4

Longitudinal Research in Program Evaluation

Greg Mason

In program evaluation, causal attribution (linking cause and effect) is an essential step in validating the relationship between program implementation and outcomes. Longitudinal studies are believed to provide a superior method of inferring causality, but the term "longitudinal research", spans a broad range of meaning. Some program evaluations use small scale follow-up studies of service recipients to decide whether specific interventions have had the intended effect. At the other end of the spectrum social policy experiments have used massive panel studies and experimental methods to evaluate the impacts of policy.

This paper will focus on the smaller scale study since this is the likely application for most evaluators. The large scale study provides a useful context for examining longitudinal research in general and will be used to illustrate some theoretical points. The paper is grounded in basic theory, but practical logistical issues are presented as well.

The first section presents the general problem of assessing causality in the context of a linear regression model. There are many approaches to analyzing longitudinal data and the regression model provides a convenient method for examining the potential for this approach. The second section reviews the specific theoretical benefits of a longitudinal design and considers how a regression model may be used to examine effects over time. Logistical considerations of how to maintain the integrity in a longitudinal study are presented in section three, and the conclusion appears in section four.

ATTRIBUTING CAUSAL RELATIONSHIPS

The form of the data sets the first limits to inference on causal relations. The set limitation is imposed by the model and the extent to which assumptions are explicit (and tested) or implicit (and untested). It is possible to identify a number data of formats within quantitative evaluation research.

TYPOLOGY OF DATA FORMATS

Each data format reflects a different ability to support causal inference.

1. **Cross-Section** data refer to quantitative data, often collected through a standardized interview, at a single point in time, on discrete units of analysis, such as individuals, households, and countries. An example is a public opinion poll.

2. **Time-series** data refer to information collected over time on the units of analysis. Unemployment rates in Canada for the years 1970 - 1990 are time series data.

3. **Cohort data** are collected over time by interviewing people in the group at specific points (e.g., a survey of "baby boomers" every 5 years). An example is a health survey which follows randomly selected respondents every five years by shifting the qualifying age up five years with each successive survey. This type of data are common to education research and demography.

4. **Panel studies** collect information over time from exactly the same respondents. Health studies have used this approach.

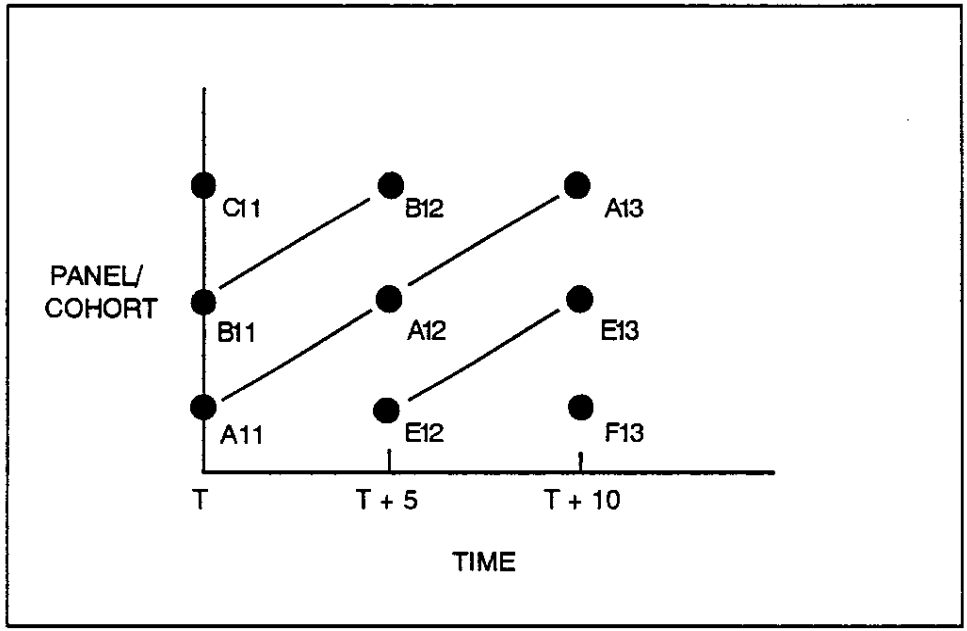
5. **Experimental** data are developed by randomly allocating respondents to a control and treatment group(s). The random allocation is designed to control for all influences except the treatment. Medical experimentation is typical of this type of research.

6. **Quasi-experimental** data involves a demarcation within cross-section or time-series data between those who were "treated" and those who were not. Multivariate analysis is used to control for potential influences beyond the particular intervention under study. Program evaluations often adopt this approach.

These data types often overlap. The large scale social policy experiments, popular in the sixties and seventies, combined an experimental and panel approach. Many quantitative program evaluations use quasi-experimental approaches with cross-sectional or time-series data. Dummy variables in a regression model are used to test whether the treatment has made a statistically significant difference in the outcome measure(s).

It is useful to visualize the types of data encountered in the social sciences. Consider Figure 1 below. An initial cross-sectional survey is specified as A1. This may

Figure 2.
Parallel Cohort Designs



as A2', A3', and so on. For example, sampling 10 year old children at time = 1, 2, and 3 is known as pooled, cross-section, time-series data and is common in economic research. One of the weaknesses of a cohort study, or any longitudinal research, is that time must pass before "results" are available. This is unavoidable in research which examines human development or disease processes which may take decades to emerge. For policy studies this time frame is a problem. One approach to produce quick research results is to do cohorts which run parallel. This is shown in Figure 2.

In this study, it is possible to undertake comparisons between two groups, say A11 and B11 which would compare 10 and 15 year old children at a given point in time or new home owners at different stages in their life-cycle (say 5 years apart in the age of heads). This type of paired cross-section provides some insight into causal processes, under the assumption that differences between the two groups are a function of intrinsic processes (life-cycle consumption) and not extrinsic forces (interest rates).

If exactly the same respondents are resurveyed, a "panel" design results. The terms "longitudinal", "panel" and "follow-up" are usually interchangeable. This re-contact of identical respondents provides a stronger basis for inferring causal

Table 2. reveals that of the 40 who stated they would support the REDS six months ago (T=0), 31 would continue to do so, 5 have gone to the BLUES and 4 to the ORANGES. Of the 23 who said they would vote BLUE six months ago, 11 would continue to do so, 3 have gone to the ORANGES and 9 to the REDS. By linking data with the same respondent through time, the development of voting intentions may be more clearly discerned.

A "recall" type question on a cross-section survey adds a time dimension which dramatically improves the insight into the evolution of and development of voting intentions. But, it is well known that people may misrepresent how they previously felt, especially if they supported a currently unpopular government. Were Table 2 derived from panel data with no recall, and respondents stated who they currently support, this distortion would be controlled. Using recall questions on cross-sectional data collection may be termed the "pseudo-panel" approach.

By attaching the same individual to information provided at two points in time, the understanding of a process is greatly enhanced. Any phenomenon which involves development and a process over time, is often enhanced with a panel design. Of course, there are other problems in panel research which are reviewed below.

CAUSAL ANALYSIS

The following conditions are usually taken to be necessary (but not sufficient) evidence of a cause and effect relationship between Y (effect) and X (cause).

- Y and X must co-vary; a change in X tends to be associated with a change in Y, and vice-versa.
- the cause X, precedes Y. (temporal precedence)
- there is no "third" factor Z, which determines both x and y simultaneously.

These rules refer to tendencies. Because of measurement error, the counter-effects of other forces, and the omission of potential influences, the link between X and Y may not be consistently observed. While correlation and temporal precedence alone do not prove causation, they are stronger evidence than simple descriptions.

The ability to infer cause and effect depends on the number of untested assumptions which must be maintained by the researcher to support a statistical model. Models which have fewer assumptions have greater power in identifying probable cause and effect, yet simple models may not express program interventions. No model, even the randomized experiment, is completely free of these assumptions.

An example of a social experiment is the Manitoba Basic Annual Income Experiment (MINCOME). Low income households were randomly assigned to treatment and control groups. The treatment groups were simply monitored as to job search and education, while the control groups were paid varying levels of a guaranteed annual income. The main objective was to evaluate the impact of guaranteed annual incomes on willingness to work. Social research can rarely test causal relations in the same way as laboratory experiments with extensive controls isolates treatments and effects. For quantitative analysis in program evaluation, the most common procedure is to create cross-sectional, quasi-experimental models supported by implicit and explicit assumptions reflecting a set of hypotheses about the interventions. The statistical model evaluates the plausibility of testing hypotheses on the coefficients. For example, the standard regression model:

$$Y = a_0 + a_1X_1 + a_2X_2 .. + a_kX_k + u$$

reflects a theory with the following specific assumptions:

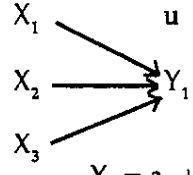
- the influences on Y are linear and additive
- these influences are comprised only of the variables X_1 - X_2 (no omitted influences on Y)
- the independent variables (X_1-X_2) are measured without error)
- the error term (u) captures all other influences in Y.

A number of specific assumptions are also made about the error term (namely that it has constant variance, a mean of 0 and that successive errors are not correlated). If x_1 is a dummy variable encoded with 0 for pre-program and 1 for post-program, then a test of the coefficient a_1 determines whether or not the intervention has had an effect.

This multivariate regression test of causality is inconclusive without additional "structure." Critics argue that the multiple regression model is little more than an elaboration of correlation. Further, there are so many assumptions which must be accepted to make the model testable, it is meaningless to call it a "test." The whole point of longitudinal analysis is to add the needed structure by *time-dating* information. This allows more complex hypotheses to be tested.

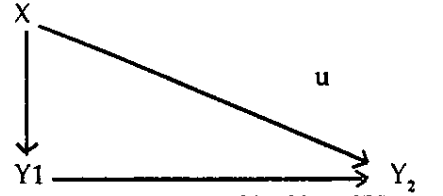
Figure 3
Cross-Section, Quasi-Experimental and Panel Designs

Period 1 Period 2 Period 3



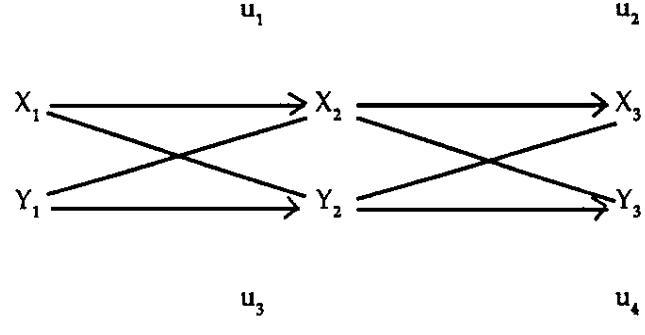
$$Y_1 = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + u = F(X_i)+u$$

Cross-Section



$$Y_1 - Y_2 = f(X) + u$$

Quasi-Experiment



$$Y2 = a_1X_1 + b_1Y_1 + u_3; \quad X2 = c_1X_1 + d_1Y_1 + u_1$$
$$Y3 = a_2X_2 + b_2Y_2 + u_4; \quad X3 = c_2X_2 + d_2Y_2 + u_2$$

Panel Design

in time. For example, an anti-poverty program may have a component which instructs social assistance recipients on the use of budgets. The use of a panel study where respondents are asked whether they are keeping a budget may result in respondents reporting they do keep a budget (some may actually start as a result of the question and not the program). The use of the same respondents produces changes which are due to the evaluation and not the program. Many evaluators are not sensitive to the ways in which social and economic research changes respondents and subjects.

One approach to dealing with panel conditioning is to replace respondents and "refresh" the sample. A design which replaces a number of the panel respondents with each survey is known as a rotating panel and is the approach employed by labour force surveys to estimate unemployment rates.

Another method for controlling panel conditioning is to conduct cross-sectional surveys and compare selected variables with those obtained from the longitudinal sample. This provides a basis for comparison and permits the researcher to evaluate the extent of conditioning.

Attrition occurs when members of the panel move, die or otherwise stop participating. If those who leave a panel are a random subset of the total panel, no corrective action is needed. This is rarely the case. For example, in a program designed to train social assistance recipients, those who leave the panel might be younger people who find work elsewhere and become untraceable. The panel becomes increasingly distorted away from the initial sample. Two choices are to replace the missing respondents by meticulous matching, or to model the attrition statistically and estimate the resulting bias. Matching is usually difficult and rarely satisfying. Statistical modelling of attrition, compiled with careful follow-up is the preferred course of action. Of course, eliminating attrition is the ideal solution and the practical mechanics of this are reviewed below when logistics of follow-up are reviewed.

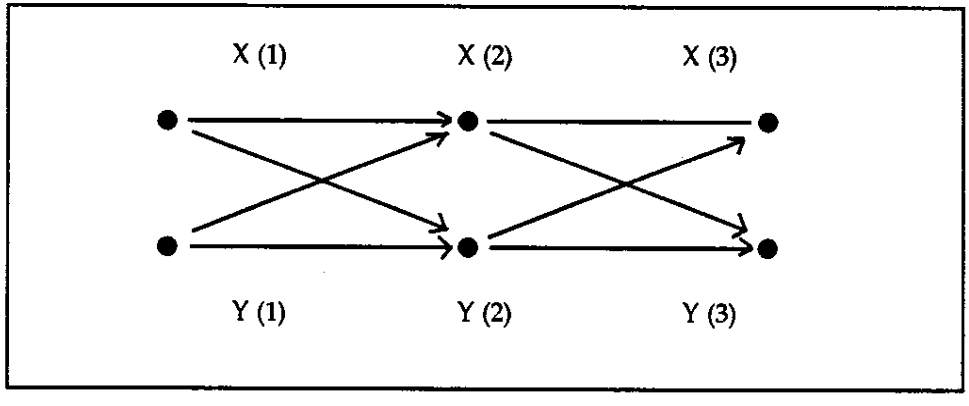
This is not the place to explore these important problems from a theoretical perspective. The general literature of sample selection as represented by the work of Heckman (1979) is a useful starting point for the interested reader.

Temporal stability is a feature of many social processes, where an attribute remains stable throughout the measurement period. Some social and economic processes develop slowly, and panel studies over a few months or even years simply may not reveal the effects of the intervention. This may be seen by the analogy of rusting iron. To understand corrosion processes, a passive observational panel study where iron bars are watched is less effective and takes too much time compared to experiments which manipulate the presumed causes of rusting. Some social processes are too slow to produce timely information.

main plausible. There are many competing hypotheses which are explained by a longitudinal data set; only by specifying a causal (structural) model in detail is it possible to initiate a process of systematically eliminating false models.

While not conclusive evidence of causality, if the two cross-lagged correlations are significantly different, then there is an additional basis for supposing that causal relation may exist. This is still an enhancement over the cross-sectional data we normally use.

Figure 4.
Cross-lagged Correlation



At the most formal level, the term "causal model or structural" has assumed a distinct meaning in social science. Combining elements of psychometrics and econometrics, causal models are based on the work of Joreskog and Sorbom (1977). A causal model is divided into two general components — a *measurement model*, usually a factor analysis, and a *structural model*, usually a system of regression equations. Econometricians are familiar with structural models while psychometricians are familiar with the measurement model. The key to the contribution of Joreskog is to unify factor analysis and structural modelling. The measurement model links latent (unmeasurable) concepts such as "consumer confidence" to measurable variables such as responses to specific questions regarding the consumer expectations or actual purchase data. The "linear structural equation model" (LISREL), which may include the latent variable as an independent or dependent variable, builds a system based on observable behaviour and integrates latent or unobserved variables through a factor analysis structure. The causal model, comprised of a set of multivariate regression questions is estimated using maximum

Sampling is always important. Initial biases produced by selective respondent participation in a cross-sectional study are magnified in a panel study. This initial bias can be further exaggerated by attrition. For example, young men are often difficult to enrol at the outset of a study, and will often leave prior to completion. Special care in drawing the sample and even over-sampling may be required to compensate for this bias. Each population needs to be assessed to counteract tendencies toward bias which may occur.

Response Burden

A panel study requires additional commitment from respondents. Researchers may find that extra inducement, and even financial compensation, is needed to maintain the sample. Questionnaires may also have to be shorter to reduce respondent burden. There is a tendency for panel studies to have very long rhetorical questionnaires. Possibly the higher cost encourages researchers to get the best value by packing all issues into one research project.

Respondent Follow-up and Tracking

Maintaining contact with respondents is obviously important. The effort in maintaining contact will often equal, or even exceed that of data entry, analysis and report preparation. It is essential that sufficient resources be allocated to follow-up, especially in studies extended over several years, and including highly mobile populations.

Recent refinements in respondent follow-up have enhanced the integrity of longitudinal panel studies. From one perspective, the panel study is simply a long survey interview. Maintaining respondent contact is a process of ensuring respondent allegiance to the aims of the study in a somewhat analogous form as ensuring respondent attention in a conventional survey. Of course, during the intervening periods between panels, even the most committed respondent will have other things on his/her mind. There are a few basic principles to maintaining the integrity of the sample:

Separation of tracking and data collection involves specializing in that function and not burdening everyone with that task. Do not require the "tracker" to also conduct interviews or undertake data entry functions. Focusing on the tracking tasks tends to produce superior results compared to sharing this function among the research group.

Tracking is a specialized function. Staff who enjoy locating respondents and

perceptions and reported behaviour in response to what is perceived as normal.

Three common approaches exist for tracking respondents: mail, telephone and in-person visit. Mailing a card and requesting it be forwarded is the cheapest, but tends to produce low recovery rate, especially for respondents who have been missing for some time. The visit is the most expensive, but works best for smaller samples where the research requires detailed information. This requires stable communities where friends and neighbours might know the whereabouts of the respondent.

Tracking by phone is a good compromise and is always the starting point for most follow-up activities. It is limited to published directory sources, or telephones of friends and relatives. These may be wrong if the tracking is done two or more years after the last contact. Of course, primary contacts such as parents, family and friends should always be called first.

One problem is that these family members may be protective and unwilling to provide this information. An option is to include these contacts in the research process by making sure they are fully informed about the nature of the project. This raises the overhead costs of the research.

The tracking strategy needs to be well defined at an early stage in the study since much of the information for these tasks will emanate from the initial contact. Not only must the respondent be "sold" on the participation, but so must friends and relatives. Later, if contact with the respondent is lost, these people need to feel comfortable in providing the "tracker" with information on the whereabouts of the respondent.

Special care is required in creating the "tracking group" and constantly empowering them to employ all data sources. Access to data bases and protection of respondent privacy are central issues to all survey research, but the tracker who accesses friends, family and other sources of information must also be very careful about ethical boundaries.

Application of modern tracking methods, and ensuring that this function is well funded can result in maintenance of over 90% of respondents over a prolonged period of time. Some follow-up studies have been able to locate most respondents interviewed 10 years previously. In many cases, a certain amount of luck is also needed. The basic rule is simply to have repeated and frequent contact with respondents, to obtain multiple identifying information at the start, and to organize the tracking function with military precision.

SUMMARY

Longitudinal survey research requires the usual attention to sound research design. It places added stress on questionnaire design and validity as well as the need to have a refined data base framework. Of special importance are attrition and panel

Applications in the Social Sciences (No. 55). Beverly Hills, CA: Sage.
Dillman, D. (1978). *Mail and Telephone Surveys: The Total Design Method.* New York: Wiley-Interscience.
Dwyer, J.A., (1983). *Statistical Models for the Social and Behavioral Sciences.* New York: Oxford University Press.
Gottfredson, M. & Hirschi, T. (1987). The Methodological Adequacy of Longitudinal Research on Crime, *Criminology*, 25 (3).
Heise, D.R. (1975). *Causal Analysis.* New York: Wiley.
Heckman, J.A. (1979). Sample Selection Bias as Specification Error, *Evaluation Studies Review Annual* (Vol. 5). Beverly Hills, CA: Sage.
Janson, C-G. (1981). Some Problems in Longitudinal Research in the Social Sciences, In F. Schulsinger, S.A. Mednick, & J. Knop, *Longitudinal Research: Methods and Uses in Behavioral Science*, London: Martinus Nijhoff .
Joreskog, K.G. & Sorbom, D. (1977). Statistical Models and Methods for the Analysis of Longitudinal Data, In D.A. Aigner & A.S. Goldberger (Eds.), *Latent Variables in Socioeconomic Models.* Amsterdam: North Holland.
Kenny, D.A. (1975). Cross-lagged panel correlations; a test for spuriousness, *Psychological Bulletin, 82*, pp. 887-903.
Lee, T.H. & Kong, C.M. (1976). *Five Thousand American Families*, Volume IV, Univer sity of Michigan: Institute for Social Research.
Long. J.S. (1983a). *Confirmatory Factor Analysis: A Preface to LISREL.* Beverly Hills, CA: Sage.
Long. J.S. (1983b). *Covariance Structure Models: An Introduction to LISREL.* Beverly Hills, CA: Sage.
Markus, G.B. (1985). *Analyzing Panel Data*, Sage series in Quantitative Applications in the Social Sciences (#18), Beverly Hills, Sage.
Morgan, J. (1973). *Five Thousand American Families* (Vol. 1). University of Michigan: Institute for Social Research.
Rossi, P.H., Wright, J.D. & Anderson, A. (1983). *Handbook of Survey Research.* New York: Academic Press.
Sudman, S. & Ferber, R. (1979). *Consumer Panels.* New York: American Marketing Association.
Thornton, A, Freedman, D.S. & Camburn, D. (1979). Obtaining Respondent Coop eration in Family Panel Studies, *Sociological Methods and Research, 11* (1).
Trochim, W. (1984). *Research Design for Program Evaluation: The Regression-Discontinuity Approach.* Beverly Hills, CA: Sage.

Greg Mason, Ph.D. is a Principal of Prairie Research Associates, Inc. in Winnipeg, Manitoba. He is also with the Department of Economics at the University of Manitoba.