these approached any interesting level of significance.

Glass¹⁰ has introduced into the social sciences a more sophisticated statistical approach, based upon the work of Box and Tiao.¹¹ This has the advantages of realistically assuming the interdependence of adjacent points and estimating a weighting parameter thereof, of avoiding the assumption of linearity (at least in a simple or direct manner), and of weighting more heavily the observations closer to the point of treatment. A number of assumptions about the nature of the data must be made, such as the absence of cycles, but these can be examined from the data. Applying this test to monthly data, he finds a drop in fatalities not quite reaching the $P < .10$ level of significance. Using a monthly difference between Connecticut's rate and that of the pool of the four control states, still less of a significant effect is found. In what he regards as the most powerful analysis available, he computes an effect parameter for each of the four comparison states and compares the effect parameter of Connecticut with this. Connecticut shows more effect, with a significance level somewhere between $P < .05$ and $P < .07$, with a one-tailed test. A more detailed description of the method and analysis of these data is given in Glass' article, immediately following in this issue of *Law & Society Review*.

Thus on the graphic evidence of steadily dropping fatality rates, and on these marginal statistical grounds, there may be an effect. This effect, it must be restated, could be due to the crackdown, or could be due to the regression effect. (Regression effects can of course produce "statistically significant" results.)

9. 399-400 (1953). Note that this invalidates the discussion of tests of significance in Campbell, *From Description to Experimentation*, *supra* note 1, at 220-30. The "Clayton test" presented there was found in the Monte Carlo simulation by Sween & Campbell to have additional errors leading it to be too optimistic.

10. A. M. MOOD, INTRODUCTION TO THE THEORY OF STATISTICS 297-98 (1950).

11. G. V. GLASS, *Analysis of Data on the Connecticut Speeding Crackdown as a Time-Series Quasi-Experiment*, 3 L. & SOC'Y REV. 55-76 (1968); T. O. MAGUIRE & G. V. GLASS, *A Program for the Analysis of Certain Time-Series Quasi-Experiments*, 27 EDUCATIONAL & PSYCHOLOGICAL MEASUREMENT 743-50 (1967); G. V. GLASS, G. C. TIAO, & T. O. MAGUIRE, *Analysis of Data on the 1900 Revision of German Divorce Laws as a Time-Series Quasi-Experiment*, 3 L. & SOC'Y REV. (1969) (in press).

12. G. E. P. BOX & G. C. TIAO, *A Change in Level of a Non-stationary Time Series*, 52 BIOMETRIKA 181-92 (1965); G. E. P. BOX, *Bayesian Approaches to Some Bothersome Problems in Data Analysis in Improving Experimental Design and Statistical Analysis* (J. C. Stanley ed. 1967).

Supplementary Analyses

In this section, we will present data that will further illustrate time-series analysis and, substantively, both indicate that the crackdown was put into effect and that it had some unanticipated and, to the policymakers, probably undesired consequences.

Figure 5 presents evidence that the crackdown was put into effect, as indicated by a great increase in suspensions of licenses for speeding. Unfortunately, we have not been able to get control state data for this and the following variables, but the single state time-series is quite convincing in itself. We regard it as confirming the appropriateness of the statistical tests that they indicate significant differences. The single-point-

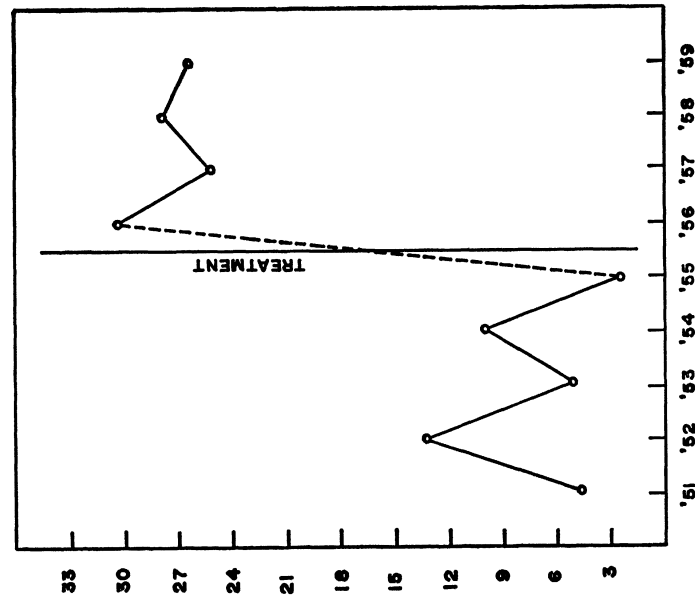


Figure 5. Suspensions of Licenses for Speeding, as a Per Cent of All Suspensions

. 48 .

extrapolation t is 4.33 with 4 degrees of freedom, where 3.75 is significant at the $P < .02$ level.

Figure 6 plots the percentage which speeding violations constitute of all traffic violations. This shows a decline, due presumably to greater conformity to speed limits, although it is possible that policemen and prosecutors were more willing, in the light of severe sanctions for speeding, to overlook minor infractions or to charge them as something else. While the graphic portrayal of declining speeding violations is convincing of a genuine effect, the statistical tests are not so emphatic. The single-point-extrapolation t is 2.66 with 4 degrees of freedom, not reaching the $P < .05$ level of 2.78.

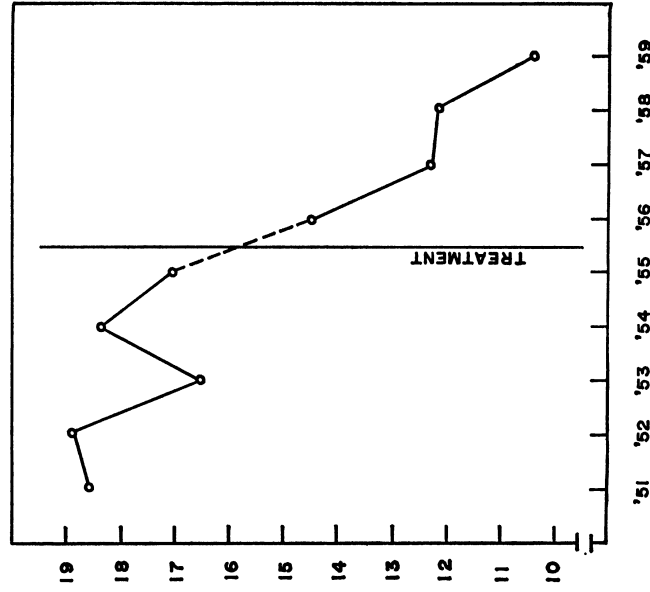


Figure 6. Speeding Violations, as Per Cent of All Traffic Violations

From Figure 5 and the reports cited in the first section of this paper it is clear that a real change in enforcement behavior resulted. It seems likely that the proportion of drivers exceeding the speed limits on Con-

. 49 .

necticut highways actually decreased. However, over and above these desired effects there are signs of unforeseen and unwanted reactions. Figure 7 concerns persons whose licenses were further suspended because they were convicted of driving with a suspended license, expressed as a percentage of all suspensions. This jumps from an almost consistent zero to some 4 to 6 per cent. Tests of significance confirm the effect. The single-point-extrapolation t reaches an incredible 130.75, due to the very small error term which the negligible variance of the pretest scores produces. (While one feels uneasy with a practically zero variance, the consistent pretest zero does genuinely make the later values unlikely.)

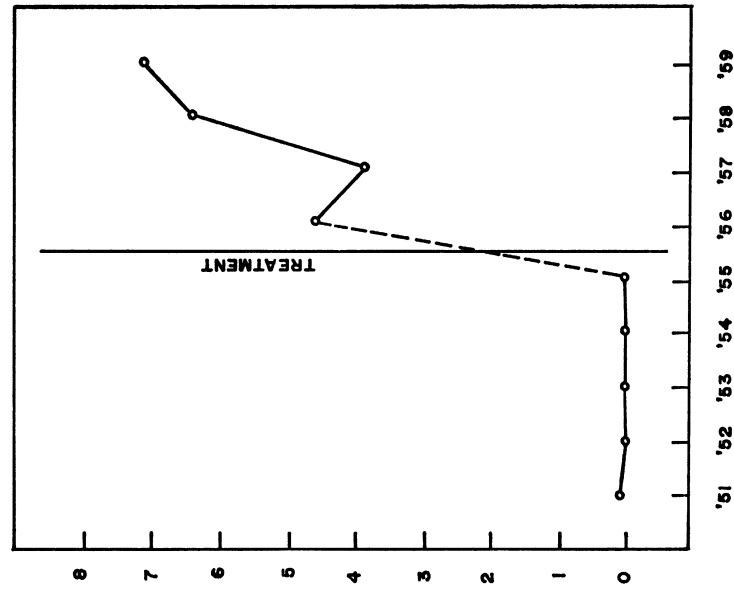


Figure 7. Arrested while Driving with a Suspended License, as Per Cent of Suspensions

• 50 •

Our interpretation of this phenomenon is that automobile transportation has become a virtual necessity for many residents of the diffusely settled megalopolitan region that includes Connecticut, and these people are willing to risk very severe sanctions in order to continue daily routines that involve driving. Since they are willing to drive with a suspended license, suspension does not have the desired restrictive effect on this group of drivers, which is probably much larger than the number apprehended and appearing in these statistics would indicate. Alternatively, of course, the increase could result, in whole or part, from more vigorous efforts at enforcement both in the crackdown itself and in special efforts at inspection comprising a followup of the crackdown effort.

Figure 8 shows a reaction on the part of the legal system. Even with fewer speeding violations reaching the courts (Figure 6), the courts were more lenient in their handling of these cases as expressed by the propor-

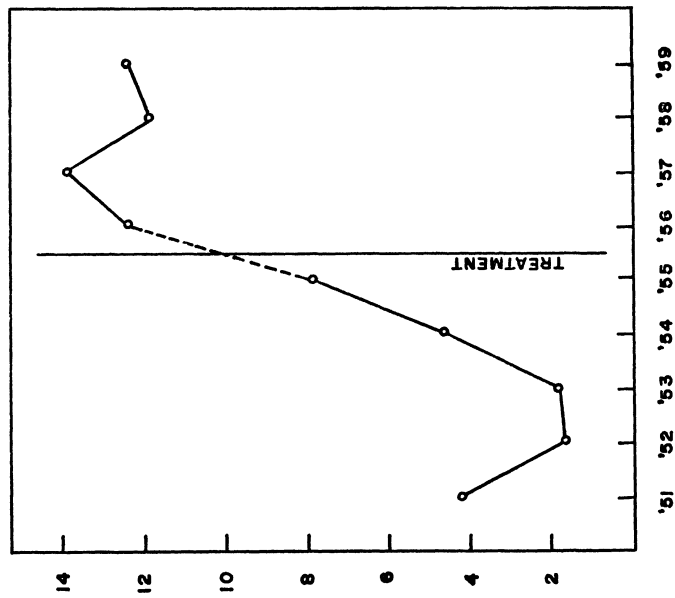


Figure 8. Per Cent of Speeding Violations Judged Not Guilty

• 51 •

tion of not guilty decisions. Tests of significance are borderline. The single-point-extrapolation t is 2.42, which with but 4 degrees of freedom fails to reach significance at the $P < .05$ level, for which 2.78 would be required. Larger proportions of not guilty judgments could be the result of more cases getting to court because of tightening of precourt standards, more generous handling by judges and prosecutors, or more vigorous defenses by the accused because more is at stake. The two effects shown in Figures 7 and 8 indicate a violation of the punitive effects of the crackdown in operation in a society where dependence on automobile transportation is acknowledged.

CONCLUSION

On the substantive side, the analysis has demonstrated that the Connecticut crackdown on speeding was a substantial enforcement effort, although some of its most punitive aspects were mitigated in practice. As to fatalities, we find a sustained trend toward reduction, but no unequivocal proof that they were due to the crackdown. The likelihood that the very high prior rate instigated the crackdown seriously complicates the inference.

We have, however, learned something about the response of the legal system to a reform bearing a harsh penal sanction. The courts, and probably also the police, are apparently unwilling to invoke penalties that might seem severe and unfamiliar in context. Moreover, the force of such penalties as are inflicted is vitiated by the willingness of the public to evade them. As in the case of white-collar crime, the effective punishment varies with the criminal.¹²

More important, we believe, than the specific findings of the study is the methodology here explored. While the social scientist cannot as a rule experiment on a societal scale, societal "experimentation" or abrupt focused social change is continually going on, initiated by government, business, natural forces, etc. The social scientist adds to his tools for understanding the social system when he attends to these events and documents their effects in as thorough a fashion as is possible. Insofar as correlational approaches differ from experimental analysis, it adds depth to the social scientist's work when he examines the fit of an experimental interpretation with full attention to the uncontrolled competing hypotheses.

12. The classic reference is E. H. SUTHERLAND, *WHITE COLLAR CRIME* (1959). See also H. L. ROSS, *Traffic Law Violation: A Folk Crime*, 8 *SOCIAL PROBLEMS* 231-41 (1961).

The methodology for such quasi-experimental analysis has a long but unsystematic history, and offers much room for development. It should be remembered that not only are the raw materials shaped by the tools, but in the long run the tools are shaped by the materials upon which they work. We should not passively accept a methodology as a revealed truth, but rather should test it in use with our materials. Methodology has in fact an empirical history and its constituents have the status of empirical discoveries. The classical control group experiment is not typical of the physical sciences, but instead emerged from psychological laboratory research, and is peculiar to the social sciences and their problems.¹³ Medical research has the placebo control group, and neurophysiology the sham operation control, as achievements of specific research traditions, not as logical dispensations from the philosophy of science or mathematical statistics. So too the methods for quasi-experimentation in settings like the present will emerge from an iteration of effort and criticism, in which many approaches will be rejected.

A final note on the treatment of uncontrolled variables is in order. On the one extreme there is that attitude often unwittingly inculcated in courses on experimental design, which looks askance at all efforts to make inferences where some variables have been left uncontrolled or where randomization has not taken place. In contrast, the quasi-experimental approach takes a radically different posture: any experiment is valid until proven invalid. The only invalidation comes from plausible rival explanations of the specific outcome. Regression effects and test-retest effects are such in many settings. An absence of randomization may in some specific way plausibly explain the obtained results. But unless one can specify such a hypothesis and the direction of its effects, it should not be regarded as invalidating. Subsequent consideration may uncover plausible rival hypotheses which have been overlooked, but such transitory validity is often the fate of laboratory experiments too.

At the other extreme is the naive attribution of cause which blithely fails to consider any explanations other than the author's favorite candidate. Such an orientation is likewise opposed. The quasi-experimentalist is obliged to search out and consider the available plausible rival hypotheses with all the vigilance at his command. While our coverage in this regard has been incomplete, we hope that we have at least illustrated such an approach.

13. E. C. BORING, *The Nature and History of Experimental Control*, 67 *AM. J. PSYCHOLOGY* 573-99 (1954).