Journal of Consulting and Clinical Psychology
1993, Vol. 61, No. 6, 907-910

0022-006X/93/$3.00

The Analysis of Change: Issues, Fallacies, and New Ideas

John M. Gottman and Regina H. Rushe

In this special section of *The Journal of Consulting and Clinical Psychology,* new ideas about how to analyze change are presented in a format that is accessible to clinicians and clinical researchers. Rogosa's (1988) myths of longitudinal research are reviewed in an attempt to familiarize psychologists with the dangers of assuming (a) that regression toward the mean is unavoidable, (b) that difference scores are unreliable, (c) that analysis of covariance is the way to analyze change, (d) that two points are adequate to measure change, and (e) that the correlation between change and initial level is always negative. An overview of the articles emphasizes what is new and improved in the design and analysis of change. The articles are preceded with a conceptual discussion of how to measure change over time when the stability of the criterion construct is high and there is little variance to predict. Other articles discuss the form of change over time and how this can be an important tool in testing specific hypotheses. Individual change over time can be described with short time-series analysis or sequential analysis of continuous data. Individual and group change over time can be described in survival analyses or cohort-sequential designs. Some articles minimize the problems of cohort-sequential designs by including cohorts of overlapping ages and comparing hierarchical models of change. Discrete-time survival analyses have intuitive appeal, can include several types of predictors in the models, and are relatively simple to compute.

The general problems associated with the analysis of change have long been considered crucial to the readership of this journal. In fact, in 1991, Newman and Howard edited a special section of this journal on new clinical research methods, and many of the articles of this special section were also concerned with the analysis of change. Recognizing the importance of this contribution, we now consider the problems associated with the study of change.

We begin this introduction with some historical perspective. In 1967, an edited book titled *Problems in Measuring Change* appeared (Harris, 1967). This book raised a number of important questions about the measurement and analysis of change, and it also suggested some new techniques such as time-series analysis of change within single subjects (the chapters by Campbell and Holtzman). Since 1967, there have been many analytic and conceptual developments, and a number of recently published volumes reflect some of these (e.g., Collins & Horn, 1991; Rovine & Von Eye, 1991; Von Eye, 1990a, 1990b). In our view, many of the new ideas summarized in these volumes are, unfortunately, largely unknown to some of our most respected researchers and journal reviewers. The lack of general dissemination of these new concepts has retarded progress on the study of change in clinical populations. It has also led reviewers to rely on reified timeworn cherished beliefs for what appears to constitute state-of-the-art sound research practice in the study of change. Unfortunately, not through any planning or design, many of these cherished beliefs are based on little more than fallacies.

John M. Gottman and Regina H. Rushe, Department of Psychology, University of Washington.

Correspondence concerning this article should be addressed to John M. Gottman, Department of Psychology, NI-25, University of Washington, Seattle, Washington 98195.

In this book, Rogosa's (1988) chapter, "Myths About Longitudinal Research," addressed a number of these crucial fallacies about the analysis of change. Unfortunately, because Rogosa's chapter is not widely known to clinical psychologists and to clinical researchers, we will briefly summarize some of it here. Rogosa's chapter was concerned with methods for the analysis of longitudinal data. He contended that longitudinal research in the behavioral and social sciences has been dominated, for the past 50 years or more, by a collection of related damaging myths and fallacies. He argued that the development and application of useful methods for the analysis of longitudinal data have been impeded by these myths. Using a very simple model of linear change over time, he challenged a number of cherished beliefs about change and showed that they were fallacies. Let us consider some of these cherished beliefs here.

Fallacy 1: In Change, Regression Toward the Mean Is an Unavoidable Law of Nature

On the contrary, Rogosa showed that regression toward the mean is not a law of nature but a mathematical tautology that comes from using the standard deviation as a metric of change. However, if the nonstandardized metric is used, regression toward the mean need not occur. Regression toward the mean implies that prescores far from the mean on either side of the mean will move in toward the mean on postmeasurement, so that prescores "squeeze in" toward the mean at postassessment. However, in many applications, one obtains instead a "fanning out" effect, which is the opposite of regression toward the mean.

Fallacy 2: The Difference Score Between Pre- and Postmeasurement Is Unreliable

Contrary to Cronbach and Furby's (1970) article, Rogosa showed that the difference score is not intrinsically unreliable.

On the contrary, under a wide range of (surprisingly) moderate test-retest reliabilities, the change score itself is quite reliable. Rogosa's discussion supported Lord's (1967) earlier landmark article, which had pointed out that the change score is the most natural metric of change and that the residual from an analysis of covariance is often confusing. Lord had written, "some people assert that deviation from the regression line is the real measure of change and that the ordinary difference between initial and final measurements is not a measure of change. This can hardly be correct. If certain individuals gained 300 ounces, this is a definite fact, not a result of an improper definition of growth" (p. 23).

This leads us to the second most common mistaken belief about the proper way to assess change, which is the analysis of covariance (and derivative methods such as path analysis) in which the prescore is the covariate.

Fallacy 3: Analysis of Covariance (ANCOVA—or Related Methods Such as Path Analysis) Is the Way to Analyze Change

Rogosa showed that residual change curves do not solve problems in the description of change over time, nor are ANCOVA matrices informative about change. Rogosa reviewed Goldstein's (1979) results from a path analysis examining longitudinal reading test scores on a nationwide British sample at three ages. Goldstein fit a regression equation between Time 1 and Time 2 scores, estimating a regression coefficient $\beta 1$ and an equation estimating Time 3 reading scores from the Time 1 and Time 2 data, estimating $\beta 2$ (the regression weight of Time 2 scores) and $\beta 3$ (the regression weight of the Time 1 scores). Goldstein found that $\beta 1 = 0.841$, $\beta 2 = 1.11$, and $\beta 3 = -0.147$. Goldstein was concerned about the negative estimate for $\beta 3$ and confused about how to interpret it. However, Rogosa showed that, if one assumes only simple straight-line change functions, $\beta 3$ is actually a parameter completely independent of the data and dependent only on the times of measurement! This astounding and alarming mathematical result shows how ANCOVA (or its derivative methods) can yield completely incorrect conclusions about change.

Fallacy 4: Two Points (Pre- and Posttest) Are Adequate for the Study of Change

The next most common fallacy is that two measurement times are really all one needs for a truly descriptive longitudinal study. It is true that the pretest–posttest design is the most common design in the study of change, and two repeated observations do indeed constitute a longitudinal study. However, two observations are not adequate for studying the form of change.[1] Two observations can only estimate the amount of change. A straight line can be passed between two points, but there is no way to assess the adequacy of the line, nor is there a way to compare the line with other functional forms of change. Rogosa also showed that the amount of change can be deceptive. If the rate of growth is not constant but depends on time, the amount of change will depend crucially on the times of measurement, and observations of individuals at a different set of two times may give contradictory results. Hence, it could be misleading to characterize growth by the amount of change. Rankings among individuals can change as a function of the times of measurement, and this will also be true of correlations of other variables with change scores. Thus, in many respects, the pretest–posttest design may be limited for the study of individual change and individual differences in change.

Fallacy 5: The Correlation Between Change and Initial Level Is Always Negative

A negative correlation between true change and true initial level is best known as the Law of Initial Values (Lacey & Lacey, 1962; Wilder, 1957).[2] It is also referred to as the "ceiling effect," because the idea is that change will be small if the initial level is high. The negative correlation is also related to regression toward the mean. Rogosa pointed out that other correlations have also been discussed. He wrote,

> "a zero correlation between change in initial status is known as the Overlap Hypothesis, which dates back to Anderson (1939) and was prominent in Bloom (1964). One interpretation of the Overlap Hypothesis is that growth occurs via independent increments (similar to the formulation of simplex models in Humphreys, 1960). A positive correlation between change and initial status corresponds to 'fanspread' where variances increase over time. The positive correlation can be described as 'them that has, gets' " (p. 182).

It turns out that the correlation between true change and true initial level depends crucially on the choice of the time when the initial level was measured. If we assume a straight line set of growth curves, the correlation between true change and true initial level will depend on when the initial level was taken. Both in psychophysiology and in studies of academic growth, widely different estimates of the correlation between true change and true initial level are obtained (see Cacioppo & Petty, 1983; Cacioppo & Tassinary, 1990; Rogosa, 1988).[3]

Overview of the Articles in This Special Section

Given that these cherished beliefs about the study of change are actually fallacies, how are we to think about change, and

---

[1] There is a sociological tradition that ignores individual differences (Coleman, 1968) and assumes that the parameters of the change function over time are constant across subjects; the emphasis here is on mean change instead of individual change. This sociological tradition makes it possible to estimate complex change curves even if the data are pretest–posttest. Data from many subjects are combined into a single change curve to estimate complex change curves from only two observations per subject (see Nielson & Rosenfeld, 1981; Salemi & Tauchen, 1982; Tuma & Hannan, 1984).

[2] Rogosa noted that, with fallible scores, the correlation between observed change and observed initial status is a poor estimate of the correlation between true change and true initial status. The estimate is negatively biased in addition to the attenuation. Rogosa wrote, "thus, because of the poor properties of this estimate, negative correlations between observed change and observed initial status are often obtained when the true-score correlation is zero or positive. The myth is stated and discussed in terms of true scores because these are of primary substantive interest; although of less interest, a similar dependence on time of initial status also holds for the observed score correlation" (p. 183).

[3] See also Figure 2 of Patterson's (1993) article, in which the fanning-out effect is also observed in antisocial behavior.

how are we to correctly analyze change? Fortunately, there are now a number of well-worked-out solutions to the problems of measuring change. The articles of this special section were chosen to highlight currently available powerful techniques for studying change over time, both within and across individuals, in which we have repeated measures over time and in which we are interested in terms of individual differences in change. In this special section, we tried to solve two problems that have existed with many previous attempts to communicate these concepts. First, the new methodological approaches seemed only remotely related to clinical issues. Second, they seemed accessible only to the most sophisticated mathematical statisticians, not to ordinary working researchers and clinicians. To deal with these two problems, whenever possible, we tried to have articles that served two functions: an illustration of the approach to clinically relevant kinds of data and a practical introductory primer for using the approach.

There are two kinds of issues addressed by these articles, those represented by a single-subject approach and those represented by more common repeated measures designs that use many subjects or groups of subjects.

Designs That Use Multiple Subjects and Multiple Groups

The second group of articles concerned with the study of change over time and individuals or across groups. Patterson's article begins with the conceptual and analytic problems of understanding the factors that contribute to change over time when the stability of the clinically significant criterion construct is high. He notes the problems with standard approaches to this problem and shows how these research problems can be turned into an opportunity for exploring the data. He then shows how one can explore the data to discover alternative trajectories toward the clinically significant criterion construct (e.g., in his article, antisocial behavior). Stoolmiller, Duncan, Bank, and Patterson's (1993) article on the study of change in psychotherapy and patterns of client resistance illustrate the theoretical importance of the form of the growth function over time in testing different models of client resistance, such as the "working through" model. The cohort-sequential design offers a most promising short-term alternative for a truly longitudinal design because approaching clinical problems with long-term longitudinal designs is difficult and expensive. Anderson's (1993) article describes a project with the cohort-sequential design and illustrates the power of the approach with an important data set. Externalizing behavior and negativity in parent–child relationships is described in terms of growth, individual differences in growth, and factors that affect the rate and pattern of growth. Anderson also notes the challenges in approximating longitudinal data through the use of overlapping cohorts. Raudenbush and Chan (1993) also discuss cohort-sequential designs and the many new problems posed in estimation and the potential lack of overlap of different cohorts. They offer a general approach to these problems with hierarchical linear modeling in an overlapping cohort design. An interesting example with data about attitudes toward deviance in adolescence is provided. After a theoretical model for individual change is formulated with a unique set of parameters, multiple models are fit to the data to describe change, compare the adequacy of linear and curvilinear growth models, estimate the correlation between initial value and rate of change, and investigate the effects of time-variant and time-invariant covariates. Willett and Singer's (1993) article on discrete-time survival analysis is a review of literature from the *Journal of Consulting and Clinical Psychology* as well as a primer on the subject. Survival analysis promises to describe the onset, recovery, relapse, and termination of clinical problems in a quantitive form that is useful to clinicians and applied researchers. Both independent and dependent variables may fluctuate over time and still be included in discrete-time models with standard logistic regression software.

Single-Case Approaches to Outcome and Process Research

Crosbie's (1993) article presents a new, powerful method for single-case analysis of change over time using the interrupted time-series design; the article shows how this can be done without needing to know sophisticated time-series modeling methods and with very few data before and after the intervention. This approach will have widespread importance in the evaluation of change in patients in clinical trials where it is possible to study people on a case-by-case basis, or in the case work of quantitatively oriented clinical practitioners. The quantitative data can be taken from self-reports of behavior (e.g., the spouse observation checklist), ratings of the severity of target symptoms (e.g., depression, severity of headaches), and so on, limited only by the imagination of the therapist. As such, Crosbie's article is a new approach to single-case outcome research that makes time-series methods available to the general clinician for the first time. Gardner's (1993) article is an equivalently powerful and elegant method for a single-case analysis of process research in treatment using sequential analysis of continuous data in dyads (such as patient and therapist). This approach makes it possible to build models for interactive structures in psychotherapy process research (or research on interpersonal relationships).

We have also selected these authors because they are all very good teachers of very complex ideas. They are also teachers who are aware of the practical problems that everyday working researchers and clinicians face in trying to design their research and analyze their data. We hope that these articles will introduce readers to ideas that really represent major breakthroughs in our ability to think about change.

References

Anderson, E. R. (1993). Analyzing change in short-term longitudinal research using cohort-sequential designs. *Journal of Consulting and Clinical Psychology, 61,* 929–940.

Anderson, J. E. (1939). The limitations of infant and preschool tests in the measurement of intelligence. *Journal of Psychology, 8,* 351–379.

Bloom, B. S. (1964). *Stability and change in human characteristics.* New York: Wiley.

Cacioppo, J. T., & Petty, R. E. (Eds.). (1983). *Social psychophysiology: A sourcebook.* New York: Guilford Press.

Cacioppo, J. T., & Tassinary, L. G. (Eds.). (1990). *Principles of psychophysiology: Physical, social, and inferential elements.* Cambridge, England: Cambridge University Press.

Campbell, D. T. (1967). From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 212–242). Madison: University of Wisconsin Press.

Coleman, J. S. D. (1968). The mathematical study of change. In H. M. Blalock & A. B. Blalock (Eds.), *Methodology in social research* (pp. 428–478). New York: McGraw-Hill.

Collins, L. M., & Horn, J. L. (1991). *Best methods for the analysis of change: Recent advances, unanswered questions, future directions.* Washington, DC: American Psychological Association.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin, 74,* 68–80.

Crosbie, J. (1993). Interrupted time-series analysis with brief single-subject data. *Journal of Consulting and Clinical Psychology, 61,* 966–974.

Gardner, W. (1993). Hierarchical continuous-time sequential analysis: A strategy for clinical research. *Journal of Consulting and Clinical Psychology, 61,* 975–983.

Goldstein, H. (1979). *The design and analysis of longitudinal studies.* San Diego, CA: Academic Press.

Harris, C. W. (Ed.). (1967). *Problems in measuring change.* Madison: University of Wisconsin Press.

Holtzman, W. H. (1967). Statistical models for the study of change in the single case. In C. W. Harris, (Ed.), *Problems in measuring change* (pp. 199–211). Madison: University of Wisconsin Press.

Humphreys, L. G. (1960). Investigations of the simplex. *Psychometrika, 25,* 313–323.

Lacey, J. I., & Lacey, B. C. (1962). The law of initial value in the longitudinal study of autonomic constitution: Reproducibility of autonomic responses and response patterns over a four year interval. In W. M. Wolf (Ed.), Rhythmic functions in the living system. *Annals of the New York Academy of Sciences, 98,* 1257–1290.

Lord, F. M. (1967). Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 199–211). Madison: University of Wisconsin Press.

Nielson, F., & Rosenfeld, R. A. (1981). Substantive interpretations of differential equation models. *American Sociological Review, 42,* 73–98.

Newman, F. L., & Howard, K. I. (1991). Introduction to the special section on seeking new clinical research methods. *Journal of Consulting and Clinical Psychology, 59,* 8–11.

Patterson, G. R. (1993). Orderly change in a stable world: The antisocial trait as a chimera. *Journal of Consulting and Clinical Psychology, 61,* 911–919.

Raudenbush, S. W., & Chan, W.-S. (1993). Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of Consulting and Clinical Psychology, 61,* 941–951.

Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171–210). New York: Springer.

Rovine, M. J., & Von Eye, A. (1991). *Applied computational statistics in longitudinal research.* San Diego, CA: Academic Press.

Salemi, M. K., & Tauchen, G. E. (1982). Estimation of nonlinear learning models. *Journal of the American Statistical Association, 77,* 725–731.

Stoolmiller, M., Duncan, T., Bank, L., & Patterson, G. R. (1993). Some problems and solutions in the study of change: Significant patterns in client resistance. *Journal of Consulting and Clinical Psychology, 61,* 920–928.

Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods.* New York: Academic Press.

VonEye, A. (Ed.). (1990a). *Statistical methods in longitudinal research: Volume I—Principles and structuring change.* Boston: Academic Press.

VonEye, A. (Ed.). (1990b). *Statistical methods in longitudinal research: Volume II—Time series and categorical longitudinal data.* Boston: Academic Press.

Wilder, J. (1957). The law of initial value in neurology and psychiatry. *Journal of Nervous and Mental Disease, 125,* 73–86.

Willett, J. B., & Singer, J. D. (1993). Investigating onset, cessation, relapse, and recovery: Why you should, and how you can, use discrete-time survival analysis to examine event occurrence. *Journal of Consulting and Clinical Psychology, 61,* 952–965.

Received January 29, 1993
Accepted January 29, 1993 ■