STRATEGIES OF RESEARCH DESIGN IN THE LEGAL IMPACT STUDY

The Control of Plausible Rival Hypotheses

RICHARD LEMPERT
University of Michigan

A legal impact study represents an attempt to ascertain how a particular law affects the conduct and attitudes of those individuals, groups or other relevant units located in jurisdictions where that law is in force. By its very nature such a study involves one essential comparison; the comparison between actual behavior patterns in jurisdictions having the law in question and the behavior patterns which would have existed in those same jurisdictions had the law in question never been enacted. Since this comparison is one which by definition cannot actually be made, the problem for the legal impact theorist is how to estimate best what the behavior patterns

---

AUTHOR'S NOTE: This work is supported by a grant from the Cooperative Research Branch of the Social Security Administration and the Welfare Administration under SSA Grant No. 201, Department of Health, Education, and Welfare, Washington, D.C., and from funds of the Public Affairs Committee of Oberlin College under a Ford Foundation Grant. Research facilities were provided by the Social Science Research Institute, University of Hawaii. The study, *Legal Interventions, Social Mobility and Dependency — A Study of Public Assistance in Housing*, is being directed by Kiyoshi Ikeda, Oberlin College, and Douglas S. Yamamura, University of Hawaii. Mr. Lempert is a Research Associate and has worked as Legal Archivist and Interview Crew Chief.

The writer would like to express his special thanks to Professors Kiyoshi Ikeda, Donald T. Campbell and Harry V. Ball for the general encouragment they have given him and for their helpful criticism of earlier drafts of this paper.

would have been in a certain jurisdiction had the law in question never existed there. The legislator or court seeking to determine the actual or probable effects of a law faces a similar problem.

There are three ways in which this comparison can be achieved. One is by comparing the same jurisdiction before and after the passage of the law in question and noting any behavioral changes which seem to have followed as a result of the passage of the law. The second is by comparing jurisdictions which have a particular law with those that do not and assuming that, if not for the law, behavior in the two sets of jurisdictions would have been the same. The third method is by combining the two approaches. This involves examining behavior patterns in a particular set of jurisdictions both before and after they passed the law in question and comparing these patterns with those found over the same period of time in a set of jurisdictions not having the law in question.

Each of these basic ways of approaching the comparison problem can be broken down further into a variety of comparative approaches, each of which can be evaluated on the basis of how well it controls for "plausible rival hypotheses." A plausible rival hypothesis in this case is a non-experimental variable which could reasonably explain the behavioral change which the investigator would like to be able to attribute to the existence of the law. The problem of controlling for such plausible rival hypotheses is vital for effective legal impact research.

Perhaps we can better grasp the nature of this problem if we look first at a concrete example of legal impact research and at the efforts which have to be made to control for plausible rival hypotheses in the field situation. One study which serves our purposes well is the Kiyoshi Ikeda-Douglas Yamamura study of public housing on the island of Oahu, Hawaii.[1] This three-year investigation sought, as a major goal, to discover the effects which different types of housing regulations have on mobility patterns among housing project families.

Initial investigation clearly showed that mobility (as measured by income increases over time and moves from project tenancy directly into home ownership) was higher among families that lived in projects without income limits (Navy Lease) than it was among families which lived in projects having low (Federal)

1. For information pertaining to this study see: K. Ikeda, H. V. Ball & D. Yamamura, *Legal Interventions, Social Mobility, and Dependence: A Study of Public Assistance in Housing for Low Income Families,* paper presented at the 1964 American Sociological Association meeting; K. Ikeda, H. V. Ball, D. Yamamura & R. Lempert, *Regulatory Norms and Occupational Conduct Among Low Income Households: A Study of Public Assistance in Housing,* paper presented at the 1966 American Sociological Association meeting.

or moderate (State) income limits. While this finding was in line with theoretical considerations, a number of plausible rival hypotheses had to be dealt with before one could begin to determine exactly what effects could be traced to the impact of the different legal-administrative regulations.

For example, it was possible that the tenants of the Navy Lease projects had started from a higher income base than had the tenants of the State or Federal projects. To check this it was necessary to go to the records of the Hawaiian Housing Authority and examine the initial incomes of sample families applying for public housing on Oahu during the time periods under study.[2] Initial income for Navy Lease families was found to fall somewhere between the initial income of Federal housing families and that of State housing families.[3]

A second possibility was that families entered into the different projects at different points in their life cycle. If this were true it might be that the very young and very old entered the projects with income limits while those in the most mobile age brackets usually applied for Navy Lease housing. This was also checked by resort to the running record.[4] In general a higher percentage of the residents of the Navy Lease project were under thirty when they applied for public housing than were residents of the other two projects, but they were similar to the State housing families in number of children and the difference was not enough to explain the different outcomes.[5]

A third and more subtle threat to the hypothesis that the law had an effect in this area was the possibility that while the three housing groups were similar wih respect to easily operationalized characteristics such as initial income, they differed significantly on such hard to measure social-psychological characteristics as their general outlook on life, the relative weights which they gave to income, residential and familial values, and on the network of significant others to whom they could turn for moral and financial support. In order to properly assess the factors which came to play in this area it was decided to interview members of some 600 families.[6] To

2. The sample was drawn from two complete cohorts of public housing applicants, those who applied between 1953 and 1957 and those who applied between 1960 and 1964.

3. K. Ikeda, H. V. Ball & D. Yamamura, *supra.* note 1, at Table 2, Appendix.

4. For a general discussion of material which can be gleaned from the running record see: E. J. WEBB, D. T. CAMPBELL, R. D. SCHWARTZ, & L. SECHREST, UNOBTRUSIVE MEASURES: NONREACTIVE RESEARCH IN THE SOCIAL SCIENCES, (1966). See especially, chs. 3 & 4.

5. K. Ikeda, H. V. Ball, & D. Yamamura, *supra* note 1, at Tables 3-5, Appendix.

increase the validity of the study the sample of families interviewed included a number of families who applied for public housing and who met the requirements for the particular projects, but who were forced to turn to the private rental market because of a lack of space in public housing. The data from the interviews and from this latter group of families are only now being processed. But early results indicate that even after controlling for the above social-psychological factors a legal impact will be demonstrable.

This is not to say that the above sketch exhausts the number of plausible rival hypotheses which threaten the validity of the Ikeda-Yamamura study. Others exist and efforts have been made to control for them. It does give one an idea, however, of some of the major threats which can be encountered when one is doing actual legal impact work. The length and intensity of the study indicate one type of work which is needed in the area. But this does not mean that the only type of legal impact work which can be valuable must be as lengthly or extensive. Many valuable but simpler studies have been done with much smaller expenditures of both time and money.[7] Enough remains to be learned in the legal impact area that there is a need for both long term and short run research.

The basic plan of this paper is to consider certain pre-experimental and quasi-experimental designs which Professors Donald Campbell and Julian Stanley set forth in a chapter from the *Handbook of Research on Teaching* entitled "Experimental and Quasi-Experimental Designs for Research on Teaching" and see how these design considerations can be helpful when dealing with sociolegal problems of the type described above.[8] For the purpose of this paper we shall follow the numbering and the illustrative systems which Messrs. Campbell and Stanley set forth in their work.

---

6. Actually around 1,000 interviews were administered because in about 350 cases both the husband and wife were interviewed for each family.

7. For a simpler study using just interview data see: G. F. Break, *Income Taxes and Incentives to Work*, 57 AM. ECON. REV. 529 (1957). For a less complex study relying mainly on questionnaire data see: S. S. Nagel, *Testing the Effects of Excluding Illegally Seized Evidence*, 1965 WIS. L. REV. 283.

8. Quasi-experimental studies, as distinguished from experimental studies, occur when the experimenter is unable to achieve full control over relevant variables. D. T. Campbell & J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research on Teaching*, in HANDBOOK OF RESEARCH ON TEACHING 204 (N. L. Gage ed. 1963). Experimental research encompasses "that portion of research in which variables are manipulated and their effects upon other variables observed." *Id.* at 171. On experimentation, see generally A. KAPLAN, THE CONDUCT OF INQUIRY 126-70 (1964).

In this illustrative system, X equals the experimental variable which for purposes of this paper will always be the particular law whose impact is being studied; O delineates each point at which an observation of the relevant behavior is taken, and a dashed line (– – – –) is used to equate two observational units.[9] For the purpose of this paper, the observational units (jurisdictions) will always be states of the United States unless some other unit is specified.

We shall also adopt Campbell and Stanley's basic criterion for the internal validity of an experimental design, namely that a design is more valid internally the better it is able to control for plausible rival hypotheses. In other words the more certain an investigator can be that the experimental variable is causally associated with perceived changes, the more successful his experiment and the more valid the experimental design he has chosen to use.[10]

The strengths and weaknesses of experimental designs are usually discussed in the abstract, and to a large extent this procedure will be followed here. However, the relative power or strength of an experimental design is always related to the situation in which it is applied. The power or strength of a design is its ability to rule out the plausible rival hypotheses available. Thus a design which might have many potential weaknesses on paper may not have them when applied in a given legal impact study. Similarly a design with many potential weaknesses may be just slightly modified in a given legal impact study and many of these weaknesses will disappear. Finally, even when the weaknesses of a particular design cannot be adequately controlled for, it is often the best design available and it is often preferable that the investigator go ahead with an imperfect design rather than leave an interesting area completely unexplored. The essential requirement in these cases is to note explicitly the potential imperfections in any reports growing out of such studies.

GENERAL CONSIDERATIONS

Before going into the specifics of the various designs there are several general points which should be discussed. In the first place the designs as set forth with respect to impact studies might be rightly referred to by the purist as non-experimental (observational) designs rather than quasi-experimental ones. As Campbell and Stanley present their quasi-experimental designs they are quasi-experimental because of the investigator's inability to equate his experimental group with his control group through some sort of random selection technique. But

9. *Id.* at 176.

10. *Id.* at 175.

even in this situation it is still the investigator who introduces the experimental variable into one group and not the other. In many legal impact studies the investigator cannot be the experimenter; he is incapable of randomly assigning laws to different populations and is also incapable, for the most part, of playing any part in the introduction of the laws or in any way controlling the setting in which they are introduced.

This difference, while creating a different set of rival hypotheses which must be controlled, does not mean that the format of the quasi-experimental and pre-experimental designs[11] cannot be applied to impact studies. The fact that the investigator himself does not introduce the experimental variable himself should not force us to call legal impact studies non-experimental. The perspective toward comparisons is the same and the rigor with which the design can be followed can also be the same and should be if the "experimental" assessment is to have the highest possible validity.[12]

Campbell and Stanley in their chapter present eight factors which are potential threats to the internal validity of an experimental design because they provide possible sources of plausible rival hypotheses. They are: history,[13] maturation,[14] testing,[15] instrumentation,[16] regression,[17] selection,[18] mortality,[19] and the interaction of certain of these factors such as selection and maturation,[20] etc. Because of the quasi-experimental nature of legal impact studies and because of the particular nature of

11. Research which is scientific in its systematic approach to the data but which lacks the controls necessary for either quasi-experimental or experimental studies.

12. D. T. Campbell, *Factors Relevant to the Validity of Experiments in Social Settings*, 54 PSYCH. BULL. 297-312 (1957).

13. "[T]he specific events [other than the introduction of the law] between the first and second measurements." D. T. Campbell & J. C. Stanley, *supra* note 8, at 175.

14. "Processes within the respondents operating as a function of the passage of time" per se (not specific to particular events.) *Ibid.*

15. "The effects of taking a test upon the scores of the second testing." *Ibid.*

16. "Changes in the calibration of a measuring instrument or changes in the observers or scores used may produce changes in the obtained measurement." *Ibid.*

17. "[Operates] where groups have been selected on the basis of their extreme scores." *Ibid.*

18. "Biases resulting in differential selection of respondents for the comparison group." *Ibid.*

19. "Differential loss of respondents from the comparison groups." *Ibid.*

20. "[The danger arises] in certain of the multiple-group quasi-experimental designs . . ., [that] such an interaction effect might be mistaken for the effect of the experimental variable." *Ibid.*

the law, many of these threats may be disregarded by the experimenter in this area, while others pose special problems which would not be present in the normal social-psychological quasi-experimental situation. In discussing the possible influences of these sources of plausible rival hypotheses, and for the purposes of general discussion throughout this paper, we shall assume a standard experimental situation in which the experimenter is interested in assessing the effects of a certain law on the behavior of people living in states of the United States which have that law. In all cases the experimenter has available to him a sizable group of states which have different laws applying to the same subject and/or a group which is without any formal regulation of the subject.

Since administrative agencies often keep statistical records concerning the incidence of the behavior which they are supposed to regulate (*e.g.*, crime rates types of cases reaching court) the effects of a particular law often may be measured by commonly available statistical data. Testing effects should therefore be non-existent and instrumentation effects will often be minimal, though there are some threats to validity in this area. (1) First, an investigator may be attempting to compare two states which use different methods of computing their statistical reports on the same subject. This can be controlled by utilizing data collected by national organizations which prescribe uniform methods of data collection in all states. (2) The second problem is that a state or subdivision thereof may change its methods of reporting data. In Chicago, for example, when a new police superintendent took office he publicily announced that he was going to institute a more thorough system for reporting criminal activities and that the crime rate should be expected to increase considerably because of this. By doing this the superintendent was in effect preventing any real comparison of the work of the police department under his supervision with its work under his predecessor. This type of change, while probably infrequent, should always be looked for when one sees a marked change in the statistics he is gathering. This must be done even if the change in the statistics follows shortly after the introduction of the law under study, because a good possibility exists that the method of collecting data was changed in conjunction with the law.

Maturation, mortality and regression should also present minimal threats to the types of experimental designs which apply in this area, though there are situations when each of these features will have to be examined in detail before they can be disregarded as possible explanations of any perceived changes. For example, if a time series study were to be made of tort damage awards in two states having different legal norms, a more rapid increase in award size in one of them than in the other might be due to a difference in certain long range secular trends (maturation effects), such as inflation, rather than to the differences in the laws of the two states.

The major threats to the validity of legal impact designs are three: history, history-selection interaction; and, one peculiar to the sociology of law, the problem of distinguishing the law as it appears on the books from the "law in fact."

Of these, the determination of what the law is in fact should be preliminary to most studies. The law as it appears on the statute books may be only a partial and sometimes misleading guide as to the administered situation which in fact exists. If the law is within the province of a particular administrative agency, the agency's interpretation of the law and its activities in reference to the law must be traced. In many cases court decisions must be examined to see what "glosses" the judges have placed on the law. In still other cases there may be differential enforcement by the police in which case the system must be examined in detail at this level. In each of these cases, there are ways of getting at the character of the law as administered.

History is another potential threat to the internal validity of sociolegal experimental designs, but ordinarily it should not be too difficult to control for. It is an especially plausible rival hypothesis in the "before-and-after" type designs, such as Design 2 and Design 7 (discussed below). A coincidental historical happening can often explain a perceived behavior change just as well as the experimental variable can. This is why it is always wise to use "before-and-after" type designs in connection with a group of control states which does not have the particular law in question and which has been subject to the same historical influences as the group of affected states. If a geographically well-distributed group of states with a given law can be compared with a similar group of states without the law, history may usually be ruled out as an explanatory cause for the differences in perceived behavior which the experimental hypothesis would suggest is caused by the law under study.

In comparing laws among states one may often want to focus on the situation in the three or four years before their passage and the situation in the years immediately after their passage. Unless the date of the passage of the laws is the same, this procedure is almost always to be avoided. Any perceived differences may actually be due to history, but the design itself will have failed to correct for it, and one reading the write-up of the intervention will likewise be unable to correct for it.

The legal impact theorist should also guard against the possibility that the enactment of a particular law may be only one of several similarly directed govern-

https://doi.org/10.2307/3053049 Published online by Cambridge University Press

mental interventions in the same historical period. In Hawaii, for instance, the legislature recently passed a law setting up a special program to encourage home ownership among public housing families.[21] Along with this the local housing authority made an administrative decision to furnish all families participating in the program with certain special services and counseling. If the number of people moving from public housing into their own homes should significantly rise in the projects affected by the law but not rise in those unaffected (or in similar projects in other states not having the particular law), it might be hypothesized that the legal stimulus regarding home ownership had indeed caused the change. However, it is possible that it is the extra services and counseling which are responsible for any rise in home ownership and that the residential mobility rate of the affected housing families would have gone up without the special program if they had only been furnished with the extra counseling and services.

Such changes, stimulated by the law but not required by it or in fact a part of it, must always be looked to by the legal impact theorist as a source of plausible rival hypotheses. Until government officials become more specific in their interventions the most the legal sociologist can say will often be that a particular law, when coupled with certain other administrative actions, will lead to a certain behavior change.

This is not to say that legal impact research is unwarranted in the situation where the law has been embellished by administrative practice. For while it is impossible to say with certainty that the law produced a specific change, it is often possible to say with a high degree of certainty that a particular legal structure plus a particular series of administrative interventions resulted in a certain change in behavior patterns. Such results are quite satisfying to the practically oriented who are primarily interested in knowing what agent or agents can produce a certain change. What changes are induced by the law and what by administrative measures depends largely on how much direction the legislature chooses to write into its particular legal mandates. What is administrative policy in one state may be written law in another. Also, as different states adopt different variations of a similar legal-administered intervention package it may become possible for later investigators to separate out exactly what the most important change agents in the package are, be they legal or otherwise.

21. Hawaii Session Laws 1964, Act 22. For a theoretical discussion of this program and of general public housing interventions see: K. Ikeda, H. V. Ball & D. Yamamura, *Legal Interventions, supra.* note 1.

https://doi.org/10.2307/3053049 Published online by Cambridge University Press

Selection is the most severe threat which operates in a manner specific to socio-legal research. This is because a law represents two things, a regulatory device and (especially, but not exclusively, in a democratic political system) an expression of the people's feelings about a particular issue. This variable is especially likely to be a factor where a legal system is being studied in isolation or where a whole nation is being studied. For example, to the best of my knowledge, every state in this nation has a law against incest, and there is very little incest in this country. The mere concurrence of these two facts does not prove that the fact that incest is made criminal is responsible for its lack of prevalence. Indeed the opposite is probably true. The American people feel incest is such a bad thing that they have made a crime of it. However, America would not be an incestuous nation even if such laws did not exist.

Arnold Rose in his book *Theory and Method in the Social Sciences* points up the fact that the French legal structure has traditionally subjected France's voluntary associations to a variety of hindering regulations. He also points out that the average Frenchman has fewer ties with voluntary associations than does his American counterpart who lives in a society where the right of association is constitutionally protected.[22] Yet these factual congruencies do not necessarily mean that Franco-American differences in voluntary association participation can be explained on the basis of the differing legal structures. The plausible rival hypothesis exists that for specific cultural reasons French and Americans are differentially motivated to enter voluntary associations and that these differences have been reflected in legal patterns rather than caused by them. Rose reaches the only safe conclusion on the subject when he says in an article treating the same subject on the Italian scene that it is at least clear that the law has not helped increase participation.[23]

Any research design purporting to deal with the impact of a particular law on the behavior of a populace will have to make certain that the law is indeed more than an expression of the popular will of the people and that the people would be acting differently without the law. There are a variety of ways in which this hypothesis can be ruled out. Harry Ball used one of them in his study on rent control.[24] In polling landlords as to what they thought of the law he found that a large proportion of them thought that the law was unfair and (presumably) would have been charging higher rents if it were not on the books. He also found that a majority of

22. A. Rose, Theory and Method in the Social Science (1954),

23. A. Rose, *Individualism and Social Responsibility* 2 Eur. J. Soc. 163-69 (1961).

24. H.V. Ball, *Social Structure and Rent Control Violations*, 65 Am. J. Soc. 598-604 (1960).

those who disapproved of the law were obeying the rules. The techniques he used strongly suggest that rent control laws were a major influence on the level of rent in Hawaii during the period in which they were enforced.

The self-selection weakness also does not enter in too strongly when dealing with some of the more technical aspects of the law. For instance, the rules of evidence vary in certain particulars from state to state. If a study were to be made to see if this affected the direction of jury verdicts in certain types of cases, it would be incorrect to say that any difference that existed occurred because the people in the state, as represented *by the jury*, had wanted certain types of cases to go preponderantly in a given direction and so had passed the particular rule of evidence. It is possible of course that the laws were passed because a certain pressure group and/or a majority of state legislators wanted the rule, but since these people are not on the jury, an experimenter can still say, given the proper control conditions and the desired results, that the rule of evidence has influenced the direction of jury verdicts in particular types of cases, *e.g.*, a rule relating to the introduction of insurance liability figures in personal injury cases.

This selection weakness can also be controlled to a large extent by using a group of states spread out over the country in the experimental group, *i.e.*, having the law, and a similarly spread group in the control group. Given the mobility in this country and the broad common grounds between people in neighboring states, if people in 25 scattered states having a particular law act one way and people in the other 25 act another way with regard to the subject matter of the statute, one can usually be safe in saying that it is the law in its regulative aspect rather than in its expressive aspect which has caused the perceived behavioral difference.

The types of validity threats which affect legal impact experiments are all such as to make the study of only one state with a particular law represent a relatively poor experiment design. Studying one state with a particular law and one state without it is much better and can be quite impressive if a time-linked comparison (Design 14 below) shows a sharp effect. However, this design grows less effective the more dissimilar the states are in other explanatory variables.

The best design will always involve a time-linked comparison of a group of states having a particular law with a similar group of states not having that law. This design is more valid the more heterogeneous each set of states is within itself and the more similar the two sets of states when each set is viewed as a whole. For instance if Mississippi and Alabama have anti-miscegenation statutes coupled with an almost non-existent rate of interracial marriages while New York and New Jersey lack such

https://doi.org/10.2307/3053049 Published online by Cambridge University Press

statutes and have a much higher rate of interracial marriage, this factual congruence will provide little ground for an investigator saying that anti-miscegenation statutes cut down on the rate of interracial marriage. However, if Mississippi and New York were to have such statutes along with a significantly lower rate of interracial marriages than Alabama and New Jersey, then the investigator would have much more reason to believe that the laws had an impact in this area. The degree of certainty would increase if Alabama and New Jersey had once had such a law and if the rate of interracial marriages had shown a marked increase once the laws were repealed or conversely if Mississippi and New York had once lacked such a law and if the rate of interracial marriage had shown a marked decrease after the law was passed.

Strictly speaking no experimental design is valid outside the particular groups which were measured in the course of the experiment, and even for those groups the internal validity would be limited to the extent to which the experimental results are reproducible. But if experimental results could not be generalized, there would be little sense in doing any but the most widely encompassing experiments.

Fortunately the results of almost all experiments can be generalized to some extent to units not involved in the experimental situation. This generalizability is what Campbell and Stanley refer to as the external validity of the design.[25] By and large we can say that the greater the external validity of a particular experimental design, the more potential applications it will have and the more it will tell one about "real world" situations. It is in this area of external validity that legal impact designs have a great potential strength. When we study the impact of a law in several states we are examining what are essentially highly heterogeneous groupings involving millions of individuals. Thus we have a strong basis for generalizing the results of such a study to people in all states across the United States; far stronger, for instance, than we might have for generalizing a study of racial attitudes among students in a sociology course to students in the same university taking an advanced biology course.

THE PRE-EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS[26]

Design 1: X O — THE ONE-SHOT CASE STUDY

Design 1 to the legal sociologist is a simple descriptive design with observations taken at one point in time. Using a Design 1 approach simply involves noting that a

---

25. Campbell and Stanley refer to external validity as "representativeness." D. T. Campbell & J. C. Stanley, *supra* note 8, at 176.

26. Design numbers as employed by Campbell and Stanley have been retained. For a

society has a particular law (The X factor) and that certain conditions exist in the area which the law is designed to regulate (The O factor). This design is neither experimental nor quasi-experimental in that even if one were to know that a particular law existed in a certain society (*e.g.* a law against incest) and that a particular condition existed in that society (*e.g.* a lack of incestuous relationships) he would have no way of knowing whether and to what extent the law was responsible for the behavior observed. An observer from another planet during the prohibition era might have noted the amount of drinking which went on in this country and, if he were to draw conclusions from a Design 1 approach, assumed that for Americans a law against drinking encourages the consumption of alcoholic beverages. In fact alcohol consumption was probably much lower during the Prohibition period than before or after, if for no other reason than that it was less readily available. The incidence of arrests for public drunkenness certainly suggests that this was the case.

Once it is established that Design 1 is not *experimental,* it must be pointed out that in a certain sense it can be *scientific;* that is to the extent that the methods of determining what the law is and what the relevant conditions are follow a scientifically oriented procedure. It also should be pointed out that there is a real need for a taxonomy of diverse legal structures. Those following a simple Design 1 pattern of observation are performing a very valuable activity when they report their findings. Such findings, however, have the limitations which inevitably characterize the case study. They cannot, by themselves, provide the basis for testing generalizations.

Design 2: O X O — ONE-GROUP PRETEST POST-TEST DESIGN

This Design differs from Design 1 in that it calls for an observation of the behavior which the law purports to regulate both before and after the passage of the law in question. In the language of the social-psychologist, the law is the *experimental variable* and the two observations are a *pretest* and a *post-test.* For the social-psychologist the pretest and post-test are often administered questionnaires designed to measure operationally the behavior which the experimental variable is supposed to affect. For the legal impact theorists the pretest and post-test equivalents will ordinarily take the form of statistical reports of the behavior which the law is designed to regulate collected at two points in time (before and after the passage of the law under study). It is in this sense that the terms pretest and post-test will be used in this paper.

---

more detailed discussion of these designs and of statistical methods which can be used in connection with them, the reader is referred to the Campbell and Stanley article. D. T. Campbell & J. C. Stanley, *supra.* note 8, at 175-176.

https://doi.org/10.2307/3053049 Published online by Cambridge University Press

Design 2 is superior to Design 1 in some respects. It would, for example, have given the interplanetary observer mentioned above a more correct picture of the relationship between drinking and the law in the United States. Nevertheless Design 2 is relatively weak in controlling plausible rival hypotheses and should be avoided where possible. Its first weakness is in controlling for independent historical factors. Indeed it is weaker in this area than it would be when applied to the normal social-psychological situation because the likelihood of independent historical variables entering as causal factors in the perceived change increases as the time lapse between the measurements increases. In the case of social-psychological experimentation this time lapse may be kept minimal, but in the case of field experiments where officially collected data are relied upon there is often at least a year between available observational points. Indeed, if the law is to show any effect, an even longer time period between observations is sometimes needed.

Design 2 is also quite weak in the area of history-selection interaction. Using it does not help the investigator decide whether a particular law caused a perceived behavioral change or whether the change would have occurred anyway with the law being merely an expression of an intent that such a change should occur. For example there was a recent case in Hawaii in which a patient with a history of dangerous activities was released from a mental hospital 90 days after being admitted. He bought a gun and shot at cars passing along one of the state's principal highways, killing a policeman and wounding several tourists. It is quite probable that this incident will result in a change in the Hawaiian insanity statutes regarding criminal and civil commitments to asylums. It is predictable that if such a law were passed the pattern of releases from mental hospitals of those who were potentially dangerous to society would change. But even without the passage of such a law it is predictable that the sniping incident and its repercussions will lead to a change in the pattern of release of such dangerous persons. The Design 2 approach cannot determine whether the law being studied caused the change in behavior, whether an intent to change behavior (which would have been carried out, law or no law) was responsible for the passage of the law, or whether the factors of law and intent are related in such a way that both will explain part of any perceived behavior change.

Where several such incidents pile up there is the added possibility that this will lead to such drastic changes at the administrative level that the whole system of record keeping will be revamped or that classificatory standards will be revised. In either case the possibility exists that such administrative changes will lead to pseudo-effects not attributable to the law in question in that they will disguise changes which the law has brought about. The sniping instance discussed above might lead to a different method of classifying people as dangerously insane which would lead to a different pattern of release among such individuals whether or not the legislature adopts stricter standards governing their release and whether or not the hospital itself imposes more stringent procedures.

Another reason why Design 2 should be avoided is that the use of statistics from only two points in time can be very misleading. Two arbitrarily chosen points in time may not be typical periods at which to measure the incidence of that behavior which is supposed to be affected by the law, and the investigator may well find a seeming change which is actually the result of regression artifacts, *i.e.*, a stochastic probability that extreme phenomena will appear less extreme on remeasurement. For instance, in Connecticut there was a tightening of the driving regulations after a year with an especially high death toll and the next year the death toll fell markedly.[27] The state officials praised the new law system and gave it much credit for the drop in deaths which a statistician could have predicted would occur (given the figures for auto deaths in the ten years previous) on the basis of regression artifacts alone.

This type of design is especially dangerous in the sociology of law for two reasons. One is that because of the responsive-expressive characteristics of the law there is a higher than chance probability that if the year before the passage of the law is chosen as one of the two points, it will have been a year of exceptionally high incidence of the type of the behavior which the law is trying to regulate, since laws often gain their necessary momentum for passage in response to just such out of the ordinary situations.

The second reason is that the Design 2 approach tends to be very misleading to the public at large, yet such studies often are publicized widely. Witness the use which politicians make of such "designs." I do not think it is pessimistic to say that the people of Connecticut are much more likely to listen to Ribicoff's "hard facts" concerning the drop in deaths in Connecticut and to attribute this difference to "his" speeding law than they are to listen to the scholarly talk of some social scientists who claim that when the whole time-series is examined the decrease in lives lost could be explained on the basis of statistical probabilities. Social scientists, if they cannot correct such political arguments, should at least try to keep their work from adding to the misinformation which the general public may receive.

27. D. T. CAMPBELL & ROSS, USE OF TIME SERIES IN EVALUATING SOCIAL CHANGE (to be published).

https://doi.org/10.2307/3053049 Published online by Cambridge University Press

Design 3: X O STATIC-GROUP COMPARISON

X O
- - - - - - - -
   O

Design 3 has been the traditional design of the great generalizers and the creators of much of the early sociology of law theory.[28] Their technique has generally been to note that certain societies have a particular law and that certain conditions exist in that society while other societies lacking that law have comparatively different situations. Usually their conclusions have run in the opposite direction from the legal impact theorists and they have tried to equate certain societal changes (as independent variables) with the legal structure of the nation (as dependent variables) rather than vice versa. Such comparisons when operating in this direction have been fruitful in stimulating the generation of the more abstract legal theory. But they are basically methods of thought rather than a systematic research design.

When it comes to actually testing hypotheses, however, and particularly when one is trying to test the hypothesis that a law has a certain influence, Design 3 has certain grave weaknesses. The gravest is the self-selection weakness. If one state has a particular law and another lacks it, even if observed conditions in the two states are different, it cannot be stated authoritatively that the law has caused the perceived difference until all the other features which the two states do not share and which could have caused the difference have been examined. This includes the possibility that because of a different history, location, or other factors, the people of the state having the legal variable would have acted in the direction that the law requires if the law had not been passed in the first place.

It seems Design 3 can approach the true experimental design of Design 6 the more numerous and the more heterogeneous a group of units that can be put in each slot. If 25 different states all have a particular law and show a certain type of behavior while 25 states without the law possess mutually similar behavior patterns which differ from those of the first set in a systematic way, then the hypothesis that the law is responsible for a particular pattern of behavior becomes much more tenable.

There remains however a major threat to validity resulting from the lack of a

28. See, *e.g.*, E. Durkheim, *Deux Lois de L'Evolution Penale*, 4 L'ANNEE SOCIOLOGIQUE 65 (1899-1900).

non-reactive pretest measurement.[29] For it is possible that the two groups of states differed along similar dimensions before the law was adopted by one set of them. The law might be an expression of a belief which for a certain set of reasons is peculiar to 25 states of the United States and not to the other 25, or it might be symptomatic of a condition which prevails in the states having the law but is not an important factor in those states without the law. In either case the Design 3 approach, even with a variety of states in each block, would yield only a reflection of differential conditions which prevailed in the states in question before the passage of the particular law rather than the effect brought about by the law. Often the investigator can ascertain if this factor is confounding his results by examining the states in question in some detail. As a rule, the more heterogeneous the "treated" and "non-treated" states, the more likely it is that possible causal beliefs and conditions will be randomly distributed throughout both groups thereby increasing the probability that Design 3 will prove to be a satisfactory design.

Design 7: O O OXO O O TIME SERIES

The investigator who is using the Design 7 approach takes a series of behavioral observations at points in time before and after the enactment of the law being studied. If the behavioral curve shows a sharp change in the predicted direction during the period after the passage of the law, it is hypothesized that the law caused the particular change. As Campbell and Stanley demonstrate, Design 7 is considerably better than Design 2 since it controls for maturation, regression and certain selection and interaction effects. If, for example, the Connecticut traffic death statistics discussed above had been plotted over time, the drop in deaths following the tightening of the laws would have been shown to be what it was – a pseudo-effect predictable from statistical regression.

Despite the fact that Design 7 is a great improvement over Design 2 and is often well worth using, it is still subject to a serious defect – the lack of a control population. The validity of a Design 7 approach may be limited because it does not control for independent historical variables, for history-selection interaction effects which are the results of the expressive function of the law, and for the fact that certain non-legal variables are often associated with the onset of a law. These, rather than the law itself, may be the primary explanatory variables.

29. A non-reactive pretest occurs when the pretest does not affect the subjects under study, thereby avoiding an effect on subsequent observations.

As an example of how these three effects may jeopardize the validity of a Design 7 type impact study, we shall give a brief description of a hypothetical study of this type and indicate the kinds of factors which would not be controlled within this design. Let us take the law passed by the Hawaiian legislature designed to encourage public housing families to move into their own homes by giving them a special subsidy if they wish to save for their own homes while in public housing. Let us suppose that our observations show that two in every hundred public housing families move into their own homes each year before the passage of the law and that this number goes up to five in a hundred during the several years following the passage of the law. The hypothesis is that housing families have been influenced by the law to increase their rate of home ownership.

Factors which threaten this hypothesis and which would not show up in the Design 7 approach include: (1) Independent Historical Variables, *e.g.*, the price of housing in Hawaii might have gone down about the same time that the law was passed; (2) History-Selection Interaction, *e.g.*, it might be that legislature enacted the program largely because of active lobbying by a group of tenants who are planning to buy their own homes in any case and wanted to see if something could be done to make things easier for them; (3) Influence of Non-Legal Variables, *e.g.*, it might be that a group of special services not required by the law are being administratively provided for these families and are responsible for the rise in home ownership.

Design 10: O X O NON-EQUIVALENT CONTROL GROUP DESIGN

```
O   X   O
-----------
O       O
```

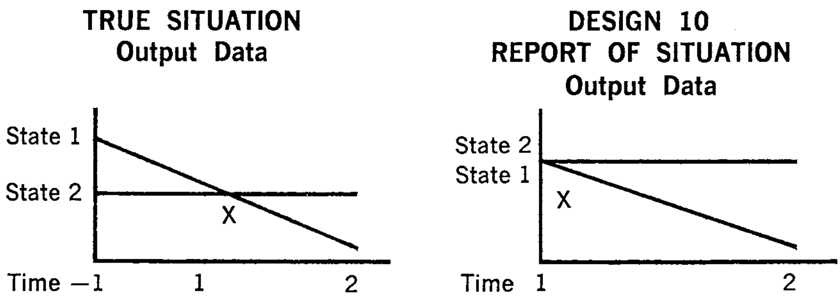
Design 10, the non-equivalent control group design, is one which, while often the best available for quasi-experimentation in a social-psychological setting, should not have to be used with great frequency by the legal impact theorist. This design calls for a pretest and a post-test measurement from both the experimental and control populations, and differs from a true experimental design only insofar as membership in the two groups is not determined by a random selection procedure. In the type of problems the legal impact theorist will deal with, the pretest and post-test scores will usually be in the form of statistical data. Where such data are available, they usually exist for and should be used for a period of years preceeding the introduction of the experimental variable as well as for a period after the introduction of that variable.

When a Design 10 approach is used in a social-psychological setting the correct procedure is to live with any differences that come to light on the pretest. For example, if an investigator has two classrooms available and can introduce the experimental variable into one of them, he should not be deterred by the fact that the two groups have different mean scores on the pretest. Most importantly, he should not take a subset of the one class with the same mean as the entire set of the other class and use it as his control population. As Campbell has convincingly shown, differences which arise between two such matched groups can often be explained by regression artifacts.[30]

Regression can similarly confound the results of a Design 10 legal impact study. The danger is especially great in such studies because the investigator is likely to choose as a comparable state one which is similar to the experimental state at the before point in time in order to make a "better" comparison. Yet there is a definite possibility that the two states chosen are only coincidentally identical in their behavior patterns in the given year and that differential changes in behavior could be expected merely on the basis of differential behavior patterns in the two states over an extended period of time.

Aside from the regression situation in which the behavioral data from one or both states is taken in a year which for that state varies greatly from a relatively stable mean level of behavior, there is also a danger that the study may be confounded by trends which are disguised in the single pretest, single post-test comparison approach.

1 = Pretest Measurement
2 = Post-test Measurement
X = Point of Introduction of Experimental Variable.
No measurement taken at this time.



30. D. T. Campbell & K. N. Clayton, *Avoiding Regression Effects in Panel Studies of Communication Impact*, 3 STUD. PUB. COMMUN. 99-118 (1961).

If, for instance, there is a situation like that graphed above, it is obvious that we have no grounds for concluding that the law has created the desired change in the behavior in State 1 and that the addition of State 2 as a control has only served to make the investigator more sure of a false conclusion.

In using Design 10, the results are not clearly interpretable because merely living with differences does not solve the problem of comparing two states. Any single year which is chosen by the investigator may be an unusual year in certain respects for a particular state and for that state a subsequent shift in behavioral patterns away from the extreme case may be seen as regression. Similarly, trend patterns may come into play when data from two states are being analyzed even if the states were chosen randomly rather than because of an initial equality in a certain respect.

The flaws in Design 10 are corrected to some extent (and this is one reason why it is a relatively powerful design in the social-psychological setting) as more units figure into the analysis. In other words, if the mean of the data from 25 states which had a certain law were to be compared with the mean of similar data from 25 states without the law, then a researcher would be in a situation to live with the differences and perhaps interpret them meaningfully. If the original intergroup difference is not great then the multiple unit variation of Design 10 may be a very successful one. However, the wider the original differences, the more likely it would be that some factor other than the laws were responsible for the subsequent differences in behavior. This is especially so because of the selection-history interaction process which has often been referred to in this paper.

Design 14: O O O OXO O O O   MULTIPLE TIME SERIES DESIGN
- - - - - - - - - - - - - - - - -
O O O O O O O O O

The best way to correct for the flaws of Design 10 is to use Design 14, the multiple time series design. This design is the design *par excellence* for impact theory experimentation, the comparison becoming more valid the larger the number of units which may be fitted into the experimental and control groups and the greater the similarity of the two time series measurements during the period prior to the introduction of the experimental variable. For example, if a group of states with a certain law and a group of states without that particular law can be compared over time, and if the behavior patterns are similar until the law is introduced in the experimental states, an investigator can attribute subsequent changes over time to the law with a high degree of certainty. It is obvious that this design controls for regression effects, trend effects, and for certain types of history. If only two states are being compared there is still the chance that the history selection interaction has been the cause of any resultant change, but is very unlikely that people of two states will be alike for an extended period of years and then undergo a sudden differentiation in their population's activity-linked attitudinal characteristics. If such differentiated behavior is perceived it is much more likely that the law has caused it than that an underlying attitude difference suddenly developed between the people of the two states. This rival hypothesis becomes less and less plausible with the inclusion of more states which are heterogeneous.

The rival hypothesis, that the extra-legal changes are causing the difference in behavior patterns, is still one which must be checked in this design. However, it too becomes less plausible as more and more states fall into the experimental and control categories, since even when laws are similar on the books, extra-legal patterns tend to be differentiated. If investigation shows that such extra-legal patterns differ in a relatively random fashion among the states in the sample, but that behavior patterns between the experimental and control states are systematically differentiated, then the investigator may again state with a high degree of certainty that the law is causing the observed differences. This design rates very high on the internal validity criteria.

The beauty of Design 14 is not only that its internal validity is inherently higher than that of the other quasi-experimental designs but that it is especially well adapted for legal impact studies. The United States federal system of government has furnished the experimenter with 50 states which, in the words of the Supreme Court, are "natural laboratories." Much of the behavioral output of these laboratories with respect to particular laws is reported in relatively available and uniform statistical form. Where such information exists it is usually available over an extended period of time. Legal sociologists can take advantage of these natural laboratories and using the Design 14 approach start intensive investigations of exactly how certain laws influence specific behavioral patterns. Once there is more information as to how this occurs there will be a more reasoned basis for determining "why" the patterns occur as they do.

While the state laboratory approach has been used in most of the examples in

---

31. For a discussion of what factors must be considered in conjunction with the formal rule structures and of possible approaches to the study of formal organizations, see P. Blau, *Formal Organization: Dimensions of Analysis*, 63 AM. J. SOC. 58 (1957).

this paper, this is not the only situation in which the impact theorist can work. Almost all politcal subdivisions, from towns through nations, have certain areas of legal variation and other areas of legal symmetry. Certain voluntary organizations have rule structures which lend themselves to the types of experimental study suggested in this paper. As a general rule comparative studies of organizations which rely upon formal enacted rule-making procedures can also be researched within this frame.[31]

CONCLUSION

In this paper an attempt has been made to indicate the manner in which some of the experimental designs described by Campbell and Stanley can be applied to the problems which confront the sociologist who wishes to study the effects of laws on behavior. We have tried to demonstrate how legal impact studies confront certain types of plausible rival hypotheses which are peculiar to them and/or are especially difficult to control for. To the extent that the Design 14 approach appears applicable it should be used since it best disposes of these alternative explanations. If Design 14 cannot be used it is almost always better to use an inferior design than to do no study at all. The important thing is that the experimenter be aware of the potential weaknesses which exist in his materials and examine the empirical situation closely to see if any such weaknesses in design may be confounding his results.

It should be emphasized that this paper has been dealing with only one area of sociology of law research. In other areas, especially where one is trying to discover how extra-legal factors affect legal relationships, the experimenter may be able selectively to control the administration of the experimental variable. Often he will be able to randomize the administration of the experimental item and so will be in a position to use a true experimental design. In particular we think that there is a great deal of room to use true experimentation in any area which might be best called the social-psychology of law, *i.e.*, working with individuals and small groups to see how people react in various law-related situations. Much of the experimentation at this level, such as the Chicago jury study, will not be of a legal impact nature, but it is possible that small group experiments could be set up to test impact-oriented hypotheses. If this were done many fruitful hypotheses could be generated. Such work would often involve true experimental designs with high internal validity. However the external validity of any observed results would always be open to question. Here again the "real world" impact study will become a necessity and here again the designs discussed above can be put to use.