Psychological Bulletin
1973, Vol. 80, No. 2, 93–105

N-OF-ONE AND N-OF-TWO RESEARCH IN PSYCHOTHERAPY

JOHN M. GOTTMAN[1]

Indiana University

This paper suggests that time-series methods can be used to make weak and strong causal inferences in N-of-one research in psychotherapy. Possible gains are discussed in process research, outcome research, and measurement design. Specific attention is given to assessing the effects of planned interventions using the interrupted time-series design; experiments are presented which illustrate the methods discussed in the paper. The major advantages of time-series methodology are that it (*a*) permits the study of the single subject and the use of subject-as-his-own-control research, (*b*) permits the study of the form of the effect of the intervention over time, and (*c*) allows one to use information over time as feedback for making decisions—a useful tool in the evaluation of psychotherapy.

Dukes (1965) found 246 single-subject studies in the period from 1940 to 1965. These studies presented diverse qualitative information and careful observations. *N*-of-one research has also involved the use of quantitative observations over time. Holtzman (1967) suggested that the use of series of observations over time deals with the problem of obtaining replications in *N*-of-one research. This application of time-series analysis to *N*-of-one research in the behavioral sciences is relatively new, although it has a long history in other sciences.

Perhaps the most promising use of time series has been proposed in the interrupted time-series design in which a series of observations denoted 000 precede and follow the introduction of an intervention, *I*. Campbell and Stanley (1970) suggested the interrupted time-series design as an excellent quasi-experimental design with special appeal for evaluating interventions in settings which typically keep archival records over time. Glass, Willson, and Gottman (1973) presented a review of time-series methods and analytic procedures for the use of time series in the behavioral sciences.

The utility of time-series designs for research in psychotherapy is discussed in this paper. The intention of this paper is to teach an approach recently developed for the analysis of time-series data with fewer observations than are usually available in the physical sciences and which are typical in behavioral investigations. Computer programs for the statistical analysis of interrupted time-series experiments are available upon request.[2]

Uses of N-of-One and N-of-Two Research in Psychotherapy

A review of the process and outcome literature in psychotherapy reveals several possible advantages in the utilization of *N*-of-one research for the construction of theories of psychotherapy. There are also important practical gains in the application of time series for the evaluation of psychotherapy and for providing feedback loops which may serve as an aid to the therapist during the course of treatment.

Possible Advantages of Time-Series Analysis in Outcome Research in Psychotherapy

In outcome research, it may be misleading to average data over individuals. For example, Bergin (1966) presented evidence for what he called "a curious and provocative finding" in a number of outcome studies which found that the variance of change scores for a group of subjects in individual therapy far exceeded the variance of change scores for an untreated group. He took this as evidence for a "deterioration effect," that is, he argued that these data showed that some patients are

[1] Requests for reprints should be sent to John M. Gottman, Department of Psychology, Indiana University, Bloomington, Indiana 47401.

[2] A manual for the analysis of the interrupted time-series experiment is available from the author. This manual presents a computer program based on the IMA (1,1) model, teaches the reader to interpret output, explains how the program works (intuitively and mathematically), presents an example, and refers the reader to other programs and sources.

TABLE 1

Variances of Discrepancy Scores on MMPI Scales for Individual Psychotherapy and Nontreatment Groups

MMPI scale	Individual psychotherapy variances[a]	Nontreatment group variances[b]	F
L	19.89	23.43	1.18
F	215.21	22.94	9.38[c]
K	55.95	31.70	1.76
Hs	127.46	64.16	1.99[d]
D	244.30	93.32	2.62[c]
Hy	113.21	87.80	1.29
Pd	155.00	89.68	1.73
Pa	111.94	68.06	1.64
Pt	208.51	73.27	2.85[c]
Sc	272.91	74.13	3.68[c]
Ma	127.79	75.34	1.68
Es	43.56	14.82	2.94[d]

Note. Adapted from an article by D. S. Cartwright published in the October 1956 issue by the *Journal of Consulting Psychology*. Copyrighted by the American Psychological Association, Inc., 1956.
[a] $n = 42$.
[b] $n = 23$.
[c] $p < .05$.
[d] $p < .01$.

greatly harmed by psychotherapy while some are greatly helped. He suggested that "we should find out whether some therapists make people better and some make them worse or whether individual therapists do both [p. 121]." Bergin's conclusion is unwarranted from the data he presents. For example, Table 1 is a reproduction of Cartwright's (1956) reanalysis of the well-known Barron and Leary (1955) study discussed by Bergin.

Two factors can affect the variance of change scores, the initial and final variances and the test–retest correlation. The largest F-ratio value in Table 1, 9.38, is for the F scale of the Minnesota Multiphasic Personality Inventory (MMPI). However, the F-ratio for the *pretest* MMPI variances can be computed from the Barron and Leary study as $F = 2.28$, which is significant at $p < .01$. This fact alone reduces the impact of the evidence for a "deterioration effect." Barron and Leary did not present test–retest correlations for the MMPI, but Schofield (1950) did. He found that the test–retest correlations were considerably reduced for some treatment groups when compared to a nontreatment group. The data in Table 2 indicate this phenomenon. A reduction in test–retest correlation would greatly inflate the variance in change scores for the treatment group. The variance of change scores, σ_D^2, is:

$$\sigma_D^2 = \sigma^2_{\text{pretest}} + \sigma^2_{\text{posttest}} - 2r\sigma_{\text{pretest}}\sigma_{\text{posttest}},$$

where r is the test–retest correlation. Assuming test and retest variances are equal to 1.0 for both treatment and no-treatment groups, and using Schofield's (1950) data, this gives:

$$\sigma_D{}^2_{\text{no treatment group}} = .6$$

and

$$\sigma_D{}^2_{\text{treatment group}} = 1.746,$$

which gives an F-ratio of 2.81 ($p < .01$).

There is another rival hypothesis to the deterioration effect, namely, a regression effect. Psychotherapy is often designed to assist

TABLE 2

MMPI Test–Retest Correlations for a Nontreatment and a Treatment Group

Test correlations

Group	n	L	F[a]	K	Hs[a]	D[a]	Hy
Nontreatment group	42	.810	.794	.663	.724	.734	.682
Hospitalized psychotics	13	.852	.375	.105	.220	.707	.436

Retest correlations

Group	n	Pd	Mf	Pa	Pt[a]	Sc[a]	Ma
Nontreatment group	42	.569	—	.585	.701	.594	.783
Hospitalized psychotics	13	.375	—	.052	.147	−.303	.608

[a] Scales for which Cartwright (1956) found significant differences in change score variances between an individual psychotherapy and a nontreatment group.

individuals at the extremes of some distribution. For example, using the variable of assertiveness, both the coward and the bully may be subjects for therapy. If psychotherapy serves to move patients closer to the mean, the major effect of psychotherapy may be an effect on the variance rather than the mean of the distribution. If therapy is effective, we would expect a decreased variance in the posttest scores. This decreased variance in the posttest scores would also result in an increased variance in the change scores for the treatment group. Schofield (1950) presented data which may support the "regression hypothesis." On two of the four scales (*F* and *Pt*) for which Cartwright (1956) found significantly different change score variances, the treatment group's posttest variance is half of its pretest variance while there has been no change in variance on these scales for the no-treatment group (see Table 3). Therefore, an analysis of the effectiveness of psychotherapy could be misleading if only means or grouped change scores were inspected. Even with randomized analysis of variance experiments, the assessment of change within individuals over time (or "*N*-of-one-at-a-time research")[3] would shed some light upon the question of psychotherapeutic outcome.

TABLE 3

PRE- AND POSTTEST VARIANCES FOR A TREATMENT AND NONTREATMENT GROUP ON FIVE OF THE MMPI SCALES FOR WHICH CARTWRIGHT (1956) FOUND SIGNIFICANTLY DIFFERENT CHANGE SCORES

Group	Variance				
	F	*Hs*	*D*	*Pt*	*Sc*
Nontreatment					
Pretest	6.8	47.6	96.0	51.8	57.8
Posttest	6.8	32.5	59.3	54.8	51.8
Hospitalized psychotics					
Pretest	88.4	176.9	121.0	130.0	256.7
Posttest	44.9	125.4	53.3	74.0	243.4

It may also be misleading to ignore the variable of time in follow-up research. Time-series designs do indeed offer an alternative when a traditional experimental design is not feasible. However, their most important contribution is that they offer a unique perspective on the assessment of interventions. Experimental designs in the Fisherian tradition may obfuscate important observations about the *form* of intervention effects across time. Simultaneous randomized designs have become so much the method of investigating treatment effects that behavioral scientists have lost sight of the fact that these designs were originally developed for use in evaluating agricultural field trials. Fisherian methodology was most appropriate for comparing agricultural treatments with respect to relative yields. The yields were crops which were harvested when ripe; it was irrelevant whether the crops grew slowly or rapidly. For social systems, however, there are no predetermined planting and harvesting times. Interventions with clients, institutions, communities, and societies do not merely have an "effect" but an "effect pattern" across time. The value of an intervention is not judged by whether the effect is observable at the fall harvest but by whether the effect occurs immediately or is delayed, whether it increases or decays, whether it is temporarily or constantly superior to the effects of alternative interventions evaluated in a cost/benefit sense. The time-series designs provide a methodology appropriate to the complexity of the effects of interventions into human systems. For example, many change processes follow an evolutionary operations curve, which is characterized by an initial extinction curve, followed by a learning curve which elevates the series to a new level. This curve has been reported anecdotally in remedial reading where old habits must be unlearned before new habits are acquired. It is also the curve describing the successful survival of a species following an adaptive mutation (Box & Draper, 1969) where the ordinate is the population of the species at time *t*. In relation to psychotherapy outcome research, if an evolutionary operations curve described the process of change, differential results would be obtained depending upon where in the change curve an assessment was performed. Again, it may be meaningful to look at change within subjects over time.

[3] The author is indebted to Alexander Buchwald and Steven Shmurak for suggesting the term "*N*-of-one-at-a-time."

TABLE 4

Time-Series Analysis of Harris, Wolf, and Baer: Data Presented by Bandura (1969) without Statistical Analysis Showing Lack of Significance of Shifts

Time periods	Student's t for shift in level	Student's t for shift in slope	df
Base line to interaction reinforced	.10	−.09	5
Interaction reinforced to solitary play reinforced	−.01	.00	4
Solitary play reinforced to interaction reinforced	.05	−.05	13

Data over time have often been presented without any statistical analysis. This practice is typical of research in behavior modification. However, an analysis of data presented by Bandura (1969, pp. 26–27) using statistical methods presented in this paper indicates that none of the observed shifts are significant at the .05 alpha level. Table 4 is a presentation of Student's t values calculated for shifts in level and slope. Therefore, it is possible that "eyeballing" time-series data may be misleading.

Time-series experiments have often been presented with incorrect statistical procedures used to assess the effects of an intervention. The use of time-series data introduces an important problem of statistical inference, namely, that observations over time are often dependent (or autocorrelated). Chassan (1967) overlooked some relevant research when he suggested that "on the basis of some preliminary theoretical statistical analysis it appears that the standard t test can be used with a reasonable validity even within a high autocorrelated, dependent series [p. 201]." Statistical tests of significance generally involve assumptions of independence. Violations of independence have been shown by Sheffé (1959, chap. 10) to be serious. Confidence intervals of 95% are reduced to 75% for even moderate autocorrelation, and the problem becomes even more severe as the autocorrelation increases. Glass et al. (1973) showed that for a simple, stationary time-series model with moderate autocorrelation of .4, the nominal 90% confidence interval in estimating the mean level of the series is .72. For a reduced autocorrelation of .3, the nominal confidence coefficient becomes .76. The effects of dependent observations on probability statements cannot be safely disregarded, unlike considerations of normality and homogeneous variances. Gastwirth and Cohen (1970) found nearly identical results for small numbers of observations using the same autoregressive model as Glass et al. did.

For the interrupted time-series experiment, denoted 000I000, it is clear that a simple t test between base line and experimental period means is misleading. An increasing trend in the base line, for example, which continues uninterrupted into the experimental period results in a significant t even though there has been no experimental effect. Similarly, a reversal of slope results in equal means for both periods, yielding a t of zero even though there has been a dramatic experimental effect. Chassan suggested using a t test on deviations from a trend line, which is a better suggestion, but it is still only a stab at a complete solution to the analytic problem.

Tyler and Brown (1968) presented the results of an experiment which illustrate the above point. They presented data which indicate an apparent shift in means from base to experimental period; by using an analysis of variance these shifts were found to be significant. However, time-series analysis indicates that the intervention did not produce a significant shift in either group (see Table 5). It is not only expedient but necessary to use appropriate statistical tools, and this paper suggests which procedures are appropriate.

TABLE 5

Time-Series Analysis of Tyler and Brown (1968) Data

Group and experimental condition	Student's t for shift in level	Student's t for shift in slope	df
Group 1: Contingent to noncontingent reinforcement	−.01	−.15	26
Group 2: Noncontingent to contingent reinforcement	−.13	.08	26

Possible Advantages of Time-Series Analysis in Process Research in Psychotherapy

In process research it may also be meaningful to investigate change within people as well as change across people. Bakan (1967) extended a paper by Sidman (1952) which he described as having dealt "a devastating criticism of a great deal of current and historical psychological research [p. 30]." Bakan and Sidman have demonstrated that if the functional relationship between two variables, x and y, is assessed for any particular individual, then even if the form of the relationship is similar across individuals, the average x values and average y values may be related in a fundamentally different way. Bakan demonstrated this by using a Taylor series expansion of $y=f(x)$ for an individual and showing that the average values do not follow the same function for learning curve data, $y=m-me^{-kx}$.

This has important implications for the construction of theories of psychotherapeutic process. Unfortunately, it may be necessary to construct minitheories specific to an individual and to attempt replications.

Parker and Fleishman (1960) illustrated this point in a study of complex tracking behavior. It is instructive to compare the individual learning curves they presented with their average learning curve. They often do not resemble each other. The averaging of data may be misleading in the study of the psychotherapeutic process.

Possible Advantages of Time-Series Analysis in Measurement of Treatment and Response Variables in Psychotherapy

Lacey (1958) found a great deal of stability within individuals over time to their physiological response pattern to stress but no consistent general patterning of responses across individuals. Physiological measures of anxiety may be highly correlated within individuals over time despite the well-established finding of low correlations across individuals (Chambers, Hopkins, & Hopkins, 1968).

Mowrer, Light, Luria, and Seleny (1953) studied tension changes during psychotherapy using a variety of indicators over time. Their objective was to validate simple, inexpensive measures of psychotherapeutic process. They wrote: "It ought to be a maxim of all scientific inquiry first to get information by simple means before employing indirect and complicated approaches [p. 362]." Using a 5-point scale, patients rated their feelings of tension and happiness before and after each session. Mowrer et al. found that self-ratings of within-session tension and happiness changes over time discriminated between persons staying in and leaving psychotherapy. They also discovered high relationships within subjects between palmar sweating and self-report measures of tension both for subjects terminating and staying in psychotherapy.

Budzynski, Stoyva, and Adler (1970) found a similar high relationship between patients' self-ratings of severity of headache over time on a five-point scale and electromyographic recordings within patients over time. These findings are in sharp contrast to the low correlations between physiological and self-report measures across subjects (Chambers et al., 1968).

If these findings have general validity, it may be feasible to develop simple, inexpensive measurement operations of high utility to evaluate the course of treatment within subjects. This notion would be important to the evaluation of psychotherapy by the practitioner. It would facilitate the linking of the practice of psychotherapy with the investigation of psychotherapy.

To summarize, it is possible that the use of time-series analysis in N-of-one research in psychotherapy may help in outcome research, in process research, in the use of appropriate statistical techniques to assess the effects of interventions, and in the design of simple measurement operations with high utility. There is one further possible gain. Research in psychotherapeutic outcome and process has often been split from the practice of psychotherapy. Practicing psychotherapists obtain little from research in psychotherapy. Psychotherapy is a complex process which has eluded definition. At best, outcome studies which globally evaluate the impact of psychotherapy and show remarkable success convey the message, "Go thou and do likewise." Time-series analysis is a methodology which can unify research and treatment and can provide the therapist with feedback on specific interventions during the course of treatment. By monitoring a patient over time and by monitoring treatment variables simultaneously, the therapist receives feedback which aids the process of therapy and also generates hypotheses for future testing. Time-series designs can offer a variety of control options for testing the hypotheses generated by continuous monitoring.

Experimental Control Options in Time-Series Designs

Series of observations of a variable across time are important data in many fields of inquiry. Each of the sciences utilizing time-series data encounters a different range of experimental control possibilities. In astronomy, for example, where no control over most experimental units is possible, the astronomer tries to decompose incoming electromagnetic radiation from stars into components which reveal the structure of the stars. In economics, more control is possible for both micro- and macrosystems. The economist also tries to decompose time series such as the Dow Jones Industrial Average into components which generate models for economic systems; for example, he is interested in determining periods of business cycles of major importance. The economist also determines the effects of various classes of events over which he has little control (strikes, wars, natural catastrophes) as well as the effects of planned interventions (changes in interest rates, wage and price freezes, governmental spending). The geophysicist studying the structure of the earth may exert control by detonating explosives, examining time-series seismographic recordings, and decomposing them on the basis of preestablished response characteristics of various geological layers he assumes to be there.

The objective in all these sciences is to simulate the natural system with maximum simplicity and minimum error. It is fortunate that time-series data have found use in sciences which span a wide range of control possibilities. The psychotherapist's activities in assessment and treatment implicitly contain objectives which simultaneously resemble those of the astronomer, the economist, and the geophysicist. Like the astronomer, the psychotherapist is interested in describing the variability in the patient's behavior over time. Like the astronomer, he is interested in making predictions and forecasts based upon past behavior. The therapist, like the economist, is interested in describing the effect of various classes of events in the patient's life and in generating hypotheses about change. He is also interested in assessing the effects of planned interventions on the patient's behavior. Finally, like the geophysicist, the psychotherapist is interested in constructing theoretical models which parsimoniously describe the system of psychotherapy.

The uses of time-series data can be classified along a dimension of experimental control options which permit varying degrees of causal inference.

Variation and Concomitant Variation

The first level of causal inference comes from describing the fluctuations of a system, its cycles, and trends which permits one to forecast[4] the future values of the series (see Box & Jenkins, 1970). The idea in forecasting is that to predict the future, something must be assumed to remain stationary. Perhaps not the original data but some function of it is assumed to remain constant. This first step in describing the variation of a series leads to a search for other series which vary concomitantly.

It is possible to learn a great deal about the system generating the series by searching for concomitant variates. Edwards and Cronbach (1952) discussed Fisher's (1921) analysis of yield of wheat in bushels per acre. They wrote:

> [Fisher] found that after he controlled variety, and fertilizer, there was considerable variation from year

[4] For computer programs which use time-series data to identify and fit models and forecast future values of the series using Box-Jenkins procedures, write to James R. Taylor, Project Administrator, University of Wisconsin, A National Program Library and Inventory Service for the Social Sciences, Room 4430, Social Science Building, Madison, Wisconsin 53706. It is important to note that, unfortunately, confidence intervals for forecasts diverge as the forecast departs from the data; forecasts are usually not projected far into the future for this reason and are continually updated as new observations are added to the series. See Box and Jenkins (1970).

to year. This variation had a slow up-and-down cycle over a seventy year period. Now Fisher set himself on the trail of the residual variation. First he studied wheat records from other sections to see if they had the trend; they did not. He considered and ruled out rainfall as an explanation. Then he started reading the records of the plots and found weeds a possible factor. He considered the nature of each species of weed and found that the response of specific weed varieties to rainfall and cultivation accounted for much of the cycle. But the large trends were not explained until he showed that the upsurge of weeds after 1875 coincided with a school-attendance act which removed cheap labor from the fields, and that another cycle coincided with the retirement of a superintendent who made weed removal his personal concern [p. 64].

Economists search for connections between time series in the hope of finding "lead indicators." A lead indicator is a series whose fluctuations are predictors of the fluctuations of another series; for example, wholesale prices are a lead indicator of retail prices. One way of studying the concomitant variation between series is by the use of a technique called "transfer functions." A transfer function is a linear equation which relates the past of one time series to the present or future of another time series.

Suppose that we have two time-series processes, the treatment process, X_t, called the *input*, and the response process, Y_t, called the *output*. If X_t is a lead indicator of Y_t, then Y_t ought to be predictable statistically from a weighted sum of the previous X_ts. If X_t leads Y_t by b time units, we ought to be able to predict Y_t from X_t, b units in advance. An equation to summarize this would be as follows:

$$Y_t = v_0 X_{t-b} + v_1 X_{t-b-1} + \cdots + v_R X_{t-b-R} + N_t,$$

where N_t represents the lack of exact predictability of Y from X and is assumed to be a random variable independent of X.

The analysis proceeds by defining a function called the *cross-correlation* function as the correlation between the two time series lagged a number of time units, k, $k=1, 2, 3, \ldots$. Using this function, we can calculate best estimates for the vs in the equation above. The solution for these weights is analogous to the solution of a regression problem.[5]

[5] Detailed statistical formulae and computer programs for this and other analyses discussed in this paper can be obtained by writing to the author.

To summarize, the first level of causal inference in time series begins by analyzing the variation and concomitant variation in the data. Concomitant variation may provide insight into possible causal connections which account for variations in the series. This can be studied by the use of *transfer functions.*

By studying the variation and concomitant variation, hypotheses can be generated about causal connection. At this point in the development of the social sciences, it is wise to consider this stage of causal inference at the level of exploratory data analysis. These analyses yield the most information in the negative case. If two series are uncorrelated, it is unlikely that they are causally connected. It is true that concomitant variation does not imply causation, but it is also true that neither does anything else. Causal connection is never demonstrated; rather, we successively eliminate rival hypotheses which militate against confidence in causal connection.

A useful method for generating hypotheses for planned interventions is now discussed.

Generating Post Hoc Hypotheses: The Annotated Time-Series Record

Hypotheses can be generated post hoc by using an historical log of events assumed to be causal in nature and scanning for shifts in the time series. Granger and Hatanaka (1964) analyzed the responses of price indices to strikes, wars, and major nonrecurring catastrophes (such as the stock market crash).

In a program for potential high school dropouts, teachers assumed they had made major breakthroughs in treatment following a long, emotional, intimate talk with a student. There were three instances of this event. Time-series analysis on three behavioral indicators, however, showed that each of the students significantly avoided the teacher subsequent to the talk, did not improve in academic performance, and significantly increased classroom participation in a disruptive manner (Gottman, 1971).

The third level of causal inference involves the use of planned interventions with various control options which permit the elimination of rival hypotheses. This paper limits itself to the discussion of two useful designs: (*a*) the

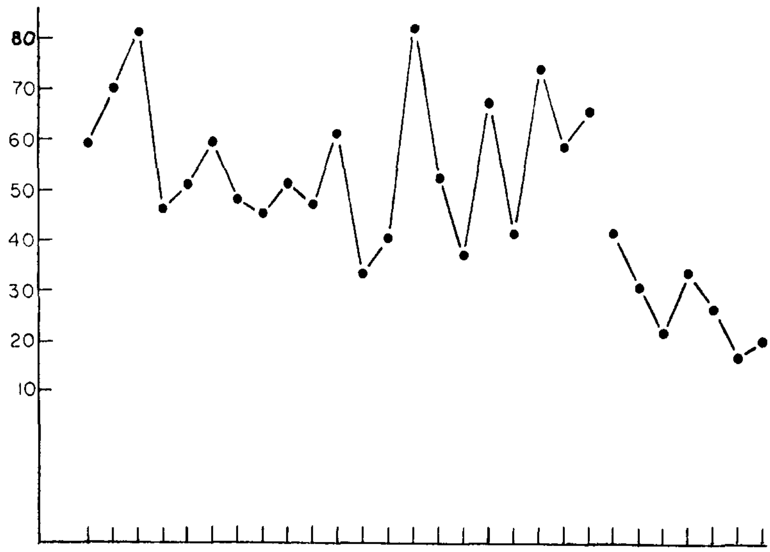

FIG. 1. Interrupted time-series experiment treating the hyperactivity of a four-year-old child using social reinforcement. (Reprinted from an article by K. E. Allen, L. B. Henke, F. R. Harris, D. M. Baer, and N. J. Reynolds published in the August 1967 issue of the *Journal of Educational Psychology*. Copyrighted by the American Psychological Association, Inc., 1967.)

interrupted time-series design and (*b*) the time-lagged control design.

Interrupted Time-Series Design

This design consists of a sequence of observations called a "base period" followed by an intervention, *I*, followed by another sequence of observations called an "experimental period." The interrupted time-series design can function as a quasi-experimental design for planned interventions in a total treatment program when a control group is not feasible. This design is an extension of the one group pretest–posttest design which is not feasible for single-subject research because it does not permit any statistical test of the hypotheses of change due to the interventions. While it is possible to develop a statistical test of the hypothesis for the interrupted time-series design, many other change-producing events may have occurred in addition to the therapist's interventions. At best, such rival hypotheses can be minimized in this design by randomly selecting the time to intervene following a sufficient number of base period observations.[6]

Figure 1 is an illustration of a simple interrupted time-series experiment treating the hyperactivity of a four-year-old child using social reinforcement (Allen, Henke, Harris, Baer, & Reynolds, 1967). For the last seven days the child was given verbal, social reinforcement for attending to a single activity for more than one minute. The statistical analysis of this experiment is discussed in the analysis section of this paper.

Time-Lagged Control Design

This design was discussed by Gottman, McFall, and Barnett (1969). An intervention is applied to one subject after a base period but withheld temporarily from a second subject. After an experimental period, the intervention is applied to the second subject and both subjects monitored for a second experi-

[6] If one is limited by cost considerations to take N observations, Glass et al. (1973) showed mathematically that the optimal intervention point is at $t = N/2$.

mental period. This design provides a powerful and highly usable design. It also ameliorates the ethical problem of withholding treatment from a patient in need. Treatment is withheld only for a time and can be administered once it has been shown to work.

Hilgard (1933) presented a time-lagged control study of two twins showing the effects of intensive training in digit memory beyond maturational trends evident in the base period.

Summary of Experimental Control Options

Two levels of causal inference in time-series analysis have been discussed. The first level relates to weak causal inferences and constitutes exploratory data analysis (Tukey, 1970). It consists of (*a*) examining concomitant variation, (*b*) examining lead indicators using transfer function, and (*c*) generating hypotheses for intervention using an annotated record and scanning for shifts post hoc. The second level relates to strong causal inference. This paper described two such planned intervention experiments: (*a*) the interrupted time-series experiments for *N*-of-one research and (*b*) the time-lagged control for *N*-of-two research.

Statistical Analysis of Time-Series Data

This section is divided into three parts. First, the construction of time-series models and its relationship to problems of statistical dependence (typical of data over time) is discussed. Second, the statistical analysis of the effects of interventions is discussed. Finally, illustrations of the analysis for shifts in level and slope are presented. This analysis is based upon a particular time-series model which has proven extremely useful in practice. The model is analogous to the straight line in regression analysis and is called the *Integrated moving average model of first order* (IMA [1,1]) *with deterministic drift.* Model building has been discussed in more detail by Box and Jenkins (1970), Gottman et al. (1969), and Glass et al. (1973). The reader is referred to these sources for a discussion of how the IMA (1,1) model is derived mathematically. This paper limits itself to a graphical discussion of two time-series models. Details of the statistics are not presented in the body of this paper in the hope that the nonmathematical reader will see that outputs of the statistical analyses are familiar t ratios (see Footnote 5). Computer programs are available upon request (see Footnote 2).

Models

If the data were assumed to be independent observations sampled from a normally distributed population, then the problem of assessing the effects of an intervention is simply solved. Shewart (1931) proposed the use of industrial quality control charts to deal with this problem.

In this case, a two standard deviation band is drawn above and below the mean in the base period (after the removal of trend from the data). If the data in the experimental period drift outside this band, a significant shift at the .05 alpha level is recorded. The data are assumed to be normally distributed.

However, most time series do not consist of independent observations. The tth observation, z_t, is usually predictable from previous observations.

A series is called *stationary* if it varies about a fixed mean and *nonstationary* if it drifts away from any mean level. In the nonstationary case, the mean has no utility as a statistic to predict where the series will be. The IMA (1,1) model drifts at a constant slope but drops or rises randomly to a new level (see Figure 2a).

The IMA (2,2) model is considerably more general and shows random components in both slope and level. The IMA (2,2) model was discussed by Glass et al. (1973). A more general computer program is available which incorporates both IMA (1,1) and IMA (2,2) models.

Assessing Shifts in the Interrupted Time-Series Design

Glass and Maguire (1968) proposed the use of the IMA (1,1) model which incorporates a slope or deterministic drift parameter. Applied to the interrupted time-series design, the n_1 base period observations are repre-

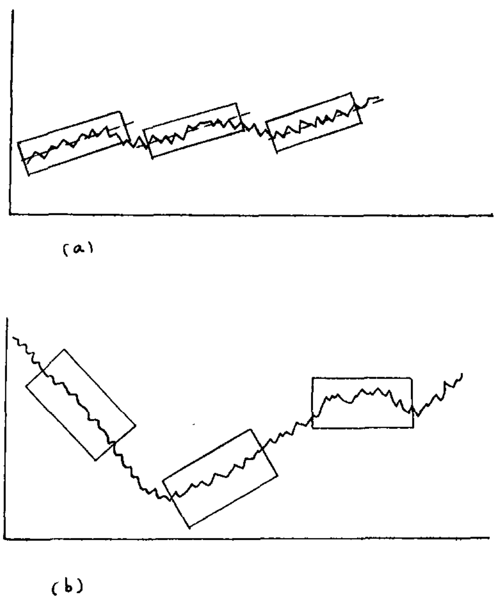

FIG. 2. Graphical illustration of two time-series models: (*a*) integrated moving average model of first order (IMA [1,1]) showing stationarity in slope but nonstationarity in level and (*b*) integrated moving average model of second order (IMA [2,2]) showing nonstationarity in both slope and level.

sented by

$$z_1 = L + b_1$$
$$z_2 = L + \gamma b_1 + b_2$$
$$z_3 = L + \gamma(b_1 + b_2) + b_3$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$z_t = L + \gamma \sum_{i=1}^{t-1} b_i + b_t. \qquad [1]$$

In Equation 1, L is a fixed but unknown initial location parameter, γ is a parameter descriptive of the degree of interdependence of the observations and takes values from 0 to 2, and b_t is a normal random variable with mean μ and variance σ^2.

The linear trend or drift in this model comes about by writing b as $\mu + a$, where a is a normal deviate with mean zero and variance σ^2. Equation 1 now becomes

$$z_t = L + \mu\gamma(t-1) + \mu + \gamma \sum_{i=1}^{t-1} a_i + a_t,$$

and we can see that the series will drift $\mu\gamma t$ units by time t.

For the $n_2 = N - n_1$, observations following the intervention I, the series is represented by the equation,

$$z_t = L + \delta + \gamma \sum_{i=1}^{t-1} b_i + b_t,$$

where δ is the change in level of the time series due to I.

Glass and Maguire's analysis consists of transforming the model into the general linear model. The general linear model has a simple least-squares solution which provides estimators for L and δ (once γ is known) which are distributed as Student's t with $N-3$ degrees of freedom when divided by appropriate error term. Although γ is not usually known, the procedure is to vary γ between 0 and 2 in increments of .01 and to calculate the likelihood function, $h(\gamma/z)$, for γ, given the observed time series. This function is a maximum when that value of γ is found which makes the sum of the squared residuals a minimum. The specific steps in this transformation are not presented here (see Footnote 5). Output from this analysis involves a graph of Student's t for shift in level and a graph of Student's t for shift in slope as a function of various levels of gamma. The user simply finds that value of gamma which maximizes the likelihood function and then looks up the values of t change in level and t change in slope for that value of gamma.

Illustrative Analysis: Shifts in Level

This example discusses only an analysis of the shifts in level for purposes of illustration. The data presented in Figure 1 on the treatment of a hyperactive four-year-old child are now analyzed. The question is, Has the intervention been effective in producing decreased activity change?

Figure 3 shows that the likelihood function for these data is a maximum when $\gamma = 0$. It would be conservative to take the range of reasonable values of γ to be from 0 to about .10. The t statistic for these values is clearly significant since its value is never greater than -4.4 for this range of γ. Glass and Maguire (1968) also showed that these data conform to the first order IMA (1,1).

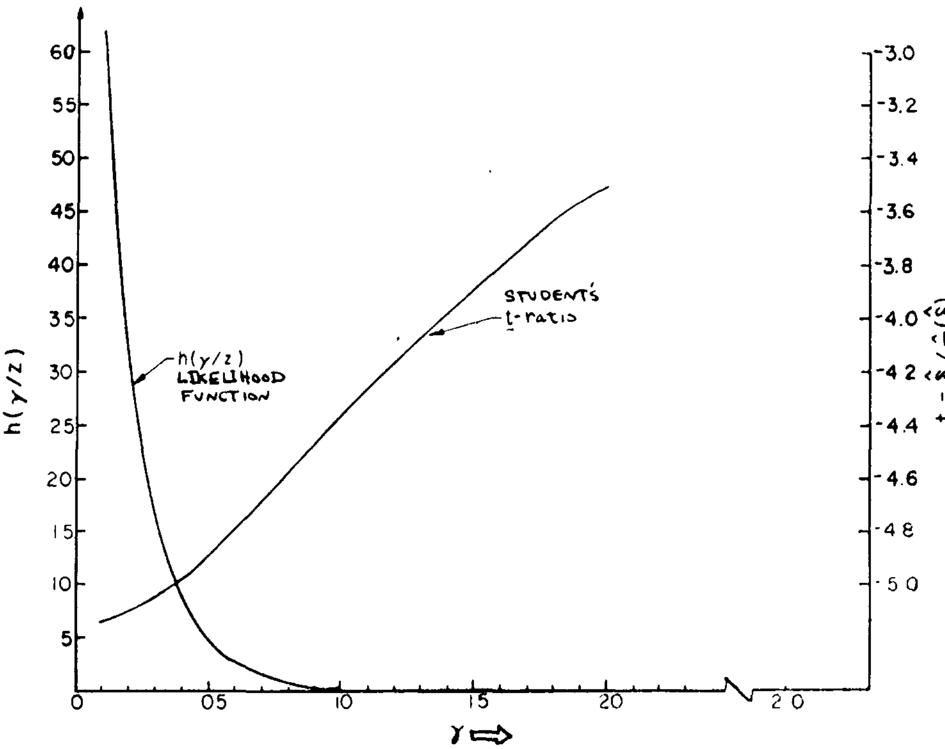

FIG. 3. Analysis of Allen et al. (1967) data (see Figure 1) showing the likelihood function, $h(\gamma/z)$, and Student's t ratio as a function. (From G. V. Glass and T. D. Maguire, 1968, and reproduced by permission.)

Illustrative Analysis: Shifts in Level and Slope

Gottman and Asher (1972) studied the effects of an intervention on the playground to modify peer social interaction in the classroom. This experiment is described in this paper because their data illustrate the interpretation of time-series data.

Method. In a third-grade classroom, 5 subjects were identified by the teacher as problem children and 5 subjects were selected at random from a class of 27 students.

An observer recorded the frequency of the following behaviors: (*a*) alone and working, (*b*) alone and not working (daydreaming, withdrawn), (*c*) positive verbal interacting with peers, (*d*) hitting, pushing, sparring with peers, (*e*) negative verbal interacting with peers, and (*f*) interacting with teacher. The observer checked the appropriate behavior every 3 seconds, scanning down a randomized list of names of the 10 subjects.[7] An acceptable interobserver agreement of 82.5% was achieved using this taxonomy of interactive behavior.

The teacher's objective was to increase positive verbal interaction and to decrease hitting and negative verbal interaction in the classroom. The children played a game on the playground they called "smear." In this game, one student got the ball; it was his job to hold on to the ball and everyone else's job to get the ball away, at any cost. It was felt that this game might be modified to increase cooperative play. It was hoped that learnings would generalize to the classroom. Students suggested a voluntary game which involved passing a ball as often as possible, a game analogous to an old game called "hot potato." Noncooperative players were not permitted to play. At first the teacher coached the children and then the children began coaching each other. The game became popular and it was the teacher's impression that there was considerable carry-over of good feelings to the classroom.

[7] An observer rated behaviors three times a week for about an hour each day. A student was located in the classroom, the appropriate behavior checked, and then the next student on the list was located. Names were randomly assigned to position on the list. After the list was scanned once, the procedure was repeated approximately 20 times. Five observations were summed so that each variable ranged 0–5; approximately 4 points were plotted each day.

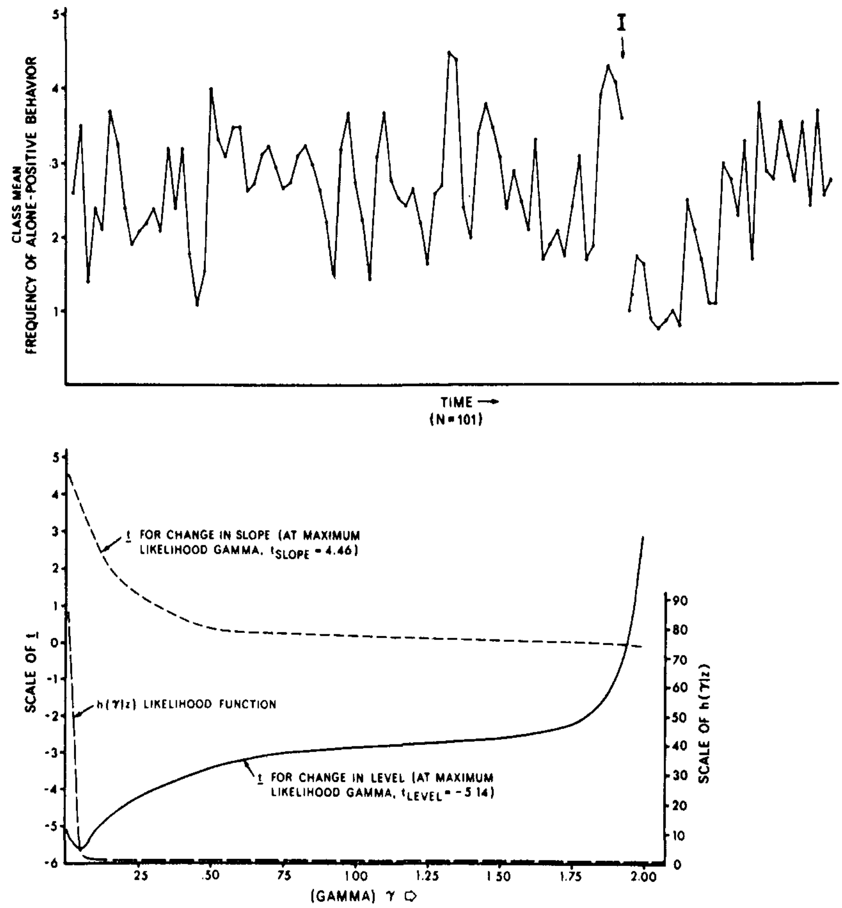

FIG. 4. Time-series experiment in peer interaction showing class means for frequency of alone-positive behavior and shifts in level and slope. (Data indicate an ephemeral effect.)

Results. The analysis of the results of this intervention have been presented in some detail by Gottman and Asher (1972). For the purpose of illustration we discuss only the results for the variable alone and working ("alone positive"). This variable is a good indicator of the effects of the intervention on increasing interaction time. As time spent alone decreased, time spent interacting with peers increased. Figure 4 shows a sharp drop in time spent alone (t level $= -5.14$, $p < .01$). The figure also indicates a return to base indicating that the effect was only temporarily effective. This return to base is indicated by a significant increase in slope after the initial drop in level (t slope $= 4.46$, $p < .01$).

SUMMARY

This paper has discussed the advantages of applying time-series analysis to the measurement of change over time for single-subject research in psychotherapy. It has presented a framework for assessing causal connection with weak and strong inference. Specific attention has been given to assessing the effects of an intervention using the interrupted time-series design.

Single-subject research in psychotherapy has advantages in outcome research, process research, measurement, and linking the practice of psychotherapy with the investigation of psychotherapy. Time-series analysis permits the study of the single subject over time and the study of effect patterns of an intervention over time. It also allows the therapist to use information as feedback for making decisions, a useful tool in the evaluation of psychotherapy. Used with larger samples as "N-of-one-at-a-time" research, time-series methods permit appropriate generalization of an effect to a population without the errors which may result when data are averaged.

REFERENCES

ALLEN, K. E., HENKE, L. B., HARRIS, F. R., BAER, D. M., & REYNOLDS, N. J. Control of hyperactivity by social reinforcement of attending behavior. *Journal of Educational Psychology*, 1967, **58**, 231–237.

BAKAN, D. *On method.* San Francisco: Jossey-Bass, 1967.

BANDURA, A. *Principles of behavior modification.* New York: Holt, Rinehart & Winston, 1969.

BARRON, F., & LEARY, T. F. Changes in psychoneurotic patients with and without psychotherapy. *Journal of Consulting Psychology*, 1955, **19**, 239–245.

BERGIN, A. E. Some implications of psychotherapy research for therapeutic practice. In G. E. Stollak, B. G. Guerney, & M. Rothberg (Eds.), *Psychotherapy research: Selected readings.* Chicago: Rand McNally, 1966.

BOX, G. E. P., & DRAPER, N. R. *Evolutionary operation.* New York: Wiley, 1969.

BOX, G. E. P., & JENKINS, G. M. *Time-series analysis: Forecasting and control.* San Francisco: Holden Day, 1970.

BUDZYNSKI, T., STOYVA, J., & ADLER, C. Feedback-induced muscle relaxation applied to tension headache. *Journal of Behavior Therapy and Experimental Psychiatry*, 1970, **1**, 205–211.

CAMPBELL, D. T., & STANLEY, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally, 1970.

CARTWRIGHT, D. S. Note on "Changes in psychoneurotic patients with and without psychotherapy." *Journal of Consulting Psychology*, 1956, **20**, 403–404.

CHAMBERS, A. C., HOPKINS, K. D., & HOPKINS, B. R. Anxiety, physiologically and psychologically measured: Its effect on mental test performance. Unpublished report, Laboratory of Educational Research, Boulder, Colorado, 1968.

CHASSAN, J. B. *Research designs in clinical psychology and psychiatry.* New York: Appleton-Century-Crofts, 1967.

DUKES, W. F. $N=1$. *Psychological Bulletin*, 1965, **64**, 74–79.

EDWARDS, A. L., & CRONBACH, L. J. Experimental design for research in psychotherapy. *Journal of Clinical Psychology*, 1952, **8**, 51–59.

FISHER, R. A. Studies in crop variation. *Journal of Agricultural Science*, 1921, **11** (Pt. 2), 8–35.

GASTWIRTH, J. L., & COHEN, M. L. Small sample behavior of robust linear estimators of location. *Journal of the American Statistical Association*, 1970, **65**, 946–973.

GLASS, G. V., & MAGUIRE, T. D. Analysis of time-series quasi-experiments. Unpublished report, Laboratory of Educational Research, Boulder, Colorado, 1968.

GLASS, G. V., WILLSON, V. K., & GOTTMAN, J. M. *The design and analysis of time-series experiments.* Boulder, Colo.: Laboratory of Educational Research Press, 1973.

GOTTMAN, J. M. Time-series analysis in the behavioral sciences and a methodology for action research. Unpublished doctoral dissertation, University of Wisconsin, 1971.

GOTTMAN, J. M., & ASHER, S. R. Modification of peer communication. Unpublished report to Madison Public Schools by the Instructional Research Laboratory, Madison, Wisconsin, 1972.

GOTTMAN, J. M., MCFALL, R. M., & BARNETT, J. T. Design and analysis of research using time series. *Psychological Bulletin*, 1969, **72**, 299–306.

GRANGER, C. W. J., & HATANAKA, M. *Spectral analysis of economic time series.* Princeton: Princeton University Press, 1964.

HILGARD, J. R. The effect of delayed practice on memory and motor performances studied by the method of co-twin control. *Genetic Psychology Monographs*, 1933, **6**, 67.

HOLTZMAN, W. Statistical models for the study of change in the single case. In C. Harris (Ed.), *Problems in measuring change.* Madison: University of Wisconsin Press, 1967.

LACEY, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy.* Vol. 1. Washington, D.C.: American Psychological Association, 1958.

MOWRER, O. H., LIGHT, D. H., LURIA, Z., & SELENY, M. P. Tension changes during psychotherapy. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research.* New York: Ronald Press, 1953.

PARKER, J. F., & FLEISHMAN, E. A. Ability factors and component performance measures as predictors of complex tracking behavior. *Psychological Monographs*, 1960, **74** (16, Whole No. 503).

SCHOFIELD, W. Changes in responses to the Minnesota Multiphasic Inventory following certain therapies. *Psychological Monographs*, 1950, **64** (5, Whole No. 311).

SHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

SHEWART, W. A. *Economic control of quality of manufactured product.* New York: Van Nostrand, 1931.

SIDMAN, M. A note on functional relations obtained from group data. *Psychological Bulletin*, 1952, **49**, 263–269.

TUKEY, J. W. *Exploratory data analysis.* Reading, Mass.: Addison-Wesley, 1970.

TYLER, V. D., & BROWN, G. D. Token reinforcement of academic performance with institutionalized delinquent boys. *Journal of Educational Psychology*, 1968, **59**, 164–168.

(Received June 21, 1972)