American Journal of Community Psychology, Vol. 9, No. 6, 1981

Contending with Rival Hypotheses in Correlation of Aggregate Time-Series (CATS): An Overview for Community Psychologists[1]

Ralph Catalano[2]
Public Policy Research Organization,
University of California, Irvine

The design issues surrounding the correlation of aggregate time-series data (CATS) are explained in an attempt to improve its application to problems of interest to community psychologists. Particular attention is given to the nature and control of third variables which give rise to rival hypotheses. Interpretive problems are also discussed. The use of CATS is encouraged when the data allow and when the inferential limits are understood.

Methods for measuring association over time between the behavior of populations and characteristics of their environments have not been widely used by community psychologists. Exceptions to this observation have suggested, however, that suicide (Vigderhaus, 1977), use of mental health facilites (Brenner, 1973), and the incidence of stressful life events, depression (Catalano & Dooley, 1977), and clinical symptoms (Dooley & Catalano, 1979) are related to such environmental factors as economic climate (Dooley & Catalano, 1980). Correlation of aggregate time-series (CATS) has also been used to evaluate the effect of mental health programs on the community (Spearly, 1980). These findings would seem to be of sufficient importance to justify a wider understanding of CATS methods than now exists among community psychologists. Greater familiarity with the technique would sharpen the debate over the existing findings and lead to more rigorous replication and to extension of the research.

[1]The author acknowledges the help of the following persons whose criticisms improved this paper: Arnold Binder, June Catalano, Ross Conner, David Dooley, Gary Evans, Robert Jackson, Stanislav Kasl, and Laurence Steinberg. This research was supported in part by NIMH grant MH 28934.
[2]All correspondence should be addressed to Ralph Catalano, PhD, Program in Social Ecology, University of California, Irvine, Irvine, California 92717.

0091-0562/81/1200-0667$03.00/0

Among the impediments to a wider understanding of CATS methods is the skepticism with which those trained in classical experimental design traditionally view correlational methods. The distrust of correlational designs is rooted in the fact that such approaches usually do not logically rule out the possibility that any association found between two phenomena is due to uncontrolled variables rather than to a causal relationship. This skepticism is deepened in the case of CATS because most time-series literature applies the method to forecasting and control problems in which causal relationships are assumed rather than tested. While preliminary attempts have been made to recast CATS methods in the framework of the controlled experiment paradigm (e.g., Cook & Campbell, 1979), these efforts must be extended if the techniques are to be more widely applied in community psychology. This paper builds on earlier attempts by describing the logic by which CATS can be rigorously applied to hypothesis-testing.

A starting definition of CATS seems appropriate before dealing specifically with its application in community psychology. A CATS analysis is a correlational test of the hypothesis that two attributes of a population or organization are related over time. The test usually involves a parametric analysis, such as multiple regression, in which the longitudinal variation in a measure of population behavior is statistically attributed to several independent variables. The "cases" in these analyses are consecutive time periods such as months or years. In addition to the independent variable of primary theoretical interest, the predictor variables usually include measures of other phenomena which may cause the dependent variable to change over time. If, for example, the hypothesis is that the unemployment rate is longitudinally related to the use of mental health facilities, the independent variables in the regression equation might include measures of other potential causes of variation in facility use. These other causes, such as weather conditions, demographic shifts, or changes in facility budgets, are included to test "rival hypotheses" or the contention that any association found between the variables of interest is due to some third variable. The control of third variables to discredit rival hypotheses is the key to rigorous CATS hypothesis tests. The principal contribution of this paper is a classification of rival hypotheses and of the strategies for discrediting them. The implications of these strategies for interpreting CATS findings are also discussed.

The remainder of this paper is divided into four sections. The general situations in which CATS is a useful analytic strategy for community psychology are described in the first section. The second is a typology of the rival hypotheses which most frequently confound CATS. The third section is a description of a control strategy devised to reduce the threat of rival hypotheses. The fourth section deals with the interpretation of CATS.

APPLICATIONS OF CATS IN COMMUNITY PSYCHOLOGY

There are three general types of problems a community psychologist may face to which CATS can apply. While all three applications are described below, the remainder of this paper is concerned primarily with the second and third.

Atheoretical Forecasting

CATS can to used to predict the future values of one time-series from the present or past values of another. An example of CATS forecasting could be the prediction of the demand for mental health services from changes in the socioeconomic or demographic structure of a catchment area. The processes connecting the variables may be complex and understood only through attitudinal and other individual level research, but the basic correlation can be a useful management tool. Such prediction ideally follows from earlier hypothesis testing which identified causal processes. Accurate prediction from CATS used for forecasting, however, is not definitive evidence that the originally hypothesized causal model was correct at the time of testing or that it remains accurate. The use of CATS to provide definitive evidence of a causal relationship implies hypothesis testing and requires that the rival hypotheses alluded to in Control Strategies (below) be controlled.

Testing Theories of Organizational Behavior

There is a seldom discussed but important difference between the use of CATS to explain person as opposed to organizational behavior. Hypotheses concerned with the latter usually involve dependent variables which cannot be logically ascribed to individuals. A germane example is the annual change in per capita expenditures for mental health services by political jurisdictions (e.g., counties or states). A hypothesis concerned with this behavior might be that temporal variation in per capita mental health expenditures is a function of revenues and of incidence of violent crime committed by persons released from overcrowded mental health facilities. While the longitudinal design and aggregate independent variable (i.e., incidence of violent crime) qualify the analysis as CATS, the dependent variable cannot be conceptualized as a characteristic of an individual. CATS, or other correlational designs, are the most efficient analyses for such problems because individuals cannot be randomly assigned to political jurisdictions in which revenues and the incidence of criminal victimization can be manipulated. Simulations of policital and fiscal constraints and of the threat of victimization in laboratory settings may be possible but would probably be no more logically compelling than well-designed CATS.

The essential characteristic of CATS used for testing theories of organizational behavior is that either the dependent or independent variable must be a descriptor of a functional organization of individuals rather than a count or proportion of individually exhibited phenomena in a population. While the obvious examples involve group-decision behavior, there are others which involve such organizational phenomena as economic performance (e.g., profit return by business organizations).

Testing Etiological Theories

The application of CATS that has caused the most controversy is its use to test theories of the etiology of disorder in individuals. These analyses usually involve a dependent variable that is an aggregate expression (e.g., frequency or proportion) of the incidence of an individually exhibited characteristic (e.g., number of persons treated at mental health facilities) in a population and an independent variable that is either another person-based variable (e.g., number of unemployed workers) or an environmental factor (e.g., change in weather conditions). Much of the controversy surrounding this type of analysis arises from the "ecological fallacy" or the assumption that those individuals in the population experiencing the assumed independent variable are also those who exhibit the dependent variable. This problem and its implications for research have been extensively described and debated in the sociological literature (Borgatta & Jackson, 1980; Gove & Hughes, 1980) and will not be elaborated upon here. The community psychologist pursuing any research (not only that using CATS) that uses aggregate data should be familiar with this debate.

CATS tests of etiological theories have also been controversial because uncontrolled third variables often give rise to rival hypotheses for discovered associations (Kasl, 1979). The following classification of third variables and of control strategies for them should help reduce this controversy or help focus it on substantive rather than methodological issues.

RIVAL HYPOTHESES

The most common rival explanation to the hypothesis that two variables are causally related over time is that a "third variable" causes the phenomena to move similarly. A third variable can be a phenomenon external to the dependent and independent variables (e.g., weather) that causes them to move similarly through time, or it can be a common internal process which is not an actual third entity but which causes two phenomena to behave similarly through time (e.g., growth processes). While third variables are most commonly cited as leading to erroneous rejection of the null hypotheses, it is also possible that they can obscure true relationships and lead the researcher to erroneously infer no association.

The following discussion of third variables is offered as a guide to conceptualizing the problem so that an appropriate control strategy can be devised. All potential third variables can be characterized on three dimensions: regularity, accessibility, and latitude. Understanding these dimensions is essential to selecting appropriate control strategies or to evaluating the integrity of reported CATS findings.

Regularity

Third variables that could cause two series to be spuriously related, or which could mask a real association, can be classified by their temporal behavior as either regular or irregular. Regular behavior refers to repeating patterns the most common of which are linear trend and seasonal cycles. A regular third variable can cause the two-subject phenomena to linearly increase or decrease or to move through similar cycles over time. Typical regular third variables include shared growth patterns, climatic seasons, or seasonal cycles in institutional activity (e.g., school openings and vacations).

In a test of the hypothesized relationship between unemployment and use of mental health facilities, for example, it is possible that bad weather increases the former by precluding out-of-doors work (e.g., construction) and the latter by actually depressing people or by driving marginal people to seek shelter by feigning need for treatment. It is also possible that unemployment has seasonal cycles unrelated to weather conditions (e.g., end of Christmas retailing peak) as might use of facilities (holiday depression). In either case, there would be a spurious correlation between unemployment and use of facilities due to coincident regular behavior.

Third variables do not, of course, have to be regular. They may vary randomly over time and cause the analyzed variables to be spuriously correlated. As discussed below, irregular third variables are difficult to detect and require the use of comparison communities as a control strategy.

Accessibility

There are two types of accessibility. First, a potentially confounding variable can be intellectually accessible. That is, it can be suspected by the researcher because it has been empirically or hypothetically linked to the dependent variable. On the other hand, there may be a confounding variable which has not been discovered or has never been suspected of influencing the dependent phenomenon and is therefore intellectually "inaccessible."

The second type of accessbility refers to whether or not a researcher can obtain time-series values for a potentially confounding phenomenon when it is intellectually accessible. If no measurement over time is possible, the potentially confounding third variable is inaccessible. The reason that these two types of accessibility were not separated is that, as discussed below, one strategy controls both.

Latitude

Third variables can influence the behavior of either or both the dependent and independent phenomenon only in the population under study or in all other similarly defined populations. The use of mental health facilities in neighboring metropolitan areas may, for example, be influenced by generally occurring phenomena such as the seasonal cycles alluded to earlier. One area, however, may experience shifts in the policies or procedures by which persons are admitted for treatment. This local third variable may produce temporal behavior in admissions which is coincident with unemployment rates or which hides a true relationship between admissions and unemployment.

There are many third variables whose effect is general. Seasonal cycles, generally mandated changes in record-keeping, and changing mores or values are typical of phenomena which will probably confound relationships across many populations. While generally occurring third variables can be controlled through techniques suggested below, local confounding phenomena pose a more difficult problem. This is particularly true if the local third variable is also inaccessible and irregular.

CONTROL STRATEGIES

The null hypothesis in a CATS test is that there is no temporal association between two time-series after all confounding third variables have been controlled. The following are four components of a rigorous control strategy that can reduce the threat of rival hypotheses arising from third variables.

Specification

If a third variable is both intellectually and physically accessible, the best strategy for control is to include it as a covariate in the test of association. If multiple regression routines are used for the test, for example, the suspected third variable would be included in the formula and its standardized beta coefficient used to compare its effect with that of the independent variable of interest.

While specification is the most direct and efficient method of control, it can be applied only when a potentially confounding variable is accessible. Because one dimension of accessibility depends on the researcher's suspicion of potential confounds, it is possible, if not probable, that many third variables will not be specified.

Decomposition

As noted above, a phenomenon can exhibit behavior which is regular over time. This behavior may defy simple specification if the external cause of the regularity is inaccessible or if there is no external cause. The conservative researcher should then assume that *any* regular behavior in either the dependent or independent variable is a potential cause of spurious association and empirically remove the trend or cycle. The process of separating a time-series into its regular and random components is called decomposition.

The first step in decomposition is to detect any regular behavior which may not be obvious by simple observation of the series. This is typically done by constructing an "autocorrelogram" which is the plotting of correlations between the series and its own lags (Kendall, 1973; McLain & McCleary, 1979; Nelson, 1973). An autocorrelogram of monthly data, for example, would be carried through at least 12 lags to test for annual as well as quarterly and semiannual cycles. The Q statistic, a chi-squarelike test, is used to determine if chance is a likely rival hypothesis to true trend or cycle in the autocorrelogram (McCleary & Hay, 1980).

The structure of an autocorrelogram yielding significant coefficients in excess of chance, or at hypothetically obvious positions (e.g., Lag 12), suggests the behavior of the regular third variable whose effect should be removed from the series of interest. Removing the regular behavior often involves mathematical routines which have been simplified for researchers by automated programs such as those based on the Box and Jenkins (1976) approach.

When the regular behavior of the series has been removed, a researcher can proceed with specifying the accessible third variables and with testing association. If the specified third variables were regular, the confidence intervals of their beta weights will include zero because their effect, along with the effect of any inaccessible regular third variables, will have been removed from the analyzed phenomena by decomposition.

Comparison

Among the remaining rival hypotheses is that a third variable is neither accessible nor regular and therefore not be controlled by either specification or decomposition. Comparison is the best strategy for controlling irregular, inaccessible confounds. This strategy involves identifying one or a group of populations or organizations which share as many of the physically inaccessible irregular third variables as possible with the test population. The researcher introduces the dependent variable as observed in the comparison group or groups into the equation as controls and assumes that the effect of any *generally occurring,* inaccessible, irregular confounds is removed (Campbell, 1968; Catalano, 1979; Steinberg, Catalano, & Dooley, 1981).

The use of comparison groups has not been common either because data describing the dependent variable have not been available for other than the test group or because the independent variable behaves identically in the test and in available comparison groups. When the latter is the case, the researcher should make sure that the test group has functional or theoretical meaning. If the test population is not defined by the action of the independent variable, there must be some theoretically compelling reason for assuming its members are affected by the processes under investigation.

Replication

After the researcher has specified the accessible third variables, removed regular behavior, and introduced, where possible, comparison groups, the last and most difficult rival hypothesis is that an irregular, inaccessible, locally occurring confound remains. The only way to reduce this threat is to replicate the findings in other populations. If the threat is real and has obscured a true, or produced a spurious, association, replication in as many populations as possible should lead to convergence on the true relationship.

Replication of a CATS finding can often be part of an analysis in which a comparison group is used. As noted above, controlling for irregular, local, inaccessible third variables requires the collection of dependent variable data in a comparison group. If the effort to collect comparison-dependent variable data is made, independent and accessible control data should, if possible, also be collected. The data would then allow a "two-way" CATS in which the groups act as controls for each other and provide replication.

The threat of rival hypotheses is minimized through specification, decomposition, comparison, and replication. If a researcher cannot apply one or more of the above controls, any descriptions of the findings should acknowledge that one or several types of the third variables listed in Control Strategies (above) remain possible threats. The degree to which the findings then gain acceptance depends on how compelling an argument the researcher can make against the type of third variable left uncontrolled.

The statistical routine used to incorporate specification, decomposition, and comparison in a CATS hypotheses test will depend on the nature of the measurements available. In terms of multiple regression, the strategy would be to construct an equation in which the independent variables would include the fitted values of the regular behavior model (e.g., Box-Jenkins estimates), any specified variables (e.g., weather), a comparison community's values of the dependent variable, and the suspected causal variable. The test of association reduces to the question of whether the suspected causal variable uniquely "explains" any variation in the dependent variable beyond an acceptable level of chance. This is a very conservative test of association because, among other reasons, it risks "throwing the baby out with the bath water" by introducing controls which may remove "true" association between the variables of interest (Mark, 1979). The technique is recommended, however, on the assumption the researcher would rather falsely accept than reject the null hypothesis of no association.

INTERPRETING CATS

Researchers unfamiliar with CATS should be aware of the major issues surrounding its interpretation. Several of these (e.g., reverse causation and ecological fallacy) are inherent in correlational and aggregate analytic strategies, while others (e.g., autocorrelated error terms and representativeness of time samples) are more typical of time-series analyses. The time-series specific issues are discussed below.

Biased Coefficients

One problem which often plagues CATS is dependence among error terms remaining after variance shared with controls and independent variables have been removed from the dependent variable. The issue of error-term dependence is important because the t or F statistics commonly used to estimate the confidence intervals of beta weights or correlation coefficients assume independent error terms.

The technique most commonly used to detect dependence among error terms is the Durbin-Watson test (Durbin & Watson, 1950; 1951). The test yields a statistic which is sensitive to either a positive or inverse relationship between error terms and their first lag. Higher order dependencies (e.g., quarterly or 6-month relationships), however, are not detected by the standard Durbin-Watson test and a variant on the test must be used when the historic behavior of the dependent variable is included in the analysis as a control variable (Koutsoyiannis, 1973). An alternative to the Durbin-Watson test is to construct an autocorrelogram (see above, Control Strategies: Decomposition) of the residuals and to inspect it for association greater than chance. The presence of association beyond chance suggests that further control is needed in the analysis.

The literature reporting CATS findings has often confused the problems of uncontrolled third variables with that of dependent error terms. Passing the Durbin-Watson test, for example, has often been cited as evidence that rival hypotheses have been discredited. It should be noted, however, that the standard Durbin-Watson test does not detect either higher order dependencies or the effect of irregular third variables. Regression error terms remaining from variables which were irregular may, moreover, be regularly patterned. As a rule, therefore, autocorrelograms of error terms should be computed and inspected for dependencies.

Representativeness of Time-Series

Another inferential problem inherent in CATS is made obvious by the question, how representative is the time period used in the analysis? This is a difficult issue which is rarely addressed even in the advanced time-series literature. The researcher who believes that the measured relationship generalizes beyond the time period analyzed must depend on replication and argue that there are no reasons to assume that the period was atypical.

Reverse Causation

Many correlation analyses offered as support for etiological theories of an association between two variables says nothing about the direction of causation. The time-series technique which allows control for reverse causation is cross-correlation (Mark, 1979). Cross-correlation involves computing not only a synchronous (the dependent and independent variable occurring in the same time period) coefficient but also equal numbers of lead (dependent variable temporally preceding the independent) and lagged (dependent variable temporally following the independent) coefficients. When the coefficients are ordered from lead, through synchronous, to lagged configurations, the pattern should conform to the theory underlying the hypothesis.

The most typical theoretical pattern is one in which the effect of change in the independent variable is manifested in subsequent behavior of the dependent variable. In such cases the number of significant coefficients among the lead and synchronous configurations should not exceed chance while there should be greater than chance significant coefficients among the lags. Should more significant coefficients than expected by chance appear among the lead and synchronous configurations, the reverse causation hypothesis can, depending on the nature of the phenomena under study, be as tenable as the primary hypothesis.

If the underlying theory predicts that the relationship between two variables should be found only in the synchronous configuration, and such a pattern is found, reverse causation is not logically ruled out. The plausability of reverse causation in such cases is established by the nature of the variables in question. A correlation between precipitation and auto accidents, for example, is less likely to be threatened by the reverse causation hypotheses than one between the incidence of stressful life events and of disease symptoms.

Strength of CATS Relationships

As may have been inferred from the above discussion of control, the variance in a dependent variable can be altered considerably by specification, decomposition, or comparison. This alteration can lead a researcher to either over- or underestimate the true effect of the independent variable if he or she attempts to use CATS to measure the strength of a relationship. Overestimation is usually caused by the use of partial correlation techniques in which the ultimate test of association is between variance remaining after control. Although the partial correlation coefficient may be very high, it is not a measure of the association between the original variables but rather between their residuals. Researchers should make this difference explicit when reporting CATS findings.

Underestimation of the effect of the independnet variable is most likely to occur when a comparison group has been used. The problem is that the independent variable may be at work in the test and comparison groups and move similarly but not identically in both. The result is that a portion of the true effect in the test group is removed by the introduction of a comparison group as a control for generally occurring third variables which are irregular and inaccessible. In multiple regression analyses, for example, the standardized beta weights could underestimate the strength of the independent variable and may lead the researcher to assume the effect is trivial.

The most prudent strategy is to consider CATS findings as hypotheses tests and not as estimates of the strength of effect. When specification, decomposition, or comparison are introduced, the researcher sacrifices information on the strength of effect to gain certainty that an effect does or does not exist. Given the potential for serious rival hypotheses, the researcher should be willing to make that sacrifice.

REFERENCES

Borgatta, E., & Jackson, D. *Aggregate data analysis and interpretation.* Beverly Hills: Sage, 1980.

Box, G., & Jenkins, G. *Time-series analysis: Forecasting and control.* San Francisco: Holden-Day, 1976.

Brenner, M. H. *Mental illness and the economy.* Cambridge, Mass.: Harvard Press, 1973.

Campbell, D. The Connecticut crackdown on speeding, time series data in quasi-experimental analysis. *Law and Society Review,* 1968, *3,* 33-53.

Catalano, R. Health costs of economic expansion: The case of manufacturing accident injuries. *American Journal of Public Health,* 1979, *69,* 789-794.

Catalano, R., & Dooley, D. Economic predictors of depressed mood and stressful life events in a metropolitan community. *Journal of Health and Social Behavior,* 1977, *18,* 292-307.

Cook, T., & Campbell, D. *Quasi-experimentation: Design and analysis for field settings.* Chicago: Rand McNally, 1979.

Dooley, D., & Catalano, R. Economic, life, and disorder changes: Time-series analyses. *American Journal of Community Psychology,* 1979, *7,* 381-396.

Dooley, D., & Catalano, R. Economic change as a cause of behavioral disorder. *Psychological Bulletin,* 1980, *87,* 450-468.

Durbin, J., & Watson, G. Testing for serial correlation in least squares regression: Part I. *Biometrika,* 1950, *37,* 409-423.

Durbin, J., & Watson, G. Testing for serial correlation in least squares regression: Part II. *Biometrika,* 1951, *38,* 159-178.

Gove, W., & Hughes, M. Reexamining the ecological fallacy: A study in which aggregate data are critical in investigating the pathological effects of living alone. *Social Forces,* 1980, *58,* 1157-1177.

Kasl, S. Mortality and the business cycle: some questions about research strategies when utilizing macro social and ecological data. *American Journal on Public Health,* 1979, *69,* 784-789.

Kendall, M. *Time-series.* London, Griffin, 1973.

Koutsoyiannis, A. *Theory of econometrics.* London: McMillan, 1973.

Mark, M. The causal analysis of concomitancies in time series. In T. Cook & D. Campbell (Eds.), *Quasi-experimentation.* Chicago: Rand McNally, 1979.

McCleary, B., & Hay, R. *Applied time series analysis for the social sciences.* Beverly Hills: Sage Publications, 1980.

McLain, L., & McCleary, R. The statistical analysis of the simple interrupted time-series quasi-experiment. In T. Cook, & D. Campbell (Eds.), *Quasi-experimentation.* Chicago: Rand McNally, 1979.

Nelson, C. *Applied time-series analysis for managerial forecasting.* San Francisco: Holden-Day, 1973.

Spearly, J. Evaluating the impact of community mental health centers on hospital admissions: An interrupted time-series analysis. *American Journal of Community Psychology,* 1980, *8,* 229-242.

Steinberg, L., Catalano, R., & Dooley, D. Economic antecedents of child abuse and neglect. *Child Development,* 1981, *52,* 260-267.

Vigderhous, G. Forecasting sociological phenomena: applications of Box-Jenkins methodology to suicide rates. In K. F. Schuessler (Ed.), *Sociological methodology 1978.* San Francisco: Jossey-Bass, 1977.